

マルチレベルモデルのベイズ推定・混合ガウス分布

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 L06(2026-05-25 Mon)

最終更新: Time-stamp: "2026-05-25 Mon 12:13 JST hig"

今日の目標

- マルチレベルモデル，階層ベイズモデルを説明できる
- 混合ガウス分布を説明できる



L05-Q1

Quiz 解答: 線形回帰モデルの最小二乗推定

$\frac{\partial S}{\partial \beta_0} = \frac{\partial S}{\partial \beta_1} = 0$ を解いて得られる.

L05-Q2

Quiz 解答: 線形回帰モデルのベイズ推定

事後分布は,

$$\begin{aligned} & f_{\dots|\dots}(\beta_0, \beta_1 | y_1, \dots, y_N, x_1, \dots, x_N) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_1 - (\beta_0 + \beta_1 x_1))^2}{2\sigma^2}} \dots \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_N - (\beta_0 + \beta_1 x_N))^2}{2\sigma^2}} \\ & \quad \times \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(\beta_0 - \mu_0)^2}{2\sigma_0^2}} \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(\beta_1 - \mu_1)^2}{2\sigma_1^2}}. \end{aligned}$$

MAP 推定は, $\frac{\partial f}{\partial \beta_0} = \frac{\partial f}{\partial \beta_1} = 0$ を考えることに相当する. 結果は

$$\sum_i (y_i - (\beta_0 + \beta_1 x_i)) + \frac{\sigma^2}{\sigma_0^2} (\beta_0 - \mu_0) = 0$$

$$\sum_i x_i (y_i - (\beta_0 + \beta_1 x_i)) + \frac{\sigma^2}{\sigma_1^2} (\beta_1 - \mu_1) = 0$$

$N\sigma^2$ が σ_0^2, σ_1^2 に比べて小さいときは第 1 項のみが残り, 最小二乗法と同じ結果になる. 一方, $N\sigma^2$ が σ_0^2, σ_1^2 に比べて大きいときは各第 2 項が効き, 事前分布の母平均値に近い値になる.

ここまで来たよ

5 回帰分布におけるベイズ推定

6 マルチレベルモデルのベイズ推定・混合ガウス分布

- マルチレベルモデル
- 混合ガウス分布

マルチレベルモデルの例 1

多数のドーナツ製造機 i は、それぞれの平均値 $\mu^{(i)}$ 、共通の (既知の) 分散 σ_x^2 の正規分布にしたがう重さのドーナツを製造する. $\mu^{(i)}$ を、そのドーナツ製造機の「固有製造重さ」という.

1 台の「ドーナツ製造機」製造機があり、ある平均値 μ_0 、既知の分散 σ_y^2 の正規分布にしたがう「固有製造重さ」のドーナツ製造機を製造する. μ_0 を設計重さという.

1 台の「ドーナツ製造機」製造機により、5 台のドーナツ製造機 $i = 1, \dots, 5$ が製造され、各製造機の製造したドーナツの重さ $x^{(i,j)}$ は、次のようだった. 1 台の μ_0 、5 台の $\mu^{(i)}$ ($i = 1, \dots, 5$) を求めたい.

製造機番号 i ドーナツの重さ $x_j^{(i)}$ ($j = 1, 2, \dots$))

1 101, 103, 102, 99, 98

2 98, 100

3 103, 105, 104, 104

4 110, 115, 109, 108

5 102, 101, 100

マルチレベルモデルの例 2

ある市の中学校 1 校 1 校の, 1 年生の数学の点数は, それぞれの平均値 $\mu^{(i)}$, 共通の (既知の) 分散 σ_x^2 の正規分布にしたがう. $\mu^{(i)}$ を学校平均点という.

その市の学校平均点は, ある平均値 μ_0 , 既知の分散 σ_y^2 の正規分布にしたがう. μ_0 を市平均点という.

5 校の生徒の点数 $x_j^{(i)}$ は, 次のようだった. 市平均点 μ_0 , 5 校の学校平均点 $\mu^{(i)}$ ($i = 1, \dots, 5$) を求めたい.

学校番号 i 点数 ($j = 1, 2, \dots$)

1 101, 103, 102, 99, 98

2 98, 100

3 103, 105, 104, 104

4 110, 115, 109, 108

5 102, 101, 100

別の表記：分散分析 ANOVA でよく見るもの

$$\mu^{(i)} = \mu + \epsilon'$$

$$x^{(i)} = \mu^{(i)} + \epsilon^{(i)}$$

値 = 全体の平均 + 群間 + 群内

$$\begin{aligned} x_j^{(i)} &= \mu_0 + (\mu^{(i)} - \mu_0) + (x_j^{(i)} - \mu^{(i)}) \\ &= \mu_0 + \epsilon' + \epsilon^{(i)} \end{aligned}$$

分散分析の問：(いずれかの) ドーナツ製造機の固有製造重さの間に差はあるか？ (いずれかの) 学校間に差はあるか？

対比 線形回帰モデル

$$y^{(x)} = (\beta_0 + \beta_1 x) + \epsilon$$

分散分析では、離散 i に依存, 目的変数は x (回帰と記号が違う)
回帰分析では、線形連続 x に依存, 目的変数は y

マルチレベルモデルって必要なの？

$\mu^{(i)}$ は、5 群に対して 5 回推定を繰り返せばいいだけでは？

μ_0 は全ドーナツ・全学生をまとめて推定すればいいだけでは？

→ 生産量が少ない i があるとき，次のように，マルチレベルモデルでまとめて考えると精度が高まる（**縮小** shrinkage）

マルチレベルモデル

$x_j^{(i)}$ 各製造機 i の製造するドーナツの重さ

分布 $f(x_j^{(i)}|\mu^{(i)})$ 例えば, $x_j^{(i)} \sim N(\mu^{(i)}, \sigma_x^2)$.

「ドーナツ製造機」の作る製造機の「固有製造重さ」 $\mu^{(i)}$

分布 $f(\mu^{(i)}|\mu_0)$ 例えば, $\mu^{(i)} \sim N(\mu_0, \sigma_y^2)$.

2 段階, 2 範囲ある, すなわち

まず製造機が製造機を製造 (試行) し, 次に製造機がドーナツを製造 (試行) する, という段階がある

市レベルと学校レベルがある,

このことからマルチレベルモデル, 階層モデル, という.

階層ベイズモデリング

「階層」は変数の構造に着目した言葉.

事前分布 μ_0 の事前分布 $f(\mu_0)$

尤度

$$\prod_i \left(f(\mu^{(i)} | \mu_0) \prod_j f(x_j^{(i)} | \mu^{(i)}) \right)$$

事後分布 これらの積が $\mu_0, \mu^{(i)}$ の事後分布

- 単レベルのモデルだったら、パラメタ $\mu^{(i)}$ が事前分布（ハイパーパラメタ μ_0 で定まる）を与えられる.
- このマルチレベルモデルでは、ハイパーパラメタ μ_0 は固定でなく、その事前分布=超事前分布（ハイパープライア）を考える.
- PyMC で何回も $\mu^{(i)}, \mu_0$ を生成することで、 $\mu^{(i)}, \mu_0$ の（独立でない）多次元事後分布が得られる.

L06-Q1

Quiz(マルチレベルモデル)

ある学年を対象に行われたテストで、各クラス i の平均点 $\mu^{(i)}$ は $N(\mu_0, 3^2)$ にしたがう、クラス i の各学生 j の点数 $x_j^{(i)}$ は $N(\mu^{(i)}, 2^2)$ にしたがうとする。

- クラス $i = 1$ の 2 名の学生の点数は、55, 59
- クラス $i = 2$ の 2 名の学生の点数は、61, 65

だった。学年平均点 μ_0 の事前分布を $N(50, 10^2)$ とする。
 $\mu_0, \mu^{(i)}$ の事後分布を求めよう。

ここまで来たよ

5 回帰分布におけるベイズ推定

6 マルチレベルモデルのベイズ推定・混合ガウス分布

- マルチレベルモデル
- 混合ガウス分布

混合ガウス分布 GMM=Gaussian mixture model

正規分布 normal distribution = ガウス分布 Gaussian distribution

岩薩林 §4.5

混合する = mix

混合ガウス分布を定義する準備として, 次の離散-連続混合型分布 $f(x, y)$ を考える.

$\pi_0 + \pi_1 = 1, \pi_y \geq 0, \pi_{\text{他}} = 0$ とする.

$$f(x, y) = \pi_y \frac{1}{(2\pi\sigma_y^2)^{1/2}} e^{-\frac{(x-\mu_y)^2}{2\sigma_y^2}} = \begin{cases} \pi_0 \cdot \frac{1}{(2\pi\sigma_0^2)^{1/2}} e^{-\frac{(x-\mu_0)^2}{2\sigma_0^2}} & (y = 0) \\ \pi_1 \cdot \frac{1}{(2\pi\sigma_1^2)^{1/2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} & (y = 1) \\ 0 & (y \neq 0, 1) \end{cases}$$
$$= \delta_{y0} \pi_0 \frac{1}{(2\pi\sigma_0^2)^{1/2}} e^{-\frac{(x-\mu_0)^2}{2\sigma_0^2}} + \delta_{y1} \pi_1 \frac{1}{(2\pi\sigma_1^2)^{1/2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}.$$

Y の周辺分布は離散型でベルヌイ分布 $B(1, \pi_1)$

$$f_Y(y) = \int f(x, y) dx = \int_{-\infty}^{+\infty} \pi_y \cdot \frac{1}{(2\pi\sigma_y^2)^{1/2}} e^{-\frac{(x-\mu_y)^2}{2\sigma_y^2}} dx = \pi_y.$$

X の周辺分布は連続型で GMM

定義 (GMM=混合ガウス分布)

連続型確率変数 X がパラメタ $(\pi_0, \pi_1, \mu_0, \mu_1, \sigma_0^2, \sigma_1^2)$ の GMM に従うとは、次の確率密度関数にしたがうことをいう。

$$f_X(x) = \sum_y f(x, y) = \pi_0 \cdot \frac{1}{(2\pi\sigma_0^2)^{1/2}} e^{-\frac{(x-\mu_0)^2}{2\sigma_0^2}} + \pi_1 \cdot \frac{1}{(2\pi\sigma_1^2)^{1/2}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}.$$

L06-Q2

Quiz(混合ガウス分布の確率密度関数)

X は混合ガウス分布 ($\pi_0 = \frac{3}{10}, \pi_1 = \frac{7}{10}, \mu_0 = 2, \mu_1 = 6, \sigma_0 = 2, \sigma_1 = \frac{1}{2}$) にしたがう. 確率密度関数 $f(x)$ のグラフの概形を描こう.
手で, または, `scipy.stats.norm` の浮動小数点計算で.

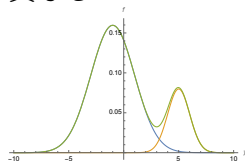
一般の混合ガウス分布: y の値が 2 個と限らず有限個

`scipy.stats` には GMM はないので, `scipy.stats.norm` を使って自分で計算する.

`sklearn.mixture.GaussianMixture`

一般の混合分布: ガウス分布とは限らない分布を重み π_y で「重ね合わせ」

注: 混合 $\pi_0 h_0(x) + \pi_1 h_1(x)$ と, 確率変数の和 $S = X_1 + X_2$ はもちろん異なる.



L06-Q3

Quiz(混合ガウス分布の確率)

X は混合ガウス分布 ($\pi_0 = \frac{3}{10}, \pi_1 = \frac{7}{10}, \mu_0 = 2, \mu_1 = 6, \sigma_0 = 2, \sigma_1 = \frac{1}{2}$) にしたがる。

- ① X の確率密度関数の $x = 4$ における値を、標準正規分布の確率密度関数 $\phi(x)$ で表そう。
- ② 確率 $P(X \leq 4)$ を、標準正規分布の累積分布関数 $\Phi(x)$ で表そう。

さらに `scipy.stats.norm` で浮動小数点数で求めてみると安心できるかも。