政治学方法論 II

第 4 回：事前分布

矢内　勇生

法学部・法学研究科

2015 年 4 月 29 日

神戸大学

今日の内容

1. 事前分布
   - 事前分布の役割
2. 共役事前分布
   - 共役事前分布 (conjugate priors)
   - パラメトリック近似
   - データを用いた事前分布の設定
3. 事後分布への影響を弱めた事前分布
   - 無情報事前分布
   - 弱情報事前分布

事前分布 (prior distributions)

- ベイズ推定に不可欠：事後分布 $\propto$ 尤度 $\times$ 事前分布
- 事前分布とデータさえあれば、半自動的に事後分布が得られる

事前分布とは何か？

1. パラメタ θ の母集団
   - θ は事前分布に従う確率変数であり、特定の θ はその実現値
2. 知識の状態
   - θ の値が事前分布からの実現値と考えられるように、知識を表現する

事前分布に対する批判

- 恣意的である
  - 科学は主観的な信念ではなく客観的な事実を扱うべき
  - 人によって異なる事前分布を想定し得るのはおかしい
  - （ベイズ分析そのものに対する懐疑の源）
- 回りくどい：事後分布を「信念」に基づいて選べばいいのでは？
- 共役事前分布の場合：「信念」ではなく、計算の簡便性で選んでいるだけでは？
- 共役でない事前分布の場合：式が美しくない（数値計算 [だけで] は信用ならない）
- etc.

事前分布を選ぶ基準

事前分布の選択の際に気をつけること：

- 知識のどの側面を事前分布に反映させるか
- 事後分布にどのような影響を与えるか

事前分布を特定することの難しさ：

- 知識自体は確率分布とは限らない
- 事前情報を確率分布に置き換える方法は 1 つではない

共役事前分布 (conjugate priors)

定義：共役分布

$p \in \mathscr{F}$ なるすべての確率分布 $p(.)$ について $p(\theta|y) \in \mathscr{F}$ となるとき、確率分布族 $\mathscr{F}$ は尤度関数（サンプリングモデル）$p(y|\theta)$ と共役 (conjugate) であるという。

- 最も単純な例：$\mathscr{F}$ をすべての確率分布とする
- $\mathscr{F}$ をできるだけ小さな集合に限定したほうが、有用性が高い
- 自然な共役分布 (natural conjugate)：関数型が同じもの

共役分布の例

表: 指数分布族の自然な共役事前分布の例

尤度関数 $f(y\|\theta)$	事前分布 $p(\theta)$	事後分布 $p(\theta\|y)$
$\mathrm{N}(\theta,\sigma^2)$	$\mathrm{N}(\mu_0,\tau_0^2)$	$\mathrm{N}\left(\frac{\tau_0^2 y+\sigma^2\mu_0}{\sigma^2+\tau_0^2}, \frac{\sigma^2\tau_0^2}{\sigma^2+\tau_0^2}\right)$
$\mathrm{N}(\mu,1/\theta)$	$\mathrm{Gamma}(\alpha,\beta)$	$\mathrm{Gamma}\left(\alpha+0.5,\beta+\frac{(\mu-y)^2}{2}\right)$
$\mathrm{Poisson}(\theta)$	$\mathrm{Gamma}(\alpha,\beta)$	$\mathrm{Gamma}(\alpha+y,\beta+1)$
$\mathrm{Gamma}(\nu,\theta)$	$\mathrm{Gamma}(\alpha,\beta)$	$\mathrm{Gamma}(\alpha+\nu,\beta+y)$
$\mathrm{Bin}(n,\theta)$	$\mathrm{Beta}(\alpha,\beta)$	$\mathrm{Beta}(\alpha+y,\beta+n-y)$
$\mathrm{Neg\text{-}bin}(m,\theta)$	$\mathrm{Beta}(\alpha,\beta)$	$\mathrm{Beta}(\alpha+m,\beta+y)$
$\mathrm{M}_k(\theta_1,\ldots,\theta_k)$	$\mathrm{Dir}(\alpha_1,\ldots,\alpha_k)$	$\mathrm{Dir}(\alpha_1+y_1,\ldots,\alpha_k+y_k)$

共役事前分布を使う理由

- 不変原理 (invariance principle)
  - 元々事前情報 $p(\theta)$ を持っている
  - データ y で情報で更新し,$p(\theta|y)$ を得る：データ y が信念を変える
  - データ y の情報は、完全ではない：分布の「構造」を変えてしまうほど信用出来ない
  - 事前分布の「構造」を残しつつ、パラメタの値を更新する
- 計算上の優位性：事後分布の形状が解析的にわかる
- 解釈上の簡便性：事前情報を追加的な（あるいは過去の）データとして扱える

共役事前分布を使わない（使えない）とき

- 知識と共役事前分布が矛盾するとき
  - 知識（事前情報）に合致する共役分布がない
  - パラメタの母集団（事前分布）が具体的に想定でき、それが共役分布ではない場合
- 知識・事前の情報を優先：共役事前分布は使わない
- 共役でない事前分布でも、情報更新の方法は同じ
- 解析的な答えが出なくても、数値計算 (computation, simulation) で対処できる

二項分布の成功確率の事前分布（1）

例：二項分布の成功確率の推定

$X_i \overset{\text{iid}}{\sim} \mathrm{Bin}(n_i, \theta)$ を、ある学期 i に政治学方法論 を受講する n_i 人のうち、単位を取得する学生の数とする。過去数年の平均合格率は 70% で、合格率の分散は 0.1 であるとする。今年の受講生のデータを使って θ を推定するとき、θ の事前分布をどのように設定する？

事前の知識と条件：

- $0 \leq \theta \leq 1$
- $\mathrm{E}(\theta) = 0.7$
- $\mathrm{var}(\theta) = 0.1$

二項分布の成功確率の事前分布（2）

- 連続一様分布？
- なぜ $\theta \sim \mathrm{U}(0,1)$ では不適切なのか？
- ベータ分布：$\theta \sim \mathrm{Beta}(\alpha,\beta)$
- ベータ分布に従う確率変数 θ のとり得る値：$0 \leq \theta \leq 1$

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

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

- 事前分布の１つの可能性：$\theta \sim \mathrm{Beta}(0.77, 0.33)$

正規分布の位置母数（平均）の事前分布

正規分布の位置母数の推定

確率変数 $X \sim \mathrm{N}(\theta, 1)$ について、事前の知識（先行研究等）により、θ の中位数（中央値）は 0、第 1 四分位数は -1, 第 3 四分位数は 1 であることがわかっている。このとき、θ の事前分布をどのように設定するか？

事前の知識と条件

- $\mathrm{Med}(\theta) = 0$
- $p(\theta)$ を θ の事前分布とすると、

$$\int_{-\infty}^{-1} p(\theta)d\theta = .25, \quad \int_{-\infty}^{1} p(\theta)d\theta = .75$$

- つまり、θ の 50%信頼区間が $[-1, 1]$

可能性 1：正規分布

- $\theta \sim \mathrm{N}(\mu, \sigma^2)$
- $\mathrm{Med}(\theta) = \mu \Rightarrow \mu = 0$
- 標準正規分布の特性により、50%のデータについて、

$$-0.674 \leq \frac{\theta - \mu}{\sigma} \leq 0.674$$

$$-0.674\sigma \leq \theta - \mu \leq 0.674\sigma$$

$\mu = 0$ だから、

$$0.674\sigma = 1 \Rightarrow \sigma \approx 1.484$$

したがって、$\sigma^2 \approx 2.20$

- 事前分布の 1 つの可能性：$\theta \sim \mathrm{N}(0, 2.2)$

可能性 2：コーシー分布

- $\theta \sim \text{Cauchy}(\mu, \phi)$
- $p(\theta) = \dfrac{1}{\pi\phi\left[1 + \left(\frac{\theta-\mu}{\phi}\right)^2\right]}$
- $\text{Med}(\theta) = \mu \Rightarrow \mu = 0$
- コーシー分布の特徴により、50%のデータについて、

$$\mu - \phi \leq \theta \leq \mu + \phi$$

$\mu = 0$ だから、

$$\phi = 1$$

- 事前分布の 1 つの可能性：$\theta \sim \text{Cauchy}(0, 1)$

事前分布の違いによる事後分布の違い

- 事前知識だけでは事前分布を特定できない：可能な分布が複数
- 異なる事前分布を選ぶと、結果が変わる：結果の違いはサンプルサイズが大きくなれば縮小
    - $\theta \sim \mathrm{N}(0, 2.2)$ の場合

$$\mathrm{E}(\theta|y) = y - \frac{y}{3.2}$$

    - $\theta \sim \mathrm{Cauchy}(0, 1)$ の場合、$|y| \geq 4$ について

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

- 頑健性・感度のチェックが必要！

腎臓がんの罹患率

BDA3 (pp.47–51) 腎臓がんの罹患率の例

1980 年から 1989 年までの 10 年間の郡 (county) 単位のデータから、アメリカ合衆国における腎臓がんの罹患率を推定する。使用するデータは、j 郡において 10 年間に観測された腎臓がん患者の数 y_j であり、y_j はポアソン分布に従うと仮定する。また、j 郡の人口は n_j であるとする。

- サンプリングモデル：$y_j \sim \mathrm{Poisson}(10 n_j \theta_j)$
- 事前分布：どう設定する？
  - ポアソン分布の自然な共役分布：ガンマ分布

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

  - α, β の値をどう設定する？

ガンマ分布の母数を特定する

- y_j の事前予測分布：負の二項分布

$$p(y_j) = \int p(y_j|\theta_j)p(\theta_j)d\theta_j = \text{Neg-bin}\left(y_j \middle| \alpha, \frac{\beta}{10n_j}\right)$$

- 負の二項分布の期待値と分散は、

$$\mathrm{E}(y_j) = 10n_j\frac{\alpha}{\beta}, \quad \mathrm{var}(y_j) = 10n_j\frac{\alpha}{\beta} + (10n_j)^2\frac{\alpha}{\beta^2}$$

- したがって、

$$\mathrm{E}\left(\frac{y_j}{10n_j}\right) = \frac{\alpha}{\beta}, \quad \mathrm{var}\left(\frac{y_j}{10n_j}\right) = \frac{1}{10n_j}\frac{\alpha}{\beta} + \frac{\alpha}{\beta^2}$$

- 観測値を代入して連立方程式を解く：$\alpha = 20, \beta = 430000$

事前分布と事後分布

- 事前分布：$\theta_j \sim \mathrm{Gamma}(20, 430000)$

$$\mathrm{E}(\theta_j) = \frac{\alpha}{\beta} = 4.65 \times 10^{-5}, \quad \mathrm{sd}(\theta_j) = \frac{\sqrt{\alpha}}{\beta} = 1.04 \times 10^{-5}$$

- 事後分布：$\theta_j | y_j \sim \mathrm{Gamma}(20 + y_j, 430000 + 10n_j)$

$$\begin{aligned} \mathrm{E}(\theta_j | y_j) &= \frac{20 + y_j}{430000 + 10n_j} \\ &= \frac{1}{430000 + 10n_j} \left(430000 \cdot (4.65 \times 10^{-5}) + 10n_j \frac{y_j}{10n_j} \right) \\ \mathrm{var}(\theta_j | y_j) &= \frac{20 + y_j}{(430000 + 10n_j)^2} \end{aligned}$$

無情報事前分布 (noninformative priors)

- こんなとき、どうする？
  - 推定する母数 θ の母集団を想定するのが困難
  - 事前分布が推定に与える影響をできるだけ小さくしたい
- 参照用事前分布 (reference priors) を使う
  - 曖昧
  - 平坦
  - ばらついている
  - 無情報 (*noninformative*)

参考：A Catalog of Noninformative Priors

無情報事前分布の例

- サンプリングモデル：$y_i \overset{\text{iid}}{\sim} \mathrm{N}(\theta, \sigma^2)$
    - σ^2 は既知
    - サンプルサイズ：n
- 事前分布：$\theta \sim \mathrm{N}(\mu_0, \tau_0^2)$
    - μ_0, τ_0^2 の知識がない or 特定の値に依存したくない
    - $\tau_0^2 = \infty$ にする
    - この事前分布（確率分布ではない）：無情報事前分布の 1 例
- 事後分布：

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

- データが事後分布を支配！

正則 (proper) 事前分布と非正則（変則, improper）事前分布

- 正則事前分布 $P(\theta)$：データに依存せず、

$$\int_{-\infty}^{\infty} p(\theta)d\theta = 1$$

- 非正則事前分布 $p(\theta)$：

$$\int_{-\infty}^{\infty} p(\theta)d\theta = \infty$$

- 正規化されいない事前分布 $p(\theta)$：

$$\int_{-\infty}^{\infty} p(\theta)d\theta = c$$

ただし、c は正の定数

  - 正規化されていない分布は、正規化すれば（密度関数を c で割れば）正則分布になる

非正則な事前分布から正則な事後分布へ

- 非正則な事前分布を使っても、正則な事後分布が得られる（常に得られるわけではない）
- $p(\theta|y) \propto p(y|\theta)p(\theta)$ で得られた事後分布そのものは正則ではない
  - y の値によらず（すべての y について)、

$$\int p(\theta|y)d\theta = c$$

となれば、c で割って正則な分布が得られる

弱情報事前分布 (weakly informative priors)

弱情報事前分布とは：

- 正則な確率分布
- 保有している知識をあえて弱めて表現したもの

弱情報事前分布を使う理由：

- モデルを単純化したい
- 保有している情報を確率分布に正確に反映するのが困難
- 計算を楽にしたい
- 事前情報の信頼性が低い

弱情報事前分布の作り方

- 無情報事前分布を基に、知識を用いて分布に制約をかける
  - 取り得ない値を除外する
  - 特定の範囲に対する信頼度を高める
  - etc.
- 情報のある事前分布を基に、不確実性を高める
  - 正則な事前分布の分散を大きくする
  - 自分が妥当性を示そうとする仮説に不利な方向に分布を拡大する
  - etc.