

# スパースとロバストに関わる最近の研究

藤澤 洋徳 数理・推論研究系 教授 (ものづくりデータ科学研究センター副センター長・  
理化学研究所AIP客員研究員・名古屋大学大学院医学系研究科客員教授)

## はじめに

目的: スパースとロバストに関わる最近5年間の研究を紹介。

**スパース**: パラメータの多くが0であると想定. 高次元データ解析に対して有効な設定.

**ロバスト**: 外れ値による悪影響に頑健.

その他のキーワード: 転移学習. 欠測. 高相関. 因果探索. 因果推論. 敵対的汚染. グラフィカルモデル. 非対称分布. EMアルゴリズム. 多重代入.

**大学院生募集**: 数理統計・統計的機械学習・医療統計が主なテーマ. 研究テーマに合わせてある程度は広く対応可能.

**共同研究&コンサルテーション**: 製造業・製薬企業など.

詳しい情報は藤澤のHPをご覧ください.

## スパース

**転移Lasso (Transfer Lasso)**: 過去推定値や熟練工の知見などを  $L_1$  罰則を通して適応的に組み込むことが可能な転移学習. Takada, Fujisawa (2020) NeurIPS. **東芝との共同研究.**

$$\frac{1}{2n} \sum_{i=1}^n (y_i - x_i^\top \beta)^2 + \lambda \left\{ \alpha \|\beta\|_1 + (1 - \alpha) \|\beta - \beta^{(0)}\|_1 \right\}$$

**高欠測に対応するLasso (HMLasso)**: 高欠測データに対しても欠測補完なしに実行可能なLasso. Takada+ (2019) IJCAI. **東芝との共同研究.**

$$\hat{\Sigma} = \arg \min_{\Sigma \geq 0} \|R^\alpha \odot (S_{xx}^{\text{pair}} - \Sigma)\|_F \quad R = (r_{jk}) = (n_{jk}/n)$$

$$\hat{\beta} = \arg \min_{\beta} \frac{1}{2} \beta^\top \hat{\Sigma} \beta - S_{xy}^\top \beta + \lambda \|\beta\|_1$$

**高相関特徴量に対応するLasso (IILasso)**: 高相関特徴量がモデルに残るとモデルの解釈的に好ましくない. 高相関特徴量の同時選択を避けやすくするスパース・モデリング. Takada+ (2018, 2020) AISTATS, Neural Comput.

**LiNGAMに基づいたスパースな因果探索**: 因果探索において有名な LiNGAM を, スパース性の下で高次元データに対しても適用可能にした手法. Harada, Fujisawa (2021+) arXiv.

**スパース主成分回帰モデリング**: 主成分を特徴量とした回帰モデリングを, 通常の2段階法ではなく, 主成分分析と回帰モデリングをリンクした1段階スパース法. Kawano+ (2015, 2018) CSDA, CSDA. **遺伝研との共同研究.**

## ロバスト

**ロバストな因果推論**: 外れ値による悪影響に頑健な因果推論. Harada, Fujisawa (2021+). 論文執筆中.

**敵対的汚染に対するロバスト推定**: 通常の外れ値よりも強力な敵対的汚染が存在した場合におけるロバスト推定の収束レート. Sasai, Fujisawa (2020) arXiv

**外れ値の割合が特徴量に依存する場合のロバスト回帰モデリング**: 外れ値の割合が特徴量に依存するときでもバイアスを十分小さくすることが可能な回帰モデリング手法. Kawashima, Fujisawa (2018) arXiv

$$d_\gamma(p_{y|x}, q_{y|x}; p_x) = -\frac{1}{\gamma} \log \int \frac{\int p_{y|x}(y|x) q_{y|x}(y|x)^\gamma dy}{\int q_{y|x}(y|x)^{1+\gamma} dy} p_x(x) dx$$

## スパース&ロバスト

**スパース性とロバスト性を併せもつ回帰モデリング**: スパース性とロバスト性を同時に併せもち, かつ, 高速にパラメータ推定が可能な手法. Kawashima, Fujisawa (2017, 2019) Entropy, JISDS.

**スパース性とロバスト性を併せもつガウシアン・グラフィカル・モデリング**: 外れ値が一般の場合とセル毎の場合. Hirose+ (2017) J. Multivariate Anal. Katayama+ (2018) Stat.

## そのほか

**モード不変な非対称分布**: 代表的な非対称分布は歪正規分布である. ただし, 歪正規分布は位置パラメータが必ずしも位置を示さない欠点がある. その問題点を克服する非対称分布の構築. Abe, Fujisawa (2019) JISDS

**歪正規分布に対する最尤推定: モメンタム構造をもつEMアルゴリズム**: 歪正規分布の潜在変数表示における過剰パラメータ構造を積極的に利用して, モメンタム構造という加速性をもつパラメータ推定アルゴリズムの構築. Abe+ (2021+) Econometrics and Statistics.

$$\mathbf{X} = \text{sgn}(Y_0) \mathbf{Y} \quad (\mathbf{Y}^\top, Y_0)^\top \sim N_{p+1}(\mathbf{0}, \Sigma)$$

$$\Sigma = \begin{pmatrix} \Omega & \tau \Omega^{1/2} \boldsymbol{\delta} \\ \tau \boldsymbol{\delta}^\top \Omega^{1/2} & \tau^2 \end{pmatrix} \quad (\tau = 1 \text{ vs. } \tau: \text{free})$$

**多重代入法に対するバイアス補正推定**: 多重代入法はそのままではパラメータ推定にバイアスをもつため, そのバイアスを適応的に減少させる手法. Tomita+ (2018) Statist. Med.

## 最近の大学院生の研究テーマ

### 現学生

笹井 健行 (社会人学生: トヨタ): ロバスト推定

原田 和治: 因果探索・因果推論

### 修了生

高田 正彬 (東芝): スパース・モデリング

川島 孝行 (東京工業大学): ロバスト性とスパース性を併せもつ回帰モデリング

富田 裕章 (民間企業): 多重代入法