

# 高次元からのデータ科学への挑戦状

矢野 恵佑

統計数理研究所 統計基盤数理研究系

@ 統計数学X情報X物質セミナー②～高次元データの計測と統計解析～

# 自己紹介

## 略歴

➤ 愛媛出身

➤ 3/2017 : 東京大学 大学院情報理工学系研究科  
(博士 情報理工学)

➤ 4/2017- 3/2020 : 東京大学 計数工学科助教

➤ 4/2020- : 統計数理研究所 准教授

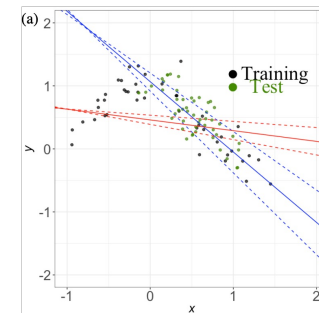
## 専門

• 統計学・機械学習→データ科学

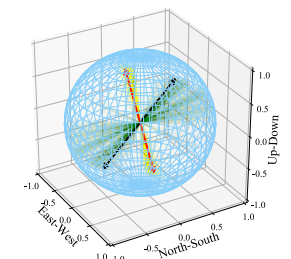
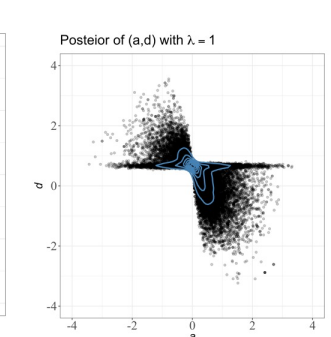
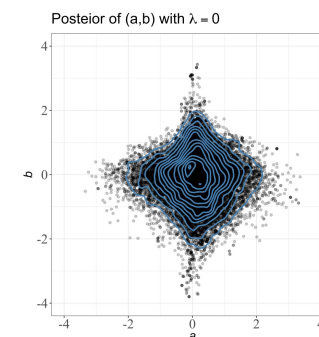
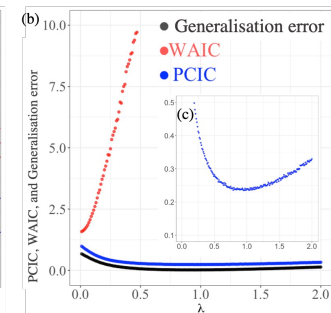
• 地震学・測地学・プラズマ物理でのデータ解析

• 「データ科学技術」とその数理メカニズムを解明するのが好き  
応用ごとに技術を開発・拡張したりするのも好き

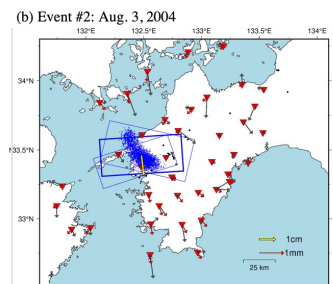
[Iba and Y. (2023) ]



[Okuno and Y. (2023) ]



Sei and Y. (2024)

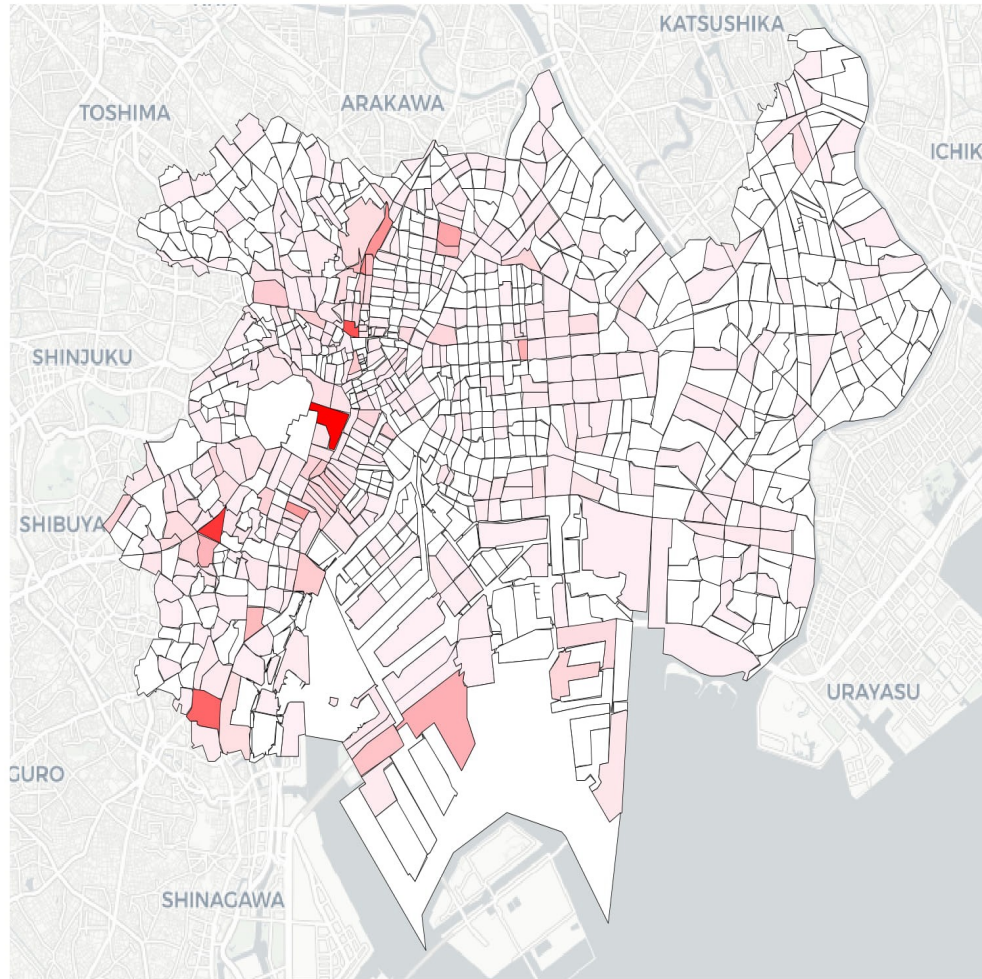


Y. and Kano (2022)

# 今日のデータ科学における「データ」

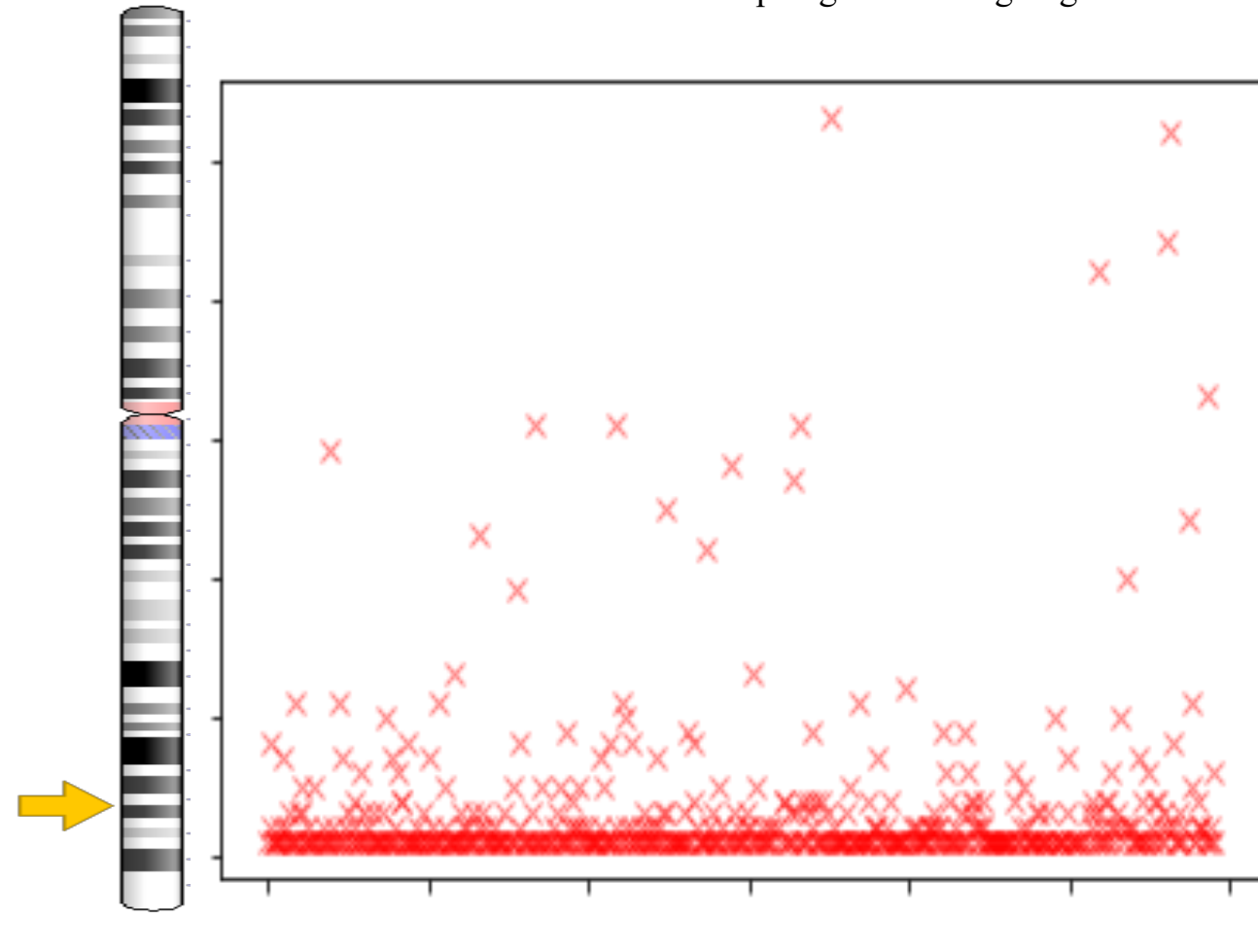
## 犯罪発生件数データ

From 東京都オープンデータ



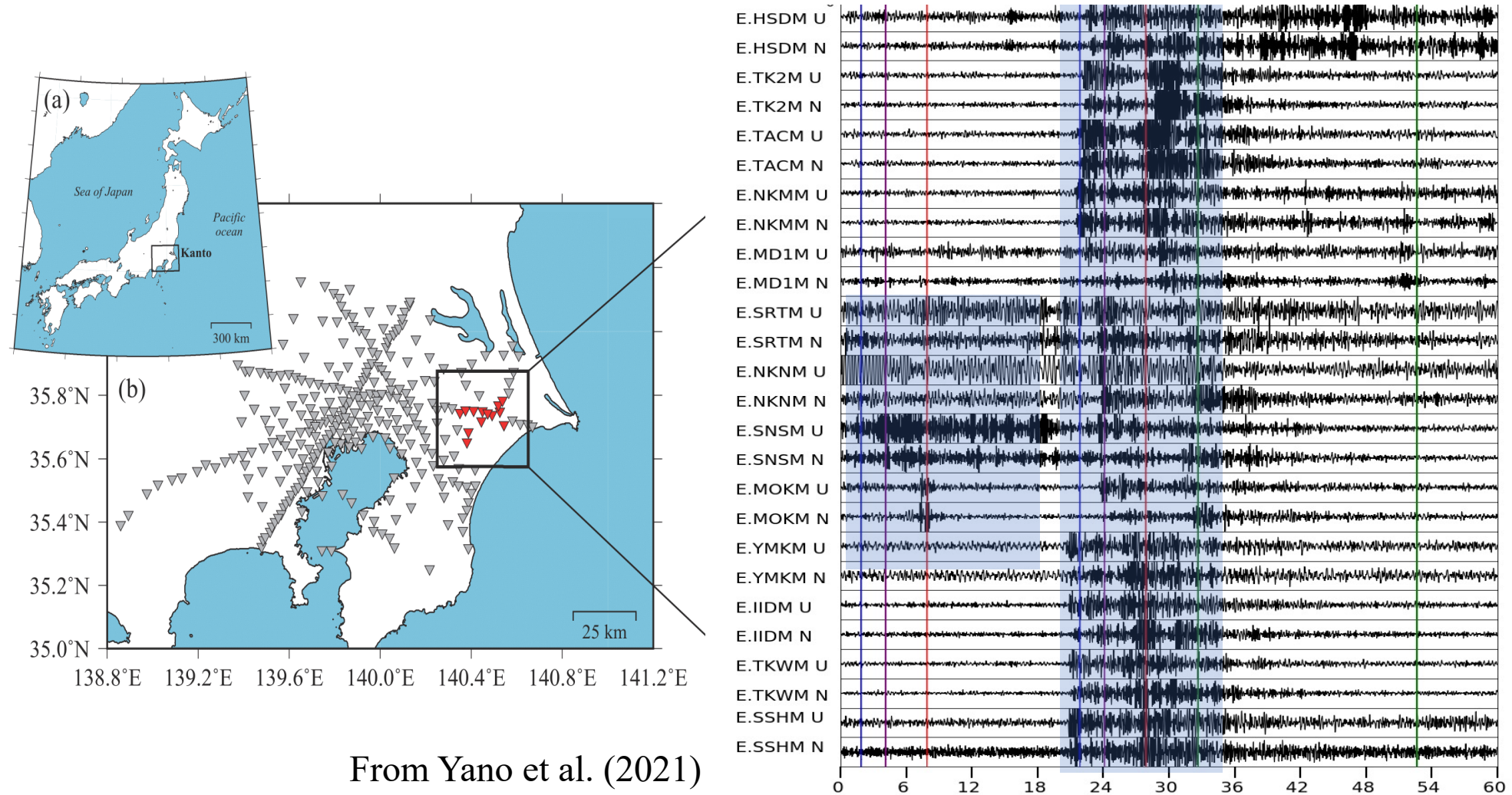
## 遺伝子発現データ

from <https://ghr.nlm.nih.gov/gene/PIK3CA>



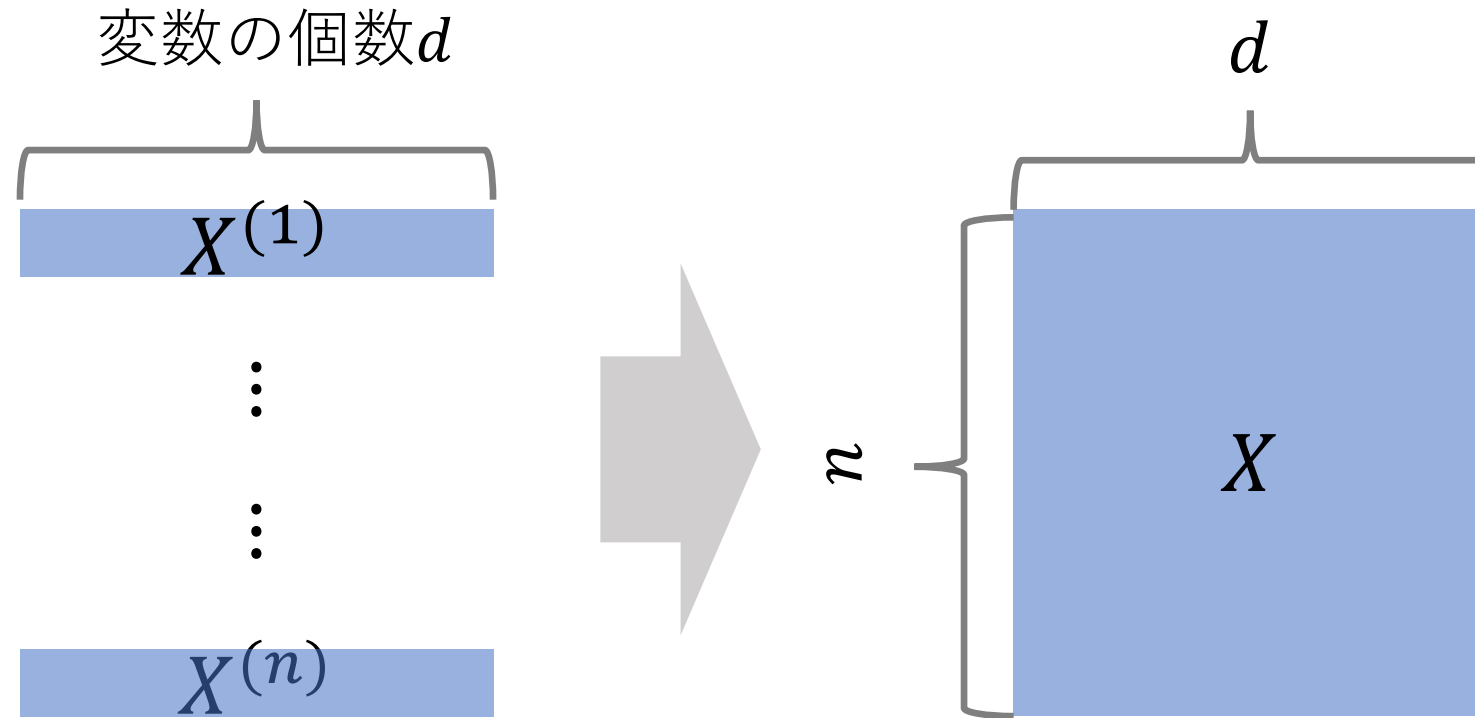
# 今日のデータ科学における「データ」

## 首都圏稠密地震観測網(MeSO-net)での連続波形記録



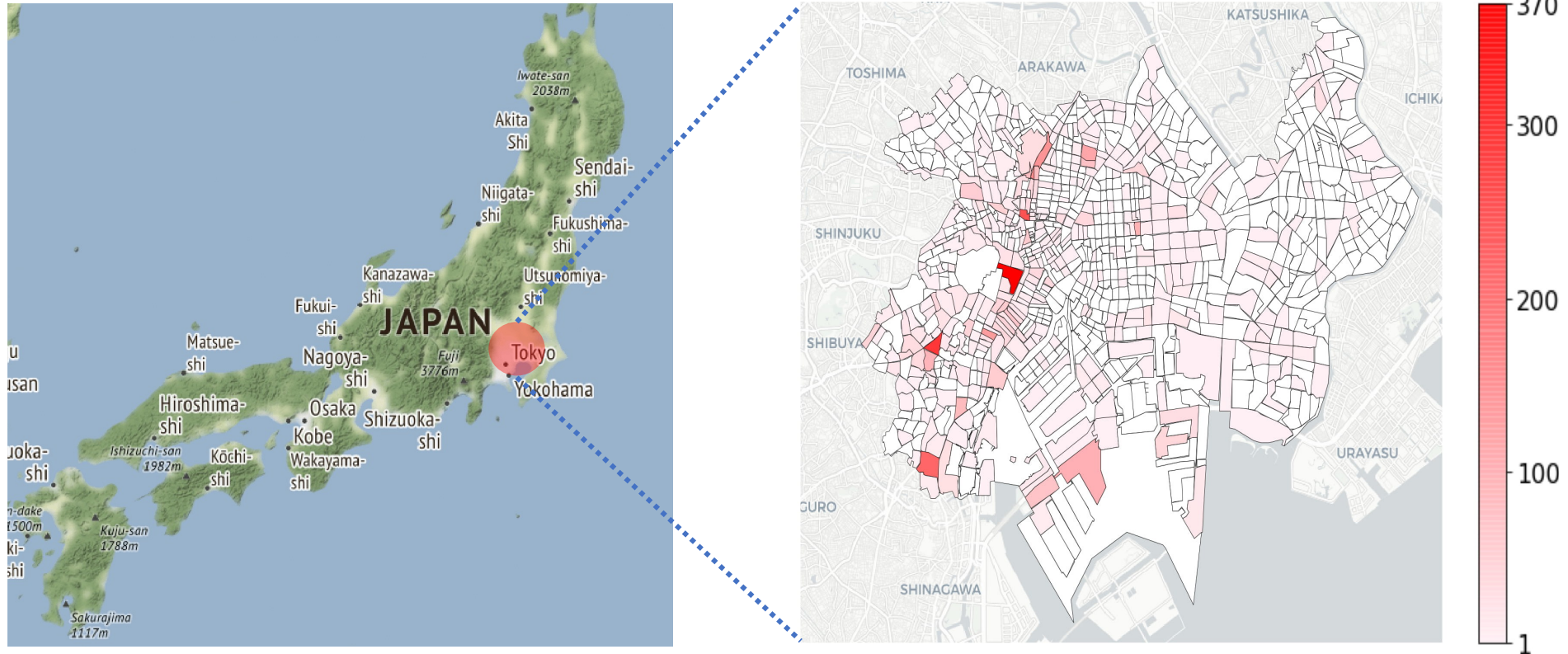
# データの整理

- 多種多様 (空間データ ・ 時間データ ・ 時空間データ ・ ・ ・)
- 「変数の個数」 (次元) と 「サンプルサイズ」 に注目して整理



# 犯罪発生件数マップ

観測  $X = (X_1, \dots, X_{978})$  : 東京8区の978町丁における犯罪発生件数

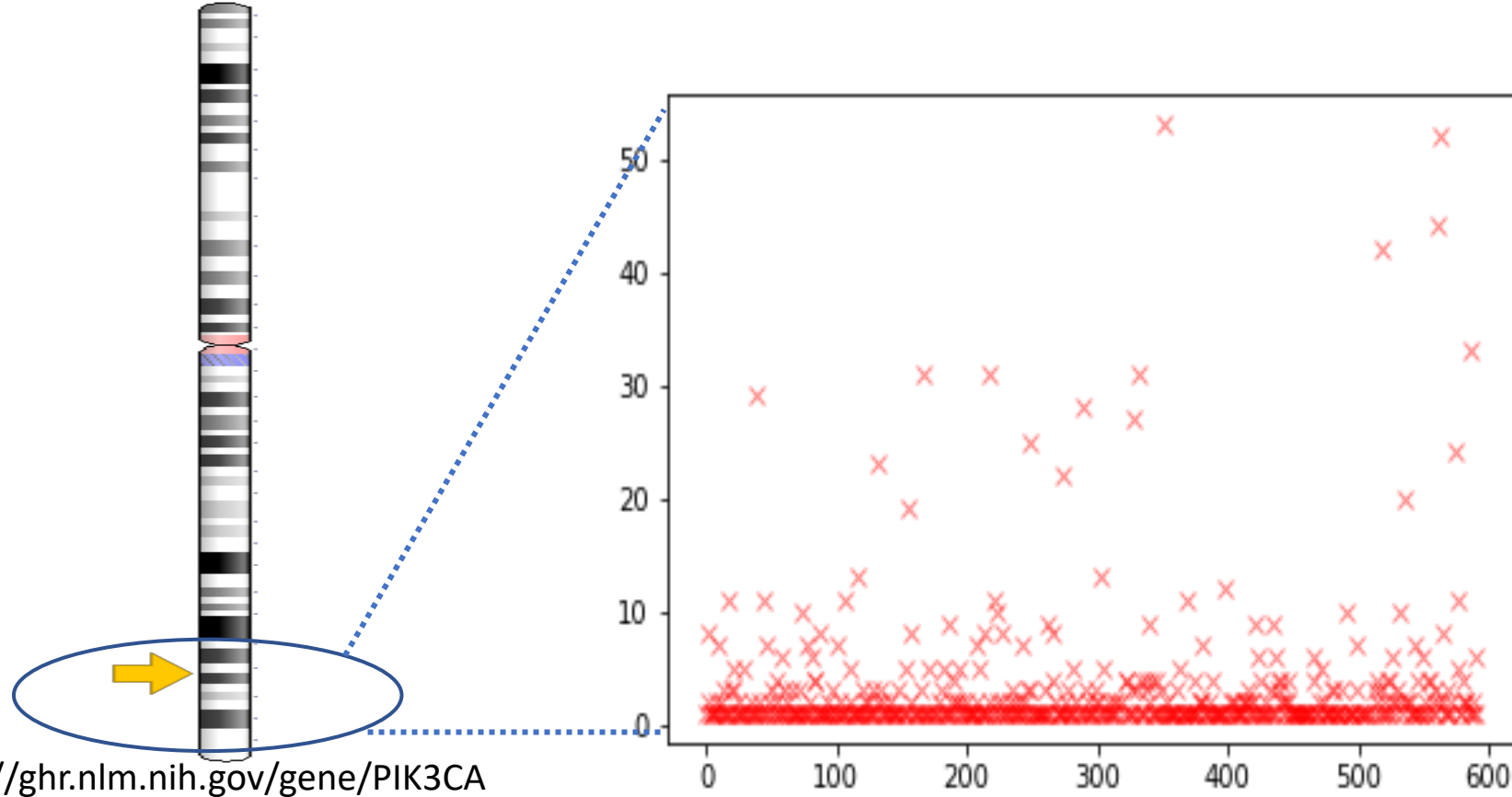


978個の変数が12年分存在

$$d = 978, n = 12$$

# 希少な遺伝子頻度

観測  $X = (X_1, \dots, X_{551})$  : 551のゲノム位置の希少遺伝子頻度

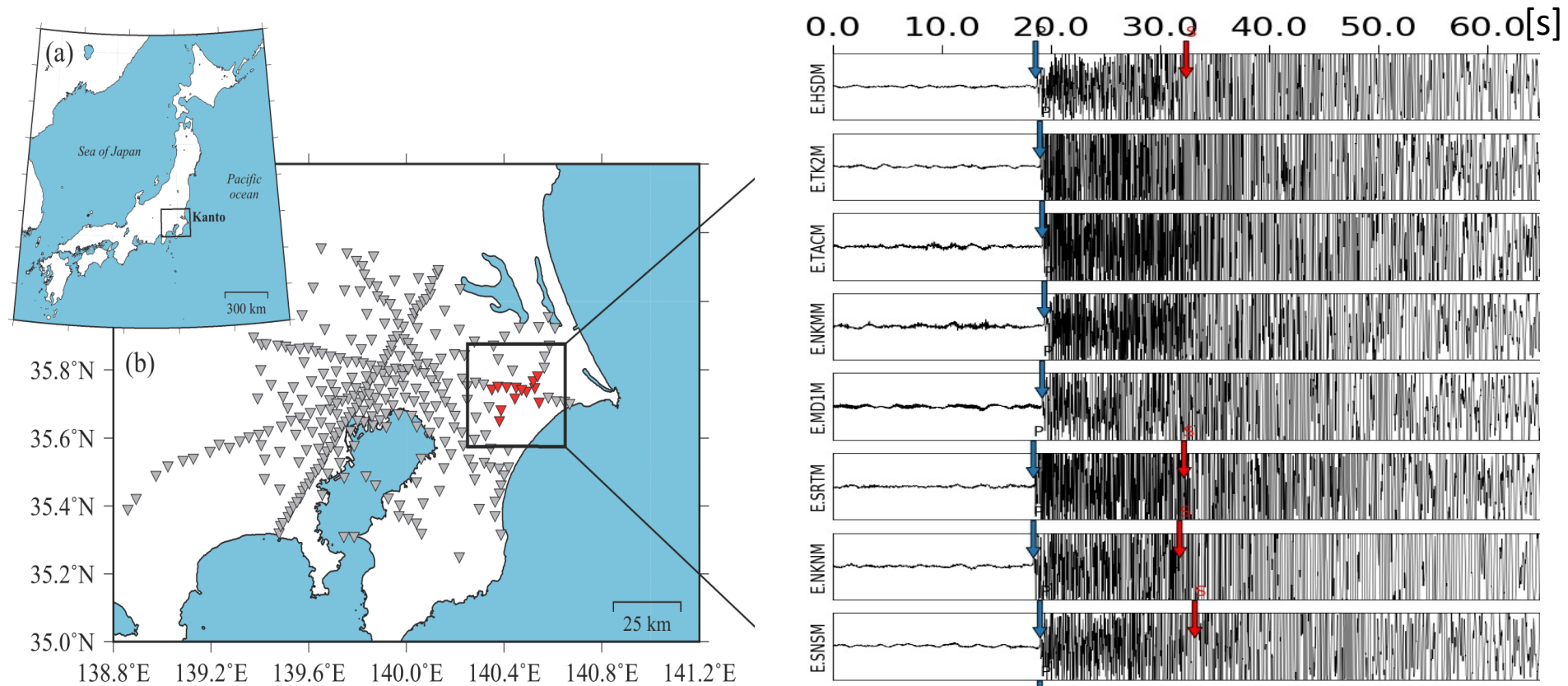


551個の変数が $10^4 \sim 10^5$ 組存在

$d = 551, n = 10^4 \sim 10^5$

# 首都圏地震観測網の連続波形

観測  $X(t) = (X_1(t), \dots, X_{296}(t))$  : 首都圏地震観測網の連続波形



0.005秒という高頻度で観測される→ 観測変数を連続な関数として扱える



# 「高次元」の定義

---

ここでの「高次元」の定義

## ■ 有限次元

変数の次元 固定, サンプルサイズ  $\rightarrow \infty$

## ■ 高次元

変数の次元  $\rightarrow \infty$

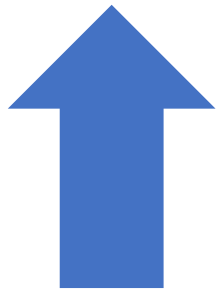
変数&サンプルサイズ  $\rightarrow \infty$

## ■ 無限次元

変数が関数自由度をもつ

# 本日の内容

可視化



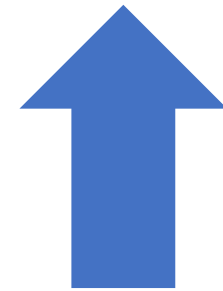
次元圧縮

発見・予測



正則化

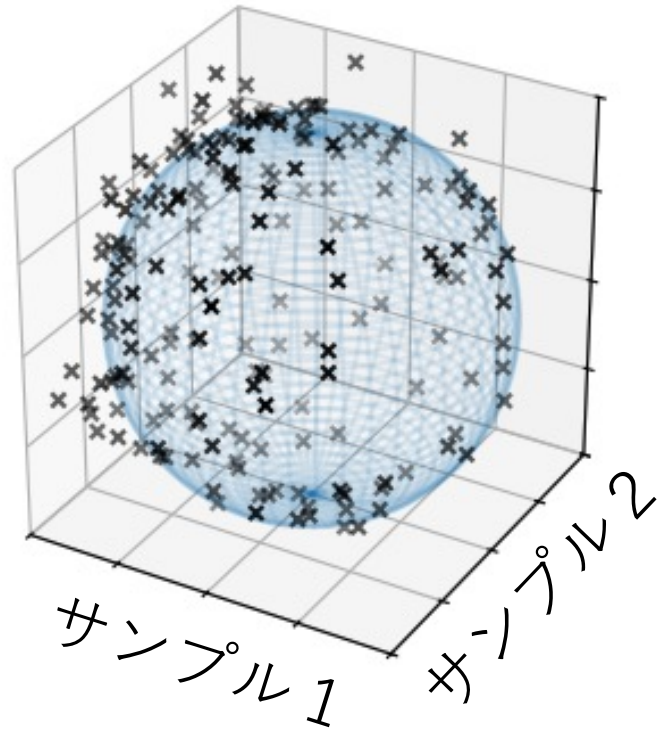
不確実性評価



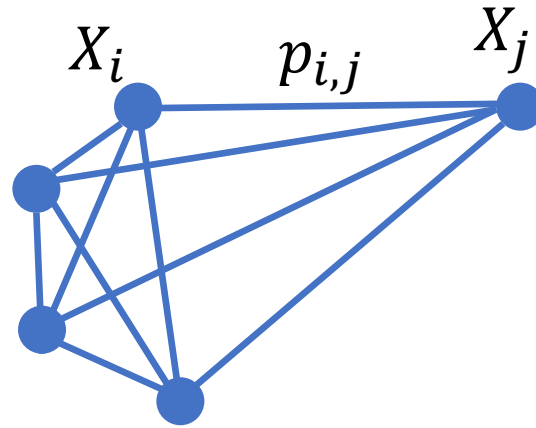
ベイズ

# 高次元のデータ解析のアイデア

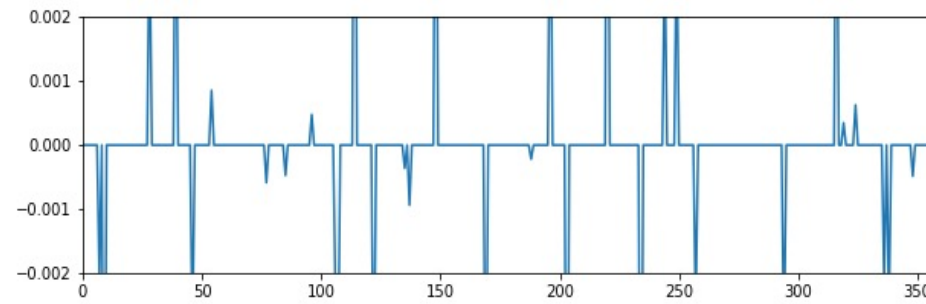
球面・軸集中現象



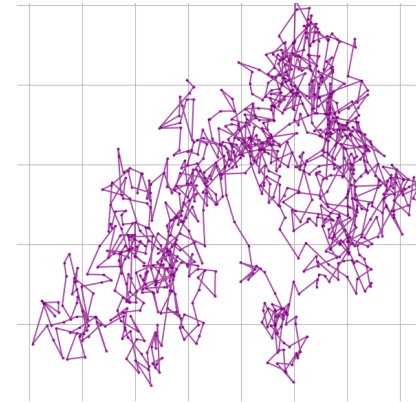
グラフ構造



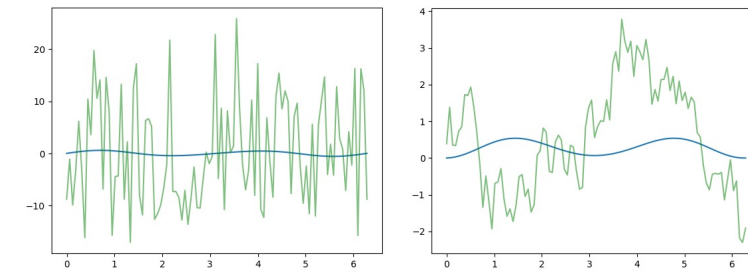
スパース性



ベイズの高次元挙動

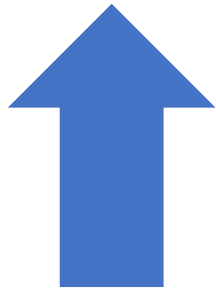


弱い位相の利用



# 本日の内容

可視化

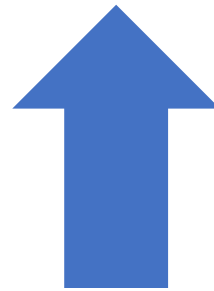


次元圧縮

主成分分析(線形次元圧縮)

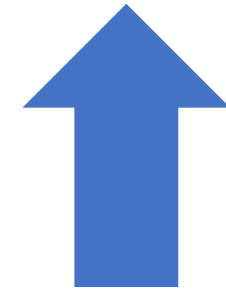
*t*-SNE,UMAP(非線形次元圧縮)

発見・予測



正則化

不確実性評価



ベイズ

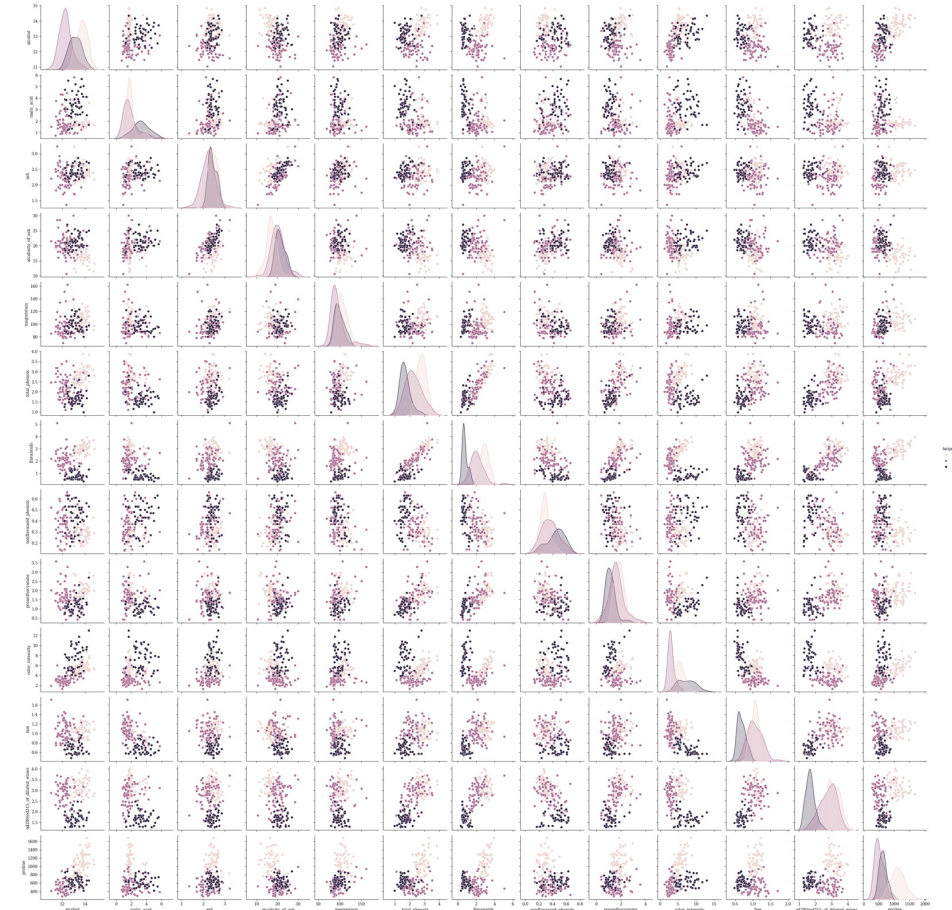
# 多変量データの可視化



From UCI Wine Dataset

変数名
アルコール度数
リンゴ酸
灰
灰のアルカリ度
マグネシウム
全フェノール
フラボノイド
非フラボノイドフェノール
プロアントシアニジン
色の濃さ
色相
希釈ワインの吸光度の比)
プロリン
ワインの点数

## ペアプロット(2変数散布図)+層別化



\*しばらく、主成分分析の復習になります

# 行列を使ったデータの整理

	アルコール度数	りんご酸	...	プロリン
1	14.23	1.71		1065
2	13.20	1.78		1050
173	14.13	4.10		560

# 特異値分解

The diagram illustrates the Singular Value Decomposition (SVD) of a matrix  $X$ . It shows the equation  $X = U \Sigma V^T$  where each term is represented by a colored square:  $X$  is a blue square,  $U$  is a red square,  $\Sigma$  is a gray square with a dashed orange diagonal line, and  $V^T$  is a yellow square. The matrices are arranged from left to right, separated by equals signs and multiplication symbols (@).

任意の行列  $X \in \mathbb{R}^{n \times d}$  は直交行列  $U \in \mathbb{R}^{n \times n}$ ,  $V \in \mathbb{R}^{d \times d}$  を用いて

$$X = U \Sigma V^T$$

\*  $\Sigma$  は対角行列  $\Delta$  とゼロ行列  $0$  で  $\Sigma = \begin{pmatrix} \Delta \\ 0 \end{pmatrix}$

# 特異値分解の見方

$$X = U @ \Sigma @ V^T$$

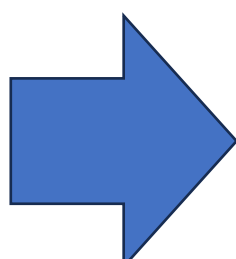
$$X = \sum_{i=1, \dots, r} \sigma_i u_i v_i^T$$



# 特異値分解の見方

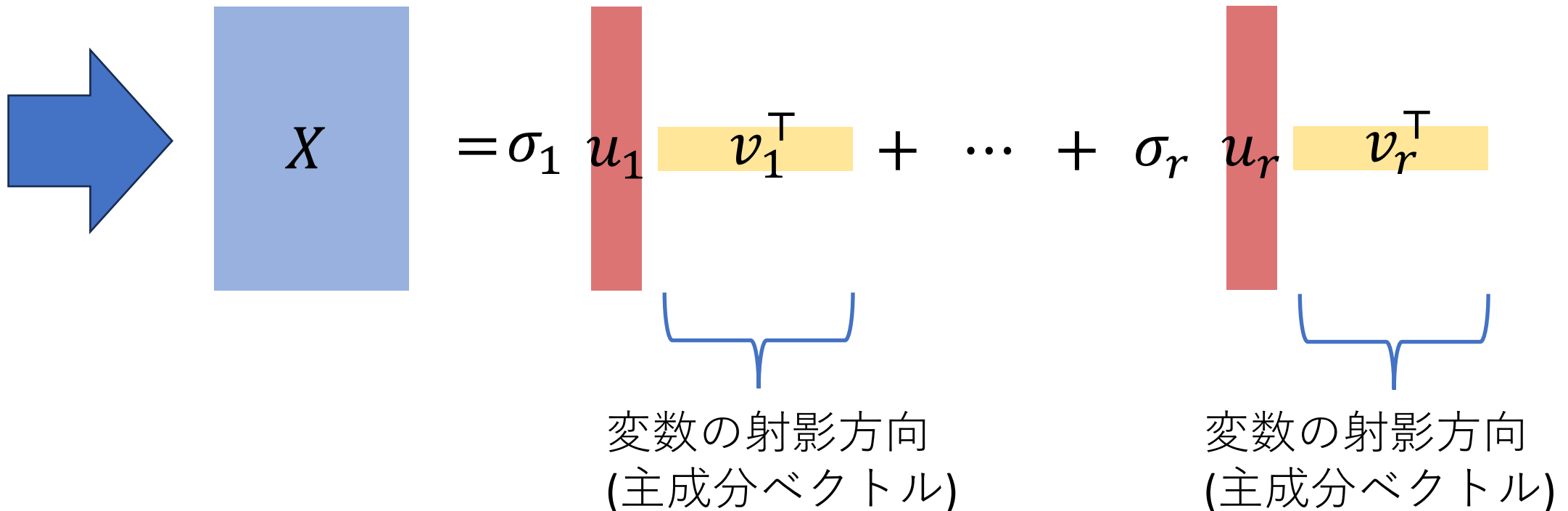
$$X = U @ \Sigma @ V^T$$

$$X = \sum_{i=1, \dots, r} \sigma_i u_i v_i^T$$


$$X = \sigma_1 u_1 v_1^T + \dots + \sigma_r u_r v_r^T$$

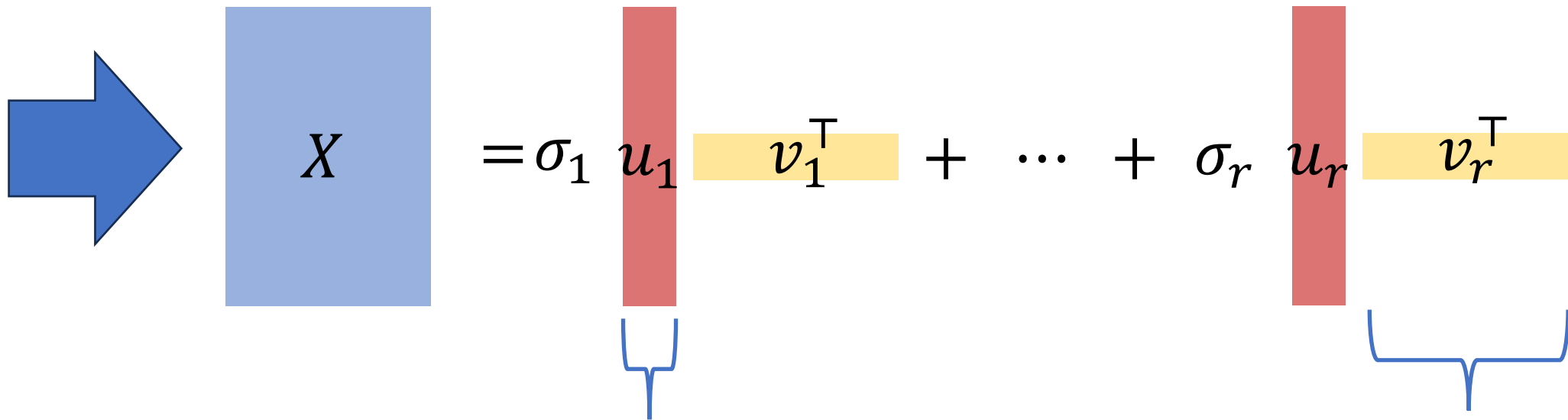
# 特異値分解における左特異ベクトル

$$X = \sum_{i=1, \dots, r} \sigma_i u_i v_i^T$$



# 特異値分解における右特異ベクトル

$$X = \sum_{i=1, \dots, r} \sigma_i u_i v_i^T$$

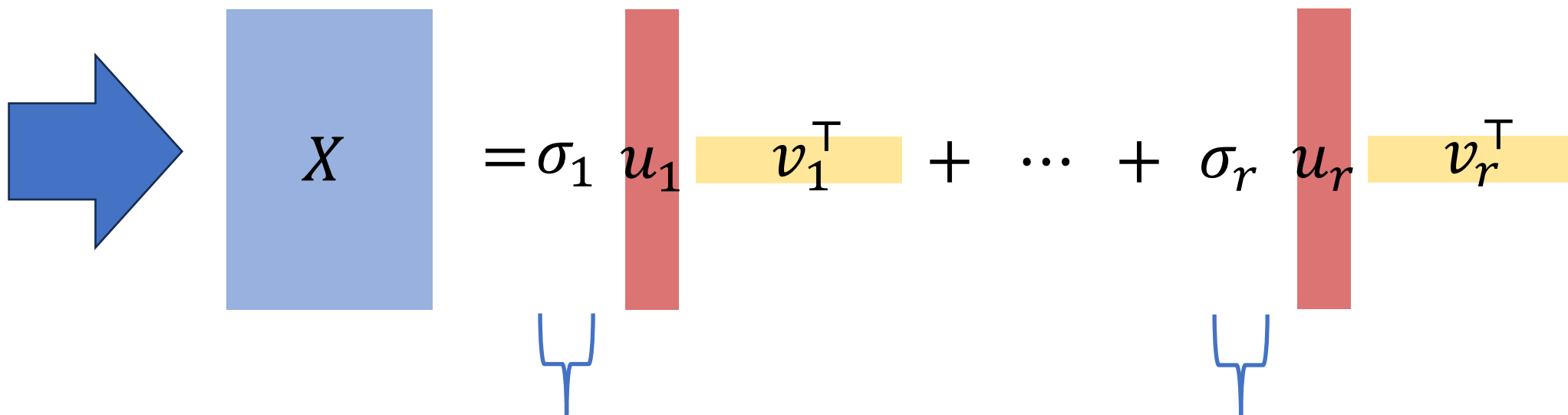


各サンプルの $v_1$ 方向の正規化座標値

各サンプルの $v_r$ 方向の正規化座標値

# 特異値分解における特異値

$$X = \sum_{i=1, \dots, r} \sigma_i u_i v_i^T$$



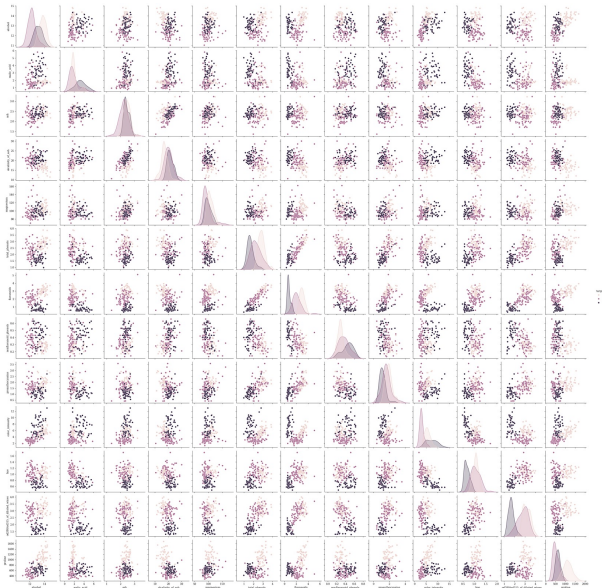
$\sigma_1^2$ :  $v_1$ 方向のサンプルのばらつき

$\sigma_2^2$ :  $v_1$ 方向のサンプルのばらつき

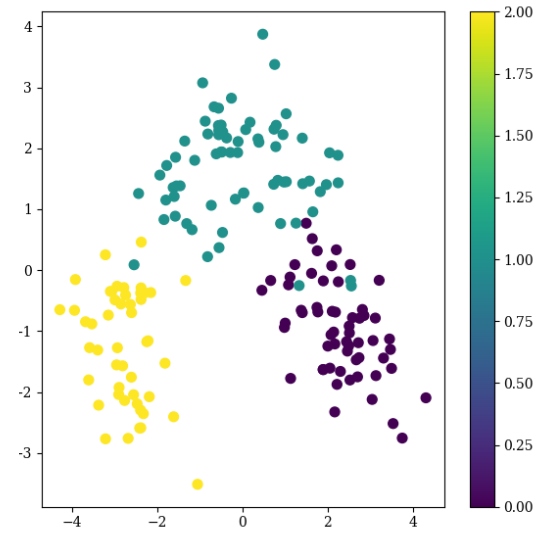
# 主成分分析 (Principal Component Analysis; PCA)

特異値の絶対値が大きい成分を使ったデータの次元縮約

$$X = \sigma_1 u_1 v_1^T + \dots + \sigma_r u_r v_r^T$$



$X_i^T v_2$ : 第二主成分



$X_i^T v_1$ : 第一主成分

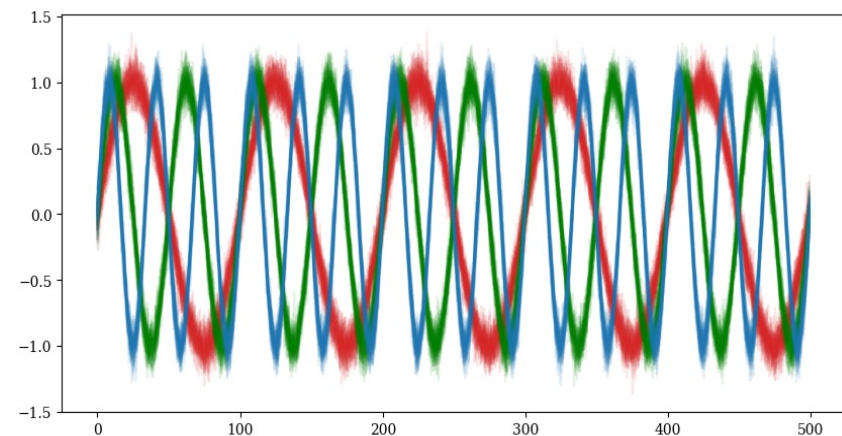
# PCAの適用例：波形特徴量の抽出

波形データや画像データに適用し、特徴量抽出を行うこともできる

データ行列

長さ500の150個の波形

$$n = 150 \ll d = 500$$



周波数が異なる三つの波形

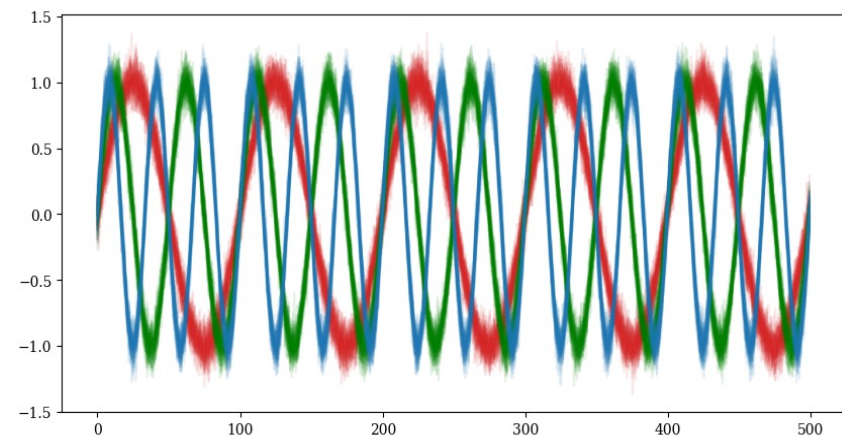
# PCAの適用例：波形特徴量の抽出

波形データや画像データに適用し、特徴量抽出を行うこともできる

## データ行列

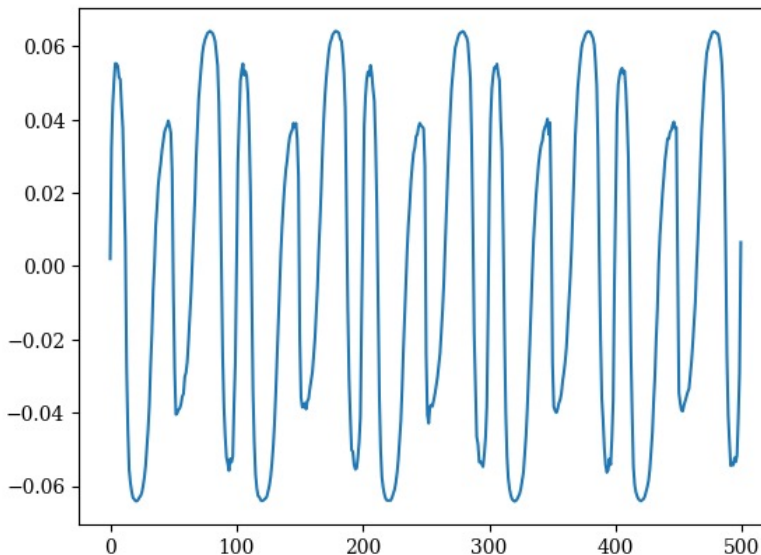
長さ500の150個の波形

$$n = 150 \ll d = 500$$



周波数が異なる三つの波形

## 主成分ベクトル



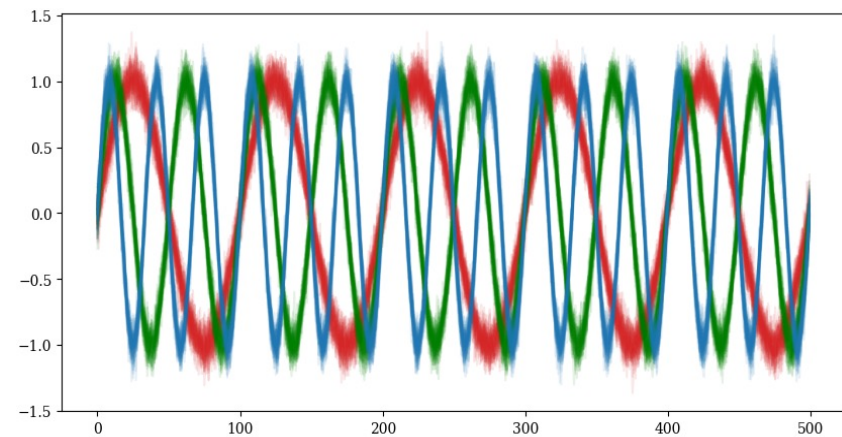
# PCAの適用例：波形特徴量の抽出

波形データや画像データに適用し、特徴量抽出を行うこともできる

## データ行列

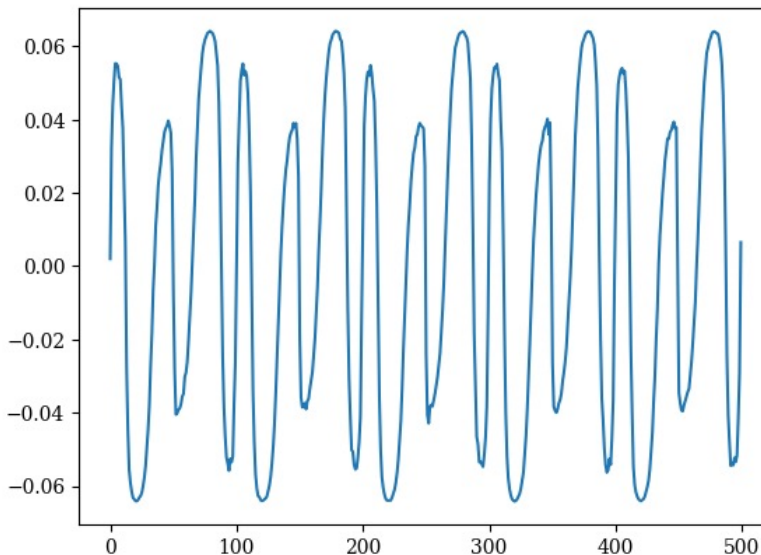
長さ500の150個の波形

$$n = 150 \ll d = 500$$

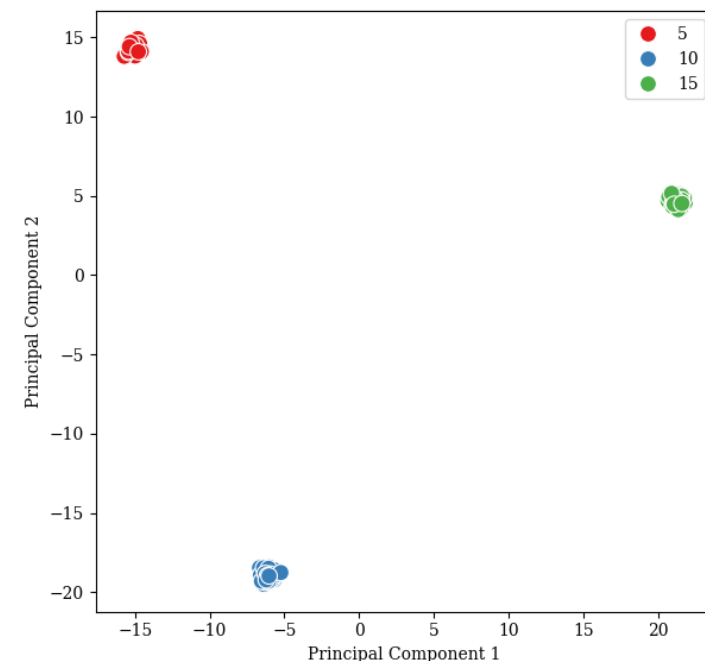


周波数が異なる三つの波形

## 主成分ベクトル



## 主成分空間



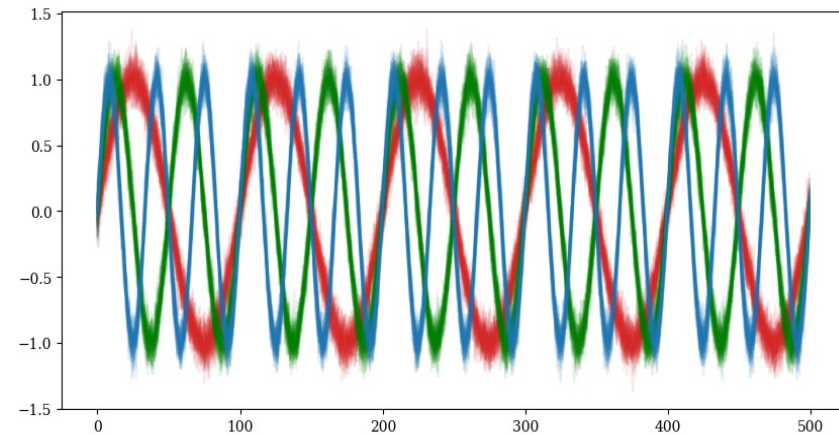


# PCAの適用例：波形特徴量の抽出

波形データや画像データに適用し、特徴量抽出を行うこともできる

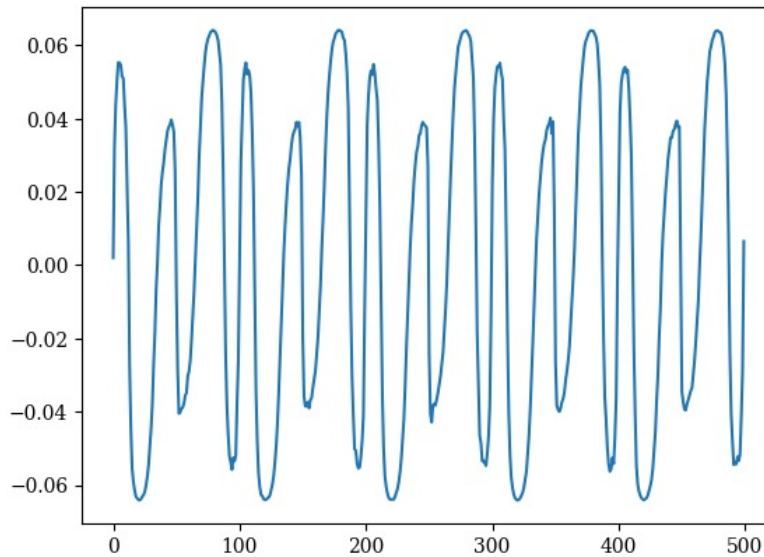
## データ行列

長さ500の150個の波形  
 $n = 150 \ll d = 500$

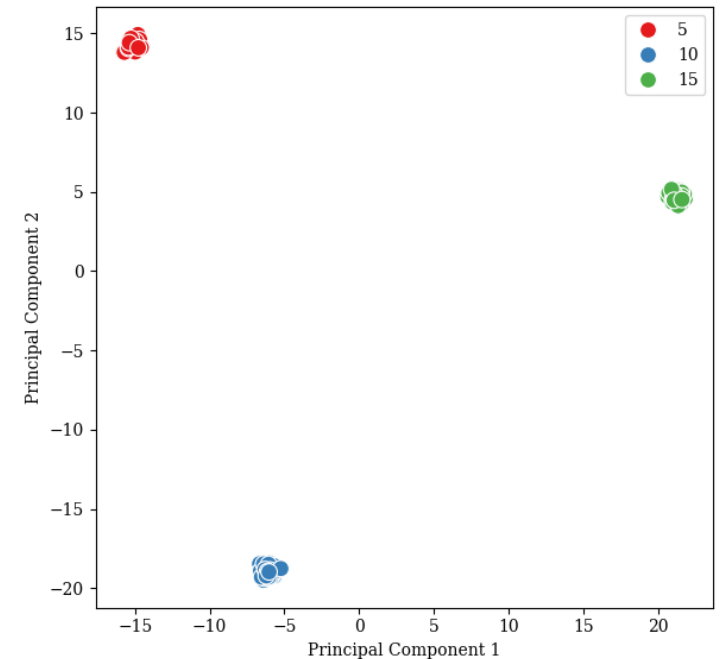


周波数が異なる三つの波形

## 主成分ベクトル



## 主成分空間



**高次元の設定 ( $d \gg n$ )** でも主成分空間で分離されている

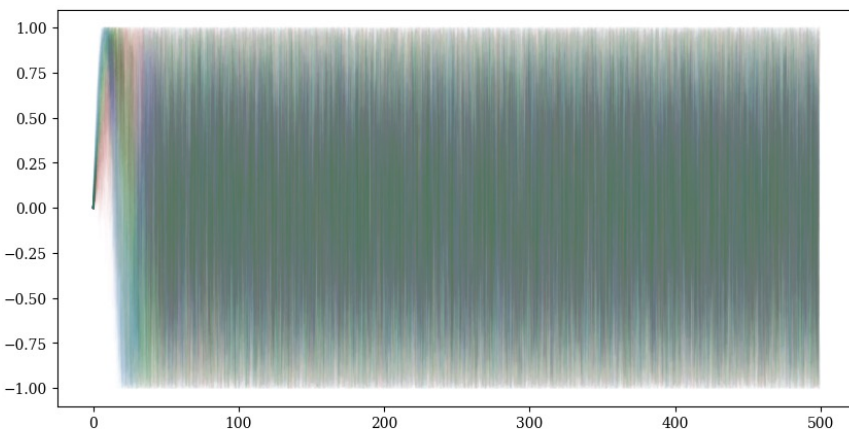
# PCAの適用例：波形特徴量の抽出（難しい例）

波形データや画像データに適用し、特徴量抽出を行うこともできる

## データ行列

150個の長さ500の波形

$$n = 150 \ll d = 500$$



先例にランダム変調

⇒ 周波数 + ガウスノイズ

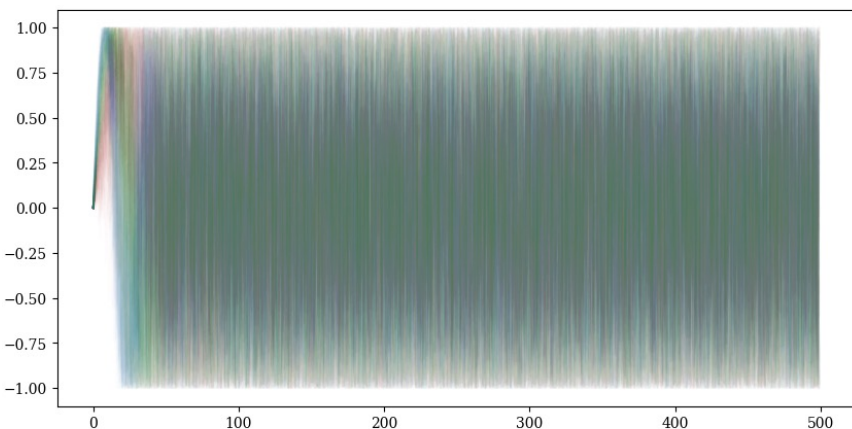
# PCAの適用例：波形特徴量の抽出（難しい例）

波形データや画像データに適用し、特徴量抽出を行うこともできる

## データ行列

150個の長さ500の波形

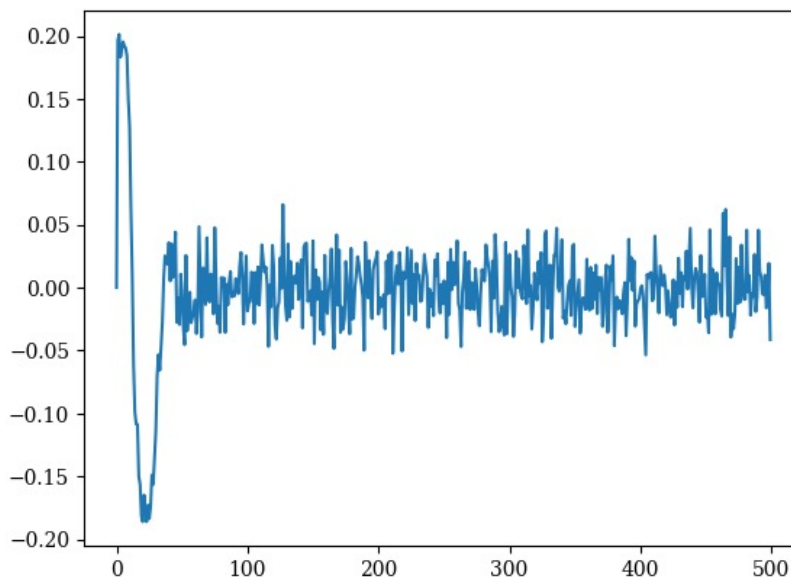
$$n = 150 \ll d = 500$$



先例にランダム変調

⇒ 周波数 + ガウスノイズ

## 主成分ベクトル



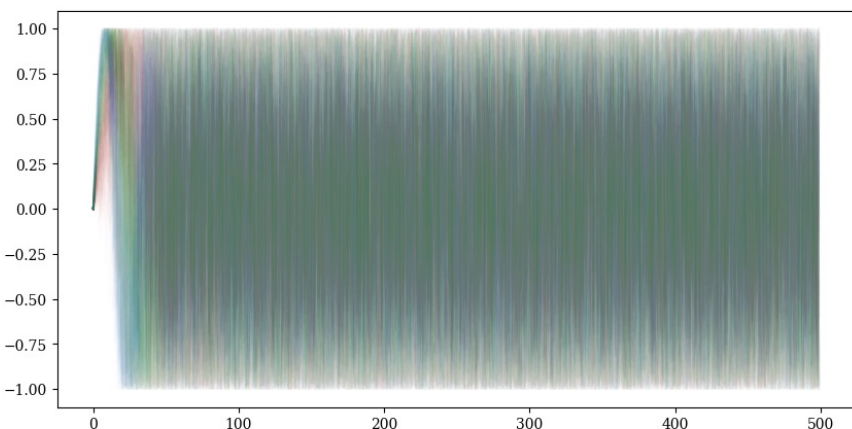
# PCAの適用例：波形特徴量の抽出（難しい例）

波形データや画像データに適用し、特徴量抽出を行うこともできる

## データ行列

150個の長さ500の波形

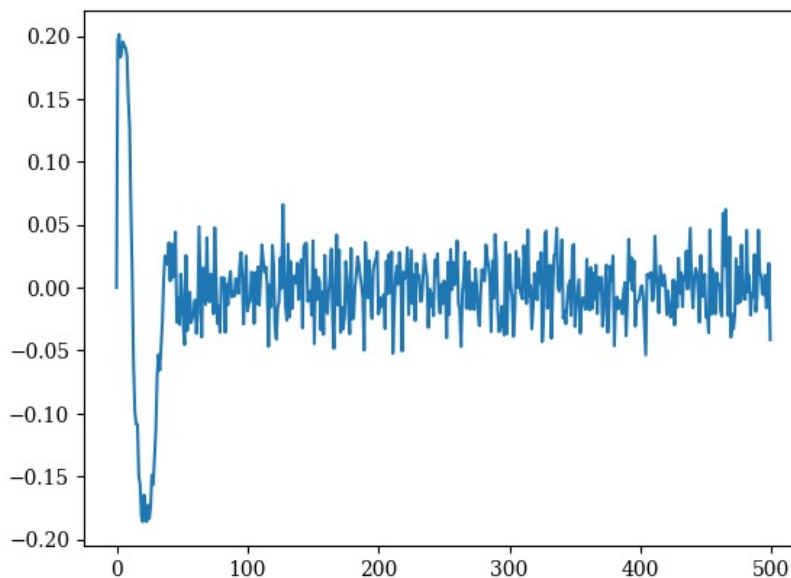
$$n = 150 \ll d = 500$$



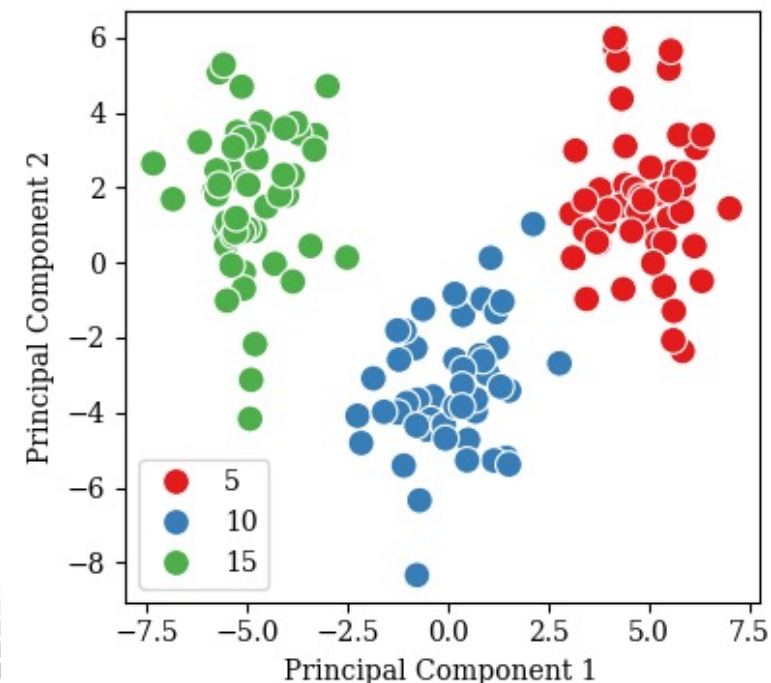
先例にランダム変調

⇒ 周波数 + ガウスノイズ

## 主成分ベクトル



## 主成分空間



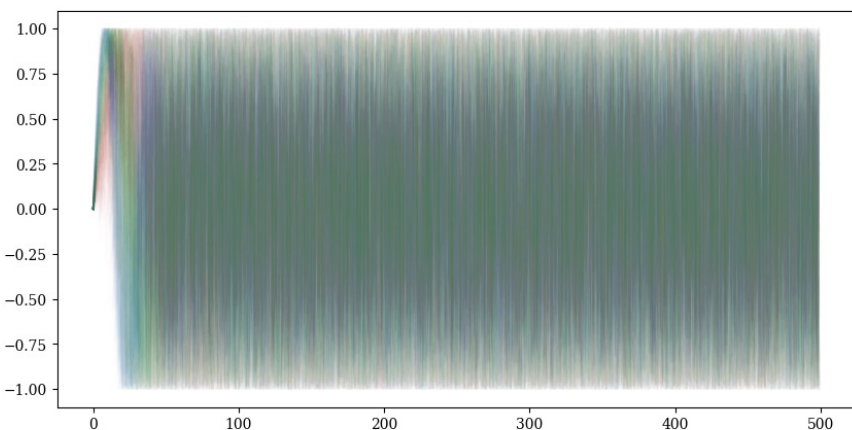
# PCAの適用例：波形特徴量の抽出（難しい例）

波形データや画像データに適用し、特徴量抽出を行うこともできる

## データ行列

150個の長さ500の波形

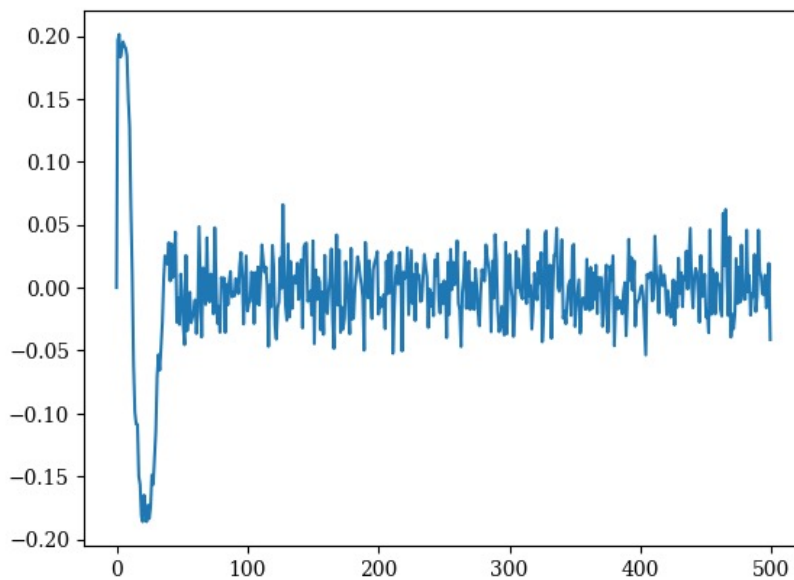
$$n = 150 \ll d = 500$$



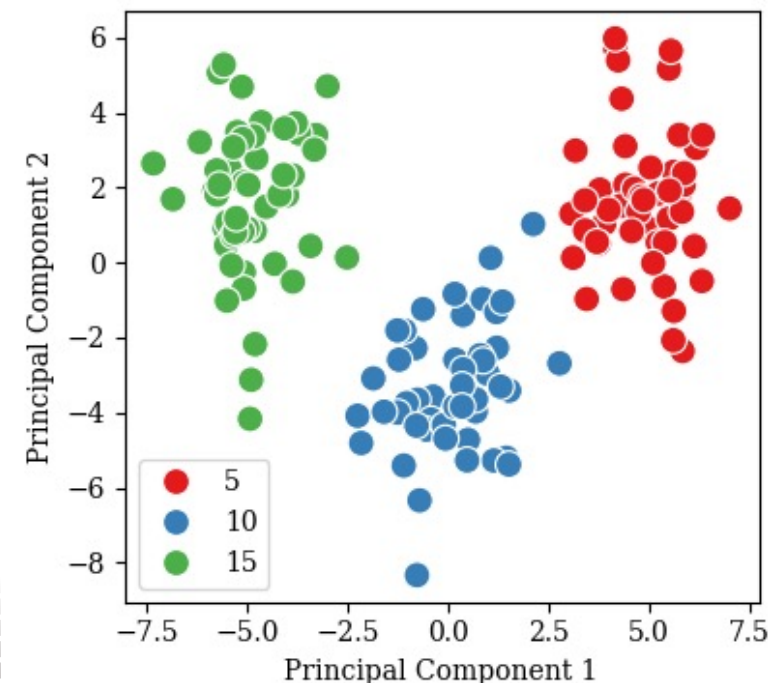
先例にランダム変調

⇒ 周波数 + ガウスノイズ

## 主成分ベクトル



## 主成分空間

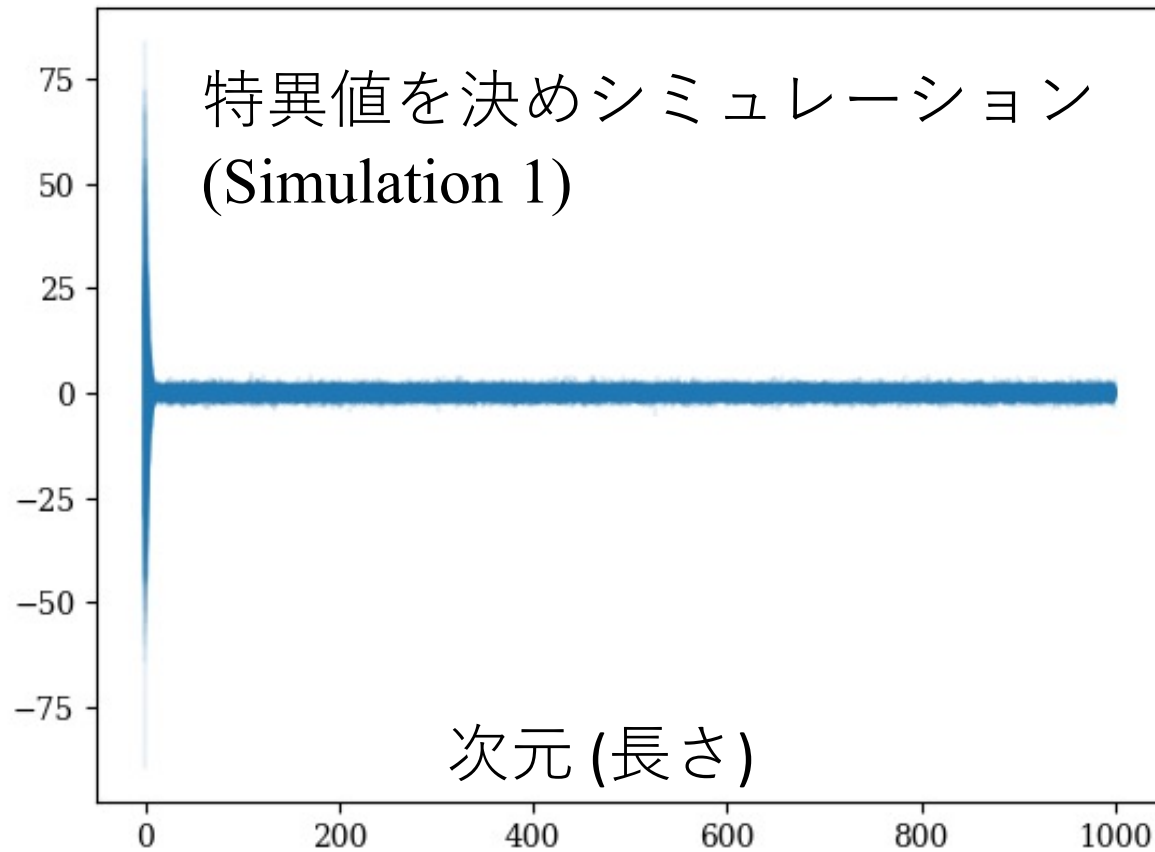


**高次元の設定 ( $d \gg n$ ) + 複雑な設定**でも一見うまくいっている

実は、**ナイーブな主成分分析**は高次元では注意が必要

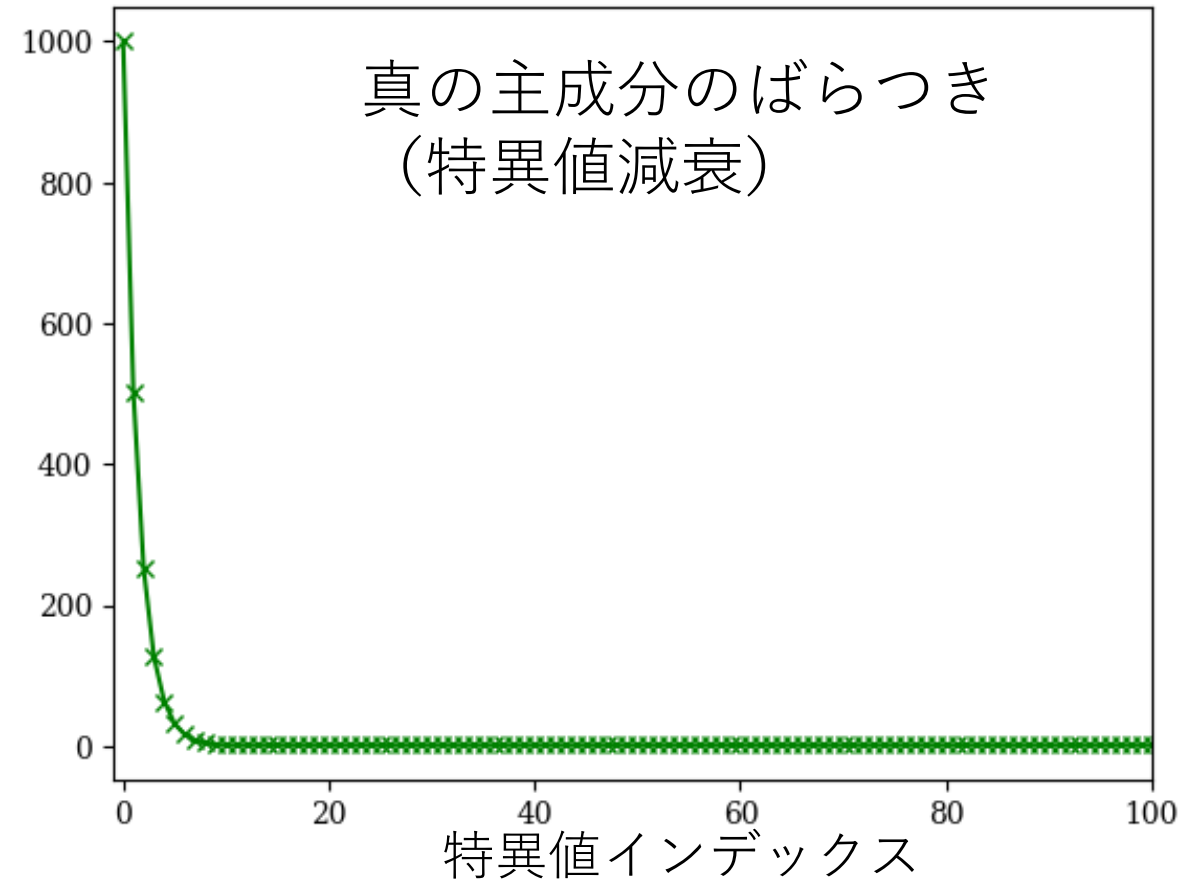
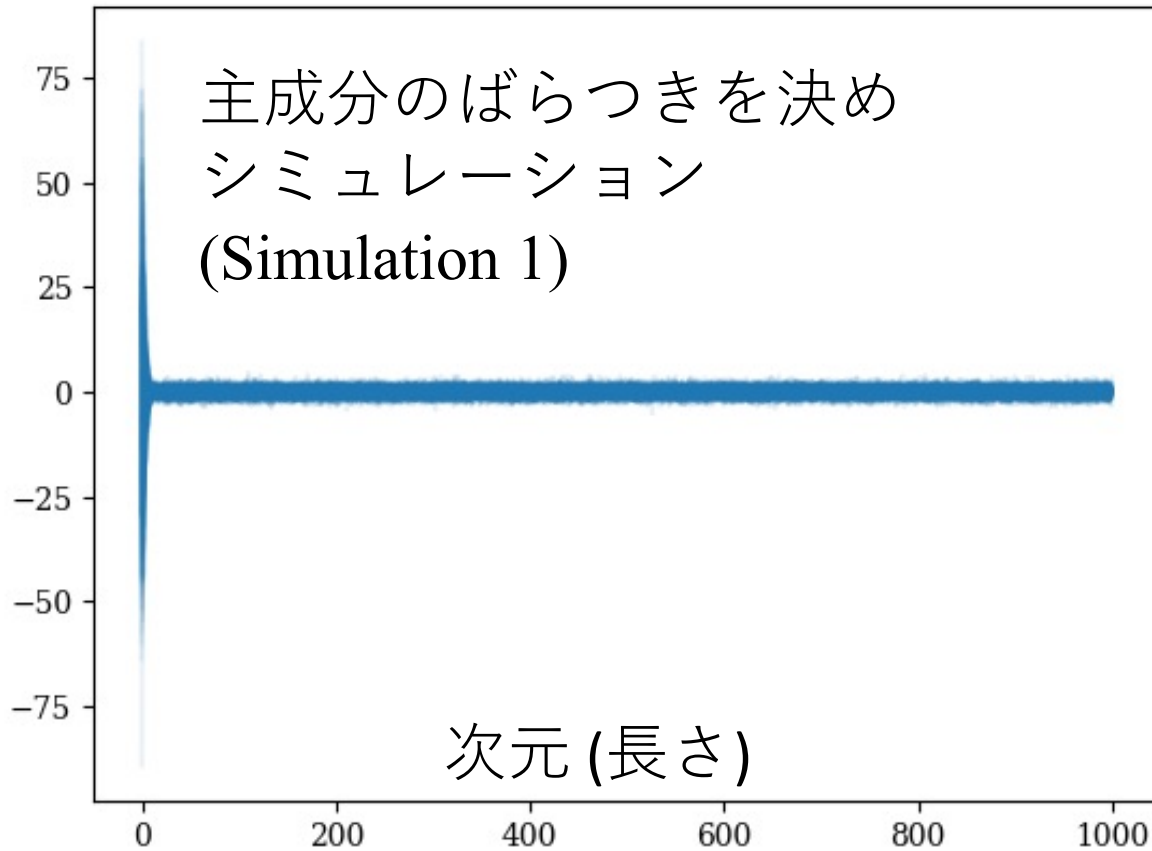
# 高次元からの挑戦状

二通りのシナリオでそのことを検証 (Simulation 1)



# 高次元からの挑戦状

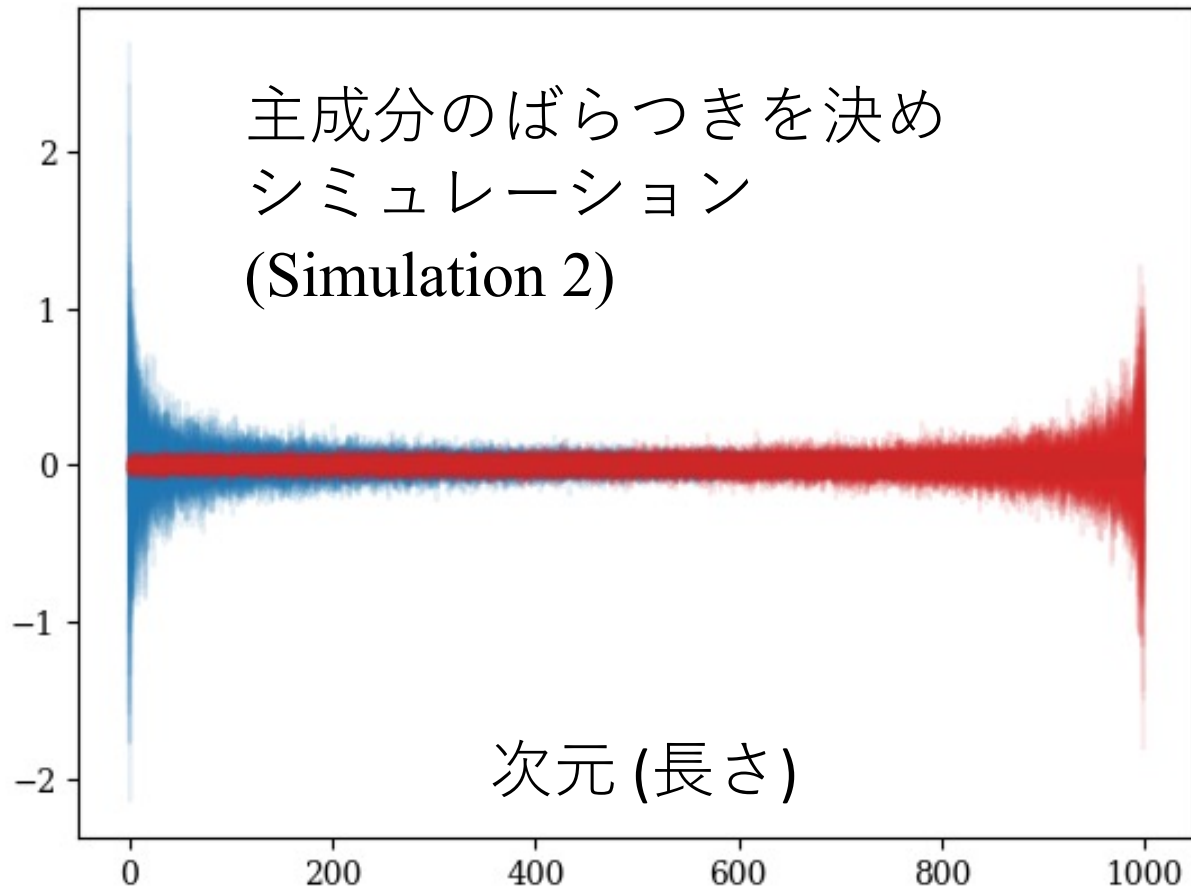
二通りのシナリオでそのことを検証 (Simulation 1)





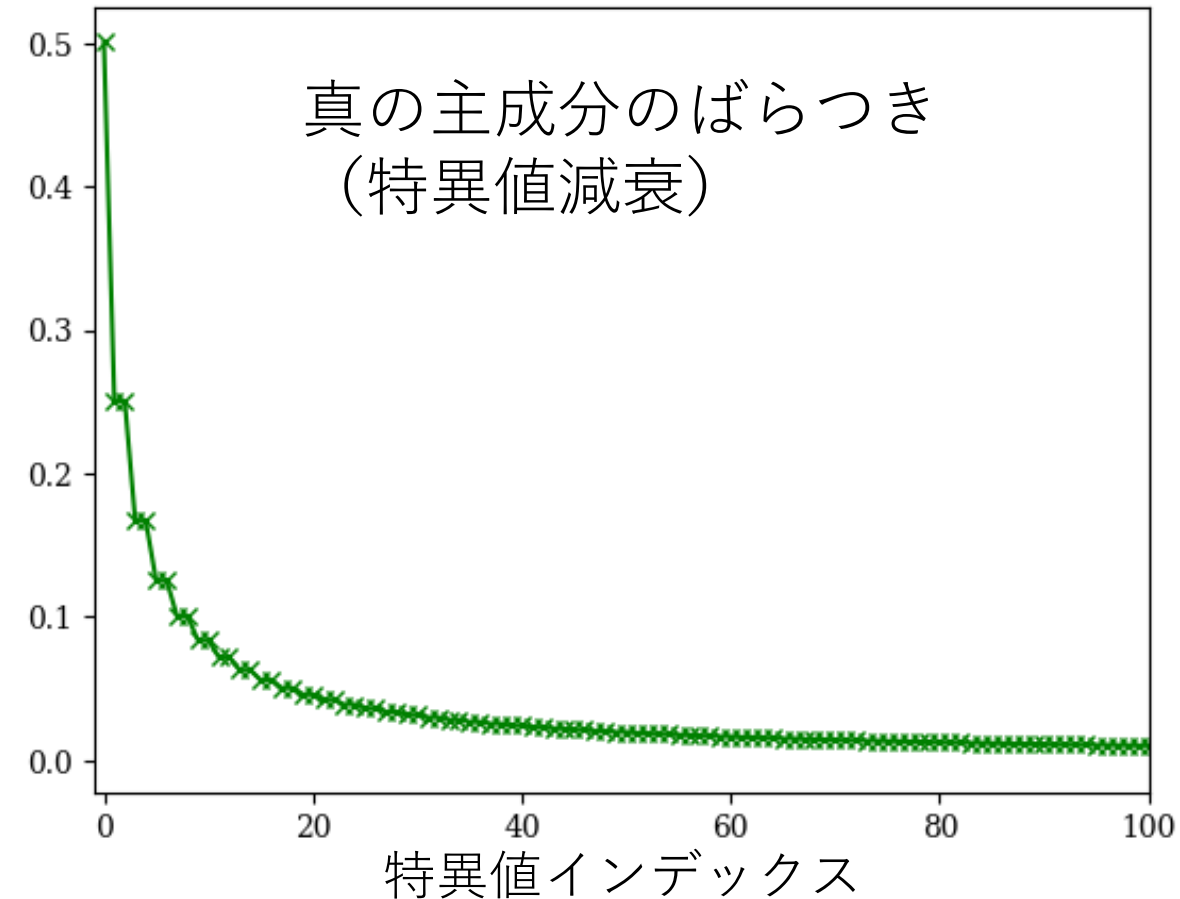
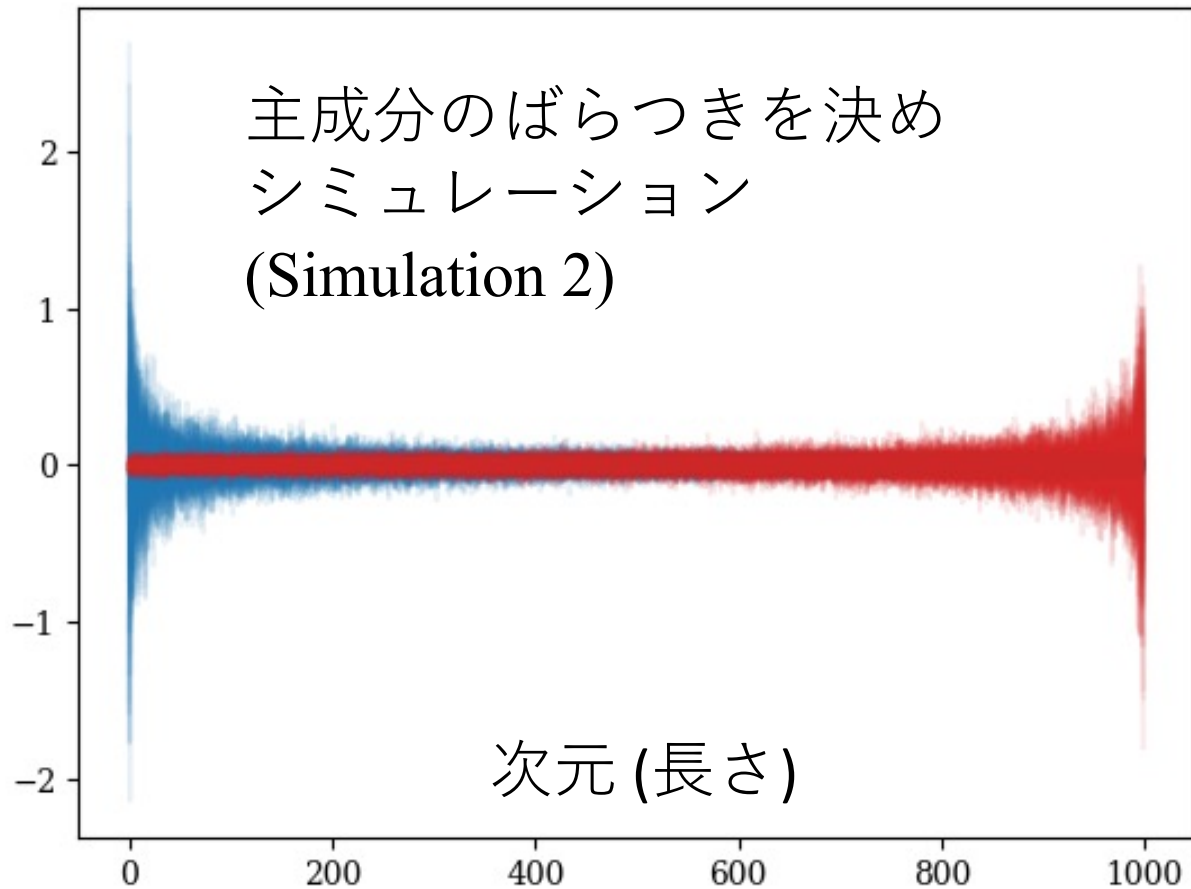
# 高次元からの挑戦状

二通りのシナリオでそのことを検証 (Simulation 2)



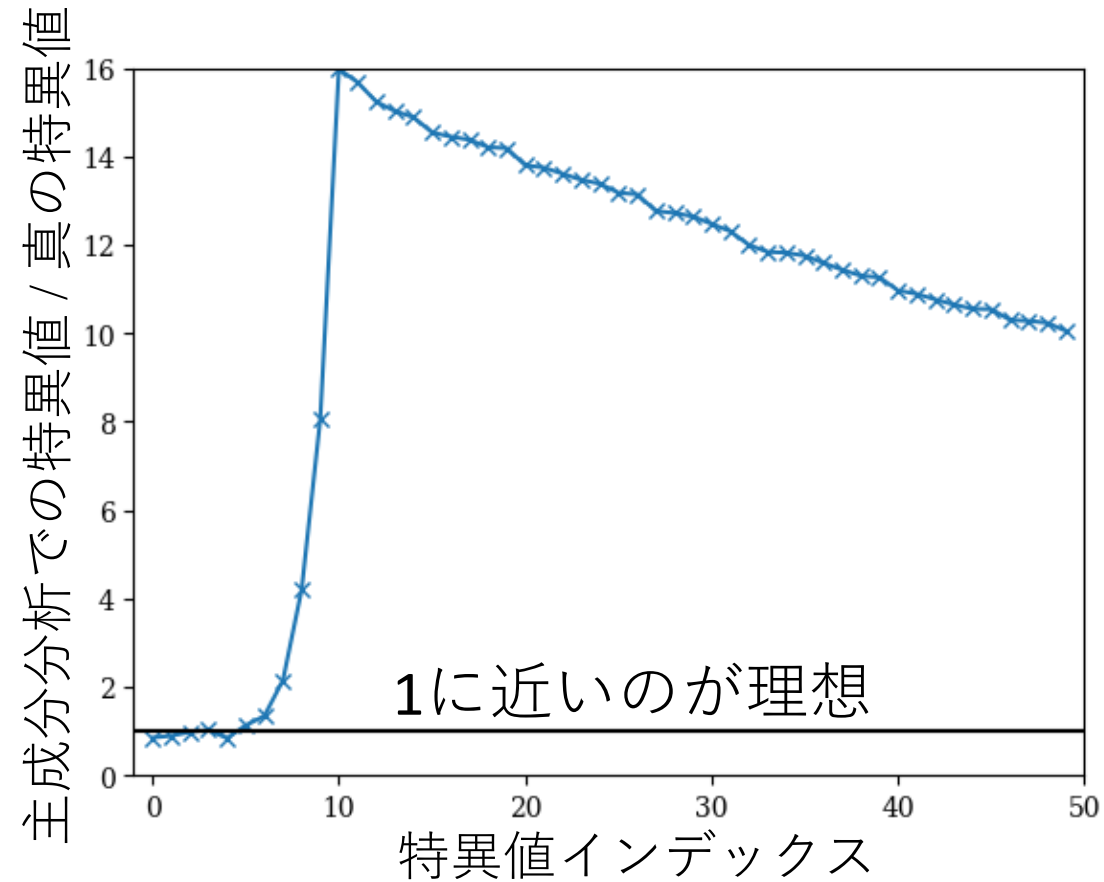
# 高次元からの挑戦状

二通りのシナリオでそのことを検証 (Simulation 2)

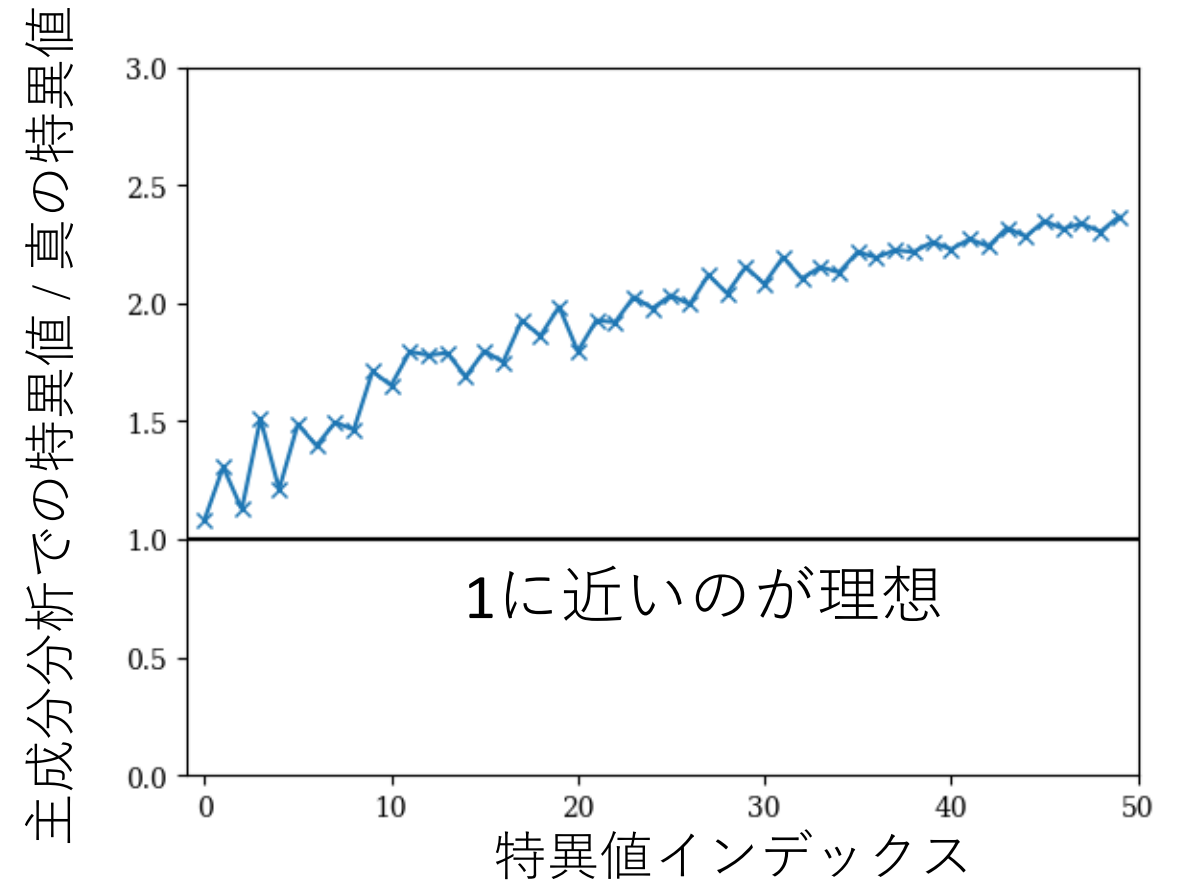


# ナイーブな主成分分析の結果

## Simulation 1の結果

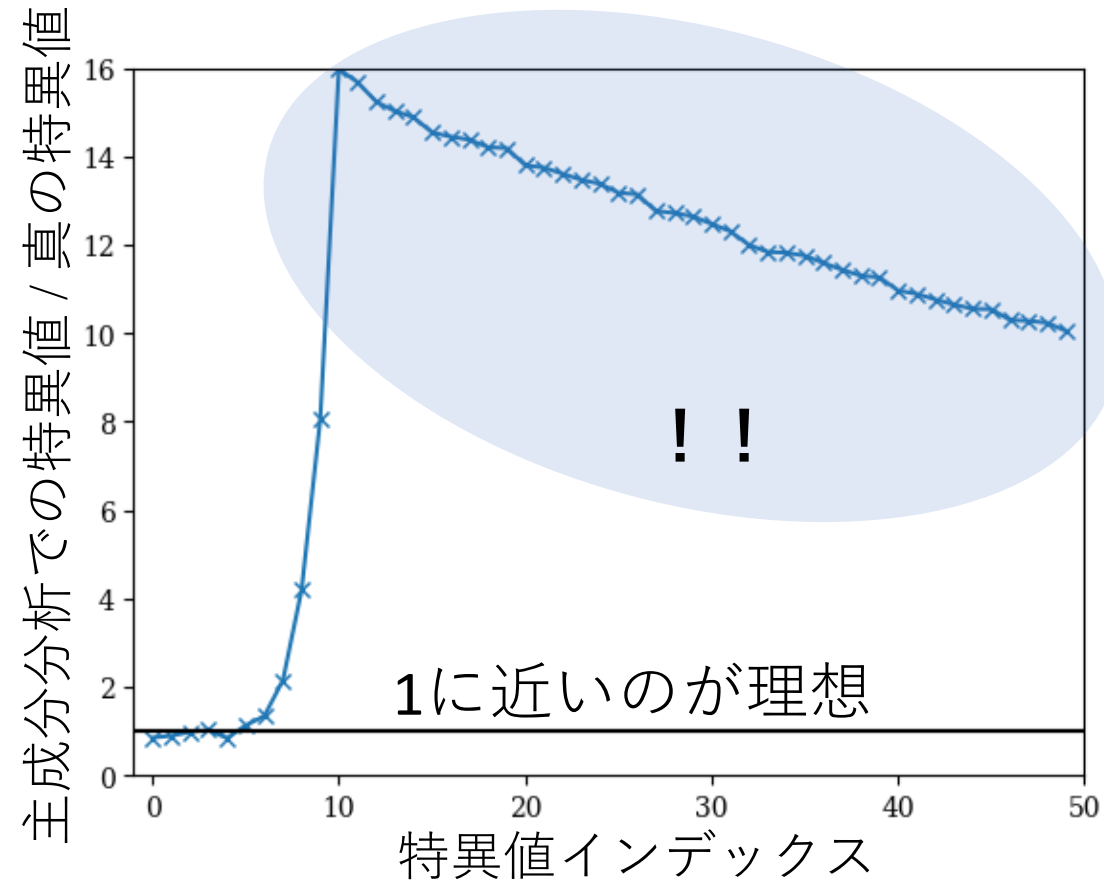


## Simulation 2の結果

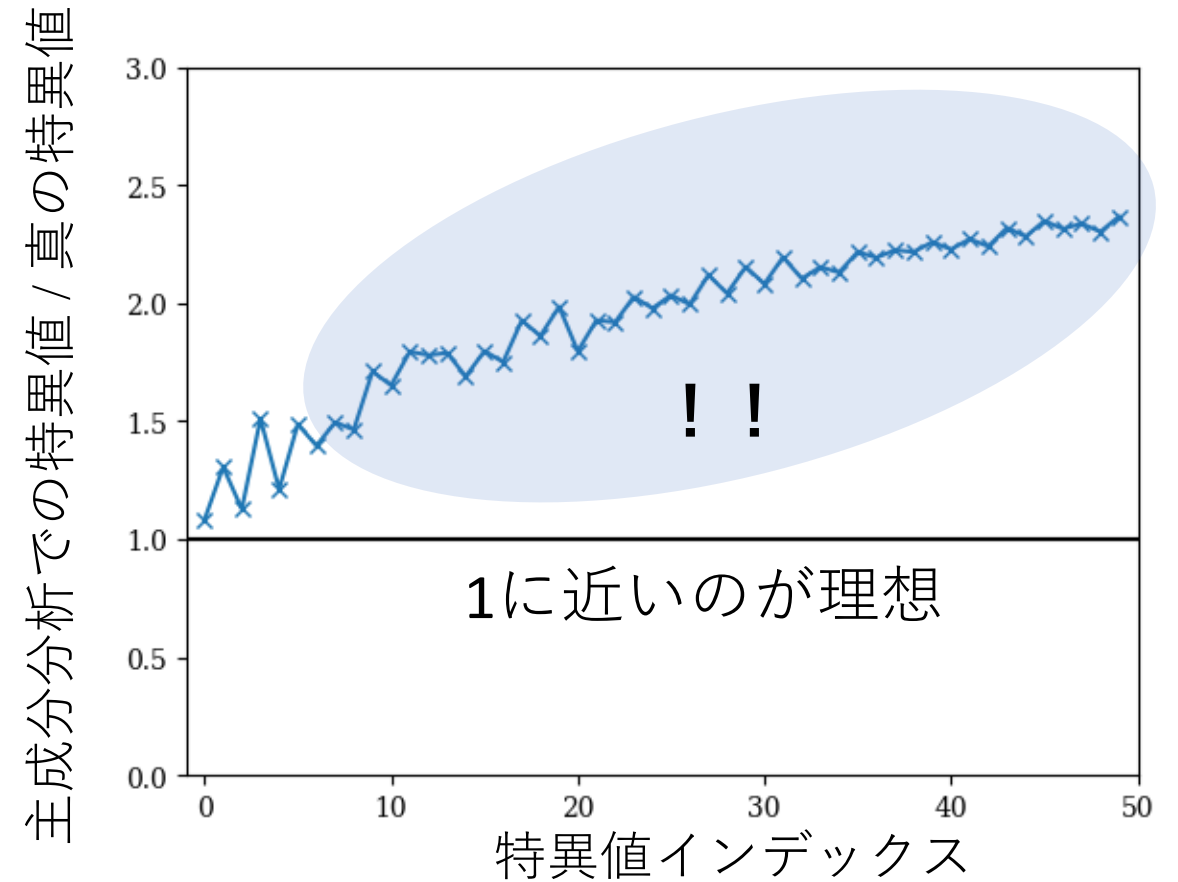


# ナイーブな主成分分析の結果

## Simulation 1の結果

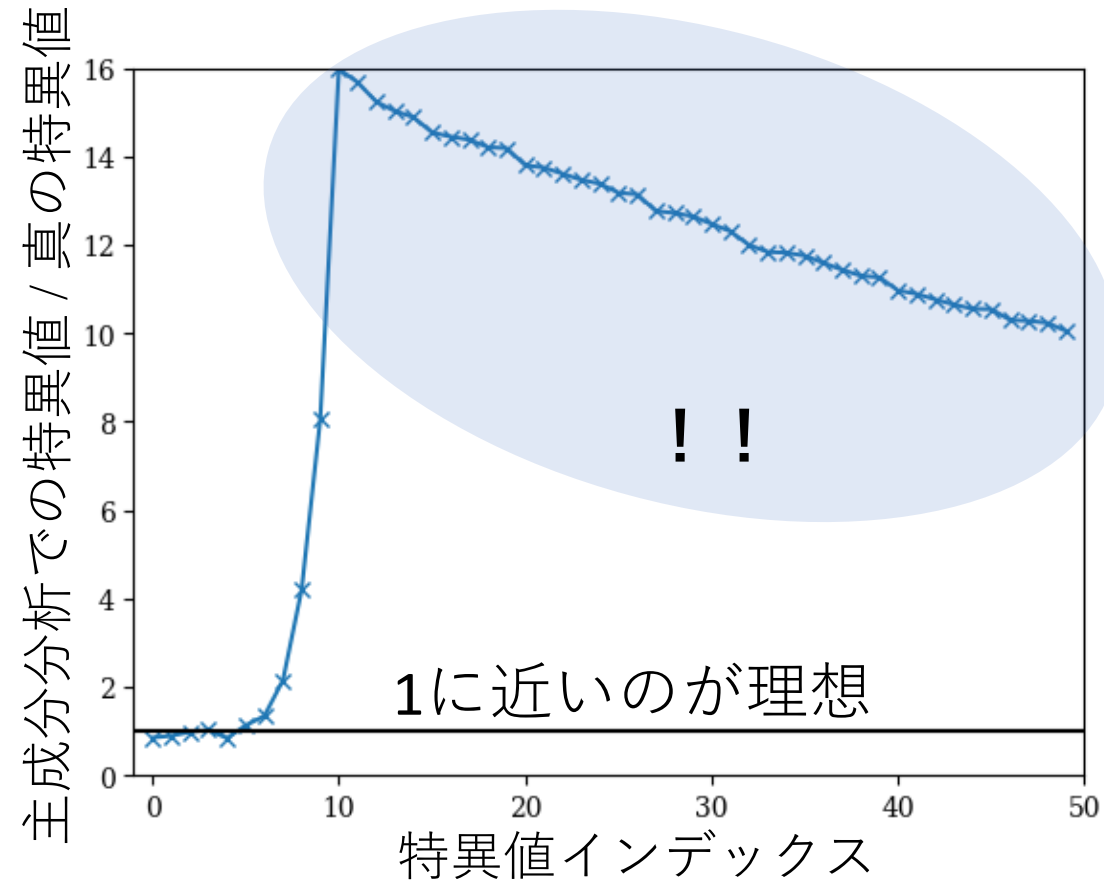


## Simulation 2の結果

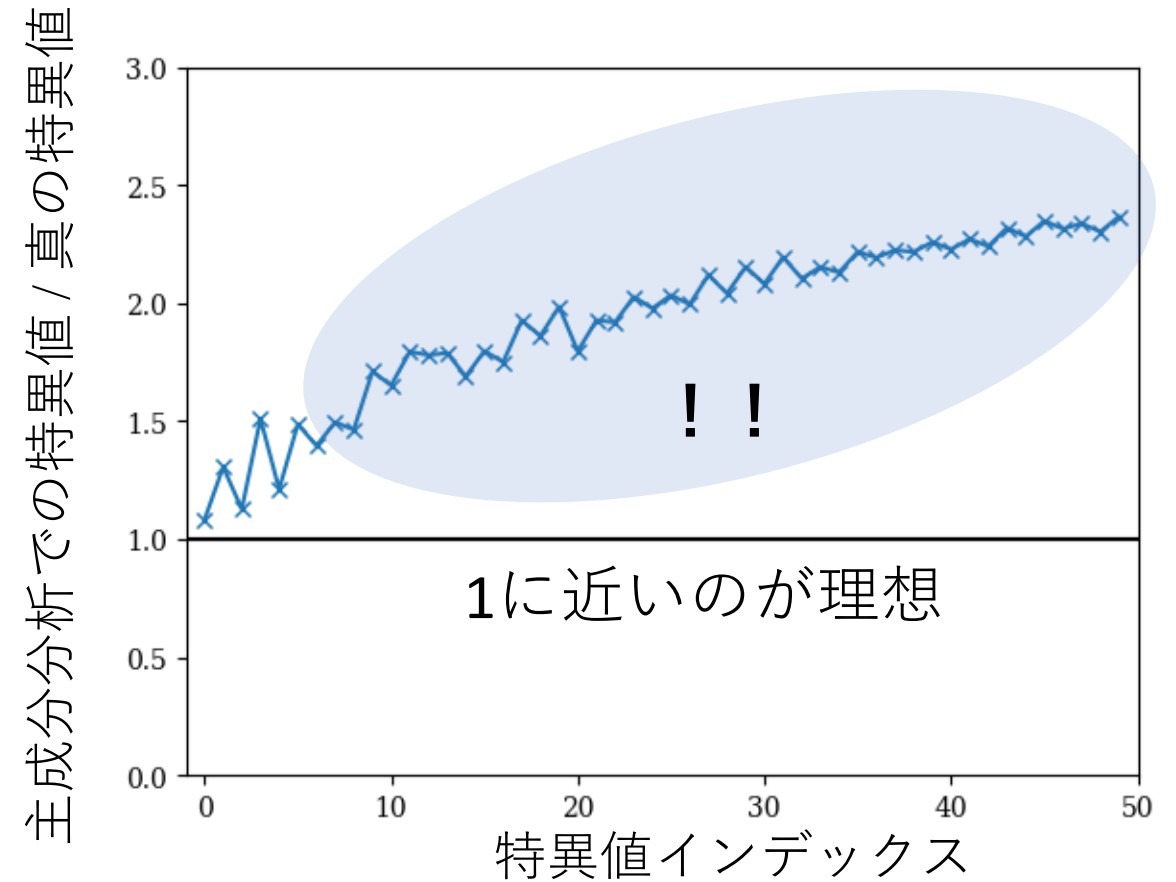


# ナイーブな主成分分析の結果

## Simulation 1の結果



## Simulation 2の結果



ナイーブな主成分分析では明らかに不一致

# 高次元小標本を生かした補正：Yata and Aoshima (2012)

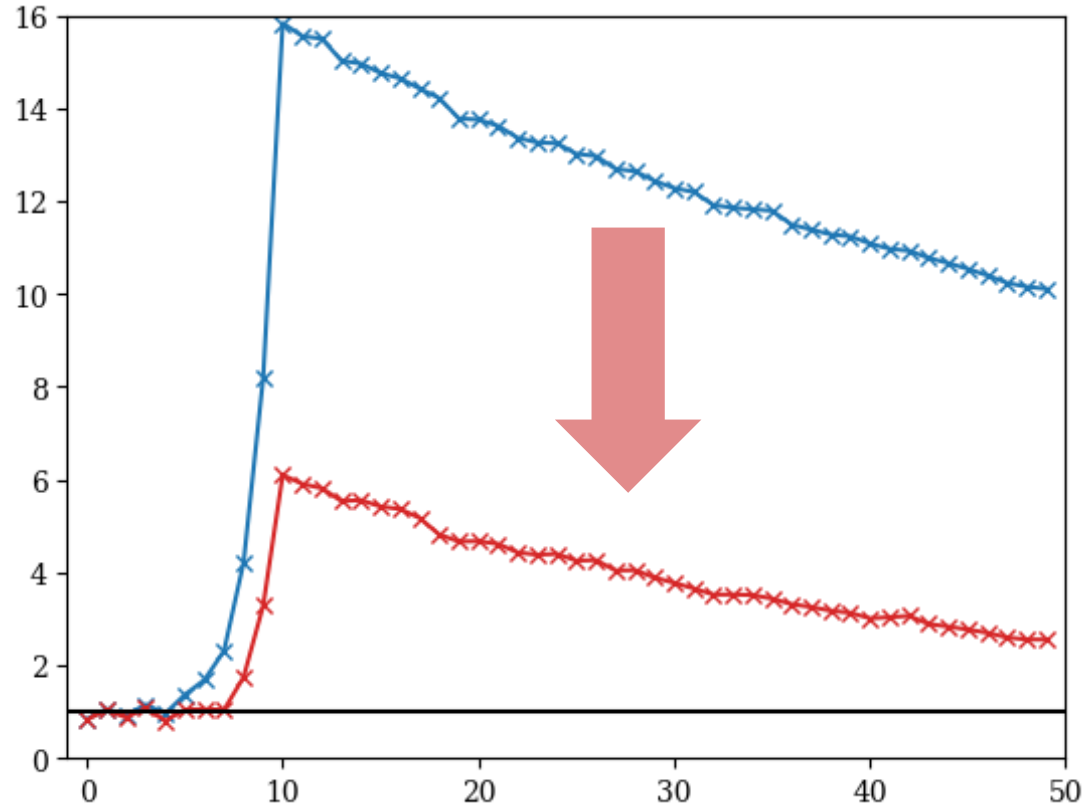
Yata and Aoshima (2012)は高次元小標本性を生かした補正法(ノイズ掃き出し法)を提案

# 高次元小標本を生かした補正：Yata and Aoshima (2012)

Yata and Aoshima (2012)は高次元小標本性を生かした補正法(ノイズ掃き出し法)を提案

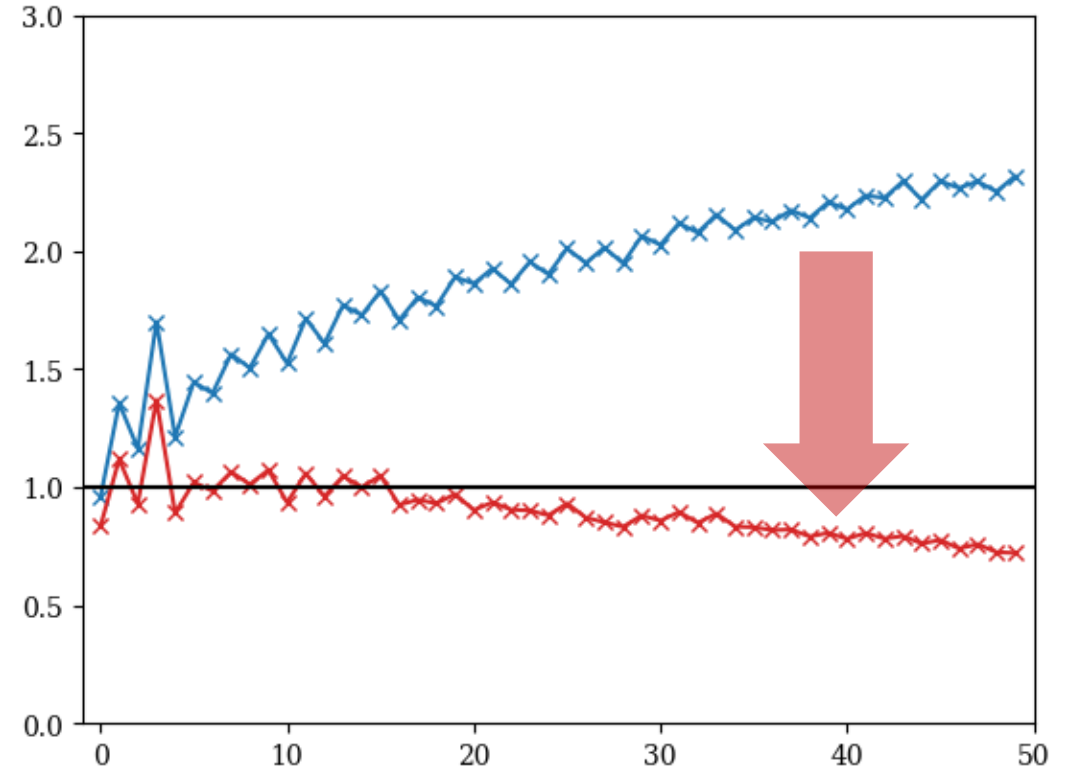
## Simulation 1の結果

主成分分析での特異値 / 真の特異値



## Simulation 2の結果

主成分分析での特異値 / 真の特異値

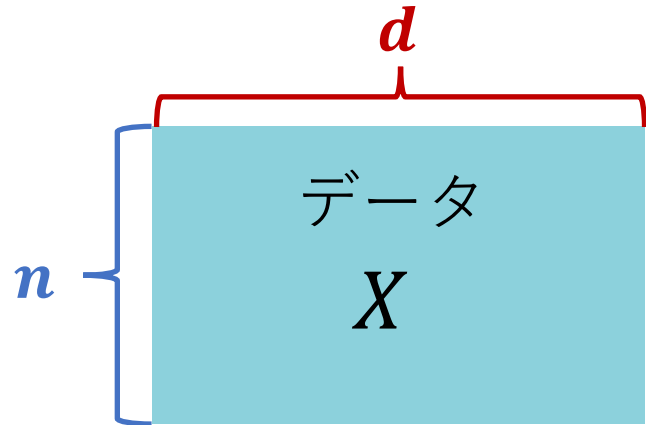


\* 青嶋・矢田(2019)「高次元の統計学」

\* Yata and Aoshima (2012) Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations

# 主成分分析における「双対表現」

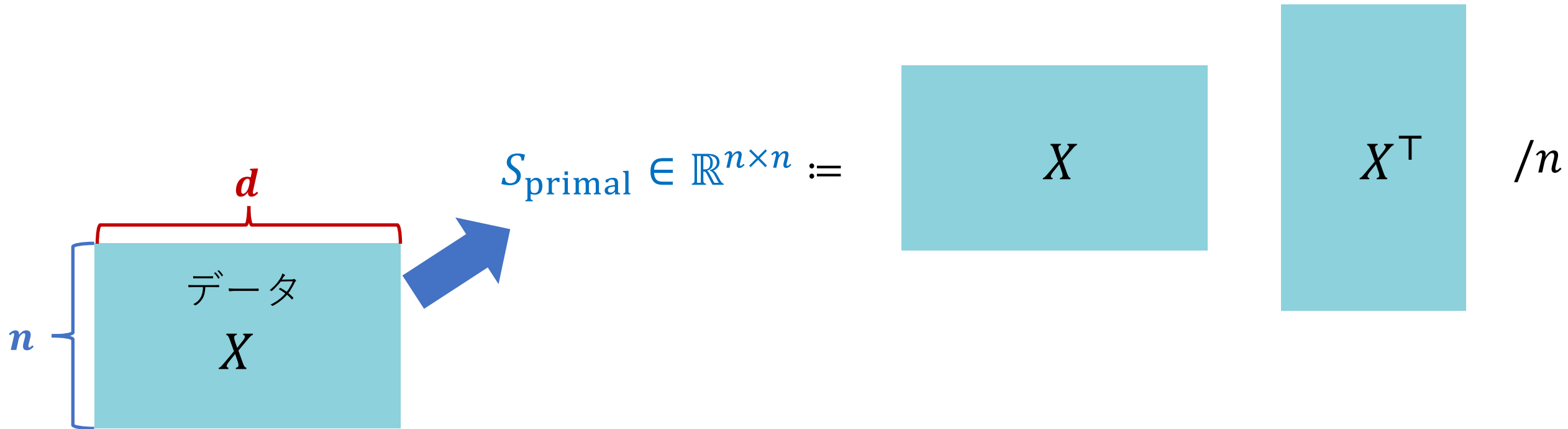
主成分分析ではデータの縮約に二つの表現が可能





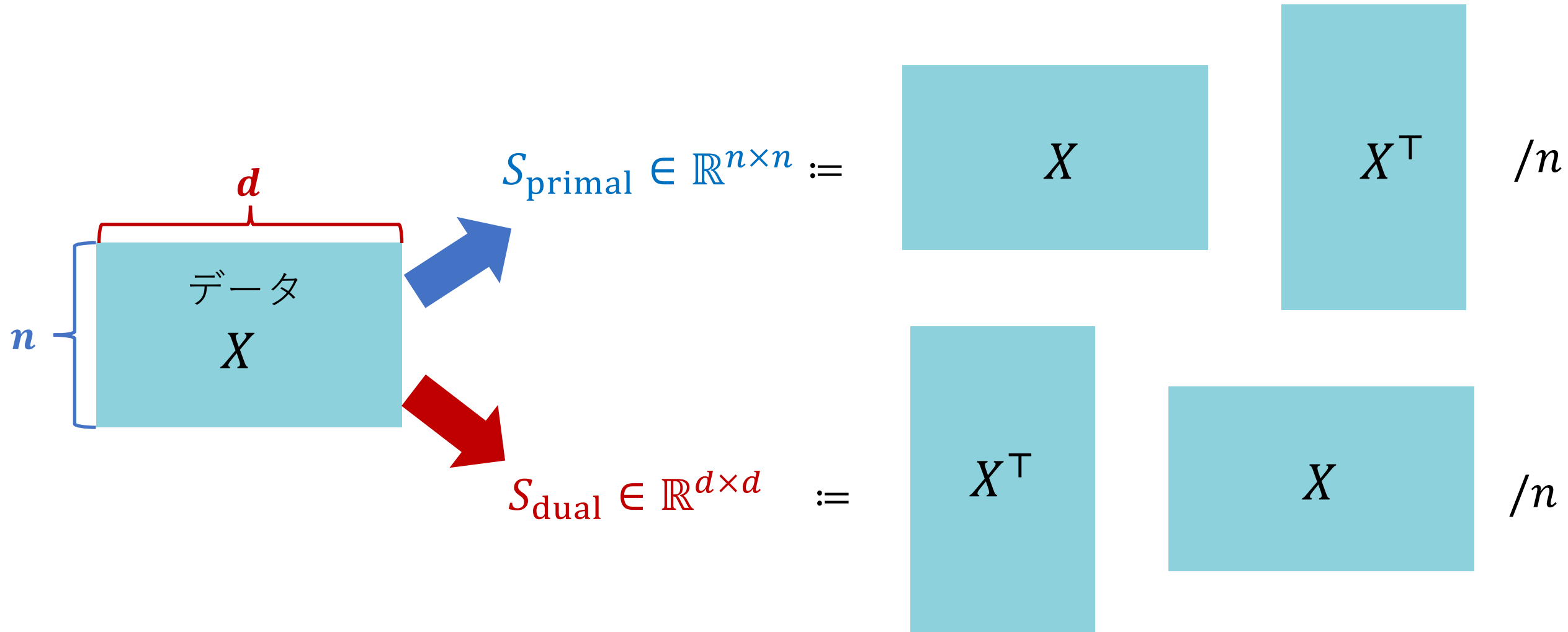
# 主成分分析における「双対表現」

主成分分析ではデータの縮約に二つの表現が可能



# 主成分分析における「双対表現」

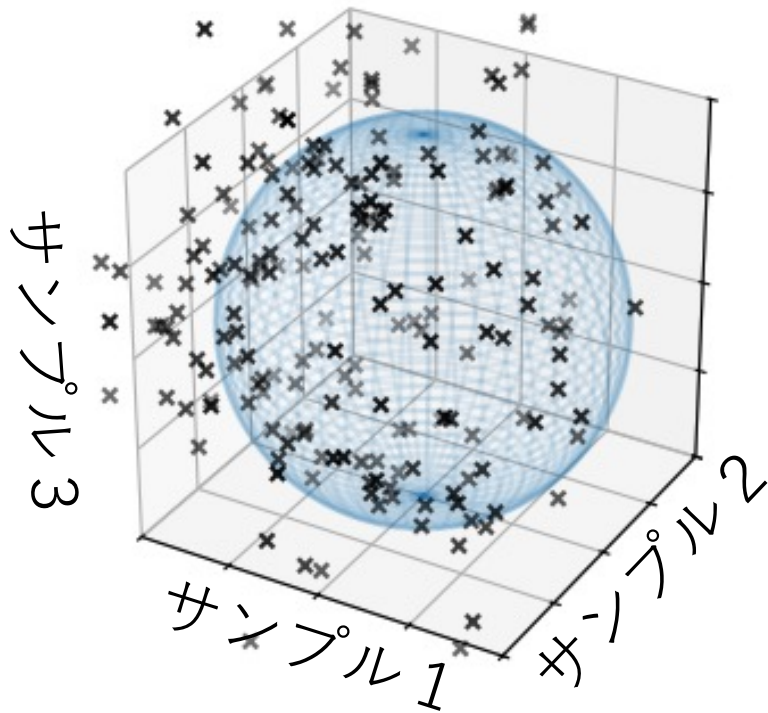
主成分分析ではデータの縮約に二つの表現が可能



ベクトル  $w_j = \left( \frac{n}{\sum_i \lambda_i} \right) S_{\text{dual}} u_j(S_{\text{dual}}) \in \mathbb{R}^n$  は  $d \rightarrow \infty$  で球に集中

ベクトル  $w_j = \left( \frac{n}{\sum_i \lambda_i} \right) S_{\text{dual}} u_j(S_{\text{dual}}) \in \mathbb{R}^n$  は  $d \rightarrow \infty$  で球に集中

$d = 100$

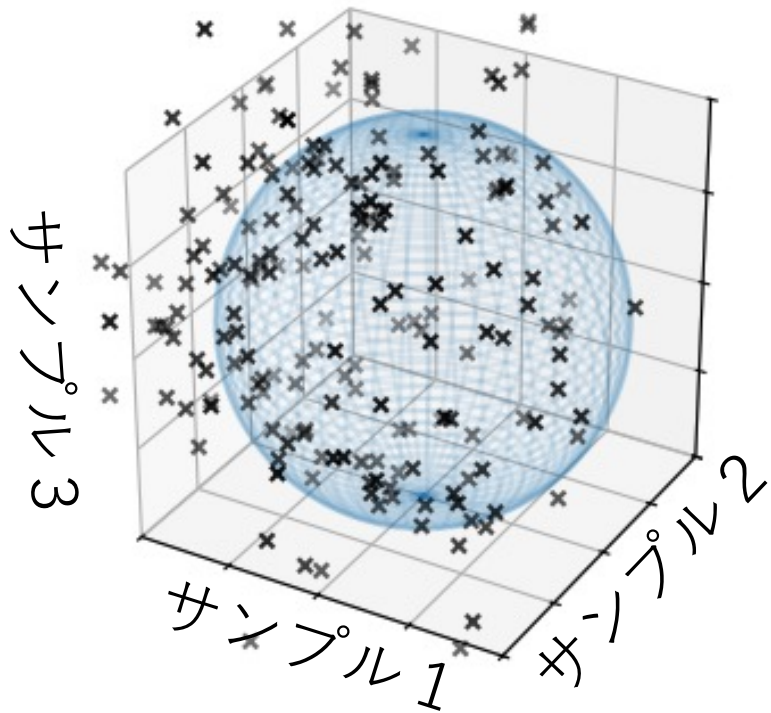


データは簡単のためガウスベクトルとする

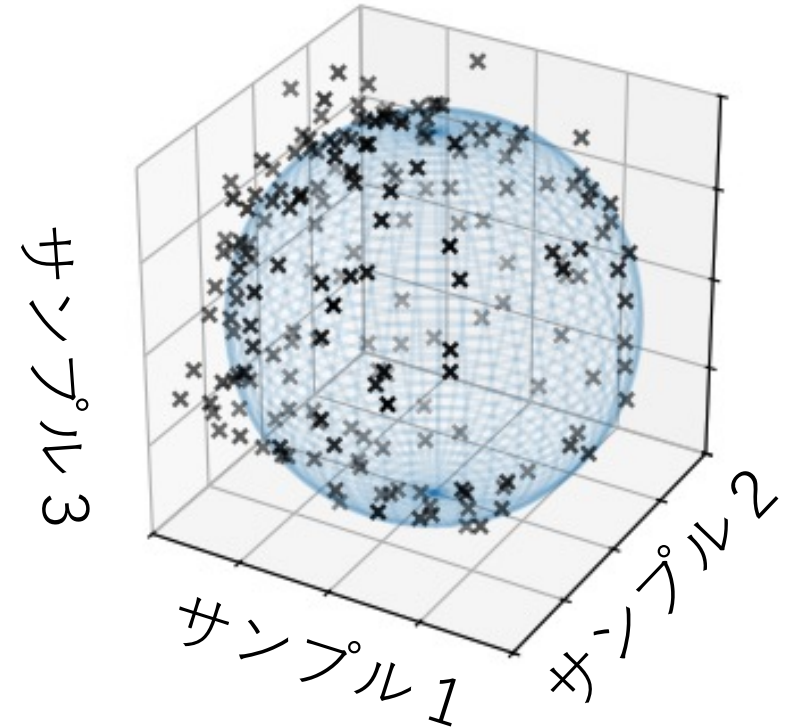
# 「高次元」データの集中現象 (Hall and Marron, 2005; Ahn et al., 2007; Yata and Aoshima, 2012)

ベクトル  $w_j = \left( \frac{n}{\sum_i \lambda_i} \right) S_{\text{dual}} u_j(S_{\text{dual}}) \in \mathbb{R}^n$  は  $d \rightarrow \infty$  で球に集中

$d = 100$



$d = 2000$



データは簡単のためガウスベクトルとする

# 「高次元」データの集中現象を利用した補正

一般に、固有値は重要な固有値, それ以外の多くの固有値に分かれる

$$S_{\text{dual}} \in \mathbb{R}^{d \times d} = \sum_{i=1, \dots, r} \sigma_i^2 u_i u_i^T + \sum_{i=r+1, \dots, d} \sigma_i^2 u_i u_i^T$$

たくさんある！（ノイズ）

# 「高次元」データの集中現象を利用した補正

一般に、固有値は重要な固有値, それ以外の多くの固有値に分かれる

$$S_{\text{dual}} \in \mathbb{R}^{d \times d} = \sum_{i=1, \dots, r} \sigma_i^2 u_i u_i^\top + \sum_{i=r+1, \dots, d} \sigma_i^2 u_i u_i^\top$$

たくさんある! (ノイズ)

ノイズ部分の集中現象を利用することで固有値・固有ベクトルを推定可能

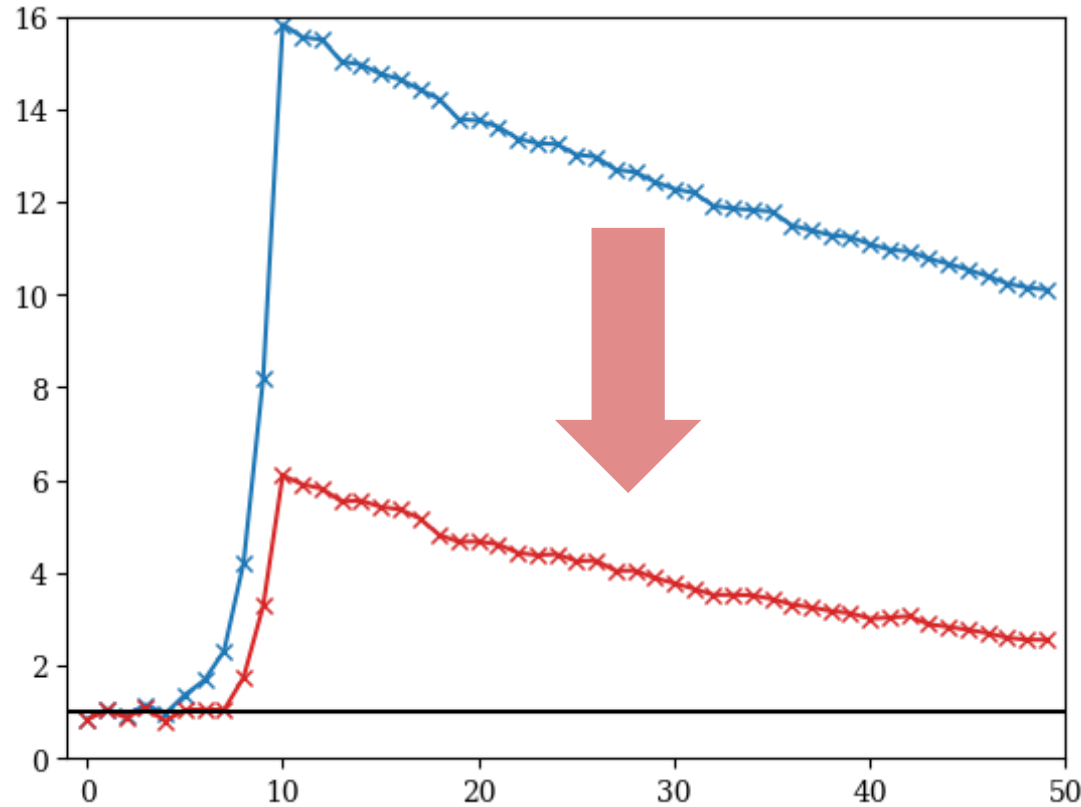
$$\check{\lambda}_i = \sigma_i^2 - \frac{\text{tr}(S_{\text{dual}}) - \sum_{j=1, \dots, i} \sigma_j^2}{n - 1 - i}$$

$$h_i = \frac{1}{\sqrt{(n-1)\check{\lambda}_i}} X u_i(S_{\text{dual}})$$

# 高次元小標本を生かした補正：Yata and Aoshima (2012)

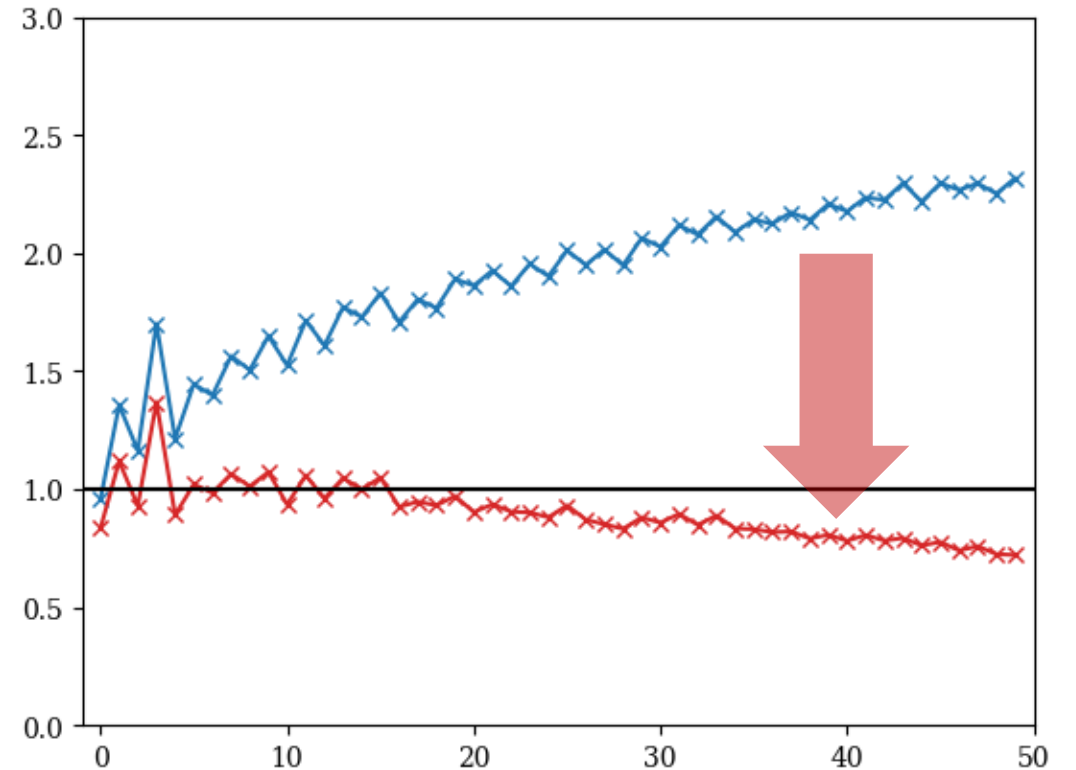
主成分分析での特異値 / 真の特異値

## Simulation 1の結果



主成分分析での特異値 / 真の特異値

## Simulation 2の結果



Yata and Aoshima (2012)は高次元小標本性を生かした補正法(ノイズ掃き出し法)を提案



# 天文学への応用：Takeuchi et al. (2024)

ALMA望遠鏡で撮影された銀河NGC 253の分光マップへの適用  
→銀河の分子ガス噴き出しをデータ駆動的に抽出

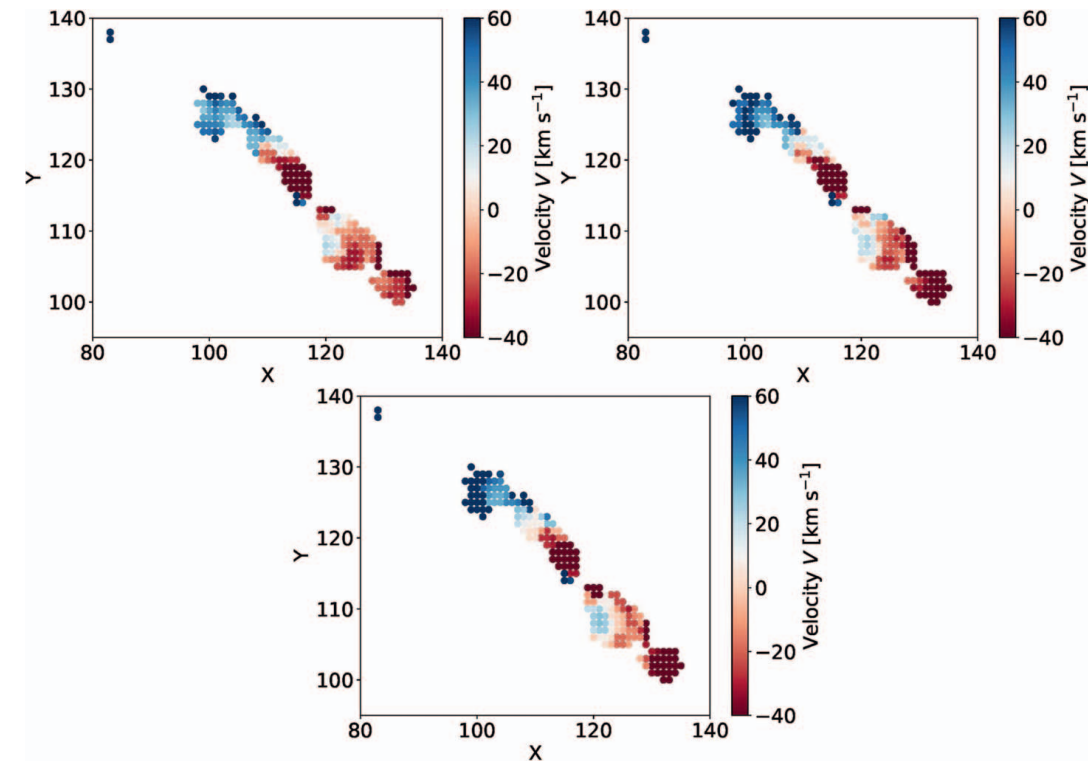


Figure 4. Velocity map of NGC 253 estimated from HCN(4-3) (top), HNC(4-3) (middle), and CS(7-6) (bottom). The systemic velocity of  $243 \text{ km s}^{-1}$  is subtracted.

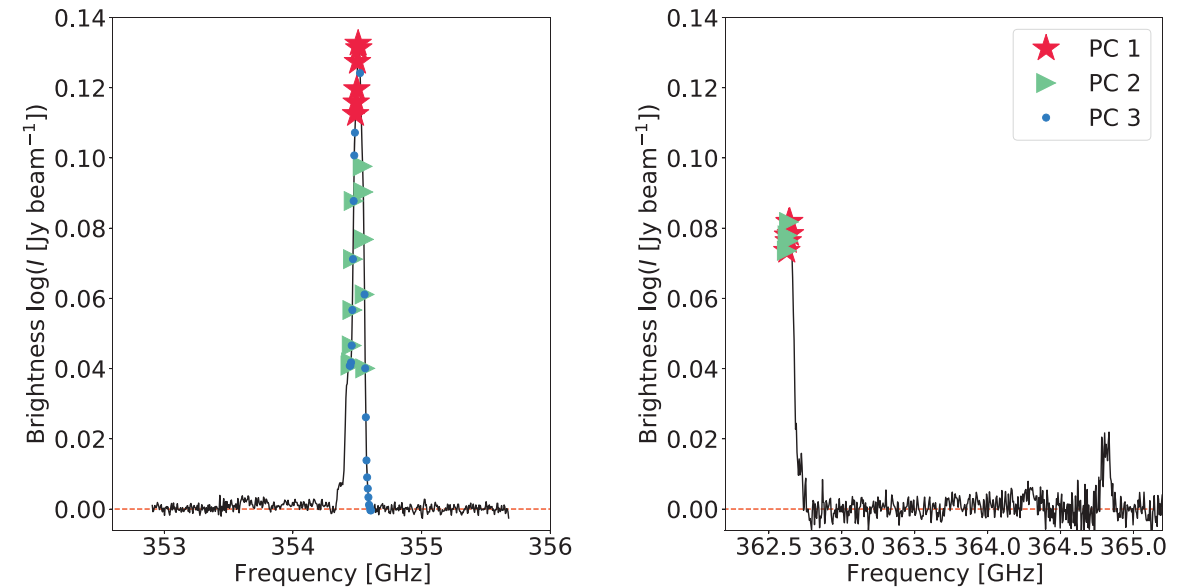


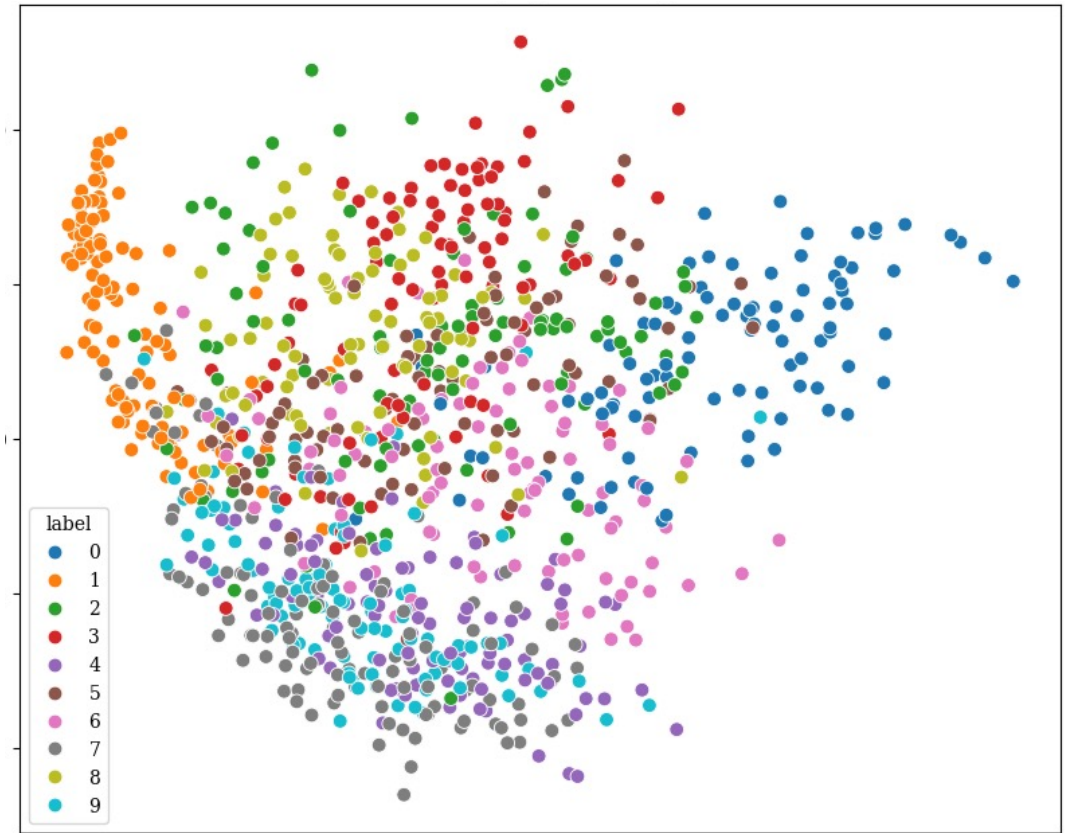
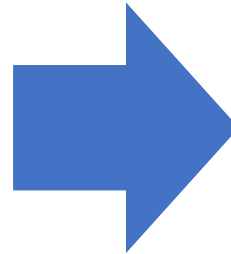
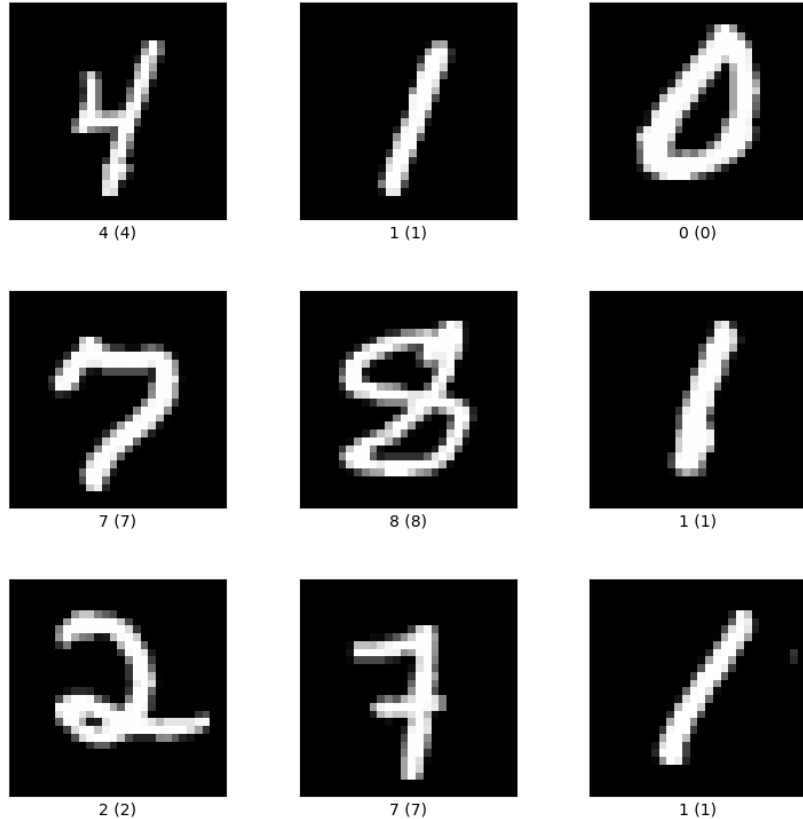
Figure 19. Responsible features to characterize PCs from the A-SPCA for the ALMA map of NGC 253, after the Doppler shift correction due to the systemic rotation. Stars and triangles represent PC1- and PC2-related responsible spectral features similar to those in Figure 10, while filled circles are PC3 related.

From Takeuchi et al. (2024) <https://doi.org/10.3847/1538-4365/ad2517>

# 線形次元圧縮から非線形の次元圧縮へ

主成分分析は強力な手法だがあくまでも線形圧縮

手書き数字認識 (MNISTデータセット)



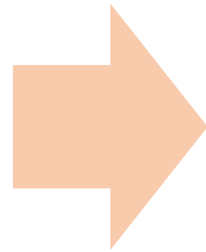
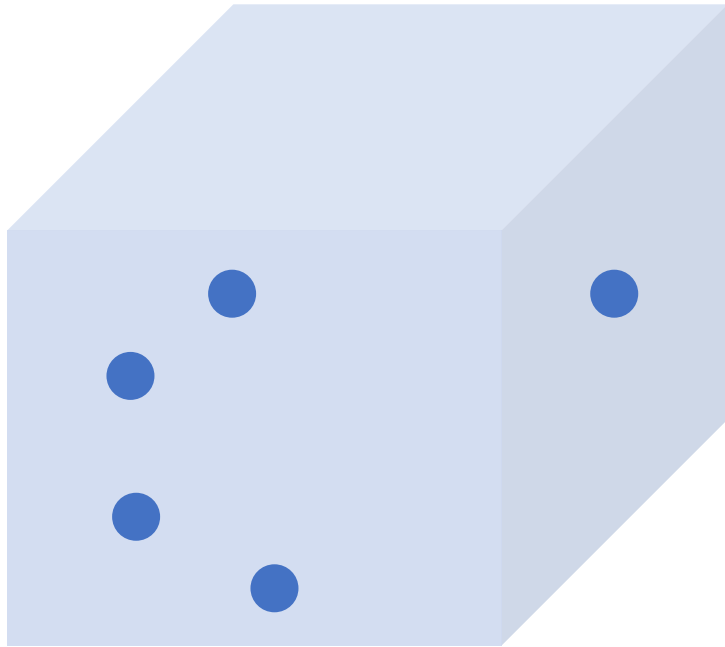
主成分だけではよくわからないことも

From <https://www.tensorflow.org/datasets/catalog/mnist?hl=ja>

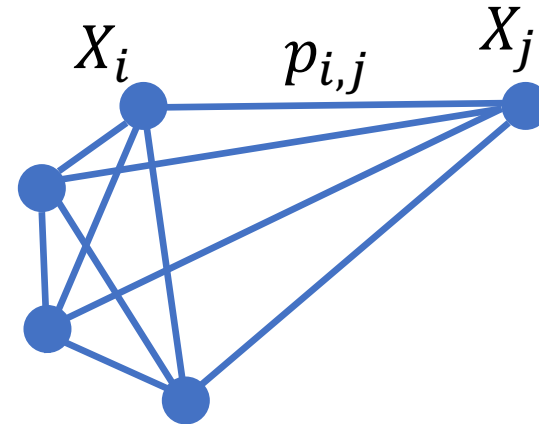
# 高次元データの「グラフ表現」

非線形圧縮で活躍するのは高次元データの「グラフ表現」

データ空間  $\mathbb{R}^d$



データのグラフ表現



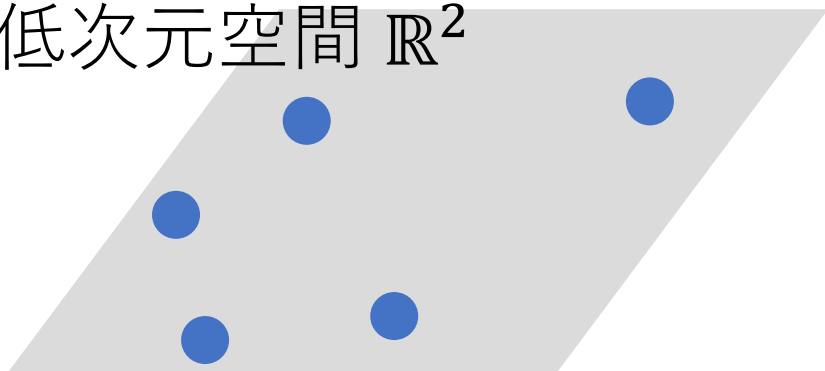
$$p_{j \rightarrow i} = \frac{\exp(-\|X_i - X_j\|^2 / (2\sigma^2))}{\sum_{k \neq j} \exp(-\|X_i - X_k\|^2 / (2\sigma^2))}$$

$$p_{i,j} = \frac{p_{j \rightarrow i} + p_{i \rightarrow j}}{2N}$$

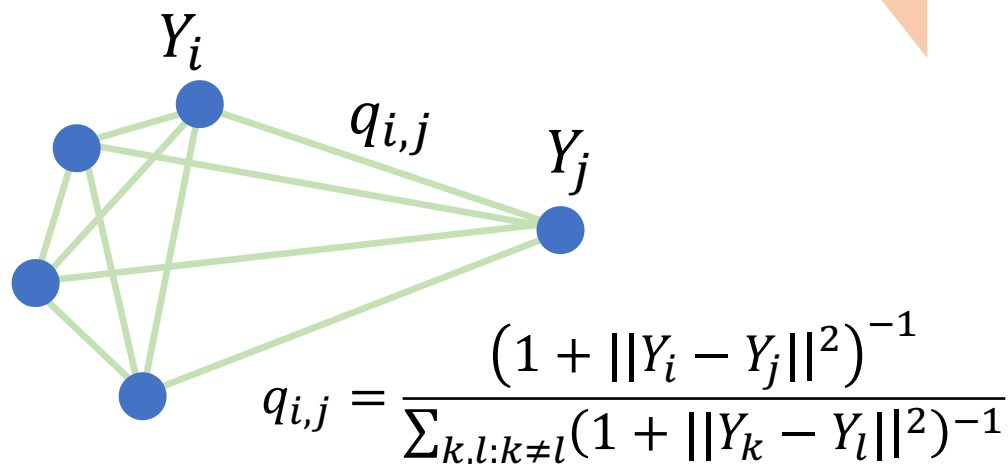
# $t$ -distributed Stochastic Neighbor Embedding ( $t$ -SNE)

データの隣接関係をできる限り保ったまま低次元空間に埋め込む手法

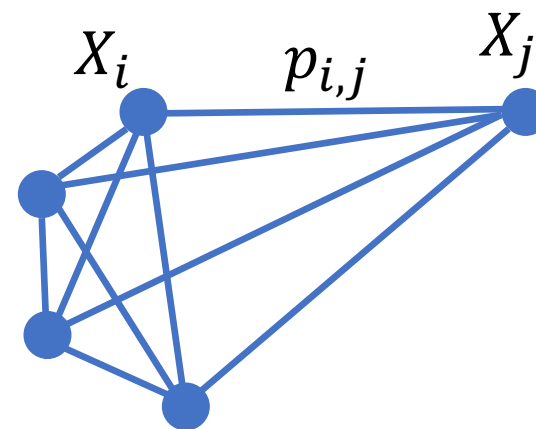
低次元空間  $\mathbb{R}^2$



低次元でのグラフ表現



データのグラフ表現

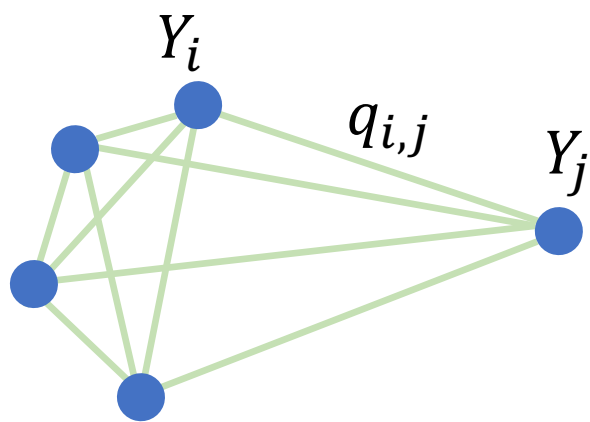


$$p_{j \rightarrow i} = \frac{\exp(-\|X_i - X_j\|^2 / (2\sigma^2))}{\sum_{k \neq j} \exp(-\|X_i - X_k\|^2 / (2\sigma^2))}$$
$$p_{i,j} = \frac{p_{j \rightarrow i} + p_{i \rightarrow j}}{2N}$$

# $t$ -distributed Stochastic Neighbor Embedding ( $t$ -SNE)

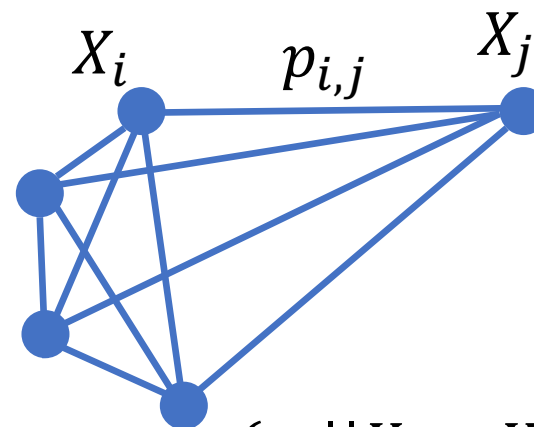
データの隣接関係をできる限り保ったまま低次元空間に埋め込む手法

低次元でのグラフ表現



$$q_{i,j} = \frac{(1 + \|Y_i - Y_j\|^2)^{-1}}{\sum_{k,l:k \neq l} (1 + \|Y_k - Y_l\|^2)^{-1}}$$

データのグラフ表現



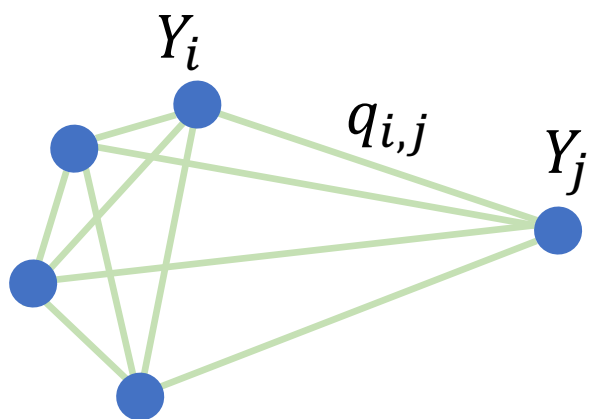
$$p_{j \rightarrow i} = \frac{\exp(-\|X_i - X_j\|^2 / (2\sigma^2))}{\sum_{k \neq j} \exp(-\|X_i - X_k\|^2 / (2\sigma^2))}$$
$$p_{i,j} = \frac{p_{j \rightarrow i} + p_{i \rightarrow j}}{2N}$$

この二つのグラフが近くなるような  $Y_1, \dots, Y_N \in \mathbb{R}^2$  (埋め込み)を見つける

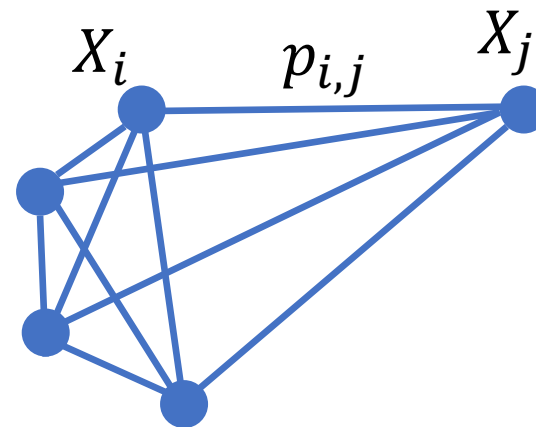
# \* $t$ -distributed Stochastic Neighbor Embedding ( $t$ -SNE)

データの隣接関係をできる限り保ったまま低次元空間に埋め込む手法

低次元でのグラフ表現



データのグラフ表現



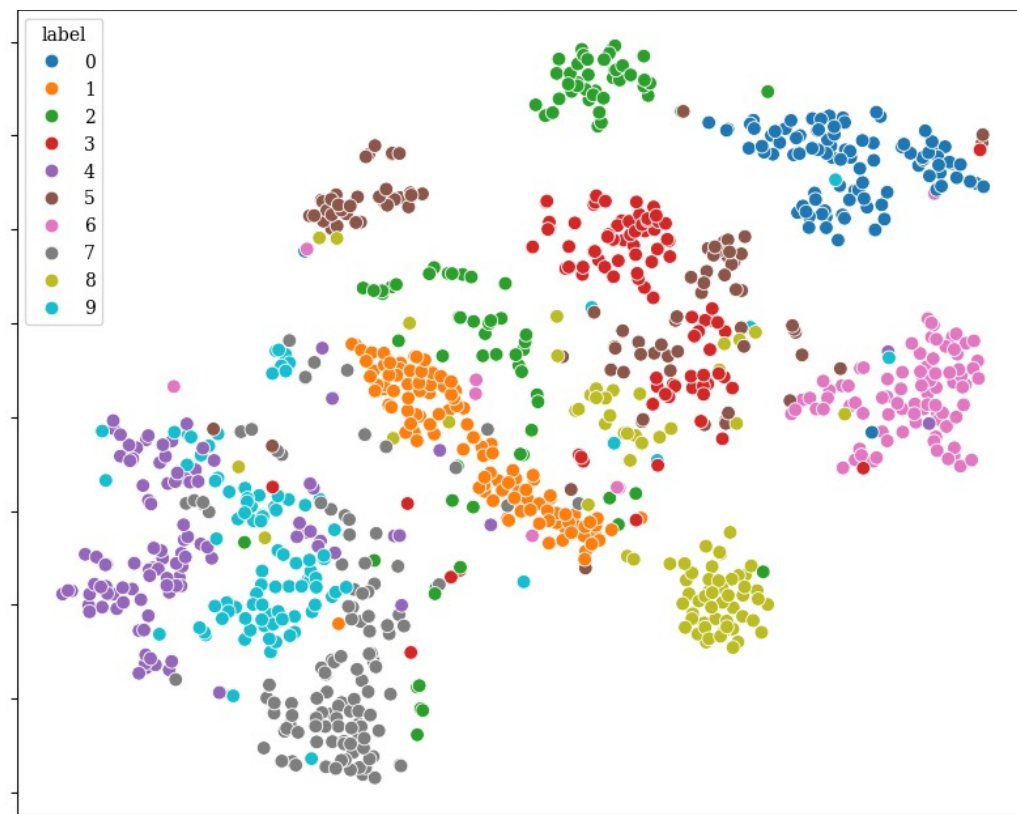
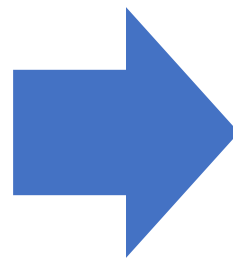
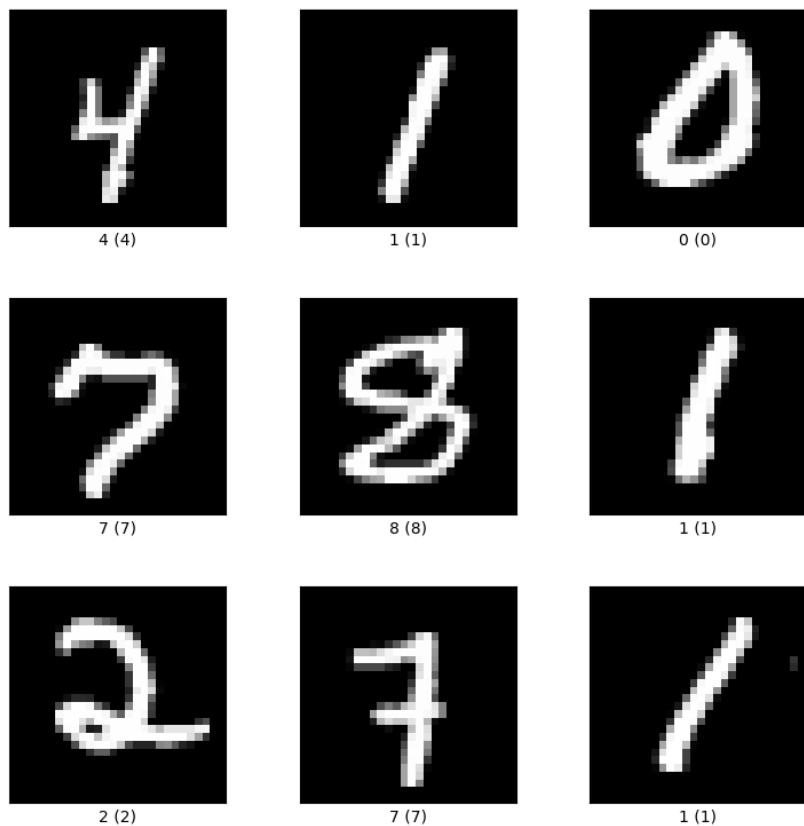
この二つのグラフが近くなるような  $Y_1, \dots, Y_N \in \mathbb{R}^2$  は

$$\min_{y_1, \dots, y_N} D(p, q) = \sum_{i,j} p_{i,j} \log \frac{p_{i,j}}{q_{i,j}}$$

を確率的勾配法によって求める

# $t$ -SNEの適用例

先ほどPCAではうまくいかなかった手書き数字認識の例で使ってみると



From <https://www.tensorflow.org/datasets/catalog/mnist?hl=ja>

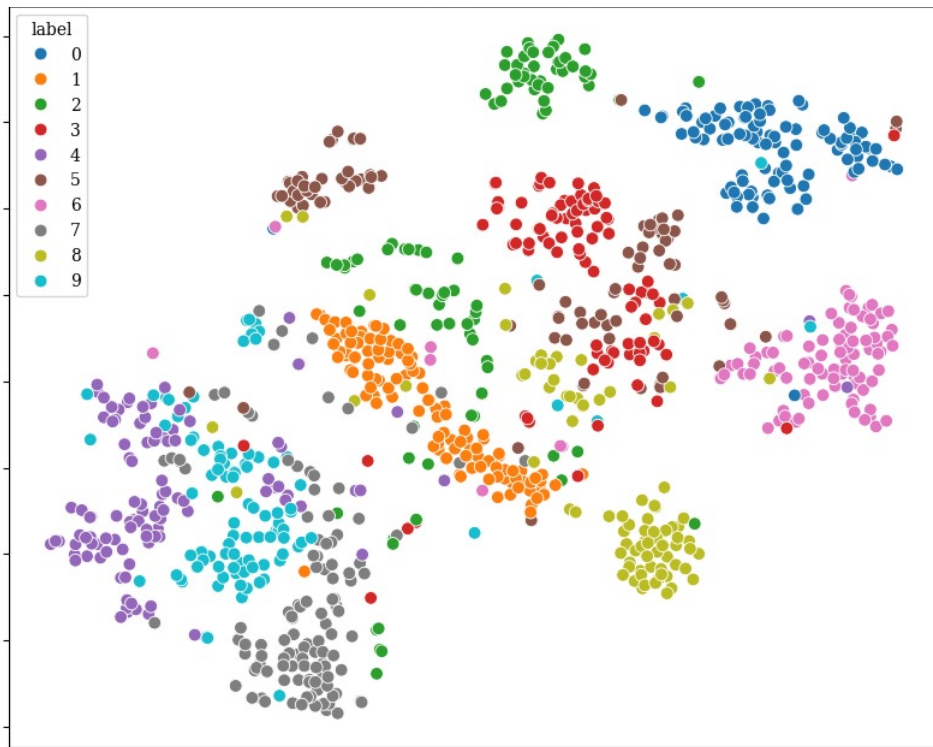
綺麗にラベルと対応した埋め込み

# Uniform Manifold Approximation and Projection (UMAP)

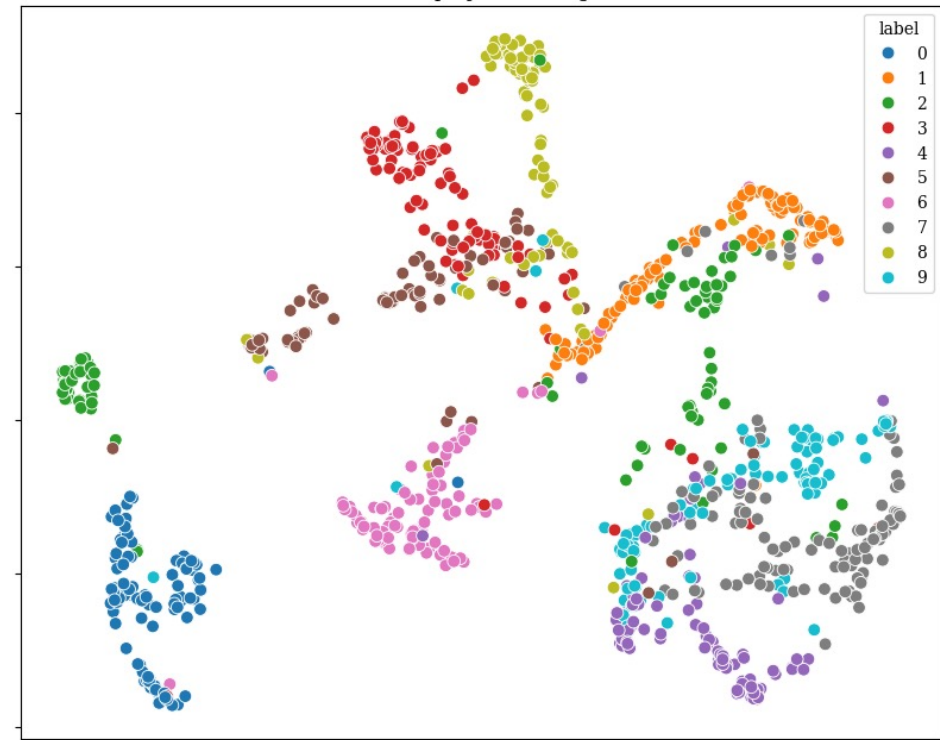
$t$ -SNEは「遠いものは遠く」なるわけではない

→この結果を解消し、解釈しやすい次元圧縮を行うのがUMAP

$t$ -SNE



UMAP



\* McInnes et al. (2018) UMAP: Uniform Manifold Approximation and Projection

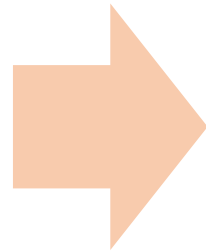
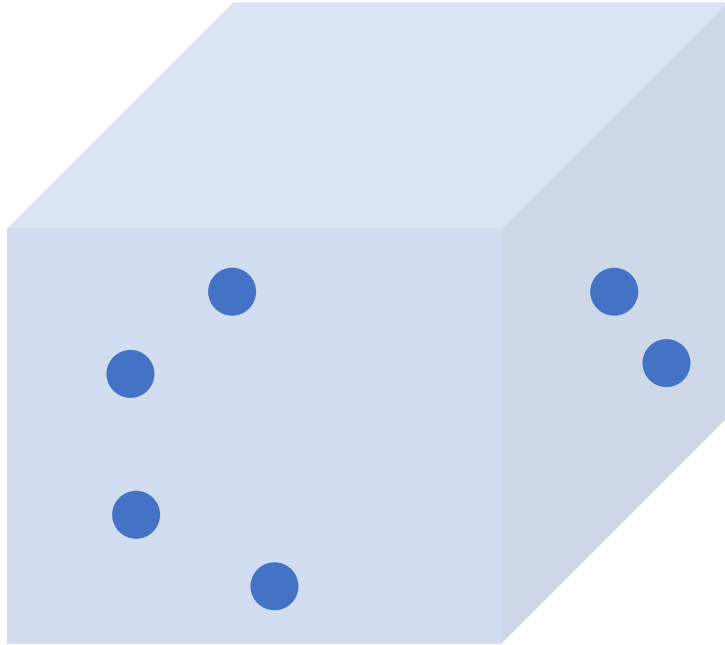
\* 酒井・寺田・高橋 (2024) 重力波観測における突発性雑音の教師なし分類



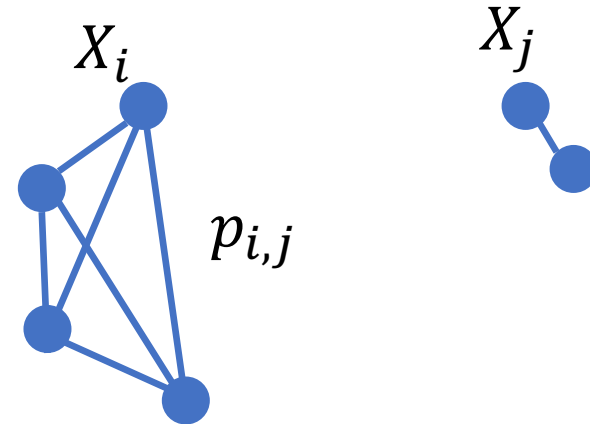
# Uniform Manifold Approximation and Projection (UMAP)

データの局所構造・大域構造を保ったまま低次元空間に埋め込む手法

データ空間  $\mathbb{R}^d$



データのグラフ表現



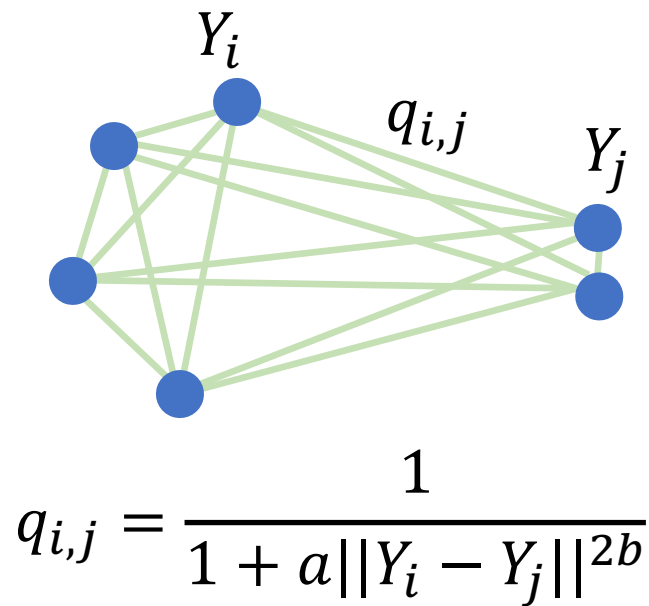
遠くに枝を張らないように  $k$ 近傍法を利用

$$p_{i,j} = \exp\left(-\frac{\|X_i - X_j\| - \rho_i}{\sigma_i}\right)$$

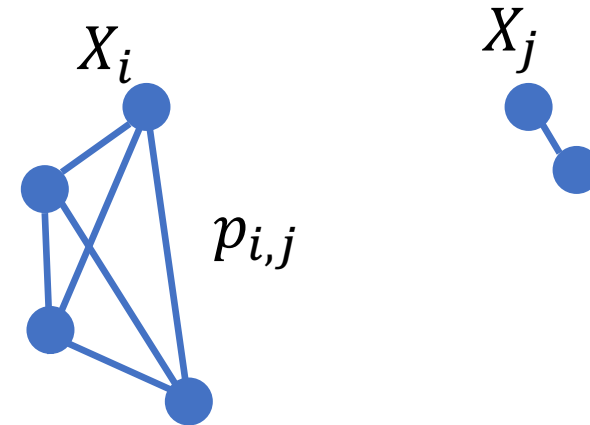
# Uniform Manifold Approximation and Projection (UMAP)

データの局所構造・大域構造を保ったまま低次元空間に埋め込む手法

低次元でのグラフ表現



データのグラフ表現



遠くに枝を張らないように  $k$  近傍法を利用

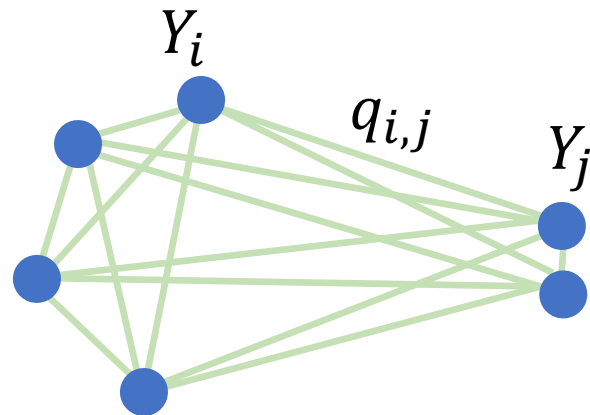
$$p_{i,j} = \exp\left(-\frac{\|X_i - X_j\| - \rho_i}{\sigma_i}\right)$$

この二つのグラフが近くなるような  $Y_1, \dots, Y_N \in \mathbb{R}^2$  (埋め込み) を見つける

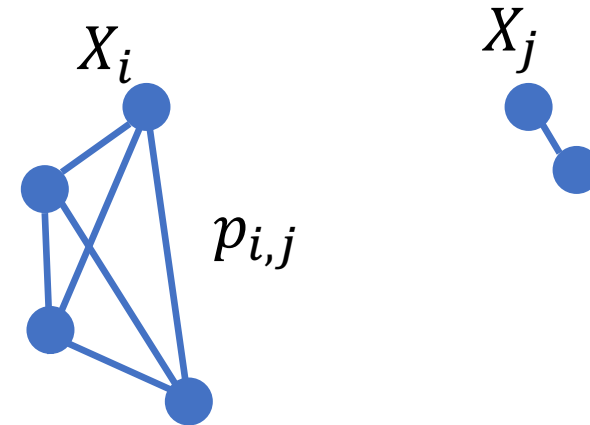
# \* Uniform Manifold Approximation and Projection (UMAP)

データの局所構造・大域構造を保ったまま低次元空間に埋め込む手法

低次元でのグラフ表現



データのグラフ表現



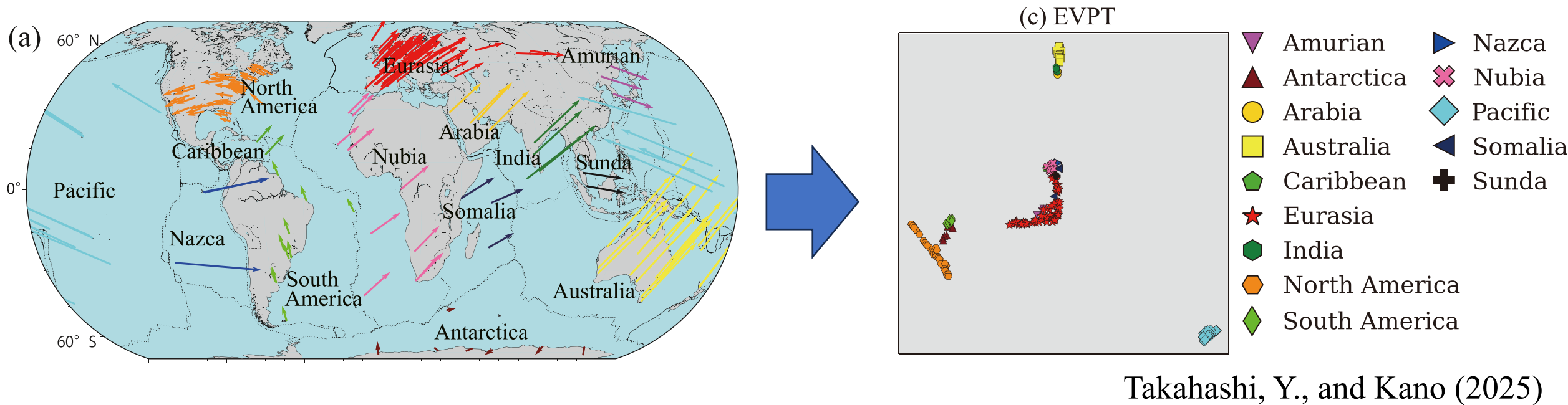
この二つのグラフが近くなるような  $Y_1, \dots, Y_N \in \mathbb{R}^2$  を

$$\min_{Y_1, \dots, Y_N} \sum_{i,j} p_{i,j} \log \frac{p_{i,j}}{q_{i,j}} + (1 - p_{i,j}) \log \frac{1 - p_{i,j}}{1 - q_{i,j}}$$

を確率的勾配法によって求める

# データ間の距離について

データ間の距離はユークリッド距離以外が適切な場合もある  
→ 変数に関する事前知識がある場合 (球面上のデータなど)

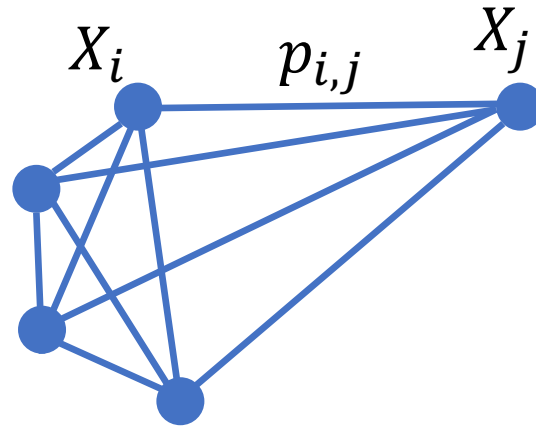
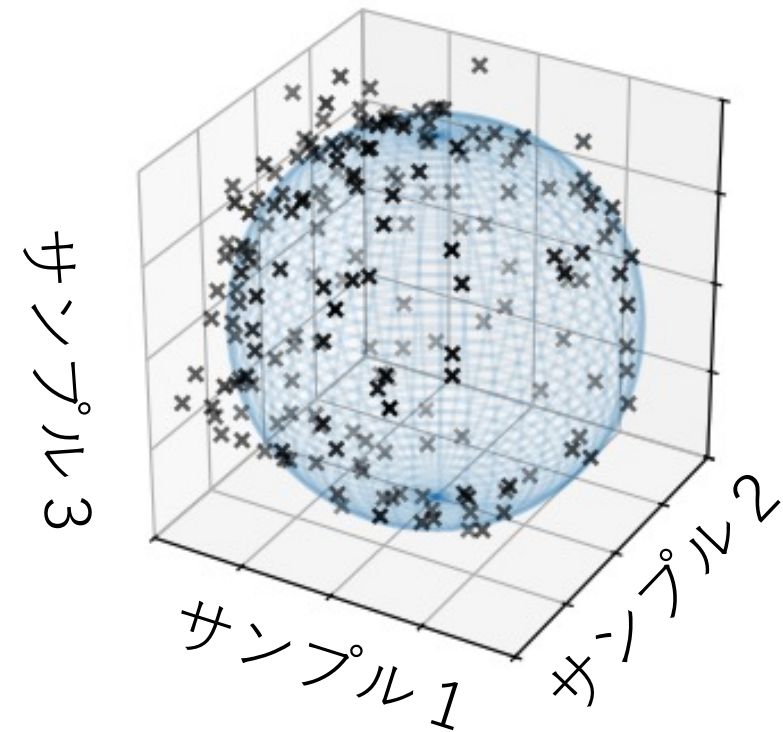


高次元データとはいえ、データに関する事前知識は重要

# 高次元のデータ解析のアイデア

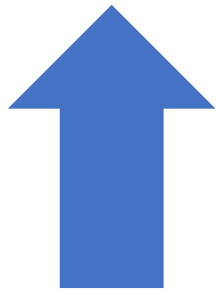
球面・軸集中現象

グラフ構造



# 本日の内容

可視化



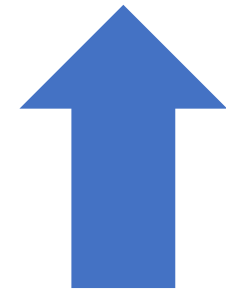
次元圧縮

発見・予測



正則化

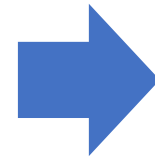
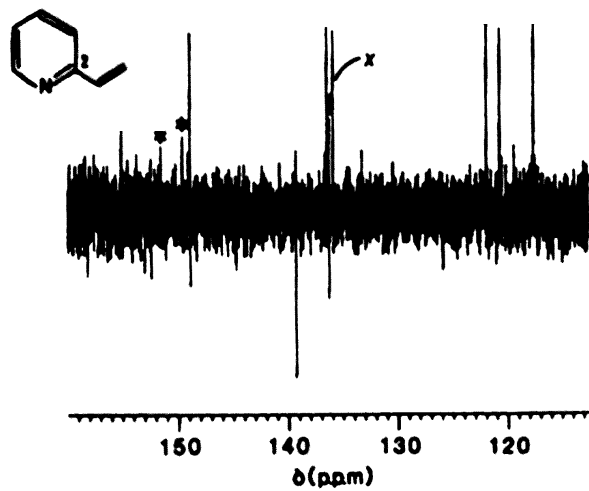
不確実性評価



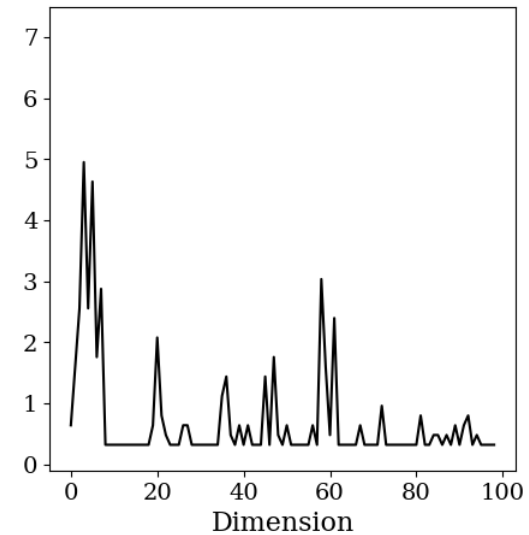
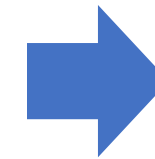
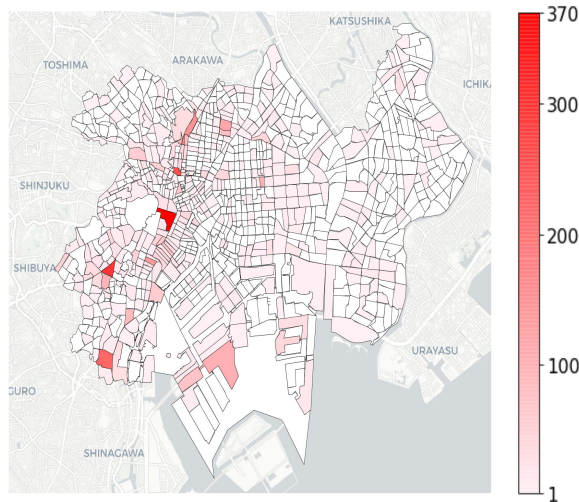
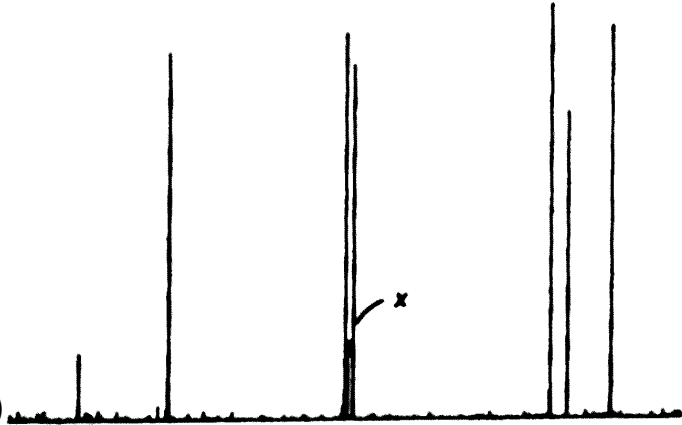
ベイズ

# スパース推定

高次元データ～「少数の有用データ」 + 「それ以外のノイズ」という仮説



From Donoho et al. (1992)



\* 川野・松井・廣瀬 (2018) 「スパース推定法による統計モデリング」

# Least Absolute Shrinkage of Selection Operators (LASSO)

回帰モデルの回帰係数のスパース推定 (Tibshirani, 1996; Chen et al., 1998)

- 1次元の予測値と  $d$ 次元の共変量の組  $\{(y_i, x_i) : i, \dots, n\}$

$$\hat{\theta}_\lambda := \operatorname{argmin}_{\theta} \left\{ \frac{1}{2n} \sum_{i=1, \dots, n} (y_i - x_i^\top \theta)^2 + \frac{\lambda}{n} \|\theta\|_1 \right\} = \operatorname{argmin}_{\theta} \left\{ \frac{1}{2n} \|Y - X\theta\|^2 + \frac{\lambda}{n} \|\theta\|_1 \right\}$$

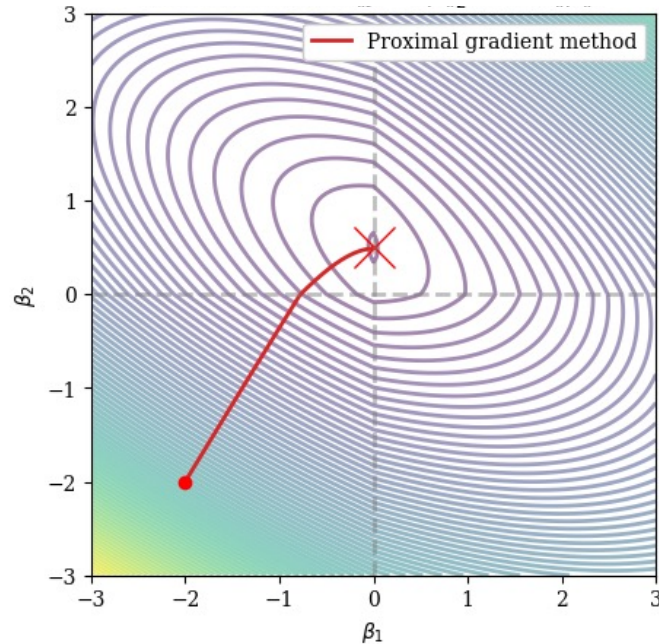


# Least Absolute Shrinkage of Selection Operators (LASSO)

回帰モデルの回帰係数のスパース推定 (Tibshirani, 1996; Chen et al., 1998)

- 1次元の予測値と  $d$ 次元の共変量の組  $\{(y_i, x_i) : i, \dots, n\}$

$$\hat{\theta}_\lambda := \operatorname{argmin}_{\theta} \left\{ \frac{1}{2n} \sum_{i=1, \dots, n} (y_i - x_i^\top \theta)^2 + \frac{\lambda}{n} \|\theta\|_1 \right\} = \operatorname{argmin}_{\theta} \left\{ \frac{1}{2n} \|Y - X\theta\|^2 + \frac{\lambda}{n} \|\theta\|_1 \right\}$$



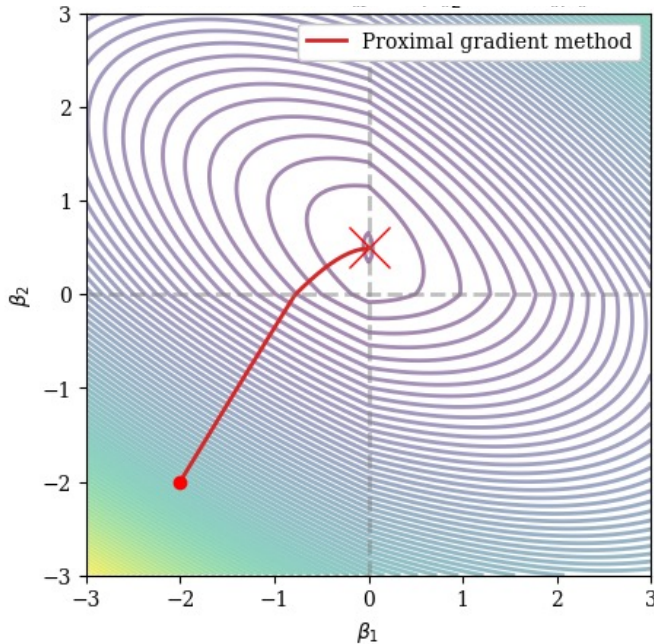
アルゴリズムの工夫により厳密に0に

# Least Absolute Shrinkage of Selection Operators (LASSO)

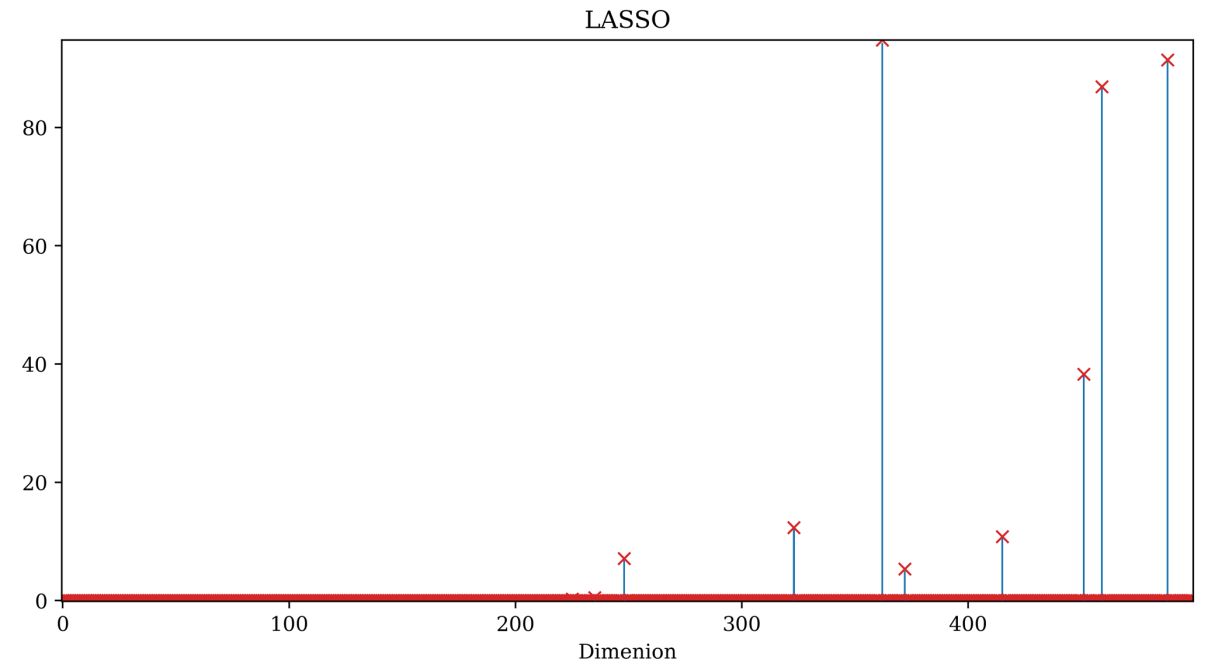
回帰モデルの回帰係数のスパース推定 (Tibshirani, 1996; Chen et al., 1998)

- 1次元の予測値と  $d$ 次元の共変量の組  $\{(y_i, x_i) : i, \dots, n\}$

$$\hat{\theta}_\lambda := \operatorname{argmin}_{\theta} \left\{ \frac{1}{2n} \sum_{i=1, \dots, n} (y_i - x_i^\top \theta)^2 + \frac{\lambda}{n} \|\theta\|_1 \right\} = \operatorname{argmin}_{\theta} \left\{ \frac{1}{2n} \|Y - X\theta\|^2 + \frac{\lambda}{n} \|\theta\|_1 \right\}$$



アルゴリズムの工夫により厳密に0に



少数な意味のある係数を高速に推定

# LASSOの仲間：Generalized LASSO

$$\operatorname{argmin}_{\theta} \left\{ \frac{1}{2n} \|Y - X\theta\|_2^2 + \frac{\lambda}{n} \|D\theta\|_1 \right\}$$

- Fused LASSO：隣接制約

$$D = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & -1 & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}$$

- $l_1$  trend filtering：区分解微分可能関数

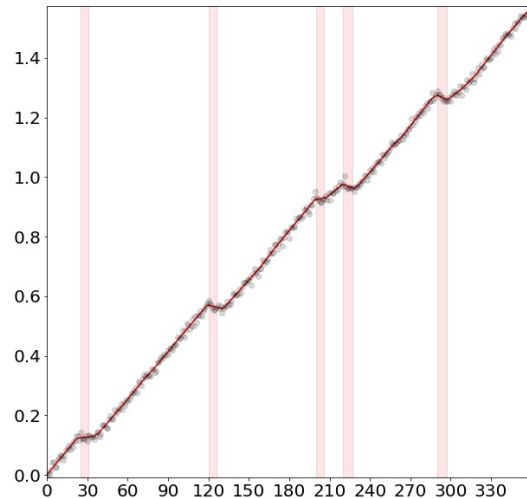
$$D = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 2 & -1 & 0 \\ 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix}$$

# LASSOの仲間：Generalized LASSO

$$\operatorname{argmin}_{\theta} \left\{ \frac{1}{2n} \|Y - X\theta\|_2^2 + \frac{\lambda}{n} \|D\theta\|_1 \right\}$$

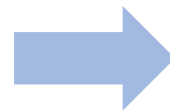
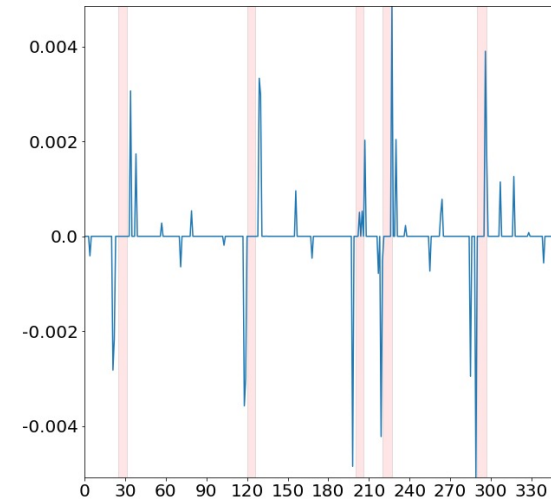
- Fused LASSO：隣接制約

$$D = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & -1 & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}$$



- $l_1$  trend filtering：区分微分可能関数

$$D = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 2 & -1 & 0 \\ 0 & 0 & \cdots & -1 & 2 & -1 \\ 0 & 0 & \cdots & 0 & -1 & 1 \end{pmatrix}$$



$D\theta$

\* LASSOの仲間：Square-root LASSO (Belloni, Chernozhukov, Wang, 2011)

- 実際のLASSOの性能は観測分散 $\sigma$ に依存する ( $\lambda$ の取り方等)
- LASSOの定式化を少し変更することでこの依存がなくせる

$$\operatorname{argmin}_{\theta} \left\{ \frac{1}{2n} \|Y - X\theta\|^2 + \frac{\lambda}{n} \|\theta\|_1 \right\} \quad \Rightarrow \quad \operatorname{argmin}_{\theta} \left\{ \frac{1}{\sqrt{2n}} \|Y - X\theta\| + \frac{\lambda}{n} \|\theta\|_1 \right\}$$

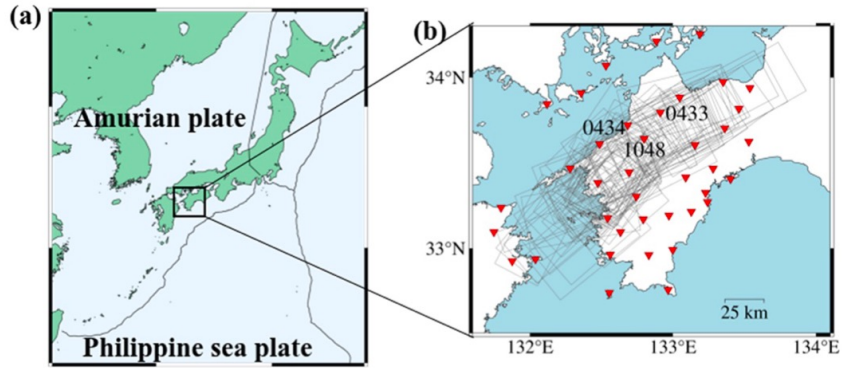
Square-root LASSO!

- Distributionally Robust Optimizationとの関連 (Blanchet, Kang, Murthy, 2019)

$$\min_{\theta} \left\{ \frac{1}{\sqrt{2n}} \|Y - X\theta\| + \frac{\lambda}{n} \|\theta\|_1 \right\} = \min_{\theta} \max_{P: D_{\infty}(P, P_n) \leq \frac{4\lambda}{n}} E_P[\|Y - X\theta\|^2]$$

$$D_{\infty}(P, P_n) := \inf_{\pi \in \Pi(P, P_n)} E_{\pi}[\|Z_1 - Z_2\|_{\infty}]$$

# $l_1$ トレンドフィルタリングを活用した現象発見 (Yano and Kano, 2023)



観測変位の中に潜むスロースリップイベントを検出



ドメイン知識：

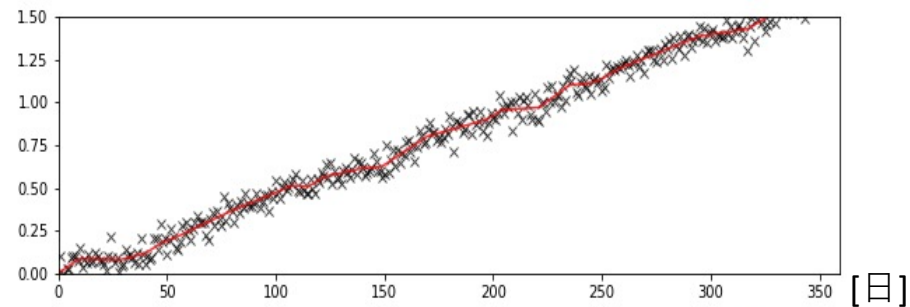
定常すべり + スロースリップの影響

→ 区分線形関数！！

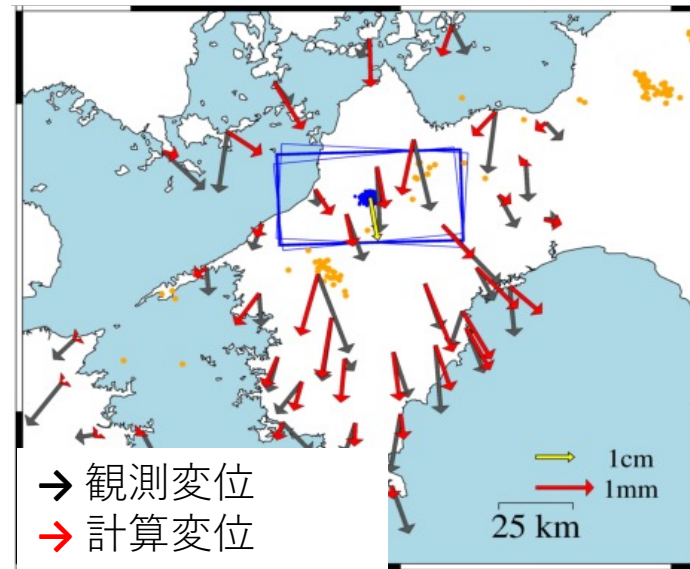
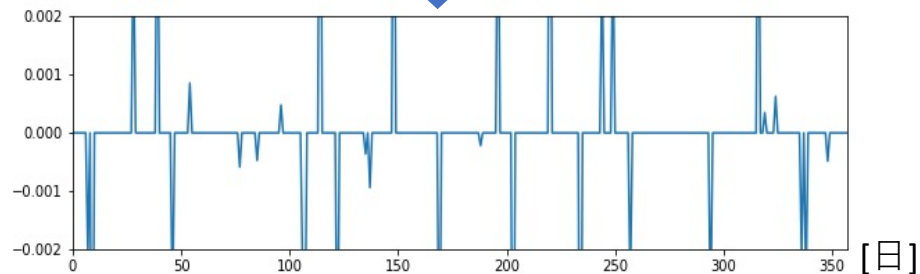


$l_1$ トレンドフィルタリング

$$\mathcal{L}(\theta) = \|Y - \theta\|^2 + \lambda \sum |\theta_{t-1} - 2\theta_t + \theta_{t+1}|$$



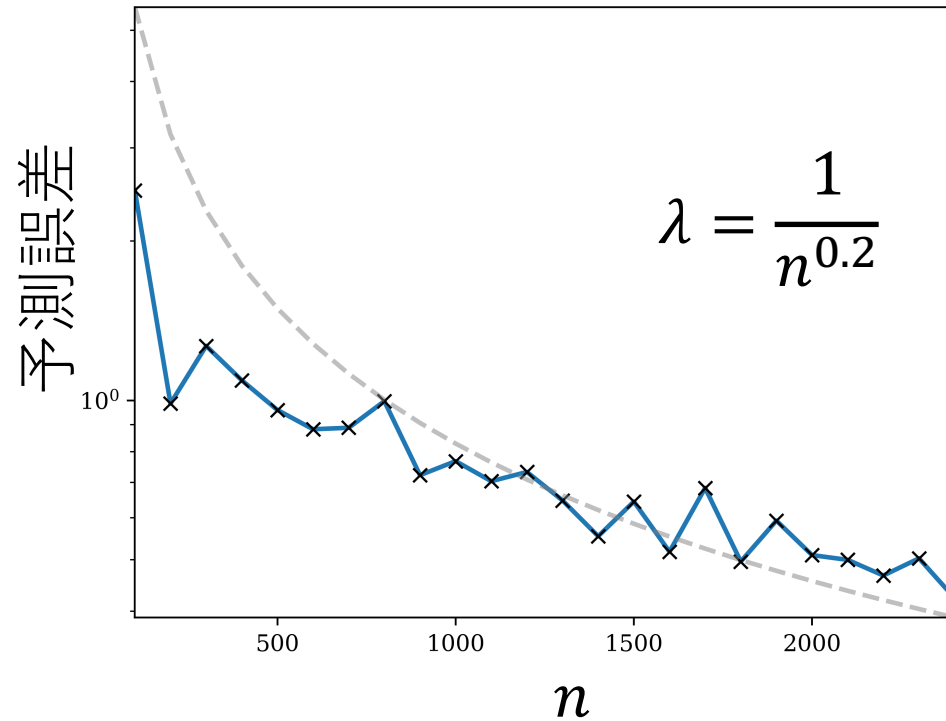
2階差分



# ハイパーパラメータ $\lambda$ の選択

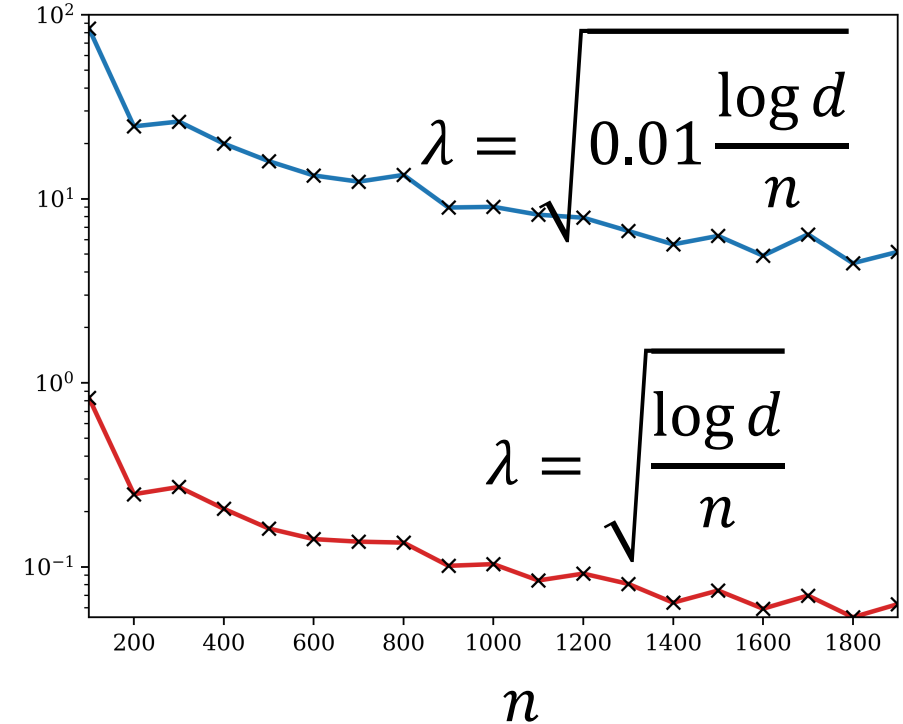
$\lambda$ の取り方が適切でないとスパース推定の技法は実際には使えない

$d = 2n$ の時の予測誤差



オーダーが違う $\lambda$ には使えない・・・

$d = 2n$ の時の予測誤差

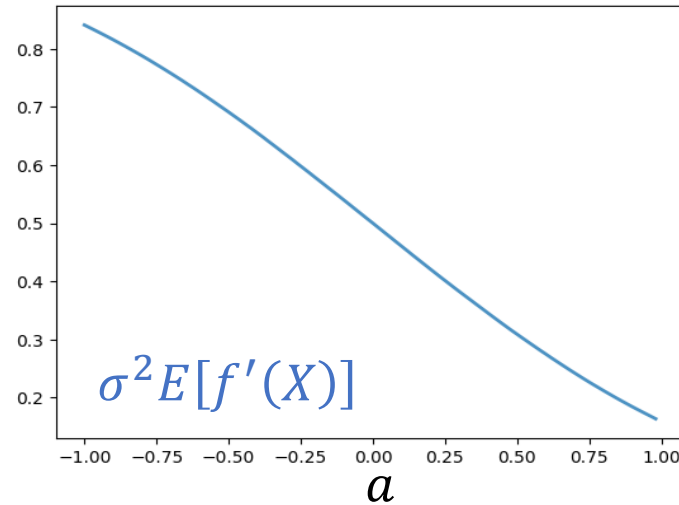
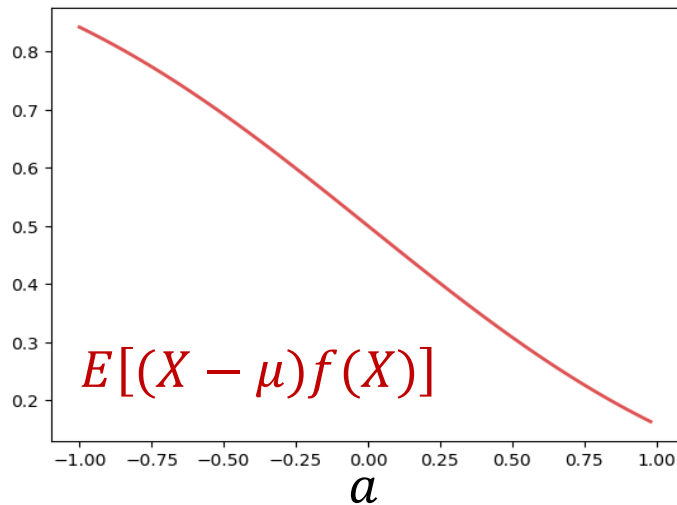


定数倍の違いなのに・・・

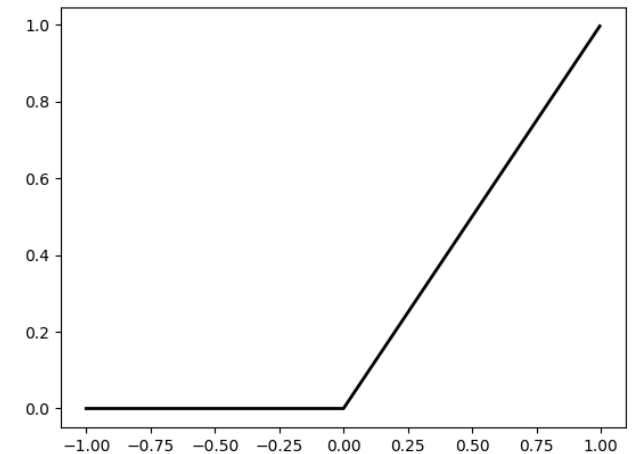
# 正規分布の部分積分：Steinの等式

- Steinの等式：正規分布に関する「部分積分」 [Stein (1973)]

$$E[(X - \mu)f(X)] = \sigma^2 E[f'(X)], \quad X \sim N(\mu, \sigma^2)$$



$$\text{reLU}(x; a) = \max(x - a, 0)$$





# Mallows' $C_p$ / Stein's unbiased risk estimate

Steinの等式を利用すると

LASSOの予測二乗誤差の「良い推定量」を作ることができる

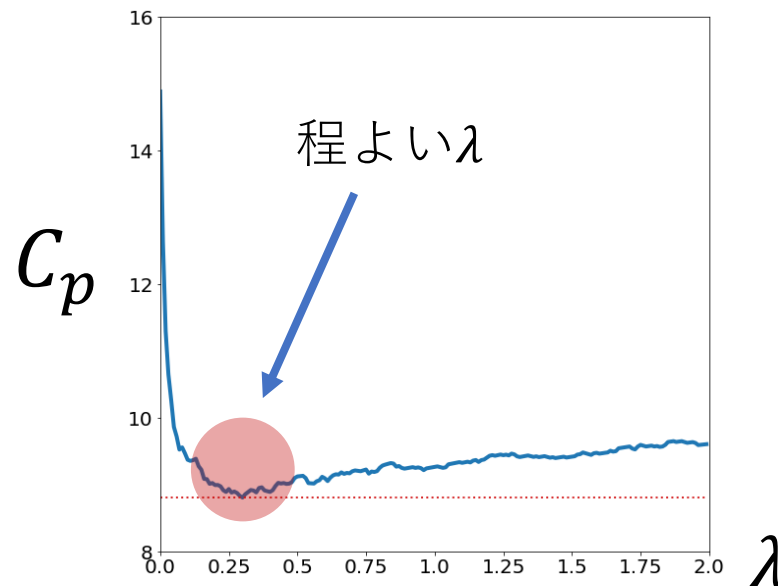
定理 (Zou, Hastie, Tibshirani, 2007)

観測ノイズがガウスであるとき、

$$E[\| \tilde{Y} - X\hat{\theta}_\lambda \|^2] = E[\| Y - X\hat{\theta}_\lambda \|^2 + 2\sigma^2 \|\hat{\theta}_\lambda\|_0]$$

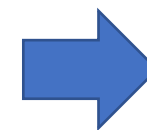
スパースにしない

$\lambda$ 小



よりスパースに

$\lambda$ 大



# LASSOの確率的挙動：LASSOの別の定式化

残差2乗和は次の最大化問題に帰着できる：

$$\frac{1}{2} \|x\|^2 = \max_u \left\{ u^\top x - \frac{1}{2} \|u\|^2 \right\}$$

 Legendre変換

LASSOはmin-max問題として定式化できる：

$$\min_{\theta \in \mathbb{R}^d} \left\{ \frac{1}{2n} \|Y - X\theta\|_2^2 + \frac{\lambda}{n} \|\theta\|_1 \right\} = \min_{\theta \in \mathbb{R}^d} \max_{u \in \mathbb{R}^n} \left\{ \frac{-1}{2n} u^\top X\theta + u^\top Y - \frac{1}{2n} \|u\|_2^2 + \frac{\lambda}{n} \|\theta\|_1 \right\}$$

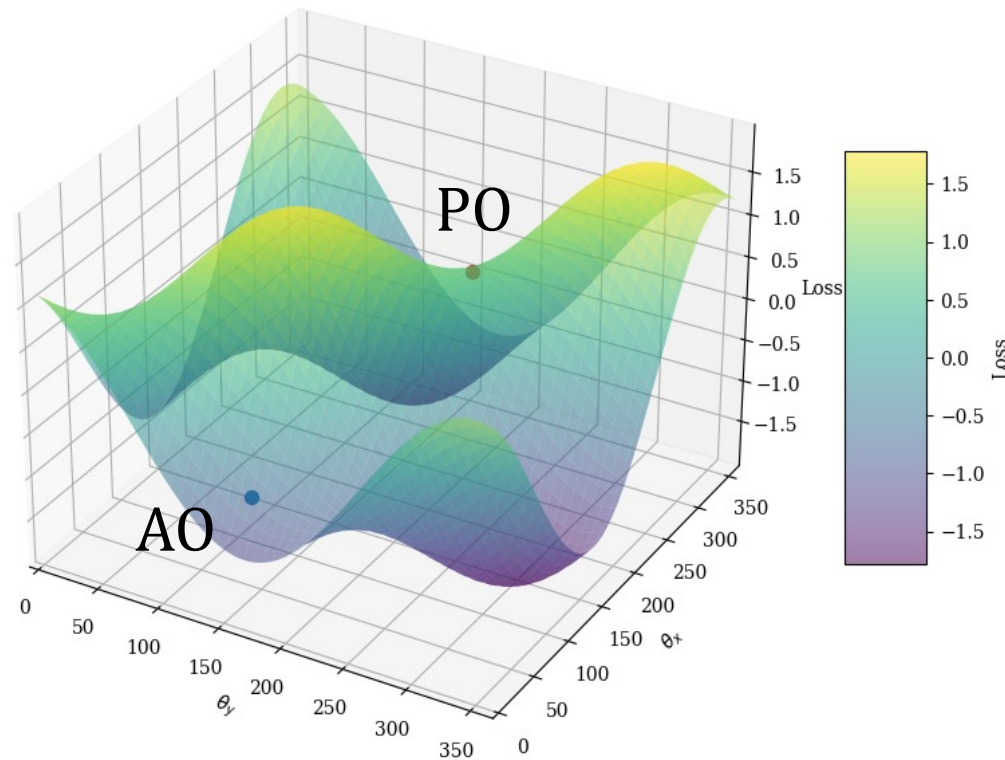


$$PO = \min_{\theta \in \mathbb{R}^d} \max_{u \in \mathbb{R}^n} \left\{ \frac{-1}{2n} u^\top X\theta + \psi_\lambda(u, \theta) \right\}$$

# LASSOの確率的挙動：LASSOの補助問題

Min-max問題の二次形式  $u^T X \theta$  は  $g, h \sim N(0, \sigma^2)$  を用いて確率的に分離可能

$$PO \sim \text{AO} = \min_{\theta \in \mathbb{R}^d} \max_{u \in \mathbb{R}^n} \left\{ \frac{1}{2n} \|u\| g^T \theta + \|\theta\| h^T u + \psi_\lambda(u, \theta) \right\}$$



\* Convex Gaussian min-max theorem (Thrampoulidis, Omyak, Hassibi, 2015)

\* Hayakawa (2023) Asymptotic Performance Prediction for ADMM-Based Compressed Sensing

# LASSOの確率的挙動：LASSOの一次元問題への帰着

補助問題は比例的高次元 ( $d/n \rightarrow \gamma$ ) で「一次元のmax-min問題」に帰着可能

$$\max_{b \geq 0} \min_{\tau \geq \sigma} \psi_\lambda(b, \tau)$$

$$\text{where } \psi_\lambda(b, \tau) := \left( \frac{\sigma^2}{\tau} + \tau \right) \frac{b}{2} - \frac{b^2}{2} + \frac{1}{\delta} \mathbb{E} \left[ \min_{\omega \in \mathbb{R}} \left\{ b \frac{\omega^2}{2\tau} - b\omega W + \lambda|\omega + \Theta| - \lambda|\Theta| \right\} \right]$$

複雑な式に見える・・・

しかし、実は解ける

$$\tau_*^2 = \sigma^2 + \frac{1}{\delta} \mathbb{E} \left[ \left\{ S_{\frac{\tau_*}{b_*}} \lambda (\Theta + \tau_* W) - \Theta \right\}^2 \right],$$
$$b_* = \tau_* \left\{ 1 - \frac{1}{\delta} \Pr \left( |\Theta + \tau_* W| \geq \frac{\tau_*}{b_*} \lambda \right) \right\}.$$

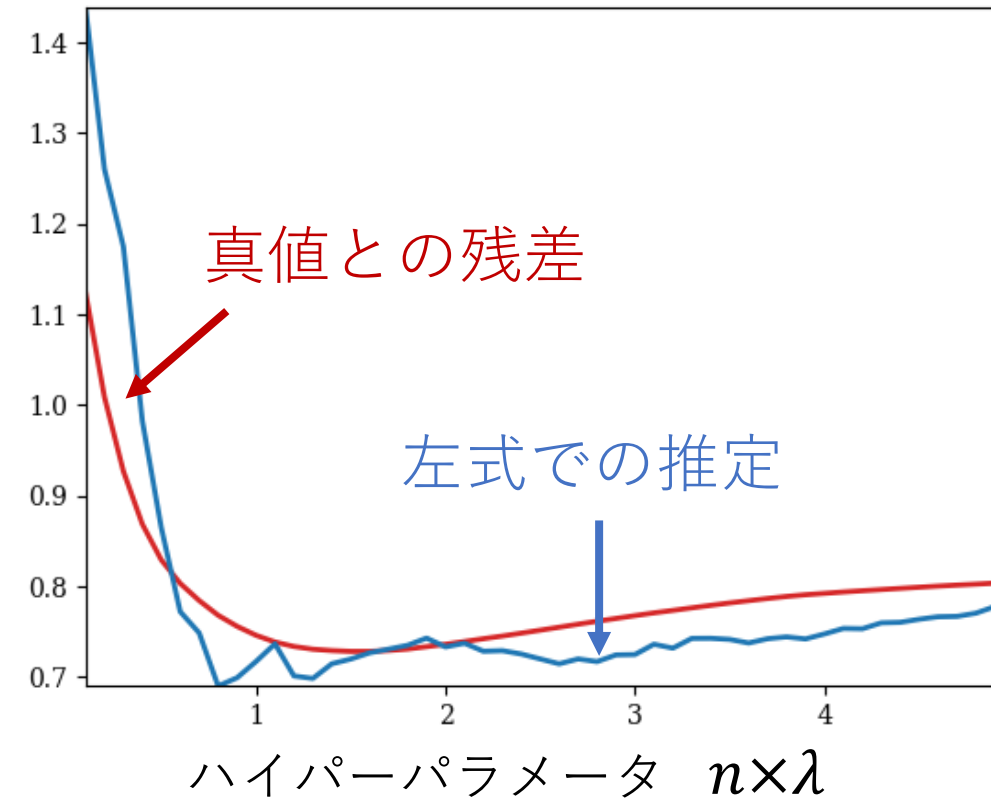
# LASSOの高次元での確率的挙動

- 先ほどの解 $\tau_*$  は次の形で簡単に推定可能

$$\hat{\tau} := \frac{\sqrt{n\|Y - X\hat{\theta}_\lambda\|^2}}{n - [\hat{\theta}_\lambda \text{の非ゼロ個数}]}$$

LASSOと真値の平均的な残差の確率挙動は比例的高次元 ( $d/n \rightarrow \gamma$ )のもと、

$$\frac{\sum_i (\hat{\theta}_\lambda - \theta^0)_i^2}{d} \rightarrow \frac{n}{d} (\hat{\tau}^2 - \sigma^2)$$



スパース推定量の(高次元での)確率的な挙動が明らかになった

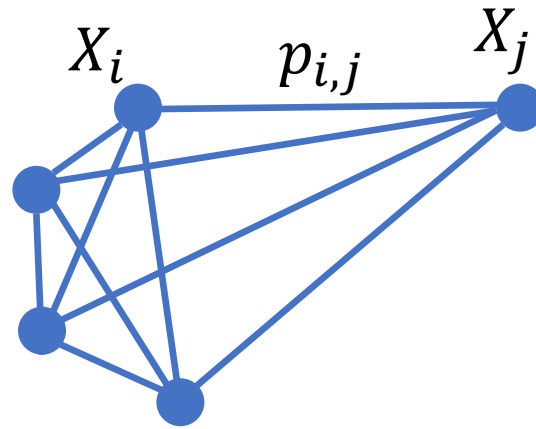
\* Han and Shen (2023) Universality of regularized regression estimators in high dimensions

\* Sawaya, Uematsu, Imaizumi (2024) HIGH-DIMENSIONAL SINGLE-INDEX MODELS

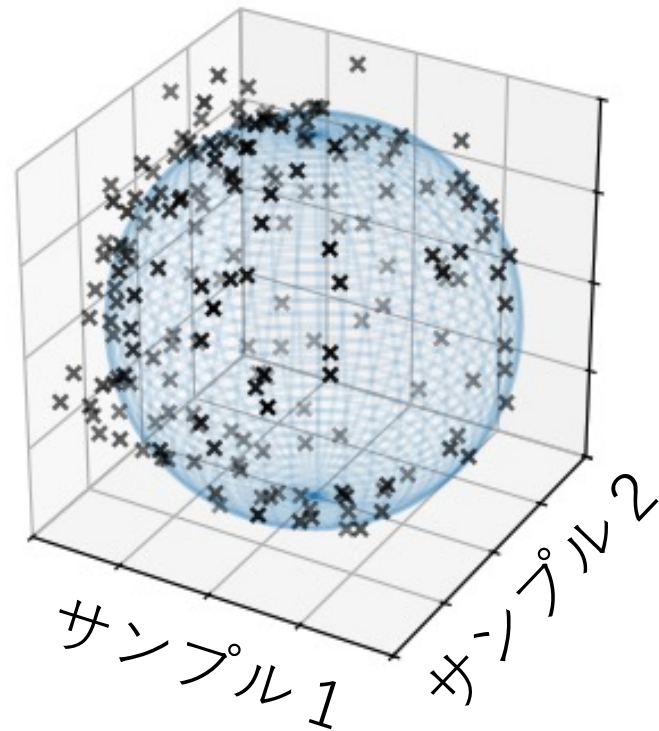
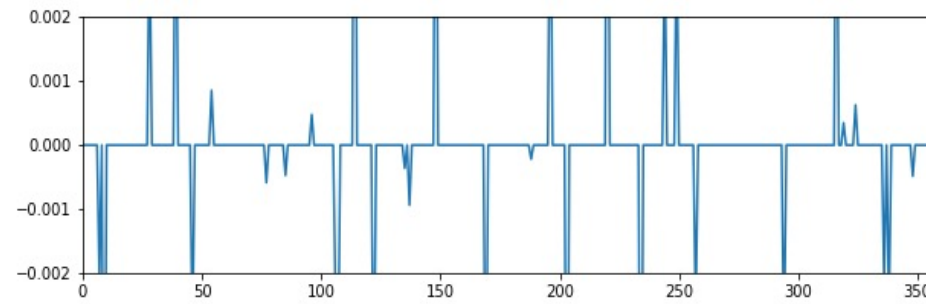
# 高次元のデータ解析のアイデア

球面・軸集中現象

グラフ構造

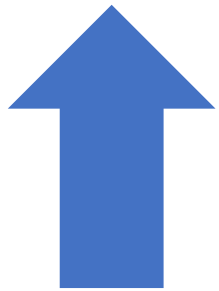


スパース性



# 本日の内容

可視化



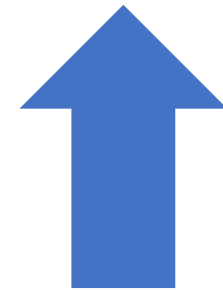
次元圧縮

発見・予測



正則化

不確実性評価



ベイズ

# ベイズ統計

「ベイズの定理」に基づき推定から不確実性評価までを一気貫通

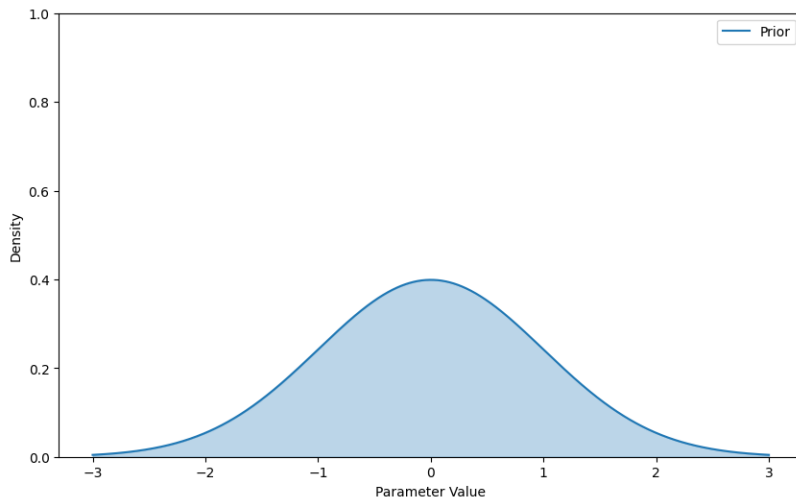
$$\pi(\theta)$$

×

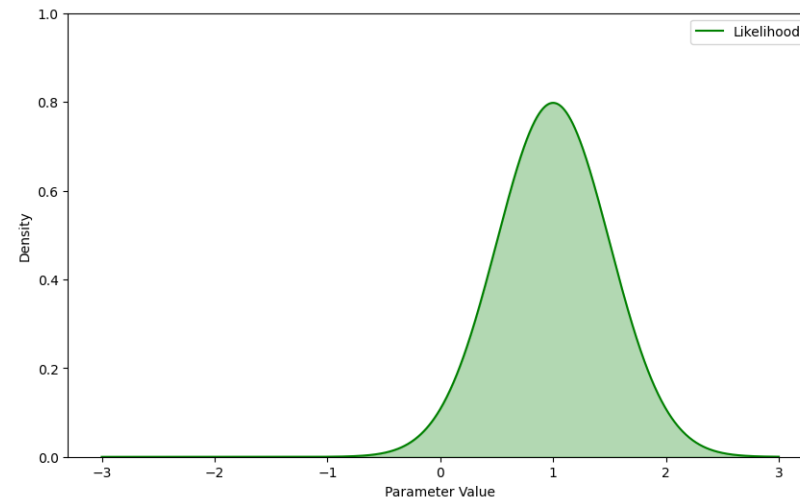
$$p(X_1, \dots, X_N | \theta)$$

$$\propto \pi(\theta | X_1, \dots, X_N)$$

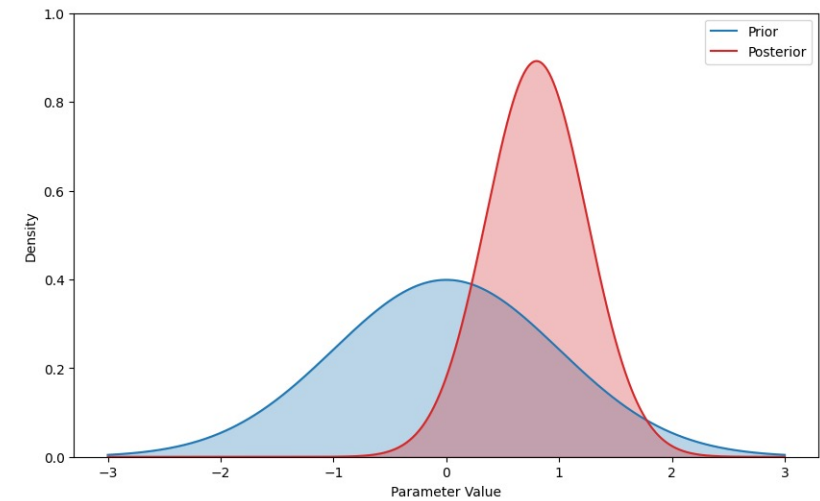
事前分布



尤度関数



事後分布





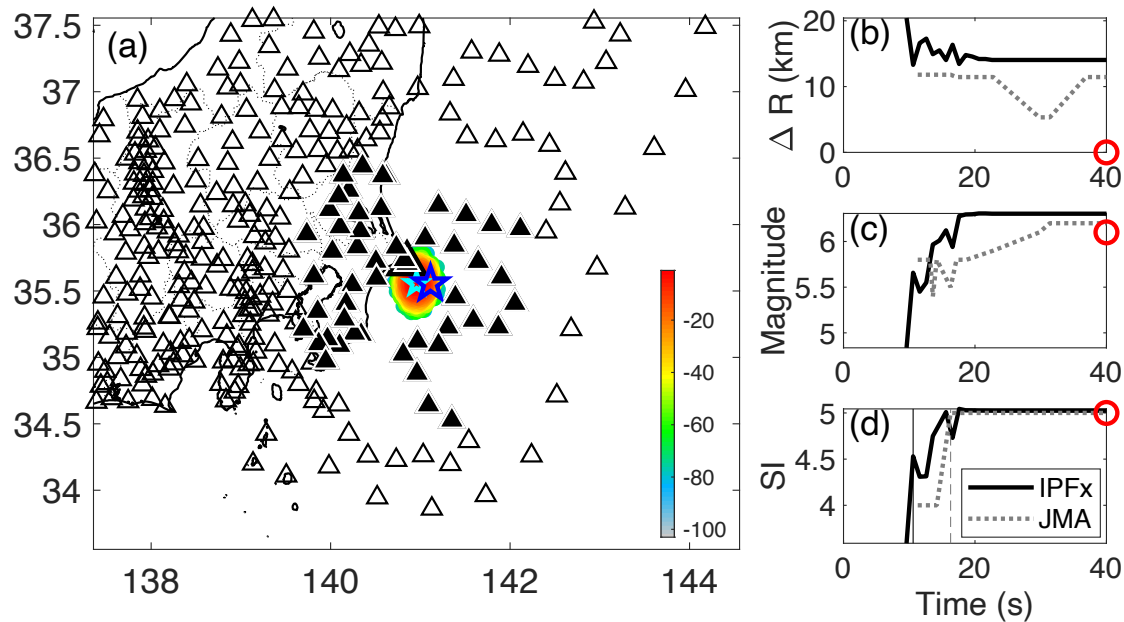
# ベイズ統計の利用：逆解析

緊急地震速報では震源をその推定の不確実性も込みで評価している

→ 推定結果の不確実性評価・多峰性の対処

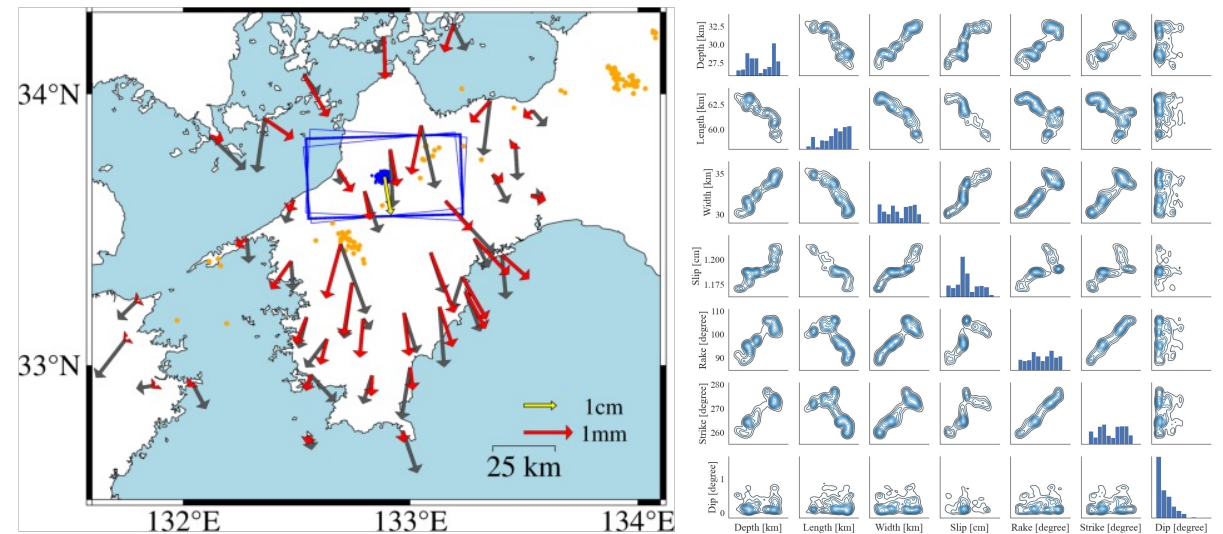
すべり推定などでも活躍

IPF法による緊急地震速報



From Yamada et al. (2021) *BSSA*

震源情報  $\theta$  (滑った断層の大きさや位置)の事後分布



From Yano and Kano, 2022, *J. Geophysical Research: Solid Earth*

# 高次元でのベイズの利用

近年は高次元逆問題や関数自由度をもつ推定問題でのベイズの利用が注目  
→推定結果の不確実性評価・安定性解析

基底関数展開

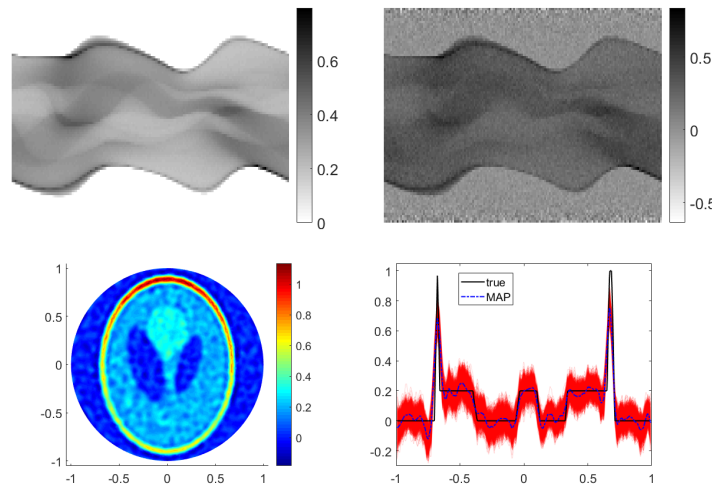
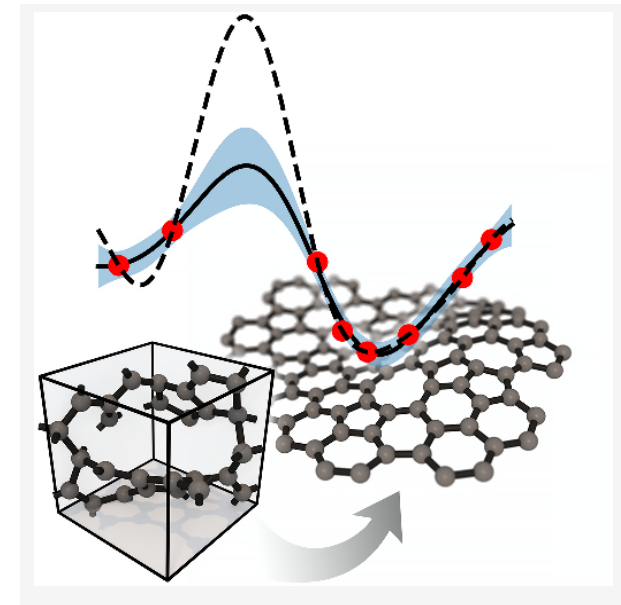


FIG 4. Example 2. Left to right. Top row:  $I_{f_1}$ ;  $I_{f_1}$  noisy (with  $\beta$  on the horizontal axis and  $\alpha$  on the vertical axis). Bottom row: posterior mean; cross-section on  $\{x_2 = 0\}$  of 2000 posterior samples.

Monard et al., AoS, 2020

ガウス過程回帰



Deringer, et al., chemical reviews, 2021

# 高次元でのベイズの利用

近年は高次元逆問題や関数自由度をもつ推定問題でのベイズの利用が注目  
→推定結果の不確実性評価・安定性解析

基底関数展開

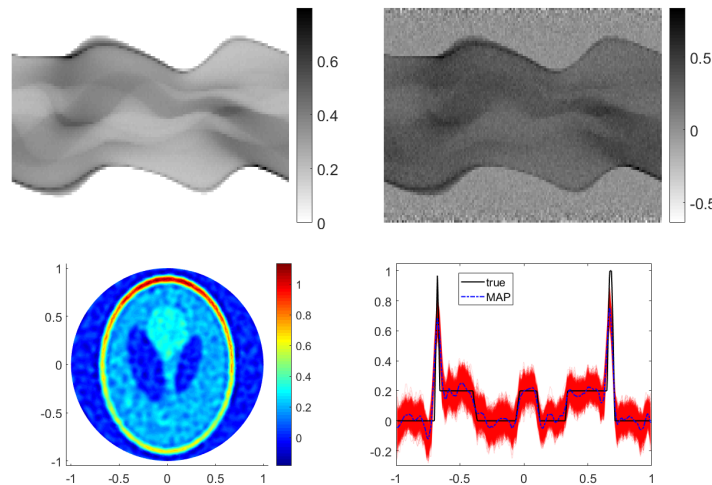
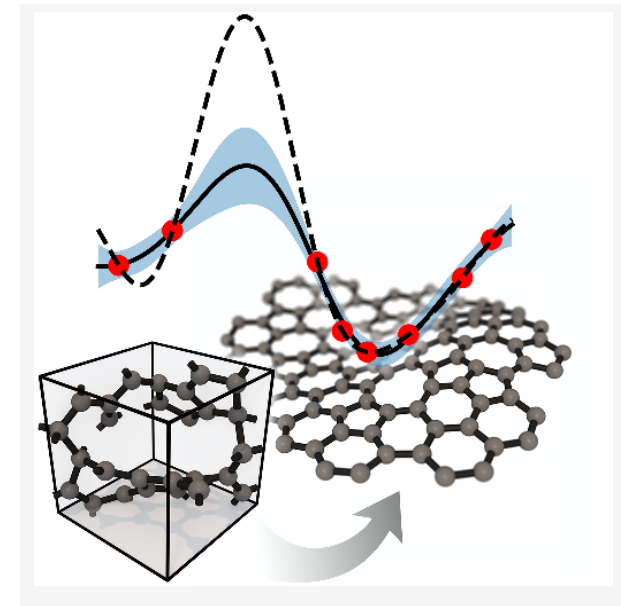


FIG 4. Example 2. Left to right. Top row:  $I_{f_1}$ ;  $I_{f_1}$  noisy (with  $\beta$  on the horizontal axis and  $\alpha$  on the vertical axis). Bottom row: posterior mean; cross-section on  $\{x_2 = 0\}$  of 2000 posterior samples.

Monard et al., AoS, 2020

ガウス過程回帰



Deringer, et al., chemical reviews, 2021

高次元・無限次元でのベイズの不確実性評価は妥当な不確実性評価？

# 事後分布と繰り返し抽出での不確実性の不一致

関数自由度をもつ事前分布を利用するとき (e.g., ガウス過程)

事後分布

$\sqrt{n}(f - \hat{f}) \mid X = x$  の分布

$\neq$

繰り返し抽出での推定量分布

$\sqrt{n}(\hat{f} - f_0) \mid f_0$  の分布

as  $n \rightarrow \infty$

# 事後分布と繰り返し抽出での不確実性の不一致

関数自由度をもつ事前分布を利用するとき(e.g., ガウス過程)

事後分布

$\sqrt{n}(f - \hat{f}) \mid X = x$ の分布

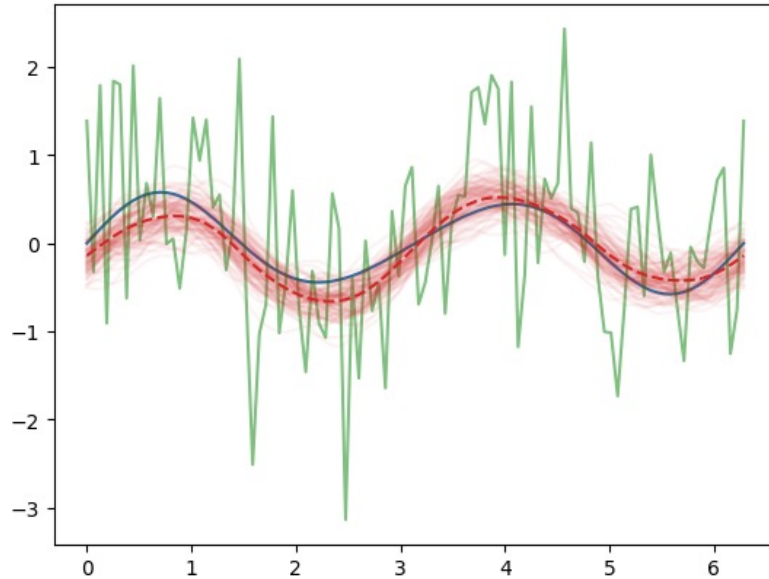
$\neq$

繰り返し抽出での推定量分布

$\sqrt{n}(\hat{f} - f_0) \mid f_0$ の分布

as  $n \rightarrow \infty$

事後分布からのサンプル



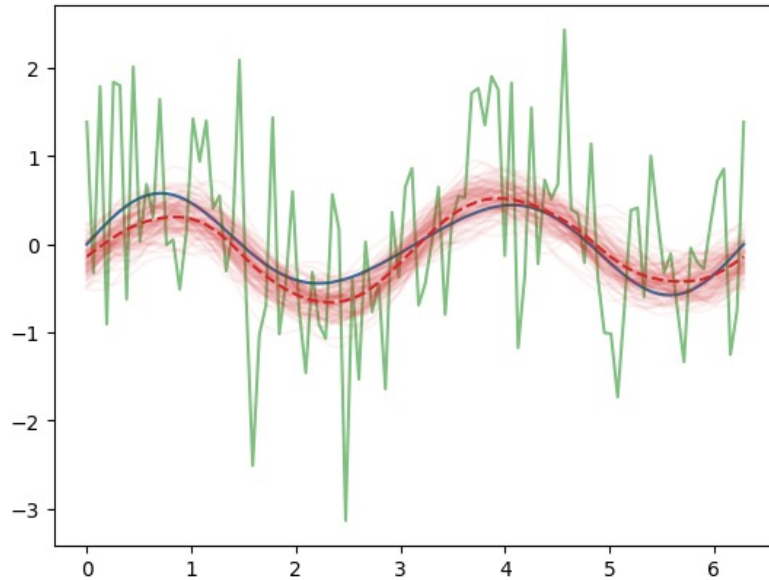
# 事後分布と繰り返し抽出での不確実性の不一致

関数自由度をもつ事前分布を利用するとき (e.g., ガウス過程)

事後分布

$\sqrt{n}(f - \hat{f}) \mid X = x$  の分布

事後分布からのサンプル

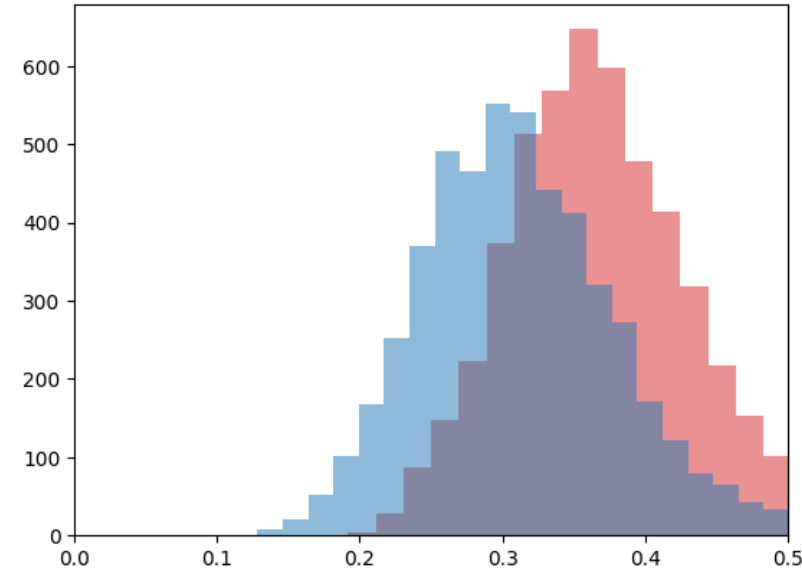


繰り返し抽出での推定量分布

$\sqrt{n}(\hat{f} - f_0) \mid f_0$  の分布

as  $n \rightarrow \infty$

$\|f - \hat{f}\|_2$  と  $\|\hat{f} - f_0\|_2$  のヒストグラム



# 事後分布と繰り返し抽出での不確実性の不一致

関数自由度をもつ事前分布を利用するとき (e.g., ガウス過程)

事後分布

$\sqrt{n}(f - \hat{f}) \mid X = x$  の分布

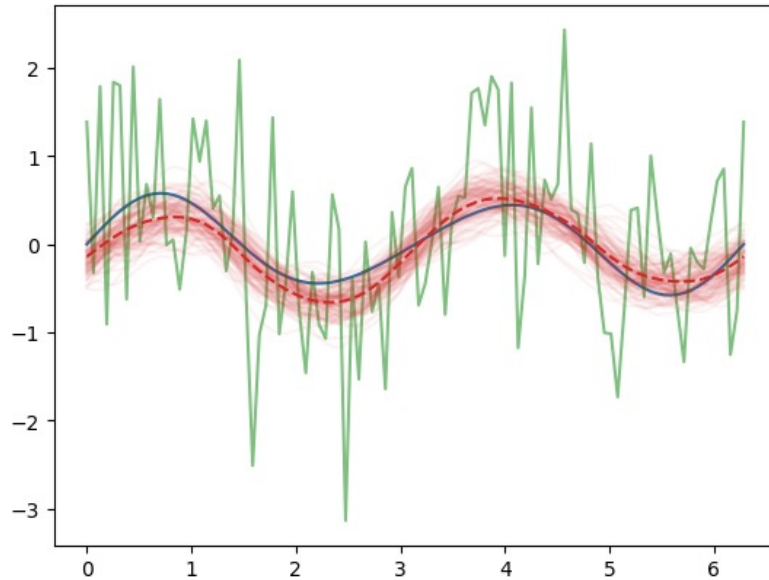
$\neq$

繰り返し抽出での推定量分布

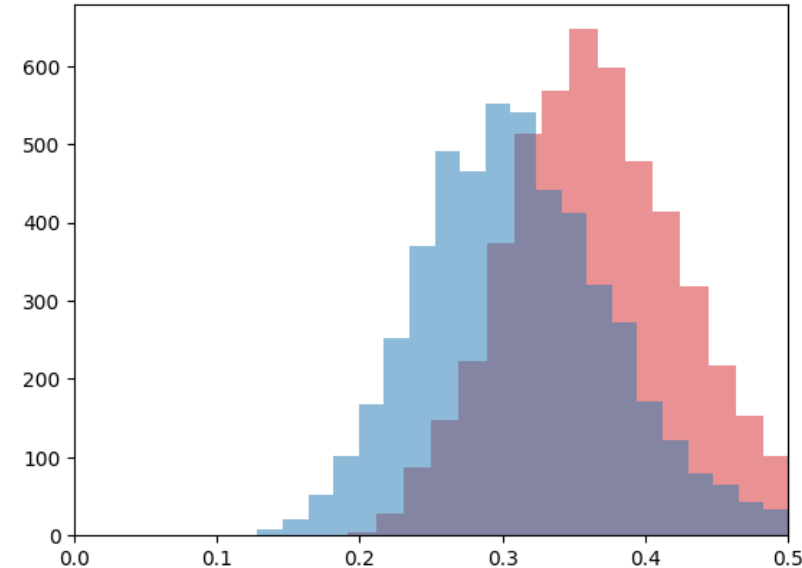
$\sqrt{n}(\hat{f} - f_0) \mid f_0$  の分布

as  $n \rightarrow \infty$

事後分布からのサンプル



$\|f - \hat{f}\|_2$  と  $\|\hat{f} - f_0\|_2$  のヒストグラム



高次元・無限次元でのベイズの不確実性評価では注意が必要

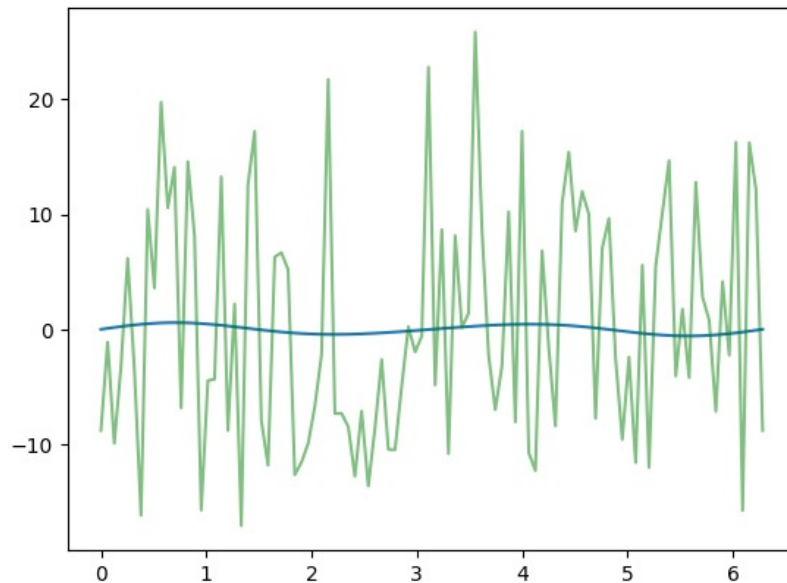
# Castillo & Nickl (2013) のアイデア：積分の利用

比較の際に「高周波成分」が邪魔をする

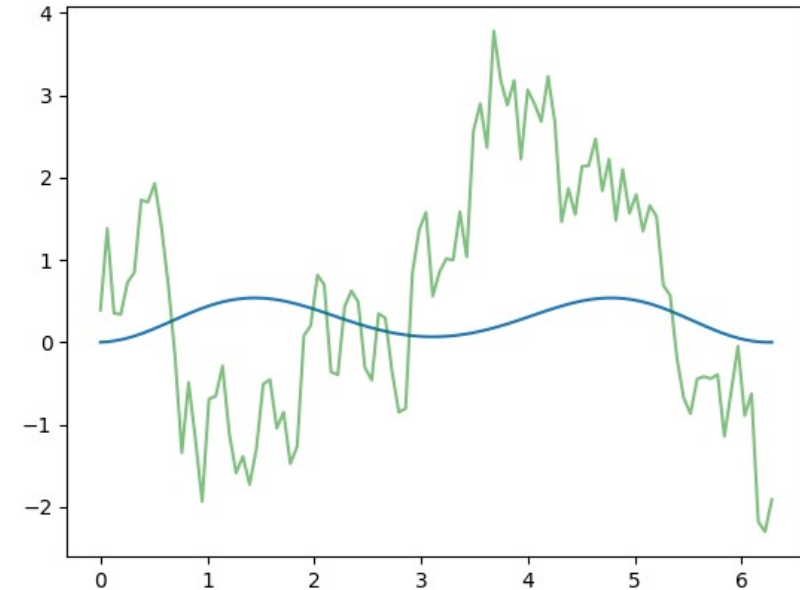
→ 積分することで高周波成分の影響を減らして考える

c.f., Sobolevの不等式  $\|F\|_2 \leq C \|\nabla F\|_2$

$f(t), dX(t)$



$\int^t f(\tau) d\tau, \int^t dX(\tau)$





# Castillo & Nickl (2013) のアイデア : weak norm

必要な回数積分する操作を一般化し、  
以下のような高周波成分を低減するようなノルムを考える：

strong inner-product  $\langle f, g \rangle_2 = \sum_i \langle f, \phi_i \rangle \langle g, \phi_i \rangle$



weak inner-product  $\|f\|_w^2 = \sum_i w_i \langle f, \phi_i \rangle \langle g, \phi_i \rangle$  with  $w_i \rightarrow 0$

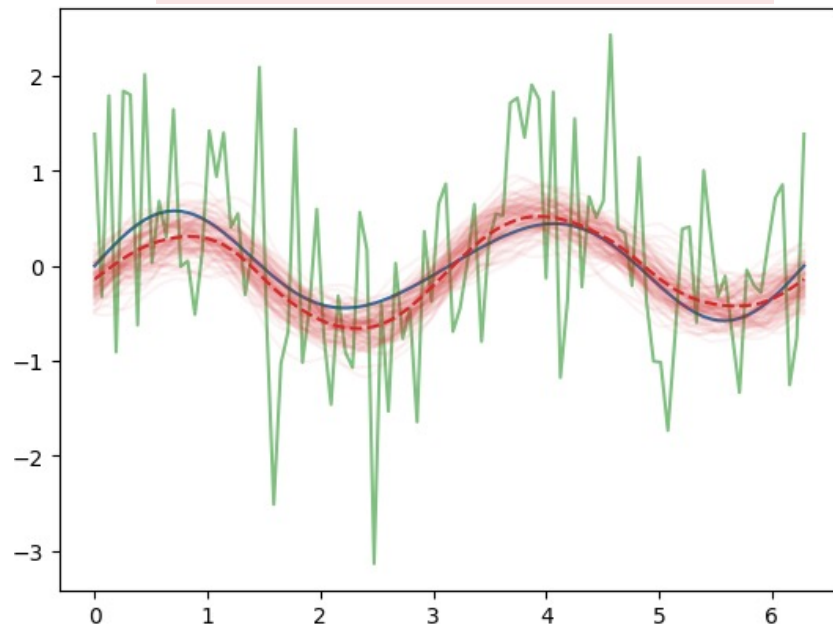
\* 弱い内積によるヒルベルト空間を  $\mathcal{H}(w)$  と書く

# 一致の「位相」を変える

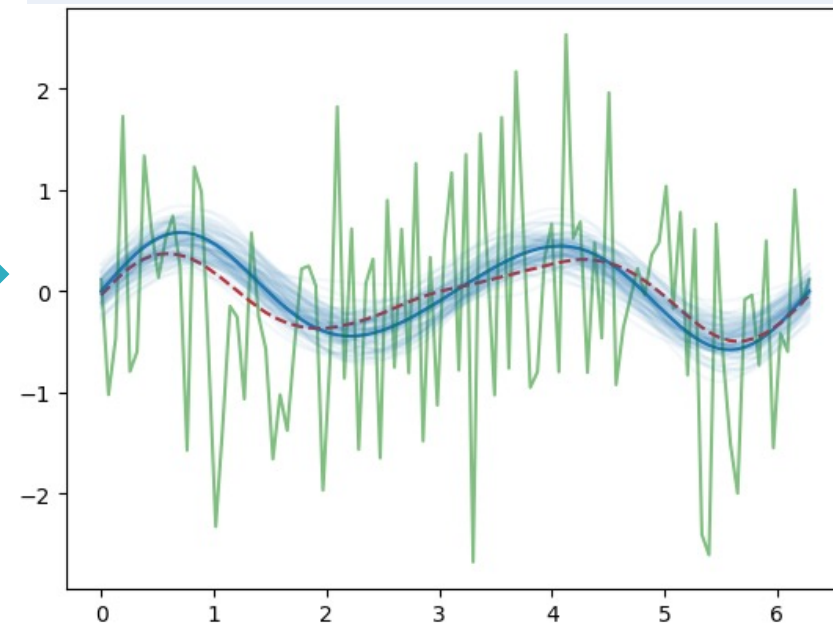
Castillo and Nickl (2013, 2014) : 分布の一致を  $\|\cdot\|_w$  で定義し直すことで

$$\sqrt{n}(f - \hat{f}) \mid X = x \text{ の分布} \quad \sim \quad \sqrt{n}(\hat{f} - f_0) \mid f_0 \text{ の分布} \quad \text{as } n \rightarrow \infty$$

事後分布  $f$  の揺らぎ

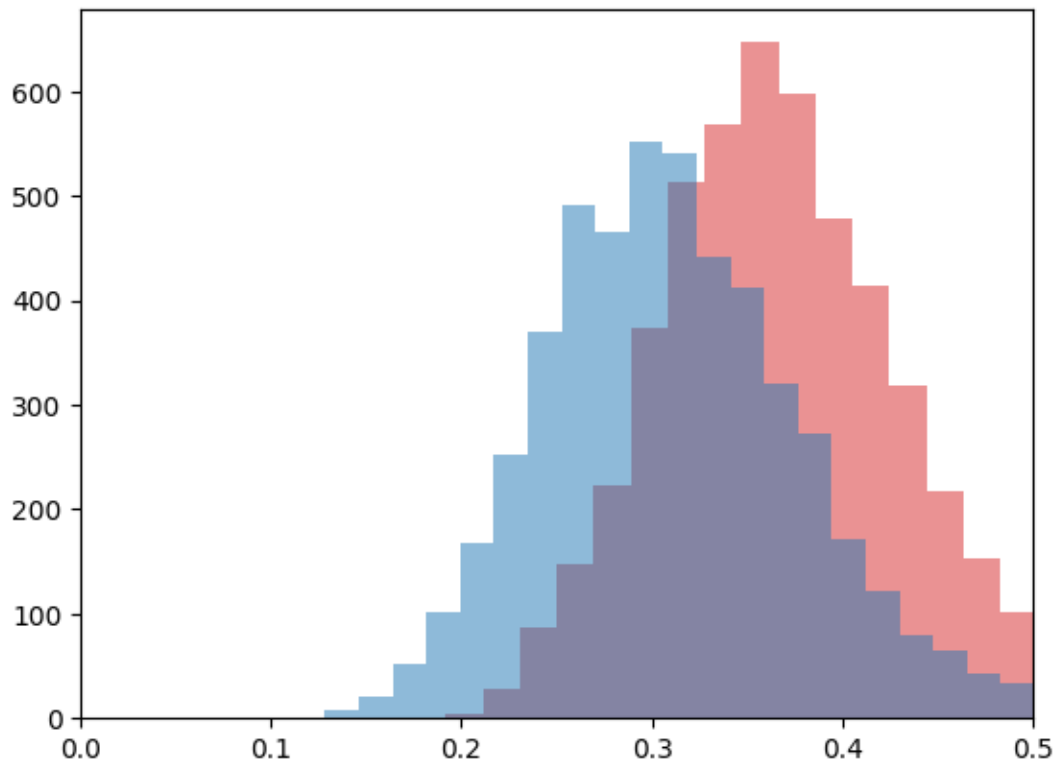


$\hat{f}$  の繰り返し抽出での揺らぎ

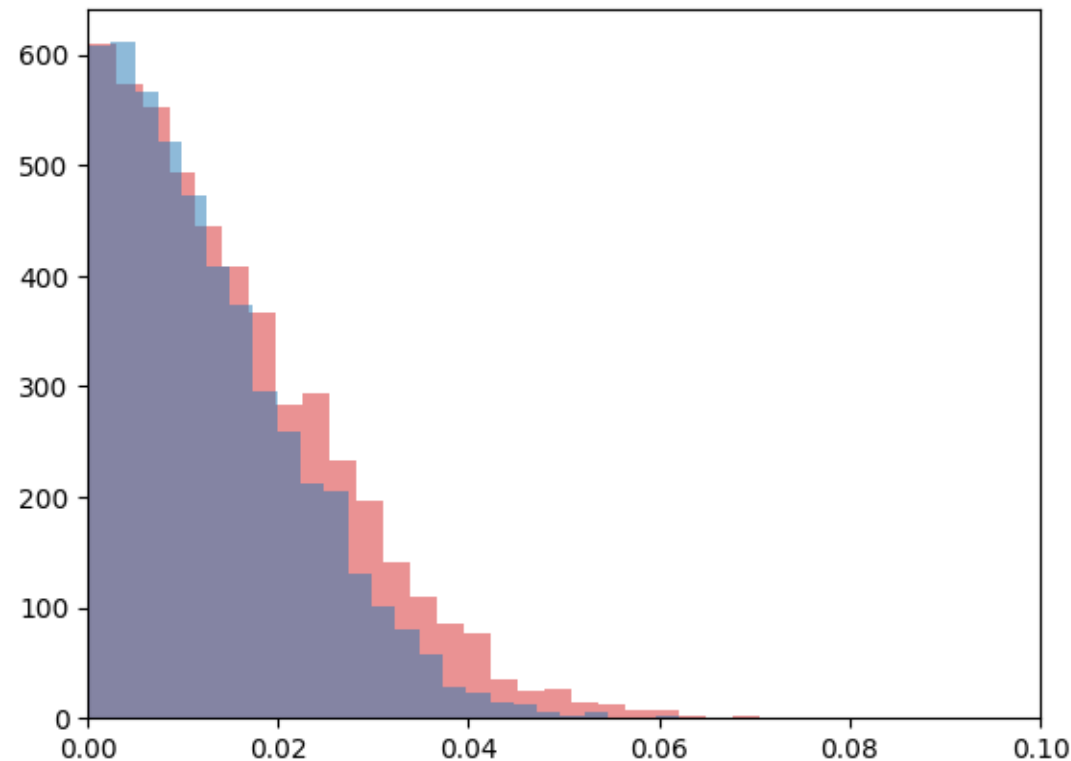


# 数値実験による検証：信用区間と信頼区間の幅の評価

$\|f - \hat{f}\|_2$  と  $\|\hat{f} - f_0\|_2$  のヒストグラム



$\|f - \hat{f}\|_w$  と  $\|\hat{f} - f_0\|_w$  のヒストグラム



幅の評価を変更するだけで正しい不確実性評価が行えていることがわかる

# ベイズにおけるモデルと事前分布の選択

ベイズでは「尤度関数」の設計と「事前分布」の設計が必要

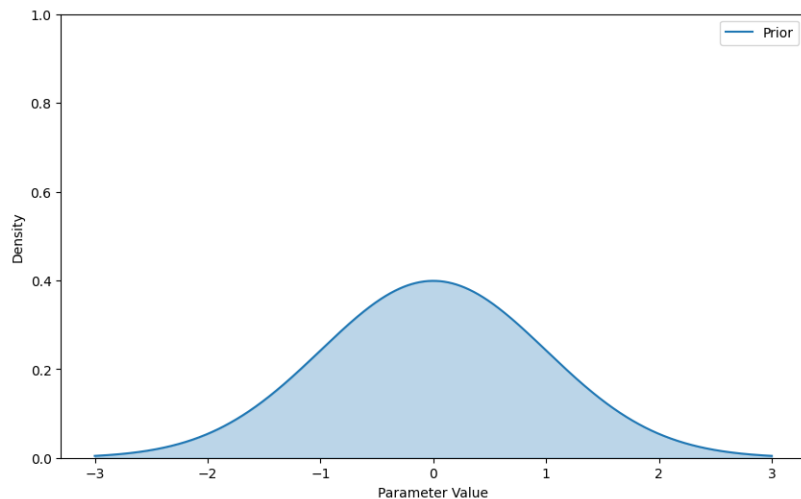
$$\pi(\theta)$$

×

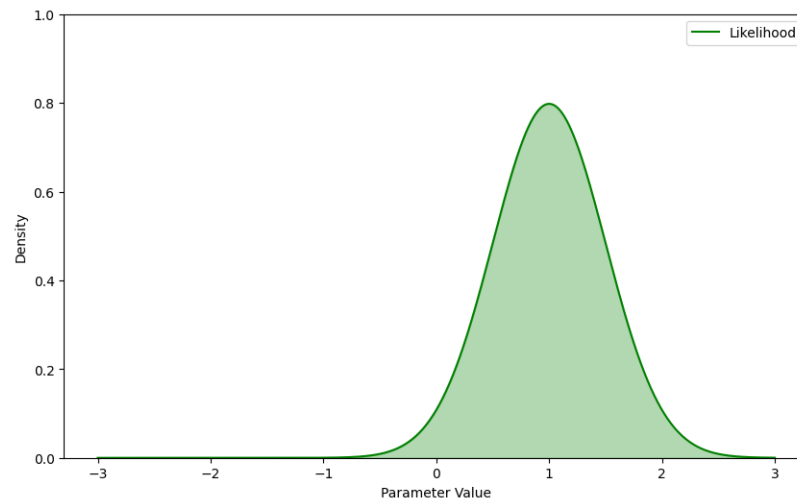
$$p(X_1, \dots, X_N | \theta)$$

$$\propto \pi(\theta | X_1, \dots, X_N)$$

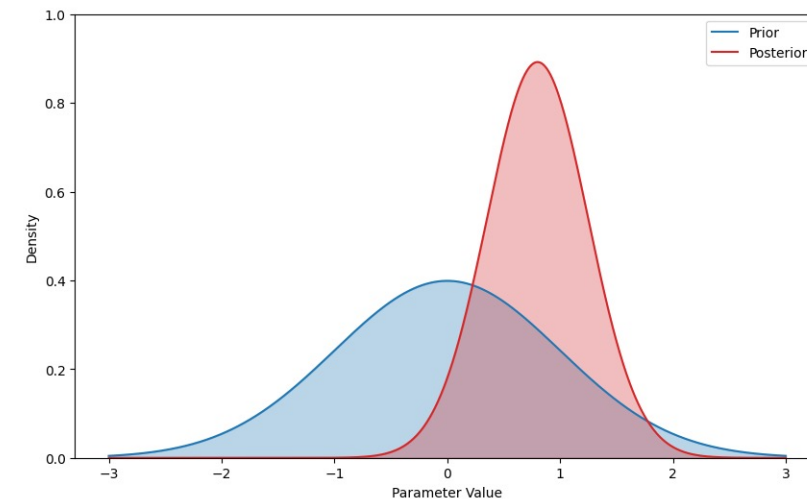
事前分布



尤度関数



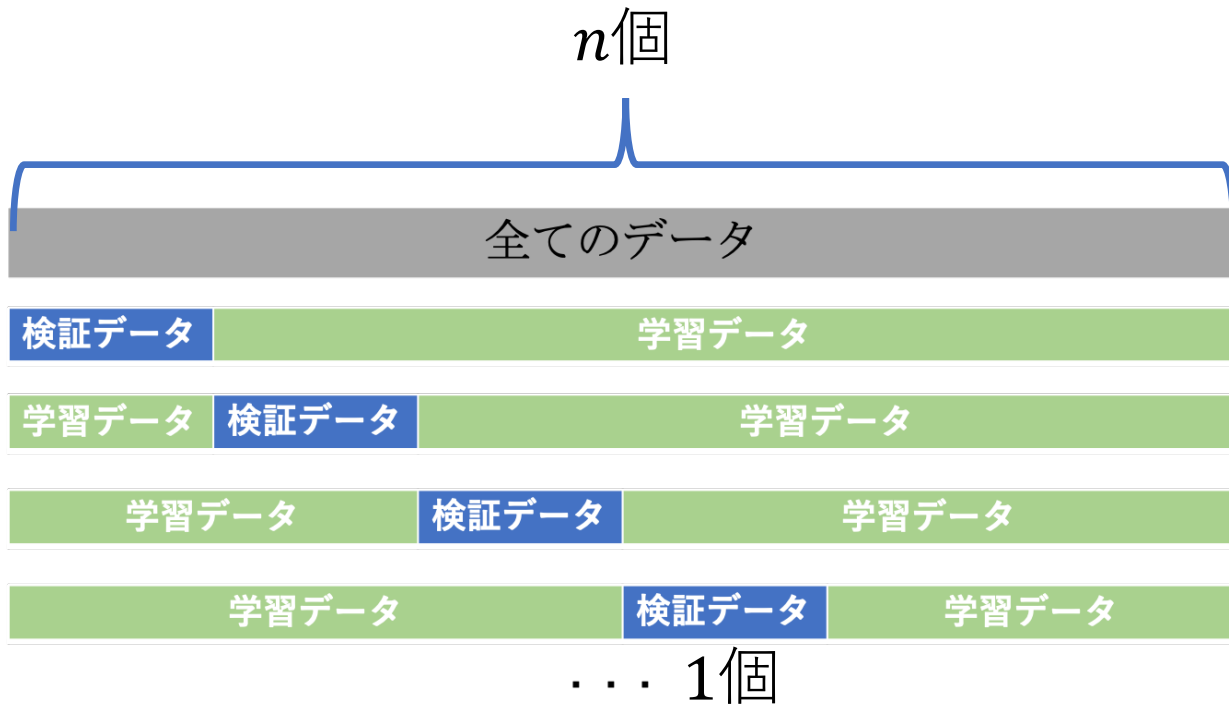
事後分布



ここではベイズの交差検証法・汎化誤差推定の観点で説明

# 交差検証法 (Leave-one-out CV; LOOCV)

- データを検証用と学習用にランダム分割し、学習用での学習結果を検証用で評価
- 機械学習モデルのチューニングパラメタ選択のデファクトスタンダード



$X^{-i}$ を $X_i$ を抜いたデータとして

$$\text{LOOCV}_B = \frac{1}{n} \sum_i \{-E_{\text{pos}}^{-i}[\log p(X_i | \theta)]\}$$

➤ 汎化損失(将来値での評価)にかなり近い

$$\mathcal{G}_B = E_{X_{n+1}} E_{X_1, \dots, X_n} [-E_{\text{pos}} \log [p(X_{n+1} | \theta)]]$$

$$E[\text{LOOCV}_B] = \mathcal{G}_B + o(1/n^2)$$

➤ ほとんど計算不可能

# ベイズモデルの評価：広く使える情報量規準 (WAIC)

WAIC (Watanabe, 2010) は汎化のためのモデル評価規準

$$\text{WAIC}_2 = \frac{1}{n} \sum_i \{ \underset{\substack{\uparrow \\ \text{事後分布による期待値}}}{E_{\text{pos}}} [-\log p(X_i | \theta)] \} + \frac{1}{n} \sum_i \underset{\substack{\uparrow \\ \text{事後分布による分散}}}{V_{\text{pos}}} [\log p(X_i | \theta)]$$

- 「全ての量が事後標本のみで計算可能」
- 多峰の事後分布でも汎化損失の妥当な推定値になっている
- 一個抜き交差検証法の近似になっている

# ベイズモデルの評価：事後共分散型情報量規準

評価する損失は対数尤度だけではなく、ユーザーが指定したい

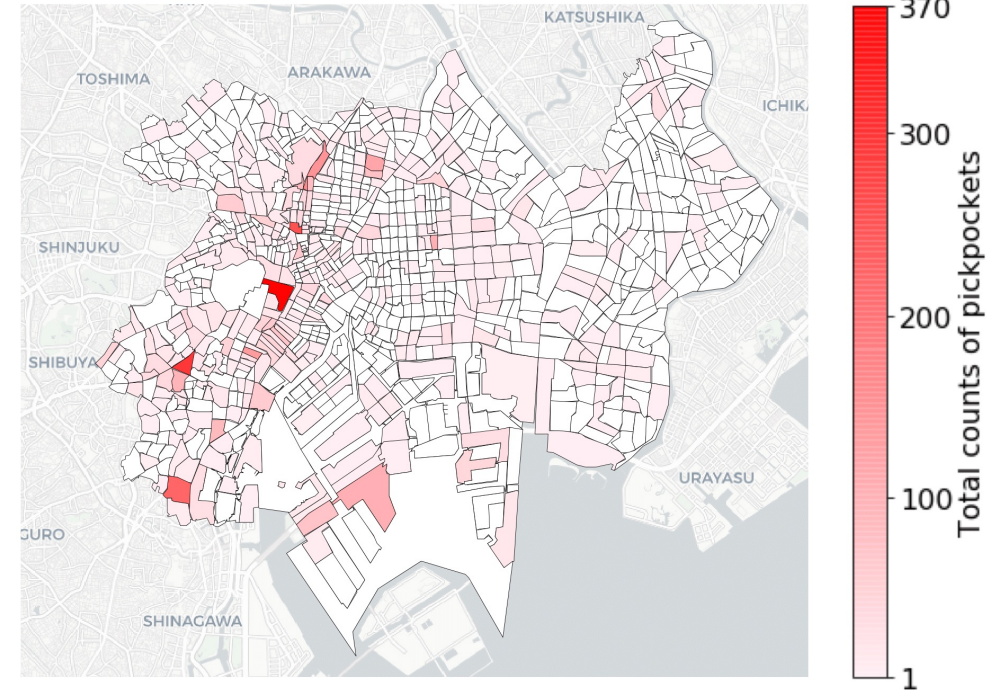
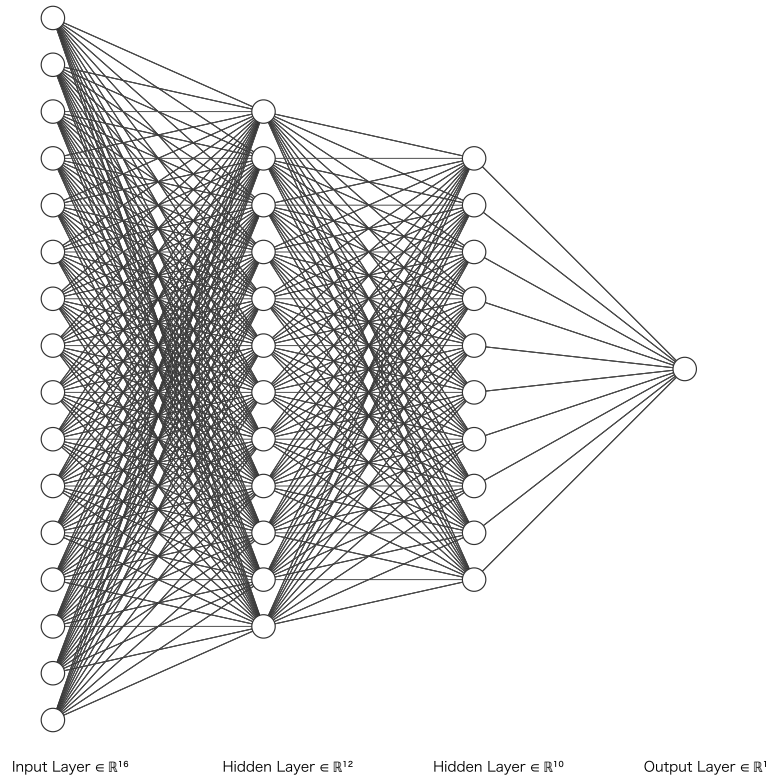
→ Iba and Y. (2023,2024) Posterior Covariance Information Criterionを提案

$$\text{PCIC}_G = \frac{1}{n} \sum_{i=1, \dots, n} \underbrace{E_{\text{pos}}[l(X_i, \theta)]}_{\text{事後分布による期待値}} - \frac{1}{n} \sum_i \underbrace{\text{Cov}_{\text{pos}}[l(X, \theta), \log p(X_i | \theta)]}_{\text{事後分布による共分散}}$$

WAICのもつ良い性質を保ったまま任意の損失を扱えるように

# 高次元への眼差し

WAICやPCICは「高次元」でも機能するのだろうか？



ベイズニューラルネットワーク・深層学習

高次元データ



# 高次元線形回帰でのWAIC

高次元線形回帰では WAICは良い汎化推定量

Theorem 1 in Okuno and Y. (2023)

$X_1, \dots, X_n$  : 独立同一に共分散行列  $\Sigma$  をもつガウス分布に従う共変量  
 $\xi = \text{tr}(\Sigma)$  &  $b = d^{1/2} \|\beta\|_\infty$

この時、任意の  $\varepsilon > 0$  に対し

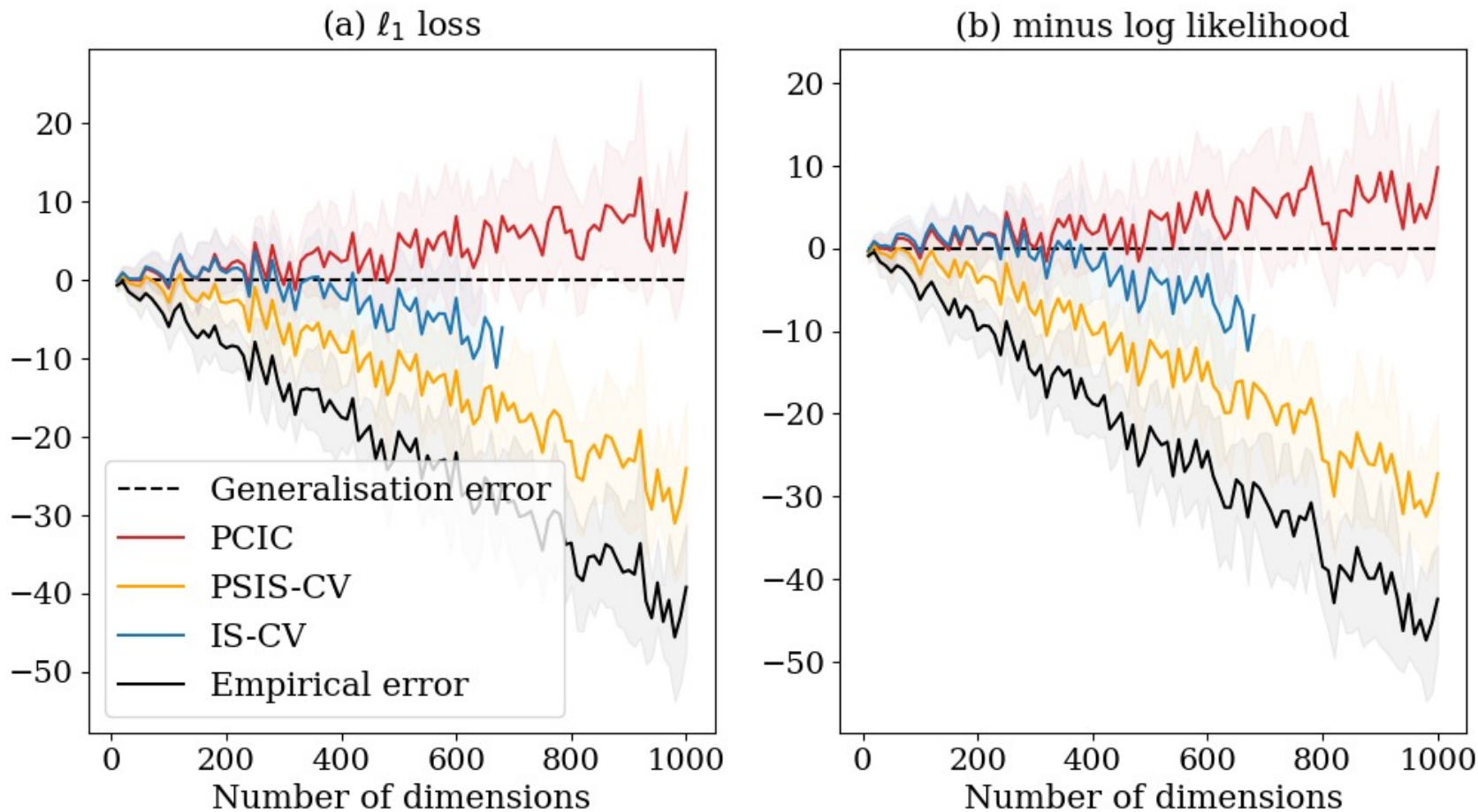
$$\Pr(|\text{WAIC} - \mathcal{G}_G| > \varepsilon) \leq C_{\xi, b} \left( \frac{1}{n} \right)$$

\* 次元数によらない収束

\* 収束スピードは共変量共分散の対角和  $\xi$  と回帰係数の大きさ  $b$  で決まる

# 高次元こそPCICやWAICが光る？

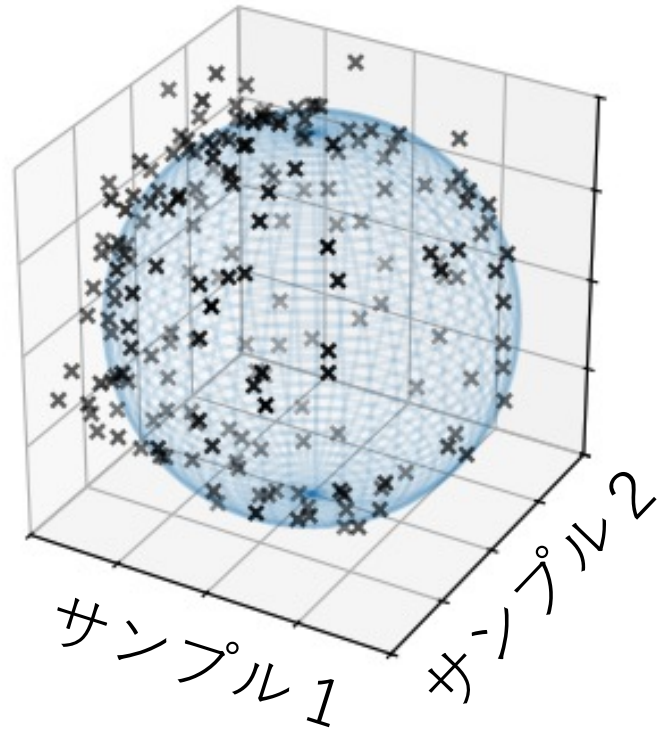
他の汎化誤差の指標との比較



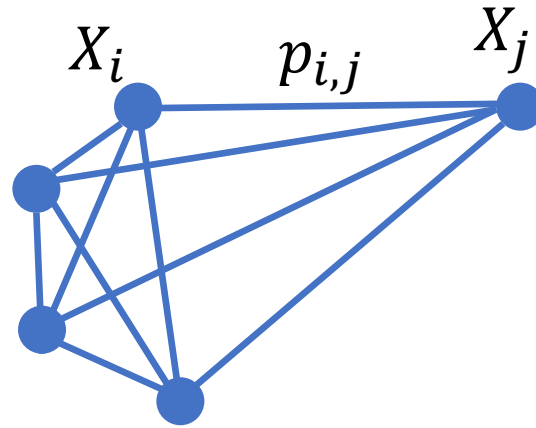
PCICやWAICは高次元でも機能する

# 高次元のデータ解析のアイデア

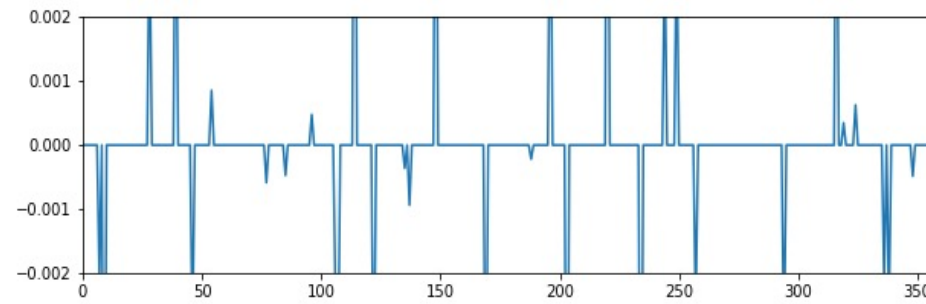
球面・軸集中現象



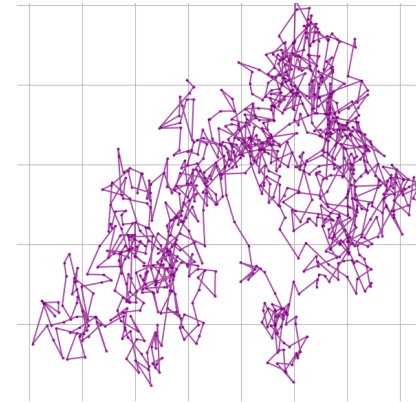
グラフ構造



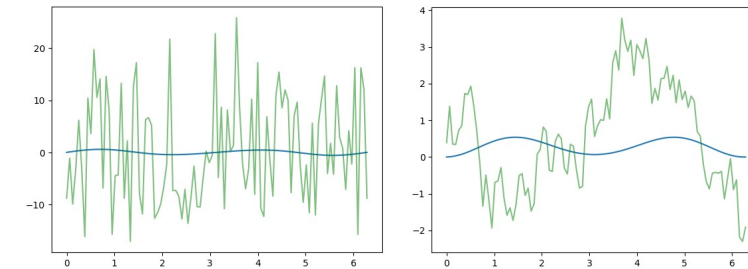
スパース性



ベイズの高次元挙動



弱い位相の利用



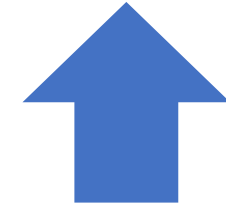
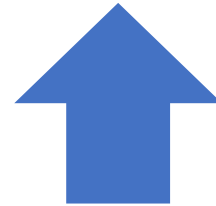
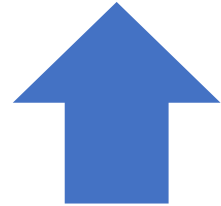
# まとめ

高次元データの解析技法を三つの観点で整理・紹介

可視化

発見・予測

不確実性評価



次元圧縮

正則化

ベイズ

