確率統計基礎講義Ａレポート課題２

レポート用紙に, 以下の, 問 1, 問 2, 問 3 (それぞれ, □ までが問題) の解答を作成して, 8 月 18 日 (金) までに, 数学事務室カウンター前のボックスに提出してください。

$(X_1, Y_1), \dots, (X_n, Y_n)$ は独立な 2 次元確率変数で,

$$\begin{aligned} &\mathrm{P}(X_i = j, Y_i = k) = p_{jk} \quad (j, k) \in \{0, 1\} \times \{0, 1\}, \\ &p_{00} + p_{01} + p_{10} + p_{11} = 1 \end{aligned}$$

とする. (X_i, Y_i) の確率関数を $f(x_i, y_i)$ とすると

$$f(x_i, y_i) = \begin{cases} p_{00}^{x_i y_i} p_{01}^{x_i(1-y_i)} p_{10}^{(1-x_i)y_i} p_{11}^{(1-x_i)(1-y_i)} & x_i, y_i \in \{0, 1\} \times \{0, 1\} \\ 0 & \text{その他} \end{cases} \tag{1}$$

と表される.

$(X_1, Y_1), \dots, (X_n, Y_n)$ の実現値を $(x_1, y_1), \dots, (x_n, y_n)$ とし

$$N_{kl} = \#\{i;\ (x_i, y_i = (k, l)),\ i = 1, 2, \dots, n\} \quad (k, l) \in \{0, 1\} \times \{0, 1\}$$

と定義する. ただし, $\#$ は集合に含まれる要素の個数を表す.

問 1. $\boldsymbol{\theta} = (p_{00}, p_{01}, p_{10})'$ の尤度関数を, p_{kl}, N_{kl} $(k, l = 0, 1)$ を用いて表せ. □

問 2. $\boldsymbol{\theta}$ の最尤推定量は

$$\hat{\boldsymbol{\theta}} = \left(\frac{N_{00}}{n}, \frac{N_{01}}{n}, \frac{N_{10}}{n}\right)'$$

で与えられることを示せ. □

X_i と Y_i が独立であるとは, $(k, l) \in \{0, 1\} \times \{0, 1\}$ に対して

$$\mathrm{P}(X_i = j, Y_i = k) = \mathrm{P}(X_i = j)\mathrm{P}(Y_i = k)$$

が成り立つことであるから,

$$\boldsymbol{\xi} = (p_{0\cdot}, p_{\cdot 0})', \quad p_{j\cdot} = \mathrm{P}(X_i = j), \quad p_{\cdot k} = \mathrm{P}(Y_i = k),$$

とおくと, X_i と Y_i が独立であるという仮説は

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\xi}) := \begin{pmatrix} p_{0\cdot} p_{\cdot 0} \\ p_{0\cdot}(1 - p_{\cdot 0}) \\ (1 - p_{0\cdot}) p_{\cdot 0} \end{pmatrix}$$

と表すことができる.

問 3　仮説 H_0 の下で, $\boldsymbol{\xi}$ の最尤推定量は

$$\hat{\boldsymbol{\xi}} = \Big(\frac{N_{00}+N_{01}}{n}, \frac{N_{00}+N_{10}}{n}\Big)'$$

で与えられることを示せ.　□

(1) の $f(x_i, y_i)$ を $f(x_i, y_i; \boldsymbol{\theta})$ と表す. $p_{11} = 1 - p_{00} - p_{01} - p_{11}$ なので

$$\frac{\partial \log f(X_i, Y_i; \boldsymbol{\theta})}{\partial p_{00}} = \frac{X_i Y_i}{p_{00}} - \frac{(1-X_i)(1-Y_i)}{p_{11}}$$
$$\frac{\partial^2 \log f(X_i, Y_i; \boldsymbol{\theta})}{\partial {p_{00}}^2} = -\frac{X_i Y_i}{p_{00}^2} - \frac{(1-X_i)(1-Y_i)}{p_{11}^2},$$
$$\frac{\partial^2 \log f(X_i, Y_i; \boldsymbol{\theta})}{\partial p_{00}\partial p_{01}} = -\frac{(1-X_i)(1-Y_i)}{p_{11}^2},$$

より,

$$-\mathrm{E}\left[\frac{\partial^2 \log f(X_i, Y_i; \boldsymbol{\theta})}{\partial {p_{00}}^2}\right] = \frac{1}{p_{00}} + \frac{1}{p_{11}}, \quad -\mathrm{E}\left[\frac{\partial^2 \log f(X_i, Y_i; \boldsymbol{\theta})}{\partial p_{00}\partial p_{01}}\right] = \frac{1}{p_{11}}$$

となる. 同様にして, 残りの 2 次導関数の期待値も計算することで, フィッシャーの情報量行列とその逆行列は

$$J(\boldsymbol{\theta}) = \begin{pmatrix} p_{00}^{-1} & & \\ & p_{01}^{-1} & \\ & & p_{10}^{-1} \end{pmatrix} + p_{11}^{-1}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix},$$
$$J(\boldsymbol{\theta})^{-1} = \begin{pmatrix} p_{00} & & \\ & p_{01} & \\ & & p_{10} \end{pmatrix} - \begin{pmatrix} p_{00} \\ p_{01} \\ p_{10} \end{pmatrix}\begin{pmatrix} p_{00} & p_{01} & p_{10} \end{pmatrix}$$

となることがわかる. 従って, 問 2, 問 3 を用いるとワルド検定統計量は

$$\begin{aligned} T_W &= n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}(\hat{\boldsymbol{\xi}}))' J(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}(\hat{\boldsymbol{\xi}})) \\ &= \frac{(N_{00} - n\hat{p}_{00}^{(0)})^2}{N_{00}} + \frac{(N_{01} - n\hat{p}_{01}^{(0)})^2}{N_{01}} + \frac{(N_{10} - n\hat{p}_{10}^{(0)})^2}{N_{10}} + \frac{(N_{11} - n\hat{p}_{11}^{(0)})^2}{n_{11}} \end{aligned}$$

となる. ただし, $\boldsymbol{\theta}(\hat{\boldsymbol{\xi}}) = (\hat{p}_{00}^{(0)}, \hat{p}_{01}^{(0)}, \hat{p}_{10}^{(0)})'$, $\hat{p}_{11}^{(0)} = 1 - \hat{p}_{00}^{(0)} - \hat{p}_{01}^{(0)} - \hat{p}_{10}^{(0)}$. である.

一方,

$$s_n(\boldsymbol{\theta}(\hat{\boldsymbol{\xi}})) = \frac{\partial}{\partial \boldsymbol{\theta}}\sum_{i=1}^{n} \log f(x_i, y_i; \boldsymbol{\theta}(\hat{\boldsymbol{\xi}})) = \left(\frac{N_{00}}{\hat{p}_{00}^{(0)}} - \frac{N_{11}}{\hat{p}_{11}^{(0)}},\ \frac{N_{01}}{\hat{p}_{01}^{(0)}} - \frac{N_{11}}{\hat{p}_{11}^{(0)}},\ \frac{N_{10}}{\hat{p}_{10}^{(0)}} - \frac{N_{11}}{\hat{p}_{11}^{(0)}}\right)'$$

となるから, スコア検定統計量は

$$\begin{aligned} T_s &= \frac{1}{n} s_n(\boldsymbol{\theta}(\hat{\boldsymbol{\xi}}))' J(\boldsymbol{\theta}(\hat{\boldsymbol{\xi}})) s_n(\boldsymbol{\theta}(\hat{\boldsymbol{\xi}})) \\ &= \frac{(N_{00} - n\hat{p}_{00}^{(0)})^2}{n\hat{p}_{00}^{(0)}} + \frac{(N_{01} - n\hat{p}_{01}^{(0)})^2}{n\hat{p}_{01}^{(0)}} + \frac{(N_{10} - n\hat{p}_{10}^{(0)})^2}{n\hat{p}_{10}^{(0)}} + \frac{(N_{11} - n\hat{p}_{11}^{(0)})^2}{n\hat{p}_{11}^{(0)}} \end{aligned}$$

となることが確かめられる.