政治学方法論 II

第 8 回：階層モデル

矢内　勇生

法学部・法学研究科

2015 年 6 月 3 日

神戸大学

今日の内容

1. 階層モデル 1：二項分布モデル
   - 共役階層モデルのベイズ推定
   - ネズミの腫瘍実験：Section 5.1 の例 (pp.102–104)

2. 階層モデル 2：正規分布モデル
   - 交換可能な母数を正規分布モデルで推定する

階層モデル

例：ネズミの腫瘍実験で考える

- 尤度：$p(y|\theta)$
- 母数分布：自然な共役自然分布 $p(\theta|\phi)$
- 超事前分布：$p(\phi)$
- 同時事後分布：

$$
\begin{aligned}
p(\theta,\phi|y) &\propto p(y|\theta,\phi)p(\theta,\phi) && (1)\\
&= p(y|\theta)p(\theta,\phi) && (2)
\end{aligned}
$$

- 解析と数値計算の両者を使って事後分布を求めてみよう！

解析によって条件付き分布と周辺分布を求める

1. 同時事後分布 $p(\theta,\phi|y)$ を、尤度、母数分布、超事前分布を使って表す
2. 条件付き確率、$p(\theta|\phi,y)$ を求める

$$p(\theta|\phi,y) = \prod_{j=1}^{J} p(\theta_j|\phi,y) \tag{3}$$

3. $p(\phi|y)$ をベイズルールによって求める

$$p(\phi|y) = \int p(\theta,\phi|y)d\theta \tag{4}$$

or

$$p(\phi|y) = \frac{p(\theta,\phi|y)}{p(\theta|\phi,y)} \tag{5}$$

シミュレーションによって事後分布を抽出する

1. $p(\phi|y)$ からベクトル ϕ を引く
2. 抽出した ϕ を基に、$p(\theta|\phi, y)$ からベクトル θ を引く：
   $p(\theta|\phi, y) = \prod_j p(\theta_j|\phi, y)$ だから、θ_j を 1 つずつ引けばよい
3. 抽出した ϕ を基に、事後予測値 $\tilde{y}$ を引く

この作業を L 回繰り返せば、事後分布が得られる

ネズミの腫瘍実験を分析する

- y_j：実験 j において、n_j 匹のネズミのうち、腫瘍をもったネズミの数
- $j = 1, \ldots, J$, $J = 71$
- サンプリングモデル（尤度）：

$$y_j \sim \mathrm{Bin}(n_j, \theta_j)$$

- θ_j はベータ分布から独立に生まれるとする：

$$\theta_j \sim \mathrm{Beta}(\alpha, \beta)$$

- (α, β) については特別な情報がないので、無情報超事前分布を使う

解析 1：同時事後分布

$$
\begin{aligned}
p(\theta,\alpha,\beta|y) &\propto p(\alpha,\beta)p(\theta|\alpha,\beta)p(y|\theta,\alpha,\beta) \\
&\propto p(\alpha,\beta) \\
&\quad \times \prod_{j=1}^{J} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_j^{\alpha-1}(1-\theta_j)^{\beta-1} \\
&\quad \times \prod_{j=1}^{J} \theta_j^{y_j}(1-\theta_j)^{n_j-y_j}
\end{aligned}
$$

- ただし、$\Gamma(.)$ はガンマ関数

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$$

- 自然数 n について、

$$\Gamma(n+1) = n!$$

解析 2：θ の条件付き分布

$$\theta_j | \alpha, \beta, y \sim \mathrm{Beta}(\alpha + y_j, \beta + n_j - y_j)$$

$$p(\theta | \alpha, \beta, y) = \prod_{j=1}^{J} \frac{\Gamma(\alpha + \beta + n_j)}{\Gamma(\alpha + y_j)\Gamma(\beta + n_j - y_j)} \theta_j^{\alpha + y_j - 1} (1 - \theta_j)^{\beta + n_j - y_j - 1}$$

解析 3：(α,β) の周辺事後分布

$$
\begin{aligned}
p(\alpha,\beta|y) &= \frac{p(\theta,\alpha,\beta|y)}{p(\theta|\alpha,\beta,y)} \\
&\propto p(\alpha,\beta)\prod_{j=1}^{J}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(\alpha+y_j)\Gamma(\beta+n_j-y_j)}{\Gamma(\alpha+\beta+n_j)}
\end{aligned}
$$

$p(\alpha,\beta)$ の設定 (1)

- 超事前分布 (α,β) を選びたい
- θ_j が従うベータ分布の平均とサンプルサイズによって特定する
- 平均：$\alpha/(\alpha+\beta)$
- サンプルサイズ：$\alpha+\beta$
- 無情報：$(-\infty,\infty)$ で考えたい
- しかし・・・

$$0<\frac{\alpha}{\alpha+\beta}<1, \quad \alpha+\beta>0$$

- 変数変換する!

$p(\alpha,\beta)$ の設定 (2)

- $(0,1) \to (-\infty,\infty)$ にする変換：ロジット変換

$$\mathrm{logit}\left(\frac{\alpha}{\alpha+\beta}\right) = \log\left(\frac{\frac{\alpha}{\alpha+\beta}}{1-\frac{\alpha}{\alpha+\beta}}\right) = \log\left(\frac{\alpha}{\beta}\right)$$

- $(0,\infty) \to (-\infty,\infty)$ にする変換：対数変換

$$\log(\alpha+\beta)$$

- $p\left(\log(\alpha/\beta), \log(\alpha+\beta)\right) \propto 1$ とする
- 残念ながら、この超事前分布は improper な事後分布を生む：使えない

$p(\alpha, \beta)$ の設定 (3)

- 代わりに、次のように設定する

$$p\left(\frac{\alpha}{\alpha+\beta}, (\alpha+\beta)^{-\frac{1}{2}}\right) \propto 1$$

- ヤコビアンを求めて変換する：

$$p(\alpha, \beta) \propto (\alpha+\beta)^{-\frac{5}{2}}$$

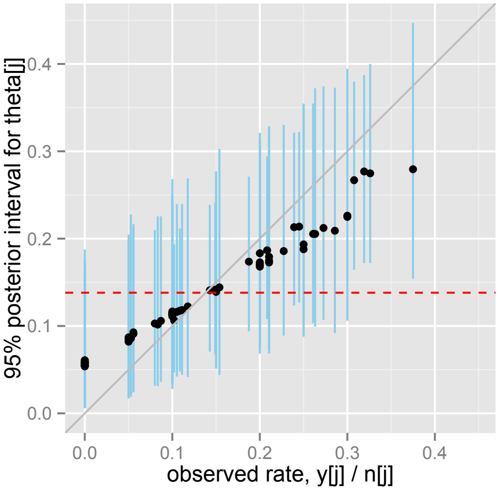
- これをさらに変換すると、

$$p\left(\log(\frac{\alpha}{\beta}), \log(\alpha, \beta)\right) \propto \alpha\beta(\alpha+\beta)^{-\frac{5}{2}}$$

事後分布の抽出

- 超事前分布が設定できたので、$p(\alpha, \beta|y)$ より (α, β) を抽出する
- (α, β) が抽出できれば、$p(\theta_j|\alpha, \beta, y)$ が特定できるので、θ_j を抽出する
- この作業を繰り返し、θ_j の事後分布を得る
- 導出の詳細は授業のウェブサイトで

推定値の収縮 (shrinkage)



問題の設定

- J 個の独立した集団（実験）
- 集団 $j = 1, \ldots, J$ における n_j 個の観測値 y_{ij} から θ_j を推定する
- 誤差の分散 σ^2 は既知であるとする

$$y_{ij} | \theta_j \sim \mathrm{N}(\theta_j, \sigma^2),$$

for $i = 1, \ldots, n_j$; $j = 1, \ldots, J$

- 集団 j における観測値の平均値とその分散は、

$$\bar{y}_{.j} = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}, \quad \sigma_j^2 = \frac{\sigma^2}{n_j}$$

尤度（サンプリングモデル）

- 十分統計量：$\bar{y}_{.j}$
- 十分統計量を使って θ_j の尤度を表すと、

$$\bar{y}_{.j} | \theta_j \sim \mathrm{N}(\theta_j, \sigma_j^2)$$

- σ^2 と n_j が既知なので、σ_j^2 も既知

理に適った推定値は？

$\theta = (\theta_1, \dots, \theta_J)$ の事後分布について、どんな推定値なら受け入れられそう？

- 集団ごとの平均値 (no pooling / complete separating)

$$\bar{y}_{.j} = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$$

- 全体の平均値 (complete pooling)

$$\bar{y}_{..} = \frac{\sum_{j=1}^{J} \frac{1}{\sigma_j^2} \bar{y}_{.j}}{\sum_{j=1}^{J} \frac{1}{\sigma_j^2}}$$

- 「伝統的な」方法：ANOVA によってどちらか選ぶ（BDA3: p.114 参照）
  - 集団間の変動 (between variance) が集団内の変動 (withing variance) に対して大きい（F 検定）とき：no pooling
  - それ以外：complete pooling

折衷案：partial pooling

- Bayesian：no pooling と complete pooling を組み合わせる

$$\hat{\theta}_j = \lambda_j \bar{y}_{.j} + (1 - \lambda)\bar{y}_{..},$$

  ただし、$\lambda_j \in [0, 1]$

- これを実現する事前分布
  1. 独立な $\theta_j \sim \mathrm{U}(-\infty, \infty) \to \lambda_j = 1$：no pooling
  2. $\theta_1 = \cdots = \theta_j = \theta \sim \mathrm{U}(\alpha, \beta) \to \lambda_j = 0$：complete pooling
  3. $\theta_j \overset{\text{iid}}{\sim} \mathrm{N}(\mu_0, \sigma_0^2) \to \lambda_j \in [0, 1]$：partial pooling
- No pooling と complete pooling は partial pooling の特殊ケース

階層モデル

- $\theta_j \overset{\text{iid}}{\sim} \mathrm{N}(\mu, \tau^2)$ と仮定する
- 超母数：μ, τ
- θ_j は (μ, τ) が与えられたときに条件付き独立

$$p(\theta_1, \ldots, \theta_J | \mu, \tau) = \prod_{j=1}^{J} \mathrm{N}(\theta_j | \mu, \tau^2)$$

$$p(\theta_1, \ldots, \theta_J) = \int \prod_{j=1}^{J} [\mathrm{N}(\theta_j | \mu, \tau^2)] p(\mu, \tau) d(\mu, \tau)$$

- τ が所与のときの μ の超事前分布を無情報と仮定する

$$p(\mu, \tau) = p(\mu | \tau) p(\tau) \propto p(\tau)$$

同時事後分布

- 十分統計量 $\bar{y}_{.j}$ を使って同時事後分布を表す

$$
\begin{aligned}
p(\theta,\mu,\tau|y) &\propto p(\mu,\tau)p(\theta|\mu,\tau)p(y|\theta) \\
&\propto p(\mu,\tau)\prod_{j=1}^{J}\mathrm{N}(\theta_j|\mu,\tau^2)\prod_{j=1}^{J}\mathrm{N}(\bar{y}_{.j}|\theta_j,\sigma_j^2)
\end{aligned}
$$

θ_j の条件付き事後分布

- (μ, τ) が与えられたとき、各 θ_j は独立
- θ_j の条件付き事後分布は、

$$\theta_j | \mu, \tau, y \sim \mathrm{N}(\hat{\theta}_j, V_j)$$

ただし、

$$\hat{\theta}_j = \frac{\frac{1}{\sigma_j^2}\bar{y}_{.j} + \frac{1}{\tau^2}\mu}{\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}}, \quad V_j = \frac{1}{\frac{1}{\sigma_j^2} + \frac{1}{\tau^2}}$$

- 事後分布の平均は、データの平均と事前分布の平均の精度 (precision) を重みとする加重平均
- 事後分布の精度は、データの精度と事前分布の精度の和

超母数の周辺事後分布

- 完全なベイズ推定のために、(μ, τ) を母数として扱う
- データから得られる情報を使って超母数の分布を得る

$$p(\mu, \tau | y) \propto p(y | \mu, \tau) p(\mu, \tau)$$

- 一般的には、$p(y|\mu, \tau)$ が特定できない：この方法は使えない
- 正規分布モデルの場合は特定できる

$$\bar{y}_{.j} | \mu, \tau \sim \mathrm{N}(\mu, \sigma_j^2 + \tau^2)$$

- 超母数の周辺事後分布は、

$$p(\mu, \tau | y) \propto p(\mu, \tau) \prod_{j=1}^{J} \mathrm{N}(\bar{y}_{.j} | \mu, \sigma_j^2 + \tau^2)$$

μ の条件付き事後分布

- τ が与えられたときの μ の条件付き事前分布を一様分布とする

$$p(\mu|\tau) \propto 1$$

- 超母数 τ が与えられたときの μ の条件付き事後分布は、

$$p(\mu|\tau, y) \propto p(\mu|\tau)p(y|\mu,\tau) \propto p(y|\mu,\tau)$$

- したがって、μ の条件付き事後分布は正規分布に従う

$$\mu|\tau, y \sim \mathrm{N}(\hat{\mu}, V_\mu)$$

ただし、

$$\hat{\mu} = \frac{\sum_{j=1}^{J} \frac{1}{\sigma_j^2+\tau_2}\bar{y}_{\cdot j}}{\sum_{j=1}^{J} \frac{1}{\sigma_j^2+\tau_2}}, \quad V_\mu = \frac{1}{\sum_{j=1}^{J} \frac{1}{\sigma_j^2+\tau^2}}$$

τ の事後分布

- τ の事後分布は、

$$p(\tau|y) = \frac{p(\mu,\tau|y)}{p(\mu|\tau,y)}$$
$$\propto \frac{p(\tau)\prod_{j=1}^{J}\mathrm{N}(\bar{y}_{.j}|\mu,\sigma_j^2+\tau^2)}{\mathrm{N}(\mu|\hat{\mu},V_\mu)}$$

- $\mu = \hat{\mu}$ としてさらに簡略化する

$$p(\tau|y) \propto \frac{p(\tau)\prod_{j=1}^{J}\mathrm{N}(\bar{y}_{.j}|\hat{\mu},\sigma_j^2+\tau^2)}{\mathrm{N}(\hat{\mu}|\hat{\mu},V_\mu)}$$
$$\propto p(\tau)V_\mu^{\frac{1}{2}}\prod_{j=1}^{J}(\sigma_j^2+\tau^2)^{-\frac{1}{2}}\exp\left(-\frac{(\bar{y}_{.j}-\hat{\mu})^2}{2(\sigma_j^2+\tau^2)}\right)$$

τ の事前分布を設定する

- τ の事前分布を設定する
- 無情報分布を仮定、

$$p(\tau) \propto 1$$

- 超事前分布が improper なので、事後分布が proper になるかどうか確認する必要がある
- $p(\log \tau) \propto 1$ にすると、事後分布は improper になってしまう

事後分布をシミュレートする

- 事後分布は以下のとおり分解できる

$$p(\theta, \mu, \tau | y) = p(\tau | y) p(\mu | \tau, y) p(\theta | \mu, \tau, y)$$

- ここまでの分析で、右辺の各要素がわっている：シミュレーションで事後分布を求める
  1. τ を抽出する
  2. 抽出した τ を用い、μ を抽出する
  3. 抽出した (μ, τ) を使い、θ を抽出する
  4. これを各集団 j に対して行う
  5. 一連の抽出を L 回行い、事後分布を得る