

政治学方法論 II – 課題 10

提出期限：2015 年 6 月 24 日（水）午前 9 時（日本時間）

提出方法：担当教員に**メールの添付ファイル**として送る

提出するメールの件名：政治学方法論 2 課題 10

注意：提出するファイルは 1 つ

- PDF ファイル：ファイル名は `rm2-hw10-NAME.pdf`

問 1. 打率を推定する (Efron and Morris 1975; Jackman 2009)

Efron and Morris (1975) は、18 人のメジャーリーガーについて、1970 年シーズンの打率について知るために、シーズン最初の 45 打席のデータを分析した。

各選手 j ($j = 1, \dots, J$; $J = 18$) について、その 45 打席の打率を y_j とする。また、観測された打率の平均の分散を、 $\sigma^2 = \frac{\bar{y}(1-\bar{y})}{n}$, $n = 45$ とする。

このとき、以下のようなモデルを考える。

$$y_j | \theta_j \sim N(\theta_j, \sigma^2)$$

$$\theta_j | \mu_0, \sigma_0^2 \sim N(\mu_0, \sigma_0^2)$$

$$\mu_0 \sim N(b_0, B_0)$$

$$\sigma_0^2 \sim \text{Inv-Gamma}(7, 0.0175)$$

ただし、 $p(\mu_0, \sigma_0^2) = p(\mu_0)p(\sigma_0^2)$ とする。

このモデルを使い、以下の各問に答えなさい。データは、R の `pscl` パッケージに、`EfronMorris` という名前で含まれている（パッケージを読み込んだ後、`data(EfronMorris)` でアクセスできる）。

1. 現実のデータ（現代の日本のプロ野球のデータでよい）に照らし、打率の 95% 信頼区間を推測で（EfronMorris データは使わずに）決めなさい（言い換えると、打率の 95% 事前区間を、関連する情報を使って求めなさい）。

2. 上で設定した 95% 事前区間の中央値を b_0 、信頼区間から計算できる分散を B_0 とする。 b_0 と B_0 を求めなさい。
3. θ の事後分布をシミュレーションによって求めなさい。
4. 各打者の打率を別々に分析した場合 (no pooling) の θ の推定値 (あるいは区間) を求めなさい。
5. すべての打者を一緒くたに扱った場合 (complete pooling) の θ の推定値 (あるいは区間) を求めなさい。
6. 3, 4, 5 の結果を比較しなさい。

問 2. 打率を推定する (2)

問 1 と同じデータを、以下の階層二項分布モデルで推定しなさい。

$$r_j | \theta_j \sim \text{Bin}(n, \theta_j)$$

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

$$\alpha \sim \text{Expon}(2)$$

$$\beta \sim \text{Expon}(2)$$

ただし、 r_j はヒットの数である。また、 $p(\alpha, \beta) = p(\alpha)p(\beta)$ であるとする。

問 3. 課題 9 の再提出

課題 9 のうち、未回答または不正解だった問題を解き、提出しなさい。

参考文献

Efron, Bradley, and Carl Morris. 1975. "Data Analysis Using Stein's Estimator and Its Generalizations." *Journal of American Statistical Association* 70: 311–319.

Jackman, Simon. 2009. *Bayesian Analysis for the Social Sciences*. Chichester: Wiley.