ベイズ統計

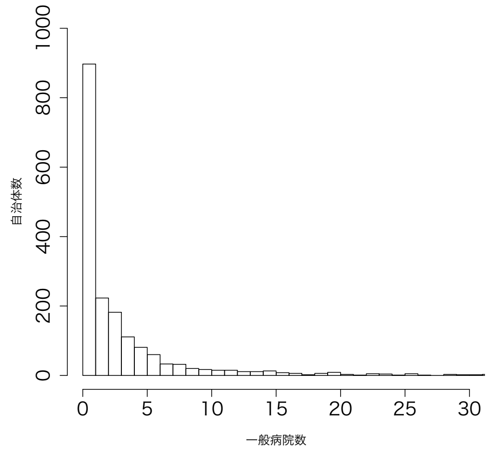
古谷知之

講義概要

- マルチレベル・モデルの演習(前回続き)
- ポアソン回帰モデル
- 負の二項分布モデル
- R演習

ポアソン分布

- 非常に多くの観測回数が繰り返されるものの、観測ケースの発生頻度が非常に低い場合に用いられる確率分布
- 一定範囲内（時間、回数、空間）である事象が発生する平均値λをもちいて計算される
- 試行回数n、発生頻度pとすると、$\lambda = np$
- ポアソン分布は次式で表される



$$Po(\lambda) = \frac{\lambda^x \cdot exp(-\lambda)}{x!}$$

ポアソン回帰モデルのベイズ推定

- ポアソン分布を説明する回帰モデルは、例えば次式のように表される。ここでy_iは被説明変数、x_1は説明変数、(β_0, β_1)は未知パラメータである。

$$y_i = Po(\lambda_i)$$
$$log(\lambda_i) = \beta_0 + \beta_1 x_{1i}$$

- 回帰係数が正規分布に従うので、例えば以下のように事前情報を設定してベイズ推定する。

$$\beta_0, \beta_1 \sim N(0, 1\times10^{-6})$$

ベルヌーイ試行と二項分布

- 0か1かしかない試行において、n回の試行でr回成功し、その確率πがわかっているとき、実験が成功する期待値は以下のベルヌーイ試行に従う
- ベルヌーイ試行の確率分布を二項分布といい、その分布は次式の確率密度関数に従う

$$Binom(n, p) = {}_nC_r \cdot \pi^r \cdot (1-\pi)^{n-r} \approx \pi^r \cdot (1-\pi)^{n-r}$$

負の二項分布

- 二項分布：n回の試行でr回成功（ベルヌーイ試行）

$$P(\ y) = {}_nC_r \cdot \pi^r \cdot (1-\pi)^{n-r}$$

- 負の二項分布：ある事象についてr回までその事象が生じなかった確率分布

$$P(\ y) = {}_{n-1}C_{r-1} \cdot \pi^r \cdot (1-\pi)^{n-r}$$

- 過大分散が生じるような現象に対して用いられる

負の二項分布

- ここで$n = y + r$と置き換えると、負の二項分布はガンマ関数を用いて以下のように表せる

$$P(y) = {}_{y+r-1}C_{r-1} \cdot \pi^r \cdot (1-\pi)^y$$
$$= \frac{\Gamma(y+r)}{y!\,\Gamma(r)} \cdot \pi^r \cdot (1-\pi)^y$$

- このとき

平均：$\frac{r(1-\pi)}{\pi}$

分散：$\frac{r(1-\pi)}{\pi^2}$

負の二項分布

- さらに$\pi = r/(\lambda + r)$と置き換えると、負の二項分布以下のように表せる

$$P(y) = \frac{\Gamma(y+r)}{y!\,\Gamma(r)} \left(\frac{r}{r+\lambda}\right)^r \left(\frac{\lambda}{r+\lambda}\right)^y$$

- このとき

平均：λ

分散：$(\lambda + r\lambda)/r$

- $(1+r)/r$を分散指標という

負の二項分布モデルのベイズ推定

- 負の二項分布モデル

$$y_i \sim NB(r, \pi_i)$$
$$\pi_i = \frac{r}{\lambda_i + r}$$

$$log(\lambda_i) = \beta_0 + \beta_1 x_{1i}$$

- 事前情報は例えば以下のようにする

$$\beta_0, \beta_1 \sim N(0, 1\times10^{-2})$$
$$r \sim Ga(1\times10^{-3}, 1\times10^{-3})$$