

パラメトリック計画法を用いた 凸クラスタリングのための選択的推論

第 42 回 IBISML 研究会

大森夢拓[†], 稲津佑[†], 竹内一郎^{†,††}

†: 名古屋工業大学

††: 理化学研究所

2021/3/2

目次

- 1 はじめに
- 2 凸クラスタリング
- 3 parametric Selective Inference
- 4 計算機実験
- 5 まとめと今後の課題

- 1 はじめに
- 2 凸クラスタリング
- 3 parametric Selective Inference
- 4 計算機実験
- 5 まとめと今後の課題

クラスタリングの信頼性

クラスタリングは幅広い分野において、パターン認識に用いられる

※ 実際に得られた結果が必ずしも真のクラスタ構造とは限らない

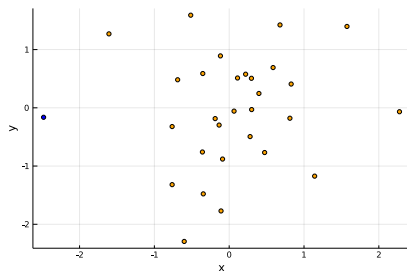
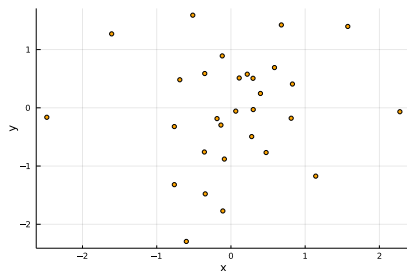


Figure: i.i.d. である標本集合 (左) に, 凸クラスタリングを適用 (右)

クラスタリングの結果が意図する結果かどうか判別したい

→ クラスタリングで得られるクラスタ間の統計的有意性の評価を行う

統計的有意性の評価

観測データ $\mathbf{X}_{obs} \in \mathbb{R}^{n \times p}$ に凸クラスタリング (Hocking et al. 2011 [2]) を適用

- n : データ数, p : 特徴数

検定問題

任意のクラス間における任意の特徴量の母平均の差の検定

$$H_0 : \mu_{C_a,k} = \mu_{C_b,k} \quad \text{vs.} \quad H_1 : \mu_{C_a,k} \neq \mu_{C_b,k}$$

- $\mu_{C_a,k}, \mu_{C_b,k}$ はクラス C_a, C_b における k 番目の特徴量の母平均

検定統計量

$$\tau = \bar{x}_{C_a,k} - \bar{x}_{C_b,k} = \boldsymbol{\eta}^\top \text{vec}(\mathbf{X}_{obs})$$

- $\bar{x}_{C_a,k}, \bar{x}_{C_b,k}$ はクラス C_a, C_b における k 番目の特徴量の標本平均

クラスタリングの統計的評価における問題点

古典的な統計手法では**選択バイアス**の発生を考慮できていない

- 選択バイアス: アルゴリズムに基づいた選択によって生じるバイアス

クラスタリングにおける選択バイアスの問題

クラスタ間の差が大きくなるようにクラスタリングされるバイアスが発生

→ 検定統計量の分布が一般的な分布 (正規分布など) に従わない

→ FPR を有意水準以下に制御できず検定としての 妥当性を失う

- FPR: 帰無仮説が正しいときに、帰無仮説を誤って棄却する割合



提案手法

parametric Selective Inference (pSI) (Duy et al. 2020 [1]) の枠組みを適用

- 1 はじめに
- 2 凸クラスタリング
- 3 parametric Selective Inference
- 4 計算機実験
- 5 まとめと今後の課題

凸クラスタリングでは、次の最適化問題を各特徴 (次元) 毎に解く

凸クラスタリングの最適化問題

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{x} - \beta\|_2^2 + \lambda \sum_{i < j} \|\beta_i - \beta_j\|_1$$

→ 一般化 lasso で定式化可能

- \mathbf{x} : X の特徴ベクトル
- β_i : x_i が所属するクラスタの中心
- λ : 正則化パラメータ
 - クラスタリングの強さを調整する役割
- クラスタ内のメンバー同士が $\beta_i = \beta_j$ を満たすようにクラスタリング

一般化 lasso

行列 D を用いて様々な正則化項を一般化したもの

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{x} - \beta\|_2^2 + \lambda \|D\beta\|_1$$

一般化 lasso は双対問題で最適化できることが知られている

双対問題

$$\hat{\beta} = \mathbf{x} - D^\top \hat{\mathbf{u}}$$

$$\min_{\mathbf{u} \in \mathbb{R}^{\binom{n}{2} \times n}} \frac{1}{2} \|\mathbf{x} - D^\top \mathbf{u}\|_2^2 \quad \text{subject to } \|\mathbf{u}\|_\infty \leq \lambda$$

- 凸クラスタリングにおける D
 - 各行が $1 \leq i < j \leq n$ における任意の組み合わせ (i, j) に対して, i 列目が 1, j 列目が -1 となるように設定

$$D \in \mathbb{R}^{\binom{n}{2} \times n} = \begin{pmatrix} 1 & -1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & \dots & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & -1 \end{pmatrix}$$

- \mathcal{A} : アクティブ集合
 - $\beta_i \neq \beta_j$ となる (i, j) に対応する D の行インデックスの集合

- 1 はじめに
- 2 凸クラスタリング
- 3 parametric Selective Inference
- 4 計算機実験
- 5 まとめと今後の課題

Selective Inference (SI) (Lee et al. 2016[3])

選択されたモデルで条件付けた推論

- 選択イベント: モデル選択の条件付けに必要なイベント
- 選択イベントの条件付けにより適切な帰無仮説の分布 (帰無分布) を導出
 - **【注意】** 選択されたモデルを直接条件付けしているわけではない

→ 選択バイアスの問題に対処可能

凸クラスタリングにおける SI では

- 選択されたモデル: X_{obs} から得られた凸クラスタリング結果
- 選択イベント: 凸クラスタリングの最適化問題から得られた \mathcal{A}
- 帰無分布: ベクトル η 上の正規分布 $N(0, \eta^\top \Sigma \eta)$ を, 選択イベントによって得られる区間 \mathcal{Z} で切断した切断正規分布 $TN(0, \eta^\top \Sigma \eta, \mathcal{Z})$

選択されたモデルに対して, 選択イベントによる条件付けが必要十分でない

検定統計量	A	クラスタリング結果
-2.0	$(u_1, u_3, u_4, u_5)^*$	$(C_1, C_2, C_3, C_3)^*$
-2.61 ~ -2.54	$(u_1, u_2, u_3, u_4, u_5, u_6)$	(C_1, C_2, C_3, C_4)
-2.54 ~ -2.37	$(u_1, u_3, u_4, u_5, u_6)$	$(C_1, C_2, C_3, C_3)^*$
-2.37 ~ -1.61	$(u_1, u_3, u_4, u_5)^*$	$(C_1, C_2, C_3, C_3)^*$
-1.61 ~ -1.44	(u_1, u_4, u_5)	$(C_1, C_2, C_3, C_3)^*$
-1.44 ~ -1.37	(u_1, u_2, u_4, u_5)	(C_1, C_2, C_3, C_4)
-1.37 ~ -0.20	$(u_1, u_2, u_4, u_5, u_6)$	(C_1, C_2, C_3, C_4)

Table: A とクラスタリング結果の対応表

- 選択されたモデルの条件付けによって得られる区間: -2.54 ~ -1.44
- 選択イベントの条件付けによって得られる区間: -2.37 ~ -1.61
 - SI が導出する区間

→ 選択されたモデルに対して選択イベントによる条件付けが過剰になる
 → **検出力低下**の原因となる！

pSI: parametric 計画法により、適切な帰無分布を過剰な条件付けなく導出する SI

- X をパラメータ z を用いて $\text{vec}(X(z)) = a + bz$ と表現
 - $a = (I - b\eta^\top)\text{vec}(X_{obs})$
 - $b = \Sigma\eta(\eta^\top\Sigma\eta)^{-1}$
 - Σ : $\text{vec}(X_{obs})$ の分散共分散行列
- z 軸上において、選択されたモデルの区間 \mathcal{Z} から帰無分布を構築
- 帰無分布から selective- p を以下のように構築

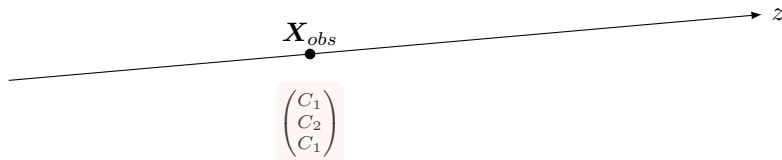
$$\text{selective-}p = 2 \min\{F_{0,\eta\Sigma\eta}^{\mathcal{Z}}(\eta^\top \text{vec}(X_{obs})), 1 - F_{0,\eta\Sigma\eta}^{\mathcal{Z}}(\eta^\top \text{vec}(X_{obs}))\}$$

- $F_{\mu,\sigma^2}^{\mathcal{Z}}(x) : \text{TN}(\mu, \sigma^2, \mathcal{Z})$ の累積分布関数

凸クラスタリングにおける parametric Selective Inference

例として, $X_{obs} \in \mathbb{R}^{3 \times 1}$, クラスタリング結果を $(C_1, C_2, C_1)^\top$ とする

- X_{obs} のクラスタリング結果と $X(z)$ のクラスタリング結果が同じとなるような z 軸上の区間から切断正規分布を構築

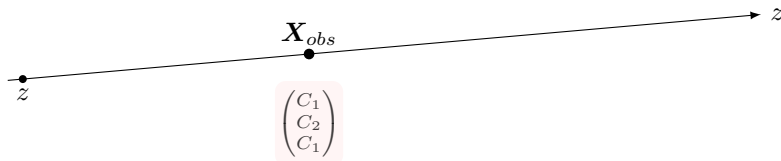


この切断正規分布の累積分布関数から selective- p 値を計算

凸クラスタリングにおける parametric Selective Inference

例として, $X_{obs} \in \mathbb{R}^{3 \times 1}$, クラスタリング結果を $(C_1, C_2, C_1)^\top$ とする

- X_{obs} のクラスタリング結果と $X(z)$ のクラスタリング結果が同じとなるような z 軸上の区間から切断正規分布を構築

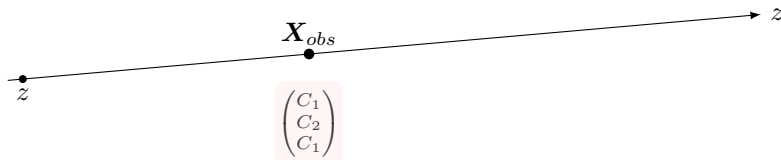


この切断正規分布の累積分布関数から selective- p 値を計算

凸クラスタリングにおける parametric Selective Inference

例として, $X_{obs} \in \mathbb{R}^{3 \times 1}$, クラスタリング結果を $(C_1, C_2, C_1)^\top$ とする

- X_{obs} のクラスタリング結果と $X(z)$ のクラスタリング結果が同じとなるような z 軸上の区間から切断正規分布を構築

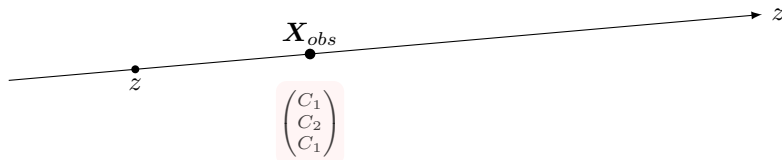


この切断正規分布の累積分布関数から selective- p 値を計算

凸クラスタリングにおける parametric Selective Inference

例として, $X_{obs} \in \mathbb{R}^{3 \times 1}$, クラスタリング結果を $(C_1, C_2, C_1)^\top$ とする

- X_{obs} のクラスタリング結果と $X(z)$ のクラスタリング結果が同じとなるような z 軸上の区間から切断正規分布を構築

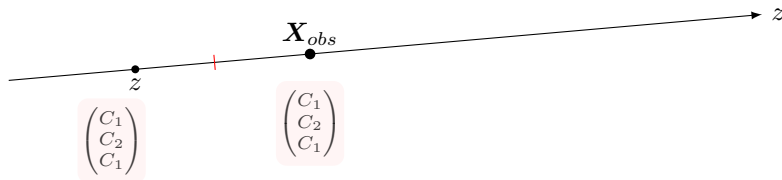


この切断正規分布の累積分布関数から selective- p 値を計算

凸クラスタリングにおける parametric Selective Inference

例として, $X_{obs} \in \mathbb{R}^{3 \times 1}$, クラスタリング結果を $(C_1, C_2, C_1)^\top$ とする

- X_{obs} のクラスタリング結果と $X(z)$ のクラスタリング結果が同じとなるような z 軸上の区間から切断正規分布を構築

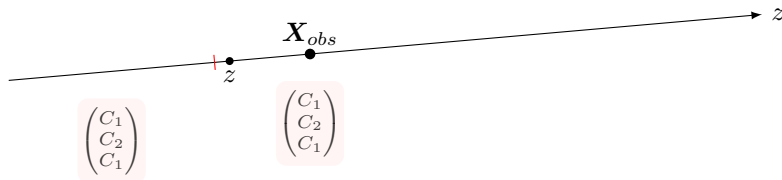


この切断正規分布の累積分布関数から selective- p 値を計算

凸クラスタリングにおける parametric Selective Inference

例として, $X_{obs} \in \mathbb{R}^{3 \times 1}$, クラスタリング結果を $(C_1, C_2, C_1)^\top$ とする

- X_{obs} のクラスタリング結果と $X(z)$ のクラスタリング結果が同じとなるような z 軸上の区間から切断正規分布を構築

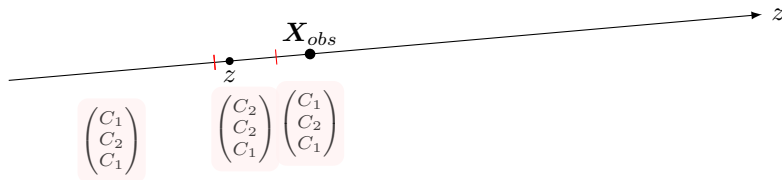


この切断正規分布の累積分布関数から selective- p 値を計算

凸クラスタリングにおける parametric Selective Inference

例として, $X_{obs} \in \mathbb{R}^{3 \times 1}$, クラスタリング結果を $(C_1, C_2, C_1)^\top$ とする

- X_{obs} のクラスタリング結果と $X(z)$ のクラスタリング結果が同じとなるような z 軸上の区間から切断正規分布を構築

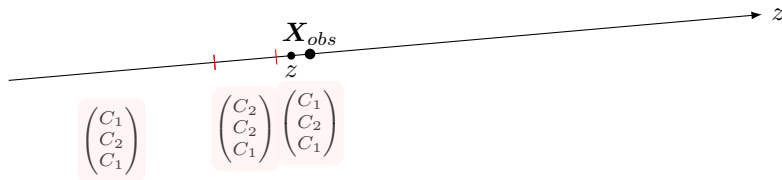


この切断正規分布の累積分布関数から selective- p 値を計算

凸クラスタリングにおける parametric Selective Inference

例として, $X_{obs} \in \mathbb{R}^{3 \times 1}$, クラスタリング結果を $(C_1, C_2, C_1)^\top$ とする

- X_{obs} のクラスタリング結果と $X(z)$ のクラスタリング結果が同じとなるような z 軸上の区間から切断正規分布を構築

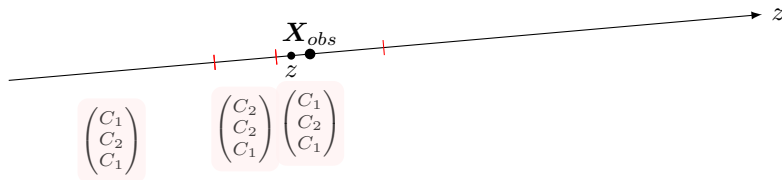


この切断正規分布の累積分布関数から selective- p 値を計算

凸クラスタリングにおける parametric Selective Inference

例として, $X_{obs} \in \mathbb{R}^{3 \times 1}$, クラスタリング結果を $(C_1, C_2, C_1)^\top$ とする

- X_{obs} のクラスタリング結果と $X(z)$ のクラスタリング結果が同じとなるような z 軸上の区間から切断正規分布を構築

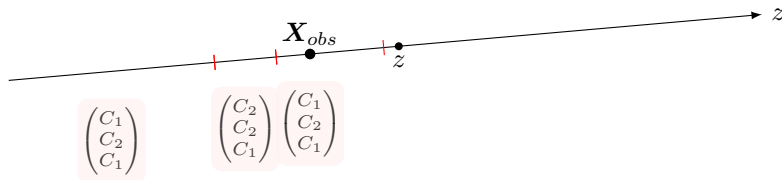


この切断正規分布の累積分布関数から selective- p 値を計算

凸クラスタリングにおける parametric Selective Inference

例として, $X_{obs} \in \mathbb{R}^{3 \times 1}$, クラスタリング結果を $(C_1, C_2, C_1)^\top$ とする

- X_{obs} のクラスタリング結果と $X(z)$ のクラスタリング結果が同じとなるような z 軸上の区間から切断正規分布を構築

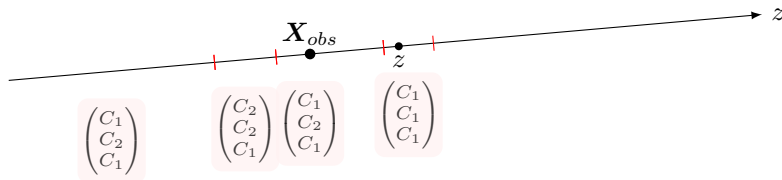


この切断正規分布の累積分布関数から selective- p 値を計算

凸クラスタリングにおける parametric Selective Inference

例として, $X_{obs} \in \mathbb{R}^{3 \times 1}$, クラスタリング結果を $(C_1, C_2, C_1)^\top$ とする

- X_{obs} のクラスタリング結果と $X(z)$ のクラスタリング結果が同じとなるような z 軸上の区間から切断正規分布を構築

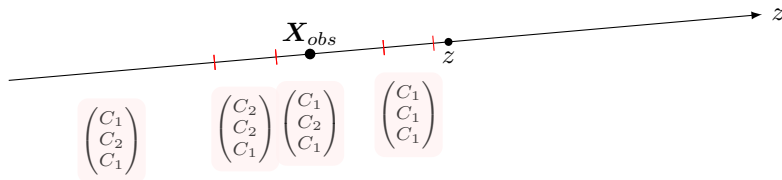


この切断正規分布の累積分布関数から selective- p 値を計算

凸クラスタリングにおける parametric Selective Inference

例として, $X_{obs} \in \mathbb{R}^{3 \times 1}$, クラスタリング結果を $(C_1, C_2, C_1)^\top$ とする

- X_{obs} のクラスタリング結果と $X(z)$ のクラスタリング結果が同じとなるような z 軸上の区間から切断正規分布を構築

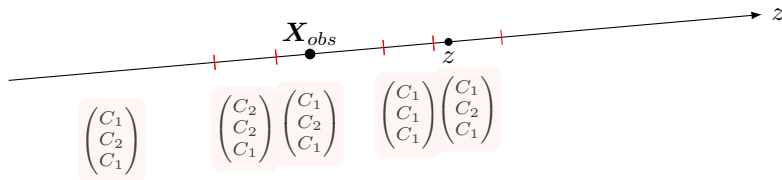


この切断正規分布の累積分布関数から selective- p 値を計算

凸クラスタリングにおける parametric Selective Inference

例として, $X_{obs} \in \mathbb{R}^{3 \times 1}$, クラスタリング結果を $(C_1, C_2, C_1)^\top$ とする

- X_{obs} のクラスタリング結果と $X(z)$ のクラスタリング結果が同じとなるような z 軸上の区間から切断正規分布を構築

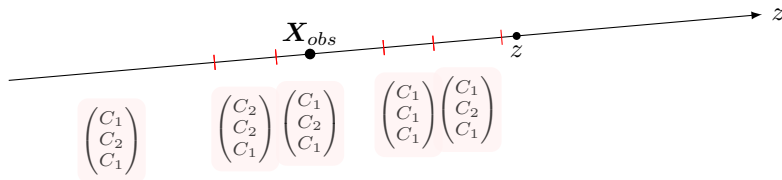


この切断正規分布の累積分布関数から selective- p 値を計算

凸クラスタリングにおける parametric Selective Inference

例として, $X_{obs} \in \mathbb{R}^{3 \times 1}$, クラスタリング結果を $(C_1, C_2, C_1)^\top$ とする

- X_{obs} のクラスタリング結果と $X(z)$ のクラスタリング結果が同じとなるような z 軸上の区間から切断正規分布を構築

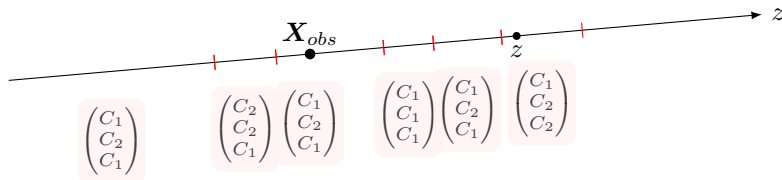


この切断正規分布の累積分布関数から selective- p 値を計算

凸クラスタリングにおける parametric Selective Inference

例として, $X_{obs} \in \mathbb{R}^{3 \times 1}$, クラスタリング結果を $(C_1, C_2, C_1)^\top$ とする

- X_{obs} のクラスタリング結果と $X(z)$ のクラスタリング結果が同じとなるような z 軸上の区間から切断正規分布を構築

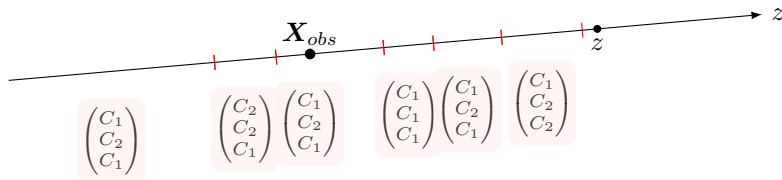


この切断正規分布の累積分布関数から selective- p 値を計算

凸クラスタリングにおける parametric Selective Inference

例として, $X_{obs} \in \mathbb{R}^{3 \times 1}$, クラスタリング結果を $(C_1, C_2, C_1)^\top$ とする

- X_{obs} のクラスタリング結果と $X(z)$ のクラスタリング結果が同じとなるような z 軸上の区間から切断正規分布を構築

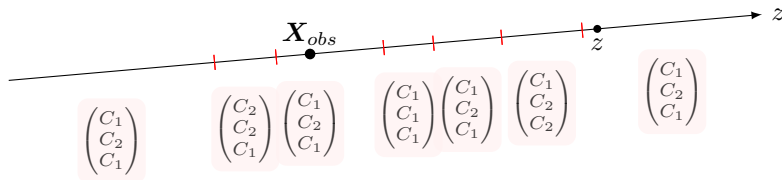


この切断正規分布の累積分布関数から selective- p 値を計算

凸クラスタリングにおける parametric Selective Inference

例として, $X_{obs} \in \mathbb{R}^{3 \times 1}$, クラスタリング結果を $(C_1, C_2, C_1)^\top$ とする

- X_{obs} のクラスタリング結果と $X(z)$ のクラスタリング結果が同じとなるような z 軸上の区間から切断正規分布を構築

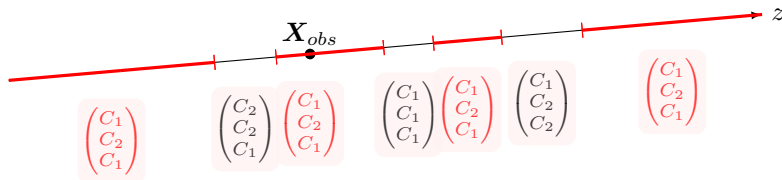


この切断正規分布の累積分布関数から selective- p 値を計算

凸クラスタリングにおける parametric Selective Inference

例として, $X_{obs} \in \mathbb{R}^{3 \times 1}$, クラスタリング結果を $(C_1, C_2, C_1)^\top$ とする

- X_{obs} のクラスタリング結果と $X(z)$ のクラスタリング結果が同じとなるような z 軸上の区間から切断正規分布を構築

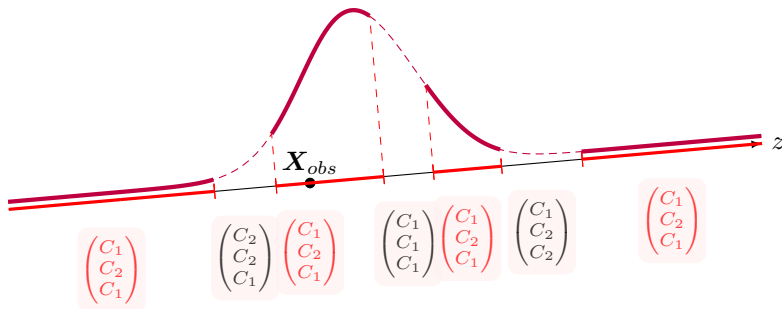


この切断正規分布の累積分布関数から selective- p 値を計算

凸クラスタリングにおける parametric Selective Inference

例として, $X_{obs} \in \mathbb{R}^{3 \times 1}$, クラスタリング結果を $(C_1, C_2, C_1)^\top$ とする

- X_{obs} のクラスタリング結果と $X(z)$ のクラスタリング結果が同じとなるような z 軸上の区間から切断正規分布を構築



この切断正規分布の累積分布関数から selective- p 値を計算

z の解パスアルゴリズムの設定

pSI にて, クラスタリング結果が変化する z を解析的に導出する解パスを提案する

- 一般化 lasso の λ の解パスを基に, z の解パスとして定式化

表記

- $D_{\mathcal{A}}$: \mathcal{A} に含まれる行インデックスから構成される D の部分行列
- s : \mathcal{A} の符号集合ベクトル
- z_h : \mathcal{A} に新たなメンバーが加わる時点の z
 - i_h : 新たに加わる D の行インデックス
- z_l : \mathcal{A} からあるメンバーが脱ける時点の z
 - i_l : 新たに脱ける D の行インデックス

z の解パスアルゴリズム

初期条件: $z = -\infty$, $\mathcal{A} = \mathcal{A}$ of Convex Clustering($\mathbf{a} + \mathbf{b}z$)
以下の手順を $z = \infty$ となるまで繰り返しながら z を更新

1. z_h, i_h を導出

$$z_h, i_h = \min_i \frac{[(D_{\bar{\mathcal{A}}} D_{\bar{\mathcal{A}}}^\top)^+ D_{\bar{\mathcal{A}}} (\lambda D_{\mathcal{A}}^\top \mathbf{s} - \mathbf{a})]_i \pm \lambda}{[(D_{\bar{\mathcal{A}}} D_{\bar{\mathcal{A}}}^\top)^+ D_{\bar{\mathcal{A}}} \mathbf{b}]_i}$$

2. z_l, i_l を導出

$$z_l, i_l = \min_i \frac{[D_{\mathcal{A}} (I - D_{\bar{\mathcal{A}}}^\top (D_{\bar{\mathcal{A}}} D_{\bar{\mathcal{A}}}^\top)^+ D_{\bar{\mathcal{A}}}) (\lambda D_{\mathcal{A}}^\top \mathbf{s} - \mathbf{a})]_i}{[D_{\mathcal{A}} (I - D_{\bar{\mathcal{A}}}^\top (D_{\bar{\mathcal{A}}} D_{\bar{\mathcal{A}}}^\top)^+ D_{\bar{\mathcal{A}}}) \mathbf{b}]_i}$$

3. $z = \min\{z_h, z_l\}$ とし, \mathcal{A} の変化を適用

$$\begin{cases} z = z_h : \mathcal{A} = \mathcal{A} \cup \{i_h\} \\ z = z_l : \mathcal{A} = \mathcal{A} \setminus \{i_l\} \end{cases}$$

- 1 はじめに
- 2 凸クラスタリング
- 3 parametric Selective Inference
- 4 計算機実験**
- 5 まとめと今後の課題

人工データを用いた FPR 実験

- $X \in \mathbb{R}^{4 \times 2}$ とし, 各要素を $N(0, 1)$ から独立に生成
- 凸クラスタリングの結果からランダムに 2 クラスタを選択して検定を試行
- 選択バイアスを考慮しない検定 (naive) と比較

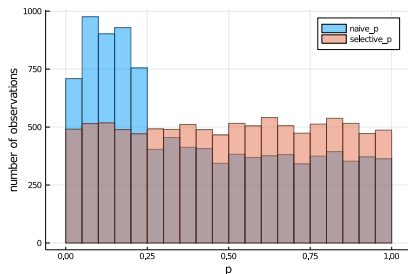


Figure: $\lambda = 0.4$ の場合の p 値の分布

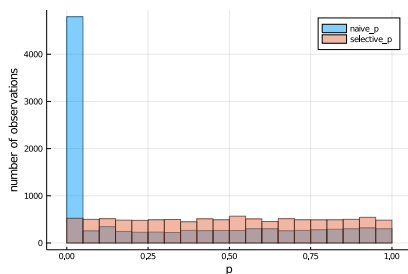


Figure: $\lambda = 0.7$ の場合の p 値の分布

selective- p の分布が一様分布 \rightarrow FPR の制御に成功

人工データを用いた検出力実験

- $X = (x_1, x_2, x_3, x_4)^T \in \mathbb{R}^{4 \times 2}$ とする
- 真のクラスタリング結果を (C_1, C_1, C_2, C_2) とする
 - x_1, x_2 の各要素は $N(0, 1)$ から, x_3, x_4 の各要素は $N(\mu, 1)$ から独立に生成

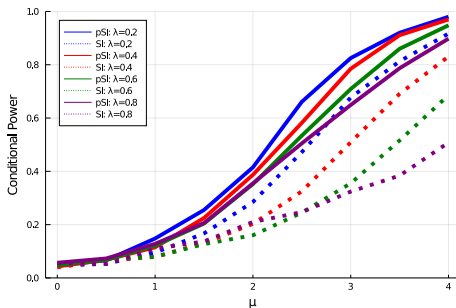


Figure: SI と pSI との検出力の比較

pSI は SI に比べて検出力が向上していることを確認

iris データを用いた実データ実験

- 150 個の標本からランダムに 50 個を選択し, $X \in \mathbb{R}^{50 \times 4}$ とする
- クラスタ数が 9 となるような λ に設定

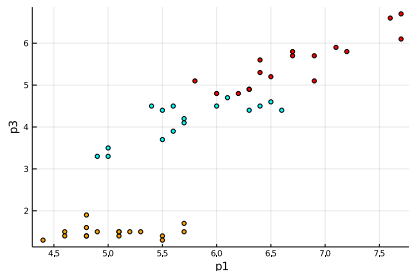


Figure: X の分布

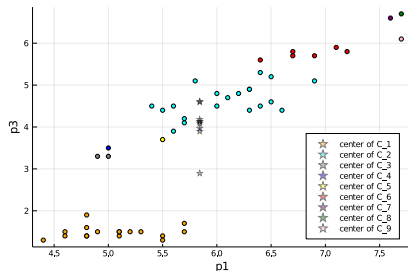


Figure: 凸クラスタリング結果

Table: pSI, SI (Conditional SI) による p 値の結果.

検定問題	pSI- p	SI- p
p3: C_1 vs C_2	0.004	0.060
p3: C_1 vs C_6	0.027	0.047

- 1 はじめに
- 2 凸クラスタリング
- 3 parametric Selective Inference
- 4 計算機実験
- 5 **まとめと今後の課題**

- まとめ
 - クラスタリング後の任意の 2 クラスタ間の検定問題を pSI の枠組みで定式化
 - 選択バイアスを考慮した妥当な検定が行えることを確認
- 今後の課題
 - さらなる実データ実験の実施

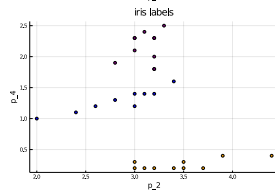
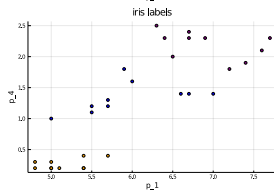
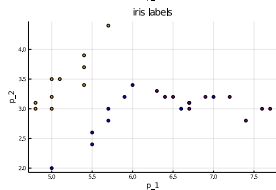
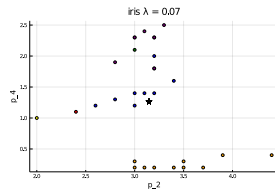
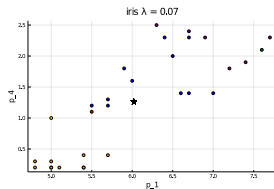
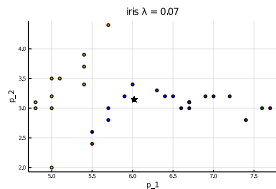
noframenumbering

- [1] Vo Nguyen Le Duy and Ichiro Takeuchi. “Parametric Programming Approach for Powerful Lasso Selective Inference without Conditioning on Signs”. In: *arXiv preprint arXiv:2004.09749* (2020).
- [2] Toby Dylan Hocking et al. “Clusterpath an algorithm for clustering using convex fusion penalties”. In: 2011.
- [3] Jason D Lee et al. “Exact post-selection inference, with application to the lasso”. In: *The Annals of Statistics* 44.3 (2016), pp. 907–927.

iris データでクラスタリング

- データ数: 30, 得られたクラス数: 6 (橙 10, 青 11, 紫 6, 赤 1, 黄 1, 緑 1)
- 設定した λ で収束した特徴量 (1, 2, 4)

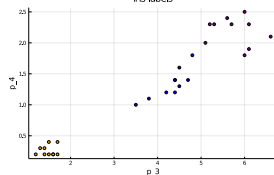
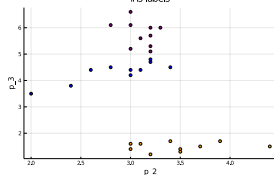
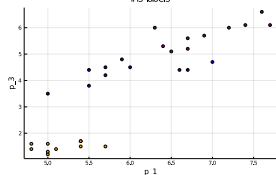
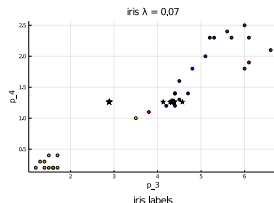
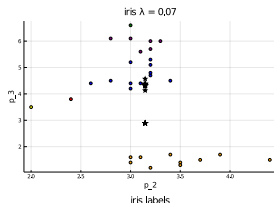
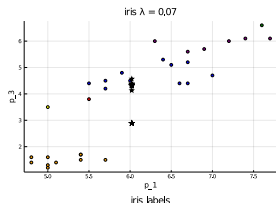
特徴 3 以外の組み合わせ



iris データでクラスタリング

特徴 3 を含む組み合わせ

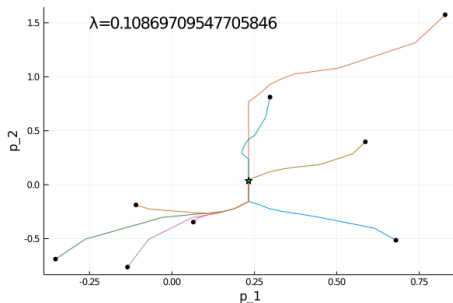
- この λ 以降はクラスタ中心 (星) が近いものから順に統合される
→ 青に対して黄や緑よりも先に紫が統合されてしまう



クラスタパス

クラスタパス: $\lambda : \lambda_{max} \rightarrow 0$ に対する β の推移

- $X \in \mathbb{R}^{8 \times 2}$

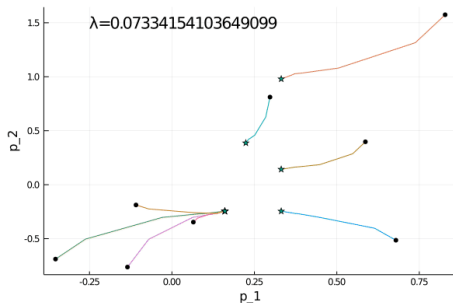


- クラスタ数: 1

クラスタパス

クラスタパス: $\lambda : \lambda_{max} \rightarrow 0$ に対する β の推移

- $X \in \mathbb{R}^{8 \times 2}$

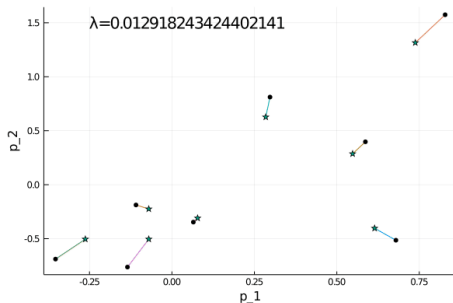


- クラスタ数: 5

クラスタパス

クラスタパス: $\lambda : \lambda_{max} \rightarrow 0$ に対する β の推移

- $X \in \mathbb{R}^{8 \times 2}$



- クラスタ数: 8