ベイズ推定

1　ベイズ統計学と非ベイズ統計学

数理統計学はベイズ統計学と非ベイズ統計学に分けられる。前者は,Tomas Bayes(1702〜1761) らによる考え方で、特徴として未知パラメータ θ の分布（これを事前分布とよぶ）および観測値 x によるパラメータの尤度の両者を、「ベイズの定理」として結合させ、パラメータの事後分布を導出する。つまり何らの予めの情報があれば、観測値を得ることでより正確な予測ができると考える。一方非ベイズ統計学は、このような事前の情報などを仮定できないとみなすものである。伝統的統計学、古典的統計学ともよばれる。著名な統計学者 K・ピアソン（1857〜1936）と R・A・フィシャー（1890〜1962）のいずれも非ベイズ統計学の立場をとっている。

2　事後分布と事前分布

まず確率のベイズの定理を述べる。事象 $A_1, A_2, \cdots, A_n$ と B において、条件 1) $A_i \cap A_j = \emptyset (i \neq j), \cup_i A_i = \Omega$ 同時に起こることはなく、互いに素なる事象。条件 2) 確率 $P(A_i)$ と条件付き確率 $P(B|A_i),\ i = 1, 2, \cdots, n$ が与えられている。この条件１)，２）から、条件付き確率 $P(A_i|B)$ が次の式で計算できる；

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_j P(A_j)P(B|A_j)}, \quad i = 1, 2, \cdots, n$$

このベイズの定理では、条件付き確率の計算であるが、未知のパラメータに対する推測では、つぎの連続型密度関数で表現される：θ を未知とし、その事前確率密度を $p(\theta)$ とするとき、観測値 $x = (x_1, x_2, \cdots, x_n)$ に対する θ の事後分布は

$$p(\theta|x) = \frac{f(x|\theta)p(\theta)}{\int f(x|\theta')p(\theta')d\theta'} \quad \text{または} \quad p(\theta|x) \propto f(x|\theta)p(\theta)$$

上式の前半で、分母に当たる部分は、確率が 1 になるように調整するだけであるから、これを省略して、後半のように比例記号 ($\propto$) を用いて表すこともある。

問１（本文資料 23 ページ）偏りのあるコイン投げを n 回繰り返す。$x_i = 1, 0$ を表か裏とする。確率に関しては全く情報をもたないとして、$p(\theta) = 1, 0 \leq \theta \leq 1.$ このとき $t = \sum_{i=1}^n x_i$ とするとき、θ の事後分布を求めよ。

3　ベイズ推定

パラメータ推定におけるベイズ統計学による接近法は、損失関数 (loss function)$L(\theta, a)$ をもちいる。未知パラメーラ θ に値 a で推定するとき、正確さの基準尺度として、この損失関数を小さくしようと考える。もし $\hat{\theta}$ を $\mathbb{E}[L(\theta, \hat{\theta})]$ (ここで期待値 $\mathbb{E}$ は事後分布 $p(\theta|x)$ に関してとる）が最小となるならば、ベイズ危険 (Bayes risk) を最小とするベイズ推定とよぶ。一般に事後分布の計算は簡単ではない。

問２（本文資料 24 ページ）未知の真値 θ を値 a で推定するとき、損失関数 $L(\theta, a)$ が次で与えられるとき、ベイズ危険を求めよ。(1) 2 乗誤差損失 $L(\theta, a) = (a - \theta)^2$ (2) 絶対誤差損失 $L(\theta, a) = |a - \theta|$

問３（本文資料 24 ページ）$X_1, \cdots, X_n$ をパラメータ λ のポアソン分布、λ をパラメータ 1 の指数分布とする。(1) λ の事後分布は $(\sum_i X_i + 1, n + 1)$ のガンマ分布になることを示せ。 (2) 2 乗誤差損失では $\hat{\theta} = (\sum_i x_i + 1)/(n + 1)$ で与えられることを示せ。 (3) 絶対誤差損失ではどうなるか。