

1. 一般モデル

統計モデル $\{f(x, \theta), \theta \in \Theta\}$ に従うデータ $\{x_1, \dots, x_n\}$ が得られたとき、Fisher の最大尤度法は次のように与えられる．対数密度関数をデータ上で平均を取ることによって対数尤度関数

$$L(\theta) = \frac{1}{n} \sum_{i=1}^n \ell(x_i, \theta)$$

を計算する．ここで $\ell(x, \theta) = \log f(x, \theta)$. この $L(\theta)$ を Θ 上で最大にする θ を最尤推定量と呼ぶ．この発表では、対数尤度関数を拡張して Ψ 尤度関数を

$$L_\Psi(\theta) = \frac{1}{n} \sum_{i=1}^n \Psi(\ell(x_i, \theta)) - b_\Psi(\theta)$$

と提案した．ここで $\Psi(\ell)$ は ℓ の単調増加関数を仮定する．典型的な例題としては、

(i) β べき尤度 ($\beta > 0$)

$$L_\beta(\theta) = \frac{1}{n} \sum_{i=1}^n \frac{f(x_i, \theta)^\beta - 1}{\beta} - b_\beta(\theta)$$

(ii) η シグモイド尤度 ($\eta > 0$)

$$L_\eta(\theta) = \frac{1}{n} \sum_{i=1}^n \log\{f(x_i, \theta) - \eta\}$$

が与えられる．あとの議論で具体的な統計モデルに適用される．一般に、 Ψ 尤度は次の Ψ ダイバージェンスと密接な関係にある．2つの正值測度が密接関数 $g(x)$ と $f(x)$ を持っているとする．この時 g から f への Ψ ダイバージェンスは、

$$D_\Psi(g, f) = \int g\{\Psi(\log g) - \Psi(\log f)\}d\nu - \int \{\Psi^*(\log g) - \Psi^*(\log f)\}d\nu$$

で定義される．ここで、 $\Psi^*(\ell) = \int^\ell \Psi'(s)e^s ds$. データ従うの分布の密度関数を $g(x)$ とする時

$$D_\Psi(g, f) \approx \frac{1}{n} \sum \{\Psi(\log g(x_i)) - \Psi(\ell(x_i, \theta))\} - \int \{\Psi^*(\log g(x)) - \Psi^*(\ell(x, \theta))\}d\nu(x)$$

なのでパラメータ θ に依る項のみに注目すると

$$D_\Psi(g, f) \approx -L_\Psi(\theta)$$

が成り立つ．ここで

$$b_\Psi(\theta) = \int \Psi^*(\ell(x, \theta)).$$

この関係から Ψ 最尤推定量 $\hat{\theta}_\Psi$ の Fisher 一致性が示される．影響関数は

$$\text{IF}(\hat{\theta}_\Psi, y) \propto \psi(\ell(y, \theta))S(y, \theta)$$

となる．ここで $\psi(\ell) = \Psi'(\ell)$, $S(y, \theta) = (\partial/\partial\theta)\ell(y, \theta)$. 通常の最尤推定の場合は 恒等的に $\psi(\ell) = 1$ なのでスコア関数 $S(y, \theta)$ が非有界であれば IF も又非有界となる．これより B-ロバストネスが無いことが議論されている．cf. Hampel (1974), JASA.

一方で $\psi(l(x, \theta))$ が x に依っている場合にスコア $S(x, 0)$ の長さが大きくなる x に対して、 $\psi(l(x, \theta))$ がそれより速く θ に近づけば影響関数の有界性が期待される。 β べき尤度と η シグモイド尤度の ψ はそれぞれ

$$\psi_\beta(\ell) = e^{\beta\ell}, \quad \psi_\eta(\ell) = \frac{e^\ell}{\eta + e^\ell}$$

となり、任意の $p > 0$ に対して

$$\lim_{\ell \rightarrow -\infty} \ell \psi(\ell) = 0$$

を充たす。この事実から、多くの統計モデルに対して有界な影響関数を持つことが示される。

2. 正規モデル

p 変量正規モデル $\{N(\mu, \Sigma) : \mu \in \mathbf{R}^p, \Sigma \in \mathbf{R}^{p \times p}\}$ を考察した。 Ψ 尤度方程式は $\theta = (\mu, \Sigma)$ に対して

$$\frac{1}{n} \sum_{j=1}^n \psi_j(\theta)(x_j - \mu) = 0, \quad \frac{1}{n} \sum_{j=1}^n \psi_j(\theta)(x_j - \mu)(x_j - \mu)^T - \Sigma = c_\Psi(\det(\Sigma))\Sigma$$

となる。この方程式は、一般的に解けないので Ψ -アルゴリズム：

$$\mu_{k+1} = \frac{\sum_{j=1}^n \psi_j(\theta_k) x_j}{\sum_{j=1}^n \psi_j(\theta_k)}, \quad \Sigma_{k+1} = \frac{\sum_{j=1}^n \psi_j(\theta_k) (x_j - \mu_k)(x_j - \mu_k)^T}{\sum_{j=1}^n \{\psi_j(\theta_k) - c_\Psi(\det(\Sigma_k))\}}$$

が提案された。ある条件下で、このアルゴリズムは Ψ -尤度関数を一様に改善することが証明できる。以下に簡単なシミュレーションを実行した。ここでは、データ分布は正規モデルから少しずれた ϵ -混入モデル ($\epsilon = 0.05$) の 2 つのケース：

$$G_{1\epsilon} = (1 - \epsilon)N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & .5 \\ .5 & 2 \end{pmatrix}\right) + \epsilon N\left(\begin{pmatrix} 5 \\ -5 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}\right)$$

$$G_{2\epsilon} = (1 - \epsilon)N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & .5 \\ .5 & 2 \end{pmatrix}\right) + \epsilon N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix} \begin{pmatrix} 9 & 9 \\ 9 & 9 \end{pmatrix}\right)$$

に設定した。このときの KL ロスによる比較は 100 回の反復の結果、分散行列 Σ の推定に着目するプロットを発表した、cf. [3].

この基礎的考察を元にして PCA, ICA, Gaussian mixture について Ψ 尤度法を適用した、cf. [1], [2], [4], [5]. また、判別分析の問題に提案されている D_w 法とも密接な関連がある。cf. [6].

参考文献

- [1] Higuchi, I. and Eguchi, S. (1998). The influence function of principal component analysis by self-organizing rule. *Neural Computation*, 10, 1435-1444
- [2] Kamiya, H. and Eguchi, S. (2001). A class of robust principal component vectors. *J. Mult. Anal.*, 77, 239-269.
- [3] Eguchi, S. and Y. Kano. (2001). Robustifying maximum likelihood estimation by psi-divergence. *ISM Research Memo* 802.
- [4] Minami, M. and S. Eguchi. (2001). Robust blind source separation by beta-Divergence. *ISM Research Memo*. 799.
- [5] Fujisawa, H. and S. Eguchi. (2001). Robust estimation in normal mixture model. (to be submitted).
- [6] Eguchi, S. and J. Copas. (2002). A class of logistic type discriminations. In press *Biometrika* 89.