Post-clustering inference under dependency

Javier González-Delgado[1], Mathis Deronzier[2], Juan Cortés[3] and Pierre Neuvial[2]

1. University of Rennes, ENSAI, CREST, 2. Institut de Mathématiques de Toulouse, 3. LAAS-CNRS

13th International Conference on Multiple Comparison Procedures (MCP–2025)

August 15, 2025

Post-clustering inference

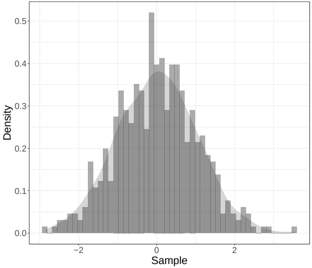
Post-clustering inference

Toy example

Post-clustering inference

Toy example

- Simulate $\mathcal{N}(0,1) + \mathcal{U}(-0.2, 0.2)$

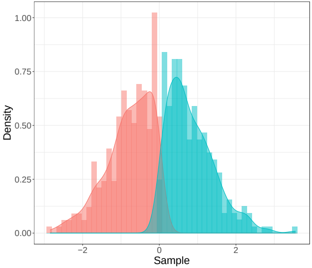


Adapted from Hivert *et al.* 2024

Post-clustering inference

Toy example

- Simulate $\mathcal{N}(0,1) + \mathcal{U}(-0.2, 0.2)$
- Ask k-means to find 2 clusters (data-driven hypothesis selection)

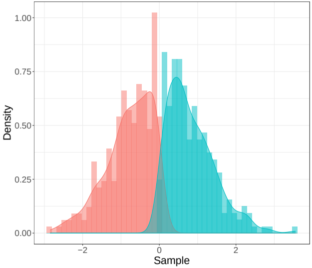


Adapted from Hivert *et al.* 2024

Post-clustering inference

Toy example

- Simulate $\mathcal{N}(0,1) + \mathcal{U}(-0.2, 0.2)$
- Ask k-means to find 2 clusters (data-driven hypothesis selection)
- Test for the difference of cluster means (inference after selection)

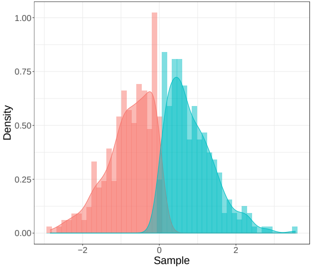


Adapted from Hivert *et al.* 2024

Post-clustering inference

Toy example

- Simulate $\mathcal{N}(0,1) + \mathcal{U}(-0.2, 0.2)$
- Ask k-means to find 2 clusters (data-driven hypothesis selection)
- Test for the difference of cluster means (inference after selection)



- Ignoring adaptive selection : $p_Z = 10^{-67}$,
- Accounting for adaptive selection : $p_{AS} = 0.84$ (Chen and Witten 2023).

Adapted from Hivert *et al.* 2024

Post-clustering inference

Framework setting

- Let $C(\cdot)$ be a clustering algorithm, $\mathbf{X}$ a $n \times p$ random matrix with $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$.

Post-clustering inference

Framework setting

- Let $C(\cdot)$ be a clustering algorithm, $\mathbf{X}$ a $n \times p$ random matrix with $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$.
- Let X_i (resp. μ_i) denote the i-th row of $\mathbf{X}$ (resp. $\boldsymbol{\mu}$) for $i \in [n] = \{1, \ldots, n\}$.

Post-clustering inference

Framework setting

- Let $C(\cdot)$ be a clustering algorithm, $\mathbf{X}$ a $n \times p$ random matrix with $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$.
- Let X_i (resp. μ_i) denote the i-th row of $\mathbf{X}$ (resp. $\boldsymbol{\mu}$) for $i \in [n] = \{1, \ldots, n\}$.
- For any $\mathcal{G} \subset \{1, \ldots, n\}$, let $\bar{X}_{\mathcal{G}} = \frac{1}{|\mathcal{G}|} \sum_{i \in \mathcal{G}} X_i$ and $\bar{\mu}_{\mathcal{G}} = \frac{1}{|\mathcal{G}|} \sum_{i \in \mathcal{G}} \mu_i$.

Post-clustering inference

Framework setting

- Let $C(\cdot)$ be a clustering algorithm, $\mathbf{X}$ a $n \times p$ random matrix with $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$.
- Let X_i (resp. μ_i) denote the i-th row of $\mathbf{X}$ (resp. $\boldsymbol{\mu}$) for $i \in [n] = \{1, \ldots, n\}$.
- For any $\mathcal{G} \subset \{1, \ldots, n\}$, let $\bar{X}_{\mathcal{G}} = \frac{1}{|\mathcal{G}|} \sum_{i \in \mathcal{G}} X_i$ and $\bar{\mu}_{\mathcal{G}} = \frac{1}{|\mathcal{G}|} \sum_{i \in \mathcal{G}} \mu_i$.
- Let $\mathcal{G}_1, \mathcal{G}_2 \subset \{1, \ldots, n\}$ be two non-overlapping groups of observations. Considering the column vector $\nu_{\mathcal{G}_1, \mathcal{G}_2} = \nu$ having as components

$$\nu_i = \mathbb{1}\{i \in \mathcal{G}_1\}/|\mathcal{G}_1| - \mathbb{1}\{i \in \mathcal{G}_2\}/|\mathcal{G}_2|,$$

for $i \in [n]$,

Post-clustering inference

Framework setting

- Let $C(\cdot)$ be a clustering algorithm, $\mathbf{X}$ a $n \times p$ random matrix with $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$.
- Let X_i (resp. μ_i) denote the i-th row of $\mathbf{X}$ (resp. $\boldsymbol{\mu}$) for $i \in [n] = \{1, \dots, n\}$.
- For any $\mathcal{G} \subset \{1, \dots, n\}$, let $\bar{X}_{\mathcal{G}} = \frac{1}{|\mathcal{G}|} \sum_{i \in \mathcal{G}} X_i$ and $\bar{\mu}_{\mathcal{G}} = \frac{1}{|\mathcal{G}|} \sum_{i \in \mathcal{G}} \mu_i$.
- Let $\mathcal{G}_1, \mathcal{G}_2 \subset \{1, \dots, n\}$ be two non-overlapping groups of observations. Considering the column vector $\nu_{\mathcal{G}_1, \mathcal{G}_2} = \nu$ having as components

$$\nu_i = \mathbb{1}\{i \in \mathcal{G}_1\}/|\mathcal{G}_1| - \mathbb{1}\{i \in \mathcal{G}_2\}/|\mathcal{G}_2|,$$

for $i \in [n]$, we can write the difference between the (empirical) group means as

$$\bar{\mu}_{\mathcal{G}_1} - \bar{\mu}_{\mathcal{G}_2} = \boldsymbol{\mu}^T \nu, \quad \text{and} \quad \bar{X}_{\mathcal{G}_1} - \bar{X}_{\mathcal{G}_2} = \mathbf{X}^T \nu.$$

Post-clustering inference

Framework setting

- Let $C(\cdot)$ be a clustering algorithm, $\mathbf{X}$ a $n \times p$ random matrix with $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$.
- Let X_i (resp. μ_i) denote the i-th row of $\mathbf{X}$ (resp. $\boldsymbol{\mu}$) for $i \in [n] = \{1, \ldots, n\}$.
- For any $\mathcal{G} \subset \{1, \ldots, n\}$, let $\bar{X}_{\mathcal{G}} = \frac{1}{|\mathcal{G}|} \sum_{i \in \mathcal{G}} X_i$ and $\bar{\mu}_{\mathcal{G}} = \frac{1}{|\mathcal{G}|} \sum_{i \in \mathcal{G}} \mu_i$.
- Let $\mathcal{G}_1, \mathcal{G}_2 \subset \{1, \ldots, n\}$ be two non-overlapping groups of observations. Considering the column vector $\nu_{\mathcal{G}_1, \mathcal{G}_2} = \nu$ having as components

$$\nu_i = \mathbb{1}\{i \in \mathcal{G}_1\}/|\mathcal{G}_1| - \mathbb{1}\{i \in \mathcal{G}_2\}/|\mathcal{G}_2|,$$

for $i \in [n]$, we can write the difference between the (empirical) group means as

$$\bar{\mu}_{\mathcal{G}_1} - \bar{\mu}_{\mathcal{G}_2} = \boldsymbol{\mu}^T \nu, \quad \text{and} \quad \bar{X}_{\mathcal{G}_1} - \bar{X}_{\mathcal{G}_2} = \mathbf{X}^T \nu.$$

We are interested in the following null hypothesis :

$$H_0^{\{\mathcal{G}_1, \mathcal{G}_2\}} : \boldsymbol{\mu}^T \nu = 0. \tag{H0}$$

Post-clustering inference

The selective type I error for clustering

Goal : Testing (H0) by controlling the selective type I error for clustering at level α, that is, by ensuring that :

$$\mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\left(\text{reject } H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}} \text{ based on } \mathbf{X} \text{ at level } \alpha \;\middle|\; \mathcal{G}_1, \mathcal{G}_2 \in \mathcal{C}(\mathbf{X})\right) \leq \alpha \quad \forall\, \alpha \in (0,1).$$

Post-clustering inference

The selective type I error for clustering

Goal : Testing (H0) by controlling the selective type I error for clustering at level α, that is, by ensuring that :

$$\mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\left(\text{reject } H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}} \text{ based on } \mathbf{X} \text{ at level } \alpha \;\middle|\; \mathcal{G}_1, \mathcal{G}_2 \in \mathcal{C}(\mathbf{X})\right) \leq \alpha \quad \forall\, \alpha \in (0,1).$$

Ideal p-value :

$$p_{\text{ideal}} = \mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\left(\text{Critical region} \;\middle|\; \mathcal{G}_1, \mathcal{G}_2 \in \mathcal{C}(\mathbf{X})\right).$$

Post-clustering inference

The selective type I error for clustering

Goal : Testing (H0) by controlling the selective type I error for clustering at level α, that is, by ensuring that :

$$\mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\left(\text{reject } H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}} \text{ based on } \mathbf{X} \text{ at level } \alpha \;\middle|\; \mathcal{G}_1, \mathcal{G}_2 \in \mathcal{C}(\mathbf{X})\right) \leq \alpha \quad \forall \alpha \in (0,1).$$

Ideal p-value :

$$p_{\text{ideal}} = \mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\left(\text{Critical region} \;\middle|\; \mathcal{G}_1, \mathcal{G}_2 \in \mathcal{C}(\mathbf{X})\right).$$

Analytically tractable p-value :

$$p_{\text{tractable}} = \mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\left(\text{Critical region} \;\middle|\; \mathcal{G}_1, \mathcal{G}_2 \in \mathcal{C}(\mathbf{X}) \cap E(\mathbf{X})\right).$$

... paying a price in power [1].

1. Jewell *et al.* 2022, Chen *et al.* 2022, Liu *et al.* 2018, Fithian *et al.* 2017.

Post-clustering inference

The selective type I error for clustering

Goal : Testing (H0) by controlling the selective type I error for clustering at level α, that is, by ensuring that :

$$\mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\left(\text{reject } H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}} \text{ based on } \mathbf{X} \text{ at level } \alpha \;\middle|\; \mathcal{G}_1, \mathcal{G}_2 \in \mathcal{C}(\mathbf{X})\right) \leq \alpha \quad \forall \alpha \in (0,1).$$

Ideal p-value :

$$p_{\text{ideal}} = \mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\left(\text{Critical region} \;\middle|\; \mathcal{G}_1, \mathcal{G}_2 \in \mathcal{C}(\mathbf{X})\right).$$

Analytically tractable p-value :

$$p_{\text{tractable}} = \mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\left(\text{Critical region} \;\middle|\; \mathcal{G}_1, \mathcal{G}_2 \in \mathcal{C}(\mathbf{X}) \cap E(\mathbf{X})\right).$$

... paying a price in power [1].

- Both p_{ideal} and $p_{\text{tractable}}$ control the selective type I error for clustering.

1. Jewell *et al.* 2022, Chen *et al.* 2022, Liu *et al.* 2018, Fithian *et al.* 2017.

Independence setting (I)

Gao, Bien and Witten 2022

Framework

Consider the model

$$\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{I}_n, \sigma^2 \mathbf{I}_p), \qquad \text{(indep)}$$

Independence setting (I)

Gao, Bien and Witten 2022

Framework

Consider the model

$$\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{I}_n, \sigma^2 \mathbf{I}_p), \qquad \text{(indep)}$$

that is, observations are independent

Independence setting (I)

Gao, Bien and Witten 2022

Framework

Consider the model

$$\mathbf{X} \sim \mathcal{MN}_{n \times p}(\boldsymbol{\mu}, \mathbf{I}_n, \sigma^2 \mathbf{I}_p), \tag{indep}$$

that is, observations are independent and distributed as $X_i \sim \mathcal{N}_p(\mu_i, \sigma^2 \mathbf{I}_p)$.

Independence setting (I)

Gao, Bien and Witten 2022

Framework

Consider the model

$$\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{I}_n, \sigma^2 \mathbf{I}_p), \tag{indep}$$

that is, observations are independent and distributed as $X_i \sim \mathcal{N}_p(\mu_i, \sigma^2 \mathbf{I}_p)$.

p-value for (H0) under (indep) (Gao, Bien and Witten 2022)

$$p(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = \mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\left(\|\mathbf{X}^T\nu\|_2 \geq \|\mathbf{x}^T\nu\|_2 \;\middle|\; \mathcal{G}_1, \mathcal{G}_2 \in \mathcal{C}(\mathbf{X}),\right. \tag{p-GBW}$$

Independence setting (I)

Gao, Bien and Witten 2022

Framework

Consider the model

$$\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{I}_n, \sigma^2 \mathbf{I}_p), \tag{indep}$$

that is, observations are independent and distributed as $X_i \sim \mathcal{N}_p(\mu_i, \sigma^2 \mathbf{I}_p)$.

p-value for (H0) under (indep) (Gao, Bien and Witten 2022)

$$p(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = \mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\Big(\|\mathbf{X}^T\nu\|_2 \geq \|\mathbf{x}^T\nu\|_2 \;\Big|\; \mathcal{G}_1, \mathcal{G}_2 \in \mathcal{C}(\mathbf{X}), \\ \boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x}, \operatorname{dir}(\mathbf{X}^T\nu) = \operatorname{dir}(\mathbf{x}^T\nu)\Big), \tag{p-GBW}$$

where $\boldsymbol{\pi}_\nu^\perp = \mathbf{I}_n - \nu\nu^T/\|\nu\|_2^2$ and $\operatorname{dir}(v) = v/\|v\|_2 \mathbb{1}\{v \neq 0\}$ for all $v \in \mathbb{R}^p$.

Independence setting (I)

Gao, Bien and Witten 2022

Framework

Consider the model

$$\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{I}_n, \sigma^2 \mathbf{I}_p), \tag{indep}$$

that is, observations are independent and distributed as $X_i \sim \mathcal{N}_p(\mu_i, \sigma^2 \mathbf{I}_p)$.

p-value for (H0) under (indep) (Gao, Bien and Witten 2022)

$$p(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = \mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\Big(\|\mathbf{X}^T\nu\|_2 \geq \|\mathbf{x}^T\nu\|_2 \;\Big|\; \mathcal{G}_1, \mathcal{G}_2 \in \mathcal{C}(\mathbf{X}), \boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x}, \operatorname{dir}(\mathbf{X}^T\nu) = \operatorname{dir}(\mathbf{x}^T\nu)\Big), \tag{p-GBW}$$

where $\boldsymbol{\pi}_\nu^\perp = \mathbf{I}_n - \nu\nu^T/\|\nu\|_2^2$ and $\operatorname{dir}(v) = v/\|v\|_2 \mathbb{1}\{v \neq 0\}$ for all $v \in \mathbb{R}^p$.

The extra conditioning event allows to rewrite :

$$\{\mathcal{G}_1, \mathcal{G}_2 \in \mathcal{C}(\mathbf{X}), \boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x}, \operatorname{dir}(\mathbf{X}^T\nu) = \operatorname{dir}(\mathbf{x}^T\nu)\} = \{\|\mathbf{X}^T\nu\|_2 \in \mathcal{S}_2(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}), \boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x}, \operatorname{dir}(\mathbf{X}^T\nu) = \operatorname{dir}(\mathbf{x}^T\nu)\}.$$

Independence setting (II)

Gao, Bien and Witten 2022

p-value for (H0) under (indep) (Gao, Bien and Witten 2022)

$$p(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = \mathbb{P}_{H_0^{\{\mathcal{G}_1, \mathcal{G}_2\}}}\Big(\|\mathbf{X}^T \nu\|_2 \geq \|\mathbf{x}^T \nu\|_2 \;\Big|\; \|\mathbf{X}^T \nu\|_2 \in \mathcal{S}_2(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}),$$

$$\boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x}, \, \mathrm{dir}(\mathbf{X}^T \nu) = \mathrm{dir}(\mathbf{x}^T \nu) \Big).$$

Independence setting (II)

Gao, Bien and Witten 2022

p-value for (H0) under (indep) (Gao, Bien and Witten 2022)

$$p(\mathbf{x};\{\mathcal{G}_1,\mathcal{G}_2\}) = \mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\Big(\|\mathbf{X}^T\nu\|_2 \geq \|\mathbf{x}^T\nu\|_2 \;\Big|\; \|\mathbf{X}^T\nu\|_2 \in \mathcal{S}_2(\mathbf{x};\{\mathcal{G}_1,\mathcal{G}_2\}), \pi_\nu^\perp \mathbf{X} = \pi_\nu^\perp \mathbf{x},\, \mathrm{dir}(\mathbf{X}^T\nu) = \mathrm{dir}(\mathbf{x}^T\nu)\Big).$$

Strategy to derive a tractable form of $p(\mathbf{x};\{\mathcal{G}_1,\mathcal{G}_2\})$

- $\mathbf{X}^T\nu \sim \mathcal{N}_p(0_p, \sigma^2\|\nu\|_2^2\, \mathbf{I}_p)$ under (H0),

Independence setting (II)

Gao, Bien and Witten 2022

p-value for (H0) under (indep) (Gao, Bien and Witten 2022)

$$p(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = \mathbb{P}_{H_0^{\{\mathcal{G}_1, \mathcal{G}_2\}}}\Big(\|\mathbf{X}^T\nu\|_2 \geq \|\mathbf{x}^T\nu\|_2 \;\Big|\; \|\mathbf{X}^T\nu\|_2 \in \mathcal{S}_2(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}), \boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x}, \, \mathrm{dir}(\mathbf{X}^T\nu) = \mathrm{dir}(\mathbf{x}^T\nu) \Big).$$

Strategy to derive a tractable form of $p(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\})$

- $\mathbf{X}^T\nu \sim \mathcal{N}_p(0_p, \sigma^2\|\nu\|_2^2 \, \mathbf{I}_p)$ under (H0),
- Choice of the norm $\|\cdot\|_2 \to \|\mathbf{X}^T\nu\|_2 \sim \sigma\|\nu\|_2 \cdot \chi_p$ under (H0),

Independence setting (II)

Gao, Bien and Witten 2022

p-value for (H0) under (indep) (Gao, Bien and Witten 2022)

$$p(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = \mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\Big(\|\mathbf{X}^T\nu\|_2 \geq \|\mathbf{x}^T\nu\|_2 \;\Big|\; \|\mathbf{X}^T\nu\|_2 \in \mathcal{S}_2(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}), \boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x}, \operatorname{dir}(\mathbf{X}^T\nu) = \operatorname{dir}(\mathbf{x}^T\nu) \Big).$$

Strategy to derive a tractable form of $p(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\})$

- $\mathbf{X}^T\nu \sim \mathcal{N}_p(0_p, \sigma^2\|\nu\|_2^2\, \mathbf{I}_p)$ under (H0),
- Choice of the norm $\|\cdot\|_2 \to \|\mathbf{X}^T\nu\|_2 \sim \sigma\|\nu\|_2 \cdot \chi_p$ under (H0),
- $\{\boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x}, \operatorname{dir}(\mathbf{X}^T\nu) = \operatorname{dir}(\mathbf{x}^T\nu)\}$ is independent of $\|\mathbf{X}^T\nu\|_2 \to$ $p(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\})$ is written as the CDF of a truncated χ_p.

Independence setting (II)

Gao, Bien and Witten 2022

p-value for (H0) under (indep) (Gao, Bien and Witten 2022)

$$p(\mathbf{x};\{\mathcal{G}_1,\mathcal{G}_2\}) = \mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\Big(\|\mathbf{X}^T\nu\|_2 \geq \|\mathbf{x}^T\nu\|_2 \;\Big|\; \|\mathbf{X}^T\nu\|_2 \in \mathcal{S}_2(\mathbf{x};\{\mathcal{G}_1,\mathcal{G}_2\}),$$
$$\boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x},\ \mathrm{dir}(\mathbf{X}^T\nu) = \mathrm{dir}(\mathbf{x}^T\nu)\Big).$$

Strategy to derive a tractable form of $p(\mathbf{x};\{\mathcal{G}_1,\mathcal{G}_2\})$

- $\mathbf{X}^T\nu \sim \mathcal{N}_p(0_p, \sigma^2\|\nu\|_2^2\,\mathbf{I}_p)$ under (H0),
- Choice of the norm $\|\cdot\|_2 \to \|\mathbf{X}^T\nu\|_2 \sim \sigma\|\nu\|_2 \cdot \chi_p$ under (H0),
- $\{\boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x},\ \mathrm{dir}(\mathbf{X}^T\nu) = \mathrm{dir}(\mathbf{x}^T\nu)\}$ is independent of $\|\mathbf{X}^T\nu\|_2 \to$ $p(\mathbf{x};\{\mathcal{G}_1,\mathcal{G}_2\})$ is written as the CDF of a truncated χ_p.

Theorem 1 in Gao, Bien and Witten 2022

$$p(\mathbf{x};\{\mathcal{G}_1,\mathcal{G}_2\}) = 1 - \mathbb{F}_p\big(\|\mathbf{X}^T\nu\|_2; \sigma\|\nu\|_2, \mathcal{S}_2(\mathbf{x};\{\mathcal{G}_1,\mathcal{G}_2\})\big)$$

where $\mathbb{F}_p(t;c,\mathcal{S})$ denotes the CDF of a $c\chi_p$ random variable truncated to the set $\mathcal{S}$.

Independence setting (II)

Gao, Bien and Witten 2022

p-value for (H0) under (indep) (Gao, Bien and Witten 2022)

$$p(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = \mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\Big(\|\mathbf{X}^T\nu\|_2 \geq \|\mathbf{x}^T\nu\|_2 \;\Big|\; \|\mathbf{X}^T\nu\|_2 \in \mathcal{S}_2(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}),$$
$$\boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x}, \, \mathrm{dir}\big(\mathbf{X}^T\nu\big) = \mathrm{dir}\big(\mathbf{x}^T\nu\big) \Big).$$

Strategy to derive a tractable form of $p(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\})$

- $\mathbf{X}^T\nu \sim \mathcal{N}_p(0_p, \sigma^2\|\nu\|_2^2\, \mathbf{I}_p)$ under (H0),
- Choice of the norm $\|\cdot\|_2 \rightarrow \|\mathbf{X}^T\nu\|_2 \sim \sigma\|\nu\|_2 \cdot \chi_p$ under (H0),
- $\{\boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x}, \, \mathrm{dir}\big(\mathbf{X}^T\nu\big) = \mathrm{dir}\big(\mathbf{x}^T\nu\big)\}$ is independent of $\|\mathbf{X}^T\nu\|_2 \rightarrow$ $p(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\})$ is written as the CDF of a truncated χ_p.

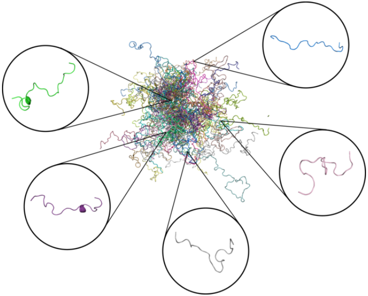
Theorem 1 in Gao, Bien and Witten 2022

$$p(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = 1 - \mathbb{F}_p\Big(\|\mathbf{X}^T\nu\|_2; \sigma\|\nu\|_2, \underbrace{\boxed{\mathcal{S}_2(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\})}}_{\text{HAC, } k\text{-means}}\Big)$$

where $\mathbb{F}_p(t; c, \mathcal{S})$ denotes the CDF of a $c\chi_p$ random variable truncated to the set $\mathcal{S}$.

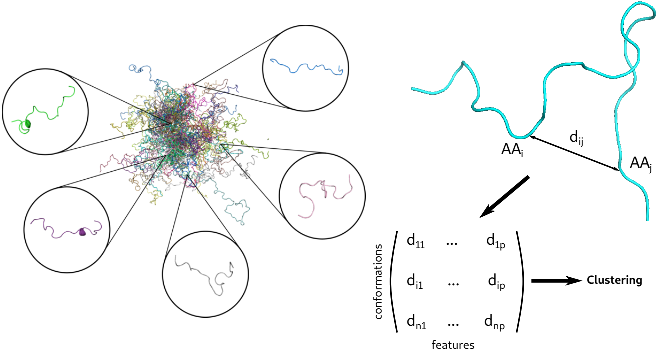
Independence is usually unrealistic

Example : clustering of flexible protein structures



Independence is usually unrealistic

Example : clustering of flexible protein structures



Ignoring dependency prevents selective type I error control

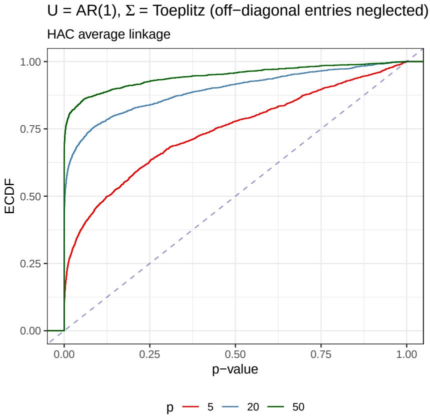
- Simulate $X_i \sim \mathcal{N}_p(\mathbf{0}_p, \mathbf{\Sigma})$ with $X_1, \ldots, X_n$ *dependent*.

Ignoring dependency prevents selective type I error control

- Simulate $X_i \sim \mathcal{N}_p(\mathbf{0}_p, \mathbf{\Sigma})$ with $X_1, \dots, X_n$ *dependent*.
- Set $\mathcal{C}$ to choose three clusters, randomly select two groups and test for the difference of their means assuming $\mathbf{\Sigma} = \sigma^2 \mathbf{I}_p$ and independent observations.

Ignoring dependency prevents selective type I error control

- Simulate $X_i \sim \mathcal{N}_p(\mathbf{0}_p, \mathbf{\Sigma})$ with $X_1, \ldots, X_n$ *dependent*.
- Set $\mathcal{C}$ to choose three clusters, randomly select two groups and test for the difference of their means assuming $\mathbf{\Sigma} = \sigma^2 \mathbf{I}_p$ and independent observations.



Arbitrary dependence setting

General strategy (I)

Framework

Consider the model

$$\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma}) \Leftrightarrow \text{vec}(\mathbf{X}) \sim \mathcal{N}_{np}(\text{vec}(\boldsymbol{\mu}), \boldsymbol{\Sigma} \otimes \mathbf{U}), \tag{dep}$$

where $\mathbf{U} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ are positive definite.

Arbitrary dependence setting

General strategy (I)

Framework

Consider the model

$$\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma}) \Leftrightarrow \text{vec}(\mathbf{X}) \sim \mathcal{N}_{np}(\text{vec}(\boldsymbol{\mu}), \boldsymbol{\Sigma} \otimes \mathbf{U}), \qquad \text{(dep)}$$

where $\mathbf{U} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ are positive definite. That means :

- $X_i \sim \mathcal{N}_p(\mu_i, U_{ii}\boldsymbol{\Sigma})$ ($\boldsymbol{\Sigma} \leftrightarrow$ dependence between features),

Arbitrary dependence setting

General strategy (I)

Framework

Consider the model

$$\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma}) \Leftrightarrow \text{vec}(\mathbf{X}) \sim \mathcal{N}_{np}(\text{vec}(\boldsymbol{\mu}), \boldsymbol{\Sigma} \otimes \mathbf{U}), \qquad \text{(dep)}$$

where $\mathbf{U} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ are positive definite. That means :

- $X_i \sim \mathcal{N}_p(\mu_i, U_{ii}\boldsymbol{\Sigma})$ ($\boldsymbol{\Sigma} \leftrightarrow$ dependence between features),
- $X^j \sim \mathcal{N}_n(\mu^j, \Sigma_{jj}\mathbf{U})$ ($\mathbf{U} \leftrightarrow$ dependence between observations).

Arbitrary dependence setting

General strategy (I)

Framework

Consider the model

$$\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma}) \Leftrightarrow \text{vec}(\mathbf{X}) \sim \mathcal{N}_{np}(\text{vec}(\boldsymbol{\mu}), \boldsymbol{\Sigma} \otimes \mathbf{U}), \qquad \text{(dep)}$$

where $\mathbf{U} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ are positive definite. That means :

- $X_i \sim \mathcal{N}_p(\mu_i, U_{ii}\boldsymbol{\Sigma})$ ($\boldsymbol{\Sigma} \leftrightarrow$ dependence between features),
- $X^j \sim \mathcal{N}_n(\mu^j, \Sigma_{jj}\mathbf{U})$ ($\mathbf{U} \leftrightarrow$ dependence between observations).

In this model, $\mathbf{X}^T\nu \sim \mathcal{N}_p(0_p, \mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2})$ under (H0), where $\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2} = \nu^T\mathbf{U}\nu\boldsymbol{\Sigma}$.

Arbitrary dependence setting

General strategy (I)

Framework

Consider the model

$$\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma}) \Leftrightarrow \text{vec}(\mathbf{X}) \sim \mathcal{N}_{np}(\text{vec}(\boldsymbol{\mu}), \boldsymbol{\Sigma} \otimes \mathbf{U}), \qquad \text{(dep)}$$

where $\mathbf{U} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ are positive definite. That means :

- $X_i \sim \mathcal{N}_p(\mu_i, U_{ii}\boldsymbol{\Sigma})$ ($\boldsymbol{\Sigma} \leftrightarrow$ dependence between features),
- $X^j \sim \mathcal{N}_n(\mu^j, \Sigma_{jj}\mathbf{U})$ ($\mathbf{U} \leftrightarrow$ dependence between observations).

In this model, $\mathbf{X}^T\nu \sim \mathcal{N}_p(0_p, \mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2})$ under (H0), where $\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2} = \nu^T\mathbf{U}\nu\boldsymbol{\Sigma}$.
Therefore, $\|\mathbf{X}^T\nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} \sim \chi_p$ under (H0), where $\|v\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} = \sqrt{v^T\mathbf{V}^{-1}_{\mathcal{G}_1,\mathcal{G}_2}v}$, $v \in \mathbb{R}^p$.

Arbitrary dependence setting

General strategy (I)

Framework

Consider the model

$$\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma}) \Leftrightarrow \text{vec}(\mathbf{X}) \sim \mathcal{N}_{np}(\text{vec}(\boldsymbol{\mu}), \boldsymbol{\Sigma} \otimes \mathbf{U}), \qquad \text{(dep)}$$

where $\mathbf{U} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ are positive definite. That means :

- $X_i \sim \mathcal{N}_p(\mu_i, U_{ii}\boldsymbol{\Sigma})$ ($\boldsymbol{\Sigma} \leftrightarrow$ dependence between features),
- $X^j \sim \mathcal{N}_n(\mu^j, \Sigma_{jj}\mathbf{U})$ ($\mathbf{U} \leftrightarrow$ dependence between observations).

In this model, $\mathbf{X}^T\nu \sim \mathcal{N}_p(0_p, \mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2})$ under (H0), where $\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2} = \nu^T\mathbf{U}\nu\boldsymbol{\Sigma}$.
Therefore, $\|\mathbf{X}^T\nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} \sim \chi_p$ under (H0), where $\|v\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} = \sqrt{v^T\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}^{-1}v}$, $v \in \mathbb{R}^p$.

Candidate p-value for (H0) under (dep)

$$p_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}; \{\mathcal{G}_1,\mathcal{G}_2\}) = \mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\Big(\|\mathbf{X}^T\nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} \geq \|\mathbf{x}^T\nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} \;\Big|\; \mathcal{G}_1,\mathcal{G}_2 \in \mathcal{C}(\mathbf{X}),$$
$$\boldsymbol{\pi}_\nu^\perp\mathbf{X} = \boldsymbol{\pi}_\nu^\perp\mathbf{x}\,,\ \mathrm{dir}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{X}^T\nu) = \mathrm{dir}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}^T\nu)\Big)$$

Arbitrary dependence setting

General strategy (II)

Candidate p-value for (H0) under (dep)

$$p_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = \mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\Big(\|\mathbf{X}^T\nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} \geq \|\mathbf{x}^T\nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} \;\Big|\; \|\mathbf{X}^T\nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} \in \mathcal{S}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}),\ \boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x},\ \mathrm{dir}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{X}^T\nu) = \mathrm{dir}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}^T\nu) \Big).$$

Arbitrary dependence setting

General strategy (II)

Candidate p-value for (H0) under (dep)

$$p_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = \mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\Big(\|\mathbf{X}^T\nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} \geq \|\mathbf{x}^T\nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} \;\Big|\; \|\mathbf{X}^T\nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} \in \mathcal{S}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}),\; \boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x},\; \mathrm{dir}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{X}^T\nu) = \mathrm{dir}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}^T\nu) \Big).$$

For any $p \times p$ symmetric positive definite matrix $\mathbf{A}$, let $\|v\|_{\mathbf{A}}^2 = v^T\mathbf{A}^{-1}v$ and $\mathrm{dir}_{\mathbf{A}}(v) = v/\|v\|_{\mathbf{A}}\mathbb{1}\{v \neq 0\}$ for all $v \in \mathbb{R}^p$.

Arbitrary dependence setting

General strategy (II)

Candidate p-value for (H0) under (dep)

$$p_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = \mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}} \Big(\|\mathbf{X}^T \nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} \geq \|\mathbf{x}^T \nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} \;\Big|\; \|\mathbf{X}^T \nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} \in \mathcal{S}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}),\; \boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x},\; \mathrm{dir}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{X}^T \nu) = \mathrm{dir}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}^T \nu) \Big).$$

For any $p \times p$ symmetric positive definite matrix $\mathbf{A}$, let $\|v\|_{\mathbf{A}}^2 = v^T \mathbf{A}^{-1} v$ and $\mathrm{dir}_{\mathbf{A}}(v) = v / \|v\|_{\mathbf{A}} \mathbb{1}\{v \neq 0\}$ for all $v \in \mathbb{R}^p$.

Proposition

(i) $\mathbf{A} = c\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}$ for some $c > 0 \overset{\text{(H0)}}{\Leftrightarrow} ||\mathbf{X}^T \nu_{\mathcal{G}_1,\mathcal{G}_2}||_{\mathbf{A}} \perp\!\!\!\perp \mathrm{dir}_{\mathbf{A}} \left(\mathbf{X}^T \nu_{\mathcal{G}_1,\mathcal{G}_2}\right)$,

Arbitrary dependence setting

General strategy (II)

Candidate p-value for (H0) under (dep)

$$p_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x};\{\mathcal{G}_1,\mathcal{G}_2\}) = \mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\Big(\|\mathbf{X}^T\nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} \geq \|\mathbf{x}^T\nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} \;\Big|\; \|\mathbf{X}^T\nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} \in \mathcal{S}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x};\{\mathcal{G}_1,\mathcal{G}_2\}),\ \boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x},\ \mathrm{dir}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{X}^T\nu) = \mathrm{dir}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}^T\nu)\Big).$$

For any $p \times p$ symmetric positive definite matrix $\mathbf{A}$, let $\|v\|_{\mathbf{A}}^2 = v^T\mathbf{A}^{-1}v$ and $\mathrm{dir}_{\mathbf{A}}(v) = v/\|v\|_{\mathbf{A}}\mathbb{1}\{v \neq 0\}$ for all $v \in \mathbb{R}^p$.

Proposition

(i) $\mathbf{A} = c\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}$ for some $c > 0 \overset{(H0)}{\Leftrightarrow} ||\mathbf{X}^T\nu_{\mathcal{G}_1,\mathcal{G}_2}||_{\mathbf{A}} \perp\!\!\!\perp \mathrm{dir}_{\mathbf{A}}\left(\mathbf{X}^T\nu_{\mathcal{G}_1,\mathcal{G}_2}\right)$,

(ii) $\mathbf{X}^T\nu_{\mathcal{G}_1,\mathcal{G}_2} \perp\!\!\!\perp \boldsymbol{\pi}^\perp_{\nu_{\mathcal{G}_1,\mathcal{G}_2}}\mathbf{X}$ for all $(\mathcal{G}_1,\mathcal{G}_2) \Leftrightarrow \mathbf{U} \in \mathcal{CS}(n)$,

Arbitrary dependence setting

General strategy (II)

Candidate p-value for (H0) under (dep)

$$p_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x};\{\mathcal{G}_1,\mathcal{G}_2\}) = \mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\Big(\|\mathbf{X}^T\nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} \geq \|\mathbf{x}^T\nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} \ \Big| \ \|\mathbf{X}^T\nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} \in \mathcal{S}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x};\{\mathcal{G}_1,\mathcal{G}_2\}),\ \boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x},\ \mathrm{dir}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{X}^T\nu) = \mathrm{dir}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}^T\nu)\Big).$$

For any $p \times p$ symmetric positive definite matrix $\mathbf{A}$, let $\|v\|_{\mathbf{A}}^2 = v^T\mathbf{A}^{-1}v$ and $\mathrm{dir}_{\mathbf{A}}(v) = v/\|v\|_{\mathbf{A}}\mathbb{1}\{v \neq 0\}$ for all $v \in \mathbb{R}^p$.

Proposition

(*i*) $\mathbf{A} = c\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}$ for some $c > 0 \overset{(H0)}{\Leftrightarrow} ||\mathbf{X}^T\nu_{\mathcal{G}_1,\mathcal{G}_2}||_{\mathbf{A}} \perp\!\!\!\perp \mathrm{dir}_{\mathbf{A}}\left(\mathbf{X}^T\nu_{\mathcal{G}_1,\mathcal{G}_2}\right)$,

(*ii*) $\mathbf{X}^T\nu_{\mathcal{G}_1,\mathcal{G}_2} \perp\!\!\!\perp \boldsymbol{\pi}^\perp_{\nu_{\mathcal{G}_1,\mathcal{G}_2}}\mathbf{X}$ for all $(\mathcal{G}_1,\mathcal{G}_2) \Leftrightarrow \mathbf{U} \in \mathcal{CS}(n)$,

where $\mathcal{CS}(n)$ is the class of compound symmetry positive definite matrices :

$$\mathcal{CS}(n) = \left\{(a-b)\mathbf{I}_n + b\mathbf{1}_{n\times n} \ : \ a \geq 0,\ -\frac{a}{n-1} < b < a\right\}.$$

Arbitrary dependence setting

Post-clustering inference for $\mathbf{U} \in \mathcal{CS}(n)$

The direct extension of the strategy of Gao *et al.* to the general model (dep) imposes a compound symmetry constraint on the dependence between observations.

Arbitrary dependence setting

Post-clustering inference for $\mathbf{U} \in \mathcal{CS}(n)$

The direct extension of the strategy of Gao *et al.* to the general model (dep) imposes a compound symmetry constraint on the dependence between observations.

Theorem

Let $\mathcal{C}$ be a clustering algorithm and $\mathbf{x}$ a realization of $\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma})$ with $\mathbf{U} \in \mathcal{CS}(n)$. Then,

$$p_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = 1 - \mathbb{F}_p\left(\|\mathbf{x}^T \nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}, \mathcal{S}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}, \{\mathcal{G}_1, \mathcal{G}_2\})\right) \qquad \text{(p-tract)}$$

where $\mathbb{F}_p(t, \mathcal{S})$ is the cumulative distribution function of a χ_p random variable truncated to the set $\mathcal{S}$.

Arbitrary dependence setting

Post-clustering inference for $\mathbf{U} \in \mathcal{CS}(n)$

The direct extension of the strategy of Gao *et al.* to the general model (dep) imposes a compound symmetry constraint on the dependence between observations.

Theorem

Let $\mathcal{C}$ be a clustering algorithm and $\mathbf{x}$ a realization of $\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma})$ with $\mathbf{U} \in \mathcal{CS}(n)$. Then,

$$p_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = 1 - \mathbb{F}_p\Big(\|\mathbf{x}^T\nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}, \underbrace{\boxed{\mathcal{S}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}, \{\mathcal{G}_1, \mathcal{G}_2\})}}_{\text{Scale trans. of } \mathcal{S}_2}\Big) \qquad \text{(p-tract)}$$

where $\mathbb{F}_p(t, \mathcal{S})$ is the cumulative distribution function of a χ_p random variable truncated to the set $\mathcal{S}$.

Arbitrary dependence setting

Post-clustering inference for $\mathbf{U} \in \mathcal{CS}(n)$

The direct extension of the strategy of Gao *et al.* to the general model (dep) imposes a compound symmetry constraint on the dependence between observations.

Theorem

Let $\mathcal{C}$ be a clustering algorithm and $\mathbf{x}$ a realization of $\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma})$ with $\mathbf{U} \in \mathcal{CS}(n)$. Then,

$$p_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = 1 - \mathbb{F}_p\Big(\|\mathbf{x}^T\nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}, \underbrace{\boxed{\mathcal{S}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}, \{\mathcal{G}_1, \mathcal{G}_2\})}}_{\text{Scale trans. of } \mathcal{S}_2}\Big) \qquad \text{(p-tract)}$$

where $\mathbb{F}_p(t, \mathcal{S})$ is the cumulative distribution function of a χ_p random variable truncated to the set $\mathcal{S}$.

- The control of the sel. type I error is robust to moderate deviations of $\mathbf{U} \in \mathcal{CS}(n)$.

What if $\mathbf{U} \notin \mathcal{CS}(n)$?

What if $\mathbf{U} \notin \mathcal{CS}(n)$?

- The projection $\boldsymbol{\pi}_{\nu}^{\perp}\mathbf{X}$ is not independent of $\mathbf{X}^T\nu$ in general.

What if $\mathbf{U} \notin \mathcal{CS}(n)$?

- The projection $\boldsymbol{\pi}_\nu^\perp \mathbf{X}$ is not independent of $\mathbf{X}^T\nu$ in general.
- Therefore, the distribution of interest is not that of $\mathbf{X}^T\nu$, but rather that of the conditioned vector :

$$\bar{\mathbf{X}}_\nu(\mathbf{x}) := \mathbf{X}^T\nu \,\big|\, \{\boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x},\, \mathrm{dir}(\mathbf{X}^T\nu) = \pm\mathrm{dir}(\mathbf{x}^T\nu)\}, \quad \text{for } \mathbf{x} \in \mathbb{R}^{n\times p},$$

What if $\mathbf{U} \notin \mathcal{CS}(n)$?

- The projection $\boldsymbol{\pi}_\nu^\perp \mathbf{X}$ is not independent of $\mathbf{X}^T\nu$ in general.
- Therefore, the distribution of interest is not that of $\mathbf{X}^T\nu$, but rather that of the conditioned vector :

$$\bar{\mathbf{X}}_\nu(\mathbf{x}) := \mathbf{X}^T\nu \,\big|\, \{\boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x},\, \mathrm{dir}(\mathbf{X}^T\nu) = \pm\mathrm{dir}(\mathbf{x}^T\nu)\}, \quad \text{for } \mathbf{x} \in \mathbb{R}^{n\times p},$$

Theorem

Let $\mathcal{C}$ be a clustering algorithm and $\mathbf{x}$ a realization of $\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma})$. Then,

$$\bar{\mathbf{X}}_\nu(\mathbf{x}) \sim \mathcal{N}_p\left(0, \boldsymbol{\Gamma}_\mathbf{x}\right),$$

under (H0),

What if $\mathbf{U} \notin \mathcal{CS}(n)$?

- The projection $\boldsymbol{\pi}_\nu^\perp \mathbf{X}$ is not independent of $\mathbf{X}^T\nu$ in general.
- Therefore, the distribution of interest is not that of $\mathbf{X}^T\nu$, but rather that of the conditioned vector :

$$\bar{\mathbf{X}}_\nu(\mathbf{x}) := \mathbf{X}^T\nu \,\big|\, \{\boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x},\, \mathrm{dir}(\mathbf{X}^T\nu) = \pm\mathrm{dir}(\mathbf{x}^T\nu)\}, \quad \text{for } \mathbf{x} \in \mathbb{R}^{n\times p},$$

Theorem

Let $\mathcal{C}$ be a clustering algorithm and $\mathbf{x}$ a realization of $\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma})$. Then,

$$\bar{\mathbf{X}}_\nu(\mathbf{x}) \sim \mathcal{N}_p\left(0, \boldsymbol{\Gamma}_\mathbf{x}\right),$$

under (H0), where

$$\boldsymbol{\Gamma}_\mathbf{x} = (\mathbf{I}_p \otimes \nu^T)(\boldsymbol{\Sigma}\otimes\mathbf{U} - (\boldsymbol{\Sigma}\otimes\mathbf{U})\mathbf{A}_\mathbf{x}{}^\top(\mathbf{A}_\mathbf{x}(\boldsymbol{\Sigma}\otimes\mathbf{U})\mathbf{A}_\mathbf{x}{}^\top)^\dagger\mathbf{A}_\mathbf{x}(\boldsymbol{\Sigma}\otimes\mathbf{U}))(\mathbf{I}_p\otimes\nu),$$

$$\mathbf{A}_\mathbf{x} = \begin{bmatrix} \boldsymbol{\pi}_{\mathbf{x}_\nu}^\perp(\mathbf{I}_p \otimes \boldsymbol{\pi}_\nu) \\ \mathbf{I}_p \otimes \boldsymbol{\pi}_\nu^\perp \end{bmatrix},$$

$$\boldsymbol{\pi}_\nu = \mathbf{I}_n - \boldsymbol{\pi}_\nu^\perp,\ \mathbf{x}_\nu = \mathrm{vec}(\boldsymbol{\pi}_\nu\mathbf{x}) \quad \text{and} \quad \boldsymbol{\pi}_{\mathbf{x}_\nu}^\perp = \mathbf{I}_{np} - \mathbf{x}_\nu^T\mathbf{x}_\nu/\|\mathbf{x}_\nu\|_2^2.$$

Arbitrary $\mathbf{U}$ structures

Candidate p-value for (H0) under arbitrary $\mathbf{U}$

$$p_{\Gamma}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = \mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\Big(\|\mathbf{X}^T\nu\|_{\Gamma_{\mathbf{x}}} \geq \|\mathbf{x}^T\nu\|_{\Gamma_{\mathbf{x}}} \;\Big|\; \mathcal{G}_1, \mathcal{G}_2 \in \mathcal{C}(\mathbf{X}),$$
$$\boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x}, \, \mathrm{dir}(\mathbf{X}^T\nu) = \pm\mathrm{dir}(\mathbf{x}^T\nu)\Big),$$

where $\|v\|^2_{\Gamma_{\mathbf{x}}} = v^T \Gamma_{\mathbf{x}}^{\dagger} v, \quad \forall\, v \in \mathbb{R}^p.$

Arbitrary $\mathbf{U}$ structures

Candidate p-value for (H0) under arbitrary $\mathbf{U}$

$$p_{\boldsymbol{\Gamma}}(\mathbf{x};\{\mathcal{G}_1,\mathcal{G}_2\}) = \mathbb{P}_{H_0^{\{\mathcal{G}_1,\mathcal{G}_2\}}}\Big(\|\mathbf{X}^T\nu\|_{\boldsymbol{\Gamma}_\mathbf{x}} \geq \|\mathbf{x}^T\nu\|_{\boldsymbol{\Gamma}_\mathbf{x}} \;\Big|\; \mathcal{G}_1,\mathcal{G}_2 \in \mathcal{C}(\mathbf{X}),$$
$$\boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x},\, \mathrm{dir}(\mathbf{X}^T\nu) = \pm\mathrm{dir}(\mathbf{x}^T\nu)\Big),$$

where $\|v\|_{\boldsymbol{\Gamma}_\mathbf{x}}^2 = v^T \boldsymbol{\Gamma}_\mathbf{x}^\dagger v, \quad \forall\, v \in \mathbb{R}^p$.

Proposition

The quantity $\|\bar{\mathbf{X}}_\nu(\mathbf{x})\|_{\boldsymbol{\Gamma}_\mathbf{x}}$ follows $\mathbf{x}$-a.s. a χ_1 distribution under (H0). Moreover,

$$p_{\boldsymbol{\Gamma}}(\mathbf{x};\{\mathcal{G}_1,\mathcal{G}_2\}) = 1 - \mathbb{F}_1\left(\|\mathbf{x}^T\nu\|_{\boldsymbol{\Gamma}_\mathbf{x}},\, \mathcal{S}_{\boldsymbol{\Gamma}_\mathbf{x}}(\mathbf{x};\{\mathcal{G}_1,\mathcal{G}_2\})\right),$$

where $\mathbb{F}_1(t,\mathcal{S})$ is the cumulative distribution function of a χ_1 random variable truncated to the set $\mathcal{S}$ and $\mathcal{S}_{\boldsymbol{\Gamma}_\mathbf{x}}(\mathbf{x};\{\mathcal{G}_1,\mathcal{G}_2\})$ is a scale transformation of $\mathcal{S}_2$.

Arbitrary $\mathbf{U}$ structures

Candidate p-value for (H0) under arbitrary $\mathbf{U}$

$$p_{\boldsymbol{\Gamma}}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = \mathbb{P}_{H_0^{\{\mathcal{G}_1, \mathcal{G}_2\}}}\Big(\|\mathbf{X}^T\nu\|_{\boldsymbol{\Gamma}_\mathbf{x}} \geq \|\mathbf{x}^T\nu\|_{\boldsymbol{\Gamma}_\mathbf{x}} \;\Big|\; \mathcal{G}_1, \mathcal{G}_2 \in \mathcal{C}(\mathbf{X}),$$
$$\boldsymbol{\pi}_\nu^\perp \mathbf{X} = \boldsymbol{\pi}_\nu^\perp \mathbf{x},\, \mathrm{dir}(\mathbf{X}^T\nu) = \pm\mathrm{dir}(\mathbf{x}^T\nu) \Big),$$

where $\|v\|^2_{\boldsymbol{\Gamma}_\mathbf{x}} = v^T \boldsymbol{\Gamma}_\mathbf{x}^\dagger v, \quad \forall\, v \in \mathbb{R}^p$.

Proposition

The quantity $\|\bar{\mathbf{X}}_\nu(\mathbf{x})\|_{\boldsymbol{\Gamma}_\mathbf{x}}$ follows $\mathbf{x}$-a.s. a χ_1 distribution under (H0). Moreover,

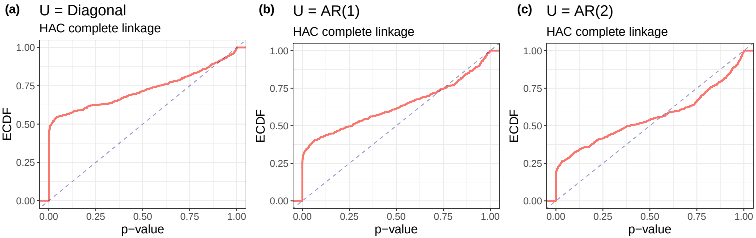
$$p_{\boldsymbol{\Gamma}}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = 1 - \mathbb{F}_1\left(\|\mathbf{x}^T\nu\|_{\boldsymbol{\Gamma}_\mathbf{x}}, \mathcal{S}_{\boldsymbol{\Gamma}_\mathbf{x}}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) \right),$$

where $\mathbb{F}_1(t, \mathcal{S})$ is the cumulative distribution function of a χ_1 random variable truncated to the set $\mathcal{S}$ and $\mathcal{S}_{\boldsymbol{\Gamma}_\mathbf{x}}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\})$ is a scale transformation of $\mathcal{S}_2$.

- Assessing whether $p_{\boldsymbol{\Gamma}}(\mathbf{X}; \{\mathcal{G}_1, \mathcal{G}_2\})$ controls the selective type I error is a challenging problem, as it requires understanding the behavior of the null distribution of $\|\bar{\mathbf{X}}_\nu(\mathbf{x})\|^2_{\boldsymbol{\Gamma}_\mathbf{x}} = \bar{\mathbf{X}}_\nu(\mathbf{x})^T \boldsymbol{\Gamma}_\mathbf{X}^\dagger \bar{\mathbf{X}}_\nu(\mathbf{x})$.

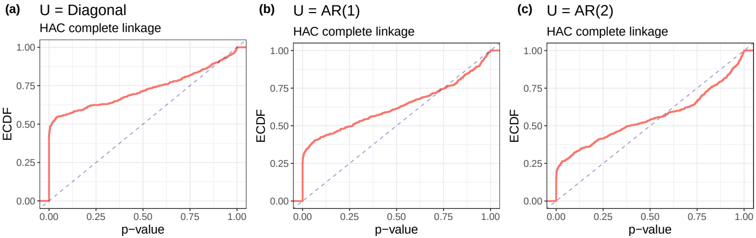
Arbitrary **U** structures

Numerical simulations suggest the unsuitability of $p_{\boldsymbol{\Gamma}}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\})$



Arbitrary $\mathbf{U}$ structures

Numerical simulations suggest the unsuitability of $p_{\mathbf{\Gamma}}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\})$



Conclusion

Defining a tractable p-value that ensures the selective type I error control requires the conditioning on events that are *independent* of the test statistic.

Estimation of unknown parameters

Independence setting (Gao *et al.* 2022)

Estimation of unknown parameters

Independence setting (Gao *et al.* 2022)

Let $\mathbf{X}^{(n)} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}^{(n)}, \mathbf{I}_n, \sigma^2 \mathbf{I}_p)$ and consider

$$\hat{p}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = 1 - \mathbb{F}_p\Big(\|\mathbf{x}^T \nu\|_2; \hat{\sigma}\|\nu\|_2, \mathcal{S}_2(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\})\Big).$$

Estimation of unknown parameters

Independence setting (Gao *et al.* 2022)

Let $\mathbf{X}^{(n)} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}^{(n)}, \mathbf{I}_n, \sigma^2 \mathbf{I}_p)$ and consider

$$\hat{p}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = 1 - \mathbb{F}_p\Big(\|\mathbf{x}^T \nu\|_2; \hat{\sigma}\|\nu\|_2, \mathcal{S}_2(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\})\Big).$$

Theorem 4 in Gao *et al.* 2022
If $\hat{\sigma}$ is an estimator of σ such that

$$\lim_{n\to\infty} \mathbb{P}_{H_0^{\{\mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)}\}}}\Big(\hat{\sigma}\big(\mathbf{X}^{(n)}\big) \geq \sigma \,\Big|\, \mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)} \in \mathcal{C}\big(\mathbf{X}^{(n)}\big)\Big) = 1, \qquad (\sigma \text{ over-est})$$

Estimation of unknown parameters

Independence setting (Gao *et al.* 2022)

Let $\mathbf{X}^{(n)} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}^{(n)}, \mathbf{I}_n, \sigma^2 \mathbf{I}_p)$ and consider

$$\hat{p}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = 1 - \mathbb{F}_p\Big(\|\mathbf{x}^T \nu\|_2; \hat{\sigma}\|\nu\|_2, \mathcal{S}_2(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) \Big).$$

Theorem 4 in Gao *et al.* 2022

If $\hat{\sigma}$ is an estimator of σ such that

$$\lim_{n\to\infty} \mathbb{P}_{H_0^{\{\mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)}\}}} \left(\hat{\sigma}\big(\mathbf{X}^{(n)}\big) \geq \sigma \,\middle|\, \mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)} \in \mathcal{C}\big(\mathbf{X}^{(n)}\big) \right) = 1, \qquad (\sigma \text{ over-est})$$

then, for any $\alpha \in [0, 1]$, it holds

$$\limsup_{n\to\infty} \mathbb{P}_{H_0^{\{\mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)}\}}} \left(\hat{p}\big(\mathbf{X}^{(n)}; \{\mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)}\}\big) \leq \alpha \,\middle|\, \mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)} \in \mathcal{C}\big(\mathbf{X}^{(n)}\big) \right) \leq \alpha.$$

Estimation of unknown parameters

Independence setting (Gao *et al.* 2022)

Let $\mathbf{X}^{(n)} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}^{(n)}, \mathbf{I}_n, \sigma^2\mathbf{I}_p)$ and consider

$$\hat{p}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = 1 - \mathbb{F}_p\Big(\|\mathbf{x}^T\nu\|_2; \hat{\sigma}\|\nu\|_2, \mathcal{S}_2(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\})\Big).$$

Theorem 4 in Gao *et al.* 2022

If $\hat{\sigma}$ is an estimator of σ such that

$$\lim_{n\to\infty} \mathbb{P}_{H_0^{\{\mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)}\}}}\left(\hat{\sigma}\big(\mathbf{X}^{(n)}\big) \geq \sigma \,\middle|\, \mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)} \in \mathcal{C}\big(\mathbf{X}^{(n)}\big)\right) = 1, \qquad (\sigma \text{ over-est})$$

then, for any $\alpha \in [0,1]$, it holds

$$\limsup_{n\to\infty} \mathbb{P}_{H_0^{\{\mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)}\}}}\left(\hat{p}\big(\mathbf{X}^{(n)}; \{\mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)}\}\big) \leq \alpha \,\middle|\, \mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)} \in \mathcal{C}\big(\mathbf{X}^{(n)}\big)\right) \leq \alpha.$$

$\rightarrow$ Gao *et al.* propose an estimator $\hat{\sigma}$ that satisfies (σ over-est) under mild assumptions on $\{\boldsymbol{\mu}^{(n)}\}_{n\in\mathbb{N}}$.

Estimation of unknown parameters

Arbitrary dependence setting

Let

$$\mathbf{X} \sim \mathcal{MN}_{n \times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma}). \tag{dep}$$

Can we estimate both $\mathbf{U}$ and $\boldsymbol{\Sigma}$?

Estimation of unknown parameters

Arbitrary dependence setting

Let

$$\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma}). \tag{dep}$$

Can we estimate both $\mathbf{U}$ and $\boldsymbol{\Sigma}$?

- Under the general model (dep), the scale matrices $\mathbf{U}$ and $\boldsymbol{\Sigma}$ are non-identifiable.

Estimation of unknown parameters

Arbitrary dependence setting

Let

$$\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma}). \tag{dep}$$

Can we estimate both $\mathbf{U}$ and $\boldsymbol{\Sigma}$?

- Under the general model (dep), the scale matrices $\mathbf{U}$ and $\boldsymbol{\Sigma}$ are non-identifiable.
- Multiple copies of $\mathbf{X}$ are needed to simultaneously estimate $\mathbf{U}$ and $\boldsymbol{\Sigma}$.

Estimation of unknown parameters

Arbitrary dependence setting

Let

$$\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma}). \tag{dep}$$

Can we estimate both $\mathbf{U}$ and $\boldsymbol{\Sigma}$?

- Under the general model (dep), the scale matrices $\mathbf{U}$ and $\boldsymbol{\Sigma}$ are non-identifiable.
- Multiple copies of $\mathbf{X}$ are needed to simultaneously estimate $\mathbf{U}$ and $\boldsymbol{\Sigma}$.
- We assume that either $\mathbf{U}$ or $\boldsymbol{\Sigma}$ is known :

$$\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma}) \Leftrightarrow \mathbf{X}^T \sim \mathcal{MN}_{p\times n}(\boldsymbol{\mu}^T, \boldsymbol{\Sigma}, \mathbf{U}).$$

Estimation of unknown parameters

Arbitrary dependence setting

Let

$$\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma}). \qquad \text{(dep)}$$

Can we estimate both $\mathbf{U}$ and $\boldsymbol{\Sigma}$?

- Under the general model (dep), the scale matrices $\mathbf{U}$ and $\boldsymbol{\Sigma}$ are non-identifiable.
- Multiple copies of $\mathbf{X}$ are needed to simultaneously estimate $\mathbf{U}$ and $\boldsymbol{\Sigma}$.
- We assume that either $\mathbf{U}$ or $\boldsymbol{\Sigma}$ is known :

$$\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma}) \Leftrightarrow \mathbf{X}^T \sim \mathcal{MN}_{p\times n}(\boldsymbol{\mu}^T, \boldsymbol{\Sigma}, \mathbf{U}).$$

$\rightarrow$ How to extend the notion of over-estimation to matrices ?

Estimation of unknown parameters

Arbitrary dependence setting

Let

$$\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma}). \qquad \text{(dep)}$$

Can we estimate both $\mathbf{U}$ and $\boldsymbol{\Sigma}$?

- Under the general model (dep), the scale matrices $\mathbf{U}$ and $\boldsymbol{\Sigma}$ are non-identifiable.
- Multiple copies of $\mathbf{X}$ are needed to simultaneously estimate $\mathbf{U}$ and $\boldsymbol{\Sigma}$.
- We assume that either $\mathbf{U}$ or $\boldsymbol{\Sigma}$ is known :

$$\mathbf{X} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma}) \Leftrightarrow \mathbf{X}^T \sim \mathcal{MN}_{p\times n}(\boldsymbol{\mu}^T, \boldsymbol{\Sigma}, \mathbf{U}).$$

$\rightarrow$ How to extend the notion of over-estimation to matrices ?

Loewner partial order $\succeq$

Let A, B be two Hermitian matrices. $A \succeq B$ if and only if $A - B$ is positive semidefinite (PSD).

Remark : If $A = \hat{\sigma}\mathbf{I}_p$ and $B = \sigma\mathbf{I}_p$, the condition $A \succeq B$ becomes $\hat{\sigma} \geq \sigma$.

Over-estimation of $\mathbf{\Sigma}$ for known $\mathbf{U}$

Let $\mathbf{X}^{(n)} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}^{(n)}, \mathbf{U}^{(n)}, \mathbf{\Sigma})$ with $\mathbf{U}^{(n)} \in \mathcal{CS}(n)$ and consider

$$p_{\hat{\mathbf{V}}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = 1 - \mathbb{F}_p\left(\|\mathbf{x}^T \nu\|_{\hat{\mathbf{V}}_{\mathcal{G}_1,\mathcal{G}_2}}; \mathcal{S}_{\hat{\mathbf{V}}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}, \{\mathcal{G}_1, \mathcal{G}_2\}) \right)$$

where $\hat{\mathbf{V}}_{\mathcal{G}_1,\mathcal{G}_2} = \nu^T \mathbf{U} \nu \hat{\mathbf{\Sigma}}(\mathbf{x})$.

Over-estimation of $\mathbf{\Sigma}$ for known $\mathbf{U}$

Let $\mathbf{X}^{(n)} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}^{(n)}, \mathbf{U}^{(n)}, \mathbf{\Sigma})$ with $\mathbf{U}^{(n)} \in \mathcal{CS}(n)$ and consider

$$p_{\hat{\mathbf{V}}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = 1 - \mathbb{F}_p\left(\|\mathbf{x}^T\nu\|_{\hat{\mathbf{V}}_{\mathcal{G}_1,\mathcal{G}_2}}; \mathcal{S}_{\hat{\mathbf{V}}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}, \{\mathcal{G}_1, \mathcal{G}_2\})\right)$$

where $\hat{\mathbf{V}}_{\mathcal{G}_1,\mathcal{G}_2} = \nu^T\mathbf{U}\nu\hat{\mathbf{\Sigma}}(\mathbf{x})$.

Theorem (extension of Theorem 4 in Gao *et al.* 2022)

Let $\mathbf{U} \in \mathcal{CS}(n)$ and $\hat{\mathbf{\Sigma}}\left(\mathbf{X}^{(n)}\right)$ be a positive definite estimator of $\mathbf{\Sigma}$ such that

$$\lim_{n\to\infty} \mathbb{P}_{H_0^{\{\mathcal{G}_1^{(n)},\mathcal{G}_2^{(n)}\}}}\left(\hat{\mathbf{\Sigma}}(\mathbf{X}^{(n)}) \succeq \mathbf{\Sigma} \,\middle|\, \mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)} \in \mathcal{C}\left(\mathbf{X}^{(n)}\right)\right) = 1, \qquad (\mathbf{\Sigma}\text{ over-est})$$

Over-estimation of $\boldsymbol{\Sigma}$ for known $\mathbf{U}$

Let $\mathbf{X}^{(n)} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}^{(n)}, \mathbf{U}^{(n)}, \boldsymbol{\Sigma})$ with $\mathbf{U}^{(n)} \in \mathcal{CS}(n)$ and consider

$$p_{\hat{\mathbf{V}}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = 1 - \mathbb{F}_p\left(\|\mathbf{x}^T \nu\|_{\hat{\mathbf{V}}_{\mathcal{G}_1,\mathcal{G}_2}}; \mathcal{S}_{\hat{\mathbf{V}}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}, \{\mathcal{G}_1, \mathcal{G}_2\}) \right)$$

where $\hat{\mathbf{V}}_{\mathcal{G}_1,\mathcal{G}_2} = \nu^T \mathbf{U} \nu \hat{\boldsymbol{\Sigma}}(\mathbf{x})$.

Theorem (extension of Theorem 4 in Gao *et al.* 2022)

Let $\mathbf{U} \in \mathcal{CS}(n)$ and $\hat{\boldsymbol{\Sigma}}\left(\mathbf{X}^{(n)}\right)$ be a positive definite estimator of $\boldsymbol{\Sigma}$ such that

$$\lim_{n\to\infty} \mathbb{P}_{H_0^{\{\mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)}\}}} \left(\hat{\boldsymbol{\Sigma}}(\mathbf{X}^{(n)}) \succeq \boldsymbol{\Sigma} \,\middle|\, \mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)} \in \mathcal{C}(\mathbf{X}^{(n)}) \right) = 1, \qquad (\boldsymbol{\Sigma} \text{ over-est})$$

then, for any $\alpha \in [0, 1]$, we have

$$\limsup_{n\to\infty} \mathbb{P}_{H_0^{\{\mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)}\}}} \left(p_{\hat{\mathbf{V}}_{\mathcal{G}_1^{(n)},\mathcal{G}_2^{(n)}}}(\mathbf{X}^{(n)}; \{\mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)}\}) \leq \alpha \,\middle|\, \mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)} \in \mathcal{C}(\mathbf{X}^{(n)}) \right) \leq \alpha.$$

Over-estimation of $\mathbf{\Sigma}$ for known $\mathbf{U}$

Let $\mathbf{X}^{(n)} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}^{(n)}, \mathbf{U}^{(n)}, \mathbf{\Sigma})$ with $\mathbf{U}^{(n)} \in \mathcal{CS}(n)$ and consider

$$p_{\hat{\mathbf{V}}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}; \{\mathcal{G}_1, \mathcal{G}_2\}) = 1 - \mathbb{F}_p\left(\|\mathbf{x}^T \nu\|_{\hat{\mathbf{V}}_{\mathcal{G}_1,\mathcal{G}_2}}; \mathcal{S}_{\hat{\mathbf{V}}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}, \{\mathcal{G}_1, \mathcal{G}_2\})\right)$$

where $\hat{\mathbf{V}}_{\mathcal{G}_1,\mathcal{G}_2} = \nu^T \mathbf{U} \nu \hat{\mathbf{\Sigma}}(\mathbf{x})$.

Theorem (extension of Theorem 4 in Gao *et al.* 2022)

Let $\mathbf{U} \in \mathcal{CS}(n)$ and $\hat{\mathbf{\Sigma}}\left(\mathbf{X}^{(n)}\right)$ be a positive definite estimator of $\mathbf{\Sigma}$ such that

$$\lim_{n\to\infty} \mathbb{P}_{H_0^{\{\mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)}\}}}\left(\hat{\mathbf{\Sigma}}(\mathbf{X}^{(n)}) \succeq \mathbf{\Sigma} \,\middle|\, \mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)} \in \mathcal{C}\left(\mathbf{X}^{(n)}\right)\right) = 1, \qquad (\mathbf{\Sigma}\text{ over-est})$$

then, for any $\alpha \in [0, 1]$, we have

$$\limsup_{n\to\infty} \mathbb{P}_{H_0^{\{\mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)}\}}}\left(p_{\hat{\mathbf{V}}_{\mathcal{G}_1^{(n)},\mathcal{G}_2^{(n)}}}(\mathbf{X}^{(n)}; \{\mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)}\}) \leq \alpha \,\middle|\, \mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)} \in \mathcal{C}\left(\mathbf{X}^{(n)}\right)\right) \leq \alpha.$$

$\rightarrow$ We propose an estimator $\hat{\mathbf{\Sigma}}$ that satisfies ($\mathbf{\Sigma}$ over-est) under mild assumptions on $\{\boldsymbol{\mu}^{(n)}\}$ and for several common models of dependence $\{\mathbf{U}^{(n)}\}$.

Thank you for your attention !

- Preprint : https://arxiv.org/abs/2310.11822,
- R package `PCIdep` at https://github.com/gonzalez-delgado/PCIdep/.

References

- Independence setting : L. L. Gao, J. Bien, and D. Witten. Selective inference for hierarchical clustering. *Journal of the American Statistical Association*, 0(0) :1–11, 2022.
- Extension to k-means : Y. T. Chen and D. M. Witten. Selective inference for k-means clustering. *Journal of Machine Learning Research*, 24(152) :1–41, 2023.
- Extension to feature-level test : B. Hivert, D. Agniel, R. Thiébaut, and B. P. Hejblum. Post-clustering difference testing : Valid inference and practical considerations with applications to ecological and biological data. *Comput. Statist. Data Anal.*, 193 :107916, May 2024.
- Alternative estimation of σ : Y. Yun and R. Foygel Barber. Selective inference for clustering with unknown variance. arXiv.2301.12999, 2023.

https://gonzalez-delgado.github.io/

Truncation sets

Independence setting

$$\hat{\mathcal{S}}_2 = \{\phi \geq 0 \, : \, \mathcal{G}_1, \mathcal{G}_2 \in \mathcal{C}(\mathbf{x}'_2(\phi))\}, \tag{1}$$

$$\mathbf{x}'_2(\phi) = \mathbf{x} + \frac{\nu}{\|\nu\|_2^2} \left(\phi - \|\mathbf{x}^T\nu\|_2\right) \mathrm{dir}(\mathbf{x}^T\nu), \tag{2}$$

Arbitrary dependence setting

$$\hat{\mathcal{S}}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} = \left\{\phi \geq 0 \, : \, \mathcal{G}_1, \mathcal{G}_2 \in \mathcal{C}\left(\mathbf{x}'_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\phi)\right)\right\}, \tag{3}$$

$$\mathbf{x}'_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\phi) = \mathbf{x} + \frac{\nu}{\|\nu\|_2^2} \left(\phi - \|\mathbf{x}^T\nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}\right) \mathrm{dir}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}(\mathbf{x}^T\nu). \tag{4}$$

Lemma (scale transformation)

$$\hat{\mathcal{S}}_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}} = \frac{\|\mathbf{x}^T\nu\|_{\mathbf{V}_{\mathcal{G}_1,\mathcal{G}_2}}}{\|\mathbf{x}^T\nu\|_2} \hat{\mathcal{S}}_2 \tag{5}$$

Asymptotic over-estimator of $\boldsymbol{\Sigma}$

Let $\mathbf{X}^{(n)} \sim \mathcal{MN}_{n \times p}(\boldsymbol{\mu}^{(n)}, \mathbf{U}^{(n)}, \boldsymbol{\Sigma})$.

For a given estimator $\hat{\boldsymbol{\Sigma}}\left(\mathbf{X}^{(n)}\right)$ of $\boldsymbol{\Sigma}$, assessing whether $\hat{\boldsymbol{\Sigma}}\left(\mathbf{X}^{(n)}\right) \succeq \boldsymbol{\Sigma}$ asymptotically strongly depends on how the sequences $\{\boldsymbol{\mu}^{(n)}\}_{n \in \mathbb{N}}$ and $\{\mathbf{U}^{(n)}\}_{n \in \mathbb{N}}$ grow up to infinity.

Asymptotic over-estimator of $\mathbf{\Sigma}$

Let $\mathbf{X}^{(n)} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}^{(n)}, \mathbf{U}^{(n)}, \mathbf{\Sigma})$.

For a given estimator $\hat{\mathbf{\Sigma}}\left(\mathbf{X}^{(n)}\right)$ of $\mathbf{\Sigma}$, assessing whether $\hat{\mathbf{\Sigma}}\left(\mathbf{X}^{(n)}\right) \succeq \mathbf{\Sigma}$ asymptotically strongly depends on how the sequences $\{\boldsymbol{\mu}^{(n)}\}_{n\in\mathbb{N}}$ and $\{\mathbf{U}^{(n)}\}_{n\in\mathbb{N}}$ grow up to infinity.

Estimator candidate

$$\hat{\mathbf{\Sigma}} = \hat{\mathbf{\Sigma}}\left(\mathbf{X}\right) = \frac{1}{n-1}\left(\mathbf{X} - \bar{\mathbf{X}}\right)^T \mathbf{U}^{-1}\left(\mathbf{X} - \bar{\mathbf{X}}\right), \qquad \text{(estimator)}$$

where $\bar{\mathbf{X}}$ is a $n \times p$ matrix having as rows the mean across rows of $\mathbf{X}$.

Asymptotic over-estimator of $\boldsymbol{\Sigma}$

Let $\mathbf{X}^{(n)} \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}^{(n)}, \mathbf{U}^{(n)}, \boldsymbol{\Sigma})$.

For a given estimator $\hat{\boldsymbol{\Sigma}}\left(\mathbf{X}^{(n)}\right)$ of $\boldsymbol{\Sigma}$, assessing whether $\hat{\boldsymbol{\Sigma}}\left(\mathbf{X}^{(n)}\right) \succeq \boldsymbol{\Sigma}$ asymptotically strongly depends on how the sequences $\{\boldsymbol{\mu}^{(n)}\}_{n\in\mathbb{N}}$ and $\{\mathbf{U}^{(n)}\}_{n\in\mathbb{N}}$ grow up to infinity.

Estimator candidate

$$\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}(\mathbf{X}) = \frac{1}{n-1}\left(\mathbf{X}-\bar{\mathbf{X}}\right)^T \mathbf{U}^{-1}\left(\mathbf{X}-\bar{\mathbf{X}}\right), \qquad \text{(estimator)}$$

where $\bar{\mathbf{X}}$ is a $n \times p$ matrix having as rows the mean across rows of $\mathbf{X}$.

$\rightarrow$ Assumptions on $\{\boldsymbol{\mu}^{(n)}\}_{n\in\mathbb{N}}$ and $\{\mathbf{U}^{(n)}\}_{n\in\mathbb{N}}$ to ensure that (estimator) a.s. asymptotically overestimates $\boldsymbol{\Sigma}$?

Assumptions on $\boldsymbol{\mu}^{(n)}$

Assumptions 1 and 2 in Gao et al. 2022 *(Assumption 1)*

For all $n \in \mathbb{N}$, there are exactly K^* distinct mean vectors among the first n observations, i.e.

$$\left\{\mu_i^{(n)}\right\}_{i=1,\ldots,n} = \{\theta_1, \ldots, \theta_{K^*}\}.$$

Besides, the proportion of the first n observations that have mean vector θ_k converges to $\pi_k > 0$, i.e.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{\mu_i^{(n)} = \theta_k\} = \pi_k, \qquad \text{(as-1)}$$

for all $k \in \{1, \ldots, K^*\}$, where $\sum_{k=1}^{K^*} \pi_k = 1$.

Assumptions on $\boldsymbol{\mu}^{(n)}$

Assumptions 1 and 2 in Gao et al. 2022 *(Assumption 1)*

For all $n \in \mathbb{N}$, there are exactly K^* distinct mean vectors among the first n observations, i.e.

$$\left\{\mu_i^{(n)}\right\}_{i=1,\dots,n} = \{\theta_1, \dots, \theta_{K^*}\}.$$

Besides, the proportion of the first n observations that have mean vector θ_k converges to $\pi_k > 0$, i.e.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{\mu_i^{(n)} = \theta_k\} = \pi_k, \quad \text{(as-1)}$$

for all $k \in \{1, \dots, K^*\}$, where $\sum_{k=1}^{K^*} \pi_k = 1$.

$\diamond$ If $\mathbf{U}^{(n)} = \mathbf{I}_n$, this is the only requirement to ensure asymp. over-estimation of $\boldsymbol{\Sigma}$.

Assumptions on $\boldsymbol{\mu}^{(n)}$

Assumptions 1 and 2 in Gao et al. 2022 *(Assumption 1)*

For all $n \in \mathbb{N}$, there are exactly K^* distinct mean vectors among the first n observations, i.e.

$$\left\{\mu_i^{(n)}\right\}_{i=1,\dots,n} = \{\theta_1, \dots, \theta_{K^*}\}.$$

Besides, the proportion of the first n observations that have mean vector θ_k converges to $\pi_k > 0$, i.e.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{\mu_i^{(n)} = \theta_k\} = \pi_k, \qquad \text{(as-1)}$$

for all $k \in \{1, \dots, K^*\}$, where $\sum_{k=1}^{K^*} \pi_k = 1$.

⋄ If $\mathbf{U}^{(n)} = \mathbf{I}_n$, this is the only requirement to ensure asymp. over-estimation of $\boldsymbol{\Sigma}$.
⋄ For **general** $\mathbf{U}^{(n)}$, the quantities

$$\frac{1}{n} \sum_{l,s=1}^{n} \left(U^{(n)}\right)^{-1}_{ls} \mathbb{1}\{\mu_l^{(n)} = \theta_k\} \, \mathbb{1}\{\mu_s^{(n)} = \theta_{k'}\}$$

are also required to **converge with explicit limit** as n tends to infinity.

One more assumption on $\boldsymbol{\mu}^{(n)}$ for non-diagonal $\mathbf{U}^{(n)}$

Assumption on $\boldsymbol{\mu}^{(n)}$ for non-diagonal $\mathbf{U}^{(n)}$ *(Assumption 2)*

If $\mathbf{U}^{(n)}$ is non-diagonal for all $n \in \mathbb{N}$, for any $k, k' \in \{1, \dots, K^*\}$, the proportion of the first n observations at distance $r \geq 1$ in $\mathbf{X}^{(n)}$ having means θ_k and $\theta_{k'}$ converges, and its limit converges to $\pi_k \pi_{k'}$ when the lag r tends to infinity. More precisely,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n-r} \mathbb{1}\{\mu_i = \theta_k\} \, \mathbb{1}\{\mu_{i+r} = \theta_{k'}\} = \pi^r_{kk'} \underset{r\to\infty}{\longrightarrow} \pi_k \, \pi_{k'} . \qquad \text{(as-2)}$$

We are asking the proportion of pairs of observations having a given a pair of means to approach the product of individual proportions (as-1) when both observations are far away in $\mathbf{X}^{(n)}$.

Assumptions on the sequence $\{\mathbf{U}^{(n)}\}_{n\in\mathbb{N}}$

Assumption on $\{\mathbf{U}^{(n)}\}_{n\in\mathbb{N}}$ *(Assumption 3)*

Every superdiagonal of $\left(\mathbf{U}^{(n)}\right)^{-1}$ defines asymptotically a convergent sequence, whose limits sum up to a real value. More precisely, for any $i \in \mathbb{N}$ and any $r \geq 0$,

$$\lim_{n\to\infty} \left(U^{(n)}\right)^{-1}_{i\,i+r} = \Lambda_{i\,i+r}, \quad \text{where} \quad \lim_{i\to\infty} \Lambda_{i\,i+r} = \lambda_r \quad \text{and} \quad \sum_{r=0}^{\infty} \lambda_r = \lambda \in \mathbb{R}.$$

Moreover, for each $r \geq 0$, the sequence $\{\left(U^{(n)}\right)^{-1}_{i\,i+r}\}_{n\in\mathbb{N}}$ satisfies any of the following conditions :

(*i*) It is dominated by a summable sequence i.e. $\left|\left(U^{(n)}\right)^{-1}_{i\,i+r} - \Lambda_{i\,i+r}\right| \leq \alpha_i \ \forall\, n \in \mathbb{N}$, with $\{\alpha_i\}_{i=1}^{\infty} \in \ell_1$,

(*ii*) For each $i \in \mathbb{N}$, it is non-decreasing or non-increasing.

Some admissible dependence models for $\{\mathbf{U}^{(n)}\}_{n\in\mathbb{N}}$

Remark 1 (Diagonal)

Let $\mathbf{U}^{(n)} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. If the sequence $\{\lambda_n\}_{n\in\mathbb{N}}$ is convergent, then the sequence $\{\mathbf{U}^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption 3.

Some admissible dependence models for $\{\mathbf{U}^{(n)}\}_{n\in\mathbb{N}}$

Remark 1 (Diagonal)

Let $\mathbf{U}^{(n)} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. If the sequence $\{\lambda_n\}_{n\in\mathbb{N}}$ is convergent, then the sequence $\{\mathbf{U}^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption 3.

Remark 2 (Compound symmetry)

Let $a, b \in \mathbb{R}$ with $b \neq a \geq 0$. If $\mathbf{U}^{(n)} = b\mathbf{1}_{n\times n} + (a - b)\,\mathbf{I}_n$, where $\mathbf{1}_{n\times n}$ is a $n \times n$ matrix of ones, then $\{\mathbf{U}^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption 3.

Some admissible dependence models for $\{\mathbf{U}^{(n)}\}_{n\in\mathbb{N}}$

Remark 1 (Diagonal)

Let $\mathbf{U}^{(n)} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. If the sequence $\{\lambda_n\}_{n\in\mathbb{N}}$ is convergent, then the sequence $\{\mathbf{U}^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption 3.

Remark 2 (Compound symmetry)

Let $a, b \in \mathbb{R}$ with $b \neq a \geq 0$. If $\mathbf{U}^{(n)} = b\mathbf{1}_{n\times n} + (a-b)\,\mathbf{I}_n$, where $\mathbf{1}_{n\times n}$ is a $n \times n$ matrix of ones, then $\{\mathbf{U}^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption 3.

Remark 3 (AR(P))

Let $\mathbf{U}^{(n)}$ be the covariance matrix of an auto-regressive process of order $P \geq 1$ such that, if $P > 2$, $\beta_k\beta_{k'} \geq 0$ for all $k, k' \in \{1, \ldots, P\}$. Then, the sequence $\{\mathbf{U}^{(n)}\}_{n\in\mathbb{N}}$ satisfies Assumption 3.

Numerical simulations

Three dependence settings

(a) $\mathbf{U} = \mathbf{I}_n$ and $\boldsymbol{\Sigma}$ is the covariance matrix of an AR(1) model, i.e. $\Sigma_{ij} = \sigma^2 \rho^{|i-j|}$, with $\sigma = 1$ and $\rho = 0.5$.

(b) $\mathbf{U}$ is a compound symmetry covariance matrix, i.e. $\mathbf{U} = b + (a - b)\mathbf{I}_n$, with $a = 0.5$ and $b = 1$. $\boldsymbol{\Sigma}$ is a Toeplitz matrix, i.e. $\Sigma_{ij} = t(|i-j|)$, with $t(s) = 1 + 1/(1+s)$ for $s \in \mathbb{N}$.

(c) $\mathbf{U}$ is the covariance matrix of an AR(1) model with $\sigma = 1$ and $\rho = 0.1$. $\boldsymbol{\Sigma}$ is a diagonal matrix with diagonal entries given by $\Sigma_{ii} = 1 + 1/i$.

Numerical simulations

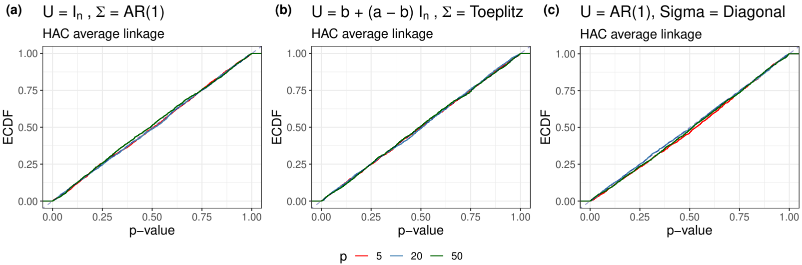
Global null hypothesis

Let $n = 100$, $\boldsymbol{\mu} = \mathbf{O}_{n \times p}$, and set $\mathcal{C}$ to choose three clusters. Then, randomly select two groups and test for the difference of their means.

Numerical simulations

Global null hypothesis

Let $n = 100$, $\boldsymbol{\mu} = \mathbf{O}_{n\times p}$, and set $\mathcal{C}$ to choose three clusters. Then, randomly select two groups and test for the difference of their means.



Numerical simulations

Conditional power

Conditional power = probability of rejecting the null for a randomly selected pair of clusters given that they are different.

Let $\boldsymbol{\mu}$ divide the $n = 50$ observations into three true clusters, for $\delta \in [4, 10.5]$:

$$\mu_{ij} = \begin{cases} -\frac{\delta}{2} & \text{if } i \leq \lfloor \frac{n}{3} \rfloor, \\ \frac{\sqrt{3}\delta}{2} & \text{if } \lfloor \frac{n}{3} \rfloor < i \leq \lfloor \frac{2n}{3} \rfloor, \\ \frac{\delta}{2} & \text{otherwise.} \end{cases} \quad \forall\, i \in \{1, \ldots, n\}, \forall\, j \in \{1, \ldots, p = 10\},$$

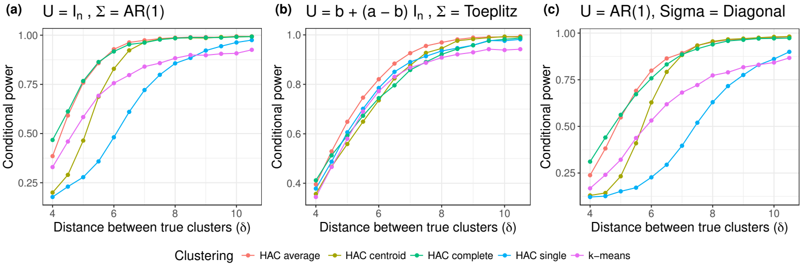
Numerical simulations

Conditional power

Conditional power = probability of rejecting the null for a randomly selected pair of clusters given that they are different.

Let $\boldsymbol{\mu}$ divide the $n = 50$ observations into three true clusters, for $\delta \in [4, 10.5]$:

$$\mu_{ij} = \begin{cases} -\frac{\delta}{2} & \text{if } i \leq \lfloor \frac{n}{3} \rfloor, \\ \frac{\sqrt{3}\delta}{2} & \text{if } \lfloor \frac{n}{3} \rfloor < i \leq \lfloor \frac{2n}{3} \rfloor, \\ \frac{\delta}{2} & \text{otherwise.} \end{cases} \quad \forall\, i \in \{1, \ldots, n\}, \forall\, j \in \{1, \ldots, p = 10\},$$



Numerical simulations

Let

$$X \sim \mathcal{MN}_{n\times p}(\boldsymbol{\mu}, \mathbf{U}, \boldsymbol{\Sigma}). \tag{dep}$$

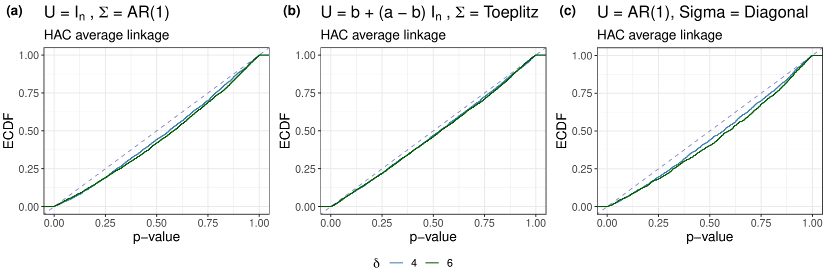
For $n = 500$ and $p = 10$, we simulated $K = 10000$ samples drawn from (dep) in settings (a), (b) and (c) with $\boldsymbol{\mu}$ being divided into two clusters :

$$\mu_{ij} = \begin{cases} \frac{\delta}{j} & \text{if } i \leq \frac{n}{2}, \\ -\frac{\delta}{j} & \text{otherwise,} \end{cases} \quad \forall\, i \in \{1, \dots, n\},\, \forall\, j \in \{1, \dots, p\},$$

with $\delta \in \{4, 6\}$.

For HAC with average linkage we set $\mathcal{C}$ to chose three clusters. Then, we kept the samples for which (H0) held when comparing two randomly selected clusters.

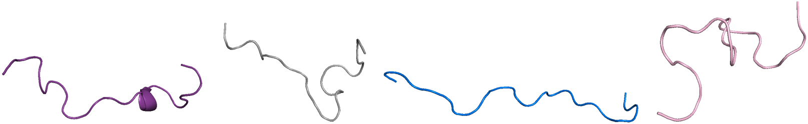
Numerical simulations



Hierarchical clustering of Hst5

Hst5 ensemble simulated with Flexible-Meccano (FM)[2] and filtered by SAXS data[3]

- $n = 2000$ conformations
- Featured by pairwise Euclidean distances of 24 amino acids $\Rightarrow p = 276$
- No temporal evolution in FM simulation : $\mathbf{U}^{(n)} = \mathbf{I}_n$
- $\boldsymbol{\Sigma}$ unknown to be estimated

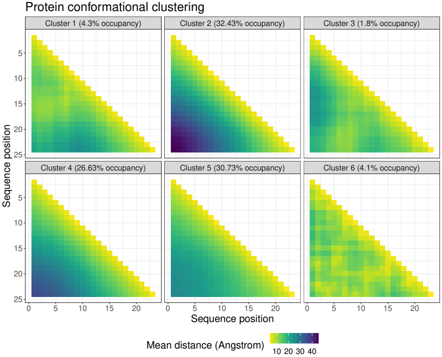


Strategy

Hierarchical clustering with average linkage, find 6 clusters.

2. Ozenne *et al. Bioinformatics* 2012, Bernadó *et al. PNAS* 2005. 3. Sagar *et al. J. Chem. Theory Comput* 2021.

Hierarchical clustering of Hst5



Pairwise p-values corrected for multiplicity (BH)

Cluster	1	2	3	4	5
2	$2.187589 \cdot 10^{-4}$				
3	$3.039844 \cdot 10^{-11}$	$1.41 \cdot 10^{-3}$			
4	$1.070993 \cdot 10^{-10}$	0.300540	$2.98464 \cdot 10^{-4}$		
5	$3.038979 \cdot 10^{-16}$	0.093018	$6.015797 \cdot 10^{-5}$	0.105446	
6	$1.729616 \cdot 10^{-6}$	0.010612	$9.290826 \cdot 10^{-9}$	$2.105 \cdot 10^{-3}$	$5.624624 \cdot 10^{-5}$