

# **A New Perspective on Pool-Based Active Classification and False Discovery Control**

Lalit Jain, Kevin Jamieson

Paul G. Allen Department of Computer Science and Engineering

University of Washington

# Contributions

- **Revisiting Classification and TPR maximization subject to FDR-control in a combinatorial setting.**
- **A common framework for classification and pure exploration for combinatorial bandits**
- **State of the art action elimination algorithms for both.**

# Formalities

$[n]$  : a finite item space

$y_i \in \{0, 1\}$ : label for  $i \in [n]$

$\mathcal{H}_0 = \{i \in [n] : y_i = 0\}, \mathcal{H}_1 = [n] \setminus \mathcal{H}_0$

## Policies:

$\Pi \subset 2^{[n]}$ : hypothesis class

Identify each  $\pi \in \Pi$  with a labeler:

$\pi \in \Pi: \pi(x) = 1 \iff x \in \pi$

## Metrics:

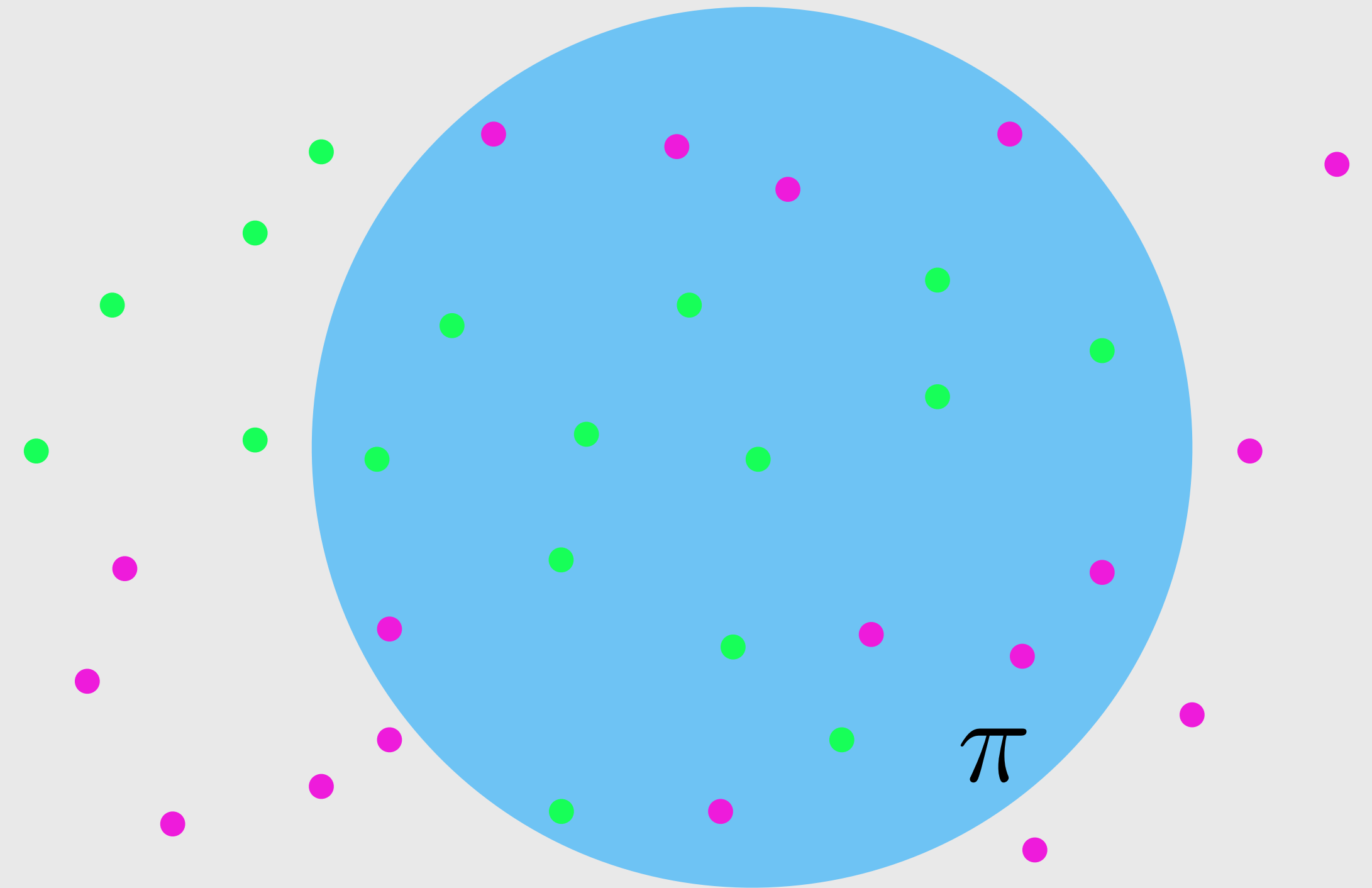
$$R(\pi) := \frac{|\pi \cap \mathcal{H}_0| + |\pi^c \cap \mathcal{H}_1|}{n}$$

$$FDR(\pi) := \frac{|\pi \cap \mathcal{H}_0|}{|\pi|}, \quad TPR(\pi) := \frac{|\pi \cap \mathcal{H}_1|}{|\mathcal{H}_1|}$$

## Example:

$$R(\pi) = .88$$

$$FDR(\pi) = .60, TPR(\pi) = .44$$



# Metrics

## Policies:

$\Pi \subset 2^{[n]}$ : hypothesis class

Identify each  $\pi \in \Pi$  with a labeler:

$$\pi \in \Pi: \pi(x) = 1 \iff x \in \pi$$

## Metrics:

$$R(\pi) := \frac{|\pi \cap \mathcal{H}_0| + |\pi^c \cap \mathcal{H}_1|}{n}$$

$$FDR(\pi) := \frac{|\pi \cap \mathcal{H}_0|}{|\pi|}, \quad TPR(\pi) := \frac{|\pi \cap \mathcal{H}_1|}{|\mathcal{H}_1|}$$

## Classification Problem:

Identify  $\pi^* := \operatorname{argmin}_{\pi \in \Pi} R(\pi)$

## FDR Control Problem:

Identify  $\pi_\alpha^* := \operatorname{argmax}_{\pi \in \Pi, FDR(\pi) \leq \alpha} TPR(\pi)$

# Revisiting Classification Loss

Re-parametrize:  $\mu_x = 2y_x - 1 = \begin{cases} 1 & y_x = 1 \\ -1 & y_x = 0 \end{cases}$

**Define:**

$$\mu_\pi := \sum_{x \in \pi} \mu_x$$

$$\begin{aligned} R(\pi) &= \frac{|\pi \cap \mathcal{H}_0| + |\pi^c \cap \mathcal{H}_1|}{n} \\ &= \frac{1}{n} \sum_{x \in \pi} \mathbf{1}\{y_x = 0\} + \frac{1}{n} \sum_{x \in \pi^c} \mathbf{1}\{y_x = 1\} \\ &= \frac{1}{n} \sum_{x \in \pi} (1 - y_x) + \frac{1}{n} \sum_{x \in \pi^c} y_x \\ &= \frac{|\mathcal{H}_1|}{n} + \sum_{x \in \pi} (1 - 2y_x) \\ &= \frac{|\mathcal{H}_1| + |\pi|}{n} - \frac{1}{n} \sum_{x \in \pi} \mu_x \end{aligned}$$

**Key Takeaway:**

$$\pi^* = \operatorname{argmin}_{\pi \in \Pi} R(\pi) = \operatorname{argmax}_{\pi \in \Pi} \mu_\pi$$

**A combinatorial interpretation of classification!**

# Detour: Pure Exploration Combinatorial Bandits

## Input:

$\nu_i, \mathbb{E}[\nu_i] = \mu_i, i \in [n]$  arm distributions

$\Pi \subset 2^{\mathcal{X}}$  collection of subsets

$$\mu_\pi := \sum_{x \in \pi} \mu_x$$

## Protocol:

In round  $t$  choose  $I_t$  and receive reward  $R_{I_t, t} \sim \nu_{I_t}$

## Return:

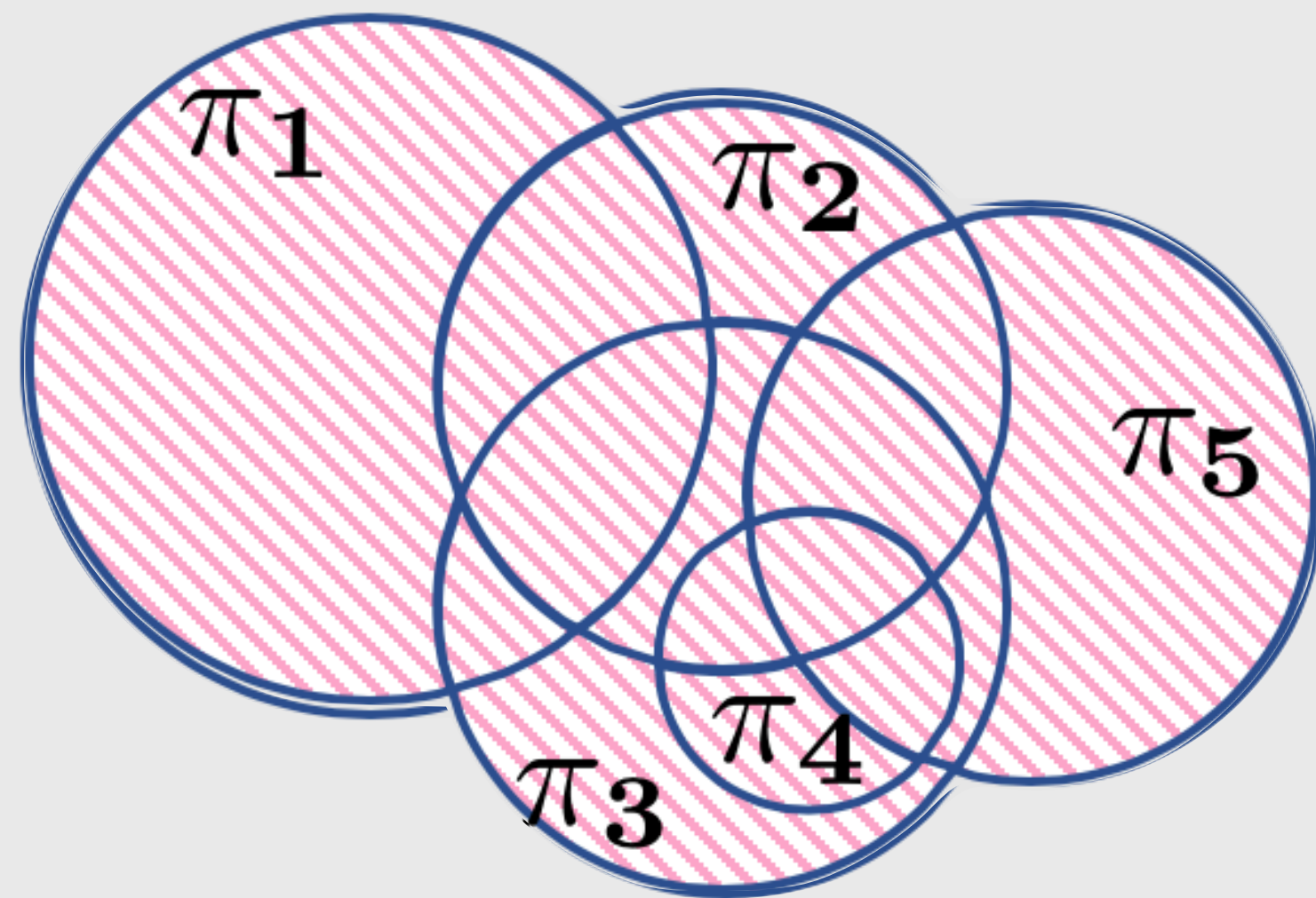
Identify  $\pi^* = \underset{\pi \in \Pi}{\operatorname{argmax}} \sum_{x \in \pi} \mu_x$

## Examples:

- $\Pi = \{\{i\} : i \in [n]\}$ . Best-Arm identification.
- $\Pi = \binom{[n]}{k}$ . Top-k.
- $\Pi = \{\text{trees in a weighted graph}\}$ . Minimal Spanning Tree
- Classification (stochastic or persistent labels).

Amazingly we can transform classification to a combinatorial bandit problem!

# Algorithm: Action Elimination



## Action Elimination:

Input:  $\delta, \Pi$

**for:**  $t = 1, 2, \dots$ ,

1. **Sample**  $I_t$  uniformly at random from  $[n]$   
(with or without replacement).

2. If  $I_s \in \cup_{\pi \in \mathcal{A}} \pi - \cap_{\pi \in \mathcal{A}} \pi$ , **observe**  $\mu_{I_t}$

3. For each  $\pi, \pi'$ , **update**  $\hat{\mu}(\pi', \pi)$

4. **Eliminate**  $\mathcal{A} = \mathcal{A} \setminus \{\pi : \exists \pi', \hat{\mu}(\pi', \pi) - C(\pi', \pi, t) > 0\}$



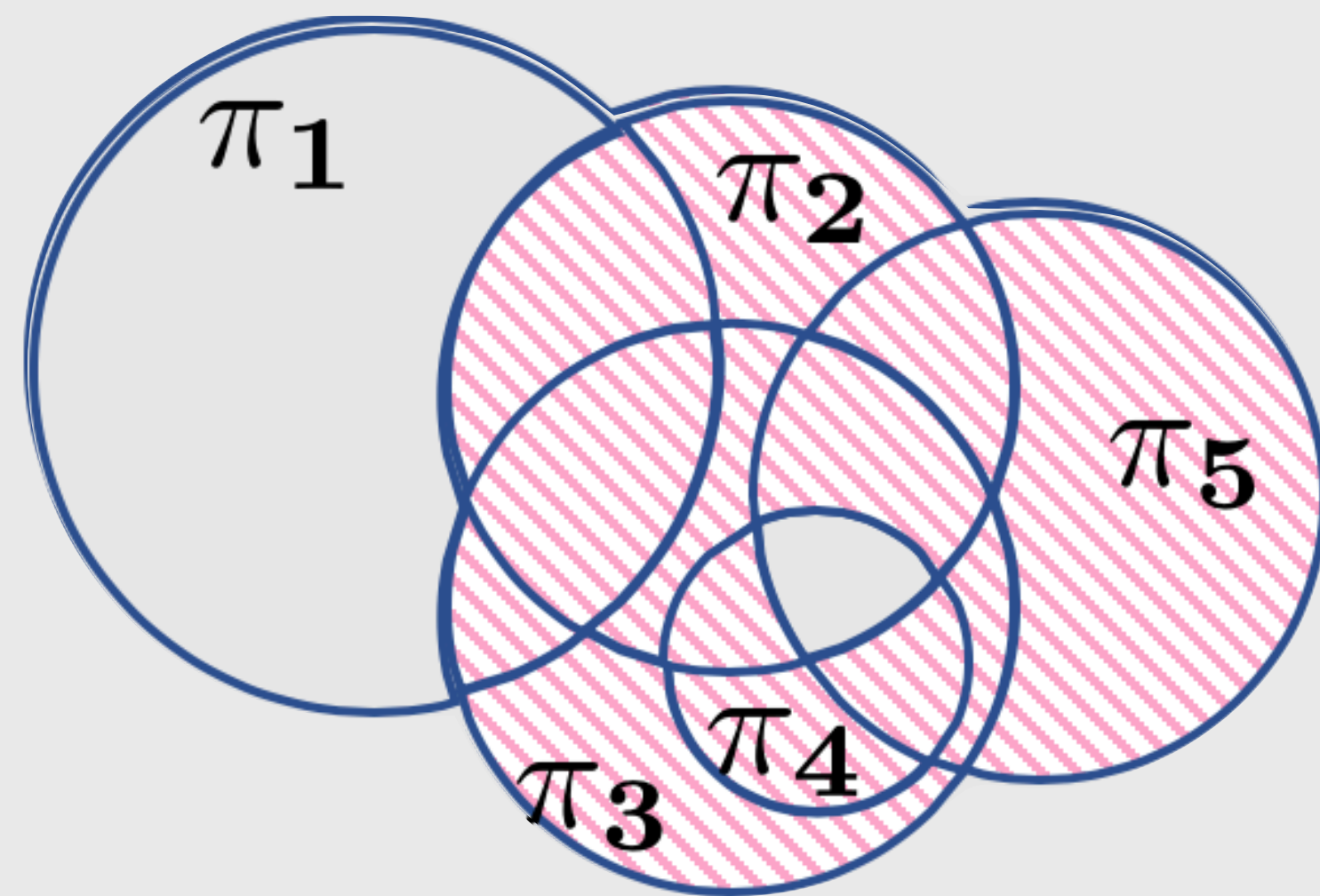
Carefully designed to  
avoid union bounds

Return  $\pi^*$

**Takeaway:** For active classification, we only need to sample points that any labeler is uncertain about.



# Algorithm: Action Elimination



## Action Elimination:

Input:  $\delta, \Pi$

**for:**  $t = 1, 2, \dots$ ,

1. **Sample**  $I_t$  uniformly at random from  $[n]$   
(with or without replacement).

2. If  $I_s \in \cup_{\pi \in \mathcal{A}} \pi - \cap_{\pi \in \mathcal{A}} \pi$ , **observe**  $\mu_{I_t}$

3. For each  $\pi, \pi'$ , **update**  $\hat{\mu}(\pi', \pi)$

4. **Eliminate**  $\mathcal{A} = \mathcal{A} \setminus \{\pi : \exists \pi', \hat{\mu}(\pi', \pi) - C(\pi', \pi, t) > 0\}$



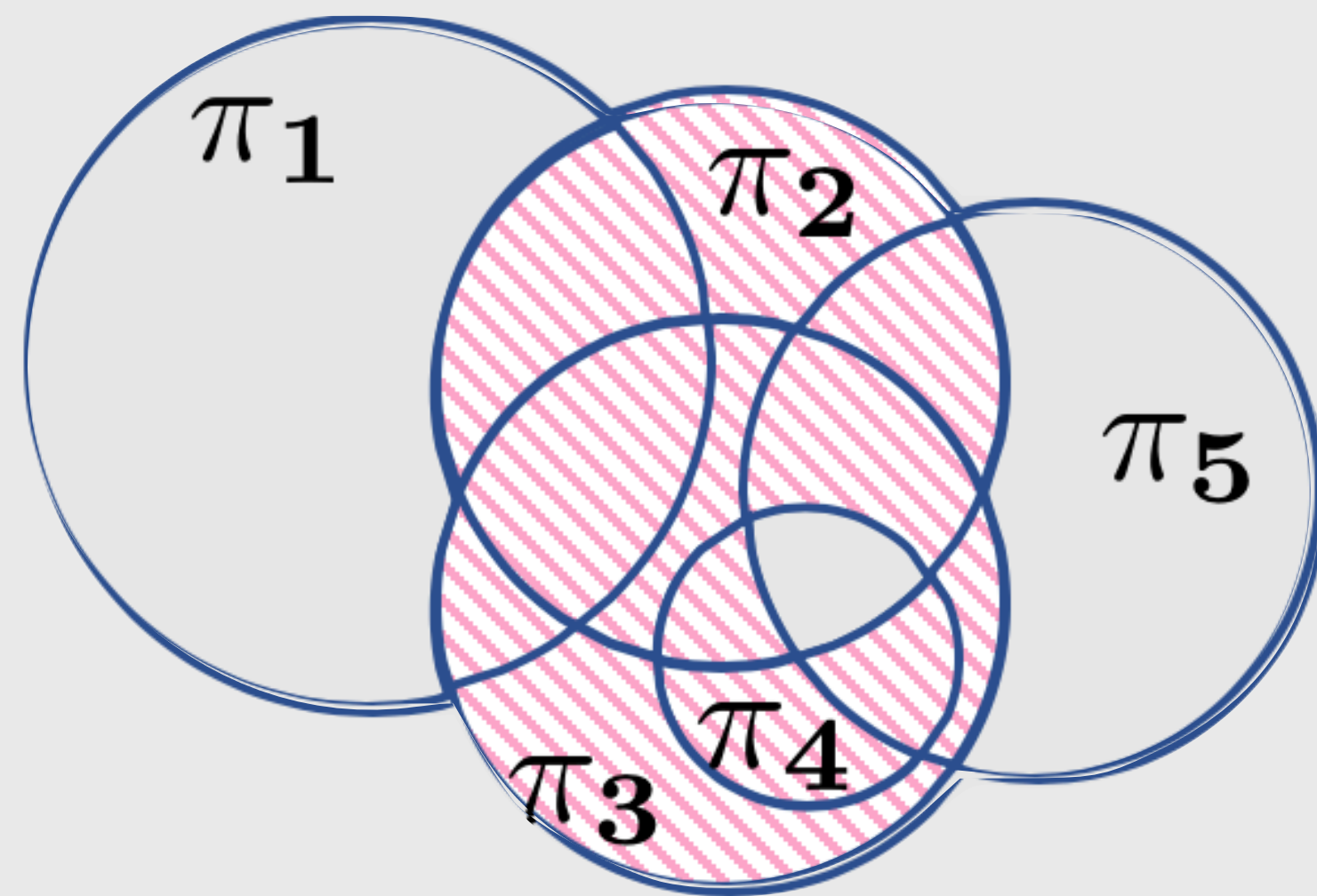
Carefully designed to  
avoid union bounds

Return  $\pi^*$

**Takeaway:** For active classification, we only need to sample points that any labeler is uncertain about.



# Algorithm: Action Elimination



$\pi_2$  knocks out  $\pi_5$

## Action Elimination:

Input:  $\delta, \Pi$

**for:**  $t = 1, 2, \dots$ ,

1. **Sample**  $I_t$  uniformly at random from  $[n]$   
(with or without replacement).

2. If  $I_s \in \cup_{\pi \in \mathcal{A}} \pi - \cap_{\pi \in \mathcal{A}} \pi$ , **observe**  $\mu_{I_t}$

3. For each  $\pi, \pi'$ , **update**  $\hat{\mu}(\pi', \pi)$

4. **Eliminate**  $\mathcal{A} = \mathcal{A} \setminus \{\pi : \exists \pi', \hat{\mu}(\pi', \pi) - C(\pi', \pi, t) > 0\}$

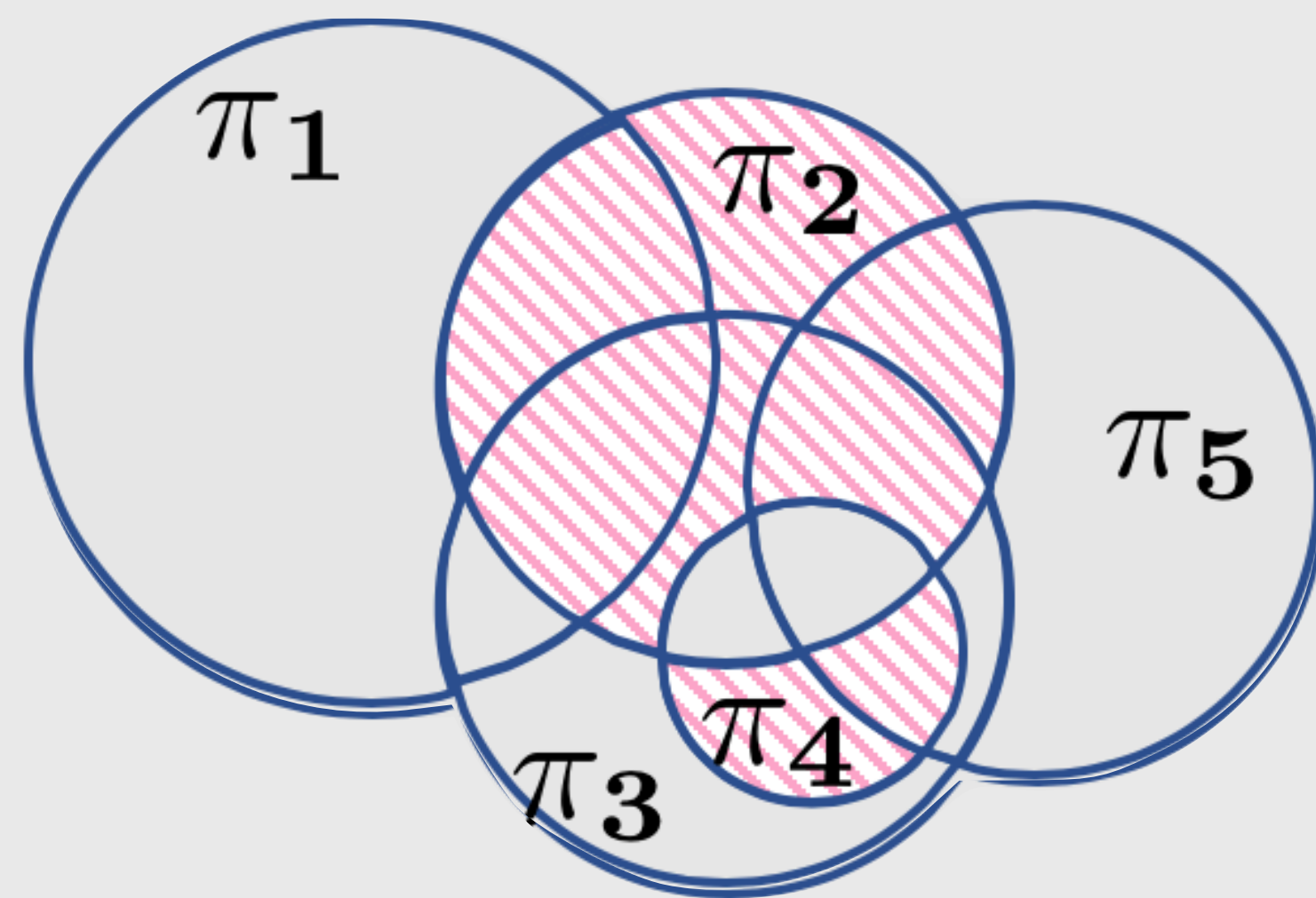


Carefully designed to  
avoid union bounds

Return  $\pi^*$

**Takeaway:** For active classification, we only need to sample points that any labeler is uncertain about.

# Algorithm: Action Elimination



## Action Elimination:

Input:  $\delta, \Pi$

**for:**  $t = 1, 2, \dots$ ,

1. **Sample**  $I_t$  uniformly at random from  $[n]$   
(with or without replacement).

2. If  $I_s \in \cup_{\pi \in \mathcal{A}} \pi - \cap_{\pi \in \mathcal{A}} \pi$ , **observe**  $\mu_{I_t}$

3. For each  $\pi, \pi'$ , **update**  $\hat{\mu}(\pi', \pi)$

4. **Eliminate**  $\mathcal{A} = \mathcal{A} \setminus \{\pi : \exists \pi', \hat{\mu}(\pi', \pi) - C(\pi', \pi, t) > 0\}$

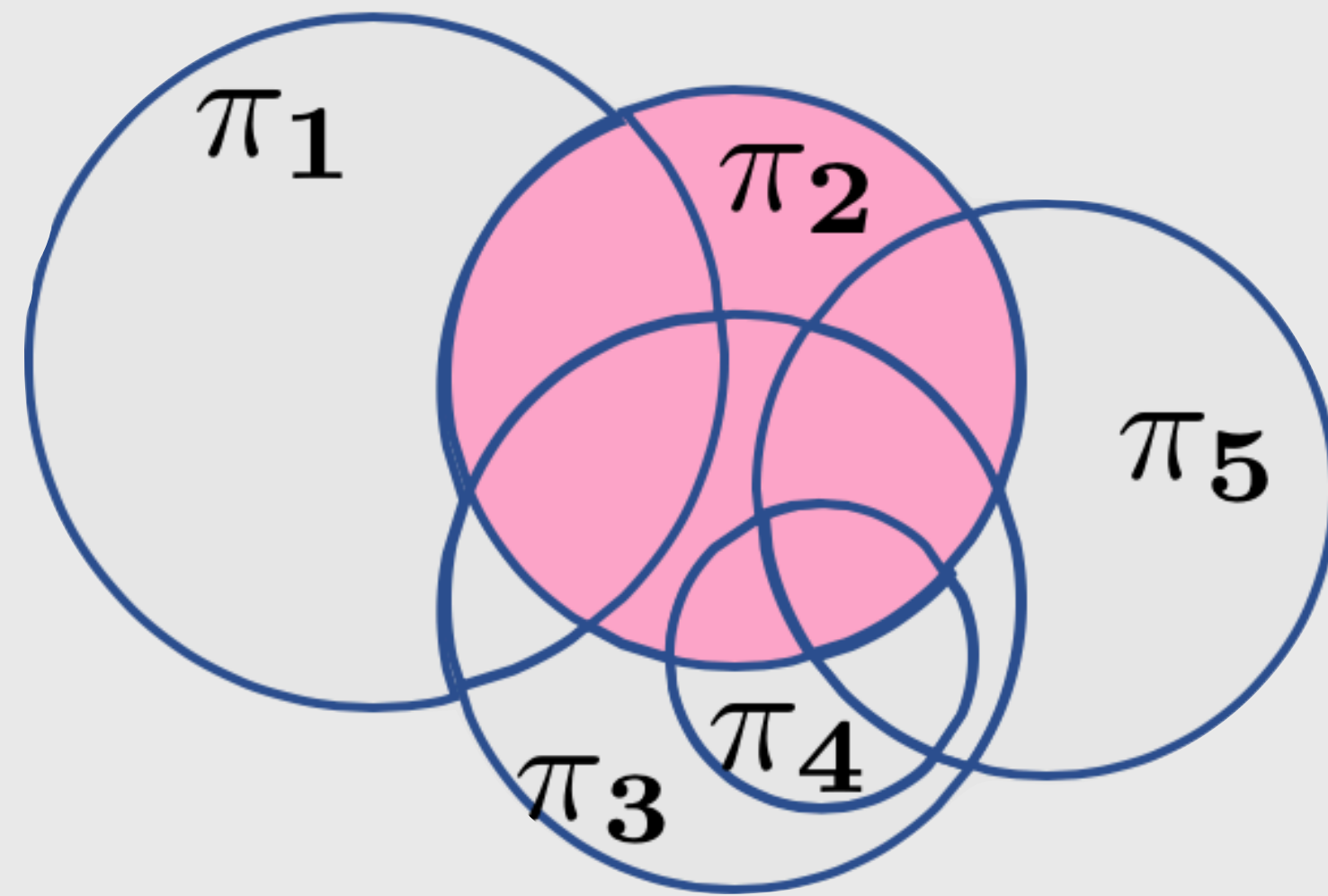


Carefully designed to  
avoid union bounds

Return  $\pi^*$

**Takeaway:** For active classification, we only need to sample points that any labeler is uncertain about.

# Algorithm: Action Elimination



$\pi_2$  knocks out  $\pi_4$

## Action Elimination:

Input:  $\delta, \Pi$

**for:**  $t = 1, 2, \dots$ ,

1. **Sample**  $I_t$  uniformly at random from  $[n]$   
(with or without replacement).

2. If  $I_s \in \cup_{\pi \in \mathcal{A}} \pi - \cap_{\pi \in \mathcal{A}} \pi$ , **observe**  $\mu_{I_t}$

3. For each  $\pi, \pi'$ , **update**  $\hat{\mu}(\pi', \pi)$

4. **Eliminate**  $\mathcal{A} = \mathcal{A} \setminus \{\pi : \exists \pi', \hat{\mu}(\pi', \pi) - C(\pi', \pi, t) > 0\}$



Carefully designed to  
avoid union bounds

Return  $\pi^*$

**Takeaway:** For active classification, we only need to sample points that any labeler is uncertain about.

# Guarantees for Action Elimination

**Theorem (J., Jamieson 2019)** Let  $\pi^* = \operatorname{argmin}_{\pi \in \Pi} R(\pi)$ ,

$$\tilde{\Delta}_\pi = \frac{|\mu_{\pi^*} - \mu_\pi|}{|\pi^* \Delta \pi|}, \text{ and } B(k, \pi^*) = \{\pi : |\pi \Delta \pi^*| = k\}$$

$$\tau_\pi = \frac{VC(B(|\pi^* \Delta \pi|, \pi^*)) \log(n \log(\tilde{\Delta}_\pi^{-2})/\delta)}{\tilde{\Delta}_\pi^2}$$

Then with probability greater than  $1 - \delta$  the Action Elimination algorithm terminates after a number of samples no more than

$$\sum_{1 \leq i \leq n} \max_{\pi: i \in \pi^* \Delta \pi} \tau_\pi$$

**Interpretation:** we can stop sampling  $i$  after  $\pi^*$  has knocked out any set  $i \in \pi^* \Delta \pi$ .

# Implications: Classification

An active classification complexity that is fundamentally not disagreement coefficient based!

Contrasts with the standard DHM algorithm:

- Like DHM we sample in the symmetric difference. Unlike DHM, we can characterize the contribution of each arm.
- For Best-Arm, DHM analysis gives the passive rate  $n\Delta_2^2$ .
- For 1-d thresholds under Tsybakov noise get the minimax rates (Castro and Nowak 2008):

$$\text{if } \alpha = 0, \log(n) \log(\log(n)/\delta)/h^2 \quad \text{if } \alpha > 0, n^{2\alpha} \log(\log(n)/\delta)/h^2$$

passive rates:

$$n^{2\alpha+1} \log(\log(n)/\delta)/h^2$$



# Remarks

- An active classification complexity that is fundamentally not disagreement coefficient based!
- Can quantify the contribution of each arm to the sample complexity
- For 1-d thresholds under Tsybakov noise get the minimax rates (Castro and Nowak 2008):
- An agnostic algorithm that matches the rates of previous binary classification !
- **No need to pull each arm once.**

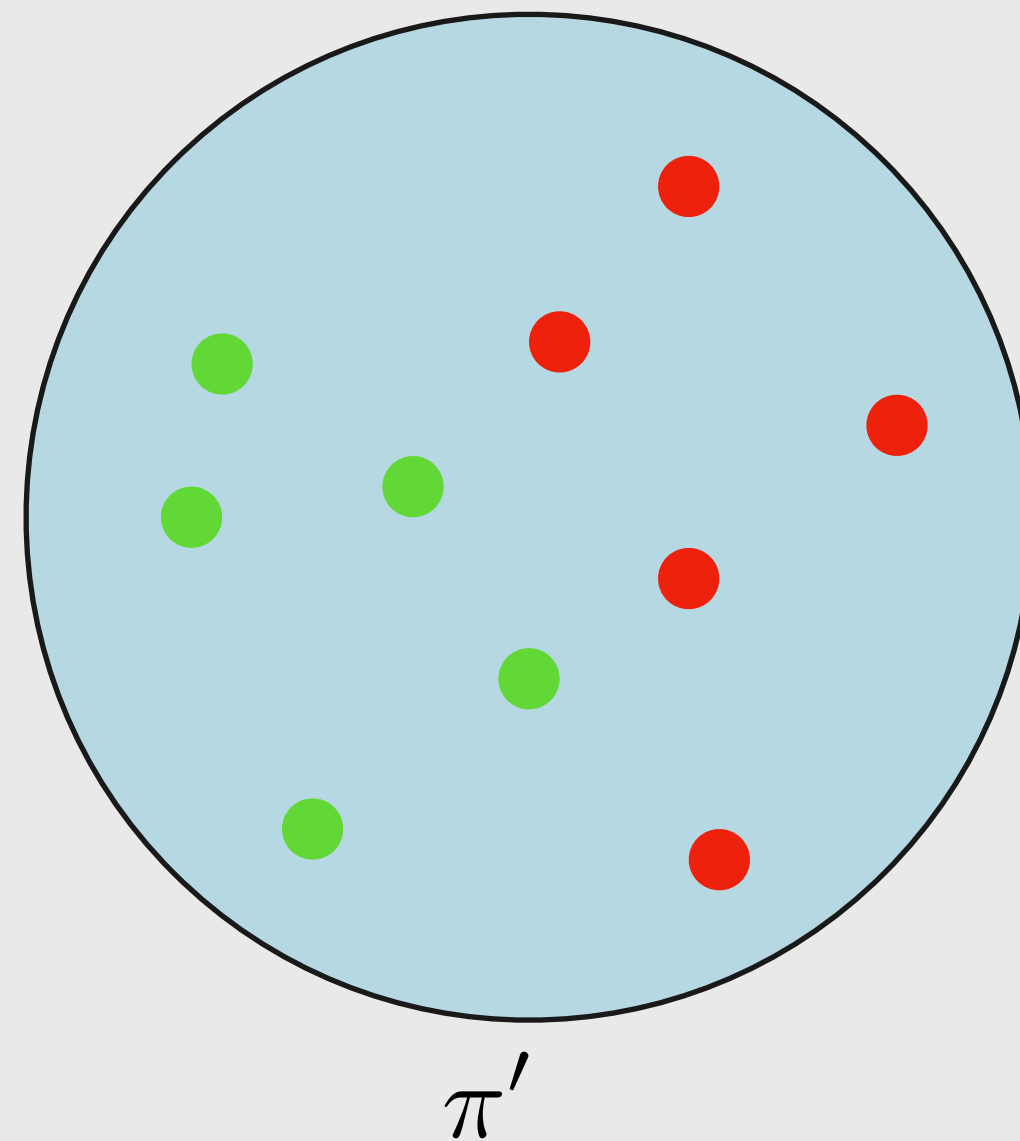
**What about active FDR-control?**

# Where to sample: FDR

$$FDR(\pi) = \frac{|\pi \cap \mathcal{H}_0|}{|\pi|} \quad TPR(\pi) = \frac{|\pi \cap \mathcal{H}_1|}{|\mathcal{H}_1|} \quad \text{Find: } \pi_{\alpha}^* = \operatorname{argmax}_{FDR(\alpha) \leq \alpha} TPR(\alpha)$$

Instead of considering the  $TPR$ , can instead consider

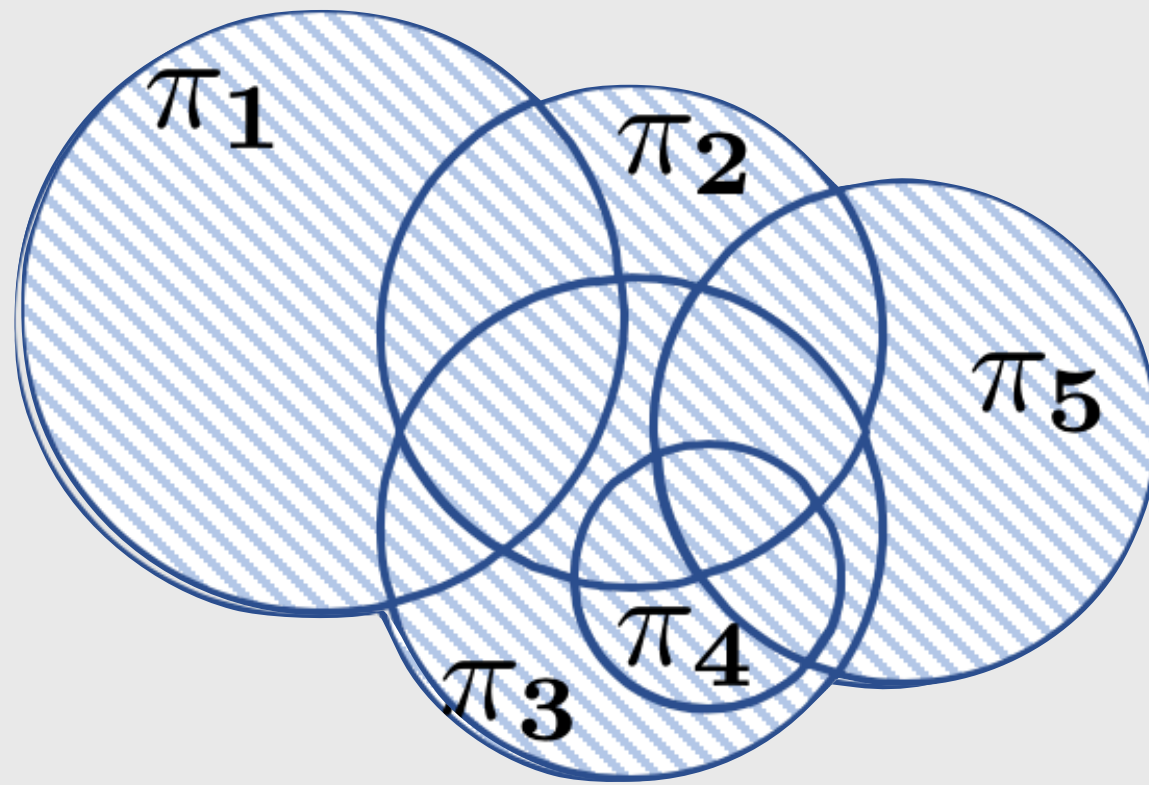
$$TP(\pi) = |\pi \cap \mathcal{H}_1| = \sum_{x \in \pi} y_x$$



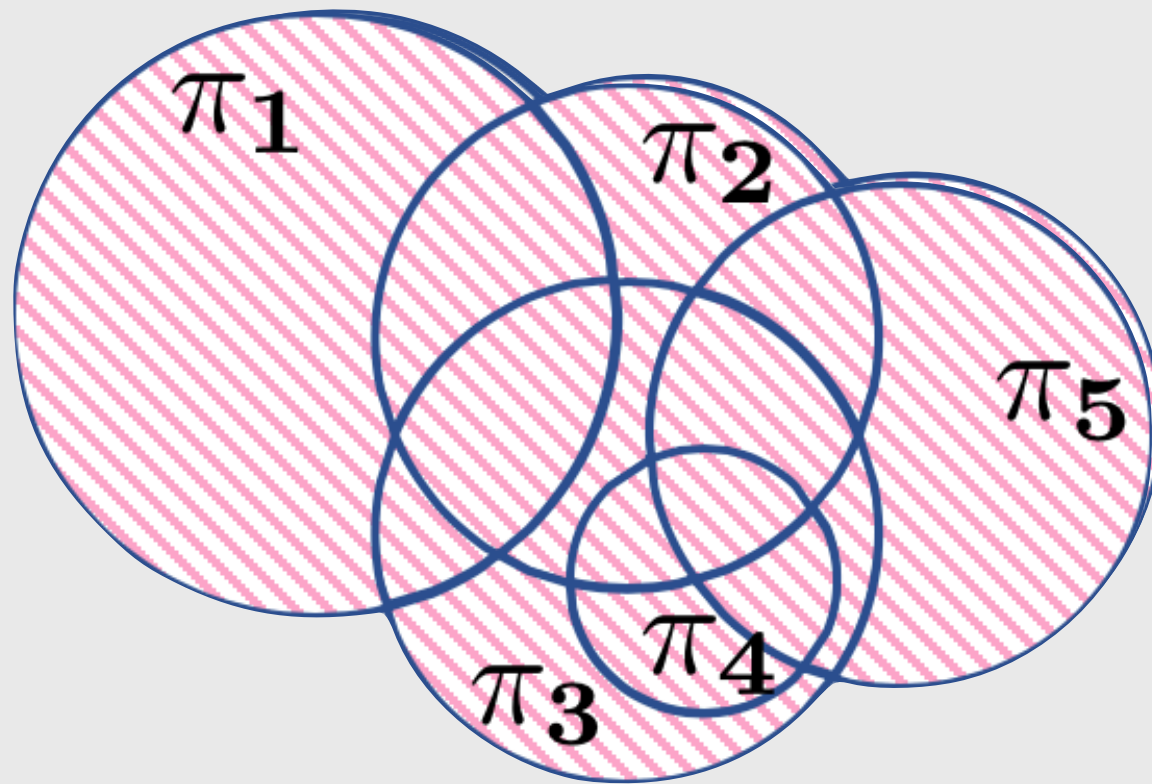


# Action-Elimination: FDR

Union



Symmetric  
Difference



## Action Elimination:

Input:  $\delta, \Pi$

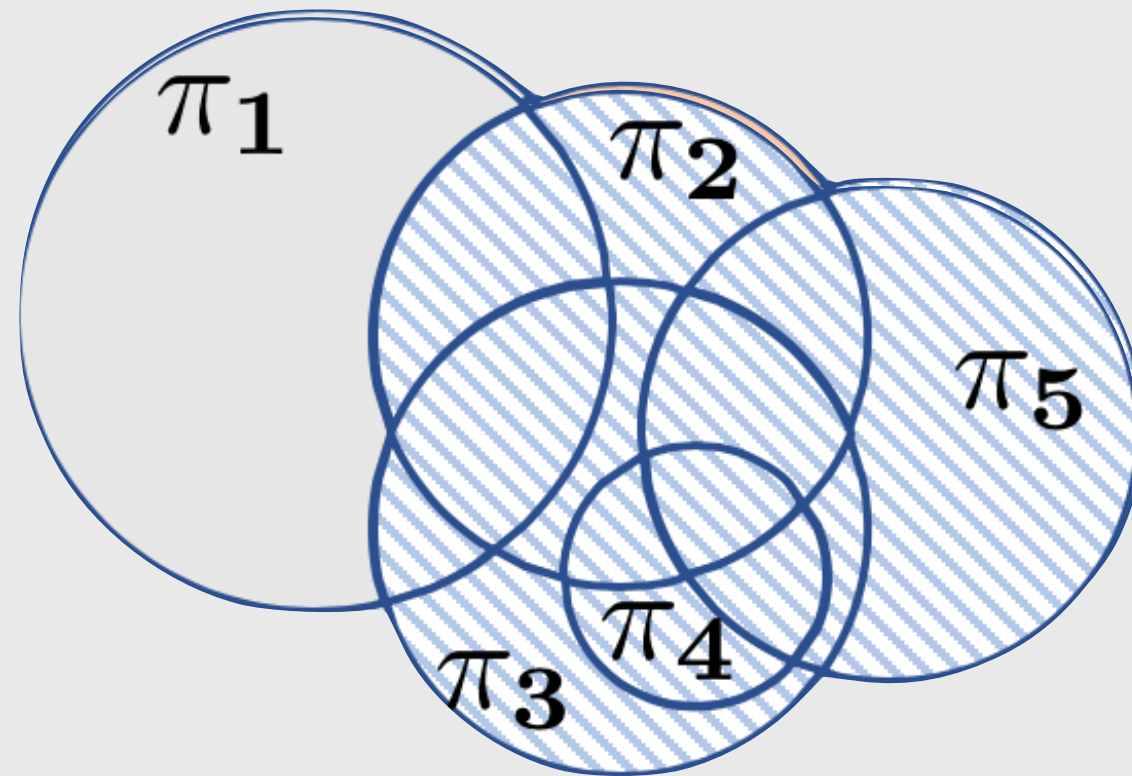
Maintain:  $\mathcal{A}$  active sets,  $\mathcal{C}$  FDR controlled sets

**for:**  $t = 1, 2, \dots$ ,

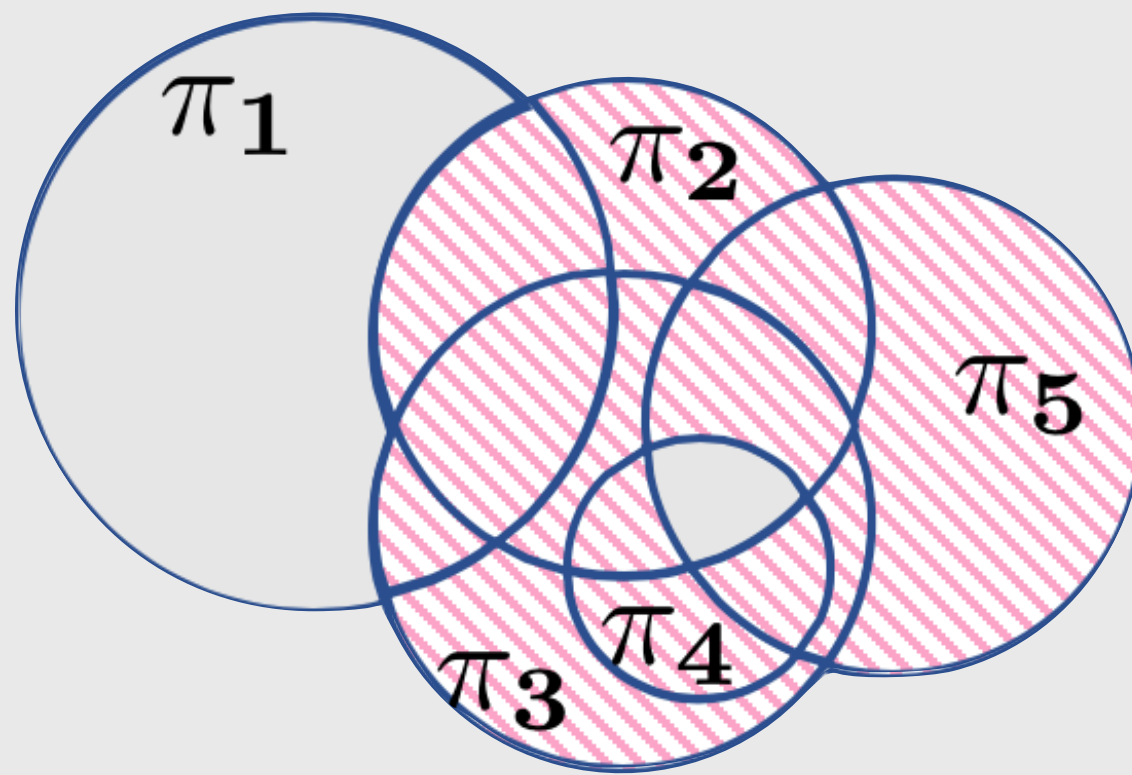
1. **Sample**  $I_t, J_t$  uniformly at random from  $[n]$  (with or without replacement).
2. If  $I_t \in \bigcup_{\pi \in \mathcal{A} \setminus \mathcal{C}} \pi$ , observe  $y_{I_t}$
3. If  $J_t \in \bigcup_{\pi \in \mathcal{A}} \pi - \bigcap_{\pi \in \mathcal{A}} \pi$ , observe  $y_{J_t}$
4. **Update**  $\mathcal{C}$  with any new FDR-controlled sets.
5. **Eliminate**  $\mathcal{A} = \mathcal{A} \setminus \{\pi : \widehat{FDR}(\pi) + C(\pi, t) < \alpha\}$
6. **Eliminate**  $\mathcal{A} = \mathcal{A} \setminus \{\pi : \widehat{TP}(\pi') - \widehat{TP}(\pi) < C(\pi', \pi, t), \pi' \in \mathcal{C}\}$

# Action-Elimination: FDR

Union



Symmetric  
Difference



$\pi_1$  not FDR $\delta$ -controlled

## Action Elimination:

Input:  $\delta, \Pi$

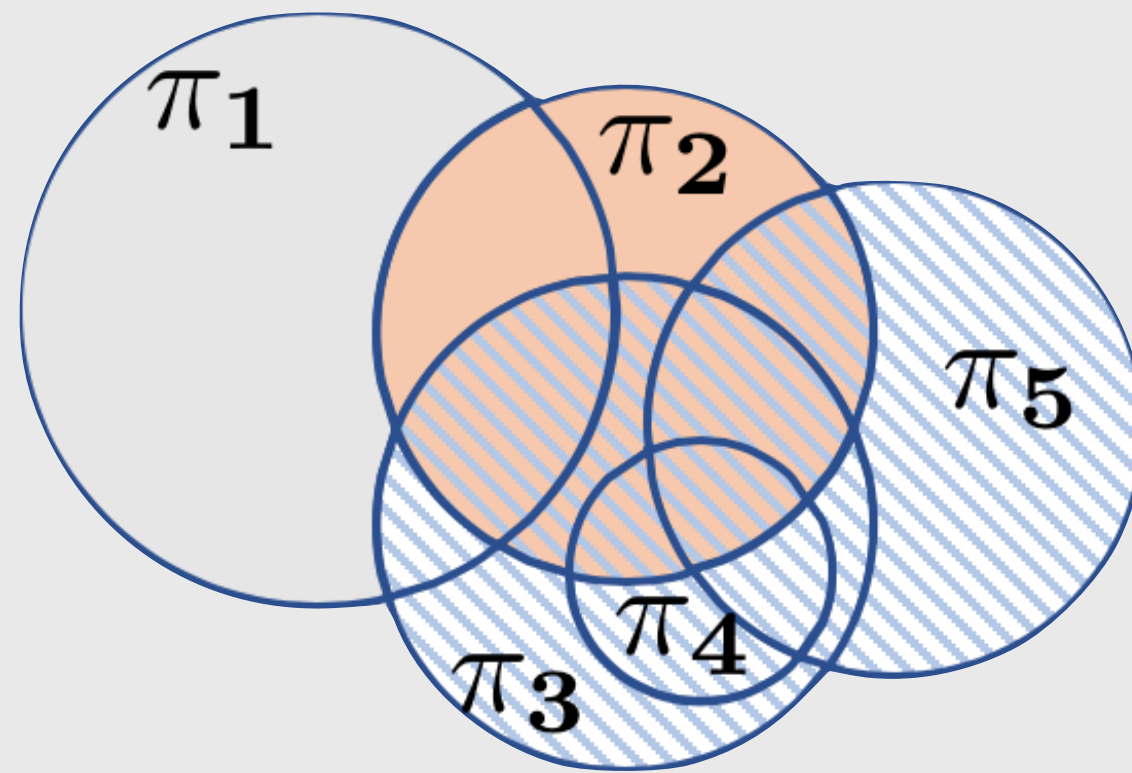
Maintain:  $\mathcal{A}$  active sets,  $\mathcal{C}$  FDR controlled sets

**for:**  $t = 1, 2, \dots$ ,

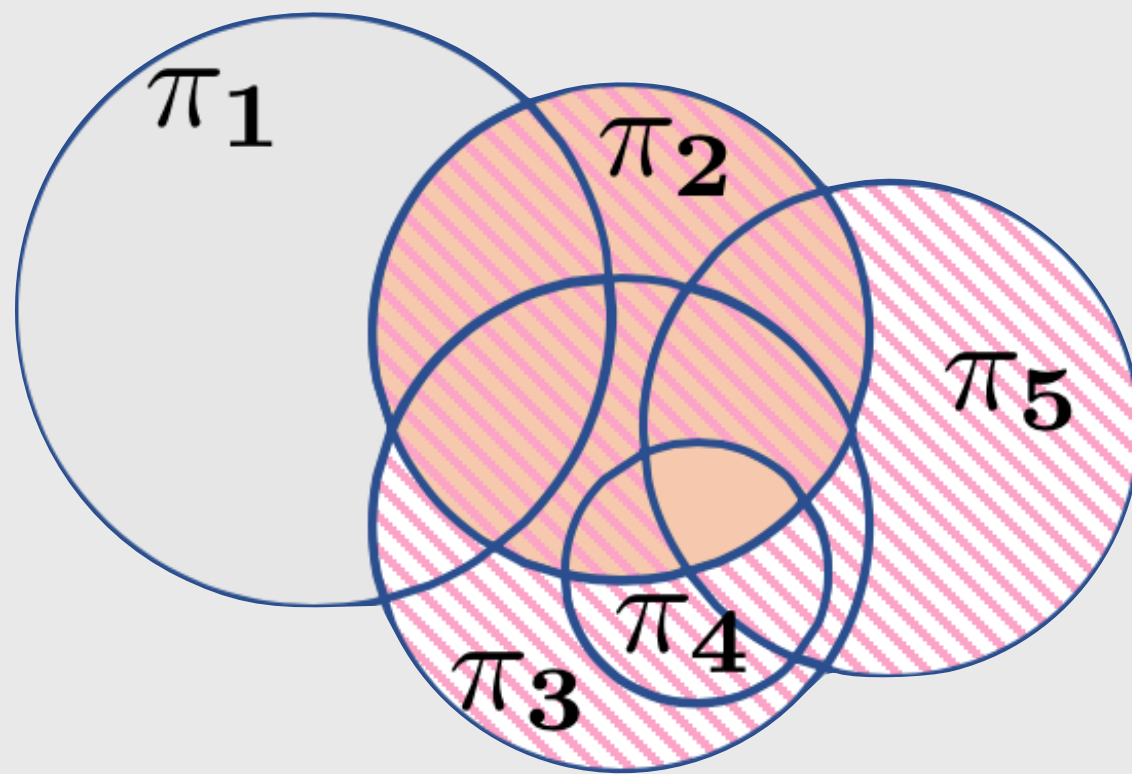
1. **Sample**  $I_t, J_t$  uniformly at random from  $[n]$  (with or without replacement).
2. If  $I_t \in \bigcup_{\pi \in \mathcal{A} \setminus \mathcal{C}} \pi$ , observe  $y_{I_t}$
3. If  $J_t \in \bigcup_{\pi \in \mathcal{A}} \pi - \bigcap_{\pi \in \mathcal{A}} \pi$ , observe  $y_{I_t}$
4. **Update**  $\mathcal{C}$  with any new FDR-controlled sets.
5. **Eliminate**  $\mathcal{A} = \mathcal{A} \setminus \{\pi : \widehat{FDR}(\pi) + C(\pi, t) < \alpha\}$
6. **Eliminate**  $\mathcal{A} = \mathcal{A} \setminus \{\pi : \widehat{TP}(\pi') - \widehat{TP}(\pi) < C(\pi', \pi, t), \pi' \in \mathcal{C}\}$

# Action-Elimination: FDR

Union



Symmetric  
Difference



$\pi_2$  FDR $\delta$ -controlled

## Action Elimination:

Input:  $\delta, \Pi$

Maintain:  $\mathcal{A}$  active sets,  $\mathcal{C}$  FDR controlled sets

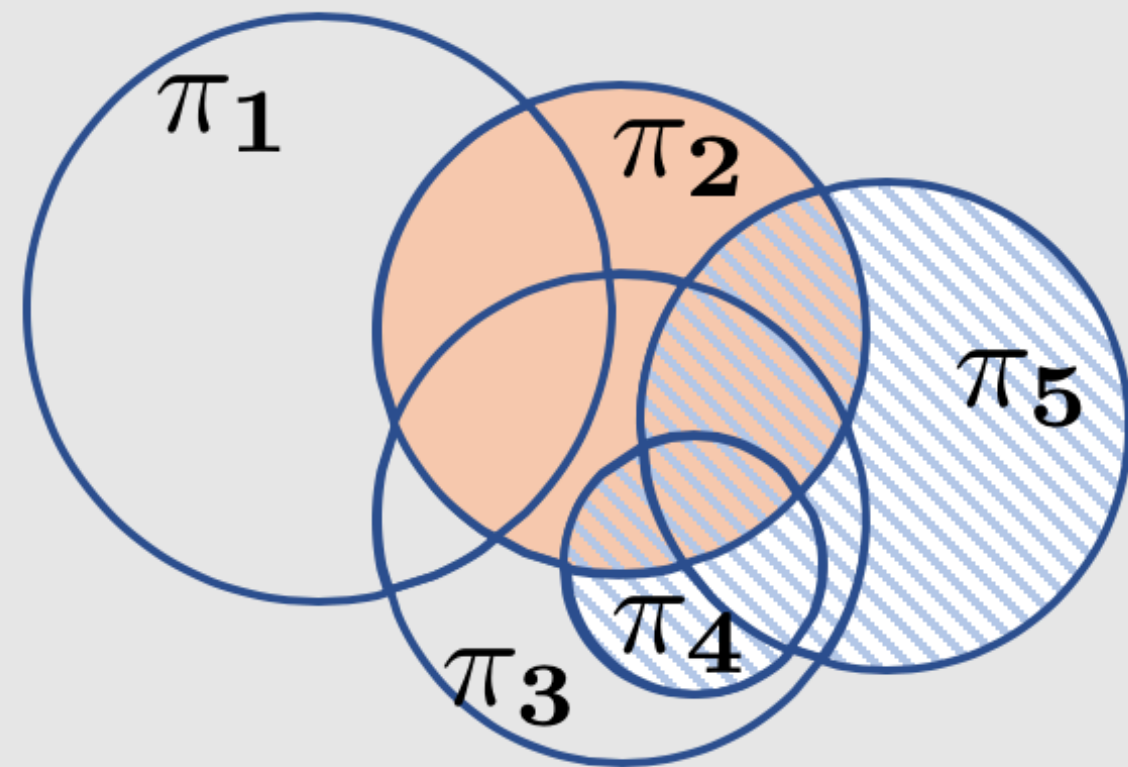
**for:**  $t = 1, 2, \dots$ ,

1. **Sample**  $I_t, J_t$  uniformly at random from  $[n]$  (with or without replacement).
2. If  $I_t \in \bigcup_{\pi \in \mathcal{A} \setminus \mathcal{C}} \pi$ , observe  $y_{I_t}$
3. If  $J_t \in \bigcup_{\pi \in \mathcal{A}} \pi - \bigcap_{\pi \in \mathcal{A}} \pi$ , observe  $y_{J_t}$
4. **Update**  $\mathcal{C}$  with any new FDR-controlled sets.
5. **Eliminate**  $\mathcal{A} = \mathcal{A} \setminus \{\pi : \widehat{FDR}(\pi) + C(\pi, t) < \alpha\}$
6. **Eliminate**  $\mathcal{A} = \mathcal{A} \setminus \{\pi : \widehat{TP}(\pi') - \widehat{TP}(\pi) < C(\pi', \pi, t), \pi' \in \mathcal{C}\}$

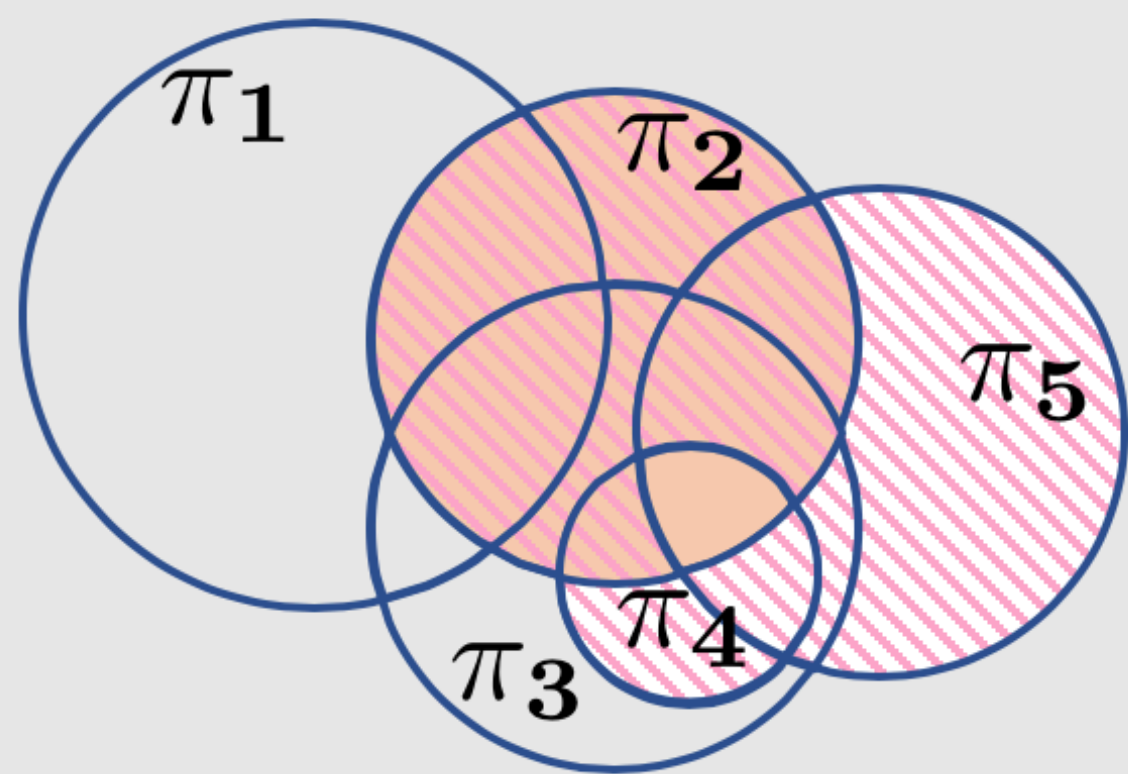


# Action-Elimination: FDR

Union



Symmetric  
Difference



$\pi_2$  eliminates  $\pi_3$

## Action Elimination:

Input:  $\delta, \Pi$

Maintain:  $\mathcal{A}$  active sets,  $\mathcal{C}$  FDR controlled sets

**for:**  $t = 1, 2, \dots$ ,

1. **Sample**  $I_t, J_t$  uniformly at random from  $[n]$  (with or without replacement).

2. If  $I_t \in \bigcup_{\pi \in \mathcal{A} \setminus \mathcal{C}} \pi$ , observe  $y_{I_t}$

3. If  $J_t \in \bigcup_{\pi \in \mathcal{A}} \pi - \bigcap_{\pi \in \mathcal{A}} \pi$ , observe  $y_{J_t}$

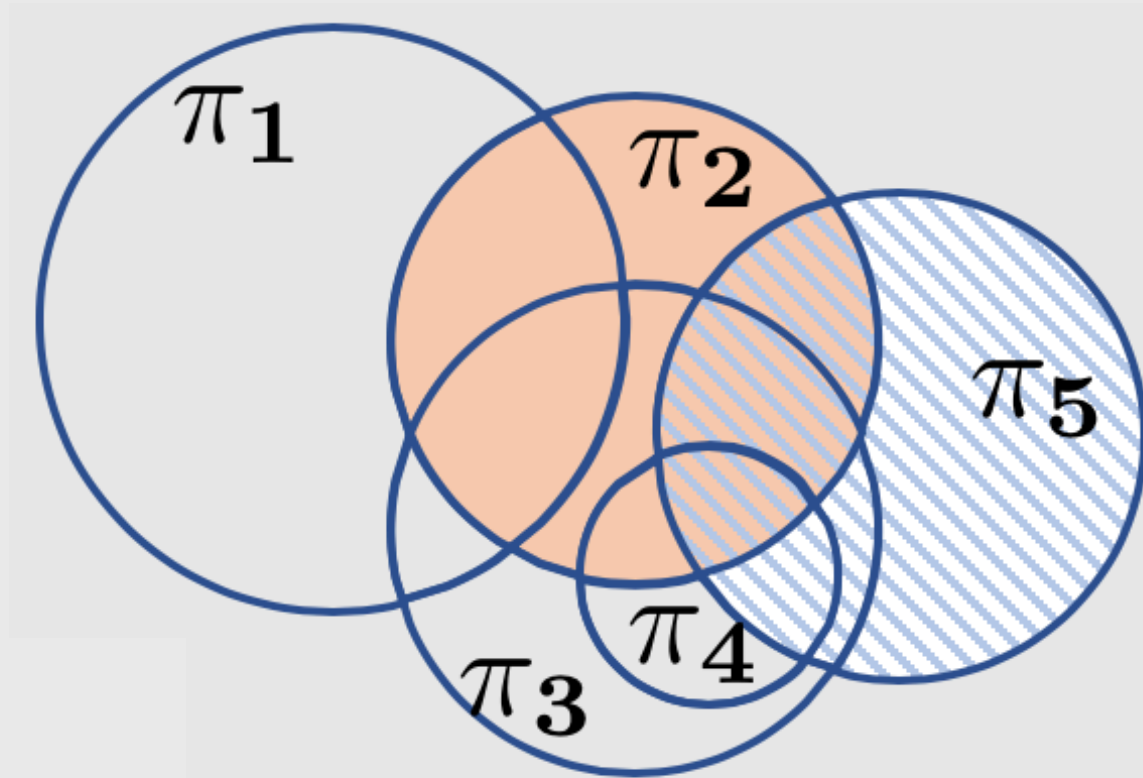
4. **Update**  $\mathcal{C}$  with any new FDR-controlled sets.

5. **Eliminate**  $\mathcal{A} = \mathcal{A} \setminus \{\pi : \widehat{FDR}(\pi) + C(\pi, t) < \alpha\}$

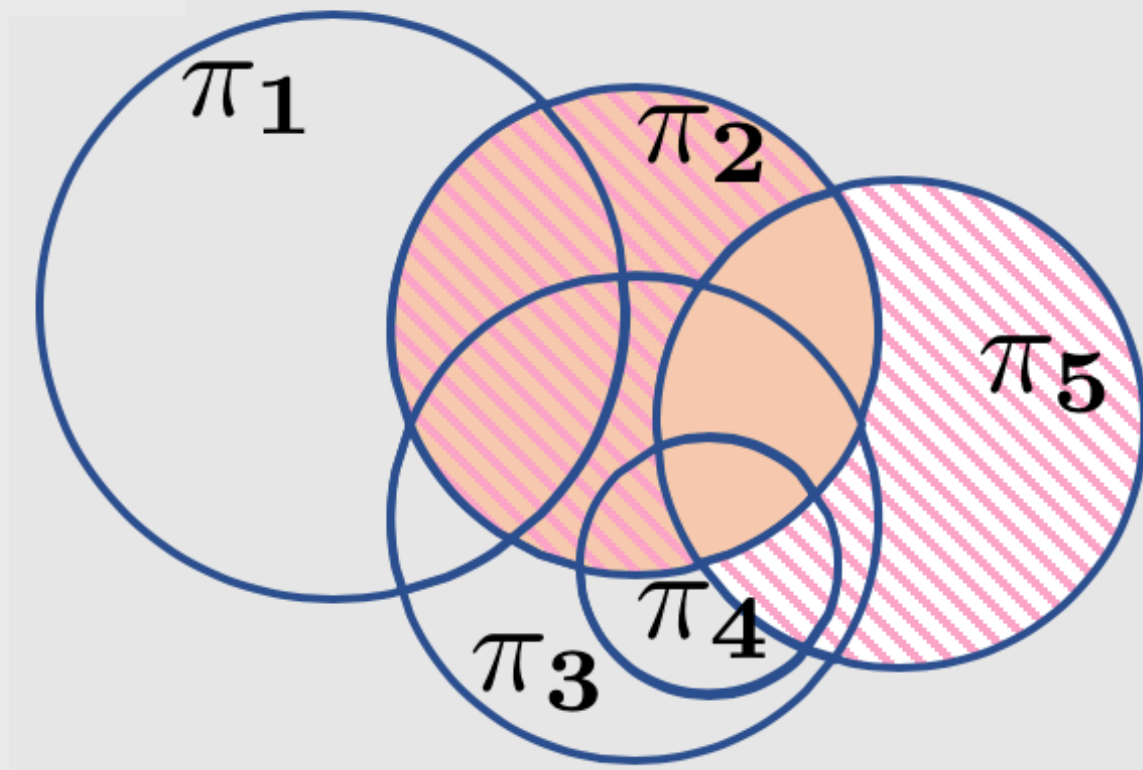
6. **Eliminate**  $\mathcal{A} = \mathcal{A} \setminus \{\pi : \widehat{TP}(\pi') - \widehat{TP}(\pi) < C(\pi', \pi, t), \pi' \in \mathcal{C}\}$

# Action-Elimination: FDR

Union



Symmetric  
Difference



$\pi_2$  eliminates  $\pi_4$  since  $\pi_4 \subset \pi_3$

## Action Elimination:

Input:  $\delta, \Pi$

Maintain:  $\mathcal{A}$  active sets,  $\mathcal{C}$  FDR controlled sets

**for:**  $t = 1, 2, \dots$ ,

1. **Sample**  $I_t, J_t$  uniformly at random from  $[n]$  (with or without replacement).

2. If  $I_t \in \bigcup_{\pi \in \mathcal{A} \setminus \mathcal{C}} \pi$ , observe  $y_{I_t}$

3. If  $J_t \in \bigcup_{\pi \in \mathcal{A}} \pi - \bigcap_{\pi \in \mathcal{A}} \pi$ , observe  $y_{I_t}$

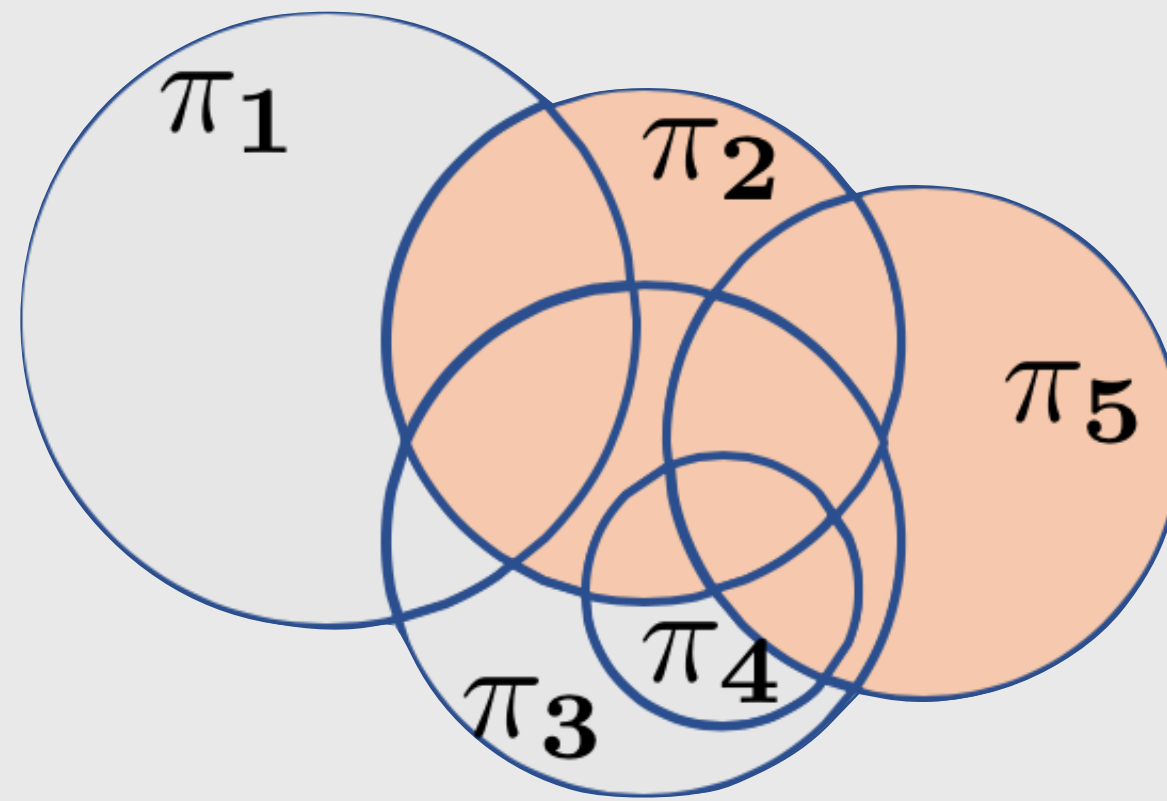
4. **Update**  $\mathcal{C}$  with any new FDR-controlled sets.

5. **Eliminate**  $\mathcal{A} = \mathcal{A} \setminus \{\pi : \widehat{FDR}(\pi) + C(\pi, t) < \alpha\}$

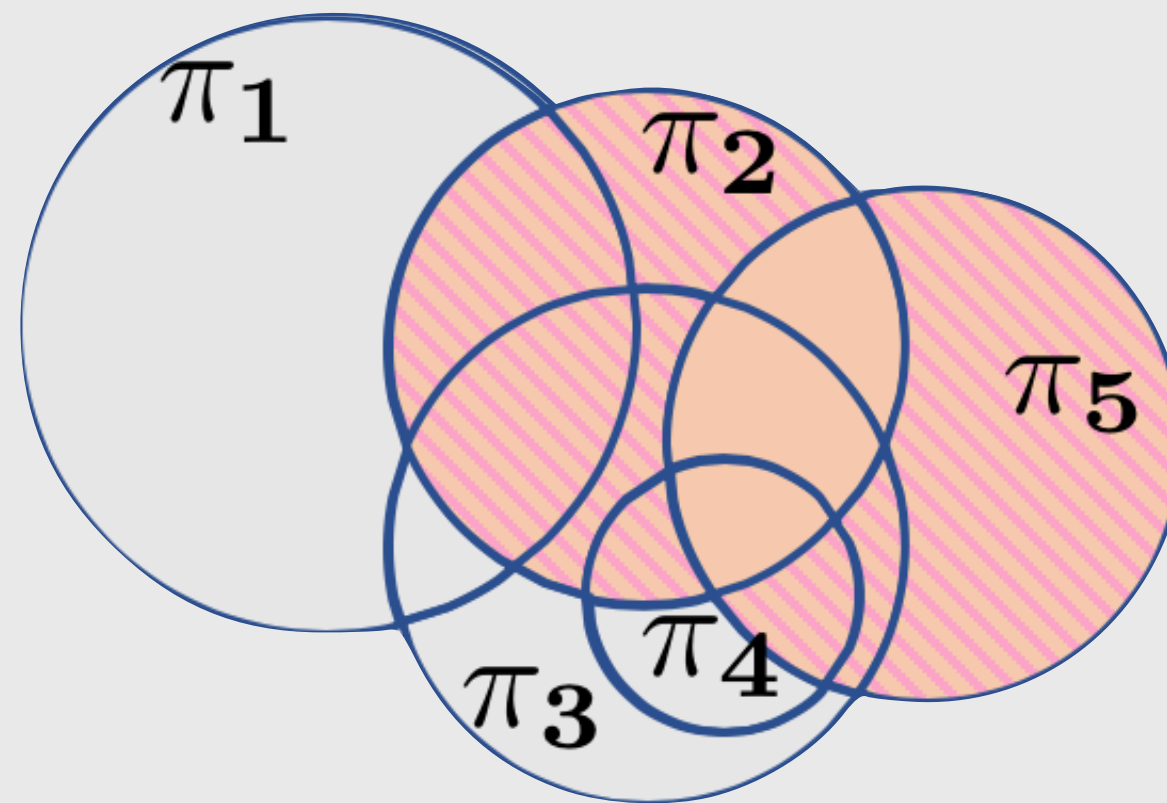
6. **Eliminate**  $\mathcal{A} = \mathcal{A} \setminus \{\pi : \widehat{TP}(\pi') - \widehat{TP}(\pi) < C(\pi', \pi, t), \pi' \in \mathcal{C}\}$

# Action-Elimination: FDR

Union



Symmetric  
Difference



## Action Elimination:

Input:  $\delta, \Pi$

Maintain:  $\mathcal{A}$  active sets,  $\mathcal{C}$  FDR controlled sets

**for:**  $t = 1, 2, \dots$ ,

1. **Sample**  $I_t, J_t$  uniformly at random from  $[n]$  (with or without replacement).
2. If  $I_t \in \bigcup_{\pi \in \mathcal{A} \setminus \mathcal{C}} \pi$ , observe  $y_{I_t}$
3. If  $J_t \in \bigcup_{\pi \in \mathcal{A}} \pi - \bigcap_{\pi \in \mathcal{A}} \pi$ , observe  $y_{I_t}$
4. **Update**  $\mathcal{C}$  with any new FDR-controlled sets.
5. **Eliminate**  $\mathcal{A} = \mathcal{A} \setminus \{\pi : \widehat{FDR}(\pi) + C(\pi, t) < \alpha\}$
6. **Eliminate**  $\mathcal{A} = \mathcal{A} \setminus \{\pi : \widehat{TP}(\pi') - \widehat{TP}(\pi) < C(\pi', \pi, t), \pi' \in \mathcal{C}\}$

# Guarantees

**How long does it take to knock out a set?** Define

$$\tilde{\Delta}_{\pi} = \frac{|TP(\pi_{\alpha}^*) - TP(\pi)|}{|\pi_{\alpha}^* \Delta \pi|}, \quad \Delta_{\pi, \alpha} = |FDR(\pi) - \alpha|,$$

and

$$s_{\pi}^{FDR} = \frac{VC(B(|\pi|)) \log(n \log(\tilde{\Delta}_{\pi, \alpha}^{-2})/\delta)}{|\pi| \tilde{\Delta}_{\pi, \alpha}^2}, \quad s_{\pi}^{TP} = \frac{VC(B(|\pi_{\alpha}^* \Delta \pi|, \pi_{\alpha}^*)) \log(n \log(\tilde{\Delta}_{\pi}^{-2})/\delta)}{|\pi_{\alpha}^* \Delta \pi| \tilde{\Delta}_{\pi}^2}$$

If  $\pi$  is not FDR-controlled:

$$\min\{s_{\pi}^{FDR}, \max\{s_{\pi^*}^{FDR}, s_{\pi}^{TPR}\}\}$$

If  $\pi$  is FDR-controlled:

$$\max\{s_{\pi^*}^{FDR}, s_{\pi}^{TPR}\}$$



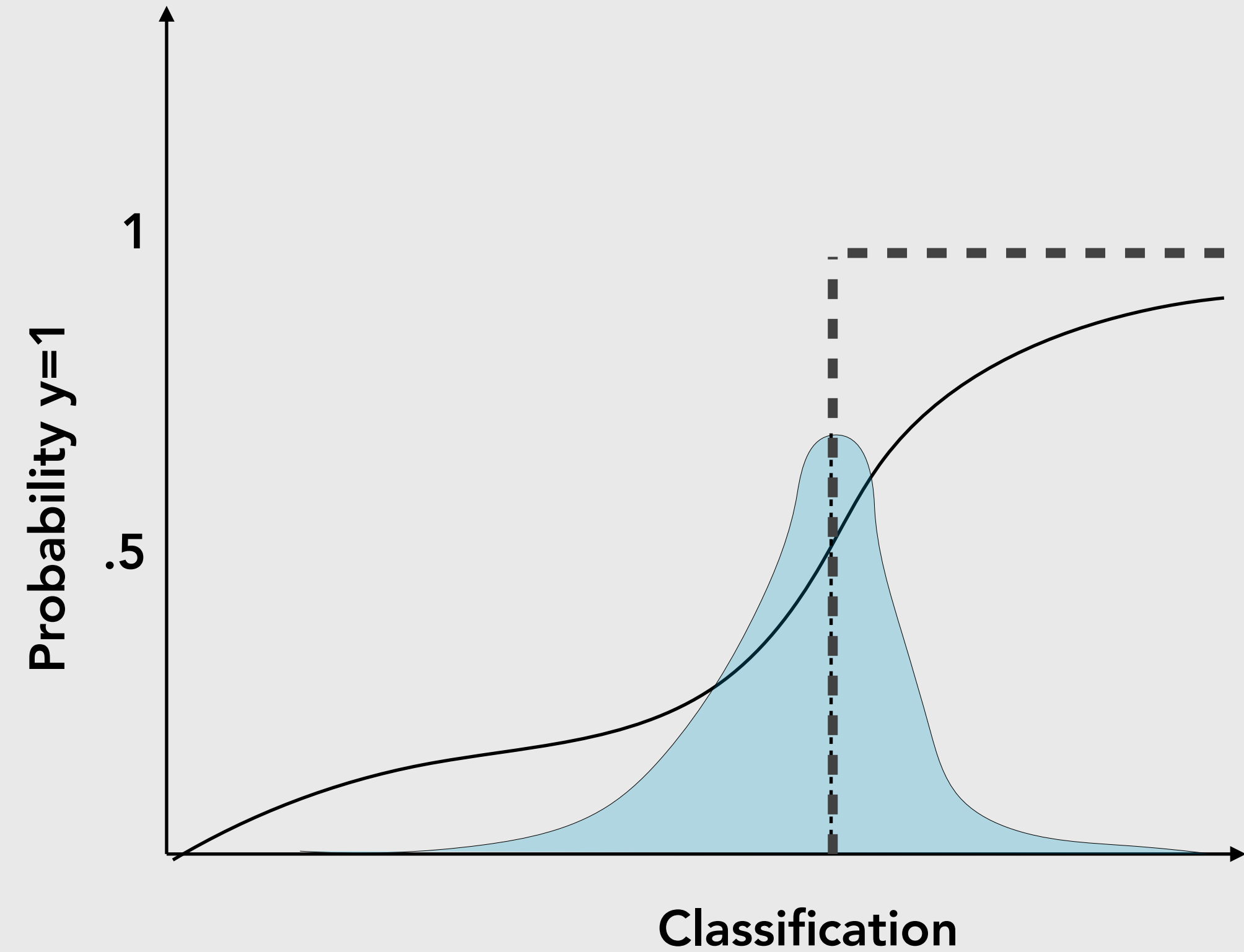
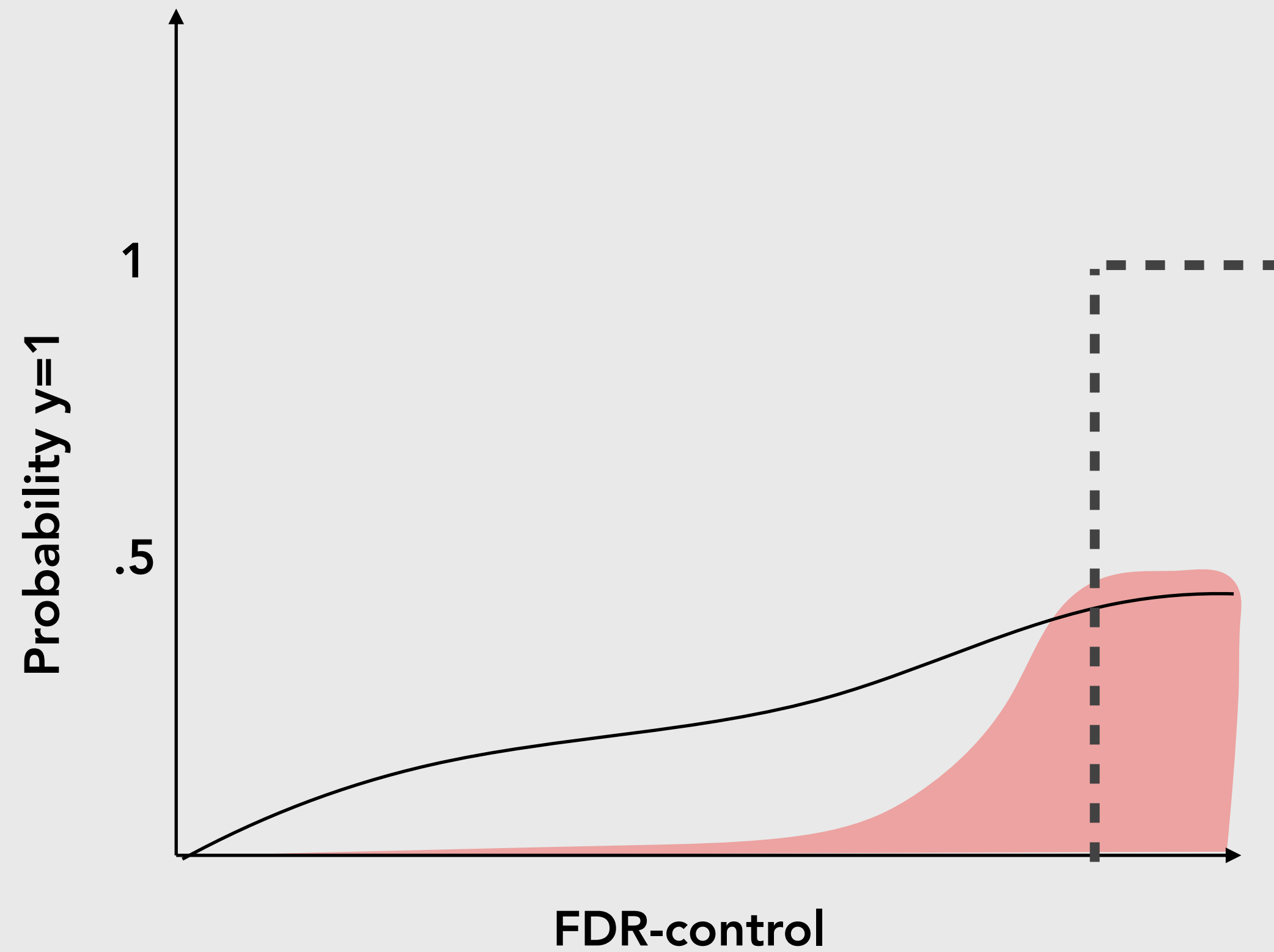
# Guarantees

**Theorem (J., Jamieson 2019)** Let With probability greater than  $1 - \delta$  the FDR Action Elimination algorithm terminates after a number of samples no more than

$$\sum_{1 \leq i \leq n} \max_{\pi: i \in \pi} \min\{s_{\pi}^{FDR}, \max\{s_{\pi^*}^{FDR}, s_{\pi}^{TPR}\}\} + \max_{\pi: i \in \pi_{\alpha}^* \Delta \pi} \max\{s_{\pi^*}^{FDR}, s_{\pi}^{TPR}\}$$

The sample complexity is at the time it takes to verify  $FDR(\pi^*) \leq \alpha$

# Contrasting Sampling Strategies



**Moral:** If you want FDR-control, it might not make sense to sample using active classification!