

# Multiple Hypothesis Testing Under Dependence

Nabaneet Das



Interdisciplinary Statistical Research Unit

Indian Statistical Institute

2025



# Indian Statistical Institute

Doctoral Thesis

---

## Multiple Hypothesis Testing Under Dependence

---

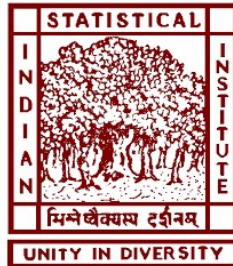
(Final Version)

*Author*

**Nabaneet Das**

*Thesis Advisor*

**Prof. Subir Kumar Bhandari**



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

Interdisciplinary Statistical Research Unit

Applied Statistics Division



# Certificate from Supervisor

This is to certify that this doctoral dissertation entitled “Multiple Hypothesis Testing Under Dependence” submitted by Nabaneet Das to Indian Statistical Institute is a record of bonafide research work under my guidance and supervision. The work contained in this dissertation is original and has not been submitted elsewhere for any degree.

---

**Prof. Subir Kumar Bhandari**

Professor

Interdisciplinary Statistical Research Unit

Indian Statistical Institute, Kolkata

# Declaration of Authorship

I, Nabaneet Das, declare that this dissertation titled “Multiple Hypothesis Testing Under Dependence” and the works presented in it are my own. I confirm the following statements.

- The 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 that 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 helped in developing different parts of this dissertation.

*Nabaneet Das*

---

Nabaneet Das

21/02/2025

Senior Research Fellow in Statistics

Interdisciplinary Statistical Research Unit

Indian Statistical Institute, Kolkata



# Abstract

We have explored various aspects of multiple hypothesis testing in the context of dependent hypotheses. Traditional methods often assume that tests are either independent or weakly dependent. However, in large-scale testing scenarios, such as those encountered in genomics, where variables like gene expressions are frequently correlated due to shared biological processes, these assumptions can be inadequate. Existing literature has not sufficiently addressed the performance of traditional algorithms under such dependence, which can lead to excessive conservatism and a consequent loss of statistical power.

One widely used method for correcting for multiplicity is the Bonferroni correction. This classical approach controls the Family-Wise Error Rate (FWER) at a specified level, regardless of the dependence among hypotheses. However, under the global null hypothesis and an equicorrelated normal model, the actual FWER of the Bonferroni procedure is constrained by a line connecting 0 and the desired significance level ( $\alpha$ ), making it overly conservative. We have examined the FWER of the Bonferroni method in nearly independent settings and demonstrated that it approximates the FWER under independence when the correlations are smaller than the order of  $\log n$ .

Our investigation also extends to the intersection of multiple testing and classification theory. The Bayes rule, which is optimal for traditional classification problems, also proves optimal for multiple testing under certain mild assumptions. Nonetheless, simplifying the test statistic derived from the Bayes rule under dependence remains challenging, limiting its practical application. We have successfully simplified the optimal test statistic in a Gaussian model and, through extensive simulations, shown that the Oracle rule significantly outperforms traditional methods, such as the Benjamini-Hochberg (BH)

procedure for False Discovery Rate (FDR) control.

Lastly, we addressed the challenge of estimating the proportion of null hypotheses in the presence of dependence. We have shown that the estimator proposed by Benjamini and Hochberg converges to 1 under independence. Our simulations evaluate this estimator's performance under various dependent structures, including  $m$ -dependent and block-dependent scenarios. We also assessed Storey's estimator and another alternative to gauge their effectiveness in these contexts.



# Acknowledgments

First and foremost, I would like to express my deepest gratitude to my Ph.D. supervisor, Prof. Subir Kumar Bhandari. His unwavering support, encouragement, and patient guidance have been instrumental throughout my Ph.D. journey. I consider myself incredibly fortunate to have him as my thesis advisor. This thesis owes its existence to his meticulous mentorship, nurturing approach, and constant enthusiasm. He not only taught me how to conduct research but also inspired me to continuously push my boundaries. I am profoundly grateful for his resilience and belief in me, even during the challenging times. The years of engaging discussions and learning experiences with him are memories I will cherish forever.

I would also like to extend my heartfelt thanks to all the previous and current members of the PhD-DSc Committee (Statistics) and the Research Fellow Advisory Committee (Statistics) for their invaluable guidance. My appreciation goes to the faculty and non-academic staff of the Interdisciplinary Statistical Research Unit for fostering an excellent academic environment.

I would like to thank the reviewers who provided useful suggestions that lead to improvement of the earlier version of this thesis.

I am especially grateful to Prof. Antar Bandyopadhyay, Prof. Probal Chaudhuri, and Prof. Ayanendranath Basu for their engaging courses, which have greatly deepened both my interest in and understanding of statistics and probability.

I am fortunate to have had the constant support and companionship of my friends, seniors, and juniors throughout this journey. I am particularly grateful to Monitirtha, Subhasis, Soumya, Rahul, Sujay da, Amarnath, and Subhankar for making my time at ISRU both

memorable and enjoyable. My gratitude also extends to my colleagues and seniors at the Reserve Bank of India for their encouragement and support.

This acknowledgment would be incomplete without recognizing the patience and unwavering support of my family. Their constant encouragement and inspiration have driven me to pursue and achieve my academic goals.

Lastly, I thank God for bestowing upon me strength, guidance, and for keeping my loved ones and me safe throughout this journey.



# Journal publications and preprints

- ◆ A version of chapter 2 has been published online as the following.

Das, Nabaneet & Bhandari, Subir Kumar, 2021. “Bound on FWER for correlated normal,” *Statistics & Probability Letters*, Elsevier, vol. 168(C)

DOI : <https://doi.org/10.1016/j.spl.2020.108943>

- ◆ A version of chapter 3 has been published online as the following.

Das, Nabaneet & Bhandari, Subir Kumar, 2025. “FWER for normal distribution in nearly independent setup” *Statistics & Probability Letters*, Elsevier, vol. 219

DOI : <https://doi.org/10.1016/j.spl.2024.110340>

- ◆ Chapter 4 is based on the following preprint.

Das, Nabaneet, and Subir K. Bhandari. “Optimal test statistic under normality assumption.”

[arXiv preprint arXiv:2306.10554](https://arxiv.org/abs/2306.10554) (2023).

- ◆ Chapter 5 is based on the following preprint.

Das, Nabaneet. “Estimation of Proportion of Null Hypotheses Under Dependence.”

[arXiv preprint arXiv:2409.04100](https://arxiv.org/abs/2409.04100) (2024).



# Contents

Certificate from Supervisor	i
Declaration of Authorship	ii
Abstract	iii
Acknowledgments	v
Journal publications and preprints	vii
<b>1 Introduction</b>	<b>1</b>
1.1 Overview of Multiple Testing Problems	2
1.1.1 Outcome of a Multiple Testing Procedure	3
1.1.2 Type I error rates in Multiple testing	3
1.1.3 Relations among different type I error rate criteria	5
1.1.4 Type II error rates	5
1.1.5 Multiple testing procedures	7
1.1.6 Single step procedures	7
1.1.7 Stepwise procedures	8
1.1.8 Dependence among Hypotheses	11
1.1.9 Existing approaches towards dependence among hypotheses	11
1.2 Motivation and brief overview of our work	13
<b>2 Asymptotic FWER for Correlated Normal distribution</b>	<b>15</b>
2.1 Introduction	15

2.2	Description of the setup . . . . .	17
2.3	Simulation results . . . . .	17
2.4	Main theorem . . . . .	22
2.4.1	An alternate form of $H(\rho)$ and it's derivatives . . . . .	23
2.4.2	Proof of the main theorem . . . . .	24
2.4.3	Proof of step 2 . . . . .	24
2.4.4	Proof of step 3 . . . . .	28
2.4.5	Appendix . . . . .	29
2.5	Simulation results . . . . .	30
2.6	FDR of BH procedure . . . . .	32
<b>3</b>	<b>FWER under nearly independent setup</b>	<b>37</b>
3.1	Introduction . . . . .	37
3.2	Description of the problem . . . . .	38
3.3	Motivation behind the nearly independent setup . . . . .	39
3.4	FWER under the nearly independent setup . . . . .	40
3.5	Simulation results . . . . .	44
<b>4</b>	<b>Optimal test statistic under Gaussian distribution</b>	<b>46</b>
4.1	Introduction . . . . .	46
4.2	Oracle Decision rule for Multivariate normal model . . . . .	48
4.2.1	Discussion on error rate criteria . . . . .	49
4.2.2	Oracle decision rule for multiple testing problem . . . . .	51
4.2.3	Implementation of the optimal test . . . . .	53
4.3	Oracle decision rule for Gaussian model . . . . .	54
4.3.1	Simulation Studies . . . . .	55
4.4	Annexure . . . . .	68
4.4.1	Proof of theorem 4.3.2 . . . . .	68
<b>5</b>	<b>Estimation of <math>\pi_0</math> under dependence</b>	<b>70</b>
5.1	Introduction . . . . .	70

5.2	The model and identifiability . . . . .	71
5.3	Inconsistency of the BH estimator . . . . .	73
5.3.1	Proof of theorem 5.3.1 . . . . .	74
5.3.2	Behaviour of $\hat{\pi}_{0,BH}$ for $0 < \pi_0 < 1$ . . . . .	79
5.3.3	Behaviour of $\hat{\pi}_{0,BH}$ under dependence . . . . .	80
5.4	Estimator proposed in Storey (2002) . . . . .	83
5.4.1	Simulation results . . . . .	85
5.5	Estimator proposed in Patra and Sen (2016) . . . . .	90
5.5.1	Mixing sequence . . . . .	93
5.5.2	Simulation results . . . . .	94
5.6	Remarks . . . . .	97
5.7	Appendix . . . . .	98
5.7.1	Proof of theorem 5.5.1 . . . . .	98
5.7.2	Proof of theorem 5.5.2 . . . . .	99
	<b>Bibliography</b> . . . . .	<b>101</b>

# List of Figures

2.1	Plots of FWER for various values of $N$ ( $\alpha = 0.01$ ) . . . . .	19
2.2	Plots of FWER for various values of $N$ ( $\alpha = 0.05$ ) . . . . .	20
2.3	Plots of FWER for various values of $N$ ( $\alpha = 0.1$ ) . . . . .	21
2.4	Plots of FDR of BH procedure for various values of $N$ ( $\alpha = 0.01$ ) . . . . .	33
2.5	Plots of FDR of BH procedure for various values of $N$ ( $\alpha = 0.05$ ) . . . . .	34
2.6	Plots of FDR of BH procedure for various values of $N$ ( $\alpha = 0.1$ ) . . . . .	35

# List of Tables

1.1	Outcomes of a multiple testing problem . . . . .	3
2.1	Simulation results . . . . .	31
3.1	FWER for level $\alpha = 0.01$ . . . . .	45
3.2	FWER for level $\alpha = 0.05$ . . . . .	45
3.3	FWER for level $\alpha = 0.1$ . . . . .	45
3.4	FWER for level $\alpha = 0.2$ . . . . .	45
4.1	FDRs of the three methods . . . . .	57
4.2	FNRs of the three methods . . . . .	60
4.3	No. of rejections of the three methods . . . . .	63
4.4	FDRs in block diagonal case . . . . .	66
4.5	FNRs in block diagonal case . . . . .	67
4.6	No. of rejections in block diagonal case . . . . .	67
5.1	$\hat{\pi}_{0,BH}$ for Gaussian mean test for $n = 10^4$ . . . . .	80
5.2	$\hat{\pi}_{0,BH}$ for Gaussian mean test for $n = 10^7$ . . . . .	80
5.3	$\hat{\pi}_{0,BH}$ for Gaussian mean test for $m = 2$ and $n = 1000$ . . . . .	81
5.4	$\hat{\pi}_{0,BH}$ for Gaussian mean test for $m = 5$ and $n = 1000$ . . . . .	81
5.5	$\hat{\pi}_{0,BH}$ for Gaussian mean test for $m = 2$ and $n = 10000$ . . . . .	82
5.6	$\hat{\pi}_{0,BH}$ for Gaussian mean test for $m = 5$ and $n = 10000$ . . . . .	82
5.7	$\hat{\pi}_{0,BH}$ for Gaussian mean test under block dependence for $n = 1000$ . . . . .	82
5.8	$\hat{\pi}_{0,BH}$ for Gaussian mean test under block dependence for $n = 10000$ . . . . .	83

5.9	$n = 1000$ and $a = 0.1, 0.2$ . . . . .	87
5.10	$n = 1000$ and $a = 0.5, 0.75$ . . . . .	87
5.11	$n = 1000$ and $m = 1, 2$ . . . . .	88
5.12	$n = 1000$ and $m = 1, 2, 5$ for the equally weighted case . . . . .	89
5.13	$n = 10^6$ and $m = 1, 2, 5$ for the equally weighted case . . . . .	89
5.14	Block dependent variables with $n = 1000$ and 10 blocks . . . . .	90
5.15	Block dependent variables with $n = 1000$ and 200 blocks . . . . .	90
5.16	AR(1) model with $a = 0.1$ and $0.2$ . . . . .	94
5.17	AR(1) model with $a = 0.5$ and $0.75$ . . . . .	95
5.18	Estimator of null proportion for m-dependent model with $m = 1, 2, 5$ . . . . .	96
5.19	Equally weighted m-dependent model with $m = 1, 2, 5$ . . . . .	96
5.20	Block dependent case with $n = 1000$ . . . . .	97

# Chapter 1

## Introduction

In the realm of scientific inquiry, the process of hypothesis testing serves as the cornerstone upon which robust conclusions are built. However, as data sets grow in complexity and size, researchers often find themselves faced with the challenge of simultaneously testing multiple hypotheses. Current statistical inference problems in areas such as astronomy, genomics etc. routinely involve the simultaneous testing of thousands, or even millions of null hypotheses. These hypotheses concern a wide range of parameters, for high dimensional multivariate distributions, with complex and unknown dependence structure among variables. Example of testing problems include the following:

- The identification of differentially expressed genes in microarray experiments.
- In drug discovery, researchers may test multiple compounds for efficacy or safety simultaneously. In clinical trials, multiple hypotheses may be tested to evaluate the effectiveness of different treatment regimens or to identify potential adverse effects.
- In finance, multiple hypothesis testing is used in portfolio management, where investors test multiple trading strategies simultaneously to maximize returns while managing risk. In economics, researchers may test multiple hypotheses to understand the impact of various policies or economic factors on outcomes.

- In environmental studies, researchers may test multiple hypotheses to understand the impact of different factors on ecosystems or environmental processes. For example, testing the effect of various pollutants on water quality or biodiversity.
- In the vast expanse of astronomy, multiple hypothesis testing is crucial for detecting exoplanets and assessing the distribution of key variables, such as metallicity and radial velocity.
- Detection of outperforming stocks in the market.

Classical theories of statistical inference, however, fail to grapple adequately with the multiplicity conundrum posed by these challenges. It is precisely these multiple testing procedures that form the crux of this thesis.

## 1.1 Overview of Multiple Testing Problems

In a classical multiple hypothesis testing framework, the problem is to test a set of  $n$  hypotheses as follows.

$$H_{0i} \text{ vs } H_{1i} \quad i \in \{1, 2, \dots, n\}$$

We denote the set of true null hypotheses by  $I_0$  and the set of false null hypotheses by  $I_1$ . If  $|I_0| = n$ , then all null hypotheses are true and this case is known as “*intersection null*” or “*global null*” (Often denoted as  $\bigcap_{i=1}^n H_{0i}$ ).

Each of the  $n$  testing problems has its own notion of type I and type II error. It is interesting to note that, if all hypotheses are tested at a level of significance  $\alpha$ , then under global null and assuming that all testing problems are independent,

$$P(\text{At least one false rejection}) = 1 - (1 - \alpha)^n > \alpha \quad \forall n > 1$$

The provided example serves to underscore a critical point: even in straightforward scenarios like the intersection null, it’s imperative to meticulously manage the overall

error rate of multiple testing problems. In the subsequent section, we delve into the concept of error rates within the realm of multiple testing.

### 1.1.1 Outcome of a Multiple Testing Procedure

For a simultaneous testing of  $n$  null hypotheses  $H_{01}, H_{02}, \dots, H_{0n}$ , the outcomes are summarized in the following table.

**Table 1.1:** Outcomes of a multiple testing problem

Hypothesis	Claimed non-significant	Claimed significant	Total
Null	$U_n$	$V_n$	$n_0$
Non-null	$T_n$	$S_n$	$n_1$
Total	$A_n$	$R_n$	$n$

In the simultaneous testing problem,  $V_n$  denote the number of type I errors (i.e. false discoveries) and  $T_n$  denote the number of type II errors (i.e. true acceptances).  $R_n$  denote the number of rejections or “discoveries” and  $|I_0| = n_0$  and  $|I_1| = n_1$ . In terms of the outcomes described in table 1.1, the error rates are defined as follows.

### 1.1.2 Type I error rates in Multiple testing

Several notions of type I error rate are available in multiple testing literature.

**Definition 1.** *Family wise error rate (FWER) =  $P(V_n \geq 1)$*

The Family-Wise Error Rate (FWER) stands as a classical measure, widely acknowledged for its significance in assessing the Type I error rate, representing the probability of at least one false rejection. Controlling FWER in the context of multiple testing has been considered in [Holm \(1979\)](#), [Hochberg \(1987\)](#), [Simes \(1986\)](#), [Hommel \(1988\)](#) and a comprehensive review is given in [Westfall and Young \(1993\)](#), [Dudoit et al. \(2003\)](#), [Efron \(2012\)](#). Despite the rich literature addressing FWER control, algorithms designed for this purpose often tend towards conservatism, warranting further exploration and refinement. A generalized version of FWER (namely k-FWER) has been proposed by [Lehmann and Romano \(2012\)](#) which is probability of at least k false rejections.

**Definition 2.**  $k$ -FWER =  $P(V_n \geq k)$

Although FWER dominates the traditional multiple comparison literature, some other measures of type I error rates are as follows.

**Definition 3.** *Per-comparison error rate (PCER)* =  $\frac{E[V_n]}{n}$

**Definition 4.** *Per-family error rate (PFER)* =  $E[V_n]$

PFER is the expected value of number of type I errors while PCER is the expected proportion of type I errors among the  $n$  tests. However, the inherent limitations of these conservative procedures underscored the need for further research in this domain. In the latter half of the twentieth century, [Benjamini and Hochberg \(1995\)](#) pioneered a groundbreaking concept: the False Discovery Rate (FDR).

**Definition 5.** *False discovery rate (FDR)* =  $E \left[ \frac{V}{R} I_{\{R>0\}} \right]$

Since its debut in the seminal 1995 paper, the False Discovery Rate (FDR) has garnered an expanding share of attention within statistical research circles. It has progressed to attain the coveted status of “accepted methodology” in scientific subject-matter journals. [Storey \(2002\)](#) introduced a notion of positive false discovery rate (pFDR) which is defined as follows.

**Definition 6.** *Positive false discovery rate (pFDR)* =  $E \left[ \frac{V}{R} \mid R > 0 \right]$

The term “positive” stems from the notion that it denotes the anticipated proportion of false discoveries, assuming that the number of rejections should be positive. [Storey \(2002\)](#) has elaborated on this concept, demonstrating that, in the case of independent observations from a mixture distribution, the positive false discovery rate (pFDR) can be construed as a Bayesian posterior probability. However, dealing with the expectation of the ratio of random variables presents substantial mathematical challenges. As an alternative with relatively more tractability, the marginal false discovery rate (mFDR) offers a similar error rate criterion.

**Definition 7.** *Marginal false discovery rate (mFDR)* =  $\frac{E[V]}{E[R]}$

Different measures of type I error rate serves different purpose in the domain of multiple testing. In experiments where even a single false positive could have severe consequences (e.g.- confirmatory clinical trials ([Dmitrienko et al. \(2009\)](#))) stringent control over the Family-Wise Error Rate (FWER) remains paramount. Conversely, in exploratory studies, where a more lenient approach may be acceptable, attention is often directed towards controlling the False Discovery Rate (FDR).

### 1.1.3 Relations among different type I error rate criteria

The following inequalities hold between the measures of type I error rates defined earlier.

- $PCER \leq FDR \leq FWER \leq PFER$
- $FDR = pFDR Pr(R > 0) \leq pFDR$
- $FDR \leq mFDR$
- $k\text{-FWER} \leq \frac{PFER}{k}$

The aforementioned inequalities shed light on the varying degrees of conservatism inherent in different criteria. For instance, given that  $FDR \leq FWER$ , an algorithm designed to maintain FWER at a predetermined level inherently manages FDR as well, although the reverse isn't necessarily true. But it is interesting to note that, under global null, FDR and FWER are the same. An extensive review of the different types of Type I error rates and their interrelationships can be found in [Dudoit and Van Der Laan \(2007\)](#), [Efron \(2012\)](#).

### 1.1.4 Type II error rates

In hypothesis testing, a prevalent strategy involves minimizing the occurrence of type II errors while maintaining a predetermined level of control over type I errors. Within the domain of multiple testing, several concepts pertaining to type II errors (or equivalently power) emerge, including the following ([Dudoit et al. \(2008\)](#)).

- The probability of rejecting at least one false null hypothesis is termed as “*AnyPwr*”.

**Definition 8.**  $AnyPwr = P(S_n \geq 1)$

- One might be interested in identifying all existing effects in a simultaneous testing study. The notion of power suitable for such scenario is “*AllPwr*” which is the probability of rejecting all false null hypotheses.

**Definition 9.**  $AllPwr = P(T_n = 0) = P(S_n = n_1)$

- The average power (*AvgPwr*) is the expected value of the proportion of true positives among the false null hypotheses.

**Definition 10.**  $AvgPwr = \frac{E[S_n]}{n_1} = 1 - \frac{E[T_n]}{n_1}$

- The expected value of the proportion of true positives among the rejected hypotheses is termed as “*true discovery rate*” (TDR).

**Definition 11.**  $True\ discovery\ rate\ (TDR) = E \left[ \frac{S_n}{R_n} I_{\{R_n > 0\}} \right]$

It is interesting to note that,  $TDR = 1 - FDR$

- Perhaps the most appropriate quantity which complements FDR is False non-discovery rate (FNR). FNR is the expected proportion of false non-discoveries among all non-discoveries.

**Definition 12.**  $False\ non-discovery\ rate\ (FNR) = E \left[ \frac{T_n}{A_n} I_{\{A_n > 0\}} \right] = E \left[ \frac{T_n}{n - R_n} I_{\{R_n < n\}} \right]$

[Genovese and Wasserman \(2002\)](#) first introduced FNR as a dual quantity to control along with FDR.

- [Storey \(2002\)](#) proposed positive false non-discovery rate (pFNR) to complement pFDR. It is simply the conditional expectation of the proportion of false non-discoveries among all non-discoveries given  $R_n < n$ .

**Definition 13.**  $Positive\ false\ non-discovery\ rate\ (pFDR) = E \left[ \frac{T_n}{A_n} | A_n > 0 \right]$

- [Sun and Cai \(2007\)](#), [Sun and Tony Cai \(2009\)](#) considered marginal false non-discovery rate to complement mFDR.

**Definition 14.** *Marginal false non-discovery rate (mFNR)*  $= \frac{E[T_n]}{E[A_n]} = \frac{E[T_n]}{E[n-R_n]}$

Sun and Tony Cai (2009), Xie et al. (2011) have established a one to one correspondence between a two-class weighted classification problem and multiple testing problem in terms of a combination of mFDR and mFNR.

We will now focus on several multiple testing procedures.

### 1.1.5 Multiple testing procedures

For simultaneous testing of  $n$  null hypotheses  $H_{01}, \dots, H_{0n}$ , a decision rule is a function on the sample space  $\mathcal{X}$  which satisfies  $\delta : \mathcal{X} \rightarrow \{0, 1\}^n$  and

$$\delta_i = \begin{cases} 1 & \text{if } i\text{-th hypothesis is rejected} \\ 0 & \text{otherwise} \end{cases}$$

Based on the type of decision rules, multiple testing procedures can be classified into two categories.

- *Margin based procedure* : In this case,  $\delta_i(\mathbf{x})$  depends only on  $x_i$ . (Where  $\delta_i(\cdot)$  is the  $i$ -th component of  $\delta = (\delta_1, \dots, \delta_n)$ )
- *Joint procedure* : In this case, the decision rule is not as simple as the marginal based procedures. The decision on  $i$ -th null hypothesis is taken based on all observations instead of the single observation  $x_i$ .

Our primary focus will be on margin-based procedures due to their simplicity in implementation. These procedures can be categorized into two main types: single-step and stepwise procedures.

### 1.1.6 Single step procedures

Single step procedures are usually of the form

$$\delta_i(\mathbf{x}) = I_{\{T_i \leq c\}}$$

where  $T = (T_1, \dots, T_n)$  are some test statistics and  $c$  is the cut-off. The following examples are based on p-values  $P_1, \dots, P_n$  of  $n$  tests and suppose  $\alpha$  is the desired level of significance.

- **Bonferroni's procedure** : Bonferroni's procedure is perhaps the oldest method of FWER control where the decision rule is as follows.

$$\delta_i(\mathbf{x}) = I_{\{P_i \leq \frac{\alpha}{n}\}}$$

This algorithm control FWER at the desired level under any configuration of null and non-null hypotheses and under any dependence structure among the hypotheses.

- **Sidak's procedure** : Since the Bonferroni's method becomes very conservative in practice, a slight improvement can be achieved by using Sidak's procedure ([Šidák \(1967\)](#)).

$$\delta_i(\mathbf{x}) = I_{\{P_i \leq 1 - (1 - \alpha)^{\frac{1}{n}}\}}$$

This procedure controls FWER at level  $\alpha$  when the hypotheses are independent.

### 1.1.7 Stepwise procedures

The stepwise algorithms of multiple testing started to gain popularity mainly in the late 1980s. In this section, we delve into a comprehensive review of such procedures.

Consider the n-simplex

$$S_n = \{(t_1, \dots, t_n) \in \mathbb{R}^n \mid 0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq 1\}$$

A step down multiple testing rule based on p values  $P_1, \dots, P_n$  makes use of the ordered p values  $P_{(1)} \leq P_{(2)} \leq \dots \leq P_{(n)}$  and a cut-off  $s = (s_1, \dots, s_n) \in S_n$ . Let,

$$k_0 = \max\{i \mid P_{(j)} \leq s_j \ \forall j = 1, \dots, i\}$$

Then, a step down procedure rejects  $H_{(01)}, \dots, H_{(0k_0)}$  (where  $H_{(i)}$  is the null hypothesis corresponding to the ordered p value  $P_{(i)}$ )

### Examples of step down procedures

- [Holm \(1979\)](#) proposed a step down procedure for controlling FWER with

$$s_i = \frac{\alpha}{n - i + 1} \quad \forall i = 1, \dots, n$$

Among the earliest step-down procedures proposed in the literature was Holm's procedure. Notably, the thresholds set by this method are higher than those of Bonferroni's method, yet it effectively maintains control over the Family-Wise Error Rate (FWER) at level  $\alpha$ , irrespective of the underlying structure of dependence among the hypotheses.

- [Benjamini and Liu \(1999a\)](#) proposed a step down method with

$$s_i = \min \left( 1, \frac{n\alpha}{(n - i + 1)^2} \right) \quad i = 1, \dots, n$$

- [Benjamini and Liu \(1999b\)](#) studied another step down method with

$$s_i = 1 - \left[ \min \left( 1, \frac{n\alpha}{(n - i + 1)} \right) \right]^{\frac{1}{(n - i + 1)}} \quad i = 1, \dots, n$$

- [Benjamini and Liu \(1999b\)](#) also mentioned a step down method with

$$s_i = 1 - (1 - \alpha)^{\frac{1}{(n - i + 1)}} \quad i = 1, \dots, n$$

Among the step down procedures a result on optimality of Holm's procedure has been shown in [Gordon and Salzman \(2008\)](#). Step down algorithms reject a null hypothesis corresponding to a ordered p value if and only if every preceding p-values fall below the designated cut-offs. The step up algorithms are a little less restrictive in the sense that, they only look for the last p-value which lies below the cut-off. A step up multiple testing rule based on p values  $P_1, \dots, P_n$  also makes use of the ordered p values  $P_{(1)} \leq P_{(2)} \leq$

$\dots \leq P_{(n)}$  and a cut-off  $s = (s_1, \dots, s_n) \in S_n$ . Let,

$$k_1 = \max\{i \mid P_{(i)} \leq s_i\}$$

Then, a step up procedure rejects  $H_{(01)}, \dots, H_{(0k_1)}$ . It is easy to see that, using a fixed cut-off  $s_n \in S_n$ , if  $k_0$  and  $k_1$  are the number of rejections by the step down and step up algorithms respectively, then  $k_1 \geq k_0$  and hence step up procedures are always less conservative compared to its step down counterpart.

### Examples of step up procedures

- [Hochberg \(1988\)](#) proposed a step up algorithm for controlling FWER with

$$s_i = \frac{\alpha}{(n - i + 1)} \quad \forall i = 1, \dots, n$$

This procedure controls FWER at level  $\alpha$  when all hypotheses are independent or there is positive regression dependence between them.

- [Benjamini and Hochberg \(1995\)](#) proposed a groundbreaking FDR controlling algorithm which is a step up procedure with

$$s_i = \frac{i\alpha}{n} \quad \forall i = 1, \dots, n$$

- [Benjamini and Yekutieli \(2001\)](#) proposed a step up method with

$$s_i = \frac{i\alpha}{n \left( \sum_{j=1}^n \frac{1}{j} \right)} \quad \forall i = 1, \dots, n$$

They have proved that, this method controls FDR at level  $\alpha$  under any dependence structure among the hypotheses.

In addition, numerous potent multiple testing algorithms exist that do not fall under the categories of step-down or step-up. One notable example is Hommel's procedure ([Hommel \(1988\)](#)), which operates as follows.

1. Compute  $j = \max\{i \in \{1, \dots, n\} \mid P_{(n-i+k)} > \frac{k\alpha}{i} \text{ for } k = 1, \dots, i\}$

### 1.1.8 Dependence among Hypotheses

The presence of dependent observations presents a significant challenge in large-scale multiple testing scenarios, as traditional methods assume independence. This challenge is particularly pronounced in fields such as microarray experiments, where thousands of genes are analyzed to identify a small subset of significant ones. In addition to traditional familywise error rate (FWER) control, the analysis of false discovery rate (FDR), introduced by Benjamini and Hochberg in 1995, has become a widely adopted approach in these scenarios.

While the original FDR procedure was designed for independent p-values, subsequent research by [Benjamini and Yekutieli \(2001\)](#), [Sarkar \(2008\)](#) demonstrated its adaptability to certain dependency structures. However, when the proportion of true null hypotheses is small, these procedures may yield undesirable outcomes, often being overly conservative, leading to fewer discoveries.

### 1.1.9 Existing approaches towards dependence among hypotheses

While procedures such as Bonferroni's method and Holm's procedure effectively control the family-wise error rate (FWER) at the desired level across different dependence setups, the options for controlling the false discovery rate (FDR) are relatively limited. The adjustment proposed by [Benjamini and Yekutieli \(2001\)](#) is notable for its conservative nature, yet it maintains control over the FDR under various dependency structures. The control over FDR of Benjamini Hochberg algorithm is studied in [Benjamini and Yekutieli \(2001\)](#) and they introduced a notion of positive regression dependency setup (PRDS) on a subset  $I_0$ . A set  $D$  is called increasing if  $x \in D$  and  $y \geq x$ , implies  $y \in D$ . Then, the property PRDS is as follows

For any increasing set  $D$  and for any  $i \in I_0$ ,  $P(X \in D \mid X_i = x)$  is non-decreasing in  $x$ .

Benjamini and Yekutieli (2001) established that, the Benjamini Hochberg procedure controls the FDR at a level  $\frac{n_0}{n}\alpha$  if the joint distribution of the test statistics is PRDS on the subset of test statistics corresponding to true null hypotheses. An equivalent explanation of PRDS property is that, if  $\psi : [0, 1]^n \rightarrow \mathbb{R}^n$  is a co-ordinate wise non-decreasing function, then PRDS property holds if

$$E[\psi(p_1, \dots, p_n) \mid p_i = u] \text{ is non-decreasing function of } u \quad \forall i \in I_0$$

Sarkar (2007) proved the control over FDR of Benjamini Hochberg algorithm under a slightly relaxed criterion.

$$E[\psi(p_1, \dots, p_n) \mid p_i \leq u] \text{ is non-decreasing function of } u \quad \forall i \in I_0$$

for any co-ordinatewise non-decreasing function  $\psi(\cdot)$ .

There are many other researches available in which the dependence problem has been studied. Blanchard and Roquain (2009) proposed adaptive step-up methods for controlling the false discovery rate (FDR) under both positive and unspecified dependencies. Fan et al. (2012) introduced a novel procedure adjusted for dependence to approximate the false discovery proportion (FDP). The asymptotic FDR and number of type I errors of Benjamini-Hochberg (BH) procedure has been studied in Finner and Roters (2001), Finner and Roters (2002), Finner et al. (2007). Efron et al. (2001) introduced an empirical Bayes approach in multiple hypothesis testing based on local false discovery rate (lfdr). This approach has been extended under dependence by Sun and Tony Cai (2009) and they have claimed that their optimal statistic in terms of mFDR and mFNR is equivalent to lfdr under independence. The hidden markov model based approach of Sun and Tony Cai (2009) has been extended by Xie et al. (2011) under Gaussian model.

This thesis is mostly focused on the problem of dependence in various aspects of multiple hypothesis testing.

## 1.2 Motivation and brief overview of our work

There are very few literature which provides an explicit mathematical work on the performance of the existing multiple testing procedures under dependence. Recognizing this void, the primary objective of this thesis is to bridge this gap by conducting a comprehensive study encompassing various facets of the simultaneous testing problem.

Chapter 2 of this thesis is focused on the asymptotic FWER of Bonferroni's procedure. It has known to be a shortcoming of this procedure for being overly conservative especially under dependence. [Proschan and Shaw \(2011\)](#) has studied the asymptotic FWER of Bonferroni's procedure under the equicorrelated normal distribution when the correlation  $\rho_n$  approaches zero as  $n \rightarrow \infty$ . However, there is no literature which studies the traditional FWER as a function of the correlation ( $\rho$ ). If the FWER lies well below the level of significance  $\alpha$ , then one can easily conclude that Bonferroni's procedure indeed becomes conservative and the level of conservativeness can be seen from the gap between actual FWER and  $\alpha$ .

Chapter 3 focuses more on the FWER of the Bonferroni's method under general correlation matrix. The result of chapter 2 was later improved by [Dey and Bhandari \(2023\)](#). They showed that the FWER is asymptotically zero for equicorrelated normal distribution. Hence, there is a discontinuity in the FWER function near independence. We have examined how far the FWER retains its behaviour similar to independence and introduced a nearly independent setup where FWER of Bonferroni's procedure is still close to the one under independence.

Chapter 4 builds on the foundational work of [Sun and Tony Cai \(2009\)](#) and [Xie et al. \(2011\)](#), who established a one-to-one correspondence between the multiple testing problem and the weighted classification problem. Consequently, the Bayes rule used in weighted classification also proves to be optimal for controlling the mFDR and mFNR in a multiple testing problem. Despite its optimality, the resulting statistic is not symmetric and requires computation involving the joint distribution of the observations.

[Sun and Tony Cai \(2009\)](#) implemented this optimal procedure within the framework of

a Hidden Markov Model. The Markov chain assumption allowed them to simplify the statistic to some extent. However, simplifying this statistic further remains challenging, which limits the practical applicability of this optimal rule.

In the context of a Gaussian model, [Xie et al. \(2011\)](#) and [Xie et al. \(2016\)](#) demonstrated an asymptotic equivalence between this test statistic and the local false discovery rate (lfdr) introduced by [Efron \(2008\)](#). This equivalence, however, is contingent upon restrictive assumptions regarding the correlation matrix. In Chapter 4, we address these limitations by simplifying the test statistic for any correlation matrix  $\Sigma$ . Our approach significantly improves upon existing methods, including the Benjamini-Hochberg procedure and the marginal procedure proposed by [Xie et al. \(2011\)](#).

In Chapter 5, we address the challenge of estimating the proportion of false null hypotheses in a multiple testing scenario. Research by [Benjamini and Yekutieli \(2001\)](#) and [Sarkar \(2008\)](#) has demonstrated that the Benjamini-Hochberg (BH) procedure often controls the false discovery rate (FDR) at a level lower than desired. Specifically, if  $\alpha$  represents the desired significance level, then under the positive regression dependency setup (PRDS), the FDR of the BH procedure is bounded by  $\pi_0\alpha$ , where  $\pi_0$  denotes the proportion of null hypotheses. Consequently, accurate estimation of  $\pi_0$  could enhance the power of the BH procedure.

Several researchers have proposed various estimators for  $\pi_0$ , including [Storey \(2002\)](#), [Efron \(2012\)](#), [Benjamini and Hochberg \(2000\)](#), [Genovese and Wasserman \(2002\)](#), [Meinshausen and Bühlmann \(2005\)](#), [Langaas et al. \(2005\)](#), and [Patra and Sen \(2016\)](#). However, most of these estimators were developed for independent samples, and there is limited literature on their performance under dependent conditions. In this chapter, we explore the asymptotic behavior of the estimator proposed by [Benjamini and Hochberg \(2000\)](#) and examine its performance under various types of dependence commonly encountered in real-life applications. Additionally, we analyze the performance of the estimators proposed by [Storey \(2002\)](#) and [Patra and Sen \(2016\)](#) under different types of dependence.

# Chapter 2

## Asymptotic FWER for Correlated Normal distribution

### 2.1 Introduction

Correlation is a fundamental measure of dependence in the analysis of normal random variables. While it doesn't capture all aspects of dependence, it serves as a crucial starting point for examining how dependence among observations impact multiple testing algorithms. [Efron \(2010\)](#) demonstrated in his study of empirical Bayes methods that the correlation penalty hinges on the root mean square (RMS) of correlations, underscoring its significance. A comprehensive review of related literature can be found in [Efron \(2012\)](#), which provides further insights. [Sarkar \(2008\)](#) offers an extensive review on methods for controlling false discovery rate (FDR) under dependence. These contributions collectively emphasize the critical need for cautious handling of error rate criteria (such as family-wise error rate (FWER) or FDR) in the presence of correlation.

In this chapter, we focus on the equicorrelated normal distribution, which is particularly relevant for modeling the lifetimes of coherent systems, as discussed by [Loperfido et al. \(2007\)](#). Some distribution free bounds on FWER under this setup can be found in [Tong](#)

(2014) using Chebyshev-type inequalities as follows.

$$FWER(\rho) \leq \frac{\left[ \{(1 + (n - 1)\rho) (c_n^2 + 1 - (n - 1)(1 - \rho))\}^{\frac{1}{2}} + c_n(n - 1)(1 - \rho)^{\frac{1}{2}} \right]^2}{n \left[ c_n^2 + \left( \frac{1 + (n - 1)\rho}{n} \right) \right]^2}$$

Where  $FWER(\rho)$  is the FWER of Bonferroni's method with correlation among hypotheses  $\rho$  and  $c_n$  is the  $(1 - \frac{\alpha}{n})$ -th quantile of  $N(0,1)$  distribution. Since  $c_n^2 = O(\log n)$  the upper bound on the right hand side increases indefinitely  $n$  gets larger and larger. As a result, the gap between the actual FWER and the upper bound is significantly high for a choice of a particular distribution (e.g. normal).

In Section 2.3, we conduct simulations and generate plots illustrating the behavior of the Family-Wise Error Rate (FWER) as a function of  $\rho$ , ranging from scenarios with a small number to a large number of null hypotheses. Through these simulations, we aim to gain insights into how the FWER behaves under varying degrees of correlation. Section 2.4 highlights the asymptotic behaviour of FWER for correlated normal distribution. This is probably the first attempt in the context of finding most classically used FWER in terms of  $\rho$  asymptotically. It is also important to mention that the lifetime of parallel systems can be conveniently modelled by the maximum of exchangeable normal random vector. The asymptotic bound on FWER provided in this chapter is actually a lower bound on the cumulative distribution function (cdf) of the failure time of the parallel systems. Loperfido et al. (2007) has shown that the distribution of the maximum of  $n$  observations from exchangeable normal distribution follows  $(n - 1)$  dimensional skew normal distribution. While the cdf of multivariate skew normal distribution is very difficult to deal with, an asymptotic bound on the cdf can be obtained by a reasoning similar to the one in this paper.

## 2.2 Description of the setup

Let  $X_1, X_2, \dots$  be a sequence of observations and the null hypotheses are

$$H_{0i} : X_i \sim N(0, 1) \quad i = 1, 2, \dots$$

Here we have considered one sided tests (This means,  $H_{0i}$  is rejected for large values of  $X_i$  (say  $X_i > c$ )). FWER is the probability of falsely rejecting at least one null hypothesis (Which happens if  $X_i > c$  for some  $i$  and the probability is computed under the global null  $H_0 = \bigcap_{i=1}^n H_{0i} = \{X_i \sim N(0, 1) \quad \forall i = 1, 2, \dots, n\}$ ). Then,

$$\mathbf{FWER} = P(\text{At least one false rejection}) = P(X_i > c \text{ for some } i \mid H_0)$$

We assume,

$$\text{Corr}(X_i, X_j) = \rho \quad \forall i \neq j \quad (\rho \geq 0)$$

For Bonferroni's method,  $c = \Phi^{-1}(1 - \frac{\alpha}{n})$  (where  $\Phi(\cdot)$  is the cdf of standard normal distribution). It has been a common shortcoming of multiple testing rules that they become conservative under positive dependence (see [Storey \(2002\)](#), [Sarkar \(2007\)](#), [Efron \(2012\)](#)). We have studied the actual level of significance of Bonferroni's method which is much smaller than the desired level  $\alpha$  and hence this procedure becomes very conservative in the presence of non-negative correlation.

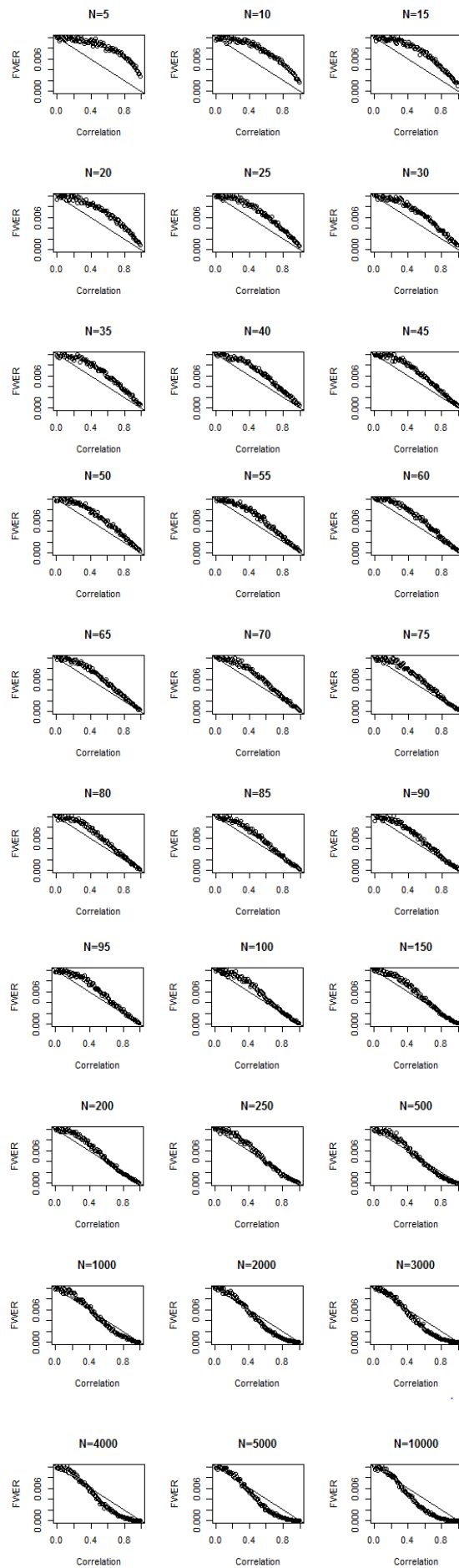
## 2.3 Simulation results

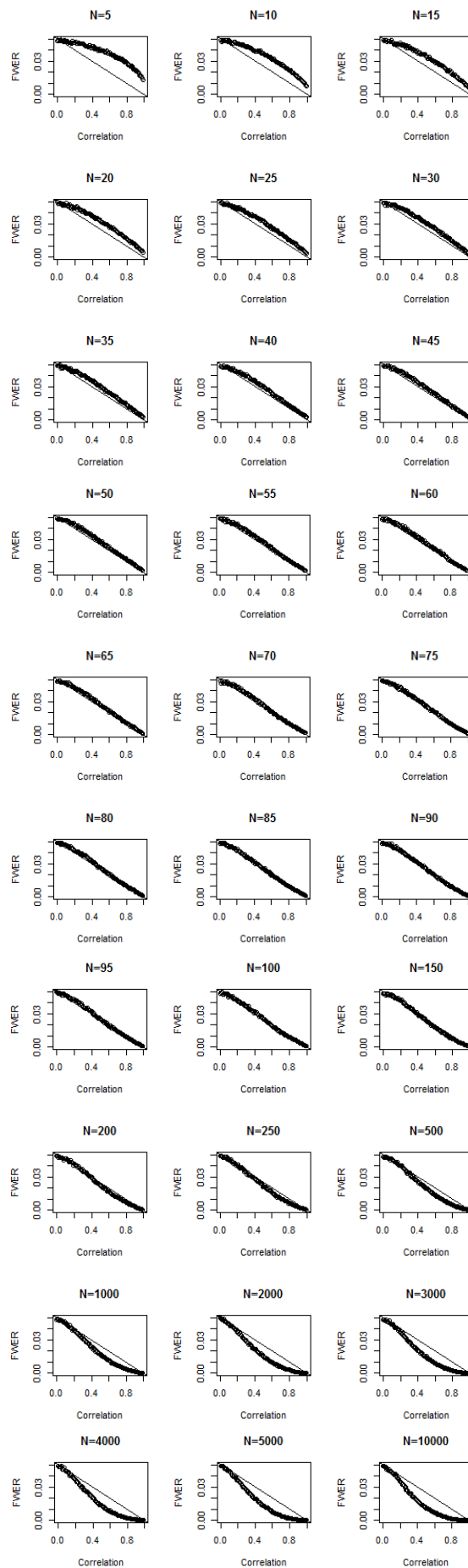
In our simulation study, we generated  $n$  observations from a  $N(0, 1)$  distribution with a constant correlation  $\rho$  between each pair of observations. Under the global null hypothesis, the Family-Wise Error Rate (FWER) is equivalent to the False Discovery Rate (FDR). Moreover, Bonferroni's method and Holm's procedure yield equivalent results for FWER. Consequently, examining Bonferroni's procedure provides valuable insights into the behavior of Holm's procedure. Various combinations of  $(n, \rho)$  were examined at significance levels

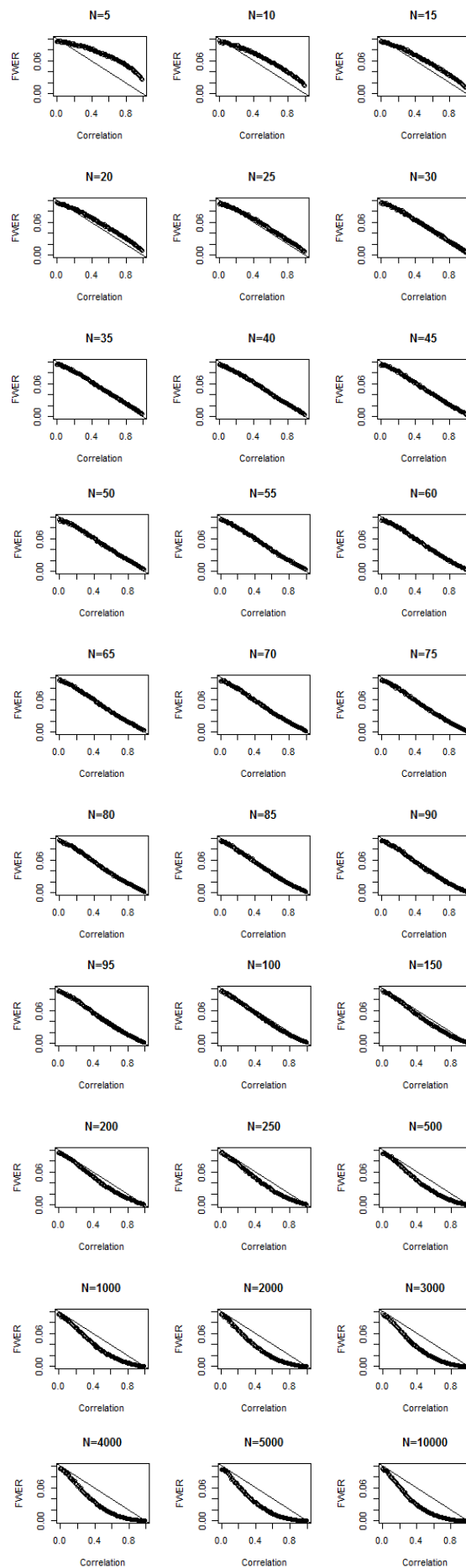
$\alpha = 0.1, 0.05,$  and  $0.01,$  which are commonly used in hypothesis testing.

We have estimated the actual FWER of the Bonferroni's procedure in the context of equicorrelated normal observations as a function of the correlation ( $\rho$ ). All FWER estimates are based on  $10^5$  replications at the significance level  $\alpha$ . Each plot includes a reference line  $L(\rho) = \alpha(1 - \rho)$ , facilitating comparison of how the FWER curves deviate from this line. Notably,  $FWER(0) \approx \alpha$  and  $FWER(1) \rightarrow 0$  as  $n \rightarrow \infty$ . Therefore, the FWER curve intersects the line  $L(\rho) = \alpha(1 - \rho)$  at  $\rho = 0$  and  $\rho = 1$  as  $n \rightarrow \infty$ .

The plots illustrating FWER as a function of correlation  $\rho$  are displayed below.

**Figure 2.1:** Plots of FWER for various values of  $N$  ( $\alpha = 0.01$ )

**Figure 2.2:** Plots of FWER for various values of  $N$  ( $\alpha = 0.05$ )

**Figure 2.3:** Plots of FWER for various values of  $N$  ( $\alpha = 0.1$ )

There is a consistent trend observed in the Family-Wise Error Rate (FWER) plots across all three  $\alpha$  values analyzed in our simulation study. Initially, for small sample sizes ( $n$ ), the FWER for Bonferroni's procedure exhibits a concave pattern. However, as the number of hypotheses increases, its behavior gradually transitions, shifting towards convexity. The rate of this transition from concavity to convexity likely varies depending on the specific value of  $\alpha$ . Observe that, in figure 2.3 (for  $\alpha = 0.1$ ) the nature changes very quickly while in figure 2.1 ( $\alpha = 0.01$ ) the convergence to convexity occurs slowly. The simulation results give us a hint that reference line  $L(\rho) = \alpha(1 - \rho)$  might be an asymptotic upper bound of  $FWER(\rho)$ . This claim is confirmed in corollary 2.4.1.1 of section 2.4.

## 2.4 Main theorem

Let,  $H(\rho) = 1 - FWER(\rho) = P(X_i \leq c, \forall i = 1, 2, \dots, n \mid H_0)$

**Theorem 2.4.1.** *Suppose each  $H_{0i}$  is being tested at size  $\alpha_n$ . If  $\lim_{n \rightarrow \infty} n\alpha_n = \alpha \in (0, 1)$  then, as  $n \rightarrow \infty$ ,  $H''(\rho) \leq 0$  and hence  $H(\rho)$  asymptotically is a concave function in  $[0, 1]$ .*

Note :-

1. For  $\rho = 0$  (Under independence), we must have,  $FWER = 1 - (1 - \alpha_n)^n \approx \alpha$ .
2. For  $\rho = 1$  (When  $X_i = X_j$  a.s.  $\forall i \neq j$ ), we must have  $FWER = \alpha_n$  (Because one rejection would imply rejection of all null hypotheses).

Suppose  $y = \mathcal{L}(\rho)$  denotes the line which joins  $(1, \alpha_n)$  and  $(0, 1 - (1 - \alpha_n)^n)$ . The following corollary describes the asymptotic behaviour of FWER as a function of  $\rho$ .

**Corollary 2.4.1.1.** *As  $n \rightarrow \infty$ ,  $FWER(\rho)$  is bounded above by the line  $\mathcal{L}(\rho)$ .*

In this section we are going to provide a proof of this theorem.

### 2.4.1 An alternate form of $H(\rho)$ and it's derivatives

Under the framework described above, we can say that under  $H_0$ , the sequence  $\{X_n\}_{n \geq 1}$  is exchangeable. (i.e.  $(X_{i_1}, \dots, X_{i_k}) \sim N_k(\mathbf{0}_k, (1-\rho)I_k + \rho J_k)$  (Where  $J_k$  is a  $k * k$  matrix of ones).

Then,  $X_k = \theta + Z_k, \forall k \geq 1$ .

Where  $\theta$  is a mean 0, normal random variable, independent of the sequence  $\{Z_n\}_{n \geq 1}$  and  $Z_i$ 's are i.i.d. normal random variables.

Since  $Cov(X_i, X_j) = \rho$ , this implies that  $Var(\theta) = \rho$

$\Rightarrow \theta \sim N(0, \rho)$  and  $Z_n \stackrel{i.i.d.}{\sim} N(0, 1-\rho) \forall n \geq 1$

Thus,

$$H(\rho) = P(\theta + Z_i \leq c \forall i = 1, 2, \dots, n) = E_\theta \left[ \Phi^n \left( \frac{c - \theta}{\sqrt{1-\rho}} \right) \right] = E \left[ \Phi^n \left( \frac{c + \sqrt{\rho}Z}{\sqrt{1-\rho}} \right) \right]$$

(Where  $Z \sim N(0, 1)$  and  $\Phi$  is the cdf of  $N(0,1)$  distribution)

If we define,  $d = \frac{c + \sqrt{\rho}Z}{\sqrt{1-\rho}}$ , then  $H(\rho) = E[\Phi^n(d)]$ .

Now, an application of dominated convergence theorem would yield,

$$H'(\rho) = E \left[ n \Phi^{n-1}(d) \phi(d) G(\rho, Z) \right] \quad (2.1)$$

(Where  $G(\rho, Z) = \frac{\partial d}{\partial \rho} = \frac{1}{2(1-\rho)^{\frac{3}{2}}} \left[ c + \frac{Z}{\sqrt{\rho}} \right]$  and  $\phi(\cdot)$  is the  $N(0,1)$  p.d.f. )

And again by D.C.T.,

$$H''(\rho) = E \left[ \frac{n}{2} \Phi^{n-2}(d) \phi(d) \left( a G^2(\rho, Z) + b G(\rho, Z) + \frac{c \Phi(d)}{4\rho(1-\rho)^{\frac{3}{2}}} \right) \right] \quad (2.2)$$

Where,

$$a = (n-1)\phi(d) - d\Phi(-d) \text{ and } b = \frac{(4\rho-1)\Phi(d)}{2\rho(1-\rho)}$$

Let's define,  $\alpha_1 = \Phi(-d)$ .

Observe that,

$$\begin{aligned} H''(\rho) &= E \left[ \frac{n}{2}(1 - \alpha_1)^{n-2} \phi(d) \left( aG^2(\rho, Z) + bG(\rho, Z) + \frac{c(1-\alpha_1)}{4\rho(1-\rho)^{\frac{3}{2}}} \right) \right] \\ &= \int_{-\infty}^{\infty} \frac{n}{2}(1 - \alpha_1)^{n-2} \phi(d) \left[ aG^2(\rho, z) + bG(\rho, z) + \frac{c(1-\alpha_1)}{4\rho(1-\rho)^{\frac{3}{2}}} \right] \phi(z) dz \end{aligned}$$

## 2.4.2 Proof of the main theorem

The proof of the main theorem involves three steps.

- **Step 1 :-** The 2nd and 3rd term in  $H''(\rho) \rightarrow 0$  as  $n \rightarrow \infty$ . (Proof is given in appendix ( lemma 2.4.2).
- **Step 2 :-** Suppose at  $z = z_0$ ,  $\alpha_1(z_0) = \Phi(-d(z_0)) = \frac{1}{n}$ . If  $G(\rho, z_0) \leq 0$  then the first term  $\rightarrow 0$  as  $n \rightarrow \infty$ .
- **Step 3 :-** If  $G(\rho, z_0) > 0$  then, the first term is asymptotically  $\leq 0$ .

Proof of step 1 is given in appendix (lemma 2.4.2). We shall proceed with the proof of step 2.

## 2.4.3 Proof of step 2

Consider the three disjoint regions  $\{\alpha_1 > \frac{6 \log n}{n}\}$ ,  $\{\frac{1}{n(\log n)^3} \leq \alpha_1 \leq \frac{6 \log n}{n}\}$  and  $\{\alpha_1 < \frac{1}{n(\log n)^3}\}$ . We'll show that the integrals in the first two regions  $\rightarrow 0$  as  $n \rightarrow \infty$  and the integral in the third region is asymptotically non-positive.

- **Case 1 :-**  $\{\alpha_1 > \frac{6 \log n}{n}\}$

If  $\alpha_1 > \frac{6 \log n}{n}$ , then  $(1 - \alpha_1)^{n-2} < (1 - \frac{6 \log n}{n})^{n-2}$ .

Using the fact that,  $(1 - \frac{x}{n})^n \leq e^{-x} \forall x > 0, \forall n \in \mathbb{N}$ , we can say that,  $(1 - \alpha_1)^{n-2} < \frac{1}{n^3}$ .

So,  $E \left[ \frac{n}{2}(1 - \alpha_1)^{n-2} \phi(d) aG^2(\rho, Z) I_{(\alpha_1 > \frac{6 \log n}{n})} \right] \leq \frac{1}{2} E \left[ \frac{aG^2(\rho, Z)}{n^2} I_{(\alpha_1 > \frac{6 \log n}{n})} \right]$

Each  $H_{0i}$  is being tested at size  $\alpha_n$  and we reject  $H_{0i}$  if  $X_i > c$ .

So,  $\alpha_n = \Phi(-c)$ .

As  $n \rightarrow \infty$ ,  $c \rightarrow \infty$  by the condition  $\lim_{n \rightarrow \infty} n\alpha_n = \alpha \in (0, 1)$ .

For large  $c$ , we have  $\Phi(-c) \sim \frac{\phi(c)}{c} = \frac{1}{\sqrt{2\pi}ce^{\frac{c^2}{2}}}$ .

So,  $n \sim \alpha\sqrt{2\pi}ce^{\frac{c^2}{2}} \dots\dots\dots(i)$

Now, observe that,  $a = (n - 1)\phi(d) - d\Phi(d) \leq n - 1 + |d| \leq n - 1 + \frac{c+\sqrt{\rho}|Z|}{\sqrt{1-\rho}}$

and  $G^2(\rho, Z) \leq \frac{c^2 + \frac{Z^2}{\rho}}{2(1-\rho)^3}$ .

$$\text{So, } E \left[ \frac{aG^2(\rho, Z)}{n^2} I_{(\alpha_1 > \frac{6 \log n}{n})} \right] \leq E \left[ \frac{\left( n-1 + \frac{c+\sqrt{\rho}|Z|}{\sqrt{1-\rho}} \right) \left( \frac{c^2 + \frac{Z^2}{\rho}}{2(1-\rho)^3} \right)}{n^2} \right] \rightarrow 0 \text{ as } n \rightarrow \infty.$$

• **Case 2** :-  $\{ \alpha_1 < \frac{1}{n(\log n)^3} \}$

For all  $d > 0$ , we must have  $\phi(d) \leq d\alpha_1$ .

This implies,  $|a| = |(n - 1)\phi(d) - d\Phi(d)| \leq n|d|\alpha_1 + |d|$ .

For large  $n$ , We should have  $n\alpha_1 < 1$  as  $n\alpha_1 < \frac{1}{(\log n)^3}$ .

This implies that, in this case,  $|a| \leq 2|d|$  for large  $n$ .

Using the fact  $\phi(d) \leq d\alpha_1$  we can deduce that,

$$E \left[ \frac{n}{2} (1 - \alpha_1)^{n-2} \phi(d) a G^2(\rho, Z) I_{\{ \alpha_1 < \frac{1}{n(\log n)^3} \}} \right] \leq E \left[ \frac{d^2 G^2(\rho, Z)}{(\log n)^3} I_{\{ \alpha_1 < \frac{1}{n(\log n)^3} \}} \right].$$

From the previous observation (in case 1) we can say that,

$$d^2 G^2(\rho, Z) \leq \frac{(c + \sqrt{\rho}|Z|)^2 (c^2 + \frac{Z^2}{\rho})}{2\rho(1 - \rho)^4}$$

**Notation** :-  $x_n = \Theta(y_n)$  if  $\exists c_1, c_2 > 0$  and  $M \in \mathbb{N}$  such that,  $c_1 y_n \leq x_n \leq c_2 y_n \quad \forall n \geq M$

(By (i)), we can say that,  $\log n = \Theta(c^2)$ .

This implies,

$$E \left[ \frac{d^2 G^2(\rho, Z)}{(\log n)^3} I_{(\alpha_1 < \frac{1}{n(\log n)^3})} \right] \rightarrow 0$$

Thus,

$$E \left[ \frac{n}{2} (1 - \alpha_1)^{n-2} \phi(d) aG^2(\rho, Z) I_{\{\alpha_1 < \frac{1}{n(\log n)^3}\}} \right] \rightarrow 0$$

- **Case 3** : -  $\{\alpha_1 \geq \frac{1}{n(\log n)^3}\}$ .

First we'll show that  $z_0$  takes a very large negative value. Observe that,

$$\Phi(-d(z_0)) = \frac{1}{n} \text{ and } \Phi(-c) = \alpha_n \sim \frac{\alpha}{n}$$

So,  $\frac{\Phi(-d(z_0))}{\Phi(-c)} \sim \frac{1}{\alpha}$ . We know that,  $x\Phi(-x) \sim \phi(x)$  for large enough  $x$ .

So, for large enough  $n$ ,  $d(z_0)\Phi(-d(z_0)) \sim \phi(-d(z_0))$  and  $c\Phi(-c) \sim \phi(c)$ .

Thus,

$$\frac{\Phi(-d(z_0))}{\Phi(-c)} \sim \frac{\frac{\phi(d(z_0))}{d(z_0)}}{\frac{\phi(c)}{c}} = \frac{ce^{\frac{c^2}{2}}}{d(z_0)e^{\frac{d^2(z_0)}{2}}} \sim \frac{1}{\alpha}$$

Since  $0 < \alpha < 1$ , we must have  $c > d(z_0)$ . Thus,  $\frac{c^2 - d^2(z_0)}{2} \leq \log(\frac{1}{\alpha})$

We can now conclude that,  $c - d(z_0) \leq \frac{\log(\frac{1}{\alpha})}{c + d(z_0)}$  for large enough  $n$ .

Since both  $c \rightarrow \infty$  and  $d(z_0) \rightarrow \infty$  as  $n \rightarrow \infty$ , this implies that  $c - d(z_0) \rightarrow 0$  as  $n \rightarrow \infty$ .

From this, it is easy to see that,  $z_0 + cT(\rho) \rightarrow 0$  as  $n \rightarrow \infty$ . (Where  $T(\rho) = \frac{1}{1 + \sqrt{1 - \rho}}$  is entirely dependent on  $\rho$  and is a positive function for all  $0 \leq \rho \leq 1$ )

This means,  $z_0$  takes a very large negative value.

First consider the region  $\{\alpha_1 > \frac{1}{n}\}$ . Since  $\alpha_1$  is a decreasing function of  $z$ , this region can also be written as  $\{z < z_0\}$

Note that,  $dG^2(\rho, Z) = p(Z, c)$  (A polynomial in  $Z$  and  $c$ ). Using this, it is easy to see that,  $\sup_{z < z_0} dG^2(\rho, z)\phi(z) \rightarrow 0$  as  $n \rightarrow \infty$ .

$$d = \frac{c + \sqrt{\rho}Z}{\sqrt{1 - \rho}} \xrightarrow{a.s.} \infty \text{ as } n \rightarrow \infty \text{ (Because } c \xrightarrow{a.s.} \infty \text{ as } n \rightarrow \infty \text{)}.$$

So,  $a = (n - 1)\phi(d) - d\Phi(d) \sim d(n\alpha_1 - 1)$ .

By lemma 2.4.3 in appendix, we can say that,

$$E \left[ \frac{n}{2} (1 - \alpha_1)^{n-2} \phi(d) a G^2(\rho, Z) \right] \sim E \left[ \frac{n}{2} (1 - \alpha_1)^{n-2} \phi(d) d (n\alpha_1 - 1) G^2(\rho, Z) \right].$$

Now observe that,

$$\begin{aligned} & E \left[ \frac{n}{2} (1 - \alpha_1)^{n-2} \phi(d) d (n\alpha_1 - 1) G^2(\rho, Z) \right] \\ &= \int_{-\infty}^{\infty} n (n\alpha_1 - 1) (1 - \alpha_1)^{n-2} \phi(d) d G^2(\rho, z) \phi(z) dz \\ &\propto \int_0^1 f_n(\alpha_1) K(\alpha_1) d\alpha_1 \end{aligned}$$

(Where  $f_n(\alpha_1) = n(n\alpha_1 - 1)(1 - \alpha_1)^{n-2}$  and  $K(\alpha_1) = dG^2(\rho, z)\phi(z)$  )

Observe that,  $f_n(\alpha_1) > (<) 0 \Leftrightarrow \alpha_1 > (<) \frac{1}{n}$  and  $\int_0^1 f_n(\alpha_1) d\alpha_1 = 0$ .

And,  $\int_{\frac{1}{n}}^1 f_n(\alpha_1) d\alpha_1 = \int_1^n (k-1) \left(1 - \frac{k}{n}\right)^{n-2} dk \leq \int_1^\infty (x-1) e^{-\frac{x}{2}} dx < \infty$

$\Rightarrow \int_0^1 |f_n(\alpha_1)| d\alpha_1$  is bounded.

Since

$$\sup_{\alpha_1 > \frac{1}{n}} K(\alpha_1) = \sup_{z < z_0} dG^2(\rho, z)\phi(z) \rightarrow 0$$

It is easy to deduce that,

$$E \left[ n (1 - \alpha_1)^{n-2} \phi(d) d (n\alpha_1 - 1) G^2(\rho, Z) I_{(\alpha_1 > \frac{1}{n})} \right] = \int_{\frac{1}{n}}^1 f_n(\alpha_1) K(\alpha_1) d\alpha_1 \rightarrow 0$$

Now we need to consider the region  $\left\{ \frac{1}{n(\log n)^3} \leq \alpha_1 \leq \frac{1}{n} \right\}$ .

$\Phi(-d(z_1)) = \frac{1}{n(\log n)^3}$  and  $z_1 = z_0 + \delta_n$ .

$$\Rightarrow \frac{\Phi(-d(z_0))}{\Phi(-d(z_1))} = (\log n)^3$$

For large  $n$ ,  $d(z_i)\Phi(-d(z_i)) \sim \phi(-d(z_i))$  for  $i = 0, 1$ .

$$\Rightarrow \frac{d(z_1) e^{-\frac{d^2(z_1)}{2}}}{d(z_0) e^{-\frac{d^2(z_0)}{2}}} \sim (\log n)^3$$

$$\Rightarrow \log\left(\frac{d(z_1)}{d(z_0)}\right) + \frac{d^2(z_1) - d^2(z_0)}{2} \sim 3 \log(\log n)$$

Since  $d(z_1) > d(z_0)$ , we can conclude that,  $\frac{(d(z_0) + d(z_1))(d(z_1) - d(z_0))}{2} \leq 3 \log(\log n)$

$$\Rightarrow d(z_1) - d(z_0) = \sqrt{\frac{\rho}{1-\rho}} \delta_n \leq \frac{6 \log(\log n)}{d(z_1)+d(z_0)}$$

Since  $d(z_1) > d(z_0) = -\Phi^{-1}(\frac{1}{n}) = \Theta(\sqrt{\log n})$  this implies,  $\frac{6 \log(\log n)}{d(z_1)+d(z_0)} \rightarrow 0$  and hence  $\delta_n \rightarrow 0$  as  $n \rightarrow \infty$ .

$$\begin{aligned} \text{So, } & E \left[ \frac{n}{2} (1 - \alpha_1)^{n-2} \phi(d) d (n\alpha_1 - 1) G^2(\rho, Z) I_{(\frac{1}{n(\log n)^3} \leq \alpha_1 \leq \frac{1}{n})} \right] \\ & \leq \int_{\frac{1}{n(\log n)^3}}^{\frac{1}{n}} n(1 - \alpha_1)^{n-2} K(\alpha_1) d\alpha_1 \end{aligned}$$

Note that,  $|K(\alpha_1)| \leq \max_{z \in [z_0, z_0 + \delta_n]} p(z, c) \phi(z)$  ( as  $K(\alpha_1) = p(z, c) \phi(z)$  )

Since  $\delta_n \rightarrow 0$  and  $p(z_0, c) \phi(z_0) \rightarrow 0$  as  $n \rightarrow \infty$ , by continuity of  $p(z, c) \phi(z)$  we can say that,  $\max_{z \in [z_0, z_0 + \delta_n]} p(z, c) \phi(z) \rightarrow 0$  and hence,  $|K(\alpha_1)| \rightarrow 0$  as  $n \rightarrow \infty$ .

$$\int_{\frac{1}{n(\log n)^3}}^{\frac{1}{n}} n(1 - \alpha_1)^{n-2} d\alpha_1 = \int_{\frac{1}{(\log n)^3}}^1 (1 - \frac{k}{n})^{n-2} dk \leq \int_0^1 e^{-\frac{k}{2}} dk \leq 1.$$

$$\text{So, } \int_{\frac{1}{n(\log n)^3}}^{\frac{1}{n}} n(1 - \alpha_1)^{n-2} K(\alpha_1) d\alpha_1 \rightarrow 0$$

$$\text{And thus, } E \left[ \frac{n}{2} (1 - \alpha_1)^{n-2} \phi(d) d (n\alpha_1 - 1) G^2(\rho, Z) I_{(\frac{1}{n(\log n)^3} \leq \alpha_1 \leq \frac{1}{n})} \right] \rightarrow 0.$$

This completes the proof of step 2.

### 2.4.4 Proof of step 3

As noted earlier,  $f_n(\alpha_1) > (<)0 \Leftrightarrow \alpha_1 > (<)\frac{1}{n}$  and  $\int_0^1 f_n(\alpha_1) d\alpha_1 = 0$ .

An argument similar to the one given in the proof of step 2, we can say that the contribution in the integrals  $\int_0^1 f_n(\alpha_1) K(\alpha_1) d\alpha_1$  and  $\int_0^1 f_n(\alpha_1) d\alpha_1$  outside the region  $\{\frac{1}{n(\log n)^3} \leq \alpha_1 \leq \frac{6 \log n}{n}\}$  is negligible as  $n \rightarrow \infty$ .

Since  $\Phi(-d(z_0)) = \frac{1}{n} < \frac{1}{2}$  for large enough  $n$ , this means,  $d(z_0) > 0$  for large  $n$ .

Also, observe that,  $\alpha_1$  is a decreasing function of  $z$ . Since  $G(\rho, z_0) > 0$ , we can say that,  $K(\alpha_1(z)) > K(\alpha_1(z')) \forall z' < z_0$  and  $z > z_0$  in the region  $\{\frac{1}{n(\log n)^3} \leq \alpha_1 \leq \frac{6 \log n}{n}\}$ .

This means,  $\int_0^{\frac{1}{n}} |f_n(\alpha_1)| K(\alpha_1) d\alpha_1 > \int_{\frac{1}{n}}^1 |f_n(\alpha_1)| K(\alpha_1) d\alpha_1$ .

Since  $f_n(\alpha_1) < 0$ ,  $\forall \alpha_1 \in (0, \frac{1}{n})$ , this means  $\int_0^1 f_n(\alpha_1) K(\alpha_1) d\alpha_1 \leq 0$

And thus,  $E\left[\frac{n}{2}(1 - \alpha_1)^{n-2} \phi(d) d(n\alpha_1 - 1) G^2(\rho, Z)\right] \leq 0$  which gives us the desired result.

This completes the proof of main theorem.

## 2.4.5 Appendix

**Lemma 2.4.2.** *The second and third term in  $H''(\rho) \rightarrow 0$  as  $n \rightarrow \infty$ .*

**Proof** :- We shall do it only for the third term. The other one follows similarly. This proof is similar to the proof of case 2 ( $G(\rho, z_0) < 0$ ) of step 2 of theorem 2.4.1.

Third term in  $H''(\rho)$  is proportional to  $E[nc(1 - \alpha_1)^{n-1} \phi(d)] = \int_{-\infty}^{\infty} nc(1 - \alpha_1)^{n-1} \phi(d) \phi(z) dz$ .

Let,  $\phi(z) = M(\alpha_1)$ .

Then,  $\int_{-\infty}^{\infty} nc(1 - \alpha_1)^{n-1} \phi(d) \phi(z) dz = c \int_0^1 (1 - \frac{k}{n})^{n-1} M(\frac{k}{n}) dk$

Since  $\phi(d) \leq d^3 \alpha_1$  for large enough  $d$ , so

$$E\left[nc(1 - \alpha_1)^{n-1} \phi(d) I_{(\alpha_1 < \frac{1}{n(\log n)^3})}\right] \rightarrow 0$$

Also, if  $\alpha_1 > \frac{6 \log n}{n}$  then  $(1 - \alpha_1)^{n-2} \leq \frac{1}{n^3}$  and hence

$$E\left[nc(1 - \alpha_1)^{n-1} \phi(d) I_{(\alpha_1 > \frac{6 \log n}{n})}\right] \rightarrow 0$$

Let,  $\Phi(-d(z_0 - \delta'_n)) = \frac{1}{n(\log n)^3}$  and  $\Phi(-d(z_0 + \delta''_n)) = \frac{6 \log n}{n}$ .

Similar to the case 2 ( $G(\rho, z_0) < 0$ ) of step 2 of theorem 2.4.1 we can say that,  $\max\{\delta'_n, \delta''_n\} \rightarrow 0$  as  $n \rightarrow \infty$ .

This implies that, if  $\frac{1}{n(\log n)^3} < \alpha_1 < \frac{6 \log n}{n}$ , then

$$cM(\alpha_1) \leq \max_{z \in [z_0 - \delta'_n, z_0 + \delta''_n]} c\phi(z) \rightarrow 0 \text{ as } n \rightarrow \infty$$

Thus,  $E \left[ nc(1 - \alpha_1)^{n-1} \phi(d) I_{\left(\frac{1}{n(\log n)^3} < \alpha_1 < \frac{6 \log n}{n}\right)} \right] \rightarrow 0$  and the third term in  $H''(\rho) \rightarrow 0$ .

**Lemma 2.4.3.**  $E \left[ \frac{n}{2} (1 - \alpha_1)^{n-2} \phi(d) G^2(\rho, Z) |a - d(n\alpha_1 - 1)| \right] \rightarrow 0$  as  $n \rightarrow \infty$ .

**Proof** :-  $|a - d(n\alpha_1 - 1)| = (n - 1) |\phi(d) - d\Phi(d)|$ .

If  $d \leq 0$ , then  $\alpha_1 \geq \frac{1}{2}$  and hence,  $(1 - \alpha_1)^{n-2} \leq \left(\frac{1}{2}\right)^{n-2}$ .

Since  $n = \Theta\left(ce^{\frac{c^2}{2}}\right)$ , this immediately implies that,

$$E \left[ \frac{n}{2} (1 - \alpha_1)^{n-2} \phi(d) G^2(\rho, Z) |a - d(n\alpha_1 - 1)| I_{(d \leq 0)} \right] \rightarrow 0$$

When  $d > 0$ ,  $|\phi(d) - d\Phi(d)| \leq K\alpha_1$  for a constant  $K > 0$ .

This implies that,

$$E \left[ \frac{n}{2} (1 - \alpha_1)^{n-2} \phi(d) G^2(\rho, Z) |a - d(n\alpha_1 - 1)| I_{(d > 0)} \right] \leq KE \left[ n^2 \alpha_1 (1 - \alpha_1)^{n-2} \phi(d) G^2(\rho, Z) I_{(d > 0)} \right]$$

An idea similar to the proof of lemma 2.4.2 will tell us that, we need to consider the region  $\left\{ \frac{1}{n(\log n)^3} < \alpha_1 < \frac{8 \log n}{n} \right\}$ . The integral corresponding to this region  $\rightarrow 0$  follows exactly in the similar manner.

## 2.5 Simulation results

Corollary 2.4.1.1 tells us for large  $n$ , FWER for Bonferroni's method (with level of significance  $\alpha$ ) is asymptotically bounded above by  $\alpha(1 - \rho)$ . In order to verify this result empirically, some simulation results have been provided in table 2.1. In our simulation experiments, we have considered  $\rho = 0, 0.1, 0.3, 0.5, 0.7, 0.9$  and  $\alpha = 0.01, 0.05, 0.1, 0.4, 0.6, 0.7$ . In each combination of  $(\rho, \alpha)$ , 10000 replications have

been made to estimate the FWER (the estimate obtained is denoted by  $\widehat{FWER}$ ). In each replication, we have generated 10000 equicorrelated normal random variables each with mean 0 and variance 1. Bonferroni's method suggests us to reject  $H_{0i}$  at level  $\alpha$  if  $Z_i > (1 - \frac{\alpha}{10000})$ -th quantile of  $N(0,1)$  distribution. In each replication we have to note whether or not any of the 10000  $Z_i$ 's exceeds that cut-off and then  $\widehat{FWER}$  is obtained accordingly from the 10000 replications.

Each  $\widehat{FWER}$  obtained at the combination  $(\rho, \alpha)$  is compared with  $\alpha(1 - \rho)$ (the upper bound mentioned in 2.4.1.1). It is impressive that in all the cases  $\widehat{FWER}$  is substantially smaller than  $\alpha(1 - \rho)$  (except at  $(\rho, \alpha) = (0.1, 0.01)$ ), although the difference is not noteworthy). All these observations suggest that in positively correlated setup, Bonferroni's method actually controls the FWER at a much smaller level than the desired level of significance which makes this method more conservative.

**Table 2.1:** Simulation results

$\rho$	$\alpha$	0.01	0.05	0.1	0.4	0.6	0.7
0.9	$\widehat{FWER}$	9.00E-05	0.00046	0.00053	0.00221	0.00324	0.0031
	$\alpha(1 - \rho)$	1.00E-03	0.005	0.01	0.04	0.06	0.07
0.7	$\widehat{FWER}$	0.00101	0.00363	0.00588	0.01617	0.02149	0.023
	$\alpha(1 - \rho)$	0.003	0.015	0.03	0.12	0.18	0.21
0.5	$\widehat{FWER}$	0.00347	0.01156	0.01918	0.04909	0.06414	0.07042
	$\alpha(1 - \rho)$	0.005	0.025	0.05	0.2	0.3	0.35
0.3	$\widehat{FWER}$	0.00683	0.02523	0.04363	0.11495	0.15013	0.16494
	$\alpha(1 - \rho)$	0.007	0.035	0.07	0.28	0.42	0.49
0.1	$\widehat{FWER}$	0.00996	0.04367	0.07978	0.23801	0.31105	0.34295
	$\alpha(1 - \rho)$	0.009	0.045	0.09	0.36	0.54	0.63
0	$\widehat{FWER}$	0.01018	0.0486	0.09424	0.32914	0.45065	0.50499
	$\alpha(1 - \rho)$	0.01	0.05	0.1	0.4	0.6	0.7

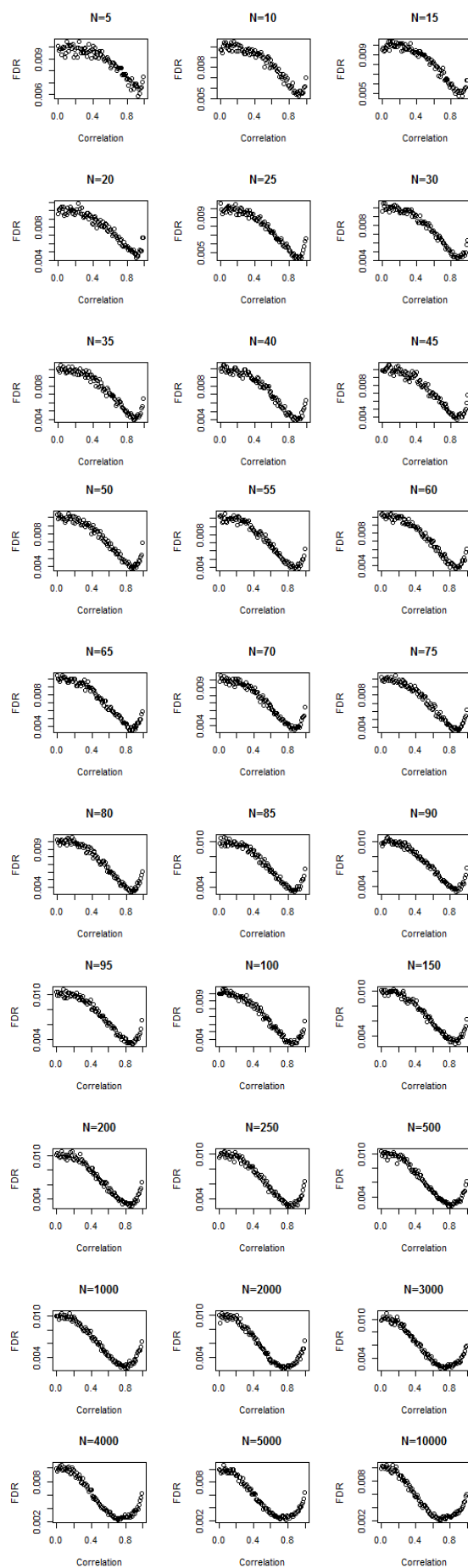
## 2.6 FDR of BH procedure

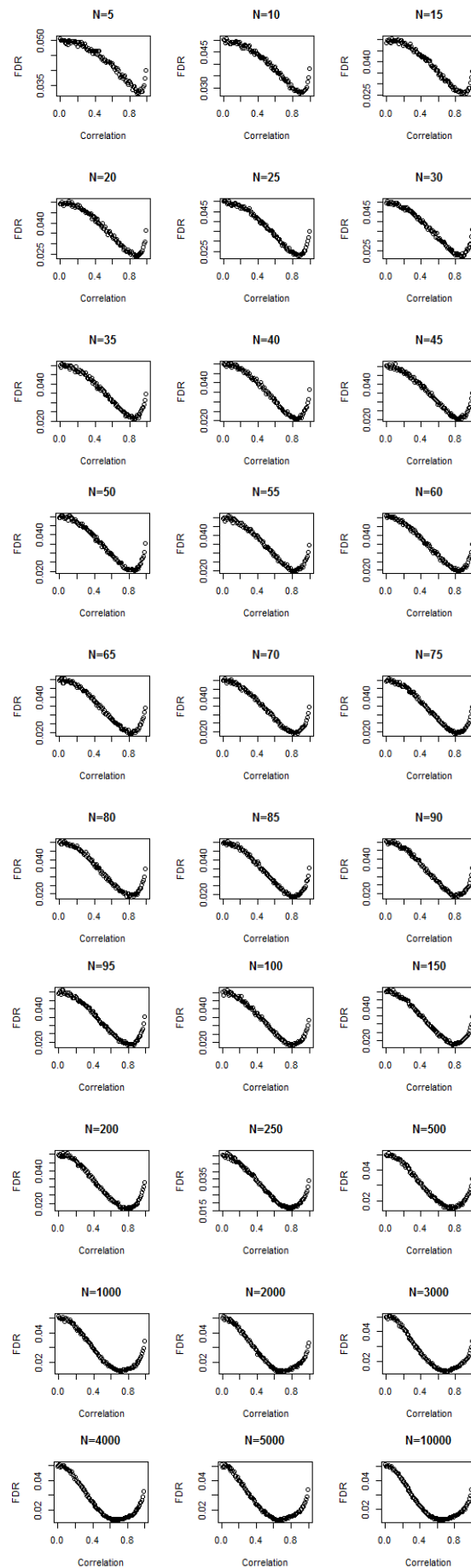
Motivated by the finding of corollary 2.4.1.1, our exploration now shifts towards the asymptotic behavior of the False Discovery Rate (FDR) concerning the Benjamini-Hochberg procedure. Herein lies our inquiry: does a simple bound akin to  $\alpha(1 - \rho)$  effectively capture the intricate dynamics of FDR under this methodology? [Finner and Roters \(2001\)](#) demonstrated that the false discovery rate (FDR) of the Benjamini-Hochberg (BH) procedure asymptotically approaches a non-zero value. However, a closed-form expression for this value as a function of  $\rho$  is not available. More recently, [Dey \(2022\)](#), [Dey and Bhandari \(2024\)](#) have further investigated the asymptotic behavior of the FDR for the BH procedure, providing additional insights into this phenomenon. They have shown that,

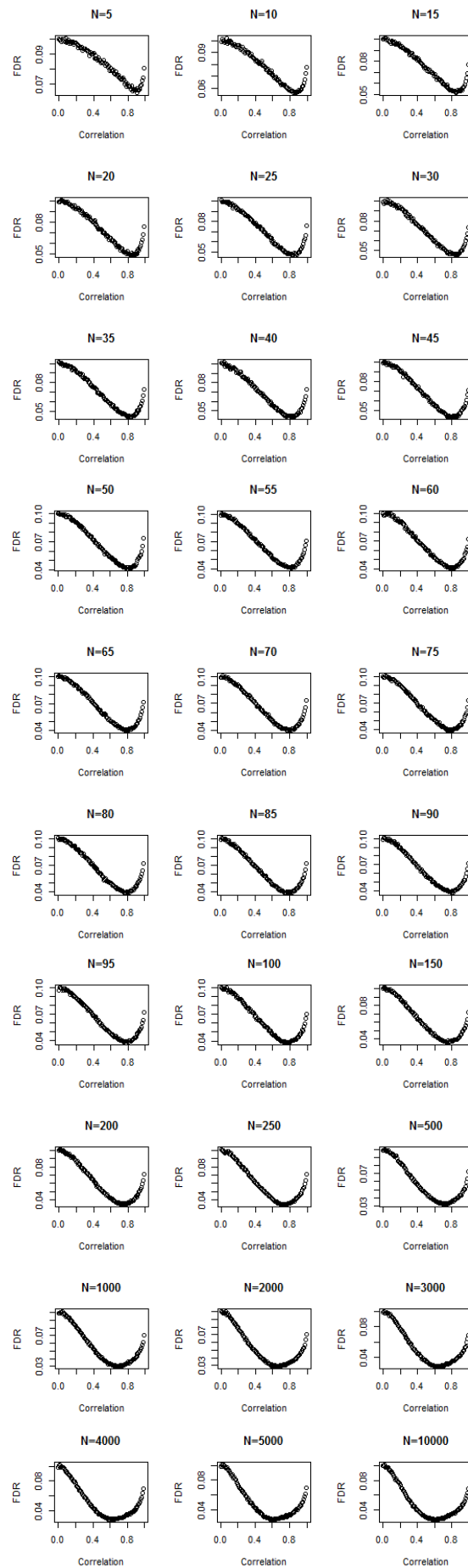
$$\lim_{n \rightarrow \infty} FDR_{BH}(n, \alpha, \rho) = 1 - \Phi \left[ \inf_{t \in (0,1)} \frac{\Phi^{-1}(1 - t\alpha) - \sqrt{\rho} \Phi^{-1}(1 - t)}{\sqrt{\rho}} \right] > 0$$

(Where  $FDR_{BH}(n, \alpha, \rho)$  denotes the FDR of BH procedure at level  $\alpha$  under an equicorrelated normal setup with correlation  $\rho$ )

Although their asymptotic limit appears to be simpler, it does not address whether the limiting FDR is below the threshold given by the line  $\mathcal{L}(\rho) = \alpha(1 - \rho)$  or not. The following graphs exhibit the asymptotic behaviour of FDR for Benjamini Hochberg method. In this study, we have again generated  $n$  observations from  $N(0, 1)$  distribution in such a way that there is a fixed correlation ( $\rho$ ) between any two observations. Different combinations of  $(n, \rho)$  are considered and they are studied at  $\alpha = 0.1, 0.05, 0.01$ .

**Figure 2.4:** Plots of FDR of BH procedure for various values of  $N$  ( $\alpha = 0.01$ )

**Figure 2.5:** Plots of FDR of BH procedure for various values of  $N$  ( $\alpha = 0.05$ )

**Figure 2.6:** Plots of FDR of BH procedure for various values of  $N$  ( $\alpha = 0.1$ )

It is clear from the plots that this procedure is not as conservative as the FWER controlling procedures. However, the FDR tends to decrease at first and then increase as  $\rho$  increases. But in all the cases, the FDR's are well below the desired level of significance ( $\alpha$ ). It is a more challenging task to provide an asymptotic upper bound on the FDR of the Benjamini-Hodgeberg procedure which enables us to provide a correlation correction in this setup. Much research is needed in order to develop a stronger upper bound on these type-I error rates under general correlation structure.

# Chapter 3

## FWER under nearly independent setup

### 3.1 Introduction

The preceding chapter delved into the control of Family-Wise Error Rate (FWER) using Bonferroni's method within the framework of equicorrelated normal distributions. Corollary 2.4.1.1 provided a crucial insight, demonstrating that as the number of hypotheses ( $n$ ) grows larger and larger, the FWER of Bonferroni's method becomes bounded above by  $\alpha(1 - \rho)$ , where  $\alpha$  represents the significance level and  $\rho$  signifies the correlation coefficient. Moreover, recent advancements by [Dey and Bhandari \(2023\)](#) have enriched our understanding of this phenomenon. They have showcased that under the assumptions of a global null hypothesis and equicorrelated normal distribution, the FWER of Bonferroni's method asymptotically approaches zero. Intriguingly, their work also reveals a captivating nuance: while the FWER remains bounded away from zero in the absence of correlation, it tends towards zero asymptotically for any non-zero correlation coefficient. This observation hints at a fascinating discontinuity within the behavior of the FWER function, particularly at  $\rho = 0$ . Consequently, it beckons researchers to explore sequences of correlation coefficients  $(\rho_n)$  that converge towards zero as  $n \rightarrow \infty$ . Such investigations aim to elucidate the rate of convergence towards zero at which the asymptotic FWER of Bonferroni's

method continues to closely resemble its behavior under independence.

While the assumption of equal correlation may be suitable for an initial exploratory analysis of FWER behavior, real-life studies often deviate from this premise. In this chapter, our objective is to address the question of to what extent the FWER of Bonferroni's method remains consistent with that under independence. Additionally, we have extended our analysis to include jointly normal distributions, focusing on approximating the FWER of Bonferroni's method within a "nearly independent" framework. Furthermore, we have introduced an asymptotic correction factor to refine the FWER of Bonferroni's method within this setup.

This chapter is organized as follows: in section 3.2, we introduce the framework of our work. In section 3.3, we have shown the motivation of considering the "nearly independent" setup and in section 3.4 we have approximated the FWER under this setup and provided a correction factor which accounts for the deviation from independence.

## 3.2 Description of the problem

Let  $X_1, X_2, \dots$  be a sequence of observations and the null hypotheses are

$$H_{0i} : X_i \sim N(0, 1) \quad i = 1, 2, \dots$$

Here we have again considered one sided tests (This means,  $H_{0i}$  is rejected for large values of  $X_i$  (say  $X_i > c$ )). Again, we consider FWER under the intersection null hypothesis  $H_0 = \bigcap_{i=1}^n H_{0i} = \{X_i \sim N(0, 1) \quad \forall i = 1, 2, \dots, n\}$ . Then,

$$\mathbf{FWER} = \text{P(At least one false rejection)} = P\left(\bigcup_{i=1}^n \{X_i > c\} \mid H_0\right)$$

Suppose,  $\text{Corr}(X_i, X_j) = \rho_{ij} \quad \forall i \neq j$  and  $\Sigma$  is the corresponding correlation matrix.

We assume the following about the correlation matrix  $\Sigma$ .

$$\rho_{ij} = O\left(\frac{1}{n^\beta}\right) \quad \forall i \neq j \text{ for some } \beta > 0$$

For convenience, we shall call this setup as “**nearly independent setup**”.

Suppose  $P(X_i > c \mid H_{oi}) = \alpha_n$ . Then, under the assumption of independence of hypotheses (i.e. if  $\Sigma$  is identity matrix), we have  $FWER = 1 - (1 - \alpha_n)^n$ . We shall study the value of the FWER under nearly independent setup. Bonferroni’s method is based on  $\alpha_n = \frac{\alpha}{n}$ . So, we shall consider the setup where  $\lim_{n \rightarrow \infty} n\alpha_n = \alpha$ .

### 3.3 Motivation behind the nearly independent setup

Efron (2007) in his study of empirical Bayes methods, has pointed out that the correlation penalty on the summary statistics based on empirical cdf depends on the root mean square (RMS) of correlations ( $\frac{1}{n(n-1)} \sum_{i \neq j} \rho_{ij}^2$ ). A natural question might arise from this result is that “does the RMS of correlations act as a correlation penalty” in our framework also? If this were true, then we would have

$$FWER \sim FWER \text{ under independence as } \frac{1}{n(n-1)} \sum_{i \neq j} \rho_{ij}^2 \rightarrow 0$$

We can construct an easy counterexample to answer this question.

Let  $M_n(\rho)$  denote the  $n \times n$  equicorrelation matrix with correlation  $\rho$  and  $\mathbf{0}_n$  denote the  $n \times n$  matrix with all zero entries. Consider the  $n^2 \times n^2$  block-diagonal matrix with  $n$  many  $M_n(\rho)$  blocks in the diagonal.

$$\text{i.e. } \Sigma = \begin{pmatrix} M_n(\rho) & \mathbf{0}_n & \dots & \mathbf{0}_n \\ \mathbf{0}_n & M_n(\rho) & \dots & \mathbf{0}_n \\ \dots & \dots & \dots & \dots \\ \mathbf{0}_n & \mathbf{0}_n & \dots & M_n(\rho) \end{pmatrix}$$

This is the correlation matrix of the  $n^2$  variables  $(X_1, \dots, X_{n^2})$  such that,

- $(X_1, \dots, X_n), (X_{n+1}, \dots, X_{2n}), \dots, (X_{n^2-n+1}, \dots, X_{n^2})$  are independent.
- In each block  $(X_{(k-1)n+1}, \dots, X_{kn})$  we have  $Corr(X_i, X_j) = \rho$  for  $i \neq j$ . (for  $k = 1, \dots, n$ )

If we apply Bonferroni ( $\alpha$ ) method in this setup then  $\alpha_n = \frac{\alpha}{n^2}$ . By the similar approach of [Das and Bhandari \(2021\)](#), we can argue that,

$$\begin{aligned} & 1 - FWER \\ & \geq \left( \left(1 - \frac{\alpha}{n^2}\right) - (1 - \rho) \left[ \left(1 - \frac{\alpha}{n^2}\right) - \left(1 - \frac{\alpha}{n^2}\right)^n \right] \right)^n \\ & = \left(1 - \frac{\alpha}{n^2}\right)^n \left[ 1 - (1 - \rho) \left[ 1 - \left(1 - \frac{\alpha}{n^2}\right)^{n-1} \right] \right]^n \end{aligned}$$

Since  $1 - (1 - x)^k \leq kx \quad \forall 0 < x < 1$ , this implies,

$$1 - FWER \geq \left(1 - \frac{\alpha}{n^2}\right)^n \left[ 1 - \frac{(1 - \rho)(n - 1)\alpha}{n^2} \right]^n \rightarrow e^{-\alpha(1 - \rho)} \text{ as } n \rightarrow \infty$$

Since  $1 - e^{-x} \leq x$  for all positive  $x$ , we finally have  $FWER \leq \alpha(1 - \rho)$  asymptotically under this setup. It is interesting to note that,

- In this setup, mean of the absolute values of correlations is  $\frac{n \binom{n}{2} \rho}{\binom{n^2}{2}} = O\left(\frac{1}{n}\right)$  and so is the mean square of correlations. Although the root mean square (RMS) of correlations becomes smaller and smaller, the FWER is not close to the FWER under independence. So, the correlation penalty is neither dependent on RMS of correlations nor the mean of their absolute values.
- It is also interesting to note that, whenever  $\max_{i \neq j} |\rho_{ij}|$  is bounded away from 0, it is possible to find a setup where FWER widely differs from the one obtained under independence.

So, if we ask ourselves “how far can we go from independence without losing much?”, then the above example suggests us to bound the absolute values of the correlations. Under the nearly independent setup, we bound the absolute values by the order  $\frac{1}{n^\beta}$  for some  $\beta > 0$ .

### 3.4 FWER under the nearly independent setup

Let's denote FWER under correlation matrix  $\Sigma$  by  $FWER(\Sigma)$ .

**Theorem 3.4.1.** Fix a sufficiently large integer  $K$ . Then, under the nearly independent setup, FWER can be approximated by the following way.

$$FWER(\Sigma) \sim \sum_{i=1}^K \frac{(-1)^{i-1} \alpha^i}{i!} + \frac{c^2 \bar{\rho}}{2} \sum_{i=2}^K (-1)^{i-1} \frac{\alpha^i}{(i-2)!}$$

Where  $\bar{\rho} = \frac{1}{n(n-1)} \sum_{i \neq j} \rho_{ij}$  is the mean of correlations.

**Remarks** :- It is interesting to note that,  $\sum_{i=1}^K \frac{(-1)^{i-1} \alpha^i}{i!} \approx 1 - e^{-\alpha}$  for sufficiently large  $K$ . Also  $\lim_{n \rightarrow \infty} 1 - (1 - \alpha_n)^n = 1 - e^{-\alpha}$  is the limiting form of FWER under independence. So, if we use large enough  $K$  for this approximation then the first term is nearly equal to the FWER for independent case. So, the second term  $\frac{c^2 \bar{\rho}}{2} \sum_{i=2}^K (-1)^{i-1} \frac{\alpha^i}{(i-2)!}$  acts as a correction factor or the amount of deviation from independence. Under the nearly independent setup,  $\bar{\rho} = O(\frac{1}{n^\beta})$  (The correction factor is very small)

Before we proceed with the proof of theorem 3.4.1, the following important theorem must be stated.

**Theorem 3.4.2.** Multivariate Mill's Ratio (Savage (1962))

Let  $\tilde{X} \sim N_m(\mathbf{0}_m, V)$  and  $M = V^{-1}$ .

$$F(\tilde{a}, M) = P(\tilde{X} > \tilde{a}) \text{ and } f(\tilde{a}, M) = \frac{|M|^{\frac{1}{2}}}{(2\pi)^{\frac{m}{2}}} \exp(-\frac{1}{2} \tilde{a}^T M \tilde{a})$$

Let  $\tilde{\Delta} = \tilde{a}^T M$  (i.e.  $\Delta_i = \sum_j a_j m_{ji}$   $i = 1, 2, \dots, m$ )

If  $\Delta_i > 0 \forall i$ , then

$$1 - \frac{1}{2} \sum_i \sum_j \frac{m_{ij}(1 + \delta_{ij})}{\Delta_i \Delta_j} < \frac{F(\tilde{a}, M)}{f(\tilde{a}, M) (\prod_{i=1}^m \Delta_i)^{-1}} < 1$$

Here  $\delta_{ij} = \text{Kronecker's Delta}$

### Proof of theorem 3.4.1

**FWER** ( $\Sigma$ )

$$= \sum_{i=1}^n P(X_i > c) - \sum_{i \neq j} P(X_i > c, X_j > c) + \sum_{i,j,k \text{ distinct}} P(X_i > c, X_j > c, X_k > c) + \dots + (-1)^{n-1} P(X_1 > c, \dots, X_n > c)$$

Fix any  $1 \leq k \leq K \ll n$  where  $K$  is a fixed positive integer.

**Lemma 3.4.3.** :- For any  $1 \leq i_1 < \dots < i_k \leq n$ ,

$$P(X_{i_1} > c, \dots, X_{i_k} > c) \sim f(c\mathbf{1}_k) \left( \prod_{i=1}^k \Delta_i \right)^{-1}$$

(Where  $\Delta = c\mathbf{1}_k^T W^{-1}$ ,  $W$  is the correlation matrix of  $(X_{i_1}, X_{i_2}, \dots, X_{i_k})$  and  $f(\cdot)$  is the density function of  $N_k(\mathbf{0}_k, W)$  distribution)

**Proof** :-

Let  $W = I + R$ .

$$\Rightarrow W^{-1} = (I + R)^{-1} = I - R + o(R)$$

By the nearly independent assumption, we have  $(I + R)^{-1} \approx (I - R)$ .

$$\text{So, } \Delta = c\mathbf{1}_k^T W^{-1} \approx c\mathbf{1}_k^T (I - R).$$

$$\Rightarrow \Delta_i = c(1 - \sum_{j \neq i} \rho_{ji}) > 0 \quad \forall i \quad (\text{By the nearly independent assumption})$$

Now it is clear that,  $1 - \frac{1}{2} \sum_i \sum_j \frac{m_{ij}(1+\delta_{ij})}{\Delta_i \Delta_j} = 1 - O(\frac{1}{c^2})$ . (This is because the sum is over a fixed no. of terms ( $k^2$  many terms)).

Since  $c \rightarrow \infty$  as  $n \rightarrow \infty$ , we can say that,  $\lim_{n \rightarrow \infty} \frac{P(X_{i_1} > c, \dots, X_{i_k} > c)}{f(c\mathbf{1}_k) \left( \prod_{i=1}^k \Delta_i \right)^{-1}} = 1$  and hence the result.

**Approximation of**  $f(c\mathbf{1}_k) \left( \prod_{i=1}^k \Delta_i \right)^{-1}$

By the nearly independent assumption on  $\Sigma$ , we can say that,  $|I + R| \sim 1$ .

Observe that,

$$f(c\mathbf{1}_k) \sim \frac{1}{(2\pi)^{\frac{k}{2}}} \exp\left(-\frac{c^2}{2} \mathbf{1}_k^T (I - R) \mathbf{1}_k\right) = \frac{1}{(2\pi)^{\frac{k}{2}}} \exp\left(-\frac{kc^2}{2}\right) \exp\left(\frac{c^2}{2} \sum_{l \neq m} \rho_{lm}\right)$$

Since  $\rho_{lm} = O(\frac{1}{n^\beta})$  and  $c^2 = O(\log n)$ , this implies,  $c^2 \sum_{l \neq m} \rho_{lm} = o(1)$ .

We know that,  $e^x \sim 1 + x$  for sufficiently small  $x$ . So,

$$\exp\left(\frac{c^2}{2} \sum_{l \neq m} \rho_{lm}\right) \sim 1 + \frac{c^2}{2} \sum_{l \neq m} \rho_{lm}$$

$$\Rightarrow f(c\mathbf{1}_k) \sim \frac{1}{(2\pi)^{\frac{k}{2}}} \exp\left(-\frac{kc^2}{2}\right) \left(1 + \frac{c^2}{2} \sum_{l \neq m} \rho_{lm}\right) \text{ and } \left(\prod_{i=1}^k \Delta_i\right) = c^k \prod_{i=1}^k \left(1 - \sum_{j \neq i} \rho_{ji}\right).$$

Since  $\rho_{ij} = O\left(\frac{1}{n^\beta}\right)$ , this implies,  $\prod_{i=1}^k \left(1 - \sum_{j \neq i} \rho_{ji}\right) \sim 1$ .

Hence,

$$f(c\mathbf{1}_k) \left(\prod_{i=1}^k \Delta_i\right)^{-1} \approx \frac{c^k \left(1 + \frac{c^2}{2} \sum_{l \neq m} \rho_{lm}\right)}{(2\pi)^{\frac{k}{2}} \exp\left(\frac{kc^2}{2}\right)}$$

Recall that,  $\Phi(-c) = \alpha_n \sim \frac{\alpha}{n}$ .

If  $\phi(\cdot)$  denote the standard normal density, then for large enough  $c$ , then

$$\phi(c) \sim c \Phi(-c) \Rightarrow n \sim \alpha \sqrt{2\pi} c e^{\frac{c^2}{2}}$$

This implies, for any finite  $k$ ,  $f(c\mathbf{1}_k) \left(\prod_{i=1}^k \Delta_i\right)^{-1} \sim \left(\frac{\alpha}{n}\right)^k \left(1 + \frac{c^2}{2} \sum_{l \neq m} \rho_{lm}\right)$

In particular,

$$P(X_{i_1} > c, \dots, X_{i_k} > c) \sim \left(\frac{\alpha}{n}\right)^k \left(1 + \frac{c^2}{2} \sum_{l \neq m \in \{i_1, \dots, i_k\}} \rho_{lm}\right)$$

Approximation of  $FWER(\Sigma)$  under the nearly independent setup

**FWER** ( $\Sigma$ )

$$= \sum_{i=1}^n P(X_i > c) - \sum_{i \neq j} P(X_i > c, X_j > c) + \sum_{i, j, k \text{ distinct}} \sum P(X_i > c, X_j > c, X_k > c) \dots + (-1)^{n-1} P(X_1 > c, \dots, X_n > c)$$

Clearly,  $P(X_i > c) = \Phi(-c) = \alpha_n$  and hence  $\sum_{i=1}^n P(X_i > c) = n\alpha_n \sim \alpha$ .

For a fixed positive integer  $K$  ( $K > 2$ ), we have

$FWER(\Sigma)$

$$= n\alpha_n + \sum_{i=2}^K \sum_{\substack{u_1, \dots, u_i \\ \text{distinct}}} (-1)^{i-1} P(X_{u_1} > c, \dots, X_{u_i} > c) + \sum_{i=K+1}^n \sum_{\substack{u_1, \dots, u_i \\ \text{distinct}}} (-1)^{i-1} P(X_{u_1} > c, \dots, X_{u_i} > c)$$

c)

Clearly,

$$\begin{aligned} & \sum_{i=2}^K \sum_{\substack{u_1, \dots, u_i \\ \text{distinct}}} (-1)^{i-1} P(X_{u_1} > c, \dots, X_{u_i} > c) \\ & \sim \sum_{i=2}^K \sum_{\substack{u_1, \dots, u_i \\ \text{distinct}}} (-1)^{i-1} \left(\frac{\alpha}{n}\right)^i \left(1 + \frac{c^2}{2} \sum_{l \neq m \in \{u_1, \dots, u_i\}} \rho_{lm}\right) \\ & \sim \alpha + \sum_{i=2}^K (-1)^{i-1} \frac{\binom{n}{i} \alpha^i}{n^i} + \frac{c^2}{2} \sum_{i=2}^K (-1)^{i-1} \left(\frac{\alpha}{n}\right)^i \binom{n-2}{i-2} \left(\sum_{i \neq j} \rho_{ij}\right) \\ & \sim \sum_{i=1}^K \frac{(-1)^{i-1} \alpha^i}{i!} + \frac{c^2}{2n^2} \sum_{i=2}^K (-1)^{i-1} \frac{\alpha^i}{(i-2)!} \left(\sum_{i \neq j} \rho_{ij}\right) \\ & \sim \sum_{i=1}^K \frac{(-1)^{i-1} \alpha^i}{i!} + \frac{c^2 \bar{\rho}}{2} \sum_{i=2}^K (-1)^{i-1} \frac{\alpha^i}{(i-2)!} \end{aligned}$$

We ignore the tail term  $\sum_{i=K+1}^n \sum_{\substack{u_1, \dots, u_i \\ \text{distinct}}} (-1)^{i-1} P(X_{u_1} > c, \dots, X_{u_i} > c)$  from  $FWER(\Sigma)$  and we approximate the FWER by the following formula.

$$FWER(\Sigma) \sim \sum_{i=1}^K \frac{(-1)^{i-1} \alpha^i}{i!} + \frac{c^2 \bar{\rho}}{2} \sum_{i=2}^K (-1)^{i-1} \frac{\alpha^i}{(i-2)!}$$

**Remarks** :- It is interesting to note that, taking  $\rho_{ij}$ 's  $O(\frac{1}{n^\beta})$  is not necessary for the proof. It works for any order faster than  $\frac{1}{c^2}$  or  $\frac{1}{\log n}$ .

## 3.5 Simulation results

Bonferroni's procedure controls FWER at desired level under independence. We examine the FWER in the nearly independent setup and compare with the corrected value of FWER as per theorem 3.4.1. For these simulations we have considered  $n = 5000$  and  $\beta = 0.4, 0.6, 0.8, 1$ . They are repeated on four levels of significance ( $\alpha$ ) (namely  $\alpha = 0.01, 0.05, 0.1, 0.2$ ). For each combination, the actual FWER is estimated based on 10,000 replications and the corrected value as per theorem 3.4.1 is computed based on  $K = 15$ .

**Table 3.1:** FWER for level  $\alpha = 0.01$ 

$\beta$	FWER	FWER under independence	FWER with correction
0.4	0.0121	0.00995	0.00993
0.6	0.011	0.00995	0.00995
0.8	0.0108	0.00995	0.00995
1	0.0089	0.00995	0.00995

**Table 3.2:** FWER for level  $\alpha = 0.05$ 

$\beta$	FWER	FWER under independence	FWER with correction
0.4	0.0543	0.0487	0.0484
0.6	0.0475	0.0487	0.0487
0.8	0.0422	0.0487	0.0488
1	0.0495	0.0487	0.0488

**Table 3.3:** FWER for level  $\alpha = 0.1$ 

$\beta$	FWER	FWER under independence	FWER with correction
0.4	0.1077	0.0952	0.0934
0.6	0.0901	0.0952	0.095
0.8	0.0967	0.0952	0.0951
1	0.0983	0.0952	0.0952

**Table 3.4:** FWER for level  $\alpha = 0.2$ 

$\beta$	FWER	FWER under independence	FWER with correction
0.4	0.0192	0.0181	0.0178
0.6	0.0182	0.0181	0.018
0.8	0.0176	0.0181	0.0181
1	0.018	0.0181	0.0181

The simulation results indicate that there is minimal deviation in the FWER between the nearly independent and independent setups when employing Bonferroni's procedure. Due to the negligible correction factor, the adjusted values closely mirror those obtained under independence. This suggests that even in equicorrelated scenarios, the asymptotic behavior of the FWER exhibits a discontinuity at zero, remaining consistently zero for any positive correlation values (Dey (2021)). This study aims to explore the permissible extent of departure from independence within a general setup without significant FWER loss. Notably, FWER equivalence to independence persists in the nearly independent setup. Further investigations in this domain could involve explicitly quantifying correction factors attributable to correlations.

# Chapter 4

## Optimal test statistic under Gaussian distribution

### 4.1 Introduction

In the realm of multiple testing, conventional approaches often revolve around the rigid process of ranking and setting thresholds on p-values. However, the landscape of statistical inference experienced a notable shift with the pioneering work of Efron, who introduced an alternative avenue through the integration of Z-values and the empirical Bayes framework. In his seminal review [Efron \(2012\)](#), Efron elucidated an empirical Bayes interpretation of the Benjamini-Hochberg algorithm, reshaping the conventional paradigm. Departing from the conventional reliance on p-values, Efron's framework introduces the concept of the local false discovery rate (LFDR), offering a nuanced perspective on multiple testing. This novel approach transcends the limitations of traditional methodologies, providing a more robust and insightful framework for statistical inference in complex scenarios.

The intersection of the multiple testing problem with traditional classification theory first garnered attention through the seminal work of [Storey \(2003\)](#). [Storey \(2002\)](#) introduced a novel type I error rate, termed the positive false discovery rate (pFDR), positing its superiority over the conventional false discovery rate (FDR) in the context of multiple testing scenarios. The way [Genovese and Wasserman \(2002\)](#) proposed a dual quantity to

FDR, Storey (2003) introduced pFNR as a dual quantity to pFDR. Through meticulous analysis, Storey (2003) showcased the intrinsic connection between the multiple testing conundrum and the classification problem, particularly in terms of pFDR and pFNR.

This groundbreaking work set the stage for further exploration by researchers such as Sun and Cai (2007), Sun and Tony Cai (2009), as well as Xie et al. (2011). Amidst the era of burgeoning microarray experiments, the prevalence of dependent observations surpassed that of independent ones. Traditional algorithms, which often prioritized validity over efficiency, fell short when confronted with such dependent data. Recognizing this challenge, Sun and Tony Cai (2009) charted a distinct course, devising an adaptive multiple testing framework tailored for controlling false discoveries in the face of dependent observations. In their paper, they have used marginal false discovery rate (mFDR) and marginal false non-discovery rate (mFNR) in place of the traditional FDR and FNRs. However, Genovese and Wasserman (2002) have established that, under the assumption of independence, these are asymptotically the same in the sense that,  $\text{mFDR} = \text{FDR} + O(\frac{1}{\sqrt{n}})$  and  $\text{mFNR} = \text{FNR} + O(\frac{1}{\sqrt{n}})$ , where  $n$  is the number of hypotheses. Such asymptotic equivalence is valid in a more general setting (Xie et al. (2011)) of short range dependency structure. Sun and Tony Cai (2009) established a one to one correspondence between weighted classification problem and multiple hypothesis testing problem under the monotone ratio condition (MRC) and introduced a new test statistic named local index of significance (LIS) which is optimal in the sense that the test based on this statistic minimizes the mFNR among all methods that control mFDR at a certain level of significance. The optimality of their test statistic is a remarkable development because of the following two reasons.

- I It does not depend on the structure of dependency of the hypotheses.
- II It has been established under MRC condition which is fairly general. As Sun and Tony Cai (2009) have highlighted that, the test statistics that are defined on the basis of z-values, such as local false discovery rate (Efron et al. (2001)), p-value and the weighted p-value vector (Genovese et al. (2006)) belong to the MRC class.

The LIS statistic reduces to local false discovery rate (LFDR) under independence and the optimality of LFDR based procedures of Efron (2012), Efron et al. (2001) is thus established (Sun and Cai (2007)). However, the closed form expression of this optimal statistic is usually very difficult to find and this poses a major challenge to its application in real data. Sun and Tony Cai (2009) considered the hidden Markov model (HMM) where the latent indicator variable of being non-null follows a homogeneous irreducible Markov Chain and developed a recursive method for implementation of the optimal statistic based test. Xie et al. (2011) have implemented this test under multivariate normal distribution model. However, their original claim that, the optimal LIS statistic and LFDR is asymptotically the same, only holds under very stringent conditions imposed on model parameters (Xie et al. (2016)). In this chapter, we have studied the same model and substantially simplified the test statistic. The reason for considering multivariate Gaussian model is its wide applicability in real life problems and the results proved in this chapter hold for any positive definite correlation matrix.

## 4.2 Oracle Decision rule for Multivariate normal model

We consider testing  $n$  null hypotheses  $H_{01}, \dots, H_{0n}$  and for  $i = 1, 2, \dots, n$

$$\theta_i = \begin{cases} 1 & \text{if } i\text{-th null hypothesis is false} \\ 0 & \text{Otherwise} \end{cases}$$

Let  $\mathbf{X} = (X_1, \dots, X_n)$  be a sequence of test statistics for testing  $\mathbf{H}_0 = (H_{01}, \dots, H_{0n})$ .

In this chapter we consider the following model.

- $\theta_1, \dots, \theta_n \stackrel{i.i.d.}{\sim} \text{Ber}(p)$  for some  $0 < p < 1$
- $\mathbf{X}|\theta \sim N_n(k\theta, \Sigma)$  where  $\theta = (\theta_1, \dots, \theta_n)$  and  $k \neq 0$ . Here  $\Sigma$  is any positive definite covariance matrix.

A decision rule in any multiple testing problem is a function  $\delta : \mathcal{X} \rightarrow \{0, 1\}^n$  such that,

$$\delta_i = I_{\{H_{0i} \text{ is rejected}\}}$$

The landscape of decision rules in statistical inference has predominantly revolved around p-value-based marginal decision rules. While these rules have exhibited robust performance under scenarios of independent hypotheses, their efficacy diminishes when confronted with dependencies among hypotheses. A decision rule  $\delta$  is symmetric if  $\delta\{\tau(\mathbf{x})\} = \tau\{\delta(\mathbf{x})\}$  for all permutation operators  $\tau$ . In scenarios characterized by the anticipation of clustered non-null hypotheses, it becomes intuitive to discern between a hypothesis enveloped by other non-null hypotheses and one ensconced amidst null hypotheses. Yet, when subjected to a symmetric rule, these two hypotheses are treated interchangeably, disregarding their contextual disparities.

Traditional false discovery rate (FDR) procedures, which hinge on thresholding p-values or local false discovery rates (LFDR), operate under the guise of symmetric rules.

Consequently, they fail to adequately account for the nuances introduced by correlated hypotheses, rendering them suboptimal in such contexts. As pointed out by [Sun and Tony Cai \(2009\)](#), the optimal procedure of [Storey \(2007\)](#), being a symmetric decision rule, becomes inefficient in a problem where hypotheses are arising from a hidden markov model (HMM). Consequently, there arises a pressing need to embrace asymmetric decision-making frameworks to attain optimality in these nuanced contexts. In this chapter, we will consider the optimal rule based on the criteria of mFDR and mFNR suggested in [Sun and Tony Cai \(2009\)](#), [Xie et al. \(2011\)](#).

### 4.2.1 Discussion on error rate criteria

For any multiple testing procedure on these  $n$  hypotheses, let  $V$ ,  $R$ ,  $W$ ,  $A$  denote the number of false rejections, number of rejections, number of false acceptances and number of acceptances respectively. The false discovery rate (FDR) and marginal false discovery rate (mFDR) are defined as below.

$$\text{FDR} = E \left[ \frac{V}{R} I_{\{R>0\}} \right] \quad \text{and} \quad m\text{FDR} = \frac{E[V]}{E[R]}$$

These are versions of type - I error in the context of multiple testing. And, the versions of type -II errors are defined as

$$\text{FNR} = E \left[ \frac{W}{A} I_{\{A>0\}} \right] \quad \text{and} \quad m\text{FNR} = \frac{E[W]}{E[A]}$$

It can be easily shown by Jensen's inequality that,

$$\text{FDR} (\text{FNR}) \leq m\text{FDR} (m\text{FNR})$$

This implies, the methods which aims to control mFDR, tend to be more conservative than the methods controlling FDR. However, [Genovese and Wasserman \(2002\)](#) has shown that,  $m\text{FDR} (m\text{FNR}) = \text{FDR} (\text{FNR}) + O(\frac{1}{\sqrt{n}})$  under independence. [Xie et al. \(2011\)](#) have established the asymptotic equivalence of mFDR (mFNR) and FDR (FNR) under short range dependency criterion.

**Theorem 4.2.1.** ([Xie et al. \(2011\)](#)) Suppose  $\mathbf{X} = (X_1, \dots, X_n)$  is a sequence of random variables with same marginal density  $f$  and  $X_i$  and  $X_j$  are independent if  $|i - j| > n^\tau$  for some  $0 \leq \tau < 1$ . Let,  $\hat{\delta}_i = I_{\{S_i \in R\}}$  be a short-ranged rule to test  $H_{0i}$ , in the sense that  $S_i$  only depends on the variables that are dependent with  $X_i$ ,

$$S_i = S(X_{i-[n^\tau]}, \dots, X_{i+[n^\tau]})$$

Further, suppose that,

$$P(S_i \in R, \theta_i = 1) \geq P(S_i \in R, \theta_i = 0),$$

and,

$$P(S_i \in R, \theta_i = 1) > 0 \quad \forall i = 1, 2, \dots, n.$$

Then, the  $FDR(FNR)$  of the rule  $\hat{\delta}$  can be approximated by the  $mFDR(mFNR)$  in the sense that,

$$mFDR = FDR + O\left(\frac{1}{n^{1-\tau}}\right) \quad \text{and} \quad mFNR = FNR + O\left(\frac{1}{n^{1-\tau}}\right)$$

**Note :-** This asymptotic equivalence does not hold when the correlation matrix is not sparse (e.g. equi-correlated case)

In this chapter, we have considered  $mFDR$  and  $mFNR$  and derived an optimal test statistic which minimizes  $mFNR$  among all methods controlling  $mFDR$  at a pre-specified level of significance.

### 4.2.2 Oracle decision rule for multiple testing problem

Consider the weighted classification problem with decision rule  $\delta = (\delta_1, \dots, \delta_n) \in \{0, 1\}^n$ , with

$$\delta_i = \begin{cases} 1 & \text{if } i\text{-th hypothesis is rejected} \\ 0 & \text{otherwise} \end{cases}$$

Consider the loss function

$$L_\lambda(\delta, \theta) = \frac{1}{n} \sum_{i=1}^n \{\delta_i(1 - \theta_i) + \lambda\theta_i(1 - \delta_i)\} \quad (4.1)$$

with  $\lambda > 0$  the weight for a false positive result. It is well-known that, if  $g(x|\theta_i = j)$  denotes the density of  $x$  when  $\theta_i = j$  ( $j = 0, 1$ ), then, the classification risk  $E[L_\lambda(\theta, \delta)]$  is minimized by the Bayes rule  $\delta(\Lambda, \lambda) = (\delta_1, \dots, \delta_n)$ , where

$$\delta_i = I\left\{\Lambda_i(x) = \frac{(1-p)g(x|\theta_i = 0)}{pg(x|\theta_i = 1)} < \lambda\right\} \quad (4.2)$$

Alternatively, if the goal is to discover as many significant hypotheses as possible while incurring a relatively low proportion of false positives, we can study a multiple testing problem where the goal is to find a decision rule  $\delta$  that has the smallest  $FNR(mFNR)$  among all  $FDR(mFDR)$  procedures at level  $\alpha$ . Consider the set of all Bayes rules for the

classification problem and name it as  $\mathcal{D}_s$ . i.e.

$$\mathcal{D}_s = \{\delta \mid \delta_i = I\{\Lambda_i(x) < \lambda\} \text{ for some } \lambda > 0\}$$

Sun and Tony Cai (2009), Xie et al. (2011) has shown that, among all procedures controlling mFDR at level  $\alpha$ , a procedure which minimizes the mFNR must belong to the class  $\mathcal{D}_s$ . The following theorem explicate the whole idea.

**Theorem 4.2.2.** ( Xie et al. (2011) ) Consider the class of decision rules  $\mathcal{D}_s = \{\delta : \delta_i = I_{\Lambda_i < \lambda}, i = 1, \dots, n\}$  where  $\Lambda = \{\Lambda_1, \dots, \Lambda_n\}$  is defined in (4.2) and  $\lambda \in \mathbb{R}$ . Given any mFDR level  $\alpha$  and a decision rule

$$\delta(S, R) = \{I_{S_1 \in R_1}, \dots, I_{S_n \in R_n}\}$$

with  $mFDR(\delta(S, R)) \leq \alpha$ . Then there exists a  $\lambda$  depending on  $\delta(S, R)$ , such that,  $\delta(\Lambda, \lambda) \in \mathcal{D}_s$  outperforms  $\delta(S, R)$  in the sense that,

$$mFDR(\delta(\Lambda, \lambda)) \leq mFDR(\delta(S, R)) \leq \alpha,$$

and

$$mFNR(\delta(\Lambda, \lambda)) \leq mFNR(\delta(S, R))$$

Theorem 4.2.2 implies that, the optimal solution of the multiple testing problem with mFDR and mFNR as the error rate criteria, belongs to the set  $\mathcal{D}_s$ . Instead of searching for all decision rules, one only needs to search in the collection  $\mathcal{D}_s$  for the optimal rule. The following result shows that, for a given  $\alpha$ , the optimal rule for the multiple testing problem is unique.

**Theorem 4.2.3.** ( Xie et al. (2011) ) Consider the optimal decision rule  $\delta(\Lambda, \lambda)$  in the weighted classification problem with the loss function (4.1). For any  $0 < \alpha < 1$ , there exists a unique  $\lambda(\alpha)$ , such that  $\delta\{\Lambda, \lambda(\alpha)\}$  controls the mFDR at level  $\alpha$  and minimizes the mFNR among all decision rules.

Theorem 4.2.2 gives us the optimal testing rule with mFDR and mFNR as the error rate criteria and theorem 4.2.3 establishes a one-to-one correspondence between the multiple testing problem and weighted classification problem. However, it is often hard to determine the  $\lambda(\alpha)$  corresponding to the given  $\alpha$ .

### 4.2.3 Implementation of the optimal test

Xie et al. (2011) have provided a method to implement the optimal test of theorem 4.2.2.

Define,

$$T_{OR,i} = P(\theta_i = 0 | x) = \frac{(1-p)g(x|\theta_i = 0)}{g(x)}$$

Clearly,  $T_{OR,i} = \frac{\Lambda_i}{1+\Lambda_i}$  increases with  $\Lambda_i$ . Thus, for a given mFDR value  $\alpha$ , one can rewrite the optimal rule as

$$\delta_{OR,i} = \delta(\Lambda, \lambda(\alpha)) = I \left\{ T_{OR,i} < \frac{\lambda(\alpha)}{1 + \lambda(\alpha)} \right\}$$

Let  $T_{OR,(i)}$  denote the  $i$ -th order statistic of  $T_{OR,i}$  and  $H_{0,(i)}$  be the corresponding null hypothesis ( $i = 1, \dots, n$ ). Sun and Tony Cai (2009) have shown that, if  $R$  denote the number of rejections, mFDR can be written as follows.

$$mFDR = E \left[ \frac{1}{R} \sum_{i=1}^R T_{OR,(i)} \right]$$

Then, according to the theorem 4 of Sun and Tony Cai (2009) or theorem 5 of Xie et al. (2011), if  $p$  and  $g$  are known, then the following method controls mFDR at level  $\alpha$  :

$$\text{Reject all } H_{0,(i)} \text{ for } i = 1, \dots, k \text{ where } k = \max \left\{ l : \frac{1}{l} \sum_{i=1}^l T_{OR,(i)} \leq \alpha \right\} \quad (4.3)$$

The final oracle rule (4.3) consists of two steps :

- Calculate the oracle statistic  $T_{OR,i}$  for  $i = 1, \dots, n$ .
- Rank the statistics and calculate the running averages to determine the cutoff. All hypotheses below the cutoff are rejected.

However, the major difficulty associated with this optimal test is that the test statistic  $T_{OR,i}$  is often very difficult to compute. A simplified expression of this test statistic is hard to find and the model parameters are difficult to estimate under dependent models. In this article, we provide a method for implementation of the optimal test under the multivariate normal model.

### 4.3 Oracle decision rule for Gaussian model

Xie et al. (2011) and Xie et al. (2016) have established the asymptotic equivalence between the oracle decision rule and the LFDR based rule under some assumptions as follows.

**Theorem 4.3.1.** (Xie et al. (2011), Xie et al. (2016))

Let,  $\theta = (\theta_1, \dots, \theta_n)$  satisfy  $\theta_1, \dots, \theta_n \stackrel{i.i.d.}{\sim} \text{Ber}(p)$  and  $X | \theta \sim N(\mu(\theta), \Sigma)$ , where  $\Sigma$  is a correlation matrix and  $\mu(\theta) = (\mu_1\theta_1, \mu_2\theta_2, \dots, \mu_n\theta_n)$ . Let  $T_{OR,i} = P(\theta_i = 0 | x)$  be the test statistic of the oracle rule as discussed before and  $T_{MG,i} = P(\theta_i = 0 | x_i)$  be the marginal test statistic which is same as local false discovery rate (LFDR).

We make the following assumptions about the distribution.

(i)  $\frac{1}{n^{\tau_1}} \leq p \leq 1 - \frac{1}{n^{\tau_1}}$  for some  $0 < \tau_1 < 1$ .

(ii)  $\mu_1, \dots, \mu_n$  are i.i.d. random variables having distribution function  $G$ .  $G(\cdot)$  satisfies

$$G((2\tau_2 \log n)^{\frac{1}{2}}) - G(-(2\tau_2 \log n)^{\frac{1}{2}}) = 0 \text{ for some } \tau_2 > \tau_1$$

This guarantees that,  $P(|\mu_i| \leq (2\tau_2 \log n)^{\frac{1}{2}}) = 1$  for  $i = 1, \dots, n$

(iii) The minimum eigenvalue of  $\Sigma$  is bounded away from zero. i.e.

$$\liminf_{n \rightarrow \infty} \lambda_{\min}(\Sigma) = \kappa > 0$$

(iv) The correlation matrix  $\Sigma$  satisfies the short range dependency structure. i.e.

$$\Sigma_{ij} = 0 \text{ for } |i - j| > n^\tau \text{ for some } 0 < \tau < 1$$

If assumptions (i)-(iv) are satisfied, then for all  $\epsilon > 0$ ,

$$\lim_{n \rightarrow \infty} P(|T_{OR,i} - T_{MG,i}| \geq \epsilon) = 0$$

Theorem 4.2.2 significantly simplifies the complex test statistic associated with the Oracle rule, given assumptions (i)-(iv) and a short-range dependency structure. The subsequent theorem further simplifies this statistic within the Gaussian model, accommodating any correlation matrix  $\Sigma$ .

**Theorem 4.3.2.** *If  $\mathbf{X} | \theta \sim N_n(k\theta, \Sigma)$  and  $\theta_1, \dots, \theta_n \stackrel{i.i.d.}{\sim} \text{Ber}(p)$ , then for  $i = 1, \dots, n$*

$$P(\theta_i = 0 | \mathbf{X}) = \frac{1}{1 + \frac{pU_i}{1-p}}$$

Where  $U_i = \exp(-(\frac{k^2}{2}t_{i,i} - k \sum_{j=1}^n t_{j,i}x_j))(\prod_{j \neq i} (pe^{-k^2t_{j,i}} + (1-p)))$  and  $t_i = (t_{1,i}, \dots, t_{n,i})$  is the  $i$ -th column of  $\Sigma^{-1}$ .

Proof of theorem 4.3.2 is given in appendix.

**Remarks** :- While the equivalence of the joint conditional probability of Xie et al. (2011) and Xie et al. (2016) was established under very restrictive assumptions on the correlation matrix, theorem 4.3.2 sufficiently simplifies the conditional probability for any covariance matrix. The result of 4.3.2 enables us to implement the Oracle decision rule for any  $p, \Sigma$ . We have performed extensive simulations with different combinations of  $p$  and  $\Sigma$  and compared the observed value of FDR and FNRs some of which will be discussed here.

### 4.3.1 Simulation Studies

In this section, we evaluate the performance of the oracle rule and compare with the BH procedure and the marginal procedure mentioned in Xie et al. (2011). We have evaluated the empirical FDR, FNR and also the number of rejections. In our simulations, we

assumed a multivariate normal model :

$$X | \theta \sim N(c\theta, \Sigma)$$

where  $\theta_i$  follows Bernoulli( $p$ ). Under this model, the non-null distribution has mean  $c$  and  $\Sigma$  is a correlation matrix. For our simulations, we have considered  $c = 2.5$  and  $\alpha = 0.05$ . Our objective is to assess the performance of these methods when there is sufficient deviation from independence. In the first case, we have considered equicorrelated  $\Sigma$ . In all the simulations, number of hypotheses ( $n$ ) have been considered to be 5000 and they are run on 10 combinations of proportion of non-null  $p = 0.01, 0.02, \dots, 0.1$ . In order to assess the performance under sufficient deviation from independence, seven combinations of correlation have been considered ( $\rho = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$ ). The results suggest that, the Oracle procedure is least conservative among the three procedures in terms of FDR. It is interesting to note that, the FDR of the Oracle rule always lies within the prescribed limit of 0.05. Maintaining this upper bound on FDR, there is a substantial gain in the FNR over both BH and marginal procedure. It is interesting to note that, the marginal procedure becomes more conservative than the other two methods. However, the FNR of the marginal procedure remains similar to the BH procedure which suggests a possibility of improvement of this method and that is achieved by considering the information of joint distribution in the Oracle procedure. The conservative nature of the marginal procedure in comparison to the BH method is possibly due to the objective of controlling mFDR instead of FDR. Since the FNR remains equivalent for these two methods, careful examination of the class  $\mathcal{D}_s$  may provide a significantly better test statistic.

**Table 4.1:** FDRs of the three methods

Marginal Procedure	BH procedure	Oracle procedure	p	Correlation
0.016	0.042	0.043	0.01	0.2
0.011	0.038	0.044	0.01	0.3
0.007	0.033	0.044	0.01	0.4
0.005	0.029	0.044	0.01	0.5
0.003	0.023	0.043	0.01	0.6
0.003	0.019	0.041	0.01	0.7
0.002	0.017	0.039	0.01	0.8
0.022	0.042	0.045	0.02	0.2
0.015	0.039	0.044	0.02	0.3
0.011	0.035	0.043	0.02	0.4
0.008	0.031	0.043	0.02	0.5
0.006	0.027	0.041	0.02	0.6
0.005	0.022	0.040	0.02	0.7
0.005	0.019	0.040	0.02	0.8
0.025	0.042	0.043	0.03	0.2
0.018	0.040	0.042	0.03	0.3
0.013	0.036	0.042	0.03	0.4
0.010	0.032	0.040	0.03	0.5
0.009	0.029	0.039	0.03	0.6
0.008	0.024	0.037	0.03	0.7
0.007	0.019	0.039	0.03	0.8
0.027	0.043	0.042	0.04	0.2
0.020	0.040	0.041	0.04	0.3
0.015	0.037	0.039	0.04	0.4
0.012	0.034	0.038	0.04	0.5
Continued on next page				

Table 4.1 (FDRs)– continued from previous page

Marginal Procedure	BH procedure	Oracle procedure	p	Correlation
0.011	0.031	0.036	0.04	0.6
0.010	0.026	0.034	0.04	0.7
0.009	0.020	0.038	0.04	0.8
0.028	0.042	0.040	0.05	0.2
0.022	0.040	0.039	0.05	0.3
0.018	0.038	0.037	0.05	0.4
0.014	0.034	0.036	0.05	0.5
0.013	0.031	0.034	0.05	0.6
0.012	0.027	0.032	0.05	0.7
0.011	0.022	0.037	0.05	0.8
0.030	0.042	0.038	0.06	0.2
0.023	0.040	0.037	0.06	0.3
0.019	0.038	0.036	0.06	0.4
0.016	0.035	0.033	0.06	0.5
0.014	0.031	0.031	0.06	0.6
0.013	0.028	0.029	0.06	0.7
0.013	0.023	0.036	0.06	0.8
0.031	0.042	0.037	0.07	0.2
0.025	0.040	0.035	0.07	0.3
0.020	0.038	0.034	0.07	0.4
0.018	0.035	0.031	0.07	0.5
0.016	0.032	0.029	0.07	0.6
0.015	0.028	0.027	0.07	0.7
0.015	0.023	0.034	0.07	0.8
0.032	0.042	0.035	0.08	0.2
0.026	0.040	0.034	0.08	0.3

Continued on next page

Table 4.1 (FDRs)– continued from previous page

Marginal Procedure	BH procedure	Oracle procedure	p	Correlation
0.022	0.038	0.032	0.08	0.4
0.019	0.036	0.029	0.08	0.5
0.017	0.032	0.027	0.08	0.6
0.017	0.030	0.025	0.08	0.7
0.016	0.025	0.033	0.08	0.8
0.032	0.041	0.034	0.09	0.2
0.027	0.039	0.032	0.09	0.3
0.023	0.037	0.030	0.09	0.4
0.020	0.036	0.027	0.09	0.5
0.019	0.033	0.025	0.09	0.6
0.019	0.030	0.023	0.09	0.7
0.018	0.025	0.032	0.09	0.8
0.033	0.041	0.032	0.1	0.2
0.028	0.039	0.031	0.1	0.3
0.024	0.038	0.028	0.1	0.4
0.022	0.036	0.026	0.1	0.5
0.020	0.033	0.023	0.1	0.6
0.020	0.030	0.021	0.1	0.7
0.019	0.025	0.030	0.1	0.8

From the FDRs, it is clear that,

- All the three methods (especially BH procedure) become more and more conservative with increasing value of the correlation.
- Marginal procedure tends to be the most conservative among the other three methods for higher correlations. (i.e.  $FDR_{MP} \leq FDR_{BH} \leq FDR_{OP}$ . However,

slight exceptions can be observed for smaller correlations and higher  $p$  where BH procedure has slightly higher FDR among all.

With the above mentioned observations on FDR, it is imperative to note the FNRs of these three methods.

**Table 4.2:** FNRs of the three methods

Marginal Procedure	BH procedure	Oracle procedure	$p$	Correlation
0.0094	0.0089	0.0085	0.01	0.2
0.0094	0.0089	0.0077	0.01	0.3
0.0094	0.0088	0.0065	0.01	0.4
0.0094	0.0088	0.0049	0.01	0.5
0.0094	0.0088	0.0030	0.01	0.6
0.0094	0.0087	0.0012	0.01	0.7
0.0094	0.0086	0.0001	0.01	0.8
0.0177	0.0170	0.0154	0.02	0.2
0.0177	0.0169	0.0135	0.02	0.3
0.0177	0.0168	0.0110	0.02	0.4
0.0177	0.0168	0.0080	0.02	0.5
0.0177	0.0166	0.0047	0.02	0.6
0.0178	0.0165	0.0017	0.02	0.7
0.0179	0.0164	0.0002	0.02	0.8
0.0253	0.0246	0.0216	0.03	0.2
0.0253	0.0244	0.0187	0.03	0.3
0.0253	0.0243	0.0150	0.03	0.4
0.0254	0.0242	0.0106	0.03	0.5
0.0254	0.0240	0.0060	0.03	0.6
0.0255	0.0239	0.0021	0.03	0.7

Continued on next page

Table 4.2(FNRs) – continued from previous page

Marginal Procedure	BH procedure	Oracle procedure	p	Correlation
0.0257	0.0236	0.0002	0.03	0.8
0.0322	0.0317	0.0274	0.04	0.2
0.0323	0.0315	0.0234	0.04	0.3
0.0324	0.0314	0.0186	0.04	0.4
0.0325	0.0312	0.0130	0.04	0.5
0.0325	0.0310	0.0073	0.04	0.6
0.0327	0.0307	0.0024	0.04	0.7
0.0331	0.0305	0.0002	0.04	0.8
0.0389	0.0385	0.0329	0.05	0.2
0.0389	0.0382	0.0280	0.05	0.3
0.0390	0.0381	0.0221	0.05	0.4
0.0392	0.0380	0.0153	0.05	0.5
0.0393	0.0376	0.0085	0.05	0.6
0.0396	0.0374	0.0028	0.05	0.7
0.0400	0.0370	0.0002	0.05	0.8
0.0451	0.0450	0.0382	0.06	0.2
0.0452	0.0448	0.0323	0.06	0.3
0.0453	0.0446	0.0254	0.06	0.4
0.0455	0.0444	0.0176	0.06	0.5
0.0456	0.0440	0.0097	0.06	0.6
0.0460	0.0437	0.0032	0.06	0.7
0.0466	0.0433	0.0002	0.06	0.8
0.0510	0.0511	0.0434	0.07	0.2
0.0511	0.0510	0.0366	0.07	0.3
0.0512	0.0507	0.0287	0.07	0.4
0.0515	0.0505	0.0198	0.07	0.5
Continued on next page				

Table 4.2(FNRs) – continued from previous page

Marginal Procedure	BH procedure	Oracle procedure	p	Correlation
0.0517	0.0503	0.0109	0.07	0.6
0.0521	0.0499	0.0035	0.07	0.7
0.0529	0.0494	0.0002	0.07	0.8
0.0565	0.0571	0.0484	0.08	0.2
0.0568	0.0570	0.0408	0.08	0.3
0.0570	0.0569	0.0320	0.08	0.4
0.0572	0.0566	0.0221	0.08	0.5
0.0575	0.0562	0.0121	0.08	0.6
0.0581	0.0559	0.0039	0.08	0.7
0.0590	0.0553	0.0002	0.08	0.8
0.0621	0.0630	0.0535	0.09	0.2
0.0622	0.0628	0.0450	0.09	0.3
0.0624	0.0626	0.0352	0.09	0.4
0.0625	0.0622	0.0244	0.09	0.5
0.0631	0.0620	0.0134	0.09	0.6
0.0636	0.0614	0.0044	0.09	0.7
0.0647	0.0608	0.0002	0.09	0.8
0.0672	0.0686	0.0585	0.1	0.2
0.0673	0.0685	0.0493	0.1	0.3
0.0675	0.0682	0.0385	0.1	0.4
0.0678	0.0680	0.0267	0.1	0.5
0.0683	0.0675	0.0148	0.1	0.6
0.0692	0.0672	0.0048	0.1	0.7
0.0705	0.0665	0.0002	0.1	0.8

As per the optimality of Oracle procedure, it has the lowest FNR among all. It is interesting to note that, while the marginal procedure was the most conservative in terms of FDR, its FNR is nearly equivalent (or even better in some cases) to the BH procedure. It is again reminded that, we are controlling the mFDR(mFNR) instead of FDR(FNR). The results suggest that there is a scope of further improvement in the class  $\mathcal{D}_s = \{\delta : \delta_i = I_{\{\Lambda_i < \lambda\}}, i = 1, \dots, n\}$  if  $\Lambda$  and  $\lambda$  can be chosen properly.

Examining the FDRs and FNRs does not entirely describe how conservative a method is. We know that, these methods become conservative with increasing value of correlation. To examine this, we have also tabulated the number of rejections of these three methods in different combinations of  $p$  and correlation.

**Table 4.3:** No. of rejections of the three methods

Marginal Procedure	BH procedure	Oracle procedure	p	Correlation
3	8	8	0.01	0.2
3	13	12	0.01	0.3
3	19	19	0.01	0.4
3	28	27	0.01	0.5
3	36	37	0.01	0.6
3	52	46	0.01	0.7
3	70	51	0.01	0.8
13	19	25	0.02	0.2
13	25	34	0.02	0.3
12	31	47	0.02	0.4
12	40	63	0.02	0.5
12	51	80	0.02	0.6
12	62	95	0.02	0.7
12	83	103	0.02	0.8
26	34	45	0.03	0.2

Continued on next page

Table 4.3(No. of rejections) – continued from previous page

Marginal Procedure	BH procedure	Oracle procedure	p	Correlation
26	40	60	0.03	0.3
26	47	80	0.03	0.4
25	55	102	0.03	0.5
25	67	125	0.03	0.6
25	79	145	0.03	0.7
24	95	155	0.03	0.8
42	51	68	0.04	0.2
42	57	89	0.04	0.3
42	64	113	0.04	0.4
41	72	142	0.04	0.5
41	85	171	0.04	0.6
40	96	195	0.04	0.7
38	111	207	0.04	0.8
61	69	92	0.05	0.2
61	75	118	0.05	0.3
61	83	149	0.05	0.4
60	92	183	0.05	0.5
60	104	217	0.05	0.6
58	117	244	0.05	0.7
56	130	259	0.05	0.8
83	89	118	0.06	0.2
82	95	149	0.06	0.3
82	103	184	0.06	0.4
81	113	223	0.06	0.5
80	123	262	0.06	0.6
78	136	294	0.06	0.7

Continued on next page

Table 4.3(No. of rejections) – continued from previous page

Marginal Procedure	BH procedure	Oracle procedure	p	Correlation
75	153	310	0.06	0.8
106	111	145	0.07	0.2
105	118	180	0.07	0.3
105	125	220	0.07	0.4
104	134	264	0.07	0.5
102	145	308	0.07	0.6
100	157	343	0.07	0.7
96	170	361	0.07	0.8
131	135	172	0.08	0.2
130	141	212	0.08	0.3
129	149	257	0.08	0.4
128	158	305	0.08	0.5
126	166	353	0.08	0.6
124	182	392	0.08	0.7
119	195	413	0.08	0.8
157	158	200	0.09	0.2
157	165	244	0.09	0.3
155	172	293	0.09	0.4
155	182	346	0.09	0.5
152	192	398	0.09	0.6
150	207	440	0.09	0.7
143	219	464	0.09	0.8
186	184	228	0.1	0.2
185	190	276	0.1	0.3
184	198	329	0.1	0.4
182	206	387	0.1	0.5

Continued on next page

**Table 4.3(No. of rejections) – continued from previous page**

Marginal Procedure	BH procedure	Oracle procedure	p	Correlation
179	215	443	0.1	0.6
175	227	488	0.1	0.7
167	241	515	0.1	0.8

Number of rejections for the Oracle procedure is significantly higher than the other three methods and hence this is the least conservative among all. However, the equicorrelated  $\Sigma$  is an unlikely scenario in real life applications. We only considered this in order to generate a scenario which is substantially different from the independent setup and compare the performances of the methods. Now we present the results on block diagonal correlation matrix. Here we have divided the correlation matrix in four equal sized blocks of equicorrelated matrices with correlation 0.15, 0.25, 0.5, 0.75. The results again suggest that, the Oracle Procedure is least conservative among the three methods in terms of FDR while maintaining the prescribed limit of 0.05. Number of rejections for Oracle procedure is significantly higher than the other two and the gain in power is also noteworthy.

**Table 4.4:** FDRs in block diagonal case

Marginal procedure	BH procedure	Oracle procedure	p
0.0185	0.0369	0.0435	0.01
0.0274	0.0385	0.0432	0.02
0.0309	0.0394	0.0416	0.03
0.0329	0.0399	0.0398	0.04
0.0346	0.0403	0.0380	0.05
0.0358	0.0403	0.0361	0.06
0.0370	0.0404	0.0345	0.07
0.0380	0.0405	0.0328	0.08
0.0388	0.0402	0.0313	0.09
0.0395	0.0401	0.0298	0.1

**Table 4.5:** FNRs in block diagonal case

Marginal procedure	BH procedure	Oracle procedure	p
0.0094	0.0090	0.0052	0.01
0.0177	0.0173	0.0091	0.02
0.0253	0.0249	0.0125	0.03
0.0323	0.0320	0.0157	0.04
0.0389	0.0388	0.0187	0.05
0.0452	0.0453	0.0216	0.06
0.0511	0.0515	0.0245	0.07
0.0567	0.0575	0.0274	0.08
0.0621	0.0633	0.0303	0.09
0.0673	0.0689	0.0332	0.1

**Table 4.6:** No. of rejections in block diagonal case

Marginal Procedure	BH Procedure	Oracle Procedure	p
3	9	25	0.01
12	19	57	0.02
26	33	92	0.03
42	49	129	0.04
61	67	166	0.05
82	87	204	0.06
106	108	242	0.07
131	131	280	0.08
157	156	318	0.09
186	181	357	0.1

As mentioned earlier, Oracle procedure exploits the information of joint distribution unlike the marginal and BH procedure. The results from the simulation studies have shown a significant improvement in FNR and the number of rejections in exchange of very little sacrifice in FDR. Hence, it is interesting to explore the class  $\mathcal{D} = \{I_{\{X_i > c\}} \mid i = 1, \dots, n\}$  and to search for a different choice of  $c$  which can provide further improvement. Also, implementation of the optimal procedure under a more general dependency setup (e.g.  $m$ -dependent structure) is still a challenging open problem.

## 4.4 Annexure

### 4.4.1 Proof of theorem 4.3.2

Let  $f(\mathbf{x}, \theta)$  denote the value of  $N(k\theta, \Sigma)$  density at  $\mathbf{x}$ . Then,

$$P(\theta_i = 0 \mid \mathbf{x}) = \frac{(1-p)E_{\theta_{0,i}}[f(\mathbf{x}, \theta_{0,i})]}{E_{\theta}[f(\mathbf{x}, \theta)]}$$

Where  $\theta_{0,i}$  has 0 in it's  $i$ -th place.

Observe that,

$$\frac{(1-p)E_{\theta_{0,i}}[f(\mathbf{x}, \theta_{0,i})]}{E_{\theta}[f(\mathbf{x}, \theta)]} = \frac{(1-p) \sum_{\theta_{0,i}} f(\mathbf{x}, \theta_{0,i}) h_0(\theta_{0,i})}{(1-p) \sum_{\theta_{0,i}} f(\mathbf{x}, \theta_{0,i}) h_0(\theta_{0,i}) + p \sum_{\theta_{1,i}} f(\mathbf{x}, \theta_{1,i}) h_1(\theta_{1,i})}$$

Where  $\theta_{1,i}$  has 1 in it's  $i$ -th place and  $h_0$  and  $h_1$  are the joint p.m.f.s of  $\theta$  given  $\theta_i = 0$  and  $\theta_i = 1$  respectively. In particular,  $\theta_{1,i} = \theta_{0,i} + e_i$  ( $e_i$  is the vector with 1 in the  $i$ -th place and 0 elsewhere)

Let  $B = \sum_{\theta_{1,i}} f(\mathbf{x}, \theta_{1,i}) h_1(\theta_{1,i})$  and  $A = \sum_{\theta_{0,i}} f(\mathbf{x}, \theta_{0,i}) h_0(\theta_{0,i})$ .

Then,  $P(\theta_i = 0 \mid \mathbf{x}) = \frac{1}{1 + \frac{pB}{(1-p)A}}$  (A monotone function in  $\frac{B}{A}$ )

Since  $\theta_i$ 's are i.i.d., we must have  $h_0 = h_1$ . Putting  $\theta_{1,i} = \theta_{0,i} + e_i$ , we get,

$$f(\mathbf{x}, \theta_{1,i}) = f(\mathbf{x}, \theta_{0,i}) \exp\left(-\frac{k^2}{2} e_i^T \Sigma^{-1} e_i\right) \exp(k(\mathbf{x} - k\theta_{0,i})^T \Sigma^{-1} e_i)$$

Let  $t_i = (t_{1,i}, \dots, t_{n,i})$  be the  $i$ -th column of  $\Sigma^{-1}$ . Then,  $e_i^T \Sigma^{-1} e_i = t_{i,i}$  and  $\mathbf{x}^T \Sigma^{-1} e_i = \sum_{j=1}^n t_{j,i} x_j$ . This implies,

$$f(\mathbf{x}, \theta_{1,i}) = f(\mathbf{x}, \theta_{0,i}) \exp\left(-\left(\frac{k^2}{2} t_{i,i} - k \sum_{j=1}^n t_{j,i} x_j\right)\right) \exp(-k^2 \theta_{0,i}^T \Sigma^{-1} e_i)$$

Observe that,

$$\begin{aligned} & \frac{\sum_{\theta_{1,i}} f(\mathbf{x}, \theta_{1,i}) h_1(\theta_{1,i})}{\sum_{\theta_{0,i}} f(\mathbf{x}, \theta_{0,i}) h_0(\theta_{0,i})} \\ &= \frac{E_{\theta_{1,i}}[f(\mathbf{x}, \theta_{1,i})]}{E_{\theta_{0,i}}[f(\mathbf{x}, \theta_{0,i})]} \\ &= \exp\left(-\left(\frac{k^2}{2} t_{i,i} - k \sum_{j=1}^n t_{j,i} x_j\right)\right) \left(\frac{E_{\theta_{0,i}}[f(\mathbf{x}, \theta_{0,i}) \exp(-k^2 \theta_{0,i}^T \Sigma^{-1} e_i)]}{E_{\theta_{0,i}}[f(\mathbf{x}, \theta_{0,i})]}\right) \end{aligned}$$

Note that,  $f(\mathbf{x}, \theta_{0,i}) = g(\mathbf{x} - \theta_{0,i}) = A$  function of  $(\mathbf{x} - k\theta_{0,i})$ . As per our model  $(\mathbf{X} - \theta_{0,i})$  is independent of  $\theta_{0,i}$  and hence, we can say that,

$$E_{\theta_{0,i}}[f(\mathbf{x}, \theta_{0,i}) \exp(-k^2 \theta_{0,i}^T \Sigma^{-1} e_i)] = E_{\theta_{0,i}}[f(\mathbf{x}, \theta_{0,i})] C$$

And thus,

$$\frac{B}{A} = \exp\left(-\left(\frac{k^2}{2} t_{i,i} - k \sum_{j=1}^n t_{j,i} x_j\right)\right) E_{\theta_{0,i}}[\exp(-k^2 \theta_{0,i}^T \Sigma^{-1} e_i)]$$

Note that,  $\theta_{0,i}^T \Sigma^{-1} e_i = \sum_{j \neq i} \theta_j t_{j,i}$  and from the independence of  $\theta_j$ 's we can conclude that,

$$E_{\theta_{0,i}}[\exp(-k^2 \theta_{0,i}^T \Sigma^{-1} e_i)] = \prod_{j \neq i} E[e^{-t_j \theta_j}] = \prod_{j \neq i} (pe^{-k^2 t_{j,i}} + (1-p))$$

Thus, we finally obtain a simplified expression of the optimal test statistic as the following

$$P(\theta_i = 0 | \mathbf{X}) = \frac{1}{1 + \frac{pU_i}{1-p}}$$

Where  $U_i = \exp\left(-\left(\frac{k^2}{2} t_{i,i} - k \sum_{j=1}^n t_{j,i} x_j\right)\right) \left(\prod_{j \neq i} (pe^{-k^2 t_{j,i}} + (1-p))\right)$  and  $t_i = (t_{1,i}, \dots, t_{n,i})$  is the  $i$ -th column of  $\Sigma^{-1}$

# Chapter 5

## Estimation of $\pi_0$ under dependence

### 5.1 Introduction

It has been shown that, the FDR controlling procedure of Benjamini-Hochberg algorithm controls FDR at a level less than or equal to  $\frac{n_0}{n}\alpha = \pi_0\alpha$  under independence ([Benjamini and Hochberg \(1995\)](#)) (Here  $\pi_0$  represents the proportion of true null hypotheses).

Traditional methods often reach their maximum Type I error rate when faced with the global null hypothesis, leaving ample room for increased statistical power, particularly when  $\pi_0 < 1$ . In an ideal scenario where  $\pi_0$  is known, adjusting the Benjamini-Hochberg procedure to precisely control the FDR at a specified  $\alpha$  level would be straightforward. However, in practice,  $\pi_0$  remains unknown and necessitates estimation from the available data. [Benjamini and Hochberg \(2000\)](#) proposed an adaptive FDR controlling procedure where  $\pi_0$  was estimated based on a graphical approach similar to [Schweder and Spjøtvoll \(1982\)](#). Building upon this foundation, [Storey \(2002\)](#) later proposed an adaptive technique grounded in a conservative estimation of  $\pi_0$ , showcasing substantial enhancements in statistical power. A bias-corrected version of Storey's estimator was later proposed by [Biswas et al. \(2022\)](#). [Genovese and Wasserman \(2004\)](#) brought attention to the identifiability challenges inherent in the estimation of  $\pi_0$ , delineating key criteria for achieving identifiability and purity in estimations. In the context of their study, a comprehensive review of existing literature was provided, synthesizing the advancements

in FDR control methodologies. Furthermore, a novel estimator for  $\pi_0$  was proposed, accompanied by a robust confidence interval construction. Notably, [Genovese and Wasserman \(2004\)](#) demonstrated the asymptotic validity of both the proposed point estimator and associated confidence intervals, reinforcing their utility in practical applications. [Efron \(2012\)](#) has addressed the identifiability issue by assuming that the support of the null density contains a portion which is disjoint from the non-null density. Many works have been done under the assumption that the distributions belong to certain parametric models (see, for example [Cohen \(1967\)](#), [Lindsay and Basak \(1993\)](#)). [Jin \(2008\)](#), [Cai and Jin \(2010\)](#) used empirical characteristic functions to estimate the non-null distribution under a semiparametric normal mixture model. In the context of multiple testing [Meinshausen and Bühlmann \(2005\)](#), [Meinshausen and Rice \(2006\)](#), [Celisse and Robin \(2010\)](#) have also suggested different estimators and confidence bounds. [Patra and Sen \(2016\)](#) have investigated the identifiability problem in complete generality and proposed a consistent estimator of  $\pi_0$ .

Although these estimators were initially conceived under the assumption of independence among test statistics (such as p-values) corresponding to each hypothesis, scant literature exists on exploring the extent to which these estimators can withstand dependence. In this chapter, we delve into an investigation of the consistency of these estimators across a spectrum of weakly dependent scenarios.

## 5.2 The model and identifiability

We consider a sample  $X_1, X_2, \dots, X_n$  which comes from a common distribution function  $F$ . In the multiple testing problem with  $n$  null hypotheses  $H_{01}, \dots, H_{0n}$ ,  $X_i$  is the test statistic corresponding to the  $i$ -th testing problem. Let  $F_0$  and  $F_1$  be the distribution under the null hypothesis and alternate hypothesis respectively ( $F_0$  and  $F_1$  does not depend on  $i$ ). Then we consider the mixture model

$$F(x) = (1 - \pi_1)F_0(x) + \pi_1F_1(x) \tag{5.1}$$

Note that, the mixture model is written as a function of proportion of non-null hypotheses ( $\pi_1 = 1 - \pi_0$ ). Usually the null distribution  $F_0(\cdot)$  is known and  $\pi_1, F_1(\cdot)$  are unknown. For example, if we consider p values for the multiple testing problem, then  $F_0$  is the *Unif*[0, 1] distribution if the p values are continuous. If we consider z values as per [Efron \(2008\)](#), then the null distribution is standard normal distribution.

When both  $\pi_1$  and  $F_1$  are unknown, the problem is non-identifiable. If the model [5.1](#) holds then for any  $0 \leq \gamma \leq \pi_0 = 1 - \pi_1$ , one can re-write as

$$F(x) = (\pi_1 + \gamma) \left( \frac{\pi_1}{\pi_1 + \gamma} F_1(x) + \frac{\gamma}{\pi_1 + \gamma} F_0(x) \right) + (1 - \pi_1 - \gamma) F_0(x)$$

So, the expression of  $F(x)$  as a convex combination of the null distribution ( $F_0(x)$ ) and some other distribution is not uniquely defined. [Genovese and Wasserman \(2004\)](#) discussed the identifiability issue when  $F_0$  is *Unif*[0, 1] distribution and  $F$  has a density. Building on this foundation, [Patra and Sen \(2016\)](#) redefine the mixing proportion as follows

$$\underline{\pi}_1 = \inf \left\{ \gamma \in (0, 1] : \left\{ \frac{F - (1 - \gamma)F_0}{\gamma} \right\} \text{ is a CDF} \right\} \quad (5.2)$$

They have further examined this  $\underline{\pi}_1$  and found the connection between  $\underline{\pi}_1$  and  $\pi_1$  satisfying model [5.1](#).

**Theorem 5.2.1.** *Patra and Sen (2016)*

Let  $F$  be as in model 5.1 and  $\underline{\pi}_1$  be defined as 5.2. Then,

$$\underline{\pi}_1 = \pi_1 - \sup\{0 \leq \epsilon \leq 1 : \pi_1 F_1 - \epsilon F_0 \text{ is a sub-CDF}\} \quad (5.3)$$

In particular, if  $F_1$  and  $F_0$  are absolutely continuous, i.e. they have densities  $f_1$  and  $f_0$  respectively, then

$$\underline{\pi}_1 = \pi_1 \left(1 - \operatorname{ess\,inf} \frac{f_1}{f_0}\right) \quad (5.4)$$

where, for any function  $g$ ,  $\operatorname{ess\,inf} g = \sup\{a \in \mathbb{R} : m\{\{x : g(x) < a\}\} = 0\}$ ,  $m$  being the Lebesgue measure.

As a consequence,  $\underline{\pi}_1 < \pi_1$  if and only if there exists  $c > 0$  such that  $f_1 \geq cf_0$ , almost everywhere  $m$ . Note that, in particular, if the support of  $F_1$  is strictly contained in that of  $F_0$ , then the problem is identifiable and we can estimate  $\pi_1$ . In the subsequent sections we will discuss different estimators of  $\pi_1$  and their consistency.

### 5.3 Inconsistency of the BH estimator

The Benjamini Hochberg step up algorithm is based on ordered p values  $P_{(1)} \leq \dots \leq P_{(n)}$  which compares each  $P_{(i)}$  to  $\frac{i\alpha}{n}$ . When the observations are independent, this algorithm controls the FDR at a level  $\pi_0\alpha \leq \alpha$ . Later an adaptive version of BH procedure was proposed by Benjamini and Hochberg (2000) and they proposed a graphical approach of estimating  $n_0$  and hence  $\pi_0 = \frac{n_0}{n}$ . The estimation works as follows

- Calculate the slopes  $S_i = \frac{(1-P_{(i)})}{(n+1-i)}$
- Let,  $j = \inf\{i \geq 2 : S_i < S_{i-1}\}$ . Then set,  $\hat{n}_0 = \min([\frac{1}{S_j} + 1], n)$  and  $\hat{\pi}_0 = \frac{\hat{n}_0}{n}$

One might expect that the estimator  $\hat{\pi}_{0,BH}$  is a consistent estimator of  $\pi_0$  in order to ensure a good performance of the adaptive BH method for large number of hypotheses.

However, the following theorem describes that  $\hat{\pi}_{0,BH}$  cannot be a consistent estimator of  $\pi_0$  as it puts non-zero mass on some point for any  $0 \leq \pi_0 \leq 1$ .

**Theorem 5.3.1.** *Let  $j$ ,  $\hat{n}_0$  and  $\hat{\pi}_{0,BH}$  be defined as per the above mentioned description. Also assume that, the  $p$  values are absolutely continuous and the alternate distribution of the  $p$  values are stochastically smaller than the null distribution ( $Unif[0, 1]$ ). Then,  $\hat{\pi}_{0,BH} \not\rightarrow \pi_0$  for all values of  $\pi_0 \in [0, 1]$  (i.e.  $\hat{\pi}_{0,BH}$  is not a consistent estimator of  $\pi_0$ )*

### 5.3.1 Proof of theorem 5.3.1

By definition,  $j \geq 2$  with probability 1.

$$\begin{aligned} & P(j > 2) \\ &= P(S_2 \geq S_1) \\ &= P\left(\frac{1-p(2)}{n-1} \geq \frac{1-p(1)}{n}\right) \end{aligned}$$

Let,  $q_i = 1 - p_i$  for  $i = 1, \dots, n$ . This implies,  $q_{(n+1-i)} = 1 - p_{(i)}$  for all  $i$ .

$$\text{So, } P(S_2 \geq S_1) = P\left(\frac{q_{(n)}}{q_{(n-1)}} \leq \frac{n}{n-1}\right)$$

Let  $g$  and  $G$  respectively denote the density and the cdf of the  $q_i$ 's. Then, from the joint density of order statistics, one can easily deduce that,

$$\begin{aligned} & P(q_{(n-1)} \geq \left(\frac{n-1}{n}\right)q_{(n)}) \\ &= n(n-1) \int_0^1 \left( \int_{\left(\frac{n-1}{n}\right)y}^y (G(x))^{n-2} g(x) dx \right) g(y) dy \\ &= \int_0^1 n \left( (G(y))^{n-1} - (G\left(\left(\frac{n-1}{n}\right)y\right))^{n-1} \right) g(y) dy \\ &= \int_0^1 \left( 1 - \left( \frac{G\left(\left(\frac{n-1}{n}\right)y\right)}{G(y)} \right)^{n-1} \right) (n(G(y))^{n-1} g(y)) dy \\ &= E \left[ 1 - \left( \frac{G\left(\left(\frac{n-1}{n}\right)q_{(n)}\right)}{G(q_{(n)})} \right)^{n-1} \right] \end{aligned}$$

The last equality holds because of the fact that  $q_{(n)}$  has density  $h(y) = n(G(y))^{n-1} g(y) I_{0 < y < 1}$ .

Now observe that,

$$\begin{aligned}
& n \left[ 1 - \frac{G\left(\left(\frac{n-1}{n}\right)q(n)\right)}{G(q(n))} \right] \\
&= n \left[ \frac{G(q(n)) - G\left(\left(\frac{n-1}{n}\right)q(n)\right)}{G(q(n))} \right] \\
&= \frac{q(n)g(\xi_n)}{G(q(n))} \text{ for some } \left(\frac{n-1}{n}\right)q(n) < \xi_n < q(n)
\end{aligned}$$

Since the alternate density is stochastically smaller than the  $Unif[0, 1]$  distribution,  $G(1) = 1$  and  $q(n) \xrightarrow{a.s.} 1$  as  $n \rightarrow \infty$ .

So, by the continuity of  $G(\cdot)$ , we can write  $G(q(n)) \xrightarrow{a.s.} G(1) = 1$  as  $n \rightarrow \infty$ .

And  $g(\xi_n) \rightarrow g(1-)$  as  $n \rightarrow \infty$ .

Thus,

$$n \left[ 1 - \frac{G\left(\left(\frac{n-1}{n}\right)q(n)\right)}{G(q(n))} \right] \rightarrow g(1-) \text{ as } n \rightarrow \infty$$

Here,  $g(1-) = \lim_{x \uparrow 1} g(x)$ .

And hence,

$$\lim_{n \rightarrow \infty} P(j > 2) = \lim_{n \rightarrow \infty} E \left[ 1 - \left( \frac{G\left(\left(\frac{n-1}{n}\right)q(n)\right)}{G(q(n))} \right)^{n-1} \right] = 1 - \exp(-g(1-))$$

**Case 1** :  $g(1-) < \infty$  and  $1 - \exp(-g(1-)) < 1$

Since  $P(j \geq 2) = 1$  and  $\lim_{n \rightarrow \infty} P(j > 2) < 1$ , this implies,

$$\lim_{n \rightarrow \infty} P(j = 2) = \exp(-g(1-)) > 0$$

And thus,

$$\lim_{n \rightarrow \infty} P(\hat{n}_0 = \frac{1}{S_2} + 1) \geq \exp(-g(1-)) > 0$$

Since  $S_2 = \frac{1-P(2)}{n-1}$ , this implies,

$$\lim_{n \rightarrow \infty} P(\hat{n}_0 = n) > 0$$

and hence

$$\lim_{n \rightarrow \infty} P(\hat{\pi}_{0,BH} = 1) > 0$$

So, even if the true value of  $\pi_0$  is strictly less than 1, the estimator asymptotically puts a positive mass at 1 and hence, this estimator is not a consistent estimator for  $\pi_0$ .

**Case 2** :  $g(1-) = \infty$  and  $1 - \exp(-g(1-)) = 1$

For  $i = 0, \dots, n-2$ , define

$$A_i = \left\{ \frac{q_{(n-i)}}{n-i} \leq \frac{q_{(n-i-1)}}{n-i-1} \right\}$$

Then,

$$P(j > k) = P(S_k \geq S_{k-1} \geq \dots \geq S_1) = P\left(\bigcap_{i=0}^{k-2} A_i\right)$$

We have already shown that  $\lim_{n \rightarrow \infty} P(A_0) = 1 - \exp(-g(1-)) = 1$ .

From the the joint density of  $(q_{n-i}, q_{n-i-1})$ , it can be deduced that,

$$\begin{aligned} & P\left(\frac{q_{(n-i)}}{n-i} \leq \frac{q_{(n-i-1)}}{n-i-1}\right) \\ &= \frac{n!}{(n-i-2)!i!} \int_0^1 \int_{\left(\frac{n-i-1}{n-i}\right)y}^y (G(x))^{i-2} (1-G(y))^i g(x)g(y) dx dy \\ &= \frac{n!}{(n-i-1)!i!} \int_0^1 (1-G(y))^i \left( (G(y))^{n-i-1} - \left(G\left(\left(\frac{n-i-1}{n-i}\right)y\right)\right)^{n-i-1} \right) g(y) dy \\ &= \frac{n!}{(n-i-1)!i!} \int_0^1 (G(y))^{n-i-1} (1-G(y))^i \left( 1 - \left(\frac{G\left(\left(\frac{n-i-1}{n-i}\right)y\right)}{G(y)}\right)^{n-i-1} \right) g(y) dy \\ &= E \left[ 1 - \left( \frac{G\left(\frac{n-i-1}{n-i} q_{(n-i)}\right)}{G(q_{(n-i)})} \right)^{n-i-1} \right] \end{aligned}$$

Note that, if  $i = o(n)$  then  $q_{(n-i)} \xrightarrow{a.s.} 1$  and  $G(q_{(n-i)}) \xrightarrow{a.s.} 1$  as  $n \rightarrow \infty$ . Since  $n \left( 1 - \frac{G\left(\frac{n-i-1}{n-i} q_{(n-i)}\right)}{G(q_{(n-i)})} \right) \xrightarrow{a.s.} g(1-) = \infty$ , this implies,  $\lim_{n \rightarrow \infty} P(A_i) = 1$ .

If  $\lim_{n \rightarrow \infty} \frac{i}{n} = \delta \in (0, 1)$ , then if  $z_\delta$  is the  $(1 - \delta)$ -th quantile of the distribution of  $q$ 's, then  $q_{(n-i)} \xrightarrow{a.s.} z_\delta$  as  $n \rightarrow \infty$ .

Again, by continuity theorem,  $G(q_{(n-i)}) \xrightarrow{a.s.} (1 - \delta)$  as  $n \rightarrow \infty$  and,

$$n \left( 1 - \frac{G\left(\frac{n-i-1}{n-i}q_{(n-i)}\right)}{G(q_{(n-i)})} \right) = \frac{z_\delta g(z_\delta -)}{(1 - \delta)^2}$$

And thus,

$$\lim_{n \rightarrow \infty} E \left[ 1 - \left( \frac{G\left(\frac{n-i-1}{n-i}q_{(n-i)}\right)}{G(q_{(n-i)})} \right)^{n-i-1} \right] = 1 - \exp\left(-\frac{z_\delta g(z_\delta -)}{(1 - \delta)^2}\right)$$

Since this is true for all  $0 < \delta < 1$ , we can surely find some  $0 < \delta_0 < 1$  for which  $1 - \exp\left(-\frac{z_\delta g(z_\delta -)}{(1 - \delta)^2}\right) < 1$  and for that  $\delta_0$ , we'll have

$$\lim_{n \rightarrow \infty} P(\hat{\pi}_{0,BH} = \delta_0) > 0$$

Existence of such  $\delta_0$  will not allow  $\hat{\pi}_{0,BH}$  to have a degenerate limit and hence it is not a consistent estimator of  $\pi_0$ .

Theorem 5.3.1 demonstrates that the BH estimator lacks consistency for  $\pi_0$ . Despite this limitation, it often serves as a conservative estimator of  $\pi_0$  (i.e.,  $\hat{\pi}_{0,BH} \geq \pi_0$ ). Consequently, the adaptive BH procedure typically maintains validity for FDR control. The following theorem illustrates this fact for the global null.

**Theorem 5.3.2.** *Under the global null (i.e.  $\pi_0 = 1$ ), the BH estimator of  $\pi_0$  satisfies*

$$\hat{\pi}_{0,BH} \xrightarrow{a.s.} 1 \text{ as } n \rightarrow \infty$$

**Proof :** If  $\pi_0 = 1$ , then all P values come from  $Unif[0, 1]$  distribution.

In this case,  $q_1, \dots, q_n \stackrel{i.i.d.}{\sim} Unif[0, 1]$  where  $q_i = 1 - P_i$  for  $i = 1, \dots, n$ .

Observe that, for  $k = 2, \dots, n$

$$\begin{aligned} & P(j > k) \\ &= P(S_k \geq S_{k-1} \geq \dots \geq S_2 \geq S_1) \\ &= P\left(\frac{q_{(n+1-k)}}{n+1-k} \geq \frac{q_{(n+2-k)}}{n+2-k} \geq \dots \geq \frac{q_{(n)}}{n}\right) \end{aligned}$$

$$\begin{aligned}
&= n! \int_0^1 \int_{\binom{n-1}{n}q(n)}^{q(n)} \cdots \int_{\binom{n+1-k}{n+2-k}q(n+2-k)}^{q(n+2-k)} \left( \int_0^{q(n+1-k)} \cdots \int_0^{q(2)} dq_{(1)} \cdots dq_{(n-k)} \right) dq_{(n+1-k)} \cdots dq_{(n-1)} dq_{(n)} \\
&= \prod_{j=n+1-k}^{n-1} \left( 1 - \left( \frac{j}{j+1} \right)^j \right) \\
&= \prod_{m=1}^{k-1} \left( 1 - \left( \frac{n-m}{n-m+1} \right)^{n-m} \right) \\
&\leq (1 - e^{-1})^{k-1}
\end{aligned}$$

If we choose a sequence  $\{x_n\}$  which is  $o(n)$ , and  $\sum_{n=2}^{\infty} (1 - e^{-1})^{x_n-1} < \infty$ , then by first Borel Cantelli lemma, we can conclude that,  $P(j \leq x_n \text{ for all but finitely many } n) = 1$ . One such example might be  $x_n = n^{0.9}$ .

So,  $j = o(n)$  almost surely. Note that, for any  $j$  with  $j = o(n)$ ,  $\frac{n+1-j}{n(1-P_{(j)})} \geq 1$  for all but finitely many  $n$ , with probability 1. Hence,

$$P(\hat{\pi}_{0,BH} = 1 \text{ for all but finitely many } n) = 1$$

This holds true not only for the global null hypothesis but also for the case where  $\lim_{n \rightarrow \infty} \pi_0 = 1$ , which is encountered most often in real-life applications. To illustrate this, consider that  $n_0$  of the  $q_i$ 's are derived from a uniform distribution  $Unif[0, 1]$ , while the remaining originate from some alternate distribution. Let  $q_{(1)}^0 \leq \cdots \leq q_{(n_0)}^0$  denote the order statistics among the sub-sample of null p-values. Let  $j_0$  represent the index among the null p-values where  $\frac{q_{(n_0+1-k)}^0}{n_0+1-k}$  first decreases. Then, in the combined sample, it is in the  $(n_0 + 1 - j_0 + x_1)$ -th place where  $x_1$  is the no. of non-null p values lying above that place. Clearly

$$(n_0 + 1 - j_0 + x_1) \leq n + 1 - j$$

This implies,

$$j \leq j_0 + (n_1 - x_1)$$

By theorem 5.3.2,  $j_0 = o(n_0) = o(n)$  and since  $\lim_{n \rightarrow \infty} \pi_0 = 1$ . this implies,  $(n_1 - x_1) = o(n)$ . Hence,  $j = o(n)$  and the conclusion of the theorem 5.3.2 remains valid.

Now we ask the question, what is the behaviour of  $\hat{\pi}_{0,BH}$  when  $0 < \pi_0 < 1$ .

### 5.3.2 Behaviour of $\hat{\pi}_{0,BH}$ for $0 < \pi_0 < 1$

**Theorem 5.3.3.** For any  $0 < \pi_0 < 1$ ,

$$\hat{\pi}_{0,BH} \xrightarrow{a.s.} 1 \text{ as } n \rightarrow \infty$$

**Proof :** Let  $h(\cdot), H(\cdot)$  denote the distribution of  $q_i = 1 - P_i$  for  $i = 1, \dots, n$ . Then, the joint density of  $q_{(1)} \leq q_{(2)} \leq \dots \leq q_{(n)}$  is given by  $h_n(x) = n! \left( \prod_{i=1}^n h(x_i) \right)$ .

Now, the proof proceeds in the similar manner as theorem 5.3.2. Again, for  $k = 2, \dots, n$

$$\begin{aligned} & P(j > k) \\ &= P(S_k \geq S_{k-1} \geq \dots \geq S_2 \geq S_1) \\ &= P\left(\frac{q_{(n+1-k)}}{n+1-k} \geq \frac{q_{(n+2-k)}}{n+2-k} \geq \frac{q_{(n)}}{n}\right) \\ &= n! \int_0^1 \int_{\left(\frac{n-1}{n}\right)q_{(n)}}^{q_{(n)}} \dots \int_{\left(\frac{n+1-k}{n+2-k}\right)q_{(n+2-k)}}^{q_{(n+2-k)}} \left( \int_0^{q_{(n+1-k)}} \dots \int_0^{q_{(2)}} \left( \prod_{i=1}^n h(x_i) \right) dq_{(1)} \dots dq_{(n-k)} \right) dq_{(n+1-k)} \dots dq_{(n-1)} dq_{(n)} \\ &= \prod_{j=n+1-k}^{n-1} \left( 1 - \left( \frac{j}{j+1} \right)^j \right) \text{ This is because } H(0) = 0 \text{ and } H(1) = 1 \\ &= \prod_{m=1}^{k-1} \left( 1 - \left( \frac{n-m}{n-m+1} \right)^{n-m} \right) \end{aligned}$$

It is very important to observe that, the quantity  $j$  is actually a distribution-free quantity.

So, we can again make a conclusion similar to theorem 5.3.2 that,  $j = o(n)$  and consequently

$$\hat{\pi}_{0,BH} \xrightarrow{a.s.} 1 \text{ as } n \rightarrow \infty$$

However, the rate of convergence is very slow and depends on the value of  $\pi_0$ . We have considered testing  $H_{0i} : \mu_i = 0$  vs  $H_{1i} : \mu_i > 0$  in the context of Gaussian distribution with unit variance. Our approach involved drawing  $n = 10000$  observations from a blend of  $N(0,1)$  and  $N(2,1)$  distributions, and then deriving  $\hat{\pi}_{0,BH}$  based on the computed p-values. Repeating this process 10000 times allowed us to scrutinize the distribution of  $\hat{\pi}_{0,BH}$ . The ensuing table showcases the mean and minimum values of  $\hat{\pi}_{0,BH}$  across these replications. Notably, for all instances where  $0 < \pi_0 < 1$ , the minimum estimators notably surpass the true  $\pi_0$  value. It indicates that the adaptive BH procedure remains valid using the

plug-in estimator.

**Table 5.1:**  $\hat{\pi}_{0,BH}$  for Gaussian mean test for  $n = 10^4$

$\pi_0$	$E[\hat{\pi}_{0,BH}]$	Median $\hat{\pi}_{0,BH}$	min $\hat{\pi}_{0,BH}$	Sd
0.5	0.76	0.76	0.69	0.021
0.75	0.91	0.91	0.88	0.011
0.8	0.94	0.94	0.91	0.008
0.9	0.98	0.98	0.96	0.004
0.95	0.99	0.99	0.98	0.002
1	1	0.998	1	6.61E-05

The table above illustrates that the value of  $\hat{\pi}_{0,BH}$  significantly exceeds  $\pi_0$  for  $n = 10000$ , confirming that the adaptive Benjamini-Hochberg (BH) procedure remains valid for practical applications. However, the primary result,  $\hat{\pi}_{0,BH} \xrightarrow{a.s.} 1$  as  $n \rightarrow \infty$ , is not evident even for  $n = 10000$  due to the slow rate of convergence. Additionally, we conducted similar simulation exercises for a larger number of hypotheses ( $n = 10^7, 10^9$ ). The subsequent table demonstrates that the BH estimator gradually converges to 1 as  $n$  increases. The convergence rate is particularly slow when  $\pi_0$  is significantly less than 1.

**Table 5.2:**  $\hat{\pi}_{0,BH}$  for Gaussian mean test for  $n = 10^7$

$\pi_0$	$E[\hat{\pi}_{0,BH}]$	Median( $\hat{\pi}_{0,BH}$ )	Min( $\hat{\pi}_{0,BH}$ )	Sd
0.5	0.8428	0.8417	0.8213	0.0093
0.75	0.9494	0.9491	0.9392	0.0037
0.8	0.9648	0.9646	0.9585	0.0027
0.9	0.9891	0.9890	0.9866	0.0011
0.95	0.9968	0.9968	0.9958	0.0004
1	1	1	0.9999	7.70E-08

### 5.3.3 Behaviour of $\hat{\pi}_{0,BH}$ under dependence

Our simulations have primarily focused on independent P-values. However, we have extended our analysis to explore the behavior of  $\hat{\pi}_{0,BH}$  under various forms of dependence. Initially, we investigate the m-dependent scenario, where  $X_i$  and  $X_j$  are independent if

$|i - j| > m$ . In this setup, we generate observations  $X_i$  from a set of independent normal variables  $\xi_1, \xi_2, \dots$ , following the formula  $X_i = \frac{1}{\sqrt{m+1}} \sum_{j=i}^{i+m} \xi_j$ .

Within this framework,  $n_1$  of the  $X_i$ 's are  $N(\mu, 1)$  and  $n_0$  of them are  $N(0, 1)$  random variables. For our simulations, we fix  $n = 1000$  and explore six combinations of  $\pi_0$ , consistent with our previous investigations. Additionally, we vary the parameter  $m$  with values of 2 and 5.

**Table 5.3:**  $\hat{\pi}_{0,BH}$  for Gaussian mean test for  $m = 2$  and  $n = 1000$

$\pi_0$	$E[\hat{\pi}_{0,BH}]$	Median( $\hat{\pi}_{0,BH}$ )	Min( $\hat{\pi}_{0,BH}$ )	Sd
0.5	0.692	0.688	0.600	0.034
0.75	0.903	0.902	0.846	0.019
0.8	0.935	0.934	0.888	0.016
0.9	0.983	0.984	0.957	0.007
0.95	0.997	0.998	0.983	0.003
1	1	1	1	0

**Table 5.4:**  $\hat{\pi}_{0,BH}$  for Gaussian mean test for  $m = 5$  and  $n = 1000$

$\pi_0$	$E[\hat{\pi}_{0,BH}]$	Median( $\hat{\pi}_{0,BH}$ )	Min( $\hat{\pi}_{0,BH}$ )	Sd
0.5	0.621	0.614	0.574	0.038
0.75	0.896	0.882	0.826	0.044
0.8	0.943	0.934	0.875	0.035
0.9	0.996	1.000	0.959	0.007
0.95	1.000	1.000	0.989	0.001
1	1	1	1	0

It is interesting to observe that, in all cases, the value of  $\hat{\pi}_{0,BH}$  lies well above the actual value. We have further simulated the same for  $n = 10000$  and the observations are presented below.

**Table 5.5:**  $\hat{\pi}_{0,BH}$  for Gaussian mean test for  $m = 2$  and  $n = 10000$ 

$\pi_0$	$E[\hat{\pi}_{0,BH}]$	Median( $\hat{\pi}_{0,BH}$ )	Min( $\hat{\pi}_{0,BH}$ )	Sd
0.5	0.746	0.742	0.681	0.027
0.75	0.931	0.930	0.896	0.013
0.8	0.956	0.955	0.926	0.010
0.9	0.991	0.991	0.978	0.004
0.95	0.999	0.999	0.993	0.001
1	1	1	1	0

**Table 5.6:**  $\hat{\pi}_{0,BH}$  for Gaussian mean test for  $m = 5$  and  $n = 10000$ 

$\pi_0$	$E[\hat{\pi}_{0,BH}]$	Median( $\hat{\pi}_{0,BH}$ )	Min( $\hat{\pi}_{0,BH}$ )	Sd
0.5	0.878	1	0.608	0.163
0.75	0.998	1	0.880	0.011
0.8	0.999	1	0.935	0.002
0.9	1	1	0.999	2.04E-05
0.95	1	1	1	0
1	1	1	1	0

As before, the values of  $\hat{\pi}_{0,BH}$  slowly increases as we increase  $n$ . So, we might expect similar result for m-dependent setup also.

Now, we explore block dependent structure. First, we have considered the problem of Gaussian mean test with 10 blocks of equicorrelated normal variables with correlations  $0.1, \dots, 0.9, 0.95$  and each block consists of 100 observations. The observations are presented in the following table. Although the mean and median of the 10000 replications

**Table 5.7:**  $\hat{\pi}_{0,BH}$  for Gaussian mean test under block dependence for  $n = 1000$ 

$\pi_0$	$E[\hat{\pi}_{0,BH}]$	Median( $\hat{\pi}_{0,BH}$ )	Min( $\hat{\pi}_{0,BH}$ )	Sd
0.5	0.713	0.712	0.349	0.082
0.75	0.886	0.891	0.547	0.051
0.8	0.915	0.922	0.635	0.044
0.9	0.965	0.972	0.690	0.030
0.95	0.985	0.991	0.745	0.023
1	0.996	1	0.765	0.018

lie well above the original value of  $\pi_0$ , the minimum observed from the 10000 replications is not above  $\pi_0$  anymore. This is possibly due to the increased level of dependence among the hypotheses. However, this gap appears to narrow progressively and the estimator increases slowly as we continually increase  $n$  from 1000 to 10000.

**Table 5.8:**  $\hat{\pi}_{0,BH}$  for Gaussian mean test under block dependence for  $n = 10000$ 

$\pi_0$	$E[\hat{\pi}_{0,BH}]$	Median( $\hat{\pi}_{0,BH}$ )	Min( $\hat{\pi}_{0,BH}$ )	Sd
0.5	0.761	0.761	0.480	0.076
0.75	0.911	0.915	0.624	0.042
0.8	0.935	0.940	0.688	0.035
0.9	0.975	0.979	0.717	0.022
0.95	0.990	0.993	0.773	0.016
1	0.998	1	0.762	0.011

## 5.4 Estimator proposed in Storey (2002)

Since the largest p-values are most likely to be uniformly distributed, Storey (2002) suggested that a conservative estimator of  $\pi_0$  is

$$\hat{\pi}_0(\lambda) = \frac{\#\{P_i > \lambda\}}{n(1 - \lambda)} = \frac{W(\lambda)}{n(1 - \lambda)} \quad (5.5)$$

for some well chosen  $\lambda$ . Since  $Pr(P_i > \lambda) = \pi_0(1 - \lambda) + \pi_1(1 - F_1(\lambda))$ , by strong law of large numbers, one can easily deduce that for any  $0 < \lambda < 1$  and for independent p values,

$$\hat{\pi}_0(\lambda) \xrightarrow{a.s.} \pi_0 + \pi_1 \left( \frac{1 - F_1(\lambda)}{1 - \lambda} \right) \text{ as } n \rightarrow \infty$$

So, the estimator is asymptotically a conservative estimator of  $\pi_0$ . Storey (2002) proposed two methods for controlling pFDR and FDR and showcased the gain in power and accuracy using the estimator  $\hat{\pi}_0(\lambda)$ . While the validity of the results were initially established under the assumption of independent p-values, it can be shown that these findings remain applicable even under conditions of weak dependence. For p values with empirical cdf and cdf  $F_n$  and  $F$  respectively, if  $\|F_n - F\|_\infty \xrightarrow{a.s.} 0$  as  $n \rightarrow \infty$ , then all theorems of Storey (2002) remains valid.

Under various forms of weak dependence among p-values, the condition  $\|F_n - F\|_\infty \xrightarrow{a.s.} 0$  persists. By invoking the Ergodic theorem as presented in Billingsley (2017), it becomes apparent that if the right shift transformation of the sequence of p-values is measure-preserving (i.e.,  $(P_1, P_2, P_3, \dots) \stackrel{d}{=} (P_2, P_3, \dots)$ ), then the empirical cumulative distribution function (ecdf)  $F_n(x)$  converges pointwise almost surely to some limit as  $n \rightarrow \infty$ . The condition

of measure preservation in the right shift transformation is synonymous with stating that the sequence of p-values is strongly stationary. The stationarity of the p-values guarantees the existence of an almost sure limit, which may or may not be degenerate.

To ensure that the pointwise limit equals to the cdf  $F(x)$ , it suffices to assume that  $F_n(x) \xrightarrow{P} F(x) \forall x$ . Convergence in probability implies that a subsequence of the original sequence is almost surely convergent. Given the original sequence's almost sure convergence, all subsequences should converge to the same limit, thereby leading to  $F_n(x)$  converging almost surely to  $F(x)$  for all  $x$ . Consequently, according to the Glivenko-Cantelli lemma, it can be asserted that  $\|F_n - F\|_\infty \xrightarrow{a.s.} 0$  as  $n \rightarrow \infty$ . These discussions are formally encapsulated into the theorem presented below.

**Theorem 5.4.1.** *Let  $P_1, P_2, \dots, P_n$  be a sequence of p values in a multiple testing problem with empirical cdf and continuous cdf  $F_n, F$  respectively. We assume the following about the joint distribution of the p values.*

I )  $P_1, P_2, \dots$  forms a strongly stationary sequence.

II )  $F_n(x) \xrightarrow{P} F(x) \forall x$  as  $n \rightarrow \infty$ .

Then,  $\|F_n - F\|_\infty \xrightarrow{a.s.} 0$  as  $n \rightarrow \infty$

The assumption (II) outlined in theorem 5.4.1 holds over a wide array of dependence structures commonly encountered in practical contexts. We highlight a few such cases where the assumptions are met.

- m-dependent and stationary p values satisfy assumption (I) and (II)
- Hidden Markov model similar to [Sun and Tony Cai \(2009\)](#). In particular. if the sequence of P values form a stationary, Ergodic Markov chain, then the assumptions (I) and (II) are met.
- All Ergodic, stationary sequence of p values satisfy assumption (I) and (II). This is because, ergodicity ensures that the limit of  $F_n(\cdot)$  is indeed the distribution function. In practice, we often assume a sequence of random variables to be asymptotically

independent. One such example of asymptotic independence is mixing sequence (Billingsley (2017)). Since all mixing sequences are ergodic (under the right shift operation), mixing sequences also satisfy assumption (I) and (II).

- The assumption of asymptotic independence can be somewhat relaxed as follows.

$$\forall x, \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i \neq j} (Pr(P_i \leq x, P_j \leq x) - (F(x))^2) = 0$$

By Chebyshev's inequality, it can be easily argued that, under the above mentioned assumption,  $F_n(x) \xrightarrow{P} F(x) \forall x$  as  $n \rightarrow \infty$ . In fact, it constitutes a necessary and sufficient criterion for assumption (II) to hold.

### 5.4.1 Simulation results

The estimator in Storey (2003) involves a tuning parameter,  $\lambda$ . Choosing  $\lambda$  requires balancing bias and variance for the estimator  $\hat{\pi}_0(\lambda)$ . According to Storey (2003), for well-behaved p-values, the bias tends to decrease as  $\lambda$  increases, reaching its minimum as  $\lambda$  approaches 1. Consequently, Storey (2003) proposed a method for selecting the optimal  $\lambda$ .

- Fix a range of  $\lambda$ 's by  $\Lambda$  (e.g.  $\Lambda = \{0, 0.01, 0.02, \dots, 0.95\}$ )
- By 5.5, compute  $\hat{\pi}_0(\lambda)$  for each  $\lambda \in \Lambda$ .
- Fit a cubic spline ( $\hat{f}$ ) on  $\hat{\pi}_0(\lambda)$  and obtain the final estimate as

$$\hat{\pi}_0 = \min\{\hat{f}(1), 1\}$$

In addition to the smoothing-based technique introduced by Storey (2003), a bootstrap-based method for selecting the optimal value of  $\lambda$  was proposed in Storey et al. (2004). This method builds on the earlier work of Storey (2002). The proposed automatic choice of  $\lambda$  aims to estimate the value that minimizes the mean squared error (MSE), balancing bias

and variance. Specifically, it seeks to minimize  $E[(\hat{\pi}_0(\lambda) - \pi_0)^2]$ . The procedure for this method is summarized below.

- For each  $\lambda \in \Lambda$ , compute  $\hat{\pi}_0(\lambda)$  as [5.5](#).
- Generate  $B$  bootstrap samples from the p values. For each  $b = 1, \dots, B$ , find the estimators  $\{\hat{\pi}_0^{*b}(\lambda)\}_{\lambda \in \Lambda}$  based on the  $b$ -th sample.
- Since  $E[\hat{\pi}_0(\lambda)] \geq p_{i_0}$  for all  $\lambda \in [0, 1)$ , a plug-in estimator of  $\pi_0$  can be taken as

$$\hat{\pi}_0 = \min_{\lambda' \in \Lambda} \{\hat{\pi}_0(\lambda')\}$$

- for each  $\lambda \in \Lambda$ , estimate its respective MSE as

$$M\hat{S}E(\lambda) = \frac{1}{B} \sum_{b=1}^B \left[ \hat{\pi}_0^{*b}(\lambda) - \min_{\lambda' \in \Lambda} \{\hat{\pi}_0(\lambda')\} \right]^2$$

- Set the final estimator as

$$\hat{\pi}_0 = \min \left\{ 1, \operatorname{argmin}_{\lambda \in \Lambda} M\hat{S}E(\lambda) \right\}$$

Based on these two methods, we examined the performance of Storey's estimator under various dependence structures. Initially, we focused on the autoregressive model of order 1 (AR(1)), where observations are generated according to the mechanism described below.

$$X_i = aX_{i-1} + \xi_i \quad \text{for } i > 1 \text{ and } X_1 = \xi_1 \quad (5.6)$$

$\xi_i$ 's are standard normal random variables. Theorem 19.1 of [Billingsley \(2013\)](#) ensures that the AR(1) process meets the necessary criteria for the validity of Storey's estimator. In our analysis, we set  $n = 1000$  and considered values of  $a$  as 0.1, 0.2, 0.5, 0.75. We also examined 10 different combinations of  $\pi_0$ , specifically  $\{0.5, 0.6, 0.65, 0.7, \dots, 0.95, 0.98\}$ . Observations were generated from a mixture of  $N(0, 1)$  and  $N(2, 1)$  random variables,

and p-values were computed for a one-sided test of the mean.

For values of  $a = 0.1$  and  $0.2$ , both the bootstrap and spline-based estimators consistently yielded results above the true values of  $\pi_0$ . There was minimal difference between these two methods in terms of selecting the optimal  $\lambda$ .

**Table 5.9:**  $n = 1000$  and  $a = 0.1, 0.2$

	$a = 0.1$		$a = 0.2$	
$\pi_0$	$\hat{\pi}_0$ (Bootstrap)	$\hat{\pi}_0$ (Smoother)	$\hat{\pi}_0$ (Bootstrap)	$\hat{\pi}_0$ (Smoother)
0.5	0.501	0.502	0.522	0.522
0.6	0.604	0.605	0.628	0.627
0.65	0.655	0.655	0.681	0.680
0.7	0.707	0.707	0.734	0.735
0.75	0.756	0.757	0.785	0.786
0.8	0.809	0.808	0.838	0.839
0.85	0.858	0.858	0.892	0.892
0.9	0.911	0.908	0.945	0.944
0.95	0.959	0.959	0.987	0.985
0.98	0.984	0.984	0.997	0.997

It is interesting to note that as the degree of dependence increases, Storey's estimator tends to become more conservative. Specifically, as we move from  $a = 0.1$  and  $a = 0.2$  to  $a = 0.5$  and  $a = 0.75$ , the estimated values approach 1 for all values of  $\pi_0$ .

**Table 5.10:**  $n = 1000$  and  $a = 0.5, 0.75$

	$a = 0.5$		$a = 0.75$	
$\pi_0$	$\hat{\pi}_0$ (Bootstrap)	$\hat{\pi}_0$ (Smoother)	$\hat{\pi}_0$ (Bootstrap)	$\hat{\pi}_0$ (Smoother)
0.5	0.692	0.691	0.977	0.977
0.6	0.827	0.828	0.998	0.996
0.65	0.893	0.895	0.998	0.997
0.7	0.962	0.961	0.999	1
0.75	0.995	0.995	1	1
0.8	1	1	1	1
0.85	1	1	1	1
0.9	1	1	1	1
0.95	1	1	1	1
0.98	1	1	1	1

In the AR(1) model, Storey's estimator exhibits increased bias as the correlation strengthens. Additionally, we evaluated the estimator for a stationary  $m$ -dependent sequence of p-values. In this case, we generated p-values from a mixture of normal distributions using the following approach:

Let  $\xi_1, \xi_2, \dots \stackrel{i.i.d.}{\sim} N(0, 1)$  and  $X_i \propto \sum_{j=i}^{i+m} ((m+1) - (j-i)) \xi_j$ .  $X_i$ 's are standardized so that, they have unit variance. In this case,  $\{X_i\}_{i \geq 1}$  forms a  $m$ -dependent sequence of random variables with the nearer observations getting higher weight. Next, we add 2 to  $n(1 - \pi_0)$  of the  $X_i$  values so that they form a mixture of  $N(0, 1)$  and  $N(2, 1)$  random variables. When computing Storey's estimator based on these p-values, the estimates tend to be very close to 1, indicating a higher level of dependence.

**Table 5.11:**  $n = 1000$  and  $m = 1, 2$

$p_{i_0}$	<b>m=1</b>		<b>m=2</b>	
	$\hat{\pi}_0$ (Bootstrap)	$\hat{\pi}_0$ (Smoother)	$\hat{\pi}_0$ (Bootstrap)	$\hat{\pi}_0$ (Smoother)
0.5	0.915	0.915	0.999	0.999
0.6	0.993	0.992	1	1
0.65	0.999	0.999	1	1
0.7	0.999	0.999	1	1
0.75	1	1	1	1
0.8	1	1	1	1
0.85	1	1	1	1
0.9	1	1	1	1
0.95	1	1	1	1
0.98	1	1	1	1

And if we move further to  $m = 5$ , then the estimator is identically 1 for all choices of  $\pi_0$ . The simulation results vary slightly in the  $m$ -dependent case, where observations are generated as  $X_i = \frac{1}{\sqrt{m+1}} \sum_{j=i}^{i+m} \xi_j$ . Unlike scenarios with decreasing weights, this setup uses equal weights for the  $m$  neighboring observations. The heightened dependence in this case results in slower convergence. Consequently, Storey's estimator is much closer to the true value of  $\pi_0$  compared to the previous case, where estimated values were near 1. Although Storey's estimator is typically biased and tends to overestimate  $\pi_0$ , this effect is less pronounced here due to the increased dependence among observations.

**Table 5.12:**  $n = 1000$  and  $m = 1, 2, 5$  for the equally weighted case

	<b>m=1</b>		<b>m=2</b>		<b>m=5</b>	
$\pi_0$	$\hat{\pi}_0$ (Bootstrap)	$\hat{\pi}_0$ (Smoother)	$\hat{\pi}_0$ (Bootstrap)	$\hat{\pi}_0$ (Smoother)	$\hat{\pi}_0$ (Bootstrap)	$\hat{\pi}_0$ (Smoother)
0.5	0.494	0.497	0.496	0.497	0.501	0.494
0.6	0.598	0.599	0.600	0.597	0.591	0.597
0.65	0.642	0.650	0.647	0.647	0.649	0.644
0.7	0.694	0.700	0.699	0.704	0.687	0.701
0.75	0.755	0.749	0.742	0.744	0.747	0.753
0.8	0.800	0.791	0.797	0.797	0.795	0.794
0.85	0.844	0.846	0.839	0.841	0.820	0.837
0.9	0.893	0.887	0.884	0.880	0.872	0.866
0.95	0.927	0.922	0.919	0.918	0.906	0.897
0.98	0.945	0.946	0.936	0.934	0.911	0.912

Due to the slight underestimation of  $\pi_0$ , the validity of the estimator seems to be questionable for  $n = 1000$ . However, as  $n$  increases, the difference appears to diminish. The bootstrap-based estimator exhibits a particularly interesting property in this context: it seems to converge to the true value of  $\pi_0$ , as evidenced by the simulation results for  $n = 10^6$ . For all values of  $m$  (i.e.,  $m = 1, 2, 5$ ), the estimated values, rounded to three decimal places, show identical figures, as is apparent from Table 5.13.

**Table 5.13:**  $n = 10^6$  and  $m = 1, 2, 5$  for the equally weighted case

	<b>m=1</b>	<b>m=2</b>	<b>m=5</b>
$\pi_0$	$\hat{\pi}_0$ (Bootstrap)	$\hat{\pi}_0$ (Bootstrap)	$\hat{\pi}_0$ (Bootstrap)
0.5	0.496	0.496	0.496
0.6	0.597	0.597	0.597
0.65	0.647	0.647	0.647
0.7	0.698	0.698	0.698
0.75	0.748	0.748	0.748
0.8	0.798	0.798	0.798
0.85	0.849	0.849	0.849
0.9	0.899	0.899	0.899
0.95	0.950	0.950	0.950
0.98	0.980	0.980	0.980

A block-dependent sequence does not meet the stationarity requirements necessary for establishing the validity of the estimator via the Strong Law of Large Numbers (SLLN). We have examined a scenario with  $n = 1000$ , where the data is divided into 10

equal-sized blocks of equicorrelated standard normal random variables with correlations 0.1, 0.2,  $\dots$ , 0.9, 0.95.

Our analysis indicates that the estimates of  $\pi_0$  are quite accurate for smaller values of  $\pi_0$ . However, as  $\pi_0$  increases, particularly for values above 0.9, the discrepancy between the estimated values and the true value of  $\pi_0$  becomes more pronounced.

**Table 5.14:** Block dependent variables with  $n = 1000$  and 10 blocks

$\pi_0$	0.5	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	0.98
$\hat{\pi}_0$ ( <b>Bootstrap</b> )	0.497	0.596	0.649	0.698	0.745	0.780	0.827	0.856	0.884	0.899
$\hat{\pi}_0$ ( <b>Smoother</b> )	0.497	0.602	0.644	0.693	0.737	0.785	0.827	0.871	0.886	0.909

Even with a slight reduction in dependence by considering 200 blocks of equicorrelated variables, each of size 5, the overall picture remains largely unchanged. In this scenario, the discrepancy between the estimated and true values of  $\pi_0$  still becomes more pronounced for higher values of  $\pi_0$ .

**Table 5.15:** Block dependent variables with  $n = 1000$  and 200 blocks

$\pi_0$	0.5	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	0.98
$\hat{\pi}_0$ ( <b>Bootstrap</b> )	0.492	0.590	0.649	0.692	0.746	0.782	0.822	0.863	0.891	0.909
$\hat{\pi}_0$ ( <b>Smoother</b> )	0.501	0.590	0.645	0.698	0.742	0.786	0.829	0.866	0.886	0.911

## 5.5 Estimator proposed in *Patra and Sen (2016)*

If  $\pi_1 = \gamma \in (0, 1]$  were known, for an independent and identically distributed sample  $X_1, X_2, \dots, X_n$  with distribution function  $F$ , a naive estimator of  $F_1$  would be

$$\hat{F}_{1,n}^\gamma = \frac{F_n - (1 - \gamma)F_0}{\gamma}$$

where  $F_n$  is the empirical CDF of the observed sample. Since  $\hat{F}_{1,n}^\gamma$  need not be a valid cdf, this naive estimator can be improved further as

$$\hat{F}_{2,n}^\gamma = \operatorname{argmin}_{W \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \{W(X_i) - \hat{F}_{1,n}^\gamma\}^2$$

where  $\mathcal{F}$  is the set of all cdfs. Essentially,  $\hat{F}_{2,n}^\gamma$  is the closest cdf to the naive estimator  $\hat{F}_{1,n}^\gamma$ . The minimizer can be easily found by the pool adjacent violators algorithm mentioned in [Robertson \(1988\)](#). In practice,  $\pi_1$  is usually unknown and using the two estimators  $\hat{F}_{1,n}^\gamma$  and  $\hat{F}_{2,n}^\gamma$  over all possible values of  $\gamma \in (0, 1]$ , [Patra and Sen \(2016\)](#) suggested the following estimator of  $\pi_1$ .

$$\hat{\pi}_1^{c_n} = \inf \left\{ \gamma \in (0, 1] : \gamma d_n(\hat{F}_{1,n}^\gamma, \hat{F}_{2,n}^\gamma) \leq \frac{c_n}{\sqrt{n}} \right\} \quad (5.7)$$

where  $c_n$  is a sequence of constants satisfying  $c_n = o(\sqrt{n})$  and  $d_n(\cdot, \cdot)$  is same as the  $\mathcal{L}_2(F_n)$  distance, i.e. if  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  are two functions, then

$$d_n^2(f, g) = \int (f(x) - g(x))^2 dF_n(x) = \frac{1}{n} \sum_{i=1}^n (f(X_{(i)}) - g(X_{(i)}))^2$$

The idea of this estimator originated from the following observations.

- If  $\gamma = 1$ , then clearly  $\hat{F}_{1,n}^1 = \hat{F}_{2,n}^1 = F_n$  and they are much different for smaller values of  $\gamma$ .
- For any  $\gamma \in (0, 1]$ , it is easy to see that,

$$d_n(F_n, \gamma \hat{F}_{2,n}^\gamma + (1 - \gamma)F_0) = \gamma d_n(\hat{F}_{1,n}^\gamma, \hat{F}_{2,n}^\gamma)$$

And

$$\lim_{\gamma \rightarrow 0^+} \gamma d_n(\hat{F}_{1,n}^\gamma, \hat{F}_{2,n}^\gamma) = d_n(F_n, F_0)$$

The quantity  $\gamma d_n(\hat{F}_{1,n}^\gamma, \hat{F}_{2,n}^\gamma)$  represents the difference between two estimators of the overall distribution function  $F$ . One estimator is the empirical cdf ( $F_n$ ), and the other is derived by using  $\hat{F}_{2,n}^\gamma$  as a plug-in estimator of  $F_1$ .

- $\gamma d_n(\hat{F}_{1,n}^\gamma, \hat{F}_{2,n}^\gamma)$  changes its behaviour at  $\underline{\pi}_1$ .

**Theorem 5.5.1.** *For  $\underline{\pi}_1 \leq \gamma \leq 1$ ,  $\gamma d_n(\hat{F}_{1,n}^\gamma, \hat{F}_{2,n}^\gamma) \leq d_n(F, F_n)$ . Then, if  $\|F_n - F\|_\infty \xrightarrow{a.s.} 0$  as  $n \rightarrow \infty$  then,*

$$\gamma d_n(\hat{F}_{1,n}^\gamma, \hat{F}_{2,n}^\gamma) \xrightarrow{a.s.} \begin{cases} 0 & \text{if } \gamma \geq \underline{\pi}_1 \\ > 0, & \text{if } \gamma < \underline{\pi}_1 \end{cases}$$

The proof of Theorem 5.5.1 can be easily done using the idea similar to [Patra and Sen \(2016\)](#). The proof is given in appendix. Theorem 5.5.1 demonstrates that the function  $\gamma d_n(\hat{F}_{1,n}^\gamma, \hat{F}_{2,n}^\gamma)$  changes behaviour at  $\underline{\pi}_1$ . For sufficiently large  $n$ , the function has a steep decline to the left of  $\underline{\pi}_1$  and a more gradual decrease near zero to the right of  $\underline{\pi}_1$ . Thus,  $\gamma d_n(\hat{F}_{1,n}^\gamma, \hat{F}_{2,n}^\gamma)$  exhibits an elbow-like structure around  $\underline{\pi}_1$ . To estimate  $\underline{\pi}_1$ , one should select the smallest  $\gamma$  for which the distance remains small. This leads to the estimator discussed in 5.7. The consistency of the estimator requires a criterion stronger than  $\|F_n - F\|_\infty \xrightarrow{a.s.} 0$  as  $n \rightarrow \infty$ . Specifically, we need  $\sqrt{n}d_n(F_n, F)$  to be stochastically bounded for the estimator in 5.7 to be consistent. The following theorem explains the consistency of the estimator in 5.7.

**Theorem 5.5.2.** *If  $\sqrt{n}d_n(F_n, F)$  is stochastically bounded and  $c_n \rightarrow \infty$  and  $\lim_{n \rightarrow \infty} \frac{c_n}{\sqrt{n}} = 0$  (i.e.  $c_n = o(\sqrt{n})$ ), then  $\hat{\pi}_1^{c_n} \xrightarrow{P} \underline{\pi}_1$  as  $n \rightarrow \infty$ .*

The proof of theorem 5.5.2 can be done using the idea similar to [Patra and Sen \(2016\)](#). The proof of this theorem is given in appendix. The criterion  $\sqrt{n}d_n(F_n, F)$  remains stochastically bounded is stronger than  $\|F_n - F\|_\infty \xrightarrow{a.s.} 0$  as  $n \rightarrow \infty$  which is required by 5.4.1. We highlight a few cases from [Billingsley \(2017\)](#) where this condition is met.

### 5.5.1 Mixing sequence

As per the definitions of [Billingsley \(2013\)](#), consider two measures of dependence between two sigma algebras  $\mathcal{F}$  and  $\mathcal{G}$ .

- ◆  $\alpha_n = \sup \{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{F}, B \in \mathcal{G}\}$
- ◆  $\rho_n = \sup \{|E[\xi\eta]| : E[\xi] = E[\eta] = 0, \|\xi\| \leq 1, \|\eta\| \leq 1, \sigma(\xi) \subseteq \mathcal{F}, \sigma(\eta) \subseteq \mathcal{G}\}$

It is well known that,  $\alpha_n \leq \rho_n$ .

- A sequence of stationary random variables  $X_1, X_2, \dots$  is said to be  $\alpha$ -mixing if for  $\mathcal{F} = \sigma(X_1, \dots, X_k)$  and  $\mathcal{G} = \sigma(X_{k+n}, X_{k+n+1}, \dots)$ ,  $\lim_{n \rightarrow \infty} \alpha_n = 0$ .
- A sequence of stationary random variables  $X_1, X_2, \dots$  is said to be  $\rho$ -mixing if  $\lim_{n \rightarrow \infty} \rho_n = 0$ .

The requirement for a mixing sequence is more rigorous than just being stationary and ergodic. For various types of mixing sequences, the criterion  $\sqrt{n}\|F_n - F\|_\infty = O_P(1)$  is satisfied. We highlight a few examples below.

- For a stationary sequence of random variables  $X_1, X_2, \dots$ , we define a sequence  $\{\rho_n(x)\}_{n \geq 1}$  as follows.

$$\rho_{|i-j|}(x) = P(X_i \leq x, X_j \leq x) - (F(x))^2$$

If for every  $x$ ,  $\sum_{n=1}^{\infty} \rho_n(x) < \infty$ , then we will have  $\sqrt{n}\|F_n - F\|_\infty = O_P(1)$  as a consequence of theorem 19.2 from [Billingsley \(2013\)](#).

- If the stationary sequence of random variables  $X_1, X_2, \dots$  is  $\alpha$ -mixing with  $\alpha_n = O(n^{-5})$ , then  $\sqrt{n}\|F_n - F\|_\infty = O_P(1)$ .

Many real-life applications frequently involve examples that meet the assumptions outlined in the previous discussion. For instance, if a stationary sequence is either m-dependent or block dependent, it will satisfy both of the criteria mentioned above. Also, by [Billingsley \(2017\)](#), a stationary markov chain with finite state space is  $\alpha$ -mixing for  $\alpha_n = K\theta^n$  for some  $\theta$  with  $|\theta| < 1$ . Thus, it also satisfies the above criterion.

### 5.5.2 Simulation results

According to Theorem 5.5.2, a sequence  $\{c_n\}$  is required such that  $c_n \rightarrow \infty$  as  $n \rightarrow \infty$  and  $\lim_{n \rightarrow \infty} \frac{c_n}{\sqrt{n}} = 0$ . *Patra and Sen (2016)* emphasized the importance of selecting an appropriate  $c_n$  for the effective performance of the estimator  $\hat{\pi}_1^{c_n}$ . Their recommendation, based on finite sample performance, suggests  $c_n = 0.1 \log(\log n)$  as a useful choice. Additionally, they proposed a cross-validation-based method for choosing the optimal  $c_n$ . For our simulations, we employed 10-fold cross-validation to determine the optimal  $c_n$ , and also tested the fixed choice  $c_n = 0.1 \log(\log n)$ . We have considered some models where  $\sqrt{n} \|F_n - F\|_\infty$  is stochastically bounded.

Our primary focus is the AR(1) model described in 5.6. We consider a sample size of  $n = 1000$  and p-values derived from a mixture of  $N(0, 1)$  and  $N(2, 1)$  distributions. For the AR(1) parameter  $a$ , we examine four values:  $a = 0.1, 0.2, 0.5,$  and  $0.75$ . The performance of the estimator for  $a = 0.1$  and  $a = 0.2$  is detailed in Table 5.16 below.

The estimator  $\hat{\pi}_1^{c_n}$  estimates  $\pi_1 = 1 - \pi_0$ . In Table 5.16, we present the estimated values of  $\pi_0$  (i.e.,  $1 - \hat{\pi}_1^{c_n}$ ), rounded to three decimal places.

**Table 5.16:** AR(1) model with  $a = 0.1$  and  $0.2$

$\pi_0$	$a = 0.1$		$a = 0.2$	
	$\hat{\pi}_0$ (Fixed)	$\hat{\pi}_0$ (CV)	$\hat{\pi}_0$ (Fixed)	$\hat{\pi}_0$ (CV)
0.5	0.544	0.500	0.547	0.512
0.6	0.636	0.596	0.637	0.599
0.65	0.682	0.652	0.685	0.644
0.7	0.729	0.698	0.724	0.700
0.75	0.770	0.729	0.773	0.738
0.8	0.818	0.795	0.816	0.786
0.85	0.860	0.841	0.859	0.835
0.9	0.899	0.886	0.898	0.883
0.95	0.945	0.935	0.940	0.927
0.98	0.963	0.969	0.966	0.953

For  $a = 0.1$  and  $a = 0.2$ , there is minimal difference between the estimated values obtained using fixed and cross-validation approaches. The estimators  $\hat{\pi}_0$  remain close to the true

value of  $\pi_0$ . However, as we increase the AR(1) coefficient to  $a = 0.5$  and  $a = 0.75$ , the gap between the true value of  $\pi_0$  and  $\hat{\pi}_0$  widens due to the increased level of dependence. This indicates that the rate of convergence slows as the coefficient  $a$  increases.

**Table 5.17:** AR(1) model with  $a = 0.5$  and  $0.75$

$\pi_0$	$a = 0.5$		$a = 0.75$	
	$\hat{\pi}_0$ (Fixed)	$\hat{\pi}_0$ (CV)	$\hat{\pi}_0$ (Fixed)	$\hat{\pi}_0$ (CV)
0.5	0.556	0.499	0.525	0.440
0.6	0.636	0.583	0.564	0.474
0.65	0.668	0.629	0.588	0.495
0.7	0.706	0.638	0.608	0.517
0.75	0.734	0.687	0.627	0.548
0.8	0.772	0.706	0.645	0.552
0.85	0.801	0.756	0.664	0.558
0.9	0.837	0.782	0.684	0.593
0.95	0.866	0.806	0.701	0.629
0.98	0.885	0.839	0.714	0.630

It is interesting to observe that the discrepancy between  $\hat{\pi}_0$  and  $\pi_0$  becomes more pronounced as  $\pi_0$  increases. The estimator's performance is satisfactory for smaller values of  $\pi_0$  (up to  $\pi_0 = 0.6$ ). Moreover, using a cross-validation-based estimator provides little or no improvement in this range. In these cases, the fixed choice of  $c_n = 0.1 \log(\log n)$  also yields good results.

We now consider the  $m$ -dependent scenario. As before, given a sample  $\{\xi_i\}_{i \geq 1}$ , observations are generated according to  $X_i \propto \sum_{j=i}^{i+m} ((m+1) - (j-i)) \xi_j$  and are standardized to have unit variance. A mixture of  $N(0, 1)$  and  $N(2, 1)$  variables is then formed from this sequence  $\{X_i\}_{i \geq 1}$ , and the p-values are computed for a one-sided test of the mean. We conduct simulations with  $n = 1000$  and consider  $m = 1, 2$ , and  $5$ . The estimated values of  $\pi_0$  are shown in the table 5.18.

Due to the very slow rate of convergence, the deviation between the true and estimated values of  $\pi_0$  is evident in this setup. This discrepancy becomes even more pronounced

**Table 5.18:** Estimator of null proportion for m-dependent model with m = 1,2,5

	<b>m=1</b>		<b>m=2</b>		<b>m=5</b>	
$\pi_0$	$\hat{\pi}_0$ ( Fixed)	$\hat{\pi}_0$ ( CV)	$\hat{\pi}_0$ ( Fixed)	$\hat{\pi}_0$ ( CV)	$\hat{\pi}_0$ ( Fixed)	$\hat{\pi}_0$ ( CV)
0.5	0.540	0.460	0.512	0.425	0.449	0.345
0.6	0.599	0.508	0.550	0.459	0.466	0.368
0.65	0.622	0.549	0.568	0.470	0.474	0.388
0.7	0.649	0.571	0.588	0.503	0.482	0.408
0.75	0.674	0.602	0.605	0.518	0.491	0.378
0.8	0.701	0.627	0.622	0.543	0.497	0.418
0.85	0.728	0.679	0.638	0.543	0.505	0.410
0.9	0.750	0.688	0.656	0.549	0.513	0.423
0.95	0.773	0.703	0.672	0.580	0.522	0.430
0.98	0.786	0.718	0.682	0.594	0.525	0.437

as the degree of dependence increases, which occurs when the value of  $m$  is raised. Additionally, in the m-dependent case where equal weights are assigned to neighbors (i.e.,  $X_i = \frac{1}{\sqrt{m+1}} \sum_{j=i}^{i+m} \xi_j$ ), the deviation between the true and estimated values is even more substantial. This is clearly illustrated in the table 5.19. It is noteworthy that, due to the increased dependence among observations, the performance of the cross-validation-based choice of  $c_n$  is inferior to the fixed choice regarding the deviation between the true and estimated values of  $\pi_0$ . The assumption of independence is crucial for the validity of cross-validation-based methods, which is why the fixed choice of  $c_n$  performs better in scenarios with a higher degree of dependence. Consequently, the extensive time spent on cross-validation appears to be unproductive in this context.

**Table 5.19:** Equally weighted m-dependent model with m = 1,2,5

	<b>m=1</b>		<b>m=2</b>		<b>m=5</b>	
$\pi_0$	$\hat{\pi}_0$ (Fixed)	$\hat{\pi}_0$ (CV)	$\hat{\pi}_0$ (Fixed)	$\hat{\pi}_0$ (CV)	$\hat{\pi}_0$ (Fixed)	$\hat{\pi}_0$ (CV)
0.5	0.443	0.381	0.449	0.398	0.444	0.386
0.6	0.501	0.448	0.498	0.436	0.496	0.429
0.65	0.516	0.482	0.522	0.472	0.515	0.463
0.7	0.550	0.495	0.546	0.499	0.537	0.467
0.75	0.564	0.528	0.570	0.523	0.561	0.513
0.8	0.588	0.546	0.585	0.553	0.589	0.525
0.85	0.608	0.572	0.611	0.563	0.608	0.538
0.9	0.632	0.591	0.627	0.576	0.608	0.581
0.95	0.648	0.614	0.640	0.595	0.636	0.580
0.98	0.658	0.621	0.652	0.615	0.649	0.597

We now shift our focus to the block-dependent setup where the requirement of stationarity is not satisfied. In this context, the estimator  $\hat{\pi}_1^{c_n}$  appears to perform better compared to the m-dependent case. We fixed  $n = 1000$  and examined two scenarios: one with 10 blocks of size 100 and another with 200 blocks of size 5. Clearly, the latter scenario is closer to independence than the former. The estimators are presented in Table 5.20. As before, we considered 10 different correlation values for the block-dependent case, namely 0.1, 0.2,  $\dots$ , 0.9, 0.95.

**Table 5.20:** Block dependent case with  $n = 1000$

$\pi_0$	Block of size 5		Block of size 100	
	$\hat{\pi}_0$ (Fixed)	$\hat{\pi}_0$ (CV)	$\hat{\pi}_0$ (Fixed)	$\hat{\pi}_0$ (CV)
0.5	0.531	0.481	0.540	0.488
0.6	0.616	0.568	0.628	0.578
0.65	0.657	0.621	0.671	0.635
0.7	0.696	0.671	0.716	0.670
0.75	0.737	0.697	0.755	0.690
0.8	0.773	0.753	0.799	0.749
0.85	0.815	0.774	0.836	0.807
0.9	0.851	0.817	0.876	0.849
0.95	0.881	0.877	0.911	0.890
0.98	0.896	0.900	0.920	0.892

Interestingly, the performance of the cross-validation-based estimator deteriorates when moving from 200 blocks to 10 blocks, due to the increased degree of dependence. In contrast, the fixed choice of  $c_n$  provides a relatively satisfactory estimate of  $\pi_0$ . However, using these estimates may not ensure conservative control over the false discovery rate (FDR), as they tend to underestimate  $\pi_0$  at higher values.

## 5.6 Remarks

The simulation results indicate that both Storey's estimator and the estimator proposed by [Patra and Sen \(2016\)](#) are sensitive to the dependence structure in the data. Storey's

estimator is known to be a biased estimator of  $\pi_0$ , whereas the estimator  $\hat{\pi}_1^{c_n}$  is consistent under conditions of weak dependence. However, its rate of convergence is notably slow, and this rate decreases even further in the presence of dependence. Even relatively simple dependence structures, such as  $m$ -dependence, can significantly distort the performance of these estimators. Therefore, further research is needed to develop a consistent estimator of the proportion of non-null hypotheses that remains reliable under weak dependence.

## 5.7 Appendix

### 5.7.1 Proof of theorem 5.5.1

**Case 1:**  $\{\gamma \geq \underline{\pi}_1\}$

Let,

$$F_{1,\gamma} = \frac{F - (1 - \gamma)F_0}{\gamma} \quad (5.8)$$

Then,  $\gamma d_n(\hat{F}_{1,n}^\gamma, \hat{F}_{1,\gamma}) = d_n(F_n, F)$ .

Consider the set  $\Gamma = \left\{ \gamma \in [0, 1] \mid \left\{ \frac{F - (1 - \gamma)F_0}{\gamma} \text{ is a valid cdf} \right\} \right\}$ . We'll first show that,  $\Gamma$  is a convex set.

Suppose  $\gamma_1, \gamma_2 \in \Gamma$  so that,  $\frac{F - (1 - \gamma_i)F_0}{\gamma_i}$  is a valid cdf for  $i = 1, 2$ .

For any  $0 \leq \alpha \leq 1$ ,

$$\begin{aligned} & \frac{F - (1 - \alpha\gamma_1 - (1 - \alpha)\gamma_2)F_0}{\alpha\gamma_1 + (1 - \alpha)\gamma_2} \\ &= \left( \frac{\alpha\gamma_1}{\alpha\gamma_1 + (1 - \alpha)\gamma_2} \right) \left( \frac{F - (1 - \gamma_1)F_0}{\gamma_1} \right) + \left( \frac{(1 - \alpha)\gamma_2}{\alpha\gamma_1 + (1 - \alpha)\gamma_2} \right) \left( \frac{F - (1 - \gamma_2)F_0}{\gamma_2} \right) \end{aligned}$$

Since convex combination of cdfs should be cdf, we can now say that,  $\frac{F - (1 - \alpha\gamma_1 - (1 - \alpha)\gamma_2)F_0}{\alpha\gamma_1 + (1 - \alpha)\gamma_2}$  is a valid cdf and hence  $\alpha\gamma_1 + (1 - \alpha)\gamma_2 \in \Gamma$ .

This implies, for any  $\underline{\pi}_1 \leq \gamma \leq 1$ ,  $\frac{F - (1 - \gamma)F_0}{\gamma}$  is a valid cdf. Thus,  $F_{1,\gamma}$  defined in 5.8 is a valid cdf.

So, by the definition of  $\hat{F}_{2,n}$ , it can be said that,

$$\gamma d_n(\hat{F}_{1,n}^\gamma, \hat{F}_{2,n}^\gamma) \leq \gamma d_n(\hat{F}_{1,n}, \hat{F}_{1,\gamma}) = d_n(F_n, F) \quad \forall \gamma \geq \underline{\pi}_1$$

Since,  $d_n(F_n, F) \leq \|F_n - F\|_\infty$ ,  $\|F_n - F\|_\infty \xrightarrow{a.s.} 0$  as  $n \rightarrow \infty$  implies,  $d_n(F_n, F) \xrightarrow{a.s.} 0$  as  $n \rightarrow \infty$ . Hence,

$$\gamma d_n(\hat{F}_{1,n}^\gamma, \hat{F}_{2,n}^\gamma) \xrightarrow{a.s.} 0 \text{ as } n \rightarrow \infty$$

**Case 2:**  $\{\gamma < \underline{\pi}_1\}$

$\hat{F}_{1,n}^\gamma$  is not a valid cdf. by the definition of  $\underline{\pi}_1$ . Also, if  $\|F_n - F\|_\infty \xrightarrow{a.s.} 0$ , then  $\hat{F}_{1,n}^\gamma \xrightarrow{a.s.} F_{1,\gamma}$  as  $n \rightarrow \infty$ .

So, for sufficiently large  $n$ , the difference between  $\hat{F}_{1,n}^\gamma$  and  $\hat{F}_{2,n}^\gamma$  is substantially high as  $\hat{F}_{1,n}^\gamma$  converges to  $F_{1,\gamma}$  which is not a valid cdf and by definition,  $\hat{F}_{2,n}^\gamma$  is a cdf.

Combining the two observations in case 1 and case 2, we can finally conclude that,

$$\gamma d_n(\hat{F}_{1,n}^\gamma, \hat{F}_{2,n}^\gamma) \xrightarrow{a.s.} \begin{cases} 0 & \text{if } \gamma \geq \underline{\pi}_1 \\ > 0, & \text{if } \gamma < \underline{\pi}_1 \end{cases}$$

### 5.7.2 Proof of theorem 5.5.2

Fix any  $\epsilon > 0$ . We need to show that,  $\lim_{n \rightarrow \infty} P(|\hat{\pi}_1^{c_n} - \underline{\pi}_1| \geq \epsilon) = 0$

We will first prove that,  $\lim_{n \rightarrow \infty} P(\hat{\pi}_1^{c_n} - \underline{\pi}_1 \leq -\epsilon) = 0$ .

If  $\underline{\pi}_1 \leq \epsilon$ , then  $\underline{\pi}_1 - \epsilon \leq 0$  and hence it is obvious. So, assume that  $\underline{\pi}_1 > \epsilon$ .

Recall that,

$$\underline{\pi}_1 = \inf \left\{ \gamma \in (0, 1] : \left\{ \frac{F - (1 - \gamma)F_0}{\gamma} \right\} \text{ is a CDF} \right\}$$

So, for  $\gamma < \underline{\pi}_1$ ,  $\frac{F - (1 - \gamma)F_0}{\gamma}$  is not a cdf and for any such  $\gamma$ ,  $\hat{F}_{1,n}^\gamma, \hat{F}_{2,n}^\gamma$  are quite different.

Consider  $\gamma = (\underline{\pi}_1 - \epsilon)$ , and we can conclude that,  $(\underline{\pi}_1 - \epsilon)d_n(\hat{F}_{1,n}^{(\underline{\pi}_1 - \epsilon)}, \hat{F}_{2,n}^{(\underline{\pi}_1 - \epsilon)})$  is bounded away from zero.

Recall that,

$$\hat{\pi}_1^{c_n} = \inf \left\{ \gamma \in (0, 1] : \gamma d_n(\hat{F}_{1,n}^\gamma, \hat{F}_{2,n}^\gamma) \leq \frac{c_n}{\sqrt{n}} \right\}$$

Consider the set  $A_n = \left\{ \gamma \in (0, 1] : \gamma d_n(\hat{F}_{1,n}^\gamma, \hat{F}_{2,n}^\gamma) \leq \frac{c_n}{\sqrt{n}} \right\}$ . Clearly  $A_n$  is a convex set.

So,  $P(\hat{\pi}_1^{c_n} \leq \underline{\pi}_1 - \epsilon) \leq P\left((\underline{\pi}_1 - \epsilon)d_n(\hat{F}_{1,n}^{(\underline{\pi}_1 - \epsilon)}, \hat{F}_{2,n}^{(\underline{\pi}_1 - \epsilon)}) \leq \frac{c_n}{\sqrt{n}}\right)$

Since,  $(\underline{\pi}_1 - \epsilon)d_n \left( \hat{F}_{1,n}^{(\underline{\pi}_1 - \epsilon)}, \hat{F}_{2,n}^{(\underline{\pi}_1 - \epsilon)} \right)$  is bounded away from zero and  $\lim_{n \rightarrow \infty} \frac{c_n}{\sqrt{n}} = 0$ , we can say that,

$$\lim_{n \rightarrow \infty} P \left( (\underline{\pi}_1 - \epsilon)d_n \left( \hat{F}_{1,n}^{(\underline{\pi}_1 - \epsilon)}, \hat{F}_{2,n}^{(\underline{\pi}_1 - \epsilon)} \right) \leq \frac{c_n}{\sqrt{n}} \right) = 0$$

Thus,  $\lim_{n \rightarrow \infty} P(\hat{\pi}_1^{c_n} \leq \underline{\pi}_1 - \epsilon) = 0$

On the other hand,  $\{\hat{\pi}_1^{c_n} - \underline{\pi}_1 > \epsilon\}$  implies,  $\left\{ (\underline{\pi}_1 + \epsilon)d_n \left( \hat{F}_{1,n}^{(\underline{\pi}_1 + \epsilon)}, \hat{F}_{2,n}^{(\underline{\pi}_1 + \epsilon)} \right) > \frac{c_n}{\sqrt{n}} \right\}$ .

So,

$$P(\hat{\pi}_1^{c_n} - \underline{\pi}_1 > \epsilon) \leq P \left( \sqrt{n}d_n \left( \hat{F}_{1,n}^{(\underline{\pi}_1 + \epsilon)}, \hat{F}_{2,n}^{(\underline{\pi}_1 + \epsilon)} \right) > \frac{c_n}{(\underline{\pi}_1 + \epsilon)} \right)$$

From the proof of theorem 5.5.1, for any  $\gamma \in (\underline{\pi}_1, 1]$ ,  $\gamma d_n(\hat{F}_{1,n}^\gamma, \hat{F}_{2,n}^\gamma) \leq d_n(F_n, F)$ .

$$\text{So, } P(\hat{\pi}_1^{c_n} - \underline{\pi}_1 > \epsilon) \leq P \left( \sqrt{n}d_n(F_n, F) > \frac{c_n}{(\underline{\pi}_1 + \epsilon)} \right)$$

Since,  $\lim_{n \rightarrow \infty} c_n = \infty$  and  $\sqrt{n}d_n(F_n, F)$  is stochastically bounded, we can conclude that,

$$\lim_{n \rightarrow \infty} P \left( \sqrt{n}d_n(F_n, F) > \frac{c_n}{(\underline{\pi}_1 + \epsilon)} \right) = 0$$

Thus,  $\lim_{n \rightarrow \infty} P(\hat{\pi}_1^{c_n} - \underline{\pi}_1 > \epsilon) = 0$  and this completes the proof.

# Bibliography

- Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.
- Y. Benjamini and Y. Hochberg. On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of educational and Behavioral Statistics*, 25(1):60–83, 2000.
- Y. Benjamini and W. Liu. A distribution-free multiple test procedure that controls the false discovery rate. *Tel Aviv, Tel Aviv University*, 1999a.
- Y. Benjamini and W. Liu. A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence. *Journal of statistical planning and inference*, 82(1-2):163–170, 1999b.
- Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *The annals of statistics*, 29(4):1165–1188, 2001.
- P. Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.
- P. Billingsley. *Probability and measure*. John Wiley & Sons, 2017.
- A. Biswas, G. Chattopadhyay, and A. Chatterjee. Bias-corrected estimators for proportion of true null hypotheses: application of adaptive fdr-controlling in segmented failure data. *Journal of Applied Statistics*, 49(14):3591–3613, 2022.
- G. Blanchard and É. Roquain. Adaptive false discovery rate control under independence and dependence. *Journal of Machine Learning Research*, 10(12), 2009.

- T. T. Cai and J. Jin. Optimal rates of convergence for estimating the null density and proportion of nonnull effects in large-scale multiple testing. 2010.
- A. Celisse and S. Robin. A cross-validation based estimation of the proportion of true null hypotheses. *Journal of Statistical Planning and Inference*, 140(11):3132–3147, 2010.
- A. C. Cohen. Estimation in mixtures of two normal distributions. *Technometrics*, 9(1):15–28, 1967.
- N. Das. Estimation of proportion of null hypotheses under dependence. *arXiv preprint arXiv:2409.04100*, 2024.
- N. Das and S. K. Bhandari. Observation on fwer and fdr for correlated normal. *arXiv preprint arXiv:2008.08366*, 2020.
- N. Das and S. K. Bhandari. Bound on fwer for correlated normal. *Statistics & Probability Letters*, 168:108943, 2021.
- N. Das and S. K. Bhandari. Optimal test statistic under normality assumption. *arXiv preprint arXiv:2306.10554*, 2023.
- N. Das and S. K. Bhandari. Fwer for normal distribution in nearly independent setup. *Statistics & Probability Letters*, 219:110340, 2025.
- M. Dey. Behaviour of familywise error rate in normal distributions. *arXiv preprint arXiv:2107.00146*, 2021.
- M. Dey. On asymptotic behaviors of stepwise multiple testing procedures. *arXiv preprint arXiv:2212.08372*, 2022.
- M. Dey and S. K. Bhandari. Fwer goes to zero for correlated normal. *Statistics & Probability Letters*, 193:109700, 2023.
- M. Dey and S. K. Bhandari. Bounds on generalized family-wise error rates for normal distributions. *Statistical Papers*, 65(4):2313–2326, 2024.

- A. Dmitrienko, A. C. Tamhane, and F. Bretz. *Multiple testing problems in pharmaceutical statistics*. CRC press, 2009.
- S. Dudoit and M. J. Van Der Laan. *Multiple testing procedures with applications to genomics*. Springer Science & Business Media, 2007.
- S. Dudoit, J. P. Shaffer, and J. C. Boldrick. Multiple hypothesis testing in microarray experiments. *Statistical Science*, 18(1):71–103, 2003.
- S. Dudoit, M. J. Van Der Laan, and M. J. van der Laan. *Multiple testing procedures with applications to genomics*. Springer, 2008.
- B. Efron. Correlation and large-scale simultaneous significance testing. *Journal of the American Statistical Association*, 102(477):93–103, 2007.
- B. Efron. Microarrays, empirical bayes and the two-groups model. 2008.
- B. Efron. Correlated z-values and the accuracy of large-scale statistical estimates. *Journal of the American Statistical Association*, 105(491):1042–1055, 2010.
- B. Efron. *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*, volume 1. Cambridge University Press, 2012.
- B. Efron, R. Tibshirani, J. D. Storey, and V. Tusher. Empirical bayes analysis of a microarray experiment. *Journal of the American statistical association*, 96(456):1151–1160, 2001.
- J. Fan, X. Han, and W. Gu. Estimating false discovery proportion under arbitrary covariance dependence. *Journal of the American Statistical Association*, 107(499):1019–1035, 2012.
- H. Finner and M. Roters. Asymptotic sharpness of product-type inequalities for maxima of random variables with applications in multiple comparisons. *Journal of statistical planning and inference*, 98(1-2):39–56, 2001.

- H. Finner and M. Roters. Multiple hypotheses testing and expected number of type I errors. *The Annals of Statistics*, 30(1):220–238, 2002.
- H. Finner, T. Dickhaus, and M. Roters. Dependency and false discovery rate: asymptotics. 2007.
- C. Genovese and L. Wasserman. Operating characteristics and extensions of the false discovery rate procedure. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):499–517, 2002.
- C. Genovese and L. Wasserman. A stochastic process approach to false discovery control. 2004.
- C. R. Genovese, K. Roeder, and L. Wasserman. False discovery control with p-value weighting. *Biometrika*, 93(3):509–524, 2006.
- A. Y. Gordon and P. Salzman. Optimality of the holm procedure among general step-down multiple testing procedures. *Statistics & probability letters*, 78(13):1878–1884, 2008.
- Y. Hochberg. Multiple comparison procedures. *Wiley Series in Probability and Statistics*, 1987.
- Y. Hochberg. A sharper bonferroni procedure for multiple tests of significance. *Biometrika*, 75(4):800–802, 1988.
- S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979.
- G. Hommel. A stagewise rejective multiple test procedure based on a modified bonferroni test. *Biometrika*, 75(2):383–386, 1988.
- J. Jin. Proportion of non-zero normal means: universal oracle equivalences and uniformly consistent estimators. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 70(3):461–493, 2008.

- M. Langaas, B. H. Lindqvist, and E. Ferkingstad. Estimating the proportion of true null hypotheses, with application to dna microarray data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(4):555–572, 2005.
- E. L. Lehmann and J. P. Romano. Generalizations of the familywise error rate. In *Selected Works of EL Lehmann*, pages 719–735. Springer, 2012.
- B. G. Lindsay and P. Basak. Multivariate normal mixtures: a fast consistent method of moments. *Journal of the American Statistical Association*, 88(422):468–476, 1993.
- N. Loperfido, J. Navarro, J. M. Ruiz, and C. J. Sandoval. Some relationships between skew-normal distributions and order statistics from exchangeable normal random vectors. *Communications in Statistics—Theory and Methods*, 36(9):1719–1733, 2007.
- N. Meinshausen and P. Bühlmann. Lower bounds for the number of false null hypotheses for multiple testing of associations under general dependence structures. *Biometrika*, 92(4):893–907, 2005.
- N. Meinshausen and J. Rice. Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. 2006.
- R. K. Patra and B. Sen. Estimation of a two-component mixture model with applications to multiple testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(4):869–893, 2016.
- M. A. Proschan and P. A. Shaw. Asymptotics of bonferroni for dependent normal test statistics. *Statistics & probability letters*, 81(7):739–748, 2011.
- T. Robertson. Order restricted statistical inference/[et al.]. 1988.
- S. K. Sarkar. Stepup procedures controlling generalized fwer and generalized fdr. *The Annals of Statistics*, 35(6):2405–2420, 2007.
- S. K. Sarkar. On methods controlling the false discovery rate. *Sankhyā: The Indian Journal of Statistics, Series A (2008-)*, pages 135–168, 2008.

- I. R. Savage. Mills' ratio for multivariate normal distributions. *J. Res. Nat. Bur. Standards Sect. B*, 66:93–96, 1962.
- T. Schweder and E. Spjøtvoll. Plots of p-values to evaluate many tests simultaneously. *Biometrika*, 69(3):493–502, 1982.
- Z. Šidák. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318):626–633, 1967.
- R. J. Simes. An improved bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754, 1986.
- J. D. Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):479–498, 2002.
- J. D. Storey. The positive false discovery rate: a bayesian interpretation and the q-value. *The annals of statistics*, 31(6):2013–2035, 2003.
- J. D. Storey. The optimal discovery procedure: a new approach to simultaneous significance testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(3):347–368, 2007.
- J. D. Storey, J. E. Taylor, and D. Siegmund. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 66(1):187–205, 2004.
- W. Sun and T. T. Cai. Oracle and adaptive compound decision rules for false discovery rate control. *Journal of the American Statistical Association*, 102(479):901–912, 2007.
- W. Sun and T. Tony Cai. Large-scale multiple testing under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):393–424, 2009.
- Y. L. Tong. *Probability inequalities in multivariate distributions*. Academic Press, 2014.

- P. H. Westfall and S. S. Young. *Resampling-based multiple testing: Examples and methods for p-value adjustment*, volume 279. John Wiley & Sons, 1993.
- J. Xie, T. T. Cai, J. Maris, and H. Li. Optimal false discovery rate control for dependent data. *Statistics and its interface*, 4(4):417, 2011.
- J. Xie, T. T. Cai, and H. Li. Correction to the paper “optimal false discovery rate control for dependent data”. *Statistics and Its Interface*, 9(1):33–35, 2016.