Isotonic subgroup selection

Manuel M. Müller

Statistical Laboratory, University of Cambridge

23rd July 2024

Novartis Stats Methods Seminar

Collaborators

Henry W. J. Reeve

University of Bristol

Timothy I. Cannings

University of Edinburgh

Richard J. Samworth

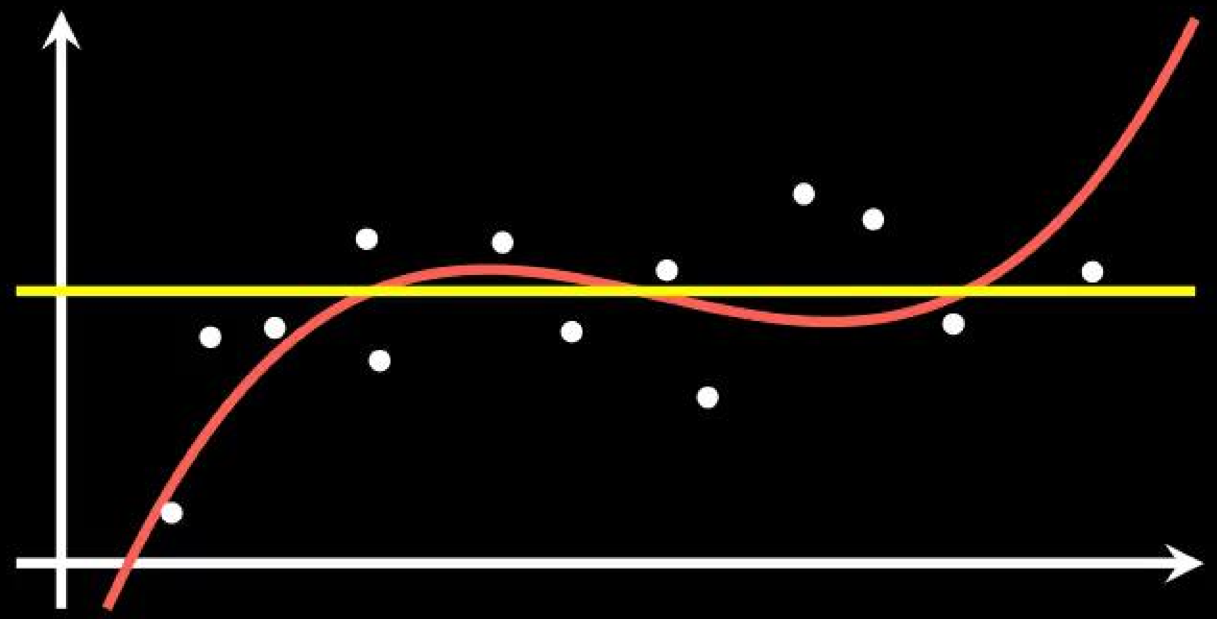
University of Cambridge

Subgroup selection

Task. *Subgroup selection* in regression: identify a subset of the covariate domain on which the regression function satisfies a particular property (e.g. exceeds a threshold).

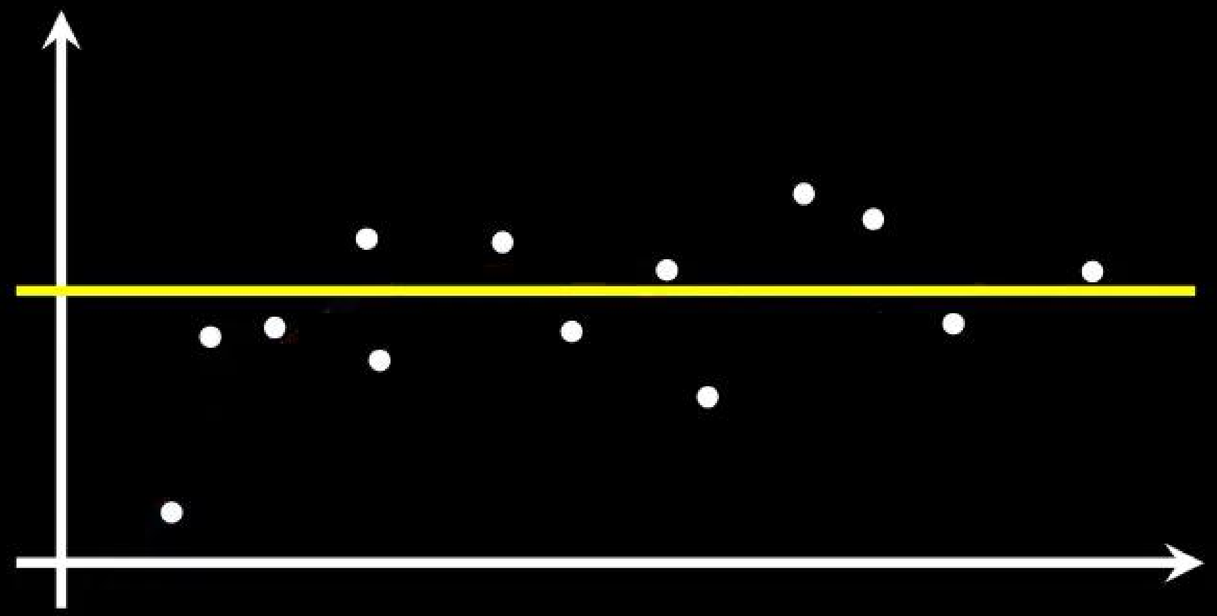
Subgroup selection

Task. *Subgroup selection* in regression: identify a subset of the covariate domain on which the regression function satisfies a particular property (e.g. exceeds a threshold).



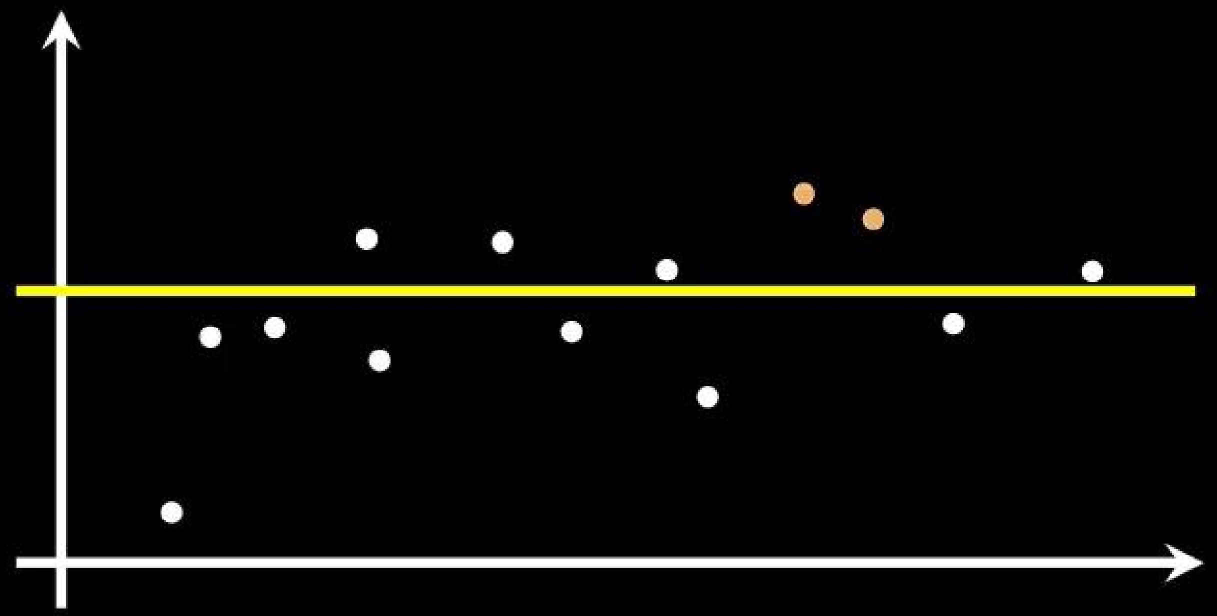
Subgroup selection

Task. *Subgroup selection* in regression: identify a subset of the covariate domain on which the regression function satisfies a particular property (e.g. exceeds a threshold).



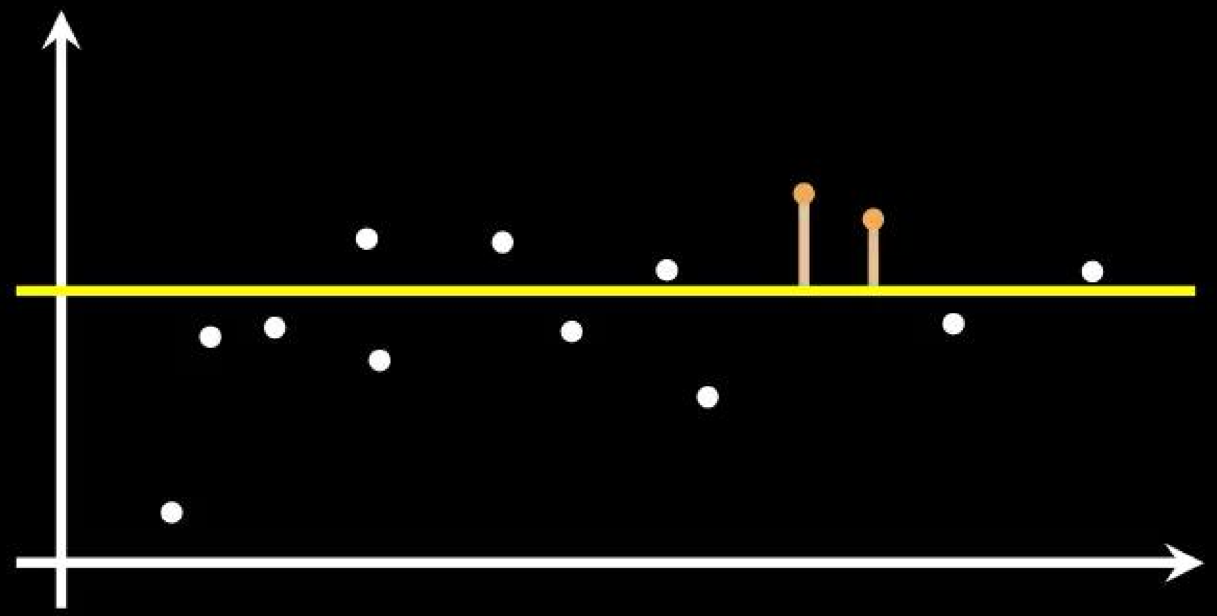
Subgroup selection

Task. *Subgroup selection* in regression: identify a subset of the covariate domain on which the regression function satisfies a particular property (e.g. exceeds a threshold).



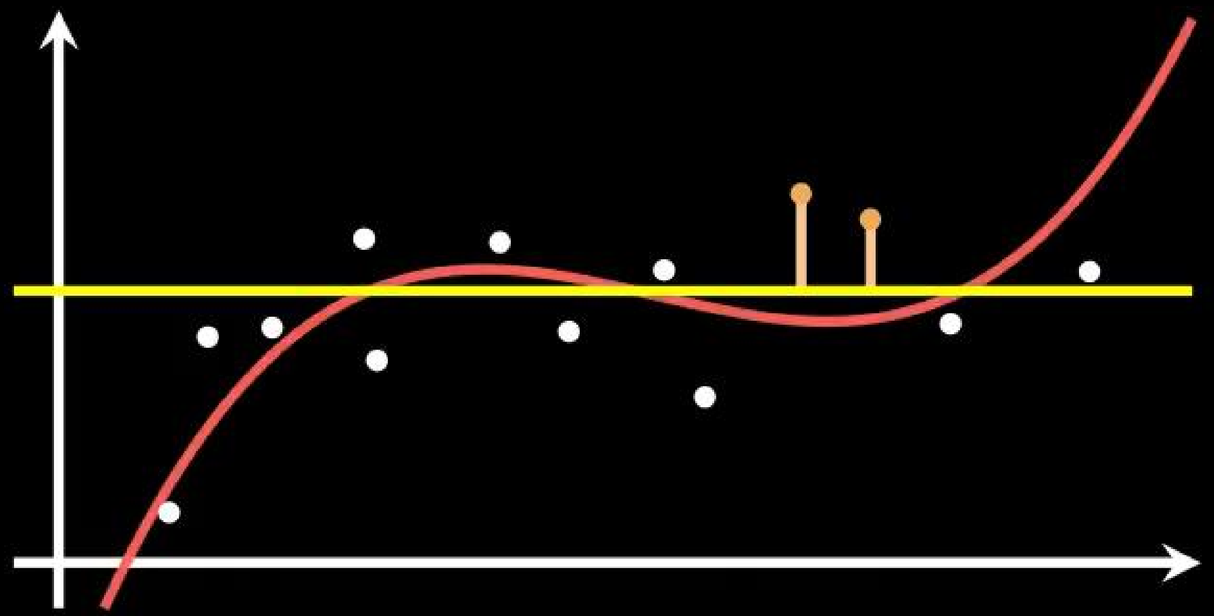
Subgroup selection

Task. *Subgroup selection* in regression: identify a subset of the covariate domain on which the regression function satisfies a particular property (e.g. exceeds a threshold).



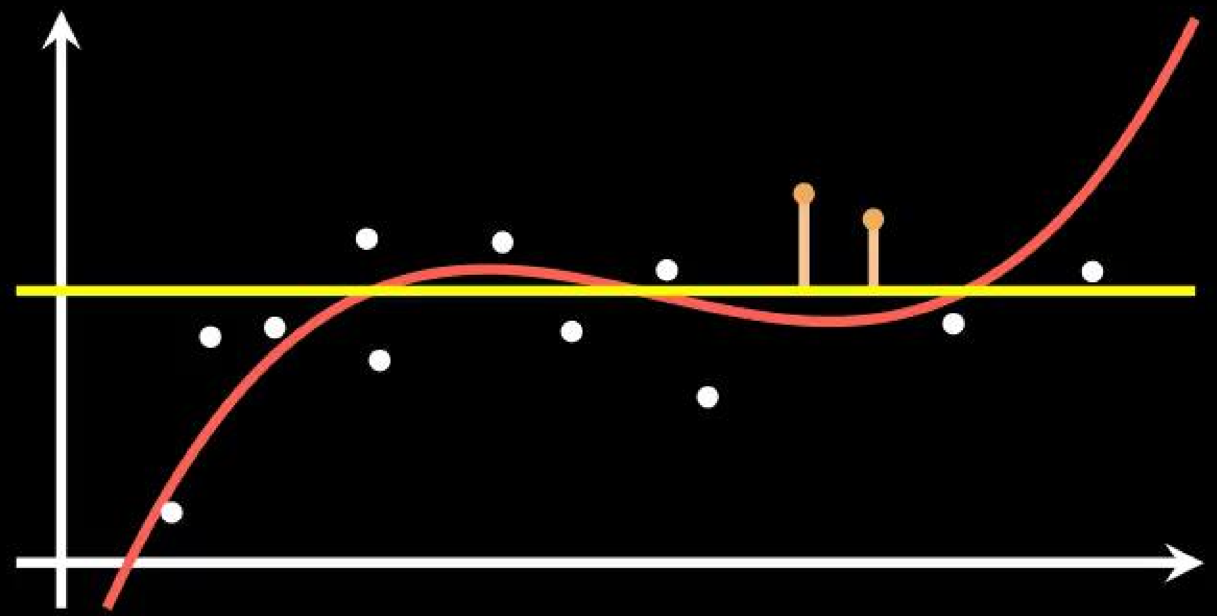
Subgroup selection

Task. *Subgroup selection* in regression: identify a subset of the covariate domain on which the regression function satisfies a particular property (e.g. exceeds a threshold).



Subgroup selection

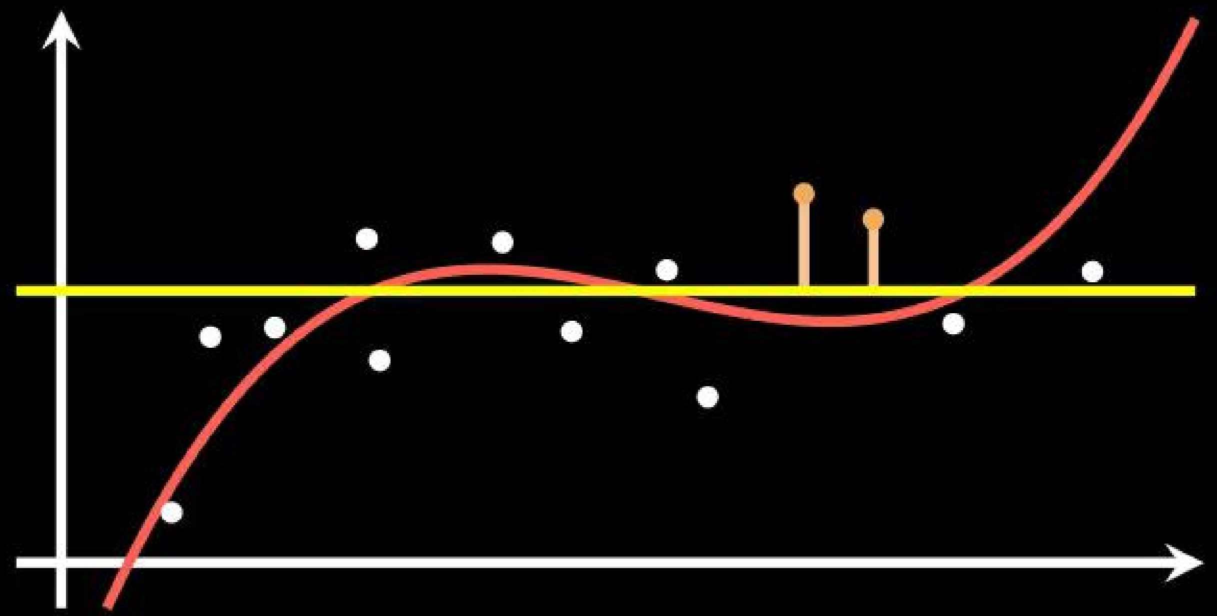
Task. *Subgroup selection* in regression: identify a subset of the covariate domain on which the regression function satisfies a particular property (e.g. exceeds a threshold).



Example. Consider patients with HIV, who haven't developed AIDS.

Subgroup selection

Task. *Subgroup selection* in regression: identify a subset of the covariate domain on which the regression function satisfies a particular property (e.g. exceeds a threshold).

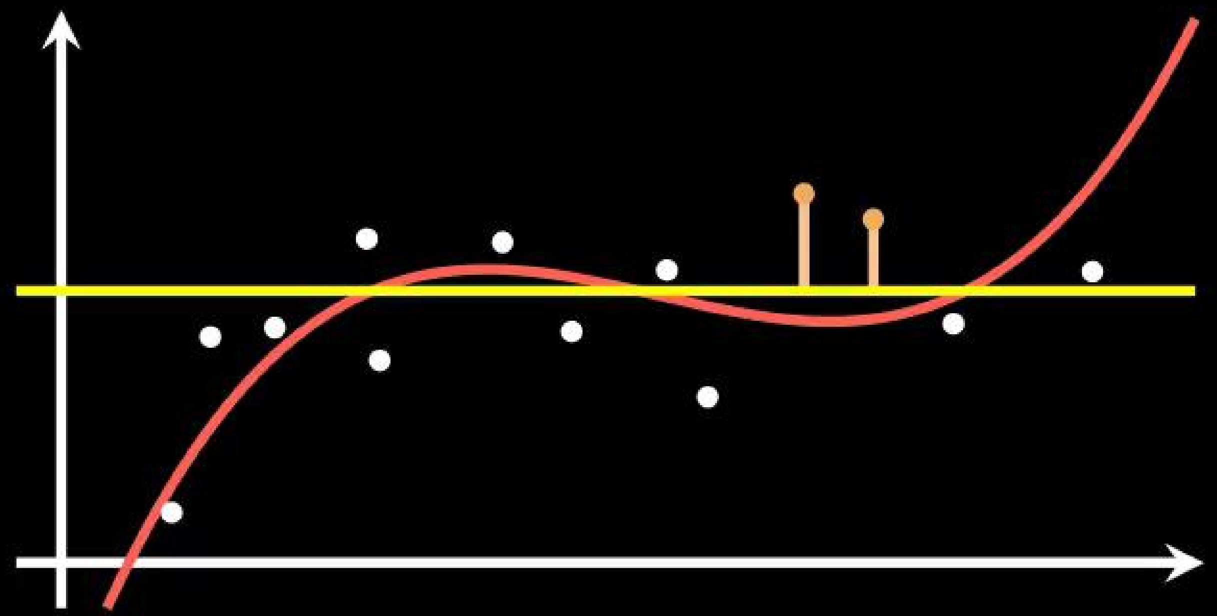


Example. Consider patients with HIV, who haven't developed AIDS.

Goal: identify those patients that are likely not to experience a negative event (sharp decrease in CD4 cell count, development of AIDS, or death) soon.

Subgroup selection

Task. *Subgroup selection* in regression: identify a subset of the covariate domain on which the regression function satisfies a particular property (e.g. exceeds a threshold).



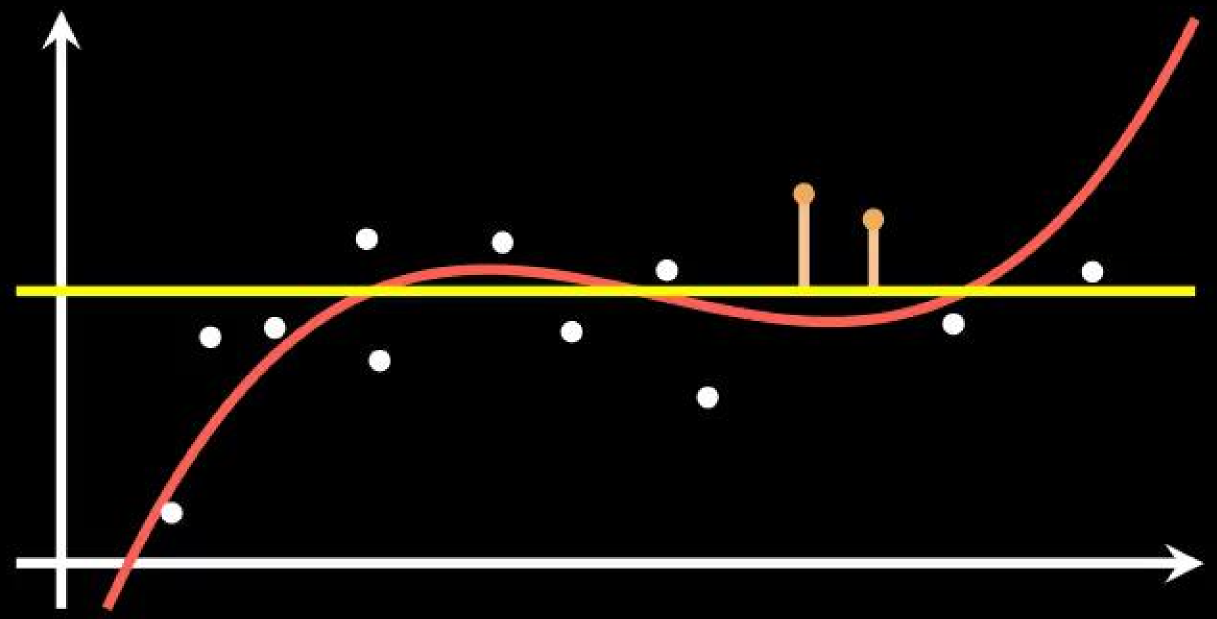
Example. Consider patients with HIV, who haven't developed AIDS.

Goal: identify those patients that are likely not to experience a negative event (sharp decrease in CD4 cell count, development of AIDS, or death) soon.

I.e. we would like to identify the subset of patients at low risk.

Subgroup selection

Task. *Subgroup selection* in regression: identify a subset of the covariate domain on which the regression function satisfies a particular property (e.g. exceeds a threshold).



Example. Consider patients with HIV, who haven't developed AIDS.

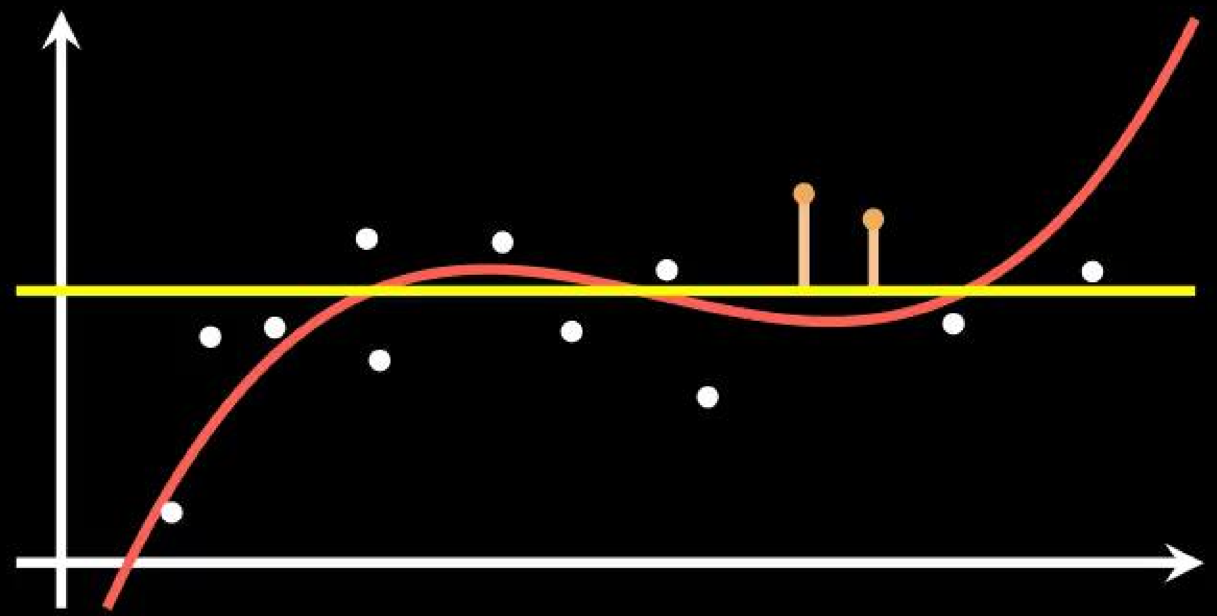
Goal: identify those patients that are likely <u>not</u> to experience a negative event (sharp decrease in CD4 cell count, development of AIDS, or death) soon.

I.e. we would like to identify the subset of patients at low risk.

But: erroneously identifying someone as having low risk must be avoided.

Subgroup selection

Task. *Subgroup selection* in regression: identify a subset of the covariate domain on which the regression function satisfies a particular property (e.g. exceeds a threshold).



Example. Consider patients with HIV, who haven't developed AIDS.

Goal: identify those patients that are likely not to experience a negative event (sharp decrease in CD4 cell count, development of AIDS, or death) soon.

I.e. we would like to identify the subset of patients at low risk.

But: erroneously identifying someone as having low risk must be avoided.

$\longrightarrow$ Asymmetry of errors

Statistical setting

Setting:

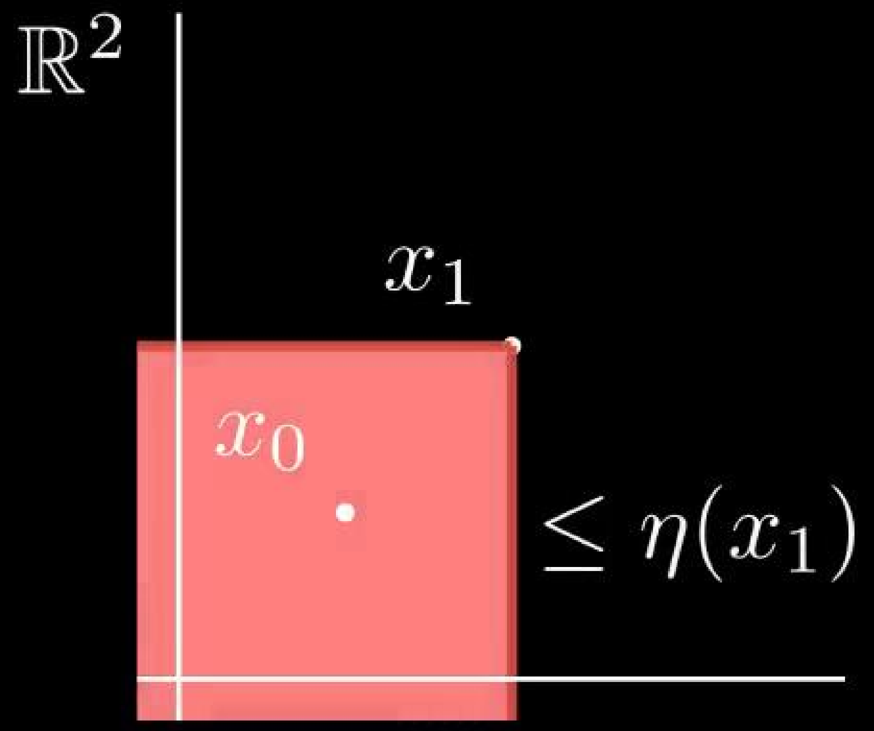
Fix $\sigma > 0$. Let $\mathcal{P}_{\mathrm{Mon},d}(\sigma)$ be the family of distributions P on $\mathbb{R}^d \times \mathbb{R}$ such that for $(X, Y) \sim P$,

Statistical setting

Setting:

Fix $\sigma > 0$. Let $\mathcal{P}_{\mathrm{Mon},d}(\sigma)$ be the family of distributions P on $\mathbb{R}^d \times \mathbb{R}$ such that for $(X, Y) \sim P$,

(i) the regression function $\eta(x) := \mathbb{E}(Y|X = x)$ is increasing on $\mathbb{R}^d$, i.e. $x_0 \preccurlyeq x_1 \implies \eta(x_0) \leq \eta(x_1)$,

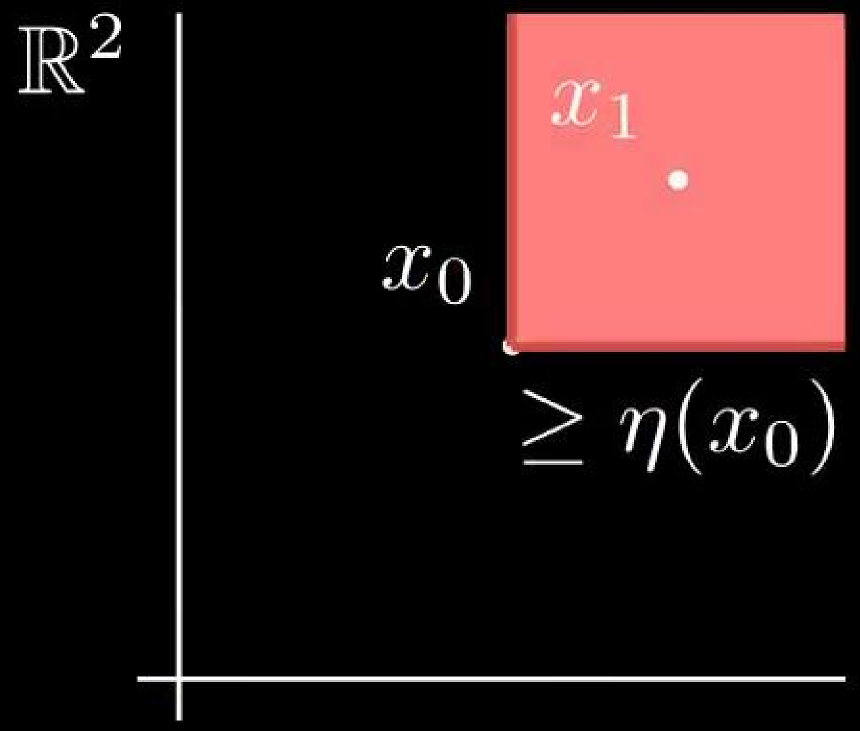


Statistical setting

Setting:

Fix $\sigma > 0$. Let $\mathcal{P}_{\mathrm{Mon},d}(\sigma)$ be the family of distributions P on $\mathbb{R}^d \times \mathbb{R}$ such that for $(X, Y) \sim P$,

(i) the regression function $\eta(x) := \mathbb{E}(Y|X = x)$ is increasing on $\mathbb{R}^d$, i.e. $x_0 \preccurlyeq x_1 \implies \eta(x_0) \leq \eta(x_1)$,



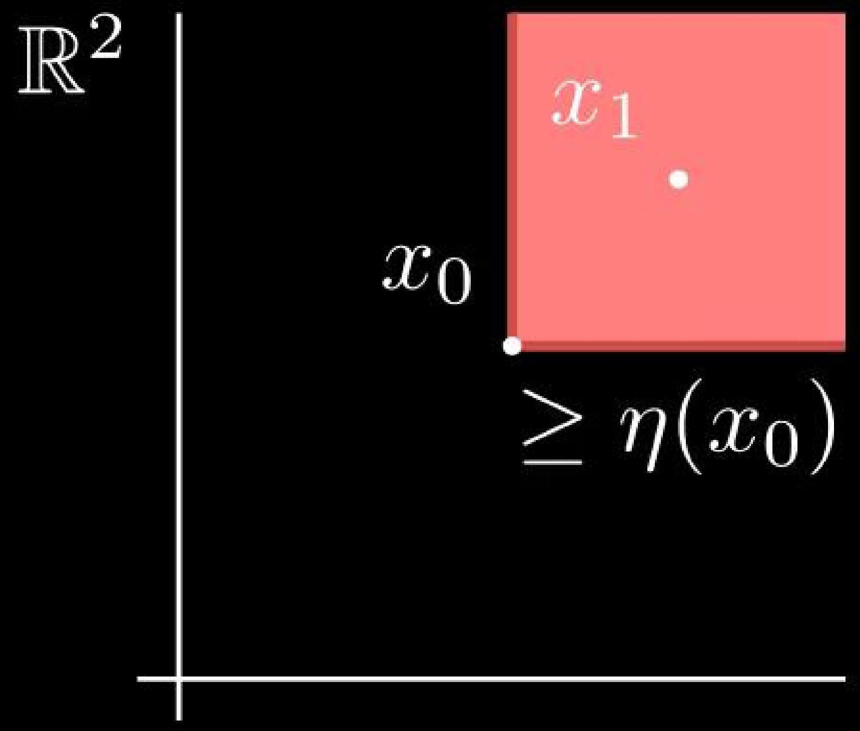
Statistical setting

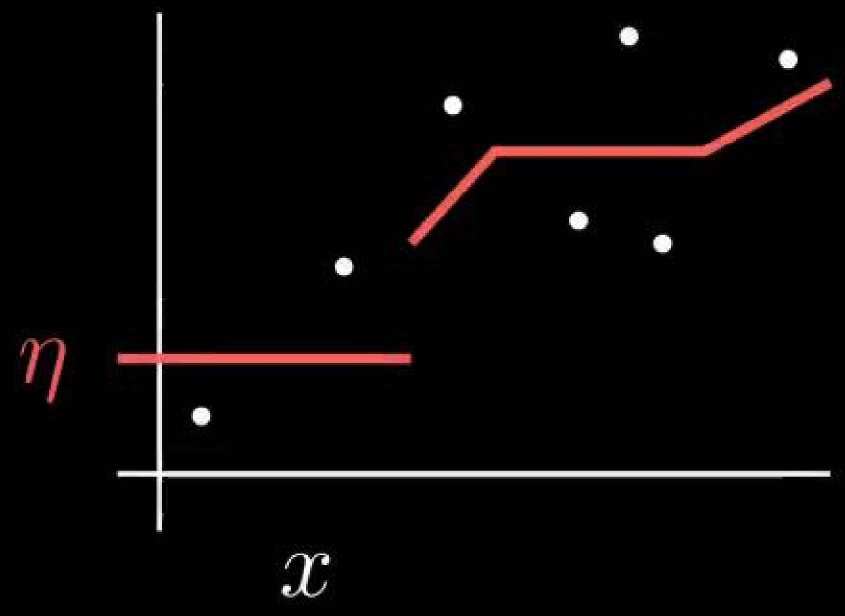
Setting:

Fix $\sigma > 0$. Let $\mathcal{P}_{\mathrm{Mon},d}(\sigma)$ be the family of distributions P on $\mathbb{R}^d \times \mathbb{R}$ such that for $(X, Y) \sim P$,

(i) the regression function $\eta(x) := \mathbb{E}(Y|X = x)$ is increasing on $\mathbb{R}^d$, i.e. $x_0 \preccurlyeq x_1 \implies \eta(x_0) \leq \eta(x_1)$,

(ii) $Y - \eta(X) \mid X = x$ is sub-Gaussian with variance parameter σ^2 for $x \in \mathbb{R}^d$.





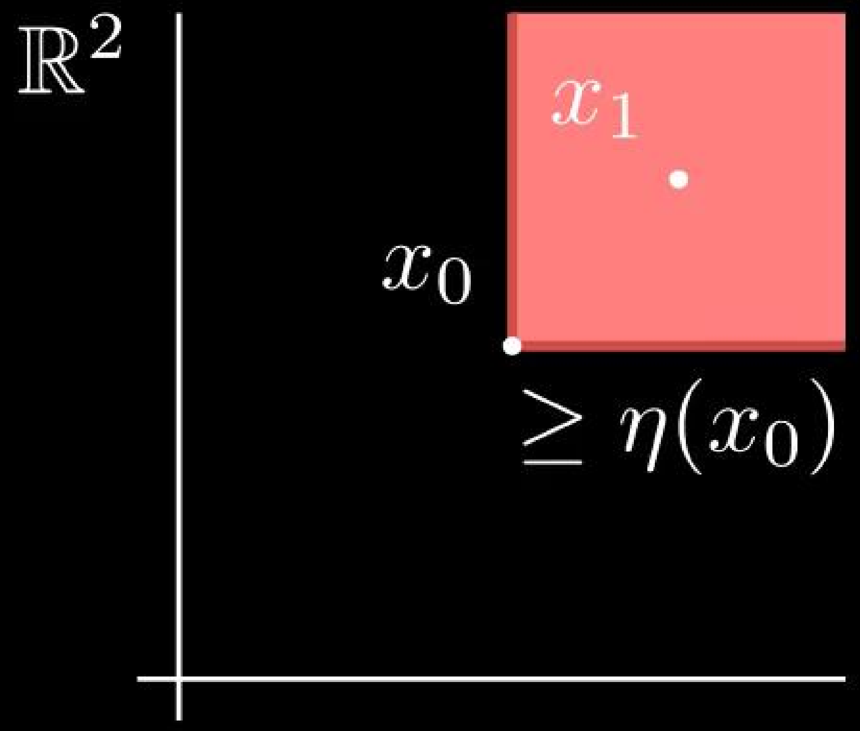
Statistical setting

Setting:

Fix $\sigma > 0$. Let $\mathcal{P}_{\mathrm{Mon},d}(\sigma)$ be the family of distributions P on $\mathbb{R}^d \times \mathbb{R}$ such that for $(X, Y) \sim P$,

(i) the regression function $\eta(x) := \mathbb{E}(Y|X = x)$ is increasing on $\mathbb{R}^d$, i.e. $x_0 \preccurlyeq x_1 \implies \eta(x_0) \leq \eta(x_1)$,

(ii) $Y - \eta(X) \mid X = x$ is sub-Gaussian with variance parameter σ^2 for $x \in \mathbb{R}^d$.

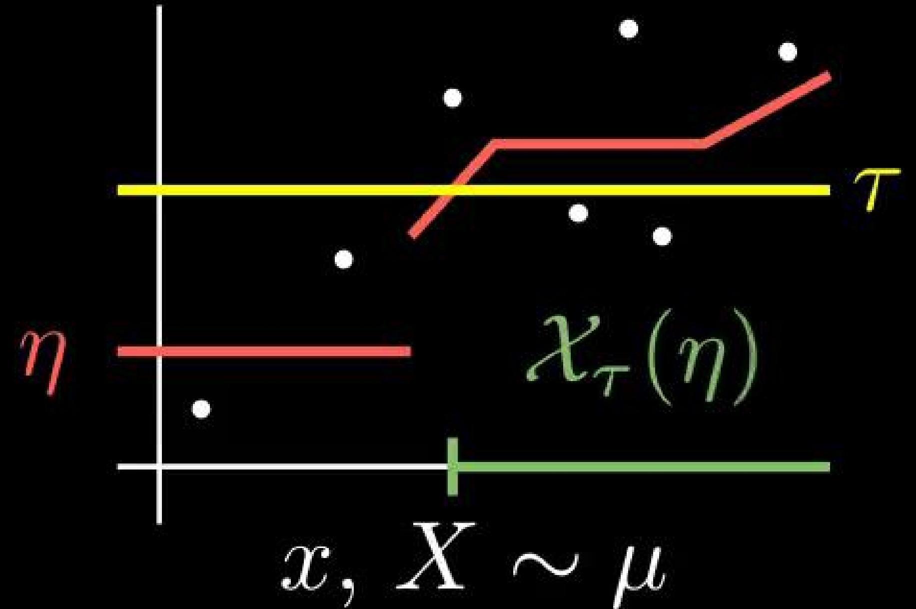


Notation:

- Fix $\tau \in \mathbb{R}$. Define τ-superlevel set by

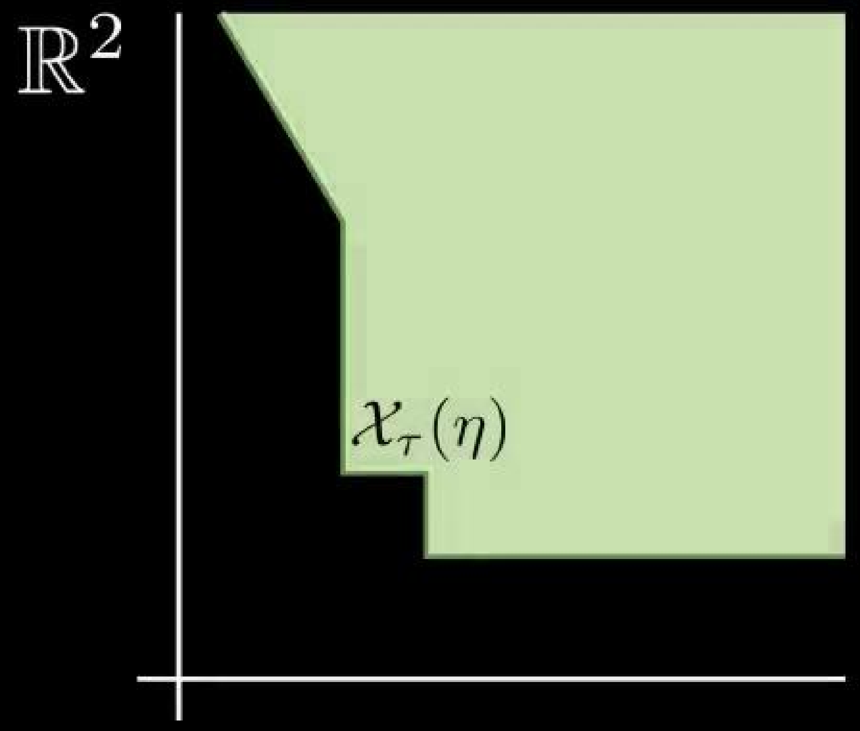
$$\mathcal{X}_\tau(\eta) := \{x \in \mathbb{R}^d : \eta(x) \geq \tau\}$$

- Denote the marginal distribution of X by μ.



Goal

Writing $\mathcal{D} := \big((X_1, Y_1), \ldots, (X_n, Y_n)\big) \sim P^n$, we want $\hat{A} : \mathcal{D} \mapsto \hat{A}(\mathcal{D}) \subseteq \mathbb{R}^d$ such that:

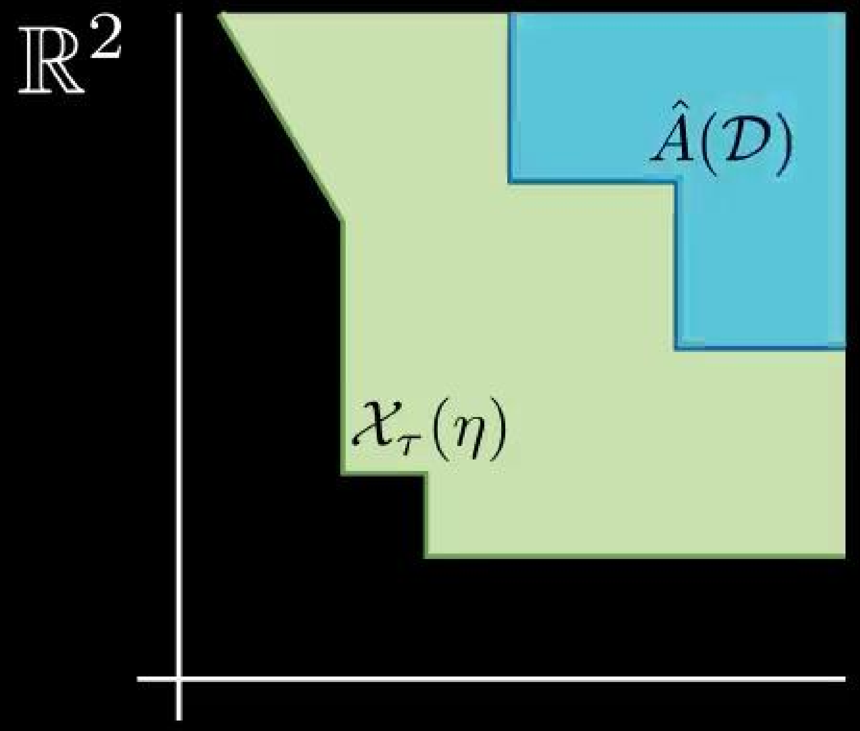


Goal

Writing $\mathcal{D} := \big((X_1, Y_1), \ldots, (X_n, Y_n)\big) \sim P^n$, we want $\hat{A} : \mathcal{D} \mapsto \hat{A}(\mathcal{D}) \subseteq \mathbb{R}^d$ such that:

Type I error control: Fix $\alpha \in (0,1)$. Require $\forall P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma)$:

$$\mathbb{P}\big\{\forall x \in \hat{A}(\mathcal{D}) : \eta(x) \geq \tau\big\} \equiv \mathbb{P}\big\{\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\big\} \geq 1 - \alpha.$$



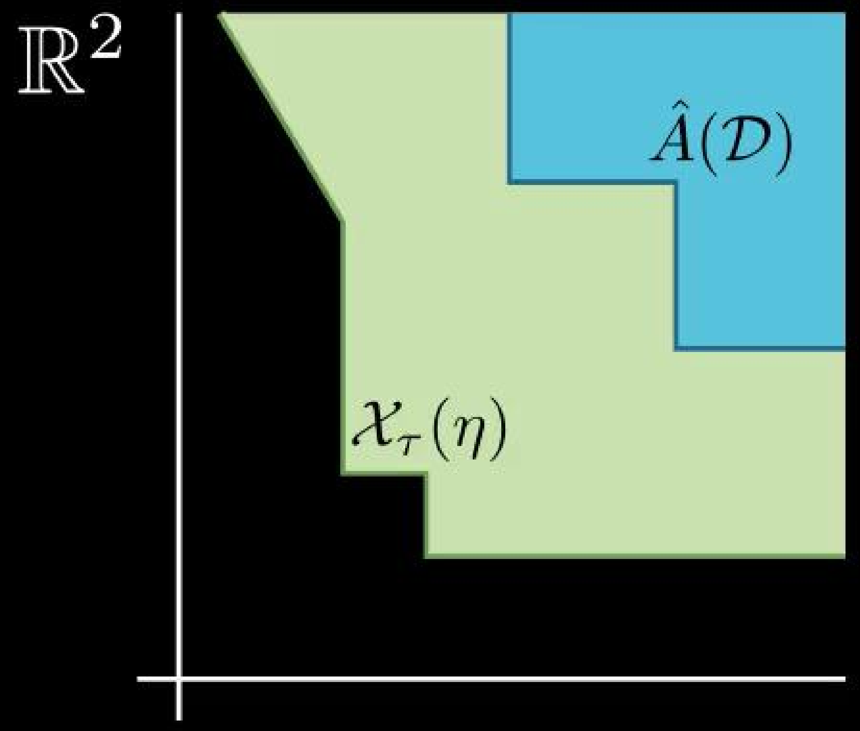
Goal

Writing $\mathcal{D} := \big((X_1, Y_1), \ldots, (X_n, Y_n)\big) \sim P^n$, we want $\hat{A} : \mathcal{D} \mapsto \hat{A}(\mathcal{D}) \subseteq \mathbb{R}^d$ such that:

Type I error control: Fix $\alpha \in (0, 1)$. Require $\forall P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma)$:

$$\mathbb{P}\big\{\forall x \in \hat{A}(\mathcal{D}) : \eta(x) \geq \tau\big\} \equiv \mathbb{P}\big\{\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\big\} \geq 1 - \alpha.$$

Power: Want small regret $\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}(\mathcal{D})\big)$.



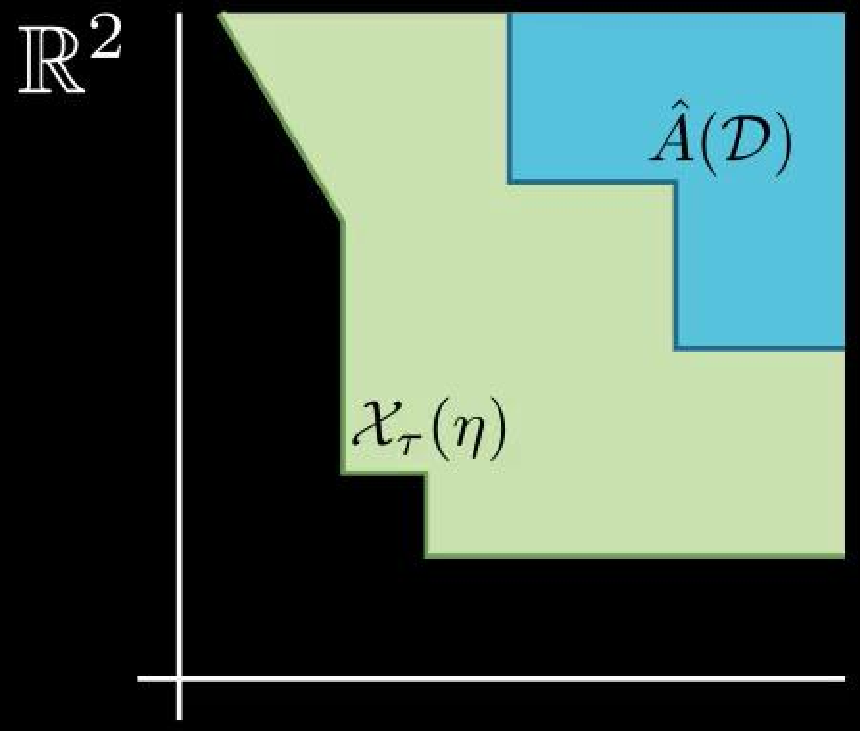
Goal

Writing $\mathcal{D} := \big((X_1, Y_1), \ldots, (X_n, Y_n)\big) \sim P^n$, we want $\hat{A} : \mathcal{D} \mapsto \hat{A}(\mathcal{D}) \subseteq \mathbb{R}^d$ such that:

Type I error control: Fix $\alpha \in (0,1)$. Require $\forall P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma)$:

$$\mathbb{P}\big\{\forall x \in \hat{A}(\mathcal{D}) : \eta(x) \geq \tau\big\} \equiv \mathbb{P}\big\{\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\big\} \geq 1 - \alpha.$$

Power: Want small regret $\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}(\mathcal{D})\big)$.

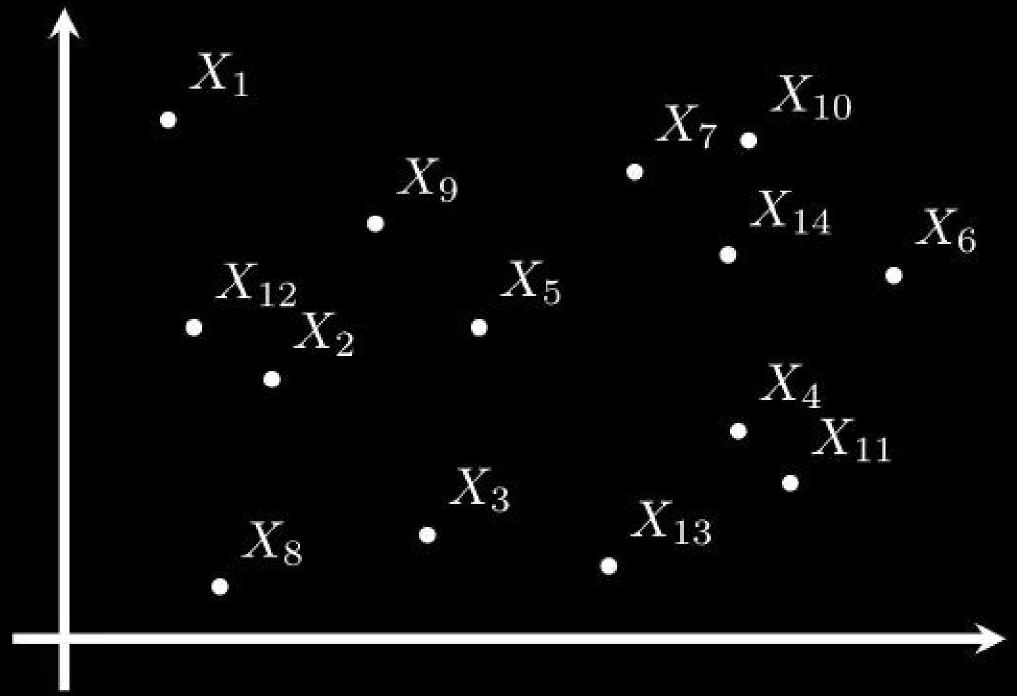


Example:

Type I error: $\hat{A}$ should contain only those HIV-patients for whom the conditional probability of not facing a negative event within the next 5 years is at least 98%.

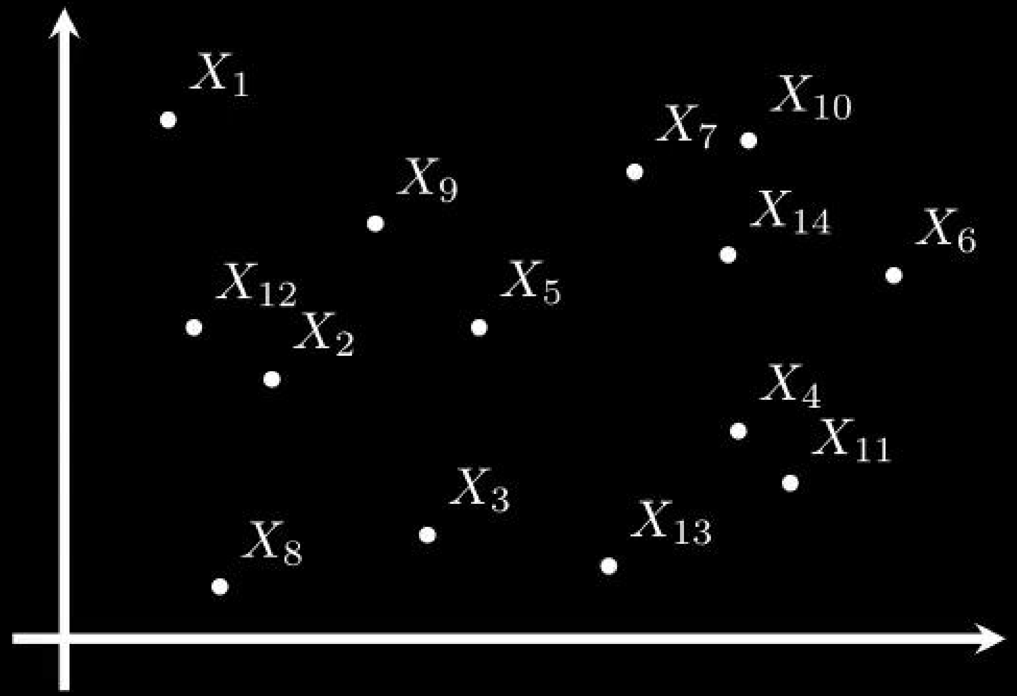
High-level strategy

For $x_0 \in \mathbb{R}^d$, define null hypothesis $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.



High-level strategy

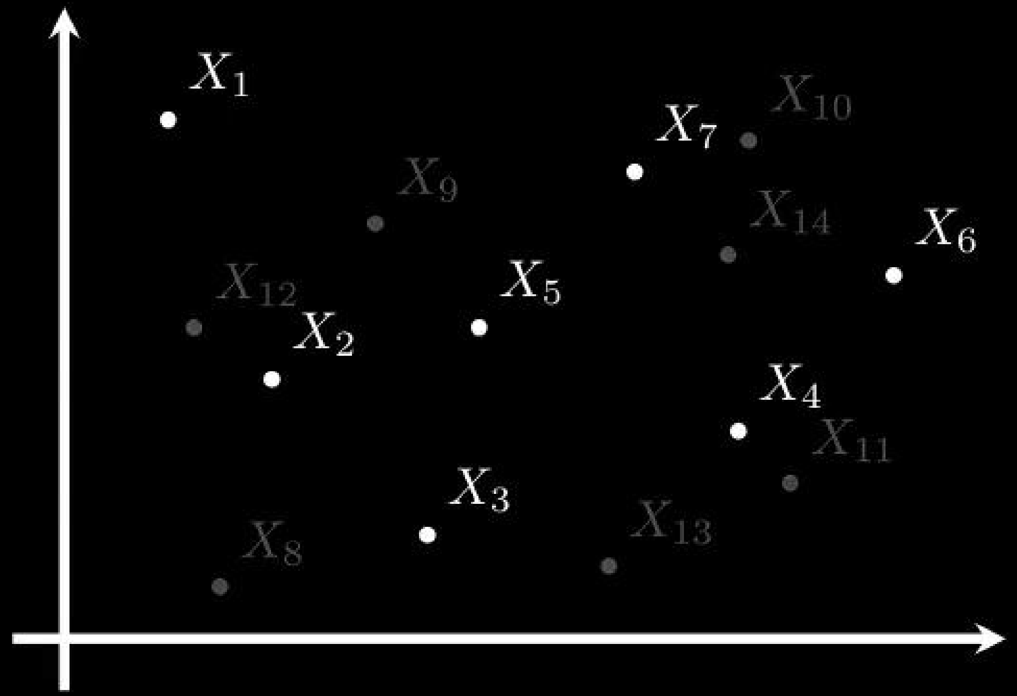
For $x_0 \in \mathbb{R}^d$, define null hypothesis $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.



High-level strategy:

High-level strategy

For $x_0 \in \mathbb{R}^d$, define null hypothesis $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.

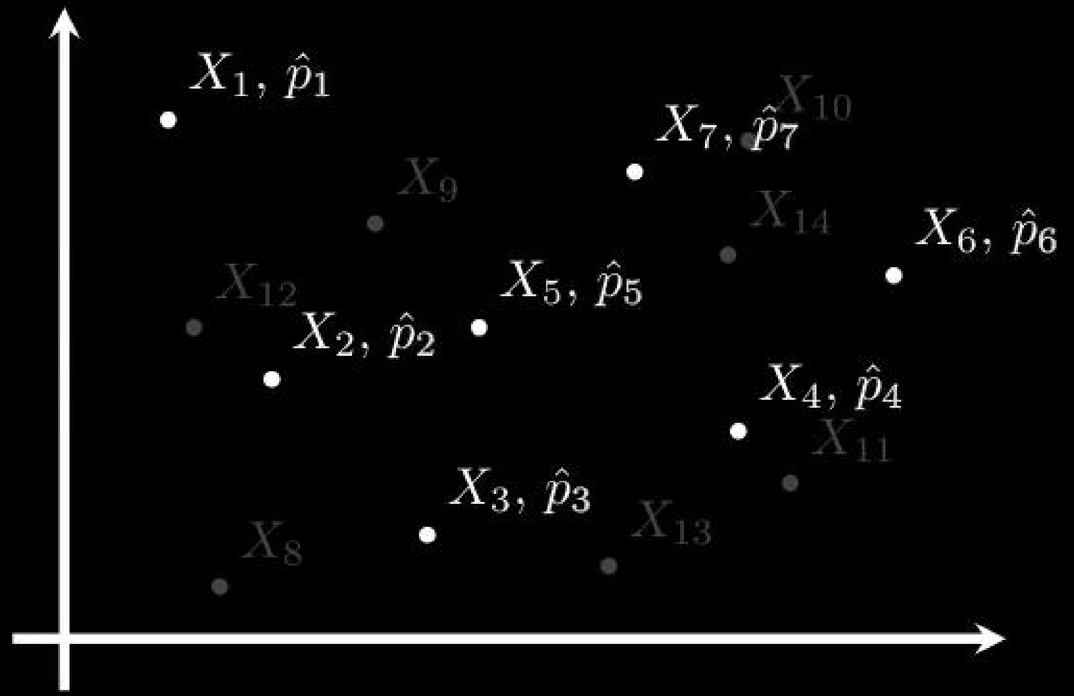


High-level strategy:

1. Subsample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;

High-level strategy

For $x_0 \in \mathbb{R}^d$, define null hypothesis $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.

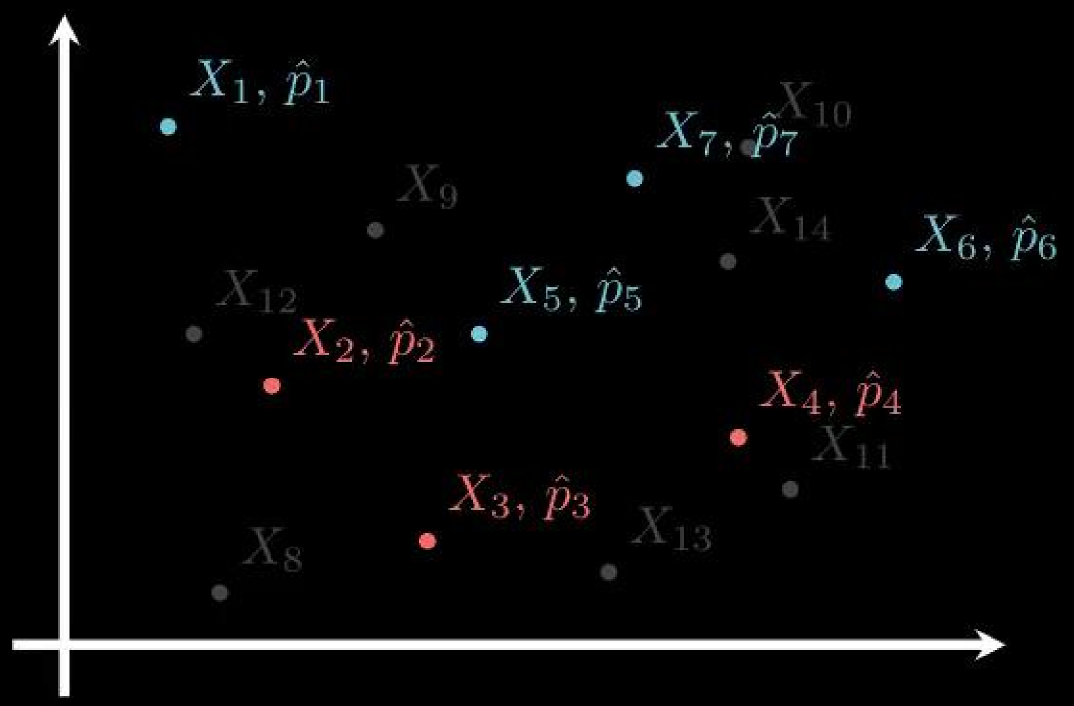


High-level strategy:

1. Subsample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Calculate *p-values* $\hat{p}_i$ for $H_0(X_i)$, $i \in [m]$;

High-level strategy

For $x_0 \in \mathbb{R}^d$, define null hypothesis $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.

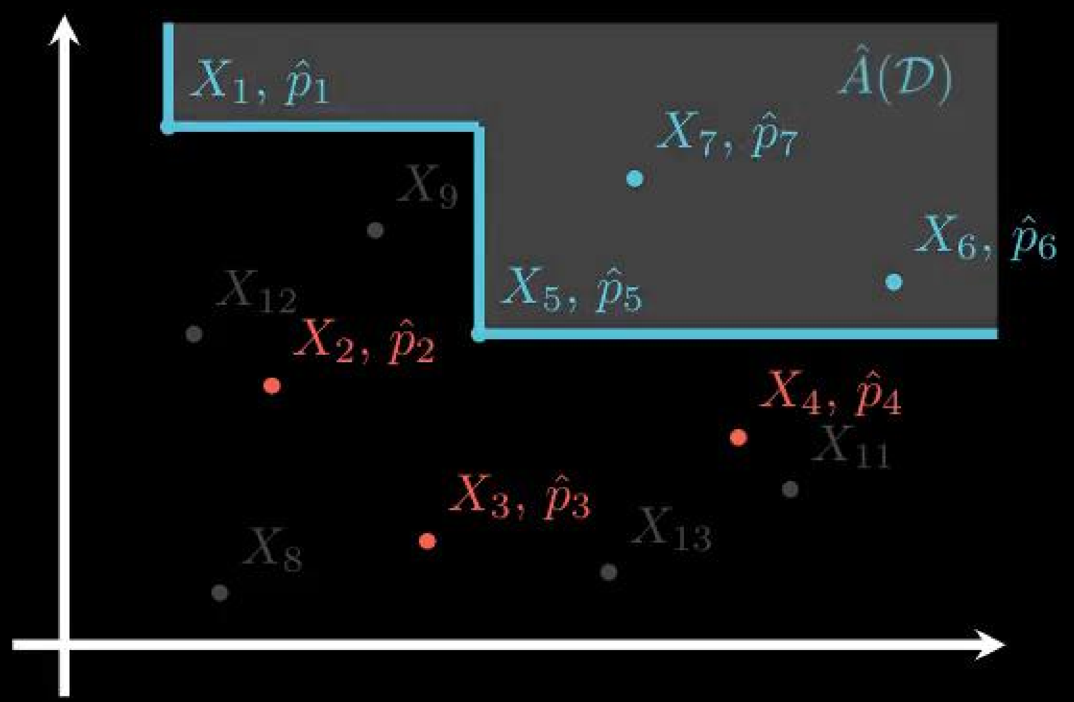


High-level strategy:

1. Subsample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Calculate *p-values* $\hat{p}_i$ for $H_0(X_i)$, $i \in [m]$;
3. Apply a *multiple testing procedure* with FWER-control at level α to reject $\mathcal{R}_\alpha \subseteq [m]$;

High-level strategy

For $x_0 \in \mathbb{R}^d$, define null hypothesis $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.

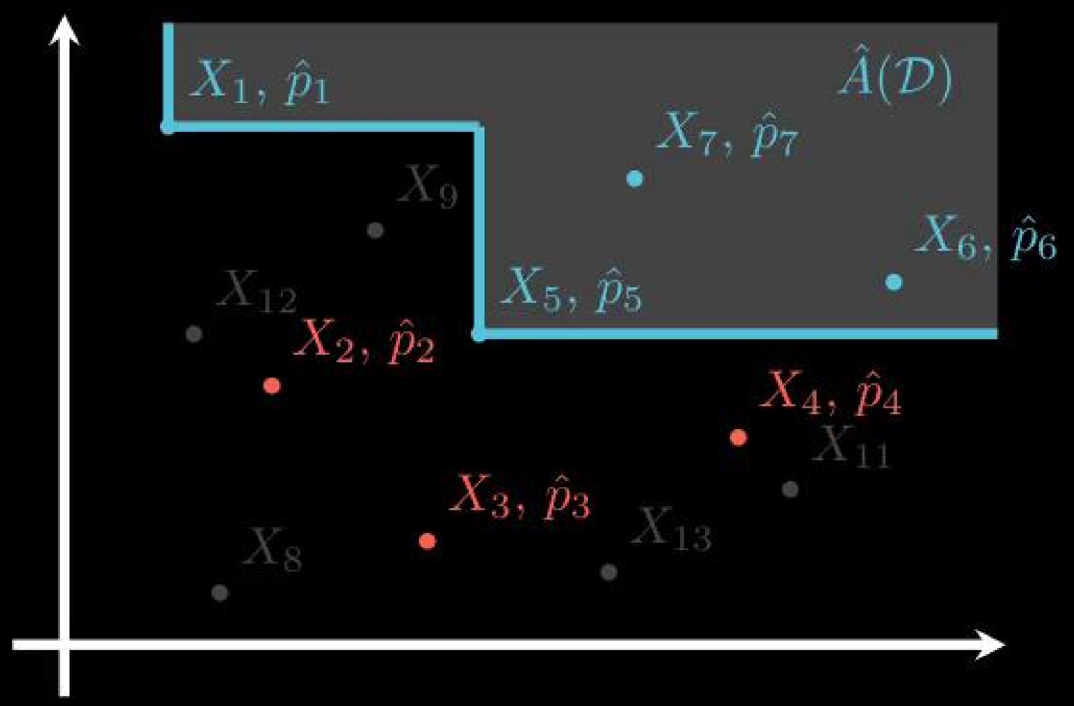


High-level strategy:

1. Subsample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Calculate *p-values* $\hat{p}_i$ for $H_0(X_i)$, $i \in [m]$;
3. Apply a *multiple testing procedure* with FWER-control at level α to reject $\mathcal{R}_\alpha \subseteq [m]$;
4. Output $\hat{A} := \{x \in \mathbb{R}^d : X_\ell \preccurlyeq x \text{ for some } \ell \in \mathcal{R}_\alpha\}$.

High-level strategy

For $x_0 \in \mathbb{R}^d$, define null hypothesis $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.



High-level strategy:

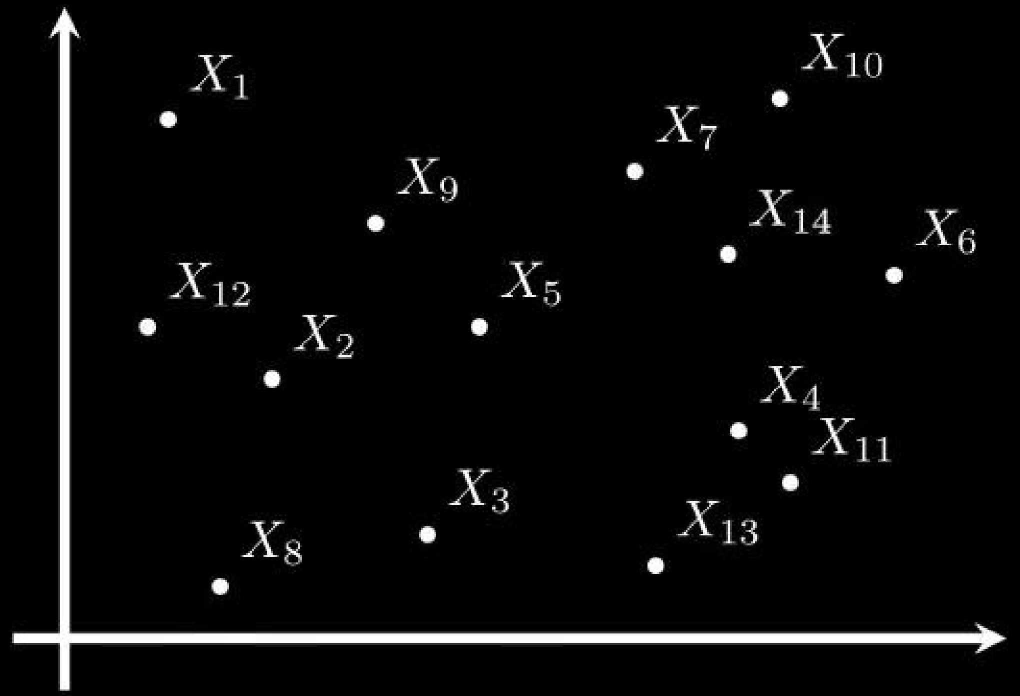
1. Subsample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Calculate *p-values* $\hat{p}_i$ for $H_0(X_i)$, $i \in [m]$;
3. Apply a *multiple testing procedure* with FWER-control at level α to reject $\mathcal{R}_\alpha \subseteq [m]$;
4. Output $\hat{A} := \{x \in \mathbb{R}^d : X_\ell \preccurlyeq x \text{ for some } \ell \in \mathcal{R}_\alpha\}$.

Construct p-values $\hat{p}_i$ for $H_0(X_i)$, $i \in [m]$

Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.

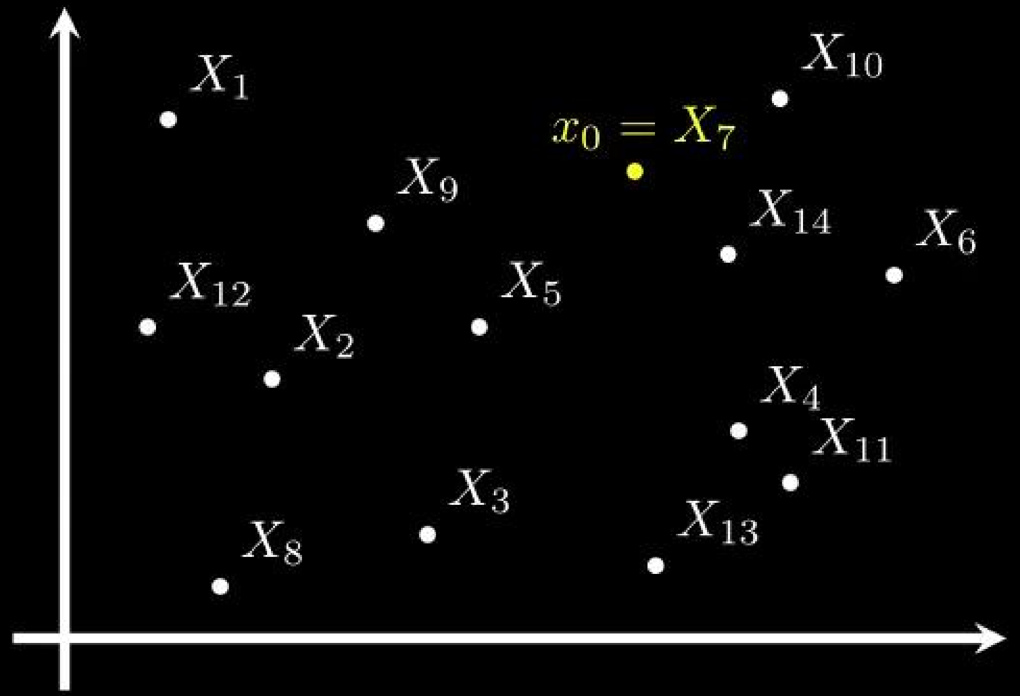
Construct p-values $\hat{p}_i$ for $H_0(X_i)$, $i \in [m]$

Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.



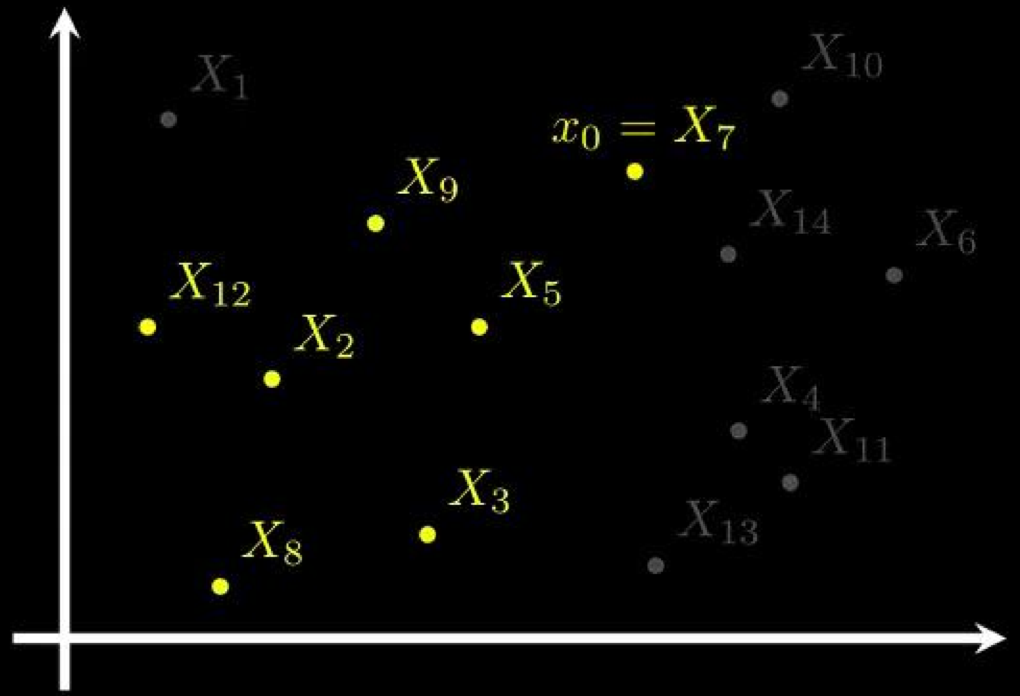
Construct p-values $\hat{p}_i$ for $H_0(X_i)$, $i \in [m]$

Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.



Construct p-values $\hat{p}_i$ for $H_0(X_i)$, $i \in [m]$

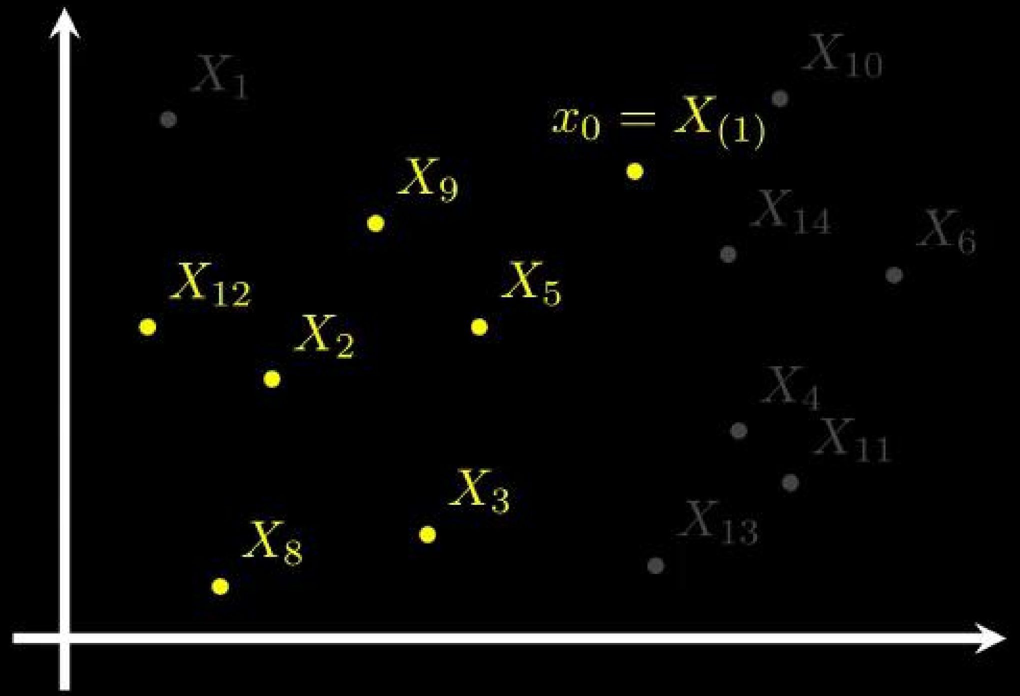
Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.

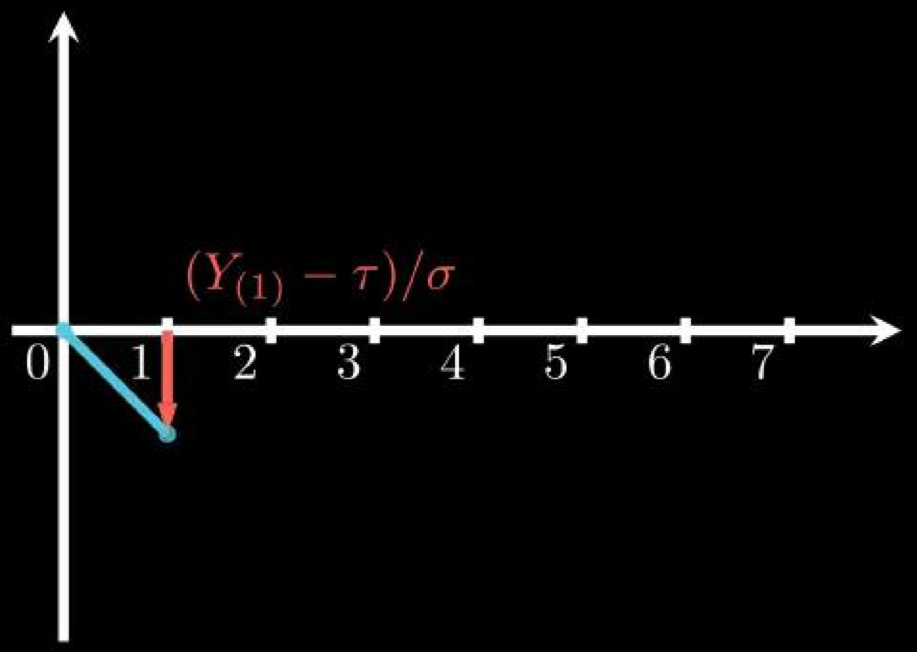


Denote $\mathcal{I}(x_0) := \{i \in [n] : X_i \preccurlyeq x_0\}$, $n(x_0) := |\mathcal{I}(x_0)|$.

Construct p-values $\hat{p}_i$ for $H_0(X_i)$, $i \in [m]$

Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.



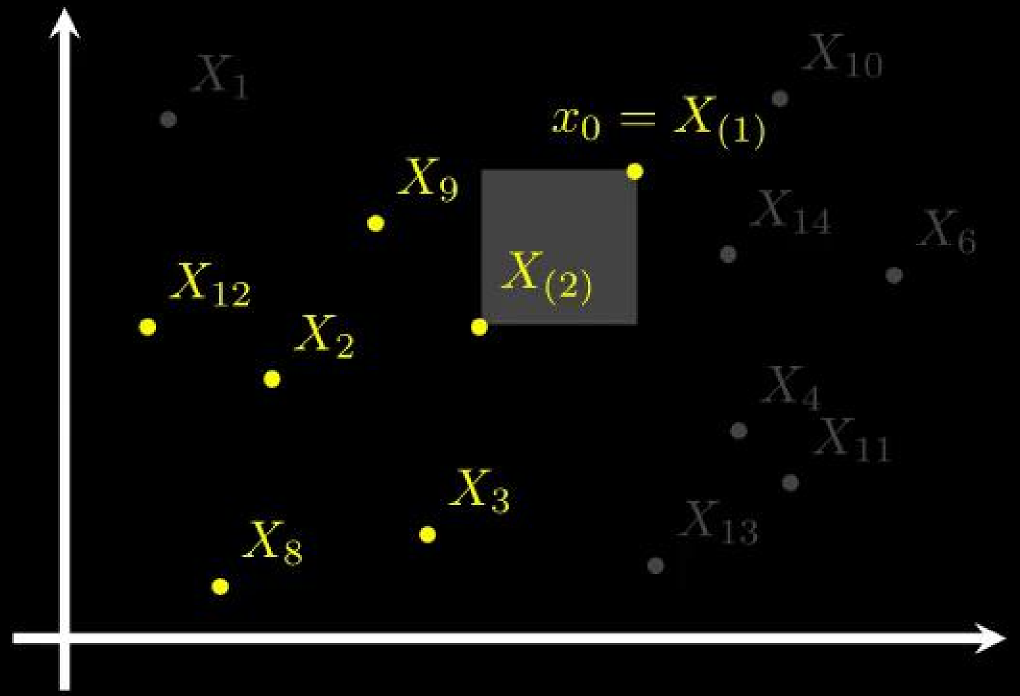


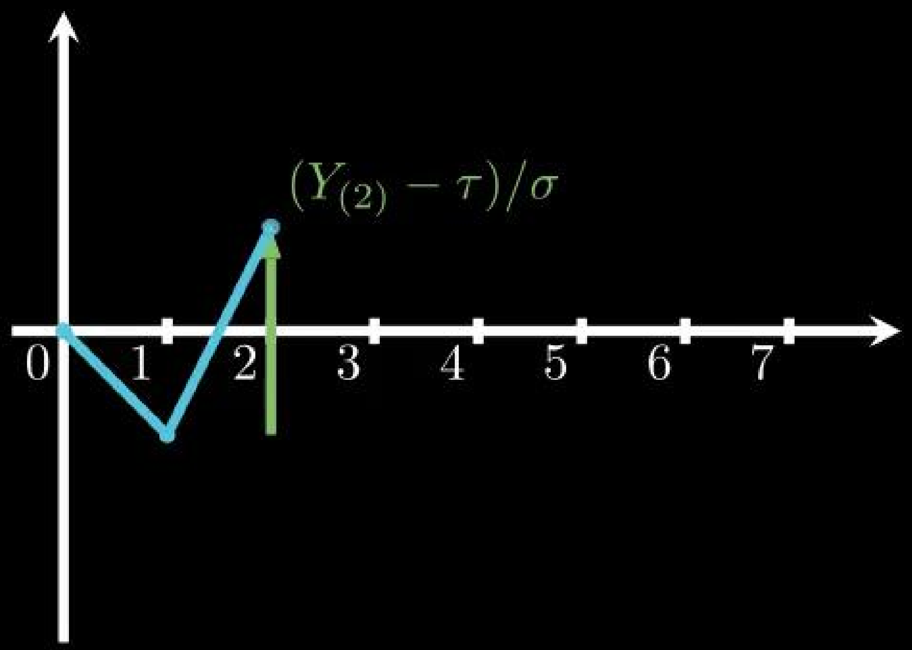
Denote $\mathcal{I}(x_0) := \{i \in [n] : X_i \preccurlyeq x_0\}$, $n(x_0) := |\mathcal{I}(x_0)|$.

Let $X_{(j)}$ be the jth nearest neighbour of x_0 among X_i, $i \in \mathcal{I}(x_0)$, in sup-norm and let $Y_{(j)}$ be the corresponding response.

Construct p-values $\hat{p}_i$ for $H_0(X_i)$, $i \in [m]$

Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.



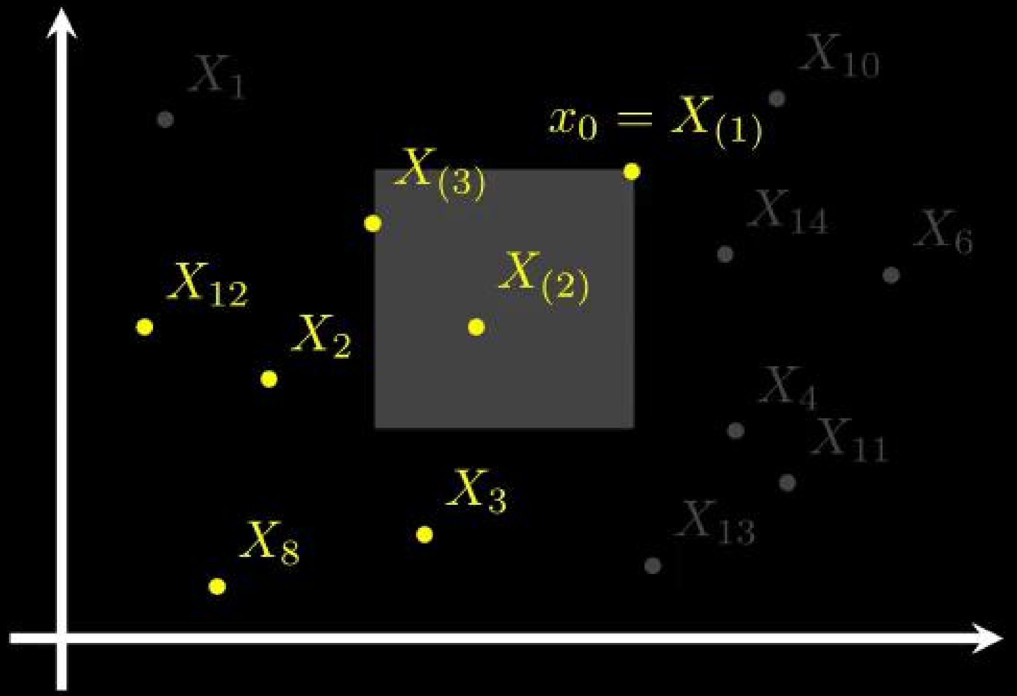


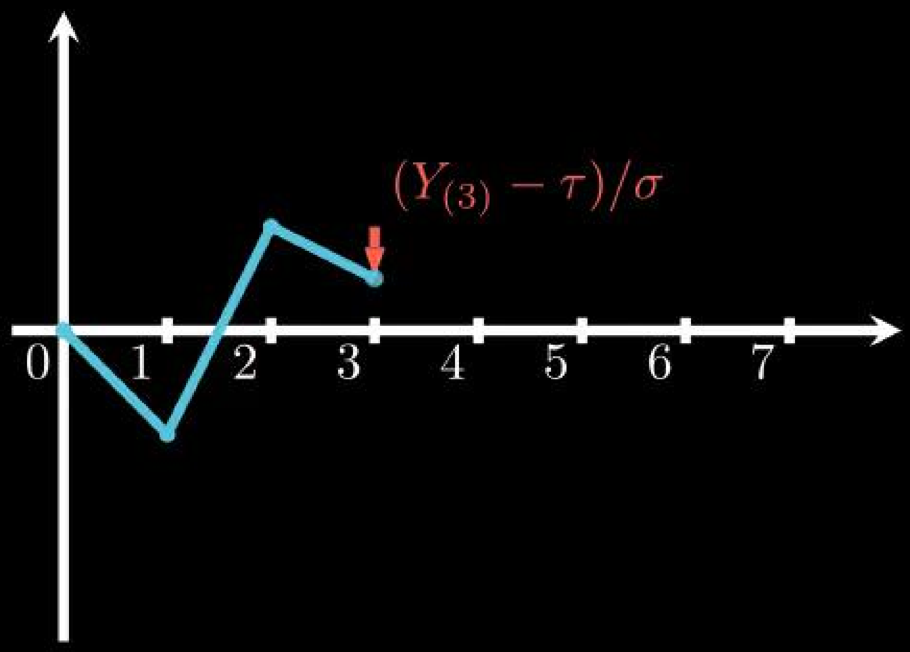
Denote $\mathcal{I}(x_0) := \{i \in [n] : X_i \preccurlyeq x_0\}$, $n(x_0) := |\mathcal{I}(x_0)|$.

Let $X_{(j)}$ be the jth nearest neighbour of x_0 among X_i, $i \in \mathcal{I}(x_0)$, in sup-norm and let $Y_{(j)}$ be the corresponding response.

Construct p-values $\hat{p}_i$ for $H_0(X_i)$, $i \in [m]$

Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.



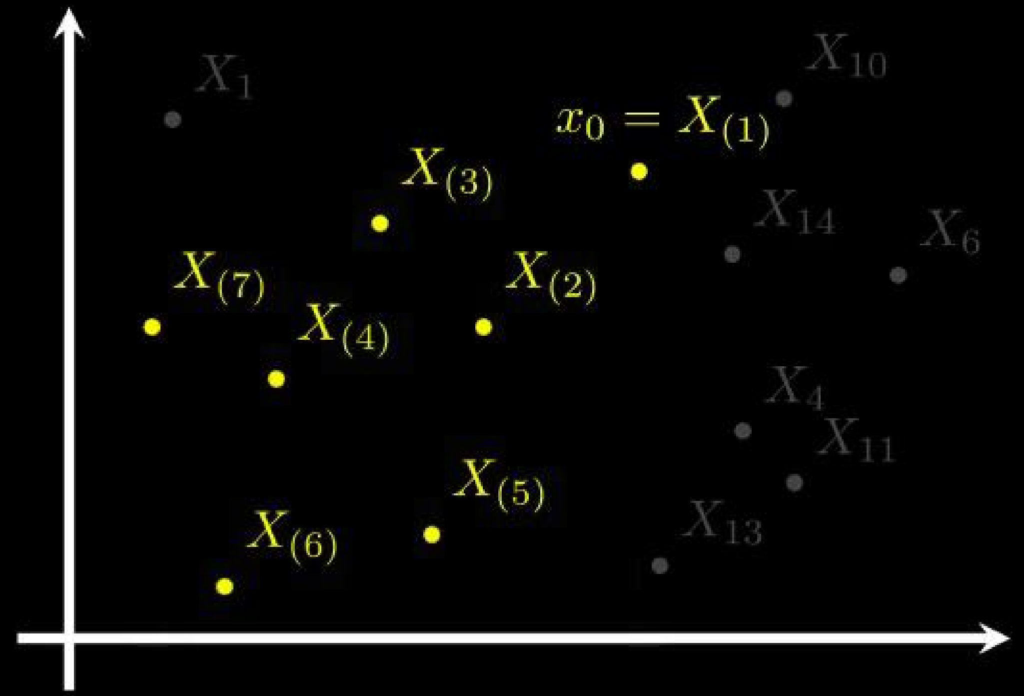


Denote $\mathcal{I}(x_0) := \{i \in [n] : X_i \preccurlyeq x_0\}$, $n(x_0) := |\mathcal{I}(x_0)|$.

Let $X_{(j)}$ be the jth nearest neighbour of x_0 among X_i, $i \in \mathcal{I}(x_0)$, in sup-norm and let $Y_{(j)}$ be the corresponding response.

Construct p-values $\hat{p}_i$ for $H_0(X_i)$, $i \in [m]$

Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.

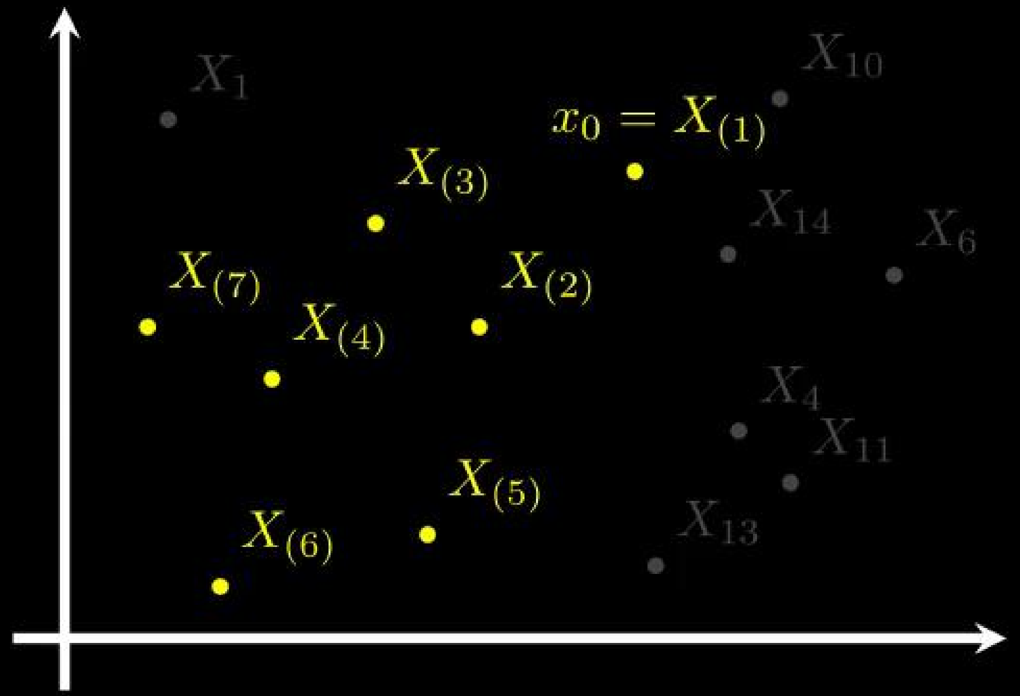


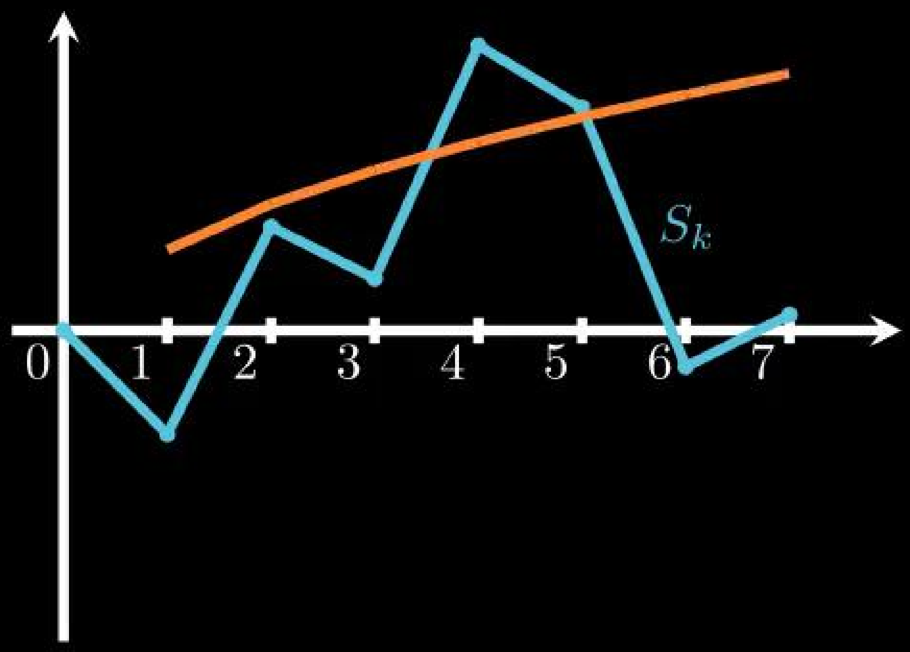
Denote $\mathcal{I}(x_0) := \{i \in [n] : X_i \preccurlyeq x_0\}$, $n(x_0) := |\mathcal{I}(x_0)|$.

Let $X_{(j)}$ be the jth nearest neighbour of x_0 among X_i, $i \in \mathcal{I}(x_0)$, in sup-norm and let $Y_{(j)}$ be the corresponding response.

Construct p-values $\hat{p}_i$ for $H_0(X_i)$, $i \in [m]$

Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.





Denote $\mathcal{I}(x_0) := \{i \in [n] : X_i \preccurlyeq x_0\}$, $n(x_0) := |\mathcal{I}(x_0)|$.

Let $X_{(j)}$ be the jth nearest neighbour of x_0 among X_i, $i \in \mathcal{I}(x_0)$, in sup-norm and let $Y_{(j)}$ be the corresponding response.

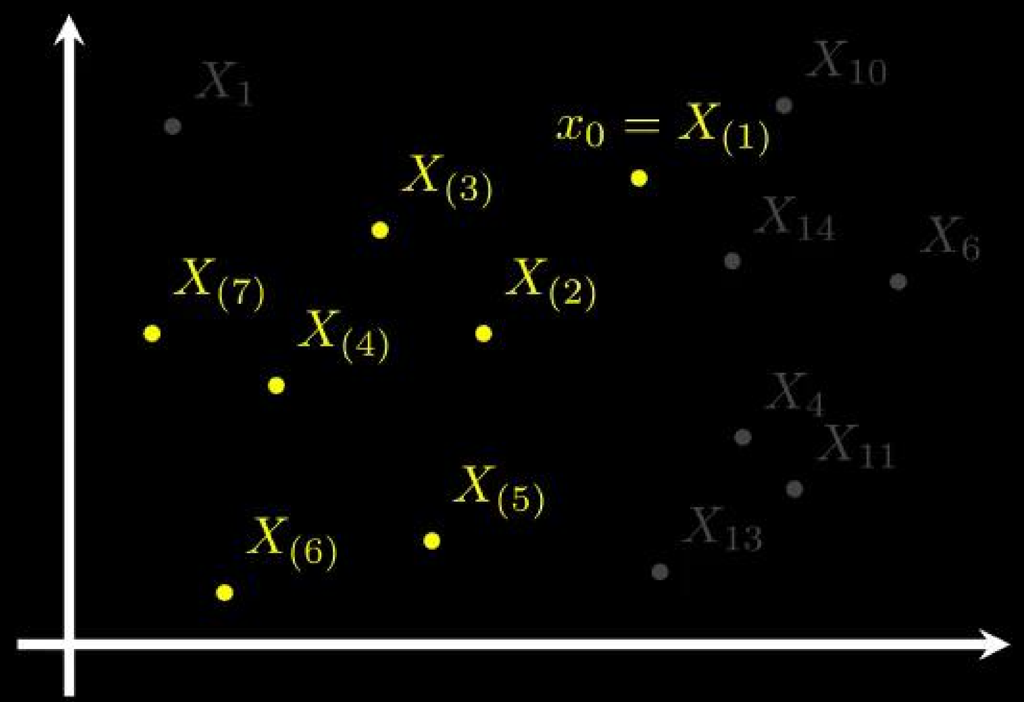
Let

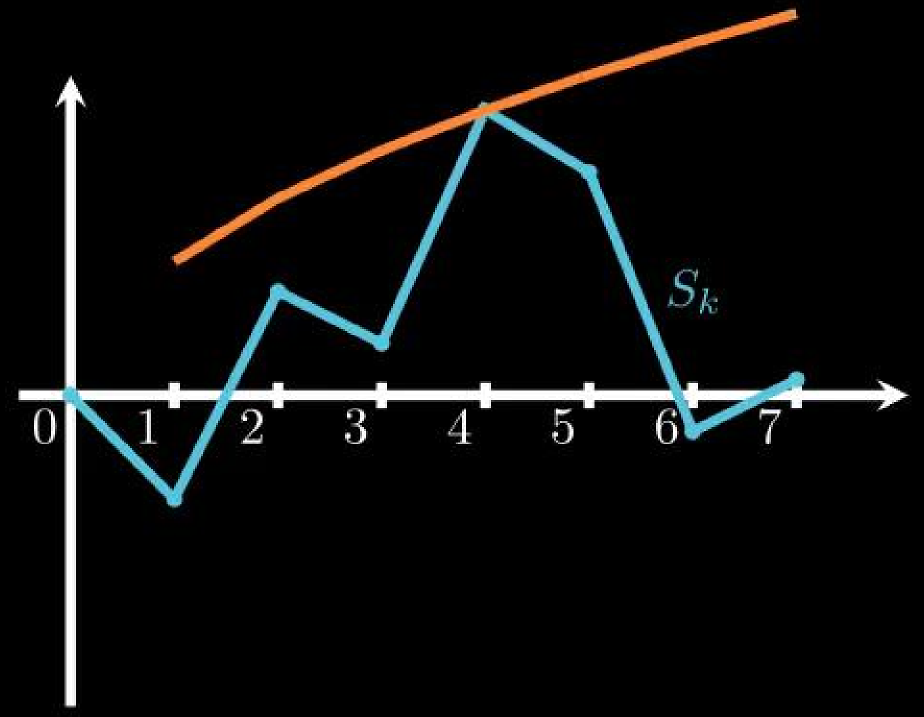
$$S_k := \sum_{j=1}^{k} \frac{Y_{(j)} - \tau}{\sigma}.$$

Then S_k is a supermartingale under $P \in H_0(x_0)$. Combination with time-uniform bounds by Howard et al. (2021) gives p-values from this martingale test (Duan et al., 2020).

Construct p-values $\hat{p}_i$ for $H_0(X_i)$, $i \in [m]$

Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.





Denote $\mathcal{I}(x_0) := \{i \in [n] : X_i \preccurlyeq x_0\}$, $n(x_0) := |\mathcal{I}(x_0)|$.

Let $X_{(j)}$ be the jth nearest neighbour of x_0 among X_i, $i \in \mathcal{I}(x_0)$, in sup-norm and let $Y_{(j)}$ be the corresponding response.

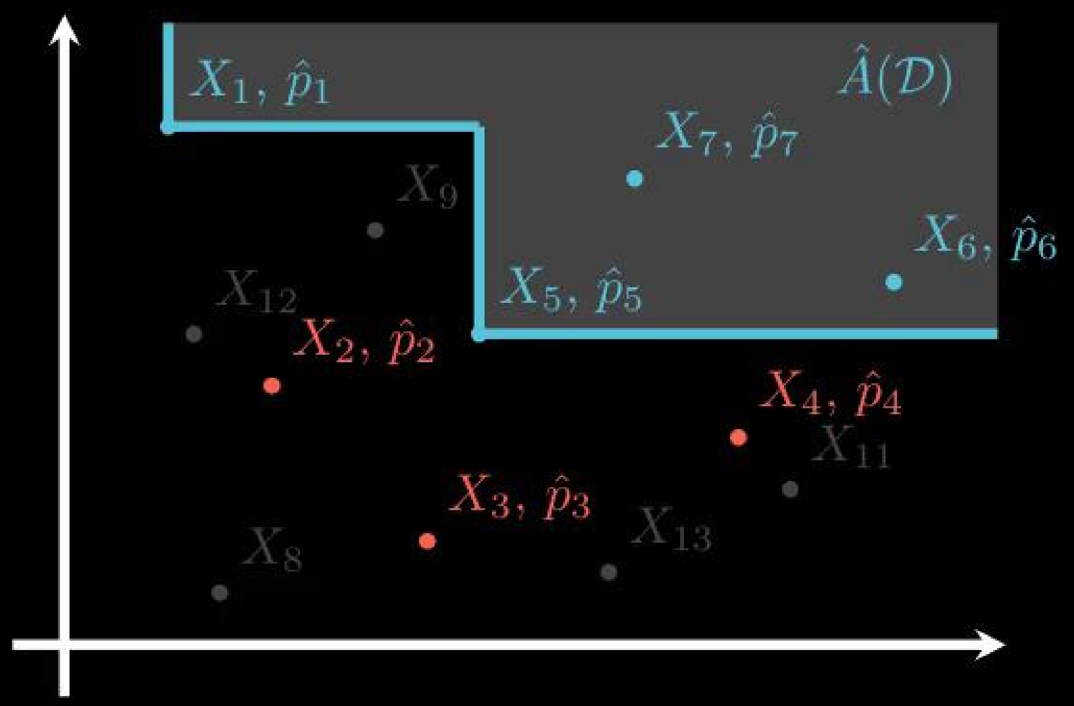
Let

$$S_k := \sum_{j=1}^{k} \frac{Y_{(j)} - \tau}{\sigma}.$$

Then S_k is a supermartingale under $P \in H_0(x_0)$. Combination with time-uniform bounds by Howard et al. (2021) gives p-values from this martingale test (Duan et al., 2020).

High-level strategy

For $x_0 \in \mathbb{R}^d$, define null hypothesis $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.

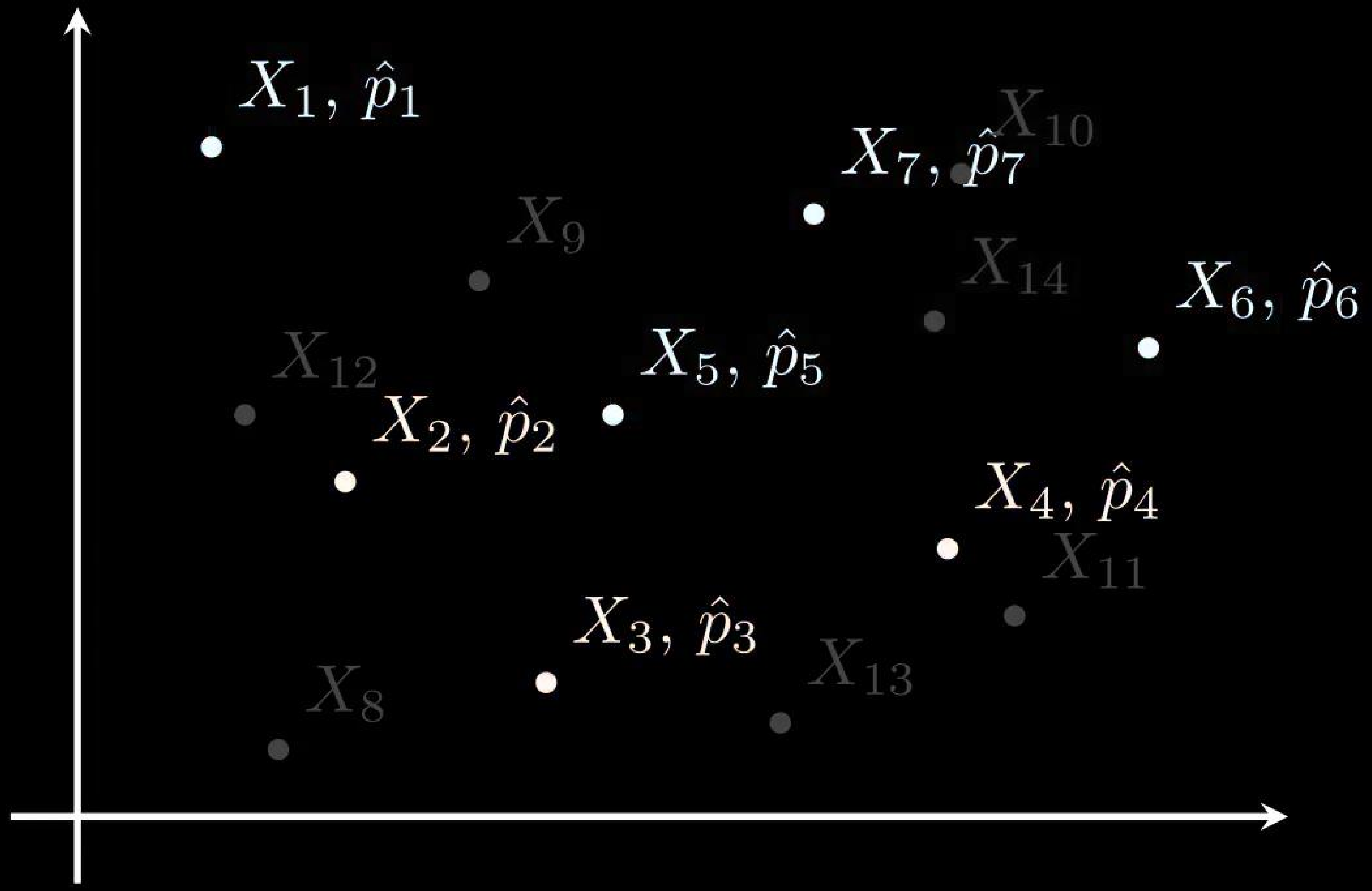


High-level strategy:

1. Subsample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Calculate *p-values* $\hat{p}_i$ for $H_0(X_i)$, $i \in [m]$;
3. Apply a *multiple testing procedure* with FWER-control at level α to reject $\mathcal{R}_\alpha \subseteq [m]$;
4. Output $\hat{A} := \{x \in \mathbb{R}^d : X_\ell \preccurlyeq x \text{ for some } \ell \in \mathcal{R}_\alpha\}$.

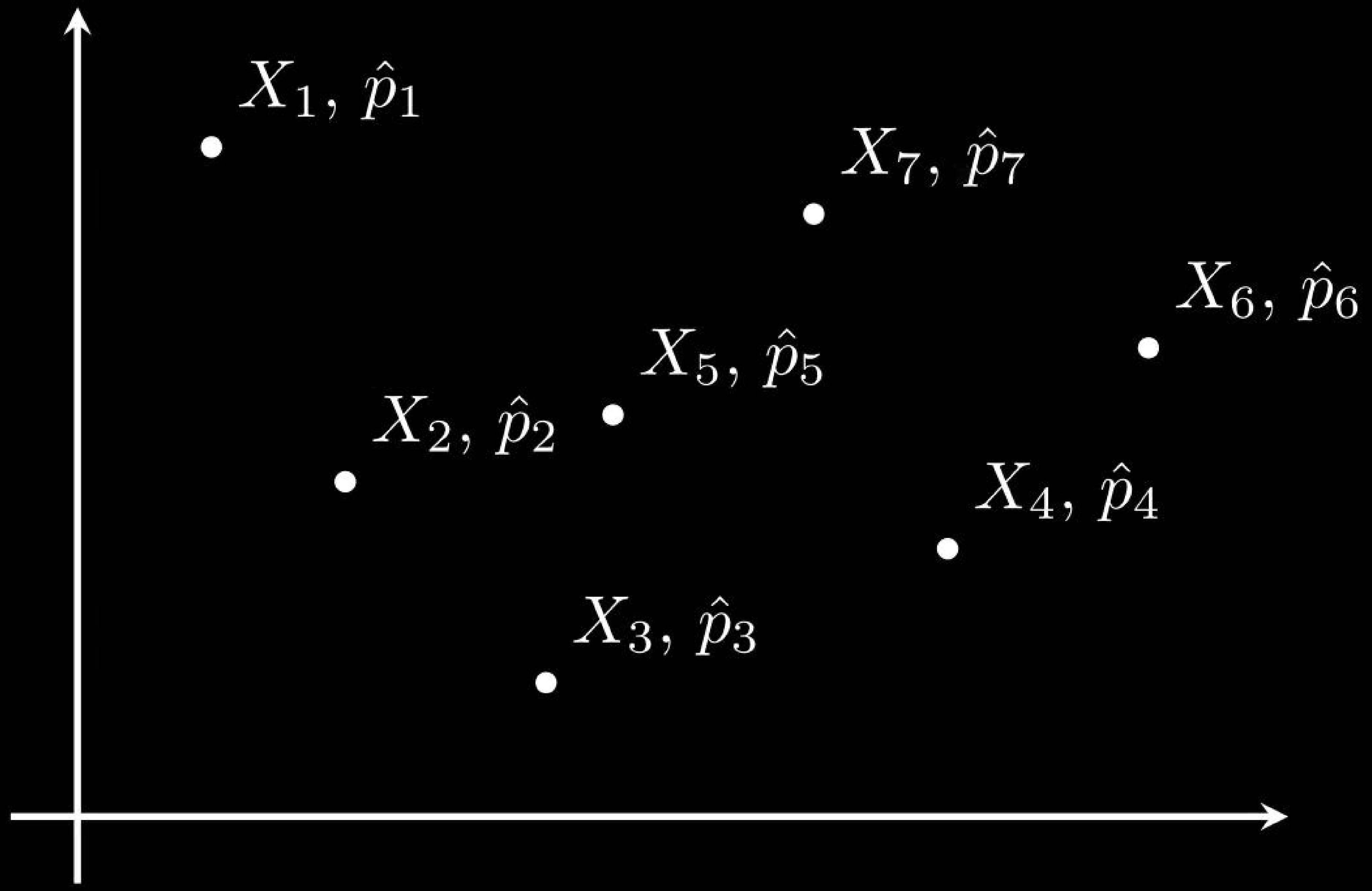
Multiple testing procedure

Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure.*



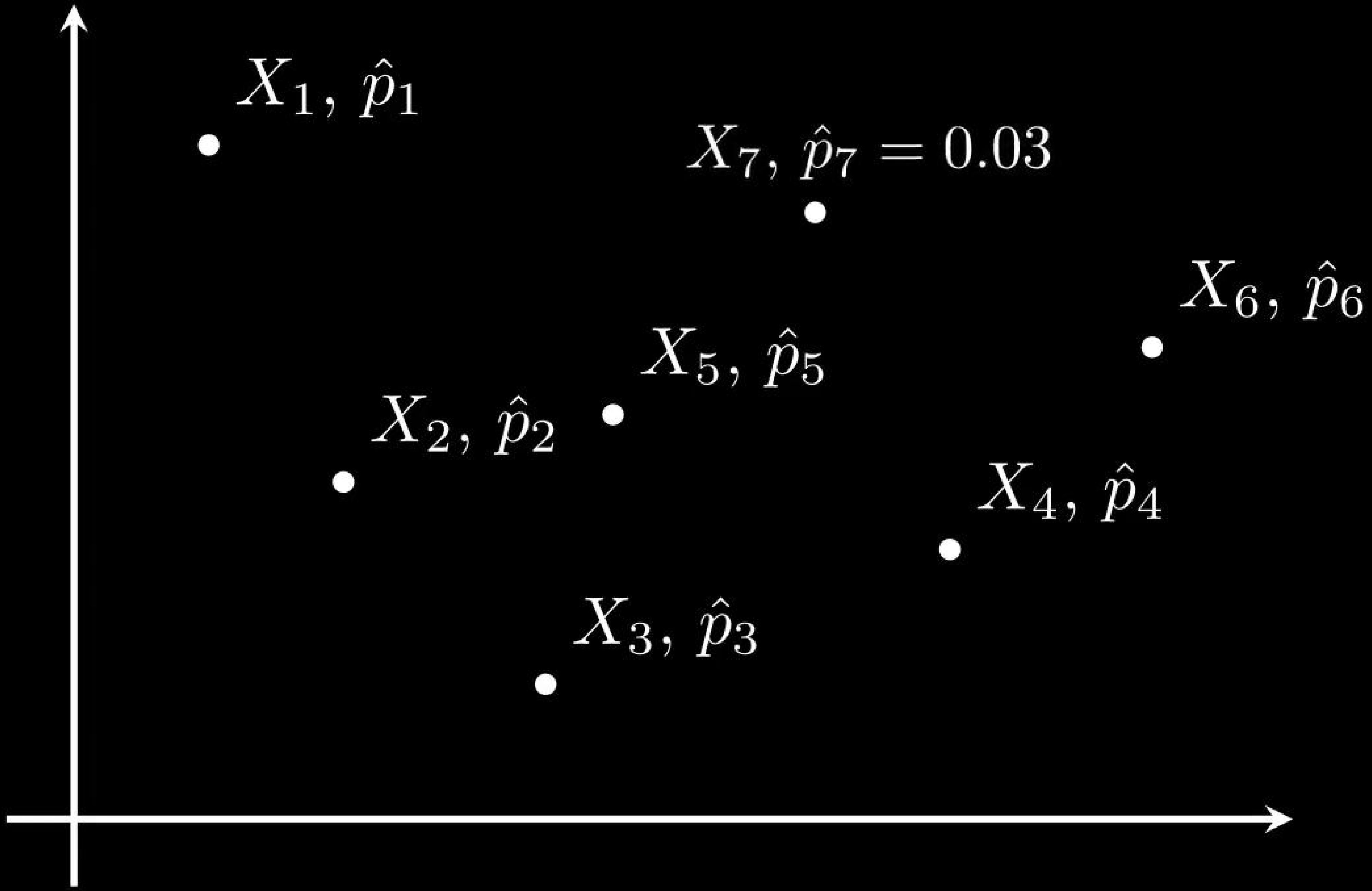
Multiple testing procedure

Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure.*



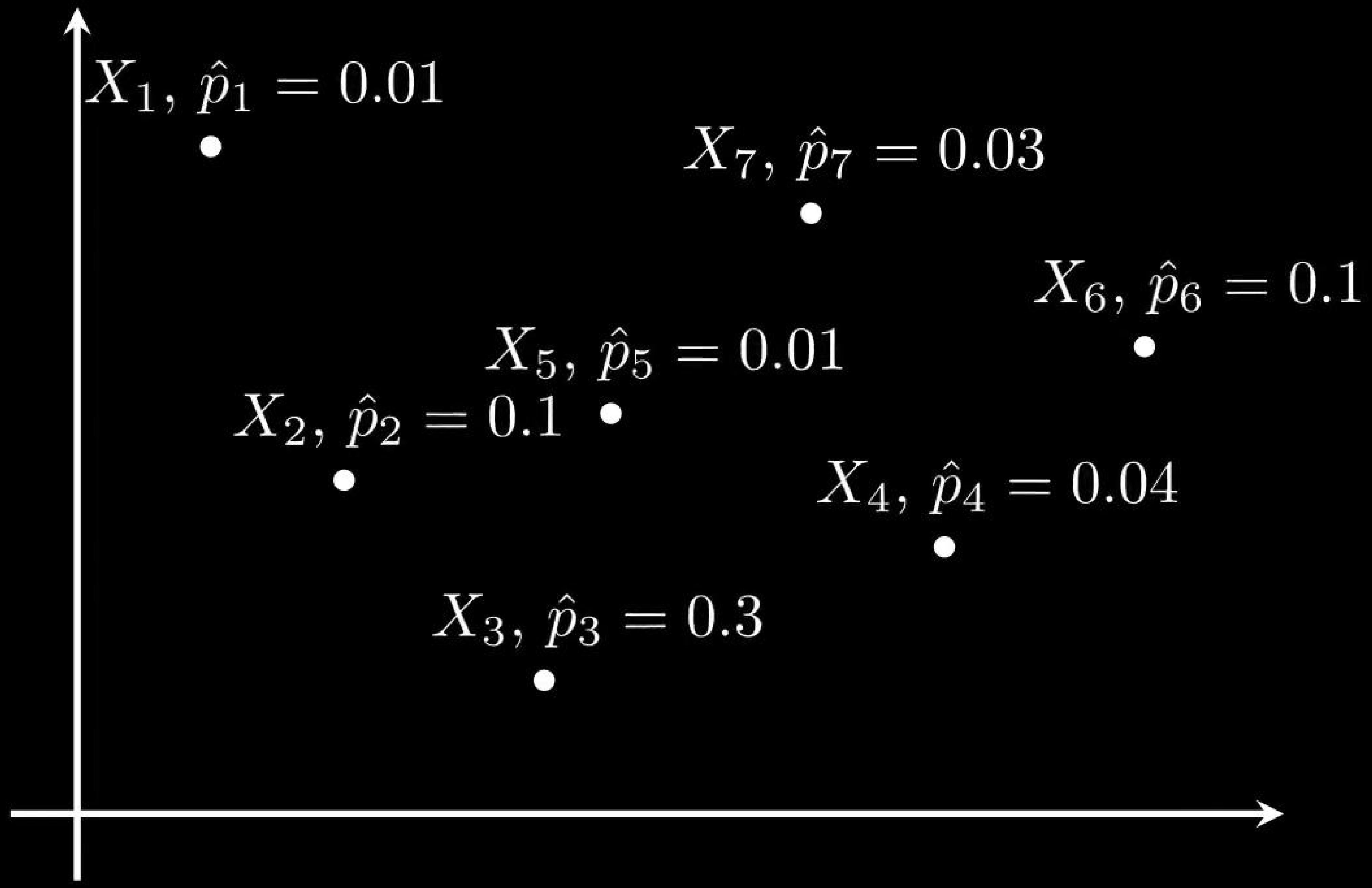
Multiple testing procedure

Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure.*



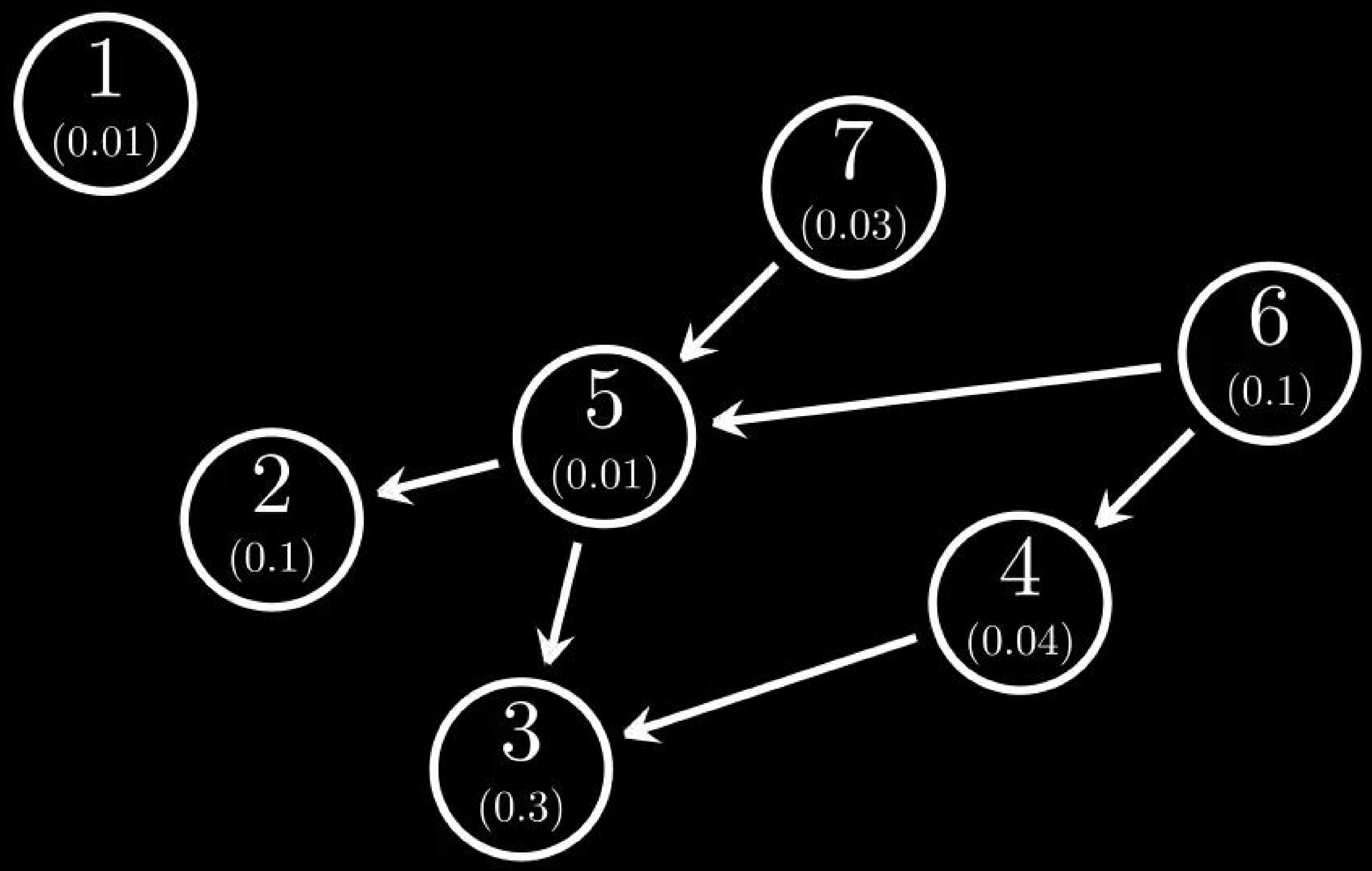
Multiple testing procedure

Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure.*



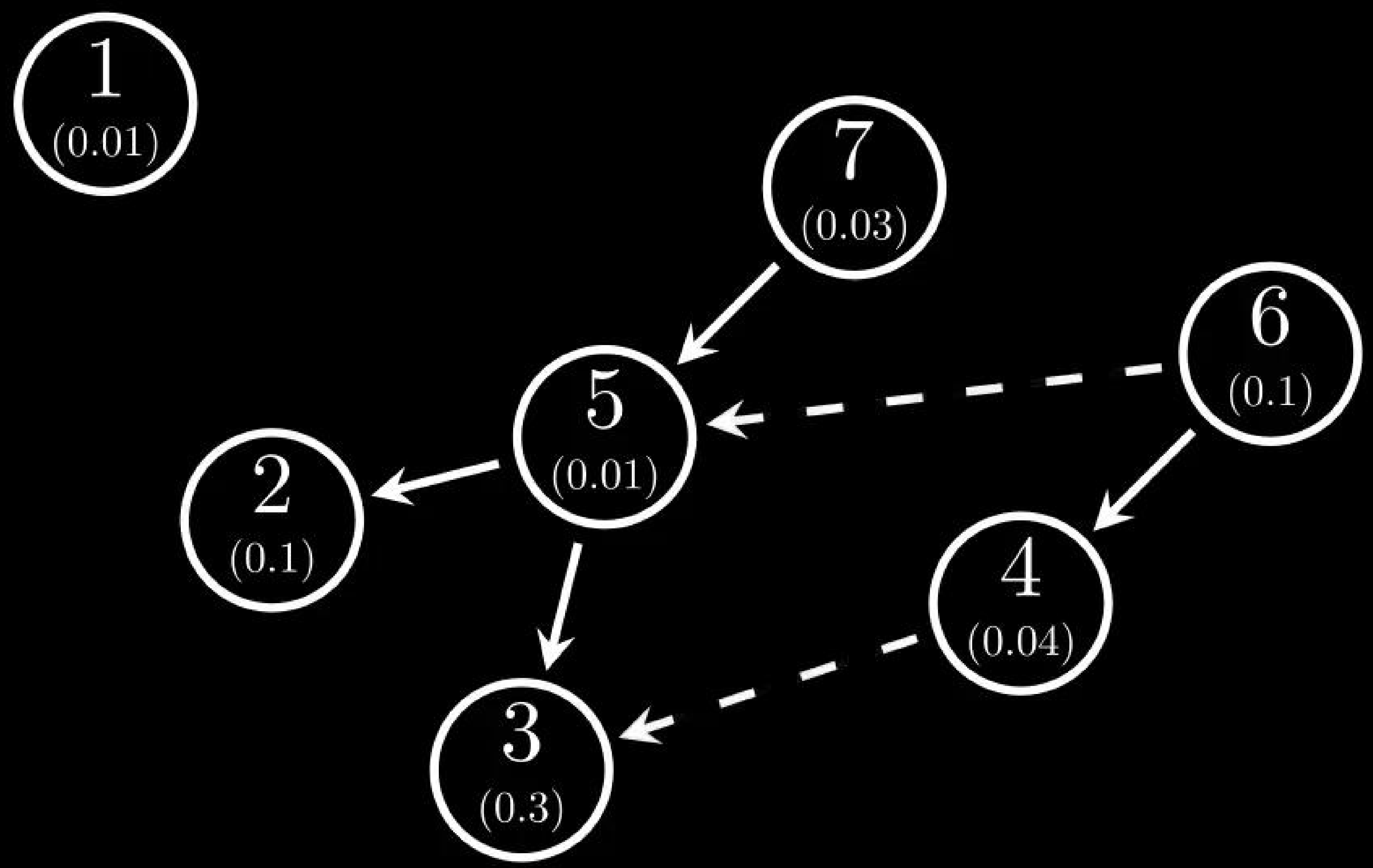
Multiple testing procedure

Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure*.



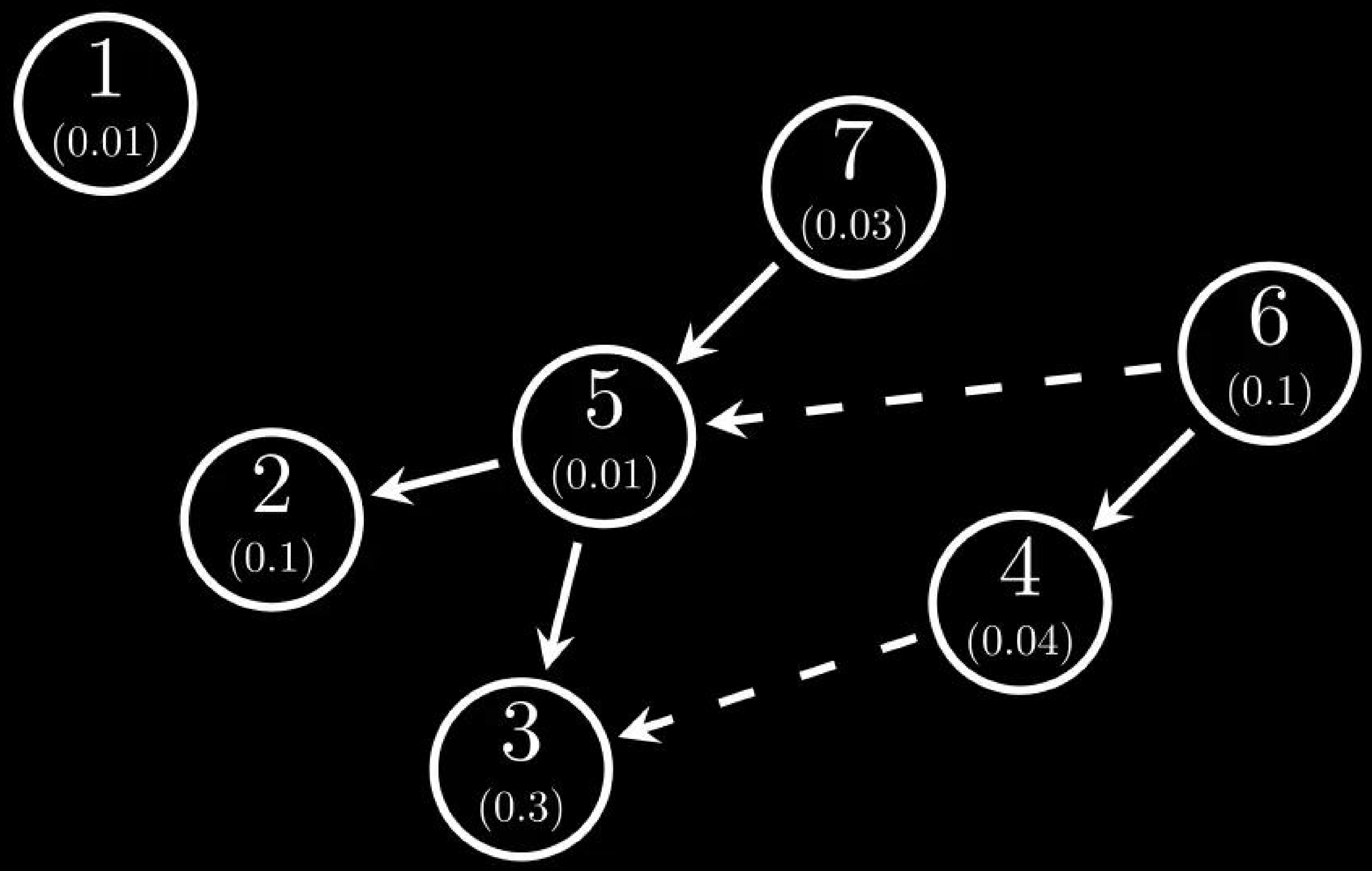
Multiple testing procedure

Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure.*



Multiple testing procedure

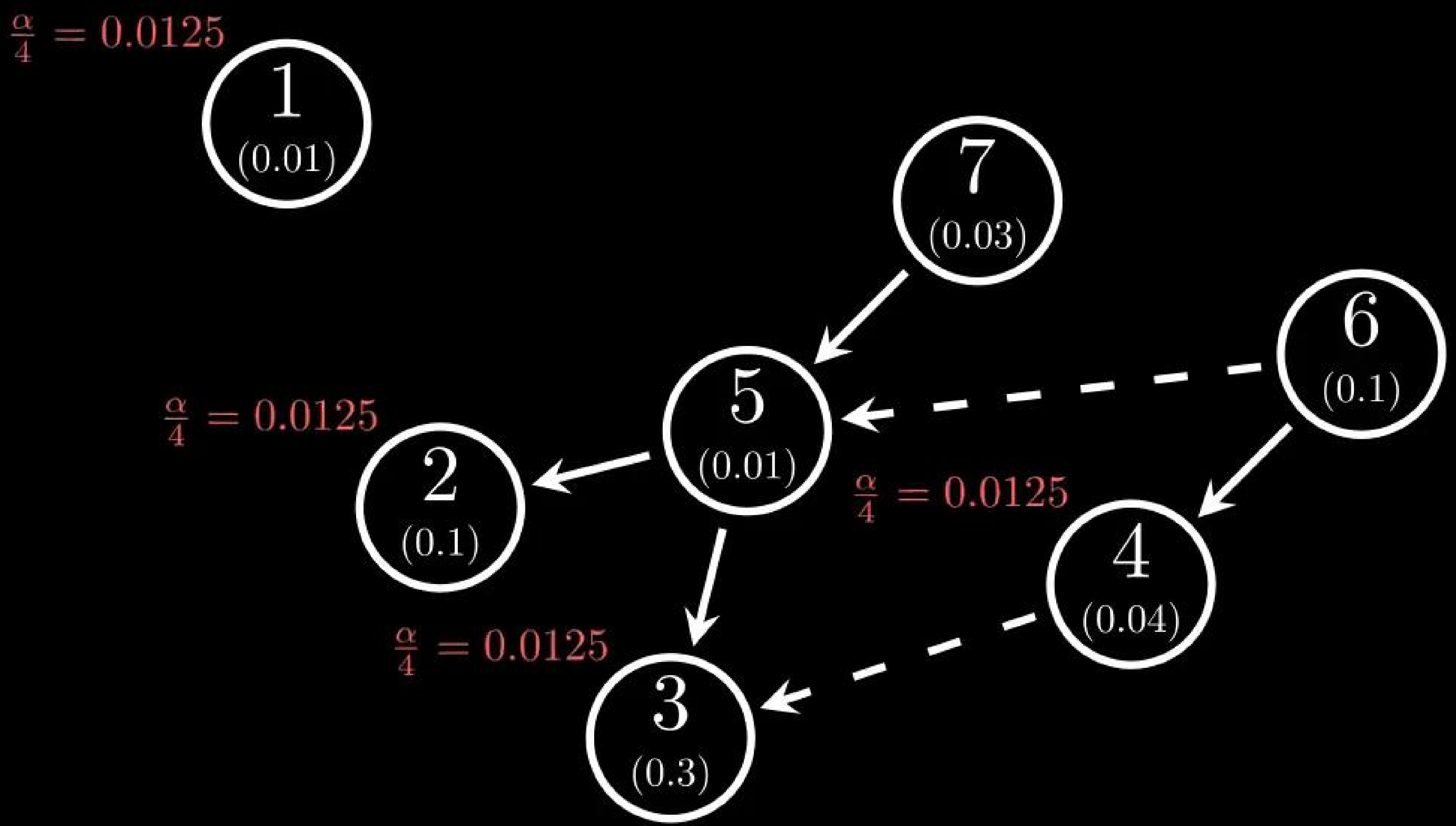
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure*.



Here: $\alpha = 0.05$.

Multiple testing procedure

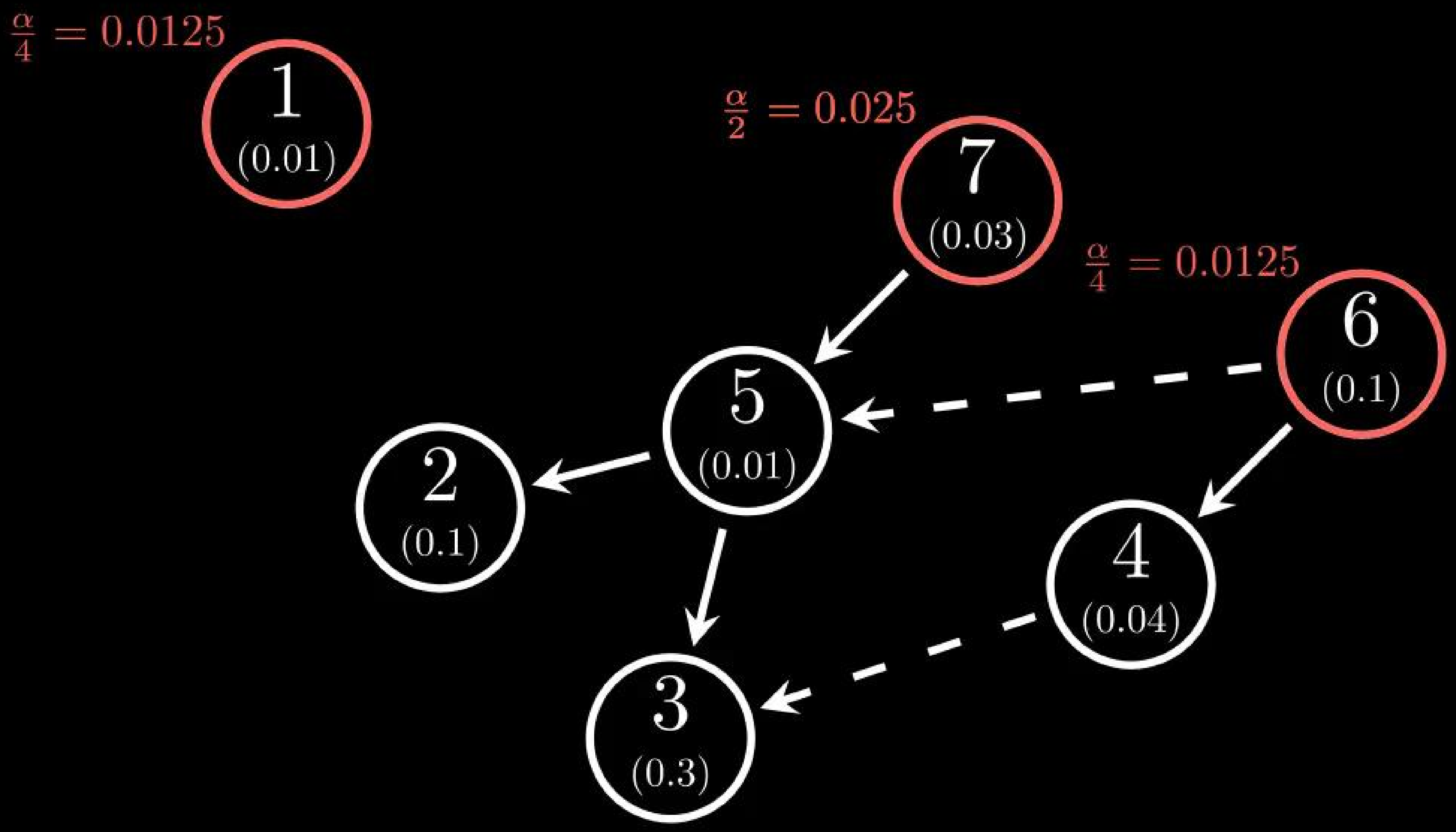
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure.*



Here: $\alpha = 0.05$.

Multiple testing procedure

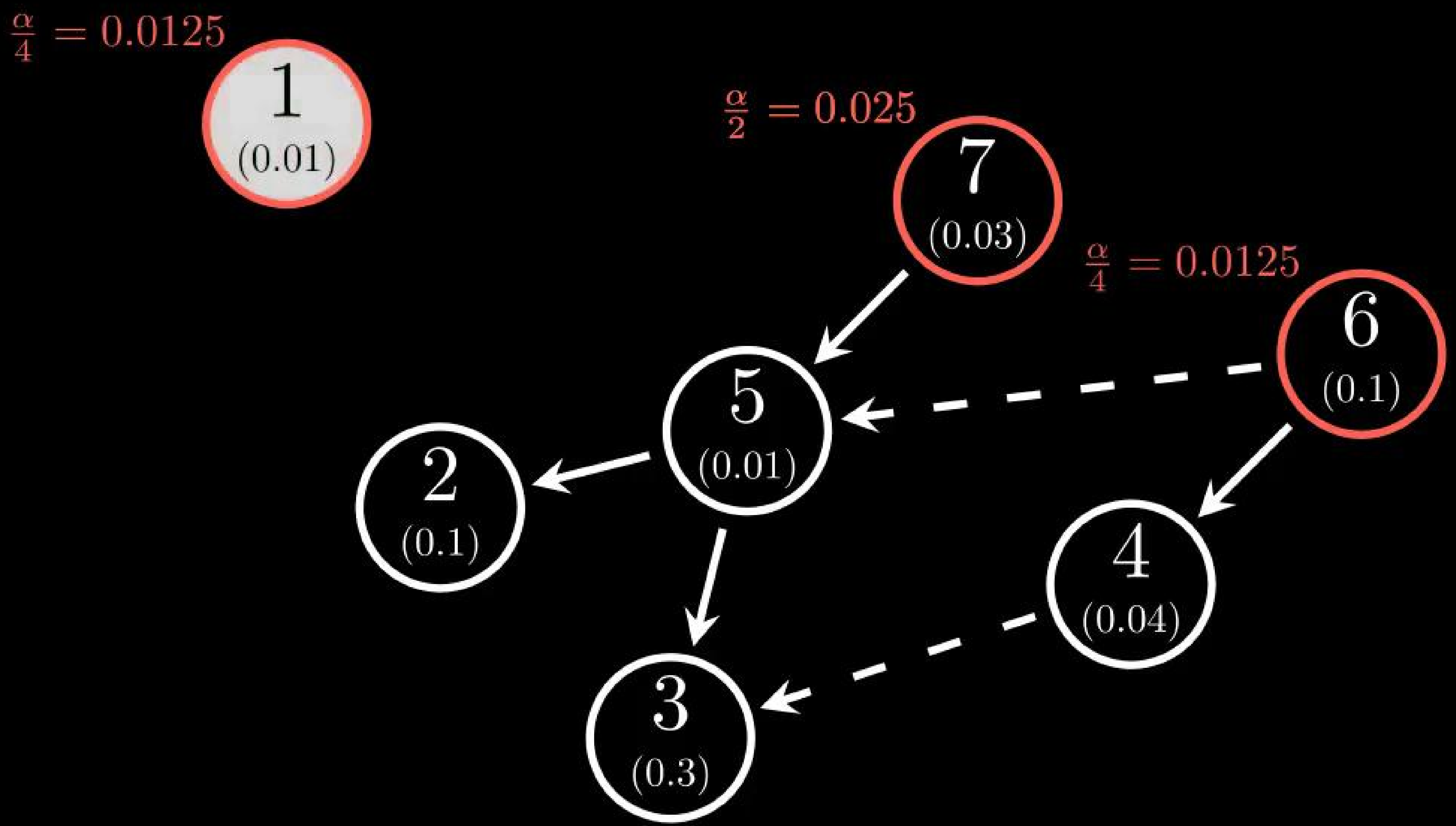
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure.*



Here: $\alpha = 0.05$.

Multiple testing procedure

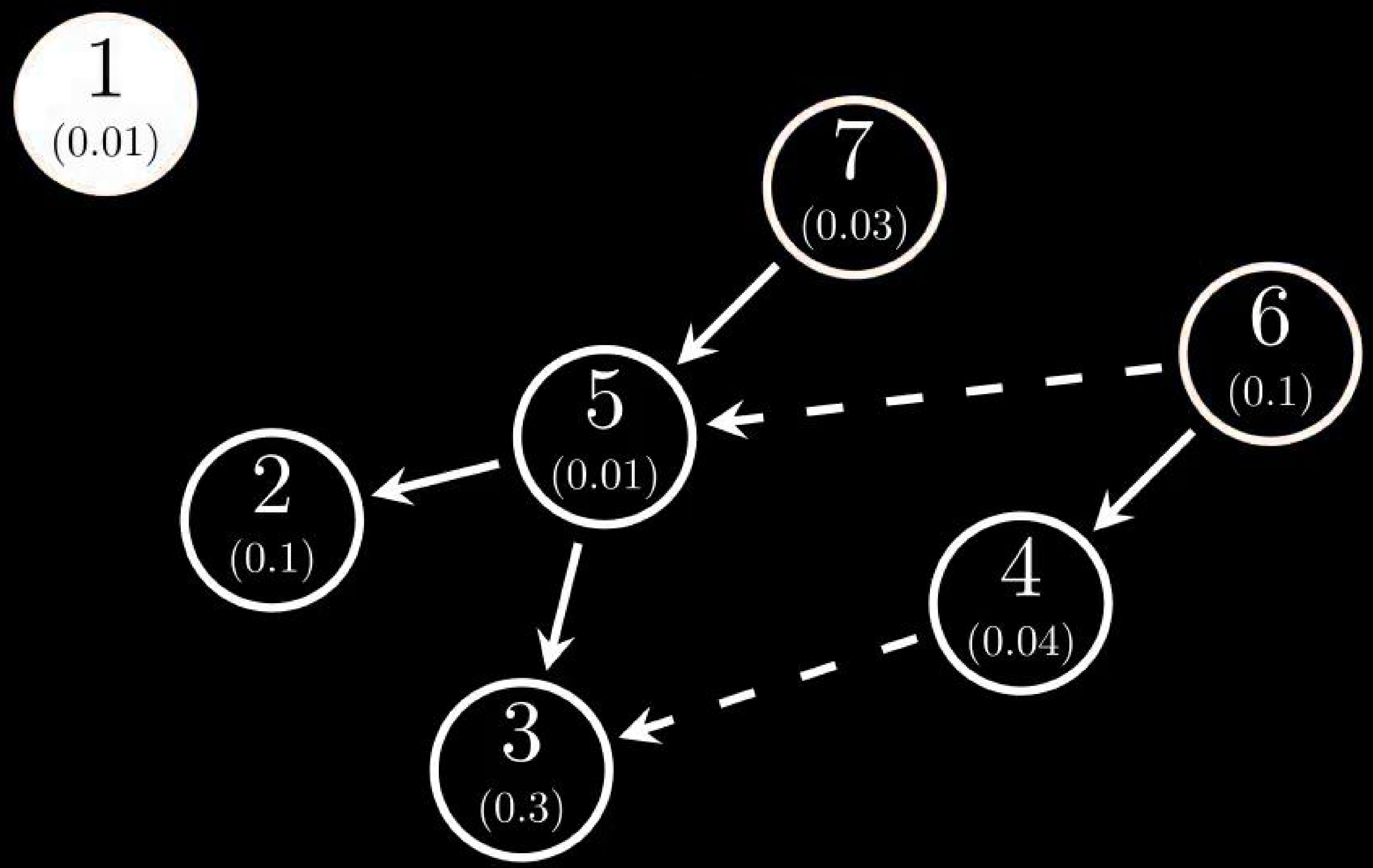
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure.*



Here: $\alpha = 0.05$.

Multiple testing procedure

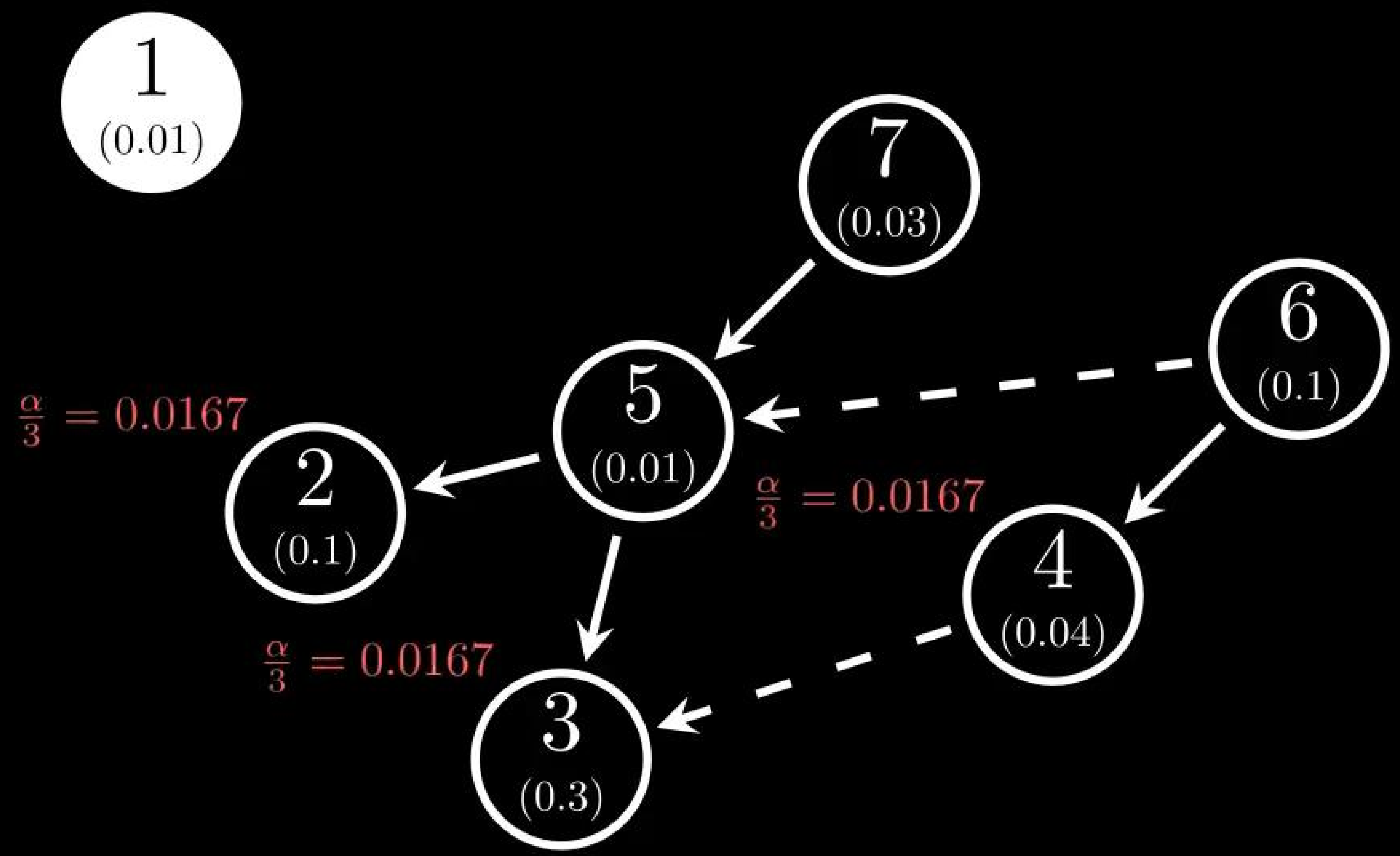
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure*.



Here: $\alpha = 0.05$.

Multiple testing procedure

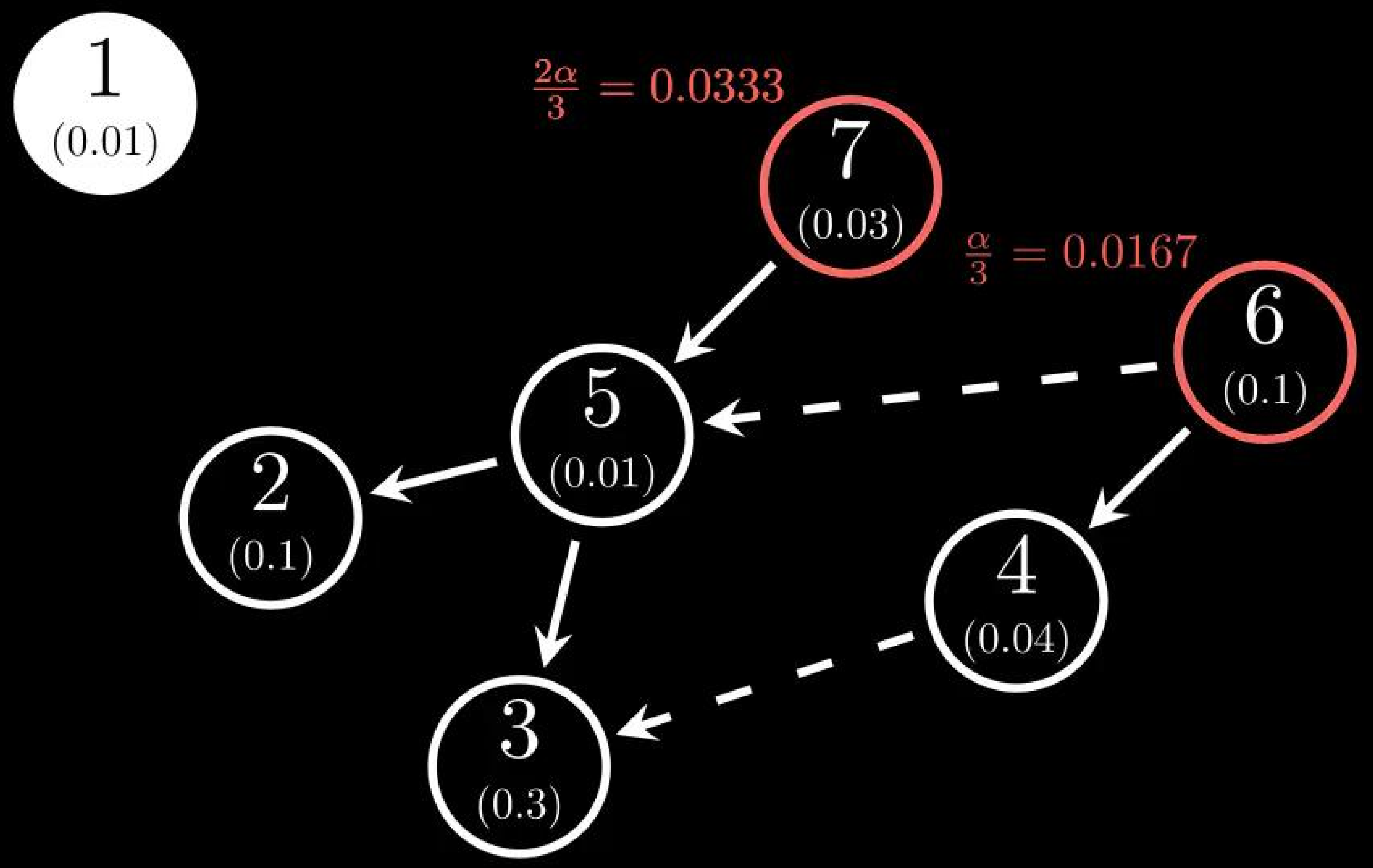
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure.*



Here: $\alpha = 0.05$.

Multiple testing procedure

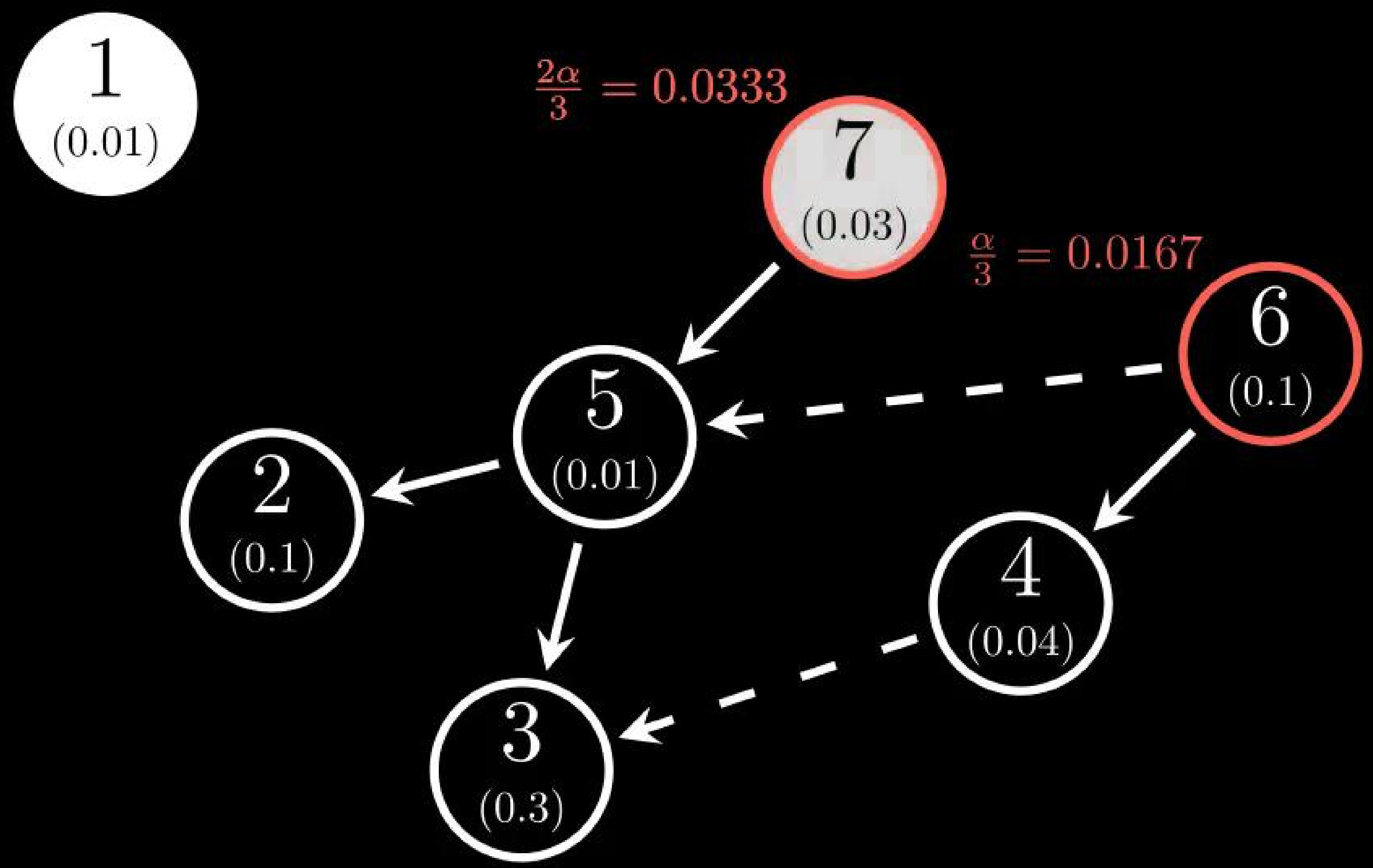
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure.*



Here: $\alpha = 0.05$.

Multiple testing procedure

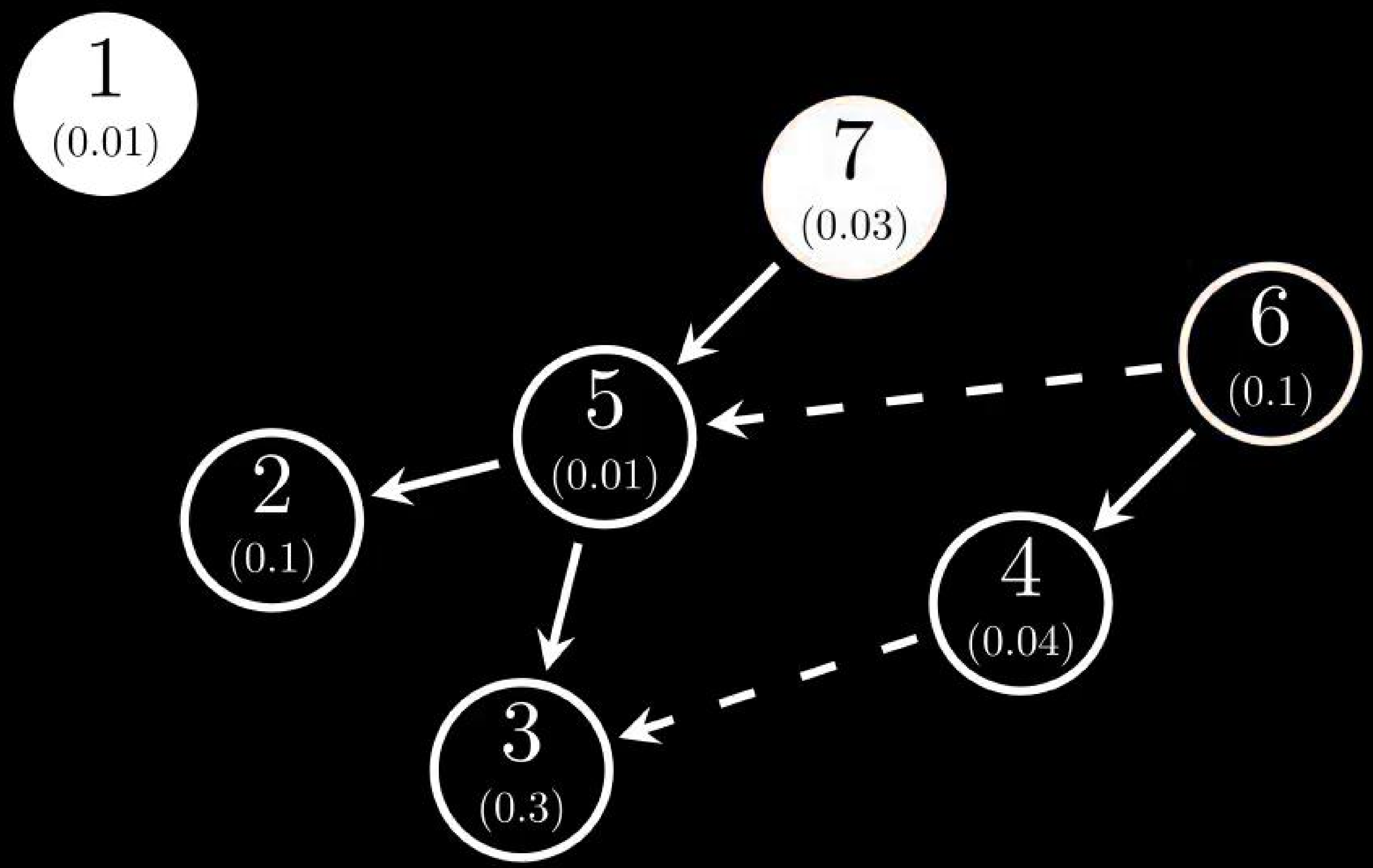
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure.*



Here: $\alpha = 0.05$.

Multiple testing procedure

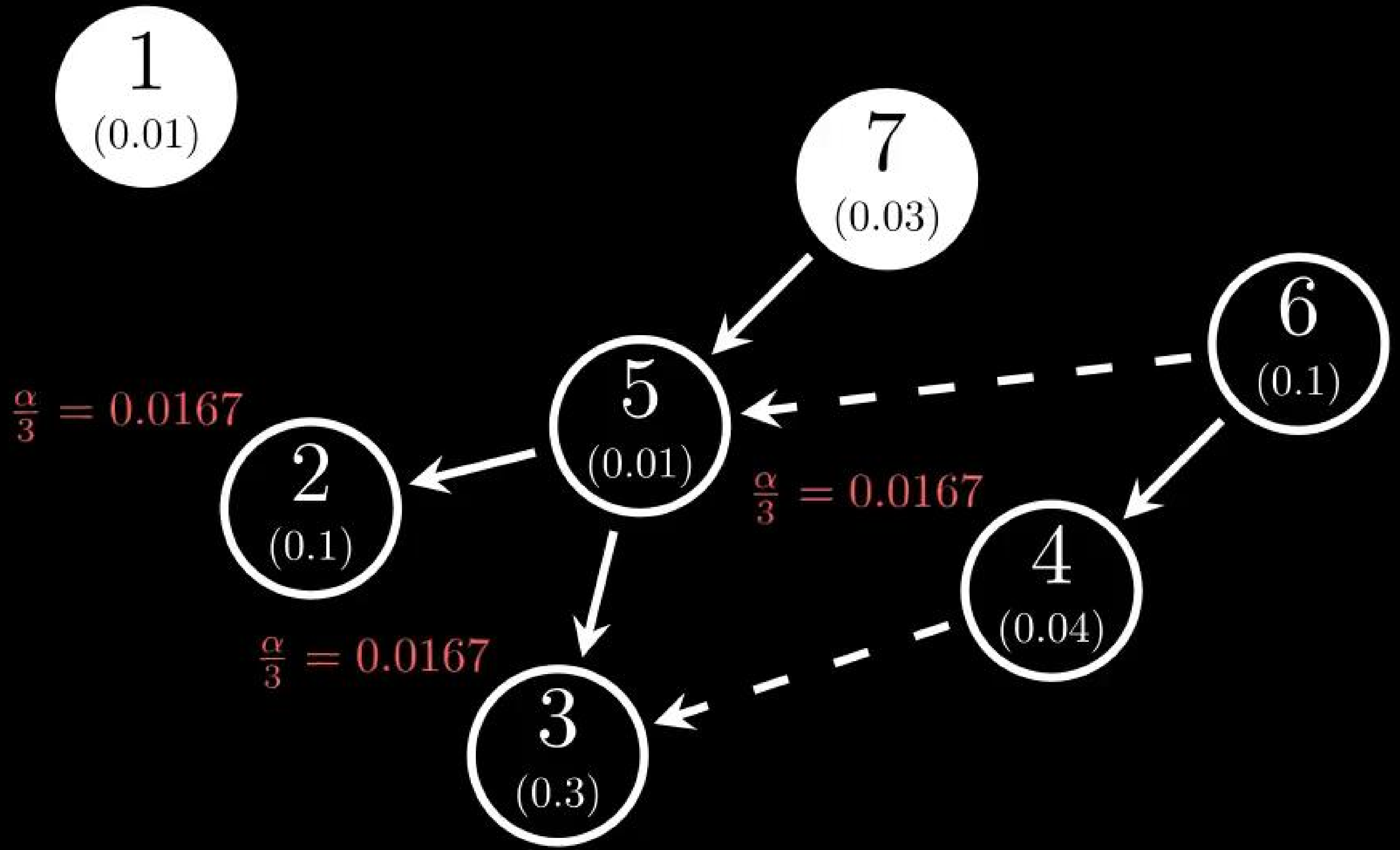
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure.*



Here: $\alpha = 0.05$.

Multiple testing procedure

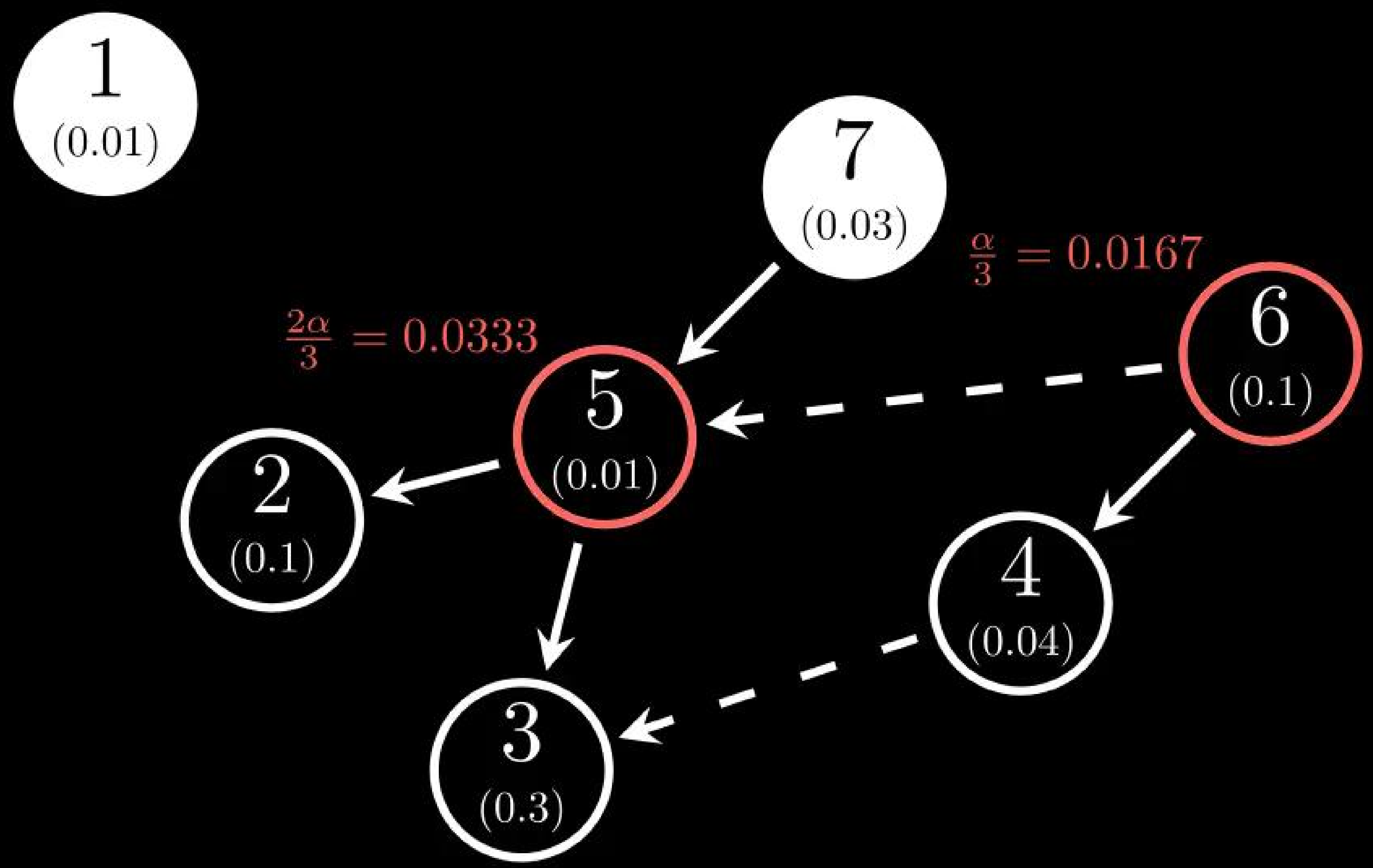
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure.*



Here: $\alpha = 0.05$.

Multiple testing procedure

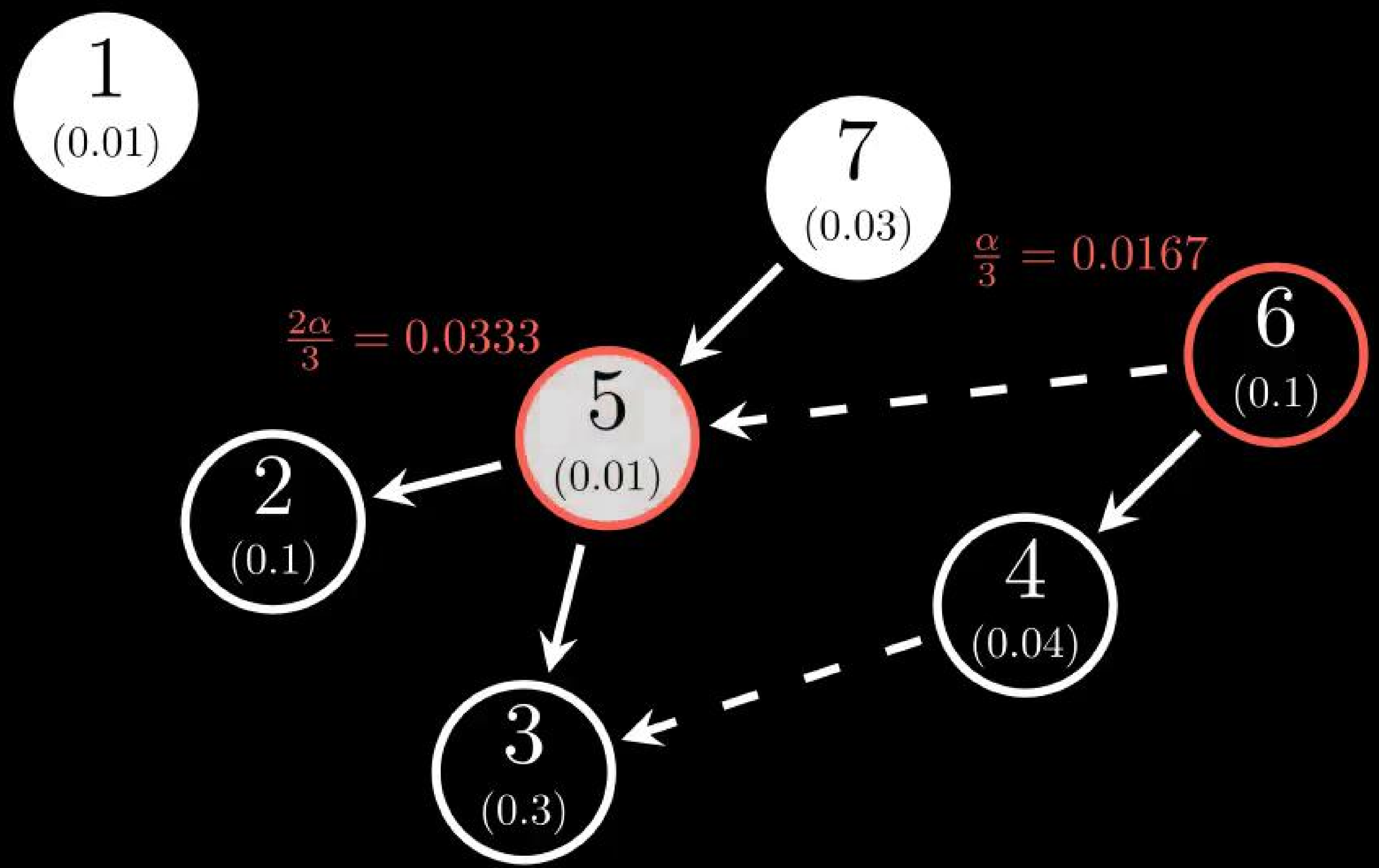
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure.*



Here: $\alpha = 0.05$.

Multiple testing procedure

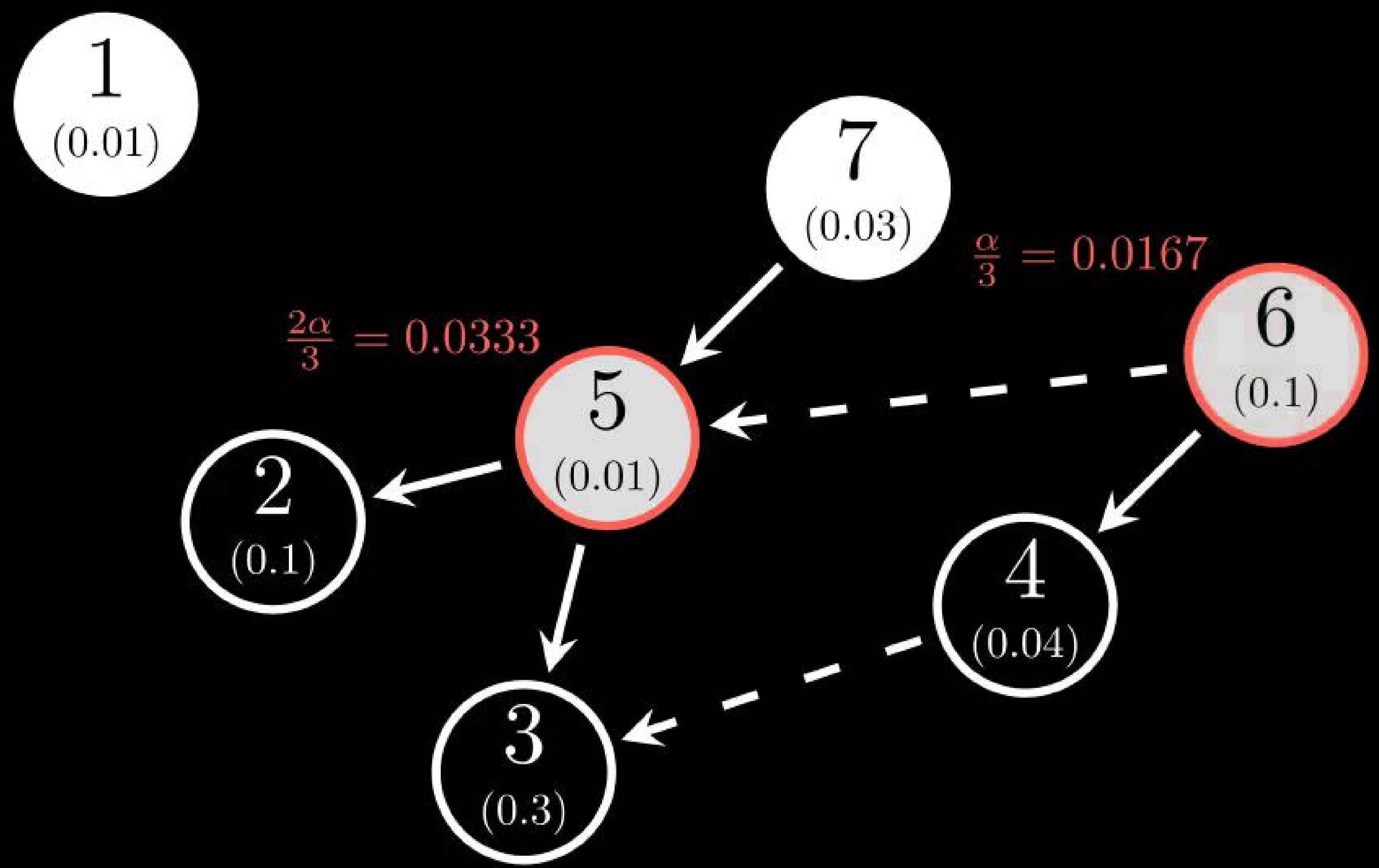
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure.*



Here: $\alpha = 0.05$.

Multiple testing procedure

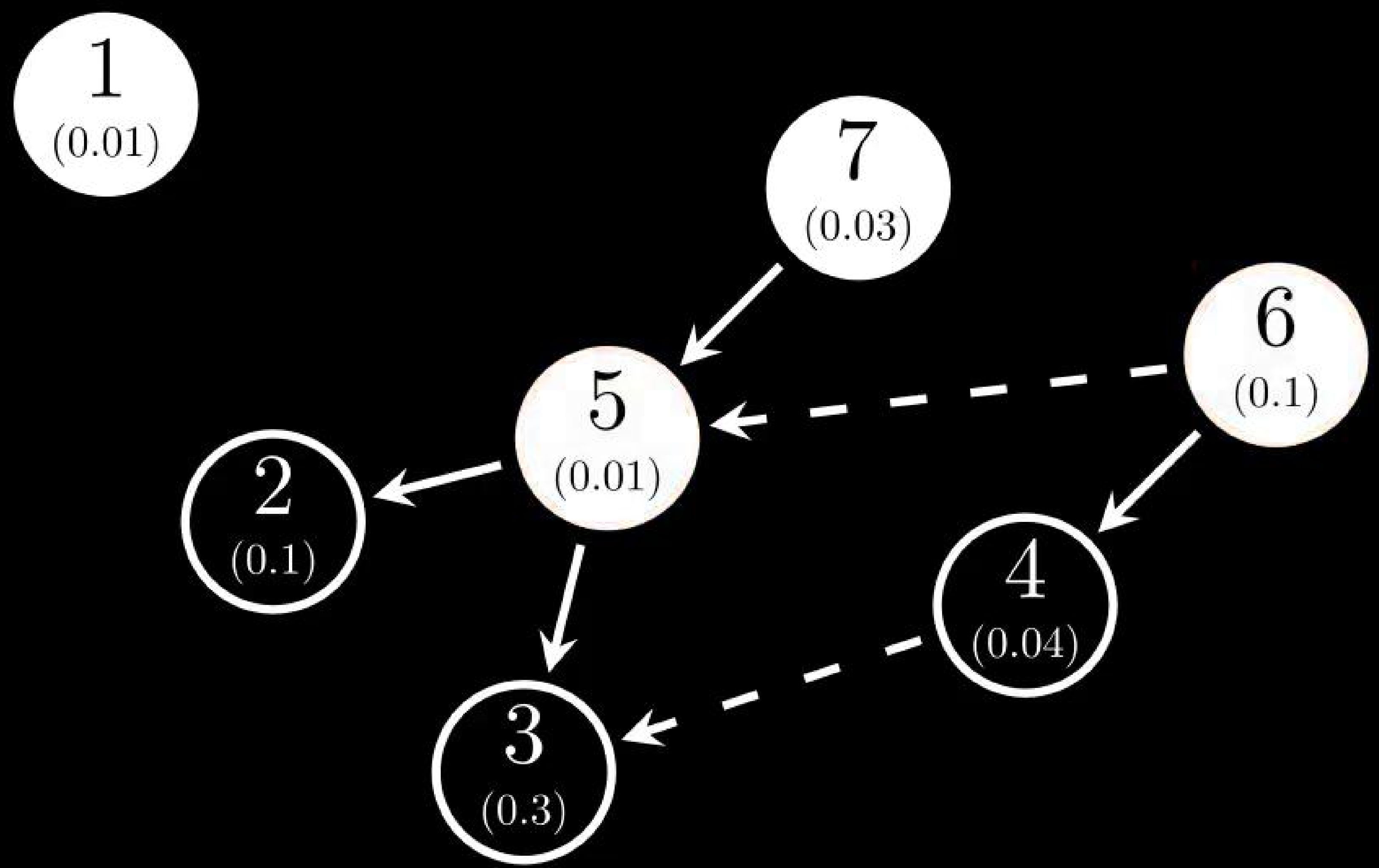
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure.*



Here: $\alpha = 0.05$.

Multiple testing procedure

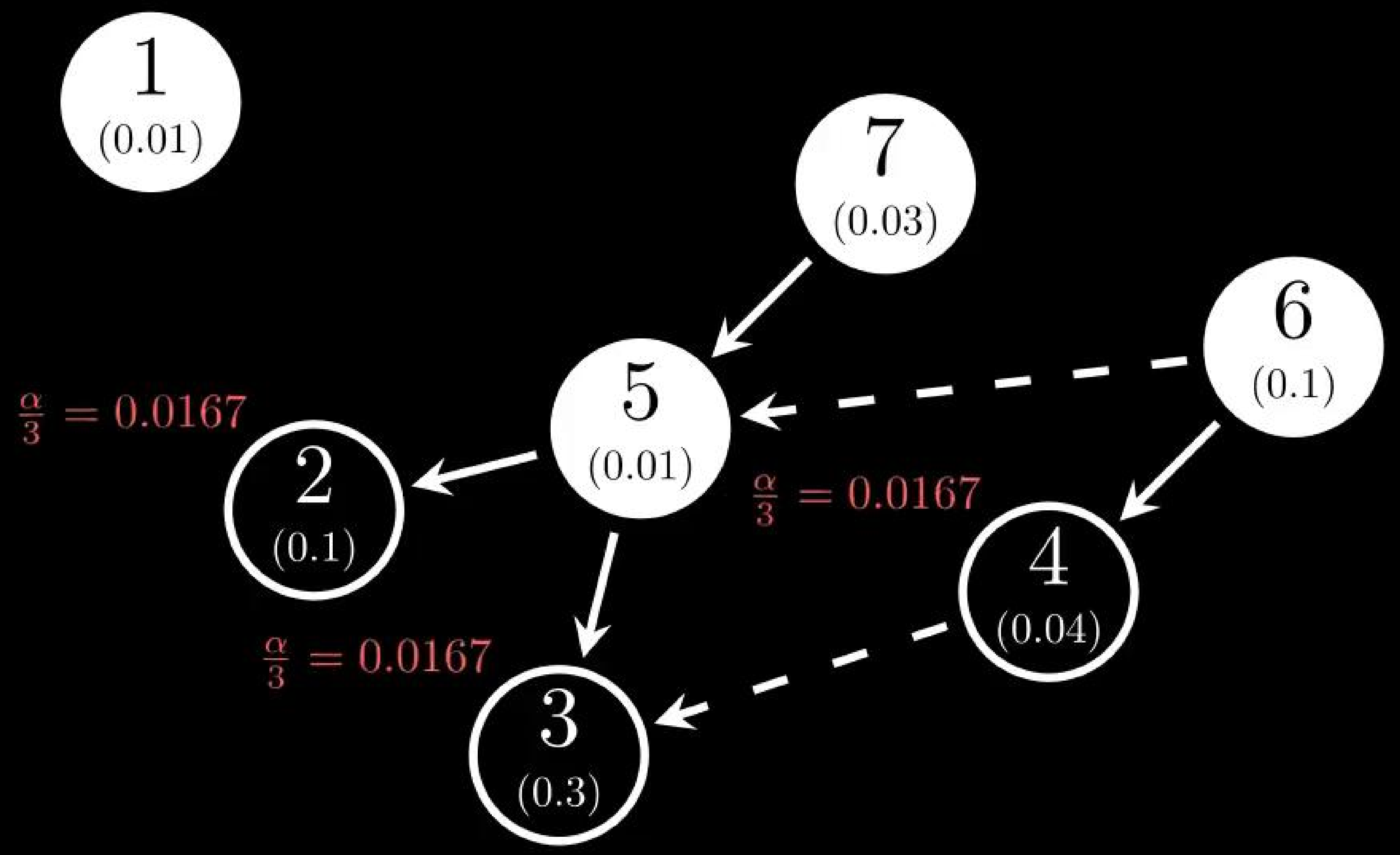
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure.*



Here: $\alpha = 0.05$.

Multiple testing procedure

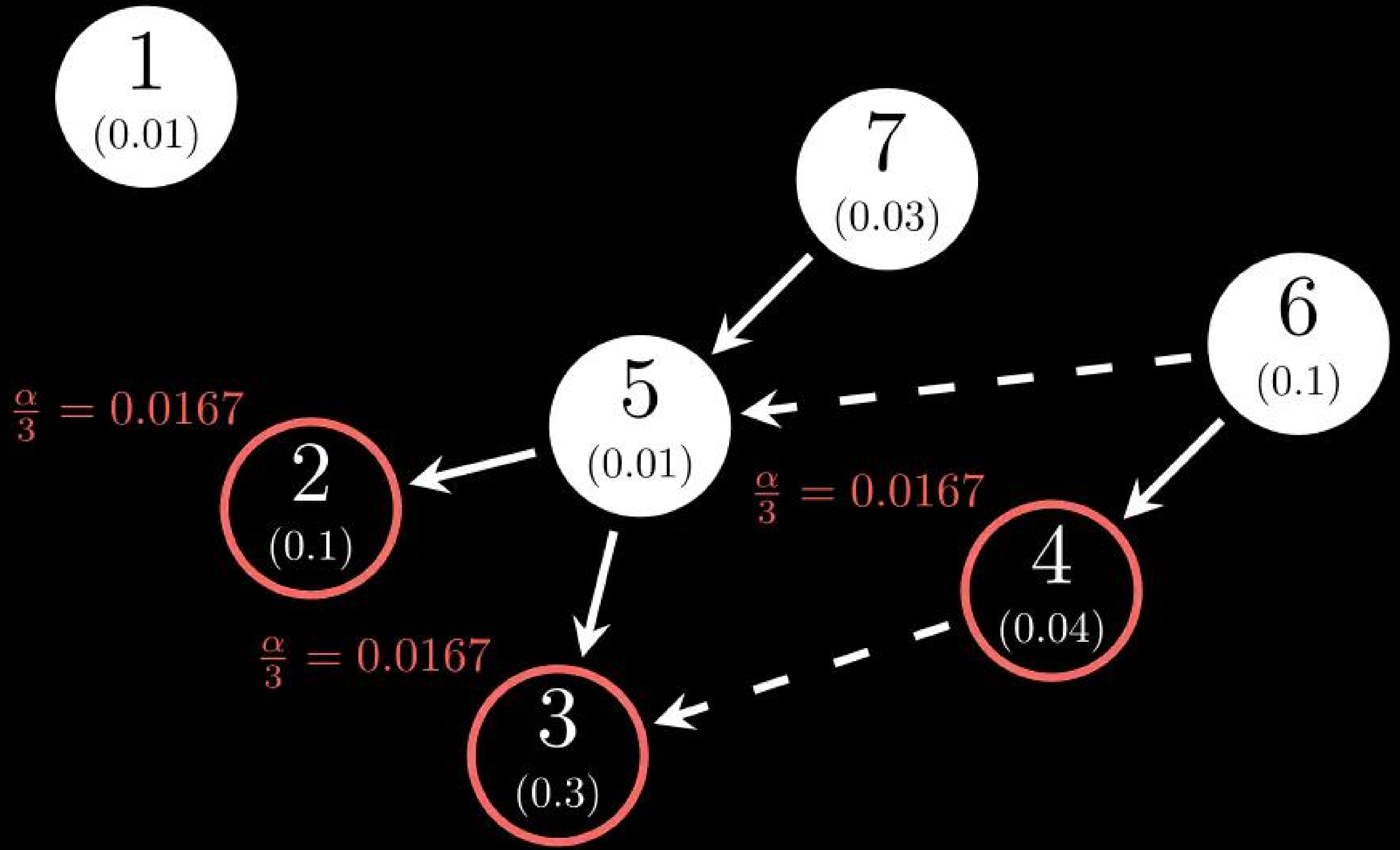
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure.*



Here: $\alpha = 0.05$.

Multiple testing procedure

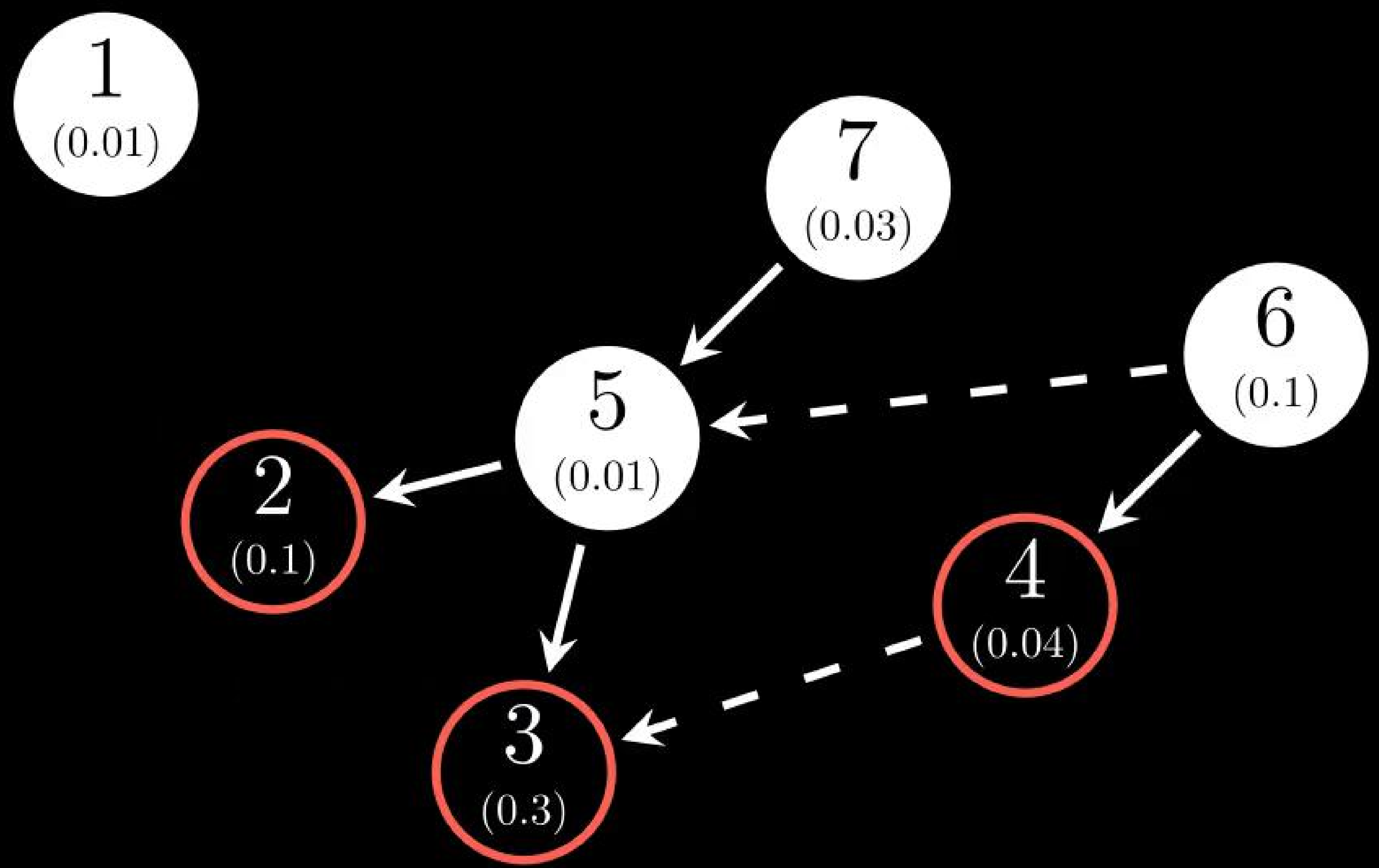
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure.*



Here: $\alpha = 0.05$.

Multiple testing procedure

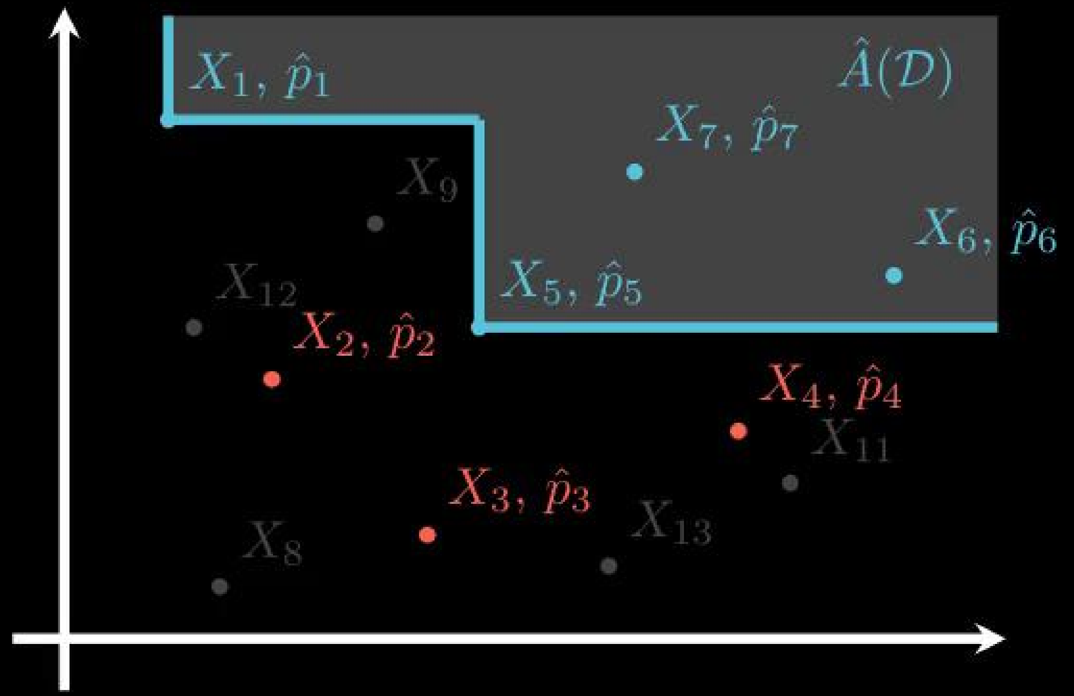
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in [m]$, induce DAG with vertex set $[m]$. We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a *DAG testing procedure.*



Here: $\alpha = 0.05$. The procedure terminates with $\mathcal{R}_\alpha = \{1, 5, 6, 7\}$.

High-level strategy

For $x_0 \in \mathbb{R}^d$, define null hypothesis $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.

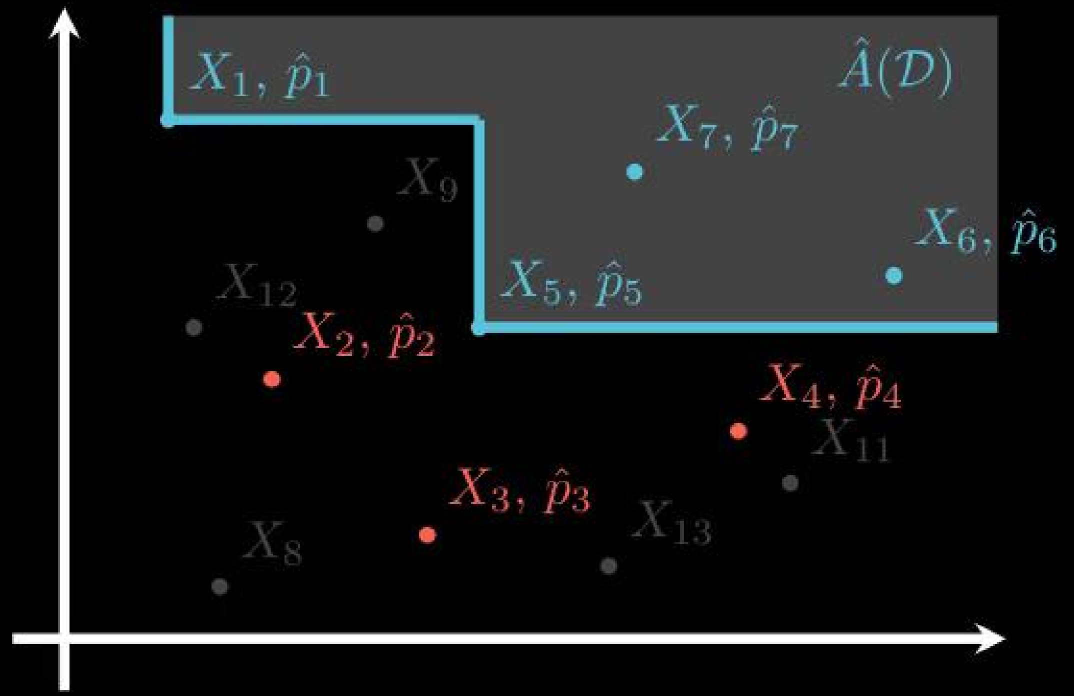


High-level strategy:

1. Subsample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Calculate *p-values* $\hat{p}_i$ for $H_0(X_i)$, $i \in [m]$;
3. Apply a *multiple testing procedure* with FWER-control at level α to reject $\mathcal{R}_\alpha \subseteq [m]$;
4. Output $\hat{A} := \{x \in \mathbb{R}^d : X_\ell \preccurlyeq x \text{ for some } \ell \in \mathcal{R}_\alpha\}$.

High-level strategy

For $x_0 \in \mathbb{R}^d$, define null hypothesis $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.



High-level strategy:

1. Subsample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Calculate *p-values* $\hat{p}_i$ for $H_0(X_i)$, $i \in [m]$;
3. Apply a *multiple testing procedure* with FWER-control at level α to reject $\mathcal{R}_\alpha \subseteq [m]$;
4. Output $\hat{A} := \{x \in \mathbb{R}^d : X_\ell \preccurlyeq x \text{ for some } \ell \in \mathcal{R}_\alpha\}$.

Theoretical guarantees

Write $\hat{A}^{\mathrm{ISS}} \equiv \hat{A}^{\mathrm{ISS}}_{\sigma,\tau,\alpha,m}(\mathcal{D})$ for the resulting selected subgroup.
Recall that $\mathcal{X}_\tau(\eta) := \big\{x \in \mathbb{R}^d : \eta(x) \geq \tau\big\}$ is the τ-superlevel set of η.

Theoretical guarantees

Write $\hat{A}^{\mathrm{ISS}} \equiv \hat{A}^{\mathrm{ISS}}_{\sigma,\tau,\alpha,m}(\mathcal{D})$ for the resulting selected subgroup.
Recall that $\mathcal{X}_\tau(\eta) := \{x \in \mathbb{R}^d : \eta(x) \geq \tau\}$ is the τ-superlevel set of η.

Theorem. For any $n \geq 1$, $m \in [n]$, $\alpha \in (0,1)$, $\sigma > 0$, we have for all $P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma)$:

$$\mathbb{P}\big(\forall x \in \hat{A}^{\mathrm{ISS}} : \eta(x) \geq \tau \mid X_1, \ldots, X_n\big) \equiv \mathbb{P}\big(\hat{A}^{\mathrm{ISS}} \subseteq \mathcal{X}_\tau(\eta) \mid X_1, \ldots, X_n\big) \geq 1 - \alpha.$$

Theoretical guarantees

Write $\hat{A}^{\mathrm{ISS}} \equiv \hat{A}^{\mathrm{ISS}}_{\sigma,\tau,\alpha,m}(\mathcal{D})$ for the resulting selected subgroup.
Recall that $\mathcal{X}_\tau(\eta) := \{x \in \mathbb{R}^d : \eta(x) \geq \tau\}$ is the τ-superlevel set of η.

Theorem. For any $n \geq 1$, $m \in [n]$, $\alpha \in (0,1)$, $\sigma > 0$, we have for all $P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma)$:

$$\mathbb{P}\big(\forall x \in \hat{A}^{\mathrm{ISS}} : \eta(x) \geq \tau \;\big|\; X_1, \ldots, X_n\big) \equiv \mathbb{P}\big(\hat{A}^{\mathrm{ISS}} \subseteq \mathcal{X}_\tau(\eta) \;\big|\; X_1, \ldots, X_n\big) \geq 1 - \alpha.$$

NB: We only use the monotonicity assumption through requiring $\mathcal{X}_\tau(\eta)$ to be an *upper set*.

Theoretical guarantees

Write $\hat{A}^{\mathrm{ISS}} \equiv \hat{A}^{\mathrm{ISS}}_{\sigma,\tau,\alpha,m}(\mathcal{D})$ for the resulting selected subgroup.
Recall that $\mathcal{X}_\tau(\eta) := \{x \in \mathbb{R}^d : \eta(x) \geq \tau\}$ is the τ-superlevel set of η.

Theorem. For any $n \geq 1$, $m \in [n]$, $\alpha \in (0,1)$, $\sigma > 0$, we have for all $P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma)$:

$$\mathbb{P}\big(\forall x \in \hat{A}^{\mathrm{ISS}} : \eta(x) \geq \tau \;\big|\; X_1,\ldots,X_n\big) \equiv \mathbb{P}\big(\hat{A}^{\mathrm{ISS}} \subseteq \mathcal{X}_\tau(\eta) \;\big|\; X_1,\ldots,X_n\big) \geq 1-\alpha.$$

NB: We only use the monotonicity assumption through requiring $\mathcal{X}_\tau(\eta)$ to be an *upper set.*

Theorem. $\hat{A}^{\mathrm{ISS}}$ is minimax optimal (in terms of expected regret) across a natural subclass of $\mathcal{P}_{\mathrm{Mon},d}(\sigma)$.

Application

Study/data: AIDS Clinical Trials Group Study 175 (Hammer et al., 1996) — focus on those 532 patients treated with zidovudine. Median follow-up duration of 143 weeks.

Application

Study/data: AIDS Clinical Trials Group Study 175 (Hammer et al., 1996) — focus on those 532 patients treated with zidovudine. Median follow-up duration of 143 weeks.

Primary endpoint in study: reduction of the CD4 cell count by 50%, development of AIDS, or death.

Application

Study/data: AIDS Clinical Trials Group Study 175 (Hammer et al., 1996) — focus on those 532 patients treated with zidovudine. Median follow-up duration of 143 weeks.

Primary endpoint in study: reduction of the CD4 cell count by 50%, development of AIDS, or death.

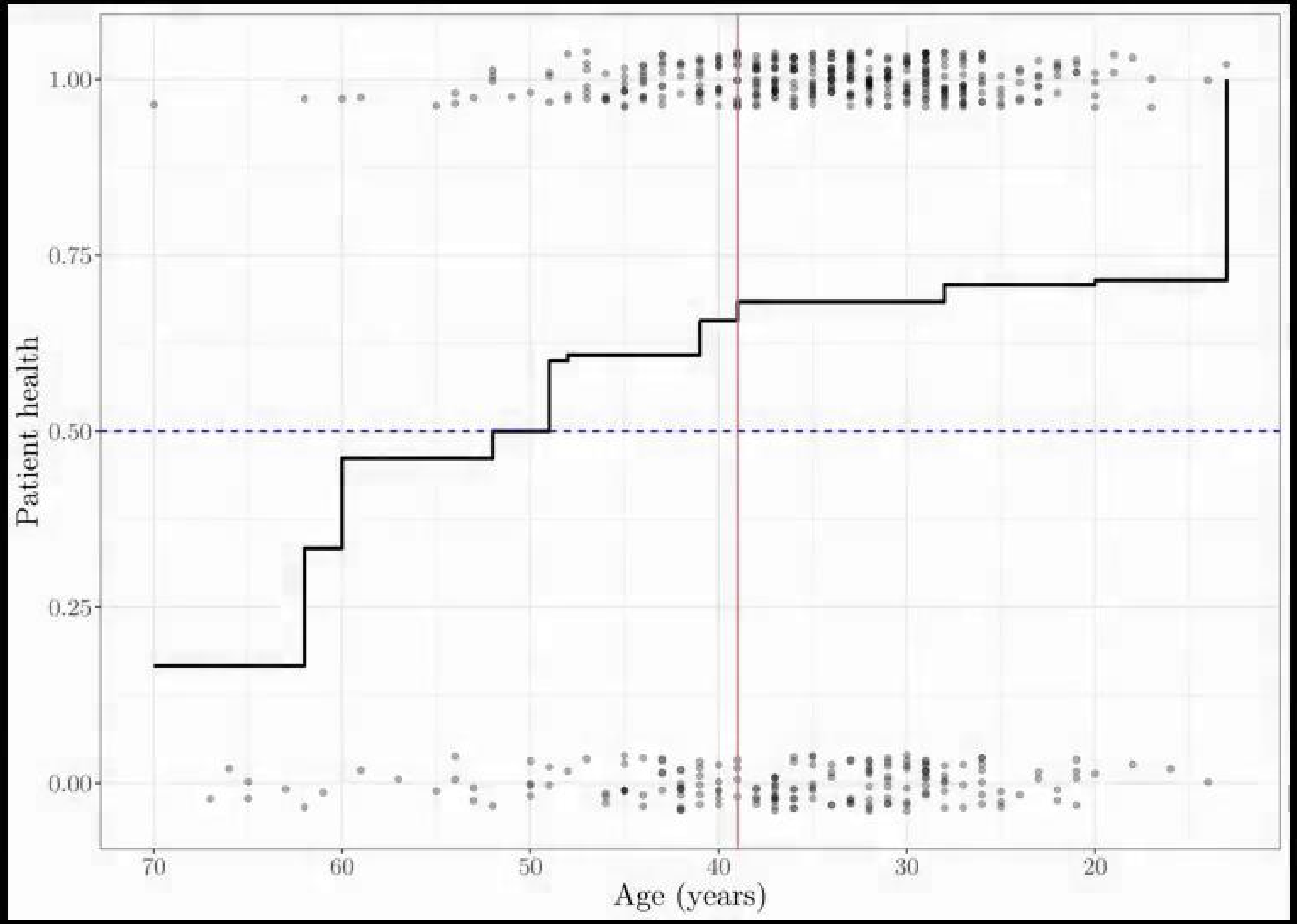
Application of our method: we set $Y_i = \mathbb{1}\{\text{patient } i \text{ reaching primary endpoint not reported}\}$, turning this into a classification setting, and let $\alpha = 0.05$, $\tau = 1/2$.

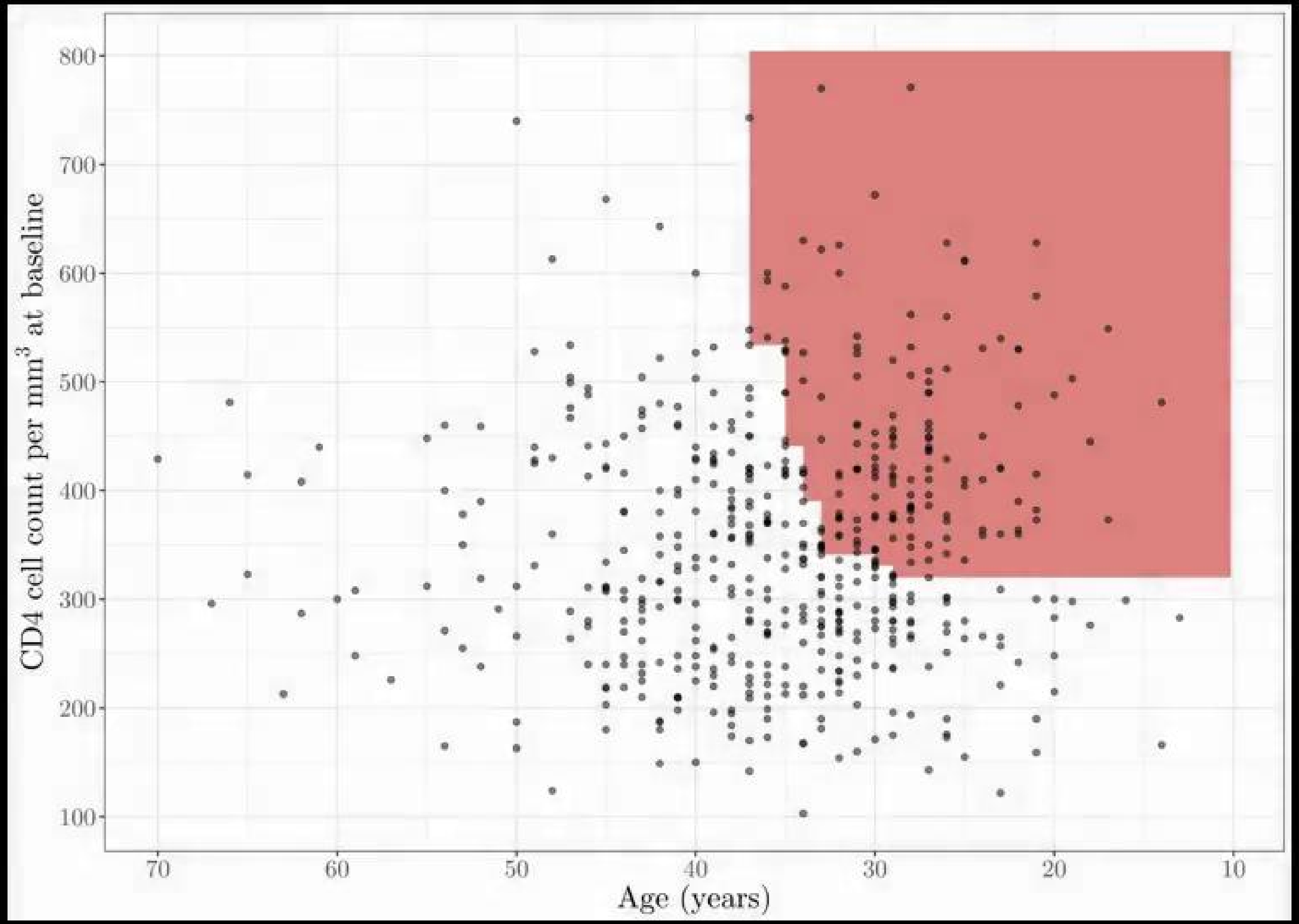
Application

Study/data: AIDS Clinical Trials Group Study 175 (Hammer et al., 1996) — focus on those 532 patients treated with zidovudine. Median follow-up duration of 143 weeks.

Primary endpoint in study: reduction of the CD4 cell count by 50%, development of AIDS, or death.

Application of our method: we set $Y_i = \mathbb{1}\{\text{patient } i \text{ reaching primary endpoint not reported}\}$, turning this into a classification setting, and let $\alpha = 0.05$, $\tau = 1/2$.





Heterogeneous treatment effects: theory

Heterogeneous treatment effects: theory

Suppose we observe n independent copies of $(X, T, \tilde{Y})$, where $T \in \{0, 1\}$ is a treatment indicator and $\tilde{Y}$ is the response.

Heterogeneous treatment effects: theory

Suppose we observe n independent copies of $(X, T, \tilde{Y})$, where $T \in \{0, 1\}$ is a treatment indicator and $\tilde{Y}$ is the response.

For $\ell \in \{0, 1\}$, let $\tilde{\eta}^{\ell}(x) := \mathbb{E}(\tilde{Y} | X = x, T = \ell)$, and define the *heterogeneous treatment effect* $\eta(x) := \tilde{\eta}^1(x) - \tilde{\eta}^0(x)$.

Heterogeneous treatment effects: theory

Suppose we observe n independent copies of $(X, T, \tilde{Y})$, where $T \in \{0, 1\}$ is a treatment indicator and $\tilde{Y}$ is the response.

For $\ell \in \{0, 1\}$, let $\tilde{\eta}^{\ell}(x) := \mathbb{E}(\tilde{Y}|X = x, T = \ell)$, and define the *heterogeneous treatment effect* $\eta(x) := \tilde{\eta}^1(x) - \tilde{\eta}^0(x)$.

Write $\pi(x) := \mathbb{P}(T = 1|X = x)$ for the *propensity score*, and consider the *inverse propensity weighted response*

$$Y := \frac{T - \pi(X)}{\pi(X)\bigl(1 - \pi(X)\bigr)} \cdot \tilde{Y},$$

so that $\mathbb{E}(Y|X = x) = \eta(x)$ for all $x \in \mathbb{R}^d$.

Heterogeneous treatment effects: theory

Suppose we observe n independent copies of $(X, T, \tilde{Y})$, where $T \in \{0, 1\}$ is a treatment indicator and $\tilde{Y}$ is the response.

For $\ell \in \{0, 1\}$, let $\tilde{\eta}^\ell(x) := \mathbb{E}(\tilde{Y}|X = x, T = \ell)$, and define the *heterogeneous treatment effect* $\eta(x) := \tilde{\eta}^1(x) - \tilde{\eta}^0(x)$.

Write $\pi(x) := \mathbb{P}(T = 1|X = x)$ for the *propensity score*, and consider the *inverse propensity weighted response*

$$Y := \frac{T - \pi(X)}{\pi(X)\bigl(1 - \pi(X)\bigr)} \cdot \tilde{Y},$$

so that $\mathbb{E}(Y|X = x) = \eta(x)$ for all $x \in \mathbb{R}^d$.

Hence, for P such that $\mathcal{D} = \bigl((X_1, Y_1), \ldots, (X_n, Y_n)\bigr) \sim P^n$, we have $\tilde{\eta}^1(x) \geq \tilde{\eta}^0(x) + \tau$ for all $x \in \hat{A}^{\mathrm{ISS}}$ with probability at least $1 - \alpha$ whenever $P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma)$.

Heterogeneous treatment effects: application

Heterogeneous treatment effects: application

Study/data: AIDS Clinical Trials Group Study 175 (Hammer et al., 1996). Treatments: zidovudine only ($T = 0$, $n = 532$) versus zidovudine together with zalcitabine ($T = 1$, $n = 524$).

Heterogeneous treatment effects: application

Study/data: AIDS Clinical Trials Group Study 175 (Hammer et al., 1996). Treatments: zidovudine only ($T = 0$, $n = 532$) versus zidovudine together with zalcitabine ($T = 1$, $n = 524$).

Application of our method: Let $\tilde{Y}$ be the difference in CD4 cell count between week 20 and week 0.
Calculate inverse propensity weighted responses Y_i with $\pi(x) \equiv 1/2$.
Apply our method with $\tau = 1/2$ (and $\alpha = 0.05$) to transformed responses $\mathbb{1}\{Y_i \geq 0\}$.
Reasoning: $\mathbb{P}(Y_i \geq 0 \mid X_i = x) \geq 1/2 \iff \eta(x) \geq 0$ (under symmetry).

Heterogeneous treatment effects: application

Study/data: AIDS Clinical Trials Group Study 175 (Hammer et al., 1996). Treatments: zidovudine only ($T = 0$, $n = 532$) versus zidovudine together with zalcitabine ($T = 1$, $n = 524$).

Application of our method: Let $\tilde{Y}$ be the difference in CD4 cell count between week 20 and week 0.
Calculate inverse propensity weighted responses Y_i with $\pi(x) \equiv 1/2$.
Apply our method with $\tau = 1/2$ (and $\alpha = 0.05$) to transformed responses $\mathbb{1}\{Y_i \geq 0\}$.
Reasoning: $\mathbb{P}(Y_i \geq 0 \mid X_i = x) \geq 1/2 \iff \eta(x) \geq 0$ (under symmetry).

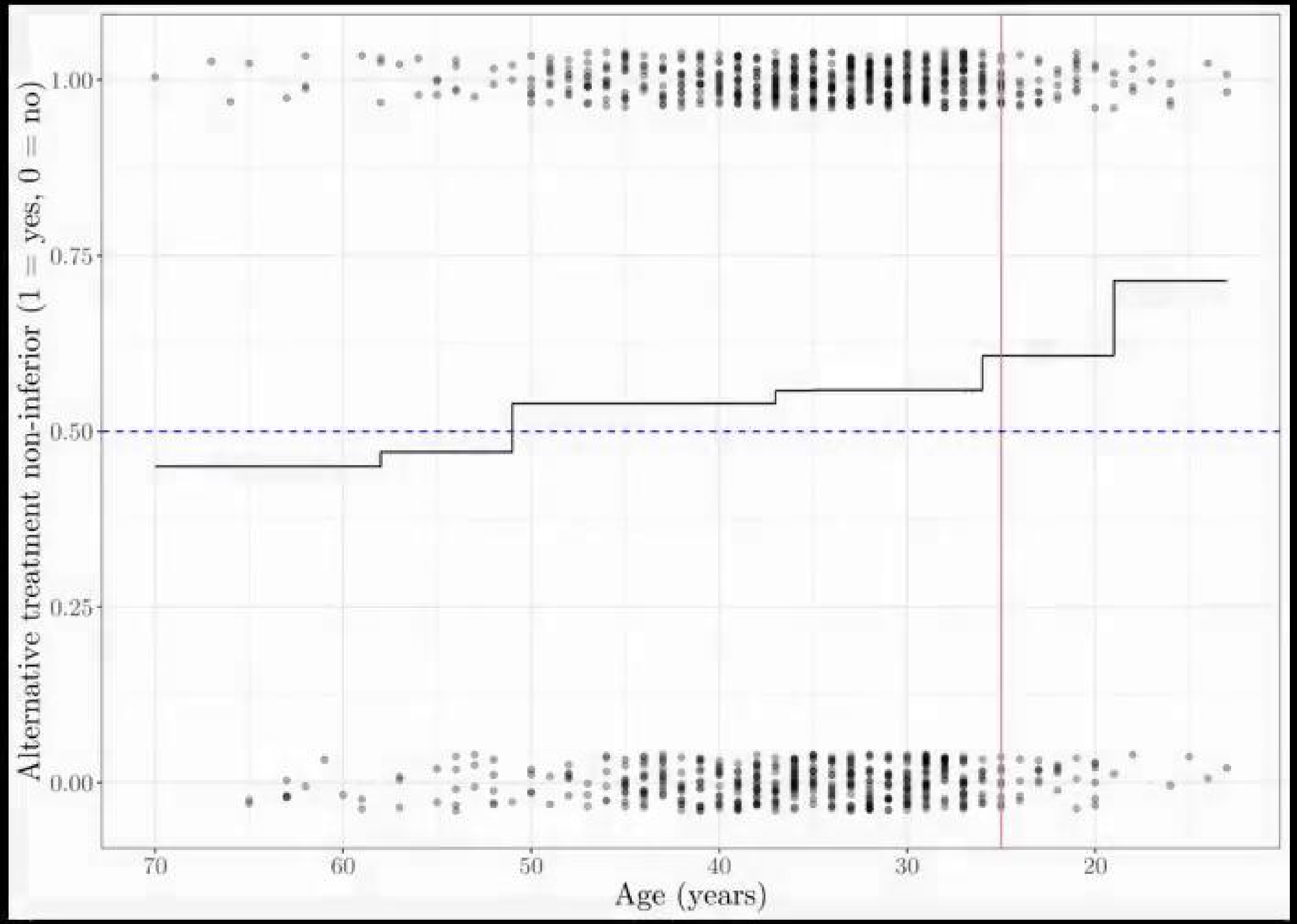
$\implies$ Non-inferiority claim.

Heterogeneous treatment effects: application

Study/data: AIDS Clinical Trials Group Study 175 (Hammer et al., 1996). Treatments: zidovudine only ($T = 0$, $n = 532$) versus zidovudine together with zalcitabine ($T = 1$, $n = 524$).

Application of our method: Let $\tilde{Y}$ be the difference in CD4 cell count between week 20 and week 0.
Calculate inverse propensity weighted responses Y_i with $\pi(x) \equiv 1/2$.
Apply our method with $\tau = 1/2$ (and $\alpha = 0.05$) to transformed responses $\mathbb{1}\{Y_i \geq 0\}$.
Reasoning: $\mathbb{P}(Y_i \geq 0 \mid X_i = x) \geq 1/2 \iff \eta(x) \geq 0$ (under symmetry).

$\implies$ Non-inferiority claim.



Extensions in the paper

- Adaptation to unknown variance of Gaussian errors.

Extensions in the paper

- Adaptation to unknown variance of Gaussian errors.

- Bounded responses such as in classification.

Extensions in the paper

- Adaptation to unknown variance of Gaussian errors.
- Bounded responses such as in classification.
- Heavy tails through isotonic quantile regression.

Extensions in the paper

- Adaptation to unknown variance of Gaussian errors.
- Bounded responses such as in classification.
- Heavy tails through isotonic quantile regression.
- Extensive simulations and further applications.

Extensions in the paper

- Adaptation to unknown variance of Gaussian errors.
- Bounded responses such as in classification.
- Heavy tails through isotonic quantile regression.
- Extensive simulations and further applications.
- `R`-package `ISS` on CRAN.

Take-home messages

Subgroup selection with strong Type I error guarantees is possible in the isotonic regression setting.

Take-home messages

Subgroup selection with strong Type I error guarantees is possible in the isotonic regression setting.

Our procedure is computationally feasible and minimax-optimal up to poly-logarithmic factors.

Take-home messages

Subgroup selection with strong Type I error guarantees is possible in the isotonic regression setting.

Our procedure is computationally feasible and minimax-optimal up to poly-logarithmic factors.

In common situations, no smoothing-parameters have to be specified.

References and acknowledgement

Duan, B., Ramdas, A., Balakrishnan, S., and Wasserman, L. (2020). Interactive martingale tests for the global null. *Electronic Journal of Statistics*, 14(2):4489–4551.

Goeman, J. J. and Solari, A. (2010). The sequential rejection principle of familywise error control. *The Annals of Statistics*, 38(6):3782–3810.

Hammer, S. M., Katzenstein, D. A., Hughes, M. D., Gundacker, H., Schooley, R. T., Haubrich, R. H., Henry, W. K., Lederman, M. M., Phair, J. P., Niu, M., et al. (1996). A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *New England Journal of Medicine*, 335(15):1081–1090.

Howard, S. R., Ramdas, A., McAuliffe, J., and Sekhon, J. (2021). Time-uniform, nonparametric, nonasymptotic confidence sequences. *The Annals of Statistics*, 49:1055–1080.

Meijer, R. J. and Goeman, J. J. (2015). A multiple testing method for hypotheses structured in a directed acyclic graph. *Biometrical Journal*, 57(1):123–143.

O'Mahony, C., Jichi, F., Pavlou, M., Monserrat, L., Anastasakis, A., Rapezzi, C., Biagini, E., Gimeno, J. R., Limongelli, G., McKenna, W. J., et al. (2014). A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM risk-SCD). *European Heart Journal*, 35(30):2010–2020.

Wasserman, L., Ramdas, A., and Balakrishnan, S. (2020). Universal inference. *Proceedings of the National Academy of Sciences*, 117(29):16880–16890.

Thank you!

Main reference:

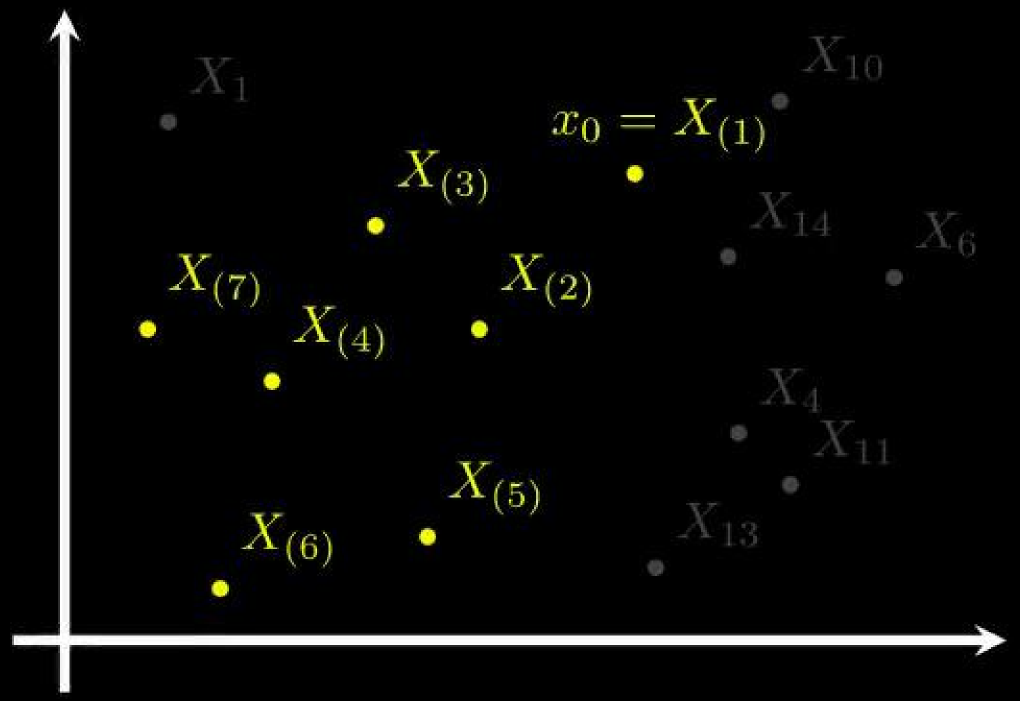
Müller, M. M., Reeve, H. W. J., Cannings, T. I. and Samworth, R. J. (2024) Isotonic subgroup selection. *J. Roy. Statist. Soc., Ser. B (to appear). arXiv:2305.04852*.

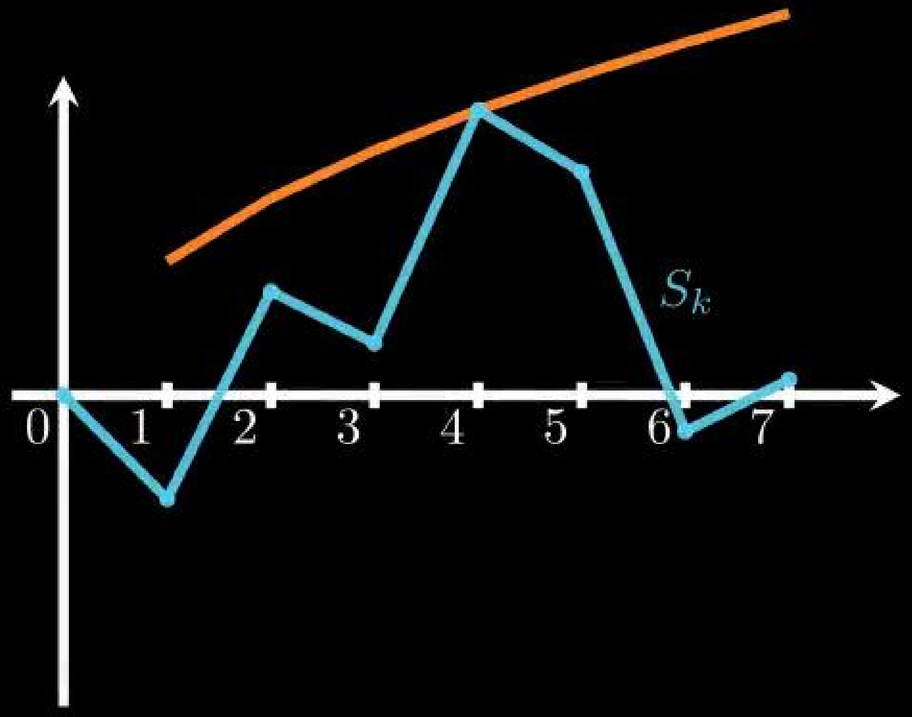
See manuelmmueller.github.io for data and R-code.

Appendix

Construct p-values $\hat{p}_i$ for $H_0(X_i)$, $i \in [m]$

Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) := \{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma) : \eta(x_0) < \tau\}$.





Denote $\mathcal{I}(x_0) := \{i \in [n] : X_i \preccurlyeq x_0\}$, $n(x_0) := |\mathcal{I}(x_0)|$.

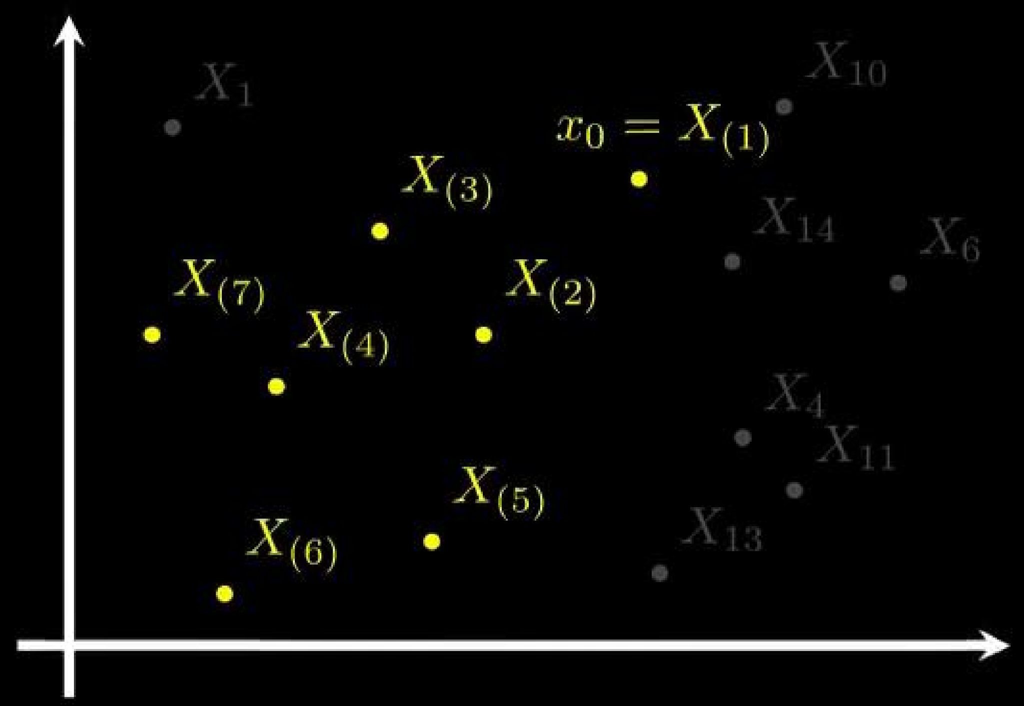
Let $X_{(j)}$ be the jth nearest neighbour of x_0 among X_i, $i \in \mathcal{I}(x_0)$, in sup-norm and let $Y_{(j)}$ be the corresponding response.

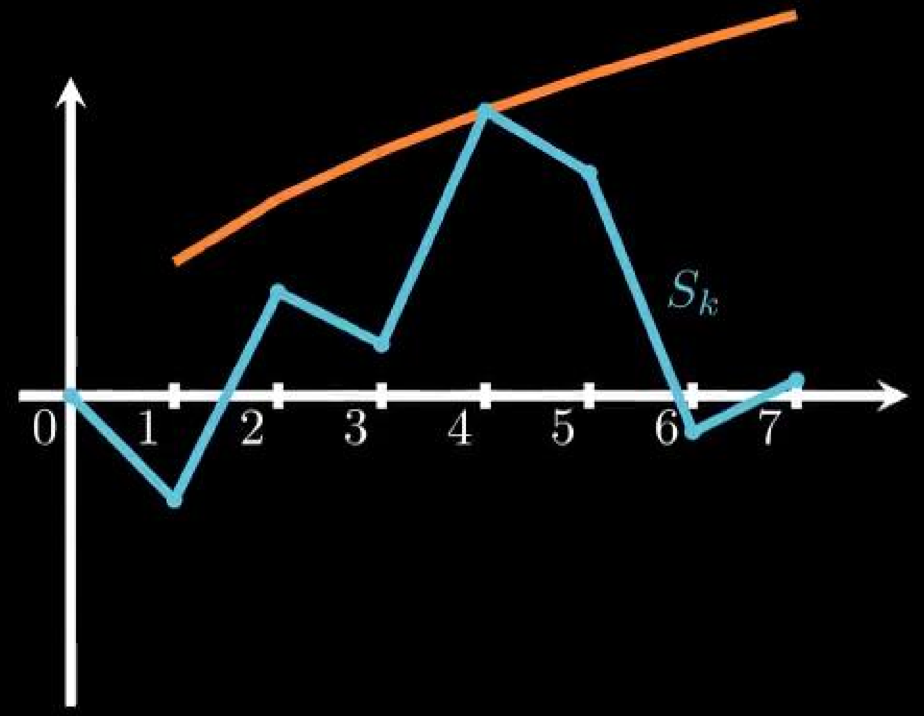
Let

$$S_k := \sum_{j=1}^{k} \frac{Y_{(j)} - \tau}{\sigma}.$$

Then S_k is a supermartingale under $P \in H_0(x_0)$. Combination with time-uniform bounds by Howard et al. (2021) gives p-values from this martingale test (Duan et al., 2020).

Construct p-values $\hat{p}_i$ for $H_0(X_i)$, $i \in [m]$



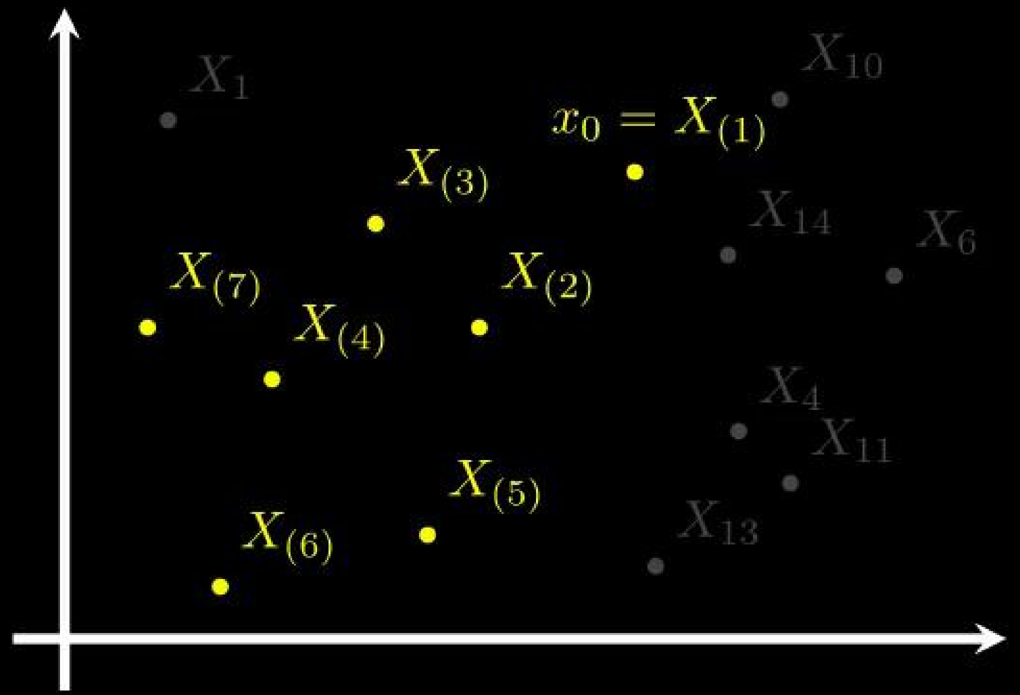


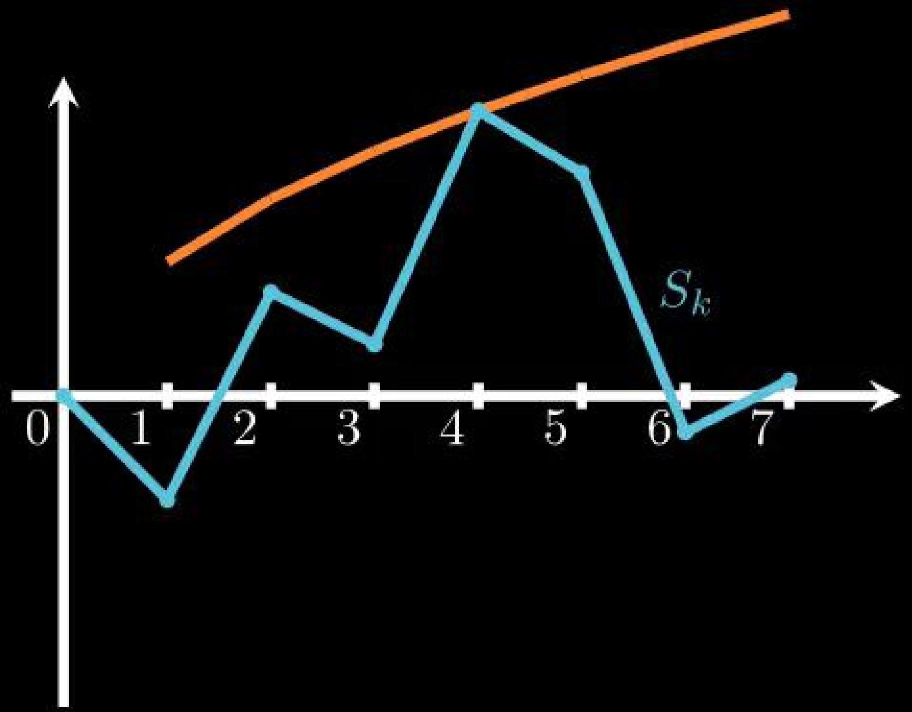
Definition. For $\sigma > 0$, $x \in \mathbb{R}^d$, let

$$\hat{p}_{\sigma,\tau}(x) := 1 \wedge \min_{k \in [n(x)]} 5.2 \exp\left\{-\frac{(S_k \vee 0)^2}{2.0808k} + \frac{\log\log(2k)}{0.72}\right\},$$

whenever $n(x) > 0$, and $\hat{p}_{\sigma,\tau}(x) := 1$ otherwise.

Construct p-values $\hat{p}_i$ for $H_0(X_i)$, $i \in [m]$





Definition. For $\sigma > 0$, $x \in \mathbb{R}^d$, let

$$\hat{p}_{\sigma,\tau}(x) := 1 \wedge \min_{k \in [n(x)]} 5.2 \exp\left\{ -\frac{(S_k \vee 0)^2}{2.0808k} + \frac{\log\log(2k)}{0.72} \right\},$$

whenever $n(x) > 0$, and $\hat{p}_{\sigma,\tau}(x) := 1$ otherwise.

Lemma. When $\eta(x) < \tau$, we have $\mathbb{P}\{\hat{p}_{\sigma,\tau}(x) \le t \mid (X_i)_{i \in [n]}\} \le t$ for all $t \in (0,1)$.

Further conditions are needed for power

Let $\hat{\mathcal{A}}_n(\tau, \alpha, \mathcal{P})$ denote the set of *data-dependent selection sets* controlling Type I error over $\mathcal{P}$.
Recall $R_\tau(\hat{A}) := \mathbb{E}\big\{\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}\big)\big\}$.

Further conditions are needed for power

Let $\hat{\mathcal{A}}_n(\tau, \alpha, \mathcal{P})$ denote the set of *data-dependent selection sets* controlling Type I error over $\mathcal{P}$.
Recall $R_\tau(\hat{A}) := \mathbb{E}\{\mu(\mathcal{X}_\tau(\eta) \setminus \hat{A})\}$.

Proposition. For any $n \in \mathbb{N}$, $\alpha \in (0,1)$, $\sigma > 0$,

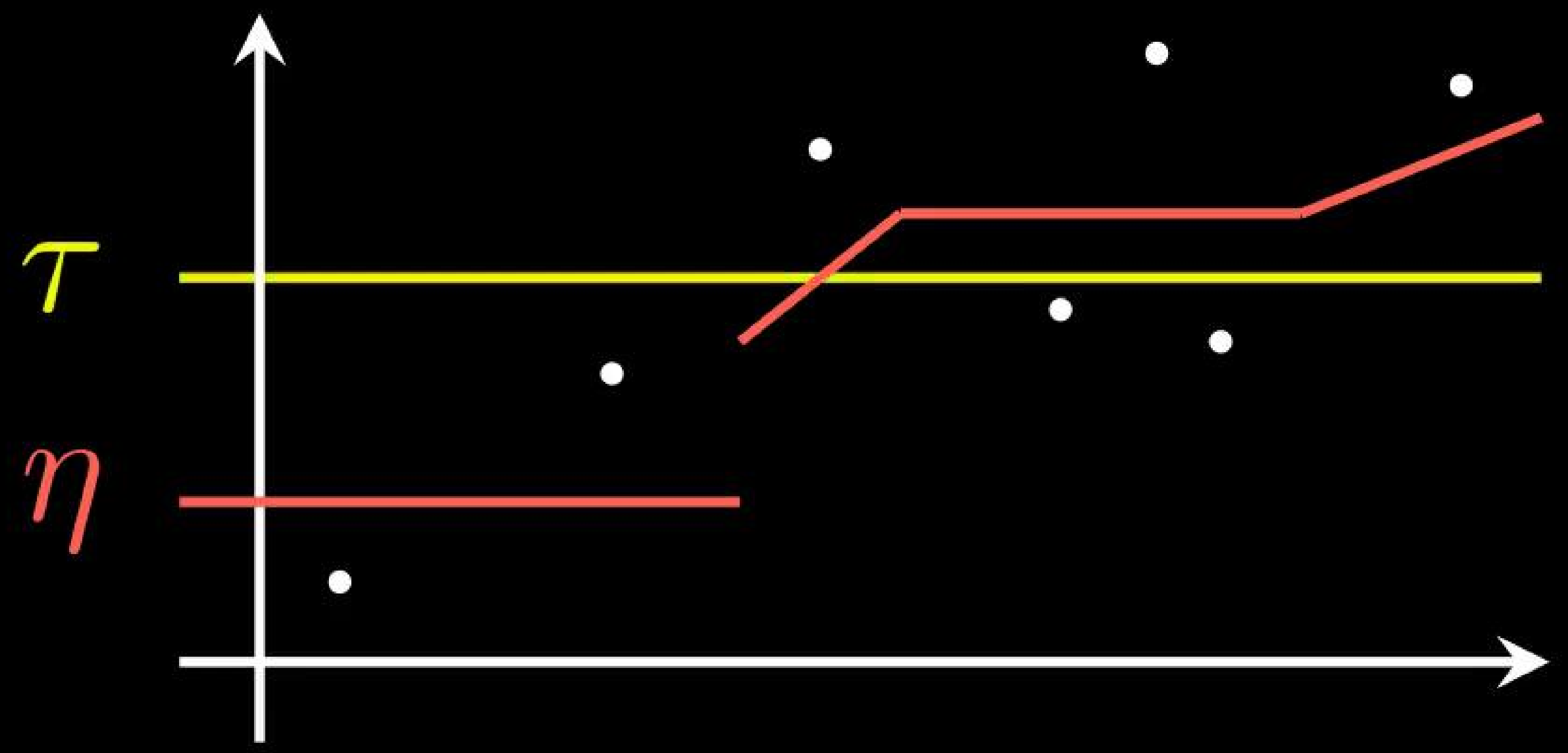
$$\sup_{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma)} \inf_{\hat{A} \in \hat{\mathcal{A}}_n(\tau, \alpha, \mathcal{P}_{\mathrm{Mon},d}(\sigma))} R_\tau(\hat{A}) \geq 1 - \alpha.$$

Further conditions are needed for power

Let $\hat{\mathcal{A}}_n(\tau, \alpha, \mathcal{P})$ denote the set of *data-dependent selection sets* controlling Type I error over $\mathcal{P}$. Recall $R_\tau(\hat{A}) := \mathbb{E}\{\mu(\mathcal{X}_\tau(\eta) \setminus \hat{A})\}$.

Proposition. For any $n \in \mathbb{N}$, $\alpha \in (0,1)$, $\sigma > 0$,

$$\sup_{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma)} \inf_{\hat{A} \in \hat{\mathcal{A}}_n(\tau, \alpha, \mathcal{P}_{\mathrm{Mon},d}(\sigma))} R_\tau(\hat{A}) \geq 1 - \alpha.$$

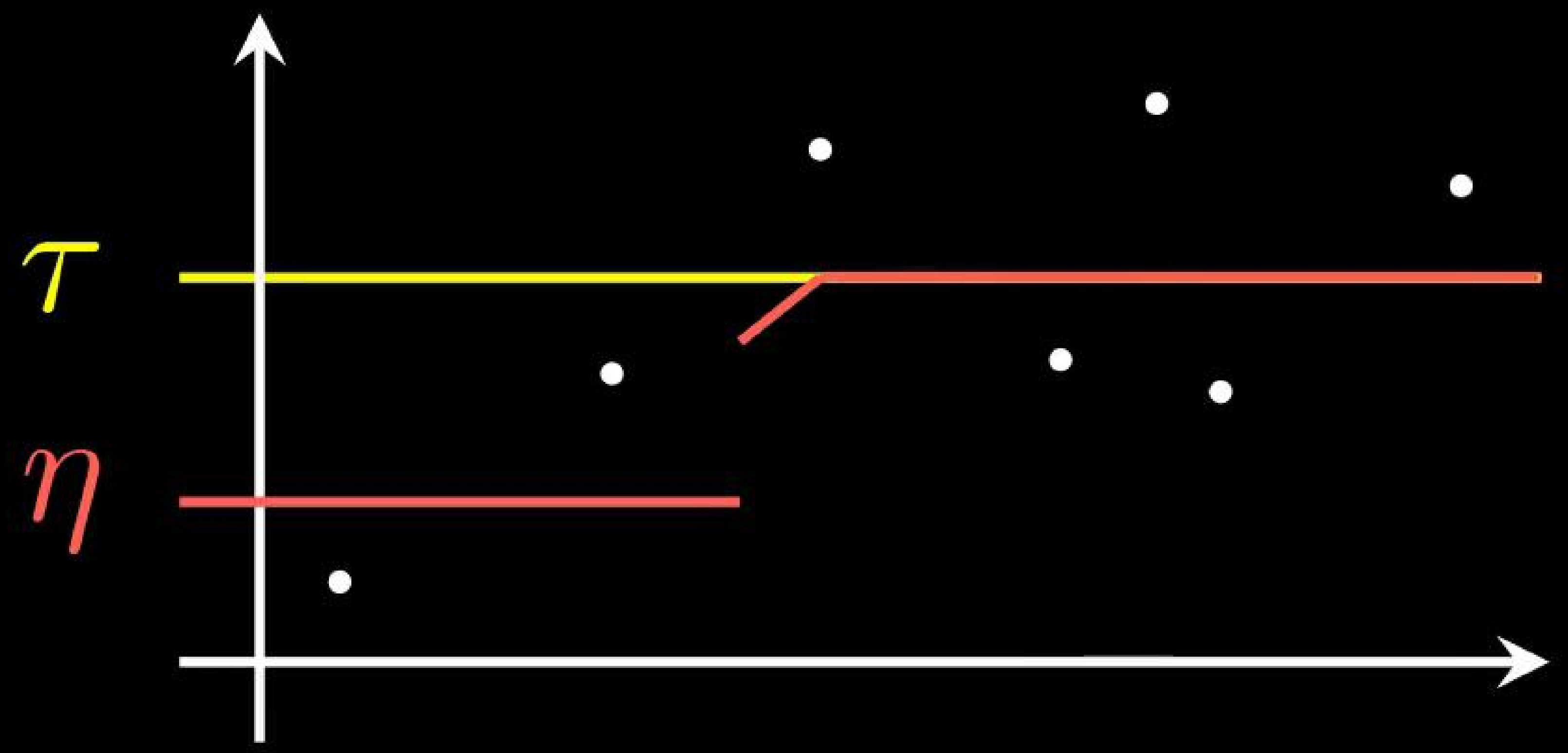


Further conditions are needed for power

Let $\hat{\mathcal{A}}_n(\tau, \alpha, \mathcal{P})$ denote the set of *data-dependent selection sets* controlling Type I error over $\mathcal{P}$. Recall $R_\tau(\hat{A}) := \mathbb{E}\{\mu(\mathcal{X}_\tau(\eta) \setminus \hat{A})\}$.

Proposition. For any $n \in \mathbb{N}$, $\alpha \in (0,1)$, $\sigma > 0$,

$$\sup_{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma)} \inf_{\hat{A} \in \hat{\mathcal{A}}_n(\tau, \alpha, \mathcal{P}_{\mathrm{Mon},d}(\sigma))} R_\tau(\hat{A}) \geq 1 - \alpha.$$



Further conditions are needed for power

Let $\hat{\mathcal{A}}_n(\tau, \alpha, \mathcal{P})$ denote the set of *data-dependent selection sets* controlling Type I error over $\mathcal{P}$. Recall $R_\tau(\hat{A}) := \mathbb{E}\{\mu(\mathcal{X}_\tau(\eta) \setminus \hat{A})\}$.

Proposition. For any $n \in \mathbb{N}$, $\alpha \in (0,1)$, $\sigma > 0$,

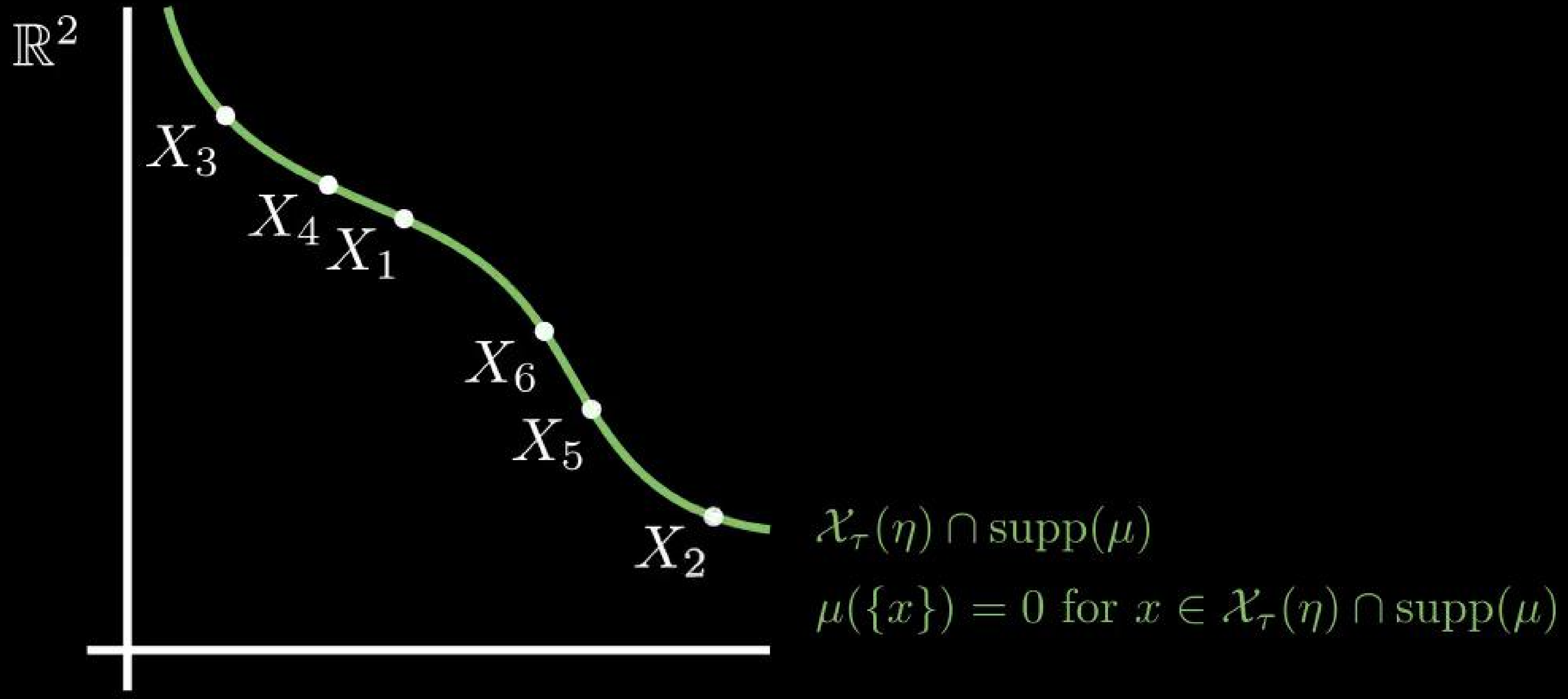
$$\sup_{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma)} \inf_{\hat{A} \in \hat{\mathcal{A}}_n(\tau,\alpha,\mathcal{P}_{\mathrm{Mon},d}(\sigma))} R_\tau(\hat{A}) \geq 1 - \alpha.$$

Further conditions are needed for power

Let $\hat{\mathcal{A}}_n(\tau, \alpha, \mathcal{P})$ denote the set of *data-dependent selection sets* controlling Type I error over $\mathcal{P}$. Recall $R_\tau(\hat{A}) := \mathbb{E}\{\mu(\mathcal{X}_\tau(\eta) \setminus \hat{A})\}$.

Proposition. For any $n \in \mathbb{N}$, $\alpha \in (0,1)$, $\sigma > 0$,

$$\sup_{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma)} \inf_{\hat{A} \in \hat{\mathcal{A}}_n(\tau, \alpha, \mathcal{P}_{\mathrm{Mon},d}(\sigma))} R_\tau(\hat{A}) \geq 1 - \alpha.$$

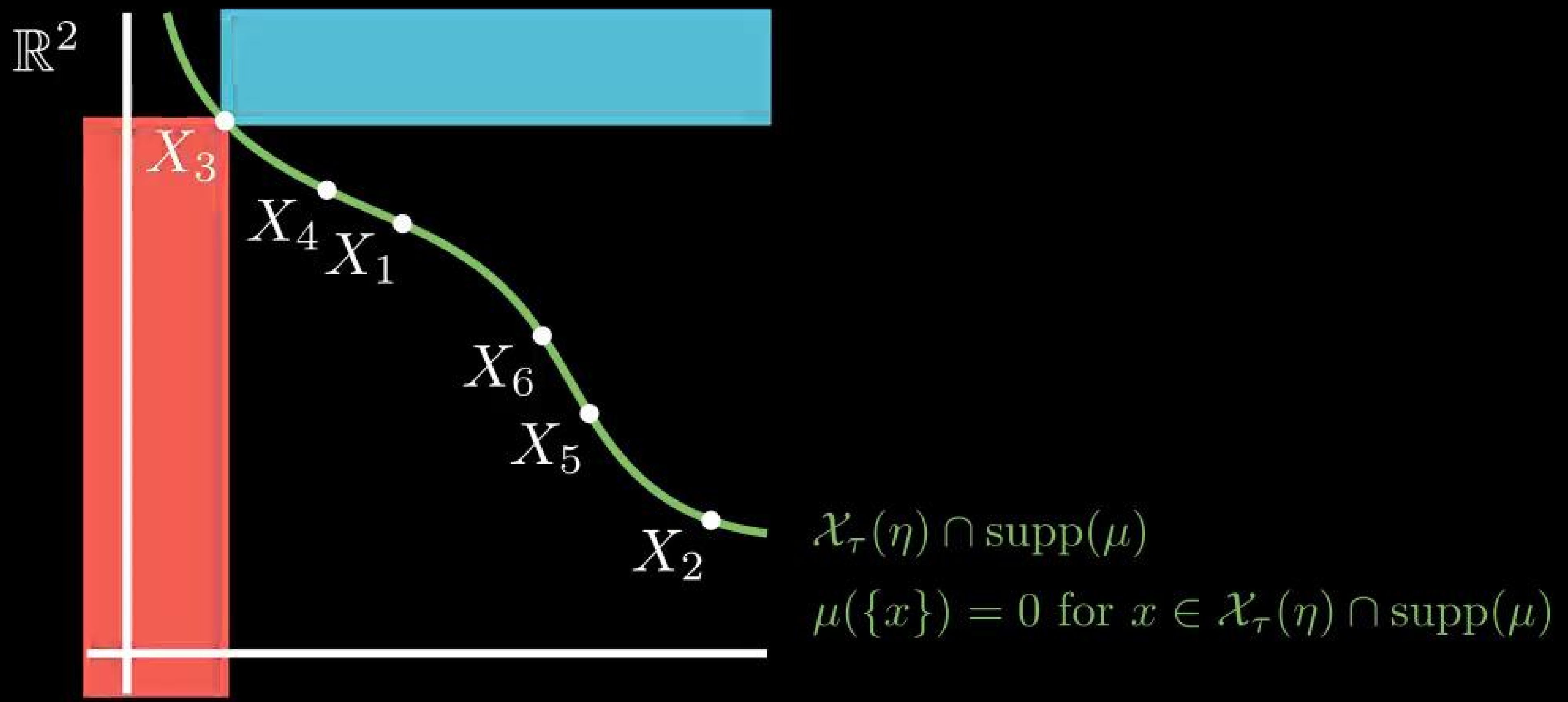


Further conditions are needed for power

Let $\hat{\mathcal{A}}_n(\tau, \alpha, \mathcal{P})$ denote the set of *data-dependent selection sets* controlling Type I error over $\mathcal{P}$. Recall $R_\tau(\hat{A}) := \mathbb{E}\{\mu(\mathcal{X}_\tau(\eta) \setminus \hat{A})\}$.

Proposition. For any $n \in \mathbb{N}$, $\alpha \in (0, 1)$, $\sigma > 0$,

$$\sup_{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma)} \inf_{\hat{A} \in \hat{\mathcal{A}}_n(\tau, \alpha, \mathcal{P}_{\mathrm{Mon},d}(\sigma))} R_\tau(\hat{A}) \geq 1 - \alpha.$$

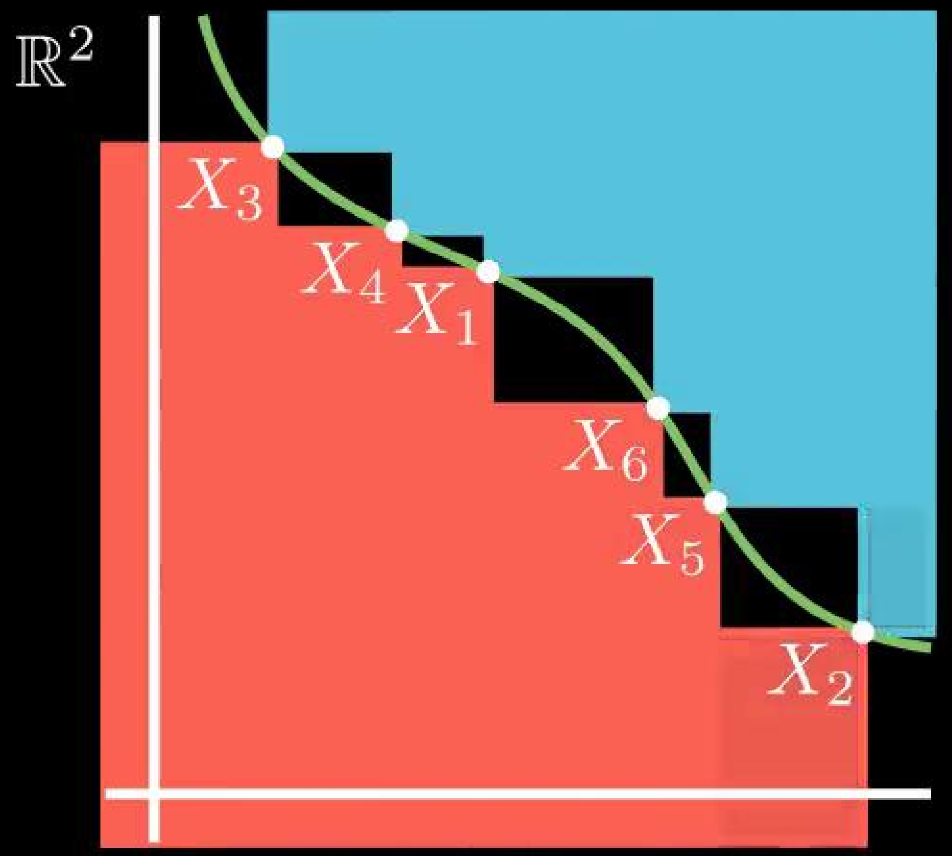


Further conditions are needed for power

Let $\hat{\mathcal{A}}_n(\tau, \alpha, \mathcal{P})$ denote the set of *data-dependent selection sets* controlling Type I error over $\mathcal{P}$.
Recall $R_\tau(\hat{A}) := \mathbb{E}\{\mu(\mathcal{X}_\tau(\eta) \setminus \hat{A})\}$.

Proposition. For any $n \in \mathbb{N}$, $\alpha \in (0,1)$, $\sigma > 0$,

$$\sup_{P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma)} \inf_{\hat{A} \in \hat{\mathcal{A}}_n(\tau,\alpha,\mathcal{P}_{\mathrm{Mon},d}(\sigma))} R_\tau(\hat{A}) \geq 1 - \alpha.$$



$\mathcal{X}_\tau(\eta) \cap \mathrm{supp}(\mu)$

$\mu(\{x\}) = 0$ for $x \in \mathcal{X}_\tau(\eta) \cap \mathrm{supp}(\mu)$

Power

Definition. Let $\sigma, \gamma, \lambda > 0$ and $\theta > 1$. $\mathcal{P}_{\mathrm{MonReg}} \equiv \mathcal{P}_{\mathrm{MonReg},d}(\sigma, \tau, \gamma, \lambda, \theta)$ denotes the set of distributions $P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma)$ for which additionally

1. $\eta\big(x + r(1, \ldots, 1)^\top\big) \geq \tau + \lambda \cdot r^\gamma$ for all $x \in \mathcal{X}_\tau(\eta) \cap \mathrm{supp}(\mu)$ and $r \in (0, 1]$;
2. $\theta^{-1} \cdot r^d \leq \mu\big(\{y \in \mathbb{R}^d : \|y - x\|_\infty \leq r\}\big) \leq \theta \cdot (2r)^d$ for all $x \in \mathcal{X}_\tau(\eta) \cap \mathrm{supp}(\mu)$, $r \in (0, 1]$.

Power

Definition. Let $\sigma, \gamma, \lambda > 0$ and $\theta > 1$. $\mathcal{P}_{\mathrm{MonReg}} \equiv \mathcal{P}_{\mathrm{MonReg},d}(\sigma, \tau, \gamma, \lambda, \theta)$ denotes the set of distributions $P \in \mathcal{P}_{\mathrm{Mon},d}(\sigma)$ for which additionally

1. $\eta\big(x + r(1, \ldots, 1)^\top\big) \geq \tau + \lambda \cdot r^\gamma$ for all $x \in \mathcal{X}_\tau(\eta) \cap \mathrm{supp}(\mu)$ and $r \in (0, 1]$;
2. $\theta^{-1} \cdot r^d \leq \mu\big(\{y \in \mathbb{R}^d : \|y - x\|_\infty \leq r\}\big) \leq \theta \cdot (2r)^d$ for all $x \in \mathcal{X}_\tau(\eta) \cap \mathrm{supp}(\mu)$, $r \in (0, 1]$.

Theorem. $\exists C(d, \theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0, 1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d, \theta) \left\{ \left(\frac{\sigma^2}{n\lambda^2} \log_+\left(\frac{m \log_+ n}{\alpha} \right) \right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m} \right)^{1/d} \right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.

Power

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta)\left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{m\log_+ n}{\alpha}\right)\right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m}\right)^{1/d}\right\},$$

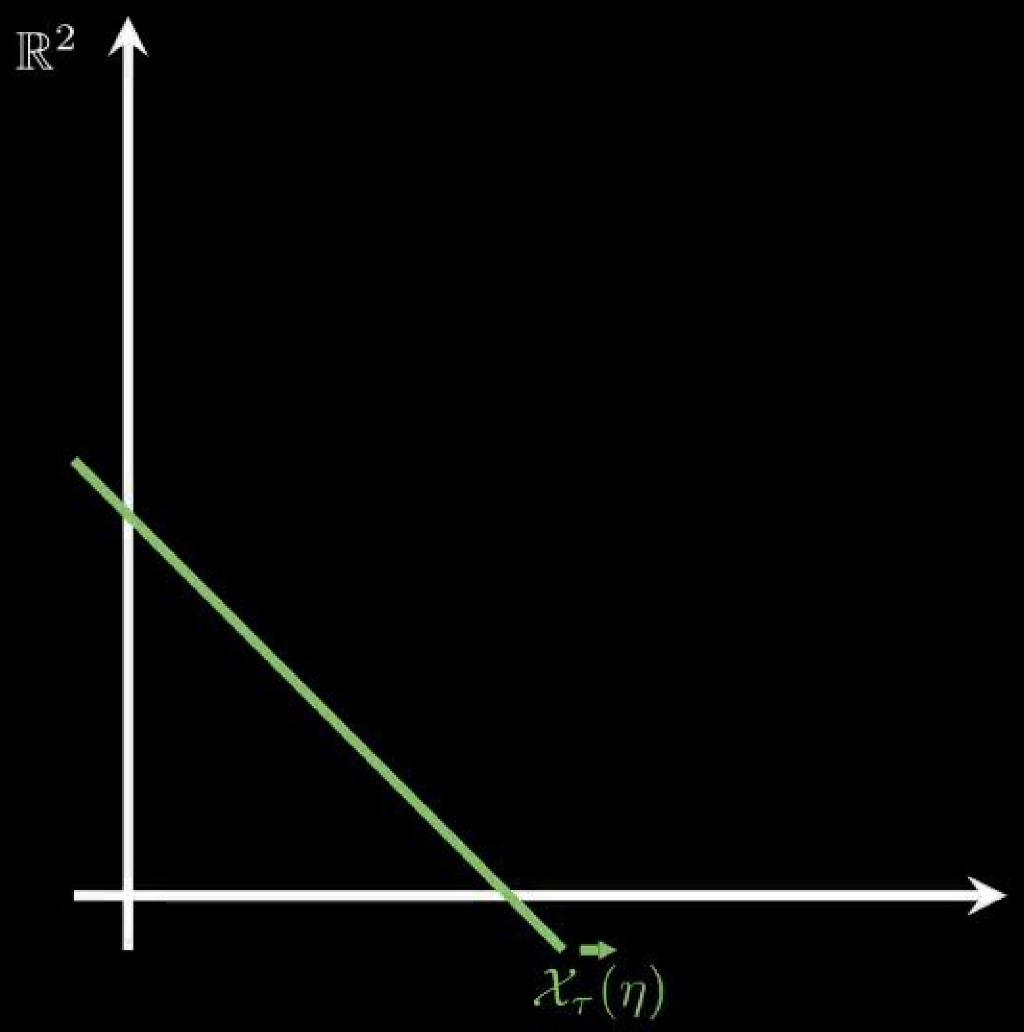
where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.

Power

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta) \left\{ \left(\frac{\sigma^2}{n\lambda^2} \log_+ \left(\frac{m \log_+ n}{\alpha} \right) \right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m} \right)^{1/d} \right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.

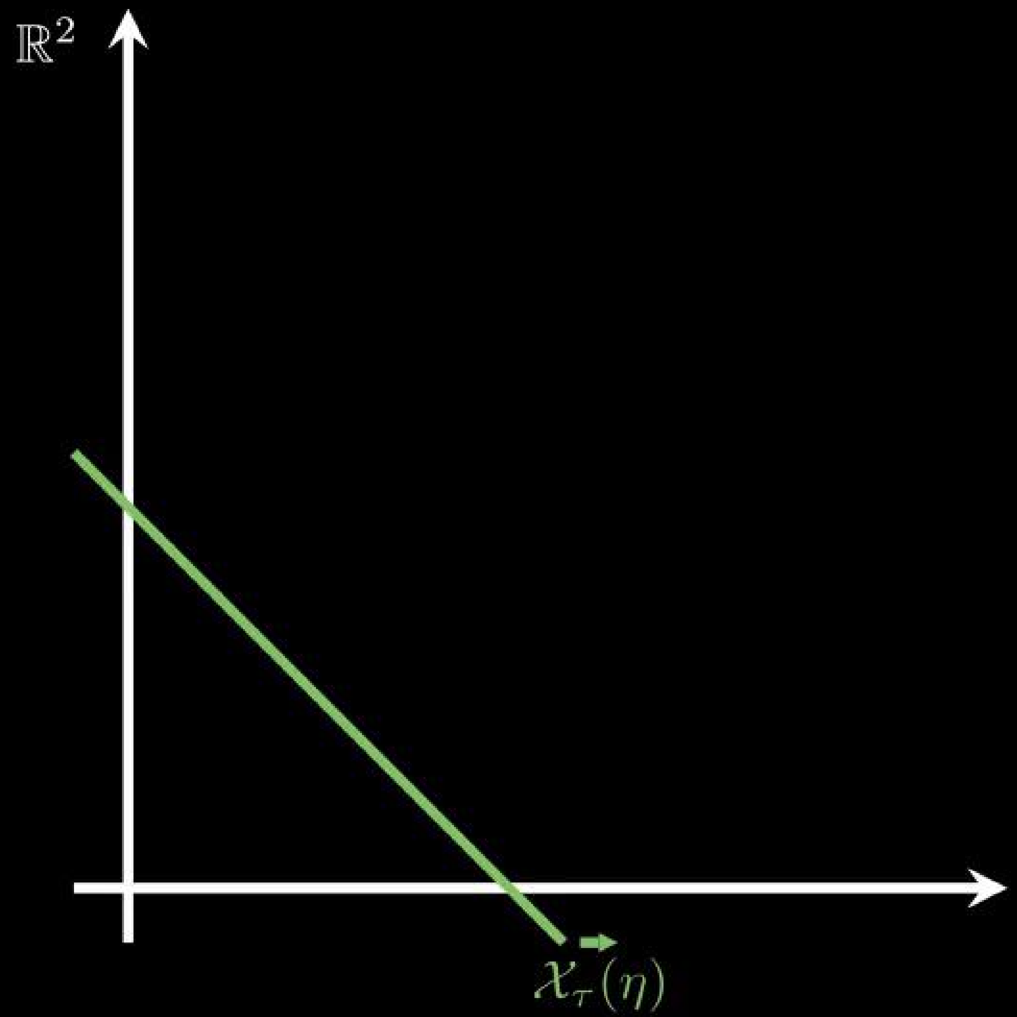


Power

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta)\left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{m\log_+ n}{\alpha}\right)\right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m}\right)^{1/d}\right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.



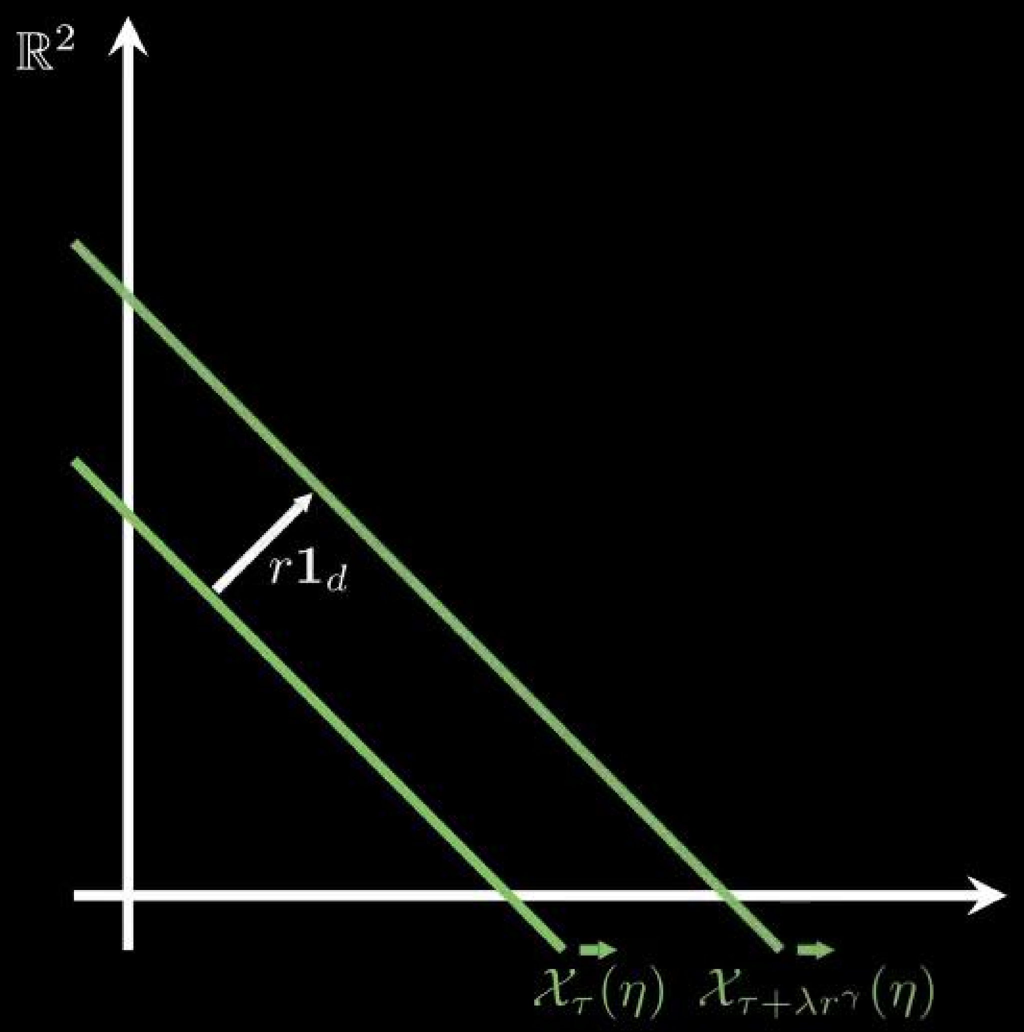
Let $r > 0$ be small.

Power

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta)\left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{m\log_+ n}{\alpha}\right)\right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m}\right)^{1/d}\right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.



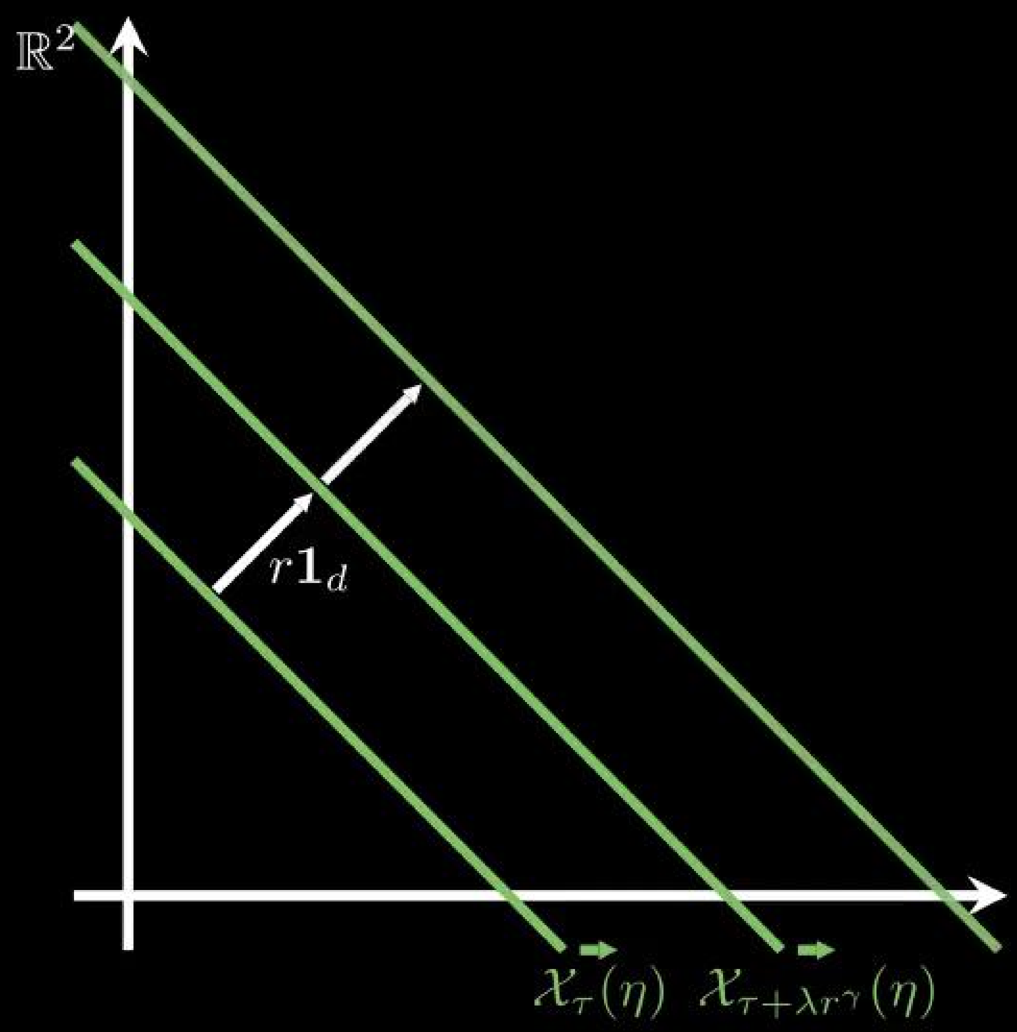
Let $r > 0$ be small.

Power

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta)\left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{m\log_+ n}{\alpha}\right)\right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m}\right)^{1/d}\right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.



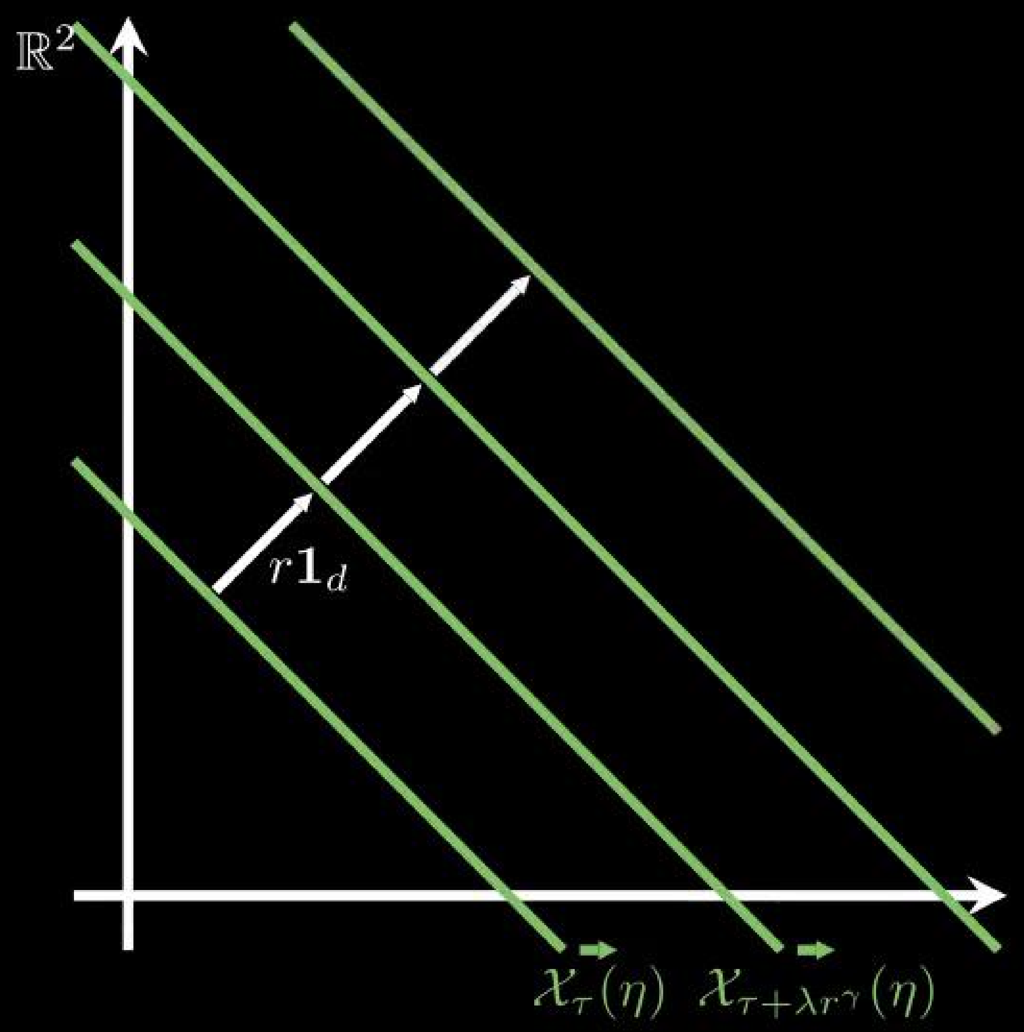
Let $r > 0$ be small.

Power

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta)\left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{m\log_+ n}{\alpha}\right)\right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m}\right)^{1/d}\right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.



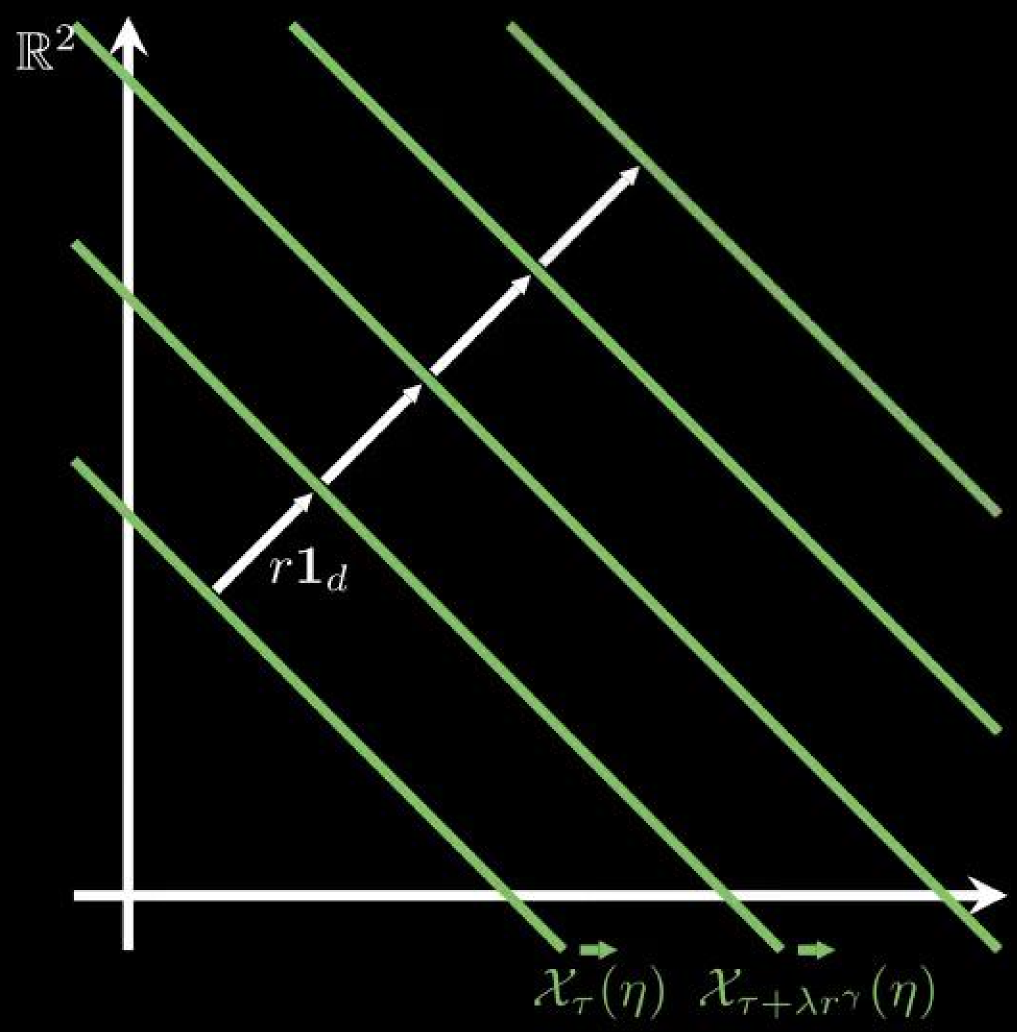
Let $r > 0$ be small.

Power

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta)\left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{m\log_+ n}{\alpha}\right)\right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m}\right)^{1/d}\right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.



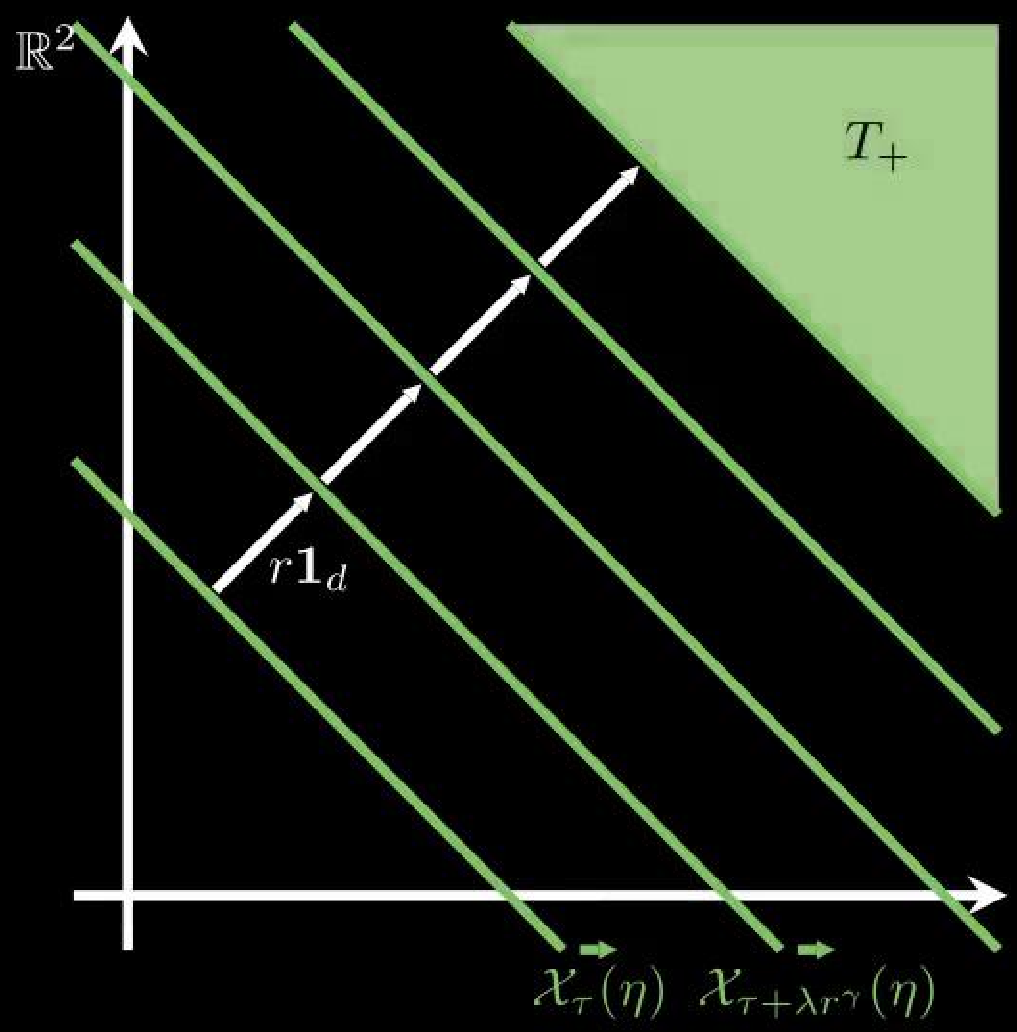
Let $r > 0$ be small.

Power

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta) \left\{ \left(\frac{\sigma^2}{n\lambda^2} \log_+ \left(\frac{m \log_+ n}{\alpha} \right) \right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m} \right)^{1/d} \right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.



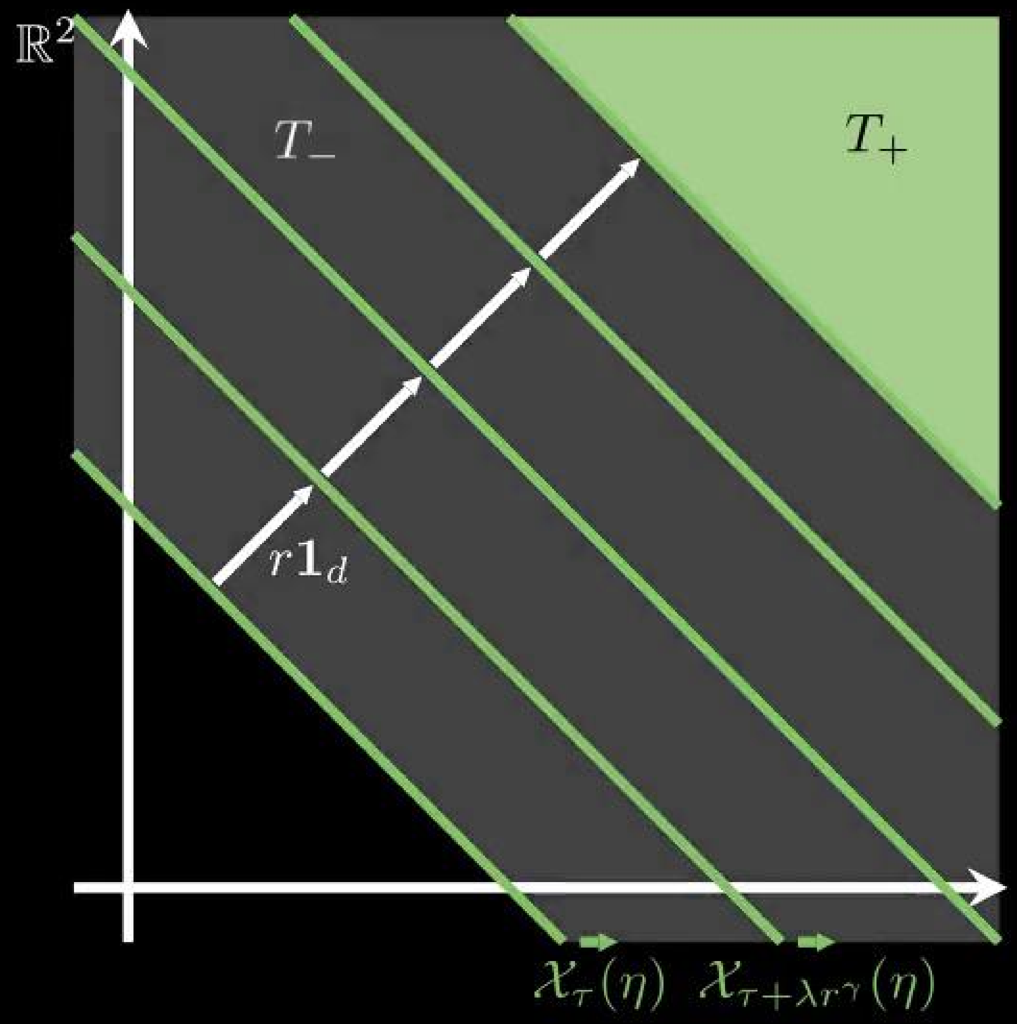
Let $r > 0$ be small.

Power

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta)\left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{m\log_+ n}{\alpha}\right)\right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m}\right)^{1/d}\right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.



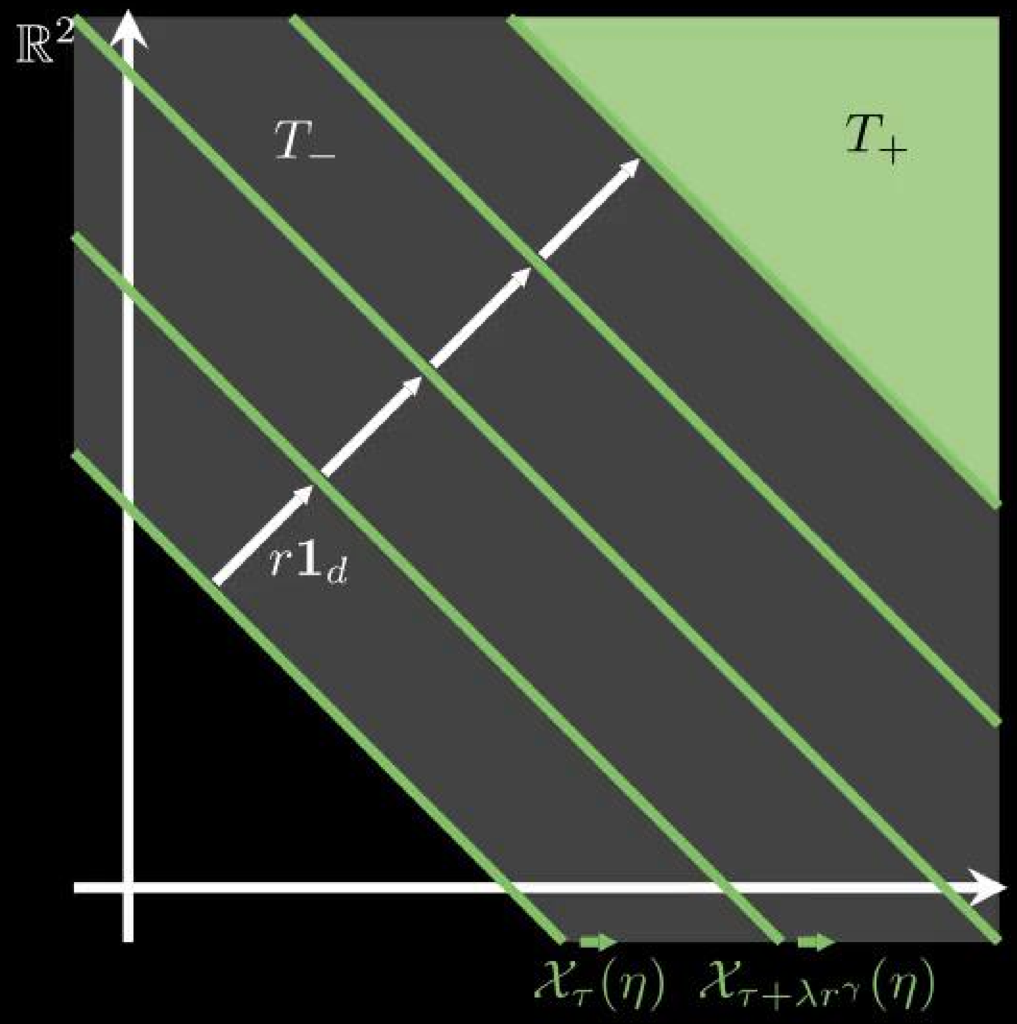
Let $r > 0$ be small.

Power

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta)\left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{m\log_+ n}{\alpha}\right)\right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m}\right)^{1/d}\right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.



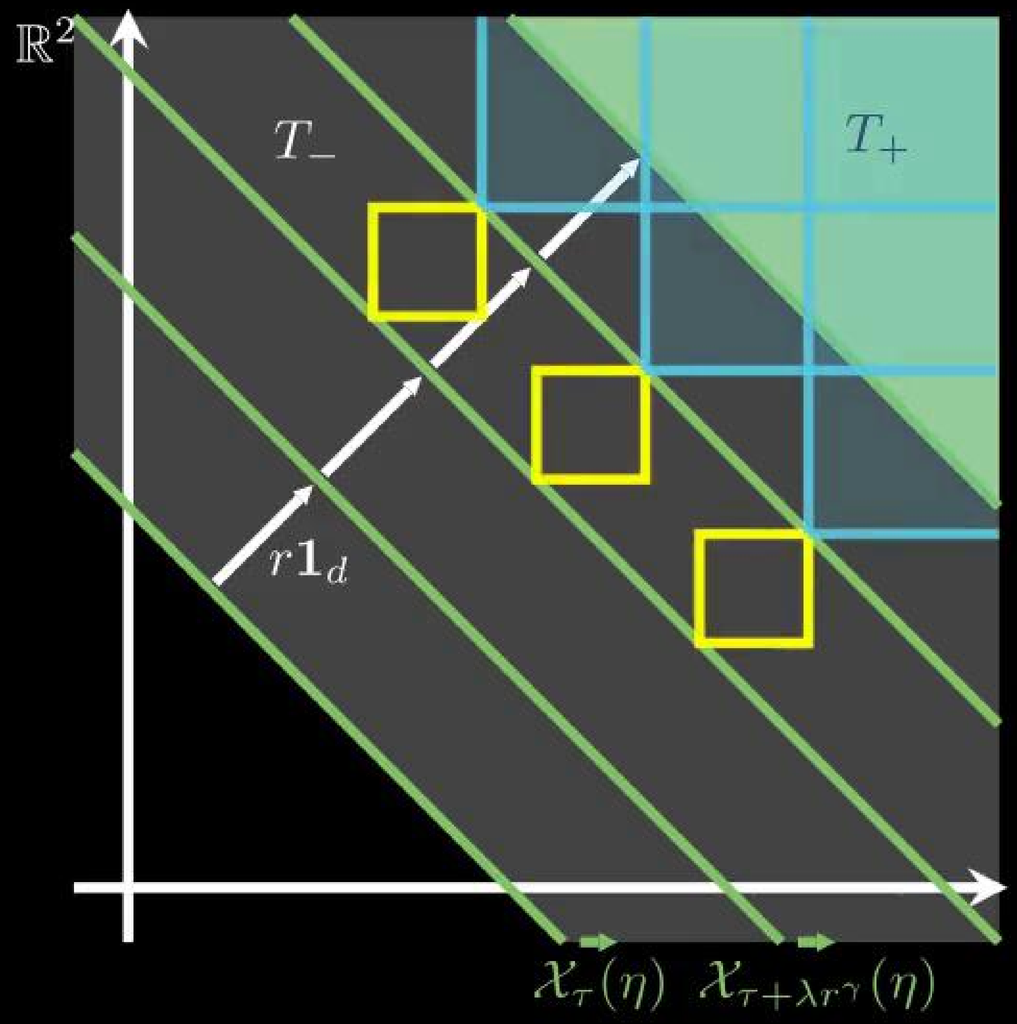
Let $r > 0$ be small.
$\mu\{\mathcal{X}_\tau(\eta) \setminus T_+\} = \mu(T_-) \approx 4r \approx r.$

Power

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta)\left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{m\log_+ n}{\alpha}\right)\right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m}\right)^{1/d}\right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.



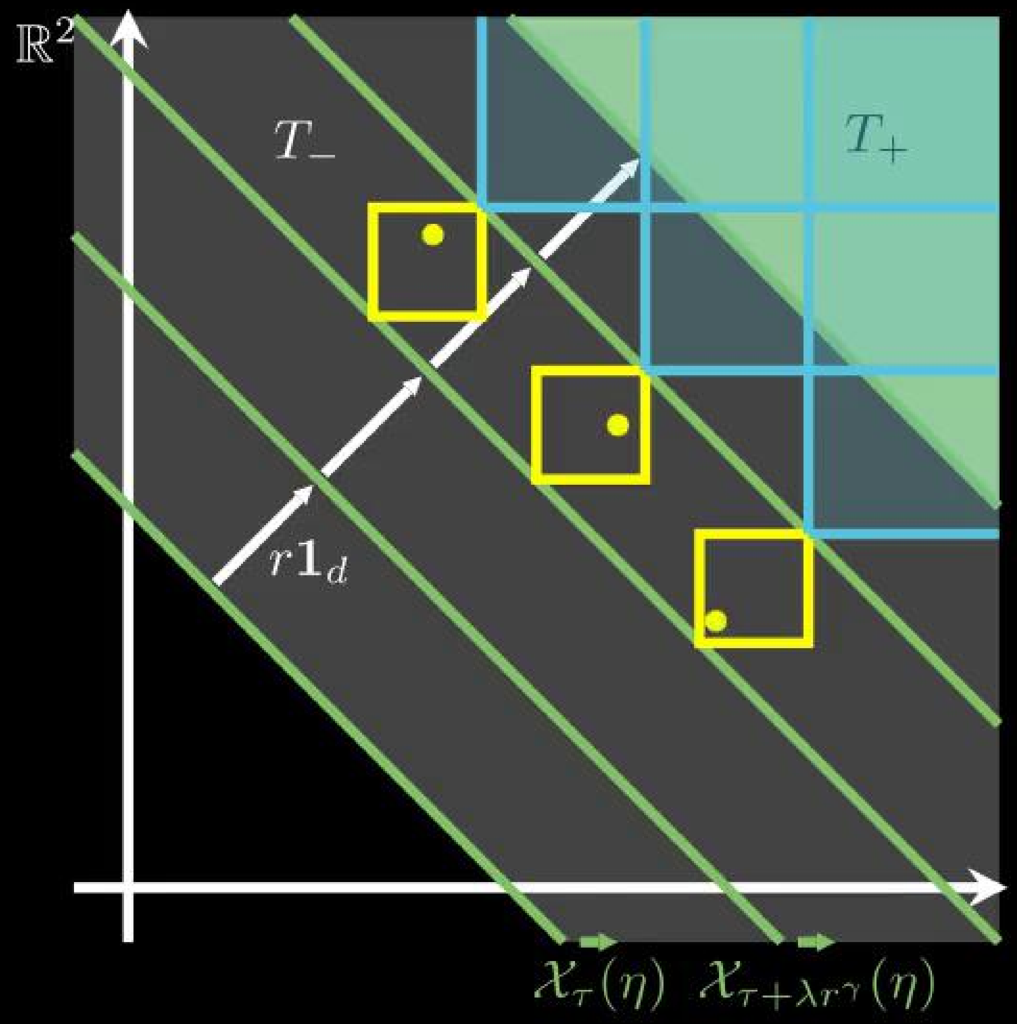
Let $r > 0$ be small.
$\mu\{\mathcal{X}_\tau(\eta) \setminus T_+\} = \mu(T_-) \approx 4r \approx r$.
Need $q :\approx 1/r^{d-1}$ boxes to cover T_+.

Power

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta)\left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{m\log_+ n}{\alpha}\right)\right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m}\right)^{1/d}\right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.



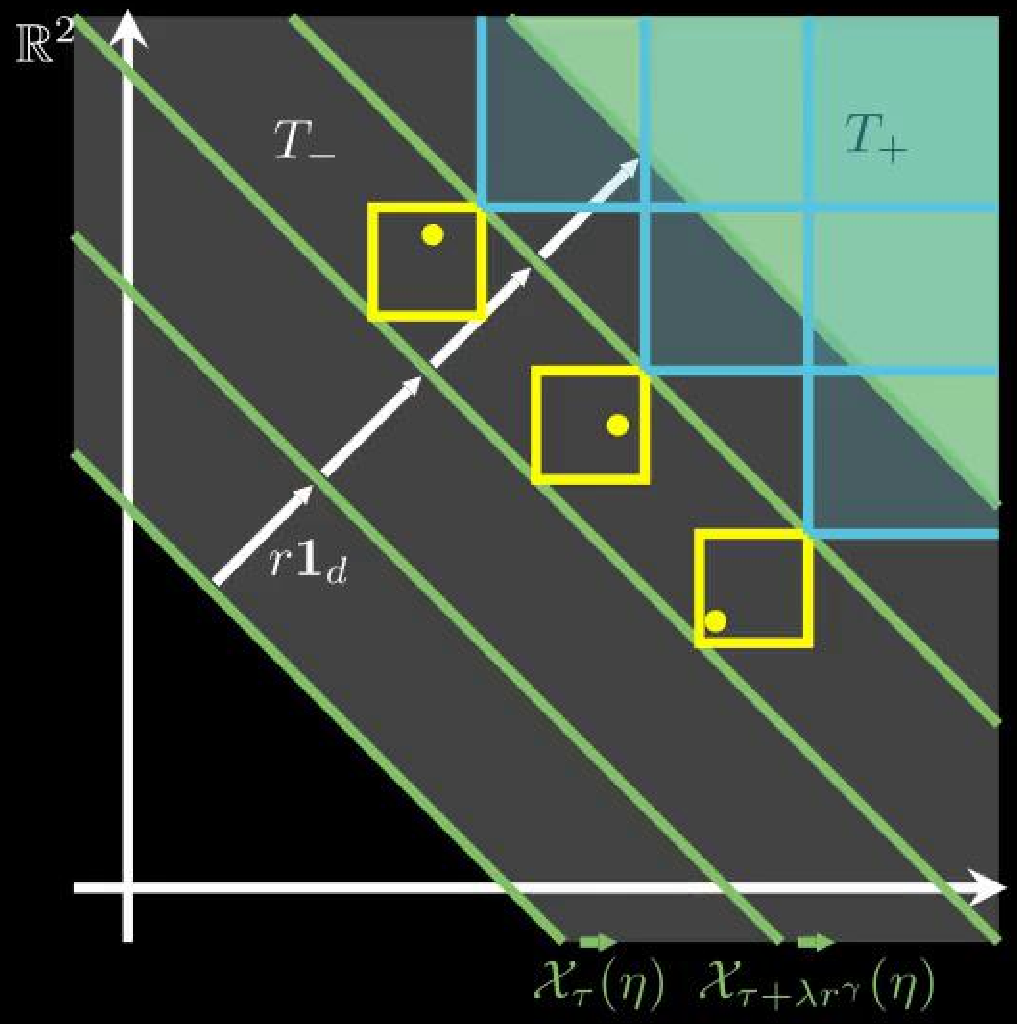
Let $r > 0$ be small.
$\mu\{\mathcal{X}_\tau(\eta) \setminus T_+\} = \mu(T_-) \approx 4r \approx r$.
Need $q :\approx 1/r^{d-1}$ boxes to cover T_+.

Power

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta) \left\{ \left(\frac{\sigma^2}{n\lambda^2} \log_+\left(\frac{m \log_+ n}{\alpha} \right) \right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m} \right)^{1/d} \right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.



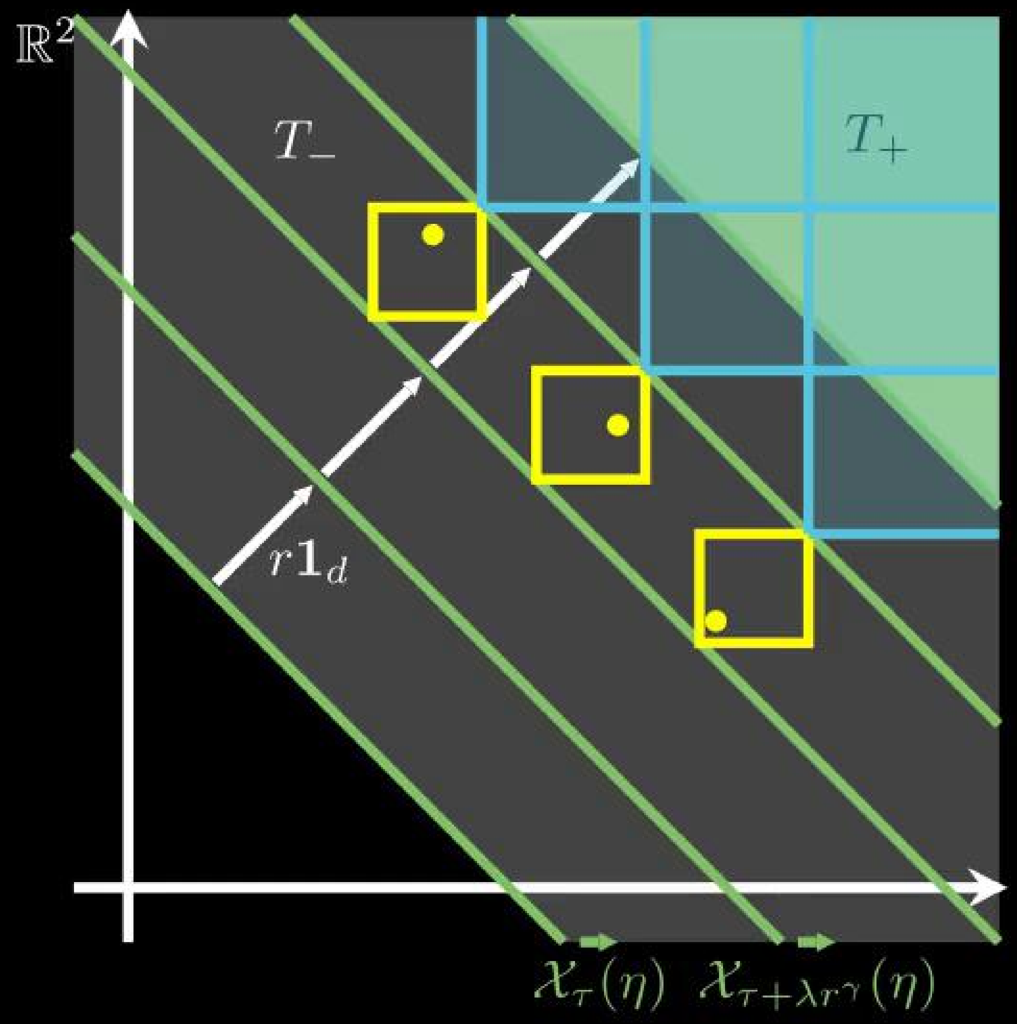
Let $r > 0$ be small.
$\mu\{\mathcal{X}_\tau(\eta) \setminus T_+\} = \mu(T_-) \approx 4r \approx r$.
Need $q :\approx 1/r^{d-1}$ boxes to cover T_+.
Out of $(X_i)_{i \in [m]}$ roughly mr^d in each box.

Power

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta) \left\{ \left(\frac{\sigma^2}{n\lambda^2} \log_+ \left(\frac{m \log_+ n}{\alpha} \right) \right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m} \right)^{1/d} \right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.



Let $r > 0$ be small.

$\mu\{\mathcal{X}_\tau(\eta) \setminus T_+\} = \mu(T_-) \approx 4r \approx r$.

Need $q :\approx 1/r^{d-1}$ boxes to cover T_+.

Out of $(X_i)_{i \in [m]}$ roughly mr^d in each box.

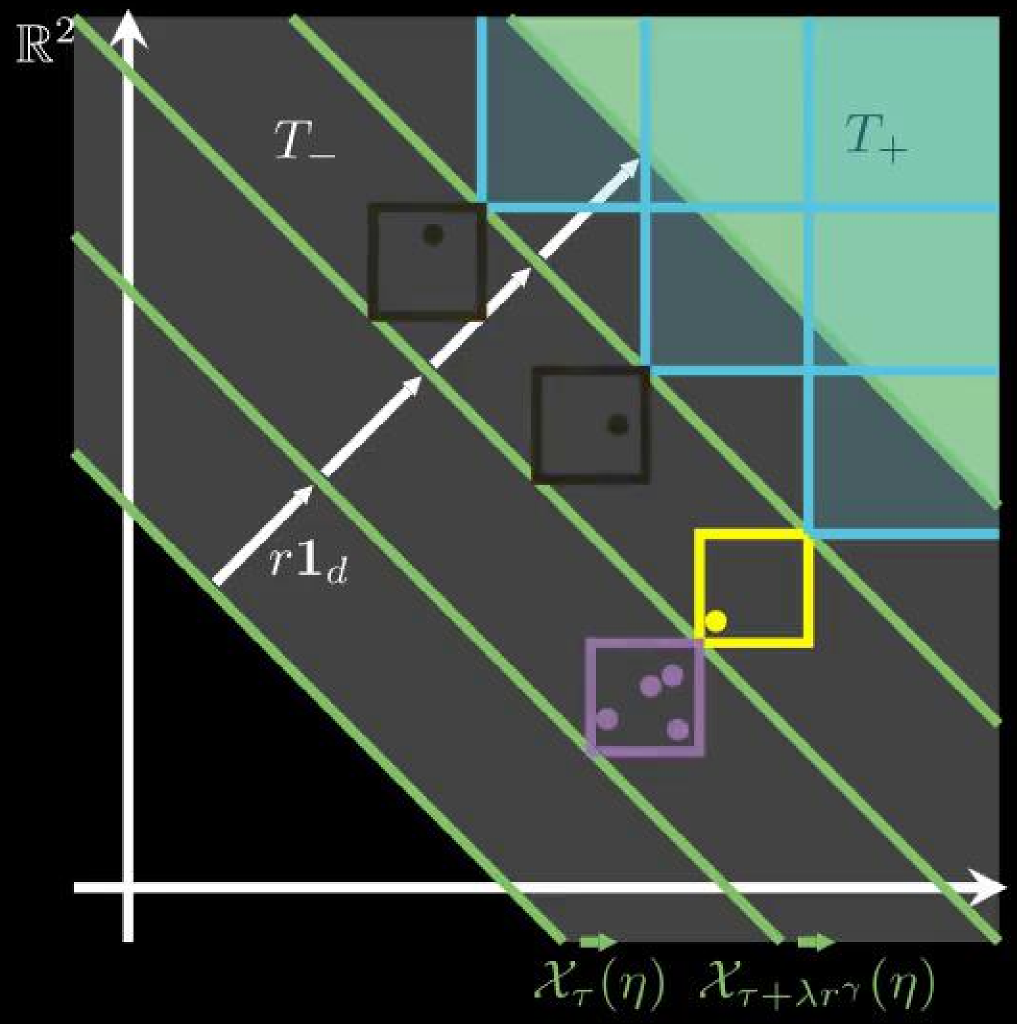
At least one per box if $mr^d \gtrsim \log q \Longleftarrow \boxed{r \overset{!}{\gtrsim} \left(\frac{\log m}{m}\right)^{1/d}}$.

Power

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta)\left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{m\log_+ n}{\alpha}\right)\right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m}\right)^{1/d}\right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.



Let $r > 0$ be small.
$\mu\{\mathcal{X}_\tau(\eta) \setminus T_+\} = \mu(T_-) \approx 4r \approx r.$
Need $q :\approx 1/r^{d-1}$ boxes to cover T_+.
Out of $(X_i)_{i\in[m]}$ roughly mr^d in each box.

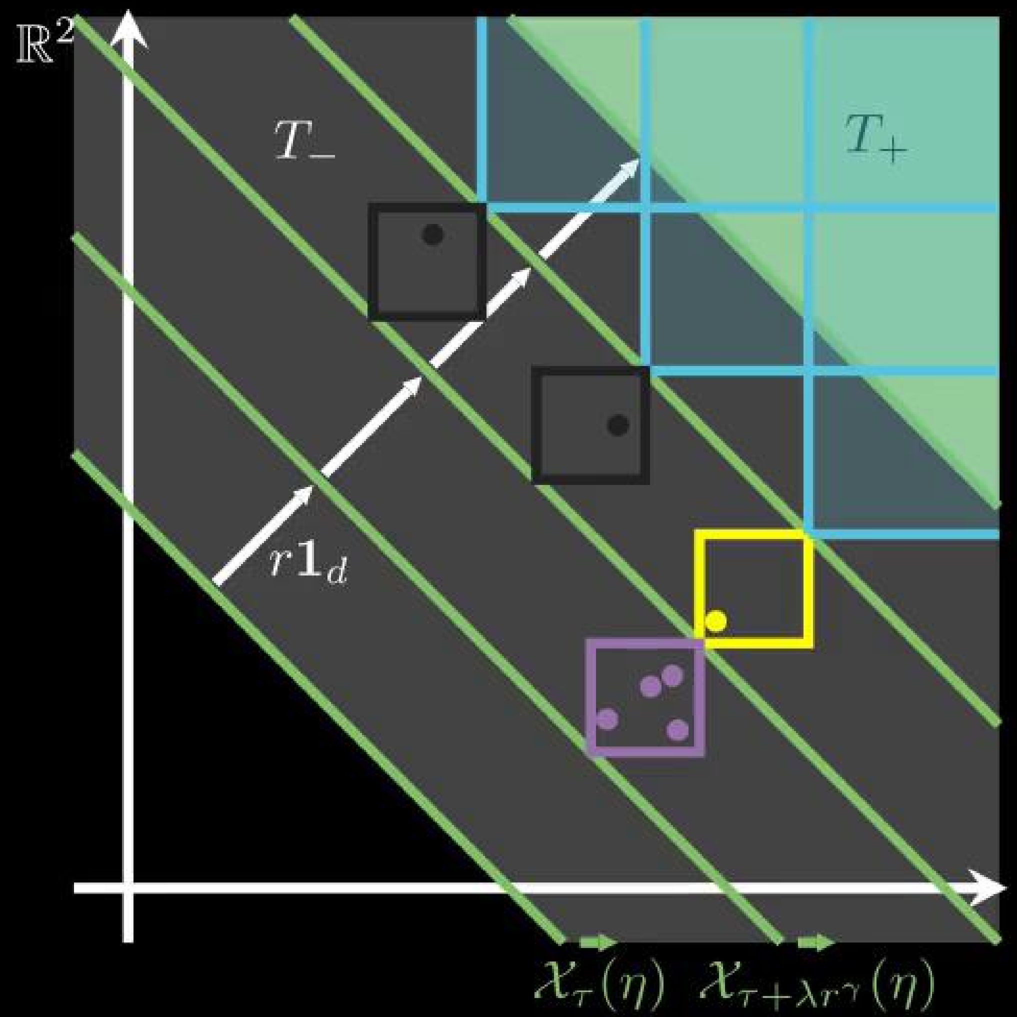
At least one per box if $mr^d \gtrsim \log q \Longleftarrow \boxed{r \overset{!}{\gtrsim} \left(\frac{\log m}{m}\right)^{1/d}}$.

Power

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta)\left\{ \left(\frac{\sigma^2}{n\lambda^2} \log_+\left(\frac{m \log_+ n}{\alpha} \right) \right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m} \right)^{1/d} \right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.



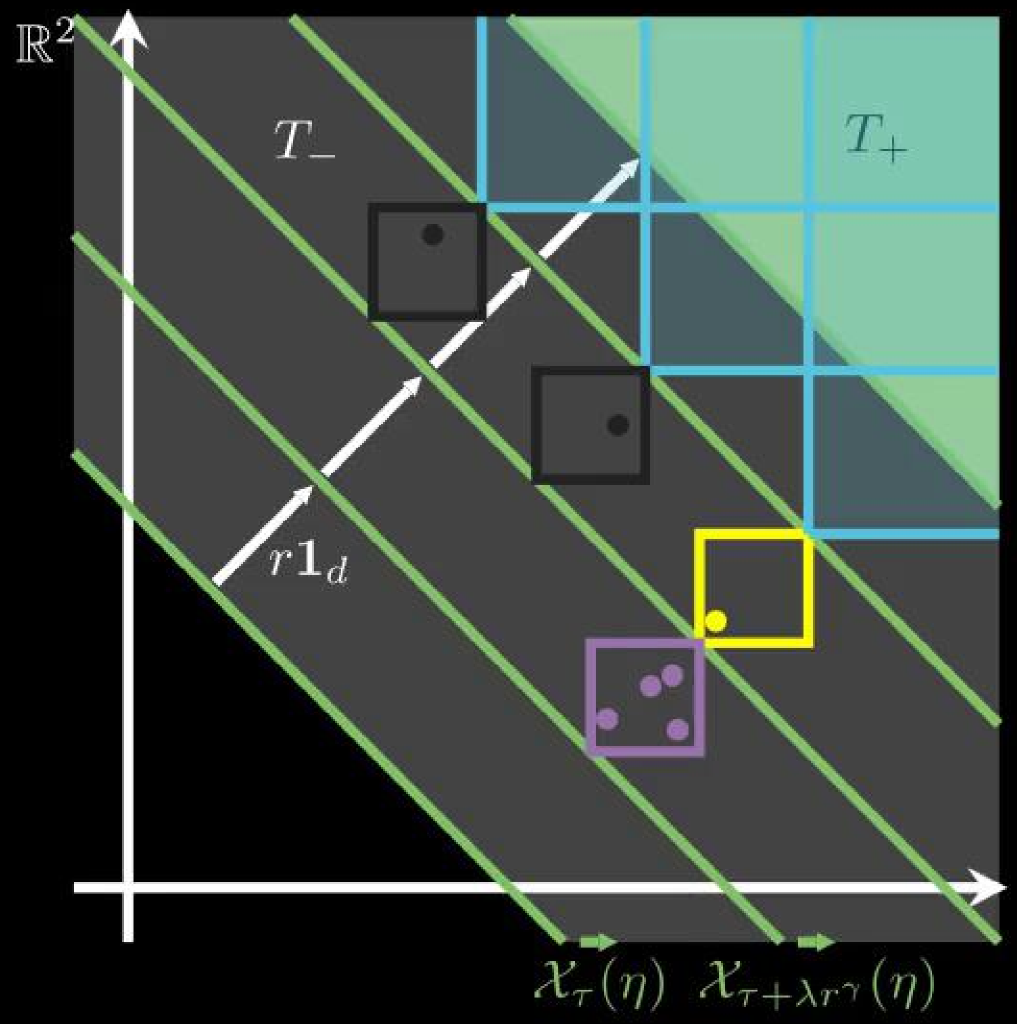
Let $r > 0$ be small.
$\mu\{\mathcal{X}_\tau(\eta) \setminus T_+\} = \mu(T_-) \approx 4r \approx r$.
Need $q :\approx 1/r^{d-1}$ boxes to cover T_+.
Out of $(X_i)_{i\in[m]}$ roughly mr^d in each box.
At least one per box if $mr^d \gtrsim \log q \Longleftarrow \boxed{r \overset{!}{\gtrsim} \left(\frac{\log m}{m}\right)^{1/d}}$.
Have $\hat{p} \leq \alpha$ if $S_k \gtrsim \sqrt{k\{\log((\log n)/\alpha)\}}$ (Howard et al, 2021).

Power

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta)\left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{m\log_+ n}{\alpha}\right)\right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m}\right)^{1/d}\right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.



Let $r > 0$ be small.

$\mu\{\mathcal{X}_\tau(\eta) \setminus T_+\} = \mu(T_-) \approx 4r \approx r$.

Need $q :\approx 1/r^{d-1}$ boxes to cover T_+.

Out of $(X_i)_{i\in[m]}$ roughly mr^d in each box.

At least one per box if $mr^d \gtrsim \log q \Longleftarrow \boxed{r \overset{!}{\gtrsim} \left(\frac{\log m}{m}\right)^{1/d}}$.

Have $\hat{p} \leq \alpha$ if $S_k \gtrsim \sqrt{k\{\log((\log n)/\alpha)\}}$ (Howard et al, 2021).

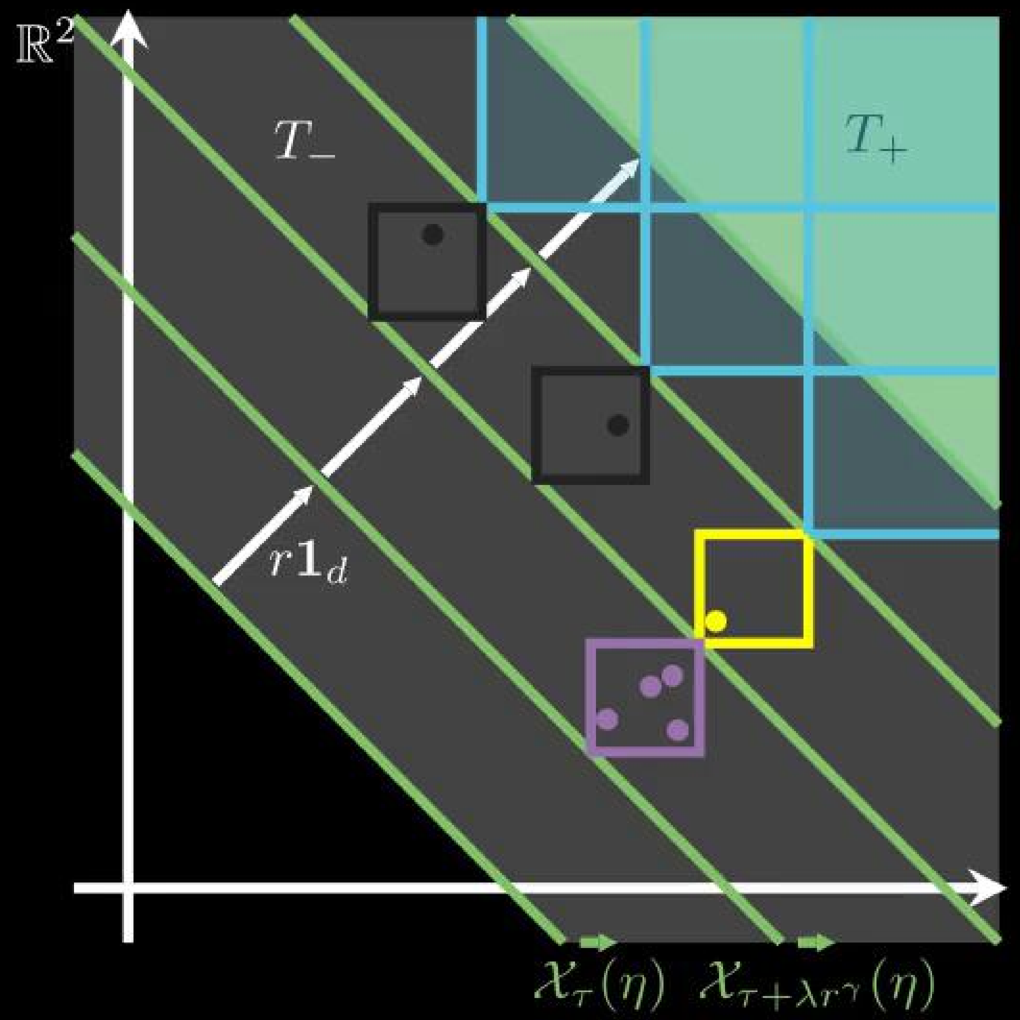
Box contains roughly nr^d points with expected response $\geq \tau + \lambda r^\gamma$.

Power

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta) \left\{ \left(\frac{\sigma^2}{n\lambda^2} \log_+ \left(\frac{m \log_+ n}{\alpha} \right) \right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m} \right)^{1/d} \right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.



Let $r > 0$ be small.

$\mu\{\mathcal{X}_\tau(\eta) \setminus T_+\} = \mu(T_-) \approx 4r \approx r$.

Need $q :\approx 1/r^{d-1}$ boxes to cover T_+.

Out of $(X_i)_{i\in[m]}$ roughly mr^d in each box.

At least one per box if $mr^d \gtrsim \log q \Longleftarrow \boxed{r \overset{!}{\gtrsim} \left(\frac{\log m}{m}\right)^{1/d}}$.

Have $\hat{p} \leq \alpha$ if $S_k \gtrsim \sqrt{k\{\log((\log n)/\alpha)\}}$ (Howard et al, 2021).

Box contains roughly nr^d points with expected response $\geq \tau + \lambda r^\gamma$.

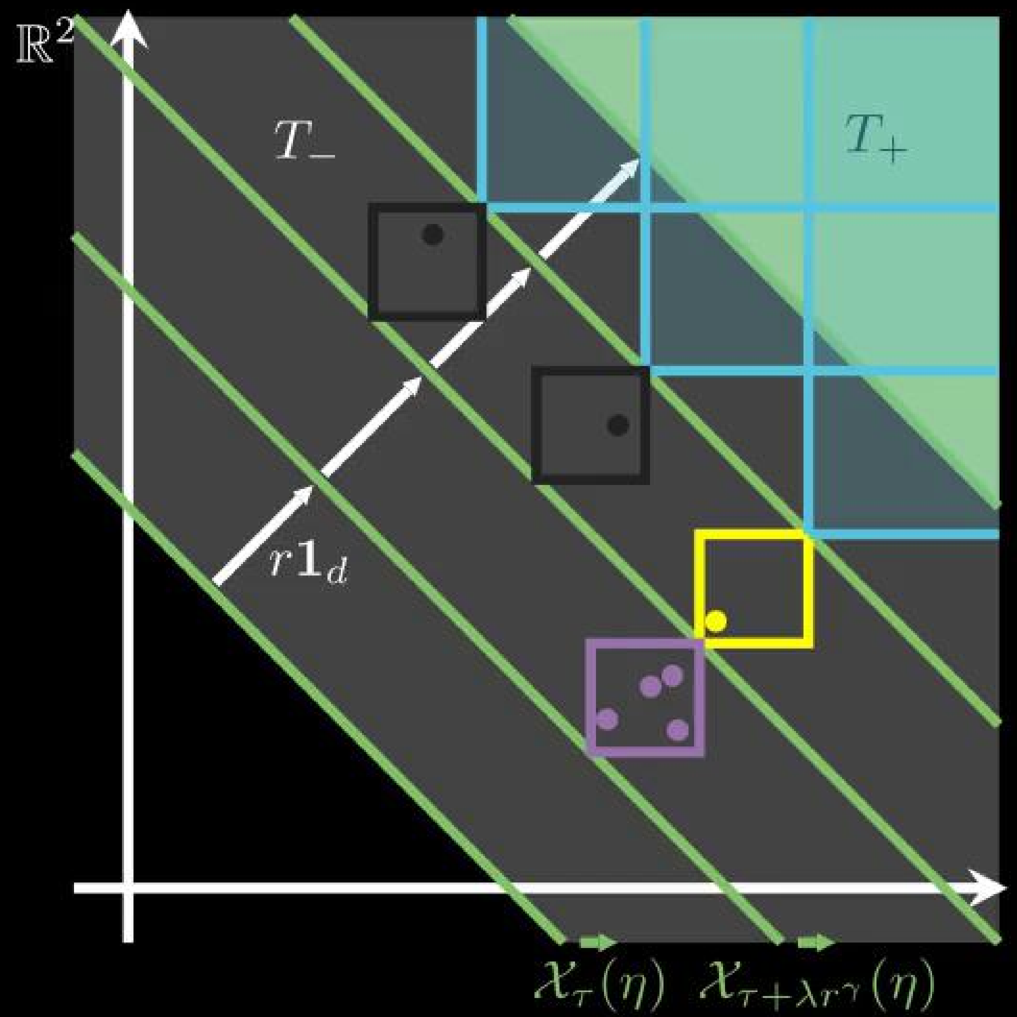
For $k \approx nr^d$, have $S_k \gtrsim nr^d \cdot \frac{\lambda r^\gamma}{\sigma} \overset{!}{\gtrsim} \sqrt{nr^d\{\log((\log n)/\alpha)\}}$.

Power

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta)\left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{m\log_+ n}{\alpha}\right)\right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m}\right)^{1/d}\right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.



Let $r > 0$ be small.

$\mu\{\mathcal{X}_\tau(\eta) \setminus T_+\} = \mu(T_-) \approx 4r \approx r$.

Need $q :\approx 1/r^{d-1}$ boxes to cover T_+.

Out of $(X_i)_{i\in[m]}$ roughly mr^d in each box.

At least one per box if $mr^d \gtrsim \log q \Longleftarrow \boxed{r \overset{!}{\gtrsim} \left(\frac{\log m}{m}\right)^{1/d}}$.

Have $\hat{p} \leq \alpha$ if $S_k \gtrsim \sqrt{k\{\log((\log n)/\alpha)\}}$ (Howard et al, 2021).

Box contains roughly nr^d points with expected response $\geq \tau + \lambda r^\gamma$.

For $k \approx nr^d$, have $S_k \gtrsim nr^d \cdot \frac{\lambda r^\gamma}{\sigma} \overset{!}{\gtrsim} \sqrt{nr^d\{\log((\log n)/\alpha)\}}$.

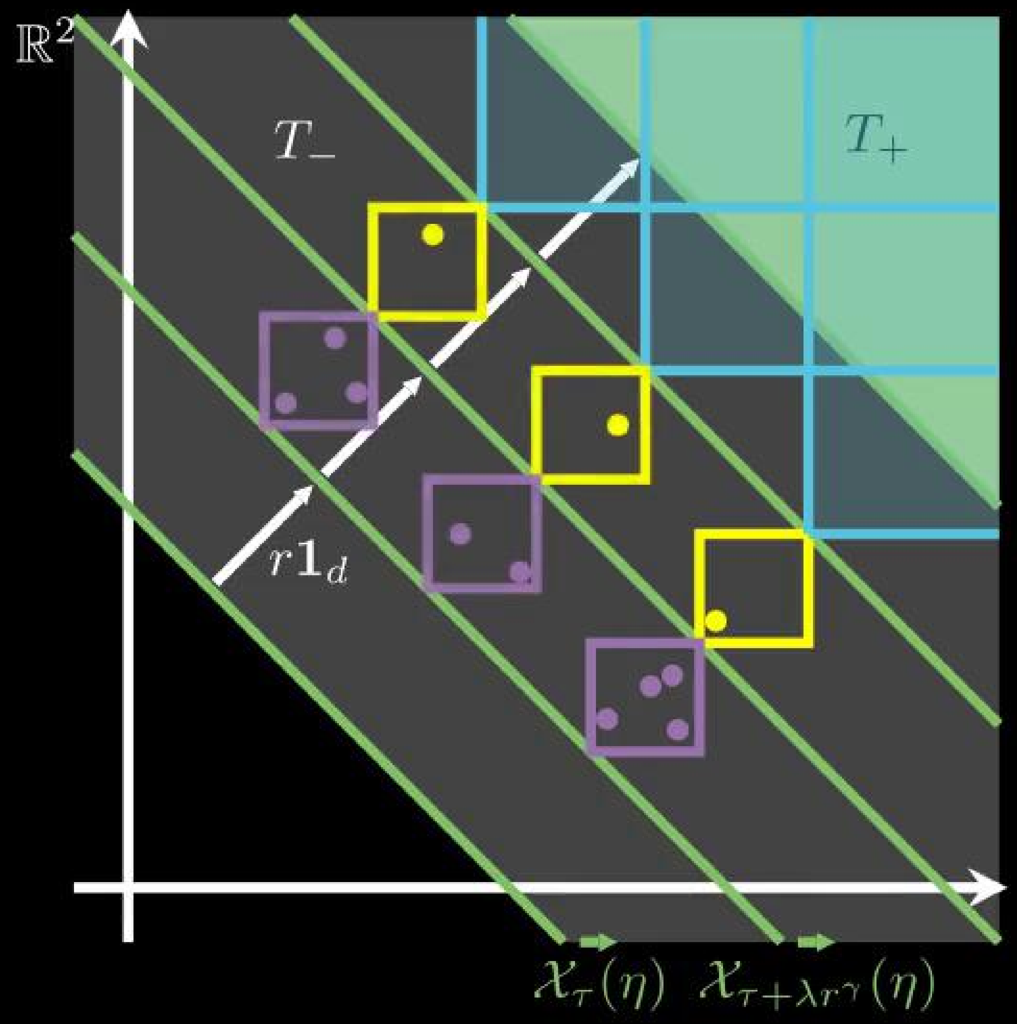
Sufficient: $\boxed{r \overset{!}{\gtrsim} \left\{\frac{\sigma^2}{n\lambda^2}\log\left(\frac{\log n}{\alpha}\right)\right\}^{1/(2\gamma+d)}}$.

Power

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta) \left\{ \left(\frac{\sigma^2}{n\lambda^2} \log_+\left(\frac{m \log_+ n}{\alpha} \right) \right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m} \right)^{1/d} \right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.



Let $r > 0$ be small.

$\mu\{\mathcal{X}_\tau(\eta) \setminus T_+\} = \mu(T_-) \approx 4r \approx r$.

Need $q :\approx 1/r^{d-1}$ boxes to cover T_+.

Out of $(X_i)_{i\in[m]}$ roughly mr^d in each box.

At least one per box if $mr^d \gtrsim \log q \Longleftarrow \boxed{r \overset{!}{\gtrsim} \left(\frac{\log m}{m}\right)^{1/d}}$.

Have $\hat{p} \leq \alpha$ if $S_k \gtrsim \sqrt{k\{\log((\log n)/\alpha)\}}$ (Howard et al, 2021).

Box contains roughly nr^d points with expected response $\geq \tau + \lambda r^\gamma$.

For $k \approx nr^d$, have $S_k \gtrsim nr^d \cdot \frac{\lambda r^\gamma}{\sigma} \overset{!}{\gtrsim} \sqrt{nr^d\{\log((\log n)/\alpha)\}}$.

Sufficient: $\boxed{r \overset{!}{\gtrsim} \left\{ \frac{\sigma^2}{n\lambda^2} \log\left(\frac{m \log n}{\alpha} \right) \right\}^{1/(2\gamma+d)}}$.

Lower bound

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m \in [n]$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta) \left\{ \left(\frac{\sigma^2}{n\lambda^2} \log_+ \left(\frac{m \log_+ n}{\alpha} \right) \right)^{1/(2\gamma+d)} + \left(\frac{\log_+ m}{m} \right)^{1/d} \right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.

Lower bound

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m = n$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta)\left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{n\log_+ n}{\alpha}\right)\right)^{1/(2\gamma+d)} + \left(\frac{\log_+ n}{n}\right)^{1/d}\right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.

Lower bound

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m = n$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta) \biggl\{ \biggl(\frac{\sigma^2}{n\lambda^2} \log_+ \Bigl(\frac{n \log_+ n}{\alpha} \Bigr) \biggr)^{1/(2\gamma+d)} + \Bigl(\frac{\log_+ n}{n} \Bigr)^{1/d} \biggr\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.

Theorem. $\exists c(d,\gamma) \in (0,1)$, such that for any $n \geq 1$ and $\alpha \in (0,1)$,

$$\inf_{\hat{A} \in \hat{\mathcal{A}}_n(\tau,\alpha,\mathcal{P}_{\mathrm{MonReg}})} \sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}) \geq c(d,\gamma) \biggl[1 \wedge \biggl\{ \biggl(\frac{\sigma^2}{n\lambda^2} \log_+ \Bigl(\frac{1}{5\alpha} \Bigr) \biggr)^{1/(2\gamma+d)} + \frac{1}{n^{1/d}} \biggr\} \biggr].$$

Lower bound

Theorem. $\exists C(d,\theta) \geq 1$, such that for any $n \geq 1$, $m = n$ and $\alpha \in (0,1)$,

$$\sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}^{\mathrm{ISS}}) \leq 1 \wedge C(d,\theta)\left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{n\log_+ n}{\alpha}\right)\right)^{1/(2\gamma+d)} + \left(\frac{\log_+ n}{n}\right)^{1/d}\right\},$$

where $\log_+ x := \log(x \vee e)$ for $x \in \mathbb{R}$.

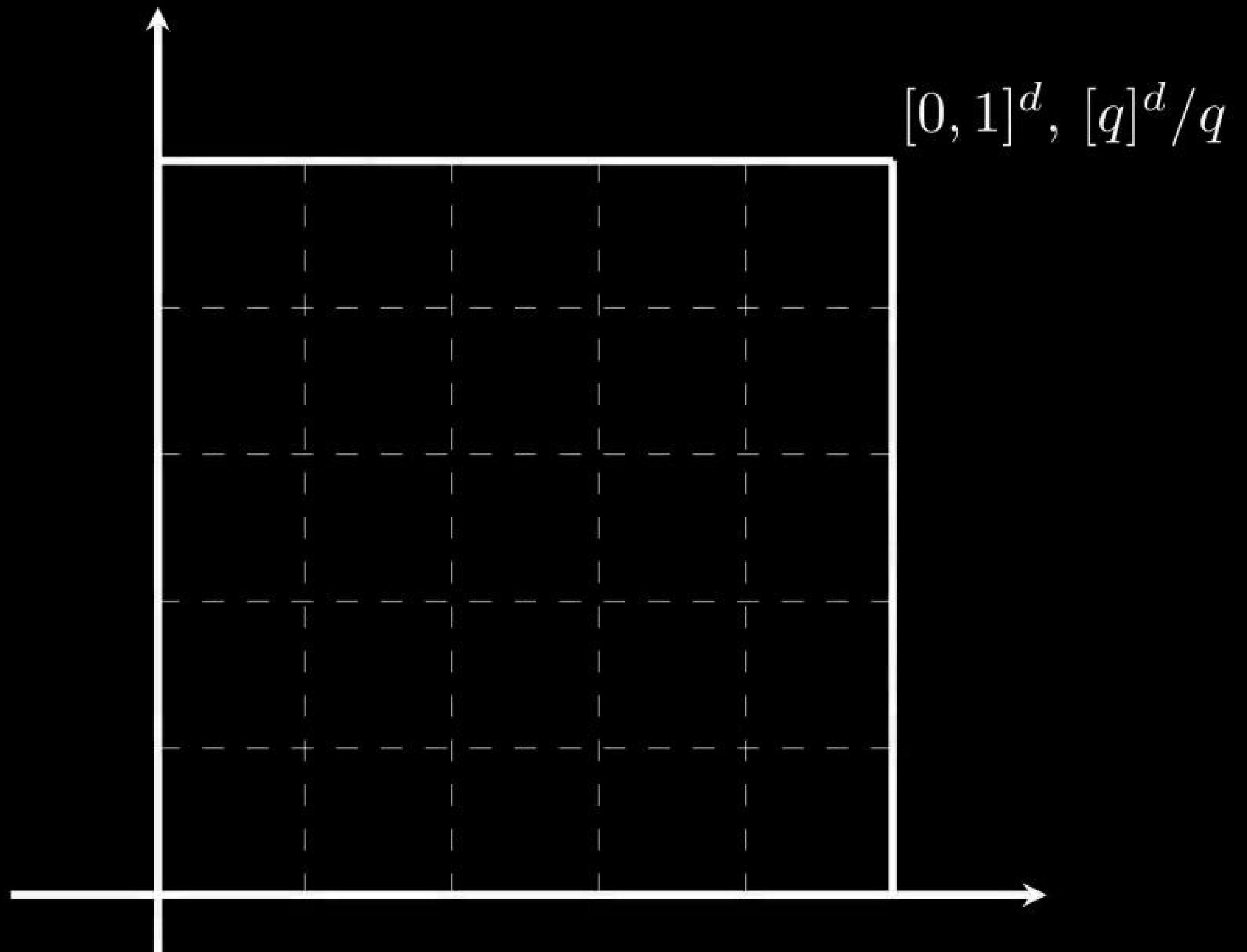
Theorem. $\exists c(d,\gamma) \in (0,1)$, such that for any $n \geq 1$ and $\alpha \in (0,1)$,

$$\inf_{\hat{A} \in \hat{\mathcal{A}}_n(\tau,\alpha,\mathcal{P}_{\mathrm{MonReg}})} \sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}) \geq c(d,\gamma)\left[1 \wedge \left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{1}{5\alpha}\right)\right)^{1/(2\gamma+d)} + \frac{1}{n^{1/d}}\right\}\right].$$

Lower bound

Theorem. $\exists c(d,\gamma) \in (0,1)$, such that for any $n \geq 1$ and $\alpha \in (0,1)$,

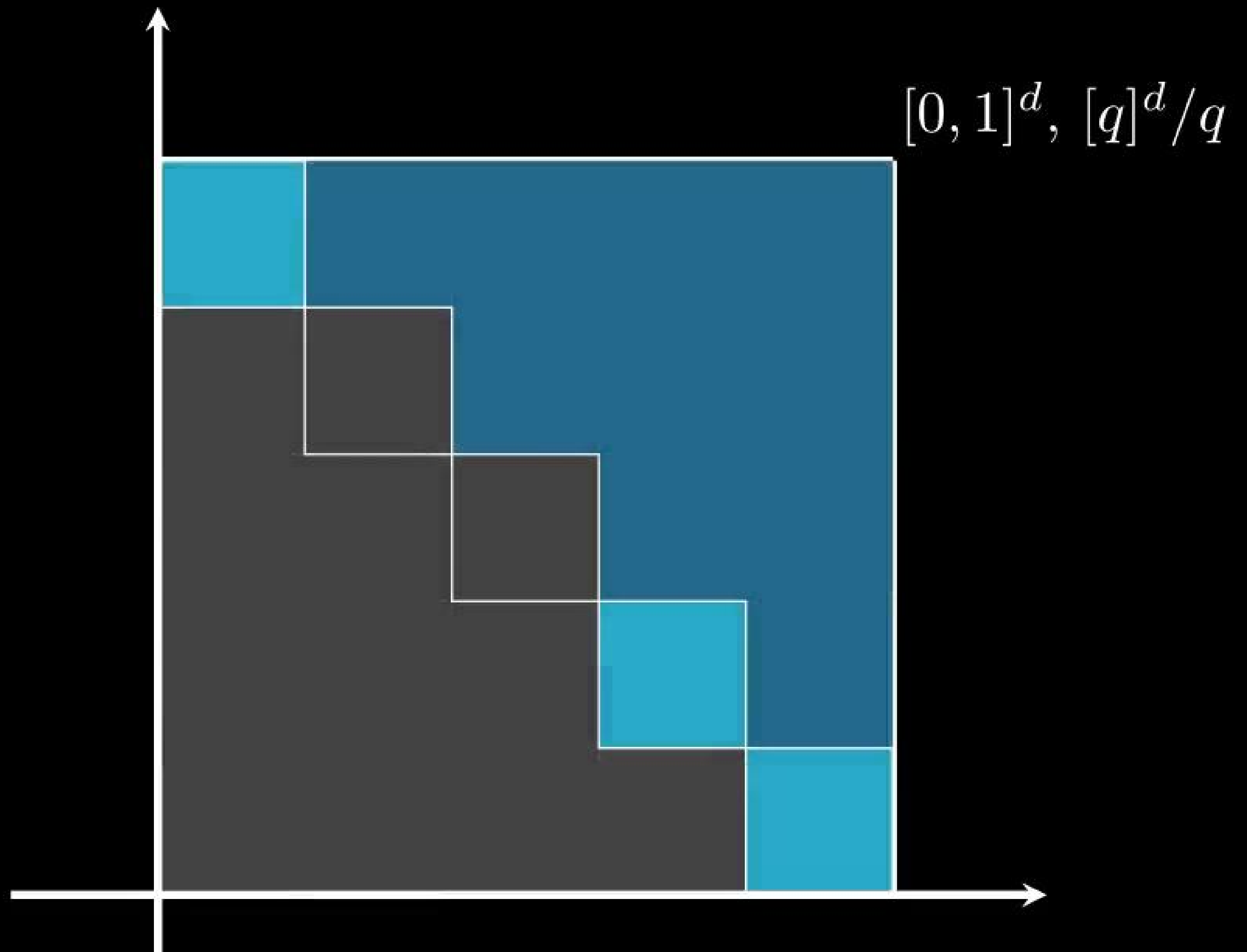
$$\inf_{\hat{A} \in \hat{\mathcal{A}}_n(\tau,\alpha,\mathcal{P}_{\mathrm{MonReg}})} \sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}) \geq c(d,\gamma)\left[1 \wedge \left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{1}{5\alpha}\right)\right)^{1/(2\gamma+d)} + \frac{1}{n^{1/d}}\right\}\right].$$



Lower bound

Theorem. $\exists c(d,\gamma) \in (0,1)$, such that for any $n \geq 1$ and $\alpha \in (0,1)$,

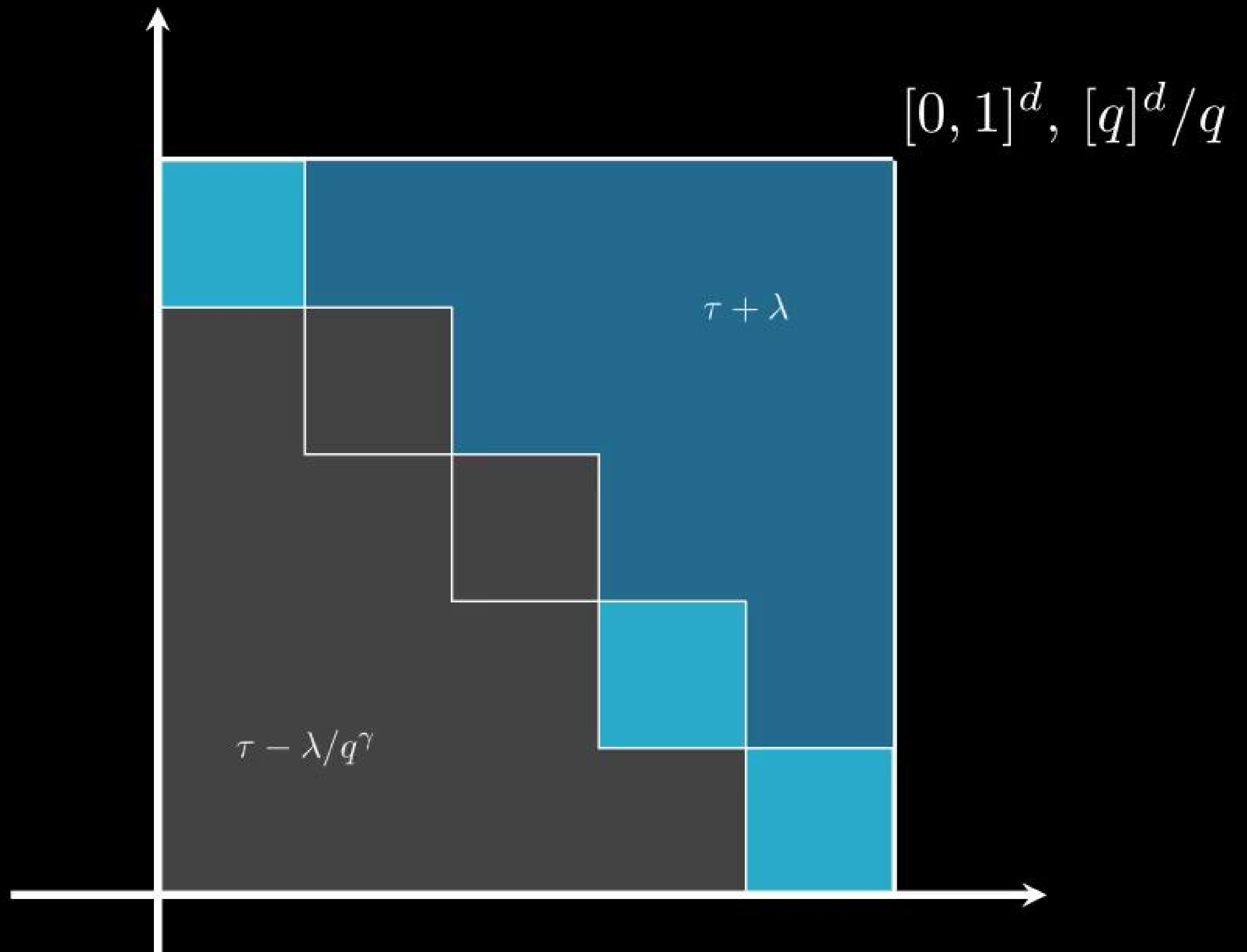
$$\inf_{\hat{A} \in \hat{\mathcal{A}}_n(\tau,\alpha,\mathcal{P}_{\mathrm{MonReg}})} \sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}) \geq c(d,\gamma)\left[1 \wedge \left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{1}{5\alpha}\right)\right)^{1/(2\gamma+d)} + \frac{1}{n^{1/d}}\right\}\right].$$



Lower bound

Theorem. $\exists c(d,\gamma) \in (0,1)$, such that for any $n \geq 1$ and $\alpha \in (0,1)$,

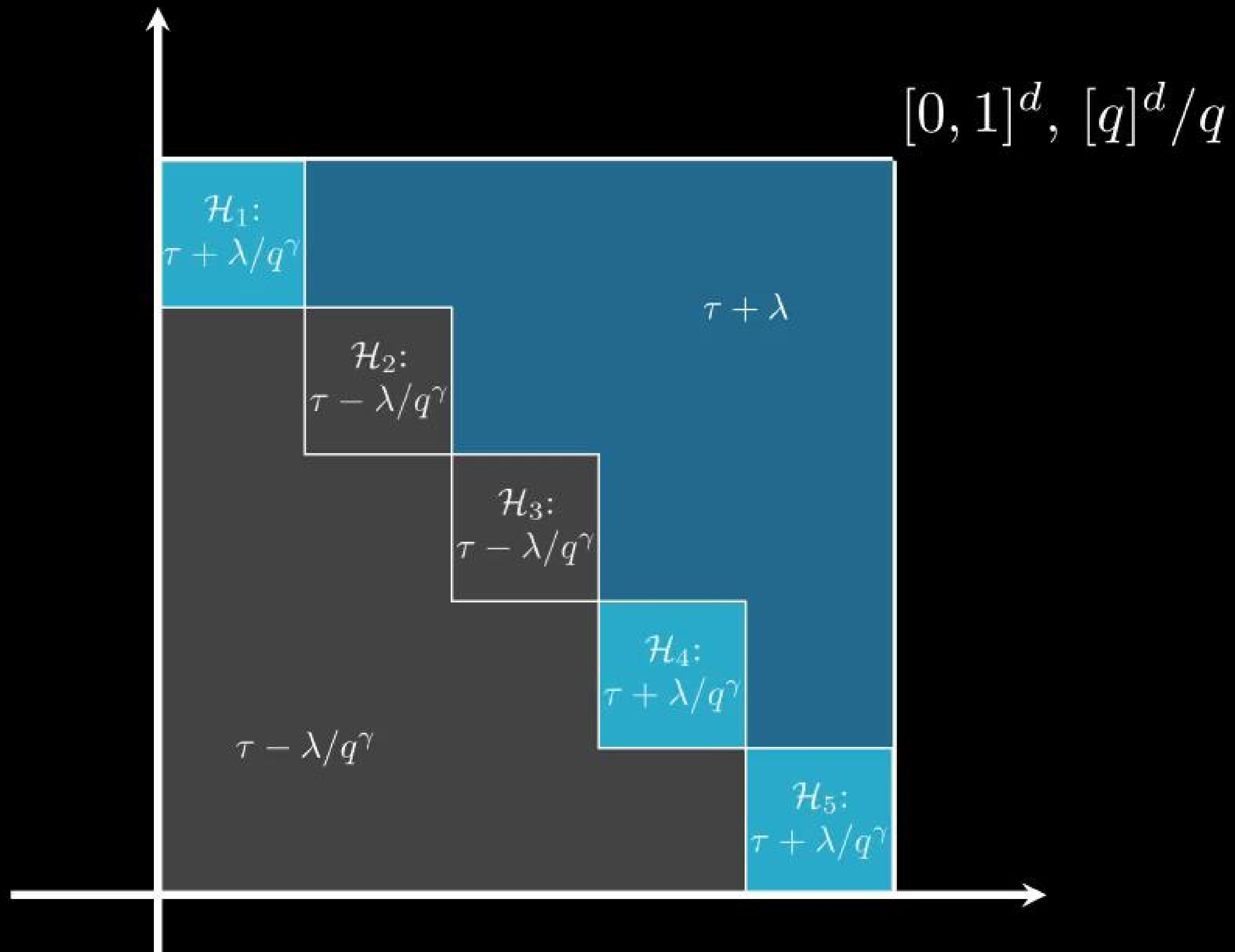
$$\inf_{\hat{A} \in \hat{\mathcal{A}}_n(\tau,\alpha,\mathcal{P}_{\mathrm{MonReg}})} \sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}) \geq c(d,\gamma)\left[1 \wedge \left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{1}{5\alpha}\right)\right)^{1/(2\gamma+d)} + \frac{1}{n^{1/d}}\right\}\right].$$



Lower bound

Theorem. $\exists c(d,\gamma) \in (0,1)$, such that for any $n \geq 1$ and $\alpha \in (0,1)$,

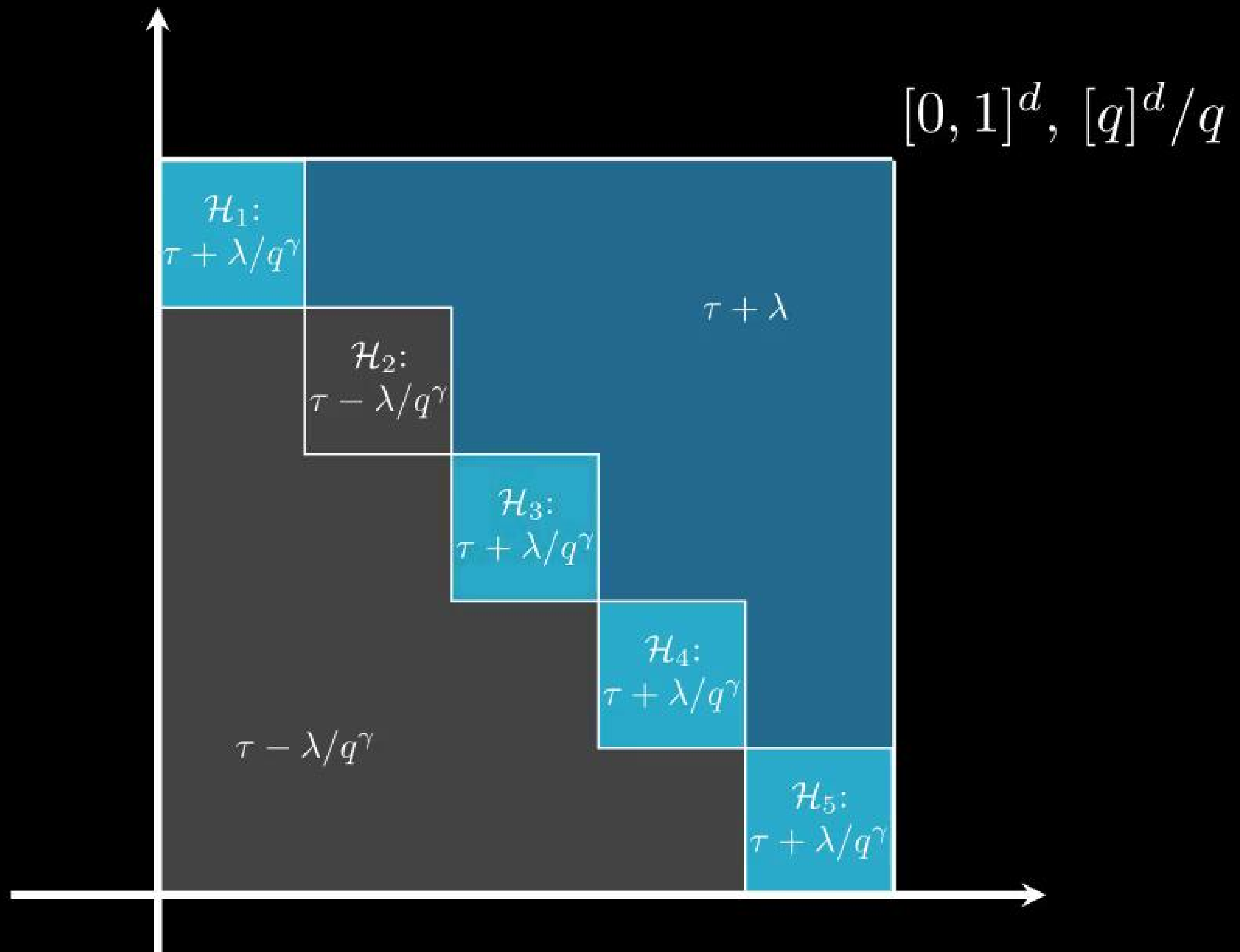
$$\inf_{\hat{A} \in \hat{\mathcal{A}}_n(\tau,\alpha,\mathcal{P}_{\mathrm{MonReg}})} \sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}) \geq c(d,\gamma)\left[1 \wedge \left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{1}{5\alpha}\right)\right)^{1/(2\gamma+d)} + \frac{1}{n^{1/d}}\right\}\right].$$



Lower bound

Theorem. $\exists c(d,\gamma) \in (0,1)$, such that for any $n \geq 1$ and $\alpha \in (0,1)$,

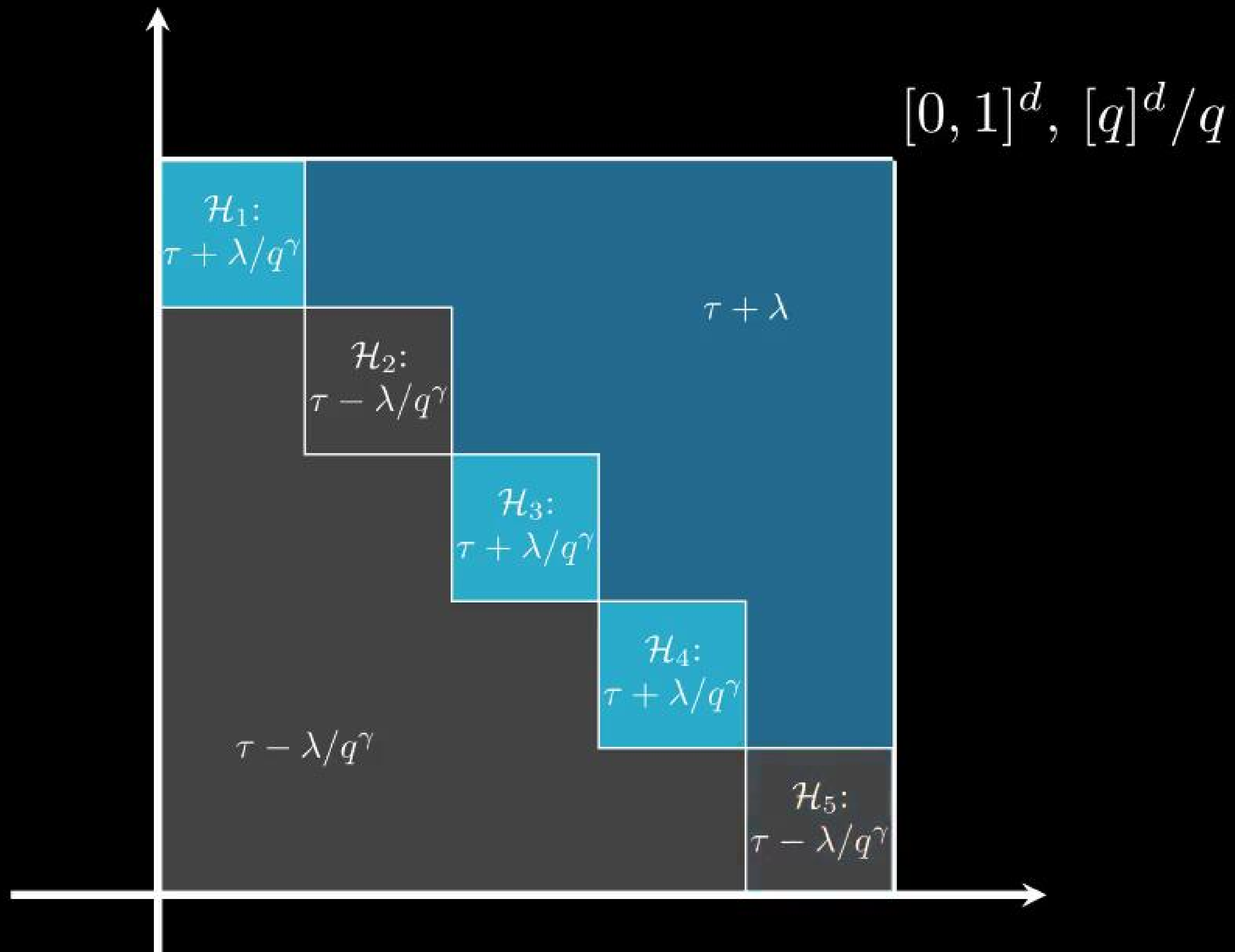
$$\inf_{\hat{A} \in \hat{\mathcal{A}}_n(\tau,\alpha,\mathcal{P}_{\mathrm{MonReg}})} \sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}) \geq c(d,\gamma)\left[1 \wedge \left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{1}{5\alpha}\right)\right)^{1/(2\gamma+d)} + \frac{1}{n^{1/d}}\right\}\right].$$



Lower bound

Theorem. $\exists c(d,\gamma) \in (0,1)$, such that for any $n \geq 1$ and $\alpha \in (0,1)$,

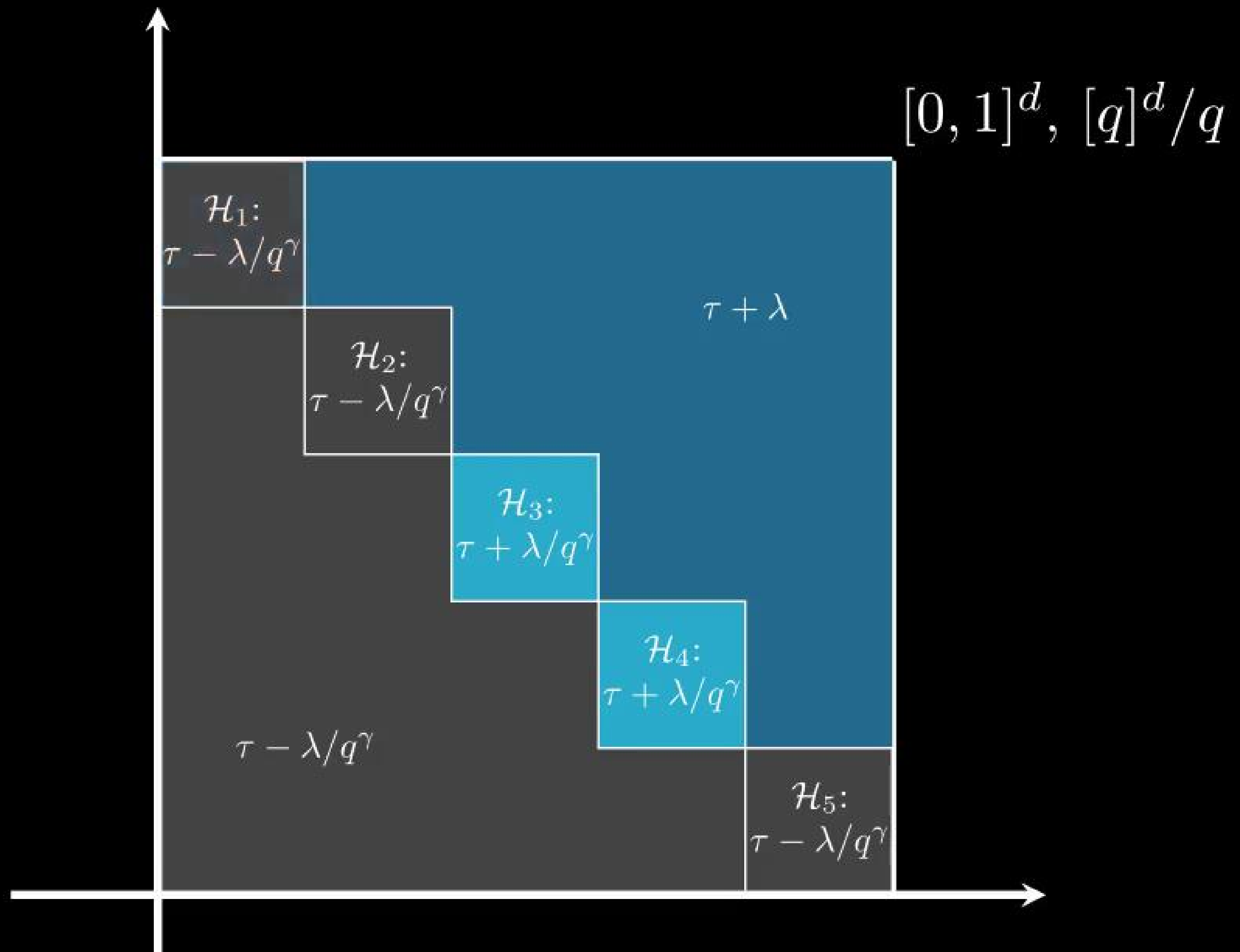
$$\inf_{\hat{A} \in \hat{\mathcal{A}}_n(\tau,\alpha,\mathcal{P}_{\mathrm{MonReg}})} \sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}) \geq c(d,\gamma)\left[1 \wedge \left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{1}{5\alpha}\right)\right)^{1/(2\gamma+d)} + \frac{1}{n^{1/d}}\right\}\right].$$



Lower bound

Theorem. $\exists c(d,\gamma) \in (0,1)$, such that for any $n \geq 1$ and $\alpha \in (0,1)$,

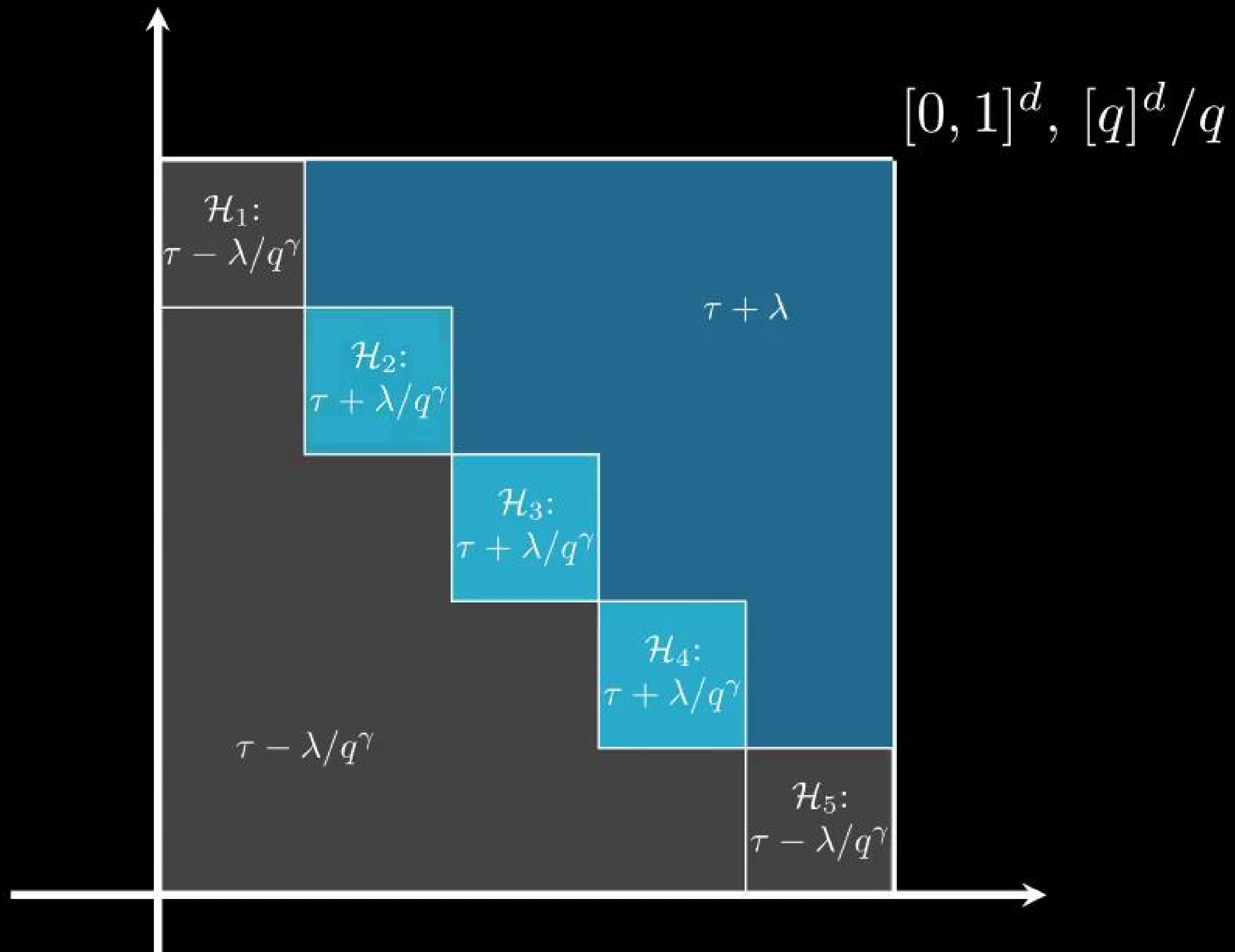
$$\inf_{\hat{A} \in \hat{\mathcal{A}}_n(\tau,\alpha,\mathcal{P}_{\mathrm{MonReg}})} \sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}) \geq c(d,\gamma)\left[1 \wedge \left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{1}{5\alpha}\right)\right)^{1/(2\gamma+d)} + \frac{1}{n^{1/d}}\right\}\right].$$



Lower bound

Theorem. $\exists c(d,\gamma) \in (0,1)$, such that for any $n \geq 1$ and $\alpha \in (0,1)$,

$$\inf_{\hat{A} \in \hat{\mathcal{A}}_n(\tau,\alpha,\mathcal{P}_{\mathrm{MonReg}})} \sup_{P \in \mathcal{P}_{\mathrm{MonReg}}} R_\tau(\hat{A}) \geq c(d,\gamma)\left[1 \wedge \left\{\left(\frac{\sigma^2}{n\lambda^2}\log_+\left(\frac{1}{5\alpha}\right)\right)^{1/(2\gamma+d)} + \frac{1}{n^{1/d}}\right\}\right].$$



Extension I: Gaussian noise

Consider adaptation to σ^2 when $Y - \eta(X)|X \sim \mathcal{N}(0, \sigma^2)$.

Extension I: Gaussian noise

Consider adaptation to σ^2 when $Y - \eta(X)|X \sim \mathcal{N}(0, \sigma^2)$.

Key idea: use an online split likelihood ratio test (Wasserman et al., 2020) for $H_0 : Y_{(k)}|X_{(k)} \sim \mathcal{N}(t_k, \sigma^2), t_k < \tau, \sigma > 0, \forall k \geq 1.$

Extension I: Gaussian noise

Consider adaptation to σ^2 when $Y - \eta(X)|X \sim \mathcal{N}(0, \sigma^2)$.

Key idea: use an online split likelihood ratio test (Wasserman et al., 2020) for $H_0 : Y_{(k)}|X_{(k)} \sim \mathcal{N}(t_k, \sigma^2), t_k < \tau, \sigma > 0, \forall k \geq 1$.

Definition. Let $\hat{\sigma}^2_{0,k} := \frac{1}{k}\sum_{j=1}^{k}\big(Y_{(j)} - \tau\big)_+^2$ and $\bar{Y}_{1,k} := \frac{1}{k}\sum_{j=1}^{k} Y_{(j)}$ for $k \in [n(x)]$ and $\hat{\sigma}^2_{1,k} := \frac{1}{k}\sum_{j=1}^{k}\big(Y_{(j)} - \bar{Y}_{1,k}\big)^2$ for $k \in \{2, \ldots, n(x)\}$. Denote $\bar{Y}_{1,0} := 0$, and $\hat{\sigma}^2_{1,k} := 1$ for $k \in \{0, 1\}$. For $k \in [n(x)]$, define

$$\bar{p}^k_\tau(x) := \frac{1}{\hat{\sigma}^k_{0,k} e^{k/2}} \cdot \prod_{j=1}^{k} \hat{\sigma}_{1,j-1} \exp\left\{ \frac{\big(Y_{(j)} - \bar{Y}_{1,j-1}\big)^2}{2\hat{\sigma}^2_{1,j-1}} \right\},$$

where $\bar{p}^k_\tau(x) := 1$ if $\hat{\sigma}_{0,k} = 0$, and $\bar{p}_\tau(x) := 1 \wedge \min_{k \in [n(x)]} \bar{p}^k_\tau(x)$.

Extension I: Gaussian noise

Consider adaptation to σ^2 when $Y - \eta(X) | X \sim \mathcal{N}(0, \sigma^2)$.

Key idea: use an online split likelihood ratio test (Wasserman et al., 2020) for $H_0 : Y_{(k)} | X_{(k)} \sim \mathcal{N}(t_k, \sigma^2), t_k < \tau, \sigma > 0, \forall k \geq 1$.

Definition. Let $\hat{\sigma}^2_{0,k} := \frac{1}{k}\sum_{j=1}^{k}\big(Y_{(j)} - \tau\big)_+^2$ and $\bar{Y}_{1,k} := \frac{1}{k}\sum_{j=1}^{k} Y_{(j)}$ for $k \in [n(x)]$ and $\hat{\sigma}^2_{1,k} := \frac{1}{k}\sum_{j=1}^{k}\big(Y_{(j)} - \bar{Y}_{1,k}\big)^2$ for $k \in \{2, \ldots, n(x)\}$. Denote $\bar{Y}_{1,0} := 0$, and $\hat{\sigma}^2_{1,k} := 1$ for $k \in \{0, 1\}$. For $k \in [n(x)]$, define

$$\bar{p}^k_\tau(x) := \frac{1}{\hat{\sigma}^k_{0,k} e^{k/2}} \cdot \prod_{j=1}^{k} \hat{\sigma}_{1,j-1} \exp\left\{ \frac{\big(Y_{(j)} - \bar{Y}_{1,j-1}\big)^2}{2\hat{\sigma}^2_{1,j-1}} \right\},$$

where $\bar{p}^k_\tau(x) := 1$ if $\hat{\sigma}_{0,k} = 0$, and $\bar{p}_\tau(x) := 1 \wedge \min_{k \in [n(x)]} \bar{p}^k_\tau(x)$.

Lemma. When $\eta(x) < \tau$, we have $\mathbb{P}\{\bar{p}_\tau(x) \leq t | \mathcal{D}_X\} \leq t$ for all $t \in (0, 1)$.

Extension II: Classification

Suppose $(X, Y) \sim P$ for some distribution P on $\mathbb{R}^d \times [0, 1]$ with increasing regression function.

Extension II: Classification

Suppose $(X, Y) \sim P$ for some distribution P on $\mathbb{R}^d \times [0, 1]$ with increasing regression function.

Consider the likelihood ratio martingale test for $H_0 : \eta(x_0) = \tau$ against $H_1 : \eta(x_0) = t$ mixed uniformly over $t \in [\tau, 1]$.

Extension II: Classification

Suppose $(X, Y) \sim P$ for some distribution P on $\mathbb{R}^d \times [0,1]$ with increasing regression function.

Consider the likelihood ratio martingale test for $H_0 : \eta(x_0) = \tau$ against $H_1 : \eta(x_0) = t$ mixed uniformly over $t \in [\tau, 1]$.

Definition. Let $\check{S}_k := \sum_{j=1}^k Y_{(j)}$ and define

$$\check{p}_\tau(x) := 1 \wedge \min_{k \in [n(x)]} \frac{\tau^{\check{S}_k}(1-\tau)^{n-\check{S}_k+1}}{\mathrm{B}(1-\tau; n-\check{S}_k+1, \check{S}_k+1)},$$

where for $z \in [0,1]$ and $a, b > 0$, we write $\mathrm{B}(z; a, b) := \int_0^z t^{a-1}(1-t)^{b-1}\,dt$ for the *incomplete beta function*.

Extension II: Classification

Suppose $(X, Y) \sim P$ for some distribution P on $\mathbb{R}^d \times [0,1]$ with increasing regression function.

Consider the likelihood ratio martingale test for $H_0 : \eta(x_0) = \tau$ against $H_1 : \eta(x_0) = t$ mixed uniformly over $t \in [\tau, 1]$.

Definition. Let $\check{S}_k := \sum_{j=1}^k Y_{(j)}$ and define

$$\check{p}_\tau(x) := 1 \wedge \min_{k \in [n(x)]} \frac{\tau^{\check{S}_k}(1-\tau)^{n-\check{S}_k+1}}{\mathrm{B}(1-\tau; n-\check{S}_k+1, \check{S}_k+1)},$$

where for $z \in [0,1]$ and $a, b > 0$, we write $\mathrm{B}(z; a, b) := \int_0^z t^{a-1}(1-t)^{b-1}\, dt$ for the *incomplete beta function.*

Lemma. If $\tau \in [0,1)$, $\eta(x) < \tau$, then $\mathbb{P}\{\check{p}_\tau(x) \le t | \mathcal{D}_X\} \le t$ for $t \in (0,1)$.

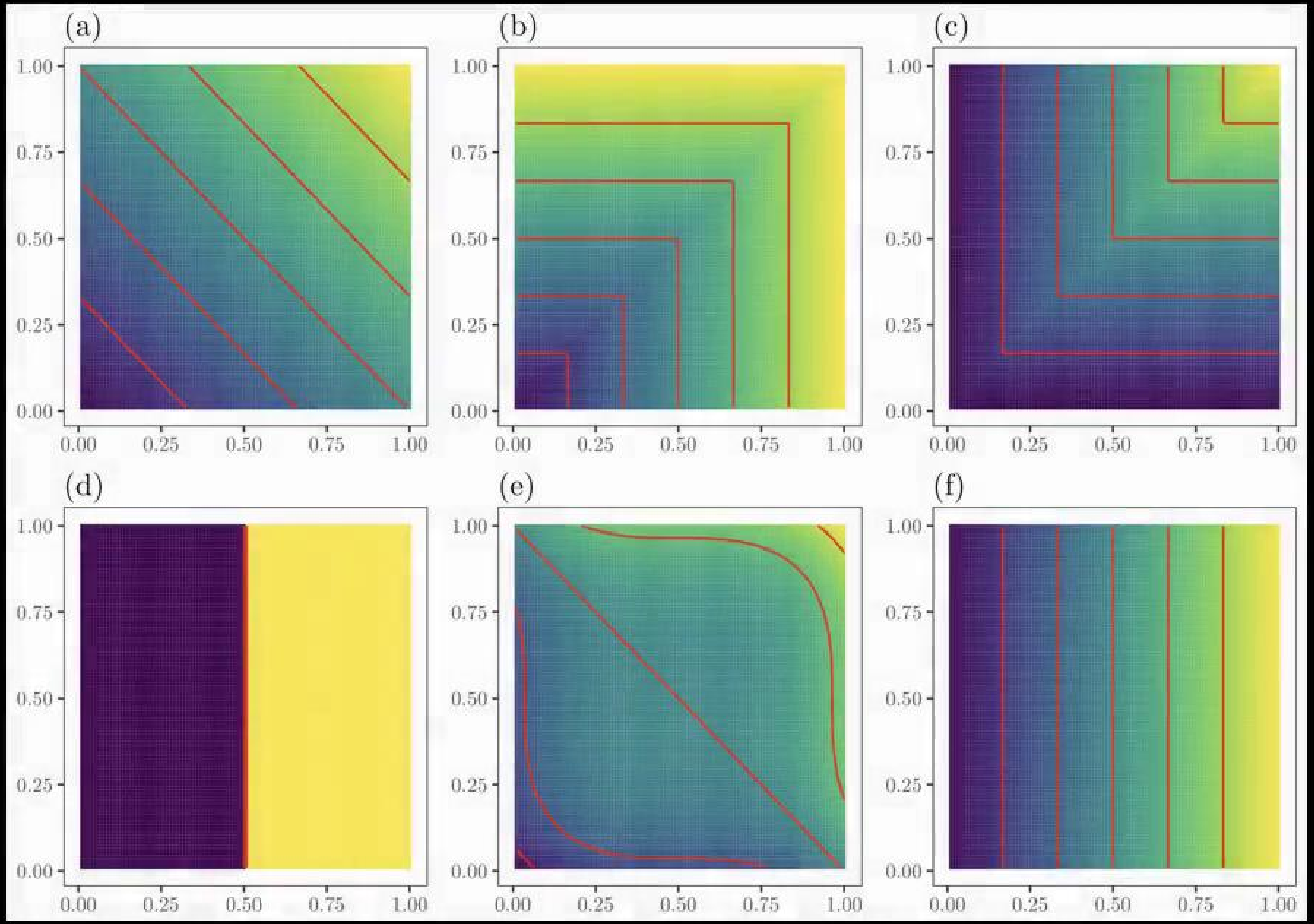
Simulations

We conduct a simulation study to compare with other choices of multiple testing procedure. We take $\mu = \mathrm{Unif}\big([0,1]^d\big)$, $Y - \eta(X)|X \sim \mathcal{N}(0, \sigma^2)$ and our regression functions η are obtained by rescaling f to $[0,1]$ on $[0,1]^d$:

Simulations

We conduct a simulation study to compare with other choices of multiple testing procedure. We take $\mu = \mathrm{Unif}\big([0,1]^d\big)$, $Y - \eta(X)|X \sim \mathcal{N}(0, \sigma^2)$ and our regression functions η are obtained by rescaling f to $[0,1]$ on $[0,1]^d$:

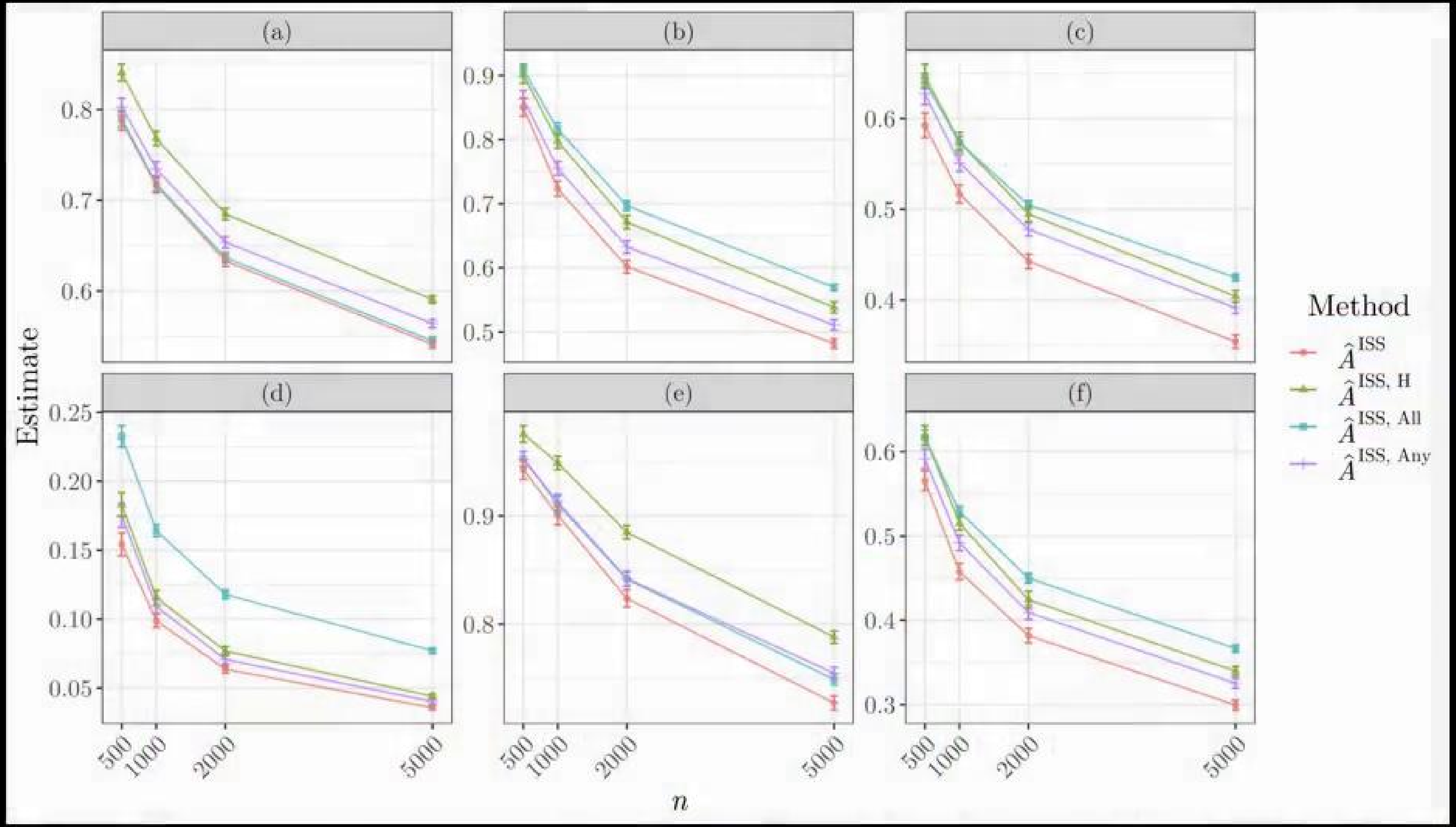
Label	Function f	τ	$\gamma(P)$
(a)	$\sum_{j=1}^d x^{(j)}$	$1/2$	1
(b)	$\max_{1\le j\le d} x^{(j)}$	$1/2^{1/d}$	1
(c)	$\min_{1\le j\le d} x^{(j)}$	$1 - 1/2^{1/d}$	1
(d)	$\mathbb{1}_{(0.5,1]}\big(x^{(1)}\big)$	$1/2$	0
(e)	$\sum_{j=1}^d \big(x^{(j)} - 0.5\big)^3$	$1/2$	3
(f)	$x^{(1)}$	$1/2$	1



Simulations

We conduct a simulation study to compare with other choices of multiple testing procedure. We take $\mu = \mathrm{Unif}\big([0,1]^d\big)$, $Y - \eta(X)|X \sim \mathcal{N}(0, \sigma^2)$ and our regression functions η are obtained by rescaling f to $[0,1]$ on $[0,1]^d$:

Label	Function f	τ	$\gamma(P)$
(a)	$\sum_{j=1}^{d} x^{(j)}$	$1/2$	1
(b)	$\max_{1\leq j\leq d} x^{(j)}$	$1/2^{1/d}$	1
(c)	$\min_{1\leq j\leq d} x^{(j)}$	$1 - 1/2^{1/d}$	1
(d)	$\mathbb{1}_{(0.5,1]}\big(x^{(1)}\big)$	$1/2$	0
(e)	$\sum_{j=1}^{d}\big(x^{(j)} - 0.5\big)^3$	$1/2$	3
(f)	$x^{(1)}$	$1/2$	1



Here, $d = 2$, $\sigma = 1/4$.
See also Meijer and Goeman (2015).