

二項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 先端理工学部研究科 数理・情報科学コース

計算科学特論 L03(2024-10-09 Wed)

最終更新: Time-stamp: "2024-10-09 Wed 13:55 JST hig"

今日の目標

- 二項モデルのパラメタをベイズ推定できる
- ベータ分布 $\text{beta}(a, b)$ を説明できる



L02-Q5

TA Prob and Sol: 連続型の条件付き確率

ある畑でとれる柿の実の $\frac{3}{10}$ は品種 A, $\frac{7}{10}$ は品種 B である.
 品種 A の実の質量は, 母平均値 200 g, 母分散 20^2 g^2 ,
 品種 B の実の質量は, 母平均値 180 g, 母分散 10^2 g^2 の, それぞれ正規分布にしたがう.

- ① 畑でとれる柿の実の質量の分布を求めよう (確率密度関数を書こう).
- ② 畑でとれたひとつの柿の実の質量が 190 g であるとき, 品種 A である確率を求めよう.

略解

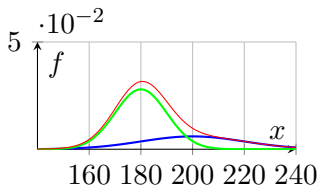
品種 A, B を $Y = 0, 1$, 実の質量を $X \text{ g}$ とする.

周辺分布は $Y \sim B(1, \frac{7}{10})$.

条件付き分布は $X|Y = 0 \sim N(200, 20^2)$, $X|Y = 1 \sim N(180, 10^2)$.

1

$$\begin{aligned}
 f_X(x) &= \sum_{y=0,1} f_{X|Y}(x|y) f_Y(y) \\
 &= \frac{1}{(2\pi 20^2)^{1/2}} e^{-\frac{(x-200)^2}{2 \cdot 20^2}} \cdot \frac{3}{10} + \frac{1}{(2\pi 10^2)^{1/2}} e^{-\frac{(x-180)^2}{2 \cdot 10^2}} \cdot \frac{7}{10}.
 \end{aligned}$$

— $f(x, 0)$ — $f(x, 1)$ — $f_X(x)$

2

$$\begin{aligned}
 p' = f_{Y|X}(0|190) &= \frac{f_{X|Y}(190|0)f_Y(0)}{f_X(190)} \\
 &= \frac{\frac{1}{(2\pi 20^2)^{1/2}} e^{-\frac{(190-200)^2}{2 \cdot 20^2}} \cdot \frac{3}{10}}{\frac{1}{(2\pi 20^2)^{1/2}} e^{-\frac{(190-200)^2}{2 \cdot 20^2}} \cdot \frac{3}{10} + \frac{1}{(2\pi 10^2)^{1/2}} e^{-\frac{(190-180)^2}{2 \cdot 10^2}} \cdot \frac{7}{10}} = 0.238.
 \end{aligned}$$

すなわち， Y の事前分布は $B(1, \frac{7}{10})$ ，事後分布は $B(1, 1 - p')$ 。

確率密度で書いたベイズの定理 I

定理 (ベイズの定理)

岩薩林 確率・統計 (2.18),(2.19)

Hoff ベイズ p.29

$$f_{X|Y}(x|y) \stackrel{\text{定義}}{=} \frac{f(x, y)}{f_Y(y)} = \frac{f_{Y|X}(y|x) f_X(x)}{\int_{-\infty}^{+\infty} f_{Y|X}(y|x') f_X(x') dx'},$$

$$f_{Y|X}(y|x) \stackrel{\text{定義}}{=} \frac{f(x, y)}{f_X(x)} = \frac{f_{X|Y}(x|y) f_Y(y)}{\int_{-\infty}^{+\infty} f_{X|Y}(x|y') f_Y(y') dy'}.$$

$f_{X|Y}(x|y)$ を $f_{Y|X}(y|x')$ (と $f_X(x')$) で書き表す式,

$f_{Y|X}(y|x)$ を $f_{X|Y}(x|y')$ (と $f_Y(y')$) で書き表す式.

定義の分子分母を, 乗法公式と全確率の公式で書き直しただけ (なので余分の記憶は不要).

$\frac{1}{f_Y(y)}$ は x に関して定数, $\frac{1}{f_X(x)}$ は y に関して定数, と思って忘れても, 比だけは求められる.

ここまで来たよ

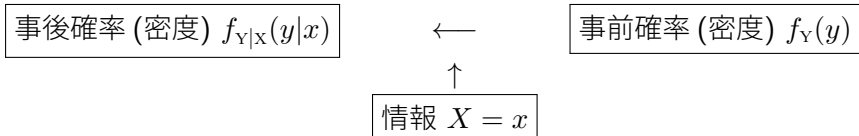
2 離散-連続混合型分布・条件付き分布・事前分布・事後分布

3 二項モデルとベイズ推定

- ベイズ推定
- 二項モデル（複数のベルヌイ分布）
- 二項モデル（二項分布）

ベイズ推定の考え方

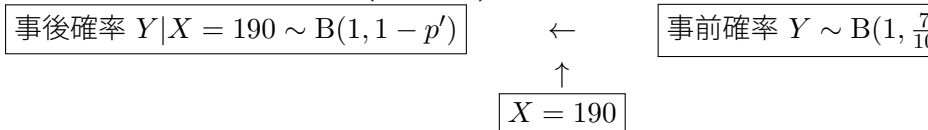
Hoff ベイズ p.29



追加の情報が得られるたびに、事後確率 (密度) は正確になっていく。

主観確率=信念 berief 事前確率密度に主観を許す考え方. 事後確率密度にも主観が含まれる.

前回の Quiz では離散 $Y = 0, 1$ (品種 AB) だったので,



二項分布

$Y \sim B(n, p)$: パラメタ n, p の**二項分布** (binomial distribution). Y は, 表の
 での確率 p のコインを n 回投げたときの表の回数. $n = 1$ はパラメタ p
 のベルヌイ分布.

ここまで来たよ

2 離散-連続混合型分布・条件付き分布・事前分布・事後分布

3 二項モデルとベイズ推定

- ベイズ推定
- 二項モデル（複数のベルヌイ分布）
- 二項モデル（二項分布）

二項モデル (複数のベルヌイ分布)

Hoff ベイズ §3.1

値が不明なパラメタ θ の確率で表 $Y = 1$ が出るコインを N 回投げた結果を, $Y_1, \dots, Y_N = 0, 1$ とする.

Y_i は独立同分布にしたがう. $Y_i \stackrel{\text{i.i.d.}}{\sim} B(1, \theta)$ ($n = 1, p = \theta$ の二項分布).
 θ は不明ということで, 分布をもつ確率変数 Θ と思って, 前回の Quiz と

- $Y_1, \dots, Y_N$ の組 $\leftrightarrow$ 柿の質量 X
- $\Theta \leftrightarrow$ 柿の品種 AB Y

と対応させ, θ の確率分布を $y_1, \dots, y_N$ の情報で更新していきたい.
 $N = 1$ のとき, $Y_1 = y_1$ を得る確率は (なぜか $\text{given } |$ を使って),

$$f_{Y|\theta}(y_1|\theta) = \theta^{y_1}(1 - \theta)^{1-y_1}.$$

一般の N のとき,

$$f_{Y|\theta}(y_1, \dots, y_N|\theta) = \theta^{\sum_i y_i} (1 - \theta)^{N - \sum_i y_i}.$$

θ の分布

θ についての主観的確率 (信念) は, 連続型確率変数 Θ で表現できる. 何の先入観も持たないときの事前確率密度は $\Theta \sim U(0, 1)$, すなわち, 一様分布

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

と考えられる.

このとき, $f_{y|\theta}(y_1, \dots, y_N|\theta)$ は条件付き確率であることが納得できる.

ベイズの定理より, $Y_1 = y_1, \dots, Y_N = y_N$ という結果を得たときの θ の事後確率密度は,

$$\begin{aligned} f_{\theta|Y}(\theta|y_1, \dots, y_N) &= \frac{f(y_1, \dots, y_N, \theta)}{f_Y(y_1, \dots, y_N)} \\ &= \frac{f_{Y|\theta}(y_1, \dots, y_N|\theta)f_{\theta}(\theta)}{f_Y(y_1, \dots, y_N)} \\ &= \frac{\theta^{\sum_i y_i} (1 - \theta)^{N - \sum_i y_i} \cdot 1}{f_Y(y_1, \dots, y_N)} \quad (0 \leq \theta \leq 1) \end{aligned}$$

分子だけ見ると θ 依存性はわかるが, 分母まで正確に求めよう.

$$f_Y(y_1, \dots, y_N) = \int_0^1 p(y_1, \dots, y_N, \theta) d\theta = \int_0^1 \theta^{\sum_i y_i} (1 - \theta)^{N - \sum_i y_i} \cdot 1 d\theta.$$

ベータ関数とガンマ関数

定義 (ベータ関数とガンマ関数)

$$B(a, b) = \int_0^1 \theta^{a-1} (1 - \theta)^{b-1} d\theta \quad (\Re a > 0, \Re b > 0)$$

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \quad (\Re z > 0)$$

命題 (ベータ関数とガンマ関数の関係)

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

$$\Gamma(z+1) = z\Gamma(z),$$

$$\Gamma(n+1) = n!, \quad (n = 0, 1, 2, \dots),$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

ベータ分布 (Beta distribution)

定義 (ベータ分布 Hoff ベイズ p.39)

連続型確率変数 θ の確率密度関数が (パラメタ $a, b > 0$)

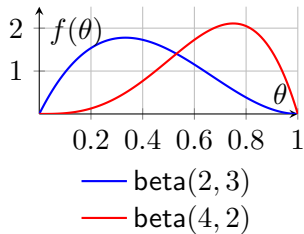
$$f(\theta) = \begin{cases} \frac{1}{B(a,b)} \theta^{a-1} (1-\theta)^{b-1} & (0 \leq \theta \leq 1) \\ 0 & (\text{他}) \end{cases}$$

のとき、 θ はパラメタ a, b のベータ分布 $\text{beta}(a, b)$ にしたがうという。

事前 (一様) 確率密度 $f_{\theta}(\theta)$ は、ベータ分布 $\text{beta}(1, 1)$.

事後確率密度は、ベータ分布 $\text{beta}(\sum_i y_i + 1, N - \sum_i y_i + 1)$.

$$f_{\theta|Y}(\theta|y_1, \dots, y_N) = \frac{\theta^{\sum_i y_i} (1-\theta)^{N-\sum_i y_i}}{\Gamma(\sum_i y_i + 1) \Gamma(N - \sum_i y_i + 1) / \Gamma(N + 2)}$$



```
1  scipy.stats.beta(a,b) # ベータ分布
2
3  scipy.special.gamma(a) # ガンマ関数
4
5  # ベータ関数は、ガンマ関数さえあれば計算できるでしょ
```

L03-Q1

Quiz(ベータ分布)

$\Theta \sim \text{beta}(a, b)$ のとき、すなわち、確率密度関数が

$$f(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1} \quad (0 \leq \theta \leq 1)$$

のとき、次を示そう。

- ① $E[\Theta] = \frac{a}{a+b}.$
- ② $V[\Theta] = \frac{ab}{(a+b)^2(a+b+1)}.$
- ③ Θ の最頻値は $\frac{a-1}{(a-1)+(b-1)}.$

情報の統合

Hoff ベイズ p.44

事前分布 $\Theta \sim \text{beta}(1, 1)$.

‘標本サイズ’= 2. 事前期待値 $E[\Theta] = \frac{1}{1+1}$

‘標本分布’ N 個中 $\sum_i y_i$ が表 ($y = 1$).

‘標本サイズ’= N , ‘標本期待値’= $\frac{\sum_i y_i}{N}$.

事後分布 $\Theta \sim \text{beta}(\sum_i y_i + 1, N - \sum_i y_i + 1)$.

事後期待値 $E[\Theta | Y_i = y_i] = \frac{\sum_i y_i + 1}{N + 2} = \frac{2}{N + 2} \frac{1}{2} + \frac{N}{N + 2} \frac{\sum_i y_i}{N}$ は両者の重み付き平均.

標本サイズが小さければ, 事前期待値の影響が残る. 標本サイズが大きければ, 標本期待値に近づく.

分散も同様に解釈できる.

ここまで来たよ

2 離散-連続混合型分布・条件付き分布・事前分布・事後分布

3 二項モデルとベイズ推定

- ベイズ推定
- 二項モデル (複数のベルヌイ分布)
- 二項モデル (二項分布)

二項モデル (二項分布)

Hoff ベイズ §3.1.1

値が不明なパラメタ θ の確率で表 $Y = 1$ が出るコインを N 回投げて、
表がでた枚数を Y とする。

$Y \sim B(N, \theta)$ ($n = N, p = \theta$ の二項分布)。

θ は不明ということで、分布をもつ確率変数 Θ と思って、前回の Quiz と

- $Y \leftrightarrow$ 柿の質量 X
- $\Theta \leftrightarrow$ 柿の品種 AB Y

と対応させ、 θ の確率分布を y の情報で更新していきたい。
条件付き確率密度は (given | を使って)、二項分布

$$f_{Y|\theta}(y|\theta) = \frac{N!}{y!(N-y)!} \theta^y (1-\theta)^{N-y}.$$

ベイズの定理より, $Y = y$ という結果を得たときの θ の事後確率密度は,

$$\begin{aligned} f_{\theta|Y}(\theta|y) &= \frac{f(y, \theta)}{f_Y(y)} \\ &= \frac{f_{Y|\theta}(y|\theta) f_{\theta}(\theta)}{f_Y(y_1, \dots, y_N)} \\ &= \frac{\frac{N!}{y!(N-y)!} \theta^y (1-\theta)^{N-y} \cdot 1}{f_Y(y)}. \end{aligned}$$

分子だけ見ると θ 依存性はわかるが, 定数まで求めたい. また積分する? しかし, $f_{\theta|Y}(\theta|y)$ は, θ に関する比例定数を除けば, ベータ分布 $\text{beta}(y+1, N-y+1)$ と同じ. どちらも確率密度関数なのだから, 全区間で積分すると 1 になる. よって比例定数は 1.

$$f_{\theta|Y}(\theta|y) \sim \text{beta}(y+1, N-y+1).$$

L03-Q2

Quiz(確率分布の規格化定数)

確率変数 X の確率密度関数が

$$f(x) = Ce^{-x^2-2x+3}$$

で与えられる．定数 C を求めよう．

共役性

定義 (共役性 Hoff ベイズ p.43, 定義 4)

θ の分布のクラス (例: ベータ分布) が確率モデル $f_{y|\theta}(y, \theta)$ に関して共役であるとは、事前分布がそのクラスに属するなら事後分布もそのクラスに属する、ことをいう。

先週扱った柿では、「ベルヌイ分布のクラスが、混合ガウス確率モデルに関して共役」だった。

十分統計量

前半の、多数の独立同分布のベルヌイ分布 Y_i の扱いで、 Y_i は、 $\sum_i Y_i$ という塊でしか現れない.

Hoff ペイズ p.40

統計量 $\sum_i Y_i$ は、データから θ を推測しようとするときの最大の情報を持っている (Y_i の 0,1 の順序には意味がない). このような状況を「 $\sum_i Y_i$ は、 θ と $p(y_1, \dots, y_n | \theta)$ に対する**十分統計量 (sufficient statistic)**である」という.