異なるモデルの尤度関数の一般化結合

江口　真透　　数理・推論研究系　教授

概要：情報幾何の基本的な考えにe-測地線とm-測地線がある．これらは確率密度関数をつなげるパスの一つであるが，このような測地線は**損失関数の空間**にも考えることができる．この考えから，負の対数尤度関数をつなげるパスを提案する．データが未知の要因から単一の統計モデルではうまく捉えられないとき，この異なるモデルの尤度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．そのために，一般化された平均の中で温度パラメータを持たせた**対数・指数和**を採用した一般化結合を考えた．**逆温度**が∞のとき特別な性質が持つことが分かった．混合分布モデルやエクスパート・ミクスチュアとも密接な関連があることが分かった．

1. 問題の設定と主な結果

データ$\mathcal{D}=\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n\}$が与えられたとき，標準的な統計モデルを仮定したいが，適切な単一のモデルの特定が困難で複数のモデル$\mathcal{M}_k=\{f_k(\boldsymbol{x},\boldsymbol{\theta}_k):\theta_k\in\Theta_k\}$ $(k=1,\cdots,K)$ が想定される状況を考える．このとき, モデル$\mathcal{M}_k$の対数密度関数の一般化結合を次のように定める.

$$E_\tau(\boldsymbol{x},\boldsymbol{\omega})=-\frac{1}{\tau}\log\sum_{k=1}^{K}\pi_k\exp\{\tau\log f_k(\boldsymbol{x},\boldsymbol{\theta}_k)\}$$

ここで $\boldsymbol{\omega}=\{(\boldsymbol{\theta}_k,\pi_k)\}_{1\le k\le K}$，$\pi_k$は混合比とし，$\tau$を逆温度パラメータとよぶ．特に，次の性質に注目する．

$$\lim_{\tau\to\infty}E_\tau(\boldsymbol{x},\boldsymbol{\omega})=-\max_{1\le k\le K}\log f_k(\boldsymbol{x},\boldsymbol{\theta}_k)$$

データ$\mathcal{D}$に対して負の対数尤度関数の結合を

$$L\tau(\boldsymbol{\omega},\mathcal{D})=\sum_{i=1}^{n}E_\tau(\boldsymbol{x}_i,\boldsymbol{\omega})+\frac{1}{\tau}\log z_\tau(\boldsymbol{\omega})$$

と定義し，その推定量$\hat{\omega}_\tau$を$L\tau(\boldsymbol{\omega},\mathcal{D})$のパラメータ$\boldsymbol{\omega}$に関する最小化と定める．ここで

$$z_\tau(\boldsymbol{\omega})=\sum_{k=1}^{K}\pi_k\int\exp\{\tau\log f_k(\boldsymbol{x},\boldsymbol{\theta}_k)\}d\boldsymbol{x}$$

とする．パラメータ$\boldsymbol{\theta}_k$の推定関数は次の負の重み付きスコア関数とバイアス補正項の和となる.

$$-\sum_{i=1}^{n}p_k(\boldsymbol{x}_i,\boldsymbol{\omega},\tau)\frac{\partial}{\partial\boldsymbol{\theta}_k}\log f_k(\boldsymbol{x}_i,\boldsymbol{\theta}_k)+\frac{\partial}{\partial\boldsymbol{\theta}_k}\log z_\tau(\boldsymbol{\omega})$$

ここで重み関数は次で与えられる．

$$p_k(\boldsymbol{x},\boldsymbol{\omega},\tau)=\frac{\pi_k\,\exp\{\tau\log f_k(\boldsymbol{x},\boldsymbol{\theta}_k)\}}{\sum_{k'=1}^{K}\pi_{k'}\,\exp\{\tau\log f_k(\boldsymbol{x},\boldsymbol{\theta}_{k'})\}}.$$

このようにして次の命題で推定量の一致性が示される．

命題 1. データ$\mathcal{D}$の真の密度関数を

$$p_{\rm true}(\boldsymbol{x},\boldsymbol{\omega}_0,\tau)=\frac{1}{z_\tau(\boldsymbol{\omega}_0)}\sum_{k=1}^{K}\pi_{0k}\exp\{\tau\log f_k(\boldsymbol{x},\boldsymbol{\theta}_{0k})\}$$

と仮定する．このとき，上の推定量$\hat{\omega}_\tau$は真値$\boldsymbol{\omega}_0$への一致性を持つ．

証明は$\mathbb{E}_{p_{\rm true}}\{E_\tau(X,\boldsymbol{\omega})\}+\log z_\tau(\boldsymbol{\omega})$ が$\boldsymbol{\omega}=\boldsymbol{\omega}_0$のときに最小になることから直ちに得られる．

2. 指数型モデル

モデル$\mathcal{M}_k$が指数型で，

$$f_k(\boldsymbol{x},\boldsymbol{\theta}_k)=\exp\{\boldsymbol{\theta}_k^\top\boldsymbol{t}_k(\boldsymbol{x})-\kappa_k(\boldsymbol{\theta}_k)\},$$

と書けるとき，正規化定数は

$$z_\tau(\boldsymbol{\omega})=\sum_{k=1}^{K}\pi_k\exp\{\kappa_k(\tau\boldsymbol{\theta}_k)-\tau\kappa_k(\boldsymbol{\theta}_k)\}$$

となり，パラメータ$\boldsymbol{\theta}_k$の推定方程式は以下のような負の重み付きスコア関数とバイアス補正項に分解される.

$$\sum_{i=1}^{n}p_k(\boldsymbol{x}_i,\boldsymbol{\omega},\tau)\Big\{\boldsymbol{t}_k(\boldsymbol{x}_i)-\frac{\partial\kappa_k(\boldsymbol{\theta}_k)}{\partial\boldsymbol{\theta}_k}\Big\}=\rho_k(\boldsymbol{\omega},\tau)\Big\{\frac{\partial\kappa_k(\tau\boldsymbol{\theta}_k)}{\partial\boldsymbol{\theta}_k}-\frac{\partial\kappa_k(\boldsymbol{\theta}_k)}{\partial\boldsymbol{\theta}_k}\Big\}$$

ここで

$$p_k(\boldsymbol{x},\boldsymbol{\omega},\tau)=\frac{\pi_k\exp[\tau\{\boldsymbol{\theta}_k^\top\boldsymbol{t}(\boldsymbol{x})-\kappa_k(\boldsymbol{\theta}_k)\}]}{\sum_{k'=1}^{K}\pi_{k'}\exp[\tau\{\boldsymbol{\theta}_{k'}^\top\boldsymbol{t}(\boldsymbol{x})-\kappa_{k'}(\boldsymbol{\theta}_{k'})\}]}$$

$$\rho_k(\boldsymbol{\omega},\tau)=\frac{\pi_k\exp\{\kappa(\tau\boldsymbol{\theta}_k)-\tau\kappa(\boldsymbol{\theta}_k)\}}{\sum_{k'=1}^{K}\pi_{k'}\exp\{\kappa(\tau\boldsymbol{\theta}_{k'})-\tau\kappa(\boldsymbol{\theta}_{k'})\}}$$

正規モデル

典型的な例題としてp次元正規分布の場合，

$$f(\boldsymbol{x},\boldsymbol{\theta}_k)=\varphi(\boldsymbol{x},\boldsymbol{\mu}_k,\boldsymbol{V}_k)$$

を考える．ここで$\boldsymbol{\theta}_k=(\boldsymbol{\mu}_k,\boldsymbol{V}_k)$. 負の対数尤度関数の結合は

$$-\frac{1}{\tau}\Big\{\sum_{i=1}^{n}\log\sum_{k=1}^{K}\pi_k\varphi(\boldsymbol{x}_i,\boldsymbol{\mu}_k,\boldsymbol{V}_k)^\tau-\log\sum_{k=1}^{K}\pi_k{\rm det}(2\pi\boldsymbol{V}_k)^{\frac{1}{2}(1-\tau)}\Big\}$$

となる．これより，$\tau=1$ならば正規混合モデルの対数尤度関数に他ならないので最尤推定量はEMアルゴリズムで求められる．τを∞に極限を取ると，つぎのように表される．

$$-\sum_{i=1}^{n}\max_{1\le k\le K}\log\varphi(\boldsymbol{x}_i,\boldsymbol{\mu}_k,\boldsymbol{V}_k)+\frac{1}{2}\min_{1\le k\le K}\log\det(2\pi\boldsymbol{V}_k)$$

一般のτに対して不動点アルゴリズム：

$$(\boldsymbol{\mu}_k,\boldsymbol{V}_k)\leftarrow\left(\frac{\sum_{i=1}^{n}p_{ki}(\boldsymbol{\omega})\boldsymbol{x}_i}{\sum_{i=1}^{n}p_{ki}(\boldsymbol{\omega})},\ \frac{\sum_{i=1}^{n}p_{ki}(\boldsymbol{\omega})(\boldsymbol{x}_i-\boldsymbol{\mu}_k)(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^\top}{\sum_{i=1}^{n}p_{ki}(\boldsymbol{\omega})+\frac{\tau-1}{2}\rho_k(\boldsymbol{\omega})}\right)$$

が与えらえる．ここで

$$p_{ki}(\boldsymbol{\omega})=\frac{\pi_k\ \varphi(\boldsymbol{x}_i,\boldsymbol{\mu}_k,\boldsymbol{V}_k)^\tau}{\sum_{\ell=1}^{K}\pi_\ell\ \varphi(\boldsymbol{x}_i,\boldsymbol{\mu}_\ell,\boldsymbol{V}_\ell)^\tau},\ \rho_k(\boldsymbol{\omega})=\frac{\pi_k\ \det(2\pi\boldsymbol{V}_k)^{\frac{1}{2}(1-\tau)}}{\sum_{\ell=1}^{K}\pi_\ell\ \det(2\pi\boldsymbol{V}_\ell)^{\frac{1}{2}(1-\tau)}}$$

特に，τを∞に極限を取ると，

$$(\boldsymbol{\mu}_k,\boldsymbol{V}_k)\leftarrow\left(\frac{\sum_{i\in I_k(\boldsymbol{\omega})}\boldsymbol{x}_i}{|I_k(\boldsymbol{\omega})|},\frac{\sum_{i\in I_k(\boldsymbol{\omega})}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^\top}{|I_k(\boldsymbol{\omega})|+\frac{1}{2}\delta(k,k_{\rm min})}\right)$$

となる．　ここで$k_{\rm min}={\rm argmin}_{1\le k\le K}\log\det(2\pi V_k)$,

$$I_k(\boldsymbol{\omega})=\{i:\varphi(\boldsymbol{x}_i,\boldsymbol{\mu}_k,\boldsymbol{V}_k)=\max_{1\le k'\le K}\varphi(\boldsymbol{x}_i,\boldsymbol{\mu}_{k'},\boldsymbol{V}_{k'})\}.$$

分散行列$\boldsymbol{V}_k$が同一で既知であると制限すると，このアルゴリズムはK平均アルゴリズムと一致する．適当な逆温度τを選択すればハードとソフトのクラスタリングの両方の良い点を持つ方法が提案できる．