On the degrees of freedom in shrinkage estimation

加藤賢悟（東京大学大学院経済学研究科）

1　設定

$y = (y_1, \ldots, y_n)'$ を目的変数，$x_j = (x_{1j}, \ldots, x_{nj})'$，$j = 1, \ldots, p$ を p 個の線形独立な説明変数，$X = [x_1 \cdots x_p]$ をデータ行列とする．このとき，次のような線形回帰モデルを考える：

$$y = X\beta + \epsilon.$$

ここで，$\beta = (\beta_1, \ldots, \beta_p)'$ は係数ベクトル，$\epsilon \sim N_n(0, \sigma^2 I_n)$ は誤差項である．

Lasso，group Lasso を一般化して，次のような最小化問題の解で定義される推定量 $\hat{\beta}_K$ を考える．$u, v \in \mathbb{R}^p$ にたいして，$\langle u, v \rangle = u'Vv$ と定める．ここで，$V = X'X$ であって，$\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$ と書く．このとき，$\hat{\beta}_K$ を次の最小化問題の解として定義する．

$$\min_{\beta} \|\beta - \hat{\beta}^{\circ}\| \ \text{ subject to } \beta \in K.$$

ここで，$\hat{\beta}^{\circ}$ は β の OLS 推定量，$K \subset \mathbb{R}^p$ は閉凸集合である．$\hat{\beta}_K$ は $\hat{\beta}^{\circ}$ を K に射影したものである．実際には，与えられた閉凸集合のコレクション $\mathcal{K}$ のなかから最適な $\hat{K}$ を選んで，$\hat{\beta}_{\hat{K}}$ を最終的な推定量とする．本報告では，K の選択に関する統一的な規準を導出する．

2　主結果

2.1　SURE 法

最適な K として，モデルの予測リスクを最小にするものを選ぶこととする．予測リスクは未知なので，推定する必要がある．

予測量 $\hat{\mu}_K = \hat{\mu}_K(y) = X\hat{\beta}_K$ を考える．y^{new} を y と同じ分布に従う独立なランダムベクトルとし，$\hat{\mu}_K$ の予測リスクを $E(\|y^{new} - \hat{\mu}_K\|_2^2)/n$ で測る．

このとき，$E(\|y^{new} - \hat{\mu}_K\|_2^2) = E(\|y - \hat{\mu}_K\|_2^2) + 2df(\hat{\mu}_K)\sigma^2$ と表すことができる．ここで，$df(\hat{\mu}_K) = \sum_{i=1}^n \mathrm{cov}(\hat{\mu}_i, y_i)/\sigma^2$ は予測量 $\hat{\mu}_K$ の自由度(degrees of freedom) と呼ばれる．Stein の補題から，自由度の不偏推定量が $\widehat{df}(\hat{\mu}_K) = \mathrm{div}\, \hat{\mu}_K$ で与えられることがわかる．従って，予測リスクの不偏推定量は，$C_p(\hat{\mu}_K) = \|y - \hat{\mu}_K\|_2^2/n + 2\widehat{df}(\hat{\mu}_K)\sigma^2/n$ で与えられる．ところで，$\hat{\beta}_K$ が $\hat{\beta}^{\circ}$ に関して (全) 微分可能であれば，$\mathrm{div}\, \hat{\mu}_K = \mathrm{tr}(\partial \hat{\beta}_K / \partial \hat{\beta}^{\circ})$ となる．

2.2 Divergence formula

$K \subset \mathbb{R}^p$ を閉凸集合とする．$x \in \mathbb{R}^p$ に対して，x_K を $\langle \cdot, \cdot \rangle$ に関する x の K への射影とする．いま，写像 $f: \mathbb{R}^p \to \mathbb{R}^p$ を $f(x) = x_K$ で定める．f のダイバージェンスを求める．

$s \in \partial K$ に対して，s における法錘は $N(K, s) = \{y - s | y_K = s\}$ で与えられる．法錘 $N(K, s)$ の次元に応じて，境界 ∂K は $\partial K = D_1 \cup \cdots \cup D_p$ と排反に分割される．ここで，$D_m = \{s \in \partial K | \dim N(K, s) = m\}$ である．いま，$E_m = \{x \in \mathbb{R}^p \setminus K | x_K \in D_m\}$ と定めると，$\mathbb{R}^p \setminus K$ の排反な分割 $\mathbb{R}^p \setminus K = E_1 \cup \cdots \cup E_p$ を得る．次の仮定をおく．

仮定 2.1. D_m は $(p-m)$ 次元の C^2 級多様体であって，有限個の相対開連結成分からなる．さらに，$E_m \backslash E_m^\circ$ の *Lebesgue* 測度は 0 である．

$\theta = (\theta^1, \ldots, \theta^{p-m})$ を D_m の C^2 級局所座標系とし，$s \in D_m$ を $s = s(\theta)$ と書く．D_m の s における接空間は $T_{s(\theta)} D_m = \mathrm{span}\,\{b_a(\theta) = \partial s / \partial \theta^a(\theta), a = 1, \ldots, p-m\}$ で与えられる．また，$T_{s(\theta)} D_m$ に直交する正規直交系 $\{n_\alpha(\theta), \alpha = 1, \ldots, m\}$ を 1 つとる．このとき，

$$(\theta, \tau) \mapsto \varphi(\theta, \tau) = (s(\theta) + \sum_{\alpha=1}^{m} \tau^\alpha n_\alpha(\theta))$$

を E_m° の C^1 級局所パラメータ付けとしてとって，f を局所座標 (θ, τ) に関して $f(\theta, \tau) = s(\theta)$ と表すことができる．

局所座標系 $\theta = (\theta^1, \ldots, \theta^{p-m})$ に付随する第 1 基本形式を $G(\theta)$，D_m の法線方向 $n_\alpha(\theta)$ に関する第 2 基本形式を $H_\alpha(\theta)$ と書く．また，$x = \varphi(\theta, \tau)$ に対して，$H(\theta, \tau) = -\sum_{\alpha=1}^{m} \tau^\alpha H_\alpha(\theta)$ とおく．これは，半正定値行列である．このとき，次の補題を得る．

補題 2.1. ダイバージェンス $\mathrm{div}\, f(x) = \sum_{j=1}^{p} \partial f_j(x) / \partial x_j$，$x \in E_m^\circ$ は

$$\mathrm{div}\, f(x) = \sum_{a=1}^{p-m} \frac{1}{1 + \kappa_a(x)}$$

で与えられる．ここで，$\kappa_a(x) = \kappa_a(\theta, \tau)$，$a = 1, \ldots p-m$ は，

$$|H(\theta, \tau) - \kappa G(\theta)| = 0 \tag{1}$$

をみたす固有値である．

2.3 Degrees of freedom

$\hat{\beta}^\circ \in E_m$ に対して，$x = \hat{\beta}^\circ$，$x_K = \hat{\beta}_K$ としたときの固有方程式 (1) の根を $\kappa_{m,a}(\hat{\beta}^\circ), a = 1, \ldots, p-m$ と書く．また，形式的に $E_0 = K$，$\kappa_{0,a} \equiv 0$，$a = 1, \ldots, p$ と定める．

定理 2.1. 以上の準備のもとで，

$$\widehat{df}(\hat{\mu}_K) = \sum_{m=0}^{p} \sum_{a=1}^{p-m} \frac{1}{1 + \kappa_{m,a}(\hat{\beta}^\circ)} I(\hat{\beta}_K \in E_m)$$

は $\hat{\mu}_K = X \hat{\beta}_K$ の自由度 $df(\hat{\mu}_K)$ の不偏推定量を与える．