Who Is Better? The Comparison of Different Procedures and Their FDR Control

Chaoan Li[*]

December 3, 2024

Abstract

As [Li and Zhang, 2024] built a unified framework of Benjamini-Hochberg (BH) procedure, Barber and Candès (BC) procedure [2015], Storey's (ST) procedure [2002], the general version Flexible BH (FBH) procedure as well as Flexible BC (FBC) procedure proposed in their paper. The author continues this research by studying the performance of all these five methods on a specific problem and proposes a reject score to guide the selection of test procedures.

Keywords: Benjamini-Hochberg procedure; Barber-Candès procedure; False discovery rate; Reject score

1 Introduction

When dealing with high-dimensional data such as the gene sequence, it is crucial to ensure that there are as few false signals as possible in the signals we detect. This usually means to control the false discovery rate (FDR). The Benjamini-Hochberg (BH) procedure is a widely used FDR-controlling procedure which rejects the hypothesis when the p-values is less than or equal to the threshold determined by the whole set of p-values. What's more, Barber and Candès proposed a model-free (BC) procedure which build the threshold by the symmetry of the p-values under the null.

Recently, Li and Zhang [2024] utilizing the e-values for the connection between the classical BH procedure and the e-BH procedure proposed in [Wang and Ramdas, 2022]. They generalized the BH and BC procedures by flexible functions why might change in different hypothesis. Also, they provided a unified viewpoint of e-BH procedure. Thus, a natural question arises, how do different methods perform under this unified framework? Who has the best FDR control?

In this work, the performance of different procedures are studied in a simple example that the null hypothesis and the alternative hypothesis are both normal distributions with different means. The author choose five representative functions in FBH and FBC procedure and nine different fixed parameters λ in ST procedure. By comparing the performance

[*]Email: chaoan@tamu.edu

of these methods, a reference is provided for selecting a specific selection method. Additionally, the author found that some methods were a bit too conservative to avoid making mistakes and missed many useful signals. So the author proposed the concept of Reject Score to reward the discovery of true positive signals. Code for reproducing experiment in this article is available at https://github.com/ChaoanLi/E-value.

2 Research Problem

Suppose there are 10,000 genes, 1% of them are truly consisted with a certain disease. In the test results, genes not related to this disease follow the standard normal distribution, genes consisted with this disease follow the normal distribution with a mean shift of 5. That is for testing $n = 10,000$ hypotheses $(H_1, \dots, H_n)$ the random variable of the result $(X_1, \dots, X_n)$, where

$$H_{0,i} \sim N(0,1), \quad H_{a,i} \sim N(5,1), \quad i \in \{1, \cdots, n\}.$$

Let $\theta = (\theta_1, \dots, \theta_n) \in \{0,1\}^n$ and $\delta = (\delta_1, \dots, \delta_n) \in \{0,1\}^n$ indicate the underlying truth and the decision rule, respectively. The FDR for the decision is defined as follows,

$$FDR(\delta) := \mathbb{E}[FDP(\delta)], \quad FDP(\delta) := \frac{\sum_{i=1}^{n}(1-\theta_i)\delta_i}{1 \vee \sum_{i=1}^{n} \delta_i}.$$

In this work, we expect to control the FDR by a real constant $\alpha \in (0,1)$. Moreover, in order to encourage the algorithm finding more signals under the FDR control. This paper heuristically introduces a Reject Score Indicator (RSI) by giving a reward $\beta_r \in (0,1)$ for every right rejection and punish $\beta_p := 1 - \beta_r$ for every false rejection,

$$RSI(\delta) := \frac{1}{1 + e^{\beta_p \sum_{i=1}^{n}(1-\theta_i)\delta_i - \beta_r \sum_{i=1}^{n} \theta_i \delta_i}}.$$

It is easy to proof that:

i. $RSI(\delta) \in (0,1)$;

ii. when $\sum_{i=1}^{n}(1-\theta_i)\delta_i$ is fixed, $RSI(\delta)$ is getting close to 1 if the right rejection grows;

iii. when $\sum_{i=1}^{n} \theta_i \delta_i$ is fixed, $RSI(\delta)$ is getting close to 0 if the false rejection grows.

Thus, we expect to get a as higher as possible RSI under the controlling of FDR. In this paper, we set $\alpha = \beta_r = 0.05$.

3 Methods

According to Li and Zhang's work [2024], we reject the ith hypothesis if $R_i(T) = 1$ with the threshold

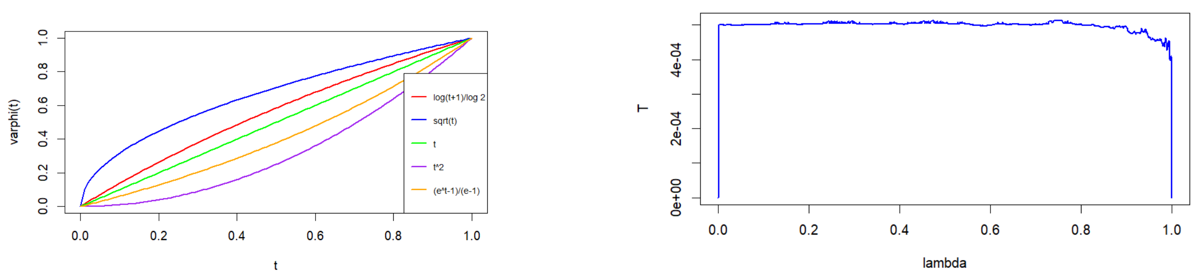
$$T := \sup\left\{ t \in \mathcal{D} : \ \frac{m(t)}{1 \vee \sum_{j=1}^{n} R_j(t)} \leq \alpha \right\}.$$

Here $\mathcal{D}$ is the domain of T, $m(t)$ is an estimate of false discoveries, and $\sum_{j=1}^{n} R_j(t)$ is the number of rejections.

In this work, we selected five representative flexible functions that all increase continuously and strictly between the starting point $(0,0)$ and the end point $(1,1)$, with different growth rates, including logarithmic, square root, functions, square, and exponential functions:

$$\varphi(t) \in \left\{ \frac{\log(t+1)}{\log 2}, \ \sqrt{t}, \ t, \ t^2, \ \frac{e^t - 1}{e - 1} \right\}.$$

We plot the relationship in Figure 1a. For the ST procedure, we found that the threshold T



(a) Flexible functions on $(0, 1)$.

(b) Threshold T in ST procedure.

Figure 1

has a sharp increase near 0; see Figure 1b. To capture this change, we took six different $\lambda \in \{0, 0.0001, \ldots, 0.0006\}$. For the flexible BH procedure, by [Li and Zhang, 2024, Proposition 2] with $g(t) := h(t) = \varphi^{-1}(t)$, controls the FDR at level α, where φ is a strictly increasing function which can differ for each i but we treat as same for simplicity. Likewise, by [Li and Zhang, 2024, Proposition 4] flexible BC procedure ensures FDR control at level α with a monotonic increasing and continuous function φ. For The selections of $\mathcal{D}$, $m(t)$ and $R(t)$ are shown in Table 1.

method	$\mathcal{D}$	$m(t)$	$R_i(t)$
BH	$(0,1]$	nt	$\mathbf{1}\{p_i \leq t\}$
BC	$(0,0.5)$	$1+\sum_{i=1}^{n}\mathbf{1}\{p_i \geq 1-t\}$	$\mathbf{1}\{p_i \leq t\}$
FBH	$(0,1]$	$n\varphi^{-1}(t)$	$\mathbf{1}\{\varphi(p_i) \leq t\}$
FBC	$(0,0.5]$	$1+\sum_{i=1}^{n}\mathbf{1}\{\varphi(1-p_i) \leq t\}$	$\mathbf{1}\{\varphi(p_i) \leq t\}$
ST	$(0,\lambda]$	$(1+n-\sum_{i=1}^{n}\mathbf{1}\{p_i \leq \lambda\})t/(1-\lambda)$	$\mathbf{1}\{p_i \leq t\}$

Table 1: The selections of $\mathcal{D}$, $m(t)$, $R_i(t)$ for different methods.

4 Results

We use subscripts to denote the parameters of methods. The comparison of methods with threshold T, FDR, and RSI are shown in Table 2, ranked in RSI. Because the reward we set is not high, all methods with high RSI can effectively control FDR. However, RSI is not inversely proportional to FDR.

When choosing the appropriate lambda value or the appropriate flexible function, ST procedure, FBH, and FBC have excellent performance. At the same time, these methods will perform poorly if the correct choice is the opposite. In the setting of this article, there is strong a symmetry, so the BC procedure performs much better than the BH procedure. The distribution of true positives and false positives in the p-values rejected by these methods can be seen in Figure 2

Method	$T*10^4$	**FDR**	**RSI**	**Method**	$T*10^4$	**FDR**	**RSI**
ST_1e-04	1	0.006	0.9	FBC_log	4.649	0.019	0.25
FBH_square	0.003	0	0.891	ST_4e-04	4	0.019	0.25
FBC_square	0.001	0	0.864	ST_6e-04	4.996	0.022	0.114
ST_2e-04	2	0.009	0.839	ST_5e-04	4.996	0.022	0.114
FBC_exp	1.876	0.009	0.832	BH	4.95	0.022	0.114
ST_0	0	0	0.5	FBH_log	7.139	0.025	0.05
BC	3.223	0.016	0.45	FBC_sqrt	179.5	0.564	0.000
ST_3e-04	3	0.016	0.45	FBH_sqrt	218.4	0.685	0.000
FBH_exp	2.761	0.016	0.45				

Table 2: Comparison of Methods with Threshold T, FDR, and RSI

5 Conclusions

This paper aims to study the performance of different procedures in controlling FDR. In statistics, our real goal is to obtain more useful signals rather than simply control FDR

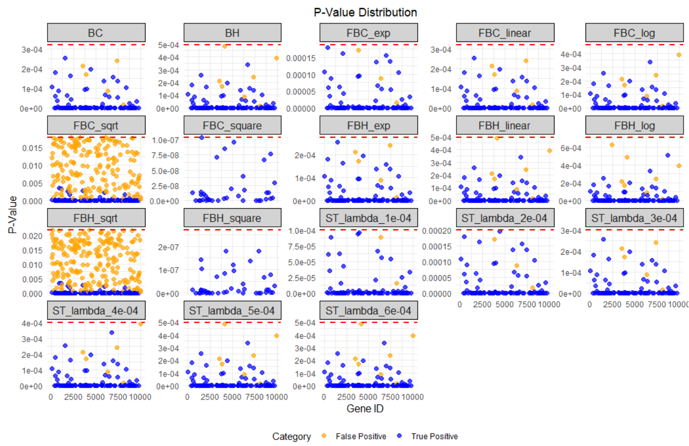


Figure 2: Comparison of Methods with False/True Positive

(if that is the case, then the method that does not reject is the best method). The work in this paper is relatively superficial, the problem background is relatively simple, and the flexible functions and lambda indicators selected are not quite enough. The author intends to observe the performance of these methods on a small scale. Therefore, readers can use this work as inspiration to specify new standards to achieve the effect of RSI. In addition, AI and other training methods can be used to select appropriate flexible functions and lambda values in more problem contexts.

References

Rina Foygel Barber and Emmanuel J Candès. Controlling the false discovery rate via knockoffs. *The Annals of statistics*, pages 2055–2085, 2015.

Guanxun Li and Xianyang Zhang. A note on e-values and multiple testing. *Biometrika*, 2024. to appear.

John D Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 64(3):479–498, 2002.

Ruodu Wang and Aaditya Ramdas. False discovery rate control with e-values. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(3):822–852, 2022.