Large-scale Asymptotics in Fixed-sample and Sequential Multiple Testing

Rahul Roy

Interdisciplinary Statistical Research Unit

Indian Statistical Institute, Kolkata

2025

INDIAN STATISTICAL INSTITUTE, KOLKATA

DOCTORAL THESIS

Large-scale Asymptotics in Fixed-sample and Sequential Multiple Testing

(Revised Version)

Author:

Rahul Roy

Supervisors:

Prof. Subir K. Bhandari

Prof. Shyamal K. De

A thesis submitted to the Indian Statistical Institute in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Statistics

in the

Interdisciplinary Statistical Research Unit

Indian Statistical Institute, Kolkata

INDIAN STATISTICAL INSTITUTE
203 Barrackpore Trunk Road,
Kolkata 700108, India

Certificate

This is to certify that this revised version of doctoral thesis entitled *"Large-scale Asymptotics in Fixed-sample and Sequential Multiple Testing"*, submitted by Rahul Roy to Indian Statistical Institute is a record of bonafide research work under our guidance and supervision. The work contained in this thesis is original and has not been submitted elsewhere for any degree.

Place: ISI Kolkata

Date: 16th June, 2026

Prof. Subir Kumar Bhandari
Professor,
Interdisciplinary Statistical
Research Unit
Indian Statistical Institute
Kolkata–700108, India.

Dr. Shyamal Krishna De
Assistant Professor,
Applied Statistics Unit
Indian Statistical Institute
Kolkata–700108, India.

Declaration of Authorship

I, Rahul Roy, declare that this thesis titled, "Large-scale Asymptotics in Fixed-sample and Sequential Multiple Testing" and the work presented in it are my own. I confirm the following statements:

- This work was done while in candidature for a research degree at Indian Statistical Institute, Kolkata.
- No part of this dissertation has been submitted for a degree or any other qualification at this institute or elsewhere.
- I have clearly cited all the previously published works I have consulted for writing this dissertation.
- I have properly acknowledged all main sources of data used in this dissertation and the methods on which I have built my work.
- I have properly acknowledged all my co-authors who have helped in developing different parts of this dissertation.

Place : ISI Kolkata
18th June, 2026

Rahul Roy
External Research Fellow
Interdisciplinary Statistical Research Unit
Indian Statistical Institute
Kolkata 700108, India

Abstract

The advent of large-scale data acquisition technologies has led to the routine emergence of massive, dynamically evolving datasets. This has created a pressing need for statistical methodologies capable of operating effectively in both static and sequential data environments. Multiple hypotheses testing, a central tool in such settings, must therefore be adapted to both fixed-sample and sequential frameworks to ensure effective decision-making while controlling error rates.

This thesis comprises two main parts. The first part addresses fixed-sample multiple testing under dependence in a Bayesian framework. Building upon the two-group mixture normal model of Bogdan et al. (2011), we extend their methodology to exchangeable multivariate normal test statistics, thereby accommodating realistic dependency structures frequently encountered in high-dimensional applications. We derive sufficient conditions under which multiple testing procedures satisfy the *Asymptotic Bayes Optimality under Sparsity* (ABOS) property in the presence of dependence. Furthermore, we show that several classical procedures, including those of Bonferroni (1936), Šidák (1967), and Benjamini and Hochberg (1995), retain the ABOS property under suitable sparsity and dependence assumptions.

The second part focuses on large-scale multiple testing in sequential settings, where data vectors or multiple data streams are observed over time rather than being available in full initially. Motivated by diverse applications, we investigate two distinct sampling termination schemes: (i) *synchronous* termination, in which sampling and testing for all hypotheses stop simultaneously; and (ii) *asynchronous* termination, in which different hypotheses may stop at different times.

For the synchronous case, we develop the *Oracle Intersection (OI)* test, based on the local false discovery rate statistic under a two-group mixture model. The OI test guarantees exact control of both the False Discovery Rate (FDR) and the False Non-Discovery Rate (FNR) at prespecified levels. We further propose a fully data-driven version that achieves asymptotic simultaneous control of FDR and FNR as the number of hypotheses $m \to \infty$. While existing sequential tests use fixed stopping boundaries, the proposed tests employ stopping boundaries that adapt quickly with the sample size, ensuring shrinkage of the continue-sampling region as the trial

progresses. As a result, stopping times for both the oracle and data-driven tests converge to a finite constant as $m \to \infty$. Moreover, the ratio of the expected sample size of the OI test to that of the Gap rule (He and Bartroff, 2021) converges to zero as $m \to \infty$. Extensive analyses of real and simulated datasets demonstrate the superiority of the proposed methods over existing approaches.

Finally, we address the case of asynchronous termination and introduce the *Oracle Stagewise (OS)* test, constructed from the local false discovery rate statistic under a two-group mixture model. Employing adaptive stopping boundaries, the OS test drops hypotheses by accepting or rejecting them at interim stages of sampling until all hypotheses are decided. This stagewise procedure achieves simultaneous control of the FDR and FNR, while substantially reducing the total sample size relative to existing sequential methods (Bartroff and Song, 2020). To address challenges arising from composite hypotheses and the instability of parameter estimation caused by shrinking active sets, we further propose the *Oracle Composite (OC)* and *Data-driven Composite (DC)* tests. These hybrid procedures combine stagewise elimination with intersection-based testing, ensuring reliable estimation and improved practical applicability. Simulation studies demonstrate that the proposed methods substantially reduce the total sample size relative to existing sequential procedures while maintaining rigorous error control.

Overall, this thesis advances the theory of large-scale multiple testing in both fixed-sample and sequential settings by establishing new optimality results under dependence and developing adaptive procedures that simultaneously achieve rigorous error control, finite stopping times, and improved sampling efficiency.

Acknowledgements

First and foremost, I would like to thank my supervisors, Professor Subir Kumar Bhandari and Dr. Shyamal Krishna De. Throughout my time as a PhD student, Professor Bhandari was not only an academic mentor but also a personal guide through many of life's thick and thin moments. He helped me become more imaginative, encouraging me to visualize complex mathematical problems using simple analogies, thereby developing deeper intuition. Professor De, on the other hand, instilled in me the importance of mathematical rigor. His guidance kept me moving forward even when no apparent solution was in sight. I would not be who I am today without their constant support and mentorship.

I am grateful to Prof. Sourabh Bhattacharya and Prof. Arijit Chakraborty for their valuable feedback and continual suggestions, which improved the quality of my research. A special mention goes to Monitirtha and Nabaneet da for innumerable discussions on multiple testing, leading to valuable insights.

I extend my sincere thanks to the faculty of ISRU, ASU, SMU, HGU, SOSU, ERU, and ACMU at the Indian Statistical Institute, Kolkata, especially Dr. Jayant Jha, Professor Anil Kumar Ghosh, Professor Debashis Sengupta, and Professor Indranil Mukhopadhyay, for always being available to address my queries.

My time at ISI would have been incomplete without a supportive peer group. Soumya, Nabaneet da, Sujay da, Pushpinder, Arijit da, Adhidev da, Subhankar, Dibyarka, Sonaksha, Sayantan, Biswadeep, Mriganka, Meghna, Upama, Sampurna, Manitirtha, Anik, Sayan, and many others made campus life both bearable and memorable. I would also like to thank Abhunandan da, Soumya da, and Mama for inspiring me to pursue goals beyond academics.

My colleagues in the Department of Statistics at St. Xavier's College deserve special appreciation for their continuous support. I thank Dr. Surabhi Dasgupta, Dr. Surupa Chakraborty, Dr. Durba Bhattacharya, Dr. Debjit Sengupta, Mrs. Madhura Das Gupta, Mrs. Pallabi Ghosh, and Dr. Sancharee Basak for their guidance and encouragement throughout this journey.

I also wish to thank my teachers from the Department of Statistics at RKMRC

Narendrapur and ISI Delhi. Their early teachings laid the foundation for this research endeavor.

The pandemic brought with it unprecedented challenges. Mental health struggles became prevalent as we were confined to our homes. I am thankful to my school friends—Debarshi, Utsab, Tuhin, Dwaipayan, Srijan, Tapajit, Arnab, Pronoy, and others—who helped me navigate those dark times.

I would like to express my deep gratitude to my late Math teacher, Mr. Ranjan Ray. Without his influence, my journey into the field of statistics might have ended far earlier. May his soul rest in peace.

There is no greater blessing than the unwavering support of one's parents. I am profoundly grateful to my parents, Mrs. Swati Ghosh Ray and Mr. Ratan Kumar Ray, for their dedication, sacrifice, and strength throughout my academic and personal journey—despite the many challenges we faced.

Finally, I express my deepest appreciation to Mrs. Debaleena Kar, my longtime partner and now my wife. Without her constant encouragement and belief in me, I may never have pursued research. I dedicate this thesis to her.

Rahul Roy

Publications and Preprints

1. A version of chapter 2 of this thesis has been published online as the following

   Roy, Rahul and Subir Kumar Bhandari (2024). "Asymptotic Bayes' optimality under sparsity for exchangeable dependent multivariate normal test statistics". In:*Statistics & Probability Letters* 207, 110030.

   DOI: https://doi.org/10.1016/j.spl.2023.110030

2. Chapter 3 is based on the following preprint

   Roy, Rahul, Shyamal Krishna De, and Subir Kumar Bhandari (2023). "Large-scale adaptive multiple testing for sequential data controlling false discovery and nondiscovery rates." https://doi.org/10.48550/arXiv.2306.05315.

3. The manuscript for chapter 4 is under preparation.

Contents

Certificate iii

Declaration of Authorship v

Abstract vii

Acknowledgements ix

Publications and Preprints xi

1 Introduction 1

1.1 Multiple Hypotheses Testing 1

1.1.1 Applications of Multiple Hypotheses Testing 2

1.1.2 Fundamentals of multiple hypotheses testing 6

1.1.3 Frequentist and Bayesian Approaches of Multiple Hypotheses Testing 8

1.1.4 Measures of Type I and Type II Errors 9

1.2 Some Notable Multiple Hypotheses Testing Procedures 12

1.2.1 Large-scale Multiple Hypotheses Testing 14

1.3 Sequential Multiple Hypotheses Testing 17

1.3.1 Asynchronous Termination of Data Streams 18

1.3.2 Simultaneous Termination of Data Streams 18

1.4 Motivation and Overview of This Thesis 19

1.4.1 Asymptotically Bayes Optimality under Sparsity in a Dependent Framework 20

1.4.2 Large-scale Sequential Multiple Hypotheses Testing 20

2 Asymptotic Bayes Optimality under Sparsity for Exchangeable Dependent Multivariate Normal Test Statistics **23**

2.1 Introduction . . . 23

2.2 A Decomposition of the Observations . . . 27

2.3 Hypothetical Rules . . . 31

2.4 Approximate Rules . . . 33

2.5 Conditions for Some Classical and Bayesian Multiple Testing Methods to be ABOS . . . 35

2.5.1 Preliminary Concepts of Multiple Hypotheses Testing . . . 35

2.5.2 Bayesian FDR Control . . . 36

2.5.3 Optimality of the Asymptotic Approximation to the BH Threshold . . . 37

2.5.4 Asymptotic Optimality Under Sparsity for Bonferroni's Procedure . . . 38

2.5.5 Asymptotic Optimality Under Sparsity for Šidák's Procedure . 38

2.5.6 Asymptotic Optimality Under Sparsity for Benjamini-Hochberg Test . . . 39

2.6 Proof of Theorems . . . 40

2.6.1 Proof of Theorem 1 . . . 40

2.7 Proof of Lemmas . . . 47

3 Large-scale Sequential Multiple Testing with Simultaneous Termination of Sampling **53**

3.1 Data Description and Problem Setup . . . 53

3.1.1 Oracle Test Statistics . . . 54

3.2 Oracle Test Procedure With Exact Error Control . . . 56

3.2.1 Oracle Intersection Test . . . 56

3.2.2 Asymptotic Properties of the Oracle Stopping Time . . . 61

3.2.3 Comparison with Gap rule . . . 63

3.3 Data-driven Test With Asymptotic Error Control . . . 65

3.4 Proof of Theorems . . . 69

3.5 Proof of Lemmas . . . 80

4 Large-scale Sequential Multiple Testing with Asynchronous Termination of Sampling **89**

4.1 Data Description and Problem Setup 89

4.2 Oracle Test Statistics 91

4.3 Oracle Tests with Exact Error Control 92

4.3.1 The Oracle Stagewise (OS) Test 93

4.3.2 Oracle Composite (OC) Test 95

4.4 Data-driven Stagewise Tests 98

4.5 Proofs of Theorems 100

4.6 Proof of Lemmas 105

5 Numerical Study **111**

5.1 Numerical Study on the OI and DI Tests 111

5.1.1 Comparison Between Sequential Tests 112

5.1.2 Comparing the DI and BH Tests 113

5.2 Real Data Analysis 116

5.3 Simulation Studies for the DS and DC Tests 118

5.3.1 Case 1: Simple Hypotheses 121

5.3.2 Case 2: Composite Hypotheses 123

6 Discussion and Future Directions **129**

6.1 ABOS Tests for the Equicorrelated Setup 129

6.2 Large-Scale Sequential Multiple Testing with Simultaneous Termination of Sampling 130

6.3 Large-Scale Sequential Multiple Testing with Asynchronous Termination of Sampling 132

Bibliography **135**

List of Figures

3.1 Sampling scheme of an OI test via adaptive stopping boundaries . . . 60
3.2 Comparison of estimated expected sample sizes (ASN) for OI test, DI test and Gap rules for the setup Ex1 in Chapter 5 68
3.3 Comparison of estimated FDR and FNR for OI test, DI test and Gap rules for the setup Ex1 in Chapter 5 68
3.4 The $Q_n'(t)$, $Q_n'^{+}(t)$ and $Q_n'^{-}(t)$ functions for $m = 10$. 83

5.1 Histogram of p-values and z-scores for the full Prostate dataset 118
5.2 Histograms of p-value and z-scores of Golub data 118

List of Tables

1.1 Truth Table for Multiple Hypotheses Testing 6

5.1 Comparing OI test, DI test and the Gap rules 114
5.2 Comparison between the DI test and the BH test 115
5.3 Comparing OS, DS and Seq-BH tests for the setup in EX-1 123
5.4 Comparing OS, DS and Seq-BH tests for the setup in EX-2 123
5.5 Comparing OS, DS and Seq-BH tests for the setup in EX-3 123
5.6 Comparing OC, DC and Seq-BH tests for the setup in EX-4 126
5.7 Comparing OC, DC and Seq-BH tests for the setup in EX-5 127
5.8 Comparing OC, DC and Seq-BH tests for the setup in EX-6 127
5.9 Comparing OC and DC test results for the setup in EX-7 127
5.10 Comparing OC and DC test results for the setup in EX-8 127

Chapter 1

Introduction

1.1 Multiple Hypotheses Testing

As we have entered a new millennium, a clear shift in the meaning of data compared to even 50 years ago has become evident. A gradual leap in science and technology and our interaction with them has driven this conceptual change (Fan, Han, and Liu (2014), Donoho et al. (2000), and Johnstone (2007)) . Today, ultra-high-definition telescopes such as the JWST can capture clear images of galaxies at unimaginably vast distances (Begley et al. (2026)). High-throughput machines can extract genetic information in the form of gene expression values (VanGuilder, Vrana, and Freeman (2008)), sequences of SNPs (Ding and Guo (2018)), or methylation patterns (Laird (2010)). The interaction and engagement of billions of people across the globe through various social media platforms generate vast amounts of data every day (Drus and Khalid (2019)).

The introduction of new types of datasets requires us to store and analyze them. Modern database management systems help in storing and processing these large volumes of data (Diène et al. (2020)). For data analysis, depending on the purpose and the nature of the dataset, various statistical (Fan, Han, and Liu (2014)), machine learning, and deep learning techniques (Rane et al. (2024)) have been proposed. Multiple hypotheses testing procedures serve as one class of statistical tools to deal with large datasets. When several hypotheses need to be tested simultaneously, a naive approach might be to test each hypothesis at a fixed significance level $\alpha \in (0, 1)$. If these hypotheses are designed to answer a broader or global question, it is appropriate to consider them as a family of hypotheses and to control an overall error metric

associated with testing that family. In such cases, testing each hypothesis at level α leads to inflation of both Type I and Type II error rates, and clearly fails to control composite or overall error measures such as the probability of making at least one Type I (or Type II) error (as discussed in detail later). Note that the problem of multiplicity arises when several hypotheses are tested simultaneously using individual level-α test procedures, and multiple hypotheses testing procedures aim to protect against overall errors for a family of hypotheses.

1.1.1 Applications of Multiple Hypotheses Testing

Multiple hypotheses testing is widely used across diverse fields, including clinical trials, genome-wide association studies, astrophysics, and weather forecasting. We discuss scenarios where multiple comparisons commonly arise, particularly in clinical trials as well as biomedical and genomic research.

Biomedical and Genomic Research

Microarray gene expression data can yield numerous biological insights through the use of various statistical tools (Schena et al. (1995)), such as clustering (Jothi, Mohanty, and Ojha (2019)), classification (Khashei, Hamadani, and Bijari (2012)), biclustering (Balamurugan, Natarajan, and Premalatha (2015)), Bayesian networks, and non-parametric regression models (Kim, Imoto, and Miyano (2004)), among others. Dudoit and Van Der Laan (2008) provides a thorough explanation of how multiple hypotheses testing can be applied in these domains. One such application is the identification of genes associated with a particular disease (Golub et al. (1999), Brown and Botstein (1999), Lockhart and Winzeler (2000), and Rosati et al. (2024)). Differentially expressed (DE) genes are those that behave differently depending on biological conditions (e.g., in cancerous versus non-cancerous tumor cells). DE genes can be identified by examining whether their expression levels differ significantly. Therefore, we can formulate a single hypothesis testing problem for each gene. Since data may be available for thousands of genes, and all corresponding hypotheses need to be tested simultaneously, a multiple hypotheses testing procedure suitable for large-scale datasets should be applied. The following example illustrates this point.

Suppose we aim to identify differentially expressed genes associated with prostate cancer from a collection of $m = 6033$ genes, for which gene expression data has been collected from $N_0 + N_1 = 102$ prostate tumor samples. Among these, $N_1 = 52$ samples correspond to malignant tumor cells, and the remaining $N_0 = 50$ correspond to benign tumor cells (Singh et al., 2002). Let x_{ni} denote the expression level of the i-th gene in the n-th benign sample, and let y_{ki} denote the expression level of the i-th gene in the k-th malignant sample, where $i \in [m]$, $n \in [N_0]$, and $k \in [N_1]$ (Here, for any $l \in \mathbb{N}$, we define $[l] = \{1, 2, \ldots, l\}$).

After appropriate pre-processing, we make the following distributional assumptions. For a differentially expressed gene i, the expression levels in malignant samples follow

$$y_{ki} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_{1i}, \sigma^2), \quad \text{for all } k \in [N_1].$$

For a non-differential gene i, the expression levels in benign samples follow

$$x_{ni} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_{0i}, \sigma^2), \quad \text{for all } n \in [N_0].$$

These assumptions imply that genes not involved in tumorigenesis exhibit similar expression profiles across both benign and malignant cells, whereas differentially expressed genes show a shift in expression in malignant samples. Let $\theta_i \in \{0, 1\}$ be an indicator variable denoting whether the i-th gene is differentially expressed ($\theta_i = 1$) or not ($\theta_i = 0$). The vector of true differential statuses is denoted by $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$. Our goal is to test the following m hypotheses simultaneously:

$$H_{0i} : \theta_i = 0 \quad \text{versus} \quad H_{1i} : \theta_i = 1, \quad \text{for each } i \in [m].$$

To test each individual hypothesis pair H_{0i} versus H_{1i}, we employ the two-sample t-test, which rejects the null hypothesis when the corresponding test statistic S_i exceeds a certain critical value. Specifically, under H_{0i}, the statistic S_i follows a t-distribution with $N_0 + N_1 - 2$ degrees of freedom, i.e.,

$$S_i \sim t_{(N_0+N_1-2)} \quad \text{under } H_{0i}.$$

Accordingly, we can perform individual tests at significance level α for each gene. Let the decision rule for the i-th test be defined as $\delta_i = \mathbb{I}\left(S_i > t_{\alpha;N_0+N_1-2}\right)$, where $t_{\alpha;N_0+N_1-2}$ denotes the $(1-\alpha)$ quantile of the t-distribution with $N_0 + N_1 - 2$ degrees of freedom. Alternatively, if we compute the p-value p_i corresponding to the statistic S_i, the decision rule can also be written as $\delta_i = \mathbb{I}(p_i < \alpha)$. This construction ensures that the probability of incorrectly rejecting a true null hypothesis is controlled at level α, i.e., $\mathbb{P}(\delta_i = 1 \mid H_{0i}) \leq \alpha$, for all $i \in [m]$.

For any $i \in [m]$, the quantity $(1-\theta_i)\delta_i$ indicates the event of rejecting the i-th null hypothesis when it is actually true—that is, a Type I error. The total number of Type I errors across all tests is denoted by $V = \sum_{i=1}^{m}(1-\theta_i)\delta_i$, as summarized in Table 1.1. The *familywise error rate* (FWER) is defined as the probability of making at least one Type I error: FWER $= \mathbb{P}(V > 0)$. Now, suppose that none of the m genes are associated with prostate cancer, that is, the null hypothesis H_{0i} is true for all $i \in [m]$. This scenario is referred to as the **global null hypothesis**, denoted by $\boldsymbol{\Omega_0} := \bigcap_{i\in[m]}\{\theta_i = 0\}$. Assuming independence among the test statistics, it can be shown that (Bonferroni, 1936; Šidák, 1967; Westfall and Young, 1993) under the global null,

$$\text{FWER} = 1 - (1-\alpha)^m > \alpha.$$

This inequality demonstrates that applying individual hypothesis testing procedures at level α, as described previously, leads to an inflation of the overall Type I error rate. This phenomenon is widely known as the *multiplicity problem* (Westfall and Young (1993) and Hochberg and Tamhane (1987)) in the multiple hypotheses testing literature. Notably, multiplicity issues can arise even when the global null does not hold, i.e., under more general configurations of the truth vector $\boldsymbol{\theta}$. The goal of a multiple hypotheses testing procedure is to control the FWER—or alternative error rates involving V (as discussed in Section 1.1.2)—and thus address the multiplicity problem.

Multiple Comparison in Clinical Trials

Clinical trials are designed to evaluate both the efficacy and potential side effects of a drug by examining specific outcomes known as *endpoints*. Endpoints are measurable

physiological responses or outcomes in human subjects that can be statistically analyzed to assess the effect of a treatment. Depending on their relevance to the primary objectives of the study, endpoints are typically classified into three categories: *primary*, *secondary*, and *exploratory*. A comprehensive discussion on the prevalence and implications of the multiplicity problem in clinical trials is provided in Dmitrienko, Tamhane, and Bretz (2009).

The primary clinical benefits of a drug can be evaluated using certain "win" criteria based on one or more endpoints. Below are examples of such criteria, adapted from Section 1.2 of Dmitrienko, Tamhane, and Bretz (2009), which illustrate the prevailing problem.

1. One primary endpoint needs to be statistically significant.
2. Among m primary endpoints, only one needs to be significant.
3. Among m primary endpoints, all need to be significant.
4. Given three specified primary endpoints $E1$, $E2$ and $E3$, either $E1$ needs to be statistically significant or both $E2$ and $E3$ need to be statistically significant.
5. Given three endpoints, $E1$, $E2$, and $E3$, either $E1$ and $E2$ or $E1$ and $E3$ need to be significant.

In case 1, we need to ensure only the significance of the given endpoint. There is no involvement of any other endpoints. Therefore, the multiplicity problem does not arise here.

In case 3, all the endpoints need to be significant individually. Here, the significance of each of these endpoints is important, and the significance of one endpoint does not influence the same for the other. Therefore, we can test each of these hypotheses at the same level of significance.

In cases 2, 4, and 5, we observe that the hypotheses (i.e., whether the individual endpoints are significant or not) jointly deliver a meaning to the underlying problem and cannot be broken down into individual hypotheses. In other words, in describing the actual problem, the individual hypotheses become intertwined in a complex amalgam (of unions and intersections). Therefore, if we try to test the single hypotheses at a certain significance level, the multiplicity problem will arise.

In practice, researchers encounter far more complex problems with much more sophisticated hypothesis. Therefore, the application of multiple hypotheses testing in clinical trials becomes inevitable.

Application of Multiple Hypotheses Testing in Other Fields

Multiple hypotheses testing also arises in many other settings such as disease mapping (Green and Richardson (2002)), astronomical surveys (Miller et al. (2001)) and public health surveillance (Castro and Singer (2006)), etc.

1.1.2 Fundamentals of multiple hypotheses testing

In the case of a single hypothesis test, there are only two possible states of nature: the null hypothesis is either true or false. Correspondingly, we can either reject or fail to reject the null hypothesis. This leads to only four possible outcomes, making the inference process relatively straightforward. However, in the context of multiple hypotheses testing involving m hypotheses evaluated simultaneously, the situation becomes significantly more complex. There are 2^m possible configurations of the true states (each hypothesis being either true or false) and 2^m potential decision combinations (each hypothesis being either rejected or not). Let m_0 denote the number of true null hypotheses, and let R be the number of null hypotheses rejected. Table 1.1 ((Dudoit and Van Der Laan, 2008; Westfall, Tobias, and Wolfinger, 2011)) summarizes the different possible outcomes and their classification.

TABLE 1.1: Truth Table for Multiple Hypotheses Testing

Decision $\downarrow$ \ **Truth** $\rightarrow$	Null is true	Null is false	Total
Reject H_0	V	S	R
Accept H_0	$m_0 - V$	W	$m - R$
Total	m_0	m_1	m

Here, V denotes the number of false rejections, that is, the null hypotheses that are incorrectly rejected (also referred to as false discoveries). Similarly, W denotes the number of false acceptances, corresponding to false null hypotheses that are incorrectly accepted (also known as false non-discoveries). In analogy with single hypothesis testing, V and W correspond to Type I and Type II errors, respectively. When the tests are conducted based on a fixed sample size (i.e., when the sampling

space cannot be altered), it is generally not possible to minimize both types of errors simultaneously.

Now, let us formally describe the multiple hypotheses testing framework. Suppose, the index i denotes the i-th hypothesis, $i = 1, 2, ..., m$ and the set $\mathcal{H} = \{1, 2, 3, ...m\}$ denotes their collection. The truth for i is given by the truth index θ_i, i.e., $\theta_i = 1(0)$ if the i-th null is false (true). The collection of these indices is denoted by $\boldsymbol{\theta} = (\theta_1, \theta_2, ...\theta_m)$. Let X be the data matrix that is generated from some underlying distribution based on the truth vector $\boldsymbol{\theta}$. X is usually assumed to have m columns, (in which case, we may presume that the i-th column relates to test i). The number of rows of X relates to the number of sampling units on which the data has been collected. Further, let $S_1, S_2, ..., S_m$ be the vector of test statistics (depending on X), where S_i is used to test the individual test i. Note that S_i may be scalar or vector valued, depending on the tests.

The aim of a multiple hypotheses testing method is to provide a decision vector $\boldsymbol{\delta} = (\delta_1, \delta_2, ... , \delta_m)$, where $\delta_i = \delta_i(S_i)$ depends on the i-th test statistic and is the decision for i, i.e. $\delta_i = 1(0)$ implies that the i-th null hypothesis is rejected (accepted). Note that,

1. $V = \sum_{i=1}^{m}(1 - \theta_i)\delta_i$,
2. $W = \sum_{i=1}^{m} \theta_i(1 - \delta_i)$,
3. $R = \sum_{i=1}^{m} \delta_i$, and
4. $m - R = \sum_{i=1}^{m}(1 - \delta_i)$.

In other words, we can define the relevant Type I and Type II error measures in terms of $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$.

In the classical framework, we construct a decision vector $\boldsymbol{\delta}(X)$ such that some measure of Type I error $\gamma(V(\boldsymbol{\theta}, \boldsymbol{\delta}), R(\boldsymbol{\delta}))$ is controlled at a certain level α. An optimal multiple hypotheses testing method further requires that a certain measure of power be maximized among all methods satisfying the said criteria.

In a Bayesian framework, we may assume the truth vector $\boldsymbol{\theta}$ to be random and observations from some appropriate distribution (say Bernoulli). A loss function may be defined in terms of the errors. For example, the sum of individual losses for

each i may be considered.

$$L(\boldsymbol{\theta}, \boldsymbol{\delta}) = \sum_{i=1}^{m} (\Delta_1(1-\theta_i)\delta_i + \Delta_2(1-\delta_i)\theta_i). \tag{1.1}$$

In the Bayesian decision-theoretic framework, we aim to find the decision vector $\boldsymbol{\delta}$ that minimizes the expected loss, i.e., the Bayes risk $R(\boldsymbol{\theta}, \boldsymbol{\delta}) = \mathbb{E}\left(L(\boldsymbol{\theta}, \boldsymbol{\delta})\right)$. The rule that yields such a decision vector is known as the optimal rule. The corresponding Bayes risk is known as the optimal Bayes risk (say R_0).

1.1.3 Frequentist and Bayesian Approaches of Multiple Hypotheses Testing

We note that, in the truth table (Table 1.1), all entries except those in the rightmost column are stochastic in nature, that is, their values may vary depending on the outcome of the experiment. In contrast, the total number of hypotheses, denoted by m, is typically fixed and non-random. An exception arises in the context of online multiple hypotheses frameworks (see, e.g., Ramdas et al. (2017), Yang et al. (2017), Ramdas et al. (2018), Weinstein and Ramdas (2020), Tian and Ramdas (2021), Xu and Ramdas (2022), and Robertson, Wason, and Ramdas (2023)), where hypotheses arrive sequentially over time and m may grow dynamically. In this work, however, we focus on a fixed-horizon multiple hypotheses testing setting. When studying the asymptotic behavior of the proposed procedure in a large-scale regime, we consider increasing values of m under varying data-generating processes. That is, asymptotics are considered with respect to increasing problem size, not by fixing a data-generating process and varying m artificially.

In the frequentist framework of multiple hypotheses testing, the primary objective is to control a pre-specified measure of Type I error. When two procedures control the same error measure (e.g., the familywise error rate or the false discovery rate) at the same nominal level, they are typically compared based on their ability to increase statistical power or, equivalently, to reduce a measure of Type II error. The field of multiple hypotheses, originally developed within this classical framework, has given rise to a wide range of methodological advances. Notable contributions include those by Bonferroni (1936), Šidák (1967), Holm (1979), Simes (1986), Hommel

(1988), Hochberg (1988), Benjamini and Hochberg (1995), Benjamini and Yekutieli (2001), Benjamini and Yekutieli (2005), Sarkar (2002), Sarkar (2004), Sarkar (2006), Sarkar (2008a), Storey (2002), Storey (2003), and Genovese, Roeder, and Wasserman (2006), among many others.

The frequentist approach to addressing the multiplicity problem has certain limitations. One key challenge is determining which hypotheses should be included in the multiplicity adjustment. In some cases, secondary or exploratory tests may be inappropriately included, leading to overly conservative corrections and potentially misleading conclusions. For instance, as discussed in the last Section (see Example 3 regarding applications of multiple hypotheses in clinical trials), an incorrect application of multiplicity adjustments to a set of m endpoints may result in erroneous inference. A strong understanding of the relevant scientific context is often necessary to make appropriate decisions in such scenarios.

Berry and Hochberg (1999) argue that such logical difficulties can be naturally resolved within the Bayesian framework (Scott and Berger (2006)) , without requiring explicit multiplicity adjustments. In that work, the authors demonstrate how Bayesian hierarchical models can be used to estimate treatment effects while appropriately accounting for multiplicity through the model structure itself. A related frequentist perspective is illustrated by Sjölander and Vansteelandt (2019) using directed acyclic graphs (DAGs) to clarify the logical relationships among hypotheses.

Bayesian approaches to multiple hypotheses testing can be broadly categorized into fully Bayesian and empirical Bayesian methods. Fully Bayesian procedures incorporate prior probabilities specified a priori or learned from the model (e.g., Müller et al. (2004) and Do, Müller, and Tang (2005), Scott and Berger (2006), Scott and Berger (2010)), while empirical Bayesian methods estimate prior probabilities directly from the observed data (e.g., Efron (2004), Efron (2007), Sun and Cai (2007a), Sun and Cai (2009), and Sun et al. (2015)).

1.1.4 Measures of Type I and Type II Errors

From a frequentist perspective, the central objective in multiple hypotheses testing is typically to control a specific measure of Type I error. A broad class of such error

measures can be expressed as a bounded function $\gamma(V, R)$, where V denotes the number of false rejections (Type I errors), and R is the total number of rejections.

One of the most commonly used error measures in this context is the *Familywise Error Rate* (FWER). To emphasize its association with Type I errors, it is sometimes referred to as *FWER-I*, and is formally defined as the probability of making at least one false rejection:

$$\text{FWER-I} = \mathbb{P}(V > 0).$$

Several classical procedures have been proposed to control the FWER-I, including those by Bonferroni (1936), Šidák (1967), Simes (1986), Hommel (1988), and Holm (1979).

When the number of hypotheses is large, controlling the Familywise Error Rate (FWER-I) becomes increasingly difficult, as the probability of making at least one false rejection tends to grow with m. In such settings, it is often desirable to adopt more lenient error measures that allow for a small number of false discoveries. One such measure is the *Generalized Familywise Error Rate of Type I* (gFWER-I), which is defined as the probability of making more than k false rejections, where $k \in \mathbb{N}_0$ is a user-specified tolerance level. In other words, we allow up to k false positives among the rejected hypotheses. This measure is denoted by gFWER-I(k), and is mathematically defined as:

$$\text{gFWER-I}(k) = \mathbb{P}(V > k).$$

Multiple hypotheses procedures for controlling gFWER-I have been proposed by Guo and Romano (2007) and Romano and Wolf (2007).

One way to address the challenge posed by a large number of hypotheses is to consider the proportion of false rejections among all rejections. This quantity is referred to as the *False Discovery Proportion* (FDP) defined as

$$\text{FDP} = \frac{V}{R \vee 1},$$

where V is the number of false discoveries, R is the total number of rejections, and $R \vee 1 := \max\{R, 1\}$. By convention, FDP $= 0$ when $R = 0$. The *False Discovery Rate*

(FDR), introduced by Benjamini and Hochberg (1995), is the expected value of the FDP:

$$\text{FDR} = \mathbb{E}(\text{FDP}).$$

A wide variety of multiple hypotheses testing procedures have been developed to control the FDR. Notable contributions include Benjamini and Hochberg (1995), Benjamini and Yekutieli (2001), Benjamini and Yekutieli (2005), and Storey (2002), and a comprehensive overview can be found in He, Gang, and Fu (2024). A related measure is the *marginal False Discovery Rate* (mFDR), defined as the ratio of the expected number of false rejections to the expected number of total rejections (Sun and Cai, 2007b):

$$\text{mFDR} = \frac{\mathbb{E}(V)}{\mathbb{E}(R)}.$$

However, controlling the FDR alone does not guarantee consistent control over the realized FDP in every experiment. To enhance interpretability and offer stronger guarantees, one may instead control the probability that the FDP exceeds a user-defined threshold β. This leads to the notion of *False Discovery Exceedance* (FDE) (Sun et al., 2015; Genovese and Wasserman, 2006), defined as

$$\text{FDE} = \mathbb{P}(\text{FDP} > \beta).$$

Controlling FDE at level α ensures that, on average, in no more than $100\alpha\%$ of repeated experiments under the same data-generating mechanism, the FDP exceeds the threshold β. This stricter form of control has several appealing properties and has been discussed extensively in the literature (see Genovese and Wasserman (2006), Genovese, Roeder, and Wasserman (2006), and Sun et al. (2015)) . Additional false discovery metrics and their theoretical properties are explored in Chapter 3 of Dudoit and Van Der Laan (2008).

For comparing two methods that control a certain measure of Type I error at a certain level, a notion of Type II error is essential. Following the definitions of measures of false positives, measures of Type II error or false negatives can be defined, i.e., it is of the form $\psi(W, m - R)$ and should be bounded. Some measures that have predominantly been used in the literature are

1. **FWER-II**(Familywise Error Rate Type II) = $\mathbb{P}(W > 0)$.
2. **gFWER-II(*k*)** Generalized Familywise Error Rate Type II) = $\mathbb{P}(W > k)$.
3. **FNR** (False Nondiscovery Rate) = $\mathbb{E}\left(\text{FNP}\right)$. (FNP is the false nondiscovery proportions, defined as FNP=$\frac{W}{\max(m-R,1)}$).
4. **mFNR** (Marginal False Nondiscovery Rate) =$\frac{\mathbb{E}(W)}{\mathbb{E}(m-R))}$.

In the context of multiple hypotheses, several measures have been proposed to quantify the ability of a procedure to detect false positives. These measures serve as analogues to statistical power in the single hypothesis testing framework. Broadly, power in this context can be categorized into two types: *disjunctive power*, which refers to the probability of rejecting at least one false null hypothesis, and *conjunctive power*, which pertains to the probability of correctly rejecting all false null hypotheses. An extensive discussion on various notions of power in the context of multiple hypotheses testing can be found in Chapter 18 of Westfall, Tobias, and Wolfinger (2011) and Chapter 1 of Dudoit and Van Der Laan (2008) .

1.2 Some Notable Multiple Hypotheses Testing Procedures

Let us now discuss several multiple hypotheses testing procedures that are widely used in practice. A central component in these procedures is the use of *p-values*, which serve as indicators of the plausibility of null hypotheses. In general, smaller p-values suggest stronger evidence against the null hypothesis. Consequently, hypotheses associated with the smallest p-values are typically rejected, while those with larger p-values are retained.

To operationalize this idea, one generally orders the p-values and defines a cutoff to determine which hypotheses to reject. multiple hypotheses testing procedures can be broadly classified based on how this cutoff is determined. One such classification is based on whether the procedure uses a single-step or a stepwise approach.

Single-step procedures. In single-step procedures, each hypothesis is tested independently of the others. The decision rule compares each p-value against a fixed

threshold c, and the i-th hypothesis is rejected if $p_i < c$, i.e.,

$$\delta_i = \mathbb{1}(p_i < c). \tag{1.2}$$

Here, p_i denotes the p-value corresponding to the i-th hypothesis. Two classical examples of single-step procedures are:

- **Bonferroni correction** (Bonferroni (1936)): uses $c = \alpha/m$;
- **Šidák's correction** (Šidák (1967)): uses $c = 1 - (1-\alpha)^{1/m}$.

Stepwise procedures. In contrast to single-step methods, stepwise procedures take into account the ordering of p-values. Let $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$ be the ordered p-values, and let $c_1 < c_2 < \cdots < c_m$ be a sequence of increasing cutoffs. Stepwise methods can be classified further into:

- **Step-up procedures:** These identify the largest index k such that $p_{(k)} \leq c_k$, and reject all hypotheses corresponding to $p_{(1)}, \ldots, p_{(k)}$. Formally,

  $$k = \max \left\{ i \in \{1, \ldots, m\} \mid p_{(i)} \leq c_i \right\}.$$

  An example is the Benjamini-Hochberg procedure (Benjamini and Hochberg (1995)), where $c_i = \frac{i\alpha}{m}$.

- **Step-down procedures:** These identify the smallest index k' such that $p_{(k')} > c_{k'}$, and reject all hypotheses corresponding to $p_{(1)}, \ldots, p_{(k'-1)}$. Formally,

  $$k' = \min \left\{ i \in \{1, \ldots, m\} \mid p_{(i)} > c_i \right\}.$$

  A prominent example is Holm's procedure (Holm (1979)), which uses $c_i = \alpha/(m-i+1)$.

Below, we briefly describe a few key procedures that are foundational in the literature and merit further attention:

1. **Bonferroni's Procedure:** One of the earliest multiple hypotheses testing methods, Bonferroni's correction (Bonferroni (1936)) controls the familywise error

rate (FWER) under the global null hypothesis. It is a single-step procedure that rejects H_{0i} if

$$p_i < \frac{\alpha}{m}, \quad \text{that is, } \delta_i = \mathbb{1}\left(p_i < \frac{\alpha}{m}\right).$$

This guarantees control of the FWER at level α when all null hypotheses are true.

2. **Šidák's Procedure:** This is another single-step procedure (Šidák (1967)), based on Šidák's inequality, which also controls the FWER under the global null. The decision rule is:

$$\delta_i = \mathbb{1}\left(p_i < 1 - (1-\alpha)^{1/m}\right).$$

3. **Benjamini–Hochberg (BH) Procedure:** Introduced by Benjamini and Hochberg (1995), this widely used step-up method aims to control the false discovery rate (FDR). It uses the cutoff values $c_i = \frac{i\alpha}{m}$ and is shown to control the FDR at level approximately α (or $\alpha m_0/m$ if m_0 nulls are true), even under certain positive dependence structures among the test statistics.

1.2.1 Large-scale Multiple Hypotheses Testing

Traditional p-value-based multiple hypotheses testing procedures are conservative (Dudoit and Van Der Laan (2008) and Westfall, Tobias, and Wolfinger (2011)), i.e., the attained value of FWER (or FDR) is much less than the intended control α. When the number of hypotheses is large (e.g., the identification of differential gene expressions pertaining to a certain disease among a large number of genes), this problem of conservative attainment of the error rates becomes more prominent. Therefore, traditional multiple tests result in smaller power. In this section, we shall talk about some multiple hypotheses testing procedures that are applicable even in such large-scale scenarios. Comprehensive reviews of large-scale multiple hypotheses testing can be found in Efron (2012).

Note that the control of FWER discourages the commitment of even a single false rejection up to the level of control. Therefore, traditional FWER controlling methods are naturally conservative. Instead, for large-scale problems, we concentrate on methods that control certain appropriate error rates, such as gFWER-I(k), FDR, and

mFDR, among others. If m is indefinitely large, a gFWER-I(k) controlling method with finite k may also be conservative. On the contrary, FDP is the ratio of V and R, and therefore, its expected value is more adaptive to changing m, in that, a proper FDR controlling procedure allows roughly a proportionate amount of false positives. This property makes such procedures considerably less conservative (Storey (2002) and Sun and Cai (2007b)). such procedures are well suited to large-scale testing problems. Naturally, another similar error rate that is applicable for large-scale treatment is the marginal FDR (mFDR).

As already mentioned, the famous BH procedure controls the FDR at level $\alpha m_0/m$. This is still conservative, especially in the case of sparse alternatives. There exist adaptive BH procedures that estimate the null proportions and thereby control the FDR exactly at the intended level (Benjamini and Hochberg (2000), Benjamini, Krieger, and Yekutieli (2006), Blanchard and Roquain (2009), Liu and Sarkar (2011), Sarkar (2008b), Sarkar, Guo, and Finner (2012), He and Sarkar (2013), and Solari and Goeman (2017)).

In Efron (2012), an empirical Bayes approach was developed to deal with the large-scale multiplicity problem occurring in experiments involving microarray data (Efron (2004) and Efron (2007)). The Oracle test statistic for testing H_{0i}, denoted as t_i is defined as the posterior probability of H_{0i} being true given the observed data. Clearly, if the value of t_i is low, we have more evidence against H_{0i}. In that case, we have an inclination to reject it. In Efron (2007), a crude testing rule has been suggested: $\delta_i = \mathbb{1}(t_i < l)$, where l is a fixed number (e.g., 0.1). Unfortunately, this quantity may involve some unknown parameters. Under a two-group mixture model assumption, Efron (2012) provides some resampling-based techniques to estimate Lfdrs. This approach paves the path for large-scale multiple hypotheses testing.

Sun and Cai (2007a), considered a two-group mixture model for the test statistics, i.e., $X_i|\theta_i \sim f = (1-\theta_i)f_0 + \theta_i f_1$ such that, conditioned on the truth vector, any pair of test statistics, (X_i, X_j), with $i, j \in [m]$ and $i \neq j$ are independent and $\theta_i \overset{iid}{\sim} Bernoulli(1-\pi_0)$. Under such assumptions, the authors showed that, corresponding to any decision rule of the form $\delta_i = \mathbb{1}(t_i < c)$, there exists a $\lambda \in \mathbb{R}$ such that the risk $R(\lambda) = \mathbb{E}(V) + \lambda \mathbb{E}(W)$ is minimized by the decision rules, $\delta_i, i \in [m]$. Consequently, the rules of the form $\delta_i = \mathbb{1}(t_i < c)$ are optimal. Further, there is a

one-to-one correspondence between λ and the attained FDR level of the given decision rule δ. Therefore, given any $\alpha \in (0,1)$, we can get an optimal oracle rule that controls the FDR at an exact level α. Now, under the two-group mixture model assumption, it can be shown that $t_i = \frac{\pi_0 f_0(x_i)}{f(x_i)}$, where x_i is the observed value of X_i. With existing methods of estimating t_i, the authors proposed an empirical Bayes method that replaces t_i with its estimate in δ_i. This method is asymptotically optimal in that, as m grows to ∞, the attained mFNR $= \beta + o_m(1)$. Here β is the attained mFNR of the optimal oracle rule. Understandably, this is well suited for large-scale problems. This idea was generalized under various dependent setups like hidden markov model (Sun and Cai (2009)), grouped mixture model (Cai and Sun (2009)), and spatial dependence (Sun et al. (2015)), etc.

It is established in experimental setups that genes that are significantly expressed against some phenotypes (e.g., prostate cancer) are sparse. In other words, very few genes are associated with a certain disease. Therefore, accommodation of sparsity in a large-scale study is important (Donoho and Jin (2004) and Jin (2009)). Specifically, we are interested in the behavior of optimal tests under sparsity and whether some popular multiple hypotheses testing methods show similar behavior under appropriate sparsity assumptions. Extensive simulation studies in Bogdan et al. (2007) and Bogdan, Ghosh, and Tokdar (2008) showed that the performance of the Benjamini-Hochberg method (Benjamini and Hochberg (1995)) and some empirical Bayes methods are equivalent under sparsity when the test statistics follow independent Gaussian distributions. Bogdan et al. (2011) considered a hierarchical two-group mixture model $X_i|\theta_i \sim (1-\theta_i)N(0,\sigma^2) + \theta_i N(0,\sigma^2+\tau^2)$, $\theta_i \overset{iid}{\sim} Bernoulli(p_m)$, $i \in [m]$. This model takes into account the differential effects of significant genes and polygenes on a certain disease. Two distinct sparsity conditions are defined: moderately sparse (for example, when $p_m \propto m^{-0.5}$) and extremely sparse (for example, when $p_m \propto m^{-1}$). Under these sparsity conditions, the optimal rules were obtained. Further, as m grows to ∞, conditions for a multiple hypotheses testing procedure to be asymptotically optimal under such sparsity conditions are derived. Tests that satisfy these conditions are termed "asymptotically Bayes optimal under sparsity" (ABOS). Later, it was proved that the Bonferroni's method is ABOS under extreme sparsity and the Benjamini-Hochberg's method is ABOS under both sparse

conditions. Similar results were obtained in Frommlet and Bogdan (2013). Several extensions of Bogdan et al. (2011) have been studied in the literature. The Subbotin family of distributions for test statistics was shown to yield similar results (Neuvial and Roquain (2012)). Using horseshoe prior, a full Bayes analysis was performed in Datta and Ghosh (2013).

1.3 Sequential Multiple Hypotheses Testing

The multiple hypotheses testing procedures discussed thus far apply to scenarios where the full dataset is available at the time of testing. Such procedures are referred to as multiple tests with *fixed-sample-size*. In contrast, many data-generating processes are inherently sequential in nature. For these *sequentially observed* or *streaming* datasets, multiple hypotheses must be tested simultaneously with the objective of stopping sampling and making decisions as quickly as possible, while controlling specified error rates (Wald (1992), Siegmund (2013), and Lai (2001)).

A prominent application of sequential multiple hypotheses testing arises in clinical trials (Pocock (2013)) with multiple endpoints or treatment arms, where patients are recruited sequentially or in groups (see, e.g., Jennison and Turnbull (2000)). At each interim analysis of such trials, one must decide whether to accept or reject a hypothesis regarding a clinical endpoint, or to collect additional data before making a decision. Other contexts where sequential or streaming data naturally occur include multichannel online (or quickest) changepoint detection (Tartakovsky, Li, and Yaralov (2003), Mei (2010)), acceptance sampling with multiple criteria (Baillie (1987)), agricultural experiments (Clements et al. (2014)), and financial market analysis (Lai and Xing (2008)). In such sequential settings, traditional fixed-sample-size multiple hypotheses testing methods, unless specifically adapted to sequential data, cannot be directly applied.

As there are sequential tests of a single hypothesis (e.g., Wald's SPRT) that can control both probabilities of Type I and Type II errors at some specified levels, a number of articles have been devoted to the development of tests that can control compound error rates of both kinds in the context of sequential multiple hypotheses testing. For instance, sequential tests for controlling the FWER-I and FWER-II or the

gFWER-I(k) and gFWER-II(l) error rates are developed in De and Baron (2012a), De and Baron (2012b), De and Baron (2015), Bartroff and Song (2014), Bartroff (2018), Song and Fellouris (2017), and Song and Fellouris (2019) and Chen et al. (2023). Sequential tests for controlling FDR and FNR, or some of their variants (e.g., pFDR, pFNR, false discovery exceedance, and nondiscovery exceedance), are developed in Bartroff and Song (2020) and He and Bartroff (2021) and Bartroff (2018).

In the literature on sequential multiple hypotheses testing, two distinct sampling schemes are commonly studied, each arising from different practical constraints and application settings. We describe these schemes in detail below.

1.3.1 Asynchronous Termination of Data Streams

In the first scheme, m data streams are observed from m independent sequential experiments, and sampling from different streams may be terminated at different time points. Such schemes, studied in Bartroff and Song (2014), Malloy and Nowak (2014), Bartroff (2018), and Bartroff and Song (2020), are suitable when the costs of sampling differ substantially across streams. In these situations, terminating more expensive experiments earlier than cheaper ones can yield substantial savings in overall sampling cost.

A motivating example is a clinical trial in which certain endpoints require costly or physically demanding medical tests, such as MRI scans, angiograms, or colonoscopies. Here, the priority is to minimize the number of such tests rather than the number of patients on which they are performed. If sufficient evidence is available to accept or reject a hypothesis for a particular endpoint, sampling for that endpoint is terminated, and no further observations are collected for it. Data collection continues only for hypotheses for which no decision has yet been reached.

1.3.2 Simultaneous Termination of Data Streams

In the second scheme, sampling from all data streams is terminated at the same time. Here, it is assumed that m-dimensional vectors are observed sequentially, each vector representing measurements from a single sampling unit (e.g., a patient), with the

m coordinates corresponding to the m data streams under investigation. Once a pre-specified stopping criterion is met, sampling of these vectors is stopped, leading to simultaneous termination of all data streams. This scheme, studied in De and Baron (2012a), De and Baron (2012b), De and Baron (2015), Song and Fellouris (2017), Song and Fellouris (2019), and He and Bartroff (2021) and Chen et al. (2023), is appropriate when the cost of obtaining each sampling unit is substantially higher than the cost of observing all m measurements from that unit.

For example, in studies of differential gene expression, the expression levels of all genes are obtained simultaneously from a single patient sample. If patient samples are collected sequentially or in groups, it is more efficient to discontinue sample collection from new patients entirely, thereby terminating observation of all gene expression values at once.

1.4 Motivation and Overview of This Thesis

This thesis is motivated by certain gaps in the existing literature and is organized into two distinct parts. The first part focuses on fixed-sample *asymptotically Bayes optimal under sparsity* (ABOS) tests, while the second part addresses sequential multiple hypotheses testing in large-scale settings.

In the first part, we investigate fixed-sample ABOS procedures. The existing literature on ABOS has largely focused on independent test statistics. For example, Bogdan et al. (2011) proposed a hierarchical model motivated by the identification of differentially expressed genes. However, dependent settings remain largely unexplored. In this work, we consider an m-variate test statistic jointly distributed as a multivariate normal vector with equal pairwise correlations, and we establish sufficient conditions under which a multiple hypotheses testing rule attains the ABOS property in such a dependent framework.

The second part of the thesis turns to sequential multiple hypotheses testing in large-scale scenarios. Existing sequential procedures are often conservative and require large stopping times to reach decisions. Assuming an indefinitely large number of hypotheses, we develop testing procedures tailored to two sampling schemes:

(i) simultaneous termination of all data streams, and (ii) asynchronous termination at different times.

Although the two parts address very different settings, they share a common theme: understanding the asymptotic behavior of multiple hypotheses testing procedures as the number of hypotheses grows without bound.

1.4.1 Asymptotically Bayes Optimality under Sparsity in a Dependent Framework

In Bogdan et al. (2011), the concept of asymptotically Bayes optimal tests under sparsity is introduced for independent and identically distributed (i.i.d.) normal test statistics. However, in real-world scenarios, independence is rarely guaranteed. To address this limitation, we extend the framework by incorporating dependence among test statistics. Given the complexity introduced by general dependence structures, we restrict our analysis to exchangeable test statistics.

This restriction is natural because exchangeability retains some of the desirable properties of the i.i.d. case. In particular, it allows the use of a common threshold rule for each hypothesis, a feature exploited by Bogdan et al. (2011).

For a multivariate normal vector to be exchangeable, the marginal means and variances must be identical, and the correlation coefficients between any pair of variables must be equal, denoted by ρ. When $\rho = 0$, the test statistics are i.i.d. by virtue of normality.

1.4.2 Large-scale Sequential Multiple Hypotheses Testing

Consider the case-control study presented in Singh et al. (2002), where $m = 6033$ two-sample tests were conducted. If patients had been recruited sequentially, a sampling scheme with simultaneous termination would have been more appropriate. This is because each newly recruited patient would contribute gene expression data for all 6033 genes at each time point.

At an interim stage, if further sampling is pursued to resolve undecided hypotheses, the gene expression levels for hypotheses already resolved (or close to resolution) would still be available at little additional cost. Therefore, prematurely discarding some hypotheses, as permitted in certain sampling schemes, might violate the sufficiency principle discussed in De and Baron (2012a).

In this setting, we require a sequential multiple hypotheses testing procedure that remains effective even when the number of hypotheses is large. Much of the existing literature on sequential multiple hypotheses testing focuses on conservative procedures, where the attained false positive and false negative rates are substantially lower than their nominal control levels. This conservatism increases the expected sample size needed to reach decisions, a drawback that becomes more pronounced as the number of hypotheses grows. A procedure whose attained error rates closely match the predetermined control levels is therefore highly desirable, as it can reduce the expected sample size and, consequently, the overall cost. In clinical trials with asynchronous sampling, reducing the number of experiments directly lowers costs. For both asynchronous and simultaneous schemes, it is therefore crucial to design testing procedures that remain effective in large-scale settings with many hypotheses.

The workflow of the thesis is as follows. In Chapter 2, we introduce the framework of *asymptotic Bayes optimality under sparsity* (ABOS) and derive sufficient conditions for a testing rule to be ABOS when the vector of test statistics follows a multivariate normal distribution with equal positive pairwise correlations. We then examine whether several well-known multiple hypotheses testing procedures, including those of Benjamini and Hochberg (1995), Bonferroni (1936), and Šidák (1967), satisfy these conditions.

Chapter 3 focuses on the simultaneously terminating sampling scheme in sequential multiple hypotheses testing. We introduce a large-scale oracle test procedure (the OI test) and establish its key properties, including finite stopping time and control of the false discovery rate (FDR) and false nondiscovery rate (FNR), along with their marginal counterparts (mFDR and mFNR). A central result, Theorem 10, demonstrates that the stopping time remains finite even as the number of hypotheses tends to infinity. We also propose a data-driven version of the OI test, showing

that it retains similar guarantees—finite stopping time and asymptotic control of FDR and FNR at the target levels.

Chapter 4 examines the asynchronous termination scheme, in which sampling from different data streams may conclude at different times. We propose both oracle (OS) and data-driven (DS) tests, and for the oracle version, we establish that the maximum stopping time is finite and that both the FDR and FNR are controlled at their respective target levels.

Chapter 5 reports numerical studies that substantiate the theoretical results from Chapters 3 and 4, and presents an application of the DI test, introduced in Chapter 3, to real datasets.

Finally, Chapter 6 offers concluding remarks and outlines potential directions for future research.

Chapter 2

Asymptotic Bayes Optimality under Sparsity for Exchangeable Dependent Multivariate Normal Test Statistics

2.1 Introduction

Let $\mathbf{X} = (X_1, X_2, \cdots, X_m)$ be the vector of test statistics. A specific coordinate of $\mathbf{X}$ is denoted by i, $i \in [m]$. Conditioned on the effects vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_m)$, $\mathbf{X}$ follows an equicorrelated multivariate normal distribution with mean $\boldsymbol{\mu}$ and the covariance matrix $\sigma_\epsilon^2 \Sigma_\rho$, where $\sigma_\epsilon^2 > 0$ is the error variance and Σ_ρ is the correlation matrix whose off-diagonal terms equal to $\rho > 0$; i.e., the correlation coefficient between each pair (X_i, X_j) is ρ, where $i, j \in [m]$ and, $i \neq j$. Hence,

$$\Sigma_\rho = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}.$$

We assume that the components of the effects vector $\boldsymbol{\mu}$ are mutually independent and normally distributed with mean zero. However, the variances are not equal for

all of the coordinates. Let $\mathcal{H}_A \subseteq [m]$ denote the set of coordinates having a non-negligible influence on the outcome of interest. This subset is referred to as the set of alternatives or positives. Consequently, if i belongs to $\mathcal{H}_A$, then we say that i-th alternative is true or H_{1i} is true. Such an index i corresponds to an alternative hypothesis. On the other hand, the coordinates of the complement of the set of $\mathcal{H}_A$ have little effect on the outcome of interest, and jointly they form the set of nulls denoted by $\mathcal{H}_0$. If i belongs to $\mathcal{H}_0$, the i-th null hypothesis or H_{0i} is said to be true. Such an index corresponds to a null hypothesis. The unobservable identity vector of the positive coordinates, introduced in Chapter 1 is defined as $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_m)$ with

$$\theta_i = \begin{cases} 1 & \text{if } i \in \mathcal{H}_A \\ 0 & \text{if } i \in \mathcal{H}_0 \end{cases}, \text{ for } i \in [m].$$

In other words, $\boldsymbol{\theta}$ determines the sets $\mathcal{H}_A$ and $\mathcal{H}_0$. We assume that for all $i \in [m]$, $\theta_i \overset{i.i.d.}{\sim}$ Bernoulli (p), where $p \in (0, 1)$.

The variance of μ_i, provided H_{0i} is true, is σ_0^2. Allowing a non-zero variance under the null accommodates the possibility of weak background effects that do not constitute meaningful signals. For example, in the context of gene expression analysis, expression levels of genes exhibiting negligible differential expression may still vary across samples. These genes are known as polygenes. On the contrary, if H_{1i} is true, the variance is assumed to be $\sigma_0^2 + \tau^2$. Here, the term τ^2 is the additional variance due to signals. This reflects the substantially higher effects of the positive genes on the phenotype mentioned above. The distribution of the vector of test statistics $\mathbf{X}$ is hierarchical:

$$\mathbf{X}|\boldsymbol{\mu} \sim \text{MVN}(\boldsymbol{\mu}, \sigma_\epsilon^2 \Sigma_\rho) \tag{2.1}$$

$$\mu_i|\theta_i \overset{\text{independent}}{\sim} (1-\theta_i)\text{N}(0, \sigma_0^2) + \theta_i \text{N}(0, \sigma_0^2 + \tau^2) \text{ for all } i \in [m] \tag{2.2}$$

$$\theta_i \overset{\text{i.i.d}}{\sim} \text{Bernoulli}(p) \text{ for all } i \in [m]. \tag{2.3}$$

Note that, for $i, j \in [m]$, and $i \neq j$, μ_i is independent of θ_j.

Our objective is to infer unobserved indicators $\{\theta_i | i \in [m]\}$ based on the observed $\mathbf{X}$. In doing so, two kinds of errors can be committed, as shown in Table 1.1.

When a null hypothesis is classified as an alternative, we make a Type I error, or a false rejection. The classification of an alternative hypothesis as a null is known as a Type II error or a false acceptance. The decision vector is $\boldsymbol{\delta} = (\delta_1, \delta_2, \cdots, \delta_m)$. Here δ_i is equal to 1(0) if we classify the i-th coordinate to be alternative (null). Note that $\boldsymbol{\delta}$ depends on the test statistics $\mathbf{X}$ and therefore is random. The number of false rejections is $V = \sum_{i=1}^{m}(1-\theta_i)\delta_i$, and the number of false acceptances is $W = \sum_{i=1}^{m}(1-\delta_i)\theta_i$.

Our goal is to minimize the overall classification error. For this, we introduce a loss function for each coordinate. The loss function assigns the loss $\Delta_0 > 0$ ($\Delta_A > 0$) to a single Type I (Type II) error. The loss of a correct decision is zero. The overall loss is then computed by summing up all the losses. The Bayes risk corresponding to this particular loss is the expected value of the overall loss. The risk is denoted by R. Therefore,

$$\begin{aligned} R =& \sum_{i=1}^{m} E(\Delta_0(1-\theta_i)\delta_i + \Delta_A\theta_i(1-\delta_i)) \\ =& \Delta_0 E(V) + \Delta_A E(W). \end{aligned} \tag{2.4}$$

The collection of decision rules that minimizes the coordinatewise risk functions with the same loss function constitutes the Bayes optimal rules minimizing R. The Bayes optimal rule for i is

$$\delta_{0i}(\mathbf{X}) = \mathbb{1}\left(\frac{f(\mathbf{X}|\theta_i = 1)}{f(\mathbf{X} \mid \theta_i = 0)} > \Delta f\right), \tag{2.5}$$

where $f = \frac{(1-p)}{p}$, $\Delta = \frac{\Delta_0}{\Delta_A}$ and for $j \in \{0,1\}$, $f(\mathbf{X} \mid \theta_i = j)$ is the conditional distribution of $\mathbf{X}$ conditioned on $\theta_i = j$. We get the optimal Bayes risk, denoted by $R_0(\mathbf{X})$, by replacing δ_i with $\delta_{0i}(\mathbf{X})$ in 2.4, i.e.,

$$R_0(\mathbf{X}) = \sum_{i=1}^{m} E(\Delta_0(1-\theta_i)\delta_{0i}(\mathbf{X}) + \Delta_A\theta_i(1-\delta_{0i}(\mathbf{X}))).$$

It is evident that the set of rules $\boldsymbol{\delta}_0(\mathbf{X}) = (\delta_{01}(\mathbf{X}), \delta_{02}(\mathbf{X}), \cdots, \delta_{0m}(\mathbf{X}))$ is not a collection of simple rules in closed form, as $f(\mathbf{X} \mid \theta_i = j)$ has a complicated expression,

namely, for $j \in \{0,1\}$,

$$f(\mathbf{X} \mid \theta_i = j) = \sum_{\mathbf{j_i} \in \{0,1\}^{m-1}} p^{\mathbf{j_i}'\mathbf{1}}(1-p)^{(m-1-\mathbf{j_i}'\mathbf{1})}\text{MVN}(\mathbf{0}, \mathbf{V_{j_i}}),$$

where $\mathbf{j_i} = (j_1, j_2, \cdots, j_{(i-1)}, j_{(i+1)}, \cdots, j_m) \in \{0,1\}^{(m-1)}$ is a particular configuration of the rest of the coordinates of the indicator vector $\boldsymbol{\theta}$, barring i, i.e., $\boldsymbol{\theta} = (j_1, j_2, \cdots, j_{(i-1)}, j, j_{(i+1)}, \cdots, j_m)$. Further, $\mathbf{V_{j_i}}$ is the $m \times m$ unconditional covariance matrix of $\mathbf{X}$ given the particular configuration of $\boldsymbol{\theta}$. Its diagonal elements are the total marginal variances $\text{Var}(X_k \mid \theta_k) = \sigma_0^2 + \sigma_\epsilon^2 + \theta_k\tau^2$. The off-diagonal elements are covariance $\rho\sigma_\epsilon^2$ for all pairs $(k,l), k \neq l$. Here, $\text{MVN}(\mathbf{0}, \mathbf{V_{j_i}})$ denotes the multivariate normal density with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{V_{j_i}}$.

Instead of striving for a set of exact optimal tests, which, as discussed above, is analytically intractable, we are interested in finding tests that are easy to implement and equivalent to the set of Bayes optimal tests ($\boldsymbol{\delta}_0$) as m grows to infinity. We adopt a suitable framework for the parameters inspired by Bogdan et al. (2011) to facilitate an asymptotic investigation. Define $\boldsymbol{\gamma} = (p, \tau, \sigma_\rho^2, \Delta_A, \Delta_0)$ to be the vector of parameters where $\sigma_\rho^2 = \sigma_0^2 + (1-\rho)\sigma_\epsilon^2$. We consider the sequence $\{\boldsymbol{\gamma}_m\}$ where $\boldsymbol{\gamma}_m = (p_m, \tau_m, \sigma_{\rho m}^2, \Delta_{Am}, \Delta_{0m})$ is the value of $\boldsymbol{\gamma}$ when the number of hypotheses is m. i.e., the m in the suffix of each of the parameters denotes the dependence of the parameters on the number of hypotheses. Henceforth, we will use these notations. Here, we consider a sparse setup, i.e., $p_m \to 0$ as m grows to infinity. These assumptions imply that only a small proportion of hypotheses correspond to true signals. Such sparsity assumptions are consistent with many high-dimensional applications of multiple testing, including genomics, signal detection, and astronomy. There will be some additional assumptions about the asymptotic framework as we proceed with our discussion. Therefore, for any set of decision rules $\boldsymbol{\delta} = (\delta_1, \delta_2, ..., \delta_m)$, the Bayes risk R_m becomes

$$R_m = \sum_{i=1}^{m} E(\Delta_{0m}(1-\theta_i)\delta_i + \Delta_{Am}\theta_i(1-\delta_i)), \tag{2.6}$$

The optimal Bayes risk R_{0m} is obtained by minimizing R_m over all the $\mathbf{X}$-measurable decision rules, which is given by

$$R_{0m} = \sum_{i=1}^{m} \left[E(\Delta_{0m}(1-\theta_i)\delta_{0i}(\mathbf{X}) + \Delta_{Am}\theta_i(1-\delta_{0i}(\mathbf{X}))) \right]. \tag{2.7}$$

Here $\delta_{0i}(\mathbf{X}) = \mathbb{1}\left(\frac{f(\mathbf{X}|\theta_i=1)}{f(\mathbf{X}|\theta_i=0)} > \Delta_m f_m\right)$ denotes the optimal Bayes rule.

A sequence of testing procedures δ with risk R_m is said to be Asymptotic Bayes Optimal under Sparsity or ABOS (Bogdan et al., 2011) if,

$$\lim_{m\uparrow\infty} \frac{R_m}{R_{0m}(\mathbf{X})} = 1. \tag{2.8}$$

Our aim is to construct an ABOS procedure that is computationally feasible and straightforward to implement. The remainder of the chapter is organized in the following way. In section 2.2, we introduce a decomposition of $\mathbf{X}$. Using this decomposition, in section 2.3, we present a set of hypothetical rules that is ABOS. These rules are hypothetical in the sense that they are based on random variables that are unobservable. In section 2.4, we propose a set of easy-to-implement approximate rules based on the test statistics $\mathbf{X}$ and show that it is ABOS. Sufficient conditions for some popular multiple testing rules to be ABOS are presented in Section 2.5. Sections 2.6 and 2.7 include the proofs of the main theorems and lemmas of this chapter.

2.2 A Decomposition of the Observations

The following lemma proposes a decomposition for each of the coordinates of $\mathbf{X}$.

Lemma 1. *Suppose* $\mathbf{X}$ *follows the hierarchical model specified in* (2.1)–(2.3). *Then there exist a latent random variable* L *and a random vector* $\mathbf{Z} = (Z_1, Z_2, \cdots, Z_m)$ *such that, for all* $i \in [m]$,

$$X_i = Z_i + L, \tag{2.9}$$

where given $\boldsymbol{\theta}$, $(Z_1, Z_2, \ldots, Z_m)$ *are mutually independent,*

$$Z_i \mid \theta_i \sim (1-\theta_i)N(0, \sigma_\rho^2) + \theta_i N(0, \sigma_\rho^2 + \tau^2) \text{ for all } i \in [m], \text{ and,}$$
$$L \sim N(0, \rho\sigma_\epsilon^2).$$

Moreover, L is independent of Z_i for every $i \in [m]$.

Since the variables $\{Z_i : i \in [m]\}$ are independent and identically distributed with $\mathbb{E}(Z_i) = 0$, the Strong Law of Large Numbers (SLLN) implies that

$$\bar{Z}_m = \frac{1}{m}\sum_{i=1}^{m} Z_i$$

converges almost surely to 0 as $m \to \infty$. If we define the sample mean of our observations as $\bar{X}_m = \sum_{i=1}^{m} X_i/m$, it follows directly from (2.9) that

$$\bar{X}_m - L \to 0, \text{ almost surely.}$$

Consequently,

$$L = \lim_{m\to\infty} \bar{X}_m \quad \text{a.s.}$$

Hence L is measurable with respect to the σ-field generated by the infinite sequence $\mathbf{X}_\infty = (X_1, X_2, \ldots)$. In particular, if the entire sequence $\mathbf{X}_\infty$ were observable, the latent factor L is identifiable with respect to $\mathbf{X}_\infty$. For finite m, the sample mean $\bar{X}_m$ serves as a strongly consistent estimator of L.

As discussed earlier, our goal is to find the set of decision rules δ that minimizes the Bayes risk (2.4). When we observe only $\mathbf{X}$, the optimal decision rules are of the form (2.5). If we consider that the latent factor L is also observable alongside $\mathbf{X}$ and we aim to find the set of decision rules that minimizes the same Bayes risk (2.4). To study this setup formally, let $\mathcal{F}_{\mathbf{X}} = \sigma(\mathbf{X})$ be the σ-field generated by $\mathbf{X}$, and let $\mathcal{F}_{\mathbf{X},L} = \sigma(\mathbf{X}, L)$ be the oracle σ-field generated by both the data $\mathbf{X}$ and the latent factor L. The term 'oracle' suggests that L is unobservable in practice; however, we are assuming it to be observable to overcome dependence. Because $\mathcal{F}_{\mathbf{X}} \subset \mathcal{F}_{\mathbf{X},L}$, every $\mathcal{F}_{\mathbf{X}}$-measurable decision rule is also $\mathcal{F}_{\mathbf{X},L}$-measurable.

If L is observed alongside $\mathbf{X}$, the optimal oracle Bayes rule for the i-th coordinate is obtained by minimizing the individual expected loss (2.4) conditioned on $(\mathbf{X}, L)$. We get the oracle decision rules as

$$\delta_{0i}(\mathbf{X}, L) = \mathbf{1}\left(\frac{f_{\mathbf{X},L|\theta_i=1}(\mathbf{X}, L)}{f_{\mathbf{X},L|\theta_i=0}(\mathbf{X}, L)} > \Delta_m f_m\right). \tag{2.10}$$

Here, $f_{\mathbf{X},L|\theta_i=j}(\mathbf{X}, L)$ denotes the joint probability density function of $\mathbf{X}$ and L conditioned on $\theta_i = j$ for $j \in \{0, 1\}$. The minimal oracle Bayes risk $R_{0m}(\mathbf{X}, L)$ is obtained by putting $\delta_i = \delta_{0i}(\mathbf{X}, L)$ in (2.4), i.e.,

$$R_{0m}(\mathbf{X}, L) = \sum_{i=1}^{m} \mathbb{E}\left[\Delta_{0m}(1-\theta_i)\delta_{0i}(\mathbf{X}, L) + \Delta_{Am}\theta_i(1-\delta_{0i}(\mathbf{X}, L))\right].$$

Now, $\sigma(\mathbf{X}) \subset \sigma(\mathbf{X}, L)$, and $R_{0m}(\mathbf{X}, L)$ is obtained by minimizing the risk (2.6) over $\sigma(\mathbf{X}, L)$. Since, minimizing a function over the larger set makes the function smaller than if it were minimized over a smaller set, the left inequality of the following result becomes immediate.

Further, suppose we replace the latent variable L in the oracle optimal Bayes rules with its consistent fixed-sample counterpart $\bar{X}_m$ to get the data-driven rules $\hat{\delta}_{0i} = \delta_{0i}(\mathbf{X}, \bar{X}_m)$. Let $\hat{R}_{0m}(\mathbf{X}, \bar{X}_m)$ denote the total expected risk achieved by replacing the optimal oracle Bayes decision rules $\delta_{0i}(\mathbf{X}, L)$ with its data-driven version i.e. $\hat{\delta}_{0i} = \delta_{0i}(\mathbf{X}, \bar{X}_m)$. Therefore,

$$\hat{R}_{0m}(\mathbf{X}, \bar{X}_m) = \sum_{i=1}^{m} \mathbb{E}\left(\Delta_{0m}\hat{\delta}_{0i}(1-\theta_i) + \Delta_{Am}\theta_i(1-\hat{\delta}_{0i})\right). \tag{2.11}$$

Then the following inequalities become true.

Lemma 2. *Under the hierarchical exchangeable distribution framework defined in (2.1)–(2.3), and for the latent variable L introduced in Lemma 1,*

$$R_{0m}(\mathbf{X}, L) \leq R_{0m}(\mathbf{X}) \leq \hat{R}_{0m}(\mathbf{X}, \bar{X}_m), \text{ for } m \in \mathbb{N}. \tag{2.12}$$

This lemma shows that for a finite size m, the optimal Bayes risk with respect to the dataset $\mathbf{X}$ is bounded below and above respectively by the optimal oracle Bayes risk and the data-driven version of the said Bayes risk. The gap arises from the additional information available to the oracle procedure through observation of the latent factor L. However, since $\bar{X}_m$ is a strongly consistent estimator of L, the data-driven procedure asymptotically reproduces the oracle decisions. Under the additional non-degeneracy condition that the average oracle risk remains bounded away from zero, this approximation error becomes negligible relative to the oracle risk, leading to the following asymptotic equivalence result.

Theorem 1. *Let* $\{\gamma_m\}_{m\geq 1}$ *denote the sequence of parameter vectors introduced in Section 2.1. Suppose* $\sup_m(\Delta_{Am} + \Delta_{0m}) \leq M$. *Suppose further that the oracle Bayes risk satisfies* $\liminf_{m\to\infty} \frac{1}{m} R_{0m}(\mathbf{X}, L) > 0$. *Then*

$$\lim_{m\to\infty} \frac{\hat{R}_{0m}(\mathbf{X}, \bar{X}_m)}{R_{0m}(\mathbf{X}, L)} = 1. \tag{2.13}$$

Therefore, Lemma 2 and Theorem 1 together guarantee that for a large number of hypotheses, the optimal testing procedure based on only the data $\mathbf{X}$ can asymptotically achieve the exact efficiency of the oracle Bayes optimal rule.

This asymptotic equivalence further provides a justification for replacing the latent factor L by its consistent estimator $\bar{X}_m$ when constructing implementable testing procedures under the ABOS framework described in the next section. We note, however, the following result.

Corollary 1. *Suppose the hierarchical model (2.1)–(2.3) is true. Then for an asymptotic setup with respect to increasing* m *defined by the sequence of parameter vector* $\{\gamma_m\}_{m\geq 1}$, *if the optimal oracle Bayes risk* $R_{0m}(\mathbf{X}, L)$ *satisfies assumptions of Theorem 1, the data-driven decision rules* $\{\hat{\delta}_{0i}(\mathbf{X}, \bar{X}_m)\}_{m\geq 1}$ *are asymptotically Bayes optimal, i.e.,*

$$\lim_{m\to\infty} \frac{\hat{R}_{0m}(\mathbf{X}, \bar{X}_m)}{R_{0m}(\mathbf{X})} = 1.$$

The proof is immediate from Lemma 2 and Theorem 1. Lemma 1 shows that the exchangeable dependence structure specified in (2.1)–(2.3) can be represented through a single latent factor L. Since L is consistently estimable from the observed data via $\bar{X}_m$, under certain non-degeneracy conditions, the risk of the optimal Bayes rule $\{\delta_{0i}(\mathbf{X}), i \in [m]\}$ becomes asymptotically equivalent to that of the data-driven rule $\{\hat{\delta}_{0i}(\mathbf{X}), i \in [m]\}$. This observation forms the basis of the "approximate rules" proposed in Section 2.4.

2.3 Hypothetical Rules

The decomposition in Lemma 1 induces the following hierarchical representation for the latent variables $\mathbf{Z}$. For all $i \in [m]$,

$$
\begin{aligned}
Z_i \mid \mu_i &\sim N(\mu_i, (1-\rho)\sigma_\epsilon^2) \text{ independently for all } i \in [m], \\
\mu_i \mid \theta_i &\sim (1-\theta_i)N(0, \sigma_0^2) + \theta_i N(0, \sigma_0^2 + \tau^2), \text{ independently, for all } i \in [m], \\
\theta_i &\overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p) \text{ for all } i \in [m].
\end{aligned}
$$

Using the decomposition in Lemma 1, it follows that:

$$
\frac{f_{\mathbf{X},L|\theta_i=1}(\mathbf{X}, L)}{f_{\mathbf{X},L|\theta_i=0}(\mathbf{X}, L)} = \frac{f_{\mathbf{X}|\theta_i=1,L=l}(\mathbf{X})}{f_{\mathbf{X}|\theta_i=0,L=l}(\mathbf{X})} = \frac{f_{Z_i|\theta_i=1}(z_i)}{f_{Z_i|\theta_i=0}(z_i)}.
$$

Here, $f_{\mathbf{X}|\theta_i=j,L=l}(\mathbf{x})$ denotes the conditional probability density function of $\mathbf{X}$ evaluated at $\mathbf{x} = (x_1, \ldots, x_m)$ given $\theta_i = j$ and $L = l$. $z_i = x_i - l$ for all $i \in [m]$. The first equality follows by dividing the numerator and the denominator by $f_L(l)$ due to the fact that L is independent of θ_i for all $i \in [m]$. The second equality follows from (2.9), and since $Z_1, Z_2, \cdots, Z_m$ form a collection of independent Gaussian random variables and Z_j is independent of θ_i for all $i \neq j$.

Consequently, the oracle Bayes rule in (2.10) can be expressed as

$$
\delta_{0i}(\mathbf{X}, L) = \mathbb{1}\left(\frac{f_{Z_i|\theta_i=1}(z_i)}{f_{Z_i|\theta_i=0}(z_i)} > \Delta_m f_m\right) \text{ for all } i \in [m]. \tag{2.14}
$$

Note that the right-hand side is independent of L due to the independence of L and $\mathbf{Z}$. Since the decision rule for the i-th problem involves only Z_i, we can express $\delta_{0i}(\mathbf{X}, L)$ as $\delta_{0i}(Z_i)$ to emphasize this fact. A direct calculation yields

$$
\delta_{0i}(Z_i) = \mathbb{1}\left(\frac{Z_i^2}{\sigma_\rho^2} > c^2\right), \text{ for all } i \in [m], \tag{2.15}
$$

with $c^2 = (1 + \frac{1}{u})\left(\log(v) + \log(1 + \frac{1}{u})\right)$, $u = \frac{\tau^2}{\sigma_\rho^2}$, and $v = u\Delta^2 f^2$.

Since the oracle rule depends on $(\mathbf{X}, L)$ only through Z_i, the corresponding optimal risk may equivalently be expressed in terms of $\mathbf{Z}$.

$$R_{0m}(\mathbf{X}, L) = R_{0m}(\mathbf{Z}) = \sum_{i=1}^{m} \mathbb{E}\left(\Delta_{0m}(1-\theta_i)\delta_{0i}(Z_i) + \Delta_{Am}\theta_i(1-\delta_{0i}(Z_i))\right). \tag{2.16}$$

Now, suppose $u_m = \frac{\tau_m^2}{\sigma_{\rho m}^2}$ and $v_m = u_m \Delta_m^2 f_m^2$ denote the dependence of u and v on m. We make the following assumptions about the asymptotic framework:

Assumption 1. *Consider the framework described in (2.1)–(2.3). Let* $\gamma_m = (p_m, \tau_m, \sigma_{\rho m}^2, \Delta_{0m}, \Delta_{Am})$ *denote the sequence of parameters related to the framework. Let* u_m, v_m, Δ_m, *and* f_m *be the parameters defined earlier. Then we assume:*

A1 $p_m \to 0$,

A2 $u_m \to \infty$,

A3 $v_m \to \infty$, *and*

A4 $\frac{\log(v_m)}{1+u_m} \to C \in [0, \infty)$.

Assumption $A1$ ensures sparsity of the alternatives. Assumptions $A2$ and $A4$ prevent the individual testing problems from becoming asymptotically degenerate and ensure non-trivial limiting power, and assumption $A3$ prevents Δ_m from vanishing too rapidly. Note that Assumption $A4$ differs slightly from the assumptions of Bogdan et al. (2011) as there the authors assumed $\log(v_m)/u_m \to C$, for some $C \in [0, \infty)$. Since $u_m \to \infty$, we have $\log(v_m)/(u_m + 1) = \log(v_m)(1 + o(1))/u_m$, i.e. the two assumptions are asymptotically equivalent.

Theorem 2. *Suppose assumptions* $A1$–$A4$ *of Assumption 1 are satisfied. Then the optimal oracle Bayes risk takes the form*

$$R_{0m}(\mathbf{Z}) = m p_m \Delta_{Am}(2\Phi(\sqrt{C}) - 1)(1 + o_m), \tag{2.17}$$

where o_m is a sequence such that $o_m \to 0$. Moreover, a sequence of testing procedures whose ith decision rule is of the form (2.15) is ABOS if and only if $c^2 = c_m^2 = \log(v_m) + l_m$, where

$$l_m = o(\log v_m), \text{ and,} \tag{2.18}$$

$$l_m + 2\log\log v_m \to \infty. \tag{2.19}$$

Therefore, Theorem 2 yields the following asymptotic representation of the oracle Bayes risk:

$$\lim_{m\to\infty} \frac{R_{0m}(\mathbf{X}, L)}{m p_m \Delta_{Am}(2\Phi(\sqrt{C}) - 1)} = 1. \tag{2.20}$$

The theorem states that if both $\mathbf{X}$ and L were observable, then the oracle rules (2.15) would constitute an implementable ABOS procedure. However, we cannot directly observe $\mathbf{Z}$ or L. Hence, the set of rules so obtained is not implementable. These oracle rules serve as a benchmark and form the basis for constructing the implementable procedures developed in the next section.

2.4 Approximate Rules

As established in Section 2.2, the sample mean $\bar{X}_m$ is a strongly consistent estimator of the latent factor L. Consequently, the quantities

$$\hat{Z}_i = X_i - \bar{X}_m, \qquad i \in [m],$$

provide natural empirical approximations of the latent variables Z_i. Motivated by this observation, we replace the unobservable quantities Z_i in the oracle rules by their empirical counterparts $\hat{Z}_i$. Since $\hat{Z}_i$ is a function of the observed data vector $\mathbf{X}$, these terms can be formally used for testing purposes. We now construct a class of common-threshold approximate rules by replacing Z_i with $\hat{Z}_i$ as follows:

$$\hat{\delta}_i = \mathbb{1}\left(\hat{Z}_i^2/\sigma_\rho^2 > \log v_m + l_m\right), \tag{2.21}$$

where l_m satisfies conditions (2.18) and (2.19). The next theorem establishes the core result of this chapter, showing that the set of approximate rules given by equation

(2.21) is ABOS. To prove this, we require an additional tracking constraint on the signal profile.

Assumption 2. *For the random vector* $\mathbf{X}$ *satisfying the framework (2.1)–(2.3), we assume that for the parameter sequence* γ_m*, and for some fixed exponent* $\lambda \in (0,1)$*, as* $m \to \infty$*:*

A5 $\tau_m p_m m^{-\lambda} \to K \in [0, \infty)$.

Assumption A5 controls the joint growth rate of the signal variance parameter τ_m and the sparsity level p_m relative to the dimension m. In particular, it prevents the aggregate signal strength from increasing too rapidly as the number of hypotheses grows.

Theorem 3. *Suppose that conditions A1–A4 of Assumption 1 and condition A5 of Assumption 2 are satisfied. Let* $\hat{R}_{am}$ *denote the Bayes risk associated with the approximate testing procedure (2.21). Then*

$$\lim_{m\to\infty} \frac{\hat{R}_{am}}{m\Delta_{Am}p_m(2\Phi(\sqrt{C})-1)} = 1. \tag{2.22}$$

Combining Theorems 2 and 3 we obtain

$$\lim_{m\to\infty} \frac{R_{0m}(\mathbf{X}, L)}{\hat{R}_{am}} = 1. \tag{2.23}$$

Now, from Lemma 2 we have $R_{0m}(\mathbf{X}, L) \leq R_{0m}(\mathbf{X})$. Further, since $R_{0m}(\mathbf{X})$ is the minimal Bayes risk based on the observed data $\mathbf{X}$, we have $R_{0m}(\mathbf{X}) \leq \hat{R}_{am}$. Combining these two inequalities and dividing them by $\hat{R}_{am}$ we get

$$\frac{R_{0m}(\mathbf{X}, L)}{\hat{R}_{am}} \leq \frac{R_{0m}(\mathbf{X})}{\hat{R}_{am}} \leq 1.$$

Therefore, (2.23) implies that the the approximate testing procedure (2.21) is ABOS. This establishes that the proposed data-driven procedure retains the asymptotic optimality of the oracle benchmark and therefore achieves asymptotic Bayes optimality under the equicorrelated sparse framework considered in this chapter.

2.5 Conditions for Some Classical and Bayesian Multiple Testing Methods to be ABOS

In this section, we examine several well-known multiple testing procedures within the framework developed in the preceding sections and derive sufficient conditions under which they are ABOS. If we could observe $\mathbf{Z}$ (or equivalently L along with $\mathbf{X}$), then by the Monotone Likelihood Ratio (MLR) property and the Neyman-Pearson lemma, rejection of H_{0i} is justified for large values of Z_i^2/σ_ρ^2. Observe that

$$\hat{Z}_i = X_i - \bar{X}_m = Z_i - \bar{Z}_m,$$

since $X_i = Z_i + L$ and $\bar{X}_m = \bar{Z}_m + L$. Therefore,

$$\hat{Z}_i^2 = (Z_i - \bar{Z}_m)^2 = Z_i^2 - 2Z_i\bar{Z}_m + \bar{Z}_m^2.$$

Since $\bar{Z}_m \to 0$ almost surely by SLLN, it follows that

$$\hat{Z}_i^2 - Z_i^2 = -2Z_i\bar{Z}_m + \bar{Z}_m^2 \longrightarrow 0 \qquad \text{a.s.}$$

for every fixed $i \in [m]$. Thus, the statistic $\hat{Z}_i^2$ provides an asymptotically equivalent surrogate to the oracle statistic Z_i^2. Hence, a large observation of $\hat{Z}_i^2/\sigma_\rho^2$ provides evidence against H_{0i}. We shall reject H_{0i} if $\hat{Z}_i^2/\sigma_\rho^2$ exceeds a chosen critical threshold sequence c_m^2. This motivates decision rules of the form

$$\hat{\delta}_i = \mathbb{1}\left(\frac{\hat{Z}_i^2}{\sigma_\rho^2} > c_m^2\right). \tag{2.24}$$

In the following subsections, we derive the explicit parameter restrictions under which standard multiple testing procedures reduce asymptotically to this optimal configuration.

2.5.1 Preliminary Concepts of Multiple Hypotheses Testing

Table 1.1 summarizes the possible classification outcomes, defining Type I errors (false rejections) and Type II errors (false acceptances). In high-dimensional regimes,

classical error rates include the Family-Wise Error Rate ($FWER$), defined as the probability of committing at least one false rejection:

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

The classical Bonferroni correction and the Šidák procedure are standard techniques designed to control the $FWER$ at a pre-specified nominal level α.

Let $R = \sum_{i=1}^{m} \delta_i$ denote the total number of rejections. We define the False Discovery Proportion (FDP) as the ratio of false rejections to total rejections, set to zero if no rejections occur:

$$FDP = \frac{V}{R \vee 1}.$$

The False Discovery Rate (FDR), introduced by Benjamini and Hochberg, 1995, is the expected value of this proportion:

$$FDR = \mathbb{E}(FDP) = \mathbb{E}\left(\frac{V}{R \vee 1}\right).$$

The Benjamini-Hochberg (BH) step-up procedure is a classical sequential thresholding method designed to bound the FDR below α.

2.5.2 Bayesian FDR Control

Following the empirical Bayes framework of Efron and Tibshirani, 2002, the Bayesian False Discovery Rate ($BFDR$) is defined as the posterior probability of a null hypothesis being true given that it has been rejected:

$$BFDR = \mathbb{P}(H_{0i} \text{ is true} \mid H_{0i} \text{ is rejected}) = \frac{(1-p)t_1}{(1-p)t_1 + p(1-t_2)}, \tag{2.25}$$

where t_1 and t_2 represent the common marginal probabilities of Type I and Type II errors across the exchangeable coordinates. For the fixed-cutoff rule (2.24), these marginal error probabilities satisfy:

$$t_1 = 2(1 - \Phi(c_m)) + o(1), \tag{2.26}$$

$$t_2 = \left(2\Phi\left(\frac{c_m}{\sqrt{u_m + 1}}\right) - 1\right) + o(1). \tag{2.27}$$

Consider a nominal control sequence $\{\alpha_m\}$ such that $\alpha_m \in (0,1)$ and $\alpha_m \to \alpha < 1$. Equating the $BFDR$ expression in (2.25) to α_m yields the boundary equation:

$$\frac{1-\Phi(c_m)}{1-\Phi\left(\frac{c_m}{\sqrt{u_m+1}}\right)} = \frac{r_{\alpha_m}}{f_m}, \tag{2.28}$$

where $r_{\alpha_m} = \alpha_m/(1-\alpha_m)$ and $f_m = (1-p_m)/p_m$. Under Assumption 1, $c_m/\sqrt{u_m+1} \to \sqrt{C}$, which implies $c_m \to \infty$. Thus, $\Phi(c_m) \to 1$, requiring that $r_{\alpha_m}/f_m \to 0$ as $m \to \infty$. This structure yields the following result:

Theorem 4. *Suppose assumptions A1–A4 of Assumption 1 and assumption A5 of Assumption 2 are satisfied. Consider the fixed-threshold rule (2.24) calibrated to control the BFDR at level α_m. Let the sequence s_m be defined via:*

$$\frac{\log(f_m\sqrt{u_m})}{\log(f_m/r_{\alpha_m})} = 1 + s_m.$$

If the parameter sequence satisfies:

$$s_m \to 0, \quad \text{as } m \to \infty, \tag{2.29}$$

and

$$2s_m \log\left(\frac{f_m}{r_{\alpha_m}}\right) - \log\log\left(\frac{f_m}{r_{\alpha_m}}\right) \to -\infty, \tag{2.30}$$

then the BFDR-controlled decision rule is ABOS.

2.5.3 Optimality of the Asymptotic Approximation to the BH Threshold

It was established by Genovese and Wasserman, 2002 that for a fixed non-zero proportion of alternatives p as $m \to \infty$, the empirical threshold of the Benjamini-Hochberg procedure can be approximated by the deterministic asymptotic cutoff $c_{GW,m}$, defined implicitly by:

$$\frac{1-\Phi(c_{GW,m})}{(1-p)(1-\Phi(c_{GW,m})) + p\left(1-\Phi\left(c_{GW,m}/\sqrt{u+1}\right)\right)} = \alpha_m. \tag{2.31}$$

The functional equation in (2.31) differs from the definition of $BFDR$ in (2.25) only by the absence of the $(1-p_m)$ multiplier in the numerator. Under extreme sparsity

where $p_m \to 0$, the scaling ratio $(1 - p_m) \to 1$, causing the $BFDR$ cutoff sequence to converge directly to the Genovese-Wasserman boundary.

Theorem 5. *Consider the constant threshold rule* $\delta_i^{GW} = \mathbb{1}\left(\hat{Z}_i^2/\sigma_\rho^2 \geq c_{GW}^2\right)$. *If conditions A1–A4 of Assumption 1 and condition A5 of Assumption 2 hold, then this rule is ABOS whenever the baseline fixed-cutoff rule defined in (2.24) is ABOS. That is,* δ_i^{GW} *is ABOS for all* $i \in [m]$ *if conditions (2.29) and (2.30) are satisfied, in which case:*

$$c_{GW}^2 = c_m^2 + o(1), \quad \textit{almost surely.} \tag{2.32}$$

2.5.4 Asymptotic Optimality Under Sparsity for Bonferroni's Procedure

For the classical Bonferroni correction, the individual testing decisions are given by $\hat{\delta}_i = \mathbb{1}\left(\hat{Z}_i^2/\sigma_\rho^2 \geq c_{\text{Bon}}^2\right)$, where the critical threshold sequence is chosen such that:

$$1 - \Phi(c_{\text{Bon}}) = \frac{\alpha}{2m}. \tag{2.33}$$

Under sufficiently strong sparsity conditions, this conservative family-wise control procedure achieves asymptotic risk optimality.

Theorem 6. *Consider the hierarchical distribution framework defined by (2.1)–(2.3), and let assumptions A1–A4 of Assumption 1 and condition A5 of Assumption 2 hold. Furthermore, assume that the sparsity parameter satisfies the constraint:*

$$mp_m \to s \in (0, \infty], \quad \textit{and} \quad \frac{\log(mp_m)}{\log(m)} \to 0 \quad \textit{as } m \to \infty. \tag{2.34}$$

Then the Bonferroni procedure controlling the FWER at level α_m *(where* $\alpha_m \to \alpha_\infty \in [0, 1)$*) is ABOS if the conditions of Theorem 4 are satisfied.*

2.5.5 Asymptotic Optimality Under Sparsity for Šidák's Procedure

The classical Šidák testing procedure defines the coordinate decision rules via $\delta_{\check{S}i} = \mathbb{1}\left(\frac{\hat{Z}_i^2}{\sigma_\rho^2} > c_{\check{S}}^2\right)$, where the threshold sequence is given by:

$$1 - \Phi(c_{\check{S}}) = 1 - (1 - \alpha)^{\frac{1}{2m}}. \tag{2.35}$$

By Bernoulli's inequality, $(1 - \frac{\alpha}{2m})^m \geq 1 - \frac{\alpha}{2}$, which implies $c_{\text{Bon}} \geq c_{\check{S}}$. For large values of m, an asymptotic expansion of the Bonferroni cutoff in (2.33) yields:

$$c_{\text{Bon}}^2 = 2\log\left(\frac{m}{\alpha}\right) - \log\left(2\log\left(\frac{m}{\alpha}\right)\right) + \log\left(\frac{2}{\pi}\right) + o(1). \tag{2.36}$$

The following lemma establishes that the difference between the Bonferroni and Šidák critical values is asymptotically negligible.

Lemma 3. *Let c_{Bon} and $c_{\check{S}}$ denote the critical thresholds defined in (2.33) and (2.35), respectively. Then:*

$$c_{Bon}^2 = c_{\check{S}}^2 + o(1). \tag{2.37}$$

We leverage this equivalence to establish the asymptotic optimality of Šidák's family-wise error control under sparse setups.

Theorem 7. *Consider the hierarchical distribution framework defined by (2.1)–(2.3), and let assumptions A1–A4 of Assumption 1, condition A5 of Assumption 2, and the extreme sparsity condition (2.34) hold. Then the Šidák procedure controlling the FWER at level α_m (where $\alpha_m \to \alpha_\infty \in [0, 1)$) is ABOS if the conditions of Theorem 4 are satisfied.*

2.5.6 Asymptotic Optimality Under Sparsity for Benjamini-Hochberg Test

To implement the sequential Benjamini-Hochberg procedure, we compute the marginal p-values using our empirical residuals. Let $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$ denote the ordered p-values calculated for the m simultaneous hypotheses. To control the FDR at a nominal target level α, the step-up procedure rejects all hypotheses whose p-values satisfy $p_{(i)} \leq p_{(k)}$, where the random stopping index k is defined by:

$$k = \max\left\{i \in [m] : p_{(i)} \leq \frac{i\alpha}{m}\right\}. \tag{2.38}$$

The following theorem establishes the sufficient conditions for this data-dependent sequential test to achieve the optimal risk profile under equicorrelated dependencies.

Theorem 8. *Consider the hierarchical distribution framework defined by (2.1)–(2.3), and let assumptions A1–A4 of Assumption 1 and condition A5 of Assumption 2 hold. Assume*

that the prior probability sequence satisfies the sparse scaling condition:

$$p_m \to 0 \quad \textit{and} \quad m p_m \to s \in (0, \infty] \quad \textit{as } m \to \infty.$$

Then the Benjamini-Hochberg step-up procedure calibrated to control the FDR at level α_m (where $\alpha_m \to \alpha_\infty < 1$) is ABOS if the critical growth conditions (2.29) and (2.30) hold.

Theorems 4–8 demonstrate that a broad class of multiple testing procedures attain asymptotic Bayes optimality under the equicorrelated sparse framework, provided the corresponding threshold sequences satisfy the growth conditions identified in Section 2.3.

2.6 Proof of Theorems

2.6.1 Proof of Theorem 1

In this theorem, we prove that the ratio of $\hat{R}_{0m}(\mathbf{X}, \bar{X}_m)$, the data-driven estimate of the oracle optimal Bayes risk based on the data $\mathbf{X}$ and $R_{0m}(\mathbf{X}, L)$, the optimal oracle Bayes risk based on the data $\mathbf{X}$ together with the latent variable L converges to 1. We have already established that due to Lemma 1, the mean of the observed data $\bar{X}_m$ converges almost surely to the latent variable L. Therefore, $\bar{X}_m$ works as a consistent estimator for L. Consequently, $\hat{R}_{0m}(\mathbf{X}, \bar{X}_m)$ provides a good estimate of the optimal oracle Bayes risk $R_{0m}(\mathbf{X}, L)$.

To analyse the asymptotic behavior, we examine the coordinate-wise expected loss functions For a given coordinate i, let $l_i(\mathbf{X}, L)$ represent the loss of the oracle rule $\delta_{0i}(\mathbf{X}, L)$, i.e., $l_i(\mathbf{X}, L) = \Delta_{0m}\delta_{0i}(\mathbf{X}, L)(1 - \theta_i) + \Delta_{Am}\theta_i(1 - \delta_{0i}(\mathbf{X}, L))$ and let $\hat{l}_i(\mathbf{X})$ represent the loss of the data-driven rule $\hat{\delta}_{0i}$), i.e., $\hat{l}_i(\mathbf{X}) = \Delta_{0m}\hat{\delta}_{0i}(1 - \theta_i) + \Delta_{Am}\theta_i(1 - \hat{\delta}_{0i})$. We can express the total risks as the sums of these expected losses:

$$R_{0m}(\mathbf{X}, L) = \sum_{i=1}^{m} \mathbb{E}\left[l_i(\mathbf{X}, L)\right],$$
$$\hat{R}_{0m}(\mathbf{X}, \bar{X}_m) = \sum_{i=1}^{m} \mathbb{E}\left[\hat{l}_i(\mathbf{X})\right].$$

Recall from Lemma 1 that each coordinate decomposes as $X_i = Z_i + L$. By the SLLN, since the coordinates Z_i are conditionally independent with $\mathbb{E}[Z_i \mid \boldsymbol{\theta}] = 0$,

the empirical average satisfies:

$$\bar{X}_m = \bar{Z}_m + L \longrightarrow L \quad \text{almost surely as } m \to \infty. \tag{2.39}$$

From the decomposition established in Lemma 1, we have

$$\frac{f_{\mathbf{X},L|\theta_i=1}(\mathbf{X}, L)}{f_{\mathbf{X},L|\theta_i=0}(\mathbf{X}, L)} = \frac{f_{Z_i|\theta_i=1}(Z_i)}{f_{Z_i|\theta_i=0}(Z_i)}.$$

Therefore, the oracle decision rule depends on $(\mathbf{X}, L)$ only through the latent variable $Z_i = X_i - L$. Define

$$g(z) = \frac{f_{Z_i|\theta_i=1}(z)}{f_{Z_i|\theta_i=0}(z)}.$$

Since $f_{Z_i|\theta_i=0}$ and $f_{Z_i|\theta_i=1}$ are Gaussian densities, the function g is continuous on $\mathbb{R}$.

Now,

$$X_i - \bar{X}_m = Z_i - \bar{Z}_m,$$

and by SLLN,

$$\bar{Z}_m \to 0 \qquad \text{a.s.}$$

Hence,

$$X_i - \bar{X}_m \to Z_i \qquad \text{a.s.}$$

for every fixed $i \in [m]$. By continuity of g the continuous mapping theorem implies,

$$g(X_i - \bar{X}_m) \to g(Z_i) \qquad \text{a.s.}$$

As the distribution of $g(Z_i)$ is continuous, the probability of the threshold event is zero, i.e., $\Pr(g(Z_i) = \Delta_m f_m) = 0$. Consequently, the indicator function converges almost surely at all continuity points of the indicator."Since Z_i has a continuous distribution, i.e.,

$$\hat{\delta}_{0i}(\mathbf{X}) = \mathbb{1}\big(g(X_i - \bar{X}_m) > \Delta_m f_m\big) \longrightarrow \mathbb{1}\big(g(Z_i) > \Delta_m f_m\big) = \delta_{0i}(\mathbf{X}, L)$$

almost surely. Consequently, for every fixed $i \in [m]$.

$$\left|\hat{l}_i(\mathbf{X}) - l_i(\mathbf{X}, L)\right| \longrightarrow 0 \quad \text{almost surely as } m \to \infty. \tag{2.40}$$

The absolute loss difference $|\hat{l}_i(\mathbf{X}) - l_i(\mathbf{X}, L)|$ is bounded by $\Delta_{0m} + \Delta_{Am}$. Since this bound is finite and does not depend on the realization of $\mathbf{X}$, the conditions for Pratt's extension of the Dominated Convergence Theorem (Pratt, 1960) are satisfied. Therefore,

$$\mathbb{E}\left[|\hat{l}_i(\mathbf{X}) - l_i(\mathbf{X}, L)|\right] \longrightarrow 0$$

for every fixed $i \in [m]$.

By symmetry of the hierarchical model and the coordinatewise construction of the decision rules, the distributions of

$$|\hat{l}_i(\mathbf{X}) - l_i(\mathbf{X}, L)|, \qquad i = 1, \ldots, m,$$

are identical across i. Therefore,

$$\mathbb{E}\left[|\hat{l}_i(\mathbf{X}) - l_i(\mathbf{X}, L)|\right] = a_m,$$

for some sequence $a_m \to 0$ that does not depend on i. Hence,

$$\frac{1}{m}\sum_{i=1}^{m} \mathbb{E}\left[|\hat{l}_i(\mathbf{X}) - l_i(\mathbf{X}, L)|\right] = a_m \longrightarrow 0.$$

Using the triangle inequality,

$$\frac{\left|\hat{R}_{0m}(\mathbf{X}, \bar{X}_m) - R_{0m}(\mathbf{X}, L)\right|}{m} \leq \frac{1}{m}\sum_{i=1}^{m} \mathbb{E}\left[|\hat{l}_i(\mathbf{X}) - l_i(\mathbf{X}, L)|\right],$$

and therefore

$$\lim_{m\to\infty} \frac{\hat{R}_{0m}(\mathbf{X}, \bar{X}_m) - R_{0m}(\mathbf{X}, L)}{m} = 0.$$

Under the non-degeneracy condition

$$\liminf_{m\to\infty} \frac{1}{m} R_{0m}(\mathbf{X}, L) > 0,$$

it follows that

$$\lim_{m\to\infty} \frac{\hat{R}_{0m}(\mathbf{X}, \bar{X}_m)}{R_{0m}(\mathbf{X}, L)} = 1. \tag{2.41}$$

This completes the proof of Theorem 1.

Proof of Theorem 2

Proof. Theorem 2 follows from theorem 3.1 and theorem 3.2 of Bogdan et al., 2011.

□

Proof of Theorem 3

Proof. Here, the Bayes risk is

$$\begin{aligned}\hat{R}_{am} &= \sum_{i=1}^{m} \Delta_{0m}\mathbb{E}(\hat{\delta}_i(1-\theta_i)) + \sum_{i=1}^{m} \Delta_{Am}\mathbb{E}((1-\hat{\delta}_i)\theta_i) \\ &= \sum_{i=1}^{m} \Delta_{0m}\hat{t}_{0mi} + \sum_{i=1}^{m} \Delta_{Am}\hat{t}_{1mi}.\end{aligned} \tag{2.42}$$

The proof proceeds by calculating $\hat{t}_{0mi}$ and $\hat{t}_{1mi}$ which represent respectively the probabilities of the Type I and Type II errors for testing the ith hypothesis. We consider a sequence $\{r_m\}$ such that, $r_m \to \infty$ and $r_m = o(\sqrt{\log v_m})$. We define $\bar{Z}_{mi} = \sum_{j\in[m],j\neq i} Z_j/(m-1)$. The law of total probability shall be applied with respect to the complimentary events $\{|\bar{Z}_{mi}| > r_m\}$ and $\{|\bar{Z}_{mi}| \leq r_m\}$.

Note that, $m^{-\lambda/2}Z_j$, $j \in [m]$, $j \neq i$ are i.i.d. random variables with mean 0 and variance $S_m^2 = m^{-\lambda}(\sigma_{\rho m}^2 + \tau_m p_m)$ which is finite due to $A5$. Therefore,

$$\begin{aligned}\mathbb{P}\left(|\bar{Z}_{mi}| > r_m\right) &= \mathbb{P}\left(\frac{\sqrt{(m-1)}m^{-\lambda/2}|\bar{Z}_{mi}|}{S_m} > \frac{\sqrt{(m-1)}m^{-\lambda/2}r_m}{S_m}\right) \\ &= 2\left(1-\Phi\left(\frac{\sqrt{(m-1)}m^{-\lambda/2}r_m}{S_m}\right)\right)(1+o(1)) \quad \text{(Due to CLT)} \\ &= \phi\left(\frac{m^{(1-\lambda)/2}r_m}{S_m}\right)\frac{S_m}{m^{(1-\lambda)/2}r_m}(1+o(1)) \\ &= \frac{1}{\sqrt{2\pi}}\frac{S_m \exp\left(-m^{(1-\lambda)}r_m^2/(2S_m^2)\right)}{m^{(1-\lambda)/2}r_m}(1+o(1)) \\ &= g_m, \quad \text{say.}\end{aligned} \tag{2.43}$$

It is evident that, $g_m = o(1)$. The equality in the third line of (2.43) follows due to the fact that for $T \sim N(0,1)$ and for some sequence $\{a_m\}$ diverging to ∞,

$$\mathbb{P}(|T| > a_m) = \frac{1}{a_m}\phi(a_m)(1+o(1)). \tag{2.44}$$

Now we define

$$\mathbb{P}\left(\frac{(Z_i-\bar{Z}_m)^2}{\sigma_\rho^2} > \log v_m + l_m \,\middle|\, |\bar{Z}_{mi}| > r_m, \theta_i = 0\right) = h_{0m}, \text{ and}$$
$$\mathbb{P}\left(\frac{(Z_i-\bar{Z}_m)^2}{\sigma_\rho^2} \leq \log v_m + l_m \,\middle|\, |\bar{Z}_{mi}| > r_m, \theta_i = 1\right) = h_{1m}.$$

Note that, h_{0m} and h_{1m} are bounded above by 1. Then

$$\begin{aligned}\hat{t}_{0i} &= \mathbb{P}\left(\frac{(Z_i-\bar{Z}_m)^2}{\sigma_\rho^2} > \log v_m + l_m \,\middle|\, \theta_i = 0\right)\\ &= \mathbb{P}\left(\frac{(Z_i-\bar{Z}_m)^2}{\sigma_\rho^2} > \log v_m + l_m \,\middle|\, \theta_i = 0, |\bar{Z}_{mi}| \leq r_m\right) \times (1-g_m) + g_m h_{0m},\end{aligned}$$

and since, $g_m = o(1)$, we can say that,

$$\hat{t}_{0i} = \mathbb{P}\left(\frac{(Z_i-\bar{Z}_m)^2}{\sigma_\rho^2} > \log v_m + l_m \,\middle|\, \theta_i = 0, |\bar{Z}_{mi}| \leq r_m\right)(1+o(1)) + o(1).$$

Now,

$$\begin{aligned}&\mathbb{P}\left(\frac{(Z_i-\bar{Z}_m)^2}{\sigma_\rho^2} > \log v_m + l_m \,\middle|\, \theta_i = 0, |\bar{Z}_{mi}| \leq r_m\right)\\ &= \mathbb{P}_0\left(\frac{(Z_i-\bar{Z}_{mi})^2}{\sigma_\rho^2} > (1-\frac{1}{m})^{-2}(\log v_m + l_m) \,\middle|\, |\bar{Z}_{mi}| \leq r_m\right)\\ &= \mathbb{P}_0\left(\frac{(Z_i-\bar{Z}_{mi})^2}{\sigma_\rho^2} > C_m \,\middle|\, |\bar{Z}_{mi}| \leq r_m\right)\\ &= \mathbb{P}_0\left(\frac{Z_i^2}{\sigma_\rho^2} > C_m - 2\sqrt{C_m}\frac{\bar{Z}_{mi}}{\sigma_\rho} + \frac{\bar{Z}_{mi}^2}{\sigma_\rho^2} \,\middle|\, |\bar{Z}_{mi}| \leq r_m\right)\\ &= \mathbb{P}_0\left(\frac{Z_i^2}{\sigma_\rho^2} > C_m(1+o(1))\right) \text{ (Since } r_m/\sqrt{\log v_m} \to 0)\\ &= \frac{\exp\left(-C_m/2\,(1+o(1))\right)}{\sqrt{2\pi C_m}(1+o(1))}(1+o\left(C_m(1+o(1))\right))\\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{v_m \log v_m}}(1+o(1)).\end{aligned}$$

Here, $\mathbb{P}_0$ implies probability under H_{0i} for any $i \in [m]$. It is easy to see that $(Z_i -$

$\bar{Z}_m) = (1 - 1/m)(Z_i - \bar{Z}_{mi})$. We consider $C_m = (1 - \frac{1}{m})^{-2}(\log v_m + l_m)$. Then $C_m = \log v_m(1 + o(1))$. Now the probability in line 3 is equivalent to finding the probability of the event $H_m = \{\frac{|Z_i - \bar{Z}_{mi}|}{\sigma_\rho} > \sqrt{C_m}\}$ subject to $|\bar{Z}_{mi}| \leq r_m$. Here, we have conditioned on the event $\{|\bar{Z}_{mi}| \leq r_m\}$, which implies that $\bar{Z}_{mi} = o(\sqrt{C_m})$. Hence, $H_m = \{\frac{|Z_i| + \bar{Z}_{mi}}{\sigma_\rho} > \sqrt{C_m}\} \cup \{\frac{|Z_i| - \bar{Z}_{mi}}{\sigma_\rho} > \sqrt{C_m}\}$. But $\bar{Z}_{mi}$ is symmetric about 0 and as a result, both the events have equal probabilities. Therefore, line 3 is equal to the probability of the event $\{\frac{|Z_i|}{\sigma_\rho} > \sqrt{C_m} - \frac{\bar{Z}_{mi}}{\sigma_\rho}\}$ given $|\bar{Z}_{mi}| \leq r_m$. Squaring both sides, we reach line 4. In the sixth line, we use the result mentioned in (2.44). Finally, we get

$$\hat{t}_{0i} = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{v_m \log v_m}}(1 + o(1)) + o(1). \tag{2.45}$$

For calculating $\hat{t}_{1i}$, we note that the initial calculations are the same albeit with reversed inequality. Therefore, we get

$$\begin{aligned}
\hat{t}_{1i} &= \mathbb{P}_1\left(\frac{Z_i^2}{\sigma_\rho^2} \leq C_m - 2\sqrt{C_m}\frac{\bar{Z}_{mi}}{\sigma_\rho} + \frac{\bar{Z}_{mi}^2}{\sigma_\rho^2} \,\middle|\, |\bar{Z}_{mi}| \leq r_m\right)(1 - g_m) + g_m h_{1m} \\
&= \mathbb{P}_1\left(\frac{Z_i^2}{\sigma_\rho^2 + \tau^2} \leq \frac{C_m}{1 + u_m} - 2\sqrt{C_m}\frac{\bar{Z}_{mi}}{(1 + u_m)\sigma_\rho} + \frac{\bar{Z}_{mi}^2}{(1 + u_m)\sigma_\rho^2} \,\middle|\, |\bar{Z}_{mi}| \leq r_m\right)(1 - g_m) \\
&\quad + g_m h_{1m} = (2\Phi(\sqrt{C}) - 1)(1 + o(1)).
\end{aligned} \tag{2.46}$$

Like before, $\mathbb{P}_1$ implies probability under H_{1i} for any $i \in [m]$. In the second line, we divide both sides of the inequality by $(1 + u_m)$. Then we note that since $r_m/\sqrt{\log v_m} = o(1)$, the right side of the inequality becomes $(\log v_m/(1 + u_m))(1 + o(1))$. Then by assumption $A4$, the result follows.

Finally if we put the values of $\hat{t}_{0i}$ and $\hat{t}_{1i}$ in (2.42), Theorem 3 is proved. □

Proof of Theorem 4

Proof. Theorem 4.1 of Bogdan et al., 2011 says that the fixed cutoff decision rules based on Z_i as the observations, i.e.,

$$\delta_{0i}(Z_i) = \mathbb{1}\left(\frac{Z_i^2}{\sigma_\rho^2} > c_m^2\right), \tag{2.47}$$

is ABOS with respect to the baseline oracle risk profile $R_{0m}(\mathbf{X}, L)$ if and only if (2.29) and (2.30) are satisfied. By invoking the global structural lower bound established in Theorem 1 (and Lemma 2), the oracle risk sets the minimal attainable risk ceiling for any data-only measurable rule configuration since $R_{0m}(\mathbf{X}) \geq R_{0m}(\mathbf{X}, L)$.

Now, since conditions (2.29) and (2.30) are true, the decision rules given by (2.47) are ABOS. Therefore, due to the uniform risk tracking established in Theorem 3 in conjunction with the bounding behavior guaranteed by Theorem 1, the empirical, data-driven set of rules (2.24) is successfully squeezed to the optimal trajectory. This proves Theorem 4. □

Proof of Theorem 5

Proof. Due to Theorem 4.2 of Bogdan et al., 2011, the set of decision rules given by $\delta^{GW}(Z_i) = \mathbb{1}\left(Z_i^2/\sigma_\rho^2 \geq c_{GW}^2\right)$ (the rule obtained by replacing Z_i in place of $\hat{Z}_i$ in δ_i^{GW}) is ABOS if and only if (2.29) and (2.30) are satisfied. By utilizing the bounding properties established under Theorem 1, the oracle baseline risk controls the lower optimization floor for the empirical framework. And therefore, with the use of Theorem 3, the set of data-driven decision rules δ_i^{GW} is successfully bounded and shown to be ABOS as well. Therefore, Theorem 5 is proved. □

Proof of Theorem 6

Proof. The proof follows from lemma 5.1 of Bogdan et al., 2011 and Theorem 3 in the same lines as that of Theorems 4 and 5. Here, the uniform convergence to the oracle ceiling guaranteed by Theorem 1 ensures that the risk inflation caused by substituting the empirical factor average vanishes asymptotically, preserving the ABOS property. □

Proof of Theorem 7

Proof. According to lemma 5.1 of Bogdan et al., 2011 and lemma 3, $c_{\check{S}}^2 = \log v_m + y_m$ with y_m following (2.18) and (2.19). By importing the structural risk squeeze from Theorem 1, the empirical Šidák cutoff rules are constrained within the optimal risk envelope. Therefore, using Theorem 3, the proof follows. The proofs are similar to proofs of Theorems 4 and 5. □

Proof of Theorem 8

Proof. Theorem 5.2 in Bogdan et al., 2011 implies that the BH rule using (2.47) is ABOS if (2.29) and (2.30) hold. Therefore, using Theorem 3, we prove this theorem. □

2.7 Proof of Lemmas

Proof of Lemma 1

Proof. This decomposition will be shown in two steps. First, the convolution property of the Bayesian prior has been used to decompose $\mathbf{X}$ as convolutions of prior $\boldsymbol{\mu}$ and $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_m)^T$ according to Theorem 2.1 of Heuvel and Klaassen, 1999. Here, $\mathbf{Y}$ follows a multivariate normal distribution with mean $\mathbf{0}$ and the same covariance matrix as $\mathbf{X}$, i.e., a covariance matrix with equal variance σ_ϵ^2 and equal correlation coefficient ρ. For all $i \in [m]$,

$$X_i = \mu_i + Y_i, \tag{2.48}$$

such that, $\mathbf{Y} \sim MVN(\mathbf{0}, \Sigma_\epsilon)$, and $\mathbf{Y}$ and $\boldsymbol{\mu}$ are independent. Now, for each $i \in [m]$, according to the hierarchical distribution structure of $\mathbf{X}$, the distribution of μ_i is determined by θ_i. Since, for each $j \in [m]$, and especially for $j = i$, Y_j is independent of μ_i, Y_j is independent of θ_i also.

Next, we write each member of the equicorrelated normal random vector $\mathbf{Y}$ as the sum of two independent normal variables, of which one remains the same for all the coordinates and represents the dependent part of the random vector $\mathbf{Y}$. The other part constitutes independently and identically distributed Gaussian normal variables and symbolizes the independent components of $\mathbf{Y}$. The following Lemma depicts the fact.

Lemma 4. *Consider a random vector* $\mathbf{U} = (U_1, U_2, \cdots, U_n)^T$ *such that*

$$\mathbf{U} \sim MVN\left(\mathbf{0},\ (1-r)I + r\mathbf{1}\mathbf{1}^T\right),$$

where $r \in (0,1)$, $\mathbf{1}^T = (1,\ 1,\ \cdots,\ 1)^{(1\times n)}$, *and* $\mathbf{0}^T = (0,\ 0,\ \cdots,\ 0)^{(1\times n)}$. *Then there exists a set of* $n+1$ *independent random variables* $(W_1,\ W_2,\ \cdots,\ W_n,\ Q)$, *such that,*

$$W_i \overset{\text{i.i.d.}}{\sim} N(0,\ (1-r)) \text{ for all } i \in [n],$$
$$Q \sim N(0,\ r),$$

and, $U_i = W_i + Q$ *for all* $i \in [n]$.

Due to lemma 4, we can write

$$Y_i = P_i + L, \tag{2.49}$$

where

$$P_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma_\epsilon^2(1-\rho)),$$
$$L \sim N(0, \rho\sigma_\epsilon^2).$$

Now, L is independent to each P_i. Therefore due to equations (2.48) and (2.49), we get

$$X_i = \mu_i + P_i + L. \tag{2.50}$$

Here, the components are independent. Assuming $\mu_i + P_i = Z_i$, we get a decomposition of the observations as:

$$X_i = Z_i + L.$$

Now, P_i is independent of μ_i and therefore of θ_i. So, the conditional distribution of $Z_i|\theta_i$ is $(1-\theta_i)N(0,\sigma_\rho^2) + \theta_i N(0, \sigma_\rho^2 + \tau^2)$. Therefore, lemma 1 is proved. □

Proof of Lemma 2 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the underlying probability space. As we defined in Section 2.2, we are working with two different information sets. The first is $\mathcal{F}_{\mathbf{X}} = \sigma(\mathbf{X})$, which represents the sigma field of our observed data. The second is $\mathcal{F}_{\mathbf{X},L} = \sigma(\mathbf{X}, L)$, which is the larger oracle sigma field where we get to see both the data and the latent factor L.

Since $\mathcal{F}_{\mathbf{X},L}$ simply adds the variable L to what we already know, it is clear that we have the nesting property:

$$\mathcal{F}_{\mathbf{X}} \subset \mathcal{F}_{\mathbf{X},L}. \tag{2.51}$$

Now, let $\mathcal{D}_{\mathbf{X}}$ and $\mathcal{D}_{\mathbf{X},L}$ be the spaces of all binary decision rules $\delta = (\delta_1, \ldots, \delta_m)^T$ that are measurable with respect to $\mathcal{F}_{\mathbf{X}}$ and $\mathcal{F}_{\mathbf{X},L}$, respectively. Because of the inclusion in (2.51), any rule we can choose using only $\mathbf{X}$ is automatically a valid rule when we have access to $(\mathbf{X}, L)$. In other words, $\mathcal{D}_{\mathbf{X}} \subset \mathcal{D}_{\mathbf{X},L}$.

Suppose $\delta_0(\mathbf{X}) = (\delta_{01}(\mathbf{X}), \ldots, \delta_{0m}(\mathbf{X})) \in \mathcal{D}_{\mathbf{X}}$ is the set of optimal decision rules that minimizes the Bayes risk (2.6). By definition, we can write this minimum risk as:

$$R_{0m}(\mathbf{X}) = \sum_{i=1}^{m} \mathbb{E}\left[\Delta_{0m}(1-\theta_i)\delta_{0i}(\mathbf{X}) + \Delta_{Am}\theta_i(1-\delta_{0i}(\mathbf{X}))\right]. \tag{2.52}$$

We want to evaluate this expectation by looking at what happens inside the larger oracle information space. By applying the Tower Property of conditional expectations and conditioning the inner layer on $\mathcal{F}_{\mathbf{X},L}$, we can rewrite the expected loss for each coordinate i as:

$$\begin{aligned}&\mathbb{E}\left[\Delta_{0m}(1-\theta_i)\delta_{0i}(\mathbf{X}) + \Delta_{Am}\theta_i(1-\delta_{0i}(\mathbf{X}))\right] \\ &\qquad = \mathbb{E}\left[\mathbb{E}\left[\Delta_{0m}(1-\theta_i)\delta_{0i}(\mathbf{X}) + \Delta_{Am}\theta_i(1-\delta_{0i}(\mathbf{X})) \mid \mathcal{F}_{\mathbf{X},L}\right]\right].\end{aligned}$$

Because $\delta_0(\mathbf{X})$ is constructed entirely from the observed data $\mathbf{X}$, it is $\mathcal{F}_{\mathbf{X}}$-measurable, which means it is also measurable with respect to the larger field $\mathcal{F}_{\mathbf{X},L}$. Therefore, we can treat the decision rule as a known constant within the inner conditional expectation and pull it out:

$$R_{0m}(\mathbf{X}) = \sum_{i=1}^{m} \mathbb{E}\left[\Delta_{0m}\delta_{0i}(\mathbf{X})\mathbb{E}[1-\theta_i \mid \mathcal{F}_{\mathbf{X},L}] + \Delta_{Am}(1-\delta_{0i}(\mathbf{X}))\mathbb{E}[\theta_i \mid \mathcal{F}_{\mathbf{X},L}]\right]. \tag{2.53}$$

Now, let $\delta_0(\mathbf{X}, L) = (\delta_{01}(\mathbf{X}, L), \ldots, \delta_{0m}(\mathbf{X}, L)) \in \mathcal{D}_{\mathbf{X},L}$ be the true oracle Bayes optimal decision rule that we defined in (2.10). By definition, the oracle rule is chosen specifically to minimize this exact posterior conditional risk pointwise for every possible value of our variables.

This means that if we substitute our data-only rule $\delta_0(\mathbf{X})$ into this inner conditional risk expression, it must give a value that is greater than or equal to what we would get using the optimized oracle rule:

$$\begin{aligned}\Delta_{0m}\delta_{0i}(\mathbf{X})\mathbb{E}[1-\theta_i \mid \mathcal{F}_{\mathbf{X},L}] + \Delta_{Am}(1-\delta_{01}(\mathbf{X}))\mathbb{E}[\theta_i \mid \mathcal{F}_{\mathbf{X},L}] \\ \geq \Delta_{0m}\delta_{0i}(\mathbf{X},L)\mathbb{E}[1-\theta_i \mid \mathcal{F}_{\mathbf{X},L}] + \Delta_{Am}(1-\delta_{0i}(\mathbf{X},L))\mathbb{E}[\theta_i \mid \mathcal{F}_{\mathbf{X},L}]. \quad (2.54)\end{aligned}$$

Since this inequality holds pointwise across the entire sample space, taking the outer expectation on both sides will preserve the inequality direction:

$$R_{0m}(\mathbf{X}) \geq \sum_{i=1}^{m} \mathbb{E}\left[\Delta_{0m}\delta_{0i}(\mathbf{X},L)\mathbb{E}[1-\theta_i \mid \mathcal{F}_{\mathbf{X},L}] + \Delta_{Am}(1-\delta_{0i}(\mathbf{X},L))\mathbb{E}[\theta_i \mid \mathcal{F}_{\mathbf{X},L}]\right]. \quad (2.55)$$

Finally, if we use the Tower Property in reverse to collapse the inner conditioning layers back down, we get:

$$R_{0m}(\mathbf{X}) \geq \sum_{i=1}^{m} \mathbb{E}\left[\Delta_{0m}(1-\theta_i)\delta_{0i}(\mathbf{X},L) + \Delta_{Am}\theta_i(1-\delta_{0i}(\mathbf{X},L))\right]. \quad (2.56)$$

The right-hand side of this expression is exactly the definition of the minimal oracle Bayes risk $R_{0m}(\mathbf{X},L)$ when optimizing over the entire set of rules $\mathcal{D}_{\mathbf{X},L}$. Therefore, we can conclude that:

$$R_{0m}(\mathbf{X}) \geq R_{0m}(\mathbf{X},L). \quad (2.57)$$

Further, note that, the data-driven version of the oracle decision rules, i.e, $\hat{\boldsymbol{\delta}} = (\hat{\delta}_{01}, \hat{\delta}_{02}, \ldots, \hat{\delta}_{0m})$ is constructed based on the data $\mathbf{X}$ only. Hence, by definition, since optimal Bayes rule based on the only the data $\mathbf{X}$, i.e., $\boldsymbol{\delta}_{\mathbf{0}}(\mathbf{X}) = (\delta_{01}(\mathbf{X}), \delta_{02}(\mathbf{X}), \ldots, \delta_{0m}(\mathbf{X}))$ minimizes the Bayes risk (2.6) over the space of all $\mathcal{F}_X$-measurable rules, its risk $R_{0m}(\mathbf{X})$ cannot exceed the risk of the data-driven rule $\hat{\boldsymbol{\delta}}(\mathbf{X})$ defined in (2.11). i.e.,

$$R_{0m}(\mathbf{X}) \leq \hat{R}_{0m}(\mathbf{X}, \bar{X}_m). \quad (2.58)$$

Therefore, combining these two properties, namely Lemma 2 and (2.58) gives us the following three-tier inequality for any finite $m \in \mathbb{N}$:

$$R_{0m}(\mathbf{X}, L) \leq R_{0m}(\mathbf{X}) \leq \hat{R}_{0m}(\mathbf{X}, \bar{X}_m). \tag{2.59}$$

This completes the proof of Lemma 2.

Proof of Lemma 3

Proof. Equation (2.33) implies that $c_{\text{Bon}} \to \infty$ as $m \uparrow \infty$. Now,

$$1 - \Phi(c_{\text{Bon}}) = \int_{c_{\text{Bon}}}^{\infty} \phi(z)dz \leq \frac{1}{c_{\text{Bon}}} \int_{c_{\text{Bon}}}^{\infty} z\phi(z)dz = \frac{\phi(c_{\text{Bon}})}{c_{\text{Bon}}}. \tag{2.60}$$

As $c_{\text{Bon}} \to \infty$, the difference between the terms in both sides of the inequality in (2.60) converges to 0. Then we can write

$$1 - \Phi(c_{\text{Bon}}) \simeq \frac{\phi(c_{\text{Bon}})}{c_{\text{Bon}}}, \tag{2.61}$$

where for 2 real sequences $\{A_m\}$ and $\{B_m\}$, $A_m \simeq B_m$ implies that $A_m / B_m \to 1$. So, equation (2.33) implies that,

$$\frac{\phi(c_{\text{Bon}})}{c_{\text{Bon}}} \simeq \frac{\alpha}{2m}. \tag{2.62}$$

Logarithm, being a bijective function, does not change the $\simeq$ relation. Hence we can write

$$\frac{c_{\text{Bon}}^2}{2} \simeq \log m. \tag{2.63}$$

Now, due to (2.35), we can again write $c_{\check{S}} \to \infty$ and therefore the result in equation (2.61) is also applicable here, i.e.,

$$1 - \Phi(c_{\check{S}}) \simeq \frac{\phi(c_{\check{S}})}{c_{\check{S}}},$$

which yields,

$$\frac{\phi(c_{\check{S}})}{c_{\check{S}}} \simeq 1 - \left(1 - \frac{\alpha}{2}\right)^{\frac{1}{m}}.$$

A simple logarithmic transformation implies:

$$-\frac{c_{\check{S}}^2}{2} \simeq \log\left(1-\left(1-\frac{\alpha}{2}\right)^{\frac{1}{m}}\right)$$

Performing an exponential transformation, subtracting both sides from 1 and taking the m-th power we get

$$\left(1-\exp\left(-\frac{c_{\check{S}}^2}{2}\right)\right)^m \simeq \left(1-\frac{\alpha}{2}\right),$$

which can be rewritten as

$$\left[\left(1-\exp\left(-\frac{c_{\check{S}}^2}{2}\right)\right)^{\exp\left(\frac{c_{\check{S}}^2}{2}\right)}\right]^{m\exp\left(-\frac{c_{\check{S}}^2}{2}\right)} \simeq \left(1-\frac{\alpha}{2}\right).$$

It is easy to see that,

$$\exp\left(-m\exp\left(-\frac{c_{\check{S}}^2}{2}\right)\right) \simeq \left(1-\frac{\alpha}{2}\right),$$

i.e.,

$$\frac{c_{\check{S}}^2}{2} \simeq \log m, \tag{2.64}$$

which proves the lemma. □

Proof of Lemma 4

Proof. The proof of lemma 4 follows from Theorem 5.3.1 of Tong, 2012. □

Chapter 3

Large-scale Sequential Multiple Testing with Simultaneous Termination of Sampling

3.1 Data Description and Problem Setup

Suppose we observe m-variate data vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots$ sequentially. Here $\mathbf{X}_n$ denotes the set of observations $(X_{n1}, X_{n2}, \ldots, X_{nm})$ corresponding to m data streams from the n-th sampling unit (e.g., patient). Alternatively, n represents a time point in the sequential scheme of sampling. In this chapter "time" and "sample size" are used interchangeably. This sequence of vectors provide m data streams where the i-th stream is given by the i-th coordinate of the m-variate data vectors as $X_{1i}, X_{2i}, X_{3i}, \ldots$, for $i = 1, \ldots, m$. Our goal is to simultaneously test hypotheses H_{0i} vs H_{1i}, related to the i-th data stream, for $i = 1, \ldots, m$. The vector of binary random variables $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_m)$ determines the true states of the hypotheses such that, if $\theta_i = 0$ then H_{0i} is true, and if $\theta_i = 1$ then H_{1i} is true. Suppose that after collecting n samples in the i-th data stream, a statistic S_{ni} is used to summarize the evidence against null hypothesis H_{0i}. The statistic S_{ni} could be a p-value or some other statistic such as a likelihood ratio based on the first n samples, or in general, some vector-valued function of first n observations $X_{1i}, X_{2i}, \ldots, X_{ni}$. For example, S_{ni} could as well be $(X_{1i}, X_{2i}, \ldots, X_{ni})$ depending on the hypothesis testing problem. We shall make necessary assumptions and construct the test statistics based on the random vector $\mathbf{S}_n = (S_{n1}, S_{n2}, \ldots, S_{nm})$ in section 3.1.1.

To formalize the setup, suppose the data vectors $\mathbf{X}_1, \mathbf{X}_2, \ldots$ are defined on a measurable space (Ω, σ), and the sequence $\{\mathbf{S}_n\}$ is defined on an increasing sequence of sigma fields $\{\sigma_n\}$ such that $\cup_{n=1}^{\infty} \sigma_n \subseteq \sigma$. For example, the natural choice for σ_n is the σ-algebra generated by $\mathbf{X}_1, \ldots, \mathbf{X}_n$. A sequential test for testing H_{0i} vs H_{1i}, for all $i = 1, \ldots, m$, is defined as the pair $(T, \boldsymbol{\delta})$ where T is a $\{\sigma_n\}$-stopping time and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m) \in \{0,1\}^m$ is a decision rule, defined on σ-field σ_T, such that δ_i equals 1 or 0 depending on whether H_{0i} is rejected or accepted. The dependency of $\boldsymbol{\delta}$ on T is suppressed for simplicity of notation. According to Table 1.1, $(T, \boldsymbol{\delta})$ makes $V = \sum_{i=1}^{m}(1-\theta_i)\delta_i$ false rejections among $R = \sum_{i=1}^{m} \delta_i$ rejections and $W = \sum_{i=1}^{m} \theta_i(1-\delta_i)$ false acceptances among $m - R = \sum_{i=1}^{m}(1-\delta_i)$ acceptances. The error metrics of our interests are FDR, FNR, mFDR and mFNR as defined in Section 1.1.2 of Chapter 1.

The primary objective of this chapter is to develop sequential tests $(T, \boldsymbol{\delta})$ such that FDR $\leq \alpha$ and FNR $\leq \beta$ for some prefixed levels $\alpha, \beta \in (0,1)$. Interestingly, it will turn out that the proposed tests controlling FDR and FNR also guarantee mFDR $\leq \alpha$ and mFNR $\leq \beta$. Among the existing sequential multiple tests, He and Bartroff (2021) and Bartroff and Song (2020) provide sequential multiple tests $(T, \boldsymbol{\delta})$ such that FDR $\leq \alpha$ and FNR $\leq \beta$. However, the premise considered in Bartroff and Song (2020) is different from that of this chapter, as they allow different coordinates to stop at different time points, and will be examined in the next chapter. The sequential tests in He and Bartroff (2021) consider stopping all data streams at some time T, and hence, are applicable to the same sequential testing framework as ours. We will compare their asymptotically optimal test with ours in section 3.2.3 and chapter 5, and show that the attained FDR and FNR levels of our methods are nearly same as the desired levels α and β respectively, providing huge savings in the average sample size as the number of hypotheses becomes large.

3.1.1 Oracle Test Statistics

Selection of an appropriate test statistic is crucial for the performance of any large-scale multiple testing procedure. Traditional fixed-sample-size multiple tests are based on p-values where hypotheses are ranked according to the ordered p-values and subsequently rejected (or accepted) based on some thresholds. Sun and Cai

(2007a) showed that p-value-based approaches do not lead to efficient multiple testing procedures in general, especially when the number of hypotheses is large. Treating multiple testing of m hypotheses as a compound decision problem, such inefficiency may be explained by the classical work of Robbins (1951) where it is argued that simple decision rules (such as p-value based marginal rules) are usually not as good as compound decision rules that can pool information from different tests. To deal with large-scale multiple testing efficiently, many authors including Sun and Cai (2007a), Sarkar, Zhou, and Ghosh (2008), Sun and Cai (2009), Sun et al. (2015) considered some oracle test statistics, which, depending on their context, essentially represented the posterior probabilities of null hypotheses given the data. Sun and Cai (2009) and Sun et al. (2015) showed that such an oracle statistic can minimize classification risk considered in their respective problem settings. Under a decision theoretic framework, Sun and Cai (2007a) developed tests based on this oracle test statistic and proved optimality in the sense of minimizing mFNR subject to a constraint on mFDR. Motivated by such optimality properties, we consider the posterior probabilities of null hypotheses given summary statistics $\mathbf{S}_n$, i.e.,

$$t_{ni} = \mathbb{P}(\theta_i = 0 \mid \mathbf{S_n}), \text{ for } i = 1, \ldots, m, \tag{3.1}$$

as the oracle test statistics for our sequential testing problem. If the conditional distribution of $\mathbf{S}_n$ given $\theta_i = j$ is $f_n^*(\mathbf{S}_n|\theta_i = j)$, for $j = 0, 1$, then

$$t_{ni} = \frac{\mathbb{P}(\theta_i = 0) f_n^*(\mathbf{S}_n|\theta_i = 0)}{\mathbb{P}(\theta_i = 0) f_n^*(\mathbf{S}_n|\theta_i = 0) + \mathbb{P}(\theta_i = 1) f_n^*(\mathbf{S}_n|\theta_i = 1)}.$$

In general, $\mathbb{P}(\theta_i = j)$ and $f_n^*(\mathbf{S}_n|\theta_i = j)$ are not known in practice, and thus, t_{ni} is referred to as an "oracle" statistic. Direct estimation of $f_n^*(\mathbf{S}_n|\theta_i = j)$, for $j = 0, 1$, is difficult without further model assumptions, and hence, different special cases of the oracle test statistics $\{t_{ni}\}$ under different assumptions are considered in the literature. For example, Sun and Cai (2009) refers $\{t_{ni}\}$ as the local index of significance (LIS) and estimates them under a hidden Markov model assumption where $\{\theta_i\}_{i=1}^m$ is a stationary, irreducible, aperiodic Markov chain and data streams are independent conditional on $\boldsymbol{\theta}$. Sun and Cai (2007a) considered an independent two-group

mixture model assumption which is as follows.

Assumption 3. *Random variables* $\theta_1, \ldots, \theta_m \overset{\text{i.i.d.}}{\sim} Bernoulli(1-\pi_0)$, *conditional on* $\boldsymbol{\theta}$, $S_{n1}, \ldots, S_{nm}$ *are independent, and* $S_{ni}|\boldsymbol{\theta} \sim (1-\theta_i)f_{0n} + \theta_i f_{1n}$, *for* $i = 1, \ldots, m$, *where* f_{0n} *and* f_{1n} *represent the null and alternative densities respectively with respect to some* σ*-finite measure.*

Under this assumption, $S_{n1}, \ldots, S_{nm}$ are i.i.d. samples from the mixture density $f_n = \pi_0 f_{0n} + (1-\pi_0)f_{1n}$, and the oracle statistic t_{ni} becomes

$$t_{ni} = \mathbb{P}(\theta_i = 0|S_{ni}) = \frac{\pi_0 f_{0n}(S_{ni})}{f_n(S_{ni})} \quad \text{for } i = 1, \ldots, m, \tag{3.2}$$

which is referred to as the Lfdr statistic by Efron et al. (2001) and Efron (2004). In recent years, numerous fixed-sample-size multiple tests controlling FDR (or some of its variants such as the mFDR) have been developed based on this Lfdr statistic. See, for example, Sun and Cai (2007a) and Sun and Cai (2009), Sun et al. (2015), Sarkar and Zhao (2022) and Basu et al. (2018). For sequentially observed data, our idea is to adopt these oracle test statistics to construct multiple tests for simultaneous control of FDR and FNR.

3.2 Oracle Test Procedure With Exact Error Control

In this section, we develop an adaptive test procedure (referred to as an oracle test) based on the oracle statistics $\{t_{ni}\}_{i=1}^m$ given in 3.1, so that exact control of FDR and FNR (or mFDR and mFNR) is achieved for testing multiple hypotheses for streaming data.

3.2.1 Oracle Intersection Test

We first discuss a possible strategy to control FDR at some prefixed level α for fixed sample size n. Since a smaller value of t_{ni} provides stronger evidence against H_{0i}, it is natural to order $\{t_{ni}\}_{i=1}^m$ from the smallest to the largest as $t_n^{(1)} \leq \cdots \leq t_n^{(m)}$ and reject, say, r_n many nulls corresponding to $t_n^{(1)}, \ldots, t_n^{(r_n)}$. The choice of r_n for FDR control can be obtained from the following calculations. If $\theta = (\theta_1, \ldots, \theta_m)$, FDR can

be expressed as

$$\begin{aligned}\text{FDR} &= \mathbb{E}_{S_n}\mathbb{E}_{\theta|S_n}\left[\frac{\sum(1-\theta_i)\,\delta_i}{(\sum\delta_i)\vee 1}\right] = \mathbb{E}_{S_n}\left[\frac{\sum\delta_i\,\mathbb{P}(\theta_i=0\mid S_n)}{(\sum\delta_i)\vee 1}\right]\\ &= \mathbb{E}_{S_n}\left[\frac{\sum\delta_i\,t_{ni}}{(\sum\delta_i)\vee 1}\right] = \mathbb{E}_{S_n}\left[\frac{1}{r_n\vee 1}\sum_{i=1}^{r_n}t_n^{(i)}\right].\end{aligned}$$

If we consider the choice of r_n as

$$r_n = \begin{cases}\max\left\{1\le q\le m : \frac{1}{q}\sum_{i=1}^{q}t_n^{(i)}\le\alpha\right\} & \text{if } t_n^{(1)}\le\alpha,\\ 0 & \text{if } t_n^{(1)}>\alpha,\end{cases} \tag{3.3}$$

then the attained FDR $\le \alpha$ for any finite m and n. In particular, if $t_n^{(1)} > \alpha$, we accept all nulls, and hence, FDR $= 0$. Note that p-value based step-up procedures (e.g., Benjamini and Hochberg (1995), Benjamini and Yekutieli (2001)) use some stringent inequalities such as the Simes inequality to guarantee exact control of FDR whereas the above rule, for fixed n, does not require any specific inequality and attains FDR $\approx \alpha$ when $\frac{1}{r_n}\sum_{i=1}^{r_n}t_n^{(i)} \approx \alpha$ even if m is extremely large.

A similar approach can be adopted to control FNR at some prefixed level β. Since a larger value of t_{ni} favors the null H_{0i}, it is natural to accept, say, a_n many nulls corresponding to $t_n^{(m-a_n+1)},\ldots,t_n^{(m)}$, where a_n may be chosen from the following calculations. Since FNR is expressed as

$$\text{FNR} = \mathbb{E}_{S_n}\left[\frac{\sum(1-\delta_i)\,(1-t_{ni})}{\sum(1-\delta_i)\vee 1}\right] = \mathbb{E}_{S_n}\left[\frac{1}{a_n\vee 1}\sum_{i=1}^{a_n}\left(1-t_n^{(m-i+1)}\right)\right],$$

the choice of

$$a_n = \begin{cases}\max\left\{1\le q\le m : \frac{1}{q}\sum_{i=1}^{q}t_n^{(m-i+1)}\ge 1-\beta\right\} & \text{if } t_n^{(m)}\ge 1-\beta,\\ 0 & \text{if } t_n^{(m)}<1-\beta,\end{cases} \tag{3.4}$$

yields the attained FNR $\le \beta$ for any finite m and n. For $t_n^{(m)} < 1-\beta$, we reject all nulls, yielding FNR $= 0$. For simultaneous control of FDR and FNR, we need to combine the above ideas of rejection and acceptance of nulls

at any interim stage n of the sequential sampling.

Recall from section 3.1 that we observe the sequence of summary statistics $\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots$, and after computing $\boldsymbol{S}_n$ based on $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, the oracle statistics $t_{n1}, \ldots, t_{nm}$ are computed and ordered from the smallest to the largest. At this n-th stage of sampling, we identify r_n many potential false nulls corresponding to $t_n^{(1)}, \ldots, t_n^{(r_n)}$ and a_n many potential true nulls corresponding to $t_n^{(m-a_n+1)}, \ldots, t_n^{(m)}$. If $r_n + a_n < m$, then there are $(m - r_n - a_n)$ many nulls yet to be decided as potential true or false nulls based on the first n observations, and hence, $(n+1)$-th observation must be collected for further evidence in favor or against these nulls. A natural stopping rule would be to stop sampling the first time when all nulls get a decision, that is, to stop sampling at

$$\tau' = \inf\{n \geq k : r_n + a_n \geq m\}, \tag{3.5}$$

where $k \in \mathbb{N}$ is some prefixed pilot sample size. Note that if $r_{\tau'} + a_{\tau'} = m$, we reject exactly $r_{\tau'}$ nulls, and hence, have a unique decision rule at stopping. However, if $r_{\tau'} + a_{\tau'} > m$, we can reject s many nulls, for any choice of s in $\{m - a_{\tau'}, \ldots, r_{\tau'}\}$, to ensure FDR and FNR control at α and β levels, respectively. For $s = r_{\tau'}$, the attained FDR is expected to be close to the target level α whereas the attained FNR may be somewhat smaller than the target level β. For $s = m - a_{\tau'}$, we expect to observe the opposite phenomena. In simulation study, we implement $s = r_{\tau'}$ to get more rejections and higher $1 - \text{FNR}$.

Existing sequential multiple tests, e.g., De and Baron (2012a), De and Baron (2012b), and De and Baron (2015), He and Bartroff (2021), and Song and Fellouris (2019), stop sampling when all test statistics under considerations cross certain stopping boundaries which typically do not depend on the sample size n. Interestingly, our stopping criteria in (3.5) can be alternatively described through some *adaptive stopping boundaries* that, unlike existing methods, depend on the sample size n. Let us define the lower and upper boundaries as

$$t_n^L = \begin{cases} t_n^{(r_n+1)} & \text{if } r_n < m, \\ 1 & \text{if } r_n = m, \end{cases} \quad \text{and} \quad t_n^U = \begin{cases} t_n^{(m-a_n)} & \text{if } a_n < m, \\ 0 & \text{if } a_n = m. \end{cases} \tag{3.6}$$

where a_n and r_n are defined in (3.4) and (3.3) respectively. At the n-th stage of sampling, we compare t_{ni} with these lower and upper boundaries with the following possibilities. If $t_{ni} < t_n^L$, then H_{0i} is considered to be a potentially false null, if $t_{ni} > t_n^U$, then H_{0i} is considered to be a potentially true null, and if $t_n^L \leq t_{ni} \leq t_n^U$, then H_{0i} is considered to be "undecided" at that stage. Thus, based on these boundaries, sampling is stopped at

$$\tau = \inf\left\{n \geq k : \, t_n^L > t_n^U\right\}. \tag{3.7}$$

The following lemma provides a comparison between stopping times τ' and τ.

Lemma 5. *(i) If $\mathbb{P}(t_{ni} = t_{nj}) = 0$ for all $i, j \in \{1, 2, \ldots, m\}$, $i \neq j$, then $r_n + a_n \geq m$ if and only if $t_n^L > t_n^U$.*
(ii) If $\mathbb{P}(t_{ni} = t_{nj}) > 0$ for some $i, j \in \{1, 2, \ldots, m\}$, $i \neq j$, then $t_n^L > t_n^U$ implies $r_n + a_n \geq m$.

A sufficient condition for avoiding ties among $\{t_{ni}\}_{i=1}^m$, i.e., ensuring $\mathbb{P}(t_{ni} = t_{nj}) = 0$ for all (i, j), is that summary statistics $\{S_{ni}\}_{i=1}^m$ are continuous. Thus, Lemma 5 leads to the following result.

Corollary 2. *Stopping times $\tau' \leq \tau$ with probability 1 where equality holds whenever $\mathbf{S}_n$ is a continuous random variable.*

Although in most applications of this article, we shall consider testing problems with continuous test statistics leading to $\tau' = \tau$, due to analytic advantages, we shall work with τ in the rest of this article. To formally describe our oracle test, let us fix $\alpha, \beta \in (0, 1)$ and a pilot sample size $k \in \mathbb{N}$. The oracle test, referred to as the **Oracle Intersection (OI)** test, for simultaneous control of FDR and FNR involves the following steps.

1. Collect $n = k$ observations and compute $t_{ni} = \mathbb{P}(\theta_i = 0 \,|\, \mathbf{S}_n)$ for $i = 1, 2, \ldots, m$.
2. Order $t_{n1}, \ldots, t_{nm}$ as $t_n^{(1)} \leq t_n^{(2)} \leq \cdots \leq t_n^{(m)}$ and compute r_n and a_n given in (3.3) and (3.4).
3. If $t_n^L \leq t_n^U$, collect $(n+1)$-th observation and compute (r_{n+1}, a_{n+1}). Else, stop sampling and set the final sample size $\tau = n$ as in (3.7).

4. After sampling terminates at τ, the decision rule $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)$ is such that $\delta_i = 1$ or 0 depending on whether $i \in \mathfrak{R}$ or $i \notin \mathfrak{R}$ where $\mathfrak{R} = \{1 \leq i \leq m : t_{\tau i} \leq t_\tau^{(s)}\}$, for any $s \in \{m - a_\tau, \ldots, r_\tau\}$.

Note that, we implement the OI test (e.g., in Figure 3.1) or its data-driven version using $s = r_\tau$ to get more rejections and higher power in terms of $1 - \text{FNR}$. The sampling scheme of the OI test is illustrated through an example in Figure 3.1 where $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \overset{\text{i.i.d.}}{\sim} N_m(\mathbf{0}, \sigma^2 \mathrm{I}_m)$ and $H_{0i} : \sigma = 1$ vs $H_{1i} : \sigma = 1.15$ is tested for $i = 1, 2, \ldots, m$. Here, $S_{ni} = (X_{i1}, \ldots, X_{in})$, $m = 500$, $\pi_0 = 0.5$, $\alpha = 0.05$, $\beta = 0.1$ and $k = 50$. Under Assumption 3, $t_{ni} = \pi_0 \prod_{j=1}^{n} \phi(X_{ij}) / \{\pi_0 \prod_{j=1}^{n} \phi(X_{ij}) + (1 - \pi_0) \prod_{j=1}^{n} \phi(X_{ij}/1.2)\}$ where ϕ denotes the density function of a $N(0,1)$ variable. We plot the stopping boundaries t_n^L, t_n^U and test statistics $\{t_{ni}\}_{i=1}^n$ at $n = 50, 100, 150$ and 218. The points plotted within $[t_n^L, t_n^U]$ represent undecided nulls at interim stages $n = 50, 100, 150$, and at stopping $\tau = 218$, $t_\tau^L > t_\tau^U$, providing accept/reject decisions for all m null hypotheses.

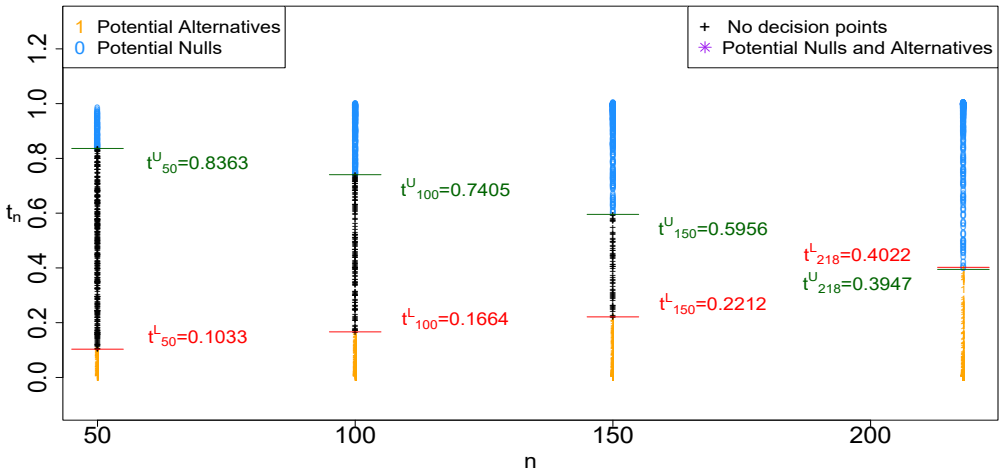

FIGURE 3.1: Sampling scheme of an OI test via adaptive stopping boundaries

To ensure sampling terminates in the OI test, we make the following mild consistency-type assumption.

Assumption 4. *As* $n \to \infty$, $t_{ni} \xrightarrow{\mathbb{P}} 1$ *under* H_{0i} *and* $t_{ni} \xrightarrow{\mathbb{P}} 0$ *under* H_{1i} *for all* $i = 1, 2, \ldots, m$.

Assumption 4 ensures that as we observe more data, the test statistics provide more evidence in favor of the null when the null is actually true and against the null when the alternative is actually true. Such an assumption holds trivially for simple-vs-simple testing with i.i.d. data sequence. Note that if $X_{i1}, \ldots, X_{in} \overset{\text{i.i.d.}}{\sim} f_0$ or f_1 under null or alternative with known densities f_0 and f_1, then $t_{ni} = [1 + ((1-\pi_0)/\pi_0)\prod_{j=1}^{n} f_1(X_{ij})/f_0(X_{ij})]^{-1} \xrightarrow{a.s.} 1$ under null since $\mathbb{E}_{f_0}[\ln f_1(X_{ij})/f_0(X_{ij})] < 0$ and $\sum_{j=1}^{n} \ln f_1(X_{ij})/f_0(X_{ij}) \xrightarrow{a.s.} -\infty$ by Jensen's inequality and the strong law of large numbers. Similarly, $t_{ni} \xrightarrow{a.s.} 0$ under alternative. The following theorem shows that the OI test terminates sampling in finite time with probability 1 and guarantees simultaneous exact control of FDR and FNR.

Theorem 9. *If Assumption 4 holds, then the OI test given by (τ, δ) yields*

(i) $\mathbb{P}(\tau < \infty) = 1$,

(ii) $FDR \leq \alpha$ *and* $FNR \leq \beta$ *for any prefixed* $\alpha, \beta \in (0,1)$,

(iii) $mFDR \leq \alpha$ *and* $mFNR \leq \beta$ *for any prefixed* $\alpha, \beta \in (0,1)$.

3.2.2 Asymptotic Properties of the Oracle Stopping Time

We investigate the asymptotic properties of the stopping time τ as the number of hypotheses m tends to infinity. Suppose, based on some fixed sample size n, we reject H_{0i} when $t_{ni} \leq t$ for some prefixed cut-off $t \in [0,1]$, i.e., $\delta_i = \mathbb{1}\{t_{ni} \leq t\}$ where $\mathbb{1}\{A\}$ denotes the indicator variable of an event A. Noting that under Assumption 3, $t_{n1}, \ldots, t_{nm}$ are i.i.d. variables, the expected false discoveries and the expected discoveries become $\mathbb{E}[V] = m\pi_0\mathbb{P}_0(t_{n1} \leq t)$ and $\mathbb{E}[R] = m\,\mathbb{P}(t_{n1} \leq t)$ respectively, where $\mathbb{P}(t_{n1} \leq t) = \pi_0\mathbb{P}_0(t_{n1} \leq t) + (1-\pi_0)\mathbb{P}_1(t_{n1} \leq t)$, $\mathbb{P}_0$ and $\mathbb{P}_1$ are the probability measures under the null and the alternative respectively. The expected false non-discoveries and the expected non-discoveries become $\mathbb{E}[W] = m(1-\pi_0)\mathbb{P}_1(t_{n1} > t)$ and $m - \mathbb{E}[R] = m\,\mathbb{P}(t_{n1} > t)$ respectively. Thus, the attained mFDR and mFNR at prefixed cut-off $t \in [0,1]$ are

$$\mathcal{Q}_n(t) = \begin{cases} \frac{\pi_0\mathbb{P}_0(t_{n1} \leq t)}{\mathbb{P}(t_{n1} \leq t)} & \text{if } t \neq 0 \\ 0 & \text{if } t = 0 \end{cases} \quad \text{and} \quad \mathcal{Q}'_n(t) = \begin{cases} \frac{(1-\pi_0)\mathbb{P}_1(t_{n1} > t)}{\mathbb{P}(t_{n1} > t)} & \text{if } t \neq 1 \\ 0 & \text{if } t = 1 \end{cases}, \tag{3.8}$$

respectively. It is shown in the supplementary material that $\mathcal{Q}_n(t)$ is nondecreasing and $\mathcal{Q}'_n(t)$ is nonincreasing in $t \in [0,1]$. Thus, the best cut-offs for mFDR control at α and mFNR control at β are

$$\mathcal{T}_n^L = \sup\{0 \le t \le 1 : \mathcal{Q}_n(t) \le \alpha\} \text{ and } \mathcal{T}_n^U = \inf\{0 \le t \le 1 : \mathcal{Q}'_n(t) \le \beta\},$$

respectively. The following Lemma establishes the limits of lower and upper stopping boundaries as $m \to \infty$.

Lemma 6. *If $\mathbf{S}_n$ is continuous and Assumptions 3 and 4 hold, then for fixed $n \in \mathbb{N}$, $t_n^L \xrightarrow{\mathbb{P}} \mathcal{T}_n^L$ and $t_n^U \xrightarrow{\mathbb{P}} \mathcal{T}_n^U$ as $m \to \infty$.*

Lemma 6 suggests that, for large m, the lower stopping boundary t_n^L is close to the theoretical cut-off $\mathcal{T}_n^L$ for mFDR control, and t_n^U is close to the theoretical cut-off $\mathcal{T}_n^U$ for mFNR control. Simultaneous use of t_n^L and t_n^U in (3.7) ensures exact control of both mFDR and mFNR given in part (iii) of Theorem 9. The following Lemma shows that the non-stochastic cut-offs $\mathcal{T}_n^L$ and $\mathcal{T}_n^U$ tend to 1 and 0 respectively as $n \to \infty$.

Lemma 7. *Under Assumptions 3 and 4, $\mathcal{T}_n^L \to 1$ and $\mathcal{T}_n^U \to 0$ as $n \to \infty$.*

Lemma 7 indicates that the set $\{n \in \mathbb{N} : \mathcal{T}_n^L > \mathcal{T}_n^U\}$ is nonempty, and hence, $n_0 = \inf\{n \in \mathbb{N} : \mathcal{T}_n^L > \mathcal{T}_n^U\}$ is a finite integer which does not depend on m. A necessary and sufficient condition for $\mathcal{T}_n^L = \mathcal{T}_n^U$ to hold for some $n < n_0$ is quite technical and is provided in Lemma 11. However, empirical evidences from a variety of simulation studies for continuous $\mathbf{S}_n$ suggest that these conditions are not very likely to hold in practice (at least for the simulation setups considered here), and hence, the assumption that $\mathcal{T}_n^L < \mathcal{T}_n^U$ for all $n < n_0$ is not so restrictive.

Theorem 10. *Suppose that the proportion of true nulls $\pi_0 \in (\alpha, 1-\beta)$ for prefixed $\alpha, \beta \in (0,1)$. If $\mathbf{S}_n$ is continuous and Assumptions 3 and 4 hold, then the stopping time τ in (3.7) satisfies the following properties.*

(i) $\mathbb{P}(\tau \le n_0) \to 1$ as $m \to \infty$.

(ii) Further, if $\mathcal{T}_n^L \ne \mathcal{T}_n^U$ for all $n < n_0$, then $\mathbb{P}(\tau = n_0) \to 1$ as $m \to \infty$.

This result is noteworthy since traditional sequential multiple tests typically uses fixed stopping boundaries, and as a result, stopping times are not bounded as $m \to$

∞. For instance, Theorem 11 proves that the stopping time of the asymptotically optimal method of He and Bartroff, 2021 is not bounded in probability. One of the attractive features of the proposed test is that it adopts stopping boundaries t_n^L and t_n^U such that the continue sampling region $[t_n^L, t_n^U]$ eventually shrinks with larger sample size n (see Figure 3.1). The adaptiveness of the stopping boundaries enables termination of sampling with finite n_0 observations even when the number of hypotheses tends to infinity.

Remark 1. *The condition $\alpha < \pi_0 < 1-\beta$, in Theorem 10, is imposed to avoid the following triviality regarding the stopping time τ. Note that the attained mFDR and mFNR of the OI test are given by $\mathcal{Q}_n(t)$ and $\mathcal{Q}'_n(t)$ in (3.8). Since $\mathcal{Q}_n(t)$ is non-decreasing in $t \in [0,1]$, the attained mFDR $\mathcal{Q}_n(t) \leq \pi_0 \leq \alpha$ for all $t \in [0,1]$ whenever $\pi_0 \leq \alpha$, implying that if we simply reject all nulls irrespective of the data observed, the mFDR will be controlled at level α. Thus, the stopping time would occur at the pilot sample size k. Similarly, since $\mathcal{Q}'_n(t)$ is non-increasing in $t \in [0,1]$, the attained mFNR $\mathcal{Q}'_n(t) \leq 1-\pi_0 \leq \beta$ for all $t \in [0,1]$ whenever $\pi_0 \geq 1-\beta$. Thus, if we simply accept all nulls irrespective of the data collected, the mFNR will be controlled at level β terminating sampling at the pilot sample size k. Therefore, the assumption $\alpha < \pi_0 < 1-\beta$ is not restrictive for investigating the asymptotic property of the stopping time τ.*

3.2.3 Comparison with Gap rule

We compare the stopping rule τ of the OI test with that of the asymptotically optimal stopping rule, originally proposed by Song and Fellouris (2017), and investigated further in He and Bartroff (2021) in the context of simultaneous control of FDR and FNR. To the best of our knowledge, their procedure is the only available multiple testing procedure controlling both FDR and FNR under the framework of sampling scheme with simultaneous termination discussed in section 1.3.2 of chapter 1.

The setup considered in He and Bartroff (2021) involves m independent data streams with the j-th stream as $X^{(j)} = (X_{1j}, X_{2j}, \dots)$ consisted of i.i.d. observations. To compare both rules under a similar setup, let us make the following assumptions.

Assumption 5. *Suppose that the distribution of the j-th data stream is determined by a binary variable θ_j, for $j = 1, \dots, m$, and $\theta_1, \dots, \theta_m \overset{\text{i.i.d.}}{\sim} Bernoulli(1-\pi_0)$. Given*

$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$, $X^{(1)}, \ldots, X^{(m)}$ are independent, and $X_{1j}, \ldots, X_{nj} \overset{\text{i.i.d.}}{\sim} (1-\theta_j) f_0 + \theta_j f_1$ $\forall\, n \in \mathbb{N}$ and $j \in \{1, \ldots, m\}$, where f_0 and f_1 are known null and alternative densities, respectively.

He and Bartroff (2021) considered testing $H_{0j} : \theta_j = 0$ vs $H_{1j} : \theta_j = 1$ based on the j-th data stream, for $j = 1, \ldots, m$, and assumed that the number of true alternative hypotheses K (hence, $\pi_0 = 1 - (K/m)$) is known. Suppose $\Lambda_{nj} := \sum_{i=1}^{n} \log[f_1(X_{ij})/f_0(X_{ij})]$ denotes the log-likelihood ratio based on n observations from the j-th stream and $\Lambda_n^{(1)} \geq \cdots \geq \Lambda_n^{(m)}$ are the corresponding ordered log-likelihood ratios. Note that, in this setup, Λ_{nj} tends to $-\infty$ and ∞ under H_{0j} and H_{1j} respectively, which ensure that Assumption 4 holds with $S_{nj} = (X_{1j}, \ldots, X_{nj})$. For prefixed $\alpha, \beta \in (0,1)$, their suggested stopping rule, referred to as the "Gap-ao" stopping rule, is given as

$$T_G^* = \inf \left\{ n \in \mathbb{N} : \Lambda_n^{(K)} - \Lambda_n^{(K+1)} \geq \log \frac{K(m-K)}{\alpha \wedge \beta} \right\}, \tag{3.9}$$

where $\alpha \wedge \beta := \min\{\alpha, \beta\}$. At stopping time T_G^*, exactly K nulls corresponding to $\Lambda_{T_G^*}^{(1)}, \ldots, \Lambda_{T_G^*}^{(K)}$ are rejected. Theorem 3.1 of He and Bartroff (2021) claims that (i) Gap-ao rule controls both FDR and FNR at levels α and β respectively, and (ii) among all tests that control FDR and FNR at levels α and β respectively, the Gap-ao rule in (3.9) is asymptotically optimal in the sense of yielding the smallest asymptotic expected sample size as $\alpha \wedge \beta \to 0$. However, the asymptotic behavior of the stopping time T_G^* was not studied as $m \to \infty$ for fixed α and β. The following Theorem establishes that, under appropriate assumptions, the sample size τ of the OI test is asymptotically better than the sample size T_G^* of the Gap-ao rule as $m \to \infty$.

Theorem 11. *For prefixed $\alpha, \beta \in (0,1)$, let T_G^* be the stopping time given in Theorem 3.1 of He and Bartroff (2021) and τ be the stopping time of the OI test in (3.7). If Assumption 5 holds, then*

$$\lim_{m \to \infty} \mathbb{P}(T_G^* > L) = 1 \quad \textit{for all } L \in \mathbb{N}. \tag{3.10}$$

Further, under the assumptions of Theorem 10, $\tau / T_G^* \xrightarrow{\mathbb{P}} 0$ as $m \to \infty$.

The average sample numbers (ASN) of the OI test and the Gap-ao rule are compared for some simulated datasets (discussed in chapter 5) in Figure 3.2 which provides a numerical validation of the claims of Theorem 11.

3.3 Data-driven Test With Asymptotic Error Control

To implement the oracle test in practice, it is necessary to estimate the oracle test statistic t_{ni} in 3.1. Since direct estimation of $\mathbb{P}(\theta_i = 0)$ and $f_n^*(\boldsymbol{S}_n|\theta_i = 0)$ in t_{ni} is difficult without further model assumptions, we resort to the two-group mixture model assumption (Assumption 3) for estimation of t_{ni} in the rest of this article. Note that, under Assumption 3, $t_{ni} = \pi_0 f_{0n}(S_{ni})/f_n(S_{ni})$ turns out to be the Lfdr statistic where the true proportion π_0 of nulls, the null density f_{0n} and the mixture density f_n may be unknown. For estimation of Lfdr, number of methods have been proposed in the literature. See, for example, Sun and Cai (2007a), Cai and Jin (2010) and Efron (2012).

In many applications, p-value, say p_{ni}, is used as the summary statistic following $U(0,1)$ under H_{0i} and the transformed summary statistic $S_{ni} = \Phi^{-1}(p_{ni})$, referred to as a z-score, follows $N(0,1)$ where Φ is the CDF of a $N(0,1)$ variable. Sun and Cai (2007a) discusses number of advantages of working with z-scores rather than p-values. In light of this, we take the summary statistic $\boldsymbol{S}_n$ to be some z-score vector (e.g., $\Phi^{-1}(p_{n1}), \ldots, \Phi^{-1}(p_{nm})$) for practical convenience and obtain the Lfdr statistics under Assumption 3 as

$$t_{ni} = \mathbb{P}(\theta_i = 0 \mid S_{ni}) = \frac{\pi_0\, \phi(S_{ni})}{f_n(S_{ni})} \text{ for } i = 1, 2, \ldots, m, \tag{3.11}$$

where ϕ is the density of $N(0,1)$. Note that, to estimate t_{ni}, only π_0 and the mixture density f_n or the alternative density f_{1n} need to be estimated.

The problem of estimation of π_0 is investigated by many authors including Storey, Taylor, and Siegmund (2004), Benjamini, Krieger, and Yekutieli (2006) and Cai and Jin (2010). In this article, we use the estimator $\widehat{\pi}_0$ of π_0 proposed by Cai and Jin (2010) whenever S_{ni} is continuous. Theorem 3.1 of Cai and Jin (2010) implies that their proposed $\widehat{\pi}_0$ is a consistent estimator of π_0 as $m \to \infty$. The mixture density f_n

is estimated by kernel density estimator $\widehat{f}_n$ which is also a consistent estimator of f_n under Assumption 3. If S_{ni} is discrete, π_0 is estimated using the method proposed in Storey, Taylor, and Siegmund (2004) and f_n is estimated by a kernel density estimator. By plugging in $\widehat{\pi}_0$ and $\widehat{f}_n$ in (3.11), we obtain the estimator $\widehat{t}_{ni}$ of t_{ni} for all $i = 1, 2, \ldots, m$.

The data-driven sequential multiple test, referred to as the **Data-driven Intersection (DI)** test, is obtained by replacing $\{t_{ni}\}_{i=1}^m$ by $\{\widehat{t}_{ni}\}_{i=1}^m$ in the OI test described in section 3.2. At any interim stage $n(\geq k)$ of the DI test, order $\{\widehat{t}_{ni}\}_{i=1}^m$ as $\widehat{t}_n^{(1)} \leq \cdots \leq \widehat{t}_n^{(m)}$ and compute the number of potential alternative $\widehat{r}_n$ and null $\widehat{a}_n$ by replacing $\{t_n^{(i)}\}_{i=1}^m$ by $\{\widehat{t}_n^{(i)}\}_{i=1}^m$ in (3.3) and (3.4) respectively. The DI test terminates sampling at $n = T$ where

$$T = \inf\left\{n \geq k : \widehat{t}_n^L > \widehat{t}_n^U\right\}, \tag{3.12}$$

and the adaptive lower and the upper boundaries are

$$\widehat{t}_n^L = \begin{cases} \widehat{t}_n^{(\widehat{r}_n+1)} & \text{if } \widehat{r}_n < m, \\ 1 & \text{if } \widehat{r}_n = m, \end{cases} \quad \text{and} \quad \widehat{t}_n^U = \begin{cases} \widehat{t}_n^{(m-\widehat{a}_n)} & \text{if } \widehat{a}_n < m, \\ 0 & \text{if } \widehat{a}_n = m, \end{cases} \tag{3.13}$$

respectively. At stopping time T, the decision rule $\widehat{\boldsymbol{\delta}} = (\widehat{\delta}_1, \ldots, \widehat{\delta}_m)$ is such that $\widehat{\delta}_i = \mathbb{1}_{\{i \in \widehat{\mathfrak{R}}\}}$ where $\widehat{\mathfrak{R}} = \left\{1 \leq i \leq m : \widehat{t}_{Ti} \leq \widehat{t}_T^{(s)}\right\}$, for any $s \in \{m - \widehat{a}_T, \ldots, \widehat{r}_T\}$.

Remark 2. *If we consider the stopping time $T' = \inf\{n \geq k : \widehat{r}_n + \widehat{a}_n \geq m\}$ in the spirit of τ', then a conclusion similar to Lemma 5 and Corollary 2 can be drawn. If $\mathbb{P}(\widehat{t}_{ni} = \widehat{t}_{nj}) = 0$ for all $i, j \in \{1, \ldots, m\}$, $i \neq j$, then $\widehat{r}_n + \widehat{a}_n \geq m$ if and only if $\widehat{t}_n^L > \widehat{t}_n^U$. If $\mathbb{P}(\widehat{t}_{ni} = \widehat{t}_{nj}) > 0$ for some $i, j \in \{1, \ldots, m\}$, $i \neq j$, then $\widehat{t}_n^L > \widehat{t}_n^U$ implies $\widehat{r}_n + \widehat{a}_n \geq m$. Therefore, $T' \leq T$ with probability 1 where equality holds whenever $\mathbf{S}_n$ is continuous.*

To ensure near-oracle performance of the DI test, we need the estimated Lfdrs $\{\widehat{t}_{ni}\}_{i=1}^m$ to be in close proximity of the oracle Lfdrs $\{t_{ni}\}_{i=1}^m$. This leads us to make the following consistency-type assumption.

Assumption 6. *For any fixed $n \geq k$, the mixture density f_n is continuous and positive on $\mathbb{R}$. The estimates $\widehat{\pi}_0$, and $\widehat{f}_n$ of π_0 and f_n, respectively, are such that $\widehat{\pi}_0 \xrightarrow{\mathbb{P}} \pi_0$ and $\mathbb{E}\|\widehat{f}_n - f_n\|^2 \to 0$ as $m \to \infty$ for fixed $n \geq k$.*

Assumption 6 readily implies that the estimated Lfdrs $\{t_{ni}\}_{i=1}^{m}$ in (3.11) are consistent, i.e., for any fixed $i \in \{1, \ldots, m\}$ and $n \geq k$, $\widehat{t}_{ni} \xrightarrow{\mathbb{P}} t_{ni}$ as $m \to \infty$. Such a consistency property of the estimated Lfdrs establishes the limits of lower and upper stopping boundaries defined in (3.13), as $m \to \infty$.

Lemma 8. *If $\mathbf{S}_n$ is continuous and Assumptions 3, 4 and 6 hold, then for fixed $n \in \mathbb{N}$, $\hat{t}_n^L \xrightarrow{\mathbb{P}} \mathcal{T}_n^L$ and $\hat{t}_n^U \xrightarrow{\mathbb{P}} \mathcal{T}_n^U$ as $m \to \infty$.*

Lemma 8 suggests that, for large m, the lower stopping boundary $\hat{t}_n^L$ is close to the theoretical cut-off $\mathcal{T}_n^L$ for mFDR control, and $\hat{t}_n^U$ is close to the theoretical cut-off $\mathcal{T}_n^U$ for mFNR control. The following theorem shows that the asymptotic behaviour of the final sample size T of the data-driven test is same as that of the oracle test.

Theorem 12. *Suppose that the proportion of true nulls $\pi_0 \in (\alpha, 1-\beta)$ for prefixed $\alpha, \beta \in (0,1)$. If Assumptions 3, 4 and 6 hold, then as $m \to \infty$, the stopping time T of the DI test satisfies the following properties.*

(i) $\mathbb{P}(T \leq n_0) \to 1$ as $m \to \infty$.

(ii) Further, if $T_n^L \neq T_n^U$ for all $n < n_0$, then $\mathbb{P}(T = n_0) \to 1$ as $m \to \infty$.

Theorem 12 readily implies that the stopping times of the oracle (OI) test and the data-driven (DI) test are asymptotically equivalent in the sense that $T - \tau \xrightarrow{\mathbb{P}} 0$ as $m \to \infty$. From Figure 3.2, it is apparent that the average sample sizes for the DI test converges to some constant as the number of hypotheses increases. Although we do not have a proof that the difference between the expected sample sizes of the DI and the OI test tends to zero, such a phenomena is observed in Figure 3.2. Note that, unlike our DI test, the traditional sequential multiple tests do not allow for adaptive stopping boundaries, and as a result, their final sample sizes may diverge to infinity as $m \to \infty$. Theorem 12 allows us to deal with large-scale sequential multiple testing problems at a low cost.

The following theorem proves that the DI test controls both false discovery and non-discovery rates asymptotically as m tends to infinity.

Theorem 13. *Suppose that the proportion of true nulls $\pi_0 \in (\alpha, 1-\beta)$ for prefixed $\alpha, \beta \in (0,1)$. If Assumptions 3, 4 and 6 hold, then the DI test given by $(T, \widehat{\boldsymbol{\delta}})$ yields, as $m \to \infty$,*

(i) $FDR \leq \alpha + o(1)$ and $FNR \leq \beta + o(1)$,

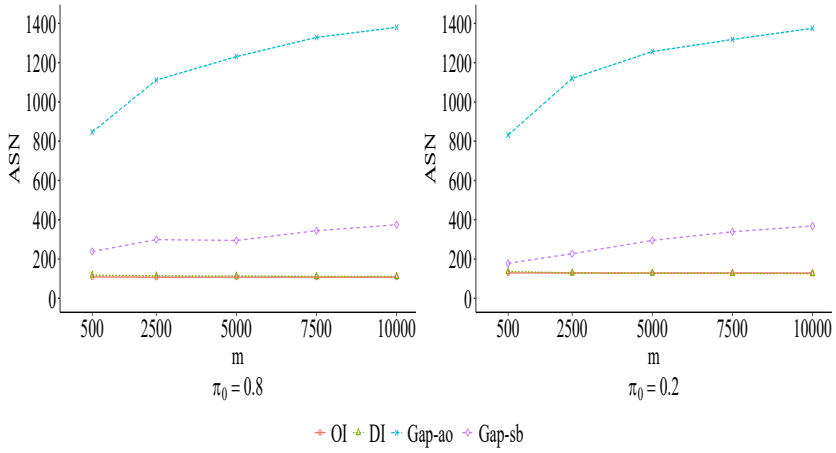

FIGURE 3.2: Comparison of estimated expected sample sizes (ASN) for OI test, DI test and Gap rules for the setup Ex1 in Chapter 5

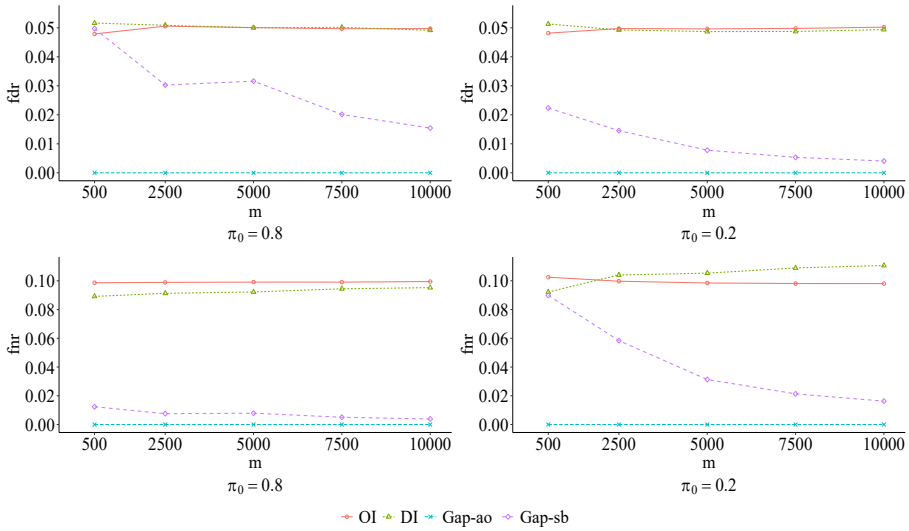

FIGURE 3.3: Comparison of estimated FDR and FNR for OI test, DI test and Gap rules for the setup Ex1 in Chapter 5

(ii) $mFDR \leq \alpha + o(1)$ *and* $mFNR \leq \beta + o(1)$.

3.4 Proof of Theorems

Proof of Theorem 9

Proof. (i): Define $[m] = \{1, 2, \ldots, m\}$. Let $\mathcal{A} = \{i \in [m] : \theta_i = 1\}$ be the set of non-nulls and $\{t_{ni}\}_{i=1}^m$ be the oracle statistics in (2.1). Note that, if for sample size $n \geq k$, $t_{ni} \leq \alpha$ for all $i \in \mathcal{A}$ and $t_{ni} \geq (1-\beta)$ for all $i \in \mathcal{A}^c$, then we must have, $\tau \leq n$. That is,

$$\begin{aligned}
\mathbb{P}(\tau \leq n) &\geq \mathbb{P}(\{\cap_{i\in\mathcal{A}}\{t_{ni} \leq \alpha\}\} \cap \{\cap_{i\in\mathcal{A}^c}\{t_{ni} \geq (1-\beta)\}\}) \\
&\geq \sum_{i\in\mathcal{A}} \mathbb{P}(t_{ni} \leq \alpha) + \sum_{i\in\mathcal{A}^c} \mathbb{P}(t_{ni} \geq (1-\beta)) - m + 1 \\
&\geq 1 - \sum_{i\in\mathcal{A}} \mathbb{P}(t_{ni} > \alpha) - \sum_{i\in\mathcal{A}^c} \mathbb{P}(t_{ni} < (1-\beta))
\end{aligned}$$

The second line comes from Bonferroni's inequality. The proof is complete by taking limit as $n \to \infty$ to both sides of the last inequality using Assumption 4.

(ii) and (iii): Now, let us state the following lemmas that are necessary to prove Theorem 9.

Lemma 9. *For any $l \in [r_\tau]$, the sequential test $(\tau, \mathbf{D})$ with $\mathbf{D} = (D_1, D_2, \ldots, D_m)$ given by $D_i = \mathbb{1}\left(t_{\tau i} \leq t_\tau^{(l)}\right)$ controls both FDR and mFDR at level $\alpha \in (0,1)$.*

Lemma 10. *For any $l \in [a_\tau]$, the sequential test $(\tau, \mathbf{d})$ with $\mathbf{d} = (d_1, d_2, \ldots, d_m)$ given by $d_i = \mathbb{1}\left(t_{\tau i} \leq t_\tau^{(m-l)}\right)$ controls both FNR and mFNR at level $\beta \in (0,1)$.*

Now, since at time τ, we must have $a_\tau + r_\tau \geq m$, the set $\{m - a_\tau, m - a_\tau + 1, \ldots, r_\tau\}$ is nonempty. Consider any s in that set. Therefore, $s \in [r_\tau]$ and $m - s \in [a_\tau]$. Therefore, using Lemma 9 and 10 the parts (ii) and (iii) of Theorem 9 are proved.

□

Proof of Theorem 10

Proof. Without loss of generality, let us consider the pilot sample size $k = 1$. Define $g_n = t_n^L - t_n^U$ and $\mathcal{G}_n = \mathcal{T}_n^L - \mathcal{T}_n^U$. Then, due to Lemma 6,

$$g_n \xrightarrow{\mathbb{P}} \mathcal{G}_n, \text{ as } m \to \infty \tag{3.14}$$

and due to Lemma 7, $n_0 = \inf\{n \in \mathbb{N} : \mathcal{T}_n^L > \mathcal{T}_n^U\}$ is a finite integer. This implies that $\mathcal{G}_n \leq 0$ when $n < n_0$, and $\mathcal{G}_{n_0} > 0$. For any $\epsilon > 0$, (3.14) implies

$$\mathbb{P}(|g_{n_0} - \mathcal{G}_{n_0}| < \epsilon) \to 1 \text{ as } m \to \infty.$$

By definition, $\mathcal{G}_{n_0} > 0$. Consider a positive number $\epsilon < \mathcal{G}_{n_0}$. If $g_{n_0} > \mathcal{G}_{n_0} - \epsilon$, g_{n_0} must be positive, i.e.,

$$\mathbb{P}(g_{n_0} > 0) > \mathbb{P}(g_{n_0} > \mathcal{G}_{n_0} - \epsilon) > \mathbb{P}(|g_{n_0} - \mathcal{G}_{n_0}| < \epsilon) \to 1 \text{ as } m \to \infty. \tag{3.15}$$

Now, $\tau > n_0$ implies that $g_i \leq 0$ for all $i \in \{1, \ldots, n_0\}$. Using (3.15),

$$\mathbb{P}(\tau > n_0) \leq \mathbb{P}(\cap_{i=1}^{n_0}\{g_i \leq 0\}) \leq \mathbb{P}(g_{n_0} \leq 0) \to 0 \text{ as } m \to \infty.$$

This proves the part (i) of Theorem 10. The following lemma is used to prove the part (ii) of Theorem 10.

Lemma 11. *If $\alpha < \pi_0 < 1 - \beta$, then for fixed $n \in \mathbb{N}$, $\mathcal{T}_n^L = \mathcal{T}_n^U$ if and only if there exists a $t_n^* \in (0, 1)$ for which the following conditions are satisfied simultaneously*

(i) $\mathbb{P}_0(t_{n1} \leq t_n^*) = \frac{\alpha(1-\pi_0-\beta)}{\pi_0(1-\alpha-\beta)}$,

(ii) $\mathbb{P}_1(t_{n1} \leq t_n^*) = \frac{(1-\alpha)(1-\pi_0-\beta)}{(1-\pi_0)(1-\alpha-\beta)}$,

(iii) $\mathcal{Q}_n(t) > \alpha$ *for all* $1 > t > t_n^*$,

(iv) $\mathcal{Q}_n'(t) > \beta$ *for all* $0 < t < t_n^*$.

Lemma 11 implies that if the conditions (i)-(iv) are not satisfied simultaneously, $\mathcal{G}_n < 0$ for all $n \in \{1, \ldots, n_0 - 1\}$. Thus, $\min\{-\mathcal{G}_1, \ldots, -\mathcal{G}_{n_0-1}, \mathcal{G}_{n_0}\}$ is positive. Let's choose $\epsilon > 0$ such that $\epsilon < \min\{-\mathcal{G}_1, \ldots, -\mathcal{G}_{n_0-1}, \mathcal{G}_{n_0}\}$. For any $\delta > 0$ and any $n \in \{1, 2, \ldots, n_0\}$, there exists a natural number $M_{n,\delta,\epsilon}$ such that

$$\mathbb{P}\left(|g_n - \mathcal{G}_n| < \epsilon\right) > 1 - \frac{\delta}{n_0}, \text{ for all } m > M_{n,\delta,\epsilon}, \tag{3.16}$$

since (3.14) holds and n_0 is finite. For all $n \leq n_0 - 1$, as $\mathcal{G}_n$ is negative and ϵ is smaller than $-\mathcal{G}_n$, we have $\mathcal{G}_n + \epsilon < 0$, i.e., $\{g_n < \mathcal{G}_n + \epsilon\}$ is a subset of $\{g_n < 0\}$. Note that

$\{|g_n - \mathcal{G}_n| \leq \epsilon\} \subseteq \{g_n < \mathcal{G}_n + \epsilon\}$. Due to (3.16), for $m > M_{n,\delta,\epsilon}$, we have

$$\mathbb{P}(g_n < 0) > \mathbb{P}(g_n < \mathcal{G}_n + \epsilon) \geq \mathbb{P}(|g_n - \mathcal{G}_n| < \epsilon) > 1 - \frac{\delta}{n_0}.$$

Similarly, noting that $\mathcal{G}_{n_0} > \epsilon > 0$, we have

$$\mathbb{P}(g_{n_0} > 0) > \mathbb{P}(g_{n_0} > \mathcal{G}_{n_0} - \epsilon) \geq \mathbb{P}(|g_{n_0} - \mathcal{G}_{n_0}| < \epsilon).$$

For $m > M_{n_0,\delta,\epsilon}$, $\mathbb{P}(|g_{n_0} - \mathcal{G}_{n_0}| < \epsilon) > 1 - \delta/n_0$ due to (3.16). Then, for all $m > M_0 := \max\{M_{1,\delta,\epsilon}, M_{2,\delta,\epsilon}, \ldots, M_{n_0,\delta,\epsilon}\}$, we have

$$\mathbb{P}(g_{n_0} > 0) > 1 - \frac{\delta}{n_0} \text{ and } \mathbb{P}(g_n < 0) > 1 - \frac{\delta}{n_0} \text{ for all } n \leq n_0 - 1. \tag{3.17}$$

The stopping time $\tau = n_0$ if and only if $g_n < 0$ for all $n \leq n_0 - 1$ and $g_{n_0} > 0$. Thus, for all $m > M_0$,

$$\begin{aligned}
\mathbb{P}(\tau = n_0) &= \mathbb{P}\left(\{\cap_{n \leq n_0 - 1}\{g_n < 0\}\} \cap \{g_{n_0} > 0\}\right) \\
&\geq \sum_{n \leq n_0 - 1} \mathbb{P}(g_n < 0) + \mathbb{P}(g_{n_0} > 0) - n_0 + 1 \\
&> n_0(1 - \frac{\delta}{n_0}) - n_0 + 1 = 1 - \delta.
\end{aligned}$$

This completes the proof of Theorem 10. □

Proof of Theorem 11

Proof. Let $A_{n,m}$ be the event that the difference between the K-the and the $K+1$-th ordered log-likelihood ratios is greater than the cutoff for GAP-ao rule given the number of hypothesis is m, i.e.,

$$A_{n,m} = \left\{\omega \in \Omega : \Lambda_n^{(K(\omega))}(\omega) - \Lambda_n^{(K(\omega)+1)}(\omega) > \log\left(\frac{K(\omega)(m - K(\omega))}{\alpha \wedge \beta}\right)\right\}.$$

For any positive integer L, if the GAP-ao method stops on or before L, then the event $A_{n,m}$ has occurred for at least one $n \leq L$, i.e., $\{T_G^* \leq L\} \subseteq \cup_{n \leq L} A_{n,m}$, which implies

$$\mathbb{P}(T_G^* \leq L) \leq \mathbb{P}(\cup_{n \leq L} A_{n,m}) \leq \sum_{n \leq L} \mathbb{P}(A_{n,m}). \tag{3.18}$$

Due to Assumption 5, for each $n \in \mathbb{N}$, the log-likelihood ratios for the data streams $\Lambda_{n1}, \ldots, \Lambda_{nm}$ are i.i.d. Assume that the CDF of Λ_{nj} is H_n. Define

$$B_{\epsilon,c,m} := \{\omega \in \Omega : |\Lambda_n^{(Z_m)}(\omega) - H_n^{-1}(\pi_1)| < \epsilon \text{ for } Z_m \in \mathbb{N}, Z_m = m\pi_1 + a_m, \\ |a_m| < c\sqrt{m}\log(m)\}.$$

Due to Lemma 10 of Bahadur (1966), there exists a natural number M_1 (depending on c and ϵ) such that $\mathbb{P}(\cap_{m \geq M_1} B_{\epsilon,c,m}) = 1$. For any $m' > M_1$, $\cap_{m \geq M_1} B_{\epsilon,c,m} \subseteq B_{\epsilon,c,m'}$. Therefore, we have

$$1 = \mathbb{P}(\cap_{m \geq M_1} B_{\epsilon,c,m}) \leq \mathbb{P}(B_{\epsilon,c,m'}) \leq 1.$$

That is, for fixed c and $\epsilon > 0$, $\mathbb{P}(B_{\epsilon,c,m}) = 1$ for $m > M_1$. For $c > 0$, define

$$D_{c,m} := \{\omega \in \Omega : |K(\omega) - m\pi_1| < c\sqrt{m}\log(m)\}.$$

Using Chebyshev's inequality, it can be shown that for fixed $c > 0$, $\mathbb{P}(D_{c,m}) \to 1$ as $m \to \infty$. For fixed $c, \epsilon > 0$, we have

$$1 \geq \mathbb{P}(D_{c,m} \cap B_{\epsilon,c,m}) \geq \mathbb{P}(D_{c,m}) + \mathbb{P}(B_{\epsilon,c,m}) - 1 \to 1 \text{ as } m \to \infty. \tag{3.19}$$

Thus, $\mathbb{P}(D_{c,m} \cap B_{\epsilon,c,m}) \to 1$ as $m \to \infty$. Now, suppose $\omega \in D_{c,m} \cap B_{\epsilon,c,m}$. Then $\omega \in D_{c,m}$ and $K(\omega)$ must lie between $(m\pi_1 - c\sqrt{m}\log(m))$ and $(m\pi_1 + c\sqrt{m}\log(m))$. Suppose $K(\omega) = m\pi_1 + \delta c\sqrt{m}\log(m)$ for some $\delta \in (-1, 1)$, i.e., both $K(\omega)$ and $K(\omega) + 1$ satisfy the defining conditions of Z_m. Since ω also belongs to $B_{\epsilon,c,m}$, we have $|\Lambda_n^{(K(\omega))}(\omega) - H_n^{-1}(\pi_1)| < \epsilon$ and $|\Lambda_n^{(K(\omega)+1)}(\omega) - H_n^{-1}(\pi_1)| < \epsilon$. Thus, $|\Lambda_n^{(K(\omega))}(\omega) - \Lambda_n^{(K(\omega)+1)}(\omega)| < 2\epsilon$ by the triangle inequality. Now, for such ω,

$$\log\left(\frac{K(\omega)(m - K(\omega))}{\alpha \wedge \beta}\right) = \log\left(\frac{(m\pi_1 + \delta c\sqrt{m}\log(m))(m\pi_0 - \delta c\sqrt{m}\log(m))}{\alpha \wedge \beta}\right) \\ = 2\log(m) + O(1) \to \infty \text{ as } m \to \infty.$$

Thus, if $\omega \in D_{c,m} \cap B_{\epsilon,c,m}$, then $\omega \notin A_{n,m}$. Therefore,

$$\mathbb{P}(A_{n,m}) = \mathbb{P}(A_{n,m} \cap (D_{c,m} \cap B_{\epsilon,c,m})) + \mathbb{P}(A_{n,m} \cap (D_{c,m} \cap B_{\epsilon,c,m})^c),$$

where the first term converges to 0 due to the discussion in the previous paragraph and the second term converges to 0 due to (3.19). So, $\mathbb{P}(A_{n,m}) \to 0$ as $m \to \infty$. Finally, we observe that $\mathbb{P}(T_G^* \leq L) \leq \sum_{n \leq L} \mathbb{P}(A_{n,m}) \to 0$ as $m \to \infty$ for all $L \in \mathbb{N}$. This proves the first part of Theorem 11.

To prove the second part, let's fix any $\epsilon > 0$ and note that

$$\begin{aligned}\mathbb{P}(\tau / T_G^* > \epsilon) &\leq \mathbb{P}(\{\tau > n_0\} \cup \{1/T_G^* > \epsilon/n_0\}) \\ &\leq \mathbb{P}(\tau > n_0) + \mathbb{P}(T_G^* < n_0/\epsilon) \to 0 \text{ as } m \to \infty.\end{aligned}$$

Note that under the assumptions of Theorem 10, $\mathbb{P}(\tau > n_0) \to 0$ and using $L = \lfloor n_0/\epsilon \rfloor + 1$ in the first part of Theorem 11, we complete the proof of the second part of Theorem 11. □

Proof of Theorem 12

Proof. Without loss of generality, let the pilot sample size k be 1. Suppose $\widehat{g}_n := \widehat{t}_n^L - \widehat{t}_n^U$ and $\mathcal{G}_n = \mathcal{T}_n^L - \mathcal{T}_n^U$. Lemma 8 implies

$$\widehat{g}_n \xrightarrow{\mathbb{P}} \mathcal{G}_n \text{ as } m \to \infty \text{ (for fixed } n). \tag{3.20}$$

By Lemma 7, n_0 is a finite integer, and by definition of n_0, $\mathcal{G}_n \leq 0$ for all $n < n_0$ and $\mathcal{G}_{n_0} > 0$. For $0 < \epsilon < \mathcal{G}_{n_0}$,

$$\mathbb{P}(\widehat{g}_{n_0} > 0) > \mathbb{P}(\widehat{g}_{n_0} > \mathcal{G}_{n_0} - \epsilon) > \mathbb{P}(|\widehat{g}_{n_0} - \mathcal{G}_{n_0}| < \epsilon) \to 1 \text{ as } m \to \infty. \tag{3.21}$$

Now, $T > n_0$ implies that $\widehat{g}_i \leq 0$ for all $i \in \{1, 2, \ldots, n_0 - 1, n_0\}$, and thus,

$$\mathbb{P}(T > n_0) \leq \mathbb{P}(\cap_{i=1}^{n_0} \{\widehat{g}_i \leq 0\}) \leq \mathbb{P}(\widehat{g}_{n_0} \leq 0) \to 0 \text{ as } m \to \infty,$$

using (3.21). This proves the first part of Theorem 12.

For the second part, we proceed along the same lines of the proof of Theorem 10. If the 4 conditions of Lemma 11 are not satisfied simultaneously, $\mathcal{G}_n < 0$ for all $n \leq n_0 - 1$. Suppose we choose $\epsilon > 0$ such that $\epsilon < \min\{-\mathcal{G}_1, \ldots, -\mathcal{G}_{n_0-1}, \mathcal{G}_{n_0}\}$. Then due to (3.20), for any $\delta > 0$, and for any $n \in \{1, 2, \ldots, n_0\}$, there exists a natural number $M_{n,\delta,\epsilon}$ such that

$$\mathbb{P}\left(|\widehat{g}_n - \mathcal{G}_n| < \epsilon\right) > 1 - \frac{\delta}{n_0} \text{ for all } m > M_{n,\delta,\epsilon}. \tag{3.22}$$

Proceeding along the same lines as in the proof of Theorem 10 and using (3.22), we can deduce that for all $m > M_0 := \max\{M_{1,\delta,\epsilon}, M_{2,\delta,\epsilon}, \ldots, M_{n_0,\delta,\epsilon}\}$,

$$\mathbb{P}(\widehat{g}_{n_0} > 0) > 1 - \frac{\delta}{n_0} \text{ and } \mathbb{P}(\widehat{g}_n < 0) > 1 - \frac{\delta}{n_0} \text{ for all } n \leq n_0 - 1 \tag{3.23}$$

Note that $T = n_0$ if and only if $\widehat{g}_n < 0$ for all $n \leq n_0 - 1$ and $\widehat{g}_{n_0}$, and thus for all $m > M_0$,

$$\begin{aligned}
\mathbb{P}(T = n_0) &= \mathbb{P}\left(\{\cap_{n \leq n_0 - 1}\{\widehat{g}_n < 0\}\} \cap \{\widehat{g}_{n_0} > 0\}\right) \\
&\geq \sum_{n \leq n_0 - 1} \mathbb{P}(\widehat{g}_n < 0) + \mathbb{P}(\widehat{g}_{n_0} > 0) - n_0 + 1 \\
&> n_0(1 - \frac{\delta}{n_0}) - n_0 + 1 = 1 - \delta.
\end{aligned}$$

The last two inequalities are due to (3.23) and the Bonferroni's inequality. □

Proof of Theorem 13

Proof. Note that

$$\begin{aligned}
\text{FDR} &= \mathbb{E}\left(\frac{\sum_{i=1}^m (1-\theta_i)\hat{\delta}_i}{(\sum_{i=1}^m \hat{\delta}_i) \vee 1}\right) = \sum_n \mathbb{E}\left(\frac{\sum_{i=1}^m (1-\theta_i)\hat{\delta}_i}{(\sum_{i=1}^m \hat{\delta}_{ni}) \vee 1} \mathbb{1}(T = n)\right) \\
&= \sum_n \mathbb{E}\left(\mathbb{E}\left(\frac{\sum_{i=1}^m (1-\theta_i)\hat{\delta}_{ni}}{(\sum_{i=1}^m \hat{\delta}_i) \vee 1} \mathbb{1}(T = n) \,\middle|\, \mathbf{S}_n\right)\right) \\
&= \sum_n \mathbb{E}\left(\frac{\sum_{i=1}^m (1 - \mathbb{E}(\theta_i|\mathbf{S}_n))\hat{\delta}_{ni}}{(\sum_{i=1}^m \hat{\delta}_i) \vee 1} \mathbb{1}(T = n)\right) \\
&\quad \text{(as } T = n \text{ is defined on } \sigma(\mathbf{S}_n)) \\
&= \sum_n \mathbb{E}\left(\frac{\sum_{i=1}^m t_{ni}\hat{\delta}_{ni}}{(\sum_{i=1}^m \hat{\delta}_{ni}) \vee 1} \mathbb{1}(T = n)\right).
\end{aligned} \tag{3.24}$$

Therefore, using (3.24), we can write

$$\begin{aligned}
\text{FDR} = & \sum_{n\le n_0} \mathbb{E}\left(\hat{Q}_n(t_n^{(s)})\,\mathbb{1}\left(T=n\right)\right) + \\
& \sum_{n\le n_0} \mathbb{E}\left(\left(\frac{\sum_{i=1}^m t_{ni}\hat{\delta}_{ni}}{(\sum_{i=1}^m \hat{\delta}_{ni})\vee 1} - \hat{Q}_n(t_n^{(s)})\right)\mathbb{1}\left(T=n\right)\right) + \\
& \sum_{n>n_0} \mathbb{E}\left(\frac{\sum_{i=1}^m t_{ni}\hat{\delta}_{ni}}{(\sum_{i=1}^m \hat{\delta}_{ni})\vee 1}\mathbb{1}\left(T=n\right)\right) \\
& \le \alpha + \sum_{n\le n_0} \mathbb{E}\left(\frac{\sum_{i=1}^m (t_{ni}-\hat{t}_{ni})\hat{\delta}_{ni}}{(\sum_{i=1}^m \hat{\delta}_{ni})\vee 1}\mathbb{1}\left(T=n\right)\right) + \\
& \sum_{n>n_0} \mathbb{E}\left(\frac{\sum_{i=1}^m t_{ni}\hat{\delta}_{ni}}{(\sum_{i=1}^m \hat{\delta}_{ni})\vee 1}\mathbb{1}\left(T=n\right)\right),
\end{aligned} \tag{3.25}$$

where $\hat{\delta}_{ni} = \mathbb{1}\left(\hat{t}_{ni} \le \hat{t}_n^{(s)}\right)$ for some $s \in [m-\hat{a}_n, \hat{r}_n]$ and

$$\hat{Q}_n(t) = \frac{\sum_{i=1}^m \hat{t}_{ni}\,\mathbb{1}\left(\hat{t}_{ni}\le t\right)}{\sum_{i=1}^m \mathbb{1}\left(\hat{t}_{ni}\le t\right)} \text{ for } t\in[0,1].$$

Here, s is well-defined since $\hat{a}_n + \hat{r}_n \ge m$ at $T=n$. Using the non-decreasing property of $\hat{Q}_n(t)$ and the definition of s and $\hat{r}_n$ at $T=n$, we have

$$\hat{Q}_n(t_n^{(s)}) \le \hat{Q}_n(t_n^{(\hat{r}_n)}) \le \alpha.$$

Note from the last line of (3.25) that

$$\sum_{n>n_0} \mathbb{E}\left(\frac{\sum_{i=1}^m t_{ni}\hat{\delta}_{ni}}{(\sum_{i=1}^m \hat{\delta}_{ni})\vee 1}\mathbb{1}\left(T=n\right)\right) \le \sum_{n>n_0} \mathbb{E}\left(\mathbb{1}\left(T=n\right)\right) = \mathbb{P}(T>n_0) \to 0,$$

as $m\to\infty$, due to Theorem 12. For any $n\le n_0$,

$$\begin{aligned}
& \mathbb{E}\left(\frac{\sum_{i=1}^m (t_{ni}-\hat{t}_{ni})\hat{\delta}_{ni}}{(\sum_{i=1}^m \hat{\delta}_{ni})\vee 1}\mathbb{1}\left(T=n\right)\right) \\
& = \mathbb{E}\left(\frac{\frac{1}{m}\sum_{i=1}^m (t_{ni}-\hat{t}_{ni})\,\mathbb{1}\left(\hat{t}_{ni}\le \hat{t}_n^{(s)}\right)}{\frac{1}{m}\left(\sum_{i=1}^m \mathbb{1}\left(\hat{t}_{ni}\le \hat{t}_n^{(s)}\right)\right)\vee 1}\mathbb{1}\left(T=n\right)\right)
\end{aligned} \tag{3.26}$$

For any i and j in $[m]$, t_{ni} and t_{nj} are i.i.d. and

$$\hat{t}_{ni} = \frac{\hat{\pi}_0\phi(S_{ni})}{\hat{f}_n(S_{ni})} \text{ and } \hat{t}_{nj} = \frac{\hat{\pi}_0\phi(S_{nj})}{\hat{f}_n(S_{nj})}.$$

Note that $\phi(S_{ni})$ and $\phi(S_{nj})$ are i.i.d. as S_{ni} and S_{nj} are i.i.d. The density estimates, $\hat{f}_n(S_{ni})$ and $\hat{f}_n(S_{nj})$ are not independent but identically distributed. Finally, $\hat{\pi}_0$ is same for both $\hat{t}_{ni}$ and $\hat{t}_{nj}$, and hence, they are identically distributed. This implies that $(t_{ni} - \hat{t}_{ni})\,\mathbb{1}\left(\hat{t}_{ni} \le \hat{t}_n^{(s)}\right)$ and $(t_{nj} - \hat{t}_{nj})\,\mathbb{1}\left(\hat{t}_{nj} \le \hat{t}_n^{(s)}\right)$ are identically distributed for any $i \neq j, i, j \in [m]$. Therefore,

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{m}\sum_{i=1}^{m}(t_{ni} - \hat{t}_{ni})\,\mathbb{1}\left(\hat{t}_{ni} \le \hat{t}_n^{(s)}\right)\right) &= \frac{1}{m^2}\sum_{i=1}^{m}\mathrm{Var}\left((t_{ni} - \hat{t}_{ni})\,\mathbb{1}\left(\hat{t}_{ni} \le \hat{t}_n^{(s)}\right)\right) \\ &\quad + \frac{1}{m^2}\sum_{i\neq j}^{m}\mathrm{cov}\left((t_{ni} - \hat{t}_{ni})\,\mathbb{1}\left(\hat{t}_{ni} \le \hat{t}_n^{(s)}\right), (t_{nj} - \hat{t}_{nj})\,\mathbb{1}\left(\hat{t}_{nj} \le \hat{t}_n^{(s)}\right)\right)\end{aligned}$$

For fixed $i, j \in [m]$ and $i \neq j$, let

$$\begin{aligned}\rho_{ij} &:= \mathrm{cov}((t_{ni} - \hat{t}_{ni})\,\mathbb{1}(\hat{t}_{ni} \le \hat{t}_n^{(s)}), (t_{nj} - \hat{t}_{nj})\,\mathbb{1}(\hat{t}_{nj} \le \hat{t}_n^{(s)})) \\ &\le \mathrm{Var}((t_{ni} - \hat{t}_{ni})\,\mathbb{1}(\hat{t}_{ni} \le \hat{t}_n^{(s)})) \\ &\le \mathbb{E}\left(((t_{ni} - \hat{t}_{ni})\,\mathbb{1}(\hat{t}_{ni} \le \hat{t}_n^{(s)}))^2\right) \le \mathbb{E}(t_{ni} - \hat{t}_{ni})^2 \to 0 \text{ as } m \to \infty.\end{aligned}$$

The last convergence follows from the dominated convergence theorem (DCT) since $|t_{ni} - \hat{t}_{ni}| \le 2$ and $\hat{t}_{ni} - t_{ni} \xrightarrow{\mathbb{P}} 0$ due to Assumption 6. Thus,

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{m}\sum_{i=1}^{m}(t_{ni} - \hat{t}_{ni})\,\mathbb{1}(\hat{t}_{ni} \le \hat{t}_n^{(s)})\right) &= \frac{1}{m^2}\sum_{i=1}^{m}\rho_{ii} + \frac{1}{m^2}\sum_{i\neq j}\rho_{ij} \\ &= \frac{\rho_{11}}{m} + \frac{m(m-1)}{m^2}\rho_{12} \to 0,\end{aligned}$$

and using Markov inequality, as $m \to \infty$,

$$\frac{1}{m}\sum_{i=1}^{m}(t_{ni} - \hat{t}_{ni})\,\mathbb{1}\left(\hat{t}_{ni} \le \hat{t}_n^{(s)}\right) - \mathbb{E}\left((t_{ni} - \hat{t}_{ni})\,\mathbb{1}\left(\hat{t}_{ni} \le \hat{t}_n^{(s)}\right)\right) \xrightarrow{\mathbb{P}} 0.$$

Noting that $\left|\mathbb{E}\left((t_{ni} - \hat{t}_{ni})\,\mathbb{1}\left(\hat{t}_{ni} \le \hat{t}_n^{(s)}\right)\right)\right| \le E(|t_{ni} - \hat{t}_{ni}|) \to 0$ due to DCT and Assumption 6, we have $\frac{1}{m}\sum_{i=1}^{m}(t_{ni} - \hat{t}_{ni})\,\mathbb{1}\left(\hat{t}_{ni} \le \hat{t}_n^{(s)}\right) \xrightarrow{\mathbb{P}} 0$ as $m \to \infty$. Now, let us deal with the denominator of RHS of (3.26). For $s = m - \hat{a}_n$, $\mathbb{1}\left(\hat{t}_{ni} \le \hat{t}_n^{(m-\hat{a}_n)}\right) =$

$\mathbb{1}\left(\hat{t}_{ni} \leq \hat{t}_n^U\right)$ due to (4.14) and

$$\operatorname{Var}\left(\frac{1}{m}\sum_{i=1}^{m}\mathbb{1}\left(\hat{t}_n^i \leq \hat{t}_n^U\right)\right) = \frac{1}{m^2}\sum_{i=1}^{m}\operatorname{Var}\left(\mathbb{1}\left(\hat{t}_{ni} \leq \hat{t}_n^U\right)\right) + \frac{1}{m^2}\sum_{i\neq j}\operatorname{cov}\left(\mathbb{1}\left(\hat{t}_{ni} \leq \hat{t}_n^U\right), \mathbb{1}\left(\hat{t}_{nj} \leq \hat{t}_n^U\right)\right). \quad (3.27)$$

Note that

$$\begin{aligned}\rho'_{ij} &:= \operatorname{cov}\left(\mathbb{1}\left(\hat{t}_{ni} \leq \hat{t}_n^U\right), \mathbb{1}\left(\hat{t}_{nj} \leq \hat{t}_n^U\right)\right)\\ &= \mathbb{E}\left(\mathbb{1}\left(\hat{t}_{ni} \leq \hat{t}_n^U\right)\mathbb{1}\left(\hat{t}_{nj} \leq \hat{t}_n^U\right)\right) - \left(\mathbb{E}\left(\mathbb{1}\left(\hat{t}_{ni} \leq \hat{t}_n^U\right)\right)\right)^2\\ &= \mathbb{P}\left(\hat{t}_{ni} \leq \hat{t}_n^U, \hat{t}_{nj} \leq \hat{t}_n^U\right) - \left(\mathbb{P}\left(\hat{t}_{ni} \leq \hat{t}_n^U\right)\right)^2.\end{aligned}$$

Since $\hat{t}_{ni} - t_{ni} \xrightarrow{\mathbb{P}} 0$ for fixed $i \in [m]$, and $\hat{t}_n^U \xrightarrow{\mathbb{P}} \mathcal{T}_n^U$ due to Lemma 8, we have $(\hat{t}_{ni} - \hat{t}_n^U) - (t_{ni} - \mathcal{T}_n^U) \xrightarrow{\mathbb{P}} 0$ as $m \to \infty$. Therefore, $\mathbb{P}\left(\hat{t}_{ni} - \hat{t}_n^U \leq 0\right) \to \mathbb{P}\left(t_{ni} - \mathcal{T}_n^U \leq 0\right)$ as $m \to \infty$. Similarly, it can be shown that

$$\mathbb{P}\left(\hat{t}_{ni} \leq \hat{t}_n^U, \hat{t}_{nj} \leq \hat{t}_n^U\right) \to \mathbb{P}\left(t_{ni} \leq \mathcal{T}_n^U, t_{nj} \leq \mathcal{T}_n^U\right) \text{ as } m \to \infty,$$

for $i, j \in \{1, 2, \ldots, m\}$ and $i \neq j$. This implies

$$\begin{aligned}\rho'_{ij} &= \mathbb{P}\left(\hat{t}_{ni} \leq \hat{t}_n^U, \hat{t}_{nj} \leq \hat{t}_n^U\right) - \left(\mathbb{P}\left(\hat{t}_{ni} \leq \hat{t}_n^U\right)\right)^2\\ &\to \mathbb{P}\left(t_{ni} \leq \mathcal{T}_n^U, t_{nj} \leq \mathcal{T}_n^U\right) - \left(\mathbb{P}\left(t_{ni} \leq \mathcal{T}_n^U\right)\right)^2 = 0 \text{ as } m \to \infty,\end{aligned}$$

since t_{ni} and t_{nj} are i.i.d. Now, 3.27 implies as $m \to \infty$,

$$\begin{aligned}\operatorname{Var}\left(\frac{1}{m}\sum_{i=1}^{m}\mathbb{1}\left(\hat{t}_{ni} \leq \hat{t}_n^U\right)\right) &\leq \frac{1}{m^2}\left(\sum_{i=1}^{m}1 + \sum_{i\neq j}\rho'_{ij}\right)\\ &= \frac{1}{m} + \frac{m(m-1)}{m^2}\rho'_{12} \to 0.\end{aligned}$$

That is, $\frac{1}{m}\sum_{i=1}^{m}\mathbb{1}\left(\hat{t}_{ni} \leq \hat{t}_n^U\right) - \mathbb{E}\left(\mathbb{1}\left(\hat{t}_{ni} \leq \hat{t}_n^U\right)\right) \xrightarrow{\mathbb{P}} 0$, which implies,

$$\frac{1}{m}\sum_{i=1}^{m}\mathbb{1}\left(\hat{t}_{ni} \leq \hat{t}_n^U\right) \xrightarrow{\mathbb{P}} \mathbb{P}\left(t_{ni} \leq \mathcal{T}_n^U\right) \text{ as } m \to \infty.$$

Now, for $s = \hat{r}_n$, $\hat{\delta}_i = \mathbb{1}\left(\hat{t}_{ni} \le \hat{t}_n^{(r_n)}\right) = \mathbb{1}\left(\hat{t}_{ni} < \hat{t}_n^L\right)$ using (4.14). Therefore, the denominator of (3.26) becomes $\frac{1}{m}\left(\sum_{i=1}^m \mathbb{1}\left(\hat{t}_{ni} < \hat{t}_n^L\right) \vee 1\right)$. As before, it can be shown that

$$\text{Var}\left(\frac{1}{m}\sum_{i=1}^m \mathbb{1}\left(\hat{t}_{ni} < \hat{t}_n^L\right)\right) \to 0, \text{ as } m \to \infty.$$

Therefore, $\frac{1}{m}\sum_{i=1}^m \mathbb{1}\left(\hat{t}_{ni} < \hat{t}_n^L\right) - \mathbb{E}\left(\mathbb{1}\left(\hat{t}_{ni} < \hat{t}_n^L\right)\right) \xrightarrow{\mathbb{P}} 0$ as $m \to \infty$. Using Assumption 4 and Lemma 4, $\mathbb{P}\left(\hat{t}_{ni} < \hat{t}_n^L\right) \to \mathbb{P}\left(t_{ni} < \mathcal{T}_n^L\right)$ as $m \to \infty$. Note that $\mathbb{P}\left(t_{ni} < t\right) = \mathbb{P}\left(t_{ni} \le t\right)$ for any $t \in (0,1)$ as S_{ni}'s are continuous.

Lemma 12. *For $\alpha < \pi_0 < 1 - \beta$ and any fixed $n \in \mathbb{N}$, we have*

$$0 < \mathbb{P}(t_{ni} \le \mathcal{T}_n^L) < 1 \text{ and } 0 < \mathbb{P}(t_{ni} \le \mathcal{T}_n^U) < 1 \text{ for all } i = 1, \ldots, m.$$

Since for any $s \in [m - \hat{a}_n, \hat{r}_n]$,

$$\frac{1}{m}\sum_{i=1}^m \mathbb{1}\left(\hat{t}_{ni} \le \hat{t}_n^{(m-\hat{a}_n)}\right) \le \frac{1}{m}\sum_{i=1}^m \mathbb{1}\left(\hat{t}_{ni} \le \hat{t}_n^{(s)}\right) \le \frac{1}{m}\sum_{i=1}^m \mathbb{1}\left(\hat{t}_{ni} \le \hat{t}_n^{(\hat{r}_n)}\right), \tag{3.28}$$

we observe using Lemma 12 that the first and the last terms in (3.28) (i.e., for $s = \hat{r}_n$ and $s = m - \hat{a}_n$) converge to positive probabilities implying that the middle term in (3.28) is bounded above and below by a positive number. Thus, the denominator of RHS of (3.26), given by

$$\frac{\frac{1}{m}\sum_{i=1}^m (t_{ni} - \hat{t}_{ni})\,\mathbb{1}\left(\hat{t}_{ni} \le \hat{t}_n^{(s)}\right)}{\frac{1}{m}\left(\sum_{i=1}^m \mathbb{1}\left(\hat{t}_{ni} \le \hat{t}_n^{(s)}\right) \vee 1\right)} \xrightarrow{\mathbb{P}} 0$$

Hence, by DCT, the second term in (3.25)

$$\sum_{n \le n_0} \mathbb{E}\left(\frac{\sum_{i=1}^m (t_{ni} - \hat{t}_{ni})\hat{\delta}_{ni}}{(\sum_{i=1}^m \hat{\delta}_{ni}) \vee 1}\,\mathbb{1}\left(T = n\right)\right) \to 0 \text{ as } m \to \infty.$$

This completes proof of the asymptotic control of FDR at level α. Now, expressing FNR in a similar way as FDR, we note that

$$\begin{aligned}\text{FNR} &= \mathbb{E}\left(\frac{\sum_{i=1}^m (1 - \hat{\delta}_i)\theta_i}{(\sum_{i=1}^m (1 - \hat{\delta}_i)) \vee 1}\right) \\ &= \sum_n \mathbb{E}\left(\frac{\sum_{i=1}^m (1 - t_{ni})(1 - \hat{\delta}_{ni})}{(\sum_{i=1}^m (1 - \hat{\delta}_{ni})) \vee 1}\,\mathbb{1}\left(T = n\right)\right).\end{aligned}$$

Thus, expanding FNR as in the case of FDR, we have

$$\begin{aligned}
\text{FNR} = & \sum_{n \le n_0} \mathbb{E}\left(\hat{Q}'_n(\hat{t}_n^{(s)})\,\mathbb{1}\left(T = n\right)\right) + \\
& \sum_{n \le n_0} \mathbb{E}\left(\left(\frac{\sum_{i=1}^m (1 - t_{ni})(1 - \hat{\delta}_{ni})}{(\sum_{i=1}^m (1 - \hat{\delta}_{ni})) \vee 1} - \hat{Q}'_n(\hat{t}_n^{(s)})\right)\mathbb{1}\left(T = n\right)\right) + \\
& \sum_{n > n_0} \mathbb{E}\left(\frac{\sum_{i=1}^m (1 - t_{ni})(1 - \hat{\delta}_{ni})}{(\sum_{i=1}^m (1 - \hat{\delta}_{ni})) \vee 1}\,\mathbb{1}\left(T = n\right)\right) \\
& \le \beta + \sum_{n \le n_0} \mathbb{E}\left(\frac{\sum_{i=1}^m (\hat{t}_{ni} - t_{ni})(1 - \hat{\delta}_{ni})}{(\sum_{i=1}^m (1 - \hat{\delta}_{ni})) \vee 1}\,\mathbb{1}\left(T = n\right)\right) + \\
& \sum_{n > n_0} \mathbb{E}\left(\frac{\sum_{i=1}^m (1 - t_{ni})(1 - \hat{\delta}_{ni})}{(\sum_{i=1}^m (1 - \hat{\delta}_{ni})) \vee 1}\,\mathbb{1}\left(T = n\right)\right), \qquad (3.29)
\end{aligned}$$

where $\hat{\delta}_{ni} = \mathbb{1}\left(\hat{t}_{ni} \le \hat{t}_n^{(s)}\right)$ for some $s \in [m - \hat{a}_n, \hat{r}_n]$ and

$$\hat{Q}'_n(t) = \frac{\sum_{i=1}^m (1 - \hat{t}_{ni})\,\mathbb{1}\left(\hat{t}_{ni} > t\right)}{\sum_{i=1}^m \mathbb{1}\left(\hat{t}_{ni} > t\right)} \text{ for } t \in [0, 1].$$

Using the non-increasing property of $\hat{Q}'_n(t)$ and the choice of s (at $T = n$), we have $\hat{Q}'_n(t_n^{(s)}) \le \hat{Q}'_n(t_n^{(m-\hat{a}_n)}) \le \beta$. The last term in (3.29) converges to 0 since $\mathbb{P}(T > n_0) \to 0$ as $m \to \infty$. Further, we can show following a similar argument (as in the case of FDR control) that

$$\frac{1}{m}\sum_{i=1}^m (\hat{t}_{ni} - t_{ni})\,\mathbb{1}\left(\hat{t}_{ni} > \hat{t}_n^{(s)}\right) \xrightarrow{\mathbb{P}} 0 \text{ as } m \to \infty. \qquad (3.30)$$

Proceeding along the same lines as in the case of FDR control, we can show that for $s = m - \hat{a}_n$, $1 - \hat{\delta}_{ni} = \mathbb{1}\left(\hat{t}_{ni} > \hat{t}_n^{(m-\hat{a}_n)}\right) = \mathbb{1}\left(\hat{t}_{ni} > \hat{t}_n^U\right)$, and

$$\frac{1}{m}\sum_{i=1}^m \mathbb{1}\left(\hat{t}_{ni} > \hat{t}_n^U\right) \xrightarrow{\mathbb{P}} \mathbb{P}\left(t_{ni} > \mathcal{T}_n^U\right) \in (0, 1) \text{ as } m \to \infty,$$

and for $s = \hat{r}_n$, $1 - \hat{\delta}_{ni} = \mathbb{1}\left(\hat{t}_{ni} > \hat{t}_n^{(\hat{r}_n)}\right) = \mathbb{1}\left(\hat{t}_{ni} \ge \hat{t}_n^L\right)$, and

$$\frac{1}{m}\sum_{i=1}^m \mathbb{1}\left(\hat{t}_{ni} \ge \hat{t}_n^L\right) \xrightarrow{\mathbb{P}} \mathbb{P}\left(t_{ni} \ge \mathcal{T}_n^L\right) \in (0, 1) \text{ as } m \to \infty.$$

Since for any $s \in [m - \hat{a}_n, \hat{r}_n]$,

$$\frac{1}{m}\sum_{i=1}^{m} \mathbb{1}\left(\hat{t}_{ni} > \hat{t}_n^{(\hat{r}_n)}\right) \leq \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}\left(\hat{t}_{ni} > \hat{t}_n^{(s)}\right) \leq \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}\left(\hat{t}_{ni} > \hat{t}_n^{(m-\hat{a}_n)}\right), \tag{3.31}$$

we observe using Lemma 12 that the first and the last terms of (3.31) converge to positive numbers, and hence, the middle term of (3.31) is bounded above and below by a positive number. Therefore, using (3.30),

$$\frac{\frac{1}{m}\sum_{i=1}^{m}(\hat{t}_{ni} - t_{ni})(1 - \hat{\delta}_{ni})}{\frac{1}{m}\left(\left(\sum_{i=1}^{m}(1 - \hat{\delta}_{ni})\right) \vee 1\right)} \xrightarrow{\mathbb{P}} 0, \quad \text{as } m \to \infty,$$

and applying DCT, the second term in the last line of (3.29), given by

$$\begin{aligned}
&\sum_{n \leq n_0} \mathbb{E}\left(\frac{\frac{1}{m}\sum_{i=1}^{m}(\hat{t}_{ni} - t_{ni})(1 - \hat{\delta}_{ni})}{\frac{1}{m}\left(\left(\sum_{i=1}^{m}(1 - \hat{\delta}_{ni})\right) \vee 1\right)} \mathbb{1}(T = n)\right) \\
&\leq \sum_{n \leq n_0} \mathbb{E}\left(\frac{\frac{1}{m}\sum_{i=1}^{m}(\hat{t}_{ni} - t_{ni})(1 - \hat{\delta}_{ni})}{\frac{1}{m}\left(\left(\sum_{i=1}^{m}(1 - \hat{\delta}_{ni})\right) \vee 1\right)}\right) \to 0 \text{ as } m \to \infty,
\end{aligned}$$

This completes the proof of FNR control at level β. The proof of part (ii) of Theorem 13 is similar to part (i), and hence, we omit the proof. □

3.5 Proof of Lemmas

Proof of Lemma 5

Proof. Note that for continuous summary statistics $\mathbf{S}_n$, $\{t_{ni}\}_{i=1}^{m}$ is continuous, and hence, $\mathbb{P}(t_{ni} = 1) = \mathbb{P}(t_{ni} = 0) = 0 \; \forall i \in [m]$.

If part: Assume $t_n^L > t_n^U$. If $t_n^U = 0$ then $a_n = m$, and if $t_n^L = 1$ then $r_n = m$ by definition of a_n and r_n leading to $r_n + a_n \geq m$ with probability 1 for these two cases. If $0 < t_n^U < t_n^L < 1$, by definitions of t_n^L and t_n^U, $t_n^L = t_n^{(r_n+1)}$ and $t_n^U = t_n^{(m-a_n)}$. Therefore, we must have $r_n + 1 > m - a_n$ and finally, $r_n + a_n \geq m$. This holds irrespective of whether $\mathbb{P}(t_{ni} = t_{nj})$ is 0 or a positive value. Hence, part (ii) of Lemma 5 is also proved.

Only if part: Assume $r_n + a_n \geq m$. If $r_n = m$, then $t_n^L = 1$ and $t_n^U < 1$ since $\mathbb{P}(t_{ni} < 1) = 1$ for all $i \in [m]$. This proves $t_n^L > t_n^U$ with probability 1. Similarly, if $a_n = m$, then $t_n^U = 0$ and $t_n^L > 0$ since $\mathbb{P}(t_{ni} > 0) = 1$ for all $i \in [m]$. In case when

$r_n < m$ and $a_n < m$, we have $t_n^L = t_n^{(r_n+1)}$ and $t_n^U = t_n^{(m-a_n)}$. Since $r_n + a_n \geq m$, i.e., $r_n + 1 > m - a_n$, and $\mathbb{P}(t_{ni} = t_{nj}) = 0$, we get $t_n^L > t_n^U$ with probability 1. This completes the proof of part (i) of Lemma 5. □

Proof of Lemma 6

Proof. For any $t \in [0,1]$, let us define

$$Q_n(t) := \begin{cases} \frac{\sum_{i=1}^m t_{ni} \mathbb{1}(t_{ni} \leq t)}{\sum_{i=1}^m \mathbb{1}(t_{ni} \leq t)} & \text{if } t_n^{(1)} \leq t \leq 1 \\ 0 & \text{if } 0 \leq t < t_n^{(1)} \end{cases}$$

For $t = t_n^{(r)}$, the value of $Q_n(t)$ is $\sum_{i=1}^r t_n^{(i)}/r$. For all intermediate values between two successive ordered oracle statistics, the value of $Q_n(t)$ is constant, i.e., $Q_n(t) = \sum_{i=1}^r t_n^{(i)}/r$ for all $t \in [t_n^{(r)}, t_n^{(r+1)})$. Since for all $r > s$, $\sum_{i=1}^r t_n^{(i)}/r \geq \sum_{i=1}^s t_n^{(i)}/s$, $Q_n(t)$ is a non-decreasing, right-continuous step function of t. For fixed $t \geq t_n^{(1)}$,

$$Q_n(t) = \frac{\sum_{i=1}^m t_{ni} \mathbb{1}(t_{ni} \leq t)/m}{\sum_{i=1}^m \mathbb{1}(t_{ni} \leq t)/m},$$

where the denominator and the numerator converge in probability to $\mathbb{P}(t_{n1} \leq t)$ and $\mathbb{E}(t_{n1}\mathbb{1}(t_{n1} \leq t)) = \pi_0 \mathbb{P}(t_{n_1} < t))$, respectively due to the weak law of large numbers (since under Assumption 3, $\{t_{ni}\}_{i=1}^m$ are i.i.d.). Thus,

$$Q_n(t) \xrightarrow{\mathbb{P}} \frac{\pi_0 \mathbb{P}(t_{n_1} < t)}{\mathbb{P}(t_{n1} \leq t)} = \mathcal{Q}_n(t) \text{ as } m \to \infty.$$

Note that if $r_n < m$, then $\sup\{t \in (0,1] : Q_n(t) \leq \alpha\} = t_n^{(r_n+1)}$, and if $r_n = m$, then $\sup\{t \in (0,1] : Q_n(t) \leq \alpha\} = 1$. So, we have

$$\sup\{t \in (0,1] : Q_n(t) \leq \alpha\} = t_n^L.$$

Therefore, due to Lemma A.5 of Sun and Cai (2007a), we have

$$t_n^L \xrightarrow{\mathbb{P}} \mathcal{T}_n^L, \text{ for fixed } n \in \mathbb{N}, \text{ as } m \to \infty. \tag{3.32}$$

To prove the second part, let us define for any $t \in [0,1]$,

$$Q'_n(t) := \begin{cases} \frac{\sum_{i=1}^m (1-t_{ni})\,\mathbb{1}(t_{ni} \geq t)}{\sum_{i=1}^m \mathbb{1}(t_{ni} \geq t)} & \text{if } 0 \leq t \leq t_n^{(m)} \\ 0 & \text{if } t_n^{(m)} < t \leq 1. \end{cases}$$

Note that $Q'_n(t)$ is a left-continuous, non-increasing step function taking values in $[0,1)$. Further, for fixed $t \in (0,1)$, $Q'_n(t) \xrightarrow{\mathbb{P}} \mathcal{Q}'_n(t)$ as $m \to \infty$, and $\inf\{t \in [0,1) : Q'_n(t) \leq \beta\} = t_n^U$ using similar arguments as in the case of $Q_n(t)$. Recall from definition that $\mathcal{T}_n^U = \inf\{t \in [0,1] : \mathcal{Q}'_n(t) \leq \beta\}$. If $Q'_n(t)$ is continuous in t, then it is easy to see that $t_n^U \xrightarrow{\mathbb{P}} \mathcal{T}_n^U$ as $m \to \infty$. However, by definition, $Q'_n(t)$ is a step function that is not continuous in t. To prove the intended result, let us define, for $t_n^{(i)} < t < t_n^{(i+1)}$,

$$Q_n^{'+}(t) := Q'_n(t_n^{(i)}) \frac{t_n^{(i+1)} - t}{t_n^{(i+1)} - t_n^{(i)}} + Q'_n(t_n^{(i+1)}) \frac{t - t_n^{(i)}}{t_n^{(i+1)} - t_n^{(i)}}$$

and

$$Q_n^{'-}(t) := Q'_n(t_n^{(i+1)}) \frac{t_n^{(i+1)} - t}{t_n^{(i+1)} - t_n^{(i)}} + Q'_n(t_n^{(i+2)}) \frac{t - t_n^{(i)}}{t_n^{(i+1)} - t_n^{(i)}},$$

where $i \in \{0,1,2,\ldots,m-1\}$ with $t_n^{(0)} := 0$ and $t_n^{(m+1)} := 1$. These two equations correspond to two continuous functions joining the points above and below the function $Q'_n(t)$ respectively (see Figure 3.4). That is,

$$Q_n^{'-}(t) \leq Q'_n(t) \leq Q_n^{'+}(t) \text{ for any } t \in (0,1). \tag{3.33}$$

For fixed $n \in \mathbb{N}$ and $t \in (0,1)$, it can be shown that

$$Q_n^{'+}(t) - Q_n^{'-}(t) \xrightarrow{a.s.} 0 \text{ as } m \to \infty. \tag{3.34}$$

Using (3.33), (3.34) and $Q'_n(t) \xrightarrow{\mathbb{P}} \mathcal{Q}'_n(t)$ as $m \to \infty$, we conclude

$$Q_n^{'+}(t) \xrightarrow{\mathbb{P}} \mathcal{Q}'_n(t) \text{ and } Q_n^{'-}(t) \xrightarrow{\mathbb{P}} \mathcal{Q}'_n(t) \text{ as } m \to \infty.$$

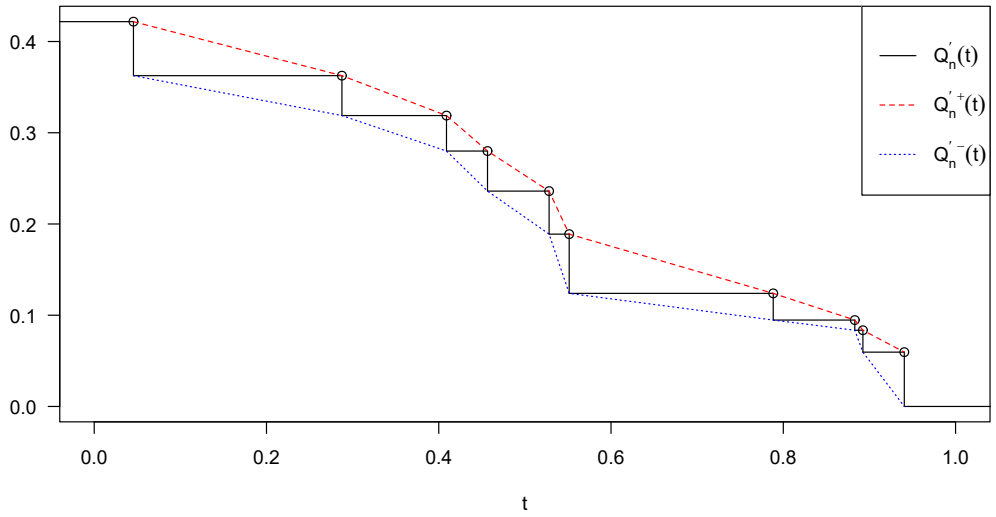

FIGURE 3.4: The $Q_n'(t)$, $Q_n'^{+}(t)$ and $Q_n'^{-}(t)$ functions for $m = 10$.

Now, let us define

$$t_n^{U+} := \inf\{t \in (0,1) : Q_n'^{+}(t) \leq \beta\} \text{ and } t_n^{U-} := \inf\{t \in (0,1) : Q_n'^{-}(t) \leq \beta\}.$$

Since $Q_n'(t)$, $Q_n'^{+}(t)$ and $Q_n'^{-}(t)$ are non-increasing functions, we can show using (3.33) that $t_n^{U+} \leq t_n^{U} \leq t_n^{U-}$. Since $Q_n'^{+}(t)$ and $Q_n'^{-}(t)$ are continuous functions of t, we have $t_n^{U+} \xrightarrow{\mathbb{P}} \mathcal{T}_n^{U}$ and $t_n^{U-} \xrightarrow{\mathbb{P}} \mathcal{T}_n^{U}$ as $m \to \infty$. This implies $t_n^{U} \xrightarrow{\mathbb{P}} \mathcal{T}_n^{U}$ as $m \to \infty$, and hence, the proof of Lemma 6 is complete. □

Proof of Lemma 7

Proof. Note that Assumption 3 implies $t_{n1}, \ldots, t_{nm}$ are i.i.d. and Assumption 4 implies, as $n \to \infty$, $\mathbb{P}_1(t_{n1} \geq t) \to 0$ for any $t \in (0,1]$ and $\mathbb{P}_0(t_{n1} \leq t') \to 0$ for any $t' \in [0,1)$. Therefore, for all $t \in (0,1)$, $\mathcal{Q}_n(t) \to 0$ and $\mathcal{Q}_n'(t) \to 0$ as $n \to \infty$. Let us fix $t_1, t_2 \in (0,1)$ such that $t_1 > t_2$. Note that by definitions of $\mathcal{T}_n^{L}$ and $\mathcal{T}_n^{U}$, $\mathcal{Q}_n(t_1) \leq \alpha$ implies $\mathcal{T}_n^{L} \geq t_1$ and $\mathcal{Q}_n'(t_2) \leq \beta$ implies $\mathcal{T}_n^{U} \leq t_2$. Since $\mathcal{Q}_n(t_1) \to 0$, there exists a natural number N_{t_1} such that $\mathcal{Q}_n(t_1) \leq \alpha$ for all $n \geq N_{t_1}$, and hence, $\mathcal{T}_n^{L} \geq t_1$ for all $n \geq N_{t_1}$. Similarly, since $\mathcal{Q}_n'(t_2) \to 0$, there exists a natural number N_{t_2} such that

$\mathcal{Q}'_n(t_2) \leq \beta$ for all $n \geq N_{t_2}$, and hence, $\mathcal{T}_n^U \leq t_2$ for all $n \geq N_{t_2}$. Therefore,

$$\mathcal{T}_n^U \leq t_2 < t_1 \leq \mathcal{T}_n^L \text{ for all } n \geq \max\{N_{t_1}, N_{t_2}\}.$$

This proves that the set $\{n \in \mathbb{N} : \mathcal{T}_n^L > \mathcal{T}_n^U\}$ is nonempty. □

Proof of Lemma 8

Proof. Let us define

$$\hat{Q}_n(t) := \frac{\sum_{i=1}^m \hat{t}_{ni} \mathbb{1}\left(\hat{t}_{ni} \leq t\right)}{\sum_{i=1}^m \mathbb{1}\left(\hat{t}_{ni} \leq t\right) \vee 1} \text{ and } \hat{Q}'_n(t) := \frac{\sum_{i=1}^m (1 - \hat{t}_{ni}) \mathbb{1}\left(\hat{t}_{ni} > t\right)}{\sum_{i=1}^m \mathbb{1}\left(\hat{t}_{ni} > t\right) \vee 1}.$$

Following similar arguments as in the proof of Lemma 6, note that $\hat{Q}_n(t)$ is a right-continuous non-decreasing function of t, and $\hat{t}_n^L = \sup\{t \in [0,1] : \hat{Q}_n(t) \leq \alpha\}$. Following the proof of Theorem 13, we can show as $m \to \infty$,

$$\text{Var}\left(\frac{1}{m}\sum_{i=1}^m \hat{t}_{ni} \mathbb{1}\left(\hat{t}_{ni} \leq t\right)\right) \to 0 \text{ and } \text{Var}\left(\frac{1}{m}\sum_{i=1}^m \mathbb{1}\left(\hat{t}_{ni} \leq t\right)\right) \to 0.$$

Using Assumption 6, it can be shown that

$$\mathbb{E}\left(\hat{t}_{n1} \mathbb{1}\left(\hat{t}_{n1} \leq t\right)\right) \to \mathbb{E}\left(t_{n1} \mathbb{1}\left(t_{n1} \leq t\right)\right) = \pi_0 \mathbb{P}_0\left(t_{n1} \leq t\right), \text{ as } m \to \infty.$$

and $\mathbb{P}\left(\hat{t}_{n1} \leq t\right) \to \mathbb{P}\left(t_{n1} \leq t\right)$, as $m \to \infty$. Combining these facts, we have for $t \in (0,1)$,

$$\hat{Q}_n(t) = \frac{\sum_{i=1}^m \hat{t}_{ni} \mathbb{1}\left(\hat{t}_{ni} \leq t\right)}{\sum_{i=1}^m \mathbb{1}\left(\hat{t}_{ni} \leq t\right) \vee 1} \xrightarrow{\mathbb{P}} \frac{\pi_0 \mathbb{P}_0\left(t_{n1} \leq t\right)}{\mathbb{P}\left(t_{n1} \leq t\right)} = \mathcal{Q}_n(t), \text{ as } m \to \infty.$$

Since $\hat{t}_n^L = \sup\{t \in [0,1] : \hat{Q}_n(t) \leq \alpha\}$, $\mathcal{T}_n^L = \sup\{t \in [0,1] : \mathcal{Q}_n(t) \leq \alpha\}$ and $\hat{Q}_n(t) \xrightarrow{\mathbb{P}} \mathcal{Q}_n(t)$, we use Lemma A.5 of Sun and Cai (2007a) to claim that $\hat{t}_n^L \xrightarrow{\mathbb{P}} \mathcal{T}_n^L$ as $m \to \infty$. To prove the second part, we can similarly show that $\hat{t}_n^U = \inf\{t \in [0,1] : \hat{Q}'_n(t) \leq \beta\}$ and for $t \in (0,1)$,

$$\hat{Q}'_n(t) = \frac{\sum_{i=1}^m (1 - \hat{t}_{ni}) \mathbb{1}\left(\hat{t}_{ni} > t\right)}{\sum_{i=1}^m \mathbb{1}\left(\hat{t}_{ni} > t\right) \vee 1} \xrightarrow{\mathbb{P}} \frac{\pi_0 \mathbb{P}_0\left(t_{n1} > t\right)}{\mathbb{P}\left(t_{n1} > t\right)} = \mathcal{Q}'_n(t), \text{ as } m \to \infty.$$

Since $\hat{Q}'_n(t)$ is a non-increasing step function that is not a continuous function of t (only left-continuous), following similar arguments as in the last part of the proof of Lemma 2, we can show that $\hat{t}_n^U \xrightarrow{\mathbb{P}} \mathcal{T}_n^U$ as $m \to \infty$. We omit the details to avoid repetitions of similar arguments. □

Proof of Lemma 9

Proof. First note that for any sample size $n \geq k$, $\sum_{i=1}^{q} t_n^{(i)}/q$, is increasing in $q \in [m]$. The attained FDR of the sequential test $(\tau, \mathbf{D})$ is

$$\begin{aligned}
\text{FDR} &= \mathbb{E}\left(\frac{\sum_{i=1}^m (1-\theta_i)D_i}{(\sum_{i=1}^m D_i) \vee 1}\right) \\
&= \mathbb{E}_{S_\tau}\left(\mathbb{E}_{\theta|S_\tau}\left(\frac{\sum_{i=1}^m (1-\theta_i)D_i}{(\sum_{i=1}^m D_i) \vee 1}\middle| S_\tau\right)\right) \\
&= \mathbb{E}_{S_\tau}\left(\frac{\sum_{i=1}^m (1-\mathbb{E}(\theta_i|S_\tau))D_i}{(\sum_{i=1}^m D_i) \vee 1}\right) \\
&= \mathbb{E}_{S_\tau}\left(\frac{\sum_{i=1}^m t_{\tau i} D_i}{(\sum_{i=1}^m D_i) \vee 1}\right) \\
&= \mathbb{E}_{S_\tau}\left(\frac{1}{l \vee 1}\sum_{i=1}^{l} t_\tau^{(i)}\right) \leq \mathbb{E}_{S_\tau}\left(\frac{1}{r_\tau \vee 1}\sum_{i=1}^{r_\tau} t_\tau^{(i)}\right) \leq \alpha.
\end{aligned}$$

The first inequality holds as $l \in [r_\tau]$ and the second inequality follows from (3.3). Following similar calculations, the attained mFDR becomes

$$\text{mFDR} = \frac{\mathbb{E}(\sum_{i=1}^m (1-\theta_i)D_i)}{\mathbb{E}(\sum_{i=1}^m D_i)} = \frac{\mathbb{E}_{S_\tau}(\sum_{i=1}^l t_\tau^{(i)})}{\mathbb{E}_{S_\tau}(l)}.$$

Since $l \leq r_\tau$, we have $\sum_{i=1}^l t_\tau^{(i)}/l \leq \sum_{i=1}^{r_\tau} t_\tau^{(i)}/r_\tau \leq \alpha$. Therefore, $\sum_{i=1}^l t_\tau^{(i)} \leq \alpha l$, i.e., $\mathbb{E}_{S_\tau}(\sum_{i=1}^l t_\tau^{(i)}) \leq \alpha \mathbb{E}_{S_\tau}(l)$. This proves mFDR $\leq \alpha$. □

Proof of Lemma 10

Proof. Note that for any sample size $n \geq k$, $\sum_{i=1}^{q}(1-t_n^{(m-i+1)})/q$ is increasing in $q \in [m]$. The attained FNR of the sequential test $(\tau, \mathbf{d})$ is

$$\begin{aligned}
\text{FNR} &= \mathbb{E}\left(\frac{\sum_{i=1}^{m}(1-d_i)\theta_i}{(\sum_{i=1}^{m}(1-d_i))\vee 1}\right) \\
&= \mathbb{E}_{S_\tau}\left(E_{\theta|S_\tau}\left(\frac{\sum_{i=1}^{m}(1-d_i)\theta_i}{(\sum_{i=1}^{m}(1-d_i))\vee 1}\middle| S_\tau\right)\right) \\
&= \mathbb{E}_{S_\tau}\left(\frac{\sum_{i=1}^{m}(1-d_i)\mathbb{E}(\theta_i|S_\tau)}{(\sum_{i=1}^{m}(1-d_i))\vee 1}\right) \\
&= \mathbb{E}_{S_\tau}\left(\frac{\sum_{i=1}^{m}(1-t_{\tau i})(1-d_i)}{(\sum_{i=1}^{m}(1-d_i))\vee 1}\right) \\
&= \mathbb{E}_{S_\tau}\left(\frac{1}{l\vee 1}\sum_{i=1}^{l}(1-t_\tau^{(m-i+1)})\right) \\
&\leq \mathbb{E}_{S_\tau}\left(\frac{1}{a_\tau\vee 1}\sum_{i=1}^{a_\tau}(1-t_\tau^{(m-i+1)})\right) \leq \beta,
\end{aligned}$$

where the first inequality holds as $l \leq a_\tau$ and the second inequality follows from (3.4). Following similar calculations, the attained mFNR becomes

$$\text{mFNR} = \frac{\mathbb{E}(\sum_{i=1}^{m}\theta_i(1-d_i))}{\mathbb{E}(\sum_{i=1}^{m}(1-d_i))} = \frac{\mathbb{E}_{S_\tau}(\sum_{i=1}^{l}(1-t_\tau^{(m-i+1)}))}{\mathbb{E}_{S_\tau}(l)}.$$

Since $l \leq a_\tau$, we have $\sum_{i=1}^{l}(1-t_\tau^{(m-i+1)})/l \leq \sum_{i=1}^{a_\tau}(1-t_\tau^{(m-i+1)})/a_\tau \leq \beta$. Therefore, $\sum_{i=1}^{l}(1-t_\tau^{(m-i+1)}) \leq \beta l$, i.e., $\mathbb{E}_{S_\tau}(\sum_{i=1}^{l}(1-t_\tau^{(m-i+1)})) \leq \beta\mathbb{E}_{S_\tau}(l)$. This concludes the proof of mFNR $\leq \beta$. □

Proof of Lemma 11

Proof. We showed in Lemma 6 that $Q_n(t) \to \mathcal{Q}_n(t)$ as $m \to \infty$ for prefixed $n \in \mathbb{N}$ and $t \in (0,1)$. Since $Q_n(t)$ is non-decreasing in t, $\mathcal{Q}_n(t)$ is also non-decreasing in t. Since by definition $\mathcal{Q}_n(t)$ is also a continuous function of t (as t_{ni} is a continuous variable), $\sup_{t\in(0,1)}\mathcal{Q}_n(t) = \mathcal{Q}_n(1) = \pi_0$. Note that $\mathcal{Q}_n(t)$ takes values in the interval $[0,\pi]$ and due to continuity, for $0 < \alpha < \pi_0$, there exists at least one point at which, $\mathcal{Q}_n(t) = \alpha$. So, there is a point $t_\alpha := \sup\{t : \mathcal{Q}_n(t) = \alpha\}$, for which $\mathcal{Q}_n(t_\alpha) = \alpha$ and $\alpha < \mathcal{Q}_n(t) \leq \pi_0$ for all $t > t_\alpha$. Clearly, $\mathcal{T}_n^L = \sup\{t \in (0,1) : \mathcal{Q}_n(t) \leq \alpha\} =$

$\sup(0, t_\alpha] = t_\alpha < 1$. Therefore, $0 < \mathcal{T}_n^L < 1$,

$$\mathcal{Q}_n(\mathcal{T}_n^L) = \alpha \text{ and } \alpha < \mathcal{Q}_n(t) \leq \pi_0 \text{ for all } t > \mathcal{T}_n^L. \tag{3.35}$$

A similar argument shows that, $0 < \mathcal{T}_n^U < 1$ and for $0 < \beta < 1 - \pi_0$,

$$\mathcal{Q}'_n(\mathcal{T}_n^U) = \beta \text{ and } \beta < \mathcal{Q}'_n(t) \leq 1 - \pi_0 \text{ for all } t < \mathcal{T}_n^U. \tag{3.36}$$

Now, suppose $\mathcal{T}_n^L = \mathcal{T}_n^U = t_n^*$, say, for some $n \in \mathbb{N}$. Then, $t_n^* \in (0, 1)$,

$$\mathcal{Q}_n(t_n^*) = \frac{\pi_0 \mathbb{P}_0(t_{n1} \leq t_n^*)}{\mathbb{P}(t_{n1} \leq t_n^*)} = \alpha \text{ and } \mathcal{Q}'_n(t_n^*) = \frac{(1-\pi_0)\mathbb{P}_1(t_{n1} > t_n^*)}{\mathbb{P}(t_{n1} > t_n^*)} = \beta.$$

This implies

$$\begin{aligned} \mathbb{P}_1(t_{n1} \leq t_n^*) &= \frac{(1-\alpha)\pi_0}{\alpha(1-\pi_0)} \mathbb{P}_0(t_{n1} \leq t_n^*) \text{ and} \\ (1 - \mathbb{P}_1(t_{n1} \leq t_n^*)) &= \frac{\beta\pi_0}{(1-\beta)(1-\pi_0)} (1 - \mathbb{P}_0(t_{n1} \leq t_n^*)). \end{aligned}$$

Using algebraic calculations, we obtain

$$\begin{aligned} \mathbb{P}_0(t_{n1} \leq t_n^*) &= \frac{\alpha(1-\pi_0-\beta)}{\pi_0(1-\alpha-\beta)} \text{ and} \\ \mathbb{P}_1(t_{n1} \leq t_n^*) &= \frac{(1-\alpha)(1-\pi_0-\beta)}{(1-\pi_0)(1-\alpha-\beta)}. \end{aligned}$$

We already showed $\alpha < \mathcal{Q}_n(t) \leq \pi_0$ for all $t > t_n^*$ and $\beta < \mathcal{Q}'_n(t) \leq (1-\pi_0)$ for all $t < t_n^*$. Hence, conditions $(i) - (iv)$ are satisfied. To prove the *only if* part, suppose that for some $n \in \mathbb{N}$ there exists a $t_n^* \in (0, 1)$ such that conditions $(i) - (iv)$ in Lemma 11 hold. Conditions (i) and (ii) yield $\mathcal{Q}_n(t_n^*) = \alpha$ and $\mathcal{Q}'_n(t_n^*) = \beta$ due to earlier calculations. Using this and the condition (iii), we have $\mathcal{Q}_n(t_n^*) = \alpha$ and $\mathcal{Q}_n(t_n^*) > \alpha$ for all $t > t_n^*$, which implies $\mathcal{T}_n^L := \sup\{t \in [0,1] : \mathcal{Q}_n(t) \leq \alpha\} = t_n^*$. Similarly, $\mathcal{Q}'_n(t_n^*) = \beta$ and $\mathcal{Q}'_n(t_n^*) > \beta$ for all $t < t_n^*$ implies $\mathcal{T}_n^U := \inf\{t \in [0,1] : \mathcal{Q}'_n(t) \leq \beta\} = t_n^*$. Hence, $\mathcal{T}_n^L = \mathcal{T}_n^U$. Hence, the proof of Lemma 11 is complete. □

Proof of Lemma 12

Proof. For any fixed $n \in \mathbb{N}$, since $\mathcal{Q}_n(t)$ is a continuous non-decreasing function on $[0,1]$ taking values in $[0,\pi_0]$ with $0 < \alpha < \pi_0 < 1$, we have from (3.35) that

$$\mathcal{Q}_n(\mathcal{T}_n^L) = \frac{\pi_0 \mathbb{P}_0(t_{n1} \leq \mathcal{T}_n^L)}{\mathbb{P}(t_{n1} \leq \mathcal{T}_n^L)} = \alpha \in (0,1). \tag{3.37}$$

This implies $\mathbb{P}_0(t_{n1} \leq \mathcal{T}_n^L) > 0$, and hence,

$$\mathbb{P}(t_{n1} \leq \mathcal{T}_n^L) = \pi_0 \mathbb{P}_0(t_{n1} \leq \mathcal{T}_n^L) + (1-\pi_0)\mathbb{P}_1(t_{n1} \leq \mathcal{T}_n^L) > 0.$$

Now, if $\mathbb{P}(t_{n1} \leq \mathcal{T}_n^L) = 1$, then $\mathbb{P}_0(t_{n1} \leq \mathcal{T}_n^L) = 1$, and hence, $\mathcal{Q}_n(\mathcal{T}_n^L) = \pi_0 = \alpha$ using (3.37). This is a contradiction since $\alpha < \pi_0$, and therefore, we have $0 < \mathbb{P}(t_{n1} \leq \mathcal{T}_n^L) < 1$. Now, for any fixed $n \in \mathbb{N}$, since $\mathcal{Q}_n'(t)$ is a continuous non-increasing function on $[0,1]$ taking values in $[0,1-\pi_0]$ with $0 < \beta < 1-\pi_0 < 1$, we have from (3.36) that

$$\mathcal{Q}_n'(\mathcal{T}_n^U) = \frac{(1-\pi_0)\mathbb{P}_1(t_{n1} > \mathcal{T}_n^U)}{\mathbb{P}(t_{n1} > \mathcal{T}_n^U)} = \beta \in (0,1). \tag{3.38}$$

Using (3.38) and similar arguments as above, $0 < \mathbb{P}(t_{n1} > \mathcal{T}_n^U) < 1$. Hence, $0 < \mathbb{P}(t_{n1} \leq \mathcal{T}_n^U) < 1$. This completes the proof of Lemma 12. □

Chapter 4

Large-scale Sequential Multiple Testing with Asynchronous Termination of Sampling

4.1 Data Description and Problem Setup

Suppose m different experiments are conducted successively on subjects arriving in sequence. For example, we may consider a clinical trial to test m endpoints with patients arriving sequentially until a certain verdict about the drug is reached. The patients may arrive one at a time or in groups. Suppose the experiments are labeled as $1, 2, 3, \ldots, m$. For each experiment i, we aim to test a null hypothesis H_{0i} against an alternative hypothesis H_{1i}. In this chapter, we use the terms "experiment" and "hypothesis" interchangeably. In the context of the clinical trial example, H_{0i} represents the hypothesis that the drug has no effect with respect to the i-th endpoint. As an outcome of m different sequential experiments, m data streams are observed where the i-th data stream is observed until enough evidence to accept/reject H_{0i} is accrued. So, at any interim stage of sampling, some hypotheses are already dropped (i.e., accepted/rejected) and the rest are "active" hypotheses (i.e., not accepted/rejected yet) for which corresponding experiments are continued with further data gathering. A similar sampling strategy is considered by De and Baron (2012a), De and Baron (2012b), De and Baron (2015), Song and Fellouris (2017), Song and Fellouris (2019), and He and Bartroff (2021) and Chen et al. (2023).

Suppose there are m data streams: $\mathbf{X}_{(1)}, \mathbf{X}_{(2)}, \ldots, \mathbf{X}_{(m)}$ for testing the m pairs of

hypotheses H_{0i} against H_{1i}, $i \in [m]$. Each data stream $\mathbf{X}_{(i)}$ is the sequence of observations corresponding to the i-th experiment, i.e., $\mathbf{X}_{(i)} = (X_{1i}, X_{2i}, \ldots, X_{ni}, \ldots)$. The observation X_{ni} is allowed to be single-valued or a vector to address the subjects that arrive in groups. At "time-point" n, first n elements from each data stream is at our disposal. A summary statistic S_{ni} is obtained as a function of $X_{1i}, X_{2i}, \ldots, X_{ni}$. Suppose I_n denotes the active set of hypotheses at time n. We test H_{0i} against H_{1i}, $i \in I_n$ based on the set of summary statistics $\{S_{ni} | i \in I_n\}$. For each of these active experiments $i \in I_n$, at time n, we have three choices. Either we accept H_{0i} or reject it or fail to make any decision and look for more evidence in favour of or against H_{0i}. In case we accept/reject H_{0i} based on n samples, we drop H_{0i} from the set of active hypotheses and stop making further observations on the i-th data stream. Active experiments in I_n, for which corresponding hypotheses are not accepted or rejected within time n, are included in I_{n+1}, the active set of hypotheses at time $n+1$. Clearly, $I_n \supseteq I_{n+1}$ for all time points $n \in \mathbb{N}$. This sampling process is continued until we have a decision for all the hypotheses, i.e., the active set of hypotheses becomes empty.

Once the testing procedure is exhausted, for each hypothesis $i \in [m]$, we have a decision rule δ_i that takes the value 1 if H_{0i} is rejected and 0 if it is accepted. The stopping time T_i is defined as the time point n when we made a decision about (either accept or reject) H_{0i}. A sequential multiple testing procedure with asynchronous termination of sampling is denoted by the pair $(\mathbf{T}, \boldsymbol{\delta})$, with $\mathbf{T} = (T_1, T_2, \ldots, T_m)$ and $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_m)$. Suppose, $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$ denote the true states of the m hypotheses. Table 1.1 and the subsequent discussion in Section 1.1.2 together refer to different error measures relating to multiple testing problems. In sequential multiple testing, it is possible to simultaneously control certain measures of both false positives and false negatives. In this chapter, our goal is to control the pair of error measures, FDR and FNR, at predetermined levels (say α and β, $0 < \alpha, \beta < 1$ respectively). In the previous chapter, where we addressed simultaneous stopping of sampling, the objective was to minimize the expected number of subjects on whom observations were made for all hypotheses. In contrast, in this chapter, our aim is to minimize the expected total number of experiments, that is, $\mathbb{E}(\sum_{i=1}^{m} T_i)$.

4.2 Oracle Test Statistics

Keeping in mind that the active set of hypotheses is shrinking with each new time point, we have to be careful about choosing the test statistics. Following the discussions of Section 3.2, we define the oracle test statistic corresponding to the i-th data stream at time point n as the probability of H_{0i} being true given that the summary statistics calculated up to n. Recall that at an interim stage n, the collection of observed summary statistics corresponding to the active hypotheses is $A_{1n} := \{S_{nj} : j \in I_n\}$. At that interim stage n, consider the set of "inactive" hypotheses (i.e., those already accepted/rejected) $[m] \setminus I_n$ and suppose that, for each $j \in [m] \setminus I_n$, the j-th inactive hypothesis was dropped at n_j-th stage, i.e., the observed stopping time for the j-th inactive hypothesis is $n_j\ (< n)$. The collection of observed summary statistics corresponding to the inactive hypotheses is $A_{2n} := \{S_{n_j j} : j \in [m] \setminus I_n\}$. Therefore, the entire collection of observed summary statistics up to the n-th stage is $A_n := A_{1n} \cup A_{2n}$. For each $i \in I_n$, we define the oracle statistic at the n-th stage as

$$t_{ni,A_n} = \mathbb{P}(\theta_i = 0 | A_n). \tag{4.1}$$

Suppose that the joint distribution of the elements of A_n conditioned on $\theta_i = j$, for $j \in \{0, 1\}$, is given by $f^*_{jn}(A_n)$. Further, let $\mathbb{P}(\theta_i = 0) = \pi_{i0}$. Then,

$$t_{ni,A_n} = \frac{\pi_{i0} f^*_{0n}(A_n)}{\pi_{i0} f^*_{0n}(A_n) + (1 - \pi_{i0}) f^*_{1n}(A_n)} \text{ for each } i \in I_n.$$

These oracle test statistics involve unknown parameters that are difficult to estimate without further assumptions and cannot be used in practice. Hence, we consider an independent two-group mixture model assumption for the summary statistics in the spirit of Assumption 3 of Chapter 3.

Assumption 7. *1.* $\theta_1, \ldots, \theta_m \overset{\text{i.i.d.}}{\sim} Bernoulli(1 - \pi_0)$.

2. *For any set of m time points, $n_1, n_2, \ldots, n_m \in \mathbb{N}$, conditioned on the parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$, the summary statistics $\{S_{n_1 1}, S_{n_2 2}, \ldots, S_{n_m m}\}$ are mutually independent.*

3. *For each* $i \in I_n$, $S_{ni}|\boldsymbol{\theta} \sim (1-\theta_i)f_{0n} + \theta_i f_{1n}$, *where* f_{0n} *and* f_{1n} *represent the null and alternate densities, respectively.*

Here, the integers $n_1, n_2, \ldots, n_m$ represent different time points. Part 1 and 2 of Assumption 7 jointly imply that the marginal distributions of the summary statistics remain independent even when they are calculated based on the corresponding data streams up to different time points. In part 3 of Assumption 7, the forms of the marginal distributions of the summary statistics are described. Under Assumption 7, the oracle test statistic t_{ni,A_n}, for each $i \in I_n$, takes a simplified form.

Lemma 13. *If Assumption 7 holds, then for each* $i \in I_n$

$$t_{ni} = \frac{\pi_0 f_{0n}(S_{ni})}{f_n(S_{ni})}, \tag{4.2}$$

where $f_n(x) = \pi_0 f_{0n}(x) + (1-\pi_0)f_{1n}(x)$ *is the marginal mixture distribution of* S_{ni}.

We note that, under assumption 7, the value of t_{ni,A_n} for $i \in I_n$ depends only on the i-th data stream, not on any other data streams. Moreover, we observe that this value of t_{ni,A_n} is the same as t_{ni} in (3.2) under Assumption 3. Throughout this chapter, we will assume that the Assumption 7 is true, and therefore, we have used the simplified notation t_{ni} to denote the oracle test statisitc at the time point n corresponding to the data stream $i \in I_n$.

It is clear from the definition that larger value of t_{ni} provides stronger evidence in favour of H_{0i} and we tend to accept H_{0i}. On the other hand, we prefer to reject H_{0i} if the value of t_{ni} is small. Therefore, after computing these statistics at any stage n, we arrange them in ascending order $t_n^{(1)} \leq t_n^{(2)} \leq \cdots \leq t_n^{(|I_n|)}$. To develop a multiple testing procedure, our strategy would be to search for a pair of upper and lower cut-off values c_U and c_L such that the set of hypotheses $\{i \in I_n | t_n^{(i)} < c_L\}$ are rejected and the set of hypotheses $\{i \in I_n | t_n^{(i)} > c_U\}$ are accepted.

4.3 Oracle Tests with Exact Error Control

In this section, we propose two adaptive oracle tests. First, an oracle stagewise test is proposed as an adaptation of the OI test described in chapter 3. Second, we propose a generalized version of the oracle stagewise test, referred to as the oracle composite

(OC) test, to allow for the development of a data-driven analog of the OC test with asymptotic control of FDR and FNR.

4.3.1 The Oracle Stagewise (OS) Test

First, let us consider the oracle intersection (OI) test described in chapter 3. This test was shown to produce the attained values for FDR and FNR (or mFDR and mFNR) very close to the predetermined levels α and β respectively. This is the consequence of the fact that "the expected value of the arithmetic mean (AM) of the oracle test statistics corresponding to the rejected hypotheses is equal to the FDR". If $\mathfrak{R}$ is the set of rejected hypotheses, $\mathfrak{A} = [m] \setminus \mathfrak{R}$ is the set of accepted hypotheses, and $R = |\mathfrak{R}|$, then

$$\text{FDR} = \mathbb{E}\left(\frac{1}{R \vee 1} \sum_{i \in \mathfrak{R}} t_{ni}\right),$$

and

$$\text{FNR} = \mathbb{E}\left(\frac{1}{(m-R) \vee 1} \sum_{i \in \mathfrak{A}} (1 - t_{ni})\right).$$

In the Oracle-Intersection (OI) test, at each time point n, we select the number of potential alternatives r_n so as to maximize the number of rejections while maintaining control over the arithmetic mean of the oracle test statistics at the desired level. This strategy ensures that the FDR is controlled at the pre-specified level. Similarly, the number of potential nulls a_n is chosen to optimize the acceptance decisions. According to the OI test, no hypothesis is accepted or rejected until all hypotheses are simultaneously decided, that is, until the first time point n for which $a_n + r_n \geq m$.

In the present sampling scheme, a natural adaptation of the OI test is to reject all r_n potential alternatives and accept all a_n potential nulls at time n from the current active set I_n. The corresponding data streams are then dropped from further consideration at the $(n+1)$-th stage. This procedure is repeated until the active set of hypotheses becomes empty.

Since decisions are made in a sequential, stagewise fashion rather than simultaneously across all data streams, we refer to this method as the *Oracle Stagewise (OS)* test procedure.

Oracle Stagewise (OS) Test Procedure

1. We start with prefixed $\alpha, \beta \in (0,1)$ and a suitable pilot sample size $n = k$. All the hypotheses are active now, i.e., $I_n = [m]$ and $\mathfrak{R} = \mathfrak{A} = \emptyset$.

2. At time point n, we calculate the oracle test statistics t_{ni} for the active set of hypotheses I_n and arrange them in ascending order: $t_n^{(1)} \leq t_n^{(2)} \leq \cdots \leq t_n^{(|I_n|)}$.

3. We define the number of potential alternatives r_n as the maximum number among the remaining $|I_n|$ hypotheses for which the cumulative mean of the ordered oracle test statistics starting from the smallest is just smaller than α, i.e.,

$$r_n = \begin{cases} \max\left\{1 \leq q \leq |I_n| \middle| \frac{1}{q}\sum_{i=1}^{q} t_n^{(i)} \leq \alpha\right\} & \text{if } t_n^{(1)} \leq \alpha, \\ 0 & \text{if } t_n^{(1)} > \alpha. \end{cases}$$

   The number of potential nulls a_n is defined as the maximum number of hypotheses among the active set such that the cumulative mean of the ordered oracle test statistics starting from the largest value is larger than β, i.e.,

$$a_n = \begin{cases} \max\left\{1 \leq q \leq |I_n| \middle| \frac{1}{q}\sum_{i=1}^{q} t_n^{(|I_n|-i+1)} \geq 1-\beta\right\} & \text{if } t_n^{(|I_n|)} \geq 1-\beta, \\ 0 & \text{if } t_n^{(|I_n|)} < 1-\beta. \end{cases}$$

4. If $a_n + r_n < |I_n|$, there are some hypotheses in the active set for which no decisions can be made at the n-th stage. In this case, we reject (accept) the null hypotheses corresponding to the r_n smallest (a_n largest) ordered oracle test statistics. The set of rejected hypotheses at time point n is $\mathcal{R}_n = \{i \in I_n | t_{ni} \leq t_n^{(r_n)}\}$. Similarly, the set of accepted hypotheses at time point n is defined as $\mathcal{A}_n = \{i \in I_n | t_{ni} \geq t_n^{(|I_n|-a_n+1)}\}$. We update the set of active hypotheses, set of accepted hypotheses and the set of rejected hypotheses accordingly, i.e., $I_{n+1} = I_n \setminus \{\mathcal{R}_n \cup \mathcal{A}_n\}$, $\mathfrak{R} = \mathfrak{R} \cup \mathcal{R}_n$, and $\mathfrak{A} = \mathfrak{A} \cup \mathcal{A}_n$. For the updated set of active data streams I_{n+1}, we observe another sample and repeat steps 2-5.

5. If $a_n + r_n \geq |I_n|$, sampling terminates as we have a potential decision for all the active data streams at time n. In this case, we choose an integer s with $|I_n| - a_n \leq s \leq r_n$ and define $\mathcal{R}_n = \{i \in I_n | t_{ni} \leq t_n^{(s)}\}$, the set of rejected

hypotheses, and $\mathcal{A}_n = \{i \in I_n | t_{ni} \geq t_n^{(s+1)}\}$, the set of accepted hypotheses. We update $\mathfrak{R} = \mathfrak{R} \cup \mathcal{R}_n$, $\mathfrak{A} = \mathfrak{A} \cup \mathcal{A}_n$ and stop.

6. In steps 4 and 5, at each time point n where a decision has been made, we define the stopping time for testing H_{0i} as $T_i := n$ if $i \in \mathcal{R}_n \cup \mathcal{A}_n$. The decision rules are defined as $\delta_i = 1(0)$ if $i \in \mathfrak{R}$ ($i \in \mathfrak{A}$) for all $i \in [m]$. The oracle stagewise test is given by the pair $(\mathbf{T}, \boldsymbol{\delta})$ with $\mathbf{T} = (T_1, T_2, \ldots, T_m)$ and $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_m)$. Total number of subjects required for completing the OS test is $T_{max} = \max\{T_1, T_2, \ldots, T_m\}$.

Note that there are no inherent fixed upper and lower boundaries (for the test statistics) associated with the OS test. However, we can propose adaptive boundaries that changes with changing time points in the spirit of the adaptive boundaries of the OI test in Section 3.2.1. If we reject r_n hypotheses at time point $n \geq k$ such that $t_{ni} \leq t_n^{(r_n)}$, then we can consider the $(r_n + 1)$-th order statistic $t_n^{(r_n+1)}$ to be the adaptive lower boundary t_n^L, and we reject each null hypotheses whose corresponding t-value lie below t_n^L. Similarly, for each time point $n \geq k$, we can consider the $(|I_n| - a_n)$-th order statistic $t_n^{(|I_n|-a_n)}$ as the adaptive upper boundary t_n^U and accept each null hypotheses whose corresponding t-value is higher than t_n^U.

4.3.2 Oracle Composite (OC) Test

For the OI test, described in Section 3.2.1, the stopping time can be restated in terms of the lower and upper cut-off values t_n^L and t_n^U (defined in (3.6)). As stated in (3.7), once we have $t_n^L > t_n^U$, the test attains termination. Lemmas 6 and 7 ensure that for a large number of hypotheses, $t_n^L > t_n^U$ is achieved in a finite amount of time (Theorem 10). Here, for example, the lower boundary t_n^L is defined as the $(r_n + 1)$-th order statistic, where r_n is the maximum number k, for which the mean of the smallest k many t_{ni} values is below α. Due to assumptions 3 and 4, the oracle test statistics corresponding to the alternative hypotheses move closer and closer to 0 as the sample size grows. This allows some larger order statistics to be included in the criteria for r_n, sending the lower boundary t_n^L towards 1. A similar "pooling effect of the extreme values" is observed for the upper boundary t_n^U towards 0. We observe this "pooling effect" illustrated in Figure 3.1. The presence of the full set of data

streams enables this "pooling effect" which, in turn, shrinks the "continue sampling region (t_n^L, t_n^U)" and thereby helps reach termination in a finite amount of time.

For the OS test, however, we keep dropping the data streams as the sampling and testing progresses, making the active set of hypotheses gradually shrink to an empty set. Especially, all the data streams with extreme t_{ni} values are dropped. Therefore, such a "pooling effect" is not observed in this case. Eventually, when the active set of hypotheses is very small, the test procedure may not make any decision for a long time as the "continue sampling region" does not shrink as quickly as before, resulting in a large maximum stopping time T_{max}. Thus, the test requires additional time and cost and has limited practical utility.

Additionally, as we will observe in the next section, the application of a data-driven version for composite alternatives requires the estimation of the oracle test statistics. But, when the set of active hypotheses I_n becomes very small (which is bound to happen for the oracle stagewise test), the estimates of oracle test statistics become unreliable.

To overcome the aforementioned limitations, we consider a prefixed integer-valued threshold m', with $0 \leq m' \leq m$. The procedure initially applies the OS test to the complete set of hypotheses. The active set is updated sequentially according to the OS testing rule until its cardinality decreases to m' or below. At this stage, the elimination process is terminated, and the oracle intersection (OI) test is applied to the remaining active data streams. This resulting hybrid testing procedure is termed the *Oracle Composite (OC) test*. The test procedure is described below.

Oracle Composite (OC) test:

1. We start with a pilot sample size $n = k(\in \mathbb{N})$. Set $I_k = [m]$, $\mathfrak{R}^C = \emptyset$, and $\mathfrak{A}^C = \emptyset$.

2. At each time point n, given the active set I_n, we calculate the oracle test statistics t_{ni} as defined in (4.2) and order them as $t_n^{(1)} \leq \cdots \leq t_n^{(|I_n|)}$. We define the number of potential rejections r_n and the number of potential acceptances a_n

respectively as:

$$r_n = \max\left\{1 \le q \le |I_n| \mid \frac{1}{q}\sum_{i=1}^{q} t_n^{(i)} \le \alpha\right\},$$
$$a_n = \max\left\{1 \le q \le |I_n| \mid \frac{1}{q}\sum_{i=1}^{q} t_n^{(|I_n|-i+1)} \ge 1-\beta\right\}.$$

Further, we define $g_n := |I_n| - a_n - r_n$.

(a) If $g_n \ge m'$ we reject the r_n hypotheses with the smallest oracle test statistics and accept the a_n hypotheses with the largest oracle test statistics. Specifically:

$$\mathcal{R}_n^C = \{i \in I_n \mid t_{ni} \le t_n^{(r_n)}\}, \quad \mathcal{A}_n^C = \{i \in I_n \mid t_{ni} \ge t_n^{(|I_n|-a_n+1)}\}.$$

Then we update the sets $\mathfrak{R}^C \leftarrow \mathfrak{R}^C \cup \mathcal{R}_n^C$, $\mathfrak{A}^C \leftarrow \mathfrak{A}^C \cup \mathcal{A}_n^C$, and $I_{n+1} = I_n \setminus (\mathcal{R}_n^C \cup \mathcal{A}_n^C)$.

(b) If $0 < g_n < m'$, we do not drop any data streams further and apply the OI test described in Chapter 3. We set $\mathcal{R}_n^C = \emptyset$ and $\mathcal{A}_n^C = \emptyset$.

(c) Finally, if $g_n \le 0$, We make decisions for the remaining active hypotheses in I_n as follows. Choose s such that $I_n - a_n \le s \le r_n$, and define:

$$\mathcal{R}_n^C = \{i \in I_n \mid t_{ni} \le t_n^{(s)}\}, \quad \mathcal{A}_n^C = \{i \in I_n \mid t_{ni} \ge t_n^{(s+1)}\}.$$

and update the sets $\mathfrak{R}^C$, $\mathfrak{A}^C$, and $I_{n+1} = I_n \setminus (\mathcal{R}_n^C \cup \mathcal{A}_n^C)$ as before.

3. If $\mathcal{R}_n^C \cup \mathcal{A}_n^C \ne \phi$, define the stopping time $T_i^C := n$ for all $i \in \mathcal{R}_n^C \cup \mathcal{A}_n^C$, and decision rule $\delta_i^C = 1$ if $i \in \mathcal{R}_n^C$ and $\delta_i^C = 0$ if $i \in \mathcal{A}_n^C$.

We define the maximum stopping time as $T_{max}^C = \max\{T_1^C, T_2^C, \ldots, T_m^C\}$.

A mild consistency property, introduced in Assumption 4 of Chapter 3, ensures that as the number of observations increases, the test statistics provide increasingly strong evidence toward the truth. We adopt the same assumption in this chapter.

Assumption 8. *As $n \to \infty$, the oracle test statistic t_{ni} satisfies*

$$t_{ni} \xrightarrow{\mathbb{P}} 1 \quad \textit{under } H_{0i}, \qquad t_{ni} \xrightarrow{\mathbb{P}} 0 \quad \textit{under } H_{1i}, \qquad \textit{for all } i = 1, 2, \ldots, m.$$

Assumption 8 characterizes the asymptotic behavior of the sequence $\{t_{ni}\}$ as $n \to \infty$. While the active set I_n eventually shrinks, this assumption guarantees that for any hypothesis i that remains active, the test statistics provide reliable evidence. This is instrumental in establishing the validity of the Oracle stagewise and Composite rules: it ensures that the procedures terminate in finite time and that both FDR and FNR remain controlled at pre-specified levels, as formalized in the following theorem.

Theorem 14. *If Assumptions 7 and 8 hold, then the OS test* $(\mathbf{T}, \boldsymbol{\delta})$ *and the OC test* $(\mathbf{T}^C, \boldsymbol{\delta}^C)$ *yield:*

(i) $\mathbb{P}(T_{max} < \infty) = 1$ *and* $\mathbb{P}(T^C_{max} < \infty) = 1$,

(ii) $FDR \leq \alpha$ *and* $FNR \leq \beta$ *for any prefixed* $\alpha, \beta \in (0, 1)$.

4.4 Data-driven Stagewise Tests

To implement the oracle tests to real datasets, one must estimate the oracle test statistics t_{ni} for each time point n and each active hypothesis $i \in I_n$. Under Assumption 7,

$$t_{ni} = \frac{\pi_0 f_{0n}(S_{ni})}{f_n(S_{ni})},$$

where π_0 denotes the proportion of true null hypotheses among the m total hypotheses, f_{0n} is the null distribution, and $f_n = \pi_0 f_{0n} + (1 - \pi_0) f_{1n}$ is the marginal distribution of the summary statistic S_{ni}. The assumption further states that the S_{ni} are independent across $i \in [m]$. Under these assumptions, at each stage of sampling, the DI test described in Chapter 3 estimates π_0 following Cai and Jin (2010) and evaluates f_n using a kernel density approach applied to all m summary statistics. Since all of the m hypotheses are tested simultaneously at each stage of sampling, π_0 and f_n are estimated *consistently* based on these large number (m) of summary statistics, a reliable *consistent* estimate of t_{ni} is available at each stage of the DI test. In this section we will attempt to employ this idea for the oracle stagewise procedure described in the last section.

Data-driven Stagewise (DS) Test: In the OS test, where asynchronous termination of sampling is inevitable, at any interim stage n, only the summary statistics $\{S_{ni} : i \in I_n\}$ corresponding to the active set $I_n \subseteq [m]$ are observed. This leads to two main issues:

1. The proportion of true null hypotheses within I_n may differ for different n causing the estimator $\hat{\pi}_0$ based on $|I_n|$ many hypotheses to be unreliable.

2. Kernel-based estimation of f_n becomes unreliable because the size of I_n decreases as n increases.

Therefore, to incorporate a data driven test based on the OS test, we start with a large *pilot sample* of size k, where $I_k = [m]$. When both m and k are large, the resulting estimate $\hat{\pi}_{0k}$ of π_0, as proposed in Cai and Jin (2010), is stable and consistent. This estimate $\hat{\pi}_{0k}$ is then held fixed for all subsequent stages of sampling. In our simulation studies, we adopt a simple-versus-simple testing framework, assuming f_{0n} and f_{1n} to be known. Thus, with $\hat{\pi}_{0k}$ fixed, we estimate the i-th oracle test statistic for $i \in I_n$ at stage n as

$$\hat{t}_{ni} = \frac{\hat{\pi}_{0k} f_{0n}(S_{ni})}{\hat{\pi}_{0k} f_{0n}(S_{ni}) + (1 - \hat{\pi}_{0k}) f_{1n}(S_{ni})}. \tag{4.3}$$

The Data-driven Stagewise Test, henceforth referred to as the **DS test**, is constructed by replacing the oracle test statistics t_{ni} in the OS test described in Section 4.3.1 with the estimated statistics $\hat{t}_{ni}$ defined in (4.3).

Note that, following the pilot sampling stage, estimation of t_{ni} requires only the computation of summary statistics S_{ni} and evaluation of the expressions $f_{0n}(S_{ni})$ and $f_{1n}(S_{ni})$ for $i \in I_n$ as a fixed estimate $\hat{\pi}_{0k}$ is used at each stage. Since, the functional forms of f_{0n} and f_{1n} are known, a reliable estimate of t_{ni} is available even when very few data streams remain active. We implement this DS test on three simulated datasets described in Section 5.3 and report their performance in Tables 5.3–5.5.

Data-driven Composite (DC) Test: The DS test is effective for many testing scenarios, especially for simple null and simple alternative hypotheses, where estimating

$\hat{t}_{ni}$ as in (4.3) is nothing but evaluating known functions f_{0n} and f_{1n} at the corresponding summary statistics S_{ni}. However, it has significant limitations when dealing with composite hypotheses. In such cases, estimation of t_{ni} will require estimating the marginal distributions f_n based on the active set of data streams I_n. As sampling progresses, the OS procedure narrows the active set I_n gradually to the empty set. Hence, eventually, the active set I_n becomes insufficient for reliable estimation of the distribution f_{0n}, which is essential for estimating the oracle test statistics t_{ni}.

For the OC test described in Section 4.3.2, a cutoff value m' is chosen so that the procedure guarantees that the size of the active set remains above m', that is, $|I_n| > m'$, until the test terminates. This allows for reliable estimation of the marginal distribution f_n as well as the null proportion π_0 at each stage.

We consider the p-values $\{p_{ni}\}$ as the summary statistics S_{ni}, for which the null distribution is $U(0,1)$. Let $\hat{f}_n$ denote the kernel density estimate of the marginal density f_n based on the active set I_n, and let $\hat{\pi}^S_{0n}$ be the estimate of π_0 obtained using Storey's method (Storey, Taylor, and Siegmund, 2004) keeping the tuning parameter $\lambda = 0.5$, that is, if the number of p-values p_{ni} which exceed 0.5 at stage n is W_n, then,

$$\hat{\pi}^S_{0n} = \frac{W_n + 1}{0.5|I_n|}, \qquad n \in \mathbb{N}.$$

Then, under the restriction $|I_n| \geq m'$, the oracle test statistic t_{ni} is estimated by

$$\hat{t}^C_{ni} = \frac{\hat{\pi}^S_{0n}}{\hat{f}_n(p_{ni})}, \qquad i \in I_n,\ n \in \mathbb{N}. \tag{4.4}$$

The **Data-driven Composite (DC) test** is constructed by replacing the oracle test statistic t_{ni} in the OC test described in Section 4.3.2 with its estimate $\hat{t}^C_{ni}$ given in (4.4). Thus, the DC test serves as a fully data-driven analogue of the OC test and is particularly suitable for applications involving composite null and alternative hypotheses.

4.5 Proofs of Theorems

Proof of Theorem 14 :

Proof. **Proof of part (i)** Note that, under the assumption 7, the oracle test statistics $t_{ni}, i \in [m]$ are mutually independent. We define T^* as the minimum sample size such that $t_{ni} < \alpha$ for all false null hypotheses (i such that $\theta_i = 1$) and $t_{ni} > 1 - \beta$ for all true null hypotheses (i such that $\theta_i = 0$). Clearly, at such a time point n, for any subset I_n of $[m]$, by definition, $r_n + a_n = m$, i.e., the oracle stagewise and composite procedures stop. Inevitably, $T_{max} \leq T^*$ and $T^C_{max} \leq T^*$. But how can we say that T^* exists? To prove the existence of T^*, it is enough to show that the set $\{n \in \mathbb{N} | t_{ni} < \alpha$ for all i such that $\theta_i = 1$ and $t_{ni} > 1 - \beta$ for all i such that $\theta_i = 0\}$ is non-empty. So far we have considered m to be a finite number. Now, due to independence, for $0 < \epsilon_i < 1,\ i \in [m]$,

$$\begin{aligned}&\mathbb{P}(\{\cap_{i:\theta_i=0}\{1 - t_{ni} < \epsilon_i\}\} \cap \{\cap_{i:\theta_i=1}\{t_{ni} < \epsilon_i\}\}) \\ &\quad = \prod_{i:\theta_i=0} \mathbb{P}(1 - t_{ni} < \epsilon_i) \times \prod_{i:\theta_i=1} \mathbb{P}(t_{ni} < \epsilon_i).\end{aligned}$$

As n grows to ∞, by the virtue of Assumption 8, all the terms inside the two products converge to 1. In turn, the original probability converges to 1 for any choice of such $(\epsilon_1, \epsilon_2, \ldots, \epsilon_m)$. Consequently, we have shown that, as n grows to infinity, given any $\boldsymbol{\theta} \in \{0,1\}^m$,

$$(t_{n1}, t_{n2}, \cdots, t_{nm}) \xrightarrow{p} (1 - \theta_1, 1 - \theta_2, \cdots, 1 - \theta_m).$$

In other words, jointly, t_{ni} converges in probability to 1 for all true null hypotheses and to 0 for all false null hypotheses in probability. Now, we have a subsequence $\{k_1, k_2, \cdots, k_n, \cdots\}$ of $\mathbb{N}$ such that,

$$(t_{k_n 1}, t_{k_n 2}, \cdots, t_{k_n m}) \xrightarrow{a.s.} (1 - \theta_1, 1 - \theta_2, \cdots, 1 - \theta_m), \text{ as } n \to \infty.$$

Therefore, there exists $n' \in \mathbb{N}$ such that, for all $n \in \{k_{n'+1}, k_{n'+2}, \cdots\}$

$$\mathbb{P}(\{\cap_{i:\theta_i=0}\{1 - t_{ni} < \beta\}\} \cap \{\cap_{i:\theta_i=1}\{t_{ni} < \alpha\}\}) = 1,$$

i.e.,

$$\{n \in \mathbb{N} | t_{ni} < \alpha \text{ for all } i \text{ such that } \theta_i = 1 \text{ and } t_{ni} > 1 - \beta \text{ for all } i \text{ such that } \theta_i = 0\} \supseteq \{k_{n'+1}, k_{n'+2}, \cdots\}$$

almost surely. Hence, the set is not empty, and the definition of T^* is valid. Now, for any $n < T^*$, either $t_{ni} \geq \alpha$ for at least one alternative or $t_{ni} \leq 1 - \beta$ for at least one null. Now, since $T_{max} \leq T^*$, we must have, $\{T_{max} > n\} \subseteq \{T^* > n\}$. Therefore,

$$\begin{aligned}
\mathbb{P}(T_{max} > n) &\leq \mathbb{P}(T^* > n) \\
&= \mathbb{P}\left(\{\cup_{i:\theta_i=0}\{t_{ni} \leq 1-\beta\}\} \cup \{\cup_{i:\theta_i=1}\{t_{ni} \geq \alpha\}\}\right) \\
&= 1 - \mathbb{P}\left(\{\cap_{i:\theta_i=0}\{t_{ni} > 1-\beta\}\} \cap \{\cap_{i:\theta_i=1}\{t_{ni} < \alpha\}\}\right) \\
&= 1 - \prod_{i:\theta_i=0} \mathbb{P}(t_{ni} > 1-\beta) \times \prod_{i:\theta_i=1} \mathbb{P}(t_{ni} < \alpha).
\end{aligned}$$

Due to Assumption 8, The right side of the equation converges to 0 as n grows to ∞. Therefore, $\lim_{n\to\infty} \mathbb{P}(T_{max} > n) = \mathbb{P}(T_{max} > \infty) = 0$. Conversely, $\mathbb{P}(T_{max} < \infty) = 1$. A similar argument can be made for the OC test as well, i.e., $\mathbb{P}(T^C_{max} < \infty) = 1$. This completes the proof of the part (i) of Theorem 14.

Proof of part (ii) Note that, FDR is defined as the expected value of the false discovery proportions (FDP). Therefore,

$$\text{FDR} = \mathbb{E}\left(\frac{\sum_{i\in[m]}(1-\theta_i)\delta_i}{(\sum_{i\in[m]}\delta_i) \vee 1}\right).$$

We condition on the summary statistics for each data streams up to the corresponding stopping times and rewrite the FDR.

$$\text{FDR} = \mathbb{E}_{S_{T_1 1}, S_{T_2 2}, \cdots, S_{T_m m}}\left(\mathbb{E}\left(\frac{\sum_{i\in[m]}(1-\theta_i)\delta_i}{(\sum_{i\in[m]}\delta_i) \vee 1}\middle| S_{T_1 1}, S_{T_2 2}, \cdots, S_{T_m m}\right)\right).$$

We note that the decision rules are functions of these summary statistics and remain unchanged. Expected conditional indicators become conditional probabilities.

$$\text{FDR} = \mathbb{E}_{S_{T_1 1}, S_{T_2 2}, \cdots, S_{T_m m}}\left(\frac{\sum_{i\in[m]}\delta_i \mathbb{P}(\theta_i = 0 | S_{T_1 1}, S_{T_2 2}, \cdots, S_{T_m m})}{(\sum_{i\in[m]}\delta_i) \vee 1}\right). \tag{4.5}$$

We propose the following lemma:

Lemma 14. *If Assumption 7 is satisfied, then,*

$$\mathbb{P}\left(\theta_i = 0 | S_{T_1 1}, S_{T_2 2}, \cdots, S_{T_m m}\right) = t_{T_i i},$$

where, $t_{T_i i}$ denotes the oracle test statisitc calculated on the basis of the i-th data stream only up to the stopping time T_i, i.e.,

$$t_{T_i i} = \mathbb{P}(\theta_i = 0 | S_{T_i i}).$$

Lemma 14 says that, the conditional probability in the above equation can be replaced with the oracle test statisitc for the i-th data stream based on observations up to time point T_i. The decision δ_i for the i-th hypotheses is given by $\delta_i = \mathbb{1}\left(t_{T_i i} \leq t_{t_i}^{(r_{T_i})}\right)$. We note that, the values of T_i lies between k and T_{max} (both inclusive.) Therefore, we can write,

$$\begin{aligned}
\text{FDR} &= E_{S_{T_1 1}, S_{T_2 2}, \cdots, S_{T_m m}}\left(\frac{\sum_{i\in[m]} \delta_i t_{T_i, i}}{\left(\sum_{i\in[m]} \delta_i\right) \vee 1}\right) \\
&= E_{\{S_{T_i i}, i\in[m]\}} \left(\frac{\begin{array}{l} \sum_{n=k}^{T_{max}-1} \sum_{i\in[m]} \mathbb{1}(T_i = n)\, \mathbb{1}(t_{ni} \leq t_n^{(r_n)}) t_{ni} + \\ \sum_{i\in[m]} \mathbb{1}(T_i = T_{max})\, \mathbb{1}(t_{T_{max} i} \leq t_{T_{max}}^{(s)}) t_{T_{max} i} \end{array}}{\begin{array}{l} \left(\sum_{n=k}^{T_{max}-1} \sum_{i\in[m]} \mathbb{1}(T_i = n)\, \mathbb{1}(t_{ni} \leq t_n^{(r_n)}) + \right. \\ \left. \sum_{i\in[m]} \mathbb{1}(T_i = T_{max})\, \mathbb{1}(t_{T_{max} i} \leq t_{T_{max}}^{(s)}) \right) \vee 1 \end{array}} \right)
\end{aligned}$$

In the numerator of the last expression, the first sum takes into account all possible values of T_is. In the second sum, we consider all the data streams. The interchange of the sums is possible because both of these additions are over a finite number of elements. Now, for any $n \in \{k, k+1, \ldots, T_{max} - 1\}$, the term $\sum_{i\in[m]} \mathbb{1}(T_i = n)\, \mathbb{1}(t_{ni} \leq t_n^{(r_n)}) t_{ni}$ represents the sum of the oracle test statistics evaluated at time point n for the data streams i with stopping time $T_i = n$. By the definition of r_n, it is clear that the said term is smaller than $r_n \alpha$. Similarly, for data streams i with $T_i = T_{max}$, by

definition of r_{Tmax}, and since $s \leq r_{T_{max}}$,

$$\frac{\sum_{i\in[m]} \mathbb{1}(T_i = T_{max})\, \mathbb{1}(t_{T_{max}i} \leq t^{(s)}_{T_{max}}) t_{T_{max}i}}{s} \leq \frac{\sum_{i\in[m]} \mathbb{1}(T_i = T_{max})\, \mathbb{1}(t_{T_{max}i} \leq t^{(r_{T_{max}})}_{T_{max}}) t_{T_{max}i}}{r_{T_{max}}} \leq \alpha.$$

Consequently $\sum_{i\in[m]} \mathbb{1}(T_i = T_{max})\, \mathbb{1}(t_{T_{max},i} \leq t^{(s)}_{T_{max}}) t_{T_{max}i} \leq s\alpha$. Further, The denominator denotes the total number of rejections for the whole procedure, which is r_n up to $n = T_{max} - 1$ and is s for $n = T_{max}$. Therefore, we have,

$$\text{FDR} \leq E_{S_{T_1,1}, S_{T_2,2}, \cdots, S_{T_m,m}} \left(\frac{\alpha(\sum_{n=k}^{T_{max}-1} r_n + s)}{(\sum_{n=k}^{T_{max}-1} r_n + s) \vee 1} \right) \leq \alpha. \tag{4.6}$$

This proves that the OS test controls FDR at the predetermined level α. The proof for FNR control, given below, is similar.

$$\begin{aligned}
\text{FNR} &= E\left(\frac{\sum_{i\in[m]}(1-\delta_i)\theta_i}{(\sum_{i\in[m]}(1-\delta_i)) \vee 1} \right) \\
&= E_{S_{T_1 1}, S_{T_2 2}, \cdots, S_{T_m m}} \left(\frac{\sum_{i\in[m]}(1-\delta_i)\mathbb{P}(\theta_i = 1 | S_{T_1 1}, S_{T_2 2}, \cdots, S_{T_m m})}{(\sum_{i\in[m]}(1-\delta_i)) \vee 1} \right) \\
&= E_{S_{T_1 1}, S_{T_2 2}, \cdots, S_{T_m m}} \left(\frac{\sum_{i\in[m]}(1-\delta_i)(1-t_{T_i i})}{(\sum_{i\in[m]}(1-\delta_i)) \vee 1} \right) \\
&= E_{S_{T_i,i}, i\in[m]} \left(\begin{array}{l} \sum_{n=k}^{T_{max}-1} \sum_{i\in[m]} \mathbb{1}(T_i = n)\, \mathbb{1}(t_{ni} \geq t_n^{(|I_n|-a_n+1)})(1-t_{ni}) + \\ \qquad \sum_{\{i|T_i = T_{max}\}} \mathbb{1}(t_{T_i,i} > t^{(s)}_{T_i})(1-t_{T_i,i}) \\ \hline \left(\sum_{n=k}^{T_{max}-1} \sum_{i\in[m]} \mathbb{1}(T_i = n)\, \mathbb{1}(t_{ni} \geq t_n^{(|I_n|-a_n+1)})(1-t_{ni}) + \right. \\ \qquad \left. \sum_{\{i|T_i = T_{max}\}} \mathbb{1}(t_{T_i i} > t^{(s)}_{T_i})(1-t_{T_i i}) \right) \vee 1 \end{array} \right) \\
&\leq E_{S_{T_1,1}, S_{T_2,2}, \cdots, S_{T_m,m}} \left(\frac{\beta(\sum_{n=k}^{T_{max}-1} a_n + |I_{T_{max}}| - s)}{(\sum_{n=k}^{T_{max}-1} a_n + |I_{T_{max}}| - s) \vee 1} \right) \leq \beta.
\end{aligned}$$

Now, for the OC test, expressing the false discovery rate as the expected value of the conditional expectation of the false discovery proportion over the set $\{S_{T_1^C 1}, S_{T_2^C 2}, \ldots, S_{T_m^C m}\}$,

and then following (4.5) we get:

$$\text{FDR} = \mathbb{E}_{S_{T_1^C 1}, S_{T_2^C 2}, \cdots, S_{T_m^C m}} \left(\frac{\sum_{i\in[m]} \delta_i \mathbb{P}\left(\theta_i = 0 | S_{T_1^C 1}, S_{T_2^C 2}, \cdots, S_{T_m^C m}\right)}{\left(\sum_{i\in[m]} \delta_i\right) \vee 1} \right). \tag{4.7}$$

Since, Lemma 14 only requires Assumption 7 to be true, the result is valid for the set of stopping times $\{T_1^C, T_2^C, \ldots, T_m^C\}$ as well. That is, we have

$$\mathbb{P}\left(\theta_i = 0 | S_{T_1^C 1}, S_{T_2^C 2}, \cdots, S_{T_m^C m}\right) = t_{T_i^C i}, \text{ for all } i \in [m].$$

Replacing this in (4.7), we get

$$\begin{aligned}
\text{FDR} &= E_{S_{T_1^C 1}, S_{T_2^C 2}, \cdots, S_{T_m^C m}} \left(\frac{\sum_{i\in[m]} \delta_i t_{T_i^C, i}}{\left(\sum_{i\in[m]} \delta_i\right) \vee 1} \right) \\
&= E_{\{S_{T_i^C i}, i\in[m]\}} \left(\frac{\sum_{n=k}^{T_{max}^C - 1} \sum_{i\in[m]} \mathbb{1}(T_i^C = n)\, \mathbb{1}(t_{ni} \le t_n^{(r_n)}) t_{ni} + \sum_{i\in[m]} \mathbb{1}(T_i^C = T_{max}^C)\, \mathbb{1}(t_{T_{max}^C i} \le t_{T_{max}^C}^{(s)}) t_{T_{max}^C i}}{\left(\sum_{n=k}^{T_{max}^C - 1} \sum_{i\in[m]} \mathbb{1}(T_i^C = n)\, \mathbb{1}(t_{ni} \le t_n^{(r_n)}) + \sum_{i\in[m]} \mathbb{1}(T_i^C = T_{max}^C)\, \mathbb{1}(t_{T_{max}^C i} \le t_{T_{max}^C}^{(s)}) \right) \vee 1} \right)
\end{aligned}$$

Now, since the definition of r_n in the OC test coincides with that in the OS test for every $n \in \mathbb{N}$, an argument analogous to that preceding (4.6) yields

$$\text{FDR} \le \mathbb{E}_{S_{T_1^C,1}, S_{T_2^C,2}, \ldots, S_{T_m^C,m}} \left(\frac{\alpha \left(\sum_{n=k}^{T_{\max}^C - 1} r_n + s \right)}{\left(\sum_{n=k}^{T_{\max}^C - 1} r_n + s \right) \vee 1} \right) \le \alpha.$$

The proof of FNR control goes through similarly.

□

4.6 Proof of Lemmas

Proof of Lemma 13:

Proof. We note that, for $i \in I_n$,

$$\begin{aligned} t_{ni,A_n} &= \mathbb{P}(\theta_i = 0 | A_n) \\ &= \frac{f(S_{nl}\,\forall l \in I_n, S_{n_j,j}\,\forall j \notin I_n, \theta_i = 0)}{f(S_{nl}\,\forall l \in I_n, S_{n_j,j}\,\forall j \notin I_n)} \end{aligned}$$

Here, $f(S_{nl}\,\forall l \in I_n, S_{n_j,j}\,\forall j \notin I_n, \theta_i = 0)$ denotes the joint distribution of the summary statistics for the active set of hypotheses I_n (calculated up to time point n) and the summary statistics for the inactive set of hypotheses calculated up to the corresponding stopping times, and the event of H_{0i} being true. For each $i \in I_n$, we use the law of total probability for the partition obtained due to different values $\{k_j \in \{0,1\} | \, j \in [m], j \neq i\} \in \{0,1\}^{m-1}$ assumed by $\{\theta_j | \, j \in [m], j \neq i\}$:

$$\begin{aligned} &f(S_{nl}\,\forall l \in I_n, S_{n_j,j}\,\forall j \notin I_n, \theta_i = 0) \\ &= \sum_{\{k_j \in \{0,1\} | j \in [m], j \neq i\}} f(S_{nl}\,\forall l \in I_n, S_{n_q q}\,\forall q \notin I_n, \theta_i = 0, \theta_j = k_j, j \in [m]) \\ &= \sum_{\{k_j \in \{0,1\} | j \in [m], j \neq i\}} f(S_{nl}\,\forall l \in I_n, S_{n_q q}\,\forall q \notin I_n | \theta_i = 0, \theta_j = k_j\,\forall j \in [m], j \neq i) \times \\ &\mathbb{P}\left(\theta_i = 0, \theta_j = k_j\,\forall j \in [m], j \neq i\right). \end{aligned}$$

Now, according to Assumption 7, given $\{\theta_i | i \in [m]\}$, $\{S_{n_i i}, i \in [m]\}$ are independently distributed and, the conditional distribution of S_{ni} is dependent only on θ_i. Further, θ_i, $i \in [m]$ are independent Bernoulli random variables, i.e.,

$$\begin{aligned} &f(S_{nl}\,\forall l \in I_n, S_{n_j,j}\,\forall j \notin I_n, \theta_i = 0) \\ &= f(S_{ni} | \theta_i = 0)\mathbb{P}(\theta_i = 0) \sum_{\{k_j \in \{0,1\} | j \in [m], j \neq i\}} \Bigg(\prod_{l \in I_n, l \neq i} f(S_{nl} | \theta_l = k_l)\mathbb{P}(\theta_l = k_l) \times \\ &\prod_{q \notin I_n} f(S_{n_q q} | \theta_q = k_q)\mathbb{P}(\theta_q = k_q) \Bigg). \end{aligned}$$

Without loss of generality, suppose, $i \neq 2, 3$ and $2 \in I_n$ and $3 \notin I_n$. Then, first expanding the sum for $j = 2$, and then for $j = 3$, we get,

$$
\begin{aligned}
&f(S_{nl}\,\forall l \in I_n, S_{n_j,j}\,\forall j \notin I_n, \theta_i = 0)\\
&= f(S_{ni}|\theta_i = 0)\mathbb{P}(\theta_i = 0) \sum_{\{k_j\in\{0,1\}|j\in[m],j\neq i,2\}} \Bigg(\prod_{l\in I_n, l\neq i,2} f(S_{nl}|\theta_l = k_l)\mathbb{P}(\theta_l = k_l)\times\\
&\prod_{q\notin I_n} f(S_{n_q q}|\theta_q = k_q)\mathbb{P}(\theta_q = k_q)\Bigg)\Bigg(f(S_{n2}|\theta_2 = 0)\mathbb{P}(\theta_2 = 0)+\\
&f(S_{n2}|\theta_2 = 1)\mathbb{P}(\theta_2 = 1)\Bigg).\\
&= f(S_{ni}|\theta_i = 0)\mathbb{P}(\theta_i = 0) \sum_{\{k_j\in\{0,1\}|j\in[m],j\neq i,2\}} \Bigg(\prod_{l\in I_n, l\neq i,2} f(S_{nl}|\theta_l = k_l)\mathbb{P}(\theta_l = k_l)\times\\
&\prod_{q\notin I_n} f(S_{n_q q}|\theta_q = k_q)\mathbb{P}(\theta_q = k_q)\Bigg) f(S_{n2}).\\
&= f(S_{ni}|\theta_i = 0)\mathbb{P}(\theta_i = 0) f(S_{n2}) \sum_{\{k_j\in\{0,1\}|j\in[m],j\neq i,2,3\}} \Bigg(\prod_{l\in I_n, l\neq i,2} f(S_{nl}|\theta_l = k_l)\times\\
&\mathbb{P}(\theta_l = k_l) \times \prod_{q\notin I_n, q\neq 3} f(S_{n_q q}|\theta_q = k_q)\mathbb{P}(\theta_q = k_q)\Bigg)\Bigg(f(S_{n_3 3}|\theta_3 = 0)\mathbb{P}(\theta_3 = 0)+\\
&f(S_{n_3 3}|\theta_3 = 1)\mathbb{P}(\theta_3 = 1)\Bigg).\\
&= f(S_{ni}|\theta_i = 0)\mathbb{P}(\theta_i = 0) f(S_{n2}) \sum_{\{k_j\in\{0,1\}|j\in[m],j\neq i,2,3\}} \Bigg(\prod_{l\in I_n, l\neq i,2} f(S_{nl}|\theta_l = k_l)\times\\
&\mathbb{P}(\theta_l = k_l) \times \prod_{q\notin I_n, q\neq 3} f(S_{n_q q}|\theta_q = k_q)\mathbb{P}(\theta_q = k_q)\Bigg) f(S_{n_3 3})\\
&= f(S_{ni}|\theta_i = 0)\mathbb{P}(\theta_i = 0) f(S_{n2}) f(S_{n_3 3}) \sum_{\{k_j\in\{0,1\}|j\in[m],j\neq i,2,3\}} \Bigg(\prod_{l\in I_n, l\neq i,2} f(S_{nl}|\theta_l = k_l)\times\\
&\mathbb{P}(\theta_l = k_l) \times \prod_{q\notin I_n, q\neq 3} f(S_{n_q q}|\theta_q = k_q)\mathbb{P}(\theta_q = k_q)\Bigg)
\end{aligned}
$$

Proceeding in this way, expanding for all the elements of I_n and $[m] - I_n$, it can be shown that, the sum in the last line equals to

$$\prod_{l\in I_n, l\neq,2} f(S_{nl}) \times \prod_{q\notin I_n, q\neq 3} f(S_{n_q q}),$$

i.e.,

$$f(S_{nl}\,\forall l \in I_n, S_{n_j,j}\,\forall j \notin I_n, \theta_i = 0) = \pi_0 f_{0n}(S_{ni}) \prod_{l\in I_n, l\neq i} f(S_{nl}) \times \prod_{q\notin I_n} f(S_{n_q q}).$$

Approaching similarly, we can show that

$$f(S_{ni}\,\forall i \in I_n, S_{n_j,j}\,\forall j \notin I_n) = \prod_{l\in I_n} f(S_{nl}) \times \prod_{q\notin I_n} f(S_{n_q q}).$$

Therefore, the oracle test statisitc takes the form

$$\begin{aligned} t_{ni,A_n} &= \frac{\pi_0 f_{0n}(S_{ni}) \prod_{l\in I_n, l\neq i} f(S_{nl}) \times \prod_{q\notin I_n} f(S_{n_q q})}{\prod_{l\in I_n} f(S_{nl}) \times \prod_{q\notin I_n} f(S_{n_q q})} \\ &= \frac{\pi_0 f_{0n}(S_{ni})}{f(S_{ni})} = \frac{\pi_0 f_{0n}(S_{ni})}{f_n(S_{ni})} = t_{ni}. \end{aligned}$$

Hence, Lemma 13 is proved. □

Proof of Lemma 14:

Proof. We define B to be the event that the value of $t_{T_i i}$ is different that that of $\mathbb{P}\left(\theta_i = 0 | S_{T_1 1, T_2 2, \cdots, T_m m}\right)$, i.e.,

$$B := \{\omega \in \Omega | t_{T_i i} \neq \mathbb{P}\left(\theta_i = 0 | S_{T_1 1}, S_{T_2 2}, \cdots, S_{T_m m}\right)\}.$$

For any m natural numbers, $n_1, n_2, \cdots, n_m \in \mathbb{N}$, define $B_{n_1,n_2,\cdots,n_m}$ to be the subset of B when $T_i = n_i$ for all $i \in [m]$, i.e.,

$$\begin{aligned} B_{n_1,n_2,\cdots,n_m} &:= B \cap \{T_i = n_i, \text{ for all } i \in [m]\} \\ &= \{\omega \in \Omega | t_{n_i i} \neq \mathbb{P}\left(\theta_i = 0 | S_{n_1 1}, S_{n_2 2}, \cdots, S_{n_m m}\right)\}. \end{aligned}$$

Further, from part (i) of Theorem 14, we can conclude that, $\mathbb{P}(T_i < \infty) = 1$ for all $i \in [m]$. Now, for any $\omega \in B$, assume that $T_i(\omega) = n_i$, for all $i \in [m]$. i.e, $\omega \in B_{n_1,n_2,\cdots,n_m}$. Therefore, from the definition of B and $B_{n_1,n_2,\cdots,n_m}$ we can say that $B \subseteq \cup_{n_1,n_2,\cdots,n_m} B_{n_1,n_2,\cdots,n_m}$. Therefore,

$$\begin{aligned} \mathbb{P}(B) &\leq \mathbb{P}(\cup_{n_1,n_2,\cdots,n_m} B_{n_1,n_2,\cdots,n_m}) \\ &\leq \sum_{n_1,n_2,\cdots,n_m} \mathbb{P}(B_{n_1,n_2,\cdots,n_m}) \end{aligned}$$

Now, note that, due to Assumption 7,

$$\begin{aligned}&\mathbb{P}(\theta_i = 0 | S_{n_1 1}, S_{n_2 2}, \ldots, S_{n_m m}) \\ &= \frac{f(S_{n_1 1}, S_{n_2 2}, \ldots, S_{n_m m}, \theta_i = 0)}{f(S_{n_1 1}, S_{n_2 2}, \ldots, S_{n_m m})}\end{aligned}$$

Now, following the proof of Lemma 13, we can show that the numerator of the expression is

$$f(S_{n_i i}, \theta_i = 0) \times \prod_{j \in [m]-i} f(S_{n_j j}) \tag{4.8}$$

And the numerator is

$$\prod_{j \in [m]} f(S_{n_j j}). \tag{4.9}$$

Finally, equations (4.8) and (4.9) together imply that

$$\begin{aligned}&\mathbb{P}(\theta_i = 0 | S_{n_1 1}, S_{n_2 2}, \ldots, S_{n_m m}) \\ &= \frac{f(S_{n_i i}, \theta_i = 0) \times \prod_{j \in [m]-i} f(S_{n_j j})}{\prod_{j \in [m]} f(S_{n_j j})} \\ &= \frac{\pi_0 f_{0 n_i}(S_{n_i i})}{f_{n_i}(S_{n_i i})} = t_{n_i i}.\end{aligned}$$

i.e., for any set $\{n_1, n_2, \ldots, n_m\} \in \mathbb{N}$, $\mathbb{P}(B_{n_1, n_2, \cdots, n_m}) = 0$. Therefore,

$$\mathbb{P}(B) \leq \sum_{n_1, n_2, \cdots, n_m} 0 = 0.$$

Since T_i are stopping times for each stream i and are almost surely finite, the sum over all combinations is taken over a countable set of indices where each term is zero, confirming $\mathbb{P}(B) = 0$. Since probability must lie in the interval $[0, 1]$, we must have,

$$\mathbb{P}(t_{T_i i} \neq \mathbb{P}\left(\theta_i = 0 | S_{T_1 1}, S_{T_2 2}, \cdots, S_{T_m m}\right)) = \mathbb{P}(B) = 0.$$

This proves the lemma. □

Chapter 5

Numerical Study

This chapter presents numerical evidence supporting the theoretical results established in Chapters 3 and 4. In Section 5.1, we evaluate the numerical performance of the OI and DI tests introduced in Chapter 3. Section 5.3 reports simulation results for the OS and DS tests, described in Chapter 4.

Throughout the simulation studies and real data applications, we fix the type I error level $\alpha = 0.05$ and the type II error level $\beta = 0.1$. To implement the OI and DI tests, we use a pilot sample of size $k = 30$. For the OS and DS procedures, a slightly larger pilot sample of size $k = 50$ is used. Performance of each test is reported using (i) the estimated expected stopping time (average sample number), denoted as ASN, (ii) the estimated FDR, denoted as fdr, and (iii) the estimated FNR, denoted as fnr.

5.1 Numerical Study on the OI and DI Tests

In this section, we compare the performance of the OI and DI tests with several existing sequential and fixed-sample-size procedures using both simulated and real datasets. For the implementation of the DI test, the proportion of true null hypotheses, denoted by π_0, is estimated based on the data type: for discrete data settings (e.g., Example 4), we use the method proposed by Storey, Taylor, and Siegmund (2004) with the default tuning parameter $\lambda = 0.5$, while for continuous data setups, π_0 is estimated using the method of Cai and Jin (2010) with the tuning parameter $\lambda = 0.1$.

Upon implementation of the OI and DI procedures, we observe that the estimated marginal false discovery rate (mFDR) and marginal false nondiscovery rate (mFNR) closely align with the conventional estimates of FDR and FNR. Due to this

similarity, the mFDR and mFNR values are omitted from the reported simulation results.

5.1.1 Comparison Between Sequential Tests

Here, we compare the performance of our tests with two versions of Gap rules discussed in He and Bartroff (2021) for controlling FDR and FNR. The stopping time of Gap rule is $T_G^* = \inf\left\{n \in \mathbb{N} : \Lambda_n^{(K)} - \Lambda_n^{(K+1)} \geq c\right\}$, where K is the known number of alternatives, $\Lambda_n^{(j)}$ is the j-th ordered log-likelihood ratio based on n observations, and c is some constant. If $c = \log K(m-K)/\alpha \wedge \beta$, then the Gap rule is referred as "Gap-ao" test. A simulation-based choice of c, say $\hat{c}$, is also considered in He and Bartroff (2021) without any theoretical guarantee of FDR and FNR control. Suppose for a given c, fdr(c) and fnr(c) denote the estimated FDR and FNR of the Gap rule respectively based on 50 Monte Carlo simulations. We examine different choices of c starting from 0.1 with an increment of 0.1 until we get some $c = \hat{c}$ that satisfies $\text{fdr}(\hat{c}) \leq \alpha$ and $\text{fnr}(\hat{c}) \leq \beta$. The Gap rule with such a $c = \hat{c}$ is referred as the "Gap-sb" test. In Figures 3.2, 3.3 and Table 5.1, ASN, fdr and fnr are computed based on 200 Monte Carlo runs.

We generate data streams based on Assumption 5 where for a prefixed π_0, we first generate $\theta_1, \ldots, \theta_m \overset{\text{i.i.d.}}{\sim} Bernoulli(1-\pi_0)$ and then for a given θ_j, we generate $X_{1j}, X_{2j}, X_{3j} \ldots \overset{\text{i.i.d.}}{\sim} (1-\theta_j)\, f_0 + \theta_j\, f_1$. To implement the Gap-sb and Gap-ao tests, the number of alternatives $K = \sum_{j=1}^m \theta_j$ need to be known and only simple null against simple alternative can be tested. This leads us to consider the following data setups for comparing our tests with the Gap-sb and Gap-ao tests.

Ex1: $f_0 \equiv N(0,1)$ and $f_1 \equiv N(0.25,1)$.

Ex2: $f_0 \equiv Exponential(\lambda = 1)$ and $f_1 \equiv Exponential(\lambda = 1.2)$ where λ is the scale parameter of the Exponential distribution.

Ex3: $f_0 \equiv N(0,1)$ and $f_1 \equiv N(0,(1.2)^2)$.

Ex4: $f_0 \equiv Bin(7,0.1)$ and $f_1 \equiv Bin(7,0.3)$.

The OI test is implemented using oracle statistics $\{t_{ni}\}_{i=1}^m$ in (3.11) where π_0 and $f_n \equiv \pi_0 f_0 + (1-\pi_0) f_1$ are known and the summary statistic $\boldsymbol{S}_n$ is a z-score vector for setups Ex1, Ex2 and Ex3. For the discrete setup Ex4, oracle statistics (3.2) is used with $S_{ni} = \sum_{j=1}^n X_{ji}$, $f_{0n} \equiv Bin(7n,0.1)$ and $f_{1n} \equiv Bin(7n,0.3)$.

For the normal location problem (setup Ex1), Figure 3.2 illustrates that ASN for the OI and DI tests remain nearly constant for moderately large m (500) to extremely large m (100000) whereas ASN corresponding to the Gap-sb and Gap-ao tests tend to diverge with larger m. This is perfectly in line with the claims of Theorems 10, 11 and 12. Figure 3.3 shows the conservative behavior of the Gap-sb and Gap-ao tests in terms of their attained FDR and FNR for large m whereas the proposed tests nearly attains the desired levels.

Table 5.1 reports ASN, fdr and fnr for the OI, DI and Gap-sb tests and only ASN for the Gap-ao test since it is so conservative for large $m(\geq 100)$ that the corresponding fdr and fnr values are nearly zero, and hence, is omitted for brevity. Indeed, the OI test provides exact control of FDR and FNR at 5% and 10% levels respectively. The attained FDR and FNR levels are nearly 5% and 10% in most cases and slightly higher in a few cases for moderate m, but decrease with larger m. The Gap-sb test is clearly conservative in terms of producing either small fdr or small fnr for large m in all scenarios. We also record savings in ASN defined as

$$s_1 = \frac{\text{ASN of Gap-sb} - \text{ASN of DI}}{\text{ASN of Gap-sb}} \text{ and } s_2 = \frac{\text{ASN of Gap-ao} - \text{ASN of OI}}{\text{ASN of Gap-ao}}$$

achieved by the DI and OI tests respectively. It is noteworthy that due to the adaptive nature of the stopping boundaries, the OI and DI tests require much smaller sample size (ASN) than the Gap rules leading to huge savings (ranging from 14% to 89%) in ASN. Although the asymptotic error control for the DI test is not theoretically proven for the discrete setup, Ex4 illustrates perfect control of the FDR and FNR in the discrete setup as well.

5.1.2 Comparing the DI and BH Tests

Now, we evaluate the performance of the DI test for some composite hypotheses and compare it with a fixed-sample-size test, namely the popular Benjamini-Hochberg (BH) test Benjamini and Hochberg (1995) which controls FDR only. For some pre-fixed π_0, we generate $\theta_1, \ldots, \theta_m \overset{\text{i.i.d.}}{\sim} Bernoulli(1-\pi_0)$ and then for a given θ_j, we generate $X_{1j}, X_{2j}, X_{3j} \ldots \overset{\text{i.i.d.}}{\sim} (1-\theta_j) f_0 + \theta_j f_1$. We consider the following setups that include both normal and non-normal settings as well as a two-sample problem.

TABLE 5.1: Comparing OI test, DI test and the Gap rules

			OI test			DI test			Gap-sb			Gap-ao
	m	π_0	ASN	fdr (%)	fnr (%)	ASN	fdr(%)	fnr (%)	ASN (s_1%)	fdr (%)	fnr (%)	ASN (s_2%)
Ex1	100	0.8	109.5	4.07	10.16	107.3	5.74	10.76	227.9 (53)	4.80	1.18	563.7 (81)
		0.2	132.9	4.67	10.2	146.1	5.95	8.55	169.7 (14)	2.35	9.60	554.8 (74)
	500	0.8	109.7	5.12	10.19	122.7	5.37	8.88	254.0 (52)	4.43	1.11	823.1 (85)
		0.2	130.6	4.89	9.83	141.5	5.42	7.96	183.8 (23)	2.15	8.64	841.4 (83)
	1000	0.8	110.2	4.82	10.03	118.6	4.65	9.09	252.4 (53)	4.57	1.13	957.4 (88)
		0.2	130.4	4.76	10.00	138.4	5.14	8.19	176.9 (22)	2.30	9.24	972.2 (86)
Ex2	100	0.8	185.04	5.03	10.14	187.55	5.89	10.77	412.2 (55)	5.06	1.23	1067.5 (82)
		0.2	246.9	4.39	10.06	258.24	5.39	10.94	309.6 (17)	2.49	10.83	1070.6 (76)
	500	0.8	192.28	4.92	9.89	210.04	5.11	9.11	454.9 (54)	4.84	1.19	1584.2 (87)
		0.2	246.22	4.97	10.22	246.97	5.14	11.19	317.7 (22)	2.56	10.34	1594.6 (85)
	1000	0.8	193.13	4.86	9.93	199.3	5.03	9.73	452.9 (56)	4.89	1.23	1808.8 (89)
		0.2	248.5	4.98	10.19	242.6	4.88	11.9	333.4 (27)	2.36	9.44	1821.9 (87)
Ex3	100	0.8	90.3	4.25	10.21	101.6	5.63	9.62	192.9 (47)	4.28	1.03	296.2 (66)
		0.2	124.6	4.61	11.0	122.8	4.86	14.25	155.1 (21)	2.39	9.66	297.8 (59)
	500	0.8	91.5	5.22	9.88	93.1	5.61	9.87	229.9 (60)	4.97	1.26	504.5 (82)
		0.2	128.1	4.93	9.87	120.5	4.63	13.99	165.4 (27)	2.37	9.85	507.9 (76)
	1000	0.8	91.3	5.21	9.89	93.1	5.14	9.94	231.16 (60)	4.89	1.20	590.7 (84)
		0.2	128.0	4.95	9.90	118.9	4.63	13.96	168.4 (30)	2.31	9.54	612.2 (81)
Ex4	100	0.8	41.7	4.30	9.63	36.7	6.05	11.02	75.88 (52)	4.09	0.96	170.7 (79)
		0.2	53.3	4.79	8.92	53.4	4.75	9.50	65.13 (18)	1.77	7.07	170.1 (69)
	500	0.8	41.5	4.91	9.82	41.4	5.55	9.00	104.9 (61)	3.15	0.8	275.2 (85)
		0.2	53.1	5.01	8.68	53.1	5.22	8.31	103.68 (49)	0.84	3.44	278.7 (63)
	1000	0.8	41.6	4.96	9.32	43.0	5.82	8.36	134.21 (68)	1.64	0.4	320.7 (87)
		0.2	53.2	5.04	9.07	53.3	5.21	8.42	133.75 (60)	0.4	1.63	324.8 (84)

Ex5: $f_0 \equiv N(\mu_0, \sigma^2)$ and $f_1 \equiv N(\mu_1, \sigma^2)$ with $\mu_0 = 0, \mu_1 = 0.25$ and $\sigma = 1$. Assuming σ unknown, we conduct one-sample student's t-tests for $H_{0j} : \mu_j = \mu_0$ vs $H_{1j} : \mu_j = \mu_1$ for $j = 1, \ldots, m$, and convert the corresponding statistics to z-scores to obtain $\{\hat{t}_{nj}\}_{j=1}^m$.

Ex6: $f_0 \equiv Cauchy(\mu_0, 1)$ and $f_1 \equiv Cauchy(\mu_1, 1)$ with $\mu_0 = 0, \mu_1 = 0.25$. For Cauchy location parameter μ_j, we conduct simple vs simple tests for $H_{0j} : \mu_j = \mu_0$ vs $H_{1j} : \mu_j = \mu_1$ for $j = 1, \ldots, m$, based on the log-likelihood ratio statistics and convert the corresponding statistics to z-scores to obtain $\{\hat{t}_{nj}\}_{j=1}^m$.

Ex7: We consider a sequential two-sample (or case-control) testing setup (see Crockett and Gebski (1984)), where for each sampling stage $n \in \mathbb{N}$ and coordinate $j = 1, \ldots, m$, we observe $X_{nj} = (U_{nj}, V_{nj})$ where $U_{1j}, U_{2j}, \ldots \overset{\text{i.i.d.}}{\sim} N(\mu_{0j}, \sigma^2)$ and $V_{1j}, V_{2j}, \ldots \overset{\text{i.i.d.}}{\sim} N(\mu_{1j}, \sigma^2)$. The goal is to conduct two-sample t-tests for $H_{0j} : \mu_{0j} = \mu_{1j}$ vs $H_{1j} : \mu_{0j} \neq \mu_{1j}$ with σ unknown. Here, we generate "control" observations as $U_{1j}, U_{2j}, \ldots \overset{iid}{\sim} N(0, 1)$. For each j, the "case" observations are either from $N(0, 1)$ (null) or from $N(0.25, 1)$ (non-null) or from $N(-0.5, 1)$ (non-null). For prefixed π_0, we generate $\theta_1, \ldots, \theta_m \overset{\text{i.i.d.}}{\sim} Bernoulli(1 - \pi_0)$ and $\eta_1, \ldots, \eta_m \overset{\text{i.i.d.}}{\sim} Bernoulli(0.75)$ and

then generate "case" observations as

$$V_{1j}, V_{2j}, \ldots \overset{\text{i.i.d.}}{\sim} (1-\theta_j)N(0,1) + \theta_j(\eta_j N(0.25,1) + (1-\eta_j)N(-0.5,1)).$$

Finally, $\{\hat{t}_{nj}\}_{j=1}^{m}$ are obtained by converting the Welch's two-sample t-statistics into z-scores. Since at each stage n, we observe a pair (U_{nj}, V_{nj}), the sample size reported (in Table 5.2) is twice the number of such pairs sampled till stopping T.

TABLE 5.2: Comparison between the DI test and the BH test

	m	π_0	DI test			BH test		
			ASN	fdr (%)	fnr (%)	$\hat{n}$ (s%)	fdr (%)	fnr (%)
Ex5	2500	0.8	121.72	5.03	8.8	151 (20)	4.28	8.81
		0.2	133.12	5.12	9.19	261 (49)	0.99	9.37
	7500	0.8	115.28	4.81	9.48	143 (19)	4.07	9.51
		0.2	130.12	5.02	9.84	258 (50)	1.00	9.88
	10000	0.8	113.7	4.94	9.65	142 (20)	3.98	9.66
		0.2	129.32	5.00	10.04	257 (50)	1.04	9.96
Ex6	2500	0.8	209.78	5.78	9.99	234 (10)	3.91	9.86
		0.2	261.22	4.96	10.33	443 (41)	0.91	10.26
	7500	0.8	206.64	5.66	10.12	231 (11)	3.65	10.29
		0.2	260.08	4.91	10.81	435 (40)	0.90	10.97
	10000	0.8	207.56	5.73	9.92	234 (11)	3.82	9.91
		0.2	275.06	5.14	8.38	460 (40)	1.00	8.35
Ex7	2500	0.8	267.76	4.97	12.7	332.0 (19)	4.46	12.2
		0.2	476.92	5.23	9.82	962.0 (50)	0.98	9.8
	7500	0.8	328.92	4.51	10.70	414.0 (21)	4.07	10.51
		0.2	469.52	5.16	10.49	954.0 (51)	0.98	10.51
	10000	0.8	357.48	5.05	9.70	428.0 (17)	3.93	9.79
		0.2	467.36	5.06	10.6	944.0 (51)	0.98	10.51

For each of the simulation setups described above, we first conduct the DI test at level $\alpha = 0.05$, and $\beta = 0.1$ and compute the ASN, fdr and fnr based on 200 Monte Carlo runs. Next, we implement the BH test based on p-values with the same α for different sample sizes $n\ (\geq k)$ and note $\text{fnr}(n)$ for each n. Finally, we select the value of $n = \hat{n}$ for which $\text{fnr}(\hat{n})$ matches closely with the fnr obtained from the DI test. Table 5.2 reports values of ASN, fdr and fnr for the DI test and the BH test with $n = \hat{n}$ as well as the savings in the expected sample size $s = (\hat{n} - \text{ASN of DI})/\hat{n}$. In all the setups, the DI test not only controls FDR and FNR at the intended levels but also leads to significant savings (ranging from 10% to 51%) in ASN. Note that the BH test is conservative for $\pi_0 = 0.2$ (producing small fdr) and thus the DI test has

more savings in ASN for these cases.

5.2 Real Data Analysis

In this section, we apply the DI test on two real datasets and compare its performance with the BH test. First, we consider the prostate dataset (Singh et al. (2002)) that contains gene expression data on 6033 genes from 102 prostates; 52 are from tumor cells, and the rest are from non-tumor cells. This is a case-control study to identify genes whose expressions are associated with prostate tumors.

This data setup is the same as the setup in **Ex7** of Section 5.1.1. Suppose U_{ij} represents the j-th gene expression associated with the i-th prostate cell with tumor, and V_{lj} represents the j-th gene expression associated with the l-th prostate cell without tumor, where $j = 1, 2, ..., 6033$, $i = 1, .., 52$ and $l = 1, ..., 50$. We assume $U_{1j}, U_{2j}, \ldots \overset{iid}{\sim} N(\mu_{1j}, \sigma^2)$ and $V_{1j}, V_{2j}, \overset{iid}{\sim} N(\mu_{2j}, \sigma^2)$. Our goal is to test$H_{0j} : \mu_{1j} = \mu_{2j}$ vs $H_{1j} : \mu_{1j} \neq \mu_{2j}$ for all $j = 1, \ldots, 6033$. After standardizing and quantile normalizing the full dataset, we start the DI test with a pilot sample of size $k = 50$ with equal $(k/2 = 25)$ numbers of tumor and non-tumor cells. In consecutive steps, we include either a new tumor or a non-tumor cell with probability 0.5. At any interim stage n of the DI test, we have n_1 tumor and n_2 non-tumor cells with $n_1 + n_2 = n$ and compute the two-sample t-statistic $S_{nj} = (\bar{U}_j - \bar{V}_j)/\hat{\sigma}_j\sqrt{n_1^{-1} + n_2^{-1}}$, where $\bar{U}_j(\bar{V}_j)$ is the sample mean for the j-th gene based on n_1 tumor cells (n_2 non-tumor cells), and $\hat{\sigma}_j$ is the pooled sample variance. We compute the z-scores as $Z_{nj} = \Phi^{-1}(F_{n-2}(S_{nj}))$ and p-values as $p_{nj} = 2\min\{F_{n-2}(S_{nj}), 1 - F_{n-2}(S_{nj})\}$ for $j = 1, \ldots 6033$, where F_n denotes the CDF of a t-distribution with n degrees of freedom. The Lfdr statistics $\{\hat{t}_{nj}\}$ are computed from the z-scores as in Section 3.3 of Chapter 4. The sampling in the DI test terminates at $n = T$ where $r_T + a_T \geq 6033$.

Based on the full data (i.e, $n = 102$ cells), we apply the BH test based on p-values and the AdaptZ test based on z-scores proposed in Sun and Cai (2007a) with same level $\alpha = 0.05$. Figure 5.1 shows the histograms of the z-scores and the p-values for the full dataset. The BH test and the AdaptZ test discovers **21** and **29** significant genes respectively. When applied with $\alpha = 0.05$ and $\beta = 0.07$, the DI test stops at $T = 87$ (with a **14.71%** saving in the sample size) and discovers **23** significant

genes that is higher than the number of discoveries made by the BH method. For $\alpha = 0.05$, $\beta = 0.1$, the DI test stops at $T = 69$ (with a **32.4%** saving in the sample size) and discovers **12** significant genes which is less than BH and AdaptZ methods. Therefore, the DI test can be viewed to have a tuning parameter β that not only controls the FNR but the final sample size and the number of discoveries.

The second dataset, given in Golub et al. (1999), consists of gene expression levels on $m = 7129$ genes from the bone marrow tissues of $n_{max} = 72$ acute leukemia patients. Among these patients, 47 were suffering from acute lymphoblastic leukemia (ALL) and the remaining 25 were acute myeloid leukemia (AML) patients. Based on such an unbalanced dataset, our goal is to discover genes that discriminate between these two types of leukemia. Similar to the prostate cancer data, we assume that gene expressions follow normal distributions with equal variance, and apply two-sample t-tests based on $\{S_{nj}\}_{j=1}^{m}$ as before. Figure 5.2 illustrates the histograms of the z-scores and the p-values for the full Golub dataset. The histogram of p-values has a peak near 0 (representing non-null cases) and seems to resemble uniform distribution (representing null cases). The Histogram of z-scores resembles a normal distribution with slightly heavy frequencies at the tail, and provides some justification for applying t-tests.

Based on the full dataset, the BH method with $\alpha = 0.05$ discovers 1280 significant genes whereas AdaptZ method with $\alpha = 0.05$ identifies 1300 genes that differentiate between these two types of leukemia. The DI test with $\alpha = 0.05$ and $\beta = 0.1$ does not terminate sampling for this dataset, that is, $T > n_{max}$. Therefore, we apply the truncated DI test discussed in Chapter 6. Considering FDR control to be our primary objective, we stop sampling at $N = \min\{T, n_{max}\}$ and reject all H_{0i} for which $\hat{t}_{Ni} \leq \hat{t}_N^{(rN)}$. This truncated DI test also identifies 1300 positive genes. Therefore, from the above data applications, we conclude that, the DI test or its truncated version can either provide savings in sample size or performs as good as the optimal fixed-sample-size AdaptZ test controlling FDR only.

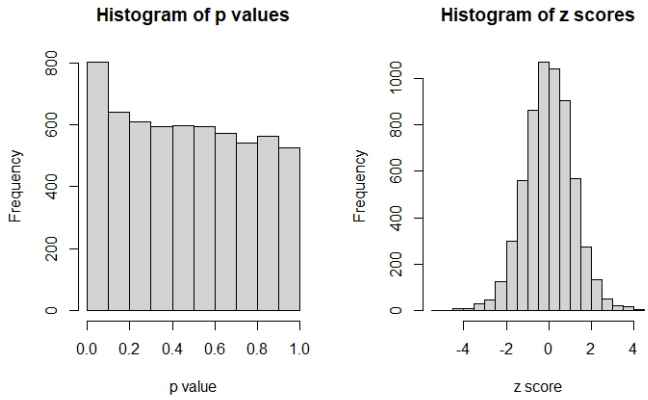

FIGURE 5.1: Histogram of p-values and z-scores for the full Prostate dataset

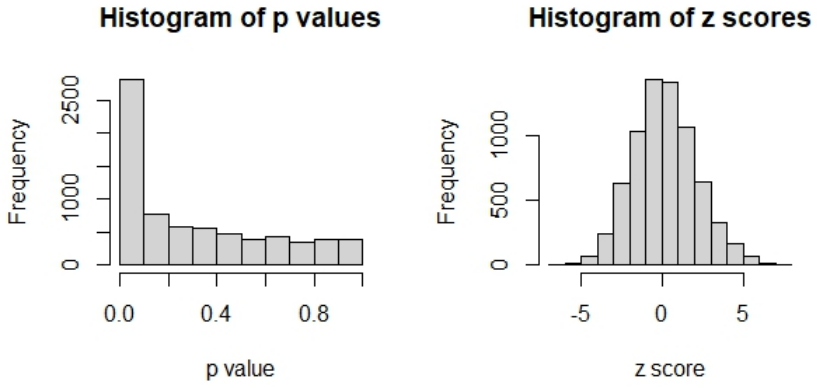

FIGURE 5.2: Histograms of p-value and z-scores of Golub data

5.3 Simulation Studies for the DS and DC Tests

In this section, first we apply the OS and DS tests to simulated datasets with 3 simple-versus-simple testing scenarios. In the simulation studies, we compare the performance of the data-driven procedure with the oracle procedure. Further, we evaluate both procedures against a sequential version of the Benjamini–Hochberg (BH) test, proposed by Bartroff and Song (2020), referred to as the sequential BH (seq-BH) test. Then, five different composite tests have been considered to illustrate the performance of the DC test. The results are compared with that of the OC test. In the first three scenarios, we also compare the results with the Sequential BH test.

A brief overview of the BH method was presented in Section 1.1.2 of Chapter 1. Below, we describe the Sequential BH test.

Sequential BH test: Suppose we have a data-generating procedure as described in Chapter 4. Then the Sequential BH procedure of Bartroff and Song (2020) may be described as the following.

Suppose, as before, m denotes the total number of hypotheses under consideration. Further, Let $I_n \subseteq [m]$ denote the active set at stage n, and let $|I_n|$ denote its cardinality.

To control the false discovery rate (FDR) and false non-discovery rate (FNR), critical values are constructed using the original number of hypotheses m. For $s \in [m]$, define

$$\alpha_s = \alpha \frac{m - s\beta}{m(m-\beta)} \quad \text{and} \quad \beta_s = \beta \frac{m - s\alpha}{m(m-\alpha)},$$

where α and β denote the target FDR and FNR levels, respectively. The corresponding lower and upper boundaries are

$$a_s = \log\left(\frac{s\beta}{(1-\alpha_s)m}\right) + \rho \quad \text{and} \quad b_s = \log\left(\frac{(1-\beta_s)m}{s\alpha}\right) - \rho, \tag{5.1}$$

where $\rho > 0$ is a continuity-correction parameter. These boundaries satisfy

$$a_1 \le a_2 \le \cdots \le a_m \le b_m \le \cdots \le b_2 \le b_1.$$

For each active hypothesis $i \in I_n$, let S_{ni} denote a sequential test statistic computed using all observations collected from stream i up to stage n. Following Bartroff and Song (2020), the statistics are transformed using a piecewise-linear standardization map ϕ, yielding standardized statistics

$$\widetilde{S}_{ni} = \phi(S_{ni}), \qquad i \in I_n.$$

Under the piecewise-linear standardization map ϕ, the boundaries a_s and b_s are transformed into the common scale

$$-(m-s+1) \quad \text{and} \quad m-s+1,$$

which yields the stopping boundaries used below. The standardized statistics are then ordered as

$$\widetilde{S}_n^{(1)} \leq \widetilde{S}_n^{(2)} \leq \cdots \leq \widetilde{S}_n^{(|I_n|)}.$$

Let a_n^* and r_n^* denote the cumulative numbers of accepted and rejected hypotheses, respectively, prior to stage n. Sampling continues until at least one ordered statistic exits the continuation region

$$\Big[-(m - a_n^* - \ell + 1),\ a_n^* + \ell \Big], \qquad \ell = 1, \ldots, |I_n|.$$

Equivalently, a stage terminates whenever

$$\widetilde{S}_n^{(\ell)} \leq -(m - a_n^* - \ell + 1) \quad \text{or} \quad \widetilde{S}_n^{(\ell)} \geq a_n^* + \ell$$

for some $\ell \in [|I_n|]$. If a lower boundary is crossed, the number of acceptances is determined by

$$a_n' = \max \left\{ k \in [|I_n|] : \widetilde{S}_n^{(k)} \leq -(m - a_n^* - k + 1) \right\}.$$

The hypotheses corresponding to the a_n' smallest ordered statistics are accepted and removed from the active set. Similarly, if an upper boundary is crossed, the number of rejections is determined by

$$r_n' = \max \left\{ k \in [|I_n|] : \widetilde{S}_n^{(|I_n|-k+1)} \geq m - r_n^* - k + 1 \right\}.$$

The hypotheses corresponding to the r_n' largest ordered statistics are rejected and removed from the active set. After each stage, accepted and rejected hypotheses are removed from I_n. One additional observation is then collected from each remaining active stream, and the procedure is repeated. The algorithm terminates when all hypotheses have been classified or when a pre-specified maximum sample size is reached.

For the OS, OC, DS, DC and the Seq-BH tests, for each hypothesis i, the stopping time

$$T_i = \inf \{ n : H_{0i} \text{ is accepted or rejected} \}$$

records the stage at which a final decision is made. To illustrate the performance of the procedures, we consider the mean stopping time ($\bar{T}$) for these stagewise procedures as

$$\bar{T} = \frac{1}{m}\sum_{i=1}^{m} T_i.$$

5.3.1 Case 1: Simple Hypotheses

We generate the data streams according to Assumption 7. First, we simulate the vector of indicators $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$, where $\theta_i \overset{\text{iid}}{\sim} \text{Bernoulli}(1-\pi_0)$ for all $i \in [m]$. Given $\theta_1, \ldots, \theta_m$, the i-th data stream is generated as

$$X_{1i}, X_{2i}, \ldots \overset{\text{iid}}{\sim} (1-\theta_i)f_0 + \theta_i f_1,$$

where f_0 and f_1 are fully known in our simulation settings. We consider the following examples:

- **Ex-1:** $f_0 \equiv N(0,1), \quad f_1 \equiv N(0.25, 1)$.
- **Ex-2:** $f_0 \equiv \text{Exponential}(\text{scale} = 1), \quad f_1 \equiv \text{Exponential}(\text{scale} = 1.2)$.
- **Ex-3:** $f_0 \equiv N(0,1), \quad f_1 \equiv N(0, (1.2)^2)$.

In the OS and DS tests, we use z-scores as the summary statistics. For the OS test, we compute $\{t_{ni}, i \in [m]\}$ assuming π_0 and $f_n = \pi_0 f_{0n} + (1-\pi_0)f_{1n}$ are known. In the DS test, π_0 is estimated once at the start of the procedure, using the method of Cai and Jin (2010) with $k = 50$ pilot samples, and this estimate $\hat{\pi}_{0k}$ is used throughout until sampling stops. We assume f_0 and f_1 are known for both the OS and DS tests in our simulations. For the Sequential BH method, the summary statistic for each data stream is taken as the likelihood ratio

$$\Lambda(X_{1i}, X_{2i}, \ldots, X_{ni}) = \prod_{j=1}^{n} \frac{f_1(X_{ji})}{f_0(X_{ji})} \tag{5.2}$$

We consider three values of m: 1000, 5000, and 10000, under two sparsity settings: $\pi_0 = 0.8$ (sparse alternative) and $\pi_0 = 0.2$ (dense setting). The OS, DS, and Sequential BH (Seq-BH) tests are each applied in 200 Monte Carlo runs for every

configuration. The performance of the procedures are reported through the measures ASN, fdr and fnr. For the stagewise procedures considered here, the ASN is the average mean stopping time ($\bar{T}$) average over 200 Monte Carlo runs. Here fdr and fnr, respectively denote the estimated FDR and FNR. Results are reported in Tables 5.3–5.5 for the setups in Ex-1, Ex-2, and Ex-3, respectively.

The savings in the ASN per data stream are reported in Tables 5.3–5.5 as s_3 (OS test) and s_4 (DS test), defined by

$$s_3 = \frac{\text{ASN of Seq-BH} - \text{ASN of OS}}{\text{ASN of Seq-BH}}, \quad s_4 = \frac{\text{ASN of Seq-BH} - \text{ASN of DS}}{\text{ASN of Seq-BH}}.$$

The simulation results indicate that both the OS and DS procedures perform satisfactorily across a wide range of configurations. For most settings, the attained FDR and FNR values remain close to their nominal levels $\alpha = 0.05$ and $\beta = 0.1$, respectively. Furthermore, the attained error rates remain relatively stable as the number of hypotheses increases from $m = 1000$ to $m = 10000$, suggesting that the procedures scale well with the dimensionality of the problem.

As expected, the OS procedure consistently achieves the smallest average sample number (ASN) among the competing methods. Since the OS procedure has access to the oracle quantities, it serves as the benchmark for evaluating the performance of the DS procedure. The results show that the DS procedure closely tracks the performance of the OS procedure in most scenarios. Although estimation of π_0 introduces some additional variability, the resulting increase in ASN is generally moderate, and the attained FDR and FNR remain comparable to those of the OS procedure.

The performance of the DS procedure is particularly encouraging in sparse alternatives, where the attained error rates are very close to their target values. In denser settings, the attained FDR tends to be somewhat conservative, while the attained FNR occasionally exceeds the nominal level. However, these departures are generally modest and become less pronounced as the number of hypotheses increases.

The simulation studies also demonstrate substantial savings in ASN relative to the Sequential Benjamini–Hochberg (Seq-BH) procedure. Across all examples considered, the OS and DS procedures require considerably fewer observations while

maintaining comparable error-rate control. The percentage reduction in ASN generally increases with the number of hypotheses, indicating that the proposed procedures become increasingly advantageous in large-scale testing problems.

Overall, the simulation results suggest that the DS procedure provides a satisfactory approximation to its oracle counterpart. Despite relying on an estimated value of π_0, the DS procedure retains the principal advantages of the OS procedure, namely effective control of FDR and FNR together with substantial reductions in sample size.

TABLE 5.3: Comparing OS, DS and Seq-BH tests for the setup in EX-1

		OS test			DS test			Seq-BH test		
m	π_0	ASN (s_3)	$fdr(\%)$	$fnr(\%)$	ASN (s_4)	$fdr(\%)$	$fnr(\%)$	ASN	$fdr(\%)$	$fnr(\%)$
1000	0.8	59.02 (61%)	4.71	10.27	58.12 (61%)	4.92	11.15	150.19	1.42	0.39
	0.2	68.24 (59%)	5.07	10.26	72.23 (56%)	3.40	13.84	165.86	0.16	2.16
5000	0.8	58.76 (62%)	4.91	9.89	57.78 (63%)	4.56	10.89	154.63	0.16	2.16
	0.2	67.54 (60%)	5.02	9.31	70.08 (59%)	4.06	12.21	169.70	0.16	2.16
10000	0.8	58.71 (62%)	4.92	10.05	58.39 (62%)	4.77	10.24	154.50	1.13	0.31
	0.2	67.54 (60%)	4.98	10.06	69.67 (59%)	4.16	12.01	170.37	0.16	1.58

TABLE 5.4: Comparing OS, DS and Seq-BH tests for the setup in EX-2

		OS test			DS test			Seq-BH test		
m	π_0	ASN (s_3)	$fdr(\%)$	$fnr(\%)$	ASN (s_4)	$fdr(\%)$	$fnr(\%)$	ASN	$fdr(\%)$	$fnr(\%)$
1000	0.8	79.53 (69%)	4.73	9.98	88.84 (69%)	2.78	9.56	284.89	1.16	0.36
	0.2	108.48 (78%)	5.00	9.82	157.83 (67%)	1.46	16.72	304.79	0.16	2.69
5000	0.8	79.02 (87%)	4.86	10.00	99.68 (83%)	3.16	7.35	292.20	0.96	0.36
	0.2	107.40 (81%)	4.98	9.88	151.44 (74%)	1.69	13.67	315.44	0.14	2.05
10000	0.8	78.91 (88%)	4.93	10.02	101.21 (84%)	3.61	7.00	294.58	0.88	0.35
	0.2	107.12 (82%)	9.89	4.99	146.64 (76%)	1.88	12.56	317.33	0.12	1.94

TABLE 5.5: Comparing OS, DS and Seq-BH tests for the setup in EX-3

		OS test			DS test			Seq-BH test		
m	π_0	ASN (s_3)	$fdr(\%)$	$fnr(\%)$	ASN (s_4)	$fdr(\%)$	$fnr(\%)$	ASN	$fdr(\%)$	$fnr(\%)$
1000	0.8	56.27 (76%)	4.26	9.98	54.68 (77%)	3.98	11.72	143.09	1.10	0.56
	0.2	67.13 (74%)	4.95	9.84	71.23 (72%)	2.83	16.81	150.89	0.13	2.75
5000	0.8	55.88 (80%)	4.89	9.99	57.34 (80%)	4.30	11.28	147.71	0.85	0.42
	0.2	67.92 (78%)	4.97	9.96	69.23 (77%)	3.29	14.77	155.04	0.12	2.47
10000	0.8	55.86 (81%)	5.02	10.00	54.87 (82%)	4.37	10.93	148.06	0.75	0.42
	0.2	66.65 (80%)	4.96	9.94	69.31 (79%)	3.46	14.17	155.99	0.11	2.36

5.3.2 Case 2: Composite Hypotheses

We evaluate the DC test across five distinct composite scenarios (Ex-4 to Ex-8), covering Z-tests, scale tests, and t-tests. Unlike simple setups where the null and alternative hypotheses are completely known, these examples require us to estimate the marginal distributions of the oracle statistics from the streaming data. In each of the

following examples, we start with a pilot sample of size $k = 50$. We consider three different values for m, namely 1000, 5000 and 10000. Two sparsity settings have been considered: $\pi_0 = 0.8$ (sparse alternative) and $\pi_0 = 0.3$ (dense alternative). The cut-off number m' was assumed to be $0.4 \times m$. Each simulation is repeated for 50 Monte Carlo runs. We report the ASN, i.e., the average mean stopping time $\bar{T}$ averaged over 50 Monte Carlo runs, fdr, i.e., the estimated FDR and fnr, the estimated FNR.

For Examples 4–8, we compare the performance of the proposed OC and DC tests. For Examples 4–6, we additionally compare their performance with that of the Sequential Benjamini–Hochberg (Seq-BH) procedure of Bartroff and Song (2020), described in Section 5.3. For the Seq-BH procedure, we employ the likelihood ratio statistics defined in (5.2) together with the cutoff values in (5.1). Although these cutoff values were originally derived for simple hypotheses, their use in the present composite settings is justified by Theorem 5.1 of Bartroff and Song (2020).

The Seq-BH procedure requires the availability of local sequential tests satisfying prescribed type I and type II error bounds. Such tests are readily available for the settings considered in Examples 4–6 through Theorem 5.1 of Bartroff and Song (2020). However, Bartroff and Song (2020) does not provide analogous local sequential tests for the one-sample and two-sample t-test settings considered in Examples 7 and 8. Constructing such procedures would require additional methodological development and is therefore beyond the scope of the present work.

For each configuration, we generate the data streams according to Assumption 7. First, we simulate the vector of indicators $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)$, where $\theta_i \overset{\text{iid}}{\sim} \text{Bernoulli}(1 - \pi_0)$ for all $i \in [m]$. The rest of the data generation procedure is detailed below.

- **Ex-4:** *Z test* $X_{ni} \overset{i.i.d.}{\sim} N(\mu, 1)$ for all $n \in \mathbb{N}$. For all $i \in [m]$, we want to test $H_{0i} : \mu = 0$ vs. $H_{1i} : \mu > \delta$ for some $\delta > 0$. Therefore, for each data stream with $\theta_i = 0$, $\mu = 0$, i.e., $X_{ni} \overset{i.i.d.}{\sim} N(0, 1)$. Here, $\delta > 0$ the indifference region between the null and alternative distributions. For data streams with $\theta_i = 1$, we assume, $X_{ni} \overset{i.i.d.}{\sim} N(0.25, 1), n \in \mathbb{N}$, i.e., we take $0 < \delta < 0.25$. The difference with Ex-1 is that, here, we actually do not know the true value of μ under H_{1i}, and we need to estimate the marginal distribution f_n to get the estimated test statistics t_{ni}^C.

- **Ex-5:** (*Normal One-Sample Scale Test*) $X_{ni} \overset{i.i.d.}{\sim} N(0, \sigma^2)$ for all $n \in \mathbb{N}$. For all $i \in [m]$, we want to test $H_{0i} : \sigma = 1$ vs. $H_{1i} : \sigma > 1 + \delta$ for some $\delta > 0$. For $\{i \in [m] | \theta_i = 1\}$, i.e., for the alternative distributions we consider $\sigma = 1.2$.

- **Ex-6:** (*Exponential Scale Test*) $X_{ni} \overset{i.i.d.}{\sim} exp(\text{ scale } = \lambda)$ for all $n \in \mathbb{N}$. For all $i \in [m]$, we want to test $H_{0i} : \lambda = 1$ vs. $H_{1i} : \lambda > 1 + \delta$ for some $\delta > 0$. For alternative distribution, we consider $\lambda = 1.2$.

- **Ex-7:** (*One-Sample t-test*) $X_{ni} \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ for all $n \in \mathbb{N}$. For all $i \in [m]$, we want to test $H_{0i} : \mu = 0$ vs. $H_{1i} : \mu > \delta$ for some $\delta > 0$. Here, $\sigma = 1$ and for alternative distribution, we consider $\mu = 0.25$.

- **Ex-8:**(*Two-Sample t-test*) Here, we adapt a sequential two-sample (case-control) testing setup motivated by Crockett and Gebski (1984). The setup is described in Example 7 of Section 5.1.2. In this particular example, we assume the null distribution to be $N(0, \sigma^2)$. The alternative distribution is a mixture distribution. $p_1 N(\mu_1, \sigma^2) + p_2 N(\mu_2, \sigma^2)$ with $\pi_0 = 1 - (p_1 + p_2)$. We consider $\mu_1 = 0.5$ and $\mu_2 = -0.5$ and $\sigma = 1$ for our simulation examples.

In tables 5.6 -5.8, where we compare the OC and DC test results with the Seq-BH test, we define the sample size savings due to the OC and DC tests respectively as

$$s_5 = \frac{\text{ASN of Seq-BH} - \text{ASN of OC}}{\text{ASN of Seq-BH}}, \quad s_6 = \frac{\text{ASN of Seq-BH} - \text{ASN of DC}}{\text{ASN of Seq-BH}}.$$

and report them alongside the ASN for OC and DC tests.

Across Examples 4–8, the OC and DC tests exhibit stable performance over a wide range of problem configurations. For both sparse and dense alternatives, the attained FDR and FNR values are generally close to the nominal levels $\alpha = 0.05$ and $\beta = 0.1$, respectively. Moreover, the attained error rates remain relatively stable as the number of hypotheses increases from $m = 1000$ to $m = 10000$, indicating that the proposed procedures scale well with the dimensionality of the problem.

As expected, the OC test consistently achieves the smallest ASN among the competing procedures. Since the OC test has access to the oracle quantities, it represents the ideal benchmark against which the data-driven procedures are evaluated. The

DC test generally requires a somewhat larger ASN than the OC test due to the additional uncertainty arising from estimation of the oracle statistics. Nevertheless, the increase in ASN is moderate in most cases, while the attained FDR and FNR remain comparable to those of the OC test. This suggests that the proposed plug-in estimation strategy provides a reasonable approximation to the oracle procedure.

In Examples 4–6, where comparison with the Sequential BH (Seq-BH) procedure is possible, both the OC and DC tests yield substantial reductions in ASN relative to Seq-BH. The savings become particularly noticeable as the number of hypotheses increases. At the same time, the Seq-BH procedure tends to attain highly conservative FDR and FNR values, indicating that the additional sampling effort does not translate into improved operating characteristics. These findings suggest that the proposed procedures are considerably more sample-efficient in large-scale testing problems while maintaining comparable error-rate control.

A slight deterioration in the attained FNR is observed in some configurations, particularly in Example 5. This phenomenon is most likely attributable to the difficulty of estimating the oracle statistics in this particular scale-testing problems. Nevertheless, even in these cases, the departures from the nominal level remain relatively small and do not materially affect the overall conclusions.

Taken together, the simulation results provide strong empirical evidence that the OC procedure achieves the desired operating characteristics with substantial savings in sample size, while the DC procedure successfully approximates the oracle performance despite relying entirely on estimated quantities.

TABLE 5.6: Comparing OC, DC and Seq-BH tests for the setup in EX-4

		OC test			DC test			Seq-BH test		
m	π_0	ASN (s_5)	$fdr(\%)$	$fnr(\%)$	ASN (s_6)	$fdr(\%)$	$fnr(\%)$	ASN	$fdr(\%)$	$fnr(\%)$
1000	0.8	107.44 (28%)	5.15	9.72	133.62 (11%)	5.4	6.96	150.19	1.42	0.39
	0.3	95.58 (42%)	4.76	9.95	129.16 (22%)	7.79	9.35	166.08	0.29	1.56
5000	0.8	107.66 (30%)	4.96	9.82	121.78 (21%)	4.79	8.56	154.63	1.1	0.33
	0.3	95.24 (44%)	4.85	9.79	140.50 (18%)	5.75	8.52	170.86	0.25	1.29
10000	0.8	106.89 (31%)	5.05	9.71	117.52 (24%)	4.97	8.68	154.50	1.13	0.31
	0.3	95.15 (44%)	4.88	10.09	142.10 (17%)	5.31	8.9	171.17	0.24	1.26

TABLE 5.7: Comparing OC, DC and Seq-BH tests for the setup in EX-5

		OC test			DC test			Seq-BH test		
m	π_0	ASN (s_5)	$fdr(\%)$	$fnr(\%)$	ASN (s_6)	$fdr(\%)$	$fnr(\%)$	ASN	$fdr(\%)$	$fnr(\%)$
1000	0.8	89.88 (37%)	5.24	9.64	100.20 (30%)	5.50	8.48	143.09	1.10	0.56
	0.3	131.78 (14%)	4.58	10.56	127.54 (17%)	5.06	10.14	153.13	0.24	2.29
5000	0.8	91.78 (38%)	4.87	10.05	97.98 (34%)	4.96	9.21	147.71	0.85	0.42
	0.3	132.18 (16%)	5.10	9.66	124.18 (21%)	4.82	12.17	156.94	0.19	1.90
10000	0.8	90.96 (39%)	5.03	9.95	95.32 (36%)	5.03	9.62	148.06	0.75	0.42
	0.3	132.06 (16%)	5.06	9.82	121.70 (23%)	4.73	13.05	157.58	0.16	1.79

TABLE 5.8: Comparing OC, DC and Seq-BH tests for the setup in EX-6

		OC test			DC test			Seq-BH test		
m	π_0	ASN (s_5)	$fdr(\%)$	$fnr(\%)$	ASN (s_6)	$fdr(\%)$	$fnr(\%)$	ASN	$fdr(\%)$	$fnr(\%)$
1000	0.8	194.06 (32%)	5.10	9.81	222.91 (22%)	5.73	7.98	284.89	1.16	0.36
	0.3	177.092 (43%)	4.70	9.76	272.80 (12%)	5.77	8.55	309.90	0.21	1.87
5000	0.8	191.28 (35%)	4.94	9.95	209.72 (28%)	5.07	8.81	292.20	0.96	0.36
	0.3	177.26 (44%)	4.95	9.98	256.58 (19%)	5.17	10.27	318.00	0.22	1.54
10000	0.8	190.65 (35%)	5.27	9.82	205.18 (30%)	5.07	8.93	294.58	0.88	0.35
	0.3	176.93 (45%)	4.95	10.06	254.81 (20%)	5.04	10.30	320.41	0.20	1.43

TABLE 5.9: Comparing OC and DC test results for the setup in EX-7

		OC test			DC test		
m	π_0	ASN	$fdr(\%)$	$fnr(\%)$	ASN	$fdr(\%)$	$fnr(\%)$
1000	0.8	109.4	5.35	9.69	137.58	5.43	6.86
	0.3	96.83	4.74	9.88	126.39	7.85	7.39
5000	0.8	110.9	4.86	10.09	125.86	4.88	8.25
	0.3	96.46	4.88	10.03	137.26	6.24	8.91
10000	0.8	111.38	4.97	10.03	121.10	4.91	8.59
	0.3	96.03	4.92	10.00	142.78	5.32	8.30

TABLE 5.10: Comparing OC and DC test results for the setup in EX-8

		OC test			DC test		
m	π_0	ASN	$fdr(\%)$	$fnr(\%)$	ASN	$fdr(\%)$	$fnr(\%)$
1000	0.8	66.64	5.17	9.92	69.11	5.68	9.85
	0.3	76.48	4.89	8.61	83.24	4.97	10.10
5000	0.8	65.54	4.99	9.99	64.20	4.87	10.33
	0.3	76.24	5.05	9.93	77.58	5.16	9.53
10000	0.8	65.24	5.03	9.92	64.73	4.86	10.16
	0.3	75.80	4.86	9.88	76.14	4.91	9.81

Chapter 6

Discussion and Future Directions

6.1 ABOS Tests for the Equicorrelated Setup

In Theorem 3 of Chapter 2, we showed that if the decision rules are of the form

$$\delta_i = \mathbb{1}\left(\frac{\hat{Z}_i^2}{\sigma_\rho^2} > c_m\right),$$

where $\hat{Z}_i = X_i - \bar{X}_m$, and c_m satisfies conditions (2.18) and (2.19), then the set of decision rules $\{\delta_i : i \in [m]\}$ is ABOS. These conditions provide a sufficient (though not necessary) framework for asymptotic Bayes optimality under sparsity. We used $\hat{Z}_i = X_i - \bar{X}_m$ as the test statistic for testing the i-th hypothesis. It is plausible that other test statistics, asymptotically equivalent to X_i, for example, statistics satisfying $S_i = X_i + o_p(1)$ may also satisfy the ABOS condition under the same cutoff rules. Establishing such results remains an open problem.

We assumed equicorrelation among all pairs X_i and X_j with $i \neq j$. To maintain the positive definiteness of the correlation matrix, ρ must lie in $(-1/m, 1)$. Since the lower bound converges to 0, as $m \to \infty$, we restrict attention to the practically relevant case of positive dependence and assume $\rho > 0$ throughout this chapter.

Multiple testing under dependence has been extensively studied in the last two decades; see Ghosh (2020) for a comprehensive review. Extending Bogdan et al. (2011) to dependent settings is challenging. The equicorrelated model is a natural starting point, preserving exchangeability similar to the independent case. However, in many applications the dependence structure depends on the ordering or spatial arrangement of the coordinates, making the equicorrelation assumption unrealistic. More realistic structures include block dependence (Xie et al. (2010)), where

the covariance matrix is given by:

$$\Sigma = \sigma_\epsilon^2 \begin{bmatrix} A & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A \end{bmatrix}, \quad A = \begin{bmatrix} 1 & \cdots & \rho \\ \vdots & \ddots & \vdots \\ \rho & \cdots & 1 \end{bmatrix}_{k \times k}, \quad \rho > 0.$$

Another example is an autoregressive structure with diminishing correlation:

$$\Sigma = \sigma_\epsilon^2 \left\{ \rho^{|i-j|} \right\}_{ij}, \quad \rho > 0.$$

More general dependence structures may also be considered.

The two-group prior used here is commonly referred to as the *spike-and-slab* prior. We assumed known parameters ρ, σ_ϵ^2, and σ_0^2. A fully Bayesian analysis incorporating priors such as the horseshoe prior or other global-local shrinkage priors (including members of the Subbotin family) may provide further insight into the ABOS properties of multiple testing procedures under dependence. For instance, Qin and Ghosh (2025) considered a multivariate normal model with general covariance under a global-local shrinkage prior, and established ABOS under sparsity of the form $p_m \propto m^{-\beta}$ with $\beta \in (0, 1)$.

6.2 Large-Scale Sequential Multiple Testing with Simultaneous Termination of Sampling

For streaming multivariate data with simultaneously terminating samples, we propose an oracle test together with its data-driven counterpart for testing a large number of hypotheses, while controlling both the false discovery rate (FDR) and the false nondiscovery rate (FNR). Controlling the FNR (or its marginal version, mFNR) at a pre-specified level β ensures that the power, measured as the expected proportion of true acceptances among all acceptances (i.e., the *true nondiscovery proportion*), is at least $1 - \beta$. Such a property is particularly desirable in large-scale testing problems, where simultaneously controlling false discoveries and false nondiscoveries is often of practical importance.

In many sequential trials including clinical trials, the maximum allowable sample size, say n_{max}, may be specified before the experiment begins due to budget constraints or some other practical reasons. In such cases, the DI test needs to stop sampling at $N := min\{T, n_{max}\}$. If $n_{max} < T$, we can view the DI test as a fixed-sample-size test with n_{max} many observations, and hence, simultaneous control of FDR and FNR at levels α and β respectively may not be guaranteed. Considering FDR control to be of our primary interest, we can consider a test, referred to as the *truncated* DI test, that rejects nulls corresponding to $\hat{t}_N^{(1)}, \ldots, \hat{t}_N^{(\hat{r}_N)}$ where $\hat{r}_N = \max\{1 \leq q \leq m : \frac{1}{q}\sum_{i=1}^{q} \hat{t}_N^{(i)} \leq \alpha\}$, if $\hat{t}_N^{(i)} \leq \alpha$, and $\hat{r}_N = 0$, otherwise. Following Theorem 11, it can be shown that the truncated DI test guarantees FDR $\leq \alpha + o(1)$ (and mFDR $\leq \alpha + o(1)$) if $n_{max} < T$, and the truncated DI test becomes equivalent to the DI test yielding both FDR and FNR control if $n_{max} \geq T$. Therefore, the truncated DI test inherits the FDR control properties of the corresponding fixed-sample-size procedure while retaining the possibility of simultaneous FDR and FNR control whenever $n_{max} \geq T$. Note that the FNR level β of this test can be viewed as a tuning parameter that regulates the power (in terms of 1 - FNR) and the sample size required for the trial.

One of the key features of the proposed OI and DI tests is that the stopping boundaries of these tests can quickly adapt with sample size as the trial progresses and continuously shrink the continue-sampling region. The adaptive nature of the boundaries ensures the desired asymptotic properties and, in particular, yields convergence of the stopping times (equivalently, the final sample sizes) to a finite natural number as $m \to \infty$. Such theoretical results are new in the literature on large-scale multiple testing. Although we do not have an explicit formula for the expected sample size, the empirical evidence indicates substantial reductions in the required sample size. In fixed-sample-size settings, Sun and Cai (2007a) proves the existence of an optimal oracle test that minimizes the mFNR among all tests that control mFDR at a prefixed level α. In the setting of simple-versus-simple hypothesis testing with sequentially observed data,, the celebrated sequential probability ratio test (SPRT) is proven to be optimal in the sense of minimizing the expected sample size among all tests controlling both probabilities of type I and type II errors. See Wald and Wolfowitz (1948) for a precise statement of the optimality property of SPRT. Theoretical

considerations and numerical evidence suggest that the proposed OI test may possess an optimality property analogous to that of the SPRT. Specifically, we conjecture that any procedure attaining smaller FDR and FNR levels than the OI test must incur a larger expected sample size. Establishing such a result remains an open problem and appears challenging because explicit expressions for the expected sample size are unavailable for sequential multiple testing procedures.

6.3 Large-Scale Sequential Multiple Testing with Asynchronous Termination of Sampling

For the sampling scheme with asynchronous termination, we proposed two oracle procedures, namely the Oracle Stagewise (OS) test and the Oracle Composite (OC) test, together with their corresponding data-driven versions (DS and DC tests, respectively). Under the assumptions of Chapter 4, both oracle procedures possess finite stopping times and guarantee simultaneous control of the false discovery rate (FDR) and the false nondiscovery rate (FNR) at predetermined levels when the number of hypotheses m is finite. An important direction for future research is to investigate the asymptotic behavior of the stopping times as $m \to \infty$. In particular, it would be of interest to study whether the maximum stopping time $T_{\max}$ converges to a finite limit under suitable choices of the cutoff parameter m'. If $m' = m^{\delta}$ for some $\delta \in (0, 1)$, one may expect the OC test to retain sufficient active hypotheses for reliable estimation while still allowing the procedure to terminate in finite time. Establishing such theoretical results remains an open problem.

A major challenge in extending the oracle procedures to practical applications is the dynamic estimation of the oracle statistics t_{ni} as data accrue over time. The Data-driven Stagewise (DS) test relies on a large pilot sample to obtain a stable estimate of the overall null proportion π_0. This estimate is then kept fixed throughout the remainder of the procedure. Such an approach works naturally when both the null and alternative distributions are simple and fully specified. However, in many practical applications, the null or alternative distributions are composite, making direct evaluation of the oracle statistics difficult. Moreover, as the stagewise procedure progresses, the active set of hypotheses shrinks rapidly. Consequently, the active

data streams available at later stages may no longer provide sufficient information for reliable estimation of the marginal distribution of the summary statistics.

This difficulty motivates the Data-driven Composite (DC) test. Unlike the DS test, the DC test incorporates a prefixed lower bound m' on the size of the active set. By preventing the active set from collapsing to a very small set of hypotheses, the procedure maintains enough observations to estimate the marginal distribution of the summary statistics and the proportion of true null hypotheses at each stage. Such a restriction appears essential whenever the oracle statistics must be estimated nonparametrically. The simulation studies suggest that maintaining a sufficiently large active set substantially improves the performance of the data-driven procedure, particularly in settings involving composite hypotheses.

Nevertheless, the estimation problem remains challenging. The proportion of true null hypotheses among the active hypotheses generally changes over time as the procedure selectively removes strong nulls and strong alternatives. Therefore, the proportion of null hypotheses within the active set at stage n, denoted by π_{0n}, need not coincide with the overall null proportion π_0. One possible approach is to combine information from the active set with the cumulative numbers of accepted and rejected hypotheses. Since the OC procedure controls FDR and FNR at levels α and β, respectively, one may approximate the number of false rejections by αR_n and the number of false acceptances by βA_n, where R_n and A_n denote the cumulative numbers of rejections and acceptances up to stage n. Such considerations lead to alternative estimators of π_0 that adapt to the evolving active set. Whether these estimators are consistent, and whether the resulting estimates of the oracle statistics retain the desirable properties of the oracle procedures, remains an open question.

Another practically important consideration is the presence of budget constraints. In many applications, a maximum allowable sample size $n_{\max}$ must be specified before the study begins. In such situations, the OC or DC procedures may not terminate naturally before reaching $n_{\max}$. This motivates a truncated version of the procedure. Sampling is continued according to the original rule until time $n_{\max}$. If the active set is still non-empty at that stage, all remaining active hypotheses are accepted. Such a truncated procedure preserves control of the FDR, although simultaneous control of the FNR may no longer be guaranteed. Consequently, the target

FNR level β may be viewed as a tuning parameter governing the trade-off between power and sampling cost. Smaller values of β typically require longer sampling and larger expected sample sizes, whereas larger values of β lead to earlier termination at the expense of reduced power.

Finally, although the oracle procedures developed in Chapter 4 provide exact finite-sample control of FDR and FNR, the theoretical understanding of their data-driven counterparts remains incomplete. Important questions include the consistency of the proposed estimators, the asymptotic behavior of the stopping times, and the preservation of error-rate control with estimated oracle statistics. Addressing these issues would significantly strengthen the theoretical foundations of large-scale sequential multiple testing with asynchronous termination and represents a promising direction for future research.

Bibliography

Bahadur, R Raj (1966). "A note on quantiles in large samples". In: *The Annals of Mathematical Statistics* 37.3, pp. 577–580.

Baillie, David H (1987). "Multivariate acceptance sampling-some applications to defence procurement". In: *Journal of the Royal Statistical Society. Series D (The Statistician)* 36.5, pp. 465–478.

Balamurugan, R, AM Natarajan, and K Premalatha (2015). "Stellar-mass black hole optimization for biclustering microarray gene expression data". In: *Applied Artificial Intelligence* 29.4, pp. 353–381.

Bartroff, Jay (2018). "Multiple hypothesis tests controlling generalized error rates for sequential data". In: *Statistica Sinica*, pp. 363–398.

Bartroff, Jay and Jinlin Song (2014). "Sequential tests of multiple hypotheses controlling type I and II familywise error rates". In: *Journal of Statistical Planning and Inference* 153, pp. 100–114. ISSN: 0378-3758.

— (2020). "Sequential tests of multiple hypotheses controlling false discovery and nondiscovery rates". In: *Sequential Analysis* 39.1, pp. 65–91.

Basu, Pallavi et al. (2018). "Weighted false discovery rate control in large-scale multiple testing". In: *Journal of the American Statistical Association* 113.523, pp. 1172–1183.

Begley, R et al. (Jan. 2026). "The JWST EXCELS survey: A spectroscopic investigation of the ionizing properties of star-forming galaxies at 1<z<8". In: *Monthly Notices of the Royal Astronomical Society* 545.1, staf1995. ISSN: 0035-8711.

Benjamini, Yoav and Yosef Hochberg (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing". In: *Journal of the Royal statistical society: series B (Methodological)* 57.1, pp. 289–300.

Benjamini, Yoav and Yosef Hochberg (2000). "On the adaptive control of the false discovery rate in multiple testing with independent statistics". In: *Journal of educational and Behavioral Statistics* 25.1, pp. 60–83.

Benjamini, Yoav, Abba M Krieger, and Daniel Yekutieli (2006). "Adaptive linear step-up procedures that control the false discovery rate". In: *Biometrika* 93.3, pp. 491–507.

Benjamini, Yoav and Daniel Yekutieli (2001). "The Control of the False Discovery Rate in Multiple Testing Under Dependency". In: *Ann. Stat.* 29.

— (2005). "False Discovery Rate–Adjusted Multiple Confidence Intervals for Selected Parameters". In: *Journal of the American Statistical Association* 100.469, pp. 71–81.

Berry, Donald A and Yosef Hochberg (1999). "Bayesian perspectives on multiple comparisons". In: *Journal of Statistical Planning and Inference* 82.1-2, pp. 215–227.

Blanchard, Gilles and Etienne Roquain (2009). "Adaptive false discovery rate control under independence and dependence." In: *Journal of Machine Learning Research* 10.12.

Bogdan, M., J. K. Ghosh, and S. T. Tokdar (2008). "A comparison of the Benjamini-Hochberg procedure with some Bayesian rules for multiple testing". In: *arXiv preprint arXiv:0805.2479*.

Bogdan, M. et al. (2007). "On the Empirical Bayes approach to the problem of multiple testing". In: *Quality and Reliability Engineering International* 23.6, pp. 727–739.

Bogdan, M. et al. (2011). "Asymptotic Bayes-optimality under sparsity of some multiple testing procedures". In: *The Annals of Statistics* 39.3, pp. 1551–1579.

Bonferroni, C.E. (1936). *Teoria statistica delle classi e calcolo delle probabilità*. Pubblicazioni del R. Istituto superiore di scienze economiche e commerciali di Firenze. Seeber.

Brown, Patrick O and David Botstein (1999). "Exploring the new world of the genome with DNA microarrays". In: *Nature genetics* 21.1, pp. 33–37.

Cai, T. Tony and J. Jin (2010). "Optimal rates of convergence for estimating the null density and proportion of nonnull effects in large-scale multiple testing". In: *The Annals of Statistics* 38.1, pp. 100 –145.

Cai, T. Tony and Wenguang Sun (2009). "Simultaneous Testing of Grouped Hypotheses: Finding Needles in Multiple Haystacks". In: *Journal of the American Statistical Association* 104.488, pp. 1467–1481.

Castro, Marcia Caldas de and Burton H Singer (2006). "Controlling the false discovery rate: a new application to account for multiple and dependent tests in local statistics of spatial association". In: *Geographical Analysis* 38.2, pp. 180–208.

Chen, Shuaiyu et al. (2023). "A stable sequential multiple test for Koopman–Darmois family". In: *Journal of Statistical Planning and Inference* 226, pp. 39–62.

Clements, Nicolle et al. (2014). "Applying multiple testing procedures to detect change in East African vegetation". In.

Crockett, NG and V Gebski (1984). "A simplified two sample sequential t-Test". In: *Journal of Statistical Computation and Simulation* 20.3, pp. 217–234.

Datta, J. and J. K. Ghosh (2013). "Asymptotic Properties of Bayes Risk for the Horseshoe Prior". In: *Bayesian Analysis* 8.1, pp. 111–132.

De, Shyamal K. and Michael Baron (2012a). "Sequential Bonferroni Methods for Multiple Hypothesis Testing with Strong Control of Family-Wise Error Rates I and II". In: *Sequential Analysis* 31.2, pp. 238–262.

— (2012b). "Step-up and step-down methods for testing multiple hypotheses in sequential experiments". In: *Journal of Statistical Planning and Inference* 142.7, pp. 2059–2070. ISSN: 0378-3758.

De, Shyamal K and Michael Baron (2015). "Sequential tests controlling generalized familywise error rates". In: *Statistical Methodology* 23, pp. 88–102.

Diène, Bassirou et al. (2020). "Data management techniques for Internet of Things". In: *Mechanical systems and signal processing* 138, p. 106564.

Ding, Xiaojun and Xuan Guo (2018). "A Survey of SNP Data Analysis". In: *Big Data Mining and Analytics* 1.3, pp. 173–190.

Dmitrienko, Alex, Ajit C Tamhane, and Frank Bretz (2009). *Multiple testing problems in pharmaceutical statistics*. CRC press.

Do, Kim-Anh, Peter Müller, and Feng Tang (2005). "A Bayesian mixture model for differential gene expression". In: *Journal of the Royal Statistical Society Series C: Applied Statistics* 54.3, pp. 627–644.

Donoho, David and Jiashun Jin (2004). "Higher criticism for detecting sparse heterogeneous mixtures". In: *The Annals of Statistics* 32.3, pp. 962 –994.

Donoho, David L et al. (2000). "High-dimensional data analysis: The curses and blessings of dimensionality". In: *AMS math challenges lecture* 1.2000, p. 32.

Drus, Zulfadzli and Haliyana Khalid (2019). "Sentiment analysis in social media and its application: Systematic literature review". In: *Procedia computer science* 161, pp. 707–714.

Dudoit, Sandrine and Mark J Van Der Laan (2008). *Multiple testing procedures with applications to genomics*. Springer.

Efron, Bradley (2004). "Large-scale simultaneous hypothesis testing: the choice of a null hypothesis". In: *Journal of the American Statistical Association* 99.465, pp. 96–104.

— (2007). "Size, power and false discovery rates". In: *The Annals of Statistics* 35.4, pp. 1351 –1377.

— (2012). *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*. Vol. 1. Cambridge University Press.

Efron, Bradley and Robert Tibshirani (2002). "Empirical Bayes methods and false discovery rates for microarrays". In: *Genetic epidemiology* 23.1, pp. 70–86.

Efron, Bradley et al. (2001). "Empirical Bayes analysis of a microarray experiment". In: *Journal of the American statistical association* 96.456, pp. 1151–1160.

Fan, Jianqing, Fang Han, and Han Liu (2014). "Challenges of big data analysis". In: *National science review* 1.2, pp. 293–314.

Frommlet, F. and M. Bogdan (2013). "Some optimality properties of FDR controlling rules under sparsity". In: *Electronic Journal of Statistics* 7, pp. 1328–1368.

Genovese, Christopher and Larry Wasserman (2002). "Operating Characteristics and Extensions of the False Discovery Rate Procedure". In: *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 64.3, pp. 499 –517.

Genovese, Christopher R, Kathryn Roeder, and Larry Wasserman (2006). "False discovery control with p-value weighting". In: *Biometrika* 93.3, pp. 509–524.

Genovese, Christopher R and Larry Wasserman (2006). "Exceedance control of the false discovery proportion". In: *Journal of the American Statistical Association* 101.476, pp. 1408–1417.

Ghosh, D. (2020). "Wavelet-based Benjamini-Hochberg procedures for multiple testing under dependence". In: *Mathematical Biosciences and Engineering* 17.1, pp. 56–72.

Golub, Todd R et al. (1999). "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring". In: *science* 286.5439, pp. 531–537.

Green, Peter J and Sylvia Richardson (2002). "Hidden Markov models and disease mapping". In: *Journal of the American statistical association* 97.460, pp. 1055–1070.

Guo, Wenge and Joseph Romano (2007). "A generalized Sidak-Holm procedure and control of generalized error rates under independence". In: *Statistical applications in genetics and molecular biology* 6.1.

He, Jianliang, Bowen Gang, and Luella Fu (2024). "False Discovery Control in Multiple Testing: A Brief Overview of Theories and Methodologies". In: *arXiv preprint arXiv:2411.10647*.

He, Li and Sanat K Sarkar (2013). "On improving some adaptive BH procedures controlling the FDR under dependence". In: *Electronic Journal of Statistics* 7, pp. 2683–2701.

He, Xinrui and Jay Bartroff (2021). "Asymptotically optimal sequential FDR and pFDR control with (or without) prior information on the number of signals". In: *Journal of Statistical Planning and Inference* 210, pp. 87–99. ISSN: 0378-3758.

Heuvel, Edwin R van den and Chris AJ Klaassen (1999). "Bayes convolution". In: *International statistical review* 67.3, pp. 287–299.

Hochberg, Yosef (1988). "A sharper Bonferroni procedure for multiple tests of significance". In: *Biometrika* 75.4, pp. 800–802. ISSN: 0006-3444.

Hochberg, Yosef and Ajit C Tamhane (1987). *Multiple comparison procedures*. John Wiley & Sons, Inc.

Holm, Sture (1979). "A Simple Sequentially Rejective Multiple Test Procedure". In: *Scandinavian Journal of Statistics* 6.2, pp. 65–70.

Hommel, G. (1988). "A stagewise rejective multiple test procedure based on a modified Bonferroni test". In: *Biometrika* 75.2, pp. 383–386. ISSN: 0006-3444.

Jennison, Christopher and Bruce W Turnbull (2000). *Group sequential methods with applications to clinical trials*. CRC Press.

Jin, Jiashun (2009). "Impossibility of successful classification when useful features are rare and weak". In: *Proceedings of the National Academy of Sciences* 106.22, pp. 8859–8864.

Johnstone, Iain M (2007). "High dimensional statistical inference and random matrices". In: *Proceedings of the International Congress of Mathematicians Madrid, August 22–30, 2006*. European Mathematical Society-EMS-Publishing House GmbH, pp. 307–333.

Jothi, R, Sraban Kumar Mohanty, and Aparajita Ojha (2019). "DK-means: a deterministic k-means clustering algorithm for gene expression analysis". In: *Pattern Analysis and Applications* 22, pp. 649–667.

Khashei, Mehdi, Ali Zeinal Hamadani, and Mehdi Bijari (2012). "A fuzzy intelligent approach to the classification problem in gene expression data analysis". In: *Knowledge-Based Systems* 27, pp. 465–474.

Kim, Sunyong, Seiya Imoto, and Satoru Miyano (2004). "Dynamic Bayesian network and nonparametric regression for nonlinear modeling of gene networks from time series gene expression data". In: *Biosystems* 75.1-3, pp. 57–65.

Lai, Tze Leung (2001). "Sequential analysis: some classical problems and new challenges". In: *Statistica Sinica*, pp. 303–351.

Lai, Tze Leung and Haipeng Xing (2008). *Statistical models and methods for financial markets*. Vol. 1017. Springer.

Laird, Peter W (2010). "Principles and challenges of genome-wide DNA methylation analysis". In: *Nature Reviews Genetics* 11.3, pp. 191–203.

Liu, Fang and Sanat K Sarkar (2011). "A new adaptive method to control the false discovery rate". In: *Recent advances in biostatistics: false discovery rates, survival analysis, and related topics*. World Scientific, pp. 3–26.

Lockhart, David J and Elizabeth A Winzeler (2000). "Genomics, gene expression and DNA arrays". In: *Nature* 405.6788, pp. 827–836.

Malloy, Matthew L and Robert D Nowak (2014). "Sequential testing for sparse recovery". In: *IEEE Transactions on Information Theory* 60.12, pp. 7862–7873.

Mei, Yajun (2010). "Efficient scalable schemes for monitoring a large number of data streams". In: *Biometrika* 97.2, pp. 419–433.

Miller, Christopher J et al. (2001). "Controlling the false-discovery rate in astrophysical data analysis". In: *The Astronomical Journal* 122.6, p. 3492.

Müller, Peter et al. (2004). "Optimal sample size for multiple testing: the case of gene expression microarrays". In: *Journal of the American Statistical Association* 99.468, pp. 990–1001.

Neuvial, P. and E. Roquain (2012). "On false discovery rate thresholding for classification under sparsity". In: *The Annals of Statistics* 40.5, pp. 2572–2600.

Pocock, Stuart J (2013). *Clinical trials: a practical approach*. John Wiley & Sons.

Pratt, John W (1960). "On interchanging limits and integrals". In: *The Annals of Mathematical Statistics* 31.1, pp. 74–77.

Qin, Zikun and Malay Ghosh (2025). "Bayes oracle property of multiple tests of multivariate normal means under sparsity". In: *Journal of Statistical Planning and Inference* 235, p. 106227.

Ramdas, Aaditya et al. (2017). "Online control of the false discovery rate with decaying memory". In: *Advances in neural information processing systems* 30.

Ramdas, Aaditya et al. (2018). "SAFFRON: an adaptive algorithm for online control of the false discovery rate". In: *International conference on machine learning*. PMLR, pp. 4286–4294.

Rane, Nitin Liladhar et al. (2024). "Machine learning and deep learning for big data analytics: A review of methods and applications". In: *Partners Universal International Innovation Journal* 2.3, pp. 172–197.

Robbins, Herbert (1951). "Asymptotically subminimax solutions of compound statistical decision problems". In: *Proceedings of the second Berkeley symposium on mathematical statistics and probability*. Vol. 2. University of California Press, pp. 131–149.

Robertson, David S, James MS Wason, and Aaditya Ramdas (2023). "Online multiple hypothesis testing". In: *Statistical science: a review journal of the Institute of Mathematical Statistics* 38.4, p. 557.

Romano, Joseph P and Michael Wolf (2007). "Control of generalized error rates in multiple testing". In.

Rosati, Diletta et al. (2024). "Differential gene expression analysis pipelines and bioinformatic tools for the identification of specific biomarkers: A review". In: *Computational and structural biotechnology journal* 23, pp. 1154–1168.

Sarkar, Sanat (2004). "FDR-Controlling Stepwise Procedures and Their False Negatives Rates". In: *Journal of Statistical Planning and Inference* 125, pp. 119–137.

— (2008a). "Two-stage stepup procedures controlling FDR". In: *Journal of Statistical Planning and Inference*, pp. 1072–1084.

Sarkar, Sanat K. (2002). "Some Results on False Discovery Rate in Stepwise multiple testing procedures". In: *The Annals of Statistics* 30.1, pp. 239 –257.

— (2006). "False discovery and false nondiscovery rates in single-step multiple testing procedures". In: *The Annals of Statistics* 34.1, pp. 394 –415.

Sarkar, Sanat K (2008b). "On methods controlling the false discovery rate". In: *Sankhyā: The Indian Journal of Statistics, Series A (2008-)*, pp. 135–168.

Sarkar, Sanat K, Wenge Guo, and Helmut Finner (2012). "On adaptive procedures controlling the familywise error rate". In: *Journal of Statistical Planning and Inference* 142.1, pp. 65–78.

Sarkar, Sanat K and Zhigen Zhao (2022). "Local false discovery rate based methods for multiple testing of one-way classified hypotheses". In: *Electronic Journal of Statistics* 16.2, pp. 6043–6085.

Sarkar, Sanat K, Tianhui Zhou, and Debashis Ghosh (2008). "A general decision theoretic formulation of procedures controlling FDR and FNR from a Bayesian perspective". In: *Statistica Sinica*, pp. 925–945.

Schena, Mark et al. (1995). "Quantitative monitoring of gene expression patterns with a complementary DNA microarray". In: *Science* 270.5235, pp. 467–470.

Scott, James G and James O Berger (2006). "An exploration of aspects of Bayesian multiple testing". In: *Journal of statistical planning and inference* 136.7, pp. 2144–2162.

— (2010). "Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem". In: *The Annals of Statistics*, pp. 2587–2619.

Siegmund, David (2013). *Sequential analysis: tests and confidence intervals*. Springer Science & Business Media.

Simes, R John (1986). "An improved Bonferroni procedure for multiple tests of significance". In: *Biometrika* 73.3, pp. 751–754.

Singh, Dinesh et al. (2002). "Gene expression correlates of clinical prostate cancer behavior". In: *Cancer cell* 1.2, pp. 203–209.

Sjölander, Arvid and Stijn Vansteelandt (2019). "Frequentist versus Bayesian approaches to multiple testing". In: *European journal of epidemiology* 34, pp. 809–821.

Solari, Aldo and Jelle J Goeman (2017). "Minimally adaptive BH: A tiny but uniform improvement of the procedure of Benjamini and Hochberg". In: *Biometrical Journal* 59.4, pp. 776–780.

Song, Yanglei and Georgios Fellouris (2017). "Asymptotically optimal, sequential, multiple testing procedures with prior information on the number of signals". In: *Electronic Journal of Statistics* 11.1, pp. 338 –363.

— (2019). "Sequential multiple testing with generalized error control: An asymptotic optimality theory". In: *The Annals of Statistics* 47.3, pp. 1776 –1803.

Storey, John D. (2002). "A direct approach to false discovery rates". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64.3, pp. 479–498.

— (2003). "The positive false discovery rate: a Bayesian interpretation and the q-value". In: *The Annals of Statistics* 31.6, pp. 2013 –2035.

Storey, John D., Jonathan E. Taylor, and David Siegmund (2004). "Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66.1, pp. 187–205.

Sun, Wenguang and T. Tony Cai (2007a). "Oracle and Adaptive Compound Decision Rules for False Discovery Rate Control". In: *Journal of the American Statistical Association* 102.479, pp. 901–912.

Sun, Wenguang and T Tony Cai (2007b). "Oracle and adaptive compound decision rules for false discovery rate control". In: *Journal of the American Statistical Association* 102.479, pp. 901–912.

Sun, Wenguang and T. Tony Cai (2009). "Large-scale multiple testing under dependence". In: *Journal of the Royal Statistical Society Series B* 71.2, pp. 393–424.

Sun, Wenguang et al. (2015). "False discovery control in large-scale spatial multiple testing". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 77.1, pp. 59–83.

Tartakovsky, Alexander G, X Rong Li, and George Yaralov (2003). "Sequential detection of targets in multichannel systems". In: *IEEE Transactions on Information Theory* 49.2, pp. 425–445.

Tian, Jinjin and Aaditya Ramdas (2021). "Online control of the familywise error rate". In: *Statistical methods in medical research* 30.4, pp. 976–993.

Tong, Yung Liang (2012). *The multivariate normal distribution*. Springer Science & Business Media.

VanGuilder, Heather D., Kent E. Vrana, and Willard M. Freeman (2008). "Twenty-Five Years of Quantitative PCR for Gene Expression Analysis". In: *BioTechniques* 44.5, pp. 619–626.

Wald, Abraham (1992). "Sequential tests of statistical hypotheses". In: *Breakthroughs in statistics: Foundations and basic theory*. Springer, pp. 256–298.

Wald, Abraham and Jacob Wolfowitz (1948). "Optimum character of the sequential probability ratio test". In: *The Annals of Mathematical Statistics*, pp. 326–339.

Weinstein, Asaf and Aaditya Ramdas (2020). "Online control of the false coverage rate and false sign rate". In: *International Conference on Machine Learning*. PMLR, pp. 10193–10202.

Westfall, Peter H, Randall D Tobias, and Russell D Wolfinger (2011). *Multiple comparisons and multiple tests using SAS*. Sas Institute.

Westfall, Peter H and S Stanley Young (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment*. John Wiley & Sons.

Xie, Jichun et al. (2010). "False Discovery Rate Control for High Dimensional Dependent Data with an Application to Large-Scale Genetic Association Studies". In: *The Annals of Applied Statistics* 41, pp. 1–28.

Xu, Ziyu and Aaditya Ramdas (2022). "Dynamic algorithms for online multiple testing". In: *Mathematical and Scientific Machine Learning*. PMLR, pp. 955–986.

Yang, Fanny et al. (2017). "A framework for multi-a (rmed)/b (andit) testing with online fdr control". In: *Advances in Neural Information Processing Systems* 30.

Šidák, Zbyněk (1967). "Rectangular Confidence Regions for the Means of Multivariate Normal Distributions". In: *Journal of the American Statistical Association* 62.318, pp. 626–633.