

# Stats 506 F20, Final Project

Zhihao Xu, xuzhihao@umich.edu

December 05, 2020

## Topic

Performance of Benjamini-Hochberg Adjustment in an atypical situation

## Introduction

In the multiple hypothesis testing problem, usually we care about the False Discovery Rate(FDR) instead of Type 1 Error. Benjamini-Hochberg procedure is commonly used to control the FDR in multiple hypothesis testing.

### Benjamini-Hochberg procedure:

Begin by ordering the p-values in ascending order.

$$p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(n)},$$

Fix a significant level  $q \in [0, 1]$ . Let  $i_0$  be the largest  $i$  for which

$$p_{(i)} \leq \frac{i}{n}q.$$

Reject all  $H_{(i)}$  with  $i \leq i_0$ . This procedure is proved in Benjamini & Yekutieli (2001). If the joint distribution of the statistics (or joint dist. of the p-values) is PRDS(positive regression dependency on each one from a subset) on the set of true nulls  $H_0$ , then the Benjamini-Hochberg procedure BH( $q$ ) controls the FDR at level  $q$ . However, in practise, we do not always know the precise correlation structure. In Benjamini & Yekutieli (2001), it also introduced a BY-Adjust procedure, which use rejection threshold  $q = q / \sum_{i=1}^n \frac{1}{i}$  to control the FDR at level  $q$  regardless of correlation structure. However, this procedure is quite conservative. In this project, I want to demonstrate the performance of original BH procedure on various correlation structure.

## Experiment Setting

Consider  $n$  i.i.d. samples  $x_t, t = 1, \dots, n, x_t = [x_{t1}, \dots, x_{tm}] \in R^m$ . Suppose  $x_{ti} \sim N(\mu_i, 1)$ . We want to test the null hypothesis  $H_{0,i} : \mu_i = 0$  for  $i = 1, \dots, m$ . Consider the following setting:

- (a) **i.i.d.**:  $x_{ti}$  for  $i = 1, \dots, m$  are i.i.d.
- (b) **Positive Correlated**:  $Cov(x_{ti}, x_{tj}) = 0.8$  for  $i \neq j$ .
- (c) **Negative Correlated**:  $Cov(x_{ti}, x_{tj}) = 0.8$  for  $1 \leq i < j \leq m/2, x_{t,i+m/2} = -x_{t,i}$  for  $1 \leq i \leq m/2$ .

Here we apply the BH-Adjust and BY-Adjust at level  $q = 0.1$  by conducting  $B = 1000$  simulations with  $n = 200$  and  $m = 100$  in each setting, and compare the FDR and power of these adjustment procedure. We will try the following cases:

- 1. Global Null:  $\mu_i = 0$
- 2. Sequential True Alternatives:  $\mu_1 = \mu_2 = 0.2, \mu_3 = \mu_4 = 0.4, \mu_5 = \mu_6 = 0.6, \mu_7 = \mu_8 = 0.8, \mu_9 = \mu_{10} = 1$ , and  $\mu_i = 0$  for  $10 < i < 50$ .

Here we test  $H_{0,i} : \mu_i = 0$  at significance level  $\alpha = 0.05$ , compute p-value by

$$p = 2 \left[ 1 - \Phi \left( \left| \frac{\bar{X}_{\cdot, i} - 0}{1/\sqrt{n}} \right| \right) \right]$$

## Simulation Result

Global Null:  $\mu_i = 0$

```
col1 = rep(c("Independent", "Positive Correlation", "Negative Correlation"),
           each = 2)
adj = rep(c("BH", "BY"), 3)
result_null = cbind(adj, col1, sprintf("%.3f", result_null))
colnames(result_null) = c("Method", "Correlation Structure", "FDR")
row.names(result_null) = NULL
knitr::kable(result_null, align="c")
```

Method	Correlation Structure	FDR
BH	Independent	0.058
BY	Independent	0.010
BH	Positive Correlation	0.035
BY	Positive Correlation	0.003
BH	Negative Correlation	0.023
BY	Negative Correlation	0.005

Sequential True Alternatives:  $\mu_1 = \mu_2 = 0.2, \mu_3 = \mu_4 = 0.6, \mu_5 = \mu_6 = 1$ , and  $\mu_i = 0$  for  $6 < i \leq 50$ .

```
result_sig = cbind(adj, col1, sprintf("%.3f", result_sig[,1]), sprintf("%.3f", result_sig[,2]))
colnames(result_sig) = c("Method", "Correlation Structure", "FDR", "power")
row.names(result_sig) = NULL
knitr::kable(result_sig, align="c")
```

Method	Correlation Structure	FDR	power
BH	Independent	0.047	0.812
BY	Independent	0.010	0.749
BH	Positive Correlation	0.024	0.807
BY	Positive Correlation	0.005	0.753
BH	Negative Correlation	0.029	0.851
BY	Negative Correlation	0.002	0.776

## Conclusion & Discussion

Through the above simulation result, we can see that although the negative correlation structure does not satisfy the PRDS assumption, we can also use BH method to control FDR in some cases. BY method can always control FDR successfully but will also leads to conservative rejection. In the above simulation, the FDR got by the BY-method is only 1/10 of our target, which will also cost the decrease in power. Hence, if the correlation structure is unknown, it is also reasonable to use BH method, and we can also use BY-method as a base line method.