修士論文 要旨

On High-Dimensional Asymptotic Properties of Error Truncated Estimators (誤差項打切を用いた推定量の高次元における漸近的な振る舞いについて)

数理情報学専攻 48206202 **安藤 瞭**

指導教員 **駒木 文保 教授**

1 はじめに

近年, 技術革新と共に高次元データを扱う機会が増えて来ている. 高次元データにおいては低次元データとは異なる振る舞いが現れることがある. 例として, 分散共分散行列の不安定化や二重降下現象 ([1, 3]), 球面集中現象 ([2]) などの現象が挙げられる. 従って, 高次元状況下特有の研究が必要とされている.

本研究においては, 線形回帰モデル

$$y_i = x_i^\top \beta + \epsilon_i, \quad i = 1, 2, ..., n \tag{1}$$

を考え, 高次元データが得られたという仮定のもとで回帰係数 $\beta \in \mathbb{R}^p$ を推定することを考える. 今回, 我々は誤差項打切を用いた推定量を用いて β を推定することを考え, その高次元における漸近的な out-of-sample リスクについて研究した.

2 準備や仮定

本研究においては, データは線形モデル (1) を仮定する. (1) において, n はサンプル数, p はデータの次元数を表し, $i = 1, 2, ..., n$ に対して, $x_i \in \mathbb{R}^p$ は特徴データ, $y_i \in \mathbb{R}$ はターゲットデータ, $\epsilon_i \in \mathbb{R}$ は誤差を表す. x_i と ϵ_i はそれぞれ独立に生成されており, それぞれ $i = 1, 2, ..., n$ に対して独立同分布に生成されていると仮定する. ここで, 誤差 ϵ_i に平均や分散の存在は仮定しない. すなわち, ϵ_i の従う分布としてコーシー分布などの裾野の重い分布を用いても良い. 以下では, $X \in \mathbb{R}^{n\times p}$ は各行が x_i の行列, $\epsilon \in \mathbb{R}^n$ は各要素が ϵ_i のベクトルを表すと仮定する.

上記の基礎的な仮定の下, β を以下のように推定することを考える.

$$\hat{\beta}_d = (X^\top D_d X)^+ X^\top D_d Y$$

ここで, $d > 0$ で,

$$D_d = \begin{pmatrix} \mathbb{1}_{\{\epsilon_1^2 \le d^2\}} & 0 & \cdots & 0 \\ 0 & \mathbb{1}_{\{\epsilon_2^2 \le d^2\}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mathbb{1}_{\{\epsilon_n^2 \le d^2\}} \end{pmatrix},$$

である. 本研究ではさらに以下の仮定を考える.

Assumption 1 (High-Dimensional Asymptotics)**.** 以下が成立すると仮定する*:*

1 データ $X \in \mathbb{R}^{n\times p}$ は $\mathbb{E}[Z_{ij}] = 0$ や $\mathrm{Var}[Z_{ij}] = 1$ であるような $n \times p$ 行列 Z と半正定値 $p \times p$ 分散共分散行列 Σ を用いて $X = Z\Sigma^{1/2}$ のように生成されている.

2 サンプル数 n とサンプルの次元 p がそれぞれ同時に $n \to \infty$, $p \to \infty$ となり, それらの比率が $p/n \to \gamma > 0$ となる.

3 Σ のスペクトル分布 F_Σ が $[0, \infty)$ 上に値を取る極限確率分布 H に収束すると仮定する. この H は母スペクトル分布 *(population spectral distribution; PSD)* と呼ばれる.

Assumption 2 (Deterministic Coefficients)**.** 回帰係数 $\beta \in \mathbb{R}^p$ は $r > 0$ に対して $\|\beta\|^2 = r^2$ を満たす.

Assumption 3. 行列 Z の各要素は $\eta > 0$ に対して

$$\mathbb{E}[Z_{ij}^3] = 0, \quad \mathbb{E}\left[Z_{ij}^{8+\eta}\right] < \infty$$

を満たす.

Assumption 4. 分散共分散行列 Σ のスペクトルノルムは $M > 0$ により上から抑えられている, すなわち, 任意の n に対して $\|\Sigma\| \le M$.

Assumption 3 の前半の仮定は先行研究 [3] などではなかったものである. 本研究ではこれを理論的な理由で仮定している. これは対称な分布であれば大体の分布では満たされる仮定ではあるため, 本研究の成立する対象が極端に狭くなるわけではないと考えている. 以下では, $\omega = \mathbb{E}\left[\mathbb{1}_{\{\epsilon_1^2 \le d^2\}}\right]$ とする. これはどの程度打ち切るかを表す指標と考えられる. これらの仮定のもとで, 以下の out-of-sample リスクの極限的な性質を調べた.

$$R_X(\hat{\beta}) \coloneqq \mathbb{E}\left[(\hat{\beta}_d^\top x_0 - \beta^\top x_0)^2 \mid X\right]$$

ここで, x_0 は新しいサンプル点である.

3 漸近的な結果

以上の仮定のもとで, 以下が成立する.

Theorem 1. *Assumption 1, 2, 3, 4* が成立すると仮定する. このとき,

$$
\begin{aligned}
&R_X(\hat{\beta}_d) \\
&\to \lim_{\lambda\to 0}\lim_{n\to\infty}\left(1+m_1\left(\frac{\lambda}{\omega}\right)\right)\beta^\top\Sigma\left(m\left(\frac{\lambda}{\omega}\right)\Sigma+I_p\right)^{-2}\beta \\
&+\lim_{\lambda\to 0}\lim_{n\to\infty}\frac{1}{n}\frac{(1+m_1\left(\frac{\lambda}{\omega}\right))\mathrm{tr}\left(\Sigma^2(m\left(\frac{\lambda}{\omega}\right)\Sigma+I_p)^{-2}\right)}{\lambda^2\left(1-\frac{1}{n\lambda}\mathrm{tr}\left(\Sigma\left(m\left(\frac{\lambda}{\omega}\right)\Sigma+I_p\right)^{-1}\right)\right)^2} \\
&\qquad\qquad \mathbb{E}\left[\mathbb{1}_{\{\epsilon_1^2\le d^2\}}\epsilon_1^2\right], \qquad (2)
\end{aligned}
$$

が成立する. ここで, $\lambda<0$ に対して, $m(\lambda)>0$ は *Stieltjes* 変換で, $m(\lambda), m_1(\lambda)>0$ には以下の関係が成立している:

$$
\lambda=\frac{1}{m(-\lambda)}-\frac{\gamma}{p\omega}\mathrm{tr}\left[\Sigma\left(m(-\lambda)\Sigma+I_p\right)^{-1}\right],
$$

$$
0=-\frac{m_1(-\lambda)}{m(-\lambda)^2}+(1+m_1(-\lambda))\frac{\gamma}{p\omega}\mathrm{tr}\left[\Sigma^2(m(-\lambda)\Sigma+I_p)^{-2}\right].
$$

上記の定理では, 先に ridge regression バージョンの誤差項打ち切り推定量の漸近的なリスクを求めてから, $\lambda\to 0$ の極限を取ることにより求めたいものを求めている. より深く上記の Theorem 1 を理解するために $\Sigma=I$ を代入しより簡単にした結果として以下のものがある.

Theorem 2. *Assumption 1, 2, 3* が成立すると仮定する. さらに $\Sigma=I$ とする. このとき *out-of-sample* リスク $R_X(\hat{\beta}_d)$ は

$$
\begin{aligned}
&R_{\mathrm{lim}}(\gamma,r,d) \\
&=\begin{cases}\frac{\gamma}{\omega(\omega-\gamma)}\mathbb{E}\left[\mathbb{1}_{\{\epsilon_1^2\le d^2\}}\epsilon_1^2\right] & (\frac{\gamma}{\omega}<1)\\ \left(1-\frac{\omega}{\gamma}\right)r^2+\frac{1}{\gamma-\omega}\mathbb{E}\left[\mathbb{1}_{\{\epsilon_1^2\le d^2\}}\epsilon_1^2\right] & (\frac{\gamma}{\omega}>1)\end{cases}
\end{aligned}
$$

に収束する.

上記の定理から, $\omega\to 0$ で $R_{\mathrm{lim}}(\gamma,r,d)\to r^2$ となることから r^2 の値が小さい場合などにおいては ω はなるべく小さくとったほうが良いことがわかる.

4 数値実験

以下では Theorem 2 の状況下で行った二つの数値実験を示す. 以下の一つ目の図は ϵ の分布として標準正規分布を, 二つ目の図は標準コーシー分布を仮定したものである. 実線が理論結果に基づくもの, 点が数値実験の結果である. ただし, ここで X の各要素は標準正規分布に従っていると仮定し, $\gamma=0.6$ としている. また, r^2 は 0.5 から 2.5 まで変動させている. 数値実験の結果と理論曲線はよく合致している. また数値実験の結果, ω は基本的に小さくした方がいいことがわかる. また, コーシー分布の方が ω を小さくした際のリスクの減少率が高いことから, 裾野が厚い分布の方が ω を小さくした際の降下が大きいのではないかと考えられる.

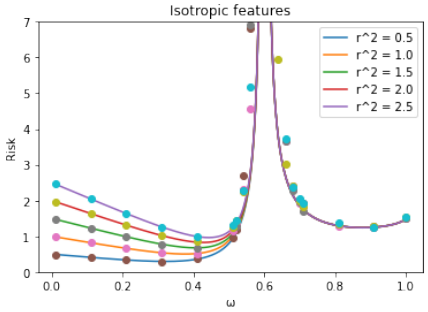

図 1. ϵ が正規分布に従っていると仮定したときの図. 実線は理論に基づくもので, 点は実験によるサンプル点である.

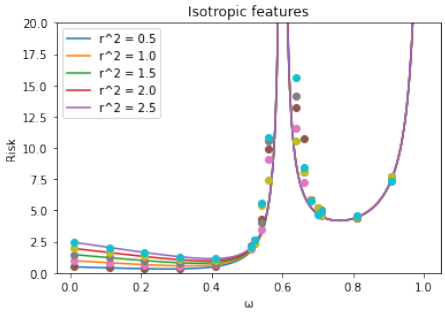

図 2. ϵ がコーシー分布に従っていると仮定したときの図. 実線は理論に基づくもので, 点は実験によるサンプル点である.

参考文献

[1] Mikhail Belkin, Alexander Rakhlin, and Alexandre B Tsybakov. Does data interpolation contradict statistical optimality? In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 1611–1619. PMLR, 2019.

[2] Peter Hall, James Stephen Marron, and Amnon Neeman. Geometric representation of high dimension, low sample size data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(3):427–444, 2005.

[3] Trevor Hastie, Andrea Montanari, Saharon Rosset, and Ryan J Tibshirani. Surprises in high-dimensional ridgeless least squares interpolation. *arXiv preprint arXiv:1903.08560*, 2019.