ガンマ・ダイバージェンスに基づく
ロバストかつスパースな回帰モデリング

総合研究大学院大学　川島　孝行
統計数理研究所　藤澤　洋徳

1　はじめに

KL ダイバージェンスに基づく回帰は，以下のように定義できる．

$$D_{KL}(g(y|x), f(y|x;\theta); g(x)) = \int D_{KL}(g(y|x), f(y|x;\theta)g(x)dx.$$

ただし，$g(y|x)$ および $g(x)$ は，データを生成する分布で，$f(y|x;\theta)$ は，パラメトリックモデルである．$g(y|x)$ および $g(x)$ を，それぞれ経験密度関数 $\bar{g}(y|x)$ と $\bar{g}(x)$ で置き換えると，最尤推定に基づく回帰に一致する．しかし，KL ダイバージェンスは，外れ値に弱い．そこで，我々は，ロバストなダイバージェンスとして知られている，γ ダイバージェンス (Fujisawa and Eguchi 2008) に基づく回帰を考える．また，これにスパース正則化を組み合わせることで，ロバストかつスパースな回帰モデリングを達成する．

2　回帰のための γ ダイバージェンス

まずは，独立同一分布の場合における γ ダイバージェンスを，以下のように回帰用へと発展させる．

$$\begin{aligned}
&D_\gamma(g(y|x), f(y|x;\theta); g(x)) \\
&= \frac{1}{\gamma(1+\gamma)} \log \int \left(\int g(y|x)^{1+\gamma} dy \right) g(x)dx \\
&\quad - \frac{1}{\gamma} \log \int \left(\int g(y|x) f(y|x;\theta)^{\gamma} dy \right) g(x)dx + \frac{1}{1+\gamma} \log \int \left(\int f(y|x;\theta)^{1+\gamma} dy \right) g(x)dx \\
&= -d_\gamma(g(y|x), g(y|x); g(x)) + d_\gamma(g(y|x), f(y|x;\theta); g(x)).
\end{aligned}$$

これの経験推定は以下で与えられる．

$$\begin{aligned}
&d_\gamma(\bar{g}(y|x), f(y|x;\theta); \bar{g}(x)) \\
&\quad = -\frac{1}{\gamma} \log \left\{ \frac{1}{n} \sum_{i=1}^{n} f(y_i|x_i;\theta)^{\gamma} \right\} + \frac{1}{1+\gamma} \log \left\{ \frac{1}{n} \sum_{i=1}^{n} \int f(y|x_i;\theta)^{1+\gamma} dy \right\},
\end{aligned} \tag{1}$$

ただし，$\bar{g}(y|x), \bar{g}(x)$ は経験密度関数を表す．また，これにスパース正則化 $P(\theta)$ を加えたものは以下で与えられる．

$$L_\gamma(\theta;\lambda) = d_\gamma(\bar{g}(y|x), f(y|x;\theta); \bar{g}(x)) + \lambda P(\theta). \tag{2}$$

パラメータの推定は，(2) の最小化を行うことで可能となる，スパース正則化を含まない (1) の場合には，効率的に推定を行うアルゴリズムが提案されている．(Fujisawa and Eguchi 2008) しかしながら，(2) の場合には，上記の方法では，効率的に推定を行うことが困難になる．そのため，我々は，MM アルゴリズム (Hunter and Lange 2004) の考えを用い，スパース正則化を加えた場合でも，効率の良い推定アルゴリズムの提案を行う．

3 ロバストかつスパースな線形回帰

ここでは，具体例として，線形回帰の γ ダイバージェンスによるロバストかつスパースな推定アルゴリズムを示す．パラメトリックモデル $f(y|x;\theta)$ として，正規モデル $\phi(y;\beta_0+x^T\beta,\sigma^2)$ を考える．MM アルゴリズムを用いることによって，(m)-回目の反復において，次の最適化問題が必要となる．

$$\beta_0^{(m+1)},\beta^{(m+1)},{\sigma^2}^{(m+1)}=\mathop{\arg\min}_{\beta_0,\beta,\sigma^2}\frac{1}{2(1+\gamma)}\log\sigma^2+\frac{1}{2}\sum_{i=1}^n\alpha_i(\theta^{(m)})\frac{(y_i-\beta_0-x_i^T\beta)^2}{\sigma^2}+\lambda||\beta||_1.$$

ただし，$\alpha_i(\theta)=\phi(y_i;\beta_0+x_i^T\beta,\sigma^2)^\gamma\Big/\sum_{k=1}^n\phi(y_k;\beta_0+x_k^T\beta,\sigma^2)^\gamma$ である．これを解くことで，実際の推定アルゴリズムが得られる．

$$\begin{aligned}
\beta_0^{(m+1)}&=\sum_{i=1}^n\alpha_i^{(m)}(y_i-{x_i}^T\beta^{(m)}),\\
\beta_j^{(m+1)}&=S\left(\sum_{i=1}^n\alpha_i^{(m)}(y_i-\beta_0^{(m+1)}-\sum_{l\neq j}x_{il}\beta_l)x_{ij},{\sigma^2}^{(m)}\lambda\right)\Bigg/\left(\sum_{i=1}^n\alpha_i^{(m)}x_{ij}^2\right)\quad(j=1,\dots,p),\\
{\sigma^2}^{(m+1)}&=(1+\gamma)\sum_{i=1}^n\alpha_i^{(m)}(y_i-\beta_0^{(m+1)}-x_i^T\beta^{(m+1)})^2.
\end{aligned}$$

ただし $S(t,\lambda)=\mathrm{sign}(t)(|t|-\lambda)_+$ である．この推定アルゴリズムは，目的関数に関して単調減少性を持つ．発表当日は，数値実験および実データに対する結果，その他の回帰手法であるロジスティック回帰やポアソン回帰への拡張についても発表を行う．