パラメトリックな確率分布間の Divergence の漸近展開

信州大学学術研究院社会科学系　椎名　洋

二つのパラメトリックな分布（密度で表現）$f(x;\theta_1)$ と $f(x;\theta_2)$ の距離を次の α-divergence $(-\infty < \alpha < \infty)$ を使って測る。

$$\overset{\alpha}{D}[\theta_1 : \theta_2] = \frac{4}{1-\alpha^2}\Big\{1 - \int_{\mathfrak{X}} f^{(1-\alpha)/2}(x;\theta_1) f^{(1+\alpha)/2}(x;\theta_2) d\mu\Big\} \tag{1}$$

基準となる分布の密度を $f(x;\theta_0)$ とした時、この分布の近傍にある $f(x;\theta)$ の振る舞いを見るために、$\overset{\alpha}{D}(\theta) = \overset{\alpha}{D}[\theta : \theta_0]$ を θ_0 の周りでテイラー展開する。

$$\begin{aligned}
&\overset{\alpha}{D}(\theta) \\
&= \sum_{i=1}^{p} \epsilon_i \overset{\alpha}{D}[\theta_0 : \theta_0](\theta^i - \theta_0^i) + \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p} \epsilon_i\epsilon_j \overset{\alpha}{D}[\theta_0 : \theta_0](\theta^i - \theta_0^i)(\theta^j - \theta_0^j) \\
&+ \frac{1}{6}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p} \epsilon_i\epsilon_j\epsilon_k \overset{\alpha}{D}[\theta_0 : \theta_0](\theta^i - \theta_0^i)(\theta^j - \theta_0^j)(\theta^k - \theta_0^k) \\
&+ \frac{1}{24}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p} \epsilon_i\epsilon_j\epsilon_k\epsilon_l \overset{\alpha}{D}[\theta_0 : \theta_0](\theta^i - \theta_0^i)(\theta^j - \theta_0^j)(\theta^k - \theta_0^k)(\theta^l - \theta_0^l) \\
&+ O(||\theta - \theta_0||^5),
\end{aligned} \tag{2}$$

但し、ϵ_i は $\overset{\alpha}{D}[\theta_1 : \theta_2]$ を θ_1^i に関して偏微分することを意味する。二次微分項、三次微分項がそれぞれ、フィッシャー計量、α-接続という幾何的な量で表現される（Eguchi's Relationship)。今回の発表では、四次微分の項の幾何的な表現（α-接続の微分と第二基本形式を使った表現）を扱う。具体的な応用として、次の三つの場合を考える。

1. 二つの多変量確率変数 X と Y の同時分布において、$Y = y^0$ の条件つきの X の分布をターゲット分布（上の $f(x;\theta_0)$ に該当）として考えて、条件 $y_i^0 \le Y_i \le y_i^0 + t_i$ 下での X の分布が、$t_i \to 0$ と共にどのようにターゲット分布に近づいてくかを考察する。

2. $lim_{u\to 0}(1+ux)^{1/u} = e^x$ の関係を使うことで指数型分布を拡張することを考える。指数型本来の分布のパラメータに u が加わることになるが、この場合の $u \to 0$ の分布の収束の速さを考える。

3. 大標本の漸近論。M.L.E $\hat{\theta}$ に基づく推定分布 $f(x;\hat{\theta})$ と真の分布 $f(x;\theta)$ の間の Divergence の平均 $E[\overset{\alpha}{D}[\hat{\theta} : \theta]]$ を上記の展開をもちいて、n^{-2}（n は標本数）のオーダーまで求める。これによって、当該パラメトリックモデルの推測の難しさが表現される。