p-filter: Multilayer False Discovery Rate for grouped hypotheses

Aaditya Ramdas

University of California, Berkeley
www.cs.berkeley.edu/~aramdas/

Joint work with (the excellent) Rina Foygel Barber, Martin Wainwright and Mike Jordan



Multiple comparisons & FDR control

Multiple comparisons & FDR control

Central Question:

When testing n different null hypotheses simultaneously, how do we determine which effects are significant?

Multiple comparisons & FDR control

Central Question:

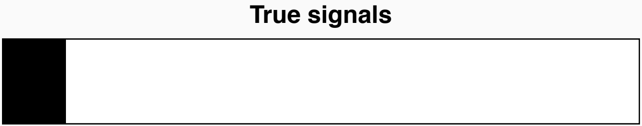
When testing n different null hypotheses simultaneously, how do we determine which effects are significant? **and** take prior structural knowledge into account while doing this?

(selective inference with structure)

Multiple comparisons & FDR control

Central Question:
When testing n different null hypotheses simultaneously, how do we determine which effects are significant? **and** take prior structural knowledge into account while doing this?
(selective inference with structure)

Example

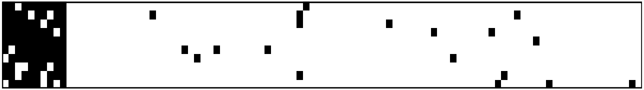


Data: $Z_i \sim \mathcal{N}(\mu, 1)$, $P_i = 1 - \Phi(Z_i)$ for 1000 hypotheses.

10×100 grid : $\mu > 0$ for 100 pixels, $\mu = 0$ for 900 pixels.

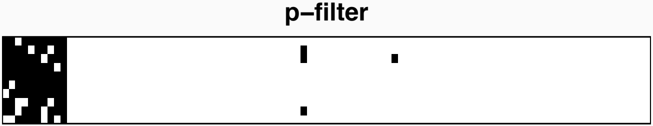
True signals have "small" p-values. Nulls have uniform p-values.

Example



Benjamini Hochberg with entry-level target FDR $= 0.2$.

Example



p-Filter with entry-level and column-level target FDR $= 0.2$

Multiple comparisons & FDR control

How can we incorporate additional information into the procedure?

Multiple comparisons & FDR control

How can we incorporate additional information into the procedure?

- If the hypotheses have a natural clustered / hierarchical structure, how can we take this into account?

Multiple comparisons & FDR control

How can we incorporate additional information into the procedure?

- If the hypotheses have a natural clustered / hierarchical structure, how can we take this into account?
- You may want to **group** together hypotheses that are likely to be null together or be non-null together.

Multiple comparisons & FDR control

How can we incorporate additional information into the procedure?

- If the hypotheses have a natural clustered / hierarchical structure, how can we take this into account?
- You may want to **group** together hypotheses that are likely to be null together or be non-null together.
- In spatio-temporal applications, it might be natural to group hypotheses by space or time or space-time blocks. "Discovery at time/location x makes discoveries around x more likely".

Multiple comparisons & FDR control

How can we incorporate additional information into the procedure?

- If the hypotheses have a natural clustered / hierarchical structure, how can we take this into account?
- You may want to **group** together hypotheses that are likely to be null together or be non-null together.
- In spatio-temporal applications, it might be natural to group hypotheses by space or time or space-time blocks. "Discovery at time/location x makes discoveries around x more likely".
- In genetics, certain genes/proteins might be known to act together, or have similar molecular structure.
- We might have some prior guess about which hypotheses are more likely to be null or non-null.

Goal (in English)

- Given n hypotheses with p-values $P := \{P_1, \ldots, P_n\}$

Goal (in English)

- Given n hypotheses with p-values $P := \{P_1, \ldots, P_n\}$
  Eg: Imagine they are placed in a $r \times c$ grid, $n = rc$.

Goal (in English)

- Given n hypotheses with p-values $P := \{P_1, \dots, P_n\}$

  Eg: Imagine they are placed in a $r \times c$ grid, $n = rc$.
- Given M partitions (disjoint subsets of P, whose union is P)

Goal (in English)

- Given n hypotheses with p-values $P := \{P_1, \ldots, P_n\}$

  Eg: Imagine they are placed in a $r \times c$ grid, $n = rc$.
- Given M partitions (disjoint subsets of P, whose union is P)

  Partition 1 could be the set of all singletons,

  Partition 2 could be the set of all rows, and

  Partition 3 could be the set of all columns.

Goal (in English)

- Given n hypotheses with p-values $P := \{P_1, \ldots, P_n\}$

  Eg: Imagine they are placed in a $r \times c$ grid, $n = rc$.
- Given M partitions (disjoint subsets of P, whose union is P)

  Partition 1 could be the set of all singletons,
  Partition 2 could be the set of all rows, and
  Partition 3 could be the set of all columns.
- Goal: select set $\widehat{S} \subseteq [n]$ such that FDR is bounded simultaneously for partition $1, 2, \ldots, M$.

  Few falsely discovered singletons,
  Few falsely discovered rows,
  Few falsely discovered columns.

p-Filter: will discover $\hat{S} \subseteq [n]$ such that FDR is simultaneously controlled for all partitions.

Multilayer FDR

$$\alpha_{\text{indiv}} = 0.1$$
$$\alpha_{\text{group}} = 0.2$$

						Simes p–value
Group 1	0.03	0.01	0.18	0.04	0.08	0.05
Group 2	0.05	0.11	0.06	0.01	0.89	0.05
Group 3	0.14	0.12	0.58	0.11	0.11	0.18
Group 4	0.88	0.24	0.09	0.66	0.45	0.45

Multilayer FDR

$$\alpha_{\text{indiv}} = 0.1$$
$$\alpha_{\text{group}} = 0.2$$

Multilayer FDR

$$\alpha_{\text{indiv}} = 0.1$$
$$\alpha_{\text{group}} = 0.2$$

The issue of “internal consistency”

Observe: If we first (somehow) select which groups to reject, and then naively run BH on hypotheses within those groups, we may not control FDR at the individual level.

The issue of “internal consistency”

Observe: If we first (somehow) select which groups to reject, and then naively run BH on hypotheses within those groups, we may not control FDR at the individual level.

Observe: Even if we corrected for selection, and then run BH, we may end up with “empty rejected groups” — we may have groups that were rejected, within which no hypotheses were rejected.

The issue of “internal consistency”

Observe: If we first (somehow) select which groups to reject, and then naively run BH on hypotheses within those groups, we may not control FDR at the individual level.

Observe: Even if we corrected for selection, and then run BH, we may end up with “empty rejected groups” — we may have groups that were rejected, within which no hypotheses were rejected.

Observe: Alternately, if we first rejected individual hypotheses, and then rejected all groups in which there was a rejected hypothesis, it may not control group FDR.

The issue of “internal consistency”

Observe: If we first (somehow) select which groups to reject, and then naively run BH on hypotheses within those groups, we may not control FDR at the individual level.

Observe: Even if we corrected for selection, and then run BH, we may end up with “empty rejected groups” — we may have groups that were rejected, within which no hypotheses were rejected.

Observe: Alternately, if we first rejected individual hypotheses, and then rejected all groups in which there was a rejected hypothesis, it may not control group FDR.

FDR control with “internal consistency”:
Every rejected hypothesis is in some rejected group, and every rejected group contains at least one rejected hypothesis.

The p-Filter algorithm

Input: n p-values, M partitions, M FDR levels

Let $t_1 = \alpha_1, \ldots, t_M = \alpha_M$. Repeat $m = 1, \ldots, M$, until no change:

The p-Filter algorithm

Input: n p-values, M partitions, M FDR levels

Let $t_1 = \alpha_1, \ldots, t_M = \alpha_M$. Repeat $m = 1, \ldots, M$, until no change:

- For the mth partition, Simes+thresholding
  - Calculate Simes p-values $P^m := \{P_1^m, \ldots, P_G^m\}$
  - Reject all groups whose $P_g^m \leq t_m$.

The p-Filter algorithm

Input: n p-values, M partitions, M FDR levels

Let $t_1 = \alpha_1, \ldots, t_M = \alpha_M$. Repeat $m = 1, \ldots, M$, until no change:

- For the mth partition, Simes+thresholding
  - Calculate Simes p-values $P^m := \{P_1^m, \ldots, P_G^m\}$
  - Reject all groups whose $P_g^m \leq t_m$.
- $\widehat{S} := \{P_i :$ in every partition, P_i's group was selected$\}$, intersect

  Let $\widehat{S}_m$ be the discovered groups in partition m, induced by $\widehat{S}$.

The p-Filter algorithm

Input: n p-values, M partitions, M FDR levels

Let $t_1 = \alpha_1, \ldots, t_M = \alpha_M$. Repeat $m = 1, \ldots, M$, until no change:

- For the mth partition, Simes+thresholding
  - Calculate Simes p-values $P^m := \{P_1^m, \ldots, P_G^m\}$
  - Reject all groups whose $P_g^m \leq t_m$.
- $\widehat{S} := \{P_i :$ in every partition, P_i's group was selected$\}$, intersect

  Let $\widehat{S}_m$ be the discovered groups in partition m, induced by $\widehat{S}$.
- Estimate FDP's for each partition: correction

$$\widehat{\mathsf{FDP}}_m = \frac{t_m \cdot G_m}{|\widehat{S}_m|} \quad \begin{array}{l} \leftarrow \text{approx. \# false discoveries} \\ \leftarrow \text{\# discoveries} \end{array}$$

If $\widehat{\mathsf{FDP}}_m > \alpha_m$, reduce t_m until $\widehat{\mathsf{FDP}}_m$ is $\leq \alpha_m$ (discrete search)

The p-Filter algorithm

Input: n p-values, M partitions, M FDR levels

Let $t_1 = \alpha_1, \ldots, t_M = \alpha_M$. Repeat $m = 1, \ldots, M$, until no change:

- For the mth partition, Simes+thresholding
  - Calculate Simes p-values $P^m := \{P^m_1, \ldots, P^m_G\}$
  - Reject all groups whose $P^m_g \le t_m$.
- $\widehat{S} := \{P_i$: in every partition, P_i's group was selected$\}$, intersect

  Let $\widehat{S}_m$ be the discovered groups in partition m, induced by $\widehat{S}$.
- Estimate FDP's for each partition: correction

$$\widehat{\mathsf{FDP}}_m = \frac{t_m \cdot G_m}{|\widehat{S}_m|} \begin{array}{l} \leftarrow \text{approx. \# false discoveries} \\ \leftarrow \text{\# discoveries} \end{array}$$

  If $\widehat{\mathsf{FDP}}_m > \alpha_m$, reduce t_m until $\widehat{\mathsf{FDP}}_m$ is $\le \alpha_m$ (discrete search)

Note: Simes and BH are special cases when $M = 1$.

p-Filter controls FDR, finds max legal threshold

Theorem 2

p-Filter finds "maximum legal threshold", and it controls FDR simultaneously $\forall m$:

$$\text{FDR for partition } m = \mathbb{E}\left[\frac{|\mathcal{H}_m^0 \cap \widehat{S}_m|}{|\widehat{S}_m|}\right] \leq \alpha_m \cdot \frac{|\mathcal{H}_m^0|}{G_m} \ \ \forall\ m.$$

Furthermore, it halts in $G_1 + G_2 + ... + G_M + 1$ outer loops.

p-Filter controls FDR, finds max legal threshold

Theorem 2

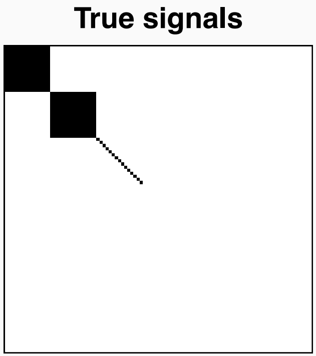
p-Filter finds "maximum legal threshold", and it controls FDR simultaneously $\forall m$:

$$\text{FDR for partition } m = \mathbb{E}\left[\frac{|\mathcal{H}_m^0 \cap \widehat{S}_m|}{|\widehat{S}_m|}\right] \leq \alpha_m \cdot \frac{|\mathcal{H}_m^0|}{G_m} \ \ \forall\ m.$$

Furthermore, it halts in $G_1 + G_2 + ... + G_M + 1$ outer loops.

Does not depend on order of specifying partitions.

Simulation results



p-Filter: entries + rows + columns (3 partitions)

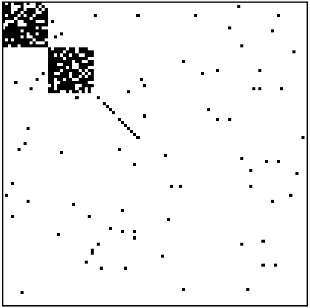
BB: entries + rows (2 partitions, constrained to be hierarchical)

BH: entries only

Target FDR: $\alpha_{\text{entries}} = \alpha_{\text{rows}} = \alpha_{\text{columns}} = 0.2$

Simulation results

BH



p-Filter: entries + rows + columns (3 partitions)

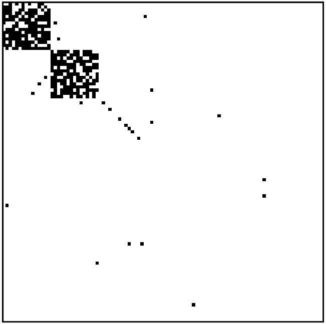
BB: entries + rows (2 partitions, constrained to be hierarchical)

BH: entries only

Target FDR: $\alpha_{\text{entries}} = \alpha_{\text{rows}} = \alpha_{\text{columns}} = 0.2$

Simulation results

p–filter



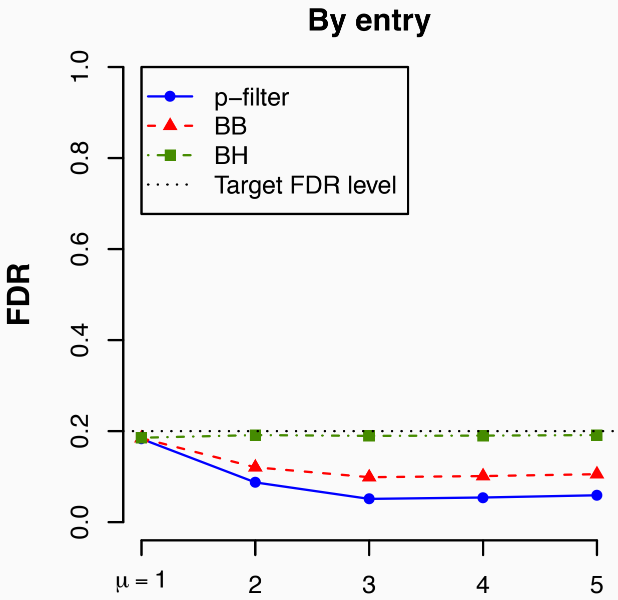
p-Filter: entries + rows + columns (3 partitions)

BB: entries + rows (2 partitions, constrained to be hierarchical)

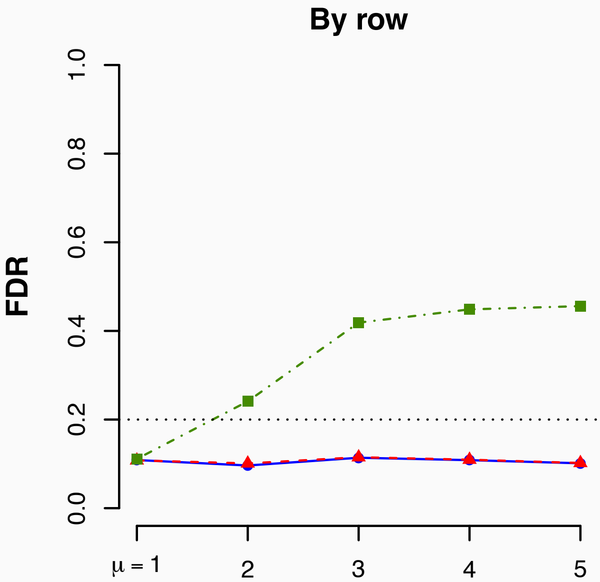
BH: entries only

Target FDR: $\alpha_{\text{entries}} = \alpha_{\text{rows}} = \alpha_{\text{columns}} = 0.2$

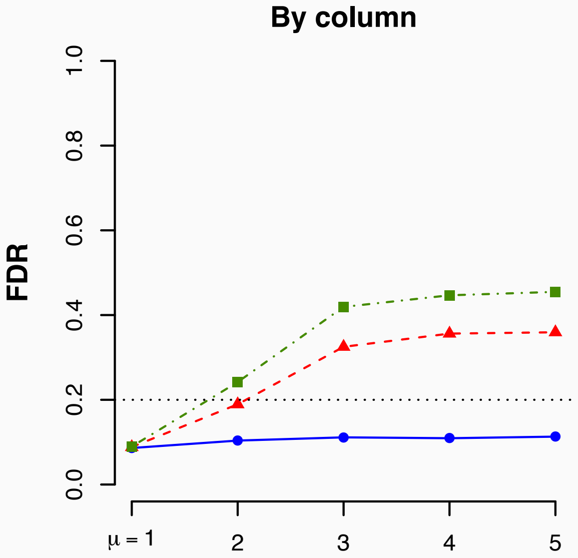
Simulation results



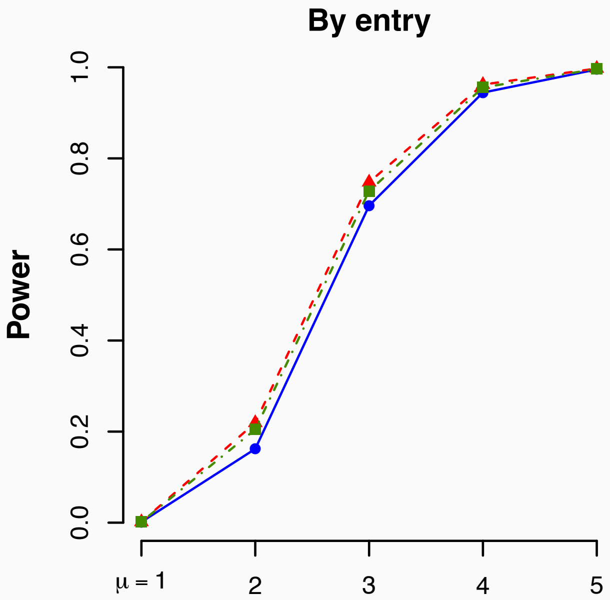
Simulation results



Simulation results



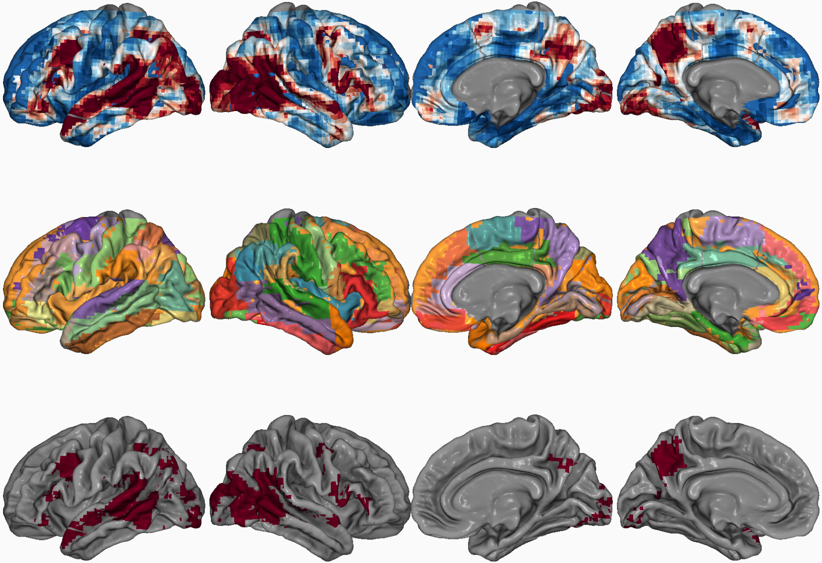
Simulation results



A neuroscience example

- 8 subjects read Chapter 9 of Harry Potter, 1 word = 0.5 sec.
- fMRI recording: 40,000 voxels, 1 scan = 2 sec.
- Consider semantic features of the text (NLP techniques).
- Try to find dependence between text presented at time t with voxel activity at time $t + \delta$, for delay $\delta = 0, 2, 4, 6, 8$ sec.
- One p-value per voxel, per delay δ.
- Can group by space, time or space-time blocks.

A neuroscience example



Big Picture[1]

Summary:

- $m = 1$: p-Filter is a strict generalization of BH and Simes.

[1] Code: `http://www.stat.uchicago.edu/~rina/pfilter.html`

Big Picture[1]

Summary:

- $m = 1$: p-Filter is a strict generalization of BH and Simes.
- $m > 1$: p-Filter controls FDR simultaneously for all partitions.

[1] Code: `http://www.stat.uchicago.edu/~rina/pfilter.html`

Big Picture[1]

Summary:

- $m = 1$: p-Filter is a strict generalization of BH and Simes.
- $m > 1$: p-Filter controls FDR simultaneously for all partitions.
- Flexibly allows the user to incorporate prior information, resulting in higher *precision* of the rejected hypotheses.

[1] Code: `http://www.stat.uchicago.edu/~rina/pfilter.html`

Big Picture[1]

Summary:

- $m = 1$: p-Filter is a strict generalization of BH and Simes.
- $m > 1$: p-Filter controls FDR simultaneously for all partitions.
- Flexibly allows the user to incorporate prior information, resulting in higher *precision* of the rejected hypotheses.
- Can incorporate weights, at hypothesis and group level.
- Can incorporate "null-proportion" estimation.
- Can have overlapping groups, incomplete partitions, etc.

Orthogonal Works (in progress):

- Improved algorithms for online FDR.

[1] Code: `http://www.stat.uchicago.edu/~rina/pfilter.html`

Big Picture[1]

Summary:

- $m = 1$: p-Filter is a strict generalization of BH and Simes.
- $m > 1$: p-Filter controls FDR simultaneously for all partitions.
- Flexibly allows the user to incorporate prior information, resulting in higher *precision* of the rejected hypotheses.
- Can incorporate weights, at hypothesis and group level.
- Can incorporate "null-proportion" estimation.
- Can have overlapping groups, incomplete partitions, etc.

Orthogonal Works (in progress):

- Improved algorithms for online FDR.
- Interactive data analysis with FDR control.

[1] Code: `http://www.stat.uchicago.edu/~rina/pfilter.html`