大学院システム情報科学府 基礎共通科目 2018 年 12 月 21 日

確率・統計特論 第12回

来嶋 秀治 (Shuji Kijima)

注意: 参照した文献等の情報を必ず記載すること.

今日の話題: ベイズ推定, MAP 推定

(再掲) ベイズの定理

事象 $A_1, \ldots, A_n$ は相互排反とし, 標本空間 Ω に対して $\bigcup_{i=1}^n A_i = \Omega$ が成り立つとする. この時

$$\Pr[A_i \mid B] = \frac{\Pr[B \mid A_i]\cdot \Pr[A_i]}{\sum_{j=1}^n \Pr[B \mid A_j]\cdot \Pr[A_j]} \tag{1}$$

が成り立つ.

ベイズ推定

確率変数 X は密度関数 $f(x \mid \theta)$ に従う[1]とする. ただし θ は (未知の) 母パラメータである. 標本 $X = z$ を得た時に,

$$w'(\theta \mid z) = \frac{w(\theta)\cdot f(z \mid \theta)}{\int_\Theta w(\theta')\cdot f(z \mid \theta')\,\mathrm{d}\theta'} \tag{2}$$

が成り立つ. ただし $w(\theta)$ は θ の従う**事前確率分布** (*prior probability distribution*) の密度関数[2], $w(\theta \mid z)$ は標本 z の下で θ の従う**事後確率分布** (*posterior probability distribution*) の密度関数を表す.

事前分布 $w(\theta)$ と事後分布 $w'(\theta \mid x)$ が同一種類の分布 (分布族) に従うとき, **共役分布** (族) (*conjugate distributions* (family)) という. 共役分布は f に依存することに注意.

共役分布の例

(i) ベータ分布. 確率変数 X は 2 項分布 $\mathrm{B}(n; p)$ に従うとし, 確率関数を $f(x \mid p)$ であらわす. 事前分布 $w(p)$ をベータ分布 $\mathrm{Be}(\alpha, \beta)$ とすると, 標本 $X = z$ に対して,

$$\begin{aligned} w'(p \mid z) &\propto w(p)\cdot L(p \mid z) = w(p)\cdot f(z \mid p) = \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1}\binom{n}{z}p^z(1-p)^{n-z} \\ &= \frac{1}{B(\alpha,\beta)} p^{\alpha-1}(1-p)^{\beta-1}\binom{n}{z}p^z(1-p)^{n-z} \propto p^{\alpha+z-1}(1-p)^{\beta+(n-z)-1} \\ &\propto \frac{1}{B(\alpha+z, \beta+(n-z))} p^{\alpha+z-1}(1-p)^{\beta+(n-z)-1} \end{aligned}$$

となり, 事後分布 $w'(p \mid z)$ はベータ分布 $\mathrm{Be}(\alpha+z, \beta+n-z)$ となる. ただし, $B(\alpha, \beta) := \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t$ は正規化定数.

(ii) ガンマ分布. 確率変数 X はポアソン分布 $\mathrm{Po}(\lambda)$ に従うとし, 確率関数を $f(x \mid \lambda)$ であらわす. 事前分布 $w(\lambda)$ をガンマ分布 $\mathrm{Ga}(\alpha, \nu)$ とすると, 標本 $X = z$ に対して, 事後分布 $w'(\lambda \mid z)$ はガンマ分布となる. ($\Longrightarrow$ 演習 3)

(ii) 正規分布. 確率変数 X は正規分布 $\mathrm{N}(\mu, \sigma^2)$ に従うとし, 確率関数を $f(x \mid \mu)$ であらわす. 事前分布 $w(\mu)$ を正規分布 $\mathrm{N}(m, s^2)$ とすると, 標本 $X = z$ に対して, 事後分布 $w'(\mu \mid z)$ は正規分布となる. ($\Longrightarrow$ 演習 3)

[1] X が離散分布の場合は $f(x \mid \theta)$ は確率関数を表す

[2] θ が離散的な値をとる場合には $w(\theta)$, $w(\theta \mid z)$ を確率関数とし, (2) 式の分母を適切に変更する.

MAP 推定
事後確率を最大化するパラメター，すなわち

$$\arg\max_{\theta} w'(\theta \mid z) \tag{3}$$

を *MAP* **推定** (*maximum a posteriori estimation*) と呼ぶ．**ベイズ推定**とは事後分布を求めることで，その事後確率を最大化するパラメータ（値）が *MAP* **推定**である．

最尤推定と MAP 推定
式 (2) から，尤度関数 $L(\theta \mid z) := f(z \mid \theta)$ を用いて，

$$w'(\theta \mid z) \propto w(\theta) \cdot L(\theta \mid z) \tag{4}$$

が成り立つことがわかる．大雑把に言えば，**最尤推定**は（同時確率ではなく）尤度関数という（人工的な）関数を最大化するのに対し，*MAP* **推定**は（人工的な事前分布を仮定したうえで）事後確率の最大化を行っている．詳細は自身で調べられたし．

演習問題

***演習 1.** 3 つのつぼ A_1, A_2, A_3 があり, それぞれ赤球と白球が $3:1$, $2:1$, $1:1$ の割合で入っている．
(i) つぼを一様ランダムに選び, 球を取り出すと赤であった．選ばれたつぼの事後確率 $\Pr(A_i \mid \text{赤球})$ をそれぞれ求めよ．
(ii) つぼを一様ランダムに選び, 復元抽出で 5 回球を取り出すと 5 回とも赤であった．選ばれたつぼの事後確率 $\Pr(A_i \mid \text{赤球} \times 5)$ をそれぞれ求めよ．
※復元抽出とは, 一度取り出した球を再びつぼに戻してから次の試行を行う方法で, すなわち各試行は (独立) 同一分布に従う．非復元抽出 (つぼに戻さない) の場合, 各試行の従う分布は残存する球の比率で変化するため, (一般には) 独立でも同一でもないことに注意せよ．

***演習 2.** ある疾病 D に罹患すると, 3 種類の症状 S_1, S_2, S_3(例えば, 頭痛, 発熱, 咳) が現れる．ただし S_1, S_2, S_3 は確率的に生じ, 独立とし, 疾病 D 以外の要因でも症状 S_1, S_2, S_3 の発現の可能性はある．疾病 D の有無による, 症状 S_1, S_2, S_3 の発現の確率を以下の表に表す．

	S_1	S_2	S_3	罹患確率
D の場合	0.1	0.7	0.6	0.23
D でない場合	0.8	0.2	0.5	0.77

症状の有無を $(s_1, s_2, s_3) \in \{0,1\}^3$ であらわすとき, すべての場合 (8 通り) についての疾病 D の事後確率 $w'(d \mid (s_1, s_2, s_3))$ を求めよ．
※例えば, $w'(d \mid (1,1,0))$ は, 症状 S_1, S_2 があり, 症状 S_3 がない場合の W に罹患している事後確率を表す．

演習 3.
(i) ポアソン分布に対して, ガンマ分布が共役分布であることを示せ．
(ii) 期待値 μ が未知の正規分布に対して, 正規分布が共役分布であることを示せ．