ベイズ統計による数理モデルの ヴァリデーションとモデル選択

総合研究大学院大学　複合科学研究科　統計科学専攻
柏屋　滋

1　モデルパラメータの推測統計

ベイズ統計学（Bayesian statistics）では、パラメータθの事前確率密度関数（prior probability density function）$\pi\left(\theta\right)$に対して、事後確率密度関数（posterior probability density function）$\pi\left(\theta \mid y\right)$を、ベイズの公式（Bayes' theorem）から計算する。

$$\pi\left(\theta \mid y\right) \propto f\left(y \mid \theta\right)\pi\left(\theta\right) \tag{1}$$

任意の実数aに対して、関数$I\left(x;a\right)$を定義関数と呼ぶ。

$$I\left(x;a\right)=1 \;\; if \;\; x \geq a \quad \left[=0 \;\; otherwise\right] \tag{2}$$

n個のデータy_iが与えられたとき、経験分布関数を定義する。

$$\hat{F}\left(y \mid \theta\right)=\frac{1}{n}\sum_{i=1}^{n} I\left(y;y_i\right) \tag{3}$$

対応する確率密度関数は、分布関数の偏微分で与えられる。

$$\hat{L}\left(\theta \mid y\right) \equiv \hat{f}\left(y \mid \theta\right)=\frac{\partial \hat{F}\left(y \mid \theta\right)}{\partial \theta} \tag{4}$$

多数の粒子（particle）「粒子」（particle）δ_aで$I\left(x;a\right)$を代表し、パラメータθの確率分布を表現する。また、事前分布$\pi\left(\theta \mid y\right)$として、超一様分布列擬似乱数を適用する。

2　ベイズ統計によるパラメータ推定

非線形（nonlinear）かつ動的（dynamic）なプロセスモデルの例として、酵素又は触媒により、尿素（NH_2CONH_2）が二酸化炭素（CO_2）とアンモニア（NH_3）に分解される化学反応を考える。

$$NH_2CONH_2 + H_2O \;\rightarrow\; CO_2 + 2NH_3 \tag{5}$$

実験データに対し、二酸化炭素濃度$C\left(t\right)$の時間変化の式を仮定する。（単位：モル%）。

$$C\left(t\right)=100\times\left[1-\exp\left(-\theta_1 t\right)\left(1+\theta_1 t\right)\right] \tag{6}$$

$$C\left(t\right)=100\times[1-\theta_2\exp\left(-\theta_1 t\right)/\left(\theta_2-\theta_1\right)+\theta_1\exp\left(-\theta_2 t\right)/\left(\theta_2-\theta_1\right)] \tag{7}$$

$$C\left(t\right)=100\times[1-\theta_2\exp\left(-\theta_1\left(t+\theta_3\right)\right)/\left(\theta_2-\theta_1\right)+\theta_1\exp\left(-\theta_2\left(t+\theta_3\right)\right)/\left(\theta_2-\theta_1\right)] \tag{8}$$

t	2	4	6	8	10	12	14	16	18	20
C	5.68	22.4	35.6	47.0	57.6	65.9	72.0	78.6	83.4	86.1

表1　Gale and Eadie による尿素加水分解の実験データ

前述のポピュレーション・モンテカルロ的手法(population Monte Carlo)により、各モデルのパラメータを推定することができる。尚、ベイズ統計では、パラメータの事後平均の計算が可能だが、MAP(maximum aposteori)推定とすれば、最尤推定と一致する。

推定方法	最尤推定			ベイズ推定:事後平均		
パラメータ	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
Model 1(式 9)	0.1910	—	—	0.1903	—	—
Model 2(式 10)	0.1098	0.5283	—	0.1118	0.4892	—
Model 3(式 11)	0.1185	0.3355	0.6948	0.1087	0.7058	−0.392

表2 Baileyらによる提案モデルのパラメータ推定

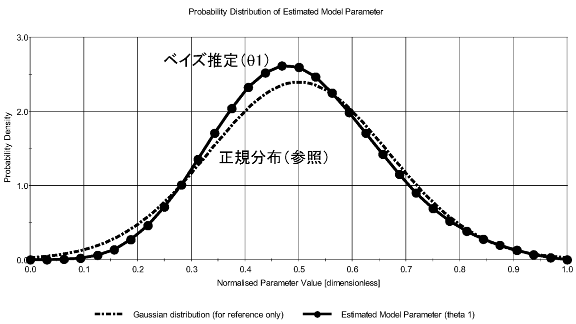

図2 パラメータθ_1の事後確率分布

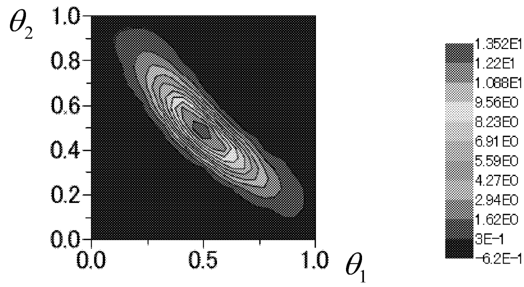

図3 パラメータθ_1, θ_2の事後確率同時分布

3 情報量基準によるモデル選択

3- 1 AIC (Akaike Information Criterion:赤池情報量基準)

$$AIC_k = -2\sum_{i=1}^{n} \log f\left(y_i \mid \hat{\theta}_k\right) + 2p_k \tag{9}$$

3- 2 BIC (Bayesian Information Criterion:ベイズ情報量基準)

$$BIC_k = -2\sum_{i=1}^{n} \log f\left(y_i \mid \hat{\theta}_k\right) + p_k \log n \tag{10}$$

3- 3 DIC (Deviance Information Criterion:偏差情報量基準)

$$DIC \equiv \bar{D}(\theta) + p_D = D\left(\bar{\theta}\right) + 2p_D \tag{11}$$

$$D(\theta) = -2\log f\left(y \mid \theta\right) + 2\log h\left(y\right) \tag{12}$$

$$p_D \equiv \bar{D}(\theta) - D\left(\bar{\theta}\right) \tag{13}$$

4- 4 モデル選択の例 (model discrimination: example)

より小さい情報量基準値ほどモデルの適合度が高いと見做す。2 項のプロセスモデルの場合、AIC, BIC, DIC 何れの情報量基準に照らしても、Model 2 が合理的モデルとして選択される。

情報量基準	AIC	BIC	DIC
Model 1 (式 9)	72.7	70.7	+0.033
Model 2 (式 10)	12.3	9.6	−0.041
Model 3 (式 11)	13.3	10.6	+0.206

表3 情報量基準によるモデルの評価