政治学方法論 II

第 3 回：1 変量のモデル

矢内　勇生

法学部・法学研究科

2015 年 4 月 22 日

神戸大学

今日の内容

1. 1 変量のベイズ推定
   - ベイズルールを利用した推定
   - 例：二項分布の母数の推定
2. 正規分布の平均値の推定
   - 観測値が 1 つの場合
   - 観測値が複数ある場合

ベイズデータ分析の手順

1. リサーチクエスチョンに答えられるデータを特定する
2. 特定したデータを記述する確率モデルを作る
3. 推定する母数の事前分布を特定する
4. ベイズルールに従い、情報を更新する
5. 事後予測分布がデータにフィットしているか確かめる

情報の更新

$p(\theta|y)$：θ の事後分布　これを知りたい！

ベイズルール

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

ベイズルールの基本

θ の事後分布 $\propto$ θ の尤度 $\times$ θ の事前分布

ベイズ推定に必要なもの

- $p(\theta)$：θ の事前分布 (prior)
- $p(y|\theta) = L(\theta|y)$：$y$ のサンプリングモデル (データ生成過程)、θ の尤度関数
- $p(y)$：y の事前予測分布 (prior predictive)

問題の設定

二項分布 (binomial distribution) の成功確率の推定

一連のベルヌーイ試行の結果から、成功確率 θ を推定する

- 観測データ（実現値）：$y_i \in \{0, 1\}$, for $i = 1, 2, \ldots, n$
- y_i は互いに独立で同一の分布に従う：y_i は交換可能 (exchangeable)
- 使用するデータ：$y = \sum_{i=1}^{n} y_i$

データを記述する確率モデル

データ生成の確率モデル、サンプリングモデル、DGP

DGP： $y_i \overset{\text{iid}}{\sim} \text{Bern}(\theta)$

$$
\begin{aligned}
p(y|\theta) &= \text{Bin}(y|n,\theta) && (3)\\
&= \binom{n}{y}\theta^y(1-\theta)^{n-y} && (4)\\
&\propto \theta^y(1-\theta)^{n-y} && (5)
\end{aligned}
$$

θ の事前分布を特定する

- θ についての事前知識が（あまり）ないとする
- θ 自体が確率：$0 \leq \theta \leq 1$
- 区間 $[0,1]$ は同様に確からしい（特定の値に他の値よりも大きな信頼をおく根拠が無い）
- 区間 $[0,1]$ の連続型一様分布

$$\theta \sim \mathrm{U}(0,1) \tag{6}$$

$$p(\theta) = \mathrm{U}(\theta|0,1) = 1 \tag{7}$$

ベイズルールに従い、情報を更新する

θ の事後分布 $\propto$ θ の尤度 $\times$ θ の事前分布

$$
\begin{aligned}
p(\theta|y) &\propto p(y|\theta)p(\theta) && (8)\\
&\propto \theta^{y}(1-\theta)^{n-y} && (9)\\
&\propto \theta^{(y+1)-1}(1-\theta)^{(n-y+1)-1} && (10)
\end{aligned}
$$

ベータ分布 (beta distributions)

ベータ分布の PDF $(0 \leq x \leq 1)$

$$p(x|\alpha,\beta) = \mathrm{Beta}(x|\alpha,\beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du} \tag{11}$$

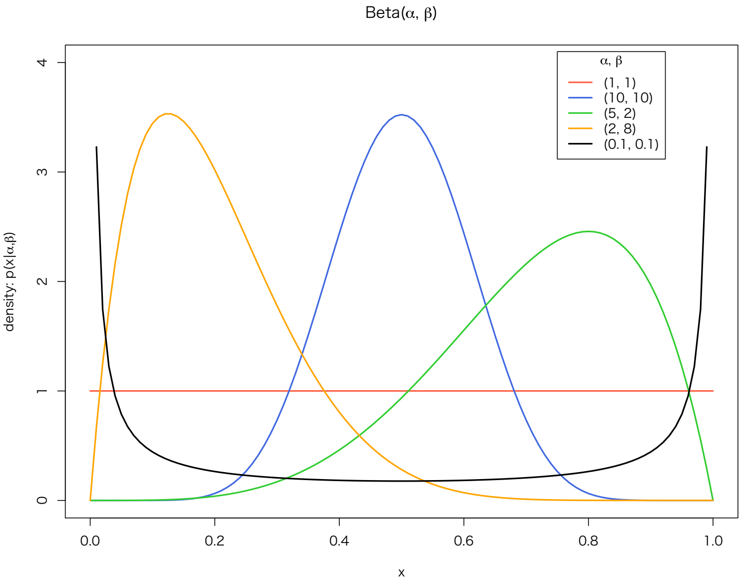
$$= \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} \tag{12}$$

パラメタ：$\alpha > 0,\ \beta > 0$

$$\mathrm{E}(X) = \frac{\alpha}{\alpha+\beta}$$

$$\mathrm{var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

例：二項分布の母数の推定



二項分布の母数 θ の事後分布

$$p(\theta|y) \propto \theta^{(y+1)-1}(1-\theta)^{(n-y+1)-1}$$

よって、

$$\theta|y \sim \mathrm{Beta}(y+1, n-y+1)$$

したがって、

$$\mathrm{E}(\theta|y) = \frac{y+1}{(y+1)+(y-n+1)} = \frac{y+1}{n+2}$$

$$\mathrm{var}(\theta|y) = \frac{(y+1)(n-y+1)}{(n+2)^2(n+3)}$$

事後情報：事前情報とデータの間

- ベイズ推定：事前情報とデータの折り合いをつける
- 事後情報：事前情報とデータの加重平均
- 重み：データのサイズが大きいほど、データの重みが増す
- サンプルサイズが十分大きくなると、事前情報の重みは取るに足らないものに
- ただし、事前情報で確率 0 が与えられた点は、事後分布から除外される：データサイズが大きくても、事前分布には意味がある

実習：授業用 GitHub リポジトリ にある `fair-coin.R` を使って、ベイズ分析の基本を理解しよう！

データの確率モデルの特定

- 1 つの観測値 y
- サンプリングモデル：

$$y \sim \mathrm{N}(\theta, \sigma^2)$$

観測値が 1 つの場合の θ の尤度

$$p(y|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y-\theta)^2}{2\sigma^2}\right] \tag{13}$$

事前分布の特定

- θ の事前情報
- アップデートすることを見越して、計算しやすい事前分布を考える：尤度と同族を使う
- 正規分布：指数分布族
- 事前分布も指数分布族で特定する

$$\theta \sim \mathrm{N}(\mu_0, \tau_0^2)$$

θ の事前分布

$$p(\theta) = \frac{1}{\sqrt{2\pi\tau_0^2}} \exp\left[-\frac{(\theta-\mu_0)^2}{2\tau_0^2}\right] \tag{14}$$

情報の更新

$$p(\theta|y) \propto p(y|\theta)p(\theta) \tag{15}$$

$$\propto \exp\left[-\frac{1}{2}\left\{\frac{(y-\theta)^2}{\sigma^2}+\frac{(\theta-\mu_0)^2}{\tau_0^2}\right\}\right] \tag{16}$$

$$\propto \exp\left[-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2\right] \tag{17}$$

ただし、

$$\mu_1 = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{1}{\sigma^2}y}{\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}}, \quad \frac{1}{\tau_1^2} = \frac{1}{\tau_0^2} + \frac{1}{\sigma^2}$$

したがって、$\theta|y \sim \mathrm{N}(\mu_1, \tau_1^2)$

データの確率モデル

- 観測値：$y = (y_1, y_2, \ldots, y_n)$
- サンプリングモデル：

$$y_i \overset{\text{iid}}{\sim} \mathrm{N}(\theta, \sigma^2)$$

データ y に対する θ の尤度関数

$$p(y|\theta) = \prod_{i=1}^{n} p(y_i|\theta) \tag{18}$$

$$\propto \prod_{i=1}^{n} \exp\left[-\frac{(y_i-\theta)^2}{2\sigma^2}\right] \tag{19}$$

事前分布と事後分布

- 事前分布：観測値が 1 つのときと同じ

$$p(\theta) \propto \exp\left[-\frac{(\theta-\mu_0)^2}{2\tau_0^2}\right]$$

$$\begin{aligned}
p(\theta|y) &\propto p(\theta)p(y|\theta) &(20)\\
&\propto \exp\left[-\frac{(\theta-\mu_0)^2}{2\tau_0^2}\right]\prod_{i=1}^{n}\exp\left[-\frac{(y_i-\theta)^2}{2\sigma^2}\right] &(21)\\
&\propto \exp\left[-\frac{1}{2}\left\{\frac{1}{\tau_0^2}(\theta-\mu_0)^2+\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\theta)^2\right\}\right] &(22)
\end{aligned}$$

事後分布は $\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$ を通じてのみ y に依存する！

十分統計量 (sufficient statistic)

- 現在検討中のデータ：すべての y_i ではなく、$\bar{y}$ だけ考えればよい

十分統計量

$\boldsymbol{x}$ について統計量 $t = t(\boldsymbol{x})$ が、以下の条件を満たすとき、t を x が与えられた条件下での θ に対する十分統計量と呼ぶ

$$p(\boldsymbol{x}|\theta) = p(t|\theta)p(x|t) \tag{23}$$

- 現在検討中の問題：$\bar{y}$ は θ に対する十分統計量である

十分統計量を使って問題を捉え直す

- $\bar{y}$ という観測値が 1 つあると考える
- サンプリングモデル：$\bar{y} \sim \mathrm{N}(\theta, \frac{\sigma^2}{n})$

したがって、

$$p(\theta|y_1, \ldots, y_n) = p(\theta|\bar{y}) = \mathrm{N}(\theta|\mu_n, \tau_n^2)$$

ただし、

$$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}, \quad \frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}$$

- n が一定で $\tau_0 \to \infty$ のとき、または t_0 が一定で $n \to \infty$ のとき、

$$p(\theta|y) \approx \mathrm{N}\left(\theta|\bar{y}, \frac{\sigma^2}{n}\right)$$