データ解析のための統計モデリング入門

10.階層ベイズモデル
GLMMのベイズモデル化

山木翔馬

これまでの流れ

- GLMM　（7章）
  - 「個体差」などを組み込んだ現実的な統計モデル
- GLMのベイズモデル化　（9章）
  - 無情報事前分布を使ったベイズ統計モデル

- MCMC　（第8章）
  - パラメータの最尤推定が困難な場合

階層ベイズモデルとMCMC

- 個体差などをくみこんだ現実的な統計モデルを構築するには、無情報事前分布だけではなく、階層事前分布が必要
- 個体差のほかにも「調査場所の差」などもモデルに含める

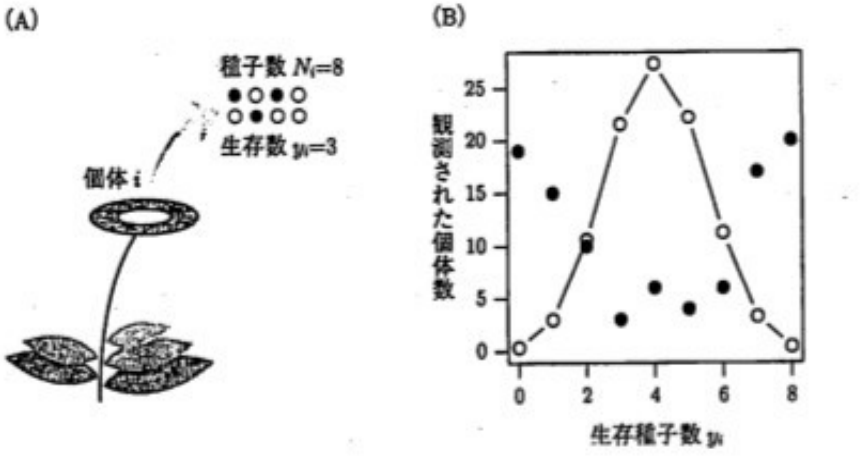
  →モデルが複雑になり、パラメータ推定も難しい

階層ベイズモデルとMCMCサンプリングによるパラメータ推定が威力を発揮する

例題

- 図（B）の●が観測データ
- 図（B）の○は生起確率0.504の二項分布

- 二項分布では観測データのばらつきをうまく説明できない



GLMMの階層ベイズモデル化

- リンク関数と線形予測子

$$logit(q_i) = \beta + r_i$$

- 尤度関数

$$p(Y|\beta, \{r_i\}) = \prod_i \left(\begin{array}{c} 8 \\ y_i \end{array} \right) q_i^{yi} (1 - q_i)^{8-yi}$$

- βは全個体に共通するパラメータ
- r_iは個体差のパラメータ

この2つのパラメータの事前分布を指定する

パラメータの事前分布

$$p(\beta) = \frac{1}{\sqrt{2\pi \times 100^2}} exp\left(\frac{-\beta^2}{2 \times 100^2}\right)$$

- 9章と同じように無情報事前分布を指定する
  （平均0、標準偏差100の正規分布）

$$p(r_i|s) = \frac{1}{\sqrt{2\pi s^2}} exp\left(\frac{-r^2}{2s^2}\right)$$

- 平均0、標準偏差sの正規分布を指定する
  - p(s) = 0から10^4までの連続一様分布
  - 事前分布の事前分布を設定　→　階層ベイズモデル

階層ベイズモデルの事後分布

$$p(\beta, s, \{r_i\}|Y) \propto p(Y|\beta, \{r_i\})p(\beta)p(s)\prod_i p(r_i|s)$$

尤度関数

$$p(Y|\beta, \{r_i\}) = \prod_i \left(\begin{array}{c} 8 \\ y_i \end{array} \right) q_i^{yi}(1-q_i)^{8-yi}$$

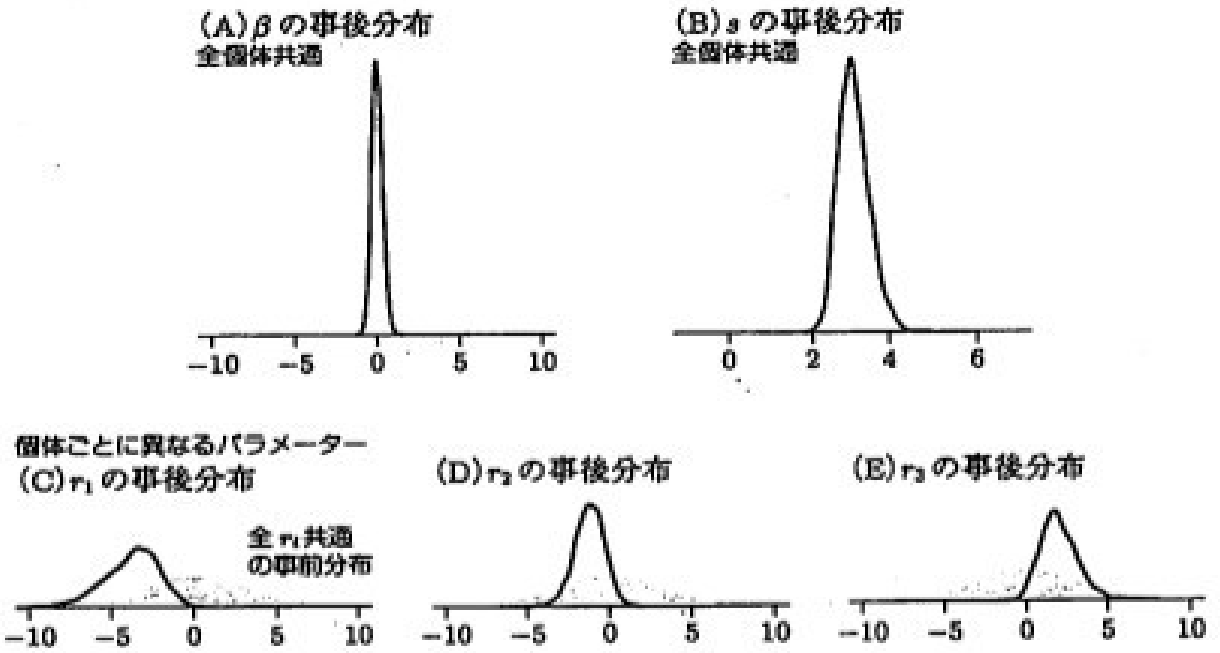
パラメータの事前分布

$$p(\beta) = \frac{1}{\sqrt{2\pi \times 100^2}} exp\left(\frac{-\beta^2}{2 \times 100^2}\right)$$

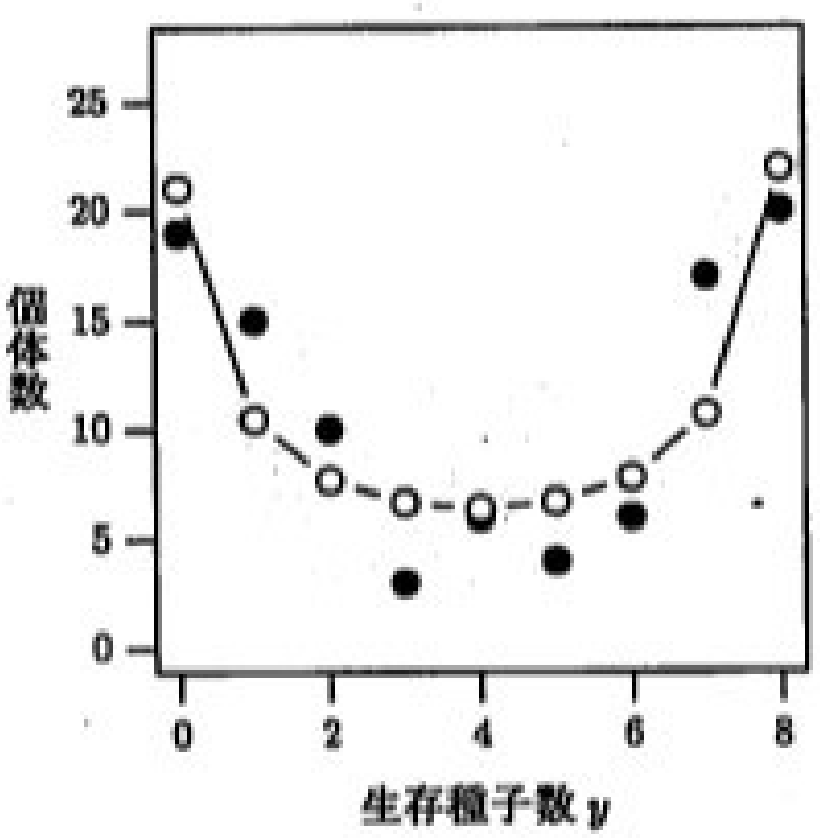
$$p(r_i|s) = \frac{1}{\sqrt{2\pi s^2}} exp\left(\frac{-r^2}{2s^2}\right)$$

パラメータの事後分布

MCMCサンプリングで得られた
パラメータごとの事後分布



yのMCMCサンプル

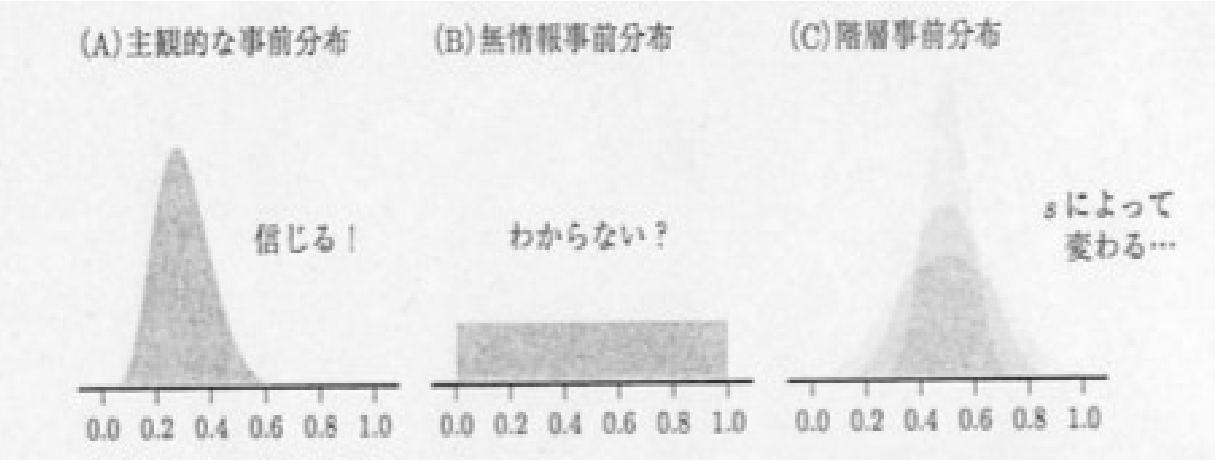


- ●が観測データ
- ○が生存種子数yの確率分布

$$p(y|\beta, s) = \int_{-\infty}^{\infty} p(y|\beta, r)p(r|s)dr$$

ベイズモデルで使う様々な事前分布

- ベイズ統計モデルの設計において事前分布の選択は重要



- よく使われる3つの事前分布

どの事前分布を使うか

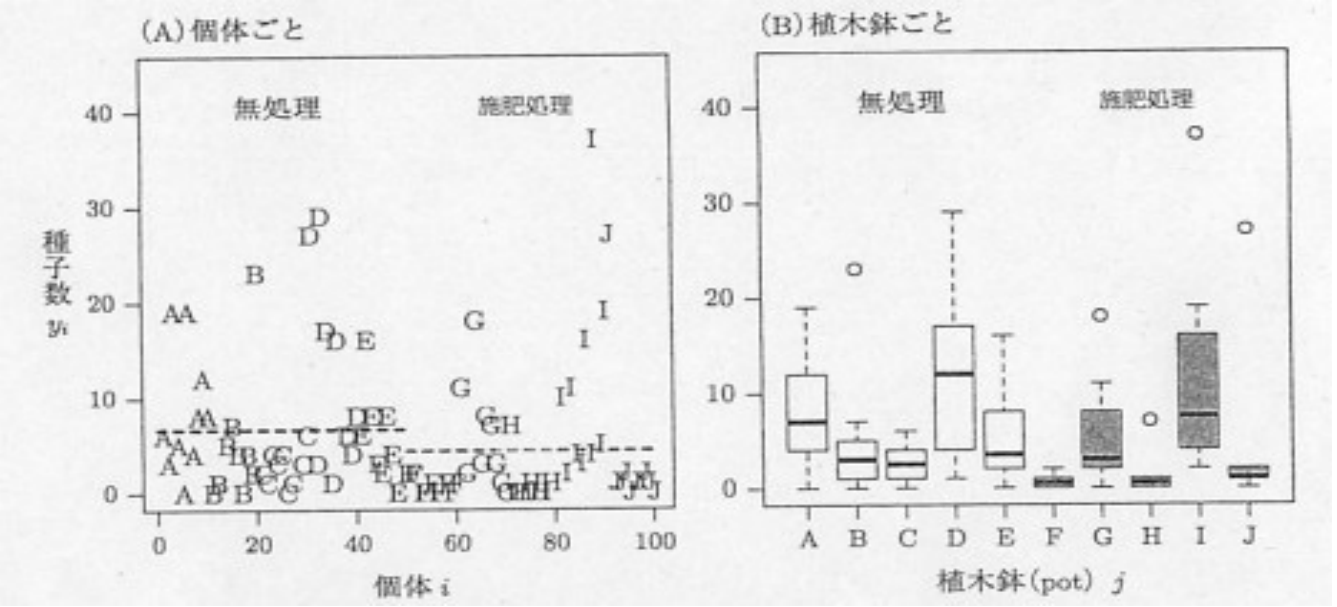
- 全体に共通する平均やばらつき
  - 例題の切片βなど
  - 無情報事前分布

- 個体ごとのずれ
  - 例題の個体差 r_i
  - 階層事前分布

個体差＋場所差の階層ベイズモデル

- 個体差や植木鉢差がある
- 例題のデータ構造は擬似反復なので個体差と植木鉢差を同時に扱う必要がある

  →GLMM化したポアソン回帰で扱える



個体iの種子数yiのばらつきを平均λiのポアソン分布で表現

$$p(y_i|\lambda_i) = \frac{\lambda_i^{yi} exp(-\lambda_i)}{y_i!}$$

平均種子数

$$log\lambda_i = \beta_1 + \beta_2 f_i + r_i + r_{j(i)}$$

β1：切片（無情報事前分布）
β2：施肥処理の有無の係数
　　（無情報事前分布）
ri：個体iの効果
　　（階層事前分布；平均0、標準偏差s）
rj(i)：植木鉢jの効果
　　（階層事前分布；平均0、標準偏差sp）

まとめ

- 階層ベイズモデルは事前分布となる確率分布のパラメータにも事前分布が指定されている
- 事前分布の選択ではパラメータが説明する範囲によって無情報事前分布・階層事前分布を選択する
- 複雑なモデリングでは階層ベイズモデルとMCMCサンプルによるパラメータ推定で対応