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
  - ベイズ情報量規準(BIC)
  - 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

線形回帰モデル（重回帰分析）

- 従属変数yとk個の独立変数$x_1, x_2, \dots, x_k$に対する標本数がn個の重回帰モデルは以下のように記述できる$(i = 1, \dots, n)$。

$$
\begin{gathered}
y_1 = \beta_0 + \beta_1 x_{11} + \cdots + \beta_k x_{1k} + \varepsilon_1 \\
y_2 = \beta_0 + \beta_1 x_{21} + \cdots + \beta_k x_{2k} + \varepsilon_2 \\
\vdots \\
y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i \\
\vdots \\
y_n = \beta_0 + \beta_1 x_{n1} + \cdots + \beta_k x_{nk} + \varepsilon_n
\end{gathered}
$$

線形回帰モデル（重回帰分析）

- 次のようなベクトルと行列を用いて、

$$
\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{i1} & \cdots & x_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{pmatrix}
$$

- 次式のように簡略化できる

$$
\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}
$$

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

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

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

尤度関数(全データ)

- データX、未知パラメータ$\boldsymbol{\beta}$、分散$\boldsymbol{\sigma}^2$が与えられた条件下で、被説明変数yが得られる条件付き確率、すなわち尤度関数は、以下のような正規分布となる

$$p(y|X;\boldsymbol{\beta},\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{(y - X\boldsymbol{\beta})^2}{2\sigma^2}\right]$$

尤度関数(全データ)

- 尤度関数の平均と分散はそれぞれ以下のとおりとなる

- 平均

$$E(y|X;\boldsymbol{\beta},\sigma^2)=X\boldsymbol{\beta}$$

- 分散

$$V(y|X;\boldsymbol{\beta},\sigma^2)=\sigma^2$$

尤度関数(全データ)

- 全てのデータについての尤度関数は

$$p(y|X;\boldsymbol{\beta},\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}exp\left[-\frac{(y-X\boldsymbol{\beta})^2}{2\sigma^2}\right]$$

$$=\frac{1}{\sqrt{2\pi\sigma^2}}exp\left[-\frac{(y-X\boldsymbol{\beta})^T(y-X\boldsymbol{\beta})}{2\sigma^2}\right]$$

対数尤度関数(全データ)

- 全てのデータについての尤度関数は

$$\ln[p(y|X;\boldsymbol{\beta},\sigma^2)] = \ln\left[\frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{(y-X\boldsymbol{\beta})^T(y-X\boldsymbol{\beta})}{2\sigma^2}\right]\right]$$

$$= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{(y-X\boldsymbol{\beta})^T(y-X\boldsymbol{\beta})}{2\sigma^2}$$

ベイズ情報量規準（BIC）

- ベイズ情報量規準は次式で与えられる

$$BIC = -2\ln\hat{L} + 2k \cdot \ln(n)$$

- 誤差項が正規分布に従う重回帰モデルに対しては、次式のようになる

$$BIC = n\left(\ln\left(2\pi\frac{\sum_{i=i}^{n}(y_i-\hat{y}_i)^2}{n}\right)+1\right) + 2k \cdot \ln(n)$$

- AICと同様に、BICがより小さいモデルが当てはまりの良いモデルといえる

尤度関数

- 精度$\tau = 1/\sigma^2$とすると尤度関数は以下のように式変形できる

$$p(y|X;\boldsymbol{\beta},\tau) = \frac{\tau^{n/2}}{(2\pi)^{n/2}} \cdot \left\{exp\left[-\frac{\tau}{2}(y - X\boldsymbol{\beta})^T(y - X\boldsymbol{\beta})\right]\right\}$$

$$\propto \tau^{\nu/2} \cdot \left\{exp\left[-\frac{\nu\hat{\sigma}^2}{2}\tau\right]\right\}$$

階層ベイズ

- 平均と分散(精度)が未知のとき、精度のガンマ分布と尤度関数の正規分布を乗じることで事後分布が得られる

精度τがガンマ分布に従う	p(τ)∝Γ( )
↓	
精度τが与えられた条件下で パラメータβが正規分布に従う	p(β|τ) ∝N( )
↓	
パラメータβと精度τが与えられた 条件下で被説明変数yが得られる	p(y|β, τ) ∝N( )

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$\boldsymbol{\beta}$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

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

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

線形回帰モデルのベイズ推定

- 共役事前分布の超（ハイパー）パラメータを導入することによりを次式のように簡略化する

$$p(\boldsymbol{\beta}|\tau, y) \sim N\left(b_0, \left(\frac{1}{\tau}\right) B_0\right)$$

$$p(\tau|y) \propto Ga\left(\frac{\nu_0}{2}, \frac{\nu_0 S_0}{2}\right)$$

- 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

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

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

$$\nu_0 = \nu$$

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

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

$$p(\boldsymbol{\beta}|\tau, y) \sim N(b_1, B_1\ )$$

$$p(\tau|y) \propto Ga\left(\frac{\nu_1}{2}, \frac{\nu_1 S_1}{2}\right)$$

- ここで、

$$b_1 = B_1(B_0^{-1} b_0 + \tau X^T y)$$
$$B_1^{-1} = B_0^{-1} + \tau X^T X$$
$$\nu_1 = \nu_0 + n$$
$$\nu_1 S_1 = \nu_0 S_0 + (y - X b_1)^T (y - X b_1)$$

線形回帰モデルのベイズ推定

- $\boldsymbol{\beta}$の周辺事後分布は以下のt分布に従う

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

$$E(\boldsymbol{\beta}|\tau, y) = b_1$$

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

線形回帰モデルのギブスサンプリング

- 繰り返し回数$s = 0, \ldots, N$とし、初期値b_0, B_0, ν_0, S_0を設定する

1. $p\left(\boldsymbol{\beta}^{\boldsymbol{s+1}}|\tau^s, y\right) \sim N(b_1^s, (1/\tau)B_1^s)$から$\boldsymbol{\beta}^{\boldsymbol{s+1}}$を生成する
2. $p\left(\tau^{s+1}|\boldsymbol{\beta}^{\boldsymbol{s+1}}, y\right) \sim Ga\left(\frac{\nu_1^s}{2}, \frac{\nu_1^s S_1^s}{2}\right)$から$\tau^{s+1}$を生成する
3. $s < N$のとき1に戻る。$s = N$のとき計算終了

線形回帰モデルのベイズ推定

- パラメータの事後平均の点推定値は、ベイズ推定の事前情報とOLSの点推定値との重み付け平均

$$E(\boldsymbol{\beta}|\tau, y) = b_1 = B_1(B_0^{-1}b_0 + \tau X^T y)$$

- 事前情報B_0^{-1}とν_0が非常に小さいときには、事後情報はどうなるだろうか？

線形回帰モデルのベイズ推定

- 事前情報$B_0^{-1} = 0$及び$\nu_0 = 0$のとき、事後情報はそれぞれ以下のようになる

$$b_1 = (X^T X)^{-1} X^T y = \widehat{\boldsymbol{\beta}}$$
$$\nu_1 = \nu_0$$
$$S_1 = \frac{1}{\nu}\left(y - X\widehat{\boldsymbol{\beta}}\right)^T\left(y - X\widehat{\boldsymbol{\beta}}\right) = \hat{\sigma}^2$$

- すなわち、最小二乗解と同じ結果となる

モデル間の比較

- いま、２つのモデル$\boldsymbol{M_0}$と$\boldsymbol{M_1}$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： $p(\boldsymbol{\beta}^{M_0}, \tau^{M_0}|y) = p(M_0|y)$

- モデルM_1： $p(\boldsymbol{\beta}^{M_1}, \tau^{M_1}|y) = p(M_1|y)$

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、

$$p(M_0|y) = p(y|M_0)p(M_0)/p(y)$$
$$p(M_1|y) = p(y|M_1)p(M_1)/p(y)$$

- ２つのモデルの比を取ると

$$\frac{p(M_0|y)}{p(M_1|y)} = \frac{p(y|M_0)}{p(y|M_1)} \cdot \frac{p(M_0)}{p(M_1)}$$

ベイズ・ファクター

- 事後分布の比＝尤度比×事前分布の比

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

- 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 BF_{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 BF_{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
$\log BF_{01}>0$	帰無仮説モデルM_0を支持
$0>\log BF_{01}>0.5$	M_1はよくも悪くもない
$0.5>\log BF_{01}>1$	M_1は十分である
$1>\log BF_{01}>2$	M_1を強く支持できる
$2>\log BF_{01}$	M_1は決定的に支持できる

ベイズファクターとBIC

- モデルM_0のBICをBIC_0、モデルM_1のBICをBIC_1とする

- このとき、 $BIC_0 - BIC_1$が$2\log(BF_{01})$に近似する

- 比較的簡便に計算できるので用いられることがあるが、ベイズファクターとは異なる指標であることに注意すること

ベイズファクターと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

推測統計学に対する懐疑論

- P値論争
  - 統計的有意性に p 値を用いることが妥当なのかどうかについて、様々な批判や議論が展開されてきた
  - アメリカ統計学会（ASA）がp値の限界などについて声明発表
  - インパクトファクターが高い学術誌などがp値の使用を禁止

- ビッグデータの普及
  - 小規模標本データを前提とした統計解析手法に縛られる必要はない

アメリカ統計学会(ASA)による声明(2016)

1. p値はデータと特定の統計モデルが矛盾する程度を示す指標の一つである
2. p値は調べている仮説が正しい確率や、データが偶然のみで得られた確率を測るものではない
3. 科学的な結論や、ビジネス、政策における決定は、p値がある値を超えたかどうかにのみ基づくべきではない
4. 適正な推測のためには、全てを報告する透明性が必要である
5. p値や統計的有意性は、効果の大きさや結果の重要性を意味しない
6. p値は、それだけでは統計モデルや仮説に関するエビデンスの良い指標とはならない

日本生物学会HPより（http://biometrics.gr.jp/news/all/ASA.pdf）

ベイズファクターとp値

- H_0とH_1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に属する個人や地域iについて、以下のような線形回帰モデルを考える

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

- ここで、誤差項は未知である
- 定数項と回帰係数について、グループ毎の違いを認めるかどうかによって、柔軟なモデル推定ができる

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

$$y_{ij} = \beta_0 + \beta_1 x_{ij1} + \cdots + \beta_k x_{ijk} + \varepsilon_{ij}$$

② 定数項のみが変動するモデル

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

③ 回帰係数のみが変動するモデル

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

④ 定数項と回帰係数の両方が変動するモデル

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

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

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

- レベル２=upper level

$$\beta_{j0} = \gamma_{00} + \gamma_{10} u_{j0} + \eta_{j0}$$
$$\beta_{j1} = \gamma_{01} + \gamma_{11} u_{j1} + \eta_{j1}$$
$$\eta_{jk} \sim N\left(0, \sigma_\eta^2\right)$$
$$\beta_{jk} \sim N\left(0, \sigma_{\beta_k}^2\right)$$
$$k = \{0, 1\}$$

マルチレベルモデル

- レベル１とレベル２を合わせると、以下のように式変形される

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

- この式は、次式のように整理できる

$$
\boldsymbol{y_i} = \boldsymbol{X_i B} + \boldsymbol{Z_i b_i}
$$
$$
\boldsymbol{b_i} \sim N(0, \sigma_{b_i}^2)
$$

マルチレベル・モデルのベイズ推定

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

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

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

マルチレベル・モデルのベイズ推定

- 未知パラメータ

$$y_i|X_i, B, \sigma_y^2 \sim N\left(X_i B, \sigma_y^2\right)$$
$$B|Z_i, \sigma_{b_i}^2, V_B \sim N(0, V_B)$$
$$V_B|\upsilon, V \sim IW(\upsilon, V)$$
$$\sigma_y^2|s_{y_i}^2 \sim \frac{\eta_j s_{y_i}^2}{\chi_{\eta_j}^2}$$
$$\boldsymbol{b_i}|\sigma_{b_i}^2 \sim N\left(0, \sigma_{b_i}^2\right)$$
$$\sigma_{b_i}^2|\eta_j \sim N\left(0, \sigma_\eta^2\right)$$

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

$$X_i \sim N(0, \Sigma)$$

- 標本分散共分散行列Wは逆ウィシャート分布に従う

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

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

- 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

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

- $\Gamma_p(a)$はp次元多変量Γ関数

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

逆ウィシャート分布

- $W(X;\Sigma,n,p)$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

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