高次元の統計学[1]

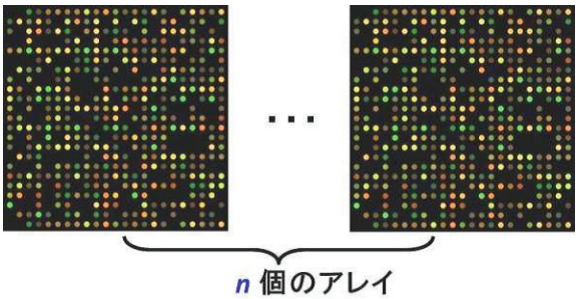
青嶋　誠
（筑波大学数理物質系・教授）

1　はじめに

ゲノム科学・情報工学・金融工学などの現代科学の一つの特徴は，データがもつ次元数の膨大さにあります．図 1 の DNA マイクロアレイのようなゲノムデータは，次元数が数万にものぼる一方で，標本数は 100 にも満たないという事例が多く見られます．これは，いわゆるビッグデータの一種で，データの次元数 d と標本数 n の大小関係が

$$d \gg n \text{（もしくは，} d > n\text{）}$$

となっていることが特徴です．



(図は https://upload.wikimedia.org/wikipedia/commons/2/2a/DNA_microarray.svg から引用)

図 1: 遺伝子発現データ（マイクロアレイデータ）．一般に，次元数（遺伝子数）$d \approx 1,000 \sim 50,000$，標本数（被験者数）$n \approx 10 \sim 100$ の巨大なデータセットになる．

従来のデータセットでは，データの次元数 d と標本数 n の大小関係は “$d < n$” が大前提です．例えば，Hand らが 1994 年に出版した有名なデータセット集 [9] を見てみましょう．ここには，当時の最先端の統計学で用いられたデータセットが，500 種類以上も掲載されています．殆どが高々10 次元で，まさしく “$d < n$” の大前提を満たしています．ところが，1990 年代後半，情報化の進展に伴って，“$d \gg n$” といった高次元データが突如出現しました．例えば，Harvard Medical School の Golub 教授らが 1999 年に Science に発表した論文 [8] には白血病患者の遺伝子発現データが与えられ，次元数は $d = 7129$，標本数は $n = 72$ でした．当時の統計学は，“$d < n$” なる条件が理論の拠り所となっていましたので，

[1]本稿は、2016 年 3 月 19 日に行った日本数学会市民講演会における講演の内容を，若干修正したものである．

高次元データの統計的推測に精度を保証することが出来ませんでした．2000 年代になり，高次元データの研究が徐々に進み，既存の統計学（多変量統計解析など）の限界が理論的に示され，新たな統計学の必要性が認識されました．2010 年代に入って，理論と応用の両面から統計学が飛躍的に向上し，新たな統計学として高次元統計解析が誕生しました．これが，2012 年以降のビッグデータのブームに繋がることになります．

本稿は，次元数 d と比べ圧倒的に少ない標本数 n でビッグデータに立ち向かう，高次元統計解析の一端をお見せしたいと思います．高次元の統計学には，従来の統計学の枠組みを超えた新しい発想が必要になることをご覧に入れます．

2　高次元データにおけるノイズの解析

一般に，次元数の増加に伴って情報が増加し，高次元データは非常に豊富な情報を有します．しかしながら，データに含まれるノイズも増加するので，高次元データは巨大なノイズを含むことになります．図 2 は，そのイメージ図です．高次元データは豊富な情報を

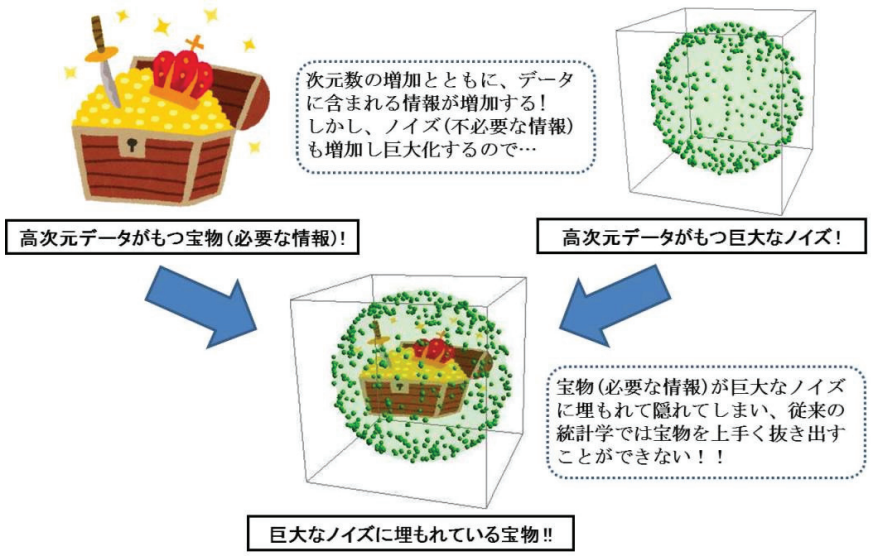


図 2: 高次元データのイメージ図

有するものの，それが巨大なノイズに埋もれています．残念ながら，多変量統計解析では巨大なノイズに太刀打ちできず，高次元データの解析に間違った結果を導くことさえあります．高次元データは，上手に扱わないとノイズしか聞こえてきません．高次元データがもつ豊富な情報を抽出するためには，従来の統計学の枠組みを超えた新しい理論と方法論が必要になるのです．高次元統計解析の第一歩は，巨大なノイズを解析することです．

図 2 のように，宝物（必要な情報）は巨大なノイズに埋もれています．巨大なノイズに比べると宝物はずっと小さいのですが，高次元データを闇雲にスパースだと思って扱うと，ノイズを拾ったり宝物を壊したりしてしまいます．高次元データの扱いには，巨大なノイズの解析が大事になるのです．その鍵は，次元数が標本数を遥かに超えた高次元空間に現れるデータの幾何学的特徴です．青嶋と矢田の一連の研究では，高次元データの様々な幾何学的特徴を見い出し，それらの幾何学的表現に基づいた統計的推測法を考案して，高次元統計解析の基礎理論を築きました．詳細は，参考文献 [1, 2, 3, 11] をご覧下さい．

簡単な例として，標本平均の高次元空間における幾何学的特徴を紹介します．平均ベクトル $\boldsymbol{\mu}$ と共分散行列 $\boldsymbol{\Sigma}$ をもつ d 次元分布から，無作為標本 $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ を抽出したとします．標本平均 $\overline{\boldsymbol{x}}_n = \sum_{j=1}^{n} \boldsymbol{x}_j/n$ について，$d/n \to 0$ (つまり，低次元大標本）の枠組みでは，よく知られているように

$$||\overline{\boldsymbol{x}}_n - \boldsymbol{\mu}|| \xrightarrow{p} 0$$

といった一致性[2]が成り立ちます．しかしながら，$d/n \to \infty$（つまり，高次元小標本）の枠組みでは，適当な条件のもとで

$$||\overline{\boldsymbol{x}}_n - \boldsymbol{\mu}|| \to \infty \text{ (確率的に)}$$

といった強不一致性[3]が現れます．これは，次元数の増加とともに，標本平均に含まれるノイズが巨大化することを意味します．ノイズが微小であることを前提とした従来の統計学では，もはや太刀打ちできないのです．

高次元空間における巨大なノイズの漸近的挙動を調べてみましょう．図 3 は，共分散行列に単位行列 $\boldsymbol{I}_d$ をもつ d 次元正規分布 $N_d(\boldsymbol{\mu}, \boldsymbol{I}_d)$ について，大きさ $n = 3$ の無作為標本による標本平均を 200 回発生させ，$\overline{\boldsymbol{x}}_n - \boldsymbol{\mu}$ を潜在空間 (固有空間) 上にプロットしたものです．次元数が $d = 4$ のときは，半径 $\sqrt{d/n} = \sqrt{4/3}$ の球の周辺に点が散らばっています．次元数が $d = 1000$ のときは，ノイズが巨大化し，半径 $\sqrt{d/n} = \sqrt{1000/3}$ の大きな球の表面に点が集中しています．この規則性を球面集中現象といいます．球面集中現象は，高次元になるとノイズ (不必要な情報) が巨大化し，データの特徴が薄れ，宝物 (必要な情報) が隠れてしまうことを意味します．しかし，高次元空間で眺めるとノイズにこういったパターンが現れてくるので，巨大なノイズの解析は高次元データを解析する上で大事になるのです．巨大なノイズを上手に扱うことができれば，宝物 (必要な情報) の本質的な特徴を捉えることができます．宝物へのアプローチは，次の 2 通りがあります．

---

[2]任意の $\varepsilon > 0$ で，$\lim_{n/d \to \infty} P(||\overline{\boldsymbol{x}}_n - \boldsymbol{\mu}|| > \varepsilon) = 0$ を意味する．ただし，$||\cdot||$ はユークリッドノルム．

[3]任意の $\varepsilon > 0$ で，$\lim_{d/n \to \infty} P(||\overline{\boldsymbol{x}}_n - \boldsymbol{\mu}|| < \varepsilon) = 0$ を意味する．

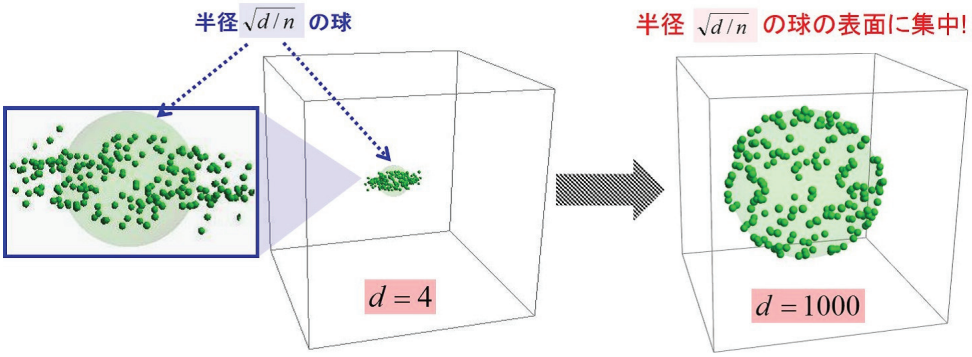


図 3: 高次元小標本の球面集中現象

1. 巨大なノイズを取り除く（応用例：高次元主成分分析・クラスター分析など）

2. 巨大なノイズを有効活用する（応用例：高次元 2 標本検定・判別分析など）

それでは，3 節で高次元主成分分析とクラスター分析について，4 節で高次元 2 標本検定と判別分析について，お話をすることにします．

3　高次元主成分分析・クラスター分析

本節では，高次元主成分分析を紹介し，その応用としてクラスター分析と実際のデータ解析を解説します．主成分分析は，多次元データがもつ情報をなるべく損失しないように次元圧縮するための手法です．例えば，図 4 のようにデータを 2 次元平面へ射影すれば，分布の特徴を簡潔に捉えることができます．

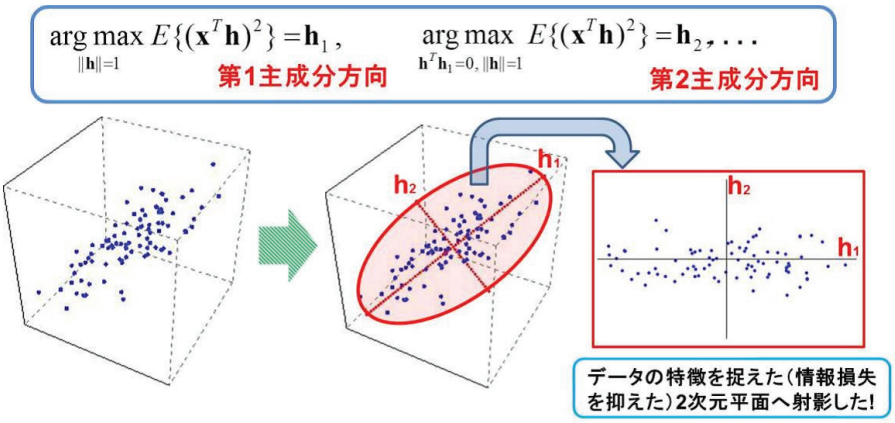


図 4: 主成分分析の例．3 次元データを 2 次元に圧縮．

それでは，高次元データに対して主成分分析を考察してみましょう．高次元データは非常に豊富な情報を有するので，上手くノイズを除去してデータの情報損失を正しく評価することができれば，必要な情報を低次元空間に縮約することができます．図 5 のように，高次元データの潜在空間 (必要な情報が詰まった空間) は巨大なノイズ空間に埋もれています．しかし，ノイズ空間が球形なので，主成分の方向を曲げることなくノイズ空間から潜在空間を抜き出すことができます．ただし，巨大なノイズの影響で固有値は過剰に見積もられるので，このままでは情報損失を正しく評価することができません．ここでは詳細を省きますが，高次元主成分分析には，クロスデータ行列法 (Yata and Aoshima [10]) やノイズ掃き出し法 (Yata and Aoshima [11]) といった巨大なノイズを取り除く手法が必要になります．

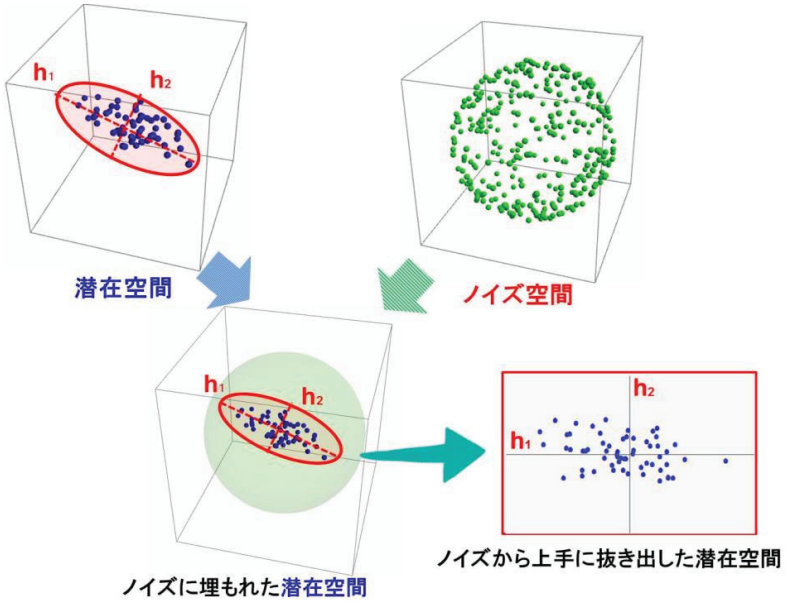


図 5: 高次元主成分分析

高次元主成分分析を使ったデータ解析をお見せしましょう．Bhattacharjee et al. [7] にある肺がんの遺伝子発現データを解析します．データセットは，次元数が $d = 3312$，標本数が $n = 58$ で，次の 3 つのタイプから構成されています．Π_1 : 肺カルチノイド（$n_1 = 20$ サンプル），Π_2 : 扁平上皮癌（$n_2 = 21$ サンプル），Π_3 : 正常肺（$n_3 = 17$ サンプル）．計 58 サンプルの 3312 次元データを高次元主成分分析で 2 次元に縮約したものが，図 6 の左下の図です．これを，タイプ別に色分けしたものが右下の図になります．

図 6 で見たように，高次元主成分分析を使うと巨大なノイズから潜在空間が抜き出され，高次元データの本質的な特徴が可視化されます．これは偶然ではなく，理論的に証明

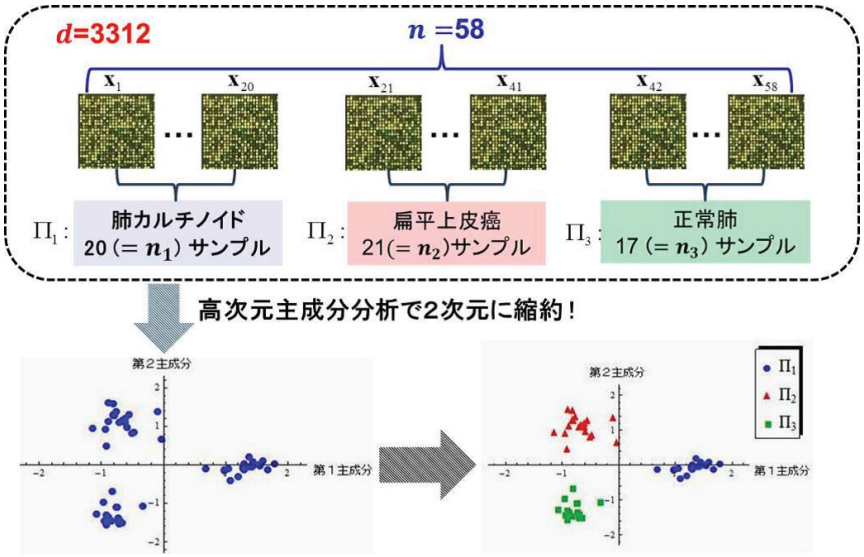


図 6: 肺がんの遺伝子発現データ ([7]) を高次元主成分分析で 2 次元に縮約.

できるのです．Yata and Aoshima [12] は，高次元主成分分析を使った高次元混合データのクラスタリングを考え，次の定理[4]を与えました．

定理 1 (高次元混合データのクラスタリング)．データが，3 個の母集団 Π_1，Π_2，Π_3 から発生するものと仮定する．第 j データの規準化した第 i 主成分を $\hat{z}_{ij}$ とおく．適当な正則条件のもとで，$d \to \infty$ のとき，各 j で次が成り立つ．

$$\operatorname*{plim}_{d\to\infty} \hat{z}_{1j} = \begin{cases} \sqrt{\dfrac{n-n_1}{n_1}} & (\boldsymbol{x}_j \in \Pi_1), \\ -\sqrt{\dfrac{n_1}{n-n_1}} & (\boldsymbol{x}_j \notin \Pi_1), \end{cases} \qquad \operatorname*{plim}_{d\to\infty} \hat{z}_{2j} = \begin{cases} 0 & (\boldsymbol{x}_j \in \Pi_1), \\ \sqrt{\dfrac{n_3}{n_2(1-n_1/n)}} & (\boldsymbol{x}_j \in \Pi_2), \\ -\sqrt{\dfrac{n_2}{n_3(1-n_1/n)}} & (\boldsymbol{x}_j \in \Pi_3). \end{cases}$$

定理 1 は，高次元混合データのクラスター構造が高次元主成分分析を使って可視化されることを，理論的に保証しています．高次元混合データのクラスターは，図 7 の左の図のような三角形の各頂点に構成される，と主張しています．先の高次元主成分分析の結果と合わせると図 7 の右の図のようになり，定理 1 が主張する通り，3 つのタイプの高次元データが各々クラスターを構成し，三角形の各頂点に集まっている様子が確認できます．このように，巨大なノイズを取り除いて潜在空間を抜き出せば，高次元データの高精度なクラスタリングが可能となるのです．

[4]Yata and Aoshima [12] の定理は，一般の k 個の母集団に対するものである．

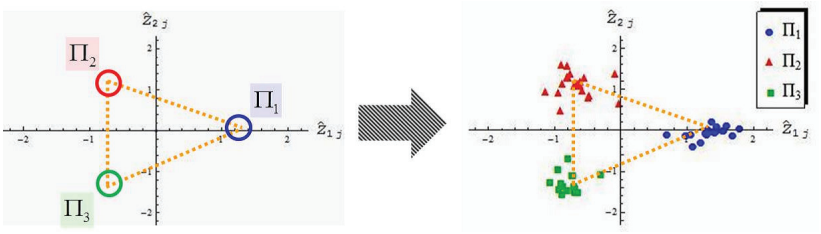


図 7: 肺がんの遺伝子発現データ ([7]) のクラスタリング

4　高次元 2 標本検定・判別分析

本節では，高次元 2 標本検定と高次元判別分析を紹介し，巨大なノイズを有効活用するアプローチを解説します．図 3 を，もう一度眺めてみましょう．高次元小標本において，$\overline{\boldsymbol{x}}_n - \boldsymbol{\mu}$ に球面集中現象が見られました．Aoshima and Yata [1, 5] は，より詳細に球面上の漸近的挙動について調べ，次の定理を証明しました．

定理 2 (球面上の中心極限定理)．適当な正則条件のもと，$d \to \infty$ のとき，次が成り立つ．

$$\frac{||\overline{\boldsymbol{x}}_n - \boldsymbol{\mu}||^2 - \mathrm{tr}(\boldsymbol{\Sigma})/n}{\sqrt{\mathrm{Var}(||\overline{\boldsymbol{x}}_n - \boldsymbol{\mu}||^2)}} \Rightarrow N(0,1).$$

定理 2 は，中心 $\boldsymbol{\mu}$，半径 $\sqrt{\mathrm{tr}(\boldsymbol{\Sigma})/n}$ の球面における $\overline{\boldsymbol{x}}_n$ の挙動が，高次元のとき正規分布に従うことを主張しています．図 8 は，d 次元正規分布 $N_d(\boldsymbol{\mu}, \boldsymbol{I}_d)$ について，大きさ $n = 3$ の無作為標本による標本平均をもとに $U_d = n(||\overline{\boldsymbol{x}}_n - \boldsymbol{\mu}||^2 - d/n)/\sqrt{2d}$ を計算し，これを 2000 回発生させてヒストグラムを作成したものです．この設定では，$\mathrm{Var}(||\overline{\boldsymbol{x}}_n - \boldsymbol{\mu}||^2) = 2d/n^2$ となります．次元数 d が大きいと，ヒストグラムの形状は $N(0,1)$ に近づくことが確認できます．つまり，球面上の微小な変動まで特定され，高次元データの巨大なノイズの法則が理論的に解明できるのです．

高次元 2 標本検定は，高次元データの巨大なノイズの法則を有効活用します．2 つの母集団 Π_i，$i = 1, 2$ を考え，d 次元の平均ベクトル $\boldsymbol{\mu}_i$ と共分散行列 $\boldsymbol{\Sigma}_i$ をもつとして，次の 2 標本検定を考えます．

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \text{vs.} \quad H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

各母集団 Π_i から，n_i 個のデータ $\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{in_i}$ を抽出し，標本平均 $\overline{\boldsymbol{x}}_{in_i} = \sum_{j=1}^{n_i} \boldsymbol{x}_{ij}/n_i$ を計算します．Aoshima and Yata [1, 5] は，2 標本について，次の高次元中心極限定理を与えました．

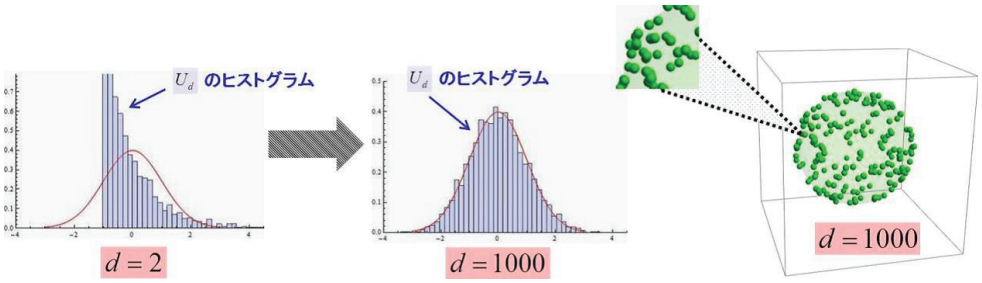


図 8: 球面上の $\overline{\boldsymbol{x}}_n$ の漸近的挙動．次元数が $d=2$ と $d=1000$ の場合について，$U_d = n(||\overline{\boldsymbol{x}}_n - \boldsymbol{\mu}||^2 - d/n)/\sqrt{2d}$ を 2000 回発生させ，ヒストグラムにした．実線は，$N(0,1)$ の確率密度関数を表す．

定理 3 (高次元 2 標本検定)．適当な正則条件のもとで，$d \to \infty$ のとき，次が成り立つ．

$$\frac{||\overline{\boldsymbol{x}}_{1n_1} - \overline{\boldsymbol{x}}_{2n_2}||^2 - \mathrm{tr}(\boldsymbol{\Sigma}_1)/n_1 - \mathrm{tr}(\boldsymbol{\Sigma}_2)/n_2 - ||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2||^2}{\sqrt{\mathrm{Var}(||\overline{\boldsymbol{x}}_{1n_1} - \overline{\boldsymbol{x}}_{2n_2} - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)||^2)}} \Rightarrow N(0,1).$$

定理 3 において，H_0 のもと $||\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2||^2 = 0$ であることに注意し，分母子のパラメータを推定すれば，高次元 2 標本検定法を構築することができます．遺伝子群検定等への応用が考えられます．

次に，高次元データの判別分析を紹介します．Π_1 か Π_2 かの判別の対象となる個体について，d 次元データを $\boldsymbol{x}_0$ とします．判別分析は，各母集団の学習データ $\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{in_i}$ $(i=1,2)$ から判別ルールを構築し，これに従って個体（$\boldsymbol{x}_0$）を Π_1 か Π_2 のどちらかに振り分けるための手法です．疾患の種類の判別や顔認証など，多くの分野で必要とされる分析です．多変量統計解析では Fisher の線形判別や 2 次判別が知られていますが，高次元小標本（$d > n_i$）においては標本共分散行列 $\boldsymbol{S}_{in_i}$ の逆行列が存在しないので，それらの判別ルールは使えません．高次元データを扱う判別分析法が色々とありますが，巨大なノイズを適切に処理しているとは言い難く，大半は精度が保証されません．Aoshima and Yata [1, 4] は，高次元データの球面集中現象に着目して，巨大なノイズの構造を利用した高次元ならではの判別分析法を考案しました．図 3 と同様に，$\boldsymbol{x}_0 \in \Pi_i$ のとき $\boldsymbol{x}_0 - \boldsymbol{\mu}_i$ は高次元で球面集中現象が見られ，球の半径は $\sqrt{\mathrm{tr}(\boldsymbol{\Sigma}_i)}$ になります．図 9 で示したように，巨大なノイズ空間の半径の違いを利用することで，高次元ならではの判別分析法が考えられます．この着想から，Aoshima and Yata [1] は，幾何学的判別法とよばれる次のような判別ルールを導き，それが高次元で完全分類を与えることを証明しました．

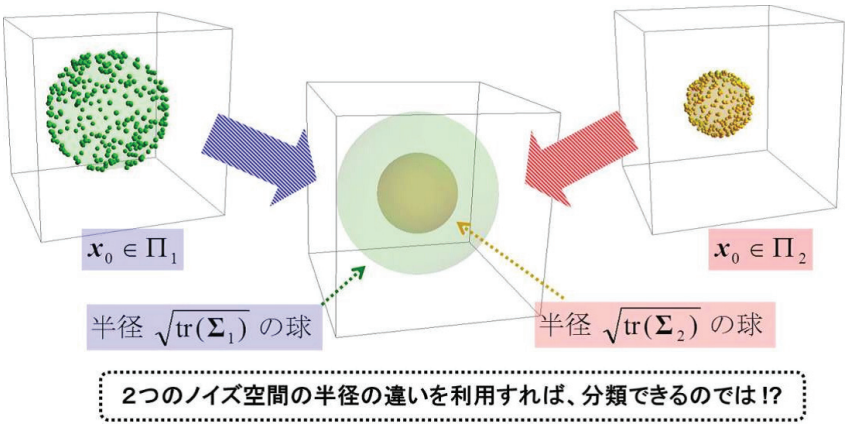


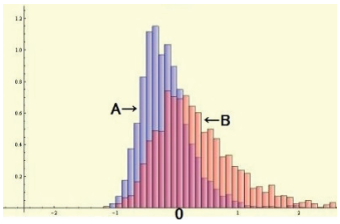
図 9: 高次元データの 2 群判別

定理 4 (幾何学的判別法). Π_1, Π_2 の学習データから $\overline{\boldsymbol{x}}_{1n_1}$, $\boldsymbol{S}_{1n_1}$, $\overline{\boldsymbol{x}}_{2n_2}$, $\boldsymbol{S}_{2n_2}$ を計算し，$\boldsymbol{x}_0$ に対して

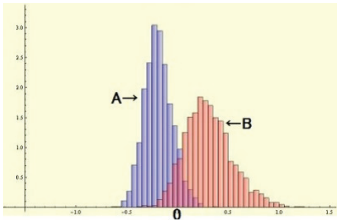
$$\omega(\boldsymbol{x}_0) = \frac{d||\boldsymbol{x}_0 - \overline{\boldsymbol{x}}_{1n_1}||^2}{\mathrm{tr}(\boldsymbol{S}_{1n_1})} - \frac{d||\boldsymbol{x}_0 - \overline{\boldsymbol{x}}_{2n_2}||^2}{\mathrm{tr}(\boldsymbol{S}_{2n_2})} - d\log\left\{\frac{\mathrm{tr}(\boldsymbol{S}_{2n_2})}{\mathrm{tr}(\boldsymbol{S}_{1n_1})}\right\} - \frac{d}{n_1} + \frac{d}{n_2}$$

とおく．判別ルールを，$\omega(\boldsymbol{x}_0) < 0$ のとき $\boldsymbol{x}_0 \in \Pi_1$，$\omega(\boldsymbol{x}_0) \geq 0$ のとき $\boldsymbol{x}_0 \in \Pi_2$ とする．そのとき，適当な条件のもとで，誤判別確率は $d \to \infty$ でゼロに収束する．

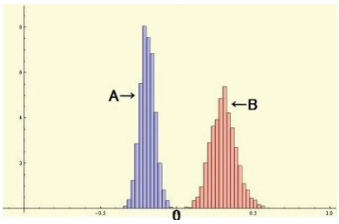
幾何学的判別法は，仮に Π_1 と Π_2 の平均に差がない状況であっても，2 つのノイズ空間の半径の違いから高精度に高次元データを分類することができます．簡単なシミュレーション実験をお見せしましょう．$\Pi_1 : N_d(\boldsymbol{0}, \boldsymbol{I}_d)$, $\Pi_2 : N_d(\boldsymbol{0}, 2\boldsymbol{I}_d)$ とし，各母集団から $n_1 = n_2 = 5$ 個の学習データを使って $\omega(\boldsymbol{x}_0)$ を計算します．次元数は $d = 8, 64, 512, 4096$ の 4 つの場合を考えます．Π_1 と Π_2 には，平均に差がありませんが，ノイズ空間の半径はそれぞれ $\sqrt{\mathrm{tr}(\boldsymbol{I}_d)} = \sqrt{d}$, $\sqrt{\mathrm{tr}(2\boldsymbol{I}_d)} = \sqrt{2d}$ になります．図 10 は，A: $\boldsymbol{x}_0 \in \Pi_1$ のときと，B: $\boldsymbol{x}_0 \in \Pi_2$ のときについて，それぞれ 2000 回のシミュレーションを行って $\omega(\boldsymbol{x}_0)/d$ のヒストグラムを作成しています．次元数が上がるにつれて，A のヒストグラムは負の値に，B のヒストグラムは正の値に，それぞれ収束していく様子が見てとれます．定理 4 の判別ルールと照らし合わせると，幾何学的判別法の誤判別確率はゼロに収束することが分かり，完全分類が確認できます．幾何学的判別法は，たとえ平均が等しくても，高次元におけるノイズ空間の半径の違いを利用して，Π_1 と Π_2 の差異を際立たせています．このように，高次元データの巨大なノイズを有効活用することで，高精度な高次元判別分析が可能となるのです．高次元判別関数のクラスを考慮した詳細な理論については，Aoshima and Yata [6] を



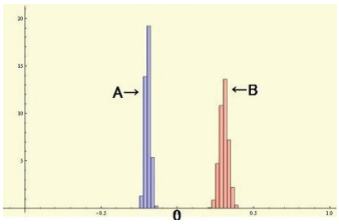
(a) $d = 8$



(b) $d = 64$



(c) $d = 512$



(d) $d = 4096$

図 10: $\omega(\boldsymbol{x}_0)/d$ のヒストグラム．A: $\boldsymbol{x}_0 \in \Pi_1(N_d(\boldsymbol{0}, \boldsymbol{I}_d))$ のとき，B: $\boldsymbol{x}_0 \in \Pi_2(N_d(\boldsymbol{0}, 2\boldsymbol{I}_d))$ のとき．

参照して下さい．

5　まとめ

高次元データは，巨大なノイズに宝物（必要な情報）が埋もれています．巨大なノイズに比べると宝物は小さいですが，最初からスパースだと思って扱うと宝物を壊してしまいます．宝物を壊すことなく本質的な特徴を抽出するためには，巨大なノイズの解析をすることが大事になります．高次元空間で眺めれば，ノイズにパターンが現れてきます．その規則性を捉えられれば，巨大なノイズを取り除くこともできますし，有効活用することもできます．従来の統計学の枠組みを超えた，新しい発想が広がります．こういった巨大なノイズの解析に基づいた新しい統計学が，高次元統計解析です．本稿は，高次元統計解析の一端を紹介しました．詳細や関連文献については，解説論文 [2, 3] や，拙 HP(http://www.math.tsukuba.ac.jp/∼aoshima-lab/jp/papers.html) をご覧下さい．

参考文献

[1] Aoshima, M. and Yata, K. (2011). Two-stage procedures for high-dimensional data. *Sequential Anal. (Editor's special invited paper)* **30**, 356-399.

[2] 青嶋　誠，矢田和善 (2013a). 論説：高次元小標本における統計的推測．数学，**65**, 225-247.

[3] 青嶋　誠，矢田和善 (2013b). 日本統計学会研究業績賞受賞者特別寄稿論文 : 高次元データの統計的方法論. 日本統計学会誌, **43**, 123-150.

[4] Aoshima, M. and Yata, K. (2014). A distance-based, misclassification rate adjusted classifier for multiclass, high-dimensional data. *Ann. Inst. Statist. Math.* **66,** 983-1010.

[5] Aoshima, M. and Yata, K. (2015a). Asymptotic normality for inference on multisample, high-dimensional mean vectors under mild conditions. *Methodol. Comput. Appl. Probab.* **17**, 419-439.

[6] Aoshima, M. and Yata, K. (2015b). High-dimensional quadratic classifiers in non-sparse settings. arXiv:1503.04549.

[7] Bhattacharjee, A. et al. (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. USA* **98**, 13790-13795.

[8] Golub, T. R. et al. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531-537.

[9] Hand, D. J. et al. (1994). *A Handbook of Small Data Sets.* Chapman and Hall, London.

[10] Yata, K. and Aoshima, M. (2010). Effective PCA for high-dimension, low-sample-size data with singular value decomposition of cross data matrix. *J. Multivariate Anal.*, **101**, 2060–2077.

[11] Yata, K. and Aoshima, M. (2012). Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. *J. Multivariate Anal.*, **105**, 193–215.

[12] Yata, K. and Aoshima, M. (2015). Principal component analysis based clustering for high-dimension, low-sample-size data. arXiv:1503.04525.

あおしま　まこと（筑波大学数理物質系）
http://www.math.tsukuba.ac.jp/∼aoshima-lab/jp/