

Locally Minimax Optimal and Dimension-Agnostic Discrete Argmin Inference

Ilmun Kim

Department of Mathematical Sciences
KAIST



Joint work with



Aaditya Ramdas
(Carnegie Mellon University)

Outline of this talk

(1) Problem Setup

(2) The Proposed Methods

(3) Theoretical Properties

(4) Empirical Results

(5) Summary

Problem Setup: Discrete Argmin Inference

- Suppose $X_1, \dots, X_{2n} \stackrel{\text{i.i.d.}}{\sim} \mathbf{P}$ in \mathbb{R}^d with mean $\mu = (\mu_1, \dots, \mu_d)^\top$ and let

$$\Theta = \Theta(\mathbf{P}) := \arg \min_{k \in [d]} \mu_k$$

Problem Setup: Discrete Argmin Inference

- Suppose $X_1, \dots, X_{2n} \stackrel{\text{i.i.d.}}{\sim} \mathbf{P}$ in \mathbb{R}^d with mean $\mu = (\mu_1, \dots, \mu_d)^\top$ and let

$$\Theta = \Theta(\mathbf{P}) := \arg \min_{k \in [d]} \mu_k$$

i.e., the set of all coordinates whose mean equals the *smallest* in μ

Problem Setup: Discrete Argmin Inference

- Suppose $X_1, \dots, X_{2n} \stackrel{\text{i.i.d.}}{\sim} \mathbf{P}$ in \mathbb{R}^d with mean $\mu = (\mu_1, \dots, \mu_d)^\top$ and let

$$\Theta = \Theta(\mathbf{P}) := \arg \min_{k \in [d]} \mu_k$$

i.e., the set of all coordinates whose mean equals the *smallest* in μ

- We would like to form a **confidence set** $\widehat{\Theta}$ for Θ

Problem Setup: Discrete Argmin Inference

- Suppose $X_1, \dots, X_{2n} \stackrel{\text{i.i.d.}}{\sim} \mathbf{P}$ in \mathbb{R}^d with mean $\mu = (\mu_1, \dots, \mu_d)^\top$ and let

$$\Theta = \Theta(\mathbf{P}) := \arg \min_{k \in [d]} \mu_k$$

i.e., the set of all coordinates whose mean equals the *smallest* in μ

- We would like to form a **confidence set** $\widehat{\Theta}$ for Θ

Each $r \in \Theta$ is included in $\widehat{\Theta}$ with high probability:

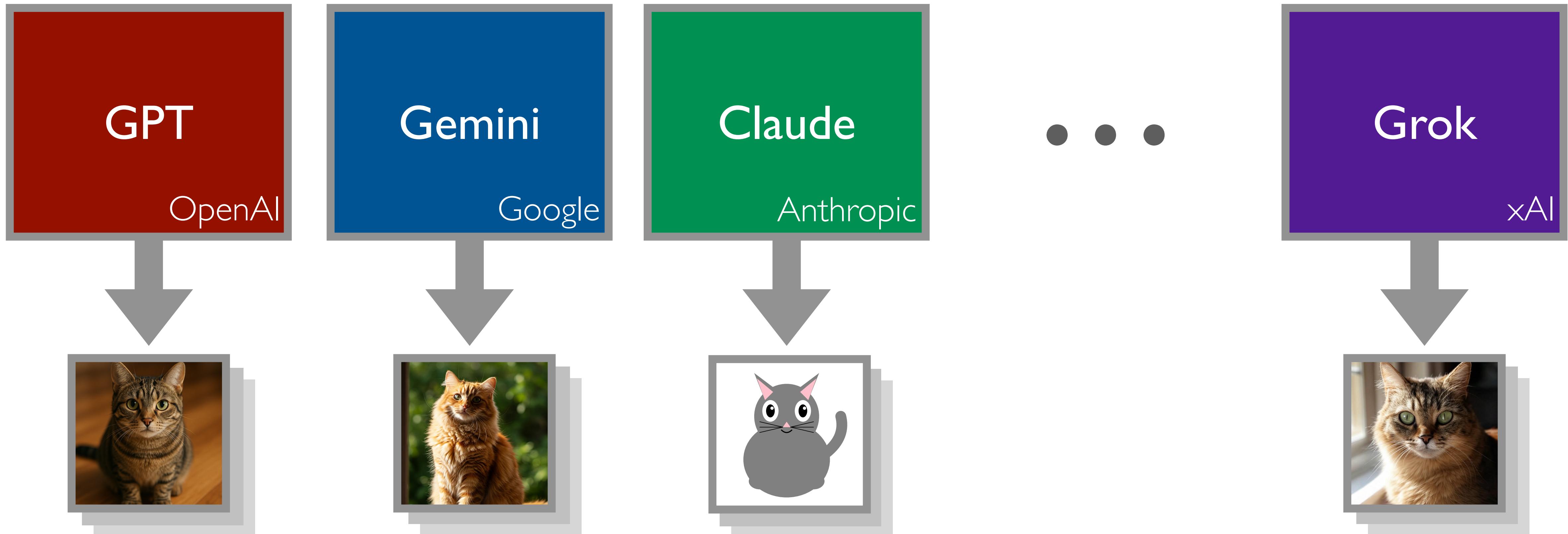
$$\inf_{\mathbf{P} \in \mathcal{P}} \inf_{r \in \Theta(\mathbf{P})} P(r \in \widehat{\Theta}) \geq 1 - \alpha$$



\mathcal{P} : class of distributions
 α : target error level

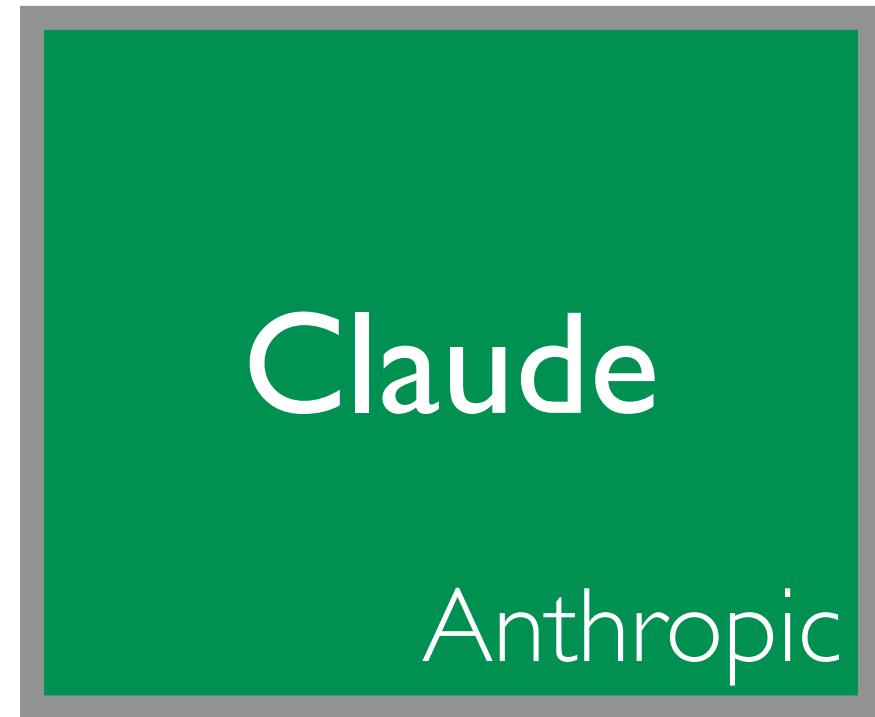
Modern Application

Prompt: “Generate cat images”



Modern Application

Prompt: “Generate cat images”



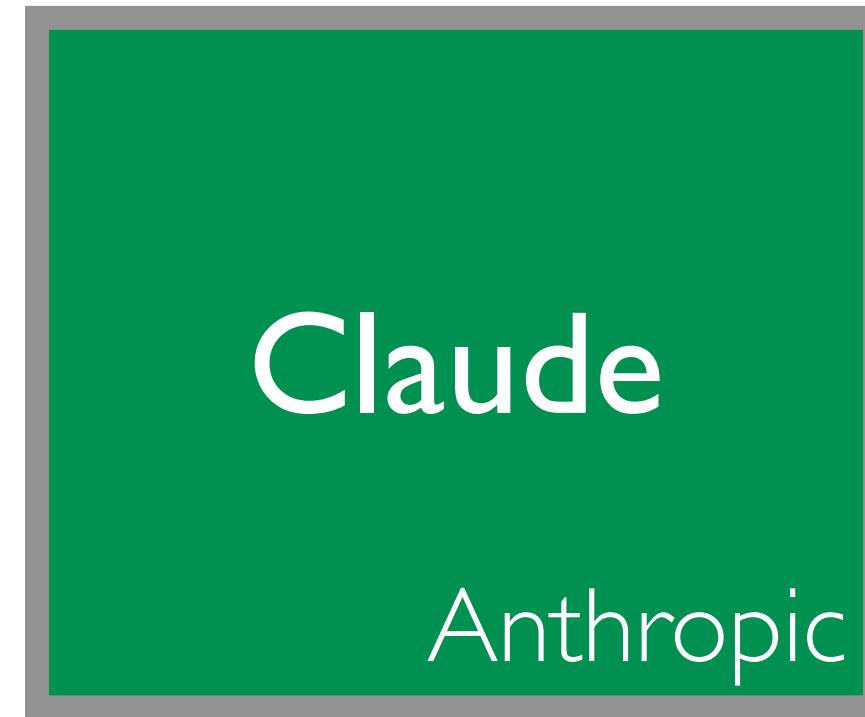
• • •



- The best model minimizes the **population risk**

Modern Application

Prompt: “Generate cat images”



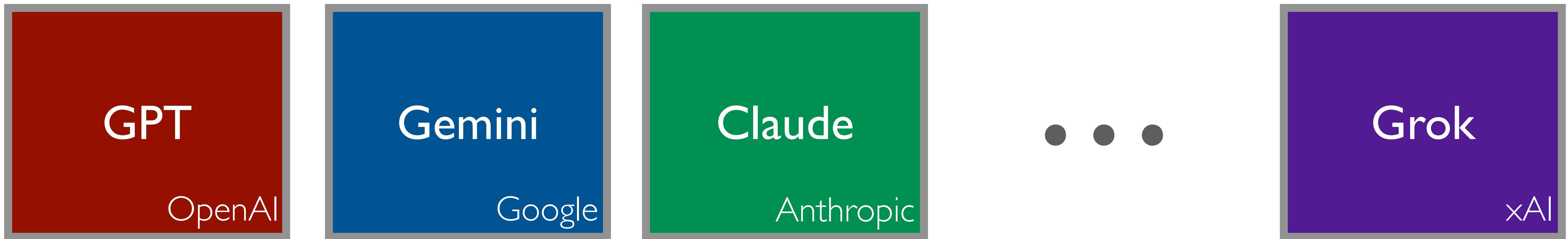
• • •



- The best model minimizes the **population risk**
- But **the population risk is unknown** → **the empirical risk**

Modern Application

Prompt: “Generate cat images”



- The best model minimizes the **population risk**
- But **the population risk is unknown** → **the empirical risk**
- We must **account for statistical uncertainty** to determine which models are plausibly **optimal with statistical confidence**

Related Work

- **Classical work:** Bechhofer (1954), Gupta (1956, 1965), Futschik and Pflug (1995) etc
→ rely on parametric models, independence between coordinates, absence of ties

Related Work

- **Classical work:** Bechhofer (1954), Gupta (1956, 1965), Futschik and Pflug (1995) etc
→ rely on parametric models, independence between coordinates, absence of ties
- **Hansen et al. (2011):** propose a sequential procedure for MCS with uniform coverage
→ widely cited (2550+), produce wide sets; extremely slow to run

Related Work

- **Classical work:** Bechhofer (1954), Gupta (1956, 1965), Futschik and Pflug (1995) etc
→ rely on parametric models, independence between coordinates, absence of ties
- **Hansen et al. (2011):** propose a sequential procedure for MCS with uniform coverage
→ widely cited (2550+), produce wide sets; extremely slow to run
- **Mogstad et al. (2024):** introduce a bootstrap approach for rank inference (can be tweaked)
→ efficiency not analyzed; results limited to fixed dimensional settings

Related Work

- **Classical work:** Bechhofer (1954), Gupta (1956, 1965), Futschik and Pflug (1995) etc
→ rely on parametric models, independence between coordinates, absence of ties
- **Hansen et al. (2011):** propose a sequential procedure for MCS with uniform coverage
→ widely cited (2550+), produce wide sets; extremely slow to run
- **Mogstad et al. (2024):** introduce a bootstrap approach for rank inference (can be tweaked)
→ efficiency not analyzed; results limited to fixed dimensional settings
- **Zhang et al. (2024):** introduce a cross-validation + privacy approach for argmin inference
→ require careful tuning and fall short of minimax optimality

We aim to develop a method for argmin inference with

I. Dimension-agnostic validity: works in both low- and high-dimensional settings

We aim to develop a method for argmin inference with

- I. Dimension-agnostic validity: works in both **low-** and **high-dimensional** settings
- 2. Powerful inference: attains **local minimax rates** across diverse regimes

We aim to develop a method for argmin inference with

- I. **Dimension-agnostic validity:** works in both **low-** and **high-dimensional** settings
2. **Powerful inference:** attains **local minimax rates** across diverse regimes
3. **Robustness to complex data:** handles **ties** and strong **dependence**

We aim to develop a method for argmin inference with

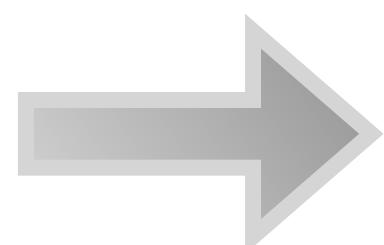
- I. **Dimension-agnostic validity:** works in both **low-** and **high-dimensional** settings
2. **Powerful inference:** attains **local minimax rates** across diverse regimes
3. **Robustness to complex data:** handles **ties** and strong **dependence**
4. **Model-agnostic applicability:** requires **no parametric** assumptions

We aim to develop a method for argmin inference with

- I. **Dimension-agnostic validity:** works in both **low-** and **high-dimensional** settings
2. **Powerful inference:** attains **local minimax rates** across diverse regimes
3. **Robustness to complex data:** handles **ties** and strong **dependence**
4. **Model-agnostic applicability:** requires **no parametric** assumptions
5. **Tuning-free implementation:** requires **no** (non-trivial) **tuning** parameters

We aim to develop a method for argmin inference with

1. **Dimension-agnostic validity:** works in both **low-** and **high-dimensional** settings
2. **Powerful inference:** attains **local minimax rates** across diverse regimes
3. **Robustness to complex data:** handles **ties** and strong **dependence**
4. **Model-agnostic applicability:** requires **no parametric** assumptions
5. **Tuning-free implementation:** requires **no (non-trivial) tuning** parameters



Sample Splitting + Studentization (Kim and Ramdas, 2024)

Dimension-Agnostic Argmin Inference

Formal (Primal) Goal.

- We seek a **dimension-agnostic confidence set** $\widehat{\Theta}$ for Θ that such that

$$\liminf_{n \rightarrow \infty} \inf_{P \in \mathcal{P}} \inf_{r \in \Theta(P)} P(r \in \widehat{\Theta}) \geq 1 - \alpha, \text{ regardless of the sequence } (d_n)_{n=1}^{\infty}$$

Formal (Primal) Goal.

- We seek a **dimension-agnostic confidence set** $\widehat{\Theta}$ for Θ that such that

$$\liminf_{n \rightarrow \infty} \inf_{P \in \mathcal{P}} \inf_{r \in \Theta(P)} P(r \in \widehat{\Theta}) \geq 1 - \alpha, \text{ regardless of the sequence } (d_n)_{n=1}^{\infty}$$

- We also seek a **confidence set** $\widehat{\Theta}$ for Θ such that its **expected cardinality** is small and **ideally optimal**

Formal (Dual) Goal.

- Given some fixed $r \in [d]$, consider the **null** and **alternative** hypotheses:

$$H_0 : r \in \Theta \quad \text{versus} \quad H_1 : r \notin \Theta$$

Formal (Dual) Goal.

- Given some fixed $r \in [d]$, consider the **null** and **alternative** hypotheses:

$$H_0 : r \in \Theta \quad \text{versus} \quad H_1 : r \notin \Theta$$

- Let $\psi_r : \{X_1, \dots, X_{2n}\} \rightarrow \{0,1\}$ denote a test function that rejects H_0 if $\psi_r = 1$

Formal (Dual) Goal.

- Given some fixed $\mathbf{r} \in [d]$, consider the **null** and **alternative** hypotheses:

$$H_0 : \mathbf{r} \in \Theta \quad \text{versus} \quad H_1 : \mathbf{r} \notin \Theta$$

- Let $\psi_{\mathbf{r}} : \{X_1, \dots, X_{2n}\} \rightarrow \{0,1\}$ denote a test function that rejects H_0 if $\psi_{\mathbf{r}} = 1$
- We seek a **dimension-agnostic test** $\psi_{\mathbf{r}}$ such that

$$\limsup_{n \rightarrow \infty} \sup_{P \in \mathcal{P}_{0,\mathbf{r}}} P(\psi_{\mathbf{r}} = 1) \leq \alpha, \quad \text{regardless of the sequence } (d_n)_{n=1}^{\infty}$$



Formal (Dual) Goal.

- Given some fixed $\mathbf{r} \in [d]$, consider the **null** and **alternative** hypotheses:

$$H_0 : \mathbf{r} \in \Theta \quad \text{versus} \quad H_1 : \mathbf{r} \notin \Theta$$

- Let $\psi_{\mathbf{r}} : \{X_1, \dots, X_{2n}\} \rightarrow \{0,1\}$ denote a test function that rejects H_0 if $\psi_{\mathbf{r}} = 1$
- We seek a **dimension-agnostic test** $\psi_{\mathbf{r}}$ such that

$$\limsup_{n \rightarrow \infty} \sup_{P \in \mathcal{P}_{0,\mathbf{r}}} P(\psi_{\mathbf{r}} = 1) \leq \alpha, \quad \text{regardless of the sequence } (d_n)_{n=1}^{\infty}$$

- We can construct a **DA confidence set** using **DA tests**



DA tests yield a DA confidence set via duality

- Suppose that each ψ_k satisfies the **dimension-agnostic (DA)** property:

$$\limsup_{n \rightarrow \infty} \sup_{P \in \mathcal{P}_{0,k}} P(\psi_k = 1) \leq \alpha, \text{ regardless of the sequence } (d_n)_{n=1}^{\infty}$$

DA tests yield a DA confidence set via duality

- Suppose that each ψ_k satisfies the **dimension-agnostic (DA)** property:

$$\limsup_{n \rightarrow \infty} \sup_{P \in \mathcal{P}_{0,k}} P(\psi_k = 1) \leq \alpha, \text{ regardless of the sequence } (d_n)_{n=1}^{\infty}$$

- We run d such **DA** tests ψ_1, \dots, ψ_d and let

$$\widehat{\Theta} = \{k \in [d] : \psi_k = 0\}$$

DA tests yield a DA confidence set via duality

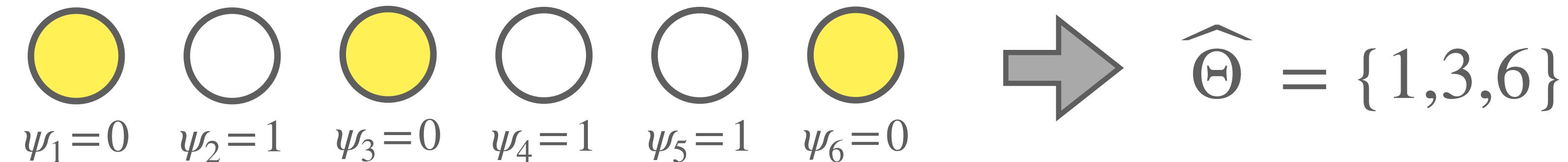
- Suppose that each ψ_k satisfies the **dimension-agnostic (DA)** property:

$$\limsup_{n \rightarrow \infty} \sup_{P \in \mathcal{P}_{0,k}} P(\psi_k = 1) \leq \alpha, \text{ regardless of the sequence } (d_n)_{n=1}^{\infty}$$

- We run d such **DA** tests ψ_1, \dots, ψ_d and let

$$\widehat{\Theta} = \{k \in [d] : \psi_k = 0\}$$

For example,



DA tests yield a DA confidence set via duality

- Suppose that each ψ_k satisfies the **dimension-agnostic (DA)** property:

$$\limsup_{n \rightarrow \infty} \sup_{P \in \mathcal{P}_{0,k}} P(\psi_k = 1) \leq \alpha, \text{ regardless of the sequence } (d_n)_{n=1}^{\infty}$$

- We run d such **DA** tests ψ_1, \dots, ψ_d and let

$$\widehat{\Theta} = \{k \in [d] : \psi_k = 0\}$$

which results in the **DA confidence set**

$$\liminf_{n \rightarrow \infty} \inf_{P \in \mathcal{P}} \inf_{r \in \Theta(P)} P(r \in \widehat{\Theta}) \geq 1 - \alpha, \text{ regardless of the sequence } (d_n)_{n=1}^{\infty}$$

Procedures

Reformulation

- Recall our (dual) goal is to test

$$H_0 : \mathbf{r} \in \Theta \quad \text{vs.} \quad H_1 : \mathbf{r} \notin \Theta$$

Reformulation

- Recall our (dual) goal is to test

$$H_0 : \mathbf{r} \in \Theta \quad \text{vs.} \quad H_1 : \mathbf{r} \notin \Theta$$

- We can reformulate this as

$$H_0 : \mu_{\mathbf{r}} - \mu_{\mathbf{s}} \leq 0 \quad \text{vs.} \quad H_1 : \mu_{\mathbf{r}} - \mu_{\mathbf{s}} > 0$$

where $\mathbf{s} = \arg \min_{k \in [d] \setminus \{\mathbf{r}\}} \mu_k$

Reformulation

- Recall our (dual) goal is to test

$$H_0 : \mathbf{r} \in \Theta \quad \text{vs.} \quad H_1 : \mathbf{r} \notin \Theta$$

- We can reformulate this as

$$H_0 : \mu_{\mathbf{r}} - \mu_{\mathbf{s}} \leq 0 \quad \text{vs.} \quad H_1 : \mu_{\mathbf{r}} - \mu_{\mathbf{s}} > 0$$

where $\mathbf{s} = \arg \min_{k \in [d] \setminus \{\mathbf{r}\}} \mu_k$

i.e., Target $\mu_{\mathbf{r}}$ is greater than the smallest mean $\mu_{\mathbf{s}}$?

Reformulation

- Recall our (dual) goal is to test

$$H_0 : \mathbf{r} \in \Theta \quad \text{vs.} \quad H_1 : \mathbf{r} \notin \Theta$$

- When \mathbf{s} is known, use a one-sided t-test

$$\frac{\bar{X}_{\mathbf{r}} - \bar{X}_{\mathbf{s}}}{\hat{\sigma}_{\mathbf{r}, \mathbf{s}}} > z_{1-\alpha}$$

- We can reformulate this as

$$H_0 : \mu_{\mathbf{r}} - \mu_{\mathbf{s}} \leq 0 \quad \text{vs.} \quad H_1 : \mu_{\mathbf{r}} - \mu_{\mathbf{s}} > 0$$

where $\mathbf{s} = \arg \min_{k \in [d] \setminus \{\mathbf{r}\}} \mu_k$

i.e., Target $\mu_{\mathbf{r}}$ is greater than the smallest mean $\mu_{\mathbf{s}}$?

Reformulation

- Recall our (dual) goal is to test

$$H_0 : \mathbf{r} \in \Theta \quad \text{vs.} \quad H_1 : \mathbf{r} \notin \Theta$$

- We can reformulate this as

$$H_0 : \mu_{\mathbf{r}} - \mu_{\mathbf{s}} \leq 0 \quad \text{vs.} \quad H_1 : \mu_{\mathbf{r}} - \mu_{\mathbf{s}} > 0$$

where $\mathbf{s} = \arg \min_{k \in [d] \setminus \{\mathbf{r}\}} \mu_k$

- When \mathbf{s} is known, use a one-sided t-test

$$\frac{\bar{X}_{\mathbf{r}} - \bar{X}_{\mathbf{s}}}{\hat{\sigma}_{\mathbf{r}, \mathbf{s}}} > z_{1-\alpha}$$

- But \mathbf{s} is unknown: one idea (Mogstad et al. 2024)

$$\max_{k \in [d] \setminus \{\mathbf{r}\}} \frac{\bar{X}_{\mathbf{r}} - \bar{X}_k}{\hat{\sigma}_{\mathbf{r}, k}} > c_{1-\alpha}$$

i.e., Target $\mu_{\mathbf{r}}$ is greater than the smallest mean $\mu_{\mathbf{s}}$?

Reformulation

- Recall our (dual) goal is to test

$$H_0 : \mathbf{r} \in \Theta \quad \text{vs.} \quad H_1 : \mathbf{r} \notin \Theta$$

- We can reformulate this as

$$H_0 : \mu_{\mathbf{r}} - \mu_{\mathbf{s}} \leq 0 \quad \text{vs.} \quad H_1 : \mu_{\mathbf{r}} - \mu_{\mathbf{s}} > 0$$

where $\mathbf{s} = \arg \min_{k \in [d] \setminus \{\mathbf{r}\}} \mu_k$

- When \mathbf{s} is known, use a one-sided t-test

$$\frac{\bar{X}_{\mathbf{r}} - \bar{X}_{\mathbf{s}}}{\hat{\sigma}_{\mathbf{r}, \mathbf{s}}} > z_{1-\alpha}$$

- But \mathbf{s} is unknown: one idea (Mogstad et al. 2024)

$$\max_{k \in [d] \setminus \{\mathbf{r}\}} \frac{\bar{X}_{\mathbf{r}} - \bar{X}_k}{\hat{\sigma}_{\mathbf{r}, k}} > c_{1-\alpha}$$

- Hard to calibrate $c_{1-\alpha}$ in high-dimensions

i.e., Target $\mu_{\mathbf{r}}$ is greater than the smallest mean $\mu_{\mathbf{s}}$?

Dimension-Agnostic Argmin Test

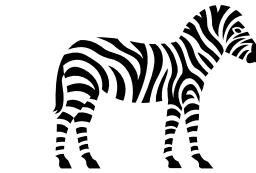
💡 Idea: use \mathcal{D}_1 to estimate s and use \mathcal{D}_2 for inference

Step I (Sample Splitting)

- Split the data into two

$$\begin{array}{c} \mathcal{D}_1 \quad \mathcal{D}_2 \\ \left[\begin{array}{c} X_1 \\ X_2 \\ \vdots \\ X_{n-1} \\ X_n \end{array} \right] \quad \left[\begin{array}{c} X_{n+1} \\ X_{n+2} \\ \vdots \\ X_{2n-1} \\ X_{2n} \end{array} \right] \end{array}$$

Dimension-Agnostic Argmin Test

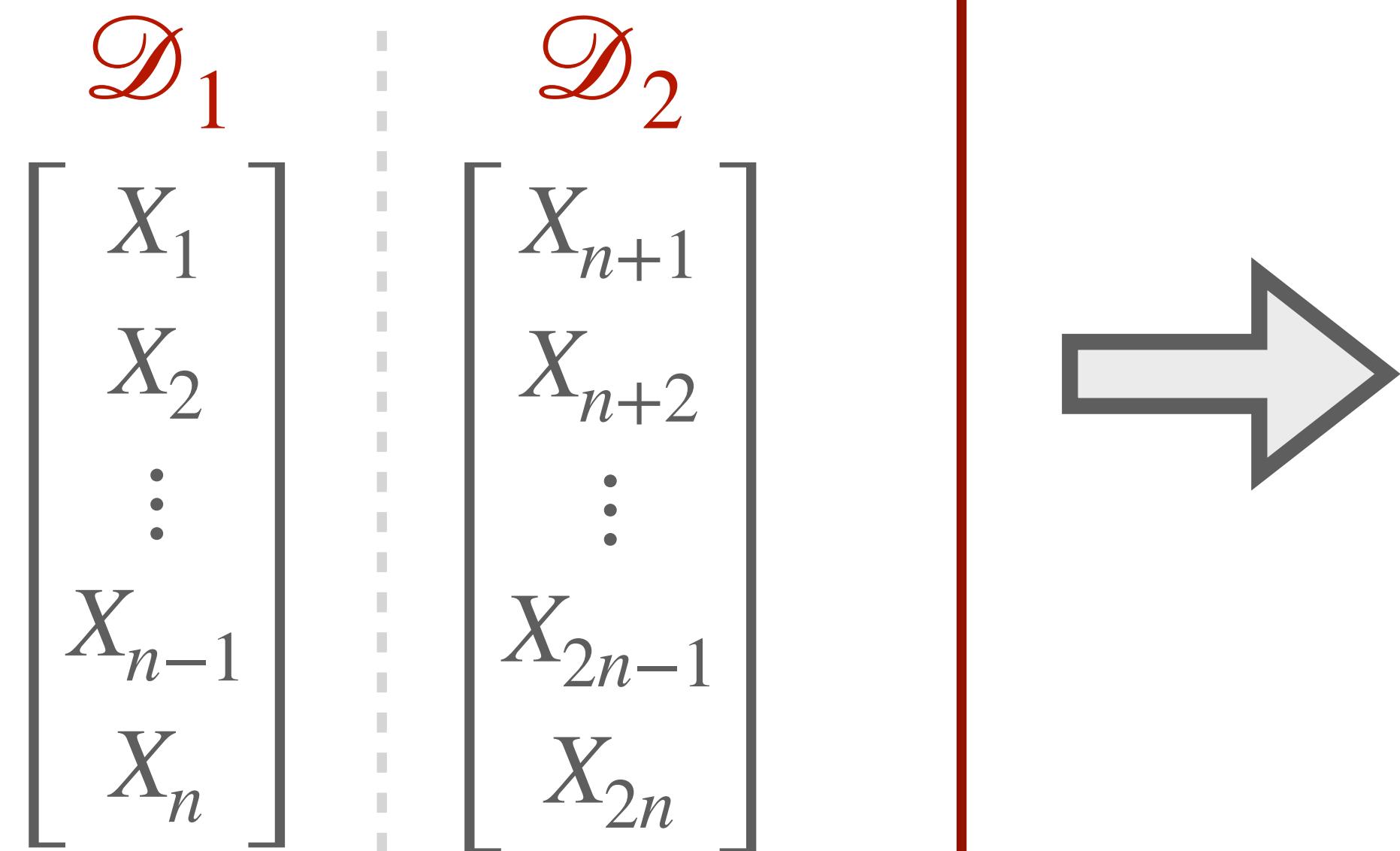


$$s = \arg \min_{k \in [d] \setminus \{r\}} \mu_k$$

Idea: use \mathcal{D}_1 to estimate s and use \mathcal{D}_2 for inference

Step I (Sample Splitting)

- Split the data into two



Step II (Model Selection)

- Based on \mathcal{D}_1 , compute

I. Plug-in version

$$\hat{s} = \arg \max_{k \in [d] \setminus \{r\}} \bar{X}_{\textcolor{red}{r}}^{(2)} - \bar{X}_k^{(2)}$$

2. Noise-adjusted version

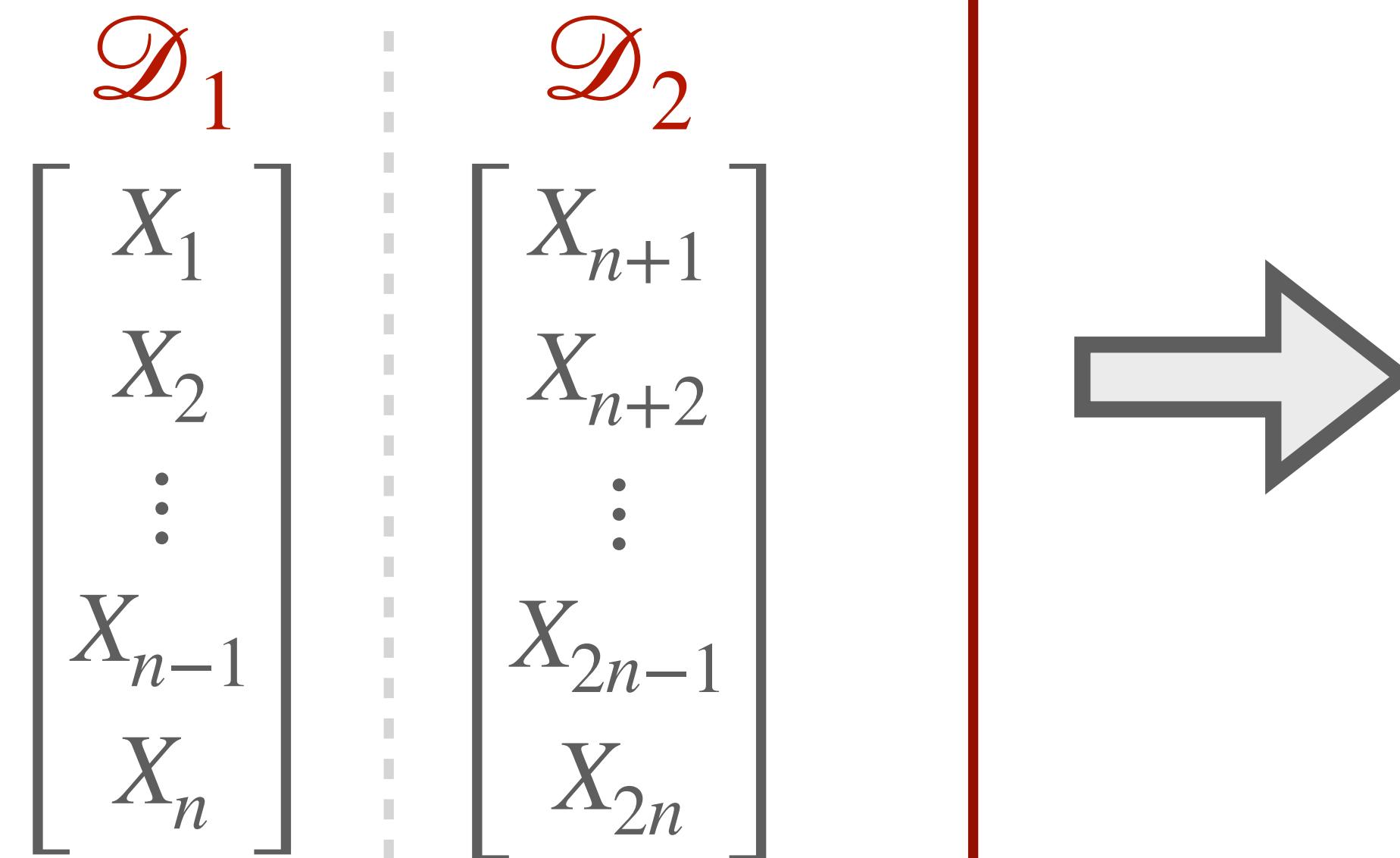
$$\hat{s} = \arg \max_{k \in [d] \setminus \{r\}} \frac{\bar{X}_{\textcolor{red}{r}}^{(2)} - \bar{X}_k^{(2)}}{\hat{\sigma}_{\textcolor{red}{r},k}^{(2)} \vee \kappa}$$

Dimension-Agnostic Argmin Test

 **Idea:** use \mathcal{D}_1 to estimate s and use \mathcal{D}_2 for inference

Step I (Sample Splitting)

- Split the data into two



Step II (Model Selection)

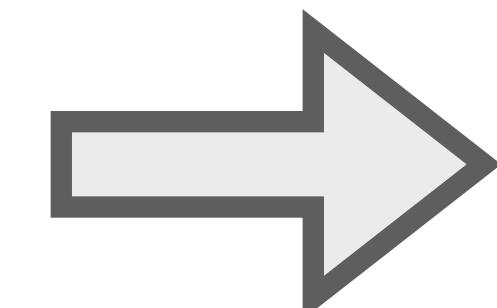
- Based on \mathcal{D}_1 , compute

I. Plug-in version

$$\hat{s} = \arg \max_{k \in [d] \setminus \{\textcolor{red}{r}\}} \bar{X}_{\textcolor{red}{r}}^{(2)} - \bar{X}_k^{(2)}$$

2. Noise-adjusted version

$$\hat{s} = \arg \max_{k \in [d] \setminus \{\textcolor{red}{r}\}} \frac{\bar{X}_{\textcolor{red}{r}}^{(2)} - \bar{X}_k^{(2)}}{\hat{\sigma}_{\textcolor{red}{r}, k}^{(2)} \vee \kappa}$$



Dimension-Agnostic Argmin Test

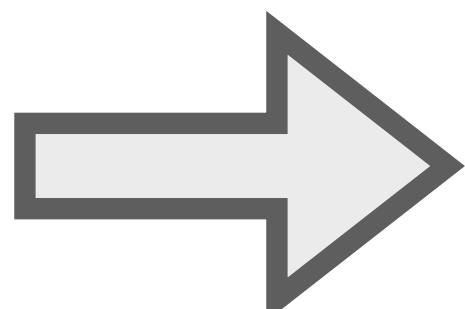
 **Idea:** use \mathcal{D}_1 to estimate s and use \mathcal{D}_2 for inference

Step III (Student's t-statistic)

- Given $\hat{s} \in [d] \setminus \{r\}$, compute a one-sided t-statistic

$$T = \frac{\bar{X}_{\color{red}r}^{(1)} - \bar{X}_{\color{blue}\hat{s}}^{(1)}}{\hat{\sigma}_{\color{red}r, \color{blue}\hat{s}}^{(1)}}$$

based on \mathcal{D}_2



Dimension-Agnostic Argmin Test

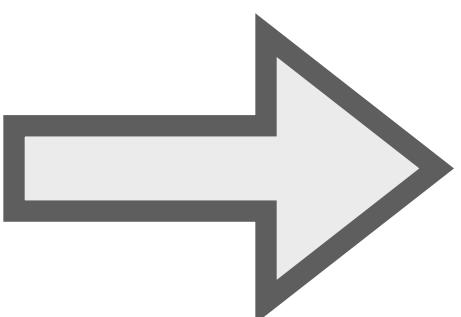
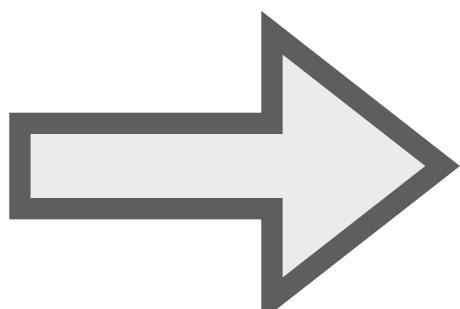
 **Idea:** use \mathcal{D}_1 to estimate s and use \mathcal{D}_2 for inference

Step III (Student's t-statistic)

- Given $\hat{s} \in [d] \setminus \{r\}$, compute a one-sided t-statistic

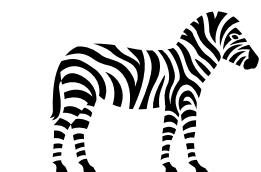
$$T = \frac{\bar{X}_{\hat{s}}^{(1)} - \bar{X}_r^{(1)}}{\hat{\sigma}_{r,\hat{s}}^{(1)}}$$

based on \mathcal{D}_2



Step IV (Decision)

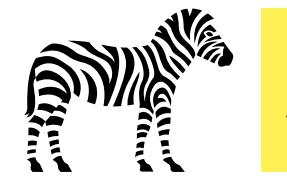
- Reject** the null if $T > z_{1-\alpha}$
- Accept** the null o.w.



$$\Phi(z_{1-\alpha}) = 1 - \alpha$$

Theoretical Properties

- ▶ I. Asymptotic Validity
- 2. Power Analysis

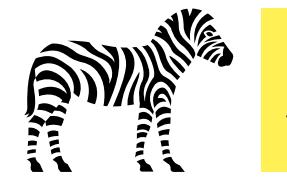


$X = (X^{(1)}, \dots, X^{(d)})$

Asymptotic Validity

Assumption (Truncated 2nd Moment Condition)

Let $W_k := (X^{(\textcolor{red}{r})} - \mu_{\textcolor{red}{r}}) - (X^{(k)} - \mu_k)$ and



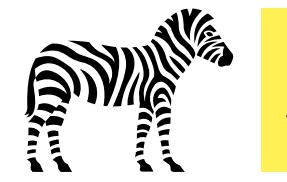
$X = (X^{(1)}, \dots, X^{(d)})$

Asymptotic Validity

Assumption (Truncated 2nd Moment Condition)

Let $W_k := (X^{(\textcolor{red}{r})} - \mu_{\textcolor{red}{r}}) - (X^{(k)} - \mu_k)$ and

$$M_k := \sup_{P \in \mathcal{P}_{0,\textcolor{red}{r}}} \mathbb{E}_P \left[\frac{W_k^2}{\mathbb{E}_P[W_k^2]} \min \left\{ 1, \frac{|W_k|}{n^{1/2}(\mathbb{E}_P[W_k^2])^{1/2}} \right\} \right]$$



$X = (X^{(1)}, \dots, X^{(d)})$

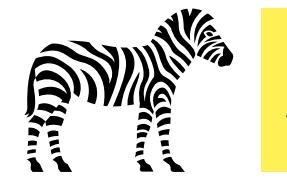
Asymptotic Validity

Assumption (Truncated 2nd Moment Condition)

Let $W_k := (X^{(\textcolor{red}{r})} - \mu_{\textcolor{red}{r}}) - (X^{(k)} - \mu_k)$ and

$$M_k := \sup_{P \in \mathcal{P}_{0,\textcolor{red}{r}}} \mathbb{E}_P \left[\frac{W_k^2}{\mathbb{E}_P[W_k^2]} \min \left\{ 1, \frac{|W_k|}{n^{1/2}(\mathbb{E}_P[W_k^2])^{1/2}} \right\} \right]$$

Assume that $\max_{k \in [d] \setminus \{\textcolor{red}{r}\}} M_k = o(1)$



$X = (X^{(1)}, \dots, X^{(d)})$

Asymptotic Validity

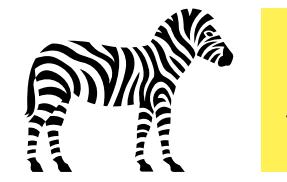
Assumption (Truncated 2nd Moment Condition)

Let $W_k := (X^{(\textcolor{red}{r})} - \mu_{\textcolor{red}{r}}) - (X^{(k)} - \mu_k)$ and

$$M_k := \sup_{P \in \mathcal{P}_{0,\textcolor{red}{r}}} \mathbb{E}_P \left[\frac{W_k^2}{\mathbb{E}_P[W_k^2]} \min \left\{ 1, \frac{|W_k|}{n^{1/2}(\mathbb{E}_P[W_k^2])^{1/2}} \right\} \right]$$

Assume that $\max_{k \in [d] \setminus \{\textcolor{red}{r}\}} M_k = o(1)$

- **Equivalent** to Lindeberg's condition (weaker than Lyapunov)



$X = (X^{(1)}, \dots, X^{(d)})$

Asymptotic Validity

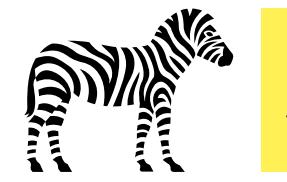
Assumption (Truncated 2nd Moment Condition)

Let $W_k := (X^{(\textcolor{red}{r})} - \mu_{\textcolor{red}{r}}) - (X^{(k)} - \mu_k)$ and

$$M_k := \sup_{P \in \mathcal{P}_{0,\textcolor{red}{r}}} \mathbb{E}_P \left[\frac{W_k^2}{\mathbb{E}_P[W_k^2]} \min \left\{ 1, \frac{|W_k|}{n^{1/2}(\mathbb{E}_P[W_k^2])^{1/2}} \right\} \right]$$

Assume that $\max_{k \in [d] \setminus \{\textcolor{red}{r}\}} M_k = o(1)$

- **Equivalent** to Lindeberg's condition (weaker than Lyapunov)
- If $X \sim N(\mu, \Sigma)$, it only requires $\text{Var}[W_k] > 0$



$X = (X^{(1)}, \dots, X^{(d)})$

Asymptotic Validity

Assumption (Truncated 2nd Moment Condition)

Let $W_k := (X^{(\textcolor{red}{r})} - \mu_{\textcolor{red}{r}}) - (X^{(k)} - \mu_k)$ and

$$M_k := \sup_{P \in \mathcal{P}_{0,\textcolor{red}{r}}} \mathbb{E}_P \left[\frac{W_k^2}{\mathbb{E}_P[W_k^2]} \min \left\{ 1, \frac{|W_k|}{n^{1/2}(\mathbb{E}_P[W_k^2])^{1/2}} \right\} \right]$$

Assume that $\max_{k \in [d] \setminus \{\textcolor{red}{r}\}} M_k = o(1)$

- **Equivalent** to Lindeberg's condition (weaker than Lyapunov)
- If $X \sim N(\mu, \Sigma)$, it only requires $\text{Var}[W_k] > 0$
- **Arbitrary dependence** among the components except $X^{(\textcolor{red}{r})}$

Asymptotic Validity

$$M_k := \sup_{P \in \mathcal{P}_{0,r}} \mathbb{E}_P \left[\frac{W_k^2}{\mathbb{E}_P[W_k^2]} \min \left\{ 1, \frac{|W_k|}{n^{1/2}(\mathbb{E}_P[W_k^2])^{1/2}} \right\} \right]$$

Assume that $\max_{k \in [d] \setminus \{r\}} M_k = o(1)$

Theorem Kim and Ramdas (2025)

Let $\mathcal{P}_{0,r}$ be the class of null distributions with $H_0 : r \in \Theta$ that satisfy the **truncated 2nd moment condition**. Then the DA test is asymptotically valid as

$$\limsup_{n \rightarrow \infty} \sup_{P \in \mathcal{P}_{0,r}} P(\psi_r = 1) \leq \alpha, \text{ regardless of the sequence } (d_n)_{n=1}^\infty$$

Asymptotic Validity

$$M_k := \sup_{P \in \mathcal{P}_{0,r}} \mathbb{E}_P \left[\frac{W_k^2}{\mathbb{E}_P[W_k^2]} \min \left\{ 1, \frac{|W_k|}{n^{1/2}(\mathbb{E}_P[W_k^2])^{1/2}} \right\} \right]$$

Assume that $\max_{k \in [d] \setminus \{r\}} M_k = o(1)$

Theorem Kim and Ramdas (2025)

Let $\mathcal{P}_{0,r}$ be the class of null distributions with $H_0 : r \in \Theta$ that satisfy the **truncated 2nd moment condition**. Then the DA test is asymptotically valid as

$$\limsup_{n \rightarrow \infty} \sup_{P \in \mathcal{P}_{0,r}} P(\psi_r = 1) \leq \alpha, \quad \text{regardless of the sequence } (d_n)_{n=1}^\infty$$

It holds for both $\left\{ \begin{array}{l} \text{I. Plug-in version} \\ \\ \hat{s} = \arg \max_{k \in [d] \setminus \{r\}} \bar{X}_r^{(2)} - \bar{X}_k^{(2)} \\ \\ \text{2. Noise-adjusted version} \\ \\ \hat{s} = \arg \max_{k \in [d] \setminus \{r\}} \frac{\bar{X}_r^{(2)} - \bar{X}_k^{(2)}}{\hat{\sigma}_{r,k}^{(2)} \vee \kappa} \end{array} \right\}$

Theoretical Properties

- I. Asymptotic Validity
- ▶ 2. Power Analysis

Confusion Set

- The problem **difficulty** depends on the cardinality of a **confusion set** \mathbb{C}_r

$$\mathbb{C}_r := \left\{ k \in [d] \setminus \{\textcolor{red}{r}\} : \frac{\mu_{\textcolor{red}{r}} - \mu_\star}{2} \leq \mu_k - \mu_\star \leq C_n \sqrt{\frac{\log(d)}{n}} \right\}$$

Confusion Set

- The problem **difficulty** depends on the cardinality of a **confusion set** \mathbb{C}_r

$$\mathbb{C}_r := \left\{ k \in [d] \setminus \{\textcolor{red}{r}\} : \frac{\mu_{\textcolor{red}{r}} - \mu_\star}{2} \leq \mu_k - \mu_\star \leq \textcolor{red}{C}_n \sqrt{\frac{\log(d)}{n}} \right\}$$

C_n : any increasing sequence

Confusion Set

$$\mu_{\star} = \min_{i \in [d]} \mu_i$$

- The problem **difficulty** depends on the cardinality of a **confusion set** \mathbb{C}_r

$$\mathbb{C}_r := \left\{ k \in [d] \setminus \{\textcolor{red}{r}\} : \frac{\mu_{\textcolor{red}{r}} - \mu_{\star}}{2} \leq \mu_k - \mu_{\star} \leq C_n \sqrt{\frac{\log(d)}{n}} \right\}$$

\mathcal{A}

- Intuition for \mathcal{A}^c

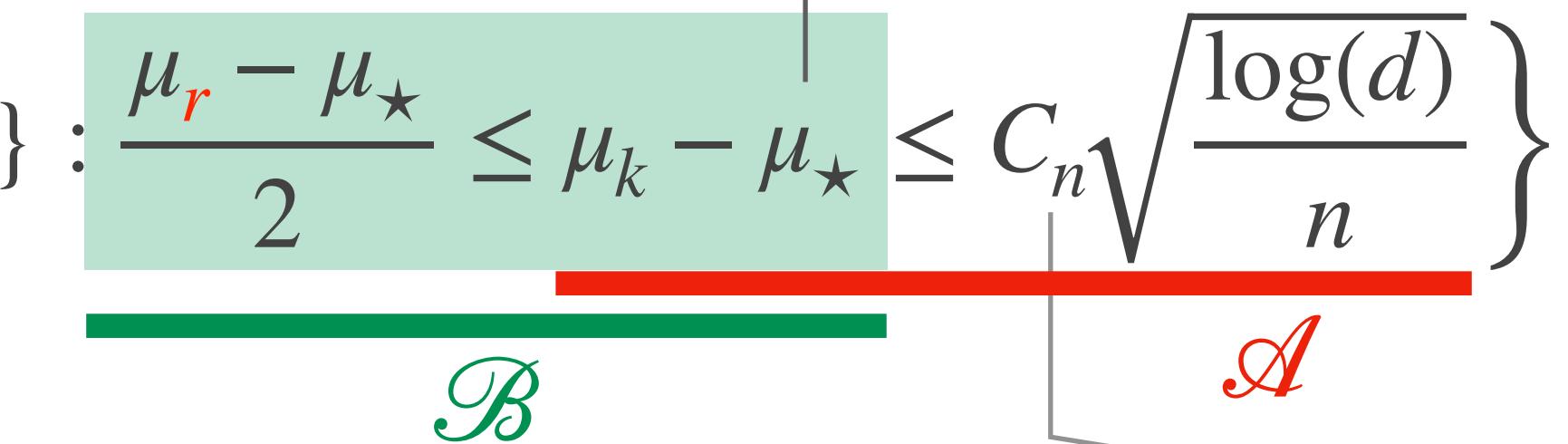
C_n : any increasing sequence

When μ_k is far from μ_{\star} , it is unlikely that $\hat{s} = k$
 \therefore Well-separated set

Confusion Set

$$\mu_{\star} = \min_{i \in [d]} \mu_i$$

- The problem **difficulty** depends on the cardinality of a **confusion set** \mathbb{C}_r

$$\mathbb{C}_r := \left\{ k \in [d] \setminus \{\textcolor{red}{r}\} : \frac{\mu_{\textcolor{red}{r}} - \mu_{\star}}{2} \leq \mu_k - \mu_{\star} \leq C_n \sqrt{\frac{\log(d)}{n}} \right\}$$


- **Intuition for \mathcal{A}^c**

When μ_k is far from μ_{\star} , it is unlikely that $\hat{s} = k$
 \therefore Well-separated set

C_n : any increasing sequence

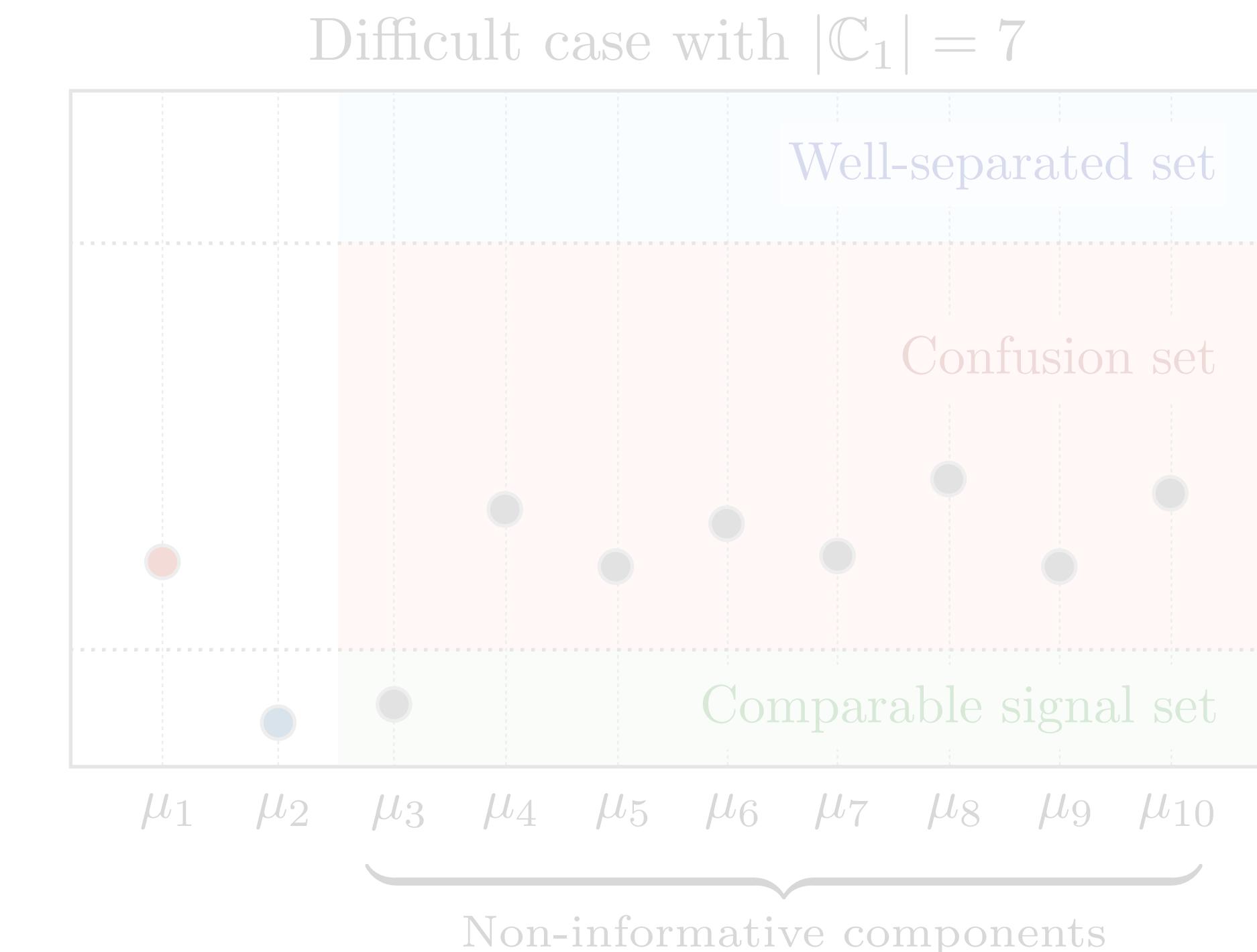
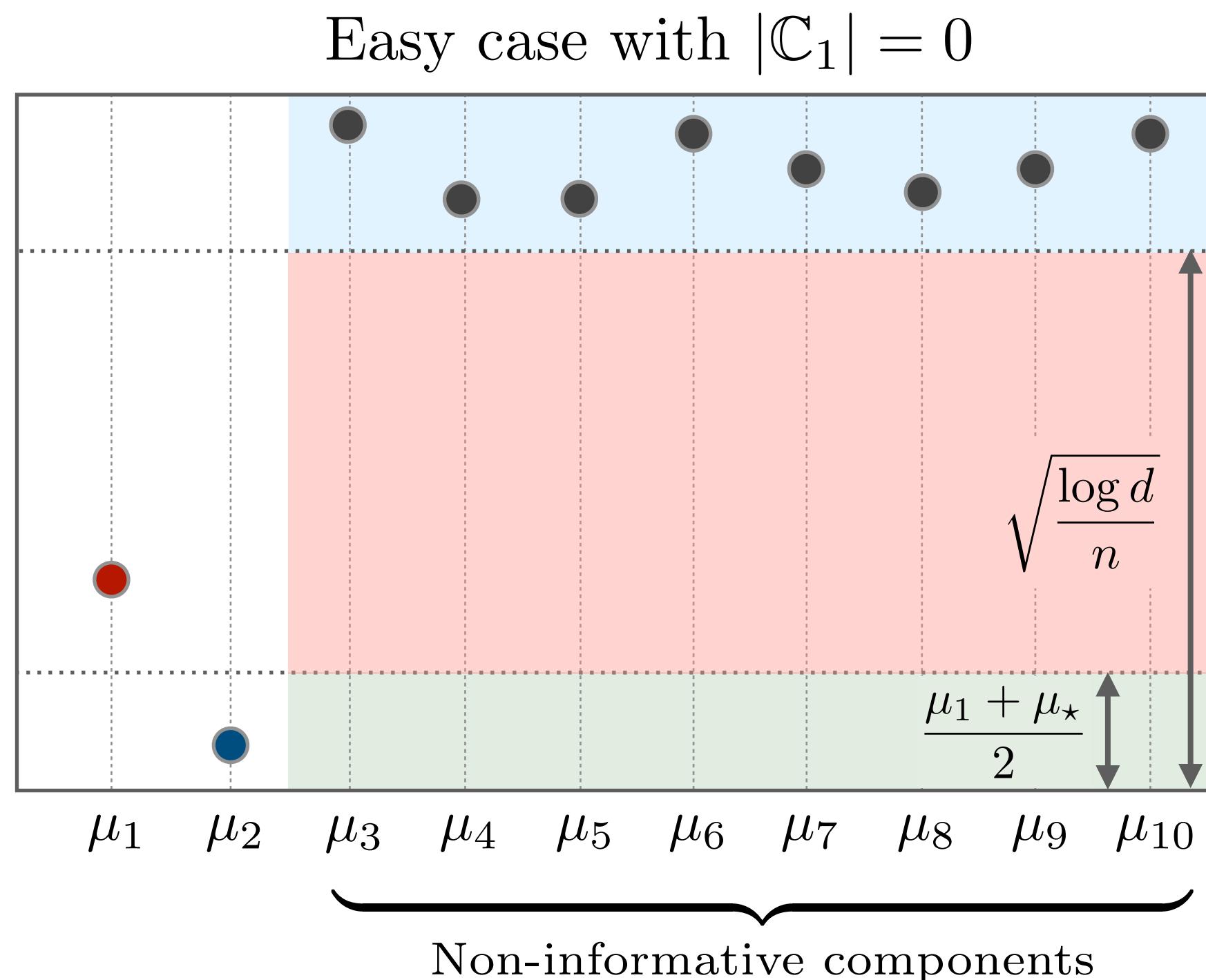
- **Intuition for \mathcal{B}^c**

$$\frac{\mu_{\textcolor{red}{r}} - \mu_{\star}}{2} > \mu_k - \mu_{\star} \iff \mu_{\textcolor{red}{r}} - \mu_k > \frac{\mu_{\textcolor{red}{r}} - \mu_{\star}}{2}$$

\therefore Comparable signal

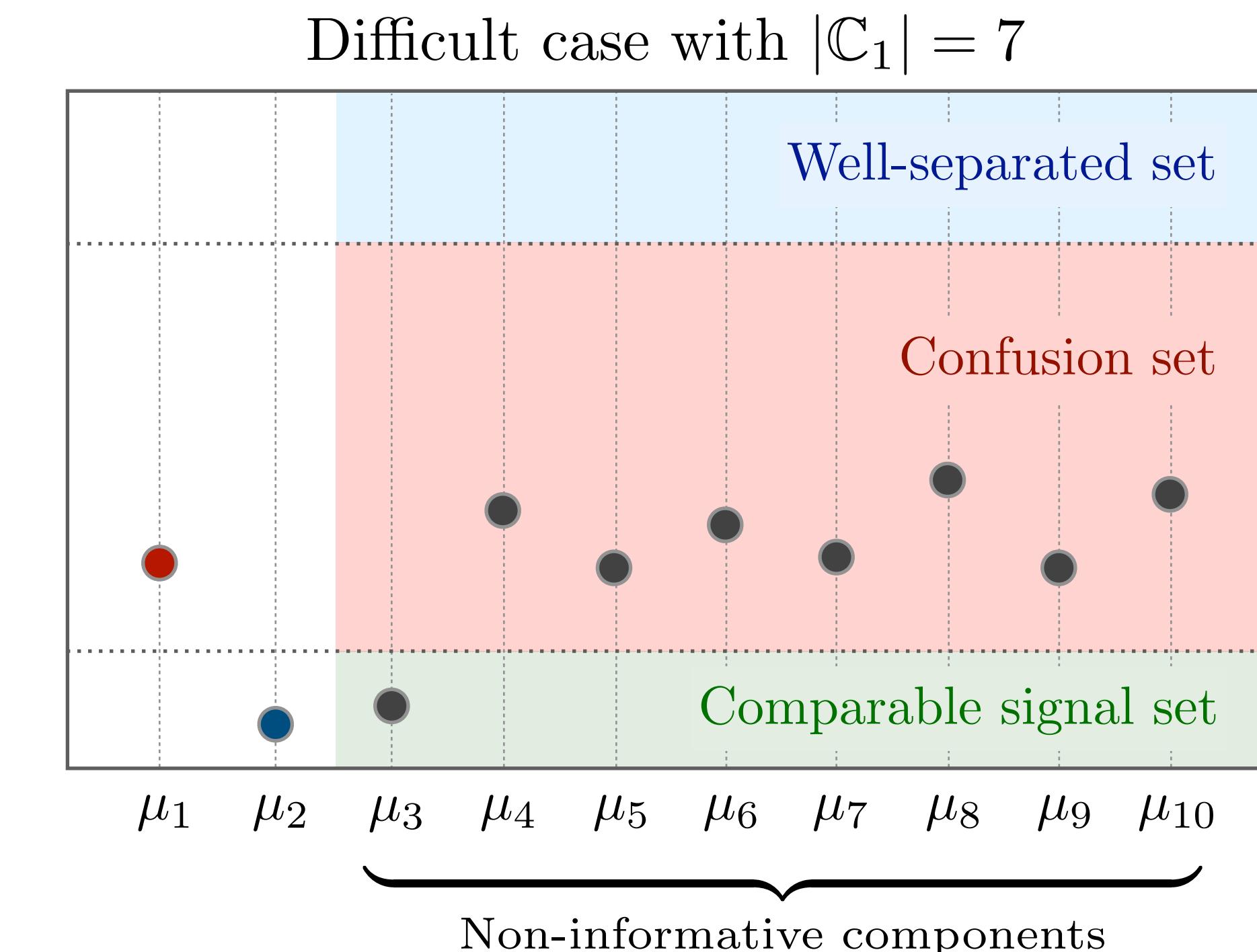
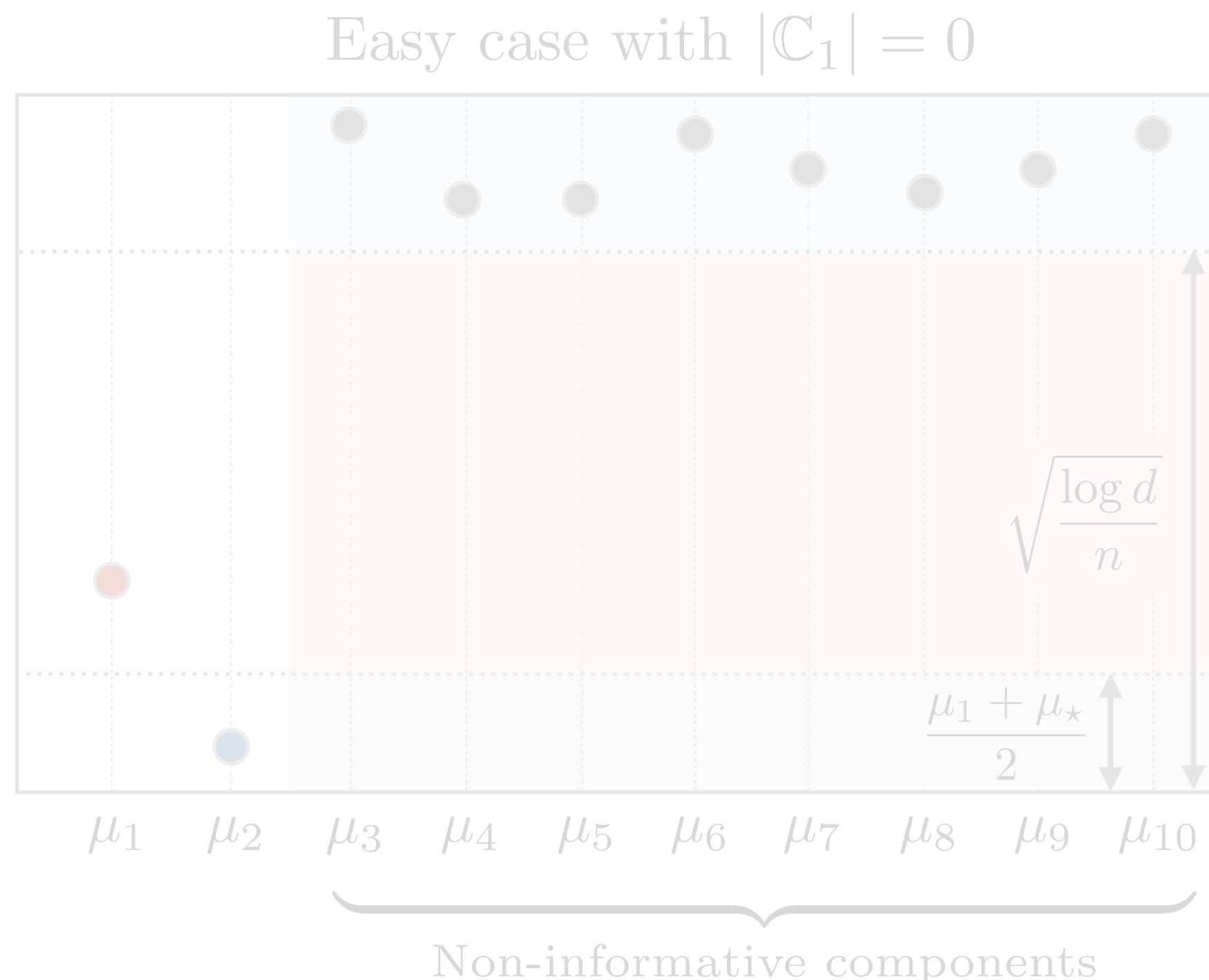
Confusion Set

- Set $\mu_r = \mu_1$ and $\mu_\star = \mu_2$



Confusion Set

- Set $\mu_r = \mu_1$ and $\mu_\star = \mu_2$



Power Analysis

- Let \mathcal{P} be a collection of **sub-Gaussian** distributions with σ^2

Power Analysis

- Let \mathcal{P} be a collection of **sub-Gaussian** distributions with σ^2
- Define a class of **alternative** distributions

$$\mathcal{P}_{1,\textcolor{red}{r}}(\varepsilon; \tau) := \left\{ P \in \mathcal{P} : \mu_{\textcolor{red}{r}} - \mu_{\star} \geq \varepsilon \text{ and } |\mathbb{C}_{\textcolor{red}{r}}| = \tau \right\}$$

Power Analysis

- Let \mathcal{P} be a collection of **sub-Gaussian** distributions with σ^2
- Define a class of **alternative** distributions

$$\mathcal{P}_{1,\textcolor{red}{r}}(\varepsilon; \tau) := \left\{ P \in \mathcal{P} : \mu_{\textcolor{red}{r}} - \mu_{\star} \geq \varepsilon \text{ and } |\mathbb{C}_{\textcolor{red}{r}}| = \tau \right\}$$

- The **critical radius** ε^{\star} is defined as

$$\varepsilon^{\star} = \varepsilon^{\star}(\tau) = \sqrt{\frac{1 \vee \log(\tau)}{n}}$$

Power Analysis

Theorem Kim and Ramdas (2025)

For any τ , suppose that $\varepsilon \gg \varepsilon^*$. Then the asymptotic uniform power of the DA test is equal to one:

$$\lim_{n \rightarrow \infty} \inf_{P \in \mathcal{P}_{1,r}(\varepsilon; \tau)} P(\psi_r = 1) = 1$$



$$\varepsilon^* = \sqrt{\frac{1 \vee \log(\tau)}{n}}$$

Power Analysis

Theorem Kim and Ramdas (2025)

For any τ , suppose that $\varepsilon \gg \varepsilon^*$. Then the asymptotic uniform power of the DA test is equal to one:

$$\lim_{n \rightarrow \infty} \inf_{P \in \mathcal{P}_{1,r}(\varepsilon; \tau)} P(\psi_r = 1) = 1$$

- Adaptivity to unknown $|\mathbb{C}_r|$
- The rate changes from $1/\sqrt{n}$ -rate to $\sqrt{\log(d)/n}$



$$\varepsilon^* = \sqrt{\frac{1 \vee \log(\tau)}{n}}$$

Power Analysis

Theorem Kim and Ramdas (2025)

For any τ , suppose that $\varepsilon \gg \varepsilon^*$. Then the asymptotic uniform power of the DA test is equal to one:

$$\lim_{n \rightarrow \infty} \inf_{P \in \mathcal{P}_{1,r}(\varepsilon; \tau)} P(\psi_r = 1) = 1$$

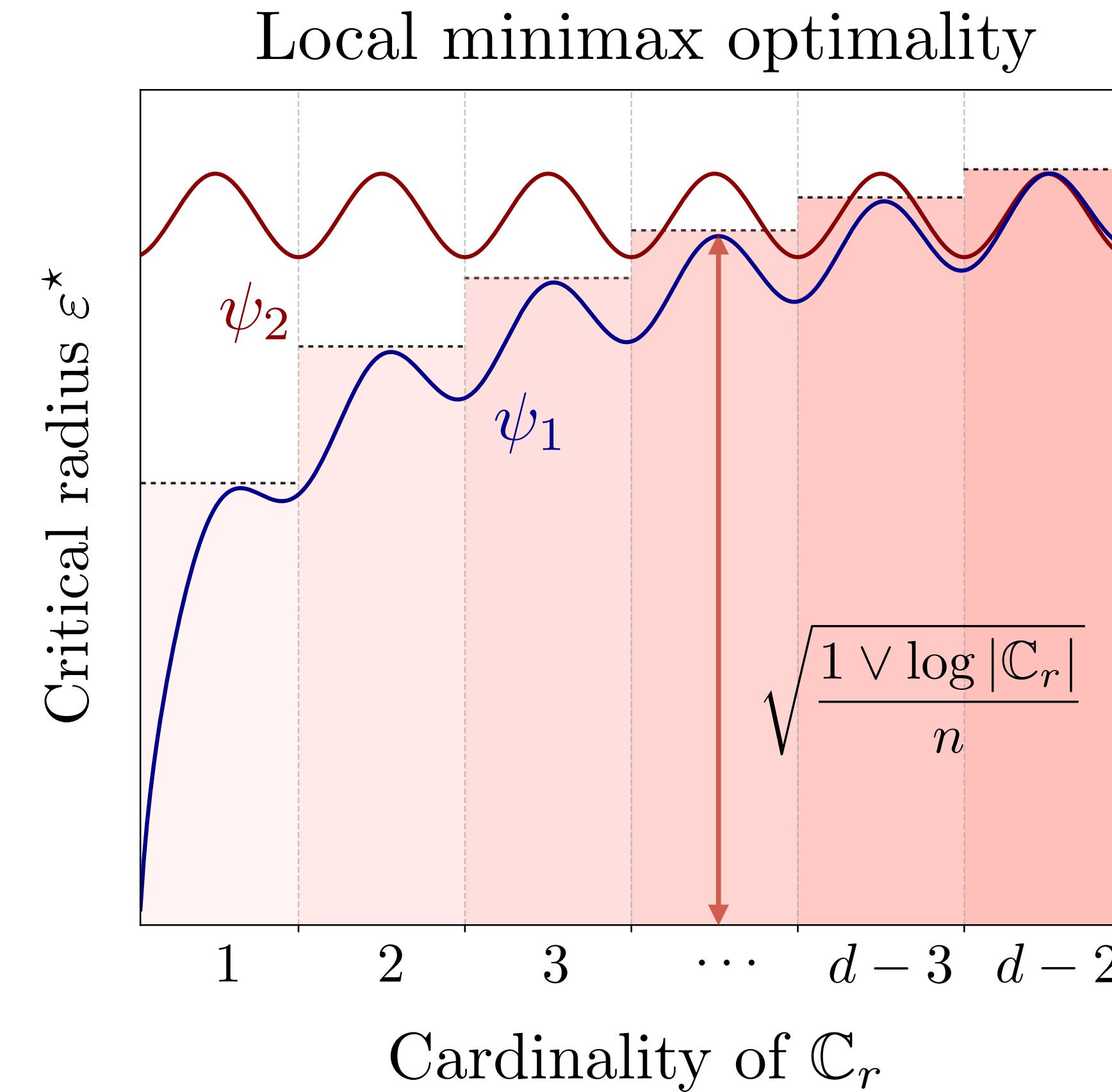
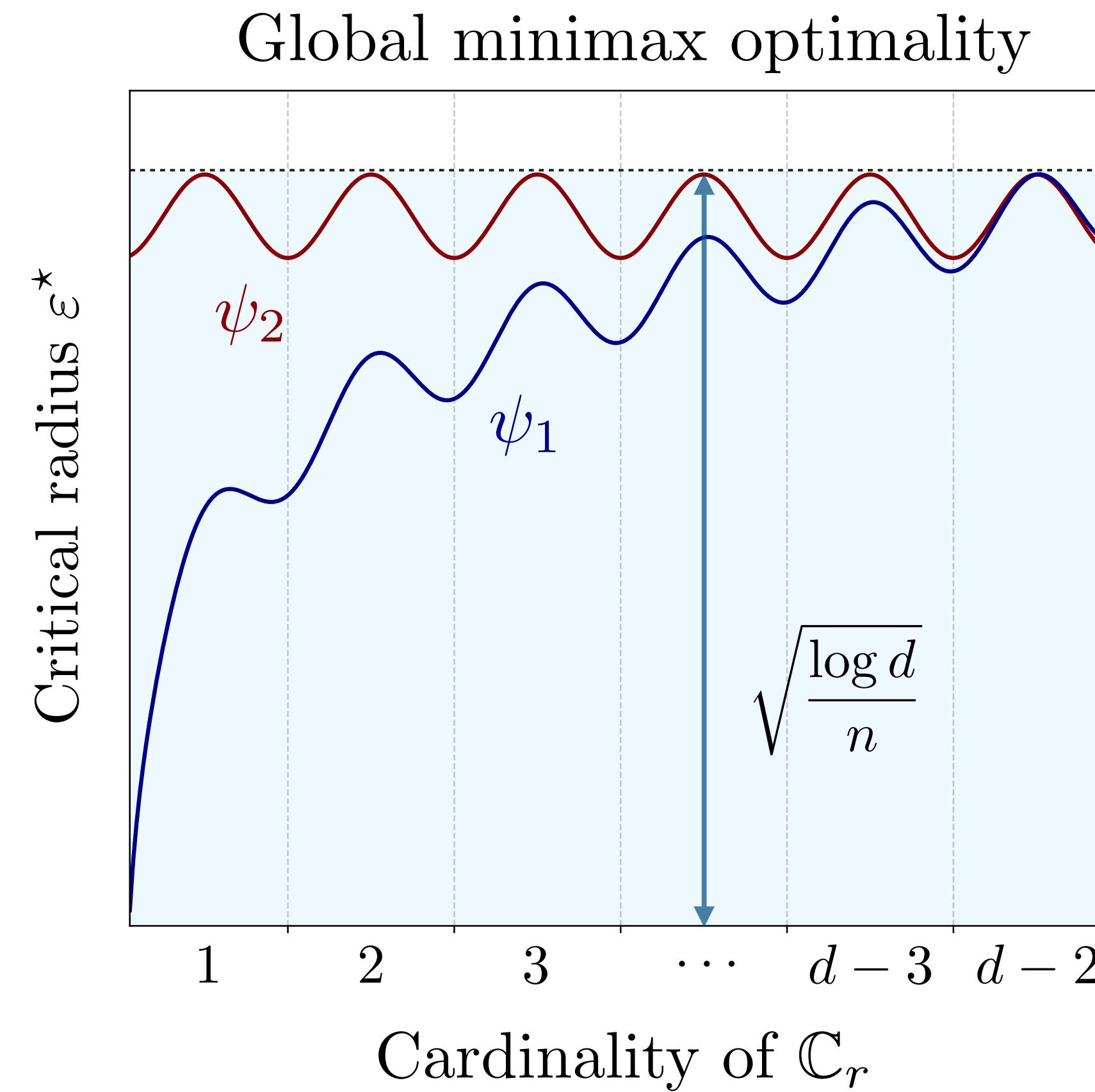
- Adaptivity to unknown $|\mathbb{C}_r|$
- The rate changes from $1/\sqrt{n}$ -rate to $\sqrt{\log(d)/n}$

Question. Can we further **improve** the **separation rate**?



$$\varepsilon^* = \sqrt{\frac{1 \vee \log(\tau)}{n}}$$

Local Minimax Optimality



Local Minimax Optimality

Theorem Kim and Ramdas (2025)

Let Ψ_α be the set of all asymptotic level- α tests over $\mathcal{P}_{0,\textcolor{red}{r}}$,

$$\Psi_{\textcolor{teal}{\alpha}} := \left\{ \psi : \limsup_{n \rightarrow \infty} \sup_{\textcolor{blue}{P} \in \mathcal{P}_{0,\textcolor{red}{r}}} \textcolor{blue}{P}(\psi = 1) \leq \alpha \right\}$$

Local Minimax Optimality

Theorem Kim and Ramdas (2025)

Let Ψ_α be the set of all asymptotic level- α tests over $\mathcal{P}_{0,\textcolor{red}{r}}$,

$$\Psi_\alpha := \left\{ \psi : \limsup_{n \rightarrow \infty} \sup_{P \in \mathcal{P}_{0,\textcolor{red}{r}}} P(\psi = 1) \leq \alpha \right\}$$

If $\varepsilon \ll \varepsilon^\star$, then the **asymptotic type II error** is at least β :

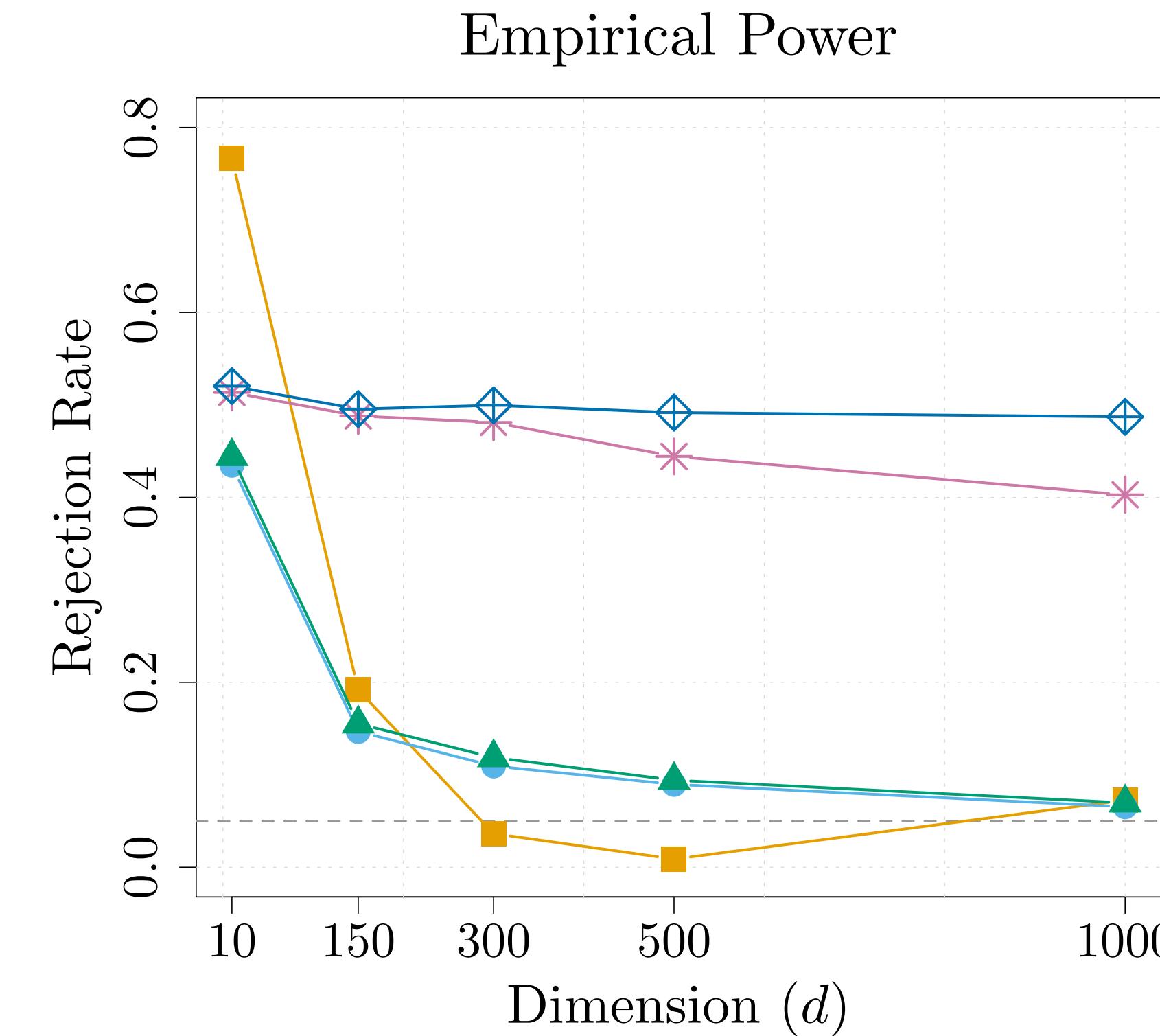
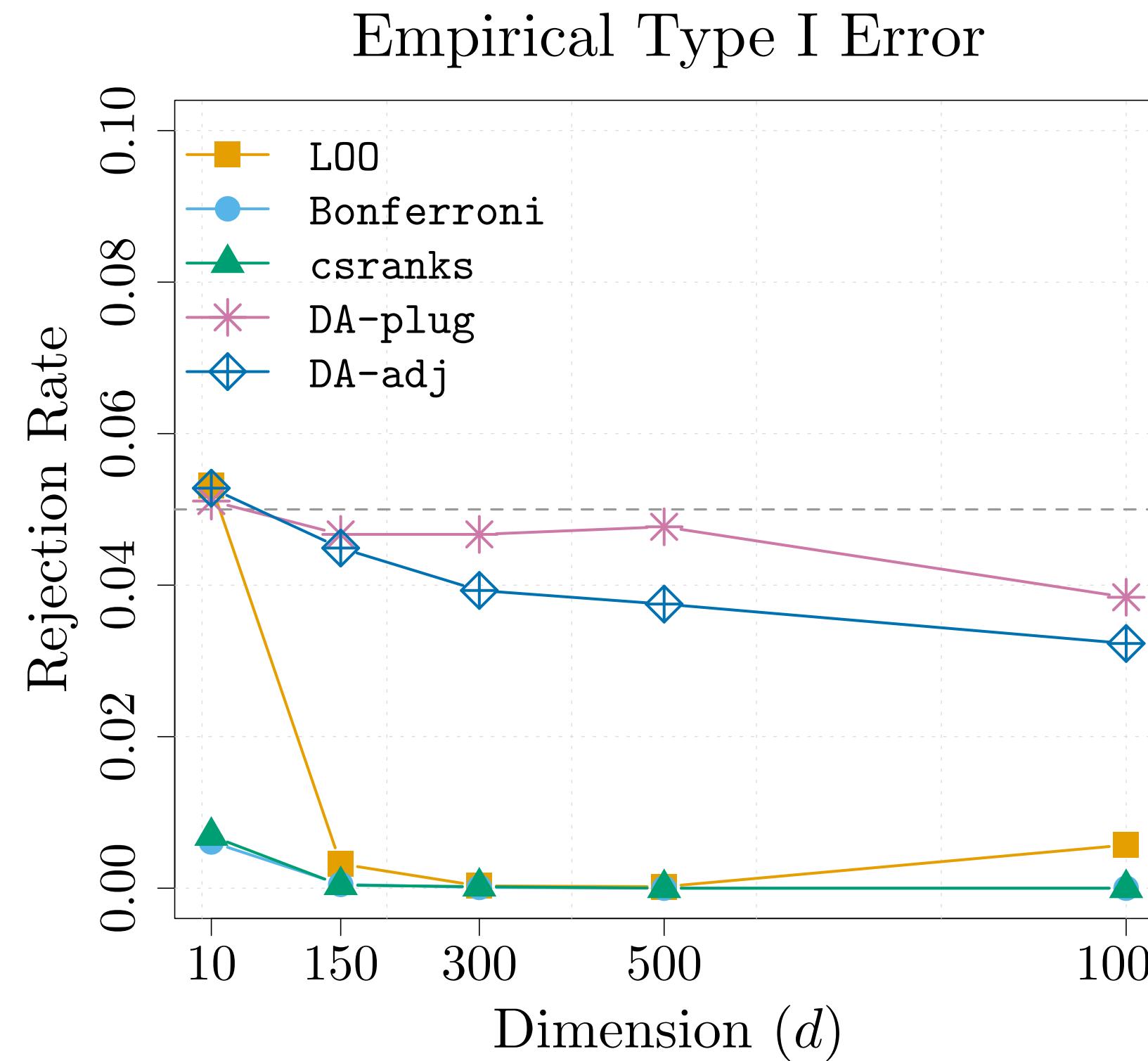
$$\liminf_{n \rightarrow \infty} \inf_{\psi \in \Psi_\alpha} \sup_{P \in \mathcal{P}_{1,\textcolor{red}{r}}(\varepsilon; \tau)} P(\psi = 0) \geq \beta$$



$$\varepsilon^\star = \sqrt{\frac{1 \vee \log(\tau)}{n}}$$

Empirical Results

Power and Validity in High-Dimensional Settings



- **L0O:** Zhang et al. (2024)
- **Bonferroni:** multiple correction
- **csranks:** Mogstad et al. (2024)
- **DA-plug:** plug-in \hat{s}
- **DA-adj:** noise-adjusted \hat{s}

• Null Setting

$$\mu = (0, 0, 1, \dots, 1)^\top$$
$$\Sigma_{11} = \Sigma_{22} = 1 \text{ & } \Sigma_{33} = \dots = \Sigma_{dd} = 20$$

• Alternative Setting

$$\mu = (0.15, 0, 1, \dots, 1)^\top$$
$$\Sigma_{11} = \Sigma_{22} = 1 \text{ & } \Sigma_{33} = \dots = \Sigma_{dd} = 20$$

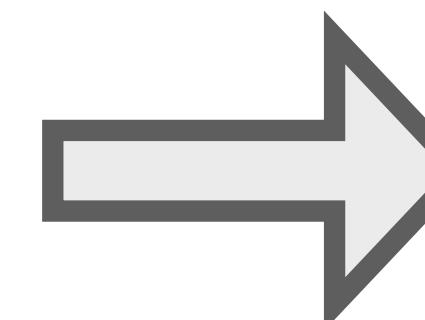
Power under Various Settings

Method	$\mu^{(a)} + \text{unequal variance}$			$\mu^{(b)} + \text{unequal variance}$			$\mu^{(c)} + \text{unequal variance}$		
	$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$	$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$	$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$
L00	0.084	0.115	0.380	0.000	0.001	0.181	0.258	0.351	0.703
Bonferroni	0.171	0.130	0.055	0.166	0.103	0.030	0.017	0.006	0.003
csranks	0.184	0.381	0.962	0.162	0.363	0.961	0.019	0.041	0.223
MCS	0.004	0.002	0.004	0.000	0.000	0.000	0.140	0.156	0.166
DA-plug	0.049	0.052	0.042	0.062	0.067	0.059	0.098	0.128	0.202
DA-plug $^{ \times 10 }$	0.050	0.052	0.050	0.080	0.080	0.073	0.125	0.145	0.240
DA-adj	0.122	0.259	0.841	0.217	0.384	0.916	0.135	0.188	0.462
DA-adj $^{ \times 10 }$	0.160	0.343	0.946	0.294	0.517	0.982	0.164	0.251	0.605

Power under Various Settings

Method	$\mu^{(a)} + \text{unequal variance}$			$\mu^{(b)} + \text{unequal variance}$			$\mu^{(c)} + \text{unequal variance}$		
	$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$	$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$	$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$
L0O	0.084	0.115	0.380	0.000	0.001	0.181	0.258	0.351	0.703
Bonferroni	0.171	0.130	0.055	0.166	0.103	0.030	0.017	0.006	0.003
csranks	0.184	0.381	0.962	0.162	0.363	0.961	0.019	0.041	0.223
MCS	0.004	0.002	0.004	0.000	0.000	0.000	0.140	0.156	0.166
DA-plug	0.049	0.052	0.042	0.062	0.067	0.059	0.098	0.128	0.202
DA-plug $^{ \times 10}$	0.050	0.052	0.050	0.080	0.080	0.073	0.125	0.145	0.240
DA-adj	0.122	0.259	0.841	0.217	0.384	0.916	0.135	0.188	0.462
DA-adj $^{ \times 10}$	0.160	0.343	0.946	0.294	0.517	0.982	0.164	0.251	0.605

- DA-plug: plug-in \hat{s}
- DA-adj: noise-adjusted \hat{s}
- DA-plug $^{ \times 10}$: 10 splits + average
- DA-adj $^{ \times 10}$: 10 splits + average



1. **Multiple splits** improve the power
2. **Noise-adjusted** version performs better

Power under Various Settings

Method	$\mu^{(a)}$ + unequal variance			$\mu^{(b)}$ + unequal variance			$\mu^{(c)}$ + unequal variance		
	$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$	$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$	$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$
L00	0.084	0.115	0.380	0.000	0.001	0.181	0.258	0.351	0.703
Bonferroni	0.171	0.130	0.055	0.166	0.103	0.030	0.017	0.006	0.003
csranks	0.184	0.381	0.962	0.162	0.363	0.961	0.019	0.041	0.223
MCS	0.004	0.002	0.004	0.000	0.000	0.000	0.140	0.156	0.166
DA-plug	0.049	0.052	0.042	0.062	0.067	0.059	0.098	0.128	0.202
DA-plug $^{ \times 10}$	0.050	0.052	0.050	0.080	0.080	0.073	0.125	0.145	0.240
DA-adj	0.122	0.259	0.841	0.217	0.384	0.916	0.135	0.188	0.462
DA-adj $^{ \times 10}$	0.160	0.343	0.946	0.294	0.517	0.982	0.164	0.251	0.605

DA-adj $^{ \times 10}$ achieves the **best or second-best** performance
 (with only a **small** margin)

Mean Structures

$$\mu^{(a)} = (0.1, 0, 0.1, \dots, 0.1)^\top \quad \mu^{(b)} = (0.1, 0.019, \dots, 0.99, 1)^\top$$

Power under Various Settings

Method	$\mu^{(a)} + \text{unequal variance}$			$\mu^{(b)} + \text{unequal variance}$			$\mu^{(c)} + \text{unequal variance}$		
	$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$	$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$	$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$
LOO	0.084	0.115	0.380	0.000	0.001	0.181	0.258	0.351	0.703
Bonferroni	0.171	0.130	0.055	0.166	0.103	0.030	0.017	0.006	0.003
csranks	0.184	0.381	0.962	0.162	0.363	0.961	0.019	0.041	0.223
MCS	0.004	0.002	0.004	0.000	0.000	0.000	0.140	0.156	0.166
DA-plug	0.049	0.052	0.042	0.062	0.067	0.059	0.098	0.128	0.202
DA-plug $^{ \times 10 }$	0.050	0.052	0.050	0.080	0.080	0.073	0.125	0.145	0.240
DA-adj	0.122	0.259	0.841	0.217	0.384	0.916	0.135	0.188	0.462
DA-adj $^{ \times 10 }$	0.160	0.343	0.946	0.294	0.517	0.982	0.164	0.251	0.605

LOO performs **best** and DA-adj $^{ \times 10 }$ performs **second best**
 But LOO **does not** control the type I error rate

Mean Structure

$$\boldsymbol{\mu}^{(c)} = (0.05, 0, 0, 0, 10, \dots, 10)^\top$$

Power under Various Settings

Method	$\mu^{(a)}$ + unequal variance			$\mu^{(b)}$ + unequal variance			$\mu^{(c,0)}$ + unequal variance		
	$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$	$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$	$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$
LOO	0.084	0.115	0.380	0.000	0.001	0.181	0.070	0.064	0.065
Bonferroni	0.171	0.130	0.055	0.166	0.103	0.030	0.002	0.001	0.001
csranks	0.184	0.381	0.962	0.162	0.363	0.961	0.002	0.001	0.002
MCS	0.004	0.002	0.004	0.000	0.000	0.000	0.042	0.042	0.034
DA-plug	0.049	0.052	0.042	0.062	0.067	0.059	0.048	0.052	0.048
DA-plug $^{ \times 10}$	0.050	0.052	0.050	0.080	0.080	0.073	0.053	0.046	0.047
DA-adj	0.122	0.259	0.841	0.217	0.384	0.916	0.054	0.052	0.050
DA-adj $^{ \times 10}$	0.160	0.343	0.946	0.294	0.517	0.982	0.053	0.049	0.050

LOO performs **best** and DA-adj $^{ \times 10}$ performs **second best**
 But LOO **does not** control the type I error rate

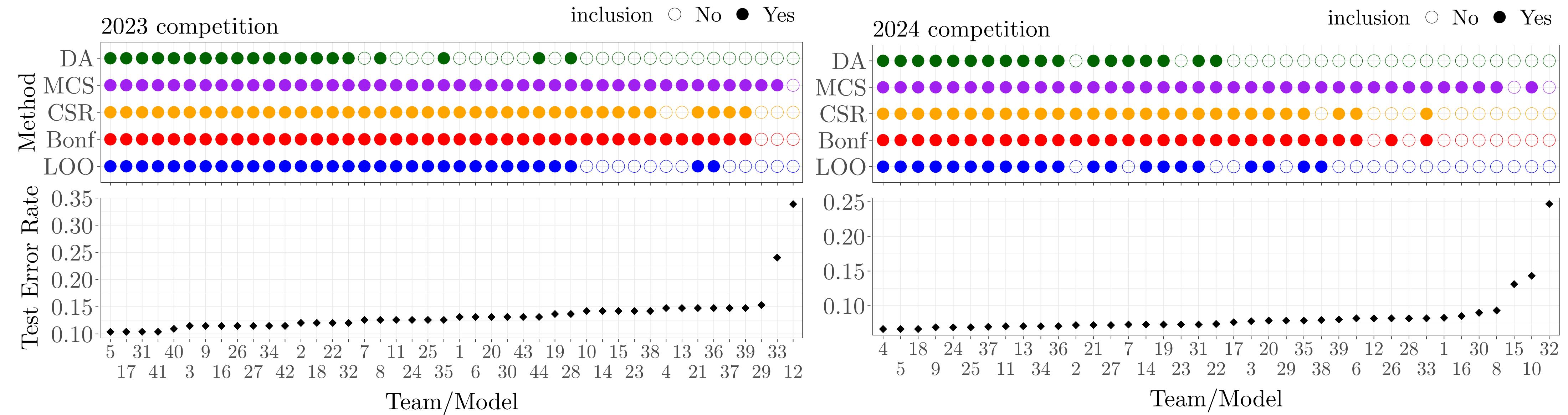
Mean Structure

$$\boldsymbol{\mu}^{(c)} = (0.05, 0, 0, 0, 10, \dots, 10)^\top$$

Real-World Data

- We revisit **CMU 36-462 prediction competition** analyzed by Zhang et al. (2024)
 - Spring 2023: $n = 183, p = 44$
 - Spring 2024: $n = 1246, p = 39$
- $X_{i,k} \in \{0,1\}$: correct/incorrect prediction by team k at test point i
- μ_k : the population prediction risk of team k (unknown)
- Methods to compare
 - Bonferroni correction (**Bonf**)
 - Hansen et al. (2011) (**MCS**)
 - Mogstad et al. (2024) (**CSR**)
 - Zhang et al. (2024) (**LOO**)

Real-World Data



- **Colored points** indicate the indices included in **each confidence set**
- Our method (**DA-adj $\times 10$**) produces confidence sets with **smallest cardinality**

Summary

- We have introduced a DA method for **high-dimensional argmin inference** problem based on **sample splitting** and **studentization**
- The proposed method achieves the **locally minimax separation rate** and adapts to the intrinsic difficulty of the problem characterized by the **confusion set**
- We have demonstrated its **strong empirical performance** under various settings

Thank you!



Coverage Guarantees

I. Pointwise coverage

Each $r \in \Theta$ is included in $\widehat{\Theta}$ with high probability:

$$\inf_{P \in \mathcal{P}} \inf_{r \in \Theta(P)} P(r \in \widehat{\Theta}) \geq 1 - \alpha$$

- **Less** demanding → **smaller** expected set size
- **Higher** power but **weaker** protection

2. Uniform coverage

The entire set Θ is contained in $\widehat{\Theta}$ with high probability:

$$\inf_{P \in \mathcal{P}} P(\Theta \subseteq \widehat{\Theta}) \geq 1 - \alpha$$

- **More** demanding → **larger** expected set size
- **Stronger** protection but **more** conservative



\mathcal{P} : class of distributions
 α : target error level

Future directions

- General **rank-k** inference
- Confidence sets for $\mu_{(k)}$
- **Computationally efficient** multiple splitting approach