DSO 699: Statistics Theory
Special Topics in Data Sciences and Operations

Week 5

Adel Javanmard

USC Marshall
Department of Data Sciences and Operations

September 22, 2023

Announcements

- HW1 is due next week (Sep 29)
- Midterm exam is in two weeks (Oct 6)
  - ✓ in-class exam
  - ✓ you can have access to slides
  - ✓ no internet

Recap from the previous class

We talked about:

- Connection/difference between global testing and multiple hypotheses testing
- Family-wise Error Rate (FWER): $\mathbb{P}(V \geq 1)$
  - ✓ Strong control
  - ✓ Weak control
- Methods for controlling FWER
  - ✓ Bonferroni's method
  - ✓ Šidák's method
  - ✓ Holm's step-down procedure
- The closure principle (valid global testing $\rightarrow$ FWER control in multiple hypotheses testing)

Outline for today

False Discovery Rate and Benjamini-Hochberg (BH) Procedure

False discovery proportion (FDP)

Outcomes in multiple testing:

	Fail to reject	Rejected	Total
True null	U	V	n_0
False null	T	S	$n - n_0$
Total	$n - R$	R	n

- We learned that FWER becomes too stringent in large scale hypothesis testing (With FWER$\leq \alpha$ the power goes down as $\#$ hypotheses grows)
- **False discovery proportion (FDP):**

$$\mathrm{FDP} = \frac{V}{\max(R, 1)} = \begin{cases} \frac{V}{R} & R \geq 1 \\ 0 & R = 0 \end{cases}$$

- FDP is unobservable. Why?
- FDP is a random quantity
- $0 \leq \mathrm{FDP} \leq 1$.

False discovery rate (FDR)

V, R are usually correlated
controling expected ratio of correlated can be challenging

- We can try to control expectation of FDP:

$$\mathrm{FDR} = \mathbb{E}[\mathrm{FDP}] = \mathbb{E}\Big[\frac{V}{\max(R,1)}\Big]$$

FWER control implies FDR control

[Benjamini-Hochberg '95]

- FDR is a weaker condition than FWER and allows for more powerful tests. Why?

$$\mathbb{I}(V \geq 1) \geq \frac{V}{\max(R,1)}$$

0 or 1 rate, so from 0 to 1.

=0 implies rhs 0, if =1 than the ratio is less or equal to 1.

- FDR gives us control *on average* across many repetitions of an experiment. Sometimes, FDP may have large variation (sizably larger than its expectation)
- False discovery exceedance:

$$\mathrm{FDX}_\gamma = \mathbb{P}(\mathrm{FDP} \geq \gamma)\,.$$

by markov inequality, we can show this is stringent than FDR

Benjamini-Hochberg (BH) procedure

1. Sort p-values $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(n)}$
2. Let $i_0 := \max\{i \in [n] : \ p_{(i)} \leq \frac{i}{n}\alpha\}$
3. Reject all p-values $p_{(i)}$ for $i \leq i_0$ (or $p_j \leq p_{(i_0)}$)

R is random, cutoff is function of R so

1) cut off is random

2)cutoff and R are correlated

Some observations:

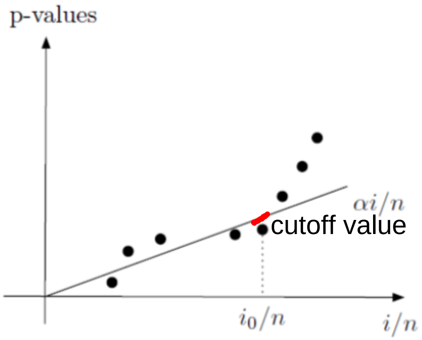
- BH cutoff depends on all p-values
- The cutoff and the total number of rejections are explicitly related:
  i_0 : total $\#$ of rejections
  cutoff:= $\frac{\alpha i_0}{n}$ explicit expression for cutoff cannot avoid dependence on i_0
- BH cutoff is a "random" quantity!

loop? this makes proof difficult

but positive feedback.

reject more (by randomness of data)-> higher cutoff-> reject more encourage more rejection

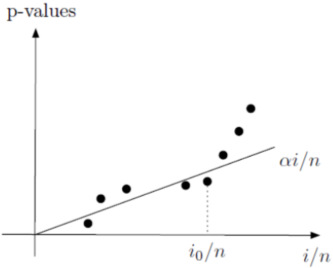
An illustration



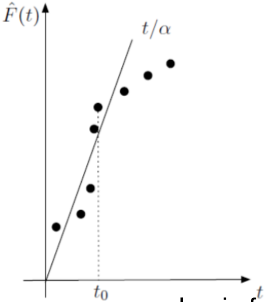
BH cutoff is shown by dashed line.

we use last index,
so we have to check all p-value

The Empirical Process Viewpoint of BH



(a) P-values on the y axis, indices on x



p-value is from 0 to 1

(b) P-values on the x axis, indices on y

moving along cutoff, check if cdf is over me

Recall the empirical CDF $\widehat{F}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(p_i \leq t)$. support = [0,1]

We then have $\widehat{F}(p_{(i)}) = \frac{i}{n}$.

The Empirical Process Viewpoint of BH (cont'd)

Recall that BH procedure rejects $p_{(1)}, \ldots, p_{(i_0)}$ where

$$i_0 = \max\left\{i: \ p_{(i)} \le \frac{\alpha i}{n}\right\}$$

The critical p-value $p^* = p_{(i_0)}$ can alternatively be written as

$$\begin{aligned}
p^* &= \max\left\{p_{(i)}: \ p_{(i)} \le \frac{\alpha i}{n}\right\} \\
&= \max\left\{p_{(i)}: \ p_{(i)} \le \alpha \widehat{F}(p_{(i)})\right\} \\
&= \max\left\{t \in \{p_1, \ldots, p_n\}: \ t \le \alpha \widehat{F}(t)\right\}
\end{aligned}$$

or max_t t/a < F(t)
max cdf that goes over the
linear line

So the BH cutoff can be written as

$$\tau_{BH} = \max\left\{t: \ \frac{t}{\widehat{F}(t)} \le \alpha\right\}$$

t need not be discrete, rejection results are the same
anyway empirical cdf is step function

Is it a natural choice?

The Empirical Process Viewpoint of BH (cont'd)

Is it a natural choice for cutoff?

n_0 is fixed constant

For given $t \in (0,1)$ consider a rule that rejects H_i iff $p_i \leq t$.

cheating

$$\mathrm{FDR}(t) = \mathbb{E}\Big[\frac{V(t)}{R(t)}\Big] \approx \frac{E[V(t)]}{R(t)}$$

linearity of E. dependence doesn't matter for E

E indicator = prob

prob of rejecting null = t

how many null? n_0

Fhat: proportion of p-values less than t

n Fhat = number of ...

$$= \frac{n_0 t}{n\widehat{F}(t)}$$

$$\leq \frac{nt}{n\widehat{F}(t)} := \widehat{\mathrm{FDR}}(t)$$

loose upper bound

So the BH cutoff can be written as

$$\tau_{BH} = \max\Big\{t : \frac{t}{\widehat{F}(t)} \leq \alpha\Big\}$$

$$= \max\Big\{t : \widehat{\mathrm{FDR}}(t) \leq \alpha\Big\}.$$

control loose upper bound by alpha which then controls FDR

BH procedure controls FDR

wrong proof:
E(V/R) = E(V) / E(R) (first mistake)
E I (p_i < a R /n) = a R/n (second mistake - R is r.v. correlated with p_i so it's not uniform prob. even conditioning on R does not make it uniform)

Theorem

If $\{p_1, \ldots, p_n\}$ are mutually independent then the BH procedure at level α satisfies

$$\mathrm{FDR} = \frac{n_0}{n}\alpha \leq \alpha$$

There are several elegant proofs for this theorem:

- Induction (Very first proof by Benjamini-Hochberg '95)
- Martingale theory
- Leave-one-out

  idea: p_i and R is correlated but not highly correlated
  e.g. if p_i is not the boundary, its value does not matter to R

  This technique is useful for weak correlation situation

Proof of FDR control

Here i is not ordered i.e. not (i)

Let $\mathcal{H}_0$ be the set of null p-values.

V_i is random variable

$$\text{FDP} = \frac{V}{\max(R,1)} = \sum_{i \in \mathcal{H}_0} \frac{V_i}{\max(R,1)}, \qquad V_i = \mathbb{I}(p_i \le \alpha R/n).$$

Let's think for individual term

(2)V_i = I[i reject]=

- *Challenge:* V_i and R are correlated.

I[i rej, R<k] U [i rej, R=k] U [i rej, R>k]

= sum of I

(1)somehow randomness moved to numerator

$$\frac{V_i}{\max(R,1)} = \sum_{k=1}^{n} \frac{V_i \mathbb{I}(R=k)}{k} = \sum_{k=1}^{n} \frac{\mathbb{I}(p_i \le \alpha k/n)\mathbb{I}(R=k)}{k}$$

transform this three times

only one term is true (nonzero)

We still have correlation between R and p_i (but a weak correlation!)

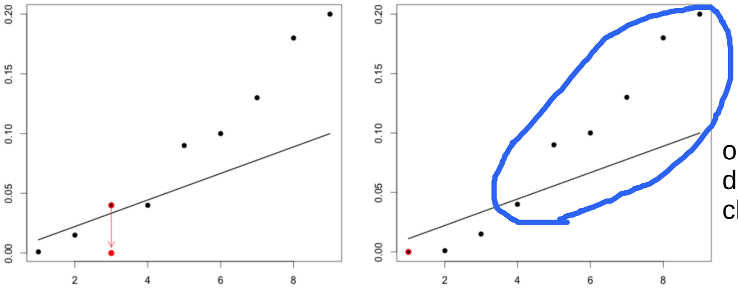
possible values:
0, 1, 1/2, 1/3, ...1/n

Proof of FDR control (cont'd)

Leave-one-out technique:

If p_i is rejected its actual value is not relevant to R as far as it is under the cutoff. i.e. number of the rejection doesnt change

(How about setting it to zero?)



order here doesnt change

this is random variable. we have i number of this kind of thing

Define $R(p_i \to 0)$ the number of rejection if you set p_i to zero and keep the others intact

Let $\mathcal{F}_i$ be (the σ-algebra) generated by $\{p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_n\}$.

Proof of FDR control (cont'd)

conditional expectation
is a random variable

$$\mathbb{E}\left(\frac{V_i}{\max(R,1)}|\mathcal{F}_i\right) = \sum_{k=1}^{n}\left(\frac{\mathbb{I}(p_i \le \alpha k/n)\,\mathbb{I}(R=k)}{k}\Big|\mathcal{F}_i\right)$$

(3) same value.
split the right term
using intersection with
left

$$= \sum_{k=1}^{n}\left(\frac{\mathbb{I}(p_i \le \alpha k/n)\,\mathbb{I}(R(p_i \to 0)=k)}{k}\Big|\mathcal{F}_i\right)$$

independece
-> E(XY)=E(X)E(Y)
& conditioned on Fi, right term is constant
& p_i is indep of other p-values
so conditioing meaningless

$$= \sum_{k=1}^{n}\frac{\mathbb{I}(R(p_i \to 0)=k)}{k}\mathbb{P}(p_i \le \alpha k/n)$$

$$= \frac{\alpha}{n}\sum_{k=1}^{n}\mathbb{I}(R(p_i \to 0)=k)$$

only one indicator
is 1

$$= \frac{\alpha}{n}.$$

(where did we use independence of p_i's?)

The proof is completed by taking another expectation (w.r.t $\mathcal{F}_i$) and summing over $i \in [n]$.

A few remarks

- BH cutoff is adaptive! (it depends on all p-values and not just a function of α and n)
- There is an inherent positive feedback: cutoff $\rightleftharpoons R$
  (Burden of proof decreases if you make more rejections)
- Under global null we have FDR = FWER (why?)

any rejection is false rejection. V=R. Then FDP = 0 (V=0) or 1(V>1) i.e. indicator random variable.
so E indicator = prob.

Comparison with Sime's: Under global null:

$$\mathrm{FWER} = \mathrm{FDR} = \frac{\alpha n_0}{n} = \alpha$$

By description of BH cutoff: BH make at least one rejection <-> simes test

$$\mathrm{FWER} = \mathbb{P}(V \geq 1) = \mathbb{P}(\exists i : \; p_{(i)} n/i \leq \alpha)$$

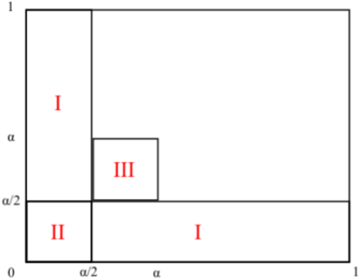
Putting together, it proves validity of Simes' procedure for global testing!

BH under Dependence

A toy example

Consider $n = 2$, $p_1, p_2 \sim \text{Unif}(0,1)$ (arbitrarily correlated).

marginal is still uniform



under global null,
same as simes

under global null, =prob of having at least one rejection (since it's global null)

$$\mathrm{FDR} = \mathrm{FWER} = \mathbb{P}(\mathrm{I}) + \mathbb{P}(\mathrm{II}) + \mathbb{P}(\mathrm{III})$$

this is not 0, due to dependence

a = (I+II)+(I+II)
so subtract II once

$$= \alpha + \mathbb{P}(\mathrm{III}) - \mathbb{P}(\mathrm{II}) \leq \alpha + \mathbb{P}(\mathrm{III})$$

don't think as ith rejection, think of as global testing

so all we can do
is upper boudning
by removing - II

Also,

$$\mathbb{P}(\mathrm{III}) = \mathbb{P}(\alpha/2 \leq p_1, p_2 \leq \alpha) \leq \mathbb{P}(\alpha/2 \leq p_1 \leq \alpha) = \alpha/2$$

a/2 < p1 < a AND a/2<p_2<a

upper bounding by marginal

So FDR$\leq 3\alpha/2$. under general dependence

we can calculate marginal
because marginal is uniform

Two theorems

Recall the harmonic sum

$$h_n := 1 + \frac{1}{2} + \ldots + \frac{1}{n} \approx \log n$$

Theorem (Benjamini and Yekuteli 2001)– upper bound

Under arbitrarily dependence among p-values, BH procedure at level α has the following control of FDR:

$$\mathrm{FDR} \le \alpha h_n \frac{n_0}{n} \le \alpha h_n .$$

still controlled, but less performance

Theorem (Guo and Rao 2008)– lower bound

which means that the above bound is tight

For any $n \ge 1$, there exists a joint distribution of p-values such that the FDR of BH at level α satisfies

$$\mathrm{FDR} \ge \min(1, \alpha h_n) .$$

Proof: (Whiteboard)

Storey's Procedure

Improving BH

Recall the empirical process view point of BH:

$$\tau_{BH} = \max\left\{t : \frac{t}{\widehat{F}(t)} \leq \alpha\right\}.$$

We were to able to exactly characterize FDR of BH(α) procedure as

$$\mathrm{FDR} = \frac{\alpha n_0}{n}$$

What if we could estimate n_0?

we can set larger threshold; by replacing n with n_0

An estimate of π_0

what we have:
if null, p-values are uniform
if not null, p-value would be generally small

Let $\pi_0 = n_0/n$ be the fraction of nulls.
Pick $\lambda \in (0,1)$ and compute   usually large lambda to get many samples

large p-values i.e. nulls

$$\widehat{\pi}_0^{\lambda} = \frac{\sum_{i=1}^{n} \mathbb{I}(p_i > \lambda)}{n(1-\lambda)} \tag{1}$$

Why is this estimate sensible?

n_0 (1-\lambda) ~~ \sum 1(_i > \lambda)
parameter   estimate

Storey's procedure

1. Pick $\lambda \in (0, 1)$ (typically $\lambda = 1/2)^{\text{larege lambda}}$
2. Estimate null proportions:

$$\widehat{\pi}_0^\lambda = \frac{1 + \sum_{i=1}^n \mathbb{I}(p_i > \lambda)}{n(1-\lambda)} \tag{2}$$

3. Similar to BH construct the cutoff:

$$\tau_{\text{storey}} = \max\left\{0 \leq t \leq \lambda : \frac{\widehat{\pi}_0^\lambda t}{\widehat{F}(t)} \leq \alpha\right\}.$$

- We only consider $0 \leq t \leq \lambda$ because in the estimate $\widehat{\pi}_0^\lambda$ we implicitly assume p-values above λ are null.

Storey's procedure

Theorem (Storey 2004)

If the p-values are independent, then for $\lambda \in (0,1)$ the Storey's procedure controls FDR as

$$\mathrm{FDR} \le (1-\lambda^{n_0})\alpha \le \alpha\,.$$

Recap

We talked about

- False discovery rate (FDR), false discovery proportion (FDP) and false discovery exceedance
- Benjamini-Hochberg procedure
  - ✓ for dependent p-values
  - ✓ for arbitrarily correlated p-values
- Storey's procedure

Don't forget about:

- HW1 (due next week)
- In-class midterm (in two weeks)

FIGHT ON!