Isotonic subgroup selection

CRiSM Seminar series in Warwick

May 2023

Henry Reeve

University of Bristol

Collaborators

Manuel Mueller

Timothy Cannings

Richard Samworth

Subgroup selection

We shall be consider in subgroup selection in the regression setting.

Subgroup selection

We shall be consider in subgroup selection in the regression setting.

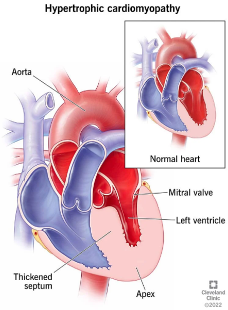
The goal is to identify a subset of the covariate domain on which the regression function satisfies a particular property of interest.

Subgroup selection

We shall be consider in subgroup selection in the regression setting.

The goal is to identify a subset of the covariate domain on which the regression function satisfies a particular property of interest.

Consider patients with hypertrophic cardiomyopathy (HC).



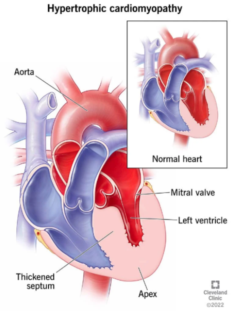
Subgroup selection

We shall be consider in subgroup selection in the regression setting.

The goal is to identify a subset of the covariate domain on which the regression function satisfies a particular property of interest.

Consider patients with hypertrophic cardiomyopathy (HC).

Patients at high risk of sudden cardiac death (SCD) require an implantable cardioverter defibrillator (O'Mahony et al., 2014).



Subgroup selection

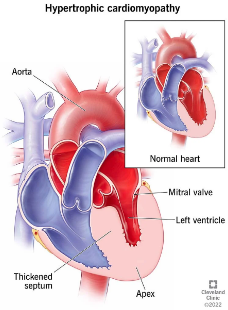
We shall be consider in subgroup selection in the regression setting.

The goal is to identify a subset of the covariate domain on which the regression function satisfies a particular property of interest.

Consider patients with hypertrophic cardiomyopathy (HC).

Patients at high risk of sudden cardiac death (SCD) require an implantable cardioverter defibrillator (O'Mahony et al., 2014).

Hence, we would like to identify the subset of HC patients which are low risk of sudden cardiac death.



Subgroup selection

The goal is to identify a subset of the covariate domain on which the regression function satisfies a particular property of interest.

We would like to identify the subset of hypertrophic cardiomyopathy patients which are low risk of sudden cardiac death.

Key features:

1. Subset selection is a post-selection inference problem since we seek inferential guarantees whilst using our data to select the region of interest.
2. Applications typically involve a strong form of asymmetry between the different types of errors.
3. Our inferential guarantees should hold for all individuals within the selected subgroup.

Statistical setting

Suppose we have a distribution P on covariate-response pairs (X, Y) in $\mathbb{R}^d \times \mathbb{R}$.

Statistical setting

Suppose we have a distribution P on covariate-response pairs (X, Y) in $\mathbb{R}^d \times \mathbb{R}$.

Let $\eta \equiv \eta_P : \mathbb{R}^d \to \mathbb{R}$ be the regression function defined by $\eta(x) := \mathbb{E}(Y|X = x)$ for $x \in \mathbb{R}^d$.

Statistical setting

Suppose we have a distribution P on covariate-response pairs (X, Y) in $\mathbb{R}^d \times \mathbb{R}$.

Let $\eta \equiv \eta_P : \mathbb{R}^d \to \mathbb{R}$ be the regression function defined by $\eta(x) := \mathbb{E}(Y|X = x)$ for $x \in \mathbb{R}^d$.

Let $\mu \equiv \mu_P$ denote the marginal distribution of the covariate X in $\mathbb{R}^d$.

Statistical setting

Suppose we have a distribution P on covariate-response pairs (X, Y) in $\mathbb{R}^d \times \mathbb{R}$.

Let $\eta \equiv \eta_P : \mathbb{R}^d \to \mathbb{R}$ be the regression function defined by $\eta(x) := \mathbb{E}(Y|X = x)$ for $x \in \mathbb{R}^d$.

Let $\mu \equiv \mu_P$ denote the marginal distribution of the covariate X in $\mathbb{R}^d$.

We would like to select a subgroup $A \subseteq \mathbb{R}^d$ such that η is above a user-specified threshold $\tau \in \mathbb{R}$ on A.

Statistical setting

Suppose we have a distribution P on covariate-response pairs (X, Y) in $\mathbb{R}^d \times \mathbb{R}$.

Let $\eta \equiv \eta_P : \mathbb{R}^d \to \mathbb{R}$ be the regression function defined by $\eta(x) := \mathbb{E}(Y|X = x)$ for $x \in \mathbb{R}^d$.

Let $\mu \equiv \mu_P$ denote the marginal distribution of the covariate X in $\mathbb{R}^d$.

We would like to select a subgroup $A \subseteq \mathbb{R}^d$ such that η is above a user-specified threshold $\tau \in \mathbb{R}$ on A.

Example: Our selected subgroup of HC patients should contain only those for whom the conditional probability of SCD is below $\tau = 1\%$.

Statistical setting

Suppose we have a distribution P on covariate-response pairs (X, Y) in $\mathbb{R}^d \times \mathbb{R}$.

Let $\eta \equiv \eta_P : \mathbb{R}^d \to \mathbb{R}$ be the regression function defined by $\eta(x) := \mathbb{E}(Y|X = x)$ for $x \in \mathbb{R}^d$.

Let $\mu \equiv \mu_P$ denote the marginal distribution of the covariate X in $\mathbb{R}^d$.

We would like to select a subgroup $A \subseteq \mathbb{R}^d$ such that η is above a user-specified threshold $\tau \in \mathbb{R}$ on A.

Example: Our selected subgroup of HC patients should contain only those for whom the conditional probability of SCD is below $\tau = 1\%$.

Hence, we are interested in subsets of the τ-super level set

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

Statistical setting

We would like to select a subgroup which is contained within

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

The practitioner has access to a sample $\mathcal{D} = ((X_1, Y_1), \ldots, (X_n, Y_n))$, consisting of n independent copies of $(X, Y) \sim P$.

Statistical setting

We would like to select a subgroup which is contained within

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

The practitioner has access to a sample $\mathcal{D} = ((X_1, Y_1), \ldots, (X_n, Y_n))$, consisting of n independent copies of $(X, Y) \sim P$.

Our algorithm returns a data-dependent subgroup $\hat{A} \equiv \hat{A}(\mathcal{D})$, which is a random subset of $\mathbb{R}^d$.

Statistical setting

We would like to select a subgroup which is contained within

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

The practitioner has access to a sample $\mathcal{D} = ((X_1, Y_1), \ldots, (X_n, Y_n))$, consisting of n independent copies of $(X, Y) \sim P$.

Our algorithm returns a data-dependent subgroup $\hat{A} \equiv \hat{A}(\mathcal{D})$, which is a random subset of $\mathbb{R}^d$.

Our data-dependent subgroup should satisfy $\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)$, with high-probability.

Type 1 error control for subgroup selection

Given a sample $\mathcal{D} = ((X_1, Y_1), \ldots, (X_n, Y_n)) \overset{\text{i.i.d.}}{\sim} P$, we aim to ensure that $\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)$, with high-probability, where $\mathcal{X}_\tau(\eta) := \{x \in \mathbb{R}^d : \eta(x) \geq \tau\}$.

Type 1 error control

Given a nominal level $\alpha \in (0, 1)$, we shall say that the data-dependent subgroup $\hat{A}$ *controls Type 1 error* at the level α, if

$$\mathbb{P}\left(\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\right) \geq 1 - \alpha.$$

Type 1 error control for subgroup selection

Given a sample $\mathcal{D} = ((X_1, Y_1), \ldots, (X_n, Y_n)) \stackrel{\text{i.i.d.}}{\sim} P$, we aim to ensure that $\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)$, with high-probability, where $\mathcal{X}_\tau(\eta) := \{x \in \mathbb{R}^d : \eta(x) \geq \tau\}$.

Type 1 error control

Given a nominal level $\alpha \in (0, 1)$, we shall say that the data-dependent subgroup $\hat{A}$ *controls Type 1 error* at the level α, if

$$\mathbb{P}\left(\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\right) \geq 1 - \alpha.$$

Example: Suppose we wish to select a subgroup consisting of low-risk HC patients for whom the conditional probability of SCD is below $\tau = 1\%$.

We aim to control the probability of a Type 1 error in which our selected subgroup contains high-risk patients.

Type 1 error and power for subgroup selection

Type 1 error control

Given a nominal level $\alpha \in (0, 1)$, we shall say that the data-dependent subgroup $\hat{A}$ *controls Type 1 error* at the level α, if

$$\mathbb{P}\left(\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\right) \geq 1 - \alpha.$$

Subject to this Type 1 error constraint, we also wish to maximise the proportion of $\mathcal{X}_\tau(\eta)$ contained within selected subgroup $\hat{A}$.

Hence, subject to the Type 1 error constraint, we aim to minimise the regret

$$R_\tau(\hat{A}) := \mathbb{E}_P\big\{\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}\big)\big\}.$$

Type 1 error and power for subgroup selection

Type 1 error control

Given a nominal level $\alpha \in (0,1)$, we shall say that the data-dependent subgroup $\hat{A}$ *controls Type 1 error* at the level α, if

$$\mathbb{P}\left(\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\right) \geq 1 - \alpha.$$

Subject to this Type 1 error constraint, we also wish to maximise the proportion of $\mathcal{X}_\tau(\eta)$ contained within selected subgroup $\hat{A}$.

Hence, subject to the Type 1 error constraint, we aim to minimise the regret

$$R_\tau(\hat{A}) := \mathbb{E}_P\big\{\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}\big)\big\}.$$

Minimising the regret $R_\tau(\hat{A})$ corresponds to maximising the power.

Type 1 error and power for subgroup selection

We seek data-dependent subgroups $\hat{A}$ which control Type 1 error at the nominal level α, i.e. $\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)$ holds with probability at least $1 - \alpha$.

Subject to this Type 1 error constraint, we also aim to minimise the regret

$$R_\tau(\hat{A}) := \mathbb{E}_P\bigl\{\mu\bigl(\mathcal{X}_\tau(\eta) \setminus \hat{A}\bigr)\bigr\}.$$

Example:

Suppose we wish to select a subgroup consisting of HC patients for whom the conditional probability of SCD is below 5%.

On the one hand, we seek to control the probability of a Type 1 error in which high-risk patients are incorrectly assigned to the low-risk subgroup.

We also wish to minimise the proportion of low-risk patients excluded from the subgroup and unnecessarily fitted with a cardioverter defibrillator.

Distributional classes

Subgroup selection has been studied in a variety of distributional regimes.

Ballarini et al. (2018) and Wan et al. (2022) consider subgroup selection in settings where the regression function η is assumed to be linear.

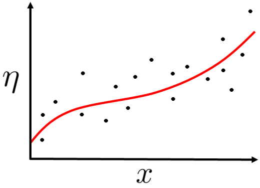
Previously, we considered a non-parametric setting in which η is assumed to belong to a Hölder class (Reeve et al. 2021).

Distributional classes

Subgroup selection has been studied in a variety of distributional regimes.

Ballarini et al. (2018) and Wan et al. (2022) consider subgroup selection in settings where the regression function η is assumed to be linear.

Previously, we considered a non-parametric setting in which η is assumed to belong to a Hölder class (Reeve et al. 2021).



Here we shall consider a setting in which η is an isotonic regression function.

Isotonic regression functions

In many applications, the regression function η is monotonic with respect to individual covariates.

Isotonic regression functions

In many applications, the regression function η is monotonic with respect to individual covariates.

Example: Age, family history of SCD, maximal left ventricular wall thickness, fractional shortening, left atrial diameter, maximal LV outflow tract gradient etc. are risk factors for SCD (O'Mahony et al., 2014).

Isotonic regression functions

In many applications, the regression function η is monotonic with respect to individual covariates.

Example: Age, family history of SCD, maximal left ventricular wall thickness, fractional shortening, left atrial diameter, maximal LV outflow tract gradient etc. are risk factors for SCD (O'Mahony et al., 2014).

More formally, we impose a partial order $\preccurlyeq$ on $\mathbb{R}^d$ by $x_0 \preccurlyeq x_1$ where $x_0 = (x_{0,j})_{j=1}^d, x_1 = (x_{1,j})_{j=1}^d \in \mathbb{R}^d$ if $x_{0,j} \leq x_{1,j}$ for each $j \in \{1, \ldots, d\}$.

Isotonic regression functions

In many applications, the regression function η is monotonic with respect to individual covariates.

Example: Age, family history of SCD, maximal left ventricular wall thickness, fractional shortening, left atrial diameter, maximal LV outflow tract gradient etc. are risk factors for SCD (O'Mahony et al., 2014).

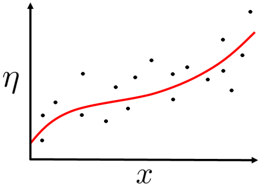
More formally, we impose a partial order $\preccurlyeq$ on $\mathbb{R}^d$ by $x_0 \preccurlyeq x_1$ where $x_0 = (x_{0,j})_{j=1}^d, x_1 = (x_{1,j})_{j=1}^d \in \mathbb{R}^d$ if $x_{0,j} \leq x_{1,j}$ for each $j \in \{1, \ldots, d\}$.

Isotonic regression functions

We say $\eta : \mathbb{R}^d \to \mathbb{R}$ is isotonic if $\eta(x_0) \leq \eta(x_1)$ for $x_0, x_1 \in \mathbb{R}^d$ with $x_0 \preccurlyeq x_1$.

Isotonic regression functions

We say $\eta : \mathbb{R}^d \to \mathbb{R}$ is isotonic if $\eta(x_0) \leq \eta(x_1)$ for $x_0, x_1 \in \mathbb{R}^d$ with $x_0 \preccurlyeq x_1$.



We write $\mathcal{P}_{\mathrm{M}}(d, \sigma)$ for the class of all distributions P on pairs (X, Y) with isotonic regression function η_P and σ^2-sub-Gaussian noise $\{Y - \eta_P(X)\}|X$.

Overall strategy

Isotonic regression functions

We say $\eta : \mathbb{R}^d \to \mathbb{R}$ is isotonic if $\eta(x_0) \leq \eta(x_1)$ for $x_0, x_1 \in \mathbb{R}^d$ with $x_0 \preccurlyeq x_1$.

We write $\mathcal{P}_{\mathrm{M}}(d, \sigma)$ for the class of all distributions P on pairs (X, Y) with isotonic regression function η_P and σ^2-sub-Gaussian noise $\{Y - \eta_P(X)\}|X$.

Our goal is to construct data-dependent subgroups which:

1. Control type 1 error at the nominal level

$$\mathbb{P}_P\big(\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\big) \geq 1 - \alpha,$$

   for all $P \in \mathcal{P}_{\mathrm{M}}(d, \sigma)$.

2. Minimise the regret $R_\tau(\hat{A}) := \mathbb{E}_P\big\{\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}\big)\big\}$, subject to 1.

Isotonic regression functions

We say $\eta : \mathbb{R}^d \to \mathbb{R}$ is isotonic if $\eta(x_0) \leq \eta(x_1)$ for $x_0, x_1 \in \mathbb{R}^d$ with $x_0 \preccurlyeq x_1$.

We write $\mathcal{P}_{\mathrm{M}}(d, \sigma)$ for the class of all distributions P on pairs (X, Y) with isotonic regression function η_P and σ^2-sub-Gaussian noise $\{Y - \eta_P(X)\}|X$.

Our objective is to choose $\hat{A} \equiv \hat{A}(\mathcal{D})$ which minimises regret $R_\tau(\hat{A})$, subject to

Type 1 error guarantee: $\inf_{P \in \mathcal{P}_{\mathrm{M}}(d,\sigma)} \mathbb{P}_P\big(\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\big) \geq 1 - \alpha$

To each $x \in \mathbb{R}^d$, we associate a null hypothesis
$H_0(x) := \{P \in \mathcal{P}_{\mathrm{M}}(d, \sigma) : \eta_P(x_0) < \tau\}$

Overall strategy

Isotonic regression functions

We say $\eta : \mathbb{R}^d \to \mathbb{R}$ is isotonic if $\eta(x_0) \leq \eta(x_1)$ for $x_0, x_1 \in \mathbb{R}^d$ with $x_0 \preccurlyeq x_1$.

We write $\mathcal{P}_{\mathrm{M}}(d, \sigma)$ for the class of all distributions P on pairs (X, Y) with isotonic regression function η_P and σ^2-sub-Gaussian noise $\{Y - \eta_P(X)\}|X$.

Our objective is to choose $\hat{A} \equiv \hat{A}(\mathcal{D})$ which minimises regret $R_\tau(\hat{A})$, subject to

Type 1 error guarantee: $\inf_{P \in \mathcal{P}_{\mathrm{M}}(d,\sigma)} \mathbb{P}_P\big(\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\big) \geq 1 - \alpha$

To each $x \in \mathbb{R}^d$, we associate a null hypothesis
$H_0(x) := \{P \in \mathcal{P}_{\mathrm{M}}(d, \sigma) : \eta_P(x_0) < \tau\} = \{P \in \mathcal{P}_{\mathrm{M}}(d, \sigma) : x_0 \notin \mathcal{X}_\tau(\eta_P)\}$.

Overall strategy

Isotonic regression functions

We say $\eta : \mathbb{R}^d \to \mathbb{R}$ is isotonic if $\eta(x_0) \leq \eta(x_1)$ for $x_0, x_1 \in \mathbb{R}^d$ with $x_0 \preccurlyeq x_1$.

We write $\mathcal{P}_{\mathrm{M}}(d, \sigma)$ for the class of all distributions P on pairs (X, Y) with isotonic regression function η_P and σ^2-sub-Gaussian noise $\{Y - \eta_P(X)\}|X$.

Our objective is to choose $\hat{A} \equiv \hat{A}(\mathcal{D})$ which minimises regret $R_\tau(\hat{A})$, subject to

Type 1 error guarantee: $\inf_{P \in \mathcal{P}_{\mathrm{M}}(d,\sigma)} \mathbb{P}_P\big(\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\big) \geq 1 - \alpha$

To each $x \in \mathbb{R}^d$, we associate a null hypothesis
$H_0(x) := \{P \in \mathcal{P}_{\mathrm{M}}(d, \sigma) : \eta_P(x_0) < \tau\} = \{P \in \mathcal{P}_{\mathrm{M}}(d, \sigma) : x_0 \notin \mathcal{X}_\tau(\eta_P)\}$.

Logical structure: If $x_0 \preccurlyeq x_1$ then $H_0(x_1) \subseteq H_0(x_0)$.

Overall strategy

Our objective is to choose $\hat{A} \equiv \hat{A}(\mathcal{D})$ which minimises regret $R_\tau(\hat{A})$, subject to
Type 1 error guarantee: $\inf_{P \in \mathcal{P}_{\mathrm{M}}(d,\sigma)} \mathbb{P}_P\big(\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\big) \geq 1 - \alpha.$

To each $x \in \mathbb{R}^d$, we associate a null $H_0(x) := \{P \in \mathcal{P}_{\mathrm{M}}(d,\sigma) : \eta_P(x_0) < \tau\}$.

High-level strategy: Given a sample $\mathcal{D} = ((X_1, Y_1), \ldots, (X_n, Y_n)) \overset{\text{i.i.d.}}{\sim} P$,

Overall strategy

Our objective is to choose $\hat{A} \equiv \hat{A}(\mathcal{D})$ which minimises regret $R_\tau(\hat{A})$, subject to
Type 1 error guarantee: $\inf_{P \in \mathcal{P}_{\mathrm{M}}(d,\sigma)} \mathbb{P}_P\big(\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\big) \geq 1 - \alpha.$

To each $x \in \mathbb{R}^d$, we associate a null $H_0(x) := \{P \in \mathcal{P}_{\mathrm{M}}(d,\sigma) : \eta_P(x_0) < \tau\}$.

High-level strategy: Given a sample $\mathcal{D} = ((X_1, Y_1), \ldots, (X_n, Y_n)) \overset{\text{i.i.d.}}{\sim} P$,

1. Sub-sample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;

Overall strategy

Our objective is to choose $\hat{A} \equiv \hat{A}(\mathcal{D})$ which minimises regret $R_\tau(\hat{A})$, subject to Type 1 error guarantee: $\inf_{P \in \mathcal{P}_{\mathrm{M}}(d,\sigma)} \mathbb{P}_P\big(\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\big) \geq 1 - \alpha.$

To each $x \in \mathbb{R}^d$, we associate a null $H_0(x) := \{P \in \mathcal{P}_{\mathrm{M}}(d,\sigma) : \eta_P(x_0) < \tau\}$.

High-level strategy: Given a sample $\mathcal{D} = ((X_1, Y_1), \ldots, (X_n, Y_n)) \overset{\text{i.i.d.}}{\sim} P$,

1. Sub-sample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Construct $\hat{p}_1, \ldots, \hat{p}_m$ so that each $\hat{p}_\ell$ is a p-value for $H_0(X_\ell)$

Overall strategy

Our objective is to choose $\hat{A} \equiv \hat{A}(\mathcal{D})$ which minimises regret $R_\tau(\hat{A})$, subject to Type 1 error guarantee: $\inf_{P \in \mathcal{P}_{\mathrm{M}}(d,\sigma)} \mathbb{P}_P\big(\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\big) \geq 1 - \alpha.$

To each $x \in \mathbb{R}^d$, we associate a null $H_0(x) := \{P \in \mathcal{P}_{\mathrm{M}}(d,\sigma) : \eta_P(x_0) < \tau\}$.

High-level strategy: Given a sample $\mathcal{D} = ((X_1, Y_1), \ldots, (X_n, Y_n)) \overset{\text{i.i.d.}}{\sim} P$,

1. Sub-sample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Construct $\hat{p}_1, \ldots, \hat{p}_m$ so that each $\hat{p}_\ell$ is a p-value for $H_0(X_\ell)$ i.e. $\mathbb{P}\big(\hat{p}_\ell \leq \alpha | (X_i)_{i=1}^m\big) \leq \alpha$ for all $P \in H_0(X_\ell)$ and $\alpha \in (0,1)$;

Overall strategy

Our objective is to choose $\hat{A} \equiv \hat{A}(\mathcal{D})$ which minimises regret $R_\tau(\hat{A})$, subject to
Type 1 error guarantee: $\inf_{P \in \mathcal{P}_{\mathrm{M}}(d,\sigma)} \mathbb{P}_P\bigl(\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\bigr) \geq 1 - \alpha.$

To each $x \in \mathbb{R}^d$, we associate a null $H_0(x) := \{P \in \mathcal{P}_{\mathrm{M}}(d,\sigma) : \eta_P(x_0) < \tau\}$.

High-level strategy: Given a sample $\mathcal{D} = ((X_1, Y_1), \ldots, (X_n, Y_n)) \overset{\text{i.i.d.}}{\sim} P$,

1. Sub-sample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Construct $\hat{p}_1, \ldots, \hat{p}_m$ so that each $\hat{p}_\ell$ is a p-value for $H_0(X_\ell)$ i.e. $\mathbb{P}\bigl(\hat{p}_\ell \leq \alpha | (X_i)_{i=1}^m\bigr) \leq \alpha$ for all $P \in H_0(X_\ell)$ and $\alpha \in (0,1)$;
3. Apply a multiple testing procedure to reject $\mathcal{R}_\alpha \subseteq \{1, \ldots, m\}$ with $\mathbb{P}_P\bigl(\mathcal{R}_\alpha \cap \{\ell \in \{1, \ldots, m\} : P \in H_0(X_\ell)\} \neq \emptyset | (X_i)_{i=1}^m\bigr) \leq \alpha$;

Overall strategy

Our objective is to choose $\hat{A} \equiv \hat{A}(\mathcal{D})$ which minimises regret $R_\tau(\hat{A})$, subject to Type 1 error guarantee: $\inf_{P \in \mathcal{P}_{\mathrm{M}}(d,\sigma)} \mathbb{P}_P\big(\hat{A}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\big) \geq 1 - \alpha.$

To each $x \in \mathbb{R}^d$, we associate a null $H_0(x) := \{P \in \mathcal{P}_{\mathrm{M}}(d,\sigma) : \eta_P(x_0) < \tau\}$.

High-level strategy: Given a sample $\mathcal{D} = ((X_1, Y_1), \ldots, (X_n, Y_n)) \overset{\text{i.i.d.}}{\sim} P$,

1. Sub-sample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Construct $\hat{p}_1, \ldots, \hat{p}_m$ so that each $\hat{p}_\ell$ is a p-value for $H_0(X_\ell)$ i.e. $\mathbb{P}\big(\hat{p}_\ell \leq \alpha | (X_i)_{i=1}^m\big) \leq \alpha$ for all $P \in H_0(X_\ell)$ and $\alpha \in (0,1)$;
3. Apply a multiple testing procedure to reject $\mathcal{R}_\alpha \subseteq \{1, \ldots, m\}$ with $\mathbb{P}_P\big(\mathcal{R}_\alpha \cap \{\ell \in \{1,\ldots,m\} : P \in H_0(X_\ell)\} \neq \emptyset | (X_i)_{i=1}^m\big) \leq \alpha$;
4. Output $\hat{A} := \big\{x \in \mathbb{R}^d : X_\ell \preccurlyeq x \text{ for some } \ell \in \mathcal{R}_\alpha\big\}$.

Overall strategy

High-level strategy: Given a sample $\mathcal{D} = ((X_1, Y_1), \ldots, (X_n, Y_n)) \overset{\text{i.i.d.}}{\sim} P$,

1. Sub-sample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Construct $\hat{p}_1, \ldots, \hat{p}_m$ so that each $\hat{p}_\ell$ is a p-value for $H_0(X_\ell)$ i.e. $\mathbb{P}\big(\hat{p}_\ell \leq \alpha | (X_i)_{i=1}^m\big) \leq \alpha$ for all $P \in H_0(X_\ell)$ and $\alpha \in (0, 1)$;
3. Apply a multiple testing procedure to reject $\mathcal{R}_\alpha \subseteq \{1, \ldots, m\}$ with $\mathbb{P}_P\big(\mathcal{R}_\alpha \cap \{\ell \in \{1, \ldots, m\} : P \in H_0(X_\ell)\} \neq \emptyset | (X_i)_{i=1}^m\big) \leq \alpha$;
4. Output $\hat{A} := \big\{x \in \mathbb{R}^d : X_\ell \preccurlyeq x \text{ for some } \ell \in \mathcal{R}_\alpha\big\}$.

Remarks:

- Any $\hat{A}$ constructed in this way controls Type 1 error at the level α;

Overall strategy

High-level strategy: Given a sample $\mathcal{D} = ((X_1, Y_1), \ldots, (X_n, Y_n)) \overset{\text{i.i.d.}}{\sim} P$,

1. Sub-sample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Construct $\hat{p}_1, \ldots, \hat{p}_m$ so that each $\hat{p}_\ell$ is a p-value for $H_0(X_\ell)$ i.e. $\mathbb{P}\big(\hat{p}_\ell \leq \alpha | (X_i)_{i=1}^m\big) \leq \alpha$ for all $P \in H_0(X_\ell)$ and $\alpha \in (0, 1)$;
3. Apply a multiple testing procedure to reject $\mathcal{R}_\alpha \subseteq \{1, \ldots, m\}$ with $\mathbb{P}_P\big(\mathcal{R}_\alpha \cap \{\ell \in \{1, \ldots, m\} : P \in H_0(X_\ell)\} \neq \emptyset | (X_i)_{i=1}^m\big) \leq \alpha$;
4. Output $\hat{A} := \big\{x \in \mathbb{R}^d : X_\ell \preccurlyeq x \text{ for some } \ell \in \mathcal{R}_\alpha\big\}$.

Remarks:

- Any $\hat{A}$ constructed in this way controls Type 1 error at the level α;
- To implement our strategy we require (2) p-values $\&$ (3) an MTP;

Overall strategy

High-level strategy: Given a sample $\mathcal{D} = ((X_1, Y_1), \dots, (X_n, Y_n)) \overset{\text{i.i.d.}}{\sim} P$,

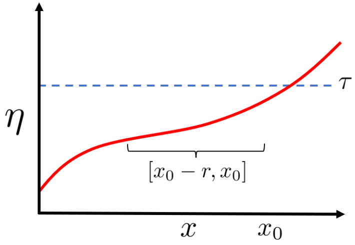
1. Sub-sample m covariate vectors $X_1, \dots, X_m$ with $m \leq n$;
2. Construct $\hat{p}_1, \dots, \hat{p}_m$ so that each $\hat{p}_\ell$ is a p-value for $H_0(X_\ell)$ i.e. $\mathbb{P}\big(\hat{p}_\ell \leq \alpha | (X_i)_{i=1}^m\big) \leq \alpha$ for all $P \in H_0(X_\ell)$ and $\alpha \in (0,1)$;
3. Apply a multiple testing procedure to reject $\mathcal{R}_\alpha \subseteq \{1, \dots, m\}$ with $\mathbb{P}_P\big(\mathcal{R}_\alpha \cap \{\ell \in \{1, \dots, m\} : P \in H_0(X_\ell)\} \neq \emptyset | (X_i)_{i=1}^m\big) \leq \alpha$;
4. Output $\hat{A} := \big\{x \in \mathbb{R}^d : X_\ell \preccurlyeq x \text{ for some } \ell \in \mathcal{R}_\alpha\big\}$.

Remarks:

- Any $\hat{A}$ constructed in this way controls Type 1 error at the level α;
- To implement our strategy we require (2) p-values $\&$ (3) an MTP;
- Care must be taken with (2) and (3) to avoid any unnecessary loss of power.

The uni-variate case

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



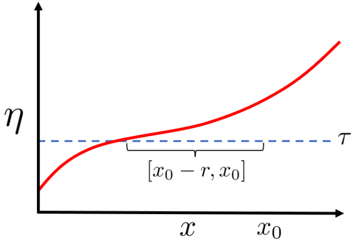
Note that for $P \in H_0(x_0)$ and $r \in [0, \infty)$ with $\{X_1, \ldots, X_n\} \cap [x_0 - r, x_0] \neq \emptyset$,

$$Z_{x_0,r} := \frac{\sum_{i=1}^{n} (Y_i - \tau) \cdot \mathbb{1}_{\{X_i \in [x_0 - r, x_0]\}}}{\sigma \cdot \sqrt{\sum_{i=1}^{n} \mathbb{1}_{\{X_i \in [x_0 - r, x_0]\}}}} \leq \frac{\sum_{i=1}^{n} (Y_i - \eta(X_i)) \cdot \mathbb{1}_{\{X_i \in [x_0 - r, x_0]\}}}{\sigma \cdot \sqrt{\sum_{i=1}^{n} \mathbb{1}_{\{X_i \in [x_0 - r, x_0]\}}}},$$

a 1 sub-Gaussian random variable.

The uni-variate case

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

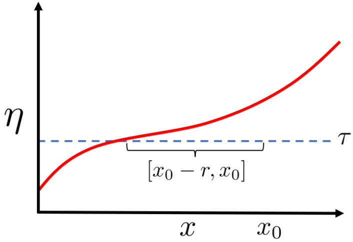


Given $r \in [0, \infty)$ with $\{X_1, \ldots, X_n\} \cap [x_0 - r, x_0] \neq \emptyset$,

$$Z_{x_0,r} \geq \frac{\sum_{i=1}^n (Y_i - \eta(X_i)) \cdot \mathbb{1}_{\{X_i \in [x_0 - r, x_0]\}}}{\sigma \cdot \sqrt{\sum_{i=1}^n \mathbb{1}_{\{X_i \in [x_0 - r, x_0]\}}}} + \frac{\eta(x_0 - r) - \tau}{\sigma} \cdot \sqrt{\sum_{i=1}^n \mathbb{1}_{\{X_i \in [x_0 - r, x_0]\}}}.$$

The uni-variate case

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



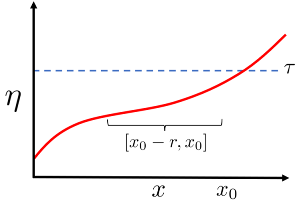
Given $r \in [0,\infty)$ with $\{X_1,\ldots,X_n\} \cap [x_0 - r, x_0] \neq \emptyset$,

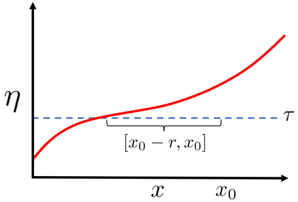
$$Z_{x_0,r} \geq \frac{\sum_{i=1}^n (Y_i - \eta(X_i)) \cdot \mathbb{1}_{\{X_i \in [x_0-r,x_0]\}}}{\sigma \cdot \sqrt{\sum_{i=1}^n \mathbb{1}_{\{X_i \in [x_0-r,x_0]\}}}} + \frac{\eta(x_0 - r) - \tau}{\sigma} \cdot \sqrt{\sum_{i=1}^n \mathbb{1}_{\{X_i \in [x_0-r,x_0]\}}}.$$

We would like to choose r to maximise power when $P \notin H_0(x_0)$.

The uni-variate case

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





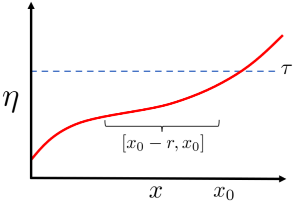
For $P \in H_0(x_0)$ and $r \in [0, \infty)$ with $\{X_1, \ldots, X_n\} \cap [x_0 - r, x_0] \neq \emptyset$,

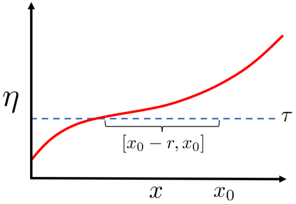
$$Z_{x_0,r} := \frac{\sum_{i=1}^n (Y_i - \tau) \cdot \mathbb{1}_{\{X_i \in [x_0 - r, x_0]\}}}{\sigma \cdot \sqrt{\sum_{i=1}^n \mathbb{1}_{\{X_i \in [x_0 - r, x_0]\}}}},$$

is stochastically dominated by a 1 sub-Gaussian random variable.

The uni-variate case

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





For $P \in H_0(x_0)$ and $r \in [0, \infty)$ with $\{X_1, \ldots, X_n\} \cap [x_0 - r, x_0] \neq \emptyset$,

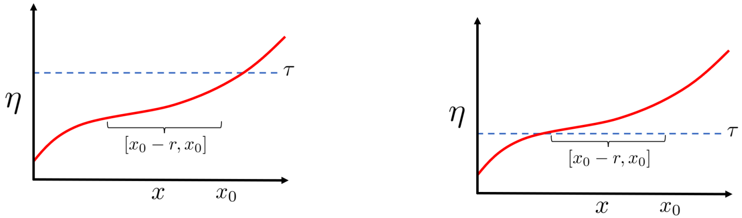
$$Z_{x_0,r} := \frac{\sum_{i=1}^n (Y_i - \tau) \cdot \mathbb{1}_{\{X_i \in [x_0 - r, x_0]\}}}{\sigma \cdot \sqrt{\sum_{i=1}^n \mathbb{1}_{\{X_i \in [x_0 - r, x_0]\}}}},$$

is stochastically dominated by a 1 sub-Gaussian random variable.

We obtain our p-value for $H_0(x_0)$ by inverting the confidence bands for the process $(Z_{x_0,r})_{r \geq 0}$.

The uni-variate case

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

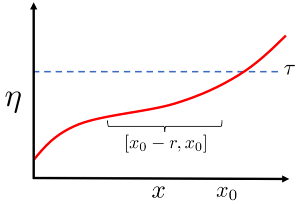


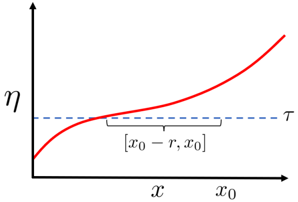
We obtain our p-value for $H_0(x_0)$ by inverting the confidence bands for the process $(Z_{x_0,r})_{r\geq 0}$, where

$$Z_{x_0,r} := \frac{\sum_{i=1}^{n}(Y_i - \tau)\cdot \mathbb{1}_{\{X_i \in [x_0-r,x_0]\}}}{\sigma \cdot \sqrt{\sum_{i=1}^{n} \mathbb{1}_{\{X_i \in [x_0-r,x_0]\}}}}.$$

We use the time-uniform confidence sequences of Howard et al. (2021).

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





We obtain our p-value for $H_0(x_0)$ by inverting the confidence bands for the process $(Z_{x_0,r})_{r\geq 0}$, where

$$Z_{x_0,r} := \frac{\sum_{i=1}^n (Y_i - \tau)\cdot \mathbb{1}_{\{X_i \in [x_0 - r, x_0]\}}}{\sigma \cdot \sqrt{\sum_{i=1}^n \mathbb{1}_{\{X_i \in [x_0 - r, x_0]\}}}}.$$

We use the time-uniform confidence sequences of Howard et al. (2021).

For each $\ell = 1, \ldots, m$, we let $\hat{p}_\ell$ be the p-value corresponding to $H_0(X_\ell)$.

The uni-variate case

For each $\ell = 1, \ldots, m$, we let $\hat{p}_\ell$ be the p-value corresponding to $H_0(X_\ell)$, i.e. $\mathbb{P}\big(\hat{p}_\ell \leq \alpha | (X_i)_{i=1}^m\big) \leq \alpha$ for all $P \in H_0(X_\ell)$ and $\alpha \in (0,1)$

Next, we require a multiple testing procedure to choose $\mathcal{R}_\alpha \subseteq \{1, \ldots, m\}$ with $\mathbb{P}_P\big(\mathcal{R}_\alpha \cap \{\ell \in \{1, \ldots, m\} : P \in H_0(X_\ell)\} \neq \emptyset | (X_i)_{i=1}^m\big) \leq \alpha$.

The uni-variate case

For each $\ell = 1, \ldots, m$, we let $\hat{p}_\ell$ be the p-value corresponding to $H_0(X_\ell)$, i.e. $\mathbb{P}\big(\hat{p}_\ell \leq \alpha | (X_i)_{i=1}^m\big) \leq \alpha$ for all $P \in H_0(X_\ell)$ and $\alpha \in (0, 1)$

Next, we require a multiple testing procedure to choose $\mathcal{R}_\alpha \subseteq \{1, \ldots, m\}$ with $\mathbb{P}_P\big(\mathcal{R}_\alpha \cap \{\ell \in \{1, \ldots, m\} : P \in H_0(X_\ell)\} \neq \emptyset | (X_i)_{i=1}^m\big) \leq \alpha$.

A simple approach is fixed sequence testing:

1. Order the null hypotheses independently of the data;
2. Test each null hypothesis sequentially at the level α;
3. Terminate the process with the first failed rejection;
4. Return the collection of rejected nulls.

The uni-variate case

For each $\ell = 1, \ldots, m$, we let $\hat{p}_\ell$ be the p-value corresponding to $H_0(X_\ell)$, i.e. $\mathbb{P}\big(\hat{p}_\ell \leq \alpha | (X_i)_{i=1}^m\big) \leq \alpha$ for all $P \in H_0(X_\ell)$ and $\alpha \in (0,1)$

Next, we require a multiple testing procedure to choose $\mathcal{R}_\alpha \subseteq \{1, \ldots, m\}$ with $\mathbb{P}_P\big(\mathcal{R}_\alpha \cap \{\ell \in \{1, \ldots, m\} : P \in H_0(X_\ell)\} \neq \emptyset | (X_i)_{i=1}^m\big) \leq \alpha$.

A simple approach is fixed sequence testing:

1. Order the null hypotheses independently of the data;
2. Test each null hypothesis sequentially at the level α;
3. Terminate the process with the first failed rejection;
4. Return the collection of rejected nulls.

Fixed sequence testing always controls the family-wise-error-rate (FWER).

The uni-variate case

For each $\ell = 1, \ldots, m$, we let $\hat{p}_\ell$ be the p-value corresponding to $H_0(X_\ell)$, i.e. $\mathbb{P}\big(\hat{p}_\ell \leq \alpha | (X_i)_{i=1}^m\big) \leq \alpha$ for all $P \in H_0(X_\ell)$ and $\alpha \in (0,1)$

Next, we require a multiple testing procedure to choose $\mathcal{R}_\alpha \subseteq \{1, \ldots, m\}$ with $\mathbb{P}_P\big(\mathcal{R}_\alpha \cap \{\ell \in \{1, \ldots, m\} : P \in H_0(X_\ell)\} \neq \emptyset | (X_i)_{i=1}^m\big) \leq \alpha$.

A simple approach is fixed sequence testing:

1. Order the null hypotheses independently of the data;
2. Test each null hypothesis sequentially at the level α;
3. Terminate the process with the first failed rejection;
4. Return the collection of rejected nulls.

Fixed sequence testing always controls the family-wise-error-rate (FWER).

Indeed, consider the first true null to be tested within the sequence. We will reject this null (and hence any subsequent null) with probability at most α.

The uni-variate case

We require a multiple testing procedure to choose $\mathcal{R}_\alpha \subseteq \{1, \ldots, m\}$ with $\mathbb{P}_P\big(\mathcal{R}_\alpha \cap \{\ell \in \{1, \ldots, m\} : P \in H_0(X_\ell)\} \neq \emptyset | (X_i)_{i=1}^m\big) \leq \alpha$.

Fixed sequence testing controls FWER at the nominal level α.

We should select our sequence to maximise power.

Logical structure: If $x_0 \preccurlyeq x_1$ then $H_0(x_1) \subseteq H_0(x_0)$.

Hence, if test our null hypotheses from right to left then we will test all of our false nulls before we test any true nulls.

We combine our p-values via a fixed sequence testing procedure

$$\mathcal{R}_\alpha := \{i \in \{1, \ldots, m\} : \hat{p}_\ell \leq \alpha \text{ whenever } X_\ell \geq X_i\}.$$

The uni-variate case

High-level strategy: Given a sample $\mathcal{D} = ((X_1, Y_1), \ldots, (X_n, Y_n)) \overset{\text{i.i.d.}}{\sim} P$,

1. Sub-sample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Construct $\hat{p}_1, \ldots, \hat{p}_m$ so that each $\hat{p}_\ell$ is a p-value for $H_0(X_\ell)$ i.e. $\mathbb{P}\big(\hat{p}_\ell \leq \alpha | (X_i)_{i=1}^m\big) \leq \alpha$ for all $P \in H_0(X_\ell)$ and $\alpha \in (0, 1)$;
3. Apply a multiple testing procedure to reject $\mathcal{R}_\alpha \subseteq \{1, \ldots, m\}$ with $\mathbb{P}_P\big(\mathcal{R}_\alpha \cap \{\ell \in \{1, \ldots, m\} : P \in H_0(X_\ell)\} \neq \emptyset | (X_i)_{i=1}^m\big) \leq \alpha$;
4. Output $\hat{A} := \big\{x \in \mathbb{R}^d : X_\ell \preccurlyeq x \text{ for some } \ell \in \mathcal{R}_\alpha\big\}$.

Our p-values $\hat{p}_\ell$ leverage the time-uniform sequences of Howard et al. (2021).

We combine our p-values with a fixed sequence testing procedure:

$$\mathcal{R}_\alpha := \{i \in \{1, \ldots, m\} : \hat{p}_\ell \leq \alpha \text{ whenever } X_\ell \geq X_i\}\,.$$

Return $\hat{A}_{\text{ISS}} = [X_{i_{\min}}, \infty)$ where $i_{\min} := \min\{i \in [m] : \hat{p}_\ell \leq \alpha \text{ for } X_\ell \geq X_i\}$.

The uni-variate case

Our construction ensures that we control type 1 error at the nominal level,

$$\mathbb{P}_P\big(\hat{A}_{\mathrm{ISS}}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\big) \geq 1 - \alpha,$$

for all $P \in \mathcal{P}_{\mathrm{M}}(d, \sigma)$.

We also wish to minimise the regret $R_\tau(\hat{A}) := \mathbb{E}_P\big\{\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}\big)\big\}$.

In order to control the regret $R_\tau(\hat{A})$ we must place some restrictions on the amount of mass μ places in regions where η is just above τ.

For $d \in \mathbb{N}$, $\tau \in \mathbb{R}$, $\beta > 0$ and $\nu > 0$, we let $\mathcal{P}_{\mathrm{R}}(d, \tau, \beta, \nu)$ denote the class of distributions P on $\mathbb{R}^d \times \mathbb{R}$ for which the marginal μ on $\mathbb{R}^d$ and the regression function $\eta : \mathbb{R}^d \to \mathbb{R}$ satisfy $\mu\big(\eta^{-1}([\tau, \tau + \nu\xi^\beta])\big) \leq \xi$ for all $\xi \in (0, 1]$.

Power bounds in the uni-variate setting

We also wish to minimise the regret $R_\tau(\hat{A}) := \mathbb{E}_P\{\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}\big)\}$.

In order to control the regret $R_\tau(\hat{A})$ we must place some restrictions on the amount of mass μ places in regions where η is just above τ.

For $d \in \mathbb{N}$, $\tau \in \mathbb{R}$, $\beta > 0$ and $\nu > 0$, we let $\mathcal{P}_{\mathrm{R}}(d, \tau, \beta, \nu)$ denote the class of distributions P on $\mathbb{R}^d \times \mathbb{R}$ for which the marginal μ on $\mathbb{R}^d$ and the regression function $\eta : \mathbb{R}^d \to \mathbb{R}$ satisfy $\mu\big(\eta^{-1}([\tau, \tau + \nu\xi^\beta])\big) \leq \xi$ for all $\xi \in (0, 1]$.

Example: Let $d = 1$ and let $P \in \mathcal{P}_M(d, \sigma)$ have uniform marginal distribution μ on $[0, 1]$. We then have $P \in \mathcal{P}_M(d, \sigma) \cap \mathcal{P}_{\mathrm{R}}(d, \tau, \beta, \nu)$ if

$$\eta(x + \xi) \geq \tau + \nu\xi^\beta,$$

for all $\xi \in (0, 1]$ and $x \in \mathcal{X}_\tau(\eta)$.

Power bounds in the uni-variate setting

Our construction ensures that we control type 1 error at the nominal level,

$$\mathbb{P}_P\big(\hat{A}_{\mathrm{ISS}}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\big) \geq 1 - \alpha,$$

for all $P \in \mathcal{P}_{\mathrm{M}}(1, \sigma)$.

For $d \in \mathbb{N}$, $\tau \in \mathbb{R}$, $\beta > 0$ and $\nu > 0$, we let $\mathcal{P}_{\mathrm{R}}(d, \tau, \beta, \nu)$ denote the class of distributions P on $\mathbb{R}^d \times \mathbb{R}$ for which the marginal μ on $\mathbb{R}^d$ and the regression function $\eta : \mathbb{R}^d \to \mathbb{R}$ satisfy $\mu\big(\eta^{-1}([\tau, \tau + \nu\xi^\beta])\big) \leq \xi$ for all $\xi \in (0, 1]$.

Theorem. Let $\sigma, \beta, \nu > 0$ and $\alpha \in (0, 1)$. There exists a universal constant $C \geq 1$ such that for any distribution $P \in \mathcal{P}_M(1, \sigma) \cap \mathcal{P}_{\mathrm{R}}(1, \tau, \beta, \nu)$ we have

$$\mathbb{E}_P\big\{\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}_{\mathrm{ISS}}(\mathcal{D})\big)\big\} \leq 1 \wedge C\left\{\left(\frac{\sigma^2}{n\nu^2}\log_+\left(\frac{\log_+ n}{\alpha}\right)\right)^{1/(2\beta+1)} + \frac{1}{n}\right\}.$$

The multi-variate case

High-level strategy: Given a sample $\mathcal{D} = ((X_1, Y_1), \ldots, (X_n, Y_n)) \overset{\text{i.i.d.}}{\sim} P$,

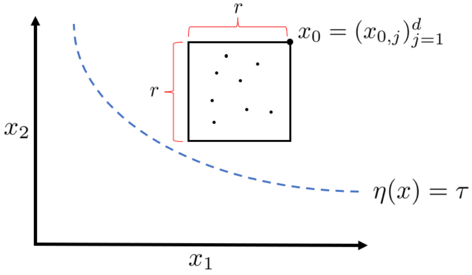
1. Sub-sample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Construct $\hat{p}_1, \ldots, \hat{p}_m$ so that each $\hat{p}_\ell$ is a p-value for $H_0(X_\ell)$ i.e. $\mathbb{P}\big(\hat{p}_\ell \leq \alpha | (X_i)_{i=1}^m\big) \leq \alpha$ for all $P \in H_0(X_\ell)$ and $\alpha \in (0,1)$;
3. Apply a multiple testing procedure to reject $\mathcal{R}_\alpha \subseteq \{1, \ldots, m\}$ with $\mathbb{P}_P\big(\mathcal{R}_\alpha \cap \{\ell \in \{1, \ldots, m\} : P \in H_0(X_\ell)\} \neq \emptyset | (X_i)_{i=1}^m\big) \leq \alpha$;
4. Output $\hat{A} := \big\{x \in \mathbb{R}^d : X_\ell \preccurlyeq x \text{ for some } \ell \in \mathcal{R}_\alpha\big\}$.

The multi-variate case

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

The multi-variate case

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



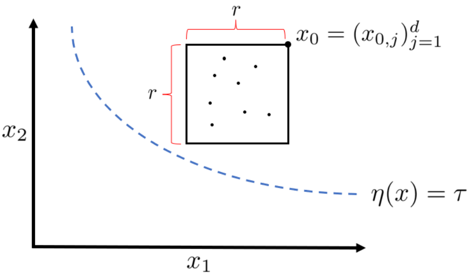
Given $x_0 = (x_{0,j})_{j=1}^d \in \mathbb{R}^d$ and $r > 0$ with $\{X_\ell\}_{\ell=1}^n \cap \prod_{j=1}^d [x_{0,j} - r, x_{0,j}] \neq \emptyset$,

$$Z_{x_0,r} := \frac{\sum_{i=1}^n (Y_i - \tau) \cdot \mathbb{1}_{\{X_i \in \prod_{j=1}^d [x_{0,j}-r, x_{0,j}]\}}}{\sigma \cdot \sqrt{\sum_{i=1}^n \mathbb{1}_{\{X_i \in \prod_{j=1}^d [x_{0,j}-r, x_{0,j}]\}}}},$$

is dominated by a 1 sub-Gaussian random variable when $P \in H_0(x_0)$.

The multi-variate case

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



We obtain our p-values $\hat{p}_\ell$ for $H_0(X_\ell)$ by inverting the confidence bands for the process $(Z_{x_0,r})_{r \geq 0}$ with $x_0 = X_\ell$.

Our p-values $\hat{p}_\ell$ leverage the time-uniform sequences of Howard et al. (2021).

The multi-variate case

High-level strategy: Given a sample $\mathcal{D} = ((X_1, Y_1), \ldots, (X_n, Y_n)) \stackrel{\text{i.i.d.}}{\sim} P$,

1. Sub-sample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Construct $\hat{p}_1, \ldots, \hat{p}_m$ so that each $\hat{p}_\ell$ is a p-value for $H_0(X_\ell)$ i.e. $\mathbb{P}\big(\hat{p}_\ell \leq \alpha | (X_i)_{i=1}^m\big) \leq \alpha$ for all $P \in H_0(X_\ell)$ and $\alpha \in (0, 1)$;
3. Apply a multiple testing procedure to reject $\mathcal{R}_\alpha \subseteq \{1, \ldots, m\}$ with $\mathbb{P}_P\big(\mathcal{R}_\alpha \cap \{\ell \in \{1, \ldots, m\} : P \in H_0(X_\ell)\} \neq \emptyset | (X_i)_{i=1}^m\big) \leq \alpha$;
4. Output $\hat{A} := \big\{x \in \mathbb{R}^d : X_\ell \preccurlyeq x \text{ for some } \ell \in \mathcal{R}_\alpha\big\}$.

To complete our procedure we require a suitable multiple testing procedure.

The multi-variate case

High-level strategy: Given a sample $\mathcal{D} = ((X_1, Y_1), \ldots, (X_n, Y_n)) \overset{\text{i.i.d.}}{\sim} P$,

1. Sub-sample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Construct $\hat{p}_1, \ldots, \hat{p}_m$ so that each $\hat{p}_\ell$ is a p-value for $H_0(X_\ell)$ i.e. $\mathbb{P}\big(\hat{p}_\ell \leq \alpha | (X_i)_{i=1}^m\big) \leq \alpha$ for all $P \in H_0(X_\ell)$ and $\alpha \in (0,1)$;
3. Apply a multiple testing procedure to reject $\mathcal{R}_\alpha \subseteq \{1, \ldots, m\}$ with $\mathbb{P}_P\big(\mathcal{R}_\alpha \cap \{\ell \in \{1, \ldots, m\} : P \in H_0(X_\ell)\} \neq \emptyset | (X_i)_{i=1}^m\big) \leq \alpha$;
4. Output $\hat{A} := \big\{x \in \mathbb{R}^d : X_\ell \preccurlyeq x \text{ for some } \ell \in \mathcal{R}_\alpha\big\}$.

To complete our procedure we require a suitable multiple testing procedure.

Our p-values $\hat{p}_1, \ldots, \hat{p}_m$ no longer have a natural sequential structure since $\{X_1, \ldots, X_m\}$ are not totally ordered by $\preccurlyeq$.

The multi-variate case

High-level strategy: Given a sample $\mathcal{D} = ((X_1, Y_1), \ldots, (X_n, Y_n)) \overset{\text{i.i.d.}}{\sim} P$,

1. Sub-sample m covariate vectors $X_1, \ldots, X_m$ with $m \leq n$;
2. Construct $\hat{p}_1, \ldots, \hat{p}_m$ so that each $\hat{p}_\ell$ is a p-value for $H_0(X_\ell)$ i.e. $\mathbb{P}\big(\hat{p}_\ell \leq \alpha | (X_i)_{i=1}^m\big) \leq \alpha$ for all $P \in H_0(X_\ell)$ and $\alpha \in (0,1)$;
3. Apply a multiple testing procedure to reject $\mathcal{R}_\alpha \subseteq \{1, \ldots, m\}$ with $\mathbb{P}_P\big(\mathcal{R}_\alpha \cap \{\ell \in \{1, \ldots, m\} : P \in H_0(X_\ell)\} \neq \emptyset | (X_i)_{i=1}^m\big) \leq \alpha$;
4. Output $\hat{A} := \big\{x \in \mathbb{R}^d : X_\ell \preccurlyeq x \text{ for some } \ell \in \mathcal{R}_\alpha\big\}$.

To complete our procedure we require a suitable multiple testing procedure.

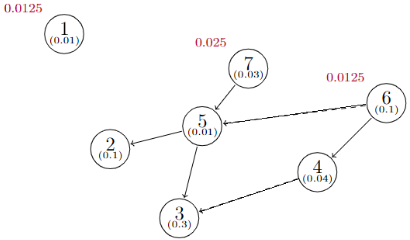
Our p-values $\hat{p}_1, \ldots, \hat{p}_m$ no longer have a natural sequential structure since $\{X_1, \ldots, X_m\}$ are not totally ordered by $\preccurlyeq$.

Logical structure: If $x_0 \preccurlyeq x_1$ then $H_0(x_1) \subseteq H_0(x_0)$.

Multiple testing procedures for DAGs

Our null hypotheses may be structured within a DAG $G = (V, E)$:

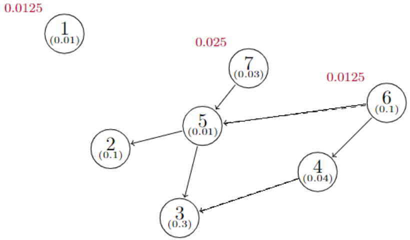
1. $V := \{1, \ldots, m\}$ with each vertex $\ell \in V$ associated to a null $H_0(X_\ell)$;
2. $E := \big\{(i_0, i_1) \in [m]^2 : i_0 \neq i_1$ and $X_{i_1} \preccurlyeq X_{i_0}$, and if $X_{i_1} \preccurlyeq X_{i_2} \preccurlyeq X_{i_0}$ then either $X_{i_2} = X_{i_0}$ and $i_0 \leq i_2$, or $X_{i_2} = X_{i_1}$ and $i_2 \leq i_1\big\}$.

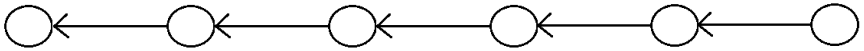


Multiple testing procedures for DAGs

Our null hypotheses may be structured within a DAG $G = (V, E)$:

1. $V := \{1, \ldots, m\}$ with each vertex $\ell \in V$ associated to a null $H_0(X_\ell)$;
2. $E := \big\{(i_0, i_1) \in [m]^2 : i_0 \neq i_1$ and $X_{i_1} \preccurlyeq X_{i_0}$, and if $X_{i_1} \preccurlyeq X_{i_2} \preccurlyeq X_{i_0}$ then either $X_{i_2} = X_{i_0}$ and $i_0 \leq i_2$, or $X_{i_2} = X_{i_1}$ and $i_2 \leq i_1\big\}$.





Multiple testing procedures for DAGs

Logical structure: If $x_0 \preccurlyeq x_1$ then $H_0(x_1) \subseteq H_0(x_0)$.

Hence, our null hypotheses have the structure of a DAG $G = (V, E)$:

1. $V := \{1, \ldots, m\}$ with each vertex $\ell \in V$ associated to a null $H_0(X_\ell)$;
2. $E := \big\{(i_0, i_1) \in [m]^2 : i_0 \neq i_1$ and $X_{i_1} \preccurlyeq X_{i_0}$, and if $X_{i_1} \preccurlyeq X_{i_2} \preccurlyeq X_{i_0}$ then either $X_{i_2} = X_{i_0}$ and $i_0 \leq i_2$, or $X_{i_2} = X_{i_1}$ and $i_2 \leq i_1\big\}$.

The logical structure between the null hypotheses is reflected within the graphical structure:

1. The truth of a null $H_0(X_\ell)$ implies the truth of all the nulls $H_0(X_{\ell'})$ such that ℓ' is a G-descendent of ℓ;

Multiple testing procedures for DAGs

Logical structure: If $x_0 \preccurlyeq x_1$ then $H_0(x_1) \subseteq H_0(x_0)$.

Hence, our null hypotheses have the structure of a DAG $G = (V, E)$:

1. $V := \{1, \ldots, m\}$ with each vertex $\ell \in V$ associated to a null $H_0(X_\ell)$;
2. $E := \big\{(i_0, i_1) \in [m]^2 : i_0 \neq i_1$ and $X_{i_1} \preccurlyeq X_{i_0}$, and if $X_{i_1} \preccurlyeq X_{i_2} \preccurlyeq X_{i_0}$ then either $X_{i_2} = X_{i_0}$ and $i_0 \leq i_2$, or $X_{i_2} = X_{i_1}$ and $i_2 \leq i_1\big\}$.

The logical structure between the null hypotheses is reflected within the graphical structure:

1. The truth of a null $H_0(X_\ell)$ implies the truth of all the nulls $H_0(X_{\ell'})$ such that ℓ' is a G-descendent of ℓ;
2. Equivalently, the falsity of a null $H_0(X_{\ell'})$ implies the falsity of all of the nulls $H_0(X_\ell)$ such that ℓ is a G-ancestor of ℓ'.

Multiple testing procedures for DAGs

Our null hypotheses have the structure of a DAG $G = (V, E)$:

1. $V := \{1, \ldots, m\}$ with each vertex $\ell \in V$ associated to a null $H_0(X_\ell)$;
2. $E := \big\{(i_0, i_1) \in [m]^2 : i_0 \neq i_1$ and $X_{i_1} \preccurlyeq X_{i_0}$, and if $X_{i_1} \preccurlyeq X_{i_2} \preccurlyeq X_{i_0}$ then either $X_{i_2} = X_{i_0}$ and $i_0 \leq i_2$, or $X_{i_2} = X_{i_1}$ and $i_2 \leq i_1\big\}$.

The logical structure between the nulls is reflected within the graphical structure: The falsity of a null implies the falsity of all of its ancestors.

A variety of DAG based multiple testing procedures (MTP) have been proposed (Bretz et al. 2009, Meijer and Goeman, 2015, Ramdas et al. 2019).

Meijer and Goeman (2015) propose two MTPs for controlling the family wise error for logically structured hypotheses within a DAG.

These MTPs follow the sequential rejection principle (Goeman and Solari, 2010).

Multiple testing procedures for DAGs

The logical structure between the nulls is reflected within the graphical structure: The falsity of a null implies the falsity of all of its ancestors.

Meijer and Goeman (2015) propose two MTPs (all-parents and any-parent) for controlling the FWER for logically structured hypotheses within a DAG.

Meijer and Goeman's all-parent method (2015): In each iteration,

Multiple testing procedures for DAGs

The logical structure between the nulls is reflected within the graphical structure: The falsity of a null implies the falsity of all of its ancestors.

Meijer and Goeman (2015) propose two MTPs (all-parents and any-parent) for controlling the FWER for logically structured hypotheses within a DAG.

Meijer and Goeman's all-parent method (2015): In each iteration,

1. The α-budget is split evenly amongst the unrejected leaf nodes;

Multiple testing procedures for DAGs

The logical structure between the nulls is reflected within the graphical structure: The falsity of a null implies the falsity of all of its ancestors.

Meijer and Goeman (2015) propose two MTPs (all-parents and any-parent) for controlling the FWER for logically structured hypotheses within a DAG.

Meijer and Goeman's all-parent method (2015): In each iteration,

1. The α-budget is split evenly amongst the unrejected leaf nodes;
2. The α-budgets are then propagated from the leaves towards currently unrejected ancestors.

Multiple testing procedures for DAGs

The logical structure between the nulls is reflected within the graphical structure: The falsity of a null implies the falsity of all of its ancestors.

Meijer and Goeman (2015) propose two MTPs (all-parents and any-parent) for controlling the FWER for logically structured hypotheses within a DAG.

Meijer and Goeman's all-parent method (2015): In each iteration,

1. The α-budget is split evenly amongst the unrejected leaf nodes;
2. The α-budgets are then propagated from the leaves towards currently unrejected ancestors. Nodes are visited in reverse topological order, with each node distributing its entire budget amongst its unrejected parents;

Multiple testing procedures for DAGs

The logical structure between the nulls is reflected within the graphical structure: The falsity of a null implies the falsity of all of its ancestors.

Meijer and Goeman (2015) propose two MTPs (all-parents and any-parent) for controlling the FWER for logically structured hypotheses within a DAG.

Meijer and Goeman's all-parent method (2015): In each iteration,

1. The α-budget is split evenly amongst the unrejected leaf nodes;
2. The α-budgets are then propagated from the leaves towards currently unrejected ancestors. Nodes are visited in reverse topological order, with each node distributing its entire budget amongst its unrejected parents;
3. Reject all hypotheses whose p-value does not exceed the assigned budget.

Multiple testing procedures for DAGs

Meijer and Goeman's all-parent method (2015): In each iteration,

1. The α-budget is split evenly amongst the unrejected leaf nodes;
2. The α-budgets are then propagated from the leaves towards currently unrejected ancestors. Nodes are visited in reverse topological order, with each node distributing its entire budget amongst its unrejected parents;
3. Reject all hypotheses whose p-value does not exceed the assigned budget.

Both the all-parent and the any-parent MTP control the FWER (M & G, 2015).

Multiple testing procedures for DAGs

Meijer and Goeman's all-parent method (2015): In each iteration,

1. The α-budget is split evenly amongst the unrejected leaf nodes;
2. The α-budgets are then propagated from the leaves towards currently unrejected ancestors. Nodes are visited in reverse topological order, with each node distributing its entire budget amongst its unrejected parents;
3. Reject all hypotheses whose p-value does not exceed the assigned budget.

Both the all-parent and the any-parent MTP control the FWER (M & G, 2015).

Both MTPs also generalise the fixed sequence testing procedure.

Multiple testing procedures for DAGs

Meijer and Goeman's all-parent method (2015): In each iteration,

1. The α-budget is split evenly amongst the unrejected leaf nodes;
2. The α-budgets are then propagated from the leaves towards currently unrejected ancestors. Nodes are visited in reverse topological order, with each node distributing its entire budget amongst its unrejected parents;
3. Reject all hypotheses whose p-value does not exceed the assigned budget.

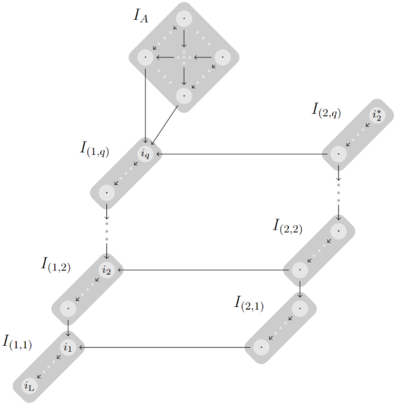
Both the all-parent and the any-parent MTP control the FWER (M & G, 2015).

Both MTPs also generalise the fixed sequence testing procedure.

Unfortunately, both MTPs are vulnerable to "bottleneck" effects whereby root nodes can be left with a very small fraction of the α-budget.

Multiple testing procedures for DAGs

Both the all-parent and the any-parent MTP control the FWER (M & G, 2015).



Unfortunately, both MTPs are vulnerable to “bottleneck” effects whereby root nodes can be left with an exponentially small fraction of the α-budget.

Multiple testing procedures for DAGs

Our null hypotheses have the structure of a DAG $G = (V, E)$:

1. $V := \{1, \ldots, m\}$ with each vertex $\ell \in V$ associated to a null $H_0(X_\ell)$;
2. $E := \big\{(i_0, i_1) \in [m]^2 : i_0 \neq i_1$ and $X_{i_1} \preccurlyeq X_{i_0}$, and if $X_{i_1} \preccurlyeq X_{i_2} \preccurlyeq X_{i_0}$ then either $X_{i_2} = X_{i_0}$ and $i_0 \leq i_2$, or $X_{i_2} = X_{i_1}$ and $i_2 \leq i_1\big\}$.

In order to avoid these "bottlenecks" we propose an alternative MTP, based on Meijer and Goeman (2015), and Goeman and Solari (2010).

Multiple testing procedures for DAGs

Our null hypotheses have the structure of a DAG $G = (V, E)$:

1. $V := \{1, \dots, m\}$ with each vertex $\ell \in V$ associated to a null $H_0(X_\ell)$;
2. $E := \big\{(i_0, i_1) \in [m]^2 : i_0 \neq i_1$ and $X_{i_1} \preccurlyeq X_{i_0}$, and if $X_{i_1} \preccurlyeq X_{i_2} \preccurlyeq X_{i_0}$ then either $X_{i_2} = X_{i_0}$ and $i_0 \leq i_2$, or $X_{i_2} = X_{i_1}$ and $i_2 \leq i_1\big\}$.

In order to avoid these “bottlenecks” we propose an alternative MTP, based on Meijer and Goeman (2015), and Goeman and Solari (2010).

We introduce an auxillary graph $F = (V, E_F)$ with nodes $V = \{1, \dots, m\}$ and $E_F \subseteq E$ chosen so that if $(i_0, i_1) \in E$ for some $i_0, i_1 \in V$, then there is exactly one $\tilde{i}_0 \in V$ with $(\tilde{i}_0, i_1) \in E_F$.

Multiple testing procedures for DAGs

Our null hypotheses have the structure of a DAG $G = (V, E)$:

1. $V := \{1, \ldots, m\}$ with each vertex $\ell \in V$ associated to a null $H_0(X_\ell)$;
2. $E := \big\{(i_0, i_1) \in [m]^2 : i_0 \neq i_1$ and $X_{i_1} \preccurlyeq X_{i_0}$, and if $X_{i_1} \preccurlyeq X_{i_2} \preccurlyeq X_{i_0}$ then either $X_{i_2} = X_{i_0}$ and $i_0 \leq i_2$, or $X_{i_2} = X_{i_1}$ and $i_2 \leq i_1\big\}$.

In order to avoid these "bottlenecks" we propose an alternative MTP, based on Meijer and Goeman (2015), and Goeman and Solari (2010).

We introduce an auxillary graph $F = (V, E_F)$ with nodes $V = \{1, \ldots, m\}$ and $E_F \subseteq E$ chosen so that if $(i_0, i_1) \in E$ for some $i_0, i_1 \in V$, then there is exactly one $\tilde{i}_0 \in V$ with $(\tilde{i}_0, i_1) \in E_F$.

That is, the graph F is a sparsification of G so that each node has at most one parent (a polyforest).

Multiple testing procedures for DAGs

Our null hypotheses have the structure of a DAG $G = (V, E)$:

1. $V := \{1, \dots, m\}$ with each vertex $\ell \in V$ associated to a null $H_0(X_\ell)$;
2. $E := \big\{(i_0, i_1) \in [m]^2 : i_0 \neq i_1$ and $X_{i_1} \preccurlyeq X_{i_0}$, and if $X_{i_1} \preccurlyeq X_{i_2} \preccurlyeq X_{i_0}$ then either $X_{i_2} = X_{i_0}$ and $i_0 \leq i_2$, or $X_{i_2} = X_{i_1}$ and $i_2 \leq i_1\big\}$.

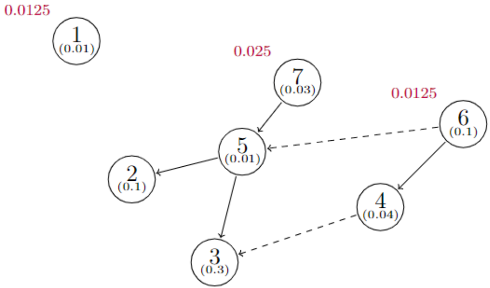
We let $F = (V, E_F)$ be a polyforest which serves as a sparsification of G.

Sparse DAG MTP: In each iteration,

1. The α-budget is distributed amongst the nodes in proportion to the number of unrejected F-leaves which are F-descended from the node;
2. Reject all hypotheses whose F parents have already been rejected and whose p-value does not exceed the assigned budget;
3. Reject also all G-ancestors of currently rejected nodes.

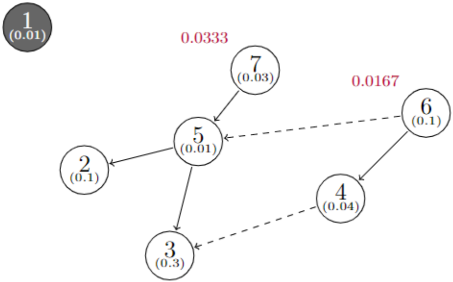
Multiple testing procedures for DAGs

In the first iteration, no hypothesis has been rejected yet and only root nodes are assigned positive α-budget.



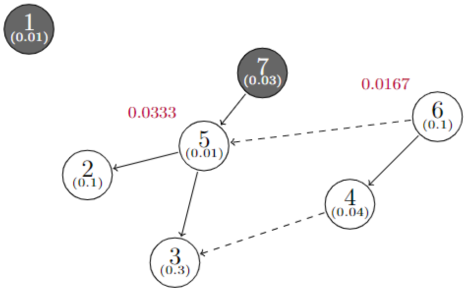
Here, nodes $1, 6$ and 7 are current rejection candidates, and 1 will be rejected, as $p_1 = 0.01 \leq 0.0125$.

After rejection of node 1 in the first step, we reallocate the α-budget, which allows us to reject node 7.



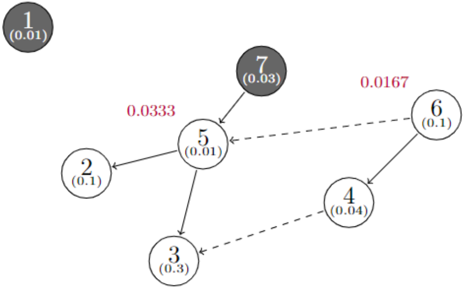
Multiple testing procedures for DAGs

Now that node 7 has been rejected, its child 5 receives α-budget sufficiently large for it to be rejected.



Multiple testing procedures for DAGs

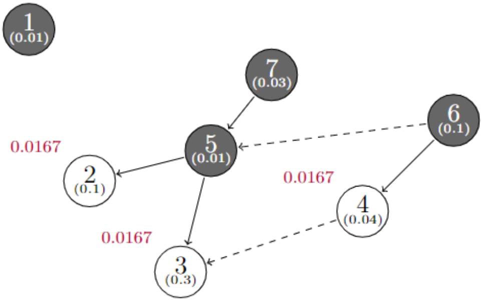
Now that node 7 has been rejected, its child 5 receives α-budget sufficiently large for it to be rejected.



Although p_6 is quite large, 6 is an ancestor of 5 in the induced DAG and will hence also be rejected.

Multiple testing procedures for DAGs

None of the remaining three nodes, which happen to be the leaf nodes, have a p-value smaller than their respective α-budgets. Hence, no further rejection is made and the procedure terminates.



Nodes 1, 5, 6 and 7 have been rejected.

The multi-variate case

High-level strategy: Given a sample $\mathcal{D} = ((X_1, Y_1), \dots, (X_n, Y_n)) \overset{\text{i.i.d.}}{\sim} P$,

1. Sub-sample m covariate vectors $X_1, \dots, X_m$ with $m \leq n$;
2. Construct $\hat{p}_1, \dots, \hat{p}_m$ so that each $\hat{p}_\ell$ is a p-value for $H_0(X_\ell)$ i.e. $\mathbb{P}\big(\hat{p}_\ell \leq \alpha | (X_i)_{i=1}^m\big) \leq \alpha$ for all $P \in H_0(X_\ell)$ and $\alpha \in (0,1)$;
3. Apply a multiple testing procedure to reject $\mathcal{R}_\alpha \subseteq \{1, \dots, m\}$ with $\mathbb{P}_P\big(\mathcal{R}_\alpha \cap \{\ell \in \{1, \dots, m\} : P \in H_0(X_\ell)\} \neq \emptyset | (X_i)_{i=1}^m\big) \leq \alpha$;
4. Output $\hat{A} := \big\{x \in \mathbb{R}^d : X_\ell \preccurlyeq x \text{ for some } \ell \in \mathcal{R}_\alpha\big\}$.

Our p-values $\hat{p}_\ell$ leverage the time-uniform sequences of Howard et al. (2021).

We combine our p-values with multiple testing procedure for DAGs which leverages an auxillary sparsified polyforest (Sparse DAG MTP).

Finally, we output the upper-hull $\hat{A}_{\text{ISS}} := \big\{x \in \mathbb{R}^d : X_\ell \preccurlyeq x \text{ for some } \ell \in \mathcal{R}_\alpha\big\}$.

Power bounds in the multi-variate setting

By adapting the approach of Goeman and Solari (2010) we see show that

$$\mathbb{P}_P\big(\hat{A}_{\mathrm{ISS}}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\big) \geq 1 - \alpha,$$

for all $P \in \mathcal{P}_{\mathrm{M}}(d, \sigma)$.

We also wish to minimise the regret $R_\tau(\hat{A}) := \mathbb{E}_P\big\{\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}\big)\big\}$.

Power bounds in the multi-variate setting

By adapting the approach of Goeman and Solari (2010) we see show that

$$\mathbb{P}_P\big(\hat{A}_{\mathrm{ISS}}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\big) \geq 1 - \alpha,$$

for all $P \in \mathcal{P}_{\mathrm{M}}(d, \sigma)$.

We also wish to minimise the regret $R_\tau(\hat{A}) := \mathbb{E}_P\big\{\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}\big)\big\}$.

Recall that in the uni-variate case we can bound regret uniformly over the class $\mathcal{P}_{\mathrm{R}}(d, \tau, \beta, \nu)$ consisting of all distributions P on $\mathbb{R}^d \times \mathbb{R}$ for which $\mu\big(\eta^{-1}([\tau, \tau + \nu\xi^\beta])\big) \leq \xi$ for all $\xi \in (0, 1]$.

Power bounds in the multi-variate setting

By adapting the approach of Goeman and Solari (2010) we see show that

$$\mathbb{P}_P\big(\hat{A}_{\mathrm{ISS}}(\mathcal{D}) \subseteq \mathcal{X}_\tau(\eta)\big) \geq 1 - \alpha,$$

for all $P \in \mathcal{P}_{\mathrm{M}}(d, \sigma)$.

We also wish to minimise the regret $R_\tau(\hat{A}) := \mathbb{E}_P\big\{\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}\big)\big\}$.

Recall that in the uni-variate case we can bound regret uniformly over the class $\mathcal{P}_{\mathrm{R}}(d, \tau, \beta, \nu)$ consisting of all distributions P on $\mathbb{R}^d \times \mathbb{R}$ for which $\mu\big(\eta^{-1}([\tau, \tau + \nu\xi^\beta])\big) \leq \xi$ for all $\xi \in (0, 1]$.

We shall see that this condition is insufficient to bound regret in the multi-variate setting.

Power bounds in the multi-variate setting

Recall that in the uni-variate case we can bound regret uniformly over the class $\mathcal{P}_{\mathrm{R}}(d, \tau, \beta, \nu)$ consisting of all distributions P on $\mathbb{R}^d \times \mathbb{R}$ for which $\mu\big(\eta^{-1}([\tau, \tau + \nu\xi^{\beta}])\big) \leq \xi$ for all $\xi \in (0, 1]$.

Proposition. Let $d \geq 2$, $\tau \in \mathbb{R}$, $\sigma, \beta, \nu > 0$ and $\alpha \in (0, 1)$. For all $n \in \mathbb{N}$,

$$\sup_{P} \inf_{\hat{A}} \mathbb{E}_P \big\{ \mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}(\mathcal{D})\big) \big\} \geq 1 - \alpha,$$

where the sup is over $P \in \mathcal{P}' = \mathcal{P}_M(d, \sigma) \cap \mathcal{P}_{\mathrm{R}}(d, \tau, \beta, \nu)$ and the inf is over procedures $\hat{A}$ which control the Type 1 error at the level α over $\mathcal{P}'$.

Power bounds in the multi-variate setting

Recall that in the uni-variate case we can bound regret uniformly over the class $\mathcal{P}_{\mathrm{R}}(d,\tau,\beta,\nu)$ consisting of all distributions P on $\mathbb{R}^d \times \mathbb{R}$ for which $\mu\big(\eta^{-1}([\tau,\tau+\nu\xi^{\beta}])\big) \leq \xi$ for all $\xi \in (0,1]$.

Proposition. Let $d \geq 2$, $\tau \in \mathbb{R}$, $\sigma,\beta,\nu > 0$ and $\alpha \in (0,1)$. For all $n \in \mathbb{N}$,

$$\sup_{P} \inf_{\hat{A}} \mathbb{E}_P\big\{\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}(\mathcal{D})\big)\big\} \geq 1-\alpha,$$

where the sup is over $P \in \mathcal{P}' = \mathcal{P}_M(d,\sigma) \cap \mathcal{P}_{\mathrm{R}}(d,\tau,\beta,\nu)$ and the inf is over procedures $\hat{A}$ which control the Type 1 error at the level α over $\mathcal{P}'$.

In essence, to bound $R_\tau(\hat{A}) := \mathbb{E}_P\big\{\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}\big)\big\}$ we must rule out the possibility that the marginal μ is concentrated on a large antichain.

Power bounds in the multi-variate setting

Given $d \in \mathbb{N}$, $\tau \in \mathbb{R}$, $\theta > 1$, $\gamma > 0$ and $\lambda \in (0,1)$, we let $\mathcal{P}_{\mathrm{Reg}}(d,\tau,\theta,\gamma,\lambda)$ denote the class of all distributions P on $\mathbb{R}^d \times \mathbb{R}$ with marginal μ on $\mathbb{R}^d$ and associated regression function η such that

(i) $\theta^{-1} \cdot r^d \leq \mu\big(B_r(x)\big) \leq \theta \cdot (2r)^d$ for all $x \in \mathcal{X}_\tau(\eta) \cap \mathrm{supp}(\mu)$ and $r \in (0,1]$;

(ii) $B_r(x) \cap \mathcal{X}_{\tau+\lambda\cdot r^\gamma}(\eta) \neq \emptyset$ for all $x \in \mathcal{X}_\tau(\eta) \cap \mathrm{supp}(\mu)$ and $r \in (0,1]$.

Power bounds in the multi-variate setting

Given $d \in \mathbb{N}$, $\tau \in \mathbb{R}$, $\theta > 1$, $\gamma > 0$ and $\lambda \in (0,1)$, we let $\mathcal{P}_{\mathrm{Reg}}(d, \tau, \theta, \gamma, \lambda)$ denote the class of all distributions P on $\mathbb{R}^d \times \mathbb{R}$ with marginal μ on $\mathbb{R}^d$ and associated regression function η such that

(i) $\theta^{-1} \cdot r^d \leq \mu\big(B_r(x)\big) \leq \theta \cdot (2r)^d$ for all $x \in \mathcal{X}_\tau(\eta) \cap \mathrm{supp}(\mu)$ and $r \in (0,1]$;

(ii) $B_r(x) \cap \mathcal{X}_{\tau + \lambda \cdot r^\gamma}(\eta) \neq \emptyset$ for all $x \in \mathcal{X}_\tau(\eta) \cap \mathrm{supp}(\mu)$ and $r \in (0,1]$.

The first condition ensures that μ is genuinely d-dimensional.

The second controls the way in which η grows around the τ-boundary.

Power bounds in the multi-variate setting

Theorem. Let $d \in \mathbb{N}$, $\tau \in \mathbb{R}$, $\sigma, \gamma > 0$, $\theta > 1$ and $\lambda \in (0,1)$. There exists $C \geq 1$, depending only on (d, θ), such that for any $P \in \mathcal{P}_{\mathrm{M}}(d,\sigma) \cap \mathcal{P}_{\mathrm{Reg}}(d,\tau,\theta,\gamma,\lambda)$, $n \in \mathbb{N}$, $\alpha \in (0,1)$ and $\mathcal{D} = \big((X_1,Y_1),\ldots,(X_n,Y_n)\big) \sim P^n$, we have for $m \in [n]$ that

$$\mathbb{E}_P\big\{\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}_{\mathrm{ISS}}(\mathcal{D})\big)\big\}$$
$$\leq 1 \wedge C\biggl\{\biggl(\frac{\sigma^2}{n\lambda^2}\log_+\Bigl(\frac{m\log_+ n}{\alpha}\Bigr)\biggr)^{1/(2\gamma+d)} + \biggl(\frac{\log_+ m}{m}\biggr)^{1/d}\biggr\}.$$

Power bounds in the multi-variate setting

Theorem. Let $d \in \mathbb{N}$, $\tau \in \mathbb{R}$, $\sigma, \gamma > 0$, $\theta > 1$ and $\lambda \in (0,1)$. There exists $C \geq 1$, depending only on (d, θ), such that for any $P \in \mathcal{P}_{\mathrm{M}}(d,\sigma) \cap \mathcal{P}_{\mathrm{Reg}}(d,\tau,\theta,\gamma,\lambda)$, $n \in \mathbb{N}$, $\alpha \in (0,1)$ and $\mathcal{D} = \big((X_1,Y_1),\ldots,(X_n,Y_n)\big) \sim P^n$, we have for $m \in [n]$ that

$$\mathbb{E}_P\big\{\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}_{\mathrm{ISS}}(\mathcal{D})\big)\big\} \\ \leq 1 \wedge C\bigg\{\bigg(\frac{\sigma^2}{n\lambda^2}\log_+\Big(\frac{m\log_+ n}{\alpha}\Big)\bigg)^{1/(2\gamma+d)} + \bigg(\frac{\log_+ m}{m}\bigg)^{1/d}\bigg\}.$$

Moreover, if we take $m_0 := n \wedge \lceil n\lambda^2/\sigma^2 \rceil$, then

$$\mathbb{E}_P\big\{\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}_{\mathrm{ISS}}(\mathcal{D})\big)\big\} \\ \leq 1 \wedge 4C\bigg\{\bigg(\frac{\sigma^2}{n\lambda^2}\log_+\Big(\frac{n\lambda^2\log_+ n}{\sigma^2\alpha}\Big)\bigg)^{1/(2\gamma+d)} + \bigg(\frac{\log_+ n}{n}\bigg)^{1/d}\bigg\}.$$

Choice of multiple testing procedure

We chose to combine our p-values with multiple testing procedure for DAGs which leverages an auxillary sparsified polyforest (Sparse DAG MTP).

We could also control the Type 1 error by combining p-values with any multiple testing which controls the family wise error.

For example, Meijer and Goeman's all-parent method, Meijer and Goeman's any-parent method, or even the classical Holm procedure.

We conduct a simulation study to compare these various choices of multiple testing procedure.

Choice of multiple testing procedure

We chose to combine our p-values with multiple testing procedure for DAGs which leverages an auxillary sparsified polyforest (Sparse DAG MTP).

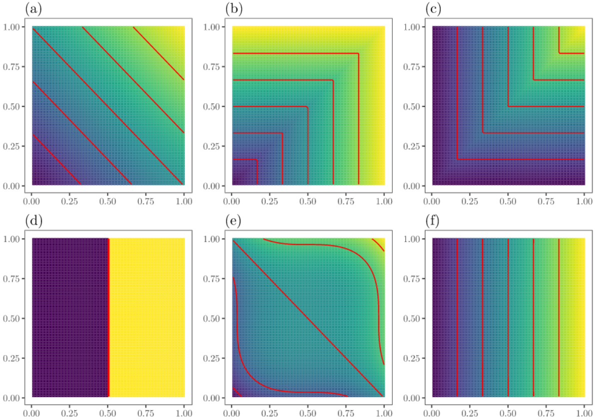
We conduct a simulation study to compare with other choices of multiple testing procedure.

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)	$\mathbb{1}_{(0.5,1]}\left(x^{(1)}\right)$	$1/2$	0
(e)	$\sum_{j=1}^{d}\left(x^{(j)}-0.5\right)^3$	$1/2$	3
(f)	$x^{(1)}$	$1/2$	1

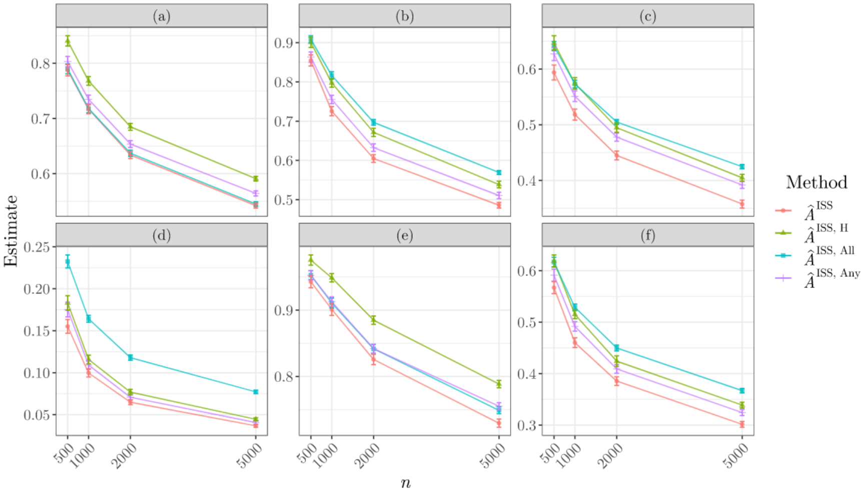
Our regression functions η are obtained by rescaling f.

Choice of multiple testing procedure

We conduct a simulation study to compare with other choices of multiple testing procedure.



Choice of multiple testing procedure



Estimated regret $\mathbb{E}_P\big\{\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}_{\mathrm{MG}}\big)\big\}$

Choice of multiple testing procedure

Recall that for $P \in \mathcal{P}_{\mathrm{M}}(d,\sigma) \cap \mathcal{P}_{\mathrm{Reg}}(d,\tau,\theta,\gamma,\lambda)$, the procedure $\hat{A}_{\mathrm{ISS}}$ achieves

$$\mathbb{E}_P\bigl\{\mu\bigl(\mathcal{X}_\tau(\eta) \setminus \hat{A}_{\mathrm{ISS}}(\mathcal{D})\bigr)\bigr\} \\ \leq 1 \wedge 4C\Biggl\{\biggl(\frac{\sigma^2}{n\lambda^2}\log_+\Bigl(\frac{n\lambda^2\log_+ n}{\sigma^2\alpha}\Bigr)\biggr)^{1/(2\gamma+d)} + \biggl(\frac{\log_+ n}{n}\biggr)^{1/d}\Biggr\}.$$

Choice of multiple testing procedure

Recall that for $P \in \mathcal{P}_{\mathrm{M}}(d,\sigma) \cap \mathcal{P}_{\mathrm{Reg}}(d,\tau,\theta,\gamma,\lambda)$, the procedure $\hat{A}_{\mathrm{ISS}}$ achieves

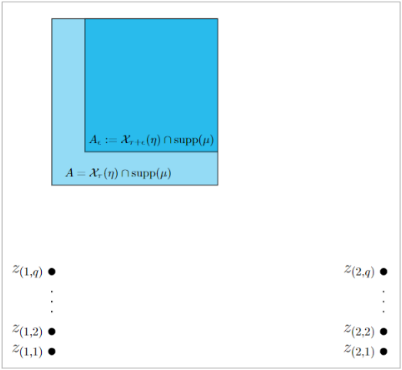
$$\begin{aligned}&\mathbb{E}_P\bigl\{\mu\bigl(\mathcal{X}_\tau(\eta) \setminus \hat{A}_{\mathrm{ISS}}(\mathcal{D})\bigr)\bigr\} \\ &\qquad \leq 1 \wedge 4C\biggl\{\biggl(\frac{\sigma^2}{n\lambda^2}\log_+\Bigl(\frac{n\lambda^2\log_+ n}{\sigma^2\alpha}\Bigr)\biggr)^{1/(2\gamma+d)} + \biggl(\frac{\log_+ n}{n}\biggr)^{1/d}\biggr\}.\end{aligned}$$

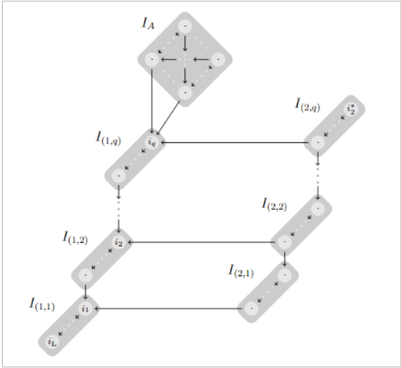
Proposition. Suppose $d \geq 2$, $\tau \in \mathbb{R}$, $\sigma, \gamma > 0$, $\lambda \in (0,1)$, $\theta \in [2^d, \infty)$, $\alpha \in (0, 1/4]$. Let $\hat{A}_{\mathrm{MG}}$ denote the data-dependent subgroup obtained via either the all-parent or the any-parent MTP of Meijer and Goeman (2015). There exists $c > 0$, depending only on d, α, σ, λ and γ, such that for every $n \in \mathbb{N}$,

$$\min_{m\in[n]} \sup_{P\in\mathcal{P}'} \mathbb{E}_P\bigl\{\mu\bigl(\mathcal{X}_\tau(\eta) \setminus \hat{A}_{\mathrm{MG}}\bigr)\bigr\} \geq \frac{c}{n^{1/(2\gamma+d+1)}(\log_+ n)^{2/d}},$$

where the sup is over $\mathcal{P}' := \mathcal{P}_{\mathrm{M}}(d,\sigma) \cap \mathcal{P}_{\mathrm{Reg}}(d,\tau,\theta,\gamma,\lambda)$.

Choice of multiple testing procedure





$$\min_{m \in [n]} \sup_{P \in \mathcal{P}'} \mathbb{E}_P\big\{\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}_{\mathrm{MG}}\big)\big\} \geq \frac{c}{n^{1/(2\gamma+d+1)}(\log_+ n)^{2/d}},$$

Minimax optimality

Theorem. Let $d \in \mathbb{N}$, $\tau \in \mathbb{R}$, $\sigma, \gamma > 0$, $\theta > 1$ and $\lambda \in (0,1)$. Then, there exists $c \in (0,1)$, depending only on (d, γ), such that for any $n \in \mathbb{N}$ and $\alpha \in (0, 1/4]$,

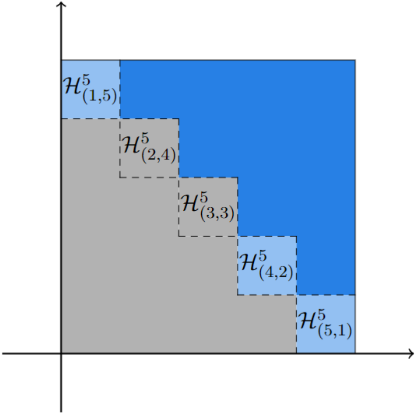
$$\inf_{\hat{A}} \sup_{P} \mathbb{E}_P \bigl\{ \mu\bigl(\mathcal{X}_\tau(\eta) \setminus \hat{A}(\mathcal{D})\bigr) \bigr\} \geq c \left[1 \wedge \left\{ \left(\frac{\sigma^2}{n\lambda^2} \log_+\Bigl(\frac{1}{5\alpha}\Bigr) \right)^{1/(2\gamma+d)} + \frac{1}{n^{1/d}} \right\} \right].$$

where the sup is over $P \in \mathcal{P}' := P_{\mathrm{M}}(d, \sigma) \cap \mathcal{P}_{\mathrm{Reg}}(d, \tau, \theta, \gamma, \lambda)$ and the inf is over procedures $\hat{A}$ which control the Type 1 error at the level α over $\mathcal{P}'$.

Recall that for $P \in \mathcal{P}_{\mathrm{M}}(d, \sigma) \cap \mathcal{P}_{\mathrm{Reg}}(d, \tau, \theta, \gamma, \lambda)$, the procedure $\hat{A}_{\mathrm{ISS}}$ achieves

$$\begin{aligned} &\mathbb{E}_P \bigl\{ \mu\bigl(\mathcal{X}_\tau(\eta) \setminus \hat{A}_{\mathrm{ISS}}(\mathcal{D})\bigr) \bigr\} \\ &\quad \leq 1 \wedge 4C \left\{ \left(\frac{\sigma^2}{n\lambda^2} \log_+\Bigl(\frac{n\lambda^2 \log_+ n}{\sigma^2 \alpha}\Bigr) \right)^{1/(2\gamma+d)} + \left(\frac{\log_+ n}{n} \right)^{1/d} \right\}. \end{aligned}$$

Minimax optimality

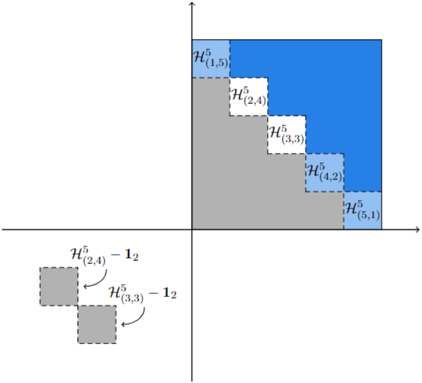


The first component of the lower bound corresponds to the difficulty of determining whether a given covariate X_i is within $\mathcal{X}_\tau(\eta)$.

Note the dependence upon α.

$$\inf_{\hat{A}} \sup_{P} \mathbb{E}_P\{\mu(\mathcal{X}_\tau(\eta)\eta \setminus \hat{A}(\mathcal{D}))\} \geq c_0 \cdot \left\{1 \wedge \left(\frac{\sigma^2}{n\lambda^2} \log_+\left(\frac{1}{5\alpha}\right)\right)^{1/(2\gamma+d)}\right\}.$$

Minimax optimality



The second component of the lower bound corresponds to the error incurred due to missing regions of the covariate space.

$$\inf_{\hat{A}} \sup_{P} \mathbb{E}_P\{\mu(\mathcal{X}_\tau(\eta)\eta \setminus \hat{A}(\mathcal{D}))\} \geq c_1 \cdot \left\{1 \wedge \left(\frac{\log_+ n}{n}\right)^{1/d}\right\}.$$

Theorem. Let $d \in \mathbb{N}$, $\tau \in \mathbb{R}$, $\sigma, \gamma > 0$, $\theta > 1$ and $\lambda \in (0,1)$. Then, there exists $c \in (0,1)$, depending only on (d,γ), such that for any $n \in \mathbb{N}$ and $\alpha \in (0, 1/4]$,

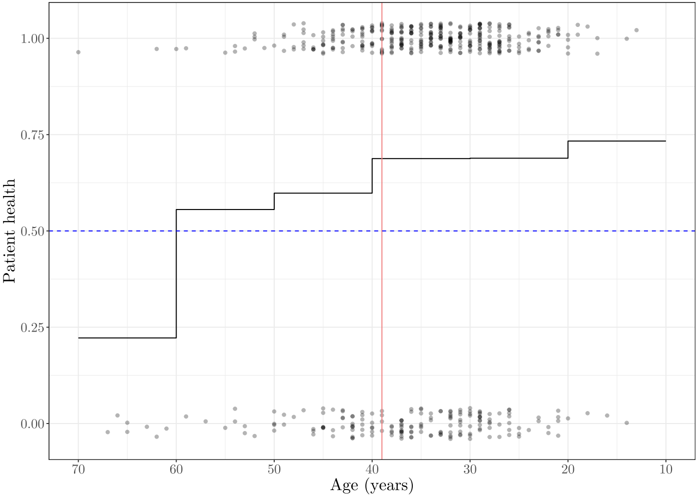
$$\inf_{\hat{A}} \sup_{P} \mathbb{E}_P\Big\{\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}(\mathcal{D})\big)\Big\} \geq c\Bigg[1 \wedge \Bigg\{\bigg(\frac{\sigma^2}{n\lambda^2}\log_+\Big(\frac{1}{5\alpha}\Big)\bigg)^{1/(2\gamma+d)} + \frac{1}{n^{1/d}}\Bigg\}\Bigg].$$

where the sup is over $P \in \mathcal{P}' := P_{\mathrm{M}}(d,\sigma) \cap \mathcal{P}_{\mathrm{Reg}}(d,\tau,\theta,\gamma,\lambda)$ and the inf is over procedures $\hat{A}$ which control the Type 1 error at the level α over $\mathcal{P}'$.

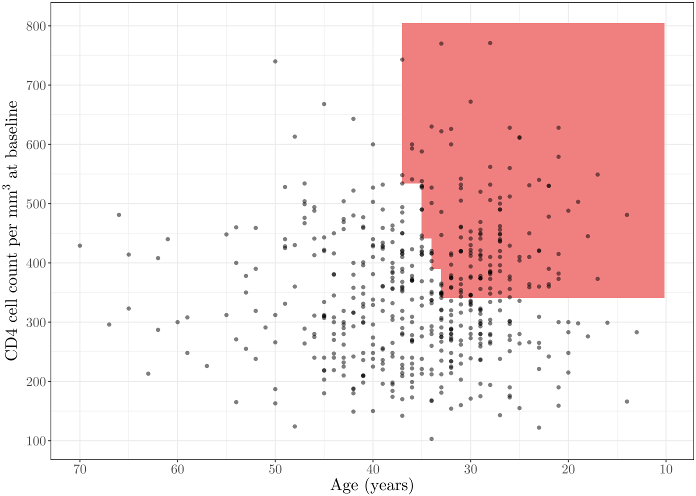
Recall that for $P \in \mathcal{P}_{\mathrm{M}}(d,\sigma) \cap \mathcal{P}_{\mathrm{Reg}}(d,\tau,\theta,\gamma,\lambda)$, the procedure $\hat{A}_{\mathrm{ISS}}$ achieves

$$\begin{aligned}&\mathbb{E}_P\Big\{\mu\big(\mathcal{X}_\tau(\eta) \setminus \hat{A}_{\mathrm{ISS}}(\mathcal{D})\big)\Big\} \\ &\qquad \leq 1 \wedge 4C\Bigg\{\bigg(\frac{\sigma^2}{n\lambda^2}\log_+\Big(\frac{n\lambda^2\log_+ n}{\sigma^2\alpha}\Big)\bigg)^{1/(2\gamma+d)} + \bigg(\frac{\log_+ n}{n}\bigg)^{1/d}\Bigg\}.\end{aligned}$$

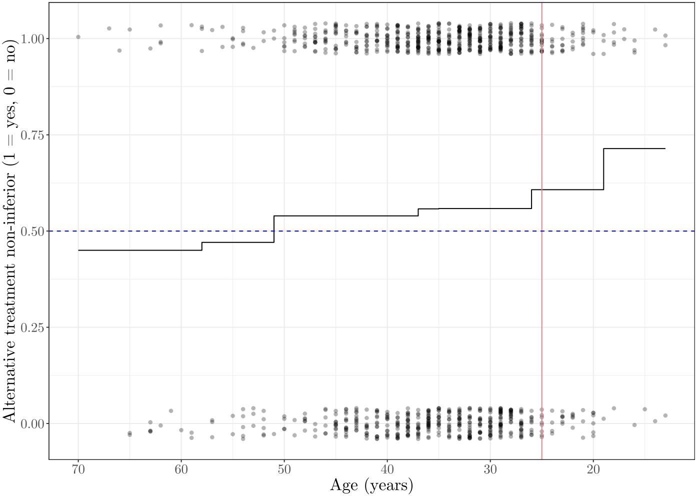
AIDS Clinical Trials Group Study 175



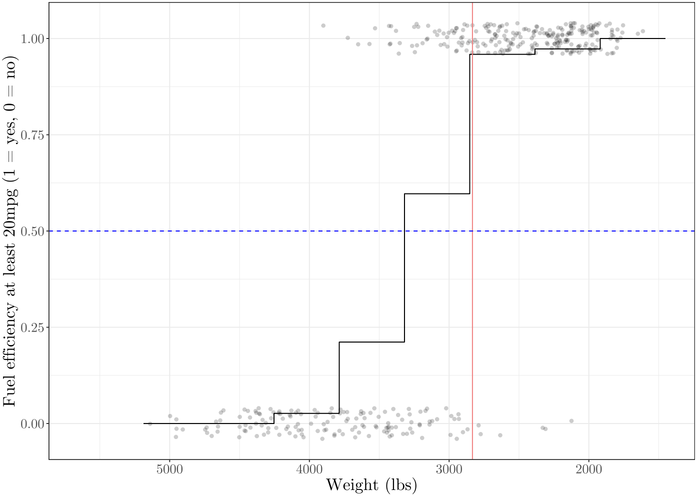
AIDS Clinical Trials Group Study 175



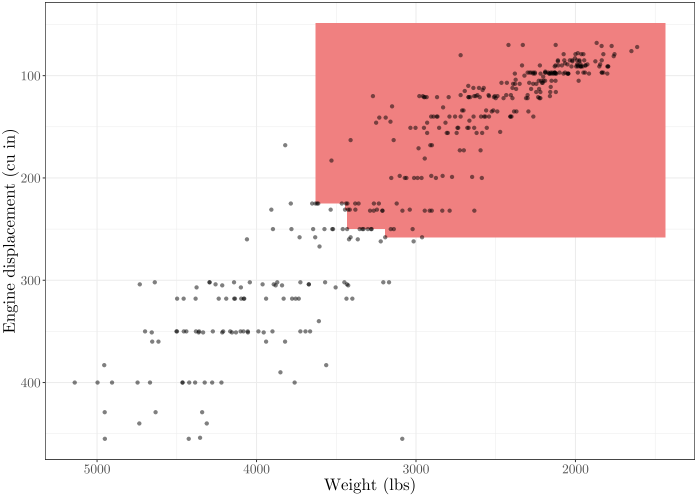
Heterogenous treatment effect



Fuel consumption dataset



Fuel consumption dataset



Summary

- We investigated subgroup selection in a non-parametric regime with a multivariate isotonic regression function.
- Our method controls Type 1 error by combining local p-values combined with multiple testing procedures.
- The choice of multiple testing procedure plays a crucial role in determining the power.
- Our regret bounds demonstrate minimax optimality up to poly-logarithmic factors under natural distributional assumptions.

Thank you for listening

References

Bretz, Maurer, Brannath, and Posch, “A graphical approach to sequentially rejective multiple test procedures” (2009).
Goeman and Solari. “The sequential rejection principle of familywise error control” (2010).

Howard, Ramdas, McAuliffe, and Sekhon. “Time-uniform, nonparametric, nonasymptotic confidence sequences.” (2021).

Lipkovich, Dmitrienko, and D’Agostino Sr. “Tutorial in biostatistics: data-driven subgroup identification and analysis in clinical trials” (2017).

Maurer, Hothorn, Lehmacher. “Multiple comparisons in drug clinical trials and preclinical assays: a priori ordered hypotheses.” (1995).

Meijer and Goeman. “A multiple testing method for hypotheses structured in a directed acyclic graph” (2015).

Mueller, Reeve, Cannings, and Samworth. “Isotonic subgroup selection” (2023), forthcoming.

Reeve, Cannings, and Samworth. “Minimax optimal subgroup selection,” (2021).

Wan, Liu and Bretz. “Confidence sets for a level set in linear regression,” (2022).