INCORPORATING CORRELATIONS TO IMPROVE MULTIPLE TESTING PROCEDURES CONTROLLING FALSE DISCOVERIES

A Dissertation
Submitted to
the Temple University Graduate Board

---

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

---

by
Li He
May, 2011

Commitee Members

Dr. Sanat K. Sarkar (Chair),

Dr. Yuexiao Dong, Dr. Zhigen Zhao

Commitee Members

Dr. Sanat K. Sarkar (Chair),

Dr. Yuexiao Dong, Dr. Zhigen Zhao

ABSTRACT

INCORPORATING CORRELATIONS TO IMPROVE MULTIPLE TESTING PROCEDURES CONTROLLING FALSE DISCOVERIES

Li He

DOCTOR OF PHILOSOPHY

Temple University, May, 2011

Commitee Members

Dr. Sanat K. Sarkar (Chair),

Dr. Yuexiao Dong, Dr. Zhigen Zhao

Professor Sanat K. Sarkar

, Chair

Multiple testing is playing an important role in analyzing data from modern scientific investigations. Some fundamentally important theoretical and methodological issues related to multiple testing still remain to be fully investigated. Often the correlation structure among test statistics involved in multiple testing is known a priori or it can be estimated from the data, yet this structure is not often

properly taken into consideration while developing multiple testing procedures, even though not doing so might result in a less powerful method than one would like to have or lead to irrelevant or misleading conclusions. This dissertation focuses on research related to improving some of the commonly used multiple testing procedures by incorporating correlations into them. We propose several new results in this dissertation and present some ideas to carry out further research.

ACKNOWLEDGEMENTS

First of all, I want to thank my advisor Dr. Sanat K. Sarkar. He introduced me to the area of multiple testing, and he guided and directed me during the course of this research. He gave me many insightful ideas and thoughtful suggestions. His comments were always to the point and gave essential ideas. Certainly, his dedication to research and his scholarly spirit will be an inspiration to me throughout my career.

I want to thank my dissertation committee members, Dr. Zhigen Zhao and Dr. Yuexiao Dong, who have given me constructive ideas and helpful suggestions in the later stages of my research.

Also, I want to thank my advisor and Dr. Boris Iglewicz for offering me the Merck fellowship which has supported me for three years of the research. I also want to thank all the professors in the Department of Statistics at Temple University who taught me statistics and who gave me kind help. Particularly, I want to thank Dr. Heiberger, who always took time to help solve any of my programming questions.

Finally, I want to thank my family who have given me endless support, understanding and encouragement during my studies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Multiple hypothesis testing has become an important statistical tool to analyze massive data sets arising in many modern scientific investigations. Unlike in testing of a single hypothesis where the type I error rate, the probability of rejecting the null hypothesis when it is true, is being controlled, a multiple testing procedure guards against an error rate that provides an overall measure of type I errors. This overall measure is not unique. The conventional one is the familywise error rate (FWER), which is the probability of making at least one type I error or making at least one false discovery, but it suffers from being often too conservative, leading to very few rejections, particularly when a large number of hypotheses are tested. Accordingly, alternative, more liberal error rates have been considered in the literature, notable among which are the k-FWER, the probability of making k or more false discoveries, the γ-FDP, the probability of the false discovery proportion (FDP) exceeding γ, and the FDR (false discovery rate), the expected FDP, with FDR being the one that has received the most attention.

Observations in large scale data sets are rarely independent. For example, in microarray analysis, genes may be correlated along some biological pathways; in clinical trials, the test statistics for different endpoints are often correlated. Despite this fact, most of the multiple testing procedures controlling the above and other measures of false discoveries were developed under the presumption that the underlying test statistics are independent, since such cases are mathematically easier to handle. The validity under certain positive dependence situations were then established for some of these procedures. These include the commonly used Sidak and Hochberg (Hochberg, 1998) procedures to control the FWER and the BH (Benjamini and Hochberg, 1995) procedure to control the FDR. Some multiple testing procedures have also been developed without any regard to the dependence structure of the test statistics, such as the well-known Bonferroni procedure to control the FWER and the BY (Benjamini and Yekutieli, 2001) procedure to control the FDR.

The correlation has a major effect on a multiple testing procedure, as noted in a number of papers. Specifically, correlation can substantially increase the variability of the number of the false discoveries and hence can make a procedure controlling false discoveries unreliable (Owen, 2005; Qiu et al, 2005; Schwartzman and Lin, 2009). While studying the correlation effect in large scale hypothesis testing and suggesting a way of properly addressing it when controlling false discoveries, Efron (2007) pointed out that although individually null distribution of the test statistics (which he calls 'theoretical null') is known, the presence of correlation can significantly narrow or widen the distribution of null test statistics ('empirical null"). Hence, the inference made based entirely on the 'theoretical null,' which is generally done, may be inaccurate and misleading.

This dissertation is motivated by the fact that often the correlation structure among the test statistics is known a priori or it can be estimated from the data, and in such cases, procedures like those mentioned above that do not directly utilize this dependence structure but are still valid (as shown, for example, in Benjamini and Yekutieli, 2001; Sarkar and Chang, 1997; Sarkar, 1998, 2002) can potentially be improved by suitably incorporating the correlations. This idea of incorporating correlations into a multiple testing procedure controlling false discoveries, of course, has been used in some papers (Cai and Sarkar, 2005; Sarkar, 2007; Sarkar, 2008; and Sun and Cai, 2009), but it is yet to be fully developed.

Procedures controlling the FWER have been studied extensively; however, the literature on procedures controlling generalized FWER's, the k-FWER and γ-FDP, is relatively much less developed, even though they were originally proposed with a view to improve the power of an FWER procedure. This motivates us to do more research on these generalized procedures in this dissertation. More specifically, the procedures controlling the k-FWER and γ-FDP that were devised so far based solely on the marginal distributions of the test statistics are further improved by suitably incorporating correlations. Some of these results will be of use in constructing improved versions of the adaptive FDR and FWER methods as discussed below.

The actual FDR of the BH method at level α is equal to $\pi_0\alpha$, under independence, and is less than or equal to $\pi_0\alpha$, under a certain type of positive dependence among the test statistics [see, for example, Banjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001; Sarkar, 2002). Thus, this method becomes more conservative than what one would hope for because of two reasons: (i) the π_0 is smaller than 1, and (ii) the test statistics are positively dependent. Hence, an improvement of the original BH method may be achieved from two different directions, adjusting it for the level of control and adjusting it for dependence.

To adjust for the level of control, the BH method has so far been adapted to the data through an estimate of π_0. Typically, there are two stages involved in such an adaptation. At the first stage, an estimator of π_0 is formed (often based on the results of some multiple testing procedure), and at the second stage, the actual procedure is adjusted using the estimator obtained at the first stage. Various adaptive BH procedures of this type have been proposed in the literature (see, for example, Benjamini et al, 2006; Storey et al, 2004; Sarkar, 2008b), with proofs that they control the FDR under independence of the test statistics. Blankard and Roquain (2009) have recently proposed adaptive BH procedures that control the FDR under positive dependence. These seem to be the first set of adaptive BH procedures proposed in the literature controlling the FDR even under positive dependence, as in the case of the original BH method. However, this control of FDR is achieved under positive dependence at the cost of making the procedures more conservative, and there is a scope of reducing this conservativeness and thus improving the power of the procedures by incorporating correlations.

So, we continue the line of research in Blankard and Roquain (2009) toward improving the BH procedure not only by adjusting it using an estimate of π_0, as done in that paper, but also by incorporating the correlations into it when they are known. We design some new adaptive BH procedures with improved power by making use of correlations when forming an estimator of π_0 at the first stage, and prove that they control the FDR under positive dependence. More specifically, instead of using the FWER as Blankard and Roquain (2009) did, we consider some generalized error rates that bring correlations into the procedures controlling them and estimate π_0 at the first stage based on these procedures. These error rates are the k-FWER and k-PCER. The k-PCER is a new form of generalized per-comparison error rate (PCER) we consider here for the first time. From our simulation studies, we see that these newer versions of adaptive BH procedures improve those proposed in Blankard and Roquain (2009) when positive correlations are present. They also improve the original BH procedure when the signals corresponding to the false nulls are high and the dependence among the test statistics is mild.

The idea of improving the FDR control of the BH method by its adaptive versions has recently been applied to improve the Bonferroni and Sidak methods for controlling the FWER (see Guo, 2009; Sarkar et al, 2010). However, these adaptive FWER procedures have been obtained only through estimating π_0. So, we consider improving them further in this dissertation by incorporating correlations, following ideas similar to what we use for adaptive BH methods. We propose new adaptive Bonferroni and adaptive Sidak procedures that control the FWER with improved power by incorporating correlations and show that they ultimately control the FWER under positive dependence.

The compound decision theoretical framework developed in Sun and Cai (2007, 2009) provides a way of incorporating the dependence structure into a multiple testing procedure. They considered the following measures of the type I and type II errors, that is, the mFDR (marginal false discovery rate), the ratio of the expected number of false rejections to the expected number of total rejections for type I, and the mFNR (marginal false nondiscovery rate), the ratio of the expected number of false nondiscoveries to the expected number of total acceptions for type II. As shown in Genovese and Wasserman (2002), when the number of hypotheses m is large, the mFDR is equivalent to the

FDR in the sense that mFDR $=$ FDR $+ o(\sqrt{m})$ under weak conditions. Sun and Cai (2007) assumed a mixture model where the test statistics are independent. Under this model, they made the connection between the compound decision problem of finding the rule that minimizes the expectation of some specific loss function, and the multiple testing problem of deriving a procedure that controls the mFDR. By using this connection, they derived an optimal procedure that minimizes the mFNR while controlling the mFDR. Under the same loss function, Sun and Cai (2009) extended their previous work in 2007 to a Hidden Markov model under which the test statistics are correlated. Specifically, they also derived a multiple testing procedure that minimizes the mFNR while controlling the mFDR. Importantly, this compound decision framework allows for the intrinsic incorporation of the dependence structure into test statistics on which their multiple testing procedure is based. Their procedure therefore accounts for the dependence structure and was shown to be more powerful in identifying the interesting nonnull hyptheses than the BH method.

The loss function considered in Sun and Cai (2007, 2009) essentially corresponds to the 0-1 losses. We propose to use a different loss function which takes into account the prior information on the signal strength. When such prior information is available, we believe that our loss function is more reasonable and can lead to more accurate decisions. Under the same mixture model as in Sun and Cai (2007), we connect the compound decision theoretical problem under this new loss function to the multiple testing problem of controlling a weighted version of the mFDR, with the set of weights indicating the signal strength. In this connection, we derive an optimal multiple testing procedure that controls the weighted mFDR and minimizes the corresponding weighted mFNR. We note that weighted multiple testing procedures have been proposed in the literature in order to incorporate prior information or to emphasize the differential importance of hypotheses. When it is done properly, incorporating weights is one way of improving the power of a multiple testing procedure (Holm, 1979; Benjamini and Hochberg, 1997; Genovese et al, 2006; and Tamhane and Liu, 2008). Both our simulation study and our real data analysis show that our new approach controlling the weighted mFDR tends to be more powerful than both the procedure derived in Sun and Cai (2007) and the BH procedure.

Ultimately, we aim to extend our result into the situation where test statistics are dependent. Specifically, we want to develop a compound decision theoretical framework under our new loss function and under a model that incorporates dependence. Under this framework, we wish to derive an optimal procedure that incorporate both the signal strength and the dependence information. In particular, our procedure would minimize the weighted mFNR subject to the constraint that the weighted mFDR is controlled. We hope this will improve the procedure derived in Sun and Cai (2009).

The dissertation contains four sets of new results, the first one is related to procedures controlling the generalized versions of the FWER, the second one is related to improved adaptive BH procedures incorporating correlations, the third one is related to improved adaptive Bonferroni and Sidak procedures, and the fourth one is related to the compound decision theoretical framework for multiple testing developed under the new loss function. These new results are reported in Chapters 6-8, respectively. Before presenting these results, we provide some background information in Chapter 2-5. Some preliminaries in terms of notations and concepts are provided in Chapter 2. Chapter 3 reviews existing procedures controlling the FWER and its generalizations, while Chapter 4 reviews the procedures controlling the FDR. In Chapter 5, we look at the effect of correlation on the multiple testing procedures. The dissertation concludes with some future research ideas presented in Chapter 9.

CHAPTER 2

PRELIMINARIES

The purpose of this chapter is to set the stage for the discussions in the succeeding chapters by introducing some background concepts and notations.

Hypothesis testing uses the observed data to make decisions about the unknown properties of the underlying data generating distributions. For any such decision, two types of errors are possible: to reject a true null hypothesis (a type I error or false positive), or to accept a false null hypothesis (a type II error or false negative). The testing procedure is based on some properly chosen test statistic. The rejection region is found based on the distribution of the test statistic so that it minimizes the type II error while controlling the type I error below some chosen significance level α.

The extension of single hypothesis testing to multiple hypothesis testing was motivated by the need in scientific investigations in order to answer multiple related questions which can be formulated as a family of hypotheses. No matter what decision rule is used, each hypothesis in the family will have the type I and type II errors attached to it. Since some conclusions are to be drawn for the family that is tested, some joint measure error control is needed to guarantee a certain level of the statistical validity. Accordingly, whereas we typically control the type I error rate for a single hypothesis test, multiple hypothesis testing involves guarding against a suitably defined joint error rate. As stated in Hochberg & Tamhane (1987), ”Statistical procedures that are designed to take into account and properly control for the multiplicity effect through some combined or joint measure of erroneous inferences are called multiple comparison procedures.”

Important aspects of multiple testing are analogous to those in the single hypothesis testing, as discussed below.

2.1 Notations

Given a probability space $\{\mathcal{X}, \mathfrak{X}, \mathcal{P}\}$, any testing problem involves making inferences on hypotheses about the probability measure $\mathcal{P}$, based on the data generated from it. Suppose we want to test a set of m hypotheses $\mathcal{H} = \{H_1, \ldots, H_m\}$. Typically, we know how to test each individual hypothesis H_i and to calculate the corresponding p-value P_i. We denote the set of true null hypotheses by $\mathcal{H}_0$ and the set of false null hypotheses by $\mathcal{H}_1 = \mathcal{H} \backslash \mathcal{H}_0$. Let $|\cdot|$ denote the cardinality of a set. Assume $|\mathcal{H}_0| = m_0$, so $|\mathcal{H}_1| = m - m_0 = m_1$. We let $\pi_0 = m_0/m$ denote the proportion of true null hypotheses, and $\pi_1 = 1 - \pi_0$ the proportion of false nulls. Further, we let $I = \{1, 2, \ldots, m\}$, I_0, and I_1 be the index sets, respectively, for $\mathcal{H}$, $\mathcal{H}_0$ and $\mathcal{H}_1$.

2.2 Joint Type I Error Rates

Table 2.1: **Classifications of m tested hypotheses**

Hypotheses	Claimed non-significant	Claimed significant	Total
Null	U_m	V_m	m_0
Nonnull	K_m	S_m	m_1
Total	A_m	R_m	m

The outcomes of any multiple testing procedure can be categorized as in Table 2.1, where R_m is the number of rejections, V_m is the number of type I errors and S_m is the number of true positives; A_m is the number of acceptions, K_m is the number of type II errors and U_m is the number of true negatives. We note R_m and A_m can be observed, while S_m, U_m, V_m and W_m implicitly depend on $\mathcal{P}$ through $\mathcal{H}_0$ and so are unobservable random variables. In multiple testing, we need to make sure that the number of Type I error is controlled in some sense through a joint error rate. We now summarize some of these error rates into the following two categories. (For simplicity, we will suppress the subscript m subsequently.)

2.2.1 Error Rates Based on V

The following error rates involve the distribution of V, the number of type I errors.

1. The familywise error rate (FWER) is the probability of making at least one type I error,

$$\text{FWER} = P(V > 0)$$

2. k-FWER is the probability of making at least k type I errors, for some user-specified integer $k \geq 1$,

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

3. The per-family error rate (PFER) is the expected value of the number of Type I errors,

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

4. The per-comparison error rate (PCER) is the expected value of the proportion of Type I errors among the m tests,

$$\text{PCER} = \frac{1}{m} E(V)$$

Classical approaches to multiple hypothesis testing call for the control of FWER, as in the well-known Bonferroni procedure. On the other hand, the control of FWER at some level α requires each of the m individual tests to be conducted at a much lower level than α. Hence insisting on an FWER control will lead to very conservative results, when m is large. Procedures controlling the PCER are generally less conservative than those controlling FWER, since, for a given multiple testing procedure, PCER $\leq$ FWER. However, since controlling PCER at level α corresponds to performing each individual hypothesis at level α, PCER does not really address the multiplicity problems. In general, one has k-FWER$\leq$ FWER$\leq$ PFER. The first inequality follows immediately from the definitions of the first two error rates and the second one from the Markov inequality.

2.2.2 Error Rates Based on FDP

Define the false discovery proportion (FDP) to be the proportion of type I errors among the rejected hypotheses, i.e. V/R, with $V/R = 0$, when $R = 0$. The following error rates are based on the distribution of FDP.

1. The γ-FDP, for a user specified $\gamma \in (0, 1)$, is defined as

$$P(\text{FDP} > \gamma)$$

2. The false discovery rate (FDR) is the expected value of FDP.

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

3. The positive false discovery rate (pFDR) is one variation of FDR. Here the term 'positive' has been added to reflect the fact that we are conditioning on the event that at least one discovery has occurred.

$$p\text{FDR} = E\left(\frac{V}{R} \mid R > 0\right)$$

4. Marginal FDR (mFDR) is the ratio of the expected number of Type I errors to the expected number of rejected hypotheses.

$$m\text{FDR} = \frac{E(V)}{E(R)}$$

In general, error rates based on the FDP are more appealing when the number m of the tested hypotheses is large. The reason is that, FDP takes into account the fact that the number of false rejections usually increases with the number of rejections. When many of the tested hypotheses are rejected, it may be preferable to control the proportion of errors rather than the probability of making even one error. In particular, the error rate FDR has gained continuous attention in the multiple testing literature, since it was introduced by Benjamini and Hochberg in their 1995 seminal paper. We note two properties of FDR. First, if all the null hypotheses are true, then FDR reduces to FWER. Second, if some null hypotheses are not true, then FDR is less than FWER. On the other hand, controlling the expected FDP cannot guarantee that the FDP will be small with high probability. Then γ-FDP can be a desirable error rate to control. By the Markov inequality, $P(\text{FDP} > \gamma) \leq E(\text{FDP})/\gamma$. Hence if a procedure controls FDR at level α, it controls the γ-FDP at level α/γ.

While the FDR is an expected value of a ratio, the mFDR is a ratio of expected values and is therefore more tractable than the FDR. They are asymptotically equivalent when the number of tests m is sufficiently large since Genovese and Wasserman (2002) showed that under weak conditions, mFDR=FDR+$o(\sqrt{m})$. mFDR and pFDP are equivalent when test statistics come from a random mixture of null and nonnull distributions (See the definition 4.1.1 of the mixture model in section 4.1).

2.3 Power

To compare the performance of multiple testing procedures, one needs to clarify the concept of power. As with type I error rates, power can be conceptualized in various ways when extending single hypothesis testing to multiple hypotheses testing. The following three definitions are among those used in the literature.

1. Average power (See Storey, 2002; Cai and Sarkar, 2005), is the expected proportion of false null hypotheses that are correctly rejected.

$$\frac{1}{m_1}E(S)$$

2. The false nondiscovery rate (FNR) (See Genovese and Wasserman, 2002; Sarkar, 2005; Storey, 2003), is the expected proportion of the falsely accepted null hypotheses among those that are accepted.

$$\text{FNR} = E\left(\frac{K}{A} \mid A > 0\right) P(A > 0)$$

3. The marginal false nondiscovery rate (mFNR) is the proportion of the expected number of false nondiscoveries to the expected number of acceptions.

$$m\text{FNR} = \frac{E(K)}{E(A)}$$

4. 1-FDR-FNR, proposed by Sarkar (2004), reflects the strength of unbiasedness of an FDR procedure.

2.4 Global versus Multiple Testing

It is important to distinguish between a multiple testing problem and a global testing problem. The latter is a single hypothesis test whose null hypothesis is the intersection of all m individual null hypotheses. Hence a global test is designed to control the type I error. The former involves testing m

single hypotheses and guarding against some compound type I error. If the intersection hypothesis is rejected by a global test, one cannot further point at the individual hypotheses, deciding which are true and which are false. This task is achieved by answering the corresponding multiple testing problem.

We say that a multiple testing procedure controls the corresponding error rate weakly if it does so only when all the null hypotheses are true; and we say that a multiple testing procedure controls the corresponding error rate strongly if it does so under all configurations of true and false null hypotheses. Unless otherwise mentioned, in this work, we refer to the control in the strong sense.

2.5 Single-step versus Stepwise Multiple Testing Procedures

In terms of the p-values, $\boldsymbol{P} = \{P_1, \ldots, P_m\}$, a multiple testing procedure involves finding a threshold D which allows us to reject H_i if $P_i \leq D$ without violating the error rate control. Depending on the way this rejection threshold is determined, multiple testing procedures can be categorized into single step procedures and sequential (stepwise) procedures. A single step procedure involves finding a constant $d \in \{0, 1\}$ such that we reject any H_i with $P_i < d$. For stepwise procedures, we decide the threshold by comparing the ordered p-values with a set of nondecreasing critical constants $\{\alpha_1, \alpha_2, \ldots, \alpha_m\}$. Hence it is data-dependent (through p-values). Depending on whether we start the comparison with the smallest p-value $P_{(1)}$ or the largest p-value $P_{(m)}$, we categorizes stepwise procedures into stepdown and stepup procedures. Specifically, thresholds D_{SD} and D_{SU}, respectively for stepdown and stepup procedures, can be defined as follows:

$$D_{SD} = \sup\{P_i,\ i \in \{1, 2, \ldots, m\} :\ P_j \leq \alpha_j, \forall j \in \{1, 2, \ldots, i\}\}$$

$$D_{SU} = \sup\{P_i,\ i \in \{1, 2, \ldots, m\} : P_i \leq \alpha_i\}$$

It is clear that, based on the same set of critical values and the same set of p-values, we have $D_{SD} \leq D_{SU}$. Hence a stepup procedure is generally less conservative than its stepdown counterpart.

In general, in terms of any test statistics, in step-down procedures, the most significant null hypotheses are considered successively. As soon as one fails to reject a null hypothesis, no further hypotheses are rejected. In contrast, for stepup procedures, the least significant null hypothesis are tested successively. As soon as one null hypothesis is rejected, all remaining hypotheses are rejected.

2.6 Testing Principles

We now review general methods for the construction of the multiple testing procedures controlling FWER.

2.6.1 The Union-Intersection Principle

Roy (1953) proposed the Union-Intersection principle which is a method for constructing a global test for an intersection null, $H_0 = \cap_{i \in I} H_i$. Suppose that a test for each H_i is available. Then, according to the Union-Intersection principle, the rejection region for H_0 is the union of the rejection regions for each H_i. Hence the intersection null hypothesis is rejected if at least one H_i

is rejected. Roy and Bose (1953) showed that if a global test is based on the Union-Intersection principle and controls the type I error at level α, then this test is also a multiple test that controls the FWER strongly at level α. Various classical FWER controlling procedures are based on this principle, such as the Bonferroni test.

2.6.2 The Closed Testing Principle

Marcus, Peritz and Gabriel (1976) proposed the following method. Form the closure of the family of hypotheses $\{H_i, i \in I\}$ by including all nonempty intersection hypotheses $H_P = \cap_{i \in P} H_i$ for $P \subseteq I$. If an α level test is available for each H_P, then the closed testing procedure rejects any H_P if and only if every H_Q is rejected by its associated α level test for all $Q \supseteq P$. The procedures constructed using this method are called the closed testing procedures. It can be shown that any closed testing procedure controls FWER strongly at level α.

It is clear that the number of tests in a closed testing procedure increases exponentially with m. However, in order to make a rejection decision on any H_i, it is not necessary to test all H_P's containing H_i. In fact, it is possible to develop a shortcut version of the closed testing procedure. For example, from any closed testing procedure, a stepup or a stepdown test can be formed as a possibly conservative shortcut of it.

2.7 Positive dependence

A multiple testing procedure is based on an m-vector test statistic, and provides a rejection region for each hypothesis. Unlike the single hypothesis testing, where the distribution of the test statistic is known under the null hypothesis, in multiple testing, the joint distribution of the test statistics is often unknown, because we do not know the true configuration of the null and nonnull hypotheses nor the dependence structure of the test statistics. In particular, the m-vector test statistic is often a correlated vector. Consequently, in order to derive the rejection regions that offer the control of some suitably defined compound error, one often makes assumptions on the the dependence structure of the test statistics, for example, the independence assumption, or the certain positive dependence assumptions of the test statistics.

Various notions of positive dependence have been introduced into the literature, some of which we give a brief review in this section. Note that often these positive dependence restrictions are imposed only on the set of test statistics corresponding to the true null hypotheses (conveniently referred to as the null test statistics), not the set corresponding to the false null hypotheses (referred to as the nonnull test statistics).

Definition 2.7.1 *The random variables in $\boldsymbol{X} = (X_1, X_2, \ldots, X_m)'$ are said to be positively associated (A), if*

$$E(h_1(\boldsymbol{X})h_2(\boldsymbol{X})) \geq E(h_1(\boldsymbol{X}))E(h_2(\boldsymbol{X}))$$

or equivalently, if

$$cov(h_1(\boldsymbol{X})h_2(\boldsymbol{X})) \geq 0$$

holds for all functions h_1 *and* h_2 *for which the expectations exist. Here* h_1 *and* h_2 *are simultaneously monotonically increasing or decreasing in each of the coordinates.*

The concept of positive dependence was first introduced by Lehmann (1966) for the bivariate case. It was later generalized to the multivariate case by Sarkar (1969).

Definition 2.7.2 *A multivariate distribution is said to have the positive regression dependence property (PRD) if for any increasing set* D*,*

$$P(\boldsymbol{X} \in D \mid X_1 = x_1, \ldots, X_i = x_i)$$

is nondecreasing in $(x_1, \ldots, x_i)$*[Sarkar(1969)].*

Recall a set D is called increasing if $x \in D$ and $y \geq x$ imply that $y \in D$ as well.

Definition 2.7.3 *A random vector* $\boldsymbol{X} = (X_1, X_2, \ldots, X_m)'$ *or its density is said to be multivariate total positivity of order 2, conveniently denoted by* MTP_2*, if*

$$f(\boldsymbol{x}) \cdot f(\boldsymbol{y}) \leq f(\boldsymbol{x} \vee \boldsymbol{y}) f(\boldsymbol{x} \wedge \boldsymbol{y}), \forall(\boldsymbol{x}, \boldsymbol{y})$$

where

$$\boldsymbol{x} = (x_1, \ldots, x_m), \boldsymbol{y} = (y_1, \ldots, y_m)$$

$$\boldsymbol{x} \wedge \boldsymbol{y} = (min(x_1, y_1), min(x_2, y_2), \ldots, min(x_m, y_m))$$

$$\boldsymbol{x} \vee \boldsymbol{y} = (max(x_1, y_1), max(x_2, y_2), \ldots, max(x_k, y_k))$$

and f *is either the joint density or the joint probability function.*

Among the above three concepts of dependence, MTP_2 is the strongest. It implies the PRD property which further implies positive association. On the other hand, while it may be generally difficult to check that a set of random variables $\boldsymbol{X} = (X_1, X_2, \ldots, X_m)'$ are positively associated, the verification of the stronger property that $\boldsymbol{X}$ is MTP_2 is often easier. For example, the multivariate normal distribution with nonnegative correlation is MTP_2 when the off-diagonals of $-\Sigma^{-1}$ are nonnegative, where Σ is the covariance matrix. See Karlin and Rinott (1980), Sarkar and Chang (1997) and Sarkar (2008) for more details.

For clarity, we summarize the implication relations as the following:

$$MTP_2 \Longrightarrow PRD \Longrightarrow A$$

Benjamini and Yekutieli (2001) introduced the following positive dependence condition called regression dependence on the subset condition (PRDS).

Definition 2.7.4 *The random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_m)'$ is said to have PRDS property in the subset I_0, $I_0 \in \{1, 2, \ldots, m\}$, if, for any increasing set D and for each $i \in I_0$, $P(\boldsymbol{X} \in D | X_i = x_i)$ is nondecreasing in x_i.*

This is a relaxed form of the positive regression dependence property in the following sense. First, monotonicity is required after conditioning only on one variable at a time. Second, the conditioning is done only on any one from a subset of the variables. Hence the following implication relations are true:

$$MTP_2 \Longrightarrow PRD \Longrightarrow PRDS$$

The following positive dependence condition was one of the sufficient conditions for the the generalized Simes inequality in Sarkar (2008):

Condition 2.7.5 *For each $\{i_1, \ldots, i_k\} \subseteq I_0 \subseteq \{1, \ldots, m\}$, and for any nondecreasing (or nonincreasing) function ϕ,*

$$E(\phi(X_1^{-\{i_1,\ldots,i_k\}}, \ldots, X_{m-k}^{-\{i_1,\ldots,i_k\}}) \mid \max\{X_{i_1}, \ldots, X_{i_k}\} \leq x)$$

is nondecreasing (or nonincreasing) in x. ([Sarkar(2008)])

Here $X_i^{-\{i_1,\ldots,i_k\}}$, $i = 1, \ldots, m-k$, denote the components of the set $\{X_1, \ldots, X_m\} \backslash \{X_{i_1}, \ldots, X_{i_k}\}$.

It is easy to see that, at $k = 1$, this condition is implied by the PRDS condition. However, as pointed out in Sarkar (2008), for $k \geq 2$, it is more restrictive than the PRDS condition. Hence the generalized Simes inequlaity holds for a smaller class of distributions than the one for which the original Simes inequality holds. On the other hand, it has been shown that this condition is implied by a symmetric MTP_2 distribution. For example, multivariate normal distribution with common nonnegative correlation satisfies this condition. See Sarkar (2008) for more details.

We now state Condition 2.7.5 for $k = 1$ and $k = 2$ by themselves, for reference convenience in the succeeding chapters.

Condition 2.7.6 *For each $i \subseteq I_0 \subseteq \{1, \ldots, m\}$ and for any nondecreasing (or nonincreasing) function ϕ,*

$$E(\phi(X_1^{\{-i\}}, \ldots, X_{m-1}^{\{-i\}} \mid X_i \leq x)$$

is nondecreasing (or nonincreasing) in x.

Condition 2.7.7 *For each* $\{i, j\} \subseteq I_0 \subseteq \{1, \ldots, m\}$ *and for any nondecreasing (or nonincreasing) function* ϕ*,*

$$E(\phi(X_1^{-\{i,j\}}, \ldots, X_{m-2}^{-\{i,j\}} \mid \max\{X_i, X_j\} \leq x)$$

is nondecreasing (or nonincreasing) in x*.*

CHAPTER 3

PROCEDURES CONTROLLING THE FWER AND GENERALIZATIONS OF THE FWER

Many commonly used procedures for testing a family of hypotheses $H_1, \ldots, H_m$ are p-value based, where the hypotheses are ranked from the most significant to the least significant according to their ordered p-values, $P_{(1)}, P_{(2)}, \ldots, P_{(m)}$. Then a cut off value is chosen along this ranking so that some joint error rate is controlled.

Unless otherwise mentioned, in this work, a stepup or stepdown procedure with critical values $\alpha_1 \leq \ldots \leq \alpha_m$ refers to a usual stepup or stepdown test based on the ordered p-values $P_{(1)}, \ldots, P_{(m)}$.

The proofs of certain multiple testing procedure depend critically on the condition that the p-values are stochastically greater than the uniform random variable under the null hypotheses, i.e.

$$Pr(P_i \leq t) \leq t, \quad \text{for } i \in I_0$$

Here and subsequently, we assume this general assumption holds.

3.1 Global Testing Procedures

We first look at some well known global tests.

3.1.1 Bonferroni Global Test

The Bonferroni procedure rejects the intersection null hypothesis if

$$P_{(1)} < \alpha/m$$

The Bonferroni test is based on the union intersection principle and it is the consequence of the Bonferroni inequality.

3.1.2 Simes Test

Simes test is a global test that rejects the intersection null hypothesis if

$$\min_i \{\frac{m}{i} P_{(i)}\} \leq \alpha, \quad i = 1, 2, \ldots, m$$

The validity of the Simes test is a consequence of the Simes inequality which provides a bound for the joint distribution of the order statistics of the random variables satisfying the PRDS Condition in terms of their marginal distributions. In terms of the p-values, it has the following form:

$$Pr(P_{(i)} \geq i\alpha/m, i = 1, \ldots, m) \geq 1 - \alpha$$

It is clear that, while controlling the type I error rate, Simes test allows easier rejection of the intersection null hypothesis than the Bonferroni test, and hence it is more powerful. Initially, Simes (1986) proved that his method controls the type I error exactly at α when the test statistics are independent and conjectured that it is conservative when they are positively dependent. This conjecture was later established by Sarkar and Chang (1997) and Sarkar (1998) for the class of distributions characterized by the MTP_2 property. They also showed many commonly encountered distributions in multiple testing satisfy the MTP_2 property. In fact, the MTP_2 condition can be further relaxed to a slightly weaker PRDS condition (See Benjamini and Yekutieli, 2001; Sarkar, 2002; Cai and Sarkar, 2005). In other words, the Simes inequality is conservative in a larger class of PRDS distributions. The importance of the result by Sarkar and Chang (1997) and Sarkar (1998) lies in the fact that it provides a theoretical basis for extending the applicability of many multiple testing procedures based on Simes inequality from the independent to the positive dependent case.

3.1.3 Modification of the Simes Test

Since the Simes test is conservative under positive dependence, it is possible to improve it under such a condition. As an attempt to incorporate the dependence structure, Cai and Sarkar (2005) modified Simes' critical values assuming the underlying test statistics are multivariate normal with a common nonnegative correlation. They proved that the modified Simes test is more powerful than the original Simes test when the assumptions hold.

3.1.4 Generalization of the Simes Test

Sarkar (2007) and Sarkar (2008) obtained a generalized form of the Simes inequality. Using this generalization, he constructed procedures controlling the k-FWER. The generalized Simes inequality provides a bound for the joint distribution of the largest $m-k+1$ order statistics, $1 \leq k \leq m$, of m random variables which satisfy Condition 2.7.5. Denote the common distribution for the maximum of any k p-values by $F_k(\cdot)$. The generalized Simes inequality follows:

$$Pr(P_{k:m} \geq \alpha_k, \ldots, P_{m:m} \geq \alpha_m) \geq 1 - F_k(\alpha_m)$$

The equality holds under independence and when $\binom{i}{k}^{-1} F_k(\alpha_i)$ is constant for $i = k, \ldots, m$. Applying this inequality, the generalized Simes test, with the critical constants $\alpha_i, i = 1, 2, \ldots, m$ satisfying

$$F_k(\alpha_i) = \begin{cases} \frac{1}{\binom{m}{k}} \frac{k(k-1)\cdots 1}{m(m-1)\cdots(m-k+1)}\alpha & i \leq k \\ \frac{\binom{i}{k}}{\binom{m}{k}} \frac{i(i-1)\cdots(i-k+1)}{m(m-1)\cdots(m-k+1)}\alpha & i > k \end{cases},$$

controls k-FWER, under the intersection null hypothesis at α, exactly when P_i's are i.i.d., and conservatively when their joint distribution satisfies Condition 2.7.5.

3.1.5 Weighted Global Tests

To incorporate prior information or to emphasize the differential importance of hypotheses, weighted intersection testing procedures were proposed. There are different ways to incorporate weights in testing. (Interested readers can refer to Benjamini and Hochberg, 1997; Genovese and Wasserman, 2006; Tamhane and Liu, 2008).

Let w_i be the weights assigned to the hypothesis H_i, $i = 1, \ldots, m$, with $\sum_{i=1}^{m} w_i = m$. Let $P_i^* = P_i/w_i$ be the weighted p-values and $P_{(1)}^* \leq \ldots, \leq P_{(m)}^*$ be the ordered weighted p-values. Let $H_{(i)}^*$ and w_i^* be, respectively, the hypothesis and the weight corresponding to $P_{(i)}^*$.

The weighted Bonferroni procedure rejects the intersection null hypothesis if

$$P_{(1)}^* < \alpha/m$$

Hochberg & Liberman (1993) extended the Simes test to the weighted case. Their weighted Simes test rejects the intersection null hypothesis if,

$$\min_i \{\frac{m}{i} P_i^*\} \leq \alpha \quad i = 1, 2 \ldots, m$$

3.2 FWER Controlling Procedures

Now we turn to multiple testing procedures controlling the FWER.

3.2.1 Bonferroni Procedure

The classical Bonferroni procedure is a single step procedure that is easiest to implement. It rejects all H_i with $P_i \leq \alpha/m$, $i = 1, 2, \ldots, m$. This procedure controls FWER strongly at level α.

It is clear that the Bonferroni procedure is extremely conservative with m large. Accordingly, there was a line of research where improvements of Bonferroni procedure were made, such as the work by Holm(1979), Hommel(1988), Hochberg(1988) and Rom(1990), discussed below.

3.2.2 Holm's Procedure

Based on the closed testing principle, Holm (1979) derived a step-down procedure with critical constants $\alpha_i = \alpha/(m-i+1), i = 1, 2, \ldots, m$. It is a shortcut version of the closed testing procedure in which the Bonferroni procedure is used to test each intersection hypothesis.

Holm's procedure requires no distribution assumptions and is uniformly more powerful than the Bonferroni single-step procedure.

3.2.3 Hommel's Procedure

Hommel's procedure is a shortcut of the closed testing procedure in which the Simes test is used to test each intersection hypothesis. Hence it controls the FWER under the PRDS condition and it is at least as powerful as Holm's test whenever the condition holds. The procedure works as follows:

Define j by

$$j = \max_i \{i : P_{(m-i+k)} > k\alpha/i, \quad k = 1, 2, \ldots, i; \qquad i = 1, 2, \ldots, m\}$$

If the maximum does not exist, reject all hypotheses. Otherwise, reject all H_i with $P_i > \alpha/j$. Note that Hommel's procedure is neither a stepup nor a stepdown procedure.

3.2.4 Hochberg Procedure

The Hochberg procedure was also derived as a shortcut of the closed testing procedure based on the Simes global test, and hence, like Hommel's procedure, only controls FWER when the PRDS condition is satisfied. Furthermore, it is a conservative simplification of the Hommel's procedure. It is a stepup procedure with the same set of critical values as those in the Holm's procedure, namely, $\alpha_i = \alpha/(m - i + 1)$, $i = 1, \ldots, m$. Clearly it is more powerful than Holm's procedure.

3.2.5 Rom's Procedure

Rom (1990) pointed out that the Hochberg procedure does not exhaust the error probability α for each intersection test. Under the independence assumption, he enlarged the rejection region by solving the set of equations

$$Pr(P_{(1_n)} \geq c_{1_n}, \ldots, P_{(m_n)} \geq c_{m_n}) = 1 - \alpha$$

recursively from $n = 1, 2, \ldots, m$.

3.2.6 Weighted Procedures

Here we will follow our notations for the weighted global tests. Holm (1979) proposed the weighted Holm procedure, which rejects $H^*_{(i)}$ when

$$P^*_{(i)} \leq \frac{\alpha}{\sum_{k=i}^{m} w^*_{(k)}}$$

When all weigths equal 1, the weighted Holm procedure reduces to the original unweighted Holm's procedure.

Hochberg & Liu (2008) considered the closed tests based on the weighted Sime test and the stepup analogue of the weighted Holm's test.

3.3 Procedures Controlling Generalizations of FWER

We now review the various procedures controlling the k-FWER and γ-FDP.

3.3.1 Lehmann & Romano (2005) Procedures

Lehmann and Romano (2005) derived a Bonferrroni type single step procedure and a Holm type improvement of it that controls k-FWER without any restriction on the dependence structure. Specifically, their single step procedure rejects any H_i for which $P_i \leq k\alpha/m$, $i = 1, \ldots, m$. Their stepdown test has the critical values defined below:

$$\alpha_i = \begin{cases} \dfrac{k\alpha}{m} & i \leq k \\ \dfrac{k\alpha}{m+k-i} & i > k \end{cases} \tag{3.1}$$

They also introduced a stepdown procedure that control the γ-FDP under some restriction on the dependence structure. Specifically, their stepdown procedure has the critical values

$$\alpha_i = \frac{(\lfloor \gamma i \rfloor) + 1)\alpha}{m + \lfloor \gamma i \rfloor + 1 - i} \qquad i = 1, \ldots, m \tag{3.2}$$

where $\lfloor x \rfloor$ denotes the smallest integer that is less than or equal to x. This procedure is valid under either of the following two conditions on the dependence structure: (1). Conditioned on the observed non-null p-values, the null p-value is stochastically dominated by the uniform distribution; (2). Condition 2.7.6.

3.3.2 Romano & Shaike (2006) Procedures

Romano and Shaike (2006) proposed stepup procedures controlling generalizations of FWER starting from any increasing set of constants without any dependence assumptions. Specifically, given $\alpha_1 \leq \ldots \leq \alpha_m$, they formed an upper bound $D_1(k, m)$ for the k-FWER of a stepup procedure, and then showed that the stepup procedure, with the critical values defined as

$$\alpha' = \alpha\alpha_i / D_1(k, m)$$

controls the k-FWER at level α. Their procedure controlling the γ-FDP was derived in a very similar fashion.

3.3.3 Procedures Based on the kth Order Joint Distribution

The procedures controlling the generalized FWER mentioned previously have the common feature that they are all based on the marginal distribution of the p-values.

In contrast, Sarkar (2007, 2008a) utilized the k-th order joint distribution of the null p-values. Specifically, Sarkar (2008a) constructed a Bonferroni type single-step procedure and its Holm type stepdown improvement under the identical k-th order joint null distribution assumption, which we refer to, respectively, as the generalized Bonferroni procedure and the generalized Holm procedure. A stepup version of his generalized stepdown procedure, referred to as the generalized Hochberg procedure, was then proved to be valid under Condition 2.7.5. Specifically, the generalized Bonferroni procedure has threshold t satisfying $F_k(t) = \alpha/\binom{m+k-i}{k}$, and the generalized Holm and Hochberg procedure have the critical values $\alpha_i, i = 1, \ldots, m$ satisfying

$$F_k(\alpha_i) = \begin{cases} \alpha/\binom{m}{k} & i \leq k \\ \alpha/\binom{m+k-i}{k} & i > k \end{cases}$$

Sarkar (2007) generalized Romano and Shaike's stepup procedure. In fact, the critical values of his stepup procedure were derived in a similar manner to those in Romano and Shaikh (2006), except that the upper bound of the k-FWER for a stepup procedure with any give set of constants $\alpha_1, \ldots, \alpha_m$ is based on the common k-th order joint null distribution $F_k(\cdot)$, instead of the marginal distribution.

3.3.4 Augmentation Procedures

Dudoit & M. van der Lann (2008) discussed the augmentation multiple testing procedures controlling generalizations of FWER, which are procedures obtained by adding suitably chosen null hypotheses to the set of hypotheses already rejected by an initial FWER controlling procedure.

Suppose one has available a procedure controlling the FWER at level α which rejects a set of hypotheses $\mathcal{R}$. Let $\mathcal{V} \subseteq \mathcal{R}$ be the set of the falsely rejected true null hypotheses. Let $|\mathcal{V}| = V$ and $|\mathcal{R}| = R$. Then the augmentation k-FWER procedure rejects the set of the hypotheses

$$\mathcal{R}^+ = \mathcal{R} \cup \mathcal{A}$$

where

$$\mathcal{A} \subseteq \mathcal{H} \setminus \mathcal{R} \qquad \text{and} \qquad |\mathcal{A}| = \min\{k-1, m-R\}$$

The proof of the validity of the above augmentation procedure is straightforward, which we now briefly describe. Let V^+ denote the number of false rejections of the augmentation procedure. Since $P(|\mathcal{A}| \leq k) = 1$, $P(V^+ \geq k) \leq P(V + k - 1 \geq k)(= P(V \geq 1))$, which is less than or equal to α by the original FWER procedure.

In order to control γ-FDP, the augmentation set $\mathcal{A}$ needs to satisfy $\mathcal{A} \subseteq \mathcal{H} \setminus \mathcal{R}$ and $|\mathcal{A}| = \min\{\lfloor \frac{\gamma R}{1-\gamma} \rfloor, m - R\}$. The proof of its validity is similar to that for controlling k-FWER and it is again straightforward.

It can be seen that there is freedom in choosing the augmentation set $\mathcal{A}$ in both of the above procedures, as long as the corresponding conditions are met. However, in order to obtain more powerful procedures, better options are to choose the additional hypotheses in $\mathcal{A}$ to be the most significant hypotheses not rejected by the initial FWER controlling procedure, where the order of the significance can be based on the p-values or any other proper test statistics.

CHAPTER 4

PROCEDURES CONTROLLING THE FDR

4.1 Mixture Model and HMM Model

Various FDR procedures are derived under the following mixture model, which we state below for later reference.

Definition 4.1.1 **[Two-class Mixture Model]** *Let θ_i be the indicator that the hypothesis H_i is false, i.e. $\theta_i = 1$ if H_i is nonnull and $\theta_i = 0$ otherwise. Suppose that θ_i, $i = 1, 2, \ldots, m$, are independent Bernoulli random variables with $P(\theta_i = 0) = \pi_0$ and $P(\theta_i = 1) = \pi_1$. Conditioned on θ_i, X_i is distributed as*

$$X_i \mid \theta_i \sim (1 - \theta_i)F_0 + \theta_i F_1 \tag{4.1}$$

Then the marginal cumulative distribution function of X is the mixture distribution $F(x) = \pi_0 F_0(x) + \pi_1 F_1(x)$*, and the probability density function is* $f(x) = \pi_0 f_0(x) + \pi_1 f_1(x)$.

The following Hidden Markov Model (HMM) model was considered by Sun and Cai (2009) to model the dependence structure among test statistics.

Definition 4.1.2 [Hidden Markov Model (HMM)] *Let* θ_i *be the indicator that the hypothesis* H_i *is false, i.e.* $\theta_i = 1$ *if* H_i *is nonnull and* $\theta_i = 0$ *otherwise. Suppose that* $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ *form a markov chain. Conditioned on* θ_i*,* X_i *is distributed as in (4.1).*

4.2 p-value Based Approach

4.2.1 The BH and BY procedures

In pioneering work, Benjamini and Hochberg (1995) introduced the error rate FDR and provided a sequential p-value method to control it. Specifically, it is a stepup procedure with the critical constants the same as those in the Simes test, namely, $\alpha_i = i\alpha/m, i = 1, \ldots, m$. They proved that, when the test statistics are independent, the BH procedure controls the FDR exactly at the level $\pi_0\alpha$. This procedure is often referred to as the BH or linear step up procedure(LSU), since the set of the critical constants are linear.

Since its proposition, the FDR has gain continuous attention. The extension of the BH method to dependence was done by Benjamini and Yekutieli (2001). They proved that if the joint distribution of the test statistics is PRDS on the subset of null test statistics, the BH procedure controls the FDR at level less than or equal to $m_0\alpha/m$. For arbitrary dependence, they derived a simple conservative modification of the BH procedure, called the BY procedure, that controls the FDR. Specifically, when the BH procedure is conducted with $\alpha/(\sum_{i=1}^{m} 1/i)$ taking the place of α, it always controls the FDR at level less than or equal to $\pi_0\alpha$.

Essentially, the BH procedure provides a random threshold D and rejects all hypotheses for which $P_i \leq D$. Wasserman and Genevose (2002) studied the asymptotic behavior of the BH method by investigating the asymptotic properties of the "deciding point" D. In particular, they showed that, under certain framework, the deciding point D converges to a limiting quantity "μ^*" for which they gave a closed form expression. The limiting threshold μ^* satisfies $\alpha/m \leq \mu^* \leq \alpha$, which indicates that the BH procedure is intermediate between the Bonferroni procedure and the uncorrected test.

4.2.2 Modifications of the BH Procedure

When the joint distribution of text statistics can be specified, a specific set of critical constants for a stepup procedure can be found, which achieves the FDR control exactly at the specified distribution. Troendle (1996) took this approach and calculated a monotone sequence of constants for test statistics which are multivariate normal with equal, nonnegative correlations. His calculations were done under the unproven assertion that, when the nonzero means are set at infinity, the FDR is maximized.

Cai and Sarkar (2005) modified Simes critical values when the underlying test statistics follow a multivariate normal distribution with a common nonnegative correlation. They proved that the stepup test based on their modified Simes critical values controls the FDR and showed that this modified BH procedure is more powerful than the original BH procedure under the normal assumption.

Yekutieli and Benjamini (1999) took into account the correlation structure of the test statistics by utilizing the information in the sample. Their resampling based procedure is more powerful, however, at the expense of numerical complexity and only approximate FDR control.

4.2.3 Adaptive BH Procedures

Since the BH method controls the FDR at level $m_0\alpha/m$, the information about m_0 in the data is neglected.

To overcome this fault, Storey (2002) took a different approach to the control of the FDR. Instead of setting the error rate at a particular level and finding the rejection region, as done in the BH procedure, he advocated an estimation approach as follows. He fixed the rejection threshold t beforehand, then found a conservative estimator of the FDR at t, denoted by $\widehat{\text{FDR}}(t)$, and then decided the rejection threshold to be $\sup\{t : \widehat{\text{FDR}}(t) \leq \alpha\}$. The conservativeness of the FDR estimators offers the analogous property of strong FDR control. He argued that this estimation approach is conceptually more direct and simpler an approach. Specifically, under the mixture model 4.1 for p-values, he formed a well-behaved point estimator for FDR by formulating an estimator for π_0.

Essentially, the operational difference between the Storey's approach and the BH procdure is the inclusion of $\hat{\pi}_0$, an estimator of the proportion of true nulls. It can be shown that Storey's approach is equivalent to the BH approach when one takes the most conservative estimator of π_0, namely, $\hat{\pi}_0 = 1$.

Storey's idea of controlling the FDR through an estimate opens up the possibility of improving the BH procedures by combining an estimate of π_0 and producing stepup critical constants of the form $i\alpha/(\hat{\pi}_0 m), i = 1, \ldots, m$. These procedures are referred to as adaptive BH procedures.

Definition 4.2.1 (Adptive BH procedure) *Let $\boldsymbol{P} = (P_1, \ldots, P_m)$ be the vector of p-values corresponding to m hypotheses. A stepup procedure with the set of critical values $\{i\alpha G(\boldsymbol{P})/m,\ i = 1, \ldots, m\}$ is called an adaptive BH procedure, where $G(\boldsymbol{P}) = 1/\hat{\pi}_0$ and $\hat{\pi}_0$ is an estimator for π_0.*

If π_0 is a known quantity, then we have the following oracle procedure.

Theorem 4.2.1 (Oracle Adptive BH procedure) *When π_0 is known, a stepup procedure with the set of critical values $\{i\alpha/(m\pi_0),\ i = 1, \ldots, m\}$ satisfies:*

$$FDR \leq \alpha$$

when $P_1, \ldots, P_m$ satisfy the PRDS condition. In particular, when $P_1, \ldots, P_m$ are independent, the equality holds.

Various estimators $G(\boldsymbol{P})$ have been introduced. We list here three which are commonly used.

1. The estimator that is used in Storey (2002) follows

$$G_1(\boldsymbol{P}) = \frac{(1-\lambda)m}{n - R(\lambda)}. \tag{4.2}$$

Here $\lambda \in (0,1)$ is a tuning parameter and $R(\lambda) = \sum_{i=1}^{m} I(P_i \leq \lambda)$.

2. Storey et al (2004) proposed a slightly modified version of the estimator mentioned above:

$$G_2(\boldsymbol{P}) = \frac{(1-\lambda)m}{n - R(\lambda) + 1}. \tag{4.3}$$

Later, in Bejamini, Krieger and Yekutieli (2006), the adaptive BH procedure based on the estimator $G_2(\boldsymbol{P})$ was proved to control the FDR under independence.

3. Bejamini, Krieger and Yekutieli (2006) suggested the following estimator:

$$G_3(\boldsymbol{P}) = \frac{(1-\lambda)m}{n - R_0}, \tag{4.4}$$

where $\lambda = \dfrac{\alpha}{\alpha+1}$ and R_0 number of rejections by a BH procedure at level λ. This procedure was also proved to control the FDR under independence.

Theoretical proof of the validity of these adaptive BH procedures under dependence is generally difficult and not available. In practical applications, the FDR may not be controlled and consequent inferences can be misleading. An adaptive procedure that improves the BH procedure and also controls the FDR is thus meaningful and desirable. Blankard and Roquain (2009) proposed two adaptive procedures controlling the FDR under dependence. Specifically, we restate them in the theorems below. The results in Blankard and Roquain (2009) were stated for a general collection of multiple testing procedures, called self-consistent procedures, with respect to a threshold collection which has a standard form $\Delta(i) = \alpha\beta(i)m$. Depending on the form of the function $\beta(i)$, their procedures offer control under positive dependence or under arbitrary dependence. In this paper, we will focus on the adaptive BH procedure, i.e. we restrict to the case where $\beta(i) = i$, and to the case of stepup procedures. Note that stepup procedures are the optimal case out of the self-consistent class.

Theorem 4.2.2 (Blankard and Roquain I) *When $P_1, \ldots, P_m$ satisfy the PRDS condition, the stepup procedure with the set of critical values $\{i\alpha_1/(m - R_0),\ i = 1, 2, \ldots, m\}$, where R_0 is the number of rejections of a multiple testing procedure that controls the FWER at level α_0, satisfies:*

$$FDR \leq \alpha_0 + \alpha_1$$

Theorem 4.2.3 (Blankard and Roquain II) *When $P_1, \ldots, P_m$ satisfy the PRDS condition, the stepup procedure with the set of critical values $\left\{\frac{i\alpha_1 G_k(R_0/m)}{m}\ i = 1, 2, \ldots, m\right\}$, where R_0 is the number of rejections of the BH procedure at level α_0, satisfies:*

$$FDR \leq k\alpha_0 + \alpha_1$$

Here

$$G_k(x) = \begin{cases} 1 & x \leq 1/k \\ \dfrac{2/k}{1 - \sqrt{1 - 4(1-x)/k}} & x > 1/k \end{cases}$$

Like usual adaptive procedures, in the above two procedures, the first step forms an estimator of π_0 through some multiple testing procedure. The second step is the actual BH adaptive procedure. Unlike usual adaptive procedures, both steps consititute an allowable error probability. For example, suppose FDR is to be controlled at level $\alpha = \alpha_1 + \alpha_0$. The actual FDR procedure has critical constants $\{i\alpha_1 G(\boldsymbol{P})/m, i = 1, 2, \ldots, m\}$, where $\alpha_1 \leq \alpha$. Thus, even if $G(\boldsymbol{P}) \geq 1$, these adaptive procedures can reject less hypotheses than the BH procedure. Indeed, in their simulation study, there are limited situations where their procedures offer an improvement of the BH procedure.

4.3 Decision Theoretical Approach

As we have seen, a multiple testing procedure is usually designed to control a certain chosen joint Type I error rate at a prespecified level. Further, some researchers have taken the decision theoretical approach and developed optimal procedures that minimize a measure of Type II errors while controlling a corresponding measure of Type I error.

The first to take such an approach was Spjotvoll (1972), who derived an optimal procedure that maximizes the number of expected true positives $E(S)$ (or that minimizes the expected number of false nondiscoveries), for each fixed number of expected false positives $E(V)$. He calculated $E(V)$ assuming all the null hypotheses are true, and calculated $E(S)$ assuming all the alternative hypotheses are true. Under such calculations, the optimal test corresponds to performing a Neyman-Pearson test for each hypothesis. As noted by Storey (2007), such a calculation maybe problematic because it does not represent the underlying reality. The null hypothesis and the alternative hypothesis can not be simultaneously true. Furthermore, no connection can be made to FDR under such calculations.

Storey (2007) derived an optimal multiple testing procedure that maximizes $E(S)$ for each fixed $E(V)$ under the calculation based on the true configuration of the hypotheses. Hence this optimality goal is connected to mFDR since $m\text{FDR} = \dfrac{E(V)}{E(V) + E(S)}$. On the other hand, both Spjotvoll's and Storey's optimal procedures depend on some unknown functions that may not be estimable from the data.

Wasserman and Genevose (2002) introduced the notion of false nondiscovery rate (FNR) and constructed an optimal procedure which, under the mixture model 4.1, minimizes the expected

FNR subject to a bound on the expected FDR. Specifically, they considered the risk function of the form

$$L^1_\lambda = \text{FNR} + \lambda\text{FDR}$$

where the constant λ is taken to be user specified. One can view λ as the Lagrange multiplier. Then minimizing the risk corresponds to minimizing the expected FNR subject to a bound on the expected FDR. Their procedure is based on the p-values for individual tests.

Under the mixture model 4.1, Sun and Cai (2007) developed a compound decision theory framework for multiple testing problems and derived an optimal procedure that minimizes the mFNR subject to $m\text{FDR} \leq \alpha$. The compound decision problem was initially proposed by Robins (1951) which involves solving m component decision problems simultaneously by minimizing some risk function. Formally, consider there are m null hypotheses $H_1, H_2, \ldots, H_m$ where we observe $\mathbf{X} = (X_1, \ldots, X_m)$. Let $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_m\}$, with $\theta_i \in \{0, 1\}$, where $\theta_i = 1$ indicates H_i is false and $\theta_i = 0$ otherwise, and let $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)$, with $\delta_i \in \{0, 1\}$, where $\delta_i = 1$ corresponds to rejecting H_i and $\delta_i = 0$ otherwise. For each hypothesis H_i, we want to make an inference regarding the unknown parameter θ_i based on the observation x_i that is generated according to θ_i. If these m decision problems are solved simultaneously, i.e. making inferences about the unknown vector $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_m\}$, based on the observed vector $\boldsymbol{x} = \{x_1, \ldots, x_m\}$, then it is called a compound decision problem and the corresponding rule is called a compound rule. In contrast, if each of the m component problems is solved separately, the corresponding rule is called a simple rule. An important point of studying multiple testing problem in a compound decision framework is that the precision of the tests can be increased by pooling information from different samples.

Note that the inferential task of testing m null hypotheses $H_1, \ldots, H_m$ can be carried out either by a multiple testing procedure or by a compound decision rule, both of which produce an outcome of the form $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)$. The compound decision rule minimizes a specific risk function, while a multiple testing procedure protects some error rate from exceeding a nominal level. Sun and Cai (2007) made the connection between the two problems. Specifically, for the ith component problem, they considered the loss function

$$\lambda L_0^i + L_1^i = \lambda \delta_i (1 - \theta_i) + \theta_i (1 - \delta_i), \tag{4.5}$$

where λ is the relative weight given to the loss for a Type I error to the loss for a Type II error. Note that L_0^i and L_1^i are essentially 0-1 losses. The compound decision rule is then to minimize the expected loss function

$$L_\lambda^2(\boldsymbol{\theta}, \boldsymbol{\delta}) = \frac{1}{m} \sum_i \{\lambda(1 - \theta_i)\delta_i + \theta_i(1 - \delta_i)\} \tag{4.6}$$

They showed that, under the mixture model 4.1 and some mild condition, the optimal multiple testing procedure that minimizes the mFNR while controlling the mFDR is based on the same test statistics on which the compound decision rule is based. Hence, to find the optimal multiple test procedure, they start with solving the compound decision problem. Specifically, their procedure is based on the local fdr, which for H_i, is defined as the posterior probability $P(\theta_i = 0 \mid X_i = x_i) = \pi_0 f_0(x_i)/f(x_i)$. We note that local fdr was introduced by Efron, Tibshirani, Storey and Tusher (2001) and it has been widely used to interpret results for individual cases.

They derived an oracle procedure (assuming all the parameters in the mixture model 4.1 are known) by ranking the null hypotheses according to their local fdr's and then finding a cutoff along this ranking. This implies that we actually rank the relative significance of the hypotheses according to their likelihood ratios, instead of their individual p-values. This procedure was compared with the optimal p-value procedure in Genovese and Wasserman (2002). The comparison shows that the p-value procedure is dominated by this procedure. A data-driven adaptive procedure was then developed that asymptotically attains the performance of the oracle procedure. Hence by studying multiple testing problem in a compound decision framework, more information in the data was used and a more powerful procedure was derived.

Sun and Cai (2009) introduced the HMM model 4.1.2 under which test statistics are correlated, and they connected the compound decision problem under the loss function (4.6) and the multiple testing problem of controlling mFDR under the HMM model. Again, they showed that the optimal multiple testing procedure that minimizes the mFNR while controlling the mFDR is based on the same test statistics on which the compound decision rule is based. The optimal statistic, found by solving the corresponding compound decision problem, is $P\{\theta_i = 0|\boldsymbol{X} = \boldsymbol{x}\}$, where the corresponding posterior probability is conditioned on the whole observed sequence. This was referred to as "the local index of significance" (LIS$_i(\boldsymbol{x})$) for the ith hypothesis in their paper. The oracle multiple testing procedure (assuming all the parameters in the HMM model are known) was then derived by ranking the null hypotheses in terms of their LIS$_i(\boldsymbol{x}), i = 1, \ldots, m$ values and then finding a cutoff along this ranking. LIS pools the information from the whole sequence of observations and takes into account the dependence structure. Hence the compound decision framework allows the intrinsic incorporation of the dependence structure. A data-driven procedure that mimics the oracle procedure is developed by plugging in consistent estimates of the unknown HMM parameters and was proven to be asymptotically optimal in the sense that it attains both the mFDR and the mFNR levels of the oracle procedure asymptotically. According to their numerical studies, their procedure is more powerful in detecting interesting nonnull cases than the BH method which does not utilize the dependence structure. This shows that when the correlation structure is appropriately incorporated, the efficiency of the test can be much increased.

CHAPTER 5

EFFECT OF CORRELATIONS ON MULTIPLE TESTING PROCEDURES

The effect of correlation on multiple testing procedures has been investigated by many researchers. Efron(2007) showed that correlation can play a major role in the correct identification of the false and true null hypotheses. Instead of using p-values for the corresponding test statistics, he converted the test statistics to z-values which theoretically and individually have a $N(0,1)$ distribution under the null hypothesis, called “theoretical null”. He studied the effect of correlation on multiple testing by studying the accurate estimation of the null distribution of the z-value under dependence. The basic idea of his paper is that although null cases individually follow the theoretical null distribution, $Z_i \sim N(0,1)$, the correlation effect can make the ensemble null distribution behave more like $N(0,\sigma^2)$, with σ different from 1. This means correlation can considerably widen or narrow the distribution of the null z-values, and unconditional use of the theoretical null can result in procedures that are too liberal or too conservative. Hence the significance levels should be judged according to the "empirical null" distribution, a normal curve fit to the central portion of the z-value histogram, instead of the "theoretical null".

Owen (2005) studied the effect of the correlation on the variance of the number of false discoveries V. He showed that correlation among the test statistics greatly inflates the variance of V. He further presented a variance formula that takes into account correlations between test statistics.

Qiu et al (2005) showed that the correlation effect can substantially deteriorate the performance of many FDR procedures. Schwartzman and Lin (2009) quantified the effect of correlation by providing explicit expressions for estimating the mean, variance, distribution and quantiles of the FDR estimators based on a negative binomial model. They showed that correlation increases both the bias and variance of the estimators substantially. In some cases, such as exchangeable distribution, correlation can cause the corresponding FDR estimators to be inconsistent.

These research further emphasizes the necessity to incorporate dependence into multiple testing procedures. In fact, the statistical methodology for multiple testing tends to increasingly concentrate on this area.

Indeed, as have been shown in the previous chapters when we review the various multiple testing procedures, when dependence is properly modeled and incorporated, powerful procedures can be derived. See Cai and Sarkar (2005), Sarkar (2007), Sarkar(2008), Sun and Cai (2009), Westfall and Young (1993), Dudoit and Van der Lann (2008), and so on.

Starting from the next chapter, we will present our procedures developed to incorporate correlations into the multiple testing procedures.

CHAPTER 6

NEW PROCEDURES BASED ON THE BIVARIATE DISTRIBUTIONS

In this chapter, we propose procedures controlling generalizations of the FWER that take correlation into account through the pairwise distribution of the null p-values. Some of these procedures are then used later to form an estimator of π_0 in our adaptive procedures.

Before we state our results, we define a new error rate which is a generalized form of the per-comparison error rate (PCER), called k-PCER, for some $k \geq 1$. It is defined as below

$$k\text{-PCER} = E\left[\frac{V}{m} I(V \geq k)\right]$$

We will develop a procedure that controls k-PCER and this procedure will be used later to help form an estimate of π_0 when we develop new adaptive procedures in chapter 7.

6.1 Controlling the k-FWER

We emphasize that our k-FWER procedures are derived for $k \geq 2$. To avoid triviality, we assume $m_0 \geq 2$, otherwise k-FWER is identically zero.

The following lemma is an extension of the Markov Inequality.

Lemma 6.1.1 *For a random variable X, with $X > 1$ and some constant $k > 1$, the following inequality holds:*

$$P(X \geq k) \leq \frac{E\left[X(X-1)\right]}{k(k-1)}$$

Since $X(X-1)$ is nondecreasing when $X > 1$, the Lemma is true. By using the Lemma, we are able to control the tail probability of the false positives through a bound based on the bivariate distribution.

We now state our single step procedure.

Theorem 6.1.1 (The Single Step Procedure) *Suppose the null p-values $\{P_i,\ i \in I_0\}$ have the identical pairwise joint distribution, denoted as $F_2(\cdot,\cdot)$. Reject all H_i with $P_i \leq t$, where t satisfies $F_2(t,t) = \dfrac{k(k-1)}{m(m-1)}\alpha$. This procedure controls the k-FWER at level α.*

Here, and subsequently, we use $P_i \vee P_j$ to denote $\max(P_i, P_j)$.

Note that, for fixed m and α, t will be a function of both k and the correlation ρ between any pair of null p-values. This contrasts with the critical constant $t = k\alpha/m$ of the single step procedure by Lehmann and Romano (2005) which is only a function of k. Table 6.1 compares

Table 6.1: **Comparison of Thresholds of Single Step Procedures**

k	Lehmann & Romano	The proposed method					
		$\rho = 0$	$\rho = 0.1$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$	$\rho = 0.9$
2	0.0010	0.0032	0.0020	0.0008	0.0003	0.0001	0.0000
6	0.0030	0.0123	0.0088	0.0044	0.0021	0.0010	0.0004
10	0.0050	0.0213	0.0160	0.0088	0.0047	0.0024	0.0010
14	0.0070	0.0303	0.0234	0.0137	0.0078	0.0042	0.0020
18	0.0090	0.0393	0.0311	0.0191	0.0114	0.0064	0.0032
22	0.0110	0.0483	0.0389	0.0248	0.0153	0.0090	0.0046
26	0.0130	0.0573	0.0468	0.0307	0.0196	0.0118	0.0063
30	0.0150	0.0663	0.0548	0.0369	0.0241	0.0150	0.0083

our single step threshold with $k\alpha/m$ under different combinations of k and correlation ρ. These thresholds are calculated for the case where $m = 100$ and the test statistics follow a multivariate normal distribution with a common correlation ρ. We can see that under independence and mild dependence, our threshold is larger, sometimes substantially. For example, when $k = 30$ and $\rho = 0.1$, our threshold is about 0.0548, almost 4 times the threshold 0.015 of Lehamnn and Romano. In particular, under independence, our threshold $\sqrt{\frac{k(k-1)}{m(m-1)}}\alpha$ will be greater than $k\alpha/m$ as long as $\alpha \leq \frac{(k-1)/k}{(m-1)/m}$. This will normally be satisfied by any reasonably small α. As can be seen from Table 6.1, as k increases, the magnitude of the gain increases. However, as correlation gets high, our threshold becomes smaller.

We now propose a Holm type improvement of the proposed single step procedure.

Theorem 6.1.2 (The Stepdown Procedure) *Suppose the null p-values $\{P_i,\ i \in I_0\}$ have identical pairwise joint distribution, denoted as $F_2(\cdot,\cdot)$. Then the stepdown procedure, with critical values $\{\alpha_i,\ i = 1, 2, \ldots, m\}$ satisfing*

$$F_2(\alpha_i, \alpha_i) = \begin{cases} \dfrac{k(k-1)}{m(m-1)}\alpha, & i \le k \\ \dfrac{k(k-1)}{(m+k-i)(m+k-i-1)}\alpha, & i > k \end{cases} \tag{6.1}$$

controls the k-FWER.

Remark 6.1.1 *Similar to the single step case, compared to the critical values of Lehmann and Romano given in (3.1), our critical values will be higher under independence and mild positive dependence, and smaller under high dependence. In particular, under independence, the critical constants will be*

$$\alpha_i = \begin{cases} \sqrt{\dfrac{k(k-1)}{m(m-1)}\alpha}, & i \le k \\ \sqrt{\dfrac{k(k-1)}{(m-i+k)(m-i+k-1)}\alpha}, & i > k \end{cases}$$

which is greater than $\dfrac{k}{(m-i+k)}\alpha$, $i = k, \ldots, n$, as long as $\alpha \le \dfrac{(k-1)/k}{(m-i+k-1)/(m-i+k)}$. This is satisfied for any reasonably small α.

We next show that the step-up analogue of the proposed stepdown procedure also controls the k-FWER, when Condition 2.7.7 and the identical pairwise distribution condition are satisfied by the null p-values. Note that independence is a special case of the Condition 2.7.7.

Theorem 6.1.3 (The Stepup Procedure) *If $\{p_i, i \in I_0\}$ satisfies Condition 2.7.7, then the stepup procedure with the critical constants of the stepdown procedure, given in 6.1, controls the k-FWER at level α.*

6.2 Controlling the γ-FDP

In this section, we present results improving some previous work on controlling the γ-FDP under both positive dependence and arbitrary dependence conditions on the null p-values (Lehmann and Romano (2005) and Romano and Shaikh (2006)). We also derive procedures that are based on the bivariate distribution of the $p$values.

We first have the following result under Condition 2.7.6:

Theorem 6.2.1 *The stepup or stepdown procedure with the critical constants*

$$\alpha_i = \frac{(\lfloor \gamma i \rfloor + 1)\alpha}{n + \lfloor \gamma i \rfloor + 1 - i}, \ i = 1, \ldots, n, \tag{6.2}$$

controls the γ-FDP at α under Condition 2.7.6.

Proof. Let $g(R) = \lfloor \gamma R \rfloor + 1$. Then, first note that

$$\begin{aligned}
\{V \geq g(R)\} &= \bigcup_{v=1}^{m_0} \left\{ \hat{P}_{(v)} \leq \alpha_R,\ g(R) \leq v,\ V = v \right\} \\
&= \bigcup_{v=1}^{m_0} \left\{ \hat{P}_{(v)} \leq \frac{g(R)\alpha}{m - R + g(R)},\ g(R) \leq v,\ V = v \right\} \\
&\subseteq \bigcup_{v=1}^{m_0} \left\{ \hat{P}_{(v)} \leq \frac{v\alpha}{m - R + v},\ V = v \right\} \\
&\subseteq \bigcup_{v=1}^{m_0} \left\{ \hat{P}_{(v)} \leq \frac{v\alpha}{m_0},\ V = v \right\} \\
&\subseteq \bigcup_{v=1}^{m_0} \left\{ \hat{P}_{(v)} \leq \frac{v\alpha}{m_0} \right\}. \qquad (6.3)
\end{aligned}$$

The probability of the event in the right-hand side of (6.3) is known to be less than or equal to α under Condition 2.7.6 from the so-called Simes' inequality (Simes (1986), Sarkar and Chang (1997), and Sarkar (1998)). Thus, we get the desired result noting that γ-FDP $= \Pr(V \geq g(R))$. ∎

Remark 6.2.1 Lehmann and Romano (2005) proposed only the stepdown procedure considered in Theorem 3.1 under the same assumption. Thus, this theorem provides an improvement of the Lehamann-Romano result, since we now have an alternative procedure under the same assumption, the step-up one, which is theoretically known to be more powerful. Moreover, we provide a simpler, unified proof of the γ-FDP control covering both our and the Lehmann-Romano original stepdown procedures. Our simulation studies indicate that this power improvement can often be quite significant (see Figure 6.1).

Besides the Condition 2.7.6, Lehmann and Romano also showed that their stepdown procedure is valid under the following dependence assumption. Where, we let $\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_{m_0}$ denote the null p-values and $\hat{Q}_1, \hat{Q}_2, \ldots, \hat{Q}_{m_1}$ denote the nonnull p-values.

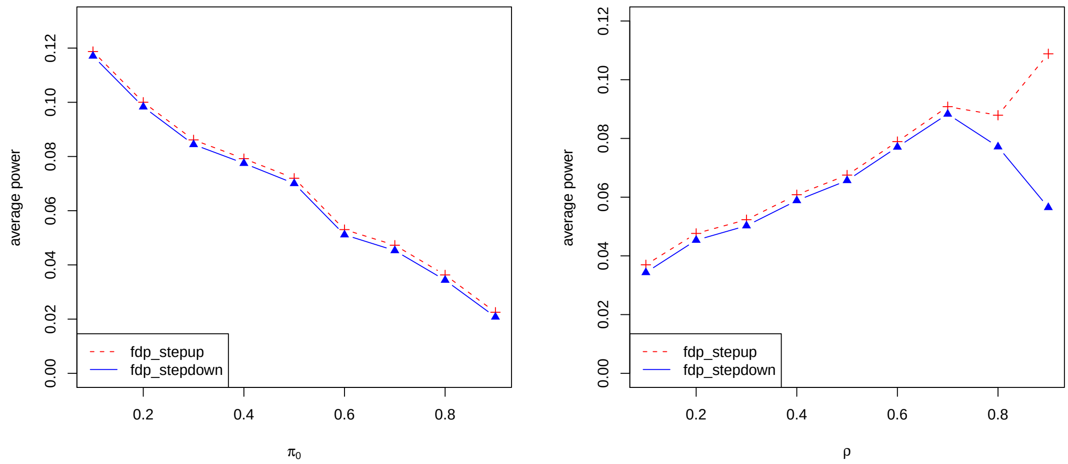


Figure 6.1: Comparison in terms of the average power of our proposed stepup procedure (red dashed) and the stepdown procedure in Lehmann and Romano (2008) (blue solid), both of which control the γFDP at level 0.05. Here $\gamma = 0.05$, the number of hypotheses is 5000, and the power is calculated over 5000 repetitions. The p-values are calculated from the set of simulated multivariate normal random variables with common correlation ρ. These normal random variables have mean equal 0 under the null hypotheses and have mean euqal 2 under the alternative. In panel (a) of this figure, we fix ρ to be 0.5 and let the proportion of true nulls π_0 move in $\{0.1, 0.2, \ldots, 0.9\}$ and in panel (b), we fix π_0 to be 0.05 and let ρ move in $\{0.1, 0.2, \ldots, 0.9\}$.

Condition 6.2.1

$$P(\hat{P}_i \leq t \mid \hat{Q}_1, \hat{Q}_2, \ldots, \hat{Q}m_1) \leq t, \qquad \textit{for} \qquad t \in (0,1) \qquad \textit{and} \qquad i \in I_0.$$

Before stating our procedure based on the bivariate distribution, we need to define the following additional condition:

Condition 6.2.2

$$P(\hat{P}_i \vee \hat{P}_j \leq t \mid \hat{Q}_1, \hat{Q}_2, \ldots, \hat{Q}m_1) \leq P(\hat{P}_i \vee \hat{P}_j \leq t), \quad \textit{for} \quad t \in (0,1) \quad \textit{and} \quad \{i,j\} \subseteq I_0.$$

For instance, when the group of null p-values and the group of nonull p-values are stochastically independent, both conditions are satisfied. We now proposed our stepdown γ-FDP procedure.

Theorem 6.2.2 *A stepdown procedure with critical values $\alpha_i, i = 1, \ldots, m$, defined as*

$$\alpha_i = \begin{cases} \dfrac{\lfloor \gamma i \rfloor + 1}{m - i + \lfloor \gamma i \rfloor + 1} & \textit{if } \lfloor \gamma i \rfloor = 0 \\ \dfrac{(\lfloor \gamma i \rfloor + 1)(\lfloor \gamma i \rfloor)}{(m - i + \lfloor \gamma i \rfloor + 1)(m - i + \lfloor \gamma i \rfloor)} & \textit{if } \lfloor \gamma i \rfloor \geq 1 \end{cases}$$

controls the γ-FDP under Condition 6.2.1 and Condition 6.2.2.

6.3 Controlling the k-PCER

We next state the following procedure that controls the k-PCER.

Theorem 6.3.1 (The Stepup Procedure Controlling k-PCER) *Suppose the set of p-values $\{p_i, i \in I_0\}$ satisfy Condition 2.7.7 and have the identical pairwise joint distribution, denoted by $F_2(\cdot,\cdot)$.*

Define the set of critical constants $\{\alpha_i, i = 1, 2, \ldots, m\}$ *as the following:*

$$F(\alpha_i, \alpha_i) = \begin{cases} \dfrac{(k-1)}{(m-1)}\alpha, & i \leq k \\ \dfrac{(k-1)}{(m-i+k-1)}\alpha, & i > k \end{cases}$$

Then, when $k \geq 2$*, a stepup procedure with this set of critical constants controls* k*-PCER at* α*.*

6.4 Simulation Study

To compare our proposed procedures with Lehmann and Romano's stepdown procedures, we performed a simulation study. We first specify the dependence model that will be used in several simulation studies in this work. This dependence structure for p-values was previously considered in Benjamini et al (2006) and Blankard & Roquain (2009).

Definition 6.4.1 (Model I)

$$Z_i \overset{\text{iid}}{\sim} N(0,1), \quad i = 0, 1, \ldots, m$$

$$Y_i = \sqrt{\rho} Z_0 + \sqrt{1-\rho} Z_i + \mu_i, \quad i = 1, ..., m$$

$$P_i = 1 - \Phi(Y_i)$$

where $\mu_i = 0$*, for* $i \in I_0$ *and* $\mu_i = \mu$*, for* $i \in I_1$*. Hence* ρ *gives the amount of the correlation, and the* μ *gives the amount of the signal.*

In the first simulation study, we did the following:

1. Generated m p-values according to the dependence structure described above. We fixed $m = 100$, $\mu = 2$ and $\pi_0 = 0.5$ in this experiment.

2. We let the parameter k vary in $\{2, 6, 14, 16, 20\}$ and let the parameter ρ vary in (0,1). For each of these configurations for k and ρ, we applied three kFWER procedures: the stepdown procedure of Lehmann and Romano (abbreviated as LR-SD), our stepdown procedure defined in Theorem 6.1.2 (abbreviated as pair-SD), and our stepup procedure defined in Theorem 6.1.3 (abbreviated as pair-SU).

3. For each configuration mentioned above, we repeated the process $n = 10000$ times.

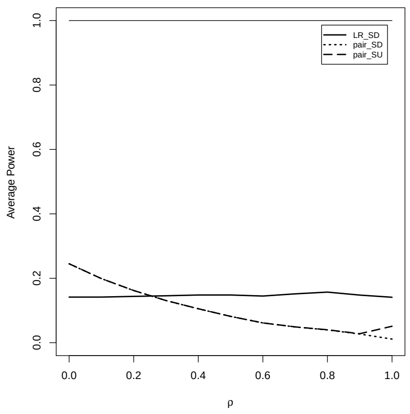
4. For each repetition and each configuration, we found the number of rejections R, the number of false rejection V, and the number of correct rejections S.

5. We then found, for each configuration of ρ and k, the simulated kFWER by calculating the proportion of incidents $V \geq k$ out of n repetitions, and found simulated average power by calculating the average of S/m_1.

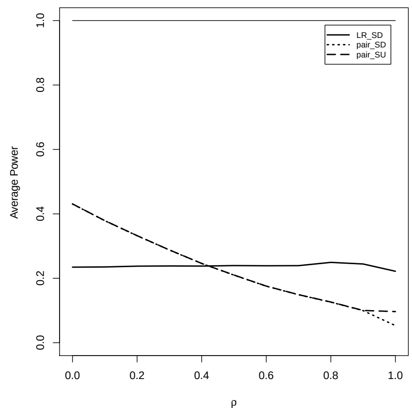
In Figure 6.2, for the chosen k-values, we plot average power versus correlation. We see that, compared to Lehmann and Romano's stepdown procedure, our procedures are more powerful under independence and mild dependence. We also see that, as k gets larger, our procedures are more powerful and there are more situations where our procedures offer improvement. However, when correlation gets high, our procedures are no longer an improvement.

We also compare the performance of our stepdown procedure defined in Theorem 6.2.2 with the stepdown procedure defined in Lehmann and Romano (2005) that control the tail probability of FDP through the following simulation study.

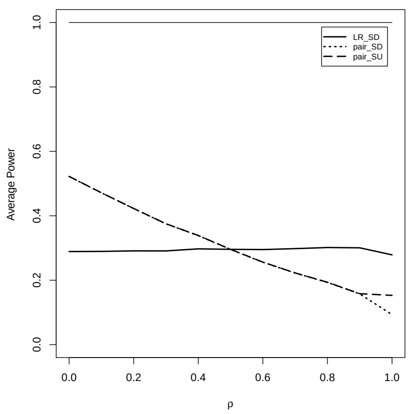
1. We independently generated the statistics for the set of the true null hypotheses and those



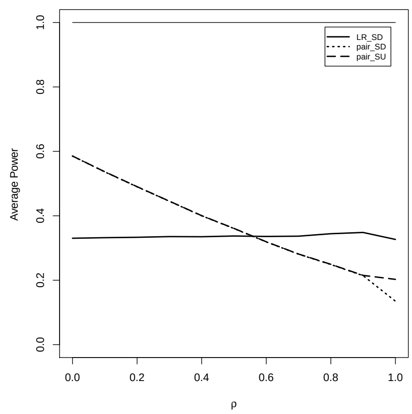
(a) $k = 2$



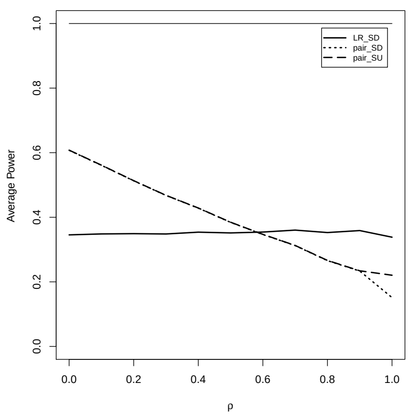
(b) $k = 6$



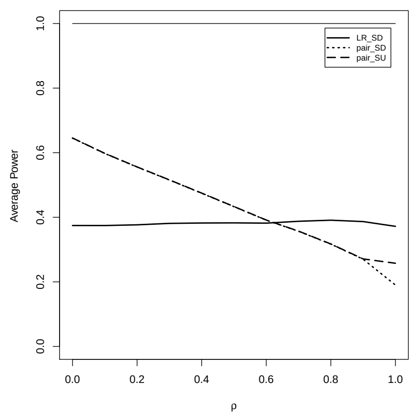
(c) $k = 10$



(d) $k = 14$

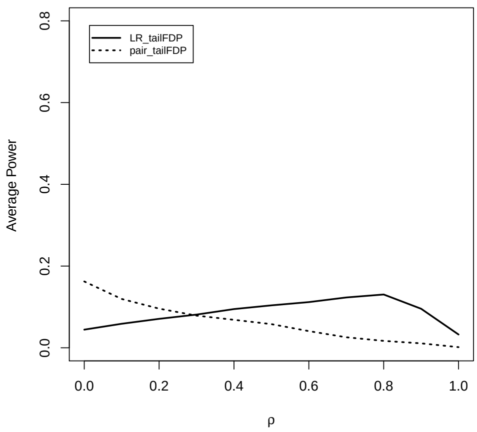


(e) $k = 16$

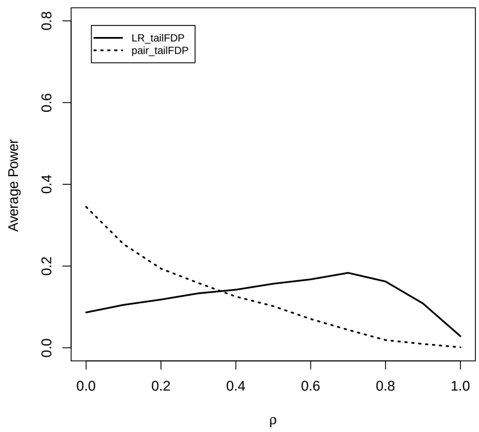


(f) $k = 20$

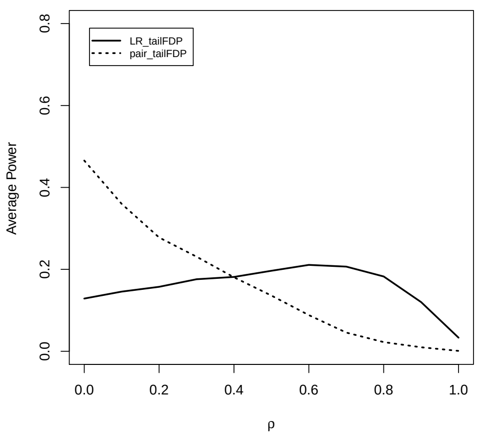
Figure 6.2: Comparison of the k-FWER Procedures



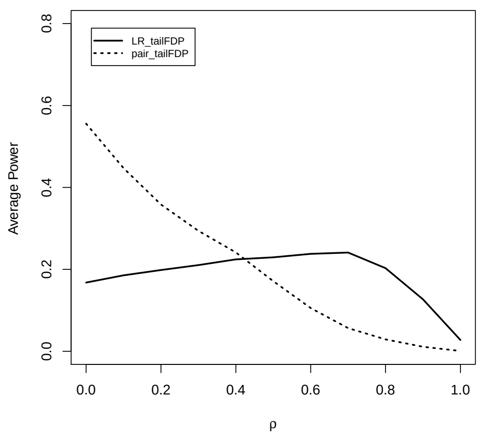
(a) $\gamma = 0.05$



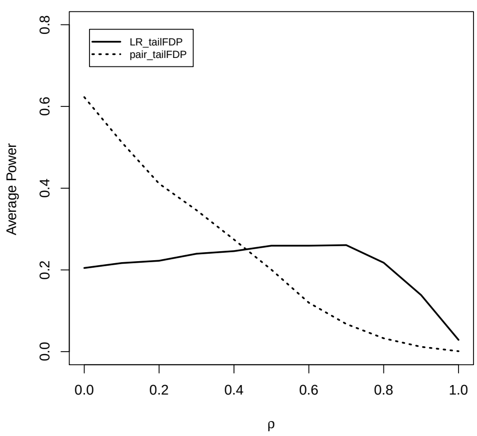
(b) $\gamma = 0.1$



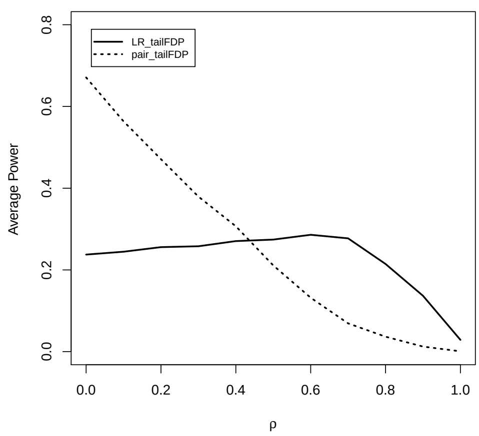
(c) $\gamma = 0.15$



(d) $\gamma = 0.2$



(e) $\gamma = 0.25$



(f) $\gamma = 0.3$

Figure 6.3: Comparison of the Procedures Controlling γ-FDP

for the set of the false null hypotheses. Within each set, the statistics are generated so that they are dependent with equal correlation. Specifically, we generated $m_1 = \pi_1 m$ dependent normal random variables $N(\mu, 1)$ with common correlation ρ, and generated separately $\pi_0 m$ dependent normal random variables $N(0, 1)$ with common correlation ρ. We fixed $m = 100$, $\mu = 2$ and $\pi_0 = 0.2$ in this experiment.

2. We let the parameter γ vary in $\{0.05, 0.1, 0.15, 0.2, 0.25, 0.3\}$ and the parameter ρ vary in (0,1). For each of these configurations for γ and ρ, we applied two procedures: the stepdown procedure of Lehmann and Romano (abbreviated as LR-tailFDP) and our stepdown procedure defined in Theorem 6.2.2 (abbreviated as pair-tailFDP).

3. For each configuration mentioned above, we repeated the process $n = 10000$ times.

4. For each repetition and each configuration, we found the number of rejections R, the number of false rejection V, and the number of correction rejections S.

5. We then found, for each configuration of ρ and k, the simulated average power by calculating the average of the S/m_1.

Both of the procedures should be valid under such dependence structures since both Condition 6.2.1 and 6.2.2 are satisfied.

In Figure 6.3, for the chosen γ-values, we plot average power versus correlation. We see that, compared to Lehmann and Romano's stepdown procedure, our procedure is more powerful under independence and mild dependence. We also see that, as γ gets larger, our procedure is more powerful and there are more situations where our procedure offers improvement. However, when correlation gets high, our procedure is no longer an improvement.

CHAPTER 7

NEW ADAPTIVE PROCEDURES

In this section, we give our results on the adaptive procedures.

7.1 Adaptive BH Procedures Controlling FDR under Dependence

7.1.1 The Procedures

We first restate Theorem 3.3 in Blankard and Roquain (2009) in terms of the adaptive BH procedure for later reference.

Lemma 7.1.1 *Under the PRDS condition, the stepup procedure with critical values*

$$\{i\alpha_1 G(\boldsymbol{P})/m, \qquad i = 1, \ldots, m\}$$

satisfies:

$$FDR \leq \alpha_1 + E\left[\frac{V}{R} I\left(R > 0, G(\boldsymbol{P}) > \frac{1}{\pi_0}\right)\right]$$

where $G(\boldsymbol{P})$ *is any estimator of* $1/\pi_0$.

The following two observations based on the procedures of Blankard and Roquain (2009), given in Theorem 4.2.2 and Theorem 4.2.2, motivate our new adaptive BH procedures. Firstly, it is possible to improve power by taking into account the dependence information. Secondly, we note that G($\boldsymbol{P}$) is a nondecreasing function of R_0 (the number of rejections at the first stage). Accordingly, as R_0 gets larger, the critical constants of the adaptive LSU procedure at the second stage also get larger. With these motivations in mind, we derived three adaptive BH procedures in this subsection.

Their adaptive procedure, given in Theorem 4.2.2, controls the FWER at the first stage. Whereas we propose to control the k-FWER at the first stage. The motivation of using k-FWER is two-fold. Firstly, the k-FWER is a more liberal error rate than the FWER and one expects to reject more hypotheses when seeking control of the k-FWER. Secondly, our proposed k-FWER controlling procedure can bring the dependence information into the picture.

Theorem 7.1.1 *Let* R_0 *be the number of rejections of a multiple testing procedure that controls the* k*FWER at level* α_0. *Then, under the PRDS condition and any other conditions that the* k*FWER procedure requires, the stepup procedure with critical values* $\left\{\frac{i\alpha_1}{m - R_0 + k - 1}, i = 1, 2, \ldots, m\right\}$ *controls the FDR at level* $\alpha_1 + \alpha_0$.

Specifically, let R_0 be the number of rejections from our stepup k-FWER procedure defined in Theorem 6.1.3. Then, the adaptive BH procedure, defined in Theorem 7.1.1, will control the FDR at level $\alpha_1 + \alpha_0$, when the following three conditions are satisfied: (i) the PRDS condition, (ii) Condition 2.7.7 and (iii) the identical pairwise joint distribution condition.

At the other side of the conservative error rate FWER lies the per-comparison error rate $E(V/m)$ (PCER). We next propose an adaptive BH procedure whose first step controls the PCER.

Theorem 7.1.2 *Let* $R_0 = \sum_{i=1}^{m}(P_i \le \alpha_0)$*. Then, under the PRDS condition, the step up adaptive procedure with critical values* $\left\{\frac{i\alpha_1 F_k\left(R_0/m\right)}{m}, i = 1, 2, \ldots, m\right\}$ *controls the FDR at level* $\alpha_1 + k\alpha_0$*, where* $k > 1$ *and*

$$F_k(x) = \begin{cases} 1 & 0 \le x \le 1/k \\ \dfrac{1-1/k}{1-x} & 1 > x > 1/k \end{cases} \tag{7.1}$$

It is possible to achieve more. We next considered the adaptive procedure where the first step controls k_1-PCER $= E\left[\frac{V}{m} I(V \ge k_1)\right]$, for some $k_1 \ge 1$.

Theorem 7.1.3 *Let* R_0 *be the number of rejections of a multiple testing procedure with* k_1*-PCER* $\le \alpha$*. Then, under the PRDS condition and under any other conditions that the procedure controlling* k_1*-PCER requires , a step up procedure with critical values* $\left\{\frac{i\alpha_1 F_{k_2}\left(R_0/m\right)}{m}, i = 1, 2, \ldots, m\right\}$ *controls FDR at level* $\alpha_1 + k_2\alpha_0$*, where* $k_1 \ge 1$*,* $k_2 > 1$ *and*

$$F_{k_2}(x) = \begin{cases} 1 & 0 \le x \le k_1/m + 1/k_2 \\ \dfrac{1-1/k_2}{1-x+k_1/m} & 1 \ge x > k_1/m + 1/k_2 \end{cases} \tag{7.2}$$

Specifically, where R_0 is the number of rejections from the our stepup procedure defined in Theorem 6.3.1, the adaptive BH procedure defined in Theorem 7.1.3 controls FDR at level $\alpha_1 +$

$k_2\alpha_0$, under the following four conditions: (i) the PRDS condition, (ii) $k_1 \geq 2$, (iii) Condition 2.7.7 and (iv) the identical pairwise joint distribution condition.

7.1.2 Simulation Study

We performed a simulation study to compare the procedure given in Theorem 4.2.2, abbreviated as BR08-dep-Holm, with two procedures: (i) our procedure defined in Theorem 7.1.1 with the first step being Lehmann and Romano's stepdown kFWER procedure, abbreviated as LR-kFWER-SD, and (ii) our procedure defined in Theorem 7.1.1 with the first step being our stepup pairwise kFWER procedure defined in Theorem 6.1.3, abbreviated as pair-kFWER-SU.

In this experiment, we did the following:

1. Under the dependence Model I, defined in Definition 6.4.1, we generated m p-values. We fixed $m = 1000$, $k = 15$ and $\pi_0 = 0.1$ in this experiment.

2. First we fixed $\rho = 0.1$ and let parameter μ vary in $\{0.5, 5\}$. For each value of μ, we applied five adaptive procedures: (i) BH(LSU) procedure, (ii) oracle adaptive procedure, (iii) BR08-dep-Holm, (vi) LR-kFWER-SD, and (v) pair-kFWER-SU. Next we fixed $\mu = 3$ and let parameter ρ vary in (0,1) and then performed the five tests for each value of ρ.

3. For each of the configurations for μ and ρ, we repeated the process $n = 1000$ times.

4. For each repetition, each configuration and each procedure, we found the number of rejections R, the number of false rejections V and the number of correction rejections S.

5. We then found the simulated average power by calculating the average of S/m_1, and we further scaled the average power of each procedure by dividing it by the average power of the oracle adaptive procedure.

6. We calculated the simulated false nondiscovery rate (FNR) by calculating the average value of $\dfrac{m_1 - S}{m - R}$.

The first column of Figure 7.1 shows the the FDR, the relative average power to the oracle adaptive BH procedure, and the FNR of the procedures, when we fix $\rho = 0.1$ and let μ vary in $\{0.5, 5\}$. We see that both LR-kFWER-SD and pair-kFWER-SU are more powerful than BR08-dep-Holm, and they both improve the BH procedure when the signal is large. Compared to LR-kFWER-SD, the procedure pair-kFWER-SU is more powerful and improves the BH procedure in more cases.

The second column of Figure 7.1 shows the FDR, the relative average power to the oracle adaptive BH procedure, and the FNR of the procedures, when we fix $\mu = 3$ and let ρ vary in $\{0, 1\}$. We see that both LR-kFWER-SD and pair-kFWER-SU are more powerful than BR08-dep-Holm for small ρ, and they both improve the BH procedure.

We performed an analogous simulation study to compare the procedure given in Theorem 4.2.3, abbreviated as BR08-dep-LSU, with two procedures: (i) our procedure defined in Theorem 7.1.2, abbreviated as dep-E(V/m) and (ii) our procedure defined in Theorem 7.1.3, abbreviated as dep-E(V/m[k1]), with the first stage being our stepup procedure defined in Theorem 6.3.1.

We fix $m = 1000$, $n = 10000$, $k_1 = 45$, $k_2 = 2$ and $\pi_0 = 0.1$ in this experiment.

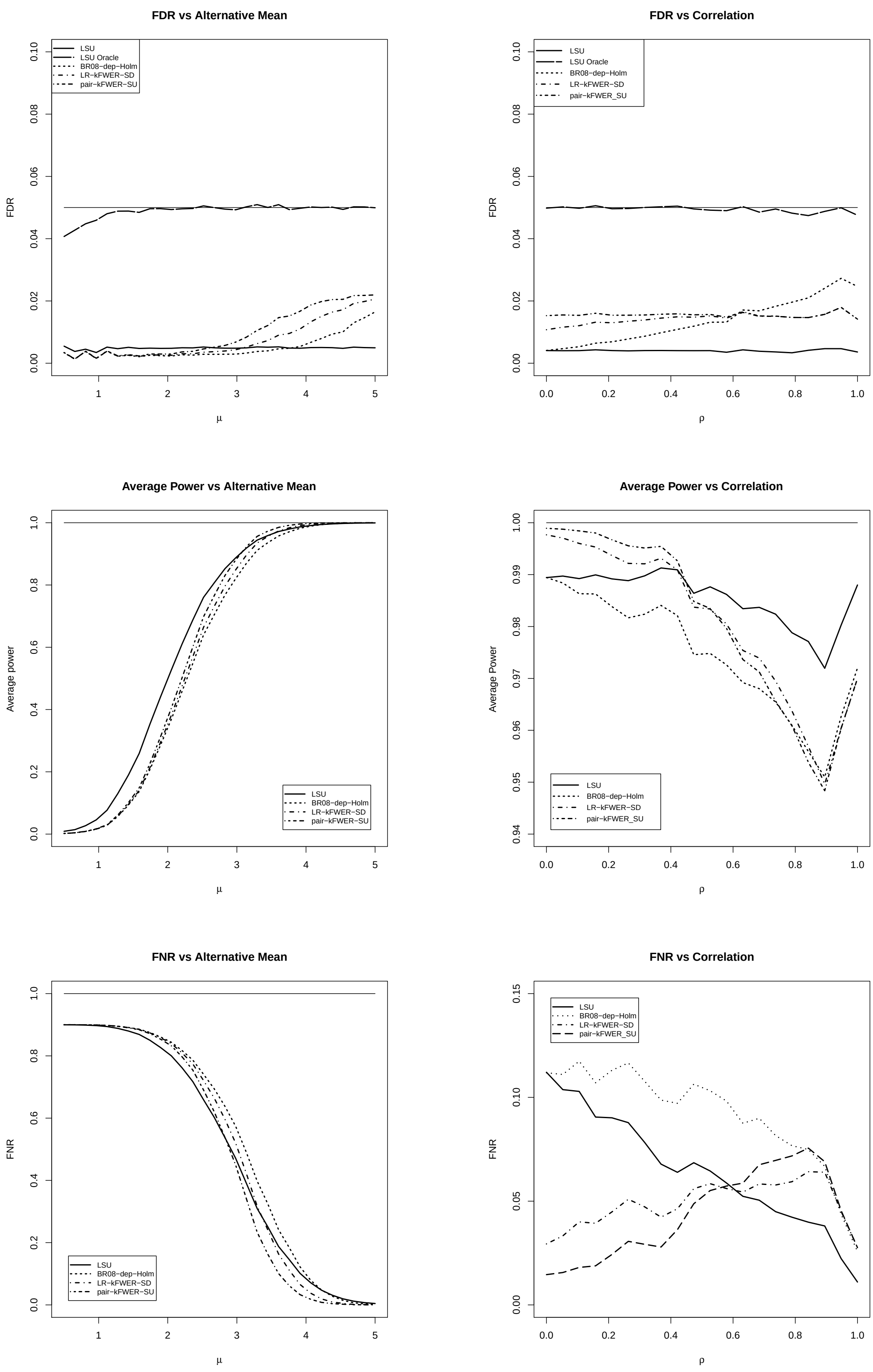


Figure 7.1: Comparisons of Adaptive BH Procedures

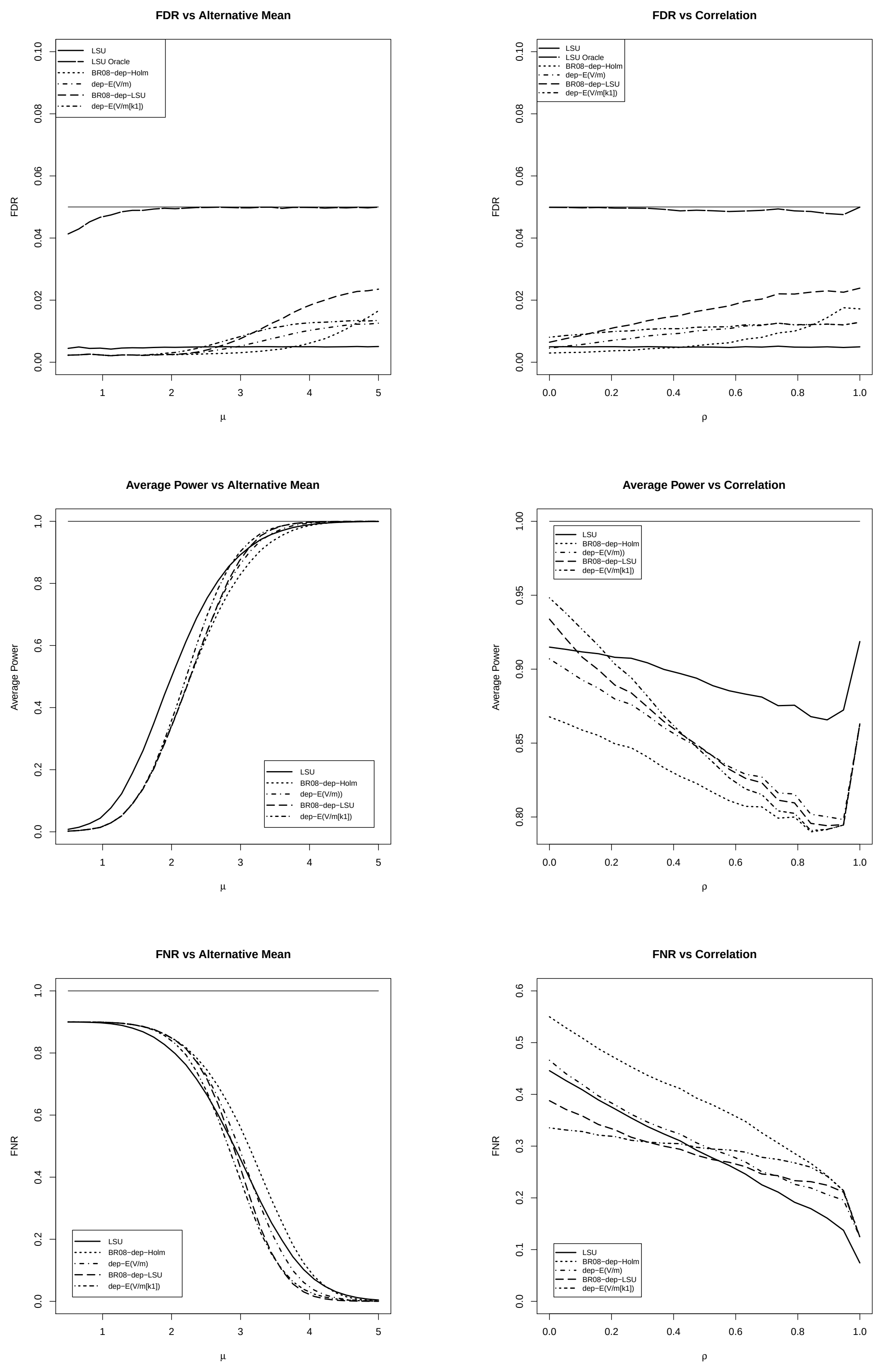


Figure 7.2: Comparisons of Adaptive BH Procedures

The first column of Figure 7.2 shows FDR, the average power, and the FNR of the procedures, when we fix $\rho = 0.1$ and let μ vary in $\{0.5, 5\}$. We see that for most values of μ, our dep-E(V/m[k1]) procedure is more powerful than BR08-dep-LSU. Also, BR08-dep-LSU improves the BH procedure when μ is high whereas our dep-E(V/m[k1]) procedure improves the BH procedure in more situations. When the signal is high, our dep-E(V/m) also improves the BH procedures, even though it seems not to be as powerful as BR08-dep-LSU.

The second column of Figure 7.2 shows FDR, the average power, and the FNR of the procedures, when we fix $\mu = 3$ and let ρ vary in $\{0, 1\}$. We see that our dep-E(V/m[k1]) procedure is more powerful than BR08-dep-LSU, when ρ is not large, say, less than 0.5. When the correlation is very high, dep-E(V/m) seems to perform better than dep-E(V/m[k1]) and BR08-dep-LSU.

7.1.3 Robustness

To see how robust are our procedures against the common pairwise distribution assumption, we performed a simulation study. We consider the following dependence structure for p-values.

Definition 7.1.1 (Model II) *Let $\boldsymbol{Y} = (Y_1, ..., Y_m)$. We have the following dependence structure for the p-values $\{P_1, \ldots, P_m\}$:*

$$\boldsymbol{Y} \sim \mathcal{N}(\mu, \Sigma)$$

$$P_i = 1 - \Phi(Y_i), \quad i = 1, 2 \ldots, m.$$

where $\mu = (\mu_1, \mu_2, \ldots, \mu_m)$ *with* $\mu_i = 0$ *for* $i \in I_0$ *and* $\mu_i = \bar{\mu}$ *for* $i \in I_1$*, and*

$$\Sigma_{m,m} = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{m-1} \\ \rho & 1 & \rho & \cdots & \rho^{m-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{m-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{m-1} & \rho^{m-2} & \rho^{m-3} & \cdots & 1 \end{pmatrix}$$

In this model, $\bar{\mu}$ *gives the amount of the signal and* ρ *gives the amount of correlation.*

Under the above structure, we generated m p-values. We fix $m = 100$, $n = 1000$, $k_1 = 4$, $k_2 = 2$ and $\pi_0 = 0.1$ in this experiment. First we fix $\rho = 0.1$ and let parameter $\bar{\mu}$ vary in $\{0.5, 5\}$. We then fix $\mu = 2.5$ and let parameter ρ vary in $(0, 1)$. For each of these configurations, we performed the following five tests: (i) BH (LSU) procedure, (ii) oracle adaptive procedure, (iii) BR08-dep-LSU, (iv) dep-E(V/m), and (v) dep-E(V/m[k1]). In this experiment, when calculting the critical constants for the procedures based on the bivariate distribution, we use the bivariate normal distribution with correlation coefficient ρ_0 to substitute for the identical pairwise distribution. Here ρ_0 is the average of all the pairwise correlations in the model.

The first column of Figure 7.3 shows the FDR, the average power, and the FNR of the procedures, when we fix $\rho = 0.1$ and let μ vary in $\{0.5, 5\}$. We see that the FDR is still controlled, and we see that, for most values of μ, our dep-E(V/m[k1]) procedure is more powerful than the BR08-dep-LSU. Both the BR08-dep-LSU procedure and our dep-E(V/m[k1]) procedure improve the BH procedure when $\bar{\mu}$ is high, whereas our our dep-E(V/m[k1]) improves the BH procedure in more situations.

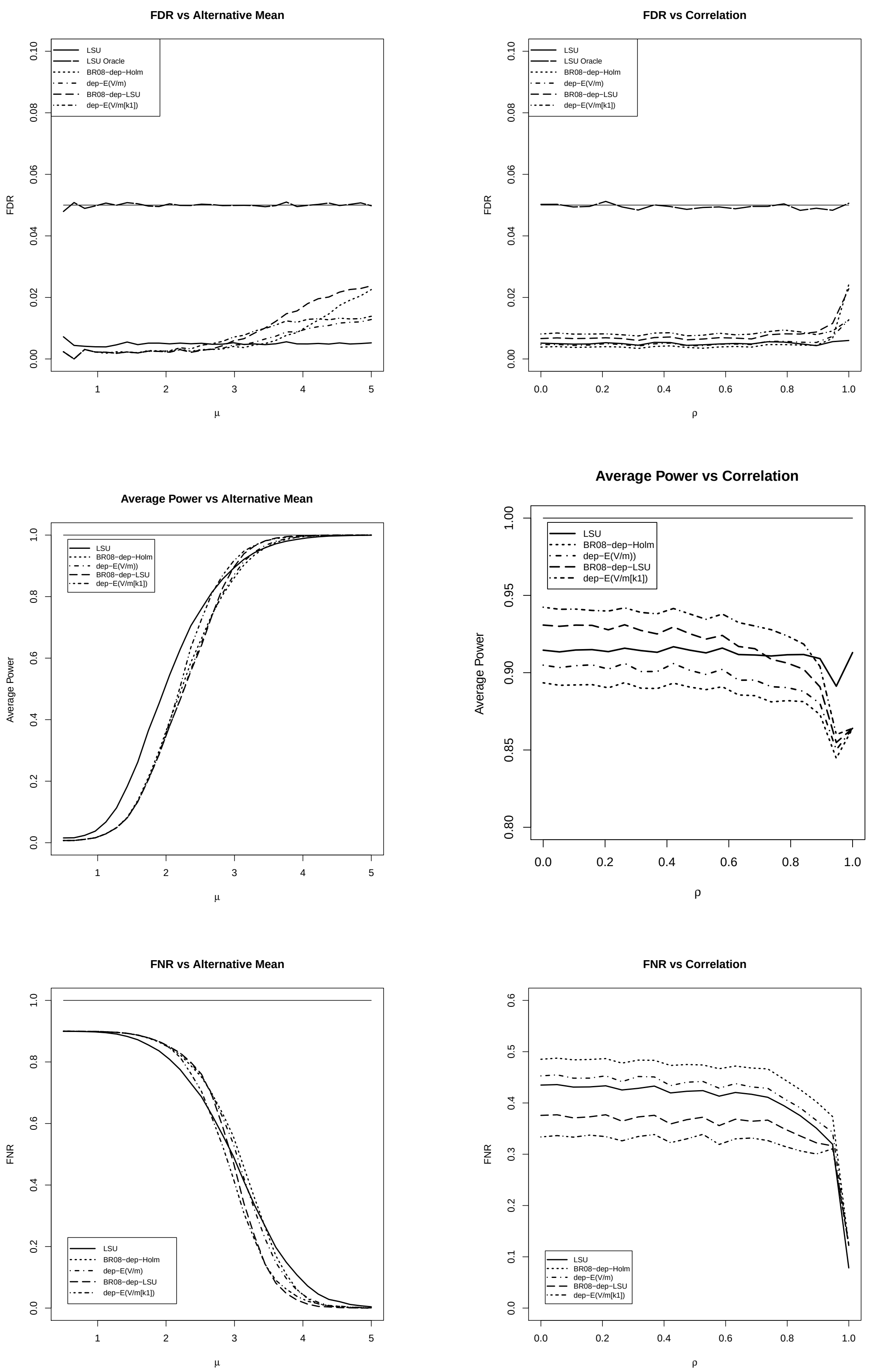


Figure 7.3: Comparisons of Adaptive BH Procedures

The second column of Figure 7.3 shows the FDR, the average power, and the FNR of the procedures, when we fix $\bar{\mu} = 3$ and let ρ vary in $\{0, 1\}$. We see that our dep-E(V/m[k1]) procedure is more powerful than the BR08-dep-LSU, and they both improve the BH procedure for almost all values of ρ, except when correlation get very high, say above 0.9.

7.2 Adaptive Procedures Controlling FWER under Dependence

7.2.1 The Procedures

The information of π_0 can also be used to improve FWER procedures.

Definition 7.2.1 (Dykstra, Hewett, and Thompson, 1973) *Random variables $X_1, \ldots, X_m$ are positively orthant dependent if*

$$P(X_1 < x_1, \ldots, X_m < x_m) \geq \prod_{i=1}^{m} Pr(X_i < x_i) \tag{7.3}$$

Lemma 7.2.1 *Let $G(\boldsymbol{P})$ be an estimate of $1/\pi_0$. The adaptive Bonferroni procedure rejects H_i whenever*

$$P_i < \alpha_1 G(\boldsymbol{P})/m$$

and the adaptive Sidak procedure rejects H_i whenever

$$P_i \leq 1 - (1 - \alpha_1)^{G(\boldsymbol{P})/m}$$

Then an adaptive Bonferroni procedure will satisfy (7.4) under arbitrary dependence, whereas an adaptive Sidak procedure satisfies (7.4) if the set of null p-values satisfy the positive orthant dependence condition defined in Definition 7.2.1

$$FWER \leq \alpha_1 + E\left[V \geq 1, G(\boldsymbol{P}) \geq \frac{1}{\pi_0}\right] \tag{7.4}$$

In a very similar fashion of adaptive BH procedures, we can develop adaptive Bonferroni and Sidak procedures that control FWER under positive dependence.

Theorem 7.2.1 *Let R_0 be the number of rejections of a multiple testing procedure that controls the k-FWER at level α_0 and let $G(\boldsymbol{P}) = m/(m - R_0 + k - 1)$. Then, under any dependence conditions that the kFWER procedure requires, our adaptive Bonferroni procedure defined in Lemma 7.2.1 controls the FWER at level $\alpha_0 + \alpha_1$, whereas our adaptive Sidak procedure defined in Lemma 7.2.1 controls the FWER at level $\alpha_0 + \alpha_1$, when the additional positive orthant dependence condition defined in Definition 7.2.1 is satisfied.*

Theorem 7.2.2 *Let $R_0 = \sum_{i=1}^{m}(P_i \leq \alpha_0)$ and let $G(\boldsymbol{P}) = F_k(R_0/m)$, where $F_k(x)$ is given in (7.1). Then, for $k > 1$, an adaptive Bonferroni procedure defined in Lemma 7.2.1 controls the FWER at level $\alpha_1 + k\alpha_0$, whereas our adaptive Sidak procedure defined in Lemma 7.2.1 controls the FWER at level $\alpha_1 + k\alpha_0$, if the additional positive orthant dependence condition defined in Definition 7.2.1 is satisfied.*

Theorem 7.2.3 *Let R_0 be the number of rejections of a multiple testing procedure with k_1-PCER $\leq \alpha_0$ and let $G(\boldsymbol{P}) = F_{k_2}(R_0/m)$ where F_{k_2} is defined in (7.2) for $k_1 \geq 1$ and $k_2 > 1$. Then, under* *any conditions that the procedure controlling k_1-PCER requires, our adaptive Bonferroni procedure defined in Lemma 7.2.1 controls FWER at level $\alpha_1 + k_2\alpha_0$, whereas our adaptive Sidak procedure defined in Lemma 7.2.1 controls FWER at level $\alpha_1 + k_2\alpha_0$, when the additional positive orthant dependence condition defined in Definition 7.2.1 is satisfied.*

7.2.2 Simulation Study

We performed a simulation study to compare the following seven adaptive Bonferroni procedures: (i) Bonferroni procedure, (ii) oracle adaptive Bonferroni procedure, (iii) our procedure defined in Theorem 7.2.1 with the first step a Holm procedure, abbreviated as BR08-dep-Holm, (iv) our procedure defined in Theorem 7.2.1 with the first stage being Lehmann and Romano's stepdown kFWER procedure, abbreviated as LR-kFWER-SD, (v) our procedure defined in Theorem 7.2.1 with the first stage being our stepup pairwise k-FWER procedure, abbreviated as pair-kFWER-SU, (vi) our procedure defined in Theorem 7.2.2, abbreviated as dep-E(V/m), and (vii) our procedure defined in Theorem 7.2.3 with the first stage being our stepup procedure defined in Theorem 6.3.1, abbreviated as dep-E(V/m[k1]).

Under the dependence model I, defined in Definition 6.4.1, we generated $m = 100$ hypotheses. We chose $k = k_1$ to be 3, 6, and 9. We fixed $k_2 = 2$, the number of simulations $n = 1000$ and the proportion of true null hypotheses $\pi_0 = 0.1$.

The first column of Figure 7.4 shows, for $k = 3, 6,$ and, 9, the scaled average power, scaled with respect to the power of the oracle adaptive Bonferroni procedure, where we fix $\rho = 0.1$ and let μ

vary in $\{0.5, 5\}$. We see that all the adaptive procedures improve the Bonferroni procedure, when the signal is relatively large. The order from the most powerful to the least powerful adaptive procedures follows: pair-kFWER-SU, dep-E(V/m[k1]), LR-kFWER-SD, dep-E(V/m), and BR08-dep-Holm.

The second column of Figure 7.4 shows, for $k = 3, 6,$ or, 9, the average power, when we fix $\mu = 3$ and let ρ vary in $\{0, 1\}$. We see that, when correlation is low and moderate, pair-kFWER-SU and dep-E(V/m[k1]) are more powerful than LR-kFWER-SD and dep-E(V/m), and when correlation is higher, it is the other way around.

7.3 Conclusions

In this chapter, we developed new adaptive BH procedures that control the FDR under positive dependence. We see that, by incorporating correlation, we derived more powerful procedures. In practice, one can apply these procedures and see whether they improve the BH procedure. If not, then one can use the BH procedure. Otherwise, one can use our procedures and assure that the FDR is controlled with the assumption that the required dependence conditions are satisfied. We also derived new adaptive Bonferroni and Sidak procedures that control the FWER under positive dependence.

Although our FDR adaptive procedures form an improvement of the procedures given in Blankard and Roquain (2009), they are still conservative and may be less powerful than the BH procedure. Note that dependence was only taken into account in the first stage where we formulated an estimator of π_0. The second stage is still conservative in the dependence point of view. More powerful

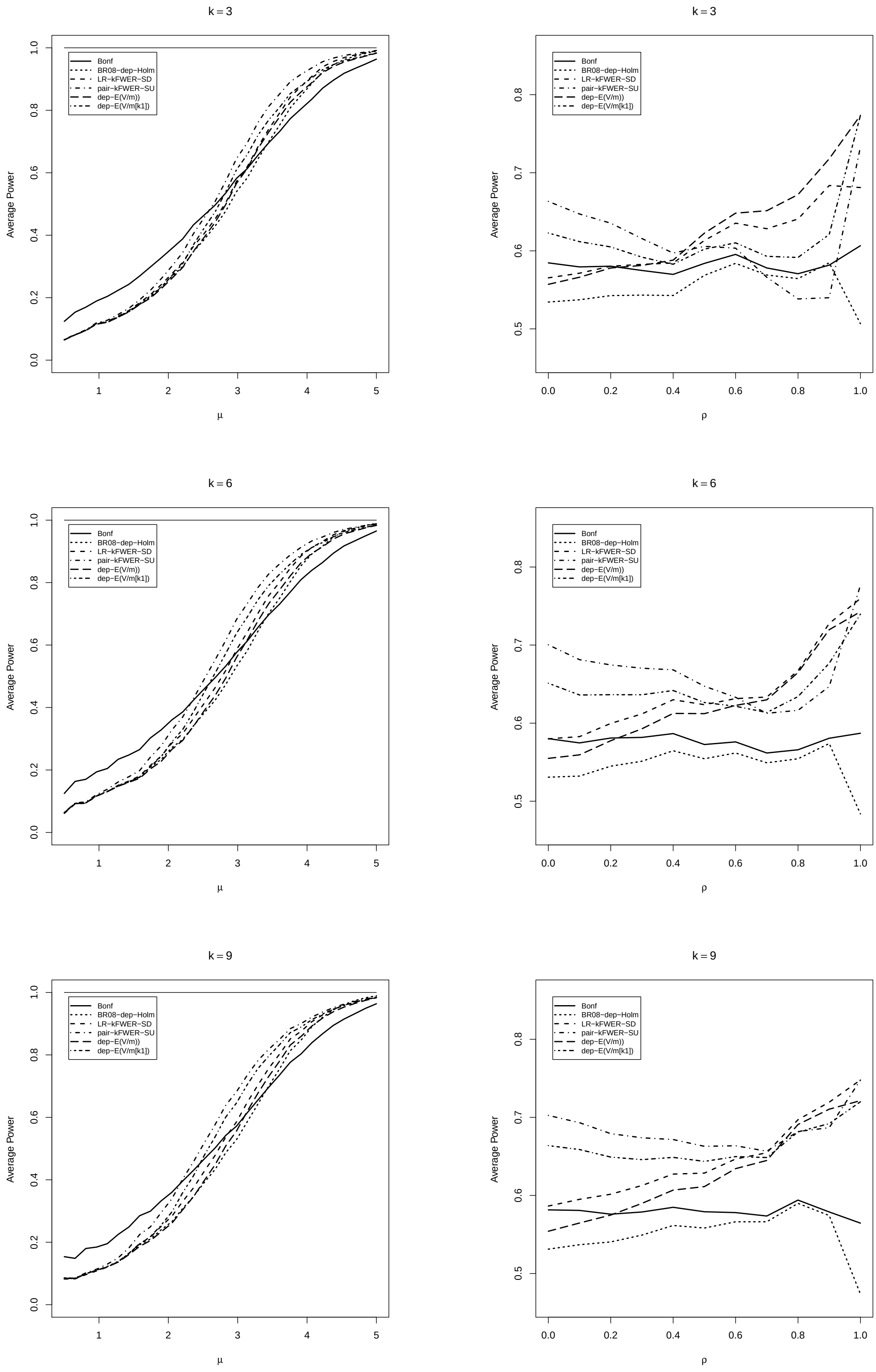


Figure 7.4: Power Comparisons of Adaptive Bonferroni Procedures

procedures may be possible and we leave this for future research.

CHAPTER 8

OPTIMAL MULTIPLE TESTING PROCEDURES UNDER A NEW LOSS FUNCTION

This chapter consists of our result on the compound decision framework for multiple testing. We first propose a more general form of loss function than (4.6). Specifically, for the ith component problem, instead of considering the 0-1 loss L_1^i in (4.5) for the type II errors , we consider a loss function that takes into account the magnitude of the signal. This new loss function has a natural connection with a weighted version of the mFDR and a weighted version of the mFNR (which we will define later), where the set of weights indicates the magnitude of the signal strength. We study the compound decision problem under this new loss function, and we connect it to the multiple testing problem of searching for the procedure that minimizes the weighted mFNR while controlling the weighted mFDR. We derive an oracle procedure through this connection. A numerical example shows that our oracle procedure which controls the weighted mFDR is more powerful than the oracle procedure in Sun and Cai (2007) that controls the mFDR. We also derive a data driven procedure which achieves the performance of our oracle procedure when the number m of hypotheses is large. A simulation study was performed to compare our data driven procedure with the data driven procedure in Sun and Cai (2007). Again, our data driven procedure is more powerful. Hence, by incorporating the prior information of the signal strength through weights, more power is gained.

8.1 The Compound Decision Rule under the New Loss Function

8.1.1 The Model and the Loss Function

We assume the following model framework. Consider testing m hypotheses $H_i : \mu_i = 0, i = 1, \ldots, m$, based on observations $X_i, i = 1, \ldots, m$, where $X_i \mid \mu_i \sim f(x_i - \mu_i)$. Again let θ_i be the indicator that H_i is false, where the θ_i's are independent Bernoulli's with $P(\theta_i = 1) = \pi_1$. Given the indicator θ_i's, the parameters μ_i's are from the following distribution

$$\mu_i \mid \theta_i = 0 \sim h_0$$

$$\mu_i \mid \theta_i = 1 \sim h(\mu)$$

where h_0 is a point mass at 0.

Then, marginally speaking, the X_i's are i.i.d. random variables following

$$X_i \mid \theta_i = 0 \sim f(x)$$
$$X_i \mid \theta_i = 1 \sim \int f(x-\mu_i)h(\mu_i)d\mu$$

This is a special case of the mixture model defined in (4.1).

Let $\boldsymbol{\delta}$ be a decision regarding the m hypotheses and $\boldsymbol{\delta}\in \{0,1\}^m$. We reject H_i if and only if the corresponding $\delta_i = 1$. Define the following loss function:

$$L_\lambda(\boldsymbol{\theta},\boldsymbol{\delta}) = \frac{1}{m}\sum_i \{\lambda(1-\theta_i)\delta_i + \mu_i^2\theta_i(1-\delta_i)\} \tag{8.1}$$

where λ is the relative weight given to the loss for a Type I error. We then find the compound decision rule $\boldsymbol{\delta}$ which minimizes the expectation of this loss function .

Note that the loss for the Type II error, for the ith hypothesis, is measured as $\theta_i(1-\delta_i)(\mu_i-\delta_i)^2$. This contrasts with the 0-1 loss, $\theta_i(1-\delta_i)(\delta_i-\theta_i)^2$, which is equal to 0 or 1, considered in Sun and Cai (2007) and many others. In this new loss function, we put more penalty on the type II errors for those hypothesis with higher signal strength. Intuitively, this is appropriate, since the higher the signal strength, the more difficult for us to tolerate the error of falsely accepting the corresponding hypothesis.

Suppose the decision rule $\boldsymbol{\delta}$ is based on the test statistics $\boldsymbol{T(x)} = (T_1(\boldsymbol{x}),\ldots,T_m(\boldsymbol{x}))$, and suppose that $T_i(\boldsymbol{x}) = T(x_i)$. In other words, we restrict our discussion to the symmetric compound decision rules. Let $T \mid \theta = 1, \mu \sim G^\mu(t)$, and $G^j(t) = P(T < t \mid \theta = j)$ for $j = 0,1$, where $G^1(t) = \int G^\mu(t)h(\mu)d\mu$, and let $g^\mu(t) = dG^\mu(t)/dt$.

Note that we will restrict ourselves to the collection of symmetric decision rules that are of the form $\boldsymbol{\delta}(\mathbf{T}, c) = I(\mathbf{T} < c\mathbf{1})$, which we denote as the collection $\mathcal{D}$.

8.1.2 Optimal Rule for the Compound Decision Problem

The following theorem gives the compound decision rule that minimizes the expected loss function (8.1).

Theorem 8.1.1 *Let* $\boldsymbol{\delta}^* = \{\delta_i^*\}_1^m$ *with* $\delta_i^* = I\left(\frac{\pi_0 f(X_i)}{\pi_1 \int \mu^2 f(X_i - \mu)h(\mu)d\mu} \leq \frac{1}{\lambda}\right)$. *Then* $\boldsymbol{\delta}^*$ *minimizes the expected loss function (8.1).*

If we replace μ by $I(\mu \neq 0)$, where $I(\cdot)$ is the indicator function, then this decision rule is equivalent to the one based on the likelihood ratio test statistic, or the local fdr statistic, which is the case for the oracle procedure in Sun and Cai (2007).

We prove the theorem below.

First we note the following:

$$\mu_i \mid x_i, \theta_i = 0 \sim h_0$$

$$\mu_i \mid x_i, \theta_i = 1 \sim \frac{f(x_i - \mu_i)h(\mu_i)}{\int f(x_i - \mu_i)h(\mu_i)d\mu_i}$$

Then we have

$$\begin{aligned} E(L_\lambda(\boldsymbol{\theta},\boldsymbol{\delta}) \mid \mathbf{x}) =& \frac{1}{m}\sum\left\{\lambda\delta_i P(\theta_i=0 \mid x_i)+(1-\delta_i)E(\mu_i^2 I(\theta_i=1)\mid x_i)\right\} \\ =& \frac{1}{m}\sum\left\{\lambda\delta_i P(\theta_i=0 \mid x_i)+(1-\delta_i)\int \mu_i^2 f(\mu_i \mid x_i,\theta_i=1)P(\theta_i=1\mid x_i)d\mu_i\right\} \\ =& \frac{1}{m}\sum\frac{P(\theta_i=1\mid x_i)}{\int f(x_i-\mu_i)h(\mu_i)d\mu_i}\int \mu_i^2 f(x_i-\mu_i)h(\mu_i)d\mu_i + \\ & \frac{1}{m}\sum\delta_i\left\{\lambda P(\theta_i=0\mid x_i)-\frac{P(\theta_i=1\mid x_i)}{\int f(x_i-\mu_i)h(\mu_i)d\mu_i}\int \mu_i^2 f(x_i-\mu_i)h(\mu_i)d\mu_i\right\} \end{aligned}$$

Then $\boldsymbol{\delta}$ minimizes the conditional risk if

$$\boldsymbol{\delta} = I\left(\lambda P(\theta_i=0|x_i) < \frac{P(\theta_i=1|x_i)}{\int f(x_i-\mu_i)h(\mu_i)d\mu_i}\int \mu_i^2 f(x_i-\mu_i)h(\mu_i)d\mu_i\right).$$

Simplifying the previous expression, one can see that the Bayes decision rule is as given in Theorem 8.1.1.

8.2 Connection to the Multiple Testing Problem

We start by defining the following set of weights. For $i=1,\ldots,m$, let

$$w_i = I(\theta_i=0)+\mu_i^2 I(\theta_i=0). \tag{8.2}$$

We then define the following weighted mFDR and weighted mFNR:

$$m\text{FDR}^* = \frac{E\sum w_i V_i}{E\sum w_i R_i} \quad \text{and} \quad m\text{FNR}^* = \frac{E\sum w_i K_i}{E\sum w_i A_i}$$

Note that both of these two measures take into account the distribution of the alternative means through the weights. In the special case where $h(\mu)$ has a point mass at 1, these two forms of error reduce to the mFDR and the mFNR. The concepts of weighted FDR and weighted FNR have been introduced in Benjamini and Hochberg (1997) and further studied in Genovese et al (2006).

We can now make the connection between the compound decision problem under the loss function (8.1) and the multiple testing problem of minimizing the mFNR* subject to a constraint on the mFDR*.

First we define the following condition for the test statistics. Let T be a test statistic, and $g^{\mu}(t)$ be the density of $T|\theta = 1, \mu$ and $g^0(t)$ be the density of $T|\theta = 0$. A test statistic $\mathbf{T}$ satisfies the weighted monotone likelihood ratio condition (wMLR), when

$$\frac{\int \mu^2 g^{\mu}(t) h(\mu) d\mu}{g^0(t)} \quad \text{is monotonically decreasing in } t$$

The monotone likelihood ratio condition (MLR) with μ^2 in the numerator being removed in the previously formula, has been considered in Sun and Cai (2007).

Let $\mathcal{T}$ be the class of test statistics $\mathbf{T}$ that satisfy the wMLR condition. We will later show that the test statistic on which the compound decision rule is based always satisfies this condition. For convenience, we define $S^* = \sum w_i S_i$, $R^* = \sum w_i R_i$, $K^* = \sum w_i K_i$ and $A^* = \sum w_i A_i$, where the w_i's are defined in (8.2). We now have the following theorems.

Theorem 8.2.1 *Consider any $\boldsymbol{\delta}(\mathbf{T}) \in \mathcal{D}$, where $\mathcal{D}$ is the collection of decision rules that are of the form $\boldsymbol{\delta}(\mathbf{T}, c) = I(\mathbf{T} < c\mathbf{1})$. If $\mathbf{T} \in \mathcal{T}$, then mFDR* is monotonically increasing in c, and mFNR* is monotonically decreasing in c. Also, $E(R^*)$ is a nondecreasing function of c, mFDR* is monotone increasing in $E(R^*)$, and mFNR* is monotonically decreasing in $E(R^*)$.*

Proof

First we show that, for any $\mathbf{T} \in \mathcal{T}$,

$$\frac{\int_{-\infty}^{c} \int \mu^2 g^{\mu}(t)h(\mu)d\mu dt}{G^0(c)} > \frac{\int \mu^2 g^{\mu}(c)h(\mu)d\mu}{g^0(c)} \tag{8.3}$$

and

$$\frac{\int_{c}^{\infty} \int \mu^2 g^{\mu}(t)g(\mu)d\mu dt}{1 - G^0(c)} < \frac{\int \mu^2 g^{\mu}(c)h(\mu)d\mu}{g^0(c)} \tag{8.4}$$

To see that (8.3) is true, we note that, by the wMLR, we have the following:

$$\begin{aligned}
\frac{\int_{-\infty}^{c} \int \mu^2 g^{\mu}(t)h(\mu)d\mu dt}{G^0(c)} &= \frac{\int_{-\infty}^{c} g^0(t)\frac{\int \mu^2 g^{\mu}(t)h(\mu)d\mu}{g^0(t)}dt}{G^0(c)} \\
&> \frac{\int \mu^2 g^{\mu}(c)(\mu)d\mu}{g^0(c)}\frac{\int_{-\infty}^{c} g^0(t)}{G^0(c)} \\
&= \frac{\int \mu^2 g^{\mu}(c)(\mu)d\mu}{g^0(c)}
\end{aligned}$$

Similarly, (8.4) can be proved.

Next, we note that $m\text{FNR}^*(c) = \frac{H(c)}{1+H(c)}$, where $H(c) = \frac{G^0(c)}{\int_{-\infty}^{c} \int \mu^2 g^{\mu}(t)h(\mu)d\mu dt}$. Then we have

$$H'(c) = \frac{g^0(c)\int_{-\infty}^{c} \int \mu^2 g^{\mu}(t)h(\mu)d\mu dt - G^0(c)\int \mu^2 g^{\mu}(c)h(\mu)d\mu}{(\int_{-\infty}^{c} \int \mu^2 g^{\mu}(t)h(\mu)d\mu dt)^2}$$

By (8.3), $H'(c) > 0$. Hence mFDR*(c) is monotonically increasing in c. Similarly, we can show that mFNR*(c) is monotonically decreasing in c.

Theorem 8.2.2 *Let* $\mathbf{\Lambda}$ *be the test statistic on which the compound decision rule minimizing the expected loss function* (8.1) *is based. Among all the tests of the form* $\boldsymbol{\delta}(\mathbf{T}, c) = I(\mathbf{T} < c\mathbf{1})$*, where*

$T \in \mathcal{T}$, the multiple testing procedure based on $\mathbf{\Lambda}$ is optimal, in the sense that, for any given level α of mFDR*, it has the smallest mFNR*.

Proof

First note that the loss function can be written as:

$$\begin{aligned} E(L_\lambda(\boldsymbol{\theta}, \boldsymbol{\delta})) &= \frac{1}{m}\left\{\lambda E(V) + E\left(\sum \mu_i^2 - (R^* - V)\right)\right\} \\ &= \frac{1}{m} E\left(\sum \mu_i^2\right) + \frac{1}{m}\left\{\lambda E(V) - E(S^*)\right\} \end{aligned} \tag{8.5}$$

Let $\mathbf{T} \in \mathcal{T}$ and let $\boldsymbol{\delta}_{\boldsymbol{T}} = (\mathbf{T} < c_T \mathbf{1})$ be a multiple testing procedure that controls mFDR* at level α. Then, by Theorem 8.2.1, there is a unique $c_T(\alpha)$. Let $r_T^* = E(R^*(\boldsymbol{\delta}_{\boldsymbol{T}}))$, also let $\boldsymbol{\delta}_{\mathbf{\Lambda}} = (\mathbf{\Lambda} < c_{\mathbf{\Lambda}} \mathbf{1})$ be the multiple testing procedure that is based on the compound test statistic $\mathbf{\Lambda}$ and has $E(R^*(\boldsymbol{\delta}_{\mathbf{\Lambda}})) = r_T^*$. It then follows from Theorem 8.2.1 that there exists a unique $c_{\mathbf{\Lambda}}(r_T^*)$ and, according to Theorem 8.1.1, a unique λ, since λ is decreasing in $c_{\mathbf{\Lambda}}$. Then according to (8.5), we must have $EV(\boldsymbol{\delta}_{\boldsymbol{T}}) \geq EV(\boldsymbol{\delta}_{\mathbf{\Lambda}})$ and $ES^*(\boldsymbol{\delta}_{\boldsymbol{T}}) \leq ES^*(\boldsymbol{\delta}_{\mathbf{\Lambda}})$. Otherwise, it contradicts the fact that $\boldsymbol{\delta}_{\mathbf{\Lambda}}$ minimizes the expected loss function (8.1) and that $E(R^*(\boldsymbol{\delta}_{\boldsymbol{T}})) = E(R^*(\boldsymbol{\delta}_{\mathbf{\Lambda}}))$.

Hence $\alpha \geq \dfrac{E(V(\boldsymbol{\delta}_{\boldsymbol{T}}))}{E(R^*(\boldsymbol{\delta}_{\boldsymbol{T}}))} \geq \dfrac{E(V(\boldsymbol{\delta}_{\mathbf{\Lambda}}))}{E(R^*(\boldsymbol{\delta}_{\mathbf{\Lambda}}))}$ and $\dfrac{E(T^*(\boldsymbol{\delta}_{\boldsymbol{T}}))}{E(A^*(\boldsymbol{\delta}_{\boldsymbol{T}}))} \geq \dfrac{E(T^*(\boldsymbol{\delta}_{\mathbf{\Lambda}}))}{E(A^*(\boldsymbol{\delta}_{\mathbf{\Lambda}}))}$. The multiple testing procedure based on $\mathbf{\Lambda}$ has the smallest mFNR* among the class $\mathcal{T}$.

8.3 The Oracle Procedure

Let $\mathbf{T_{OR}}(\mathbf{X}) = \left\{T_{OR}(X_i) = \frac{\pi_0 f(X_i)}{\pi_0 f(X_i) + \pi_1 \int \mu^2 f(X_i - \mu)h(\mu)d\mu}, i = 1, \ldots, m\right\}$. Assume $T_{OR}(X)|\theta = 1, \mu \sim G^{\mu}_{OR}$, $T_{OR}(X)|\theta = j \sim G^{j}_{OR}$ with $j = 0, 1$, $g^{\mu}_{OR}(t) = dG^{\mu}_{OR}(t)/dt$, and $g^{j}_{OR}(t) = dG^{j}_{OR}(t)/dt$. We first show that $\mathbf{T_{OR}}$ satisfies the wMLR condition. Consider the decision rule $\boldsymbol{\delta}(\mathbf{T_{OR}}, t) = I(\mathbf{T_{OR}} < t\mathbf{1})$. Then the risk of this rule is the following:

$$\frac{1}{m}\sum\{\lambda\pi_0 G^0_{OR}(t) + \pi_1 \int \mu^2(1 - G^{\mu}_{OR}(t))h(\mu)d\mu\}.$$

The optimal cutoff t^* that minimizes this risk must satisfy

$$\frac{\int \mu^2 g^{\mu}(t^*)h(\mu)d\mu}{g^0(t^*)} = \frac{\pi_0}{\pi_1}\lambda$$

By Theorem 8.1.1, we know that $t^* = \frac{1}{1+\lambda}$. Then we have,

$$\frac{\int \mu^2 g^{\mu}(t^*)h(\mu)d\mu}{g^0(t^*)} = \frac{\pi_0}{\pi_1}(\frac{1}{t^*} - 1)$$

Hence the ratio is decreasing in t^*, and $\mathbf{T_{OR}} \in \mathcal{T}$.

Theorem 8.3.1 (The oracle multiple testing procedure) *The oracle multiple testing procedure that minimizes mFNR* among all procedures in $\mathcal{T}$ that controls the mFDR* at α is $\boldsymbol{\delta}(\mathbf{T_{OR}}, \lambda_{OR}\mathbf{1}) = \{I(T_{OR}(X_i) < \lambda_{OR}), i = 1, \ldots, m\}$, where*

$$\lambda_{OR} = \sup\{t : \frac{\pi_0 G^0_{OR}(t)}{\pi_0 G^0_{OR}(t) + \pi_1 \int \mu^2 G^{\mu}_{OR}(t)h(\mu)d\mu} \leq \alpha\} \tag{8.6}$$

We compare our oracle procedure with the oracle procedure in Sun and Cai (2007). Specifically, we calculate the acceptance regions, mFNR, and mFNR* for both approaches under the following model, which was also used in Example 1 in Section 3.2 of Sun and Cai (2007). Specifically,

$$Z \sim \pi_0 N(0,1) + \pi_{11} N(\mu_1, 1) + \pi_{12} N(\mu_2, 1), \tag{8.7}$$

where $\pi_0 + \pi_{11} + \pi_{12} = 1$. Let $\pi_{11} + \pi_{12} = \pi_1$. This is a special form of our model with $h(\mu_1) = \dfrac{\pi_{11}}{\pi_1}$ and $h(\mu_2) = \dfrac{\pi_{12}}{\pi_1}$. Under this model, the oracle testing procedure $T_{\text{OR}} \leq \lambda_{\text{OR}}$ corresponds to $\{z : Z \leq c_l \quad \text{or} \quad Z \geq c_u\}$. Analogous to Sun and Cai (2007), we can find the rejection region in the following way. For a given t, solve the following equation for z to obtain c_l and c_u:

$$t[\pi_{11}\mu_1^2 exp(\mu_1 z - \frac{1}{2}\mu_1^2) + \pi_{12}\mu_2^2 exp(\mu_2 z - \frac{1}{2}\mu_2^2)] - \pi_0(1-t) = 0$$

Then calculate the mFDR* and mFNR* as

$$m\text{FDR}^* = \frac{\pi_0[\Phi(c_l) + \overline{\Phi}(c_u)]}{\pi_0[\Phi(c_l) + \overline{\Phi}(c_u)] + \pi_{11}\mu_1^2[\Phi(c_l - \mu_1) + \overline{\Phi}(c_u - \mu_1)] + \pi_{12}\mu_2^2[\Phi(c_l - \mu_2) + \overline{\Phi}(c_u - \mu_2)]}$$

$$m\text{FNR}^* = \frac{\pi_{11}\mu_1^2[\Phi(c_u - \mu_1) - \Phi(c_l - \mu_1)] + \pi_{12}\mu_2^2[\Phi(c_u - \mu_2) - \Phi(c_l - \mu_2)]}{\pi_0[\Phi(c_u) - \Phi(c_l)] + \pi_{11}\mu_1^2[\Phi(c_u - \mu_1) - \Phi(c_l - \mu_1)] + \pi_{12}\mu_2^2[\Phi(c_u - \mu_2) - \Phi(c_l - \mu_2)]}$$

Repeat the process until we find t such that mFDR* converges to α. As in Sun and Cai (2007), we chose $\pi_0 = 0.8$, $\mu_1 = -3$, $\mu_2 = 4$ and let π_{11} vary in $(0, 0.2)$. The result of this numerical study is shown in Figure 8.1 and Figure 8.2. In Figure 8.1, for different π_{11} values, we found, at level 0.05, the acceptance regions in terms of the z values for our procedure controlling the mFDR* and that for the procedure of Sun and Cai (2007) controlling the mFDR. It is clearly seen that the proposed approach results in much wider rejection regions. In Figure 8.2, we have further reported the values of the mFNR and the mFNR* of two approaches under the same setting. The proposed approach

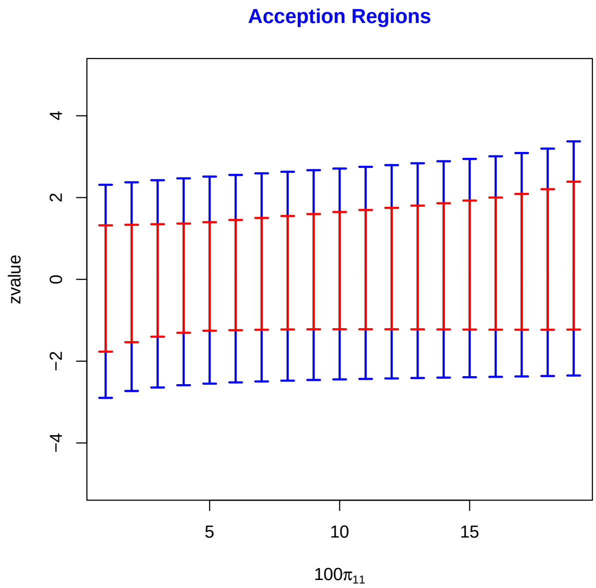


Figure 8.1: The acceptance regions: oracle procedure controlling mFDR at level 0.05 (blue) (Sun and Cai, 2007), oracle procedure controlling mFDR* at level 0.05 (red). The data are generated according to (8.7) with $\pi_0 = 0.8$, and π_{11} varying from 0 to 0.2. The parameters μ_1 and μ_2 are set to be -3 and 4 respectively.

results in a much smaller mFNR and mFNR*, for each value of π_{11}. For instance, the ratio of the mFNR* of the proposed approach to that of the oracle procedure in Sun and Cai (2007) can be as small as 0.15. It is thus demonstrated that the proposed approach is more powerful.

8.4 The Data Driven Procedure

Previously, we have given the oracle procedure and demonstrated its performance. The oracle procedure depends on unknown quantities, such as $\int \mu^2 f(x-\mu)h(\mu)d\mu$, which need to be estimated from the data. In Section 8.4.3, we provide consistent estimates for such unknown quantities and derive a data driven procedure as described in Section 8.4.1. In Section 8.4.2, we show that the

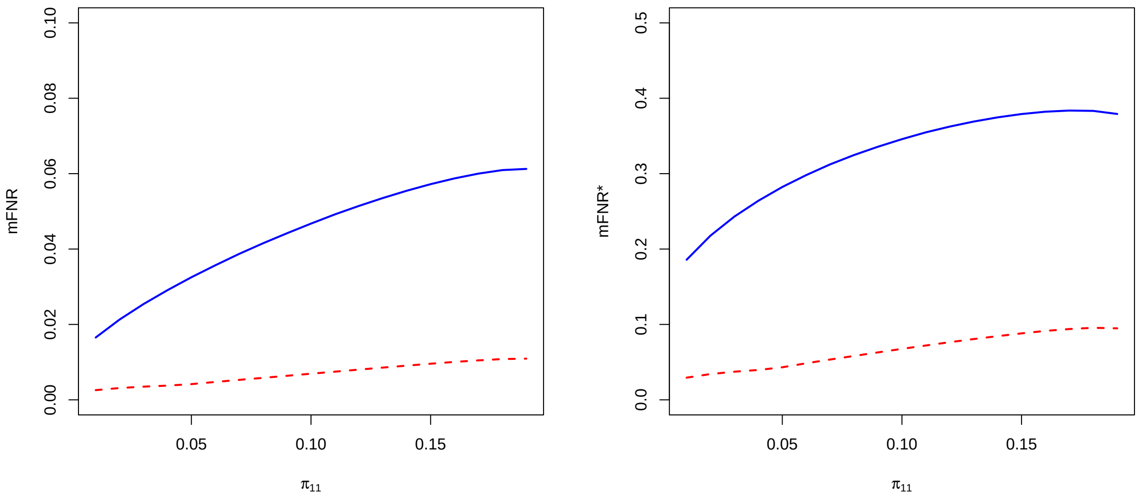


Figure 8.2: Comparisons of the two oracle procedures in terms of their mFNR and mFNR*: controlling mFDR at level 0.05 (blue solid), controlling mFDR* at level 0.05(red dashed). The setting is the same as described in the caption of Figure 8.1.

mFDR* and the mFNR* of our data driven procedure converge to those of the oracle procedure asymptotically.

8.4.1 The Procedure

First, for the oracle procedure $\boldsymbol{\delta}(\mathbf{T_{OR}}, \lambda_{OR}\mathbf{1}) = \{I(T_{OR}(X_i) < \lambda_{OR}), \quad i = 1, \ldots, m\}$, its mFDR* can be written as:

$$m\text{FDR*}(\lambda_{OR}) = Q^*(\lambda_{OR}) = \frac{\int I(T_{OR}(x) < \lambda_{OR}) T_{OR}(x)(\pi_0 f(x) + \pi_1 \int \mu^2 f(x-\mu)h(\mu)d\mu)dx}{\int I(T_{OR}(x) < \lambda_{OR})(\pi_0 f(x) + \pi_1 \int \mu^2 f(x-\mu)h(\mu)d\mu)dx}$$

Let $q(x) = \pi_0 f(x) + \pi_1 \int f(x-\mu)h(\mu)d\mu$ be the marginal distribution of x, and let $q^*(x) = \pi_0 f(x) + \pi_1 \int \mu^2 f(x-\mu)h(\mu)d\mu$. Suppose $\hat{q}$, $\hat{q}^*$ and $\hat{T}_{OR}$ are estimators of q, q^* and T_{OR}. Then an estimator of the mFDR* follows:

$$\hat{Q}^*(c) = \frac{\frac{1}{m}\sum I(\hat{T}_{OR}(x_i) < c)\hat{T}_{OR}(x_i)\frac{\hat{q}^*(x_i)}{\hat{q}(x_i)}}{\frac{1}{m}\sum I(\hat{T}_{OR}(x_i) < c)\frac{\hat{q}^*(x_i)}{\hat{q}(x_i)}}$$

Let $\hat{\lambda}_{OR} = \sup\{t : \hat{Q}^*(t) \leq \alpha\}$. It is equivalent to consider only the set of the discrete thresholds in the set of estimated estimators $\{\hat{T}_{OR}(x_i), 1 = 1, \ldots, m\}$. Then a data driven procedure can be given as the following.

Definition 8.4.1 (The Data Driven Procedure) *Define* $c_i = \frac{\hat{q}^*(x_i)}{\hat{q}(x_i)}$. *Let* $\hat{T}_{OR(1)}, \ldots, \hat{T}_{OR(m)}$ *be the ordered test statistics and* $c_{(1)}, \ldots, c_{(m)}$ *be the corresponding* c_i*'s. Let*

$$k = \max\left\{j : \frac{\sum_{i=1}^{j}\hat{T}_{OR(i)}c_{(i)}}{\sum_{i=1}^{j}c_{(i)}} \leq \alpha\right\} \tag{8.8}$$

Then reject all $H_{(i)}, i = 1, \ldots, k$.

If we had chosen the weights $w_i = 1, \forall i$, then q^* would reduce to q and our oracle and adaptive procedures would reduce to the oracle and adaptive procedures derived by Sun amd Cai (2007). By giving weights $w_i = \mu_i^2$ for the false nulls, we take into account the distribution of signal strength when ordering the hypotheses from the most significant to the least significant. Hence our threshold will also be different.

8.4.2 Asymptotic Properties

Theorem 8.4.1 (Asymptotic Validity) *Suppose* $X_1, \ldots, X_m$ *are generated according to our model given in Section 8.1.1. Assume both* $q(x)$ *and* $q^*(x)$ *are continuous and positive on* $\mathcal{R}$. *If* $\hat{\pi}_0 \to_p \pi_0$

and $E \parallel \hat{f} - f \parallel^2 \rightarrow 0$, $E \parallel \hat{q}^* - q^* \parallel^2 \rightarrow 0$, *then the* m*FDR* of the data driven procedure defined in (8.8) converges to* $Q^*(\lambda_{OR})$, *the* m*FDR* of the oracle procedure defined in* (8.6), *as* $m \rightarrow \infty$.

Theorem 8.4.2 (Asymptotic Optimality) *Assume* $X_1, \ldots, X_m$, q, q^*, $\hat{\pi}_0$, $\hat{f}$ *and* $\hat{q}^*$ *satisfy the same conditions as in Theorem 8.4.1. Let the* m*FNR* of the oracle procedure defined in (8.6) be* $\tilde{Q}^*(\lambda_{OR})$. *Then the* m*FNR* of the data driven procedure defined in (8.8) converges to* $\tilde{Q}^*(\lambda_{OR})$ *as* $m \rightarrow \infty$.

8.4.3 Estimates Under the Normal Assumption

Previously, we have shown that the mFDR* and mFNR* of the data driven procedure defined in Definition 8.4.1 converge respectively to those of the oracle procedure if there exists consistent estimators for the unknown quantities $q(x)$ and $q^*(x)$. Note that $q(x)$ is simply the marginal density and can be estimated by using the kernel density estimation. The estimation of $q^*(x)$ is more challenging because it involves $\int \mu^2 f(x-\mu)h(\mu)d\mu$. When assuming $f(x-\mu) = \phi(x-\mu)$, where ϕ is the density for the standard normal random variable, we can derive the following:

$$\begin{aligned}
&\int \mu^2 \phi(x-\mu)h(\mu)d\mu \\
&= \int [(\mu - x)^2 + 2x(\mu - x) + x^2]\phi(x-\mu)h(\mu)d\mu \\
&= \int ((\mu - x)^2 \phi(x-\mu)h(\mu)d\mu + 2x \int \phi'(x-\mu)h(\mu)d\mu + x^2 q_1(x) \\
&= \int [(\phi''(x-\mu) + \phi(x-\mu)]h(\mu)d\mu + 2x \int \phi'(x-\mu)h(\mu)d\mu + x^2 q_1(x) \\
&= q_1''(x) + 2xq_1'(x) + (x^2+1)q_1(x) \qquad (8.9)
\end{aligned}$$

Here $q_1(x) = \int f(x-\mu)h(\mu)d\mu$. It is thus sufficient to estimate $q_1(x)$ and its first and second derivative. Let $\hat{\pi}_0$ is an estimate of π_0. Let $\hat{q}(x)$, be the kernel density estimates of $q(x)$. According to Brown and Greenshtein (2009), we can easily get $\hat{q}'(x)$ and $\hat{q}''(x)$, estimators of the derivatives of the density. Since

$$q(x) = \pi_0 f_0(x) + \pi_1 q_1(x),$$

one can thus estimate $q_1(x)$, $q_1'(x)$, and $q_1''(x)$ by $\hat{q}_1(x) = (\hat{q}(x) - \hat{\pi}_0\phi(x))/(1-\hat{\pi}_0)$, $\hat{q}_1'(x) = (\hat{q}'(x) - \hat{\pi}_0\phi'(x))/(1-\hat{\pi}_0)$ and $\hat{q}_1''(x) = (\hat{q}''(x) - \hat{\pi}_0\phi''(x))/(1-\hat{\pi}_0)$, respectively. We further write $q_1^*(x) = \int \mu^2 f(x-\mu)h(\mu)d\mu$. According to (8.9), one can further estimate q_1^* and q^*. Specifically, we estimate q_1^* as $\hat{q}_1^*(x) = \hat{q}_1''(x) + 2x\hat{q}_1'(x) + (x^2+1)\hat{q}_1(x)$ and $\hat{q}^*(x) = \hat{\pi}_0\phi(x) + (1-\hat{\pi}_0)\hat{q}_1^*(x)$. Eventually, we derive an estimate of $T_{OR}(x)$ as $\hat{T}_{OR} = \dfrac{\hat{\pi}_0\phi(x)}{\hat{q}^*(x)}$.

The calculation (8.9) relies heavily on the normal assumption. In general applications, the observations x_i's may follow some other distribution, such as T_d under the null hypothesis. One can simply use the transformation $y_i = \Phi^{-1}(T_d(x_i))$ to get the normally distributed working observation y_i's.

In a series of influential papers including Efron et al (2001) and Efron (2008), the necessity of estimating the empirical null was emphasized, i.e., estimating μ_0, and σ_0^2, the parameters of the null distribution. We can also utilize such information by replacing $\phi(x)$ by $\phi(\dfrac{x-\mu_0}{\sigma_0})$ when reliable estimators are available (Jin and Cai, 2007; and Cai and Jin, 2010).

8.5 Simulation Study and Real Data Analysis

8.5.1 Data-driven Procedure

In this subsection, we perform two numerical studies to compare the performance of our data driven procedure with the data driven procedure in Sun and Cai (2007). In both studies, we let the observations be $X_i \sim N(\mu_i, 1)$ and let θ_i be indicator for $\mu_i = 0$, $i = 1, \dots, m$. Further, conditioned on $\theta_i = 0$, we have $\mu_i = 0$, and conditioned on $\theta_i = 1$, we have $\mu_i \sim h(\mu)$. In the first study, we let $h(\mu) = \phi(\mu - 0.9)$, and we consider testing $H_i : \mu_i = 0$ against $\mu_i > 0$, $i = 1, \dots, m$. In the second one, we let $h(\mu) = \pi_{11}\phi(\mu - 2) + (1 - \pi_{11})\phi(\mu + 3)$, and we consider testing $H_i : \mu_i = 0$ against $\mu_i \neq 0$, $i = 1, \dots, m$. Here ϕ is the density function of the standard normal random variable, and $\pi_{11} \in [0, 1]$. We fix the number hypotheses m to be 3000 and the number of repetitions n to be 3000 in both studies.

Simulation Study I

We let the proportion of true null hypotheses π_0 move in $\{0.1, 0.2, \dots, 0.9\}$. For each π_0, we generate m i.i.d. θ_i's from the Bernoulli$(1 - \pi_0)$ distribution. We then generate the mean vector $\boldsymbol{\mu}$ conditioned on the realized θ_i's; that is, if $\theta_i = 0$, then $\mu_i = 0$; if $\theta_i = 1, \mu_i$ is generated from the $N(2, 1)$ distribution. The observed X_i's are generated from the $N_m(\boldsymbol{\mu}, I)$ distribution, which is the multivariate normal distribution with the mean $\boldsymbol{\mu}$ and the covariance matrix the identity matrix. At level 0.05, we apply our data driven procedure, the data driven procedure in Sun and Cai (2007), and the oracle adaptive BH procedure defined in 4.2.1. We repeat the above process n times, and

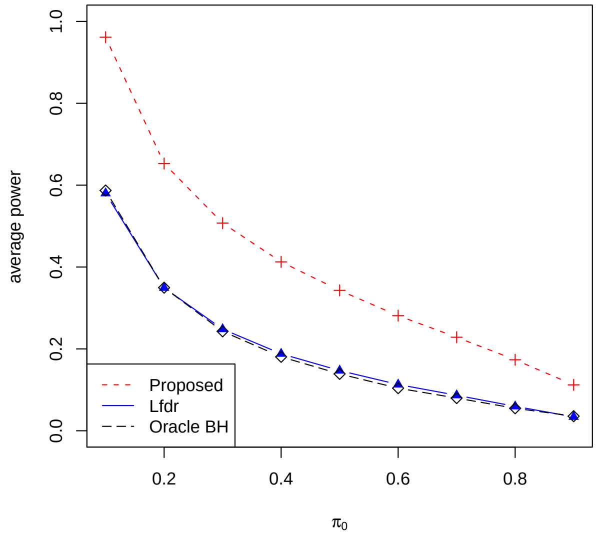


Figure 8.3: Comparison of the Data Driven Procedures in terms of Average Power: data driven procedure controlling mFDR at level 0.05 (blue solid) (Sun and Cai, 2007), the proposed data driven procedure controlling mFDR* at level 0.05 (red dashed), and the oracle adaptive BH procedure (long-dashed black).

calculate the average power, mFDR, mFNR, mFDR*, and mFNR* of the three procedures. Figure 8.3 shows the average power of the three procedures for the different π_0 values. We see that our procedure is more powerful for all values of π_0. Figure 8.4 shows the mFDR, mFDR*, mFNR, and mFNR* of the three procedures for the different π_0 values. We see that our procedure has mFDR* controlled at level 0.05 and has lowest mFNR and mFNR*.

Simulation Study II

We fix $\pi_0 = 0.5$ and let π_{11} move in $\{0.1, 0.2, \ldots, 0.9\}$, where π_{11} is the proportion of the alternative means which follow $N(2, 1)$ distribution. For each π_{11} value, we generate m i.i.d. θ_i's from the Bernoulli(0.5) distribution. If $\theta_i = 0$, then $\mu_i = 0$, and if $\theta_i = 1$, we generate γ_i from the

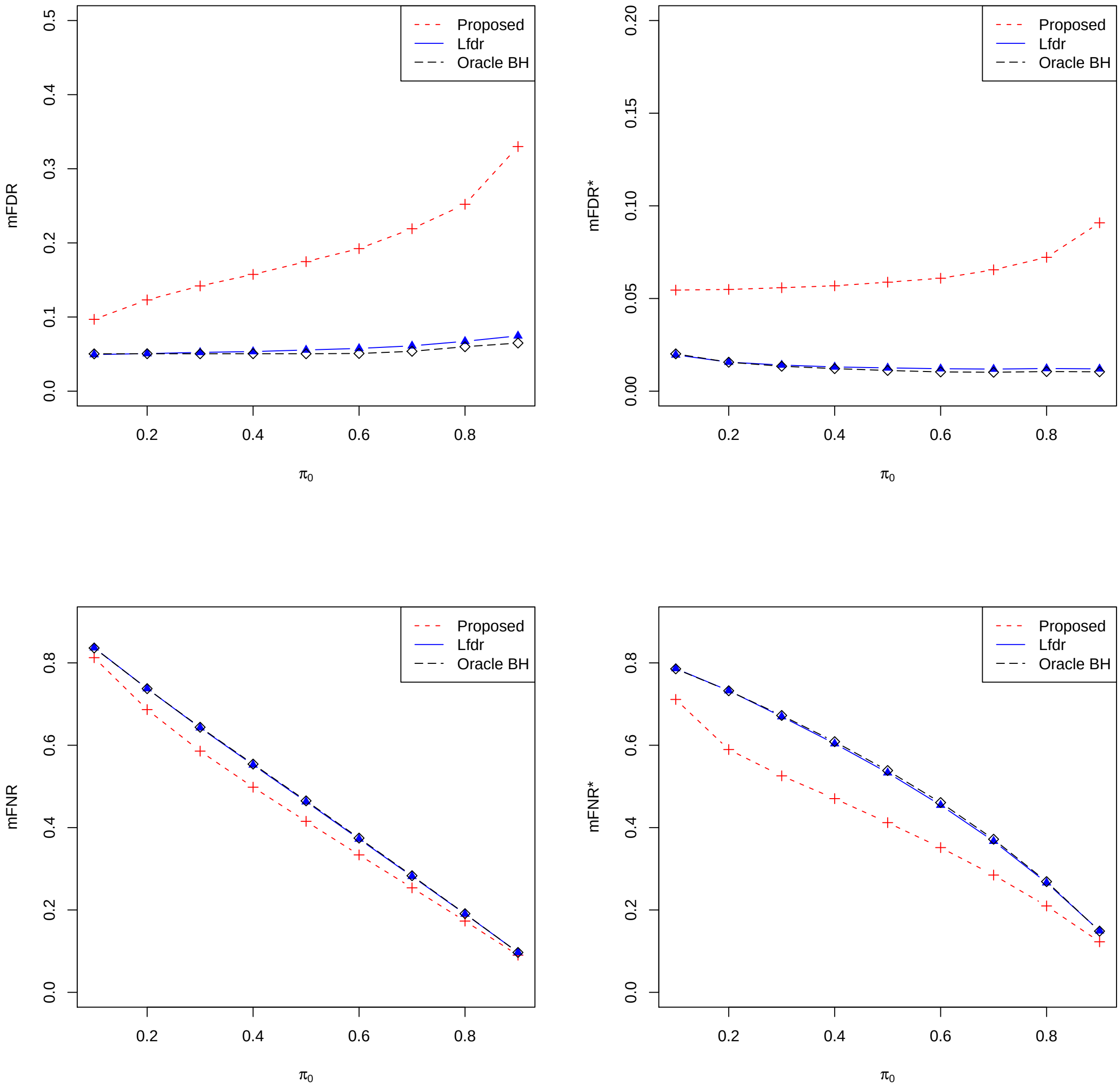


Figure 8.4: Comparison of the Data Driven Procedures: data driven procedure controlling mFDR at level 0.05 (blue solid) (Sun and Cai, 2007), the proposed data driven procedure controlling mFDR* at level 0.05 (red dashed), and the oracle adaptive BH procedure (long-dashed black).

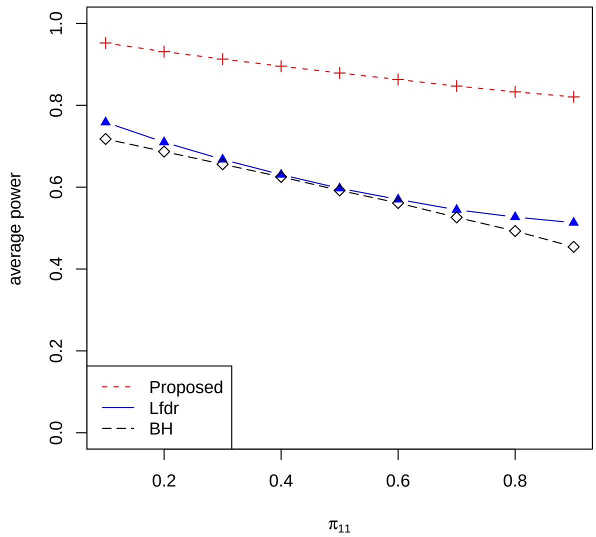


Figure 8.5: Comparison of the Data Driven Procedures in terms of Average Power: data driven procedure controlling mFDR at level 0.05 (blue solid) (Sun and Cai, 2007), the proposed data driven procedure controlling mFDR* at level 0.05 (red dashed), and the oracle BH procedure (long-dashed black).

Bernoulli(π_{11}) distribution. Conditioned on $\gamma_i = 1$, we generate μ_i from the $N(2, 1)$ distribution, and conditioned on $\gamma_i = 0$, μ_i is generated according to the $N(-3, 1)$ distribution. The observed X_i's are again generated from the $N_m(\boldsymbol{\mu}, I)$ distribution, conditioned on the realized mean vector $\boldsymbol{\mu}$. Figures 8.5 and 8.6 illustrate the results of the this simulation study. Figure 8.5 shows the average power of the three procedures for the different π_{11} values. We see that our procedure is more powerful for all values of π_{11}, and the procedure in Sun and Cai (2007) dominates the oracle adaptive BH procedure. Figure 8.6 shows the mFDR, mFDR*, mFNR, and mFNR* of the three procedures for the different π_0 values. We see that our procedure has mFDR* controlled at level 0.05 and has lowest mFNR and mFNR*, and the procedure in Sun and Cai (2007) has better performance than the oracle BH procedure.

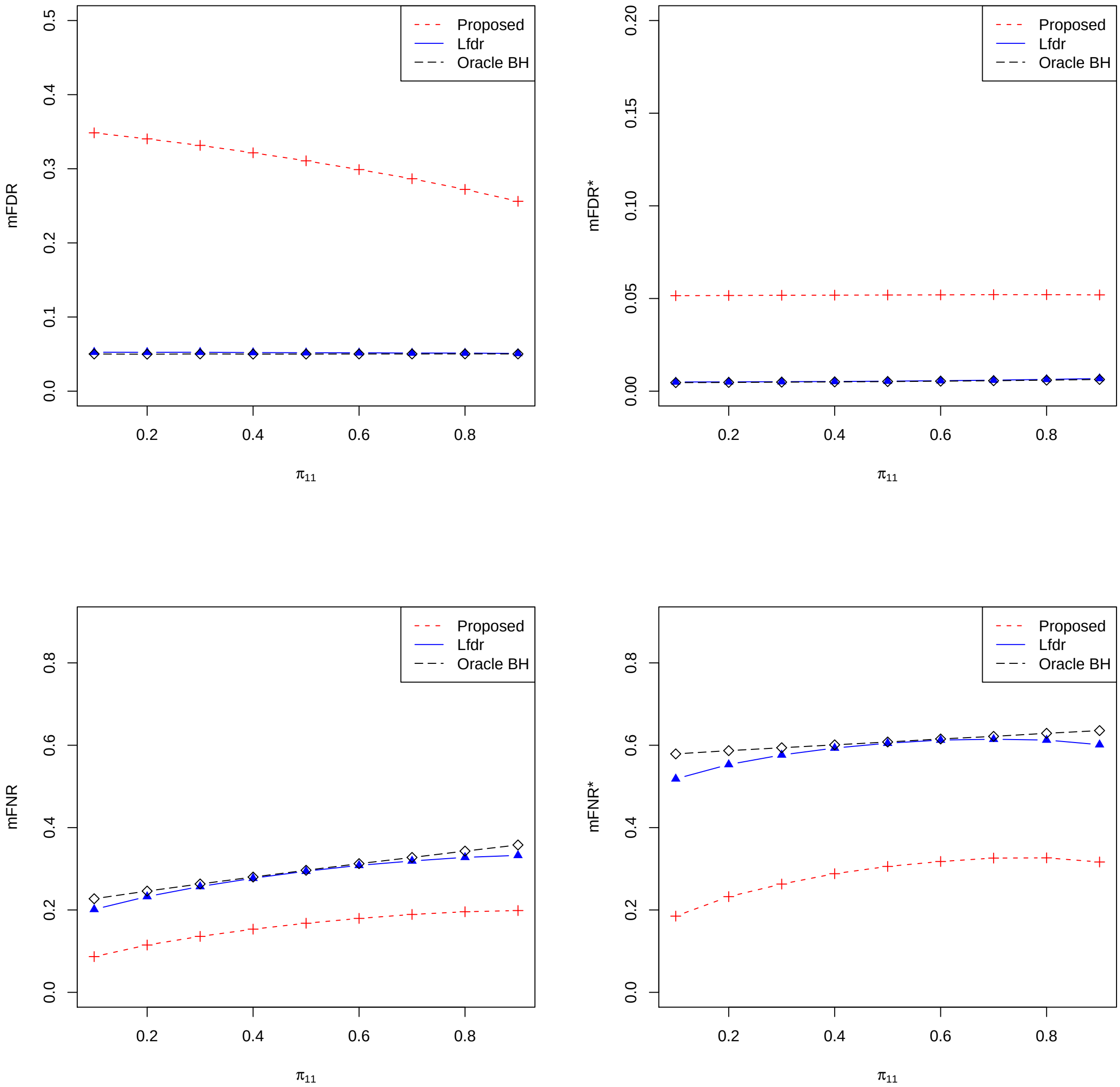


Figure 8.6: Comparison of the Data Driven Procedures: data driven procedure controlling mFDR at level 0.05 (blue solid) (Sun and Cai, 2007), the proposed data driven procedure controlling mFDR* at level 0.05 (red dashed), and the oracle BH procedure (long-dashed black).

8.5.2 Applications to a Real Data Set

We applied our proposed data driven procedure to the HIV data considered in Sun and Cai (2007) and Efron (2007). The HIV data compares 4 HIV positive patients versus 4 negative controls which provide microarray of expression levels on $m = 7680$ genes (Van't Wout et al. 2003). The two-sample t-statistic for each gene, which compares the expression levels of positive and negative HIV patients, is transformed to the z-value using the G_0 with 6 degrees of freedom. At level 0.05, we applied both our data driven procedure controlling the mFDR*, and the data driven procedures controlling the mFDR in Sun and Cai (2007). Using our data driven procedure to control the mFDR*, we found 73 significant genes, whereas 21 significant genes were found by using the method of Sun and Cai (2007) to control the mFDR.

8.6 Concluding Remarks

In this chapter, we proposed a compound decision framework for multiple testing under a new loss function (8.1) with the loss for a Type II error measured according to the corresponding signal strength. We found that the compound decision problem under the new loss function has a natural connection with the multiple testing procedure that minimizes the mFNR*, the weighted mFNR, while controlling the mFDR*, the weighted mFDR, with the set of weights indicating the strength of the signals. We derived an oracle procedure and then derived a data driven procedure that performs as good as the oracle procedure when the number of hypotheses under testing is very large. Both our simulation study and the real data analysis show that this new approach tends to be more powerful than the procedure of Sun and Cai (2007) and the BH procedure.

In this chapter of our research, we derived a new powerful testing procedure and a new error criteria, mFDR*, which seems to be less conservative than the traditional mFDR. It would be interesting to theoretically investigate its properties. Another extension of the current work would be the extension of it to the cases where the test statistics are dependent (see Sun and Cai, 2009; Efron, 2007). We leave these as future projects.

CHAPTER 9

FUTURE RESEARCH

9.1 The Compound Decision Framework

9.1.1 Incorporating Correlation Through a Hierarchy Model

Sun and Cai (2009) modeled dependence through an HMM model and formed a multiple testing procedure that accounts for dependence under this model. However, their assumption in this model may not be reasonable in some settings. Particularly, in genomics research, it may not be sensible to assume the genes are correlated like a Markov chain. We propose the following hierarchy model to incorporate dependence.

Definition 9.1.1 **[Hierarchy Model]** *Let $\theta_1, \ldots, \theta_m$ be the indicator variables that the hypotheses $H_1, \ldots, H_m$ are false, i.e. $\theta_i = 1$ if the hypothesis H_i is false and $\theta_i = 0$ otherwise. Let $X_i, i =$*

$1, \ldots, m$*, be the corresponding test statistics. Consider the following hierarchy model:*

$$\begin{aligned} P &\sim \textit{Beta}(\alpha, \beta) \\ \theta_i \mid P = p &\overset{\text{iid}}{\sim} \textit{Bernoulli}(p), \qquad i = 1, \ldots, m \\ X_i \mid \theta_i &\sim (1 - \theta_i)F_0 + \theta_i F_1, \qquad i = 1, \ldots, m \end{aligned}$$

Conditioned upon P, the components of $\boldsymbol{\theta}$ are assumed to be i.i.d. distributed as Bernoulli(p). It can be shown, unconditionally, that $\theta_1, \ldots, \theta_m$ are correlated Bernoulli random variables and, for any $i \neq j$

$$\begin{aligned} Cov(\theta_i, \theta_j) &= var(P) \geq 0 \qquad and \\ Corr(\theta_i, \theta_j) &= \frac{1}{\alpha + \beta + 1} \end{aligned}$$

Under the loss function defined in (8.1) and the above hierarchy model, the optimal classification rule that minimizes the risk is based on the test statistics $\boldsymbol{\Lambda} = (\Lambda_i(\boldsymbol{x}), i = 1, \ldots, m)$, where

$$\Lambda_i(\boldsymbol{x}) = \frac{P(\theta_i = 0 \mid \boldsymbol{x})}{P(\theta_i = 1 \mid \boldsymbol{x})}.$$

Under the compound decision theoretical framework, the optimal multiple testing procedure that controls mFDR at level α with the smallest mFNR among the collection of tests satisfying MRC is also based on $\boldsymbol{\Lambda}$.

Let the local index of the significance (LIS) be $\text{LIS}_i(\boldsymbol{x}) = P(\theta_i = 0 \mid \boldsymbol{x})$, as defined in Sun and Cai (2009). Consider the situation where the necessary parameter of the hierarchy model is known to us. Then the optimal oracle procedure that controls FDR at level α follows.

Theorem 9.1.1 *Let $LIS_{(1)}, \ldots, LIS_{(m)}$ be the ranked LIS values and $H_{(1)}, \ldots, H_{(m)}$ be the corresponding hypotheses. Let*

$$k = max\{i : \frac{1}{i}\sum_{j=1}^{i} LIS_{(j)}(\boldsymbol{x}) \leq \alpha\}$$

Then reject $H_{(i)}, i = 1, \ldots, k$.

Simulation studies are expected to be done to investigate the performance of the oracle procedure defined in 9.1.1. To make the procedure applicable in real data, the unknown parameters of the hierarchy model need to be estimated and a data-driven procedure that can mimic the performance of the oracle procedure is necessary.

9.1.2 Compound Decision Framework under the New Loss Function and Dependence

We aim to extend our result in Chapter 8 to the case where the test statistics are dependent. Specifically, we want to develop a compound decision framework under our new loss function and under a model that induces dependence. Under this framework, we wish to derive an optimal procedure that incorporate both the signal strength and the dependence information. In particular, our procedure would minimize the weighted mFNR subject to a constraint to the weighted mFDR which we hope will improve the procedure derived in Sun and Cai (2009).

9.2 Adaptive procedures using Efron's estimator of π_0

Efron (2007) advocated using the "empirical" null distribution for the multiple testing inference when correlation is present. The "empirical" null can be used when forming an estimate of π_0 which can in turn be used in forming an adaptive procedure. Efron estimated the number of true nulls in the same spirit as Storey (2002), except that he uses z-values rather than p-values and uses the empirical null $N(0, \sigma_0^2)$ for z-values instead of the theoretical null $N(0, 1)$ for z-values. In the unpublished manuscript of Efron (2007), Efron gave an estimator of π_0 as the following:

$$\hat{\pi}_0 = \frac{\hat{P}_0}{P_0(\hat{\sigma}_0)}$$

where

$$\hat{P}_0 = \sum_{i=1}^{m} (z_i \in [-x_0, x_0]) \tag{9.1}$$

$$P_0(\hat{\sigma}_0) = 2\Phi(x_0/\hat{\sigma}_0) - 1 \tag{9.2}$$

Here x_0 is a fixed threshold, $\hat{\sigma}_0$ is an estimate of σ_0 which is an adjustment for correlation, and $\Phi(\cdot)$ is the cumulative distribution function of $N(0, 1)$. The new adaptive BH procedure, based on the Efron's estimate of π_0, is a step-up procedure with the critical values $\alpha_i = \dfrac{i\alpha}{m\hat{\pi}_0}$. We want to investigate the performance of this adaptive procedure and compare it with some other adaptive procedures.

CHAPTER 10

PROOFS

Proof of Theorem 6.1.1

Proof The proof follows from Lemma 6.1.1.

$$
\begin{aligned}
P\{V \geq k\} &\leq \frac{E\left[V(V-1)\right]}{k(k-1)} \\
&= \frac{1}{k(k-1)} E\left[\sum_{i \in I_0} I(P_i < t)\left[\sum_{i \in I_0} I(P_i < t) - 1\right]\right] \\
&= \frac{1}{k(k-1)} \sum_{i \neq j \in I_0} P\{P_i \vee P_j < t\} \\
&= \frac{m_0(m_0-1)}{k(k-1)} F_2(t,t) \\
&= \frac{m_0(m_0-1)}{m(m-1)} \alpha \\
&\leq \alpha
\end{aligned}
$$

∎

Proof of Theorem 6.1.2

Proof Let j be the smallest (random) index such that $P_{(j)} = P_{k:m_0}$, where $P_{k:m_0}$ is the k-th smallest null p-value. Then

$$k \leq j \leq m - m_0 + k \quad \text{and} \quad \frac{k(k-1)}{(m+k-j)(m+k-j-1)} \leq \frac{k(k-1)}{m_0(m_0-1)}$$

This is because that the largest j occurs when the smallest $m - m_0$ p-values are nonnull p-values and the next k smallest p-values are null p-values. The stepdown procedure rejects at least k true null hypotheses $(V \geq k)$, if and only if

$$P_{(1)} \leq \alpha_1,\ P_{(2)} \leq \alpha_2,\ \ldots,\ P_{(j)} \leq \alpha_j$$

Then we have

$$\begin{aligned} P\{V \geq k\} &\leq P\{P_{(j)} \leq \alpha_j\} \\ &= P\{P_{k:m_0} \leq \alpha_j\} \\ &= P\left\{\sum_{i \in I_0} I(P_i < \alpha_j) \geq k)\right\} \\ &\leq \frac{m_0(m_0-1)F_2(\alpha_j, \alpha_j)}{k(k-1)} \\ &\leq \alpha \end{aligned}$$

∎

Proof of Theorem 6.1.3

Proof First note the following implication relations

$$k \leq V \leq m_0 \Rightarrow (V-k)(m_0-V) \geq 0 \Rightarrow V(m_0-V+k) \geq m_0 k \quad \text{and}$$

$$k \leq V \leq m_0 \Rightarrow (V-1)(m_0-V+k-1) \geq (m_0-1)(k-1)$$

Also note that $n - R \geq m_0 - V$. Although independence is a special case of Condition 2.7.7, we give a separate proof for this case, since it is particularly simple.

$$\begin{aligned}
&P\{V \geq k\} \\
&= \sum_{r=k}^{m} E\left[I(V \geq k, R=r)\right] \\
&\leq \sum_{r=k}^{m} E\left[\frac{(m-r+k)(m-r+k-1)V(V-1)}{m_0 k(m_0-1)(k-1)} I(V \geq k, R=r)\right] \\
&= \frac{1}{m_0(m_0-1)} \sum_{r=k}^{m} E\Bigg[\frac{(m-r+k)(m-r+k-1)}{k(k-1)} \sum_{i \neq j \in I_0} I\left(P_i \vee P_j \leq \alpha_r, V^{(-i,-j)} \geq k-2,\right. \\
&\quad \left. R^{(-i,-j)} = r-2\right)\Bigg] \\
&= \frac{1}{m_0(m_0-1)} \sum_{r=k}^{m} \frac{(m-r+k)(m-r+k-1)}{k(k-1)} \sum_{i \neq j \in I_0} F_2(\alpha_r, \alpha_r) P\left\{V^{(-i,-j)} \geq k-2, R^{(-i,-j)} = r-2\right\} \\
&= \frac{\alpha}{m_0(m_0-1)} \sum_{i \neq j \in I_0} \sum_{r=k}^{m} P\left\{V^{(-i,-j)} \geq k-2, R^{(-i,-j)} = r-2\right\} \\
&\leq \frac{\alpha}{m_0(m_0-1)} \sum_{i \neq j \in I_0} P\{R^{(-i,-j)} \geq k-2\} \\
&\leq \alpha
\end{aligned}$$

Here and subsequently, we use $R^{(-i,-j)}$ to denote the number of rejections based on $\{P_1, \ldots, P_n\}/\{P_i, P_j\}$ and critical constants $\{\alpha_3, \ldots, \alpha_m\}$, and $V^{(-i,-j)}$ to denote the number of false rejections among $R^{(-i,-j)}$.

In general, for dependent p-values satisfying Condition 2.7.7 and common pairwise distribution, we have the following proof:

$$
\begin{aligned}
P(V \geq k) &= \sum_{r=k}^{m} E\{I(V \geq k, R = r)\} \\
&\leq \frac{1}{m_0(m_0-1)} \sum_{r=k}^{m} E\left[\frac{(m-r+k)(m-r+k-1)}{k(k-1)} \sum_{i \neq j \in I_0} F_2(\alpha_r, \alpha_r)\right. \\
&\quad \left. P\left\{V^{(-i,-j)} \geq k-2,\ R^{(-i,-j)} = r-2 \mid P_i \vee P_j \leq \alpha_r\right\}\right] \\
&= \frac{\alpha}{m_0(m_0-1)} \sum_{i \neq j \in I_0} \sum_{r=k}^{m} P\left\{V^{(-i,-j)} \geq k-2,\ R^{(-i,-j)} = r-2 \mid P_i \vee P_j \leq \alpha_r\right\} \\
&\leq \frac{\alpha}{m_0(m_0-1)} \sum_{i \neq j \in I_0} \sum_{r=k}^{m} P\left\{R^{(-i,-j)} = r-2 \mid P_i \vee P_j \leq \alpha_r\right\} \\
&\leq \alpha
\end{aligned}
$$

The last inequality follows from the identity

$$
\sum_{r=k}^{n} P\left\{R^{(-i,-j)} = r-2 \mid P_i \vee P_j \leq \alpha_r\right\} \leq 1 \tag{10.1}
$$

which we prove below:

$$
\begin{aligned}
&\sum_{r=k}^{m} P\left\{R^{(-i,-j)} = r-2 \mid P_i \vee P_j \leq \alpha_r\right\} \\
&= \sum_{r=k}^{n} P\left\{R^{(-i,-j)} \geq r-2 \mid P_i \vee P_j \leq \alpha_r\right\} - \sum_{r=k}^{n-1} P\left\{R^{(-i,-j)} \geq r-1 \mid P_i \vee P_j \leq \alpha_r\right\} \\
&\leq \sum_{r=k}^{n} P\left\{R^{(-i,-j)} \geq r-2 \mid P_i \vee P_j \leq \alpha_r\right\} - \sum_{r=k}^{n-1} P\left\{R^{(-i,-j)} \geq r-1 \mid P_i \vee P_j \leq \alpha_{r+1}\right\} \\
&= P\left\{R^{(-i,-j)} \geq k-2 \mid P_i \vee P_j \leq \alpha_k\right\} \\
&\leq P\left\{R^{(-i,-j)} \geq 0 \mid P_i \vee P_j \leq \alpha_k\right\} \\
&= 1
\end{aligned}
$$

The first inequality follows from the Condition 2.7.7 and the fact that $I\left(R^{(-i,-j)} \geq r-1\right)$ is a coordinatewise decreasing function of $\{P_1,\ldots,P_n\}/\{P_i,P_j\}$. ∎

Proof of Theorem 6.2.2

Proof Assume $m_0 > 0$, otherwise there is nothing to prove. Let $\alpha_0 = 0$ and $\alpha_m = 1$. Then $\{(\alpha_{i-1},\alpha_i], i = 1,\ldots,m\}$ form a partition of the set $(0,1]$. Let S_i be the number of $\hat{Q}_i$'s, i.e., the number of true positives, in the interval $(\alpha_{i-1},\alpha_i]$. Define

$$J = \min\left\{l \in I : \, l - \sum_{i=1}^{l} S_i > l\gamma\right\} \tag{10.2}$$

i.e. J is the smallest random index where it is possible to have FDP $> \gamma$. Defined as such, $J = j$ is uniquely determined conditional on the set of observed $\hat{q}_1,\hat{q}_2,\ldots,\hat{q}_{m_1}$. write $k(j) = \lfloor \gamma j \rfloor + 1$.

We now argue, that given the set of observed $\hat{q}_1,\hat{q}_2,\ldots,\hat{q}_{n_1}$, a stepdown procedure defined in Theorem 6.2.2 will have the tail probability of FDP controlled below α. First note, as argued in Lehmann and Romano (2005), that by definition, j must satisfy:

$$j - \sum_{1}^{j} S_i = \lfloor \gamma j \rfloor + 1 = k(j)$$

i.e. at index j we must have $k(j)$ of true null p-values being less than α_j. Also note that

$$m_0 \leq m - \sum_{1}^{j} S_i = m - j + k(j) \tag{10.3}$$

Let V_j denote the number of true nulls p-values that are smaller than α_j, i.e. $V_j = \sum_{i \in I_0} (\hat{P}_i \leq \alpha_j)$. Then for a stepdown procedure,

$$P\{FDP > \gamma \mid \hat{q}_1,\hat{q}_2,\ldots,\hat{q}_{m_1}\} \leq P\left\{\text{at least } k(j) \text{ of } \hat{P}_i \leq \alpha_j \mid \hat{q}_1,\hat{q}_2,\ldots,\hat{q}_{m_1}\right\}$$

in the case where $k(j) = 1$, we have

$$\begin{aligned}
& P\{V_j \geq 1 \mid \hat{q}_1, \ldots, \hat{q}_{m_1}\} \\
& \leq E(V_j \mid \hat{q}_1, \ldots, \hat{q}_{m_1}) \\
& = m_0 P(\hat{P}_i \leq \alpha_j \mid \hat{q}_1, \hat{q}_2, \ldots, \hat{q}_{m_1}) \\
& \leq m_0 \frac{\alpha}{m - j + 1} \\
& \leq \alpha
\end{aligned} \tag{10.4}$$

On the other hand, if $k(j) \geq 2$, we have

$$\begin{aligned}
& P\{V_j \geq k(j) \mid \hat{q}_1, \ldots, \hat{q}_{m_1}\} \\
& \leq \frac{1}{k(j)(k(j)-1)} E(V_j(V_j - 1) \mid \hat{q}_1, \ldots, \hat{q}_{m_1}) \\
& = \frac{m_0(m_0-1)}{k(j)(k(j)-1)} P(\hat{P}_i \vee \hat{P}_j \leq \alpha_j \mid \hat{q}_1, \hat{q}_2, \ldots, \hat{q}_{m_1}) \\
& \leq \frac{m_0(m_0-1)}{k(j)(k(j)-1)} \frac{(\lfloor \gamma j \rfloor + 1)(\lfloor \gamma j \rfloor)}{(m - j + \lfloor \gamma j \rfloor + 1)(m - j + \lfloor \gamma i \rfloor)} \alpha \\
& \leq \frac{m_0(m_0-1)\alpha}{(m - j + \lfloor \gamma j \rfloor + 1)(m - j + \lfloor \gamma i \rfloor)} \\
& = \frac{m_0(m_0-1)\alpha}{(m - j + k(j))(m - j + k(j) - 1)} \\
& \leq \alpha
\end{aligned} \tag{10.5}$$

The last inequality in both 10.4 and 10.5 follows from 10.3. In any case, we always have

$$P(FDP > \gamma \mid \hat{q}_1, \hat{q}_2, \ldots, \hat{q}_{m_1}) \leq \alpha$$

which obviously implies $P(FDP > \gamma) \leq \alpha$. ∎

Proof of Theorem 6.3.1

Proof First note the following implication relation.

$$k \le V \le m_0 \Rightarrow (V-1)(m_0 - V + k - 1) \ge (m_0 - 1)(k-1)$$

Then we have

$$\begin{aligned}
& E\left[\frac{V}{m} I(V \ge k)\right] \\
& \le E\left[\frac{V}{m}\frac{(m-R+k-1)(V-1)}{(k-1)(m_0-1)} I(V \ge k)\right] \\
& \le \frac{1}{m(m_0-1)}\frac{m-r+k-1}{k-1}\sum_{i \ne j \in I_0}\sum_{r=k}^{m} Pr\left\{P_i \vee P_j \le \alpha_r, V^{(-i,-j)} \ge k-2, R^{(-i,-j)} = r-2\right\} \\
& \le \frac{1}{m(m_0-1)}\frac{m-r+k-1}{k-1}\sum_{i \ne j \in I_0}\sum_{r=k}^{m} Pr\left\{P_i \vee P_j \le \alpha_r, R^{(-i,-j)} = r-2\right\} \\
& \le \frac{1}{m(m_0-1)}\frac{m-r+k-1}{k-1}\sum_{i \ne j \in I_0}\sum_{r=k}^{m} Pr\left\{R^{(-i,-j)} = r-2 \mid P_i \vee P_j \le \alpha_r\right\} P\{P_i \vee P_j \le \alpha_r\} \\
& \le \frac{m_0(m_0-1)}{m(m_0-1)}\alpha \sum_{r=k}^{m} Pr\left\{R^{(-i,-j)} = r-2 \mid P_i \vee P_j \le \alpha_r\right\} \\
& \le \frac{m_0}{m}\alpha
\end{aligned}$$

The last inequality is due to Identity 10.1 ∎

Proof of Lemma 7.1.1

Proof

$$
\begin{aligned}
FDR &= E\left[\frac{V}{R}I(R>0)\right] \\
&= E\left[\frac{V}{R}I\left(R>0, G\left(\boldsymbol{P}\right)<\frac{1}{\pi_0}\right)\right] + E\left[\frac{V}{R}I\left(R>0, G_2\left(\boldsymbol{P}\right)>\frac{1}{\pi_0}\right)\right]
\end{aligned}
$$

The first part of the right hand side:

$$
\begin{aligned}
& E\left[\frac{V}{R}I\left(R>0, G\left(\boldsymbol{P}\right)<\frac{1}{\pi_0}\right)\right] \\
&= \sum_{r=1}^{m}\sum_{i\in I_0}\frac{1}{r}Pr\left\{P_i<\frac{r\alpha_1 G\left(\boldsymbol{P}\right)}{m}, R=r, G\left(\boldsymbol{P}\right)<\frac{1}{\pi_0}\right\} \\
&\leq \sum_{r=1}^{m}\sum_{i\in I_0}\frac{1}{r}Pr\left\{P_i<\frac{r\alpha_1}{m_0}, R=r\right\} \\
&\leq \frac{\alpha_1}{m_0}\sum_{r=1}^{m}\sum_{i\in I_0}Pr\left\{R=r \mid P_i<\frac{r\alpha_1}{m_0}\right\} \\
&\leq \alpha
\end{aligned}
$$

∎

Proof of Theorem 7.1.1

Proof From Lemma 7.1.1, we have, for the adaptive procedure defined in Theorem 7.1.1,

$$
FDR = \alpha_1 + E\left[\frac{V}{R}I\left(R>0, \frac{m-R_0+k-1}{m}>\frac{1}{\pi_0}\right)\right]
$$

for the second part of the right hand side of the inequality, we have

$$E\left[\frac{V}{R}I\left(R>0,\frac{m}{m-R_0+k-1}>\frac{1}{\pi_0}\right)\right] \leq E\left[\frac{m-R_0+k-1}{m}<\pi_0\right]$$
$$\leq P(R_0>m-m_0+k-1)$$
$$\leq P(V\geq k)$$
$$\leq \alpha_0$$

∎

Proof of Theorem 7.1.2

Proof First note that it is obvious if we reject each H_i when $P_i \leq \alpha_0$, then

$$E\left[\frac{V_0}{m}\right] = \frac{1}{m}\sum_{i\in I_0} Pr(P_i\leq\alpha_0) \leq \pi_0\alpha_0$$

Hence

$$\pi_0\alpha \geq E\left[\frac{V_0}{m}\right] \geq E\left[\frac{R_0-m_1}{m}\right]$$

Again, by Lemma 7.1.1, we have

$$FDR = \alpha_1 + E\left[\frac{V}{R}I\left(R>0, F_k\left(\frac{R_0}{m}\right)>\frac{1}{\pi_0}\right)\right]$$

Let the generalized inverse of $F_k(x)$ defined in Theorem 7.1.2 to be $F_k^{-1}(y) = \inf\{x \mid F_k(x) > y\}$. Since the function is nondecreasing, continuous function of x, $F_k(x) > y$ implies $x > F_k^{-1}(y)$.

Then we have,

$$
\begin{aligned}
E(I\left[F_k\left(\frac{R_0}{m}\right) > \frac{1}{\pi_0}\right]) &\leq P\left(\frac{R_0}{m} \geq F_k^{-1}\left(\frac{1}{\pi_0}\right)\right) \\
&= P\left(\frac{R_0}{m} - \pi_1 \geq F_k^{-1}\left(\frac{1}{\pi_0}\right) - \pi_1\right) \\
&\leq \frac{E\left[\frac{R_0}{m} - \pi_1\right]}{F_k^{-1}\left(\frac{1}{\pi_0}\right) - \pi_1} \\
&\leq \frac{\pi_0\alpha_0}{F_k^{-1}\left(\frac{1}{\pi_0}\right) - \pi_1} \\
&\leq k\alpha_0
\end{aligned}
$$

$\blacksquare$

Proof of Theorem 7.1.3

Proof Consider first step controlling $E\left[\frac{V}{m}(V \geq k_1)\right]$ using the procedure defined in Theorem 7.1.3. Then we have

$$\pi_0\alpha \geq E\left[\frac{V}{m}I\left(V \geq k_1\right)\right]$$

Let the generalized inverse of $F_{k_2}(x)$ defined in Theorem 7.1.3 to be $F_{k_2}^{-1}(y) = \inf\{x \mid F_{k_2}(x) > y\}$.

$$\begin{aligned}
P\left\{\frac{R_0}{m} \geq F_{k_2}^{-1}\left(\frac{1}{\pi_0}\right)\right\} &= P\left\{\frac{R_0}{m} - \pi_1 \geq F_{k_2}^{-1}\left(\frac{1}{\pi_0}\right) - \pi_1\right\} \\
&= P\left\{\left(\frac{R_0}{m} - \pi_1\right) I\left(V \geq k_1\right) + \left(\frac{R_0}{m} - \pi_1\right) I\left(V < k_1\right) \geq F_{k_2}^{-1}\left(\frac{1}{\pi_0}\right) - \pi_1\right\} \\
&\leq P\left\{\left(\frac{R_0}{m} - \pi_1\right) I\left(V \geq k_1\right) + \frac{k_1}{m} \geq F_{k_2}^{-1}\left(\frac{1}{\pi_0}\right) - \pi_1\right\} \\
&\leq \frac{E\left[\left(\frac{R_0}{m} - m_1\right) I\left(V \geq k_1\right)\right]}{F_{k_2}^{-1}\left(\frac{1}{\pi_0}\right) - \pi_1 - \frac{k_1}{m}} \\
&\leq \frac{E\left[\left(\frac{V}{m} I\left(V \geq k_1\right)\right)\right]}{F_{k_2}^{-1}\left(\frac{1}{\pi_0}\right) - \pi_1 - \frac{k_1}{m}} \\
&\leq k_2 \alpha_0
\end{aligned}$$

Hence we have

$$\begin{aligned}
FDR &\leq \alpha_1 + P\left\{\frac{R_0}{m} \geq F_{k_2}^{-1}\left(\frac{1}{\pi_0}\right)\right\} \\
&\leq \alpha_1 + k_2 \alpha_0
\end{aligned}$$

∎

Proof for Lemma 7.2.1

Proof

$$\begin{aligned} FWER &= E\left[I\left(V \geq 1\right)\right] \\ &= E\left[I\left(V \geq 1, G(\boldsymbol{P}) \leq \frac{1}{\pi_0}\right)\right] + E\left[I\left(V \geq 1, G(\boldsymbol{P}) > \frac{1}{\pi_0}\right)\right] \end{aligned} \tag{10.6}$$

The first part of the right hand side:

$$\begin{aligned} E\left[I\left(V \geq 1, G(\boldsymbol{P}) \leq \frac{1}{\pi_0}\right)\right] &\leq \sum_{i \in I_0} E\left[P_i < \frac{\alpha_1 G(\boldsymbol{P})}{m}, G(\boldsymbol{P}) \leq \frac{1}{\pi_0}\right] \\ &\leq \sum_{i \in I_0} Pr\left\{P_i < \frac{\alpha_1}{m_0}\right\} \\ &\leq \alpha_1 \end{aligned}$$

∎

Lemma 10.0.1 *Assume* $q(x) = \pi_0 f(x) + \pi_1 \int f(x-\mu)h(\mu)d\mu$ *is positive and continuous on* $\mathcal{R}$ *and* $q(x) = \pi_0 f(x) + \pi_1 \int \mu^2 f(x-\mu)h(\mu)d\mu$ *is continuous on* $\mathcal{R}$*. Let* $T = \frac{\pi_0 f}{q^*}$ *and* $\hat{T} = \frac{\hat{\pi}_0 \hat{f}}{\hat{q}^*}$*. If* $\hat{\pi}_0 \rightarrow_p \pi_0$*,* $E \parallel \hat{f} - f \parallel^2 \rightarrow 0$ *and* $E \parallel \hat{q}^* - q^* \parallel^2 \rightarrow 0$*, then* $E \parallel \hat{T} - T \parallel^2 \rightarrow 0$*.*

Proof

1. First It can be shown that $f, \hat{f}, q^*, \hat{q}^*$ are all bounded except for an event with small probability.

Here q is the marginal density function of X and q is continuous and positive. Hence $P(|X| \geq K) \rightarrow 0$, as $K \rightarrow \infty$. Choose K_1 large and let $I = [-K_1, K_1]$. Let $l = \inf_{x \in I} q^*(x)$ and $u = \sup_{x \in I} q^*(x)$. Let $A_\epsilon^1 = \{x : |q^* - \hat{q}^*| \geq l/2\}$. Then $(l/2)^2 P(A_\epsilon^1) \leq E \parallel q^* - \hat{q}^* \parallel^2$. By the assumption that $E \parallel \hat{q}^* - q^* \parallel^2 \to 0$, we have $P(A^1_\epsilon) \to 0$, as the number of hypotheses $m \to \infty$. Note that $l/2 \leq \hat{q}^* \leq l/2 + u$ for $x \notin A^1_\epsilon$. Similarly, $f, \hat{f}$ are bounded except in a set with small probability. Let A_ϵ be the set such that $q^*, q, f, \hat{f}$ are all bounded for $x \in A^c_\epsilon$. Since we also have $E \parallel \hat{f} - f \parallel^2 \to 0$, then $P(A_\epsilon) \to 0$, as $m \to \infty$.

2. Note $T - \hat{T} = \dfrac{\hat{f} q^*(\pi_0 - \hat{\pi}_0) + (1 - \pi_0) q^*(\hat{f} - f) + (1 - p) f (q^* - \hat{q}^*)}{\hat{q}^* q^*}$. We also know that $f, \hat{f}, q^*, \hat{q}^*$ are all bounded in A^c_ϵ. This implies that

$$(T - \hat{T})^2 \leq c_1(\pi_0 - \hat{\pi}_0)^2 + c_2(\hat{f} - f)^2 + c_3(q^* - \hat{q}^*)^2$$

in A^c_ϵ. Also, it can be seen that $E \parallel T - \hat{T} \parallel^2$ is bounded above, say, by L. Then

$$E \parallel T - \hat{T} \parallel^2 \leq LP(A_\epsilon) + (c_1 E(\pi_0 - \hat{\pi}_0)^2 + c_2 E \parallel \hat{f} - f \parallel^2 + c_3 E \parallel q^* - \hat{q}^* \parallel^2) P(A^c_\epsilon)$$

Since $\hat{\pi}_0 \to_p \pi_0$ implies $E(\hat{\pi}_0 - \pi_0)^2 \to 0$, then by the assumption that $E \parallel \hat{f} - f \parallel^2 \to 0$ and $E \parallel \hat{q}^* - q^* \parallel^2 \to 0$, we have $E \parallel T - \hat{T} \parallel^2 \longrightarrow 0$. ∎

We assume the assumptions in Lemma 10.0.1 hold in Lemma 10.0.2 to Lemma 10.0.6

Lemma 10.0.2 $E \parallel \hat{T} - T \parallel^2 \to 0$ *implies that* $\hat{T}(X) \to_p T(X)$. *Similarly,* $E \parallel \hat{q} - q \parallel^2 \to 0$ *implies that* $\hat{q}(X) \to_p q(X)$, *and* $E \parallel \hat{q}^* - q^* \parallel^2 \to 0$ *implies that* $\hat{q}^*(X) \to_p q^*(X)$.

Proof

$\epsilon^2 P(|\hat{T} - T| \geq \epsilon) \leq E \parallel T - \hat{T} \parallel^2$. Since $\parallel T - \hat{T} \parallel^2 \to 0$, we have $P(|\hat{T} - T| \geq \epsilon) \to 0$. The rest of the proof follows. ∎

Lemma 10.0.3 *Suppose* $\hat{T}(X) \rightarrow_p T(X)$*,* $\hat{q}(X) \rightarrow_p q(X)$*, and* $\hat{q}^*(X) \rightarrow_p q^*(X)$*. Then we have the following:*

$$(1).\ E\left(I(\hat{T}(X) < c)\hat{T}(X)\frac{\hat{q}^*(X)}{\hat{q}(X)}\right) \rightarrow E\left(I(T(X) < c)T(X)\frac{q^*(X)}{q(X)}\right);$$
$$(2).\ E\left(I(\hat{T}(X) < c)\frac{\hat{q}^*(X)}{\hat{q}(X) \vee d}\right) \rightarrow E\left(I(T(X) < c)\frac{q^*(X)}{q(X) \vee d}\right), \forall d > 0.$$

Proof

(1). First we note the following: $\hat{T}(X) \rightarrow_p T(X)$ implies that $I(\hat{T}(X) < c) \rightarrow_p I(T(X) < c)$; $\hat{q}(X) \rightarrow_p q(X)$ implies that $\frac{1}{\hat{q}(X)} \rightarrow_p \frac{1}{q(X)}$, since $q(\cdot) > 0$; Also, $\hat{q}^*(X) \rightarrow_p q^*(X)$. Hence we have

$$I(\hat{T}(X) < c)\hat{T}(X)\frac{\hat{q}^*(X)}{\hat{q}(X)} \rightarrow_p I(T(X) < c)T(X)\frac{q^*(X)}{q(X)};$$

which implies that

$$I(\hat{T}(X) < c)\hat{T}(X)\frac{\hat{q}^*(X)}{\hat{q}(X)} \rightarrow_d I(T(X) < c)T(X)\frac{q^*(X)}{q(X)}.$$

Also note that $I(\hat{T}(X) < c)\hat{T}(X)\frac{\hat{q}^*(X)}{\hat{q}(X)} = I(\hat{T}(X) < c)\frac{\hat{\pi}_0 f}{\hat{q}}$, which is bounded by 1. Then by the Portmanteau theorem,

$$E\left(I(\hat{T}(X) < c)\hat{T}(X)\frac{\hat{q}^*(X)}{\hat{q}(X)}\right) \rightarrow E\left(I(T(X) < c)T(X)\frac{q^*(X)}{q(X)}\right) \tag{10.7}$$

(2). $\hat{q}(X) \rightarrow_p q(X)$ implies that $\hat{q}(X) \vee d \rightarrow_p q(X) \vee d$, for any $d > 0$.

$$I(\hat{T}(X) < c)\frac{\hat{q}^*(X)}{\hat{q}(X) \vee d} \rightarrow_p I(T(X) < c)\frac{q^*(X)}{q(X) \vee d}, \forall d.$$

Since $E||\hat{q}^* - q^*||^2 \to 0$, we have $E||\hat{q}^*||^2 \leq E||q^*||^2 + C$ where C is some constant for sufficiently large m. Further, one knows that $P(|\hat{q}^* - q^*| > \epsilon) \leq \frac{1}{\epsilon^2} E||\hat{q}^* - q^*||^2$. Let $A_{n,\epsilon} = \{|\hat{q}^* - q^*| > \epsilon\}$, then

$$EI(\hat{T}(x) < c)\frac{\hat{q}^*(x)}{\hat{q}(x) \vee d} \leq E1_{A_{n,\epsilon}} I(\hat{T}(x) < c)\frac{q^*(x) + \epsilon}{\hat{q}(x) \vee d} + E1_{A_{n,\epsilon}} \frac{1}{d}\hat{q}^*(x),$$

with

$$E1_{A_{n,\epsilon}} \frac{1}{d}\hat{q}^*(x) \leq \frac{1}{d} E1_{A_{n,\epsilon}} E|\hat{q}^*(x)|^2 \leq \frac{1}{d}(E|q^*(x)|^2 + C)E1_{A_{n,\epsilon}},$$

which goes to zero as $m \to \infty$.

Further, $E1_{A^c_{n,\epsilon}} I(\hat{T}(x) < c)\frac{q^*(x) + \epsilon}{\hat{q}(x) \vee d} \to EI(T(x) < c)\frac{q^*(x) + \epsilon}{q(x) \vee d}$ because the function $1_{A^c_{n,\epsilon}} \frac{q^*(x) + \epsilon}{\hat{q}(x) \vee d}$ is bounded. Let $\epsilon \to 0$, then we know that

$$\limsup EI(\hat{T}(x) < c)\frac{\hat{q}^*(x)}{\hat{q}(x) \vee d} \leq EI(T(x) < c)\frac{q^*(x)}{q(x) \vee d}.$$

One can similarly show that

$$\liminf EI(\hat{T}(x) < c)\frac{\hat{q}^*(x)}{\hat{q}(x) \vee d} \geq EI(T(x) < c)\frac{q^*(x)}{q(x) \vee d}.$$

Consequently,

$$E\left(I(\hat{T}(X) < c)\frac{\hat{q}^*(X)}{\hat{q}(X) \vee d}\right) \to E\left(I(T(X) < c)\frac{q^*(X)}{q(X) \vee d}\right), \forall d. \tag{10.8}$$

∎

Lemma 10.0.4 *Write* $G^{0*}(c) = E\left(I(T(X) < c)T(X)\frac{q^*(X)}{q(X)}\right)$, $G^*_d(c) = E\left(I(T(X) < c)\frac{q^*(X)}{q(X) \vee d}\right)$ *and* $Q^*_d(c) = \frac{G^{0*}(c)}{G^*_d(c)}$. *Let* $\hat{G}^{0*}(c) = \frac{1}{m}\sum I(\hat{T}(X_i) < c)\hat{T}(X_i)\frac{\hat{q}^*(X_i)}{\hat{q}(X_i)}$, $\hat{G}^*_d(c) = \frac{1}{m}\sum I(\hat{T}(X_i) < c)\frac{\hat{q}^*(X_i)}{\hat{q}(X_i) \vee d}$, *and* $\hat{Q}^*_d(c) = \frac{\hat{G}^{0*}(c)}{\hat{G}^*_d(c)}$. *Then* $\hat{Q}^*_d(c) \to_p Q^*_d(c)$.

Proof

1. Let $s_m = \sum_{i=1}^{m} I(\hat{T}(X_i) < c)\hat{T}(X_i)\frac{\hat{q}^*(X_i)}{\hat{q}(X_i)}$. First note that the following is true.

$$I(\hat{T}(X_i) < c)\hat{T}(X_i)\frac{\hat{q}^*(X_i)}{\hat{q}(X_i)}I(\hat{T}(X_j) < c)\hat{T}(X_j)\frac{\hat{q}^*(X_j)}{\hat{q}(X_j)}$$
$$\to_d I(T(X_i) < c)T(X_i)\frac{q^*(X_i)}{q(X_i)}I(T(X_j) < c)T(X_j)\frac{q^*(X_j)}{q(X_j)}$$

Also, we know that $I(\hat{T}(X_i) < c)\hat{T}(X_i)\frac{\hat{q}^*(X_i)}{\hat{q}(X_i)}I(\hat{T}(X_j) < c)\hat{T}(X_j)\frac{\hat{q}^*(X_j)}{\hat{q}(X_j)}$ is bounded. Hence, we have

$$E\left(I(\hat{T}(X_i) < c)T(X_i)\frac{\hat{q}^*(X_i)}{\hat{q}(X_i)}I(\hat{T}(X_j) < c)T(X_j)\frac{\hat{q}^*(X_j)}{\hat{q}(X_j)}\right)$$
$$\to E\left(I(T(X_i) < c)T(X_i)\frac{q^*(X_i)}{q(X_i)}I(T(X_j) < c)T(X_j)\frac{q^*(X_j)}{q(X_j)}\right) \quad \text{and}$$

$$\rho_m = cov\left(I(\hat{T}(X_i) < c)\hat{T}(X_i)\frac{\hat{q}^*(X_i)}{\hat{q}(X_i)}, I(\hat{T}(X_j) < c)\hat{T}(X_j)\frac{\hat{q}^*(X_j)}{\hat{q}(X_j)}\right)$$
$$\to cov\left(I(T(X_i) < c)T(X_i)\frac{q^*(X_i)}{q(X_i)}, I(T(X_j) < c)T(X_i)\frac{q^*(X_j)}{q(X_j)}\right) = 0$$

2. Also, we have the following:

$$Var\left(I(\hat{T}(X) < c)\hat{T}(X)\frac{\hat{q}^*(X)}{\hat{q}(X)}\right) \le E\left(I(\hat{T}(X) < c)\hat{T}(X)\frac{\hat{q}^*(X)}{\hat{q}(X)}\right)^2$$
$$E\left(I(\hat{T}(X) < c)\hat{T}(X)\frac{\hat{q}^*(X)}{\hat{q}(X)}\right)^2 \to E\left(I(T(X) < c)T(X)\frac{q^*(X)}{q(X)}\right)^2 < \infty.$$

Hence $var(s_m/m) \to 0$.

3. By the weak law of large numbers, $\frac{1}{m}\sum I(\hat{T}(X) < c)\hat{T}(X)\frac{\hat{q}^*(X)}{\hat{q}(X)} \to_p E\left(I(\hat{T}(X) < c)\hat{T}(X)\frac{\hat{q}^*(X)}{\hat{q}(X)}\right)$.

4. By (10.7),

$$\frac{1}{m}\sum I(\hat{T}(x_i) < c)\hat{T}(x_i)\frac{\hat{q}^*(x_i)}{\hat{q}(x_i)} \to_p G^{0*}(c).$$

Similarly,

$$\frac{1}{m}\sum I(\hat{T}(x_i) < c)(x_i)\frac{\hat{q}^*(x_i)}{\hat{q}(x_i)} \to_p G_d^*(c).$$

Hence $\hat{Q}_d^*(c) = \dfrac{\hat{G}^{0*}(c)}{\hat{G}_d^*(c)} \to_p Q_d^*(c)$. ∎

Lemma 10.0.5 *Let $G^{0*}(c)$, $\hat{G}^{0*}(c)$, $\hat{Q}_d^*(c)$, and $Q_d^*(c)$ be defined as in Lemma 10.0.4. Define $G^*(c) = E\left(I(T(X) < c)\dfrac{q^*(X)}{q(X)}\right)$, $Q^*(c) = \dfrac{G^{0*}(c)}{G_d^*(c)}$, $\hat{G}^*(c) = \dfrac{1}{m}\sum I(\hat{T}(X_i) < c)\dfrac{\hat{q}^*(X_i)}{\hat{q}(X_i)}$, and $\hat{Q}^*(c) = \dfrac{\hat{G}^{0*}(c)}{\hat{G}^*(c)}$. The we have that $\hat{Q}_d^*(c) \to_p Q_d^*(c)$ implies $\hat{Q}^*(c) \to_p Q^*(c)$.*

Proof

First note that we have the following:

$$\begin{aligned} |\hat{Q}^*(c) - Q^*(c)| &\leq |\hat{Q}^*(c) - \hat{Q}_d^*(c)| + |\hat{Q}_d^*(c) - Q_d^*(c)| + |Q_d^*(c) - Q^*(c)| \\ &= \hat{Q}_d^*(c) - \hat{Q}^*(c) + |\hat{Q}_d^*(c) - Q_d^*(c)| + |Q_d^*(c) - Q^*(c)|. \end{aligned}$$

Since $Q_d^*(c)$ is a increasing function with respect to d, then $Q_d^*(c) \to Q^*(c)$ as $d \to 0$. Also we assume that $\hat{Q}_d^*(c) \to_p Q_d^*(c)$. Then for any $\epsilon_1, \epsilon_2 > 0$, there exists $d_0 > 0$ such that $\forall d < d_0$,

$$0 < Q_d^*(c) - Q^*(c) < \epsilon_1.$$

and

$$P(|\hat{Q}_d^*(c) - Q_d^*(c)| < \epsilon_1) > 1 - \epsilon_2.$$

Further, for any ω from the underlying probability space, $\hat{Q}^*_d(c)$ is increasing with respect to d. Consequently,

$$\lim_{d\to 0}\hat{Q}^*_d(c) = \hat{Q}^*(c).$$

In other words, $\hat{Q}^*_d(c) \to \hat{Q}^*(c)$ pointwise, which implies that $\hat{Q}^*_d(c) \to_p \hat{Q}^*(c)$. Consequently, for any ϵ_1, and ϵ_2, there exists d_1, such that $for all d < d_1$,

$$P(|\hat{Q}^*_d(c) - \hat{Q}^*(c)| < \epsilon_1) > 1 - \epsilon_2.$$

All together, we know that for any $\epsilon_1, \epsilon_2 > 0$, there exists a $D = \min(d_0, d_1)$, such that $\forall d < D$,

$$\begin{aligned} & P(|\hat{Q}^*(c) - Q^*(c)| < 3\epsilon_1) \\ > \; & P(|\hat{Q}^*_d(c) - \hat{Q}^*(c)| < \epsilon_1, |\hat{Q}^*_d(c) - Q^*_d(c)| < \epsilon_1) \\ \geq \; & 1 - 2\epsilon_2. \end{aligned}$$

We therefore conclude that $\hat{Q}^*(c) \to Q^*(c)$ in probability. ∎

Lemma 10.0.6 *Assume that $\hat{Q}^*(c)$ and $Q^*(c)$ are as defined in Lemma 10.0.5. Let $\hat{\lambda} = \sup\{c \in (0,1) : \hat{Q}^*(c) \leq \alpha\}$ and $\lambda = \sup\{c \in (0,1) : Q^*(c) \leq \alpha\}$. If $\hat{Q}^*(c) \to_p Q^*(c)$, then $\hat{\lambda} \to_p \lambda$.*

Proof

1. Let $T_{(i)}, i = 1, \ldots, m$ be the ordered values of $T(X_i), i = 1, \ldots, m$. Note that $\hat{Q}^*(c)$ is not continuous. Construct two functions as below. For $T_{(k)} < c < T_{(k+1)}$, define

$$\hat{Q}^{*-}(c) = \hat{Q}^*(T_{(k)}) + (\hat{Q}^*(T_{(k)}) - \hat{Q}^*(T_{(k-1)}))\frac{T_{(k)} - c}{T_{(k+1)} - T_{(k)}} \quad \text{and}$$

$$\hat{Q}^{*+}(c) = \hat{Q}^*(T_{(k)}) + (\hat{Q}^*(T_{(k+1)}) - \hat{Q}^*(T_{(k)}))\frac{T_{(k)} - c}{T_{(k+1)} - T_{(k)}}$$

2. Let $c_i = \dfrac{\hat{q}^*(x_i)}{\hat{q}(x_i)}$ and let $c_{(i)}$ be the value corresponding to $T_{(i)}$. Then

$$\begin{aligned}
\hat{Q}^*(T_{(k+1)}) - \hat{Q}^*(T_{(k)}) &= \frac{\sum_{i=1}^{k+1} c_{(i)} T_{(i)}}{\sum_{i=1}^{k+1} c_{(i)}} - \frac{\sum_{i=1}^{k} c_{(i)} T_{(i)}}{\sum_{i=1}^{k} c_{(i)}} \\
&= \frac{c_{(k+1)} T_{(k+1)}}{\sum_{i=1}^{k+1} c_{(i)}} + \sum_{i=1}^{k} \left(\frac{1}{\sum_{i=1}^{k+1} c_{(i)}} - \frac{1}{\sum_{i=1}^{k} c_{(i)}} \right) c_{(i)} T_{(i)} \\
&= \frac{c_{(k+1)} T_{(k+1)}}{\sum_{i=1}^{k+1} c_{(i)}} - \sum_{i=1}^{k} \frac{c_{(k+1)}}{\sum_{i=1}^{k+1} c_{(i)} \sum_{i=1}^{k} c_{(i)}} c_{(i)} T_{(i)} \\
&\geq 0
\end{aligned}$$

Hence $\hat{Q}^{*-}(c) \leq \hat{Q}^*(c) \leq \hat{Q}^{*+}(c)$. Let $\hat{\lambda}^- = \sup\{c \in (0,1) : \hat{Q}^{*-}(c) \leq \alpha\}$ and $\hat{\lambda}^+ = \sup\{c \in (0,1) : \hat{Q}^{*+}(c) \leq \alpha\}$. Then $\hat{\lambda}^- \geq \hat{\lambda} \geq \hat{\lambda}^+$.

3. It can be shown, as done is Lemma A.4 in Sun and Cai (2007), that $\hat{Q}^{*-}(c) \rightarrow_{a.s.} \hat{Q}^{*+}(c)$, $\hat{\lambda}^- \rightarrow_p \lambda$, and $\hat{\lambda}^+ \rightarrow_p \lambda$. Hence $\hat{\lambda} \rightarrow_p \lambda$. ∎

Proof of Theorem 8.4.1

By the Lemma 10.0.1, we have $\hat{T}_{OR} \to_p T_{OR}$ and by the Lemma 10.0.2 to Lemma 10.0.6, $\hat{\lambda}_{OR} \to_p \lambda_{OR}$. Then we have

$$\int I(\hat{T}_{OR}(x) < \hat{\lambda}_{OR}) f(x) dx \to \int I(T_{OR}(x) < \lambda_{OR}) f(x) dx,$$
$$\int I(\hat{T}_{OR}(x) < \hat{\lambda}_{OR}) q^*(x) dx \to \int I(T_{OR}(x) < \lambda_{OR}) q^*(x) dx.$$

The mFDR of the data driven procedure is $\frac{\pi_0 \int I(\hat{T}_{OR}(x) < \hat{\lambda}_{OR}) f(x) dx}{\int I(\hat{T}_{OR}(x) < \hat{\lambda}_{OR}) q^*(x) dx}$, which converges to $Q^*(\lambda_{OR})$. ∎

Proof of Theorem 8.4.2

Recall $q_1^*(x) = \int \mu^2 f(x-\mu) h(\mu) d\mu$. By Lemma 10.0.1 to Lemma 10.0.6, we have the following:

$$\int I(\hat{T}_{OR}(x) > \hat{\lambda}_{OR}) q_1^*(x) dx \to \int I(T_{OR}(x) > \lambda_{OR}) q_1^*(x) dx;$$
$$\int I(\hat{T}_{OR}(x) > \hat{\lambda}_{OR}) q^*(x) dx \to \int I(T_{OR}(x) > \lambda_{OR}) q^*(x) dx.$$

The mFNR of the data driven procedure is $\frac{(1-\pi_0) \int I(\hat{T}_{OR}(x) > \hat{\lambda}_{OR}) q_1^*(x) dx}{\int I(\hat{T}_{OR}(x) > \hat{\lambda}_{OR}) q^*(x) dx}$, which converges to $\tilde{Q}^*(\lambda_{OR})$, where $\tilde{Q}^*(OR)$ is the the mFNR of the oracle procedure. ∎

REFERENCES

[1] Benjamini, Y and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289-300.

[2] Benjamini, Y. & Hochberg, Y. (1997). Multiple hypothesis testing with weigths. Scand. J. Statist. **24** 407-418.

[3] Benjamini, Y. and Hochberg, Y. (2000) On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J. Educ. Behav. Statist.* **25** 60Ű83.

[4] Benjamini, Y., Krieger, A. M. and Yekutieli, D. (2006). Adap- tive linear step-up false discovery rate controlling procedures. *Biometrika* **93** 491-507.

[5] Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165-1188.

[6] Blanchard, G. and Roquain, E. (2008b). Self-consistent multiple testing procedures. ArXiv preprint math.ST/0802.1406v1.

[7] Blanchard, G. and Roquain, E. (2009). Adaptive FDR control under dependence and independence. *J. Mach. Learn. Res.* To appear.

[8] Brown, L. D. and Greenshtein, E. (2009). Nonparametric empirical bayes and compound decision approaches to estimation of a high dimensional vector of normal means. *Ann. Statist.* **37** 1685-1704.

[9] Cai, T. and Jin, J. (2010). Optimal rates of convergence for estimating the null density and proportion of nonnull effects in large-scale multiple testing. *Ann. Statist.* **38** 100Ű145.

[10] Cai, G. and Sarkar, S. K. (2005). Modified Simes' critical values under independence. Technical report, Temple Univ. Available at http://astro.temple.edu/sanat/reports/Modified-Simes Independence.pdf.

[11] Cai, G. and Sarkar, S. K. (2006). Modified Simes' critical values under positive dependence. *J. Statist. Plann. Inference* **136** 4129Ű4146.

[12] Dudoit, S. & Van der Lann, M. J. (2008). *Multiple testing procedures with applications to Genomics.* New York: 2008 Springer Science Business Media, LLC.

[13] Dykstra, R. L., Hewett, J. E., and Thompson, W. A., Jr. (1973). Events which are almost independent. *The Annals of Statistics* **1** 674-681.

[14] Efron, B., Tibshirani, R. , Storey, J. & Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. *J. Am. Statist. Ass.* **96** 1151-1160.

[15] Efron, B. (2004). Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. *J. Am. Statist. Ass.* **99** 96-104.

[16] Efron, B. (2007). Correlation and large-scale simultaneous significance testing. *J. Am. Statist. Ass.* **102** 93-103.

[17] Efron, B. (2008). Microarrays, empirical Bayes and the two-groups model. *Statist. Sci.* **23** 1Ű22.

[18] Efron, B. (2010) Large-scale inference, empirical Bayes methods for estimation, testing, and prediction. Cambridge University Press, 2010.

[19] Gabriel. K. R. (1969). Simultaneous test procedures-Some theory of multiple comparisons. *Ann. Math. Statist.,* **40** 224-250.

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

[21] Genovese, C. and Wasserman, L. (2002) Operating characteristics and extensions of the false discovery rate procedure. *J. R. Statist. Soc. B* **64** 499-517.

[22] Genovese, C. and Wasserman, L. (2004) A stochastic process approach to false discovery control. *Ann. Statist.* **32** 1035-1061.

[23] Genovese, C., Roeder, K. & Wasserman, L. (2006) False discovery control with p value weighting. *Biometrika* **93** 509-524

[24] Guo, W. (2009). A note on adaptive Bonferroni and Holm procedures under dependence. *Biometrika* **96** 1012-1018.

[25] Hochberg, Y. (1998). A sharper Bonferroni procedure for multiple significance testing. *Biometrika* **75** 800-802.

[26] Hochberg, Y. & Liberman, U. (1994). An extended Simes test. *Statist. Prob. Lett.* **21** 101-105.

[27] Hochberg, Y. and Tamhane, A.C. (1987). *Muiltiple comparison procedure.* New York: John Wiley and Sons.

[28] Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Statist.* **6** 65-70.

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

[30] Jin, J and Cai, T. (2007). Estimating the null and the proportional of nonnull effects in large-scale multiple comparisons. *J. Amer. Statist. Assoc.*, **102** 495-506.

[31] Karlin, S. and Rinott, Y (1980). Classes of orderings of measures and related correlation inequalitiesI: Multivariate totally positive distributions.*J. Multi. Anal* **10** 467-498.

[32] Lehmann, E.L. (1966). Some concepts of dependence. *Ann. Math. Statist. Soc. B***37** 1137-1153.

[33] Lehmann, E. L. Romano, J. P. (2005). Genalizations of the familywise error rate. *Ann. Statist* **33** 1138-1154.

[34] Liu, W. (1996). Multiple test of a non-hierarchical finite family of hypotheses. *J. R. Statist. Soc. B* **58**, 455-461.

[35] Owen, A. B. (2005) Variance of the number of false discoveries. *J. R. Statist. Soc. B* **67** 411Ű426.

[36] Pitt, L.D. (1982). Positively correlated normal variables are associated. *Ann. Probab*. **10** 496-499.

[37] Qiu, X., Klebanov, L. and Yakovlev, A. (2005) Correlation between gene expression levels and limitations of the empirical Bayes methodology for finding differentially expressed genes. *Statist. Appl. Genet. Molec. Biol., 4,* article 34

[38] Qiu, X and Yakovlev, A. (2006) Some comments on instability of false discovery rate estimation. *Journal of Bioinformatics and Computational Biology*. **4** 1057-1068.

[39] Robbins, H. (1951) Asymptotically subminimax solutions of compound statistical decision problems. *In Proc. 2nd Berkeley Symp. Mathematical Statistics and Probability.* Berkeley: University of California Press.

[40] Rom, D. (1990). A sequentiall rejective test procedure based on a modified Bonferroni inequality. *Biometrika* **77** 663-665.

[41] Romano, J. P. and Shaikh, A. M. (2006). Stepup procedures for control of the generalizations of the familywise error rate. *Ann Statist.* **34** 1850-1873.

[42] Roy, S.N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Statist.* **24** 220-238.

[43] Roy, S.N. and Bose, R.C. (1953). Simultaneous confidence interval estimation. *Ann. Math. Statist.* **24** 513-536

[44] Sarkar, S. K. (1998). Some probability inequalities for the ordered MTP2 random variables: A proof of the Simes conjecture. Ann. Statist. **26** 494-504.

[45] Sarkar, S. K. (2007). Stepup procedures controlling generalized FWER and generalized FDR. *Ann. Statist.*

[46] Sarkar, S. K. (2008a). Generalizing Sime's test and Hochberg's stepup procedure. *Ann. Statist.* **36** 337-363.

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

[48] Sarkar, S. K. and Chang, C. (1997) The Simes method for multiple hypothesis testing with positively dependent test statistics. *J. Amer. Statist. Assoc.* **92** 1601-1608

[49] Sarkar, S. K., Guo, W. and Finner, H. (2010). On adaptive pro- cedures controlling the familywise error rate. Submitted to Annals of Statistics.

[50] Sarkar, T.K. (1969). Some lower bounds of reliability. Technical report, 124, Department of operation research and statistics, Stanford Univ.

[51] Schwartzman, A. & Lin, X. (2009). The effect of correlation in false discovery rate estimation. Havard University Biostatics working paper series, paper 106.

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

[53] Storey, J.D. (2002). A direct approach to false discovery rates. *J. R. Statist. Soc. B* **64** 479-498.

[54] Storey, J. D. (2007) The optimal discovery procedure: a new approach to simultaneous significance testing. *J. R. Statist. Soc. B* **69** 347-368.

[55] Storey, J. D., Taylor, J.E. & Siegmund, D. (2004). Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach. *J. Roy. Statist. Soc. B* **66** 187-205.

[56] Sun, W. and Cai, T. (2007) Oracle and adaptive compound decision rules for false discovery rate control. *J. Am. Statist. Ass.* **102** 901Ű912.

[57] Sun, W. and Cai, T. (2009). Large-scale multiple testing under dependence. *J. R. Statist. Soc. B***71** 393-424.

[58] Tamhane, A.C. & Liu, L. (2008). On weighted Hochberg procedures. *Biometrika* **95** 279-294.

[59] van't Wout, A., Lehrman, G., Mikheeva, S., O'Keeffe, G., Katze, M., Bumgarner, R., Geiss, G., and Mullins, J. (2003). Cellular Gene Expression Upon Human Immunodeficiency Virus Type 1 Infection of CD4+-T-Cell Lines. *Journal of Virology*. **77** 1392-1402.

[60] Westfall,P.H. and Young, S.S.(1993). On adjusting P-Values for multiplicity. *Biometrics* **49** 941-945.

[61] Wu, W. (2008) On false discovery control under dependence. *Ann. Statist.* **36** 364-380.

[62] Yekutieli, D. & Benjamini, Y. (1999) Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *J. Statist. Planng Inf.* **82** 171-196.