

異なるモデルの尤度関数の結合

統計数理研究所 数理・推論研究系 江口 真透
はこだて未来大学, 理化学研究所 AIP 竹之内 高志

1. はじめに

負の対数尤度関数を含む一般のロス関数をつなげるパスを提案する．データが単一の統計モデルではうまく捉えられないとき，異なるモデルのロス関数を一般化結合し，データの持っている異質性を柔軟に反映する方法を開発したい．更にガンマ・ロス関数を特別な結合でつなぐと混合モデルのロバスト推定が得られることが分かった．

2. ロス関数の一般化結合

データ $\mathcal{D} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ が与えられたとき，標準的な統計モデル $\mathcal{M}_1 = \{f_1(\mathbf{x}, \boldsymbol{\theta}_1) : \boldsymbol{\theta}_1 \in \Theta_1\}$ を仮定して，ロス関数 $L_1(\boldsymbol{\theta}_1, \mathcal{D}) = \sum_{i=1}^n \ell_1(\mathbf{x}_i, \boldsymbol{\theta}_1)$ の最小化によって $\boldsymbol{\theta}_1$ の推定量を得ることが通常の方法である．しかし，適切な単一のモデルの特定が困難で複数のモデル $\mathcal{M}_k = \{f_k(\mathbf{x}, \boldsymbol{\theta}_k) : \boldsymbol{\theta}_k \in \Theta_k\}$ ($k = 1, \dots, K$) が想定される状況を考えなければいけないことがしばしば起こる．このとき，モデル $\mathcal{M}_k$ 上のロス関数 $L_k(\boldsymbol{\theta}_k, \mathcal{D})$ の一般化平均による統合ロス関数

$$L_\phi(\boldsymbol{\omega}, \mathcal{D}) = - \sum_{i=1}^n \phi^{-1} \left(\sum_{k=1}^K \pi_k \phi(-\ell_k(\mathbf{x}_i, \boldsymbol{\theta}_k)) \right) \quad (1)$$

を定義する．ここで $\boldsymbol{\omega} = (\boldsymbol{\pi}, \boldsymbol{\theta}) = \{(\pi_k, \boldsymbol{\theta}_k)\}_{1 \leq k \leq K}$ とし， π_k は $\pi_k \geq 0, \sum_{k=1}^K \pi_k = 1$ を満たす混合比を表す．混合比ベクトル $\boldsymbol{\pi}$ の k -成分が 1 となるときの，この統合ロス関数 $L_\phi(\boldsymbol{\omega}, \mathcal{D})$ は k 番目のロス関数 $L_k(\boldsymbol{\theta}_k, \mathcal{D})$ に帰着されるが，一般の $\boldsymbol{\pi}$ では K 個のロス関数を混合した形になっている．統合ロス関数 (1) のパラメータ $\boldsymbol{\omega}$ に関する最小化によって定まる推定量を $\hat{\boldsymbol{\omega}}_\phi$ と書く．

逆温度パラメータ τ に対して $\phi_\tau(t) = \exp(\tau t)$ と取ると，統合ロス関数 (1) は

$$L_\tau^{\text{exp}}(\boldsymbol{\omega}, \mathcal{D}) = - \sum_{i=1}^n \log \left(\sum_{k=1}^K \pi_k \exp\{\tau \log f_k(\mathbf{x}_i, \boldsymbol{\theta}_k)\} \right)$$

で与えられる．このとき，逆温度が $\tau = 1$ のときは単に有限混合モデルの負の対数尤度関数になるが， τ を極限 ∞ にとると $L_\tau^{\text{exp}}(\boldsymbol{\omega}, \mathcal{D})$ は $-\sum_{i=1}^n \log \{ \max_{1 \leq k \leq K} f_k(\mathbf{x}_i, \boldsymbol{\theta}_k) \}$ に収束する性質が注目される．

つぎに，ガンマ・ロス関数とベキ指数生成関数を

$$\ell_\gamma(\mathbf{x}, \boldsymbol{\theta}) = -\frac{1}{\gamma} \left\{ \left(\frac{f(\mathbf{x}, \boldsymbol{\theta})}{\|f(\cdot, \boldsymbol{\theta})\|_\gamma} \right)^\gamma - 1 \right\}, \quad \phi_{\gamma, \tau}(t) = (1 + \gamma \tau t)^{\frac{1}{\gamma}}$$

と定めるとき， K 個の $\ell_\gamma(\mathbf{x}, \boldsymbol{\theta}_k)$ を $\phi_{\gamma, \tau}(t)$ でつないだ統合ロス関数 (1) はつぎのようになる：

$$L_{\gamma, \tau}(\boldsymbol{\omega}) = - \sum_{i=1}^n \frac{1}{\gamma} \left[\left(\sum_{k=1}^K \pi_k \left\{ \frac{f_k(\mathbf{x}_i, \boldsymbol{\theta}_k)}{\|f_k(\cdot, \boldsymbol{\theta}_k)\|_\gamma} \right\}^\tau \right)^{\frac{\gamma}{\tau}} - 1 \right]$$

ここで $\|\cdot\|_\gamma$ は $p = \gamma + 1$ の L_p ノルムを表す．このようにして次の命題で推定量の一致性が得られる．

定理 1. データ $\mathcal{D}$ が 次の密度関数を持つ分布からのランダムサンプルとする．

$$p_\gamma(\mathbf{x}, \boldsymbol{\omega}_0) = \frac{1}{z(\boldsymbol{\omega}_0)} \left[\sum_{k=1}^K \pi_{0k} \left\{ \frac{f_k(\mathbf{x}, \boldsymbol{\theta}_{0k})}{\|f_k(\cdot, \boldsymbol{\theta}_{0k})\|_\gamma} \right\}^{\gamma+1} \right]^{\frac{1}{\gamma+1}}$$

ここで $\boldsymbol{\omega}_0 = (\boldsymbol{\pi}_0, \boldsymbol{\theta}_0)$ ， $z(\boldsymbol{\omega}_0)$ は規格化定数とする．このとき，期待ロスが次を満たす：

$$\boldsymbol{\omega}_0 = \operatorname{argmin}_{\boldsymbol{\omega}} \mathbb{E}_{p_\gamma(\cdot, \boldsymbol{\omega}_0)} \{L_{\gamma, \gamma+1}(\boldsymbol{\omega})\}.$$

上の定理において一般の統合ロス関数 $L_{\gamma, \tau}(\boldsymbol{\omega})$ では推定量の一致性が云えなくて，逆温度が $\tau = \gamma + 1$ のときに限り，一致性が成立することに注意する．特に γ の極限 0 をとると $p_\gamma(\mathbf{x}, \boldsymbol{\omega})$ は有限混合モデルになり， $L_{\gamma, \gamma+1}(\boldsymbol{\omega})$ は負の対数尤度関数になるので有限混合モデルの最尤推定に他ならないが，一般の $L_{\gamma, \gamma+1}(\boldsymbol{\omega})$ では γ を自由にとってロバスト性と有効性の調整が可能になる．

正規分布モデルの場合， $\boldsymbol{\theta}_k$ を平均ベクトル $\boldsymbol{\mu}_k$ と分散行列 $\mathbf{V}_k$ すると， $L_{\gamma, \gamma+1}(\boldsymbol{\omega})$ の最小化アルゴリズムは $\boldsymbol{\mu}_k$ と $\mathbf{V}_k$ に対して **EM** アルゴリズムと同様な形で与えられる．しかも $f(\mathbf{x}_i, \boldsymbol{\theta}_k)^\gamma$ の重み付けが加わるので， $\gamma > 0$ の適切な選択により有限混合正規モデルのロバスト推定が可能になる．