False Discovery Rate Control using Covariates

Kyurhi Kim

Seoul National University

November 12, 2023

Contents

False Discovery Rate

Multiple testing

- conduct m hypothesis tests simultaneously

$$H_{0i} \quad \text{vs} \quad H_{1i}, \qquad i = 1, 2, \ldots, m$$

- **Issue** : How can we control type I error in multiple testing?

	H0 retained	H0 rejected	Total
Actual H0	TN	FP	$T0$
Actual H1	FN	TP	$T1$
Total	N	P	m

Multiple testing

- **FWER** : $\mathbb{P}(\mathrm{FP} > 0)$
  - Bonferroni gaurantees $\mathbb{P}(\mathrm{FP} > 0) \leq \alpha$
  - the resulting thresholds often suffer from low power
- **FDR** control is suggested to increase power while maintaining some threshold on error

False Discovery Rate

- **False Discovery Proportion (FDP)**: proportion of false discoveries among total rejections

$$\text{FDP(t)} = \frac{\text{number of False Positives}}{\text{number of Positives}} = \frac{FP}{P}$$

- **False Discovery Rate (FDR)**: expected FDP

$$\text{FDR} = \mathbb{E}\left(\text{FDP(t)}\right)$$

Benjamini-Hochberg Procedure

BH procedure step

1. Sort the p-values and give the smallest p-value rank 1: $p_1, p_2, ..., p_m$
2. Compute BH-critical value $(\frac{i}{m}\alpha)$ for each p-value

   alpha: desired false discovery rate

   i: the rank

   m: the total number of p-values
3. Find the largest i for which the p-value is less than the corresponding critical value
4. Let k be the largest i s.t. $p_{(i)} \leq \frac{i}{m}\alpha$, then reject hypotheses 1,...,k

Local FDR

Local FDR

- From a Bayesian perspective, it is natural to conditioning on the actual value of z, in other words

$$\mathbb{P}(H_{0j}|z_j = z)$$

- Efron et al(2001), Newton et al. proposed the local false discovery rate as

$$\mathbb{P}(H_{0j}|z_j = z) = \frac{\pi_0 f_0(z_j)}{f(z_j)}$$

where $f(z) = \pi_0 f_0(z) + \pi_1 f_1(z)$ is the marginal density of z-values and $f_0(z)$ is the null density, and π_0 is proportion of genes that are not differently expressed

Local FDR

- posterior probability of gene j being not differentially expressed

$$\begin{aligned}\tau_0(z_j) &= \mathbb{P}(\text{jth gene is null}|z_j) \\ &= \frac{\pi_0 f_0(z_j)}{f(z_j)} \\ &= \frac{\pi_0 f_0(z_j)}{\pi_0 f_0(z_j) + (1-\pi_0) f_1(z_j)}\end{aligned}$$

Local FDR

local fdr procedure

1. Obtain the z-score for each of the genes

$$z_j = \Phi^{-1}(1 - p_j)$$

2. Rank the genes on the basis of the z-scores, starting with the largest ones
3. The posterior probability of non-differential expresison of gene j, is given by $\tau_0(z_j)$
4. Conclude gene j to be differentially expressed if $\widehat{\tau_0}(z_j) < c_0$

FDR control using covariates

FDR control using covariates

- for each hypothesis H_i, $i \in [n]$ observe not only a p-value $p_i \in [0, 1]$ but also a predictor x_i lying in some generic space χ
- Unlike p_i, x_i carries only indirect information about the hypothesis i.e. capture some side information

AdaPT: *adaptive p-value thresholding*

- AdaPT(*adaptive p-value thresholding*) is a iterative method for FDR control with general side information

Algorithm 1 AdaPT

Input: predictors and p-values $(x_i, p_i)_{i\in[n]}$, initialization s_0, target FDR level α

Procedure:

for $t = 0, 1, \ldots$ **do**
  $\widehat{\mathrm{FDP}}_t \leftarrow \frac{1+A_t}{R_t \vee 1}$;
  if $\widehat{\mathrm{FDP}}_t \leq \alpha$ **then**
    Reject $\{H_i : p_i \leq s_t(x_i)\}$;
    Return s_t;
  end if
  $s_{t+1} \leftarrow \mathrm{UPDATE}\left((x_i, \tilde{p}_{t,i})_{i\in[n]}, A_t, R_t, s_t\right)$;
end for

Figure: algorithm of AdaPT

AdaPT procedure

- For each step t=0,1,2,...
  consider rejection threshold at step t: $s_t(x)$, and compute:

$$R_t = |i : p_i \leq s_t(x_i)|$$
$$A_t = |i : p_i \geq 1 - s_t(x_i)|$$
$$\hat{\text{FDP}}_t = \frac{1 + A_t}{R_t \vee 1}$$

- If $\hat{\text{FDP}}_t \leq \alpha$, procedure stop and the set of R_t is returned
  (i.e. reject $\{H_i : p_i \leq s_t(x_i)\}$)
- Otherwise, update the rejection threshold under two protocols , and continue
  1. rejection threshold must be more stringent : $s_{t+1}(x_i) \leq s_t(x_i)$
  2. partially mask the p-values determining R_t and A_t

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

AdaPT procedure

- To select $s_{t+1}(x)$, $\begin{cases} x_1, x_2, ..., x_n \\ \widetilde{p}_{t,1}, ..., \widetilde{p}_{t,n} \\ A_t, R_t \end{cases}$

  any such update rule is OK

- AdaPT procedure repeats by estimating FDP and updating the threshold until the target FDR level is reached: $\hat{\text{FDP}}_t \leq \alpha$ or $R_t = 0$

Visualize AdaPT

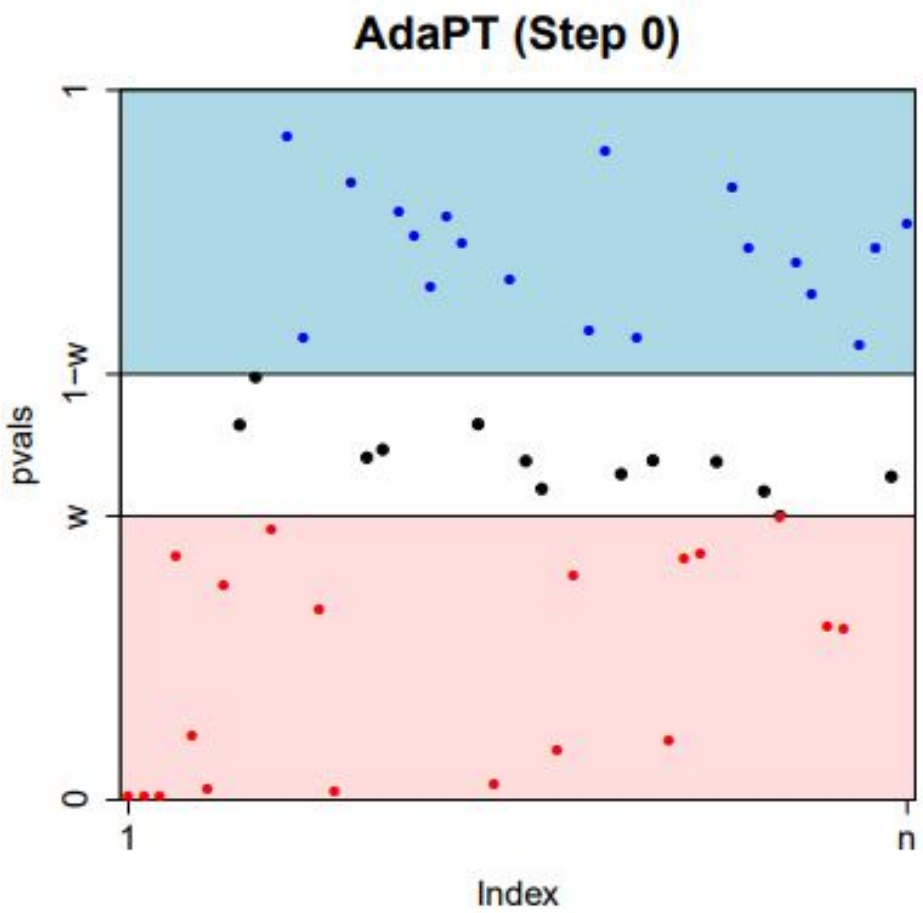

Figure: AdaPT(step0)

Visualize AdaPT

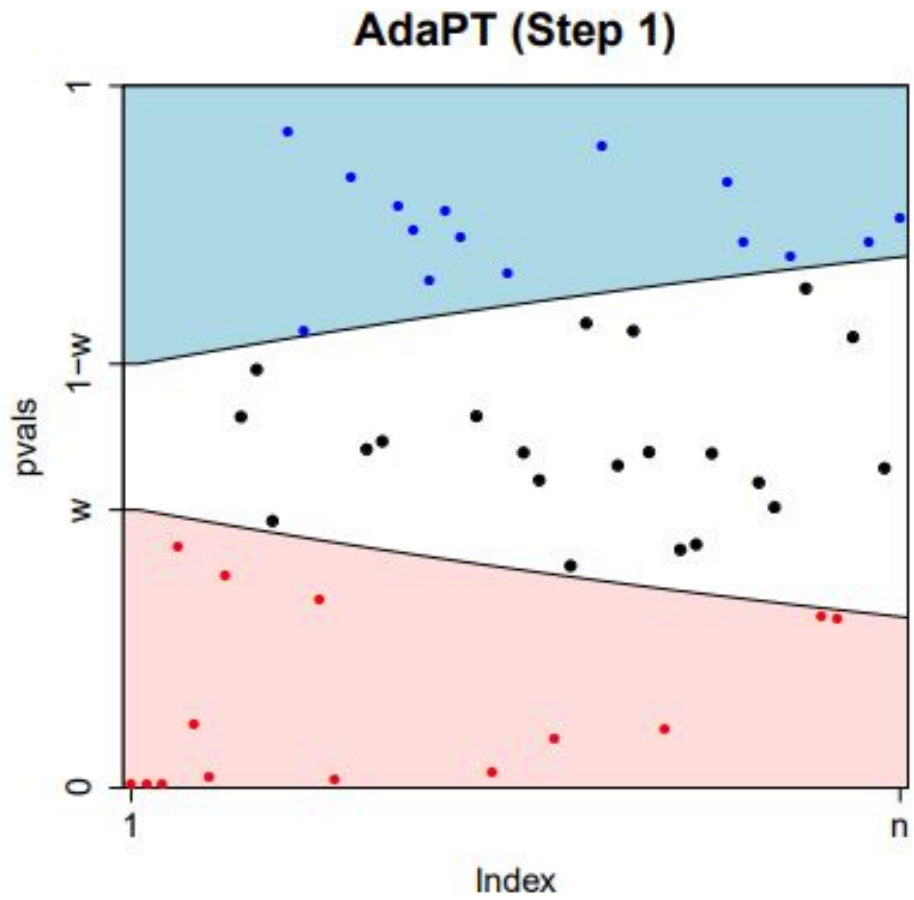

Figure: AdaPT(step1)

Visualize AdaPT

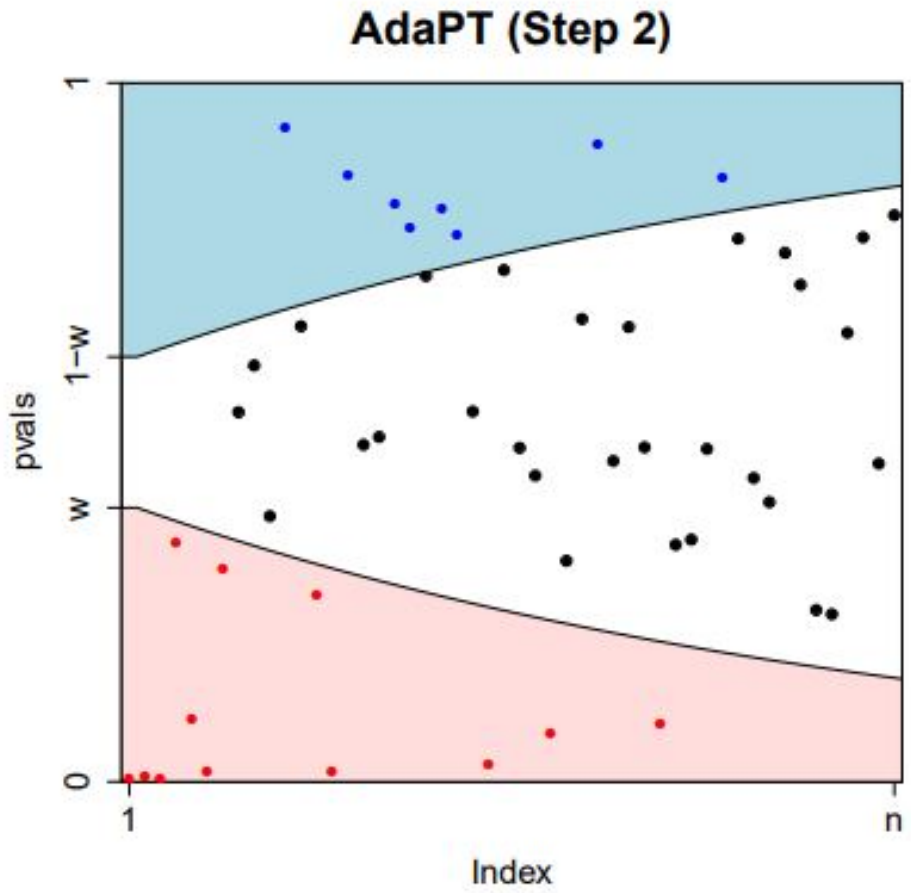

Figure: AdaPT(step2)

Visualize AdaPT

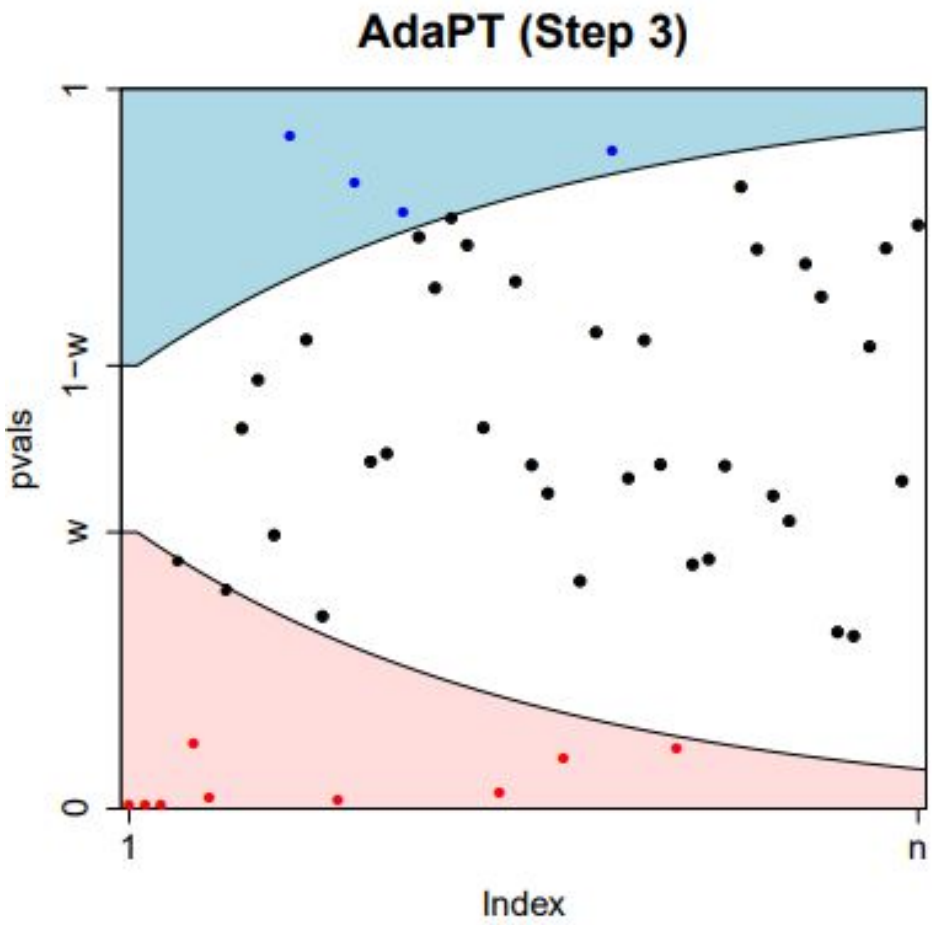

Figure: AdaPT(step3)

Visualize AdaPT

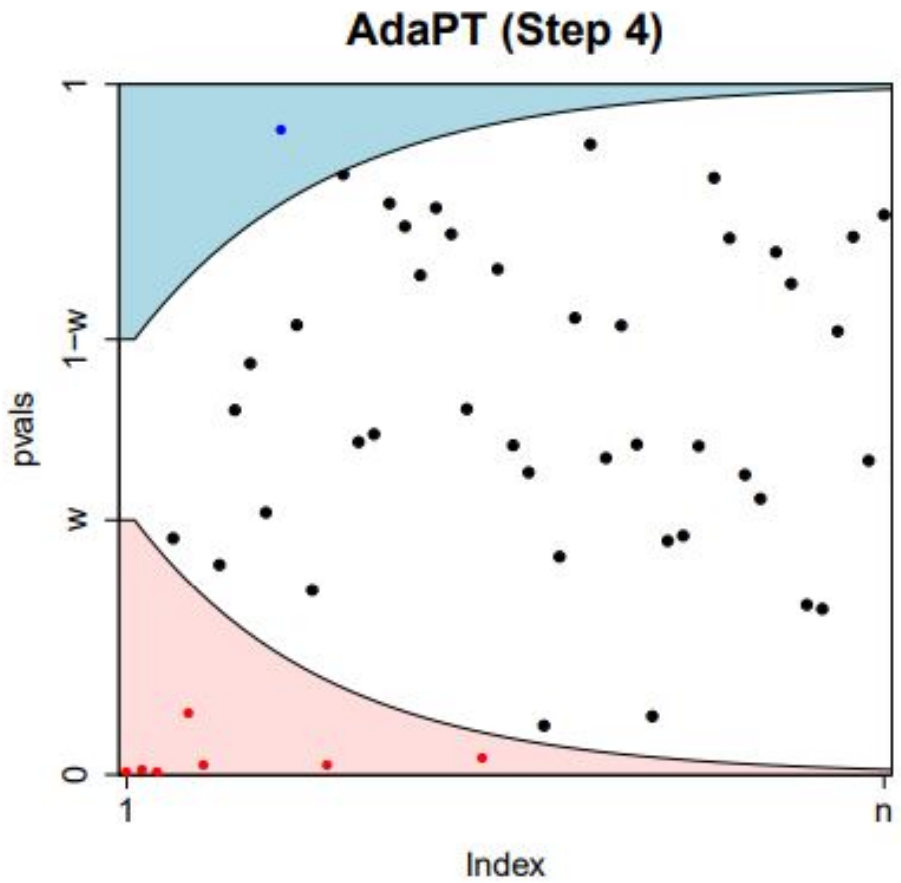

Figure: AdaPT(step4)

AdaPT

Theorem1 (Lei and F, 2016)

Assume that

1. null p-values are independent of each other and of the non-null p-values (i.e. $\{p_i : i = 1, ..., n\}$ are independent)
2. the null p-values are uniform or mirror conservative

Then the AdaPT procedure controls the FDR at level α,
conditional on $(x_i)_{i=1}^n$ and $(p_i)_{i \notin H_0}$

AdaPT: How to choose thresholding rules?

- Although AdaPT controls FDR at level α, its power depends on the quality of the updates
- Assume a two groups model conditional on the predictors x_i, and we assume:

$$H_i|x_i \sim \text{Bernoulli}(\pi_1(x_i))$$

$$p_i|H_i, x_i \sim \begin{cases} f_0(p|x_i) \text{ (i.e. Unif)}, & \text{if } H_i = 0 \\ f_1(p|x_i), & \text{if } H_i = 1 \end{cases}$$

- Also, assume that (x_i, H_i, p_i) are independent for $i \in [n]$

Thresholding rules: two-groups model and local fdr

- define the conditional mixture density

$$f(p \mid x) = (1 - \pi_1(x))f_0(p \mid x) + \pi_1 f_1(p \mid x) = 1 - \pi_1(x) + \pi_1(x)f_1(p \mid x)$$

  and define the conditional local fdr

$$fdr(p \mid x) = \mathbb{P}(H_i \text{is null} \mid x_i = x, p_i = p) = \frac{1 - \pi_1(x)}{f(p \mid x)}$$

- Unless $f_1(p \mid x)$ is known a priori, it is possible to make the conservative identifying asumption that

$$1 - \pi_1(x) = \inf_{p \in [0,1]} f(p \mid x) = f(1 \mid x)$$

  attributing as many observations as possible to the null hypothesis
- then, estimate

$$\widehat{fdr}(p \mid x) = \frac{\widehat{f}(1 \mid x)}{\widehat{f}(p \mid x)}$$

optimal threshold under the two groups model

Theorem2 (Lei and F, 2016)

Assume that

1. $f_1(p \mid x_i)$ is continuously non-increasing, $f_0(p \mid x_i)$ is continuously non-decreasing, and uniformly bounded away from ∞
2. probability measure ν on χ is a discrete measure supported on $x_1, ..., x_n$ with $\nu(x_i : fdr(0 \mid x_i) < \alpha, fdr(0 \mid x_i) > 0) > 0$

Then $\max_s Pow(s; \nu)$ s.t. $\text{FDR}(s; \nu) \leq \alpha$ has at least a solution,
and all solutions are level surfaces of local FDR

(note that Pow(s;ν) $= \mathbb{P}(\text{H is rejected} \mid \text{H} = 1) = \mathbb{P}(P \leq s(X) \mid \text{H} = 1)$)

EM to estimate $\pi_1(\cdot)$ and $\mu(\cdot)$ based on $D_t = (x_i, \widetilde{p}_{t,i})$

- To estimate local FDR, we need to estimate $\widehat{f}(p \mid x)$

Algorithm 2 EM algorithm to estimate $\pi_1(\cdot)$ and $\mu(\cdot)$ based on $D_t = (x_i, \tilde{p}_{t,i})_{i\in[n]}$

Input: data D_t, number of iterations m, initialization $\hat{\theta}^{(0)}, \hat{\beta}^{(0)}$;

for $r = 1, 2, \ldots, m$ **do**

$(E\text{-}step)$:

$$\widehat{H}_i^{(r)} \leftarrow \mathbb{E}_{\hat{\theta}^{(r-1)},\hat{\beta}^{(r-1)}}[H_i \mid D_t], \quad i \in [n];$$

$$\hat{y}_i^{(r,1)} \leftarrow \mathbb{E}_{\hat{\theta}^{(r-1)},\hat{\beta}^{(r-1)}}[y_i H_i \mid D_t, H_i = 1]/\widehat{H}_i^{(r)}, \quad i \in [n];$$

$(M\text{-}step)$:

$$\hat{\theta}^{(r)} \leftarrow \texttt{glm}\left(\widehat{H}^{(r)} \sim \phi_\pi(x),\ \texttt{family = binomial}\right);$$

$$\hat{\beta}^{(r)} \leftarrow \texttt{glm}\left(\hat{y}^{(r,1)} \sim \phi_\mu(x),\ \texttt{family = ...(link} = \zeta\texttt{)},\ \texttt{weights = } \widehat{H}^{(r)}\right);$$

end for

Output: $\hat{\pi}_1(x) = \left(1 + e^{-\phi_\pi(x)'\hat{\theta}^{(m)}}\right)^{-1}, \quad \hat{\mu}(x) = \zeta^{-1}\left(\phi_\mu(x)'\hat{\beta}^{(m)}\right).$

Figure: EM algorithm

Updating the threshold

- For the previous model, level surfaces of the local FDR are given by

$$c = \frac{f(1|x)}{f(s(x)|x)} = \frac{\pi_1(x)h(1;\mu(x)) + 1 - \pi_1(x)}{\pi_1(x)h(s(x);\mu(x)) + 1 - \pi_1(x)}$$

- For various widely-used exponential families, $h(p;\mu)$ is decreasing w.r.t. p, in which case,

$$s(x;c) = f^{-1}(\frac{h(1;\mu(x))}{c} + \frac{1-\pi_1(x)}{\pi_1(x)}\frac{1-c}{c};\mu(x))$$

- Given a chosen local FDR level c, s_t can be evolved by $s_{t+1} = \min(s_t(x), s(x;c))$
- Choice of c can be computed in the following way
  - Estimate local FDR for each $p'_{t,i}$ where $p'_{t,i}$ is the minimum element in $\widetilde{p}_{t,i}$)
  - Set c as the largest value of lfdr among all partially masked p-values

applying AdaPT

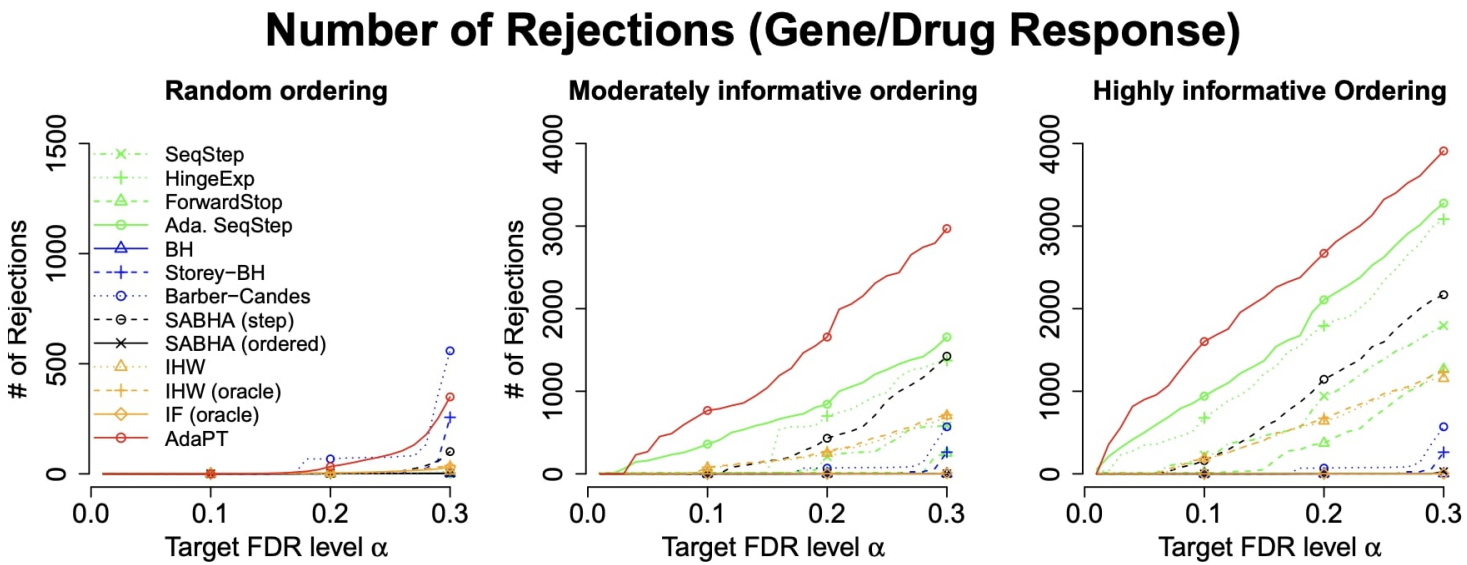

Figure: AdaPT

AdaPT summary and limitation

- AdaPT: iteratively estimates the p-value thresholds using partially censored p-values
- However, AdaPT is computationally intensive and may suffer from significant power loss when the signal is sparse, and covariate is not very informative
- Zhang and Chen, 2020, proposed a new procedure to incorporate covariate information: Covariate Adaptive Multiple Testing (CAMT)

CAMT

- Consider m hypotheses $H_i i = 1, 2, .., m$ and corresponding p_i
- Consider two-component mixture model

$$H_i|x_i \sim \text{Bernoulli}(1 - \pi_{0i})$$
$$p_i|H_i, x_i \sim (1 - H_i)f_0 + H_i f_{1,i}$$

- Optimal Rejection rule based on local fdr takes the form of

$$\frac{f_{1,i}(p_i)}{f_0(p_i)} \geq \frac{(1-t)\pi_{0i}}{t(1-\pi_{0i})}$$

where t $\in (0, 1)$ is a cutoff value

CAMT

- Since $f_{1,i}$ is unidentifiable without extra assumptions on its form, and consistent estimation of the decision rule is difficult, this paper suggested a rejection rule that can mimic some characteristics of the optimal rule
- First, replace $\frac{f_{1,i}}{f_0}$ by a surrogate function h_i which satisfy

$$\begin{cases} h_i(p) > 0 \text{ for p} \in [0,1] \\ \int_0^1 h_i(p)dp = 1 \\ \text{h is decreasing} \end{cases}$$

- In this article, beta density is suggested

$$h_i(p) = (1-k_i)p^{-k_i}, 0 < k_i < 1$$

where k_i is a parameter that depends on x_i

- Based on the surrogate likelihood ratio, the rejection rule is given by

$$h_i(p_i) \geq w_i(t) := \frac{(1-t)\pi_i}{t(1-\pi_i)}$$

for some weights π_i to be determined later

CAMT

- In this article, EM-type algorithm is proposed to estimate π_i and k_i. and model both π_i and k_i as functions of the covariate x_i
- Suppose

$$p_i|H_i, x_i \sim (1 - H_i)f_0 + H_i f_{1,i}$$
$$x_i|H_i \sim (1 - H_i)g_0 + H_i g_1$$

  where $H_i \overset{iid}{\sim} \text{Bernoulli}(1 - \pi_0)$

- Using Bayes rule, we have $f(p_i \mid x_i) = \pi(x_i)f_0(p_i) + (1 - \pi(x_i))f_{1,i}(p_i)$ where $\pi(x) = \frac{g_0(x)\pi_0}{g_0(x)\pi_0 + g_1(x)(1-\pi_0)} = f(H_i = 0|x_i = x)$
- Therefore, π_i is the conditional probability that the ith hypothesis is under the null give the covariate x_i

CAMT

- To motivate estimation procedure for π_i and k_i, let's define

$$\pi_\theta(x) = \frac{1}{1+e^{-\theta_0-\theta_1' x}} \text{ and } k_\beta(x) = \frac{1}{1+e^{-\beta_0-\beta_1' x}}$$

- Conditonal on x_i and marginalizing over H_i,

$$\begin{aligned} f(p_i \mid x_i) &= \pi_\theta(x_i) f_0(p_i) + (1-\pi_\theta(x_i)) f_{1,i}(p_i) \\ &= f_0(p_i) \left\{ \pi_\theta(x_i) + (1-\pi_\theta(x_i) \frac{f_{1,i}(p_i)}{f_0(p_i)}) \right\} \end{aligned}$$

- Replacing $\frac{f_{1,i}}{f_0}$ by surrogate likelihood ratio whose parameters k_i depend on x_i,

$$\widetilde{f}(p_i|x_i) = f_0(p_i) \left\{ \pi_\theta(x_i) + (1-\pi_\theta(x_i)(1-k_\beta(x_i)) p_i^{-k_\beta(x_i)} \right\}$$

CAMT

- Take a log scale and summing up the individual log likelihoods, null density is a nuisance parameter that does not depend on θ and β:

$$\sum_{i=1}^{m} log\widetilde{f}(p_i|x_i) = \sum_{i=1}^{m} log\left\{\pi_\theta(x_i) + (1-\pi_\theta(x_i)(1-k_\beta(x_i))p_i^{-k_\beta(x_i)}\right\} + C_0$$

where $C_0 = \sum_{i=1}^{m} logf_0(p_i)$

- Above discussions motivate the following optimization problem for estimating the unknown parameters

$$\underset{\theta=(\theta_0,\theta_1)'\in\Theta,\beta=(\beta_0,\beta_1)'\in\mathfrak{B}}{max} \sum_{i=1}^{m} log\{\pi_i + (1-\pi_i)(1-k_i)p^{-k_i}\}$$

where $log(\frac{\pi_i}{1-\pi_i}) = \theta_0 + \theta_1' x_i$, $log(\frac{k_i}{1-k_i}) = \beta_0 + \beta_1' x_i$

- This can be solved using the EM algorithm together with the Newton's method in its M-step

Last Algorithm

$$\hat{\pi}_i = W(1/(1+e^{-\tilde{x}_i'\hat{\theta}}), \epsilon_1, \epsilon_2)$$
$$:= \begin{cases} \epsilon_1, & \text{if } 1/(1+e^{-\tilde{x}_i'\hat{\theta}}) \le \epsilon_1, \\ 1/(1+e^{-\tilde{x}_i'\hat{\theta}}), & \text{if } \epsilon_1 < 1/(1+e^{-\tilde{x}_i'\hat{\theta}}) < 1-\epsilon_2, \\ 1-\epsilon_2, & \text{otherwise,} \end{cases}$$

and $\hat{k}_i = 1/(1+e^{-\tilde{x}_i'\hat{\beta}})$ with $\tilde{x}_i = (1, x_i')'$, and

$$\hat{w}_i(t) = \frac{(1-t)\hat{\pi}_i}{t(1-\hat{\pi}_i)}.$$

$$\hat{t} = \max\left\{ t \in [0,1] : \frac{1 + \sum_{i=1}^m \mathbf{1}\{(1-\hat{k}_i)(1-p_i)^{-\hat{k}_i} > \hat{w}_i(t)\}}{1 \vee \sum_{i=1}^m \mathbf{1}\{(1-\hat{k}_i)p_i^{-\hat{k}_i} \ge \hat{w}_i(t)\}} \le \alpha \right\}.$$

- Use winsorization to prevent $\widehat{\pi}_i$ from being too close to zero
- In numerical studies, $\epsilon_1 = 0.1$ and $\epsilon_2 = 10^{-5}$ appeared to perform reasonably well
- Then, reject the ith hypothesis if $(1-\widehat{k}_i)p_i^{-\widehat{k}_i} \ge \widehat{w}_i(\widehat{t})$

Compare AdaPT vs CAMT

AdaPT

- use partially censored p-values to determine the threshold, which can discard useful information about the alternative distribution of p-values
- requires multiple stages
- Unknown whether it controls FDR with dependent p-values

CAMT

- use all the p-values to determine the threshold, and therefore exhibit more power when signal is sparse
- a single-stage procedure
- achieves asymptotic FDR control even when p-values are dependent

CAMT

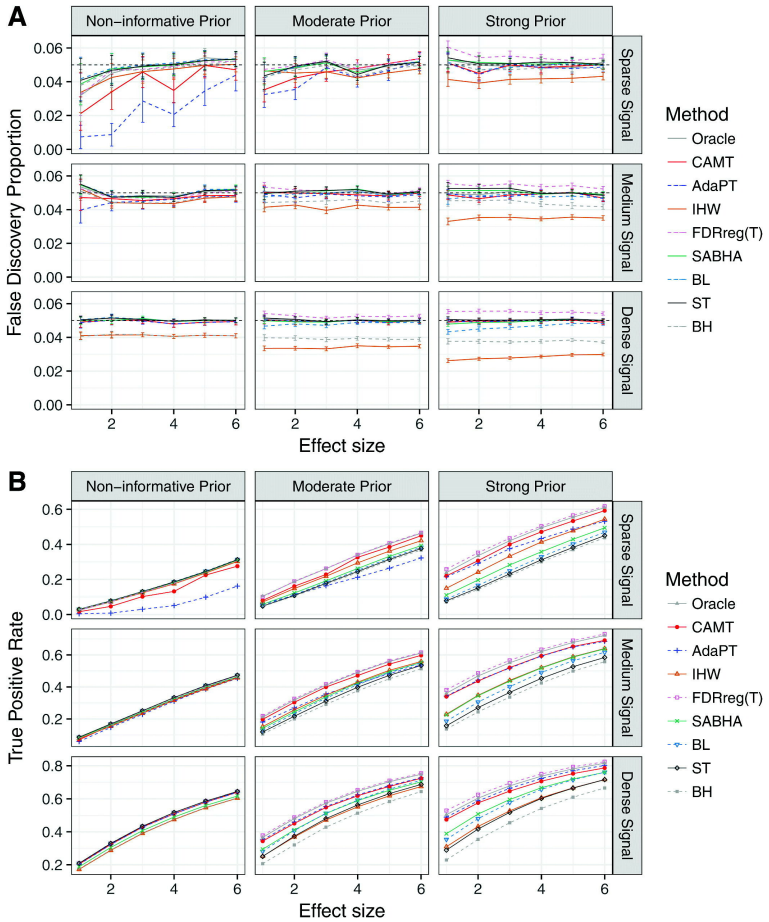

Figure: CAMT

CAMT

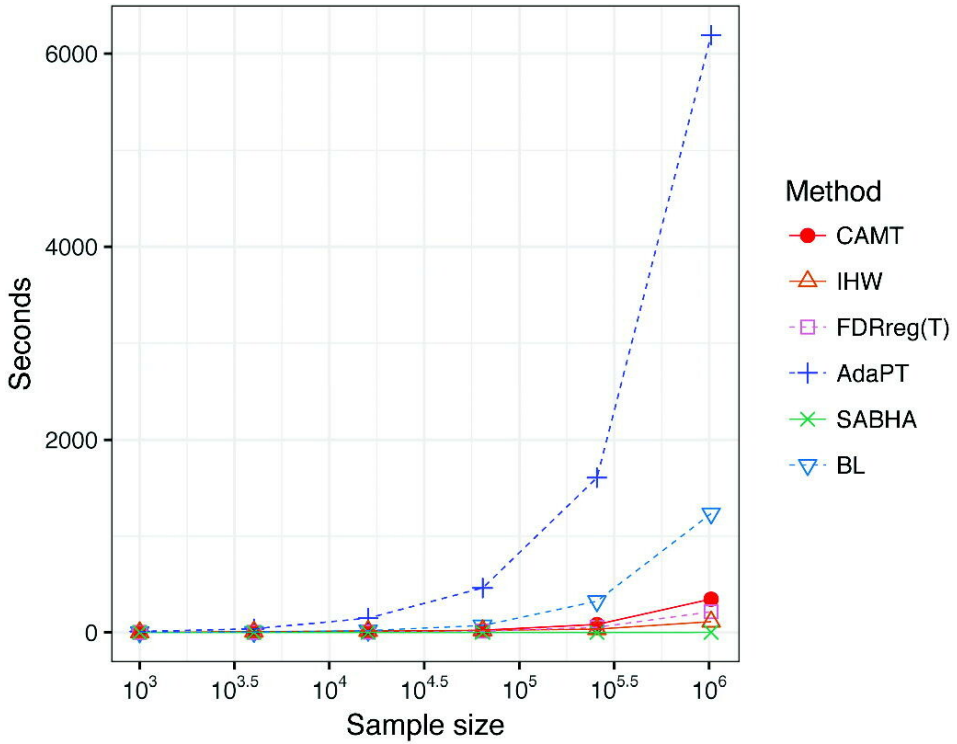

Figure: CAMT

References

- Christopher R. Genovese, "A Tutorial on False Discovery Control", Department of Statistics Carnegie Mellon University
- Bradely Efron and Robert Tibshirani, "Empirical Bayes Methods and False Discovery Rates for Microarrays", Genet Epidemiol. 2002 Jun;23(1):70-86. doi: 10.1002/gepi.1124.
- Lihua Lei and William Fithian, "AdaPT: An interactive procedure for multiple testing with side information", J. R. Statist. Soc.B (2018)80,Part4,pp.649–679
- R. Yurko, M. G'Sell, K. Roeder, and B. Devlin, "A selective inference approach for false discovery rate control using multiomics covariates yields insights into disease risk", PNAS, 2020 Jun; vol. 117 (26) 15028-15035
- X. Zhang and J. Chen, "Covariate Adaptive False Discovery Rate Control With Applications to Omics-Wide Multiple Testing", Journal of The American Statistical Association, 2022, vol.117 no.537,411-427