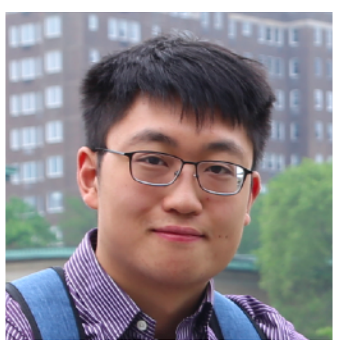
Bringing Closure to FDR Control With a Uniform Improvement of the e-Benjamini-Hochberg Procedure

Intenational Seminar on Selective Inference
May 27, 2025



Neil Xu

Joint work with:

Aldo Solari

Lasse Fischer

Rianne de Heide

Aaditya Ramdas

Jelle Goeman

Outline

1. Multiple testing with error risk control
2. The Closure Principle for FWER control
3. The e-Closure Principle
    3.1. Improving the eBH procedure
        3.1.1. When does $\overline{\mathrm{eBH}}$ beat eBH?
    3.2. Improving the BY procedure
    3.3. Computing e-Closure procedures
4. Simulations
5. Conclusion

Multiple testing w/ general risk metrics

Setup: Statistical model M, where $H_1, \ldots, H_m$ are m hypotheses (with $H_i \subseteq M$)

True (null) hypotheses for distribution $\mathrm{P} \in M$: $N_{\mathrm{P}} := \{i \in [m] : \mathrm{P} \in H_i\}$

Intersection hypothesis $H_S := \cap_{i \in S} H_i$ for each $S \in 2^{[m]}$

Output **discovery (or rejection) set** $\mathbf{R} \in 2^{[m]}$ or set of discovery sets $\mathscr{R} \subseteq 2^{[m]}$

Nonnegative **loss** $\mathrm{f}_N(R) : 2^{[m]} \times 2^{[m]} \to [0,\infty)$, e.g,

$\mathrm{f}_N(R) = \mathbf{1}\{ |N \cap R| > 0\}$ — single false discovery

$\mathrm{f}_N(R) = \mathrm{FDP}_N(R) := \dfrac{|N \cap R|}{|R| \vee 1}$ (false discovery proportion)

Desiderata: ER_{f} (error rate) is controlled at (fixed) level $\alpha \in [0,1]$

$$\mathbb{E}_{\mathrm{P}}\left(\mathrm{f}_{N_{\mathrm{P}}}(\mathbf{R})\right) \leq \alpha$$

(classical control)

or

$$\mathbb{E}_{\mathrm{P}}\left(\max_{R \in \mathscr{R}} \mathrm{f}_{N_{\mathrm{P}}}(R)\right) \leq \alpha$$

(simultaneous control)

The Closure Principle for FWER control

Family-wise error rate (FWER): $\mathbf{R}$ controls FWER if $\mathrm{P}(|N_{\mathrm{P}} \cap \mathbf{R}| > 0) \leq \alpha$ for all $\mathrm{P} \in M$.

Theorem (Marcus '76, Sonneman '82, '08) **:** $\mathbf{R}$ is FWER controlling iff it can be written as

$$\mathrm{R}(\boldsymbol{\varphi}) = \{i \in [m] : \varphi_S = 1 \text{ for all } i \in S\}$$

where $\boldsymbol{\varphi} = (\varphi_S)_{S \in 2^{[m]}}$ is a set of local intersection tests.

$\varphi_S \in \{0,1\}$ is a **local intersection test** iff $\mathrm{P}(\varphi_S = 1) \leq \alpha$ for all $\mathrm{P} \in H_S$.

Given p-values $P_1, \ldots, P_m$ for m hypotheses, two FWER controlling procedures are:

$$\mathbf{R}^{\mathsf{B}} = \{i \in [m] : P_i \leq \alpha/m\} \qquad \mathbf{R}^{\mathsf{HB}} = \mathrm{R}(\boldsymbol{\varphi}^{\mathsf{B}}) = \{i \in [m] : P_i \leq t^{\mathsf{HB}}\}$$

$$\varphi_S^{\mathsf{B}} = \mathbf{1}\left\{\min_{i \in S} P_i \leq \alpha/|S|\right\} \qquad t^{\mathsf{HB}} = \max\{P_{(i)} : i \in [m] \text{ and } P_{(j)} \leq \alpha/(m-j+1) \text{ for all } j \leq i\} \cup \{0\}$$

Since $t^{\mathsf{HB}} \geq \alpha/m$, Holm-Bonferroni uniformly improves Bonferroni.

Classical Closure also capture all methods with simultaneous probabilistic bounds on the FDP (Genovese + Wasserman '06, Goeman + Solari '11, Goeman+ '21).

E-values

A nonnegative random variable $\mathbf{e}$ is an **e-value** w.r.t. to a distribution $\mathrm{P} \in M$ if and only if $\mathbb{E}_{\mathrm{P}}(\mathbf{e}) \leq 1$.

Given e-values $\mathbf{e}_1, \ldots, \mathbf{e}_m$ for $H_1, \ldots, H_m$, define:

$$\mathscr{R}^{\mathrm{SC}} := \left\{ R \in 2^{[m]} : \min_{i \in R} \mathbf{e}_i \geq \frac{m}{\alpha |R|} \right\}$$

self-consistency

$$\mathbf{R}^{\mathrm{eBH}} = \operatorname{argmax}_{R \in \mathscr{R}^{\mathrm{SC}}} |R|$$

eBH procedure

(eBH = Benjamini-Hochberg (BH) applied to $\mathbf{e}_1^{-1}, \ldots, \mathbf{e}_m^{-1}$)

Theorem (Wang + Ramdas '20) **:** $\mathscr{R}^{\mathrm{SC}}$ and simultaneous FDR control (and $\mathbf{R}^{\mathrm{eBH}}$ has FDR control).

<u>"If we have (compound) e-values, then we robustness to dependence, selection rule, etc."</u>
(Wang + Ramdas '20, X+ '21, Hartog + Lei '25, etc.)
Motivated by many developments in e-values (sequential testing, universal inference, Chernoff bounds, etc.)

<u>"X was (compound) e-values (+eBH) all along"</u>

1. Ren + Barber '24 (knockoffs)
2. Bashari+ '24, Lee + Ren '24 (conformal)
3. Li + Zhang '25 (BH)

(among others)

<u>"All FDR controlling procedures were compound e-values + eBH all along"</u>
(Banerjee+ '23, Ignatiadis+ '25)

Specific to FDR, and not easy to derive improvements.

The generalized e-Closure Principle

$\mathbf{E} := (\mathbf{e}_S)_{S \in 2^{[m]}}$ is a **e-collection** if $\mathbf{e}_{N_\mathrm{P}}$ is an e-value for each $\mathrm{P} \in M$.

Theorem (ours): $\mathscr{R}$ has simultaneous ER_f control iff there exists an e-collection $\mathbf{E}$ and

$$\mathscr{R}^{\mathrm{ER}_\mathrm{f}}(\mathbf{E}) = \left\{ R \in 2^{[m]} : \mathbf{e}_S \geq \frac{\mathrm{f}_S(R)}{\alpha} \text{ for all } S \in 2^{[m]} \right\}$$ such that $\mathscr{R} \subseteq \mathscr{R}^{\mathrm{ER}_\mathrm{f}}(\mathbf{E})$.

Proof:

(e-collection $\to \mathrm{ER}_\mathrm{f}$)

If $R \in \mathscr{R}^{\mathrm{ER}_\mathrm{f}}(\mathbf{E})$, then $\mathbb{E}_\mathrm{P}\left(\max_{R \in \mathscr{R}^{\mathrm{ER}_\mathrm{f}}(\mathbf{E})} \mathrm{f}_{N_\mathrm{P}}(R) \right) \leq \mathbb{E}_\mathrm{P}\left(\alpha \mathbf{e}_{N_\mathrm{P}} \right) \leq \alpha$.

($\mathrm{ER}_\mathrm{f} \to$ e-collection)

If we are given $\mathscr{R}$ that has ER_f control, then define

$$\mathbf{e}_S = \frac{\max_{R \in \mathscr{R}} \mathrm{f}_S(R)}{\alpha}$$

If $R \in \mathscr{R}$ then $\frac{\mathrm{f}_S(R)}{\alpha} \leq \mathbf{e}_S$ for all $S \in 2^{[m]}$ by definition of $\mathbf{e}_S$ (thus $\mathscr{R} \subseteq \mathscr{R}^{\mathrm{ER}_\mathrm{f}}(\mathbf{E})$).

$$\mathbb{E}_\mathrm{P}\left(\mathbf{e}_{N_\mathrm{P}}\right) = \frac{\mathbb{E}_\mathrm{P}\left(\max_{R \in \mathscr{R}} \mathrm{f}_{N_\mathrm{P}}(R)\right)}{\alpha} \leq 1$$ by $\mathscr{R}$ being ER_f controlling.

Let $\mathrm{ER}_\mathrm{f} = \mathrm{FWER}$, i.e.,

$$\mathrm{f}_N(R) = \mathbf{1}\{ |N \cap R| > 0 \}$$

If we set

$$\varphi_S = \mathbf{1}\{\mathbf{e}_S \geq \alpha^{-1}\}$$

or

$$\mathbf{e}_S = \varphi_S \cdot \alpha^{-1}$$

We have that

$$\mathrm{R}(\boldsymbol{\varphi}) = \operatorname{argmax}_{R \in \mathscr{R}^{\mathrm{FWER}}(\mathbf{E})} |R|$$

Recovers standard Closure Principle via all or nothing e-values

Theorem (ours): All FDR controlling procedures are characterized by the e-Closure Principle with $\mathrm{f}_S(R) = \mathrm{FDP}_S(R)$.

One view: an application of beyond Neyman-Pearson paradigm to multiple testing (Grünwald ’24).

$\overline{\text{eBH}}$ (closed eBH) : Improving the eBH procedure

Assume we have m arbitrarily dependent e-values $\mathbf{e}_1, \ldots, \mathbf{e}_m$.

Define $\mathbf{E}$ with $\mathbf{e}_S = \frac{1}{|S|} \sum_{i \in S} \mathbf{e}_i$ Only admissible symmetric e-merging function (Vovk + Wang '21)

Define the novel procedures:

$$\mathscr{R}^{\mathrm{MC}} := \mathscr{R}^{\mathrm{FDR}}(\mathbf{E}) = \left\{ R \in 2^{[m]} : \mathbf{e}_S \geq \frac{\mathrm{FDP}_S(R)}{\alpha} \text{ for all } S \in 2^{[m]} \right\}$$ **mean-consistency**

$\mathbf{R}^{\overline{\text{eBH}}} := \text{argmax}_{R \in \mathscr{R}^{\mathrm{MC}}} |R|$ **closed eBH procedure**

Theorem (ours): $\mathscr{R}^{\mathrm{MC}}$ has simultaneous FDR control via the e-Closure Principle.
Further, $\mathscr{R}^{\mathrm{SC}} \subseteq \mathscr{R}^{\mathrm{MC}}$ and consequently $\mathrm{R}^{\mathrm{eBH}} \subseteq \mathrm{R}^{\overline{\mathrm{eBH}}}$ (uniform improvements).

Proof: Let $R \in \mathscr{R}^{\mathrm{SC}}$, then

$$\mathbf{e}_S = \frac{1}{|S|} \sum_{i \in S} \mathbf{e}_i \geq \frac{|S \cap R| \cdot \min_{i \in S \cap R} \mathbf{e}_i}{|S|} \geq \frac{|S \cap R| \cdot \min_{i \in R} \mathbf{e}_i}{|S|} \geq \frac{|S \cap R|}{|R| \vee 1} \cdot \frac{m}{|S|\alpha} \geq \frac{\mathrm{FDP}_S(R)}{\alpha}$$

Thus, $R \in \mathscr{R}^{\mathrm{MC}}$.

(The minimally adaptive eBH of Ignatiadis+ '23 is also contained in $\mathscr{R}^{\mathrm{MC}}$ via similar argument.)

When does $\overline{\text{eBH}}$ beat eBH?

$m = 3$ hypotheses, $\alpha^{-1} = 20$ $(\alpha = 0.05)$

Minimally adaptive eBH rejects
$$\begin{cases} \text{eBH at level } \alpha \cdot \dfrac{m}{m-1} & \text{if } \mathbf{e}_{[m]} \geq \alpha^{-1} \\ \text{nothing} & \text{otherwise} \end{cases}$$

minimally adaptive eBH rejects 2: $\mathbf{e}_3 < \dfrac{2}{3\alpha} = 13\frac{1}{3}$

eBH rejects 1: $\mathbf{e}_2 < \dfrac{3}{2\alpha} = 30$

$\mathbf{e}_1 = 60$ $\quad$ $\mathbf{e}_2 = 29$ $\quad$ $\mathbf{e}_3 = 11$

$\mathbf{e}_{\{1,2\}} = 44.5$ $\quad$ $\mathbf{e}_{\{2,3\}} = 20$

$\mathbf{e}_{\{1,2,3\}} = 33\frac{1}{3}$

$\dfrac{\mathrm{FDP}_S(\mathbf{R}^{\overline{\text{eBH}}})}{\alpha}$

$\dfrac{1}{3\alpha} = 6\frac{2}{3}$

$\dfrac{2}{3\alpha} = 13\frac{1}{3}$

$\dfrac{1}{\alpha} = 20$

$\overline{\text{eBH}}$ rejects all 3

$\overline{\text{eBH}}$ adapts to non-null e-values.
(observed in FWER/FDP controlling methods of Vovk + Wang ’23 and Hartog + Lei ’25)

$\overline{\mathrm{BY}}$ (closed BY): improving the BY procedure

Assume we have m arbitrarily dependent p-values $\mathbf{p}_1, \ldots, \mathbf{p}_m$.

BY procedure (Benjamini + Yekutieli '01)

$$\mathscr{R}^{\mathrm{BY}} := \left\{ R \in 2^{[m]} : \max_{i \in R} \mathbf{p}_i \leq \frac{\alpha |R|}{m h_m} \right\}$$

$\mathbf{R}^{\mathrm{BY}} :=$ argmax$_{R \in \mathscr{R}^{\mathrm{BY}}} |R|$

Theorem (Su '18, Benjamini + Yekutieli '01) **:** $\mathscr{R}^{\mathrm{BY}}$ has simultaneous FDR control (and $\mathbf{R}^{\mathrm{BY}}$ has FDR control).

Define p-to-e calibrator $\mathrm{e}_k(p) := \dfrac{k \cdot \mathbf{1}\{p \leq \alpha / h_k\}}{\alpha \cdot \lceil p \cdot k h_k \alpha^{-1} \rceil}$.

$\mathbf{R}^{\mathrm{BY}}$ is equivalent to applying eBH to $\mathbf{e}_i := \mathrm{e}_m(\mathbf{p}_i)$.

(Vovk + Wang '21, X + Wang + Ramdas '24)

Define e-collection $\mathbf{E}$ with

$$\mathbf{e}_S = \frac{1}{|S|} \sum_{i \in S} \mathrm{e}_{|S|}(\mathbf{p}_i).$$

closed BY procedure

$\mathscr{R}^{\overline{\mathrm{BY}}} := \mathscr{R}^{\mathrm{FDR}}(\mathbf{E})$ $\mathbf{R}^{\overline{\mathrm{BY}}} :=$ argmax$_{R \in \mathscr{R}^{\overline{\mathrm{BY}}}} |R|$

Theorem (ours) : $\mathscr{R}^{\overline{\mathrm{BY}}}$ has simultaneous FDR control (and $\mathbf{R}^{\overline{\mathrm{BY}}}$ has FDR control) via the e-Closure Principle. Further, $\mathscr{R}^{\mathrm{BY}} \subseteq \mathscr{R}^{\overline{\mathrm{BY}}}$ and $\mathbf{R}^{\mathrm{BY}} \subseteq \mathbf{R}^{\overline{\mathrm{BY}}}$.

Proof: Let $R \in \mathscr{R}^{\mathrm{BY}}$ and $R \neq \varnothing$

$$\alpha \cdot \mathbf{e}_S \geq \frac{\alpha}{|S|} \sum_{i \in S \cap R} \mathrm{e}_{|S|}(\mathbf{p}_i) = \sum_{i \in S \cap R} \frac{\mathbf{1}\{\mathbf{p}_i \leq \alpha / h_{|S|}\}}{\lceil \mathbf{p}_i |S| h_{|S|} \alpha^{-1} \rceil} \geq \sum_{i \in S \cap R} \frac{1}{\lceil \mathbf{p}_i |S| h_{|S|} \alpha^{-1} \rceil} \geq \sum_{i \in S \cap R} \frac{1}{\lceil \alpha |R| (m h_m)^{-1} |S| h_{|S|} \alpha^{-1}) \rceil}$$

$$\geq \frac{|S \cap R|}{|R|} = \mathrm{FDP}_S(R).$$

Thus, $R \in \mathscr{R}^{\overline{\mathrm{BY}}}$.

Computing e-Closure procedures

Computing $\overline{\mathrm{eBH}}$ discovery set efficiently:

Define $\mathbf{s}_i = \sum_{j=1}^{i} \mathbf{e}_{(i)}$ —- we reject the r largest e-values if

$g(r, n_1, n_2) := \mathbf{s}_m - \mathbf{s}_{n_1} + \mathbf{s}_r - \mathbf{s}_{n_2} - \dfrac{(m - n_1 + r - n_2)(r - n_2)}{r\alpha} \geq 0$ for all $n_2 \in [r-1], n_1 \in \{r, \ldots, m\}$

g is convex in n_2 via $\mathbf{e}_{(i)}$ being decreasing in i.

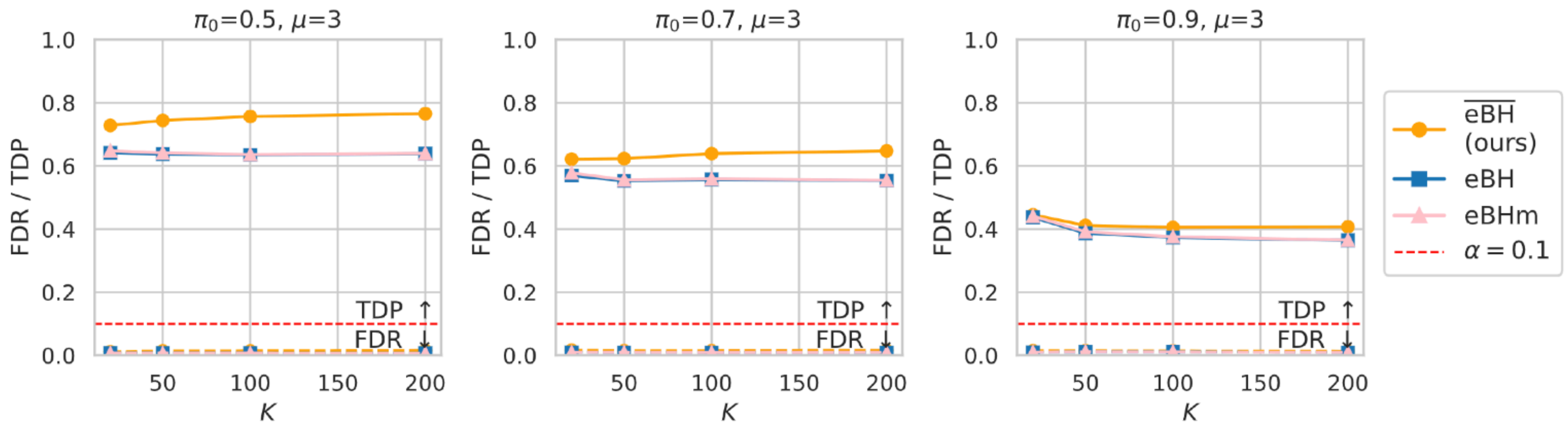
Since there are $O(m^2)$ pairs of r and n_1 to check, overall complexity of $O(m^2 \log m)$.

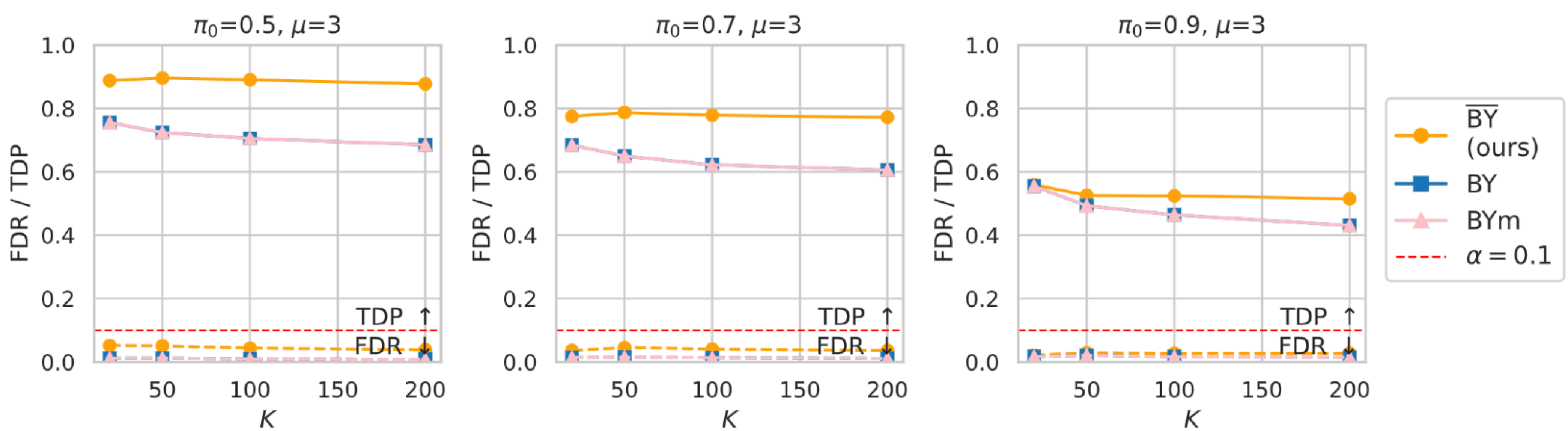
Computing $\overline{\mathrm{BY}}$ discovery set —- $O(m^3)$ total to find worst case null for each discovery set consisting of k smallest p-values (similar to Dobriban '20 for classical Closure Principle).

Slightly less powerful but fast version: apply $\overline{\mathrm{eBH}}$ to $\mathbf{e}_i = \mathrm{e}_m(\mathbf{p}_i)$.

Simulations

H_i: $\mu_i = 0$ (else $\mu_i = 3$)
$\mathbf{X} := (X_1, \ldots, X_m) \sim \mathcal{N}(\mu, \Sigma)$
$\mathbf{e}_i = \exp(\lambda X_i - \lambda^2/2)$ where $\lambda = 3$ and ($\Sigma = I_d$)
$\mathbf{p}_i = 1 - \Phi(X_i)$ where $|\Sigma_{i,j}| = \exp(-|i-j|/10)/5$ and switching pos. and neg.





Conclusion

- The e-Closure Principle generalizes the classical Closure Principle to characterize general multiple testing risk control.
  - Explicit derivation for tail bounds on FDP as well.
- We derive uniform improvements of eBH, BY, and Su's self-consistent procedure under positive dependence (see paper).
- Post-hoc validity, boosting, randomization, and alternative e-merging functions in paper as well.

Combined paper on arXiv soon.

Thanks!

e-Closure and the universality of eBH

Define compound e-values $\tilde{\mathbf{e}}_1, \ldots, \tilde{\mathbf{e}}_m$ for $\mathrm{P} \in M$ as satisfying $\sum_{i \in N_\mathrm{P}} \mathbb{E}_{N_\mathrm{P}}(\tilde{\mathbf{e}}_i) \leq m$

Define $\mathbf{E}$ with $\mathbf{e}_S = \frac{1}{m} \sum_{i \in S} \tilde{\mathbf{e}}_i$ and $\tilde{\mathscr{R}}^{\mathrm{MC}} := \mathscr{R}^{\mathrm{FDR}}(\mathbf{E})$

$\tilde{\mathscr{R}}^{\mathrm{SC}} := \left\{ R \in 2^{[m]} : \min_{i \in R} \tilde{\mathbf{e}}_i \geq \frac{m}{\alpha |R|} \right\}$

Theorem (ours): $\tilde{\mathscr{R}}^{\mathrm{MC}} = \tilde{\mathscr{R}}^{\mathrm{SC}}$

Proof: $\tilde{\mathscr{R}}^{\mathrm{SC}} \subseteq \tilde{\mathscr{R}}^{\mathrm{MC}}$ since $\sum_{i \in S} \tilde{\mathbf{e}}_i \geq |S \cap R| \min_{i \in R} \tilde{\mathbf{e}}_i$ for all $R \in \tilde{\mathscr{R}}^{\mathrm{SC}}$

Let $R \in \tilde{\mathscr{R}}^{\mathrm{MC}}$. For all $i \in R$

$\mathbf{e}_{\{i\}} = \frac{\tilde{\mathbf{e}}_i}{m} \geq \frac{\mathrm{FDP}_{\{i\}}(R)}{\alpha} = (\alpha |R|)^{-1}$. Thus $\tilde{\mathbf{e}}_i \geq \frac{m}{\alpha |R|}$ for all $i \in R$ and $R \in \tilde{\mathscr{R}}^{\mathrm{SC}}$.