NEW RESULTS ON THE FALSE DISCOVERY RATE

A Dissertation
Submitted to
the Temple University Graduate Board

---

in Partial Fulfillment
of the Requirements for the Degree of
DOCTOR OF PHILOSOPHY

---

by
Fang Liu
June, 2010

Examining Committee Members:

Sanat K. Sakar, Advisory Chair, Statistics
Damaraju Raghavarao, Statistics
Milton Parnes, Statistics
Keith A. Soper, External Member, Merck & Co., Inc.

ABSTRACT

NEW RESULTS ON THE FALSE DISCOVERY RATE

Fang Liu

DOCTOR OF PHILOSOPHY

Temple University, June, 2010

Professor Sanat K. Sakar, Chair

The false discovery rate (FDR) introduced by Benjamini and Hochberg (1995) is perhaps the most standard error controlling measure being used in a wide variety of applications involving multiple hypothesis testing. There are two approaches to control the FDR - the fixed error rate approach of Benjamini and Hochberg (1995) where a rejection region is determined with the FDR below a fixed level and the estimation based approach of Storey (2002) where the FDR is estimated for a fixed rejection region before it is controlled. In this dissertation, we concentrate on both these approaches and propose new, improved versions of some FDR controlling methods available in the literature.

A number of adaptive methods have been put forward in the literature, each attempting to improve the method of Benjamini and Hochberg (1995), the BH method, by incorporating into this method an estimate of n_0, the unknown number true null hypotheses. Among these, the method of Benjamini, Krieger and Yekutieli (2006), the BKY method, has been receiving a lot of attention recently. In this dissertation, a variant of the BKY method is pro-

posed by considering a different estimate of n_0, which often outperforms the BKY method in terms of the FDR control and power.

Storey's (2002) estimation based approach to controlling the FDR has been developed using a class of conservatively biased point estimates of the FDR under a mixture model for the underlying p-values and a fixed rejection threshold for each null hypothesis. An alternative class of point estimates of the FDR with uniformly smaller conservative bias is proposed under the same setup. Numerical evidence is provided to show that the mean squared error (MSE) is also often smaller for this new class of estimates. Storey's (2002) q-value method has been modified based on this new class of point estimates of the FDR, which appears to be more powerful than Storey's original q-value method.

ACKNOWLEDGEMENTS

First and foremost I offer my sincerest gratitude to my advisor, Dr. Sanat K. Sarkar, who has supported me thoughout my thesis with his patience and knowledge whilst allowing me the room to work in my own way. His insights into many aspects of multiple testing and logical way of thinking have been of great value for me. Without him, this thesis would not have been completed or written. I am indebted to him more than he knows.

Besides my advisor, I am deeply grateful to the rest of my thesis committee members: Dr. Damaraju Raghavarao, Dr. Milton Parnes and Dr. Keith Soper. Their kind support and guidance have been of great value in this study and have made this dissertation definitely better.

A great deal of thanks goes to all the professors in Statistics Department, Temple University, especially those who have taught me statistics: Drs. Richard M. Heiberger, Burt Holland, Francis Hsuan, Boris Iglewicz, Damaraju Raghavarao and Jagbir Singh. I would like to thank Dr. Woollcott Smith for the offer of department scholarships, which had supported me during my first three years of study.

I would like to express my special thanks to my friends and my family for their unconditional support and encouragement, for listening to my complaints and frustration, and for believing me.

The financial support of this research by NSF Grant DMS-0603868 is gratefully acknowledged.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Recently, scientific advances made in different areas, such as biology, astronomy, economics and finance, have led to experiments that often require thousands of hypotheses to be simultaneously tested. As a result, multiple testing has become a topic of renewed interest and undergone tremendous development in the recent decades. A large part of this development has taken place since the introduction of the False Discovery Rate (FDR) by Benjamini and Hochberg (1995). The FDR (defined in Section 2.1) is a notion of Type-I error rate while testing multiple null hypotheses that is more preferable to the traditional familywise (Type-I) error rate, the FWER, as the FWER is too conservative when the number of hypotheses is large. The primary goal of this research project is to further develop the theory and methodologies of the FDR.

A number of methods controlling the FDR are available in the literature. These have been developed either (i) taking the so called fixed error rate approach [Benjamini and Hochberg (1995)] where a rejection region is determined

with the error rate, in this case the FDR, less than a fixed level, or (ii) taking the estimation based approach [Storey (2002)] where the FDR is estimated considering the rejection region to be fixed before controlling it at the desired level. In this dissertation, we revisit two of such methods, one based on the fixed error rate approach and the other estimation based, and propose some improved versions of them.

The method of Benjamini and Hochberg (1995), the BH method (see, Section 3.1 for details), is the most popular fixed error rate approach to control the FDR. For multiple testing of n null hypotheses using their respective p-values, the FDR of this method, given the level α at which it is to be controlled, is exactly $n_0\alpha/n$, where n_0 is the unknown number of true null hypotheses, when the p-values are independent and is less than or equal to $n_0\alpha/n$ when these p-values are positively dependent in a certain sense [Benjamini and Hochberg (1995), Benjamini and Yekutieli (2001) and Sarkar (2002b)]. Since n_0 is unknown, estimating it and then modifying the BH method using this estimate can potentially make the FDR control of the BH method less conservative and thus more powerful. Such modifications are referred to as adaptive BH methods. A number of such methods have been proposed in the literature, each based on a particular type of estimate of n_0. Among these, the method in Benjamini, Krieger and Yekutieli (2006), to be referred to as the BKY method (see Section 3.6.3 for details), has received some attention recently.

We go back to the BKY method in our present research and investigate if it can be improved by considering a different method of estimating n_0 while modifying the BH method. Indeed, we have been able to construct an alternative version of the BKY method. As in the case of the BKY method, we establish

the FDR control of this newer method theoretically under the independence of the p-values and provide numerical evidence to show that it may control the FDR under normal distributional setting with equal positive correlation. In fact, as our simulations indicate, our proposed method can often outperform the BKY method in the sense of providing a better FDR control and being more powerful when the correlation between the test statistics is moderately low or the proportion of true null hypotheses is high. An application to a microarray data also indicates that the proposed method is more powerful than the BKY method. These results are due to appear in Liu and Sarkar (2010).

Storey's (2002) estimation based approach to controlling the FDR is based on a class of conservatively biased point estimates of the FDR (see Section 3.2 for details) that he developed assuming a mixture model for the p-values and considering a fixed rejection threshold for each hypothesis. We revisit this approach in our present research to investigate if the FDR controlling method it generates can be improved. Again, we have been able to show that an improvement is indeed possible. By focusing on estimation of the FDR under the same setup as in Storey (2002), we show that a different class of point estimates of the FDR can be formed, each of which has uniformly smaller conservative bias than the corresponding estimate in Storey (2002), and, as we numerically verify, the mean square error (MSE) is also quite often smaller. This class is actually generated by deriving the uniformly minimum variance unbiased estimate (UMVUE) of the FDR with a known n_0 and then substituting into this UMVUE an estimate of n_0. A similar class of estimates of the positive false discovery rate (pFDR) (See, Section 2.1 for definition) is then proposed, before modifying Storey's q-value method of controlling FDR that

is based on the pFDR. When applied to a gene-expression data, the modified q-value method identifies a few more significant genes than Storey's original q-value method at certain thresholds. These results have appeared in Liu and Sarkar (2009).

The layout of this dissertation is as follows. In Chapter 2, we give some basic notations and definitions used in multiple testing and basic formulas for the FDR of stepwise multiple testing methods. We review some FDR controlling procedures that are relevant to our research and their properties in Chapter 3. As estimation of n_0 is one of the important steps toward the development of our improved version of the BKY method, we devote one full chapter (Chapter 4) to it. Chapter 5 presents the derivation of our proposed alternative to the BKY method based on the estimate of n_0 developed in Chapter 4. Also included in this chapter are the results of simulation studies investigating its FDR controlling property and power performance relative to the BKY method, and an application of it to a real microarray data. In Chapter 6, we construct the aforementioned new class of point estimates of the FDR and present some numerical results comparing the performances of these estimates with those of Storey (2002) in terms of bias and MSE. Chapter 7 presents the new class of estimates of pFDR corresponding the FDR estimates proposed in Chapter 6, and a new q-value method based on these pFDR estimates. We also show in this chapter how our proposed q-value method compares with Storey's original q-value method when they are applied to a microarray data. We conclude the dissertation by presenting a direction of future research.

CHAPTER 2

BASIC NOTATIONS AND DEFINITIONS

In this chapter, we formalize the framework for multiple testing methods in this research by providing the basic notations and definitions.

2.1 Types of Error Rates

Consider simultaneous testing of n null hypotheses $H_1, \ldots, H_n$, each against a certain alternative, using their respective observed p-values $p_1, \ldots, p_n$. The main objective of a multiple testing procedure is to decide for each null hypothesis if it is true or false, which is typically done by comparing the corresponding p-value with a critical value that may or may not depend on the other p-values. These critical values are determined subject to a control of an overall measure of false rejections of null hypotheses. This measure when testing a single null hypothesis is usually the Type-I error rate, the probability of falsely rejecting the null hypothesis, but, in the context of multiple

Table 2.1: The different outcomes in testing n null hypothesis

Number of	Number accepted	Number rejected	
True null hypotheses	U	V	n_0
False null hypotheses	T	S	n_1
	A	R	n

testing, the Type-I error rate can be conceptualized in many different ways. Similarly, while comparing different multiple testing procedures, each controlling a Type-I error rate, there is no unique concept of power generalizing the same from single testing to multiple testing that can be used. Different kinds of Type-I error rate and power for multiple testing have been defined in the literature. All these measures can be described based on the entries in Table 2.1. In this table, n is known, n_0 and n_1 are unknown parameters, R and A are observable, U, V (number of Type-I errors), T (number of Type-II errors) and S are unobservable.

In this dissertation, we will use only one kind of power: Average Power (AvePower), which is defined as the expected proportion of correct rejections among all false null hypotheses, i.e.

$$\text{AvPower}=\begin{cases} \frac{E(S)}{n_1}, & \text{if } n_1 > 0, \\ 1, & \text{if } n_1 = 0. \end{cases}$$

We will describe different notions of Type-I error rate in the following, even though we will be mainly concerned in our present research with only one kind of Type-I error rate, the False Discovery Rate (FDR).

- Per-Family Error Rates (PFER): The expected number of false positives,

$$\text{PFER} = E(V).$$

- Per-comparison Error Rate (PCER): The expected proportion of false positives,

$$\text{PCER} = \frac{E(V)}{n}.$$

- Familywise Error Rate(FWER): The probability of at least one false positive, i.e.

$$\text{FWER} = Pr(V \geq 1).$$

- Generalized Familywise Error Rate (k-FWER): The probability of at least k false rejections,

$$k\text{-FWER} = Pr(V \geq k).$$

  Evidently, the 1-FWER is the FWER.

  The rest of the error rates are based on the False Discovery Proportion (FDP), which is the proportion of false positives among all rejections, i.e.,

$$\text{FDP} = \begin{cases} \frac{V}{R}, & \text{if } R > 0, \\ 0, & \text{if } R = 0. \end{cases}$$

- False Discovery Rate (FDR): The expected proportion of false rejections among the rejected hypotheses,

$$\text{FDR} = E\left(\text{FDP}\right) = E\left(\frac{V}{R} \mid R > 0\right) \cdot Pr(R > 0).$$

When all the null hypotheses are true, the FDR is equivalent to FWER.

- Positive False Discovery Rate (pFDR): The expected proportion of false rejections among all rejections given there is at least one rejection,

$$\text{pFDR} = E\left(\text{FDP}|R > 0\right) = E\left(\frac{V}{R}|R > 0\right).$$

- The Exceedance Probability of False Discovery Proportion (γ-ExFDP): The probability of FDP exceeding a fixed value $\gamma \in (0, 1)$,

$$\gamma\text{-ExFDP} = Pr\left(\text{FDP} > \gamma\right).$$

- Generalized False Discovery Rate (k-FDR): The expected proportion of k or more false rejections among all rejections,

$$k\text{-FDR} = E\left(k\text{-FDP}\right),$$

where

$$k\text{-FDP} = \begin{cases} \frac{V}{R}, & \text{if } V \geq k, \\ 0, & \text{if } V < k. \end{cases}$$

Evidently, the 1-FDP is the FDP.

There is a fundamental difference between strong and weak control of a particular Type-I error rate.

- *Strong control* implies that the Type-I error measure is controlled at a pre-specified level no matter what the configurations of true and false null hypotheses are.

- *Weak control* implies that the Type-I error measure is controlled when all the null hypotheses are assumed to be true.

The FWER, PFER and PCER are the three traditional Type-I error rates, among which the FWER has been the most commonly used; see, for example, Hochberg and Tamhane (1987). However, while testing a large number of hypotheses, as in most of the modern applications of multiple testing, the FWER is too conservative, often allowing a very few rejections. The FDR, k-FWER and γ-FDP have been introduced as less conservative notions of Type-I error rates than the FWER. Benjamini and Hochberg (1995) introduced the FDR, Korn et al. (2004) introduced the k-FWER, and Genovese and Wasserman (2001) and Dudoit et al. (2003) independently proposed to control the γ-ExFDP. The concept of pFDR is due to Storey (2002) who proposed to use an analog of p-value in terms of the pFDR for multiple testing. Sarkar (2007) introduced the k-FDR.

The choice of different Type-I error rates is a trade-off between power and Type-I error control. It is straightforward to see that

$$\begin{aligned} \text{PCER} &\leq \min(\text{FDR}, \gamma - \text{ExFDP}) \leq \max(\text{FDR}, \gamma - \text{ExFDP}) \\ &\leq FWER \leq PFER \end{aligned}$$

So, if a procedure controls the FWER in a strong sense, it will also control the FDR and γ-ExFDP. However, if a procedure controls the FDR or γ-ExFD, it controls the FWER only in a weak sense.

2.2 Types of Multiple Testing Procedure

There are basically two types of multiple testing procedures — stepwise and single-step. These can be described in terms of $p_{1:n} \leq \cdots \leq p_{n:n}$, the

ordered p-values.

1. Stepwise Procedure: There are two types of stepwise procedures, step-down and step-up, in each of which the hypotheses are tested sequentially. Let $H_{i:n}$ be the null hypothesis corresponding to $p_{i:n}$, $i = 1, \ldots, n$.

   - Step-Down Procedure (SDP): Starting with the most significant p-value, it continues rejecting null hypotheses as long as their corresponding p-values are small. More formally, given a set of critical values $\alpha_1 \leq \cdots \leq \alpha_n$, the SDP rejects the null hypotheses $H_{1:n}, \cdots, H_{r:n}$, where $r = \max\{1 \leq i \leq n : p_{j:n} \leq \alpha_j \text{ for all } j \leq i\}$, provided the maximum exists; otherwise, accepts all the null hypotheses.

   - Step-Up Procedure (SUP): Starting with the least significant p-value, it continues accepting null hypotheses as long as their corresponding p-values are large. More formally, given a set of critical values $\alpha_1 \leq \cdots \leq \alpha_n$, the SUP rejects the null hypotheses $H_{1:n}, \cdots, H_{r:n}$, where $r = \max\{1 \leq i \leq n : p_{i:n} \leq \alpha_i\}$, if the maximum exists; otherwise, accepts all null hypotheses.

2. Single-Step Procedure: All the hypotheses are tested by comparing the p-values with one critical value. So, given a common critical value, say c, the hypothesis H_i will be rejected if $p_i \leq c$, for $i = 1, \ldots, n$ A stepwise procedure will reduce to a single-step procedure if all the critical values are same.

With the same sets of critical values, a SUP is more powerful than the corresponding SDP, as the SUP rejects at least the same number of null hy-

potheses as the SDP does. However, it does not mean that a SUP is better than the corresponding SDP, unless the Type-I error rate they are designed to control is exactly the same for both of them.

There are other types of multiple procedures that are extensions of stepwise procedures.

- Generalized Step-Up-Step-Down Procedure (SUSDP): Given a set of critical values $\alpha_1 \leq \cdots \leq \alpha_n$, a generalized step-up-step-down test procedure of order r starts with $p_{n-r+1:n}$. If $p_{n-r+1:n} > \alpha_{n-r+1:n}$, the procedure accepts the null hypotheses $H_{n-r+1:n}, \ldots, H_{n:n}$ and continues to test the remaining hypotheses in a step-up manner based on the corresponding p-values and the critical values; otherwise, the procedure rejects the null hypotheses $H_{1:n}, \ldots, H_{n-r+1:n}$ and continues testing the remaining ones in a step-down manner using the corresponding p-values and the critical values.

  When $r = 1$ (or n), a generalized step-up-step-down procedure of order r reduces to an ordinary step-up (or step-down) procedure. The SUSDP was introduced by Tamhane, Liu and Dunnett(1998); also see Sarkar (2002b).

- Multi-Stage Procedure: Different stepwise procedures are applied at different stages with the critical values at each stage being modified based on information available at the previous stage.

2.3 Formulas for the FDR of Stepwise Procedures

Several different formulae for the FDR of stepwise procedures, step-up, step-down or single-step, have been considered in different papers [see, for example, Benjamin and Yekutielli (2001), Sarkar (2002b, 2006)]. However, we will present alternative expressions for the FDR, given recently in Sarkar (2008b), that provide better insight and will be of use in this dissertation.

For any multiple testing method, the FDR is the expectation of the following:

$$\text{FDP} = \frac{V}{\max(1,R)} = \sum_{i\in I_0}\sum_{r=1}^{n}\frac{1}{r}I\left(H_i \text{ is rejected, } R=r\right), \tag{2.1}$$

where I_0 is the set of indices of true null hypotheses. For a step-up procedure, this expectation can be written more explicitly as follows, with P_i denoting the random variable corresponding to the observed p-value p_i.

Formula 2.1. *For a step-up method of testing n null hypotheses $H_1,\ldots,H_n$ using their p-values $p_1,\ldots,p_n$, respectively, and the critical constants $\alpha_1 \leq \cdots \leq \alpha_n$, the FDR is given by*

$$\text{FDR} = \sum_{i\in I_0} E\left[\frac{I\left(P_i \leq \alpha_{R^{(-i)}_{SU,n-1}(\alpha_2,\ldots,\alpha_n)+1}\right)}{R^{(-i)}_{SU,n-1}(\alpha_2,\ldots,\alpha_n)+1}\right],$$

where $R^{(-i)}_{SU,n-1}(\alpha_2,\ldots,\alpha_n)$ is the number of rejections in testing the $n-1$ null hypotheses other than H_i using the step-up method based on their p-values and the critical constants $\alpha_2 \leq \cdots \leq \alpha_n$.

By taking $\alpha_i = c$, for all $i = 1,\ldots,n$, in the above formula, one gets the

following formula for a single-step procedure that rejects H_i if $p_i \leq c$:

$$\text{FDR} = \sum_{i \in I_0} E\left[\frac{I\left(P_i \leq c\right)}{R_{n-1}^{(-i)}(c)+1}\right],$$

where $R_{n-1}^{(-i)}(c)$ is the number of rejections in testing the $n-1$ null hypotheses other than H_i using the single-step procedure based on the p-values other than p_i.

Formula 2.2 *For a stepdown method of testing n null hypotheses $H_1, \ldots, H_n$ using their p-values $p_1, \ldots, p_n$, respectively, and the critical constants $\alpha_1 \leq \cdots \leq \alpha_n$, the FDR satisfies the following inequality:*

$$\text{FDR} \leq \sum_{i \in I_0} E\left[\frac{I\left(P_i \leq \alpha_{R_{SD,n-1}^{(-i)}(\alpha_1,\ldots,\alpha_{n-1})+1}\right)}{R_{SD,n-1}^{(-i)}(\alpha_1,\ldots,\alpha_{n-1})+1}\right],$$

where $R_{SD,n-1}^{(-i)}(\alpha_1, \ldots, \alpha_{n-1})$ is the number of rejections in testing the $n-1$ null hypotheses other than H_i using the stepdown method based on their p-values and the critical constants $\alpha_1 \leq \cdots \leq \alpha_{n-1}$.

2.4 Q-value Method

Q-value, which is introduced by Storey (2002) along with the pFDR, gives a measure of significance for each observed statistic in terms of the pFDR. For an observed p-value p_{obs}, $q(p_{obs})$ is defined as

$$q(p_{obs}) = \inf_{t \geq p_{obs}}\{\text{pFDR}(t)\}, \tag{2.2}$$

the minimum pFDR we can achieve for the rejection region $(0, p_{obs}]$ among all rejection regions of the form $(0, t]$ containing it.

The q-value provides a measure of significance while automatically taking multiple testing into account. To apply the q-value method, Storey (2002) proposed to estimate $q(p_{obs})$ using the class of estimates of pFDR. Please see Section 3.2 for details of Storey's q-value method.

CHAPTER 3

A REVIEW OF SOME FDR CONTROLLING METHODS

In this chapter, we will focus on reviewing the BH method and its adaptive versions utilizing an estimate of the number of true null hypotheses modifying the BH method.

3.1 The BH Method

The BH method is the one that Benjamini and Hochberg (1995) introduced for the first time as an FDR controlling method. It is a step-up procedure with the critical values $\alpha_i = i\alpha/n$, $i = 1, \ldots, n$; that is, it rejects the null hypotheses $H_{1:n}, \ldots, H_{r:n}$ and accepts the rest, where

$$r = \max\left\{1 \leq i \leq n : p_{i:n} \leq \frac{i}{n}\alpha\right\}, \tag{3.1}$$

provided this maximum exists; otherwise, accepts all the null hypotheses.

The above critical values are the same ones as Simes (1986) originally con-

sidered while testing the global null hypotheses $H_0 = \cap_{i=1}^{n} H_i$. He also proposed to use them in a step-up manner for multiple testing of the H_i's upon rejection of the global null hypothesis. However, this multiple testing method controls the FWER at level α weakly, and it does so when the p-values are either independent [Simes (1986)] or positively dependent in a certain sense [Sarkar and Chang (1997); Sarkar (1998, 2008a)]. It does not control the FWER strongly, even when the p-values are independent [Hommel (1988)]. Benjamini and Hochberg (1995) showed that this step-up method can be used to control the FDR in a strong sense, at least when the p-values are independent. In particular, they proved that FDR $\leq n_0\alpha/n$ for this procedure when the p-values are independent, with the p-values corresponding to the null hypotheses having $U(0,1)$ distribution.

Later, it was proved that for the BH method the FDR is actually equal to $n_0\alpha/n$ under the independence of the p-values [Benjamini and Yekutieli (2001), Finner and Roters (2001), Sarkar (2002b, 2008b), Storey, Taylor and Siegmund (2004)], and it is less than or equal to $n_0\alpha/n$ under the following type of positive dependence among the p-values:

$$\mathrm{E}\left\{\psi\left(P_1,\ldots,P_n\right) \mid P_i = u\right\} \text{ is non-decreasing in } u \text{ for each } i \in I_0, \qquad (3.2)$$

for any (coordinatewise) non-decreasing function ψ [Benjamini and Yekutieli (2001), Sarkar (2002b, 2008b)]. This positive dependence condition, referred to as the positive regression dependence on subset (PRDS) condition, is satisfied by a number of multivariate distributions arising in many multiple testing situations, among which the multivariate normal with non-negative correlations is the most common. Other commonly arising multivariate distributions for which the BH method works are multivariate t with the associated multi-

variate normal with non-negative correlations (when $\alpha \leq 1/2$), absolute valued multivariate t with the associated normals being independent and some type of multivariate F [Benjamini and Yekutieli (2001), Sarkar (2002b, 2004)].

Sarkar (2002b) proved that the step-down analog of the BH method, that is, the method that rejects the null hypotheses $H_{1:n}, \ldots, H_{r:n}$ and accepts the rest, where

$$r = \max \left\{ 1 \leq i \leq n : p_{j:n} \leq \frac{j}{n}\alpha \text{ for all } j = 1, \ldots, i \right\},$$

provided this maximum exists, otherwise, accepts all the null hypotheses, also controls the FDR under the independence or the same type of positive dependence of the p-values as above. It has been noted recently in a number of papers [Finner, Dickhaus and Roters (2009) and Sarkar(2008b)] that the above positive dependence condition can be slightly relaxed to

$$\mathrm{E}\left\{\psi\left(P_1, \ldots, P_n\right) \mid P_i \leq u\right\} \text{ is non-decreasing in } u \text{ for each } i \in I_0, \quad (3.3)$$

for any (coordinatewise) non-decreasing function ψ.

3.2 Estimation Based Approach and Adaptive BH Methods

Storey (2002) put forward his estimation based approach to controlling the FDR based on a class of point estimates of the FDR for a single-step test rejecting each H_i if $p_i \leq t$, for some fixed threshold t, that he developed under the following model:

Mixture Model. *Let P_i denote the random p-value corresponding to p_i*

and $H_i = 0$ *or* 1 *according to the associated null hypothesis is true or false. Let* (P_i, H_i), $i = 1, \ldots, n$, *be independently and identically distributed with* $\Pr(P_i \leq u \mid H_i) = (1 - H_i)u + H_i F_1(u)$, $u \in (0, 1)$, *for some continuous cdf* $F_1(u)$, *and* $\Pr(H_i = 0) = \pi_0 = 1 - \Pr(H_i = 1)$.

Having proved that the FDR of the above single-step test under the mixture model is given by

$$\mathrm{FDR}_n(t) = \frac{\pi_0 t}{F(t)} \Pr\{R_n(t) > 0\}, \tag{3.4}$$

where

$$R_n(t) = \sum_{i=1}^{n} I(p_i \leq t), \; F(t) = \Pr(P_i \leq t) = \pi_0 t + (1 - \pi_0) F_1(t), \tag{3.5}$$

Storey (2002) considered the following class of point estimates of the $\mathrm{FDR}_n(t)$:

$$\widehat{\mathrm{FDR}}_{\lambda,n}(t) = \frac{\hat{\pi}_0(\lambda) t}{\hat{F}_n(t)}, \quad \lambda \in [0, 1) \tag{3.6}$$

where

$$\hat{F}_n(t) = \frac{1}{n} \max\{R_n(t), 1\} \quad \text{and} \quad \hat{\pi}_0(\lambda) = \frac{n - R_n(\lambda)}{n(1 - \lambda)}. \tag{3.7}$$

The estimate of π_0 was originally suggested by Schweder and Spjotvoll (1982) in a different context. Storey (2002) proved that $E(\widehat{\mathrm{FDR}}_{n,\lambda}(t)) \geq \mathrm{FDR}_n(t)$, that is, $\widehat{\mathrm{FDR}}_{n,\lambda}(t)$ is conservatively biased as an estimate of $\mathrm{FDR}_n(t)$, which he argued is desirable, because by controlling it one can control the true $\mathrm{FDR}_n(t)$. Storey (2002) suggested using

$$t_\alpha = \sup\left\{0 \leq t \leq 1 : \widehat{\mathrm{FDR}}_{n,\lambda}(t) \leq \alpha\right\} \tag{3.8}$$

to threshold the p-values, that is, to use it as the cut-off point below which a p-value should be declared significant at a level α.

Storey (2002) modified the FDR by considering it conditional on at least one rejection, which he called the positive false discovery rate (pFDR)(See, Section 2.1 for definition) and noted that for the single-step test with the threshold t under mixed model it is actually equal to

$$\frac{\pi_0 t}{F(t)} = Pr(H_1 = 0 | P_1 \leq t),$$

a 'Bayesian p-value'. This motivated him to introduce a pFDR analog of the p-value, called the q-value, as a test-specific measure of significance in multiple testing. For a test with the observed p-value p_{obs}, it is defined as

$$q(p_{obs}) = \inf_{t \geq p_{obs}} \{\text{pFDR}(t)\}, \tag{3.9}$$

the minimum pFDR that can be achieved for a rejection region $(0, p_{obs}]$ among all rejection regions of the form $(0, t]$ containing it. To apply his so called q-value method, Storey (2002) proposed to estimate $q(p_{obs})$ using the following class of estimates of the $\text{pFDR}_n(t)$ that is derived from that of the $\text{FDR}_n(t)$ in 3.6:

$$\widehat{\text{pFDR}}_{n,\lambda}(t) = \frac{\widehat{\text{FDR}}_{n,\lambda}(t)}{1 - (1-t)^n}, \quad \lambda \in [0, 1). \tag{3.10}$$

Thus, the estimated q-value to be used in the q-value method, as Storey(2002) has suggested, is

$$\hat{q}(p_{obs}) = \inf_{t \geq p_{obs}} \{\widehat{\text{pFDR}}(t)\}, , \quad \lambda \in [0, 1). \tag{3.11}$$

Storey (2002) has made a connection between his and the fixed rejection region approaches through his class of estimates. He pointed out that if one thresholds the p-values at $p_{\hat{l}(\lambda):n}$, that is, rejects the null hypotheses $H_{1:n}, \ldots, H_{\hat{l}(\lambda):n}$, where

$$\hat{l}(\lambda) = \max\{1 \leq i \leq n : \widehat{\text{FDR}}_{n,\lambda}(p_{i:n}) \leq \alpha\}, \tag{3.12}$$

then one gets the BH method when $\lambda = 0$. This is obvious, because $\widehat{\text{FDR}}_{n,\lambda=0}(p_{i:n}) = np_{i:n}/i$, so (3.12) when $\lambda = 0$ is the same as

$$\hat{l}(0) = \max\{1 \leq i \leq n : p_{i:n} \leq i\alpha/n\}, \tag{3.13}$$

that provides the BH method [see, (3.1)]. Notice that thresholding the p-values at $p_{\hat{l}^*(\lambda):n}$, instead of at $p_{\hat{l}(\lambda):n}$, where

$$\hat{l}^*(\lambda) = \max\left\{1 \leq j \leq n : \widehat{\text{FDR}}_{n,\lambda}(p_{i:n}) \leq \alpha \text{ for all } i = 1, \ldots, j\right\}, \tag{3.14}$$

is equivalent to the step-down version of the BH method considered in Sarkar (2002b) when $\lambda = 0$.

Storey's idea of thresholding the p-values through estimated FDRs, as in (3.8) or (3.10), yields a class of step-up or step-down methods, each modifying the BH method or its step-down analog by incorporating into it an estimate of $n_0 = n\pi_0$ based on the same set of available p-values. Such modification of the BH method was originally proposed by Benjamini and Hochberg (2000), and is referred to as an adaptive BH method in the literature. We will review some specific adaptive BH methods in the following.

3.2.1 The Adaptive BH Method of Benjamini & Hochberg

Benjamini and Hochberg (2000) introduced this adaptive BH method for independent p-values based on an estimate of n_0 developed using the so called the lowest slope (LSL) method.

When all the null hypotheses are true and the test statistics are independent, the p-values should be iid as $U(0, 1)$ with the expectations of the ordered p-values as $E(P_{i:n}) = i/(n+1)$, $i = 1, \ldots, n$. Therefore, the plot of $p_{i:n}$ versus i

should exhibit a linear relationship, along the line with the slope $S = 1/(n+1)$ and passing through the origin and the point $(n+1, 1)$ (assuming $p_{n+1:n} = 1$).

When $n_0 \leq n$, the p-values corresponding to the false null hypotheses tend to be small, so they concentrate on the left side of the above plot. The relationship over the right side of the plot remains approximately linear with the slope $\beta = 1/(n_0+1)$. Therefore, using a suitable set of the largest p-values, a straight line through the point $(n+1, 1)$ can be fitted with slope $\hat{\beta}$ and n_0 can be estimated as $\hat{n}_0 = 1/\hat{\beta}$. Benjamini and Hochberg (2000) suggested estimating n_0 using the LSL method and their adaptive procedure as follows:

1. Apply the original BH method If none is rejected, then accept all hypotheses and stop; otherwise continue.

2. Calculate the slopes $S_i = (1 - p_{i:n}/(n+1-i))$.

3. Starting with $i = 1$, proceed as long as $S_i \geq S_{i-1}$ and stop when the first time $S_j < S_{j-1}$. Let $\hat{n}_0 = \min[n, 1/S_j + 1]$.

4. Apply the BH method with $\alpha_i = i\alpha/\hat{n}_0$.

Though there is no theoretical proof that their procedure guarantees an FDR control, simulation studies indicate that it does.

3.2.2 The Adaptive BH Method of Storey, Taylor and Siegmund

Storey, Taylor and Siegmund (2004) modified Storey's (2002) original estimate $\widehat{\mathrm{FDR}}_{n,\lambda}(t)$ of $\mathrm{FDR}_n(t)$, when $0 < \lambda < 1$, to

$$\widehat{\mathrm{FDR}}^{STS}_{n,\lambda}(t) = \begin{cases} \frac{n\hat{\pi}_0^{STS}(\lambda)t}{\max\{R_n(t),1\}} & \text{if } t \le \lambda, \\ 1 & \text{if } t > \lambda, \end{cases} \tag{3.15}$$

with

$$\hat{\pi}_0^{STS}(\lambda) = \frac{n - R_n(\lambda) + 1}{n(1-\lambda)}, \tag{3.16}$$

and suggested thresholding the p-values based on this new estimate as follows:

$$t_\alpha(\widehat{FDR}^{STS}_{n,\lambda}) = \sup\{0 \le t \le 1 : \widehat{FDR}^{STS}_{n,\lambda}(t) \le \alpha\}. \tag{3.17}$$

The adaptive BH method corresponding to this new estimate, to be called the STS method, rejects $H_{1:n}, \ldots, H_{r:n}$ where

$$r = \max\left\{0 \le i \le n : p_{i:n} \le \min\left\{\frac{i\alpha}{\hat{n}_0^{STS}}, \lambda\right\}\right\}, \tag{3.18}$$

$$\hat{n}_0^{STS}(\lambda) = \frac{n - R_n(\lambda) + 1}{1-\lambda}. \tag{3.19}$$

The STS controls the FDR under the independence of the p-values [Benjamini, Krieger and Yekutieli (2006), Storey, Taylor and Siegmund (2004), Sarkar(2004, 2008b)], as well as under certain form of weak dependence asymptotically as $n \to \infty$ [Storey, Taylor and Siegmund (2004)].

3.2.3 The Adaptive BH Method of Benjamini, Krieger and Yekutieli (2006)

Unlike Storey (2002) or Storey, Taylor and Siegmund (2004) where n_0 is estimated based on the number of significant p-values observed in a single-step test with an arbitrary critical value λ, Benjamini, Krieger and Yekutieli (2006) considered estimating n_0 using the BH method at level $\alpha/(1+\alpha)$. Their adaptive version of the BH method, to be called the BKY method, runs as follows:

1. Apply the BH method at level $q = \frac{\alpha}{1+\alpha}$. Let r_1 be the number of rejections. If $r_1 = 0$, accept all the null hypotheses and stop; if $r_1 = n$, reject all the null hypotheses and stop; otherwise continue to the next step.

2. Estimate n_0 as

$$\hat{n}_0^{BKY} = \frac{n - r_1}{1 - q} = (n - r_1)(1 + \alpha). \tag{3.20}$$

3. Apply the BH method with the critical values $\alpha_i = i\alpha/\hat{n}_0^{BKY}$, $i = 1, \ldots, n$.

As Benjamini, Krieger and Yekutieli (2006) have proved, the BKY method controls the FDR at α under independence of the p-values. While it is less powerful than the adaptive procedure proposed in Storey et al.(2004) when the p-values are independent, simulation studies have shown that, with the p-values generated from multivariate normals with common positive correlations, it can also control the FDR. Benjamini, Krieger and Yekutieli (2006) also extended the BKY method to a multiple-stage procedure (MST) by repeating the two-stage procedure as long as more hypotheses are rejected, which is stated as follows:

1. Let $r = \max\{i : \text{ for all } j \leq i, \text{ there exists } l \geq j \text{ so that } p_{l:n} \leq \alpha l/[n + 1 - j(1 - \alpha)]\}$.

2. If such an r exists, reject $p_{1:n}, \ldots, p_{r:n}$; otherwise reject no hypotheses.

This multiple-stage procedure is a combination of step-up and step-down methods. They offered no analytical proof of its FDR control. Benjamini, Krieger and Yekutieli (2006) also mentioned that a multiple-stage step-down procedure

(MSD) can be developed by choosing $l = j$ in MST. They provided numerical results showing that the MST method can also control the FDR, the theoretical justification of which is given later in Gavrilov, Benjamini and Sarkar (2009) to be reviewed in the following section.

3.2.4 The Adaptive Method of Gavrilov, Benjamini and Sarkar (2009)

As mentioned above, Gavrilov, Benjamini and Sarkar (2009) reexamined the multiple-stage step-down procedure, the MSD method, mentioned in Benjamini, Krieger and Yekutieli (2006) and proved that this multiple-stage step-down procedure can control the FDR under the independence of the p-values. The following is the MSD method:

Find $k = \max\{1 \leq i \leq n : p_{j:n} \leq j\alpha/(n+1-j(1-\alpha)) \text{ for all } j = 1, \ldots, i\}$ and reject $H_{1:n}, \ldots, H_{k:n}$ if k exists; otherwise reject no hypotheses.

Although it has been referred to as a multiple-stage stepdown method by Benjamini, Krieger and Yekutieli (2006), it is actually, as Sarkar (2008b) argued, an adaptive version of the stepdown analog of the BH method considered in Sarkar (2002b). To see this, first note that, under the same setup involving the mixture model and a constant rejection threshold t for each p-value as in Storey (2002) or Storey, Taylor and Siegmund (2004), one can consider estimating n_0 based on the number of significant p-values compared to the t, rather than a different arbitrary constant λ. In other words, by considering the Storey, Taylor and Siegmund (2004) type estimate of $n_0 = n\pi_0$ with $\lambda = t$ and using this estimate in $\widehat{\mathrm{FDR}}_{n,\lambda}(t)$, Storey's original estimate of the $\mathrm{FDR_n(t)}$,

one can develop the following alternative estimate of $\mathrm{FDR_n(t)}$:

$$\widehat{\mathrm{FDR}}_n^*(t) = \frac{[n - R_n(t) + 1]t}{(1-t)\max\{R_n(t), 1\}}.$$

A step-down method developed through this estimate, that is, the one that rejects $H_{1:n}, \ldots, H_{r:n}$ where

$$\begin{aligned} r &= \max\left\{1 \le i \le n : \widehat{\mathrm{FDR}}_n^*(p_{j:n}) \le \alpha \text{ for all } j = 1, \ldots, i\right\} \\ &= \max\left\{1 \le i \le n : \frac{p_{j:n}}{1 - p_{j:n}} \le \frac{j\alpha}{n - j + 1} \text{ for all } j = 1, \ldots, i\right\}, \end{aligned} \tag{3.21}$$

which is the same as the MSD, is an adaptive version of the step-down analog of the BH method.

Gavrilov, Benjamini and Sarkar (2009) also conducted simulation studies comparing the above three FDR controlling adaptive procedures, the BKY, MSD and STS. The STS is the most powerful one when the test statistics are independent, with the MSD taking the second place, although sometimes the power is very close to that of the STS. Under the setting of equicorrelated normal test statistics, the BKY method is the only one that seems to control the FDR [Romano, Sheikh and Wolf (2004)]. The MSD in this case also appears to perform well and its control over the FDR does not break down by much from the desired level.

There are other methods of estimating n_0 proposed in the literature, such as parametric beta-uniform mixture model by Pounds and Morris (2003), the Spacing LOESS Histogram (SPLOSH) method by Pounds and Cheng (2004), the nonparametric MLE method by Langaas and Lindqvist (2005), the moment generating function approach by Broberg (2005), and the resampling

strategy by Lu and Perkins (2007). These n_0 estimates could also be used while developing adaptive versions of the BH method or its step-down analog. But, does any of these control the FDR theoretically, at least when the p-values are independent? The answer is not known yet.

CHAPTER 4

NEW ESTIMATE OF NUMBER OF TRUE NULL HYPOTHESES

We present in this chapter a new estimate of n_0 and the results of a simulation study comparing this estimate to $\hat{n}_0^{STS}$ and $\hat{n}_0^{BKY}$, before we use it to propose our version of adaptive BH method in the next chapter.

4.1 New n_0 Estimate

Our estimate of n_0 is developed somewhat along the line of that in the BKY method. However, instead of deriving it from the number of significant p-values in the original BH method at level $q = \alpha/(1+\alpha)$, which is being done in the BKY method, we consider deriving it from the number of significant p-values in the step-down analog of the BH method at the same level q but using a formula that is similar to that in Storey, Taylor and Siegmund (2004).

More specifically, our proposed estimate of n_0 is given by:

$$\hat{n}_0^{NEW} = \frac{n - R_{SD}(\gamma_1, \ldots, \gamma_n) + 1}{1 - \gamma_{R_{SD}(\gamma_1,\ldots,\gamma_n)+1}}, \tag{4.1}$$

where $R_{SD}(\gamma_1, \ldots, \gamma_n)$ the number of rejections in the step-down version of the BH method with the critical values $\gamma_i = i\gamma/n$, for $i = 1, \ldots, n$, where $\gamma = \alpha/(1+\alpha)$ and $\gamma_{n+1} \in [\gamma, (1+\gamma)/2)$. The choice of γ_{n+1} in this particular interval, as discussed in Chapter 5, is dictated that for such γ_{n+1} the FDR of the corresponding adaptive BH method can be controlled at α, at least when the p-values are independent.

The results presented in the following section favoring $\hat{n}_0^{NEW}$ as an estimate of n_0 over $\hat{n}_0^{BKY}$ provide some rationale for our choice of this new estimate.

4.2 Simulation Study

We ran a simulation study to investigate numerically how $\hat{n}_0^{NEW}$ performs compared to $\hat{n}_0^{STS}$ (with $\lambda = 0.5$) and $\hat{n}_0^{BKY}$ as an estimate of n_0. We generated n dependent random variables $X_i \sim N(\mu_i, 1)$, $i = 1, \ldots, n$, with a common non-negative correlation ρ, and determined their p-values for testing $\mu_i = 0$ against $\mu_i > 0$. We repeated this 10,000 times by setting n at 5000, ρ at $0, 0.25, 0.5, 0.75$, the proportion of the true null hypotheses π_0 at $0, 0.25, 0.5, 0.75$ and 1, the value of μ_i for each false null hypothesis at 1, and the value of α at 0.05. Each time, we calculated the values of the three estimates. From these 10,000 values, we constructed the boxplots and calculated the estimated mean and variance for each estimate. We present these boxplots in Figure 4.1 and the estimated means and variances in Table 1 only for

Table 4.1: The estimated mean and variance of $\hat{n}_0^{NEW}$, $\hat{n}_0^{STS}$ and $\hat{n}_0^{BKY}$ for the cases of $n = 5000$, $n_0 = 2500$ and $\rho = 0, 0.25, 0.5$ and 0.75.

	mean			variance		
	NEW	BKY	STS	NEW	BKY	STS
$\rho = 0$	4996	5242	3296	33.87	55.82	3782
$\rho = 0.25$	4927	5166	3284	69133	81661	2539704
$\rho = 0.5$	4862	5088	3280	279038	325556	5257668
$\rho = 0.75$	4881	5008	3276	448641	712831	8263566

$\pi_0 = 0.5$, as they provide very similar comparative pictures for other values of π_0.

As seen from Figure 4.1 and Table 4.1, $\hat{n}_0^{NEW}$ is a better estimate of n_0 than $\hat{n}_0^{BKY}$. Looking at $\hat{n}_0^{STS}$ and comparing it to the other two, one notices that although it is more centrally located at the true n_0, it is more variable, and the variability increases quite dramatically with increasing ρ. The variabilities of both $\hat{n}_0^{NEW}$ and $\hat{n}_0^{BKY}$, on the other hand, remain relatively more stable with increasing ρ.

The above findings seem to suggest that the adaptive BH method based on our estimate $\hat{n}_0^{NEW}$ may perform well compared to that based on $\hat{n}_0^{BKY}$ in some situations. Moreover, both these adaptive BH methods seem to behave similarly in terms of the FDR control and power compared to that based on $\hat{n}_0^{STS}$. For instance, like the BKY method, the adaptive BH method based on $\hat{n}_0^{NEW}$ which controls the FDR under independence, which we will prove in the next chapter, can also control the FDR under positive dependence, which we will verify also in the next chapter.

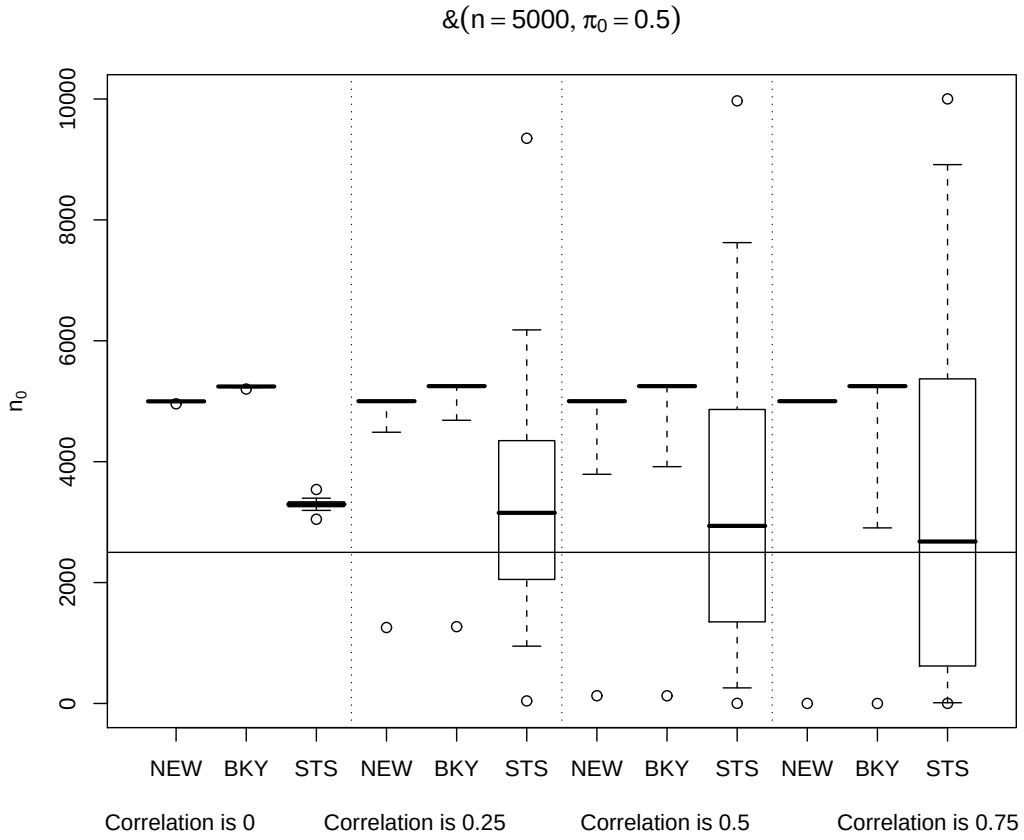

Figure 4.1: The simulated distribution of $\hat{n}_0^{NEW}$, $\hat{n}_0^{STS}$ and $\hat{n}_0^{BKY}$ for the cases of $n = 5000$, $n_0 = 2500$ and $\rho = 0, 0.25, 0.5$ and 0.75. Each box displays the median and quartiles as usual. The whiskers extend to the 5% and 95% quartiles. The circles are the extreme values, i.e. the 0.01% and 99.99% percentiles.

CHAPTER 5

NEW ADAPTIVE METHOD CONTROLLING THE FDR

In this chapter, we present our adaptive version of the BH method based on the estimate $\hat{n}_0^{NEW}$ of n_0. We will prove that the FDR of this adaptive BH method is controlled under independence of the p-values and numerically show that this control continues to hold even when the p-values are positively dependent under normal distributional setting with equal positive correlation. The performance of this adaptive procedure is examined by comparing it to the BKY method.

5.1 The New Adaptive FDR Controlling Method

The following is our proposed adaptive BH method:

PROCEDURE 5.1.

1. *Observe $R_{SD}(\gamma_1, \ldots, \gamma_n)$, the number of rejections in a step-down method*

with the critical values $\gamma_i = i\gamma/n$, $i = 1, \ldots, n$, *with* $\gamma = \alpha/(1+\alpha)$, *and calculate*

$$\hat{n}_0^{NEW} = \frac{n - R_{SD}(\gamma_1, \ldots, \gamma_n) + 1}{1 - \gamma_{R_{SD}(\gamma_1,\ldots,\gamma_n)+1}}, \tag{5.1}$$

with an arbitrary $\gamma_{n+1} \in [\gamma, (1+\gamma)/2)$.

2. *Apply the step up procedure with the critical values* $\alpha_i = i\alpha/\hat{n}_0^{NEW}$, $i = 1, \ldots, n$, *for testing the null hypotheses.*

Theorem 5.1 *Procedure 5.1 controls the FDR at* α *when the p-values are independent.*

The following two lemmas will facilitate our proof of this theorem. These lemmas will be proved later in this section.

Lemma 5.1. *Let* $U \sim U(0,1)$. *Then, for any non-increasing function* $\phi(U) > 0$ *and a constant* $c > 0$, *we have*

$$E\left\{\frac{I\left(U \le c\phi(U)\right)}{\phi(U)}\right\} \le c. \tag{5.2}$$

Lemma 5.2. *Let* $R_{n-1,SD}^{(-i)}(c_1, \ldots, c_{n-1})$ *be the number of rejections in a step-down method based on the* $n-1$ *p-values other than* p_i, *where* $i \in I_0$, *and a set of critical values* $0 < c_1 \le \cdots \le c_{n-1} < 1$. *Then, under independence of the p-values, we have*

$$\sum_{i \in I_0} E\left\{\frac{1 - c_{R_{SD,n-1}^{(-i)}(c_1,\ldots,c_{n-1})+1}}{n - R_{SD,n-1}^{(-i)}(c_1, \ldots, c_{n-1})}\right\} \le 1 - Pr\left\{P_{1:n} \le c_1, \ldots, P_{n:n} \le c_n\right\}, \tag{5.3}$$

for an arbitrary, fixed $c_n \in [c_{n-1}, 1)$.

Proof of Theorem 5.1. Using Formula 2.1, we first note that

$$\begin{aligned} \text{FDR} &= \sum_{i\in I_0} E\left[\frac{I\left(P_i \le \alpha_{R^{(-i)}_{SU,n-1}(\alpha_2,\ldots,\alpha_n)+1}\right)}{R^{(-i)}_{SU,n-1}(\alpha_2,\ldots,\alpha_n)+1}\right] \\ &= \sum_{i\in I_0} E\left[\frac{I\left(P_i \le \left\{R^{(-i)}_{SU,n-1}(\alpha_2,\ldots,\alpha_n)+1\right\}\alpha_1\right)}{R^{(-i)}_{SU,n-1}(\alpha_2,\ldots,\alpha_n)+1}\right], \end{aligned} \tag{5.4}$$

with

$$\begin{aligned} \alpha_i &= \frac{i\alpha\left(1-\gamma_{R_{SD}(\gamma_1,\ldots,\gamma_n)+1}\right)}{n-R_{SD}(\gamma_1,\ldots,\gamma_n)+1} \\ &= \begin{cases} \frac{i\alpha[n-\gamma(R_{SD}(\gamma_1,\ldots,\gamma_n)+1)]}{n[n-R_{SD}(\gamma_1,\ldots,\gamma_n)+1]} & \text{when } R_{SD}(\gamma_1,\ldots,\gamma_n)=0,\ldots,n-1, \\ i\alpha\left(1-\gamma_{n+1}\right) & \text{when } R_{SD}(\gamma_1,\ldots,\gamma_n)=n, \end{cases} \end{aligned} \tag{5.5}$$

$i = 1,\ldots,n$.

Now, notice that $R_{SD}(\gamma_1,\ldots,\gamma_n)$, with fixed $(\gamma_1,\ldots,\gamma_n)$, is a decreasing function of each of the p-values, and as a function of $R_{SD}(\gamma_1,\ldots,\gamma_n)$, α_i is an increasing function if $\gamma \le n/(n+2)$ and $\gamma_{n+1} \le (1+\gamma)/2$. But, $\gamma \le n/(n+2)$ means that $\alpha \le n/2$, which is obviously true, since $n \ge 2$. Thus, as long as $\gamma_{n+1} \le (1+\gamma)/2$, each α_i is a (componentwise) decreasing function of $\mathbf{P} = (P_1,\ldots,P_n)$. So, by letting $P_i \to 0$ in α_1 we see that

$$\alpha_1 \le \frac{\alpha\left(1-\gamma_{R^{(-i)}_{SD,n-1}(\gamma_2,\ldots,\gamma_n)+2}\right)}{n-R^{(-i)}_{SD,n-1}(\gamma_2,\ldots,\gamma_n)}, \tag{5.6}$$

since $R_{SD}(\gamma_1,\ldots,\gamma_n) \to R^{(-i)}_{SD,n-1}(\gamma_2,\ldots,\gamma_n)+1$ as $P_i \to 0$. Let us define $g(\mathbf{P}) = R^{(-i)}_{SU,n-1}(\alpha_2,\ldots,\alpha_n)+1$ and $h(\mathbf{P}^{(-i)})$ equal the right-hand side of (21),

with $\mathbf{P}^{(-i)} = (P_1, \ldots, P_n) \setminus \{P_i\}$. Then, we have

$$\begin{aligned}
\text{FDR} &\leq \sum_{i \in I_0} E\left[\frac{I\left(P_i \leq g(\mathbf{P})\, h\left(\mathbf{P}^{(-i)}\right)\right)}{g(\mathbf{P})}\right] \\
&= \sum_{i \in I_0} E\left[E\left\{\frac{I\left(P_i \leq g(\mathbf{P})\, h\left(\mathbf{P}^{(-i)}\right)\right)}{g(\mathbf{P})} \mid \mathbf{P}^{(-i)}\right\}\right] \\
&\leq \sum_{i \in I_0} E\left\{h\left(\mathbf{P}^{(-i)}\right)\right\} \\
&\leq \alpha\left[1 - Pr\left\{P_{1:n} \leq \gamma_2, \ldots, P_{n:n} \leq \gamma_{n+1}\right\}\right] \\
&\leq \alpha, \qquad (5.7)
\end{aligned}$$

with the second and third inequalities following from Lemmas 5.1 and 5.2 respectively. Thus, the theorem is proved.

We will now give proofs of Lemmas 5.1 and 5.2.

Proof of Lemma 5.1. Consider the function $\psi(u) = u - c\phi(u)$. Since this is non-decreasing, there exists a constant c^* such that $\{\psi(u) \leq 0\} \subseteq \{u \leq c^*\}$ and $\psi(c^*) \leq 0$, that is, $c^* \leq c\phi(c^*)$. Since $\phi(u) \geq \phi(c^*)$ when $u \leq c^*$, we have

$$E\left\{\frac{I\left(U \leq c\phi(U)\right)}{\phi(U)}\right\} \leq E\left\{\frac{I\left(U \leq c^*\right)}{\phi(c^*)}\right\} = \frac{c^*}{\phi(c^*)} \leq \frac{c\phi(c^*)}{\phi(c^*)} = c.$$

Thus, the lemma is proved.

Proof of Lemma 5.2.

$$
\begin{aligned}
& \sum_{i\in I_0} E\left\{\frac{1-c_{R^{(-i)}_{SD,n-1}(c_1,\ldots,c_{n-1})+1}}{n-R^{(-i)}_{SD,n-1}(c_1,\ldots,c_{n-1})}\right\} \\
= & \sum_{i\in I_0}\sum_{r=0}^{n-1}\frac{1-c_{r+1}}{n-r}Pr\left\{R^{(-i)}_{SD,n-1}(c_1,\ldots,c_{n-1})=r\right\} \\
= & \sum_{i\in I_0}\sum_{r=0}^{n-1}\frac{1}{n-r}Pr\left\{P^{(-i)}_{1:n-1}\le c_1,\ldots,P^{(-i)}_{r:n-1}\le c_r,P^{(-i)}_{r+1:n-1}>c_{r+1},P_i>c_{r+1}\right\} \\
\le & \sum_{i=1}^{n}\sum_{r=0}^{n-1}\frac{1}{n-r}Pr\left\{P^{(-i)}_{1:n-1}\le c_1,\ldots,P^{(-i)}_{r:n-1}\le c_r,P^{(-i)}_{r+1:n-1}>c_{r+1},P_i>c_{r+1}\right\} \\
= & \sum_{r=0}^{n-1}Pr\left\{P_{1:n}\le c_1,\ldots,P_{r:n}\le c_r,P_{r+1:n}>c_{r+1}\right\} \\
= & 1-Pr\left\{P_{1:n}\le c_1,\ldots,P_{n:n}\le c_n\right\}, \qquad (5.8)
\end{aligned}
$$

where $P^{(-i)}_{1:n-1} \le \ldots \le P^{(-i)}_{n-1:n-1}$ are the ordered components of $\mathbf{P}^{(-i)}$. The third equality in (23) follows from results on ordered random variables given in Sarkar (2002b). Thus, the lemma is proved.

Note that any value of γ, as long as $0 \le \gamma \le \frac{n}{n+2}$, can ensure that the new adaptive procedure controls FDR under independence. In this thesis, we only consider the situation with $\gamma = \alpha/(1+\alpha)$, which is equal to the choice of q in the BKY method.

5.2 Simulation Study

A simulation study was performed to compare the FDR control and power of our proposed method with those of the BKY method. The study consisted of two parts, the first part was designed for small number of hypotheses, while the second part was designed for relatively large number of hypotheses as seen

in most applications of the FDR.

In the first part of the study, we generated n dependent random variables $X_i \sim N(\mu_i, 1)$, $i = 1, \ldots, n$, with a common non-negative correlation ρ, and applied both the BKY and our proposed methods to test $\mu_i = 0$ against $\mu_i > 0$, simultaneously for $i = 1, \ldots, n$ at a level α. We repeated this 10,000 times by setting n at $4, 8, 16, 64, 128, 256$ and 512, the value of ρ at $0, 0.1, 0.25$ and 0.5, the proportion of the true null hypotheses π_0 at $0, 0.25, 0.5, 0.75$ and 1, α at 0.05, and μ_i at 1 for each false null hypothesis, to simulate the FDR and average power (the expected proportion of alternative μ_i's that are correctly identified) for both methods.

Figure 5.1 compares the FDR control and Table 5.1 lists the ratios of power of both methods to the 'Oracle' method when $n = 32, 128$ and 512. The 'Oracle' method is the BH method based on the critical values $\alpha_i = i\alpha/n_0$, which controls the FDR at the exact level α under the independence of the test statistics. Obviously, it is not implementable in practice as n_0 is unknown, but it serves as a benchmark against which other methods can be compared. As seen in Figure 5.1, our proposed method, which is known to control the FDR at the desired level $\alpha = 0.05$ under independence, can continue to maintain a control over the FDR even under positive dependence, like the BKY method, although ours is often less conservative. Also in terms of power, as seen from Table 5.1, our method appears to be more powerful than the BKY method in most of the cases considered, especially when the correlation is not very high.

The second part of the study was conducted by setting $n = 5000$. The simulated FDR and power were also based on 10,000 iterations. The comparison between simulated FDR of the two methods is presented in Figure 5.2. Again,

Table 5.1: Estimated power for $n = 32, 128, 512$ and $\rho = 0, 0.1, 0.25, 0.5$

		$\pi_0 = 0.25$			$\pi_0 = 0.5$			$\pi_0 = 0.75$		
n		32	128	512	32	128	512	32	128	512
$\rho = 0$	new/oracle	0.1881	0.1160	0.0737	0.4661	0.3935	0.3111	0.7553	0.7188	0.7088
	BKY/oracle	0.1867	0.1105	0.0682	0.4618	0.3746	0.2925	0.7460	0.6914	0.6692
$\rho = 0.1$	new/oracle	0.2281	0.1769	0.1479	0.4956	0.4142	0.3550	0.7868	0.7060	0.6408
	BKY/oracle	0.2293	0.1714	0.1403	0.4937	0.3979	0.3354	0.7796	0.6794	0.6024
$\rho = 0.25$	new/oracle	0.2846	0.2530	0.2373	0.5291	0.4949	0.4582	0.7542	0.7502	0.6979
	BKY/oracle	0.2908	0.2495	0.2305	0.5333	0.4829	0.4423	0.7467	0.7294	0.6650
$\rho = 0.5$	new/oracle	0.3618	0.3391	0.3288	0.5942	0.5729	0.5551	0.7808	0.7895	0.7618
	BKY/oracle	0.3755	0.3438	0.3305	0.6096	0.5725	0.5478	0.7876	0.7781	0.7432

there is evidence that our method can continue to control the FDR under positive dependence, at least when the p-values are equally correlated. The power comparisons in this case are displayed in Figures 5.3 and 5.4. Figure 5.3 indicates that the proposed method is more powerful than the BKY when the correlation between the test statistics is moderately low. Figure 5.4 compares the power of the two methods under the condition of high proportion of true null, $\pi_0 \geq 0.9$, which is often the case in modern multiple testing situations. The proposed method seems to be more powerful then the BKY method in such situations.

In conclusion, the simulation study seems to indicate that the new proposed method can control the FDR under positive dependence of the p-values. It is more powerful than the BKY method under positive but not very high correlations between the test statistics. When there is a large proportion of true null hypotheses, the new method appears to perform better than the BKY method even in the case of high correlations.

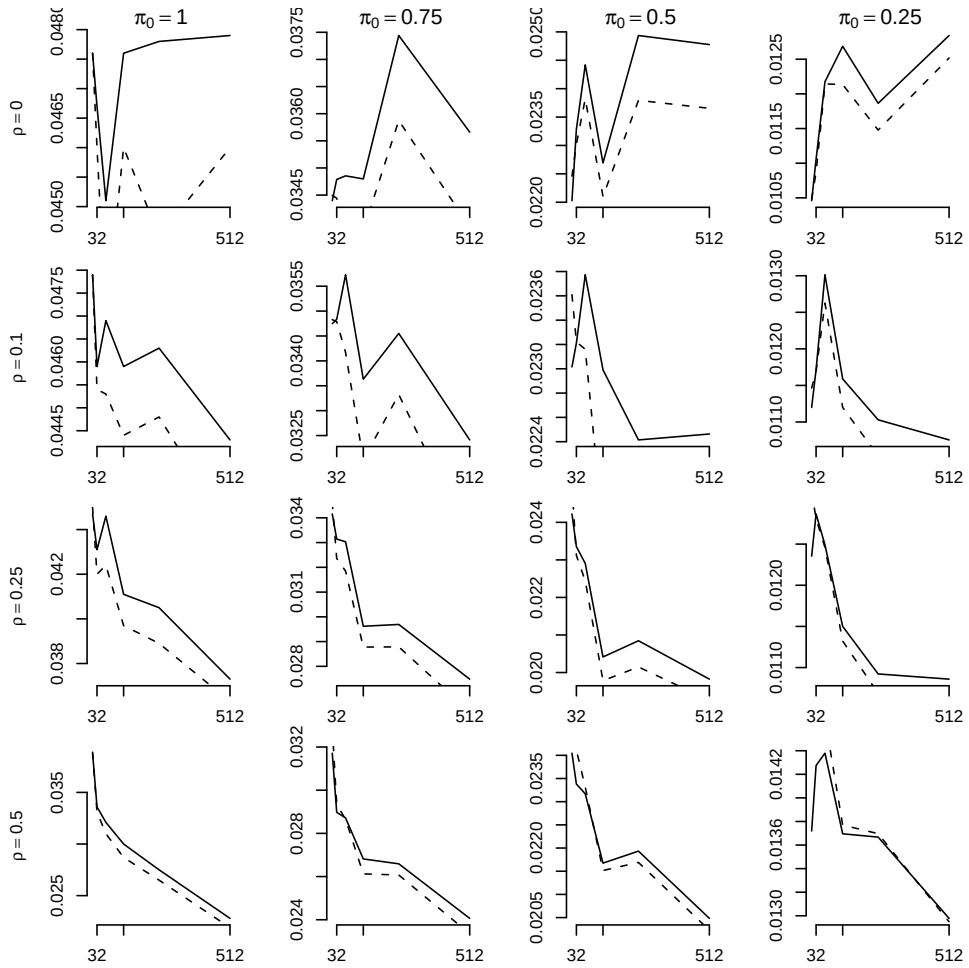

Figure 5.1: Estimated FDR values for $n = 16, 32, \ldots, 512$ and $\rho = 0, 0.1, 0.25, 0.5$. Legend: NEW — solid line; BKY — dashed line.

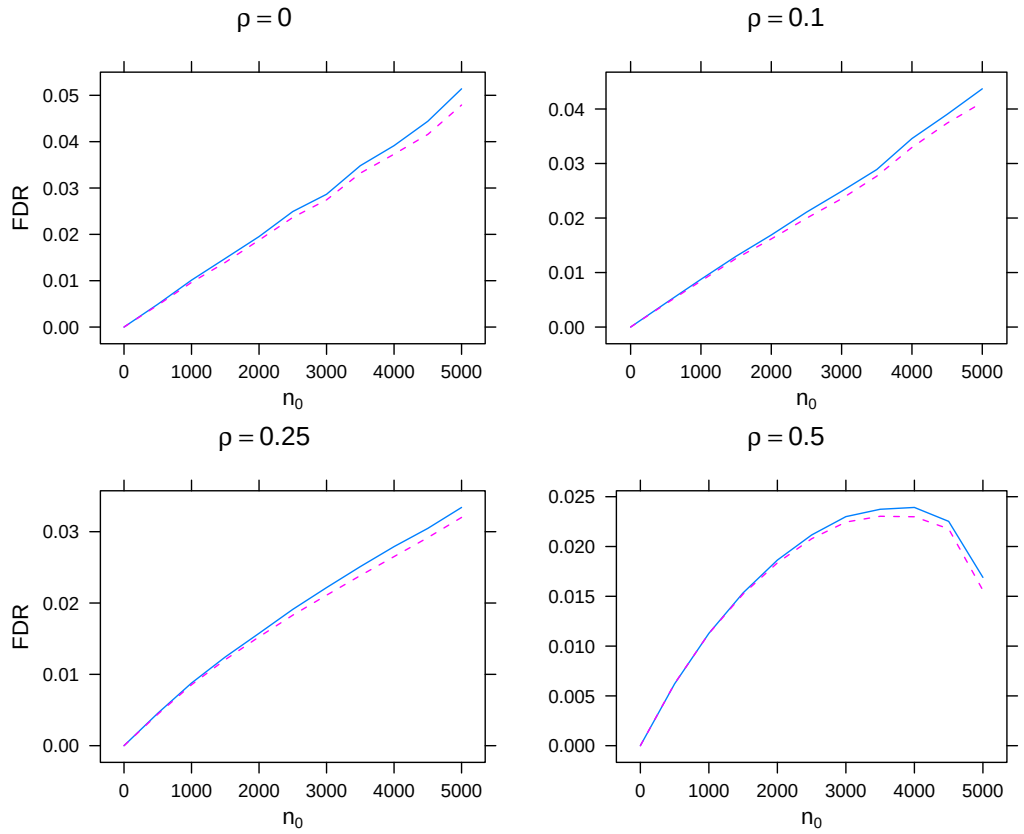

Figure 5.2: Estimated FDR values for $n = 5000$ and $\rho = 0, 0.1, 0.25, 0.5$. Legend: NEW — solid line; BKY — dashed line.

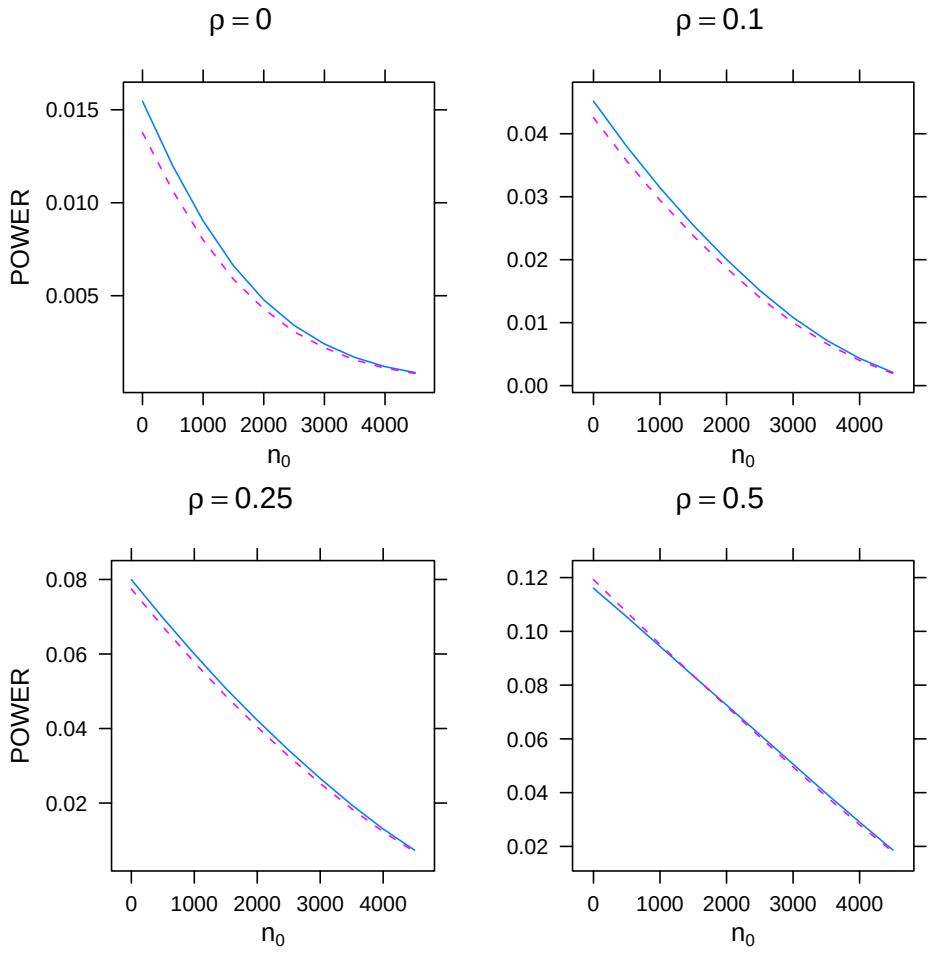

Figure 5.3: Estimated power for $n = 5000$ and $\rho = 0, 0.1, 0.25, 0.5$. Legend: NEW — solid line; BKY — dashed line.

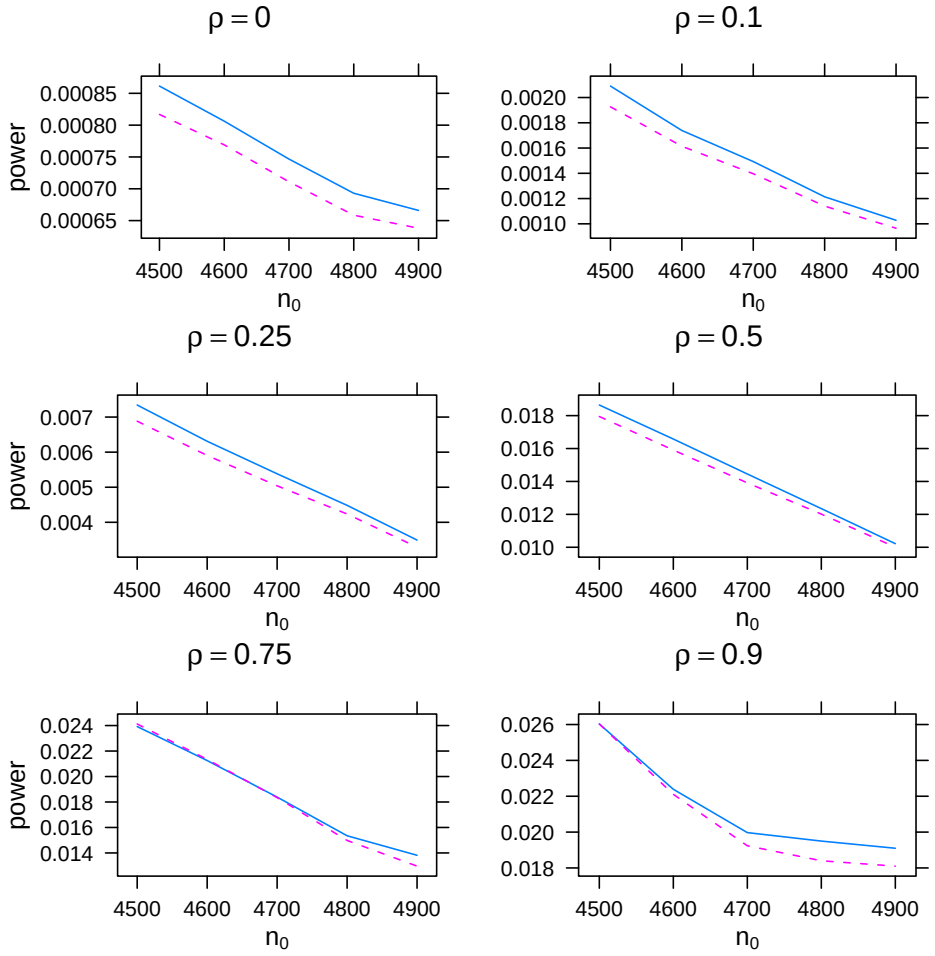

Figure 5.4: Estimated power for $n = 5000$, $\pi_0 = 0.9, 0.92, 0.94, 0.96, 0.98$ and $\rho = 0, 0.1, 0.25, 0.5, 0.75$ and 0.9. Legend: NEW — solid line; BKY — dashed line.

5.3 An Application to Breast Cancer Data

We applied both the new adaptive BH and the BKY methods to the breast cancer data of Hedenfalk et al. (2001) available at http://www.nejm.org/general/content/supplemental/hedenfalk/index.html; see also Storey and Tibshirani (2003) from http://genomine.org/qvalue/results.html. The results are presented in this section.

The data consists of 3,226 genes on 7 BRCA1 arrays, 8 BRCA2 arrays and 7 sporadic tumors. The goal of the study is to establish differences in gene expression patterns between these tumor groups. Here we analyzed this data with permutation t-test to compare BRCA1 and BRCA2. The data were entered into R and all analysis were done using R. As Storey and Tibshirani (2003) did, if any gene had one or more measurement ($\log_2$ expression value) exceeding 20, then this gene was eliminated. This left $n = 3170$ genes for permutation t-test.

We tested each gene for differential expression between BRCA1 and BRCA2 by using a two-sample t-test. The p-values were calculated using a permutation method as in Storey and Tibshirani (2003). We did $B = 100$ permutations for each gene and got a set of null statistics $t_1^{0b}, \ldots, t_n^{0b}$, $b = 1, \ldots, B$. The p-value of the permutation t-test for gene i was calculated by

$$p_i = \sum_{b=1}^{B} \frac{\#\{j : |t_j^{0b}| \geq |t_i|, j = 1, \ldots, n\}}{nB}$$

The new adaptive method identifies 94 significant genes at the 0.05 level of false discovery rate, whereas, the BKY method gets 93 significant genes. This additional significant gene picked up by our method is intercellular adhesion

molecule 2 (clone 471918).

5.4 Remarks

The new adaptive BH method we propose in this chapter competes well with the BKY method. Like the BKY method, it controls the FDR with independent p-values and, as can be seen numerically, continues to maintain the control with the same type of positively dependent p-values as in the BKY method. More importantly, it can perform better than the BKY method in some instances, especially when the proportion of true null hypotheses is very large, as what happens in many applications.

We further explored the property of FDR control of our new adaptive BH method and the BKY method in two ways:

First, we have considered using $\lambda = 0.5$ in the STS method, since this is what Storey, Taylor and Siegmund (2004) have suggested, even though it may not control the FDR under positive dependence, and $\alpha/(1+\alpha)$ for γ in our procedure, since this is what Benjamini, Krieger and Yekutieli (2006) have also considered for the q in their method. All these methods can be proven to control the FDR under independence if other different values are chosen for λ, q and γ. But, the BKY as well as our procedures may not continue to control the FDR under positive dependence with these other values of γ and q.

Second, we checked the FDR control of our new adaptive BH method and the BKY method under block dependence through simulation study. In our simulation study, we generated 5000 dependent random variables $X_i \sim N(\mu_i, 1)$, $i = 1, \ldots, 5000$ with the following covariance structure:

- The variables are grouped into 500 blocks, with every 100 variables being treated as a block. Within each block, the 100 variables follow $AR(1)$ correlation structure with non-negative correlation ρ, that is, $\rho_{i,j} = \rho^{|i-j|}$, for $i, j = 1, \ldots, 100$.

- Variables from different blocks are independent.

The simulation study shows that BKY method can still control FDR under the block dependence. Our new adaptive BH method also shows FDR control when $\pi_0 < 1$, while it breaks down a little from the desired level when $\pi_0 = 1$. After carefully investigating the performance of our new adaptive BH method, we made a slight modification in our new adaptive BH method. The modified version is as follows:

1. Observe $R_{SD}(\gamma_1, \ldots, \gamma_n)$, the number of rejections in a step-down method with the critical values $\gamma_i = i\gamma/n$, $i = 1, \ldots, n$, with $\gamma = \alpha/(1+\alpha)$, and calculate

$$\hat{n}_0^{NEW^*} = \begin{cases} \frac{n - R_{SD}(\gamma_1,\ldots,\gamma_n)+1}{1-\gamma_{R_{SD}(\gamma_1,\ldots,\gamma_n)+1}}, & R_{SD}(\gamma_1, \ldots, \gamma_n) < n, \\ n(1+\alpha), & R_{SD}(\gamma_1, \ldots, \gamma_n) = n. \end{cases}$$

   .

2. Apply the step up procedure with the critical values $\alpha_i = i\alpha/\hat{n}_0^{NEW^*}$, $i = 1, \ldots, n$, for testing the null hypotheses.

Actually, this modification only changes the critical values of new adaptive BH method in Step 2 to the critical values of the BKY method in Step 2 when all the hypotheses are rejected in the Step 1. It is easy to prove that the modified version of our new adaptive BH method can control FDR under

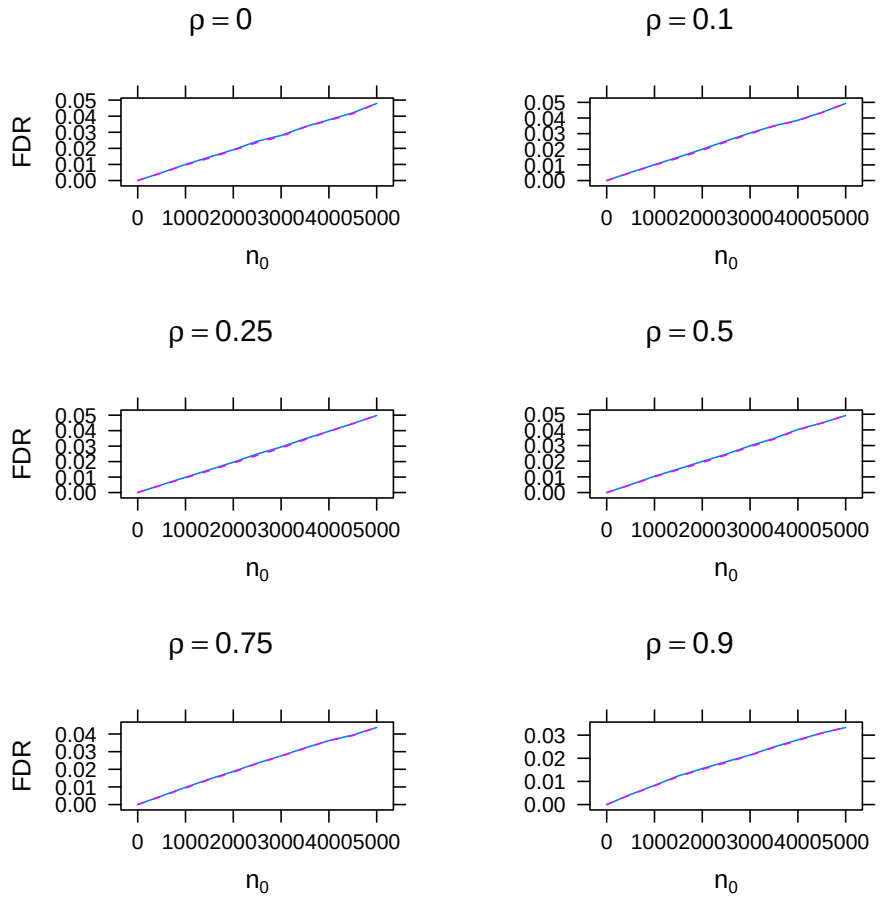

Figure 5.5: Estimated FDR for $n = 5000$ and $\rho = 0, 0.1, 0.25, 0.5, 0.75, 0.9$ under block dependence. Legend: MODIFIED NEW — solid line; BKY — dashed line.

independence and it also shows FDR control in simulation when the p-values are positively dependent under normal distributional setting with equal positive correlation. More over, simulation study shows that the modified version controls FDR under the block dependence and is more powerful than the BKY method, which is displayed in Figures 5.5 and 5.6.

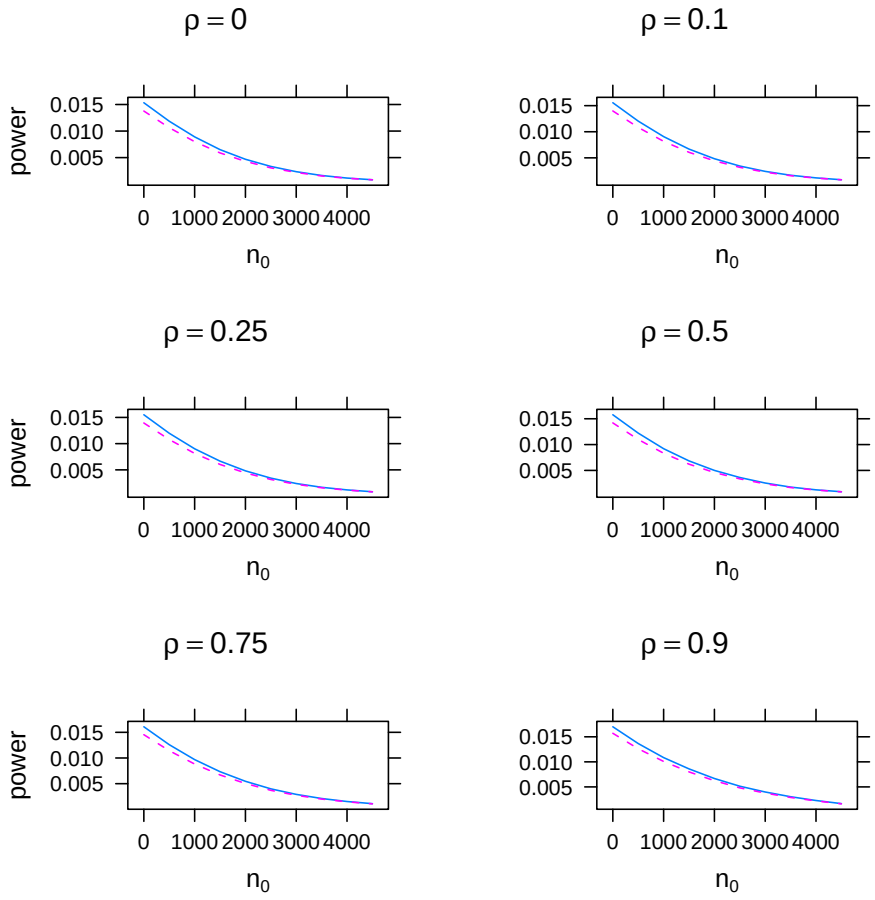

Figure 5.6: Estimated power for $n = 5000$ and $\rho = 0, 0.1, 0.25, 0.5, 0.75, 0.9$ under block dependence. Legend: MODIFIED NEW — solid line; BKY — dashed line.

CHAPTER 6

NEW CLASS OF FDR ESTIMATE

In this chapter, first we present our class of estimates of the FDR which is also based on mixture model assumption as the one in Storey (2002). Then we report the results of a simulation study that we conducted to numerically investigate the extent of improvements we get by using our new FDR estimates over Storey's in terms of the conservative bias and MSE.

6.1 New FDR Estimate

Our new point estimate of the FDR is derived from an identity for the FDR given in the following theorem.

Theorem 6.1 *Under the mixture model described in Section 3.2, the FDR of a single-step test rejecting each H_i if $p_i \leq t$, for some fixed $t \in [0,1]$, is given by*

$$\mathrm{FDR}_n(t) = (n+1)\pi_0 t E\left(\frac{1}{max\{R_n(t), \frac{1}{n}\}+1}\right). \tag{6.1}$$

Proof. Let $R_{n-1}^{(-i)}(t) = \sum_{j(\neq i)=1}^{n} I(P_j \leq t)$,

$$\begin{aligned}
\mathrm{FDR}_n(t) &= E\{\tfrac{V_n(t)}{R_n(t)} I(R_n(t) > 0)\} \\
&= E\{\sum_{i=1}^{n}\sum_{r=1}^{n} \frac{1}{r} I(P_i \leq t, H_i = 0, R_n(t) = r)\} \\
&= \sum_{i=1}^{n} Pr(P_i \leq t, H_i = 0) \sum_{r=1}^{n} \frac{1}{r} Pr(R_n^{(-i)}(t) = r-1) \\
&= \pi_0 t E\left(\sum_{i=1}^{n} \frac{1}{R_n^{(-i)}(t)+1}\right)
\end{aligned} \tag{6.2}$$

When $R_n^{(-i)}(t) > 0$, since $R_n^{(-i)}(t) = R_n(t) - I(P_i \leq t)$,

$$\begin{aligned}
\sum_{i=1}^{n} \frac{1}{R_n^{(-i)}(t)+1} &= \textstyle\sum_{i:P_i \leq t} \frac{1}{R_n^{(-i)}(t)+1} + \sum_{i:P_i > t} \frac{1}{R_n^{(-i)}(t)+1} \\
&= \frac{R_n(t)}{R_n(t)} + \frac{n - R_n(t)}{R_n(t)+1} \\
&= \frac{n+1}{R_n(t)+1}.
\end{aligned}$$

When $R_n^{(-i)}(t) = 0$, $\sum_{i=1}^{n} \frac{1}{R_n^{(-i)}(t)+1} = n$. Therefore, the right-hand side in (6.2) is equal to

$$(n+1)\pi_0 t E\left(\frac{1}{\max\{R_n(t), \frac{1}{n}\}+1}\right).$$

This proves the theorem.

It is easy to see that the expression (6.1) is the same as the expression (3.4), which is commonly seen in the literature. Nevertheless, the new expression provides directly an unbiased estimate of $\mathrm{FDR}_n(t)$ if π_0 is known, which is given by

$$\widehat{\mathrm{FDR}}_n(t|\pi_0) = \frac{(n+1)\pi_0 t}{\max\{R_n(t), \frac{1}{n}\} + 1}. \tag{6.3}$$

Since $R_n(t)$ is complete sufficient for $F(t)$, (6.3) is the uniformly minimum variance unbiased estimate (UMVUE) of the FDR_n with a known π_0.

Remark 6.1. It is important to note the difference between the approaches taken here and in other papers, like Dalmasso et al. (2005) and Pounds and Cheng (2006), towards reducing the conservative bias of the estimate $\widehat{\mathrm{FDR}}_{n,\lambda}(t)$ of $\mathrm{FDR}_n(t)$. We start with the best unbiased estimate of $\mathrm{FDR}_n(t)$ assuming π_0 known before using a conservatively biased point estimate of π_0, while the others have started with the estimate $\pi_0 t/\hat{F}_n(t)$ when π_0 is known that is already conservatively biased before developing an estimate of π_0 better than Storey's, thus attempting to improve $\widehat{\mathrm{FDR}}_{n,\lambda}(t)$. It would be interesting to see what happens if one uses the type of estimates of π_0 suggested in these other papers in our UMVUE, instead of using the same estimate as Storey originally used. While we expect that this would produce better performing estimates of $\mathrm{FDR}_n(t)$ than ours when compared to $\widehat{\mathrm{FDR}}_{n,\lambda}(t)$, we have to rely on simulations to see it, as seeing it theoretically seems difficult. The only reason we have kept Storey's estimate of π_0 while replacing the π_0 in the UMVUE is that we can theoretically see why the resulting estimate is in fact less conservatively biased than (3.6). As seen numerically in the next sub-section, such

an estimate also often outperforms (3.6) in terms of the MSE, especially in small-scale multiple testings.

With unknown π_0, we use the same estimate of it as in Storey (2002) (3.7) and derive the following class of estimates of $\mathrm{FDR}_n(t)$:

$$\begin{aligned} \widehat{\mathrm{FDR}}^{*}_{n,\lambda}(t) &= \frac{(n+1)\hat{\pi}_0(\lambda)t}{\max\{R_n(t), \frac{1}{n}\}+1} \\ &= \frac{(n+1)[n-R_n(\lambda)]t}{n(1-\lambda)[\max\{R_n(t), \frac{1}{n}\}+1]}, \qquad \lambda \in [0,1]. \end{aligned} \tag{6.4}$$

The proposed estimate $\widehat{\mathrm{FDR}}^{*}_{n,\lambda}(t)$ has better finite sample properties than the estimate $\widehat{\mathrm{FDR}}_{n,\lambda}(t)$ given in Storey (2002), where

$$\widehat{\mathrm{FDR}}_{n,\lambda}(t) = \frac{\hat{\pi}_0(\lambda)t}{\{R(t) \bigvee 1\}/n},$$

as given in (3.6), for any fixed λ and t, though they become asymptotically equivalent as $n \to \infty$. More specifically, we have the following theorem.

Theorem 6.2 *For any fixed* $\lambda, t \in [0,1]$,

$$\mathrm{FDR}_n(t) \le E(\widehat{\mathrm{FDR}}^{*}_{n,\lambda}(t)) \le E(\widehat{\mathrm{FDR}}_{n,\lambda}(t)). \tag{6.5}$$

Proof. Let us suppress the subscript n in R_n.

$$E(\widehat{\mathrm{FDR}}^{*}_{n,\lambda}(t)) = (n+1)tE\left\{\hat{\pi}_0(\lambda)E\left(\frac{1}{\max\{R(t), \frac{1}{n}\}+1}|R(\lambda)\right)\right\}. \tag{6.6}$$

Since,

$$\begin{aligned} &R(t)|R(\lambda) \sim \text{binomial}(R(\lambda), \tfrac{\mathrm{F}(t)}{\mathrm{F}(\lambda)}), && \text{when} \quad t \le \lambda; \\ &n-R(t)|R(\lambda) \sim \text{binomial}(n-R(\lambda), \tfrac{1-\mathrm{F}(t)}{1-\mathrm{F}(\lambda)}), && \text{when} \quad t > \lambda, \end{aligned}$$

given $R(\lambda)$, $R(t)$ is stochastically increasing in $R(\lambda)$. Therefore, $E\left(\frac{1}{\max\{R(t), \frac{1}{n}\}+1}|R(\lambda)\right)$ is a decreasing function of $R(\lambda)$. Also, $\hat{\pi}_0(\lambda)$ is a

decreasing function of $R(\lambda)$, which implies that the outer expectation in (6.6) is greater than or equal to

$$E\{\hat{\pi}_0(\lambda)\}E\left\{E\left\{\left(\frac{1}{\max\{R(t),\frac{1}{n}\}+1}|R(\lambda)\right)\right\}\right\}$$

$$= E\{\hat{\pi}_0(\lambda)\}E\left\{\left(\frac{1}{\max\{R(t),\frac{1}{n}\}+1}\right)\right\} \tag{6.7}$$

$$\geq \pi_0 E\left\{\left(\frac{1}{\max\{R(t),\frac{1}{n}\}+1}\right)\right\},$$

with the inequality following from the fact that $\hat{\pi}_0$ is upward biased as shown in Storey (2002). Thus, we have

$$E(\widehat{\mathrm{FDR}}^{*}_{n,\lambda}(t)) \geq (n+1)t\pi_0 E\left\{\frac{1}{\max\{R(t),\frac{1}{n}\}+1}\right\} = \mathrm{FDR}_n(t), \tag{6.8}$$

which is the first inequality of (6.5). The second inequality follows from the fact that $\widehat{\mathrm{FDR}}^{*}_{n,\lambda}(t) \leq \widehat{\mathrm{FDR}}_{n,\lambda}(t)$ with probability one.

Comparing Storey's estimate $\widehat{\mathrm{FDR}}_{n,\lambda}(t)$ (3.6) with our new FDR estimate $\widehat{\mathrm{FDR}}^{*}_{n,\lambda}(t)$ (6.4), it is easy to see that for smaller R, the difference between $\widehat{\mathrm{FDR}}_{n,\lambda}(t)$ and $\widehat{\mathrm{FDR}}^{*}_{n,\lambda}(t)$ will be larger. As π_0 increases, there is greater possibility that $R_n(t)$ will decrease and $\widehat{\mathrm{FDR}}^{*}_{n,\lambda}(t)$ will show more advantage than $\widehat{\mathrm{FDR}}_{n,\lambda}(t)$.

6.2 Simulation Study

We conducted a simulation study to investigate the reduction in bias that our estimate of $\widehat{\mathrm{FDR}}^{*}_{n,\lambda}(t)$ offer compared to Storey's estimate $\widehat{\mathrm{FDR}}_{n,\lambda}(t)$. In particular, we generated n independent pairs of observations (X_i, Z_i), $i =$

$1, \cdots, n$, where Z_i is the outcome of a Bernoulli experiment with $1 - \pi_0$ as the success probability and $X_i \sim N(\mu_i, 1)$, with $\mu_i = \delta Z_i$ for a fixed $\delta \geq 0$. For each X_i, we calculated the p-value $p_i = 1 - \Phi(X_i)$, where Φ is the cdf of $N(0, 1)$, and considered testing $\mu_i = 0$ against $\mu_i = \delta$ based on the rejection region $\{p_i \leq t\}$. It was repeated over 20000 times before simulating the bias and MSE for each estimate and for each combination of values of π_0, t, n and δ selected from the range $0.1, 0.2, \ldots, 0.9$, in increments of 0.1, for π_0, the two values 0.001 and 0.01 for t, the two values 100 and 500 for n, and the value 1 for δ. For λ, we considered the values 0.5, which Storey (2002) originally used, and 0.95, which is often used in practice. The formula (3.4) was used to compute the true FDR for the bias and MSE calculations.

The numerical findings are summarized in Tables 6.4 (for λ=0.5) and 6.4 (for λ=0.95). The proposed estimate $\widehat{\mathrm{FDR}}^{*}_{n,\lambda}(t)$ is seen to have better performance than Storey's both in terms of the conservative bias and MSE, with the improvement in terms of the bias appearing quite significant particularly when π_0 is large.

6.3 A Connection To the BH method

We modify the BH method in this section by redefining it in terms of the new threshold $t_\alpha(\widehat{\mathrm{FDR}}^{*}_{n,\lambda=0})$, where

$$t_\alpha(\widehat{\mathrm{FDR}}^{*}_{n,\lambda}) = \sup\{0 \leq t \leq 1 : \widehat{\mathrm{FDR}}^{*}_{n,\lambda}(t) \leq \alpha\}.$$

Clearly, this is more powerful than the original BH method, since $t_\alpha(\widehat{\mathrm{FDR}}_{n,\lambda=0}) \leq t_\alpha(\widehat{\mathrm{FDR}}^{*}_{n,\lambda=0})$, and, as stated in the following theorem, it is often a valid FDR-controlling method.

Theorem 6.3 *Assume that* $Pr(P_i \leq u) \leq \mu$, *for each* P_i *when the corresponding null hypothesis is true, and the p-values are independent or positively dependent in the following sense:*

$$\mathrm{E}\left\{\psi\left(P_1, \ldots, P_n\right) \mid P_i \leq u\right\} \uparrow u \in (0,1), \tag{6.9}$$

for each P_i *that corresponds to a null hypothesis and any increasing (coordinatewise) function* ψ. *Also, assume that the proportion of true null hypotheses,* π_0, *is random and distributed as* $U(0,1)$. *Then, the FDR of the method that rejects* H_i *if* $p_i \leq t_\alpha(\widehat{\mathrm{FDR}}^*_{n,\lambda=0})$ *is less than or equal to* α.

Before we proceed to explain how to prove this theorem, it is important to keep in mind that the FDR here is actually the unconditional FDR defined as

$$\text{Unconditional FDR} = \int_0^1 \mathrm{FDR}(\pi_0) d\pi_0,$$

where $\mathrm{FDR}(\pi_0)$ is the conditional FDR given as π_0.

This theorem can be proved somewhat in the same way, that is, using the following lemma, as in providing the FDR control of the BH method with a fixed π_0 under the same type of positive dependency as assumed in this theorem, once we make the same kind of arguments as in Lemma 1 or 2 of Storey et al. (2004) and see that thresholding the p-values at $t_\alpha(\widehat{\mathrm{FDR}}^*_{n,\lambda=0})$ is equivalent to the following stepup test: Reject H_i if $p_i \leq p_{(l)}$, where

$$l = \max\Big\{1 \leq i \leq n : p_{(n)} \leq \frac{(i+1)\alpha}{n+1}\Big\},$$

provided the maximum exists, otherwise, accept all null hypotheses. Several versions of a proof of this lemma can be seen in the literature [e.g., Benjamini and Yekutieli (2001), Sarkar (2002), Blanchard and Roquain (2008)], so we ignore proving it here.

Lemma 6.3. *Consider a stepup test with the critical values* $\alpha_1 \leq \ldots \leq \alpha_n$, *that is, reject* H_i *if* $p_i \leq p_{(r)}$, *where*

$$r = max\{1 \leq i \leq n : p_{(i)} \leq \alpha_i\},$$

provided the maximum exists, otherwise, accept all null hypotheses. Given n_0 *true null hypotheses, the FDR of this method is less than or equal to* $n_0\alpha_1$ *under the positive dependency assumed in Theorem 6.3 if* α_i/i *is non-increasing in* i.

Proof of Theorem 6.3. From Lemma 6.3, we see that the unconditional FDR of the method in this theorem satisfies the following:

$$\text{Unconditional FDR} = \int_0^1 \text{FDR}(\pi_0)d\pi_0 \leq \frac{2n\alpha}{n+1}\int_0^1 \pi_0 d\pi_0 \leq \alpha.$$

Thus, the theorem is proved.

6.4 About Black's Comments

Black (2004) examined the performance of Storey's FDR estimate $\widehat{\text{FDR}}_{n,\lambda}(t)$ (3.6) and found two main deficiencies of this estimate. First, the bias between $\hat{\pi_0}$ and true π_0 is very large when the distance between the distributions associated with the two hypotheses is small. Second, Black compared the fixed rejection region approach using $\widehat{\text{FDR}}_{n,\lambda}(t)$ with the adaptive BH method to control FDR at level $\alpha = \widehat{\text{FDR}}_{n,\lambda}(t)/\hat{\pi}_0$. The simulation showed that adaptive method outperforms the fixed rejection region method a little, although Storey (2002) commented that these two methods should be identical. Black mentioned that the differences between these two methods arise as a consequence of various inaccuracies in the estimates of π_0 and FDR(t). This line needs to be clarified.

As we use the same estimate of π_0 as Storey (2002), the same problem of the $\hat{\pi_0}$ still exists. However, the new FDR estimate lightened the deficiency from the $\hat{\pi_0}$ as the bias of $\widehat{\mathrm{FDR}}^{*}_{n,\lambda}(t)$ is smaller than $\widehat{\mathrm{FDR}}_{n,\lambda}(t)$.

The second deficiency comes from the misunderstanding of the fixed rejection region method. In Storey (2002), he described a new method to fix the rejection region first and then estimate the FDR, which is called the estimation based approach. As we can see, the estimate of FDR may not be monotone with the threshold t, which means that with a larger t, the estimate of FDR may be smaller. Therefore, we cannot use the fixed rejection region method even to control FDR at level $\widehat{\mathrm{FDR}}_{\lambda}(t)$. The adaptive step-up BH method to control FDR at level $\alpha/\hat{\pi}_0$ is the same as the procedure with which rejecting $H_{1:m}, \ldots, H_{r:m}$, such that

$$r = \max\{i : \widehat{\mathrm{FDR}}_{\lambda}(P_{i:m}) \leq \alpha\}.$$

Since

$$\widehat{\mathrm{FDR}}_{\lambda}(P_{r:m}) = \frac{\hat{\pi}_0 P_{r:m}}{r/m},$$

this is equivalent to

$$r = \max\{i : P_{i:m} \leq \frac{i\alpha}{m\hat{\pi}_0}\}. \tag{6.10}$$

Now suppose that there exists t_1, t_2 and $t_1 < t_2$, so that $\widehat{\mathrm{FDR}}_{\lambda}(t_1) = \widehat{\mathrm{FDR}}_{\lambda}(t_2) = \alpha$. If we set $t = t_1$ in the fixed rejection region procedure, we will get $\widehat{\mathrm{FDR}}_{\lambda}(t_1) = \alpha$. With the adaptive step-up BH method to control FDR at level $\widehat{\mathrm{FDR}}_{\lambda}(t_1)/\hat{\pi}_0$, actually we will reject hypotheses with $P_i \leq t_2$. Therefore, the adaptive step-up BH method to control FDR at level $\widehat{\mathrm{FDR}}_{\lambda}(t)/\hat{\pi}_0$ will always be the same or more powerful than the fixed rejection region method.

Table 6.1: Numerical comparison between $\widehat{\mathrm{FDR}}_{n,\lambda}(t)$ and $\widehat{\mathrm{FDR}}^{*}_{n,\lambda}(t)$ in terms of the bias and MSE ($\lambda = 0.5$)

		$n = 500$					$n = 100$			
		Bias		MSE			Bias		MSE	
π_0	FDR	$\widehat{\mathrm{FDR}}_{\lambda}(t)$	$\widehat{\mathrm{FDR}}^{*}_{\lambda}(t)$	$\widehat{\mathrm{FDR}}_{\lambda}(t)$	$\widehat{\mathrm{FDR}}^{*}_{\lambda}(t)$	FDR	$\widehat{\mathrm{FDR}}_{\lambda}(t)$	$\widehat{\mathrm{FDR}}^{*}_{\lambda}(t)$	$\widehat{\mathrm{FDR}}_{\lambda}(t)$	$\widehat{\mathrm{FDR}}^{*}_{\lambda}(t)$
$t = 0.01$										
0.1	0.012	0.03511	0.03405	0.00131	0.00123	0.012	0.04140	0.03441	0.00268	0.00166
0.2	0.026	0.03515	0.03361	0.00138	0.00126	0.026	0.04460	0.03407	0.00384	0.00201
0.3	0.044	0.03533	0.03305	0.00150	0.00132	0.044	0.04983	0.03361	0.00630	0.00266
0.4	0.067	0.03572	0.03233	0.00175	0.00146	0.067	0.05797	0.03291	0.01102	0.00388
0.5	0.098	0.03665	0.03144	0.00229	0.00180	0.097	0.07073	0.03212	0.02029	0.00670
0.6	0.140	0.03861	0.03019	0.00354	0.00257	0.138	0.09158	0.03066	0.03709	0.01131
0.7	0.202	0.04299	0.02834	0.00692	0.00464	0.196	0.12113	0.02819	0.06396	0.02084
0.8	0.302	0.05457	0.02551	0.01845	0.01093	0.283	0.16329	0.02352	0.10229	0.03732
0.9	0.493	0.09397	0.01998	0.08883	0.04049	0.418	0.20844	0.01485	0.13695	0.06374
0.95	0.672	0.16051	0.01628	0.29442	0.10346	0.513	0.22469	0.00946	0.14044	0.07928
$t = 0.001$										
0.1	0.006	0.02124	0.01735	0.00068	0.00040	0.005	0.02248	0.01413	0.00069	0.00033
0.2	0.013	0.02322	0.01722	0.00103	0.00049	0.010	0.02354	0.01347	0.00079	0.00037
0.3	0.023	0.02616	0.01699	0.00169	0.00066	0.017	0.02433	0.01264	0.00088	0.00042
0.4	0.035	0.03094	0.01669	0.00301	0.00099	0.024	0.02478	0.01162	0.00093	0.00047
0.5	0.051	0.04026	0.01736	0.00597	0.00188	0.032	0.02451	0.01033	0.00094	0.00053
0.6	0.074	0.05334	0.01692	0.01087	0.00323	0.042	0.02368	0.00890	0.00089	0.00059
0.7	0.108	0.07214	0.01626	0.01854	0.00603	0.053	0.02164	0.00709	0.00077	0.00062
0.8	0.160	0.09688	0.01439	0.02824	0.01086	0.065	0.01808	0.00507	0.00057	0.00061
0.9	0.246	0.11543	0.01022	0.03189	0.01779	0.079	0.01260	0.00262	0.00033	0.00053
0.95	0.310	0.11072	0.00610	0.02601	0.02025	0.087	0.00887	0.00135	0.00022	0.00044

Table 6.2: Numerical comparison between $\widehat{\mathrm{FDR}}_{n,\lambda}(t)$ and $\widehat{\mathrm{FDR}}^{*}_{n,\lambda}(t)$ in terms of the bias and MSE ($\lambda = 0.95$)

		$n = 500$					$n = 100$			
		Bias		MSE			Bias		MSE	
π_0	FDR	$\widehat{\mathrm{FDR}}_{\lambda}(t)$	$\widehat{\mathrm{FDR}}^{*}_{\lambda}(t)$	$\widehat{\mathrm{FDR}}_{\lambda}(t)$	$\widehat{\mathrm{FDR}}^{*}_{\lambda}(t)$	FDR	$\widehat{\mathrm{FDR}}_{\lambda}(t)$	$\widehat{\mathrm{FDR}}^{*}_{\lambda}(t)$	$\widehat{\mathrm{FDR}}_{\lambda}(t)$	$\widehat{\mathrm{FDR}}^{*}_{\lambda}(t)$
$t = 0.01$										
0.1	0.012	0.00942	0.00894	0.00021	0.00019	0.012	0.01230	0.00910	0.00119	0.00074
0.2	0.026	0.00976	0.00885	0.00034	0.00031	0.026	0.01531	0.00909	0.00255	0.00146
0.3	0.044	0.01034	0.00878	0.00056	0.00050	0.044	0.02037	0.00915	0.00529	0.00262
0.4	0.067	0.01119	0.00861	0.00095	0.00083	0.067	0.02837	0.00913	0.01079	0.00472
0.5	0.098	0.01252	0.00824	0.00168	0.00142	0.097	0.04001	0.00836	0.02041	0.00811
0.6	0.140	0.01512	0.00781	0.00332	0.00270	0.138	0.06107	0.00827	0.04145	0.01572
0.7	0.202	0.02077	0.00742	0.00754	0.00574	0.196	0.09230	0.00798	0.07868	0.03074
0.8	0.302	0.03455	0.00707	0.02125	0.01424	0.283	0.13888	0.00643	0.14050	0.05729
0.9	0.493	0.07789	0.00580	0.09861	0.04917	0.418	0.19080	0.00286	0.21361	0.10564
0.95	0.672	0.14930	0.00688	0.32371	0.12495	0.513	0.21397	0.00183	0.24647	0.14011
$t = 0.001$										
0.1	0.006	0.00636	0.00459	0.00013	0.00007	0.005	0.00746	0.00372	0.00029	0.00015
0.2	0.013	0.00808	0.00456	0.00032	0.00014	0.010	0.00939	0.00355	0.00049	0.00025
0.3	0.023	0.01084	0.00452	0.00077	0.00030	0.017	0.01132	0.00335	0.00072	0.00039
0.4	0.035	0.01529	0.00440	0.00175	0.00060	0.024	0.01312	0.00314	0.00098	0.00058
0.5	0.051	0.02381	0.00501	0.00409	0.00138	0.032	0.01415	0.00256	0.00121	0.00078
0.6	0.074	0.03642	0.00484	0.00839	0.00272	0.042	0.01498	0.00220	0.00145	0.00104
0.7	0.108	0.05564	0.00486	0.01598	0.00564	0.053	0.01489	0.00171	0.00165	0.00131
0.8	0.160	0.08222	0.00449	0.02688	0.01118	0.065	0.01333	0.00115	0.00177	0.00159
0.9	0.246	0.10536	0.00317	0.03436	0.01999	0.079	0.00997	0.00035	0.00184	0.00183
0.95	0.310	0.10468	0.00166	0.03144	0.02420	0.087	0.00745	0.00008	0.00187	0.00193

CHAPTER 7

New Q-value method

In this chapter, we utilize our new classe of FDR estimates in Chapter 6 to develop a new q-value method.

7.1 Estimating the pFDR

Under the mixture model, Storey (2002) extended his FDR estimate to a conservative pFDR estimate as

$$\widehat{\mathrm{pFDR}}_{n,\lambda}(t) = \frac{\widehat{\mathrm{FDR}}_{n,\lambda}(t)}{1-(1-t)^n}, \quad \lambda \in [0,1). \tag{7.1}$$

Using the same lower bound $1-(1-t)^n$ to estimate $\Pr\{R(t)>0\}$ as in Storey (2002), a new class of pFDR estimate can be derived based on our new FDR estimate (6.4) under mixture model as

$$\widehat{\mathrm{pFDR}}^{*}_{n,\lambda}(t) = \frac{\widehat{\mathrm{FDR}}^{*}_{n,\lambda}(t)}{1-(1-t)^n} = \frac{(n+1)\hat{\pi}_0(\lambda)t}{[\max\{R_n(t), \frac{1}{n}\}+1]\{1-(1-t)^n\}}, \quad \lambda \in [0,1), \tag{7.2}$$

with the same $\hat{\pi}_0(\lambda)$ as in (3.7).

The new class of pFDR estimate (7.2) has uniformly smaller bias than Storey's pFDR estimate (7.1), as shown in the following theorem.

Theorem 7.1 *For any fixed* $\lambda, t \in [0,1]$,

$$\mathrm{pFDR}_n(t) \leq E(\widehat{\mathrm{pFDR}}^*_{n,\lambda}(t)) \leq E(\widehat{\mathrm{pFDR}}_{n,\lambda}(t)). \tag{7.3}$$

Proof. The result is obvious by Theorem 6.2 and the fact that $\Pr(R > 0) \geq 1 - (1-t)^n$ for any t.

We conducted a simulation study to investigate the reductions in bias that our estimates of $\mathrm{pFDR}_n(t)$ offer compared to Storey's estimate. We do not report the simulation results here as it is similar to the results of the FDR estimate comparison in Section 6.2.

7.2 The modified q-value method and an application

Now we have an estimate of the pFDR less conservatively biased than what Storey (2002) used while estimating this q-value, it would be natural to use this alternative estimate to modify Storey's q-value method. The modified estimated q-value of an observed p-value p_{obs} is

$$\hat{q}^*(p_{obs}) = \inf_{t \geq p_{obs}} \{\widehat{\mathrm{pFDR}}^*_{n,\lambda}(t)\}. \tag{7.4}$$

We applied both the original Storey's and our modified q-value method to the breast cancer data which is introduced in Section 5.3. The data consist of 3,226 genes on $n_1 = 7$ BRCA1 arrays, $n_2 = 8$ BRCA2 arrays and $n_3 = 7$ sporadic tumors, which have been analyzed in numerous other articles, often

using only the BRCA1 and BRCA2 groups and sometimes using all three groups. Here, we considered both these cases, and used a permutation t-test to compare BRCA1 and BRCA2 and permutation F test to compare the three groups. As in Storey and Tibshirani (2003), if any gene had one or more measurement ($\log_2$ expression value) exceeding 20, then this gene was eliminated. This left $n = 3170$ genes for permutation t-test and $n = 3169$ genes for permutation F test.

The p-values of the permutation t-test were calculated with formula (5.3).The p-values of permutation F test were calculated in a similar way.

For both the t and F tests, we calculated the q-values using Storey's original estimate (3.11) as well as our new estimate(7.4). Data was entered into R and all the analysis was conducted with R. Storey's q-value method is applied with the "qvalue" function in R library "qvalue".

For both cases, λ is decided in three different ways.

- Fixed λ method: $\lambda = 0.5$ (Storey (2002)).
- Smoother method (Storey and Tibshirani (2003)).
- Bootstrap method (Storey, Taylor & Siegmund (2004)).

The results of the permutation t-test are shown in Table 7.1. As expected, the new q-value method picks up at least the same number of significant genes as Storey's method. For example, with the fixed $\lambda = 0.5$, by thresholding the q-value at 0.03, 0.05 or 0.07 finds 76, 161 and 231 significant genes, respectively with our new method, while these numbers are 76, 159 and 229, respectively, for the Storey's method. Figure 7.1 shows the significantly expressed genes

Table 7.1: Number of Significant Genes for Permutation T-test

λ	threshold	Storey	New	More Significant Genes
0.5	0.03	76	76	
	0.05	159	161	M-phase phosphoprotein 4 (clone 785816), v-yes-1 Yamaguchi sarcoma viral oncogene homolog 1 (clone 273435)
	0.07	229	231	actin, beta (clone 34357), ligase I (clone 29627)
	0.1	314	314	
Smooth	0.03	76	76	
Spline	0.05	162	162	
	0.07	233	233	
	0.1	319	319	
Bootstrap	0.03	76	76	
	0.05	161	162	acyl-Coenzyme A oxidase (clone 210862)
	0.07	231	232	thyroid hormone receptor interactor 7 (clone 781704)
	0.1	317	319	KIAA0303 protein (clone 768031), ESTs, Weakly similar to putative p150 (clone 137417)

Table 7.2: Number of Significant Genes for Permutation F Test

λ	threshold	Storey	New	More Significant Genes
0.5	0.05	77	77	
	0.07	180	181	EphB4 (clone 75009)
	0.1	293	300	aldo-keto reductase family 1 gene (clone 33008), ataxia-telangiectasia group D-associated protein gene (clone 377275), actin (clone 34357), chromosome 14 open reading frame 2 (clone 110772), fibrogenic lymphokine (clone 366580), tyrosine kinase 2 (clone 756452), synuclein (clone 40764)
Smooth	0.05	124	138	KIAA0265 protein (clone 140716), adenylosuccinate lyase (clone 813280), Ras suppressor protein 1 (clone 687397), gamma-aminobutyric acid (GABA) A receptor, pi (clone 563598), ESTs, Weakly similar to trg [R.norvegicus] (clone 246749), ESTs, Weakly similar to B0495.6 [C.elegans] (clone 144926), protein phosphatase 2 (formerly 2A), regulatory subunit A (clone 205490), MAD (mothers against decapentaplegic, Drosophila) homolog 4 (clone 140827), cAMP responsive element binding protein 1 (clone 362332), DKFZP434D1335 protein (clone 161195), PDZ domain containing guanine nucleotide exchange factor 1 (clone 824895), lung resistance-related protein (clone 591281), cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4) (clone 291057), ESTs, Weakly similar to ORF YDL040c [S.cerevisiae] (clone 196866)
Spline	0.07	212	212	
	0.1	381	385	thyroid hormone receptor interactor 6 (clone 811108), sarcoma amplified sequence (clone 361692), transcription elongation factor A (SII), 1 (clone 257458), LIM and SH3 protein 1 (clone 44050)
Bootstrap	0.05	138	139	ribosomal protein L38 (clone 839594)
	0.07	218	218	
	0.1	392	393	mitogen-activated protein kinase kinase 4 (clone 726147)

under the cutoff 0.05. The two triangles above the line q-value$= 0.05$ are the two more genes that our new method found significant, the 160th and 161th most significant genes. They are M-phase phosphoprotein 4 (Symbol:ILF3, clone 785816) gene and v-yes-1 Yamaguchi sarcoma viral oncogene homolog 1 (Symbol:YES1, clone 273435) gene. These two genes are already found to be related with breast cancer in the literature. Vumbaca et al.(2008) found that ILF3 promotes breast cancer growth and angiogenesis in vivo and Sommera et al.(2005) stated that changes in the expression of YES1 have been associated with the aggressiveness of human breast cancer cells. There is a possibility that these two genes are only related to one of the BRCA1 and BRCA2 mutation, which needs to be further proved biologically.

Similarly, for the permutation F-test, the number of significant genes identified by the new q-value method is more than those found by Storey's method, as shown in Table 7.2. When we set the cutoff at 0.1 with $\lambda = 0.5$, we find 7 more significant genes with the new q-value method than the Storey's method. Figure 7.2 displays the q-values from both sets with $\lambda = 0.5$. The q-values of our new method are uniformly smaller than Storey's.

Therefore, our new q-value method is uniformly powerful than Storey's as the new method can achieve at least the same number of rejections as Storey's.

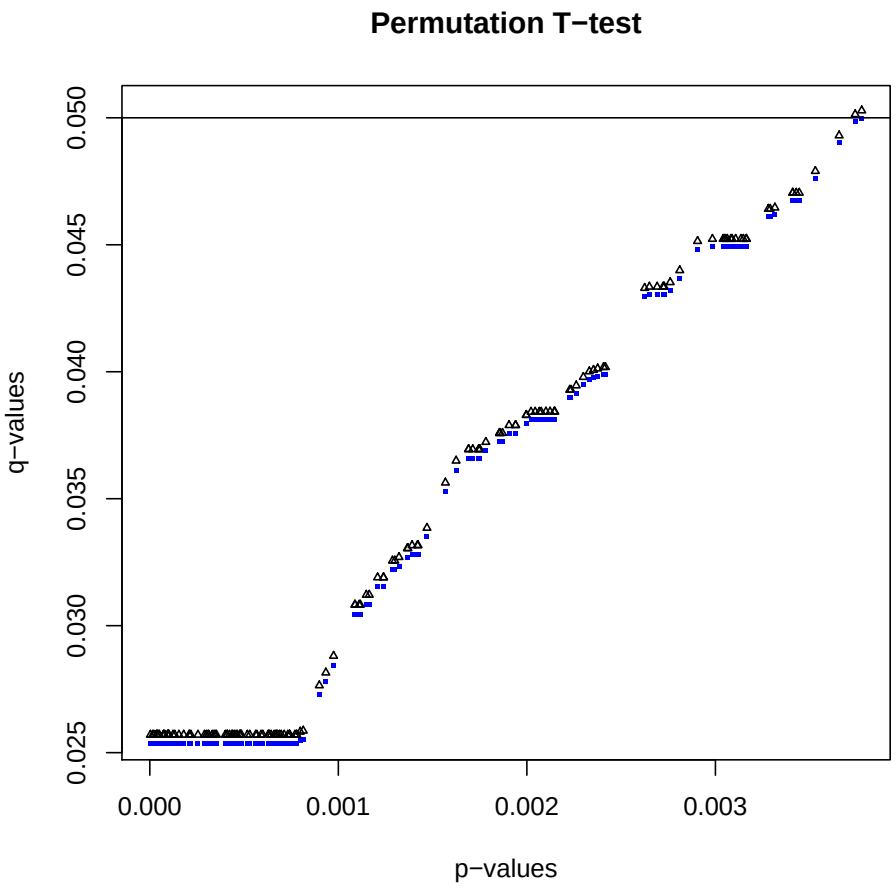

Figure 7.1: Q-value of Permutation T test to compare BRCA1 and BRCA2. Legend: Storey's method — triangles; New q-value method — dots .

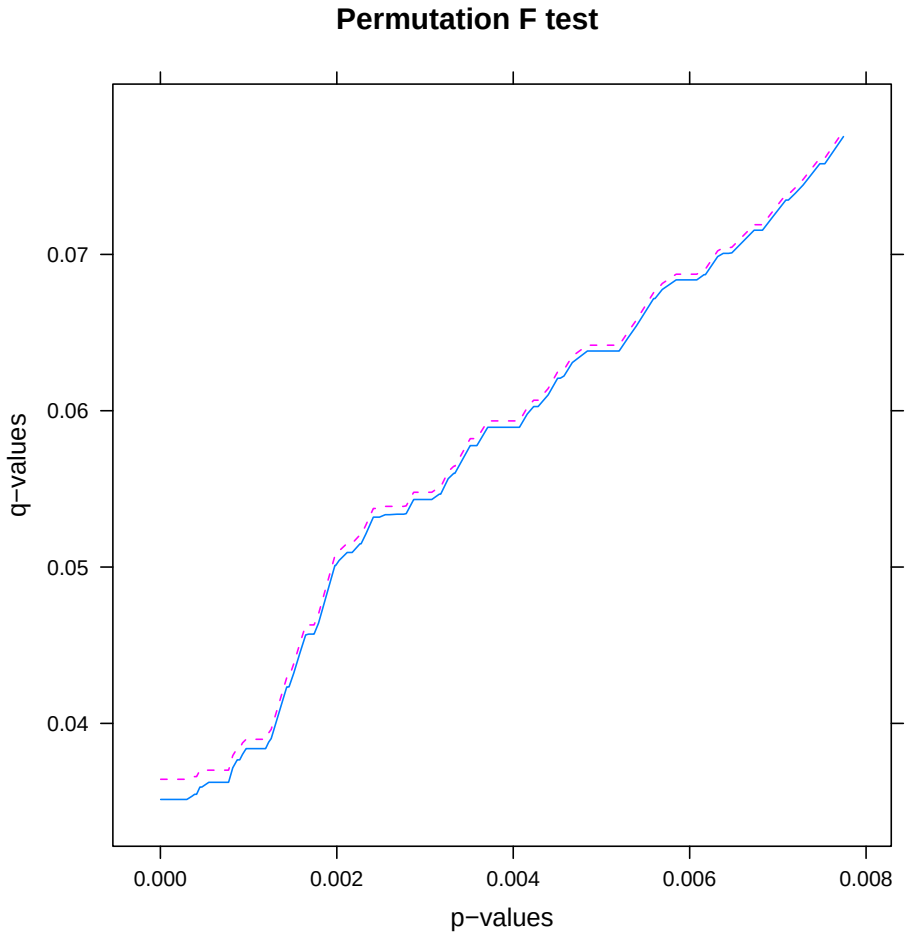

Figure 7.2: Q-value of Permutation F test to compare the three groups. Legend: Storey's method — dashed line; New q-value method — solid line.

CHAPTER 8

CONCLUSION AND FUTURE RESEARCH

8.1 Contribution of Dissertation

In this dissertation, we proposed two new FDR controlling methods, one based on the fixed error rate approach and the other estimation based.

- We constructed an alternative version of the BKY method by considering a different estimate of n_0. The new adaptive BH method competes well with the BKY method. Like the BKY method, it controls the FDR with independent p-values and, as can be seen numerically, continues to remain the control with the same type of positively dependent p-values as in the BKY method. More importantly, it can perform better than the BKY method in some instances, especially when the proportion of true null hypotheses is very large, as happens in many applications.

- We developed a new class of point estimates of the FDR, which has

uniformly smaller conservative bias than what Storey (2002) considered under the same setup while introducing his estimation based approach to controlling the FDR. The corresponding class of estimates of the pFDR provides a slightly different version of the q-value method than the original one that Storey proposed. This newer q-value mthod performs better than the original one. Also, these FDR estimates suggest a BH type method controlling the FDR in commonly encountered situations where the proportion of true null hypotheses can reasonably be assumed to follow a uniform distribution over the $(0, 1)$ interval.

8.2 Future Work

Some ideas of future work that could possibly fine-tune the work we have done in this thesis are described in the following.

First, the BH method is the only method which is proved to control the FDR under certain type of dependence so far. Though a number of simulation studies showed that both our new adaptive method and the BKY method can control the FDR under certain dependence structure, a theoretical proof of it is not yet available in the literature. It will be a great achievement if a proof can be offered.

Second, Table 5.1 shows that the ratios of power of both the BKY method and the new adaptive method to the 'Oracle' method. The 'oracle' method is the BH method based on the critical values $\alpha_i = i\alpha/n_0$, which should controls the FDR exactly at α under the independence of the p-values. So if n_0 were known, the 'oracle' method can be used as a benchmark against which an

FDR controlling method should be judged. As we can see from Table 5.1, the smallest ratio of power is 6.8%. Therefore, there is still much room to improve the current adaptive methods so that more power can be achieved.

REFERENCES

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological) 57*, 289-300.

Benjamini, Y. and Y. Hochberg (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of educational and behavioral statistics 25*, 60-83.

Benjamini, Y., A. Krieger, and D. Yekutieli (2006): Adaptive linear step-up procedures that control the false discovery rate. *Biometrica 93*, 491-507.

Benjamini, Y. and D. Yekutieli (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics 29*, 1165-1188.

Black, M. A. (2004). A note on the adaptive control of false discovery rate. *Journal of the Royal Statistical Society, Series B (Methodological) 66*, 297-304.

Broberg, P. (2005). A comparative review of estimates of the proportion unchanged genes and the false discovery rate. *BMC Bioinformatics 6*, 199.

Dalmasso, A. C., P. Broet and T. Moreau (2005). A simple procedure for estimating the false discovery rate. *Bioinformatics 21*, 660-668.

Dudoit, S., J. P. Shaffer and J. C. Boldrick (2003). Multiple hypothesis testing in microarray experiment. *Statistical Science 18(1)*, 71-103.

Finner, H. and M. Roters (2001). On the false discovery rate and expected type I errors. *Biometrical Journal 43*, 985-1005.

Finner, H., T. Dickhaus and M. Roters (2009). On the false discovery rate and an asymptotically optimal rejection curve. *Annals of Statistics 37*, 596-618.

Gavrilov, Y., Y. Benjamini and S. K. Sarkar(2009). An adaptive step-down procedure with proven FDR control. *Annals of Statistics 37*, 619-629.

Genovese, C. and L. Wasserman (2001). A large-sample approach to controlling false discovery rates. Technical report, Carnegie Mellon University.

Genovese, C. and L. Wasserman (2002). Operating characteristics and extensions of the false discovery rate procedure. *Journal of the Royal Statistical Society, Series 64*, 499-517.

Hedenfalk, I., Duggan, D., Chen, Y. D., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Kallioniemi and O. P., et al. (2001). Gene-expression profiles in hereditary breast cancer. *New England Journal of Medicine 344*, 539-548.

Hochberg, Y. and A.C.Tamhane (1987). *Multiple Comparison Procedure.* John Wiley & Sons.

Hommel, G. (1988). A stepwise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika 75*, 383-386.

Korn, E., J. Troendle, L. Mcshane and R. Simon (2004). Controlling the number of false discoveries: Application to high-dimensional genomic data. *Journal of Statistical Planning and Inference 124*, 379-398.

Langaas, M. and B. H. Lindqvist (2005). Estimating the proportion of true null hypotheses, with application to DNA microarray data. *Journal of the Royal Statistical Society, Series B (Methodological) 67*, 555-572.

Liu, F. and S. K. Sarkar (2009). A new adaptive method to control the false discovery rate. *Series in Biostatistics Vol. 4, World Scientific*, to appear.

Liu, F. and S. K. Sarkar (2010). A Note on Estimating the False Discovery Rate. *Journal of Statistical Planning and Inference 140*, 1601-1609.

Lu, X and D. L. Perkins (2007). Re-sampling strategy to improve the estimation of number of null hypotheses in FDR control under strong correlation structures. *BMC Bioinformatics 8*, 157.

Pounds, S. and S. W. Morris (2003). Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values. *Bioinformatics 19(10)*, 1236-1242.

Pounds, S. and C. Cheng (2004). Improving false discovery rate estimation. *Bioinformatics 20(11)*, 1737-1745.

Pounds, S. and C. Cheng (2006). Robust estimation of the false discovery rate. *Bioinformatics 22*, 1979-1987.

Romano, J. P., A. M. Shaikh and M. Wolf(2008). Control of the false discovery rate under dependence using the bootstrap and subsampling. *Test 17*, 417-442.

Sarkar, S. K. (1998). Some probability inequalities for ordered MTP_2 random variables: a proof of the simes conjecture. *Annals of Statistics 26(2)*, 494-504.

Sarkar, S. K. (2002a). Recent advances in multiple testing. *Journal of Statistical Studies special volume in honor of Prof. M.M. Ali's 65th birthday*, 293-306.

Sarkar, S. K. (2002b). Some results on false discovery rate in stepwise multiple testing procedures. *Annals of Statistics 30*, 239-257.

Sarkar, S. K. (2004). FDR-controlling procedures and their false negatives rates. *Journal of Statistical Planning and Inference 125*, 119-139.

Sarkar, S. K. (2006). False discovery and false non-discovery rates in single-step multiple testing procedures. *Annals of Statistics 34*, 394-415.

Sarkar, S. K. (2007). Stepup procedures controlling generalized FWER and generalized FDR. *Annals of Statistics 35*, 2405-2420.

Sarkar, S. K. (2008a). On the Simes inequality and its generalization. *IMS Collections Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen 1*, 231-242.

Sarkar, S. K. (2008b). On methods controlling the False Discovery Rate (with discussions). *Sankhya 70*, 135-168.

Sarkar, S. K. and C.-K. Chang (1997). The simes method for multiple hypothesis testing with positively dependent test statistics. *Journal of the American Statistical Association 92(440)*, 1601-1608.

Schweder, R. and E. Spjotvoll (1982). Plots of p-values to evaluate many tests simultaneously. *Biometrika 69*, 493-502.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika 73*, 751-754.

Sommera, S., Cuia, Y., Brewerb, G. and Fuqua, S. A.W.(2005). The c-Yes 3-UTR contains adenine/uridine-rich elements that bind AUF1 and HuR involved in mRNA decay in breast cancer cells. *The Journal of Steroid Biochemistry and Molecular Biology 97*, 219-229

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B (Methodological) 64*, 479-498.

Storey, J. D. (2003). The positive false discovery rate: a bayesian interpretation and the q-value. *Annals of Statistics 31*, 2013-2035.

Storey, J. D., Taylor, J. and Siegmund, D. (2004). Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach. *Journal of the Royal Statistical Society, Series B (Methodological) 66*, 187-205.

Storey, J. D. and Tibshirani, R. (2003). Statistical Significance for genome-wide studies. *Proc. Nat. Acad. Sci 100*, 9440-9445.

Tamhane, A. C., W. Liu and C. W. Dunnett (1998). A generalized step-up-down multiple test procedure. *Canadian Journal of Statistics 26*, 353-363.

Vumbaca, F., Phoenix, K. N., Rodriguez-Pinto, D., Han, D. K. and Claffey, K. P. (2007). Double-Stranded RNA-Binding Protein Regulates Vascular Endothelial Growth Factor mRNA Sta-

bility, Translation, and Breast Cancer Angiogenesis. *Molecular and Cellular Biology, 28*, 772-783.