ベイズ統計

古谷知之

講義概要

- 線形回帰モデルのベイズ推定(復習)
- モデル選択
  – ベイズファクター(BF)
- BFとp値を巡る議論
- マルチレベル・モデルのベイズ推定
- R演習
  – R2jags

線形回帰モデル

- サンプル数n (i=1, …, n)、説明変数の数k
- 被説明変数y_i、説明変数x_{ik}、未知パラメータβ_k、誤差項ε_i
- 誤差項ε_iは平均0、分散σ^2の正規分布に従う⇒分散は未知

$$y_i = \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{ik}\beta_k + \varepsilon_i,$$

$$\varepsilon_i \sim N\left(0, \sigma^2\right)$$

線形回帰モデル

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\downarrow \qquad\qquad \downarrow \qquad\qquad \downarrow \qquad \downarrow$$

$$y \qquad\qquad X \qquad\qquad \beta \qquad \varepsilon$$

$$\Rightarrow \quad y = X\beta + \varepsilon$$

線形回帰モデルのベイズ推定

- 線形回帰モデルは、未知パラメータβと精度τについて共役事前分布を与えることによりベイズ推定できる
- 自然共役事前分布

$$p(\beta \mid \tau, y) = N\left(\left(X^T X\right)^{-1} X^T y, \left(1/\tau\right)\left(X^T X\right)^{-1}\right)$$

$$p(\tau \mid y) \propto \Gamma\left(a + \frac{\nu}{2}, b + \frac{\nu}{2}\hat{\sigma}^2\right) \propto \Gamma\left(\frac{\nu}{2}, \frac{\nu}{2}\hat{\sigma}^2\right)$$

線形回帰モデルのベイズ推定

- 共役事前分布の超パラメータを導入することにより次式のように簡略化する

$$p\left(\beta \mid \tau, y\right) \sim N\left(b_0, \left(1/\tau\right) B_0\right)$$

$$p\left(\tau \mid y\right) \sim \Gamma\left(\frac{\nu_0}{2}, \frac{\nu_0 S_0}{2}\right)$$

- 超パラメータb_0, B_0, ν_0, S_0は適当に設定しても問題ない

線形回帰モデルのベイズ推定

- 超パラメータを以下のように置き換える

$$b_0 = \left(X^T X\right)^{-1} X^T y$$

$$B_0 = \left(X^T X\right)^{-1}$$

$$\nu_0 = \nu$$

$$S_0 = \hat{\sigma}^2 = \frac{1}{\nu}\left(y - x\hat{\beta}\right)^T \left(y - x\hat{\beta}\right)$$

線形回帰モデルのベイズ推定

- 以下のような事後分布が得られる

$$p\left(\beta \mid \tau, y\right) \sim N\left(b_1, B_1\right)$$

$$p\left(\tau \mid y\right) \sim \Gamma\left(\frac{\nu_1}{2}, \frac{\nu_1 S_1}{2}\right)$$

$$b_1 = B_1\left(B_0^{-1} b_0 + \tau X^T y\right),\ B_1^{-1} = B_0^{-1} + \tau X^T X,$$

$$\nu_1 = \nu_0 + n,\ \nu_1 S_1 = \nu_0 S_0 + \left(y - X b_1\right)^T \left(y - X b_1\right)$$

線形回帰モデルのベイズ推定

- β の周辺事後分布は以下のt分布に従う

$$p\left(\beta \mid \tau, y\right) \sim t\left(b_1, \hat{\sigma}^2 S_1, \nu_1\right)$$

$$E\left(\beta \mid \tau, y\right) = b_1$$

$$V\left(\beta \mid \tau, y\right) = \frac{\nu_1 \hat{\sigma}^2}{\nu_1 - 2} S_1$$

```
> print(swiss.jags, digits=3)
Inference for Bugs model at "/Library/Frameworks/R.framework/Versions/3.0/
Resources/library/R2jags/model/model1.txt", fit using jags,
 3 chains, each with 10000 iterations (first 1000 discarded), n.thin = 9
 n.sims = 3000 iterations saved
         mu.vect sd.vect    2.5%     25%     50%     75%   97.5%  Rhat n.eff
beta0     67.105  13.990  45.178  59.293  67.151  74.183  89.391 1.003  1400
beta1     -0.172   0.111  -0.312  -0.220  -0.171  -0.122  -0.026 1.129  1900
beta2     -0.264   0.354  -0.766  -0.434  -0.258  -0.090   0.257 1.094  3000
beta3     -0.870   0.190  -1.227  -0.992  -0.873  -0.747  -0.505 1.001  3000
beta4      0.104   0.039   0.035   0.080   0.104   0.128   0.176 1.009  3000
beta5      1.071   0.402   0.298   0.812   1.068   1.339   1.858 1.001  2100
tau        0.020   0.004   0.012   0.016   0.019   0.022   0.029 1.009  1200
deviance 319.807   7.890 313.920 316.642 318.926 321.825 329.118 1.023  1400

For each parameter, n.eff is a crude measure of effective sample size,
and Rhat is the potential scale reduction factor (at convergence, Rhat=1).

DIC info (using the rule, pD = var(deviance)/2)
pD = 31.1 and DIC = 350.9
DIC is an estimate of expected predictive error (lower deviance is better).
```

```
> gelman.diag(as.mcmc(swiss.jags))
Potential scale reduction factors:

         Point est. Upper C.I.
beta0         1.003       1.01
beta1         1.004       1.02
beta2         1.000       1.00
beta3         1.002       1.01
beta4         0.999       1.00
beta5         1.002       1.01
deviance      1.004       1.01
tau           1.000       1.00

Multivariate psrf

1.01
```

```
> geweke.diag(as.mcmc(swiss.jags))
[[1]]

Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

   beta0    beta1    beta2    beta3    beta4    beta5 deviance      tau
  1.0412  -0.6132  -0.2245  -1.2326  -1.7890  -0.7599   1.1214  -1.2407


[[2]]

Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

   beta0    beta1    beta2    beta3    beta4    beta5 deviance      tau
-0.59821  1.39684 -1.31319  1.67479 -1.04701  0.74338  1.25556 -0.07713


[[3]]

Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5

   beta0    beta1    beta2    beta3    beta4    beta5 deviance      tau
  0.4123  -0.7620  -0.6936   0.7896   1.1846   0.9047   1.3075  -1.9717
```

```
> raftery.diag(as.mcmc(swiss.jags))
[[1]]

Quantile (q) = 0.025
Accuracy (r) = +/- 0.005
Probability (s) = 0.95

You need a sample size of at least 3746 with these values of q, r and s

[[2]]

Quantile (q) = 0.025
Accuracy (r) = +/- 0.005
Probability (s) = 0.95

You need a sample size of at least 3746 with these values of q, r and s

[[3]]

Quantile (q) = 0.025
Accuracy (r) = +/- 0.005
Probability (s) = 0.95

You need a sample size of at least 3746 with these values of q, r and s
```

モデル間の比較

- いま、２つのモデルM_0とM_1を推定しどちらがよいか比較したい
- ２つのモデルの事後分布を簡略化してそれぞれ以下のように表す
  - モデルM_0： $p\left(\beta^{M_0},\tau^{M_0} \mid y\right) = p\left(M_0 \mid y\right)$
  - モデルM_1： $p\left(\beta^{M_1},\tau^{M_1} \mid y\right) = p\left(M_1 \mid y\right)$

モデル間の比較

- ベイズの定理から、事後分布を尤度と事前分布を用いて書き換えると、

$$p\left(M_0 \mid y\right) = p\left(y \mid M_0\right) p\left(M_0\right) \Big/ p\left(y\right)$$

$$p\left(M_1 \mid y\right) = p\left(y \mid M_1\right) p\left(M_1\right) \Big/ p\left(y\right)$$

- ２つのモデルの比を取ると

$$\frac{p\left(M_0 \mid y\right)}{p\left(M_1 \mid y\right)} = \frac{p\left(y \mid M_0\right)}{p\left(y \mid M_1\right)} \cdot \frac{p\left(M_0\right)}{p\left(M_1\right)}$$

ベイズ・ファクター

- 事後分布の比＝尤度比×事前分布の比

$$\underbrace{\frac{p(M_0 \mid y)}{p(M_1 \mid y)}}_{\text{Posterior odds}} = \underbrace{\frac{p(y \mid M_0)}{p(y \mid M_1)}}_{\substack{\text{Bayes factor} \\ (B_{01})}} \cdot \underbrace{\frac{p(M_0)}{p(M_1)}}_{\text{Prior odds}}$$

- ２つのモデルを比較して、モデルM_0とM_1のどちらが優勢かは事前にわからない
  $p(M_0)=P(M_1)=1/2$ので、ベイズ・ファクターB_{01}の値がモデルの優劣を決める

ベイズ・ファクター

- ベイズ・ファクターB_{01}の値により、帰無仮説モデルM_0と提案モデルM_1の指示度がかわる

Jeffreys (1961)

ベイズ・ファクター	解釈
$\log B_{01} < 0$	帰無仮説モデルM_0を支持
$0 < \log B_{01} < 0.5$	M_1はよくも悪くもない
$0.5 < \log B_{01} < 1$	M_1は十分である
$1 < \log B_{01} < 2$	M_1を強く支持できる
$2 < \log B_{01}$	M_1は決定的に支持できる

BFとp値

- p値はしばしば帰無仮説の事後確率と誤解されているのではないか
- 単純な帰無仮説に対して、p値は帰無仮説のエビデンスをかなり水増ししているのではないか
- 対立仮説は帰無仮説の片側分布に対応しているので、p値を用いるなら両側検定より片側検定が適しているのではないか

BFとp値

- H0とH1に等しい事前確率を与えた場合、

両側検定	p値 0.05	0.01	0.001
t値	1.96	2.58	3.29
BF	0.15	0.04	0.004
min(H_0|x)	12.8%	3.5%	0.4%

片側検定	p値 0.05	0.01	0.001
t値	1.64	2.33	3.09
BF	0.26	0.07	0.008
min(H_0|x)	20.5%	6.3%	0.8%

Edwards et al. (1963)

p値にまつわる12の誤解

S. Goodman (2008)

ELSEVIER

Seminars in
HEMATOLOGY

A Dirty Dozen: Twelve *P*-Value Misconceptions

Steven Goodman

Table 1 Twelve *P*-Value Misconceptions

1	*If P = .05, the null hypothesis has only a 5% chance of being true.*
2	*A nonsignificant difference (eg, P ≥.05) means there is no difference between groups.*
3	*A statistically significant finding is clinically important.*
4	*Studies with P values on opposite sides of .05 are conflicting.*
5	*Studies with the same P value provide the same evidence against the null hypothesis.*
6	*P = .05 means that we have observed data that would occur only 5% of the time under the null hypothesis.*
7	*P = .05 and P ≤.05 mean the same thing.*
8	*P values are properly written as inequalities (eg, "P ≤.02" when P = .015)*
9	*P = .05 means that if you reject the null hypothesis, the probability of a type I error is only 5%.*
10	*With a P = .05 threshold for significance, the chance of a type I error will be 5%.*
11	*You should use a one-sided P value when you don't care about a result in one direction, or a difference in that direction is impossible.*
12	*A scientific conclusion or treatment policy should be based on whether or not the P value is significant.*

マルチレベル・モデル

- グループj(=1, …, m)に属する個人i(=1, … n)を考える
- 説明変数の数k、被説明変数y_i、説明変数x_{ik}、未知パラメータβ_k、誤差項ε_i
- 誤差項ε_iは平均0、分散σ^2の正規分布に従う⇒分散は未知
- このとき、

$$y_{ij} = \beta_0 + x_{ij1}\beta_1 + x_{ij2}\beta_2 + \cdots + x_{ijk}\beta_k + \varepsilon_i,$$

$$\varepsilon_i \sim N\left(0, \sigma_y^2\right)$$

マルチレベル・モデル

① 定数項と回帰係数が変動しないモデル

$$y_{ij} = \beta_0 + x_{ij1}\beta_1 + x_{ij2}\beta_2 + \cdots + x_{ijk}\beta_k + \varepsilon_i$$

② 定数項のみ変動するモデル

$$y_{ij} = \beta_{j0} + x_{ij1}\beta_1 + x_{ij2}\beta_2 + \cdots + x_{ijk}\beta_k + \varepsilon_i$$

③ 回帰係数のみ変動するモデル

$$y_{ij} = \beta_0 + x_{ij1}\beta_{j1} + x_{ij2}\beta_{j2} + \cdots + x_{ijk}\beta_{jk} + \varepsilon_i$$

④ 定数項と回帰係数が変動するモデル

$$y_{ij} = \beta_{j0} + x_{ij1}\beta_{j1} + x_{ij2}\beta_{j2} + \cdots + x_{ijk}\beta_{jk} + \varepsilon_i$$

ランダム効果と固定効果

- ランダム効果
  - 回帰係数や定数項について、個人や地域・グループでの変動を認めるモデル
- 固定効果
  - 上記のような変動を認めないモデル
- 混合効果
  - ランダム効果と固定効果が混在したモデル

マルチレベル・モデル

- レベル1 = lower level

$$y_{ij} = \beta_0 + x_{ij1}\beta_1 + \varepsilon_{ij},\ \varepsilon_{ij} \sim N\left(0, \sigma_y^2\right)$$

$$y_{ij} \sim N\left(\beta_0 + x_{ij1}\beta_1, \sigma_y^2\right)$$

- レベル2 = upper level

グループによる変動を受けない（固定効果）

グループによる変動を受ける（ランダム効果）

$$\beta_{j0} = \gamma_{00} + \gamma_{10}u_{j0} + \eta_{j0},$$

$$\beta_{j1} = \gamma_{01} + \gamma_{11}u_{j0} + \eta_{j1}$$

$$\eta_{jk} \sim N\left(0, \sigma_\eta^2\right),\ \beta_{jk} \sim N\left(\mu_k, \sigma_{\beta_k}^2\right),\ k = \left\{0,1\right\}$$

マルチレベル・モデル

- レベル１とレベル２をあわせると

$$
\begin{aligned}
y_{ij} &= \beta_0 + x_{ij1}\beta_1 + \varepsilon_{ij} \\
&= \left(\gamma_{00} + \gamma_{10}u_{j0}\right) + \left(\gamma_{01} + \gamma_{11}u_{j0}\right)x_{ij1} + \eta_j + \varepsilon_{ij} \\
&= \underbrace{\left(\gamma_{00} + \gamma_{01}x_{ij1}\right)}_{\text{固定効果}} + \underbrace{\left(\gamma_{10}u_{j0} + \gamma_{11}u_{j0}x_{ij1}\right)}_{\text{ランダム効果}} + \eta_j + \varepsilon_{ij}
\end{aligned}
$$

$\Leftrightarrow$

$$y_i = X_i B + Z_i b_i$$

$$b_i \sim N\left(0,\ \sigma_{b_i}^2\right)$$

マルチレベル・モデルのベイズ推定

$$y_i \sim N\left(X_i B, \sigma_y^2\right)$$

$$\begin{pmatrix} \beta_{j0} \\ \beta_{j1} \end{pmatrix} \sim N\left(\begin{pmatrix} \gamma_{00} + \gamma_{10} u_{j0} \\ \gamma_{01} + \gamma_{11} u_{j1} \end{pmatrix}, \begin{pmatrix} \sigma_{\beta_0}^2 & \rho\sigma_{\beta_0}\sigma_{\beta_1} \\ \rho\sigma_{\beta_0}\sigma_{\beta_1} & \sigma_{\beta_1}^2 \end{pmatrix} \right)$$

$$B \sim N\left(0, V_B\right)$$

$$\varepsilon_{ij} \sim N\left(0, \sigma_y^2\right),\ \eta_{jk} \sim N\left(0, \sigma_\eta^2\right)$$

マルチレベル・モデルのベイズ推定

- 未知パラメータ

$$y_i \mid X_i, B, \sigma_y^2 \sim N\left(X_i B, \sigma_y^2\right)$$

$$B \mid Z_i, \sigma_{b_i}^2, V_B \sim N\left(0, V_B\right)$$

$$V_B \mid \upsilon, V \sim IW\left(\upsilon, V\right)$$

$$\sigma_y^2 \mid s_{y_i}^2, \upsilon \sim \eta_j s_{y_i}^2 \Big/ \chi_{\eta_j}^2$$

$$b_i \mid \sigma_{b_i}^2 \sim N\left(0, \sigma_{b_i}^2\right)$$

$$\sigma_{b_i}^2 \mid \eta_i \sim N\left(0, \sigma_{b_i}^2\right)$$

ウィシャート分布

- x_iがn個のp次元多変量正規分布に従うとき

$$x_i \sim N\left(0, \Sigma\right)$$

- 標本分散共分散行列Wは逆ウィシャート分布に従う

$$X = \sum_{i=1}^{n} x_i \cdot x_i^T$$

$$X \sim W\left(X; \Sigma, n, p\right)$$

- ベイズ統計では多変量正規分布の共分散行列の自然共役事前分布として用いられる

ウィシャート分布

- ウィシャート行列はカイ二乗分布を多次元化したものである

$$W\left(X;\Sigma,n,p\right)=\frac{\left|\Sigma\right|^{-\frac{n}{2}}\left|X\right|^{\frac{n-p-1}{2}}}{2^{\frac{np}{2}}\cdot\Gamma_p\left(\frac{n}{2}\right)}\cdot\exp\left(-\frac{1}{2}tr\Sigma^{-1}X\right)$$

- Γ_p(a)はp次元多変量Γ関数

$$\Gamma_p\left(p\right)=\pi^{\frac{p\left(p-1\right)}{4}}\prod_{j=1}^{p}\left[a-\frac{1}{2}\left(j-1\right)\right]$$

逆ウィシャート分布

- W(Σ,n,p)に従うXをV=X^{-1}とすると、Vの分布を逆ウィシャート分布という

$$IW\left(X;\Sigma,n,p\right)=\frac{\left|\Sigma\right|^{-\frac{n}{2}}\left|V\right|^{-\frac{n+p+1}{2}}}{2^{\frac{np}{2}}\cdot\Gamma_p\left(\frac{n}{2}\right)}\cdot\exp\left(-\frac{1}{2}tr\Sigma^{-1}V^{-1}\right)$$