5 重回帰モデル

確率変数ベクトル $(Y, \boldsymbol{X})$ が次の $p+1$ 変量正規分布に従うとする．

$$\begin{pmatrix} Y \\ \boldsymbol{X} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_Y \\ \boldsymbol{\mu_X} \end{pmatrix}, \begin{pmatrix} \sigma_Y & \boldsymbol{\sigma}' \\ \boldsymbol{\sigma} & \boldsymbol{\Sigma_X} \end{pmatrix}\right)$$

このとき，$\boldsymbol{X} = \boldsymbol{x}$ を与えたときの Y の条件付き分布は，

$$Y|\boldsymbol{X} = \boldsymbol{x} \sim N\left(\mu_Y + \boldsymbol{\sigma}'\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}(\boldsymbol{x} - \boldsymbol{\mu_X}), \sigma^2\right)$$

である．ただし，

$$\sigma^2 = \sigma_Y - \boldsymbol{\sigma}'\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\boldsymbol{\sigma}$$

である．ここで，$\beta_0 = \mu_Y - \boldsymbol{\sigma}'\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}\boldsymbol{\mu_X}$，$\beta_i$ を $\boldsymbol{\sigma}'\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}$ の第 i 成分（$i = 1, \ldots, p$），$\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ とすると，

$$\begin{aligned} E(Y|\boldsymbol{X} = \boldsymbol{x}) &= \mu_Y + \boldsymbol{\sigma}'\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}(\boldsymbol{x} - \boldsymbol{\mu_X}) \\ &= \beta_0 + \boldsymbol{\beta}'\boldsymbol{x} \end{aligned}$$

である．つまり，$\boldsymbol{X}$ への Y の回帰関数は，$\beta_0 + \boldsymbol{\beta}'\boldsymbol{x}$ である．また，$\boldsymbol{X}$ を用いて Y を予測することを考えるならば，第 8 回：重相関係数の講義資料より

$$\mu_Y + \boldsymbol{\sigma}'\boldsymbol{\Sigma}_{\boldsymbol{X}}^{-1}(\boldsymbol{x} - \boldsymbol{\mu_X})$$

は Y の最良線形予測量に $\boldsymbol{X} = \boldsymbol{x}$ を代入したものであることがわかる．つまり，Y を $\boldsymbol{X} = \boldsymbol{x}$ を用いて予測する際には $\beta_0 + \boldsymbol{\beta}'\boldsymbol{x}$ の形式が良さそうであることがわかる．

$\boldsymbol{X} = \boldsymbol{x}$ を与えたとき，

$$Y = \beta_0 + \boldsymbol{\beta}'\boldsymbol{x} + \epsilon \tag{16}$$

とも表せる．ただし，$\epsilon \sim N(0, \sigma^2)$ である．式 (16) は，$\boldsymbol{X} = \boldsymbol{x}$ を与えたときに，Y は $\boldsymbol{x}$ で説明できる部分（すなわち，$\beta_0 + \boldsymbol{\beta}'\boldsymbol{x}$）と $\boldsymbol{x}$ では説明できない部分をまとめた誤差 ϵ との和で記述できるというモデル（仮説）を連想させる．これらのことから，次の重回帰モデルを考える．

定義 8. 重回帰モデルとは

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon$$

となるような Y と $(x_1, \ldots, x_p)$ の関係が成り立つことである．ただし，$E(\epsilon) = 0$，$V(\epsilon) = \sigma^2$ である[1]．

互いに独立な n 個の標本を $(Y_i, X_{i1}, \ldots, X_{ip})'$（$i = 1, \ldots, n$）とする．ただし，$n > p+1$ とする．このとき，重回帰モデルを仮定すると，$i = 1, \ldots, n$ に対して

$$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i \tag{17}$$

が成り立つ．ただし，$\{\epsilon_i\}$ は互いに無相関（$Cov(\epsilon_i, \epsilon_j) = 0,\ i \neq j$）であり，$E(\epsilon_i) = 0$，$V(\epsilon_i) = \sigma^2$ である．また，重回帰モデルでは $\{X_{ij}\}$ を非確率的変数値として扱うことから，小文字 $\{x_{ij}\}$ を用いた．式 (17) はベクトルと行列を用いて

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} \tag{18}$$

と表せる．ここで，$\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$，$\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)'$，$\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_p)'$，

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}$$

とすると，式 (18) は，

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{19}$$

と表せる．ここに，$\boldsymbol{Y}$ を目的変数ベクトル，$\boldsymbol{X}$ を説明変数行列，$\boldsymbol{\beta}$ を回帰係数ベクトル，$\boldsymbol{\epsilon}$ を誤差ベクトルという．また，$E(\boldsymbol{\epsilon}) = \boldsymbol{0}$，$V(\boldsymbol{\epsilon}) = \sigma^2 \boldsymbol{I}_n$ であることに注意する．ここでの関心は，互いに独立な n 個の標本を用いて，未知パラメータ $\boldsymbol{\beta}$ を推定することである．そのために，次のベクトル微分を用いる．

[1] ϵ が正規分布に従うことを仮定していないことに注意しよう．

ベクトル微分

ベクトル変数 $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ の実数値関数 $S(\boldsymbol{\beta})$ のベクトル微分は，

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \left(\frac{\partial S(\boldsymbol{\beta})}{\partial \beta_1}, \ldots, \frac{\partial S(\boldsymbol{\beta})}{\partial \beta_p} \right)'$$

で定義される．

- $\dfrac{\partial \boldsymbol{c}'\boldsymbol{\beta}}{\partial \boldsymbol{\beta}} = \dfrac{\partial \boldsymbol{\beta}'\boldsymbol{c}}{\partial \boldsymbol{\beta}} = \boldsymbol{c}$

  確認：$\boldsymbol{c}'\boldsymbol{\beta} = \boldsymbol{\beta}'\boldsymbol{c} = c_1\beta_1 + \cdots + c_p\beta_p$ より

$$\frac{\partial \boldsymbol{c}'\boldsymbol{\beta}}{\partial \beta_1} = \frac{\partial \boldsymbol{\beta}'\boldsymbol{c}}{\partial \beta_1} = c_1, \ldots, \frac{\partial \boldsymbol{c}'\boldsymbol{\beta}}{\partial \beta_p} = \frac{\partial \boldsymbol{\beta}'\boldsymbol{c}}{\partial \beta_p} = c_p$$

  であるから，

$$\frac{\partial \boldsymbol{c}'\boldsymbol{\beta}}{\partial \boldsymbol{\beta}} = \frac{\partial \boldsymbol{\beta}'\boldsymbol{c}}{\partial \boldsymbol{\beta}} = (c_1, \ldots, c_p)' = \boldsymbol{c}$$

- $\dfrac{\partial \boldsymbol{\beta}'\boldsymbol{A}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}} = (\boldsymbol{A} + \boldsymbol{A}')\,\boldsymbol{\beta}$

  確認：$\boldsymbol{\beta}'\boldsymbol{A}\boldsymbol{\beta} = \sum_i \sum_j a_{ij}\beta_i\beta_j$ より

$$\frac{\partial \boldsymbol{\beta}'\boldsymbol{A}\boldsymbol{\beta}}{\partial \beta_k} = \sum_j a_{kj}\beta_j + \sum_i a_{ik}\beta_i = (\boldsymbol{a}_k + \boldsymbol{a}_k^t)\boldsymbol{\beta}$$

  である．ただし，$\boldsymbol{a}_k$ と $\boldsymbol{a}_k^t$ はそれぞれ $\boldsymbol{A}$ と $\boldsymbol{A}'$ の第 k 行に対応するベクトルである．したがって，

$$\frac{\partial \boldsymbol{\beta}'\boldsymbol{A}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}} = \begin{pmatrix} \boldsymbol{a}_1 + \boldsymbol{a}_1^t \\ \vdots \\ \boldsymbol{a}_p + \boldsymbol{a}_p^t \end{pmatrix} \boldsymbol{\beta} = (\boldsymbol{A} + \boldsymbol{A}')\boldsymbol{\beta}$$

- もし $\boldsymbol{A}$ が対称行列（$\boldsymbol{A} = \boldsymbol{A}'$）ならば，$\dfrac{\partial \boldsymbol{\beta}'\boldsymbol{A}\boldsymbol{\beta}}{\partial \boldsymbol{\beta}} = 2\boldsymbol{A}\boldsymbol{\beta}$

$\boldsymbol{\beta}$ の推定量を**最小 2 乗法**（method of least squares）を用いて構成しよう．標本 $(Y_i, X_{i1}, \ldots, X_{ip})'$（$i = 1, \ldots, n$）の実現値を $(y_i, x_{i1}, \ldots, x_{ip})'$ とする．最小 2 乗法は誤差の 2 乗和を最小にするように $\boldsymbol{\beta}$ を定める方法である．$\boldsymbol{y} = (y_1, \ldots, y_n)'$ とすると，誤差の 2 乗和は

$$\begin{aligned}\epsilon_1^2 + \cdots + \epsilon_n^2 &= \boldsymbol{\epsilon}'\boldsymbol{\epsilon} \\ &= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\ &= \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}\end{aligned}$$

であるから，$\boldsymbol{\beta}$ でベクトル微分すると

$$\frac{\partial \boldsymbol{\epsilon}'\boldsymbol{\epsilon}}{\partial \boldsymbol{\beta}} = \boxed{\text{読者の演習}} = -2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}$$

である．ここで，$\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{0}$ を満たす $\boldsymbol{\beta}$ を $\hat{\boldsymbol{\beta}}$ とする．これは $\boldsymbol{\beta}$ に関する連立 1 次方程式であり，**正規方程式**（normal equation）と呼ばれる．$\boldsymbol{X}'\boldsymbol{X}$ の階数（ランク）が $p+1$ ならば $\boldsymbol{X}'\boldsymbol{X}$ は正則行列であるから，

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

を得る．また，$\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ のとき，誤差の 2 乗和は最小値に到達する．したがって，$\boldsymbol{\beta}$ の**最小 2 乗推定量**（leasts squares estimator）は

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{20}$$

である．$E(\boldsymbol{Y}) = E(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) = \boldsymbol{X}\boldsymbol{\beta} + E(\boldsymbol{\epsilon}) = \boldsymbol{X}\boldsymbol{\beta}$ であるから，

$$\begin{aligned}E\left(\hat{\boldsymbol{\beta}}\right) &= E\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}\right) \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{Y}) \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \\ &= \boldsymbol{\beta}\end{aligned}$$

より最小 2 乗推定量 $\hat{\boldsymbol{\beta}}$ は $\boldsymbol{\beta}$ の不偏推定量である．

5.1 演習問題

問 1 行列 $\boldsymbol{P}$ が対称性 $\boldsymbol{P} = \boldsymbol{P}'$ かつ冪等性 $\boldsymbol{P} = \boldsymbol{P}^2$ を満たすとき，$\boldsymbol{P}$ を直交射影行列という．$n > p+1$ と $\boldsymbol{X}'\boldsymbol{X}$ の階数（ランク）が $p+1$ であることを仮定して，式 (19) の $\boldsymbol{X}$ を用いた行列のなかで，直交射影行列であるものはどれか．次の ① ～ ④ のうちから適切なものを一つ選べ． $\boxed{\mathbf{1}}$

① $\boldsymbol{X}'\boldsymbol{X}$ ② $(\boldsymbol{X}'\boldsymbol{X})^{-1}$

③ $\boldsymbol{X}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}$ ④ $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$

問 2 式 (20) の $\boldsymbol{\beta}$ の最小 2 乗推定量 $\hat{\boldsymbol{\beta}}$ を用いて，

$$e_i = Y_i - \hat{Y}_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip}) \quad (i = 1, \ldots, n)$$

を残差という．残差ベクトル $\boldsymbol{e} = (e_1, \ldots, e_n)'$ を表したものとして正しいものはどれか．次の ① ～ ④ のうちから適切なものを一つ選べ． $\boxed{\mathbf{2}}$

① $\boldsymbol{Y} - \boldsymbol{\beta}$ ② $\boldsymbol{Y} - \hat{\boldsymbol{\beta}}$

③ $\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}$ ④ $(\boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{Y}$

問 3 式 (20) の $\boldsymbol{\beta}$ の最小 2 乗推定量 $\hat{\boldsymbol{\beta}}$ に対して，

$$E\left((\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})\right)$$

はいくらか．次の ① ～ ④ のうちから最も適切なものを一つ選べ．
$\boxed{\mathbf{3}}$

① 0 ② $\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}$

③ $\boldsymbol{\beta}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{\beta}$ ④ 1