2023 年 5 月 26 日 統計数理研究所 オープンハウス

歪正規ノイズに基づくLasso回帰

小山 和輝　総合研究大学院大学 統計科学専攻 博士課程(3年次編入学) 4年

1　研究の概要

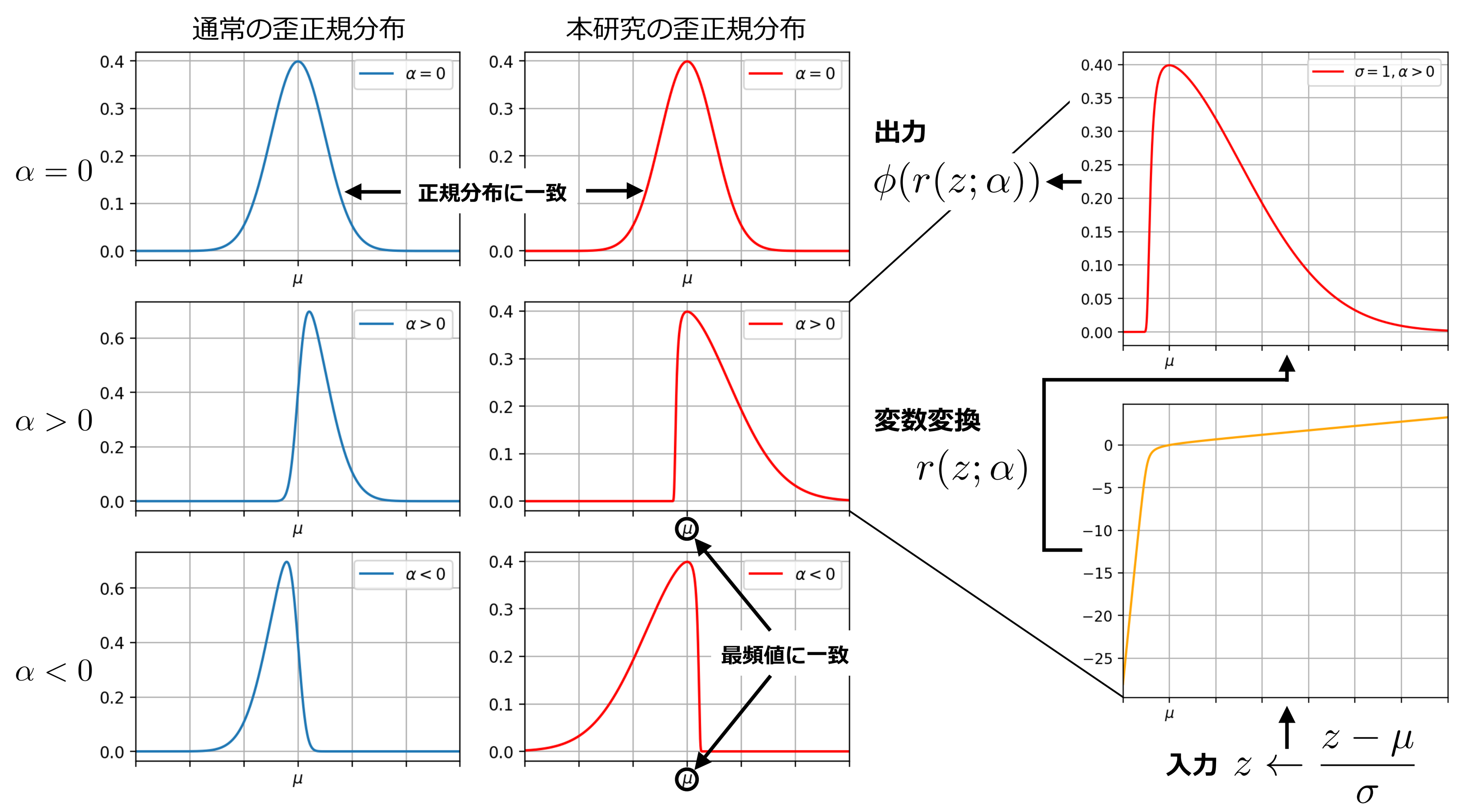
- ノイズが**歪正規分布**に従う線形回帰モデル $y = X\boldsymbol{\beta} + \epsilon$ を扱う
  - ➡ $X\boldsymbol{\beta}$ を分布の平均値や最頻値に回帰させたいが…
  - ➡ 一般的な歪正規分布では解析解が得られない [Azzalini, 2013]
- 本研究では**性質の良い歪正規分布** [Fujisawa and Abe, 2015] を用いる
  - ➡ 位置パラメータ μ が常に最頻値に一致する分布
  - ➡ $X\boldsymbol{\beta} \to \mu$ で最頻値への回帰が可能で回帰に違和感がない
- 今回は特にL1罰則をつけた**Lasso回帰**について考察する
  - ➡ Excess Risk 及び パラメータの推定誤差に関する理論を構築
  - ➡ 数値実験では回帰スコア＆変数選択性能の両面で良好な結果

2　提案手法

- 本研究の歪正規分布　[Jones, 2014; Fujisawa and Abe, 2015]

$$f(z;\mu,\sigma,\alpha) := \frac{1}{\sigma}\phi\left(r\left(\frac{z-\mu}{\sigma};\alpha\right)\right) \quad \to \mu \text{ が最頻値に一致}$$

ここで，ϕ は標準正規分布の確率密度関数，α は歪度を調整するパラメータ
また，変換 r は $s(w) := r^{-1}(w)$ として $s'(w) + s'(-w) = 2,\ r'(z) > 0$ を満たす



- 提案手法の定式化

$$\underset{\boldsymbol{\beta},\sigma,\alpha}{\operatorname{argmin}}\ \underbrace{\ell(\boldsymbol{\beta},\sigma,\alpha)}_{\text{誤差項}} + \lambda\|\boldsymbol{\beta}\|_1$$

ここで，$\boldsymbol{\beta} \in \mathbb{R}^P, \sigma \in \mathbb{R}, \alpha \in \mathbb{R}$ で
パラメータ空間と共変量は有界とする

$$\ell(\boldsymbol{\beta},\sigma,\alpha) := -\frac{1}{N}\log\prod_{n=1}^{N} f(y_n; X_n\boldsymbol{\beta},\sigma,\alpha)$$

$$f(y; X\boldsymbol{\beta},\sigma,\alpha) := \frac{1}{\sigma}\phi\left(r\left(\frac{y-X\boldsymbol{\beta}}{\sigma};\alpha\right)\right)$$

- 提案手法の最適化（非凸最適化問題）
  - ➡ $\boldsymbol{\beta}$ に関する最適化（σ,α 固定）
    誤差項は $\boldsymbol{\beta}$ に対して性質の良い凸関数
    $0 \preceq \nabla^2_{\boldsymbol{\beta}}\ell(\boldsymbol{\beta},\sigma,\alpha) \preceq C_{\sigma,\alpha}\Sigma$
    → 例えばMMアルゴリズムなどで効率的に最適化できる
  - ここで，$\Sigma := \frac{1}{N}X^\top X$ であり
    $C_{\sigma,\alpha}$ は σ と α に応じた定数
  - 反復
  - ➡ σ に関する最適化（$\alpha,\boldsymbol{\beta}$ 固定）と α に関する最適化（$\boldsymbol{\beta},\sigma$ 固定）
    高々1変数の最適化なので種々のソルバーを利用可能

3　推定誤差に関する理論

定理：推定誤差理論（提案手法の場合）

共変量 $X_1,\dots,X_N$ を固定し，スパース推定の一般的な仮定をおく
最適化では，ある定数 T と λ_0 を用いて正則化係数を $\lambda \geq 2T\lambda_0$ で選ぶ

このとき，（十分大きな確率で）以下の評価が成り立つ

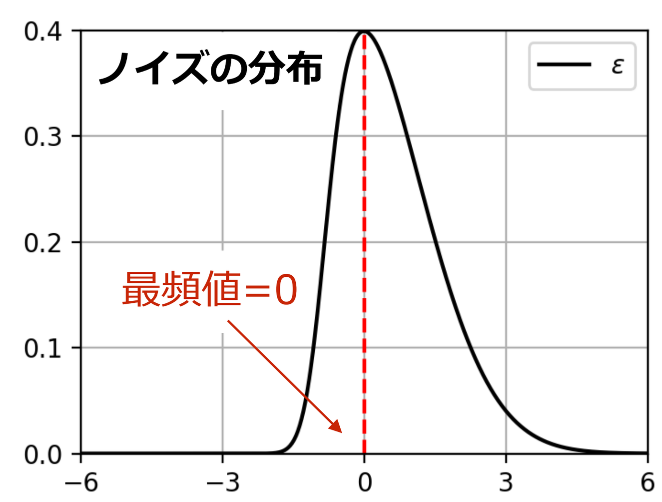
$$\bar{\mathscr{E}}(\hat f \mid f^0) + 2(\lambda - T\lambda_0)\left\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\right\|_1 \leq O\left(\kappa^2 s_0(\log(N))^4\log(P\vee N)/N\right)$$

ここで，λ_0 のレートは $\lambda_0 \asymp (\log(N))^2\sqrt{\dfrac{\log(P\vee N)}{N}}$ であり，
s_0 は真の非ゼロ係数の個数，κ は制限固有値条件に応じた定数である
$\bar{\mathscr{E}}$ は $f^0 := f(y; X^\top\boldsymbol{\beta}^0,\sigma^0,\alpha^0)$ に対する $\hat f := f(y; X^\top\hat{\boldsymbol{\beta}},\hat\sigma,\hat\alpha)$ のAverage Excess Riskで

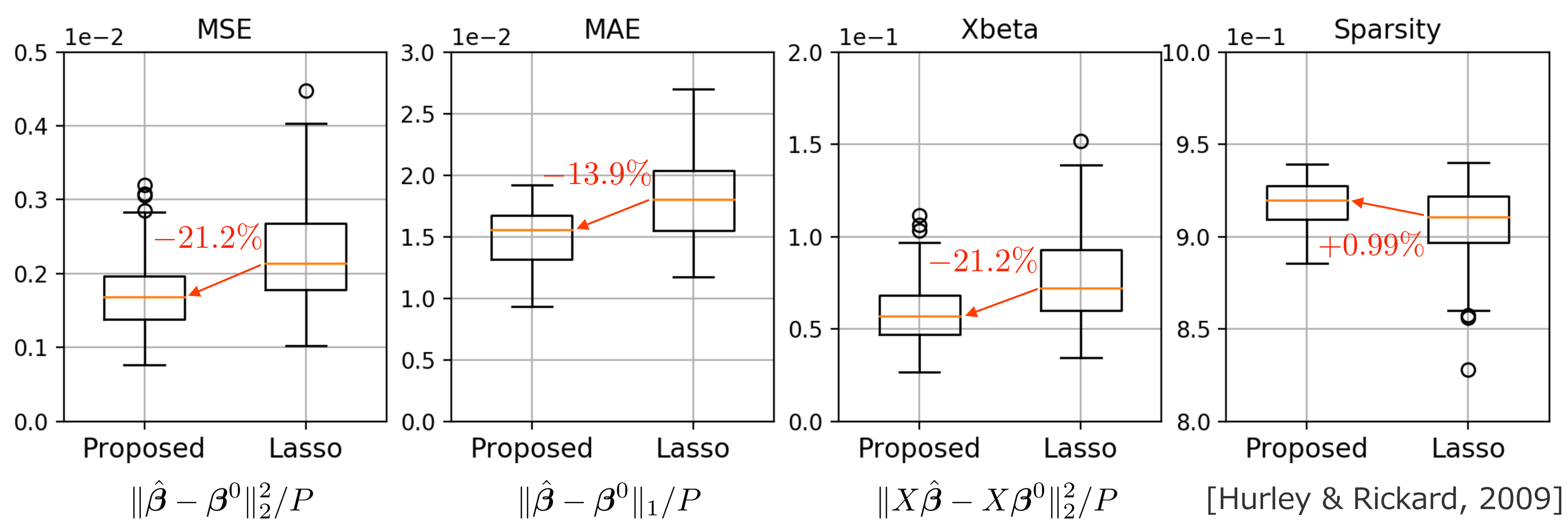
$$\bar{\mathscr{E}}(\hat f \mid f^0) := \frac{1}{N}\sum_{n=1}^{N} D_{\mathrm{KL}}(f(y_n; X_n^\top\boldsymbol{\beta}^0,\sigma^0,\alpha^0)\,\|\,f(y_n; X_n^\top\hat{\boldsymbol{\beta}},\hat\sigma,\hat\alpha)) \quad (D_{\mathrm{KL}}\text{はKL情報量})$$

4　数値実験

- データセット
  - ➡ スパースなパラメータを持つ線形回帰を考える
    説明変数：$X \in \mathbb{R}^{500\times 100},\ x_{np} \sim U(-1,1)$　（U：一様分布）
    目的変数：$\boldsymbol{y} \in \mathbb{R}^{500},\quad y_n = X_n\boldsymbol{\beta}^0 + \epsilon_n$
    パラメータ：$\boldsymbol{\beta}^0 = [1.0, -0.9, 0.8, \dots, -0.1, 0.0, \dots, 0.0]^\top \in \mathbb{R}^{100}$
    $\beta_p \neq 0$：10個　$\beta_p = 0$：90個
    ノイズ：$\epsilon_n \sim \phi(r(z;1))$　本研究の歪正規分布　$\sigma = 1, \alpha = 1$,（$\mu = 0$）に相当
  - ➡ 実験条件
    - 訓練データ（100セット）＋テストデータ
    - 比較手法は通常のLassoモデル
    - 正則化係数は5-分割交差検証で調整
      （CVの評価指標は尤度の最大化に基づく）



- 実験結果①：パラメータ $\boldsymbol{\beta}$ の推定精度
  - ➡ テストデータに対する評価（各100個）をボックスプロット



[Hurley & Rickard, 2009]

- 実験結果②：パラメータ $\boldsymbol{\beta}$ の変数選択性能
  - ➡ 事前に正規化 $\hat{\boldsymbol{\beta}}' := \hat{\boldsymbol{\beta}}/\|\hat{\boldsymbol{\beta}}\|_\infty$
  - ➡ 閾値 δ を基準に変数選択を行う
    $\mathbb{1}\{|\hat{\boldsymbol{\beta}}'| > \delta\}$ と $\mathbb{1}\{|\boldsymbol{\beta}^0| \neq 0\}$ の比較