

のように表わされる。ここで

$$\begin{aligned}\bar{p} &= ap + (1-a)\bar{p} \\ \bar{q} &= aq + (1-a)\bar{q} \\ \bar{d} &= ad + (1-a)\bar{d} \\ e &= a(1-a)(p-\bar{p})(q-\bar{q})\end{aligned}$$

例えば、大竜巻のような天災が起きるか否かの予想と、実際に起きたか否かにより結果を分類した表に、完璧な予測に対応する対角モデル $D^+(p)$ と、当てずっぽな予測に対応する独立モデル $I(p, q)$ との混合モデルを対峙させると、 a はその予想のパターンの“的中率”を示す。

また、ある期間に結婚した夫婦の前歴で、夫が初婚か再婚か、妻が初婚か再婚かで分類した表には、三角モデル $U^+(p, q)$ と独立モデル $I(p, q)$ 又は三角モデル $S\left(\frac{p+q}{2}, p - \left(\frac{p+q}{2}\right)^2\right)$ との混合モデルを対応させると、 a は初婚の夫が再婚の妻との結婚を忌避する尺度を示すものと考えられる。

この他、各種の混合モデルを対峙させて、分割表に潜在している「構造」に探りを入れることができる。

ベイズ型密度関数推定モデルについて

坂元 慶行

滑らかな密度関数をもつ限り適用可能な密度関数の推定法を [1] で提案した。この方法は、赤池の GALTHY の考え方に基礎を置き、具体的には次の手順によっている。

- ① データ x を、適当な分布関数 $Q(x)$ を用いて、 $y=Q(x)$ に変換
- ② 区間 $[0, 1]$ 上のヒストグラムのベイズ推定値 $\hat{h}(y)$ を求める
- ③ $\hat{h}(y)$ を 2 次スプラインで平滑化し、 $\tilde{h}(y)$ を求める
- ④ x の密度関数の推定値として $f(x)=\tilde{h}(Q(x))q(x)$ をとる

この方法においては、②で、尤度 $L(\mathbf{h})$ の \mathbf{h} に対して事前分布 $\pi(\mathbf{h} | u^2, h_{-1}, h_0)$ を仮定し、ハイパー・パラメータを、

$$ABIC = (-2) \log \int L(\mathbf{h}) \pi(\mathbf{h} | u^2, h_{-1}, h_0) d\mathbf{h} + 2 \quad (\text{ハイパー・パラメータ数})$$

によって決定する。この方法は、従来順位統計量によって処理されることの多かった“(2 標本)ノンパラメトリック検定”に対する新たな情報量統計学的アプローチの可能性をひらくと考えられる。すなわち、

$$g(y | u^2, h_{-1}, h_0) = \int L(\mathbf{h}) \pi(\mathbf{h} | u^2, h_{-1}, h_0) d\mathbf{h}$$

が確率分布をなす場合、分布型未知の 2 つの母集団分布 g_1, g_2 から m, n 個の標本 $\{y_i^{(1)}\}, \{y_j^{(2)}\}$ がとられたとき、2 つのモデル、

$$\begin{aligned}\text{MODEL (0): } y_i^{(j)} &\sim g(y | u^2, h_{-1}, h_0) & j=1, 2 \\ \text{MODEL (1): } y_i^{(j)} &\sim g(y | u_{(j)}^2, h_{(j)}^2, h_0^{(j)}), & j=1, 2\end{aligned}$$

を想定し、ABIC の最小値を与えるモデルを採択することによって実現できると考えられる。

いくつかの実験結果から、正答率には遜色ないことがわかった。しかし、この方法の特長は分布型そのものの良好な推定が可能にある。

[1] Ishiguro, M. and Sakamoto, Y. (1984). A Bayesian Approach to the Probability Density Estimation, *Ann. Inst. Statist. Math.*, 36, 3, 523-538.

[2] 坂元慶行 (1985). カテゴリカルデータのモデル分析, 共立出版。(印刷中)