GLMのベイズモデル化と事後分布の推定（導入）

太田研４年　和田

統計モデル(線形予測子込み)のベイズ化
＋事後分布からのMCMCサンプリング

7章：GLMM

$$logit(q_i) = \beta_1 + \beta_2 x_i \boxed{+ r_i}$$

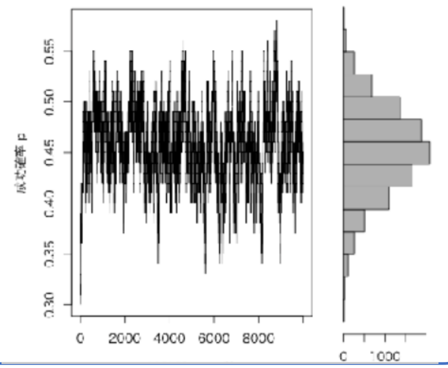
・観測され得ないデータを変数に

・確率分布を入れ子にできる

8章前半：MCMC

・最尤推定の方法の一種

・試行回数を重ねると最良のパラを中心に分布



8章後半：ベイズ統計モデル

・世の中の現象は確率分布で表せる

・事後分布 ∝ 尤度×事前分布(定数)

<u>一つに統合！！</u>

・線形予測子　×　事後分布公式

・事後分布　×　MCMC

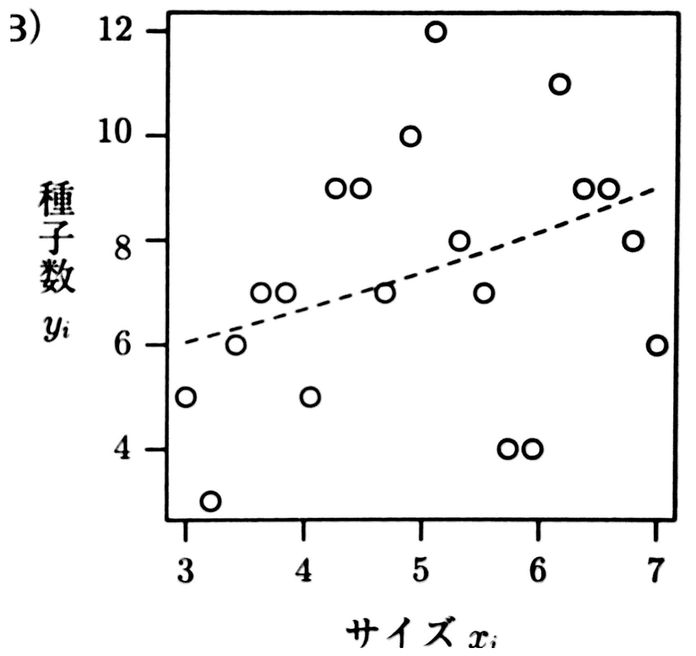
9.1 今回扱う例題（注：too simple）

- 植物が20個体いる(個体i, i=1~20)
- 各個体の種子数y_i（データあり）
- ↑体の大きさx_iに依存（x_iが種子数平均を決定）

《仮定》

- 個体差は存在しない
  - ランダム効果はなし
- 種子数yiのばらつきはλiのポアソン分布
- $\lambda_i = \exp(\beta_1 + \beta_2 x_i)$

→普通のポアソン回帰GLMで算出可能



9.2. GLMのベイズモデル化

種子数yiのばらつきはλiのポアソン分布
$$\lambda_i = \exp(\beta_1 + \beta_2 x_i)$$

- 尤度関数　$\mathrm{L}(\beta_1, \beta_2) = \prod p(yi|\lambda_i)$　$= \prod p(yi|\beta_1, \beta_2, x_i)$
  - ポアソン分布なので；離散確率分布！
  - $\mathrm{p}(\beta_1, \beta_2 \,|\, Y) = \mathrm{L}(\beta_1, \beta_2)$
- 事後分布 ∝ 尤度 × 事前分布　より
- $\mathrm{p}(\beta_1, \beta_2 \,|\, Y) \propto \mathrm{p}(Y \,|\beta_1, \beta_2)\boxed{\mathrm{p}(\beta_1)\mathrm{p}(\beta_2)}$

それぞれβ_1, β_2の事前分布

これを適切に指定すると…

GLMのベイズモデル化！

9.3 無情報事前分布
ー事前分布をどう設定するのかー

- 事前分布は「観測データが得られてない時の、パラの確率分布」

？！？！そんなのわからないでしょ！！！

→ [-∞,∞]の範囲で好きな値をとっていい　…「<u>**無情報事前分布**</u>」

- 無限区間の一様分布？　→（確率分布なのに）積分が1にならない

妥協案１
$-10^9 < \beta < 10^9$の範囲の一様分布

妥協案2
異常に平べったい正規分布
(e.g. 平均0, $\sigma = 100$)

疑問
なぜ好きな値なのに一様？
なぜ正規分布？