

線形回帰モデルにおけるベイズ推定

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 L05(2026-05-18 Mon)

最終更新: Time-stamp: "2026-08-06 Thu 08:42 JST hig"

今日の目標

- ベイズ推定の枠組みで線形回帰を説明できる



L04-Q1

TA Prob and Sol: 正規分布の母平均値のベイズ推定

確率変数 X は、正規分布にしたがう。すなわち、

$$f(x|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

母数 μ の事前分布を次として、 μ のベイズ推定を考える。

$$f(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}}.$$

- ① X のサイズ 1 の標本として、 x_1 が得られたとき、 μ の事後分布を求めよう。 μ の母平均値, 母分散を求めよう。
- ② X のサイズ 2 の標本として、 x_1, x_2 が得られたとき、 μ の事後分布を求めよう。 μ の母平均値, 母分散を求めよう。

略解

① $x = x_1$ とおく.

$$\begin{aligned}
 f_{\mu|x}(\mu|x) &= (\text{定数}) e^{-\frac{(x-\mu)^2}{2\sigma^2}} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}} \\
 &= (\text{定数})' e^{-\frac{1}{2}(\sigma^{-2} + \sigma_0^{-2})(\mu - \frac{\sigma^{-2}x + \sigma_0^{-2}\mu_0}{\sigma^{-2} + \sigma_0^{-2}})^2} \\
 &= \frac{1}{\sqrt{2\pi(\sigma^{-2} + \sigma_0^{-2})^{-1}}} e^{-\frac{1}{2}(\sigma^{-2} + \sigma_0^{-2})(\mu - \frac{\sigma^{-2}x + \sigma_0^{-2}\mu_0}{\sigma^{-2} + \sigma_0^{-2}})^2}
 \end{aligned}$$

よって、正規分布で、 $E[\mu|X] = \frac{\sigma^{-2}x + \sigma_0^{-2}\mu_0}{\sigma^{-2} + \sigma_0^{-2}}$. $V[\mu|X] = (\sigma^{-2} + \sigma_0^{-2})^{-1}$.
 $(\text{定数})'$ は、計算しなくても、分散さえわかれば、全確率が1であることから、 $\frac{1}{\sqrt{2\pi \text{分散}}}$ と定まる.

$$\begin{aligned}
& f_{\mu|x_1x_2}(\mu|x_1, x_2) \\
&= (\text{定数}) e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} e^{-\frac{(x_2-\mu)^2}{2\sigma^2}} e^{-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}} \\
&= (\text{定数})' e^{-\frac{1}{2}(2\sigma^{-2}+\sigma_0^{-2})(\mu-\frac{\sigma^{-2}\sum_i x_i+\sigma_0^{-2}\mu_0}{2\sigma^{-2}+\sigma_0^{-2}})^2} \\
&= \frac{1}{\sqrt{2\pi(2\sigma^{-2}+\sigma_0^{-2})^{-1}}} e^{-\frac{1}{2}(2\sigma^{-2}+\sigma_0^{-2})(\mu-\frac{2\sigma^{-2}\frac{1}{2}\sum_i x_i+\sigma_0^{-2}\mu_0}{2\sigma^{-2}+\sigma_0^{-2}})^2}
\end{aligned}$$

よって、正規分布で、 $E[\mu|X_1X_2] = \frac{2\sigma^{-2}\frac{1}{2}\sum_i x_i+\sigma_0^{-2}\mu_0}{2\sigma^{-2}+\sigma_0^{-2}}.$

$V[\mu|X_1X_2] = (2\sigma^{-2}+\sigma_0^{-2})^{-1}.$

L04-Q2

TA Prob and Sol: 正規分布の母平均値のベイズ推定

あるドーナツ屋さんの作るオールドファッションドーナツの重さ X (g) は、未知の母平均値 μ g, 既知の標準偏差 3 g の正規分布 $N(\mu, 3^2)$ にしたがう。母平均値 μ の事前分布は、 $N(100, 5^2)$ g と考えている。オールドファッションドーナツ 10 個の重さを測定したところ、標本平均値は 102 g だった。

- ① μ の事後分布を求めよう（改めて積分しなくても、共役事前分布についての「公式」を使ってよい）
- ② μ を MAP 推定しよう。
- ③ μ を EAP 推定しよう。
- ④ μ の、信頼係数 0.95 の HPD を求めよう。
- ⑤ μ が 100 より大きい確率を求めよう。
- ⑥ μ を MLE 推定（最尤推定）しよう。

略解

- ① 正規分布は共役事前分布であり，事後分布も正規分布．その母平均値は

$$\frac{\sigma_0^{-2}\mu_0 + N\sigma^{-2}\frac{1}{N}\sum_i x_i}{\sigma_0^{-2} + N\sigma^{-2}} = \frac{5^{-2} \cdot 100 + 10 \cdot 3^{-2} \cdot 102}{5^{-2} + 10 \cdot 3^{-2}} = \frac{26400}{259} = 101.93$$

その分散は，

$$(\sigma_0^{-2} + N\sigma^{-2})^{-1} = (5^{-2} + 10 \cdot 3^{-2})^{-1} = \frac{225}{259} = 0.869.$$

すなわち， $N(\frac{26400}{259}, \frac{225}{259})$.

- ② 正規分布の確率密度関数が最大となるのは母平均値で，
 $\frac{26400}{259} = 101.93$.
- ③ 正規分布の母平均値で， $\frac{26400}{259} = 101.93$.

- ④ 標準正規分布の累積分布関数を Φ とすると,
 $[\frac{26400}{259} - \Phi(0.975)(\frac{225}{259})^{1/2}, \frac{26400}{259} + \Phi(0.975)(\frac{225}{259})^{1/2}]$, すなわち,
 $[100.09, 103.78]$.
- ⑤ $P(\mu > 100) = 1 - \Phi(\frac{100 - \frac{26400}{259}}{\sqrt{225/259}}) = 0.981$.
- ⑥ 標本平均値 102 g.

今年までのあらすじ

x, y は非対称.

データ分析 2022 データ (x_i, y_i) ($i = 1, \dots, N$) があるとき

- (標本) 共分散 S_{xy} , (標本) 相関係数 r が計算できる
- 回帰直線 $y = ax + b$ を, 最小二乗法 $\arg \min_{a,b} \sum_i (y_i - (ax_i + b))^2$ で求められる. 正規方程式 $\leadsto a = \frac{S_{xy}}{S_{xx}}$

確率統計 I(2023)L? 確率変数 X, Y があるとき,

- (母) 共分散, 相関係数が計算できる

L05-Q1

Quiz(線形回帰モデルの最小二乗推定)

線形回帰モデルにしたがう観測値 (x_1, y_1) ($i = 1, \dots, N$) が与えられている。二乗誤差

$$S(x_1, \dots, y_1, \dots; \beta_0, \beta_1) = \sum_{i=1}^N (y_i - (\beta_0 + \beta_1 x_i))^2$$

を最小にする β_0, β_1 が，正規方程式

$$\begin{aligned} N\beta_0 + \left(\sum_i x_i\right)\beta_1 &= \sum_i y_i \\ \left(\sum_i x_i\right)\beta_0 + \left(\sum_i x_i^2\right)\beta_1 &= \sum_i x_i y_i \end{aligned}$$

を満たすことを示そう。

(確率変数でない) 変数 x に依存する確率変数 Y

確率統計 I(2023)L14

岩薩林 §9, §8.5

このドーナツ製造機の作るドーナツの重さ Y は, 温度 x によるらしい.
次の線形回帰モデルを仮定する.

$$Y = \beta_0 + \beta_1 \cdot x + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$

Y, ϵ : 連続型確率定数, β_0, β_1 : 回帰係数, $\sigma > 0$: 定数=パラメタ=母数

Y : 目的変数 (従属変数) ここでは確率変数

x : 説明変数 (独立変数) ここでは確率変数でない

標本から求めた r, a, b は, 母ナントカの推定値

ノイズ・誤差 $\epsilon = Y - (\beta_0 + \beta_1 x) \sim N(0, \sigma^2)$ より, ドーナツの重さ y を得る確率密度は,

$$f(y|x, \beta_0, \beta_1, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y - (\beta_0 + \beta_1 x))^2}{2\sigma^2}}$$

回帰係数の最尤推定

(正確に) 指定した温度 x_i ($i = 1, \dots, n$) で製造したときの重さが y_i ($i = 1, \dots, n$) である確率密度は, 独立分布なので積.

$$\begin{aligned} f(y_1, y_2, \dots, y_n | x_1, \dots, x_n, \beta_0, \beta_1, \sigma) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \beta_0 - \beta_1 \cdot x_i)^2}{2\sigma^2}} \\ &= (2\pi\sigma^2)^{-n/2} e^{-\sum_{i=1}^n \frac{(y_i - \beta_0 - \beta_1 \cdot x_i)^2}{2\sigma^2}} \\ &= L(\beta_0, \beta_1, \sigma) \end{aligned}$$

測定済のデータ $x_i, y_i \ i = 1, \dots, n$ を定数と思ったときの, 3 変数関数

$L(\beta_0, \beta_1, \sigma) = f(y_1, y_2, \dots, y_n | x_1, \dots, x_n, \beta_0, \beta_1, \sigma)$ の最大値は? $\leadsto$ 偏微分

微積分 II

$$0 = \frac{\partial L}{\partial \beta_0}(\beta_0, \beta_1, \sigma) = \frac{\partial L}{\partial \beta_1}(\beta_0, \beta_1, \sigma) = \frac{\partial L}{\partial \sigma}(\beta_0, \beta_1, \sigma)$$

ここでは最初の 2 つの等式だけ解く.

$$0 = \frac{\partial L}{\partial \beta_0} = (\text{定数}) \sum_i \frac{1}{\sigma^2} (y_i - \beta_0 - \beta_1 x_i) \times e^{\text{同じ}}$$

$$0 = \frac{\partial L}{\partial \beta_1} = (\text{定数}) \sum_i \frac{1}{\sigma^2} x_i (y_i - \beta_0 - \beta_1 x_i) \times e^{\text{同じ}}$$

しよせん, β_0, β_1 の連立 1 次方程式

正規方程式 岩薩林 §9.2

$$\begin{aligned} n\beta_0 + \left(\sum_i x_i\right)\beta_1 &= \sum_i y_i \\ \left(\sum_i x_i\right)\beta_0 + \left(\sum_i x_i^2\right)\beta_1 &= \sum_i x_i y_i \end{aligned}$$

加減法 $\rightsquigarrow$

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}},$$

$$\hat{\beta}_0 = \bar{y} - \frac{s_{xy}}{s_{xx}} \bar{x}$$

$$y = \frac{s_{xy}}{s_{xx}} x + \bar{y} - \frac{s_{xy}}{s_{xx}} \bar{x}$$

$$y - \bar{y} = \frac{s_{xy}}{s_{xx}} (x - \bar{x})$$

ここで,

$$\bar{x} = \frac{1}{n} \sum_i x_i \quad \text{平均値っぽい形}$$

$$\bar{y} = \frac{1}{n} \sum_i y_i \quad \text{平均値っぽい形}$$

$$s_{xy} = \frac{1}{n} \sum_i x_i y_i - \bar{x} \cdot \bar{y} \stackrel{\text{岩薩林 定理 1.5}}{=} \frac{1}{n} \sum_i (x_i - \bar{x})(y_i - \bar{y}) \quad \text{共分散っぽい形}$$

$$s_{xx} = \frac{1}{n} \sum_i x_i^2 - \bar{x}^2 \stackrel{\text{岩薩林 定理 1.2}}{=} \frac{1}{n} \sum_i (x_i - \bar{x})^2 \quad \text{分散っぽい形}$$

$Y = \hat{\beta}_0 + \hat{\beta}_1 \cdot x + \epsilon$ ($\hat{\beta}_0, \hat{\beta}_1$: 回帰係数) に代入.

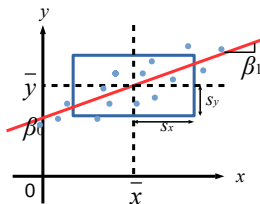
回帰直線

岩薩林 §9.1

推定結果 $\hat{\beta}_0, \hat{\beta}_1$ を係数とする xy 平面の直線

データ分析

$$y - \bar{y} = \frac{s_{xy}}{s_{xx}}(x - \bar{x})$$



回帰係数, 予測値の信頼区間

確率統計 II,III

進んだ内容（頻度主義の）

- 相関係数の区間推定
- 回帰係数 a, b の区間推定
- $r = 0$ を帰無仮説とする検定（無相関検定） $\leadsto$ 対立仮説 x, y には相関がある
- $a = 0$ を帰無仮説とする検定 $\leadsto$ 対立仮説 y は x による

x は確率変数なの？

- x を人間が設定して y を測定することもある
- X が確率変数の場合もある

両方を統合して、以下では X の条件付き確率で考える．

線形回帰モデルをベイズの枠組みで

ベイズの、観測値 X , 母数 Y という変数名はリセットして、回帰の X, Y を使う **確率変数と分布**

- X 分布未指定
- $\epsilon \sim N(0, \sigma^2)$, X と独立.
- $Y = \beta_0 + \beta_1 X + \epsilon$

$\epsilon = Y - (\beta_0 + \beta_1 X)$ なので, ϵ を消去して次のように書いても同じこと

$$Y|X = x_0 \sim N(\beta_0 + \beta_1 x_0, \sigma^2)$$

母数

- 回帰係数 β_0, β_1 , 分散 σ^2 について事前事後分布を考える
- 回帰係数 β_0, β_1 について事前事後分布を考える (分散 σ^2)

観測値 (y_i, x_i) ($i = 1, \dots, N$).

モデル=尤度=条件付き分布

$$f_{y|x\beta_0\beta_1\sigma}(y_1, \dots, y_N | x_1, \dots, x_N, \beta_0, \beta_1, \sigma) \\ = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_1 - (\beta_0 + \beta_1 x_1))^2}{2\sigma^2}} \dots \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_N - (\beta_0 + \beta_1 x_N))^2}{2\sigma^2}}.$$

事前分布 (例えば)

$$\beta_0 \sim N((0 \text{ や主観的な予想}), (\text{大きい値})^2),$$

$$\beta_1 \sim N((0 \text{ や主観的な予想}), (\text{大きい値})^2),$$

$$(\sigma \sim \text{逆ガンマ分布} \quad \text{or} \quad \text{半正規分布}(0, (\text{大きい値})^2))$$

$Z \sim N(0, 1^2)$ のとき, $|Z|$ は半正規分布にしたがう.

事後分布同時分布

$$f_{\dots|\dots}(\beta_0, \beta_1 | y_1, \dots, y_N, x_1, \dots, x_N)$$

$$f_{\dots|\dots}(\beta_0, \beta_1, \sigma | y_1, \dots, y_N, x_1, \dots, x_N)$$

σ 固定のとき, β_0, β_1 の事前分布が独立な正規分布でも, 事後分布は独立でない (共分散行列が単位行列でない) 2次元正規分布になる
したがって, σ 固定 (既知) のとき, 2次元正規分布はこのモデルの共役事前分布.

σ 未知のとき, 2次元正規分布-逆ガンマ分布がこのモデルの共役事前分布.

L05-Q2

Quiz(線形回帰モデルのベイズ推定)

線形回帰モデルにしたがう観測値 (x_1, y_1) ($i = 1, \dots, N$) が与えられている.

σ を定数として、尤度が

$$f(y|x, \beta_0, \beta_1) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y - (\beta_0 + \beta_1 x))^2}{2\sigma^2}}$$

であり、事前分布が

$$f(\beta_0, \beta_1) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(\beta_0 - \mu_0)^2}{2\sigma_0^2}} \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(\beta_1 - \mu_1)^2}{2\sigma_1^2}}$$

であるとき、 β_0, β_1 の MAP 推定値は、どのような方程式を満たすか。

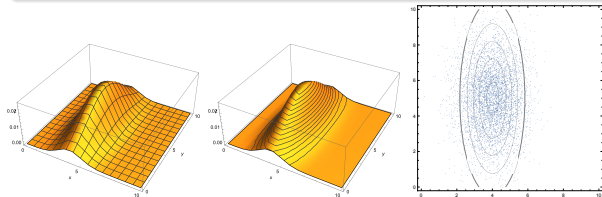
2次元正規分布 (X, Y が独立でない)

定義 (一般の2次元正規分布)

連続型確率変数 X, Y が 次の形 (指数関数の肩が x, y の2次式) の同時確率密度関数を持つとき, X, Y は2次元正規分布にしたがうという.

$$f(x, y) = C \times e^{-ax^2 - by^2 + cxy + px + qy}$$

$a, b > 0, c, p, q$ は, 全確率の積分が収束するような数. C は全確率の積分が1になるような値.



一般の2次元正規分布を行列形で

定義 (一般の2次元正規分布の同時確率密度関数)

2次元正規分布 $N(\boldsymbol{\mu}, \Sigma)$ の同時確率密度関数は、確率変数を $\mathbf{X} = {}^t(X, Y)$ とするとき、

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^2 \det \Sigma}} e^{-\frac{1}{2} {}^t(\mathbf{x}-\boldsymbol{\mu}) \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})}$$

$$\text{母平均値 (ベクトル)} \boldsymbol{\mu} = \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix} \stackrel{\text{要証明}}{=} \begin{pmatrix} E[X] \\ E[Y] \end{pmatrix}$$

$$\text{母共分散行列} \Sigma = \begin{pmatrix} S_{XX} & S_{XY} \\ S_{XY} & S_{YY} \end{pmatrix} \stackrel{\text{要証明}}{=} \begin{pmatrix} V[X] & \text{Cov}[X, Y] \\ \text{Cov}[X, Y] & V[Y] \end{pmatrix}$$

- たしかに正規分布
- $\text{Cov}[X, Y] = 0$ のとき独立なケースに帰着.

`pymc.MvNormal(mu=[ $\mu_X, \mu_Y$ ], cov=[[ $S_{XX}, S_{XY}$ ], [ $S_{XY}, S_{YY}$ ]])`

予測分布

β_0, β_1 の事後分布, $x = x_*$ という条件のもとでの y の分布 $Y_* | x_*, x_i, y_i$.

$$f(y_*) = \int d\beta_0 d\beta_1 \beta_1 \cdot f(x = \frac{y_* - \beta_0}{\beta_1} | \beta_0, \beta_1) f(\beta_0, \beta_1 | x_i, y_i)$$

σ 固定, β_0, β_1 の事前分布が正規なら正規になる.

PyMC 使用の実用的注意

PyMC は、モデルに従って、 $(\beta_0, \beta_1, \sigma)$ を多数生成する（サンプルを出力する）ので、それが標本になるような母分布，と思う．PyMC は事後分布の周辺分布も表示する．

- `plot_trace()` は、事後分布（の周辺分布）に加えて、推定の経過を表示するため、推定の信頼性の情報を与える．
- 左側は、3 回の「試行」でのヒストグラムをなめらかにした事後分布（の周辺分布）．3 回とも近い分布になっていれば、正しい推定である可能性がある．
- 左は横軸を「モンテカルロステップ」とした時系列．水平な一直線のまわりに揺れていれば、正しい推定である可能性がある．複数の一直線のつぎはぎの場合は疑う余地がある．
- 左右とも、黒い縦線は「divergence」が生じたステップ-値．その個数は `sample()` の結果に表示される．多い場合は疑う余地がある．減らすには、`sample()` のオプション `target_accept` を 1 に近い値にする（デフォルトは 0.80）．このオプションはモンテカルロの「受理」確率の目標値を指定する．