回帰モデルの予測関数の情報幾何

統計数理研究所・数理推論研究系　江口 真透
総合研究大学院・統計科学専攻　大前 勝弘

1. はじめに

一般化線形モデル (GLM) の枠組みの中で回帰モデルの情報幾何を考える．従来の情報幾何の枠組みは確率密度関数の空間に構築されているが，回帰モデルでは（条件付き）密度関数だけでなく回帰関数と予測関数を考慮しなければならない．$\boldsymbol{X}$ を p 次元の説明変数ベクトル，Y を応答変数とすると，回帰関数 $\mu(\boldsymbol{x}) = \mathbb{E}(Y|\boldsymbol{X} = \boldsymbol{x})$ に対して，線形予測関数 $f(\boldsymbol{x}) = \boldsymbol{\beta}^\top \boldsymbol{x} + \alpha$ は期待値リンク関数 ℓ によって $\mu(\boldsymbol{x}) = \ell(f(\boldsymbol{x}))$ と一対一で結ばれると仮定する．Y の $\boldsymbol{X}$ を与えたときの条件付き密度関数は

$$p(y|\boldsymbol{x}, f, \delta) = \exp\Big\{\frac{yg(f(\boldsymbol{x})) - \psi(g(f(\boldsymbol{x})))}{\delta} - \gamma(\boldsymbol{x}, \delta)\Big\}, \tag{1}$$

と仮定される．ここで g を標準リンク関数と呼ぶ．このように GLM では正規回帰，ロジスティック回帰，ポアソン回帰，ガンマ回帰，逆ガウス回帰などを統一的に扱い，多くの応用がある (McCullagh & Nelder, 1989)．

2. セミパラメトリック GLM

前節の GLM の枠組みでは線形予測関数を考えていたが，ここでは可能なすべての予測関数の空間 $\mathcal{F}$ を考察しよう．これによって回帰関数の空間 $\mathcal{R} = \ell(\mathcal{F})$ と Y の条件付き密度の空間 $\mathcal{P} = \{p(y|\boldsymbol{x}, f, \delta) : f \in \mathcal{F}\}$ が考えられる．$\mathcal{F}$ の 2 つの予測関数 $f_0(\boldsymbol{x})$ と $f_1(\boldsymbol{x})$ をつなぐ測地線を次のように定義する．

$$f_t^{(\phi)}(\boldsymbol{x}) = \phi^{-1}\big((1-t)\phi(f_0(\boldsymbol{x})) + t\phi(f_1(\boldsymbol{x}))\big), \quad t \in [0,1] \tag{2}$$

ここで ϕ は一対一変換とする．特に ϕ を標準リンク関数 g と選ぶとき，測地線 $f_t^{(g)}(\boldsymbol{x})$ は $\mathcal{P}$ の中へ $p(y|\boldsymbol{x}, f_t^{(g)}, \delta)$ を誘導する．これは 2 つの条件付き密度関数 $p(y|\boldsymbol{x}, f_0, \delta)$ と $p(y|\boldsymbol{x}, f_1, \delta)$ をつなぐ**指数測地線**に他ならない：

$$p(y|\boldsymbol{x}, f_t^{(g)}, \delta) = z(t, \boldsymbol{x})\{p(y|\boldsymbol{x}, f_0, \delta)\}^{(1-t)}\{p(y|\boldsymbol{x}, f_1, \delta)\}^t \tag{3}$$

一方で ϕ が期待値リンク ℓ のときは $f_t^{(\ell)}(\boldsymbol{x})$ は $\mathcal{R}$ の中へ**混合測地線** $(1-t)\mu_0(\boldsymbol{x}) + t\mu_1(\boldsymbol{x})$ を誘導する．更に測地線上の KL ダイバージェンスを $\mathcal{F}$ に引き戻して考えると $\mathcal{F}$ 上にピタゴラスの恒等式が成立することが示せる (Eguchi & Omae, 2017).

3. 対数・指数平均

セミパラメトリック GLM の設定において予測関数空間 $\mathcal{F}$ の上で考えた特別な 2 つ測地線が，それぞれ $\mathcal{P}$ 上の指数測地線と $\mathcal{R}$ 上の混合測地線を導くことが分かった．しかし，回帰分析の目的を考えると必ずしも 2 つの場合だけでは十分ではない．説明変数 $\boldsymbol{X}$ から応答変数 Y への柔軟なモデリングのために線形予測関数を超えたものを考えたい．例えば，K 個の線形予測関数 $\{f_k(\boldsymbol{x}) = \boldsymbol{\beta}_k^\top \boldsymbol{x} + \alpha\}_{k=1}^K$ に対して次のような結合を考えよう．

$$f_{\tau,\pi}^{(\exp)}(\boldsymbol{x}) = \frac{1}{\tau}\log\Big(\sum_{k=1}^K \pi_k \exp\{\tau f_k(\boldsymbol{x})\}\Big), \quad \pi = (\pi_1, \cdots, \pi_K) \in \mathcal{S}_{K-1}$$

ここで τ はスケールパラメータで，$f_{\tau,\pi}^{(\exp)}$ は $\tau = 0$ のときは算術平均 $\sum_{k=1}^K \pi_k f_k$ となり，$\tau = \infty$ のときは最大値 $\max_{1\le k\le K} f_k$ となるように，K 個の線形予測関数 f_k を柔軟につなぐことが分かる．特に説明変数ベクトル $\boldsymbol{X}$ が K 個の異なる要因を表すグループ $\{\boldsymbol{X}_{(k)}\}_{k=1}^K$ に分かれていて $\boldsymbol{X} = (\boldsymbol{X}_{(1)}, \cdots, \boldsymbol{X}_{(K)})$ となるとき，$\{f_k(\boldsymbol{x}) = \boldsymbol{\beta}_{(k)}^\top \boldsymbol{x}_{(k)} + \alpha\}_{k=1}^K$ を結合した $f_{\tau,\pi}^{(\exp)}$ を考えることができる．遺伝子発現の予測問題では被験者グループに異なる潜在的なグループのマーカーがある場合この予測関数モデルは線形予測を超える予測性能を持つことが示されている (Omae et al. 2017).

ニューラルネットワークモデルにおいて最近，活性化関数を新しく ReLU, max 関数を選ぶことによって**深層学習**が可能になって来ているが，この予測関数の結合と密接に関連している．とくに τ を無限大にすると MaxOut ニューラルネットと制限ボルツマンマシンとの深い関連がある．密度関数のように正値をとる関数空間の情報幾何は深く研究されているが，予測関数のように実軸全体上の値を取る関数空間の情報幾何は今後の発展方向の一つに考えられる．