AdaPT: Interactive Multiple Testing with Side Information

Lihua Lei & Will Fithian

Department of Statistics, UC Berkeley

March 8, BSTARS 2018

Side Information

Setting: hypotheses $H_1, \ldots, H_n$ with p-values $p_1, \ldots, p_n$

Observe side information $x_i \in \mathcal{X}$ for each H_i

Examples:

- Identifying differentially expressed genes with "prioritization scores" (priors, auxiliary information, etc.)
- Spatiotemporal testing;
- Clinical meta-analysis;

Side Information

Setting: hypotheses $H_1, \ldots, H_n$ with p-values $p_1, \ldots, p_n$

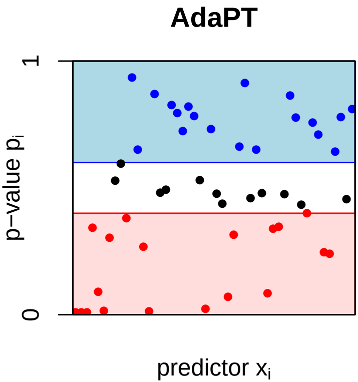
Observe side information $x_i \in \mathcal{X}$ for each H_i

Examples:

- Identifying differentially expressed genes with "prioritization scores" (priors, auxiliary information, etc.)
- Spatiotemporal testing;
- Clinical meta-analysis;

Goal: produce hypothesis-specific p-value thresholds s_i's and control False Discovery Rate (FDR).

AdaPT, Visualized



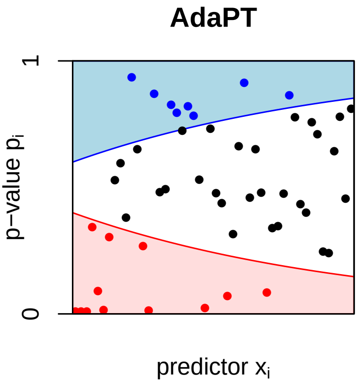
Step t:

Covariate-dependent threshold $s_t(x)$

Mirror image $1 - s_t(x)$

$$\widehat{\mathrm{FDP}}_t = \frac{\#\text{ blue points} + 1}{\#\text{ red points} \vee 1}$$

AdaPT, Visualized



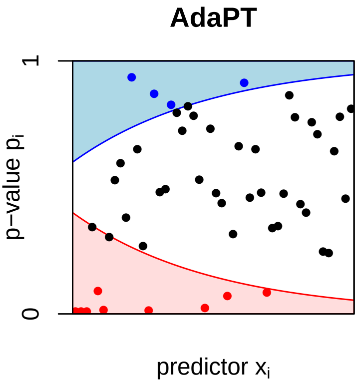
Step t:

Covariate-dependent threshold $s_t(x)$

Mirror image $1 - s_t(x)$

$$\widehat{\mathrm{FDP}}_t = \frac{\#\text{ blue points} + 1}{\#\text{ red points} \vee 1}$$

AdaPT, Visualized



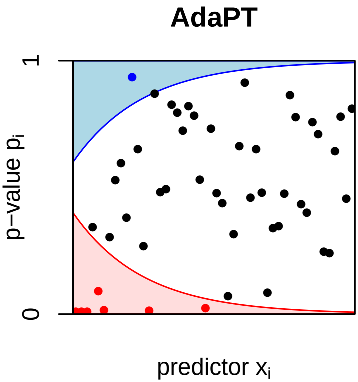
Step t:

Covariate-dependent threshold $s_t(x)$

Mirror image $1 - s_t(x)$

$$\widehat{\mathrm{FDP}}_t = \frac{\#\text{ blue points} + 1}{\#\text{ red points} \vee 1}$$

AdaPT, Visualized



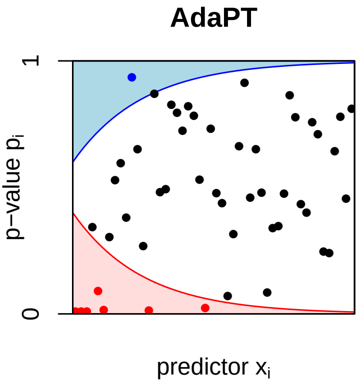
Step t:

Covariate-dependent threshold $s_t(x)$

Mirror image $1 - s_t(x)$

$$\widehat{\mathrm{FDP}}_t = \frac{\#\text{ blue points} + 1}{\#\text{ red points} \vee 1}$$

AdaPT, Visualized



Step t:

Covariate-dependent threshold $s_t(x)$

Mirror image $1 - s_t(x)$

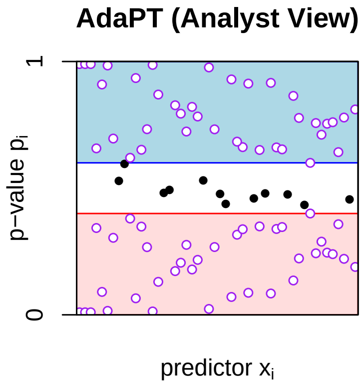
$$\widehat{\mathrm{FDP}}_t = \frac{\text{\# blue points} + 1}{\text{\# red points} \vee 1}$$

Stop when $\widehat{\mathrm{FDP}}_t \leq \alpha$, and reject all red points

AdaPT, "Analyst View"

Define partially masked p-values:

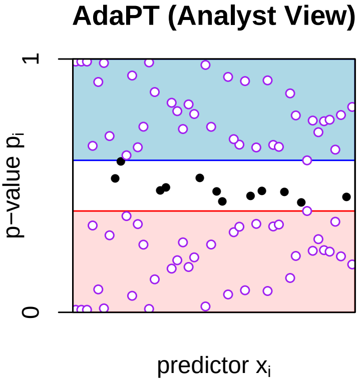
$$\tilde{p}_{t,i} = \begin{cases} p_i & s_t(x_i) < p_i < 1 - s_t(x_i) \\ \{p_i,\ 1 - p_i\} & \text{otherwise.} \end{cases}$$



AdaPT, "Analyst View"

Define partially masked p-values:

$$\tilde{p}_{t,i} = \begin{cases} p_i & s_t(x_i) < p_i < 1 - s_t(x_i) \\ \{p_i,\ 1 - p_i\} & \text{otherwise.} \end{cases}$$



To select $s_{t+1}(x)$, we can use:

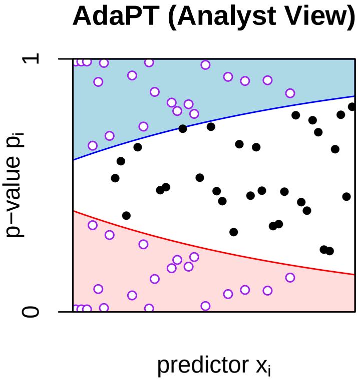
- $x_1, \ldots, x_n$
- $\tilde{p}_{t,1}, \ldots, \tilde{p}_{t,n}$

Any such update rule is OK

AdaPT, "Analyst View"

Define partially masked p-values:

$$\tilde{p}_{t,i} = \begin{cases} p_i & s_t(x_i) < p_i < 1 - s_t(x_i) \\ \{p_i,\ 1 - p_i\} & \text{otherwise.} \end{cases}$$



To select $s_{t+1}(x)$, we can use:

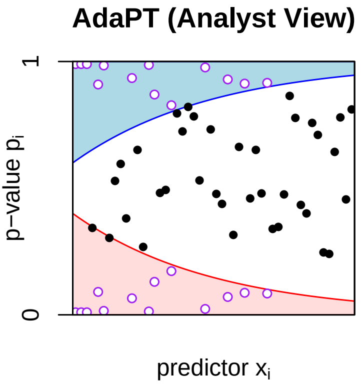
- $x_1, \ldots, x_n$
- $\tilde{p}_{t,1}, \ldots, \tilde{p}_{t,n}$

Any such update rule is OK

AdaPT, "Analyst View"

Define partially masked p-values:

$$\tilde{p}_{t,i} = \begin{cases} p_i & s_t(x_i) < p_i < 1 - s_t(x_i) \\ \{p_i,\ 1 - p_i\} & \text{otherwise.} \end{cases}$$



To select $s_{t+1}(x)$, we can use:

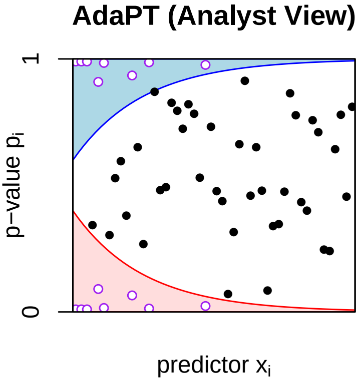
- $x_1, \ldots, x_n$
- $\tilde{p}_{t,1}, \ldots, \tilde{p}_{t,n}$

Any such update rule is OK

AdaPT, "Analyst View"

Define partially masked p-values:

$$\tilde{p}_{t,i} = \begin{cases} p_i & s_t(x_i) < p_i < 1 - s_t(x_i) \\ \{p_i,\ 1 - p_i\} & \text{otherwise.} \end{cases}$$



To select $s_{t+1}(x)$, we can use:

- $x_1, \ldots, x_n$
- $\tilde{p}_{t,1}, \ldots, \tilde{p}_{t,n}$

Any such update rule is OK

AdaPT: Finite-Sample FDR Control

Theorem 1 (Lei and Fithian, 2016).

*Under standard assumptions in literature, AdaPT controls FDR at level α, **regardless of the update rule**.*

AdaPT: Finite-Sample FDR Control

Theorem 1 (Lei and Fithian, 2016).

*Under standard assumptions in literature, AdaPT controls FDR at level α, **regardless of the update rule**.*

AdaPT is flexible in an unusually strong sense in that the update rule can be **arbitrary**,

AdaPT: Finite-Sample FDR Control

Theorem 1 (Lei and Fithian, 2016).

*Under standard assumptions in literature, AdaPT controls FDR at level α, **regardless of the update rule**.*

AdaPT is flexible in an unusually strong sense in that the update rule can be **arbitrary**,

- no matter how misspecified our model is;

AdaPT: Finite-Sample FDR Control

Theorem 1 (Lei and Fithian, 2016).

*Under standard assumptions in literature, AdaPT controls FDR at level α, **regardless of the update rule**.*

AdaPT is flexible in an unusually strong sense in that the update rule can be **arbitrary**,

- no matter how misspecified our model is;
- no matter how misguided our priors are (if we use a Bayesian method);

AdaPT: Finite-Sample FDR Control

Theorem 1 (Lei and Fithian, 2016).

*Under standard assumptions in literature, AdaPT controls FDR at level α, **regardless of the update rule**.*

AdaPT is flexible in an unusually strong sense in that the update rule can be **arbitrary**,

- no matter how misspecified our model is;
- no matter how misguided our priors are (if we use a Bayesian method);
- no matter how we select a model or tuning parameter;

AdaPT: Finite-Sample FDR Control

Theorem 1 (Lei and Fithian, 2016).

*Under standard assumptions in literature, AdaPT controls FDR at level α, **regardless of the update rule**.*

AdaPT is flexible in an unusually strong sense in that the update rule can be **arbitrary**,

- no matter how misspecified our model is;
- no matter how misguided our priors are (if we use a Bayesian method);
- no matter how we select a model or tuning parameter;
- no matter how crazy we are.

AdaPT: Finite-Sample FDR Control

Theorem 1 (Lei and Fithian, 2016).

*Under standard assumptions in literature, AdaPT controls FDR at level α, **regardless of the update rule**.*

AdaPT is flexible in an unusually strong sense in that the update rule can be **arbitrary**,

- no matter how misspecified our model is;
- no matter how misguided our priors are (if we use a Bayesian method);
- no matter how we select a model or tuning parameter;
- no matter how crazy we are.

Don't worry. Use your favorite model to guide the update!

THANKS!