『社会科学のためのベイズ統計モデリング』第３章

Hiroshi Hamada
Tohoku University

May, 2020
at Tohoku University

3.1 同時分布と条件付き確率

同時確率分布：

2 つの確率変数の組 (X, Y) 確率密度関数 $p(x, y)$

周辺確率分布（marginal probability distribution）：

$$p(x) = \int p(x, y) dy, \qquad p(y) = \int p(x, y) dx$$

同時確率分布から片方の変数を積分消去して得た確率分布（確率密度関数）

X が与えられたときの Y の条件付き確率分布

$$p(y|x) = \frac{p(x, y)}{p(x)}$$

ベイズ推測にはパラメータ θ が与えられたときの X の条件付き確率分布 $p(x|\theta)$ と θ の確率分布 $\varphi(\theta)$ が必要

$$p(x|\theta) : 確率モデル, \qquad \varphi(\theta) : 事前分布$$

3.2 事後分布

定義 (事後分布)

パラメータ θ の事後分布（posterior distribution）を

$$p(\theta|x^n) = \frac{1}{Z_n}\prod_{i=1}^{n} p(x_i|\theta)\varphi(\theta), Z_n = \int_S \prod_{i=1}^{n} p(x_i|\theta)\varphi(\theta)\, d\theta$$

と定義する．積分の範囲 S はパラメータ θ のとりうる範囲（$\theta \in S$）．Z_n を周辺尤度 (marginal likelihood) または分配関数（partition function）という．

直感的意味：データを条件とした場合のパラメータの条件付き分布

周辺尤度 Z_n は事後分布の正規化定数であり，かつ，サンプル（確率変数）

$$X^n = (X_1, X_2, \ldots, X_n)$$

の確率密度関数．Z_n を $(x_1, x_2, \ldots, x_n)$ で積分すると

$$\begin{aligned}
&\iint \cdots \int Z_n dx_1 dx_2 \cdots dx_n \\
&= \int_S \varphi(\theta) d\theta \prod_{i=1}^{n} \left(\int p(x_i|\theta) dx_i \right) = \int_S \varphi(\theta) d\theta \cdot 1 = 1
\end{aligned}$$

3.3 予測分布

定義 (予測分布)

事後分布により確率モデルを平均した

$$p^*(x) = \mathbb{E}_{p(\theta|X^n)}[p(x|\theta)] = \int p(x|\theta)p(\theta|X^n)d\theta$$

を予測分布（predictive distribution）という

ベイズ推測とは

「真の分布 $q(x)$ は，おおよそ予測分布 $p^*(x)$ であろう」

と推測することをいう．

3.4 ベイズ推測の具体例

データを生成する未知の分布を仮のパラメータ q のベルヌーイ分布（確率モデルとして）で推測する.

パラメータ q のもとで，データ $y^n = (y_1, y_2, \ldots, y_n)$ を観測する n 人分の同時確率

$$p(y^n|q) = \prod_{i=1}^{n} q^{y_i}(1-q)^{1-y_i}$$

を確率モデルとおく

事前分布

q の事前分布として実現値が $[0,1]$ の区間に収まる分布 $\mathsf{Beta}(a,b)$ を仮定

$$q \sim \mathsf{Beta}(a,b).$$

ベータ分布 $\mathsf{Beta}(a,b)$ の確率密度関数 $\varphi(q)$ は

$$\varphi(q) = \frac{1}{\mathsf{B}(a,b)} q^{a-1}(1-q)^{b-1}$$

$\mathsf{B}(a,b)$ はベータ関数と呼ばれる定数

$$\mathsf{B}(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx.$$

事後分布の計算

$$
\begin{aligned}
&\frac{1}{Z_n}\left(\prod_{i=1}^{n} q^{y_i}(1-q)^{1-y_i}\right)\varphi(q) \\
&= \frac{1}{Z_n}\left(\prod_{i=1}^{n} q^{y_i}(1-q)^{1-y_i}\right)\underbrace{\frac{q^{a-1}(1-q)^{b-1}}{\mathrm{B}(a,b)}}_{\text{ベータ分布の確率密度関数}} \\
&= \frac{1}{Z_n}\left(q^{\sum y_i}(1-q)^{n-\sum y_i}\right)\frac{q^{a-1}(1-q)^{b-1}}{\mathrm{B}(a,b)} \\
&= \frac{1}{Z_n\cdot\mathrm{B}(a,b)} q^{\sum y_i+a-1}(1-q)^{n-\sum y_i+b-1}
\end{aligned}
$$

- 事後分布 $p(q|y^n)$ は，ベータ分布

$$\frac{q^{\sum y_i+a-1}(1-q)^{n-\sum y_i+b-1}}{\mathrm{B}(a+\sum y_i, b+n-\sum y_i)}$$

  パラメータは $a+\sum y_i, b+n-\sum y_i$

- 事前分布 $\varphi(\theta)$ と事後分布 $p(\theta|x^n)$ が同じ種類の確率分布になるよう設定した事前分布を共役事前分布という.
- 共役事前分布を設定した場合，事後分布は解析的に計算できる（できない場合→ MCMC）

事後分布の具体例

Table: 10 人分の保険加入記録データ（再掲）

個人	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
行動	0	1	0	0	0	0	1	0	0	0

これを使って最尤法（2 章）ではなく，ベイズ推測する (3 章).

事後分布

q の事後分布はサンプルの実現値 y^n のもとで

$$
\begin{aligned}
&\mathrm{Beta}(a+\sum y_i, b+n-\sum y_i) \\
&= \mathrm{Beta}(1+\sum y_i, 1+n-\sum y_i) \\
&= \mathrm{Beta}(1+2, 1+10-2) \\
&= \mathrm{Beta}(3, 9)
\end{aligned}
$$

q の事後確率分布であるベータ分布 Beta$(3, 9)$ をグラフで確認

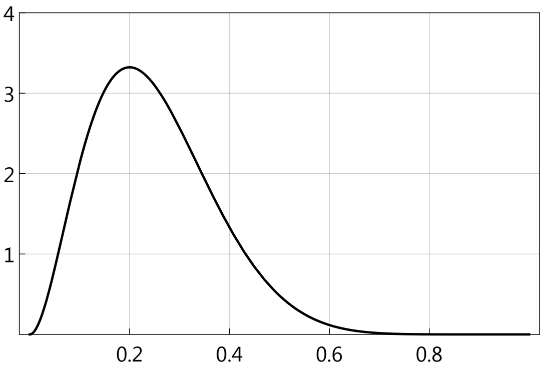

Figure: Beta$(3, 9)$ の確率密度関数（q の事後分布）．$q = 0.2$ は最尤推定値

直感的イメージ

- 最尤法：q の推定値は 1 点（$q = 0.2$）.
  →予測分布は確率モデルに推定値を代入して作る
- ベイズ推測：q の事後分布は $\mathrm{Beta}(3, 9)$.
  →予測分布は確率モデルを事後分布で平均して作る

予測分布

行動 $y \in \{0, 1\}$ の予測分布は

$$p^*(y) = \int_0^1 \underbrace{p(y|q)}_{\text{確率モデル}} \underbrace{p(q|y^n)}_{\text{事後分布}} dq$$

予測分布は、事後分布による期待値

$Y = 1$ となる予測確率 $p^*(y)$ は,

$$\begin{aligned} p^*(y) &= p^*(1) \\ &= \int_0^1 P(Y=1) \cdot p(q|y^n) dq \\ &= \int_0^1 q \cdot p(q|y^n) dq \\ &= \frac{a + \sum y_i}{a + b + n} \end{aligned}$$

a, b は事前分布（ベータ分布）のパラメータ

ベイズ推測の結果得た予測分布は，パラメータ q の事後分布の期待値 $(a+\sum y_i)/(a+b+n)$ をパラメータとするベルヌーイ分布と見なせる．つまり

$$Y_{n+1} \sim \text{Bernoulli}\left(\frac{a+\sum y_i}{a+b+n}\right)$$

データに基づいて具体的にパラメータの値を計算すると

$$\frac{a+\sum y_i}{a+b+n} = \frac{1+2}{1+1+10} = \frac{3}{12} = 0.25$$

より

$$Y_{n+1} \sim \mathrm{Bernoulli}(0.25)$$

3 章まとめ

- ベイズ推測に必要な道具は，サンプルの確率モデル，パラメータの事前分布，パラメータの事後分布である
- パラメータの事後分布は確率モデルと事前分布から計算できる．ただし解析的に解けない場合がある
- 事後分布を解析的に求めることができるような便利な事前分布を共役事前分布という
- 事後分布による確率モデルの平均がベイズ推測の予測分布である．ベイズ推測とは，事後分布を使って計算した予測分布により，真の分布を推測することをいう