次元縮約の問題における逆説的な現象について

逸見 昌之　　データ科学研究系 准教授

【はじめに】

高次元のデータに対する次元縮約の方法には様々なものがありますが、p次元の共変量ベクトル $\mathbf{X}$ が与えられた下での1次元の結果変量 Y の条件付き分布（あるいは条件付き期待値）が、$\beta^T\mathbf{X}$（但し β はある$p\times d$ 行列で$p>d$）というpよりも低次元の変量を通してのみ $\mathbf{X}$ に依存するという仮定の下で定式化された次元縮約の問題は、Li (1991) によって層別逆回帰法（sliced inverse regression (SIR)）が提案されて以来、統計学の分野で様々な研究が成されてきました。これまで提案されてきた方法の多くは、必要に応じてさらにいくつかの仮定を置きながら個々に工夫を施すというものでしたが、近年、Ma and Zhu (2012) はこの問題をセミパラメトリック推定の問題と見なし、その一般論による統一的な方法論の枠組みを提示して、これまでの提案法の多くが、その特別な場合として見なせることを示しました。しかしながら、そこである不思議な現象が起こることが述べられています。それは、これまで多くの方法で付加されていた条件が仮に成り立っていたとしても、それを無視して（用いずに）推定を行った方が、推定効率が良くなる、ということです。この逆説的な現象は、後にMa and Zhu (2013) によって厳密に証明されていますが、ここでは、幾何的な観点からこのような現象が起こるメカニズムについて考察します。なお本研究は、サウスカロライナ大学のYanyuan Ma教授と共同で行ったものです。

【次元縮約の問題】

次元縮約の条件 $P(Y\le y\mid\mathbf{X})=P(Y\le y\mid\boldsymbol{\beta}^T\mathbf{X})\ (\forall y\in\mathbf{R})$（ $\boldsymbol{\beta}: p\times d$ 行列, $p>d$） 中心化部分空間（central subspace）$S_{Y|\mathbf{X}}$: $\boldsymbol{\beta}$ の列ベクトルで張られる最小の部分空間

$E(Y\mid\mathbf{X})=E(Y\mid\boldsymbol{\beta}^T\mathbf{X})$ 中心化平均部分空間（central mean subspace）$S_{E(Y|\mathbf{X})}$: $\boldsymbol{\beta}$ の列ベクトルで張られる最小の部分空間

一般性を失うことなく、$E(\mathbf{X})=\mathbf{0}, Cov(\mathbf{X})=I_p$ とし、$\boldsymbol{\beta}:=\begin{pmatrix} I_d \\ \boldsymbol{\beta}_l \end{pmatrix}$ とする。（但し、I_d はd 次単位行列で、$\boldsymbol{\beta}_l$ が推定すべきパラメータとなる。）

今回は、中心化部分空間の方について考える。

【推定方程式】

セミパラメトリックモデル $p_{\mathbf{X},Y}(\mathbf{x},y;\boldsymbol{\beta},\eta)=\eta_1(\mathbf{x})\eta_2(y,\boldsymbol{\beta}^T x)\ (\eta=(\eta_1,\eta_2))$

セミパラメトリック推定理論に基づく推定方程式 $\sum_{i=1}^n\sum_{j=1}^q \mathbf{g}_j(Y_i)\{\mathbf{a}_j(\mathbf{X}_i)-\mathbf{m}_j(\boldsymbol{\beta}^T\mathbf{X}_i)\}^T=\mathbf{0}$ （但し、$\{(\mathbf{X}_i,Y_i)\}_{i=1}^n$ は $p_{\mathbf{X},Y}(\mathbf{x},y;\boldsymbol{\beta},\eta)$ からのランダムサンプル）

$\mathbf{g}_j$: p_g 次元ベクトル値関数, $\mathbf{a}_j$: p_a 次元ベクトル値関数, $p_g p_d=(p-d)d=\dim\boldsymbol{\beta}_l$, $\mathbf{m}_j(\boldsymbol{\beta}^T\mathbf{X}):=E\{\mathbf{a}_j(\mathbf{X})\mid\boldsymbol{\beta}^T\mathbf{X}\}\ (j=1,\ldots,q)$

【既存法との関係】

例えば、SIR (Li, 1991) では、$\mathbf{g}_j(Y)=D_jE(\mathbf{X}\mid Y), \mathbf{a}_j(\mathbf{X})=X_j, \mathbf{X}=(X_1,\ldots,X_p)^T, D_j:(p-d)d\times p$ 行列

特に、$E(\mathbf{X}\mid\boldsymbol{\beta}^T\mathbf{X})=\boldsymbol{\beta}(\boldsymbol{\beta}^T\boldsymbol{\beta})^{-1}\boldsymbol{\beta}^T\mathbf{X}$（線型条件）を仮定、つまり $\mathbf{m}_j(\boldsymbol{\beta}^T\mathbf{X}):=E\{\mathbf{a}_j(\mathbf{X})\mid\boldsymbol{\beta}^T\mathbf{X}\}=E\{X_j\mid\boldsymbol{\beta}^T\mathbf{X}\}$ は既知とする。

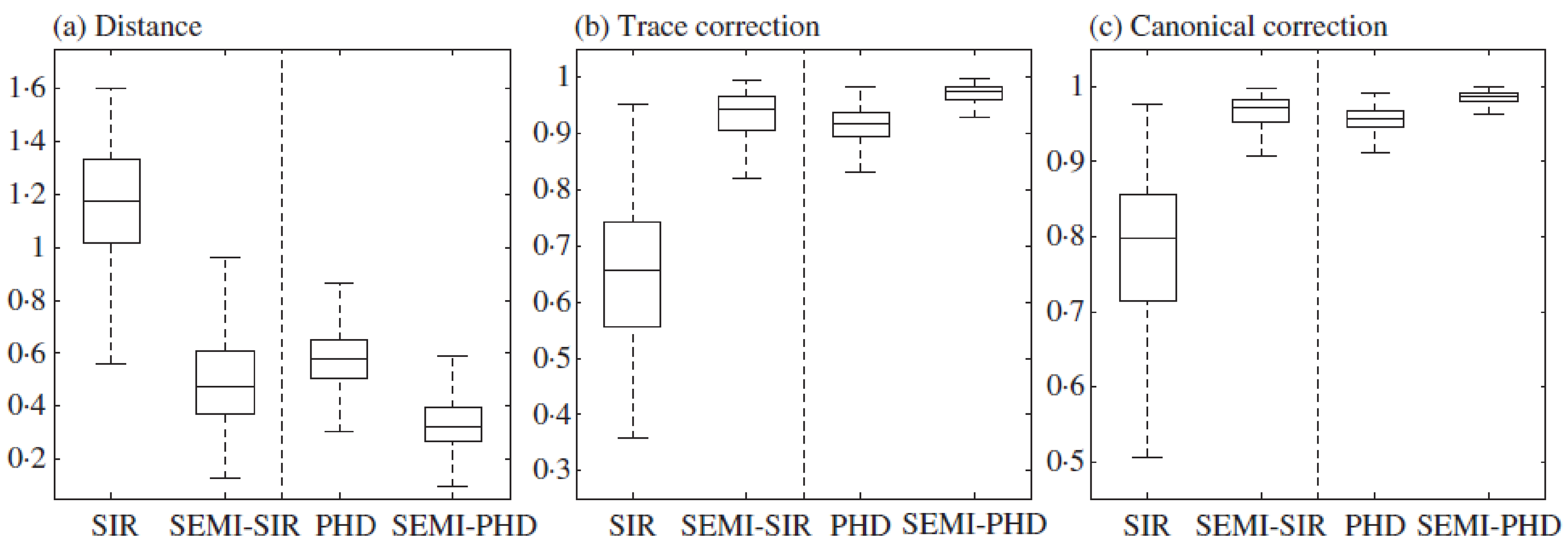
【関数mが既知の場合と未知の場合】

推定方程式 $\sum_{i=1}^n \mathbf{g}(Y_i)\{\mathbf{a}(\mathbf{X}_i)-\mathbf{m}(\boldsymbol{\beta}^T\mathbf{X}_i)\}^T=\mathbf{0}$ （以下、q=1の場合で考える。）

$\mathbf{m}(\mathbf{X},\boldsymbol{\beta})$ が未知の場合は、カーネルによるノンパラメトリック推定を行い、$\tilde{\boldsymbol{\beta}},\hat{\boldsymbol{\beta}}$ をそれぞれ $\mathbf{m}(\mathbf{X},\boldsymbol{\beta})$ が既知の場合、未知の場合の（推定方程式の解としての）$\boldsymbol{\beta}$ の推定量とすると、

$\mathrm{Avar}\{vecl(\hat{\boldsymbol{\beta}})\}\le\mathrm{Avar}\{vecl(\tilde{\boldsymbol{\beta}})\}$ (Ma and Zhu, 2013) 但し、Avar は漸近分散共分散行列を表し、$vecl$ は行列の下側部分をベクトル化したものを表す。

【シミュレーション】



線型条件が成り立っている下で、
サンプル数200、シミュレーション回数1000
(Ma and Zhu, 2013)

【パラメトリックモデルとプラグイン推定量】

$\mathbf{m}(\boldsymbol{\beta}^T\mathbf{X},\boldsymbol{\alpha})$: $E\{\mathbf{a}(\mathbf{X})\mid\boldsymbol{\beta}^T\mathbf{X}\}$ に対するパラメトリックモデル

回帰モデル $\mathbf{a}(\mathbf{X})=\mathbf{m}(\boldsymbol{\beta}^T\mathbf{X},\boldsymbol{\alpha})+\boldsymbol{\varepsilon}, E(\boldsymbol{\varepsilon}\mid\boldsymbol{\beta}^T\mathbf{X})=\mathbf{0}, V(\boldsymbol{\varepsilon}\mid\boldsymbol{\beta}^T\mathbf{X})=:\mathbf{Q}(\boldsymbol{\beta}^T\mathbf{X})$

推定方程式 $\sum_{i=1}^n \mathbf{A}(\boldsymbol{\beta}^T\mathbf{X}_i)\{\mathbf{a}(\mathbf{X}_i)-\mathbf{m}(\boldsymbol{\beta}^T\mathbf{X}_i,\boldsymbol{\alpha})\}=\mathbf{0}$ によってパラメータ $\boldsymbol{\alpha}=\boldsymbol{\alpha}(\boldsymbol{\beta})$ を推定 （但し、$\mathbf{X}_1,\ldots,\mathbf{X}_n$ はランダムサンプルで、$\mathbf{A}(\boldsymbol{\beta}^T\mathbf{X}_i)$ は $p_\alpha\times p_a$ 行列）

特に、最適な重み付け最小二乗推定方程式 $\sum_{i=1}^n \mathbf{m}_{\boldsymbol{\alpha}}^T(\boldsymbol{\beta}^T\mathbf{X}_i,\boldsymbol{\alpha})Q^{-1}(\boldsymbol{\beta}^T\mathbf{X}_i)\{\mathbf{a}(\mathbf{X}_i)-\mathbf{m}(\boldsymbol{\beta}^T\mathbf{X}_i,\boldsymbol{\alpha})\}=\mathbf{0}$ の解を $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta})$ として、

$\boldsymbol{\beta}$ の推定方程式 $\sum_{i=1}^n \mathbf{g}(Y_i)\{\mathbf{a}(\mathbf{X}_i)-\mathbf{m}(\boldsymbol{\beta}^T\mathbf{X}_i,\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}))\}^T=\mathbf{0}\ \Longrightarrow\ \boldsymbol{\beta}$ のプラグイン推定量 $\breve{\boldsymbol{\beta}}$

【3つの推定量の比較】

$\sqrt{n}vecl(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})=1/\sqrt{n}\sum_{i=1}^n\Sigma_A^{-1}\{\mathbf{u}(\mathbf{X}_i,Y_i,\boldsymbol{\beta})+\mathbf{v}(\mathbf{X}_i,Y_i,\boldsymbol{\beta})\}+o_p(1)$

$\sqrt{n}vecl(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})=1/\sqrt{n}\sum_{i=1}^n\Sigma_A^{-1}\mathbf{u}(\mathbf{X}_i,Y_i,\boldsymbol{\beta})+o_p(1)$

$\sqrt{n}vecl(\breve{\boldsymbol{\beta}}-\boldsymbol{\beta})=1/\sqrt{n}\sum_{i=1}^n\Sigma_A^{-1}\{\mathbf{u}(\mathbf{X}_i,Y_i,\boldsymbol{\beta})+\mathbf{v}(\mathbf{X}_i,Y_i,\boldsymbol{\beta})+\mathbf{w}(\mathbf{X}_i,Y,\boldsymbol{\beta})\}+o_p(1)$

但し、$\mathbf{u}(\mathbf{X},Y,\boldsymbol{\beta}):=vec([\mathbf{g}(Y)-E\{\mathbf{g}(Y)\mid\boldsymbol{\beta}^T\mathbf{X}\}][\mathbf{a}(\mathbf{X})-E\{\mathbf{a}(\mathbf{X})\mid\boldsymbol{\beta}^T\mathbf{X}\}]^T)$

$\mathbf{v}(\mathbf{X},Y,\boldsymbol{\beta}):=vec([E\{\mathbf{g}(Y)\mid\boldsymbol{\beta}^T\mathbf{X}\}][\mathbf{a}(\mathbf{X})-E\{\mathbf{a}(\mathbf{X})\mid\boldsymbol{\beta}^T\mathbf{X}\}]^T)$

$\mathbf{w}(\mathbf{X},Y,\boldsymbol{\beta}):=-\mathbf{B}_2\mathbf{B}_3^{-1}\mathbf{A}(\boldsymbol{\beta}^T\mathbf{X})[\mathbf{a}(\mathbf{X})-E\{\mathbf{a}(\mathbf{X})\mid\boldsymbol{\beta}^T\mathbf{X}\}]$

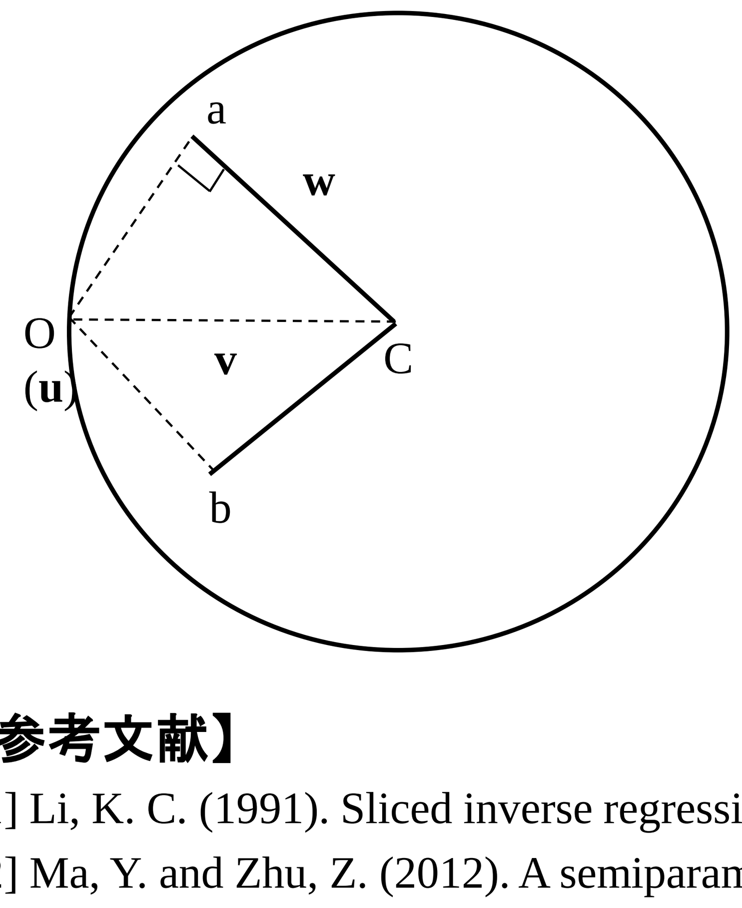
$\Sigma_A:=E([\mathbf{a}(\mathbf{X})-E\{\mathbf{a}(\mathbf{X})\mid\boldsymbol{\beta}^T\mathbf{X}\}]\otimes\partial E\{\mathbf{g}(Y)\mid\boldsymbol{\beta}^T\mathbf{X}\}/\partial vecl(\boldsymbol{\beta})^T])$

$\mathbf{B}_2:=E[\mathbf{m}_{\boldsymbol{\alpha}}\{\mathbf{X},\boldsymbol{\beta},\boldsymbol{\alpha}_0(\boldsymbol{\beta})\}\otimes\mathbf{g}(Y)]$

$\mathbf{B}_3:=E[\mathbf{m}_{\boldsymbol{\alpha}}^T\{\mathbf{X},\boldsymbol{\beta},\boldsymbol{\alpha}_0(\boldsymbol{\beta})\}\mathbf{Q}^{-1}(\boldsymbol{\beta}^T\mathbf{X})\mathbf{m}_{\boldsymbol{\alpha}}\{\mathbf{X},\boldsymbol{\beta},\boldsymbol{\alpha}_0(\boldsymbol{\beta})\}]$

$\mathbf{m}_{\boldsymbol{\alpha}}\{\mathbf{X},\boldsymbol{\beta},\boldsymbol{\alpha}_0(\boldsymbol{\beta})\}:=\partial\mathbf{m}(\mathbf{X},\boldsymbol{\beta},\boldsymbol{\alpha})/\partial\boldsymbol{\alpha}^T\ |_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0(\boldsymbol{\beta})}$

ここで、$\mathbf{u}(\mathbf{X},Y,\boldsymbol{\beta})$ は $\mathbf{v}(\mathbf{X},Y,\boldsymbol{\beta}),\mathbf{w}(\mathbf{X},Y,\boldsymbol{\beta})$ に直交し、$\mathbf{w}(\mathbf{X},Y,\boldsymbol{\beta})$ は $\mathbf{v}(\mathbf{X},Y,\boldsymbol{\beta})+\mathbf{w}(\mathbf{X},Y,\boldsymbol{\beta})$ に直交する。



$(\mathbf{v}+\mathbf{w})\perp\mathbf{w}$

$\Longrightarrow$

$E\{(\mathbf{v}+\mathbf{w})^T(\mathbf{v}+\mathbf{w})\}\le E(\mathbf{v}^T\mathbf{v})$

$\mathrm{Avar}\{vecl(\breve{\boldsymbol{\beta}})\}\le\mathrm{Avar}\{vecl(\tilde{\boldsymbol{\beta}})\}$（を示唆）

【ノンパラメトリック推定】

$\boldsymbol{\beta}^T\mathbf{X}=\boldsymbol{\beta}^T\mathbf{x}_0$ のとき、$\mathbf{m}(\boldsymbol{\beta}^T\mathbf{X}):=E\{\mathbf{a}(\mathbf{X})\mid\boldsymbol{\beta}^T\mathbf{X}\}=\mathbf{c}_0$ として

推定方程式 $\sum_{i=1}^n I(\boldsymbol{\beta}^T\mathbf{X}_i=\boldsymbol{\beta}^T\mathbf{x}_0)Q^{-1}(\boldsymbol{\beta}^T\mathbf{X}_i)\{\mathbf{a}(\mathbf{X}_i)-\mathbf{c}_0\}=\mathbf{0}$

$$\Longrightarrow\quad \mathbf{c}_0=\frac{\sum_{i=1}^n I(\boldsymbol{\beta}^T\mathbf{X}_i=\boldsymbol{\beta}^T\mathbf{x}_0)\mathbf{a}(\mathbf{X}_i)}{\sum_{i=1}^n I(\boldsymbol{\beta}^T\mathbf{X}_i=\boldsymbol{\beta}^T\mathbf{x}_0)}$$

期待値関数の滑らかさを仮定して

$$\mathbf{c}_0=\frac{\sum_{i=1}^n K_h(\boldsymbol{\beta}^T\mathbf{X}_i-\boldsymbol{\beta}^T\mathbf{x}_0)\mathbf{a}(\mathbf{X}_i)}{\sum_{i=1}^n K_h(\boldsymbol{\beta}^T\mathbf{X}_i-\boldsymbol{\beta}^T\mathbf{x}_0)}$$

（Nadaraya-Watsonノンパラメトリック
カーネル回帰推定量）

【参考文献】

[1] Li, K. C. (1991). Sliced inverse regression for dimension reduction, *Journal of the American Statistical Association*, 86, 316-342.

[2] Ma, Y. and Zhu, Z. (2012). A semiparametric approach to dimension reduction, *Journal of the American Statistical Association*, 107, 168-179.

[3] Ma, Y. and Zhu, Z. (2013). Efficiency Loss Caused by Linearity Condition in Dimension Reduction, *Biometrika*, 100, 371-383.