自発的なデータ学習について

統計数理研究所・数理推論研究系　　江口 真透

1.　はじめに．真の分布がモデルに属すときの統計的推定理論は深く考察され高い完成の域に達している．この発表では**ノンパラメトリック**な設定において，この理論が崩壊する挙動について着目する．真の分布は多峰まで想定するが統計モデルは標準的な単峰モデルを考える．一方で，推定はできるだけ多様な推定量を考えよう．ある特別な推定方法を選択するとモデルから乖離した真の分布の情報を取り出すことができる (**推定量選択**)．これを自発的なデータ学習と呼ぶ．

例えば，真の分布を多変量正規混合分布で多峰であるとし，ワーキングモデルとして正規平均モデルを採用し，最小ベキダイバージェンスで推定を考える．ただしベキパラメータは自由に動かす．ベキダイバージェンスの正規平均に対する経験ロス関数は**複数個の局所最小値**を持つことが分かる．一方で負の尤度関数は凸関数となり最尤推定量は標本平均ベクトルに他ならない．ベキパラメータを適切に選択すると，その局所最小解の集合が正規混合分布の中の正規分布の平均の集合に一致することが後で示せる．このようにして定めるクラスタリングは自発的にクラスター数が局所最小解の数として推定される [1]．さらにノンパラメトリックな密度推定ができることが分かる．従来のパラメトリックモデルの誤特定の議論とは全く違う観点から自発データ学習の新たな可能性について考察したい．

2. ベキロス関数．　標準的なモデル $M = \{f_\theta(x) : \theta \in \Theta\}$ を考える．このとき真の密度関数 $g(x)$ に対して M 上のベキダイバージェンスによる期待ロス関数は次のように定まる：

$$L_\gamma(\theta) = c_\gamma(\theta) \int f_\theta(x)^\gamma g(x) dx$$

ここで $c_\gamma(\theta)$ は正規化定数 $-(1/\gamma)\big(\int f_\theta(x)^{\gamma+1}dx\big)^{-\gamma/(\gamma+1)}$ とする．経験ロス関数は $L_\gamma(\theta)$ において g に関する期待値の代わりにデータによる標本期待値に置き換えることに得られる．真の密度がモデルに属すとき，すなわち $g = f_{\theta_0}$ のとき，$L_\gamma(\theta)$ は $\theta = \theta_0$ で最小値を持つ（フィッシャー一致性）．

モデル M を正規平均モデル $f_\theta \sim N(\theta, I_d)$ とし，真の密度関数が正規混合 $g(x) = \sum_{j=1}^J p_j f_{\theta_j}(x)$ とするとき，期待ロス関数は，次のように表される：

$$L_\gamma(\theta) \propto -\textstyle\sum_{j=1}^J p_j f_{\theta_j}(\theta)^{\frac{\gamma}{\gamma+1}}.$$

これより，γ が十分大きいときは期待ロス関数は $-g(\theta)$ に漸近する．$L_\gamma(\theta)$ の局所最小は真の密度関数 $g(x)$ のモードで起こることを意味する。このことは次の考察で拡張できる．

定理 1．$\varphi(x)$ を x が原点のみにモードを持つ d 変量の密度関数とし，ロケーションモデル $M = \{f_\theta(x) := \varphi(x-\theta) : \theta \in \mathbb{R}^d\}$ を考える．このとき，任意の真の密度関数 $g(x)$ に対して

$$\lim_{\gamma\to\infty} b_\gamma L_\gamma(\theta) = -g(\theta).$$

が成立する．ここで $b_\gamma = \varphi(0)^{-\gamma}\big(\int \varphi(x)^{\gamma+1}dx\big)^{\gamma/(\gamma+1)}$ とする．

証明は極限を取ると非積分関数が θ で退化するディラック関数 $\delta(x-\theta)$ と $-g(x)$ の積に表されることから示される．このことよりモデル M と無関係に真の密度 g が γ-漸近的に得られることが分かる．真の分布がモデルに含まれるなら任意の γ でベキロス関数の最小化は一致推定を保証し，$\gamma = 0$ (最尤推定) で漸近有効になる．一方で，真の分布がモデルから乖離していると γ を適切に選択すると自発的にデータを学習する機能が見つかった．この機能について推定量選択の文脈で更なる解明と応用について考えたい．

[1]. A. Notsu, O. Komori and S. Eguchi. *Neural Computation* 26, 2 (2014) 421-448.