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
  - 通常最小二乗法(OLS)との比較
  - 尤度関数と自然共役事前分布
  - 平均・分散未知の場合の階層ベイズ推定
- 無情報事前分布と非正則事前分布
- 偏差情報量基準
- R演習

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

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

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

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

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

尤度関数

- データX、未知パラメータβ、分散σ^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は
- $x_i = \left(1, x_{i1}, \ldots, x_{ik}\right)$ とするとき
- 尤度関数は正規分布となる

$$p\left(y_i \mid x_i; \beta, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\left(y_i - x_i\beta\right)^2}{2\sigma^2}\right]$$

尤度関数

- 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

$$E\left(y_i \mid x_i;\beta,\sigma^2\right) = x_i\beta$$

- 分散

$$V\left(y_i \mid x_i;\beta,\sigma^2\right) = \sigma^2$$

尤度関数

- 全てのiについての尤度関数は

$$p\left(y \mid x;\beta,\sigma^2\right)$$

$$=\prod_{i=1}^{n} p\left(y_i \mid x_i;\beta,\sigma^2\right)$$

$$=\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{\sum_{i=1}^{n}\left(y_i - x_i\beta\right)^2}{2\sigma^2}\right]$$

対数尤度関数

- 尤度関数の自然対数をとると

$$\ln\left[p\left(y \mid x;\beta,\sigma^2\right)\right]$$

$$=\prod_{i=1}^{n}\ln\left[p\left(y_i \mid x_i;\beta,\sigma^2\right)\right]$$

$$=-\frac{n}{2}\ln 2\pi-\frac{n}{2}\ln\sigma^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i-x_i\beta\right)^2$$

最尤法と最小二乗法

- 最尤法では(対数)尤度関数を最大化することで未知パラメータを得る⇒対数尤度関数は上に凸となる関数
- 対数尤度関数を最大化することは、次式を最小化することと同じ

$$\sum_{i=1}^{n}\left(y_i - x_i\beta\right)^2$$

最小二乗法による解

- 次式を最小化することにより得られる未知パラメータはそれぞれ以下のようになる

$$\sum_{i=1}^{n}\left(y_i - x_i\beta\right)^2$$

- 最小二乗解

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

$$\hat{\sigma}^2 = \frac{1}{n-\left(k+1\right)}\sum_{i=1}^{n}\left(y_i - x_i\hat{\beta}\right)^2$$

自由度: n-(k+1)

尤度関数(全データ)

- データX、未知パラメータβ、分散σ^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は
- 尤度関数は正規分布となる

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

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

$$E\left(y \mid X; \beta, \sigma^2\right) = X\beta$$

- 分散

$$V\left(y \mid X; \beta, \sigma^2\right) = \sigma^2$$

尤度関数(全データ)

- 全てのデータについての尤度関数は

$$
p\left(y \mid X; \beta, \sigma^2\right)
$$

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

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

$$\ln p\left(y \mid X; \beta, \sigma^2\right)$$

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

最小二乗法による解(全データ)

- 対数尤度関数を最大化⇔最小二乗法による不偏推定量が得られる

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

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

$$\nu = n - \left(k + 1\right)$$ …自由度

↑定数項を加えた変数の数

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

$$p(y \mid X; \beta, \tau)$$

$$= \frac{\tau^{1/2}}{(2\pi)^{n/2}} \cdot \left\{ \exp\left[-\frac{\tau}{2} (y - X\beta)^T (y - X\beta) \right] \right\}$$

$$\propto \tau^{1/2} \cdot \left\{ \exp\left[-\frac{\nu\hat{\sigma}^2}{2} \tau \right] \right\}$$

自然共役事前分布と事後分布

- 正規分布の自然共役事前分布にガンマ分布があった
- 精度 $\tau = 1/\sigma^2$ が未知のとき、その事前分布にガンマ分布Γ(τ |a, b)を乗じると事後分布(ガウス-ガンマ分布)が得られる

$$p(y \mid X; \beta, \tau) \cdot p(\tau \mid a, b)$$

$$\propto \tau^{\nu/2} \cdot \exp\left[-\frac{\nu\hat{\sigma}^2}{2}\tau\right] \qquad \text{…ガウス分布}$$

$$\cdot \tau^{a-1} \exp(-b\tau) \qquad \text{…ガンマ分布}$$

自然共役事前分布と事後分布

- 尤度関数(正規分布)に自然共役事前分布(ガンマ分布)を乗じて得られる事後分布(ガンマ分布)は以下のようになる（正規化係数のガンマ分布はキャンセル）

$$p\left(y \mid X;\beta,\tau\right)\cdot p\left(\tau \mid a,b\right)$$

$$\propto \tau^{\nu/2}\cdot\tau^{a-1}\cdot\exp\left[-b\tau-\frac{\nu\hat{\sigma}^2}{2}\tau\right]$$

$$\propto \tau^{\left(a+\nu/2\right)-1}\cdot\exp\left[-\left\{b+\frac{\nu\hat{\sigma}^2}{2}\right\}\tau\right]$$

自然共役事前分布と事後分布

- 事後分布

$$p\left(y \mid X;\beta,\tau\right)\cdot p\left(\tau \mid a,b\right)$$

$$\propto \tau^{\left(a+\nu/2\right)-1}\cdot\exp\left[-\left\{b+\frac{1}{2}\left(y-X\beta\right)^{T}\left(y-X\beta\right)\right\}\tau\right]$$

$$\propto \tau^{\left(a+\nu/2\right)-1}\cdot\exp\left[-\left\{b+\frac{\nu}{2}\hat{\sigma}^{2}\right\}\tau\right]$$

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

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	$p(\tau) \propto \Gamma(\ )$
↓	
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	$p(\beta \mid \tau) \propto N(\ )$
↓	
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	$p(y \mid \beta, \tau) \propto N(\ )$

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

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

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

- 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

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

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

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

$$\nu_1 = \nu_0 + n,\ \nu_1 S_1 = \nu_0 S_0 + \left(y - Xb_1\right)^T \left(y - Xb_1\right)$$

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

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

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

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

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, …, Nとし、初期値b_0, B_0, ν_0, S_0を設定する

1. $p\left(\beta^{s+1} \mid \tau^{s}, y\right) \sim N\left(b_1^s, \left(1/\tau\right) B_1^s\right)$ から β^{s+1} を生成する
2. $p\left(\tau^{s+1} \mid \beta^{s+1}, y\right) \sim N\left(b_1^s, \left(1/\tau\right) B_1^s\right)$ から τ^{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

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

- 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

$$b_1 = \left(X^T X\right)^{-1} X^T y = \hat{\beta}$$

$$\nu_1 = n$$

$$S_1 = \frac{1}{\nu_1}\left(y - Xb_1\right)^T \left(y - Xb_1\right) = \hat{\sigma}^2$$

- すなわち、最小二乗解と同じ結果!!

線形回帰モデルのベイズ推定

- 推測統計学による線形回帰モデルの推定は、ベイズ推定において事前情報を与えない場合と同じ結果である
- 事前情報なしでベイズ推定する場合の事前分布を**無情報事前分布**(non-informative prior)という
- ベイズ統計では未知パラメータβがランダム変数であると考えるが、推測統計学では$\hat{\beta}$をランダム変数と考えているにほかならない

線形回帰モデルのベイズ推定

- 無情報事前分布は、積分しても1とならない(定数に収束しない)
- このような事前分布を非正則事前分布(improper prior)という
- 非正則事前分布の例として、一様分布が挙げられる
- 推測統計学による線形回帰モデルの不偏推定量は、事前情報に一様分布を与えた場合のベイズ推定とみなすこともできる

偏差情報量基準 (Deviance Information Criterion: DIC)

- モデルの当てはまりの良さを示す指標
- 未知パラメータθが得られた条件下でデータyが求められる確率p(y|θ)⇒つまり尤度関数
- 対数尤度を用いて得られる次式をデヴィアンスとよぶ

$$D(\theta) = -2\log p(y \mid \theta)$$

偏差情報量基準

- パラメータ点推定値に対するデヴィアンス

$$D\left(\bar{\theta}\right) = -2\log p\left(y \mid \bar{\theta}\right)$$

- デヴィアンスの平均値

$$\overline{D}\left(\theta\right) = E_{\theta|y}\left[-2\log p\left(y \mid \theta\right)\right]$$

偏差情報量基準

- 有効なパラメータ数

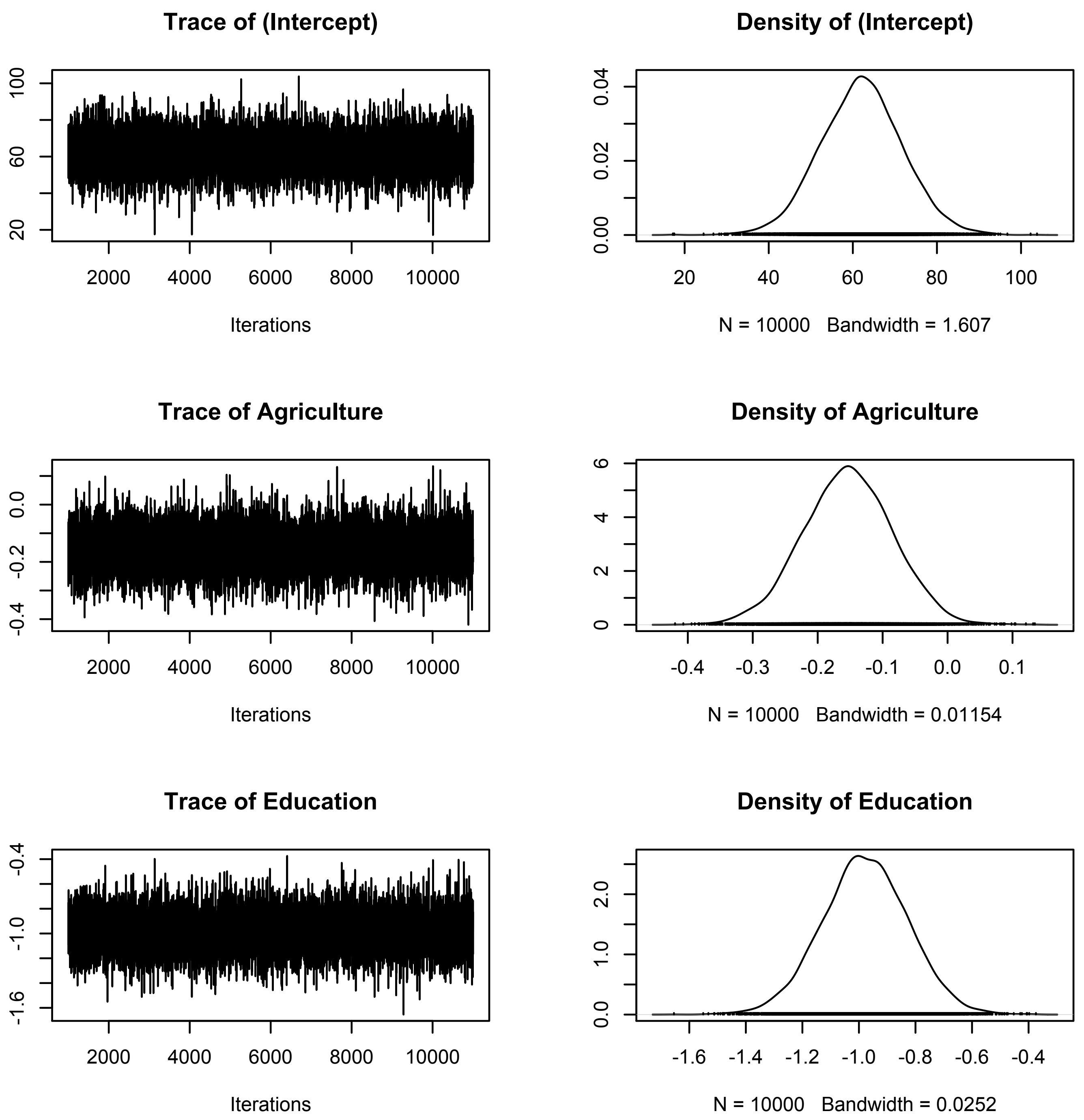
$$pD = \overline{D}\left(\theta\right) - D\left(\overline{\theta}\right)$$

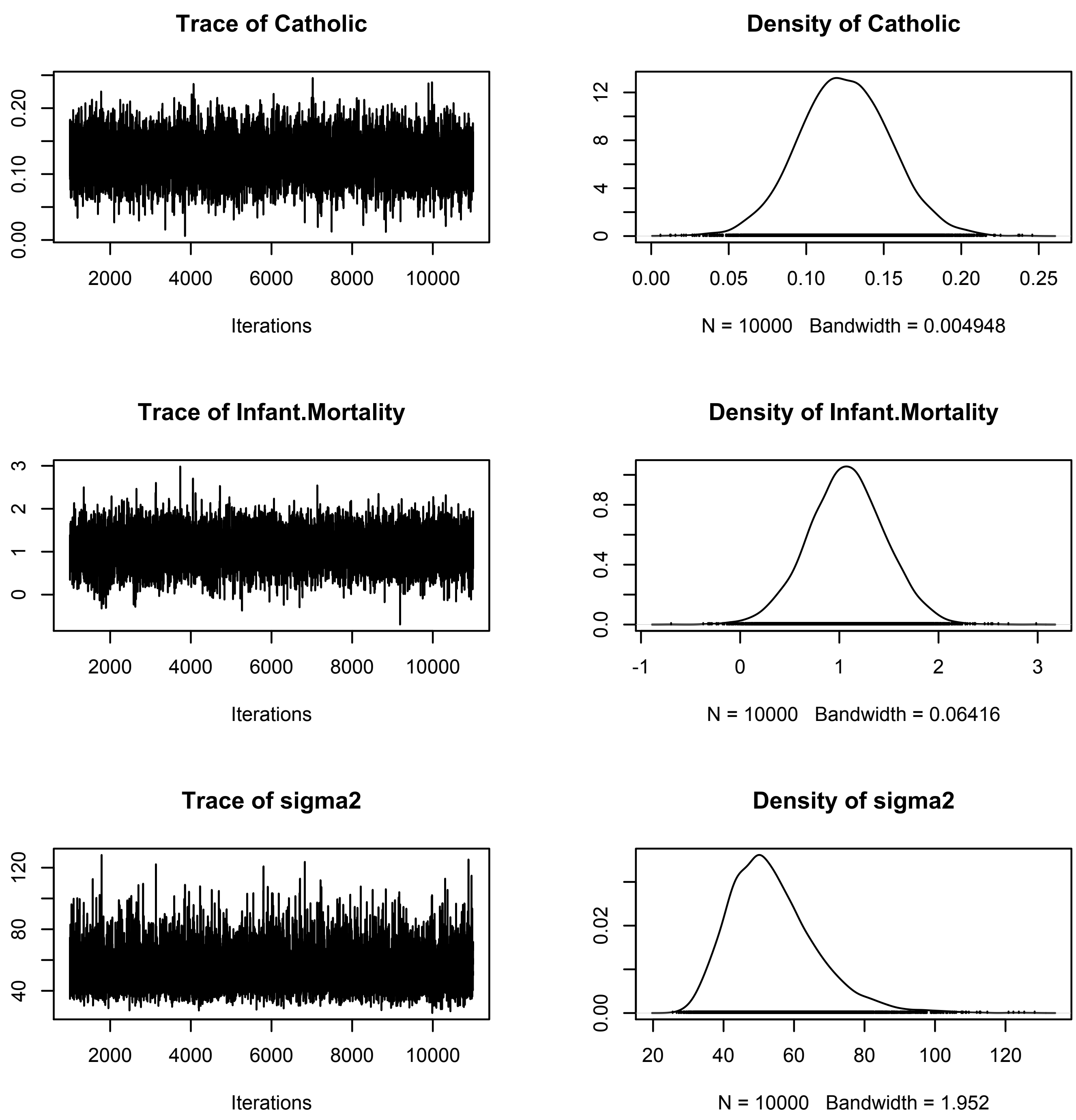
- 偏差情報量基準(DIC)

$$DIC = \overline{D}\left(\theta\right) + pD$$
$$= D\left(\overline{\theta}\right) + 2pD$$

線形回帰モデルのベイズ推定例

```
>  swiss.post <- MCMCregress(Fertility~Agriculture+Education+Catholic
+Infant.Mortality, data=swiss, burnin = 1000, mcmc = 10000, thin = 1)
>  summary(swiss.post)

Iterations = 1001:11000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                   Mean       SD  Naive SE Time-series SE
(Intercept)     62.0820  9.75835 0.0975835      0.0975835
Agriculture     -0.1538  0.06894 0.0006894      0.0006723
Education       -0.9782  0.15212 0.0015212      0.0015212
Catholic         0.1246  0.02945 0.0002945      0.0002945
Infant.Mortality 1.0753  0.38785 0.0038785      0.0038785
sigma2          53.9003 12.37564 0.1237564      0.1399208

2. Quantiles for each variable:

                    2.5%     25%     50%     75%    97.5%
(Intercept)     42.96315 55.5923 62.0890 68.4122 81.46942
Agriculture     -0.29126 -0.1995 -0.1535 -0.1074 -0.02082
Education       -1.27996 -1.0778 -0.9786 -0.8768 -0.68357
Catholic         0.06697  0.1049  0.1243  0.1444  0.18293
Infant.Mortality 0.30498  0.8190  1.0739  1.3307  1.83538
sigma2          34.79708 45.0713 52.1649 60.6398 83.16363
```


Trace of (Intercept)
Iterations
Density of (Intercept)
N = 10000 Bandwidth = 1.607
Trace of Agriculture
Iterations
Density of Agriculture
N = 10000 Bandwidth = 0.01154
Trace of Education
Iterations
Density of Education
N = 10000 Bandwidth = 0.0252


Trace of Catholic
Iterations
Density of Catholic
N = 10000   Bandwidth = 0.004948
Trace of Infant.Mortality
Iterations
Density of Infant.Mortality
N = 10000   Bandwidth = 0.06416
Trace of sigma2
Iterations
Density of sigma2
N = 10000   Bandwidth = 1.952

```
> geweke.diag(swiss.post2)

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

     (Intercept)      Agriculture        Education         Catholic
          0.5021          -0.2424           1.0638           0.9029
Infant.Mortality           sigma2
         -0.8999           1.1620

>
> raftery.diag(swiss.post2)

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

                  Burn-in  Total Lower bound  Dependence
                  (M)      (N)   (Nmin)       factor (I)
 (Intercept)      4        7482  3746         2.00
 Agriculture      4        7420  3746         1.98
 Education        4        7482  3746         2.00
 Catholic         4        7542  3746         2.01
 Infant.Mortality 4        7420  3746         1.98
 sigma2           4        7482  3746         2.00
```