

スパース・ロバスト・歪分布に関わる最近の研究

藤澤 洋徳 統計基盤数理研究系 教授 (理研AIP客員研究員・名古屋大医客員教授)

はじめに

目的：スパースとロバストの最近5年間程度の研究紹介。

スパース：パラメータの多くの真値が0であると想定。高次元データ解析に対して有効。

$$y = x^T \beta^* + e, \quad \beta^* = (3, 1, 0, \dots, 0)^T$$

ロバスト：外れ値による悪影響に頑健。例：中央値。

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。異常検知。

大学院生募集：統計的推論・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。最近では社会人学生が多い。

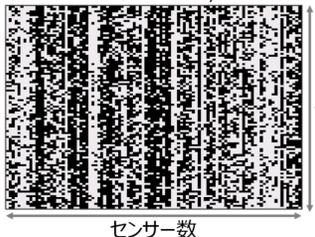
共同研究&コンサルテーション：製造業・製薬企業など。詳しい情報は藤澤のHPをご覧ください。

スパース

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) **東芝との共同研究**。

$$\hat{\Sigma} = \arg \min_{\Sigma \geq 0} \|R^\alpha \odot (S_{xx}^{\text{pair}} - \Sigma)\|_F$$
$$R = (r_{jk}) = (n_{jk}/n)$$
$$\hat{\beta} = \arg \min_{\beta} \frac{1}{2} \beta^T \hat{\Sigma} \beta - S_{xy}^T \beta + \lambda \|\beta\|_1$$

データのイメージ (黒：欠測, 白：観測)



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) **東芝との共同研究**。

過去情報 (m:多い) → 有効利用 → 現在情報 (n:多くない)

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

$\beta_j^{(0)}$: 初期推定値 (過去情報で構築)

適応的転移 Lasso (Adaptive Transfer Lasso)：Adaptive Lasso の考え方を組み込んだ Transfer Lasso。両方の良さを併せもつ。Takada, Fujisawa (2024) **東芝との共同研究**。

$$\frac{1}{2n} \sum_{i=1}^n (y_i - x_i^T \beta)^2 + \frac{\lambda}{n} \sum_j \frac{1}{|\beta_j^{(0)}|^{\gamma_1}} |\beta_j| + \frac{\eta}{n} \sum_j |\beta_j^{(0)}|^{\gamma_2} |\beta_j - \beta_j^{(0)}|$$

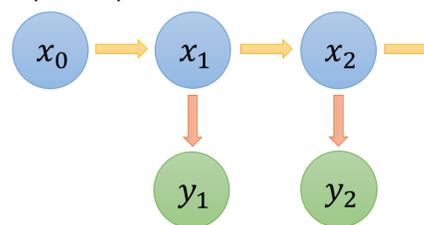
LINGAMに基づいたスパースな因果探索：因果探索において有名な LINGAM を、スパース性の下で高次元データに対しても適用可能にした手法。Harada, Fujisawa (2021)

ロバスト

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 (2023)

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

状態空間モデルに対するロバスト推定：理論的保証があり計算効率が高い、状態空間モデルに対するロバスト推定。Ishizuka, Fujisawa (2023)。



ロバストな因果推論：逆確率重み法や二重頑健推定において平均処置効果を、外れ値にロバストな推定を行う方法。Harada, Fujisawa (2024)。

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)

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) Katayama+ (2018)

敵対的汚染に対してロバストでスパースな推定：通常の外れ値よりも強力な敵対的汚染に対してロバストでスパースな推定の収束レート。Sasai, Fujisawa (2023)

歪分布

歪ノイズに対応可能なモード不変なスパース回帰モデリング：ノイズが左右対称でない場合の回帰モデリング。回帰モデルはモードであり解釈が容易。マイナス対数尤度関数は回帰パラメータに関して凸なので高次元特徴量でも対応しやすい。Koyama+ (2024)

$$f(y) = \frac{1}{\sigma} \phi \left(r_\alpha \left(\frac{y - x^T \beta}{\sigma} \right) \right) \quad \alpha: \text{歪度パラメータ}$$
$$r_\alpha: \text{特殊な関数}$$

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)