

一般化線形モデルとベイズ推定

樋口さぶろお <https://hig3.net>

龍谷大学大学院 先端理工学研究科 数理・情報科学コース

計算科学特論 L08(2026-06-08 Mon)

最終更新: Time-stamp: "2026-06-08 Mon 17:51 JST hig"

今日の目標

- 一般化線形モデルを説明できる
- 一般化線形モデルのベイズ化を説明できる



L07-Q1

Quiz 解答: 多項分布の確率

- ① 四面体サイコロを 5 回振ったときの目の出方は、各面の出方の確率に基づく多項分布に従います。具体的には、各目 $i = 1, 2, 3, 4$ の出る回数 x_i は次のように表されます。

$$(x_1, x_2, x_3, x_4) \sim M(5, 0.3, 0.2, 0.4, 0.1)$$

②

$$P(x_1 = 3, x_2 = 0, x_3 = 1, x_4 = 1) = \frac{5!}{3!0!1!1!} \cdot 0.3^3 \cdot 0.2^0 \cdot 0.4^1 \cdot 0.1^1 = 0.02$$

L07-Q2

Quiz(ディリクレ分布の確率)

p_1, p_2, p_3 は, パラメタ $(1, 1, 1)$ のディリクレ分布にしたがう.

- ① 確率密度関数 $f(p_1, p_2, p_3)$ が 0 にならない $(p_1, p_2, p_3) \in \mathbb{R}^3$ は, $\mathbb{R}^3$ のどこにあるか.
- ② $P(p_1 \geq \frac{1}{2})$ を求めよう.

緑の教科書にしたがって，一般化線形モデル (GLM=Generalized Linear Models) を説明する．

- 久保拓弥，データ解析のための統計モデリング入門，岩波書店，2012

GLM は，線形モデル=正規分布に基づくモデル（線形回帰とか）を拡張し，ポアソン回帰，ロジスティック回帰などを一括して扱う枠組み．一般化線形モデルの（階層）ベイズ化（=母数の事前事後分布を導入）をやっておくと，適用範囲が広い

ここまで来たよ

7 潜在変数モデルのベイズ推定・ディリクレ分布

8 一般化線形モデルとベイズ推定

- ポアソン分布
- ポアソン回帰
- 一般化線形混合モデル

種子の個数の分布

(ある地区に生えている) ある植物の, 1 本のつける種子の個数の分布は単峰 (unimodal) だという.
どんな分布でモデル化?

- 正規分布? (線形モデル)



- 二項分布?



ポアソン分布

単位時間（連続）あたりの平均発生回数が λ であるような事象の発生回数 Y のとり離散分布．このとき，間隔（時間）の分布は指数分布にしたがう．

定義 (ポアソン分布)

次の確率関数を持つ離散型確率変数 Y を，パラメタ λ のポアソン分布 $\text{Po}(\lambda)$ にしたがうという．

$$p(y; \lambda) = \frac{e^{-\lambda}}{y!} \lambda^y \quad (y \geq 0)$$

L08-Q1

ポアソン分布の母期待値・母分散

$$E[Y^0]=1, E[Y^1]=, V[Y]=,$$

例：単位時間あたりのサッカーの得点，単位質量単位時間あたりの放射性元素の崩壊

L08-Q2

Quiz(ポアソン分布)

ある県では, 交通死亡事故が, 平均すると 1 日に 3 件起きるという. 1 日の事故の件数はパラメタ $\lambda = 3$ のポアソン分布に従う.

- ① 特定の 1 日に, 交通死亡事故が 0 件である確率を求めよう.
- ② 特定の 1 日に, 交通死亡事故が 6 件である確率を求めよう.
- ③ 1 日に起きる交通死亡事故の件数の分散を求めよう.

L08-Q3

Quiz(ポアソン分布)

あるサッカーチームは、90 分のゲームで平均 3 点得点できる。

- ① ハーフ 45 分間に 0 点である確率は?
- ② ハーフ (前半) 0 点 かつ ハーフ (後半) 3 点である確率は?
- ③ ゲーム 90 分で 3 点であるときに、ハーフ (前半) 0 点, ハーフ (後半) 3 点である確率は?

ここまで来たよ

7 潜在変数モデルのベイズ推定・ディリクレ分布

8 一般化線形モデルとベイズ推定

- ポアソン分布
- **ポアソン回帰**
- 一般化線形混合モデル

植物の全高と種子の個数

(ある地区に生えている) ある植物の, 1 本のつける種子の個数 Y は, その株の全高 x が高い方が多いらしい. どんな統計モデル?

- 線形回帰 $y = \beta_0 + \beta_1 x + \epsilon$, $\epsilon \sim N(0, \sigma^2)$?



ポアソン回帰

$\lambda = \beta_0 + \beta_1 x.$ ノイズ項 ϵ なし

$$p(y|\lambda) = \frac{e^{-\lambda}}{y!} \lambda^y \quad \text{ここにランダム効果}$$



正の量である λ を $\exp(\text{ナントカ})$ で表す

ポアソン回帰

$$\lambda = e^{\beta_0 + \beta_1 x}$$

$$\text{or } \log(\lambda) = \beta_0 + \beta_1 x$$

$$Y \sim \text{Po}(\lambda)$$

$\leadsto$ ベイズ化 $p(y|\beta_0, \beta_1)$, 事前分布 $f(\beta_0, \beta_1)$.

一般化線形モデル

一般化線形モデル (GLM=Generalized Linear Model)

$$\log(\lambda) = \beta_0 + \beta_1 x, \quad (x : \text{説明変数})$$

リンク関数(母数) = 線形予測子

目的変数 $Y \sim \text{Po}(\lambda)$

誤差構造

リンク関数，線形予測子，誤差構造の3個組で定まる
線形回帰

- リンク関数
- 線形予測子
- 誤差構造

さまざまな線形予測子

$\lambda = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots$ 説明変数について線形という意味ではない

$\lambda = \beta_0 + \beta_1 x_1 + \beta_{11} x_1 x_2 + \dots$ 複数の説明変数

$\lambda = \beta_0 + \beta_1 x + \beta_2 d$

...

ダミー変数 $d = \begin{cases} 0 & (\text{品種 A}) \\ 1 & (\text{品種 B}) \end{cases}$.

ロジスティック回帰

植物の全高と種子の個数

(ある地区に生えている) ある植物は, (固定) n 個の花をつける. そのうち $0 \leq y \leq n$ 個に実がつく. 1 本のつける実の個数 Y は, その株の全高 x が高い方が多いらしい. どんな統計モデル?

誤差構造

0 以上 n 以下, 離散, 花 1 個ずつの勝負だろうから, 母数 p で,

$$Y \sim B(n, p)$$

線形予測子

$$p = \beta_0 + \beta_1 x?$$

リンク関数

$$p = \frac{1}{1 + e^{-z}}, z = \beta_0 + \beta_1 x$$

$0 \leq p \leq 1$ という要素を入れて, ロジスティック関数で.

逆に解いて, リンク関数は, ロジット関数 $z = \text{logit}(p) = \log \frac{p}{1-p}$.

L08-Q4

ロジット関数

$z = \text{logit}(p) = \frac{p}{1-p}$ を示そう．グラフを描こう．

→ ベイズ化 $p(y|\beta_0, \beta_1)$, 事前分布 $f(\beta_0, \beta_1)$.

L08-Q5

小数の法則

二項分布 $B(n, p)$ は、極限 $n \rightarrow +\infty$, $np = \lambda$ (一定) でポアソン分布 $Po(\lambda)$ になることを示そう.

ここまで来たよ

7 潜在変数モデルのベイズ推定・ディリクレ分布

8 一般化線形モデルとベイズ推定

- ポアソン分布
- ポアソン回帰
- 一般化線形混合モデル

一般化線形混合モデル (Generalized Linear Mixed Model=GLMM)

(とりあえずは Gaussian Mixed Model=GMM の mixed とは違う意味)
植物の話で、株ごとに (たまたま虫に食われたとか、栄養分に行き当たったとかで) λ の値はランダムに変わると考える.

$$\log \lambda_i = \beta_0 + \beta_1 x_i + r_i, \quad r \sim N(0, \sigma^2)$$

- $\beta_0 + \beta_1 x_i$ 固定効果
- $r_i \sim N(0, \sigma^2)$ ランダム効果 (の混合)
- ランダム性は、ポアソン誤差構造が分担してたんじゃないの？
 - ▶ 現象に即したランダム性の入れ方
 - ▶ ポアソンでは $E[Y] = V[Y]$ だが、分散だけを増やせる
 - ▶ 母数が β_0, β_1, σ が増える

ベイズ化 $\rightsquigarrow$ 階層ベイズ

$$\begin{aligned} &\swarrow \beta_0 \\ \lambda_i &\leftarrow \beta_1 \\ &\swarrow r_i \leftarrow \sigma^2 \end{aligned}$$

事前分布 $f(\beta_0, \beta_1)$.

ハイパー事前分布 $f(\sigma)$.

マルチレベルモデル

地区 j ごとに育ち具合が違ふことを考慮.

$$\log \lambda_{ij} = \beta_0 + \beta_1 x_i + s_j, \quad s_j \sim N(0, \sigma^2)$$

事前分布 $f(\beta_0, \beta_1)$.

ハイパー事前分布 $f(\sigma)$.