

2025年度科学研究費シンポジウム

大規模複雑データの理論と方法論の深化：最前線の動向と新たな展開

科学研究費補助金 基盤研究(A) 25H01107「大規模複雑データの理論と方法論の深化と展開」(研究代表者: 青嶋誠)によるシンポジウムを下記のように催しますので、ご案内申し上げます。

青嶋 誠 (筑波大学)
矢田和善 (筑波大学)
石井 晶 (東京理科大学)
江頭健斗 (東京理科大学)

記

日時: 2025年10月30日(木)~31日(金)

場所: イーアスホール (イーアスつくば内 <https://tsukuba.iias.jp/iiashall>)
〒305-0817 茨城県つくば市研究学園5丁目19番地

プログラム

10月30日(木)

10:45~10:50 開会

10:50~11:20 海野 哲也 (筑波大・数理物質科学研究群)

矢田 和善 (筑波大・数理物質系)

青嶋 誠 (筑波大・数理物質系)

高次元信号行列に対する自動スペース推定について

11:25~11:55 矢島 豪太 (日本大・生物資源科学研究科)

東出 大志 (石川県立大・生物資源環境学部)

中島 啓裕 (日本大・生物資源科学部)

深層学習モデルも間違える: 不完全な個体識別を伴う動物の密度推定

12:00~12:40 大田 浩史 (東京大・大学院総合文化研究科)

置換行列が作用した統計モデルの推測理論

12:40~14:00 昼食

14:00~14:40 仲北 祥悟 (東京大・大学院総合文化研究科)

今泉 允聰 (東京大・大学院総合文化研究科, RIKEN AIP)

生成拡散モデルの報酬によるアライメントに対する次元非依存汎化誤差評価

14:45~15:25 佃 康司 (九州大・大学院数理学研究院)

松浦 峻 (慶應義塾大・理工学部)

多変量 allometric 回帰モデルにおける共分散行列の第一固有ベクトル推定

15:30～16:10 寺田 吉壱 (大阪大・大学院基礎工学研究科)

行列分解因子分析の正体について

16:25～17:05 竹内 努 (名古屋大・理学研究科)

Application of high-dimensional statistical analysis to astrophysical problems

(天体物理学的問題における高次元統計解析の応用)

17:10～17:50 張 元宗 (日大・統計数理研究所)

篠崎 信雄 (慶應義塾大)

Shrinkage estimators improving for a wide class of restricted subsets

18:30～20:30 懇親会

10月31日(金)

10:45～11:15 浦崎 航 (東京理科大・創域理工学研究科)

二変量正規分布を潜在的にもつ分割表に対する連関尺度の提案

11:20～11:50 屋良 淳朝 (大阪大・大学院基礎工学研究科)

寺田 吉壱 (大阪大・大学院基礎工学研究科)

ノンパラメトリック・ロジスティック回帰の統計理論

11:55～12:35 中川 智之 (明星大・データサイエンス学環, RIKEN CBS)

鶴田 靖人 (明治大・経営学部)

文 翔 (東京理科大・創域理工学研究科)

田畠 耕治 (東京理科大・創域理工学部)

方向データにおけるロバストなベイズ推定

12:35～14:00 昼食

14:00～14:40 入江 薫 (東京大・大学院経済学研究科)

多変量モデルの頑健な事後分析

14:45～15:25 小池 祐太 (東京大・大学院数理科学研究科)

林 高樹 (慶應義塾大・大学院経営管理研究科)

塩谷 天章 (東京大・大学院数理科学研究科)

点過程に対するリード・ラグ解析

15:30～16:10 江頭 健斗 (東京理科大・創域理工学部)

矢田 和善 (筑波大・数理物質系)

青嶋 誠 (筑波大・数理物質系)

A cumulative sum-based change-point detection method for high-dimensional data

16:10～16:15 閉会

高次元信号行列に対する自動スパース推定について

筑波大学・数理物質科学研究群 海野 哲也

筑波大学・数理物質系 矢田 和善

筑波大学・数理物質系 青嶋 誠

1 はじめに

本講演では、高次元データに対する低ランク行列の復元問題について考えた。未知である $d \times n$ の低ランク行列 $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_n]$ に対し、そのランク数を r とする。今、 $d \times n$ のデータ行列 \mathbf{X} が以下で与えられているとする。

$$\mathbf{X} = \sqrt{n}\mathbf{A} + \mathbf{W}. \quad (1)$$

ここで、 $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_n]$ はノイズ行列であり、 $\mathbf{w}_1, \dots, \mathbf{w}_n$ は独立同分布に、 $E(\mathbf{w}_l) = \mathbf{0}$ かつ $Cov(\mathbf{w}_l) = \Sigma_W$ を持つ分布に従うとする。信号行列とも呼ばれる低ランク行列 \mathbf{A} の推定は、ゲノム科学や画像解析等、高次元データを扱う多くの分野において重要な技術であり、これまでにいくつかの先行研究がある。Shabalin and Nobel [1] は \mathbf{W} に正規性を仮定し、 $d/n \rightarrow c > 0$ の枠組みのもとで、ランダム行列理論に基づく \mathbf{A} の推定量を提案した。Yata and Aoshima [4] は \mathbf{W} に正規性を課さない場合における標本特異値・特異ベクトルの高次元漸近的性質を示し、特異値分解に基づく従来の推定量が高次元で巨大なノイズを内包することを示した。その解決策として、Yata and Aoshima [4] は Yata and Aoshima [3] で提案されたノイズ掃き出し法を特異値の推定に応用することで、 $d/n \rightarrow \infty$ となる高次元小標本設定においても高精度な特異値の推定を可能とした。しかしながら、高次元小標本設定において d 次元の特異ベクトルを高精度に推定する手法は、未だ確立されていない。

最近、Yata and Aoshima [5] は高次元主成分ベクトルの推定に対し、自動スパース推定と呼ばれる方法論を提案し、これが高次元小標本設定においても主成分ベクトルを高精度に推定できることを示した。さらに、Umino et al. [2] は自動スパース推定を高次元相互共分散行列の推定に応用した。本講演では、高次元特異ベクトルの推定に自動スパース推定を応用することで、次元数が標本数を凌駕する状況においても高精度な信号行列の再構築手法を提案した。

2 高次元特異値・特異値ベクトルの一貫性

\mathbf{A} の特異値分解を $\mathbf{A} = \sum_{i=1}^r \lambda_{i(A)}^{1/2} \mathbf{u}_{i(A)} \mathbf{v}_{i(A)}^T$ とする。ただし、 $\lambda_{1(A)}^{1/2} \geq \dots \geq \lambda_{r(A)}^{1/2} (> 0)$ は \mathbf{A} の特異値であり、 $\mathbf{u}_{i(A)} \in \mathbf{R}^d$ 、 $\mathbf{v}_{i(A)} \in \mathbf{R}^n$ はそれぞれ $\lambda_{i(A)}^{1/2}$ に対応する左特異ベクトル、右特異ベクトルである。また、 \mathbf{X} も $\mathbf{X}/\sqrt{n} = \sum_{i=1}^{\min\{d,n\}} \hat{\lambda}_i^{1/2} \hat{\mathbf{u}}_i \hat{\mathbf{v}}_i^T$ と特異値分解する。 $\|\cdot\|_F$ をフロベニウスノルムとし、 $d \times n$ 行列 $\hat{\mathbf{A}}$ に対する損失を $L(\hat{\mathbf{A}}|\mathbf{A}) = \|\hat{\mathbf{A}} - \mathbf{A}\|_F^2$ で定める。このとき、 \mathbf{X} の特異値分解に基づく \mathbf{A} の推定量 $\hat{\mathbf{A}}_r = \sum_{i=1}^r \hat{\lambda}_i^{1/2} \hat{\mathbf{u}}_i \hat{\mathbf{v}}_i^T$ に対し、Yata and Aoshima [4] は適当な

正則条件のもと,

$$\begin{aligned}\hat{\lambda}_i/\lambda_{i(A)} &= 1 + \kappa_i + o_p(1), \quad \|\hat{\mathbf{u}}_i - \mathbf{u}_{i(A)}\|^2 = 2\{1 - (1 + \kappa_i)^{-1/2}\} + o_p(1), \quad \text{and} \\ L(\hat{\mathbf{A}}_r | \mathbf{A}) &= r \text{tr}(\Sigma_W)/n + o_p(\lambda_{r(A)}) \quad \text{as } d, n \rightarrow \infty.\end{aligned}$$

が成り立つことを示した。ただし, $\|\cdot\|$ はユークリッドノルムであり, $\kappa_i = \text{tr}(\Sigma_W)/(n\lambda_{i(A)})$ である。この結果から, $d/n \rightarrow \infty$ なる高次元小標本設定において, $\hat{\lambda}_i$, $\hat{\mathbf{u}}_i$ を用いて $\lambda_{i(A)}$, $\mathbf{u}_{i(A)}$ を推定することは困難となり, $\hat{\mathbf{A}}_r$ は巨大なノイズを内包することがわかる。

一方, $\lambda_{i(A)}$, $\mathbf{u}_{i(A)}$ の推定にノイズ書き出し法を適用して得られる推定量 $\tilde{\lambda}_{i(r)}$, $\tilde{\mathbf{u}}_{i(r)}$ に対し, 適当な正則条件のもと以下が成立する。

$$\tilde{\lambda}_{i(r)}/\lambda_{i(A)} = 1 + o_p(1) \quad \text{and} \quad \|\tilde{\mathbf{u}}_{i(r)} - \mathbf{u}_{i(A)}\|^2 = \kappa_i\{1 + o_p(1)\} + o_p(1) \quad \text{as } d, n \rightarrow \infty.$$

したがって $\tilde{\lambda}_{i(r)}$ は $\lambda_{i(A)}$ の精度良い推定量となるものの, $\tilde{\mathbf{u}}_{i(r)}$ のノルム損失は依然として κ_i に依存することがわかる。Yata and Aoshima [4] は $\tilde{\lambda}_{i(r)}$ を用いた \mathbf{A} の推定量 $\tilde{\mathbf{A}}_r = \sum_{i=1}^r \tilde{\lambda}_{i(r)}^{1/2} \hat{\mathbf{u}}_i \hat{\mathbf{v}}_i^T$ を考案し, 漸近的に $L(\tilde{\mathbf{A}}_r | \mathbf{A}) \leq L(\hat{\mathbf{A}}_r | \mathbf{A})$ となることを示したものの, $L(\tilde{\mathbf{A}}_r | \mathbf{A}) = o_p(\lambda_{r(A)})$ の意味での一致性を有するには依然として $\lim_{d,n \rightarrow \infty} \kappa_r = o(1)$ なる条件が必要であった。

本講演では, $\tilde{\mathbf{u}}_{i(r)}$ が高次元空間で有する幾何学的性質に着目することで, $\lim_{d,n \rightarrow \infty} \kappa_r = o(1)$ の条件無しに

$$\|\tilde{\mathbf{u}}_{i(r)} - \mathbf{u}_{i(A)}\|^2 = o_p(1)$$

の意味での一致性を有する高次元特異ベクトルの推定量 $\tilde{\mathbf{u}}_{i(r)}$ を提案すると共に, この一致推定に基づく新たな信号行列の再構築手法を提案し, 理論的かつ数値的に既存手法と比較した。さらに, \mathbf{A} のランク数 r を推定するための方法論も与えた上で, その性能を実際の高次元データを用いて確認した。

参考文献

- [1] Shabalin, A. A. and Nobel, A. B. (2013). Reconstruction of a low-rank matrix in the presence of Gaussian noise. *Journal of Multivariate Analysis*, 118:67–76.
- [2] Umino, T., Yata, K., and Aoshima, M. (2025). Automatic sparse estimation of the high-dimensional cross-covariance matrix. *Revised in Journal of Multivariate Analysis*.
- [3] Yata, K. and Aoshima, M. (2012). Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. *Journal of Multivariate Analysis*, 105(1):193–215.
- [4] Yata, K. and Aoshima, M. (2016). Reconstruction of a high-dimensional low-rank matrix. *Electronic Journal of Statistics*, 10(1):895–917.
- [5] Yata, K. and Aoshima, M. (2025). Automatic sparse pca for high-dimensional data. *Statistica Sinica*, 35:1069–1090.

深層学習モデルも間違える：

不完全な個体識別を伴う動物の密度推定

矢島豪太（日本大学） 東出大志（石川県立大学） 中島啓裕（日本大学）

1. はじめに

野生動物の個体数密度推定は、生態学・保全生物学で重要な課題の一つである。特に空間明示型捕獲再捕獲（SECR）モデルは、空間情報と検出過程を統合して信頼性の高い密度推定を可能にする手法として広く用いられている。ただし、個体識別が前提となるため、カメラトラップ画像を用いる場合は識別作業が大きな負担となる。

近年、個体識別の自動化に畳み込みニューラルネットワーク（CNN）による深層距離学習が活用されている（Vidal et al. 2021）。この手法は、画像から個体固有の特徴を多次元ベクトル（画像特徴量）として抽出し、同一個体間の類似度を最大化するように学習する。海域から陸域まで様々な動物種へ適用が進む一方で、深層学習の識別結果には不確実性があり、これを無視すると SECR の推定にバイアスが生じる。従って、その不確実性を統計的に取り込む枠組みが求められる。

この点で参考となるのが Unmarked SECR モデル（Chandler and Royle 2013）で、個体識別の不確実性を確率的に扱い、空間的制約を利用して個体数や活動中心を推定する。本研究ではこの発想を発展させ、画像特徴量を確率的なシグナルとして扱う新たな SECR モデルを開発した。特徴量の変動をパラメータ化し、空間情報と統合して個体帰属確率を推定する構造とした。実データとしてツキノワグマ画像に深層距離学習を適用し、モンテカルロシミュレーションで本モデルの性能を検証した。

2. 統計モデル

本モデルは、1) トラップと個体の空間情報を考慮した撮影プロセスと、2) 深層学習による画像特徴量の観測プロセスを階層ベイズモデルとして統合した。

2.1. 個体の撮影プロセス

調査領域 $A \in \mathbb{R}^2$ において J 箇所のカメラトラップを設置し、個体 $i = 1, \dots, N$ の活動中心 \mathbf{s}_i が A 内に潜在変数として一様分布すると仮定した。各トラップ位置 $\mathbf{x}_j, j = 1, \dots, J$ における個体 i の検出回数 $y_{i,j}$ （捕獲履歴）の期待値を、半正規関数として以下のように定義した：

$$\lambda(\mathbf{s}_i, \mathbf{x}_j) = \lambda_0 \exp\left(-\frac{\|\mathbf{s}_i - \mathbf{x}_j\|_2^2}{2\sigma_\lambda^2}\right)$$

ここで λ_0 は、 $\|\mathbf{s}_i - \mathbf{x}_j\|_2^2$ が 0 の時の平均検出回数、 σ_λ はスケールパラメータである。

2.2. 特徴量の観測プロセス

L 次元特徴空間 E 上に一様分布するクラスター中心 $\mathbf{g}_i^{true} \in \mathbb{R}^L$ を導入し、 \mathbf{g}_i^{true} から観測特徴量 $g_{m,l}^{obs}$ が不確実性を伴い生成されるものとして表現した。サンプル m に対応する個体の ID を $ID_m \in \{1, \dots, N\}$ として、観測される特徴量 $g_{m,l}^{obs}$ の生成過程を以下のように記述した：

$$g_{m,l}^{obs} | ID_m = i \sim \text{Normal}(g_{i,l}^{true}, \sigma_g^2)$$

ここで σ_g は同一個体の異なる画像から得られる特徴量の変動を表し、深層学習モデルの識別性能と撮影条件の変動を統合した不確実性パラメータである。

3. 推論アルゴリズム

我々は確定的な個体識別情報が得られないことを想定するため、 $y_{i,j}$ を潜在変数として扱った。そこで、マルコフ連鎖モンテカルロ（MCMC）法によるデータ拡大法によって、 $y_{i,j}$ のサンプリングと、 $y_{i,j}$ が既知のもとでの他のパラメータの更新を逐次的に行つた。

3.1. 空間情報のみに基づく捕獲履歴の構成

捕獲履歴 $y_{i,j}$ は、個体数 (i 方向の次元数 N) と各セルの値の 2 つが欠測するため、これらを補完する必要がある。1 つ目については、 N より十分大きい M を既知の定数として置き換え、潜在変数 $z_i \sim \text{Bernoulli}(\psi)$ を導入した (すなわち $N = \sum_i z_i$)。この二項指示変数により個体数の次元を M で固定した。2 つ目については、ポアソン分布を和で条件つけると多項分布になる性質を利用した。具体的には、 $y_{i,j}$ が未知であるが、トラップ j における総観測回数 $n_j = \sum_i y_{i,j}$ は既知であるため、以下の関係性が成り立つ:

$$\{y_{1,j}, \dots, y_{M,j}\} \sim \text{Multinomial}(n_j, \{\pi_{1,j}, \dots, \pi_{M,j}\}), \pi_{i,j} \propto \lambda(\mathbf{s}_i, \mathbf{x}_j) z_i$$

この多項分布によって $y_{i,j}$ のサンプリングが可能となる。

3.2. 画像特徴空間への拡張

上記のサンプリング式に画像特徴量を組み込むために、捕獲履歴 $y_{i,j}$ をイベント単位に分解し、多項分布をカテゴリカル分布に置き換えた。同様に、構成対象も $y_{i,j}$ から ID_m に置き換えた。イベント単位で定義される $\pi_{i,m}$ は、サンプル m と個体 i の空間的な近さと特徴量中心への近さを考慮して以下のように定義した:

$$\pi_{i,m} \propto \lambda(\mathbf{s}_i, \mathbf{x}_{j(m)}) z_i d(\mathbf{g}_i^{true}, \mathbf{g}_m^{obs})$$

ここで、 $j(m)$ はサンプル m の観測トラップを示し、 $d(\mathbf{g}_i^{true}, \mathbf{g}_m^{obs})$ は、正規分布の確率密度を用いて、 $\prod_{l=1}^L \text{Normal}(g_{m,l}^{obs}; g_{i,l}^{true}, \sigma_g^2)$ とした。この確率密度が高い場合は、 $g_{m,l}^{obs}$ に対する $g_{i,l}^{true}$ の割り当てが適切であることを示し、それは $g_{m,l}^{obs}$ と $g_{i,l}^{true}$ の距離が近いことを示す。MCMC の各イテレーションで、観測データを条件として、潜在変数 ID_m をサンプリングし、これに基づいて捕獲履歴 $y_{i,j}$ を構築した。その他のパラメータについては、 $y_{i,j}$ を既知の変数として扱い、モンテカルロ・ヘイスティング法によってサンプリングした。

当日の発表では、モデル及び推論アルゴリズムの詳細とシミュレーション結果を報告した。

4. 引用文献

- Chandler, R. B., and J. A. Royle. 2013. Spatially Explicit Models for Inference about Density in Unmarked or Partially Marked Populations. *The Annals of Applied Statistics* 7: 936–954.
- Vidal, M., N. Wolf, B. Rosenberg, B. P. Harris, and A. Mathis. 2021. Perspectives on Individual Animal Identification from Biology and Computer Vision. *Integrative and Comparative Biology* 61:900–916.

INFERENCE ON PERMUTED STATISTICAL MODELS

HIROFUMI OTA

We study two fundamental statistical models in which permutation matrices act on the underlying data structures, producing irregular and highly nonconvex inference problems. The first target is permuted linear regression, where responses and covariates are mismatched at the row level for a small subset of observations. The second is graph matching under additive Gaussian noise, where two weighted graphs are aligned only up to an unknown vertex permutation. Both problems share a common feature: the parameter space of possible permutations is discrete and combinatorial, and its complexity grows extremely fast as the sample size increases. The discrete and combinatorial nature makes standard asymptotic inference difficult and motivates the development of computationally efficient algorithms to reduce the combinatorial burden inherent in related optimization methods.

To address these challenges, we develop a unified finite-sample inference framework based on the repro samples method. The key idea is to construct a data-dependent candidate set of permutations that contains the true alignment with high probability. Within this localized candidate set, pivotal distributions can be used for testing and constructing confidence sets, avoiding the need for complicated conditional distributions or asymptotic approximations as utilized in the standard selective/post-selection inference methodology. This approach allows us to perform valid finite-sample inference even when the underlying model involves such discrete structures.

In the permuted linear regression model, we study the case where the correspondence between covariates and responses is partially broken. Even when only a small fraction of data points are mismatched, the number of possible label configurations becomes enormous. Our framework first constructs a data-driven candidate set of sparse permutations that is guaranteed to include the true permutation with high probability. Based on this set, we develop a finite-sample valid hypothesis test for detecting the presence of mismatches, as well as a post-selection confidence set for the regression coefficients that maintains nominal coverage levels.

From a computational perspective, we propose a score-weighted Hungarian algorithm, a scalable surrogate that approximates the combinatorial objective efficiently while retaining theoretical guarantees of equivalence under mild conditions. The framework is further extended to handle partially permuted designs and ridge-regularized regressions, both of which preserve the same finite-sample validity.

Simulation studies confirm that our procedures achieve accurate coverage and stable performance across a wide range of sample sizes and noise levels. These results provide the first comprehensive inferential framework in the permuted linear regression models, especially for permutation matrices.

For the graph matching, we consider the weighted graphs that share a common latent structure but are observed with additive Gaussian noise and possible correlation between the two observations. The inferential goal is a finite-sample confidence set for the unknown vertex permutation that aligns the graphs, motivated by reliability and interpretability needs in applications such as neural connectivity analysis, social-network de-anonymization, and comparative genomics. The main obstacles are also the discrete, combinatorial parameter space and the lack of reasonable asymptotics for permutation-valued parameters.

To make inference on permutation structures, we construct a pivotal statistic based on the geometric angle between the edge-space difference at a candidate alignment and an independent Gaussian direction. At the true alignment, this statistic has an exact Beta-law that is free of nuisance parameters (e.g., noise scales or correlation). Applying a single independent repro direction to the candidate set yields a union confidence set for the permutation with finite-sample coverage guarantees.

On computation, solving the exact selection objective over all permutations is NP-hard due to the nature of quadratic assignment problems so that the Monte Carlo approximation of the test statistic is highly infeasible. To address the limitation, we propose the novel and efficient algorithm named the “repro-assisted first order method” that approximates the original objective function by the mean-field alternatives, reducing the quadratic assignment to a linear assignment per each Gaussian random projection.

Both parts are joint works with Masaaki Imaizumi (University of Tokyo / RIKEN).

(H. Ota) GRADUATE SCHOOL OF ARTS AND SCIENCES, THE UNIVERSITY OF TOKYO, JAPAN.
Email address: hirofumi-ota@g.ecc.u-tokyo.ac.jp

生成拡散モデルの報酬によるアライメントに対する 次元非依存汎化誤差評価

仲北 祥悟^{*}, 今泉 允聰[†]

概要

報酬に基づくファインチューニングは、新たな課題や目的に対する事前学習済み拡散モデルのアライメントを実現する上で確立された戦略であり、例えばテキストから画像への生成品質の向上や生成画像の審美性の強化といった目的に用いられてきた。従来の多くの研究では、ファインチューニングにおける任意の仮説に対する期待報酬が利用可能であると仮定してきた。この仮定はプロンプト分布が既知であることを前提としているが、実際にはプロンプト分布は通常未知であるため、期待報酬の代替として経験報酬が用いられる。しかし、この代替がいつ、どのように統計的に正当化されるのかについては明らかとなっていない。

この問題に対処するために、条件付き拡散モデルに対する報酬誘導型アライメントについて統計的理論保証を確立する。具体的には、(i) 状態空間およびプロンプト空間の次元、(ii) 事前学習済みモデルおよびファインチューニング仮説の構造、(iii) 陽でない定数のいずれにも依存しない汎化誤差の集中上界を導出する。これらの上界を得るために、新たなカップリング PAC-Bayes 上界と拡散モデルのタイトなカップリングを導入する。さらに、重要な視点の転換として、事前学習済み拡散モデルを事前分布とみなし、ファインチューニング後のモデルを対応する事後分布とみなすことで、PAC-Bayes の議論から所望の上界を導出する。

1 導入

拡散モデルのアライメントは近年大きな関心を集めている。その目的は多岐にわたり、生成画像の美的品質の向上、人間の選好への適合、テキストから画像への生成品質の改善などが含まれる。特に、報酬誘導型ファインチューニング (reward-guided fine tuning, RFT) は、このような問題に取り組むための確立された手法となっている (Clark et al., 2024; Uehara et al., 2024; Han et al., 2025; Kawata et al., 2025)。RFT の利点の 1 つはその柔軟性であり、報酬関数を適切に設計することで、拡散モデルを多様な目的や文脈に合わせて調整することができる。

拡散モデルの報酬誘導型ファインチューニングに関する従来の研究の多くは、ファインチューニングでの各仮説に対して、プロンプトとファインチューニング後の拡散モデルの両方に対する報酬の期待値が利用可能であると仮定している。具体的には、次の期待報酬関数が既知であると想定している。

$$\mathcal{R}(\gamma) = \mathbb{E}_c \left[\mathbb{E}_{x(c)}^\gamma [r(x_0^\gamma(c), c)] \right].$$

ここで、 γ はファインチューニングの仮説を表し、 c はプロンプトを表す確率変数、 $x_0^\gamma(c)$ は与えられた c に対するファインチューニング後の拡散モデルの出力、 r は c を条件とした $x_0^\gamma(c)$ に割り当てられる報酬を表す。しかし、実際にはこの関数は通常未知である。その主因の一つはプロンプト分布ならびにその分布に対する期待値が通常アприオリには知られていない点にある。ファインチューニングの目的は新しいタスクへのアライメントであるにもかかわらず、プロンプト分布が既知であるとする仮定は過剰に制約的である。さらにテキスト埋め込みのようにプロンプ

^{*}東京大学. nakakita@g.ecc.u-tokyo.ac.jp

[†]東京大学, 理化学研究所. imaizumi@g.ecc.u-tokyo.ac.jp

トが連続値であるならば、期待値評価は計算困難である。この未知性に対処するために、Uehara et al. (2024) および Han et al. (2025) が指摘するように、プロンプトサンプリングによる期待報酬の経験近似が広く用いられる。しかし、そのような経験近似に対する統計的保証は欠如しており、依然として重要な未解決問題である。いくつかの研究はそのような保証を与えていたが (Tian and Shen, 2024; Cheng et al., 2025; Mou, 2025)，それらの解析は特定の報酬関数に依存する、あるいは高次元設定において緩い上界となるなど、一般の報酬関数における高次元設定でも有効な統計原理はまだ明らかでない。

1.1 主たる貢献

本研究は、プロンプトサンプリングに関する拡散モデルの報酬誘導型ファインチューニングに対する基本的な統計原理を確立する。特に、得られる汎化誤差上界はファインチューニングの仮説の ℓ^2 ノルムによって特徴づけられる。結果として、この上界は次元非依存かつアーキテクチャ非依存であり、さらに上界に現れる定数はすべて明示的かつ小さい。特に、プロンプト空間の次元からの独立性は自明ではなく、直感に反する性質である。これらの性質は、より良い汎化性能を得るためににはファインチューニングの ℓ^2 ノルムのみを考慮すればよいことを示している。さらに、本研究は理論的なものであるが、実際的なファインチューニング手法に対しても示唆を与える。すなわち、ファインチューニング後の拡散モデルにおける破滅的忘却 (catastrophic forgetting) ならびに過学習を緩和するために提案してきたエントロピー正則化 (Uehara et al., 2024; Tang, 2024) は、汎化性能を確かに改善することがわかる。

謝辞

本研究は JSPS 科研費 JP24K02904, JST CREST JPMJCR21D2, JST さきがけ JPMJPR24K7 の助成を受けたものです。

参考文献

- Cheng, Z., Xie, T., Zhang, S., and Zhang, C. (2025). Provable Sample-Efficient Transfer Learning Conditional Diffusion Models via Representation Learning. *arXiv preprint arXiv:2502.04491*.
- Clark, K., Vicol, P., Swersky, K., and Fleet, D. J. (2024). Directly Fine-Tuning Diffusion Models on Differentiable Rewards. In *The Twelfth International Conference on Learning Representations*.
- Han, Y., Razaviyayn, M., and Xu, R. (2025). Stochastic Control for Fine-tuning Diffusion Models: Optimality, Regularity, and Convergence. In *Forty-second International Conference on Machine Learning*.
- Kawata, R., Oko, K., Nitanda, A., and Suzuki, T. (2025). Direct Distributional Optimization for Provable Alignment of Diffusion Models. In *The Thirteenth International Conference on Learning Representations*.
- Mou, W. (2025). Is RL fine-tuning harder than regression? A PDE learning approach for diffusion models. *arXiv preprint arXiv:2509.02528*.
- Tang, W. (2024). Fine-tuning of diffusion models via stochastic control: entropy regularization and beyond. *arXiv preprint arXiv:2403.06279*.
- Tian, X. and Shen, X. (2024). Enhancing Accuracy in Generative Models via Knowledge Transfer. *arXiv preprint arXiv:2405.16837*.
- Uehara, M., Zhao, Y., Black, K., Hajiramezanali, E., Scalia, G., Diamant, N. L., Tseng, A. M., Biancalani, T., and Levine, S. (2024). Fine-Tuning of Continuous-Time Diffusion Models as Entropy-Regularized Control. *arXiv preprint arXiv:2402.15194*.

多変量 allometric 回帰モデルにおける共分散行列の第一固有ベクトル推定

佃 康司（九州大学），松浦 峻（慶應義塾大学）

独立な n 個体における p -次元変量の観測 $\mathbf{y}_1, \dots, \mathbf{y}_n$ について，多変量重回帰モデル

$$\mathbf{y}_i = \boldsymbol{\mu} + \mathbf{B}^\top \mathbf{x}_i + \mathbf{e}_i, \quad \mathbf{e}_i \sim \mathcal{N}_p(\mathbf{0}_p, \boldsymbol{\Sigma}), \quad i \in \{1, \dots, n\}$$

において， $\boldsymbol{\alpha}$ をある $q \times 1$ ベクトルとして $\mathbf{B} = \boldsymbol{\alpha}\boldsymbol{\gamma}_1^\top$ を想定する allometric 回帰モデル

$$\mathbf{y}_i = \boldsymbol{\mu} + (\boldsymbol{\alpha}^\top \mathbf{x}_i)\boldsymbol{\gamma}_1 + \mathbf{e}_i, \quad \mathbf{e}_i \sim \mathcal{N}_p(\mathbf{0}_p, \boldsymbol{\Sigma}), \quad i \in \{1, \dots, n\} \quad (1)$$

を議論した。ここで， \mathbf{x}_i は個体 $i \in \{1, \dots, n\}$ において観測される中心化された説明変数を表す $q \times 1$ 定数ベクトル（中心化されていることより $\sum_{i=1}^n \mathbf{x}_i = \mathbf{0}_q$ ）。ただし， $q < n - 1$ とし，必要に応じて別の仮定を行う。 $p \times 1$ 定数ベクトルである $\boldsymbol{\mu}$ ， $q \times p$ 行列である \mathbf{B} ， $p \times p$ 正定値行列である $\boldsymbol{\Sigma}$ は未知のパラメータである。 $\lambda_1, \dots, \lambda_p$ を降順に並べた $\boldsymbol{\Sigma}$ の固有値とし， $\lambda_1 > \lambda_2$ を仮定する。また，各 $i \in \{1, \dots, p\}$ について $\boldsymbol{\gamma}_i$ を λ_i に対応する $\boldsymbol{\Sigma}$ の長さ 1 の固有ベクトルとする。Allometric 回帰モデルは，特に相違なる $i, j \in \{1, \dots, n\}$ の組について $E[\mathbf{y}_i] - E[\mathbf{y}_j] = \{\boldsymbol{\alpha}^\top (\mathbf{x}_i - \mathbf{x}_j)\}\boldsymbol{\gamma}_1$ が $\boldsymbol{\Sigma}$ の第一固有ベクトル $\boldsymbol{\gamma}_1$ の定数倍となるランク 1 の回帰モデルであり，Tarpey and Ivey [1] で提案された。 $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)^\top$ ， $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ ， $\mathbf{E} = (\mathbf{e}_1, \dots, \mathbf{e}_n)^\top$ ， $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p)$ ， $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_p)$ とおく。このとき， $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^\top$ とかけ，(1) を

$$\mathbf{Y} = \mathbf{1}_n \boldsymbol{\mu}^\top + \mathbf{X} \boldsymbol{\alpha} \boldsymbol{\gamma}_1^\top + \mathbf{E}, \quad \text{vec}(\mathbf{E}) \sim \mathcal{N}_{np}(\mathbf{0}_{np}, \boldsymbol{\Gamma} \boldsymbol{\Lambda} \boldsymbol{\Gamma}^\top \otimes \mathbf{I}_n)$$

とかける。以降， $\mathbf{X}^\top \mathbf{X}$ が正則とする。 $\hat{\mathbf{B}}_{\text{OLS}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$ として，

$$\mathbf{S}_T = \mathbf{Y}^\top (\mathbf{I}_n - n^{-1} \mathbf{1}_n \mathbf{1}_n^\top) \mathbf{Y}, \quad \mathbf{S}_R = \hat{\mathbf{B}}_{\text{OLS}}^\top \mathbf{X}^\top \mathbf{X} \hat{\mathbf{B}}_{\text{OLS}}, \quad \mathbf{S}_E = \mathbf{S}_T - \mathbf{S}_R$$

とおく。

講演では，Tsukuda and Matsuura [2] にもとづいて，allometric 回帰モデル (1) において重要な量である $\boldsymbol{\gamma}_1$ の推定を議論した。具体的には次の内容を報告した。 $w \in [0, 1]$ について， $\hat{\boldsymbol{\gamma}}_1(w)$ を

$$\mathbf{S}(w) = (1 - w) \mathbf{S}_R + w \mathbf{S}_E$$

の長さ 1 の第一固有ベクトルとして（方向は適当にとる），推定量のクラス $\{\hat{\boldsymbol{\gamma}}_1(w)\}_{w \in [0, 1]}$ を考えた。固定した w について $\hat{\boldsymbol{\gamma}}_1(w)$ の MSE(up to sign) の非漸近的上界を特定された形で与え，上界を最小化する $w = w^*$ を求めたうえでその推定量 \hat{w}^* を用いる $\hat{\boldsymbol{\gamma}}_1(\hat{w}^*)$ を新たに提案した。いくつかの漸近レジームのもとで推定量の一致性を議論した。数値実験により各推定量を比較した。また，橭円対称分布を仮定した MSE(up to sign) の非漸近的上界に関する議論も一部行った。

Acknowledgments

This study was supported in part by Japan Society for the Promotion of Science KAKENHI Grant Numbers 20K11713, 21K13836, and 25K07133.

参考文献

- [1] Tarpey, T., Ivey, C. T. (2006). Allometric extension for multivariate regression. *Journal of Data Science* **4**, no.4, 479–495.
- [2] Tsukuda, K., Matsuura, S. (2025). Estimators for multivariate allometric regression model. *J. Multivariate Anal.* **210**, 105482.

行列分解因子分析の正体について

大阪大学大学院基礎工学研究科, 理化学研究所革新知能統合研究センター 寺田 吉壱

1 はじめに

因子分析では, 最尤法をはじめとして, 様々な推定方法が提案されており, それらの統計的性質は詳細に議論されている (e.g., Anderson and Rubin, 1956). 因子分析は, p 次元の観測 x に対して, 以下のような統計モデルを考える.

$$x = \mu + \Lambda f + \epsilon$$

ここで, $\mu \in \mathbb{R}^p$ は母平均ベクトル, $\Lambda \in \mathbb{R}^{p \times m}$ は因子負荷行列, $m (< p)$ は因子数, f は共分散行列が単位行列の中心化された m 次元確率ベクトル, ϵ は f とは無相間な中心化された p 次元確率ベクトルである. f と ϵ は, それぞれ共通因子, 独自因子と呼ばれ, 独自因子の各要素は無相間であると仮定する. 観測を中心化することで, 以下では $\mu = 0_p$ とする. 独自因子の共分散行列を $\text{Var}(\epsilon) = \Psi^2$ と記すと, 因子分析モデル (直交モデル) の下で x の共分散行列は,

$$\text{Var}(x) = \Lambda \Lambda^\top + \Psi^2$$

と表される. 多くの因子分析法は, 標本共分散行列 \hat{S}_n とモデル化した共分散行列 $\Sigma(\Phi) := \Lambda \Lambda^\top + \Psi^2$ の差を最小化することでパラメータ $\Phi := [\Lambda, \Psi]$ を推定する方法として定式化される. このような方法で得られる推定量は, MDF 推定量 (Minimum Discrepancy Function estimator) と呼ばれる.

2 行列分解因子分析 (MDFA)

2000 年代はじめに, 従来の因子分析法とは大きく異なる行列分解に基づく因子分析が提案された. この方法は, 行列分解因子分析 (Matrix Decomposition Factor Analysis, MDFA) と呼ばれる. n 個の観測 x_1, \dots, x_n から構成されるデータ行列 $X_n = (x_1, \dots, x_n)^\top$ に対して, MDFA は以下の損失関数の最小化として定式化される.

$$\mathcal{L}_n(\Lambda, \Psi, F, E) := \frac{1}{n} \|X_n - (F \Lambda^\top + E \Psi)\|_F^2 = \frac{1}{n} \sum_{i=1}^n \|x_i - (\Lambda f_i + \Psi e_i)\|^2$$

ここで, $F = (f_1, \dots, f_n)^\top \in \mathbb{R}^{n \times m}$ と $E = (e_1, \dots, e_n)^\top \in \mathbb{R}^{n \times p}$ は以下の制約を満たす行列である.

$$1_n^\top F = 0_m^\top, 1_n^\top E = 0_p^\top, \frac{1}{n} F^\top F = I_m, \frac{1}{n} E^\top E = I_p, \text{ and } F^\top E = O_{m \times p}$$

以下では, 上述の制約を満たす Z の集合を Θ_Z と記す. MDFA は, シンプルな交互最適化アルゴリズムによって, 容易かつ安定的に解くことができる. MDFA の詳細に関しては, Adachi and Trendafilov (2018) を参照されたい.

3 行列分解因子分析の正体

MDFA は、サンプルサイズ n に依存する F と E をパラメータに含んでおり、以下のセミパラメトリック・モデルに対する最尤推定量として捉えることができる。

$$x_i = \Lambda f_i + \Psi e_i + \xi_i \quad (i = 1, \dots, n)$$

ここで、 $f_1, \dots, f_n, e_1, \dots, e_n$ は上述の制約を満たす定数ベクトルであり、 ξ_1, \dots, ξ_n は独立に分散既知の p 次元の正規分布 $N_p(0_p, \tau_0^2 I_p)$ に従う確率ベクトルである。このセミパラメトリック最尤推定において $Z = [F, E]$ を局外母数と考え、有限次元パラメータ $\Phi = [\Lambda, \Psi]$ に対する新たなプロファイル尤度を考えることで、行列分解因子分析の正体が明らかとなる。

Proposition. 任意の $\Phi \in \Theta_\Phi$ に対して、

$$\min_{Z \in \Theta_Z} \mathcal{L}_n(\Phi, Z) = \text{tr}(\widehat{S}_n) + \text{tr}\{\Sigma(\Phi)\} - 2\text{tr}\left\{\left(\widehat{S}_n^{1/2}\Sigma(\Phi)\widehat{S}_n^{1/2}\right)^{1/2}\right\} =: d_{\text{BW}}^2(\widehat{S}_n, \Sigma(\Phi))$$

が成り立つ。ここで、 $\Sigma(\Phi) := \Phi\Phi^T$ であり、 $d_{\text{BW}}(A, B)$ は正定値行列 A と B の Bures-Wasserstein 距離 (e.g., Bhatia et al. (2019)) を表す。

この命題は、MDFA 推定量が Bures-Wasserstein 距離に基づく MDF 推定量として定式化できることを示している。2つの正定値行列 Σ_1 と Σ_2 に対して、 $d_{\text{BW}}(\Sigma_1, \Sigma_2)$ は Σ_1 と Σ_2 を共分散行列にもつ中心化された2つ楕円分布間の L_2 -Wasserstein 距離を表している (Gelbrich, 1990)。より具体的には、以下の密度関数で定義される楕円分布を考える。

$$f(x | \mu, \Sigma) := \frac{1}{|\Sigma|^{1/2}} g\left((x - \mu)^\top \Sigma^{-1}(x - \mu)\right).$$

ここで、 μ は平均ベクトル、 Σ は共分散行列である。このとき、2つの楕円分布 $f_1(\cdot) = f(\cdot | \mu_1, \Sigma_1)$ と $f_2(\cdot) = f(\cdot | \mu_2, \Sigma_2)$ の間の L_2 -Wasserstein 距離は以下で与えられる。

$$W_2^2(f_1, f_2) := \inf_{X_1 \sim f_1, X_2 \sim f_2} \mathbb{E}[\|X_1 - X_2\|_2^2] = \|\mu_1 - \mu_2\|_2^2 + \text{tr}\left(\Sigma_1 + \Sigma_2 - 2\{\Sigma_1^{1/2}\Sigma_2\Sigma_1^{1/2}\}^{1/2}\right).$$

興味深い事実として、 L_2 -Wasserstein 距離は関数 g に依存しない。つまり、MDFA 推定量は一般の楕円対称分布族を想定した頑健な推定量と捉えることができる。

参考文献

- [1] Anderson, T. and Rubin, H. (1956). Statistical Inference in Factor Analysis. In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability 111*, University of California Press.
- [2] Adachi, K. and Trendafilov, N. T. (2018). Some mathematical properties of the matrix decomposition solution in factor analysis. *Psychometrika*, **83**, 407–424.
- [3] Bhatia, R., Jain, T. and Lim, Y. (2019). On the Bures-Wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, **37**, 165–191.
- [4] Gelbrich, M. (1990). On a formula for the L^2 -Wasserstein metric between measures on Euclidean and Hilbert spaces. *Mathematische Nachrichten*, **147**, 186–203.

Application of High-Dimensional Statistical Analysis to Astrophysical Problems

Tsutomu T. Takeuchi (Division of Particle and Astrophysical Science, Nagoya University, Japan), K. Yata, M. Aoshima, K. Egashira, A. Ishii, Y. Ogane, R. R. Kano, H.-X. Ma, S. A. Matsui, N. Harada, K. Nakanishi, K. Yoshikawa, S. Cooray, H. Okubo, and K. Kohno

1. Introduction

A galaxy is a complex system consisting of stars, interstellar gas and dust (interstellar medium; ISM), and dark matter, all interacting over a wide range of spatial and temporal scales. Star formation occurs in molecular clouds (few pc), while galaxies extend to tens of kpc. Despite this scale gap, global galaxy properties are tightly correlated with their star-formation activity, implying the existence of intermediate (“meso-scale”) physical processes linking the ISM to galactic dynamics. Spectroscopy is the key to probing these processes because spectral lines encode the chemical and dynamical information of remote astrophysical systems. However, recent multi-dimensional spectroscopic surveys, such as ALMA data cubes or integral-field spectroscopy, produce data with thousands of spectral channels for each spatial element, resulting in extremely high-dimensional data structures.

2. Methodology: High-Dimensional PCA and Its Extensions

In classical PCA, meaningful information is extracted from the eigenstructure of the sample covariance matrix. When $d \gg n$, the eigenvalues and eigenvectors are dominated by noise. High-dimensional PCA circumvents this by constructing a dual covariance representation in an $n \times n$ space that preserves the essential eigenvalue information. Theoretical work by Yata and Aoshima (2012, 2024) shows that data vectors in such spaces concentrate either on a hypersphere (Gaussian populations) or along coordinate axes (non-Gaussian populations), explaining why naïve PCA fails. By correcting for eigenvalue bias caused by cumulative noise from higher dimensions, High-dimensional PCA can recover the true principal components. Furthermore, automatic sparse PCA (A-SPCA) selectively extracts a minimal set of variables (spectral channels) that contribute most strongly to each component, enhancing interpretability.

3. Application to the Starburst Galaxy NGC 253

The nearby starburst galaxy NGC 253 is a prototypical target to study intense star-formation activity. ALMA observations provide three-dimensional spectral data cubes with a spatial dimension of $n = 231$ and a spectral dimension of $d = 2248$ (Ando et al. 2017), precisely an HDLSS case. Classical PCA fails to converge, but High-dimensional PCA successfully identifies the dominant spectral components. The first two eigenvalues already account for most of the variance, indicating that the huge spectral information is largely governed by two physical modes. A-SPCA further reveals that these modes are characterized by a small number of molecular lines, such as HCN(4–3), HNC(4–3), and CS(7–6).

Doppler shift corrections (Takeuchi et al. 2024a) remove systemic rotation effects, clarifying the physical meaning of principal components.

4. Further Developments: ARCHEMI Data and Error Detection

New wide-band spectral data of NGC 253 have been obtained with the ARCHEMI instrument (85–163 GHz), providing $n \approx 2000$ spatial elements and $d = 50,702$ spectral channels (Ogane et al. 2025). The data exhibit high spectral but limited spatial resolution and are partially contaminated by strong atmospheric absorption. Applying HD-PCA efficiently identified and removed corrupted frequency ranges, leading to a clean, physically meaningful data set.

5. Application to the HI Forest and Cosmological Simulations

At much earlier cosmic epochs, neutral hydrogen (HI) leaves a forest of absorption lines in quasar spectra — the “HI forest” (Ciardi et al. 2013). These lines trace the large-scale distribution of primordial galaxies and their evolution into modern structures. Even with next-generation instruments such as the Square Kilometre Array (SKA), only a few tens of background quasars are expected, while the number of absorption features is enormous — again an HDLSS situation. Using data from the IllustrisTNG cosmological simulation, we constructed matrices of HI optical depths as functions of redshift. Applying HD-PCA to these matrices (Cang et al. 2025) reveals evolutionary trends in the principal components.

6. Summary and Outlook

Spectroscopic mapping has become indispensable for understanding the physics of the interstellar medium, star formation, and galaxy evolution. However, such data sets are inherently high-dimensional and often limited in sample size. The application of high-dimensional statistical methods, particularly high-dimensional PCA and its sparse, enables the extraction of physically relevant signals while mitigating noise and bias. In the case of NGC 253, these methods successfully revealed dominant molecular components and kinematic structures from ALMA data. For cosmological absorption systems, high-dimensional PCA provided a robust means to characterize the evolution of the HI forest. Future developments include kernel PCA and manifold-learning approaches to capture nonlinear structures.

References

- Ando R. et al. 2017, ApJ, 849, 81
- Ciardi B. et al. 2013, MNRAS, 428, 1755
- Takeuchi T. T. et al. 2024a, ApJS, 271, 44
- Takeuchi T. T. et al. 2024b, Proc. Inst. Stat. Math.(統計数理), 72(2), 273

Shrinkage estimators improving for a wide class of restricted subsets

Yuan-Tsung Chang (ISM, Mejiro University) and Shinozaki Nobuo (Keio University)
Grant-in-Aid for Scientific Research (A) 25H01107 Symposium, Oct. 30-31/2025 at Tsukuba

Abstract In estimating a multivariate normal mean vector Sengupta and Sen (1991) have shown that Stein-type shrinkage estimator improves upon MLE when the shrinkage is made in a positively homogeneous subset. In this paper it is shown that Stein-type shrinkage estimator gives improvement for a wide class of restricted subsets by simply using an integration by parts after applying the transformation to polar coordinates. Further, it is generalized to the case where the shrinkage factor depends on not only the norm but also the direction of the observed vector. Some applications to the estimators shrinking towards a ball and a hyperplane are given.

1 Introduction

Let \mathbf{X} be a p -dimensional random vector which is distributed as the normal distribution $N_p(\boldsymbol{\mu}, \mathbf{I})$ with mean vector $\boldsymbol{\mu}$ and covariance matrix \mathbf{I} . In simultaneous estimation of $p (\geq 3)$ means James and Stein (1961) have shown that an estimator, which we call Stein-type estimator, has uniformly smaller risk than the maximum likelihood estimator (MLE) \mathbf{X} under quadratic loss.

Chang (1981) was the first to discuss the cases where linear inequality constraints are given on p normal means and give Stein-type estimators which improve upon the restricted maximum likelihood estimator (RMLE). Since RMLE is quite different depending on whether the observation vector satisfies the linear inequalities, we have to partition the sample space and construct the improved estimator for each subset separately.

In a systematic and unified manner Sengupta and Sen (1991) have given Stein-type shrinkage estimators which dominate RMLE when the mean vector is restricted to a polyhedral cone. Amirdjanova and Woodroffe (2004) have given related results which allow a more flexible form of shrinkage when a multivariate normal mean belongs to a polyhedral cone. Ouassou and Strawderman (2002), Fourdrinier et al. (2003, 2006) have given improved estimators of a location vector restricted to a convex cone in the general case of a spherically symmetric distribution. Tsukuma and Kubokawa (2008) have shown that the generalized Bayes estimator against the uniform prior distribution over a polyhedral cone is improved upon by a shrinkage estimator. It is also necessary to partition the sample space and construct the improved estimator separately when various types of vague prior information on the mean vector are available. Bock (1982) considered the case where one suspects that the normal mean vector satisfies several linear inequalities and belongs to a closed convex polyhedron K . The proposed estimator shrinks \mathbf{X} in the direction of the orthogonal projection onto K and the dimension used for shrinkage depends on \mathbf{X} .

Kuriki and Takemura (2000) have treated the general case where Stein-type estimator of a multivariate normal mean vector shrinks towards a closed convex set with a smooth or piecewise smooth boundary. Especially, the shrinkage estimator towards a ball is given. Further, it is noticed that Sclove et al. (1972) have given a shrinkage estimator which shrinks \mathbf{X} only when the hypothesis $\boldsymbol{\mu} = \mathbf{0}$ is rejected and improves upon the preliminary-test estimator.

One basic result given by Sengupta and Sen (1991) shows that Stein-type shrinkage estimators dominate MLE when shrinkage is made in a positively homogeneous subset, although it is not explicitly given as a proposition and the proof given in the paper is not simple. In Section 2, it is shown that Stein-type estimator gives smaller averaged loss than \mathbf{X} for a wide class of subsets which includes positively homogeneous ones. Its simple proof is an application of Stein's integration by parts technique (Stein, 1981) after applying the transformation into polar coordinates. Two examples give a new class of improved shrinkage estimators towards a ball. In Section 3 a generalization is given to the case where the amount of shrinkage depends on not only the norm $\|\mathbf{X}\|$ but also the direction $\mathbf{X}/\|\mathbf{X}\|$. Two examples give shrinkage estimators which shrink towards a hyperplane and a face of a p -dimensional cube.

2 Improved estimators shrinking in a straight homogeneous subset

In this section we first give a general theorem which will be useful to show the dominance of the Stein-type shrinkage estimator when shrinkage is made in a class of subsets. Suppose that a p -dimensional random vector $\mathbf{X} = [X_1 \cdots X_p]^t$ is distributed as the p -dimensional normal distribution with mean vector $\boldsymbol{\mu} = [\mu_1 \cdots \mu_p]^t$ and covariance matrix \mathbf{I} . We estimate $\boldsymbol{\mu}$ by $\hat{\boldsymbol{\mu}}$ under the loss function $(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})^t(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu})$. Assuming that $p \geq 3$, we consider a class of estimators of the form

$$\delta(\mathbf{X}) = \mathbf{X} - \frac{\phi(\mathbf{X}^t \mathbf{X})}{\mathbf{X}^t \mathbf{X}} \mathbf{X}, \quad (1)$$

where

- (i) $\phi(\cdot)$ is a non-decreasing and absolutely continuous,
- (ii) $0 \leq \phi(\cdot) \leq 2(p-2)$ and $0 < \phi(\cdot) < 2(p-2)$ on a set of positive measure.

It is well-known that if $\phi(\cdot)$ satisfies the conditions (i) and (ii), then $\delta(\mathbf{X})$ has uniformly smaller risk than \mathbf{X} . (See, for example, Fourdrinier et al. (2018), Section 2.4.) We notice here that the latter condition of (ii) should be read as $0 < \phi(\cdot) < 2(p-2)$ on a subset of \mathcal{S} with positive measure when the domain of $\delta(\mathbf{X})$ is limited to $\mathcal{S} \subset R^p$.

A subset $\mathcal{S} \subset R^p$ is positively homogeneous (p.h.) if $\mathbf{x} \in \mathcal{S}$ implies $c\mathbf{x} \in \mathcal{S}$ for any $c > 0$. Sengupta and Sen (1991) have essentially shown the basic result that the averaged loss of $\delta(\mathbf{X})$ over a p.h. subset is smaller than that of \mathbf{X} , although their results are more general and complicated. However, the basic result is not given as a proposition, and its proof is not simple. Here we generalized it and give a simple proof by using Stein's method of integration by parts. We first give the following.

Definition 1. $\mathcal{S} \subset R^p$ is said to be straight homogeneous (s.h.) if $\mathbf{x} \in \mathcal{S}$ implies $c\mathbf{x} \in \mathcal{S}$ for any $c > 1$. If \mathcal{S} is p.h., then it is s.h. Thus, a cone with vertex the origin is s.h. However, some subsets which are separate from the origin are s.h. The outside of a p -dimensional ball to which the origin belongs is s.h. We can easily see the following properties.

In the proofs of the main results (Theorems 1 and 2) of this paper we use the transformation into polar coordinates:

$$\begin{aligned} x_1 &= r \cos \theta_1, \\ x_2 &= r \sin \theta_1 \cos \theta_2, \\ &\vdots \\ x_{p-1} &= r \sin \theta_1 \sin \theta_2 \cdots \sin \theta_{p-2} \cos \theta_{p-1}, \\ x_p &= r \sin \theta_1 \sin \theta_2 \cdots \sin \theta_{p-2} \sin \theta_{p-1}, \end{aligned} \tag{2}$$

where $0 \leq r, 0 \leq \theta_i \leq \pi, i = 1, \dots, p-2$ and $0 \leq \theta_{p-1} < 2\pi$. We put $\boldsymbol{\theta} = [\theta_1 \cdots \theta_{p-1}]^t$ and $\Theta = \{\boldsymbol{\theta} \mid 0 \leq \theta_i \leq \pi, i = 1, \dots, p-2, 0 \leq \theta_{p-1} < 2\pi\}$ and express the transformation as $x_i = ru_i(\boldsymbol{\theta}), i = 1, \dots, p$. Putting $\mathbf{x} = [x_1 \cdots x_p]^t$ and $\mathbf{u}(\boldsymbol{\theta}) = [u_1(\boldsymbol{\theta}) \cdots u_p(\boldsymbol{\theta})]^t$, we have $\mathbf{x} = r\mathbf{u}(\boldsymbol{\theta})$. Now we show the following.

Theorem 1. Suppose that $\mathbf{X} \sim N_p(\boldsymbol{\mu}, I)$ and that $p \geq 3$. Let $\mathcal{S} \subset R^p$ be s.h. If $\phi(\cdot)$ satisfies the conditions (i) and (ii), then the averaged loss over \mathcal{S} of the estimator (1) is smaller than that of \mathbf{X} .

Now we give the following examples.

Example 1. Shrinkage estimators towards a ball.

Example 2. Improving preliminary-test estimators.

3 Improved estimators shrinking towards a hyperplane

In Example 1 we have given a class of improved estimators that shrink towards a sphere by limiting the shrinkage. Here we first show that it is possible to give improved estimators that shrink in the direction of the origin but does not shrink beyond the s.h. subset \mathcal{S} . For that purpose, we consider the estimators of the form

$$\delta(\mathbf{X}) = \left\{ 1 - \frac{\phi(\mathbf{X})}{\mathbf{X}^t \mathbf{X}} \right\} \mathbf{X}, \tag{3}$$

where $\phi(\mathbf{x}) = \xi(\|\mathbf{x}\|^2, \mathbf{x}/\|\mathbf{x}\|)$, which is $\xi(r^2, \mathbf{u}(\boldsymbol{\theta}))$ when the transformation (2) is applied. We assume that $\phi(\mathbf{x})$ satisfies the following two conditions:

- (i') $\xi(r^2, \mathbf{u}(\boldsymbol{\theta}))$ is a non-decreasing and absolutely continuous function of r for every fixed $\mathbf{u}(\boldsymbol{\theta})$,
- (ii) $0 \leq \phi(\mathbf{x}) \leq 2(p-2)$ and $0 < \phi(\mathbf{x}) < 2(p-2)$ on a set of positive measure.

Therefore, the shrinkage factor may depend on not only r but also $\mathbf{u}(\boldsymbol{\theta})$. Then we have the following.

Theorem 2. Suppose that $\mathbf{X} \sim N_p(\boldsymbol{\mu}, I)$ and that $p \geq 3$. Let $\mathcal{S} \subset R^p$ be s.h. If $\phi(\mathbf{x}) = \xi(\|\mathbf{x}\|^2, \mathbf{x}/\|\mathbf{x}\|)$ satisfies the conditions (i') and (ii), then the averaged loss over \mathcal{S} of the estimator (5) is smaller than that of \mathbf{X} .

Example 3. Let $\mathcal{S} = \{\mathbf{x} = [x_1 \cdots x_p]^t \mid x_i > 0, i = 1, \dots, p, \sum_{i=1}^p x_i \geq 1\}$. An improved estimator is given which shrinks $\mathbf{x} \in \mathcal{S}$ towards the hyperplane $\sum_{i=1}^p x_i = 1$. We should notice that shrinkage is made in the direction of the origin and not in the direction of the closest point on the hyperplane (Bock, 1982).

Example 4. Shrinkage estimator shrinking to $2p$ faces of the p -dimensional cube or more generally rectangular parallelepiped.

二変量正規分布を潜在的にもつ分割表に対する 連関尺度の提案

浦崎 航¹

¹東京理科大学大学院 創域理工学研究科

1 はじめに

カテゴリカル変数は医学、教育学、社会科学などの様々な分野で見られ、解析の対象となることが多い。また、このようなカテゴリカルデータに対する解析は1世紀以上前から現在に至るまで多くの議論や研究の対象となってきた。特に、 r 個のカテゴリと c 個のカテゴリから構成される2種類のカテゴリカル変数を考えたとき、それらの組み合わせから得られる r 行 c 列の表は二元分割表と呼ばれている。このような分割表の解析目的の1つに、2つの変数に関係性があるかを調べることがあり、連関性の解析として様々な手法が提案してきた。そのなかには「連関尺度」というものがあり、サンプルサイズに影響を受けずに関係性の強さを一定の区間内で定量化する手法である。連関尺度の例に、クラメール係数 V^2 (Cramér, 1946) などが古くから存在するが、現在でも様々な場面で用いられ、研究の対象となることも多々にある。

様々な連関性の解析手法が提案される一方、分割表の構成方法に関する議論の1つとして、本来は連続的な潜在変数が閾値で区切られた結果が分割表として観測されるデータである、というものがある。そのため、潜在分布を仮定した分割表に対する提案が様々にされており、連関尺度の枠組みにおいてはポリコリック相関係数などがある。そこで本研究の目的は、Urasaki et al. (2024) で示された独立性における離散型 divergence と潜在分布に二変量正規分布を仮定時の相関係数 ρ との関係に基づき、divergenceに基づく新たな連関尺度の提案にある。また、潜在分布を仮定する状況下で構成される divergence 型の尺度には、どのような特徴が示されるかを提供し、漸近分布に基づく近似信頼区間の導入、提案尺度を用いて実データ解析を行なった結果などを示した。

2 近似的な divergence と相関係数の関係に基づく連関尺度の提案

Urasaki et al. (2024) で示した離散 divergence と潜在分布における二変量正規分布の相関係数との近似的な関係から、パラメータ $\lambda = 0, 1$ を与えた場合に ρ^2 について解くことで、潜在分布に二変量正規分布が仮定される分割表に対する連関尺度とする。したがって、次の2つを提案する：

$$\begin{aligned}\rho_{KL}^2 &= 1 - \exp\{-2I_{KL}(\{p_{ij}\}; \{p_i \cdot p_{\cdot j}\})\}, \\ \rho_P^2 &= 1 - \{2I_P(\{p_{ij}\}; \{p_i \cdot p_{\cdot j}\}) + 1\}^{-1},\end{aligned}$$

ここで, $I_{KL}(\cdot; \cdot)$ および $I_P(\cdot; \cdot)$ のそれぞれは, KL-divergence および Pearson divergence であり, power-divergence に $\lambda = 0, 1$ を適用した場合と一致するため, 以下のように表される:

$$I_{KL}(\{p_{ij}\}; \{p_{i \cdot} p_{\cdot j}\}) = \sum_{i=1}^r \sum_{j=1}^c p_{ij} \log \frac{p_{ij}}{p_{i \cdot} p_{\cdot j}}, \quad I_P(\{p_{ij}\}; \{p_{i \cdot} p_{\cdot j}\}) = \frac{1}{2} \sum_{i=1}^r \sum_{j=1}^c \frac{(p_{ij} - p_{i \cdot} p_{\cdot j})^2}{p_{i \cdot} p_{\cdot j}}.$$

これらの提案尺度はクラメール係数 V^2 などと同様に, divergence に基づいて構成された連関尺度となっている. また, これらの尺度については次の定理を得る.

定理 1. 行と列変数の *Shannon entropy* を $H(X), H(Y)$ とするとき, ρ_{KL}^2 は次の性質をもつ:

- 1 $0 \leq \rho_{KL}^2 \leq 1 - \exp\{-2 \min(H(X), H(Y))\}$.
- 1-1 $r \geq c$ かつ $p_{i \cdot} = r^{-1}$ のとき, $0 \leq \rho_{KL}^2 \leq 1 - r^{-2}$.
- 1-2 $r < c$ かつ $p_{\cdot j} = c^{-1}$ のとき, $0 \leq \rho_{KL}^2 \leq 1 - c^{-2}$.
- 2 $\rho_{KL}^2 = 0$ のとき, 行と列変数には完全な独立構造がある.
- 3 ρ_{KL}^2 が最大値をとるとき, 行と列変数には完全な連関構造がある.
- 4 $r, c \rightarrow \infty$ のとき, $0 \leq \rho_{KL}^2 < 1$.

定理 2. ρ_P^2 は次の性質をもつ:

- 1 $0 \leq \rho_P^2 \leq 1 - \min(r, c)^{-1}$.
- 2 $\rho_P^2 = 0$ のとき, 行と列変数には完全な独立構造がある.
- 3 $\rho_P^2 = 1 - \min(r, c)^{-1}$, 行と列変数には完全な連関構造がある.
- 4 $r, c \rightarrow \infty$ のとき, $0 \leq \rho_P^2 < 1$.

これらの定理から見てとれるように, 提案尺度の上限値が周辺確率やカテゴリ数に依存していることから, 既存の divergence 型尺度とは異なる性質を備えていることがわかる.

本講演では, 上記で提案する 2 つの尺度に関する特徴を深掘りしたとともに, 提案尺度の推定量における漸近分布に基づき構成された近似信頼区間を紹介した. また, 潜在的に仮定された二変量正規分布の相関係数とどの程度近くなるのかを確認するシミュレーション実験を行い, 本研究で提案された連関尺度がカテゴリ数および標本数の増加に伴って相関係数の値に近づいていくことが示された. そのほか, 実データ解析の例から直感的な関係性の度合いを提供できることも示した.

参考文献

- Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton university press.
- Urasaki, W., Nakagawa, T., Momozaki, T., and Tomizawa, S. (2024). Generalized cramér's coefficient via f-divergence for contingency tables. *Advances in Data Analysis and Classification*, 18(4):893–910.

ノンパラメトリック・ロジスティック回帰の統計理論

大阪大学大学院基礎工学研究科 屋良淳朝
大阪大学大学院基礎工学研究科, 理研 AIP 寺田吉壱

はじめに

深層学習は、様々な実世界の複雑なタスクに対して極めて高い性能を発揮しており、注目を浴びている。なぜ深層学習による推定量が他の推定量を優越するかの解明は、我々が深層学習をどのように利用するのが適切かについての指針を与えるため、重要な課題である。この課題を解決するために、様々な観点から深層学習の理論研究が行われている。特に、回帰問題や分類問題を中心に、深層学習を用いたノンパラメトリック推定量の理論的性質の解明が進んでいる。

回帰問題においては、多くの研究が、深層学習の優れた適応能力を明らかにしており、深層学習がカーネル法などの他のノンパラメトリック推定量を優越する理由が部分的に解明されている。さらに、いくつかの研究では、深層学習を用いた推定量は次元の呪いを回避できる可能性があることを示唆している (e.g., Schmidt-Hieber, 2020)。

分類問題では、誤判別率の収束など判別性能に関する研究が進んでいる。一般に、分類問題では、経験的な誤判別率の最小化を実施することは難しいため、分類器（推定量）はヒンジ損失やロジスティック損失などの凸な代理損失を最小化することによって得られる。近年、様々な代理損失に対して、深層学習を用いた分類法における誤判別率の収束レートが明らかとなり、分類問題においても深層学習の利点が解明されている (e.g., Kim et al., 2021)。

一方で、ロジスティック回帰分析のように、分類だけでなくクラス所属確率（条件付き確率）の推定を同時に使う場合も多い。このような場合、推定された分類器の分類性能だけでなく、条件付き確率の推定精度も分類問題における重要な話題の一つである。通常、深層学習による分類器は、出力層にロジスティックシグモイド関数やソフトマックス関数を用いることで各クラスの所属確率を出力する。条件付き確率の推定は、単にクラスの所属を予測するよりも多くの情報を与えるので、現実世界のデータ分析においても有用である。

本研究では、ロジスティック回帰による条件付き確率のノンパラメトリック推定を考える。具体的には、条件付き確率のノンパラメトリック最尤推定量 (nonparametric maximum likelihood estimator, NPMLE) の理論的性質について解析する。NPMLE の一致性を示すには、単に推定量と真の条件付き確率のロジスティック損失に関する差の期待値の収束を考えればよく、ロジスティック回帰においては推定量と真の条件付き確率の間の Kullback-Leibler (KL) 情報量の期待値を考えれば良い。しかし、条件付き確率のノンパラメトリック推定では、真の条件付き確率との推定量の KL 情報量は容易に発散することが知られている。具体的には、推定量の台が真の条件付き確率の台を覆っていないければ発散する。従って、ノンパラメトリックロジスティック回帰において KL 情報量に関する収束を導くためには、「真の条件付き確率が bounded away from zero である」といった仮定が必要となる (e.g., Ohn and Kim, 2022)。

そこで、本研究では van de Geer (2000) のアプローチを用いて真の条件付き確率とその推定量の間の Hellinger 距離を直接評価し、多クラスのロジスティック回帰分析におけるノンパラメトリック最尤推定量のオラクル不等式を導出する。Hellinger 距離の収束を考えることで、KL 情報量の評価に比べて、現実的な仮定の下で推定量の性能を評価することができる。また、オラクル不等式の重要な応用として、真のクラス条件付き確率が composition structured function (Schmidt-Hieber, 2020) であるときに深層学習による推定量の Hellinger 距離に関する収束レートを導出する。さらに、その収束レートがほとんどミニマックス最適であることを示す。

主結果

K クラスの分類問題を考える。 $\mathcal{X} \subset \mathbb{R}^d$ を入力空間、 $\mathcal{Y} = \{\mathbf{e}_i\}_{i=1}^K$ をラベルの集合とする。ここで、 $\mathbf{e}_1, \dots, \mathbf{e}_K$ は K 次元の標準基底ベクトルである。データ $(\mathbf{X}, \mathbf{Y}) \in \mathcal{X} \times \mathcal{Y}$ は以下のモデルから生成されると仮定する。

$$Y_k \mid \mathbf{X} = \mathbf{x} \sim \text{Bernoulli}(\eta_k(\mathbf{x})), \quad \mathbf{X} \sim P_{\mathbf{X}}, \quad k = 1, \dots, K. \quad (1)$$

ここで, $\eta_k(\mathbf{x}) := \mathbb{P}(\mathbf{Y} = \mathbf{e}_k \mid \mathbf{X} = \mathbf{x})$ は真のクラス条件付き確率であり, $P_{\mathbf{X}}$ は入力空間 \mathbf{X} 上の未知の分布である. \mathbf{X} と \mathbf{Y} の同時分布を P と書く. $\mathcal{D}_n = \{(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_n, \mathbf{Y}_n)\}$ を母集団分布 P からのサイズ n の i.i.d. サンプルとする. 本研究では, 条件付き確率 $\boldsymbol{\eta}$ のノンパラメトリック推定に焦点を当てる.

データ \mathcal{D}_n が与えられたとき, 条件付き確率 $\mathbf{p} = (p_1(\mathbf{x}), \dots, p_K(\mathbf{x}))^\top$ の負の対数尤度関数は

$$L_n(\mathbf{p}) := -\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K Y_{ik} \log p_k(\mathbf{X}_i) = -\frac{1}{n} \sum_{i=1}^n \mathbf{Y}_i^\top \log \mathbf{p}(\mathbf{X}_i) \quad (2)$$

で与えられ, 最尤推定量 (maximum likelihood estimator, MLE) は $\hat{\mathbf{p}}_n \in \arg \min_{\mathbf{p} \in \mathcal{F}_n} L_n(\mathbf{p})$ で定義される. ここで, \mathcal{F}_n は推定量の候補となる関数の集合 (仮説空間) である. 損失関数 (2) に超過リスクは

$$\mathbb{E}_{\mathbf{X}} \left[\boldsymbol{\eta}(\mathbf{X})^\top \log \frac{\boldsymbol{\eta}(\mathbf{X})}{\hat{\mathbf{p}}(\mathbf{X})} \right] = \mathbb{E}_{\mathbf{X}} [\text{KL}(\boldsymbol{\eta}(\mathbf{X}) \parallel \hat{\mathbf{p}}(\mathbf{X}))] \quad (3)$$

で表される. ここで, $\text{KL}(\cdot \parallel \cdot)$ は KL 情報量を表す. MLE $\hat{\mathbf{p}}_n$ の一致性を示すには, 超過リスク (3) を評価するのが自然に思えるが, モデル \mathcal{F}_n として“小さい”クラスを考えたときでさえ超過リスク (3) が発散してしまう例が存在する. そのため, 本研究では元々の超過リスク (3) の代わりに以下の量を評価する:

$$R(\boldsymbol{\eta}(\mathbf{X}), \hat{\mathbf{p}}(\mathbf{X})) := \mathbb{E}_{\mathbf{X}} [H^2(\boldsymbol{\eta}(\mathbf{X}), \hat{\mathbf{p}}(\mathbf{X}))]. \quad (4)$$

ここで, H^2 は Hellinger 距離を表す.

上記の設定の下で, 以下の不等式は真の条件付き確率と NPMLE との Hellinger 距離に関する評価を与える. 詳細は Yara and Terada (2025) を参照されたい.

定理 1 (オラクル不等式). K クラスの分類問題 (1) を考える. 適当な条件の下で, 普遍定数 c と $\sqrt{n}\delta_n^2 \geq c\Psi(\delta_n)$ なる任意の $\delta \geq \delta_n$ に対して

$$\mathbb{E}_{\mathcal{D}_n} [R(\hat{\mathbf{p}}_n, \boldsymbol{\eta})] \leq 514(1 + c_0^2)(\delta^2 + R(\tilde{\mathbf{p}}_n, \boldsymbol{\eta})) + \frac{c^3}{n} \quad (5)$$

が成り立つ. ここで, $\tilde{\mathbf{p}}_n$ は \mathcal{F}_n の任意の元であり, $\mathbb{E}_{\mathcal{D}_n}$ は学習データ \mathcal{D}_n についての期待値を表す.

このオラクル不等式によって, 様々な真の条件付き確率が属する関数クラスと推定モデルの設定のもとで NPMLE の収束レートを導出することができる. 例えば, 推定モデルとして適当な大きさの DNN を用いた場合, $\boldsymbol{\eta}$ が合成関数の構造を持つ場合 (Schmidt-Hieber (2020) を見よ) は

$$\mathbb{E}_{\mathcal{D}_n} [R(\hat{\mathbf{p}}_n, \boldsymbol{\eta})] \lesssim \phi_n \log(n)^3, \quad \phi_n := \max_{i=0, \dots, q} n^{-\frac{\beta_i^*}{\beta_i^* + t_i}}, \quad \beta_i^* := \beta_i \prod_{l=i+1}^q (\beta_l \wedge 1)$$

が成立する. この収束レートは入力の次元に依存しておらず, 次元の呪いを回避していることがわかる.

紙面の都合上省略するが, 当日は上記の収束レートのミニマックス最適性や, 真の条件付き確率が非等方 Besov 空間に属する場合の収束レートについても紹介する.

参考文献

- Yongdai Kim, Ilsang Ohn, and Dongha Kim. Fast convergence rates of deep neural networks for classification. *Neural Networks*, 138:179–197, 2021.
- Ilsang Ohn and Yongdai Kim. Nonconvex sparse regularization for deep neural networks and its optimality. *Neural computation*, 34:476–517, 2022.
- Johannes Schmidt-Hieber. Nonparametric regression using deep neural networks with ReLU activation function. *Annals of Statistics*, 48:1875 – 1897, 2020.
- Sara van de Geer. *Empirical Processes in M-estimation*. Cambridge university press, 2000.
- Atsutomo Yara and Yoshikazu Terada. Nonparametric logistic regression with deep learning. *Bernoulli*, 2025. page in press.

方向データにおけるロバストなベイズ推定

中川 智之¹ 鶴田 靖人² 文 翔³ 田畠 耕治³

¹ 明星大学データサイエンス学環/理化学研究所

² 明治大学経営学部

³ 東京理科大学創域理工学部

導入

$p - 1$ 次元単位球面 $\mathcal{S}_p = \{x \in \mathbb{R}^p \mid \|x\| = 1\}$ 上の観測 x_1, \dots, x_n は、さまざまな科学分野で現れる。ここで $\|\cdot\|$ はユークリッドノルムである。例えば、風向、動物の方向性、山火事の延焼方向などは \mathcal{S}_2 上に分布する [5]。 $p = 3$ の典型例としては、岩石の古地磁気の方向や地球から星への方向が挙げられる [5]。テキストデータやゲノム配列の表現は、高次元単位球面上の観測とみなされる [2]。球面データの解析のために、文献においていくつかの確率分布が提案してきた。その中でも中心的な役割を果たしてきたモデルは、Langevin 分布とも呼ばれる von Mises-Fisher 分布である。方向パラメータを μ 、集中パラメータを κ とすると von Mises-Fisher 分布の密度関数は次で与えられる。

$$f(x \mid \mu, \kappa) = \frac{\kappa^{(p-2)/2}}{(2\pi)^{p/2} I_{(p-2)/2}(\kappa)} \exp\{\kappa \mu^\top x\}, \quad x \in \mathcal{S}_p$$

ここで、 $I_{(p-2)/2}(\cdot)$ は第一種の修正ベッセル関数である。本研究では、von Mises-Fisher 分布における位置または集中パラメータのロバスト推定に焦点を当てる。Agostinelli[1] は、円周上のデータに対して、両方のパラメータを同時に推定するための別の 2 つの方法を提案している。また、Kato and Eguchi[4] は von Mises-Fisher 分布に対し、density power-divergence および γ -divergence を用いた M -推定量を導入し、そのロバスト性を検討した。しかしながら、これらの手法はカーネル密度推定などが必要で計算コストが高く p が大きい場合の推定が不安定になる。さらに、信頼区間など不確実性の評価が難しい。

そこで本研究では、ベイズ統計の枠組みを用いて、外れ値に対してロバストな事後分布を、density power-divergence および γ -divergence を用いて提案する。このアプローチは、ベイズ統計の枠組みを用いることで、特に標本サイズが小さい場合にも推定の不確実性を容易に評価できる利点がある。さらに、本研究では事後平均を推定するための重み付きベイズブートストラップ法を用いた近似的な事後分布からのサンプリングアルゴリズムを提示する。

方向データにおけるロバストなベイズ推定

方向データにおいて、Kato and Eguchi[4] は von Mises-Fisher 分布に対し、density power-divergence および γ -divergence に基づいた推定量を導入し、そのロバスト性を検討した。一方で、ベイズ的枠組みにおいて、Ghosh and Basu[3] および Nakagawa and Hashimoto[6] は、それぞれ density power-divergence および γ -divergence に基づく一般化事後分布 $\pi^{(\alpha)}(\xi \mid x_{1:n})$

および $\pi^{(\gamma)}(\boldsymbol{\xi} \mid \mathbf{x}_{1:n})$ を提案している。ただし、 $\mathbf{x}_{1:n} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ は観測データである。von Mises-Fisher 分布の場合、これらは次のように表される。

$$\pi^{(\alpha)}(\boldsymbol{\xi} \mid \mathbf{x}_{1:n}) = \frac{\pi(\boldsymbol{\xi}) \exp(-nd_\alpha(\bar{g}, f_{\boldsymbol{\xi}}))}{\int \pi(\boldsymbol{\xi}) \exp(-nd_\alpha(\bar{g}, f_{\boldsymbol{\xi}})) d\boldsymbol{\xi}}, \quad \pi^{(\gamma)}(\boldsymbol{\xi} \mid \mathbf{x}_{1:n}) = \frac{\pi(\boldsymbol{\xi}) \exp(-n\tilde{d}_\gamma(\bar{g}, f_{\boldsymbol{\xi}}))}{\int \pi(\boldsymbol{\xi}) \exp(-n\tilde{d}_\gamma(\bar{g}, f_{\boldsymbol{\xi}})) d\boldsymbol{\xi}}$$

ここで、 $\boldsymbol{\xi} = \kappa \boldsymbol{\mu}$ であり、 $\alpha, \gamma > 0$ はロバスト性を制御するためのチューニングパラメータであり、 $d_\alpha(\bar{g}, f_{\boldsymbol{\xi}}) = n^{-1} \sum_{i=1}^n \ell_\alpha(\mathbf{x}_i, \boldsymbol{\xi})$ 、および $\tilde{d}_\gamma(\bar{g}, f_{\boldsymbol{\xi}}) = -\gamma^{-1} \{ \exp(-\gamma d_\gamma(\bar{g}, f_{\boldsymbol{\xi}})) - 1 \} = n^{-1} \sum_{i=1}^n \ell_\gamma(\mathbf{x}_i, \boldsymbol{\xi})$ はそれぞれ次のように定義される。

$$\begin{aligned}\ell_\alpha(\mathbf{x}_i, \boldsymbol{\xi}) &= -\frac{1}{\alpha} \frac{\exp(\alpha \boldsymbol{\xi}^\top \mathbf{x}_i)}{K_p(\boldsymbol{\xi})^\alpha} + \frac{1}{\alpha+1} \frac{K_p((1+\alpha)\boldsymbol{\xi})}{K_p(\boldsymbol{\xi})^{\alpha+1}}, \\ \ell_\gamma(\mathbf{x}_i, \boldsymbol{\xi}) &= -\frac{1}{\gamma} \frac{\exp(\gamma \boldsymbol{\xi}^\top \mathbf{x}_i)}{K_p((1+\gamma)\boldsymbol{\xi})^{\gamma/(1+\gamma)}} + \frac{1}{\gamma}\end{aligned}$$

ロバスト性

影響関数 (Influence Function; IF) は、ロバスト統計学において推定量のロバスト性を評価するために用いられる概念であり、データに対する微小な変化に対する感度を定量化するものである。すなわち、ある推定量 $T_n(G)$ が外れ値や小さな摂動にどの程度影響を受けるかを評価する尺度として次のように定義される。

$$IF(\mathbf{y}, T_n, G) = \lim_{\varepsilon \rightarrow 0} \frac{T_n((1-\varepsilon)G + \varepsilon \Delta_{\mathbf{y}}) - T_n(G)}{\varepsilon}$$

ここで、 ε は汚染比率、 $\Delta_{\mathbf{y}}$ は点 $\mathbf{y} \in \mathcal{S}_p$ に集中する退化分布である。方向データの場合、データ空間がコンパクト集合上に存在するため、影響関数は常に有界となる。このような場合には、スケールおよびパラメータ化の影響を排除するために、標準化影響関数 (Standardized Influence Function; SIF) を考慮するのが有用である。このとき、SIF のノルムは次式で定義される。

$$SIF(\mathbf{y}, T_n, G) = \sqrt{IF(\mathbf{y}, T_n, G)^\top S(G)^{-1} IF(\mathbf{y}, T_n, G)}$$

ここで、 $S(G)$ は漸近分散共分散行列を表す。この指標は推定量の外れ値に対する感度を統一的なスケールで評価できるため、ロバストベイズ推定法の比較や性能評価において重要である。当日は数値実験の結果を含めて詳細を報告し、有意義なディスカッションができた。

参考文献

- [1] C. Agostinelli. Robust estimation for circular data. *Computational Statistics & Data Analysis*, 51(12):5867–5875, 2007.
- [2] A. Banerjee, I. S. Dhillon, J. Ghosh, and Sra. Clustering on the unit hypersphere using von mises-fisher distributions. *Journal of Machine Learning Research*, 6(9):1345–1382, 2005.
- [3] A. Ghosh and A. Basu. Robust bayes estimation using the density power divergence. *Annals of the Institute of Statistical Mathematics*, 68(2):413–437, 2016.
- [4] S. Kato and S. Eguchi. Robust estimation of location and concentration parameters for the von Mises–Fisher distribution. *Statistical Papers*, 57:205–234, 2016.
- [5] K. V. Mardia, P. E. Jupp, and K. Mardia. *Directional statistics*, volume 2. Wiley Online Library, 2000.
- [6] T. Nakagawa and S. Hashimoto. Robust Bayesian inference via γ -divergence. *Communications in Statistics-Theory and Methods*, 49(2):343–360, 2020.

多変量モデルの頑健な事後分析

東京大学経済学部 入江 薫

概要

一変量回帰モデルのパラメータの事後分布を外れ値から保護することを目的として、事後分布の頑健性が数学的に定義されている。過去10年間のうちに頑健性達成のための簡易な十分条件の研究が進み、位置尺度モデル (Desgagné, 2015), 回帰モデル (Gagnon et al., 2020), 階層モデル (Hamura et al., 2022), ポアソン・負の二項回帰モデル (Hamura et al., 2025+) の頑健な事後分析の理論が明らかになっている。これらの研究はすべて、誤差分布を裾厚にするという古典的なロバスト統計のモデリングに基づくものだが、裾の厚さが t 分布よりも厚い（密度関数が対数項を含み、対数正規変動である）誤差分布でないと事後頑健性を達成できないという点が特徴である。

本研究ではこれまでの研究を多変量へと拡張し、多変量回帰モデルにおける事後頑健性の十分条件を与える。具体的には、 p 次元の観測値 \mathbf{y}_i に対して、 $\mathbf{y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \boldsymbol{\Sigma})$ というモデルを考える。このクラスは共分散行列 $\boldsymbol{\Sigma}$ の推定に关心があるグラフィカルモデルを含む。いま、 \mathbf{y}_i の成分（の一部）に外れ値の存在が疑われるときに、元のモデルをいかに変更すれば事後頑健性を達成できるかという問い合わせに対して、我々は以下のモデルを提案する：

$$\mathbf{y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{T}_i\boldsymbol{\Sigma}\mathbf{T}_i), \quad \mathbf{T}_i = \text{diag}(t_{i1}, \dots, t_{ip}).$$

ここで、 t_{ik} には独立に超裾厚な分布を仮定し、そのサポートは $(-\infty, -1) \cup (1, \infty)$ とする。各成分 y_{ik} の分散は $t_{ik}^2 \Sigma_{kk}$ であるから、外れ値に対して t_{ik}^2 が大きくなり誤差分散を増加させ、外れ値を誤差項で説明するという点は変わらない。特徴的であるのは t_{ik} が負になりうるという点である。これにより、 y_{ik} と $y_{ik'}$ の相関は

$$\text{sgn}(t_{ik}t_{ik'}) \frac{\Sigma_{kk'}}{\Sigma_{kk}\Sigma_{k'k'}},$$

となり、外れ値が存在する ($t_{ik} \neq 1$ または $t_{ik'} \neq 1$) 時には相関の符号が変わり得ることが分かる。逆にいえば、他のモデルでは外れ値の存在によって相関の絶対値が減少したり、あるいは外れ値が生じるときも通常と同じ相関を用いることになり不合理である。本講演では、このような共分散構造が事後頑健性の達成にとって本質的である点について議論する。また、類似のモデル、特に \mathbf{T}_i に負値を許すモデルの研究はほとんど見当たらないことについて、参考文献を挙げつつ議論する。

本研究は羽村靖之氏（京都大学）および菅澤翔之助氏（慶應義塾大学）との共同研究である。

研究発表の報告

研究報告においては、提案する誤差分布 $\varepsilon_i \sim N(\mathbf{0}, \mathbf{T}_i\boldsymbol{\Sigma}\mathbf{T}_i)$ の特徴を解説するため、 $p = 2$ 次元の場合の密度関数を描画した。ただし、 $\Sigma_{11} = \Sigma_{22} = 1$ かつ $\Sigma_{12} = 0.9$ とし、異常値は確率 0.05 で片側または両側対数ペレート分布から生成されるとし、モンテカルロ法を用いている。 t_{ik} が正の値を取る場合（上図）と、正負両方の値を取る場合（下図）の違いは明らかである。この例では、通常時には誤差に正の相関が見られるが、上図のモデルでは異常値が生じる際にも変数間の正の相関が維持されてしまう。下図のモデルでは、異常値が生じる場合には通常とは異なる相関にもある程度の確率密度を与えていた。

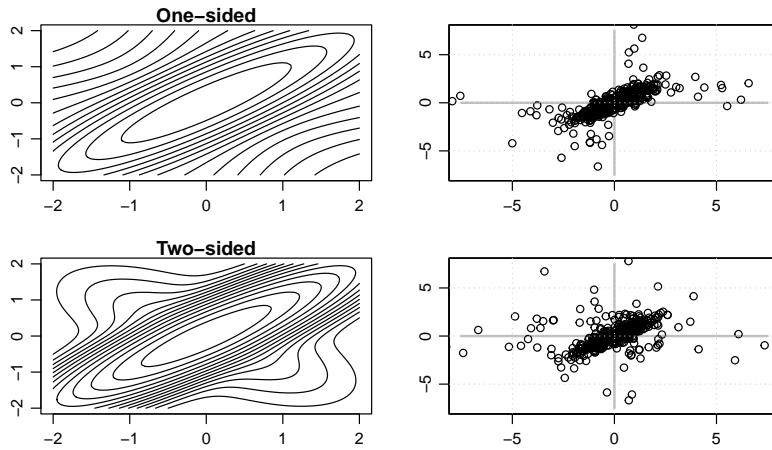


図 1 $p = 2$ の場合の誤差分布の密度

講演後の議論では、仮に異常値が正の相関を持っていた場合には、むしろ上図のモデルの方が効率性の点で優れるのではという指摘があった。確かにその通りであるが、そのためにはデータから異常値の生成構造について推定できなければならない。そもそも異常値の数が少ないという前提からいえば、異常値の生成構造を推定することは一般には困難であると思われる。提案する下図のモデルは、異常値の構造によらず、関心のあるパラメータの事後分布を頑健にすることを目標としている。

- Desgagné, A. (2015). "Robustness to outliers in location-scale parameter model using log-regularly varying distributions." *The Annals of Statistics*, 43(4).
- Gagnon, P., Desgagné, A., & Bédard, M. (2020). "A new Bayesian approach to robustness against outliers in linear regression." *Bayesian Analysis*, 15(2), 389-414.
- Hamura, Y., Irie, K., & Sugashawa, S. (2022). "Log-regularly varying scale mixture of normals for robust regression." *Computational Statistics & Data Analysis*, 173, 107517.
- Hamura, Y., Irie, K., & Sugashawa, S. (2025+). "Robust hierarchical modeling of counts under zero-inflation and outliers: Theoretical robustness and efficient computation." *Journal of the American Statistical Association*, published online.
- Hamura, Y., Irie, K., and Sugashawa, S. (2025), "Outlier-robust Bayesian multivariate analysis with correlation-intact sandwich mixture," arXiv:2508.18004.

点過程に対するリード・ラグ解析

塩谷天章 * 林高樹 † 小池祐太 *

2つの金融時系列データ間のリード・ラグ関係(時間差を伴って生じる相関関係)に関する実証研究は、ファイナンス分野では古くから研究されてきたテーマの1つである。本研究では、ティックデータ(1日内にあった約定・注文を全て記録したデータ)を用いて金融資産間のリード・ラグ関係を解析する方法について議論する。ティックデータの特徴として、約定・注文が生じた時刻ごとにデータが記録されていることに起因する、観測時刻の不規則性がある。従って、ティックデータではタイムスタンプが非常に重要であり、分析のためには明示的にモデルに組み込む必要がある。その際、連続時間確率過程の枠組みでモデル化するのが便利である。そのようなアプローチに基づきティックデータを用いてリード・ラグ関係を分析する方法を提案した研究として代表的なものに、Hoffmann et al. (2013) と Dobrev and Schaumburg (2015) がある。前者は価格系列に基づく分析手法、後者はタイムスタンプに基づく分析手法をそれぞれ提案している。価格系列どうしのリード・ラグ関係の検出は、裁定取引戦略の構築や市場の価格発見機能の特定に直接つながると考えられることもあり重要ではあるが、ティックデータにおいては市場の取引メカニズムから生じる「ノイズ」(いわゆるマーケット・マイクロストラクチャーノイズ)の存在により、リード・ラグ関係の推定結果が不安定になる傾向がある。そのため、Hoffmann et al. (2013) の方法による推定結果はしばしば解釈が難しい。Dobrev and Schaumburg (2015) により提案されたタイムスタンプのみに基づく方法はこの問題を回避しており、実際に非常に安定したリード・ラグ関係が推定されることがいくつかの実証研究で報告されている。Dobrev and Schaumburg (2015) および林 (2017) を参照のこと。

一方で、Dobrev-Schaumburg の方法は記述統計的なアプローチであり、彼らの方法が一体何を「推定」しているのか不明瞭であるという問題がある。また、実際には DS 推定量がうまく機能しないケースも存在し、特に、データ件数が相対的に少ないようなデータにおいてはその傾向が顕著である。加えて、DS 推定量の実装のためにはデータ区間を等間隔のバケットに分割する必要があるが、バケットサイズの選択が推定結果に大きく影響することが確認できる。しかし、Dobrev-Schaumburg の方法は「モデルフリー」であるが故に、これらの現象が生じてしまう原因に対する統計的な解釈を与えてくれない。

この問題を解決するために、本研究では、タイムスタンプ間のリード・ラグ関係を推定する問題を、定常2変量点過程の cross-pair correlation function (CPCF) のピーク位置を推定する問題として定式化する。この定式化の下で、Dobrev-Schaumburg の cross-market activity measure は本質的に

* 東京大学 大学院数理科学研究科

† 慶應義塾大学 大学院経営管理研究科

は CPCF の推定量を与えており, 従って DS 推定量は CPCF の maximizer をリード・ラグ時間として推定しているものと解釈できることを示す. その結果, Dobrev–Schaumburg の方法が不安定となる状況は, バケットの取り方が, データの潜在的な性質から決まる許容範囲に比して狭すぎることに起因することが示唆される. この問題を解消するために, 本研究では, Dobrev–Schaumburg の方法よりもある意味で「自然」なリード・ラグ時間の推定量を提案し, Dobrev–Schaumburg の方法がうまくいかないような状況でも安定した推定結果が得られることを見る.

参考文献

- Dobrev, D., Schaumburg, E. 2015. High-frequency cross-market trading: Model free measurement and applications. Working paper.
- 林高樹 2017. 「高頻度注文板データの統計解析: 異市場・同一株式価格間の先行遅行関係」, 『統計数理』, 65 (1), 113–139.
- Hoffmann, M., Rosenbaum, M., Yoshida, N. 2013. Estimation of the lead-lag parameter from non-synchronous data. Bernoulli, 19 (2), 426–461.

A cumulative sum-based change-point detection method for high-dimensional data

Kento Egashira^a, Kazuyoshi Yata^b, Makoto Aoshima^b

^aDepartment of Information Sciences, Tokyo University of Science

^bInstitute of Mathematics, University of Tsukuba

Change-point analysis can generally be classified into two broad approaches. The first deals with a given dataset to investigate the existence, number, and locations of change-points, while the second monitors data that are observed sequentially in order to detect a change-point as quickly as possible. The latter is commonly referred to as online change-point detection, whereas the former is known as offline change-point analysis. In this talk, we focused on the offline setting. Even within this framework, two distinct situations can be considered. In the first, the existence of change-points is assumed, and the objective is to estimate both their number and locations. In the second, the existence of change-points is itself uncertain, requiring both a statistical test for their presence and, if present, estimation of their number and locations.

In recent years, the scope of change-point analysis has been extended to high dimensional data, motivated by the increasing availability of large-scale observations in fields such as genomics, finance, neuroscience, and network analysis. Classical change-point detection methods are often designed under the assumption that the data dimension is fixed and relatively small compared to the sample size. However, such methods may fail or become unreliable when the number of variables is comparable to or even exceeds the number of observations. This high-dimensional regime introduces substantial challenges, including the need to handle strong dependence among variables. Accordingly, a growing body of research has been devoted to developing high-dimensional change-point detection methods that leverage modern statistical and computational tools such as principal component analysis, factor models, estimating covariance structure, and sparse regularization techniques. These approaches aim to detect structural changes not only in the mean vector but also in the covariance or latent factor structure, providing a more comprehensive understanding of change-point detection in high-dimensional settings. A number of studies have been conducted on change-point detection in high-dimension, low-sample-size (HDLSS) settings. Among them, the cumulative sum (CUSUM) statistic, a representative approach for detecting mean shifts, has been widely extended to the high-dimensional framework. For instance, Liu et al. [3] developed a CUSUM-type procedure based on U-statistics for high-dimensional data.

Despite the extensive literature on high-dimensional change-point analysis, most existing methods rely on specific structural assumptions - such as sparsity, independence, or normality - that may not hold in practice. Meanwhile, in the broader development of high-dimensional statistical analysis, the importance of the eigenvalue structure of the covariance matrix has been increasingly recognized. Empirically, Yata and Aoshima [4]

reported that the leading eigenvalues of covariance matrices often grow as power functions of the dimension. In response to these findings, Aoshima and Yata [1] proposed a classification of high-dimensional covariance structures based on the contribution of the leading eigenvalues to introduce the strongly spiked eigenvalue (SSE) and the non-SSE (NSSE) model for high dimensional covariance matrix. This dichotomy provides a fundamental classification for analyzing high-dimensional data, since the covariance matrix and the inference procedures differ drastically between these two regimes.

Building upon these developments, it is natural to consider change-point analysis within the framework of the SSE model. In addition, since many real datasets fall within the scope of the SSE model, developing theoretical frameworks that explicitly account for the SSE model can be expected to enhance methodological performance. Consequently, the importance of developing change-point detection procedures under the SSE model has become increasingly evident.

In this talk, we introduced a high-dimensional multivariate CUSUM procedure presented in Liu et al. [2], and derived its asymptotic properties in the framework of HDLSS, including the SSE model. Based on the asymptotic properties obtained, we proposed a modification that ensures consistency under the SSE model without additional structural assumptions. Furthermore, we introduced a change-point detection procedure that captures not only shifts in mean vectors but also changes in covariance structures.

Acknowledgments

This research of the first author was partially supported by Grants-in-Aid for Early-Career Scientists, JSPS, under Contract Number 24K20748. This research of the second author was partially supported by a Grant-in-Aid for Scientific Research (C), JSPS, under Contract Number 22K03412. This research of the third author was partially supported by Grant-in-Aid for Scientific Research (A), JSPS, under Contract Number 25H01107.

References

- [1] Aoshima, M., Yata, K. (2018). Two-sample tests for high-dimension, strongly spiked eigenvalue models. *Statistica Sinica*, 28, 43–62.
- [2] Liu, B., Zhang, X., Liu, Y. (2022). High dimensional change point inference: recent developments and extensions. *Journal of Multivariate Analysis*, 188, 104833.
- [3] Liu, B., Zhou, C., Zhang, X., Liu, Y. (2020). A unified data-adaptive framework for high dimensional change point detection. *Journal of the Royal Statistical Society, Series B*, 82, 933–963.
- [4] Yata, K., Aoshima, M. (2013). PCA consistency for the power spiked model in high-dimensional settings. *Journal of Multivariate Analysis*, 122, 334–354.