§6.3 ベイズ回帰モデル
§6.4 階層モデル

相原　功昌

茨城大学

2008 年 10 月 17 日

ベイズ統計学によるアプローチ

1. 標本回帰係数 $\hat{\beta}_2$ に対して $P(\hat{\beta}_2 < 0)$
(通常の数理統計学的アプローチ)
$\longrightarrow$ 済
2. 母回帰係数 β_2 に対して $P(\beta_2 < 0)$
(ベイズ統計学的アプローチ)
$\longrightarrow$ これから

ベイズ統計学によるアプローチをするにあたって

- $\hat{\beta}_2$ の分布：

$$\mathcal{N}\left(\beta_2, \frac{\sigma^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}\right) \tag{6.2.4}$$

- $\hat{\beta}_1$ の分布：

$$\mathcal{N}\left(\beta_1, \frac{\sum x_i^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sigma^2\right) \tag{6.3.2}$$

回帰分析

$\hat{\beta}_1, \hat{\beta}_2$ の共分散：

$$Cov(\hat{\beta}_1, \hat{\beta}_2) = \frac{-\bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sigma^2 \tag{6.3.2}$$

以上は、回帰分析の標準的結果

$(\hat{\beta}_1, \hat{\beta}_2)$ の分布 $\rightarrow$ 2 次元正規分布
期待値：

$$(\beta_1, \beta_2) \tag{6.3.3}$$

分散・共分散行列：

$$\sigma^2 \cdot \begin{pmatrix} \frac{\sum x_i^2}{n\sum(x_i-\bar{x})^2} & \frac{-\hat{x}}{\sum(x_i-\bar{x})^2} \\ \frac{-\hat{x}}{\sum(x_i-\bar{x})^2} & \frac{1}{\sum(x_i-\bar{x})^2} \end{pmatrix} \tag{6.3.4}$$

同時確率分布

$(\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2)$ の同時確率分布
$\rightarrow$ 式 (6.3.5)

$$f(\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2 | \beta_1, \beta_2, \sigma^2) = \frac{1}{\sigma^n} \exp\left(-\frac{\mathcal{Q}}{2\sigma^2} \right) \tag{6.3.5}$$

$$\begin{aligned} \mathcal{Q} &= \mathcal{Q}(\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2) \\ &\equiv n(\hat{\beta}_1 - \beta_1)^2 + 2(\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2) \sum x_i \\ &\quad + (\hat{\beta}_2 - \beta_2)^2 \sum x_i^2 + (n-2)\hat{\sigma}^2 \end{aligned}$$

$\beta_1, \beta_2, \sigma^2$：未知の母数

$\rightarrow$ $\beta_1, \beta_2, \sigma^2$ をベイズ的に求める

事前分布

未知の母数に対して、事前分布をおく

- 共役事前分布
  $\longrightarrow$ 代表的だが、このケースには用意されていない
- 無情報事前分布
  $\longrightarrow$ 今回はこちらを用いる

β_1, β_2 の無情報事前分布→一様分布

$$w(\beta_1) \equiv c_1 \qquad -\infty < \beta_1 < \infty \qquad (c_1 > 0) \tag{6.3.6}$$

$$w(\beta_2) \equiv c_2 \qquad -\infty < \beta_2 < \infty \qquad (c_2 > 0) \tag{6.3.7}$$

c_1, c_2 は何でもよい

$$\int_{-\infty}^{\infty} w(\beta_1) d\beta_1 = \infty, \int_{-\infty}^{\infty} w(\beta_2) d\beta_2 = \infty, \tag{6.3.8}$$

事前情報

σ の無情報事前分布

$$w(\sigma) = \frac{1}{\sigma} \qquad (0 < \sigma < \infty) \tag{6.3.9}$$

$$\int_0^\infty w(\sigma)d\sigma = \int_0^\infty \frac{d\sigma}{\sigma} = \infty$$

式 (6.3.6)〜(6.3.9) は、変則な事前分布として認める
→ これらで、”何もわからない ”ことを表現

事後分布

ベイズの定理より、β_1, β_2, σ の事後分布
$\rightarrow$ 式 (6.3.10)

$$w(\beta_1)w(\beta_2)w(\sigma) \cdot f(\hat{\beta}_1, \hat{\beta}_2 | \beta_1, \beta_2, \sigma^2) \tag{6.3.10}$$

から、式 (6.3.11)

$$\begin{aligned} w'(\beta_1, \beta_2, \sigma^2 | \hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2) \propto \frac{1}{\sigma^{n+1}} & \\ & \exp\Big[-\frac{1}{2\sigma^2}\{n(\beta_1 - \hat{\beta}_1)^2 \\ & + 2(\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2)\sum x_i \\ & + (\beta_2 - \hat{\beta}_2)^2 \sum x_i^2 \\ & + (n-2)\hat{\sigma}^2\}\Big] \end{aligned} \tag{6.3.11}$$

が求まる

β_1 の分布

式 (6.3.11) から、単独に β_1 の分布を求める
$\rightarrow \sigma^2$ で積分消去

$$
\begin{aligned}
w'(\beta_1, \beta_2 | \hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2) &= \int_0^\infty w'(\beta_1, \beta_2, \sigma^2 | \hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2) d\sigma^2 \\
&\propto \Big\{ (n-2)\hat{\sigma}^2 + n(\beta_1 - \hat{\beta}_1)^2 \\
&+ 2(\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2) \sum x_i \\
&+ (\beta_2 - \hat{\beta}_2)^2 \sum x_i^2 \Big\}^{-\frac{n}{2}} \qquad (6.3.12)
\end{aligned}
$$

→「2 変量 t 分布」の 2 次元への拡張

β_1,β_2 に関する推論 1

β_1,β_2 をそれぞれ積分消去
→式 (6.3.13)、(6.3.14)

$$w'(\beta_1|\hat{\beta}_1,\hat{\beta}_2,\hat{\sigma}^2) \propto \left\{(n-2) + \frac{\sum(x_i-\bar{x})^2}{\frac{\hat{\sigma}^2\sum x_i^2}{n}}(\beta_1-\hat{\beta}_1)^2\right\}^{-\frac{(n-1)}{2}}$$
$$,(-\infty < \beta_1 < \infty) \quad (6.3.13)$$

$$w'(\beta_2|\hat{\beta}_1,\hat{\beta}_2,\hat{\sigma}^2) \propto \left\{(n-2) + \frac{\sum(x_i-\bar{x})^2}{\hat{\sigma}^2}(\beta_2-\hat{\beta}_2)^2\right\}^{-\frac{(n-1)}{2}}$$
$$,(-\infty < \beta_2 < \infty) \quad (6.3.14)$$

β_1,β_2 に関する推論 2

変換　$\beta_1 \to t, \beta_2 \to t'$

$$t = \left\{ \frac{\sum (x_i - \bar{x})^2}{\frac{\hat{\sigma}^2 \sum x_i^2}{n}} \right\}^{\frac{1}{2}} (\beta_1 - \hat{\beta}_1) \tag{6.3.15}$$

$$t' = \left\{ \frac{\sum (x_i - \bar{x})^2}{\hat{\sigma}^2} \right\}^{\frac{1}{2}} (\beta_2 - \hat{\beta}_2) \tag{6.3.16}$$

とすると、確率変数 t, t' は共に自由度 $n-2$ の t 分布 $t(n-2)$ を持つことがわかる
→ t 分布表より、β_1, β_2 に関する推論ができる

このように確率が計算できる
→ベイズ統計学のメリット

階層モデル

ベイズ統計学は確率分布が2階建てになっているモデルと考えられる

例1　事前分布を導入した回帰分析（前節）

1. データ：$\beta_1, \beta_2, \sigma^2$ が決まることから生まれる
2. $\beta_1, \beta_2, \sigma^2$：ある確率分布から決まる

例2　平均が θ の正規分布 $\mathcal{N}(\theta, \sigma^2)$ (σ^2 は決まっているとする)

3階　θ の事前分布にも超パラメータがあるはずなので
それの確率分布があるはず

2階　θ が出る事前分布によって、θ が求まる

1階　正規分布 $\mathcal{N}(\theta, \sigma^2)$ に従って y が求まる

このような考え→「階層モデル」

階層モデルとしてのベイズ統計学

階層モデルのベイズ統計学の展開の各場面

- 分散分析の変量効果モデル（分散成分）
- 因子分析のような潜在変数モデル

と考えられる
→１つの統一的視覚と言える

階層モデルは

- 無限に階を重ねるわけにはいかない
- モデルは本来単純であるべき
- '３階建て' が最も基本的
- '階' の代わりに「層」「レベル」を用いる

変量効果モデルとの関係

ある分散成分モデルを考える

- 平均の部分にクロス分類 (i, j) が入っている
- その相互作用がある

1 層 $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$

2A 層 $\alpha_i \sim \mathcal{N}(0, \tau_\alpha^2)$

2B 層 $\beta_j \sim \mathcal{N}(0, \tau_\beta^2)$

2C 層 $\gamma_{ij} \sim \mathcal{N}(0, \tau_\gamma^2)$

2D 層 $e_{ijk} \sim \mathcal{N}(0, \tau_\sigma^2)$

3 層 $\mu \sim w_\mu, \tau_\alpha^2 \sim w_{\tau\alpha},\ \tau_\beta^2 \sim w_{\tau\beta}, \tau_\gamma^2 \sim w_{\tau\gamma},\ \sigma^2 \sim w_\sigma$

- w は、全てわかっている確立分布
- '$\sim$' は' ○○分布に従う' を意味する

→ 効果が線型でちょうど交互作用のある二元配置分散分析の変量効果モデル

二元配置分散分析の変量効果モデル

分散分析とは
→ 変数の各水準の母平均の違いを「分散」の差で分析するもの

二元配置
グループを識別する要素を２つにして、分散分析するもの
例

	アメリカ	フランス	日本
男性	○○ cm	○○ cm	○○ cm
女性	○○ cm	○○ cm	○○ cm

グループを識別する要素：性別、父親の身長

例のデータを考えた場合，分析の目的は

1. 性別によって平均身長に違いはあるのか
2. 父親の身長の高低によって平均身長に違いはあるのか
3. ２つの要素による相乗効果はあるのか (交互作用)

一元配置変量効果モデル

線型の変量効果モデル (最も簡単な階層モデル) を考える

$y_{ij} = \mu + \alpha_i + e_{ij}$, $\alpha_i \sim \mathcal{N}(0, \sigma_A^2)$, $e_{ij} \sim \mathcal{N}(0, \sigma^2)$
$(i = 1, 2, \ldots, I; j = 1, 2, \ldots, J)$

具体例
ランダムにとられた $(I = 6)$ 通りのバッチからランダムにとられた $(J = 5)$ 観測値 → 表 6.2

表 6.2　最も簡単な階層モデル

バッチ	$(i =)1$	2	3	4	5	6
$(j = 1)$	1545	1540	1595	1445	1595	1520
2	1440	1555	1550	1440	1630	1455
3	1440	1490	1605	1595	1515	1450
4	1520	1560	1510	1465	1635	1480
5	1580	1495	1560	1545	1625	1445
$\bar{y}_{i.}$	1505	1528	1564	1498	1600	1470
$\bar{y}_{..}$	1527.5					(g)

一元配置変量効果モデルの具体例

一元配置分散分析の変量効果モデルでみると

$$S_0 = \sum_i \sum_j (y_{ij} - \bar{y})^2, S_A = J \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2, S_E = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2$$

として

$$E\left(\frac{S_A}{I-1}\right) = \sigma^2 + J\sigma_A^2, E\left(\frac{S_E}{IJ-1}\right) = \sigma^2$$

が示される

前式の片方に σ_A^2 が含まれるので、その後の扱いが異なり

$$\hat{\sigma}^2 = \frac{S_E}{IJ-1}, \hat{\sigma}_A^2 = \frac{1}{J}\left\{\left(\frac{S_A}{I-1} - \frac{S_E}{IJ-1}\right)\right\}$$

となる

計算

数値を計算する → 表 6.4

表 6.4 計算結果の一部

変動原因	変動 (平方和)	自由度	平均平方和	同期待値
バッチ間	$S_A = 56357.50$	5	11271.50	$\sigma^2 + 5\sigma_A^2$
バッチ内	$S_E = 58830.50$	29	2028.62	σ^2

$$\hat{\sigma}^2 = 2028.62, \hat{\sigma}_A^2 = \frac{11271.50 - 2028.62}{5} = 1848.58$$

平均平方和を比較すると $\dfrac{11271.50}{2028.62} = 5.56$

→ バッチ間で $\sigma_A^2 > 0$ であることが十分に可能な大きい値

バッチ間の平均平方和はバッチ内を大きく上回らなければならない

難点を含むケース

変量効果モデルの分析法には難点がある

<u>具体例</u>

表 6.3 のデータ：$I = 6, J = 5$

$$S_A = 41.6816, S_E = 358.7014, \hat{\sigma}^2 = 12.3690$$

$$\hat{\sigma}_A^2 = \frac{8.3363 - 12.3690}{5} = -0.80654$$

表 6.3　最も簡単な階層モデル (難点を含むケース)

バッチ	$(i=)1$	2	3	4	5	6
$(j=1)$	7.298	5.220	0.110	2.212	0.282	1.722
2	3.846	6.556	10.386	4.852	9.014	4.782
3	2.434	0.608	13.434	7.092	4.458	8.106
4	9.566	11.788	5.510	9.288	9.446	0.758
5	7.990	-0.892	8.166	4.980	7.198	3.758
$\bar{y}_{i.}$	6.2268	4.6560	7.5212	5.6848	6.0796	3.8252
$\bar{y}_{..}$	5.6656					

難点

平均平方和： $\dfrac{41.6816/5}{358.7014/29} = 0.674$
バッチ間とバッチ内の平均平方和の差が小さい

このような場合は分析を放棄するのではなく、階層モデルを用いれば十分に有意味な解析が可能
→ 階層モデルが十分な一般性をもつことの１つの証拠

色つきは、テキストと違うところを表す