Isotonic subgroup selection

Manuel M. Müller

Statistical Laboratory, University of Cambridge

Basel Biometric Society Workshop on Controlled Subgroup Discovery

29th August 2024

Collaborators

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

Henry W. J. Reeve

University of Bristol

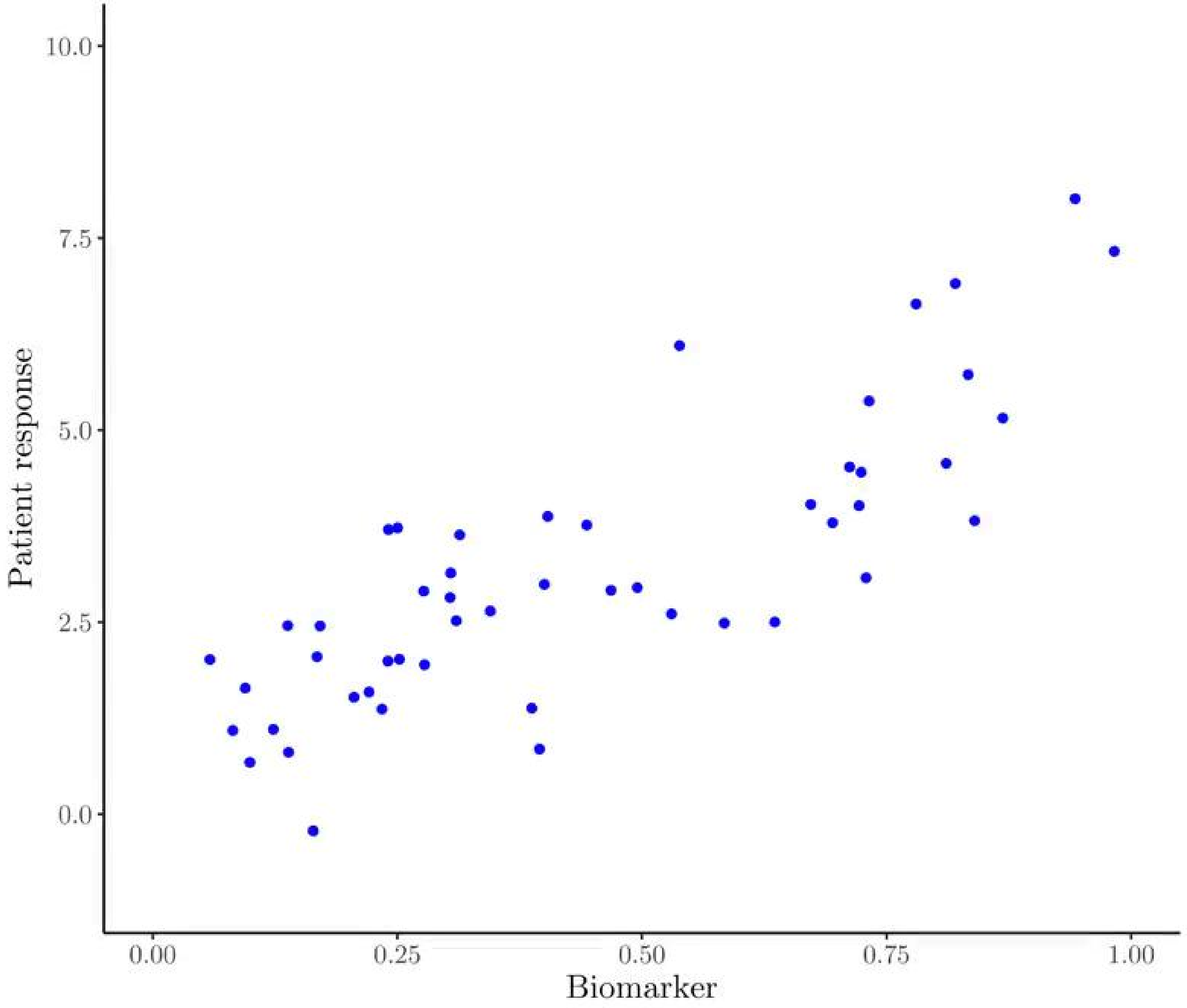
Timothy I. Cannings

University of Edinburgh

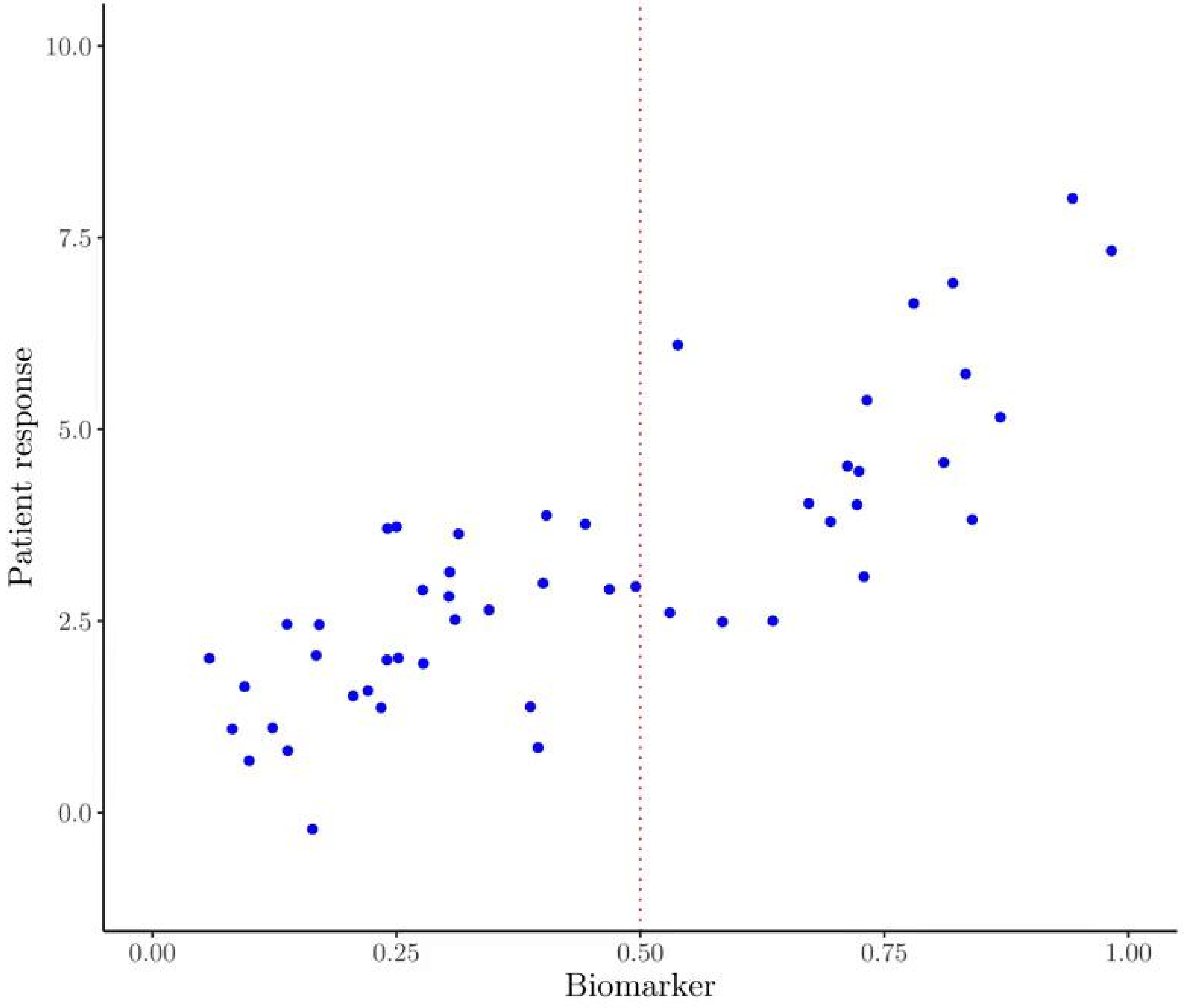
Richard J. Samworth

University of Cambridge

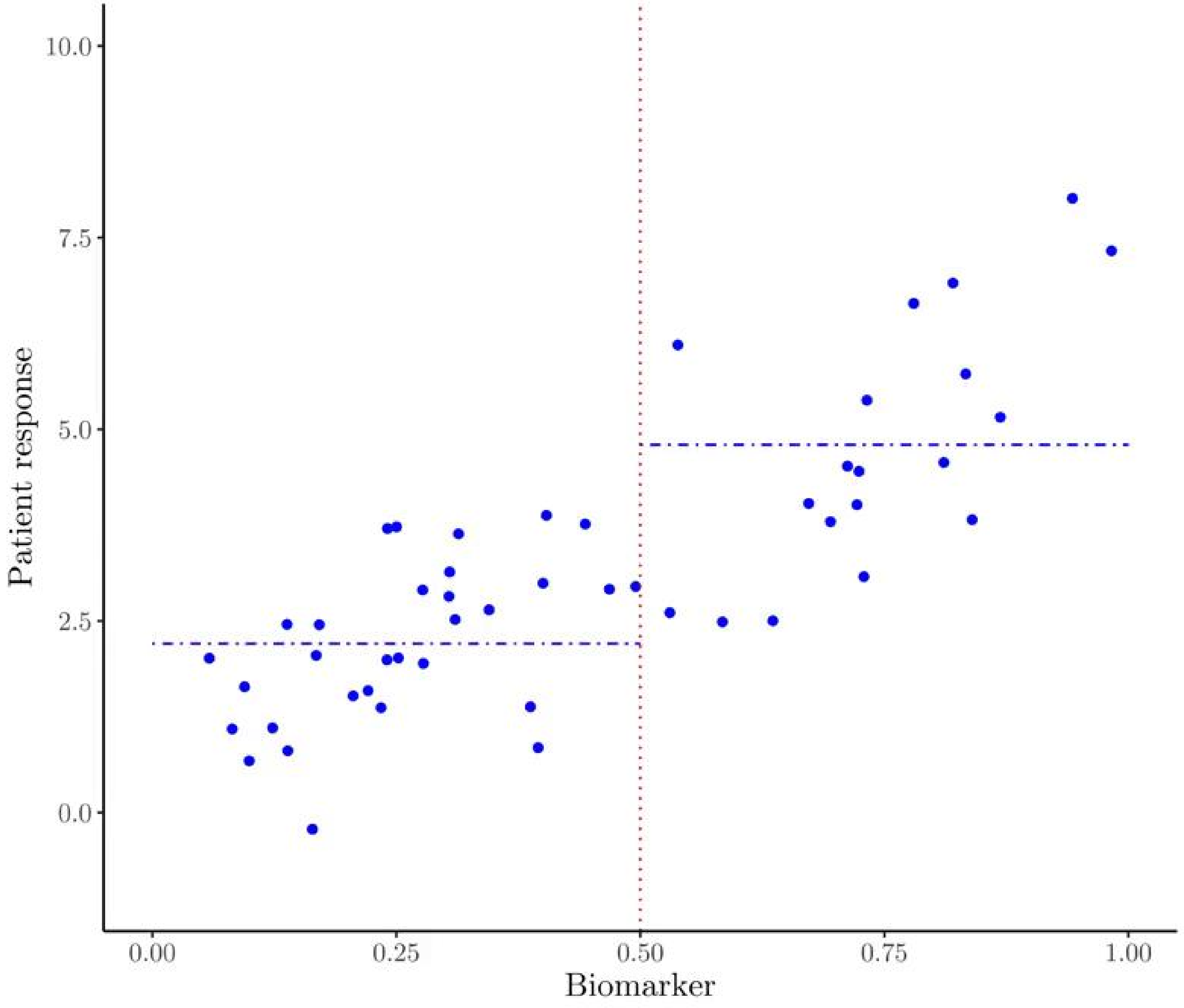
What is a subgroup?



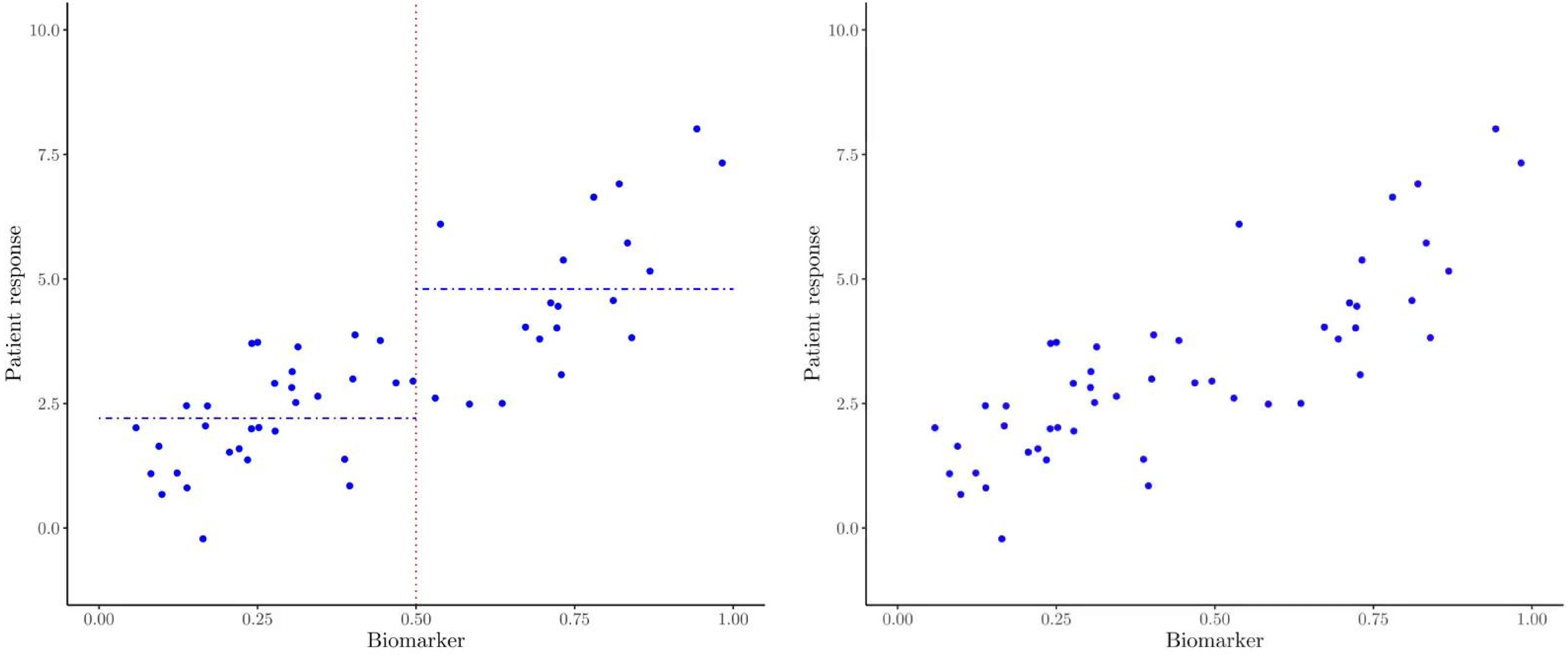
Patient response
10.0
7.5
5.0
2.5
0.0
0.00
0.25
0.50
0.75
1.00
Biomarker



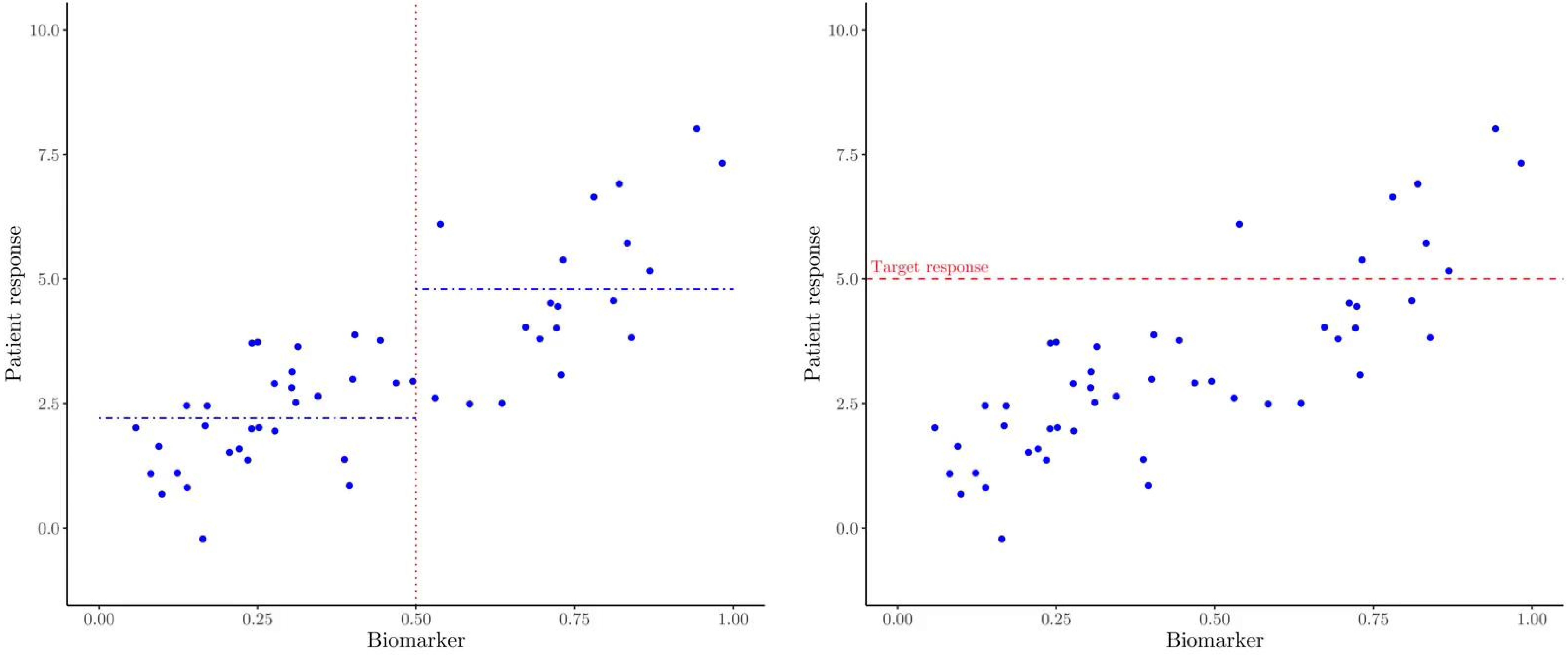
10.0
7.5
5.0
2.5
0.0
Patient response
0.00
0.25
0.50
0.75
1.00
Biomarker

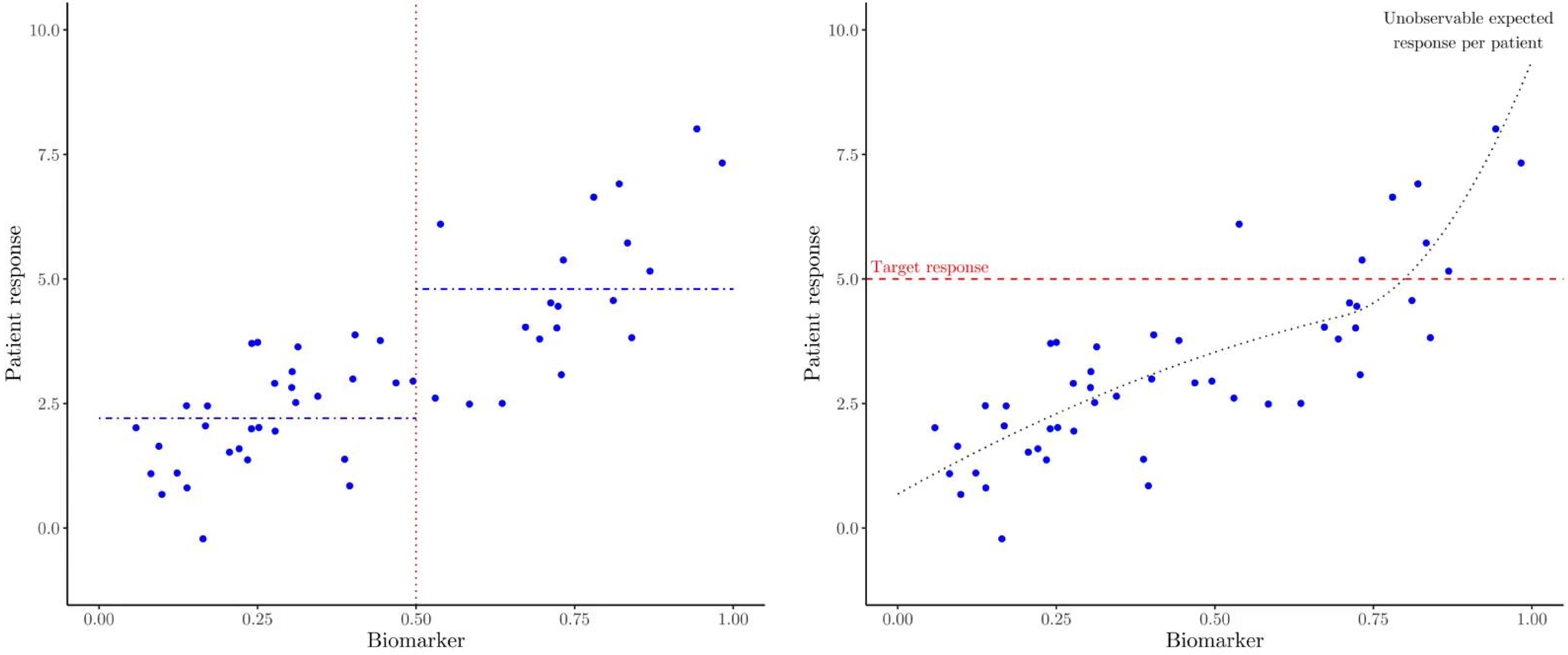


10.0
7.5
5.0
2.5
0.0
Patient response
0.00
0.25
0.50
0.75
1.00
Biomarker

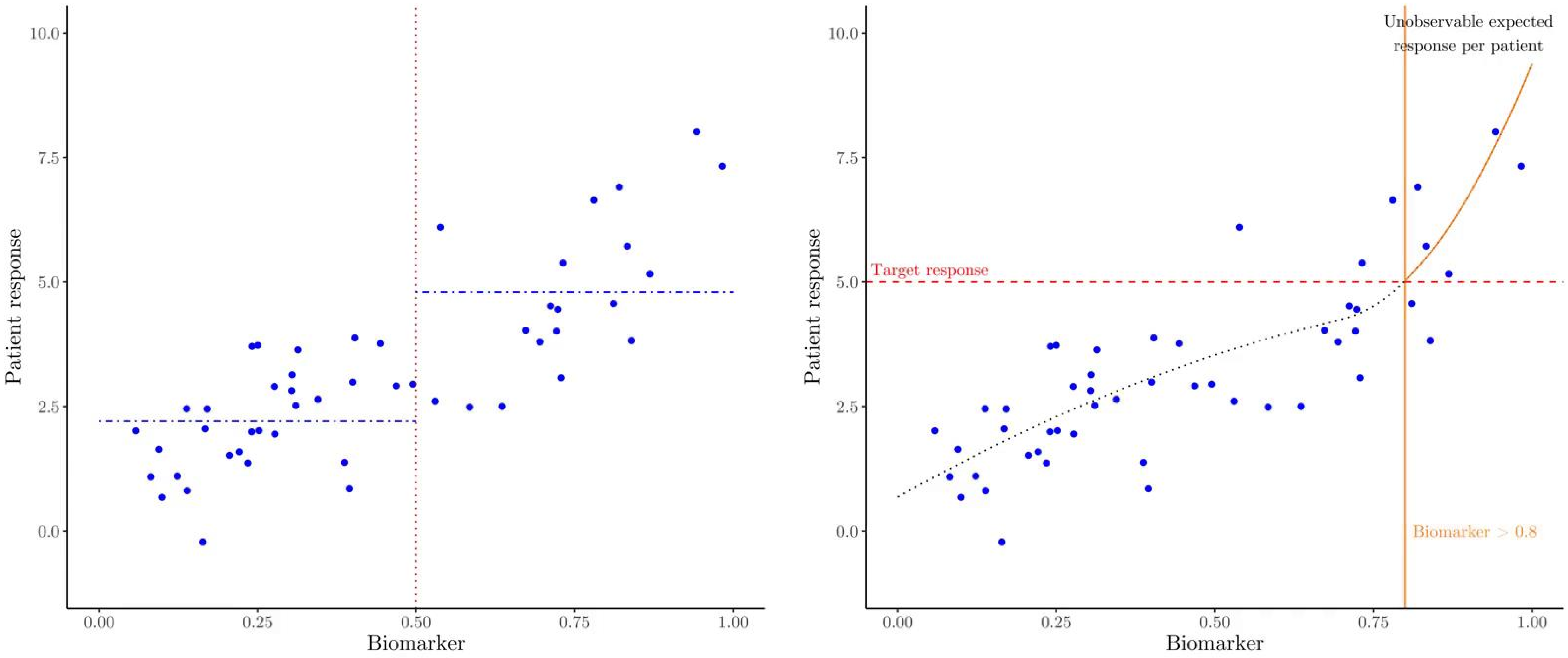


Patient response
10.0
7.5
5.0
2.5
0.0
0.00
0.25
0.50
0.75
1.00
Biomarker
Patient response
10.0
7.5
5.0
2.5
0.0
0.00
0.25
0.50
0.75
1.00
Biomarker


10.0
7.5
5.0
2.5
0.0
Patient response
0.00
0.25
0.50
0.75
1.00
Biomarker
Target response

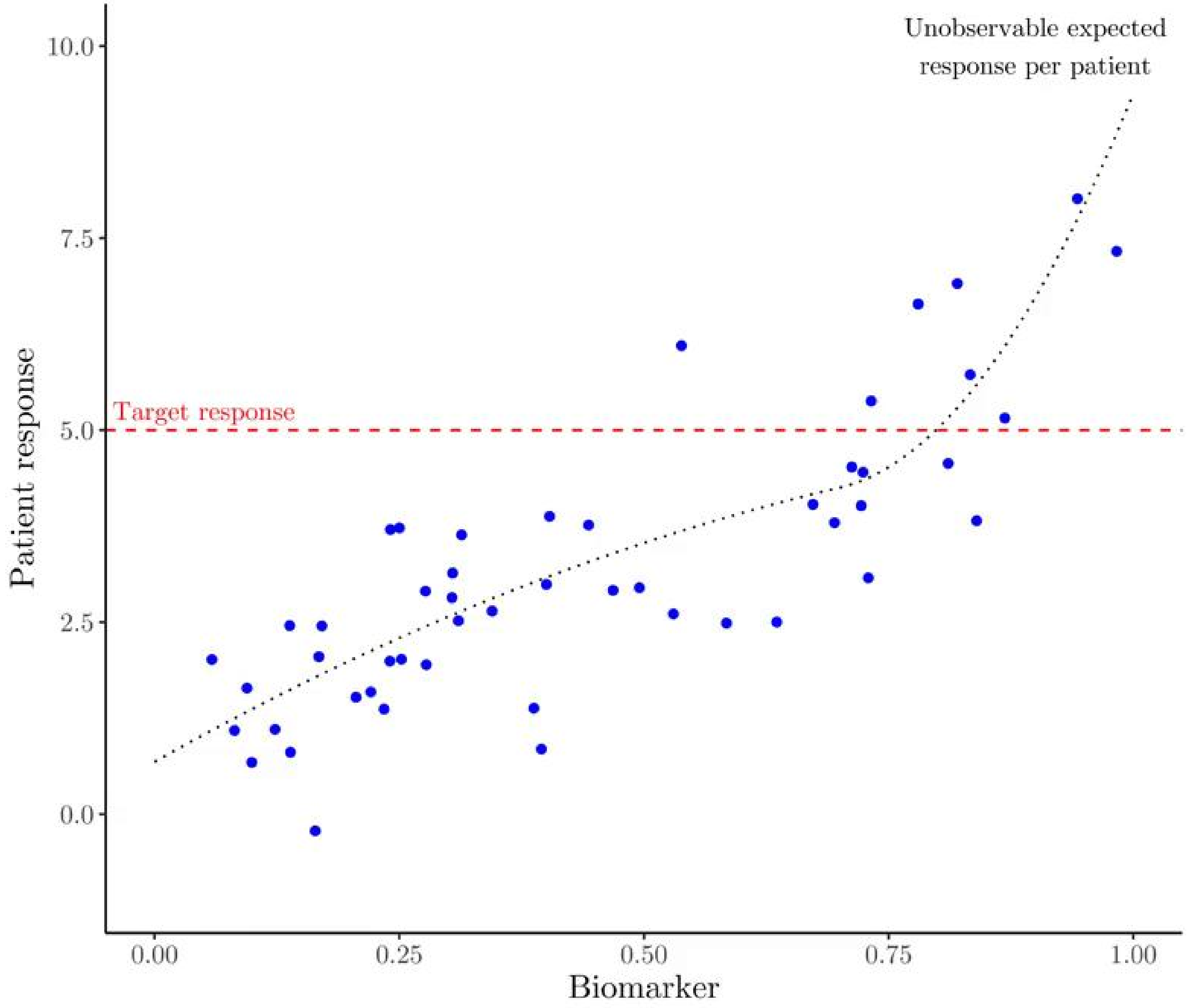


Patient response
Biomarker
0.00
0.25
0.50
0.75
1.00
0.0
2.5
5.0
7.5
10.0
Target response
Unobservable expected response per patient



Patient response
Biomarker
10.0
7.5
5.0
2.5
0.0
0.00
0.25
0.50
0.75
1.00
Unobservable expected
response per patient
Target response
Biomarker > 0.8

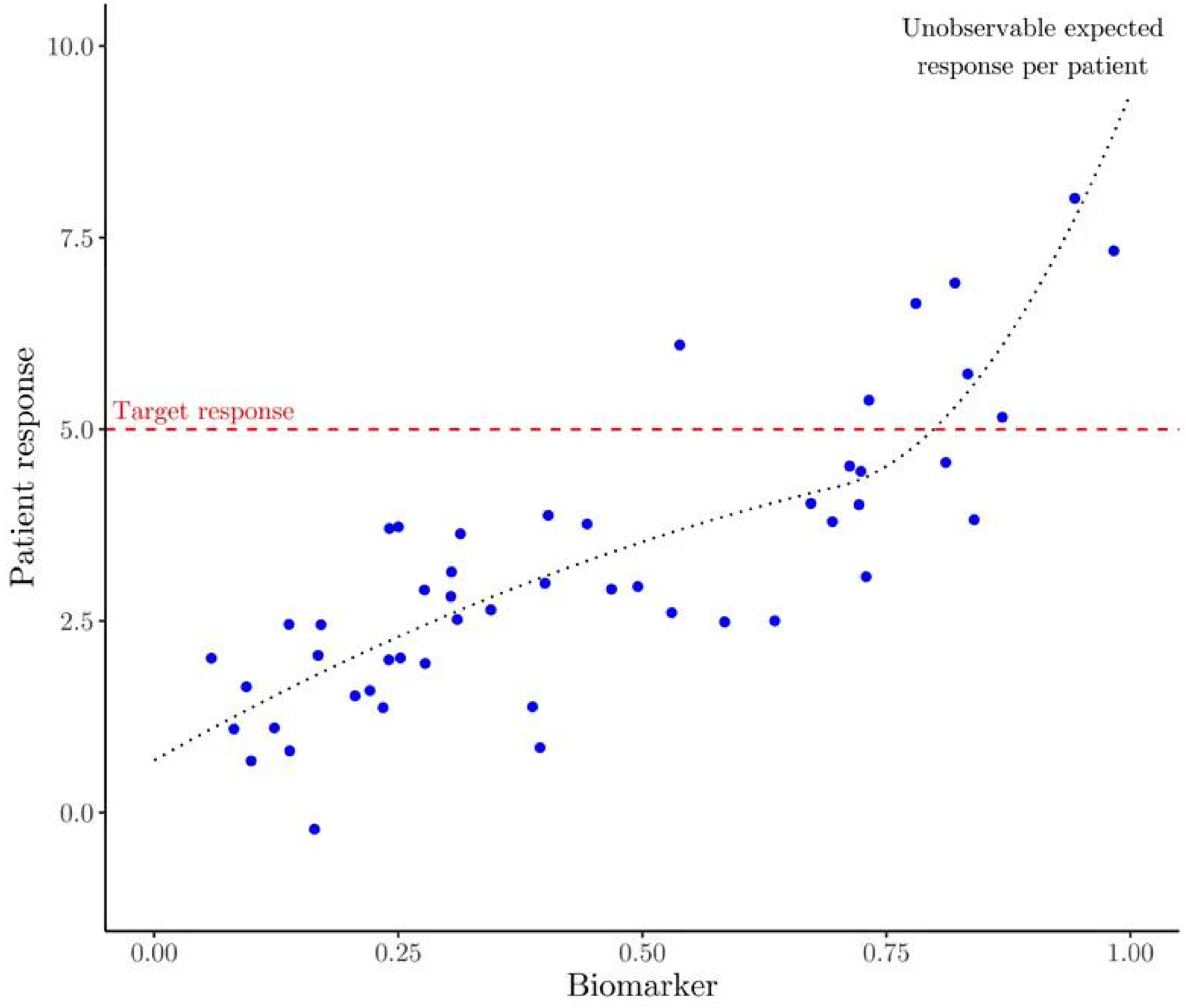
Subgroup selection



Subgroup selection

Data. Independent and identically distributed covariate-response pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ with values in $\mathbb{R}^d \times \mathbb{R}$ and unknown population regression function

$$\eta(x) := \mathbb{E}(Y_1 | X_1 = x).$$

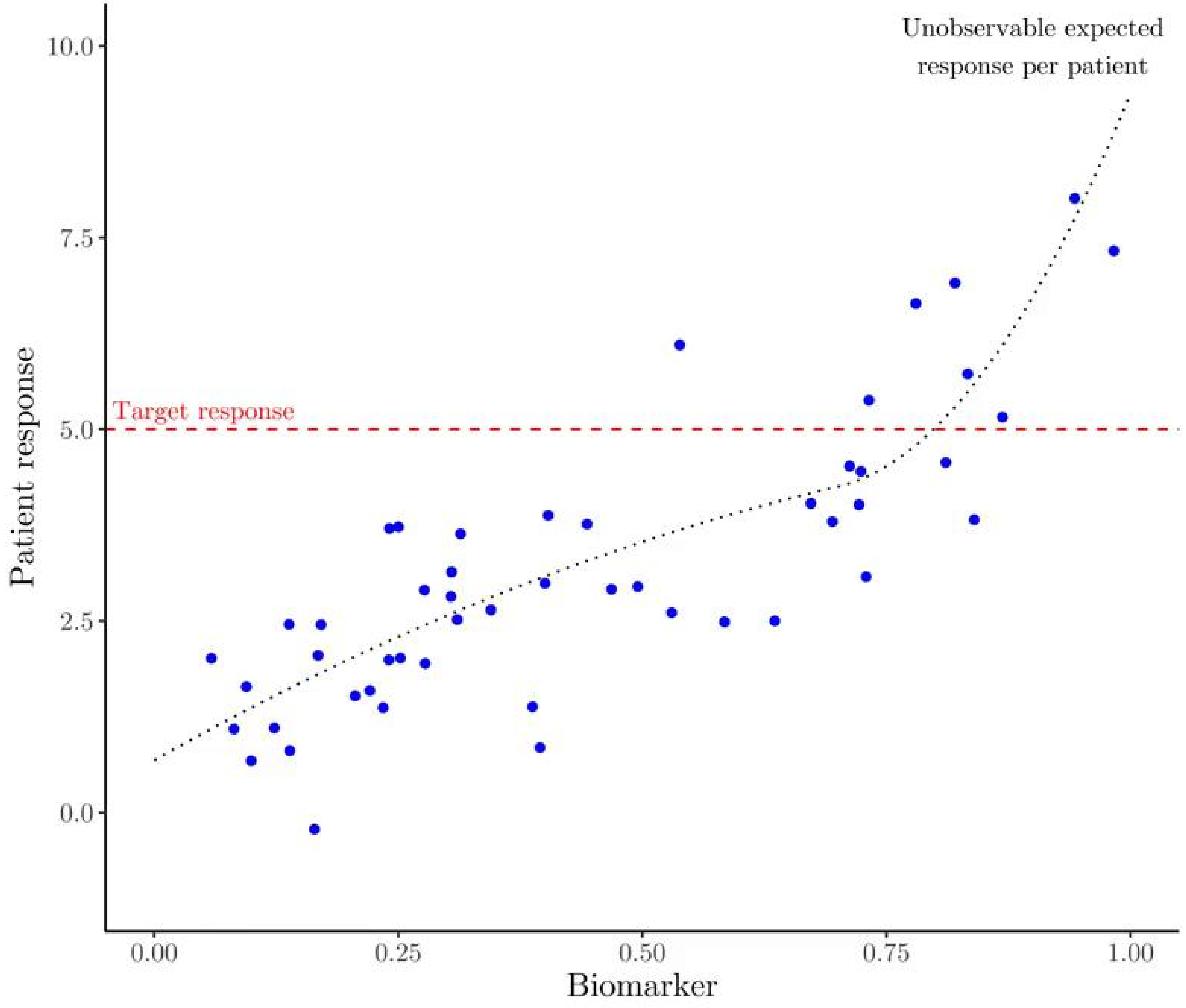


Subgroup selection

Data. Independent and identically distributed covariate-response pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ with values in $\mathbb{R}^d \times \mathbb{R}$ and unknown population regression function

$$\eta(x) := \mathbb{E}(Y_1 | X_1 = x).$$

User-specified: Threshold $\tau \in \mathbb{R}$ and maximal Type I error rate $\alpha \in (0, 1)$.

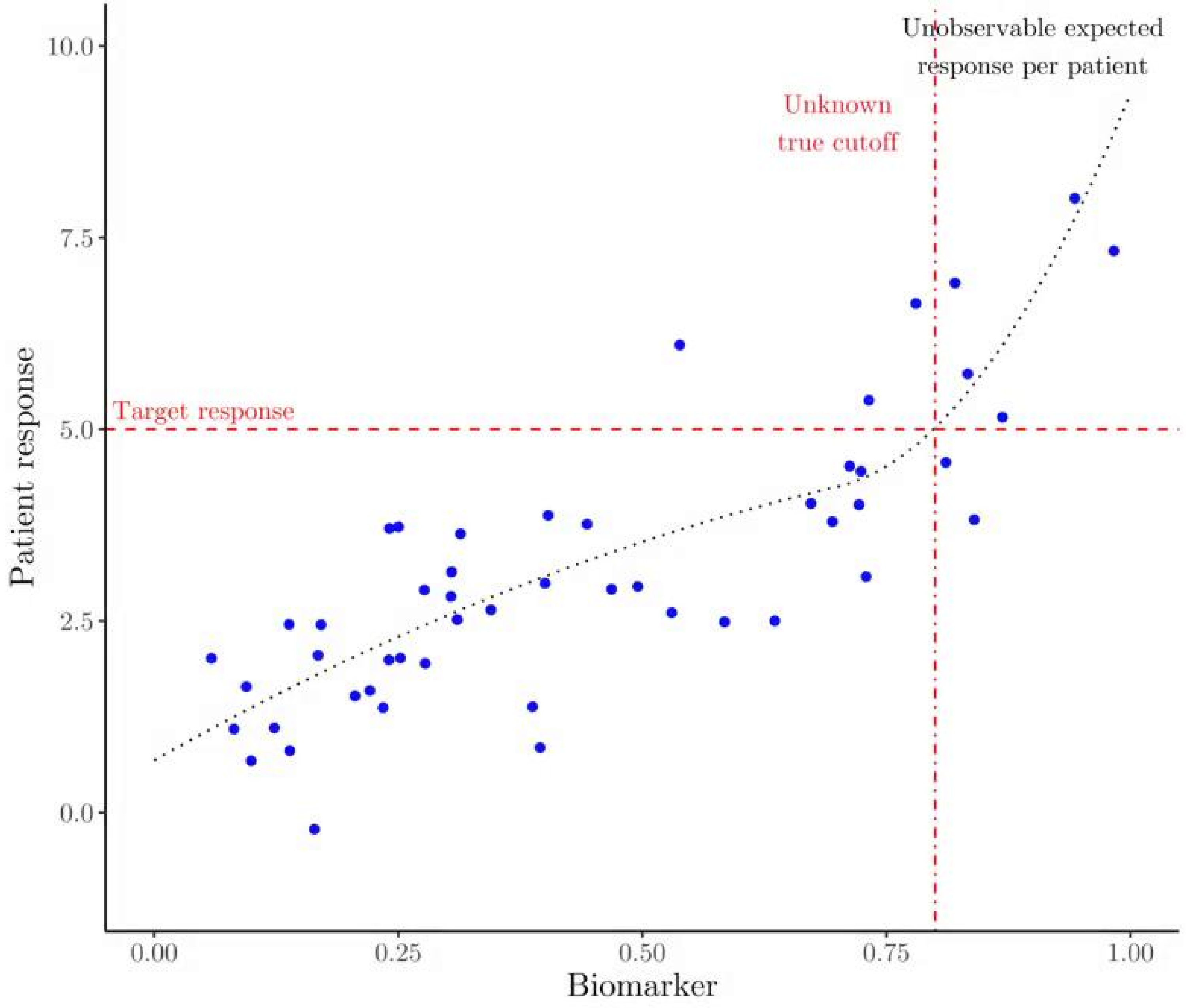


Subgroup selection

Data. Independent and identically distributed covariate-response pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ with values in $\mathbb{R}^d \times \mathbb{R}$ and unknown population regression function

$$\eta(x) := \mathbb{E}(Y_1 | X_1 = x).$$

User-specified: Threshold $\tau \in \mathbb{R}$ and maximal Type I error rate $\alpha \in (0, 1)$.

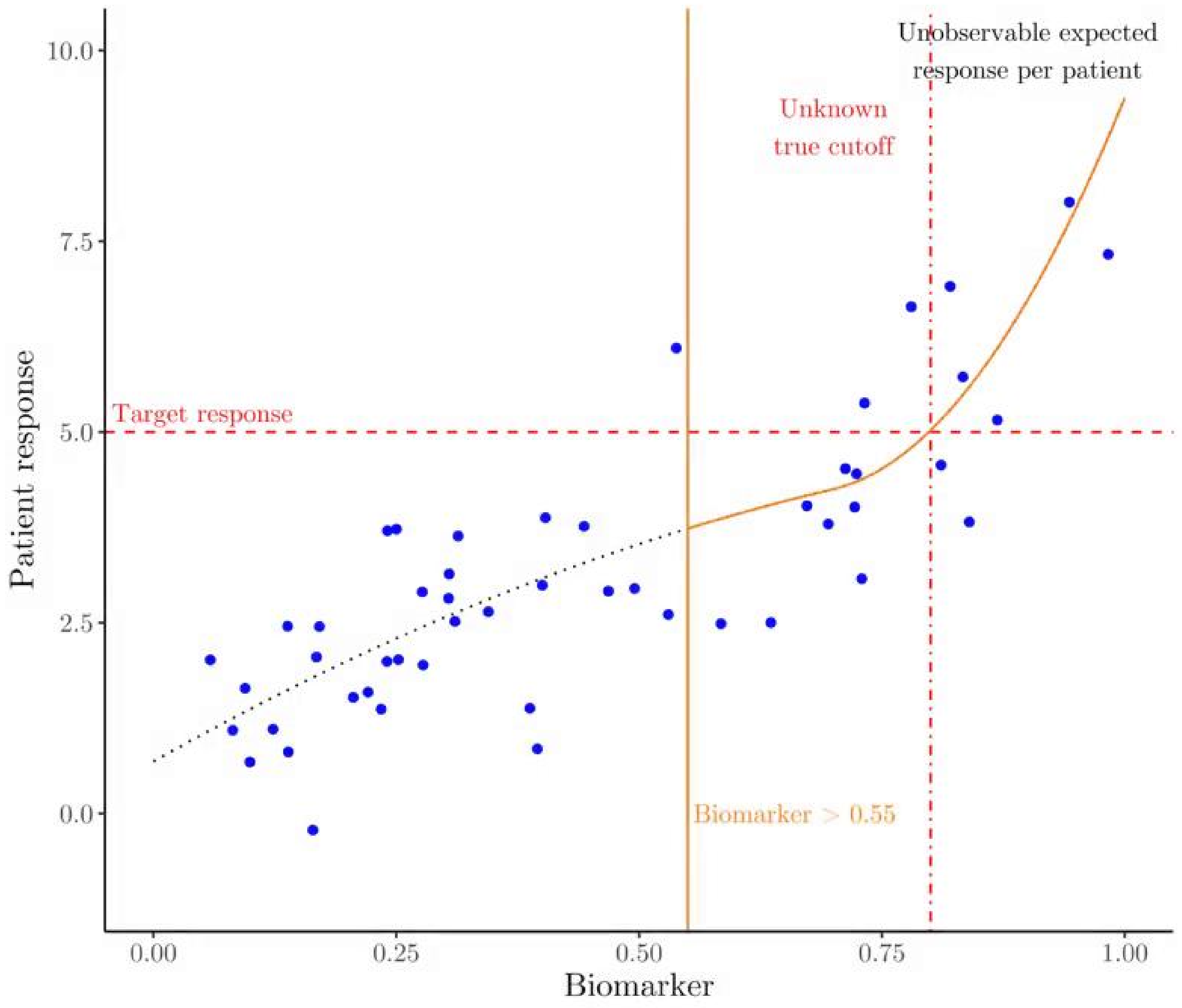


Subgroup selection

Data. Independent and identically distributed covariate-response pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ with values in $\mathbb{R}^d \times \mathbb{R}$ and unknown population regression function

$$\eta(x) := \mathbb{E}(Y_1 | X_1 = x).$$

User-specified: Threshold $\tau \in \mathbb{R}$ and maximal Type I error rate $\alpha \in (0, 1)$.

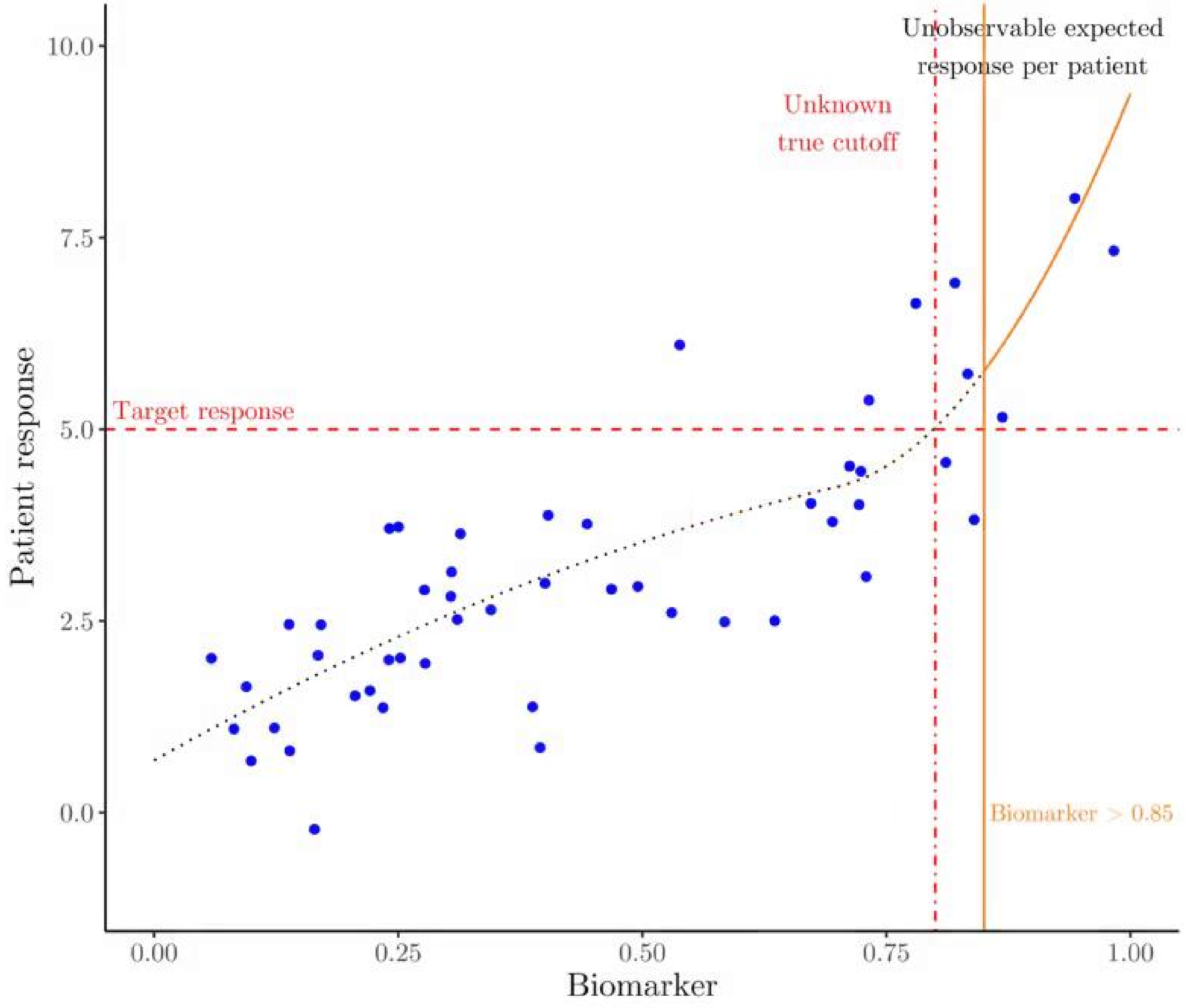


Subgroup selection

Data. Independent and identically distributed covariate-response pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ with values in $\mathbb{R}^d \times \mathbb{R}$ and unknown population regression function

$$\eta(x) := \mathbb{E}(Y_1 | X_1 = x).$$

User-specified: Threshold $\tau \in \mathbb{R}$ and maximal Type I error rate $\alpha \in (0, 1)$.



Subgroup selection

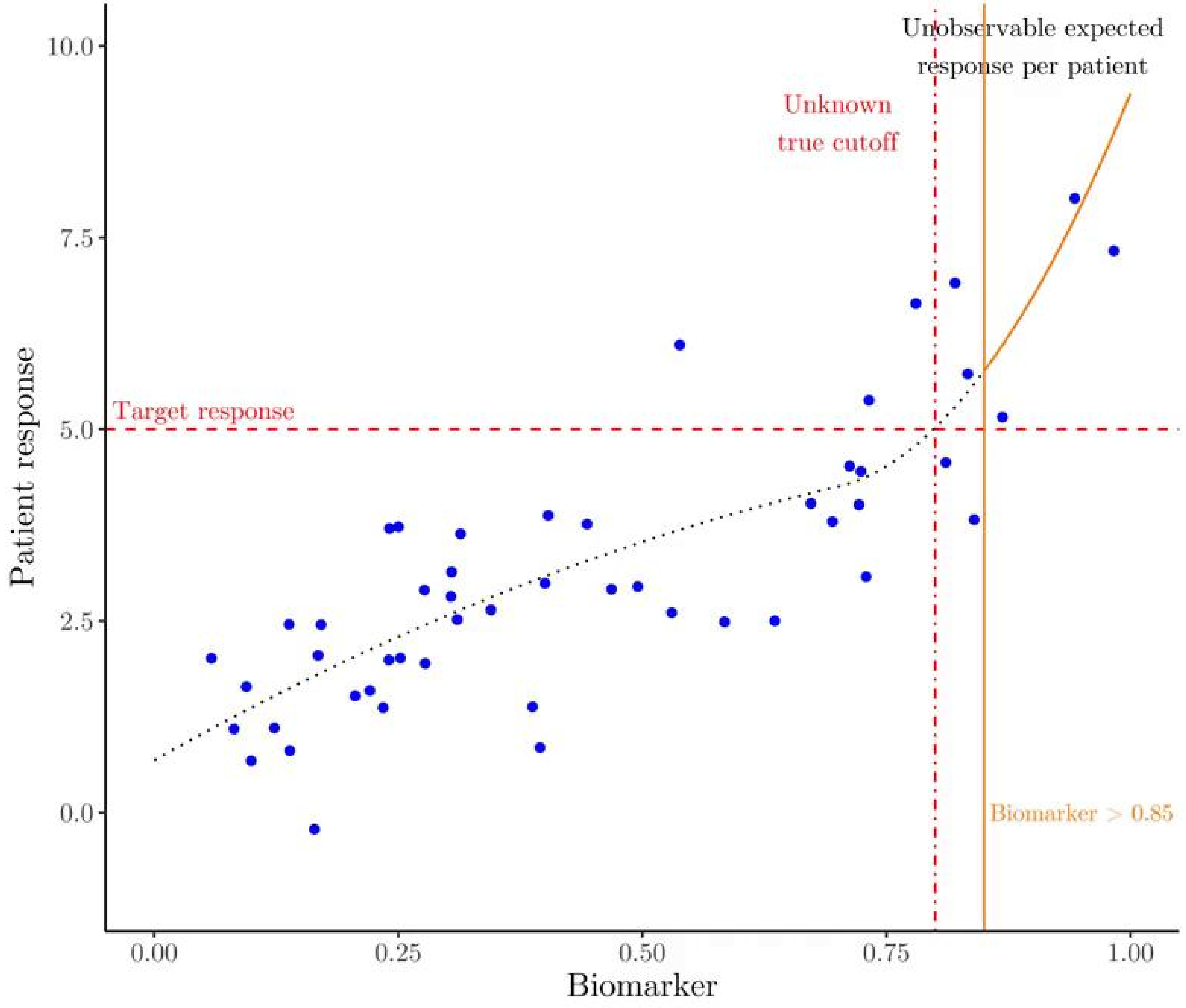
Data. Independent and identically distributed covariate-response pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ with values in $\mathbb{R}^d \times \mathbb{R}$ and unknown population regression function

$$\eta(x) := \mathbb{E}(Y_1 | X_1 = x).$$

User-specified: Threshold $\tau \in \mathbb{R}$ and maximal Type I error rate $\alpha \in (0, 1)$.

Task: Identify a subset $\hat{A}$ of $\mathbb{R}^d$, such that:

$$\mathbb{P}\big(\forall x \in \hat{A} : \eta(x) \geq \tau\big) \geq 1 - \alpha.$$



Subgroup selection

Data. Independent and identically distributed covariate-response pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$ with values in $\mathbb{R}^d \times \mathbb{R}$ and unknown population regression function

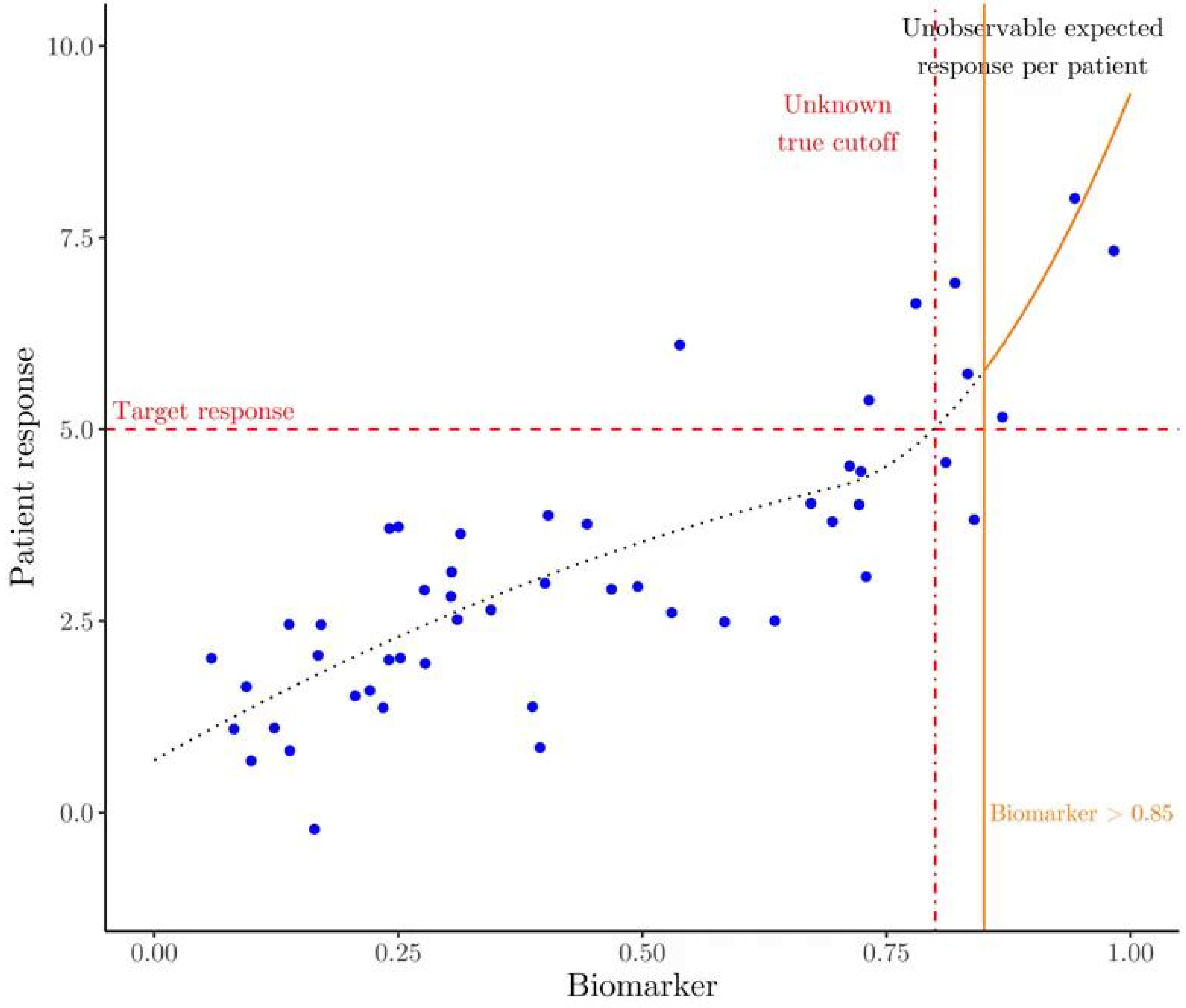
$$\eta(x) := \mathbb{E}(Y_1 | X_1 = x).$$

User-specified: Threshold $\tau \in \mathbb{R}$ and maximal Type I error rate $\alpha \in (0, 1)$.

Task: Identify a subset $\hat{A}$ of $\mathbb{R}^d$, such that:

$$\mathbb{P}\big(\forall x \in \hat{A} : \eta(x) \geq \tau\big) \geq 1 - \alpha.$$

This will be called Type I error rate control.



Subgroup selection

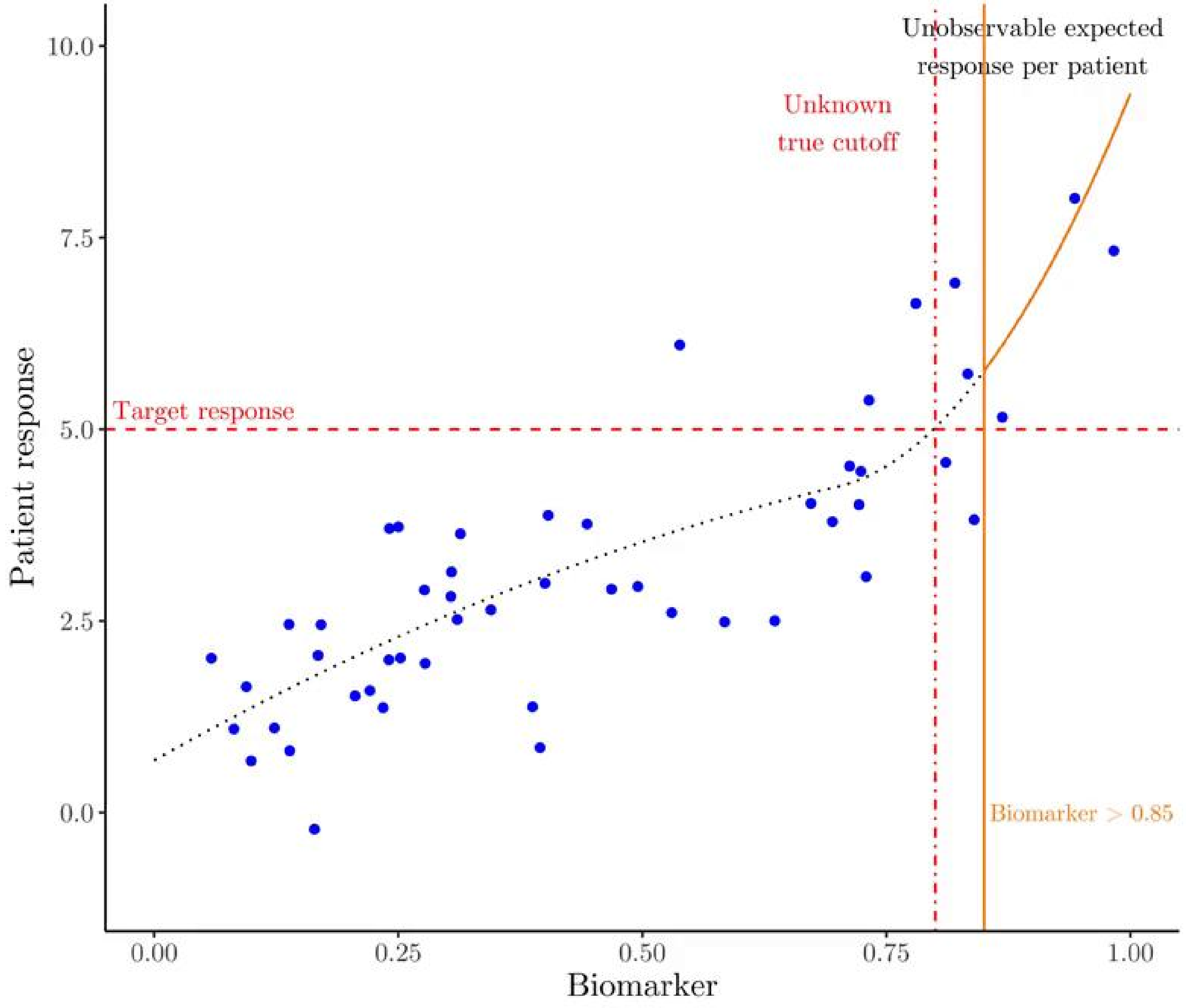
Subject to satisfying

$$\mathbb{P}\big(\forall x \in \hat{A} : \eta(x) \geq \tau\big) \geq 1 - \alpha,$$

we would like $\hat{A}$ to be as large as possible, i.e. for a new covariate point X we want

$$\mathbb{P}\big(X \in \hat{A} \mid \eta(X) \geq \tau\big)$$

to be large!



Subgroup selection

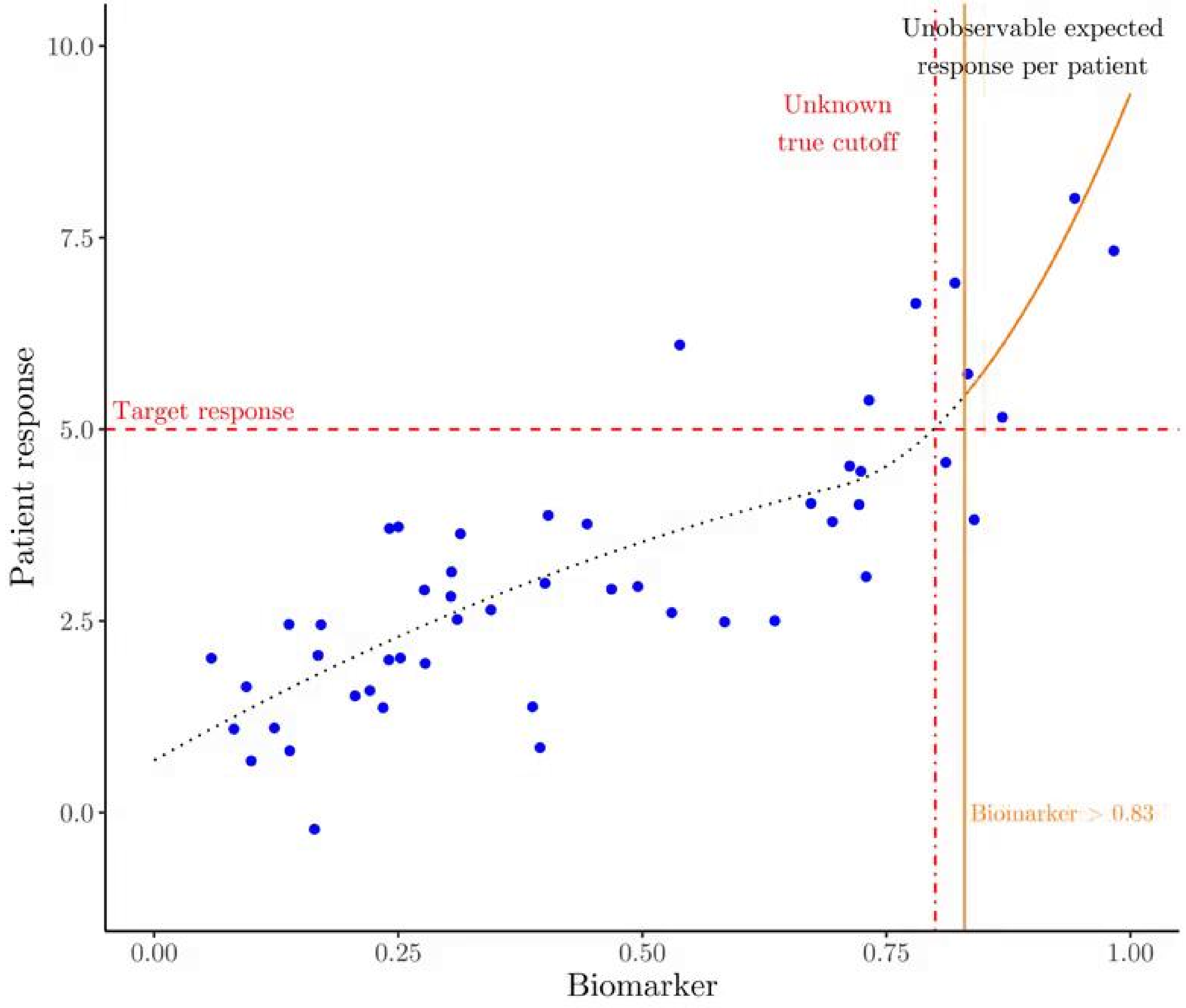
Subject to satisfying

$$\mathbb{P}\big(\forall x \in \hat{A} : \eta(x) \geq \tau\big) \geq 1 - \alpha,$$

we would like $\hat{A}$ to be as large as possible, i.e. for a new covariate point X we want

$$\mathbb{P}\big(X \in \hat{A} \mid \eta(X) \geq \tau\big)$$

to be large!



Subgroup selection

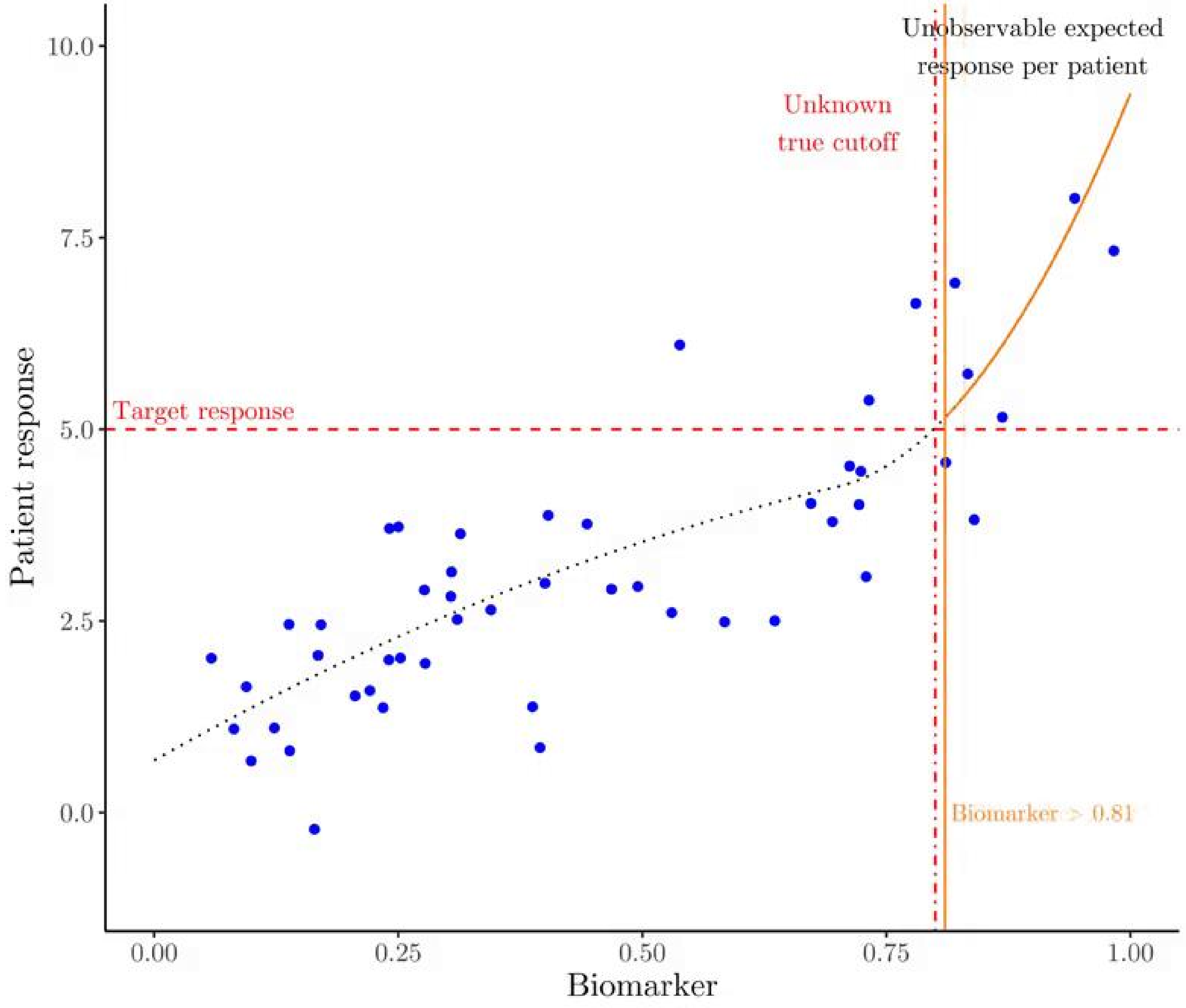
Subject to satisfying

$$\mathbb{P}\big(\forall x \in \hat{A} : \eta(x) \geq \tau\big) \geq 1 - \alpha,$$

we would like $\hat{A}$ to be as large as possible, i.e. for a new covariate point X we want

$$\mathbb{P}\big(X \in \hat{A} \mid \eta(X) \geq \tau\big)$$

to be large!



Subgroup selection

Subject to satisfying

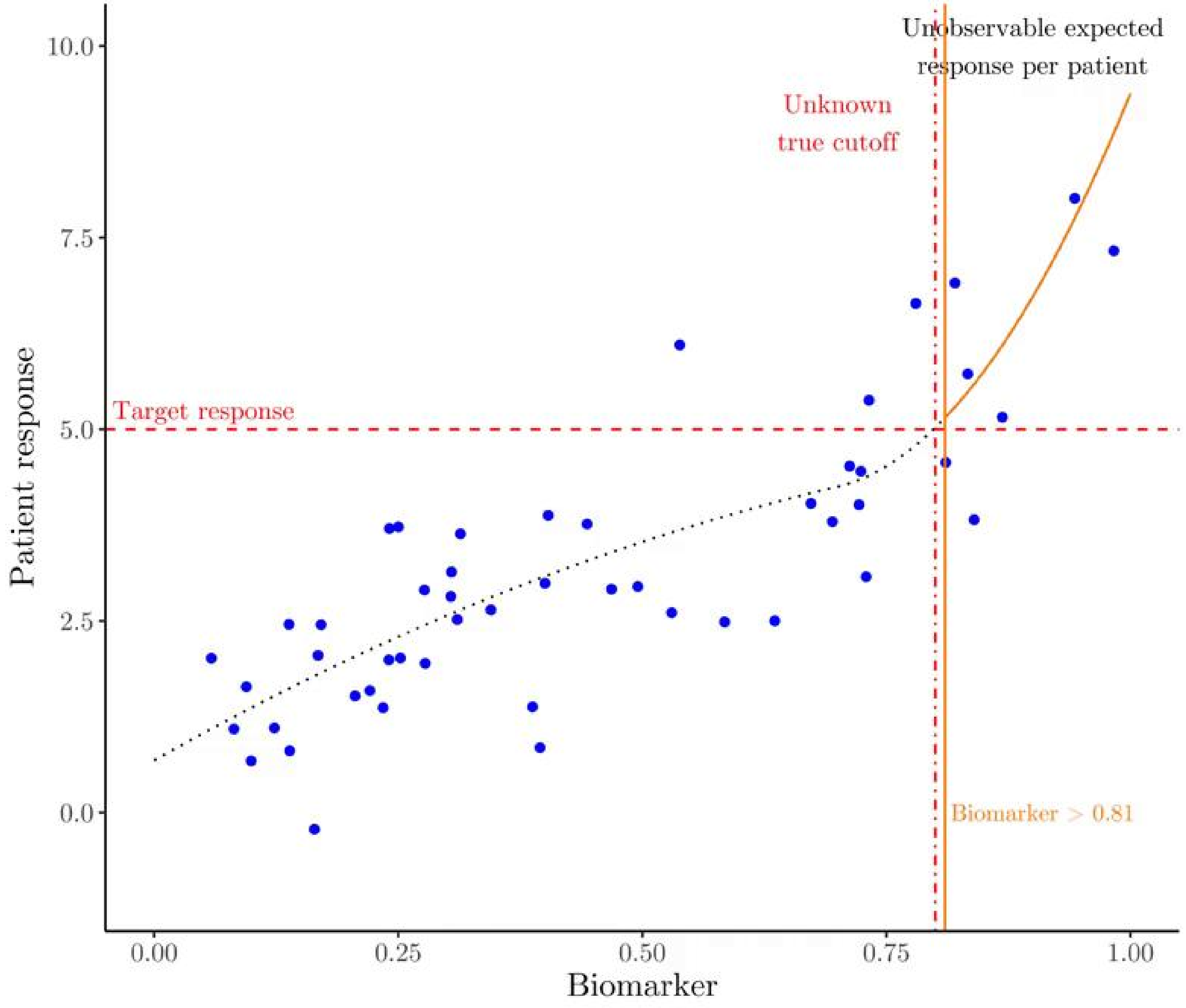
$$\mathbb{P}\big(\forall x \in \hat{A} : \eta(x) \geq \tau\big) \geq 1 - \alpha,$$

we would like $\hat{A}$ to be as large as possible, i.e. for a new covariate point X we want

$$\mathbb{P}\big(X \in \hat{A} \mid \eta(X) \geq \tau\big)$$

to be large!

We call this the procedure's power.



Subgroup selection

Subject to satisfying

$$\mathbb{P}\big(\forall x \in \hat{A} : \eta(x) \geq \tau\big) \geq 1 - \alpha,$$

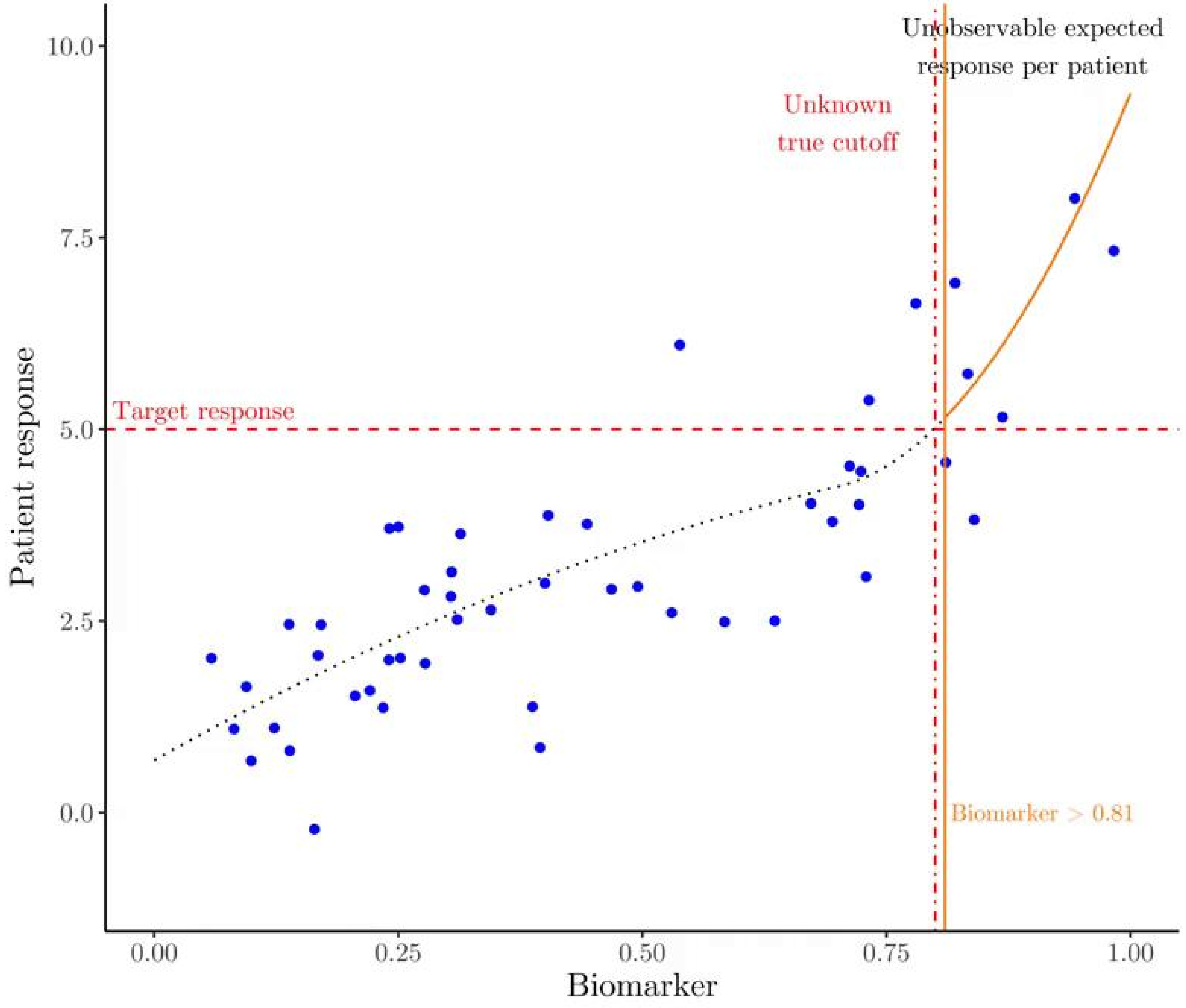
we would like $\hat{A}$ to be as large as possible, i.e. for a new covariate point X we want

$$\mathbb{P}\big(X \in \hat{A} \mid \eta(X) \geq \tau\big)$$

to be large!

We call this the procedure's power.

Theorem: *Any procedure controlling Type I error over all possible distributions will inevitably have trivial power.*



Subgroup selection

Subject to satisfying

$$\mathbb{P}\big(\forall x \in \hat{A} : \eta(x) \geq \tau\big) \geq 1 - \alpha,$$

we would like $\hat{A}$ to be as large as possible, i.e. for a new covariate point X we want

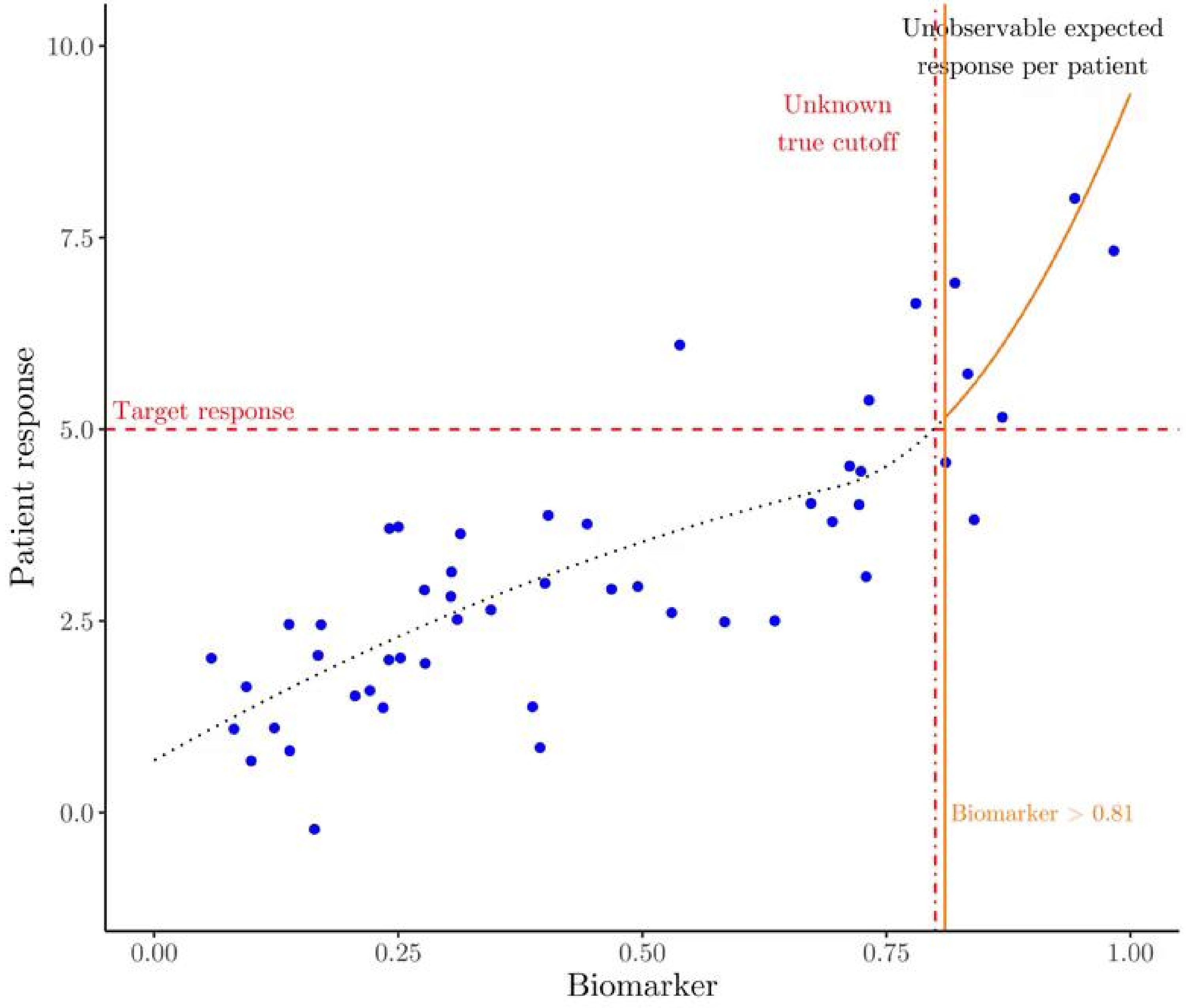
$$\mathbb{P}\big(X \in \hat{A} \mid \eta(X) \geq \tau\big)$$

to be large!

We call this the procedure's power.

Theorem: *Any procedure controlling Type I error over all possible distributions will inevitably have trivial power.*

$\Longrightarrow$ Assumptions are unavoidable.



Application: Risk group estimation

Background: In a Phase 2 study, about 250 patients received a new drug with varying dose. Some patients faced adverse events (AE). Can we predict which patients are at risk of AEs?

Application: Risk group estimation

Background: In a Phase 2 study, about 250 patients received a new drug with varying dose. Some patients faced adverse events (AE). Can we predict which patients are at risk of AEs?

Application of subgroup selection: we set $Y_i := \mathbb{1}\{\text{patient } i \text{ does not report AE}\}$, turning this into a classification setting.

Application: Risk group estimation

Background: In a Phase 2 study, about 250 patients received a new drug with varying dose. Some patients faced adverse events (AE). Can we predict which patients are at risk of AEs?

Application of subgroup selection: we set $Y_i := \mathbb{1}\{\text{patient } i \text{ does not report AE}\}$, turning this into a classification setting.

$\hat{A}$ then only contains covariate configurations with probability of **not** observing an AE exceeding τ.

Application: Risk group estimation

Background: In a Phase 2 study, about 250 patients received a new drug with varying dose. Some patients faced adverse events (AE). Can we predict which patients are at risk of AEs?

Application of subgroup selection: we set $Y_i := \mathbb{1}\{\text{patient } i \text{ does not report AE}\}$, turning this into a classification setting.

$\hat{A}$ then only contains covariate configurations with probability of **not** observing an AE exceeding τ.

E.g. $\tau = 0.95$ and $\alpha = 0.05$.

Application: Risk group estimation

Background: In a Phase 2 study, about 250 patients received a new drug with varying dose. Some patients faced adverse events (AE). Can we predict which patients are at risk of AEs?

Application of subgroup selection: we set $Y_i := \mathbb{1}\{\text{patient } i \text{ does not report AE}\}$, turning this into a classification setting.

$\hat{A}$ then only contains covariate configurations with probability of **not** observing an AE exceeding τ.

E.g. $\tau = 0.95$ and $\alpha = 0.05$.

Decision process once we have computed $\hat{A}$ and observe a new patient with covariate values X:

Application: Risk group estimation

Background: In a Phase 2 study, about 250 patients received a new drug with varying dose. Some patients faced adverse events (AE). Can we predict which patients are at risk of AEs?

Application of subgroup selection: we set $Y_i := \mathbb{1}\{\text{patient } i \text{ does not report AE}\}$, turning this into a classification setting.

$\hat{A}$ then only contains covariate configurations with probability of **<u>not</u>** observing an AE exceeding τ.

E.g. $\tau = 0.95$ and $\alpha = 0.05$.

Decision process once we have computed $\hat{A}$ and observe a new patient with covariate values X:

- If $X \in \hat{A}$: patient can be expected to not face AEs since $\eta(X) \geq \tau$ (with probability $1 - \alpha$).

Application: Risk group estimation

Background: In a Phase 2 study, about 250 patients received a new drug with varying dose. Some patients faced adverse events (AE). Can we predict which patients are at risk of AEs?

Application of subgroup selection: we set $Y_i := \mathbb{1}\{\text{patient } i \text{ does not report AE}\}$, turning this into a classification setting.

$\hat{A}$ then only contains covariate configurations with probability of **not** observing an AE exceeding τ.

E.g. $\tau = 0.95$ and $\alpha = 0.05$.

Decision process once we have computed $\hat{A}$ and observe a new patient with covariate values X:

- If $X \in \hat{A}$: patient can be expected to not face AEs since $\eta(X) \geq \tau$ (with probability $1 - \alpha$).
- If $X \notin \hat{A}$: patient might need further attention.

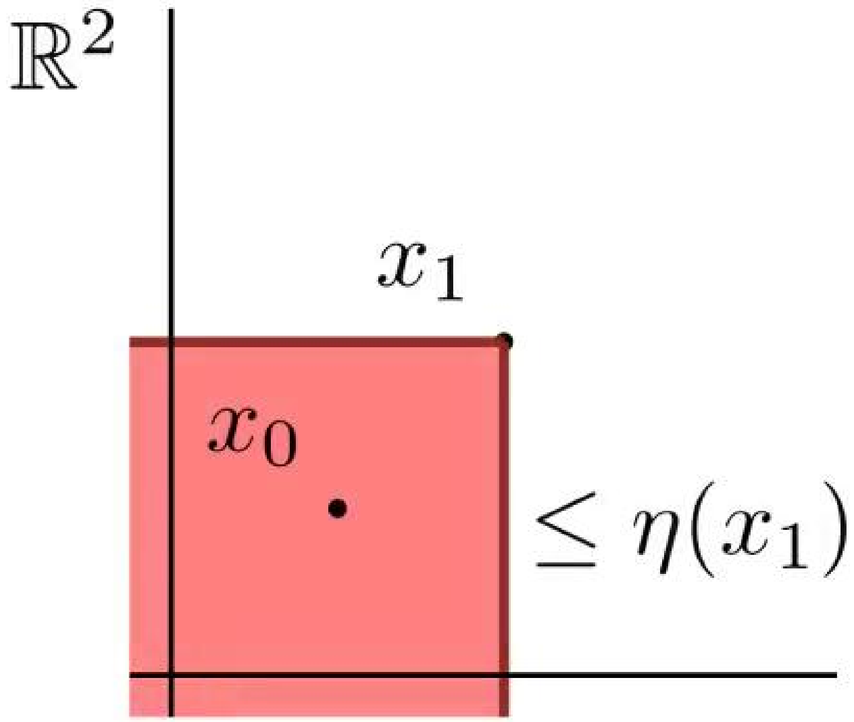
Statistical setting

Assumptions:

Statistical setting

Assumptions:

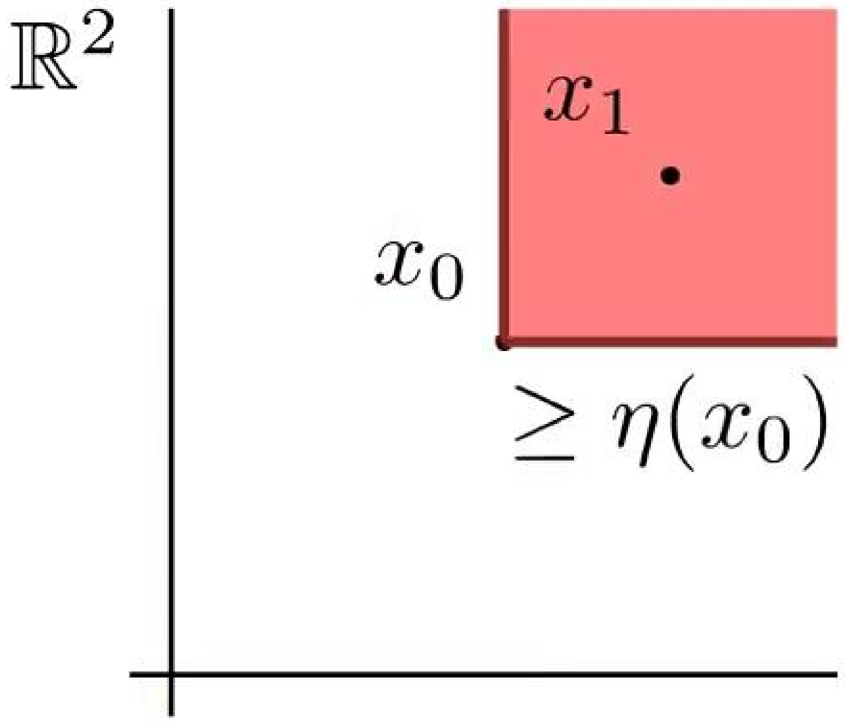
(i) the regression function $\eta(x) := \mathbb{E}(Y|X = x)$ is increasing on $\mathbb{R}^d$, i.e. $x_0 \preccurlyeq x_1 \Longrightarrow \eta(x_0) \leq \eta(x_1)$,



Statistical setting

Assumptions:

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

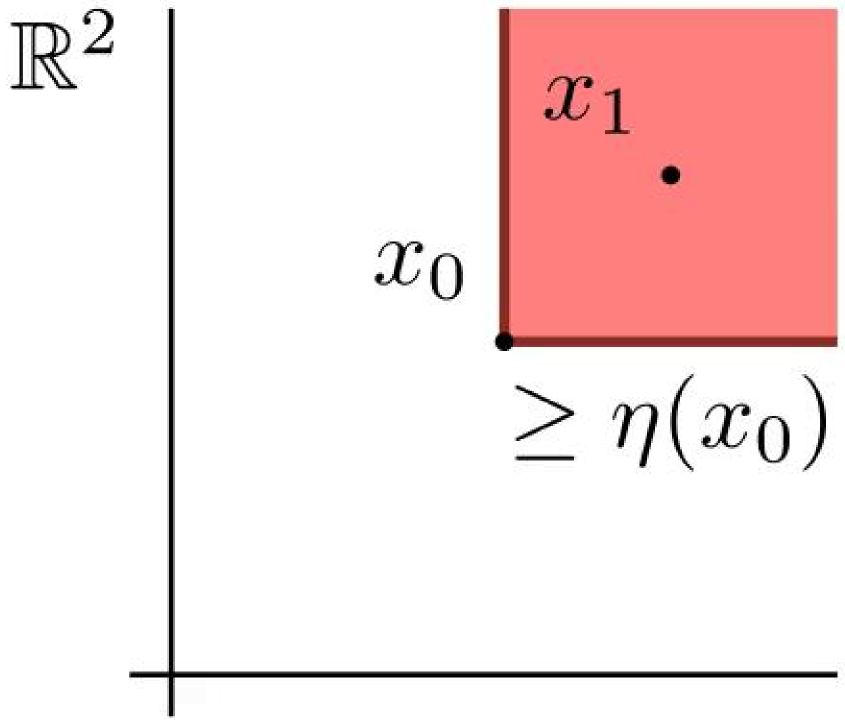


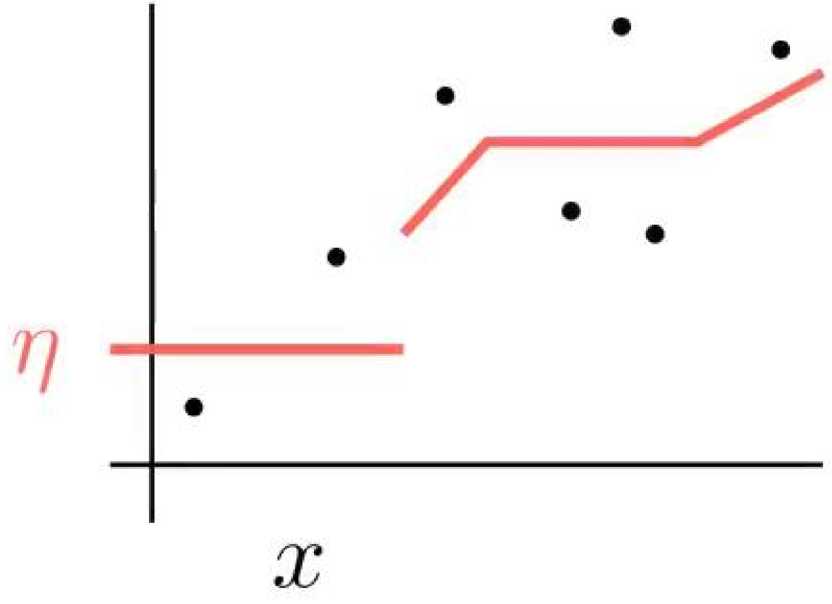
Statistical setting

Assumptions:

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

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



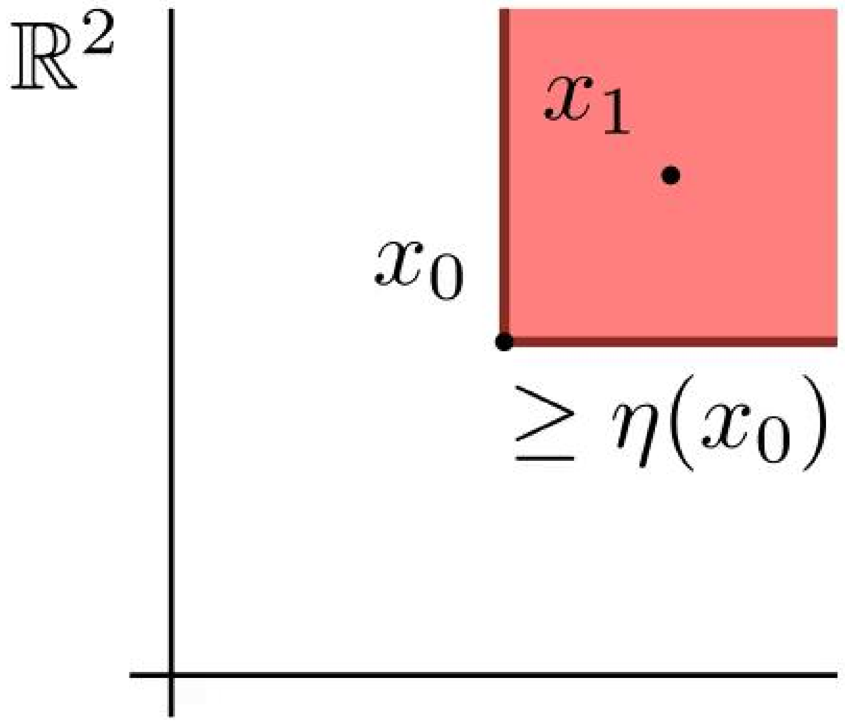


Statistical setting

Assumptions:

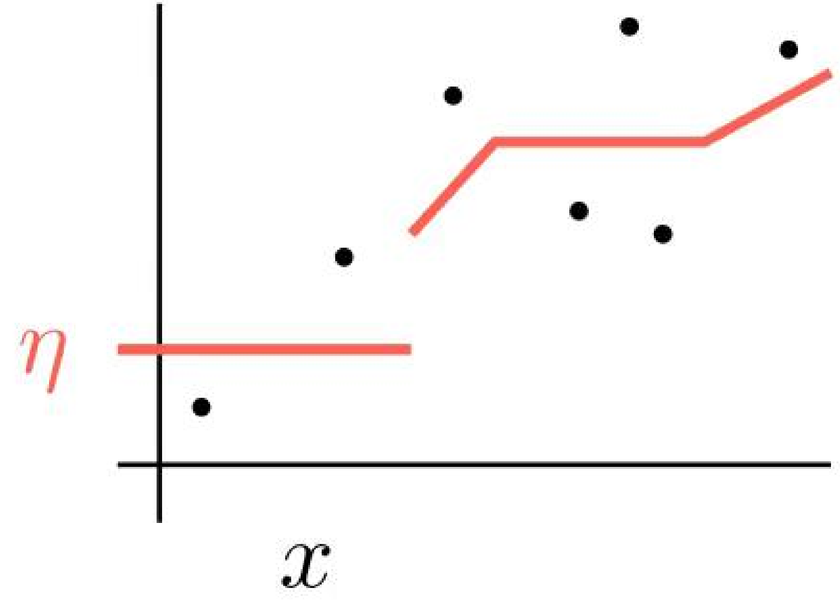
(i) the regression function $\eta(x) := \mathbb{E}(Y|X=x)$ is increasing on $\mathbb{R}^d$, i.e. $x_0 \preccurlyeq x_1 \Longrightarrow \eta(x_0) \leq \eta(x_1)$,

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



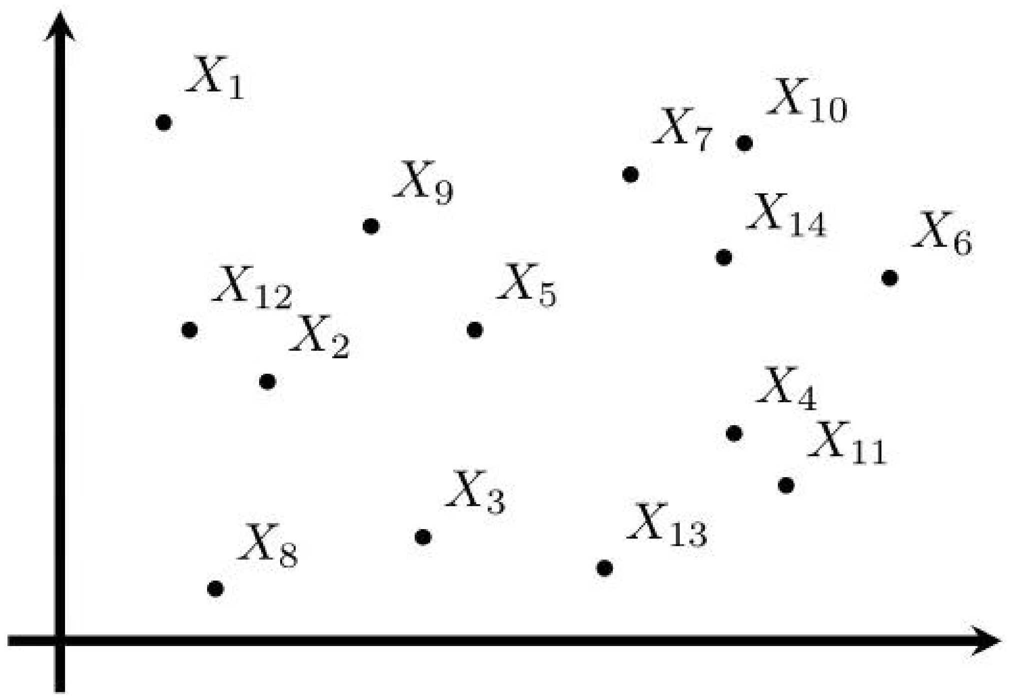
Alternative settings:

Hölder-smoothness of η (Reeve et al., 2023), Generalized Linear Models (GLMs) (Wan et al., 2024).



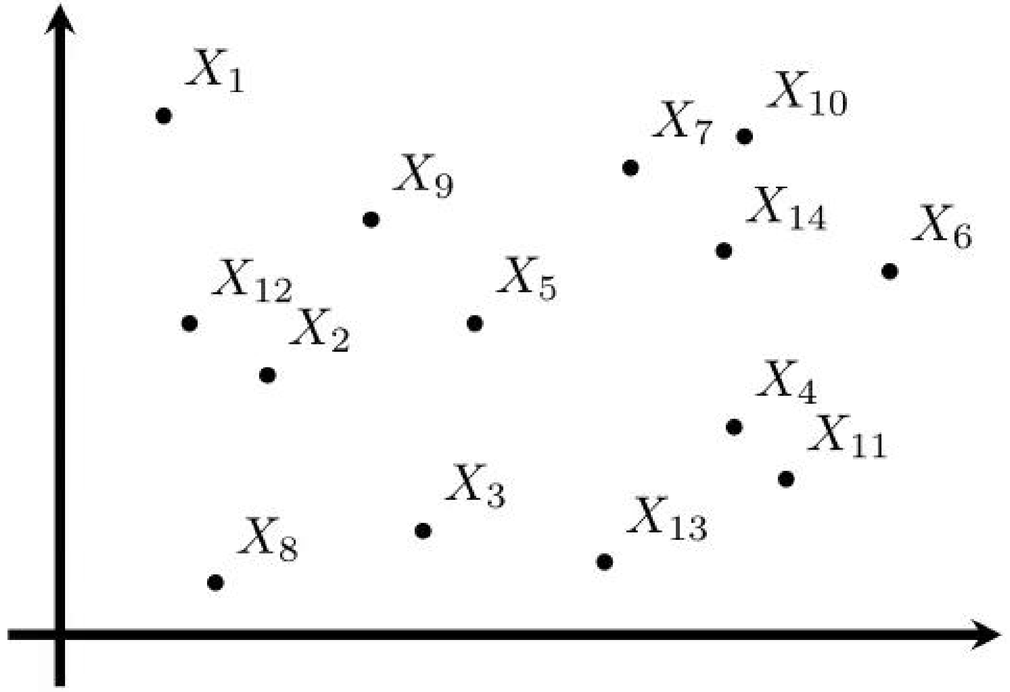
High-level strategy

For $x_0 \in \mathbb{R}^d$, define null hypothesis $H_0(x_0) : \eta(x_0) < \tau$.



High-level strategy

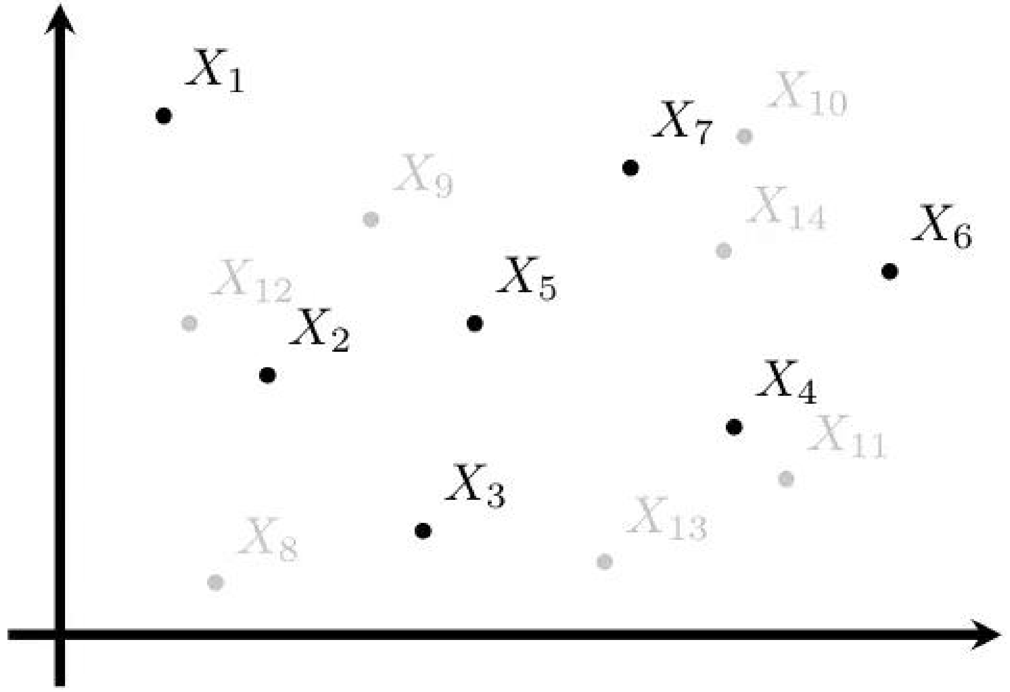
For $x_0 \in \mathbb{R}^d$, define null hypothesis $H_0(x_0) : \eta(x_0) < \tau$.



High-level strategy:

High-level strategy

For $x_0 \in \mathbb{R}^d$, define null hypothesis $H_0(x_0) : \eta(x_0) < \tau$.

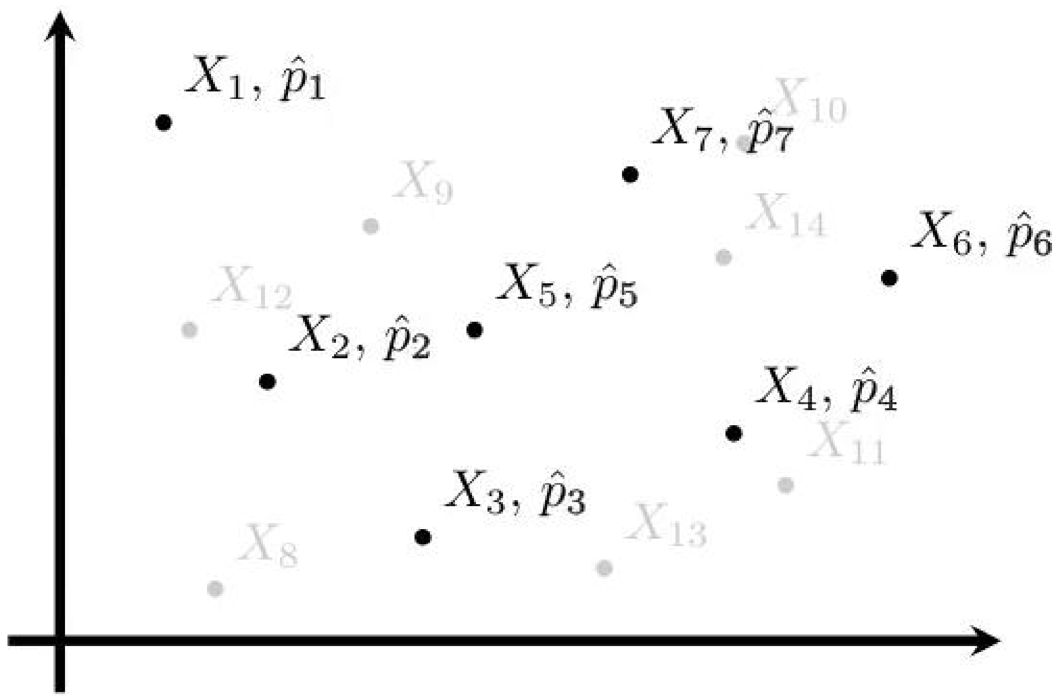


High-level strategy:

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 $H_0(x_0) : \eta(x_0) < \tau$.

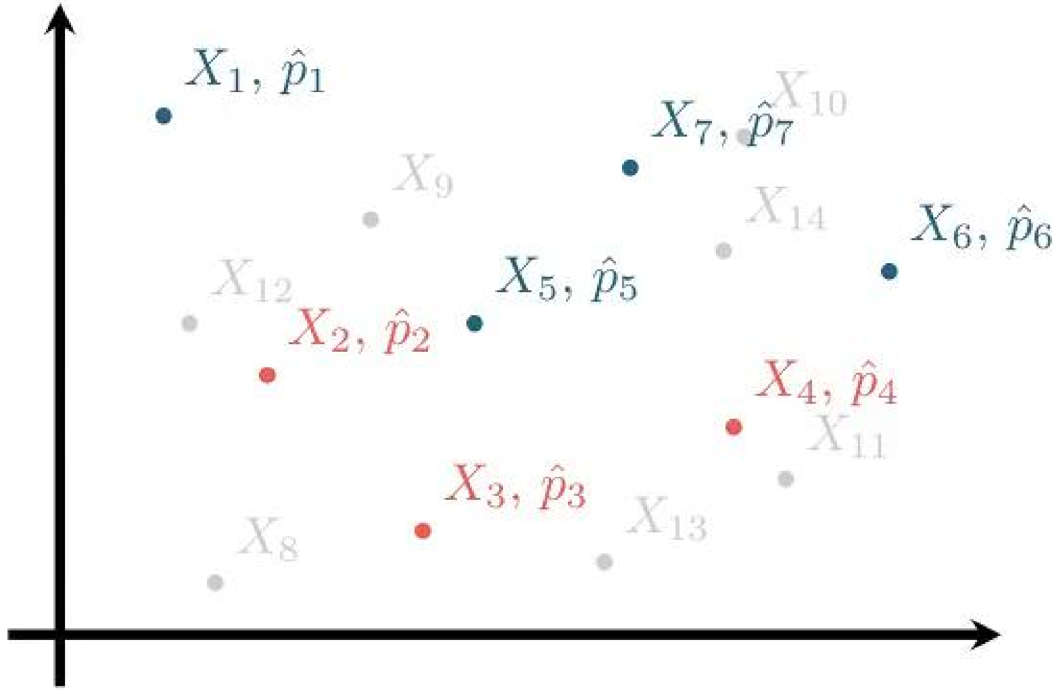


High-level strategy:

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

High-level strategy

For $x_0 \in \mathbb{R}^d$, define null hypothesis $H_0(x_0) : \eta(x_0) < \tau$.

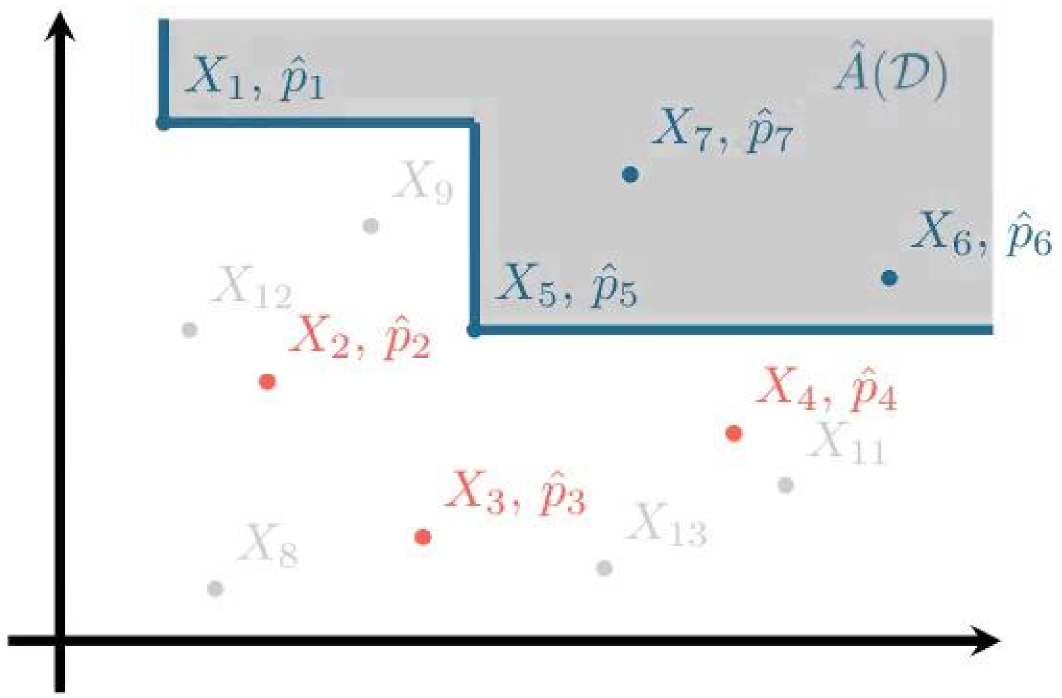


High-level strategy:

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

High-level strategy

For $x_0 \in \mathbb{R}^d$, define null hypothesis $H_0(x_0) : \eta(x_0) < \tau$.

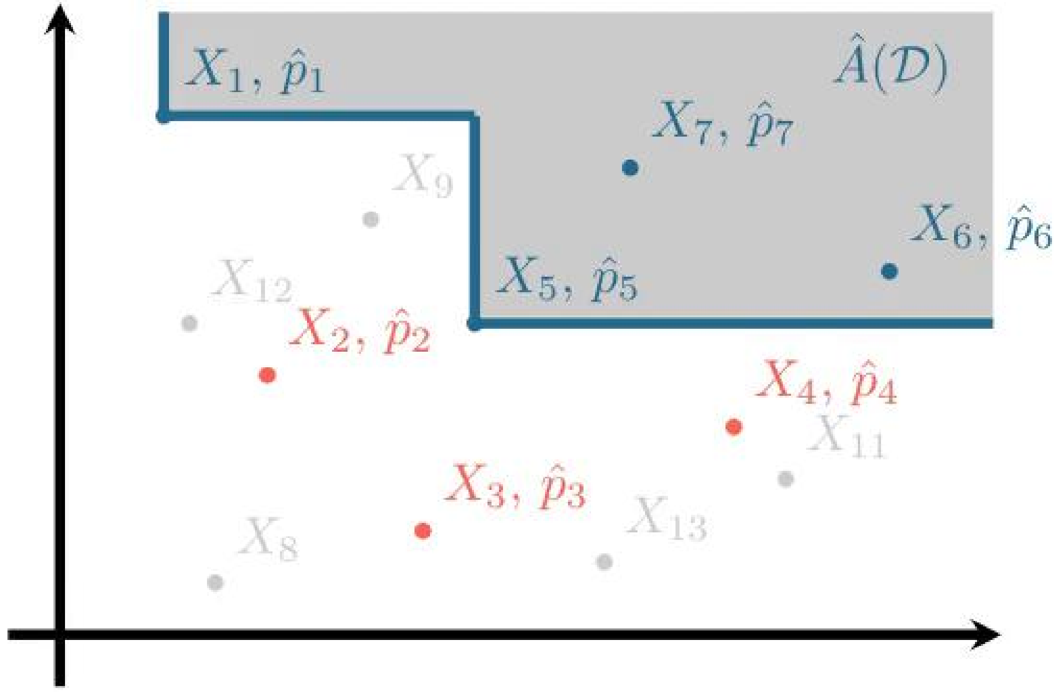


High-level strategy:

1. Subsample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Calculate *p-values* $\hat{p}_i$ for $H_0(X_i)$, $i \in \{1, \ldots, m\}$;
3. Apply a *multiple testing procedure* with FWER-control to reject $\mathcal{R}_\alpha \subseteq \{1, \ldots, m\}$;
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 $H_0(x_0) : \eta(x_0) < \tau$.



High-level strategy:

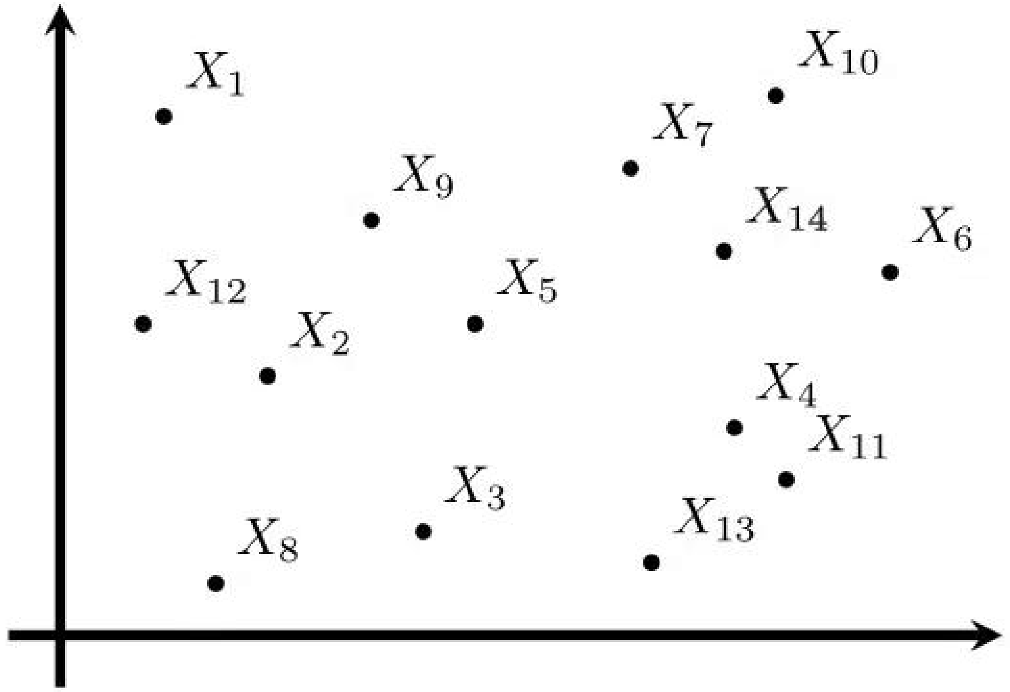
1. Subsample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Calculate *p-values* $\hat{p}_i$ for $H_0(X_i)$, $i \in \{1, \ldots, m\}$;
3. Apply a *multiple testing procedure* with FWER-control to reject $\mathcal{R}_\alpha \subseteq \{1, \ldots, m\}$;
4. Output $\hat{A} := \{x \in \mathbb{R}^d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

Construct p-values $\hat{p}_i$ for $H_0(X_i)$, $i \in \{1, \ldots, m\}$ (sub-Gaussian case)

Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) : \eta(x_0) < \tau$.

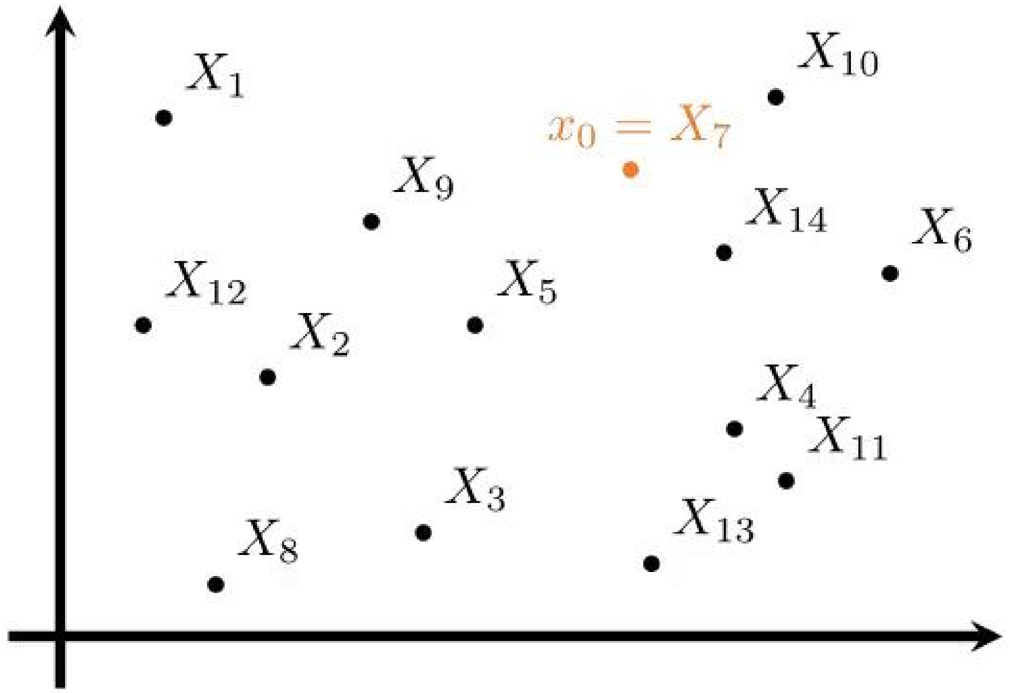
Construct p-values $\hat{p}_i$ for $H_0(X_i)$, $i \in \{1, \ldots, m\}$ (sub-Gaussian case)

Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) : \eta(x_0) < \tau$.



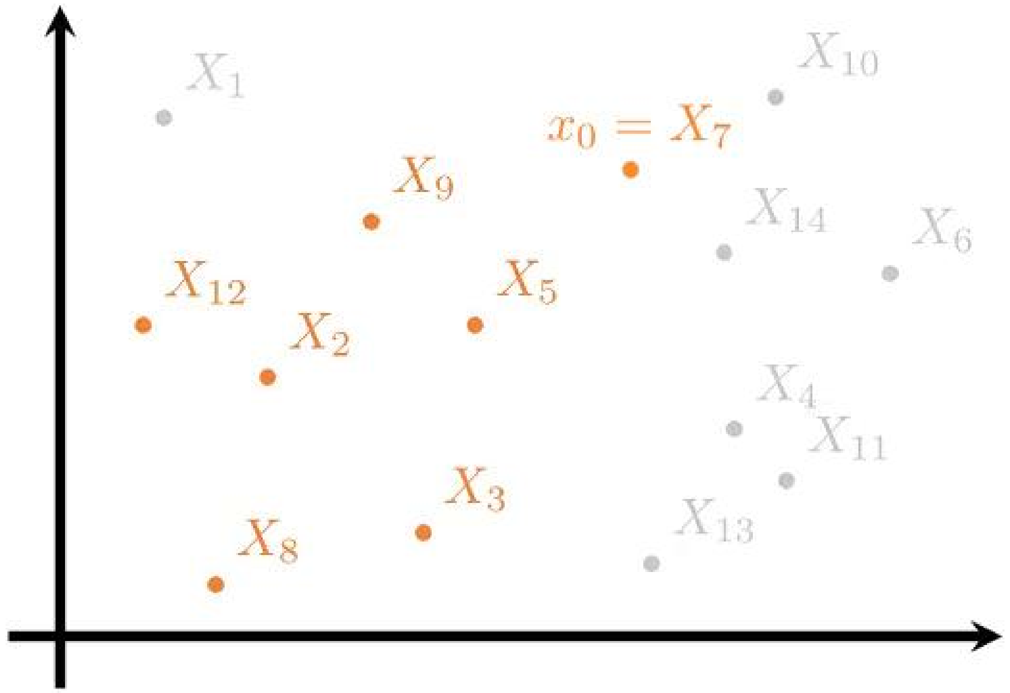
Construct p-values $\hat{p}_i$ for $H_0(X_i)$, $i \in \{1, \ldots, m\}$ (sub-Gaussian case)

Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) : \eta(x_0) < \tau$.



Construct p-values $\hat{p}_i$ for $H_0(X_i)$, $i \in \{1, \ldots, m\}$ (sub-Gaussian case)

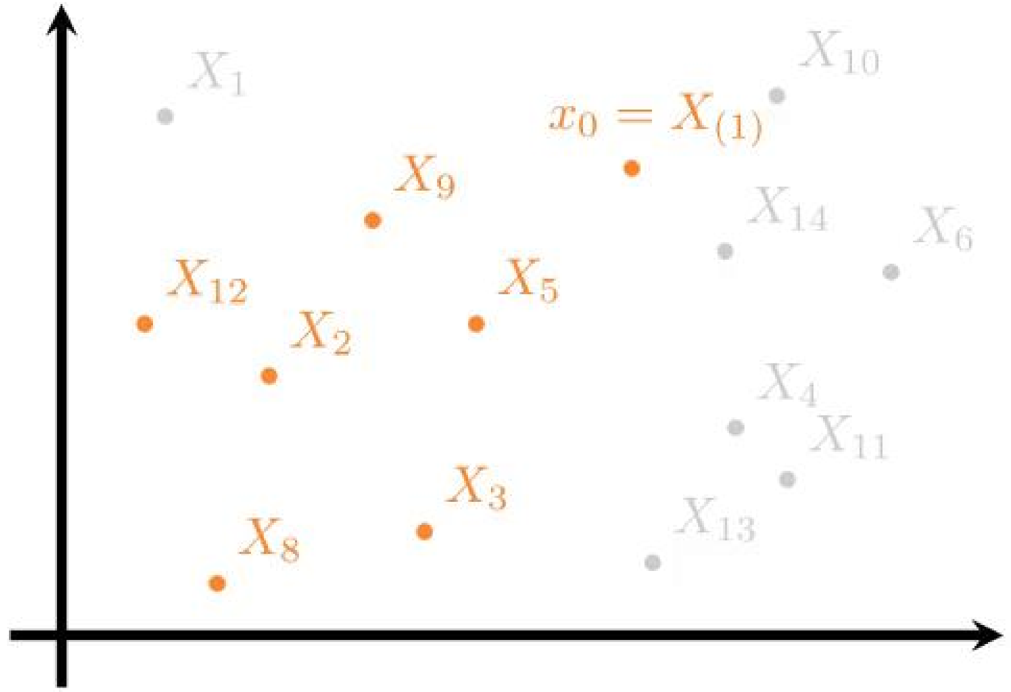
Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) : \eta(x_0) < \tau$.

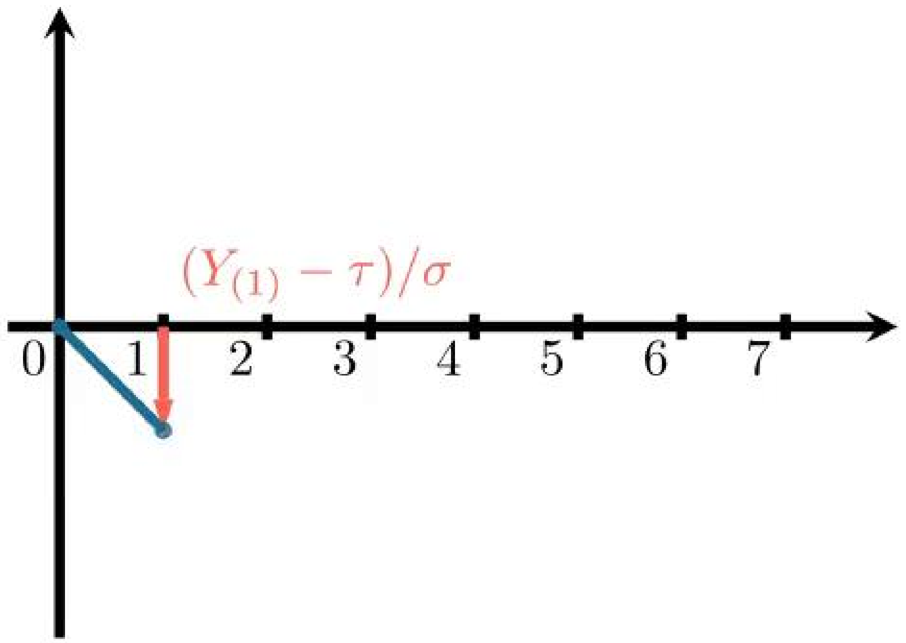


Denote $\mathcal{I}(x_0) := \big\{i \in \{1, \ldots, n\} : X_i \preccurlyeq x_0\big\}$, $n(x_0) := |\mathcal{I}(x_0)|$.

Construct p-values $\hat{p}_i$ for $H_0(X_i)$, $i \in \{1, \ldots, m\}$ (sub-Gaussian case)

Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) : \eta(x_0) < \tau$.



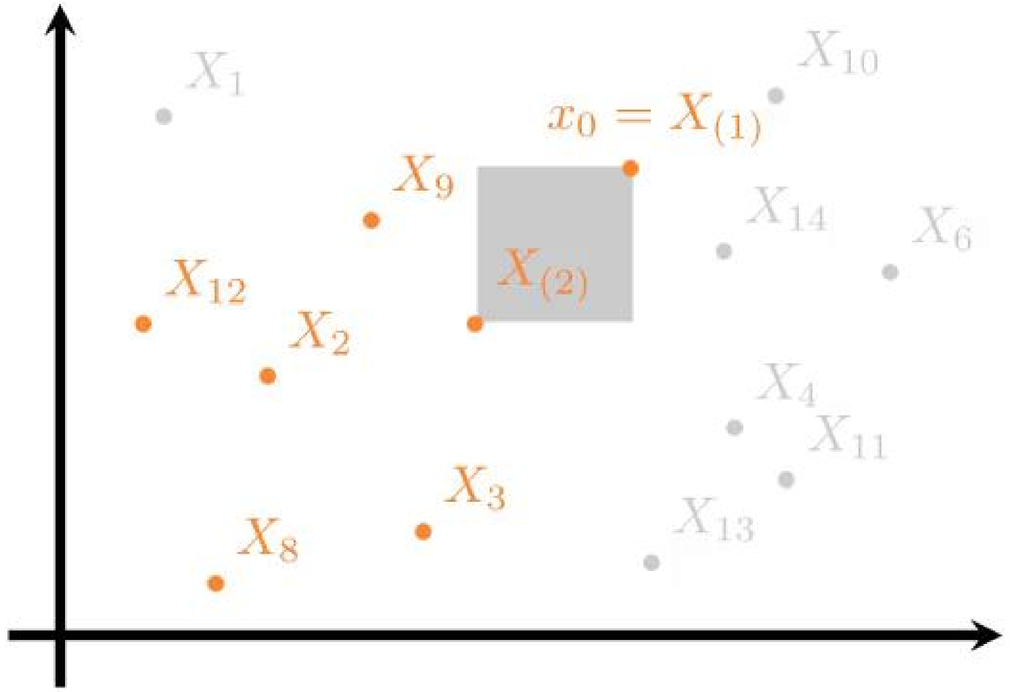


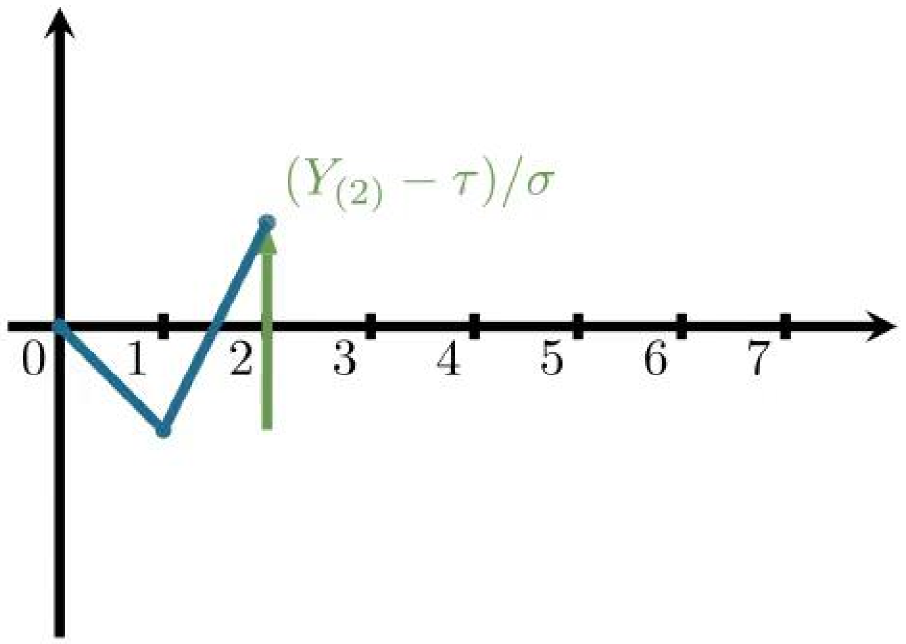
Denote $\mathcal{I}(x_0) := \big\{i \in \{1, \ldots, n\} : X_i \preccurlyeq x_0\big\}$, $n(x_0) := |\mathcal{I}(x_0)|$.

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 $H_0(X_i)$, $i \in \{1,\ldots,m\}$ (sub-Gaussian case)

Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) : \eta(x_0) < \tau$.



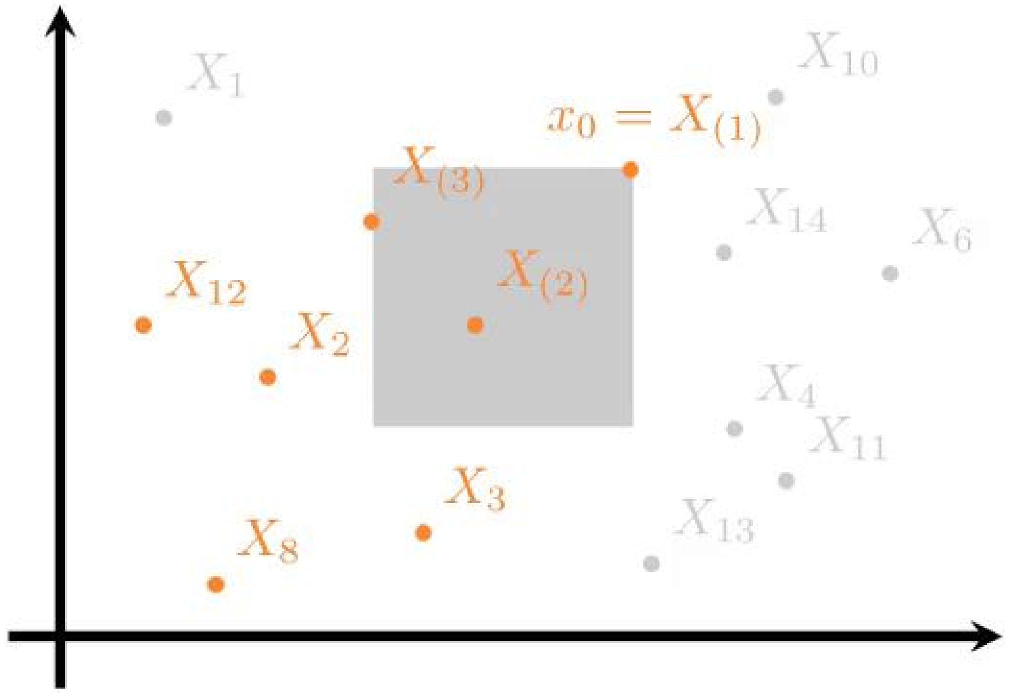


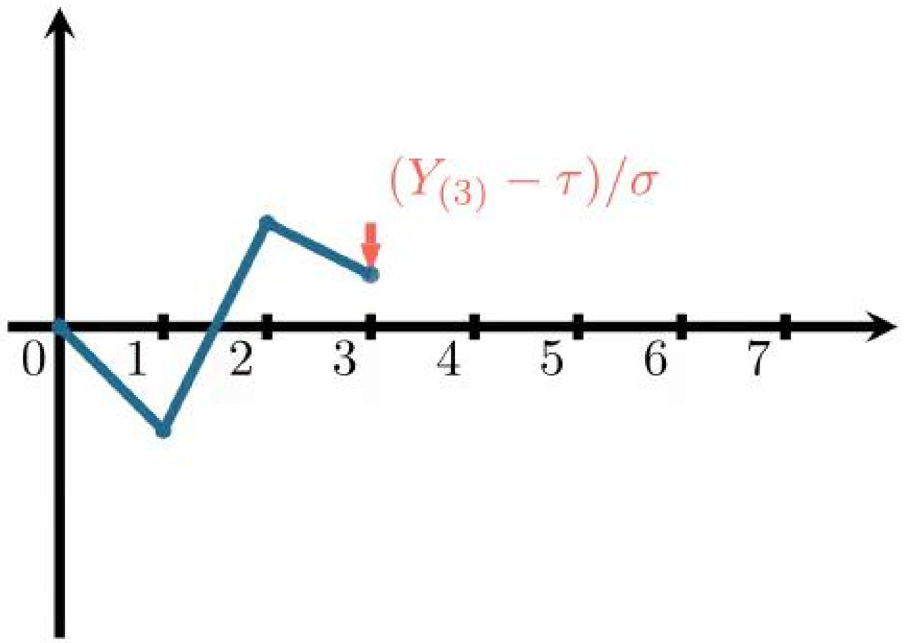
Denote $\mathcal{I}(x_0) := \big\{ i \in \{1,\ldots,n\} : X_i \preccurlyeq x_0 \big\}$, $n(x_0) := |\mathcal{I}(x_0)|$.

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 $H_0(X_i)$, $i \in \{1, \ldots, m\}$ (sub-Gaussian case)

Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) : \eta(x_0) < \tau$.



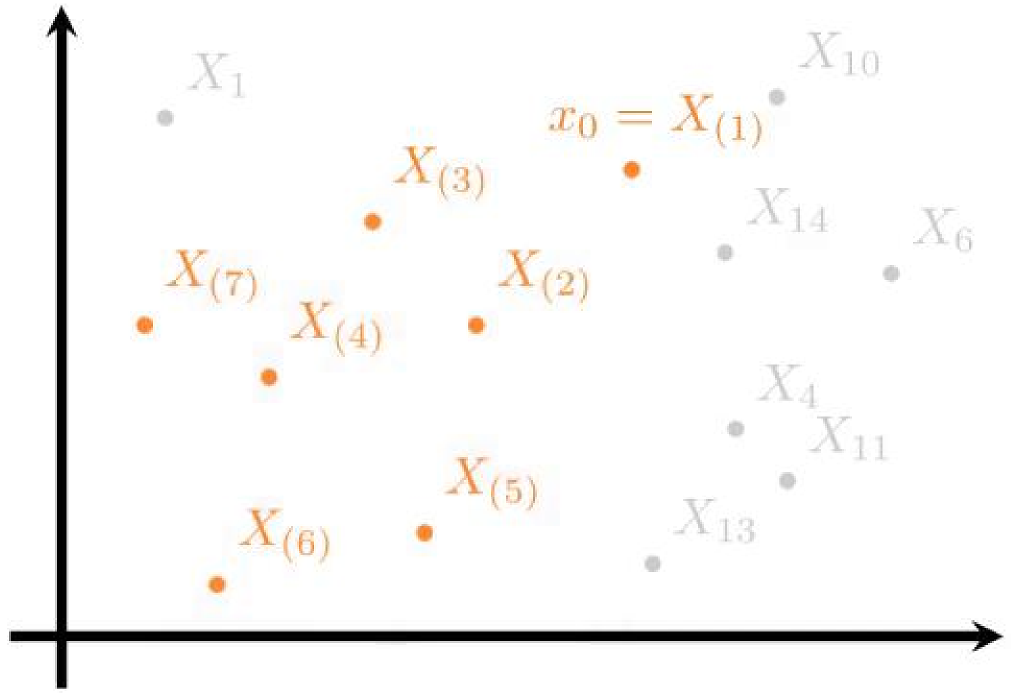


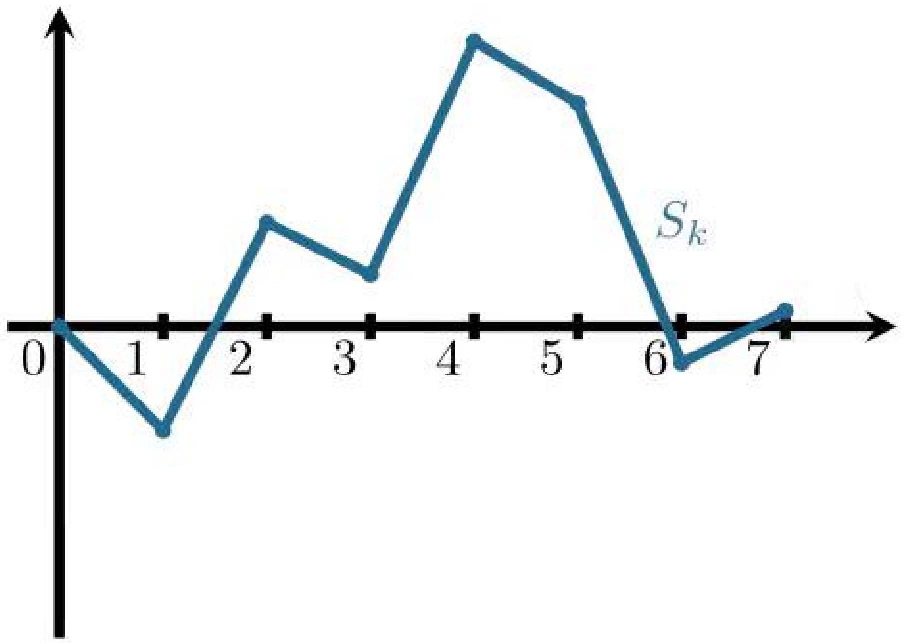
Denote $\mathcal{I}(x_0) := \{i \in \{1, \ldots, n\} : X_i \preccurlyeq x_0\}$, $n(x_0) := |\mathcal{I}(x_0)|$.

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 $H_0(X_i)$, $i \in \{1, \ldots, m\}$ (sub-Gaussian case)

Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) : \eta(x_0) < \tau$.



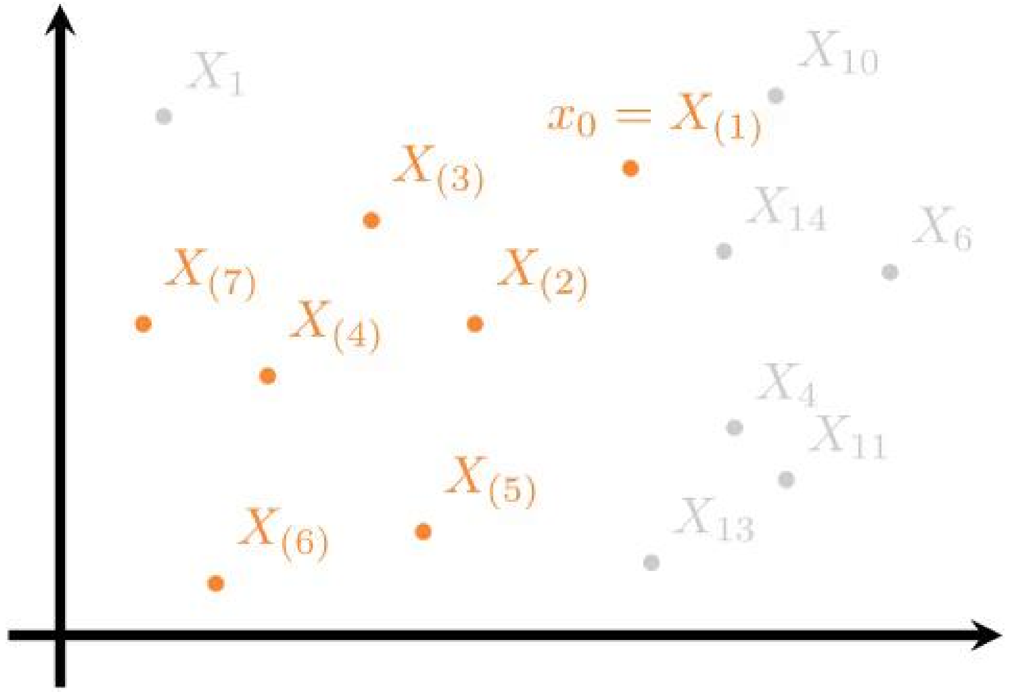


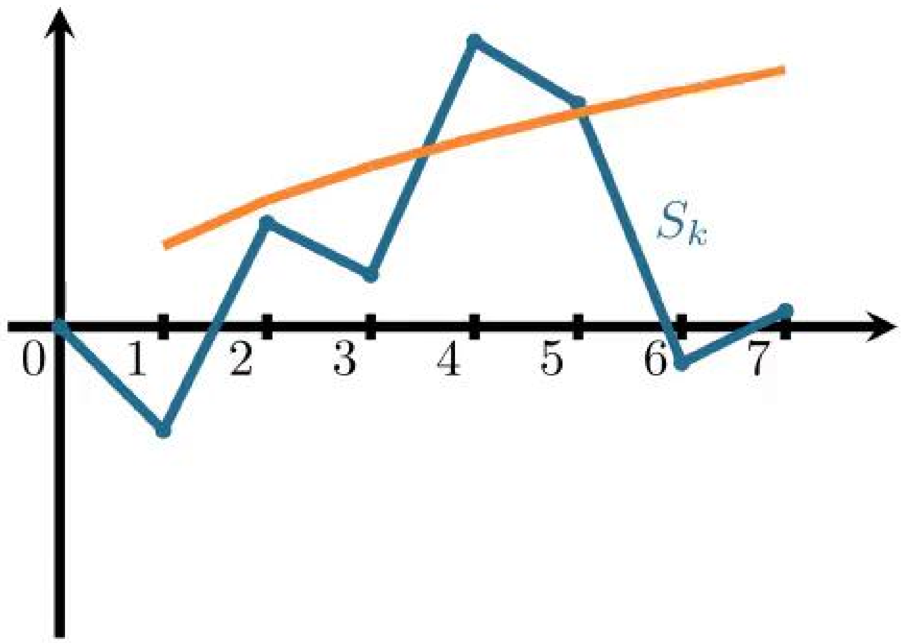
Denote $\mathcal{I}(x_0) := \big\{i \in \{1, \ldots, n\} : X_i \preccurlyeq x_0\big\}$, $n(x_0) := |\mathcal{I}(x_0)|$.

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 $H_0(X_i)$, $i \in \{1, \ldots, m\}$ (sub-Gaussian case)

Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) : \eta(x_0) < \tau$.





Denote $\mathcal{I}(x_0) := \big\{i \in \{1, \ldots, n\} : X_i \preccurlyeq x_0\big\}$, $n(x_0) := |\mathcal{I}(x_0)|$.

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.

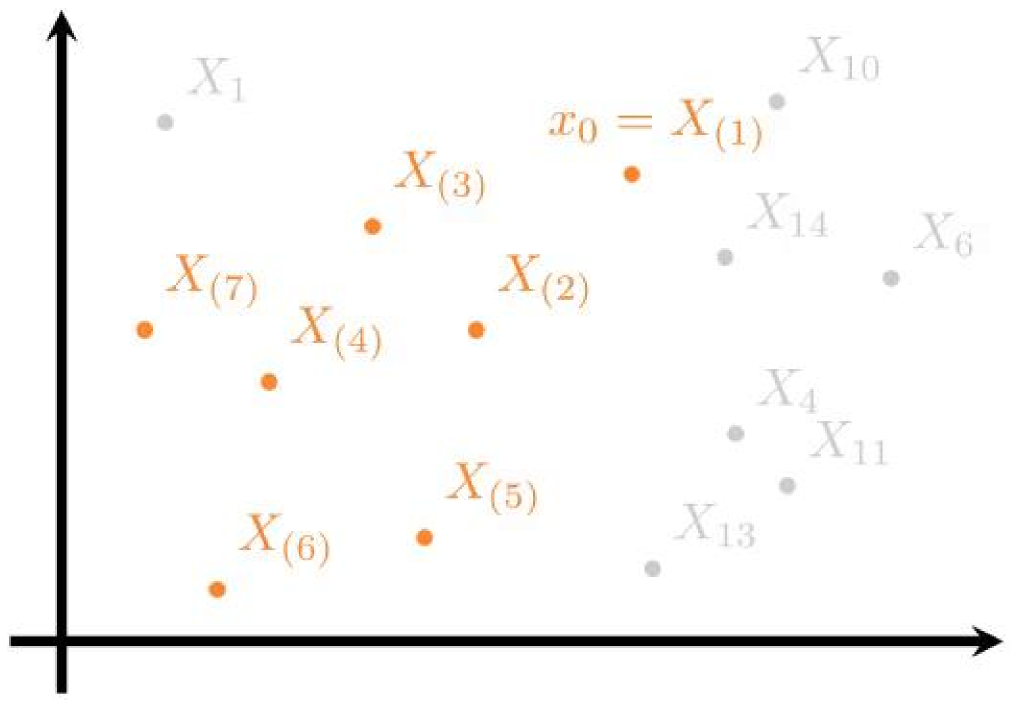
Let

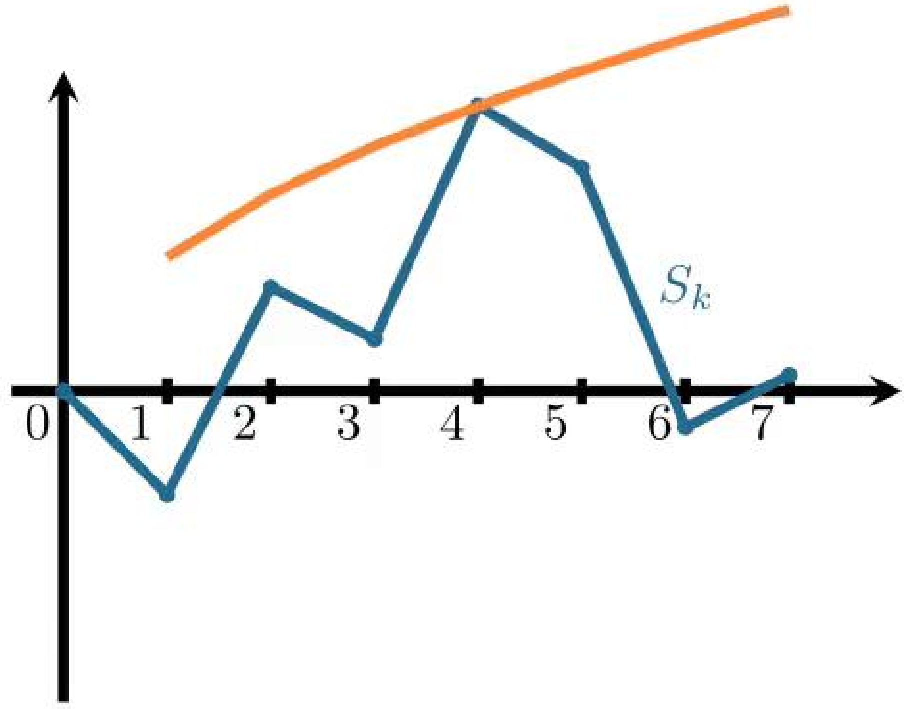
$$S_k := \sum_{j=1}^{k} \frac{Y_{(j)} - \tau}{\sigma}.$$

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 $H_0(X_i)$, $i \in \{1, \dots, m\}$ (sub-Gaussian case)

Given $x_0 \in \mathbb{R}^d$, we seek a p-value for $H_0(x_0) : \eta(x_0) < \tau$.





Denote $\mathcal{I}(x_0) := \big\{i \in \{1, \dots, n\} : X_i \preccurlyeq x_0\big\}$, $n(x_0) := |\mathcal{I}(x_0)|$.

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.

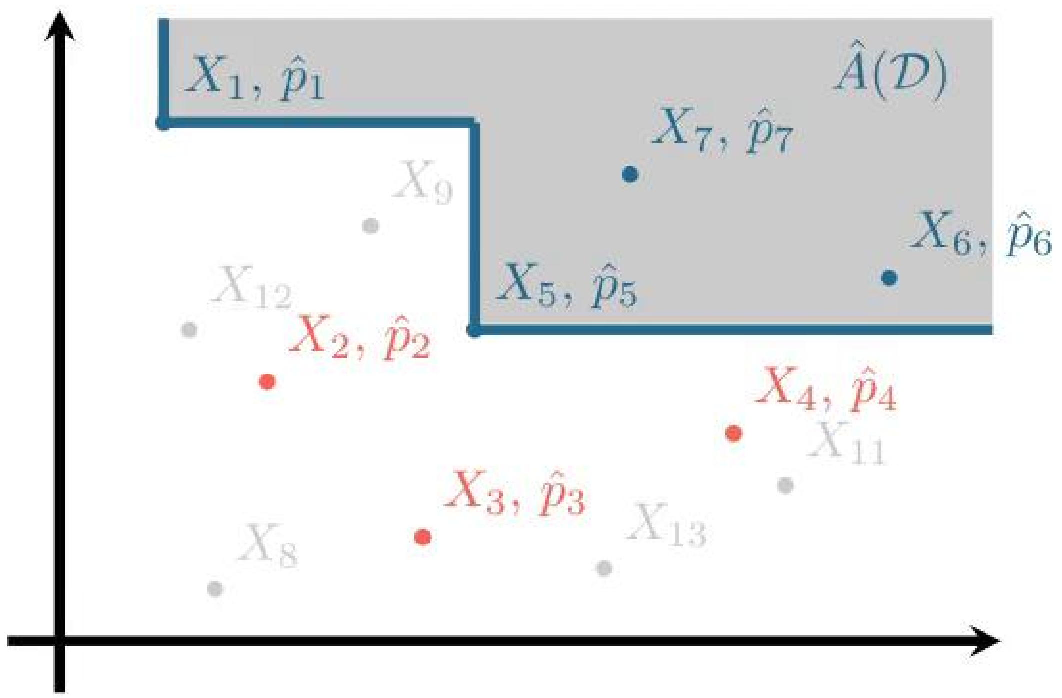
Let

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

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 $H_0(x_0) : \eta(x_0) < \tau$.

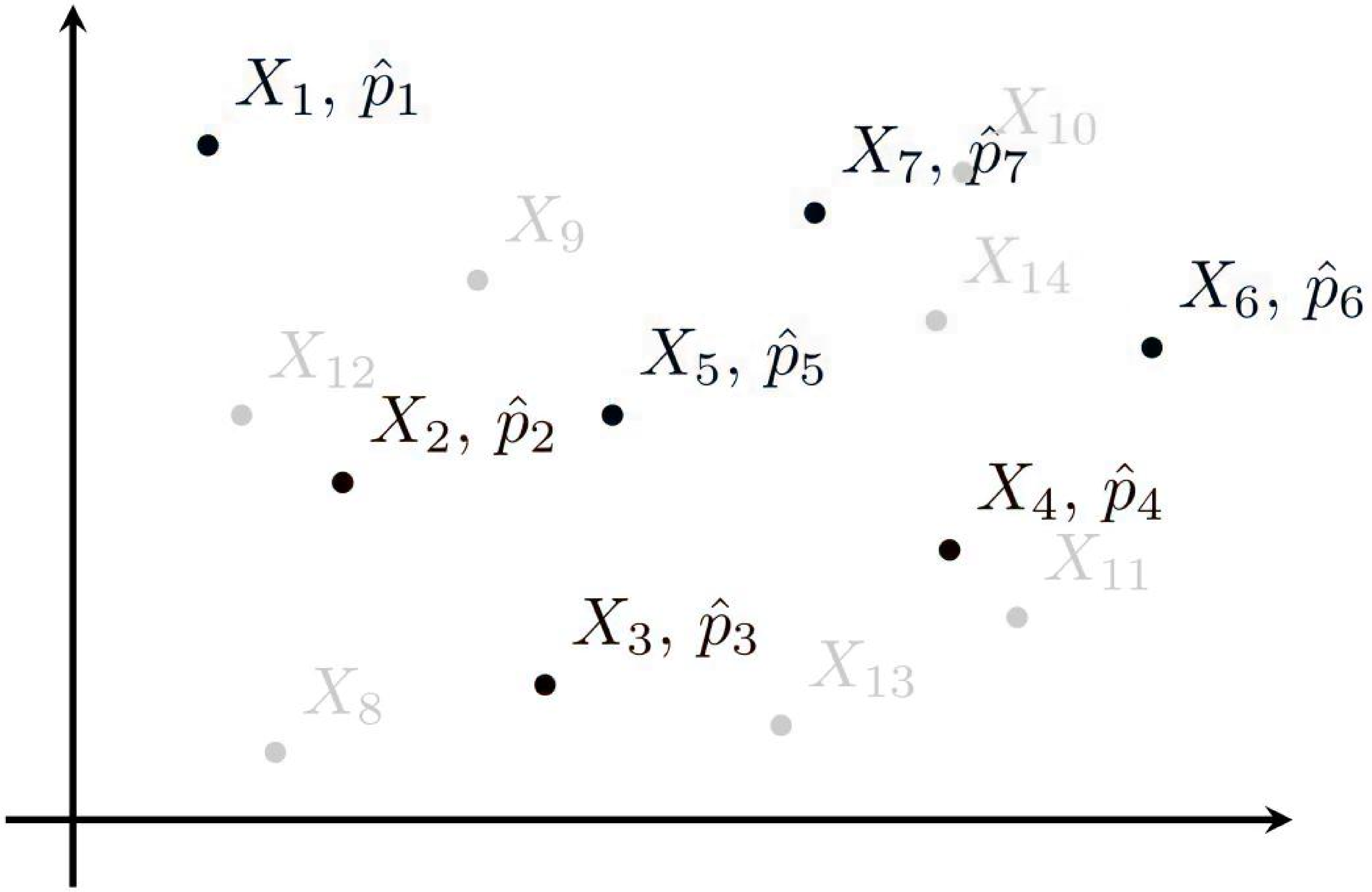


High-level strategy:

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

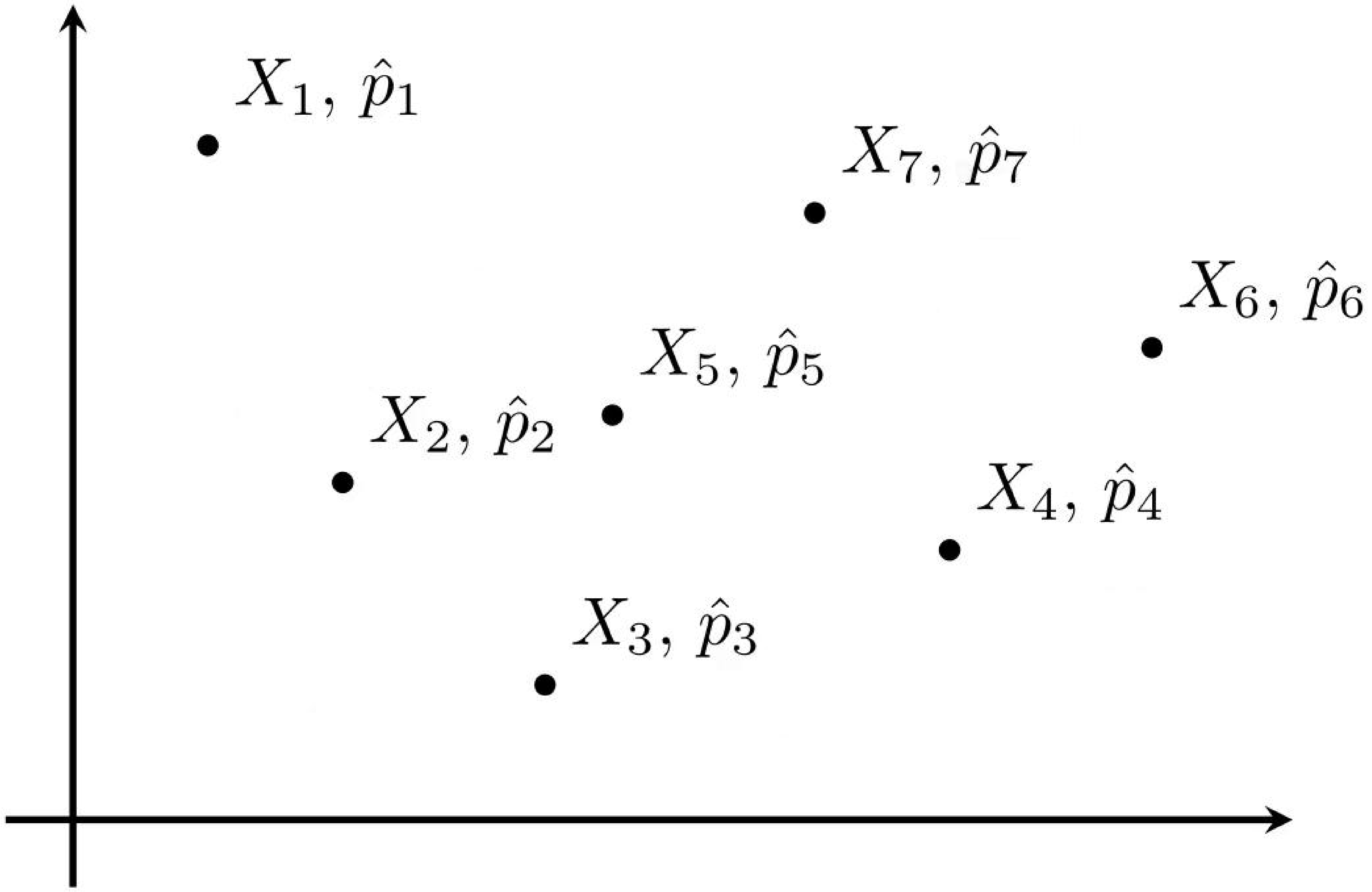
Multiple testing procedure

Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \ldots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



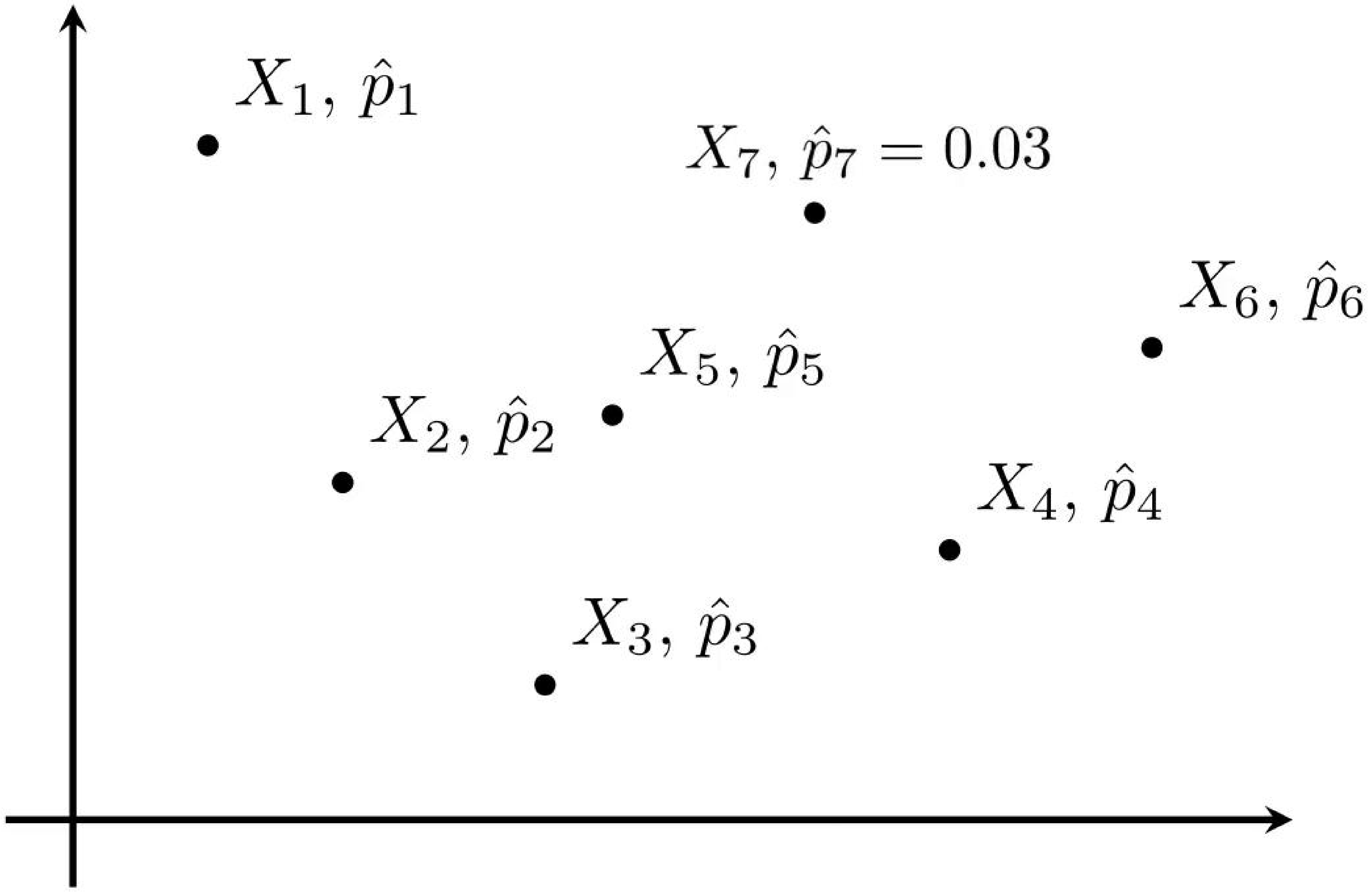
Multiple testing procedure

Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \dots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



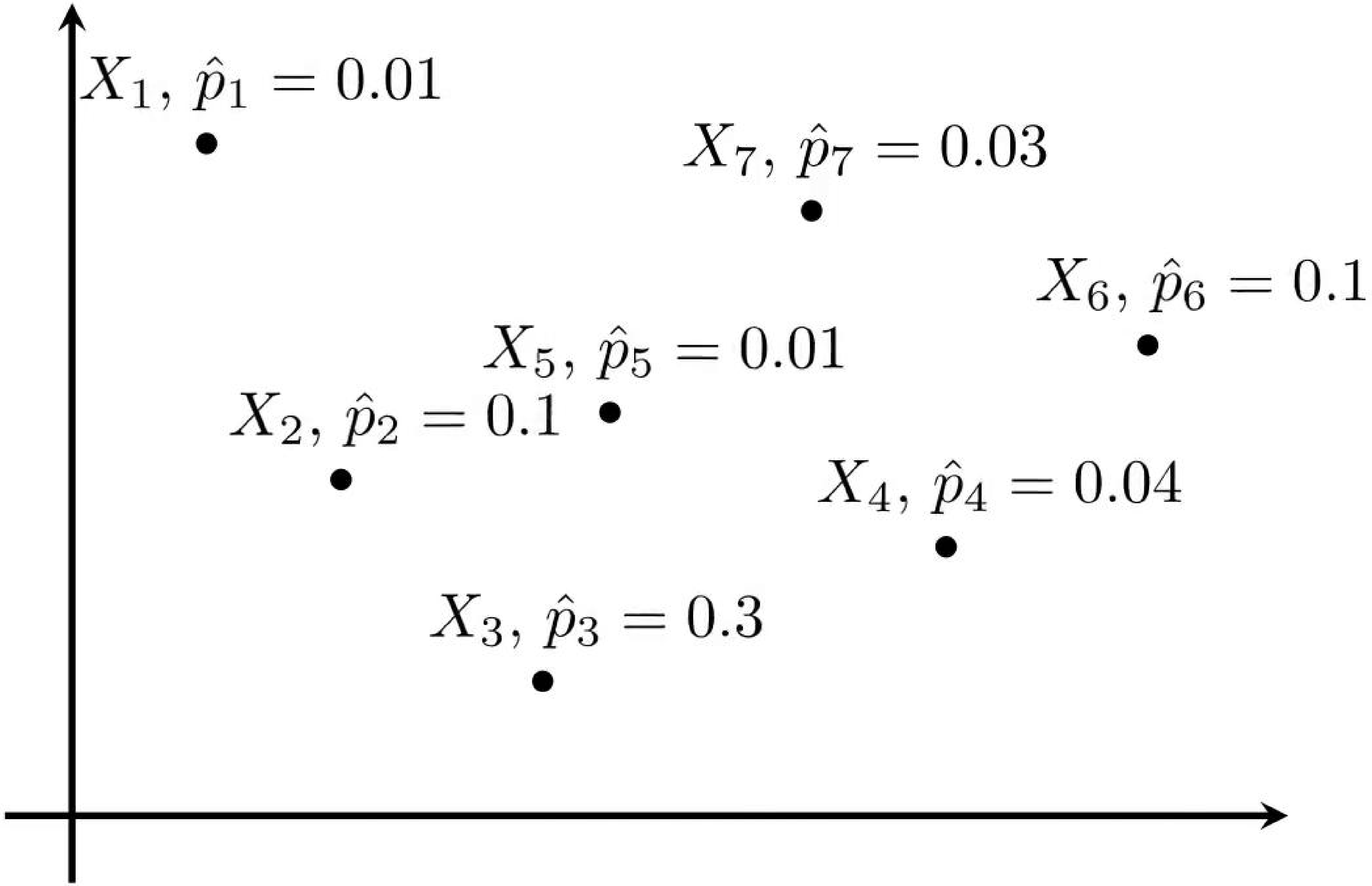
Multiple testing procedure

Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \ldots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



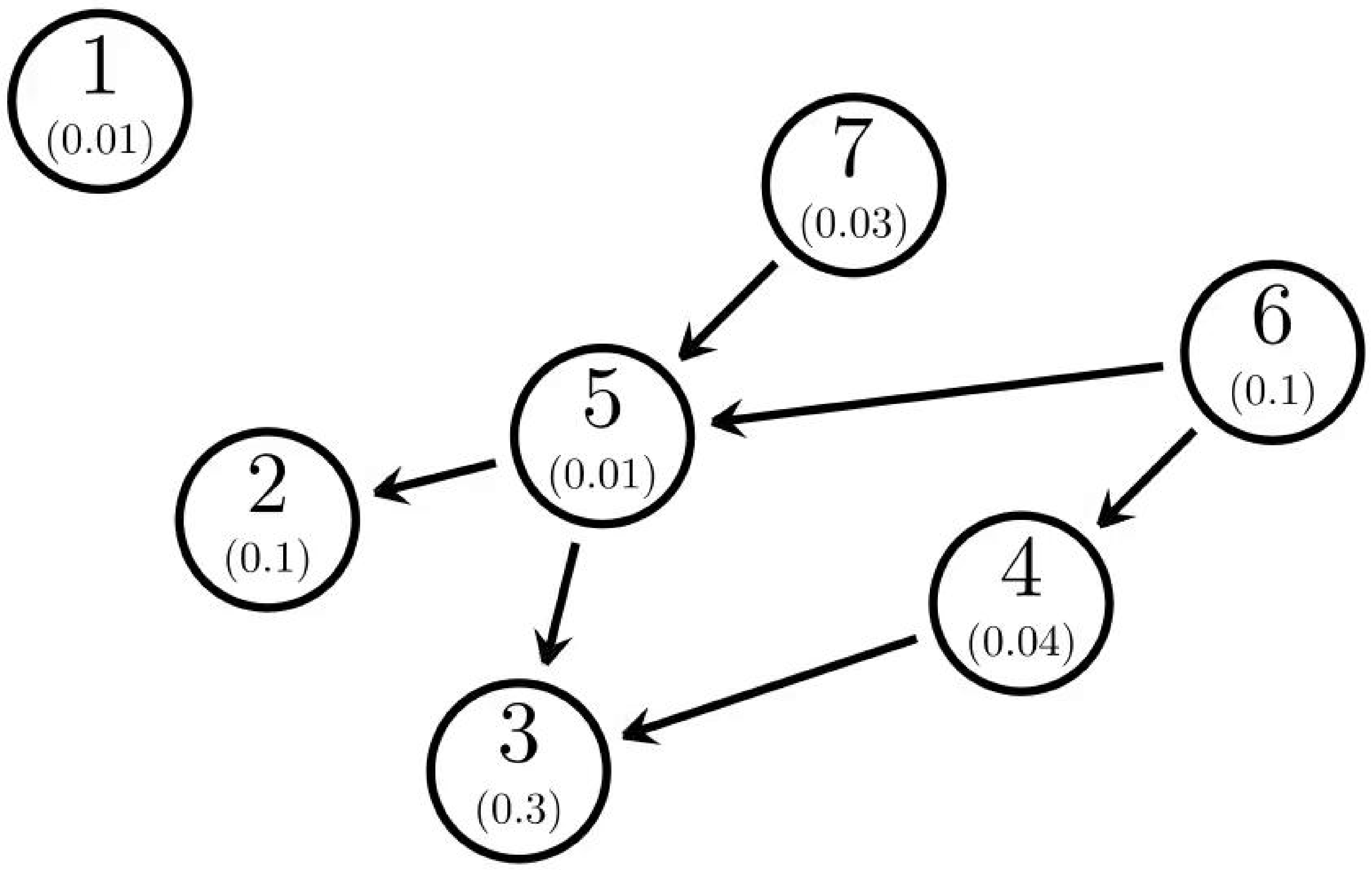
Multiple testing procedure

Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \ldots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



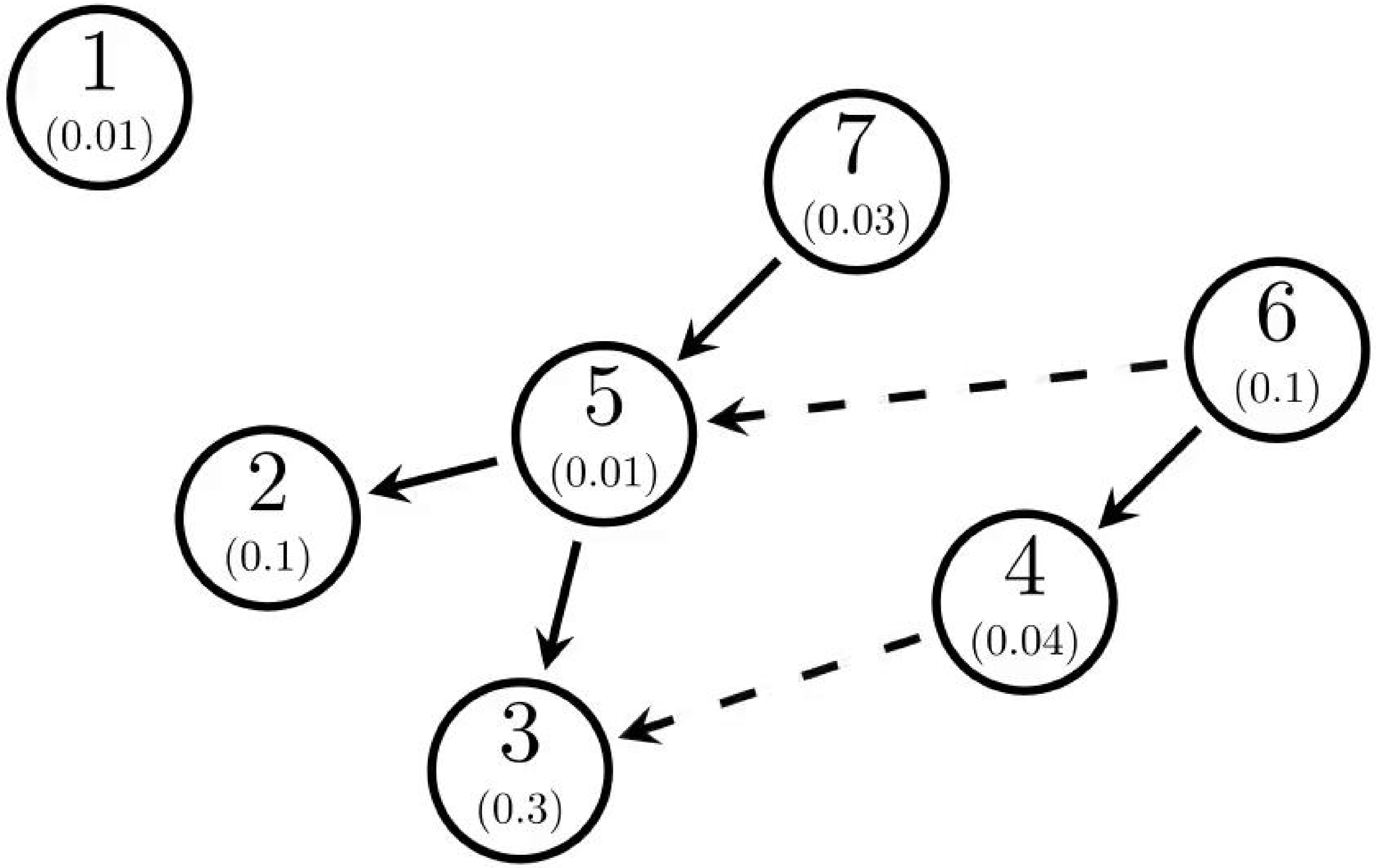
Multiple testing procedure

Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \ldots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



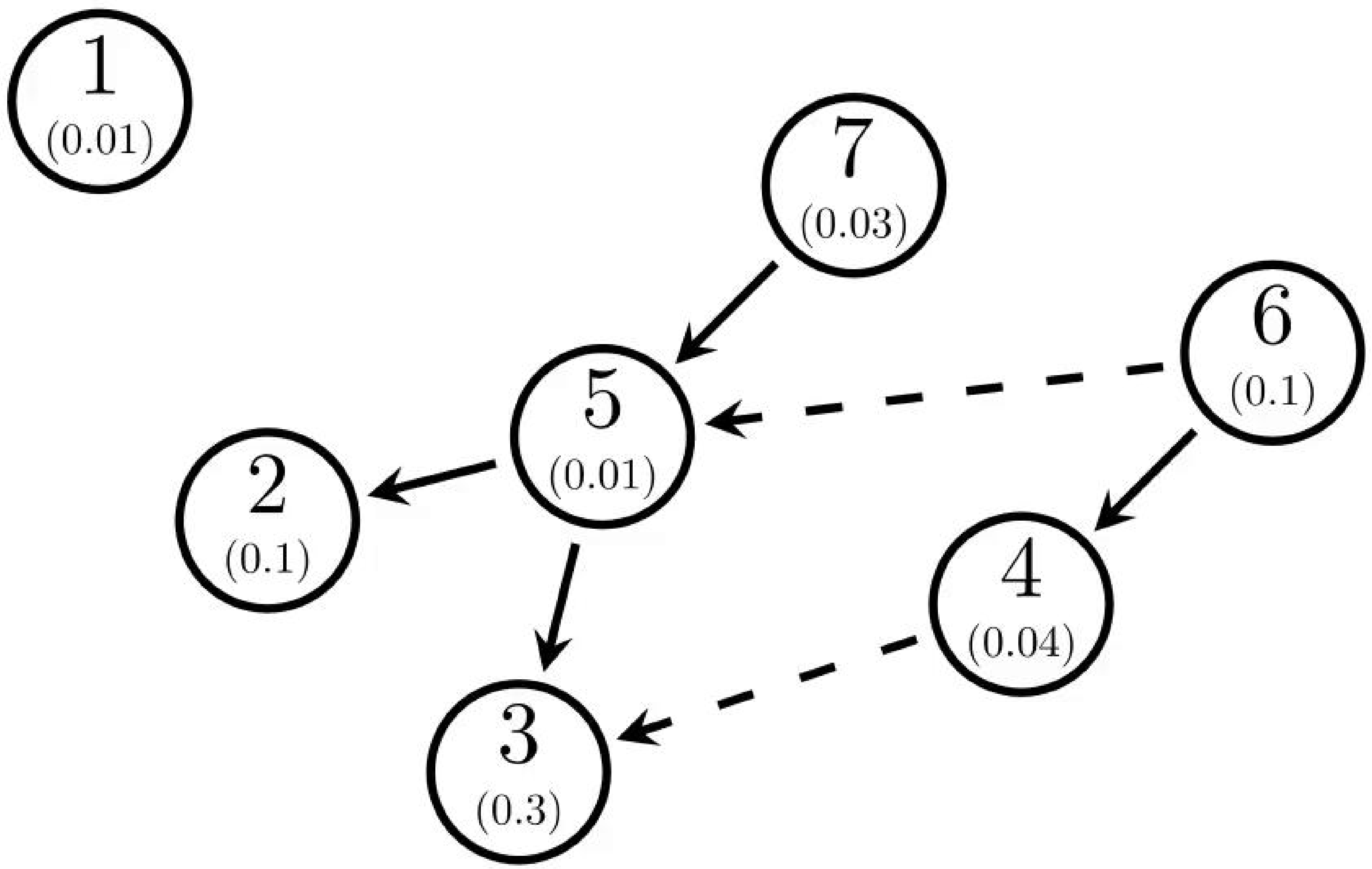
Multiple testing procedure

Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \ldots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



Multiple testing procedure

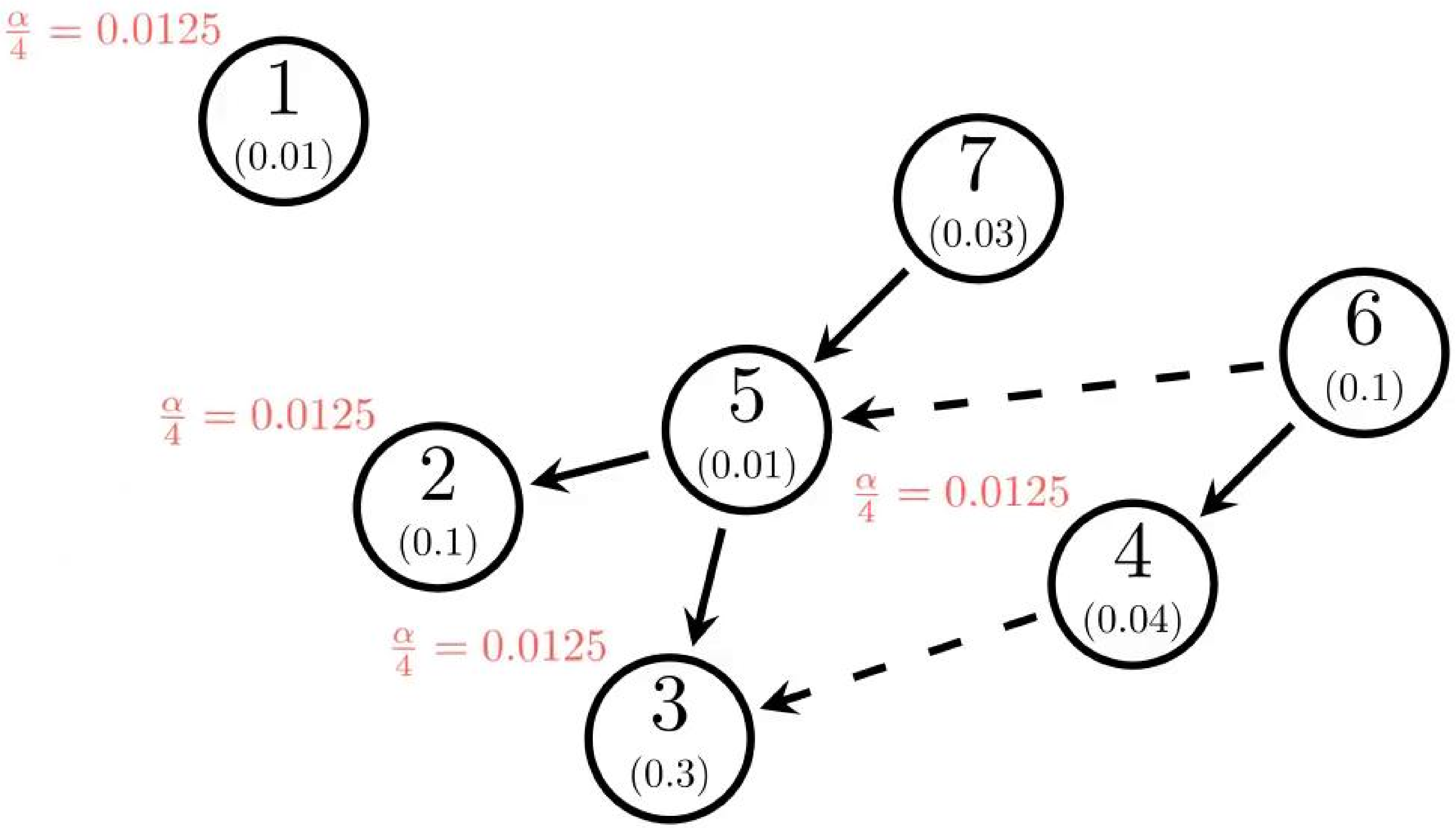
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \ldots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



Here: $\alpha = 0.05$.

Multiple testing procedure

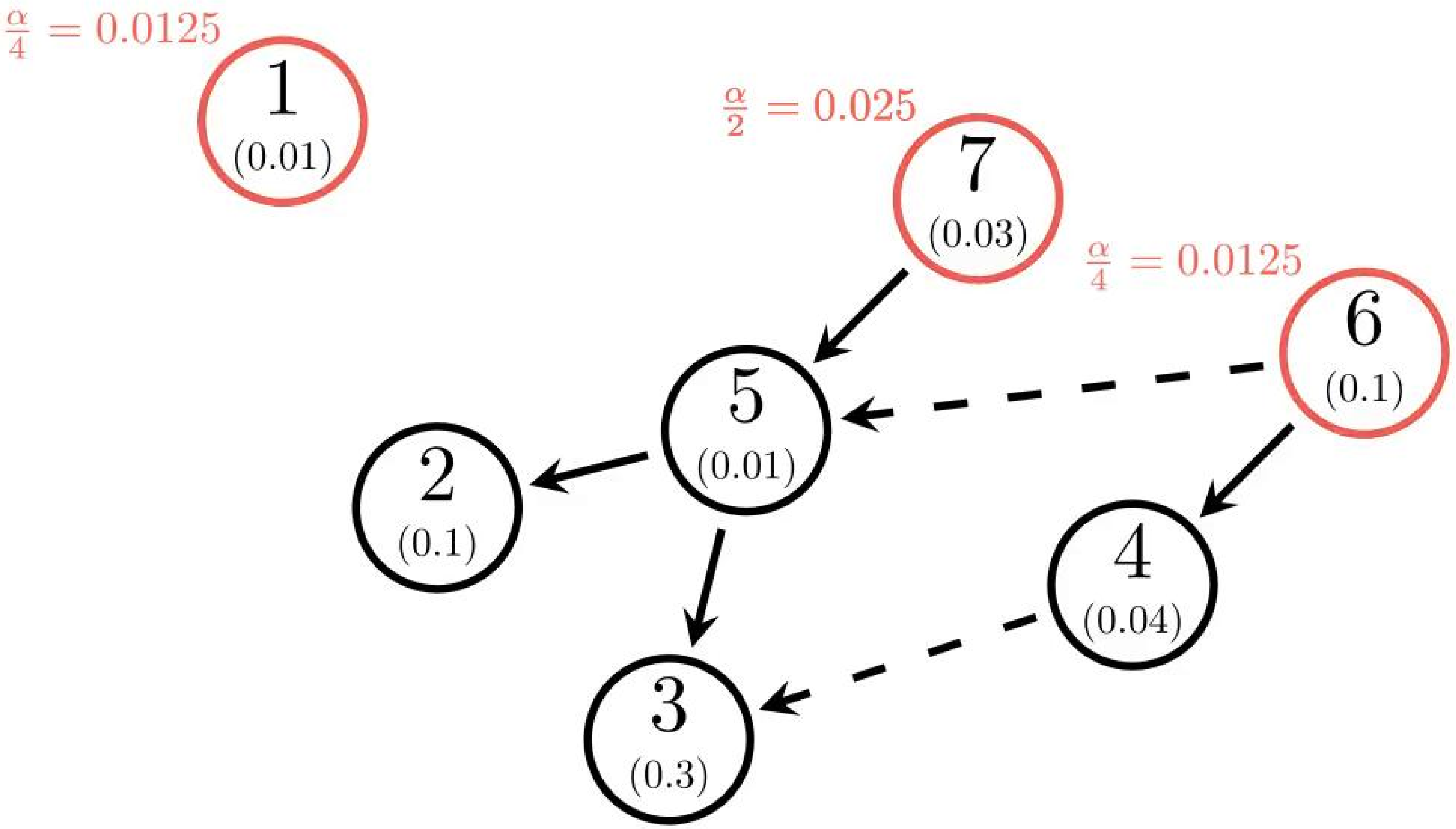
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \ldots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



Here: $\alpha = 0.05$.

Multiple testing procedure

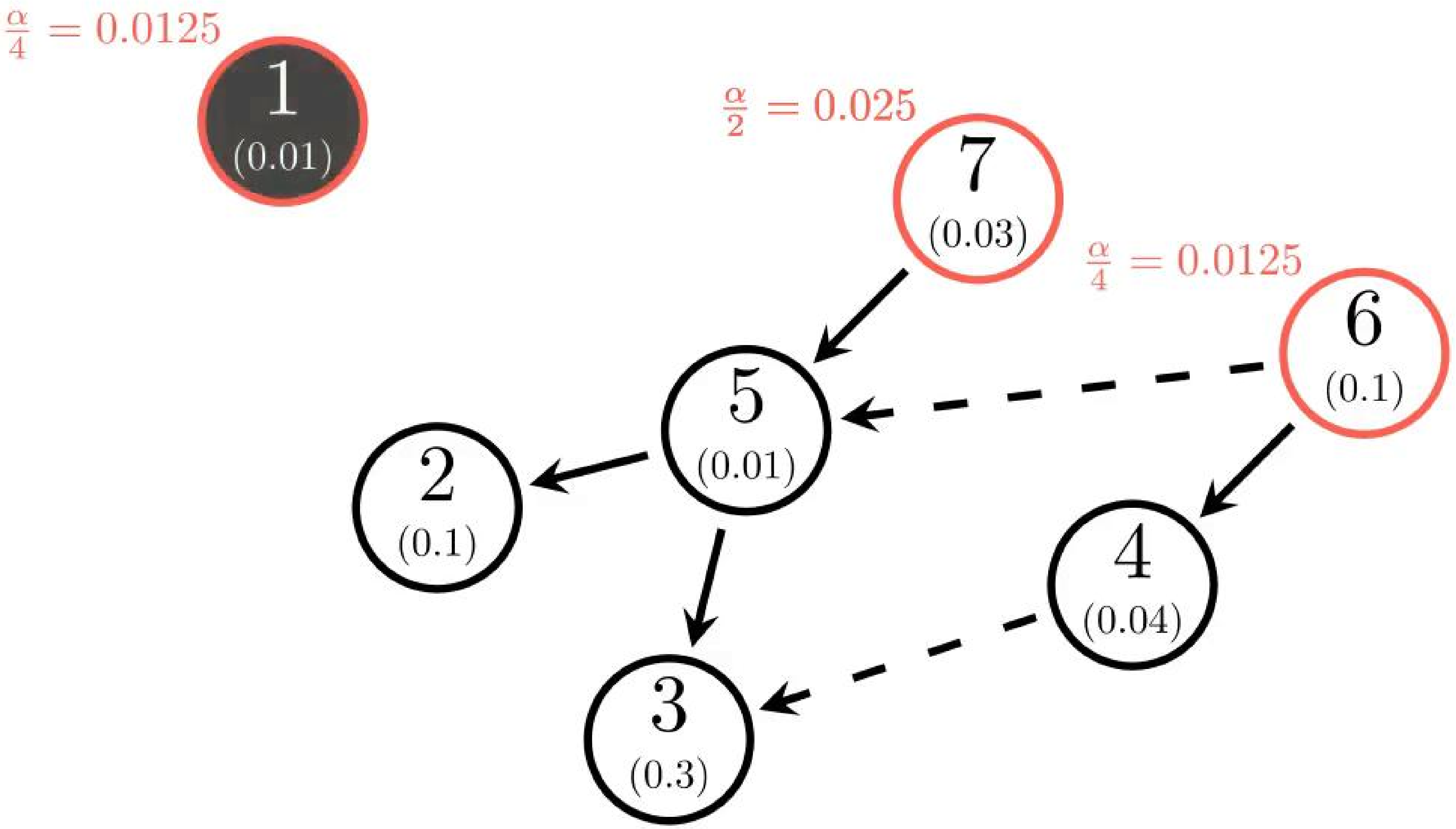
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \dots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



Here: $\alpha = 0.05$.

Multiple testing procedure

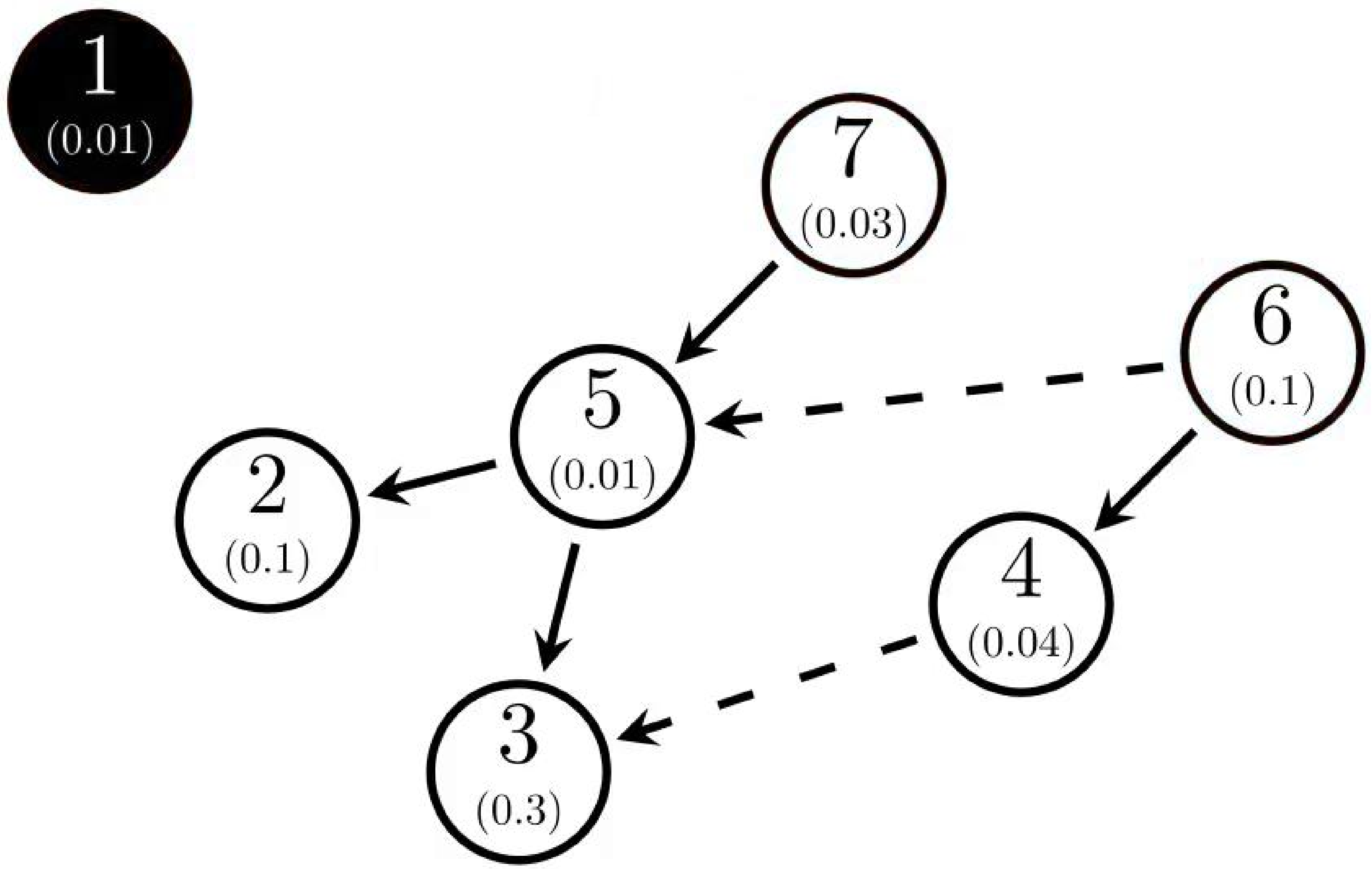
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \ldots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



Here: $\alpha = 0.05$.

Multiple testing procedure

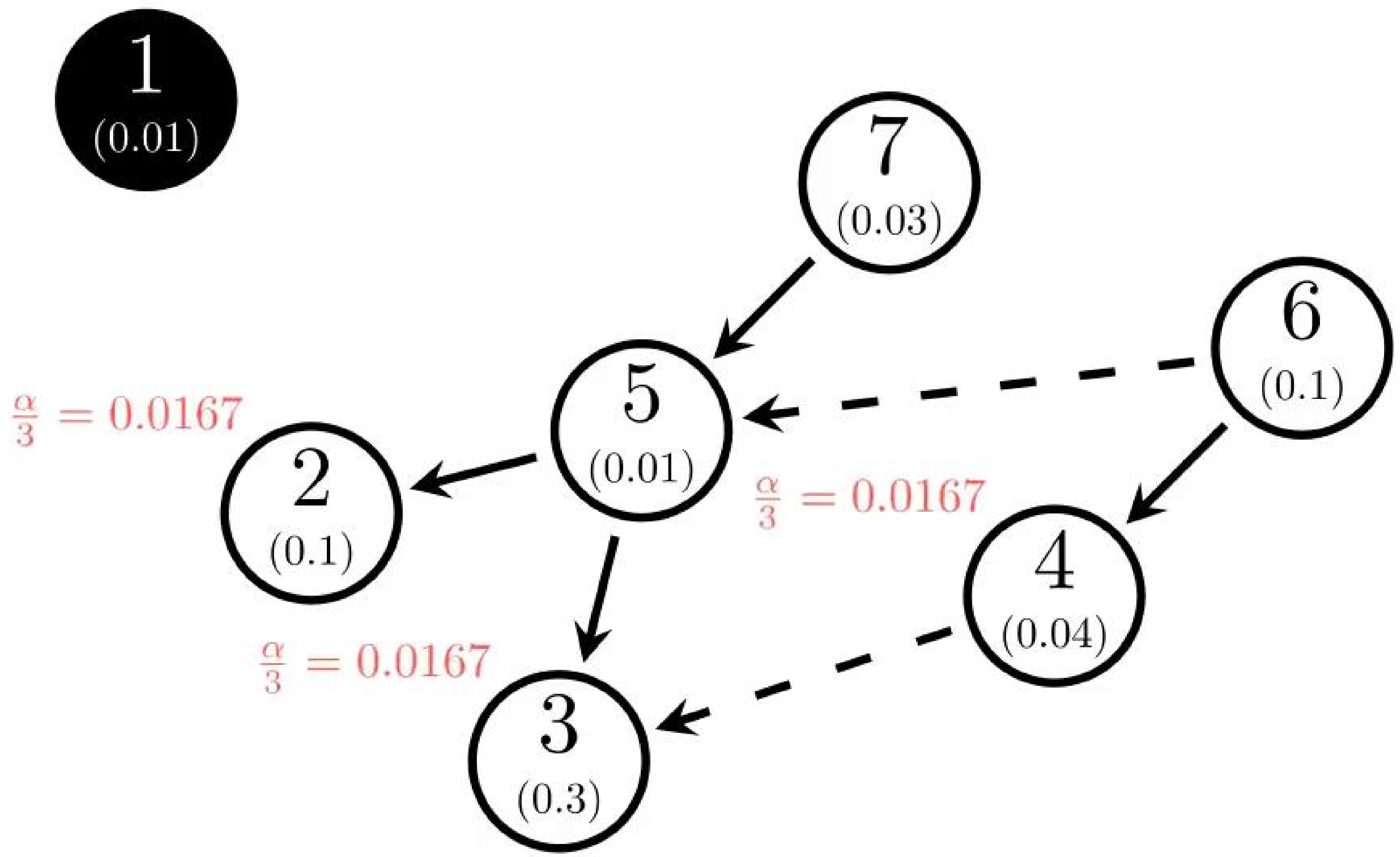
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \ldots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



Here: $\alpha = 0.05$.

Multiple testing procedure

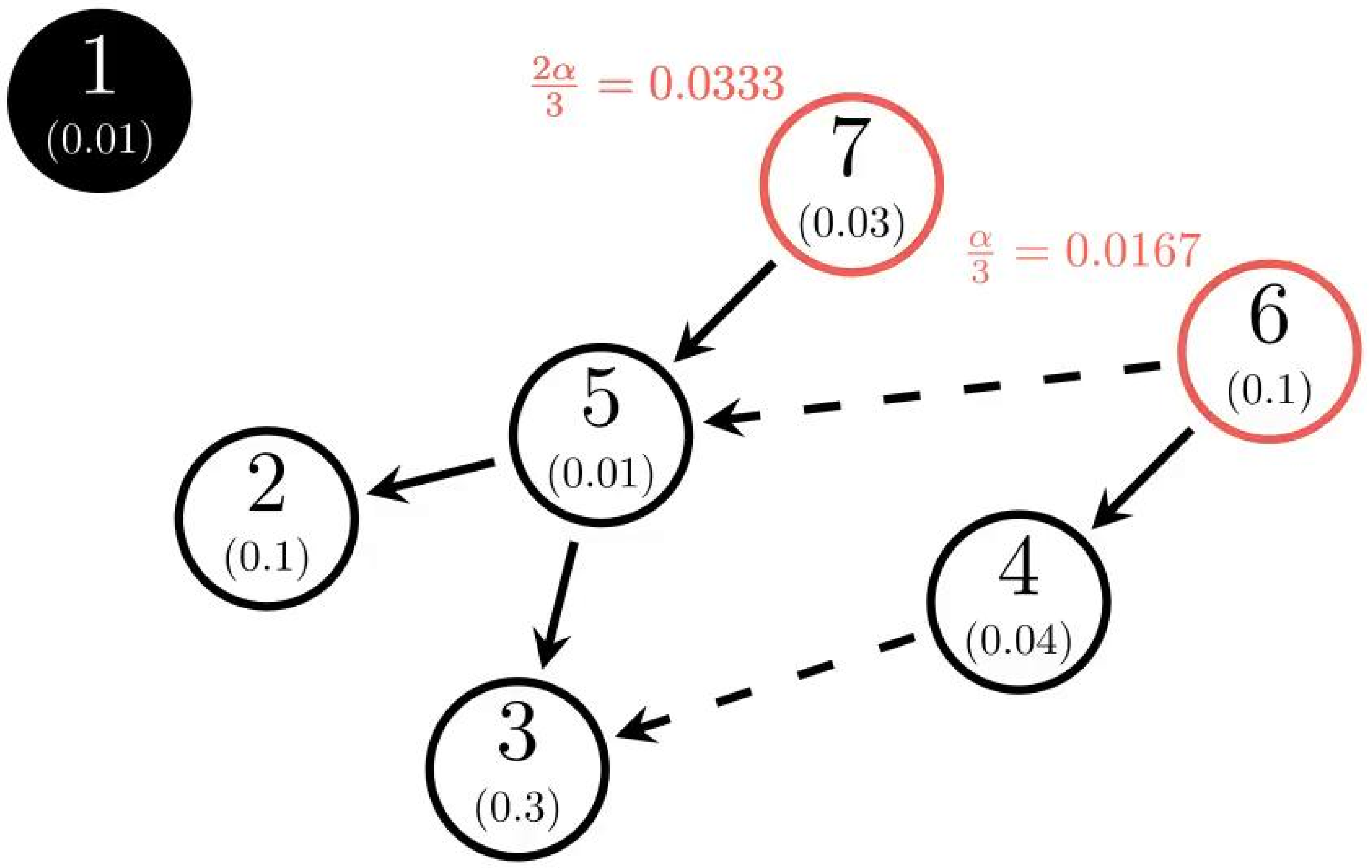
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \ldots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



Here: $\alpha = 0.05$.

Multiple testing procedure

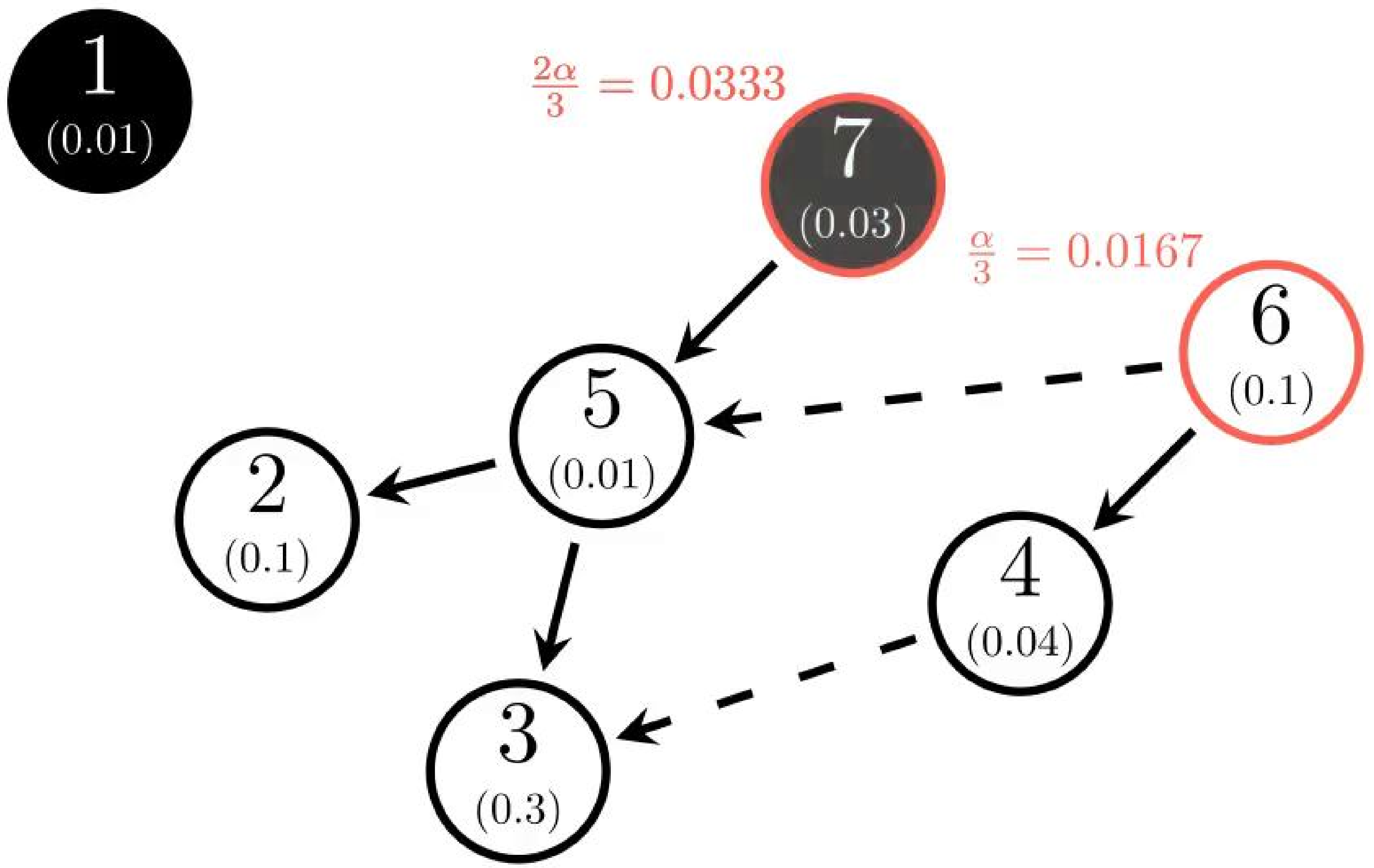
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \ldots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



Here: $\alpha = 0.05$.

Multiple testing procedure

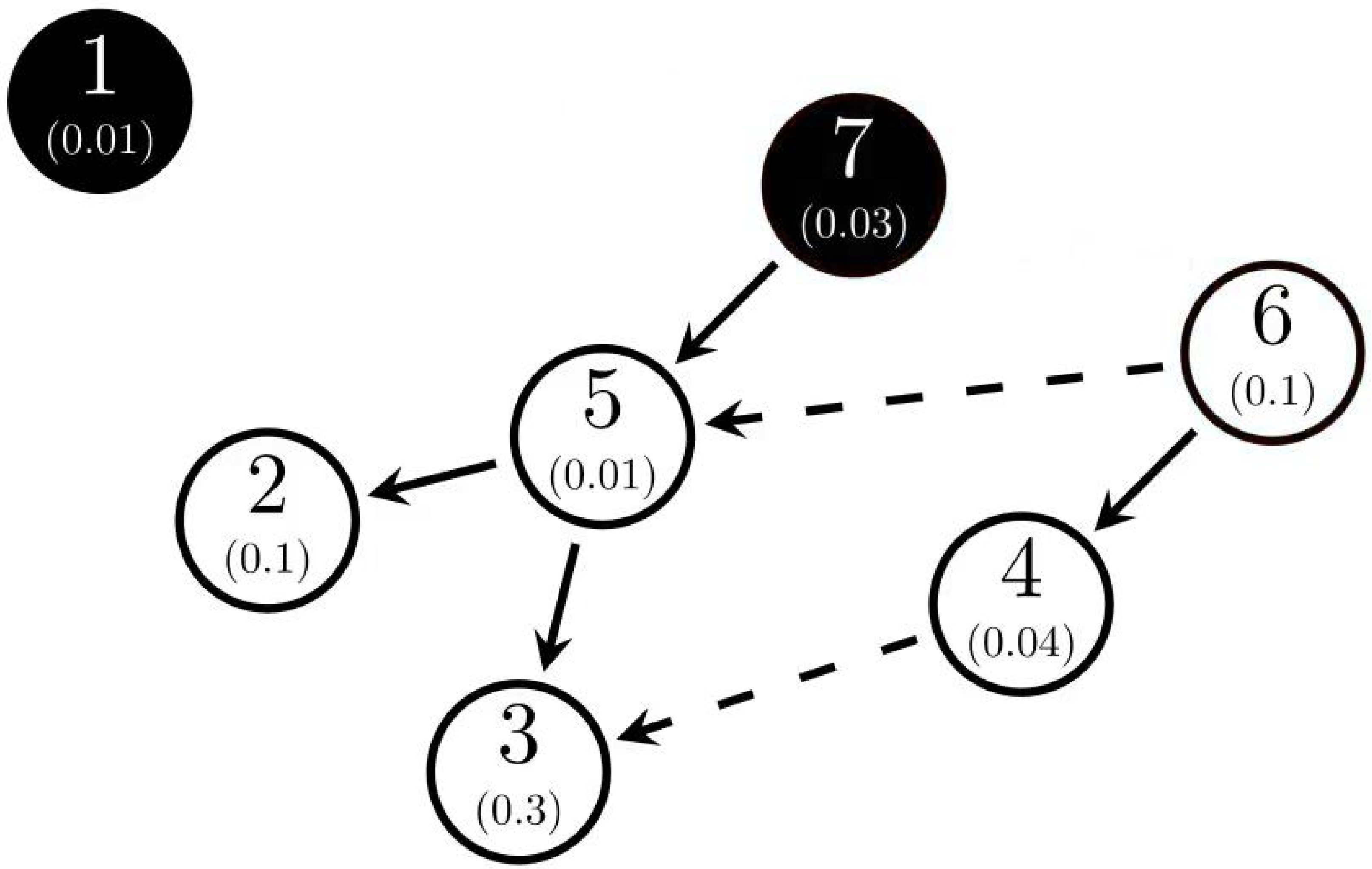
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \ldots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



Here: $\alpha = 0.05$.

Multiple testing procedure

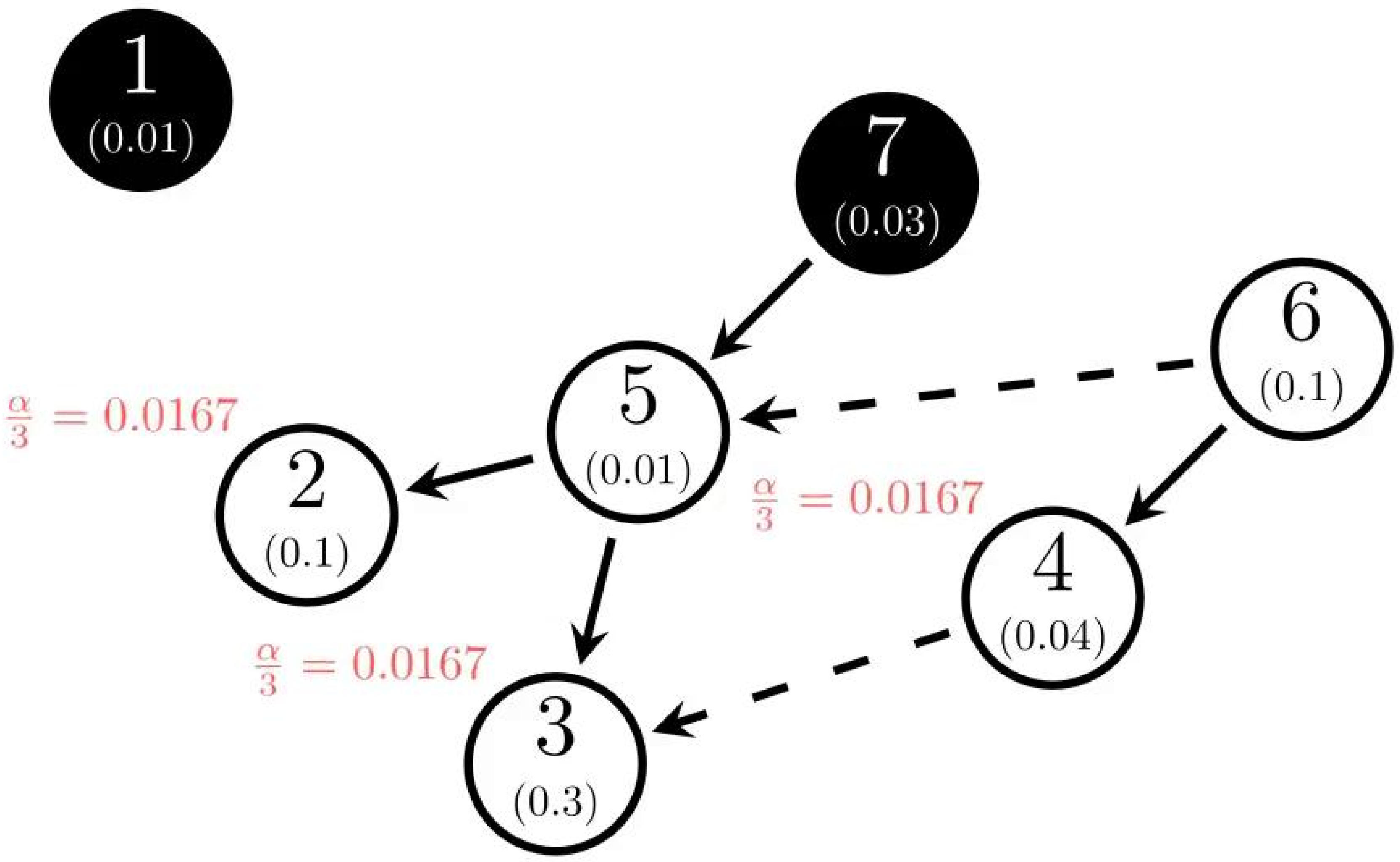
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \ldots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



Here: $\alpha = 0.05$.

Multiple testing procedure

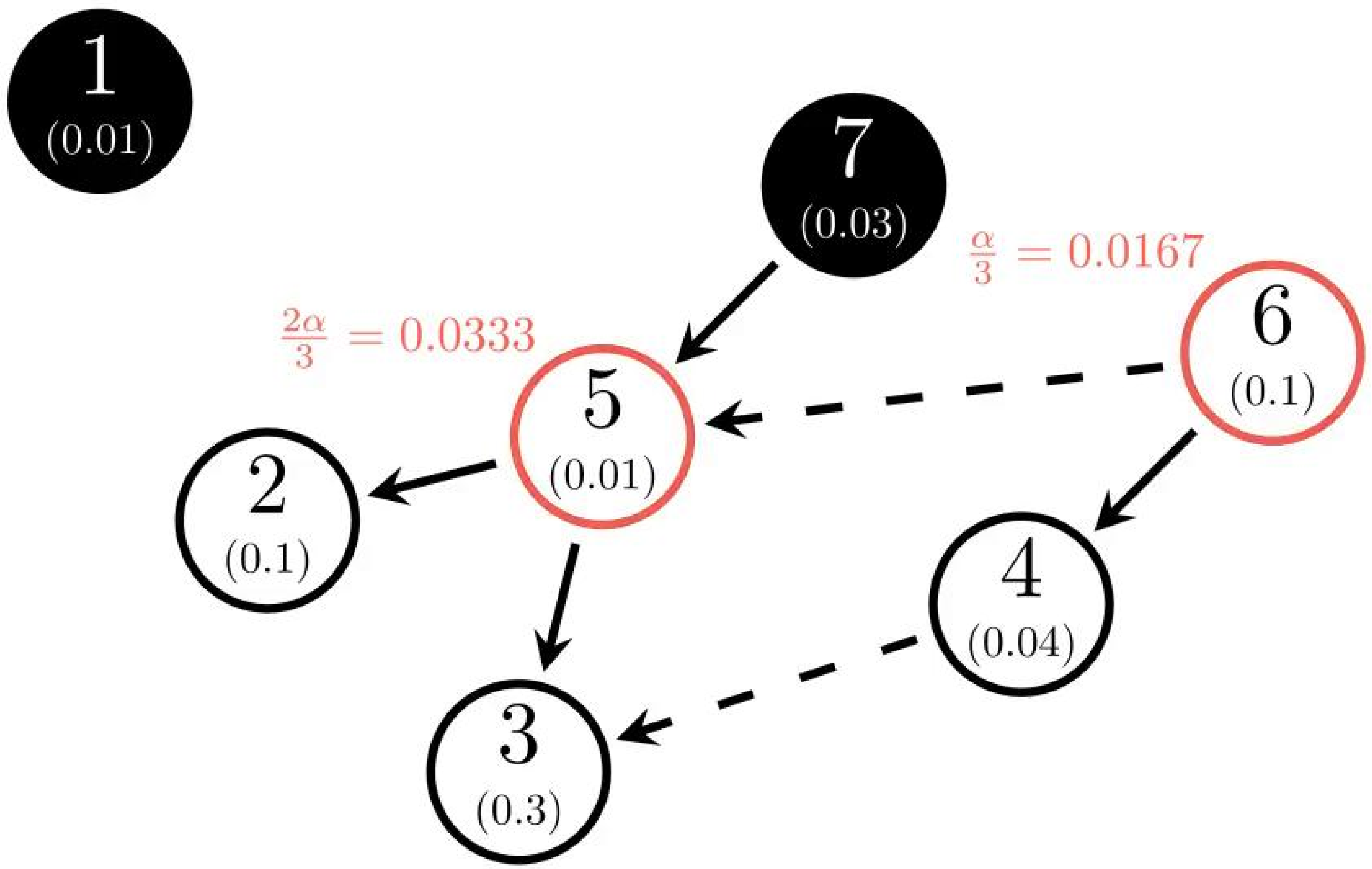
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \ldots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



Here: $\alpha = 0.05$.

Multiple testing procedure

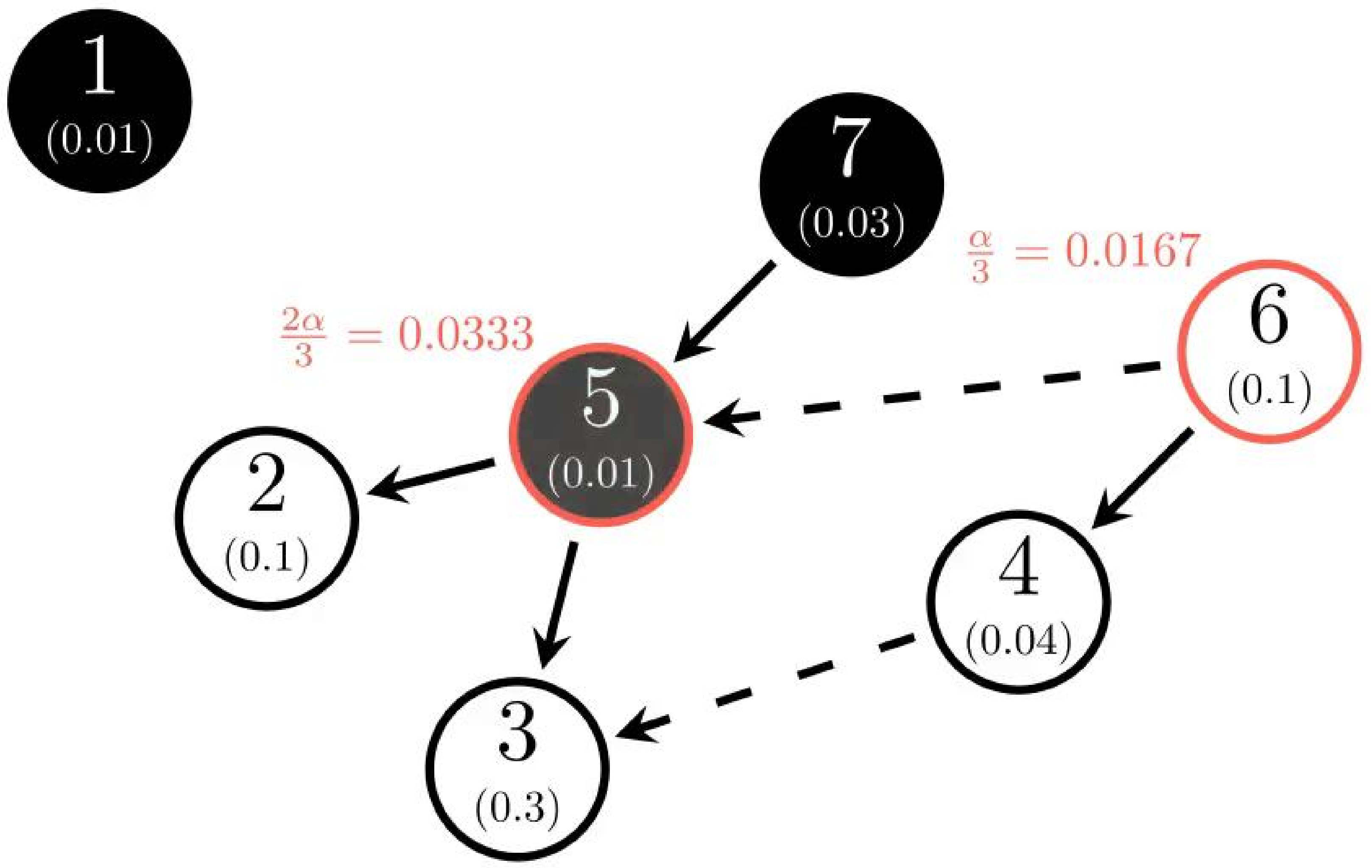
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \dots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



Here: $\alpha = 0.05$.

Multiple testing procedure

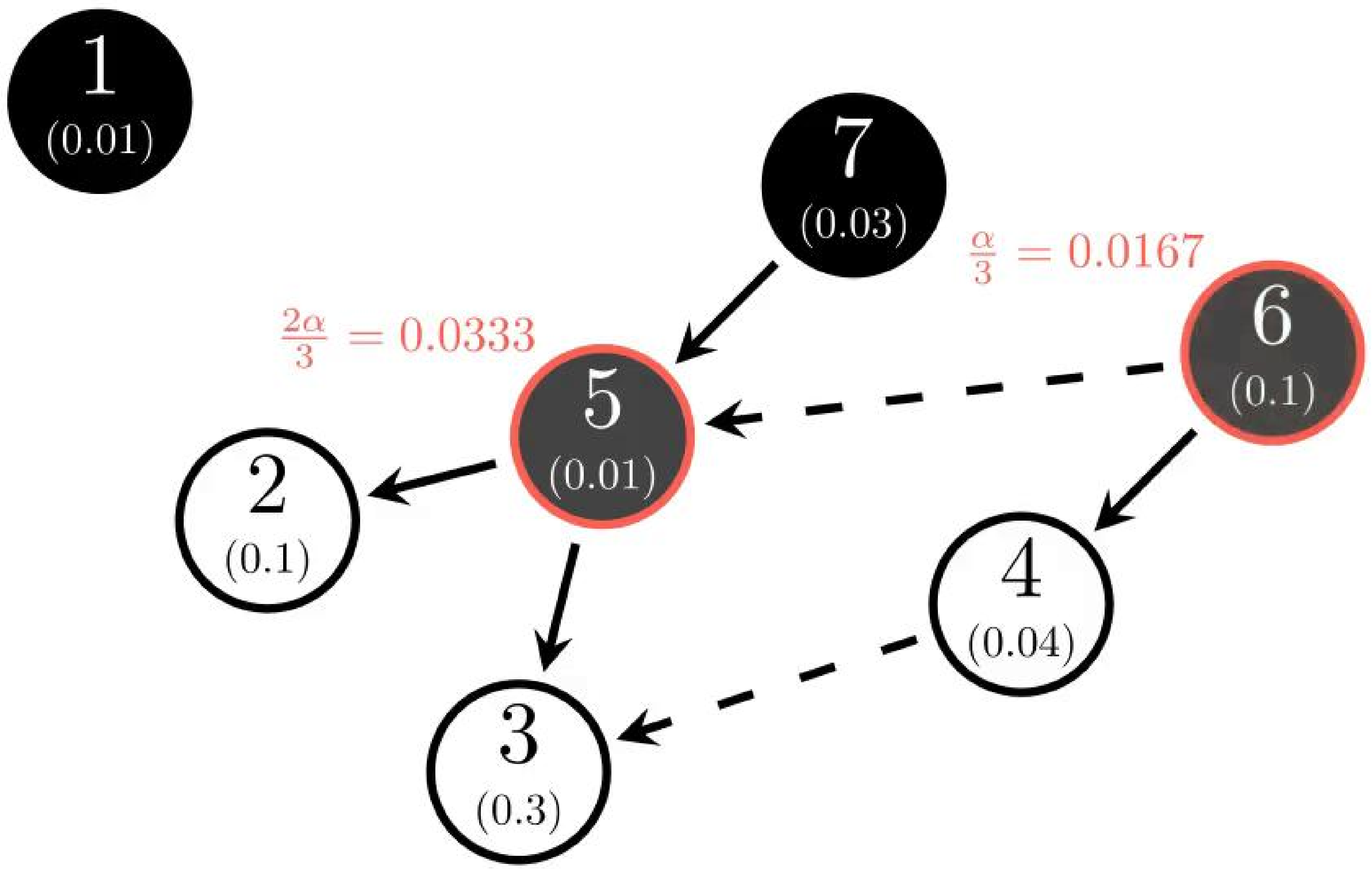
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \ldots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



Here: $\alpha = 0.05$.

Multiple testing procedure

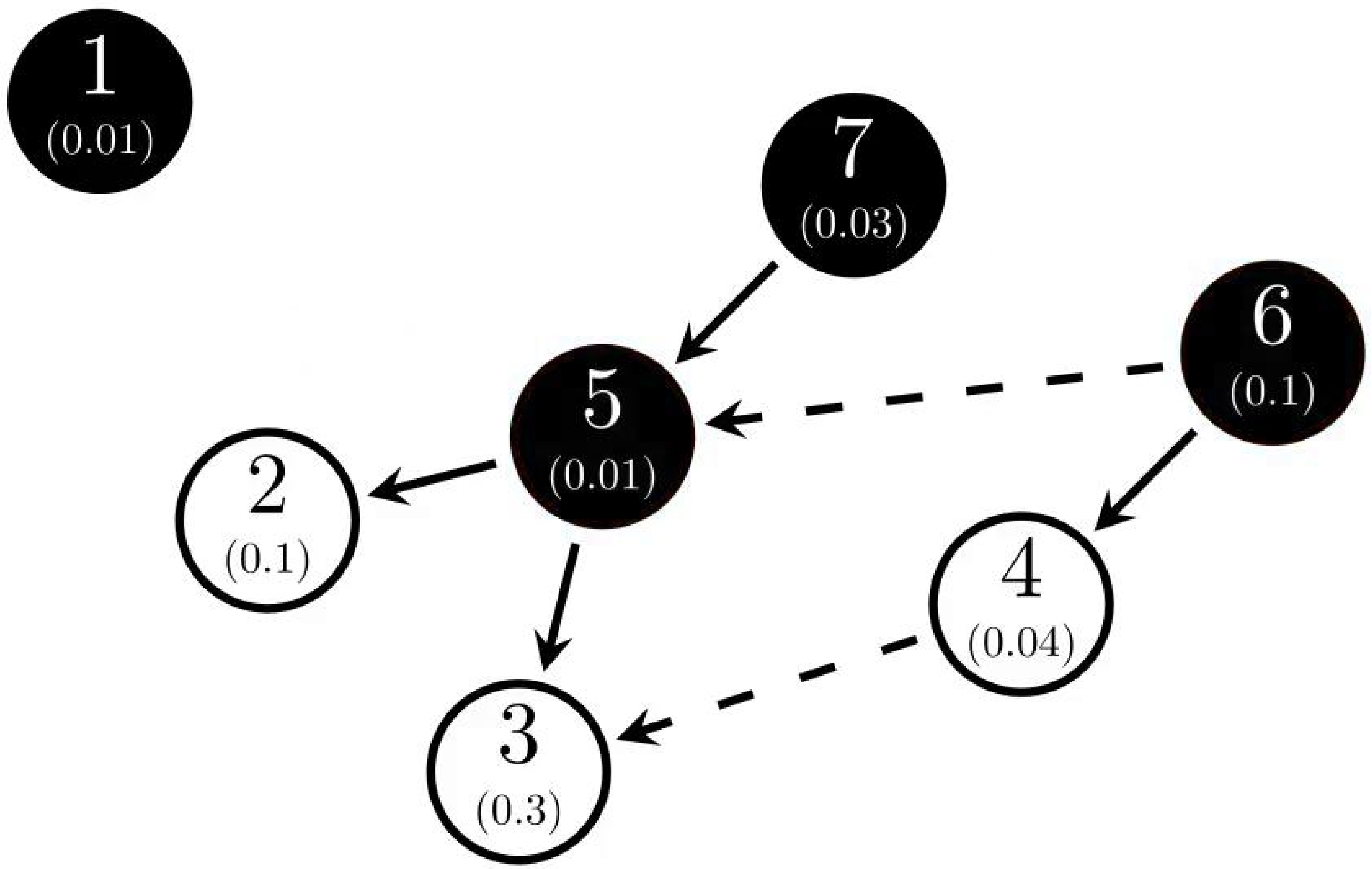
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \ldots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



Here: $\alpha = 0.05$.

Multiple testing procedure

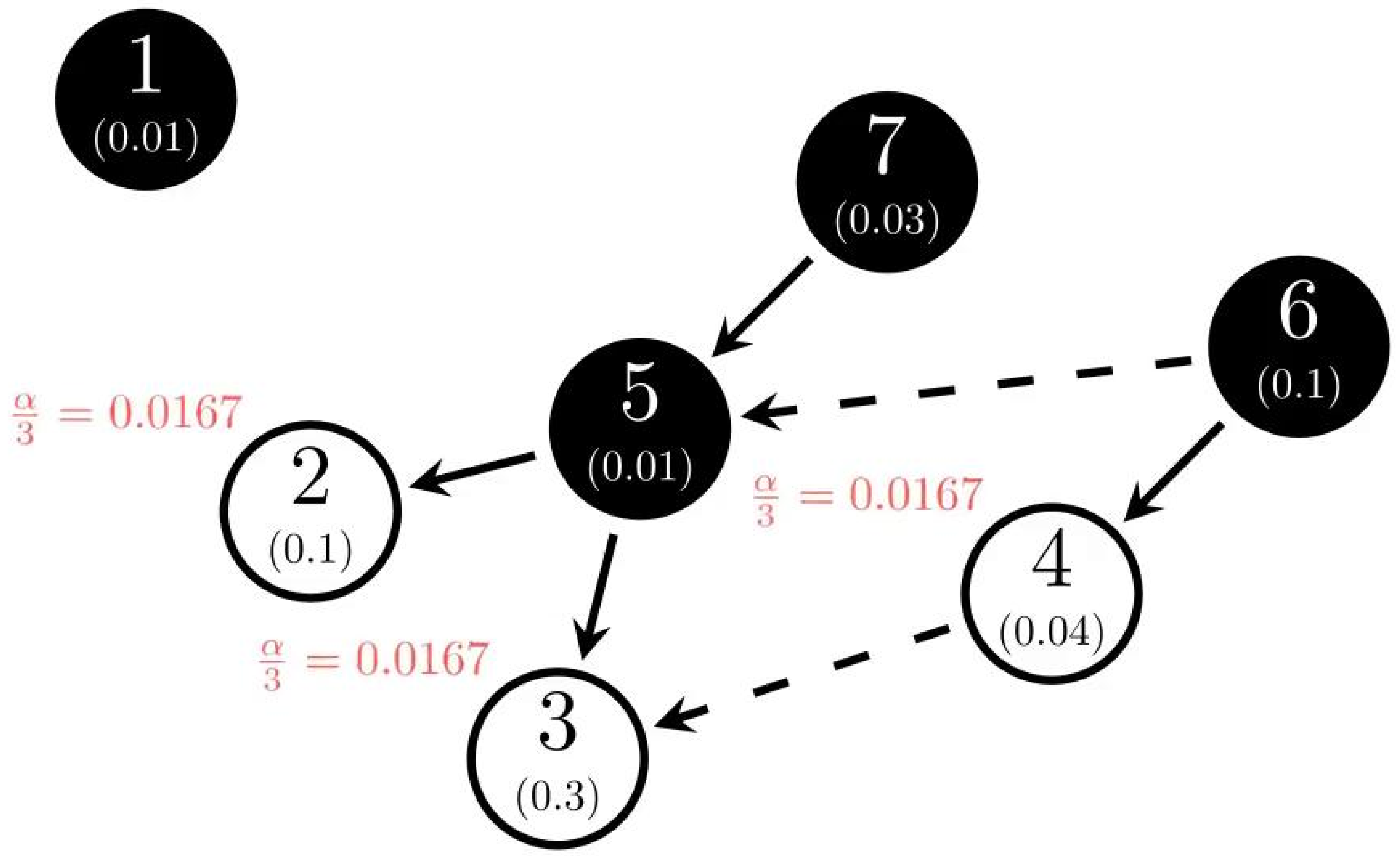
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \ldots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



Here: $\alpha = 0.05$.

Multiple testing procedure

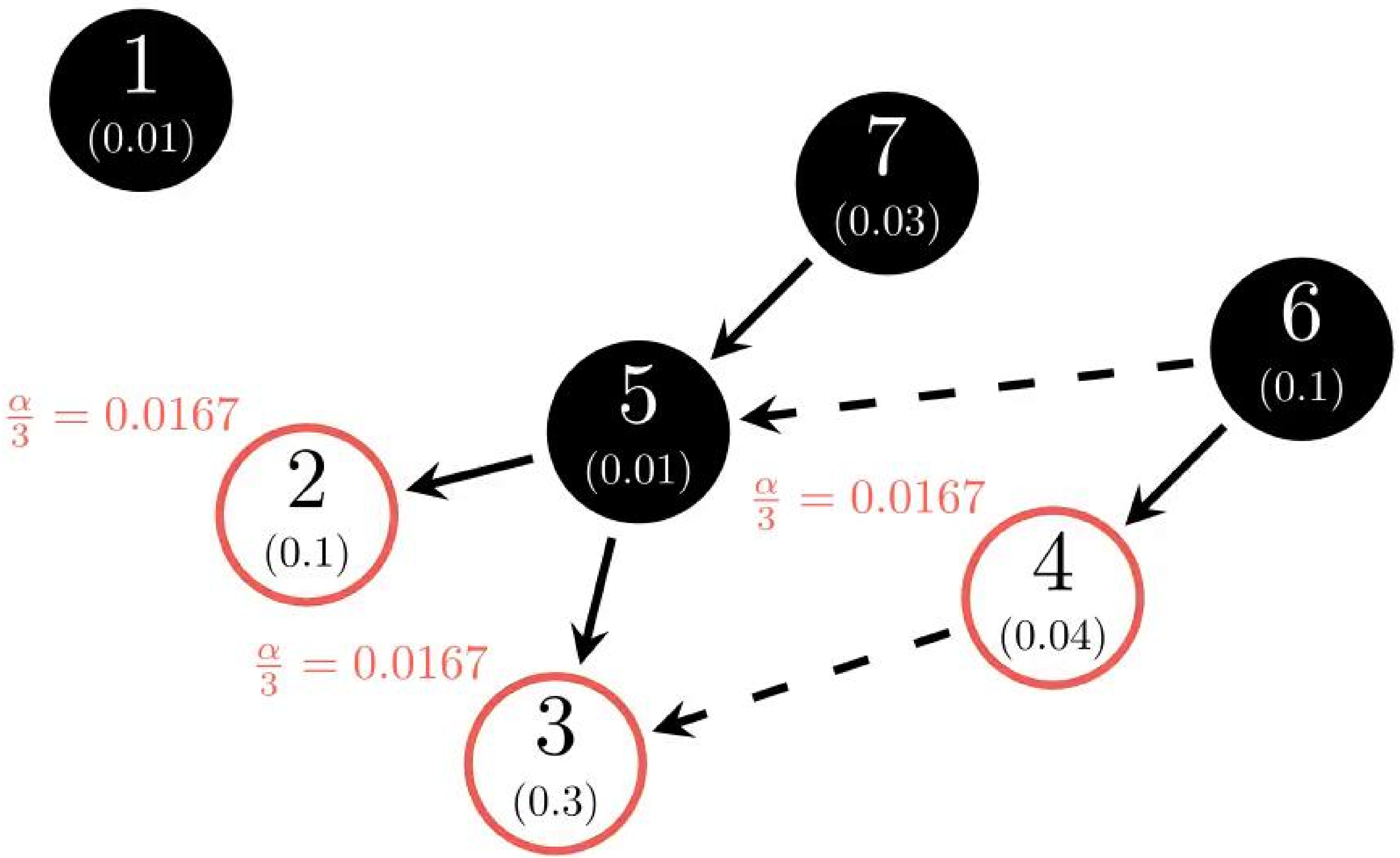
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \ldots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



Here: $\alpha = 0.05$.

Multiple testing procedure

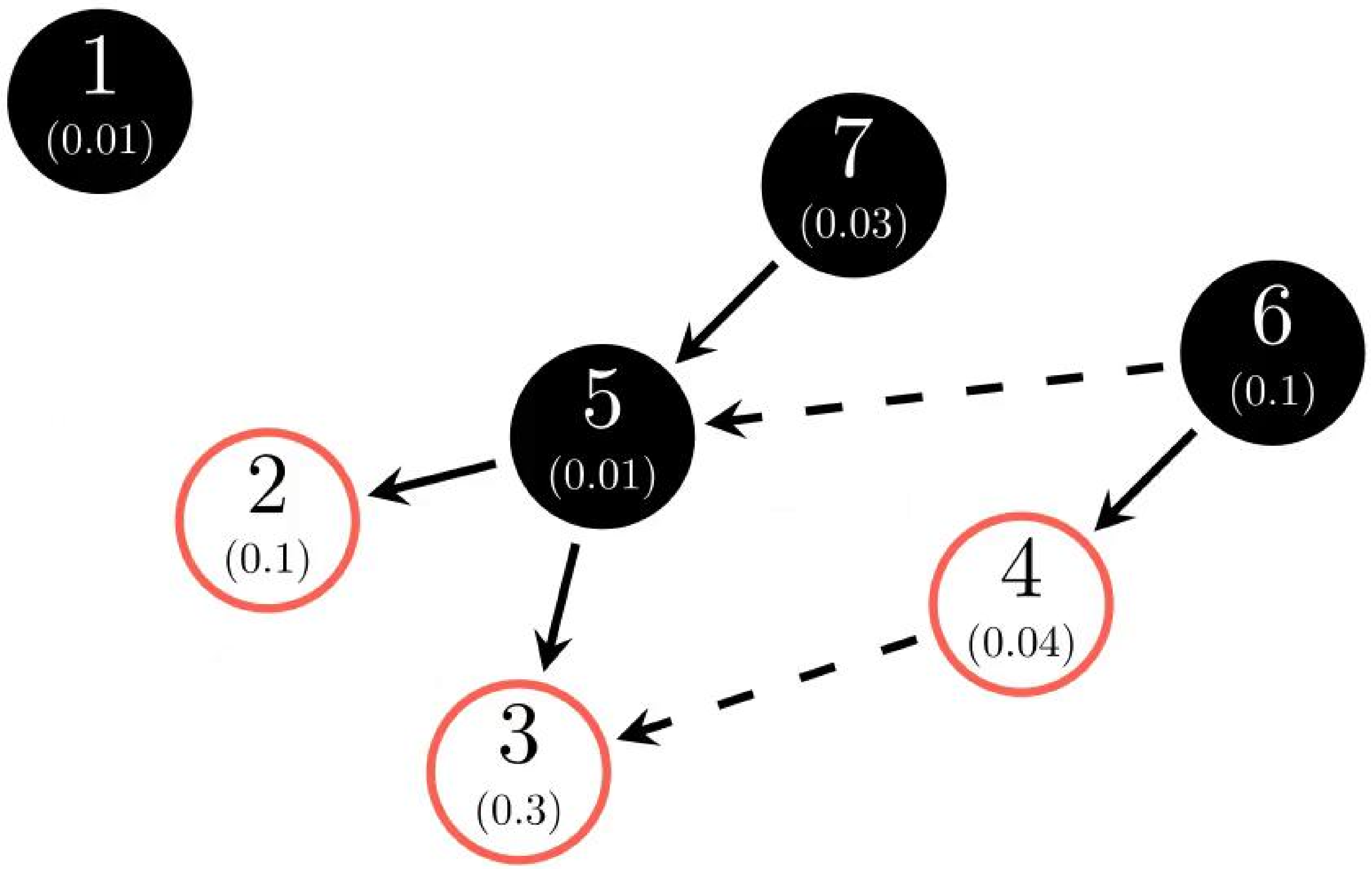
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \ldots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



Here: $\alpha = 0.05$.

Multiple testing procedure

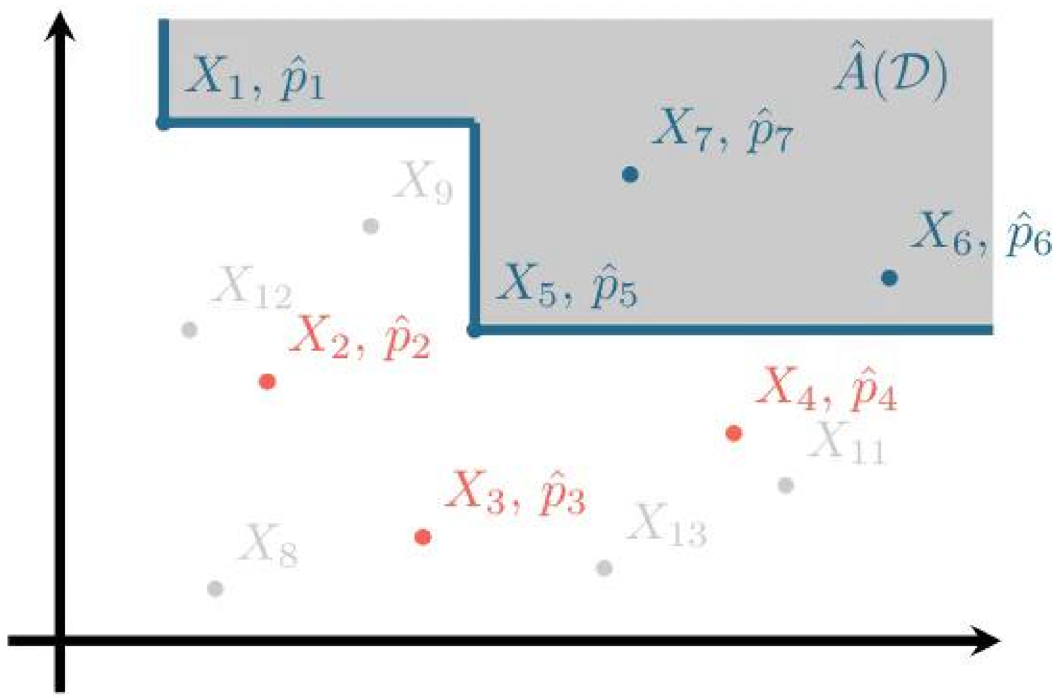
Key idea: logical relationships of hypotheses $H_0(X_i)$, $i \in \{1, \ldots, m\}$, induce a directed acyclic graph (DAG). We combine the sequential rejection principle (Goeman and Solari, 2010) with careful α-budget allocation to construct a procedure similar to Bretz et al. (2009).



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 $H_0(x_0) : \eta(x_0) < \tau$.

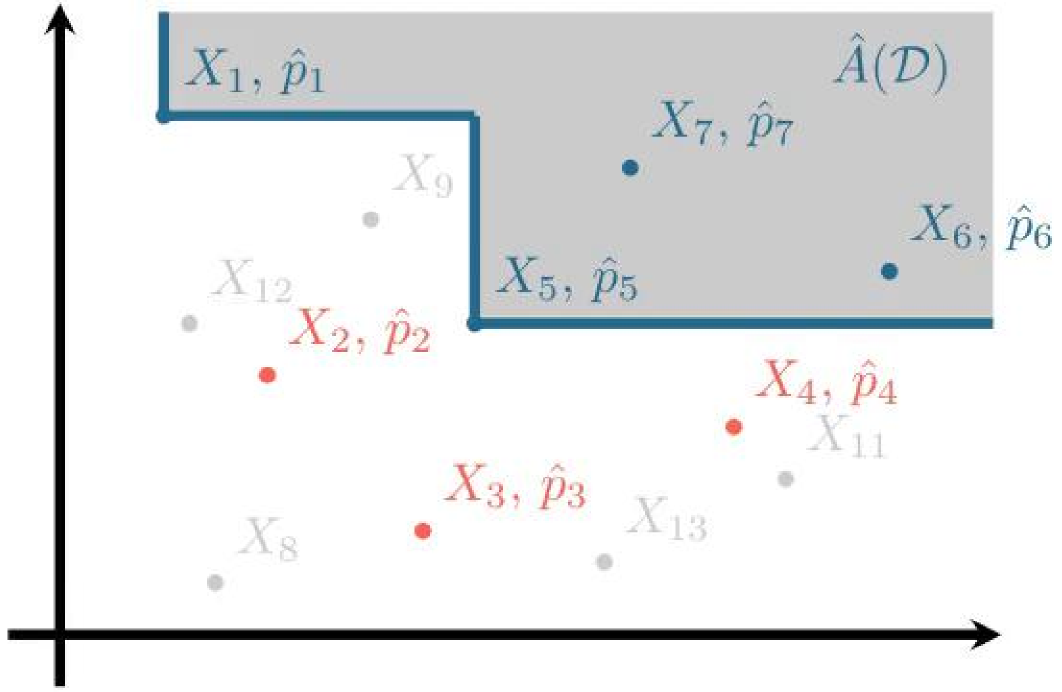


High-level strategy:

1. Subsample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Calculate *p-values* $\hat{p}_i$ for $H_0(X_i)$, $i \in \{1, \ldots, m\}$;
3. Apply a *multiple testing procedure* with FWER-control to reject $\mathcal{R}_\alpha \subseteq \{1, \ldots, m\}$;
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 $H_0(x_0) : \eta(x_0) < \tau$.



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 $H_0(X_i)$, $i \in \{1, \ldots, m\}$;
3. Apply a *multiple testing procedure* with FWER-control to reject $\mathcal{R}_\alpha \subseteq \{1, \ldots, m\}$;
4. Output $\hat{A} := \big\{x \in \mathbb{R}^d : X_\ell \preccurlyeq x \text{ for some } \ell \in \mathcal{R}_\alpha\big\}$.

Theoretical guarantees

Write $\hat{A}^{\mathrm{ISS}}$ for the resulting selected subgroup.

Theoretical guarantees

Write $\hat{A}^{\mathrm{ISS}}$ for the resulting selected subgroup.

Theorem. For any $n \geq 1$, $m \leq n$, $\alpha \in (0,1)$, $\sigma > 0$, we have:

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

Theoretical guarantees

Write $\hat{A}^{\mathrm{ISS}}$ for the resulting selected subgroup.

Theorem. For any $n \geq 1$, $m \leq n$, $\alpha \in (0,1)$, $\sigma > 0$, we have:

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

Note: this still holds if there are certain violations of monotonicity and this guarantee ensures we are robust against covariate-shifts.

Theoretical guarantees

Write $\hat{A}^{\mathrm{ISS}}$ for the resulting selected subgroup.

Theorem. For any $n \geq 1$, $m \leq n$, $\alpha \in (0,1)$, $\sigma > 0$, we have:

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

Note: this still holds if there are certain violations of monotonicity and this guarantee ensures we are robust against covariate-shifts.

Theorem. $\hat{A}^{\mathrm{ISS}}$ is minimax optimal (in terms of power) across a natural subclass of distributions in the sub-Gaussian setting.

Application I: Risk group estimation

Application I: Risk group estimation

Background: In a Phase 2 study, about 250 patients received a new drug with varying dose. Some patients faced adverse events (AE). Can we predict which patients are at risk of AEs?

Application of subgroup selection: we set $Y_i := \mathbb{1}\{\text{patient } i \text{ does not report AE}\}$, turning this into a classification setting.

$\hat{A}$ then only contains covariate configurations with probability of **not** observing an AE exceeding τ.

E.g. $\tau = 0.95$ and $\alpha = 0.05$.

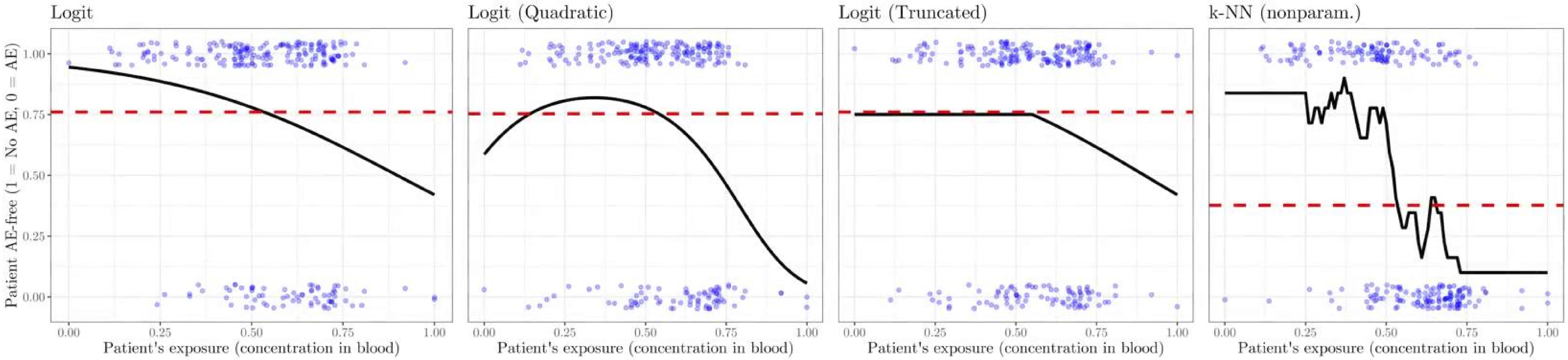
Decision process once we have computed $\hat{A}$ and observe a new patient with covariate values X:

- If $X \in \hat{A}$: patient can be expected to not face AEs since $\eta(X) \geq \tau$ (with probability $1 - \alpha$).
- If $X \notin \hat{A}$: patient might need further attention.

Application I: Simulation setup

Application I: Simulation setup

Using the R package *synthpop* (Nowok et al., 2016) we sample from the covariate distribution of the study. We then sample the responses according to the probabilities given by the following functions, which are also motivated by the real data:



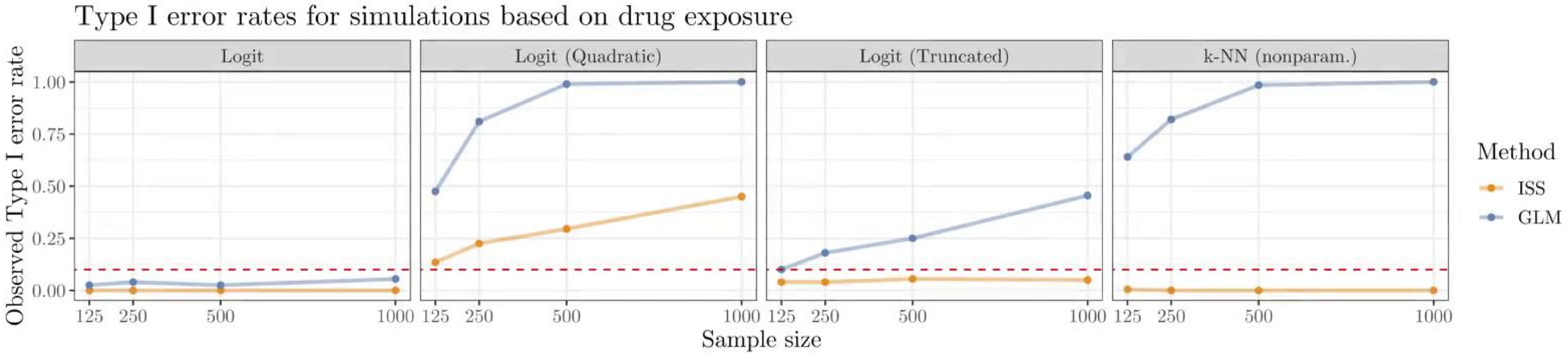
The threshold $\tau \in [0, 1]$ is chosen such that roughly 50% of patients fall into the subgroup defined by it.

Application I: Results

We compare ISS to the parametric method that assumes a GLM by Wan et al. (2024).

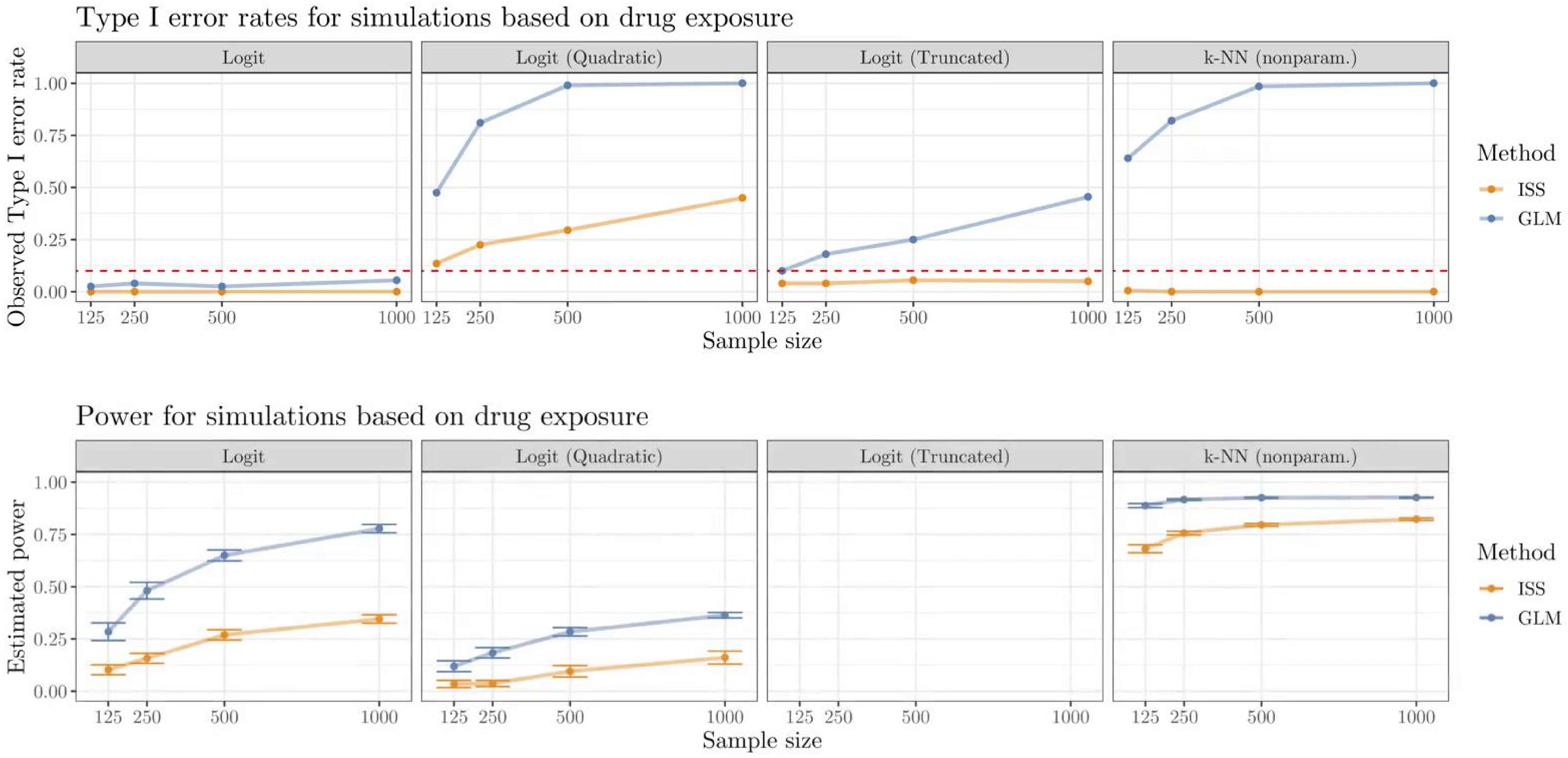
Application I: Results

We compare ISS to the parametric method that assumes a GLM by Wan et al. (2024).



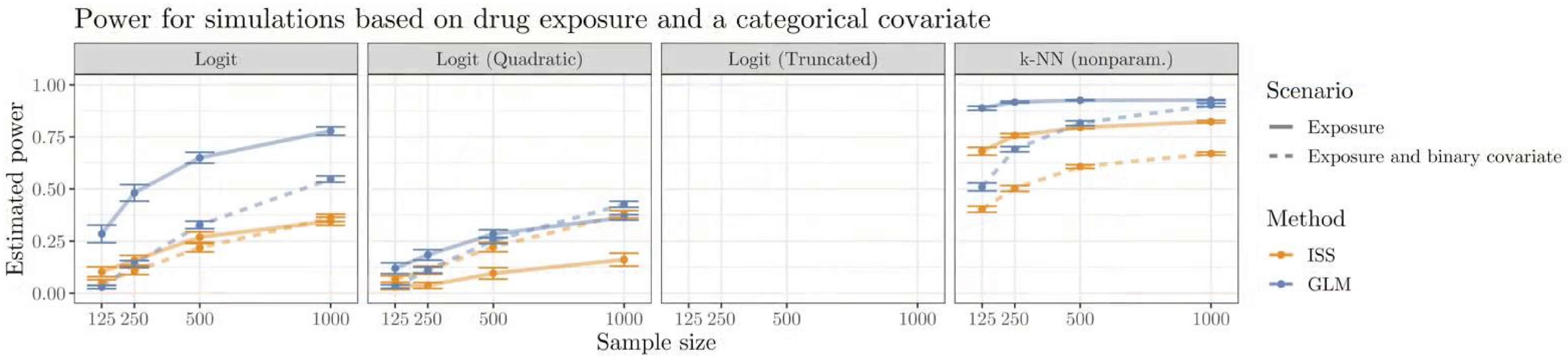
Application I: Results

We compare ISS to the parametric method that assumes a GLM by Wan et al. (2024).



Application I: Results

We compare ISS to the parametric method that assumes a GLM by Wan et al. (2024).



Application II: treatment effects

Application II: Treatment effects

Application II: Treatment effects

Background: We use the package `R` package *benchtm* (Sun et al., 2024) designed specifically to simulate data inspired by clinical trials with treatment effect heterogeneity.

Application II: Treatment effects

Background: We use the package `R` package *benchtm* (Sun et al., 2024) designed specifically to simulate data inspired by clinical trials with treatment effect heterogeneity.

Estimand: We let $Y(1)$ be the potential outcome under treatment and $Y(0)$ the potential outcome under control. Then, the conditional average treatment effect (CATE) is given by:

$$\mathrm{CATE}(x) := \mathbb{E}\big(Y(1) - Y(0)|X = x\big).$$

Application II: Treatment effects

Background: We use the package `R` package *benchtm* (Sun et al., 2024) designed specifically to simulate data inspired by clinical trials with treatment effect heterogeneity.

Estimand: We let $Y(1)$ be the potential outcome under treatment and $Y(0)$ the potential outcome under control. Then, the conditional average treatment effect (CATE) is given by:

$$\mathrm{CATE}(x) := \mathbb{E}\big(Y(1) - Y(0) | X = x\big).$$

Subgroup selection: Given an efficacy threshold τ, identify patients with CATE of at least τ.

Application II: Treatment effects

Background: We use the package `R` package *benchtm* (Sun et al., 2024) designed specifically to simulate data inspired by clinical trials with treatment effect heterogeneity.

Estimand: We let $Y(1)$ be the potential outcome under treatment and $Y(0)$ the potential outcome under control. Then, the conditional average treatment effect (CATE) is given by:

$$\mathrm{CATE}(x) := \mathbb{E}\big(Y(1) - Y(0) | X = x\big).$$

Subgroup selection: Given an efficacy threshold τ, identify patients with CATE of at least τ.

Difficulty: For patient i, we observe $Y_i(1)$ if they have been assigned to treatment ($T_i = 1$) or $Y_i(0)$ if they have been assigned to control ($T_i = 0$), but never both.

Application II: Treatment effects

Background: We use the package `R` package *benchtm* (Sun et al., 2024) designed specifically to simulate data inspired by clinical trials with treatment effect heterogeneity.

Estimand: We let $Y(1)$ be the potential outcome under treatment and $Y(0)$ the potential outcome under control. Then, the conditional average treatment effect (CATE) is given by:

$$\mathrm{CATE}(x) := \mathbb{E}\big(Y(1) - Y(0) | X = x\big).$$

Subgroup selection: Given an efficacy threshold τ, identify patients with CATE of at least τ.

Difficulty: For patient i, we observe $Y_i(1)$ if they have been assigned to treatment ($T_i = 1$) or $Y_i(0)$ if they have been assigned to control ($T_i = 0$), but never both.

Solution: We use the double-robust learning approach to generate pseudo-observations mimicking $Y_i(1) - Y_i(0)$ (Kennedy, 2023).

Application II: Simulation setup

Application II: Simulation setup

We simulate data using the *benchtm* `R` package (Sun et al., 2024).

Application II: Simulation setup

We simulate data using the *benchtm* `R` package (Sun et al., 2024).

This package provides realistic covariate distributions and responses distributed standard-normally around $g_0(x)$ in the control arm and around $g_0(x) + CATE(x)$ in the treatment arm:

Application II: Simulation setup

We simulate data using the *benchtm* `R` package (Sun et al., 2024).

This package provides realistic covariate distributions and responses distributed standard-normally around $g_0(x)$ in the control arm and around $g_0(x) + CATE(x)$ in the treatment arm:

Scenario	$g_0(x) := \mathbb{E}(Y \mid T = 0, X = x)$	CATE(x)	Label
(1)	$0.5\mathbb{1}\{x^{(1)} = \text{“}Y\text{”}\} + x^{(2)}$	$\beta_0 + \beta_1\Phi\big(20(x^{(2)} - 0.5)\big)$	$\Longrightarrow$ *GaussCDF*
(2)	$x^{(3)} - \mathbb{1}\{x^{(4)} = \text{“}N\text{”}\}$	$\beta_0 + \beta_1 x^{(3)}$	$\Longrightarrow$ *Linear*
(3)	$\mathbb{1}\{x^{(1)} = \text{“}N\text{”}\} - 0.5x^{(5)}$	$\beta_0 + \beta_1\mathbb{1}\{x^{(3)} > 0.25 \text{ and } x^{(1)} = \text{“}N\text{”}\}$	$\Longrightarrow$ *‘And’-condition*
(4)	$x^{(2)} - x^{(3)}$	$\beta_0 + \beta_1\mathbb{1}\{x^{(3)} > 0.3 \text{ or } x^{(6)} = \text{“}Y\text{”}\}$	$\Longrightarrow$ *‘Or’-condition*

$x = \big(x^{(1)}, \dots, x^{(6)}\big)$ denotes 6 different covariates and Φ the standard normal CDF. $g_0(x)$ is given up to constant factors and $\beta_0, \beta_1 \in \mathbb{R}$ (differing from row to row).

Application II: Simulation setup

We simulate data using the *benchtm* `R` package (Sun et al., 2024).

This package provides realistic covariate distributions and responses distributed standard-normally around $g_0(x)$ in the control arm and around $g_0(x) + CATE(x)$ in the treatment arm:

Scenario	$g_0(x) := \mathbb{E}(Y \mid T = 0, X = x)$	$\mathrm{CATE}(x)$	Label
(1)	$0.5\mathbb{1}\{x^{(1)} = \text{“}Y\text{”}\} + x^{(2)}$	$\beta_0 + \beta_1 \Phi\big(20(x^{(2)} - 0.5)\big)$	$\Longrightarrow$ *GaussCDF*
(2)	$x^{(3)} - \mathbb{1}\{x^{(4)} = \text{“}N\text{”}\}$	$\beta_0 + \beta_1 x^{(3)}$	$\Longrightarrow$ *Linear*
(3)	$\mathbb{1}\{x^{(1)} = \text{“}N\text{”}\} - 0.5x^{(5)}$	$\beta_0 + \beta_1 \mathbb{1}\{x^{(3)} > 0.25 \text{ and } x^{(1)} = \text{“}N\text{”}\}$	$\Longrightarrow$ *‘And’-condition*
(4)	$x^{(2)} - x^{(3)}$	$\beta_0 + \beta_1 \mathbb{1}\{x^{(3)} > 0.3 \text{ or } x^{(6)} = \text{“}Y\text{”}\}$	$\Longrightarrow$ *‘Or’-condition*

$x = \big(x^{(1)}, \ldots, x^{(6)}\big)$ denotes 6 different covariates and Φ the standard normal CDF. $g_0(x)$ is given up to constant factors and $\beta_0, \beta_1 \in \mathbb{R}$ (differing from row to row).

We aim to select the subgroup based on predictive covariates and consider cases...

Application II: Simulation setup

We simulate data using the *benchtm* `R` package (Sun et al., 2024).

This package provides realistic covariate distributions and responses distributed standard-normally around $g_0(x)$ in the control arm and around $g_0(x) + CATE(x)$ in the treatment arm:

Scenario	$g_0(x) := \mathbb{E}(Y \vert T = 0, X = x)$	$\mathrm{CATE}(x)$	Label
(1)	$0.5\mathbb{1}\{x^{(1)} = \text{“}Y\text{”}\} + x^{(2)}$	$\beta_0 + \beta_1 \Phi\big(20(x^{(2)} - 0.5)\big)$	$\Longrightarrow$ *GaussCDF*
(2)	$x^{(3)} - \mathbb{1}\{x^{(4)} = \text{“}N\text{”}\}$	$\beta_0 + \beta_1 x^{(3)}$	$\Longrightarrow$ *Linear*
(3)	$\mathbb{1}\{x^{(1)} = \text{“}N\text{”}\} - 0.5x^{(5)}$	$\beta_0 + \beta_1 \mathbb{1}\{x^{(3)} > 0.25 \text{ and } x^{(1)} = \text{“}N\text{”}\}$	$\Longrightarrow$ *‘And’-condition*
(4)	$x^{(2)} - x^{(3)}$	$\beta_0 + \beta_1 \mathbb{1}\{x^{(3)} > 0.3 \text{ or } x^{(6)} = \text{“}Y\text{”}\}$	$\Longrightarrow$ *‘Or’-condition*

$x = \big(x^{(1)}, \ldots, x^{(6)}\big)$ denotes 6 different covariates and Φ the standard normal CDF. $g_0(x)$ is given up to constant factors and $\beta_0, \beta_1 \in \mathbb{R}$ (differing from row to row).

We aim to select the subgroup based on predictive covariates and consider cases…

- …where $Y_i(1) - Y_i(0)$, $i \in \{1, \ldots, n\}$, are observed (as reference point).

Application II: Simulation setup

We simulate data using the *benchtm* `R` package (Sun et al., 2024).

This package provides realistic covariate distributions and responses distributed standard-normally around $g_0(x)$ in the control arm and around $g_0(x) + CATE(x)$ in the treatment arm:

Scenario	$g_0(x) := \mathbb{E}(Y \mid T = 0, X = x)$	CATE(x)	Label
(1)	$0.5\mathbb{1}\{x^{(1)} = \text{"}Y\text{"}\} + x^{(2)}$	$\beta_0 + \beta_1\Phi(20(x^{(2)} - 0.5))$	$\Longrightarrow$ *GaussCDF*
(2)	$x^{(3)} - \mathbb{1}\{x^{(4)} = \text{"}N\text{"}\}$	$\beta_0 + \beta_1 x^{(3)}$	$\Longrightarrow$ *Linear*
(3)	$\mathbb{1}\{x^{(1)} = \text{"}N\text{"}\} - 0.5x^{(5)}$	$\beta_0 + \beta_1\mathbb{1}\{x^{(3)} > 0.25 \text{ and } x^{(1)} = \text{"}N\text{"}\}$	$\Longrightarrow$ *'And'-condition*
(4)	$x^{(2)} - x^{(3)}$	$\beta_0 + \beta_1\mathbb{1}\{x^{(3)} > 0.3 \text{ or } x^{(6)} = \text{"}Y\text{"}\}$	$\Longrightarrow$ *'Or'-condition*

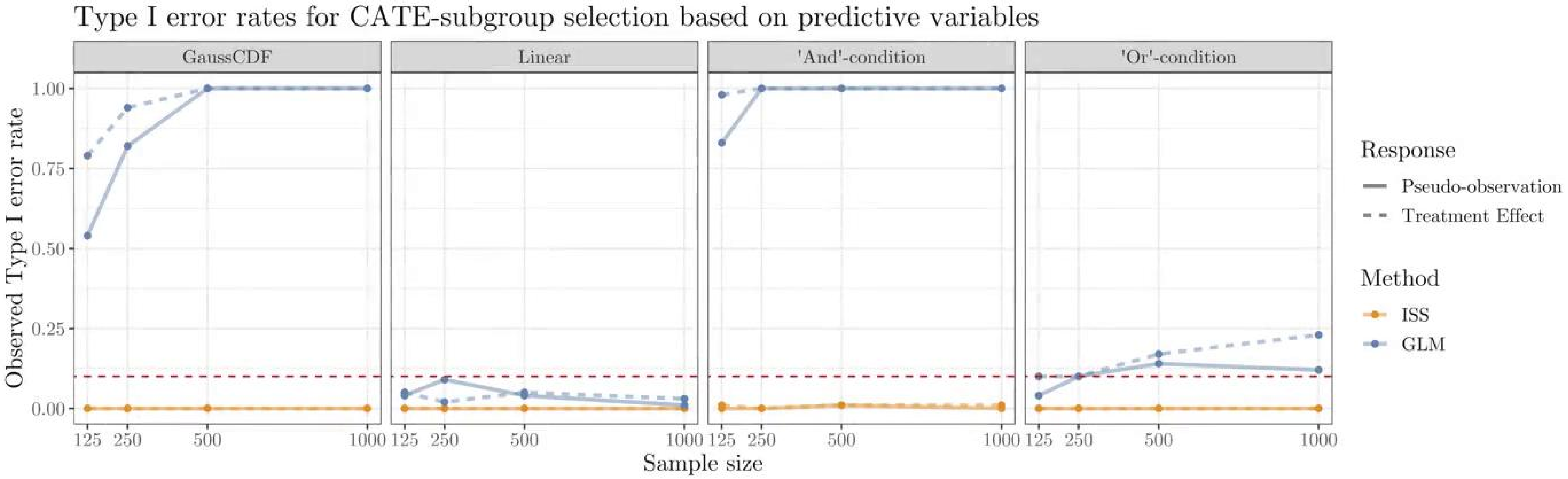
$x = \left(x^{(1)}, \ldots, x^{(6)}\right)$ denotes 6 different covariates and Φ the standard normal CDF. $g_0(x)$ is given up to constant factors and $\beta_0, \beta_1 \in \mathbb{R}$ (differing from row to row).

We aim to select the subgroup based on predictive covariates and consider cases...

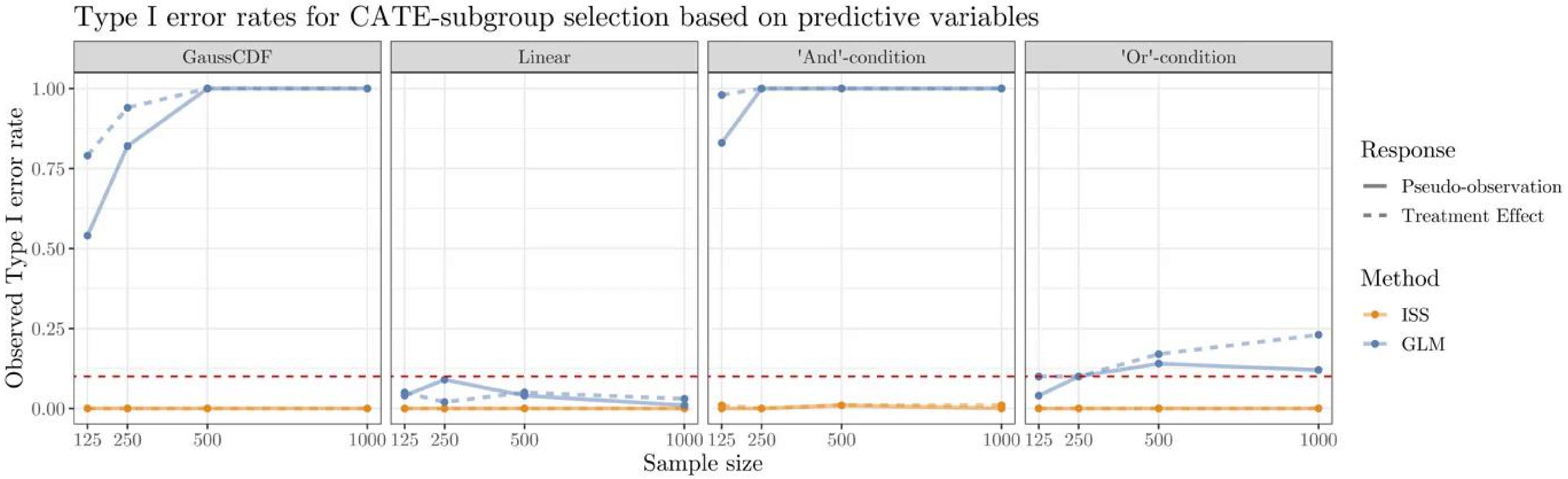
- ...where $Y_i(1) - Y_i(0)$, $i \in \{1, \ldots, n\}$, are observed (as reference point).
- ...where we have to use pseudo-observations.

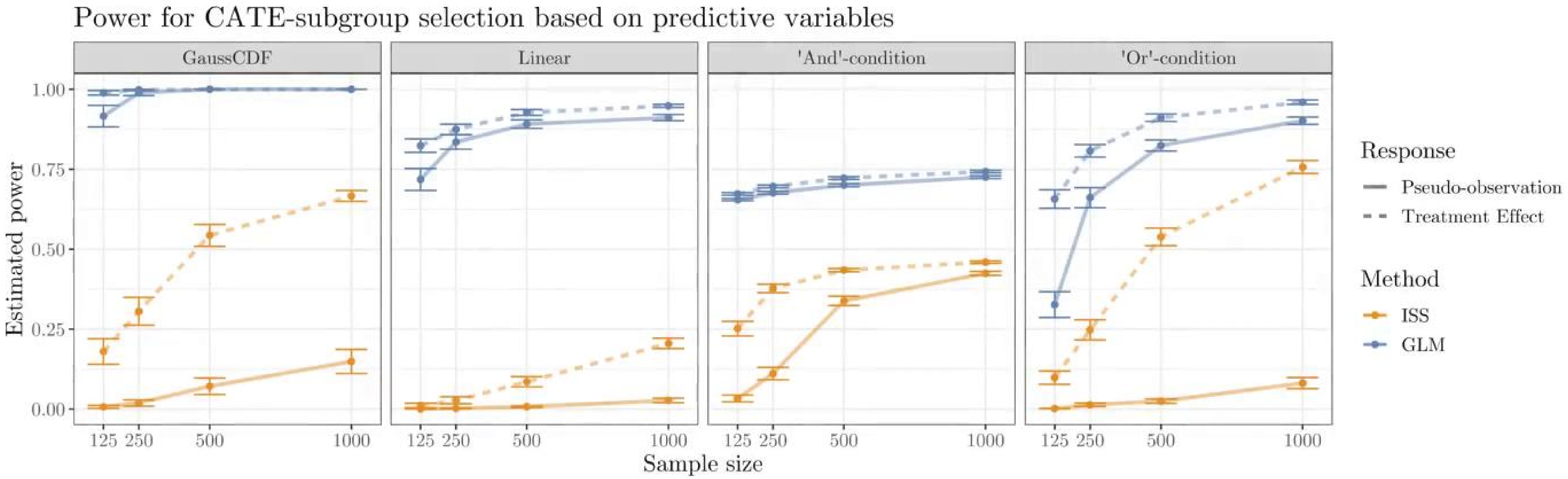
Application II: Results

Application II: Results



Application II: Results





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

Bretz, F., Maurer, W., Brannath, W., and Posch, M. (2009). A graphical approach to sequentially rejective multiple test procedures. *Statistics in Medicine*, 28:586-604.

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.

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

Kennedy, E. H. (2023). Towards optimal doubly robust estimation of heterogeneous causal effects. *Electronic Journal of Statistics*, 17(2):3008-3049.

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.

Nowok, B., Raab, G.M., and Dibben, C. (2016), synthpop: Bespoke creation of synthetic data in R. *Journal of Statistical Software*, 74:1-26.

Sun, S., Sechidis, K., Chen, Y., Lu, J., Ma, C., Mirshani, A., Ohlssen, D., Vandemeulebroecke, M., and Bornkamp, B. (2024). Comparing algorithms for characterizing treatment effect heterogeneity in randomized trials. *Biometrical Journal*, 66, 2100337.

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

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 $t \in (0,1)$.

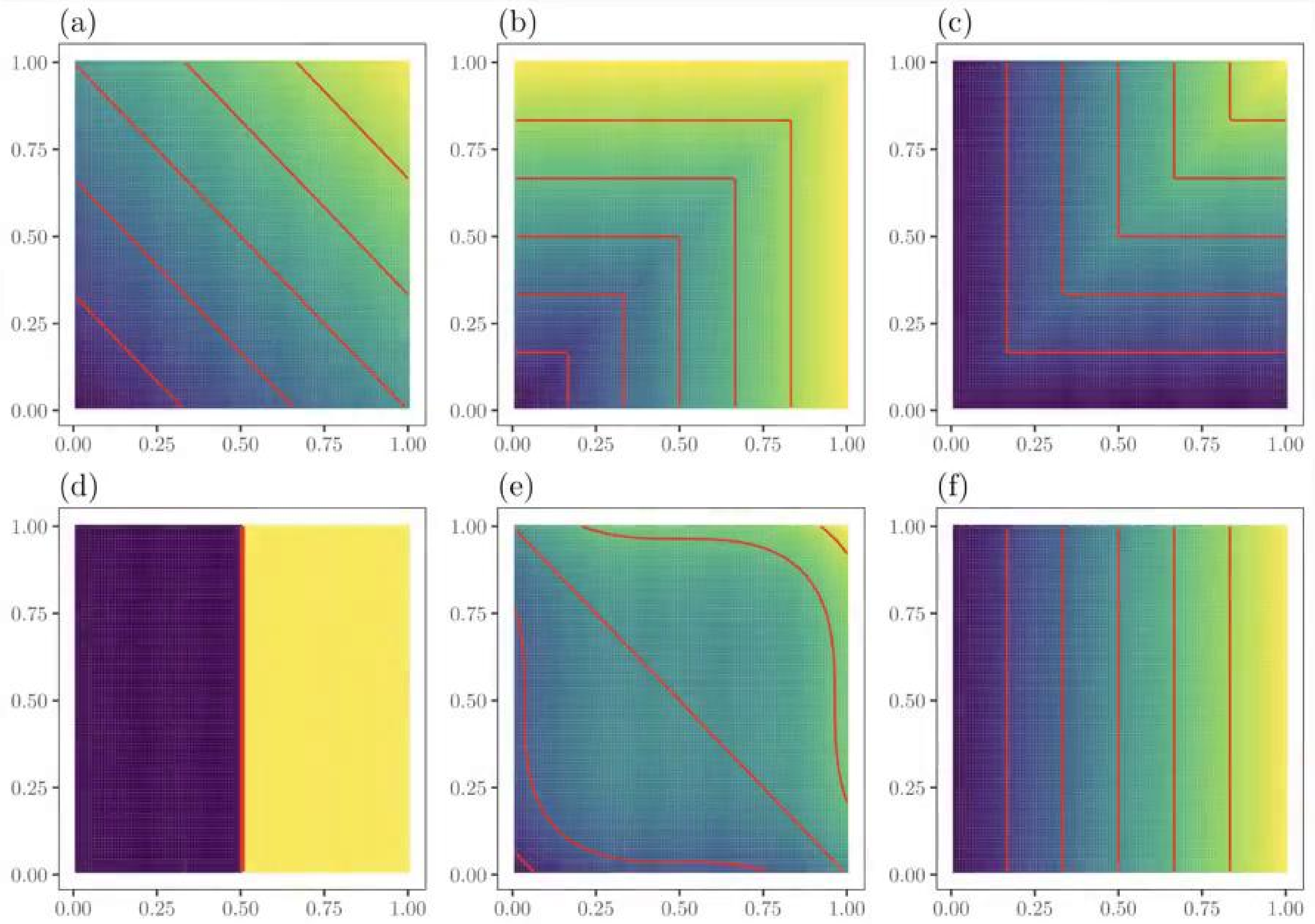
Simulations

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 $[0,1]^d$:

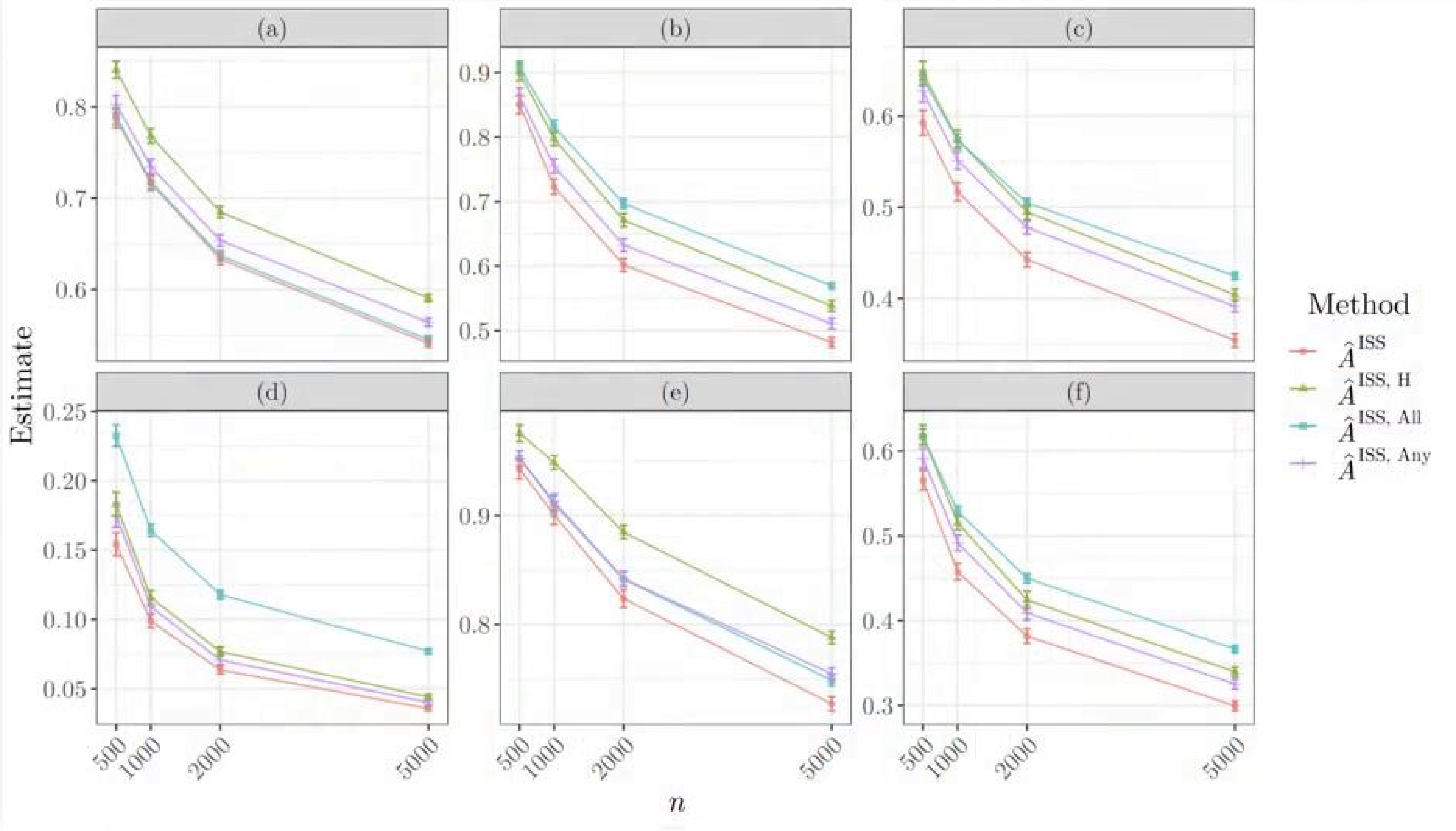
Label	Function f	τ	$\gamma(P)$
(a)	$\sum_{j=1}^d x^{(j)}$	$1/2$	1
(b)	$\max_{1\le j\le d} x^{(j)}$	$1/2^{1/d}$	1
(c)	$\min_{1\le j\le d} x^{(j)}$	$1-1/2^{1/d}$	1
(d)	$\mathbb{1}_{(0.5,1]}\big(x^{(1)}\big)$	$1/2$	0
(e)	$\sum_{j=1}^d \big(x^{(j)}-0.5\big)^3$	$1/2$	3
(f)	$x^{(1)}$	$1/2$	1



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)	$\max_{1\le j\le d} x^{(j)}$	$1/2^{1/d}$	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



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