

Stats 506 F20, Final Project

Zhihao Xu, xuzhihao@umich.edu

December 14, 2020

Topic

Performance of Benjamini-Hochberg Adjustment in an atypical situation. Click link here to view the Github Repo of this project.

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 in multiple hypothesis testing to control the FDR.

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 Benjamini-Yekutieli procedure. Under dependence, the BHq procedure controls at level $q \cdot S(n)$. In fact, $FDR \leq qS(n) \cdot \frac{n_0}{n}$, where $S_n = 1 + 1/2 + 1/3 + \cdots + 1/n$. Hence, we can use $q / \sum_{i=1}^n \frac{1}{i}$ to take the place of q in Benjamini-Hochberg Procedure, which can always 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, and compare the FDR and power of BH procedure and BY procedure.

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

1. Global Null: $\mu_i = 0$

Table 1: Simulation Result under Global Null

Method	Correlation Structure	FDR
BH	Independent	0.049
BY	Independent	0.011
BH	Positive Correlation	0.033
BY	Positive Correlation	0.004
BH	Negative Correlation	0.025
BY	Negative Correlation	0.004

2. 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 $6 < i \leq 50$.

Table 2: Simulation Result for Sequential True Alternatives

Method	Correlation Structure	FDR	power
BH	Independent	0.044	0.816
BY	Independent	0.012	0.753
BH	Positive Correlation	0.021	0.817
BY	Positive Correlation	0.003	0.737
BH	Negative Correlation	0.025	0.843
BY	Negative Correlation	0.005	0.765

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 lead to 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.