Some Contributions to Multiple Hypotheses Testing under Dependence

Monitirtha Dey

Interdisciplinary Statistical Research Unit
Indian Statistical Institute
2024

INDIAN STATISTICAL INSTITUTE

DOCTORAL THESIS

Some Contributions to Multiple Hypotheses Testing under Dependence

---

(Final Version)

Author.
MONITIRTHA DEY

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

Indian Statistical Institute

Declaration

I declare that

(a) The research work presented in this thesis titled '*Some Contributions to Multiple Hypotheses Testing under Dependence*' is original and has been done by myself under the general guidance of my advisor, Prof. Subir Kumar Bhandari.

(b) This thesis represents my ideas in my own words. Whenever I have used ideas or materials (text, data, theoretical analysis) from other sources, I have adequately cited and referenced the original sources.

(c) I have adhered to all principles of academic honesty and integrity in my submission.

(d) This work has not been submitted to any other Institute or University for a degree.

(e) I have followed the guidelines provided by the Institute in writing this thesis.

Place : ISI Kolkata
26^{th} April, 2024

Monitirtha Dey
Senior Research Fellow (Statistics)
Interdisciplinary Statistical Research Unit
Indian Statistical Institute
Kolkata 700108, India

Indian Statistical Institute

Certificate

This is to certify that this revised version of doctoral thesis entitled '*Some Contributions to Multiple Hypotheses Testing under Dependence*', submitted by Monitirtha Dey to Indian Statistical Institute is a record of bonafide research work under my guidance and supervision. The work contained in this thesis is original and has not been submitted elsewhere for any degree.

Place : ISI Kolkata
26^{th} April, 2024

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

Acknowledgements

If I have seen further, it is by standing on the shoulders of giants. - Isaac Newton

The landmarks we cross in life are not just our own. They are the fruits of immense contributions from people around us. People who believed in us. Who encouraged us. Who inspired us. On the eve of completing my doctoral journey, I would like to express my heartfelt acknowledgements to those beacons in my life.

I consider myself extremely fortunate to have Prof. Subir Kumar Bhandari as my thesis advisor. This thesis would not exist in its present form without his patient guidance, nurturing, and constant enthusiasm. The times spent with him in all these years, those intriguing discussions about mathematics, research, and life are memories I shall cherish lifelong. I thank him for giving me so much of his time and care.

I am immensely grateful to Prof. Tapas Samanta for his constant encouragement and guidance throughout this journey. I express my gratitude to him, Prof. Arijit Chakrabarti, Prof. Debapriya Sengupta, and Prof. Probal Chaudhuri for those inference courses in my M.Stat. and doctoral coursework that, in a way, have added to my interest in the field.

My heartfelt thanks go to Prof. Diganta Mukherjee for his advice, support, and remarkable teaching. I am thankful to Dr. Ekata Saha for her inspirational teaching and positivity. I have learned a lot about teaching from their courses.

Over the last two and half years, teaching has been a pivotal part of my academic life, and I have enjoyed every bit of it. My sincere thanks to Prof. Smarajit Bose and Prof. Diganta Mukherjee for giving the teaching opportunities in their courses. I have enjoyed teaching to the fullest while co-teaching four courses with my thesis advisor - when preparing lecture notes for students, when they asked questions, or when they gave new or different proofs! A lot of thanks to my advisor and the students.

I thank the present and former Dean of Studies (Prof. Gopal Krishna Basak and Prof. Amita Pal, respectively) and the Dean's office for helping with various formalities. I sincerely thank our research fellow advisory committee chairs, Prof. Bimal Roy and Prof. Ayanendranath Basu, and the conveners, Dr. Partha Sarathi Mukherjee and Dr. Abhik Ghosh, for their helpful guidance. I also thank all the faculties and non-academic staff of the Interdisciplinary Statistical Research Unit for providing an excellent academic environment to work in. I thank the PhD-DSc committee, especially Dr. Rituparna Sen and Prof. Anil Kumar Ghosh, for helping whenever needed.

I was fortunate to have the constant company, humorous gossip, and support from my friends and seniors throughout this journey. It is my pleasure to thank Anik da, Sayantan da, Biswadeep da, Sayan da, Meghna di, Chirayata da, Debika, Soutik, Javed da, Subhankar da, Mriganka Da, Rahul da, Nabaneet da, Arijit da, Sujay da, Amarnath da, Upama, Sampurna, Arijit, and Shuvrarghya for making my time in ISRU and ASU memorable. Those tea sessions and after-tea sessions are the emotional mementos I take with me. Thanks a ton, Nadim, Rounak, Subhrajyoty, Aritra, Souvik, and Ashirwad, for those enjoyable discussions and planning sudden meets. I especially thank Prithaj, Arnab, Bikram, and Shatantik for those rejuvenating football matches and engaging weekend evenings. I thank Nistha for always being there and rooting for me - *thanks for everything!*

I dedicate this thesis to my parents and my elder brother. My mother passed away eleven years back, yet her all-time smiling face and unconditional acceptance continue to inspire me to date. My father is my first teacher in mathematics and is the one who motivated me to study statistics. His unparalleled encouragement and a plethora of good advice have often helped me during this journey. I am proud of my parents for their good deeds and genuinely feel that I am enjoying its fruits. I learned the basics of inequalities and got interested in it from my elder brother. It is heartening as many portions of this thesis use inequalities. He has always helped me whenever I got stuck with any technicality. It was my aunt who kept our family sane with her love and support after my mother passed away. Mere words can do no justice to their contributions to my life.

Revisiting the starting quote, I acknowledge those intellectual minds whose earlier works have helped me build this thesis brick by brick. I would also like to thank the editors, associate editors, and reviewers of different journals for their insightful observations and suggestions. I also thank the two anonymous reviewers of this thesis for their constructive comments and wise remarks that led to this improved version. I thank the Indian Statistical Institute for providing me with a research fellowship and facilities during this tenure. Thank you, ISI, for embracing me over the last eight years and for the countless rewarding and exhilarating experiences during my B.Stat., M.Stat., and PhD. We obtained the numerical results through the R software, and I sincerely thank its contributors for keeping it freely available.

Finally, I thank God for empowering me and keeping me and my loved ones safe and sound. Life would have been much harder otherwise. *It was the best Ph.D. life I could have asked for!*

Monitirtha Dey

Indian Statistical Institute, Kolkata

Abstract

Some Contributions to Multiple Hypotheses Testing under Dependence

Monitirtha Dey

The field of simultaneous statistical inference has attracted several statisticians for decades for its interesting theory and paramount applications. A potpourri of different methodologies exists to control various error rates, e.g., the false discovery rate (FDR) or the family-wise error rate (FWER). Most of these classical procedures were proposed under independence or some form of weak dependence among the concerned variables. However, large-scale multiple testing problems in various scientific disciplines often study correlated variables simultaneously. For example, in microRNA expression data, several genes may cluster into groups through their transcription processes and possess high correlations. The data observed from different locations and periods in public health studies are generally spatially or serially correlated. fMRI studies and multistage clinical trials also involve variables with complex and unknown dependencies. Consequently, the study of the effect of correlation on dependent test statistics in simultaneous inference problems has attracted considerable attention recently.

However, the existing literature lacks the study of the performances of FWER or generalized FWER controlling procedures under dependent setups. For these reasons, this thesis concentrates mainly on FWER and generalized FWER controlling procedures. We consider the correlated Gaussian sequence model as our underlying framework.

FWER has been a prominent error criterion in simultaneous inference for decades. The Bonferroni method is the earliest and one of the most popular methods for controlling FWER. However, we find little literature that illustrates the magnitude of the conservativeness of Bonferroni's procedure in the correlated framework with small or moderate dimensions. We address this research gap in a unified manner by establishing upper bounds on Bonferroni FWER in equicorrelated and non-negatively correlated non-asymptotic Gaussian sequence model setups.

We also derive similar upper bounds for the generalized FWERs and propose an improved k-FWER controlling procedure. Towards this, we establish an inequality related to the probability that at least k among n events occur, which extends and sharpens the classical ones. The computation of this probability arises in various contexts, e.g.,

in reliability problems of communication networks. Our probabilistic results might be insightful in those areas, too.

We also study the limiting behavior of Bonferroni FWER as the number of hypotheses approaches infinity. We prove that in the equicorrelated Gaussian setup with positive equicorrelation, Bonferroni FWER tends to zero asymptotically. These results elucidate that Bonferroni's procedure becomes extremely conservative for large-scale multiple-testing problems under correlated frameworks. We extend this result for generalized FWERs and to non-negatively correlated Normal frameworks where the limiting infimum of the correlations is strictly positive. Our proposed approximation of FWER also provides an estimate of the c.d.f of the failure time of the parallel systems.

We then move to the general class of stepwise multiple testing procedures (MTPs). The role of correlation on the limiting behavior of the FWER for stepwise procedures is less studied. Also, the existing literature lacks theoretical justifications for why FWER methods fail in large-scale problems. We address this problem by theoretically investigating the limiting FWER values of general step-down procedures under the correlated Gaussian setup. These results provide new insights into the behavior of step-down decision procedures. By establishing the limiting performances of commonly used step-up methods, e.g., the Benjamini-Hochberg (BH) and the Hochberg method, we have elucidated that the class of step-up procedures does not possess a similar universal asymptotic zero result as obtained in the case of step-down procedures. It is also noteworthy that most of our results are very general since they accommodate any combination of true and false null hypotheses. We have also obtained the limiting powers of the stepwise procedures.

Our results elucidate that, at least under the correlated Gaussian sequence model with many hypotheses, Holm's MTP and Hochberg's MTP do not have significantly different performances since they both asymptotically have zero FWER and zero power. It is also astonishing to note that, among all the procedures studied in this thesis, the BH method is the only one which can hold the FWER at a strictly positive level asymptotically under the equicorrelated Gaussian setup.

Finally, we consider the simultaneous inference problem in a sequential framework in *Chapter 6*. The mainstream sequential simultaneous inference literature has traditionally focused on the independent setup. However, there is little work studying the multiple inference problem in a sequential framework where the observations corresponding to the various streams are dependent. We consider the classical means-testing problem in an equicorrelated Gaussian and sequential framework. We focus on sequential test procedures that control the type I and type II familywise error probabilities at pre-specified levels. We establish that our proposed MTPs have the optimal average sample numbers under every possible signal configuration asymptotically, as the two types of familywise error probabilities approach zero at arbitrary rates. Towards this, we elucidate that the ratio of the expected sample size of our proposed rule and that of the classical SPRT goes to one asymptotically, thus illustrating their connection. Generalizing this, we show that our proposed procedures, with suitably modified cutoffs, are asymptotically optimal for controlling any multiple testing error criteria lying between multiples of FWER in a certain sense. This class of criteria includes FDR/FNR and pFDR/pFNR among others.

The results in this thesis illuminate that dependence might be a blessing or a curse, subject to the type of dependence or the underlying paradigm. Several popular and widely used procedures fail to hold the FWER at a positive level asymptotically under positively correlated Gaussian frameworks. On the contrary, the expected sample size of the asymptotically optimal sequential multiple testing rule is a decreasing function in the common correlation under the equicorrelated framework. Thus, correlation plays a dual role in the classical fixed-sample size and the sequential paradigms.

Journal Publications and Preprints

A version of Chapter 2 of this thesis has been published online as the following:

M. Dey (2024) Behavior of FWER in Normal Distributions, *Communications in Statistics - Theory and Methods*, 53(9), 3211–3225, DOI: 10.1080/03610926.2022.2150826.

A version of Chapter 3 has been published online as the following:

M. Dey, S. K. Bhandari (2023) Bounds on generalized family-wise error rates for normal distributions, *Statistical Papers*, DOI: https://doi.org/10.1007/s00362-023-01487-0.

A version of Chapter 4 has been published online as the following:

M. Dey, S. K. Bhandari (2023) FWER goes to zero for correlated normal, *Statistics & Probability Letters*, 193:109700, DOI: https://doi.org/10.1016/j.spl.2022.109700.

Chapter 5 is based on the following preprint and submitted to a journal.

M. Dey (2023) On Asymptotic Behaviors of Stepwise Multiple Testing Procedures, https://arxiv.org/abs/2212.08372.

Chapter 6 is based on the following preprint and submitted to a journal.

M. Dey, S. K. Bhandari (2023) Asymptotically Optimal Sequential Multiple Testing Procedures for Correlated Normal, https://arxiv.org/abs/2309.16657.

Contents

List of Tables and Figures

Notations and Abbreviations

n	The number of hypotheses tests performed.
α	The desired level of overall control of an error rate.
$N(\mu, \sigma^2)$	The univariate normal distribution with mean μ and variance σ^2.
$\mathbf{N}_d(\mu, \mathbf{\Sigma})$	The d-dimensional normal distribution with mean μ and covariance matrix $\mathbf{\Sigma}$.
$\phi(\cdot)$	The probability density function of $N(0, 1)$.
$\Phi(\cdot)$	The cumulative distribution function of $N(0, 1)$.
MTP	Multiple testing procedure.
$M_n(\rho)$	The $n \times n$ correlation matrix with each off-diagonal entry $\rho \in [0, 1]$.
Σ_n	The $n \times n$ correlation matrix with (i, j)'th entry $\rho_{ij} \in [0, 1]$, $i \neq j$.
$\mathcal{I}$	The set of first n positive integers.
H_{0i}	The i'th null hypothesis, $i \in \mathcal{I}$.
H_{1i}	The i'th alternative hypothesis, $i \in \mathcal{I}$.
H_0	The global null hypothesis.
$\mathcal{A}$	The set of true nulls, i.e, indices for which H_{0i} is true.
n_0	The number of true nulls, i.e, the cardinality of $\mathcal{A}$.
n_1	The number of true alternatives, i.e, $n - n_0$.
$R_n(T)$	The number of rejected null hypotheses by a MTP T.
$V_n(T)$	The number of falsely rejected null hypotheses by a MTP T.
$S_n(T)$	The number of correctly rejected null hypotheses by a MTP T.
FWER	Family-wise error rate.
k-FWER	Generalized familywise error rate.
FDR	False discovery rate.
BH method	Benjamini Hochberg Procedure.
$FWER_T$	The family-wise error rate of a MTP T.
FDR_T	The false discovery rate of a MTP T.
$AnyPwr_T$	The disjunctive power of a MTP T.
$AllPwr_T$	The conjunctive power of a MTP T.
$AvgPwr_T$	The average power of a MTP T.

Chapter 1

Introduction

1.1 Background

The more questions one asks, the more wrong answers one receives on average - even if the source of every single of them is genuinely authentic. We can often formalize the relevant questions in many scientific disciplines by statistical hypothesis tests. In that case, the data-driven information provides the answers. This topic is known as *multiple hypotheses testing* or, in more generality, simultaneous statistical inference.

However, this modern paradigm of simultaneous inference fundamentally differs from the classical inference theory built by Pearson, Fisher, Neyman, and others. The classical approach to statistics starts with forming a question or hypothesis and then conducting relevant experiments to collect data. Modern science uses the reverse order - the questions are asked after the data collection, thanks to the ever-evolving computational power and improving technologies that allow us to obtain, store, and analyze massive datasets cheaply and efficiently. To quote Efron (2010a),

"...now the flood of data is accompanied by a deluge of questions, perhaps thousands of estimates or hypothesis tests that the statistician is charged with answering together; not at all what the classical masters had in mind."

The scientific community's response to this foray into data has witnessed an extreme explosion of statistical theory, methodologies, and applications. Many of these new-era concepts have been sophisticated theories while many others are application-specific and tailored for solving simultaneous inference problems in changed circumstances. Before going into the formalism of theoretical discussion, we consider a few specific contexts involving multiple testing problems.

(a) *Differential Gene expression studies through DNA microarrays.*

Large-scale genomic studies are most effectively carried out through DNA microarray methodologies (Trevino et al., 2007). Microarrays or sequencing experiments assess the gene expression levels of the cancer patients and the normals. Perhaps the foremost question here is: which genes have significantly different mean expression levels in these two populations ? Identifying these interesting genes is also the principal goal of differential gene expression studies in oncology. One way to discover this set is to test equality of means for each gene and note those that pass some significance cutoff. Naturally, a multiple-testing situation comes in!

(b) *Recommending the best drug in confirmatory clinical trials.*

The regulatory agencies require the pharmaceutical companies to provide conclusive evidence that the proposed treatment is better than the existing drug (control). Confirmatory clinical trials test several candidate drugs against the control (Dmitrienko et al., 2010). A type I error in any of these comparisons results in the wrong recommendation of an inferior drug. With an increasing number of new drugs in the research, the probability of wrong recommendations also increases. Thus, the company should control the chance of making at least one type I error to monitor the occurrence of the wrong recommendations.

(c) *Detecting outperforming stocks in the market.*

Consider the top 200 companies in Nifty500. Suppose one is interested in detecting which of them, if any, outperformed the market in a certain period. One can perform 200 separate t-tests to compare the differences between these 200 stocks' mean returns and the mean return of the market index, each at a pre-specified precision level α (say .05). Such multiple t-tests are used quite frequently in practice. However, this method does not address the multiplicity effect arising from several hypotheses. Note that even if no company actually outperforms the market index, about 5% or ten companies might have an outstanding performance by chance. In words of Grinold and Kahn (2000),

"The fundamental goal of performance analysis is to separate skill from luck. But, how do you tell them apart? In a population of 1000 investment managers, about 5 percent, or 50, should have exceptional performance by chance alone. None of the successful managers will admit to being lucky; all of the unsuccessful managers will cite bad luck."

Gauriot and Page (2019) raised a similar concern using a dataset on the performance of individual football players. They also investigated specific situations where luck may be overrewarded in performance evaluation.

These examples elucidate simultaneous inference problems frequently arise in many scientific avenues and also illustrate that these problems can not be addressed by simply applying the classical methods separately. They need new statistical approaches which address the multiplicity effect also. These multiple testing procedures are the basic premise of this thesis.

1.2 Family of Comparisons

Intuitively, it is much more efficient, statistically and economically, to have one large experiment to address many related questions rather than having individual experiments for them. The previous section also illustrates that testing many hypotheses separately often fails because it leads to too many false discoveries.

Tukey (1953) introduced the term "family". Hochberg and Tamhane (1987) remark that a family is "a collection of inferences for which it is meaningful to take into account some combined measure of errors". A concise review of the principles and issues in family selection is in Saunders (2014). In the microarrays example, the genes and, in the second example, all new drugs tested against the control constitute the family. In the stock market example, the 200 companies build the family.

A family of hypotheses is called *hierarchical* (Rom and Holland, 1995; Guo, 2007) if there are at least two hypotheses such that one implies the other. Otherwise, the family is called *non-hierarchical*. We consider non-hierarchical families in this thesis.

The following section introduces the existing error rates in simultaneous inference while its subsequent section reviews the popular multiple testing procedures.

1.3 Error Rates

In the classical multiple testing framework, the problem is to test the n hypotheses

$$H_{0i} \quad \text{vs} \quad H_{1i}, \quad i \in \mathcal{I} := \{1, \dots, n\},$$

based on some dataset. Here the i-th testing problem concerns testing the i'th null H_{0i} against the i'th alternative H_{1i}. The intersection null hypothesis (also called the global null) $H_0 = \bigcap_{i=1}^{n} H_{0i}$ states that each H_{0i} is true. Let $\mathcal{A}$ be the set of true nulls. So, under the global null, $\mathcal{A}$ is $\mathcal{I}$.

1.3.1 True and False Discoveries

The simultaneous testing of $H_{01}, H_{02}, \ldots, H_{0n}$ is implemented though some decision rule $\mathcal{D}$ that, for each of the n tests, accepts or rejects it. Such decision rules are called *multiple testing procedures*, which reject some of the hypotheses based on the observations. The general decision pattern of a multiple testing procedure (MTP, henceforth) for n tests is outlined in Table 1.1.

Table 1.1: Classification of the n hypotheses by a MTP $\mathcal{D}$

	Null not rejected	Null Rejected	**Total**
Actual Null	U_n	V_n	n_0
Actual Non-null	T_n	S_n	n_1
Total	$n - R_n$	R_n	n

In Table 1.1 and throughout this dissertation, V_n denotes the number of type I errors (i.e, *false discoveries*). Likewise, T_n denotes the number of type II errors (i.e, *false non-discoveries*). R_n is the total number of hypotheses rejected. Here, n is fixed and known. Each of the quantities U_n, V_n, T_n, S_n, R_n is a random variable. However, among these, we can only observe the value of R_n. The quantities U_n, V_n, T_n, S_n all are unobservable. Hence, the row marginals n_0 and n_1 re unknown parameters. Thus, in a sense, Table 1.1 portrays a hypothetical classification of $\mathcal{D}$'s performance based on an omniscient oracle: n_0 cases were true null, n_1 were actually non-null, out of the actual nulls $\mathcal{D}$ had V_n incorrect decisions while out of the non-nulls, $\mathcal{D}$ made T_n incorrect decisions.

1.3.2 Type I Error Rates and their Control

Several different definitions (or metrics) for the type I error rate based on V_n and R_n have been proposed in the simultaneous inference literature. In the following, T denotes a MTP.

- *Family-wise error rate (FWER).*

This is the most classic notion of type-I error in simultaneous statistical inference. This is the probability of making at least one type I error, i.e.,

$$FWER_T := \mathbb{P}(V_n(T) \geq 1).$$

FWER is a widely considered frequentist approach in multiple testing. Controlling FWER has been a traditional concern in many multiple testing problems. This tradition is reflected in the books by Hochberg and Tamhane (1987), Hsu (1996), Miller (1981), Westfall and Young (1993), and the review by Tamhane (1996). Although a large portion of this thesis focuses on studying the finite-sample and asymptotic performances of various FWER controlling procedures, we mention some other type I error criteria for completeness.

- *Per-comparison error rate (PCER).*

  It is the expected value of the proportion of false positives among the n tests, i.e.,

$$PCER_T := \frac{\mathbb{E}[V_n(T)]}{n}.$$

- *Per-family error rate (PFER).*

  It is the expected number of false positives, i.e.,

$$PFER_T := \mathbb{E}[V_n(T)].$$

- *False discovery proportion (FDP).*

  Instead of focusing on V_n, one might be interested in the proportion of false discoveries among total discoveries:

$$FDP_T := \frac{V_n(T)}{\max\{R_n(T), 1\}}.$$

- *False discovery rate (FDR).*

  Dickhaus (2014) remarks that multiple testing witnessed the start of a new era when Benjamini and Hochberg (1995) introduced the FDR:

$$FDR_T := \mathbb{E}(FDP_T) = \mathbb{E}\left[\frac{V_n(T)}{R_n(T)} \mid R_n(T) > 0\right] \cdot \mathbb{P}(R_n(T) > 0).$$

  Since its inception in their pioneering work, FDR has enjoyed considerable attention in mainstream statistical research, and appears to have gained "accepted methodology" stature in scientific journals (Efron, 2010a).

- *Positive False discovery rate (pFDR).*

  Storey (2002, 2003) introduced the notion of pFDR:

$$pFDR_T := \mathbb{E}\left[\frac{V_n(T)}{R_n(T)} \mid R_n(T) > 0\right].$$

  The term *positive* stands for the conditioning on at least one positive finding (i.e, discoveries). Storey (2003) showed that when the test statistics have a mixture distribution, then the pFDR may be represented as a Bayesian posterior probability.

- *Marginal False discovery rate (mFDR).*

Another closely related type-I error rate criterion to FDR is the following:

$$mFDR_T := \frac{\mathbb{E}[V_n(T)]}{\mathbb{E}[R_n(T)]}.$$

It is also called proportion of expected false positives (PEFP) (Dudoit and Laan, 2008).

- *Generalized family-wise error rate (k-FWER).*
  Lehmann and Romano (2005) generalized the classical FWER to the following:

$$k\text{-}FWER_T := \mathbb{P}(V_n(T) \geq k).$$

Here $k \in \mathcal{I}$ is pre-specified or user-supplied. Putting $k = 1$ gives the FWER. Larger values of k give less conservative results.

- *Generalized false discovery rate (k-FDR).*
  Sarkar (2007) introduced $k\text{-}FDR = \mathbb{E}(k\text{-}FDP)$ where

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

$k = 1$ gives FDR.

Measures like FDR, k-FWER and k-FDR are often considered as less stringent approaches than the FWER for finding the significant few from the insignificant many effects tested. However, the various error criteria are suitable in different situations. The FDR, being a more liberal criteria, is appropriate in exploratory studies. The FWER or k-FWER are more often used in confirmatory studies or cases where a type I error incurs a huge loss, e.g., confirmatory clinical trials (Dmitrienko et al., 2010; Ren, 2021).

1.3.3 Strong and Weak Control

Let $e(V_n, R_n)$ be a general type-I error rate. A MTP T controls this error rate *weakly* at level $\alpha \in (0, 1)$ if $e(V_n, R_n)_T \leq \alpha$ under the global null hypothesis.

Alternatively, we say that T controls $e(V_n, R_n)$ *strongly* at level $\alpha \in (0, 1)$ if $e(V_n, R_n)_T \leq \alpha$ under any configuration of true and false null hypotheses.

In general, controlling an error rate strongly is desirable, since the configuration of null hypotheses is usually unknown.

1.3.4 Relations among type I error rates

From the earlier definitions, we note that the following inequalities hold for any MTP T and any configuration of true and false null hypotheses:

$$
\begin{aligned}
&PCER_T \le FDR_T \le FWER_T \le PFER_T,\\
&k\text{-}FDR_T \le FDR_T \le pFDR_T,\\
&k\text{-}FDR_T \le k\text{-}FWER_T \le FWER_T,\\
&PCER_T \le mFDR_T,\\
&k\text{-}FWER_T \le \frac{PFER_T}{k}.
\end{aligned}
$$

The above inequalities present the relative conservativeness of different criteria. For example, any FWER controlling procedure also controls FDR and k-FWER for $k > 1$. The smaller error criteria typically induce less stringent methods in the sense of allowing more rejections. Also, under the global null hypothesis, FDR and FWER are same. So, any FDR controlling procedure also controls FWER weakly.

1.3.5 Type II Error Rates

One often wishes to compare MTPs which control the same type I error criteria at level α. For this, we additionally need a concept of type II error rate, or equivalently, power. A few commonly used notions of power (Dudoit and Laan, 2008; Grandhi, 2015; Ramsey, 1978) are:

- *Disjunctive power (AnyPwr).*

  This is the probability of rejecting at least one false null:

$$AnyPwr_T := \mathbb{P}(S_n(T) \ge 1).$$

This is appropriate in studies aiming to identify at least one existing effect, e.g. in union intersection settings (Bretz et al., 2011).

- *Conjunctive power (AllPwr).*

  This is the probability of rejecting all false null hypotheses:

$$AllPwr_T := \mathbb{P}(S_n(T) = n_1).$$

This is appropriate in studies aiming to identify all existing effects, e.g. in intersection union settings (Bretz et al., 2011).

- *Average power (AvgPwr).*

It is defined as

$$AvgPwr_T := \frac{\mathbb{E}[S_n(T)]}{n_1}, \quad \text{when } n_1 > 0.$$

These three notions of power have also been termed minimal power, complete power, and proportional power, respectively (Westfall et al., 1999). We also note that, for any MTP T,

$$AllPwr_T \leq AnyPwr_T \quad and \quad AllPwr_T \leq AvgPwr_T.$$

- *True discovery rate (TDR).*

  It is the expected value of the proportion of true discoveries among total discoveries.

$$TDR_T := \mathbb{E}\left[\frac{S_n(T)}{R_n(T)} \mid R_n(T) > 0\right] \cdot \mathbb{P}(R_n(T) > 0) = \mathbb{P}\left(R_n(T) > 0\right) - FDR_T.$$

In a sense, TDR is a power analogue of the FDR.

- *False non-discovery rate (FNR).*

  Genovese and Wasserman (2002) considered the expected value of the proportion of non-discoveries among total non-discoveries,

$$FNR_T := \mathbb{E}\left[\frac{T_n(T)}{n - R_n(T)} \mid R_n(T) < n\right] \cdot \mathbb{P}(R_n(T) < n).$$

They studied new MTPs that incorporate both FDR and FNR. Sarkar (2004) referred to this as the *false negatives rate*.

Storey (2003) proposed *positive false non-discovery rate* (pFNR). This is the conditional expectation $\mathbb{E}\left[\frac{T_n(T)}{n-R_n(T)} \mid R_n(T) < n\right]$. He established a connection between multiple testing and classification theory in terms of a combination of pFDR and pFNR.

Sun and Cai (2007) considered minimization of the following quantity:

$$mFNR_T := \frac{\mathbb{E}[T_n(T)]}{\mathbb{E}[n - R_n(T)]},$$

subject to controlling mFDR at a given level.

1.4 Multiple Testing Procedures

A MTP for the simultaneous testing of n hypotheses is a decision rule providing rejection regions for each of the n hypotheses. So, for each of the n tests, a MTP gives a set of values (i.e, the critical region for the corresponding null).

1.4.1 Types of Multiple Testing Procedures

The literature on MTPs is ever-increasing. However, we can still classify the proposed methods based on some general considerations. For example, MTPs can be classified based on their different approaches in defining rejection regions. Some MTPs are built based on the marginal distributions of the statistics while some others also consider their dependence. These are respectively *Margin-based MTPs* and *Multivariate MTPs*:

(a) *Margin-based (or Marginal) MTPs.*

This class involves modelling the marginal distributions of the n test statistics. (Dickhaus, 2014).

(b) *Multivariate (or Joint) MTPs.*

This class considers the joint distribution of all the statistics. Their decision rules (i.e, the rejection regions) involve exact or approximate calculations of the quantiles of this joint distribution, obtained through multivariate CLTs or resampling (Dickhaus, 2014).

This thesis focuses on the class of marginal MTPs, which include single-step and step-wise procedures.

1.4.2 Single-step Procedures

Single-step MTPs test individual null hypothesis at (local) level α^*, where $\alpha^* \leq \alpha$ is due to a multiplicity correction of α. These procedures are extremely easy to implement: we just need to calculate marginal p-values $P_1, \ldots, P_n$ and we reject H_{0i} if and only if $P_i < \alpha^*$.

- *The Bonferroni Procedure.*

This is one of the most widely used FWER-controlling method. This method uses $\alpha^* = \alpha/n$:

$$FWER_{Bon} = \mathbb{P}\left(\bigcup_{i \in \mathcal{A}} \{P_i \leq \alpha/n\}\right).$$

The Bonferroni method controls the FWER under any dependency structure among the n p-values. This simple procedure requires no distributional assumption. However, it has the disadvantage that α/n is very small for large n.

- *The Sidak Procedure (Šidák, 1967).*

This method uses $\alpha^* = 1 - (1-\alpha)^{1/n}$:

$$FWER_{Sidak} = \mathbb{P}\left(\bigcup_{i \in \mathcal{A}} \{P_i \leq 1 - (1-\alpha)^{1/n}\}\right).$$

It controls FWER under independence.

1.4.3 Stepwise Procedures

Single-step MTPs compare the individual test statistics to the corresponding cut-offs simultaneously, and they stop after performing this simultaneous 'joint' comparison. Often stepwise methods (Holm, 1979; Rom, 1990) possess greater power than the single-step procedures, while still controlling FWER (or, in general, the error rate under consideration) at the desired level.

Consider the set

$$\mathcal{S}_n = \{\mathbf{t} = (t_1, \ldots, t_n) \in \mathbb{R}^n : 0 \leq t_1 \leq \ldots \leq t_n \leq 1\}.$$

A p-value based step-down MTP uses a vector of cutoffs $\mathbf{u} = (u_1, \ldots, u_n) \in \mathcal{S}_n$, and works as follows. The step-down MTP compares the most significant p-value $P_{(1)}$ with the smallest u-value u_1 at first and so on. More formally, let $m_1 = \max\left\{i : P_{(j)} \leq u_j \text{ for all } j = 1, \ldots, i\right\}$. Then the step-down MTP based on critical values $\mathbf{u}$ rejects $H_{0(1)}, \ldots, H_{0(m_1)}$.

Example 1. The Bonferroni method is a step-down MTP with $u_i = \alpha/n$, $i \in \mathcal{I}$.

Example 2. The Sidak method is a step-down MTP with $u_i = 1 - (1-\alpha)^{1/n}$, $i \in \mathcal{I}$.

Example 3. Holm (1979) proposed a popular step-down MTP with $u_i = \alpha/(n-i+1), i \in \mathcal{I}$. This method strongly controls FWER at α under arbitrary dependence.

Example 4. Benjamini and Liu (1999a) introduced a step-down MTP with

$$u_i = \min\left(1, \frac{n\alpha}{(n-i+1)^2}\right), \quad i \in \mathcal{I}.$$

Example 5. Benjamini and Liu (1999b) studied another step-down MTP with

$$u_i = 1 - \left[1 - \min\left(1, \frac{n\alpha}{n-i+1}\right)\right]^{1/(n-i+1)}, \quad i \in \mathcal{I}.$$

Example 6. Benjamini and Liu (1999b) mentioned a Holm-type procedure with critical values

$$u_i = 1 - (1-\alpha)^{1/(n-i+1)}, \quad i \in \mathcal{I}.$$

The step-up MTP also utilizes set of critical values, say $\mathbf{u} = (u_1, \ldots, u_n) \in \mathcal{S}_n$. But the step-up method is inherently different from the step-down method because it starts by comparing the least significant p-value $P_{(n)}$ with the largest u-value u_n and so on. Formally, the step-up MTP based on critical values $\mathbf{u}$ rejects the hypotheses

$H_{0(1)}, \ldots, H_{0(m_2)}$, where $m_2 = \max\left\{i : P_{(i)} \leq u_i\right\}$. If such a m_2 does not exist, then the MTP accepts each null hypothesis.

We make a remark on the nomenclature of the step-down and step-up MTPs. The direction 'up' or 'down' refers to the order of the significance of the p-values in which the method proceeds. The step-down MTP steps toward the less significant p-values (i.e, it 'steps down'), while the step-up MTP steps toward more significant p-values (i.e, it 'steps up').

Example 7. The Bonferroni correction is also a step-up MTP, where $u_i = \alpha/n, i \in \mathcal{I}$.

Example 8. The Sidak method is a step-up procedure with $u_i = 1 - (1-\alpha)^{1/n}$, $i \in \mathcal{I}$.

Example 9. Hochberg (1988) proposed a popular step-up MTP which uses the same cut-offs as Holm MTP (i.e, $u_i = \alpha/(n-i+1)$). This strongly controls FWER at α under independence or positive regression dependence.

Example 10. The classic Benjamini-Hochberg method (Benjamini and Hochberg, 1995) is a step-up MTP with $u_i = i\alpha/n$. Benjamini and Hochberg (1995) originally showed a conservative control of the FDR of their procedure at $n_0\alpha/n$ under independence of the underlying test statistics. The exact control was shown later in Benjamini and Yekutieli (2001); Finner and Roters (2001b); Sarkar (2002).

Example 11. Benjamini and Yekutieli (2001) proved that the step-up MTP with

$$u_i = \frac{i\alpha}{n} \cdot \frac{1}{D}, \quad D = \sum_{i=1}^{n} \frac{1}{i},$$

controls FDR at level α under arbitrarily dependent test statistics.

1.5 Dependence and Our Multiple Testing Framework

Several authors have proposed a plethora of different methodologies to control various types of error rates. Most of these classical procedures for controlling FDR or FWER were introduced under independence or some form of weak dependence among the concerned variables. However, large-scale multiple testing problems arising in various scientific disciplines often study correlated variables simultaneously. For example, in microRNA expression data, several genes may cluster into groups through their transcription processes and possess high correlations. The data observed from different locations and time periods in public health studies are generally spatially or serially correlated. fMRI studies and multistage clinical trials also involve variables with complex and unknown dependencies. Consequently, the study of the effect of correlation on dependent test statistics in simultaneous inference problems has attracted considerable attention recently.

1.5.1 Existing Approaches towards Dependence

Benjamini and Yekutieli (2001) introduced a special dependency property *positive regression dependency on a subset (PRDS)*. The notion of PRDS involves increasing sets.

Definition 1. We call $D \subset \mathbb{R}^k$ to be an *increasing set* if $\mathbf{a} \in D$ and $\mathbf{b} \geq \mathbf{a}$ imply that $\mathbf{b} \in D$.

Definition 2. Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be the vector of test statistics. We say that the *PRDS property* holds on $\mathcal{A}$ if for any increasing set D, and for each $i \in \mathcal{A}$,

$$\mathbb{P}\{\mathbf{X} \in D \mid X_i = x\}$$

is nondecreasing in x.

Benjamini and Yekutieli (2001) established that the BH method controls FDR under the PRDS property. Let $\psi : [0,1]^n \to \mathbb{R}$ be any coordinate-wise non-decreasing function. Then the PRDS property holds if

$$\mathbb{E}\left[\psi\left(p_1, \ldots, p_m\right) \mid p_i = u\right] \text{ is non-decreasing in } u \;\; \forall i \in \mathcal{A}.$$

Sarkar (2008) gave a slightly more relaxed condition, termed as *positive dependence*, which holds if

$$\mathbb{E}\left[\psi\left(p_1, \ldots, p_m\right) \mid p_i \leq u\right] \text{ is non-decreasing in } u \;\; \forall i \in \mathcal{A}.$$

He showed that BH method also controls the FDR under positive dependence. Sarkar (2002) established some general results on FDR control under dependence.

Storey and Tibshirani (2003) studied the estimation of FDR under dependence. Blanchard and Roquain (2009) proposed FDR-controlling adaptive step-up methods under positive dependence and unspecified dependence. Finner and Roters (2001a) studied the expected type I errors of single-step MTPs based on exchangeable test statistics. Fan et al. (2012) introduced a method of tackling dependent test statistics with a known dependence structure. Fan and Han (2016) extended this to unknown dependence structures. Qiu et al. (2005) demonstrated that many FDR controlling procedures lose power significantly under dependence. Leek and Storey (2008) developed a framework for large-scale simultaneous inference under dependence.

1.5.2 Motivation and a Brief Overview of Our Work

Controlling FDR is appropriate in many scenarios, e.g., in "preliminary studies where it is acceptable for a certain proportion of discoveries to be false in exchange for an increase in overall discoveries as compared to FWER controlling methods" (Saunders, 2014). FWER, on the contrary, has a confirmatory nature (Hochberg and Tamhane, 1987). Also, the existing literature lacks the study of the performances of FWER or generalized FWER controlling procedures under dependent setups. For these reasons, this thesis concentrates mainly on FWER and generalized FWER controlling procedures. However, in the subsequent chapters, we shall see that similar asymptotic results also hold for FDR.

Guo and Sarkar (2020) presented adaptive versions of the BH and the Bonferroni MTP in a block dependence structure which control the FDR and FWER, respectively. They elucidated that their proposed methods can handle the underlying covariance structure more efficiently than the corresponding traditional adaptive method.

Proschan and Shaw (2011) studied the asymptotics of Bonferroni procedure for equicorrelated normal test statistics, where the correlation ρ_n approaches 0 as n approaches infinity. Das and Bhandari (2021) have established that the Bonferroni FWER is asymptotically convex in correlation ρ under the equicorrelated Gaussian framework (discussed in the following subsection). Consequently, they show that the Bonferroni FWER is bounded above by $\alpha(1-\rho)$, α being the target level. However, this bound fails in equicorrelated setups with small and moderate dimensions. There is little literature that elucidates the magnitude of the conservativeness of Bonferroni's method in those scenarios. In *Chapter 2*, we address this research gap in a unified manner by establishing upper bounds on Bonferroni FWER in non-negatively correlated non-asymptotic setups.

In *Chapter 3*, We derive similar upper bounds for the generalized FWERs and propose an improved k-FWER controlling procedure. Towards this, we establish an inequality related to the probability that at least k among n events occur, which extends and sharpens the classical ones. The computation of this probability arises in various contexts, e.g., in reliability problems of communication networks. Our results might be insightful in those areas, too.

In *Chapter 4*, we improve the main result of Das and Bhandari (2021) by showing that the Bonferroni FWER asymptotically goes to zero for any strictly positive ρ. We also extend this to arbitrarily correlated setups where the limiting infimum of the correlations is strictly positive. These results elucidate that Bonferroni's procedure becomes extremely conservative for large-scale multiple-testing problems under dependence.

Huang and Hsu (2007) remark that stepwise decision rules based on modeling of the dependence structure are in general superior to their counterparts that do not consider the correlation. Finner and Roters (2002) studied the number of falsely rejected hypotheses in single-step, step-down and step-up methods under independence. However, the role of correlation on the limiting behavior of the FWER for stepwise procedures is less studied. Also, the existing literature lacks theoretical justifications for why FWER methods fail in large-scale problems. *Chapter 5* addresses this problem by theoretically investigating the limiting FWER values of general step-down procedures under the correlated normal setup. These results provide new insights into the behavior of step-down decision procedures. By establishing the limiting performances of commonly used step-up methods, e.g., the BH MTP and the Hochberg MTP, we have elucidated that the class of step-up procedures does not possess a similar universal asymptotic zero result as obtained in the case of step-down procedures. It is also noteworthy that most of our results are very general since they accommodate any combination of false and true null hypotheses. We have also obtained the limiting powers of the stepwise procedures.

Finally, we consider the simultaneous inference problem in a sequential framework in *Chapter 6*. The mainstream sequential simultaneous inference literature has traditionally focused on the independent setup. However, there is little work studying the multiple inference problem in a sequential framework where the observations corresponding to the various streams are dependent. We consider the classical means-testing problem in an equicorrelated Gaussian and sequential framework. We focus on sequential test procedures that control the type I and type II familywise error probabilities at pre-specified levels. We establish that our proposed MTPs have the optimal average sample numbers asymptotically, as the two types of familywise error probabilities approach zero at arbitrary rates. Towards this, we elucidate that the ratio of the average sample number of the proposed rule and that of the classical SPRT goes to one asymptotically, thus illustrating their connection.

1.5.3 Correlated Gaussian sequence model

This thesis views the simultaneous inference problem through a Gaussian sequence model framework (Das and Bhandari, 2021; Delattre and Roquain, 2011; Finner et al., 2007, 2009; Proschan and Shaw, 2011). Suppose we have n observations

$$X_i \sim N\left(\mu_i, 1\right), \quad i \in \mathcal{I} := \{1, \ldots, n\},$$

where the X_i 's are dependent. The variances are considered to be unity since the literature on the multiple testing theory often assumes that the variances are known (see, e.g., Abramovich et al. (2006); Bogdan et al. (2011); Das and Bhandari (2021, 2023); Delattre and Roquain (2011); Donoho and Jin (2004); Proschan and Shaw (2011)).

We wish to test:

$$H_{0i} : \mu_i = 0 \quad \text{vs} \quad H_{1i} : \mu_i > 0, i \in \mathcal{I}.$$

The global null $H_0 = \bigcap_{i=1}^{n} H_{0i}$ states that each μ_i is zero. This thesis studies the finite-sample and asymptotic properties of a broad class of MTPs under two dependent setups:

(a) The equicorrelated setup:

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

(b) The non-negatively correlated setup:

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

The equicorrelated setup (Cohen et al., 2009; Das and Bhandari, 2020, 2021; Delattre and Roquain, 2011; Dickhaus, 2008; Finner and Roters, 2001a; Finner et al., 2007, 2009; Proschan and Shaw, 2011; Roy and Bhandari, 2024; Sarkar, 2007) is the intraclass covariance matrix model, characterizing the exchangeable situation. Simultaneous testing of normal means under equicorrelated frameworks has witnessed considerable attention in recent years. Although the equicorrelated setup is a special case of the non-negatively correlated case, we are mentioning them separately since, often the proof of a result in the general case is based on the corresponding results in the equicorrelated case.

The equicorrelated setup also encompasses the problem of comparing a control against several treatments. To see this, let $\bar{Y}_i \sim N\left(\gamma_i, \sigma^2/a_i\right), i = 0, \ldots, n$, be independent sample means with known variance $\sigma^2 > 0$, $a_1 = \cdots = a_n$ and $\gamma_i \geq \gamma_0$ for $i = 1, \ldots, n$. Suppose we wish to test

$$\tilde{H}_i : \gamma_i = \gamma_0 \quad \text{vs} \quad \tilde{K}_i : \gamma_i > \gamma_0 \quad \text{for} \quad i = 1, \ldots, n$$

based on the test statistics

$$T_i = \left[\frac{1}{a_0} + \frac{1}{a_1}\right]^{-1/2} \frac{\left(\bar{Y}_i - \bar{Y}_0\right)}{\sigma}, \quad i = 1, \ldots, n.$$

Then,

$$\mathbb{E}(T_i) = \left[\frac{1}{a_0} + \frac{1}{a_1}\right]^{-1/2} \frac{(\gamma_i - \gamma_0)}{\sigma} = \mu_i \;\; \text{(say)}.$$

Also, $\text{Var}\,[T_i] = 1$ and

$$\text{Cov}\,(T_i, T_j) = a_1/\left(a_1 + a_0\right) = \rho \quad \text{(say)}.$$

However, many scientific disciplines involve variables with more complex dependence structure (e.g., fMRI studies). These complex dependence scenarios need to be tackled with more general covariance matrices. The arbitrarily correlated setup also includes the successive correlation covariance matrix, which covers change point problems (Cohen et al., 2009).

Throughout this thesis, $M_n(\rho)$ denotes the $n \times n$ matrix having each diagonal entry 1 and each off-diagonal entry $\rho \in [0, 1]$. Also, Σ_n denotes the $n \times n$ correlation matrix with (i, j)'th entry equal to $\rho_{ij} \in [0, 1]$, $i \neq j$. We denote the pdf and the cdf of $N(0, 1)$ by $\phi(\cdot)$ and $\Phi(\cdot)$ respectively. The desired significance level is denoted by $\alpha \in (0, 1)$.

Chapter 2

Non-asymptotic Behaviors of FWER in Correlated Normal Distributions [1]

2.1 Introduction

This chapter considers the equicorrelated normal distribution with positive correlation ρ at first. Das and Bhandari (2021) have found that under this setup, FWER(ρ) is convex in ρ as the number of hypotheses approaches infinity. Consequently, they show that the Bonferroni FWER is bounded above by $\alpha(1-\rho)$, α being the desired level. In Chapter 4, we shall show that the Bonferroni FWER(ρ) approaches zero asymptotically for any positive ρ. These works explicate the fact that Bonferroni's procedure becomes very conservative for large-scale multiple testing problems under correlated setups. However, the convergence of FWER to zero is extremely slow and we find little literature which elucidates the magnitude of the conservativeness of Bonferroni's method in a dependent setup with small or moderate dimensions. In this chapter, we bridge this research gap in a unified manner by establishing upper bounds on Bonferroni FWER in the equicorrelated and non-negatively correlated non-asymptotic setups.

Order statistics for exchangeable normal random variables have applications in biometrics (Olkin and Viana, 1995; Viana, 1998). Also, the maximum of exchangeable normal random vector can be used to model the lifetime of parallel systems conveniently. The non-asymptotic bounds on FWER proposed in this chapter provide lower bounds on the cdf of the failure time of parallel systems. It is known that the maximum of n observations from equicorrelated normal distribution has a $(n-1)$ dimensional skew normal distribution

[1]This chapter is based on the publication M. Dey (2024) Behavior of FWER in Normal Distributions, *Communications in Statistics - Theory and Methods*, 53(9), 3211-3225, DOI: 10.1080/03610926.2022.2150826.

(Loperfido et al., 2007). Although finding the cdf of multivariate skew normal distribution is difficult, non-asymptotic bounds on the cdf may be obtained from the results derived in this chapter.

In Section 2.2, we set up the framework and introduce the necessary notation. Section 2.3 contains theoretical results about the bounds on FWER in equicorrelated normal setup. Section 2.4 extends these results to non-negatively dependent setups. We also propose an improved multiple testing procedure utilizing those bounds in Section 2.5. Section 2.6 presents simulation findings. We end this chapter with a brief conclusion in Section 2.7.

2.2 Preliminaries

We consider the Gaussian sequence model introduced in Section 1.5.3:

$$X_i \sim N(\mu_i, 1), \quad i \in \mathcal{I}.$$

The global null $H_0 = \bigcap_{i=1}^{n} H_{0i}$ states that each μ_i is zero, while under the global alternative, at least one μ_i is positive. We reject H_{0i} for large values of X_i (say $X_i > c$ for some cut-off c). Under the global null,

$$FWER = \mathbb{P}_{H_0}\left(\bigcup_{i=1}^{n} \{X_i > c\}\right).$$

We have considered equicorrelated setup at first (in Section 2.3) whereas in Section 2.4, we have dealt with non-negatively correlated setup. In our one-sided setting, the Bonferroni method rejects H_{0i} if $X_i > \Phi^{-1}(1-\alpha/n)(= c_{\alpha,n}, \text{say})$. So, under the global null, the Bonferroni FWER (for the covariance matrix Σ_n) is defined by

$$FWER_{Bon}(n, \alpha, \Sigma_n) = \mathbb{P}_{\Sigma_n}\left(X_i > c_{\alpha,n} \text{ for some } i \mid H_0\right) = \mathbb{P}_{\Sigma_n}\left(\bigcup_{i=1}^{n} \{X_i > c_{\alpha,n} \mid H_0\}\right).$$

For $\rho \in [0, 1]$, We denote $FWER_{Bon}(n, \alpha, M_n(\rho))$ as $FWER_{Bon}(n, \alpha, \rho)$ for simpler notation. Das and Bhandari (2021) show the following under the equicorrelated setup:

Theorem 2.2.1. *Suppose we test each H_{0i} at size α_n. Suppose $\lim_{n\to\infty} n\alpha_n = \alpha \in (0, 1)$. Then, $FWER_{Bon}(n, \alpha, \rho)$ asymptotically is a convex function in $\rho \in [0, 1]$.*

For Bonferroni's procedure, $\alpha_n = \alpha/n$ and thus Theorem 2.2.1 also applies for Bonferroni's method. Moreover, Theorem 2.2.1 results in the following corollary.

Corollary 2.2.1. *Given any* $\alpha \in (0, 1)$ *and* $\rho \in [0, 1]$, $FWER_{Bon}(n, \alpha, \rho)$ *is asymptotically bounded by* $\alpha(1 - \rho)$.

Corollary 2.2.1 shows that Bonferroni procedure controls FWER at a much smaller level than α, when n is very large. We shall later prove a much stronger result than Corollary 2.2.1 (Theorem 4.3.3) in Chapter 4.

Corollary 2.2.1 highlights the fundamental drawback of Bonferroni MTP. We shall call the setup with very large number of hypotheses an asymptotic setup while setups with small or moderate number of hypotheses will be referred to as non-asymptotic setups. We summarize known and new results on FWER of Bonferroni's MTP under various dependent normal setups in Table 2.1.

Table 2.1: Results on Bonferroni FWER

Dependent Setup	**Results on FWER**
Equicorrelated Asymptotic	Corollary 2.2.1 (Das and Bhandari 2021), Theorem 4.3.1, Theorem 4.3.3
General Asymptotic	Theorem 4.3.4
Equicorrelated Non-asymptotic	Theorem 2.3.1,2.3.3,2.3.4,2.3.5,2.3.6 and Corollary 2.3.2
General Non-asymptotic	Theorem 2.4.1,2.4.2,2.4.3,2.4.4 and Corollary 2.4.2

2.3 Bounds on FWER in Equicorrelated Non-asymptotic Setup

The $\alpha(1 - \rho)$ bound fails in equicorrelated setups with small and moderate dimensions, (we shall see this in detail in Section 2.6). We need large number of hypotheses, e.g 100 million to get values of FWER close to zero. Hence, establishing upper bounds on FWER in problems with small and moderate number of hypotheses become relevant. The following result will be crucial towards this.

Theorem 2.3.1. *Under the equicorrelated normal set-up,*

$$FWER_{Bon}(n, \alpha, \rho) \leq \alpha - \frac{n-1}{n} \cdot \frac{\alpha^2}{n} - \frac{n-1}{2\pi} \int_0^\rho \frac{1}{\sqrt{1-z^2}} e^{\frac{-\Phi^{-1}(1-\frac{\alpha}{n})^2}{1+z}} dz.$$

It is noteworthy that this bound holds for any choice of (n, α) and any $\rho \geq 0$. We need two results to establish this theorem.

Lemma 2.3.1. *(Kwerel, 1975) Let $A_1, A_2, \ldots, A_n$ be n events. Let $S_1 = \sum_{i=1}^{n} \mathbb{P}(A_i)$ and $S_2 = \sum_{1 \le i < j \le n} \mathbb{P}(A_i \cap A_j)$. Then, $\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \le S_1 - \frac{2}{n} S_2$.*

This bound on the union of n events is also called the Sobel-Uppuluri upper bound and is the optimal linear bound in S_1 and S_2 (Chen, 2015). The second lemma is regarding the joint distribution function of a bivariate normal distribution:

Lemma 2.3.2 (Monhor (2013))**.** *Suppose $(X, Y) \sim$ Bivariate Normal$(0, 0, 1, 1, \rho)$ with $\rho \ge 0$. Then, for all $x > 0$,*

$$\mathbb{P}(X \le x, Y \le x) = [\Phi(x)]^2 + \frac{1}{2\pi} \int_0^{\rho} \frac{1}{\sqrt{1-z^2}} e^{\frac{-x^2}{1+z}} dz.$$

Proof of Theorem 2.3.1. For $i = 1, \ldots, n$, we define the event $A_i = \{X_i > \Phi^{-1}(1 - \alpha/n) | H_0\}$. So, $\mathbb{P}(A_i) = \mathbb{P}_{H_0}\left[X_i > \Phi^{-1}(1-\alpha/n)\right] = \alpha/n$. This gives $S_1 = \sum_{i=1}^{n} \mathbb{P}(A_i) = n \cdot \alpha/n = \alpha$. Now,

$$\begin{aligned}
&\mathbb{P}(A_i \cap A_j) \\
&= 1 - \mathbb{P}(A_i^c) - \mathbb{P}(A_j^c) + \mathbb{P}(A_i^c \cap A_j^c) \\
&= 1 - (1-\alpha/n) - (1-\alpha/n) + \mathbb{P}_{H_0}\Big(X_i \le \Phi^{-1}(1-\alpha/n), X_j \le \Phi^{-1}(1-\alpha/n)\Big) \\
&= \frac{2\alpha}{n} - 1 + (1-\alpha/n)^2 + \frac{1}{2\pi}\int_0^{\rho} \frac{1}{\sqrt{1-z^2}} e^{\frac{-\Phi^{-1}(1-\frac{\alpha}{n})^2}{1+z}} dz \quad \text{(using Lemma 2.3.2)} \\
&= \frac{\alpha^2}{n^2} + \frac{1}{2\pi}\int_0^{\rho} \frac{1}{\sqrt{1-z^2}} e^{\frac{-\Phi^{-1}(1-\frac{\alpha}{n})^2}{1+z}} dz
\end{aligned}$$

This gives

$$S_2 = \binom{n}{2} \cdot \left[\frac{\alpha^2}{n^2} + \frac{1}{2\pi}\int_0^{\rho} \frac{1}{\sqrt{1-z^2}} e^{\frac{-\Phi^{-1}(1-\frac{\alpha}{n})^2}{1+z}} dz\right].$$

The rest is obvious from Lemma 2.3.1 since $FWER_{Bon}(n, \alpha, \rho) = \mathbb{P}(\bigcup_{i=1}^{n} A_i)$. □

Remark 1. *Yang et al. (2016) established the following two upper bounds on the union of n events, for any $\mathbf{c} = (c_1, \ldots, c_n)$ with $\sum_{i=1}^{n} c_i = 1$ and $c_i > 0$ for each i:*

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \le \left(\frac{1}{\min_j c_j} + 1\right) \sum_{i=1}^{n} c_i \mathbb{P}(A_i) - \frac{1}{\min_j c_j} \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \mathbb{P}(A_i \cap A_j).$$

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \le \min_i \left\{\frac{\sum_j c_j \mathbb{P}(A_i \cap A_j) - (\min_j c_j)\, \mathbb{P}(A_i)}{1 - \min_j c_j}\right\}$$

$$- \frac{1}{(\min_j c_j)(1 - \min_j c_j)} \left[\sum_{i=1}^{n} c_i \mathbb{P}(A_i) - \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \mathbb{P}(A_i \cap A_j) \right].$$

They conjectured that the optimal upper bounds in the above two classes are achieved at $\mathbf{c} = \frac{1}{n}\mathbf{1}$. *In the proof of Lemma 2.3.2 we have seen that* $\mathbb{P}(A_i)$ *is same for each* i *and* $\mathbb{P}(A_i \cap A_j)$ *is same for each pair* (i, j), $i \neq j$ *in the equicorrelated setup. In this case the above two upper bounds by Yang et al. become identical and reduce to the quantity*

$$\mathbb{P}(A_i) + \frac{\mathbb{P}(A_i) - \mathbb{P}(A_i \cap A_j)}{\min_j c_j} \tag{2.1}$$

which is a decreasing function in $\min_j c_j$. *Therefore, their conjecture is proved affirmatively when* $\mathbb{P}(A_i)$ *is same for each* i *and* $\mathbb{P}(A_i \cap A_j)$ *is same for each pair* (i, j), $i \neq j$. *It is also noteworthy that the quantity in (2.1) is same as the upper bound given by Lemma 2.3.1 in this case.*

Corollary 2.3.1. *Under the equicorrelated normal set-up, if* $\rho \leq \alpha/n$,

$$FWER_{Bon}(n, \alpha, \rho) \leq \alpha - \frac{n-1}{n} \cdot \alpha\rho.$$

Hence, throughout this chapter, we assume that $\rho \geq \alpha/n$. We observe that the bound mentioned in Theorem 2.3.1 involves a definite integral which is very difficult to evaluate analytically. As we are interested in obtaining upper bounds for FWER, it is enough if we can find a lower bound to the integral. Towards this, we show the following theorem which will be crucial to obtain a lower bound to the integral mentioned in Theorem 2.3.1.

Theorem 2.3.2. *Suppose* $(X, Y) \sim \text{Bivariate Normal}(0, 0, 1, 1, \rho)$ *with* $\rho \geq 0$. *Then, for all* $x \geq 2$,

$$\mathbb{P}(X \leq x, Y \leq x) \geq [\Phi(x)]^2 + \frac{1}{2\pi} \cdot \sin^{-1}\rho \cdot e^{-\frac{x^2}{1+\frac{\rho}{2}}}.$$

We use two well-known inequalities to prove this theorem.

Lemma 2.3.3 (Chebyshev Integral Inequality). *Let* f *and* g *be two nonnegative integrable functions and synchronous on a bounded interval* $[a, b]$, *i.e*

$$\forall x, y \in [a, b], \qquad [f(x) - f(y)] \cdot [g(x) - g(y)] \geq 0.$$

Then,

$$(b - a) \cdot \int_a^b f(x)g(x)dx \geq \int_a^b f(x)dx \cdot \int_a^b g(x)dx.$$

Lemma 2.3.4 (Hermite-Hadamard Integral Inequality)**.** *Let $f : [a, b] \to \mathbb{R}$ be a convex function. Then,*

$$\int_a^b f(x)dx \geq (b-a) \cdot f\left(\frac{a+b}{2}\right).$$

Proof of Theorem 2.3.2. Suppose $(X, Y) \sim$ Bivariate Normal$(0, 0, 1, 1, \rho)$ with $\rho \geq 0$. Then, from Lemma 2.3.2,

$$\forall x > 0 \qquad \mathbb{P}(X \leq x, Y \leq x) = [\Phi(x)]^2 + \frac{1}{2\pi}\int_0^\rho \frac{1}{\sqrt{1-z^2}} e^{\frac{-x^2}{1+z}} dz.$$

It can be easily shown that the functions $\dfrac{1}{\sqrt{1-z^2}}$ and $e^{\frac{-x^2}{1+z}}$ have same monotony in $z \in [0, 1]$, i.e are synchronous on $[0, 1]$. Using lemma 2.3.3, we obtain

$$\int_0^\rho \frac{1}{\sqrt{1-z^2}} e^{\frac{-x^2}{1+z}} dz \geq \frac{1}{\rho}\int_0^\rho \frac{1}{\sqrt{1-z^2}} dz \cdot \int_0^\rho e^{\frac{-x^2}{1+z}} dz = \frac{\sin^{-1}\rho}{\rho} \cdot \int_0^\rho e^{\frac{-x^2}{1+z}} dz. \qquad (*)$$

The function $e^{\frac{-x^2}{1+z}}$ is convex in z if $z \leq \frac{x^2}{2} - 1$. Now, $0 \leq z \leq \rho \leq 1$. So, $z \leq \frac{x^2}{2} - 1$ holds if $x \geq 2$. Hence, $e^{\frac{-x^2}{1+z}}$ is convex in $z \in [0, 1]$ for $x \geq 2$. Applying Lemma 2.3.4 on this function, we get,

$$\forall x \geq 2, \qquad \int_0^\rho e^{\frac{-x^2}{1+z}} dz \geq \rho \cdot e^{-\frac{x^2}{1+\frac{\rho}{2}}}.$$

Combining this with (*), we get, for $x \geq 2$,

$$\int_0^\rho \frac{1}{\sqrt{1-z^2}} e^{\frac{-x^2}{1+z}} dz \geq \sin^{-1}\rho \cdot e^{-\frac{x^2}{1+\frac{\rho}{2}}}.$$

The rest is obvious from Lemma 2.3.2. □

Remark 2. *Monhor (2013) obtained the following inequality for positively correlated bivariate normal distribution function using Lemma 2.3.2.*

$$\mathbb{P}(X \leq x, Y \leq x) \geq [\Phi(x)]^2 + \frac{1}{2\pi} \cdot \sin^{-1}\rho \cdot e^{-x^2} \quad \forall x > 0.$$

Theorem 2.3.2 provides a sharper inequality for $x \geq 2$.

Theorem 2.3.2 can be used to establish the following.

Corollary 2.3.2. *Under the equicorrelated normal set-up, if $x = \Phi^{-1}(1 - \frac{\alpha}{n}) \geq 2$,*

$$FWER_{Bon}(n, \alpha, \rho) \leq \alpha - \frac{n-1}{n} \cdot \frac{\alpha^2}{n} - \frac{n-1}{2\pi} \cdot \sin^{-1}\rho \cdot e^{-\frac{x^2}{1+\frac{\rho}{2}}}.$$

We shall write x for $\Phi^{-1}(1-\alpha/n)$ from now on. We observe from simulation study that, the upper bound $\alpha(1-\rho)$ given by Corollary 2.2.1 holds for any nonnegative value of ρ when $n \geq 10000$ and $\alpha \geq 0.01$. When $n = 10000$ and $\alpha = 0.01$, we have $x = 4.42$. This, along with the findings from our simulations suggest that the bound holds for $x \geq 4.42$. Therefore, here we restrict ourselves to the case $x \leq 4.42$.

We also observe that, when $\rho \geq 0.5$, the bound $\alpha(1-\rho)$ works when $n \geq 900$ and $\alpha \geq 0.01$. When $n = 900$ and $\alpha = 0.01$, we have $x = 4.23$. This, along with the findings from our simulations suggest that, when $\rho \geq .5$, the bound works for $x \geq 4.23$. Therefore, when $\rho \geq .5$, we restrict ourselves to the case $x \leq 4.23$.

We also assume $\rho \geq 0.01$ for the rest of this chapter. We shall derive upper bounds on $FWER_{Bon}(n,\alpha,\rho)$ in each of the following four cases separately:

1. $4.23 \geq x \geq 2, \rho \geq .5$
2. $4.42 \geq x \geq 2, .01 \leq \rho < .5$
3. $x \leq 2, \rho \geq .5$
4. $x \leq 2, \rho < .5$

Case 1. $4.23 \geq x \geq 2, \rho \geq .5$

Theorem 2.3.3. *Let* $4.23 \geq x \geq 2$ *and* $\rho \geq .5$*. Then,*

$$\forall x \in [x_l, x_{l+1}], \qquad FWER_{Bon}(n,\alpha,\rho) \leq \alpha - \frac{n-1}{n}\cdot\frac{\alpha^2}{n} - \frac{n-1}{n}\cdot\frac{\alpha\rho}{6}\cdot C_{x_l}$$

where x_l*'s and* C_{x_l}*'s are as follows:*

l	1	2	3	4	5	6	7
x_l	2	2.56	3.06	3.33	3.71	3.93	4.23
C_{x_l}	1	$\frac{1}{2}$	$\frac{1}{\pi}$	$\frac{1}{2\pi}$	$\frac{1}{\pi^2}$	$\frac{1}{6\pi}$	-

Proof of Theorem 2.3.3. We have from Corollary 2.3.2, for each $x \geq 2$,

$$\begin{aligned} FWER_{Bon}(n,\alpha,\rho) &\leq \alpha - \frac{n-1}{n}\cdot\frac{\alpha^2}{n} - \frac{n-1}{2\pi}\cdot\sin^{-1}\rho\cdot e^{-\frac{x^2}{1+\frac{\rho}{2}}} \\ &\leq \alpha - \frac{n-1}{n}\cdot\frac{\alpha^2}{n} - \frac{n-1}{2\pi}\cdot\frac{2\pi\rho}{6}\cdot e^{-\frac{x^2}{1+.25}}. \end{aligned}$$

The last step follows since $\rho \geq .5$ implies $\frac{\sin^{-1}\rho}{\rho} \geq \frac{\pi}{3}$. Hence it is enough to show that

$$\forall x \in [x_l, x_{l+1}] \qquad e^{-\frac{x^2}{1.25}} \geq \frac{\alpha}{n}\cdot C_{x_l}(n).$$

Now, $\frac{\alpha}{n} = 1 - \Phi(x)$. Let, $M(x) = \frac{e^{-\frac{x^2}{1.25}}}{1-\Phi(x)}$. Using computational tools, we get that $\forall x \in [x_l, x_{l+1}]$, $M(x) \geq C_{x_l}(n)$ and the proof is completed. □

Case 2. $4.42 \geq x \geq 2, .01 \leq \rho < .5$

Theorem 2.3.4. *Let $4.42 \geq x \geq 2$ and $.01 \leq \rho < .5$. Let $I_1 = [\frac{1}{3}, .5)$, $I_2 = [\frac{1}{2\pi}, \frac{1}{3})$ and $I_3 = [0.01, \frac{1}{2\pi})$. Then, for $\rho \in I_i$ with $i = 1, 2, 3$ and for $x \in [x_m(i), x_{m+1}(i)]$,*

$$FWER_{Bon}(n, \alpha, \rho) \leq \alpha - \frac{n-1}{n} \cdot \frac{\alpha^2}{n} - \frac{n-1}{n} \cdot \frac{\alpha\rho}{2\pi} \cdot D_{x_m}$$

where x_m's and D_{x_m}'s are as follows:

m	1	2	3	4	5	6	7	8	9	
D_{x_m}	1	$\frac{1}{2}$	$\frac{1}{\pi}$	$\frac{1}{2\pi}$	$\frac{1}{\pi^2}$	$\frac{1}{\pi^3}$	$\frac{1}{\pi^4}$	$\frac{1}{4\pi^4}$	$\frac{1}{16\pi^4}$	
$x_m(1)$	2	2.3	2.76	3	3.36	3.56	4	4.42		
$x_m(2)$		2	2.49	2.72	3.04	3.23	3.66	4.03	4.42	
$x_m(3)$		2	2.28	2.5	2.8	2.97	3.37	3.72	4.1	4.42

Its proof is identical to the preceding proof and hence omitted.

Case 3. $x \leq 2, \rho \geq .5$

Theorem 2.3.5. *Let $x \leq 2$ and $\rho \geq .5$. Then,*

$$FWER_{Bon}(n, \alpha, \rho) \leq \alpha - \frac{n-1}{n} \cdot \frac{\alpha^2}{n} - \frac{n-1}{n} \cdot \frac{\alpha\rho}{6}.$$

Proof of Theorem 2.3.5. We have, from Theorem 2.3.1,

$$FWER_{Bon}(n, \alpha, \rho) \leq \alpha - \frac{n-1}{n} \cdot \frac{\alpha^2}{n} - \frac{n-1}{2\pi} \int_0^\rho \frac{1}{\sqrt{1-z^2}} e^{\frac{-x^2}{1+z}} dz.$$

Now,

$$\begin{aligned}
& \frac{n-1}{2\pi} \int_0^\rho \frac{1}{\sqrt{1-z^2}} e^{\frac{-x^2}{1+z}} dz \\
\geq\ & \frac{n-1}{2\pi} \cdot \frac{1}{\rho} \int_0^\rho \frac{1}{\sqrt{1-z^2}} dz \cdot \int_0^\rho e^{\frac{-x^2}{1+z}} dz \quad \text{(using Lemma 2.3.3)} \\
=\ & \frac{n-1}{2\pi} \cdot \frac{\sin^{-1}\rho}{\rho} \cdot \left[\int_0^{\rho/2} e^{\frac{-x^2}{1+z}} dz + \int_{\rho/2}^\rho e^{\frac{-x^2}{1+z}} dz \right] \\
\geq\ & \frac{n-1}{2\pi} \cdot \frac{\sin^{-1}\rho}{\rho} \cdot \frac{\rho}{2} \left[e^{-x^2} + e^{-\frac{x^2}{1+\rho/2}} \right] \quad \text{(since } e^{\frac{-x^2}{1+z}} \text{ is increasing in } z\text{)} \\
=\ & \frac{\sin^{-1}\rho}{2\pi} \cdot (n-1) \cdot \left[\frac{e^{-x^2} + e^{-\frac{x^2}{1+\rho/2}}}{2} \right]
\end{aligned}$$

$$\begin{aligned}
&\geq \frac{\sin^{-1}\rho}{2\pi} \cdot (n-1) \cdot \left[\frac{e^{-x^2} + e^{-\frac{x^2}{1+.25}}}{2}\right] \quad (\text{since } \rho \geq .5) \\
&= \frac{\sin^{-1}\rho}{2\pi} \cdot (n-1) \cdot G(x) \quad (\text{suppose})
\end{aligned}$$

Now, we have $\frac{\sin^{-1}\rho}{2\pi} \geq \frac{\rho}{6}$ since $\rho \geq .5$. Also, $G(x) \geq 1 - \Phi(x) = \frac{\alpha}{n}$ for $x \leq 2.2$. The rest follows from Theorem 2.3.1. □

Case 4. $x \leq 2, \rho < .5$

Theorem 2.3.6. *Let $x = \Phi^{-1}(1 - \frac{\alpha}{n}) \leq 2$. Then,*

$$FWER_{Bon}(n, \alpha, \rho) \leq \alpha - \frac{n-1}{n} \cdot \frac{\alpha^2}{n} - \frac{n-1}{n} \cdot \frac{2\alpha\rho}{5\pi}.$$

It is mention-worthy that Theorem 2.3.6 is valid for any non-negative ρ.

Proof of Theorem 2.3.6. $\Phi^{-1}(1 - \frac{\alpha}{n}) \leq 2$ implies $\frac{\alpha}{n} \geq 1 - \Phi(2) = 0.02275$. Therefore, $\rho \geq 0.02275$. Now, along the same lines of the preceding proof, we have,

$$\begin{aligned}
&\frac{n-1}{2\pi} \int_0^{\rho} \frac{1}{\sqrt{1-z^2}} e^{\frac{-x^2}{1+z}} dz \\
&\geq \frac{\sin^{-1}\rho}{2\pi} \cdot (n-1) \cdot \left[\frac{e^{-x^2} + e^{-\frac{x^2}{1+\rho/2}}}{2}\right] \\
&\geq \frac{\rho}{2\pi} \cdot (n-1) \cdot \left[\frac{e^{-x^2} + e^{-\frac{x^2}{1+.011375}}}{2}\right] \quad (\text{since } \rho \geq .02275) \\
&= \frac{\rho}{2\pi} \cdot (n-1) \cdot H(x) \quad (\text{suppose})
\end{aligned}$$

Now, $H(x) \geq \frac{4}{5}(1 - \Phi(x)) = \frac{4\alpha}{5n}$ for $x \leq 2$. The rest is obvious from Theorem 2.3.1. □

2.4 Bounds on FWER in General Non-asymptotic Setup

We have considered an equicorrelated dependence structure so far. However, problems involving variables with more general dependence structure need to be tackled with more general correlation matrices. Hence, the study of the behavior of FWER in arbitrarily correlated normal setups becomes crucial. Towards this, we consider the same *Gaussian sequence model* as in Section 2, but now we assume $\mathbb{C}\mathrm{orr}\,(X_i, X_j) = \rho_{ij}$ for $i \neq j$ with

$\rho_{ij} \geq 0$. We recall the definition of FWER for this setup:

$$FWER_{Bon}(n, \alpha, \Sigma_n) = \mathbb{P}_{\Sigma_n}\left(\bigcup_{i=1}^{n}\{X_i > c_{\alpha,n} \mid H_0\}\right) = \mathbb{P}_{\Sigma_n}\left(\bigcup_{i=1}^{n} A_i\right).$$

where $A_i = \{X_i > \Phi^{-1}(1 - \frac{\alpha}{n})|H_0\}$ for $i \in \mathcal{I}$.

In the equicorrelated setup, we use Kwerel's inequality (Lemma 2.3.1) to find an upper bound to FWER:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i) - \frac{2}{n} \sum_{1 \leq i < j \leq n} \mathbb{P}(A_i \cap A_j).$$

That approach can be used to obtain bounds on FWER in the non-negatively correlated setup also. However, one observes that the above inequality gives equal importance to all the intersections. Therefore, it might be advantageous to use some other probability inequality which involves the intersections with higher probabilities only. We mention such an inequality below:

Lemma 2.4.1 (Kounias (1968))**.** *Let A_1, A_2, $\ldots$, A_n be n events. Then,*

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i) - \max_{1 \leq i \leq n} \sum_{j=1, j \neq i}^{n} \mathbb{P}(A_i \cap A_j).$$

Evidently Kounias's inequality is sharper than Kwerel's inequality and they are equivalent when $\mathbb{P}(A_i \cap A_j)$ is same for all $i \neq j$. We state a generalization of Theorem 2.3.1:

Theorem 2.4.1. *Consider the non-negatively correlated normal set-up with covariance matrix Σ_n. Then,*

$$FWER_{Bon}(n, \alpha, \Sigma_n) \leq \alpha - \frac{n-1}{n} \cdot \frac{\alpha^2}{n} - \frac{1}{2\pi} \sum_{j=1, j \neq i^*}^{n} \int_0^{\rho_{i^*j}} \frac{1}{\sqrt{1-z^2}} e^{\frac{-\Phi^{-1}(1-\frac{\alpha}{n})^2}{1+z}} dz$$

where $i^* = \arg\max_i \sum_{j=1, j \neq i}^{n} \rho_{ij}$.

We observe that Theorem 2.4.1 reduces to Theorem 2.3.1 when $\rho_{ij} = \rho$ for all $i \neq j$.

Proof of Theorem 2.4.1. We have $\mathbb{P}(A_i) = \frac{\alpha}{n}$ where $A_i = \{X_i > \Phi^{-1}(1 - \frac{\alpha}{n})|H_0\}$, for $i = 1, \ldots, n$. One can show, along the similar lines of the proof of Theorem 2.3.1, the following:

$$\mathbb{P}_{\Sigma_n}(A_i \cap A_j) = \frac{\alpha^2}{n^2} + \frac{1}{2\pi}\int_0^{\rho_{ij}} \frac{1}{\sqrt{1-z^2}} e^{\frac{-\Phi^{-1}(1-\frac{\alpha}{n})^2}{1+z}} dz \qquad \forall i \neq j.$$

Hence, $\mathbb{P}_{\Sigma_n}(A_i \cap A_j)$ is an increasing function of ρ_{ij}. Therefore,

$$\arg\max_i \sum_{j=1, j\neq i}^{n} \mathbb{P}_{\Sigma_n}(A_i \cap A_j) = \arg\max_i \sum_{j=1,j\neq i}^{n} \rho_{ij} = i^* \quad \text{(say)}.$$

Hence, applying Lemma 2.4.1, we get

$$\begin{aligned} FWER_{Bon} = \mathbb{P}_{\Sigma_n}\left(\bigcup_{i=1}^{n} A_i\right) &\leq \sum_{i=1}^{n} \mathbb{P}(A_i) - \max_{1\leq i\leq n} \sum_{j=1,j\neq i}^{n} \mathbb{P}_{\Sigma_n}(A_i \cap A_j) \\ &= \alpha - \frac{n-1}{n}\cdot\frac{\alpha^2}{n} - \frac{1}{2\pi}\sum_{j=1,j\neq i^*}^{n} \int_0^{\rho_{i^*j}} \frac{1}{\sqrt{1-z^2}} e^{\frac{-\Phi^{-1}(1-\frac{\alpha}{n})^2}{1+z}} dz \end{aligned}$$

, completing the proof. □

Corollary 2.4.1. *Consider the non-negatively correlated normal setup with covariance matrix Σ_n. Let $i^* = \arg\max_i \sum_{j=1,j\neq i}^{n} \rho_{ij}$ and $j_* = \arg\min_j \rho_{i^*j}$. Then, if $\rho_{i^*j_*} \leq \frac{\alpha}{n}$, $FWER_{Bon}(n,\alpha,\Sigma_n) \leq \alpha - \dfrac{n-1}{n}\cdot \alpha\rho_{i^*j_*}$.*

Hence, we assume $\rho_{i^*j_*} > \frac{\alpha}{n}$ from now on. Suppose $\bar{\rho}_{i^*} = \dfrac{1}{n-1}\displaystyle\sum_{j=1,j\neq i^*}^{n} \rho_{i^*j}$. We have the following two generalizations of Theorem 2.3.5 and Theorem 2.3.6 respectively:

Theorem 2.4.2. *Consider the non-negatively correlated normal setup with covariance matrix Σ_n. Let $\Phi^{-1}(1-\frac{\alpha}{n}) \leq 2, \rho_{i^*j_*} \geq .5$. Then,*

$$FWER_{Bon}(n,\alpha,\Sigma_n) \leq \alpha - \frac{n-1}{n}\cdot\frac{\alpha^2}{n} - \frac{n-1}{n}\cdot\frac{\alpha\bar{\rho}_{i^*}}{6}.$$

Theorem 2.4.3. *Consider the non-negatively correlated normal setup with covariance matrix Σ_n. Let $\Phi^{-1}(1-\frac{\alpha}{n}) \leq 2$. Then,*

$$FWER_{Bon}(n,\alpha,\Sigma_n) \leq \alpha - \frac{n-1}{n}\cdot\frac{\alpha^2}{n} - \frac{n-1}{n}\cdot\frac{2\alpha\bar{\rho}_{i^*}}{5\pi}.$$

Proof of Theorem 2.4.2. For any $j \in \mathcal{I}\setminus\{i\}$,

$$\begin{aligned} &\frac{1}{2\pi}\int_0^{\rho_{i^*j}} \frac{1}{\sqrt{1-z^2}} e^{\frac{-x^2}{1+z}} dz \\ \geq\ &\frac{1}{2\pi}\cdot\frac{1}{\rho_{i^*j}}\int_0^{\rho_{i^*j}} \frac{1}{\sqrt{1-z^2}} dz \cdot \int_0^{\rho_{i^*j}} e^{\frac{-x^2}{1+z}} dz \quad \text{(using Lemma 2.3.3)} \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2\pi}\cdot\frac{\sin^{-1}\rho_{i^*j}}{\rho_{i^*j}}\cdot\left[\int_0^{\rho_{i^*j}/2} e^{\frac{-x^2}{1+z}}dz + \int_{\rho_{i^*j}/2}^{\rho_{i^*j}} e^{\frac{-x^2}{1+z}}dz\right]\\
&\geq \frac{1}{2\pi}\cdot\frac{\sin^{-1}\rho_{i^*j}}{\rho_{i^*j}}\cdot\frac{\rho_{i^*j}}{2}\left[e^{-x^2}+e^{-\frac{x^2}{1+\rho_{i^*j}/2}}\right] \quad \text{(since } e^{\frac{-x^2}{1+z}} \text{ is increasing in } z)\\
&= \frac{\sin^{-1}\rho_{i^*j}}{2\pi}\cdot\left[\frac{e^{-x^2}+e^{-\frac{x^2}{1+\rho_{i^*j}/2}}}{2}\right]\\
&\geq \frac{\sin^{-1}\rho_{i^*j}}{2\pi}\cdot\left[\frac{e^{-x^2}+e^{-\frac{x^2}{1+.25}}}{2}\right] \quad \text{(since } \rho_{i^*j}\geq .5)\\
&= \frac{\sin^{-1}\rho_{i^*j}}{2\pi}\cdot G(x) \quad \text{(suppose)}\\
&\geq \frac{\rho_{i^*j}}{6}\cdot\frac{\alpha}{n} \quad \text{(since } \rho_{i^*j}\geq .5 \text{ and } G(x)\geq \alpha/n \text{ for } x\leq 2.2)
\end{aligned}$$

Summing over j, we obtain,

$$\frac{1}{2\pi}\sum_{j=1,j\neq i^*}^{n}\int_0^{\rho_{i^*j}}\frac{1}{\sqrt{1-z^2}}e^{\frac{-\Phi^{-1}(1-\frac{\alpha}{n})^2}{1+z}}dz \geq \frac{n-1}{n}\cdot\frac{\alpha\bar{\rho}_{i^*}}{6}.$$

The rest follows from Theorem 2.4.1. □

Proof of Theorem 2.4.3. $\Phi^{-1}(1-\frac{\alpha}{n})\leq 2$ implies $\frac{\alpha}{n}\geq 1-\Phi(2)=0.02275$. Therefore, $\rho\geq 0.02275$. Now, along the same lines of the preceding proof, we have,

$$\begin{aligned}
&\frac{n-1}{2\pi}\int_0^{\rho}\frac{1}{\sqrt{1-z^2}}e^{\frac{-x^2}{1+z}}dz\\
&\geq \frac{\sin^{-1}\rho}{2\pi}\cdot(n-1)\cdot\left[\frac{e^{-x^2}+e^{-\frac{x^2}{1+\rho/2}}}{2}\right]\\
&\geq \frac{\rho}{2\pi}\cdot(n-1)\cdot\left[\frac{e^{-x^2}+e^{-\frac{x^2}{1+.011375}}}{2}\right] \quad \text{(since } \rho\geq .02275)\\
&= \frac{\rho}{2\pi}\cdot(n-1)\cdot H(x) \quad \text{(suppose)}
\end{aligned}$$

Now, $H(x)\geq\frac{4}{5}(1-\Phi(x))=\frac{4\alpha}{5n}$ for $x\leq 2$. The rest is obvious from Theorem 2.3.1. □

In the proof of Theorem 2.3.2, we show that, for any $\rho\geq 0$,

$$\forall x\geq 2, \qquad \int_0^{\rho}\frac{1}{\sqrt{1-z^2}}e^{\frac{-x^2}{1+z}}dz \geq \sin^{-1}\rho\cdot e^{-\frac{x^2}{1+\frac{\rho}{2}}}.$$

This inequality leads to the following.

Corollary 2.4.2. *Consider the non-negatively correlated normal setup with covariance matrix Σ_n. Let $x = \Phi^{-1}(1 - \frac{\alpha}{n}) \geq 2$. Then,*

$$FWER_{Bon}(n, \alpha, \Sigma_n) \leq \alpha - \frac{n-1}{n} \cdot \frac{\alpha^2}{n} - \frac{1}{2\pi} \sum_{j=1, j \neq i^*}^{n} \sin^{-1} \rho_{i^*j} \cdot e^{-\frac{x^2}{1+\frac{\rho_{i^*j}}{2}}}$$

where $i^* = \arg\max_i \sum_{j=1, j\neq i}^{n} \rho_{ij}$.

One can derive results similar to Theorem 2.3.3 or Theorem 2.3.4 using the above corollary by imposing certain conditions on the values of the correlations in the i^*-th row of R. For example, we have the following if we assume that $\rho_{i^*j_*} \geq .5$:

Theorem 2.4.4. *Consider the non-negatively correlated normal setup with covariance matrix Σ_n. Let $4.23 \geq x \geq 2$ and $\rho_{i^*j_*} \geq .5$. Then,*

$$\forall x \in [x_l, x_{l+1}] \qquad FWER(n, \alpha, \Sigma_n) \leq \alpha - \frac{n-1}{n} \cdot \frac{\alpha^2}{n} - \frac{n-1}{n} \cdot \frac{\alpha \bar{\rho}_{i^*}}{6} \cdot C_{x_l}(n)$$

where x_l's and $C_{x_l}(n)$'s are as follows:

l	1	2	3	4	5	6	7
x_l	2	2.56	3.06	3.33	3.71	3.93	4.23
$C_{x_l}(n)$	1	$\frac{1}{2}$	$\frac{1}{\pi}$	$\frac{1}{2\pi}$	$\frac{1}{\pi^2}$	$\frac{1}{6\pi}$	-

This can be established along the same lines of the proof of Theorem 2.3.3.

2.5 An Improved Multiple Testing Procedure

In the preceding section we have obtained the following upper bound on Bonferroni FWER in the the non-negatively correlated normal set-up with covariance matrix Σ_n (Theorem 2.4.1):

$$\begin{aligned} FWER_{Bon}(n, \alpha, \Sigma_n) &\leq \alpha - \frac{n-1}{n} \cdot \frac{\alpha^2}{n} - \frac{1}{2\pi} \sum_{j=1, j\neq i^*}^{n} \int_0^{\rho_{i^*j}} \frac{1}{\sqrt{1-z^2}} e^{\frac{-\{\Phi^{-1}(1-\frac{\alpha}{n})\}^2}{1+z}} dz \\ &= f_{n,\Sigma_n}(\alpha) \quad \text{(say)} \end{aligned}$$

where $i^* = \arg\max_i \sum_{j=1, j\neq i}^{n} \rho_{ij}$. This upper bound can be used to adjust the critical points in the targeted multiple test problem. This enables us to obtain a more powerful test than the existing ones, e.g., the Bonferroni procedure. Towards this, let

$$\alpha^\star := \arg\max_{\beta \in (0,1)} \Big\{ f_{n,\Sigma_n}(\beta) \leq \alpha \Big\}.$$

Evidently $\alpha^\star \geq \alpha$. Then, we can decrease the Bonferroni cutoff $\Phi^{-1}(1-\alpha/n)$ to $\Phi^{-1}(1-\alpha^\star/n)$ and thus significantly improve the ability to detect false hypotheses. In other words, for our modified method, under the global null,

$$FWER_{modified}(n, \alpha, \Sigma_n) = \mathbb{P}_{\Sigma_n}\left(X_i > \Phi^{-1}(1-\alpha^\star/n) \quad | \, H_0\right).$$

The definition of $\alpha^\star$ itself ensures that $FWER_{modified}(n, \alpha, \Sigma_n)$ is indeed controlled at level α. We note that this proposed method controls FWER under any covariance matrix with non-negative entries. We present the values of $\alpha^\star$ for some combinations of $(n, \alpha = .05, \rho)$ in Table 2.2. We observe that $\alpha^\star$ is much larger than α for larger values of ρ. Hence the proposed method will have more rejection at the same FWER level and make better inference.

Table 2.2: Values of $\alpha^\star$ for different choices of $(n, \alpha = .05, \rho)$

(n, α)	$x = \Phi^{-1}(1-\alpha/n)$	Correlation (ρ)	0.1	0.3	0.5	0.7	0.9
$(10, 0.05)$	2.5758	$\alpha^\star$	0.05	.0515	.0550	.0636	.0970
$(100, 0.05)$	3.2905	$\alpha^\star$	0.05	.0505	.0520	.0581	.0879

2.6 Simulation Study

The bound by Das and Bhandari (2021) provides a significant gain in power for Bonferroni method for large number of hypotheses. However, for equicorrelated setups with small or moderate dimensions, their bound fails as mentioned earlier. We verify this through simulations. Our simulation scheme, for fixed (n, α) is as follows:

(a) For each $\rho \in \{0, .025, .050, .075, \ldots, 1\}$, we generate 10000 n-variate equicorrelated multivariate normal observations (each with mean 0 and variance 1; common correlation coefficient being ρ).

(b) For each ρ,

- in each of the 10000 replications, we note whether or not any of the generated n components exceeds the cutoff $\Phi^{-1}(1-\alpha/n)$.

- the estimated FWER (for that ρ) is obtained accordingly from the 10000 replications.

The plots after running these simulations for $(n, \alpha) = (100, .01)$ and $(500, .05)$ are given in Figure 2.2 (the blue line represents the straight line $\alpha(1-\rho)$).

Figure 2.3: FWER Plots for (n, α)=(100,.01) and (500,.05)

We can see that the $\alpha(1-\rho)$ bound fails in these cases. Also, FWER is not a convex function of ρ in these cases. We present the simulation results for some choices of (n, α, ρ) along with our proposed bounds in Table 2.4. It is mention worthy that in each case the estimated FWER is smaller than our proposed bounds.

One can see that our bounds give good results for small values of equicorrelation ρ and tend to become weak for large values of ρ. This is in contrast to the method of Das and Bhandari (2021) whose bound works in the large ρ case. Therefore, in a way, our bounds and the $\alpha(1-\rho)$ bound are complementary to each other in depicting the behaviour of FWER in equicorrelated normal setups.

2.7 Concluding Remarks

This work is probably the first attempt in studying the effect of correlation on FWER for small and moderate number of hypotheses.

Table 2.4: Estimates of FWER(n, α, ρ)

(n, α)	x	Correlation (ρ)	0.1	0.3	0.5	0.7	0.9
$(10, 0.3)$	1.8808	$\widehat{\text{FWER}}(n, \alpha, \rho)$	.2132	.2053	.1688	.1242	.0733
		Bound	.2885	.2816	.2747	.2678	.2610
$(100, 0.05)$	3.2905	$\widehat{\text{FWER}}(n, \alpha, \rho)$	.0456	.0355	.0265	.0153	.0005
		Bound	.0499	.0495	.0479	.0432	.0294
$(500, 0.05)$	3.7190	$\widehat{\text{FWER}}(n, \alpha, \rho)$	.0451	.0319	.0198	.0081	.0028
		Bound	.0499	.0498	.0488	.0451	.0318

The proofs of our results heavily use the fact that FWER can be regarded as $\mathbb{P}(\cup_{i=1}^{n} A_i)$ for suitably defined events A_i, $1 \leq i \leq n$. Accurate computation of this probability is difficult because, in practice (as in multiple testing), the complete dependence between the events $(A_1, \ldots, A_n)$ is often unknown or unavailable (in our case ρ is unknown and we have only some idea about ρ), unless the events A_j are independent. The available information is often the marginal probabilities and joint probabilities up to level $m(m << n)$. In these situations, one aims to compute a bound using only a limited amount of information. We concentrated on individual and pairwise intersection probabilities. From dependence analysis standpoint, pairwise intersection probabilities convey important information. Hence, the second order Bonferroni inequalities are most adequate general tool. On the other hand, taking specific bounds developed for concrete distributions into account leads to more concrete and better results. With this attitude in mind, for low dimensional normal distributions, by the refining and applying of Monhor's representation formula (Lemma 2.3.2) and inequality for correlated bivariate normal distribution, the role of correlation coefficient, hence, of dependence expressed in probability content becomes more explicit. In general, the correlated bivariate normal distribution is key distribution for probabilistic analysis of dependence. Our refinement to Monhor's inequality improves the tightness of the bound for some areas of arguments of distribution function. In various stochastic modelling the univariate and multivariate normal distributions are most frequent (Hutchinson and Lai, 1990; Monhor, 2011). The probabilistic results in this chapter may be useful not only in FWER questions, but other areas, too.

Throughout this chapter, we have considered multivariate normal setup. In various areas of stochastic modeling, the multivariate normal distribution is frequent (Hutchinson and Lai, 1990; Olkin and Viana, 1995; Monhor, 2011). However, one interesting extension would be to study the behavior of k-FWER under more general distributional setups.

Chapter 3

Non-asymptotic Behaviors of Generalized FWERs in Correlated Normal Distributions[1]

3.1 Introduction

The preceding chapter focused on establishing bounds on FWER under correlated normal setups (Theorem 2.3.1, Theorem 2.4.1). However, in many scientific avenues where the number of hypotheses n is moderately large, FWER control is stringent. So the deviations from null have little chance of being identified. For this reason, other error criteria are proposed in the literature.

Lehmann and Romano (2005) consider the k-FWER (or gFWER), the probability of rejecting at least k true null hypotheses in a simultaneous testing problem. This is pertinent in settings where several type I errors are allowed, provided the number of type I errors is controlled. Thus k-FWER controls false rejections less severely, but in doing so detects false null hypotheses better and consequently provides better power. k-FWER is especially relevant in those areas where the number of hypotheses is large e.g microarray data analysis.

The usual Bonferroni procedure uses the cutoff $\Phi^{-1}(1-\alpha/n)$ to control FWER at level α. Lehmann and Romano (2005) remark that controlling k-FWER allows one to decrease this cutoff to $\Phi^{-1}(1-k\alpha/n)$, and thus significantly increase the ability to identify false

[1]This chapter is based on the publication M. Dey and S. K. Bhandari (2023) Bounds on generalized family-wise error rates for normal distributions, *Statistical Papers*, DOI: https://doi.org/10.1007/s00362-023-01487-0.

hypotheses. Thus, for their Bonferroni-type procedure, under the global null,

$$k\text{-FWER}(n,\alpha,\Sigma_n) = \mathbb{P}_{\Sigma_n}\left(X_i > \Phi^{-1}(1-k\alpha/n) \text{ for at least } k \ i\text{'s} \mid H_0\right).$$

Evidently, when $k=1$, Lehmann-Romano procedure simplifies to the Bonferroni method and k-FWER reduces to the usual FWER.

The existing literature lacks a theory on the extent of the conservativeness of gFWER controlling procedures under dependent frameworks with a moderate number of hypotheses. This chapter tackles this problem in a unified manner by theoretically establishing upper bounds on the gFWER of the Lehmann-Romano procedure under correlated Gaussian setups. Towards this, we derive a new and quite general probability inequality which, in turn, extends a classical inequality. Our results also generalize the results (e.g. Theorem 2.4.1) of the preceding chapter.

We first formally introduce the framework with relevant notations. We derive some inequalities on the probability of occurrence of at least k among n events in Section 3.3. We also propose an improved multiple testing procedure utilizing those inequalities. We analyze the performance of our proposed k-FWER controlling procedure in a real dataset in Section 3.4. Simulation studies are presented in Section 3.5. We conclude with a brief discussion in Section 3.6.

3.2 Preliminaries

We address the multiple testing problem through a *Gaussian sequence model*:

$$X_i \sim N(\mu_i, 1), \quad i \in \{1,\dots,n\}$$

where $\mathrm{Corr}\,(X_i, X_j) = \rho_{ij}$ for each $i \neq j$ $(\rho_{ij} \geq 0)$. We wish to test:

$$H_{0i} : \mu_i = 0 \quad vs \quad H_{1i} : \mu_i > 0, \quad 1 \leq i \leq n.$$

The global null $H_0 = \bigcap_{i=1}^n H_{0i}$ hypothesizes that each mean is zero.

3.2.1 Inequalities on Probabilities of Events

Probability bounding has been a traditional problem in probability theory and has witnessed numerous applications in statistics, reliability theory, and stochastic programming. Before we delve into the proofs of our proposed inequalities and multiple testing procedure, we review two classical probability inequalities. We also utilized these inequalities to obtain upper bounds on FWER in the preceding chapter.

Towards this, let $A_1, \ldots, A_n$ denote n events. Suppose $S_1 = \sum_{i=1}^{n} \mathbb{P}(A_i)$ and $S_2 = \sum_{1 \le i < j < n} \mathbb{P}(A_i \cap A_j)$ respectively denote the sum of individual probabilities and the sum of probabilities of pairwise intersections. Moreover, let

$$S_2' = \max_{1 \le i \le n} \sum_{j=1, j \ne i}^{n} \mathbb{P}(A_i \cap A_j).$$

Kwerel (1975) (Lemma 2.3.1) obtained the following:

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \le S_1 - \frac{2}{n} S_2.$$

We utilized the following main technical tool in the proof of Theorem 2.4.1 (Kounias, 1968) (Lemma 2.4.1):

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \le S_1 - S_2'.$$

Evidently, $S_2' \ge \frac{2}{n} S_2$ and therefore Kounias's inequality is stronger than Kwerel's inequality.

3.3 Main Results

3.3.1 Some New Inequalities

In Chapter 2, we had remarked that the derivation of upper bounds on FWER heavily uses the fact that FWER is the probability of the union of some suitably defined events. We have also previously seen that k-FWER is $\mathbb{P}(\text{at least } k \text{ out of } n\ A_i\text{'s occur})$ for suitably defined events A_i, $1 \le i \le n$. It thus seems natural that a similar upper bound on k-FWER can be derived if we have an extension of Kounias's inequality for probabilities of the form $\mathbb{P}(\text{at least } k \text{ out of } n\ A_i\text{'s occur})$. Accurate computation of this probability requires knowing the complete dependence between the events $(A_1, \ldots, A_n)$, which we typically do not know unless they are independent. As mentioned in Chapter 2, the available information is often the marginal probabilities and joint probabilities up to level $m(m << n)$. In those situations, one aims to compute upper bounds which require only the marginal and pairwise probabilities, as in Kounias's inequality. Towards finding such an easily computable upper bound on the probability that at least k among n events occur, we generalize Kounias's inequality in the following:

Lemma 3.3.1. *Let A_1, A_2, ..., A_n be n events. Then,*

$$\mathbb{P}(\textit{at least } k \textit{ out of } n\ A_i\textit{'s occur}) \le \frac{S_1 - S_2'}{k} + \frac{k-1}{k} \cdot \max_{1\le i\le n} \mathbb{P}(A_i).$$

Proof of Lemma 3.3.1. Let $I_i(w)$ be the indicator random variable of the event A_i for $1 \le i \le n$. Then the random variable $\max I_{i_1}(w)\cdots I_{i_k}(w)$ is the indicator of the event that at least k among n A_i's occur. Here the maximum is taken over all tuples $(i_1, \dots, i_k)$ with $i_1, \dots, i_k \in \{1, \dots, n\}$, $i_1 < \dots < i_k$. Now, for any $i = 1, \dots, n$,

$$\max I_{i_1}(w)\cdots I_{i_k}(w) \le \frac{1}{k}[1 - I_i(w)] \sum_{j=1}^{n} I_j(w) + I_i(w).$$

Taking expectations in above, we obtain

$$\mathbb{P}(\text{at least } k \text{ out of } n\ A_i\text{'s occur}) \le \frac{1}{k}\cdot\sum_{j=1}^{n}\mathbb{P}(A_j) - \frac{1}{k}\cdot\sum_{j=1, j\ne i}^{n}\mathbb{P}(A_i\cap A_j) + \mathbb{P}(A_i)\cdot\frac{k-1}{k}.$$

The rest follows by observing that the above holds for any $i = 1, \dots, n$. □

Note that Lemma 3.3.1 reduces to Lemma 2.4.1 for $k = 1$. We propose now another inequality:

Lemma 3.3.2. *Let A_1, A_2, ..., A_n be n events. Then, for each $k \ge 2$,*

$$\mathbb{P}(\textit{at least } k \textit{ out of } n\ A_i\textit{'s occur}) \le \frac{2S_2}{k(k-1)}.$$

Proof of Lemma 3.3.2. Let $Q_m, 0 \le m \le n$, denote the probability that exactly m among n events occur. Then,

$$\begin{aligned}\mathbb{P}(\text{at least } k \text{ out of } n\ A_i\text{'s occur}) = \sum_{m=k}^{n} Q_m &\le \sum_{m=k}^{n}\frac{\binom{m}{2}}{\binom{k}{2}}Q_m \\ &\le \frac{2}{k(k-1)}\sum_{m=2}^{n}\binom{m}{2}Q_m \\ &\le \frac{2}{k(k-1)}\mathbb{E}\left[\binom{T_n}{2}\right]\end{aligned}$$

where, in the last step, T_n denotes the number of events occurring. Now,

$$\mathbb{E}\left[\binom{T_n}{2}\right] = \mathbb{E}\left[\sum_{1\le i<j\le n} I_i(w)I_j(w)\right] = \sum_{1\le i<j\le n}\mathbb{P}(A_i\cap A_j) = S_2,$$

completing the proof. $\square$

The following result follows directly from the preceding two lemmas and is crucial in establishing an upper bound on k-FWER.

Corollary 3.3.1. *Let $A_1, A_2, \ldots, A_n$ be n events. Then, for each $k \geq 2$,*

$$\mathbb{P}(\textit{at least } k \textit{ among } n \ A_i\textit{'s occur}) \leq \min\left\{\frac{S_1 - S_2'}{k} + \frac{k-1}{k} \cdot \max_{1\leq i\leq n} \mathbb{P}(A_i), \frac{2S_2}{k(k-1)}\right\}.$$

Lemma 3.3.3. *Let $A_1, A_2, \ldots, A_n$ be n events such that $\mathbb{P}(A_i) = c \in (0,1)$ for each i. Then,*

$$\min\left\{\frac{S_1 - S_2'}{k} + \frac{k-1}{k} \cdot \max_{1\leq i\leq n} \mathbb{P}(A_i), \frac{2S_2}{k(k-1)}\right\} \leq \frac{S_1}{k}.$$

Proof of Lemma 3.3.3. If the first term in the above parentheses is less than or equal to S_1/k then we have nothing left to prove. If the first term is strictly greater than S_1/k then we show that the second term must be less than S_1/k. Now,

$$\begin{aligned}
&\frac{S_1 - S_2'}{k} + \frac{k-1}{k} \cdot \max_{1\leq i\leq n} \mathbb{P}(A_i) \geq \frac{S_1}{k} \\
\Longrightarrow\ &\frac{k-1}{k} \cdot c \geq \frac{S_2'}{k} \\
\Longrightarrow\ &\frac{k-1}{k} \cdot c \geq \frac{2S_2}{nk} \\
\Longrightarrow\ &\frac{S_1}{k} \geq \frac{2S_2}{k(k-1)} \quad (\text{since, } S_1 = nc).
\end{aligned}$$

This completes the proof. $\square$

3.3.2 Bounds on k-FWER

We suppose now that $(X_1, X_2, \ldots, X_n)$ have covariance matrix $\Sigma_n = ((\rho_{ij}))$ with $\rho_{ij} \geq 0$ for all $i \neq j$. We define $A_i = \{X_i > \Phi^{-1}(1 - k\alpha/n)\}$ for $1 \leq i \leq n$. This implies

$$k\text{-FWER}(n, \alpha, \Sigma_n) = \mathbb{P}_{\Sigma_n}(\text{at least } k \ A_i\text{'s occur} \mid H_0)\,.$$

Now,

$$\mathbb{P}_{H_0}(A_i) = \mathbb{P}_{H_0}\left[X_i > \Phi^{-1}(1 - k\alpha/n)\right] = 1 - \mathbb{P}_{H_0}\left[X_i \leq \Phi^{-1}(1 - k\alpha/n)\right] = k\alpha/n.$$

So, $S_1 = k\alpha$. To apply Corollary 3.3.1, we need to find an expression on $\mathbb{P}(A_i \cap A_j)$, which is given in the representation theorem by Monhor (2013) (Lemma 2.3.2). Now,

$$\begin{aligned}
&\mathbb{P}_{H_0}(A_i \cap A_j) \\
&= 1 - \mathbb{P}_{H_0}(A_i^c) - \mathbb{P}_{H_0}(A_j^c) + \mathbb{P}_{H_0}(A_i^c \cap A_j^c) \\
&= 1 - (1 - k\alpha/n) - (1 - k\alpha/n) + \mathbb{P}_{H_0}\Big(X_i \leq \Phi^{-1}(1 - k\alpha/n), X_j \leq \Phi^{-1}(1 - k\alpha/n)\Big) \\
&= \frac{2k\alpha}{n} - 1 + (1 - k\alpha/n)^2 + \frac{1}{2\pi}\int_0^{\rho_{ij}} \frac{1}{\sqrt{1-z^2}} e^{\frac{-\{\Phi^{-1}(1-\frac{k\alpha}{n})\}^2}{1+z}} dz \quad \text{(using Lemma 2.3.2)} \\
&= \frac{k^2\alpha^2}{n^2} + \frac{1}{2\pi}\int_0^{\rho_{ij}} \frac{1}{\sqrt{1-z^2}} e^{\frac{-\{\Phi^{-1}(1-\frac{k\alpha}{n})\}^2}{1+z}} dz.
\end{aligned}$$

So,

$$S_2 = \sum_{1\leq i<j\leq n} \mathbb{P}_{H_0}(A_i \cap A_j) = \sum_{1\leq i<j\leq n} \left[\frac{k^2\alpha^2}{n^2} + \frac{1}{2\pi}\int_0^{\rho_{ij}} \frac{1}{\sqrt{1-z^2}} e^{\frac{-\{\Phi^{-1}(1-\frac{k\alpha}{n})\}^2}{1+z}} dz\right].$$

Thus,

$$\begin{aligned}
\frac{2S_2}{k(k-1)} &= \frac{(n-1)k}{n(k-1)} \cdot \alpha^2 + \frac{1}{\pi k(k-1)} \sum_{1\leq i<j<n} \int_0^{\rho_{ij}} \frac{1}{\sqrt{1-z^2}} e^{\frac{-\{\Phi^{-1}(1-\frac{k\alpha}{n})\}^2}{1+z}} dz. \\
&= f_{n,k,\Sigma_n}(\alpha) \quad \text{(say)}.
\end{aligned}$$

We note that f is undefined for $k = 1$. Since we are interested in deriving upper bounds of a probability, we define $f_{n,1,\Sigma_n}(\alpha) = 1$. We also note that, $\mathbb{P}_{H_0}(A_i \cap A_j)$ is an increasing function of ρ_{ij}. Therefore,

$$\arg\max_i \sum_{j=1,j\neq i}^{n} \mathbb{P}_{H_0}(A_i \cap A_j) = \arg\max_i \sum_{j=1,j\neq i}^{n} \rho_{ij} = i^* \quad \text{(say)}.$$

So,

$$S_2' = \frac{n-1}{n} \cdot \frac{k^2\alpha^2}{n} + \frac{1}{2\pi} \sum_{j=1,j\neq i^*}^{n} \int_0^{\rho_{i^*j}} \frac{1}{\sqrt{1-z^2}} e^{\frac{-\{\Phi^{-1}(1-\frac{k\alpha}{n})\}^2}{1+z}} dz.$$

Hence, Lemma 3.3.1 gives that k-FWER(n, α, Σ_n) is bounded above by g where

$$\begin{aligned}
g_{n,k,\Sigma_n}(\alpha) &= \frac{S_1 - S_2'}{k} + \frac{k-1}{k} \cdot \max_{1\leq i\leq n} \mathbb{P}_{H_0}(A_i) \\
&= \alpha \cdot \frac{n+k-1}{n} - \frac{n-1}{n} \cdot \frac{k\alpha^2}{n} - \frac{1}{2\pi k} \sum_{j=1,j\neq i^*}^{n} \int_0^{\rho_{i^*j}} \frac{1}{\sqrt{1-z^2}} e^{\frac{-\{\Phi^{-1}(1-\frac{k\alpha}{n})\}^2}{1+z}} dz.
\end{aligned}$$

We apply Corollary 3.3.1 to obtain an extension of Theorem 2.4.1:

Theorem 3.3.1. *Let Σ_n be the correlation matrix of $X_1, \ldots, X_n$ with (i,j)'th entry ρ_{ij}. Suppose $\rho_{ij} \geq 0$ for all $i \neq j$. Then, for each $k > 1$,*

$$k\text{-}FWER(n, \alpha, \Sigma_n) \leq \min\{f_{n,k,\Sigma_n}(\alpha), g_{n,k,\Sigma_n}(\alpha)\}.$$

Remark 1. *We note that $\mathbb{P}_{H_0}(A_i) = k\alpha/n$ for each i, where $A_i = \{X_i > \Phi^{-1}(1-k\alpha/n)\}$. Lemma 3.3.3 gives us the following:*

$$\min\{f_{n,k,\Sigma_n}(\alpha), g_{n,k,\Sigma_n}(\alpha)\} \leq \alpha.$$

This implies that our upper bound is indeed sharper than the existing ones and the Lehmann-Romano procedure controls k-FWER at a level smaller than α.

It is mention-worthy that Theorem 3.3.1 is a quite general result in the non-asymptotic setup, because it tackles both more than one false rejections and general correlation matrices simultaneously. We have the following immediate corollary under the equicorrelated setup, i.e when $\rho_{ij} = \rho$ for all $i \neq j$:

Corollary 3.3.2. *Consider the equicorrelated normal set-up with correlation $\rho \geq 0$. Then,*

$$\begin{aligned} k\text{-}FWER(n, \alpha, \rho) \leq min\Bigg\{ & \frac{(n-1)k}{n(k-1)} \cdot \alpha^2 + \frac{n}{k-1} \cdot \frac{n-1}{2\pi k} \int_0^{\rho} \frac{1}{\sqrt{1-z^2}} e^{\frac{-\{\Phi^{-1}(1-\frac{k\alpha}{n})\}^2}{1+z}} dz, \\ & \alpha \cdot \frac{n+k-1}{n} - \frac{n-1}{n} \cdot \frac{k\alpha^2}{n} - \frac{n-1}{2\pi k} \int_0^{\rho} \frac{1}{\sqrt{1-z^2}} e^{\frac{-\{\Phi^{-1}(1-\frac{k\alpha}{n})\}^2}{1+z}} dz \Bigg\}. \end{aligned}$$

One can establish case-specific upper bounds on k-FWER depending on the values of $\Phi^{-1}(1 - k\alpha/n)$ and ρ_{ij}'s in general and equicorrelated normal setups using Theorem 3.3.1 and Corollary 3.3.2 respectively, in the same way as in Chapter 2.

3.3.3 An Improved Multiple Testing Procedure

The upper bound on k-FWER obtained in the preceding section can be used to adjust the critical points in the targeted multiple test problem. This enables us to obtain a more powerful test than the existing ones, e.g., the Lehmann-Romano procedure. Towards this, let

$$\alpha^\star := \underset{\beta \in (0,1)}{\arg\max} \Big\{ \min\{f_{n,k,\Sigma_n}(\beta), g_{n,k,\Sigma_n}(\beta)\} \leq \alpha \Big\}.$$

Remark 1 gives $\alpha^\star \geq \alpha$. Then, we can decrease the Lehmann-Romano cutoff $\Phi^{-1}(1 - k\alpha/n)$ to $\Phi^{-1}(1 - k\alpha^\star/n)$ and thus significantly improve the ability to detect false hypotheses. In other words, for our modified method, under the global null,

$$k\text{-FWER}_{modified}(n, \alpha, \Sigma_n) = \mathbb{P}_{\Sigma_n}\left(X_i > \Phi^{-1}(1 - k\alpha^\star/n) \text{ for at least } k \ \ i\text{'s} \mid H_0\right).$$

The definition of $\alpha^\star$ itself ensures that k-FWER$_{modified}(n, \alpha, \Sigma_n)$ is indeed controlled at level α. We note that this proposed method controls k-FWER under any covariance matrix with non-negative entries.

When the test statistics are independent, we have a nice relationship between α and $\alpha^\star$, as provided in the following result:

Lemma 3.3.4. *Let $k > 1$. Suppose $X_1, \ldots, X_n$ are independent. Moreover, let $\alpha \leq \frac{n(k-1)}{(n-1)k}$. Then, $\alpha^\star = \sqrt{\frac{n(k-1)\alpha}{(n-1)k}}$.*

Proof of Lemma 3.3.4. Under independence, we have

$$f_{n,k,I_n}(\alpha) = \frac{(n-1)k}{n(k-1)} \cdot \alpha^2,$$

$$g_{n,k,I_n}(\alpha) = \alpha \cdot \frac{n+k-1}{n} - \frac{n-1}{n} \cdot \frac{k\alpha^2}{n}.$$

Simple algebraic manipulations yield that

$$f_{n,k,I_n}(\alpha) \leq g_{n,k,I_n}(\alpha) \iff \alpha \leq \frac{n(k-1)}{(n-1)k}.$$

Thus, under the conditions of this theorem,

$$\alpha^\star = \underset{0<\beta<1}{\arg\max}\left\{\frac{(n-1)k}{n(k-1)} \cdot \beta^2 \leq \alpha\right\} = \sqrt{\frac{n(k-1)\alpha}{(n-1)k}}.$$

This completes the proof. $\square$

Lemma 3.3.4 implies that, under independence of the test statistics, $\alpha^\star$ can be taken close to $\sqrt{\alpha}$. This actually greatly increases the ability to reject false null hypotheses as we shall in Table 3.1.

3.4 Applications

Our proposed upper bound on the probability of occurrence of at least k among n events may be used in reliability. Subasi et al. (2017) remarks

"The typical application is to estimate the reliability evaluations of k-out-of-n systems such as multistate networks (oil and gas supply systems, communication networks, power generation and transmission systems, etc.) and fault tolerant systems (multidisplay system in a cockpit, multiengine system in an airplane, and multipump system in a hydraulic control system, etc.)."

We now discuss the practical relevance of our proposed multiple testing procedure through a real dataset, called the *prostate cancer dataset* (Singh et al., 2002). This dataset contains gene expression measurements of of $n = 6033$ genes for $N = 102$ individuals: 52 prostate cancer patients and 50 healthy persons.

Suppose y_{ij} denotes the expression level for gene i on individual j, $1 \le i \le n$, $1 \le j \le N$. Then the prostate cancer data is a $n \times N$ matrix $\mathbf{Y}$ with $1 \le j \le 50$ for the healthy persons and $51 \le j \le 102$ for the cancer patients. Let $\bar{y}_i(1)$ and $\bar{y}_i(2)$ be the averages of y_{ij} for these two groups respectively. It is important to identify genes whose levels vary between the cancer patients and healthy individuals (Efron, 2010a). This leads to testing

$$H_{0i} : y_{ij} \text{ follows the same distribution for the two groups of patients.}$$

One might use the usual t statistic $t_i = \frac{\bar{y}_i(2)-\bar{y}_i(1)}{s_i}$ to test H_{0i}. Here,

$$s_i^2 = \frac{\sum_{j=1}^{50} (y_{ij} - \bar{y}_i(1))^2 + \sum_{j=51}^{102} (y_{ij} - \bar{y}_i(2))^2}{100} \cdot \left(\frac{1}{50} + \frac{1}{52}\right).$$

We would reject H_{0i} at $\alpha = .05$ (based on usual normality assumptions) if $|t_i|$ exceeds 1.98, i.e, the two-tailed 5% point for a Student-t random variable with 100 d.f. Since we have viewed the multiple testing problem from a Gaussian sequence model framework in this thesis, we transform the t values to X values:

$$X_i = \Phi^{-1}\left(F\left(t_i\right)\right),$$

where F denotes the cdf of t_{100} distribution. We have,

$$H_{0i} : X_i \sim N(0,1).$$

We have $n = 6033$ genes to test. For a given k, the Lehmann-Romano procedure at level $\alpha = .05$ rejects any test having X value greater than $\Phi^{-1}(1 - k\alpha/n)$. Our proposed procedure rejects any test having X value greater than $\Phi^{-1}(1-k\alpha^\star/n)$. The observed value of the usual estimate of the correlation coefficient between the 6033 X values is less than

.001. For this reason and for technical simplicity, we take $\rho = 0$ in our computations. The number of rejected hypotheses for different values of k by Lehmann-Romano procedure and our proposed procedure is given in Table 3.1.

Table 3.1: Number of rejected hypotheses for different values of k

k	2	4	6	8	10	12	14	16	18	20
Lehmann-Romano Method	6	9	12	13	13	13	14	15	16	17
Proposed Method	12	15	19	22	26	27	27	28	31	34

We observe that our proposed method rejects significantly more number of hypotheses than the Lehmann-Romano method for each value of k. This also elucidates the significance of our proposed bound on the probability of occurrence of at least k among n events.

3.5 Simulations and Discussion

We adopt the following simulation scheme for fixed values of (n, k, α):

(a) For each $\rho \in \{.1, .3, .5, .7, .9\}$, we generate 10000 n-variate equicorrelated multivariate normal observations (each of the n components having zero mean, unit variance and each pair of components having common correlation ρ).

(b) For each $\rho \in \{.1, .3, .5, .7, .9\}$,

- in each of the 10000 replications, we note whether at least k many of the generated n components exceeds the cutoff $\Phi^{-1}(1 - k\alpha/n)$.
- The k-FWER for that ρ is estimated as the number of times at least k many exceeds the cutoff, divided by 10000.

We present the simulation results for k-FWER for different values of k and $(n = 100, \alpha = .05)$ along with our proposed bounds (given by Corollary 3.3.2) in Table 3.2.

Simulation results for k-FWER for various values of k and $(n = 500, \alpha = .05)$ along with our proposed bounds (given by Corollary 3.3.2) are presented in Table 3.3.

It is mentionworthy from the two tables that in each case our proposed bounds are significantly smaller than $\alpha = .05$, even for small values of ρ. Also, the estimated k-FWER is smaller than our proposed bounds. We also note that the estimated values of k-FWER decrease in k for small values of ρ, whereas increase in k for large values of ρ.

Table 3.2: Estimates of k-FWER$(n = 100, \alpha = .05, \rho)$

k	ρ	0.1	0.3	0.5	0.7	0.9
2	$k\text{-}\widehat{\text{FWER}}(\rho)$	.0104	.0177	.0172	.0137	.0069
	Bound	.0140	.0498	.0478	.0426	.0287
4	$k\text{-}\widehat{\text{FWER}}(\rho)$	.0015	.0090	.0132	.0135	.0089
	Bound	.0083	.0365	.048	.0423	.0283
8	$k\text{-}\widehat{\text{FWER}}(\rho)$	.0001	.0052	.0122	.0145	.0115
	Bound	.0063	.0232	.049	.0428	.0289

Table 3.3: Estimates of k-FWER$(n = 500, \alpha = .05, \rho)$

k	ρ	0.1	0.3	0.5	0.7	0.9
2	$k\text{-}\widehat{\text{FWER}}(\rho)$	.0121	.0169	.0127	.0082	.0027
	Bound	.0186	.0498	.0486	.0446	.0309
4	$k\text{-}\widehat{\text{FWER}}(\rho)$	.0027	.0117	.0110	.0090	.0038
	Bound	.0110	.0498	.0484	.0438	.0299
8	$k\text{-}\widehat{\text{FWER}}(\rho)$	.0003	.0086	.0105	.0097	.0051
	Bound	.0084	.0468	.0482	.0431	.0291

3.6 Concluding Remarks

Finner et al. (2007) remark that false discoveries are challenging to tackle in models with complex dependence structures, e.g., arbitrarily correlated normal models. This chapter proposes new bounds on the k-FWER of the Lehmann-Romano method in general dependent Gaussian scenarios. Towards this, we establish an inequality related to the probability that at least k among n events occur. This arises in various contexts, e.g., in transportation and communication networks. Our probabilistic results might be insightful in those areas, too.

Throughout the chapter, we have considered multivariate normal setup. In various areas of stochastic modeling, the multivariate normal distribution is frequent (Hutchinson and Lai, 1990; Olkin and Viana, 1995; Monhor, 2011). However, one interesting extension would be to study the behavior of k-FWER under more general distributional setups.

We investigate the limiting behaviors of step-wise multiple testing procedures under correlated normal setups in Chapter 5. It would be interesting to derive similar upper bounds for the FWER of step-wise decision rules under dependent normal frameworks.

Chapter 4

Asymptotic Behaviors of FWER and Generalized FWERs in Correlated Normal Distributions[1]

4.1 Introduction

In the preceding two chapters, we have studied the behaviors of FWER and generalized FWERs in a non-asymptotic correlated Gaussian sequence model setup. In this chapter, we consider the same under the asymptotic framework.

Das and Bhandari (2021) have shown that under the equicorrelated normal setup with non-negative correlation ρ , FWER(ρ) is a convex in ρ as the number of hypotheses approaches infinity. Consequently, they prove that the Bonferroni FWER is bounded above by $\alpha(1-\rho)$ where α is the target level. Here we establish that the Bonferroni FWER(ρ) approaches zero asymptotically for any positive ρ. This, combined with the well-known fact that for $\rho = 0$, FWER goes to $1-e^{-\alpha}$ asymptotically implies that limiting Bonferroni FWER is a convex function in ρ, though discontinuous at 0. Thus, the main result of Das and Bhandari (2021) follows from our result.

This chapter is organized as follows. Section 4.2 introduces the framework formally. Section 4.3 contains theoretical results about the limiting behavior of the FWER in equicorrelated and non-negatively correlated normal setups. Section 4.4 presents an extension of our main contribution to generalized FWERs. We study the asymptotic power of Bonferroni's procedure in Section 4.5. Section 4.6 includes simulation findings

[1]This chapter is based on the publication M. Dey, S. K. Bhandari (2023) FWER goes to zero for correlated normal, *Statistics & Probability Letters*, 193:109700, DOI: https://doi.org/10.1016/j.spl.2022.109700.

that empirically demonstrate our results. We end the chapter with a brief discussion in Section 4.7.

4.2 Preliminaries

We consider the correlated Gaussian sequence model discussed in Section 1.5.3. The FWER of a MTP T is defined as

$$FWER_T = \mathbb{P}(V_n(T) \geq 1). \tag{4.1}$$

where $V_n(T)$ denotes the number of false rejections made by T. It is mention-worthy that FWER is *not* the probability of making any false rejections under the global null. To control FWER at α, we need to ensure the probability of making any false rejection be less than α under any configuration of the n null hypotheses. However, for technical simplicity we shall compute the probability in r.h.s of Equation (4.1) under the global null H_0 at first (and consider that as the definition of FWER) and then extend the results obtained in this case to arbitrary configurations of the n hypotheses.

The Bonferroni FWER for the equicorrelated normal setup under the global null is given by

$$FWER_{Bon}(n, \alpha, \rho) = \mathbb{P}_{H_0}\left(\bigcup_{i=1}^{n}\{X_i > c_{\alpha,n}\}\right). \tag{4.2}$$

Das and Bhandari (2021) have shown that, $FWER_{Bon}(n, \alpha, \rho)$ is a convex function in ρ as $n \to \infty$ (Theorem 2.2.1). Consequently, they prove that $FWER_{Bon}(n, \alpha, \rho)$ is bounded above by $\alpha(1-\rho)$ asymptotically where α is the target level (Corollary 2.2.1).

To prove Theorem 2.2.1, Das and Bhandari (2021) consider the function

$$H_n(\rho) = 1 - FWER_{Bon}(n, \alpha, \rho).$$

The sequence $\{X_r\}_{r\geq 1}$ is exchangeable under H_0 for the equicorrelated set-up. For each $i \geq 1$, $X_i = \theta + Z_i$ where $\theta \sim N(0, \cdot)$ is independent of $\{Z_n\}_{n\geq 1}$ and $Z_i \sim N(0, \cdot)$. The equicorrelation structure gives $\mathrm{Var}(\theta) = \rho$. This implies $\theta \sim N(0, \rho)$ and $Z_i \overset{iid}{\sim} N(0, 1-\rho)$ for each $i \geq 1$. Thus,

$$\begin{aligned} H_n(\rho) &= \mathbb{P}\left(\theta + Z_i \leq c_{\alpha,n} \quad \forall i = 1, 2, \ldots, n \mid H_0\right) \\ &= \mathbb{E}_\theta\left[\Phi\left(\frac{c_{\alpha,n} - \theta}{\sqrt{1-\rho}}\right)^n\right] \\ &= \mathbb{E}\left[\Phi\left(\frac{c_{\alpha,n} + \sqrt{\rho}Z}{\sqrt{1-\rho}}\right)^n\right] \quad \text{(where } Z \sim N(0,1)\text{).} \end{aligned} \tag{4.3}$$

Thus,

$$FWER_{Bon}(n,\alpha,\rho) = 1 - \mathbb{E}_Z\left[\Phi\left(\frac{c_{\alpha,n}+\sqrt{\rho}Z}{\sqrt{1-\rho}}\right)^n\right] \quad (\text{where } Z \sim N(0,1)). \tag{4.4}$$

Remark 3. *An anonymous reviewer pointed out that one can generalize (4.4) to unequal correlations having a product structure: $\rho_{ij} = \lambda_i\lambda_j$ where $\lambda_i \geq 0$ for each i. This correlation structure arises, e.g., in one-way ANOVA with $n+1$ groups with a_0 observations on the control group ($i=0$) and a_i observations on the i-th test group ($i=1,\ldots,n$). To see this, let $\bar{Y}_i \sim N(\gamma_i, \sigma^2/a_i), i=0,\ldots,n$, be independent sample means with known variance $\sigma^2 > 0$ and $\gamma_i \geq \gamma_0$ for $i=1,\ldots,n$. Note that this problem is an extension of the ANOVA example considered in Section 1.5.3.*

Suppose we wish to test

$$\tilde{H}_i : \gamma_i = \gamma_0 \quad vs \quad \tilde{K}_i : \gamma_i > \gamma_0 \quad for \quad i = 1,\ldots,n$$

based on the test statistics

$$T_i = \left[\frac{1}{a_0}+\frac{1}{a_i}\right]^{-1/2}\frac{\left(\bar{Y}_i - \bar{Y}_0\right)}{\sigma}, \quad i=1,\ldots,n.$$

These control vs. test group contrasts have product correlation structure

$$\mathrm{Cor}\left(T_i, T_j\right) = \rho_{ij} = \sqrt{\frac{a_i a_j}{\left(a_0+a_i\right)\left(a_0+a_j\right)}}.$$

This implies

$$\lambda_i = \sqrt{\frac{a_i}{a_0+a_i}}, \quad i=1,\ldots,n.$$

In this case it is easy to show that the Bonferroni FWER can be represented as

$$1 - \mathbb{E}_Z\left[\prod_{i=1}^{n}\Phi\left(\frac{c_{\alpha,n}+\lambda_i Z}{\sqrt{1-\lambda_i^2}}\right)\right] \quad (\textit{where } Z \sim N(0,1)). \tag{4.5}$$

This essentially expresses the exact value of the Bonferroni FWER for a more general class of covariance matrices but in a very restrictive set-up.

We study the limiting behavior of $H_n(\rho)$ as $n \to \infty$ in the next section.

4.3 Asymptotic Behavior of FWER

Firstly, we state the foremost theoretical result of this chapter.

Theorem 4.3.1. $\lim_{n\to\infty} FWER_{Bon}(n,\alpha,\rho) = 0$ *for all* $\alpha \in (0,1)$ *and* $\rho \in (0,1]$.

Theorem 4.3.1, a much stronger result than Corollary 2.2.1, highlights the fundamental problem of Bonferroni method as a MTP.

Remark 4. *One observes from the definition of* $H_n(\rho)$ *that* $H_n(0) = (1-\alpha/n)^n \to e^{-\alpha}$ *as* n *approaches infinity. So,* $\lim_{n\to\infty} FWER(n,\alpha,0) = 1 - e^{-\alpha}$. *Combining this with Theorem 4.3.1, we obtain that the limiting Bonferroni FWER is convex in* ρ, *though discontinuous at* 0. *Thus, Theorem 4.3.1 implies Theorem 2.2.1 established in Das and Bhandari (2021).*

We note that proving Theorem 4.3.1 is equivalent to showing $\lim_{n\to\infty} H_n(\rho) = 1$. We establish this in the following steps:

(a) Finding an approximation for $c_{\alpha,n}$ for large n.

(b) With the help of the approximation, showing for each α and ρ in $(0,1)$,

$$\lim_{n\to\infty}\left[\Phi\left(\frac{c_{\alpha,n}}{\sqrt{1-\rho}}\right)\right]^n = 1.$$

(c) Showing that for any fixed real number t and for each α, ρ in $(0,1)$,

$$\lim_{n\to\infty}\left[\Phi\left(\frac{c_{\alpha,n}+t}{\sqrt{1-\rho}}\right)\right]^n = 1.$$

We explicate the steps in three lemmas.

Lemma 4.3.1. *Given any* $\alpha \in (0,1)$, $c_{\alpha,n} \le \sqrt{2\ln(n)}$ *for all sufficiently large* n. *Also,* $\frac{c_{\alpha,n}}{\sqrt{2\ln(n)}} \longrightarrow 1$ *as* $n \to \infty$.

The lemma follows from utilizing the following well-known result by Gordon (1941) once one replaces x by $c_{\alpha,n}$ and observes that $\Phi(c_{\alpha,n}) = 1 - \alpha/n$.

Theorem 4.3.2. *(Gordon, 1941) For arbitrary positive number* $x > 0$, *the inequalities*

$$\frac{x\phi(x)}{1+x^2} < 1 - \Phi(x) < \frac{\phi(x)}{x}$$

hold. In particular,

$$\lim_{x\to\infty}\frac{x(1-\Phi(x))}{\phi(x)} = 1.$$

Lemma 4.3.2. *For each α and ρ in* $(0, 1)$,

$$\lim_{n\to\infty}\left[\Phi\left(\frac{c_{\alpha,n}}{\sqrt{1-\rho}}\right)\right]^n = 1. \tag{4.6}$$

Proof of Lemma 4.3.2. Observe that

$$\begin{aligned}
\lim_{n\to\infty}\left[\Phi\left(\frac{\sqrt{2\ln n}}{\sqrt{1-\rho}}\right)\right]^n &= \lim_{n\to\infty}\left[\Phi\left(\sqrt{2\ln\left(n^{\frac{1}{1-\rho}}\right)}\right)\right]^n \\
&= \lim_{n\to\infty}\left[\left[\Phi\left(\sqrt{2\ln\left(n^{\frac{1}{1-\rho}}\right)}\right)\right]^{n^{\frac{1}{1-\rho}}}\right]^{n^{\frac{-\rho}{1-\rho}}} \\
&= \lim_{n\to\infty}\left[\left[\Phi(\sqrt{2\ln m})\right]^m\right]^{n^{\frac{-\rho}{1-\rho}}} \quad (\text{here } m = n^{\frac{1}{1-\rho}}). \qquad (4.7)
\end{aligned}$$

Invoking Lemma 4.3.1, we obtain from (4.7)

$$\begin{aligned}
\lim_{n\to\infty}\left[\Phi\left(\frac{\sqrt{2\ln n}}{\sqrt{1-\rho}}\right)\right]^n &\geq \lim_{n\to\infty}\left[\left[\Phi(c_{\alpha,m})\right]^m\right]^{n^{\frac{-\rho}{1-\rho}}} \\
&= \lim_{n\to\infty}\left[e^{-\alpha}\right]^{n^{\frac{-\rho}{1-\rho}}}, \qquad (4.8)
\end{aligned}$$

where the last step emanates from the definition of $c_{\alpha,n}$ and the fact that $(1-\alpha/m)^m \to e^{-\alpha}$ as m goes to infinity.

For $\rho = 0$, $\lim_{n\to\infty}\left[e^{-\alpha}\right]^{n^{\frac{-\rho}{1-\rho}}} = e^{-\alpha}$ and for $\rho \in (0, 1)$, $\lim_{n\to\infty}\left[e^{-\alpha}\right]^{n^{\frac{-\rho}{1-\rho}}} = 1$. Therefore,

$$\lim_{n\to\infty}\left[\Phi\left(\frac{\sqrt{2\ln n}}{\sqrt{1-\rho}}\right)\right]^n = 1 \quad \text{for each } \rho \in (0, 1). \tag{4.9}$$

Lemma 4.3.1 enables us to choose $\rho_1 \in (0, \rho]$ such that

$$\left[\Phi\left(\frac{c_{\alpha,n}}{\sqrt{1-\rho}}\right)\right]^n \geq \left[\Phi\left(\frac{\sqrt{2\ln n}}{\sqrt{1-\rho_1}}\right)\right]^n \quad \text{for all sufficiently large } n.$$

The rest follows from (4.9). $\square$

We shall now prove the following generalization of Lemma 4.3.2:

Lemma 4.3.3. *For any fixed real number t, we have*

$$\lim_{n\to\infty}\left[\Phi\left(\frac{c_{\alpha,n}+t}{\sqrt{1-\rho}}\right)\right]^n = 1 \quad \forall\alpha\in(0,1),\forall\rho\in(0,1). \tag{4.10}$$

Proof of Lemma 4.3.3. For positive values of t, the result is immediate from Lemma 4.3.2 using the increasing property of $\Phi(\cdot)$. For negative values of t, we choose $\rho_2 \in (0,\rho]$ such that

$$\left[\Phi\left(\frac{c_{\alpha,n}+t}{\sqrt{1-\rho}}\right)\right]^n \geq \left[\Phi\left(\frac{c_{\alpha,n}}{\sqrt{1-\rho_2}}\right)\right]^n \quad \text{for all sufficiently large } n.$$

Such a ρ_2 can always be chosen because t is fixed and $c_{\alpha,n} \longrightarrow \infty$ as $n \to \infty$ for any $\alpha \in (0,1)$. An application of Lemma 4.3.2 gives us (4.10). □

Now we establish Theorem 4.3.1.

Proof of Theorem 4.3.1. Let $Z \sim N(0,1)$ and $\Phi^n(\cdot)$ denote $[\Phi(\cdot)]^n$. We obtain from (4.9) and (4.10),

$$\lim_{n\to\infty}\left[\Phi^n\left(\frac{c_{\alpha,n}+\sqrt{\rho}Z}{\sqrt{1-\rho}}\right) - \Phi^n\left(\frac{c_{\alpha,n}}{\sqrt{1-\rho}}\right)\right] = 0 \quad \text{almost everywhere in } Z,$$

for each α and ρ in $(0,1)$. Since $\Phi(\cdot)$ is a bounded function, an application of dominated convergence theorem yields

$$\lim_{n\to\infty}\left[\mathbb{E}\left[\Phi^n\left(\frac{c_{\alpha,n}+\sqrt{\rho}Z}{\sqrt{1-\rho}}\right)\right] - \mathbb{E}\left[\Phi^n\left(\frac{c_{\alpha,n}}{\sqrt{1-\rho}}\right)\right]\right] = 0.$$

Applying Lemma 4.3.2, we have $\lim_{n\to\infty} H_n(\rho) = 1$ for each $\rho \in (0,1)$. Definition of $H_n(\rho)$ gives

$$\lim_{n\to\infty} FWER_{Bon}(n,\alpha,\rho) = 0 \quad \forall\rho\in(0,1).$$

Thus, the only thing remaining to show is $\lim_{n\to\infty} FWER(n,\alpha,1) = 0$. For $\rho = 1$, $X_i = X_j$ almost surely $\forall i \neq j$. Consequently, one rejection implies rejection of all null hypotheses and $H_n(\rho) = \mathbb{P}(X_1 \leq c_{\alpha,n}) = 1 - \alpha/n$. Hence, for $\rho = 1$ also, $H_n(\rho)$ tends to 1 as $n \to \infty$, completing the proof. □

Remark 5. *Proof of Theorem 4.3.1 provides us with the following approximation to FWER of Bonferroni procedure for large n:*

$$FWER_{Bon}(n,\alpha,\rho) \approx 1 - e^{-\alpha\cdot n^{\frac{-\rho}{1-\rho}}}.$$

We now mention an extension of Theorem 4.3.1 to general configurations of the null hypotheses in the following theorem:

Theorem 4.3.3. *Bonferroni $FWER$ tends to zero asymptotically as $n \longrightarrow \infty$ under any configuration of true and false null hypotheses.*

Proof of Theorem 4.3.3. Without loss of generality, we may assume that the set of true null hypotheses $\mathcal{A}$ is given by $\{1, \ldots, n_0\}$, where n_0 is the number of true nulls.

Under the equicorrelated setup, for each $i \in \mathcal{A}$, $X_i = \theta + Z_i$ where $\theta \sim N(0, \rho)$ is independent of $\{Z_n\}_{n \geq 1}$ and $Z_i \overset{iid}{\sim} N(0, 1-\rho)$. Now,

$$
\begin{aligned}
1 - FWER_{Bon}(n, \alpha, \rho) =& \mathbb{P}\left(X_i \leqslant c_{\alpha,n} \quad \forall i = 1, 2, \ldots, n_0\right) \\
=& \mathbb{P}\left(\theta + Z_i \leqslant c_{\alpha,n} \quad \forall i = 1, \ldots, n_0\right) \\
=& \mathbb{E}_\theta\left[\Phi^{n_0}\left(\frac{c_{\alpha,n} - \theta}{\sqrt{1-\rho}}\right)\right] \\
=& \mathbb{E}\left[\Phi^{n_0}\left(\frac{c_{\alpha,n} + \sqrt{\rho} Z}{\sqrt{1-\rho}}\right)\right] \quad (Z \sim N(0,1)) \\
\geqslant& \mathbb{E}\left[\Phi^{n}\left(\frac{c_{\alpha,n} + \sqrt{p} Z}{\sqrt{1-\rho}}\right)\right] \quad (\text{since } 0 \leqslant \Phi(\cdot) \leqslant 1 \text{ and } 1 \leqslant n_0 \leqslant n) \\
=& H_n(\rho) \longrightarrow 1 \text{ as } n \to \infty.
\end{aligned}
$$

The rest follows. □

We have considered an equicorrelated normal setup so far. However, problems involving variables with a more general dependence structure need to be addressed with more general correlation matrices. Hence, the study of the limiting behavior of FWER in arbitrarily correlated normal setups becomes crucial. Towards this, we consider the same *Gaussian sequence model* as in Section 4.2, but now we assume $\mathbb{C}\mathrm{orr}\left(X_i, X_j\right) = \rho_{ij}$ for $i \neq j$ where $\rho_{ij} \in (0, 1]$. Let $\mathbf{\Sigma}_n$ be the correlation matrix of $X_1, \ldots, X_n$ and $FWER(n, \alpha, \mathbf{\Sigma}_n)$ denote the FWER of Bonferroni's method under this setup. The following result generalizes Theorem 4.3.1 for this setup.

Theorem 4.3.4. *Suppose* $\liminf \rho_{ij} = \delta > 0$. *Then, for any* $\alpha \in (0, 1)$,

$$\lim_{n \to \infty} FWER_{Bon}(n, \alpha, \mathbf{\Sigma}_n) = 0$$

under any configuration of true and false null hypotheses.

We establish this using a famous inequality due to Slepian (1962):

Theorem 4.3.5. *Let* $\mathbf{X}$ *follow* $\mathbf{N}_k(\mathbf{0}, \mathbf{\Sigma})$*, where* $\mathbf{\Sigma}$ *is a* $k \times k$ *correlation matrix. Let* $\mathbf{a} = (a_1, \ldots, a_k)' \in \mathbb{R}^k$ *be arbitrary. Consider the quadrant probability*

$$g(k, \mathbf{a}, \mathbf{\Sigma}) = \mathbb{P}_{\mathbf{\Sigma}}\left[\bigcap_{i=1}^{k} \{X_i \leqslant a_i\}\right].$$

Let $\mathbf{R} = (\rho_{ij})$ *and* $\mathbf{T} = (\tau_{ij})$ *be two correlation matrices. If* $\rho_{ij} \geqslant \tau_{ij}$ *holds for all* i, j*, then* $g(k, \mathbf{a}, \mathbf{R}) \geq g(k, \mathbf{a}, \mathbf{T})$*, i.e*

$$\mathbb{P}_{\mathbf{\Sigma}=\mathbf{R}}\left[\bigcap_{i=1}^{k} \{X_i \leqslant a_i\}\right] \geqslant \mathbb{P}_{\mathbf{\Sigma}=\mathbf{T}}\left[\bigcap_{i=1}^{k} \{X_i \leqslant a_i\}\right]$$

holds for all $\mathbf{a} = (a_1, \ldots, a_k)'$*. Moreover, we have a strict inequality if* $\mathbf{R}, \mathbf{T}$ *are positive definite and if* $\rho_{ij} > \tau_{ij}$ *holds for some* i, j*.*

Proof of Theorem 4.3.4. We prove this only under the global null hypothesis. The proof under general configuration can be done similarly as the preceding proof. For fixed $n \in \mathbb{N}$, suppose

$$\mathcal{M}_n := \Big\{ i \in \{1, \ldots, n\} : \forall j \neq i, \rho_{ij} \geq \delta \Big\}.$$

Now,

$$\begin{aligned}
& FWER_{Bon}(n, \alpha, \mathbf{\Sigma}_n) \\
=& \mathbb{P}_{\Sigma_n}\Bigg(\bigcup_{i=1}^{n} \{X_i > c_{\alpha,n}\}\Bigg) \\
=& \mathbb{P}_{\Sigma_n}\Bigg(\bigcup_{i \in \mathcal{M}_n} \{X_i > c_{\alpha,n}\} \bigcup \bigcup_{i \notin \mathcal{M}_n} \{X_i > c_{\alpha,n}\}\Bigg) \\
\leq& \mathbb{P}\Bigg(\bigcup_{i \in \mathcal{M}_n} \{X_i > c_{\alpha,n}\}\Bigg) + \mathbb{P}\Bigg(\bigcup_{i \notin \mathcal{M}_n} \{X_i > c_{\alpha,n}\}\Bigg) \quad \text{(using Boole's inequality)} \\
\leq& \mathbb{P}\Bigg(\bigcup_{i \in \mathcal{M}_n} \{X_i > c_{\alpha,n}\}\Bigg) + [n - |\mathcal{M}_n|] \cdot \frac{\alpha}{n} \quad \text{(using Boole's inequality)} \\
\leq& \mathbb{P}\Bigg(\bigcup_{i \in \mathcal{M}_n} \{X_i > c_{\alpha,|\mathcal{M}_n|}\}\Bigg) + [n - |\mathcal{M}_n|] \cdot \frac{\alpha}{n} \quad \text{(since } n \geq |\mathcal{M}_n|\text{)} \\
=& 1 - g(|\mathcal{M}_n|, \mathbf{a}, \mathbf{\Sigma}_{\mathcal{M}_n}) + [n - |\mathcal{M}_n|] \cdot \frac{\alpha}{n}
\end{aligned}$$

where g is as defined in Theorem 4.3.5, $\mathbf{\Sigma}_{\mathcal{M}_n}$ is the covariance matrix of $(X_i : i \in \mathcal{M}_n)$, and $a_i = \Phi^{-1}(1 - \alpha/|\mathcal{M}_n|)$ for $i \in \mathcal{M}_n$.

Theorem 4.3.5 gives $g(|\mathcal{M}_n|, \mathbf{a}, \Sigma_{\mathcal{M}_n}) \geq g(|\mathcal{M}_n|, \mathbf{a}, M_{|\mathcal{M}_n|}(\delta))$. So,

$$\begin{aligned} FWER_{Bon}(n, \alpha, \mathbf{\Sigma}_n) &\leq 1 - g(|\mathcal{M}_n|, \mathbf{a}, M_{|\mathcal{M}_n|}(\delta)) + [n - |\mathcal{M}_n|] \cdot \frac{\alpha}{n} \\ &= FWER_{Bon}(|\mathcal{M}_n|, \alpha, \delta) + [n - |\mathcal{M}_n|] \cdot \frac{\alpha}{n}. \end{aligned}$$

Since $\liminf \rho_{ij} = \delta > 0$, we also have $n - |\mathcal{M}_n|$ is finite. The rest follows from Theorem 4.3.3 by taking $n \to \infty$. □

Remark 6. *Slepian's inequality is valid for any n. So it can be used to extend the upper bounds established in Chapter 2 to non-negatively correlated set-ups by letting $\rho^* = \min_{1 \leq i,j \leq n} \rho_{ij}$ as the common correlation. However the following example (given by an anonymous reviewer) illustrates the conservatism involved in replacing all the ρ_{ij} by their minimum ρ^*. Consider a 4×4 correlation matrix with the ρ_{ij} having a product structure (as in Remark 3) with $\lambda_1 = 0.1$ and $\lambda_2 = \lambda_3 = \lambda_4 = 0.9$. Then*

$$\rho_{12} = \rho_{13} = \rho_{14} = 0.09 \quad \text{and} \quad \rho_{23} = \rho_{24} = \rho_{34} = 0.81 \tag{4.11}$$

and $\rho^* = \min_{1 \leq i,j \leq n} \rho_{ij} = 0.09$. If we choose $c_{.05,4} = 2.24$, which is approximately the 5% critical point of the four-variate standard normal with common correlation $= 0.10$ then we get the upper bound $= 0.0498$ calculated using (4.4). But the exact probability calculated using (4.5) is 0.0385. So the upper bound is conservative by almost 30%.

4.4 Generalized familywise error rates

Consider the arbitrarily correlated setup described in Section 4.3. Let n_0 be the cardinality of the set of true null hypotheses. Without loss of generality, we may assume that the true null hypotheses are given by H_{0i}, $1 \leq i \leq n_0$. For the Lehmann and Romano (2005) k-FWER controlling procedure,

$$k\text{-}FWER(n, \alpha, \mathbf{\Sigma}_n) = \mathbb{P}_{\mathbf{\Sigma}_n}\left(X_i > \Phi^{-1}(1 - k\alpha/n) \text{ for at least } k \ \ i\text{'s}, 1 \leq i \leq n_0\right).$$

We now extend Theorem 4.3.4 for generalized familywise error rates.

Corollary 4.4.1. *Suppose $\liminf \rho_{ij} = \delta > 0$. Then, for any $\alpha \in (0,1)$ and any positive integer k satisfying $k\alpha < 1$, $\lim_{n\to\infty} k\text{-}FWER(n, \alpha, \mathbf{\Sigma}_n) = 0$.*

This follows from Theorem 4.3.4 once one observes that, for any natural number k with $k\alpha < 1$,

$$k\text{-}FWER(n, \alpha, \mathbf{\Sigma}_n) \leq FWER_{Bon}(n, \alpha^* = k\alpha, \mathbf{\Sigma}_n).$$

We have assumed that the value of k is fixed so far. However, in many applications it may be more practical to consider a setting where the k is a function of n. The following theorem shows that Corollary 4.4.1 still holds under such settings.

Theorem 4.4.1. *Suppose* $\liminf \rho_{ij} = \rho > 0$. *Then, for any* $\alpha \in (0,1)$ *and any positive integer valued sequence* $\{k_n\}_{n\in\mathbb{N}}$ *satisfying* $k_n = \mathrm{o}(n^\rho)$,

$$\lim_{n\to\infty} k_n\text{-}FWER(n,\alpha,\mathbf{\Sigma}_n) = 0.$$

Note that $k_n = \mathrm{o}(n^\rho)$ implies $k_n\alpha/n < 1$ for all sufficiently large n.

Proof of Theorem 4.4.1. We have

$$\begin{aligned}
& k_n FWER(n,\alpha,\Sigma_n) \\
= \;& \mathbb{P}_{\Sigma_n}\big(X_i > \Phi^{-1}(1-k_n\alpha/n) \text{ for at least } k\ i\text{'s}, 1\le i\le n_0\big) \\
\le \;& \mathbb{P}_{\Sigma_n}\big(X_i > \Phi^{-1}(1-k_n\alpha/n) \text{ for at least one } i, 1\le i\le n_0\big) \\
\le \;& \mathbb{P}_{\Sigma_n}\big(X_i > \Phi^{-1}(1-k_n\alpha/n) \text{ for at least one } i, 1\le i\le n\big) \\
\le \;& \mathbb{P}_{M_n(\rho)}\big(X_i > \Phi^{-1}(1-k_n\alpha/n) \text{ for at least one } i, 1\le i\le n\big) \text{ (using Theorem 4.3.5)} \\
= \;& 1-\mathbb{E}\left[\Phi^n\left(\frac{c_{\alpha,n/k_n}+\sqrt{\rho}Z}{\sqrt{1-\rho}}\right)\right]. \qquad (4.12)
\end{aligned}$$

We observe that

$$\begin{aligned}
\lim_{n\to\infty}\left[\Phi\left(\sqrt{\frac{2\ln(n/k_n)}{1-\rho}}\right)\right]^n &= \lim_{n\to\infty}\left[\Phi\left(\sqrt{2\ln\left(n^{\star\frac{1}{1-\rho}}\right)}\right)\right]^n \quad (\text{where } n^\star = n/k_n) \\
&= \lim_{n\to\infty}\left[\left[\Phi\left(\sqrt{2\ln\left(n^{\star\frac{1}{1-\rho}}\right)}\right)\right]^{n^{\star\frac{1}{1-\rho}}}\right]^{\frac{n}{n^{\star\frac{1}{1-\rho}}}} \\
&= \lim_{n\to\infty}\left[\left[\Phi(\sqrt{2\ln p})\right]^p\right]^{\frac{n}{n^{\star\frac{1}{1-\rho}}}} \quad (\text{here } p = (n^\star)^{\frac{1}{1-\rho}}).
\end{aligned}$$

Proceeding similarly as in Theorem 4.3.1, one can show that

$$\lim_{n\to\infty}\mathbb{E}\left[\Phi^n\left(\frac{c_{\alpha,n/k_n}+\sqrt{\rho}Z}{\sqrt{1-\rho}}\right)\right] = 1 \quad \text{if} \quad \frac{n}{n^{\star\frac{1}{1-\rho}}} \to 0.$$

Consequently, from (4.12) we have $k_n FWER(n,\alpha,\Sigma_n)$ goes to zero if

$$\frac{n}{n^{\star\frac{1}{1-\rho}}} \to 0 \iff \frac{k_n^{\frac{1}{1-\rho}}}{n^{\frac{\rho}{1-\rho}}} \to 0 \iff \frac{k_n}{n^\rho} \to 0 \iff k_n = \mathrm{o}(n^\rho),$$

completing the proof. □

4.5 Power Analysis

We discuss now the asymptotic power of Bonferroni's procedure under the equicorrelated normal setup. We consider the following notion of power for a MTP T described in Section 1.3.5:

$$AnyPwr_T = \mathbb{P}(S_n(T) \geq 1),$$

where $S_n(T)$ denotes the number of true positives in MTP T.

Without loss of generality, we assume $X_i \sim N(\mu_i, 1)$ $(\mu_i > 0)$ for $1 \leq i \leq n_1$ and for $n_1 < i \leq n$, $X_i \sim N(0,1)$. The following two results describe the asymptotic power of Bonferroni's method.

Theorem 4.5.1. *Consider the equicorrelated normal setup with equicorrelation $\rho \in (0,1)$. Suppose $\sup \mu_i$ is finite. Then, for any $\alpha \in (0,1)$, $AnyPwr_{Bon}$ goes to zero as $n \to \infty$.*

Proof of Theorem 4.5.1. We have,

$$\begin{aligned}
1 - AnyPwr_{Bon} =& \mathbb{P}(S_n(Bonferroni) = 0)\\
=& \mathbb{P}\left(X_i \leqslant c_{\alpha,n} \quad \forall i = 1,2,\ldots,n_1\right)\\
=& \mathbb{P}\left(X_i - \mu_i \leqslant c_{\alpha,n} - \mu_i \quad \forall i = 1,2,\ldots,n_1\right)\\
=& \mathbb{P}\left(\theta + Z_i \leqslant c_{\alpha,n} - \mu_i \quad \forall i = 1,\ldots,n_1\right)
\end{aligned}$$

where the last step utilizes the fact that, under our setup, for each $1 \leq i \leq n_1$, $X_i - \mu_i = \theta + Z_i$ where $\theta \sim \mathcal{N}(0,\rho)$ is independent of $\{Z_n\}_{n\geq 1}$ and $Z_i \overset{iid}{\sim} \mathcal{N}(0, 1-\rho)$. Thus,

$$\begin{aligned}
1 - AnyPwr_{Bon} =& \mathbb{E}_\theta\left[\prod_{i=1}^{n_1} \Phi\left(\frac{c_{\alpha,n} - \theta - \mu_i}{\sqrt{1-\rho}}\right)\right]\\
=& \mathbb{E}\left[\prod_{i=1}^{n_1} \Phi\left(\frac{c_{\alpha,n} + \sqrt{\rho}Z - \mu_i}{\sqrt{1-\rho}}\right)\right] \quad (Z \sim N(0,1))\\
\geq& \mathbb{E}\left[\Phi^{n_1}\left(\frac{c_{\alpha,n} + \sqrt{\rho}Z - \mu^\star}{\sqrt{1-\rho}}\right)\right] \quad (\text{where } \mu^\star = \sup \mu_i < \infty)\\
\geq& \mathbb{E}\left[\Phi^{n}\left(\frac{c_{\alpha,n} + \sqrt{\rho}Z - \mu^\star}{\sqrt{1-\rho}}\right)\right] \quad (\text{since } 0 \leqslant \Phi(\cdot) \leqslant 1 \text{ and } 1 \leqslant n_0 \leqslant n).
\end{aligned}$$

Proceeding similarly as in Theorem 4.3.1, one obtains

$$\lim_{n\to\infty} \mathbb{E}\left[\Phi^n\left(\frac{c_{\alpha,n}+\sqrt{\rho}Z-\mu^\star}{\sqrt{1-\rho}}\right)\right] = 1.$$

The rest is obvious. □

Theorem 4.5.2. *Suppose $n_1 \to \infty$ and $n_1/n \to p_1 \in (0,1]$ as $n \to \infty$. Then, $AnyPwr_{Bon}$ goes to one as $n \to \infty$ if $\frac{\sqrt{2\log n_1}}{\mu_{n_1}} \longrightarrow 0$ as $n_1 \to \infty$.*

Proof of Theorem 4.5.2. We have,

$$\begin{aligned} c_{\alpha,n} = \Phi^{-1}\left(1-\frac{\alpha}{n}\right) &= \Phi^{-1}\left(1-\frac{p_1\alpha}{n_1}\cdot\frac{n_1}{np_1}\right) \\ &\approx \Phi^{-1}\left(1-\frac{p_1\alpha}{n_1}\right) \quad \text{for sufficiently large } n\text{, since } n_1/n \to p_1 \\ &= c_{p_1\alpha,n_1}. \end{aligned}$$

The proof of the preceding theorem gives,

$$\begin{aligned} 1 - AnyPwr_{Bon} =& \mathbb{E}\left[\prod_{i=1}^{n_1}\Phi\left(\frac{c_{\alpha,n}+\sqrt{\rho}Z-\mu_i}{\sqrt{1-\rho}}\right)\right] \quad (Z\sim N(0,1)) \\ \leq& \mathbb{E}\left[\Phi\left(\frac{c_{\alpha,n}+\sqrt{\rho}Z-\mu_{n_1}}{\sqrt{1-\rho}}\right)\right] \quad (\text{since } 0\leqslant\Phi(\cdot)\leqslant 1) \\ \approx& \mathbb{E}\left[\Phi\left(\frac{c_{p_1\alpha,n_1}+\sqrt{\rho}Z-\mu_{n_1}}{\sqrt{1-\rho}}\right)\right] \\ \longrightarrow& 0 \text{ as } n_1\to\infty \quad (\text{since } c_{p_1\alpha,n_1}-\mu_{n_1}\to-\infty \text{ as } n_1\to\infty). \end{aligned}$$

The rest follows. □

4.6 Simulation Study

We have mentioned in Section 4.2 that under H_0, $X_i = \theta + Z_i$ where $\theta \sim N(0,\rho)$, independent of $\{Z_n\}_{n\geq 1}$ and $Z_i \overset{iid}{\sim} N(0,1-\rho)$. Equation (4.2) gives us

$$\begin{aligned} FWER_{Bon}(n,\alpha,\rho) &= \mathbb{P}_{H_0}\left(\bigcup_{i=1}^{n}\{Z_i+\theta > c_{\alpha,n}\}\right) \\ &= \mathbb{P}_{H_0}\left(\max_{1\leq i\leq n} Z_i > c_{\alpha,n}-\theta\right) \\ &= \mathbb{P}_{H_0}\left(\max_{1\leq i\leq n} W_i > \frac{c_{\alpha,n}-\sqrt{\rho}\beta}{\sqrt{1-\rho}}\right) \quad (\text{where } W_i = Z_i/\sqrt{1-\rho},\ \theta=\sqrt{\rho}\beta) \end{aligned}$$

$$= \mathbb{E}_\beta \left[\mathbb{I}\left\{ \max_{1\le i\le n} W_i > C_{\beta,\rho} \right\} \right] \tag{4.13}$$

where $\mathbb{I}\{A\}$ is the indicator variable of event A and $C_{\beta,\rho}$ is the quantity it is replacing. Note that $\beta \sim N(0,1)$ independent of $W_i \sim N(0,1)$ under H_0.

Equation (4.13) illustrates an elegant and computationally less expensive simulation scheme of estimating FWER given (n, α, ρ). Firstly, we generate 100000 independent observations from $N(0,1)$ (these are the β variables, i.e the repetitions). Given ρ, we compute the cutoff $C_{\beta_i,\rho}$ for each of the simulated β_i's, $1 \le i \le 100000$. Given n, for each β_i, we generate n independent observations from $N(0,1)$ (these are the W_i's) and compute the maximum of these n observations. We note for how many i's, the maximum (obtained from i'th sample of n independent standard normals) exceeds the cutoff $C_{\beta_i,\rho}$. An estimate of FWER(n, α, ρ) is obtained accordingly from the 100000 repetitions.

Table 4.1: Estimates of $FWER_{Bon}(n, \alpha = .05, \rho)$

Correlation	Number of hypotheses (n)					
(ρ)	100	1000	10000	100000	1 Million	10 Million
0	.04878247	.04877176	.04877069	.04877059	.04877058	.04877058
0.1	.04697	.04601	.0431	.04295	.03965	.03893
0.2	.04285	.039	.03473	.03166	.02703	.02496
0.3	.03749	.03116	.0252	.02146	.01601	.01445
0.4	.03149	.02332	.01741	.01314	.00915	.00748
0.5	.02532	.01656	.01113	.00716	.00464	.00341
0.6	.01982	.01145	.00653	.00374	.00206	.00142
0.7	.01435	.00678	.0034	.0016	.00078	.00049
0.8	.00904	.00356	.00139	.00052	.00029	.00001
0.9	.00448	.00132	.0004	.00006	.00003	0

Table 4.1 presents the estimated FWERs under the equicorrelated normal setup for $\alpha = .05$ and some values of n. It also gives the exact values of $FWER_{Bon}(\cdot,\cdot,0)$. We observe that, for each positive ρ, FWER values decrease as n grows while for each n, the values decrease as ρ increases. Also, the rate of decay is much faster for higher values of ρ.

Let

$$B_n(\rho_1,\rho_2,\rho_3) = \begin{pmatrix} M_n(\rho_1) & 0_n & 0_n \\ 0_n & M_n(\rho_2) & 0_n \\ 0_n & 0_n & M_n(\rho_3) \end{pmatrix}$$

where 0_n denotes the $n \times n$ matrix of all zeroes. Table 4.2 presents the FWER values under the $B_n(\rho_1,\rho_2,\rho_3)$ for different combinations of (ρ_1,ρ_2,ρ_3) for $\alpha = .05$ and for different

Table 4.2: Estimates of $FWER_{Bon}(B_n(\rho_1, \rho_2, \rho_3))$

Block Correlation Values	Number of hypotheses ($3n$)				
(ρ_1, ρ_2, ρ_3)	900	9000	90000	900000	9 Million
(.1, .1, .1)	.04672	.04606	.04606	.04542	.04345
(.1, .1, .5)	.03982	.03682	.03484	.03237	.03113
(.1, .1, .9)	.03246	.03093	.03082	.02982	.02922
(.1, .5, .5)	.03269	.02754	.02350	.02056	.01863
(.1, .5, .9)	.02528	.02160	.01943	.01799	.01670
(.1, .9, .9)	.01822	.01608	.01533	.01517	.01497
(.5, .5, .5)	.02519	.01801	.01250	.00823	.00553
(.5, .5, .9)	.01772	.01201	.00838	.00562	.00358
(.5, .9, .9)	.01061	.00643	.00423	.00277	.00182
(.9, .9, .9)	.00310	.00098	.00024	.00009	.00001

choices of n. One observes that for each (ρ_1, ρ_2, ρ_3), FWER values decrease as n grows. The simulation scheme for the block-equicorrelated setup is similar to the equicorrelated setup and hence omitted.

4.7 Concluding Remarks

FWER and k-FWER are widely used in DNA microarray analyses. It is well-known that Bonferroni's method becomes very stringent for large-scale multiple testing problems under the classical i.i.d framework. However, there is very little literature which elucidates the extent of conservativeness of Bonferroni's method under dependence. Our results address this gap in a unified manner. Even more importantly, by considering k-FWER instead of FWER, we encompass a much broader class of error rates.

Chapter 5

Asymptotic Behaviors of Stepwise Multiple Testing Procedures

5.1 Introduction

In chapters 2 and 3, we have obtained upper bounds on the Bonferroni FWER and the generalized FWERs (of Lehmann-Romano procedure) in the equicorrelated and general Gaussian setups with small and moderate dimensions. In Chapter 4, we have improved this result by showing that the Bonferroni FWER and the Lehmann-Romano gFWER asymptotically go to zero for any strictly positive ρ. We have also extended this to arbitrarily correlated setups where the limiting infimum of the correlations is positive.

We have focused on Bonferroni-type single-step procedures so far. Stepwise methods constitute one of the most successful approaches to FWER control (Efron, 2010a). Finner and Roters (2002) studied the number of false rejections in single-step, step-down and step-up methods under independence. However, the role of correlation on the limiting behavior of the FWER for stepwise procedures is little-known. Also, the existing literature lacks theoretical justifications for why FWER methods fail in large-scale problems.

The chapter addresses this problem by theoretically investigating the limiting FWER values of general step-down procedures under the correlated Gaussian setup. These results provide new insights into the behavior of step-down decision procedures. By establishing the limiting performances of commonly used step-up methods, e.g., the Benjamini-Hochberg method and the Hochberg method, we have elucidated that the class of step-up procedures does not possess a similar *universal asymptotic zero* result as obtained in the case of step-down procedures. It is also noteworthy that most of our results are pretty general since they accommodate any combination of the null hypotheses.

We have also obtained the limiting powers of the stepwise procedures.

This chapter is structured as follows. We first formally introduce the framework with relevant notations in the next section. Section 5.3 presents the limiting behaviors of the FWER of step-down procedures in equicorrelated and non-negatively correlated Normal setups. Section 5.4 discusses similar results on Hochberg's and Benjamini-Hochberg's procedures. Hommel's stepwise MTP is studied in Section 5.5. We outline our contributions and discuss related problems briefly in Section 5.6.

5.2 Preliminaries

We discuss the simultaneous inference problem through a *Gaussian sequence model*:

$$X_i \sim N(\mu_i, 1), \quad i \in \{1, \ldots, n\}$$

where $(X_1, \ldots, X_n)$ have covariance matrix Σ_n. Suppose

$$H_{0i} : \mu_i = 0 \quad vs \quad H_{1i} : \mu_i > 0, \quad 1 \leq i \leq n.$$

The FWER of a MTP T is defined as

$$FWER_T(n, \alpha, \Sigma_n) = \mathbb{P}_{\Sigma_n}(V_n(T) \geq 1) \tag{5.1}$$

where $V_n(T)$ denotes the number of false rejections made by T. For simplicity, we write $FWER_T(n, \alpha, M_n(\rho))$ as $FWER_T(n, \alpha, \rho)$. As in the previous chapter, we shall consider the probability in the r.h.s of (5.1) under the intersection null H_0 at first (and take that as the definition of FWER) for technical simplicity. Then we shall extend the results obtained in this case to any combination of false and true null hypotheses.

In Chapter 4, we have shown in Theorem 4.3.1 that the Bonferroni FWER asymptotically approaches zero for any strictly positive ρ. The proof of Theorem 4.3.1 utilizes the exchangeability of the sequence $\{X_r\}_{r\geq 1}$ under H_0 for the equicorrelated set-up. For each $i \geq 1$, $X_i \stackrel{d}{=} \theta + Z_i$ where $\theta \sim N(0, \rho)$ is independent of $\{Z_n\}_{n\geq 1}$ and $Z_i \stackrel{i.i.d}{\sim} N(0, 1-\rho)$. This representation of X_i as the sum of two independent components θ and Z_i would be repeatedly applied in the proofs of the results of this chapter also.

Throughout this chapter, P_i denotes the p-value for H_{0i}, $1 \leq i \leq n$. Also, let $P_{(1)} \leqslant \ldots \leqslant P_{(n)}$ be the ordered p-values. We denote the null hypothesis corresponding to the p-value $P_{(i)}$ as $H_{0(i)}$, $1 \leq i \leq n$.

5.3 Limiting FWER and Power of Step-down MTPs

Holm's method (Holm, 1979), one of the earliest example of a step-down procedure, uses modified critical values and utilizes the Bonferroni inequality. It rejects $H_{0(i)}$ if

$$\forall j \in \{1, \dots i\}, \qquad P_{(j)} \leq \frac{\alpha}{n-j+1}.$$

Holm's step-down procedure controls FWER under any dependence. We have the following result on the limiting FWER of Holm's method under the equicorrelated normal framework:

Theorem 5.3.1. *Suppose* $\mu^\star = \sup \mu_i < \infty$. *Then, under any configuration of false and true null hypotheses, we have*

$$\lim_{n\to\infty} FWER_{Holm}(n, \alpha, \rho) = 0 \quad \textit{for all } \alpha \in (0,1) \textit{ and } \rho \in (0,1].$$

Proof of Theorem 5.3.1. We have,

$$\begin{aligned} FWER_{Holm}(n, \alpha, \rho) =& \mathbb{P}_{M_n(\rho)}(V_n(Holm) \geq 1) \\ \leq& \mathbb{P}_{M_n(\rho)}(R_n(Holm) \geq 1) \\ =& \mathbb{P}_{M_n(\rho)}(P_{(1)} \leq \alpha/n) \\ =& \mathbb{P}_{M_n(\rho)}(X_{(n)} \geq c_{\alpha,n}) \\ =& 1 - \mathbb{P}_{M_n(\rho)}\left(X_i \leqslant c_{\alpha,n} \quad \forall i = 1, 2, \dots, n\right). \end{aligned}$$

Without loss of generality, we assume (in this and the next proof) $X_i \sim N(\mu_i, 1)$ $(\mu_i > 0)$ for $1 \leq i \leq n_1$ and for $n_1 < i \leq n$, $X_i \sim N(0,1)$. This gives,

$$\begin{aligned} FWER_{Holm}(n, \alpha, \rho) \leq& 1 - \mathbb{P}\left[\bigcap_{i=1}^{n_1} \{\theta + Z_i + \mu_i \leqslant c_{\alpha,n}\} \bigcap \bigcap_{i=n_1+1}^{n} \{\theta + Z_i \leqslant c_{\alpha,n}\}\right] \\ =& 1 - \mathbb{E}_\theta\left[\prod_{i=1}^{n_1} \Phi\left(\frac{c_{\alpha,n} - \theta - \mu_i}{\sqrt{1-\rho}}\right) \Phi^{n-n_1}\left(\frac{c_{\alpha,n} - \theta}{\sqrt{1-\rho}}\right)\right] \\ \leq& 1 - \mathbb{E}_\theta\left[\Phi^n\left(\frac{c_{\alpha,n} - \theta - \mu^\star}{\sqrt{1-\rho}}\right)\right] \\ \longrightarrow& 1 - 1 = 0 \text{ as } n \to \infty \quad (\text{since } \mu^\star < \infty). \end{aligned}$$

□

We now extend this result to arbitrary correlated normal setups:

Theorem 5.3.2. *Let Σ_n be the correlation matrix of $X_1, \ldots, X_n$ with (i,j)'th entry ρ_{ij} such that* $\liminf \rho_{ij} = \delta > 0$. *Suppose $\mu^\star = \sup \mu_i < \infty$. Then, for any $\alpha \in (0,1)$,*

$$\lim_{n\to\infty} \mathbb{P}_{\Sigma_n}(R_n(Holm) \geq 1) = \lim_{n\to\infty} \mathbb{P}_{\Sigma_n}(P_{(1)} \leq \alpha/n) = 0$$

under any configuration of false and true null hypotheses. Consequently,

$$\lim_{n\to\infty} FWER_{Holm}(n, \alpha, \Sigma_n) = 0 \quad \textit{and} \quad \lim_{n\to\infty} AnyPwr_{Holm}(n, \alpha, \Sigma_n) = 0.$$

Proof of Theorem 5.3.2. We have,

$$\begin{aligned}
&\mathbb{P}_{\Sigma_n}(R_n(Holm) \geq 1)\\
=&\mathbb{P}_{\Sigma_n}(P_{(1)} \leq \alpha/n)\\
=&\mathbb{P}_{\Sigma_n}(X_{(n)} \geq c_{\alpha,n})\\
\leq&\mathbb{P}_{M_n(\rho)}(X_{(n)} \geq c_{\alpha,n}) \quad \text{(using Theorem 4.3.5)}\\
=&1 - \mathbb{P}_{M_n(\rho)}\left(X_i \leqslant c_{\alpha,n} \quad \forall i = 1, 2, \ldots, n\right)\\
=&1 - \mathbb{P}_{M_n(\rho)}\left[\bigcap_{i=1}^{n_1}\{\theta + Z_i + \mu_i \leqslant c_{\alpha,n}\} \bigcap \bigcap_{i=n_1+1}^{n} \{\theta + Z_i \leqslant c_{\alpha,n}\}\right]\\
=&1 - \mathbb{E}_\theta\left[\left\{\prod_{i=1}^{n_1} \Phi\left(\frac{c_{\alpha,n} - \theta - \mu_i}{\sqrt{1-\rho}}\right)\right\} \cdot \Phi^{n-n_1}\left(\frac{c_{\alpha,n} - \theta}{\sqrt{1-\rho}}\right)\right]\\
\leq&1 - \mathbb{E}_\theta\left[\Phi^n\left(\frac{c_{\alpha,n} - \theta - \mu^\star}{\sqrt{1-\rho}}\right)\right].
\end{aligned}$$

The last quantity above tends to zero asymptotically since $\mu^\star < \infty$. The rest follows by noting that $R_n(Holm) \geq \max\{V_n(Holm), S_n(Holm)\}$. $\square$

We extend Theorem 5.3.2 to any step-down MTP below:

Theorem 5.3.3. *Let Σ_n be the correlation matrix of $X_1, \ldots, X_n$ with (i,j)'th entry ρ_{ij} such that* $\liminf \rho_{ij} = \delta > 0$. *Suppose* $\sup \mu_i$ *is finite and T is any step-down MTP controlling FWER at $\alpha \in (0,1)$. Then, for any $\alpha \in (0,1)$,*

$$\lim_{n\to\infty} \mathbb{P}_{\Sigma_n}\Big(R_n(T) \geq 1\Big) = 0$$

under any configuration of false and true null hypotheses. Consequently,

$$\lim_{n\to\infty} FWER_T(n, \alpha, \Sigma_n) = 0 \quad \textit{and} \quad \lim_{n\to\infty} AnyPwr_T(n, \alpha, \Sigma_n) = 0.$$

We establish Theorem 5.3.3 using the following result due to Gordon and Salzman (Gordon and Salzman, 2008).

Theorem 5.3.4. *Let T be a step-down MTP based on the set of cut-offs $\mathbf{u} \in \mathcal{S}_n$. If $FWER_T \leq \alpha < 1$, then $u_i \leq \alpha/(n-i+1), i = 1, \dots, n$.*

Proof of Theorem 5.3.3. We have,

$$\begin{aligned} FWER_T(n, \alpha, \Sigma_n) = \mathbb{P}_{\Sigma_n}(V_n(T) \geq 1) \leq \mathbb{P}_{\Sigma_n}(R_n(T) \geq 1) =& \mathbb{P}_{\Sigma_n}(P_{(1)} \leq u_1) \\ \leq& \mathbb{P}_{\Sigma_n}(P_{(1)} \leq \alpha/n). \end{aligned}$$

The last step above follows since we have $u_1 \leq \alpha/n$ from Theorem 5.3.4. The rest is obvious from Theorem 5.3.2. □

Theorem 5.3.3 can be viewed as a *universal asymptotic zero* result since it encompasses all step-down FWER controlling MTPs and also accommodates any configuration of the nulls.

5.4 Limiting FWER of Some Step-up MTPs

Let us consider a step-down MTP T_1 and a step-up MTP T_2 having the identical vector of cutoffs $\mathbf{u} = (u_1, \dots, u_n) \in \mathcal{S}_n$. We always have $m_1(T_1) \leq m_2(T_2)$ where $m_1(T_1) = \max\left\{i : P_{(j)} \leq u_j \text{ for all } j = 1, \dots, i\right\}$ and $m_2(T_2) = \max\left\{i : P_{(i)} \leq u_i\right\}$. This implies that the step-up MTP is at least as rejective as the step-down MTP (which uses the same cutoffs). This observation steers that we might not get a similar *universal asymptotic zero* result for the class of step-up MTPs as obtained for step-down MTPs (Theorem 5.3.3). This is indeed the case as we shall show in the next two subsections the following:

(a) Under the equicorrelated Gaussian sequence model, the FWER of Hochberg MTP (Hochberg, 1988) asymptotically approaches zero as the number of tests becomes arbitrarily large.

(b) Under the equicorrelated Gaussian sequence model and under H_0, the BH method (Benjamini and Hochberg, 1995) with a pre-specified FDR level controls FDR at some strictly positive quantity which is a function of the chosen FDR level and the common correlation, even when the number of tests approaches infinity.

We have considered Hochberg's MTP in particular because it uses the same vector of cutoffs as Holm's MTP (note that Holm's MTP has the 'optimal' critical values in the class of step-down procedures). Benjamini-Hochberg method, on the other hand, has been one of the most eminent MTPs proposed in the literature and also possesses some optimality properties both in frequentist and Bayesian paradigms of simultaneous inference (see Bogdan et al. (2011); Guo and Rao (2008)).

5.4.1 Hochberg's Procedure

Hochberg's (Hochberg, 1988) MTP and Holm's sequentially rejective procedure use the same set of cutoffs; and hence, as mentioned earlier, Hochberg's MTP is sharper than Holm's MTP. The following result depicts the limiting behavior of the FWER of Hochberg's method under the correlated Gaussian sequence model:

Theorem 5.4.1. *Consider the correlated normal setup with common correlation $\rho \in [0, 1)$. Then,*

(a) When $\rho = 0$ (i.e., the independent normal setup), we have

$$\lim_{n\to\infty} FWER_{Hochberg}(n, \alpha, 0) \in [1 - e^{-\alpha}, \alpha]$$

under the global null.

(b) When $\rho \in (0, 1)$, we have

$$\lim_{n\to\infty} FWER_{Hochberg}(n, \alpha, \rho) = 0$$

for any $\alpha \in (0, 1/2)$, under the global null hypothesis.

Proof of Theorem 5.4.1. We have, under the global null,

$$FWER_{Hochberg}(n, \alpha, 0) = \mathbb{P}_{I_n}\left[\bigcup_{i=1}^{n}\left\{P_{(i)} \leqslant \frac{\alpha}{n-i+1}\right\}\right] \geqslant \mathbb{P}_{I_n}\left[P_{(1)} \leqslant \frac{\alpha}{n}\right] \xrightarrow[n\to\infty]{} 1 - e^{-\alpha}.$$

Also, Hochberg's procedure controls FWER at level α (Hochberg, 1988). So,

$$1 - e^{-\alpha} \leqslant \lim_{n\to\infty} FWER_{Hochberg}(0) \leqslant \alpha.$$

Also, $\lim_{\alpha\to 0} \frac{1-e^{-\alpha}}{\alpha} = 1$. thus, we have, as $\alpha \to 0$, $\lim_{n\to\infty} \frac{FWER_{Hochberg}(0)}{\alpha} = 1$. This completes the proof of the first part.

When $\rho \in (0,1)$, we have,

$$\begin{aligned}
&P_{(i)} \leqslant \frac{\alpha}{n-i+1} \\
\Longleftrightarrow &1 - \frac{\alpha}{n-i+1} \leq \Phi\left(X_{(n-i+1)}\right) \\
\Longleftrightarrow &\Phi^{-1}\left(1 - \frac{\alpha}{n-i+1}\right) \leqslant U + Z_{(n-i+1)} \\
\Longleftrightarrow &\Phi^{-1}\left(1 - \frac{\alpha}{n-i+1}\right) \leqslant U + \sqrt{1-\rho}\cdot\Phi^{-1}\left(1-\frac{i}{n}\right) \text{ (for all sufficiently large } n).
\end{aligned}$$

Therefore, for all sufficiently large values of n, we have

$$\begin{aligned}
&P_{(i)} \leqslant \frac{\alpha}{n-i+1} \\
\Longleftrightarrow &-U - \sqrt{1-\rho}\cdot\Phi^{-1}\left(1-\frac{i}{n}\right) \leqslant -\Phi^{-1}\left(1-\frac{\alpha}{n-i+1}\right) \\
\Longleftrightarrow &-U + \sqrt{1-\rho}\cdot\Phi^{-1}\left(\frac{i}{n}\right) \leqslant \Phi^{-1}\left(\frac{\alpha}{n-i+1}\right) \\
\Longleftrightarrow &\frac{-U}{\Phi^{-1}\left(\frac{\alpha}{n-i+1}\right)} + \frac{\sqrt{1-\rho}\cdot\Phi^{-1}\left(\frac{i}{n}\right)}{\Phi^{-1}\left(\frac{\alpha}{n-i+1}\right)} \geqslant 1 \quad (\text{since } \alpha \in (0,1/2)) \\
\Longleftrightarrow &\lim_{n\to\infty} \frac{\Phi^{-1}\left(\frac{i}{n}\right)}{\Phi^{-1}\left(\frac{\alpha}{n-i+1}\right)} \geq \frac{1}{\sqrt{1-\rho}}.
\end{aligned}$$

Thus, we have $i/n < 1/2$, because otherwise the limiting ratio of $\Phi^{-1}\left(\frac{i}{n}\right)$ and $\Phi^{-1}\left(\frac{\alpha}{n-i+1}\right)$ can not be positive. So, we have

$$\frac{i}{n} < \frac{\alpha}{n-i+1} < 1/2.$$

This implies $i(n-i+1) < \alpha \cdot n$. But this is not valid for any value of i in $\{1,\ldots,n\}$. Consequently, the limiting FWER is zero. $\square$

Theorem 5.4.2. *Consider the multiple testing problem under the equicorrelated normal setup with equicorrelation $\rho \in (0,1)$. Suppose $\sup \mu_i$ is finite. Then,*

$$\lim_{n\to\infty} \mathbb{P}_{M_n(\rho)}\Big(R_n(Hochberg) \geq 1\Big) = 0$$

for any $\alpha \in (0,1/2)$. Consequently, for any $\alpha \in (0,1/2)$,

$$\lim_{n\to\infty} FWER_{Hochberg}(n,\alpha,M_n(\rho)) = \lim_{n\to\infty} AnyPwr_{Hochberg}(n,\alpha,M_n(\rho)) = 0.$$

Proof of Theorem 5.4.2. This proof is similar to the preceding proof. Here, we have $X_i = U + W_i$ for each $i \geq 1$. Here $U \sim N(0,\rho)$ is independent of $\{W_n\}_{n\geq 1}$ and $W_i = \mu_i + V_i$.

Here V_i's are i.i.d $N(0, 1-\rho)$ and μ_i is zero if $i \in \mathcal{A}$ and positive otherwise. So, W_i always lies in $[V_i, V_i + \sup \mu_i]$. This implies, for all $i \in \{1, \ldots, n\}$,

$$W_{(n-i+1)} \leq V_{(n-i+1)} + \sup \mu_i.$$

We have,

$$\begin{aligned}
\mathbb{P}_{M_n(\rho)}\Big(R_n(Hochberg) \geq 1\Big) =& \mathbb{P}_{M_n(\rho)}\left[\bigcup_{i=1}^{n}\left\{P_{(i)} \leqslant \frac{\alpha}{n-i+1}\right\}\right] \\
=& \mathbb{P}_{M_n(\rho)}\left[\bigcup_{i=1}^{n}\left\{\Phi^{-1}\left(1-\frac{\alpha}{n-i+1}\right) \leq U + W_{(n-i+1)}\right\}\right] \\
\leq& \mathbb{P}_{M_n(\rho)}\left[\bigcup_{i=1}^{n}\left\{\Phi^{-1}\left(1-\frac{\alpha}{n-i+1}\right) \leq U + V_{(n-i+1)} + \sup \mu_i\right\}\right].
\end{aligned}$$

Therefore, for all sufficiently large values of n, we have

$$\begin{aligned}
& P_{(i)} \leqslant \frac{\alpha}{n-i+1} \\
\Longrightarrow & -U - \sqrt{1-\rho} \cdot \Phi^{-1}\left(1-\frac{i}{n}\right) - \sup \mu_i \leqslant -\Phi^{-1}\left(1-\frac{\alpha}{n-i+1}\right) \\
\Longleftrightarrow & -U + \sqrt{1-\rho} \cdot \Phi^{-1}\left(\frac{i}{n}\right) - \sup \mu_i \leqslant \Phi^{-1}\left(\frac{\alpha}{n-i+1}\right) \\
\Longleftrightarrow & \frac{-U}{\Phi^{-1}\left(\frac{\alpha}{n-i+1}\right)} + \frac{\sqrt{1-\rho} \cdot \Phi^{-1}\left(\frac{i}{n}\right)}{\Phi^{-1}\left(\frac{\alpha}{n-i+1}\right)} - \frac{\sup \mu_i}{\Phi^{-1}\left(\frac{\alpha}{n-i+1}\right)} \geqslant 1 \quad (\text{since } \alpha \in (0, 1/2)) \\
\Longleftrightarrow & \lim_{n\to\infty} \frac{\Phi^{-1}\left(\frac{i}{n}\right)}{\Phi^{-1}\left(\frac{\alpha}{n-i+1}\right)} \geq \frac{1}{\sqrt{1-\rho}}.
\end{aligned}$$

Proceeding exactly in the same way as in the earlier proof, one obtains that the probability of rejecting any null hypothesis approaches zero. □

5.4.2 Benjamini-Hochberg Procedure

The BH MTP is the first FDR controlling method (Finner et al., 2007). Let $i_{\max}$ be the largest such i for which $p_{(i)} \leqslant i\alpha/n$. The BH procedure rejects $H_{0(i)}$ if $i \leqslant i_{\max}$ and accepts $H_{0(i)}$ otherwise. The following result evaluates the limiting FDR of Benjamini-Hochberg method:

Theorem 5.4.3. *Consider the multiple testing problem under the equicorrelated normal setup with correlation ρ. Then, under the global null, for all $\alpha \in (0, 1)$ and for all*

$\rho \in (0,1)$,

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

Also, $\lim_{n\to\infty} FDR_{BH}(n,\alpha,0) = \lim_{n\to\infty} FDR_{BH}(n,\alpha,1) = \alpha$.

Proof of Theorem 5.4.3. We have

$$FDR_{BH} = \mathbb{E}\left[\frac{V_n(BH)}{\max\{R_n(BH),1\}}\right] = \mathbb{E}\left[\frac{V_n(BH)}{R_n(BH)} \mid V_n(BH) > 0\right]\mathbb{P}\left(V_n(BH) > 0\right).$$

Under the global null H_0, each rejection is a false rejection and FDR equals FWER. We shall work with $FWER$ for the rest of this proof.

Suppose exactly n_0 null hypotheses are true. Then, it is a well-known fact (Benjamini and Hochberg, 1995; Efron, 2010a; Sarkar, 2002) that, under the independent setup,

$$FDR_{BH}(n,\alpha,\rho) = \frac{n_0}{n}\alpha.$$

So, under the global null, $FDR_{BH}(n,\alpha,0) = FWER_{BH}(n,\alpha,0) = \alpha$. Now,

$$\begin{aligned} p_{(i)} \leqslant \frac{i\alpha}{n} &\Longleftrightarrow 1 - \Phi\left(X_{(n-i+1)}\right) \leqslant \frac{i\alpha}{n} \\ &\Longleftrightarrow 1 - \frac{i\alpha}{n} \leqslant \Phi\left(X_{(n-i+1)}\right) \\ &\Longleftrightarrow X_{(n-i+1)} \geqslant \Phi^{-1}\left(1 - \frac{i\alpha}{n}\right). \end{aligned}$$

Consequently,

$$\begin{aligned} FWER_{BH}(\rho) =& \mathbb{P}_{M_n(\rho)}\left[\bigcup_{i=1}^{n}\left\{P_{(i)} < \frac{i\alpha}{n}\right\}\right] \\ =& \mathbb{P}_{M_n(\rho)}\left[\bigcup_{i=1}^{n}\left\{X_{(n-i+1)} \geqslant \Phi^{-1}\left(1 - \frac{i\alpha}{n}\right)\right\}\right]. \end{aligned}$$

When $\rho = 1$, $X_i = X_j$ w.p 1. This implies

$$FWER_{BH}(\rho) = \mathbb{P}\left[\bigcup_{i=1}^{n}\left\{X \geqslant \Phi^{-1}\left(1 - \frac{i\alpha}{n}\right)\right\}\right] = \mathbb{P}\left[X \geqslant \Phi^{-1}(1-\alpha)\right] = \alpha,$$

where $X \sim N(0,1)$.

Consider the case $0 < \rho < 1$ now. Then, $X_i = U_\rho + Z_i$ where $U_\rho \sim N(0,\rho)$ is independent of $Z_i \sim N(0, 1-\rho)$. Here Z_i's are i.i.d. So, under the global null,

$$\begin{aligned}
FWER_{BH}(\rho) =& \mathbb{P}_{M_n(\rho)}\left[\bigcup_{i=1}^{n}\left\{X_{(n-i+1)} \geqslant \Phi^{-1}\left(1-\frac{i\alpha}{n}\right)\right\}\right] \\
=& \mathbb{P}_{M_n(\rho)}\left[\bigcup_{i=1}^{n}\left\{U_\rho + Z_{(n-i+1)} \geqslant \Phi^{-1}\left(1-\frac{i\alpha}{n}\right)\right\}\right] \\
=& \mathbb{P}_{M_n(\rho)}\left[\bigcup_{i=1}^{n}\left\{U_\rho > \Phi^{-1}\left(1-t_i\alpha\right) - Z_{(n-nt_i+1)}\right\}\right] \quad \text{where } t_i = i/n. \\
=& \mathbb{P}_{M_n(\rho)}\left[\bigcup_{i=1}^{n}\left\{M > \frac{\Phi^{-1}\left(1-t_i\alpha\right) - Z_{(n-nt_i+1)}}{\sqrt{\rho}}\right\}\right],
\end{aligned}$$

where in the last step above, $M = U_\rho/\sqrt{\rho} \sim N(0,1)$. For $t \in (0,1)$, $Z_{(n-nt+1)} = Z_{\left(n\left(1-t+\frac{1}{n}\right)\right)}$ converges in probability to $(1-t)$'th quantile of the distribution of Z_i (i.e. $\sqrt{1-\rho}\cdot\Phi^{-1}(1-t)$) as $n \to \infty$. So,

$$\begin{aligned}
\lim_{n\to\infty} FWER_{BH}(\rho) &= \mathbb{P}\left[\bigcup_{t\in(0,1)}\left\{M > \frac{\Phi^{-1}\left(1-t\alpha\right) - \sqrt{1-\rho}\cdot\Phi^{-1}(1-t)}{\sqrt{\rho}}\right\}\right] \\
&= \mathbb{P}\left[\bigcup_{t\in(0,1)}\left\{M > s(t)\right\}\right] \quad \text{(say)} \\
&= \mathbb{P}\left[M > \inf_{t\in(0,1)} s(t)\right] \\
&= 1 - \Phi\left[\inf_{t\in(0,1)} s(t)\right].
\end{aligned}$$

Now, $\inf_{t\in(0,1)} s(t) \leq s(.5) < \infty$. So, $\Phi\left[\inf_{t\in(0,1)} s(t)\right] < 1$. Thus, $\lim_{n\to\infty} FWER_{BH}(\rho) > 0$ for $\rho \in (0,1)$. $\square$

Remark 7. *Since we are considering the infimum of the function $s(\cdot)$ and since $s(0) = \infty = s(1)$, the previous result still holds good if one considers the closed interval $[0,1]$ in place of the open interval $(0,1)$.*

Remark 8. *Finner et al. (2007) studied the (limiting) empirical distribution function of the p-values and used those to study limiting behaviors of FDR. Their results are derived under general distributional setups and different values of ξ_n where ξ_n denotes the proportion of the true nulls. Our elementary proof, in contrast, uses standard analytic tools and provides a simple closed-form expression for the limiting FDR under the global null.*

5.4.3 Other Step-up MTPs

We have discussed the limiting FWER values of two step-up procedures so far. In this subsection, we shall provide an upper bound on the limiting FWER of any step-up method satisfying some properties. Towards this, we consider a special dependency property of test statistics introduced by Benjamini and Yekutieli (2001) (see Section 1.5.1, Definition 2).

Guo and Rao (2008) showed the following.

Lemma 5.4.1. *Let T be any step-up MTP having vector of critical values $\alpha_T \in \mathcal{S}_n$. Then we have the following inequality under the PRDS property:*

$$\sum_{k=1}^{n} \mathbb{P}\left(R_n(T) = k \mid P_i \leq \alpha_k\right) \leq 1, \quad \text{for } i \in \mathcal{A}. \tag{5.2}$$

Moreover, this inequality becomes an equality under the independence of the test statistics.

They also constructed an example of joint distribution of the p-values, under which the PRDS property fails to hold although the inequality (5.2) holds. Thus, it turns out (5.2) is a relaxed condition than PRDS. Guo and Rao (2008) further showed the following optimality property of the BH method:

Theorem 5.4.4. *Let $\mathcal{T}$ be the class of all step-up procedures with vector of cutoffs belonging to $\mathcal{S}_n$ and satisfying the inequality (5.2). Then, the BH method is optimal in the class $\mathcal{T}$. In other words, for any step-up method $T \in \mathcal{T}$ with vector of cutoffs $\alpha_T \in \mathcal{S}_n$, which controls FDR at α, then $\alpha_k \leq k\alpha/n$ for each $k \in \{1, \ldots, n\}$.*

Theorem 5.4.3 and Theorem 5.4.4 result in the following:

Theorem 5.4.5. *Let $\mathcal{T}$ be the class of all step-up procedures with vector of cutoffs belonging to $\mathcal{S}_n$ and satisfying the inequality (5.2). Let $T \in \mathcal{T}$ be such that it controls the FDR at $\alpha \in (0,1)$. Consider the equicorrelated normal setup with correlation ρ. Then, under the global null, for all $\alpha \in (0,1)$ and for all $\rho \in (0,1)$,*

$$\lim_{n\to\infty} FDR_T(n, \alpha, \rho) \leq 1 - \Phi\left[\inf_{t\in(0,1)} \frac{\Phi^{-1}(1-t\alpha) - \sqrt{1-\rho}\cdot\Phi^{-1}(1-t)}{\sqrt{\rho}}\right].$$

5.5 Hommel's Procedure

We have focused on step-down and step-up procedures so far. However, many powerful MTPs proposed in the literature do not belong to the step-down or step-up categories.

The Hommel (Hommel, 1988) procedure is such a p-value based MTP that controls the FWER:

Step 1. Compute $j = \max\left\{i \in \{1,\dots,n\} : P_{(n-i+k)} > k\alpha/i \text{ for } k = 1,\dots,i\right\}$.

Step 2. If the maximum does not exist in Step 1, reject each null hypothesis.

Otherwise, reject all H_i with $P_i \leqslant \alpha/j$.

Hommel's MTP is uniformly more powerful than the Bonferroni, Holm, and Hochberg methods (Gou et al., 2014). The following two results depict the limiting behavior of the FWER of Hommel's procedure under the independent normal setup and under the positively equicorrelated normal setup, respectively.

Theorem 5.5.1. *Consider the multiple testing problem under the independent normal setup. Under the global null, we have*

$$\lim_{n\to\infty} FWER_{Hommel}(n,\alpha,0) = 1 - e^{-\alpha}.$$

Theorem 5.5.2. *Consider the multiple testing problem under the equicorrelated normal framework with correlation $\rho \in (0,1)$. Then, for any $\alpha \in (0,1)$,*

$$\lim_{n\to\infty} FWER_{Hommel}(n,\alpha,\rho) = 0$$

with probability one under the global null hypothesis.

Theorem 5.5.3. *Consider the equicorrelated normal setup with equicorrelation $\rho \in (0,1)$. Suppose $\sup \mu_i$ is finite. Then, for any $\alpha \in (0,1)$,*

$$\lim_{n\to\infty} \mathbb{P}_{M_n(\rho)}\Big(R_n(Hommel) \geq 1\Big) = 0$$

with probability one under any configuration of false and true null hypotheses. Consequently, $FWER_{Hommel}(n,\alpha,\rho)$ and $AnyPwr_{Hommel}(n,\alpha,M_n(\rho))$ tend to zero with probability one as $n \to \infty$.

Proof of Theorem 5.5.1. Note that $P_{(i)} = 1 - \Phi\left(X_{(n-i+1)}\right)$. Putting $i = n - j + k$ (here $1 \leq j \leq n$ and $1 \leq k \leq j$) gives $P_{(n-j+k)} = 1 - \Phi\left(X_{(j-k+1)}\right)$, $k \leqslant j$. Now,

$$\begin{aligned}
P_{(n-j+k)} > \frac{k\alpha}{j} &\iff 1 - \Phi\left(X_{(j-k+1)}\right) > \frac{k\alpha}{j} \\
&\iff \Phi^{-1}\left(1 - \frac{k\alpha}{j}\right) > X_{(j-k+1)} \\
&\iff \Phi^{-1}\left(1 - \frac{s\alpha}{t}\right) > X_{(n(t-s)+1)} \quad \text{where } s = k/n \text{ and } t = j/n.
\end{aligned}$$

For any $r \in (0,1)$, $X_{(nr)}$ converges in probability to r'th quantile of the distribution of X_1 as $n \to \infty$. This implies, $X_{(n(t-s)+1)}$ converges in probability to $\Phi^{-1}(t-s)$ as $n \to \infty$. Thus, as $n \to \infty$,

$$\begin{aligned} P_{(n-j+k)} > \frac{k\alpha}{j} &\iff \Phi^{-1}\left(1 - \frac{s\alpha}{t}\right) > \Phi^{-1}(t-s) \\ &\iff 1 - \frac{s\alpha}{t} > t - s \\ &\iff t - s\alpha > t(t-s) \\ &\iff t(1-t) > s(\alpha - t). \end{aligned}$$

We have $t \geq s$ and $1 > \alpha$. So, $t(1-t) > s(\alpha - t)$ always holds. This means that the largest t for which $t(1-t) > s(\alpha - t)$ holds for each $s \in (0, t]$ is 1. This in turn implies that, as $n \to \infty$, the largest integer $j \leq n$ satisfying $P_{(n-j+k)} > \frac{k\alpha}{j}$ for $k \in \{1, \dots, j\}$ is n with probability one. Thus, the Hommel's MTP is same as the Bonferroni's MTP as $n \to \infty$. Hence,

$$\lim_{n\to\infty} FWER_{Hommel}(n, \alpha, 0) = 1 - e^{-\alpha}.$$

□

Proof of Theorem 5.5.2. Consider the equicorrelated normal framework with correlation $\rho \in (0,1)$. Under the global null, we have $X_i = U + Z_i$ for each $i \geq 1$. Here $U \sim N(0, \rho)$ is independent of $\{Z_n\}_{n\geq 1}$ and $Z_i \overset{iid}{\sim} N(0, 1-\rho)$.

We establish Theorem 5.5.2 in the following steps:

(a) Showing that as $n \to \infty$,

$$P_{(n-j+k)} > \frac{k\alpha}{j} \quad \text{for all } k = 1, \dots, j \iff U < \min_{0<s<t} f(s)$$

where $f(s) = \Phi^{-1}(1 - s\alpha/t) - \sqrt{1-\rho} \cdot \Phi^{-1}(t-s)$.

(b) Showing that

$$\Phi\left(\frac{-U - \Phi^{-1}(\alpha)}{\sqrt{1-p}}\right) > t \text{ implies } U < \min_{0<s<t} f(s).$$

(c) Showing that, for each positive integer m,

$$FWER_{Hommel}(n, \alpha, \rho) \leq \mathbb{P}\left[P_{(1)} \leqslant \frac{1}{t_0} \cdot \frac{\alpha}{n}\right] + \mathbb{P}(U \geq m)$$

where $t_0 = \max_t \{t \in (0,1) : \min_{0<s<t} f(s) > U\}$.

We explicate the steps now. Similar to the previous proof, we have

$$\begin{aligned}
P_{(n-j+k)} > \frac{k\alpha}{j} &\iff 1 - \Phi\left(X_{(j-k+1)}\right) > \frac{k\alpha}{j} \\
&\iff \Phi^{-1}\left(1 - \frac{k\alpha}{j}\right) > X_{(j-k+1)} \\
&\iff \Phi^{-1}\left(1 - \frac{k\alpha}{j}\right) > U + Z_{(j-k+1)} \\
&\iff \Phi^{-1}\left(1 - \frac{s\alpha}{t}\right) > U + Z_{(n(t-s)+1)} \quad \text{where } s = k/n, t = j/n.
\end{aligned}$$

For any $r \in (0,1)$, $Z_{(nr)}$ converges in probability to r'th quantile of the distribution of Z_1 as $n \to \infty$. This implies, $Z_{(n(t-s)+1)}$ converges in probability to $\sqrt{1-\rho} \cdot \Phi^{-1}(t-s)$ as $n \to \infty$. Thus, as $n \to \infty$,

$$P_{(n-j+k)} > \frac{k\alpha}{j} \iff U < \Phi^{-1}\left(1 - \frac{s\alpha}{t}\right) - \sqrt{1-\rho} \cdot \Phi^{-1}(t-s).$$

This means, as $n \to \infty$,

$$P_{(n-j+k)} > \frac{k\alpha}{j} \text{ for all } k = 1, \ldots, j \iff U < \min_{0<s<t} f(s) \tag{5.3}$$

completing the proof of step 1.

Now, $t > t - s$ as $s > 0$. This implies $\Phi^{-1}(t) > \Phi^{-1}(t-s)$. Consequently, for each $s > 0$, $f(s) > g(s)$ where $g(s) = \Phi^{-1}\left(1 - \frac{s\alpha}{t}\right) - \Phi^{-1}(t)$. Thus,

$$g(s) > U \implies f(s) > U.$$

Now,

$$\begin{aligned}
g(s) > U &\iff \Phi^{-1}\left(1 - \frac{s\alpha}{t}\right) - \sqrt{1-\rho} \cdot \Phi^{-1}(t) > U \\
&\iff \Phi^{-1}\left(1 - \frac{s\alpha}{t}\right) > U + \sqrt{1-\rho} \cdot \Phi^{-1}(t) \\
&\iff 1 - \frac{s\alpha}{t} > \Phi\left(U + \sqrt{1-\rho} \cdot \Phi^{-1}(t)\right) \\
&\iff \frac{\Phi\left(-U - \sqrt{1-\rho} \cdot \Phi^{-1}(t)\right)}{\alpha} > \frac{s}{t}.
\end{aligned}$$

Therefore, if $\frac{\Phi\left(-U-\sqrt{1-\rho}\cdot\Phi^{-1}(t)\right)}{\alpha} > 1$ then $\forall s \in (0,t)$, $g(s) > U$. Hence, $\Phi\left(-U - \sqrt{1-\rho} \cdot \Phi^{-1}(t)\right)/\alpha > 1$ implies $f(s) > U$ for all $s \in (0,t)$.

Now,

$$\frac{\Phi\left(-U-\sqrt{1-\rho}\cdot\Phi^{-1}(t)\right)}{\alpha}>1 \iff -U-\sqrt{1-\rho}\cdot\Phi^{-1}(t)>\Phi^{-1}(\alpha)$$
$$\iff \Phi\left(\frac{-U-\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right)>t.$$

Therefore, we have established the following:

$$\Phi\left(\frac{-U-\Phi^{-1}(\alpha)}{\sqrt{1-p}}\right)>t \text{ implies } U<\min_{0<s<t} f(s), \tag{5.4}$$

completing step 2.

Thus,

$$t_0:=\max_t\left\{t\in(0,1):\min_{0<s<t} f(s)\geqslant U\right\}\geqslant\max_{t\in(0,1)}\left\{\Phi\left(\frac{-U-\Phi^{-1}(\alpha)}{\sqrt{1-\rho}}\right)>t\right\}.$$

Now, $U<r$ implies $t_0\geqslant\varepsilon_r$ where

$$\varepsilon_r=\Phi\left(\frac{-r-\Phi^{-1}(\alpha)}{\sqrt{1-p}}\right).$$

So, for every $m\in\mathbb{N}$, there exists $\varepsilon_m>0$ such that $t_0>\varepsilon_m$ if $U<m$. In other words, there is ε_m such that $t_0>\varepsilon_m>0$ with probability at least $\mathbb{P}(U<m)$. This implies, t_0 is bounded away from zero with probability one. Now, let

$$j_0=\max_{1\leqslant j\leqslant n}\left\{P_{(n-j+k)}>\frac{k\alpha}{j} \text{ for } k=1,\ldots,j\right\}.$$

Evidently, $j_0\geqslant nt_0$. Consequently, under the global null,

$$\begin{aligned}
FWER_{Hommel}(n,\alpha,\rho)=&\mathbb{P}\left[\bigcup_{i=1}^{n}\left\{P_i\leqslant\frac{\alpha}{j_0}\right\}\right]\\
\leqslant&\mathbb{P}\left[\bigcup_{i=1}^{n}\left\{P_i\leqslant\frac{\alpha}{nt_0}\right\}\right]+\mathbb{P}(U\geqslant m)\\
=&\mathbb{P}\left[P_{(1)}\leqslant\frac{1}{t_0}\cdot\frac{\alpha}{n}\right]+\mathbb{P}(U\geqslant m).
\end{aligned}$$

This completes the proof of Step 3. Now, $\mathbb{P}(U\geq m)\leq\epsilon$ for all $\epsilon>0$ as $m\to\infty$. We claim now that

$$\mathbb{P}\left[P_{(1)}\leqslant\frac{1}{t_0}\cdot\frac{\alpha}{n}\right]\longrightarrow 0 \quad\text{as } n\to\infty.$$

Its proof is precisely the same as the proof of Theorem 4.3.1 and we therefore omit it.

The rest is obvious. $\square$

Remark 9. *Suppose $a > 0$. The proof of Theorem 4.3.1 also culminates in the following:*

$$\mathbb{P}_{M_n(\rho)}\left[P_{(1)} \leqslant a \cdot \frac{\alpha}{n}\right] \longrightarrow 0 \quad \text{as } n \to \infty$$

for each $\rho \in (0,1)$. Then, invoking Slepian's inequality, we have the following: Let Σ_n be the correlation matrix of $X_1, \ldots, X_n$ having (i,j)'th entry ρ_{ij} with $\liminf \rho_{ij} = \delta > 0$. Suppose $\mu^\star = \sup \mu_i < \infty$. Then, for any $\alpha \in (0,1)$,

$$\mathbb{P}_{\Sigma_n}\left[P_{(1)} \leqslant a \cdot \frac{\alpha}{n}\right] \longrightarrow 0 \quad \text{as } n \to \infty.$$

Note that this is a much stronger result than Theorem 5.3.2.

Proof of Theorem 5.5.3. This proof is similar to the earlier proof. We consider the equicorrelated normal framework with correlation $\rho \in (0,1)$. We have $X_i = U + W_i$ for each $i \geq 1$. Here $U \sim N(0,\rho)$ is independent of $\{W_n\}_{n\geq 1}$ and $W_i = \mu_i + V_i$. Here V_i's are i.i.d $N(0, 1-\rho)$ and μ_i is 0 if $i \in \mathcal{A}$ and positive otherwise.

We establish Theorem 5.5.3 in the following steps:

(a) Showing that as $n \to \infty$,

$$P_{(n-j+k)} > \frac{k\alpha}{j} \quad \text{for all } k = 1, \ldots, j \Longleftarrow U < \min_{0<s<t} f_2(s)$$

where $f_2(s) = \Phi^{-1}(1 - s\alpha/t) - \sqrt{1-\rho} \cdot \Phi^{-1}(t-s) - \mu^*$, $\mu^* := \sup \mu_i$.

(b) Showing that

$$\Phi\left(\frac{-U - \Phi^{-1}(\alpha) - \mu^*}{\sqrt{1-p}}\right) > t \text{ implies } U < \min_{0<s<t} f_2(s).$$

(c) Showing that, for each positive integer m,

$$FWER_{Hommel}(n, \alpha, \rho) \leq \mathbb{P}\left[P_{(1)} \leqslant \frac{1}{t_1} \cdot \frac{\alpha}{n}\right] + \mathbb{P}(U \geq m)$$

where $t_1 = \max_t \{t \in (0,1) : \min_{0<s<t} f_2(s) > U\}$.

We explicate the steps now.

Similar to the preceding proof, we have

$$\begin{aligned}
P_{(n-j+k)} > \frac{k\alpha}{j} &\iff 1-\Phi\left(X_{(j-k+1)}\right) > \frac{k\alpha}{j} \\
&\iff \Phi^{-1}\left(1-\frac{k\alpha}{j}\right) > X_{(j-k+1)} \\
&\Longleftarrow \Phi^{-1}\left(1-\frac{k\alpha}{j}\right) > U + V_{(j-k+1)} + \mu^* \\
&\iff \Phi^{-1}\left(1-\frac{s\alpha}{t}\right) > U + V_{(n(t-s)+1)} + \mu^* \quad \text{where } s = k/n, t = j/n.
\end{aligned}$$

For any $r \in (0,1)$, $V_{(nr)}$ converges in probability to r'th quantile of the distribution of V_1 as $n \to \infty$. This implies, $V_{(n(t-s)+1)}$ converges in probability to $\sqrt{1-\rho}\cdot\Phi^{-1}(t-s)$ as $n \to \infty$. Thus, as $n \to \infty$,

$$P_{(n-j+k)} > \frac{k\alpha}{j} \Longleftarrow U < \Phi^{-1}\left(1-\frac{s\alpha}{t}\right) - \sqrt{1-\rho}\cdot\Phi^{-1}(t-s) - \mu^*.$$

This means, as $n \to \infty$,

$$P_{(n-j+k)} > \frac{k\alpha}{j} \text{ for all } k = 1,\ldots,j \iff U < \min_{0<s<t} f_2(s) \tag{5.5}$$

completing the proof of step 1.

Now, $t > t-s$ as $s > 0$. This implies $\Phi^{-1}(t) > \Phi^{-1}(t-s)$. Consequently, for each $s > 0$, $f_2(s) > g_2(s)$ where $g_2(s) = \Phi^{-1}\left(1-\frac{s\alpha}{t}\right) - \Phi^{-1}(t) - \mu^*$. Thus,

$$g_2(s) > U \implies f_2(s) > U.$$

Now,

$$\begin{aligned}
g_2(s) > U &\iff \Phi^{-1}\left(1-\frac{s\alpha}{t}\right) - \sqrt{1-\rho}\cdot\Phi^{-1}(t) - \mu^* > U \\
&\iff \Phi^{-1}\left(1-\frac{s\alpha}{t}\right) > U + \sqrt{1-\rho}\cdot\Phi^{-1}(t) + \mu^* \\
&\iff 1-\frac{s\alpha}{t} > \Phi\left(U + \sqrt{1-\rho}\cdot\Phi^{-1}(t) + \mu^*\right) \\
&\iff \frac{\Phi\left(-U - \sqrt{1-\rho}\cdot\Phi^{-1}(t) - \mu^*\right)}{\alpha} > \frac{s}{t}.
\end{aligned}$$

Therefore, if $\frac{\Phi\left(-U-\sqrt{1-\rho}\cdot\Phi^{-1}(t)-\mu^*\right)}{\alpha} > 1$ then $\forall s \in (0,t)$, $g(s) > U$. Hence,

$\Phi\left(-U-\sqrt{1-\rho}\cdot\Phi^{-1}(t)-\mu^*\right)/\alpha > 1$ implies $f(s) > U$ for all $s \in (0,t)$. Now,

$$\begin{aligned}\frac{\Phi\left(-U-\sqrt{1-\rho}\cdot\Phi^{-1}(t)-\mu^*\right)}{\alpha} > 1 &\iff -U-\sqrt{1-\rho}\cdot\Phi^{-1}(t)-\mu^* > \Phi^{-1}(\alpha)\\ &\iff \Phi\left(\frac{-U-\Phi^{-1}(\alpha)-\mu^*}{\sqrt{1-\rho}}\right) > t.\end{aligned}$$

Therefore, we have established the following:

$$\Phi\left(\frac{-U-\Phi^{-1}(\alpha)-\mu^*}{\sqrt{1-p}}\right) > t \text{ implies } U < \min_{0<s<t} f_2(s), \tag{5.6}$$

completing step 2.

Thus,

$$t_1 := \max_t\left\{t \in (0,1) : \min_{0<s<t} f_2(s) \geqslant U\right\} \geqslant \max_{t\in(0,1)}\left\{\Phi\left(\frac{-U-\Phi^{-1}(\alpha)-\mu^*}{\sqrt{1-\rho}}\right) > t\right\}.$$

Now, $U < r$ implies $t_1 \geqslant \varepsilon_r$ where

$$\varepsilon_r = \Phi\left(\frac{-r-\Phi^{-1}(\alpha)-\mu^*}{\sqrt{1-p}}\right).$$

So, for every $m \in \mathbb{N}$, there exists $\varepsilon_m > 0$ such that $t_0 > \varepsilon_m$ if $U < m$. In other words, there is ε_m such that $t_1 > \varepsilon_m > 0$ with probability at least $\mathbb{P}(U < m)$. This implies, t_1 is bounded away from zero with probability one.

Now, let

$$j_1 = \max_{1\leqslant j\leqslant n}\left\{P_{(n-j+k)} > \frac{k\alpha}{j} \text{ for all } k = 1,\ldots,j\right\}.$$

Evidently, $j_1 \geqslant nt_1$. Consequently,

$$\begin{aligned}\mathbb{P}_{M_n(\rho)}\Big(R_n(Hommel) \geq 1\Big) &\leqslant \mathbb{P}\left[\bigcup_{i=1}^n\left\{P_i \leqslant \frac{\alpha}{j_1}\right\}\right]\\ &\leqslant \mathbb{P}\left[\bigcup_{i=1}^n\left\{P_i \leqslant \frac{\alpha}{nt_1}\right\}\right] + \mathbb{P}(U \geqslant m)\\ &= \mathbb{P}\left[P_{(1)} \leqslant \frac{1}{t_1}\cdot\frac{\alpha}{n}\right] + \mathbb{P}(U \geqslant m).\end{aligned}$$

This completes the proof of Step 3. Now, $\mathbb{P}(U \geq m) \leq \epsilon$ for all $\epsilon > 0$ as $m \to \infty$. We have earlier seen that

$$\mathbb{P}\left[P_{(1)} \leqslant \frac{1}{t_1}\cdot\frac{\alpha}{n}\right] \longrightarrow 0 \quad \text{as } n \to \infty.$$

The rest is obvious. □

5.6 Concluding Remarks

The preceding chapter sheds light on the extent of the conservativeness of the Bonferroni method under dependent setups. However, there is little literature on the effect of correlation on general step-down or step-up procedures. This chapter addresses this gap in a unified manner by investigating the limiting behaviors of several testing rules under the correlated Gaussian sequence model. We have proved asymptotic zero results for some popular MTPs controlling FWER at a pre-specified level. Specifically, we have shown that the limiting FWER approaches zero for any step-down rule provided the infimum of the correlations is strictly positive.

Huang and Hsu (2007) elucidate that both Holm's and Hochberg's methods are special cases of partition testing:

"while Holm's method tests each partition hypothesis using the largest order statistic, setting a critical value based on the Bonferroni inequality, Hochberg's method tests each partition hypothesis using all the order statistics, setting a series of critical values based on Simes' inequality."

It is natural to expect partition testing utilizing the joint distribution is sharper than partition testing based on probability inequalities. Our results elucidate that, at least under the correlated Gaussian sequence model setup with many hypotheses, Holm's MTP and Hochberg's MTP do not have significantly different performances in that they both have asymptotic zero FWER and asymptotic zero power.

The Benjamini-Hochberg procedure has been one of the most studied MTP and has several desirable optimality properties (Bogdan et al., 2011; Guo and Rao, 2008). It is astonishing to note that, among all the methods studied in this chapter, the BH procedure is the only one which can hold the FWER at a strictly positive level asymptotically under the equicorrelated normal setup. An interesting problem would be to study the limiting power of the Benjamini-Hochberg method.

Hommel's method is more rejective than Hochberg's MTP (and consequently, Holm's and Bonferroni's methods) Gou et al. (2014). Yet, within our chosen asymptotic framework, this has asymptotic zero FWER and asymptotic zero power.

Finally, there are possible scopes of interesting extensions in several directions. One extension is to consider more general distributional setups. Another is to study the limiting behaviors of Hochberg, Hommel, and Benjamini-Hochberg procedures under general dependent normality. The primary tool in establishing universal asymptotic zero results for the step-down MTPs is Slepian's inequality which compares the quadrant probabilities of two normal random vectors. However, for the step-up procedures, the FWERs become functions of several order statistics. Hence we can not directly apply Slepian's inequality in these scenarios. Indeed, Finner et al. (2007) remark that it is challenging to deal with false discoveries in models with complicated dependence structures, e.g., in a multivariate Gaussian model with a general covariance matrix. It is also interesting to theoretically investigate whether similar asymptotic results hold for other classes of MTPs, e.g., the class of consonant procedures (Westfall et al., 1999).

Chapter 6

Asymptotically Optimal Sequential Multiple Testing Procedures for Correlated Normal

6.1 Introduction

Simultaneous inference has been a cornerstone in the statistics methodology literature, particularly because of its fundamental theory and paramount applications. The mainstream multiple testing literature has traditionally considered two frameworks:

(a) The sample size is deterministic, i.e., the full data is available while testing. The classical Bonferroni method or the multiple testing procedures (MTPs henceforth) proposed by Benjamini and Hochberg (1995), Holm (1979), Hochberg (1988), Hommel (1988) - each of them is valid under the fixed sample size paradigm.

(b) The test statistics corresponding to the various tests are independent. For example, the BH method was initially shown to control FDR under independence (Benjamini and Hochberg, 1995). The Sidak's procedure is valid under independence and for test statistics with certain parametric distributions.

However, in many modern applications, these assumptions are routinely violated:

(a) Quite often, the data is streaming or arriving sequentially. In multiple-endpoint clinical trials, patients are collected sequentially. At each interim step, the researcher has to decide whether to collect more observations or stop. Of late, MTPs that can handle these kinds of sequential data have been proposed and studied.

When working in simultaneous inference problems in sequential setups, one might consider the natural generalization of Wald's sequential framework where all the data streams are terminated simultaneously. This setup finds application in multiple access wireless communication (Rappaport, 2002) and multisensor surveillance platforms (Foresti et al., 2003). The last decade has witnessed significant progress on this line of work: De and Baron (2012a,b, 2015), Song and Fellouris (2017, 2019), Song (2019), He and Bartroff (2021), Roy et al. (2023).

(b) Simultaneous inference problems arising in many disciplines often involve correlated observations. For example, in microRNA expression data, several genes may cluster into groups through their transcription processes and exhibit high correlations. Functional magnetic resonance imaging (fMRI) studies and multistage clinical trials also concern dependent observations. Simultaneous testing methods under dependence have been studied by Efron (2007, 2010b), Liu et al. (2016), Sun and Cai (2009), Xie et al. (2011), among others. Fan et al. (2012) proposed a novel approach of tackling dependent test statistics with a known correlation structure. They capture the association between correlated statistics using the principal eigenvalues of the covariance matrix. Fan and Han (2016) extended this work to unknown covariance structures. Qiu et al. (2005) demonstrated that many FDR controlling procedures lose power significantly under dependence. Huang and Hsu (2007) mention that stepwise decision rules based on modeling of the dependence structure are generally superior to their counterparts that do not consider the correlation.

However, we find little literature which studies the multiple testing problem in a sequential framework where the test statistics corresponding to the various streams are dependent. This chapter fills this gap in a unified way by considering the classical means-testing problem in a equicorrelated Gaussian and sequential framework.

We organize the chapter as follows. The next section introduces the framework with necessary notations and mentions some existing results on the sequential test rules. We discuss the asymptotic expansion of average sample number required by the sequential probability ratio test in Section 6.3. We propose feasible sequential rules and establish their asymptotic optimality in Section 6.4. Section 6.5 extends our asymptotic results to a wide class of error rates. We end the chapter with a brief discussion in Section 6.6.

As we shall discuss several hypothesis testing problems (HTPs henceforth) in this chapter, we shall name each problem for notational convenience.

6.2 Preliminaries

6.2.1 The Testing Framework

We consider $K \geq 2$ data streams:

$$\begin{aligned}&X_{11}, X_{12}, \ldots, X_{1n}, \ldots\\&X_{21}, X_{22}, \ldots, X_{2n}, \ldots\\&\vdots\\&X_{K1}, X_{K2}, \ldots, X_{Kn}, \ldots.\end{aligned}$$

Here X_{ij} denotes the j'th observation of i'th data stream, $i \in [K] := \{1, \ldots, K\}$. We assume throughout this chapter that the elements of a given stream are i.i.d but not necessarily the streams are independent of each other. For each $i \in [K]$, we consider two simple hypotheses:

HTP 1. $H_{0i} : X_{ij} \sim N(0,1)$ for each $j \in \mathbb{N}$ $\quad vs \quad$ $H_{1i} : X_{ij} \sim N(\mu, 1)$ for each $j \in \mathbb{N}$

where $\mu > 0$.

We assume that for each $j \in \mathbb{N}$, $(X_{1j}, \ldots, X_{Kj})$ follows a multivariate normal distribution with variance covariance matrix $M_K(\rho)$ for some $\rho \geq 0$. We say that there is "noise" in ith stream if H_{0i} is true and there is "signal" in the ith stream otherwise.

For $n \in \mathbb{N}$, let $\mathcal{S}_n^i$ denote the σ-field generated by the first n observations of the i th data stream, i.e., $\sigma(X_{i1}, \ldots, X_{in})$. Let $\mathcal{S}_n$ be the σ-field generated by the first n observations in all streams, that is, $\sigma(\mathcal{S}_n^i, i \in [K])$. The data in all streams are observed sequentially. We wish to solve the K decision problems by terminating sampling as soon as possible, subject to controlling relevant error criteria.

We define a *sequential MTP* as a pair (T, d) where T is an $\{\mathcal{S}_n\}$-stopping time at which we stop sampling in each stream, and d an $\mathcal{S}_T$-measurable vector of Bernoulli random variables, $(d_1, \ldots, d_K)$. $d_i = 1$ corresponds to selecting H_{1i} over H_{0i}, and $d_i = 0$ means selecting H_{0i} over H_{1i}. So $\mathcal{D} := \{i \in [K] : d_i = 1\}$ is the collection of streams in which we reject the null. Suppose $\mathcal{A}$ is the true subset of indices for which the null hypothesis is true. For any sequential MTP (T, d), one has

$$\{(\mathcal{D} \backslash \mathcal{A}^c) \neq \emptyset\} = \bigcup_{j \notin \mathcal{A}^c} \{d_j = 1\}, \quad \text{and} \quad \{(\mathcal{A}^c \backslash \mathcal{D}) \neq \emptyset\} = \bigcup_{k \in \mathcal{A}^c} \{d_k = 0\}.$$

For any subset $\mathcal{B} \subset [K]$ let $\mathbb{P}_{\mathcal{B}}$ be defined as the distribution of $\{\mathbf{X}_n, n \in \mathbb{N}\}$ when $\mathcal{B}$ is the true subset of nulls. Thus, the two types of familywise error rates are given by

$$FWER_{I,\mathcal{A}}(T,d) = \mathbb{P}_{\mathcal{A}}((\mathcal{D}\backslash\mathcal{A}^c) \neq \emptyset) = \mathbb{P}_{\mathcal{A}}((T,d) \text{ makes at least one false rejection}),$$
$$FWER_{II,\mathcal{A}}(T,d) = \mathbb{P}_{\mathcal{A}}((\mathcal{A}^c\backslash\mathcal{D}) \neq \emptyset) = \mathbb{P}_{\mathcal{A}}((T,d) \text{ makes at least one false acceptance}).$$

We write $FWER_{i,\mathcal{A}}(T,d)$ as $FWER_i(T,d)$ for $i = I, II$. For pre-specified precision levels $\alpha \in (0,1)$ and $\beta \in (0,1)$, we are here concerned with sequential MTPs (T,d) satisfying

$$FWER_I(T,d) \leq \alpha \text{ and } FWER_{II}(T,d) \leq \beta \text{ for every } \mathcal{A}.$$

Note that incorporating a prior information about the true subset of signals is same as assuming that $\mathcal{A}$ belongs to a class $\mathcal{P}$ of subsets of $[K]$. We consider the class

$$\Delta_{\alpha,\beta}^{FWER}(\mathcal{P}) := \{(T,d) : FWER_I(T,d) \leq \alpha \text{ and } FWER_{II}(T,d) \leq \beta \quad \forall \mathcal{A} \in \mathcal{P}\}.$$

In this chapter, two classes $\mathcal{P}$ will be considered. In the first case, we know beforehand that $\mathcal{A}^c$ has exactly m members, where $1 \leq m \leq K-1$. In the second case, although the exact number of signals might not be known, strict lower and upper bounds for the same are available. These two cases respectively correspond to the classes

$$\mathcal{P}_m := \{\mathcal{A} \subset [K] : |\mathcal{A}^c| = m, 0 < m < K\}, \quad \mathcal{P}_{\ell,u} := \{\mathcal{A} \subset [K] : 0 < \ell < |\mathcal{A}^c| < u < K\}.$$

6.2.2 Asymptotic Optimality for controlling FWER

We are interested in finding sequential MTPs belonging to $\Delta_{\alpha,\beta}^{FWER}(\mathcal{P}_m)$ or $\Delta_{\alpha,\beta}^{FWER}(\mathcal{P}_{\ell,u})$ which are optimal in the natural sense of Wald's sequential framework, i.e., which achieve the least average sample number under each possible signal configuration.

Definition 3. (Song and Fellouris, 2017; Song, 2019) Let $\mathcal{P}$ be a given class of subsets and let (T^*,d^*) be a sequential MTP belonging to $\Delta_{\alpha,\beta}^{FWER}(\mathcal{P})$ for any given $\alpha,\beta \in (0,1)$. (T^*,d^*) is called *asymptotically optimal in the class $\mathcal{P}$ for controlling FWER*, if for every $\mathcal{A} \in \mathcal{P}$ we have,

$$\lim_{\alpha,\beta\to 0} \frac{\mathbb{E}_{\mathcal{A}}[T^*]}{\inf\limits_{(T,d)\in\Delta_{\alpha,\beta}^{FWER}(\mathcal{P})} \mathbb{E}_{\mathcal{A}}[T]} = 1$$

where $\mathbb{E}_{\mathcal{A}}$ denotes the expectation under $\mathbb{P}_{\mathcal{A}}$.

We shall often write $x \sim y$ to mean $x/y \to 1$. In this chapter, we shall discuss sequential MTPs that are asymptotically optimal in the classes $\mathcal{P}_m$ and $\mathcal{P}_{\ell,u}$.

6.2.3 Existing Results under Independent Setup

Song and Fellouris (2017) consider, for each $i \in [K]$, the following simple vs simple testing problem:

$$\textbf{HTP 2.}\ H_{0i} : X_{ij} \sim \mathbb{P}_{0i} \text{ for each } j \in \mathbb{N} \quad vs \quad H_{1i} : X_{ij} \sim \mathbb{P}_{1i} \text{ for each } j \in \mathbb{N},$$

where P_{0i} and P_{1i} are distinct probability measures. Note that HTP 1 can be thought of as a special case of HTP 2. They suppose that, for each stream $i \in [K]$, the observations $\{X_{ij}, j \in \mathbb{N}\}$ follow, independently of each other, common density f_{0i} and f_{1i} w.r.t a σ-finite measure μ_i under $\mathbb{P}_{0i}$ and $\mathbb{P}_{1i}$ respectively. They consider Kullback Leibler information numbers

$$D_{0i} := \int \log\left(\frac{f_{0i}}{f_{1i}}\right) f_{0i} d\mu_i, \quad D_{1i} := \int \log\left(\frac{f_{1i}}{f_{0i}}\right) f_{1i} d\mu_i.$$

For each subset $\mathcal{B} \subset [K]$, let

$$\eta_1^{\mathcal{B}} := \min_{i \in \mathcal{B}} D_{1i}, \quad \eta_0^{\mathcal{B}} := \min_{i \notin \mathcal{B}} D_{0i}.$$

Let $\lambda_i(n)$ be the cumulative log-likelihood ratio corresponding to the first n observations in i'th data stream. These, when ordered, are denoted as

$$\lambda_{(1)}(n) \geq \cdots \geq \lambda_{(K)}(n).$$

Following Song and Fellouris (2017), we denote the corresponding stream indices by $i_1(n), ..., i_K(n)$, i.e.,

$$\lambda_{(k)}(n) = \lambda_{i_k(n)}(n), \quad \forall k \in [K].$$

We denote cardinality by $|\cdot|$ and, for $x, y \in \mathbb{R}$, we write $x \wedge y = \min\{x, y\}$ and $x \vee y = \max\{x, y\}$.

6.2.3.1 Prior information on number of signals

Suppose that it is known that the cardinality of $\mathcal{A}^c$ is exactly m, $1 \leq m \leq K-1$. So, $\mathcal{P} = \mathcal{P}_m$. We observe that for any $\mathcal{A} \in \mathcal{P}_m$ and (T, d) such that $|\mathcal{D}| = m$, we have

$$FWER_I(T, d) = FWER_{II}(T, d) = \mathbb{P}_{\mathcal{A}}\{\mathcal{D} \neq \mathcal{A}^c\}.$$

$\mathbb{P}_{\mathcal{A}}\{\mathcal{D} \neq \mathcal{A}^c\}$ is simply the probability that the sequential test rule (T, d) commits an incorrect selection, and therefore will alternatively be denoted as $PICS(T, d)$. Hence, the class of feasible sequential tests is given by

$$\Delta_{\alpha,\beta}^{FWER}(\mathcal{P}_m) = \{(T,d) : PICS_{\mathcal{A}}(T,d) \leq \alpha \wedge \beta \text{ for every } \mathcal{A} \in \mathcal{P}_m\}.$$

Song and Fellouris (2017) propose the following sequential decision rule (which is commonly referred to as the *gap rule*) in this setup. The stopping time of their rule is

$$T_{Gap}(m,c) := \inf\left\{n \geq 1 : \lambda_{(m)}(n) - \lambda_{(m+1)}(n) \geq c\right\}.$$

The procedure rejects the nulls having the highest m log-likelihood ratios at time $T_{Gap}(m,c)$. We write $T_{Gap}(m,c)$ as T_{Gap} for simpler notation. Song and Fellouris (2017) establish the following:

Theorem 6.2.1. *Suppose $\mathcal{A} \in \mathcal{P}_m$. For each threshold $c > 0$, we have $\mathbb{P}_{\mathcal{A}}(T_{Gap} < \infty) = 1$ and*

$$PICS_{\mathcal{A}}(T_{Gap}, d_{Gap}) \leq m(K-m)e^{-c}.$$

Consequently, $(T_{Gap}, d_{Gap}) \in \Delta_{\alpha,\beta}^{FWER}(\mathcal{P}_m)$ when

$$c = |\log(\alpha \wedge \beta)| + \log(m(K-m)). \tag{6.1}$$

Theorem 6.2.2. *Suppose $\mathcal{A}^c \in \mathcal{P}_m$ and the threshold c in the gap rule is chosen as in (6.1). Then, we have as $\alpha, \beta \to 0$*

$$\mathbb{E}_{\mathcal{A}}[T_{Gap}] \sim \frac{|\log(\alpha \wedge \beta)|}{\eta_1^{\mathcal{A}^c} + \eta_0^{\mathcal{A}^c}} \sim \inf_{(T,d)\in\Delta_{\alpha,\beta}^{FWER}(\mathcal{P}_m)} \mathbb{E}_{\mathcal{A}}[T].$$

Song and Fellouris (2017) also derived asymptotically optimal sequential tests when it is known that $\ell \leq |\mathcal{A}^c| \leq u$ for some $0 \leq \ell < u \leq K$.

6.2.3.2 General Error Rate controlling Procedures

He and Bartroff (2021) showed that the procedures proposed by Song and Fellouris (2017), with suitably modified cut-offs, are asymptotically optimal for controlling any multiple testing error criterion that lies between multiples of FWER. We shall establish similar optimality properties of our proposed procedures in section 6.5.

6.3 Asymptotic Expansion of Expected Sample Size of SPRT

This chapter aims to propose feasible sequential test procedures that require minimum expected sample size in the classes $\mathcal{P}_m$ and $\mathcal{P}_{\ell,u}$ asymptotically as $\alpha, \beta \longrightarrow 0$. We shall establish this asymptotic optimality by comparing the ratios of the expected sample sizes of our proposed procedure and Wald's sequential probability ratio test (SPRT). Therefore, we at first focus on the expected sample size (or average sample number) of SPRT.

Consider the following hypothesis testing problem:

HTP 3. $H_0 : U_i \sim N(\theta_0, \sigma^2) \quad vs \quad H_1 : U_i \sim N(\theta_1, \sigma^2), \quad \theta_0 < \theta_1.$

Let γ and δ denote the desired levels of type I and type II error probabilities, respectively. At the n-th stage, the SPRT with strength (γ, δ) for this testing problem is as follows:

(a) accept H_0, if

$$\sum_{i=1}^{n} U_i - \frac{n\,(\theta_1 + \theta_0)}{2} \leqslant \frac{b\sigma^2}{\theta_1 - \theta_0}.$$

(b) reject H_0, if

$$\sum_{i=1}^{n} U_i - \frac{n\,(\theta_1 + \theta_0)}{2} \geqslant \frac{a\sigma^2}{\theta_1 - \theta_0}.$$

(c) continue sampling otherwise,

where $a = \log[(1-\delta)/\gamma]$ and $b = \log[\delta/(1-\gamma)]$.

The following result depicts the optimality of SPRT among all fixed-sample-size or sequential tests:

Theorem 6.3.1. *(see, e.g., Theorem 3.3.1 of Mukhopadhyay and Silva (2019)) Consider the class of all fixed-sample-size or sequential tests for which the type I and type II error probabilities are less than γ and δ respectively, and for which $\mathbb{E}(T)$ (T denotes the stopping time) is finite both under the null and under the alternative. The classical SPRT with error probabilities γ and δ minimizes $\mathbb{E}(T)$ both under the null and under the alternative in this class when $\gamma + \delta < 1$.*

We denote the SPRT for the above hypothesis testing problem as $SPRT_{(\gamma,\delta)}(\theta_0, \theta_1, \sigma^2)$.

Theorem 6.3.2. *(see, e.g., Rao (1973) or Mukhopadhyay and Silva (2019)) The approximate expressions for the ASN of $SPRT_{(\gamma,\delta)}(\theta_0, \theta_1, \sigma^2)$ under H_0 and H_1 are given* *by*

$$\mathbb{E}_{H_0}(T_{SPRT}) \sim \frac{(1-\gamma)\log\left(\frac{\delta}{1-\gamma}\right) + \gamma\log\left(\frac{1-\delta}{\gamma}\right)}{\frac{-(\theta_1-\theta_0)^2}{2\sigma^2}},$$

$$\mathbb{E}_{H_1}(T_{SPRT}) \sim \frac{\delta\log\left(\frac{\delta}{1-\gamma}\right) + (1-\delta)\log\left(\frac{1-\delta}{\gamma}\right)}{\frac{(\theta_1-\theta_0)^2}{2\sigma^2}},$$

respectively. Here $x \sim y$ implies $x/y \to 1$ as $\gamma, \delta \to 0$.

Let, $L = \frac{(\theta_1-\theta_0)^2}{2\sigma^2}$. Then,

$$\begin{aligned} 0 &< \frac{\mathbb{E}_{H_0}(T_{SPRT})}{-\log(\gamma)} = \frac{1}{L}\cdot\left[(1-\gamma)\frac{\log(\delta)}{\log(\gamma)} - \gamma\right] \leq \frac{1}{L}\cdot(1-\gamma) \\ \implies 0 &\leq \lim_{\gamma\to 0}\frac{\mathbb{E}_{H_0}(T_{SPRT})}{-\log(\gamma)} \leq \frac{1}{L}. \end{aligned}$$

Without loss of generality, we may assume $\gamma \leq \delta$. This gives,

$$\begin{aligned} \frac{1}{L}\cdot(1-2\delta) &\leq \frac{\mathbb{E}_{H_1}(T_{SPRT})}{-\log(\gamma)} \leq \frac{1}{L}\cdot(1-\delta). \\ \implies \lim_{\delta\to 0}\frac{\mathbb{E}_{H_1}(T_{SPRT})}{-\log(\gamma)} &= \frac{1}{L}. \end{aligned}$$

Therefore, we obtain

$$\begin{aligned} \lim_{\gamma,\delta\to 0}\frac{\max\{\mathbb{E}_{H_0}(T_{SPRT}), \mathbb{E}_{H_1}(T_{SPRT})\}}{-\log(\gamma)} &= \frac{1}{L} \\ \implies \lim_{\gamma,\delta\to 0}\frac{\mathbb{E}(T_{SPRT})}{-\log(\gamma)} = \frac{1}{L} \end{aligned}$$

where $\mathbb{E}(T_{SPRT}) = \max\{\mathbb{E}_{H_0}(T_{SPRT}), \mathbb{E}_{H_1}(T_{SPRT})\}$ is the ASN of $SPRT_{\gamma,\delta}(\theta_0, \theta_1, \sigma^2)$. Thus, we establish the following:

Theorem 6.3.3. *The approximate expression for the ASN of $SPRT_{(\gamma,\delta)}(\theta_0, \theta_1, \sigma^2)$ when $\gamma, \delta \to 0$ is*

$$\mathbb{E}(T_{SPRT}) \sim \frac{2\sigma^2}{(\theta_1-\theta_0)^2}\cdot|\log(\gamma\wedge\delta)|.$$

Remark 10. *Theorem 6.3.1 concerns both-sided SPRTs. However in many scenarios, one faces the hypothesis testing problem*

$$\textbf{\textit{HTP 4.}}\ H_0 : U_i \sim N(\theta_0, \sigma^2) \quad vs \quad H_1 : U_i \sim N(\theta_1, \sigma^2), \quad \theta_0 < \theta_1,$$

where sampling should be terminated as soon as possible if there is enough evidence against H_0 and in favor of H_1. A feasible solution to this sequential hypothesis testing problem is *the one-sided SPRT:*

$$T_{SPRT} = \inf\{n \geq 1 : \lambda(n) \geq A\}, \quad \inf\{\phi\} = \infty,$$

where $A > 1$ is a fixed threshold and $\{\lambda(n)\}$ is the corresponding likelihood-ratio process. In these problems, $\delta = 0$. Along the same lines of the preceding proof, we obtain:

Theorem 6.3.4. *The approximate expression for the ASN of $SPRT_{(\gamma,0)}(\theta_0, \theta_1, \sigma^2)$ when $\gamma \to 0$ is*

$$\mathbb{E}(T_{SPRT}) \sim \frac{2\sigma^2}{(\theta_1 - \theta_0)^2} \cdot |\log(\gamma)|.$$

6.4 Main Results for FWER Controlling Tests

6.4.1 A distributionally equivalent representation of the vector observations

Let $\mathbf{X}_n = (X_{1n}, \ldots, X_{kn})$ denote the K-dimensional vector storing all the observations collected at time n. Suppose $\mu = (\mu_1, \ldots, \mu_K)'$ where

$$\mu_i = \begin{cases} 0, & \text{if } i \in \mathcal{A} \\ \mu, & \text{if } i \notin \mathcal{A} \end{cases}.$$

Then, for each $n \geq 1$, we have

$$\mathbf{X}_n \sim MVN_K(\mu, M_K(\rho)).$$

This implies,

$$\mathbf{X}_n \stackrel{d}{=} \mathbf{Z}_n + V_n \cdot \mathbf{1}_K$$

where $\mathbf{Z}_n = (Z_{1n}, \ldots, Z_{kn}) \sim MVN_K(\mu, (1-\rho) \cdot I_K)$ and $V_n \sim N(0, \rho)$ are independent. Here $\mathbf{1}_K$ denotes the K dimensional vector of all ones. Thus, for each $i \in [K]$ and for each $j \geq 1$,

$$X_{ij} \stackrel{d}{=} Z_{ij} + V_j.$$

This gives,

$$\sum_{j=1}^{n} X_{ij} \stackrel{d}{=} \sum_{j=1}^{n} Z_{ij} + \sum_{j=1}^{n} V_j$$

$$\Longrightarrow S_{i,n} \stackrel{d}{=} R_{i,n} + \sum_{j=1}^{n} V_j \quad \text{(say)}$$

$$\implies S_{i,n} - S_{i',n} \stackrel{d}{=} R_{i,n} - R_{i',n}.$$

We write $S_{i,n}$ and $R_{i,n}$ as S_i and R_i respectively for simpler notation. We also note that

$$S_{(m)} - S_{(m+1)} \stackrel{d}{=} R_{(m)} - R_{(m+1)}$$

where $S_{(1)} \geq \cdots \geq S_{(K)}$ and $R_{(1)} \geq \cdots \geq R_{(K)}$. This implies, although we can not directly observe Z_{ij}'s or R_i's, we can observe the quantities $R_i - R_{i'}$ and $R_{(m)} - R_{(m+1)}$. The distributions of V_j or $\sum_j V_j$ do not depend on μ. Therefore, inference on μ built on R_i would disseminate the same amount of information as the inference built on S_i would have.

6.4.2 Log-likelihood ratio statistics

Consider the i'th hypothesis test for the Z_{ij}'s:

HTP 5. $H^\star_{0i} : Z_{ij} \sim N(0, 1-\rho)$ for each $j \in \mathbb{N}$ $\quad vs \quad$ $H^\star_{1i} : Z_{ij} \sim N(\mu, 1-\rho)$ for each $j \in \mathbb{N}$

where $\mu > 0$. The log-likelihood ratio statistic for this test is

$$\lambda^\star_i(n) = \frac{\mu}{1-\rho}\Big[R_i - \frac{n\mu}{2}\Big].$$

This gives, for $i \neq i' \in [K]$, $\lambda^\star_i(n) - \lambda^\star_{i'}(n) = \frac{\mu}{1-\rho}\Big[R_i - R_{i'}\Big]$. Thus,

$$\lambda^\star_{(m)}(n) - \lambda^\star_{(m+1)}(n) = \frac{\mu}{1-\rho}\Big[R_{(m)} - R_{(m+1)}\Big].$$

However, our original hypotheses tests are based on X_{ij}'s and not on Z_{ij}'s. Also, Z_{ij}'s are unobserved variables. Yet, from the previous subsection, we obtain

$$\frac{\mu}{1-\rho}\Big[S_{(m)} - S_{(m+1)}\Big] \stackrel{d}{=} \lambda^\star_{(m)}(n) - \lambda^\star_{(m+1)}(n).$$

6.4.3 Proposed procedure for known number of signals

Mimicking the gap rule proposed by Song and Fellouris (2017), we propose the following stopping time:

$$T^\star_{Gap}(m, c) := \inf\left\{n \geq 1 : S_{(m)} - S_{(m+1)} \geq \frac{1-\rho}{\mu} \cdot c\right\}.$$

where c is defined as in (6.1). Let $G = (1-\rho)\cdot c/\mu$ be the r.h.s of the above inequality. Evidently, an incorrect selection happens at the stopping time (say, n) if there is some $j \in \mathcal{A}$ and some $i \in \mathcal{A}^c$ such that $R_j - R_i \geq G$. This is because in this case the gap rule declares $i \in \mathcal{A}$ and $j \in \mathcal{A}^c$.

The above discussion elucidates that the probability of incorrect selection (which is same as $FWER_I$ and $FWER_{II}$ in this setup with known m) of our gap rule is given by

$$PICS(T^\star_{Gap}, d^\star_{Gap}) = \mathbb{P}_{\mathcal{A}}\left[\bigcup_{j\in\mathcal{A}, i\in\mathcal{A}^c}\{R_j - R_i \geq G\}\right] \quad (|\mathcal{A}^c| = m). \tag{6.2}$$

We write $PICS(T^\star_{Gap}, d^\star_{Gap})$ as $PICS(T^\star_{Gap})$ for simpler notation.

Consequently we obtain that the proposed gap rule makes an error if it declares $R_j - R_i \sim N(n\mu, 2(1-\rho)n)$ when actually $R_j - R_i \sim N(-n\mu, 2(1-\rho)n)$. This observation motivates us to study the following classical one-vs-one hypothesis testing problem in a sequential framework:

HTP 6. $H_0 : T \sim N(-\mu, 2(1-\rho)) \quad vs \quad H_1 : T \sim N(\mu, 2(1-\rho)), \quad \mu > 0.$

Here we wish to terminate sampling as soon as possible if there is sufficient evidence against H_0 and in favor of H_1. The desired level of type I error is given by $\frac{\alpha\wedge\beta}{m(K-m)}$. The optimal test for this problem is $SPRT_{\frac{\alpha\wedge\beta}{m(K-m)},0}(-\mu, \mu, 2(1-\rho))$ where α and β are the target precision levels of type I and II error respectively. Theorem 6.3.4 implies

$$\mathbb{E}\big[T_{SPRT_{\frac{\alpha\wedge\beta}{m(K-m)},0}(-\mu,\mu,2(1-\rho))}\big] \sim \frac{1-\rho}{\mu^2}\cdot|\log(\alpha\wedge\beta)|. \tag{6.3}$$

In other words,

$$\log\big(PICS(SPRT)\big) = -\frac{\mu^2}{1-\rho}\cdot\mathbb{E}(T_{SPRT}) + o(\mathbb{E}(T_{SPRT})), \tag{6.4}$$

We observe that $SPRT_{\frac{\alpha\wedge\beta}{m(K-m)},0}(-\mu, \mu, 2(1-\rho)$ also uses G as the critical point. We have the following inequalities from (6.2)

$$PICS(SPRT_{(\frac{\alpha\wedge\beta}{m(K-m)},0)}) \leq PICS(T^\star_{Gap}) \leq m(K-m)PICS(SPRT_{(\frac{\alpha\wedge\beta}{m(K-m)},0)}). \tag{6.5}$$

Since $m(K-m) < \infty$, this gives

$$\log\big(PICS(T^\star_{Gap})\big) = -\frac{\mu^2}{1-\rho}\cdot\mathbb{E}(T_{SPRT}) + o(\mathbb{E}(T_{SPRT})). \tag{6.6}$$

This establishes the following theorem:

Theorem 6.4.1. *Suppose $\mathcal{A} \in \mathcal{P}_m$. We have as $\alpha, \beta \to 0$*

$$\mathbb{E}_{\mathcal{A}}\left[T^{\star}_{Gap}\right] \sim \frac{1-\rho}{\mu^2} \cdot |\log(\alpha \wedge \beta)| \sim \inf_{(T,d)\in\Delta^{FWER}_{\alpha,\beta}(\mathcal{P}_m)} \mathbb{E}_{\mathcal{A}}[T].$$

6.4.4 Proposed procedure when upper and lower bounds on the number of signals are available

Consider now the setup it is known beforehand that $\ell < |\mathcal{A}^c| < u$ for some $0 < l \leq u < K$. This corresponds to considering the class $\mathcal{P}_{\ell,u}$. We propose the gap rule

$$T^{\star\star}_{Gap}(l,u,e) := \inf\left\{n \geq 1 : \max_{\ell<i<u}\left(S_{(i)} - S_{(i+1)}\right) \geq e\right\}$$

where e is suitably defined. Let p be the index where the above maximum occurs at time $T^{\star\star}_{Gap}(l,u,e)$. The set of rejected nulls is given by

$$d^{\star\star}_{Gap} := \left\{i_1(T^{\star\star}_{Gap}), \ldots, i_p(T^{\star\star}_{Gap})\right\}.$$

Now,

$$\begin{aligned}
FWER_I(T^{\star\star}_{Gap}, d^{\star\star}_{Gap}) =& \mathbb{P}_{\mathcal{A}}\left(\bigcup_{i\in\mathcal{A}}\{d^{\star\star}_{Gap,i} = 1\}\right) \\
\leq& \sum_{i\in\mathcal{A}} \mathbb{P}_{\mathcal{A}}\left(d^{\star\star}_{Gap,i} = 1\right) \\
\leq& (K-l)\mathbb{P}_{\mathcal{A}}\left(d^{\star\star}_{Gap,i} = 1\right) \quad \text{(for any } i \in \mathcal{A}) \\
\leq& 2(K-l)\mathbb{P}_{\mathcal{A}}\left(\bigcup_{j\in\mathcal{A}, j\neq i}\{S_i - S_j \geq e\}\right) \\
\leq& 2(K-l)(K-l-1)\mathbb{P}_{\mathcal{A}}(S_i - S_j \geq e) \quad \text{(for } j \in \mathcal{A}, j \neq i) \\
=& 2(K-l)(K-l-1)\Phi\left(\frac{-e}{\sqrt{2(1-\rho)n}}\right) \\
=& 2(K-l)(K-l-1)\mathbb{P}_{\mathcal{A}}\left(S_i \geq \frac{e}{\sqrt{2}}\right).
\end{aligned}$$

Hence, the chosen e should satisfy

$$\begin{aligned}
2(K-l)(K-l-1)\mathbb{P}_{\mathcal{A}}\left(S_i \geq \frac{e}{\sqrt{2}}\right) &\leq \alpha \\
\implies \mathbb{P}_{\mathcal{A}}\left(S_i \geq \frac{e}{\sqrt{2}}\right) &\leq \frac{\alpha}{2(K-l)(K-l-1)}.
\end{aligned}$$

Thus, we can choose e to be

$$\frac{e}{\sqrt{2}} = \frac{1-\rho}{\mu} \cdot |\log\left(\frac{\alpha}{2(K-l)(K-l-1)}\right)| + n\mu/2.$$

This is because this is the cutoff used by $SPRT_{(\frac{\alpha}{2(K-l)(K-l-1)},0)}(0,\mu,1-\rho)$.

One can also show the following exactly as in the same way as above:

$$\begin{aligned} FWER_{II}(T^{\star\star}_{Gap}, d^{\star\star}_{Gap}) &\leq 2u(u-1)\mathbb{P}_{\mathcal{A}}(S_i - S_j \geq e) \quad (\text{for } i,j \notin \mathcal{A}, j \neq i) \\ &= 2u(u-1)\mathbb{P}_{\mathcal{A}}\left(S_i \geq \frac{e}{\sqrt{2}}\right). \end{aligned}$$

Hence, the chosen e should satisfy

$$\mathbb{P}_{\mathcal{A}}\left(S_i \geq \frac{e}{\sqrt{2}}\right) \leq \frac{\beta}{2u(u-1)}.$$

This leads to the choice

$$\frac{e}{\sqrt{2}} = \frac{1-\rho}{\mu} \cdot |\log\left(\frac{\beta}{2u(u-1)}\right)| + n\mu/2.$$

Therefore, we choose the following e:

$$e = \frac{1-\rho}{\mu} \cdot \max\left\{|\log\left(\frac{\alpha}{2(K-l)(K-l-1)}\right)|, |\log\left(\frac{\beta}{2u(u-1)}\right)|\right\} + n\mu/2.$$

The previous derivations result in the following:

$$\begin{aligned} FWER_I(T^{\star\star}_{Gap}, d^{\star\star}_{Gap}) &\leq 2(K-l)(K-l-1)PICS\Big[SPRT_{\left(\frac{\alpha}{2(K-l)(K-l-1)},0\right)}(0,\mu,(1-\rho)\Big], \\ FWER_{II}(T^{\star\star}_{Gap}, d^{\star\star}_{Gap}) &\leq 2u(u-1)PICS\Big[SPRT_{\left(\frac{\beta}{2u(u-1)},0\right)}(0,\mu,(1-\rho)\Big]. \end{aligned}$$

However, those derivations also give the following inequalities:

$$\begin{aligned} FWER_I(T^{\star\star}_{Gap}, d^{\star\star}_{Gap}) &\geq PICS\Big[SPRT_{\left(\frac{\alpha}{2(K-l)(K-l-1)},0\right)}(0,\mu,(1-\rho)\Big], \\ FWER_{II}(T^{\star\star}_{Gap}, d^{\star\star}_{Gap}) &\geq PICS\Big[SPRT_{\left(\frac{\beta}{2u(u-1)},0\right)}(0,\mu,(1-\rho)\Big]. \end{aligned}$$

Combining all these four inequalities, we obtain,

$$\mathbb{E}_{\mathcal{A}}(T^{\star\star}_{Gap}) = \max\left\{\mathbb{E}\left(T_{SPRT_{\left(\frac{\alpha}{2(K-l)(K-l-1)},0\right)}(0,\mu,(1-\rho)}\right), \mathbb{E}\left(T_{SPRT_{\left(\frac{\beta}{2u(u-1)},0\right)}(0,\mu,(1-\rho)}\right)\right\}.$$

This gives, as $\alpha, \beta \to 0$, and when $\log\alpha \sim \log\beta$,

$$\mathbb{E}_{\mathcal{A}}\left[T^{\star\star}_{Gap}\right] \sim \frac{2(1-\rho)}{\mu^2} \cdot |\log(\alpha \wedge \beta)|.$$

Theorem 6.4.2. *Suppose $\mathcal{A} \in \mathcal{P}_{\ell,u}$. We have as $\alpha, \beta \to 0$ and when $\log\alpha \sim \log\beta$,*

$$\mathbb{E}_{\mathcal{A}}\left[T^{\star\star}_{Gap}\right] \sim \frac{2(1-\rho)}{\mu^2} \cdot |\log(\alpha \wedge \beta)| \sim \inf_{(T,d)\in\Delta^{FWER}_{\alpha,\beta}(\mathcal{P}_{\ell,u})} \mathbb{E}_{\mathcal{A}}[T].$$

6.5 General Error Rate controlling Procedures

As mentioned earlier, He and Bartroff (2021) studied asymptotic optimality of general multiple testing error metrics under the independent streams framework. In this section, we focus on deriving similar results under dependence. Suppose

$$\mathrm{MTE} = (\mathrm{MTE}_1, \mathrm{MTE}_2)$$

denotes a generic multiple testing error metric. In other words, MTE is a pair of functions from the set of MTPs onto $[0,1]$. The type I and type II familywise error probabilities are given by

$$FWER_{1,\mathcal{A}} = \mathbb{P}_{\mathcal{A}}(V \geq 1), \quad FWER_{2,\mathcal{A}} = \mathbb{P}_{\mathcal{A}}(W \geq 1)$$

where V and W respectively denote the number of type I errors and type II errors respectively. For a generic metric MTE, let

$$\Delta^{\mathrm{MTE}}_{\alpha,\beta}(\mathcal{P}) = \left\{(T,d) : \mathrm{MTE}_{1,\mathcal{A}}(T,d) \leq \alpha \text{ and } \mathrm{MTE}_{2,\mathcal{A}}(T,d) \leq \beta \text{ for all } \mathcal{A}^c \in \mathcal{P}\right\}.$$

We now mention a result on asymptotic optimality of general error metric controlling procedures.

Theorem 6.5.1. *Consider the equicorrelated streams setup with common correlation $\rho > 0$. Fix $1 \leq m \leq K-1$ and let $\left(T^{\star}_{Gap}(c), d^{\star}_{Gap}(c)\right)$ denote our gap rule with number of signals m and threshold $c > 0$. Suppose MTE is a multiple testing error metric satisfying:*

(i) there is a constant C_1 such that

$$\mathrm{MTE}_{i,\mathcal{A}}\left(T^{\star}_{Gap}(c), d^{\star}_{Gap}(c)\right) \leq C_1 \cdot FWER_{i,\mathcal{A}}\left(T^{\star}_{Gap}(c), d^{\star}_{Gap}(c)\right) \tag{6.7}$$

for $i = 1$ and 2 , for all $\mathcal{A} \in \mathcal{P}_m$, and for all $c > 0$, and

(ii) there is a constant C_2 such that

$$\mathrm{MTE}_{i,\mathcal{A}}(T,d) \geq C_2 \cdot FWER_{i,\mathcal{A}}(T,d) \tag{6.8}$$

for $i = 1$ and 2 , for all $\mathcal{A} \in \mathcal{P}_m$, and for all procedures (T, d).

Given $\alpha, \beta \in (0,1)$, let $\left(T'_{Gap}, d'_{Gap}\right)$ be our proposed gap rule with number of signals m and threshold

$$c = |\log\left((\alpha/C_1) \wedge (\beta/C_1)\right)| + \log(m(K-m)).$$

Then we have the following.

(1) The procedure $\left(T'_{Gap}, d'_{Gap}\right)$ is admissible for MTE control. That is,

$$\left(T'_{Gap}, d'_{Gap}\right) \in \Delta^{\text{MTE}}_{\alpha,\beta}(\mathcal{P}_m). \tag{6.9}$$

(2) The MTP $\left(T'_{Gap}, d'_{Gap}\right)$ is asymptotically optimal for MTE control in class $\mathcal{P}_m$. In other words, for all $\mathcal{A} \in \mathcal{P}_m$, as $\alpha, \beta \to 0$,

$$\mathbb{E}_{\mathcal{A}}\left(T'_{Gap}\right) \sim \frac{1-\rho}{\mu^2} \cdot |\log(\alpha \wedge \beta)| \sim \inf_{(T,d) \in \Delta^{\text{MTE}}_{\alpha,\beta}(\mathcal{P}_m)} \mathbb{E}_{\mathcal{A}}(T). \tag{6.10}$$

Proof. For the first part, we fix arbitrary $\mathcal{A} \in \mathcal{P}_m$. We have

$$\text{FWER}_{1,\mathcal{A}}\left(T'_{Gap}, d'_{Gap}\right) \leq \alpha/C_1 \text{ and } \text{FWER}_{2,\mathcal{A}}\left(T'_{Gap}, d'_{Gap}\right) \leq \beta/C_1.$$

Applying (6.7) yields

$$\text{MTE}_{1,\mathcal{A}}\left(T'_{Gap}, d'_{Gap}\right) \leq \alpha \quad \text{and} \quad \text{MTE}_{2,\mathcal{A}}\left(T'_{Gap}, d'_{Gap}\right) \leq \beta.$$

Hence (6.9) is established. For the second part, again we fix arbitrary $\mathcal{A} \in \mathcal{P}_m$. We have, as $\alpha, \beta \to 0$,

$$\mathbb{E}_{\mathcal{A}}\left(T'_{Gap}\right) \sim \frac{1-\rho}{\mu^2} \cdot |\log(\alpha \wedge \beta)|. \tag{6.11}$$

It remains to show that the r.h.s in the above equation is also a lower bound for any MTP in $\Delta^{\text{MTE}}_{\alpha,\beta}(\mathcal{P}_m)$. (6.8) gives

$$\Delta^{\text{MTE}}_{\alpha,\beta}(\mathcal{P}_m) \subseteq \Delta^{\text{FWER}}_{\alpha/C_2,\beta/C_2}(\mathcal{P}_m).$$

This implies

$$\inf_{(T,d) \in \Delta^{\text{MTE}}_{\alpha,\beta}(\mathcal{P}_m)} \mathbb{E}_{\mathcal{A}}(T) \geq \inf_{(T,d) \in \Delta^{\text{FWER}}_{\alpha/C_2,\beta/C_2}(\mathcal{P}_m)} \mathbb{E}_{\mathcal{A}}(T).$$

The latter is of the order $\frac{1-\rho}{\mu^2} \cdot |\log(\alpha \wedge \beta)|$. This, combined with (6.11) gives the desired result. □

Remark 11. *(FDR, FNR) and $(pFDR, pFNR)$ satisfy the conditions mentioned in*

Theorem 6.5.1. This is because of the following inequalities (He and Bartroff, 2021):

$$\frac{1}{K} \cdot FWER_1 \leq FDR \leq FWER_1, \quad \frac{1}{K} \cdot FWER_2 \leq FNR \leq FWER_2,$$

$$\frac{1}{K} \cdot FWER_1 \leq pFDR, \quad \frac{1}{K} \cdot FWER_2 \leq pFNR,$$
$$pFDR(T^{\star}_{Gap}) \leq FWER_1(T^{\star}_{Gap}), \quad pFNR(T^{\star}_{Gap}) \leq FWER_2(T^{\star}_{Gap}).$$

We also have the following similar asymptotic optimality result for general error metrics in the case when lower and upper bounds for the number of signals is available:

Theorem 6.5.2. *Consider the equicorrelated streams setup with common correlation $\rho > 0$. Fix integers $0 < \ell < u < K$ and let $T^{\star\star}_{Gap}(e)$ denote the gap-intersection rule with strict bounds ℓ, u on the number of signals and threshold e. Let MTE be a multiple testing error metric satisfying:*

(i) there exists $C_1 \in \mathbb{R}$ for which

$$\mathrm{MTE}_{i,\mathcal{A}}\left(T^{\star\star}_{Gap}(e), d^{\star\star}_{Gap}(e)\right) \leq C_1 \cdot FWER_{i,\mathcal{A}}\left(T^{\star}_{Gap}(e), d^{\star\star}_{Gap}(e)\right) \tag{6.12}$$

for $i = 1$ and 2 , for all $\mathcal{A} \in \mathcal{P}_{\ell,u}$, and for all $e > 0$, and

(ii) there exists $C_2 \in \mathbb{R}$ for which

$$\mathrm{MTE}_{i,\mathcal{A}}(T, d) \geq C_2 \cdot FWER_{i,\mathcal{A}}(T, d) \tag{6.13}$$

for $i = 1$ and 2 , for all $\mathcal{A} \in \mathcal{P}_{\ell,u}$, and for all MTPs (T, d).

Given $\alpha, \beta \in (0, 1)$, let $\left(T''_{Gap}, d''_{Gap}\right)$ be our proposed gap rule with bounds ℓ, u on the number of signals and threshold

$$e = \frac{1-\rho}{\mu} \cdot \max\left\{|\log\left(\frac{\alpha/C_1}{2(K-\ell)(K-\ell-1)}\right)|, |\log\left(\frac{\beta/C_1}{2u(u-1)}\right)|\right\} + n\mu/2.$$

Then we have the following.

(1) The procedure $\left(T''_{Gap}, d''_{Gap}\right)$ is admissible for MTE control. That is,

$$\left(T''_{Gap}, d''_{Gap}\right) \in \Delta^{\mathrm{MTE}}_{\alpha,\beta}(\mathcal{P}_{\ell,u}). \tag{6.14}$$

(2) The procedure $\left(T''_{Gap}, d''_{Gap}\right)$ is asymptotically optimal for MTE control in class $\mathcal{P}_{\ell,u}$.

In other words, for all $\mathcal{A} \in \mathcal{P}_{\ell,u}$,

$$\mathbb{E}_{\mathcal{A}}\left(T''_{Gap}\right) \sim \frac{2(1-\rho)}{\mu^2} \cdot |\log(\alpha \wedge \beta)| \sim \inf_{(T,d)\in\Delta^{\mathrm{MTE}}_{\alpha,\beta}(\mathcal{P}_{\ell,u})} \mathbb{E}_{\mathcal{A}}(T) \tag{6.15}$$

as $\alpha, \beta \to 0$.

The proof is identical to the preceding and hence omitted.

6.6 Concluding Remarks

Our results (e.g., Theorem 6.4.1 and Theorem 6.4.2) elucidate that the asymptotically optimal average sample numbers are decreasing in the common correlation ρ. Large values of ρ indicate that the streams are more positively correlated to each other and hence one might expect that it should require less number of samples on average to detect the signals. Our results illustrate this remarkable blessing of dependence. This result is in contrast to the fixed sample size paradigm, as we have seen in the earlier chapters that several popular and widely used procedures fail to hold the FWER at a positive level asymptotically under positively correlated Gaussian frameworks. Thus, correlation plays a dual role in the classical fixed-sample size and the sequential paradigms.

Finner et al. (2007) remark that false discoveries are challenging to tackle in models with complex dependence structures, e.g., arbitrarily correlated Gaussian models. An interesting problem is to explore if there are connections between the SPRT and the optimal sequential test rules under general dependencies. Throughout this chapter, we have considered multivariate Gaussian setup, frequent in various areas of stochastic modeling (Hutchinson and Lai, 1990; Olkin and Viana, 1995; Monhor, 2011). However, one interesting extension would be to study the signal detection problem under general distributions and to see whether similar connections with SPRT exist in those cases too.

Bibliography

F. Abramovich, Y. Benjamini, D. L. Donoho, and I. M. Johnstone. Special invited lecture: Adapting to unknown sparsity by controlling the false discovery rate. *The Annals of Statistics*, 34(2):584–653, 2006. URL http://www.jstor.org/stable/25463431. [14]

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. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1995.tb02031.x. [5, 11, 63, 67, 79]

Y. Benjamini and W. Liu. A distribution-free multiple test procedure that controls the false discovery rate. Technical report, Tel Aviv University, Technical report RP-SOR-99-3., 1999a. [10]

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):163–170, 1999b. URL https://www.sciencedirect.com/science/article/pii/S0378375899000403. [10]

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. URL https://doi.org/10.1214/aos/1013699998. [11, 12, 69]

G. Blanchard and E. Roquain. Adaptive false discovery rate control under independence and dependence. *Journal of Machine Learning Research*, 10(97):2837–2871, 2009. URL http://jmlr.org/papers/v10/blanchard09a.html. [12]

M. Bogdan, A. Chakrabarti, F. Frommlet, and J. K. Ghosh. Asymptotic Bayes-optimality under sparsity of some multiple testing procedures. *The Annals of Statistics*, 39(3):1551 – 1579, 2011. URL https://doi.org/10.1214/10-AOS869. [14, 64, 77]

F. Bretz, T Hothorn, and P. Westfall. *Multiple Comparisons Using R*. CRC Press, 2011. [7]

John Tuhao Chen. *Multivariate Bonferroni-Type Inequalities Theory and Applications*. CRC Press, 2015. [20]

A. Cohen, Sackrowitz H. B., and M. Xu. A new multiple testing method in the dependent case. *The Annals of Statistics*, 37(3):1518 – 1544, 2009. URL https://doi.org/10.1214/08-AOS616. [15, 16]

N. Das and S. K. Bhandari. Observation on FWER and FDR for correlated normal, 2020. URL https://arxiv.org/abs/2008.08366. [15]

N. Das and S. K. Bhandari. Bound on FWER for correlated normal. *Statistics & Probability Letters*, 168(C):108943, 2021. URL https://www.sciencedirect.com/science/article/abs/pii/S0167715220302467. [13, 14, 15, 17, 18, 30, 31, 45, 46, 48]

N. Das and S. K. Bhandari. Correction factor of FWER for normal distribution in nearly independent setup, 2023. URL https://arxiv.org/pdf/2306.10546. [14]

S. K. De and M. Baron. Step-up and step-down methods for testing multiple hypotheses in sequential experiments. *Journal of Statistical Planning and Inference*, 142(7): 2059–2070, 2012a. URL https://www.sciencedirect.com/science/article/pii/S037837581200050X. [80]

S. K. De and M. Baron. Sequential bonferroni methods for multiple hypothesis testing with strong control of family-wise error rates i and ii. *Sequential Analysis*, 31(2):238–262, 2012b. URL https://doi.org/10.1080/07474946.2012.665730. [80]

S. K. De and M. Baron. Sequential tests controlling generalized familywise error rates. *Statistical Methodology*, 23:88–102, 2015. URL https://www.sciencedirect.com/science/article/pii/S1572312714000744. [80]

S. Delattre and E. Roquain. On the false discovery proportion convergence under gaussian equi-correlation. *Statistics & Probability Letters*, 81(1):111–115, 2011. URL https://www.sciencedirect.com/science/article/pii/S0167715210002750. [14, 15]

T. Dickhaus. *False Discovery Rate and Asymptotics, PhD Thesis*. Heinrich-Heine-University Düsseldorf, 2008. [15]

T. Dickhaus. *Simultaneous Statistical Inference With Applications in the Life Sciences*. Springer, 2014. [5, 9]

A. Dmitrienko, A.C. Tamhane, and Frank Bretz. *Multiple Testing Problems in Pharmaceutical Statistics*. CRC Biostatistics Series. CRC Press, 2010. [2, 6]

D. Donoho and J. Jin. Higher criticism for detecting sparse heterogeneous mixtures. *The Annals of Statistics*, 32(3):962–994, 2004. URL https://doi.org/10.1214/009053604000000265. [14]

S. Dudoit and Mark J. Laan. *Multiple Testing Procedures with Applications to Genomics*. Springer Series in Statistics. Springer, 2008. [6, 7]

B. Efron. Correlation and large-scale simultaneous significance testing. *Journal of the American Statistical Association*, 102(477):93–103, 2007. URL https://doi.org/10.1198/016214506000001211. [80]

B. Efron. *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Institute of Mathematical Statistics Monographs. Cambridge University Press, 2010a. URL https://doi.org/10.1017/CBO9780511761362. [1, 5, 41, 59, 67]

B. Efron. Correlated z-values and the accuracy of large-scale statistical estimates. *Journal of the American Statistical Association*, 105(491):1042–1055, 2010b. URL https://doi.org/10.1198/jasa.2010.tm09129. [80]

J. Fan and X. Han. Estimation of the False Discovery Proportion with Unknown Dependence. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(4):1143–1164, 09 2016. URL https://doi.org/10.1111/rssb.12204. [12, 80]

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. URL https://doi.org/10.1080/01621459.2012.720478. [12, 80]

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):39–56, 2001a. URL https://www.sciencedirect.com/science/article/pii/S0378375800003049. [12, 15]

H. Finner and M. Roters. On the false discovery rate and expected type i errors. *Biometrical Journal*, 43(8):985–1005, 2001b. URL https://doi.org/10.1002/1521-4036(200112)43:8<985::AID-BIMJ985>3.0.CO;2-4. [11]

H. Finner and M. Roters. Multiple hypotheses testing and expected number of type I. errors. *The Annals of Statistics*, 30(1):220 – 238, 2002. URL https://doi.org/10.1214/aos/1015362191. [14, 59]

H. Finner, T. Dickhaus, and M. Roters. Dependency and false discovery rate: Asymptotics. *The Annals of Statistics*, 35(4):1432–1455, 2007. URL https://www.jstor.org/stable/25464546. [14, 15, 43, 66, 68, 78, 95]

H. Finner, T. Dickhaus, and M. Roters. On the false discovery rate and an asymptotically optimal rejection curve. *The Annals of Statistics*, 37(2):596–618, 2009. URL https://doi.org/10.1214/07-AOS569. [14, 15]

G. Foresti, C. Regazzoni, and P.K. Varshney. *Multisensor Surveillance Systems: The Fusion Perspective*. Springer, New York, 07 2003. ISBN 978-1-4615-0371-2. [80]

R. Gauriot and L. Page. Fooled by Performance Randomness: Overrewarding Luck. *The Review of Economics and Statistics*, 101(4):658–666, 2019. URL https://doi.org/10.1162/rest_a_00783. [2]

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, 08 2002. URL https://doi.org/10.1111/1467-9868.00347. [8]

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. URL https://www.sciencedirect.com/science/article/pii/S0167715208000540. [62]

R. D. Gordon. Values of Mills' Ratio of Area to Bounding Ordinate and of the Normal Probability Integral for Large Values of the Argument. *The Annals of Mathematical Statistics*, 12(3):364 – 366, 1941. URL https://doi.org/10.1214/aoms/1177731721. [48]

J. Gou, A. C. Tamhane, D. Xi, and D. Rom. A class of improved hybrid Hochberg-Hommel type step-up multiple test procedures. *Biometrika*, 101(4):899–911, 2014. URL http://www.jstor.org/stable/43304695. [70, 77]

A. Grandhi. *Multiple testing procedures for complex structured hypotheses and directional decisions, PhD Thesis*. New Jersey Institute of Technology, 2015. [7]

R. C. Grinold and R. N. Kahn. *Active Portfolio Management: A Quantitative Approach for Providing Superior Returns and Controlling Risk*. McGraw-Hill, 2000. [2]

W. Guo. *Generalized Error Control in Multiple Hypothesis Testing, PhD Thesis*. University of Cincinnati, 2007. [3]

W. Guo and M. B. Rao. On optimality of the Benjamini–Hochberg procedure for the false discovery rate. *Statistics & Probability Letters*, 78(14):2024–2030, 2008. URL https://www.sciencedirect.com/science/article/pii/S0167715208000734. [64, 69, 77]

W. Guo and S. Sarkar. Adaptive controls of FWER and FDR under block dependence. *Journal of Statistical Planning and Inference*, 208:13–24, 2020. URL https://www.sciencedirect.com/science/article/pii/S0378375819301181. [13]

X. He and J. Bartroff. Asymptotically optimal sequential FDR and pFDR control with (or without) prior information on the number of signals. *Journal of Statistical Planning and Inference*, 210:87–99, 2021. URL https://www.sciencedirect.com/science/article/pii/S0378375820300604. [80, 84, 92, 94]

Y. Hochberg. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, 75(4):800–802, 12 1988. URL https://doi.org/10.1093/biomet/75.4.800. [11, 63, 64, 79]

Y. Hochberg and A. C. Tamhane. *Multiple Comparison Procedures*. Wiley, 1987. [3, 4, 13]

S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979. URL http://www.jstor.org/stable/4615733. [10, 61, 79]

G. Hommel. A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, 75(2):383–386, 06 1988. URL https://doi.org/10.1093/biomet/75.2.383. [70, 79]

J. C. Hsu. *Multiple Comparisons Theory and Methods*. CRC Press, 1996. [4]

Y. Huang and J. C. Hsu. Hochberg's step-up method: Cutting corners off Holm's step-down method. *Biometrika*, 94(4):965–975, 2007. URL http://www.jstor.org/stable/20441429. [13, 77, 80]

T. P. Hutchinson and C. D. Lai. *Continuous bivariate distributions, emphasising applications*. Rumsby Scientific Publishing Adelaide, S. Aust, 1990. ISBN 0731672062. [32, 43, 95]

E. G. Kounias. Bounds for the probability of a union, with applications. *The Annals of Mathematical Statistics*, 39(6):2154–2158, 1968. URL http://www.jstor.org/stable/2239318. [26, 35]

S. M. Kwerel. Most stringent bounds on aggregated probabilities of partially specified dependent probability systems. *Journal of the American Statistical Association*, 70(350):472–479, 1975. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1975.10479893. [19, 35]

J. T. Leek and J. D. Storey. A general framework for multiple testing dependence. *Proceedings of the National Academy of Sciences*, 105(48):18718–18723, 2008. URL https://www.pnas.org/doi/abs/10.1073/pnas.0808709105. [12]

E. L. Lehmann and Joseph P. Romano. Generalizations of the familywise error rate. *The Annals of Statistics*, 33(3):1138 – 1154, 2005. URL https://doi.org/10.1214/009053605000000084. [6, 33, 53]

J. Liu, C. Zhang, and D. Page. Multiple testing under dependence via graphical models. *The Annals of Applied Statistics*, 10(3):1699 – 1724, 2016. URL https://doi.org/10.1214/16-AOAS956. [80]

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. URL https://doi.org/10.1080/03610920601126175. [18]

R. Miller. *Simultaneous Statistical Inference*. Springer-Verlag, 1981. [4]

D. Monhor. A new probabilistic approach to the path criticality in stochastic pert. *European Journal of Operations Research*, 19(4):615–633, 2011. URL https://doi.org/10.1007/s10100-010-0151-x. [32, 43, 95]

D. Monhor. Inequalities for correlated bivariate normal distribution function. *Probability in the Engineering and Informational Sciences*, 27(1):115–123, 2013. URL https://doi.org/10.1017/S0269964812000332. [20, 22, 38]

N Mukhopadhyay and B. M. de Silva. *Sequential Methods and Their Applications*. Chapman and Hall, 2019. [85]

I. Olkin and M. Viana. Correlation analysis of extreme observations from a multivariate normal distribution. *Journal of the American Statistical Association*, 90(432):1373–1379, 1995. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476642. [17, 32, 43, 95]

M. A. Proschan and P. A. Shaw. Asymptotics of Bonferroni for dependent normal test statistics. *Statistics & Probability Letters*, 81(7):739–748, 2011. URL https://www.sciencedirect.com/science/article/pii/S0167715210003329. Statistics in Biological and Medical Sciences. [13, 14, 15]

X. Qiu, L. Klebanov, and A. Yakovlev. Correlation between gene expression levels and limitations of the empirical bayes methodology for finding differentially expressed genes. *Statistical Applications in Genetics and Molecular Biology*, 4(1), 2005. URL https://doi.org/10.2202/1544-6115.1157. [12, 80]

P. H. Ramsey. Power differences between pairwise multiple comparisons. *Journal of the American Statistical Association*, 73(363):479–485, 1978. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1978.10480038. [7]

C. R. Rao. *Linear Statistical Inference and its Applications: Second Editon*. Linear Statistical Inference and its Applications: Second Editon, 1973. [85]

T. S. Rappaport. *Wireless communications: Principles and practice*. Prentice Hall, Upper Saddle River, NJ., 2002. [80]

Z. Ren. *Model-free Methods for Multiple Testing and Predictive Inference , PhD Thesis*. Stanford University, 2021. [6]

D. M. Rom. A sequentially rejective test procedure based on a modified Bonferroni inequality. *Biometrika*, 77(3):663–665, 09 1990. URL https://doi.org/10.1093/biomet/77.3.663. [10]

D. M. Rom and B. Holland. A new closed multiple testing procedure for hierarchical families of hypotheses. *Journal of Statistical Planning and Inference*, 46(3):265–275, 1995. URL https://www.sciencedirect.com/science/article/pii/037837589400116D. [3]

R. Roy and S. K. Bhandari. Asymptotic bayes' optimality under sparsity for exchangeable dependent multivariate normal test statistics. *Statistics & Probability Letters*, 207:110030, 2024. URL https://www.sciencedirect.com/science/article/pii/S0167715223002535. [15]

R. Roy, S. K. De, and S. K. Bhandari. Large-scale adaptive multiple testing for sequential data controlling false discovery and nondiscovery rates, 2023. URL https://arxiv.org/abs/2306.05315. [80]

S. K. Sarkar. Some Results on False Discovery Rate in Stepwise multiple testing procedures. *The Annals of Statistics*, 30(1):239– 257, 2002. URL https://doi.org/10.1214/aos/1015362192. [11, 12, 67]

S. K Sarkar. Fdr-controlling stepwise procedures and their false negatives rates. *Journal of Statistical Planning and Inference*, 125(1):119–137, 2004. URL https://www.sciencedirect.com/science/article/pii/S037837580400059X. The Third International Conference on Multiple Comparisons. [8]

S. K. Sarkar. Stepup procedures controlling generalized FWER and generalized FDR. *The Annals of Statistics*, 35(6):2405 – 2420, 2007. URL https://doi.org/10.1214/009053607000000398. [6, 15]

S. K. Sarkar. On methods controlling the false discovery rate. *Sankhyā: The Indian Journal of Statistics, Series A*, 70(2):135–168, 2008. URL http://www.jstor.org/stable/41234409. [12]

G. Saunders. *Family-wise Error Rate Control in Quantitative Trait Loci (QTL) Mapping and Gene Ontology Graphs with Remarks on Family Selection, PhD Thesis*. Utah State University, 2014. [3, 13]

D. Singh, P. G. Febbo, K. Ross, D. G. Jackson, J. Manola, C. Ladd, P. Tamayo, A. A. Renshaw, A. V. D'Amico, J. P. Richie, E. S. Lander, M. Loda, P. W. Kantoff, T. R. Golub, and W. R. Sellers. Gene expression correlates of clinical prostate cancer behavior. *Cancer cell*, 1(2):203—209, March 2002. URL https://doi.org/10.1016/s1535-6108(02)00030-2. [41]

D. Slepian. The one-sided barrier problem for gaussian noise. *The Bell System Technical Journal*, 41(2):463–501, 1962. URL https://doi.org/10.1002/j.1538-7305.1962.tb02419.x. [51]

Y. Song. *Some Topics in Sequential Analysis, PhD Thesis*. University of Illinois at Urbana-Champaign, 2019. [80, 82]

Y. Song and G. Fellouris. Asymptotically optimal, sequential, multiple testing procedures with prior information on the number of signals. *Electronic Journal of Statistics*, 11(1): 338 – 363, 2017. URL https://doi.org/10.1214/17-EJS1223. [80, 82, 83, 84, 88]

Y. Song and G. Fellouris. Sequential multiple testing with generalized error control: An asymptotic optimality theory. *The Annals of Statistics*, 47(3):1776 – 1803, 2019. URL https://doi.org/10.1214/18-AOS1737. [80]

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. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00346. [5]

J. D. Storey. The positive false discovery rate: A bayesian interpretation and the q-value. *The Annals of Statistics*, 31(6):2013–2035, 2003. URL http://www.jstor.org/stable/3448445. [5, 8]

J. D. Storey and R. Tibshirani. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 100(16):9440–9445, 2003. URL https://www.pnas.org/doi/abs/10.1073/pnas.1530509100. [12]

M. M. Subasi, E. Subasi, A. Binmahfoudh, and A. Prékopa. New bounds for the probability that at least k-out-of-n events occur with unimodal distributions. *Discrete*

Applied Mathematics, 226:138–157, 2017. URL https://www.sciencedirect.com/science/article/pii/S0166218X17301440. [40]

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. URL https://doi.org/10.1198/016214507000000545. [8]

W. Sun and T. T. Cai. Large-scale multiple testing under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):393–424, 2009. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2008.00694.x. [80]

A. C. Tamhane. Multiple comparisons. In *Design and Analysis of Experiments*, volume 13 of *Handbook of Statistics*, pages 587–630. Elsevier, 1996. URL https://www.sciencedirect.com/science/article/pii/S0169716196130200. [4]

V. Trevino, F. Falciani, and H. A. Barrera-Saldaña. DNA microarrays: a powerful genomic tool for biomedical and clinical research. *Molecular Medicine*, 13(9-10):527–541, 2007. URL https://doi.org/10.2119/2006-00107.Trevino. [2]

J. W. Tukey. *The Problem of Multiple Comparisons*. Mimeographed monograph, Princeton University, 1953. [3]

M. Viana. Linear combinations of ordered symmetric observations with applications to visual acuity. In *Order Statistics: Applications*, volume 17 of *Handbook of Statistics*, pages 513–524. Elsevier, 1998. URL https://www.sciencedirect.com/science/article/pii/S0169716198170219. [17]

P. H. Westfall and S. S. Young. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. Wiley, 1993. [4]

P. H. Westfall, R. D. Tobias, D. Rom, R. D. Wolfinger, and Y. Hochberg. *Multiple Comparisons and Multiple Testing Using the SAS System*. SAS Institute, Inc., 1999. [8, 78]

J. Xie, T. T. Cai, J. Maris, and H. Li. Optimal false discovery rate control for dependent data. *Statistics and its Interface*, 4(4):417–430, 2011. [80]

J. Yang, F. Alajaji, and G. Takahara. On bounding the union probability using partial weighted information. *Statistics & Probability Letters*, 116:38–44, 2016. URL https://www.sciencedirect.com/science/article/pii/S0167715216300323. [20]

Z. Šidák. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318):626–633, 1967. URL https://doi.org/10.1080/01621459.1967.10482935. [9]