ベイズ推定

樋口さぶろお https://hig3.net

龍谷大学大学院 理工学研究科 数理情報学専攻

理論物理学特論　L05 (2022-10-19 Wed)

最終更新: Time-stamp: ”2022-10-31 Mon 10:44 JST hig”

今日の目標

- 分布のパラメタを確率変数とみなして事後分布を求められる
- マルコフ連鎖の定義と例を説明できる

略解 I

L04-Q1

Quiz 解答: 混合ガウス分布のナイーブベイズ

$$P(Y=0|X=1)=\frac{\frac{1}{(2\pi 2^2)^{1/2}}\mathrm{e}^{-\frac{(1-2)^2}{2\cdot 2^2}}\cdot\frac{3}{10}}{\frac{1}{(2\pi 2^2)^{1/2}}\mathrm{e}^{-\frac{(1-2)^2}{2\cdot 2^2}}\cdot\frac{3}{10}+\frac{1}{(2\pi(1/2)^2)^{1/2}}\mathrm{e}^{-\frac{(1-6)^2}{2\cdot(1/2)^2}}\cdot\frac{7}{10}}$$

L04-Q2

Quiz 解答: 混合ガウス分布のナイーブベイズ推定 (訓練, 検証)

$y=0,1$ それぞれについて標本平均値, 不偏標本分散を求めて, 母ナントカが次の値であるとして考える.

$$\pi_0=\frac{2}{5},\mu_0=-2,\sigma_0^2=2.\pi_1=\frac{3}{5},\mu_1=2,\sigma_1^2=1.$$

略解 II

$x = 3$ という条件のもとでの $y = 0$ の条件付き確率は,

$$\frac{\frac{2}{5}\frac{1}{\sqrt{2\pi\cdot 2}}\mathrm{e}^{-\frac{1}{2\cdot 2}(0-(-2))^2}}{\frac{2}{5}\frac{1}{\sqrt{2\pi\cdot 2}}\mathrm{e}^{-\frac{1}{2\cdot 2}(0-(-2))^2}+\frac{3}{5}\frac{1}{\sqrt{2\pi\cdot 1}}\mathrm{e}^{-\frac{1}{2\cdot 1}(0-2)^2}} = 0.56 > \frac{1}{2}.$$

$y = 0$ と推定する.

L04-Q3

Quiz 解答: ベイズ推定

1 回の検査の結果 $X = \begin{cases} 80 & \text{(陽性)} \\ 20 & \text{(陰性)} \end{cases}$, 病状 $Y = \begin{cases} 100 & \text{(病気)} \\ 0 & \text{(病気でない)} \end{cases}$ で表す.

略解 III

1.

$$f_{\mathrm{Y}}(100) = 0.005,\, f_{\mathrm{X|Y}}(80|100) = 0.99,\, f_{\mathrm{X|Y}}(80|0) = 0.02$$

1 回目の検査結果の同時確率分布

$y\backslash x$	80	20	計
100	真陽性 0.99×0.005	偽陰性 0.01×0.005	0.005
0	偽陽性 0.02×0.995	真陰性 0.98×0.995	0.995

$$\begin{aligned} f_{\mathrm{Y|X}}(100|80) =& \frac{f_{\mathrm{X|Y}}(80|100) \times f_{\mathrm{Y}}(100)}{f_{\mathrm{X|Y}}(80|100) \times f_{\mathrm{Y}}(100) + f_{\mathrm{X|Y}}(80|0) \times f_{\mathrm{Y}}(0)} \\ =& \frac{0.99 \times 0.005}{0.99 \times 0.005 + 0.02 \times 0.995} = 0.199195. \end{aligned}$$

略解 IV

2. $f_{\mathrm{Y}}(100) = 0.199195, f_{\mathrm{Y}}(0) = 1 - 0.199195$ として, 同じ計算を繰り返す.

$$
\begin{aligned}
f_{\mathrm{Y}|\mathrm{X}}(100|80) =& \frac{f_{\mathrm{X}|\mathrm{Y}}(80|100) \times f_{\mathrm{Y}}(100)}{f_{\mathrm{X}|\mathrm{Y}}(80|100) \times f_{\mathrm{Y}}(100) + f_{\mathrm{X}|\mathrm{Y}}(80|0) \times f_{\mathrm{Y}}(0)} \\
=& \frac{0.99 \times 0.199195}{0.99 \times 0.199195 + 0.02 \times (1 - 0.199195)} = 0.924884.
\end{aligned}
$$

別解 2 回中の陽性の回数 Z を考えると,
$f_{\mathrm{Z}|\mathrm{Y}}(2|100) = 0.99^2, f_{\mathrm{Z}|\mathrm{Y}}(2|80) = 0.01^2$ である.

$$
\begin{aligned}
f_{\mathrm{Y}|\mathrm{Z}}(100|2) =& \frac{f_{\mathrm{Z}|\mathrm{Y}}(2|100) \times f_{\mathrm{Y}}(100)}{f_{\mathrm{Z}|\mathrm{Y}}(2|100) \times f_{\mathrm{Y}}(100) + f_{\mathrm{Z}|\mathrm{Y}}(2|0)) \times f_{\mathrm{Y}}(0)} \\
=& \frac{0.99^2 \cdot 0.005}{0.99^2 \cdot 0.005 + 0.02^2 \cdot 0.995} = 0.924884.
\end{aligned}
$$

ここまで来たよ

4 ベイズ推定

5 ベイズ推定
- ベイズ推定
- マルコフ連鎖
- 確率ベクトル, 転置確率行列
- 定常分布, 分布の時間発展

ベイズ推定の考え方

$$\boxed{\text{事後確率}\ f_{\mathrm{Y}|\mathrm{X}}(y|x)} \longleftarrow \boxed{\text{事前確率}\ f_{\mathrm{Y}}(y)}$$

$$\uparrow$$

$$\boxed{\text{情報}\ X = x}$$

事前 (prior), 事後 (posterior)
主観 (subjective) 確率
事前, 事後確率は, その時点で観察者の持つ情報に依存する (主観的である) という考え方.

推定すべき母分布のパラメタを確率変数と考える

未知パラメタ μ, σ の正規分布 $X \sim \mathrm{N}(\mu, \sigma^2)$ から, X の標本 $\{X\}$ または $\{X_i\}_{i=1}^n$ から 未知パラメタ μ, σ を推定することを考える.
従来の方法 不偏標本平均値, 不偏標本分散, 信頼区間
ベイズ推定 μ, σ を確率変数 Y であるかのように考え, (主観的な) 事前分布 $f(\mu, \sigma)$ が, X の標本の情報で, 事後分布 $f(\mu, \sigma|x)$ に修正される, と見る.

(病気の例では, 被験者が病気であるないを確率変数 $Y = 0, 100$ とみなしたようなもの. Y はベルヌーイ分布 $\mathrm{B}(1, p)$ にしたがう. $p \leftrightarrow (\mu, \sigma)$.)

説明の簡単のため, なぜか $\sigma^2 = 2^2$ はわかっていて固定されているとして, $\mu = Y$ のみが確率変数であるとする. パラメタによく使う記号 $Y = \mu = \Theta, \theta$.
X の分布=条件付き分布

$$X \sim \mathrm{N}(\theta, 2^2) \rightsquigarrow f_{\mathrm{X}|\Theta}(x|\theta) = \frac{1}{\sqrt{2\pi \cdot 2^2}} \mathrm{e}^{-\frac{1}{2 \cdot 2^2}(x-\theta)^2}$$

事前分布 $\theta = \mu$ の事前分布は，主観的に，または領域知識から自由に決めていいのだが，いま，これも正規分布にしたがうとする．

$$\Theta \sim \mathrm{N}(\mu_0, \sigma_0^2) \rightsquigarrow f(\theta) = \frac{1}{\sqrt{2\pi \cdot \sigma_0^2}} \mathrm{e}^{-\frac{1}{2\cdot\sigma_0^2}(\theta-\mu_0)^2}$$

さらに気分を出すため，$\mu_0 = -2, \sigma_0^2 = 3^2$ と固定しよう．
事後分布 X の標本サイズ 1 の標本 $\{-3\}$ が得られたときの，事後分布

$$f_{\Theta|\mathrm{X}}(\theta|-3) = \frac{\frac{1}{\sqrt{2\pi\cdot 2^2}}\mathrm{e}^{-\frac{1}{2\cdot 2^2}(-3-\theta)^2} \cdot \frac{1}{\sqrt{2\pi\cdot 3^2}}\mathrm{e}^{-\frac{1}{2\cdot 3^2}(\theta-(-2))^2}}{(\text{分子の } \theta \text{ 積分})} = C\mathrm{e}^{A}$$

L05-Q1

事後分布

この事後分布が正規分布であることを示そう．母平均値と母分散を求めよう．

この場合はたまたま事後分布が (事前分布と同じ) 正規分布にもどった!
事前分布 (この場合正規分布) が「共役 (conjugate) 事前分布である」と表現する

一般に，サイズ n の標本 $\{x_i\}_{i=1,\ldots,n}$ が得られたときの Θ の事後分布を考える．

$$f_{\Theta|\{X\}}(\theta|\{x_i\}_{i=1}^n) = \frac{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\cdot\sigma^2}} \mathrm{e}^{-\frac{1}{2\cdot\sigma^2}(x_i-\theta)^2} \cdot \frac{1}{\sqrt{2\pi\cdot\sigma_0^2}} \mathrm{e}^{-\frac{1}{2\cdot\sigma_0^2}(\mu-\mu_0)^2}}{(\text{分子の}\ \theta\ \text{積分})} = C\mathrm{e}^A$$

$$A = -\frac{1}{2\frac{\sigma^2}{n}\cdot(1+\frac{\sigma^2}{n\sigma_0^2})^{-1}}\left(\theta - \frac{\frac{1}{n}\sum x_i + \frac{1}{n}\frac{\sigma^2}{\sigma_0^2}\mu_0}{1+\frac{1}{n}\frac{\sigma^2}{\sigma_0^2}}\right)^2$$

また，事後分布も正規分布で，

- 母平均値 自分の言葉で
- 母分散 自分の言葉で

事後分布が式で書けたので，その母比率，母期待値なども式で計算できる．
一般には，マルコフ連鎖モンテカルロ法などで数値計算する必要がある．

信用区間

(伝統的な=頻度主義での) 母ナントカの区間推定では, 信頼係数 (confidence coefficient) $1-\alpha = 0.99, 0.95$ の信頼区間 (confidence interval) を考えた.

- 正しい理解「同様の区間推定を繰り返し行ったとき, 真の母ナントカが (各回得られる) 区間に含まれる確率」
- よくある誤解「真の母ナントカがこの (1 個の) 区間に含まれる確率」

ベイズ推定の立場では, 母ナントカ ($Y=\mu$) が確率変数なので,

$$P(a<\mu<b)=0.99$$

という後者の主張に意味がある.

定義 (信用集合, 信用区間)

$$\int_C f_{\Theta|\mathrm{X}}(\theta|x)\mathrm{d}\theta \geq 1-\alpha$$

となる集合 C を $(1-\alpha)$ 信用集合 (credible set) という. $C \subset \mathbb{R}$ で C が区間のとき, 信用区間 (credible interval) という.

信用区間は一意ではない. 上側片側, 下側片側, 両側=等裾 (とうきょ) に相当するものや, それ以外のものが許される. 等裾がもっともよく使われる.

定義 (最高事後密度信用集合)

信用集合 C で, ある $c \in \mathbb{R}$ で次の形に書けるもの. 最高事後密度 =HPD=highest posterior density.

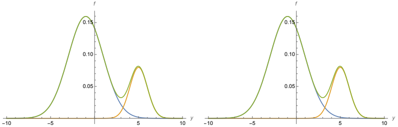
$$C = \{\theta | f_{\Theta|\mathrm{X}}(\theta|x) \geq c\}$$

L05-Q2

Quiz(ベイズ推定の信用集合)

図示したパラメタの事後分布を考える.

1. 95%等裾信用集合を求めよう
2. 95% 最高事後密度信用集合を求めよう



混合ガウス分布のベイズ推定

混合ガウス分布

$$f_{\mathrm{Y}}(y) = \theta f(y;\mu_0,\sigma_0) + (1-\theta) f(y;\mu_1,\sigma_1)$$

を考える. ここで, パラメタ $0 \le \theta \le 1$ を確率変数 X であるかのように扱う. $\mu_0, \sigma_0, \mu_1, \sigma_1$ は固定, 既知. ガウス分布でなくても以下の話はできる.
事前分布 $f_\Theta(\theta)$ を決める. 特に事前の情報がないとき,

$$f_\Theta(\theta) = \begin{cases} 1 & (0 \le \theta \le 1) \\ 0 & (\text{他}) \end{cases}$$

としてもよい.
Y のデータ $\{y_i\}_{i=1}^n$ が得られたとき, θ の情報を得る試みとして, 事後分布 $f_{\Theta|\mathrm{Y}}(\theta|\{y_i\})$ を得たい.

ベイズの定理から

$$f_{\Theta|\mathrm{Y}}(\theta|\{y_i\}) = \frac{\prod_{i=1}^{n}(\theta f(y_i;\mu_0,\sigma_0) + (1-\theta)f(y_i;\mu_1,\sigma_1))}{\int_{-\infty}^{+\infty}\prod_{i=1}^{n}(\theta f(y_i;\mu_0,\sigma_0) + (1-\theta)f(y_i;\mu_1,\sigma_1))\mathrm{d}\theta}.$$

この確率密度関数を具体的に積分して確率や期待値を求めるのはたいへん
共役事前分布ではなかった
$\rightsquigarrow$ 伏線: マルコフ連鎖モンテカルロ法 (のギブズサンプリング)

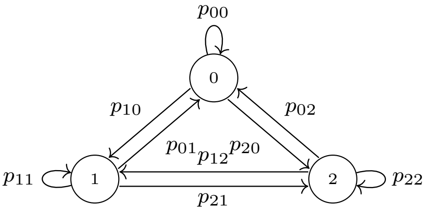
ここまで来たよ

4 ベイズ推定

5 ベイズ推定
- ベイズ推定
- マルコフ連鎖
- 確率ベクトル, 転置確率行列
- 定常分布, 分布の時間発展

マルコフ連鎖の例 I

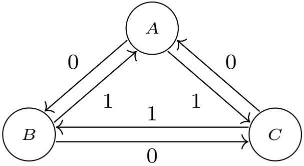
時間 t と共に変化する猫の状態
0: 食べる 1: 寝る 2: 遊ぶ を考える.
状態空間 $S = \{0, 1, 2\}$
状態 $x = 0$: 食べる,...
時刻 t の猫の状態 $X(t) = 0, 1, 2.$



推移確率　　条件付き確率

$$p_{先\ 元} = p_{xy} = P(X(t) = x | X(t-1) = y)$$

比較: オートマトンの状態遷移図では, 矢印の数値は確率でなく入力



上の量を, 世の中では p_{xy} でなく p_{yx} と書くこともある.
マルコフ連鎖「このような」状態空間と, 推移確率の組.

$p(x,t)$ の漸化式

略記 $p(x,t) = P(X(t) = x)$.
条件付き確率の全確率の法則から,

$$\begin{aligned} p(0,t) =& p_{00} \cdot p(0,t-1) + p_{01} \cdot p(1,t-1) + p_{02} \cdot p(2,t-1) \\ p(1,t) =& p_{10} \cdot p(0,t-1) + p_{11} \cdot p(1,t-1) + p_{12} \cdot p(2,t-1) \\ p(2,t) =& p_{20} \cdot p(0,t-1) + p_{21} \cdot p(1,t-1) + p_{22} \cdot p(2,t-1) \end{aligned}$$

$$\begin{pmatrix} p(0,t) \\ p(1,t) \\ p(2,t) \end{pmatrix} = \begin{pmatrix} p_{00} & p_{01} & p_{02} \\ p_{10} & p_{11} & p_{12} \\ p_{20} & p_{21} & p_{22} \end{pmatrix} \begin{pmatrix} p(0,t-1) \\ p(1,t-1) \\ p(2,t-1) \end{pmatrix}$$

$$p(x,t) = \sum_{y=0,1,2} p_{xy} \cdot p(y,t-1). \qquad (x = 0,1,2)$$

転置推移確率行列 M, 分布=ベクトル $\vec{p}(t)$ で書くと (x, y が成分番号)

$$\vec{p}(t) = M\vec{p}(t-1).$$

この漸化式を解くと, $\vec{p}(t) = M\vec{p}(t-1) = M^2\vec{p}(t-2) = \cdots = M^t\vec{p}(0)$.

L05-Q3

Quiz(マルコフ連鎖の転置推移確率行列)

状態 $x = 0, 1, 2, 3$ からならる状態空間を持つマルコフ連鎖を考える. 時刻 $t = 0$ に状態は $x = 1$ である. 時刻 t に状態が x であるという条件のもとで, 時刻 $t + 1$ での状態は,

- 確率 $\frac{1}{7}$ で状態 $x - 1$,
- 確率 $\frac{2}{7}$ で状態 $x + 1$,
- 確率 $\frac{4}{7}$ で状態 x.

ただし, 上の条件付き確率で, 状態 $x = -1$ とは状態 $x = 3$ のこと, 状態 $x = 4$ とは状態 $x = 0$ のことをいう.

1. 推移図を書こう.
2. 転置推移確率行列を書こう.
3. 時刻 $t = 0$ における分布 $\vec{p}(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ であるとき,$\vec{p}(1)$ を求めよう.

L05-Q4

Quiz(マルコフ連鎖の転置推移確率行列)

状態 $x = 0, 1, 2, 3$ からなる状態空間を持つマルコフ連鎖を考える. 時刻 $t = 0$ に状態は $x = 1$ である. 時刻 t に状態が x であるという条件のもとで, 時刻 $t + 1$ での状態は,

- 確率 $\frac{1}{7}$ で状態 $x - 1$,
- 確率 $\frac{2}{7}$ で状態 $x + 1$,
- 確率 $\frac{4}{7}$ で状態 x.

ただし, 上の条件付き確率で, 状態 $x = -1$ とは状態 $x = 0$ のこと, 状態 $x = 4$ とは状態 $x = 3$ のことをいう.

1. 推移図を書こう.
2. 転置推移確率行列を書こう.
3. 時刻 $t = 0$ における分布 $\vec{p}(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ であるとき,$\vec{p}(1)$ を求めよう.

ここまで来たよ

4 ベイズ推定

5 ベイズ推定
- ベイズ推定
- マルコフ連鎖
- 確率ベクトル, 転置確率行列
- 定常分布, 分布の時間発展

線形代数のりで確率ベクトル, 転置確率行列の言葉を準備 I

定義 (非負ベクトル)

m 次元ベクトル $\begin{pmatrix} p_0 \\ p_1 \\ \vdots \\ p_{m-1} \end{pmatrix}$ が $p_i \geq 0$ を満たすとき, 非負ベクトルという.

定義 (確率ベクトル)

非負ベクトル $\begin{pmatrix} p_0 \\ p_1 \\ \vdots \\ p_{m-1} \end{pmatrix}$ が, 規格化 $\sum_{i=0}^{m-1} p_i = 1$ を満たすとき, 確率ベクトルという.

線形代数のりで確率ベクトル, 転置確率行列の言葉を準備 II

離散型確率分布 $f(x)$ は, 確率ベクトルと 1 対 1 に対応.

$$\vec{p} = \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} f(0) \\ f(1) \\ f(2) \\ \vdots \end{pmatrix}$$

定義 (非負行列 ($n \times m = 3 \times 2$ の例で))

実行列 $\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \\ p_{20} & p_{21} \end{pmatrix}$ が $p_{ij} \geq 0$ を満たすとき**非負行列**という.

定義 (転置確率行列)

行列 $M = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \\ p_{20} & p_{21} \end{pmatrix}$ の各列ベクトルが確率ベクトルであるとき, つまり
$\forall j \sum_{i=0}^{n-1} p_{ij} = 1$ であるとき, M を**転置確率行列**という.

転置推移確率行列

定義 (転置推移確率行列)

マルコフ連鎖に現れる, 転置確率行列

$$M = \begin{pmatrix} p_{00} & p_{01} & p_{02} \\ p_{10} & p_{11} & p_{12} \\ p_{20} & p_{21} & p_{22} \end{pmatrix}$$

を, マルコフ連鎖の**転置推移確率行列**という. $m \times m$ 正方行列.

世の中では, それぞれの転置で (${}^{\mathrm{t}}M$: **推移確率行列**) 次のように書くのがふつう.

$${}^{\mathrm{t}}\vec{p}(t) = {}^{\mathrm{t}}\vec{p}(t-1)\, {}^{\mathrm{t}}M$$

漸化式 $\vec{p}(t) = M\vec{p}(t-1)$ の解

$$\vec{p}(t) = M\vec{p}(t-1) = M^2\vec{p}(t-2) = \cdots = M^t\vec{p}(0).$$

命題 (転置確率行列と確率ベクトルの積)

M が転置確率行列, $\vec{p}$ が確率ベクトルのとき, $M\vec{p}$ も確率ベクトル

証明:

確率過程とマルコフ連鎖

確率過程　時間 t に依存する確率変数

もちろん反復試行から来る独立な確率変数群 $R(t)$ も時間 t に依存するけど, 独立じゃない場合がおもしろい

離散時間マルコフ連鎖は, 次の性質を満たす確率過程.

離散時間　t が離散的. $t = 0, 1, 2, \ldots$

マルコフ Markov　$\vec{p}(t)$ が直前の時刻の分布 $\vec{p}(t-1)$ だけから決まる. 転置推移確率行列 p_{xy} で表現できる.

連鎖 chain　状態空間 $S \ni x$ が離散的.

いま考えてる, 時間空間離散のランダムウォークや猫は離散時間マルコフ連鎖の典型例.

離散時間マルコフ連鎖の (確率) 分布は, 確率ベクトルで表現できる.

離散時間マルコフ連鎖の遷移確率は, 転置確率行列で表現できる.

ここまで来たよ

4 ベイズ推定

5 ベイズ推定
- ベイズ推定
- マルコフ連鎖
- 確率ベクトル, 転置確率行列
- 定常分布, 分布の時間発展

定義 (定常分布)

$\vec{p} = M\vec{p}$ となる分布 $\vec{p}$ を定常分布という.

意味 自分の言葉でどうぞ

線形代数の言葉で言うと, 定常分布は, 転置推移確率行列 M の

自分の言葉でどうぞ

建設的心配性大爆発

- 定常分布っていつでもある? $\rightsquigarrow$ Yes
- 固有値 (の絶対値) が 1 より大きな固有ベクトルはあるの? $\rightsquigarrow$ No

状態数 m が有限のとき, ペロン・フロベニウスの定理から言える.

命題 (固有値 1 の存在)

転置確率行列は 固有値 1 を持つ.

証明