数学カフェ「機械学習の数理」

関係データ解析へのランダム行列理論の応用

渡邊千紘

NTT コミュニケーション科学基礎研究所
東京大学大学院 数理情報学専攻 第 6 研究室

※本日の内容は東京大学 鈴木大慈准教授との共著になります

数学カフェ HP：https://mathcafe.net/

目次

目次

自己紹介

- 渡邊千紘（Chihiro Watanabe）
- 現在の所属
  - NTT コミュニケーション科学基礎研究所 研究員@厚木
  - 社会人博士課程@東京大学大学院 情報理工学系研究科 数理情報学専攻
- 経歴
  - 東京大学 工学部 計数工学科 システム情報工学コース（数理で卒論）
  - 東京大学大学院 情報理工学系研究科 システム情報学専攻
- 興味：関係データ解析，解釈可能性などなど
- twitter: @chihiro_ribbon

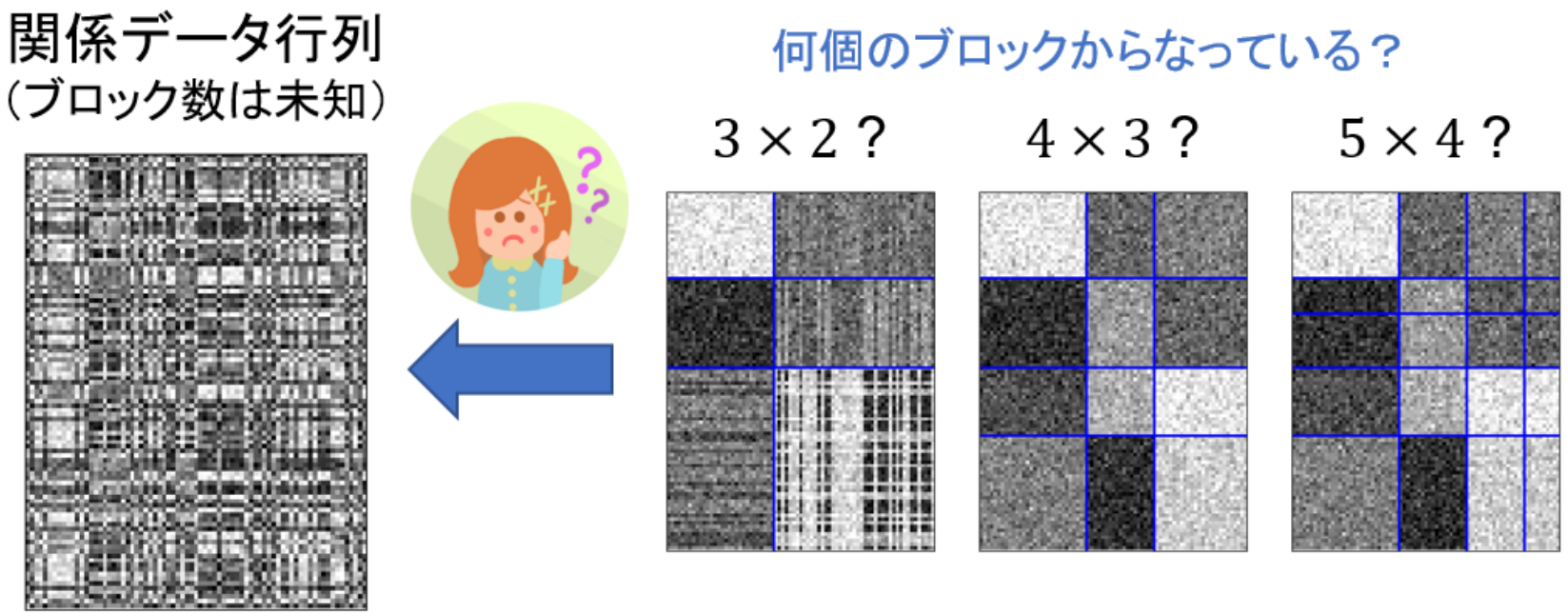
本発表の内容についての詳細情報

本発表の内容（図表含む）は以下の論文に基づいています.

- Chihiro Watanabe, Taiji Suzuki. “Goodness-of-fit Test for Latent Block Models,” Computational Statistics & Data Analysis, Vol. 154 (2021), pp. 107090. https://doi.org/10.1016/j.csda.2020.107090.
- arXiv:1906.03886, 2019. https://arxiv.org/abs/1906.03886.

本研究の概要

- 与えられた関係データ行列を表現するのに適切なクラスタ数を，統計的検定に基づいて決める手法を提案



目次

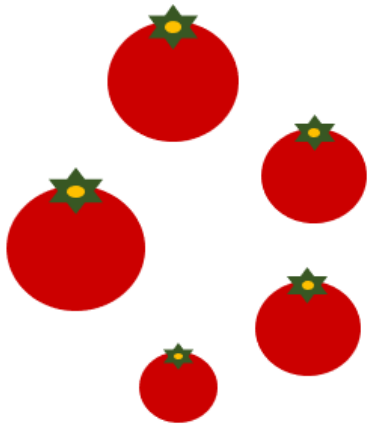
準備 (1) 統計的検定（仮説検定）

- 何か検証したい仮説がある時に，実際に得られたデータに基づいて，その仮説が妥当そうかどうかを検討するためのもの（具体的な手順は次ページ）
- 例：去年と今年のトマト栽培で肥料を変更してみた．今年採れたトマトは去年よりも有意に大きいか？

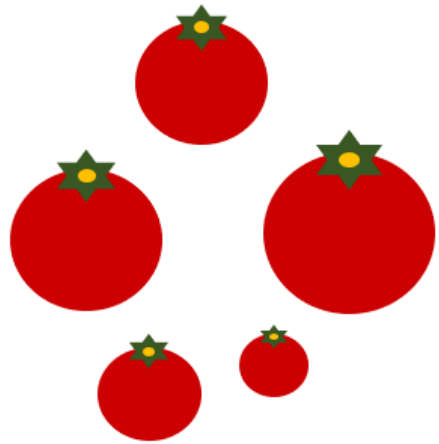
トマト栽培の例

- 実際に去年と今年のトマトの大きさを測ってみればどちらの平均値が大きいかは分かるが…
- たとえ去年と今年で同じ分布からトマトの大きさが決まっているとしても，得られたデータの平均値をとってみれば，たまたま今年の方が大きくなる可能性はある

去年のトマトのデータ

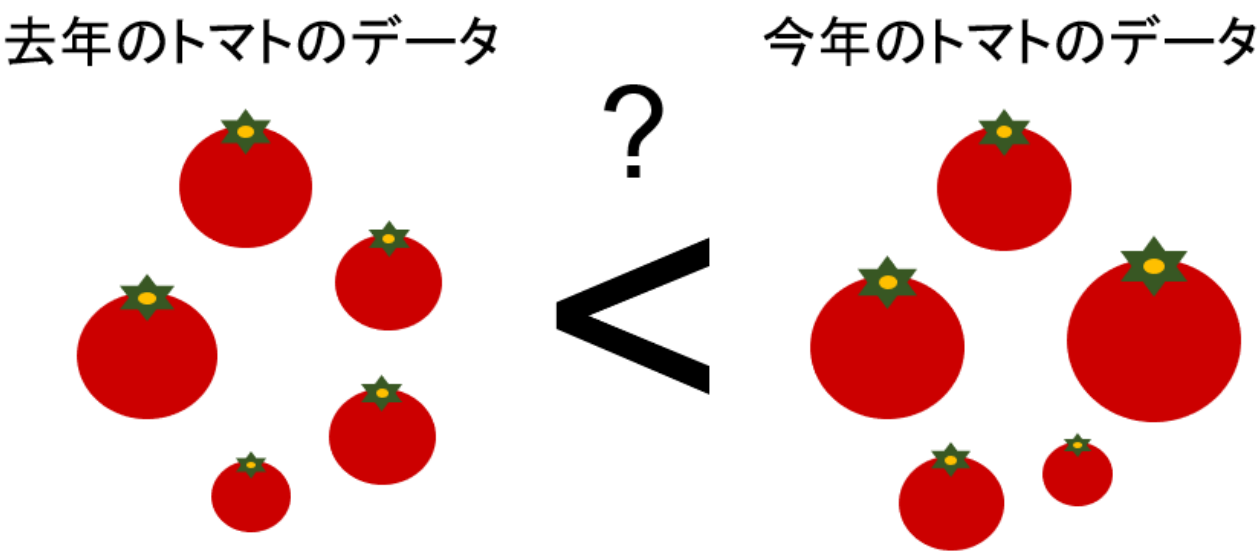


今年のトマトのデータ



トマト栽培の例

- 昨年と今年のトマトの大きさがそれぞれ平均 μ_1, μ_2 の分布から決まっていると仮定すると…
- 知りたいことは$\mu_1 = \mu_2$ なのか，それとも $\mu_1 < \mu_2$ なのか（たまたま得られたデータから計算される平均値の大小ではない）



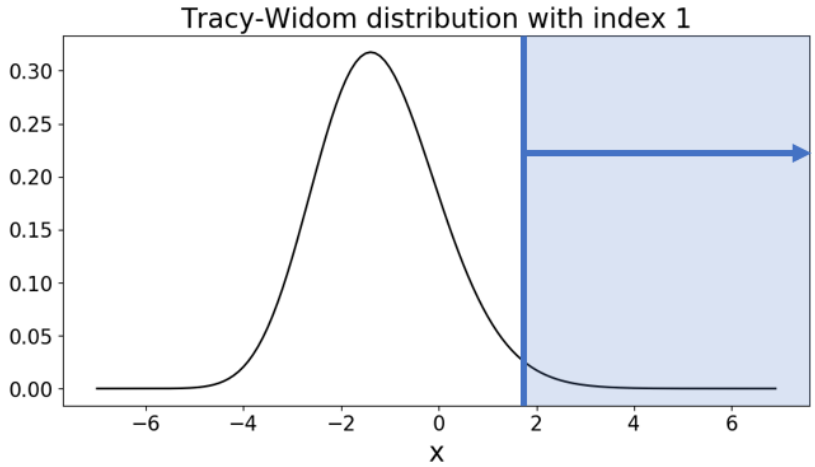
統計的検定の考え方

- 帰無仮説（$\mu_1 = \mu_2$, トマトの大きさは変わっていない）と対立仮説（$\mu_1 < \mu_2$, 今年のトマトの方が大きい）を定義
- ♣「帰無仮説が正しい場合に成り立つこと」を考える
- 実際のデータと照らし合わせて，♣ が成り立っているというのは無理があるかどうかを考える
- 無理がありそうなら帰無仮説を棄却（対立仮説を受容）
- 無理があると言えない場合は帰無仮説を受容（対立仮説を棄却）

※具体的な手順は次ページ

一般的な統計的検定の流れ

1. 帰無仮説・対立仮説と有意水準 α を設定
2. 検定に使う統計量（検定統計量）T を定義
3. 帰無仮説の下で T が従う分布（帰無分布）を導出
4. 実際に得られたデータから T の具体的な値を計算
5. 帰無分布と実際に計算された T を比べる
6. （例えば片側検定の場合）T が帰無分布の α-upper quantile に含まれる（＝極端な値を取っている）なら帰無仮説を棄却
7. そうでなければ帰無仮説を受容



α-upper quantile
（確率密度関数の
右側，面積がα）

統計的検定における注意事項

- 検定統計量 T は帰無分布が導出できるようなものならば何でも好きに決めてよい
  - ただし，同じ有意水準 α の設定において，検出力（対立仮説が正しいときに，帰無仮説を棄却できる確率）がより高いものが望ましい
- 有意水準 α の意味
  - 帰無仮説が正しい場合に，それにもかかわらず帰無仮説を棄却してしまう確率
  - 上記の失敗をどれくらい許せるかに応じて，ユーザが設定するもの（0.1, 0.05, 0.01 に設定することが多い）
  - 「帰無仮説が正しい/正しくない確率」そのものではない

準備 (2) 確率収束，法則収束

- 確率変数の列 $X_1, X_2, \ldots$ が確率変数 X に確率収束するとは，
  $\forall \epsilon > 0, \forall \delta > 0, \exists M \in \mathbb{N}, \forall m \geq M, \Pr(|X_m - X| > \epsilon) < \delta.$
- 確率変数の列 $X_1, X_2, \ldots$ が確率変数 X に法則収束（分布収束）するとは，任意の連続有界関数 $f(x)$ について，$m \to \infty$ で
  $\mathbb{E}[f(X_m)] \to \mathbb{E}[f(X)]$.
- 確率変数の列 $X_1, X_2, \ldots$ が確率変数 X に確率収束するならば，確率変数の列 $X_1, X_2, \ldots$ は確率変数 X に法則収束する．

準備 (3) 中心極限定理

- 確率変数 $X_1, \ldots, X_n$ が独立に同一の確率分布（期待値 μ, 標準偏差 σ）に従うとすると，$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma}$ は漸近的に標準正規分布 $\mathcal{N}(0, 1)$ に従う（法則収束）.

準備 (4) オーダ記法

$$x_m = O\left[f(m)\right]. \Leftrightarrow \ \exists C > 0, M > 0, \ \forall m \geq M, \ Cf(m) \geq |x_m|.$$
$$x_m = \Omega\left[f(m)\right]. \Leftrightarrow \ \exists C > 0, M > 0, \ \forall m \geq M, \ Cf(m) \leq |x_m|.$$
$$x_m = \Theta\left[f(m)\right]. \Leftrightarrow \ \exists C_1, C_2 > 0, M > 0, \ \forall m \geq M,$$
$$C_1 f(m) \leq |x_m| \leq C_2 f(m).$$

$$X_m = O_p\left[f(m)\right]. \Leftrightarrow \ \forall \epsilon > 0, \ \exists C > 0, M > 0, \ \forall m \geq M,$$
$$\Pr\left[Cf(m) \geq |X_m|\right] \geq 1 - \epsilon.$$
$$X_m = \Omega_p\left[f(m)\right]. \Leftrightarrow \ \forall \epsilon > 0, \ \exists C > 0, M > 0, \ \forall m \geq M,$$
$$\Pr\left[Cf(m) \leq |X_m|\right] \geq 1 - \epsilon.$$
$$X_m = \Theta_p\left[f(m)\right].$$
$$\Leftrightarrow \ \forall \epsilon > 0, \ \exists C_1, C_2 > 0, M > 0, \ \forall m \geq M,$$
$$\Pr\left[C_1 f(m) \leq |X_m| \leq C_2 f(m)\right] \geq 1 - \epsilon.$$

準備 (5) ノルムの定義

行列 $A = (A_{ij})_{1 \le i \le n, 1 \le j \le p} \in \mathbb{R}^{n \times p}$ の作用素ノルム $\|\cdot\|_{\mathrm{op}}$ とフロベニウスノルム $\|\cdot\|_{\mathrm{F}}$

$$\|A\|_{\mathrm{op}} = \sup_{\boldsymbol{u} \in \mathbb{R}^p} \frac{\|A\boldsymbol{u}\|}{\|\boldsymbol{u}\|}, \qquad \|A\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{p} A_{ij}^2}, \tag{1}$$

準備 (6) Tracy-Widom 分布 (index 1) I

- 累積密度関数 $F_1(x)$ が以下で定義される確率分布．本発表を通し，TW_1 分布とも表記する．

$$
\begin{aligned}
F_1(x) &= E(x)F(x), \\
E(x) &= \exp\left(-\frac{1}{2}\int_x^\infty q(y)\mathrm{d}y\right) \\
F(x) &= \exp\left(-\frac{1}{2}\int_x^\infty (y-x)q(y)^2\mathrm{d}y\right).
\end{aligned} \tag{2}
$$

ただし，$\mathrm{Ai}(x) = \frac{1}{\pi}\int_0^\infty \cos\left(\frac{t^3}{3}+xt\right)\mathrm{d}t$ として，$q(y)$ は境界条件 $q(x) \sim \mathrm{Ai}(x), x \to \infty$ を満たす Painlevé II-型方程式 $\frac{\mathrm{d}^2 q}{\mathrm{d}x^2} = xq + 2q^3$ の唯一の解．

- 確率密度関数は陽には求まらないため，近似式 [15] が提案されている

準備 (6) Tracy-Widom 分布 (index 1) II

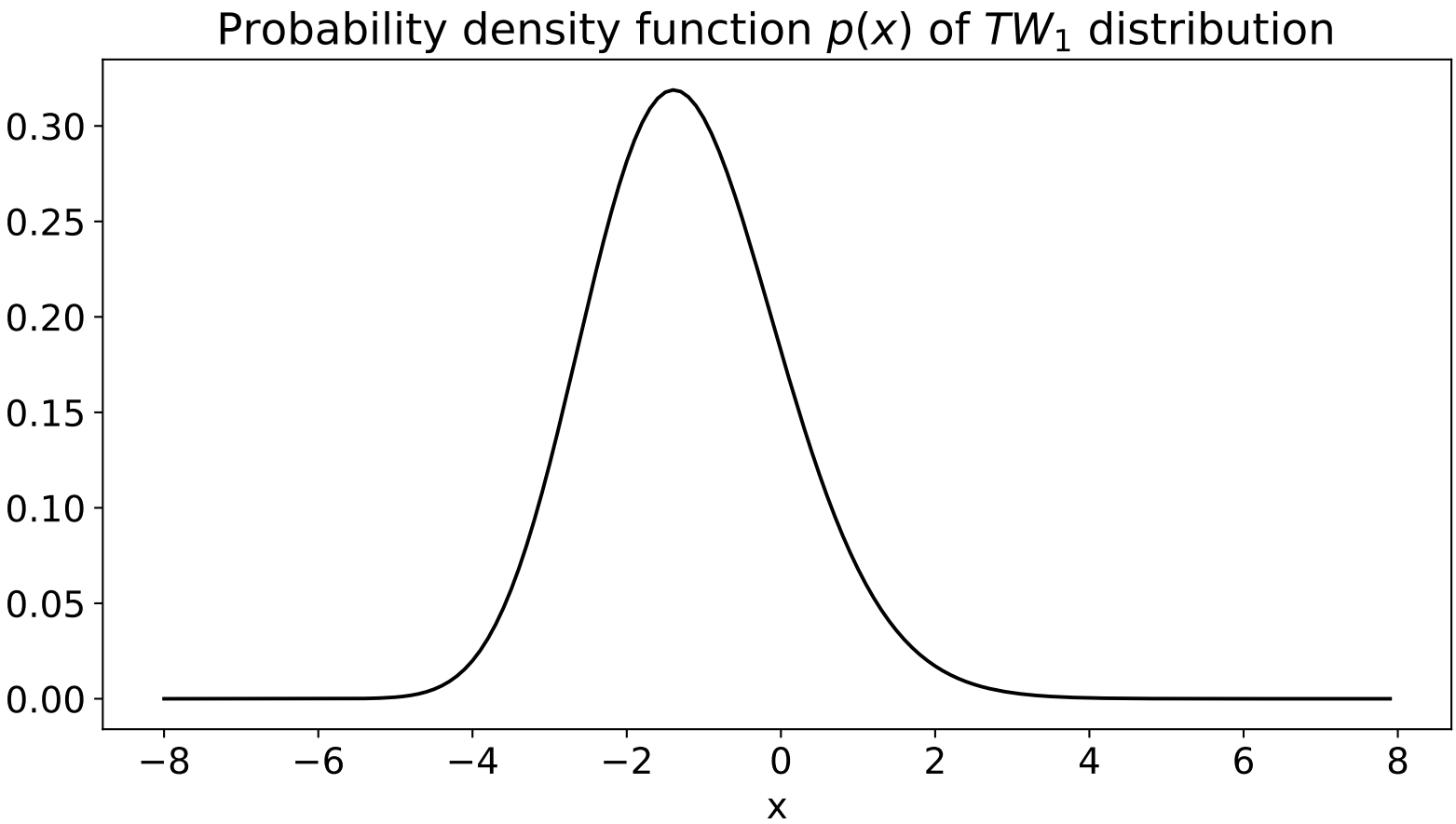
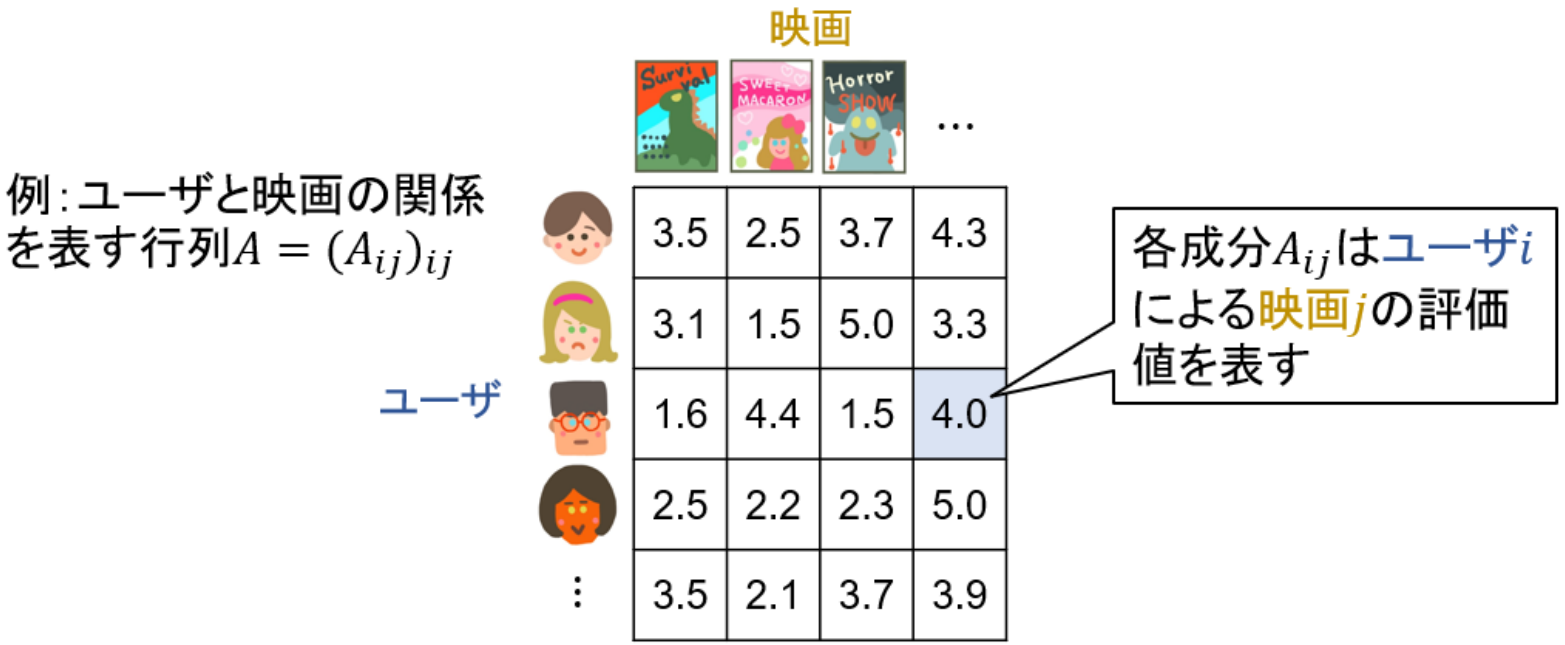


Figure: 近似式 [15] による TW_1 分布の確率密度関数のプロット.

目次

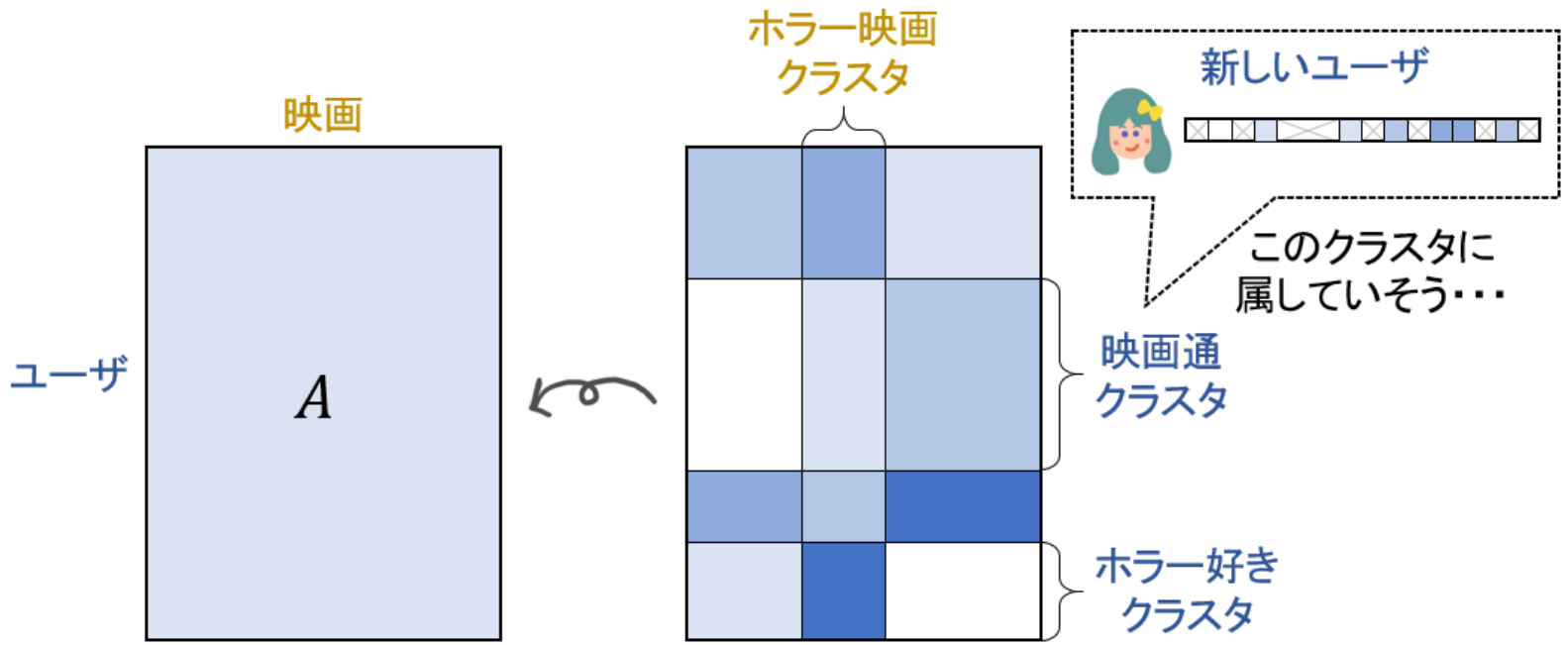
研究の背景：関係データの解析

- 関係データ：一般に異なる 2 つのオブジェクト間の関係を表す行列データ
- 我々の身の回りには，様々な関係データが存在する
  - ユーザ/映画の関係（評価），ユーザ/商品の関係（取引状況），文書/単語の関係など



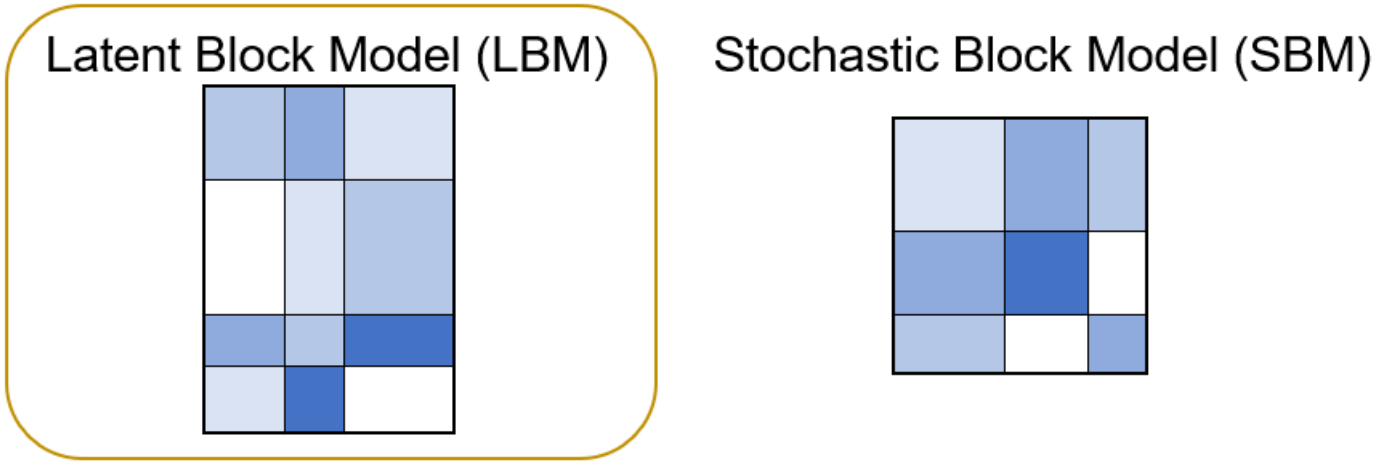
関係データ解析におけるブロックモデルの有用性

- 関係データの解析において，ブロックモデルが有効であることが知られている [6, 10, 13]
  - 行と列をそれぞれクラスタリング（共クラスタリングと呼ばれる）
  - 応用例：映画や商品の推薦システム



Latent Block Model (LBM)

- 関係データにおける共クラスタリングの構造を表現したブロックモデル
  - 仮定：行と列のクラスタ構造が与えられた条件のもとで，各成分のデータが独立にブロック内で同一の分布から生成される
  - 各成分が従う分布の仮定（e.g., 正規分布，Bernoulli 分布）を変えることで，多様な行列データの表現が可能
  - 対称な正方行列に対するブロックモデル（Stochastic Block Model, SBM）と区別して上記のように呼ぶ



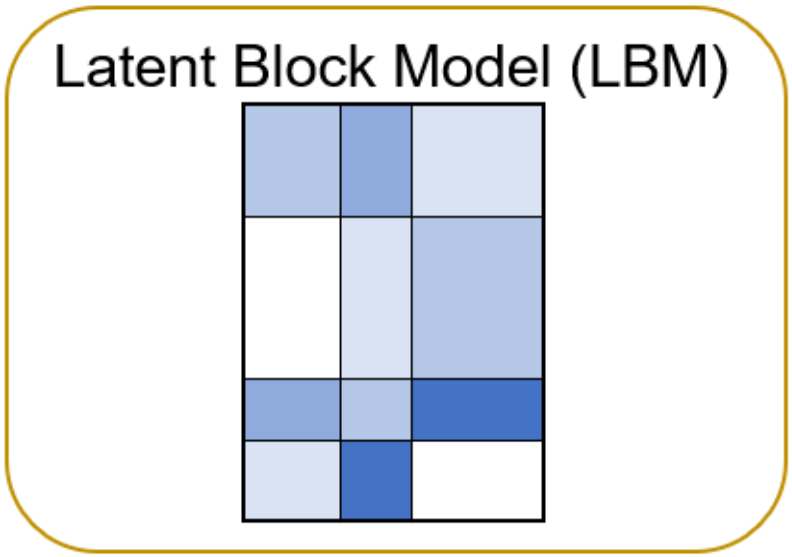
Latent Block Model を用いる際の課題

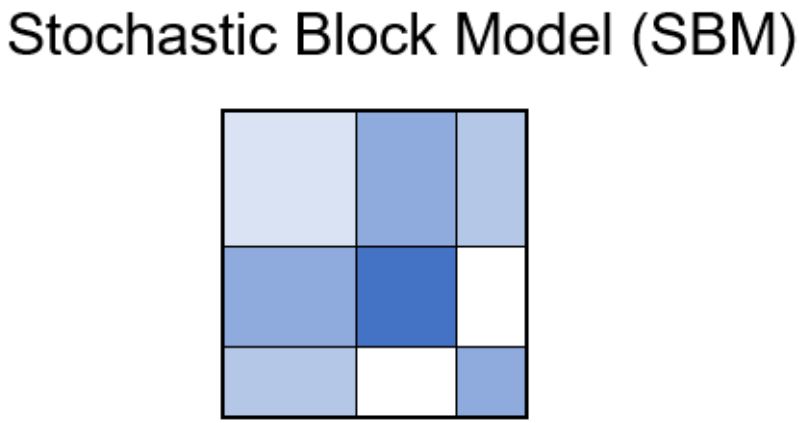
- 行・列のクラスタ数を事前に決めておく必要がある
  - 設定したクラスタ数に応じて異なるクラスタリング結果が得られる
- 一般にクラスタ数が事前に分かっているとは限らないため，行列データからクラスタ数を決める手法が必要
  - 統計的検定や情報量規準などによるアプローチが可能
  - 使いたい場面に応じて適切な手法を使い分ける必要あり

目次

関連研究

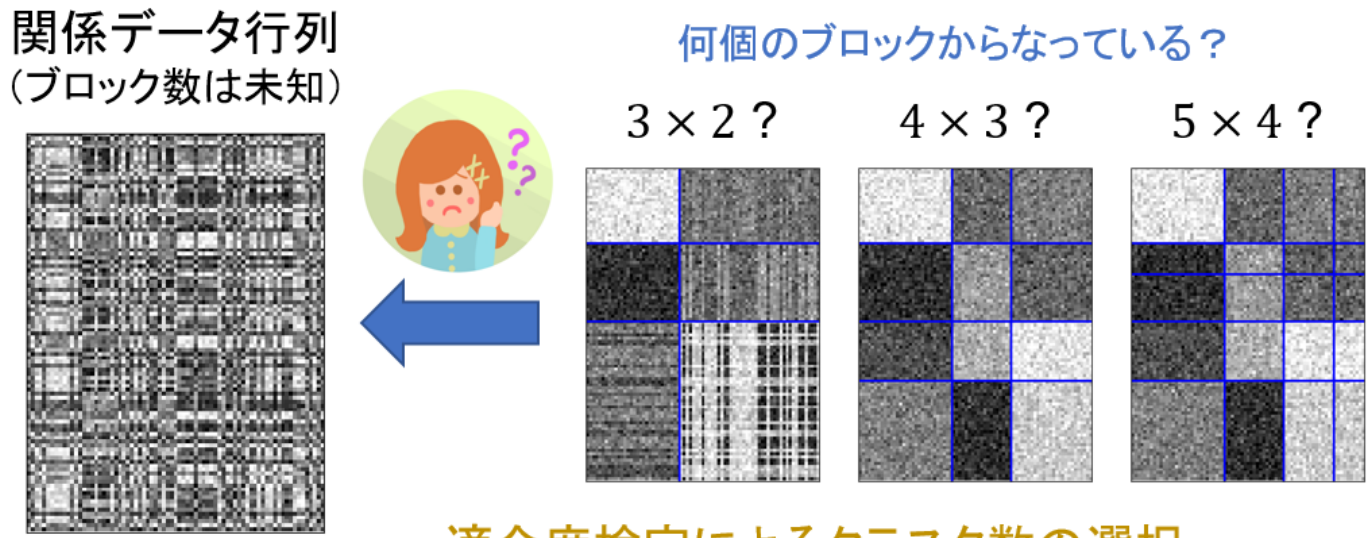
- ブロックモデルのクラスタ数に対する検定 [2, 9, 11]
  - クラスタ数に関する帰無仮説・対立仮説を立てて検定
  - 対称な正方行列に対するブロックモデル（Stochastic Block Model, SBM）に対する統計的検定しか提案されていなかった
  - 本研究では，LBM の設定（行と列が一般に異なるオブジェクトに対応）で使える検定を初めて提案





本研究による貢献

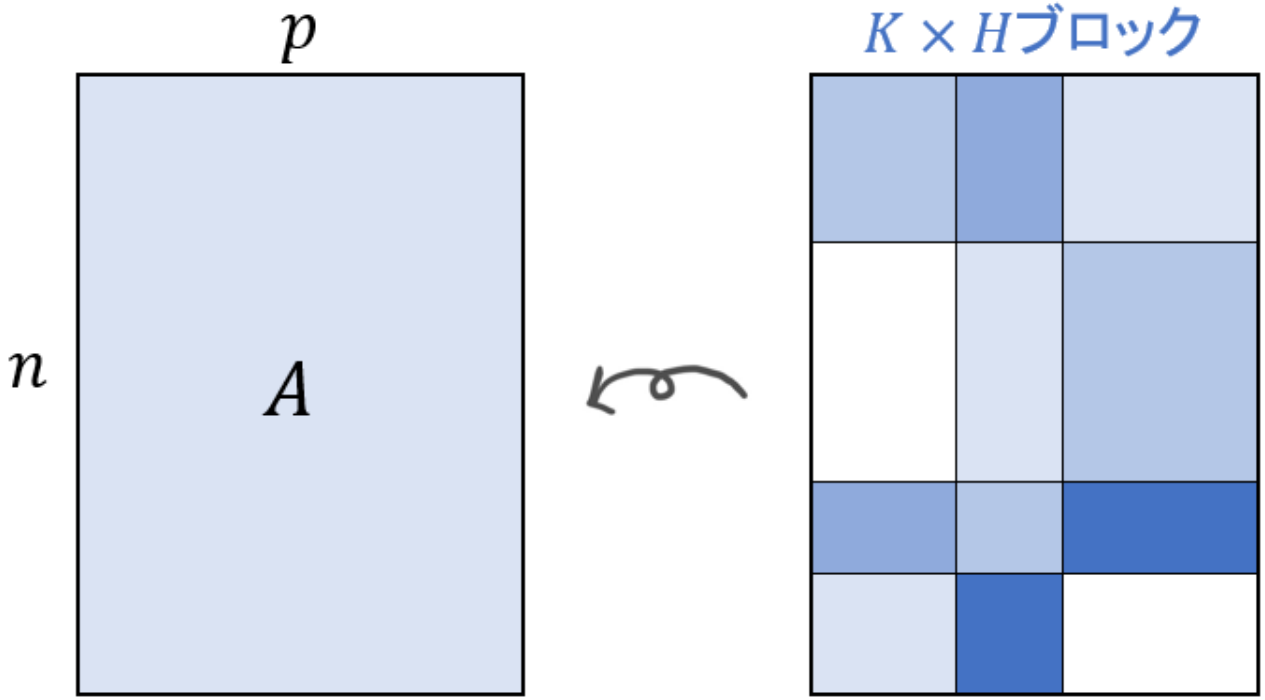
- Latent Block Model の問題設定（共クラスタリング）に適用可能な統計的検定手法を構築



目次

Latent Block Model に対する適合度検定

- 観測データ： $A = (A_{ij})_{1 \leq i \leq n, 1 \leq j \leq p} \in \mathbb{R}^{n \times p}$
- $K \times H$ 個のブロックからなるクラスタ構造を持つことを仮定（K: 未知の行クラスタ数，H: 未知の列クラスタ数）
- K, H は未知なので，仮説クラスタ数 K_0, H_0 を持つと思ってクラスタ構造の推定を行い，その結果を用いて (K_0, H_0) についての検定を行う



記法と仮定 I

- 観測データ： $A = (A_{ij})_{1 \le i \le n, 1 \le j \le p} \in \mathbb{R}^{n \times p}$
- 行列サイズ $n, p \propto m$ と仮定し，$m \to \infty$ の極限を考える
- クラスタ数 K, H は行列サイズ m によらない定数と仮定する
- 各クラスタに含まれる最小の行・列サイズ $n_{\min}, p_{\min}$ は $n_{\min} = \Omega(m)$, $p_{\min} = \Omega(m)$ を満たすと仮定する
- $(K, H) = (K_0, H_0)$ のとき実現可能であると呼ぶ．また，$K > K_0$ か $H > H_0$ の少なくとも一方が成り立つとき，実現不可能であると呼ぶ．
- 行クラスタの割り当て：$g^{(1)} = \left(g_i^{(1)}\right)_{1 \le i \le n} \in \{1, \ldots, K\}$．$g_i^{(1)}$ は i 行目のクラスタ番号を表す
- 列クラスタの割り当て：$g^{(2)} = \left(g_j^{(2)}\right)_{1 \le j \le p} \in \{1, \ldots, H\}$．$g_j^{(2)}$ は j 列目のクラスタ番号を表す

記法と仮定 II

- A の各要素はブロックごとに同一の分布から独立に生成されている．(k, h) ブロックの平均を B_{kh}, 標準偏差を $S_{kh} > 0$ とする $(k = 1, \ldots, K.h = 1, \ldots, H)$
- A の各要素の平均を格納した行列 $P = (P_{ij})_{1 \leq i \leq n, 1 \leq j \leq p},\ P_{ij} = B_{g_i^{(1)} g_j^{(2)}}$.
- A の各要素の標準偏差を格納した行列 $\sigma = (\sigma_{ij})_{1 \leq i \leq n, 1 \leq j \leq p},\ \sigma_{ij} = S_{g_i^{(1)} g_j^{(2)}}$.
- $\mathbb{E}[A_{ij}] = P_{ij},\ \mathbb{E}[(A_{ij} - P_{ij})^2] = \sigma_{ij}^2$.
- 平均 0, 分散 1 に標準化された観測行列 Z を以下で定義する．Z は sub-exponential decay を持つ（i.e., ある $\vartheta > 0$ が存在して $x > 1$ について $\Pr(|Z_{ij}| > x) \leq \vartheta^{-1} \exp(-x^{\vartheta})$ を満たす）と仮定する．

$$Z = (Z_{ij})_{1 \leq i \leq n, 1 \leq j \leq p},\ Z_{ij} = (A_{ij} - P_{ij})/\sigma_{ij}. \tag{3}$$

記法と仮定 III

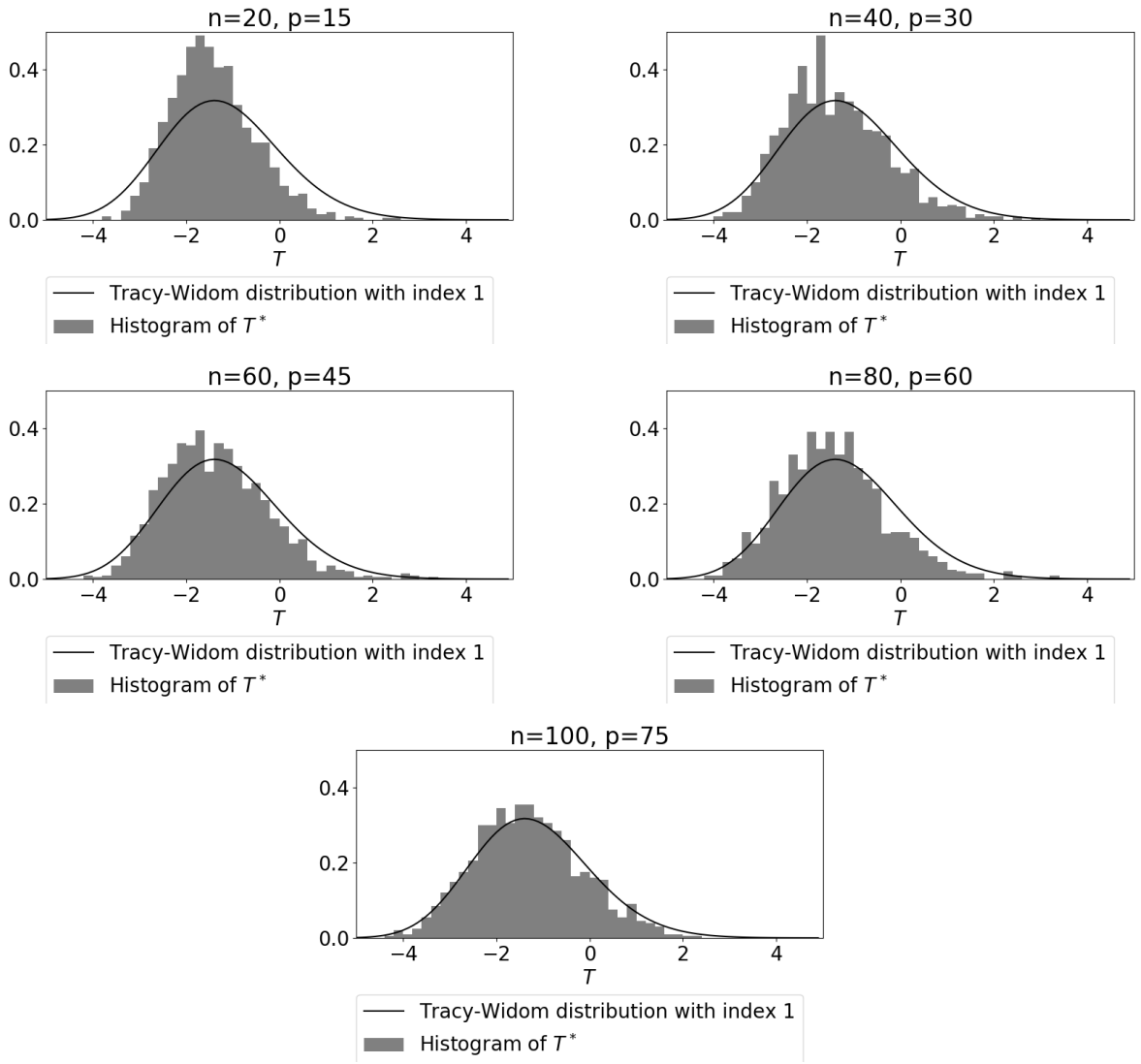
- 行列 $Z^\top Z$ の最大固有値を λ_1 とする．[12] より，$m \to \infty$ で以下の T^* は index 1 の Tracy-Widom 分布 (TW_1 分布) に法則収束することが知られている．

$$T^* = \frac{\lambda_1 - a^{\mathrm{TW}}}{b^{\mathrm{TW}}}, \ T^* \rightsquigarrow TW_1 \ (\text{Convergence in law}). \tag{4}$$

ただし，

$$a^{\mathrm{TW}} = (\sqrt{n} + \sqrt{p})^2, \ b^{\mathrm{TW}} = (\sqrt{n} + \sqrt{p}) \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p}} \right)^{\frac{1}{3}}. \tag{5}$$

記法と仮定 IV


n=20, p=15
n=40, p=30
n=60, p=45
n=80, p=60
n=100, p=75
0.0
0.2
0.4
−4
−2
0
2
4
T
Tracy-Widom distribution with index 1
Histogram of T^*

記法と仮定 V

- 仮説のクラスタ数 (K_0, H_0) として推定された行クラスタの割り当てを $\hat{g}^{(1)} = \left(\hat{g}_i^{(1)}\right)_{1 \le i \le n} \in \{1, \ldots, K_0\}$, 列クラスタの割り当てを $\hat{g}^{(2)} = \left(\hat{g}_j^{(2)}\right)_{1 \le j \le p} \in \{1, \ldots, H_0\}$ とする．k 番目の行クラスタに属すると推定された行集合を I_k, h 番目の列クラスタに属すると推定された列集合を J_h とする．

$$I_k = \left\{ i : \hat{g}_i^{(1)} = k \right\}, \ J_h = \left\{ j : \hat{g}_j^{(2)} = h \right\}. \tag{6}$$

- 実現可能な時，ブロック構造（行・列のクラスタ割り当て）を推定するアルゴリズムは 一致性 を持つと仮定する（i.e., $m \to \infty$ で正しいブロック構造を出力する確率が 1 に収束する）．いくつかの共クラスタリングアルゴリズムが一致性を持つことが知られている [1, 4, 8].

記法と仮定 VI

- 各推定ブロックごとのサンプル平均・標準偏差を格納した行列を $\hat{P}$, $\hat{\sigma}$ とする.

$$\hat{B} = (\hat{B}_{kh})_{1\leq k\leq K_0, 1\leq h\leq H_0},\ \hat{B}_{kh} = \frac{1}{|I_k||J_h|}\sum_{i\in I_k, j\in J_h} A_{ij},$$

$$\hat{P} = (\hat{P}_{ij})_{1\leq i\leq n, 1\leq j\leq p},\ \hat{P}_{ij} = \hat{B}_{\hat{g}_i^{(1)}\hat{g}_j^{(2)}},$$

$$\hat{S} = (\hat{S}_{kh})_{1\leq k\leq K_0, 1\leq h\leq H_0},\ \hat{S}_{kh} = \sqrt{\frac{1}{|I_k||J_h|}\sum_{i\in I_k, j\in J_h}\left(A_{ij} - \hat{P}_{ij}\right)^2},$$

$$\hat{\sigma} = (\hat{\sigma}_{ij})_{1\leq i\leq n, 1\leq j\leq p},\ \hat{\sigma}_{ij} = \hat{S}_{\hat{g}_i^{(1)}\hat{g}_j^{(2)}}, \tag{7}$$

- 標準化観測行列の推定版を $\hat{Z}$ とする.

$$\hat{Z} = (\hat{Z}_{ij})_{1\leq i\leq n, 1\leq j\leq p},\ \hat{Z}_{ij} = \frac{A_{ij} - \hat{P}_{ij}}{\hat{\sigma}_{ij}}. \tag{8}$$

記法と仮定 VII

- $\hat{Z}^\top \hat{Z}$ の最大固有値 $\hat{\lambda}_1$ から，検定統計量 T を以下で定義する．

$$T = \frac{\hat{\lambda}_1 - a^{\mathrm{TW}}}{b^{\mathrm{TW}}}. \tag{9}$$

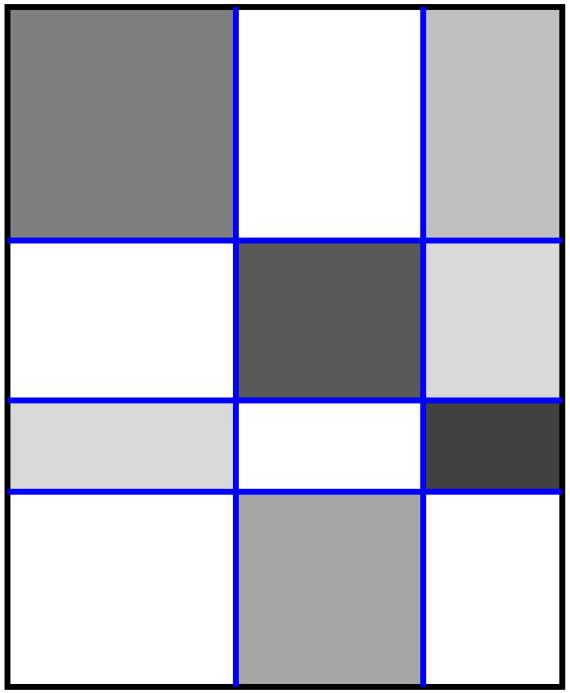
- また，(k, h) 番目の帰無ブロック（帰無仮説が正しいとしたときの正しいブロック構造におけるブロック）におけるサンプル平均・標準偏差を $\tilde{B}_{kh}$, $\tilde{S}_{kh}$ とする．

$$\begin{aligned}
\tilde{P} &= (\tilde{P}_{ij})_{1 \le i \le n, 1 \le j \le p},\ \tilde{P}_{ij} = \tilde{B}_{g_i^{(1)} g_j^{(2)}}, \\
\tilde{\sigma} &= (\tilde{\sigma}_{ij})_{1 \le i \le n, 1 \le j \le p},\ \tilde{\sigma}_{ij} = \tilde{S}_{g_i^{(1)} g_j^{(2)}}, \\
\tilde{Z} &= (\tilde{Z}_{ij})_{1 \le i \le n, 1 \le j \le p},\ \tilde{Z}_{ij} = \frac{A - \tilde{P}_{ij}}{\tilde{\sigma}_{ij}}.
\end{aligned} \tag{10}$$

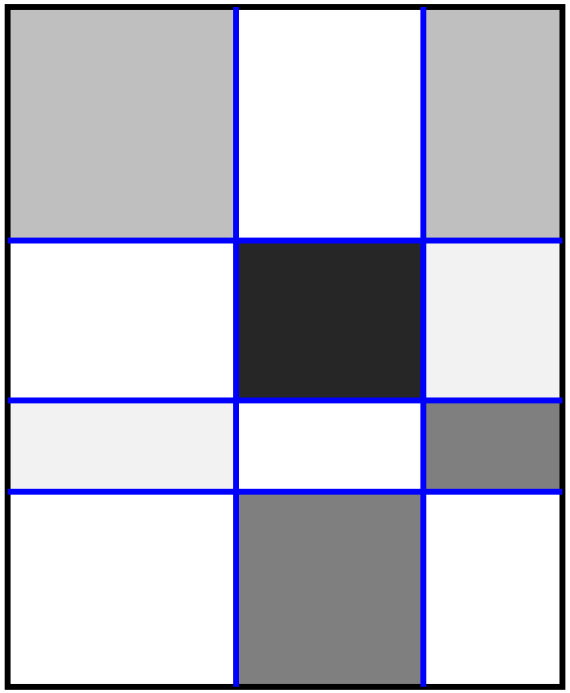
- $\tilde{Z}^\top \tilde{Z}$ の最大固有値を $\tilde{\lambda}_1$ とする．

記法と仮定 VIII

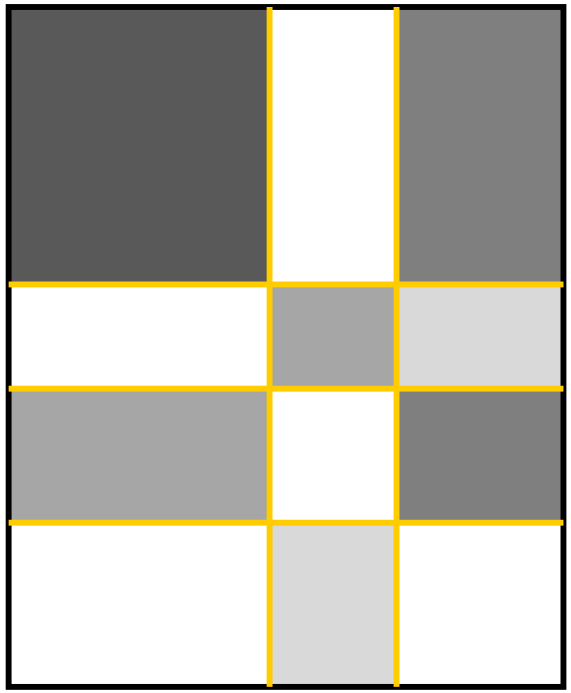
- Z: **正しいブロック構造** gと**母平均・標準偏差**で標準化された観測行列
- T^*: $Z^{\mathrm{T}}Z$ の最大固有値λ_1を正規化した統計量



- $\tilde{Z}$: **正しいブロック構造** gと**標本平均・標準偏差**で標準化された観測行列
- $\tilde{\lambda}_1$: $\tilde{Z}^{\mathrm{T}}\tilde{Z}$の最大固有値



- $\hat{Z}$: **推定ブロック構造**$\hat{g}$と**標本平均・標準偏差**で標準化された観測行列
- T: $\hat{Z}^{\mathrm{T}}\hat{Z}$の最大固有値$\hat{\lambda}_1$を正規化した統計量



主結果：検定統計量の漸近的な性質

Theorem 1 (実現可能な場合)

前記の仮定と帰無仮説（*i.e.*, 実現可能）$(K_0, H_0) = (K, H)$ の下で，
$m \to \infty$ で T は TW_1 分布に法則収束する．

$$T \rightsquigarrow TW_1 \text{ (Convergence in law)}. \tag{11}$$

Theorem 2 (実現不可能な場合)

前記の仮定と対立仮説（*i.e.*, 実現不可能，$K_0 < K$ か $H_0 < H$ の少なくとも一方が成り立つ）の下で，

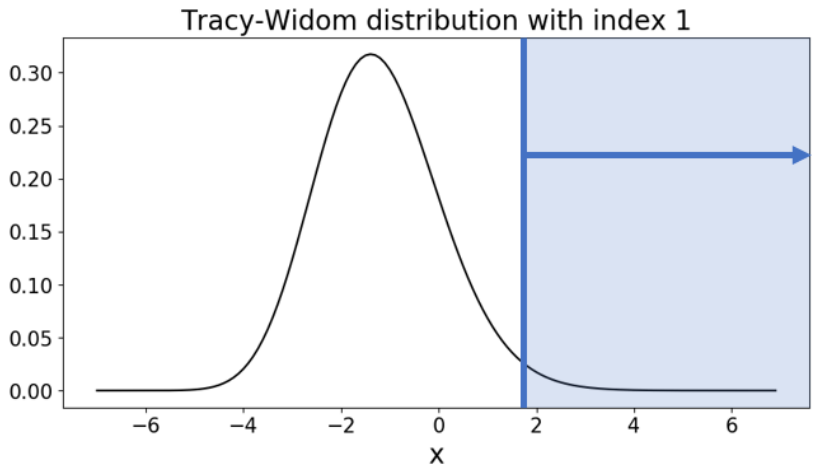
$$T = \Theta_p\left(m^{\frac{5}{3}}\right). \tag{12}$$

与えられた仮説クラスタ数 (K_0, H_0) についての検定

- 上記の Theorem 1, 2 から，仮説クラスタ数 (K_0, H_0) に関する有意水準 α の検定を以下のように定義する．

$$\text{Reject null hypothesis } ((K, H) = (K_0, H_0)), \text{ if } T \geq t(\alpha), \quad (13)$$

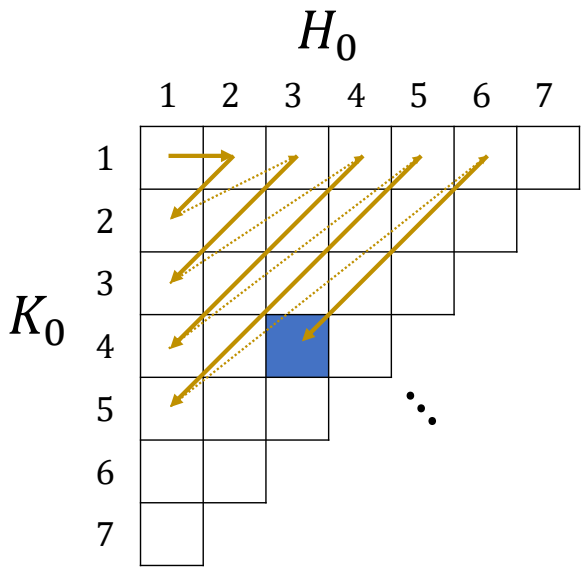
ただし，$t(\alpha)$ を TW_1 分布の α upper quantile とする．上記のルール (13) に基づき様々な仮説クラスタ数について検定を行うことで，クラスタ数の選択ができる（次ページ）．



α-upper quantile
（確率密度関数の
右側，面積がα）

検定の流れ

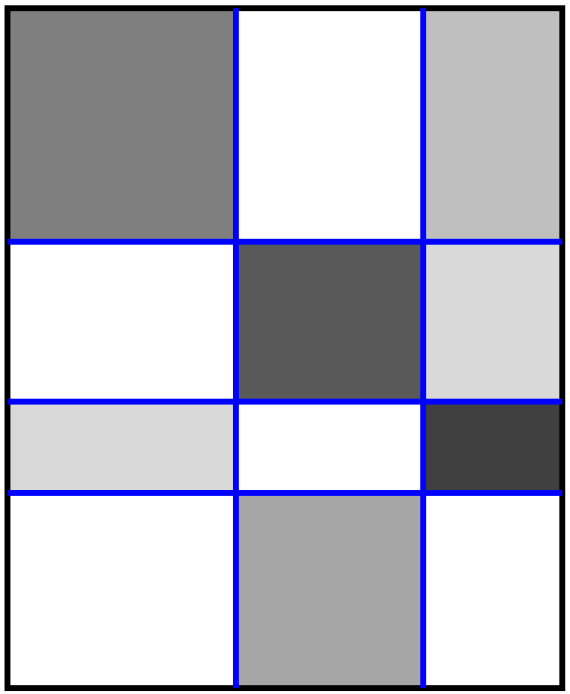
- 帰無仮説：$(K, H) = (K_0, H_0)$（観測データは $K_0 \times H_0$ 個のブロックからなる）
- 対立仮説：$K > K_0$ または $H > H_0$（行・列の少なくともどちらかでクラスタ数が足りない）
- 以下の順番で検定し，帰無仮説受容時の設定 $(\hat{K}, \hat{H})$ を選択結果とする
  1. $(K_0, H_0) = (1, 1)$,
  2. $(K_0, H_0) = (1, 2), (2, 1)$,
  3. $(K_0, H_0) = (1, 3), (2, 2), (3, 1), \cdots$



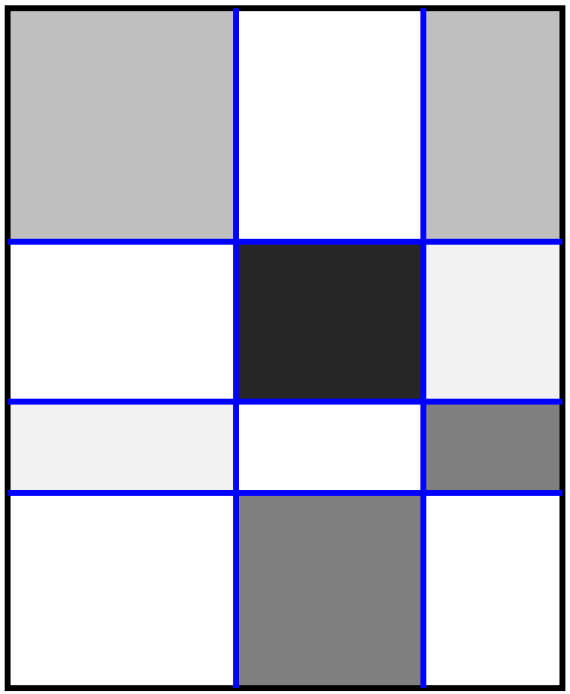
～5 分休憩タイム～

定理 1 の証明 I

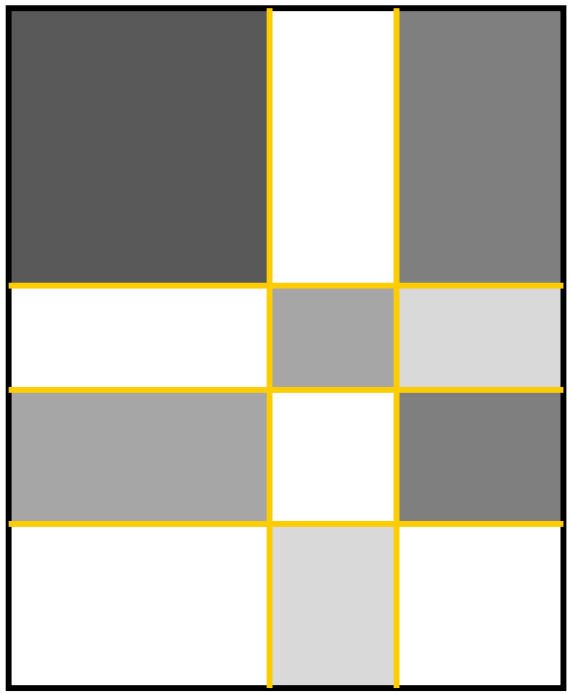
- Z: **正しいブロック構造** g と**母平均・標準偏差**で標準化された観測行列
- T^*: $Z^{\mathrm{T}}Z$ の最大固有値 λ_1 を正規化した統計量



- $\tilde{Z}$: **正しいブロック構造** g と**標本平均・標準偏差**で標準化された観測行列
- $\tilde{\lambda}_1$: $\tilde{Z}^{\mathrm{T}}\tilde{Z}$ の最大固有値



- $\hat{Z}$: **推定ブロック構造** $\hat{g}$ と**標本平均・標準偏差**で標準化された観測行列
- T: $\hat{Z}^{\mathrm{T}}\hat{Z}$ の最大固有値 $\hat{\lambda}_1$ を正規化した統計量



定理 1 の証明 II

大まかな証明方針

- $\left|\lambda_1 - \tilde{\lambda}_1\right|$ が小さいことを示す（具体的には $\left|\lambda_1 - \tilde{\lambda}_1\right| = O_p\left(m^{\frac{2}{7}+\epsilon}\right)$ for all $\epsilon > 0$）．ここでランダム行列の固有ベクトルの delocalization property (後述) を用いる．
- 上記とブロック構造推定アルゴリズムの一致性の仮定から，$\left|\lambda_1 - \hat{\lambda}_1\right| = O_p\left(m^{\frac{2}{7}+\epsilon}\right)$ for all $\epsilon > 0$ も示せる．
- [12] より，$m \to \infty$ で $T^* = \frac{\lambda_1 - a^{\mathrm{TW}}}{b^{\mathrm{TW}}}$ は TW_1 分布に法則収束することから，$T = \frac{\hat{\lambda}_1 - a^{\mathrm{TW}}}{b^{\mathrm{TW}}} = T^* + \frac{\hat{\lambda}_1 - \lambda_1}{b^{\mathrm{TW}}}$ も TW_1 分布に法則収束することを示せる．

定理 1 の証明 III

まずは正しいブロック構造における母平均（標準偏差）B_{kh} (S_{kh}) と標本平均（標準偏差）$\tilde{B}_{kh}$ $(\tilde{S}_{kh})$ の差を求める.
仮定から各帰無ブロック内の要素数は m^2 に比例するので，中心極限定理から $\sqrt{m^2}\left(B_{kh} - \tilde{B}_{kh}\right)$ は漸近的に正規分布 $\mathcal{N}(0, S_{kh}^2)$ に従う（法則収束）.
Prokhorov の定理 [14]（i.e., 法則収束するならば $O_p(1)$）より，

$$\left|\tilde{B}_{kh} - B_{kh}\right| = O_p\left(\frac{1}{m}\right). \tag{14}$$

また，標準偏差について，以下が成り立つことを示すことができる（証明略）.

$$\left|\tilde{S}_{kh} - S_{kh}\right| = O_p\left(\frac{1}{m}\right). \tag{15}$$

定理 1 の証明 IV

ここからは $\tilde{\lambda}_1$ と λ_1 の差について考える．

[12] より，$m \to \infty$ で $T^* = \frac{\lambda_1 - a^{\mathrm{TW}}}{b^{\mathrm{TW}}}$ は TW_1 分布に法則収束する． よって，Prokhorov の定理 [14] から，$\lambda_1 = O_p(m)$. また，Z の最大特異値は $\sqrt{\lambda_1} = \|Z\|_{\mathrm{op}} = O_p(\sqrt{m})$.
ある行列 X の (k, h) 番目の帰無ブロック（部分行列）を $X^{(k,h)}$ と書き，このサイズを $n_k \times p_h$ とする．Z と $\tilde{Z}$ の定義より，

$$Z^{(k,h)} = \frac{A^{(k,h)} - P^{(k,h)}}{S_{kh}}, \ \tilde{Z}^{(k,h)} = \frac{A^{(k,h)} - \tilde{P}^{(k,h)}}{\tilde{S}_{kh}}. \tag{16}$$

これとフロベニウスノルムは作用素ノルム以上になること，(14), (15), [12] を用いると，以下を導ける．

$$\|Z^{(k,h)} - \tilde{Z}^{(k,h)}\|_{\mathrm{op}} = O_p(1). \tag{17}$$

定理 1 の証明 V

ある行列の作用素ノルムはその部分行列の作用素ノルムの和以下であることと，ブロック数が行列サイズ m によらない定数であることから，

$$\|Z - \tilde{Z}\|_{\mathrm{op}} \leq \sum_{k=1}^{K} \sum_{h=1}^{H} \|Z^{(k,h)} - \tilde{Z}^{(k,h)}\|_{\mathrm{op}} = O_p(1). \tag{18}$$

一方，作用素ノルムの劣加法性より，

$$\left| \|Z\|_{\mathrm{op}} - \|\tilde{Z}\|_{\mathrm{op}} \right| \leq \|Z - \tilde{Z}\|_{\mathrm{op}} = O_p(1). \tag{19}$$

定理 1 の証明 VI

次に，ブロック構造のずれについて考える．「$\tilde{Z} = \hat{Z}$ が成り立つ」という事象を $\mathcal{F}_m$ と，「$\left| \|Z\|_{\mathrm{op}} - \|\tilde{Z}\|_{\mathrm{op}} \right| \leq C$ が成り立つ」という事象を $\mathcal{G}_{m,C}$ とすると，これらの同時確率は

$$\Pr\left(\mathcal{F}_m \cap \mathcal{G}_{m,C}\right) \geq 1 - \Pr\left(\mathcal{F}_m^{\mathrm{C}}\right) - \Pr\left(\mathcal{G}_{m,C}^{\mathrm{C}}\right). \tag{20}$$

ただし，$\mathcal{A}^{\mathrm{C}}$ は $\mathcal{A}$ の余事象を表す．
実現可能な（＝ブロック数が足りている）設定において，ブロック構造推定アルゴリズムが一致性を持つという仮定より，$m \to \infty$ で $\Pr\left(\mathcal{F}_m^{\mathrm{C}}\right) \to 0$.
このことと (19) より，

$$\forall \epsilon > 0,\ \exists C > 0, M > 0,\ \forall m \geq M,\ \Pr\left(\mathcal{F}_m \cap \mathcal{G}_{m,C}\right) \geq 1 - \epsilon. \tag{21}$$

よって，以下も言える．

$$\left| \|Z\|_{\mathrm{op}} - \|\hat{Z}\|_{\mathrm{op}} \right| = O_p(1). \tag{22}$$

定理 1 の証明 VII

このことを用いて，以下を示せる．

$$\forall \epsilon > 0,\ \frac{|\lambda_1 - \hat{\lambda}_1|}{b^{\mathrm{TW}}} = O_p\left(m^{-\frac{1}{21}+\epsilon}\right). \tag{23}$$

これと [12] の結果より $m \to \infty$ で $T^* = \frac{\lambda_1 - a^{\mathrm{TW}}}{b^{\mathrm{TW}}}$ は TW_1 分布に法則収束することから，Slutsky の定理（i.e., $X_n \rightsquigarrow X$ かつ Y_n が c に確率収束するならば $X_n + Y_n \rightsquigarrow X + c$）より，$\epsilon < \frac{1}{21}$ に設定することで，

$$T = \frac{\hat{\lambda}_1 - a^{\mathrm{TW}}}{b^{\mathrm{TW}}} = T^* + \frac{\hat{\lambda}_1 - \lambda_1}{b^{\mathrm{TW}}} \rightsquigarrow TW_1 \ (\text{Convergence in law}). \tag{24}$$

以下では，(23) を示す．

定理 1 の証明 VIII

Lemma 3

$Z^\top Z$ と $\tilde{Z}^\top \tilde{Z}$ の最大固有値をそれぞれ $\lambda_1, \tilde{\lambda}_1$ とすると,

$$\forall \epsilon > 0,\ \lambda_1 \leq \tilde{\lambda}_1 + O_p\left(m^\epsilon\right). \tag{25}$$

$Z^\top Z$ と $\tilde{Z}^\top \tilde{Z}$ の最大固有値に対応する固有ベクトルをそれぞれ $\boldsymbol{v}, \tilde{\boldsymbol{v}}$ とおく.

$$\begin{aligned} Z^\top Z \boldsymbol{v} &= \lambda_1 \boldsymbol{v}, \quad \|\boldsymbol{v}\| = 1, \\ \tilde{Z}^\top \tilde{Z} \tilde{\boldsymbol{v}} &= \tilde{\lambda}_1 \tilde{\boldsymbol{v}}, \quad \|\tilde{\boldsymbol{v}}\| = 1. \end{aligned} \tag{26}$$

$\sqrt{\tilde{\lambda}_1}$ は $\tilde{Z}$ の最大特異値なので,

$$\sqrt{\tilde{\lambda}_1} = \sup_{\boldsymbol{u} \in \mathbb{R}^p} \frac{\|\tilde{Z}\boldsymbol{u}\|}{\|\boldsymbol{u}\|} \geq \frac{\|\tilde{Z}\boldsymbol{v}\|}{\|\boldsymbol{v}\|} = \|\tilde{Z}\boldsymbol{v}\| \iff \tilde{\lambda}_1 \geq \|\tilde{Z}\boldsymbol{v}\|^2. \tag{27}$$

定理 1 の証明 IX

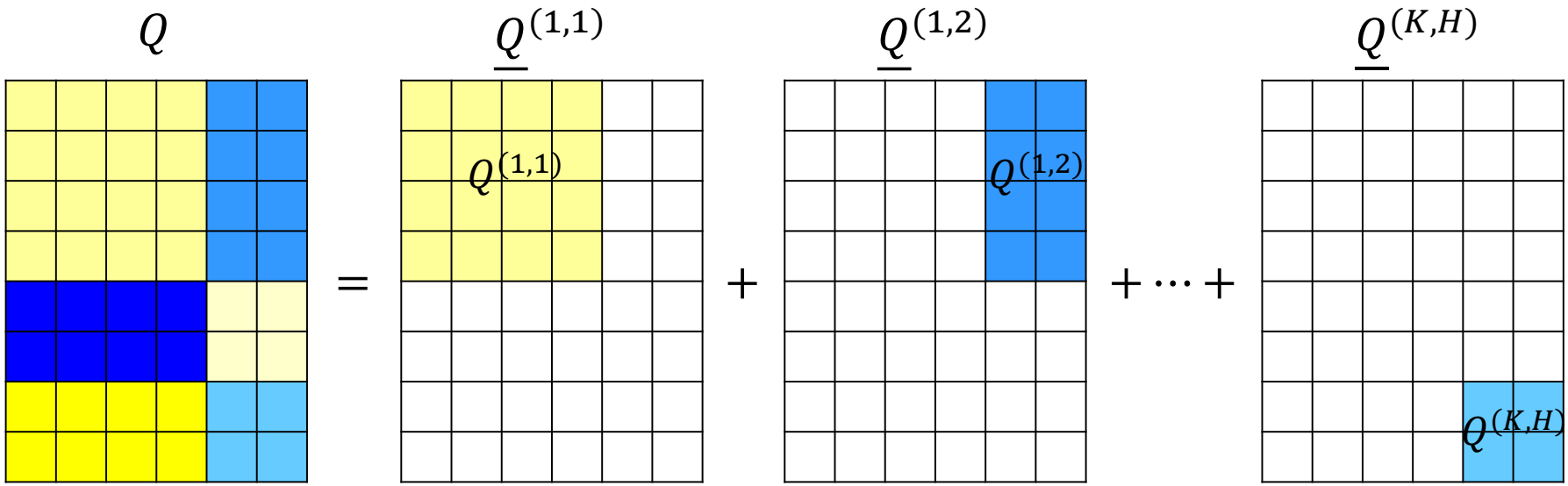
また，各 (k,h) ブロックに対応する行列 $Q^{(k,h)}$ を以下で定義する．

$$
\begin{aligned}
Q^{(k,h)} &\equiv Z^{(k,h)} - \frac{\tilde{S}_{kh}}{S_{kh}} \tilde{Z}^{(k,h)} = \dots \\
&= \left(\frac{1}{n_k p_h} \sum_{i=1}^{n_k} \sum_{j=1}^{p_h} Z_{ij}^{(k,h)} \right) \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{bmatrix}. \qquad (28)
\end{aligned}
$$

ここで，(k,h) 番目の帰無ブロック（※帰無仮説における正しいブロック構造におけるブロックのこと）のサイズを $n_k \times p_h$ とした．
さらに，$\underline{Z}^{(k,h)}$, $\underline{\tilde{Z}}^{(k,h)}$, $\underline{Q}^{(k,h)}$ を (k,h) 番目の帰無ブロックの部分が $Z^{(k,h)}$, $\tilde{Z}^{(k,h)}$, $Q^{(k,h)}$ でそれ以外の要素がゼロであるような $n \times p$ の行列とする．

定理 1 の証明 X

行列 Q を以下で定義する．$Q \equiv \sum_{k=1}^{K} \sum_{h=1}^{H} \underline{Q}^{(k,h)}$.



定理 1 の証明 XI

(27) で $\tilde{\lambda}_1 \geq \|\tilde{Z}\boldsymbol{v}\|^2$ だったので，定義に基づいて展開していくと

$$\begin{aligned}\tilde{\lambda}_1 \geq \|\tilde{Z}\boldsymbol{v}\|^2 &= \|\sum_{k=1}^{K}\sum_{h=1}^{H}\underline{\tilde{Z}}^{(k,h)}\boldsymbol{v}\|^2 = \dots \\ &\geq \lambda_1 - 2\sqrt{\lambda_1}\|Q\boldsymbol{v}\| - 2(\sqrt{\lambda_1} + \|Q\boldsymbol{v}\|) \\ &\left[\sum_{k=1}^{K}\sum_{h=1}^{H}\left|1 - \frac{S_{kh}}{\tilde{S}_{kh}}\right|\left(\sqrt{\lambda_1^{(k,h)}} + \|\underline{Q}^{(k,h)}\boldsymbol{v}\|\right)\right]. \qquad (29)\end{aligned}$$

ここで，$\lambda_1^{(k,h)}$ を $\left(Z^{(k,h)}\right)^\top Z^{(k,h)}$ の最大固有値とした．
ここから，$Z^\top Z$ の固有ベクトルの delocalization property を用いて以下を示す．

- $\|Q\boldsymbol{v}\| = O_p\left(\frac{1}{\sqrt{m}}\right)$.
- $\|\underline{Q}^{(k,h)}\boldsymbol{v}\| = O_p\left(\frac{1}{\sqrt{m}}\right)$.

定理 1 の証明 XII

v_j をベクトル $\boldsymbol{v}$ の j 番目の要素とし，

$$\nu_{kh} \equiv \frac{1}{n_k p_h} \sum_{i=1}^{n_k} \sum_{j=1}^{p_h} z_{ij}^{(k,h)},\ \omega_{hh'} \equiv \sum_{k=1}^{K} n_k \nu_{kh} \nu_{kh'},\ \zeta_h \equiv \sum_{h'=1}^{H} \omega_{hh'} \sum_{j \in J_{h'}} v_j. \tag{30}$$

とすると，Q の (k,h) 番目のブロック，$Q^\top Q$ の (h,h') 番目のブロック，

$Q^\top Q\boldsymbol{v}$ の h 番目のブロックはそれぞれ $\nu_{kh} \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$,

$\omega_{hh'} \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$, and $\zeta_h \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ で与えられる．

定理 1 の証明 XIII

[3] の定理 2.17 より，$Z^\top Z$ の j 番目の固有ベクトル $\boldsymbol{v}_j$ $(\|\boldsymbol{v}_j\| = 1)$ は以下の *delocalization property* を持つ．任意の $\tilde{d} \in \mathbb{N}$ と，$\|\boldsymbol{w}^{(i)}\| = 1$ を満たす任意の定数ベクトル $\{\boldsymbol{w}^{(i)}\}$ $(i = 1, \ldots, m^{\tilde{d}})$ に対し，

$$\max_{i \in \{1,\ldots,m^{\tilde{d}}\}} \max_{j \in \{1,\ldots,p\}} |\boldsymbol{v}_j^\top \boldsymbol{w}^{(i)}| = O_p\left(m^{-\frac{1}{2}+\epsilon}\right), \text{ for all } \epsilon > 0. \tag{31}$$

$\boldsymbol{u}^{(h)} \in \mathbb{R}^p$ を，h 番目の列クラスタに当たる要素が $\frac{1}{\sqrt{p_h}}$ でそれ以外の要素がゼロであるようなベクトルとすると，上記の性質より，任意の $\epsilon > 0$ について，$|\boldsymbol{v}^\top \boldsymbol{u}^{(h)}| = O_p\left(m^{-\frac{1}{2}+\epsilon}\right)$ が成り立つ．ここで，

- $Q^\top Q \boldsymbol{v} = \sum_{h=1}^H \zeta_h \sqrt{p_h} \boldsymbol{u}^{(h)}$.
- $\nu_{kh} = O_p\left(\frac{1}{m}\right)$, $\omega_{hh'} = O_p\left(\frac{1}{m}\right)$.
- $\forall \epsilon > 0$, $\zeta_h = \sum_{h'=1}^H \omega_{hh'} \sqrt{p_{h'}} \boldsymbol{v}^\top \boldsymbol{u}^{(h')} = O_p\left(m^{-1+\epsilon}\right)$.

定理 1 の証明 XIV

より，

$$\|Q\boldsymbol{v}\| = \sqrt{\sum_{h=1}^{H} \zeta_h \sqrt{p_h} \boldsymbol{v}^\top \boldsymbol{u}^{(h)}} = O_p\left(m^{-\frac{1}{2}+\epsilon}\right), \text{ for all } \epsilon > 0. \tag{32}$$

同様に，行列 $(\underline{Q}^{(k,h)})^\top \underline{Q}^{(k,h)}$ は，(h,h) ブロックの部分が

$n_k \nu_{kh}^2 \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{bmatrix}$ で，それ以外の部分の要素がゼロの行列であるため，

$(\underline{Q}^{(k,h)})^\top \underline{Q}^{(k,h)} \boldsymbol{v} = n_k \nu_{kh}^2 \left(\sum_{j \in J_h} v_j\right) \sqrt{p_h} \boldsymbol{u}^{(h)}$ が成り立ち，このことから以下が導ける．

$$\|\underline{Q}^{(k,h)} \boldsymbol{v}\| = O_p\left(m^{-\frac{1}{2}+\epsilon}\right), \text{ for all } \epsilon > 0. \tag{33}$$

定理 1 の証明 XV

さらに,

$$\left|1 - \frac{S_{kh}}{\tilde{S}_{kh}}\right| = O_p\left(\frac{1}{m}\right). \tag{34}$$

が成り立つ（証明略）ことから，(32), (33), (29) より，$\epsilon < \frac{1}{2}$ とすれば,

$$\begin{aligned}\tilde{\lambda}_1 &\geq \lambda_1 - O_p\left(m^{\epsilon}\right) - O_p\left(m^{\frac{1}{2}}\right)\left[\sum_{k=1}^{K}\sum_{h=1}^{H} O_p\left(m^{-1}\right) O_p\left(m^{\frac{1}{2}}\right)\right] \\ &= \lambda_1 - O_p\left(m^{\epsilon}\right). \end{aligned} \tag{35}$$

が証明できた.

定理 1 の証明 XVI

Lemma 4

$Z^\top Z$ と $\tilde{Z}^\top \tilde{Z}$ の最大固有値をそれぞれ $\lambda_1, \tilde{\lambda}_1$ とすると，

$$\tilde{\lambda}_1 \leq \lambda_1 + O_p\left(m^{\frac{2}{7}+\epsilon}\right), \text{ for all } \epsilon > 0. \tag{36}$$

$Z^\top Z$ の固有値（降順）と対応する固有ベクトルを $\{\lambda_j\}, \{\boldsymbol{v}_j\}$ ($\|\boldsymbol{v}_j\| = 1$, $j = 1, \ldots, p$) とする．また，$\tau_{kh} \equiv \frac{S_{kh}}{\tilde{S}_{kh}}$ とおく．(34) から $|\tau_{kh} - 1| = O_p\left(\frac{1}{m}\right)$ が成り立つ．

定理 1 の証明 XVII

$Z^\top Z$ は対称行列なので，固有ベクトル $\{\boldsymbol{v}_j\}$ は正規直交系をなし，以下を満たす唯一の係数 $\{c_j\}$ が存在する．

$$\tilde{\boldsymbol{v}} = \sum_{j=1}^{p} c_j \boldsymbol{v}_j = \tilde{\boldsymbol{v}}_1 + \tilde{\boldsymbol{v}}_2, \tag{37}$$

$$\tilde{\boldsymbol{v}}_1 \equiv \sum_{j=1}^{t} c_j \boldsymbol{v}_j, \quad \tilde{\boldsymbol{v}}_2 \equiv \sum_{j=t+1}^{p} c_j \boldsymbol{v}_j,$$

$$\lambda_t \geq \lambda_1 - n^d, \quad \lambda_{t+1} < \lambda_1 - n^d, \quad d = \frac{5}{7}. \tag{38}$$

定理 1 の証明 XVIII

このことと，K, H が行列サイズ m によらない定数であることを用いて，以下を導ける．

$$\begin{aligned}\tilde{\lambda}_1 &= \tilde{\boldsymbol{v}}^\top \tilde{Z}^\top \tilde{Z} \tilde{\boldsymbol{v}} = \dots \\ &\leq \|Z\tilde{\boldsymbol{v}}\|^2 + O_p(1) - 2\sum_{k=1}^{K}\sum_{h=1}^{H} \tau_{kh} \tilde{\boldsymbol{v}}^\top Z^\top \underline{Q}^{(k,h)} \tilde{\boldsymbol{v}} \\ &+ \left[O_p\left(\frac{1}{\sqrt{m}}\right) + \sum_{k=1}^{K}\sum_{h=1}^{H} \tau_{kh} \|\underline{Q}^{(k,h)} \tilde{\boldsymbol{v}}\| \right]^2 . \end{aligned} \tag{39}$$

さらに，(39) の最後の項については，固有ベクトル $\{\boldsymbol{v}_j\}$ の delocalization property [3] を用いて以下を導ける．

$$\|\underline{Q}^{(k,h)} \tilde{\boldsymbol{v}}\| = O_p\left(m^{\epsilon}\right), \quad \text{for all } \epsilon > 0. \tag{40}$$

定理 1 の証明 XIX

K, H が行列サイズ m によらない定数であることと，(40),
$\tau_{kh} = 1 + O_p\left(\frac{1}{m}\right)$ を (39) に代入して，

$$\tilde{\lambda}_1 \leq \|Z\tilde{\boldsymbol{v}}\|^2 - 2\sum_{k=1}^{K}\sum_{h=1}^{H}\tau_{kh}\tilde{\boldsymbol{v}}^\top Z^\top \underline{Q}^{(k,h)}\tilde{\boldsymbol{v}} + O_p\left(m^{2\epsilon}\right), \quad \text{for all } \epsilon > 0. \tag{41}$$

定理 1 の証明 XX

さらに，(38) の定義を用いることで，右辺の最初の 2 項について以下の bound を導ける．

$$\|Z\tilde{\boldsymbol{v}}\|^2 - 2\sum_{k=1}^{K}\sum_{h=1}^{H}\tau_{kh}\tilde{\boldsymbol{v}}^\top Z^\top \underline{Q}^{(k,h)}\tilde{\boldsymbol{v}} = \dots$$

$$\leq \lambda_1 - n^d\|\tilde{\boldsymbol{v}}_2\|^2 - 2\sum_{k=1}^{K}\sum_{h=1}^{H}\tau_{kh}\tilde{\boldsymbol{v}}^\top Z^\top \underline{Q}^{(k,h)}\tilde{\boldsymbol{v}}_1$$

$$- 2\sum_{k=1}^{K}\sum_{h=1}^{H}\tau_{kh}\tilde{\boldsymbol{v}}^\top Z^\top \underline{Q}^{(k,h)}\tilde{\boldsymbol{v}}_2. \tag{42}$$

(42) の第 3 項は，固有ベクトル $\{\boldsymbol{v}_j\}$ の delocalization property [3] を用いて，以下の不等式で抑えられる．

$$-\tilde{\boldsymbol{v}}^\top Z^\top \underline{Q}^{(k,h)}\tilde{\boldsymbol{v}}_1 \leq \cdots \leq \sqrt{t}\; O_p\left(m^\epsilon\right), \quad \text{for all } \epsilon > 0. \tag{43}$$

定理 1 の証明 XXI

また，(42) の第 4 項は，以下の不等式で抑えられる．

$$-\tilde{\boldsymbol{v}}^\top Z^\top \underline{Q}^{(k,h)} \tilde{\boldsymbol{v}}_2 \leq \cdots \leq \|\tilde{\boldsymbol{v}}_2\| O_p(\sqrt{m}). \tag{44}$$

(43) と (44) を (42) に代入して，

$$\begin{aligned} &\|Z\tilde{\boldsymbol{v}}\|^2 - 2\sum_{k=1}^{K}\sum_{h=1}^{H} \tau_{kh} \tilde{\boldsymbol{v}}^\top Z^\top \underline{Q}^{(k,h)} \tilde{\boldsymbol{v}} \leq \ldots \\ &\leq \lambda_1 - n^d \|\tilde{\boldsymbol{v}}_2\|^2 + \sqrt{t}\ O_p\left(m^\epsilon\right) + \|\tilde{\boldsymbol{v}}_2\| O_p(\sqrt{m}). \end{aligned} \tag{45}$$

(38) の定義で，t は $\lambda_j \geq \lambda_1 - n^d$ を満たす j の数であったが，これについて [12] の結果から以下が導ける（証明略）．

$$t = O_p\left(m^{\frac{3}{2}d - \frac{1}{2}}\right). \tag{46}$$

定理 1 の証明 XXII

(45) に上記と (38) の仮定で具体的に $d = \frac{5}{7}$ としていたことを代入すると，任意の $\epsilon > 0$ について，以下が成り立つ．

$$\begin{aligned}
&\|Z\tilde{\boldsymbol{v}}\|^2 - 2\sum_{k=1}^{K}\sum_{h=1}^{H}\tau_{kh}\tilde{\boldsymbol{v}}^\top Z^\top \underline{Q}^{(k,h)}\tilde{\boldsymbol{v}} \\
&\leq \lambda_1 + \|\tilde{\boldsymbol{v}}_2\|\left[n^{\frac{1}{2}}\varpi - n^d\|\tilde{\boldsymbol{v}}_2\|\right] + O_p\left(m^{\frac{2}{7}+\epsilon}\right), \\
&\varpi \equiv n^{-\frac{1}{2}}\|Z\|_{\mathrm{op}}\|\underline{Q}^{(k,h)}\|_{\mathrm{F}} = O_p(1). \qquad (47)
\end{aligned}$$

ここで，(a) $n^{\frac{1}{2}}\varpi - n^d\|\tilde{\boldsymbol{v}}_2\| \leq 0$ の場合は，

$$\|Z\tilde{\boldsymbol{v}}\|^2 - 2\sum_{k=1}^{K}\sum_{h=1}^{H}\tau_{kh}\tilde{\boldsymbol{v}}^\top Z^\top \underline{Q}^{(k,h)}\tilde{\boldsymbol{v}} \leq \lambda_1 + O_p\left(m^{\frac{2}{7}+\epsilon}\right). \qquad (48)$$

定理 1 の証明 XXIII

一方，(b) $n^{\frac{1}{2}}\varpi - n^d\|\tilde{\boldsymbol{v}}_2\| > 0$ の場合は，$\|\tilde{\boldsymbol{v}}_2\| < n^{\frac{1}{2}-d}\varpi$ なので，

$$\|\tilde{\boldsymbol{v}}_2\|\left[n^{\frac{1}{2}}\varpi - n^d\|\tilde{\boldsymbol{v}}_2\|\right] \leq n^{1-d}\varpi^2. \tag{49}$$

(38) の仮定で $d = \frac{5}{7}$ としていたので，$n^{1-d}\varpi^2 = O_p\left(m^{\frac{2}{7}}\right)$ となり，この場合も (48) が成り立つ．つまり，(48) はいつでも成り立つ．(41) に代入して，

$$\tilde{\lambda}_1 \leq \lambda_1 + O_p\left(m^{\frac{2}{7}+\epsilon}\right) + O_p\left(m^{2\epsilon}\right), \quad \text{for all } \epsilon > 0. \tag{50}$$

よって，$\epsilon < \frac{2}{7}$ と設定することにより，

$$\tilde{\lambda}_1 \leq \lambda_1 + O_p\left(m^{\frac{2}{7}+\epsilon}\right), \quad \text{for all } \epsilon > 0. \tag{51}$$

が示せた．

定理 1 の証明 XXIV

Lemma 5

$Z^\top Z$ と $\hat{Z}^\top \hat{Z}$ の最大固有値をそれぞれ $\lambda_1, \hat{\lambda}_1$ とすると，

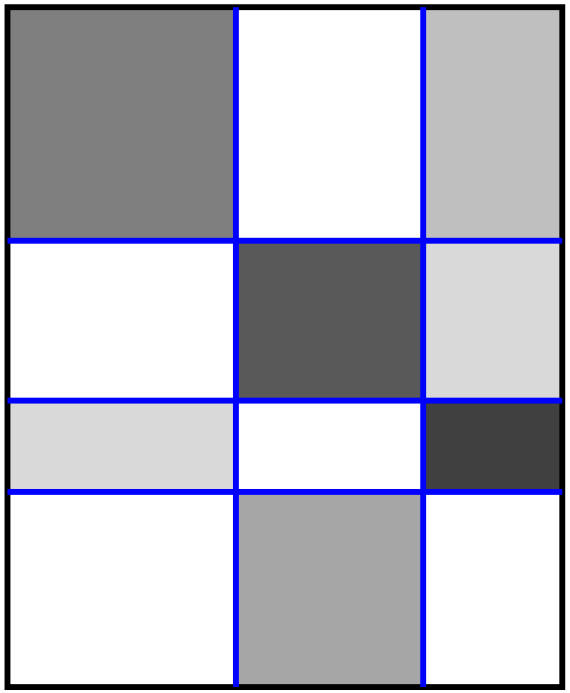
$$\frac{|\lambda_1 - \hat{\lambda}_1|}{b^{\mathrm{TW}}} = O_p\left(m^{-\frac{1}{21}+\epsilon}\right), \quad \text{for all } \epsilon > 0. \tag{52}$$

Lemma 3, 4 から，以下が成り立つことは既に分かっている．

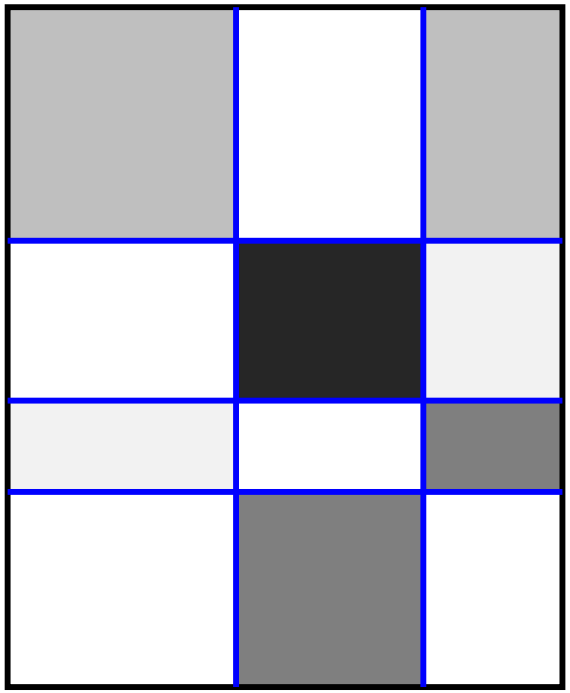
$$\frac{|\lambda_1 - \tilde{\lambda}_1|}{b^{\mathrm{TW}}} = O_p\left(m^{-\frac{1}{21}+\epsilon}\right), \quad \text{for all } \epsilon > 0. \tag{53}$$

定理 1 の証明 XXV

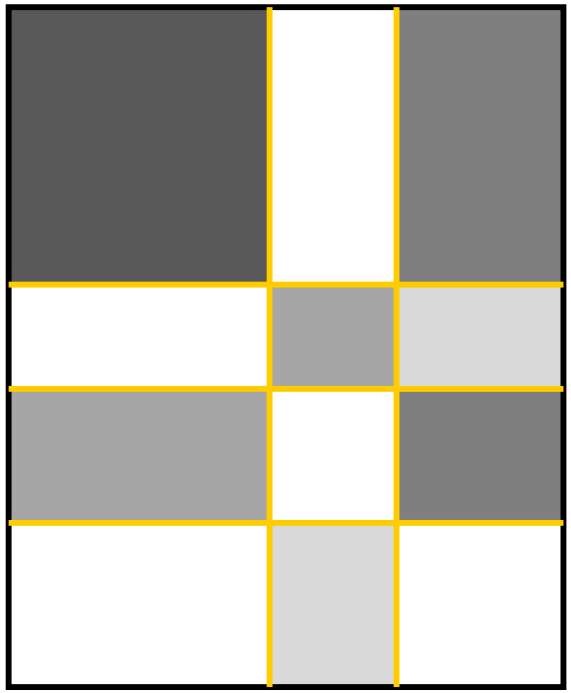
- Z: **正しいブロック構造** g と**母平均・標準偏差**で標準化された観測行列
- T^*: $Z^{\mathrm{T}}Z$ の最大固有値λ_1を正規化した統計量



- $\tilde{Z}$: **正しいブロック構造** g と**標本平均・標準偏差**で標準化された観測行列
- $\tilde{\lambda}_1$: $\tilde{Z}^{\mathrm{T}}\tilde{Z}$の最大固有値



- $\hat{Z}$: **推定ブロック構造**$\hat{g}$と**標本平均・標準偏差**で標準化された観測行列
- T: $\hat{Z}^{\mathrm{T}}\hat{Z}$の最大固有値$\hat{\lambda}_1$を正規化した統計量



定理 1 の証明 XXVI

「$\tilde{Z} = \hat{Z}$ が成り立つ」という事象 $\mathcal{F}_m$ と，「$\frac{|\lambda_1 - \tilde{\lambda}_1|}{b^{\mathrm{TW}}} \leq Cm^{-\frac{1}{21}+\epsilon}$ が成り立つ」という事象 $\mathcal{G}_{m,C}$ の同時確率は以下の不等式を満たす．

$$\Pr\left(\mathcal{F}_m \cap \mathcal{G}_{m,C}\right) \geq 1 - \Pr\left(\mathcal{F}_m^{\mathrm{C}}\right) - \Pr\left(\mathcal{G}_{m,C}^{\mathrm{C}}\right). \tag{54}$$

ここで，$\mathcal{A}^{\mathrm{C}}$ は $\mathcal{A}$ の余事象を表す．
ブロック構造推定アルゴリズムの一致性の仮定より，実現可能な場合，$m \to \infty$ で $\Pr\left(\mathcal{F}_m^{\mathrm{C}}\right) \to 0$. このことと (53) より，

$$\forall \tilde{\epsilon} > 0,\ \exists C > 0, M > 0,\ \forall m \geq M,\ \Pr\left(\mathcal{F}_m \cap \mathcal{G}_{m,C}\right) \geq 1 - \tilde{\epsilon}. \tag{55}$$

であり，(52) が示せた．

定理 2 の証明 I

Theorem 6 (実現不可能な場合の lower bound)

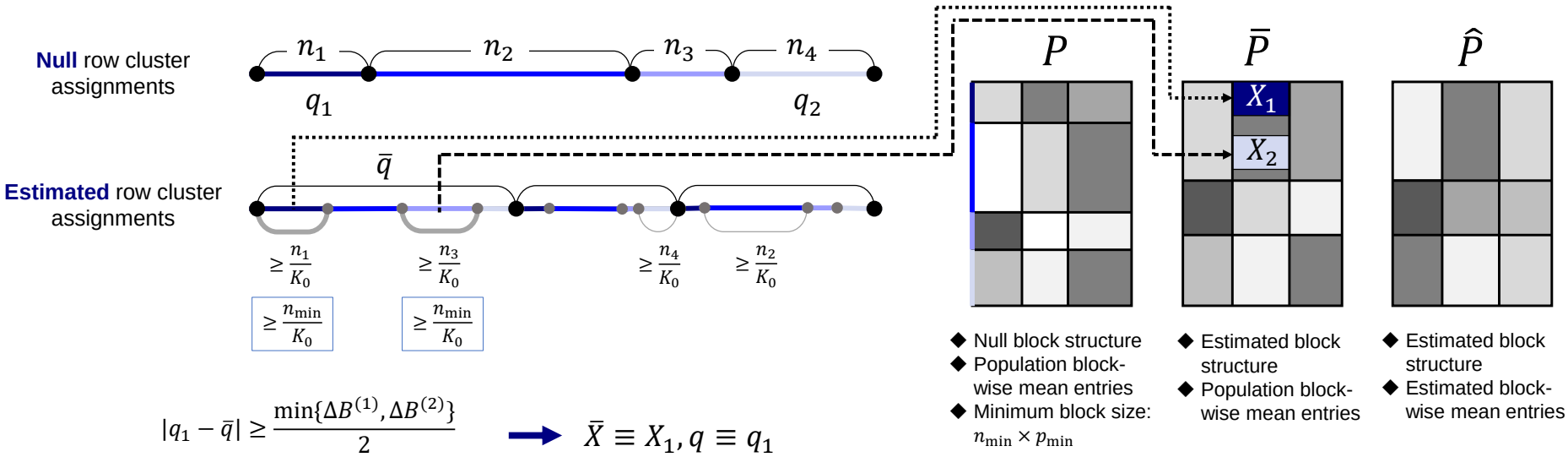
$K_0 < K$ か $H_0 < H$ の少なくとも一方が成り立つ時,

$$T = \Omega_p\left(m^{\frac{5}{3}}\right). \tag{56}$$

$\bar{P}$ を，推定ブロック構造からなり，各ブロックごとの平均（各帰無ブロックの母平均 P から計算される）を格納した行列とする．P と $\hat{P}$ の差を考えるために，まずは P と $\bar{P}$ の差について考える．$K_0 < K$ の場合を考えても一般性を失わない．k 番目の帰無行クラスタのサイズを n_k, h 番目の帰無列クラスタのサイズを p_h とする．

定理 2 の証明 II

全ての $k \in \{1, \ldots, K\}$ について，少なくとも 1 つの推定行クラスタが存在して，k 番目の帰無行クラスタに属する行を n_k/K_0 個以上含む．$K_0 < K$ なので，少なくとも 1 つの推定ブロックが存在して，「帰無行クラスタが互いに異なり，全て $n_{\min}/K_0$ 個以上の行サイズを持つ」ような行集合を 2 つ以上含む（$n_{\min}$ は最小の帰無行クラスタサイズ）．

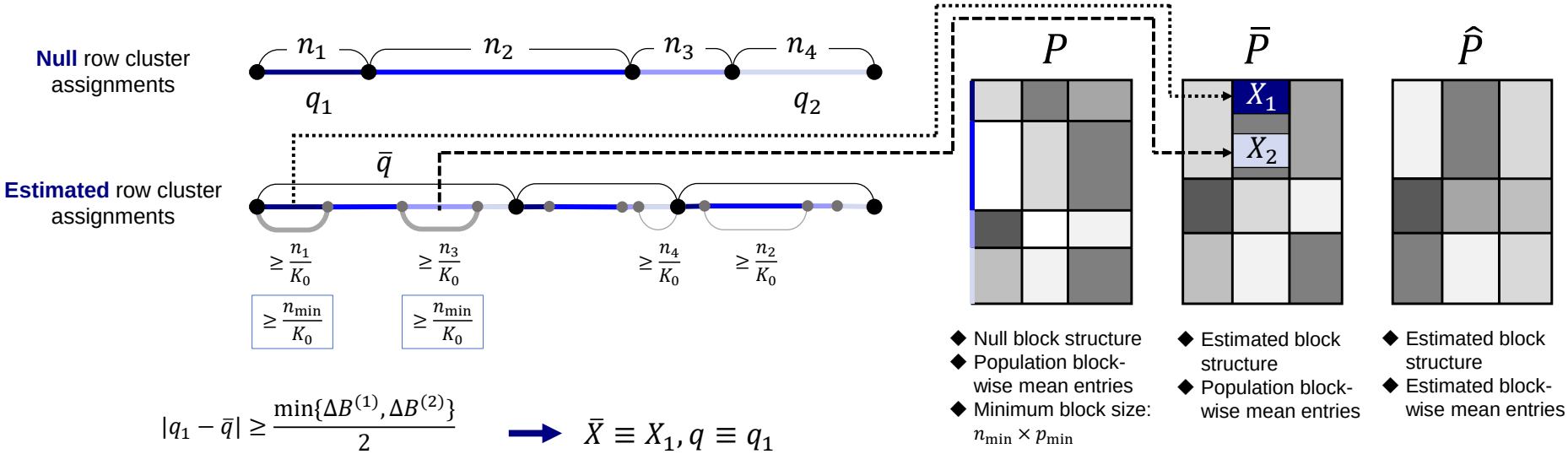


$$|q_1 - \bar{q}| \geq \frac{\min\{\Delta B^{(1)}, \Delta B^{(2)}\}}{2} \quad \rightarrow \quad \bar{X} \equiv X_1, q \equiv q_1$$

定理 2 の証明 III

同様に，全ての $h \in \{1, \ldots, H\}$ について，少なくとも 1 つの推定列クラスタが存在して，h 番目の帰無列クラスタに属する列を p_h/H_0 個以上含む．

これらにより，少なくとも 1 つの推定ブロックが存在して，「全て $(n_{\min}/K_0) \times (p_{\min}/H_0)$ 以上のサイズを持ち，互いに帰無ブロックが異なる（それぞれブロック平均 q_1, q_2）」ような 2 つ以上の部分行列 X_1, X_2 を含む．$q_1 > q_2$ を仮定して一般性を失わない．



$$|q_1 - \bar{q}| \geq \frac{\min\{\Delta B^{(1)}, \Delta B^{(2)}\}}{2} \quad \rightarrow \quad \bar{X} \equiv X_1, q \equiv q_1$$

定理 2 の証明 IV

$\bar{P}$（推定ブロック構造を持つ）は，X_1 と X_2 のどちらの部分行列においても同じ $\bar{q}$ の値を持つ．ここで，$\bar{q} \geq \frac{q_1+q_2}{2}$ ならば $|q_2 - \bar{q}| \geq \frac{|q_1-q_2|}{2}$ が，そうでなければ $|q_1 - \bar{q}| \geq \frac{|q_1-q_2|}{2}$ が成り立つ．
よって，いかなる $\bar{q}$ に対しても，少なくとも $(n_{\min}/K_0) \times (p_{\min}/H_0)$ 以上のサイズを持つ部分行列 $\bar{X}$（X_1 もしくは X_2）が少なくとも 1 つ存在して，P の中では q の値（q_1 もしくは q_2）を持ち，

$$
\begin{aligned}
|q - \bar{q}| &\geq \frac{\min\{\Delta B^{(1)}, \Delta B^{(2)}\}}{2}, \\
\Delta B^{(1)} &\equiv \min_{k,k' \in \{1,\ldots,K\}, h \in \{1,\ldots,H\}} |B_{kh} - B_{k'h}|, \\
\Delta B^{(2)} &\equiv \min_{k \in \{1,\ldots,K\}, h,h' \in \{1,\ldots,H\}} |B_{kh} - B_{kh'}|. \qquad (57)
\end{aligned}
$$

定理 2 の証明 V

$\bar{P}$ と $\hat{P}$（どちらも推定ブロック構造を持つ）の差について，仮定 $n_{\min}, p_{\min} = \Omega_p(m)$ と $\|Z\|_{\mathrm{op}} = O_p(\sqrt{m})$ から以下を導ける．

$$|\hat{q} - \bar{q}| = \cdots = O_p\left(\frac{1}{\sqrt{m}}\right). \tag{58}$$

任意の q, $\bar{q}$, $\hat{q}$ について，

$$\Big||q - \bar{q}| - |q - \hat{q}|\Big| \leq |\hat{q} - \bar{q}|. \tag{59}$$

が成り立つため，「$|q - \bar{q}| - C/\sqrt{m} \leq |q - \hat{q}|$ が成り立つ」という事象 $\mathcal{E}_{m,C}$ を考えると，(58) と (59) から，

$$\forall \epsilon > 0,\ \exists C > 0, M > 0,\ \forall m \geq M,\ \Pr(\mathcal{E}_{m,C}) \geq 1 - \epsilon. \tag{60}$$

部分行列 $\bar{X}$ のサイズを $\bar{n}_1 \times \bar{p}_1$ とおく．A, P, $\bar{P}$, $\hat{P}$, Z, $\hat{Z}$ の $\bar{X}$ に当たる部分行列をそれぞれ A^*, P^*, $\bar{P}^*$, $\hat{P}^*$, Z^*, $\hat{Z}^*$ とおく．また，σ, $\hat{\sigma}$ の $\bar{X}$ に当た

定理 2 の証明 VI

る部分行列の要素（定数）をそれぞれ $\sigma^*, \hat{\sigma}^*$ とおく．部分行列の作用素ノルムは元の行列の作用素ノルム以下であることと，定義 (8) から，以下を導ける．

$$\|\hat{Z}\|_{\mathrm{op}} \geq \|\hat{Z}^*\|_{\mathrm{op}} = \cdots = \frac{1}{\hat{\sigma}^*} \left| \sigma^* \|Z^*\|_{\mathrm{op}} - \|P^* - \hat{P}^*\|_{\mathrm{op}} \right|. \tag{61}$$

ここで，

$$\hat{\sigma}^* = O_p(1). \tag{62}$$

が成り立つ（証明略）．
定数行列 $\left(P^* - \hat{P}^*\right)$ の唯一の非ゼロの特異値（i.e., 最大特異値）は $\sqrt{\bar{n}_1 \bar{p}_1}\,|q - \hat{q}|$ であるため，

$$\|P^* - \hat{P}^*\|_{\mathrm{op}} = \sqrt{\bar{n}_1 \bar{p}_1}\,|q - \hat{q}| \geq \sqrt{\frac{n_{\min}}{K_0} \frac{p_{\min}}{H_0}}\,|q - \hat{q}|. \tag{63}$$

定理 2 の証明 VII

よって，(57) で $|q-\bar{q}| \geq \frac{\min\{\Delta B^{(1)}, \Delta B^{(2)}\}}{2}$ だったので，
「$|q-\bar{q}| - C/\sqrt{m} \leq |q-\hat{q}|$ が成り立つ」という事象 $\mathcal{E}_{m,C}$ が起こった時，
以下も成り立つことに注意する．

$$\sqrt{\frac{n_{\min}}{K_0}\frac{p_{\min}}{H_0}}\left(\frac{\min\{\Delta B^{(1)}, \Delta B^{(2)}\}}{2} - \frac{C}{\sqrt{m}}\right) \leq \|P^* - \hat{P}^*\|_{\mathrm{op}}, \tag{64}$$

(60) で $\forall \epsilon > 0,\ \exists C > 0, M > 0,\ \forall m \geq M,\ \Pr(\mathcal{E}_{m,C}) \geq 1-\epsilon$ が成り立っていたことを思い出すと，$\|P^* - \hat{P}^*\|_{\mathrm{op}} = \Omega_p(m)$.
さらに，$\|Z^*\|_{\mathrm{op}} \leq \|Z\|_{\mathrm{op}} = O_p(\sqrt{m})$ であることと，(62), (64) を (61) に代入して，

$$\hat{\lambda}_1 = \|\hat{Z}\|_{\mathrm{op}}^2 = \Omega_p\left(m^2\right). \tag{65}$$

定理 2 の証明 VIII

ここで，a^{TW}, b^{TW} の定義から，

$$a^{\mathrm{TW}} = \Theta(m), \quad b^{\mathrm{TW}} = \Theta\left(m^{\frac{1}{3}}\right). \tag{66}$$

これらと定義 $T = \frac{\hat{\lambda}_1 - a^{\mathrm{TW}}}{b^{\mathrm{TW}}}$, (65) より，

$$T = \Omega_p\left(m^{\frac{5}{3}}\right). \tag{67}$$

が示せた.

定理 2 の証明 IX

Theorem 7 (実現不可能な場合の upper bound)

$K_0 < K$ か $H_0 < H$ の少なくとも一方が成り立つ時,

$$T = O_p\left(m^{\frac{5}{3}}\right). \tag{68}$$

行列の作用素ノルムはその全ての部分行列の作用素ノルムの和以下になることを用いて，以下を導ける.

$$\|\hat{Z}\|_{\mathrm{op}} \leq \cdots \leq K_0 H_0 \sqrt{np} = O_p(m). \tag{69}$$

定義 $T = \frac{\hat{\lambda}_1 - a^{\mathrm{TW}}}{b^{\mathrm{TW}}}$, $\hat{\lambda}_1 = \|\hat{Z}\|_{\mathrm{op}}^2 = O_p(m^2)$, (66) で $a^{\mathrm{TW}} = \Theta(m)$, $b^{\mathrm{TW}} = \Theta\left(m^{\frac{1}{3}}\right)$ であったことから，$T = O_p(m^2/m^{\frac{1}{3}}) = O_p(m^{\frac{5}{3}})$ が示せた.

主結果の証明で用いた主なランダム行列理論の結果

- [12] の結果：平均 0, 分散 1 で sub-exponential decay を持つ（e.g., ある $\vartheta > 0$ が存在して $x > 1$ について $\Pr(|Z_{ij}| > x) \leq \vartheta^{-1} \exp(-x^{\vartheta})$ を満たす）行列 $Z \in \mathbb{R}^{n \times p}$ について，$Z^\top Z$ の最大固有値を λ_1 とする．

$$T^* = \frac{\lambda_1 - a^{\mathrm{TW}}}{b^{\mathrm{TW}}}, \ T^* \rightsquigarrow TW_1 \ (\text{Convergence in law}), \ m \to \infty. \tag{70}$$

ただし，$a^{\mathrm{TW}} = (\sqrt{n} + \sqrt{p})^2$, $b^{\mathrm{TW}} = (\sqrt{n} + \sqrt{p}) \left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p}} \right)^{\frac{1}{3}}$.

- [3] の定理 2.17, $Z^\top Z$ の固有ベクトル $\boldsymbol{v}_j$ $(\|\boldsymbol{v}_j\| = 1)$ の delocalization property: 任意の $\tilde{d} \in \mathbb{N}$ と，$\|\boldsymbol{w}^{(i)}\| = 1$ を満たす任意の定数ベクトル $\{\boldsymbol{w}^{(i)}\}$ $(i = 1, \ldots, m^{\tilde{d}})$ に対し，

$$\max_{i \in \{1, \ldots, m^{\tilde{d}}\}} \max_{j \in \{1, \ldots, p\}} |\boldsymbol{v}_j^\top \boldsymbol{w}^{(i)}| = O_p\left(m^{-\frac{1}{2} + \epsilon} \right), \ \text{for all } \epsilon > 0. \tag{71}$$

まとめ：提案する適合度検定の流れ

- 帰無仮説：$(K, H) = (K_0, H_0)$（観測データは $K_0 \times H_0$ 個のブロックからなる）
- 対立仮説：$K > K_0$ または $H > H_0$（行・列の少なくともどちらかでクラスタ数が足りない）
- 各仮説クラスタ数 (K_0, H_0) について検定統計量 T を計算し，$T \geq t(\alpha)$ ならば帰無仮説を棄却（$t(\alpha)$ は TW_1 分布の α upper quantile）
- 以下の順番で検定し，帰無仮説受容時の設定 $(\hat{K}, \hat{H})$ を選択結果とする
  1. $(K_0, H_0) = (1, 1)$,
  2. $(K_0, H_0) = (1, 2), (2, 1)$,
  3. $(K_0, H_0) = (1, 3), (2, 2), (3, 1), \cdots$

目次

実験 (1) 検定統計量 T の TW_1 分布への法則収束 I

- 実現可能な設定において，検定統計量 T が TW_1 分布に法則収束することを確認
- 実験設定
  - $(K, H) = (4, 3)$ の真のクラスタ構造を持つ人工データ
  - Gaussian LBM (G-LBM): 各ブロックの要素は平均 $B = \begin{pmatrix} 0.9 & 0.1 & 0.4 \\ 0.2 & 0.7 & 0.3 \\ 0.3 & 0.2 & 0.8 \\ 0.6 & 0.9 & 0.1 \end{pmatrix}$, 標準偏差 $S = \begin{pmatrix} 0.08 & 0.06 & 0.15 \\ 0.14 & 0.12 & 0.07 \\ 0.09 & 0.1 & 0.11 \\ 0.16 & 0.13 & 0.05 \end{pmatrix}$ の正規分布に従う
  - Bernoulli LBM (B-LBM): 各ブロックの要素は平均 $B = \begin{pmatrix} 0.9 & 0.1 & 0.4 \\ 0.2 & 0.7 & 0.3 \\ 0.3 & 0.2 & 0.8 \\ 0.6 & 0.9 & 0.1 \end{pmatrix}$ の Bernoulli 分布に従う

実験 (1) 検定統計量 T の TW_1 分布への法則収束 II

- Poisson LBM (P-LBM): 各ブロックの要素は平均
$$B = \begin{pmatrix} 9.0 & 1.0 & 4.0 \\ 2.0 & 7.0 & 3.0 \\ 3.0 & 2.0 & 8.0 \\ 6.0 & 9.0 & 1.0 \end{pmatrix}$$
の Poisson 分布に従う

- 行列サイズ：$(n, p) = (300 \times i, 225 \times i),\ i = 1, \ldots, 10.$
- 各行・列の正解クラスタはそれぞれ $\{1, 2, 3, 4\}$ 上・$\{1, 2, 3\}$ 上の一様分布に基づき定義
- 観測データを 1000 個生成．それぞれに対し階層的クラスタリング [16] によりクラスタ構造を推定，検定統計量 T を計算

検定統計量 T の裾確率の漸近的なふるまい

- 3 種類の有意水準 α に対し，検定統計量 T が $T \geq t(\alpha)$ を満たした試行回数の割合をプロット（ただし，$t(\alpha)$ は TW_1 分布の α upper quantile）
- 漸近的に TW_1 分布の裾確率に収束

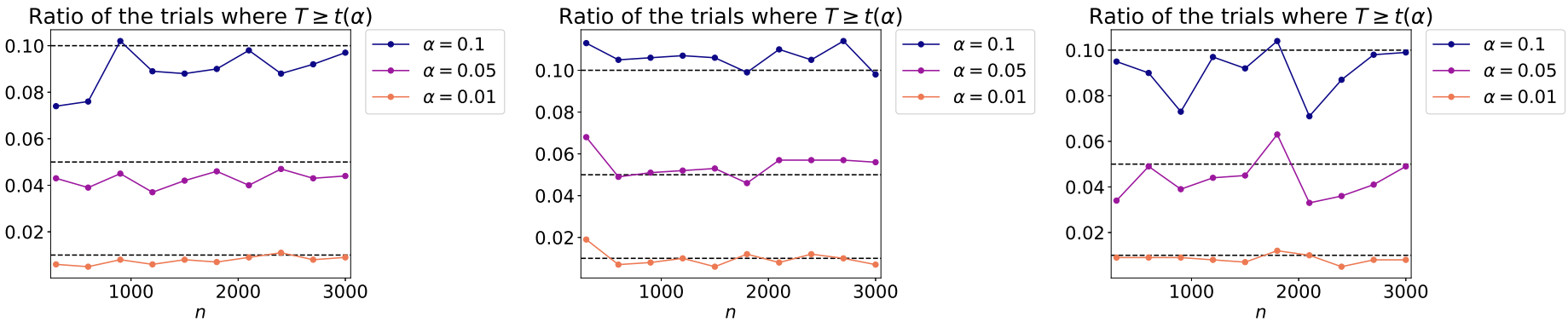


Figure: $T \geq t(\alpha)$ を満たした試行回数の割合．左/中央/右の図はそれぞれ G/B/P-LBM の結果を表す．横軸は行サイズ n を表す．

Kolmogorov-Smirnov 検定 [5]

- 検定統計量 T が従う分布が TW_1 分布かどうかを検定
- T の経験分布と TW_1 分布の累積密度関数の差の絶対値の最大値を D として，$D\sqrt{r}$ の値に基づき検定を行う（r：標本数）

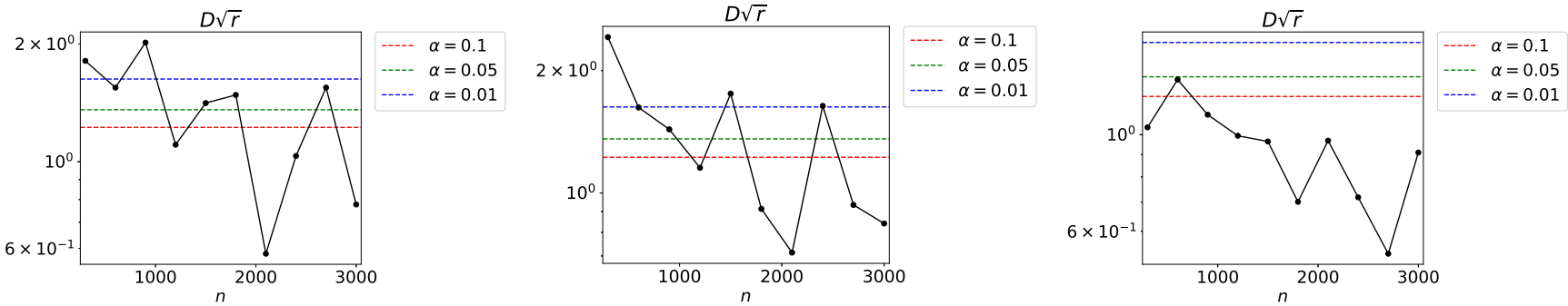


Figure: Kolmogorov-Smirnov 検定の検定統計量 $D\sqrt{r}$．左/中央/右の図はそれぞれ G/B/P-LBM の結果を表す．横軸は行サイズ n を表す．検定統計量の値が破線より上側ならば T が TW_1 分布に従うという帰無仮説を棄却，そうでなければ帰無仮説を受容．

実験 (2) 実現不可能な場合の検定統計量 T のふるまい

- 実現不可能な場合，検定統計量 T が $T = \Theta_p\left(m^{\frac{5}{3}}\right)$ を満たすことを確認
- 実験設定
  - $(K, H) = (4, 3)$ の真のクラスタ構造を持つ人工データ
  - ブロックごとの平均 B, 標準偏差 S と正解クラスタの定義は実験 (1) と同じ
  - 行列サイズ：$(n, p) = (200 \times i, 150 \times i)$, $i = 1, \ldots, 10$.
  - 観測データを 100 個生成．それぞれに対し階層的クラスタリング [16] によりクラスタ構造を推定，検定統計量 T を計算

実現不可能な場合の T の漸近的なふるまい

- 実現不可能な設定において，検定統計量 T は漸近的に $n^{\frac{5}{3}}$ に比例して増加（n：行サイズ．列サイズは $p = \frac{3}{4}n$）
- T の平均値に $n^{-\frac{5}{3}}$ を乗じた値をプロット
- 異なる線は仮説として設定するクラスタ数の違いに対応

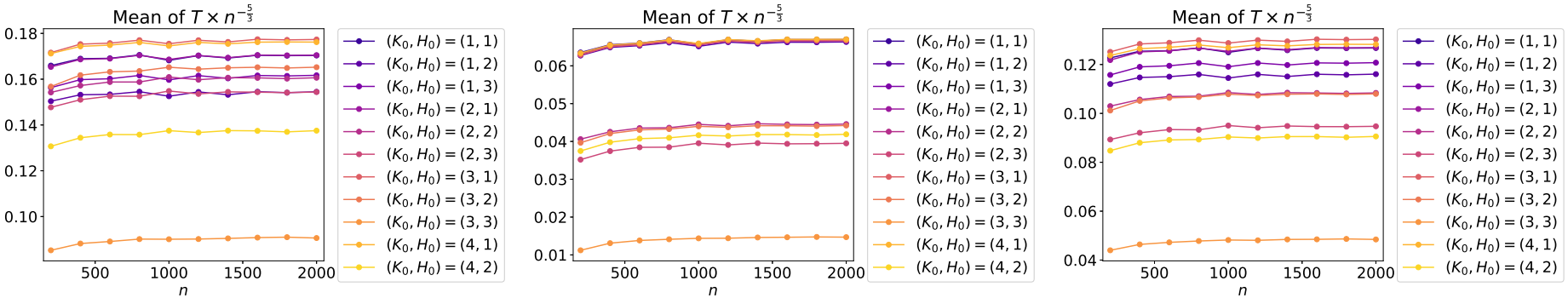


Figure: T の平均値に $n^{-\frac{5}{3}}$ を乗じた値．左/中央/右の図はそれぞれ G/B/P-LBM の結果を表す．横軸は行サイズ n を表す．

実験 (3) 提案手法によるクラスタ数推定の精度 I

- 提案手法を用いた際に，選択されたクラスタ数が真のクラスタ数と一致した試行回数の割合を確認
- 実験設定
  - $(K, H) = (4, 3)$ の真のクラスタ構造を持つ人工データ
  - 正解クラスタの定義は実験 (1) と同じ
  - G-LBM:

$$B' = \begin{pmatrix} 0.9 & 0.1 & 0.4 \\ 0.2 & 0.7 & 0.3 \\ 0.3 & 0.2 & 0.8 \\ 0.6 & 0.9 & 0.1 \end{pmatrix}, \quad S = \begin{pmatrix} 0.08 & 0.06 & 0.15 \\ 0.14 & 0.12 & 0.07 \\ 0.09 & 0.1 & 0.11 \\ 0.16 & 0.13 & 0.05 \end{pmatrix},$$

$$\forall k, h,\ B_{kh} = \left(1 - \frac{t}{10}\right)(B'_{kh} - 0.5) + 0.5, \quad \text{for } t = 0, \ldots, 9, \qquad (72)$$

実験 (3) 提案手法によるクラスタ数推定の精度 II

- B-LBM:

$$B' = \begin{pmatrix} 0.9 & 0.1 & 0.4 \\ 0.2 & 0.7 & 0.3 \\ 0.3 & 0.2 & 0.8 \\ 0.6 & 0.9 & 0.1 \end{pmatrix},$$

$$\forall k, h,\ B_{kh} = \left(1 - \frac{t}{10}\right)(B'_{kh} - 0.5) + 0.5, \quad \text{for } t = 0, \ldots, 9. \tag{73}$$

- P-LBM:

$$B' = \begin{pmatrix} 9.0 & 1.0 & 4.0 \\ 2.0 & 7.0 & 3.0 \\ 3.0 & 2.0 & 8.0 \\ 6.0 & 9.0 & 1.0 \end{pmatrix},$$

$$\forall k, h,\ B_{kh} = \left(1 - \frac{t}{10}\right)(B'_{kh} - 5) + 5, \quad \text{for } t = 0, \ldots, 9. \tag{74}$$

- 行列サイズ：$(n, p) = (40 \times i, 30 \times i),\ i = 1, \ldots, 10.$

実験 (3) 提案手法によるクラスタ数推定の精度 III

- 上記の各設定に対し，観測データを 1000 個生成．それぞれに対し階層的クラスタリング [16] によりクラスタ構造を推定，有意水準 $\alpha = 0.01$ として帰無仮説が受容されるまでクラスタ数を検定

人工的に生成した観測データの例 (G-LBM)

- 10 種類のブロック平均 B の設定における観測データの例（行列サイズは全て $(40, 30)$）
- t が大きい（異なるブロックにおける真の平均値の差が小さい）ほど，クラスタ数の検定が困難になる

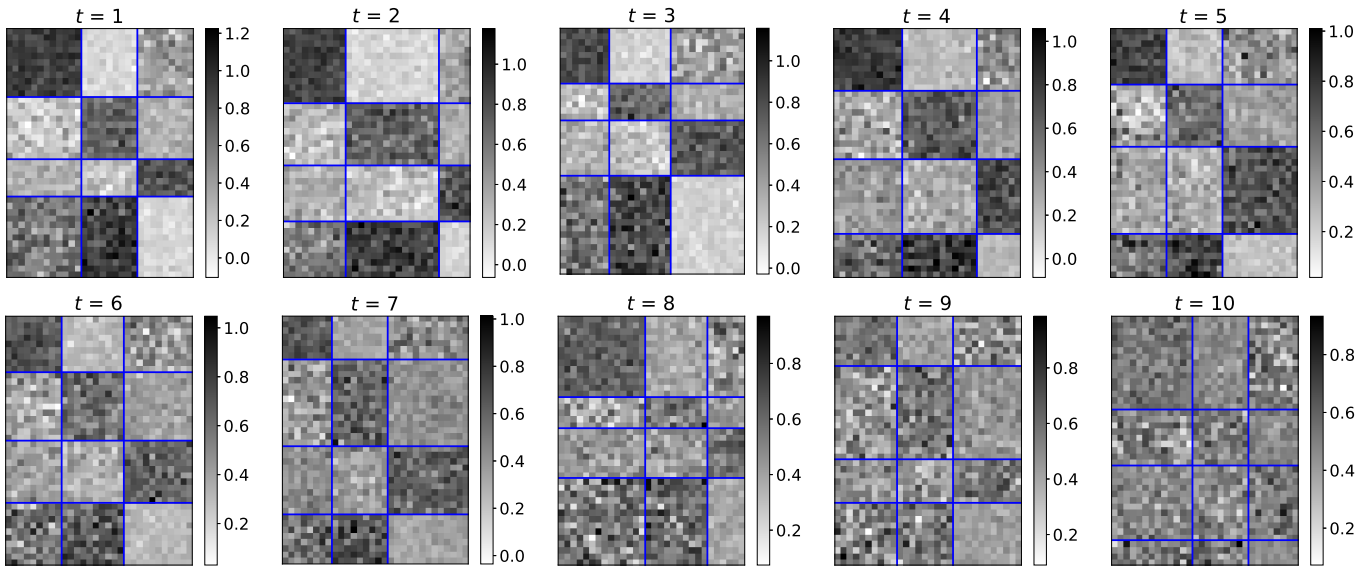


Figure: 異なる B の設定 $(t = 1, \ldots, 10)$ において，G-LBM から生成された観測行列の例．行と列の番号はクラスタごとにソートしてある．

人工的に生成した観測データの例 (B-LBM)

- 10 種類のブロック平均 B の設定における観測データの例（行列サイズは全て $(40, 30)$）
- t が大きい（異なるブロックにおける真の平均値の差が小さい）ほど，クラスタ数の検定が困難になる

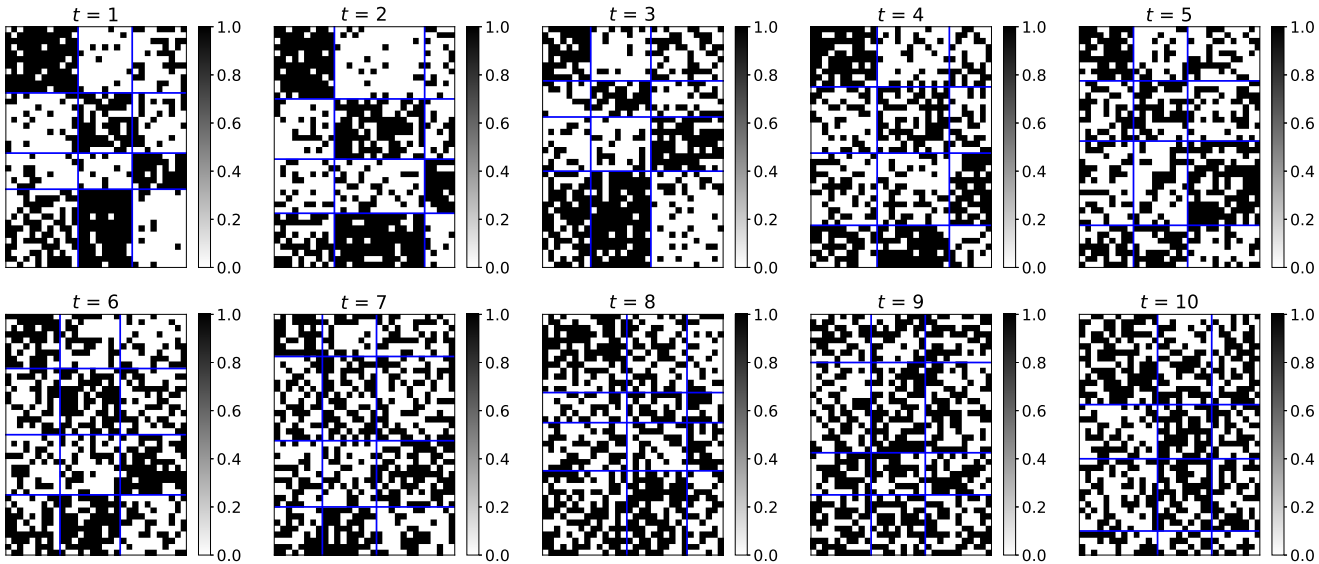


Figure: 異なる B の設定 $(t = 1, \ldots, 10)$ において，B-LBM から生成された観測行列の例．行と列の番号はクラスタごとにソートしてある．

人工的に生成した観測データの例 (P-LBM)

- 10 種類のブロック平均 B の設定における観測データの例（行列サイズは全て $(40, 30)$）
- t が大きい（異なるブロックにおける真の平均値の差が小さい）ほど，クラスタ数の検定が困難になる

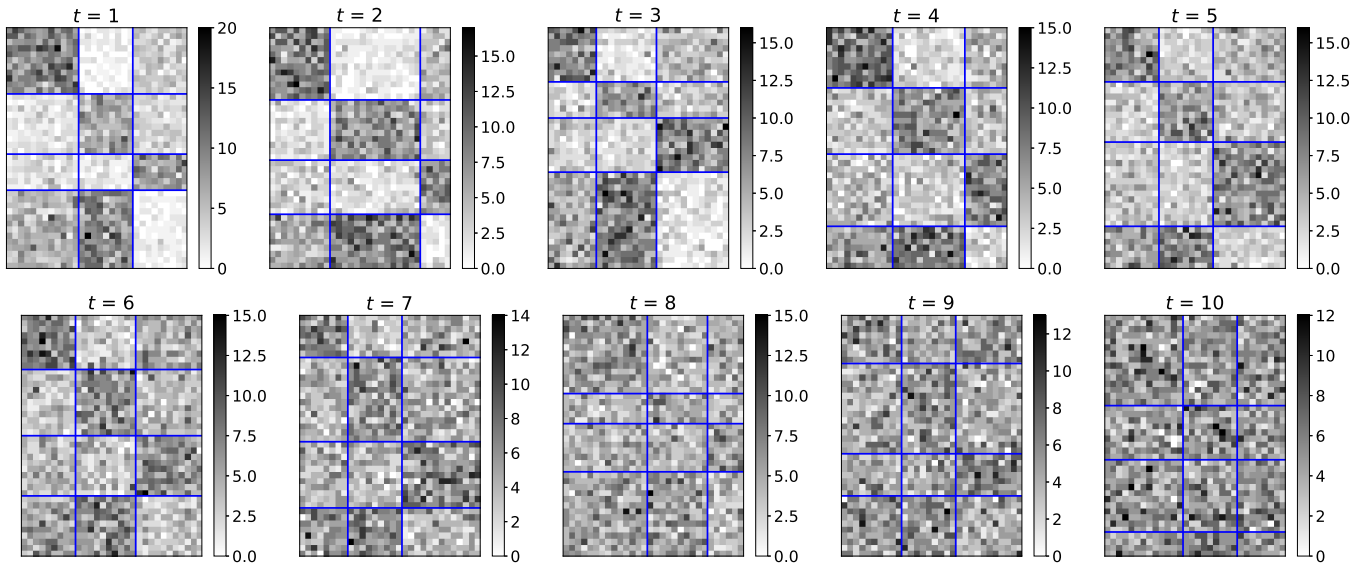


Figure: 異なる B の設定 $(t = 1, \ldots, 10)$ において，P-LBM から生成された観測行列の例．行と列の番号はクラスタごとにソートしてある．

提案手法によるクラスタ数の検定の精度

- 異なる行列サイズ・ブロック平均 B に対し，選択されたクラスタ数が真のクラスタ数と一致した試行回数の割合をプロット
- 行列サイズ n が大きいほどクラスタ数を当てやすい
- ブロック平均 B_{kh} 間の差が大きいほどクラスタ数を当てやすい

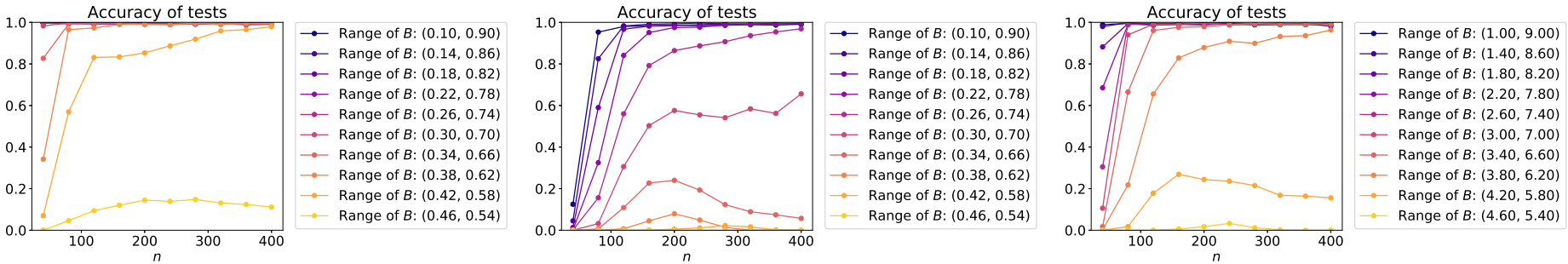


Figure: 選択されたクラスタ数が真のクラスタ数と一致した試行回数の割合．左/中央/右の図はそれぞれ G/B/P-LBM の結果を表す．横軸は行サイズ n を表す．

実験 (4) 実データへの適用

- 1984 United States Congressional Voting Records Database, UCI Machine Learning Repository [7] を用いた
- 各行が連邦議会議員，各列がその特徴を表す（435×16 の行列）
- 各要素は “yea,” “nay,” and unknown のうちいずれか． “yea” を 1, それ以外を 0 として解析
- 階層的クラスタリング [16] によりクラスタ構造を推定，有意水準 $\alpha = 0.01$ として帰無仮説が受容されるまでクラスタ数を検定

実験結果（$(\hat{K}, \hat{H}) = (9, 14)$ で帰無仮説受容）

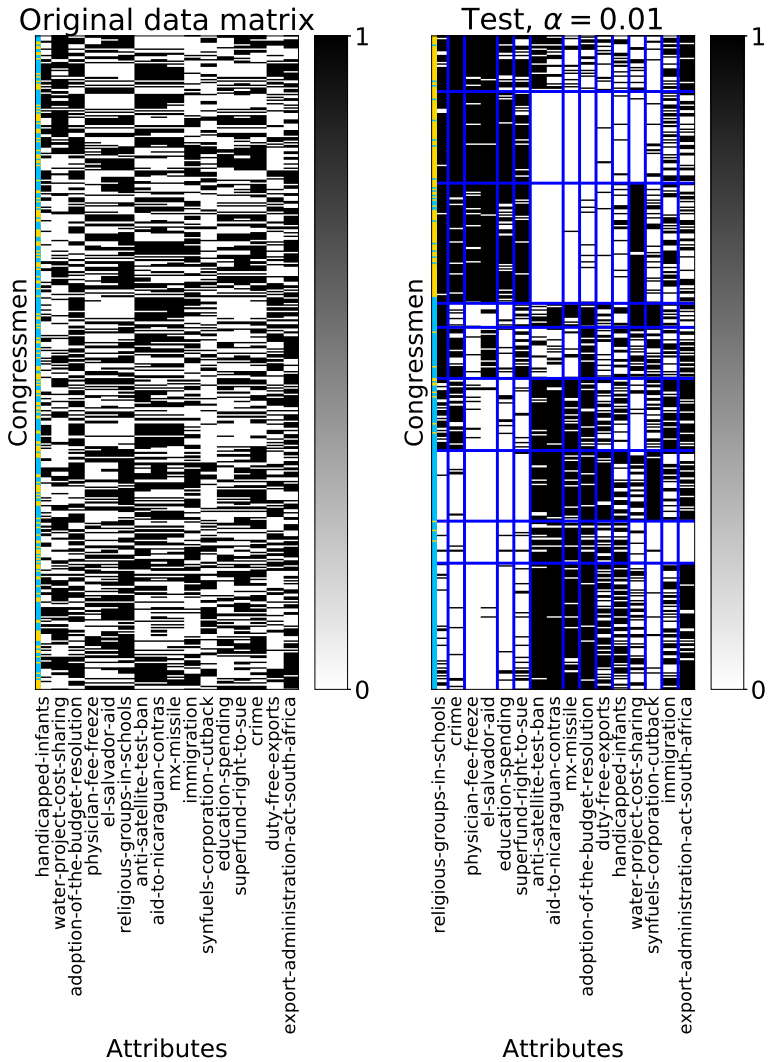


Figure: 左：観測データ行列．右：帰無仮説受容時の推定クラスタ構造．

目次

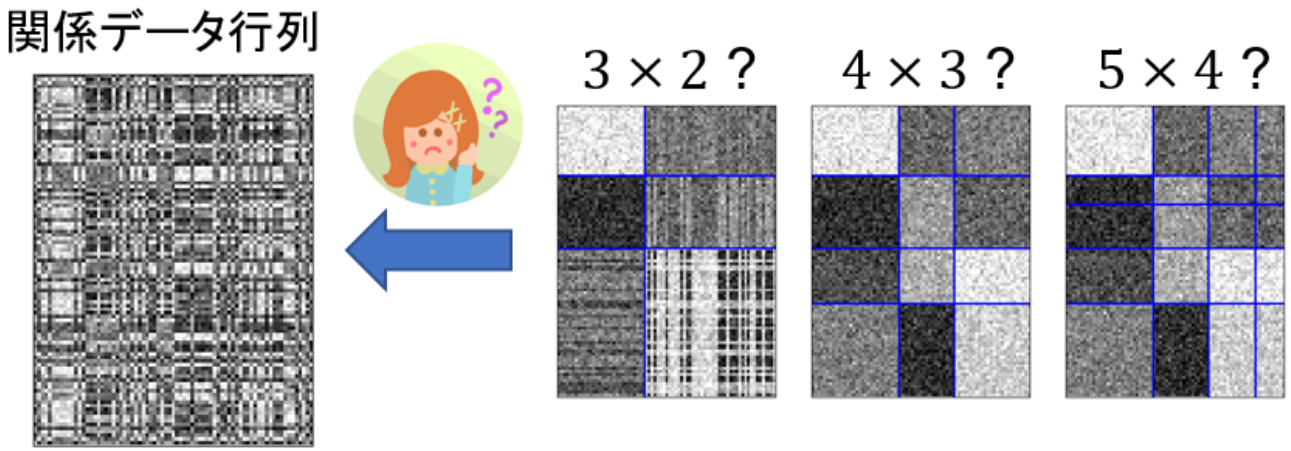
考察

- より良い検定の構築は可能か？
  - 対立仮説が複数のモデルを含む本問題設定においては，最強力検定は存在しない．しかし，有限の大きさの行列に対して収束率の意味でより良い検定は存在しうる．
  - 今回構成した検定統計量の TW_1 分布への収束率は不明
- 問題設定の拡張
  - 格子状のブロック構造で表されるとは限らないケース・行列サイズに応じてクラスタ数が増加する設定 [17]
  - ブロック数ではなくブロック構造についての検定 [18]

目次

まとめ

- 与えられた関係データ行列を表現するのに適切なクラスタ数を，統計的検定に基づいて決める手法を提案
  - 一致性を満たすクラスタリングアルゴリズムにより得られたクラスタリング結果に基づき，各成分を標準化した行列を $\hat{Z}$ とし，$\hat{Z}^\top \hat{Z}$ の最大固有値 $\hat{\lambda}_1$ から検定統計量 T を定義
  - 実現可能な場合，検定統計量 T が行列サイズ $m \to \infty$ の極限で TW_1 分布に法則収束することを利用



参考文献 I

[1] B. P. W. Ames. Guaranteed clustering and biclustering via semidefinite programming. *Mathematical Programming*, 147(1):429–465, 2014.

[2] P. J. Bickel and P. Sarkar. Hypothesis testing for automated community detection in networks. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(1):253–273, 2016.

[3] A. Bloemendal, A. Knowles, H.-T. Yau, and J. Yin. On the principal components of sample covariance matrices. *Probability Theory and Related Fields*, 164:459–552, 2016.

[4] V. Brault and A. Channarond. Fast and consistent algorithm for the latent block model. arXiv:1610.09005, 2016.

[5] W. J. Conover. *Practical Nonparametric Statistics*. John Wiley & Sons, New York, 1999.

[6] I. S. Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 269–274, 2001.

[7] D. Dua and C. Graff. UCI machine learning repository. `http://archive.ics.uci.edu/ml`, 2017. University of California, Irvine, School of Information and Computer Sciences.

[8] C. J. Flynn and P. O. Perry. Profile likelihood biclustering. *Electronic Journal of Statistics*, 14(1):731–768, 2020.

参考文献 II

[9] J. Hu, J. Zhang, H. Qin, T. Yan, and J. Zhu. Using maximum entry-wise deviation to test the goodness of fit for stochastic block models. *Journal of the American Statistical Association*, 0(0):1–10, 2020.

[10] C. Keribin, V. Brault, G. Celeux, and G. Govaert. Estimation and selection for the latent block model on categorical data. *Statistics and Computing*, 25:1201–1216, 2015.

[11] J. Lei. A goodness-of-fit test for stochastic block models. *The Annals of Statistics*, 44(1):401–424, 2016.

[12] N. S. Pillai and J. Yin. Universality of covariance matrices. *Annals of Applied Probability*, 24(3):935–1001, 2014.

[13] H. Shan and A. Banerjee. Bayesian co-clustering. In *Proceedings of the 8th IEEE International Conference on Data Mining*, pages 530–539, 2008.

[14] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

[15] J. D. Vlok and J. C. Olivier. Analytic approximation to the largest eigenvalue distribution of a white Wishart matrix. *IET Communications*, 6(12):1804–1811, 2012.

[16] J. H. Ward, Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.

参考文献 III

[17] C. Watanabe and T. Suzuki. A goodness-of-fit test on the number of biclusters in a relational data matrix. arXiv:2102.11658, 2021.

[18] C. Watanabe and T. Suzuki. Selective inference for latent block models. *Electronic Journal of Statistics*, 15(1):3137–3183, 2021.