High Dimensional Statistical Testing With Applications to Gene Significance Detection

Hongyuan Cao

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research (Statistics).

Chapel Hill
2010

Approved by

Michael R. Kosorok, advisor

Amarjit Budhiraja, reader

Andrew Nobel, reader

Yufeng Liu, reader

Donglin Zeng, reader

ABSTRACT

HONGYUAN CAO: High Dimensional Statistical Testing With Applications to Gene Significance Detection

(Under the direction of Michael R. Kosorok)

High-throughput screening has become an important mainstay for contemporary biomedical research. A standard approach is to use a large number of t-tests simultaneously and then select p-values in a manner that controls false discovery rate (FDR). Existing methods require very strong assumptions on the distribution of the data and the distribution of the p-values. We propose an asymptotically valid, data-driven procedure to find critical values for the t-statistics which requires minimal assumptions. A new asymptotically consistent estimate for the proportion of alternatives has been developed along the way. We demonstrate that our approach has improved computational efficiency and power over existing approaches while requiring fewer assumptions. The method controls the k-family wise error rate (k-FWER), the tail probability of false discovery proportion (FDTP) and false discovery rate (FDR). Simulation studies support our theoretical results and demonstrate the favorable performance of our new multiple testing procedure. We also apply our method to analyze cancer microarray studies.

One feature of our approach is that it takes the alternative into account. Existing approaches take the alternative into account as well. However, we found that a standard concavity assumption on the p-value distribution for the alternative is violated under certain circumstances. A more general concept is the monotone likelihood ratio condition (MLRC) introduced in Sun and Cai (2007). We show that the concavity assumption can be violated for (i) a simple heteroscedastic normal mixture model and (ii) dependent tests. Some interesting implications, including the choice of test statistics, existing FDR control procedures (step-up and step-down) and the power definition, are discussed.

ACKNOWLEDGEMENTS

This dissertation could not have been written without my advisor, Dr. Michael Kosorok, who patiently guided me through the dissertation process. It was my great fortune and privilege to work with him. I wish to express my deepest appreciation to him for his support, encouragement and mentoring throughout my doctoral studies. Most importantly, I learned what a great scholar should be from him.

I also would like to thank the rest of my committee members, Drs. Amarjit Budhiraja, Andrew Nobel, Yufeng Liu and Donglin Zeng for their careful reading, insightful discussions and stimulating suggestions which significantly improved this dissertation. I owe many thanks to Dr. Donglin Zeng for his encouraging inspirations, kind guidance, and enormously helpful discussions in completing this dissertation.

I am deeply grateful to Dr. Fred Wright for his in-depth knowledge of statistical genetics, valuable advice and kindness throughout this research. I also appreciate very much discussions with Dr. Wenguang Sun; working with him has been a wonderful experience for me. They are both great collaborators and my dearest friends.

I was very fortunate to have the opportunity to work at NC TraCs Institute under the supervision of Drs. Rosalie C. Dominik and Denise Esserman. A special thanks goes to them for the enjoyable working experience and financial support of my graduate studies and research.

I am also deeply indebted to Dr. Michael C. Wu for his enormous help during my job hunting process, whose friendship and support have made this journey more enjoyable and memorable. My sincere thanks go to Drs. Edward Carlstein, Yun Li and Wei Sun for improving my presentation skills. I would also like to thank our high dimensional statistical learning group as well as my friend Andrey Shabalin for elucidating comments and suggestions.

Finally, it's impossible to have completed this journey without the love, support and en-

couragement from my father Jirong Cao and my mother Yunqing Bai. This dissertation is dedicated to them.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction and Background

With the advancement of modern technology, it is now easier to get access to large data sets. For example, microarrays in genomics, functional Magnetic resonance imaging (MRI) in imagine analysis, astronomical surveys and many contemporary scientific endeavors. Compared with traditional ones, such data has very different structures. First, the number of features is huge, usually of the orders of tens of thousands; second, the number of observations is modest, usually of the orders of dozens; and third, very few individual features are related to the outcome, the so-called sparsity issue. The scientific objective is to do statistical inference about the true association between outcomes and relevant features. People have referred to it vividly as "finding needles in a haystack".

There are several inter-related problems that are of interest. First, we want to ask if there are any features in the data that are of interest to the scientist. This is a signal detection problem. Second, we would like to know what is the fraction of the features that contain the signals. This involves proportion estimation; third, after we know that there are certain features, we want to ask where are the features? This is a large scale multiple testing problem; and fourth, it is of interest to know the sizes of the features, this is a high-dimensional model selection and related coefficient estimation issue. In this dissertation, we focus on the second and third topics, large scale multiple testing and the related proportion estimation.

This dissertation is composed of three parts. In the first part, we propose an asymptotically valid, data-driven procedure to find critical values for t-statistics with minimal assumptions. The method controls the k-family wise error rate (k-FWER), the tail probability of false discovery proportion (FDTP) and false discovery rate (FDR). A new asymptotically consistent estimate for the proportion of alternatives has been developed along the way. In the second

part, we examine a standard assumption (monotone likelihood ratio condition) placed on the alternative hypotheses that is required for the optimal testing procedures. We exhibited a counter example situation under which this assumption does not hold. Some interesting implications, including step-up, step-down procedures, the choice of test statistics and power definition in multiple testing scenario are discussed. In the previous two parts, there is an underlying exchangeability assumption for all the tests, which means that each test is equally important. While in practice, some tests are more important than others. Therefore, in the third project, instead of doing individual tests, we put tests into different groups and study the joint association of each group with a phenotype of interest. The tests are grouped by some prior knowledge, for example, the inherent pathways by the underlying biological functioning in gene expression data. In the literature, the absolute association strength is evaluated, which favors larger groups at the expense of smaller groups. This motivated us to use the relative measure — the proportion of significant tests in a group—as comparison criterion. The proportion estimates are derived for the t-test, F- test and χ^2-test. This approach is shown to be robust to the size of the groups. Subsampling and the bootstrap are used to do inference.

In many areas of application, particularly in bioinformatics, conclusions are drawn by simultaneous testing of a large number of hypotheses. In these high-dimensional situations, common single inference approaches are well known to fail, leaving open the problem of making a small number of false discoveries by controlling a suitable error rate, and maximizing the power of each test at the same time. Such problems of simultaneous inference is usually refereed to as multiple testing. Applications of multiple testing include identifying neuronal activity in the living brain or the identification of differentially expressed genes in DNA microarray experiments. For a review of multiple testing methods in the context of microarray data analysis, see Dudoit, Shaffer and Boldrick (2003) and Sebastiani, Gussoni, Kohane and Ramoni (2003) for an excellent review of genomics and statistical challenges in genomics. Among the other possible applications, there are general medicine, pharmacology, epidemiology, psychometrics and even marketing. Moreover, multiple tests can be used as a key part of statistical procedures, like variable selection, item-response modeling, structural equation modeling, decision trees, wavelet thresholding, and so on.

Let's start with a motivating example. The dataset is from a microarray gene expression

study, see Golub (1999). There are 72 samples, of which 47 are from class ALL (acute lymphoblastic leutemia) and 25 are from class AML (acute myeloid leukemia). Each array was measured on the expression level for the same 7129 genes. Our interest is which genes are differentially expressed (d.e.) between these two types of tumors? The dataset can be represented in table 1 as follows:

Table 1.1: Gene expression data structure

d.e.(0/1)		ALL			AML		t-stat	p-value
indicator	X_1	$\cdots$	X_{47}	Y_1	$\cdots$	Y_{25}	T	P
H_1	$x_{1,1}$	$\cdots$	$x_{1,47}$	$y_{1,1}$	$\cdots$	$y_{1,25}$	t_1	p_1
H_2	$x_{2,1}$	$\cdots$	$x_{2,47}$	$y_{2,1}$	$\cdots$	$y_{2,25}$	t_2	p_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
H_{7129}	$x_{7129,1}$	$\cdots$	$x_{7129,47}$	$y_{7129,1}$	$\cdots$	$y_{7129,25}$	t_{7129}	p_{7129}

Consider a multiple testing situation in which m ($m = 7129$) tests are being performed. Suppose m_0 of the m hypotheses are true, and m_1 are false. Table 1 summarizes the possible outcomes: we denote with R the number of rejections, with V and F the exact (unknown) number of errors made after testing; and with U and S the number of correctly retained and rejected null hypotheses. The number of rejected hypotheses R is random, while m_0 and m_1 can either be considered as random or just not observable, depending on the specific application. In this dissertation, we treat m_0 and m_1 as unknown parameters.

Table 1.2: Outcomes when testing m hypotheses.

Hypothesis	Accept	Reject	Total
Null true	U	V	m_0
Alternative true	F	S	m_1
Total	W	R	m

In the usual (single) test setting, one controls the probability of false rejection (Type I error) while looking for a procedure that possibly minimizes the probability of observing a false negative (Type II error).

In the multiple case, despite the fact that each uncorrected level γ test falsely rejects the null hypothesis with small probability (namely, γ), as m increases the number of false positives can explode. For instance, if $m = 1000$ true null hypotheses are simultaneously tested at level $\gamma = 0.05$, around $R = 50$ false discoveries are expected. The consequences of so high a number

of false discoveries in real applications would usually be extremely disturbing to investigators. From a different point of view it can be said that a p-value around, for instance, 0.05 is unlikely to be correspondent to a true discovery, since it is very likely under the null hypothesis that such a small p-value will occur when many are computed at once.

Corrections arise from the control of specific type I error measures, and there are a variety of functions of the counts of false positives V that can serve as possible generalizations of the probability of Type I error. Control of the chosen Type I error rate can be loosely defined to be achieved when the error rate is bounded above by a pre-specified $\gamma \in (0, 1)$. The most classical multiple Type I error rate is based only on the distribution of V, that is, on what happens for the tests corresponding to the true null hypotheses. Here and in what follows, unless stated otherwise, probability and expectations are computed conditionally on the true parameter configurations, that is, on which and how many hypotheses are true.

FWER (FWER $= P(V \geq 1)$) control is desirable when the number of tests is small, so that a good number of rejections can be made, and all can be trusted to be true findings. But in modern applications, the number of tests can be very large. In these settings, FWER controlling procedures tend to become conservative and finally lead to rejection of a very limited number of hypotheses, if any. One way around this is to increase the number k of false rejections one is willing to tolerate. This results in a relaxed version of FWER, k-FWER $= P(V \geq k)$, defined as the chance of at least k type I errors.

Benjamini and Hochberg (1995) (BH) pioneered an alternative. Define the false discovery proportion (FDP) to be the number of false rejections divided by the number of rejections (FDP $= V/(R \vee 1)$). The only effect of the $R \cup 1$ in the denominator is that the ratio V/R is set to zero when $R = 0$. Without loss of generality, we treat FDP $= V/R$. The FDP is based on the distribution of R, that is, on what happens for the hypotheses for which H_0 is false. Define the false discovery tail probability FDTP $= P(V \geq \alpha R)$, where α is pre-specified based on the application. van der Laan, Dudoit and Pollard (2004) and independently Genovese and Wasserman (2006) along similar lines propose to control FDTP. Several papers have developed procedures for FDTP control. We shall not attempt a complete review here but mention the following: van der Laan, Dudoit and Pollard (2004) proposed an augmentation-based procedure, Lehmann and Romano (2005*b*) derived a step-down procedure and Genovese and Wasserman

(2004) suggested an inversion-based procedure, which is equivalent to the van der Laan, Dudoit and Pollard (2004) procedure under mild conditions (Genovese and Wasserman (2004)).

The false discovery rate (FDR) is the expected FDP. Benjamini and Hochberg (1995) provided a distribution-free, finite sample method for choosing a p-value threshold that guarantees that the FDR is less than a target level γ. The first to consider this error measure was probably Seeger (1968) who advocated control of FWER but with additional checking of the proportion of false nulls. Control of FDTP or FDR is justified by the idea that any researcher is prepared to bear a higher number of Type I errors when more rejections are made. In practical high-dimensional data analysis, the goal is to reduce a vast set of possibilities to a much smaller set of scientifically interesting prospects, which fits into the definition of FDR. Since this publication, there has been considerable research on both the theory and application of FDR control. Benjamini and Hochberg (2000) and Benjamini and Yekutieli (2001) extended the BH method to a class of dependent tests. Further generalizations of the FWER and FDR and proposed in Efron and Tibshirani (2002), Storey (2002) and Lehmann and Romano (2005*a*).

Storey for instance introduced the positive FDR defined as $pFDR = E[FDP|R > 0]$. Control of this error measure is more appropriate when the probability of making no rejections is high, so that FDR control may be misleading; and can moreover lead to more powerful multiple testing procedures in certain situations. Note that for any number of rejected hypotheses FDR $\leq$ pFDR. Storey suggested how to estimate and thus control pFDR using a fixed rejection region, and introduced the q-value, a pFDR analogue of the p-value. An interpretation of the pFDR and q-value as Bayesian posterior probabilities is in Storey (2003), who also shows connections to classification theory. A discussion of weighted FDR controlling procedures, included in Benjamini and Hochberg (1997) and Genovese, Roeder and Wasserman (2006), also shows how to give different importance to each hypothesis, and also how to enhance power by weighting.

It is straightforward to see that FDR and FDTP control is also a weak control on the FWER in the sense that FWER is controlled if all the null hypotheses are true.

FDTP and FDR are closely related, being functionals of the same random variable, namely, the FDP. It is straightforward to see that in general if FDTP is controlled at level γ, then FDR is controlled at level $\alpha + (1 - \alpha)\gamma$. A partial converse is given by an application of

Markov's inequality, which shows that if FDR $< \gamma$, then FDTP $< \gamma/\alpha$. Moreover, note that FDR $= E[\text{FDTP}] = \int_0^1 P(V > \alpha R) d\alpha$, that is, FDR control is a control on the average FDTP (with respect to Lebesgue measure). Following this statement, we can apply the mean value theorem and prove that at least asymptotically there exist $\eta \in [0, 1]$ such that FDTP(η) = FDR. That is, if FDR $\leq \gamma$, there exist $\eta \in [0, 1]$ for which FDTP $\leq \gamma$ for any $\alpha > \eta$.

A Bayesian mixture model approach to obtain multiple testing procedures controlling the FDR is considered in Efron, Tibshirani, Storey and Tusher (2001), Storey (2002), Storey (2003), Storey and Tibshirani (2003), Storey, Tibshirani and Siegmund (2004). Wu (2008) considered the conditional dependence model under the assumption of Donsker properties of the indicator function of the true state for each hypothesis and derived asymptotic properties of false discovery proportions and numbers of rejected hypotheses. A systematic study on multiple testing procedures is given in a book by Dudoit and van der Laan (2008). Other related work can be found in Chi (2007) and Chi and Tan (2008).

One challenge in multiple hypothesis testing is that many procedures depend on the proportion of null hypotheses which is not known in reality. Estimating the proportion has long been known as a difficult problem. There have been some interesting developments recently, for example, an approach by Meinshausen and Rice (2006) (see also Efron, Tibshirani, Storey and Tusher (2001), Genovese and Wasserman (2004), Meinshausen and Bühlmann (2005), and Langaas, Lindqvist and Ferkingstad (2005)). Roughly speaking, these approaches are only successful under a condition which Genovese and Wasserman (2004) called the "purity" condition. Unfortunately, the purity condition depends on p-values and is hard to check in practice.

The general framework for k-FWER, FDTP and FDR control and the estimation of proportion of alternative hypotheses is based on p-values which are assumed to be known in advance or can be accurately approximated. However, the assumption that p-values are always available is not realistic. In some special settings, approximate p-values have been shown to be asymptotically equivalent to exact p-values for controlling FDR (Fan, Hall and Yao (2007) and Kosorok and Ma (2007)). But these approximations are only helpful in certain simultaneous error control settings and are not universally applicable. Moreover, if the p-values are not reliable, any procedures derived afterwards are problematic.

This motivates us to propose a method to find critical values directly for rejection regions to

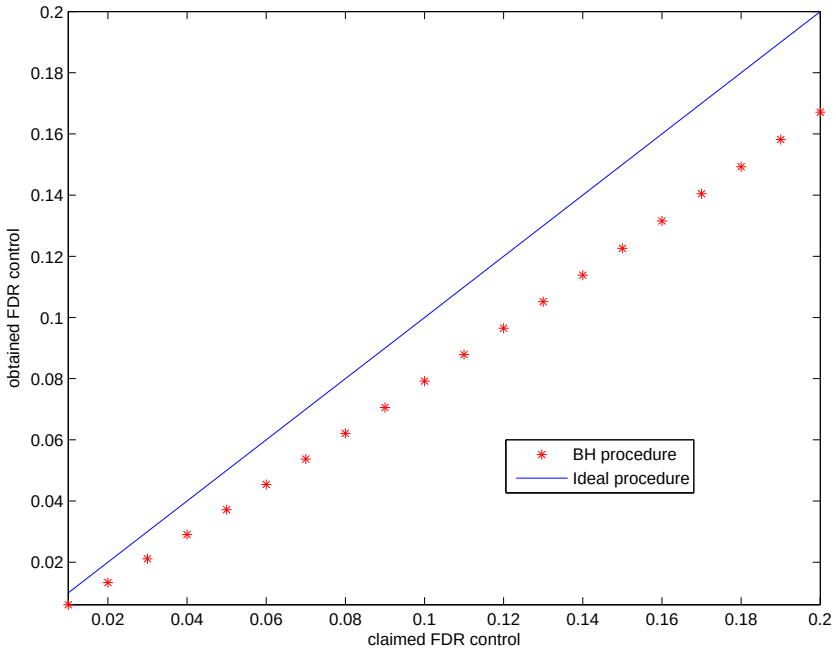
control k-FWER, FDTP and FDR by using one-sample and two-sample t-statistics. The advantage of using t-tests is that they require minimum conditions on the population, only existence of the fourth moment, which is relatively easy to be satisfied by most statistical distributions, rather than other stringent conditions such as the existence of the moment generating function. In addition, we approximate tail probabilities of both null and alternative hypotheses accurately, rather than p-value approaches that only consider the case under the null hypotheses. Thus a better ranking of hypotheses is obtained. Furthermore, we propose a consistent estimate of the proportion of alternative hypotheses which only depends on test statistics. As long as the asymptotic distribution of the test statistic is known under the null hypothesis, we can apply our method to get this proportion estimated, resulting in more precise cutoffs.

The BH procedure controls the FDR conservatively at $\pi_0\gamma$, where π_0 is the proportion of null hypotheses and γ is the targeted significance level. If π_0 is much smaller than 1, the statistical power is greatly compromised. The power we use in this paper is NDR $= E[S]/m_1$ as defined in Craiu and Sun (2008). We also discuss a parallel concept called false non-discovery rate (FNR) first proposed by Genovese and Wasserman (2002) and independently by Sarkar (2002) and the rationale for using NDR as a power definition. In the situation that t-statistics can be used, our procedure gives a better approximation, and more accurate critical values can be obtained, by plugging in the estimate of π_0. The validity of our approach is guaranteed by empirical process methods and recent theoretical advances of self-normalized moderate deviations in combination with Berry-Esseen type bounds for central and non-central t-statistics.

To illustrate, we simulate a Markov chain as in Sun and Cai (2009) of Bernoulli variables $(H_i), i = 1, \cdots, 5000$ to indicate the true state of each hypothesis test ($H_i = 1$ if the alternative is true; $H_i = 0$ if the null is true). Conditional on the indicator, observations $x_{ij}, i = 1, \cdots, 5000, j = 1, \cdots, 80$ are generated according to the model $x_{ij} = \mu_i + \epsilon_{ij}$. The one-sample t-statistic is used to perform simultaneous hypothesis testing. Figure 1.1 shows the plot of 10000 MCMC results of the realized and nominal FDR control based on the BH method for different control levels. From this plot, we can see that as the control level increases, the BH procedure becomes more and more conservative. For instance, the actual obtained FDR is 0.167 when the nominal level is set at 0.2, reflecting a significant loss in power.

The three methods of multiple testing control we utilize are k-FWER, FDTP and FDR.

Figure 1.1: Claimed and obtained FDR control using BH procedure



The criterion for using k-FWER is asymptotically

$$P(V \geq k) \leq \gamma. \tag{1.1}$$

Since we only apply our method when there are discoveries ($R > 0$), we need for the FDTP with a given proportion $0 < \alpha < 1$ and significance level $0 < \gamma < 1$, asymptotically, to satisfy

$$P(V \geq \alpha R) \leq \gamma. \tag{1.2}$$

Similarly, the criterion for using FDR is asymptotically

$$FDR \leq \gamma \quad \text{or} \quad \int_0^1 P(V \geq \alpha R) d\alpha \leq \gamma. \tag{1.3}$$

For each hypothesis test, we claim it is significant if $|T_i| \geq t$, where T_i is the ith test statistic under consideration from $i = 1, \ldots, m$ distinct hypotheses. In this dissertation, the T_is are one-sample or two-sample t-statistics unless otherwise noted. Our goal is to choose t such that control of k-FWER FDTP and FDR is asymptotically guaranteed simultaneously over all

$i = 1, \ldots, m$.

In evaluating the efficiency of multiple testing procedures, the FNR is used as a criterion similar to type II error in single hypothesis testing. Defined as FNR $= E[F/W]$, formally, FNR is the expected value of missed discoveries divided by the total number of accepted hypotheses. Basically, it evaluates the detection ability of a multiple testing procedure. In the literature, the multiple testing problem is framed as a weighted classification problem minimizing FNR + λFDR, see Genovese and Wasserman (2002), Storey (2003) and Sun and Cai (2007). The optimal procedure is the one that minimizes E(FNR) subject to E(FDR) $\leq \alpha$. Under the assumption that the distribution of p-values under the alternative is concave, FNR is shown to be a monotone decreasing function with respect to FDR see Sun and Cai (2007). So the best procedure is the one that satisfies E(FDR) $= \alpha$ and E(FNR) is automatically minimized. The concavity assumption is intuitively appealing—under the null hypothesis, the p-value has a Unif$(0, 1)$ distribution, and under the alternative, the cumulative distribution function of the p-value is stochastically smaller than the cumulative distribution of Unif$(0, 1)$ since small p-values indicate significance of the alternative. However, the monotone decreasing relationship between FDR and FNR does not necessarily hold in general as will be illustrated in one of our simulation studies in chapter 5. On the other hand, NDR defined as NDR $= E[F/m_1]$ is monotonically decreasing as more true alternatives are rejected (S decreases and m_1 is fixed if F increases.) So we use NDR as the detection ability measure in this dissertation.

An FDR procedure based on the test statistic T in general has the following form:

$$\boldsymbol{\delta}(T, c) = \{I[T < c] : i = 1, \cdots, m\}.$$

T is any test statistic, not necessarily a t-statistic. Note that we omit the dependent relationship of T on the data X and simply write it as T. In the multiple testing literature, it is often assumed that

$$\text{the FDR of } \boldsymbol{\delta}(T, c) \text{ is monotonically increasing in the cutoff } c. \tag{1.4}$$

When p-values are used, a sufficient condition for (5.1) to hold is that

$$G_P^1(t) \text{ is concave in t,} \tag{1.5}$$

where $G_P^1(t)$ is the p-value distribution under the alternative. The concavity of $G_P^1(t)$ has been assumed in Storey (2003), Genovese and Wasserman (2004) and Kosorok and Ma (2007). A generalized condition was considered for a family of test statistics $\mathcal{T}$ in Sun and Cai (2007). Specifically, let G_T^0 and G_T^1 be the conditional cdf of T under the null and alternative, respectively. Denote by g_T^0 and g_T^1 the corresponding density functions. A sufficient condition for (5.1) to hold is the following monotone likelihood ratio condition (MLRC):

$$g_T^1(c)/g_T^0(c) \text{ is monotonically decreasing in } c. \tag{1.6}$$

Note that G_P^0 is uniform, and it is easy to verify that (5.2) implies (5.3) when the p-value is used. In some applications, it may be important to impose less structure under the alternative.

The main contribution of this dissertation is as follows: 1. Moderate deviation results which only require the finiteness of fourth moment from which the statistic is computed in probability theory are applied in multiple testing. Thus the applicability of this procedure is dramatically expanded—it can deal with non-normal populations and even highly skewed populations. 2. The critical values for rejection regions are computed directly, which circumvents the intermediate p-value step. 3. An asymptotically consistent estimation of the proportion of alternative is developed for multiple testing procedures under very general conditions. 4. A non-monotonicity phenomenon of FDR in terms of cut-off value is noted, with a counter example to show the consequence of this violation in implementing different testing procedures.

The remainder of the dissertation is organized as follows. In chapter 2, we present the basic data structure, our goals, the procedures and theoretical results for the one-sample t-test. Two-sample t-test results are discussed in chapter 3. Chapter 4 is devoted to numerical investigations using simulation, and we apply our procedure to detect significantly expressed genes in a microarray study of leukemia cancer. In chapter 5, we provide a counter example in violation of the monotonicity assumption of FDR analytically and numerically, discuss the implication on different testing procedures as well as the interpretation of the testing result. Group testing is discussed in Chapter 6 and concluding remarks are given in chapter 7. Proofs of results from Sections 2 and 3 are given in the Appendix.

CHAPTER 2

One-sample t-test

2.1 One-sample t-test

In this section, we first introduce the basic framework for simultaneous hypothesis testing followed by our main results. Estimation of the unknown proportion of alternative hypotheses π_1 is presented next. We conclude this section by presenting theoretical results for the special case of completely independent observations. This special setting is the basis for the more general main results and also is of independent interest since fairly precise rates of convergence can be obtained.

2.1.1 Basic framework

As a specific application of multiple hypothesis testing in very high dimensions, we use gene expression microarray data to illustrate. At the level of single genes, researchers seek to establish whether each gene in isolation behaves differently in a control versus a treatment situation. If the transcripts are pair-wise under two conditions, we can use a one-sample t-statistic to test for differential expression.

The mathematical model is

$$X_{ij} = \mu_i + \epsilon_{ij}, \quad 1 \leq j \leq n, \quad 1 \leq i \leq m. \tag{2.1}$$

It should be noted that the following discussion is under this model and does not hold in general. Here X_{ij} represents the expression level in the ith gene and jth array. Since the subjects are independent, for each i, $\epsilon_{i1}, \epsilon_{i2}, \cdots \epsilon_{in}$ are independent random variables with mean zero and variance σ_i^2. The null hypothesis is $\mu_i = 0$ and the alternative hypothesis is $\mu_i \neq 0$. For the

relationship between different genes, we propose the conditional independence model: Let (H_i) be a 0/1 valued stationary process, and, given $(H_i)_{i=1}^m$, $X_{ij}, i = 1, \cdots, m$ are independently generated. The dependence is imposed on the hypothesis (H_i), where $H_i = 0$ if the null hypothesis is true and $H_i = 1$ if the alternative is true. From Table 1, we can see that $\sum_{i=1}^m H_i = m_1$ and $\sum_{i=1}^m (1 - H_i) = m_0$. It is assumed that $(H_i)_{i=1}^m$ satisfy a strong law of large numbers:

$$\frac{1}{m}\sum_{i=1}^{m} H_i \to \pi_1 \in (0,1) \quad a.s. \qquad \text{as} m \to \infty. \tag{2.2}$$

This condition is satisfied in a variety of scenarios, for example, the independent case, Markov models, stationary ergodic models, etc. Consider the one-sample t-statistic

$$T_i = \sqrt{n}\bar{X}_i / S_i,$$

where

$$\bar{X}_i = \frac{1}{n}\sum_{j=1}^{n} X_{ij}, \quad S_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)^2.$$

If we use t as a cut-off, then the number of rejected hypotheses, and the number of false discoveries are

$$R = \sum_{i=1}^{m} 1_{\{|T_i| \geq t\}}, \quad V = \sum_{i=1}^{m}(1 - H_i)1_{\{|T_i| \geq t\}}. \tag{2.3}$$

Under the null hypothesis, it is well known that T_i follows a student t-distribution with $n-1$ degrees of freedom if the sample is from a normal distribution. Asymptotic convergence to a standard normal distribution holds when the population is completely unknown provided it has finite fourth moment under the null hypothesis. Moreover, under the alternative hypothesis, T_i can also be approximated by a normal distribution but with a shift in location. We will show that

$$F_0(t) := P(|T_i| \geq t | H_i = 0) = P(|Z| \geq t)(1 + o(1)) = 2\bar{\Phi}(t)(1 + o(1)) \quad \text{as} n \to \infty, \tag{2.4}$$

$$F_1(t) := P(|T_i| \geq t | H_i = 1) = \mathrm{E}[P(|Z + \sqrt{n}\mu_i/\sigma_i| \geq t | \mu_i, \sigma_i)](1 + o(1)) \quad \text{as} n \to \infty, \tag{2.5}$$

uniformly for $t = o(n^{1/6})$ under some regularity conditions, where Z denotes the standard

normal random variable, $\bar{\Phi}$ is the tail probability of the standard normal distribution and that the critical values $t_{n,m}$ that control the FDTP and FDR asymptotically at prescribed level γ are bounded. These assumptions are fairly realistic in practice. We do not require the critical value for k-FWER to be bounded. Although we do not typically know m_1, $F_0(t)$ or $F_1(t)$ in practice, we need the following theorem—the proof of which is given in the Appendix—as the first step. We will shortly extend this result, in Theorem 2.1.2 below, to permit estimation of the unknown quantities.

Theorem 2.1.1. *Assume that* $E(\epsilon_{ij}|\mu_i, \sigma_i^2) = 0$, $Var(\epsilon_{ij}|\mu_i, \sigma_i^2) = \sigma_i^2$, $\sup_{i,j} E\epsilon_{ij}^4 < \infty$, $0 < \pi_1 < 1 - \alpha$ *and (2.2) is satisfied. Also assume that there exist* $\epsilon_0 > 0$ *and* $c_0 > 0$ *such that*

$$P(|\sqrt{n}\mu_i/\sigma_i| \geq \epsilon_0 | H_i = 1) \geq c_0 \ \ \forall n \geq 1. \tag{2.6}$$

Let

$$\mu_m(t) = \alpha m_1 F_1(t) - (1 - \alpha) m_0 F_0(t), \tag{2.7}$$

and

$$\sigma_m^2(t) = \alpha^2 m_1 F_1(t)(1 - F_1(t)) + (1 - \alpha)^2 m_0 F_0(t)(1 - F_0(t)). \tag{2.8}$$

(i) If $t_{n,m}^{fdtp}$ *is chosen such that*

$$t_{n,m}^{fdtp} = \inf\{t : \mu_m(t)/\sigma_m(t) \geq z_\gamma\}, \tag{2.9}$$

where z_γ *is the* γ*th quintile of standard normal distribution, then*

$$\lim_{m \to \infty} P(FDP \geq \alpha) = \lim_{m \to \infty} P(V \geq \alpha R) \leq \gamma \tag{2.10}$$

holds.

(ii) If $t_{n,m}^{fdr}$ *is chosen such that*

$$t_{n,m}^{fdr} = \inf\{t : \frac{m_0 F_0(t)}{m_0 F_0(t) + m_1 F_1(t)} \leq \gamma\}, \tag{2.11}$$

then

$$\lim_{m\to\infty} FDR = \lim_{m\to\infty} E(V/R) \leq \gamma \tag{2.12}$$

holds.

(iii) If $t_{n,m}^{k-FWER}$ *is chosen such that*

$$t_{n,m}^{k-FWER} = \inf\{t : P(\eta(t) \geq k) \leq \gamma\}, \tag{2.13}$$

where $\eta(t) \sim Poisson(\theta(t))$ *and*

$$\theta(t) = m_o F_0(t),$$

then

$$\lim_{m\to\infty} k\text{-}FWER = \lim_{m\to\infty} P(V \geq k) \leq \gamma \tag{2.14}$$

holds.

Remark. In the next section, we use a Gaussian approximation for $F_0(t)$ and $F_1(t)$ for both FDTP and FDR, for which the critical values are shown to be bounded. In this case, m can be arbitrarily large while the critical value remains bounded. Due to sparsity, we use a Poisson approximation for k-FWER, for which the critical value is no longer bounded as $m \to \infty$, and we require $\log m = o(n^{1/3})$.

2.1.2 Main Results

Note that in Theorem 2.1.1, there are unknown parameter m_1 and unknown functions $F_0(t)$ and $F_1(t)$ involved in $\mu_m(t)$ and $\sigma_m(t)$. For practical settings, we need to estimate these quantities. We will begin by assuming that we have a strongly consistent estimate of π_1, and we will then provide one such estimate in the next section. Given $\mathcal{H}$, note that $p(t) = P(|T_i| \geq t) = (1 - H_i)P(|T_i| \geq t|H_i = 0) + H_i P(|T_i| \geq t|H_i = 1)$ can be estimated from the empirical distribution $\hat{p}_m(t)$ of $\{|T_i|\}$, where

$$\hat{p}_m(t) = \frac{1}{m}\sum_{i=1}^{m} I_{\{|T_i|\geq t\}} \tag{2.15}$$

and that $P(|T_i| \geq t|H_i = 0)$ is close to $P(|Z| \geq t)$ when n is large by (2.4). The next theorem, proven in the Appendix, provides a consistent estimate of the critical value $t_{n,m}$.

Theorem 2.1.2. *Let*

$$\nu_m(t) = \alpha\hat{p}_m(t) - 2(1-\hat{\pi}_1)\bar{\Phi}(t) \tag{2.16}$$

and

$$\begin{aligned} \tau_m^2(t) &= \alpha^2(\hat{p}_m(t) - 2(1-\hat{\pi}_1)\bar{\Phi}(t))(1 - \frac{1}{\hat{\pi}_1}(\hat{p}_m(t) - 2(1-\hat{\pi}_1)\bar{\Phi}(t))) \\ &+ 2(1-\alpha)^2(1-\hat{\pi}_1)\bar{\Phi}(t)(1-2\bar{\Phi}(t)), \end{aligned} \tag{2.17}$$

where $\hat{\pi}_1$ is a strongly consistent estimate of π_1. Assume that the conditions of Theorem 2.1.1 are satisfied.

(i) If $\hat{t}_{n,m}^{fdtp}$ is chosen such that

$$\hat{t}_{n,m}^{fdtp} = \inf\{t : \frac{\sqrt{m}\nu_m(t)}{\tau_m(t)} \geq z_\gamma\}, \tag{2.18}$$

then

$$|\hat{t}_{n,m}^{fdtp} - t_{n,m}^{fdtp}| = o(1) \ \ a.s.. \tag{2.19}$$

(ii) If $\hat{t}_{n,m}^{fdr}$ is chosen such that

$$\hat{t}_{n,m}^{fdr} = \inf\{t : \frac{2(1-\hat{\pi}_1)\bar{\Phi}(t)}{\hat{p}_m(t)} \leq \gamma\} \tag{2.20}$$

then

$$|\hat{t}_{n,m}^{fdr} - t_{n,m}^{fdr}| = o(1) \ \ a.s.. \tag{2.21}$$

(iii) If $\hat{t}_{n,m}^{k-FWER}$ is chosen such that

$$\hat{t}_{n,m}^{k-FWER} = \inf\{t : P(\zeta(t) \geq k)\} \leq \gamma \tag{2.22}$$

where $\zeta(t) \sim Poisson(\bar{\theta}(t))$ *and*

$$\bar{\theta}(t) = 2m(1 - \hat{\pi}_1)\bar{\Phi}(t),$$

then as long as $\log m = o(n^{1/3})$

$$|\hat{t}_{n,m}^{k-FWER} - t_{n,m}^{k-FWER}| = o(1) \ \ a.s.. \tag{2.23}$$

Remark. This theorem deals with the general dependence case, where $(H_i)_1^m$ is assumed to follow a two state hidden model and the data are generated independently conditional on $(H_i)_1^m$. The proof is mainly based on the independence case, which we present in Section 2.4 below, plus a conditioning argument.

2.1.3 Estimating π_1

In the previous section, we assumed that $\hat{\pi}_1$ was a consistent estimator of π_1. Now we develop one such estimator. By the two group nature of multiple testing, the test statistic is essentially a mixture of null and alternative hypotheses with proportion as a parameter. By virtue of moderate deviations, the distribution of t-statistics can be accurately approximated under both null and alternative hypotheses. But for the alternative approximation, an unknown mean and variance are involved. So we think of a functional transformation of the t-statistics which has a ceiling at 1 to get a conservative estimate of π first which is consistent under certain conditions. Let $c > 0$ and define $g_c(x) = \min(|x|, c)/c$. It is easy to see that g_c is a decreasing function of c, bounded by 1 and that the derivative $\frac{dg_c}{dc}$ is bounded by $1/c$. Hence the function class $\{g_c\}$ indexed by c is a Donsker class and thus also Glivenko-Cantelli. Let

$$\hat{g}_c = \frac{1}{m}\sum_{i=1}^{m} g_c(T_i). \tag{2.24}$$

Theorem 2.1.3. *We have*

$$\pi_1 \geq \lim_{m\to\infty, n\to\infty} \sup_{c>0} \frac{\hat{g}_c - E(g_c(Z))}{1 - E(g_c(Z))} \ \ a.s.$$

If, in addition, we assume that

$$\sqrt{n}\mu_i/\sigma_i \to \infty \text{ for all } i \text{ with } H_i = 1, i = 1, \cdots, m, \quad a.s., \quad \text{as } n \to \infty, \tag{2.25}$$

then

$$\pi_1 = \lim_{m\to\infty, n\to\infty} \sup_{c>0} \frac{\hat{g}_c - E(g_c(Z))}{1 - E(g_c(Z))} \quad a.s.,$$

where

$$E(g_c(Z)) = \frac{2}{c\sqrt{2\pi}}(1 - e^{-c^2/2}) + 2\bar{\Phi}(c).$$

Proof. We can write

$$\begin{aligned} \hat{g}_c &= \frac{\sum_{i=1}^m 1_{\{H_i=0\}}}{m} \frac{\sum_{i=1}^m g_c(T_i)1_{\{H_i=0\}}}{\sum_{i=1}^m 1_{\{H_i=0\}}} + \frac{\sum_{i=1}^m 1_{\{H_i=1\}}}{m} \frac{\sum_{i=1}^m g_c(T_i)1_{\{H_i=1\}}}{\sum_{i=1}^m 1_{\{H_i=1\}}} \\ &:= \frac{m_0}{m} I + \frac{m_1}{m} II. \end{aligned}$$

Let $\mathcal{H} = \{H_i, 1 \le i \le m\}$. Conditional on $\mathcal{H}$, $T_i, 1 \le i \le m$, are independent random variables. We consider I first. Let

$$A_m(c) = \frac{\sum_{i=1}^m g_c(T_i|\mathcal{H})1_{\{H_i=0\}}}{\sum_{i=1}^m 1_{\{H_i=0\}}} - \frac{\sum_{i=1}^m E(g_c(T_i|\mathcal{H})1_{\{H_i=0\}}}{\sum_{i=1}^m 1_{\{H_i=0\}}},$$

and let E be the infinite sequence $1_{\{H_1=0\}}, 1_{\{H_2=0\}}, \ldots$, and let F be the event that $\sum_{i=1}^m 1_{\{H_i=0\}} \to \infty$ as $m \to \infty$. By the assumption (2.2), we know that $P(F) = 1$. Thus

$$P\left(lim_{m\to\infty} \sup_{c>0} |A_m(c)| = 0\right) = E\left[P\left(\lim_{m\to\infty} \sup_{c>0} |A_m(c)| = 0 \middle| E\right)\right] = 1,$$

where the second equality follows from the fact that, conditional on E, the terms in the sum are i.i.d., and thus the standard Glivenko-Cantelli theorem applies. Arguing similarly based on conditioning on the sequence $1_{\{H_1=1\}}, 1_{\{H_2=1\}}, \ldots$, we can also establish that

$$\sup_{c>0} \left| \frac{\sum_{i=1}^m g_c(T_i|\mathcal{H})1_{\{H_i=1\}}}{\sum_{i=1}^m 1_{\{H_i=1\}}} - \frac{\sum_{i=1}^m E(g_c(T_i|\mathcal{H})1_{\{H_i=1\}}}{\sum_{i=1}^m 1_{\{H_i=1\}}} \right| \to 0, a.s..$$

Now note that $II \le 1$. Thus, since $m_0/m \to (1 - \pi_1)$ a.s. and $m_1/m \to \pi_1$ a.s., we have that

when $m \to \infty, n \to \infty$,

$$\begin{aligned} \hat{g}_c &\leq (1-\pi_1)E(g_c(Z)) + \pi_1 \quad a.s. \\ &= E(g_c(Z)) + (1 - E(g_c(Z)))\pi_1. \end{aligned}$$

We now have the following lower bound for π_1:

$$\pi_1 \geq \lim_{m\to\infty, n\to\infty} \sup_{c>0} \frac{\hat{g}_c - E(g_c(Z))}{1 - E(g_c(Z))} \quad a.s. \tag{2.26}$$

Define

$$\Delta_1 := (1-\pi_1)E(g_c(Z)) + \pi_1 \frac{1}{m_1} \sum_{i=1}^{m} E(g_c(T_i)|\mathcal{H}) 1_{\{H_i=1\}},$$
$$\Delta_2 := (1-\pi_1)E(g_c(Z)) + \pi_1 \frac{\sum_{i=1}^{m} E(g_c(Z + \frac{\sqrt{n}\mu_i}{\sigma_i})) 1_{\{H_i=1\}}}{\sum_{i=1}^{m} 1_{\{H_i=1\}}}.$$

Letting $n \to \infty$, we have $\sup_{c>0} |\Delta_1 - \Delta_2| \to 0$ a.s.. Also,

$$\begin{aligned} \Delta_2 &= (1-\pi_1)E(g_c(Z)) + \pi_1 \frac{1}{\sum_{i=1}^{m} 1_{\{H_i=1\}}} \sum_{i=1}^{m} E(g_c(Z + \frac{\sqrt{n}\mu_i}{\sigma_i})(I_{\{|Z+\frac{\sqrt{n}\mu_i}{\sigma_i}|\geq c\}} + I_{\{|Z+\frac{\sqrt{n}\mu_i}{\sigma_i}|<c\}}))H_i \\ &\geq (1-\pi_1)E(g_c(Z)) + \pi_1 \frac{\sum_{i=1}^{m} P(|Z + \frac{\sqrt{n}\mu_i}{\sigma_i}| \geq c)H_i}{\sum_{i=1}^{m} 1_{\{H_i=1\}}} \\ &\geq (1-\pi_1)E(g_c(Z)) + \pi_1 \\ &= E(g_c(Z)) + \pi_1(1 - E(g_c(Z))). \end{aligned}$$

Note that

$$\sup_c |\hat{g}_c - \Delta_1| \to 0 \text{ a.s., as } m \to \infty, n \to \infty.$$

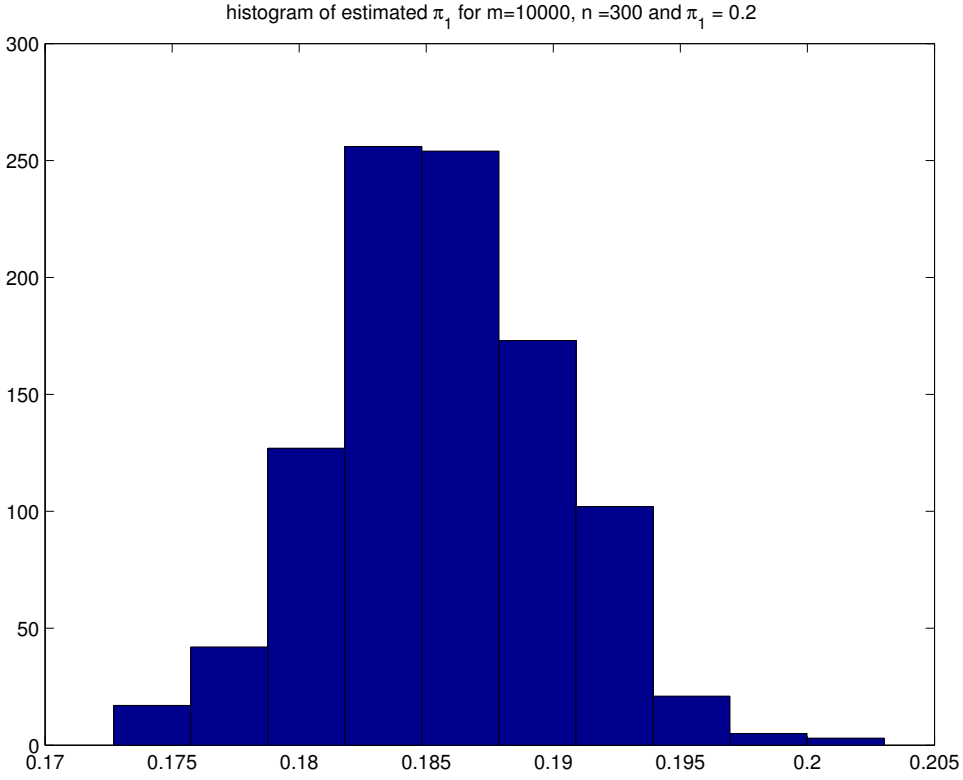
Therefore,

$$\hat{g}_c \geq E(g_c(Z)) + \pi_1(1 - E(g_c(Z))) \text{ a.s., as } m \to \infty, n \to \infty.$$

Thus we obtain

$$\pi_1 \leq \lim_{m\to\infty, n\to\infty} \sup_{c>0} \frac{\hat{g}_c - E(g_c(Z))}{1 - E(g_c(Z))} \ a.s.. \quad \square \tag{2.27}$$

Figure 2.1: Histogram of π_1 estimate



In consequence of this theorem, we propose the following estimate of π_1:

$$\hat{\pi}_1 := \sup_{c>0} \frac{\hat{g}_c - E(g_c(Z))}{1 - E(g_c(Z))}, \tag{2.28}$$

where

$$E(g_c(Z)) = \frac{2}{c\sqrt{2\pi}}(1 - e^{-c^2/2}) + 2\bar{\Phi}(c).$$

The histogram of a simulation study on the accuracy of this estimate is plotted.

Remark. If we use $\hat{\pi}_1$ as given in (2.28), then theorem 2.1.2 yields a fully automated procedure to do multiple hypothesis testing in very high dimensions in practical data settings.

2.1.4 Consistency and rate of convergence under independence

In order to prove the main results in the general, possibly dependent t-test setting we need results under the assumption of independence between t-tests. Specifically, we assume in this section that $(T_i, H_i), i = 1, \cdots, m$ are independent, identically distributed random variables,

with $\pi_1 = P(T_i = 1)$. This independence assumption can also yield stronger results than the more general setting and is of independent interest.

The next theorem, proven in the Appendix, provides a strong consistent estimate of the critical value $t_{n,m}$ as well as its rate of convergence:

Theorem 2.1.4. *Let*

$$\nu_m(t) = \alpha\hat{p}_m(t) - 2(1-\pi_1)\bar{\Phi}(t) \tag{2.29}$$

and

$$\begin{aligned}\tau_m^2(t) &= \alpha^2\hat{p}_m(t)(1-\hat{p}_m(t)) + 4\alpha(1-\pi_1)\hat{p}_m(t)\bar{\Phi}(t) \\ &\quad +2(1-\pi_1)\bar{\Phi}(t)(1-2\alpha-2(1-\pi_1)\bar{\Phi}(t)).\end{aligned}$$

Assume the conditions of Theorem 2.1.1 with (2.2) replaced by the assumption that $(T_i, H_i), i = 1, \cdots, m$ are i.i.d. and $\pi_1 = P(T_i = 1)$. Let $\mathcal{J} = \{i : H_i = 1\}$ be the set that contains the indices of alternative hypotheses. Also assume that μ_i, σ_i are i.i.d. for $i \in \mathcal{J}$.

(i) If $\hat{t}_{n,m}^{fdtp}$ is chosen such that

$$\hat{t}_{n,m}^{fdtp} = \inf\{t : \frac{\sqrt{m}\nu_m(t)}{\tau_m(t)} \geq z_\gamma\}, \tag{2.30}$$

then

$$|\hat{t}_{n,m}^{fdtp} - t_{n,m}^{fdtp}| = O(n^{-1/2} + m^{-1/2}(\log\log m)^{1/2}) \quad a.s. \tag{2.31}$$

and

$$|\hat{t}_{n,m}^{fdtp} - t_{n,m}^{fdtp}| = O(n^{-1/2} + m^{-1/2}) \quad \textit{in probability.} \tag{2.32}$$

Here $t_{n,m}^{fdtp}$ is the critical value defined in (A.17).

(ii) If $\hat{t}_{n,m}^{fdr}$ is chosen such that

$$\hat{t}_{n,m}^{fdr} = \inf\{t : \frac{2(1-\pi_1)\bar{\Phi}(t)}{\hat{p}_m(t)} \leq \gamma\}, \tag{2.33}$$

then

$$|\hat{t}^{fdr}_{n,m} - t^{fdr}_{n,m}| = O(n^{-1/2} + m^{-1/2}(\log\log m)^{1/2}) \quad a.s. \tag{2.34}$$

and

$$|\hat{t}^{fdr}_{n,m} - t^{fdr}_{n,m}| = O(n^{-1/2} + m^{-1/2}) \quad \textit{in probability.} \tag{2.35}$$

Here $t^{fdr}_{n,m}$ is the critical value defined in (A.19).

(iii) If $\hat{t}^{k-FWER}_{n,m}$ is chosen such that

$$\hat{t}^{k-FWER}_{n,m} = \inf\{t : P(\zeta(t) \geq k)\} \leq \gamma \tag{2.36}$$

where $\zeta(t) \sim Poisson(\bar{\theta}(t))$ and

$$\bar{\theta}(t) = 2m(1 - \hat{\pi}_1)\bar{\Phi}(t),$$

then

$$|\hat{t}^{k-FWER}_{n,m} - t^{k-FWER}_{n,m}| = O((\log m)^{-1/2}) \quad a.s.. \tag{2.37}$$

Here $t^{k-FWER}_{n,m}$ is the critical value defined in (A.21).

Remark. If $\alpha = \gamma$ in theorem 2.1.4, then it is not difficult to see that $\hat{t}^{fdtp}_{n,m} - \hat{t}^{fdr}_{n,m} = O(m^{-1/2})$ *a.s.*. Therefore (2.31) and (2.32) remain valid with $\hat{t}^{fdtp}_{n,m}$ replaced by $\hat{t}^{fdr}_{n,m}$. This shows that controlling FDTP is asymptotically equivalent to controlling FDR. This is also true in the more general dependence case. Thus we will focus primarily on FDR in our numerical studies.

Remark. Note that π_1 is assumed to be known in order to get a precise rate of convergence for FDTP and FDR. If $\hat{\pi}_1$ is estimated with rate of convergence r_n, then the correct convergence rate for the in probability result for FDR and FDTP would involve an additional term $O(r_n)$ added in (2.32) and (2.35). It is unclear what the correction would be for the almost sure rate in (2.31) and (2.34). These corrections are beyond the scope of this paper and will not be pursued further here. Note that the rate of $\hat{\pi}_1$ is not needed in the main results presented in Sections 2.1–2.3.

CHAPTER 3

Two-sample t-test

3.1 Two-sample t-test

In this section, the results of the previous section are extended to the two-sample t-test setting. The estimator of the unknown parameter π_1 remains the same as in the one-sample case but with T_i in (2.24) being the two-sample rather than one-sample t-statistic. Theoretical results for the rates of convergence under independence are also presented as in the previous section.

3.1.1 Basic set-up and results

When two groups such as a control and experimental group are independent, which we assume here, a natural statistic to use is the two-sample t-statistic. We adopt the same notation used in the one-sample case, as much as possible, and assume that (2.2) holds. We observe the random variables

$$X_{ij} = \mu_i + \epsilon_{ij}, \quad 1 \leq j \leq n_1,\ 1 \leq i \leq m, \quad Y_{ij} = \nu_i + \omega_{ij}, \quad 1 \leq j \leq n_2,\ 1 \leq i \leq m,$$

with the index i denoting the ith gene, j indicating the jth array, μ_i representing the mean effect for the ith gene from the first group, and ν_i representing the mean effect for the ith gene from the second group. The sampling processes for the two groups are assumed to be independent of each other. The sample sizes n_1 and n_2 are assumed to be of the same order, i.e. $0 < b_1 \leq n_1/n_2 \leq b_2 < \infty$. We will also assume that for each i, $\epsilon_{i1}, \epsilon_{i2}, \cdots \epsilon_{in_1}$ are independent random variables with mean zero and variance σ_i^2; $\omega_{i1}, \omega_{i2}, \cdots \omega_{in_2}$ are independent random variables with mean zero and variance τ_i^2. The null hypothesis is $\mu_i = \nu_i$, the alternative hypothesis is $\mu_i \neq \nu_i$, and the dependence is assumed to be generated in the same manner as

the dependence in the one-sample setting. Consider the two-sample t-statistic

$$T_i^* = \frac{\bar{X}_i - \bar{Y}_i}{\sqrt{S_{1i}^2/n_1 + S_{2i}^2/n_2}},$$

where

$$\begin{aligned} \bar{X}_i &= \frac{1}{n_1}\sum_{j=1}^{n_1} X_{ij}, \quad \bar{Y}_i = \frac{1}{n_2}\sum_{j=1}^{n_2} Y_{ij}, \\ S_{1i}^2 &= \frac{1}{n_1 - 1}\sum_{j=1}^{n_1}(X_{ij} - \bar{X}_i)^2, \; S_{2i}^2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}(Y_{ij} - \bar{Y}_i)^2 . \end{aligned}$$

Then

$$R = \sum_{i=1}^{m} 1_{\{|T_i^*| \geq t\}}, \qquad V = \sum_{i=1}^{m}(1 - H_i)1_{\{|T_i^*| \geq t\}}. \tag{3.1}$$

The two-sample t-statistic is one of the most commonly used statistics to construct confidence intervals and do hypothesis testing for the difference between two means. There are several premises underlying the use of two-sample t-tests. It is assumed that the data has been derived from populations with normal distributions. Based on the fact that $S_{1i} \to \sigma_i, S_{2i} \to \tau_i$ a.s., with moderate violation of the assumption, quite often statisticians recommend using the two sample t-test provided the samples are not too small and the samples are of equal or nearly equal size. When the populations are not normally distributed, it is a consequence of the central limit theorem that two-sample t-tests remain valid. A more refined confirmation of this validity under non-normality based on moderate deviations is shown in in Cao (2007). Furthermore, under the alternative hypothesis, the asymptotic results still hold but with a shift in location similar to the one sample case under certain conditions, i.e.,

$$\begin{aligned} P(|T_i^*| \geq t | H_i = 0) &= P(|Z| \geq t)(1 + o(1)), \\ P(|T_i^*| \geq t | H_i = 1) &= P(|Z + \frac{\mu_i - \nu_i}{B_{n_1,n_2}}| \geq t)(1 + o(1)), \end{aligned}$$

uniformly in $t = o(n^{1/6})$, where $B_{n_1,n_2}^2 = \sigma_i^2/n_1 + \tau_i^2/n_2$. Under the assumption of (2.2), asymptotic critical values to control FDTP, FDR and k-FWER are very similar to the one-sample t-test case with the one-sample t-statistic T_i replaced by the two-sample t-statistic

T_i^*. The following theorem, proved in the Appendix, is analogous to Theorem 2.1.1 and is a necessary first step:

Theorem 3.1.1. *Assume that* $E(\epsilon_{ij}|\mu_i,\ \sigma_i^2) = 0$, $E(\omega_{ij}|\nu_i,\ \tau_i^2) = 0$, $Var(\epsilon_{ij}|\mu_i, \sigma_i^2) = \sigma_i^2$, $Var(\omega_{ij}|\nu_i, \tau_i^2) = \tau_i^2$, $\limsup E\epsilon_{ij}^4 < \infty$, $\limsup E\tau_{i,j}^4 < \infty$, $0 < \pi_1 < 1 - \alpha$ *and (2.2) is satisfied. Assume that there exist* ϵ_0 *and* c_0*, such that*

$$P(|\frac{\mu_i - \nu_i}{B_{n_1,n_2}}| \geq \epsilon_0 | H_i = 1) \geq c_0 \quad \textit{for all } n_1, n_2. \tag{3.2}$$

Then the conclusions of Theorem 2.1.1 hold with the one-sample t-statistic T_i *replaced by the two-sample t-statistic* T_i^*.

3.1.2 Main Results

The unknown parameter m_1 and functions $F_0(t)$ and $F_1(t)$ in Theorem 3.1.1 are estimated similarly as in the one-sample case with the one-sample t-statistic replaced by its two-sample counterpart. The following theorem, the proof of which is given in the Appendix, gives our main results for two-sample t-tests:

Theorem 3.1.2. *Assume the conditions in Theorem 3.1.1 are satisfied. Replace the one-sample t-statistic* T_i *by the two-sample t-statistic* T_i^* *in Theorem 2.1.2. Let* $\hat{\pi}_1$ *be a strong consistent estimate of* π_1 *as in (2.28) using the two-sample t-statistic* T_i^*.

(i) If $\hat{t}_{n,m}^{fdtp}$ *is chosen such that*

$$\hat{t}_{n,m}^{fdtp} = \inf\{t : \frac{\sqrt{m}\nu_m(t)}{\tau_m(t)} \geq z_\gamma\}, \tag{3.3}$$

then

$$|\hat{t}_{n,m}^{fdtp} - t_{n,m}^{fdtp}| = o(1) \quad a.s. \tag{3.4}$$

(ii) If $\hat{t}_{n,m}^{fdr}$ *is chosen such that*

$$\hat{t}_{n,m}^{fdr} = \inf\{t : \frac{2(1-\hat{\pi}_1)\bar{\Phi}(t)}{\hat{p}_m(t)} \leq \gamma\} \tag{3.5}$$

$$|\hat{t}_{n,m}^{fdr} - t_{n,m}^{fdr}| = o(1) \quad a.s. \tag{3.6}$$

(iii) If $\hat{t}_{n,m}^{k-FWER}$ *is chosen such that*

$$\hat{t}_{n,m}^{k-FWER} = \inf\{t : P(\zeta(t) \geq k)\} \leq \gamma \tag{3.7}$$

where $\zeta(t) \sim Poisson(\bar{\theta}(t))$ *and*

$$\bar{\theta}(t) = 2m(1 - \hat{\pi}_1)\bar{\Phi}(t),$$

then as long as $\log m = o(n^{1/3})$

$$|\hat{t}_{n,m}^{k-FWER} - t_{n,m}^{k-FWER}| = o(1) \quad a.s. \tag{3.8}$$

Remark. $\hat{\pi}_1$ can be estimated through (2.28) by using two-sample t-statistics. Theorem 2.1.3 is applicable in the two-sample setting as well as in the one-sample case, and consistency follows. Thus theorem 3.1.2 gives a fully automated procedure to conduct multiple hypothesis testing using two-sample t-statistics after we plug in the $\hat{\pi}_1$ given in (2.28).

3.1.3 Consistency and rate of convergence under independence

Results for the independence setting are needed for the proofs of the main results, as was the case for one-sample t-tests. We can, once again, obtain more precise estimation compared with the general dependence case. The following theorem, proven in the Appendix, gives us conditions and conclusions using two-sample t-statistics for controlling FDTP and FDR asymptotically as well as rates of convergence under the assumption that (T_i, H_i) are independent of each other for $1 \leq i \leq m$. Assume π_1 is the proportion of the alternative hypotheses among m hypothesis test, i.e., $\pi_1 = P(H_i = 1)$. Let $\mathcal{J} = \{i : H_i = 1\}$.

Theorem 3.1.3. *Assume the conditions of Theorem 3.1.1 are satisfied. Rather than (2.2), we assume that* (T_i, H_i) *are independent and identically distributed. In addition,* $\pi_1 = P(T_1 = 1)$ *and* μ_i, σ_i *are i.i.d. for* $i \in \mathcal{J}$*. Let*

$$p(t) = P(|T_1^*| \geq t), \tag{3.9}$$

$$a_1(t) = \alpha p(t) - (1-\pi_1)P(|T_1^*| \geq t | H_1 = 0), \tag{3.10}$$

$$\begin{aligned} b_1^2(t) &= \alpha^2 p(t)(1-p(t)) + 2\alpha(1-\pi_1)p(t)P(|T_1^*| \geq t|H_1=0) \\ &\quad +(1-\pi_1)P(|T_1^*| \geq t|H_1=0)(1-2\alpha-(1-\pi_1)P(|T_1^*| \geq t|H_1=0)), \end{aligned}$$

$$\hat{p}_m(t) = \frac{1}{m}\sum_{i=1}^{m} I_{\{|T_i^*| \geq t\}}, \tag{3.11}$$

$$\nu_m(t) = \alpha \hat{p}_m(t) - 2(1-\pi_1)\bar{\Phi}(t), \tag{3.12}$$

and

$$\begin{aligned} \tau_m^2(t) &= \alpha^2 \hat{p}_m(t)(1-\hat{p}_m(t)) + 4\alpha(1-\pi_1)\hat{p}_m(t)\bar{\Phi}(t) \\ &\quad +2(1-\pi_1)\bar{\Phi}(t)(1-2\alpha-2(1-\pi_1)\bar{\Phi}(t)). \end{aligned}$$

Then the conclusions of Theorem 2.1.4 hold with the one-sample t-statistics T_i replaced by the two-sample t-statistics T_i^*.

Remark. In the above sections, we developed our theorems based on two-sided tests. The results for the case of one sided tests are very similar but with rejection region $\{T_i \geq t\}$ for each test. We omit the details.

CHAPTER 4

Numerical Studies

4.1 Simulations

In this section, we present numerical studies based on simulated data and compare the power of our approach with Benjamini and Hochberg (1995)(BH) and Storey and Tibshirani (2003)(ST) approaches using one-sample t-statistics. The results for using two-sample t-statistics are very similar and we omit the details here.

4.1.1 Asymptotic sample path

We investigate the results for the i.i.d. case first. Recall the model

$$X_{ij} = \mu_i + \epsilon_{ij}, \quad 1 \le i \le m, \quad 1 \le j \le n.$$

We set the signal using $\mu_i \sim Unif(0.5, 1)$ or $\mu_i \sim Unif(-1, -0.5)$, which is of the right order for the standardized error term. Here, the number of hypothesis tests is $m = 10,000$, which is the same for all following simulation studies unless otherwise noted, the proportion of alternatives $\pi_1 = 0.2$ and the error term $t(4)$ are used just to illustrate the asymptotic results. We vary the number of arrays n from 20, 50 to 300 to evaluate our asymptotic approximation. Empirical distributions of FDTP, FDR and k-FWER based on $100,000$ repetitions are treated as the gold standard since it has almost negligible Monte Carlo error. The samples are generated to evaluate our proposed method based on asymptotic theory. Specifically, for each sample, we calculate the sample paths of the following quantities indexed by t: $\sqrt{m}\nu_m(t)/\tau_m(t)$ for studying FDTP, $2(1-\hat{\pi}_1)\bar{\Phi}(t)/\hat{p}_m(t)$ for studying FDR and $P(Poisson(2m(1-\hat{\pi}_1)\bar{\Phi}(t)) \ge 10)$ for studying 10-FWER (Here we pick $k = 10$ just for illustration). $\hat{\pi}_1$ is defined as in (2.28).

Figure 4.1: Overlay of true and 100 random estimated sample paths with respect to cut-off t for the three procedures under differing sample sizes.

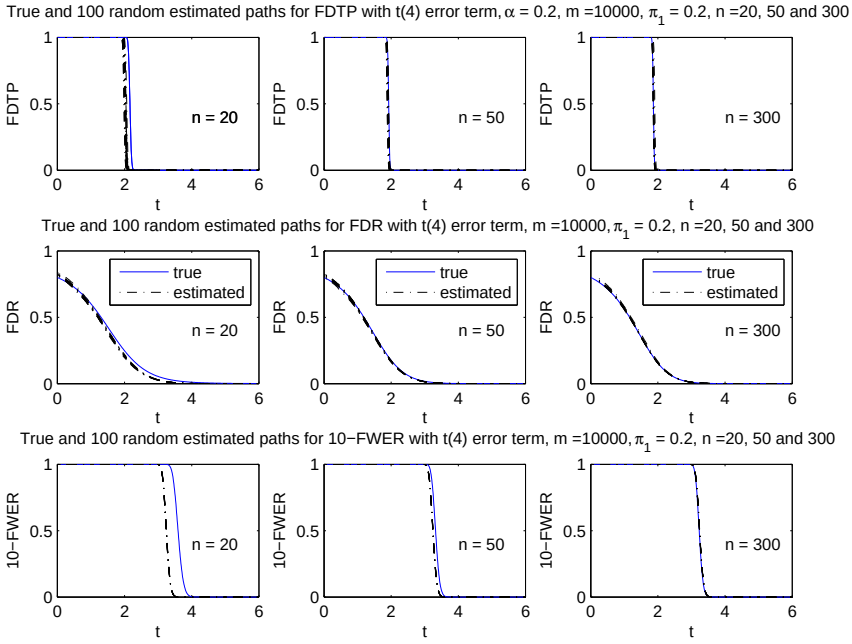
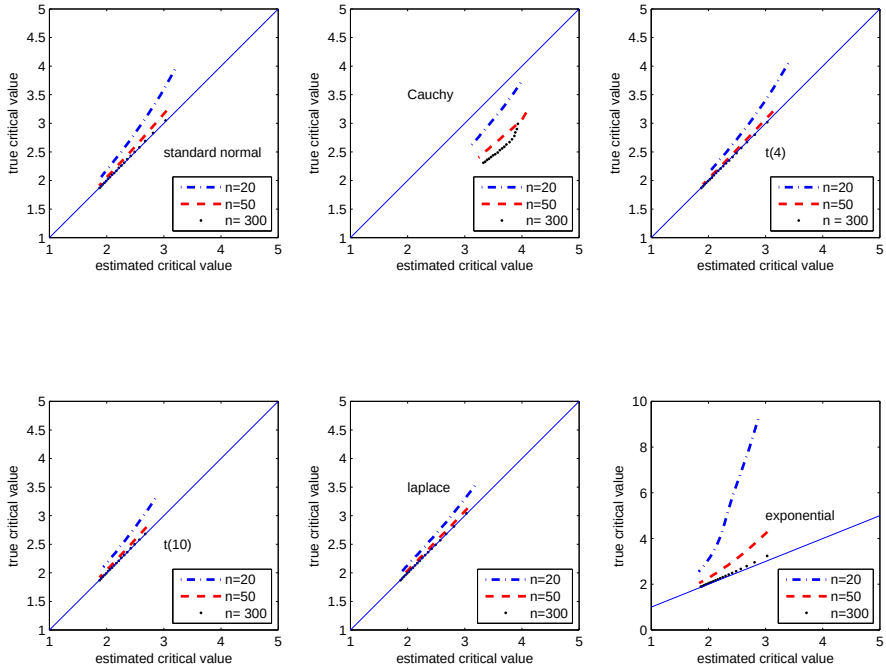


Figure 4.1 shows the overlay of the true path and 100 random estimated paths for FDTP FDR and k-FWER respectively. As n increases, we see that the true path and estimated paths are pretty close to each other, which in turn validates our asymptotic theory. We can see that the slope of FDTP and 10-FWER are very steep, which means a small change in the critical value results in a large change in the level of control, while the FDR has a flatter trend.

4.1.2 Robustness to different error terms

Under the same setup as in the previous section, we simulate data with different error terms: standard normal($N(0,1)$), student t with 1 degree of freedom (Cauchy), student t with 4 degree of freedom ($t(4)$), student t with 10 degree of freedom ($t(10)$), Laplace and exponential. Note that except for the Cauchy error term, all the remaining error terms satisfy the condition of finite 4th moment. Empirical distributions of FDTP FDR and k-FWER based on $100,000$ repetitions are treated as the gold standard to obtain true critical values. Each scenario is repeated 1000 times to evaluate our proposed method for estimating the critical value based on asymptotic theory. We control FDR at different levels (from 0.01 to 0.2) to get true and estimated critical values. Asymptotically, the estimated critical value $\hat{t}$ based on our theory should be very close

Figure 4.2: Comparison of true and estimated critical values using FDR for different error terms and numbers of arrays n.

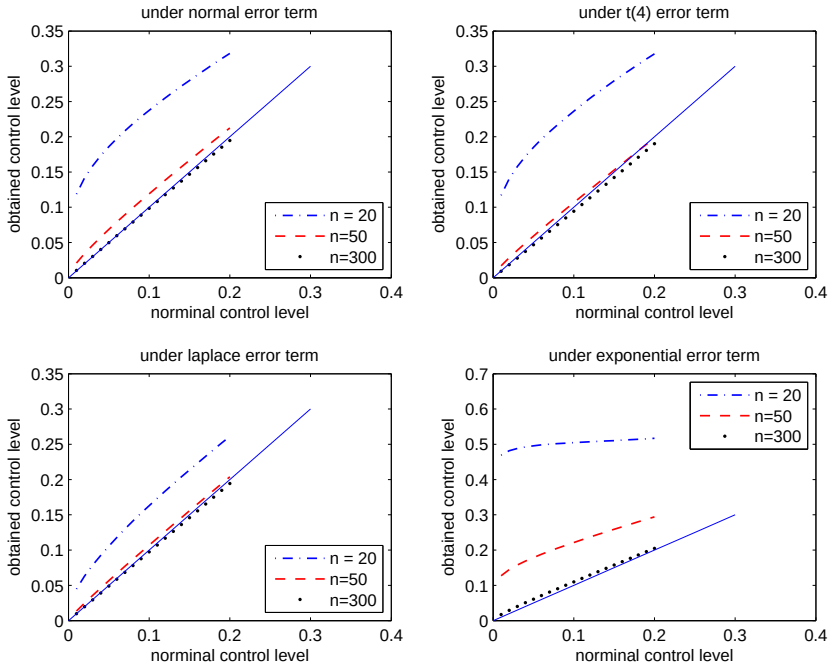


to the true critical value t and lie on a diagonal line of the square. From Figure 3, the estimated critical values $\hat{t}$ do not match the true critical value t under the Cauchy error since the Cauchy distribution does not have finite 4th moment. For the Cauchy distribution, even the central limit theorem does not hold since it does not have finite mean. As the number of arrays n increases, the estimated critical values $\hat{t}$ match the true critical values t better under symmetric error terms ($N(0,1), t(4), t(10)$ and Laplace) but not quite so well under asymmetric errors (e.g., exponential errors). The difficulty with the exponential error terms suggest the value of conducting research to derive higher order approximations. We plan on undertaking this in the near future.

4.1.3 Prescribed FDR control

The above results are from the independent test setting. We did similar simulation studies for the dependent setting, and found that the corresponding plots are quite similar to the above results and the same conclusions can be drawn. To see whether our proposed method obtains the claimed level of control, we use a hidden Markov chain to generate dependent indicators $H_i, i = 1, \cdots, m$. Conditional on $H_i, i = 1, \cdots, m$, the data is generated independently. The

Figure 4.3: Comparison of nominal and obtained control level for different error terms and numbers of arrays n.



transition probability of the hidden Markov chain is set to

$$\begin{pmatrix} 1-p_1 & p_1 \\ p_0 & 1-p_0 \end{pmatrix},$$

where p_1 is the transition probability from 0 to 1 and p_0 is the transition probability from 1 to 0. In the simulation, $p_0 = 0.8$ and $p_1 = 0.2$. Based on the limiting stationary distribution, the alternative proportion should be $\pi_1 = p_1/(p_0 + p_1)$. Under the null hypothesis, we simulate data from four error terms ($N(0,1)$, $t(4)$, Laplace and exponential); and under the alternative hypothesis, we simulate data with mean effects half from $Unif(0.1, 0, 5)$ and half from $Unif(-0.5, -0.1)$ plus the same four error terms. Figure 4.3 uses FDR as the control criterion. For different control levels γ, we compare the claimed level of control and the actually obtained level of control based on our method for different numbers of arrays: small ($n = 20$), medium ($n = 50$) and large ($n = 300$).

From Figure 4.3, we can see that when the number of arrays n is small ($n = 20$), we do not in general achieve the claimed level of control. If we have a medium sample size ($n = 50$), the

obtained level of control is very close to the nominal level of control and the results are almost perfect if we have a large number of arrays ($n = 300$), even for the asymmetric exponential error term. This strongly supports our theoretical predictions but suggests that higher order approximations would be useful in some settings.

To see the performance of our method using 10-FWER Table 4.1 summarizes the actually obtained control level for different error terms and numbers of arrays n when the nominal control level is 0.05. The obtained control level is incorrect when the number of arrays n is small, which

Table 4.1: Obtained control level using 10-FWER with nominal control level 0.05.

n	$N(0,1)$	$t(4)$	Laplace	exponential
20	0.998 (9.0e-05)	0.90 (7.0e-03)	0.81 (1.1e-02)	1(0)
50	0.52 (1.2e-02)	0.14 (9.1e-03)	0.17 (1.2e-02)	1 (0)
300	0.076 (3.8e-03)	0.031 (2.8e-03)	0.05 (2.7e-03)	0.82 (4.6e-03)

can be deduced from the samples paths of 10-FWER given in Figure 1.1. It has a very steep slope, so that when n is small, the approximation is crude and there is a noticeable difference between the estimated critical value and the true critical value, yielding a big difference in the control level. For large sample sizes, the obtained control level is reasonably good because our asymptotic theory begins to take effect. The exponential error setting appears to not perform as well as the other error settings.

4.1.4 Estimation accuracy of π_1

All previous numerical studies involve the alternative proportion estimate $\hat{\pi}_1$ defined in (2.28). In this section, we investigate numerically how this estimate is affected by number of arrays n and compare with the alternative estimate proposed by Storey and Tibshirani (2003). The first simulation setup is similar to the one in the previous section. We drew $N = 1000$ sets of data as follows. Dependent indicators $H_i, i = 1, \cdots, m$ are generated from a hidden Markov chain with the limiting alternative proportion $\pi_1 = 0.2$. Conditional on these, a vector of expected values, $\mu = (\mu_1, \cdots, \mu_m)$, was constructed. The expected values for the true null hypotheses were set to 0 with standard normal noise, whereas the expected values for the alternative hypotheses were draw from $Unif(0.1, 0.5)$ plus standard normal noise. Correspondingly, 1000 replications

of the proportion estimate $\hat{\pi}_1$ were calculated by using (2.28). The RMSE is given as

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(\hat{\pi}_1^{(n)} - \pi_1^{(n)})^2},$$

where $\hat{\pi}_1^{(n)}$ is the estimate of π_1 for the nth simulated data set and $\pi_1^{(n)}$ is the truth. Table 4.2 summarizes the effect of n. As the number of arrays n increases, the RMSE gets smaller, which validates our asymptotic prediction.

Table 4.2: RMSE for $N = 1000$ estimated values of π_1.

n	20	50	300
RMSE	0.0156	0.0136	0.0104

In the second simulation, we compare our proportion estimate with the one using spline smoothing proposed by Storey and Tibshirani (2003). Recall the proportion estimate $\pi_0(\lambda) = \#\{p_i > \lambda; i = 1, \cdots, m\}/(m(1-\lambda))$. The smoothing approach proceeds as follows: first $\pi_0(\lambda)$ are calculated over a (fine) grid of λ; then, a natural cubic spline y with 3 degree of freedom is fitted to $(\lambda, \hat{\pi}_0(\lambda))$; finally, π_0 is estimated by $\hat{\pi}_0 = y(1)$. The simulation setup is similar to the previous one except that we have two groups here with $n1 = 70$ and $n2 = 80$. We change the alternative proportion to compare the performances of our approach (π_1^{ck}) with the spline smoothing approach (π_1^{st}) in Table 4.3. They produce very similar results, both are conservative, with less bias using our approach and less variance using the spline smoothing approach. The advantage of our approach is that it is computationally very fast, while the spline smoothing approach requires obtaining p-values using permutation first, which is computationally much more intensive than our approach which can be computed directly from the t-statistics.

Table 4.3: Proportion estimate comparison

π_1	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
$\hat{\pi}_1^{ck}$	0.044	0.091	0.141	0.182	0.217	0.255	0.289	0.335	0.365
$\hat{\pi}_1^{st}$	0.041	0.081	0.125	0.161	0.195	0.236	0.276	0.323	0.355
$sd(\hat{\pi}_1^{ck})$	0.042	0.043	0.041	0.040	0.046	0.041	0.047	0.042	0.038
$sd(\hat{\pi}_1^{st})$	0.039	0.041	0.036	0.040	0.041	0.038	0.034	0.036	0.031

4.2 Comparison with BH and ST procedure

In this section, we compare our approach with the BH and ST procedures under the dependence structure described in Wu (2008)'s paper. We also use a Hidden Markov model to simulate the indicator function $H_i, i = 1, \cdots, m$. Conditional on $H_i, i = 1, \cdots, m$, the data is generated independently. The number of hypotheses tested $m = 5000$ and the number of arrays $n = 80$. The data generating mechanism is otherwise the same as in the independence case. First, we construct a one-sample t-statistic and apply our procedure to get the critical value for the rejection region. We then obtain p-values and q-values, and apply the BH and ST procedures to decide which genes are significantly expressed. We now briefly describe the BH procedure. Let p_i be the marginal p-value of the ith test, $1 \leq i \leq m$, and let $p_{(1)} \leq \cdots \leq p_{(m)}$ be the order statistics of $p_1, \cdots p_m$. Given a control level $\gamma \in (0, 1)$, let

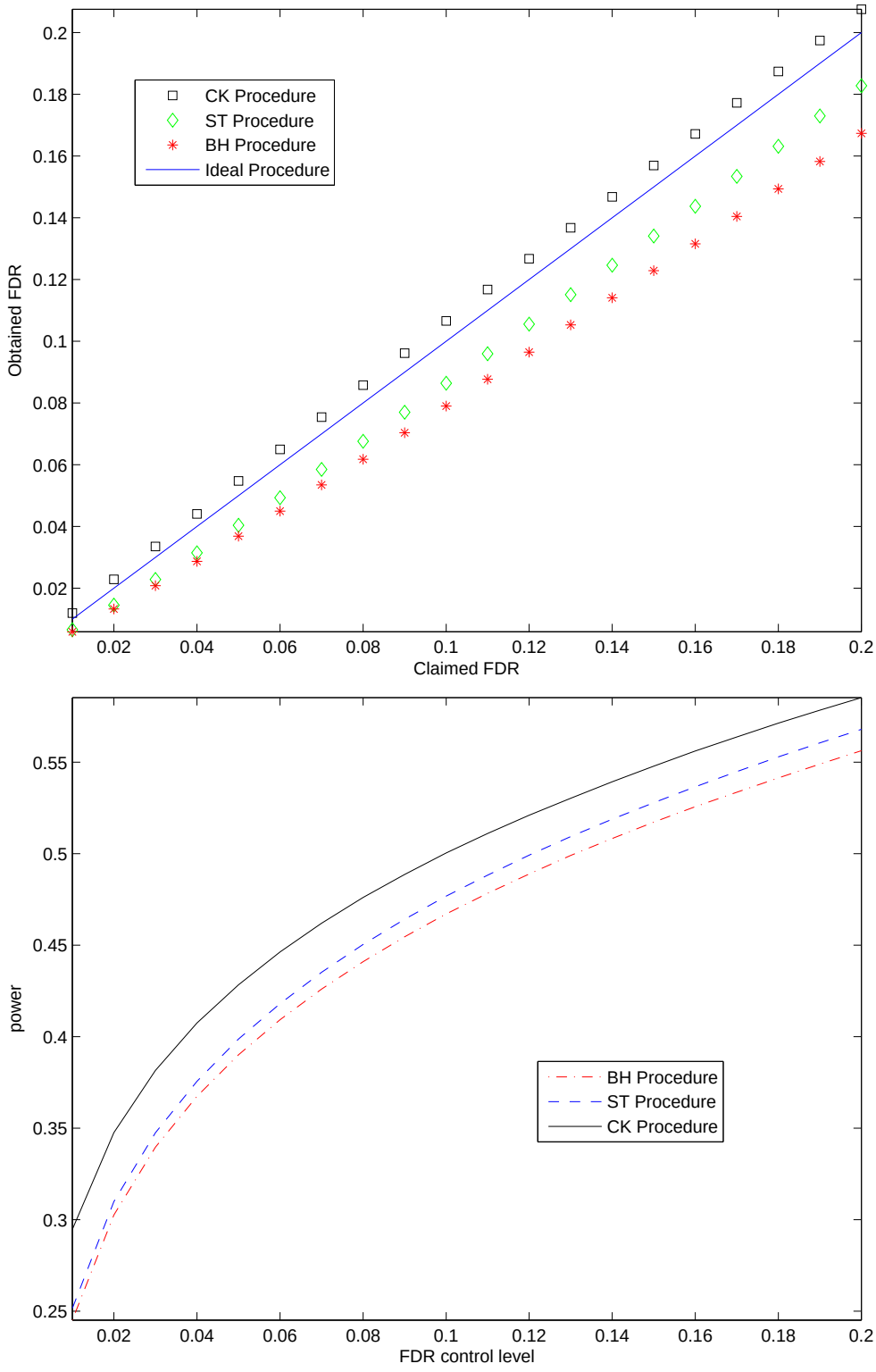
$$r = \max\{i \in \{0, 1, \cdots, m+1\} : p_{(i)} \leq \gamma i/m\},$$

where $p_0 = 0$ and $p_{(m+1)} = 1$. The BH procedure rejects all hypotheses for which $p_{(i)} \leq p_{(r)}$. If $r = 0$, then all hypotheses are accepted. The q-value in ST's paper is similar though to the well known p-value, except it is a measure of significance in terms of FDR rather than type I error and an estimate of alternative proportion is plugged in based on available p-values as described in the previous section. We revisit the motivating example and give a plot of the claimed FDR and actually obtained FDR by using the proposed critical value method. From Figure 4.4, we can see that our procedure controls the FDR at the claimed level asymptotically, though a little bit liberally for finite samples, and has better power at the same target FDR level compared with the BH and ST procedures. especially when the proportion of alternatives exceeds 0.1.

4.3 Applications to microarray analysis

We now apply the proposed procedure to the analysis of a leukemia cancer data set (Golub (1999)) in order to identify differentially expressed genes between AML and ALL. For the original data, please see `http://www.broad.mit.edu/cgi-bin/cancer/datasets.cgi`. In this analysis, we use the methodology developed for the dependence case. The raw data consist of $m = 7129$ genes and 72 samples coming from two classes: 47 in class ALL (acute myeloid

Figure 4.4: Power comparison and FDR control



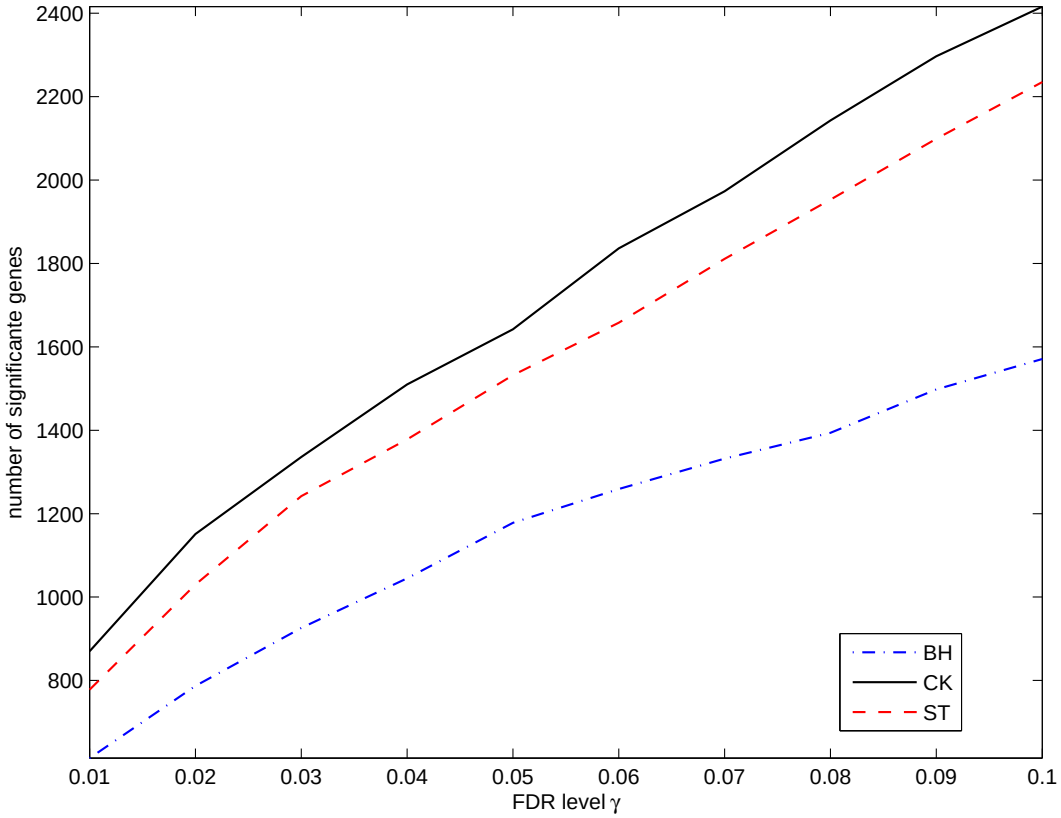
leukemia) and 25 in class AML (acute lymphoblastic leukemia). Our simulation results showed reasonable performance of the procedure for moderate sample size in this range. For each gene location, the two-sample t-statistic comparing the 47 ALL responses with the 25 AML responses was computed. Using our proposed approach for the dependent case, we find the critical value for controlling FDR at level γ:

$$\hat{t}_{n,m}^{fdr} = \inf\{t : \frac{2(1-\hat{\pi}_1)\bar{\Phi}(t)}{\hat{p}_m(t)} \leq \gamma\},$$

where $\hat{p}_m = \sum_{i=1}^{m} 1_{\{|T_i| \geq t\}}/m$ and $\hat{\pi}_1$ is estimated by (2.28).

In Figure 4.5, we plot the FDR level and the number of significantly expressed genes by our procedure (CK), BH procedure and the q-value based Storey Tibshirani (ST) procedure. From the plot, we can see that our procedure detects the largest number of significant genes, followed by the ST procedure and then the BH procedure, which is the most conservative one. At FDR level 0.01, we detected 870 genes, the ST procedure detected 778 gens and the BH procedure detected 614 genes. Using the two-sample t-test, similar to the higher power of our approach in simulation studies, we detected all of the genes that the other two approaches detected. The BH procedure is very conservative at the expense of power loss. The ST procedure requires permutation to get p-values, while our procedure gets the critical value directly, and thus is faster in terms of computation. The estimation of π_1 is 0.467 by our procedure and 0.477 by the ST procedure. These results can serve as a first exploration step for more refined analyses concerning these significant genes. Another issue may be that the critical value approach based on asymptotic FDR control may not be conservative enough in some settings.

Figure 4.5: Comparison between our procedure (CK), the ST procedure and the BH procedure in real data.



CHAPTER 5

Counter example for monotone likelihood ratio condition

5.1 Background

The concavity of p-value distribution under the alternative has been a standard condition for developing many FDR procedures: Storey; 2003, Genovese and Wasserman; 2004, Kosorok and Ma; 2007. A more general concept is the monotone likelihood ratio condition (MLRC) introduced in Sun and Cai (2007). We show in this chapter that the concavity assumption can be violated for (i) a simple heteroscedastic normal mixture model and (ii) dependent tests. Some interesting implications, including different testing procedures (step-up vs step down), the choice of test statistic and the power definition in multiple testing are discussed.

Consider a random mixture model

$$T_i \overset{i.i.d}{\sim} (1-\pi_1)F_0 + \pi_1 F_1, \quad i = 1, \cdots, m.$$

We can think of this asymptotically, T_i has limiting distribution F_0 under the null and F_1 under the alternative with a prior belief that among the m hypotheses, π_1 are from the alternative. Let $H_1, \cdots, H_m$ be the associated unknown states with $H_i = I(T_i \text{ comes from } F_1)$. A FDR procedure based on test statistics T in general has the following form

$$\boldsymbol{\delta}(T, c) = \{I[T > c] : i = 1, \cdots, m\}.$$

5.2 The monotone likelihood ratio condition

In the multiple testing literature, it is often assumed that

$$\text{the FDR of } \boldsymbol{\delta}(T, c) \text{ is monotonically decreasing in the cutoff } c. \tag{5.1}$$

When p-values are used, a sufficient condition for (5.1) to hold is that

$$G_P^1(t) \text{ is concave,} \tag{5.2}$$

where $G_P^1(t)$ is the p-value distribution under the alternative. The concavity of $G_P^1(t)$ has been assumed in Storey (2003) and Genovese and Wasserman (2004). A generalized condition was considered for a family of test statistics $\mathcal{T}$ in Sun and Cai (2007). Specifically, let G_T^0 and G_T^1 be the conditional cdf of T under the null and alternative, respectively. Denote by g_T^0 and g_T^1 the corresponding density functions. A sufficient condition for (5.1) to hold is the following monotone likelihood ratio condition (MLRC):

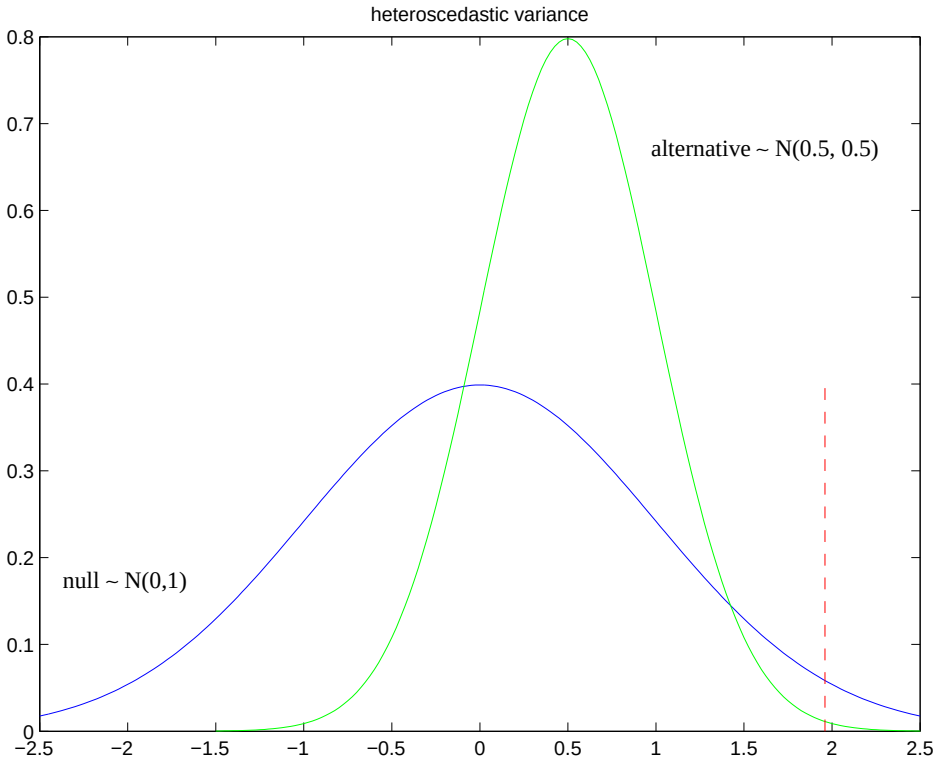
$$g_T^1(c)/g_T^0(c) \text{ is monotonically decreasing in } c. \tag{5.3}$$

Note that G_P^0 is uniform, and it is easy to verify that (5.2) implies (5.3) when p-values are used.

This MLRC holds if the null and alternative distribution have the same spread. However, for heteroscedastic variance, it fails. We use a single hypothesis test to illustrate. Suppose under the null, that the test statistic follows a $N(0, 1)$ distribution and under the alternative, the test statistic follows a $N(0.5, 0.5)$. If we control the traditional 0.025 tail probability, the critical value is 1.96. But at 1.96, the probability that the observation comes from the alternative is 0.0018, much smaller than the probability that it comes from the null which is 0.025. In fact, it is 13.89 times more likely that the observation comes from the null rather than the alternative.

The concavity of the alternative distribution is useful in deriving the optimal testing procedure—maximizing the power at the same level of FDR control. The monotone likelihood assumption is a generalization for the concavity of p-value distribution under the alternative.

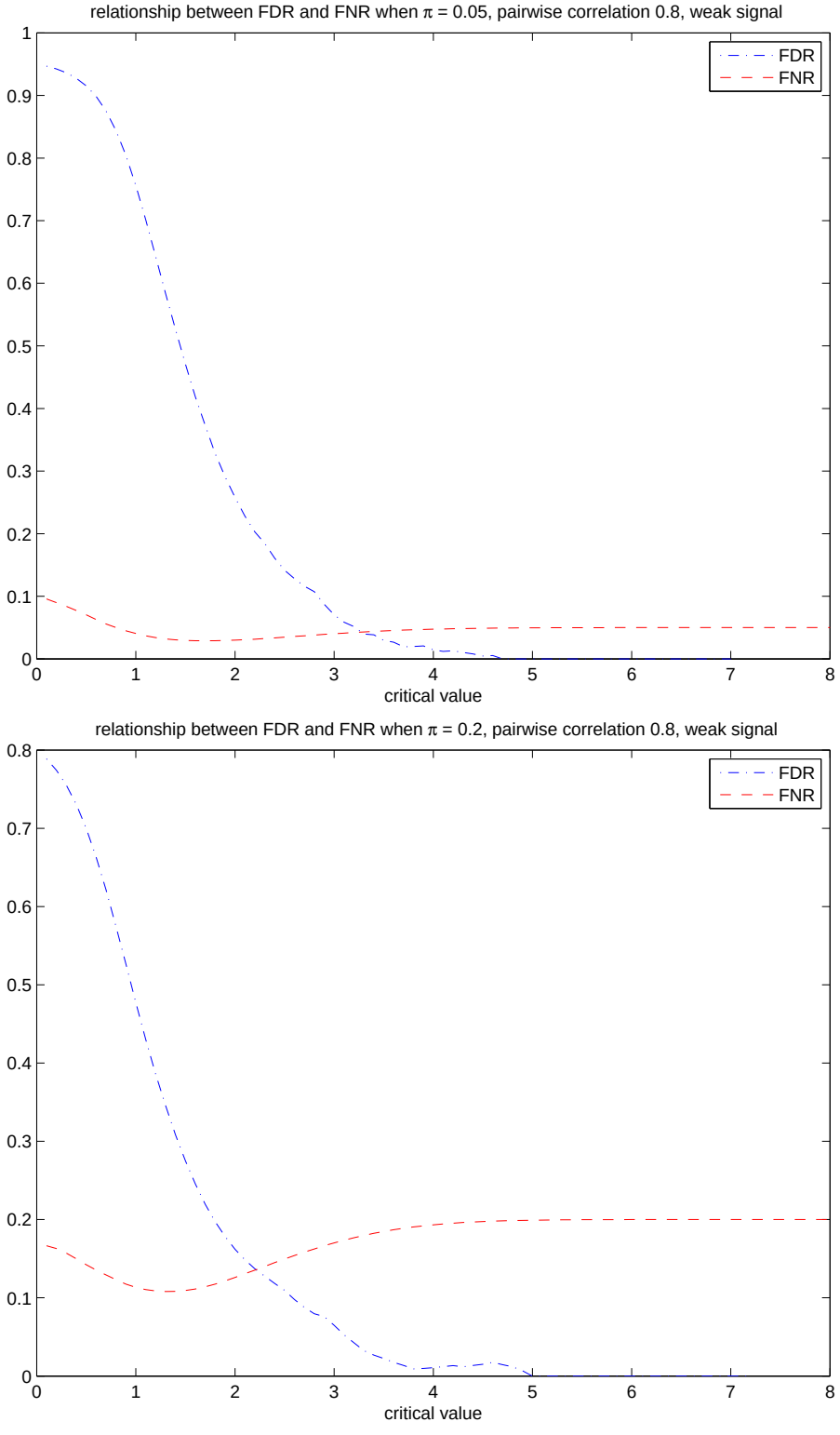
Figure 5.1: Heteroscedastic variance between null and alternative



Genovese and Wasserman (2002) and independently Sarkar (2002) proposed to use false non-discovery rate FNR defined as $E(F/W)$ as a detection ability criterion in multiple testing. Later on, various procedures were developed to optimize this FNR on the same level of FDR control, see Sun and Cai (2007), etc. But the key assumption for these procedure is that the FNR is a monotone non-increasing function of FDR which does not necessarily hold, especially when the signal is weak no matter what the proportion of alternatives is.

In a simulation, we now study the trajectory of FDR and FNR with respect to critical value t based on the model $x_{ij} = \mu_i + \epsilon_{ij}$ with null hypothesis $\mu_i = 0$ and alternative hypothesis $\mu_i \neq 0$. Here $\epsilon_{ij} \sim N(0,1)$ marginally and $Corr(\epsilon_{ij}, \epsilon_{ik}) = \rho$ when $j \neq k$. For this simulation, we use multivariate normal to simulate the data with pairwise correlation 0.8, the proportion for alternatives is 0.05 and 0.2 respectively and we use Unif$(0.1, 0.5)$ as a weak alternative signal and Unif$(0.5, 1)$ as a strong alternative signal. From Figure 5.2 , we can see that the FDR decreases with respect to critical value while FNR decreases first and then increase with respect to critical value for both small and large proportions of alternatives when the signal is

Figure 5.2: Non-monotonicity between FDR and FNR under positive correlation for weak signal



weak under strong correlation.

When the signal is strong, the monotonicity relationship between FDR and FNR holds and under weak signal, it holds if the tests are independent or the correlation is weak. See Figure 5.3.

Due to this non-monotonicity between FDR and FNR we propose to use the non-discovery rate (NDR) defined as NDR $= E(F/m_1)$ for the power definition in multiple testing. The denominator m_1 is a fixed but unknown parameter and the denominator F represents the total number of missed discoveries, which decreases monotonically as the critical value increases, and correspondingly the total number of true rejections increases.

5.3 Counter examples

5.3.1 One-sided Z test

This section gives an example where the concavity of G_P^1 does not hold in a multiple testing situation. Consider a two component normal mixture

$$T_i \overset{i.i.d.}{\sim} (1-\pi_1)N(0,1) + \pi_1 N(\mu, \sigma^2). \tag{5.4}$$

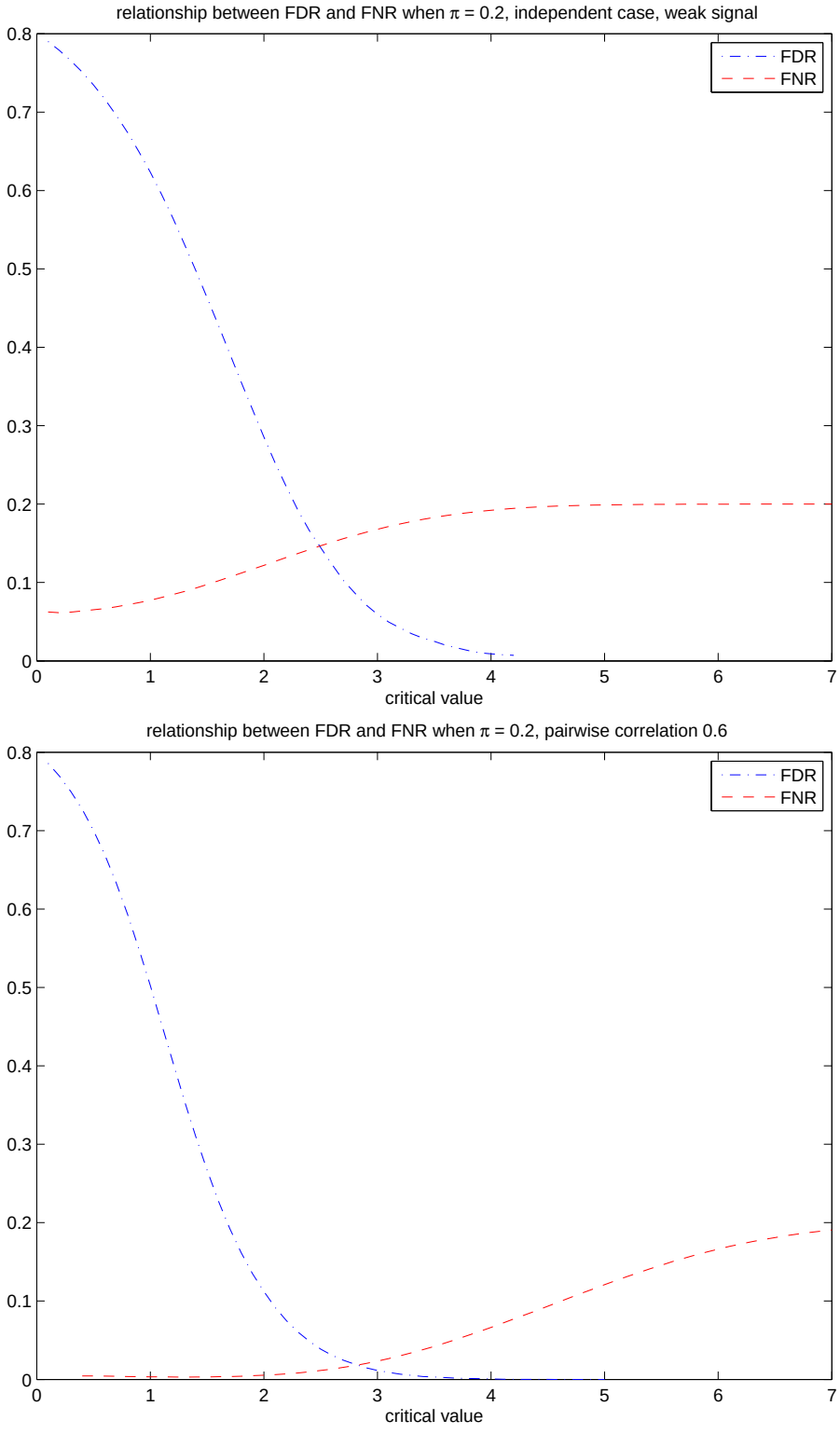
The null hypothesis $\mu = 0 \quad vs \quad \mu > 0$ and take $P_i = P\{N(0,1) > T_i\}$. Denote by Φ and ϕ the cdf and pdf of a standard normal variable, respectively. Observe that

$$G_P^1(t) \quad = \quad P_{\theta_i=1}(P_i < t) = \Phi\left\{\frac{\Phi^{-1}(t)+\mu}{\sigma}\right\},$$

the conditional pdf of the p-value is

$$\begin{aligned} g_P^1(t) &= \frac{1}{\sigma}\phi\left\{\frac{\Phi^{-1}(t)+\mu}{\sigma}\right\}/\phi\{\Phi^{-1}(t)\} \\ &= \begin{cases} (1/\sigma)\exp\left[-\frac{1-\sigma^2}{2\sigma^2}\left\{\Phi^{-1}(t)+\frac{\mu}{1-\sigma^2}\right\}^2 + \frac{\mu^2}{2(1-\sigma^2)}\right] & \text{if } \sigma < 1 \\ (1/\sigma)\exp\left[\frac{\sigma^2-1}{2\sigma^2}\left\{\Phi^{-1}(t)-\frac{\mu}{\sigma^2-1}\right\}^2 - \frac{\mu^2}{2(\sigma^2-1)}\right] & \text{if } \sigma > 1 \\ \exp\left\{-\Phi^{-1}(t)\mu - \frac{1}{2}\mu^2\right\} & \text{if } \sigma = 1 \end{cases} \end{aligned}$$

Figure 5.3: Monotonicity between FDR and FNR under independence and positive correlation for strong signal



The critical region for inference is the interval $t \in (0, \eta)$, where η is usually very small. In order to guarantee that $G_P^1(t)$ is concave, $g_P^1(t)$ should be decreasing in t. It is easy to see that $g_P^1(t)$ is a decreasing function for $t \in (0, \eta)$ when $\sigma \geq 1$. However, $g_P^1(t)$ is increasing in t for $t < \Phi\left\{-\mu/(1-\sigma^2)\right\}$ when $\sigma < 1$, which implies that G_P^1 is not concave. ($\Phi^{-1}(t) < \Phi^{-1}(\eta) < \Phi^{-1}(1/2) = 0$) The MLRC (5.3) also fails. Some further analysis reveals that the FDR can be decreasing in t.

5.3.2 Two-sided Z test

Consider a three component normal mixture

$$X_i \overset{i.i.d.}{\sim} (1-p_1-p_2)N(0,1) + p_1 N(\mu_1, \sigma_1^2) + p_2 N(\mu_2, \sigma_2^2). \tag{5.5}$$

Define the two-sided p-value $P_i = P\{|N(0,1)| > |X_i|\} = 2\Phi(-|X_i|)$. Therefore

$$\begin{aligned} G_P^0(t) &= P_0\{|X_i| > -\Phi^{-1}(t/2)\} \\ &= P_0\{X_i < \Phi^{-1}(t/2)\} + P_0\{X_i > -\Phi^{-1}(t/2)\} \\ &= t. \end{aligned}$$

and

$$\begin{aligned} G_P^1(t) &= \frac{p_1}{p_1+p_2}\left[\Phi\left\{\frac{\Phi^{-1}(t/2)-\mu_1}{\sigma_1}\right\} + \Phi\left\{\frac{\Phi^{-1}(t/2)+\mu_1}{\sigma_1}\right\}\right] \\ &\quad + \frac{p_2}{p_1+p_2}\left[\Phi\left\{\frac{\Phi^{-1}(t/2)-\mu_2}{\sigma_2}\right\} + \Phi\left\{\frac{\Phi^{-1}(t/2)+\mu_2}{\sigma_2}\right\}\right] \end{aligned}$$

Similarly, for $\sigma_i \neq 1$, we have

$$\begin{aligned} g_P^1(t) &= \frac{p_1}{p_1+p_2}\left[\frac{1}{2\sigma_1}\exp\left\{-\frac{1-\sigma_1^2}{2\sigma_1^2}\left(\Phi^{-1}\left(\frac{t}{2}\right) - \frac{\mu_1}{1-\sigma_1^2}\right)^2 + \frac{\mu_1^2}{2(1-\sigma_1^2)}\right\}\right] \\ &\quad + \frac{p_1}{p_1+p_2}\left[\frac{1}{2\sigma_1}\exp\left\{-\frac{1-\sigma_1^2}{2\sigma_1^2}\left(\Phi^{-1}\left(\frac{t}{2}\right) + \frac{\mu_1}{1-\sigma_1^2}\right)^2 + \frac{\mu_1^2}{2(1-\sigma_1^2)}\right\}\right] \\ &\quad + \frac{p_2}{p_1+p_2}\left[\frac{1}{2\sigma_2}\exp\left\{-\frac{1-\sigma_2^2}{2\sigma_2^2}\left(\Phi^{-1}\left(\frac{t}{2}\right) - \frac{\mu_2}{1-\sigma_2^2}\right)^2 + \frac{\mu_2^2}{2(1-\sigma_2^2)}\right\}\right] \end{aligned}$$

$$+\frac{p_2}{p_1+p_2}\left[\frac{1}{2\sigma_2}\exp\left\{-\frac{1-\sigma_2^2}{2\sigma_2^2}\left(\Phi^{-1}\left(\frac{t}{2}\right)+\frac{\mu_1}{1-\sigma_2^2}\right)^2+\frac{\mu_2^2}{2(1-\sigma_2^2)}\right\}\right]$$

For $\sigma_1=\sigma_2=1$, we have

$$\begin{aligned} g_P^1(t) &= \frac{p_1}{p_1+p_2}\left[\frac{1}{2}\exp\left\{\Phi^{-1}(t/2)\mu_1-\mu_1^2/2\right\}+\frac{1}{2}\exp\left\{-\Phi^{-1}(t/2)\mu_1-\mu_1^2/2\right\}\right] \\ &\quad +\frac{p_2}{p_1+p_2}\left[\frac{1}{2}\exp\left\{\Phi^{-1}(t/2)\mu_2-\mu_2^2/2\right\}+\frac{1}{2}\exp\left\{-\Phi^{-1}(t/2)\mu_2-\mu_2^2/2\right\}\right] \end{aligned}$$

Specifically for the case of $\sigma_1=\sigma_2=1$, it follows that

$$\begin{aligned} (g_P^1)'(t) &= \frac{p_1\exp(-\mu_1^2/2)\mu_1}{4(p_1+p_2)\phi\{\Phi^{-1}(t/2)\}}\left[e^{\Phi^{-1}(t/2)\mu_1}-e^{-\Phi^{-1}(t/2)\mu_1}\right] \\ &\quad +\frac{p_1\exp(-\mu_2^2/2)\mu_2}{4(p_1+p_2)\phi\{\Phi^{-1}(t/2)\}}\left[e^{\Phi^{-1}(t/2)\mu_2}-e^{-\Phi^{-1}(t/2)\mu_2}\right]. \end{aligned}$$

Hence we have $(g_P^1)'(t)<0$, implying that the p-value cdf is always concave. Two sided tests still depend on σ: there is nothing special for this case.

Remark. In practice, the Z-test is seldom used since it involves the unknown variance which has to be estimated from the data. However the asymptotic distributions of a fair number of test statistics are asymptotically normal under both the null and alternative. The sign test is one of them.

5.3.3 Sign test

Suppose we are interested to see if a continuous random variable Y is symmetrically distributed around 0. We can use a sign test to do the analysis. Assume that we have n independent realizations of $Y, Y_1, Y_2, \cdots, Y_n$, and let

$$\text{sign}(Y_i)=\begin{cases} 1 & \text{if } Y_i>0 \\ -1 & \text{if } Y_i<0. \end{cases}$$

The test statistic we use is $T_n=n^{-1}\sum_{i=1}^n \text{sign}(Y_i)=n^{-1}\sum_{i=1}^n[2I_{\{Y_i>0\}}-1]$. The null hypothesis is $H_0: p=P(Y_i>0)=0.5$ and the alternative hypothesis is $H_1: p\neq 0.5$. Under

the null hypothesis, T_n has expected value $E(\text{sign}(Y_i)) = P(Y_i > 0) - P(Y_i < 0) = 0$ and variance $Var(\text{sign}(T_n)) = n^{-2}\sum_{i=1}^{n} 4\text{Var}(I_{\{Y_i>0\}}) = 4P(Y_i > 0)(1 - P(Y_i > 0))/n = 1/n$. So the standardized test statistic $\sqrt{n}T_n \to N(0,1)$ under the null hypothesis. Under the alternative hypothesis, $E(\text{sign}(Y_i)) = P(Y_i > 0) - P(Y_i < 0) = 2P(Y_i > 0) - 1$, $Var(\text{sign}(Y_i)) = 4P(Y_i > 0)(1 - P(Y_i > 0)) = -4P(Y_i > 0)^2 + 4P(Y_i > 0)$, and so

$$\frac{\sqrt{n}T_n - \sqrt{n}[2P(Y_i > 0) - 1]}{\sqrt{-4P(Y_i > 0)^2 + 4P(Y_i > 0)}} \sim N(0,1) \quad as \quad n \to \infty.$$

So under the alternative, for the unstandardized test statistic $\sqrt{n}T_n$, the asymptotic distribution is normal with a shift in location and shrinkage of the variance (it is smaller than 1). This can serve as an example of the normal mixture counter example for practical testing problems.

To fix the problem, we recommend to use the standardized version of the sign test by plugging in the consistent estimate of the parameter $p = P(Y_i > 0)$ by $\hat{p} = \frac{\sum_{i=1}^{n} I_{\{Y_i\}}}{n}$ in the variance to get the estimate.

Remark. We conjecture that the heteroscedasticity phenomenon exists for central and non-central χ^2 test as well, but this is a future research topic.

5.3.4 Numerical studies

We use a normal mixture model to do the simulation. The summary statistic T has a $N(0,1)$ distribution under the null and a $N(\mu, \sigma)$ distribution under the alternative. Suppose the proportion of alternatives is π_1. We have summary statistic $T_i, i = 1, \cdots, m$ for each test and the mixture model:

$$T_i \sim (1 - \pi_1)N(0,1) + \pi_1 N(\mu, \sigma).$$

In the simulation study, the number of tests $m = 1000$ and we have 1001 replications. We first study the effect of μ on the shape of FDR for $\sigma \geq 1$ and $\sigma < 1$.

In the first simulation, the proportion of alternatives is 0.2 and the alternative standard deviation is 0.3. We change the mean effect from small to large (0.5 to 3) and see the shape of the FDR. Figure 5.4 is the plot for the case $\sigma < 1$. When $\sigma \geq 1$, the FDR is a monotone decreasing function with respect to the critical value t. In other words, the MLRC holds in this case. See Figure 5.5.

Figure 5.4: FDR with respect to alternative mean for fixed alternative proportion and small variance

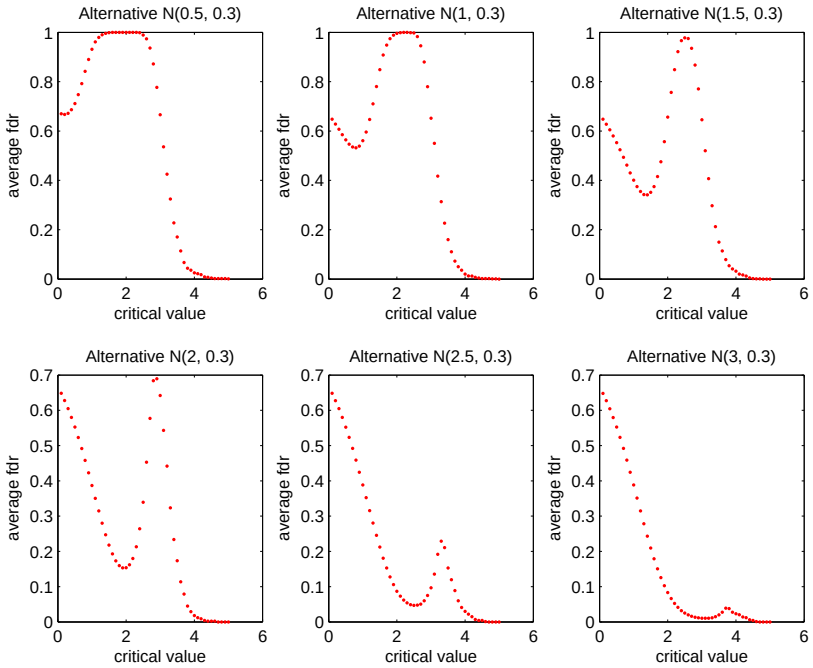


Figure 5.5: FDR with respect to alternative mean for fixed alternative proportion and large variance

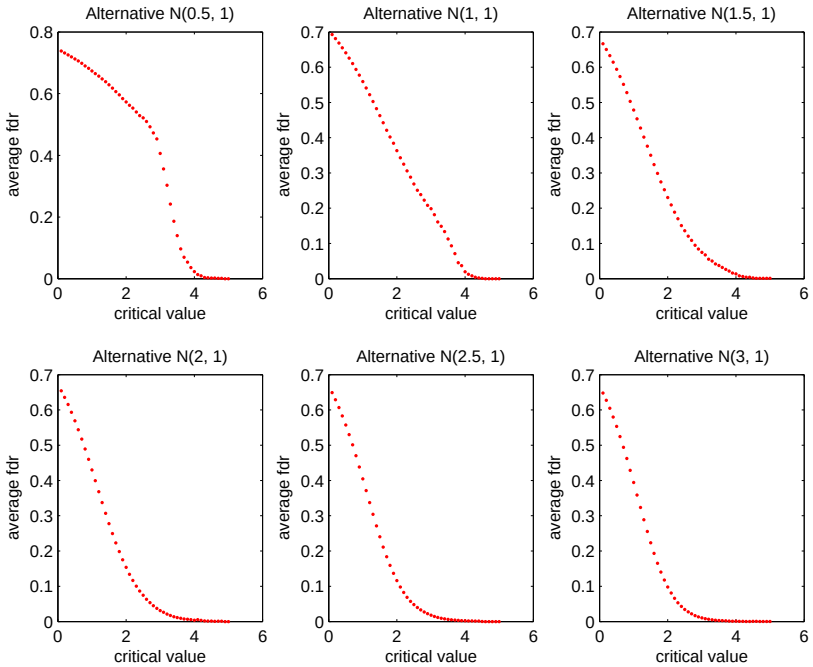
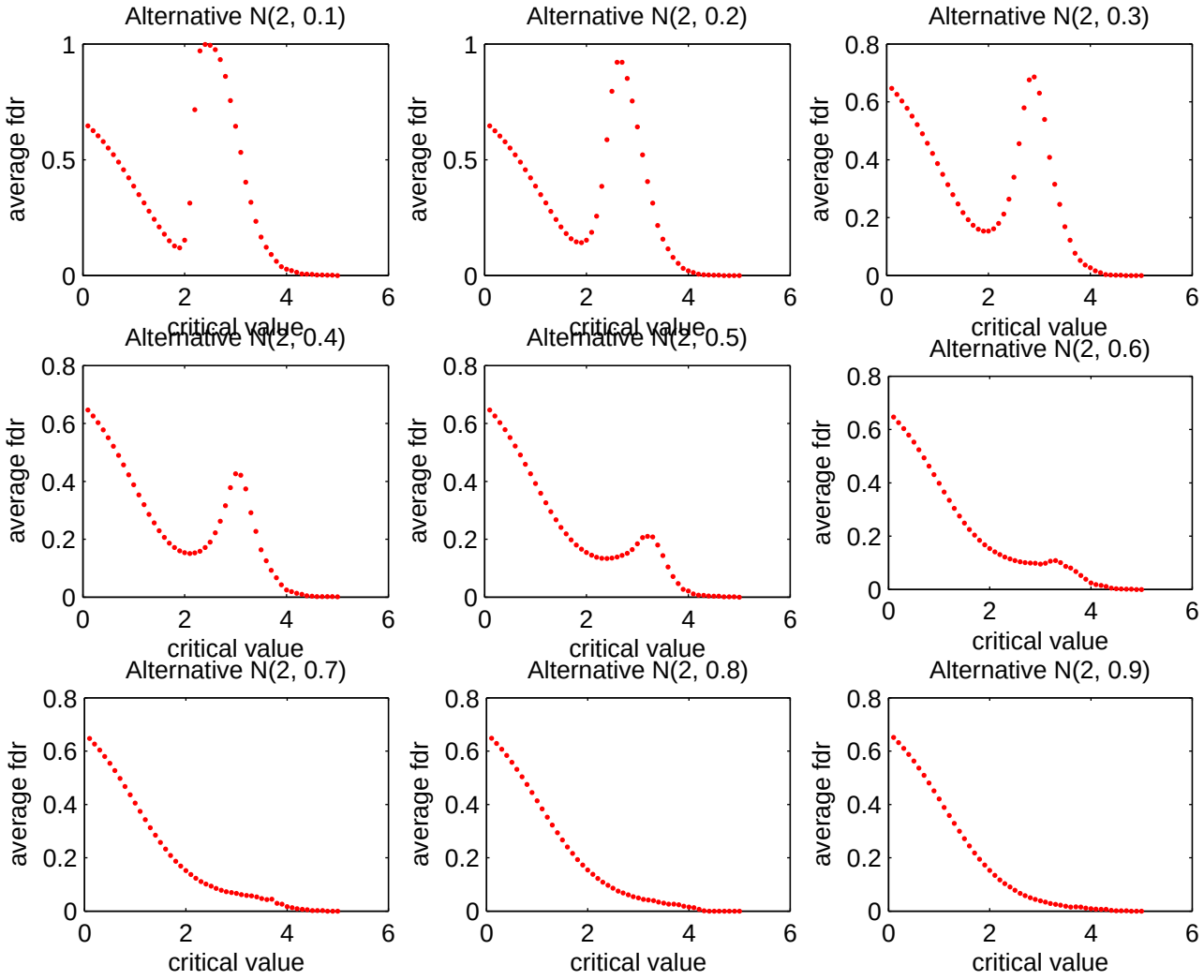


Figure 5.6: FDR with respect to small alternative variance for fixed alternative proportion and mean



In the second simulation, we fix the proportion of alternatives to be the same 0.2, the alternative mean $\mu = 2$ and change the alternative stand deviation from small to large (0.1 to 0.9). The FDR is not monotone with respect to the critical value as in Figure 5.6. When we change the alternative standard deviation $\sigma \geq 1$, the FDR is a monotone decreasing function with respect to the critical value t. In other words, the monotonicity of FDR holds in this case as can be seen from Figure 5.7.

In the third simulation, we fix the alternative mean $\mu = 2$, set a small alternative standard deviation $\sigma = 0.2$ and change the proportion of alternatives from small to large (0.01 to 0.25).

Figure 5.7: FDR with respect to large alternative variance for fixed alternative proportion and mean

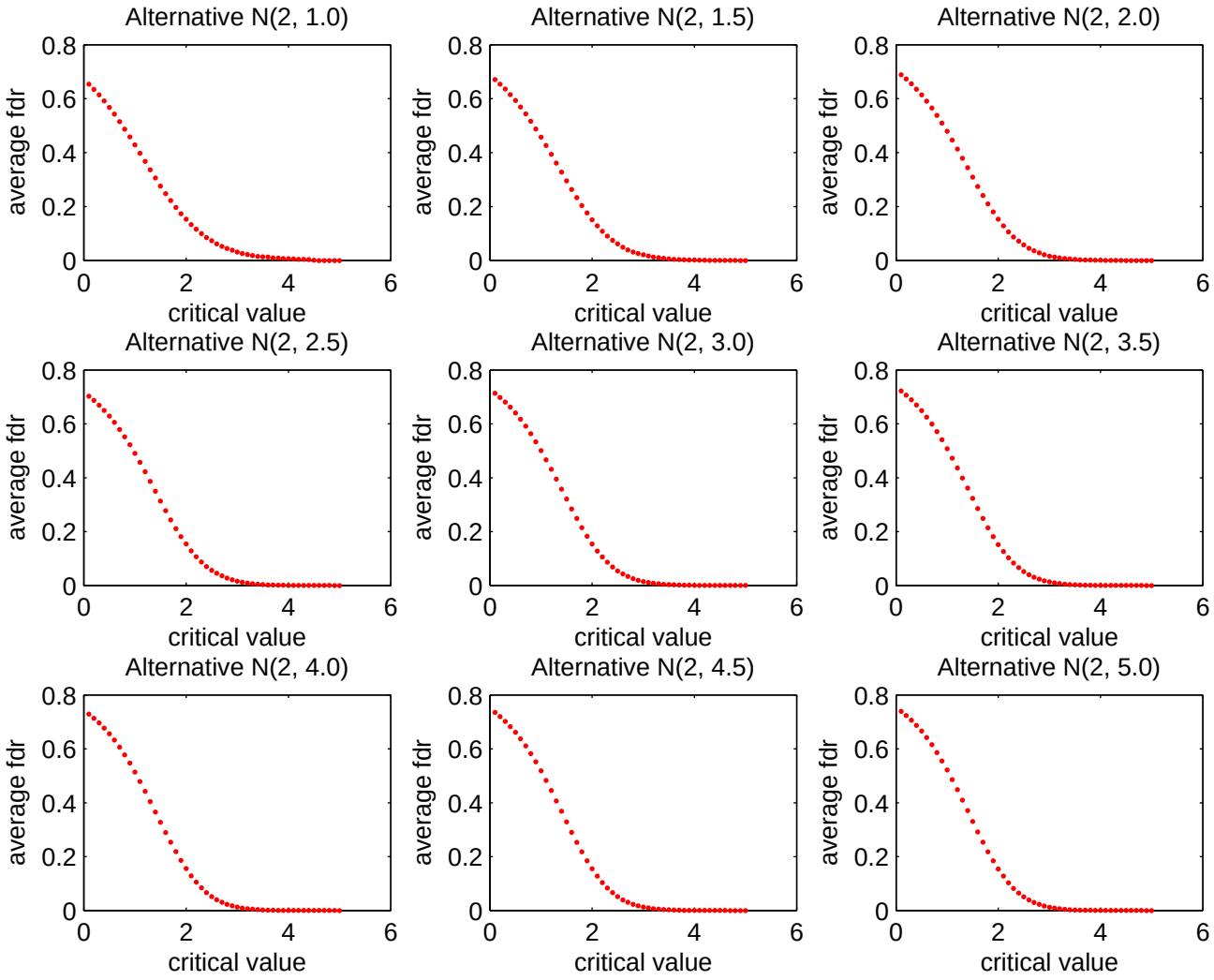
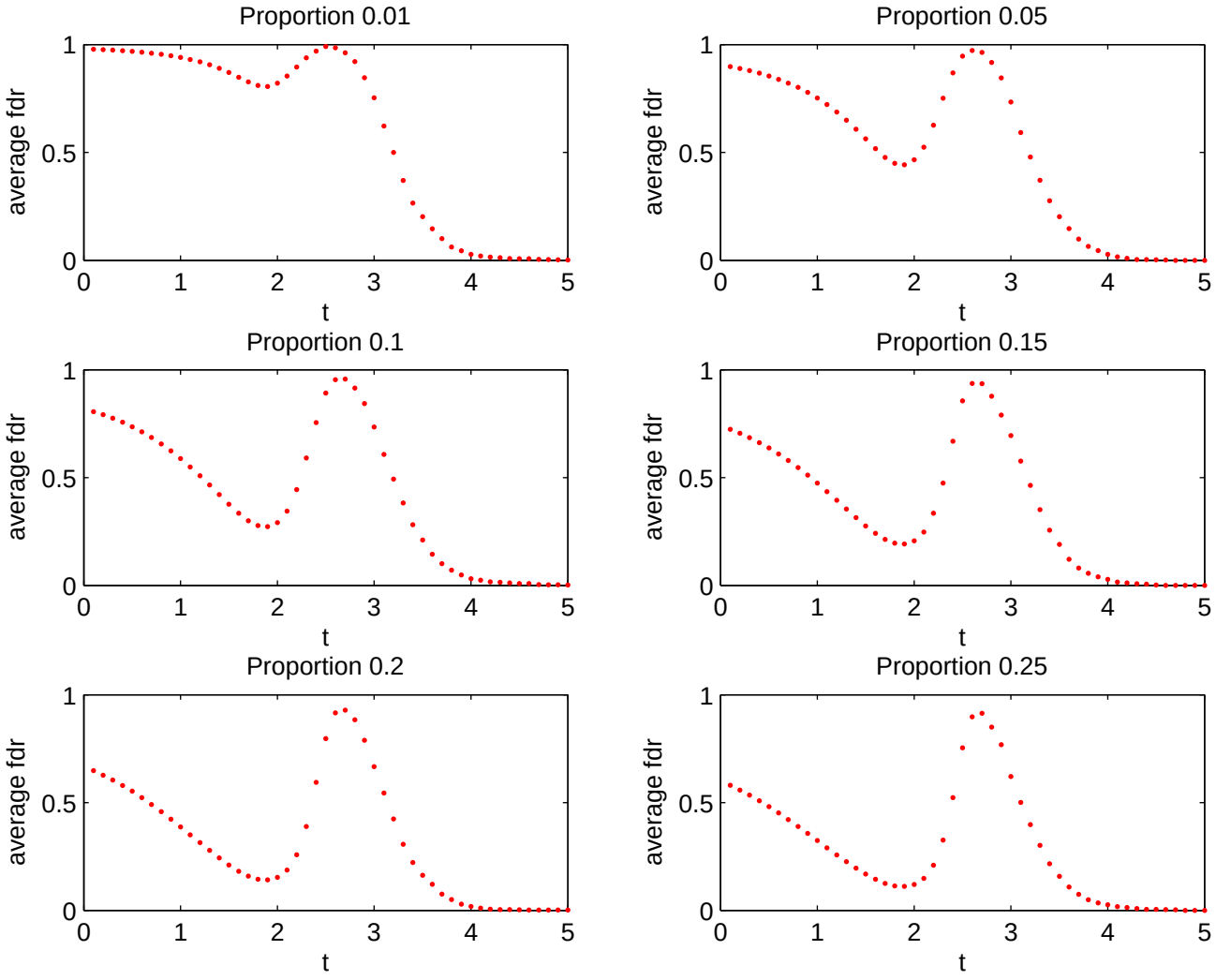


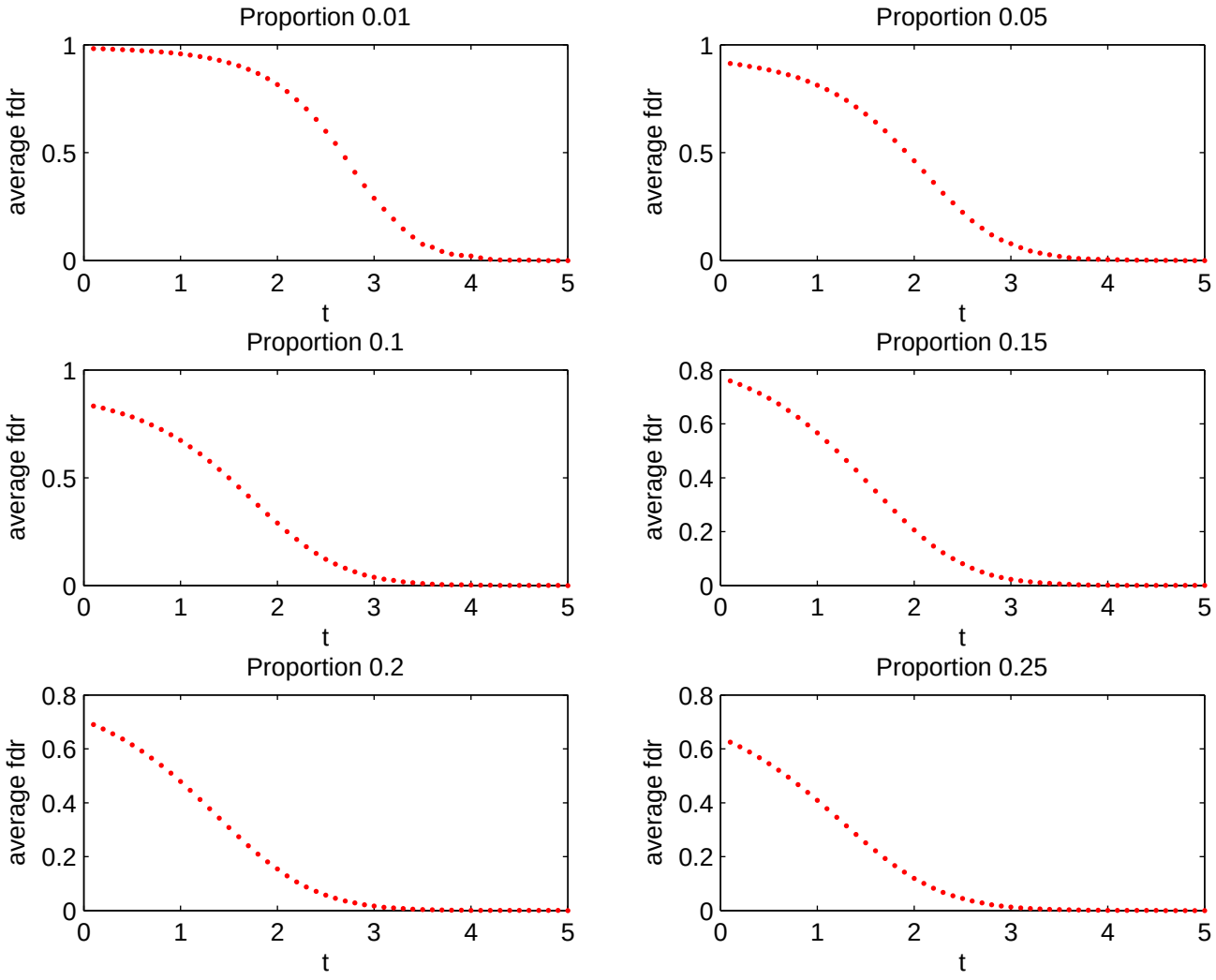
Figure 5.8: FDR with respect to alternative proportion for fixed alternative mean and small variance



From Figure 5.8, the FDR is not monotone for any proportion of alternatives. Next, we change the alternative standard deviation $\sigma \geq 1$, and the FDR is a monotone decreasing function with respect to the critical value t as illustrated in Figure 5.9.

In observing of the heteroscedasticity phenomenon of the variance under null and alternative hypotheses, in practical multiple testing implementation, we would like to use test statistics that do not have such issues. The t-statistics are good candidates.

Figure 5.9: FDR with respect to alternative proportion for fixed alternative mean and large variance



5.4 t-statistics

5.4.1 one-sided test

Lemma 5.4.1. *Suppose $X, X_i,\ i = 1, \cdots, n$ are independent identically distributed random variables. Let*

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}, \qquad s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2.$$

If X satisfies $E|X|^4 < \infty$, $E(X^2) = \sigma^2 > 0$ and $E(X) = 0$, then

$$P(\frac{\sqrt{n}(\bar{X} + c)}{s_n} \geq t) = \big(1 - \Phi(t - c\sqrt{n}/\sigma)\big)(1 + o(1)) \tag{5.6}$$

uniformly in $|c\sqrt{n}/\sigma| \leq t/5$ and $t = o(n^{1/6})$. We remark that following the same lines as their proof, we can see that (A.13) remains valid for $-t/5 \leq c\sqrt{n}/\sigma \leq t$.

If we use t-statistics, consider the one-sided test first:

$$\begin{aligned} P_i = P(T_i > t_i^{\text{observed}}) &= P(Z > T_i^{\text{observed}})(1 + o(1)) \\ &= (1 - \Phi(T_i))(1 + o(1)). \end{aligned}$$

The distribution of p-values under the null hypothesis is

$$\begin{aligned} G_p^0(t) &= P((1 - \Phi(T_i))(1 + o(1)) < t) \\ &= P(\Phi(T_i)(1 + o(1)) > 1 - t + o(1)) \\ &= P(T_i > \Phi^{-1}(1 - t + o(1))) \\ &= P(-T_i < \Phi^{-1}(t + o(1))) = t + o(1) \end{aligned}$$

Under the alternative hypothesis,

$$G_p^1(t) = P(T_i > \Phi^{-1}(1 - t + o(1)))$$

$$
\begin{aligned}
&= P(\frac{\sqrt{n}(\bar{X}-\mu+\mu)}{s_n} > \Phi^{-1}(1-t+o(1))) \\
&= \bar{\Phi}(-\Phi^{-1}(t+o(1)) - \frac{\sqrt{n}\mu}{\sigma}) \\
&= \Phi(\Phi^{-1}(t+o(1)) + \frac{\sqrt{n}\mu}{\sigma})
\end{aligned}
$$

Take derivative with respect to t, we get the pdf of p-values under the alternative hypothesis.

$$
\begin{aligned}
g_p^1(t) &= \frac{\phi(\Phi^{-1}(t+o(1)) + \frac{\sqrt{n}\mu}{\sigma})}{\phi(\Phi^{-1}(t+o(1)))} \\
&= exp(-\frac{1}{2}\frac{n\mu^2}{\sigma^2} - \frac{\sqrt{n}\mu\Phi^{-1}(t+o(1))}{\sigma}),
\end{aligned}
$$

which is a monotone decreasing function with respect to t.

5.4.2 two-sided test

Lemma 5.4.2. *Suppose $X, X_i,\ i = 1, \cdots, n$ are independent identically distributed random variables. Let*

$$
\bar{X} = \frac{\sum_{i=1}^n X_i}{n}, \qquad s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2.
$$

If X satisfies $E|X|^4 < \infty$, $E(X^2) = \sigma^2 > 0$ and $E(X) = 0$, then

$$
P(|\frac{\sqrt{n}(\bar{X}+c)}{s_n}| \geq t) = P(|Z + c\sqrt{n}/\sigma| \geq t)(1+o(1)) \tag{5.7}
$$

uniformly in c and $t = o(n^{1/6})$. Here and in the sequel, Z denotes a standard normal random variable. Note that

$$
\begin{aligned}
P_i = P(|T_i| > |t_i^{\text{observed}}|) &= P(|Z| > |T_i|^{observed})(1+o(1)) \\
&= 2\bar{\Phi}(|T_i|)(1+o(1)) = 2\Phi(-|T_i|)(1+o(1)).
\end{aligned}
$$

The distribution of p-values under the null hypothesis is

$$\begin{aligned} G_p^0(t) &= P(\Phi(-|T_i|)(1+o(1)) < t/2) \\ &= P(-|T_i| < \Phi^{-1}(t/2+o(1))) \\ &= P(|T_i| > -\Phi^{-1}(t/2+o(1))) \\ &= P(|Z| > -\Phi^{-1}(t/2+o(1))) \\ &= 2\Phi(\Phi^{-1}(t/2+o(1))) = t+o(1). \end{aligned}$$

Under alternative hypothesis,

$$\begin{aligned} G_p^1(t) &= P(|T_i| > -\Phi^{-1}(t/2+o(1))) \\ &= P(|\frac{\sqrt{n}(\bar{X}-\mu+\mu)}{s_n}| > -\Phi^{-1}(t/2+o(1))) \\ &= P(|Z+\frac{\sqrt{n}\mu}{\sigma}| > -\Phi^{-1}(t/2+o(1))) \\ &= P(Z > -\frac{\sqrt{n}\mu}{\sigma} - \Phi^{-1}(t/2+o(1))) + P(Z < \Phi^{-1}(t/2+o(1)) - \frac{\sqrt{n}\mu}{\sigma}) \\ &= \Phi(\Phi^{-1}(t/2+o(1)) + \sqrt{n}\mu/\sigma) + \Phi(\Phi^{-1}(t/2+o(1)) - \sqrt{n}\mu/\sigma). \end{aligned}$$

Taking derivative with respect to t, we get the pdf of p-values under the alternative hypothesis:

$$\begin{aligned} g_p^1(t) &= \frac{1}{2}[\frac{\phi(\Phi^{-1}(t/2+o(1)) + \sqrt{n}\mu/\sigma)}{\phi(\Phi^{-1}(t/2+o(1)))} + \frac{\phi(\Phi^{-1}(t/2+o(1)) - \sqrt{n}\mu/\sigma)}{\phi(\Phi^{-1}(t/2+o(1)))}] \\ &= \frac{1}{2}[exp(-\frac{1}{2}\frac{n\mu^2}{\sigma^2} - \frac{\sqrt{n}\mu}{\sigma}\Phi(t/2+o(1))) + exp(-\frac{1}{2}\frac{n\mu^2}{\sigma^2} + \frac{\sqrt{n}\mu}{\sigma}\Phi(t/2+o(1)))] \\ &= \frac{1}{2}exp^{-\frac{1}{2}\frac{n\mu^2}{\sigma^2}}[exp(\frac{\sqrt{n}\mu}{\sigma}\Phi(t/2+o(1))) + exp(-\frac{\sqrt{n}\mu}{\sigma}\Phi(t/2+o(1)))], \end{aligned}$$

which is monotone decreasing. So the p-value distributions under the alternative is a concave function.

Remarks

- *Step-up vs. Step-down Procedures.* If the FDR is not monotonically increasing in c,

the step-up and step-down procedure will for sure produce different results. One is too conservative and the other is two liberal in terms of FDR control. Further numerical studies are in progress for evaluating different procedures in face of this non-monotonicity issue. For the comparison literature of these two procedures, see Lehmann, Romano and Shaffer (2005).

- *On the choice of test statistics.* A test statistic that does not satisfy the MLRC seems to be inappropriate and counter-intuitive. In contrast, it can be shown using similar techniques in the proof of Corollary 1 in Sun and Cai (2009) that when the local fdr statistic $\text{Lfdr}(X_i)$ is used, that the MLRC (5.3) always holds. We've proved that the MLRC holds for t-tests as well. The scenarios considered in Fan, Hall and Yao (2007) still hold.

- *Implications on existing FDR procedures.* The concavity assumption has been extensively used in Storey (2003) and Genovese and Wasserman (2004). It is a convenient assumption to obtain desired results. The conclusions still hold as long as the tail distribution of the alternative is concave.

- *Efficiency of multiple testing.* As long as p-values are uniformly distributed under the null hypothesis, the BH procedure and some variants can be applied for testing with valid FDR control. But if MLRC does not hold, many procedures claimed to control FDR at a specified level and minimize FNR are not valid.

- *Dependent case.* It is of interest to investigate analytically whether the MLRC holds when tests are dependent. This is a future research topic. We plan to start with pairwise correlated tests or with a hidden Markov model assumption.

CHAPTER 6

Group testing under dependence

6.1 Background

With the easy access to massive datasets, it is increasingly important to extract useful features, which at most are only a small portion of the high throughput data from the vast amounts of data. In genomics microarrays, usually the available dataset are two matrices $X_{p\times n}$, and $Y_{n\times k}$, where p is the total number of genes, n is the number of subjects, and k is the number of covariates for each subject. In the simplest treatment and control situation, $k = 1$ with dichotomous categories. Usually, the number of genes p is of the order of thousands, while the number of arrays is at most hundreds. The interest lies in finding relevant genes that contribute to the different phenotypes, which can be casted into a multiple testing framework.

Many methods have been proposed in the literature. Usually an appropriate test statistic is calculated for each gene and used to assign a parametric or permutation-based p-value (Tusher et al., 2001; Dudoit et al., 2002b; Newton et al., 2004). Once a test statistic has been chosen, the primary statistical obstacle is accounting for multiple comparisons. Ranked lists of genes with small p-values are typically produced and subjected to an appropriate form of error rate control, such as the family-wise error rate (FWER) or the false discovery rate (FDR). However, this usual mode of analysis has been found to have several limitations. In particular, individual Gene analysis is often too conservative due to the need to control for a large number of multiple comparisons and correlation among genes, and results are subject to poor interpretability and reproducibility. An alternative approach is to incorporate prior biological information. Specifically, it is known that biological phenomena occur through the concerted expression of multiple genes. Thus, we can use our prior knowledge of what genes belong to various pathways to focus

our analysis on groups of functionally related genes called gene sets. The logic behind this type of analysis is that several functionally related genes demonstrating moderate differences between experimental conditions may be more important than a single, possibly spurious, highly significant gene. Instead of considering individual genes, the pathway approach treats the gene set as a single unit to be tested. This approach is becoming increasingly popular as it addresses various issues associated with individual gene analysis and provides more directly interpretable and reproducible results. Therefore, recent efforts have focused on the discovery of biological pathways rather than individual gene function, with the development of methods that are robust to the inaccuracies of specific gene estimates and which provide a more expansive view of the underlying processes.

A good approach for finding significant pathways depends on two components: (i) an accurate and powerful statistical method to discover significant patterns for a group of genes and (ii) a comprehensive and well-characterized pathway information mapped to microarray probes. The overall objective of the analysis is to test whether a group of genes has a coordinated association with a phenotype of interest. In terms of formal statistical language, there are two ways to formulate the null hypothesis. First, the competitive Null, H_0^{comp}: The genes in a gene set show the same pattern of associations with the phenotype compared with the rest of the genes. Second, the self-contained Null, H_0^{self}: The gene set does not contain any genes whose expression levels are associated with the phenotype of interest.

An essential difference between H_0^{comp} and H_0^{self} is that H_0^{comp} compares the association strength for genes in a gene set with the association strength for genes outside the gene set, whereas H_0^{self} only focuses on the associations of genes within the gene set. As a ranking criterion, H_0^{self} has its own limitation: When there is a significant proportion of genes associated with the phenotype of interest, large gene sets corresponding to irrelevant pathways could contain many genes associated with the phenotype by chance and be ranked highly according to H_0^{self}. To circumvent such problems, we propose a gene sets testing approach based on the proportion of significant genes in the gene pathway rather than the absolute number of significant genes. Therefore, we can produce more robust results irrespective of the size of gene sets.

Suppose we have G gene sets (which is available from the Gene Ontology or KEGG

database), with different sizes ranging from 30 to 200. For each gene set, we use $\pi_g, g = 1, \cdots, G$ to denote the proportion of differentially expressed genes in the pathway and $\pi_g^c, g = 1, \cdots, G$ to denote the proportion of differentially expressed genes in the complement pathway. We want to see if this pathway is special compared with the whole gene lists:

$$H_0 : \pi_g = \pi_g^c, \quad H_1 : \pi_g \neq \pi_g^c, \quad \text{where} \quad g \in \{1, \cdots, G\}.$$

Based on previous work, if we use the two sample t statistic as the testing statistic, under certain assumptions, the proportion can be written as

$$\pi_g = lim_{m_g \to \infty, n \to \infty} sup_{c>0} \frac{\hat{g}_c - E(g_c(Z))}{1 - E(g_c(Z))} \quad a.s.,$$

where m_g is the total number of genes in gene pathway G_g, n is the number of arrays, and $Z \sim N(0,1)$. Here $g_c(x) = min(|x|, c)/c$, $\hat{g}_c = \sum_{l=1}^{m_i} g_c(T_l)$ and $E(g_c(Z)) = \frac{2}{c\sqrt{2\pi}}(1 - e^{-c^2/2}) + 2\bar{\Phi}(c)$.

For the proportion estimate $\hat{\pi} = \frac{\hat{g}_c - Eg_c(Z)}{1 - Eg_c(Z)}$, under the assumption $\sqrt{n}\mu_i/\sigma_i \to \infty \quad a.s., \quad as \ n \to \infty$ for all i that represent the differentially expressed genes, we have $P(|Z + \frac{\sqrt{n}\mu_i}{\sigma_i}| \geq c) = 1$. So

$$\begin{aligned} E\hat{\pi} &= \frac{E\hat{g}_c - Eg_c(Z)}{1 - Eg_c(Z)} \\ &= \frac{\pi E g_c(T)|H = 1 + (1-\pi)Eg_c(T)|H = 0}{1 - Eg_c(Z)} \\ &\approx \frac{\pi g_c(Z + \sqrt{n}\mu_i/\sigma_i) + (1-\pi)Eg_c(Z)}{1 - Eg_c(Z)} \\ &\approx \frac{\pi(1 - Eg_c(Z))}{1 - Eg_c(Z)} = \pi. \end{aligned}$$

suppose the maximum is obtained at value c_0 in $\hat{\pi}$. Now we calculate the variance of this proportion estimate. We assume that for the alternative hypotheses, $H_i = 1, \sqrt{n}\mu_i/\sigma_i \to \infty \ a.s.$ We first try the pairwise correlation between test statistics $cov(T_i, T_j) = \rho$.

So

$$\begin{aligned} var(\hat{g}_{c_0}) &= \frac{1}{m}\{(1-\pi)Eg_{c_0}(Z)^2 - (1-\pi)(Eg_{c_0})^2 + \pi(1-\pi)(1 - Eg_{c_0}(Z))^2\} \\ &+ \frac{2(m-1)}{m}\{\pi^2 + 2\pi(1-\pi)Eg_{c_0}(Z) + (1-\pi)^2 Eg_{c_0}(X)g_{c_0}(Y) - [(1-\pi)Eg_{c_0}(Z) + \pi]^2\}, \end{aligned}$$

where X, Y are bivariate normal with variance σ_1^2, σ_2^2 and correlation ρ and $Eg_{c_0}(Z)^2 = 2\bar{\Phi}(c_0) + \frac{2}{c_0^2}\Phi(c_0) - \frac{1}{c_0^2} - \frac{2e^{-c_0^2/2}}{\sqrt{2\pi}c_0}$. Recall that $g_c(x) = min\{\frac{|x|}{c}, 1\}$, we have

$$Eg_{c_0}(X)g_{c_0}(Y) \quad = \quad 4\int_0^{c_0}\int_0^{c_0}\frac{xy}{c_0^2}f_\rho(x,y)dxdy + 4\int_{c_0}^{\infty}\int_{c_0}^{\infty}f_\rho(x,y)dxdy + 8\int_0^{c_0}\int_{c_0}^{\infty}\frac{x}{c_0}f_\rho(x,y)dxdy,$$

where $f_\rho(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho)^2}(x^2/\sigma_1^2 + y^2/\sigma_2^2)}$. Eventually, if the number of test m is big,

$$var(\hat{\pi}) = O(1/m) + \frac{2}{(1 - Eg_{c_0}(Z))^2}\{(1-\pi)^2 Eg_{c_0}(X)g_{c_0}(Y) + 2\pi(1-\pi)(Eg_{c_0}(Z) - Eg_{c_0}(Z)^2)^2\}.$$

Suppose within a gene set group $g(g = 1, \cdots, G)$, the model is as follows:

$$x_{ji} = \mu_i + \epsilon_{ji}, \quad i = 1, \cdots, m_g;\ j = 1, \cdots, n,$$

where i indicates different gene location within the group and j indicates different samples. μ_i is the fixed effect of expression level of gene i, and ϵ_{ji} is the random noise. Let ϵ_j, represent the genes corresponding to sample i and $\epsilon_j \sim N(0, \Sigma)$. For now, we assume that Σ is of the form

$$\begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \rho & \cdots & 1 & \rho \\ \rho & \rho & \cdots & 1 \end{pmatrix}.$$

In other words, for the sample subject, we assume the genes are dependent with a pairwise correlation ρ. Compared with the within-group correlation, the correlation between different gene sets is very weak, and we ignore it for now.

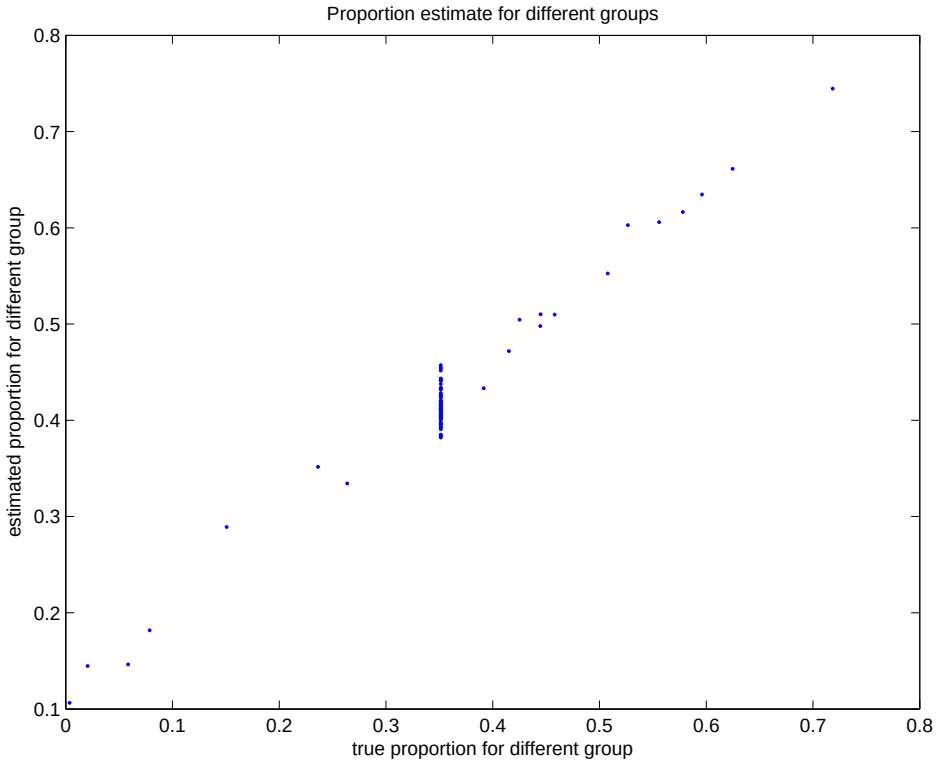
6.2 Simulation

In our numerical studies, we want to see the performance of this proportion estimate under dependence and compare this proportion with its complement for each gene set to test for the groups that are different from the rest.

The set-up is as follows. We have $G = 100$ different groups, with group size $mG \sim$ Unif$(30, 200)$. The group size is modest between $30 - 200$ for the asymptotics to work and

the consideration that larger groups always tend to be significant. The proportion of significant groups is 0.2, which is 20 in our case. Within the significant groups, the proportion of alternative tests is generated from a $\beta(1,3)$ distribution. For the remaining 80 groups, the proportions of alternative tests are generated by the weighted average of the 20 significant proportions—the same value for all 80 groups. So the average of the overall proportion is the same as the average of the null 80 groups. For the 20 significant group, the proportion of the group and its complement is different, while for the null 80 groups, the proportion is the same as their complements. Within each gene set, the signal is generated according to Unif$(0.5, 1)$ and the noise is generated according to a multivariate normal with mean 0 and pairwise correlation 0.7. Based on the two-sample t-statistic, the proportion π_g is calculated for each group as well as its complement π_g^c for each gene set. The point estimate of π_g is reasonable as can be seen in Figure 6.1. In order to get the confidence interval of $\pi_g - \pi_g^c$, we use the bootstrap method to permute the column from two different groups, calculate $\pi_g - \pi_g^c$ for each gene set based on 100 permutation, and evaluate whether 0 falls into the empirical 90% percentile of the permuted $\pi_g - \pi_g^c$. This is a per-comparison test, which did not adjust for multiple comparison. We combine the results across different gene sets and calculate the FDR. For the nominal significance level of 0.1, the overall FDR is controlled around 20% based on this empirical study. We tried to implement the BH procedure by calculating bootstrap p-values for each set, ranking and thresholding to adjust for the multiplicity. The performance is very poor due to the reason that in our case, under the null hypothesis, the p-value is not Unif$(0, 1)$ distributed. Further investigation in this approach will be pursued by possibly smoothing the p-value under the null to be Unif$(0, 1)$ or using non-parametric multiple comparison adjustment which doesn't require knowing the distribution of p-values.

Figure 6.1: Proportion estimate for group testing under dependence



CHAPTER 7

Concluding Remarks

7.1 Overview

In this dissertation, we have presented a groundbreaking new approach to large scale multiple testing which improves the power at the same FDR level but requires minimum assumptions compared with standard approaches, developed a new asymptotically consistent estimate for the proportion of alternatives, pointed out a new non-monotone phenomenon of FDR with respect to the threshold, its practical implications for existing FDR procedures, and a group testing method using relative measure of error with applications in genomics. The general idea is to incorporate alternatives into decision making rather than the p-value approach which only takes into account of the null with specific test statistics and in certain applications. Our work was motivated by high-throughput techniques in genomics, recent theoretical advancement in the field of moderate deviations, concentration inequalities and empirical processes.

We have presented a new approach for the significance analysis of thousands of features in high-dimensional biological studies. The approach is based on estimating the critical values of the rejection regions for high dimensional multiple hypothesis testing rather than the conventional p-value approaches in the literature. We developed a detailed method that can be used to identify differentially expressed genes in microarray experiments. The proposed procedure performs well for large samples, reasonably good for intermediate samples and not quite as good for small samples, and appears to perform better than existing alternatives under realistic sample sizes. Our method is also computationally faster than the competing approaches. The potential for improvement in small sample performance motivates the need for a second order expansion of our theoretical work. In addition, we have proposed a new consistent estimate of

the proportion of alternative hypotheses under certain conditions. Numerical studies demonstrate that our methodology fits the truth well and improves the statistical power in multiple testing.

The non-monotonicity of FDR with respect to the threshold is interesting since it opens a new door for the interpretation of high-throughput screening in terms of statistical analysis. The traditional ranking does not work well in this circumstance — the features corresponding to large test statistic values are important, followed by a gap which jumps over the intermediate value and the features corresponding to some relatively small test statistics are picked up as important, while the traditional FDR approach ranks and thresholds only features corresponding to the large test statistics.

7.2 Future Research

Extensions of the current work can be done in several directions.

First, as we said before, the precision of the asymptotic approximations has room for improvement in small to moderately small sample sizes, suggesting that a second order expansion would be valuable. Second, under the dependence case, it would be of interest to see how the rate of convergence could be derived under various assumptions on the form of the dependence. Thirdly, the plug-in estimator π_1 is consistent but somewhat ad-hoc. Complete, theoretical properties of this estimator remain to be explored. Fourth, we only considered a fixed proportion π_1 of alternative hypotheses. It is of great interest to consider also the sparsity setting, in which $\pi_1 \to 0$ as $m \to \infty$, and see what patterns emerge. When the number of hypothesis tests is of the order of millions, the number of signals doesn't change much, so it is not realistic to assume the proportion of alternatives is fixed when the number of tests increase. Higher Criticism is shown to be useful in rare proportions and weak signal modeling. See Donoho and Jin (2004) and Donoho and Jin (2006). But no proportion estimates have been derived in the higher criticism approach and the assumption of sharing the same signal strength for all alternatives is not realistic. We plan to add some prior on the alternative mean, model the alternative variances as random variables coming from an underlying smooth function, and then explore the multiplicity calibration when the proportion of alternatives goes to 0.

APPENDIX A

Preliminary lemmas

A.1 Berry-Esseen bound for non-central t-statistics

Our main tools are limit theorems of empirical processes, Berry-Esseen bounds, and self-normalized moderate deviations for one and two sample t-statistics.

We first state a non-uniform Berry-Esseen inequality for non-linear statistics:

Lemma A.1.1. *Chen and Shao (2007). Let $\xi_1, \xi_2, \ldots, \xi_n$ be independent random variables with $E\xi_i = 0$, $\sum_{i=1}^n E\xi_i^2 = 1$ and $E|\xi_i|^3 < \infty$. Let $W_n = \sum_{i=1}^n \xi_i$ and $\Delta = \Delta(\xi_1, \ldots, \xi_n)$ be a measurable function of $\{\xi_i\}$. Then*

$$\begin{aligned} |P(W_n + \Delta \le z) - \Phi(z)| &\le P(|\Delta| > (|z|+1)/3) \\ &+ C(|z|+1)^{-3}\big(||\Delta||_2 + \sum_{i=1}^n (E\xi_i^2)^{1/2}(E(\Delta - \Delta_i)^2)^{1/2} \quad \text{(A.1)} \\ &+ \sum_{i=1}^n E|\xi_i|^3\big) \quad \text{(A.2)} \end{aligned}$$

This is Theorem 2.2 in Chen and Shao (2007) and the proof can be found therein. The next lemma gives a Berry-Esseen bound for non-central t-statistics:

Lemma A.1.2. *Let $X, X_1, \cdots, X_n$ be i.i.d. random variables with $E(X) = 0$, $\sigma^2 = EX^2$ and $EX^4 < \infty$. Let*

$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i, \quad s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2.$$

Then

$$|P(\frac{\sqrt{n}(\bar{X}+c)}{s_n} \le x) - \Phi(x - \sqrt{n}c/\sigma)| \le K\frac{(1+|x|)}{(1+|x-\sqrt{n}\,c/\sigma|)\sqrt{n}} \quad \text{(A.3)}$$

for any c and x, where K is a finite constant that may depend on σ and EX^4.

Proof. Without loss of generality, assume $x \ge 0$ and $\sigma = 1$. Using

$$1 - |t| \le (1+t)^{1/2} \le 1 + |t| \quad \text{for} \quad t \ge -1, \quad \text{(A.4)}$$

we have

$$xs_n = x(1 + s_n^2 - 1)^{1/2} \le x(1 + |s_n^2 - 1|). \tag{A.5}$$

and

$$xs_n \ge x(1 - |s_n^2 - 1|). \tag{A.6}$$

Therefore

$$\begin{aligned} P(\frac{\sqrt{n}(\bar{X} + c)}{s_n} \le x) &= P(\sqrt{n}(\bar{X} + c) \le x\, s_n) \\ &\le P\big(\sqrt{n}\bar{X} \le x - \sqrt{n}\, c + x|s_n^2 - 1|\big). \end{aligned} \tag{A.7}$$

We now apply $(A.2)$ with $\xi_i = X_i/\sqrt{n}$, $W_n = \sqrt{n}\,\bar{X}$, and

$$z = x - \sqrt{n}\, c,\ \Delta = -x|s_n^2 - 1|,\quad \Delta_i = -x|s_{n,i}^2 - 1|,$$

where $s_{n,i}^2$ is defined as s_n^2 with 0 to replace X_i.

Noting that

$$s_n^2 - 1 = \frac{1}{n-1}(\sum_{j=1}^{n}(X_j^2 - 1) - n\bar{X}^2) + \frac{1}{n-1},$$

$$s_{n,i}^2 - 1 = \frac{1}{n-1}(\sum_{j\ne i}(X_j^2 - 1) - n(\bar{X} - X_i/n)^2),$$

we have

$$E|s_n^2 - 1|^2 \le K\, EX^4/n \tag{A.8}$$

and

$$\begin{aligned} E(s_n^2 - s_{n,i}^2)^2 &= \frac{1}{(n-1)^2}E\Big((X_i^2 - 1) - n\bar{X}^2 + n(\bar{X} - X_i/n)^2 + 1\Big)^2 \\ &= \frac{1}{(n-1)^2}E\Big((X_i^2 - 1) - X_i(2(\bar{X} - X_i/n) + X_i/n) + 1\Big)^2 \\ &\le \frac{2}{(n-1)^2}E\Big(2(X_i^2 - 1)^2 + 2 + X_i^2(2(\bar{X} - X_i/n) + X_i/n)^2\Big) \\ &\le \frac{2}{(n-2)^2}\Big(4EX^4 + 6 + EX_i^2(8(\bar{X} - X_i/n)^2 + 2EX_i^2/n)\Big) \\ &\le KEX^4/n^2. \end{aligned} \tag{A.9}$$

It follows from $(A.8)$ and $(A.9)$ that

$$\begin{aligned} ||\Delta||_2 &\leq K\frac{|x|\sqrt{EX^4}}{\sqrt{n}}, \\ P(|\Delta| > \frac{|z|+1}{3}) &\leq K\frac{|x|\sqrt{EX^4}}{\sqrt{n}(1+|z|)}, \\ \sum_{i=1}^{n}(E\xi_i^2)^{1/2}(E(\Delta-\Delta_i)^2)^{1/2} &\leq K\frac{|x|\sqrt{EX^4}}{\sqrt{n}}, \end{aligned}$$

and

$$\sum_{i=1}^{n} E|\xi_i|^3 \leq \frac{EX^3}{\sqrt{n}}.$$

Therefore, by $(A.2)$,

$$|P\big(\sqrt{n}\bar{X} \leq x - \sqrt{n}\,c + x|s_n^2 - 1|\big) - \Phi(x - \sqrt{n}\,c)| \leq \frac{K(1+|x|)}{(1+|x-\sqrt{n}c|)\sqrt{n}}. \tag{A.10}$$

Similarly,

$$P(\frac{\sqrt{n}(\bar{X}+c)}{s_n} \leq x) \geq P\big(\sqrt{n}\bar{X} \leq x - \sqrt{n}\,c - x|s_n^2 - 1|\big)$$

and

$$|P\big(\sqrt{n}\bar{X} \leq x - \sqrt{n}\,c - x|s_n^2 - 1|\big) - \Phi(x - \sqrt{n}\,c)| \leq \frac{K(1+|x|)}{(1+|x-\sqrt{n}c|)\sqrt{n}}. \tag{A.11}$$

This proves $(A.3)$. $\square$

A.2 Moderate deviation for non-central t-statistics

We also need a moderate deviation for the non-central t-statistics:

Lemma A.2.1. *Suppose $X, X_i,\ i = 1, \cdots, n$ are independent identically distributed random variables. Let*

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}, \qquad s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

If X satisfies $E|X|^4 < \infty$, $E(X^2) = \sigma^2 > 0$ and $E(X) = 0$, then

$$P(|\frac{\sqrt{n}(\bar{X}+c)}{s_n}| \geq t) = P(|Z + c\sqrt{n}/\sigma| \geq t)(1 + o(1)) \tag{A.12}$$

uniformly in c and $t = o(n^{1/6})$. Here and in the sequel, Z denotes a standard normal random variable.

Proof. When t is bounded, $(A.12)$ follows from Lemma A.1.2. Consider large t with $t = o(n^{1/6})$. We need the following result of Wang and Hall (2009) and Wang (2008):

$$P(\frac{\sqrt{n}(\bar{X}+c)}{s_n} \geq t) = \big(1 - \Phi(t - c\sqrt{n}/\sigma)\big)(1 + o(1)) \tag{A.13}$$

uniformly in $|c\sqrt{n}/\sigma| \leq t/5$ and $t = o(n^{1/6})$. We remark that following the same lines as their proof, we can see that $(A.13)$ remains valid for $-t/5 \leq c\sqrt{n}/\sigma \leq t$. Write

$$P(|\frac{\sqrt{n}(\bar{X}+c)}{s_n}| \geq t) = P(\frac{\sqrt{n}(\bar{X}+c)}{s_n} \geq t) + P(\frac{\sqrt{n}(-\bar{X}-c)}{s_n} \geq t).$$

By $(A.13)$, the remark above and the fact that

$$1 - \Phi(t + x) = o(1 - \Phi(t - x))$$

for $x \geq 1$ (recall here we assume t is large), $(A.12)$ holds for $-t \leq c\sqrt{n}/\sigma \leq t$. Now assume $|c|\sqrt{n}/\sigma > t$. Then by $(A.3)$

$$|P(|\frac{\sqrt{n}(\bar{X}+c)}{s_n}| \geq t) - P(|Z + c\sqrt{n}/\sigma| \geq t)| = o(1).$$

Since $|c|\sqrt{n}/\sigma > t$, we have $P(|Z + c\sqrt{n}/\sigma| \geq t) \geq 1/2$ and hence

$$P(|\frac{\sqrt{n}(\bar{X}+c)}{s_n}| \geq t) = P(|Z + c\sqrt{n}/\sigma| \geq t)(1 + o(1)).$$

This completes the proof of $(A.12)$. □

A.3 Results under i.i.d. assumption

The following i.i.d. results are essential for the general results.

Lemma A.3.1. *Assume the conditions of Theorem 2.1.1 with (2.2) replaced by the assumption that* $(T_i, H_i), i = 1, \cdots, m$ *are i.i.d. and* $\pi_1 = P(T_i = 1)$. *Let* $\mathcal{J} = \{i : H_i = 1\}$ *be the set that contains the indices of alternative hypotheses. Also assume that* μ_i, σ_i *are i.i.d. for* $i \in \mathcal{J}$. *Let*

$$p(t) = P(|T_1| \geq t), \tag{A.14}$$

$$a_1(t) = \alpha p(t) - (1 - \pi_1)F_0(t), \tag{A.15}$$

and

$$b_1^2(t) = \alpha^2 p(t)(1 - p(t)) + 2\alpha(1 - \pi_1)p(t)F_0(t) + (1 - \pi_1)F_0(t)(1 - 2\alpha - (1 - \pi_1)F_0(t)). \tag{A.16}$$

(i) If $t_{n,m}^{fdtp}$ *is chosen such that*

$$t_{n,m}^{fdtp} = \inf\{t : \sqrt{m}\, a_1(t)/b_1(t) \geq z_\gamma\}, \tag{A.17}$$

then

$$\lim_{m\to\infty} P(FDP \geq \alpha) = \lim_{m\to\infty} P(V \geq \alpha R) \leq \gamma \tag{A.18}$$

holds.

(ii) If $t_{n,m}^{fdr}$ *is chosen such that*

$$t_{n,m}^{fdr} = \inf\{t : \frac{(1 - \pi_1)F_0(t)}{p(t)} \leq \gamma\}, \tag{A.19}$$

then

$$\lim_{m\to\infty} FDR = \lim_{m\to\infty} E(V/R) \leq \gamma \tag{A.20}$$

holds.

(iii) If $t_{n,m}^{k-FWER}$ *is chosen such that*

$$t_{n,m}^{k-FWER} = \inf\{t : P(\eta(t) \geq k) \leq \gamma\}, \tag{A.21}$$

where $\eta(t) \sim Poisson(\theta(t))$ *and*

$$\theta(t) = m(1-\pi_1)F_0(t),$$

then

$$\lim_{m\to\infty} k\text{-}FWER = \lim_{m\to\infty} P(V \geq k) \leq \gamma \tag{A.22}$$

holds.

Proof. We first prove the i.i.d. case for one-sample t-statistic. By (2.3),

$$\begin{aligned}
\alpha R - V &= \alpha \sum_{i=1}^{m} I_{\{|T_i|\geq t\}} - \sum_{i=1}^{m}(1-H_i)I_{\{|T_i|\geq t\}} \\
&= \sum_{i=1}^{m}(H_i + \alpha - 1)I_{\{|T_i|\geq t\}} \\
&= \sum_{i=1}^{m} \alpha I_{\{|T_i|\geq t\}} I_{\{H_i=1\}} + \sum_{i=1}^{m}(\alpha-1)I_{\{|T_i|\geq t\}} I_{\{H_i=0\}} \\
&= \sum_{i=1}^{m} \alpha I_{\{|T_i|\geq t\}}(1 - I_{\{H_i=0\}}) + \sum_{i=1}^{m}(\alpha-1)I_{\{|T_i|\geq t\}} I_{\{H_i=0\}} \\
&= \sum_{i=1}^{m}(\alpha I_{\{|T_i|\geq t\}} - I_{\{|T_i|\geq t\}} I_{\{H_i=0\}}) \\
&= \sum_{i=1}^{m} \xi_i,
\end{aligned}$$

where

$$\xi_i := \xi_i(t) = \alpha I_{\{|T_i|\geq t\}} - I_{\{|T_i|\geq t\}} I_{\{H_i=0\}}.$$

is obviously a Donsker class indexed by t (Kosorok (2008)). Hence

$$P(V \geq \alpha R) = P(\sum_{i=1}^{m} \xi_i(t) \leq 0). \tag{A.23}$$

Note that since ξ_i are independent random variables, we can apply the uniform central limit theorem to choose t so that

$$P(\sum_{i=1}^{m} \xi_i(t) \leq 0) \leq \gamma. \tag{A.24}$$

To this end, we need to have the mean and variance of ξ_i. Without loss of generality, we use ξ_1 as an example, since ξ_i are i.i.d. random variables. Thus

$$\begin{aligned} E\xi_1 &= \alpha P(|T_1| \geq t) - P(|T_1| \geq t, H_1 = 0) \\ &= \alpha P(|T_1| \geq t) - P(H_1 = 0)P(|T_1| \geq t|H_1 = 0) \\ &= \alpha P(|T_1| \geq t) - (1 - \pi_1)P(|T_1| \geq t|H_1 = 0). \end{aligned} \tag{A.25}$$

Similarly,

$$\begin{aligned} E\xi_1^2 &= E(\alpha^2 I_{\{|T_1| \geq t\}} + (1 - 2\alpha) I_{\{|T_1| \geq t\}} I_{\{H_1 = 0\}}) \\ &= \alpha^2 P(|T_1| \geq t) + (1 - 2\alpha)(1 - \pi_1)P(|T_1| \geq t|H_1 = 0) \end{aligned} \tag{A.26}$$

and

$$\begin{aligned} \mathrm{Var}(\xi_1) &= E\xi_1^2 - (E\xi_1)^2 \\ &= \alpha^2 P(|T_1| \geq t) + (1 - 2\alpha)(1 - \pi_1)P(|T_1| \geq t|H_1 = 0) \\ &\quad -\{\alpha P(|T_1| \geq t) - (1 - \pi_1)P(|T_1| \geq t|H_1 = 0)\}^2 \\ &= \alpha^2 P(|T_1| \geq t)(1 - P(|T_1| \geq t)) \\ &\quad +(1 - \pi_1)P(|T_1| \geq t|H_1 = 0)(1 - 2\alpha - (1 - \pi_1)P(|T_1| \geq t|H_1 = 0)) \\ &\quad +2\alpha(1 - \pi_1)P(|T_1| \geq t)P(|T_1| \geq t|H_1 = 0). \end{aligned} \tag{A.27}$$

Now define

$$t_{n,m} = \inf\{t : \frac{\sqrt{m}\, E\xi_1(t)}{(\mathrm{Var}(\xi_1(t)))^{1/2}} \geq z_\gamma\}. \tag{A.28}$$

By Lemma A.4.1, $t_{n,m}$ is bounded and hence the uniform central limit theorem yields

$$\begin{aligned} & P(\sum_{i=1}^{m} \xi_i(t_{n,m}) \le 0) \\ = \;& P\Big(\frac{\sum_{i=1}^m (\xi_i(t_{n,m}) - E\xi_i(t_{n,m}))}{(\sum_{i=1}^m \mathrm{Var}(\xi_i(t_{n,m})))^{1/2}} \le -\frac{\sum_{i=1}^m E\xi_i(t_{n,m})}{(\sum_{i=1}^m \mathrm{Var}(\xi_i(t_{n,m})))^{1/2}}\Big) \\ \le \;& P\Big(\frac{\sum_{i=1}^m (\xi_i(t_{n,m}) - E\xi_i(t_{n,m}))}{(\sum_{i=1}^m \mathrm{Var}(\xi_i(t_{n,m})))^{1/2}} \le -z_\gamma\Big) \\ \to \;& \Phi(-z_\gamma) = \gamma. \end{aligned} \tag{A.29}$$

This proves $(A.18)$.

Note that

$$\begin{aligned} FDR &= \int_0^1 P(\mathrm{FDTP} \ge x)dx \\ &= \int_0^1 P(V \ge xR)dx \\ &= \int_0^1 P(\sum_1^m \xi_i \le 0)dx \\ &= \int_0^1 P(N(0,1) \le \frac{-\sqrt{m}E\xi_1}{\sqrt{Var\xi_1}}))dx. \end{aligned}$$

Let $m \to +\infty$, $P(N(0,1) \le -\sqrt{m}E\xi_1/\sqrt{Var\xi_1})$ is either 0 or 1 depending on the sign of $E\xi_1$. Thus the range of x that makes this probability 1 satisfies

$$E\xi_1 = xP(|T_1| \ge t) - (1-\pi_1)P(|T_1| \ge t|H_1 = 0) < 0$$

and the corresponding $x < (1-\pi_1)P(|T_1| \ge t|H_1 = 0)/P(|T_1| \ge t)$. In order to control FDR at level γ, we require

$$\frac{(1-\pi_1)P(|T_1| \ge t|H_1 = 0)}{P(|T_1| \ge t)} \le \gamma.$$

This proves $(A.19)$.

For the k-FWER, we use the characteristic function method. Letting $\eta_i = (1-H_i)I_{\{|T_i| \ge t\}}$,

we have

$$\begin{aligned} Ee^{is\sum_{i=1}^{m}\eta_i} &= \prod_{i=1}^{m} Ee^{is\eta_i} \\ &= \prod_{i=1}^{m}[e^{is}(1-\pi_1)F_0 + 1 - (1-\pi_1)F_0] \\ &= [1 + \frac{1}{m}m(1-\pi_1)F_0(e^{is}-1)]^m \\ &\to e^{\lambda(e^{is}-1)}, \end{aligned}$$

where $m_0F_0 \to \lambda$ as $m \to \infty$, and λ is the parameter for Poisson distribution, such that

$$P(\text{Poiss}(\lambda) \geq k) \leq \gamma. \qquad \square$$

The following functional central limit theorem is needed in the proof of theorem 2.1.1:

Lemma A.3.2. *Suppose the triangular array $\{f_{ni}(\omega,t), i = 1, \cdots, m_n, t \in T\}$ consists of independent processes within rows and is AMS. Let*

$$X_n(\omega,t) \equiv \sum_{i=1}^{m_n}[f_{ni}(\omega,t) - Ef_{ni}(.,t)]. \tag{A.30}$$

Assume:

(A) the $\{f_{ni}\}$ are manageable, with envelopes $\{F_{ni}\}$ which are also independent within rows;

(B) $H(s,t) = \lim_{n\to\infty} EX_n(s)X_n(t)$ exists for every $s,t \in T$;

(C) $limsup_{n\to\infty}\sum_{i=1}^{m_n} E^*F_{ni}^2 < \infty$;

(D) $lim_{n\to\infty}\sum_{i=1}^{m_n} E^*F_{ni}^2 1\{F_{ni} > \epsilon\} = 0$, for each $\epsilon > 0$;

(E) $\rho(s,t) = lim_{n\to\infty}\rho_n(s,t)$, *where*

$$\rho_n(s,t) \equiv (\sum_{i=1}^{m_n} E|f_{ni}(.,s) - f_{ni}(.,t)|^2)^{1/2},$$

exists for every $s,t \in T$, *and for all deterministic sequences* $\{s_n\}$ *and* $\{t_n\}$ *in* T, *if* $\rho(s_n,t_n) \to 0$ *then* $\rho_n(s_n,t_n) \to 0$.

Then X_n *converges weakly on* $l^\infty(T)$ *to a tight mean zero Gaussian process* X *concentrated on* $UC(T,\rho)$, *with covariance* $H(s,t)$.

The definitions involved in this lemma and the proof can be found in Theorem 11.16 of Kosorok (2008). Below, we verify that conditional on $\mathcal{H}$, $f_{ni}(\omega,t) = \xi_i(\omega,t)/\sqrt{m}$ satisfy the conditions in Lemma A.3.2. Since $\xi_i(\omega,t)$ is the difference between two monotone bounded functions, it is clear that conditional on $\mathcal{H}$, $\xi_i(\omega,t)/\sqrt{m}$ is AMS, manageable and has envelopes $\alpha/\sqrt{m}$. Also,

$$\begin{aligned}
EX_n(s)X_n(t) &= EE[X_n(s)X_n(t)|\mathcal{H}] \\
&= EE[\frac{\sum_{i=1}^m(\xi_i(s)|\mathcal{H} - E\xi_i(s)|\mathcal{H})}{\sqrt{m}}\frac{\sum_{j=1}^m(\xi_j(t)|\mathcal{H} - E\xi_j(t)|\mathcal{H})}{\sqrt{m}}] \\
&= EE\frac{\sum_{i=1}^m(\xi_i(s)|\mathcal{H} - E\xi_i(s)\mathcal{H})(\xi_i(t)|\mathcal{H} - E\xi_i(t)\mathcal{H})}{m} \\
&= \frac{1}{m}E\sum_{i=1}^m E(\xi_i(s)|\mathcal{H})(\xi_i(t)|\mathcal{H}) - \sum_{i=1}^m E(\xi_i(s)|\mathcal{H})E(\xi_i(t)|\mathcal{H}) \\
&= \frac{1}{m}E\sum_{i=1}^m(\alpha^2H_i + (1-\alpha)^2(1-H_i))EI_{\{|T_i|\geq t\cup s|\mathcal{H}\}} \\
&- \sum_{i=1}^m[\alpha H_i + (1-\alpha)(1-H_i)]^2EI_{\{|T_i|\geq s\mathcal{H}\}}EI_{\{|T_i|\geq t|\mathcal{H}\}} \\
&= \frac{1}{m}E\sum_{i=1}^m(\alpha^2H_iF_1(t\cup s) + (1-\alpha)^2(1-H_i)F_0(t\cup s)) \\
&- \sum_{i=1}^m[\alpha^2H_i + (1-\alpha)^2(1-H_i)][H_iF_1(s) + (1-H_i)F_0(s)][H_iF_1(t) + (1-H_i)F_0(t)] \\
&= \frac{1}{m}E\sum_{i=1}^m[\alpha^2H_i(F_1(t\cup s) - F_1(t)F_1(s)) + (1-\alpha)^2(1-H_i)(F_0(t\cup s) - F_0(t)F_0(s))] \\
&\to \pi_1\alpha^2(F_1(t\cup s) - F_1(t)F_1(s)) + (1-\pi_1)(1-\alpha)^2(F_0(t\cup s) - F_0(t)F_0(s)) \\
&\equiv H(s,t),
\end{aligned}$$

which is the same as $q^2(t)$ when $s = t$. (C) is easily satisfied. $\forall \epsilon > 0$, there exists a N_0 such that $\alpha/N_0 < \epsilon$ so $lim_{m\to\infty} \sum_{i=1}^{m} E\alpha^2/m 1\{\alpha/\sqrt{m} > \epsilon\} = lim_{m\to\infty} \sum_{i=1}^{N_0-1} \alpha^2/m = 0$, which verifies (D). Similarly we can show that (E) is satisfied and thus the functional central limit theorem holds. $\square$

Let

$$\begin{aligned} G(t) &= \alpha\pi_1 EP(|Z + \sqrt{n}\mu_1/\sigma_1| \geq t) - (1-\alpha)(1-\pi_1)P(|Z| \geq t) \\ &= \alpha\pi_1 EP(|Z + \sqrt{n}|\mu_1|/\sigma_1| \geq t) - (1-\alpha)(1-\pi_1)P(|Z| \geq t) \end{aligned}$$

and

$$t_1 = \inf\{t : G(t) = 0\}. \tag{A.31}$$

The following lemma is needed in the proof of consistency.

Lemma A.3.3. *Assume* $0 < \pi_1 < 1 - \alpha$ *and* $(A.40)$ *is satisfied. Then*

$$G(t) \begin{cases} < 0 & \text{for } t < t_1, \\ = 0 & \text{for } t = t_1, \\ > 0 & \text{for } t > t_1. \end{cases} \tag{A.32}$$

Moreover, $G'(t_1) \geq e^{-t_0^2/2}/\sqrt{2\pi}$.

Proof: We first observe that $0 < t_1 \leq t_0$ by the fact that $G(0) < 0$, $G(t_0) > e^{-t_0^2/2} > 0$ in $(A.48)$ and $G(t)$ is a continuous function.

To prove $(A.32)$, it suffices to show that there exists a $t_2 > t_1$ such that $G(t)$ is increasing in $[0, t_2]$ and decreasing in $[t_2, \infty)$. To this end, consider the derivative of G:

$$\begin{aligned} G'(t) &= -\alpha\pi_1 E\Big(\phi(t - \sqrt{n}|\mu_1|/\sigma_1) + \phi(t + \sqrt{n}|\mu_1|/\sigma_1)\Big) + 2(1-\alpha)(1-\pi_1)\phi(t) \\ &= \frac{e^{-t^2/2}}{\sqrt{2\pi}}\Big\{ - \alpha\pi_1 E\Big(\exp\big(- \frac{n\mu_1^2}{2\sigma_1^2} + \frac{\sqrt{n}|\mu_1|t}{\sigma_1}\big) + \exp\big(- \frac{n\mu_1^2}{2\sigma_1^2} - \frac{\sqrt{n}|\mu_1|t}{\sigma_1}\big)\Big) \\ &\quad +2(1-\alpha)(1-\pi_1)\Big\}. \end{aligned} \tag{A.33}$$

Let

$$H(t) = -\alpha\pi_1 E\Big(\exp\big(-\frac{n\mu_1^2}{2\sigma_1^2} + \frac{\sqrt{n}|\mu_1|t}{\sigma_1}\big) + \exp\big(-\frac{n\mu_1^2}{2\sigma_1^2} - \frac{\sqrt{n}|\mu_1|t}{\sigma_1}\big)\Big) + 2(1-\alpha)(1-\pi_1).$$

Then

$$\begin{aligned} H'(t) &= -\alpha\pi_1 E\Big\{ \frac{\sqrt{n}|\mu_1|}{\sigma_1} \exp\big(\frac{\sqrt{n}|\mu_1|t}{\sigma_1} - \frac{n\mu_1^2}{2\sigma_1^2}\big) \\ &\qquad -\frac{\sqrt{n}|\mu_1|}{\sigma_1} \exp\big(-\frac{\sqrt{n}|\mu_1|t}{\sigma_1} - \frac{n\mu_1^2}{2\sigma_1^2}\big)\Big\} \\ &= -\alpha\pi_1 E\Big\{ \frac{\sqrt{n}|\mu_1|}{\sigma_1} e^{-\frac{n\mu_1^2}{2\sigma_1^2}} \Big(\exp\big(\frac{\sqrt{n}|\mu_1|t}{\sigma_1}\big) - \exp\big(-\frac{\sqrt{n}|\mu_1|t}{\sigma_1}\big)\Big)\Big\} < 0 \qquad \text{(A.34)} \end{aligned}$$

for all $t > 0$. Therefore, $H(t)$ is monotone decreasing. Taking into account the fact that $H(0) > 0$ by assumption, and $\pi_1 < 1 - \alpha$ and $H(+\infty) < 0$, we conclude that $H(t)$ has only one zero point, say, t_2. Moreover, $H(t) > 0$ for $t < t_2$ and $H(t) < 0$ for $t > t_2$. This is also true for $G'(t)$ by $(A.33)$. Hence, $G(t)$ is increasing for $t < t_2$ and decreasing for $t > t_2$. Notice that since $G(0) < 0, G(t_0) > 0$ and $G(+\infty) = 0$, we can see that $G(t)$ has a unique zero point t_1 and $t_2 > t_1$. Since $G(t)$ is increasing for $0 < t < t_2$, we have $G'(t_1) > 0$. We now prove that $G'(t_1) \geq e^{-t_0^2/2}/\sqrt{2\pi}$. It follows from the proof of $(A.48)$ that

$$G(t_0) \geq e^{-t_0^2/2}. \qquad \text{(A.35)}$$

Recalling that $G'(t) = \frac{e^{-t^2/2}}{\sqrt{2\pi}} H(t)$ and H is decreasing, we have

$$\begin{aligned} G(t_0) &= G(t_0) - G(t_1) = \int_{t_1}^{t_0} G'(s)ds \\ &\leq \int_{t_1}^{t_0} \frac{e^{-s^2/2}}{\sqrt{2\pi}} H(t_1)ds \\ &\leq H(t_1)(1 - \Phi(t_1)) \leq H(t_1)e^{-t_1^2/2} = G'(t_1)\sqrt{2\pi}. \qquad \text{(A.36)} \end{aligned}$$

This proves $G'(t_1) \geq e^{-t_0^2/2}/\sqrt{2\pi}$. $\square$

Now, let's go back to show our main theorem under dependence. Let $\mathcal{H} = \{H_i,\ 1 \leq i \leq m\}$. To prove (i), following along the same lines as the proof of lemma A.3.1, we need to obtain the

asymptotic distribution of

$$P(V \geq \alpha R) = P(\sum_{i=1}^{m} \xi_i(t) \leq 0), \tag{A.37}$$

where

$$\xi_i(t) = \alpha I_{\{|T_i| \geq t\}} - I_{\{|T_i| \geq t\}} I_{\{H_i = 0\}} = (\alpha + H_i - 1) I_{\{|T_i| \geq t\}} = [\alpha H_i - (1-\alpha)(1-H_i)] I_{\{|T_i| \geq t\}}.$$

Note that

$$P(|T_i| \geq t | \mathcal{H}) = (1 - H_i) P(|T_i| \geq t | H_i = 0) + H_i P(|T_i| \geq t | H_i = 1).$$

Given $\mathcal{H}$, $\xi_i(t), 1 \leq i \leq m$ are independent random variables. The conditional mean equals

$$\begin{aligned}
& E(\sum_{i=1}^{m} \xi_i | \mathcal{H}) \\
= & \sum_{i=1}^{m} \Big\{ \alpha E(I_{\{H_i=0\}} | \mathcal{H}) P(|T_i| \geq t | H_i = 0) + \alpha E(I_{\{H_i=1\}} | \mathcal{H}) P(|T_i| \geq t | H_i = 1) \\
& - E(I_{\{H_i=0\}} | \mathcal{H}) P(|T_i| \geq t | H_i = 0) \Big\} \\
= & \sum_{i=1}^{m} \Big\{ \alpha (1 - H_i) P(|T_i| \geq t | H_i = 0) + \alpha H_i P(|T_i| \geq t | H_i = 1) \\
& - (1 - H_i) P(|T_i| \geq t | H_i = 0) \Big\} \\
= & \alpha \sum_{i=1}^{m} \Big\{ H_i P(|T_i| \geq t | H_i = 1) \Big\} - (1 - \alpha) \sum_{i=1}^{m} \Big\{ (1 - H_i) P(|T_i| \geq t | H_1 = 0) \Big\} \\
= & \alpha m_1 F_1(t) - (1 - \alpha) m_0 F_0(t).
\end{aligned}$$

Next we calculate the conditional variance of $\sum_{i=1}^{m} \xi_i(t)$, given $\mathcal{H}$:

$$\begin{aligned}
& var(\sum_{i=1}^{m} \xi_i(t) | \mathcal{H}) \\
& \qquad = var(\sum_{i=1}^{m} [\alpha H_i - (1-\alpha)(1-H_i)] I_{\{|T_i| \geq t | \mathcal{H}\}}) \\
& \qquad = \sum_{i=1}^{m} (\alpha^2 H_i + (1-\alpha)^2 (1 - H_i)) var(I_{\{|T_i| \geq t | \mathcal{H}\}}) \\
& \qquad = \alpha^2 m_1 F_1(t)(1 - F_1(t)) + (1-\alpha)^2 m_0 F_0(t)(1 - F_0(t)).
\end{aligned}$$

From (2.7) and (2.8),

$$\frac{\mu_m(t)}{\sigma_m(t)} = \sqrt{m}\frac{\mu_m(t)/m}{\sqrt{\sigma_m^2(t)/m}}.$$

By the fact that $m_1/m \to \pi_1 \quad a.s.$, we have

$$\mu_m(t)/m \to \alpha\pi_1 F_1(t) - (1-\alpha)(1-\pi_1)F_0(t) \quad a.s. \tag{A.38}$$

and

$$\sigma_m^2(t)/m \to \alpha^2\pi_1 F_1(t)(1-F_1(t)) + (1-\alpha)^2(1-\pi_1)F_0(t)(1-F_0(t)) = q^2(t) \;\; a.s., \tag{A.39}$$

which is smaller than $var(\xi_1(t))$ due to the fact that

$$varX = E(var(X|Y)) + var(E(X|Y))$$

for any two random variables X and Y. By (A.43), we can see that the critical value defined at (2.9) is bounded. Thus conditional on $\mathcal{H}$, we can use the functional central limit theorem on $\sum_{i=1}^m \xi_i(t)/\sqrt{m}$ by virtue of lemma A.3.2. The limit is a Gaussian process with continuous sample paths. Hence

$$\begin{aligned}
P(\sum_{i=1}^m \xi_i(t) \le 0) &= E(E1_{\{\sum_{i=1}^m \xi_i(t)/\sqrt{m} \le 0\}}|\mathcal{H}) \\
&= E\Big\{P\Big(\sum_{i=1}^m \xi_i/\sqrt{m} - \sum_{i=1}^m E(\xi_i|\mathcal{H})/\sqrt{m} \le \frac{-\sum_{i=1}^m E(\xi_i|\mathcal{H})\sigma_m(t)}{\sqrt{m}\sigma_m(t)}|\mathcal{H}\Big)\Big\} \\
&\le E\Big\{P\Big(\sum_{i=1}^m \xi_i/\sqrt{m} - \sum_{i=1}^m E(\xi_i|\mathcal{H})/\sqrt{m} \le \frac{-\sum_{i=1}^m E(\xi_i|\mathcal{H})}{\sigma_m(t)}\frac{\sigma_m(t)}{\sqrt{m}}|\mathcal{H}\Big)\Big\} \\
&\le E\Big\{P\Big(N(0,1)q(t) \le -z_\gamma q(t)\Big)\Big\} \\
&\to P(N(0,1) \le -z_\gamma) = \gamma \; as \;\; m \to \infty.
\end{aligned}$$

This proves (2.9).

(ii) can be proved similarly. The characteristic function method can be used to prove (iii).

□

A.4 Boundedness of critical values

The lemma below shows that $t_{n,m}$ defined in $(A.17)$ under independence is bounded:

Lemma A.4.1. *Assume that there exist $\varepsilon_0 > 0$ and $c_0 > 0$ such that*

$$P(|\sqrt{n}\mu_1/\sigma_1| \geq \varepsilon_0) \geq c_0. \tag{A.40}$$

Let $t_{n,m}$ satisfy $(A.28)$. Then

$$t_{n,m} \leq t_0, \tag{A.41}$$

where t_0 is the solution to

$$\alpha\pi_1 c_0 \exp((t_0 - \varepsilon_0)\varepsilon_0) = 12(1 + t_0 - \varepsilon_0). \tag{A.42}$$

Proof. It suffices to show that

$$\sqrt{m}\, E\xi_1(t_0) \geq (\mathrm{Var}(\xi_1(t_0)))^{1/2} z_\gamma. \tag{A.43}$$

It is easy to see that $P(|Z + a| \geq t_0)$ is a monotone increasing function of $a > 0$. Hence

$$\begin{aligned}
&P(|Z + \sqrt{n}\mu_1/\sigma_1| \geq t_0) \\
&\quad \geq \quad P(|Z + \sqrt{n}\mu_1/\sigma_1| \geq t_0, |\sqrt{n}\mu_1/\sigma_1| \geq \varepsilon_0) \\
&\quad \geq \quad P(|Z + \varepsilon_0| \geq t_0) P(|\sqrt{n}\mu_1/\sigma_1| \geq \varepsilon_0) \\
&\quad \geq \quad c_0 P(|Z + \varepsilon_0| \geq t_0) \geq c_0(1 - \Phi(t_0 - \varepsilon_0)) \\
&\quad \geq \quad \frac{c_0}{3(1 + t_0 - \varepsilon_0)} \exp(-(t_0 - \varepsilon_0)^2/2) \\
&\quad \geq \quad \frac{c_0}{3(1 + t_0 - \varepsilon_0)} \exp(-t_0^2/2 + (t_0 - \varepsilon_0)\varepsilon_0),
\end{aligned} \tag{A.44}$$

Here we use the fact that

$$\frac{1}{2} e^{-x^2/2} \geq 1 - \Phi(x) \geq \frac{1}{\sqrt{2\pi}(1 + x)} e^{-x^2/2} \quad \text{for} \quad x \geq 0.$$

Under the null hypothesis $H_1 = 0$, which corresponds to $\mu_i = 0$, we apply Lemma A.2.1 and obtain

$$P(|T_1| \geq t | H_1 = 0) = P(|Z| \geq t)(1 + o(1)). \tag{A.45}$$

uniformly in $t = o(n^{1/6})$.

Under the alternative hypothesis $H_1 = 1$, we apply Lemma A.2.1 to $X_{ij} - \mu_i$ and obtain

$$\begin{aligned} P(|T_1| \geq t | H_1 = 1) &= P(|\sqrt{n}(\bar{X}_1 - \mu_1 + \mu_1)/s_1| \geq t | H_1 = 1) \\ &= E[P(|Z + \sqrt{n}\mu_1/\sigma_1)| \geq t \mid \mu_1, \sigma_1)](1 + o(1)) \\ &= P(|Z + \sqrt{n}\mu_1/\sigma_1)| \geq t)(1 + o(1)) \end{aligned} \tag{A.46}$$

uniformly in $t = o(n^{1/6})$.

Also note that

$$\begin{aligned} P(|T_1| \geq t) &= P(|T_1| \geq t, H_1 = 0) + P(|T_1| \geq t, H_1 = 1) \\ &= (1 - \pi_1)P(|T_1| \geq t | H_1 = 0) + \pi_1 P(|T_1| \geq t | H_1 = 1) \\ &= (1 - \pi_1)P(|Z| \geq t)(1 + o(1)) + \pi_1 P(|Z + \sqrt{n}\mu_1/\sigma_1| \geq t)(1 + o(1)). \end{aligned} \tag{A.47}$$

By $(A.25)$, $(A.45)$, $(A.47)$ and $(A.44)$,

$$\begin{aligned} E\xi_1(t_0) &= \alpha(1 - \pi_1)P(|Z| \geq t_0)(1 + o(1)) + \alpha\pi_1 P(|Z + \sqrt{n}\mu_1/\sigma_1| \geq t_0)(1 + o(1)) \\ &\quad -(1 - \pi_1)P(|Z| \geq t_0)(1 + o(1)) \\ &\geq \alpha\pi_1 \frac{c_0}{6(1 + t_0 - \varepsilon_0)} \exp(-t_0^2/2 + (t_0 - \varepsilon_0)\varepsilon_0) - 2P(Z \geq t_0) \\ &\geq \frac{\alpha\pi_1 c_0}{6(1 + t_0 - \varepsilon_0)} \exp(-t_0^2/2 + (t_0 - \varepsilon_0)\varepsilon_0) - e^{-t_0^2/2} \\ &= e^{-t_0^2/2}\Big(\frac{\alpha\pi_1 c_0}{6(1 + t_0 - \varepsilon_0)} \exp((t_0 - \varepsilon_0)\varepsilon_0) - 1\Big) \\ &= e^{-t_0^2/2} \end{aligned} \tag{A.48}$$

by $(A.42)$ and the definition of t_0. It is easy to see that $E\xi_1^2 \leq 1$ and $\mathrm{Var}(\xi_1(t_0)) \leq 1$ in

particular. Thus, by $(A.48)$,

$$\frac{\sqrt{m}\, E\xi_1(t_0)}{(\mathrm{Var}(\xi_1(t)))^{1/2}} \geq \sqrt{m} e^{-t_0^2/2} \geq z_\gamma, \tag{A.49}$$

provided that m is large enough. This proves $(A.43)$. $\square$

APPENDIX B

Proof of Theorems in chapter 2

B.1 Proof of Theorem 2.1.2

We first prove (i), and (ii) follows along the same lines as the independent case plus a conditional argument. Without loss of generality, we use T_1 as a representative that comes from the alternative. We have to show that

$$|\hat{t}_{n,m} - t_{n,m}| = o(1) \quad a.s.. \tag{B.1}$$

We first prove

$$|\hat{t}_{n,m} - t_1| = o(1) \quad a.s., \tag{B.2}$$

where t_1 is defined as in (A.31). It suffices to show that for any $\varepsilon > 0$,

$$\frac{\sqrt{m}\nu_m(t_1+\varepsilon)}{\tau_m(t_1+\varepsilon)} \geq z_\gamma \tag{B.3}$$

and

$$\frac{\sqrt{m}\nu_m(s)}{\tau_m(s)} < z_\gamma \quad \text{for all} \quad s \leq t_1 - \varepsilon. \tag{B.4}$$

Recall $\hat{p}_m(t) = \frac{1}{m}\sum_{i=1}^m I_{\{|T_i|\geq t\}}$. Given $\mathcal{H}$, by the uniform law of the iterated logarithm (see e.g., Dudley and Philipp (1983)),

$$\hat{p}_m(t) - \frac{1}{m}\sum_{i=1}^m \Big\{(1-H_i)F_0(t) + H_i F_1(t)\Big\} = o(m^{-1/2}(loglogm)^{1/2}) \quad a.s.,$$

combined with

$$\frac{1}{m}\sum_{i=1}^m \Big\{(1-H_i)F_0(t) + H_i F_1(t)\Big\} \to (1-\pi_1)F_0(t) + \pi_1 F_1(t) \quad a.s., \tag{B.5}$$

by $(A.3)$, our strong consistent estimate $\hat{\pi}_1$ described in Section 2.3 and the continuous mapping

theorem, we have

$$\sup_t |\nu_m(t) - \{\alpha((1-\pi_1)F_0(t) + \pi_1 F_1(t)) - (1-\pi_1)P(|Z| \geq t)\}| \to 0 \quad a.s., \tag{B.6}$$

which together with $(A.47)$ and the definition of G implies

$$\sup_{0 \leq t \leq 1+t_0} |\nu_m(t) - G(t)| \to 0 \quad a.s.. \tag{B.7}$$

In particular, since $G(t_1 + \varepsilon) > 0$ for $0 < \varepsilon < t_2 - t_1$, we have

$$\nu_m(t_1 + \varepsilon) \geq G(t_1 + \varepsilon)/2 \quad a.s., \tag{B.8}$$

for sufficiently large m, and therefore $\sqrt{m}\nu_m(t_1 + \epsilon) \geq z_\gamma \tau_m(t_1 + \epsilon)$. This proves $(B.3)$.

Similarly, since $G(t)$ is increasing and $G(t_1 - \varepsilon) < 0$, we have

$$\max_{s \leq t_1 - \varepsilon} \nu_m(s) \leq G(t_1 - \varepsilon)/2 \quad a.s., \tag{B.9}$$

for sufficiently large m. Hence, $(B.4)$ holds. This proves $(B.2)$.

Following the same lines as the proof of $(B.2)$, we have

$$|t_{n,m} - t_1| = o(1). \tag{B.10}$$

This completes the proof of $(B.1)$.

For k-FWER, let η_0 be the number that satisfies $P(\text{Poiss}(\eta_0) \geq k) \leq \gamma$. Let $t_{0,m} = t_{n,m}^{\text{k-FWER}}$ and $t_m = \hat{t}_{m,n}^{\text{k-FWER}}$. Thus, by definition, $t_{0,m}$ is the t that satisfies $(1-\pi_1)mF_o(t) = \eta_0$ and t_m is the t that satisfies $2(1-\hat{\pi}_1)m\bar{\Phi}(t) = \eta_0$. Then we have $\frac{(1-\pi_1)F_0(t_{0,m})}{(1-\hat{\pi}_1)2\bar{\Phi}(t_m)} = 1$ which implies

$$\begin{aligned}
\frac{F_0(t_{0,m})}{2\bar{\Phi}(t_m)} = \frac{1-\hat{\pi}_1}{1-\pi_1} &= 1 + o_P(1) \Rightarrow \\
\frac{\bar{\Phi}(t_{0,m})}{\bar{\Phi}(t_m)}(1 + O(n^{-1/2})) &= 1 + o_P(1) \Rightarrow \\
\frac{\bar{\Phi}(t_{0,m})}{\bar{\Phi}(t_m)} &= 1 + o_P(1) \Rightarrow
\end{aligned}$$

$$\frac{t_m}{t_{0,m}} e^{-t_{0,m}^2/2+t_m^2/2} = 1 + o_P(1) \Rightarrow$$
$$Re^{-t_{0,m}^2/2+R^2 t_{0,m}^2/2} = Re^{-(1-R^2)t_{0,m}^2/2} = 1 + o_P(1).$$

Hence $R = t_m/t_{0,m} \to 1$ in probability. Thus

$$t_{0,m}^2 - t_m^2 = o_P(1) \Rightarrow |t_{0,m} - t_m| = \frac{o_P(1)}{1 + |t_{0,m} + t_m|} = O_p((logm)^{-1/2}),$$

since $t_m = o_P(n^{1/6})$ and $\log m = o(n^{1/3})$. $\square$

B.2 Proof of Theorem 2.1.4

In this section, we give the proof of the rate of convergence for the i.i.d. case by using the one-sample t-statistic. Let $p(t) = P(|T_1| \geq t)$ and let

$$\hat{p}_m(t) = \frac{1}{m} \sum_{i=1}^{m} I_{\{|T_i| \geq t\}}.$$

By the Glivenko-Cantelli theorem,

$$\sup_t |\hat{p}_m(t) - p(t)| \to 0 \quad a.s., \tag{B.11}$$

and, by the Donsker theorem,

$$\sup_t |\hat{p}_m(t) - p(t)| = O(m^{-1/2}) \text{ in probability.} \tag{B.12}$$

By the uniform law of the iterated logarithm,

$$\sup_t |\hat{p}_m(t) - p(t)| = O(m^{-1/2}(loglogm)^{1/2}) \quad a.s.. \tag{B.13}$$

We define strong consistent estimators of $E\xi_1(t)$ and $\text{Var}(\xi_1(t))$ by $\nu_m(t)$ and $\tau_m^2(t)$ respectively, where

$$\nu_m(t) = \alpha \hat{p}_m(t) - (1 - \pi_1)P(|Z| \geq t) \tag{B.14}$$

and

$$\begin{aligned}\tau_m^2(t) &= \alpha^2\hat{p}_m(t)(1-\hat{p}_m(t)) + 2\alpha(1-\pi_1)\hat{p}_m(t)P(|Z|\geq t) \\ &\quad +(1-\pi_1)P(|Z|\geq t)(1-2\alpha-(1-\pi_1)P(|Z|\geq t)). \end{aligned} \tag{B.15}$$

Now we define an estimator of $t_{n,m}$ by

$$\hat{t}_{n,m} = \inf\{t : \frac{\sqrt{m}\nu_m(t)}{\tau_m(t)} \geq z_\gamma\}. \tag{B.16}$$

For FDTP, we have to show that

$$|\hat{t}_{n,m} - t_{n,m}| = O(\frac{1}{\sqrt{n}} + (\frac{\log\log m}{m})^{1/2}) \quad a.s. \tag{B.17}$$

and

$$|\hat{t}_{n,m} - t_{n,m}| = O(n^{-1/2} + m^{-1/2}) \text{ in probability.} \tag{B.18}$$

Below we prove $(B.17)$ and $(B.18)$. We will show that

$$|\hat{t}_{n,m} - t_1| = O((\frac{1}{n})^{1/2} + (\frac{\log\log m}{m})^{1/2}) \quad a.s., \tag{B.19}$$

$$|t_{n,m} - t_1| = O((\frac{1}{n})^{1/2} + (\frac{\log\log m}{m})^{1/2}) \quad a.s.. \tag{B.20}$$

By the uniform law of the iterated logarithm,

$$\sup_t |\hat{p}_m(t) - p(t)| = O((\frac{\log\log m}{m})^{1/2}) \quad a.s.. \tag{B.21}$$

So we have

$$\sup_t |v_m(t) - [\alpha p(t) - (1-\pi_1)P(|Z|\geq t)]| = O((\frac{\log\log m}{m})^{1/2}) \quad a.s.. \tag{B.22}$$

Note that

$$\alpha p(t) - (1-\pi_1)P(|Z|\geq t) - G(t)$$

$$
\begin{aligned}
&= \alpha(1-\pi_1)(P(|T_1| \geq t|H_1 = 0) - P(|Z| \geq t)) \\
&+ \alpha\pi_1(P(|T_1| \geq t|H_1 = 1) - EP(|Z + \sqrt{n}\mu_1/\sigma_1| \geq t)).
\end{aligned}
$$

From $(A.3)$, we obtain

$$
P(|T_1| \geq t|H_1 = 0) - P(|Z| \geq t) = O(\frac{1}{\sqrt{n}}) \quad a.s. \tag{B.23}
$$

and

$$
P(|T_1| \geq t|H_1 = 1) - EP(|Z + \sqrt{n}\mu_1/\sigma_1| \geq t) = O(\frac{1}{\sqrt{n}}) \quad a.s.. \tag{B.24}
$$

Thus we have

$$
\sup_t |\alpha p(t) - (1-\pi_1)P(|Z| \geq t) - G(t)| = O(\frac{1}{\sqrt{n}}) \quad a.s.. \tag{B.25}
$$

Taking into account $(B.22)$, we have

$$
\sup_t |v_m(t) - G(t)| \leq c_2(\frac{1}{\sqrt{n}} + (\frac{\log\log m}{m})^{1/2}) \quad a.s. \tag{B.26}
$$

for some constant $0 < c_2 < \infty$. Below we show that there exists a finite constant $c_3 > 0$ such that

$$
t_1 - c_3(\frac{1}{\sqrt{n}} + (\frac{\log\log m}{m})^{1/2}) < \hat{t}_{n,m} < t_1 + c_3(\frac{1}{\sqrt{n}} + (\frac{\log\log m}{m})^{1/2}). \tag{B.27}
$$

Recalling $(B.26)$, we have, for $\epsilon = c_3(\frac{1}{\sqrt{n}} + (\frac{\log\log m}{m})^{1/2})$, that

$$
\begin{aligned}
v_m(t_1+\epsilon) &\geq G(t_1+\epsilon) - c_2(\frac{1}{\sqrt{n}} + (\frac{\log\log m}{m})^{1/2}) \\
&= G(t_1) + \epsilon G'(t_1+\theta_1) - c_2(\frac{1}{\sqrt{n}} + (\frac{\log\log m}{m})^{1/2}) \\
&\geq c_1\epsilon - c_2(\frac{1}{\sqrt{n}} + (\frac{\log\log m}{m})^{1/2}) > 2\Big(\frac{\log\log m}{m}\Big)^{1/2},
\end{aligned}
$$

provided that c_3 is chosen large enough: here $0 \leq \theta_1 \leq \epsilon$ and we used Lemma A.3.3. For sufficiently large m, we have

$$
\sqrt{m}v_m(t_1+\epsilon) > \tau_m(t_1+\epsilon)z_\gamma.
$$

This proves

$$\hat{t}_{n,m} - t_1 \leq c_3((\frac{1}{n})^{1/2} + (\frac{\log\log m}{m})^{1/2}) \quad a.s..$$

Similarly, we have

$$\hat{t}_{n,m} - t_1 \geq -c_3((\frac{1}{n})^{1/2} + (\frac{\log\log m}{m})^{1/2}) \quad a.s..$$

This proves $(B.19)$.

Following the same line of proof, we have

$$|t_{n,m} - t_1| = O(\frac{1}{\sqrt{n}} + (\frac{\log\log m}{m})^{1/2}) \quad a.s..$$

If we use

$$\sup_t |\hat{p}_m(t) - p(t)| = O(m^{-1/2}) \quad \text{in probability} \tag{B.28}$$

based on the Donsker theorem instead of $(B.21)$, using the same line of the proof of the a.s. convergence rate, we can obtain the rate of convergence in probability, which is

$$|\hat{t}_{n,m} - t_{n,m}| = O(n^{-1/2} + m^{-1/2}) \qquad \text{in probability.}$$

This completes the proof of $(B.17)$.

Similarly, the critical value for FDR control is bounded due to the fact that

$$EP(|Z + \frac{\sqrt{n}\mu_1}{\sigma_1}| \geq t) \leq 1.$$

By (B.12), (B.13), (B.23) and (B.24), we have

$$\sup_t |\frac{m_0 F_0(t)}{m_0 F_0(t) + m_1 F - 1(t)} - \frac{2(1-\pi_1)\bar{\Phi}(t)}{\hat{p}_m(t)}| = O(n^{-1/2} + (\frac{\log\log m}{m})^{1/2}) \quad a.s.$$
$$\sup_t |\frac{m_0 F_0(t)}{m_0 F_0(t) + m_1 F - 1(t)} - \frac{2(1-\pi_1)\bar{\Phi}(t)}{\hat{p}_m(t)}| = O(n^{-1/2} + (m)^{-1/2}) \quad \text{in probability.}$$

Noting that $2(1-\pi_1)\bar{\Phi}(t)/[2(1-\pi_1)\bar{\Phi}(t) + EP(|Z + \sqrt{n}\mu_1/\sigma_1| \geq t)]$ is a monotone decreasing continuous function with respect to t combined with the definition of $(t^{fdr}_{n,m})$ and $(\hat{t}^{fdr}_{n,m})$, (2.34) and (2.35) hold.

The proof of k-FWER is the same as that given in Theorem 2.1.2. □

APPENDIX C

Proof of Theorems in chapter 3

C.1 Proof of Theorem 3.1.1

For the two-sample t-statistic, the only part we need to show is the boundedness of $t_{n,m}$ under independence, which will imply the boundedness in the general dependence case as happens with the one-sample t-statistic. The remaining results follow along the same lines as the proof in the one sample t-statistic setting. Based on lemma C.1.1 below, plus (3.1.1), and using the same line of proof as in the one-sample t-statistic case, the boundedness of $t_{n,m}$ holds for two-sample t-statistics.

The proof of the boundedness of $t_{n,m}$ is based on the following asymptotic distribution of T_i^* under the alternative hypothesis:

Lemma C.1.1. *Suppose $X, X_1, \cdots, X_{n_1}$ are independent and identically distributed random variables from a population with mean μ_1 and variance σ_1^2; $Y, Y_1, \cdots, Y_{n_2}$ are independent and identically distributed random variables from another population with mean μ_2 and variance σ_2^2. Assume the sampling processes are independent of each other. Assume also that there are $0 < c_1 \leq c_2 < \infty$ such that $c_1 \leq n_1/n_2 \leq c_2$. Let*

$$T^* = \frac{\bar{X} - \bar{Y}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}, \tag{C.1}$$

where

$$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i, \qquad \bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i, \tag{C.2}$$

$$s_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \bar{X})^2, \quad \textit{and} \quad s_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2. \tag{C.3}$$

If $EX^4 < \infty$ and $EY^4 < \infty$, then

$$P(|T^*| \geq t) = P(|Z + \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}| \geq t)(1 + o(1)) \tag{C.4}$$

uniformly in $t = o(n^{1/6})$, where $n = \max\{n_1, n_2\}$.

The proof of this lemma is very similar to the proof of Lemma A.2.1 and we omit the details. □.

C.2 Proof of Theorem 3.1.2

This follows the same arguments as in the one-sample t-statistic case by virtue of lemma C.1.1.

C.3 Proof of Theorem 3.1.3

When we plug in an estimator of $P(|T_i^*| \geq t)$,

$$\hat{p}_m(t) = \frac{1}{m}\sum_{i=1}^{m} I_{\{|T_i^*| \geq t\}},$$

the proof of the two-sample t-statistic case is along the same lines as its one-sample counterpart except that we have to show the rate of convergence under the alternative hypothesis for the two-sample t-statistic. This follows from the following lemma which completes the proof of Theorem 3.1.3.

Lemma C.3.1. *Let $X, X_1, \cdots, X_{n_1}$ be i.i.d. random variables from a population with mean μ_1 and variance σ_1^2; $Y, Y_1, \cdots, Y_{n_2}$ be i.i.d. random variables from another population with mean μ_2 and variance σ_2^2. The sampling processes are assumed to be independent of each other. Assume that there are $0 < c_1 \leq c_2 < \infty$ such that $c_1 \leq n_1/n_2 \leq c_2$. Let T^* be defined as in Lemma C.1.1. If $E|X|^4 < \infty$ and $E|Y|^4 < \infty$, then*

$$|P(T^* \leq x) - \Phi(x - \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}})| \leq \frac{K(1+|x|)}{(1+|x - \frac{\mu_1-\mu_2}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}|)\sqrt{min\{n_1, n_2\}}}. \quad \text{(C.5)}$$

where K is a finite constant that may depend on $\sigma_1^2, \sigma_2^2, E|X|^3, E|Y|^3, EX^4$ and EY^4.

Proof. Without loss of generality, we assume $n_1 = b_1 n$, $n_2 = b_2 n$, $b_1 + b_2 = 1$ with $b_1 > 0$ and $b_2 > 0$. Note that

$$\begin{aligned}
P(T^* \leq x) &= P(\frac{\bar{X} - \mu_1 - (\bar{Y} - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} + \frac{\mu_1 - \mu_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \leq x) \\
&= P(\frac{\bar{X} - \mu_1 - (\bar{Y} - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} + \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \leq x\frac{\sqrt{s_1^2/n_1 + s_2^2/n_2}}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}) \\
&\leq P(\frac{\bar{X} - \mu_1 - (\bar{Y} - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \leq x - \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} + x|\frac{s_1^2/n_1 + s_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2} - 1|),
\end{aligned}$$

where we make use of $(A.4)$. Now we apply $(A.2)$ with $\xi_i = \frac{(X_i - \mu_1)/n_1}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}$ for $1 \leq i \leq n_1$ and $\xi_i = -\frac{(Y_i - \mu_2)/n_2}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}$ for $n_1 + 1 \leq i \leq n_1 + n_2$. Let

$$z = x - \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}, \quad \Delta = -x|\frac{s_1^2/n_1 + s_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2} - 1|,$$

$$\Delta_i = -x|\frac{s_{1,i}^2/n_1 + s_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2} - 1|,$$

for $1 \leq i \leq n_1$, and

$$\Delta_i = -x|\frac{s_1^2/n_1 + s_{2,i}^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2} - 1|,$$

for $n_1 + 1 \leq i \leq n_1 + n_2$, where $s_{1,i}^2$ is defined as s_1^2 with 0 to replace X_i and $s_{2,i}^2$ is defined as s_2^2 with 0 to replace Y_i. Noting that

$$\frac{s_1^2/n_1 + s_2^2/n_2}{\sigma_1^2/n_1 + \sigma_2^2/n_2} - 1 = \frac{1}{\sigma_1^2/n_1 + \sigma_2^2/n_2}[(s_1^2 - \sigma_1^2)/n_1 + (s_2^2 - \sigma_2^2)/n_2],$$

we have by $(A.8)$ that

$$E|\frac{s_1^2/n_1+s_2^2/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2}-1|^2 \le K\frac{EX^4+EY^4}{n}.$$

For $1 \le i \le n_1$,

$$\begin{aligned} & E(\frac{s_1^2/n_1+s_2^2/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2}-\frac{s_{1i}^2/n_1+s_2^2/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2})^2 \\ = & \frac{1}{n_1^2(\sigma_1^2/n_1+\sigma_2^2/n_2)^2}E(s_1^2-s_{1i}^2)^2 \le \frac{K\,EX^4}{n^2} \end{aligned}$$

by $(A.9)$. Similarly for $n_1+1 \le i \le n_1+n_2$, we have

$$\begin{aligned} & E(\frac{s_1^2/n_1+s_2^2/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2}-\frac{s_1^2/n_1+s_{2i}^2/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2})^2 \\ = & \frac{1}{n_2^2(\sigma_1^2/n_1+\sigma_2^2/n_2)^2}E(s_2^2-s_{2i}^2) \le \frac{K\,EY^4}{n^2}. \end{aligned}$$

It follows that

$$\begin{aligned} ||\Delta||_2 &\le K\frac{|x|\sqrt{EX^4+EY^4}}{\sqrt{n}}, \\ P(|\Delta| > \frac{|z|+1}{3}) &\le K\frac{E|\Delta|}{|z|+1} \le K\frac{||\Delta||_2}{|z|+1} \le K\frac{|x|\sqrt{EX^4+EY^4}}{\sqrt{n}(|z|+1)}, \\ \sum_{i=1}^n (E\xi_i^2)^{1/2}(E(\Delta-\Delta_i)^2)^{1/2} &\le K\frac{\sqrt{(\sigma_1^2+\sigma_2)(EX^4+EY^4)}}{\sqrt{n}}, \\ \sum_{i=1}^n E|\xi_i|^3 &\le K\frac{E|X|^3+E|Y|^3}{\sqrt{n}}. \end{aligned}$$

Therefore, by $(A.2)$,

$$\begin{aligned} |P(\frac{\bar{X}-\mu_1-(\bar{Y}-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}} &\le x-\frac{\mu_1-\mu_2}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}+x|\frac{s_1^2/n_1+s_2^2/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2}-1|) \\ - & \Phi(x-\frac{\mu_1-\mu_2}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}})| \le K\frac{1+|x|}{(1+|x-\frac{\mu_1-\mu_2}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}|)\sqrt{n}}. \end{aligned}$$

Similarly,

$$\begin{aligned} P(T^* \le x) &= P(\frac{\bar{X}-\mu_1-(\bar{Y}-\mu_2)}{\sqrt{s_1^2/n_1+s_2^2/n_2}}+\frac{\mu_1-\mu_2}{\sqrt{s_1^2/n_1+s_2^2/n_2}} \le x) \\ &\ge P(\frac{\bar{X}-\mu_1-(\bar{Y}-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}} \le x-\frac{\mu_1-\mu_2}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}-x|\frac{s_1^2/n_1+s_2^2/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2}-1|) \end{aligned}$$

and

$$\begin{aligned} |P(\frac{\bar{X}-\mu_1-(\bar{Y}-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}} &\le x-\frac{\mu_1-\mu_2}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}-x|\frac{s_1^2/n_1+s_2^2/n_2}{\sigma_1^2/n_1+\sigma_2^2/n_2}-1|) \\ - & \Phi(x-\frac{\mu_1-\mu_2}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}})| \le K\frac{1+|x|}{(1+|x-\frac{\mu_1-\mu_2}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}|)\sqrt{n}}. \end{aligned}$$

This proves $(C.5)$. $\square$

BIBLIOGRAPHY

Benjamini, Y. and Hochberg, Y. (1995), 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *J.R.Statist.Soc. (B)* **57**, 289–300.

Benjamini, Y. and Hochberg, Y. (1997), 'Multiple hypothesis testing with weights', *Scandinavian Journal of Statistics* **24**, 407–418.

Benjamini, Y. and Hochberg, Y. (2000), 'On the adaptive control of the false discovery rate in multiple testing with independent statistics', *Journal of Educational and Behavioral Statistics* **25**, 60–83.

Benjamini, Y. and Yekutieli, D. (2001), 'The control of the false discovery rate in multiple testing under dependency', *Ann. Statist.* **29**(4), 1165–1188.

Cao, H. (2007), 'Moderate deviations for two sample t-statistics', *ESAIM P & S* **11**, 264–271.

Chen, L. and Shao, Q. (2007), 'Normal approximation for nonlinear statistics using a concentration inequality approach', *Bernoulli* **13**, 581–599.

Chi, Z. (2007), 'On the performance of fdr control: constraints and a partial solution', *Ann. Statist* **35**, 1409–1431.

Chi, Z. and Tan, Z. (2008), 'Positive false discovery proportions: intrinsic bounds and adaptive control', *Statist. Sinica* **18**, 837–860.

Craiu, R. and Sun, L. (2008), 'Choosing the lesser evil: trade-off between false discovery rate and non-discovery rate', *Statist. Sinica.* **18**, 861–879.

Donoho, D. and Jin, J. (2004), 'Higher criticism for detecting sparse heterogeneous mixtures', *Ann. Statist.* **32**, 962–994.

Donoho, D. and Jin, J. (2006), 'Asymptotic minimaxity of false discovery rate thresholding for sparse exponential data', *Ann. Statist.* **34**, 2980–3018.

Dudley, R. and Philipp, W. (1983), 'Invariance principles for sums of banach space valued random elements and empirical processes', *Z. Wahrsch. Verw. Gebiete* **62**, 509–552.

Dudoit, S., Shaffer, P. and Boldrick, J. (2003), 'Multiple hypothesis testing in microarray genomics', *Statistical Science* **18**, 71–103.

Dudoit, S. and van der Laan, M. (2008), *Multiple testing procedures with applications to genomics*, Springer, New York.

Efron, B. and Tibshirani, R. (2002), 'Empirical bayes methods and false discovery rates for microarrays', *Genetics Epidemiology* **23**, 70–86.

Efron, B., Tibshirani, R., Storey, J. D. and Tusher, V. (2001), 'Empirical Bayes analysis of a microarray experiment', *J. Amer. Statist. Assoc.* **96**, 1151–1160.

Fan, J., Hall, P. and Yao, Q. (2007), 'To how many simultaneous hypothesis tests can normal, students t or bootstrap calibration be applied?', *JASA* **19**, 1282–1288.

Genovese, C., Roeder, K. and Wasserman, L. (2006), 'False discovery control with p-value weighting', *Biometrika* **93**, 509–524.

Genovese, C. and Wasserman, L. (2002), 'Operating characteristics and extensions of the false discovery rate procedure', *J. R. Stat. Soc. B* **64**, 499–517.

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

Genovese, C. and Wasserman, L. (2006), 'Exceedance control of the false discovery proportion', *JASA* **101**, 1408–1417.

Golub, T. R. e. (1999), 'Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring', *Science* **286**, 531–537.

Kosorok, M. (2008), *Introduction to Empirical Processes and Semiparametric Inference*, Springer, New York.

Kosorok, M. and Ma, S. (2007), 'Marginal asymptotics for the "large p, small n" paradigm: with application to microarray data.', *Ann. Statist.* **35**, 1456–1486.

Langaas, M., Lindqvist, B. H. and Ferkingstad, E. (2005), 'Estimating the proportion of true null hypotheses, with application to dna microarray data.', *J. Roy. Statist. Soc. B* **67**, 555572.

Lehmann, E. L. and Romano, J. P. (2005*a*), 'Generalizations of the familywise error rate', *Ann. Statist.* **33**(3), 1138–1154.

Lehmann, E. L. and Romano, J. P. (2005*b*), *Testing statistical hypotheses*, Springer Texts in Statistics, third edn, Springer, New York.

Lehmann, E. L., Romano, J. P. and Shaffer, J. P. (2005), 'On optimality of stepdown and stepup multiple test procedures', *Ann. Statist.* **33**, 1084–1108.
URL: *http://dx.doi.org/10.1214/009053605000000066*

Meinshausen, N. and Bühlmann, P. (2005), 'Lower bounds for the number of false null hypotheses for multiple testing of associations', *Biometrika* **92**, 893–907.

Meinshausen, N. and Rice, J. (2006), 'Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses.', *Ann. Statist.* **34**, 373–393.

Sarkar, S. K. (2002), 'Some results on false discovery rate in stepwise multiple testing procedures', *Ann. Statist.* **30**, 239–257.

Sebastiani, P., Gussoni, E., Kohane, I. and Ramoni, M. (2003), 'Statistical chanllenges in functional genomics', *Statistical Science* **18**, 33–70.

Seeger, P. (1968), 'A note on a method for the analysis of significane en masse', *Technometrics* **10**, 586–593.

Storey, J. D. (2002), 'A direct approach to false discovery rates', *J. R. Stat. Soc. B* **64**, 479–498.

Storey, J. D. (2003), 'The positive false discovery rate: a Bayesian interpretation and the q-value', *Ann. Statist.* **31**, 2013–2035.

Storey, J. D. and Tibshirani, R. (2003), 'Statistical significance for genomewide studies', *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9440–9445.

Storey, J. D., Tibshirani, R. and Siegmund, D. (2004), 'Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach', *J.R. Statist. Soc. (B)* **66**, 187–205.

Sun, W. and Cai, T. T. (2007), 'Oracle and adaptive compound decision rules for false discovery rate control', *J. Amer. Statist. Assoc.* **102**, 901–912.

Sun, W. and Cai, T. T. (2009), 'Large-scale multiple testing under dependence', *J. R. Stat. Soc. B* **71**, 393–424.

van der Laan, M., Dudoit, S. and Pollard, K. (2004), 'Augmentation procedures for control of the generalized family-wise error rate and tail probabilities for the proportion of false positives', *Statistical Applications in Genetics and Molecular Biology* .

Wang, Q. (2008), Absolute and relative errors in central limit theorem for self-normalized sums: review and new results.

Wang, Q. and Hall, P. (2009), 'Relative errors in central limit theorem for student's t statistics with applications', *Statist. Sinica.* **19**, 343–354.

Wu, W. B. (2008), 'On false discovery control under dependence', *Ann. Statist.* **36**(1), 364–380.