False discovery rate control with compound p-values

Rina Foygel Barber
(joint work w/ Richard Samworth)

http://rinafb.github.io

Introduction

Intro: multiple testing with false discovery rate (FDR) control

The aim of FDR control:

Given data for testing null hypotheses $H_{0,1}, \dots, H_{0,m}$,
select subset $\widehat{S} \subseteq [m]$ such that

$$\mathsf{FDR} = \mathbb{E}\left[\mathsf{FDP}(\widehat{S})\right] \leq \alpha$$

where

set of true nulls

$$\mathsf{FDP}(\widehat{S}) = \frac{|\widehat{S} \cap \mathcal{H}_0|}{1 \vee |\widehat{S}|} = \text{false discovery proportion}$$

The Benjamini–Hochberg procedure

Given p-values $p_1, \ldots, p_m$ testing null hypotheses $H_{0,1}, \ldots, H_{0,m}$:

BH procedure[1] for FDR control at level α

Rank $p_{(1)} \leq \cdots \leq p_{(m)}$,

& reject $p_{(1)}, \ldots, p_{(k)}$ for largest k such that $p_{(k)} \leq \alpha k/m$

[1] Benjamini & Hochberg 1995, *Controlling the false discovery rate: a practical and powerful approach to multiple testing*

The Benjamini–Hochberg procedure

Theoretical guarantees:

- For independent p-values,[2] FDR $\leq \alpha$
- For PRDS (positive dependence) p-values,[3] FDR $\leq \alpha$

[2] Benjamini & Hochberg 1995, *Controlling the false discovery rate: a practical and powerful approach to multiple testing*

[3] Benjamini & Yekutieli 2001, *The control of the false discovery rate in multiple testing under dependency*

The Benjamini–Hochberg procedure

Theoretical guarantees:

- For independent p-values,[2] FDR $\leq \alpha$
- For PRDS (positive dependence) p-values,[3] FDR $\leq \alpha$
- For arbitrarily dependent p-values,[3] FDR $\leq \alpha \mathsf{h}_m$

$\nearrow$

$\mathsf{h}_m = 1 + \frac{1}{2} + \cdots + \frac{1}{m} \approx \log m$

[2] Benjamini & Hochberg 1995, *Controlling the false discovery rate: a practical and powerful approach to multiple testing*

[3] Benjamini & Yekutieli 2001, *The control of the false discovery rate in multiple testing under dependency*

The Benjamini–Hochberg procedure

Theoretical guarantees:

- For independent p-values,[2] FDR $= \alpha \cdot \frac{|\mathcal{H}_0|}{m}$
- For PRDS (positive dependence) p-values,[3] FDR $\leq \alpha \cdot \frac{|\mathcal{H}_0|}{m}$
- For arbitrarily dependent p-values,[3] FDR $\leq \alpha \mathrm{h}_m \cdot \frac{|\mathcal{H}_0|}{m}$

$\mathrm{h}_m = 1 + \frac{1}{2} + \cdots + \frac{1}{m} \approx \log m$

[2] Benjamini & Hochberg 1995, *Controlling the false discovery rate: a practical and powerful approach to multiple testing*

[3] Benjamini & Yekutieli 2001, *The control of the false discovery rate in multiple testing under dependency*

From p-values to compound p-values

Definition of a p-value:

$$\underbrace{\mathbb{P}\{p_i \leq t\} \leq t}_{\text{requires knowledge of } H_{0,i}} \quad \text{for all } i \in \mathcal{H}_0 \text{ and all } t \in [0,1]$$

From p-values to compound p-values

Definition of a p-value:

$$\underbrace{\mathbb{P}\{p_i \leq t\} \leq t}_{\text{requires knowledge of } H_{0,i}} \text{ for all } i \in \mathcal{H}_0 \text{ and all } t \in [0,1]$$

Compound p-values[4]

$$\frac{1}{m}\sum_{i\in\mathcal{H}_0} \mathbb{P}\{p_i \leq t\} \leq t \text{ for all } t \in [0,1]$$

[4] Armstrong 2022, *False discovery rate adjustments for average significance level controlling tests*
Ignatiadis & Sen 2025, *Empirical partially Bayes multiple testing and compound* χ^2 *decisions*
Ignatiadis, Wang, Ramdas 2024, *Asymptotic and compound e-values: multiple testing and empirical Bayes*

From p-values to compound p-values

Definition of a p-value:

$$\underbrace{\mathbb{P}\{p_i \le t\} \le t}_{\text{requires knowledge of } H_{0,i}} \text{ for all } i \in \mathcal{H}_0 \text{ and all } t \in [0,1]$$

Compound p-values[4]

$$\frac{1}{m}\sum_{i\in\mathcal{H}_0} \mathbb{P}\{p_i \le t\} \le t \text{ for all } t \in [0,1]$$

Intuition for this relaxation:[5] for multiple testing,

estimate of the null only needs to be accurate "on average"

---

[4] Armstrong 2022, *False discovery rate adjustments for average significance level controlling tests*
Ignatiadis & Sen 2025, *Empirical partially Bayes multiple testing and compound* χ^2 *decisions*
Ignatiadis, Wang, Ramdas 2024, *Asymptotic and compound e-values: multiple testing and empirical Bayes*
[5] Efron 2007, *Size, power and false discovery rates*

Theoretical results

BH with compound p-values

Key question:

Bound on FDR when BH is applied to compound p-values?

Existing result: (Armstrong 2022)

If $p_1, \ldots, p_m$ are compound p-values (w/ arbitrary dependence),

$$\mathrm{FDR} \leq \alpha \mathrm{h}_m$$

BH with compound p-values

Key question:

Bound on FDR when BH is applied to compound p-values?

Existing result: (Armstrong 2022)

If $p_1, \ldots, p_m$ are compound p-values (w/ arbitrary dependence),

$$\mathrm{FDR} \leq \alpha \mathrm{h}_m$$

- Same as the result for BH under arbitrary dep.
- Is $\mathrm{FDR} \leq \alpha$ or $\leq \mathcal{O}(\alpha)$ if assume indep. or PRDS?

BH with compound p-values

Theorem

If $p_1, \ldots, p_m$ are independent compound p-values, then

$$\mathrm{FDR} \leq 1.93\alpha$$

for the BH procedure.

Interpretation: FDR inflation is bounded by a constant factor (rather than $h_m \approx \log m$)

BH with compound p-values

A matching lower bound:

Theorem

There exists a joint distribution such that $p_1, \ldots, p_m$ are independent compound p-values, for which

$$\text{FDR} \geq \frac{7}{6}\alpha$$

for the BH procedure, if we assume $\frac{1}{2\alpha}$ is $\leq \frac{m}{3}$ and is an integer.

BH with compound p-values

A special case: the global null ($\mathcal{H}_0 = [m]$)

Theorem

If $p_1, \ldots, p_m$ are independent compound p-values, then under the global null,

$$\text{FDR} \leq \alpha + 2\alpha^2$$

for the BH procedure.

BH with compound p-values

A special case: the global null ($\mathcal{H}_0 = [m]$)

Theorem

If $p_1, \ldots, p_m$ are independent compound p-values, then under the global null,

$$\mathrm{FDR} \leq \alpha + 2\alpha^2$$

for the BH procedure.

Moreover, there exists a joint distribution with

$$\mathrm{FDR} \geq \alpha + \frac{1}{4}\alpha^2,$$

if we assume $\alpha \leq \frac{2}{3}$ and $m \geq 2$.

BH with compound p-values

Under the PRDS condition...

- For p-values, same FDR bound as for independence
- But for compound p-values, a different story....

BH with compound p-values

Under the PRDS condition...

- For p-values, same FDR bound as for independence
- But for compound p-values, a different story....

Theorem

There exists a joint distribution such that $p_1, \ldots, p_m$ are PRDS compound p-values, for which

$$\mathrm{FDR} \geq \frac{3}{8}\alpha \mathrm{h}_m$$

for the BH procedure, if we assume $\alpha \mathrm{h}_m \leq 1$.

BH with compound p-values

Existing results for p-values

	Upper bound = lower bound
Independence	α (Benjamini & Hochberg 1995)
PRDS	α (Benjamini & Yekutieli 2001)
Arbitrary dependence	$\alpha \mathrm{h}_m$ (Benjamini & Yekutieli 2001; Guo & Rao 2008)

BH with compound p-values

Existing results for p-values

	Upper bound = lower bound
Independence	α (Benjamini & Hochberg 1995)
PRDS	α (Benjamini & Yekutieli 2001)
Arbitrary dependence	αh_m (Benjamini & Yekutieli 2001; Guo & Rao 2008)

Existing & new results for compound p-values

	Upper bound	Lower bound
Independence	1.93α	$\frac{7}{6}\alpha$
Indep. + global null	$\alpha + 2\alpha^2$	$\alpha + \frac{1}{4}\alpha^2$
PRDS	αh_m	$\frac{3}{8}\alpha h_m$
Arbitrary dependence	αh_m	αh_m

(Armstrong 2022) (Guo & Rao 2008)

Proof idea

Recall the theorem....

If $p_1, \dots, p_m$ are independent compound p-values, then

$$\text{FDR} \leq 1.93\alpha$$

for the BH procedure.

In order to have $\text{FDP} = \frac{|\widehat{S} \cap \mathcal{H}_0|}{|\widehat{S}|} = \frac{i}{k}$, need

$$\sum_{j \in \mathcal{H}_0} \mathbb{1}\left\{p_j \leq \frac{\alpha k}{m}\right\} = i,$$

along with $k - i$ many non-null $p_j \leq \frac{\alpha k}{m}$

Proof idea

Let $\mathcal{E}_i =$ the event that $\sum_{j \in \mathcal{H}_0} \mathbb{1}\left\{p_j \leq \frac{\alpha k}{m}\right\} \geq i$

↖

independent Bernoulli r.v.'s,

with total prob. $\leq \alpha k$

Proof idea

Let $\mathcal{E}_i =$ the event that $\sum_{j\in\mathcal{H}_0} \mathbb{1}\left\{p_j \leq \frac{\alpha k}{m}\right\} \geq i$

$\nwarrow$ independent Bernoulli r.v.'s, with total prob. $\leq \alpha k$

$$\Rightarrow \quad \mathbb{P}\left\{\mathcal{E}_i\right\} \leq \mathbb{P}\left\{\text{Binom}\left(m, \tfrac{\alpha k}{m}\right) \geq i\right\} \leq \mathbb{P}\left\{\text{Poisson}(\alpha k) \geq i\right\}$$

Proof idea

Proof also uses identities for the Poisson distribution:

Lemma

For any $c > 0$ and $k \geq 1$,

$$\arg\max \left\{ \mathbb{P}\left\{\text{Poisson}(t) \geq k\right\} - c \cdot \mathbb{P}\left\{\text{Poisson}(t) > k\right\} \right\} = \frac{k}{c}$$

Lemma

For any $t \in (0, 1)$,

$$\sum_{k=1}^{\infty} \mathbb{P}\left\{\text{Poisson}(tk) > k\right\} = \frac{1}{2}\left(\frac{t}{1-t}\right)^2$$

Examples of compound p-values

Example 1: weighted p-values

Let $p_1^*, \ldots, p_m^*$ be p-values.

- The BH procedure treats the p_i^*'s symmetrically
- In practice, may want to incorporate prior information

Example 1: weighted p-values

Let $p_1^*, \ldots, p_m^*$ be p-values.

- The BH procedure treats the p_i^*'s symmetrically
- In practice, may want to incorporate prior information

BH with weighted p-values:[6] $p_i = \frac{p_i^*}{w_i} \wedge 1$
(higher weight $\rightsquigarrow$ greater chance of rejection)

[6] Genovese, Roeder, Wasserman 2006, *False discovery control with p-value weighting*

Example 1: weighted p-values

Let $p_1^*, \ldots, p_m^*$ be p-values.

- The BH procedure treats the p_i^*'s symmetrically
- In practice, may want to incorporate prior information

BH with weighted p-values:[6] $p_i = \frac{p_i^*}{w_i} \wedge 1$

(higher weight $\rightsquigarrow$ greater chance of rejection)

The weighted p-values p_i are compound p-values[7]

[6] Genovese, Roeder, Wasserman 2006, *False discovery control with p-value weighting*

[7] Ignatiadis, Wang, Ramdas 2024, *Asymptotic and compound e-values: multiple testing and empirical Bayes*

Example 2: decreasing densities

Independent data points: $X_1 \sim f_1,\ X_2 \sim f_2,\ \ldots,\ X_m \sim f_m$

$$H_{0,i} :\ f_i \text{ is a decreasing density}$$

- Impossible to produce meaningful p-values:

  e.g., observe $X_i = 1.5 \rightsquigarrow$ evidence against $H_{0,i}$?

Example 2: decreasing densities

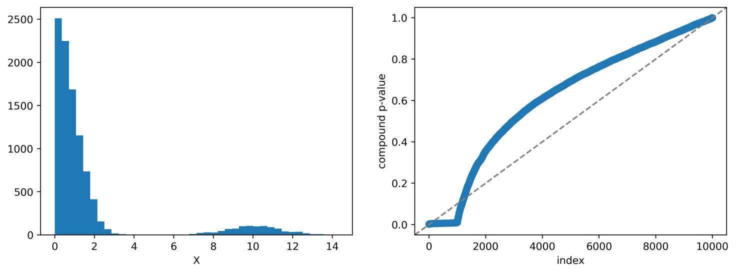
Independent data points: $X_1 \sim f_1,\ X_2 \sim f_2,\ \ldots,\ X_m \sim f_m$

$$H_{0,i}:\ f_i \text{ is a decreasing density}$$

- Impossible to produce meaningful p-values:

  e.g., observe $X_i = 1.5 \rightsquigarrow$ evidence against $H_{0,i}$?

- Can construct compound p-values by estimating $\frac{1}{m}\sum_{i \in \mathcal{H}_0} f_i$ (see paper for details)



Example 3: permutation tests

	Treatment					Placebo				
Trial 1	$X_{1,1}$	$X_{1,2}$	$\dots$	$\dots$		$\dots$	$\dots$	$\dots$	X_{1,n_1}	
$\vdots$										
Trial m	$X_{m,1}$	$X_{m,2}$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	X_{m,n_m}	

- Permutation test p-values (low power if n_i's are small)

$$p_i^{\text{perm}} = \frac{1}{n_i!} \sum_{\sigma \in \mathcal{S}_{n_i}} \mathbb{1}\left\{ T_i\big(\sigma(X_i)\big) \geq T_i(X_i) \right\}$$

test statistic, e.g., 2-sample t-test

Example 3: permutation tests

	Treatment				Placebo			
Trial 1	$X_{1,1}$	$X_{1,2}$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	X_{1,n_1}
$\vdots$								
Trial m	$X_{m,1}$	$X_{m,2}$	$\dots$	$\dots$ $\dots$	$\dots$	$\dots$	$\dots$	X_{m,n_m}

- Permutation test p-values (low power if n_i's are small)

$$p_i^{\text{perm}} = \frac{1}{n_i!} \sum_{\sigma \in \mathcal{S}_{n_i}} \mathbb{1}\left\{ T_i\big(\sigma(X_i)\big) \geq T_i(X_i) \right\}$$

test statistic, e.g., 2-sample t-test

- Compound p-values ("pooled controls"):

$$p_i = \frac{1}{m} \sum_j \left[\frac{1}{n_j!} \sum_{\sigma \in \mathcal{S}_{n_j}} \mathbb{1}\left\{ T_j\big(\sigma(X_j)\big) \geq T_i(X_i) \right\} \right]$$

Experiment

Experiment

Experiment on the `Upworthy` data set:[8]

Compare click rates for different headlines for the same article, e.g.,

```
BAM! Nurses Explain Obamacare In 90 Seconds
Did You Know These 15 Facts About Obamacare?
Nurses Explain How Obamacare Will Affect Your Future
```

Question:
Do headlines containing a numerical digit have a higher click rate?

[8] https://upworthy.natematias.com/

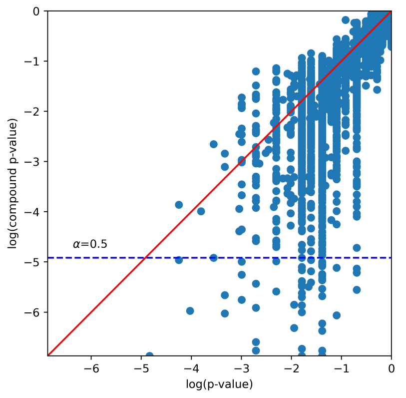
Experiment

- $m = 3{,}893$ articles, with 17,033 headlines in total
- Test statistic: compare click counts for articles with vs. without a numerical digit, using Fisher's exact test (one-sided)
- Compute p-values and compound p-values via permutation test (permute headlines)

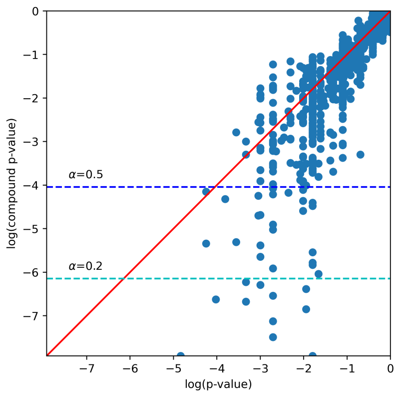
Experiment

Results:

All articles
($m = 3{,}893$)



Articles with > 5 headlines
($m = 1{,}025$)



Discussion

Related work: asymptotic FDR control

Our work: finite-sample FDR bounds for compound p-values

A complementary perspective: the asymptotic regime

- Classical results for p-values: (Storey 2002, Storey et al 2004)
  If empirical distrib. of null p_i's $\to U[0,1]$, then FDP $\to \alpha$
- Analogous results for compound p-values (Armstrong 2022)

Related work: BH with e-values

Random variable $e_j \geq 0$ is an e-value[9] for null hypothesis $H_{0,j}$ if

$$\mathbb{E}[e_j] \leq 1 \text{ for all } j \in \mathcal{H}_0.$$

e-BH procedure[10] for FDR control at level α

Rank $e_{(1)} \geq \cdots \geq e_{(m)}$,
& reject $e_{(1)}, \ldots, e_{(k)}$ for largest k such that $e_{(k)} \geq m/(\alpha k)$.

[9] Vovk & Wang 2021, *E-values: calibration, combination and applications*

[10] Wang & Ramdas 2022, *False discovery rate control with e-values*

Related work: BH with e-values

Random variable $e_j \geq 0$ is an e-value[9] for null hypothesis $H_{0,j}$ if

$$\mathbb{E}[e_j] \leq 1 \text{ for all } j \in \mathcal{H}_0.$$

e-BH procedure[10] for FDR control at level α

Rank $e_{(1)} \geq \cdots \geq e_{(m)}$,

& reject $e_{(1)}, \ldots, e_{(k)}$ for largest k such that $e_{(k)} \geq m/(\alpha k)$.

Theoretical guarantee: $\text{FDR} \leq \alpha$

- Arbitrary dependence between $e_1, \ldots, e_m$
- Sufficient to assume *compound e-values*:

$$\frac{1}{m} \sum_{j \in \mathcal{H}_0} \mathbb{E}[e_j] \leq 1.$$

[9] Vovk & Wang 2021, *E-values: calibration, combination and applications*

[10] Wang & Ramdas 2022, *False discovery rate control with e-values*

Summary & questions

Our results for BH on compound p-values:

- Under independence, at most constant inflation of FDR,
  but under PRDS, can have $\asymp \log m$ inflation
- New examples & constructions of compound p-values

Summary & questions

Our results for BH on compound p-values:

- Under independence, at most constant inflation of FDR,
  but under PRDS, can have $\asymp \log m$ inflation
- New examples & constructions of compound p-values

Open questions:

- Conditions (different from PRDS) to ensure FDR $\leq \mathcal{O}(\alpha)$?
- A local FDR framework for compound p-values?