

Interfaces Between Bayesian and Frequentist Multiple Testing

by

Shih-Han Chang

Department of Statistical Science
Duke University

Date: _____

Approved:

James O. Berger, Supervisor

Surya Tokdar

David Banks

Yin Xia

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Statistical Science
in the Graduate School of Duke University
2015

ABSTRACT

Interfaces Between Bayesian and Frequentist Multiple Testing

by

Shih-Han Chang

Department of Statistical Science
Duke University

Date: _____

Approved:

James O. Berger, Supervisor

Surya Tokdar

David Banks

Yin Xia

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in the Department of Statistical Science
in the Graduate School of Duke University
2015

Copyright © 2015 by Shih-Han Chang
All rights reserved except the rights granted by the
Creative Commons Attribution-Noncommercial Licence

Abstract

This thesis investigates frequentist properties of Bayesian multiple testing procedures in a variety of scenarios and depicts the asymptotic behaviors of Bayesian methods. Both Bayesian and frequentist approaches to multiplicity control are studied and compared, with special focus on understanding the multiplicity control behavior in situations of dependence between test statistics.

Chapter 2 examines a problem of testing mutually exclusive hypotheses with dependent data. The Bayesian approach is shown to have excellent frequentist properties and is argued to be the most effective way of obtaining frequentist multiplicity control without sacrificing power. Chapter 3 further generalizes the model such that multiple signals are acceptable, and depicts the asymptotic behavior of false positives rates and the expected number of false positives. Chapter 4 considers the problem of dealing with a sequence of different trials concerning some medical or scientific issue, and discusses the possibilities for multiplicity control of the sequence. Chapter 5 addresses issues and efforts in reconciling frequentist and Bayesian approaches in sequential endpoint testing. We consider the conditional frequentist approach in sequential endpoint testing and show several examples in which Bayesian and frequentist methodologies cannot be made to match.

To my family, for their unwavering faith in me.

Contents

Abstract	iv
List of Tables	ix
List of Figures	x
List of Abbreviations and Symbols	xii
Acknowledgements	xiii
1 Introduction	1
1.1 Multiple testing	1
1.1.1 Applications	2
1.1.2 A review of multiplicity adjustment	2
1.1.3 Family-wise error rate (FWER) control	4
1.1.4 False discovery rate (FDR) control	5
1.1.5 Other approaches	6
1.2 Outline and Contributions of the Thesis	6
2 Comparison of Bayesian and frequentist multiplicity correction for testing mutually exclusive hypotheses under data dependence	8
2.1 Introduction	8
2.2 Frequentist Multiplicity Control	10
2.2.1 An Ad hoc Procedure	11
2.2.2 Likelihood Ratio Test	12
2.3 A Bayesian Test	15
2.4 The situation as the correlation goes to 1	17

2.5	Asymptotic frequentist properties of Bayesian procedures	21
2.5.1	Posterior probabilities	22
2.5.2	False positive probability	27
2.6	A Type II maximum likelihood approach	29
2.7	Power analysis for the Type II ML Bayesian procedure	36
2.8	Analysis as the information grows	38
2.9	Conclusions	42
2.10	Appendix	42
2.10.1	Normal Theory	42
2.10.2	Type II MLE	44
3	Frequentist multiplicity control of Bayesian model selection with spike-and-slab priors	51
3.1	The standard Bayesian model of multiple testing	52
3.1.1	Bayes factors	53
3.1.2	Posterior model probabilities	53
3.2	Posterior probability of the null model as n grows	54
3.3	Asymptotic frequentist properties of Bayesian procedures	56
3.3.1	Inclusion probabilities and the decision rule	56
3.3.2	Expected number of false positives	60
3.3.3	FPPs for various structures of prior probabilities	61
3.4	Conclusion	66
3.5	Appendix	67
4	Bayesian multiple testing for a sequence of trials	72
4.1	Introduction	72
4.2	Bayesian sequence multiplicity control	73
4.2.1	Recursive formula	75
4.3	HIV vaccines	75

4.4	Analysis when only p -values are available	77
5	Reconciling frequentist and Bayesian methods in sequential endpoint testing	79
5.1	Introducion	79
5.2	Conditional frequentist testing	81
5.2.1	Testing of two simple hypotheses	81
5.2.2	Conditioning partitions for the sequential endpoint problem	82
5.2.3	Conditional frequentist error probabilities for the random partition	84
5.3	Assessing the possibility of Bayesian/frequentist agreement	86
5.3.1	Testing two simple hypotheses	87
5.3.2	Testing composite hypotheses	91
5.4	Appendix	95
5.4.1	Conditional frequentist analysis for the fixed partition	95
5.4.2	Conditional frequentist test	96
	Bibliography	99
	Biography	104

List of Tables

1.1	Decisions and errors	3
5.1	Frequentist error probabilities, conditional on being in the random partition.	84
5.2	Decisions and reported errors in sequential endpoint tests	87
5.3	Decisions and reported frequentist errors for two endpoints.	87
5.4	Decisions and reported errors	88
5.5	Decision rules	95
5.6	Conditional probabilities up to a normalizing constant	95

List of Figures

2.1	Ratio of estimated and true posterior probability of M_1 as n grows under the null model and fixed τ, r , different ρ . Each subplot is for different correlations and contains 200 simulations.	24
2.2	Estimated (red line) and true posterior probability (blue line) of M_1 for different τ under the null model, for fixed $n = 2000, \rho = r = 0.5$	25
2.3	Convergence of $P(M_0 \mathbf{x})$ to the prior probability (0.5) under the null model. Each subplot has a different correlation and contains 50 simulations.	27
2.4	Comparison of the simulated FPP and its asymptotic approximation when $n = 10^6, p = 1/3; r = 0.5, \rho = 0$ as n varies from 10^1 to 10^8 (x-axis in exponential scale)	34
2.5	threshold probability (p) versus the number of hypotheses (n) for fixed FPP(= 0.05) and prior probability of null model (= 0.5).	35
2.6	Solution of $\frac{1}{k^*} - \frac{1}{2}$ (y-axis) versus $\frac{p}{(1-p)(1-r)}$ (x-axis).	35
2.7	$\hat{\tau}^2$ versus θ_i and fixed $n = 10^4, r = 0.5, \rho = 0$, and $p = 1/3$. Each point is the average $\hat{\tau}^2$ among 10^4 independent draws from the multivariate normal with the constraint $\tau^2 > 1$	36
2.8	Power versus θ_i and fixed $n = 10^4, r = 0.5, \rho = 0, p = 1/3$. Each point is the average acceptance rate of the true non-null model with respect to θ in the x-axis range.	37
2.9	θ_i versus n for fixed power 0.95 (above) 0.75 (below), $\rho = 0, p = 1/3$	38
3.1	Box plot of the posterior probability of the null model under the null model, where $P(M_0) = 0.5, k_{max} = 4, \tau = 2$. The green horizontal line is $y = 0.5$. For each m (number of hypotheses), 3000 iterations have been performed.	56

3.2	For $k_{max} = 4$, $\tau = 2$, and various numbers of hypotheses m , box plots of the logarithm of the ratio of the true inclusion probability to the estimated inclusion probability, under the null model. For each m , 3000 iterations have been carried out. The red horizontal line is $y = 0$ (ratio = 1).	58
3.3	Box plots of empirical false positive rates with median (red line), mean (green dashed line) and the theoretical false positive probabilities (black lines), when $k_{max} = 2$, the decision threshold is $q = 0.05$, and $\tau = 1$	60
3.4	False positive probability based on the threshold in Theorem 3.3.10. The blue curve is numerical FPP (average FPP at given m); the green curve is $1 - (1 - \frac{\alpha}{m})^m$ where $\alpha = 0.05$	64
3.5	Box plots of inclusion probability $P(\mu_i \neq 0 X)$ versus signal size θ_i . $m = 100$, $k_{max} = 3$, $\tau = 1$. For each θ_i , 100 iterations were generated to get the box plot.	65
3.6	Posterior of p versus different noise ratios. $m = 20$, $\tau = 2$. For each plot, k_{max} is indicated in the subtitle, and simulate data \mathbf{x} according to the top right true noise ratio legend.	66

List of Abbreviations and Symbols

Symbols

\mathbf{X}	a vector of i.i.d random variables.
X_i	i^{th} element of \mathbf{X} .
\mathbf{x}	a vector of observations.
γ_i	an indicator for nonzero means.
$\boldsymbol{\gamma}$	$(\gamma_1, \dots, \gamma_n)$
M_0^i	i^{th} null model.
M_1^i	i^{th} alternative model.
$\Phi(x)$	cumulative density function of Gaussian distribution.
$\phi(x)$	probability density function of Gaussian distribution.
α	Type-I error.
ρ	correlation

Abbreviations

FPP	False Positive Probability.
FDR	False Discovery Rate
FWER	Family-wise Error Rate.
i.i.d.	independent and identically distributed.

Acknowledgements

There were many people involved in the process of my PhD that have assisted me greatly to reach this point. First, I am extremely grateful to the advice and support from professor Jim Berger, who guided me extensively through the process of finding a research topic and accomplishing high-level research by thinking critically and thoroughly. His feedback is always insightful. I appreciate his time and assistance over the years. I would also like to thank my committee members: David Banks, Surya Tokdar, Yin Xia and Barbara Engelhardt, and statistics professors Robert Wolpert, David Dunson for all the feedback and support they provided during my graduate studies.

Thank you to my fellow classmates - I have been fortunate enough to have you from our First Year Exam to ISBA Cancún, from discussing research to career development, I have benefited greatly by having you there with me through it all. I also owe a great deal of thanks to my friends from the Duke Taiwanese Student Association, the Duke Table Tennis Varsity Team, and my internship in California. My Duke life and internships were colorful and enjoyable because of you.

To all my friends from the Duke Chapel and FYTER congregations, especially Karen and Michael, Carol and Richard, Steve and Sara, and Howard and Joann I always feel your warmhearted hospitality and love. You are like family and gave me a place where I can feel like home. Special thanks to pastors Luke and Carol for sharing your wisdom; every sermon has been inspiring and enlightening. Thank you to Chapel conductor Rodney for constantly bringing us great music - every Handel's

Messiah and Christmas Eve Service has been so beautiful and memorable.

Finally, I would like to thank my parents Yu-Miao and Chang-Wang, brother Li-Ying, uncle Yumin and aunt Yolanda for their consistent care and support; thank you to my wonderful wife Chien-Kuang for your love, care, and companionship. Without you, none of this would have been possible.

Introduction

1.1 Multiple testing

Multiple testing has been an important topic for decades, and as technology advances, more and more scientific disciplines require testing a large number of hypotheses simultaneously or sequentially. Simply testing each hypothesis at some significance level α and reporting findings results in false discoveries and irreproducible outcomes in multiple testing situations (Ioannidis (2005), Banks et al. (2011), Heller et al. (2014)). In clinical trials or pharmaceutical studies, such errors may lead to wasting resources or potentially costing lives.

The Bayesian approach to controlling for multiplicity operates through the prior probabilities assigned to hypotheses. For example, suppose the observations come from

$$Y_i \sim pF_0 + (1 - p)F_1,$$

where p is the probability that the null model is true, and F_0, F_1 denote the probability distributions corresponding to the null and the alternative model, respectively. By giving a non-degenerate prior on p , if strong signals are discovered, p could shift towards zero. This automatically controls for multiplicity, as discussed extensively

in Scott and Berger (2006), Scott and Berger (2010) for a two-groups model and variable-selection in linear models. This simplicity of Bayesian multiplicity control makes it highly attractive, the hope being that it can lead to fully powered multiplicity control (to be defined later) even in scenarios of high dependence amongst test statistics (where frequentist multiplicity control can be difficult).

To be widely accepted, however, the frequentist properties of Bayesian multiplicity control procedures need to be understood. The central goal of this thesis, therefore, is to investigate frequentist properties of Bayesian multiplicity control in a variety of scenarios. It aims to answer when a Bayesian procedure can be considered to be a valid frequentist procedure, and depicts the asymptotic behaviors of Bayesian procedures as the number of tests grows.

1.1.1 Applications

Multiple testing methodology has wide applications including microarrays (Dudoit and Van Der Laan (2007), Noble (2009)), data quality (Karr et al. (2006)), neuroimage classification (Genovese et al. (2002)), image processing (White and Ghosal (2014)), astrophysics Miller et al. (2001), high-dimensional hypothesis testing (Cai et al. (2013), Tony Cai et al. (2014)) etc. It is also worth noting the strong connection between multiplicity and sparsity, which is discussed in Malloy and Nowak (2011), Abramovich et al. (2006) Datta et al. (2013), Bogdan et al. (2008b), Cai and Xia (2014).

1.1.2 A review of multiplicity adjustment

Many multiple testing procedures have been developed to control type I or other errors variants. For conducting n hypothesis tests, the i^{th} test consists of testing the null hypothesis M_0^i versus the alternative M_1^i . For each individual test i , there are four possible scenarios : (1) rejecting the null when the null is true, (2) rejecting the null when the alternative is true, (3) failing to reject the null when the null is true

or (4) failing to reject the null when the alternative is true. Scenarios (2) and (3) are correct. Because we are conducting multiple tests, we need to consider the fact that each test will yield one of the four outcomes, and hence may have a correct or incorrect result. The following table shows the counts for the outcomes for n tests in terms of these four scenarios:

Table 1.1: Decisions and errors

	Null True	Alternative True	Total
Fail to Reject Null	U	T	$n - R$
Reject Null	V	S	R
	N_0	$n - N_0$	n

Recall that only two of these scenarios (rejecting the null when the alternative is true and failing to reject the null when the null is true) are correct. We can then compute error rates for the n tests based on the counts in the table for the other two scenarios. The types of errors that have been often considered are:

Per-comparison error rate (PCER): $E(V)/n$

Per-family error rate (PFER): expected number of Type-I errors, $E(V)$

Family-wise error rate (FWER): $P(V \geq 1)$

False discovery rate (FDR): $E(V/R \mid R > 0)P(R > 0)$

Positive false discovery rate (pFDR): $E(V/R \mid R > 0)$

A relationship between these errors is

$$PCER \leq FDR \leq pFDR \leq FWER \leq PFER, \quad (1.1)$$

cf. Ge et al. (2003), Goeman and Solari (2014) for details.

We use the standard term *weak control* to refer to computing any of these error rates assuming all n null hypotheses are true; *strong control* refers to computing these error rates under any combination of true and false nulls.

It may be desirable to control the rate of one or more of these type of errors. In

the following sections, we will discuss a few methods for controlling the error rate for specific error types.

1.1.3 Family-wise error rate (FWER) control

Three possible methods for controlling FWER are as follows:

1. The *Bonferroni procedure* rejects a hypothesis if its p -value is less than α/n , which directly controls the FWER error. This is equivalent to rejecting if

$$\tilde{p}_i = \min\{np_i, 1\} \leq \alpha.$$

2. Another single-step FWER adjustment is Sidk's method:

$$\tilde{p}_i = 1 - (1 - p_i)^n.$$

3. The previous two methods adjust for all p -values simultaneously. Holm (1979) instead attempts to gain more power by successively making smaller adjustments to the p -values. He denotes $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(n)}$ as the ordered raw p -values, and rejects a hypothesis if:

$$\tilde{p}_{(i)} = \max_{k \in \{1, \dots, i\}} \{\min((n - k + 1)p_{(k)}, 1)\} \leq \alpha.$$

All of these methods suffer when the test statistics are not independent. Westfall and Young (1993)'s step-down resampling-based algorithm controls FWER without the independence assumption.

However, in many applications, such as RNA sequencing, which have millions of hypotheses, controlling FWER is too conservative. With millions of tests being run, it is highly likely that at least one of them will result in a Type 1 Error. Such a strong control as FWER, which attempts to eliminate Type 1 Error from all tests, is overly conservative.

1.1.4 False discovery rate (FDR) control

FDR is one alternative error metric that typically reduces the FWER without sacrificing power (cf (1.1)).

If all null hypotheses are true and test statistics are independent, then 5% of p -values are expected to be less than the 0.05 threshold since p -values are uniformly distributed under null. Motivated by this, Benjamini and Hochberg (1995) considered dividing each observed p -value by its percentile rank in order to estimate FDR. In this approach, they declare significance if

$$\tilde{p}_{(i)} \leq \min_{k \in \{1, \dots, n\}} \{\min(n p_{(k)}/k, 1)\}.$$

Alternatively, Efron et al. (2001) proposed the two-groups approach: let Z_i be the z-value of i^{th} hypothesis testing. Assume Z_1, \dots, Z_n follows the below mixture cumulative distribution functions (cdf):

$$F(z) = p_0 F_0(z) + (1 - p_0) F_1(z), \quad (1.2)$$

where p_0 is the null prior probability, and F_0, F_1 are the cdfs of null and alternative hypotheses, respectively. From Bayes rule,

$$Fdr(z) = P(\text{null} \mid Z \leq z) = p_0 F_0(z) / (F_0(z) + F_1(z)),$$

$$\text{local FDR} = fdr(z) = P(\text{null} \mid Z = z) = p_0 f_0(z) / (f_0(z) + f_1(z)),$$

where F_0, F_1 are probability distribution functions corresponding to null and alternative models. Note that local FDR is calculated with respect to a single score, which is useful for checking the false positive probability of a single case. Although (1.2) is a strong assumption, the two-groups approach is easy to interpret, deals with dependency (Efron (2004)) and links to empirical Bayes.

1.1.5 Other approaches

We have mentioned some Bayesian work in the beginning of this chapter which attempts to control error rates in multiple testing situations. There are other Bayes and Empirical Bayes approaches such as parametric empirical Bayes (Bogdan et al. (2007), Bogdan et al. (2008a), Dutta et al. (2015)), decision theoretical FDR (Muller et al. (2006), Bogdan et al. (2011)) and, in different contexts, subgroup analysis (Wang et al. (2007), Berger et al. (2014)), or time series (Scott (2009)).

1.2 Outline and Contributions of the Thesis

Chapter 2 examines a problem of testing mutually exclusive hypotheses with dependent data. Both Bayesian and frequentist approaches to multiplicity control are studied and compared, and this work helps us gain understanding as to the effect of data dependence on each approach. The Bayesian approach is shown to have excellent frequentist properties and is argued to be the most effective way of obtaining frequentist multiplicity control without sacrificing power.

Chapter 3 generalizes the model so that multiple signals are acceptable. The challenging part is the exponential growth in model size (i.e. 2^n models for n tests). We make a very mild assumption on the model complexity and determine the asymptotic behavior of false positives rates and the expected number of false positives.

Chapter 4 applies Bayesian methods to clinical trials, where multiplicity has been neglected in many studies. We demonstrate how to use Bayesian methodology to control for multiplicity and exemplify the technique on HIV vaccine data. In addition, since many existing studies have been conducted via frequentist methods, we show possible ways to obtain Bayesian answers by converting p -values to Bayes factors.

Chapter 5 addresses issues and efforts in reconciling frequentist and Bayesian approaches in the sequential endpoint testing problem. Conditional frequentist methodology has been one of the most successful ways to reconcile Bayes and frequentist

approaches in the single hypothesis testing scenario. However, we show that no reconciliation appears to be possible, in general, for sequential endpoint testing.

Comparison of Bayesian and frequentist multiplicity correction for testing mutually exclusive hypotheses under data dependence

2.1 Introduction

Modern scientific experiments often require considering a large number of hypotheses simultaneously (Efron (2004), Noble (2009)) and has led to extensive interest in controlling for multiple testing (henceforth, just termed controlling for multiplicity). Many multiplicity control methods have been proposed in the frequentist literature, such as the Bonferroni procedure which controls the family-wise error rates, and various versions of false discovery rates (cf. Benjamini and Hochberg (1995) and Storey (2003)) which control for the fraction of false discoveries to stated discoveries. Asymptotic behavior of false discovery rate have been studied in Abramovich et al. (2006).

The Bayesian approach to controlling for multiplicity operates through the prior probabilities assigned to hypotheses. For instance, in the scenario that is considered herein of testing mutually exclusive hypotheses (only one of n considered hypotheses

can be true), one can simply assign each hypothesis prior probability equal to $1/n$ and carry out the Bayesian analysis; this automatically controls for multiplicity. That multiplicity is controlled through prior probabilities of hypotheses or models is extensively discussed in Scott and Berger (2006), Scott and Berger (2010) for a two-groups model and variable-selection in linear models.

One of the appeals of the Bayesian approach to multiplicity control is that it does not depend on the dependence structure of the data; the Bayes procedure will automatically adapt to the dependence structure through Bayes theorem, but the prior probability assignment that is controlling for multiplicity is unaffected by data dependence. In contrast, frequentist approaches to multiplicity control are usually highly affected by data dependence. For instance, the Bonferroni correction is fine if the test statistics for the hypotheses being tested are independent, but can be much too conservative (losing detection power), if the test statistics are dependent.

An interesting possibility for frequentist multiplicity control in data dependence situations is thus to develop the procedure in a Bayesian fashion and verify that the procedure has sufficient control from a frequentist perspective. This has the potential of yielding optimally powered frequentist procedures for multiplicity control. There have been other papers that study the frequentist properties of Bayesian multiplicity control procedures (Bogdan et al. (2008a), Guo and Heitjan (2010), Abramovich and Angelini (2006)), but they have not focused on the situation of data dependence.

We investigate the potential for this program by an exhaustive analysis of the simplest multiple testing problem which exhibits data dependence. The data $\mathbf{X} = (X_1, \dots, X_n)'$ arises from the multivariate normal distribution

$$\mathbf{X} \sim \text{multinorm} \left(\begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{pmatrix}, \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} \right), \quad (2.1)$$

where ρ is the correlation between the observations. Consider testing the n hypotheses $M_0^i : \theta_i = 0$ versus $M_1^i : \theta_i \neq 0$, but under the assumption that at most one alternative hypothesis could be true. (It is possible that no alternative is true.) Although our study of this problem is pedagogical in nature, such testing problems can arise in signal detection, when a signal could arise in one and only one of n channels, and there is common background noise in all channels, leading to the equal correlation structure. We will, for convenience in exposition, use this language in referring to the situation.

In subsection 2 we introduce two natural frequentist procedures for multiplicity control in this problem and, in subsection 3, we introduce a natural Bayesian procedure. subsection 4 explores a highly curious phenomenon that is encountered in the situation where ρ is near 1; when $n > 2$, the Bayesian procedure finds the true alternative hypothesis with near certainty, while an ad hoc frequentist procedure fails to do so. subsections 5 and 6 study the frequentist properties of the original Bayesian procedure and the Type-II MLE approach, showing that, as $n \rightarrow \infty$, the Bayesian procedures have strong frequentist control of error. subsection 7 discusses power of the Type II maximum likelihood procedure, and argues that it is essentially an optimal frequentist procedure. Subsection 8 considers the situation in which there is a data sample of growing size m for each θ_i .

2.2 Frequentist Multiplicity Control

Two natural frequentist procedures are considered.

2.2.1 An Ad hoc Procedure

Declare channel i to have the signal if $\max_{1 \leq j \leq n} |X_j| > c$, where c is determined to achieve overall family-wise error control

$$\alpha = P \left(\max_{1 \leq j \leq n} |X_j| > c \mid \theta_i = 0 \forall i \right). \quad (2.2)$$

Lemma 2.2.1. (2.2) can be expressed as

$$\alpha = 1 - \mathbb{E}^Z \left\{ \left[\Phi \left(\frac{c - \sqrt{\rho}Z}{\sqrt{1-\rho}} \right) - \Phi \left(\frac{-c - \sqrt{\rho}Z}{\sqrt{1-\rho}} \right) \right]^n \right\},$$

where the expectation is with respect to $Z \sim N(0, 1)$.

Proof. By Lemma 2.21, under the null model, X_i can be written as $X_i = \sqrt{\rho}Z + \sqrt{1-\rho}Z_i$, where the Z and the Z_i are independent standard normal random variables. Thus

$$\begin{aligned} & P \left(\max_{1 \leq j \leq n} |X_j| > c \mid \theta_i = 0 \forall i \right) \\ &= 1 - \mathbb{E}^Z \left\{ P \left(\text{for all } j, |\sqrt{\rho}Z + \sqrt{1-\rho}Z_j| < c \mid Z \right) \right\} \\ &= 1 - \mathbb{E}^Z \left\{ \prod_1^n P \left(\frac{-c - \sqrt{\rho}Z}{\sqrt{1-\rho}} < Z_j < \frac{c - \sqrt{\rho}Z}{\sqrt{1-\rho}} \mid Z \right) \right\} \\ &= 1 - \mathbb{E}^Z \left\{ \left[\Phi \left(\frac{c - \sqrt{\rho}Z}{\sqrt{1-\rho}} \right) - \Phi \left(\frac{-c - \sqrt{\rho}Z}{\sqrt{1-\rho}} \right) \right]^n \right\}. \end{aligned}$$

□

Corollary 2.2.2.

- When $\rho = 0$, $\Phi(c) = 1 + \frac{\log(1-\alpha)}{2n} + O(1/n^2)$, essentially calling for the Bonferroni correction.

- When $\rho \rightarrow 1$, $\Phi(c) \rightarrow 1 - \frac{\alpha}{2}$, so the critical region is the same as that for a single test.

Proof. If $\rho = 0$, $\alpha = 1 - (\Phi(c) - \Phi(-c))^n = 1 - (2\Phi(c) - 1)^n$, from which it follows that

$$\Phi(c) = \frac{1 + (1 - \alpha)^{1/n}}{2} = \frac{1 + 1 + \frac{\log(1-\alpha)}{n} + O(1/n^2)}{2}.$$

If $\rho \rightarrow 1$, by Lemma 2.21,

$$\begin{aligned} & \lim_{\rho \rightarrow 1} P \left(\max_{1 \leq j \leq n} |X_j| > c \mid \theta_i = 0 \forall i \right) \\ &= 1 - \lim_{\rho \rightarrow 1} \mathbb{E}^{Z_1, \dots, Z_n} \left\{ P \left(|\sqrt{\rho}Z + \sqrt{1-\rho}Z_j| < c \mid Z_1, \dots, Z_n \right) \right\} \\ &= 1 - \mathbb{E}^{Z_1, \dots, Z_n} \left\{ \lim_{\rho \rightarrow 1} P \left(|\sqrt{\rho}Z + \sqrt{1-\rho}Z_j| < c \mid Z_1, \dots, Z_n \right) \right\} \\ &= 1 - (\Phi(c) - \Phi(-c)) \\ &= 2(1 - \Phi(c)). \end{aligned}$$

□

The extreme effect of dependence in the data on frequentist multiplicity correction is clear here; the correction ranges from full Bonferroni correction to no correction, as the correlation ranges from 0 to 1.

2.2.2 Likelihood Ratio Test

A more principled frequentist procedure would be the likelihood ratio test (LRT):

Theorem 2.2.3. *The test statistic arising from the likelihood ratio test is*

$$T = \max_j \left[\sqrt{1-\rho}x_j + n\rho \left(\frac{x_j - \bar{\mathbf{x}}}{\sqrt{1-\rho}} \right) \right]^2$$

and the LRT would be to reject the null hypothesis if $T > c$, where c satisfies $\alpha = P(T > c \mid \theta_i = 0 \forall i)$.

Proof. Denote

$$\Sigma_0 = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}$$

and its inverse

$$\Sigma_0^{-1} = \begin{pmatrix} a & b & \cdots & b \\ b & a & \cdots & b \\ \vdots & \vdots & \ddots & \vdots \\ b & b & \cdots & a \end{pmatrix} \text{ where } \begin{cases} a = a_n = \frac{1+(n-2)\rho}{(1+(n-1)\rho)(1-\rho)} \\ b = b_n = \frac{-\rho}{(1+(n-1)\rho)(1-\rho)} \end{cases} .$$

The likelihood ratio is then, letting $f(\cdot)$ denote the density of X and $\tilde{\mathbf{x}}_i = (x_1, \dots, x_{i-1}, x_i - \theta_i, x_{i+1}, \dots, x_n)'$,

$$LR = \frac{f(\mathbf{x} \mid \theta_i = 0, \forall i)}{\max_{i, \theta_i} f(\mathbf{x} \mid \theta_i \neq 0, \theta_{-i} = 0)} = \frac{(\det \Sigma_0)^{-\frac{1}{2}} \exp \left\{ \frac{-1}{2} \mathbf{x}^T \Sigma_0^{-1} \mathbf{x} \right\}}{\max_i (\det \Sigma_0)^{-\frac{1}{2}} \exp \left\{ \frac{-1}{2} \sup_{\theta_i} \tilde{\mathbf{x}}_i^T \Sigma_0^{-1} \tilde{\mathbf{x}}_i \right\}} . \quad (2.3)$$

Computation yields, defining $u_i = \sum_{j \neq i}^n x_j$,

$$\hat{\theta}_i = \arg \max_{\theta_i} \frac{-1}{2} \tilde{\mathbf{x}}_i^T \Sigma_0^{-1} \tilde{\mathbf{x}}_i = x_i + \frac{b}{a} u_i ,$$

from which it is immediate that

$$\begin{aligned}
LR &= \min_i \frac{\exp \left\{ \frac{-1}{2} \left((a-b) \left(\sum_1^n x_j^2 \right) + b \left(\sum_1^n x_j \right)^2 \right) \right\}}{\exp \left\{ \frac{-1}{2} \left((a-b) \left(\frac{b^2}{a^2} u_i^2 + \sum_{j \neq i}^n x_j^2 \right) + b \left(-\frac{b}{a} u_i + \sum_{j \neq i}^n x_j \right)^2 \right) \right\}} \\
&= \min_i \exp \left\{ \frac{-1}{2} \left[(a-b) \left(x_i^2 - \frac{b^2}{a^2} u_i^2 \right) + b \left((u_i + x_i)^2 - u_i^2 \left(\frac{b}{a} - 1 \right)^2 \right) \right] \right\} \\
&\qquad\qquad\qquad \left(\text{since } \sum_1^n x_j = u_i + x_i \right) \\
&= \min_i \exp \left\{ \frac{-1}{2} \left[a x_i^2 + 2 b u_i x_i + \frac{b^2}{a} u_i^2 \right] \right\} \\
&= \min_i \exp \left\{ \frac{-1}{2a} (a x_i + b u_i)^2 \right\}.
\end{aligned}$$

Noting that

$$\begin{aligned}
&\frac{1}{a} (a x_j + b u_j)^2 \\
&= \frac{1}{(1 + (n-1)\rho)(1 + (n-2)\rho)(1 - \rho)} \left((1 + (n-2)\rho) x_j - \rho \sum_{k \neq j} x_k \right)^2 \\
&= \frac{1}{(1 + (n-1)\rho)(1 + (n-2)\rho)(1 - \rho)} \left[(1 - \rho) x_j + n \rho (x_j - \bar{\mathbf{x}}) \right]^2 \\
&= \frac{1}{(1 + (n-1)\rho)(1 + (n-2)\rho)} \left[\sqrt{1 - \rho} x_j + n \rho \left(\frac{x_j - \bar{\mathbf{x}}}{\sqrt{1 - \rho}} \right) \right]^2,
\end{aligned}$$

it is immediate that LR is equivalent to the test statistic T .

The rejection region is $LR \leq k$ for some k , which is clearly equivalent to $T \geq c$ for appropriate critical value c . □

When $\rho = 0$, $T = \max_i x_i^2$, and the LRT reduces to the ad hoc testing procedure in the previous section. On the other hand, as $\rho \rightarrow 1$, $T \approx \max_i n^2 \left(\frac{x_i - \bar{\mathbf{x}}}{\sqrt{1 - \rho}} \right)^2$, which exhibits a quite different behavior that will be discussed later.

2.3 A Bayesian Test

On the Bayesian side, it is convenient to view this as the model selection problem of deciding between the $n + 1$ exclusive models

$$\begin{aligned} M_0 : \theta_1 = \dots = \theta_n = 0 \text{ (null model)} \\ M_i : \theta_i \neq 0, \theta_{(-i)} = \mathbf{0} \end{aligned} \quad (2.4)$$

where $\theta_{(-i)}$ is the vector of all θ_j except θ_i .

The simplest prior assumption computationally is that, for the nonzero θ_i (if there is one),

$$\theta_i \sim N(0, \tau^2);$$

initially we will assume τ^2 to be known, but later will consider it to be unknown.

Then under model M_i , the marginal likelihood of model M_i is

$$\begin{aligned} m_0(\mathbf{x}) &\sim N(\mathbf{0}, \Sigma_0) \\ m_i(\mathbf{x}) &= \int f(\mathbf{x} | \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} \sim N(\mathbf{0}, \Sigma_i), \end{aligned} \quad (2.5)$$

where

$$\Sigma_i = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & 1 + \tau^2 & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}.$$

The posterior probability of M_i (that the i^{th} channel has the signal) is then

$$P(M_i | \mathbf{x}) = \frac{m_i(\mathbf{x})P(M_i)}{\sum_{j=0}^n m_j(\mathbf{x})P(M_j)},$$

where $P(M_j)$ is the prior probability of model M_j .

Theorem 2.3.1. *For any $\rho \in [0, 1)$ and positive integer $n > 1$, the null posterior probability is:*

$$P(M_0 | \mathbf{x}) = \left\{ 1 + \left(\frac{1-r}{nr} \right) \frac{1}{\sqrt{1+\tau^2 a}} \sum_{i=1}^n \exp \left\{ \frac{\tau^2}{2(1+\tau^2 a)} \left(\frac{x_i}{1-\rho} + bn\bar{x} \right)^2 \right\} \right\}^{-1},$$

and the posterior probability of an alternative model M_i is

$$P(M_i | \mathbf{x}) = \left\{ \begin{array}{l} \sqrt{1 + a\tau^2} \left(\frac{nr}{1-r} \right) \exp \left\{ \frac{-\tau^2}{2(1+\tau^2a)} \left(\frac{x_i}{1-\rho} + n\bar{\mathbf{x}}b \right)^2 \right\} \\ + \sum_{k=1}^n \exp \left\{ \frac{-\tau^2}{2(1+\tau^2a)} \left(\frac{x_i^2 - x_k^2}{(1-\rho)^2} + \frac{2b}{1-\rho} (n\bar{\mathbf{x}})(x_i - x_k) \right) \right\} \end{array} \right\}^{-1}.$$

Proof. The posterior probability of M_i is

$$\begin{aligned} P(M_i | \mathbf{x}) &= \frac{m_i(\mathbf{x})P(M_i)}{\sum_{j=0}^n m_j(\mathbf{x})P(M_j)} \\ &= \frac{\frac{1-r}{n} |\Sigma_i|^{-\frac{1}{2}} \exp \left\{ \frac{-1}{2} \mathbf{x}' \Sigma_i^{-1} \mathbf{x} \right\}}{r |\Sigma_0|^{-\frac{1}{2}} \exp \left\{ \frac{-1}{2} \mathbf{x}' \Sigma_0^{-1} \mathbf{x} \right\} + \sum_{j=1}^n \frac{1-r}{n} |\Sigma_j|^{-\frac{1}{2}} \exp \left\{ \frac{-1}{2} \mathbf{x}' \Sigma_j^{-1} \mathbf{x} \right\}} \quad (2.6) \\ &= \left\{ \begin{array}{l} \left(\frac{nr}{1-r} \right) \left| \frac{\Sigma_0}{\Sigma_i} \right|^{-\frac{1}{2}} \exp \left\{ \frac{-1}{2} \mathbf{x}' (\Sigma_0^{-1} - \Sigma_i^{-1}) \mathbf{x} \right\} \\ + 1 + \sum_{j \neq i}^n \exp \left\{ \frac{-1}{2} \mathbf{x}' (\Sigma_j^{-1} - \Sigma_i^{-1}) \mathbf{x} \right\} \end{array} \right\}^{-1}. \end{aligned}$$

The expression can be simplified by further computing $\Sigma_i^{-1}, (\Sigma_i^{-1} - \Sigma_k^{-1})$ and $\det(\Sigma_i)$.

First notice that by the Cholesky decomposition

$$\Sigma_0^{-1} = \begin{pmatrix} a & b & \cdots & b \\ b & a & \cdots & b \\ \vdots & \vdots & \ddots & \vdots \\ b & b & \cdots & a \end{pmatrix} = \mathbf{L}\mathbf{L}^T,$$

for some lower triangular matrix \mathbf{L} . Then by the Woodbury identity, the difference of two inverse matrices can be obtained:

$$\begin{aligned} \Sigma_i^{-1} &= (\Sigma_0 + \boldsymbol{\tau}_i \boldsymbol{\tau}_i^T)^{-1} = \Sigma_0^{-1} - \Sigma_0^{-1} \boldsymbol{\tau}_i (1 + \boldsymbol{\tau}_i^T \Sigma_0^{-1} \boldsymbol{\tau}_i)^{-1} \boldsymbol{\tau}_i^T \Sigma_0^{-1} \\ &= \Sigma_0^{-1} \frac{-\tau^2}{1 + \tau^2 a} \Sigma_0^{-1} \begin{pmatrix} b & \cdots & b & 0 \\ & & & a \\ & & & 0 & b & \cdots & b \end{pmatrix}, \end{aligned}$$

where $\boldsymbol{\tau}_i = (0, \dots, \tau, \dots, 0)^T$ (the i^{th} element is τ^2).

Therefore,

$$\begin{aligned}\mathbf{x}'(\Sigma_0^{-1} - \Sigma_i^{-1})\mathbf{x} &= \frac{\tau^2}{1 + \tau^2 a} \left(x_i(a - b) + bn\bar{\mathbf{x}} \right)^2 = \frac{\tau^2}{1 + \tau^2 a} \left(\frac{x_i}{1 - \rho} + bn\bar{\mathbf{x}} \right)^2 \\ \mathbf{x}'(\Sigma_k^{-1} - \Sigma_i^{-1})\mathbf{x} &= \frac{\tau^2}{1 + \tau^2 a} (x_i^2 - x_k^2)(a - b)^2 + 2b(a - b)(n\bar{\mathbf{x}})(x_i - x_k).\end{aligned}$$

Also the ratio of two determinants is

$$\begin{aligned}\det(\Sigma_i)/\det(\Sigma_0) &= \det(\Sigma_1)/\det(\Sigma_0) = \det(I + \Sigma_0^{-1}\boldsymbol{\tau}_1\boldsymbol{\tau}_1^T) \\ &= \det(I + \mathbf{L}\mathbf{L}^T\boldsymbol{\tau}_1\boldsymbol{\tau}_1^T) \\ &= \det(I + \boldsymbol{\tau}_1^T\mathbf{L}^T\mathbf{L}\boldsymbol{\tau}_1) \\ &= (1 + \tau^2 L_{11}^2) \\ &= (1 + \tau^2 a).\end{aligned}$$

By plugging back these quantities into (2.6), the proof is complete. \square

Remark 2.3.2. (2.24) gives the full expression for $P(M_i | \mathbf{x})$, without using a, b , and is utilized in the subsequent proofs.

Corollary 2.3.3. *In particular, when $n = 2$, the null posterior probability is:*

$$P(M_0 | \mathbf{x}) = \left\{ 1 + \frac{1 - r}{2r} \sqrt{\frac{1 - \rho^2}{1 - \rho^2 + \tau^2}} \sum_{i \in \{1, 2\}} \exp \left\{ \frac{\tau^2}{2(1 - \rho^2 + \tau^2)} \left(\frac{x_i - \rho x_{(-i)}}{\sqrt{1 - \rho^2}} \right)^2 \right\} \right\}^{-1},$$

and the posterior probability of the alternative $M_i, i \in \{1, 2\}$ is:

$$P(M_i | \mathbf{x}) = \left\{ \begin{aligned} &\sqrt{\frac{1 - \rho^2 + \tau^2}{1 - \rho^2}} \left(\frac{2r}{1 - r} \right) \exp \left\{ \frac{-\tau^2}{2(1 - \rho^2 + \tau^2)} \frac{(x_i - \rho x_{(-i)})^2}{1 - \rho^2} \right\} \\ &+ 1 + \exp \left\{ \frac{-\tau^2}{2(1 - \rho^2 + \tau^2)} (x_i^2 - x_{(-i)}^2) \right\} \end{aligned} \right\}^{-1}.$$

2.4 The situation as the correlation goes to 1

The following theorem shows the surprising result that, when the dimension is greater than 2, the Bayesian method can correctly select the true model when the correlation goes

to one. In two dimensions, however, there is nonzero probability of choosing the wrong alternative model if a non-null model is true.

Theorem 2.4.1.

When $n = 2$, $i \in \{1, 2\}$, $\rho \rightarrow 1$, then:

$$P(M_0 | \mathbf{X}) \rightarrow 1 \quad \text{under null model,}$$

$$P(M_i | \mathbf{X}) \rightarrow \left(1 + \exp \left\{ \frac{-1}{2} (X_i^2 - X_{(-i)}^2) \right\} \right)^{-1} \quad \text{under } M_1 \text{ or } M_2.$$

When $n > 2$, $i, j \in \{0, 1, \dots, n\}$, $\rho \rightarrow 1$, under model M_j :

$$P(M_i | \mathbf{X}) \rightarrow \delta_i^j = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{else.} \end{cases}$$

Proof. By Lemma 2.21, x_i can be written as $x_i = \theta_i + \sqrt{\rho}z + \sqrt{1-\rho}z_i$.

Case I: $n=2$:

$$\begin{aligned} \frac{x_i - \rho x_{(-i)}}{\sqrt{1-\rho^2}} &= \frac{(\theta_i - \theta_{(-i)}) + \sqrt{1-\rho}(z_i - \rho z_{(-i)}) + z\sqrt{\rho}(1-\rho)}{\sqrt{1-\rho^2}} \\ &= \frac{\theta_i - \theta_{(-i)}}{\sqrt{1-\rho^2}} + \frac{z_i - \rho z_{(-i)}}{\sqrt{1+\rho}} + \sqrt{1-\rho} \frac{\sqrt{\rho}z}{\sqrt{1+\rho}}. \end{aligned} \tag{2.7}$$

If M_0 is true, since both $\theta_i, \theta_{(-i)}$ are zero, the dominant term of (2.7) is $O(1)$, so the null posterior probability (Corollary 2.3.3) becomes:

$$P(M_0 | \mathbf{x}) = (1 + O(\sqrt{1-\rho^2})) = 1 + o(1).$$

If M_i is true, the dominant term of (2.7) is $O(1/\sqrt{1-\rho^2})$; hence as $\rho \rightarrow 1$:

$$\begin{aligned} P(M_i | \mathbf{x}) &= \left\{ \frac{\sqrt{\frac{\tau^2}{2(1-\rho^2)}} \left(\frac{2r}{1-r} \right) \exp \left\{ \frac{-1}{2} \frac{\theta_i^2}{2(1-\rho^2)} + O\left(\frac{1}{\sqrt{1-\rho}}\right) \right\}}{+1 + \exp \left\{ \frac{-1}{2} \theta_i (\theta_i + 2z) \right\}} \right\}^{-1} \\ &= \left(\frac{1}{1 + \exp \left\{ \frac{-1}{2} \theta_i (\theta_i + 2z) \right\}} \right) (1 + o(1)) \\ &= \frac{1}{1 + \exp \left\{ \frac{-1}{2} (x_i^2 - x_{(-i)}^2) \right\}} (1 + o(1)), \end{aligned}$$

using the fact that $e^{-1/\sqrt{1-\rho}}$ goes to zero faster than $1/\sqrt{1-\rho}$ goes to infinity.

If $M_{(-i)}$ is true: similar to the previous case, (2.7) is $O(1/\sqrt{1-\rho})$, so only the last term in (2.3.3) remains.

Case II: $n > 2$: denote $\mathbf{z} = (z_1, \dots, z_n)$, $\bar{z} = 1/n \sum_1^n z_i$;

If M_0 is true : take $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n) = \mathbf{0}$ in (2.24), let $c = c(\rho) = \frac{1+(n-1)\rho}{(1-\rho+\tau^2)(1+(n-1)\rho)-\tau^2\rho}$, and note that $\lim_{\rho \rightarrow 1} c(\rho) = \frac{n}{\tau^2(n-1)}$. Then

$$\begin{aligned}
& P(M_i | \mathbf{x}) \\
&= \left[\begin{aligned} & \sqrt{\frac{1}{c(1-\rho)}} \left(\frac{nr}{1-r} \right)^* \\ & \exp \left\{ \frac{-\tau^2}{2} c \left((z_i - \bar{z}) + \underbrace{\frac{\sqrt{1-\rho}}{1+(n-1)\rho} (\sqrt{\rho}z + \sqrt{1-\rho}\bar{z})}_{O(\sqrt{1-\rho})} \right)^2 \right\} + 1 + \\ & \sum_{k \neq i}^n \exp \left\{ \frac{-\tau^2}{2} c \left\{ 2 \underbrace{\frac{\frac{(z_i+z_k-2\bar{z})(z_i-z_k)}{1-\rho} + (1-\rho)\bar{z}(z_i-z_k) + \sqrt{\rho(1-\rho)}z(z_i-z_k)}{1+(n-1)\rho}}_{O(\sqrt{1-\rho})} \right\} \right\} \end{aligned} \right]^{-1} \\
&= \left[\begin{aligned} & \sqrt{\frac{\tau^2(n-1)/n}{1-\rho}} \left(\frac{nr}{1-r} \right) \exp \left\{ \frac{-n}{2(n-1)} (z_i - \bar{z})^2 \right\} + 1 + \\ & \sum_{k \neq i}^n \exp \left\{ \frac{-n}{2(n-1)} \frac{(z_i+z_k-\bar{z})(z_i-z_k)}{1-\rho} \right\} \end{aligned} \right]^{-1} (1 + o(1)) \\
&\leq \left[\sqrt{\frac{\tau^2(n-1)/n}{1-\rho}} \left(\frac{nr}{1-r} \right) \exp \left\{ \frac{-n}{2(n-1)} (z_i - \bar{z})^2 \right\} \right]^{-1} (1 + o(1)) \\
&= \sqrt{1-\rho} \sqrt{\frac{1-r}{r\tau^2}} (1 + o(1)) \rightarrow 0.
\end{aligned}$$

If M_j is true ($j > 0$): Take $\theta_k = 0 \forall k \neq j$ in (2.24):

$$\begin{aligned}
& P(M_j | \mathbf{x}) \\
&= \left\{ \begin{aligned} & \sqrt{\frac{1}{c(1-\rho)}} \left(\frac{nr}{1-r} \right) \exp \left\{ \frac{-\tau^2 c}{2(1-\rho)} \left((1-2/n)\theta_j + \sqrt{1-\rho}(z_i - \bar{z}) + O(1-\rho) \right)^2 \right\} \\ & + 1 + \sum_{k \neq j}^n \exp \left\{ \frac{-\tau^2 c}{2} \left(\frac{(1-2/n)\theta_j + \sqrt{1-\rho}(z_j + z_k - 2\bar{z})(\theta_j + \sqrt{1-\rho}(z_i - z_k))}{1-\rho} + O(1) \right) \right\} \end{aligned} \right\}^{-1} \\
&= \left\{ \begin{aligned} & \sqrt{\frac{\frac{n-1}{n}\tau^2}{1-\rho}} \left(\frac{nr}{1-r} \right) \exp \left\{ \frac{-n}{2(n-1)} \left(\frac{(1-2/n)^2 \theta_j^2}{1-\rho} + O\left(\frac{1}{\sqrt{1-\rho}}\right) \right) \right\} + 1 \\ & + (n-1) \exp \left\{ \frac{-n}{2(n-1)} \left(\frac{(1-2/n)\theta_j^2}{1-\rho} + O\left(\frac{1}{\sqrt{1-\rho}}\right) \right) \right\} \end{aligned} \right\}^{-1} (1 + o(1)) \\
&\rightarrow 1 \text{ since } \lim_{\rho \rightarrow 1} \sqrt{1/(1-\rho)} \exp\{-1/(1-\rho)\} = 0.
\end{aligned}$$

□

Theorem 2.4.2. *The likelihood ratio test (Theorem 2.2.3) is fully powered (i.e., rejects the null with probability 1 under an alternative hypothesis) when $\rho \rightarrow 1$ and $n > 2$, but (as with the Bayesian test) is not fully powered when $n = 2$.*

Proof. By Lemma 2.22:

$$\begin{cases} \frac{x_i - \bar{x}}{\sqrt{1-\rho}} = z_i - \bar{z} & \text{under the null model } M_0, \\ \frac{x_i - \bar{x}}{\sqrt{1-\rho}} = \frac{\theta_j(\delta_j^i - 1/n)}{\sqrt{1-\rho}} + z_i - \bar{z} & \text{under an alternative model } M_j. \end{cases}$$

When $n = 2$, under M_i , when $\rho \rightarrow 1$:

$$\begin{aligned}
\lim_{\rho \rightarrow 1} T &= \lim_{\rho \rightarrow 1} \max_{j \in \{1,2\}} \left[\sqrt{1-\rho} x_j + 2\rho \left(\frac{x_j - \bar{x}}{\sqrt{1-\rho}} \right)^2 \right] \\
&= 2 \lim_{\rho \rightarrow 1} \max_{j \in \{1,2\}} \rho \left(\frac{x_j - x_{(-j)}}{2\sqrt{1-\rho}} \right)^2,
\end{aligned}$$

For both $j = i$ or $j = (-i)$, the corresponding likelihood ratios go to infinity at the same asymptotic rate since $[\theta_i/(2\sqrt{1-\rho}) + (z_i - z_{(-i)})/2]^2 = [-\theta_i/(2\sqrt{1-\rho}) - (z_i - z_{(-i)})/2]^2$. Hence, one cannot distinguish whether θ_i or $\theta_{(-i)}$ is nonzero. There is a positive probability choosing an incorrect alternative model.

When $n > 2$, under M_j , when $\rho \rightarrow 1$:

$$\begin{aligned} & \max_i \left[\sqrt{1-\rho} x_i + n\rho \left(\frac{x_i - \bar{x}}{\sqrt{1-\rho}} \right)^2 \right] \\ &= \max_i n \left[\frac{\theta_i - \theta_j/n}{\sqrt{1-\rho}} + z_i - \bar{z} \right]^2 + o(1) \\ &= n \left[\frac{\theta_j(1-1/n)}{\sqrt{1-\rho}} + z_j - \bar{z} \right]^2 + o(1). \end{aligned}$$

In this case, the true alternative model has largest likelihood ratio ($=\infty$), hence, LRT is fully powered. \square

From Theorem 2.4.1 and 2.4.2, when the correlation goes to 1 and the dimension is larger than 2, both the Bayesian procedure and the LRT are fully powered. This surprising behavior as the correlation goes to one can be explained by the following observations.

When $n = 2$, $\rho \rightarrow 1$:

$$x_i - x_j = \begin{cases} 0 & \text{under null model} \\ \theta_i \text{ or } -\theta_j & \text{else.} \end{cases}$$

Hence, one can correctly distinguish the null model if it is true, but can not declare which non-null model is true when $x_i - x_j$ is not 0.

When $n > 2$, $\rho \rightarrow 1$: If all pairs $x_i - x_j$ are zero, then the null model is true. If there are pairs $x_i - x_j$, $x_j - x_i$ that are nonzero, we can further check whether $x_i - x_k$ ($k \neq j$) equals zero or not to see whether θ_i or θ_j is nonzero.

Note that the ad hoc frequentist test does not have this behavior. As $\rho \rightarrow 1$, the test

- still has probability α of incorrectly rejecting a true M_0 ;
- still has positive probability of not detecting a signal when M_i is true.

2.5 Asymptotic frequentist properties of Bayesian procedures

In this section, we will be studying the false positive probability (FPP) theoretically and numerically. We first need to obtain asymptotic posteriors.

2.5.1 Posterior probabilities

Lemma 2.5.1. *As $n \rightarrow \infty$ under the null model,*

$$P(M_i | \mathbf{x}) = \left(1 + \frac{n}{1-r} \sqrt{\frac{1-\rho+\tau^2}{1-\rho}} \exp \left\{ \frac{-\tau^2}{2(1-\rho+\tau^2)} \left(\frac{x_i - \bar{\mathbf{x}}}{\sqrt{1-\rho}} \right)^2 \right\} \right)^{-1} (1 + o(1)) \quad (2.8)$$

almost surely.

Proof. Take $\boldsymbol{\theta} = 0$ in (2.24):

$P(M_i | \mathbf{x})$

$$= \left\{ \begin{aligned} & \sqrt{\frac{1}{c(1-\rho)}} \left(\frac{nr}{1-r} \right) \exp \left\{ \frac{-\tau^2}{2} c \left(\underbrace{z_i - \bar{\mathbf{z}} + \frac{\sqrt{1-\rho}}{1+(n-1)\rho} (\sqrt{\rho}z + \sqrt{1-\rho}\bar{\mathbf{z}})}_I \right)^2 \right\} + 1 + \\ & \sum_{k \neq i}^n \exp \left\{ \frac{-\tau^2}{2} c \left(\underbrace{(z_i + z_k - 2\bar{\mathbf{z}})(z_i - z_k) + 2 \frac{(\sqrt{\rho}z + \sqrt{1-\rho}\bar{\mathbf{z}})(\sqrt{1-\rho}(z_i - z_k))}{1+(n-1)\rho}}_{II} \right) \right\} \end{aligned} \right\}^{-1} \quad (2.9)$$

Without loss of generality, assuming $|z_i| \leq n^{1/2-\epsilon}$ for all i , which holds almost surely by Lemma 2.10.4, asymptotic analysis of (2.9) yields:

$$\begin{aligned} I &= \left((1-1/n)^2 z_i^2 + \underbrace{\bar{\mathbf{z}}_{(-i)}^2}_{O(n^{-1})} - 2(1-1/n) \underbrace{z_i \bar{\mathbf{z}}_{(-i)}}_{O(n^{-\epsilon})} \right) \quad (\text{where } \bar{\mathbf{z}}_{(-i)} = 1/n \sum_{k \neq i} z_k) \\ &+ 2 \underbrace{\left(\frac{\sqrt{1-\rho}}{1+(n-1)\rho} \right)}_{O(n^{-1})} \underbrace{\left((1-1/n)z_i - \bar{\mathbf{z}}_{(-i)} \right)}_{O(n^{1/2-\epsilon})} \underbrace{\left(\sqrt{\rho}z + \sqrt{1-\rho}\bar{\mathbf{z}} \right)}_{O(1)} \\ &+ \underbrace{\left(\frac{\sqrt{1-\rho}}{1+(n-1)\rho} \right)^2}_{O(n^{-2})} \underbrace{\left(\sqrt{\rho}z + \sqrt{1-\rho}\bar{\mathbf{z}} \right)^2}_{O(1)} \\ &= (1-1/n)^2 z_i^2 + O(n^{-\epsilon}). \end{aligned}$$

Therefore,

$$\begin{aligned}
& \sqrt{\frac{1}{c(1-\rho)}} \left(\frac{nr}{1-r} \right) \exp \left\{ \frac{-\tau^2}{2} cI \right\} \\
&= \sqrt{\frac{1}{c(1-\rho)}} \left(\frac{nr}{1-r} \right) \exp \left\{ \frac{-\tau^2}{2} c(1-1/n)^2 z_i^2 \right\} \left(1 + O(n^{-\epsilon}) \right) \\
&= \sqrt{\frac{1-\rho+\tau^2}{1-\rho}} \left(\frac{nr}{1-r} \right) \exp \left\{ \frac{-\tau^2}{2(1-\rho+\tau^2)} z_i^2 \right\} (1 + o(1)).
\end{aligned} \tag{2.10}$$

$$\begin{aligned}
II &= \underbrace{(z_i^2 - z_k^2)}_{O(n^{1-2\epsilon})} (1 - 2/n) - \underbrace{(2/n \sum_{l \notin \{i,k\}} z_l)}_{O(n^{-1/2})} \underbrace{(z_i - z_k)}_{O(n^{1/2-\epsilon})} \\
&+ \underbrace{\left(\frac{2}{1 + (n-1)\rho} \right)}_{O(n^{-1})} \left\{ \begin{array}{l} \sqrt{\rho(1-\rho)} \underbrace{z_i z_k}_{O(n^{1/2-\epsilon})} + (1-\rho) \underbrace{z_i/n(-z_k)}_{O(n^{-1/2-\epsilon})} \\ + (1-\rho) \underbrace{z_i^2/n}_{O(n^{-2\epsilon})} + \underbrace{(\sqrt{\rho}z + \sqrt{1-\rho}\bar{z}_{(-i)})(-z_k)}_{O(n^{1/2-\epsilon})} \end{array} \right\} \\
&= (z_i^2 - z_k^2) (1 - 2/n) + O(n^{-\epsilon}).
\end{aligned}$$

The summation term in (2.9) becomes:

$$\begin{aligned}
& \sum_{k \neq i}^n \exp \left\{ \frac{-\tau^2}{2} cII \right\} \\
&= \exp \left\{ -c \frac{\tau^2}{2} (1 - 2/n) z_i^2 \right\} (1 + O(n^{-\epsilon})) \sum_{k=1}^n \exp \left\{ c \frac{\tau^2}{2} (1 - 2/n) z_k^2 \right\} \\
&= \exp \left\{ \frac{-\tau^2}{2(1-\rho+\tau^2)} z_i^2 \right\} n \left(\mathbb{E}^Z \left[\exp \left\{ \frac{\tau^2}{2(1-\rho+\tau^2)} Z^2 \right\} \right] + o(1) \right) (1 + o(1)) \\
&\quad \text{by Law of Large Number and } Z \sim N(0, 1) \\
&= \exp \left\{ \frac{-\tau^2}{2(1-\rho+\tau^2)} z_i^2 \right\} n \sqrt{\frac{1-\rho+\tau^2}{1-\rho}} (1 + o(1)).
\end{aligned} \tag{2.11}$$

The proof is completed by summing (2.10), (2.11). \square

Remark 2.5.2. Note that, by Lemma 2.22,

$$z_i = z_i(\mathbf{x}) = \frac{x_i - \bar{\mathbf{x}}}{\sqrt{1 - \rho}} + O(1/\sqrt{n}), \quad (2.12)$$

so that Lemma 2.5.1 can be written with respect to z_i :

$$P(M_i | \mathbf{x}) = \left(1 + \frac{n}{1-r} \sqrt{\frac{1-\rho+\tau^2}{1-\rho}} \exp \left\{ \frac{-\tau^2 z_i^2}{2(1-\rho+\tau^2)} \right\} (1+o(1)) \right)^{-1}.$$

Remark 2.5.3. Figure 2.1 shows the ratio of the estimated $P(M_1|\mathbf{x})$ (from Lemma 2.8) and the true probability (from Theorem 2.3.1), as n grows. Each plot contains 200 different ratio curves based on independent simulations with fixed $\rho, P(M_0)$ and τ . As can be seen, the ratio goes to 1 when n grows and the convergence rate indeed depends on the correlation.

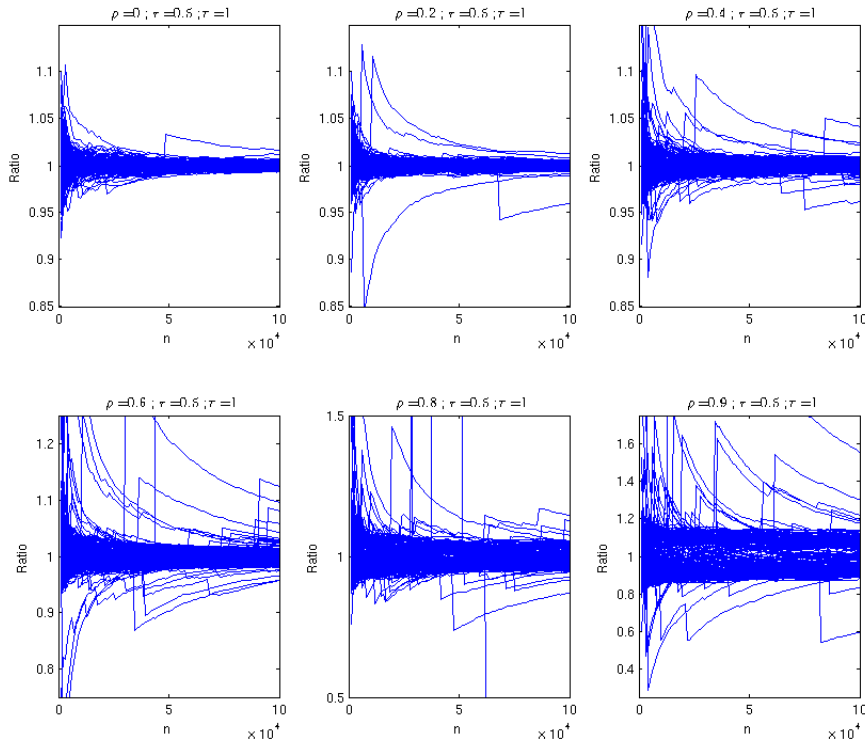


FIGURE 2.1: Ratio of estimated and true posterior probability of M_1 as n grows under the null model and fixed τ, r , different ρ . Each subplot is for different correlations and contains 200 simulations.

Remark 2.5.4. Figure 2.2 gives the estimated and true posterior probability of M_1 under the assumption that the null model is true, for fixed $r = \rho = 0.5$ and n , but varied τ . Notice that, for fixed n , the estimated probability is closer to the true probability when τ is small but is worse for larger τ , indicating that larger n is required for obtaining the same precision as τ grows.

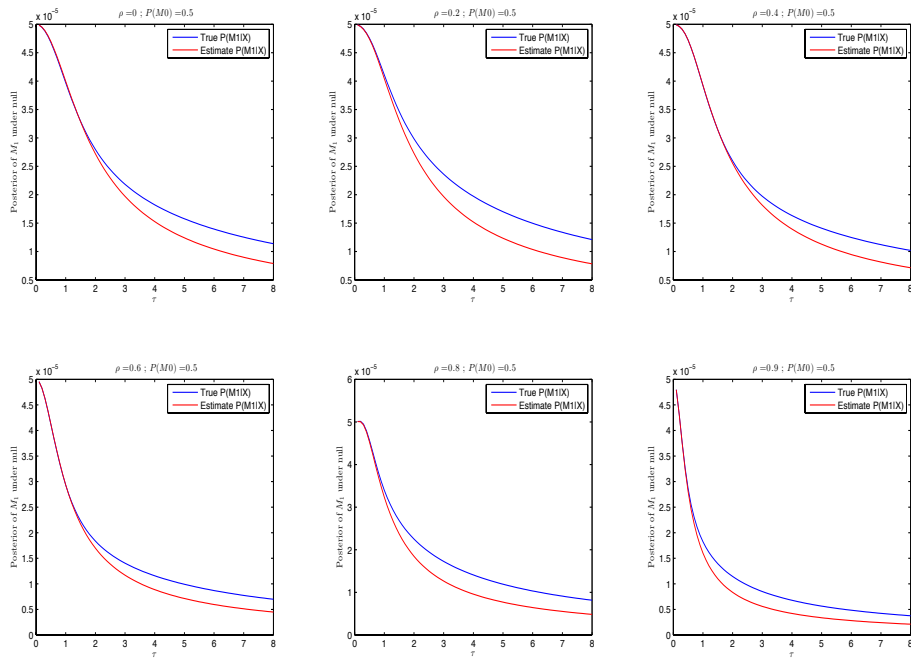


FIGURE 2.2: Estimated (red line) and true posterior probability (blue line) of M_1 for different τ under the null model, for fixed $n = 2000$, $\rho = r = 0.5$.

The following theorem shows the surprising result that, as n grows when the null model is true, the posterior probability of the null model converges to its prior probability. Thus one cannot learn that the null model is true.

Theorem 2.5.5. *As $n \rightarrow \infty$ and $\rho \in [0, 1)$, under the null model,*

$$P(M_0 | \mathbf{X}) \rightarrow P(M_0).$$

Proof. First note that

$$\begin{cases} a_n = \frac{1}{1-\rho} + \frac{-\rho}{(1-\rho)(1+(n-1)\rho)} = \frac{1}{1-\rho} + O(1/n), \\ nb_n = \frac{-1}{1-\rho} + \frac{1-\rho}{(1-\rho)(1+(n-1)\rho)} = \frac{-1}{1-\rho} + O(1/n). \end{cases} \quad (2.13)$$

The summation term in the null posterior (Theorem 2.3.1) becomes

$$\begin{aligned} & \left(\frac{1-r}{nr} \right) \frac{1}{\sqrt{1+\tau^2/(1-\rho)}} \sum_1^n \exp \left\{ \frac{\tau^2}{2(1+\tau^2/(1-\rho))} \left[\frac{x_i - \mathbf{x}}{1-\rho} \right]^2 \right\} (1+o(1)) \\ &= \left(\frac{1-r}{r} \right) 1/n \sqrt{\frac{1-\rho}{1-\rho+\tau^2}} \sum_1^n \exp \left\{ \frac{\tau^2}{2(1-\rho+\tau^2)} z_i^2 \right\} (1+o(1)) \quad (\text{by Lemma 2.22}) \\ &\rightarrow \frac{1-r}{r} \quad (\text{by the Strong Law of Large Numbers}). \end{aligned}$$

Therefore, $P(M_0 | \mathbf{X}) \rightarrow (1 + (1-r)/r)^{-1} = r = P(M_0)$.

□

Remark 2.5.6. Figure 2.3 shows simulations of the null posterior probability for different numbers of hypotheses and different correlations. Interestingly, by Theorem 2.4.1, the Bayes procedure identifies the correct model (here the null model) when n is fixed and the correlation goes to 1, resulting in higher initial posterior probability of the null model for highly correlated cases. On the other hand, by Theorem 2.5.5, this posterior probability converges to its prior probability regardless of the correlation. This convergence can be seen in Figure 2.3.

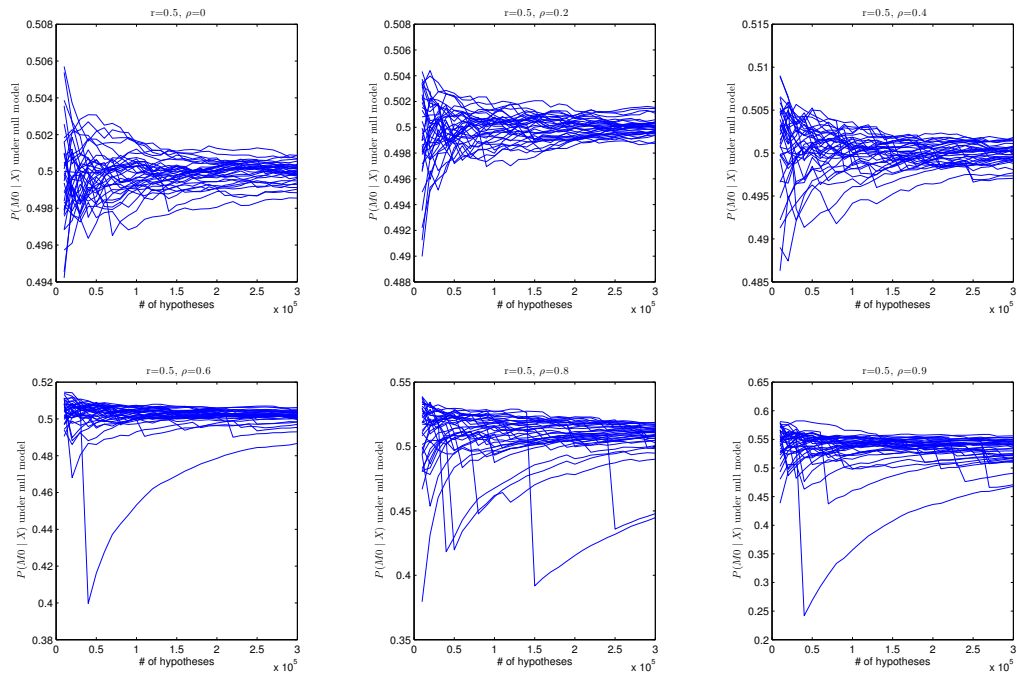


FIGURE 2.3: Convergence of $P(M_0 | \mathbf{x})$ to the prior probability (0.5) under the null model. Each subplot has a different correlation and contains 50 simulations.

2.5.2 False positive probability

Here we focus on the major goal, to find the frequentist false positive probability under the null model of the Bayesian procedure. To begin, we must formally define the Bayesian procedure for detecting a signal.

Definition 2.5.7 (Bayesian detection criterion). Accept model M_i if its posterior probability $P(M_i | \mathbf{x})$, is greater than a specified threshold $p \in (0, 1)$. If multiple models pass this threshold, choose the one with largest posterior probability.

Definition 2.5.8 (False positive probability, FPP). Under the null model, the FPP is the frequentist probability of accepting a non-null model.

Theorem 2.5.9 (False positive probability). *Under the null model, as $n \rightarrow \infty$,*

$$P(\text{false positive} | r, \rho, \tau^2) = O(n^{-\frac{1-\rho}{\tau^2}} (\log n)^{-1/2}).$$

Proof. Under the null model, by (2.8), $P(M_i | \mathbf{x}) \geq p$ is equivalent to

$$\begin{aligned} z_i^2 &\geq 2 \left(\frac{1-\rho+\tau^2}{\tau^2} \right) \ln \left(\frac{n}{1-r} \frac{p}{1-p} \sqrt{\frac{1-\rho+\tau^2}{1-\rho}} (1+o(1)) \right) \\ &= 2 \left(\frac{1-\rho+\tau^2}{\tau^2} \right) \ln \left(\frac{n}{1-r} \frac{p}{1-p} \sqrt{\frac{1-\rho+\tau^2}{1-\rho}} \right) + o(1). \end{aligned} \tag{2.14}$$

By Fact 5.4.1:

$$\begin{aligned} \mathbb{P} \left(|Z_i| \geq \underbrace{\sqrt{2 \left(\frac{1-\rho+\tau^2}{\tau^2} \right) \ln \left(\frac{n}{1-r} \frac{p}{1-p} \sqrt{\frac{1-\rho+\tau^2}{1-\rho}} \right) + o(1)}}_{\gamma_n} \right) \\ = 1/n \left(\frac{\frac{2}{\sqrt{2\pi}} n^{-\frac{1-\rho}{\tau^2}} \left(\frac{1}{1-r} \frac{p}{1-p} \sqrt{\frac{1-\rho+\tau^2}{1-\rho}} \right)^{-(1+\frac{1-\rho}{\tau^2})} (1+o(1))}{\underbrace{\sqrt{2 \left(\frac{1-\rho+\tau^2}{\tau^2} \right) \ln \left(\frac{n}{1-r} \frac{p}{1-p} \sqrt{\frac{1-\rho+\tau^2}{1-\rho}} \right) + o(1)}}_{d_n}} \right) + O \left(\frac{1}{n (\log n)^2} \right). \end{aligned}$$

$$\begin{aligned} \mathbb{P}(\text{any false positive} | M_0) &= 1 - \prod_i^n \mathbb{P}(|Z_i| < \gamma_n) \\ &= 1 - (1 - \mathbb{P}(|Z_1| \geq \gamma_n))^n = 1 - \left(1 - \frac{d_n}{n} \right)^n = 1 - (1 - d_n) + O(d_n^2) \\ &= \left(\frac{1}{n^{\frac{1-\rho}{\tau^2}}} \right) \left(\frac{|\tau|}{\sqrt{\pi}(1-\rho+\tau^2)^{1+\frac{1-\rho}{2\tau^2}}} \right) \left(\frac{(1-r)(1-p)}{p} \right)^{1+\frac{1-\rho}{\tau^2}} \\ &\quad \left(\log \frac{n}{1-r} + \log \left(\frac{p}{1-p} \sqrt{\frac{1-\rho+\tau^2}{1-\rho}} \right) \right)^{\frac{-1}{2}} (1+o(1)) \\ &= O(n^{-(\frac{1-\rho}{\tau^2})} (\log n)^{\frac{-1}{2}}). \end{aligned}$$

□

This is a surprising and unsettling result: a standard Bayesian procedure yields a false positive probability that goes to zero at a polynomial rate. This is much too strong error probability control from a frequentist perspective; that it happens on the Bayesian side is

surely indication that assuming we know τ^2 is too strong an assumption. Hence we turn to a more flexible approach in the next section.

2.6 A Type II maximum likelihood approach

The type II maximum likelihood approach to choice of the prior under the alternative model replaces a pre-specified τ^2 with that prior variance which maximizes the marginal likelihood over all possible τ^2 ; see Berger (1985) for discussion of this approach.

Lemma 2.6.1. *Let $\tilde{L}_n(\tau^2)$ be the marginal likelihood of τ^2 given (x_1, \dots, x_n) :*

$$\tilde{L}_n(\tau^2) = \sum_{i=0}^n P(M_i) m_i(\mathbf{x} \mid \tau^2),$$

from which it follows that

$$\arg \max_{\tau^2} \tilde{L}_n(\tau^2) = \arg \max_{\tau^2} \frac{1}{\sqrt{1 + \tau^2 a}} \frac{1}{n} \sum_{i=1}^n \exp \left\{ \frac{\tau^2}{2(1 + \tau^2 a)} \left(\frac{x_i}{1 - \rho} + bn\bar{\mathbf{x}} \right)^2 \right\}.$$

Proof.

$$\begin{aligned} \tilde{L}_n &= r |\Sigma_0|^{-\frac{1}{2}} \exp \left\{ \frac{-1}{2} \mathbf{x}' \Sigma_0^{-1} \mathbf{x} \right\} + \frac{1-r}{n} |\Sigma_1|^{-\frac{1}{2}} \sum_{i=1}^n \exp \left\{ \frac{-1}{2} \mathbf{x}' \Sigma_i^{-1} \mathbf{x} \right\} \\ &= r |\Sigma_0|^{-\frac{1}{2}} \exp \left\{ \frac{-1}{2} \mathbf{x}' \Sigma_0^{-1} \mathbf{x} \right\} + \frac{1-r}{n} |\Sigma_0|^{-\frac{1}{2}} (1 + \tau^2 a)^{-\frac{1}{2}} \\ &\quad * \exp \left\{ \frac{-1}{2} \mathbf{x}' \Sigma_0^{-1} \mathbf{x} \right\} \sum_{i=1}^n \exp \left\{ \frac{\tau^2}{2(1 + \tau^2 a)} \left(x_i(a - b) + bn\bar{\mathbf{x}} \right)^2 \right\} \\ &= |\Sigma_0|^{-\frac{1}{2}} \exp \left\{ \frac{-1}{2} \mathbf{x}' \Sigma_0^{-1} \mathbf{x} \right\} * \left\{ \begin{aligned} &r + \frac{1-r}{n} (1 + \tau^2 a)^{-\frac{1}{2}} \\ &* \sum_{i=1}^n \exp \left\{ \frac{\tau^2}{2(1 + \tau^2 a)} \left(x_i(a - b) + bn\bar{\mathbf{x}} \right)^2 \right\} \end{aligned} \right\}. \end{aligned}$$

Note that a, b, Σ_0 and $\mathbf{x}' \Sigma_0^{-1} \mathbf{x}$ are independent of τ^2 and the result follows. □

For simplicity, denote $L_n = \frac{1}{n\sqrt{1 + \tau^2 a}} \sum_{i=1}^n \exp \left\{ \frac{\tau^2}{2(1 + \tau^2 a)} \left(\frac{x_i}{1 - \rho} + bn\bar{\mathbf{x}} \right)^2 \right\}$, then

$$\begin{aligned} \hat{\tau}_n^2 &= \arg \max_{\tau^2} \tilde{L}_n(\tau^2) = \arg \max_{\tau^2} L_n(\tau^2) \\ &= \arg \max_{\tau^2} \sqrt{\frac{1 - \rho}{1 - \rho + \tau_n^2}} \frac{1}{n} \sum_{i=1}^n \exp \left\{ \frac{\hat{\tau}_n^2 z_i^2}{2(1 - \rho + \tau_n^2)} \right\} (1 + o(1)) \text{ by (2.25)}. \end{aligned}$$

The following theorem gives the Type II ML estimator and the corresponding FPP:

Theorem 2.6.2 (Type II ML false positive probability). *Given null model prior probability r , correlation ρ , and decision threshold p , as $n \rightarrow \infty$:*

$$\mathbb{P}(\text{false positive} \mid \text{null model}, \hat{\tau}_n^2) = \frac{1}{\log n} \left(\frac{1}{k^*} - \frac{1}{2} \right) (1 + o(1)),$$

where k^* satisfies:

$$-2 \log \left(\sqrt{\pi} \left(\frac{1}{k^*} - \frac{1}{2} \right) \right) = \log k^* + 2 \log \left(\frac{p}{(1-p)(1-r)} \right) + 2 \left(\frac{1}{k^*} \right). \quad (2.15)$$

Proof. Without loss of generality, assume $\max_i z_i^2 = z_1^2$. By the model selection criteria

(2.14), z_1 is a false positive if:

$$\begin{aligned} z_1^2 &\geq 2 \left(1 + \frac{1-\rho}{\hat{\tau}_n^2} \right) \log \left(n \frac{p}{(1-p)(1-r)} \sqrt{\frac{1-\rho + \hat{\tau}_n^2}{1-\rho}} \right) + o(1) \\ &= 2 \log n + \log \left(\frac{1-\rho + \hat{\tau}_n^2}{1-\rho} \right) + \underbrace{2 \log \left(\frac{p}{(1-p)(1-r)} \right)}_u \\ &\quad + 2(1-\rho) \frac{\log n}{\hat{\tau}_n^2} + \frac{1-\rho}{\hat{\tau}_n^2} \log \left(\frac{1-\rho + \hat{\tau}_n^2}{1-\rho} \right) + \frac{(1-\rho)}{\hat{\tau}_n^2} u + o(1), \\ \lim_{n \rightarrow \infty} L_n(\hat{\tau}_n^2) &= \lim_{n \rightarrow \infty} \left(\underbrace{\sqrt{\frac{1-\rho}{1-\rho + \hat{\tau}_n^2}} 1/n \exp \left\{ \frac{\hat{\tau}_n^2 z_1^2}{2(1-\rho + \hat{\tau}_n^2)} \right\}}_I + \right. \\ &\quad \left. \underbrace{\sqrt{\frac{1-\rho}{1-\rho + \hat{\tau}_n^2}} 1/n \sum_{i=2}^n \exp \left\{ \frac{\hat{\tau}_n^2 z_i^2}{2(1-\rho + \hat{\tau}_n^2)} \right\}}_{II} \right) (1 + o(1)). \end{aligned} \quad (2.16)$$

If $\frac{\hat{\tau}_n^2}{\log n} \rightarrow 0$, then $L_n(\hat{\tau}_n^2) \rightarrow 1$ since

$$z_1^2 = 2 \log n + 2(1-\rho)/\hat{\tau}_n^2 \log n + \log \hat{\tau}_n^2 + u + o(1),$$

$$I = \frac{1}{n\sqrt{\hat{\tau}_n^2}} \exp \left\{ \frac{\hat{\tau}_n^2 + 1 - \rho}{1 - \rho + \hat{\tau}_n^2} \log n + \frac{\hat{\tau}_n^2 \log \hat{\tau}_n^2}{2(1 - \rho + \hat{\tau}_n^2)} + \frac{\hat{\tau}_n^2 u}{2(1 - \rho + \hat{\tau}_n^2)} \right\} (1 + o(1))$$

$$= O((\hat{\tau}_n^2)^{-(1-\rho)/2/(1-\rho+\hat{\tau}_n^2)}) = o(1) \text{ since}$$

$$\log(\hat{\tau}_n^2)^{-\frac{1-\rho}{2(1-\rho+\hat{\tau}_n^2)}} = \frac{-(1-\rho)}{2(1-\rho+\hat{\tau}_n^2)} \log \hat{\tau}_n^2 \rightarrow 0.$$

$II \rightarrow 1$ by Corollary 2.10.7.

If $\frac{\hat{\tau}_n^2}{\log n} \rightarrow \infty$, $L_n(\hat{\tau}_n^2) \rightarrow e^{u/2}$ since

$$z_1^2 = 2 \log n + \log \hat{\tau}_n^2 + u + o(1),$$

$$I = (\hat{\tau}_n^2)^{-1/2}/n \exp \left\{ \frac{\hat{\tau}_n^2 \log n}{1 - \rho + \hat{\tau}_n^2} + \frac{\hat{\tau}_n^2 \log \hat{\tau}_n^2}{2(1 - \rho + \hat{\tau}_n^2)} + \frac{\hat{\tau}_n^2 u}{2(1 - \rho + \hat{\tau}_n^2)} \right\} (1 + o(1))$$

$$= e^{\frac{u\hat{\tau}_n^2}{2(1-\rho+\hat{\tau}_n^2)}} (\hat{\tau}_n^2)^{-\frac{1-\rho}{2(1-\rho+\hat{\tau}_n^2)}} n^{-\frac{1-\rho}{2(1-\rho+\hat{\tau}_n^2)}}$$

$$= e^{u/2} + o(1) \text{ since}$$

$$\begin{cases} \log(\hat{\tau}_n^2)^{-\frac{1-\rho}{2(1-\rho+\hat{\tau}_n^2)}} = \frac{-(1-\rho)}{2(1-\rho+\hat{\tau}_n^2)} \log \hat{\tau}_n^2 \rightarrow 0, \\ \log(n^{-(1-\rho)/(1-\rho+\hat{\tau}_n^2)}) = -\frac{1-\rho}{1-\rho+\hat{\tau}_n^2} \log n \rightarrow 0. \end{cases}$$

$II \rightarrow 0$ by Corollary 2.10.7.

There exists an $\epsilon > 0$ and $k \in (0, \infty)$ such that $\lim_{n \rightarrow \infty} L_n((1 - \rho)k \log n) > \max\{1, e^{u/2}\} + \epsilon$

since: $\tilde{z}^2 = 2 \log n + \log \log n + \log k + u + 1/k + o(1)$ by (2.16),

$$I = \frac{1}{n\sqrt{k \log n}} \exp \left\{ \frac{k \log n}{1 + k \log n} \left(\log n + \frac{\log \log n}{2} + \frac{u}{2} + \frac{1}{k} \right) \right\} (1 + o(1))$$

$$= n^{-1/(1+k \log n)} \sqrt{k \log n}^{-1/(1+k \log n)} e^{u/2} e^{1/k} (1 + o(1))$$

$$\rightarrow e^{u/2} \text{ since } \begin{cases} \log((k \log n)^{-1/(2(1+k \log n))}) = \frac{-\log \log n}{2 \log n} \rightarrow 0, \\ \log(n^{-1/(1+k \log n)}) = \frac{-\log n}{1+k \log n} \rightarrow -1/k. \end{cases}$$

$II \rightarrow 2\Phi(\sqrt{2/k}) - 1$ by Corollary 2.10.7.

Choose $\epsilon = 0.5 \min\{1, e^{u/2}\}$. Because (1) $I + II \rightarrow e^{u/2} + 2\Phi(\sqrt{2/k}) - 1$, (2)

$\lim_{k \rightarrow 0} 2\Phi(\sqrt{2/k}) - 1 = 1$, there exists k such that $I + II > \max\{1, e^{u/2}\} + \epsilon$. Also k must

be bounded away from zero and infinity otherwise it violates conclusions from the previous two cases. Hence $\hat{\tau}_n^2 \in (k_1 \log n, k_2 \log n)$, where $0 < k_1 < k_2 < \infty$.

To find the exact rejection region, we solve for the critical value in (2.16), defined by exact equality holding; denote this critical value by \tilde{z}^2 . Then

$$\tilde{z}^2 = 2 \log n + \log \log n + c \quad (0 < c < \infty).$$

For each c , by Lemma 2.10.9,

$$\begin{cases} \hat{\tau}_n^2 &= (1 - \rho)k(c) \log n(1 + o(1)), \\ k(c) &= \left(\frac{1}{2} + \frac{1}{\sqrt{\pi}} \exp\{-\frac{c}{2}\}\right)^{-1}. \end{cases}$$

maximizes L_n . With such $\hat{\tau}_n^2$, by (2.16), $\tilde{z}^2 = 2 \log n + \log \log n + (\log k + 2 \log u + 2/k)$, c also needs to satisfy $c = \log k + 2 \log u + 2/k$. Therefore, the boundary region is defined, if it exists, by

$$\begin{aligned} c^* &= \log k^* + 2 \log u + 2/k^* \\ &= -\log\left(\frac{1}{2} + \frac{\exp\{-c^*/2\}}{\sqrt{\pi}}\right) + 2 \log u + 1 + 2\frac{\exp\{-c^*/2\}}{\sqrt{\pi}}, \end{aligned} \tag{2.17}$$

where c^*, k^* are the c, k when the boundary condition holds. Since

$$c^*(k^*) = -2 \log(\sqrt{\pi}(1/k^* - 1/2)),$$

(2.15) \Leftrightarrow (2.17). With the above exact rates $\hat{\tau}_n^2, z_1^2$, one can compute the FPP similar to Theorem 2.5.9:

$P(\text{false positive} \mid \hat{\tau}_n^2, p, r)$

$$\begin{aligned}
&= 1 - \left\{ 1 - 2 \left\{ \frac{\frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} 2 \left(1 + \frac{1-\rho}{\hat{\tau}^2} \right) \log \left(\frac{n}{1-r} \frac{p}{1-p} \sqrt{\frac{1-\rho+\hat{\tau}^2}{1-\rho}} \right) + o(1) \right\}}{\sqrt{2 \left(1 + \frac{1-\rho}{\hat{\tau}^2} \right) \log \left(\frac{np}{(1-r)(1-p)} \sqrt{\frac{1-\rho+\hat{\tau}^2}{1-\rho}} \right) + o(1)}} + o \left(\frac{1}{n(\log n)^2} \right) \right\} \right\}^n \\
&= 1 - \left\{ 1 - \frac{\left(\frac{np}{(1-p)(1-r)} \sqrt{1 + k^* \log n} \right)^{-\left(1 + \frac{1}{k^* \log n} \right)} (1 + o(1))}{\sqrt{\pi \left(1 + \frac{1}{k^* \log n} \right) \log \left(\frac{np}{(1-r)(1-p)} \sqrt{1 + k^* \log n} \right) + o(1)}} \right\}^n \\
&= 1 - \left\{ 1 - \frac{1}{n} \frac{\left(\frac{(1-p)(1-r)}{p} \right) \left(\frac{pn}{(1-p)(1-r)} \left(1 + k^* \log n \right) \right)^{\frac{-1}{1+k^* \log n}}}{\underbrace{\sqrt{\pi \left(k^* \log n + 2 + \frac{1}{k^* \log n} \right) \log \left(\frac{np}{(1-p)(1-r)} \sqrt{1 + k^* \log n} \right)}}_{d_n}} (1 + o(1)) \right\}^n \\
&= d_n (1 + o(1)) + O \left(\frac{1}{(\log n)^2} \right) \\
&= \frac{1}{\sqrt{\pi k^*}} \exp \left\{ \frac{-1}{k^*} \right\} \left(\frac{(1-p)(1-r)}{p} \right) \frac{1}{\log n} (1 + o(1)) \\
&= \frac{1}{\sqrt{\pi k^*}} \sqrt{\pi k^*} \left(\frac{1}{k^*} - \frac{1}{2} \right) \frac{p}{(1-p)(1-r)} \left(\frac{(1-p)(1-r)}{p} \right) \frac{1}{\log n} (1 + o(1)) \quad \text{by (2.17)} \\
&= \frac{1}{\log n} \left(\frac{1}{k^*} - \frac{1}{2} \right) (1 + o(1)).
\end{aligned}$$

□

Note that (2.15) can be solved numerically. When $p = 1/3; r = 0.5$, $k^* \approx 1.37$ and $c^* \approx 1.78$. Solution of $\frac{1}{k^*} - \frac{1}{2}$ with respect to $\frac{p}{(1-r)(1-p)}$ will be given in Figure 2.6.

The Type II ML FPP converges to 0 at a logarithmic rate in n , which is much less conservative than the Bayesian procedure with fixed τ^2 . In addition, the asymptotic FPP does not depend on ρ .

Remark 2.6.3. Figure 2.4 provides the simulated (red curve) and theoretical (in blue) false positive probability (FPP) with respect to the number of hypotheses (denoted by n). As expected, the simulated results match the theoretical prediction, the rate of convergence being around $1/\sqrt{\log n}$. Note that the FPP does not become extremely small even for very large n .

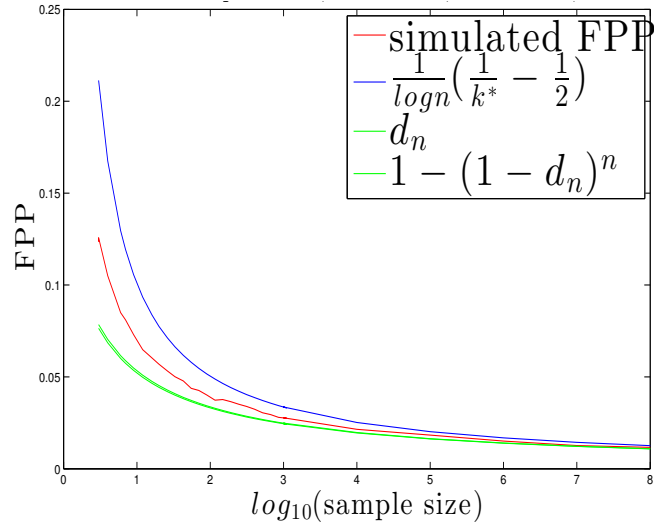


FIGURE 2.4: Comparison of the simulated FPP and its asymptotic approximation when $n = 10^6$, $p = 1/3$; $r = 0.5$, $\rho = 0$ as n varies from 10^1 to 10^8 (x-axis in exponential scale)

Remark 2.6.4. Figure 2.5 demonstrates how the threshold p (Definition 2.5.7) varies as a function of the sample size, given that the false positive probability (FPP) is a fixed number (0.05). Due to the multiplicity correction (2.8), the posterior probability of each non-null model is penalized by $1/n$, resulting in higher posterior probability of the null model and lower FPP. Therefore, to reach the same FPP, a higher acceptance rate or lower threshold probability is needed, which can be seen from the downward trend in this figure.

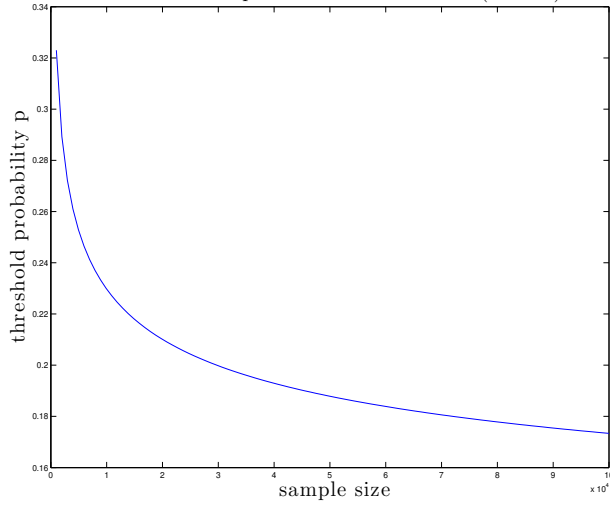


FIGURE 2.5: threshold probability (p) versus the number of hypotheses (n) for fixed FPP(= 0.05) and prior probability of null model (= 0.5).

Remark 2.6.5. Figure 2.6 shows the value of $\frac{1}{k^*} - \frac{1}{2}$ for different $\frac{p}{(1-r)(1-p)}$. For fixed dimension n , larger threshold p or prior probability of null model r has smaller false positive probability.

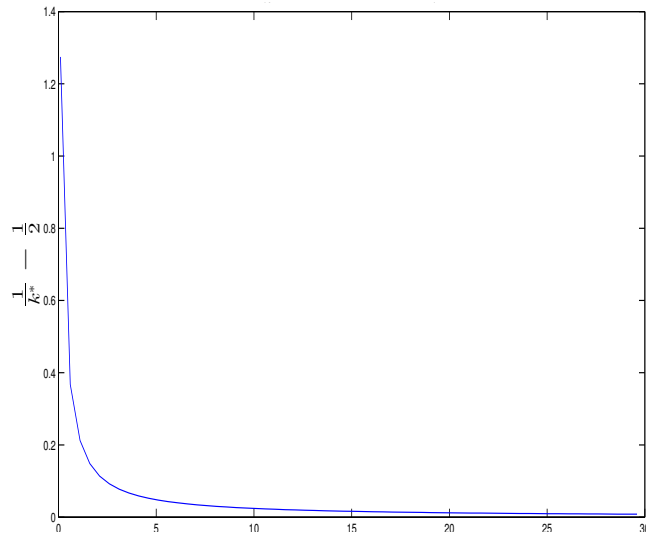


FIGURE 2.6: Solution of $\frac{1}{k^*} - \frac{1}{2}$ (y-axis) versus $\frac{p}{(1-r)(1-p)}$ (x-axis).

2.7 Power analysis for the Type II ML Bayesian procedure

The FPP in the previous section was computed under the assumption that the null model is true. For power analysis, we assume the alternative model M_i is true, so that θ_i is nonzero. Figure 2.7 shows the numerical solution for the Type II ML $\hat{\tau}^2$, as a function of θ , for the specified scenario.

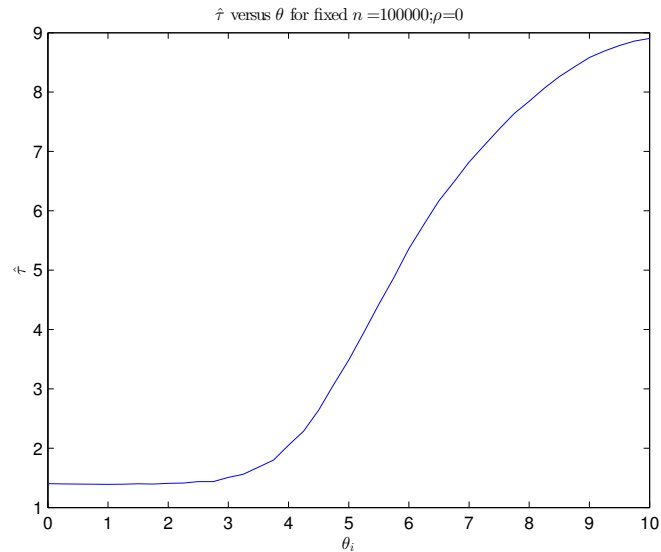


FIGURE 2.7: $\hat{\tau}^2$ versus θ_i and fixed $n = 10^4$, $r = 0.5$, $\rho = 0$, and $p = 1/3$. Each point is the average $\hat{\tau}^2$ among 10^4 independent draws from the multivariate normal with the constraint $\tau^2 > 1$.

Figure 2.8 demonstrates the how the detection power (computed numerically) varies when the signal size increases, for the specified situation.

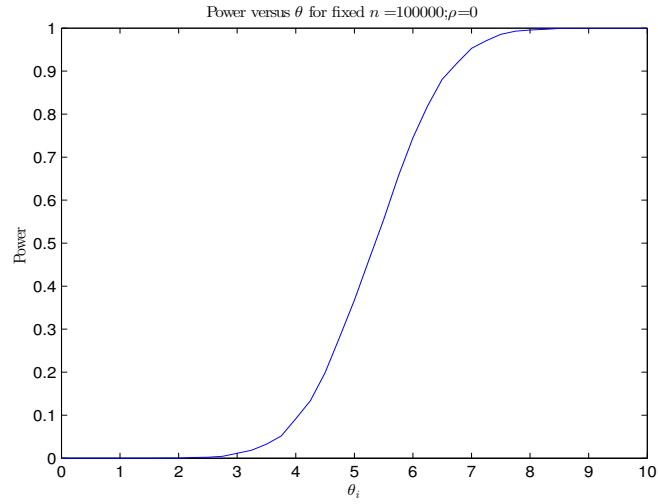


FIGURE 2.8: Power versus θ_i and fixed $n = 10^4$, $r = 0.5$, $\rho = 0$, $p = 1/3$. Each point is the average acceptance rate of the true non-null model with respect to θ in the x-axis range.

Figure 2.9 shows that, when the number of hypotheses grows, the size of the signal (θ_i) must also increase to guarantee the same power.

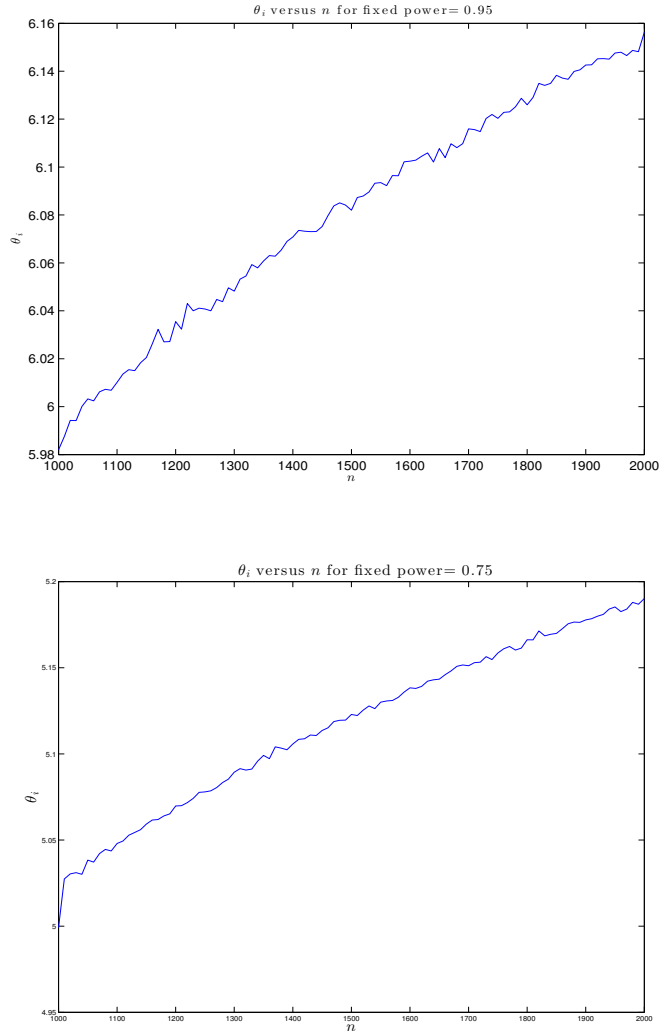


FIGURE 2.9: θ_i versus n for fixed power 0.95 (above) 0.75 (below), $\rho = 0$, $p = 1/3$.

2.8 Analysis as the information grows

In this section, we generalize model (2.4) to the scenario where each channel has m i.i.d observations. Then the sample mean satisfies

$$\bar{\mathbf{X}} \sim \text{multinorm} \left(\begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{pmatrix}, \frac{1}{m} \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} \right). \quad (2.18)$$

Hence, m can be seen as the precision of $\bar{\mathbf{X}}$. More generally, we will replace $1/m$ by a general function σ_n^2 , where σ_n^2 decreases to zero as n grows. The theorem below gives the rate of decrease which guarantees consistency.

Theorem 2.8.1. *Consider model (2.4) with the following covariate matrix:*

$$\mathbf{X} \sim \text{multinorm} \left(\begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{pmatrix}, \sigma_n^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} \right), \quad (2.19)$$

When $\sigma_n^2 \log n \rightarrow 0$, consistency holds for both the null and alternative models.

When $\sigma_n^2 \log n \rightarrow d \in (0, \infty)$,

$$\text{Under } M_0: P(M_0 | \mathbf{X}) \rightarrow \left(1 + \frac{1-r}{r} \left[2\Phi\left(\frac{(1-\rho)d}{\tau^2}\right) - 1\right]\right)^{-1}.$$

Under an alternative model M_j , if $d \in (0, \frac{\theta_j^2}{2(1-\rho)})$, consistency holds for M_j , whereas consistency does not hold otherwise.

When $\sigma_n^2 \log n = o(\log n)$ and $\sigma_n^2 \log n \rightarrow \infty$, consistency does not hold for any model. In addition, when the null hypothesis is true,

$$P(M_0 | \mathbf{X}) \rightarrow P(M_0).$$

Proof. Write $\sigma_n^2 = d_n/\log n$, $X_i^* = X_i/\sigma_n$, and $\theta_i^* = \theta_i/\sigma_n$. Then a nonzero θ_i^* has prior $N(0, \tau^2/\sigma_n^2)$ and the generalized model 2.19 becomes:

$$\mathbf{X}^* \sim \text{multinorm} \left(\begin{pmatrix} \theta_1^* \\ \theta_2^* \\ \vdots \\ \theta_n^* \end{pmatrix}, \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} \right).$$

Under the null model: by Theorem 2.3.1,

$$\begin{aligned}
& P(M_0 | \mathbf{x})^{-1} \\
&= 1 + \left(\frac{1-r}{nr} \right) \sqrt{\frac{\sigma_n^2}{\sigma_n^2 + \tau^2 a_n} *} \\
&\quad \sum_{i=1}^n \exp \left\{ \frac{\tau^2}{2(\sigma_n^2 + \tau^2 a_n)} \left(\frac{z_i}{\sqrt{1-\rho}} + \frac{\sqrt{\rho}z}{1-\rho} + b_n n \sqrt{1-\rho} \bar{z} + b_n n \sqrt{\rho} z \right)^2 \right\} \\
&= 1 + \left(\frac{1-r}{nr} \right) \sqrt{\frac{\sigma_n^2}{\sigma_n^2 + \tau^2/(1-\rho)}} \sum_{i=1}^n \exp \left\{ \frac{\tau^2}{2(\sigma_n^2 + \tau^2/(1-\rho))} \frac{z_i^2}{(1-\rho)} \right\} \left(1 + O\left(\frac{1}{\sqrt{n}}\right) \right) \\
&= 1 + \left(\frac{1-r}{nr} \right) \sqrt{1-c_n} \sum_{i=1}^n \exp \left\{ \frac{c_n}{2} z_i^2 \right\} \left(1 + O(1/\sqrt{n}) \right),
\end{aligned} \tag{2.20}$$

where

$$\begin{cases} a_n = \frac{1}{1-\rho} + \frac{-\rho}{(1-\rho)(1+(n-1)\rho)} = \frac{1}{1-\rho} + O(1/n), \\ nb_n = \frac{-1}{1-\rho} + \frac{1-\rho}{(1-\rho)(1+(n-1)\rho)} = \frac{-1}{1-\rho} + O(1/n), \\ c_n = \frac{\tau^2}{(1-\rho)\sigma_n^2 + \tau^2} = \frac{\tau^2}{(1-\rho)d_n/\log n + \tau^2}. \end{cases}$$

Since $d_n = o(\log n)$, $1 - c_n = o(1)$, and $0 < c_n < 1$, one can apply Theorem 2.10.6 to get the asymptotic analysis of (2.20):

$$\begin{aligned}
& 2\Phi \left(\frac{2(1-c_n)}{c_n} \log \frac{n}{\sqrt{1-c_n}} \right) - 1 \\
&= \frac{(1-\rho)\sigma_n^2}{\tau^2} \log \left(n \sqrt{\frac{(1-\rho)\sigma_n^2 + \tau^2}{(1-\rho)\sigma_n^2}} \right) \\
&= \frac{(1-\rho)d_n}{\tau^2 \log n} \left[\log \left(n \sqrt{\frac{\log n}{d_n}} \right) + O(1) \right] \\
&= \frac{1-\rho}{\tau^2} \left[d_n + \frac{1}{2} \left(\frac{d_n}{\log n} \right) \log \left(\frac{\log n}{d_n} \right) \right] (1 + O(1)) \\
&\rightarrow \begin{cases} \infty & \text{if } d_n \rightarrow \infty, \\ \frac{1-\rho}{\tau^2} d & \text{if } d_n \rightarrow d, \\ 0 & \text{if } d_n \rightarrow 0. \end{cases}
\end{aligned}$$

Hence, under the null hypothesis,

$$P(M_0 | \mathbf{X}) \rightarrow \begin{cases} P(M_0) & \text{if } d_n \rightarrow \infty, \\ (1 + (\frac{1-r}{\tau})(2\Phi(\frac{1-\rho}{\tau^2}d) - 1))^{-1} & \text{if } d_n \rightarrow d, \\ 1 & \text{if } d_n \rightarrow 0. \end{cases}$$

Under the alternative model M_j : by Theorem 2.3.1,

$$\begin{aligned} P(M_j | \mathbf{x})^{-1} &= \sqrt{1 + \frac{a_n \tau^2}{\sigma_n^2} \left(\frac{n r}{1-r} \right)} \exp \left\{ \frac{-\tau^2}{2(\sigma_n^2 + a_n \tau^2)} \left[\frac{\theta_j^2}{\sigma_n^2 (1-\rho)^2} + O\left(\frac{1}{\sigma_n}\right) \right] \right\} \\ &+ 1 + \sum_{k \neq j}^n \exp \left\{ \frac{-\tau^2}{2(\sigma_n^2 + a_n \tau^2)} \left[\frac{\theta_j^2}{2(1-\rho)^2 \sigma_n^2} + O\left(\frac{1}{\sigma_n}\right) \right] \right\}, \end{aligned}$$

the first term:

$$\sqrt{1 + \frac{\tau^2/(1-\rho)}{\sigma_n^2} \left(\frac{n r}{1-r} \right)} \exp \left\{ \frac{-\tau^2}{2(\sigma_n^2 + \tau^2/(1-\rho))} \left[\frac{\theta_j^2}{(1-\rho)^2 \sigma_n^2} \right] \right\} \left(1 + O\left(\sqrt{\frac{\log n}{d_n}}\right) \right)$$

$$= \sqrt{\tau^2/(1-\rho)} \sqrt{\frac{\log n}{d_n}} \left(\frac{r}{1-r} \right) n^{1 - \frac{\theta_j^2}{2(1-\rho)d_n}} \left(1 + O\left(\sqrt{\frac{\log n}{d_n}}\right) \right)$$

$$\rightarrow 0 \text{ if and only if } \lim_{n \rightarrow \infty} d_n < \frac{\theta_j^2}{2(1-\rho)}.$$

And the last term:

$$\sum_{k \neq j}^n \exp \left\{ \frac{-\tau^2}{2(\sigma_n^2 + \tau^2/(1-\rho))} \left[\frac{\theta_j^2}{(1-\rho)^2 \sigma_n^2} \right] \right\} \left(1 + O\left(\sqrt{\frac{\log n}{d_n}}\right) \right)$$

$$= n \exp \left\{ -\frac{\theta_j^2}{2(1-\rho)} \frac{\log n}{d_n} \right\} \left(1 + O\left(\sqrt{\frac{\log n}{d_n}}\right) \right)$$

$$= n^{1 - \frac{\theta_j^2}{2(1-\rho)d_n}} \left(1 + O\left(\sqrt{\frac{\log n}{d_n}}\right) \right)$$

$$\rightarrow 0 \text{ when } \lim_{n \rightarrow \infty} d_n < \frac{\theta_j^2}{2(1-\rho)}.$$

□

Interestingly, Theorem 2.5.5 can be seen as the limit case of Theorem 2.19 when $\sigma_n^2 = 1$. And this result is true in the more general model selection problem.

2.9 Conclusions

We have studied the comparison of Bayesian and frequentist multiplicity correction under a scenario of data dependence, where there is (at most) one true hypothesis, but the number of potential hypotheses grows. We first discussed the surprising result that, as the correlation $\rho \rightarrow 1$, the likelihood ratio test and the Bayesian procedure have similar performance, both choosing the correct model with certainty (when $n \geq 2$). In addition, when $n \rightarrow \infty$, the Bayes with fixed τ^2 has false positive probability converging to 0 at a polynomial (logarithmic) rate in 0.

For the Type II ML procedure and the default $p = r = 0.5$ and typical large n (but not extreme n), the FPP ranges between 0.02 and 0.04. The suggestion from this example is that Bayesian and Type II ML procedures have very strong control of FWER while having excellent power for detecting a signal, even with highly dependent data.

2.10 Appendix

2.10.1 Normal Theory

Lemma 2.10.1.

$$\mathbf{X} \sim \text{multinorm} \left(\begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{pmatrix}, \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} \right)$$

is equivalent to

$$X_i = \theta_i + \sqrt{\rho}Z + \sqrt{1-\rho}Z_i \quad \forall i \in \{1, 2, \dots, n\}, \quad (2.21)$$

where $Z, Z_1, \dots, Z_n \sim \text{iid } N(0, 1)$. Furthermore, if $\theta_j = 0 \quad \forall j$, then when $n \rightarrow \infty$,

$$\begin{cases} \frac{\bar{\mathbf{x}}}{\sqrt{\rho}} = z + O\left(\frac{1}{\sqrt{n}}\right) \\ \frac{x_i - \bar{\mathbf{x}}}{\sqrt{1-\rho}} = z_i + O\left(\frac{1}{\sqrt{n}}\right) \end{cases} \quad (2.22)$$

Proof. Expectation and covariance of (2.21) can be derived straightforwardly: $\mathbb{E}(X_i) = \theta_i$

and

$$\mathbb{E}(X_i) = \theta_i$$

$$Cov(X_i, X_j) = \begin{cases} 1 & \text{if } i = j \\ \rho & \text{if } i \neq j \end{cases}$$

(2.22) can be obtained accordingly by definition and central limit theorem. \square

Fact 2.10.2 (Normal tail probability). Let $\Phi(t)$ be cumulative distribution function of standard normal, then

$$\frac{t \frac{1}{\sqrt{2\pi}} e^{-t^2/2}}{t^2 + 1} \leq 1 - \Phi(t) \leq \frac{1}{\sqrt{2\pi}} \frac{e^{-t^2/2}}{t}$$

$$1 - \Phi(t) = \frac{1}{\sqrt{2\pi}} \frac{e^{-t^2/2}}{t} + O\left(\frac{e^{-t^2/2}}{t^3}\right)$$

Proof can be found in Durrett (2010).

Expand a, b in Theorem 2.3.1, one can get the following explicit form of posteriors:

Corollary 2.10.3. *The posterior of any non-null model M_i is:*

$$P(M_i | \mathbf{x}) = \left\{ \begin{aligned} & \sqrt{\frac{(1-\rho+\tau^2)(1+(n-1)\rho)-\tau^2\rho}{(1+(n-1)\rho)(1-\rho)}} \left(\frac{n r}{1-r}\right) * \\ & \exp\left\{\frac{-\tau^2}{2} \frac{1+(n-1)\rho}{[(1-\rho+\tau^2)(1+(n-1)\rho)-\tau^2\rho](1-\rho)} \left((x_i - \bar{\mathbf{x}}) + \frac{(1-\rho)\bar{\mathbf{x}}}{1+(n-1)\rho}\right)^2\right\} \\ & + \sum_{k=1}^n \exp\left[\frac{-\tau^2}{2} \frac{1+(n-1)\rho}{[(1-\rho+\tau^2)(1+(n-1)\rho)-\tau^2\rho](1-\rho)} \left((x_i + x_k - 2\bar{\mathbf{x}})(x_i - x_k) + 2\frac{\bar{\mathbf{x}}(x_i - x_k)(1-\rho)}{1+(n-1)\rho}\right)\right] \end{aligned} \right\}^{-1} \quad (2.23)$$

And in terms of $z(x)$, $z_i = z_i(\mathbf{x})$:

$$P(M_i | \mathbf{x}) = \left\{ \begin{aligned} & \sqrt{\frac{(1-\rho+\tau^2)(1+(n-1)\rho)-\tau^2\rho}{(1+(n-1)\rho)(1-\rho)}} \left(\frac{n r}{1-r}\right) * \\ & \exp\left[\frac{-\tau^2}{2} \frac{1+(n-1)\rho}{[(1-\rho+\tau^2)(1+(n-1)\rho)-\tau^2\rho](1-\rho)} \left(\theta_i - \bar{\theta} + \sqrt{1-\rho}(z_i - \bar{z}) + (\bar{\theta} + \sqrt{\rho}z + \sqrt{1-\rho}\bar{z})\right)^2\right] \\ & + \sum_{k=1}^n \exp\left[\frac{-\tau^2}{2} \frac{1+(n-1)\rho}{[(1-\rho+\tau^2)(1+(n-1)\rho)-\tau^2\rho]} \right. \\ & \quad \left. \left(\frac{[\theta_i + \theta_k - 2\bar{\theta} + \sqrt{1-\rho}(z_i + z_k - 2\bar{z})][\theta_i - \theta_k + \sqrt{1-\rho}(z_i - z_k)]}{1-\rho} + 2\frac{[\bar{\theta} + \sqrt{\rho}z + \sqrt{1-\rho}\bar{z}][\theta_i - \theta_k + \sqrt{1-\rho}(z_i - z_k)]}{1+(n-1)\rho}\right)\right] \end{aligned} \right\}^{-1} \quad (2.24)$$

Lemma 2.10.4. Suppose $i \in \{1, 2, \dots, n\}$, Z_i are i.i.d. standard normals, then

$$|Z_i| \leq n^{1/2-\epsilon} \quad \forall i \quad \text{holds almost surely.}$$

Proof. By Fact 5.4.1:

$$\begin{aligned} & P(\text{ for all } i, |Z_i| \leq n^{1/2-\epsilon}) \\ &= \left(1 - P(|Z_1| \geq n^{1/2-\epsilon}) \right)^n \\ &= \left(1 - 2 \frac{\frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}n^{1-2\epsilon}\}}{n^{1/2-\epsilon}} + O\left(\exp\left\{ \frac{-n^{1-2\epsilon}}{2} \right\} \right) \right)^n \\ &= \left(1 - \frac{\frac{2n^{1/2-\epsilon}}{\sqrt{2\pi}} \exp\{-\frac{1}{2}n^{1-2\epsilon}\}}{n} + o(n^{-2}) \right)^n \\ &= 1 - O\left(2n^{1/2+\epsilon} \exp\{-\frac{1}{2}n^{1-2\epsilon}\} \right) \\ &= 1 + o(1). \end{aligned}$$

□

2.10.2 Type II MLE

Fact 2.10.5 (Weak law for triangular arrays). For each n , let $X_{n,i}$, $1 \leq k \leq n$ be independent. Let $\beta_n > 0$ with $\beta_n \rightarrow \infty$ and let $\bar{x}_{n,k} = X_{n,k} 1_{\{|X_{n,k}| \leq \beta_n\}}$. Suppose that as $n \rightarrow \infty$:

$$\sum_{k=1}^n P(|X_{n,k}| > \beta_n) \rightarrow 0 \text{ and } 1/\beta_n^2 \sum_{k=1}^n E\bar{X}_{n,k}^2 \rightarrow 0. \text{ then}$$

$$\frac{(S_n - \alpha_n)}{\beta_n} \rightarrow 0 \text{ in probability}$$

$$\text{where } S_n = X_{n,1} + \dots + X_{n,n} \text{ and } \alpha_n = \sum_{k=1}^n E\bar{X}_{n,k}.$$

See Durrett (2010) for proof.

Theorem 2.10.6. *If $c_n \in (0, 1) \forall n$, $1 - c_n = o(1)$, then*

$$\lim_{n \rightarrow \infty} 1/n \sqrt{1 - c_n} \sum_{i=1}^n \exp \left\{ \frac{c_n}{2} z_i^2 \right\} = \lim_{n \rightarrow \infty} 2\Phi \left(\sqrt{\frac{2(1 - c_n)}{c_n} \log \frac{n}{\sqrt{1 - c_n}}} \right) - 1$$

in probability.

Proof. Take $X_{n,i} = \exp \left\{ \frac{c_n}{2} z_i^2 \right\}$; $\beta_n = \frac{n}{\sqrt{1 - c_n}}$ in Fact 3.5.2.

The first assumption:

$$\begin{aligned} P(|X_{n,i}| > \beta_n) &= P\left(|z_i| > \sqrt{\frac{2}{c_n} \log \frac{n}{\sqrt{1 - c_n}}}\right) \\ &= 2 \frac{\frac{1}{2\pi} \exp \left\{ \frac{-1}{2} \frac{2}{c_n} \log \frac{n}{\sqrt{1 - c_n}} \right\}}{\sqrt{\frac{2}{c_n} \log \frac{n}{\sqrt{1 - c_n}}}} + O\left(\frac{\left(\frac{n}{\sqrt{1 - c_n}}\right)^{-\frac{1}{c_n}}}{\left(\frac{1}{c_n} \log \frac{n}{\sqrt{1 - c_n}}\right)^3}\right) \\ &= \frac{1}{\sqrt{\pi}} \frac{\sqrt{1 - c_n}^{\frac{1}{c_n}}}{n^{\frac{1}{c_n}}} \frac{1}{\sqrt{\log \frac{n}{\sqrt{1 - c_n}}}} (1 + o(1)) \\ &< \frac{1}{\sqrt{\pi}} \frac{\sqrt{1 - c_n}^{\frac{1}{c_n}}}{n^{\frac{1}{c_n}}} \frac{1}{\sqrt{\log n}} (1 + o(1)). \end{aligned}$$

Therefore,

$$\begin{aligned} \sum_{i=1}^n P(|X_{n,k}| > \beta_n) &= nP(|X_{n,k}| > \beta_n) \\ &< n^{1 - \frac{1}{c_n}} (1 - c_n)^{\frac{1}{2c_n}} \frac{1}{\sqrt{\log n}} \\ &= n^{-\frac{1 - c_n}{c_n}} (1 - c_n)^{\frac{1}{2c_n}} \frac{1}{\sqrt{\log n}} \rightarrow 0 \end{aligned}$$

The second assumption: since $\lim_{n \rightarrow \infty} c_n \rightarrow 1$, without loss of generality, assume $c_n > 3/4$.

$$\begin{aligned}
& \frac{1}{\beta_n^2} \sum_{k=1}^n E \bar{X}_{n,k}^2 \\
&= \frac{1-c_n}{n^2} n \int_{|z| < \sqrt{\frac{2}{c_n} \log \frac{n}{\sqrt{1-c_n}}}} \exp\{c_n z^2\} \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{-1}{2} z^2\right\} dz \\
&= \frac{1-c_n}{n} \frac{1}{\sqrt{2\pi}} \left\{ \int_{1 < |z| < \sqrt{\frac{2}{c_n} \log \frac{n}{\sqrt{1-c_n}}}} \exp\left\{\left(c_n - \frac{1}{2}\right) z^2\right\} dz + \int_{|z| < 1} \exp\left\{\left(c_n - \frac{1}{2}\right) z^2\right\} dz \right\} \\
&\leq \frac{1-c_n}{n} \frac{1}{\sqrt{2\pi}} \left\{ \int_{1 < |z| < \sqrt{\frac{2}{c_n} \log \frac{n}{\sqrt{1-c_n}}}} z \exp\left\{\left(c_n - \frac{1}{2}\right) z^2\right\} dz + d \right\} \\
&= \frac{1-c_n}{n} \frac{1}{\sqrt{2\pi}} \left\{ \frac{1}{2\left(c_n - \frac{1}{2}\right)} \exp\left\{\left(c_n - \frac{1}{2}\right) \left(\frac{2}{c_n}\right) \log \frac{n}{\sqrt{1-c_n}}}\right\} + d' \right\} \\
&= \frac{1}{\sqrt{2\pi}} \left(\frac{1}{2c_n - 1}\right) \frac{1-c_n}{n} \left(\frac{n}{\sqrt{1-c_n}}\right)^{2-\frac{1}{c_n}} + o(1) \\
&= \frac{1}{\sqrt{2\pi}} \underbrace{\left(\frac{1}{2c_n - 1}\right)}_{\leq 2} n^{1-\frac{1}{c_n}} (1-c_n)^{\frac{1}{2c_n}} + o(1) \\
&\leq \frac{2}{\sqrt{2\pi}} n^{-\frac{1-c_n}{c_n}} (1-c_n)^{\frac{1}{2c_n}} + o(1) \\
&= o(1).
\end{aligned}$$

Therefore, both assumptions hold. The limit is:

$$\begin{aligned}
\frac{\sqrt{1-c_n}}{n}\alpha_n &= \frac{1-c_n}{n} \sum_{i=1}^n E\bar{X}_{n,i} \\
&= (1-c_n) \int_{|z| < \sqrt{\frac{2}{c_n} \log \frac{n}{\sqrt{1-c_n}}}} e^{\frac{c_n z^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \\
&= 2 \left(\Phi \left(\frac{\sqrt{\frac{2}{c_n} \log \frac{n}{\sqrt{1-c_n}}}}{\sqrt{\frac{1}{1-c_n}}} \right) - \frac{1}{2} \right),
\end{aligned}$$

by Fact 3.5.2:

$$\frac{S_n - \alpha_n}{\beta_n} = \frac{\sum_{i=1}^n e^{c_n z_i^2} - \alpha_n}{\frac{n}{\sqrt{1-c_n}}} = \frac{\sqrt{1-c_n} \sum_{i=1}^n e^{c_n z_i^2}}{n} - \frac{\sqrt{1-c_n}}{n} \alpha_n \rightarrow 0.$$

in probability. And the result follows. \square

Corollary 2.10.7. Let $c_n = \frac{\hat{\tau}_n^2}{1-\rho+\hat{\tau}_n^2}$, then

$$\frac{1}{n\sqrt{1-c_n}} \sum_{i=1}^n \exp \left\{ \frac{c_n}{2} z_i^2 \right\} \rightarrow \begin{cases} 1 & \text{if } \frac{\log n}{\hat{\tau}_n^2} \rightarrow \infty, \\ 2\Phi \left(\sqrt{\frac{2}{k}} \right) - 1 & \text{if } \frac{\log n}{\hat{\tau}_n^2} \rightarrow \frac{1}{(1-\rho)k}, \\ 0 & \text{if } \frac{\log n}{\hat{\tau}_n^2} \rightarrow 0 \end{cases}$$

in probability.

Proof. By Theorem 2.10.6:

Case I: $\frac{\log n}{\hat{\tau}_n^2} \rightarrow \infty$

$$\frac{\sqrt{1-c_n}}{n}\alpha_n \Rightarrow 2\Phi \left(\sqrt{\frac{2(1-\rho)}{\hat{\tau}_n^2} \log \left(n \sqrt{\frac{1-\rho+\hat{\tau}_n^2}{1-\rho}} \right)} \right) - 1 \rightarrow 1.$$

Case II: $\frac{\log n}{\hat{\tau}_n^2} \rightarrow \frac{1}{(1-\rho)k}$

$$\frac{\sqrt{1-c_n}}{n}\alpha_n \Rightarrow 2\Phi \left(\sqrt{\frac{2(1-\rho)}{\hat{\tau}_n^2} \log \left(n \sqrt{\frac{1-\rho+\hat{\tau}_n^2}{1-\rho}} \right)} \right) - 1 \rightarrow 2\Phi \left(\sqrt{\frac{2}{k}} \right) - 1.$$

Case III: $\frac{\log n}{\hat{\tau}_n^2} \rightarrow 0$

$$\frac{\sqrt{1-c_n}}{n} \alpha_n \Rightarrow 2\Phi \left(\sqrt{\frac{2(1-\rho)}{\hat{\tau}_n^2} \log \left(n \sqrt{\frac{1-\rho+\hat{\tau}_n^2}{1-\rho}} \right)} \right) - 1 \rightarrow 0.$$

□

Lemma 2.10.8.

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n\sqrt{1+\tau_n^2 a}} \sum_{i=1}^n \exp \left\{ \frac{\tau_n^2}{2(1+\tau_n^2 a)} \left(\frac{x_i}{1-\rho} + bn\bar{x} \right)^2 \right\} \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sqrt{\frac{1-\rho}{1-\rho+\tau_n^2}} \sum_{i=1}^n \exp \left\{ \frac{\tau_n^2 z_i^2}{2(1-\rho+\tau_n^2)} \right\} \quad a.s. \end{aligned} \quad (2.25)$$

Proof. Expand the coefficients:

$$\begin{aligned} \frac{1}{1+\tau_n^2 a} &= \left(1 + \frac{\tau_n^2(1+(n-2)\rho)}{(1+(n-1)\rho)(1-\rho)} \right)^{-1} \\ &= \frac{1-\rho}{1-\rho+\tau_n^2 \left(1 + \frac{-\rho}{1+(n-1)\rho} \right)} = \frac{1-\rho}{1-\rho+\tau_n^2} (1+o(1)), \end{aligned}$$

and

$$\begin{aligned} \left(\frac{x_i}{1-\rho} + bn\bar{x} \right)^2 &= \frac{1}{(1-\rho)^2} \left(x_i + \frac{-\rho n\bar{x}}{1+(n-1)\rho} \right)^2 \\ &= \frac{1}{(1-\rho)^2} \left(x_i - \bar{x} \left(1 - \frac{1-\rho}{1-\rho+\rho n} \right) \right)^2 \\ &= \frac{1}{(1-\rho)^2} \left(\underbrace{\sqrt{1-\rho} z_i + \sqrt{\rho} z}_{O(1/n)} + \underbrace{\sqrt{1-\rho} \bar{z}}_{O(1/\sqrt{n})} \left(-1 + \frac{1-\rho}{1-\rho+\rho n} \right) \right)^2 \\ &= \frac{z_i^2}{1-\rho} + O((\log n)/\sqrt{n}). \end{aligned}$$

Therefore,

$$\begin{aligned} & \frac{1}{\sqrt{1+\tau_n^2 a}} \frac{1}{n} \sum_i \exp \left\{ \frac{\tau_n^2}{2(1+\tau_n^2 a)} \left(\frac{x_i}{1-\rho} + bn\bar{x} \right)^2 \right\} \\ &= \sqrt{\frac{1-\rho}{1-\rho+\tau_n^2+o(1)}} \frac{1}{n} \sum_i \exp \left\{ \frac{\tau_n^2}{2} \left[\frac{z_i^2}{1-\rho+\tau_n^2} + o(1) \right] \right\}. \end{aligned}$$

□

Lemma 2.10.9 (Type II ML estimator). *Under null model, suppose*

$$\max_j \left(\frac{x_j - \bar{\mathbf{x}}}{\sqrt{1 - \rho}} \right)^2 = 2 \log(n) + \log \log(n) + c,$$

then

$$\arg \max_{\hat{\tau}_n^2 \in \{k' \log n, k' \in \mathbb{R}^+\}} L(\hat{\tau}_n^2) = (1 - \rho) k(c) \log n,$$

where $k(c) = \left(\frac{1}{2} + \frac{1}{\sqrt{\pi}} \exp\left\{-\frac{c}{2}\right\}\right)^{-1}$.

Proof. Without loss of generality, $\max |z_i| = |z_1|$. By (2.25),

$$\begin{aligned} \max_{\tau_n^2} L(\tau_n^2) &= \max_{\tau_n^2} \sqrt{\frac{1 - \rho}{1 - \rho + \tau_n^2}} \frac{1}{n} \sum_i \exp \left\{ \frac{\tau_n^2 z_i^2}{2(1 - \rho + \tau_n^2)} \right\} (1 + o(1)) \\ &= \max_{c_n} \frac{\sqrt{1 - c_n}}{n} \sum_i \exp \left\{ \frac{c_n z_i^2}{2} \right\} (1 + o(1)) \\ &= \max_{c_n} \left\{ \underbrace{1/n \sqrt{1 - c_n} \exp \left\{ \frac{c_n z_1^2}{2} \right\}}_{O(1)} + 2\Phi \left(\sqrt{2 \left(\frac{1}{c_n} - 1 \right) \log \left(\frac{n}{\sqrt{1 - c_n}} \right)} \right) - 1 \right\} (1 + o(1)), \end{aligned}$$

the last equality holds since Corollary 2.10.7. Next step is finding critical points of the above function, define f, g, h as:

$$f(c_n) = \underbrace{1/n \sqrt{1 - c_n} e^{\frac{c_n}{2} z_1^2}}_g + \underbrace{2\Phi \left(\sqrt{2 \left(\frac{1}{c_n} - 1 \right) \log \left(\frac{n}{\sqrt{1 - c_n}} \right)} \right) - 1}_h,$$

since $c_n = c(\tau_n) = \frac{\tau_n^2}{1 - \rho + \tau_n^2}$ is monotone, solving (2.25) is equivalent to find $\arg \max_{c_n} f(c_n)$:

$$\sqrt{1-c_n} \frac{dg}{dc_n} = \frac{1}{2} \left(2 \frac{1-\rho}{k} - 1 \right) \left(\log n \right)^{\frac{1}{2}} \exp \left\{ \frac{c}{2} \right\} \exp \left\{ -\frac{1-\rho}{k} \right\} (1 + o(1))$$

$$\begin{aligned} \sqrt{1-c_n} \frac{dh}{dc_n} &= -\frac{1}{\sqrt{\pi}} \exp \left\{ -\frac{1-\rho}{k \log n} \log \left(n \sqrt{\frac{1-\rho+k \log n}{1-\rho}} \right) \right\} \\ &\quad * \sqrt{\log \left(n \sqrt{\frac{1-\rho+k \log n}{1-\rho}} \right)} (1 + o(1)) \\ &= \frac{-1}{\sqrt{\pi}} \exp \left\{ -\frac{1-\rho}{k} \right\} \sqrt{\log n} (1 + o(1)) \end{aligned}$$

$\frac{df}{dc_n} = 0$ if and only if k satisfies:

$$\begin{aligned} \frac{1}{2} \left(2 \frac{1-\rho}{k} - 1 \right) \left(\log n \right)^{\frac{1}{2}} \exp \left\{ \frac{c}{2} \right\} &= \frac{1}{\sqrt{\pi}} \sqrt{\log n} (1 + o(1)), \\ \left(\frac{2(1-\rho)}{k} - 1 \right) &= \frac{2}{\sqrt{\pi}} \exp \left\{ \frac{-c}{2} \right\} (1 + o(1)). \end{aligned}$$

Therefore,

$$k(c) = \frac{(1-\rho)}{\frac{1}{2} + \frac{1}{\sqrt{\pi}} \exp \left\{ -\frac{c}{2} \right\}} (1 + o(1)).$$

□

Frequentist multiplicity control of Bayesian model selection with spike-and-slab priors

Much of scientific research requires testing thousands to millions of hypotheses simultaneously, from genomewide studies (Storey and Tibshirani (2003)) to pharmaceutical research (Dmitrienko et al. (2009)). The goal is to find true signals amongst multiple hypotheses, but also control the occurrence of false positives, including controlling family-wise error rate (FWER) (Benjamini and Hochberg (1995)), false discovery rate (FDR) (Efron et al. (2001), Efron (2004)), and positive false discovery rate (pFDR) (Storey (2003)), etc..

There also has been increased interest in the Bayesian approach to handling multiple testing, including Cui and George (2008), Guindani et al. (2009), Bogdan et al. (2008a). It has been shown that Bayes and empirical Bayes methods control multiplicity through specification of the prior probabilities of hypotheses (Scott and Berger (2006), Scott and Berger (2010)). However, exact false positive rates for Bayesian decision-theoretic approaches have yet to be found.

In this chapter, we consider the standard Bayesian approach to multiple hypothesis testing, involving spike-and-slab priors, focusing on the frequentist performance of the Bayesian procedures. In particular, we study the frequentist false positive rate (under the

null model of no effect) of the Bayesian procedure, asymptotically and through simulation.

In variable selection, one would like to include variables which are good predictors for responses; each variable is either selected or not. Hence, an intuitive prior in this scenario is the spike-slab prior (George and McCulloch (1993), Chipman et al. (2001), Clyde et al. (2011)), which includes a point mass at zero (indicating not selected) and a continuous normal distribution for nonzero variables. The corresponding posterior has good interpretability by having posterior probability on the point mass (chance of not selected) and elsewhere.

3.1 The standard Bayesian model of multiple testing

Let

$$X_i \sim N(\mu_i, 1) \quad \text{for } i \in \{1, \dots, m\}, \quad (3.1)$$

where $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)$ are unknown a priori with the spike-slab prior:

$$\mu_i \stackrel{i.i.d.}{\sim} p\delta_0 + (1-p)N(0, \tau^2), \quad (3.2)$$

where δ_0 is a point mass at zero, $p = P(\mu_i = 0)$. We focus on the testing question of which of the μ_i are zero, and approach this through the Bayesian model selection approach.

Following the usual convention, we use $\boldsymbol{\gamma}$ to index models:

$$\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_m) \in \{0, 1\}^m,$$

$$\gamma_i = \begin{cases} 0, & \text{if } \mu_i = 0, \\ 1, & \text{if } \mu_i \neq 0; \end{cases}$$

thus $M_{\boldsymbol{\gamma}}$ represents the model with the indicated zero and nonzero means, i.e.

$$M_{\boldsymbol{\gamma}} = \{ \mathbf{X} \sim N(\boldsymbol{\mu}, I_n) \text{ where } \mu_i \sim (1 - \gamma_i)\delta_0 + \gamma_i N(0, \tau^2) \}, \quad (3.3)$$

with I_n being the n by n identity matrix.

Model size or model complexity of $M_{\boldsymbol{\gamma}}$ is defined as $|\boldsymbol{\gamma}| = \sum_{i=1}^m \mathbb{1}_{\gamma_i \neq 0}$. We consider models having size up to k_{max} (between 0 and m) and k_{max} is allowed to grow with m .

The corresponding model space is:

$$\mathcal{M}_{k_{max}} = \{M_\gamma \mid \gamma \in \{0, 1\}^m, |\gamma| \leq k_{max}\}.$$

Given $p_{(0)}, p_{(1)}, \dots, p_{(k_{max})}$ satisfying $p_{(k)} \geq 0 \forall k$ and $\sum_{k=0}^{k_{max}} p_{(k)} = 1$, one can assign model priors by:

$$P(M_\gamma) = p_{(|\gamma|)} \frac{1}{\binom{m}{|\gamma|}}. \quad (3.4)$$

In other words, models having the same complexity k share the weight $p_{(k)}$ equally.

This prior generalizes the following intuitive prior from the Beta distribution:

$$p = P(\gamma_i \neq 0) \sim Be(1, 1) \quad \forall 1 \leq i \leq m,$$

$$P(M_\gamma) = \int_0^1 p(M_\gamma \mid p) \pi(p) dp = \frac{1}{(k_{max} + 1)} \frac{1}{\binom{m}{|\gamma|}}, \quad (3.5)$$

which is the special case of (3.4) when $p_{(0)} = p_{(1)} = \dots = p_{(k_{max})} = \frac{1}{k_{max} + 1}$.

Before we dig into the asymptotic analysis, first consider the explicit form of the Bayes factor and the posterior model probabilities.

3.1.1 Bayes factors

The marginal likelihood of M_γ is:

$$m_\gamma(\mathbf{x}) = \int f(\mathbf{x} \mid \boldsymbol{\mu}) p(\boldsymbol{\mu} \mid \gamma, \tau) d\boldsymbol{\mu} \quad (f \text{ is the likelihood function of } \mathbf{x})$$

$$= \prod_{i:\gamma_i \neq 0} \frac{\exp\{-x_i^2/(2(1+\tau^2))\}}{\sqrt{2\pi(1+\tau^2)}} \prod_{i:\gamma_i=0} \frac{\exp\{-x_i^2/2\}}{\sqrt{2\pi}}.$$

Under the null model, the Bayes factor of M_γ to the null model is:

$$B_{\gamma 0}(\mathbf{x}) = \frac{m_\gamma(\mathbf{x})}{m_{\mathbf{0}}(\mathbf{x})} = \prod_{i:\gamma_i \neq 0} \frac{1}{\sqrt{1+\tau^2}} \exp\left\{\frac{\tau^2}{2(1+\tau^2)} x_i^2\right\}. \quad (3.6)$$

3.1.2 Posterior model probabilities

The posterior model probabilities are

$$P(M_\gamma \mid \mathbf{x}) = \frac{P(M_\gamma) B_{\gamma 0}(\mathbf{x})}{\sum_{\gamma \in \mathcal{M}_{k_{max}}} P(M_\gamma) B_{\gamma 0}(\mathbf{x})} = \frac{P(M_\gamma) B_{\gamma 0}(\mathbf{x})}{\sum_{k=0}^{k_{max}} (p_{(k)} \frac{1}{\binom{m}{k}}) \sum_{\boldsymbol{\eta}: |\boldsymbol{\eta}|=k} B_{\boldsymbol{\eta} 0}(\mathbf{x})}. \quad (3.7)$$

3.2 Posterior probability of the null model as n grows

In Theorem 2.5.5, we saw the interesting fact that, for the mutually exclusive testing scenario and under the null model, the null posterior probability converges to its prior probability as the number of tests grows. In this section, we show the same result for the generalized model.

Theorem 3.2.1. *Suppose $k_{max} = O(m^r)$ where $r \in [0, \frac{1}{1+\tau^2}]$ and $\tau^2 > 1/3$. Then, under the null model,*

$$P(M_0 | \mathbf{X}) \rightarrow P(M_0) \text{ when } m \rightarrow \infty.$$

Proof. The posterior can be written in term of Bayes factors:

$$P(M_0 | \mathbf{X}) = \frac{P(0)}{\sum_{k=0}^{k_{max}} p(k) \frac{1}{\binom{m}{k}} \sum_{|\gamma|=k} B_{\gamma 0}(\mathbf{X})}. \quad (3.8)$$

The key step is to show

$$\frac{1}{\binom{m}{k}} \sum_{|\gamma|=k} B_{\gamma 0} = 1 + o(1) \quad \forall k = \{1, 2, \dots, k_{max}\}, \quad (3.9)$$

which would complete the proof by observing that the denominator of (3.8) is

$$\sum_{k=1}^{k_{max}} \frac{p(k)}{p(0)} \left(1 + o(1)\right) = \frac{1}{p(0)} \sum_{k=1}^{k_{max}} p(k) (1 + o(1)) = \frac{1 - p(0)}{p(0)} + o(1).$$

To prove (3.9), first denote $Y_j = \frac{1}{\sqrt{1+\tau^2}} e^{X_j^2 \tau^2 / 2(1+\tau^2)}$ and expand $(\frac{1}{m} \sum_1^m Y_j)^k$:

$$\begin{aligned} \left[\frac{\sum_1^m Y_j}{m} \right]^k &= \sum_{k_1 + \dots + k_m = k} \binom{k}{k_1, k_2, \dots, k_m} \prod_{1 \leq j \leq m} \frac{Y_j^{k_j}}{m^{k_j}} \\ &= \underbrace{\sum_{\substack{k_1 + \dots + k_m = k \\ \max |k_i| = 1}} \frac{k!}{m^k} \prod_{1 \leq j \leq m} Y_j^{k_j}}_I + \underbrace{\sum_{\substack{k_1 + \dots + k_m = k \\ \max |k_i| \geq 2}} \frac{k!}{k_1! \dots k_m!} \prod_{1 \leq j \leq m} \frac{Y_j^{k_j}}{m^{k_j}}}_{II}. \end{aligned} \quad (3.10)$$

Notice that

$$\begin{aligned}
I &= \frac{(m-1)\dots(m-k+1)}{m^{k-1}} \frac{1}{\binom{m}{k}} \sum_{\substack{k_1+\dots+k_m=k \\ \max|k_i|=1}} \prod_{1 \leq j \leq m} \exp\left\{\frac{\tau^2}{2(1+\tau^2)} k_j X_j^2\right\} \\
&= \frac{(m-1)\dots(m-k+1)}{m^{k-1}} \frac{1}{\binom{m}{k}} \sum_{|\gamma|=k} B_{\gamma 0} \\
&= \left[\frac{1}{\binom{m}{k}} \sum_{|\gamma|=k} B_{\gamma 0}(x) \right] (1 + o(1)); \\
II &= \frac{1}{m^k} \sum_{\substack{k_1+\dots+k_m=k \\ \max|k_i| \geq 2}} \frac{k!}{k_1! \dots k_m!} \prod_{1 \leq j \leq m} Y_j^{k_j} \\
&\leq \frac{k(k-1)}{m^k} \left(\sum_1^m y_i^2 \right) \sum_{\tilde{k}_1+\dots+\tilde{k}_m=k-2} \frac{(k-2)!}{\tilde{k}_1! \dots \tilde{k}_m!} Y_1^{\tilde{k}_1} \dots Y_m^{\tilde{k}_m} \\
&= \left(\frac{k(k-1)}{m^{2/(1+\tau^2)}} \right) \left(\frac{\sum_1^m Y_i^2}{m^{2\tau^2/(1+\tau^2)}} \right) \left(\frac{\sum_1^m Y_i}{m} \right)^{k-2}.
\end{aligned}$$

To conclude that $II \rightarrow 0$ when $m \rightarrow \infty$, note that $\frac{k(k-1)}{m^{2/(1+\tau^2)}}$ converges to a constant because of the $k = O(m^r)$ assumption;

By Lemma 3.5.3, when $\tau^2 > 1/3$, $\frac{\sum_1^m Y_i^2}{m^{2\tau^2/(1+\tau^2)}} = o(1)$;

By Lemma 3.5.5, $\left(\frac{\sum_1^m Y_i}{m}\right)^{k-2} = 1 + o(1)$.

Hence, the right hand side of (3.10) converges to $1/\binom{m}{k} \sum_{|\gamma|=k} B_{\gamma 0}(1 + o(1))$. By Lemma 3.5.5, the left hand side goes to 1. Hence,

$$\lim_{m \rightarrow \infty} \frac{1}{\binom{m}{k}} \sum_{|\gamma|=k} B_{\gamma 0} = 1.$$

□

Remark 3.2.2. Figure 3.1 plots the null posterior probability under the null model for different numbers of hypotheses. Here we use the prior $p_{(0)} = 0.5$; $p_{(1)} = p_{(2)} = \dots = p_{(k_{max})} = \frac{1-p_{(0)}}{k_{max}}$. As can be seen, under the null model, the posterior of null model converges to its prior probability when m grows, as established in Theorem 3.2.1.

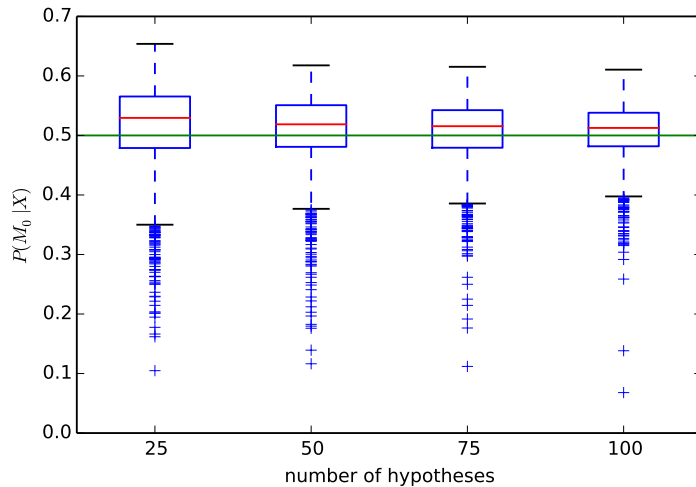


FIGURE 3.1: Box plot of the posterior probability of the null model under the null model, where $P(M_0) = 0.5$, $k_{max} = 4$, $\tau = 2$. The green horizontal line is $y = 0.5$. For each m (number of hypotheses), 3000 iterations have been performed.

3.3 Asymptotic frequentist properties of Bayesian procedures

This section is devoted to studying the asymptotic false positive probability and the expected number of false discoveries under the null model. We begin by discussing rates of the inclusion probability and the decision criterion.

3.3.1 Inclusion probabilities and the decision rule

Definition 3.3.1. The posterior inclusion probability for the i^{th} mean is

$$P(\gamma_i = 1 | \mathbf{X}) = \sum_{\gamma} \mathbb{1}_{\gamma_i=1} P(M_{\gamma} | \mathbf{X}).$$

Theorem 3.3.2. Suppose $k_{max} = O(m^r)$ where $r \in [0, \frac{1}{1+\tau^2}]$ and $\tau^2 > 1/3$. Then, under the null model and as $m \rightarrow \infty$,

$$P(\gamma_i = 1 | \mathbf{X}) = \frac{\frac{1}{\sqrt{1+\tau^2}} \exp\left\{\frac{\tau^2 X_i^2}{2(1+\tau^2)}\right\} \left(\frac{1}{m} \sum_1^{k_{max}} k p(k)\right)}{1 + \left(\frac{1}{m} \sum_1^{k_{max}} k p(k)\right) \left(\frac{1}{\sqrt{1+\tau^2}} \exp\left\{\frac{\tau^2 X_i^2}{2(1+\tau^2)}\right\} - 1\right)} (1 + o(1)). \quad (3.11)$$

Proof.

$$P(\gamma_i = 1 \mid \mathbf{X}) = \frac{\sum_{k=1}^{k_{max}} p(k) \frac{1}{\binom{m}{k}} \sum_{\substack{|\gamma|=k \\ \gamma_i=1}} B_{\gamma_0}}{p(0) + \sum_{k=1}^{k_{max}} p(k) \left[\frac{1}{\binom{m}{k}} \sum_{\substack{|\gamma|=k \\ \gamma_i=1}} B_{\gamma_0} + \frac{1}{\binom{m}{k}} \sum_{\substack{|\gamma|=k \\ \gamma_i=0}} B_{\gamma_0} \right]}. \quad (3.12)$$

From the explicit expression for Bayes factors (3.6):

$$B_{\gamma_0} = \frac{1}{\sqrt{1+\tau^2}} \exp \left\{ \frac{\tau^2 X_i^2}{2(1+\tau^2)} \right\} \sum_{|\tilde{\gamma}|=k-1} B_{\tilde{\gamma}_i 0}(\tilde{\mathbf{X}}_i), \quad (3.13)$$

where

$$\begin{aligned} \tilde{\gamma}_i &= (\gamma_1, \dots, \gamma_{i-1}, \gamma_{i+1}, \dots, \gamma_m), \\ \tilde{\mathbf{X}}_i &= (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_m). \end{aligned}$$

By (3.9),

$$\frac{1}{\binom{m-1}{k-1}} \sum_{|\tilde{\gamma}_i|=k-1} B_{\tilde{\gamma}_i 0}(\tilde{\mathbf{X}}_i) \rightarrow 1 \forall i.$$

Therefore,

$$\begin{aligned} \frac{1}{\binom{m}{k}} \sum_{\substack{|\gamma|=k \\ \gamma_i=1}} B_{\gamma_0} &= \frac{\binom{m-1}{k-1}}{\binom{m}{k}} \frac{1}{\sqrt{1+\tau^2}} \exp \left\{ \frac{\tau^2 X_i^2}{2(1+\tau^2)} \right\} (1 + o(1)) \\ &= \frac{k}{m} \frac{1}{\sqrt{1+\tau^2}} \exp \left\{ \frac{\tau^2 X_i^2}{2(1+\tau^2)} \right\} (1 + o(1)). \end{aligned}$$

Similarly, one can show

$$\frac{1}{\binom{m}{k}} \sum_{\substack{|\gamma|=k \\ \gamma_i=0}} B_{\gamma_0} = \frac{\binom{m-1}{k}}{\binom{m}{k}} (1 + o(1)) = \frac{m-k}{m} (1 + o(1)).$$

Combining all of these, the inclusion posterior probability (3.12) becomes:

$$P(\gamma_i = 1 \mid \mathbf{X}) = \frac{\frac{1}{\sqrt{1+\tau^2}} \exp \left\{ \frac{\tau^2 X_i^2}{2(1+\tau^2)} \right\} \left(\frac{1}{m} \sum_1^{k_{max}} k p(k) \right)}{1 + \left(\frac{1}{m} \sum_1^{k_{max}} k p(k) \right) \left(\frac{1}{\sqrt{1+\tau^2}} \exp \left\{ \frac{\tau^2 X_i^2}{2(1+\tau^2)} \right\} - 1 \right)} (1 + o(1)). \quad (3.14)$$

□

Remark 3.3.3. Figure 3.2 is the box plot of the ratio of the true inclusion probability (3.12) to the approximated one (3.11) under the null model. As expected, the ratio concentrates at 1, indicating that the estimated inclusion probability matches with the true one. In addition, the empirical convergence rate can be obtained by running the same simulation for multiple numbers of hypotheses.

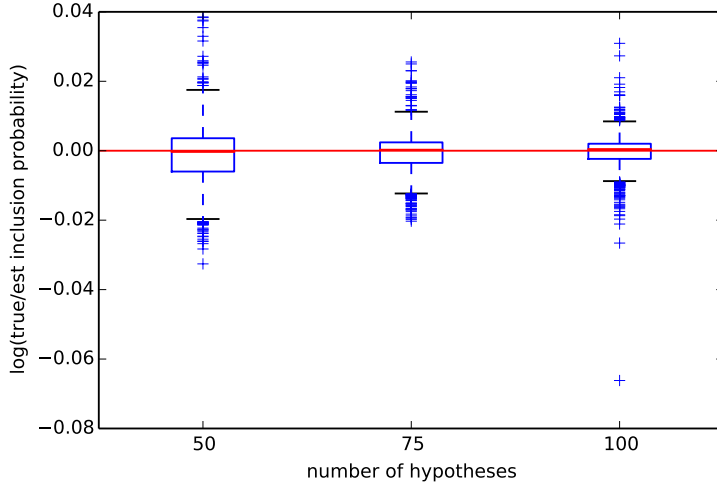


FIGURE 3.2: For $k_{max} = 4$, $\tau = 2$, and various numbers of hypotheses m , box plots of the logarithm of the ratio of the true inclusion probability to the estimated inclusion probability, under the null model. For each m , 3000 iterations have been carried out. The red horizontal line is $y = 0$ (ratio = 1).

Definition 3.3.4 (Decision rule). The i^{th} channel is claimed to have a signal if the posterior inclusion probability $P(\gamma_i = 1 \mid \mathbf{x})$ is greater than a preassigned threshold $q \in (0, 1)$.

Theorem 3.3.5 (False positive probability). *If the maximum model size $k_{max} = O(m^r)$ where $r \in [0, \frac{1}{1+\tau^2}]$, $\tau^2 > 1/3$, then under the null model, the false positive probability (FPP) asymptotically equals to:*

$$\frac{\left(\sum_{k=1}^{k_{max}} kp(k)\right)^{\frac{1+\tau^2}{\tau^2}}}{m^{\frac{1}{\tau^2}}} \frac{(1-q)^{\frac{1+\tau^2}{\tau^2}}}{\sqrt{\log \left[\frac{q}{1-q} \sqrt{1+\tau^2} \left(\frac{m}{\sum_{k=1}^{k_{max}} kp(k)} \right) \right]}} \frac{|\tau|}{\sqrt{\pi q} \frac{1+\tau^2}{\tau^2} (1+\tau^2)^{1+\frac{1}{2\tau^2}}}.$$

Proof. A false positive occurs when this probability is greater than p , from Theorem 3.3.2:

$$\begin{aligned}
& P\left(|X_i| > \underbrace{\sqrt{2\left(\frac{1+\tau^2}{\tau^2}\right) \log \left[\frac{q}{1-q} \sqrt{1+\tau^2} \left(\frac{m}{\sum_{k=1}^{k_{max}} kp(k)}\right)\right]}}_{\gamma_m} (1+o(1))\right) \\
&= \left(\frac{\sum_{k=1}^{k_{max}} kp(k)}{m}\right)^{\frac{1+\tau^2}{\tau^2}} \frac{1}{\sqrt{\log \left(\frac{m}{\sum_{k=1}^{k_{max}} kp(k)}\right)}} \frac{|\tau|}{\sqrt{\pi}(1+\tau^2)^{1+\frac{1}{2\tau^2}}} \left(\frac{1-q}{q}\right)^{\frac{1+\tau^2}{\tau^2}} (1+o(1)) \\
&= \left(\frac{\sum_{k=1}^{k_{max}} kp(k)}{m}\right)^{\frac{1+\tau^2}{\tau^2}} \frac{1}{\sqrt{\log \left[\frac{q}{1-q} \sqrt{1+\tau^2} \left(\frac{m}{\sum_{k=1}^{k_{max}} kp(k)}\right)\right]}} \\
&\quad \frac{|\tau|}{\sqrt{\pi}(1+\tau^2)^{1+\frac{1}{2\tau^2}}} \left(\frac{1-q}{q}\right)^{\frac{1+\tau^2}{\tau^2}} (1+o(1)) \text{ by Fact 5.4.1.}
\end{aligned} \tag{3.15}$$

Hence, the FPP is:

$$\begin{aligned}
& P(\text{any false positive} \mid M_0) \\
&= 1 - P(\text{no false positive} \mid M_0) = 1 - \prod_{i=1}^m P\left(|X_i| \leq \gamma_m\right) \\
&= 1 - \prod_{i=1}^m [1 - P(|X_i| > \gamma_m)] = 1 - [1 - P(|X_i| > \gamma_m)]^m \\
&= mP(|X_1| > \gamma_m)(1+o(1)) \\
&= \frac{\left(\sum_{k=1}^{k_{max}} kp(k)\right)^{\frac{1+\tau^2}{\tau^2}}}{m^{\frac{1}{\tau^2}}} \frac{1}{\sqrt{\log \left[\frac{q}{1-q} \sqrt{1+\tau^2} \left(\frac{m}{\sum_{k=1}^{k_{max}} kp(k)}\right)\right]}} \\
&\quad \frac{|\tau|}{\sqrt{\pi}(1+\tau^2)^{1+\frac{1}{2\tau^2}}} \left(\frac{1-q}{q}\right)^{\frac{1+\tau^2}{\tau^2}} (1+o(1)).
\end{aligned}$$

□

Remark 3.3.6. Figure 3.3 demonstrates the convergence rate of the false positive probabilities as the number of hypotheses increases. At each m , 300 m -dimensional multivariate

normals were generated. FPP is the number of multivariate normals having at least one false positive over 300. To get the distribution of FPP, repeat the above procedure a hundred times and include all FPPs in the box plot. Furthermore, the theoretical FPP based on Theorem 3.3.5 is plotted in black lines. As m grows, the simulated FPP converges to the theoretical FPP as expected.

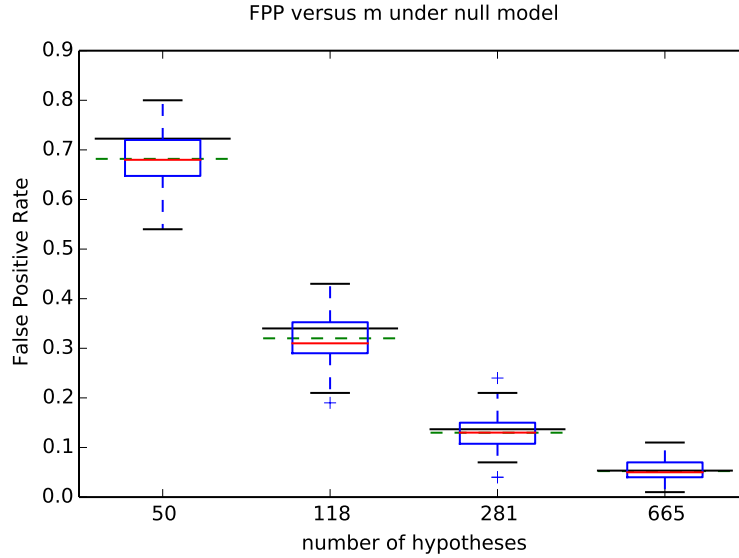


FIGURE 3.3: Box plots of empirical false positive rates with median (red line), mean (green dashed line) and the theoretical false positive probabilities (black lines), when $k_{max} = 2$, the decision threshold is $q = 0.05$, and $\tau = 1$.

3.3.2 Expected number of false positives

Theorem 3.3.7. *Suppose $k_{max} = O(m^r)$, $r \in [0, \frac{1}{1+\tau^2}]$, and $\tau^2 > 1/3$. Then under the null model, the expected number of false positives is asymptotically the same as the false positive probability.*

Proof.

$$\begin{aligned}
& \mathbb{E} \left[\sum_{i=1}^m \mathbb{1}_{P(\gamma_i=1|\mathbf{X})>1/2} \mid M_0 \right] \\
&= \sum_{i=1}^m P \left((P(\gamma_i = 1 \mid \mathbf{X}) > 1/2) \mid M_0 \right) \\
&= mP \left(|X_i| > \sqrt{2 \left(\frac{1+\tau^2}{\tau^2} \right) \log \left[\frac{q}{1-q} \sqrt{1+\tau^2} \left(\frac{m}{\sum_{k=1}^{k_{max}} kp_{(k)}} \right) \right]} (1+o(1)) \right) \\
&= O \left[\frac{\left(\sum_{k=1}^{k_{max}} kp_{(k)} \right)^{\frac{1+\tau^2}{\tau^2}}}{m^{\frac{1}{\tau^2}}} \frac{1}{\sqrt{\log \left(\frac{m}{\sum_{k=1}^{k_{max}} kp_{(k)}} \right)}} \right].
\end{aligned} \tag{3.16}$$

□

3.3.3 FPPs for various structures of prior probabilities

The false positive probability (Theorem 3.3.5) depends on the prior structure $p_{(i)} = \sum_{|\gamma|=i} P(M_\gamma)$. In this section, we consider several configurations for this prior structure and compare the resulting FPPs.

1. Trivially, no false positive occurs if a prior does not allow any positive models, i.e. $p_{(0)} = 1$. This is the case when $\sum_{k=1}^{k_{max}} kp_{(k)}$ is 0. This case can be categorized as $k_{max} = 0$ as well.
2. If $k_{max} = 1$, $p_{(0)} = r$, then $p_{(1)} = 1 - r$. In this case, the multiple signals model reduces to the mutually exclusive model. As expected, the FPP rate in the mutually exclusive model (Theorem 2.5.9) is a special case of FPP rate in the generalized model (Theorem 3.3.5); and the inclusion probability (3.12) is essentially the same as the posterior probability in Lemma 2.5.1.
3. At the other extreme, the maximum of $\sum_{k=1}^{k_{max}} kp_{(k)}$ can be obtained straightforwardly by Cauchy's inequality:

Fact 3.3.8. When $p_{(k)} = \frac{2k}{k_{max}(k_{max}+1)}(1 - p_{(0)})$, $\sum_{k=1}^{k_{max}} kp_{(k)}$ attains its maximum at

$$\frac{2k_{max} + 1}{3}(1 - p_{(0)}).$$

4. The Beta-induced prior $p_{(0)} = p_{(1)} = p_{(2)} = \dots = p_{(k_{max})}$ has similar asymptotic behavior to the previous case since $\sum_{k=1}^{k_{max}} kp_{(k)}$ has the same order:

Fact 3.3.9. When the prior probabilities on all levels are the same: $p_{(0)} = p_{(1)} = p_{(2)} = \dots = p_{(k_{max})} = \frac{1}{k_{max}+1}$, then

$$\sum_{k=1}^{k_{max}} kp_{(k)} = \frac{k_{max}}{2}.$$

For cases 3 and 4 above, if $k_{max} = O(m^r)$, the corresponding FPP is:

$$O\left(m^{\frac{1}{\tau^2}(r(1+\tau^2)-1)} \frac{1}{\sqrt{\log m}}\right).$$

In particular, when $r = \frac{1}{1+\tau^2}$, the FPP is $O\left(\frac{1}{\sqrt{\log m}}\right)$.

Theorem 3.3.10. *Suppose $k_{max} = O(m^r)$ where $r \in [0, \frac{1}{1+\tau^2}]$ and there is a non-zero weight on an alternative model, i.e. $p_{(0)} < 1$. Then for any $\alpha \in (0, 1)$, the following decision threshold*

$$q = q(m, \alpha) = \left(1 + \frac{m^{1/(1+\tau^2)}}{\sum_{k=1}^{k_{max}} kp_{(k)}} (\pi\alpha^2 \log m)^{\frac{\tau^2}{2(1+\tau^2)}} \sqrt{1 + \tau^2}\right)^{-1}$$

guarantees the false positive probability to be α .

Proof. FPP can be obtained from criteria (3.15) with the aforementioned $p(m, \alpha)$:

$$\begin{aligned}
& mP \left\{ |X_i| > \sqrt{2 \left(\frac{1 + \tau^2}{\tau^2} \right) \log \left[\frac{q(m, \alpha)}{1 - q(m, \alpha)} \sqrt{1 + \tau^2} \left(\frac{m}{\sum_{k=1}^{k_{max}} k p(k)} \right) \right]} (1 + o(1)) \right\} \\
&= mP \left\{ |X_i| > \sqrt{2 \log \left[m \left(\frac{1}{\sqrt{\pi \alpha^2 \log m}} \right) \right]} (1 + o(1)) \right\} \\
&= mP \left\{ |X_i| > \sqrt{\left[2 \log m - \log \left(\pi \alpha^2 \log m \right) \right]} (1 + o(1)) \right\} \tag{3.17} \\
&= m \frac{2 \frac{1}{m} \sqrt{\pi \alpha^2 \log m}}{\sqrt{2\pi} \sqrt{2 \log m - \log \log m - \log(\pi \alpha^2)}} (1 + o(1)) \quad \text{by Fact 5.4.1} \\
&= \alpha(1 + o(1)).
\end{aligned}$$

□

Remark 3.3.11. Figure 3.4 is a simulation study of Theorem 3.3.10 for fixed $\alpha = 0.05$. The red curve is $1 - (1 - \frac{\alpha}{m})^m$ which converges to α when m grows. FPP is generated from the same procedure as Remark 3.3.6. As can be seen, numerical FPP converges to $1 - (1 - \alpha/m)^m$ (and hence converges to α) slowly.

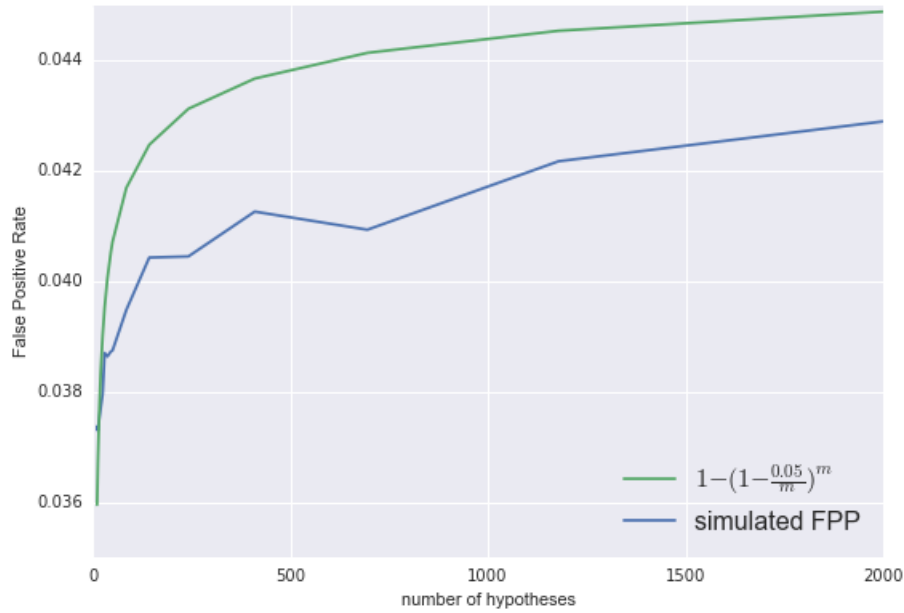


FIGURE 3.4: False positive probability based on the threshold in Theorem 3.3.10. The blue curve is numerical FPP (average FPP at given m); the green curve is $1 - (1 - \frac{\alpha}{m})^m$ where $\alpha = 0.05$.

Remark 3.3.12. Interestingly, from the numerical study Figure 3.4, the real FPP tends to be smaller than the theoretical FPP, indicating that the testing procedure usually performs better than the preassigned FPP.

Remark 3.3.13. Figure 3.5 shows the power with respect to different signal size (θ_i). For each θ , box plot of 100 iterations of posterior inclusion probabilities is shown.

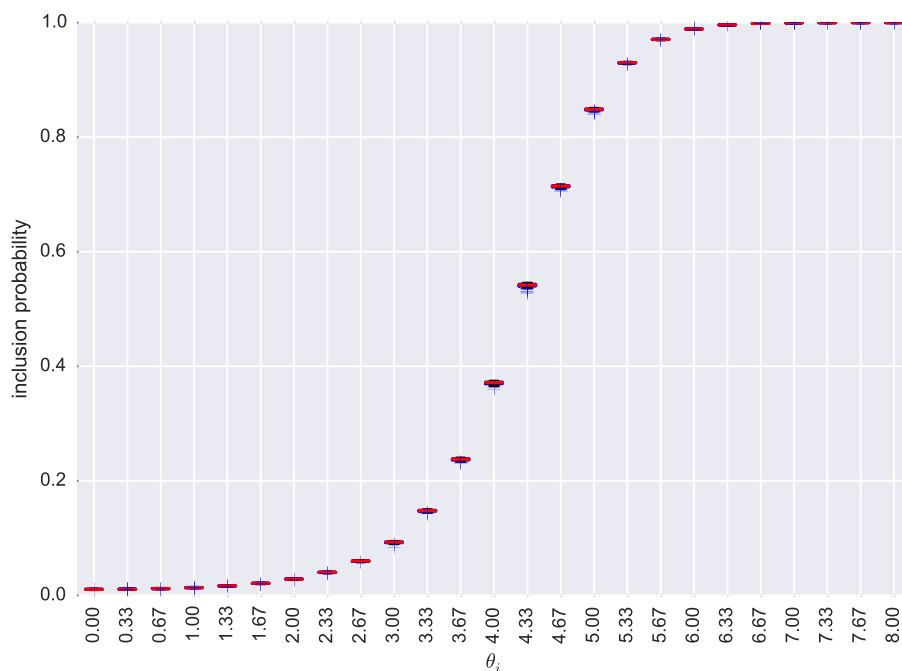


FIGURE 3.5: Box plots of inclusion probability $P(\mu_i \neq 0 \mid X)$ versus signal size θ_i . $m = 100, k_{max} = 3, \tau = 1$. For each θ_i , 100 iterations were generated to get the box plot.

Remark 3.3.14. Figure 3.6 demonstrates how $\pi(p \mid \mathbf{x})$ varies with respect to different expected and true noise ratios. The model noise ratio is k_{max}/m , and the true noise ratio (the legend on the top right of each plot) is the number of nonzero means over all means ($= m$) in the data generation process. As can be seen, when the model size is not restricted ($k_{max} = m$), the posterior of p concentrates at the true noise ratio. On the other hand, $\pi(p \mid \mathbf{x})$ tends to not surpass the model noise ratio k_{max}/m even when the true ratio is much higher.

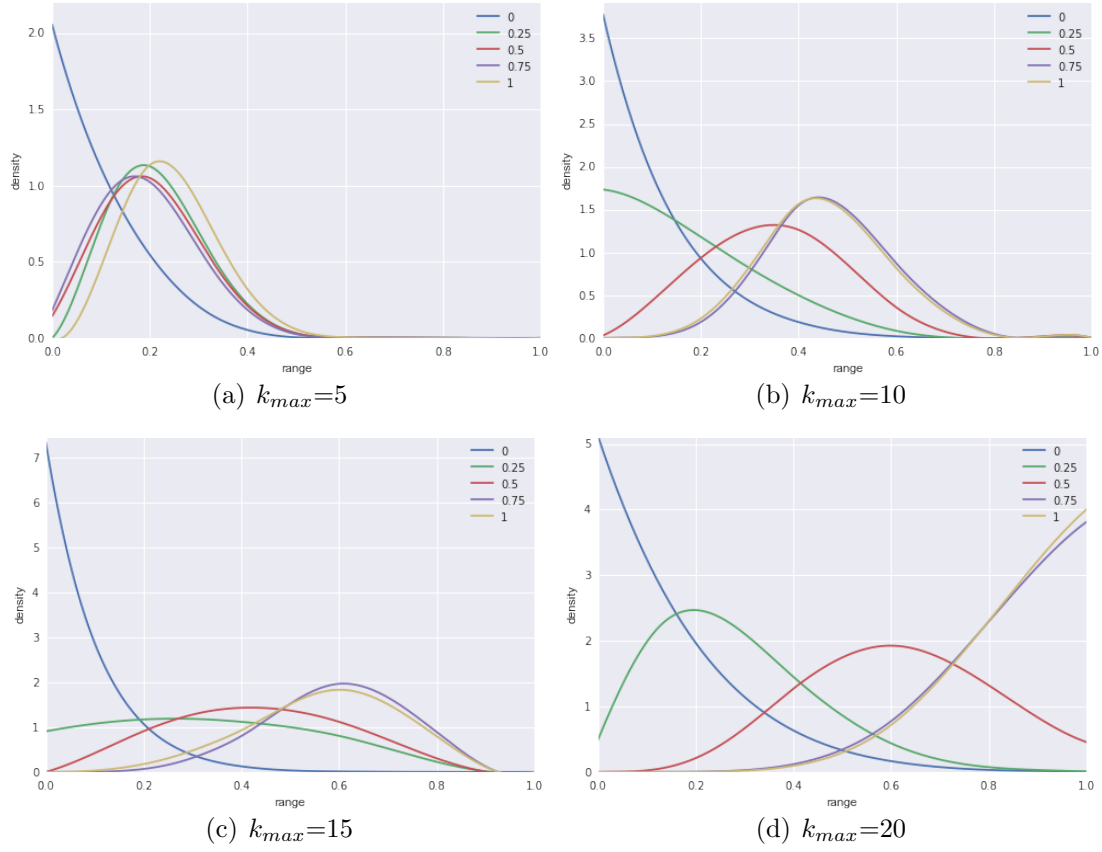


FIGURE 3.6: Posterior of p versus different noise ratios. $m = 20$, $\tau = 2$. For each plot, k_{max} is indicated in the subtitle, and simulate data \mathbf{x} according to the top right true noise ratio legend.

3.4 Conclusion

In this chapter, we described the asymptotic behavior of frequentist properties of the standard Bayesian model selection procedure, providing explicit rates for the false positive probability and the expected number of false positives. For those who prefer having a pre-specified constant false positive rate, the Bayesian decision threshold was provided that guarantees this.

Future work on this topic includes extending weak control to strong control; doing all computations under the null model does not reflect situations such as DNA sequencing data, where more than one gene will typically affect phenotypes. Another generalization

is to consider more general covariance structures for \mathbf{X} . It should be possible to apply a matrix decomposition method, similar to Theorem 2.3.1, and obtain the relevant frequentist properties.

3.5 Appendix

Fact 3.5.1 (Normal tail probability). Let $\Phi(t)$ be cumulative distribution function of standard normal. Then

$$\frac{t \frac{1}{\sqrt{2\pi}} e^{-t^2/2}}{t^2 + 1} \leq 1 - \Phi(t) \leq \frac{\frac{1}{\sqrt{2\pi}} e^{-t^2/2}}{t},$$

$$1 - \Phi(t) = \frac{\frac{1}{\sqrt{2\pi}} e^{-t^2/2}}{t} (1 + o(1)).$$

The proof can be found in Durrett (2010).

Fact 3.5.2 (Weak law for triangular arrays). For each m , let $X_{m,k}$, $1 \leq k \leq m$ be independent random variables. Let $\beta_m > 0$ with $\beta_m \rightarrow \infty$ and let $\bar{X}_{m,k} = X_{m,k} 1_{\{|X_{m,k}| \leq \beta_m\}}$.

Suppose that as $m \rightarrow \infty$

$$\sum_{k=1}^m P(|X_{m,k}| > \beta_m) \rightarrow 0,$$

$$\frac{1}{\beta_m^2} \sum_{k=1}^m E \bar{X}_{m,k}^2 \rightarrow 0.$$

Denote $S_m = X_{m,1} + \dots + X_{m,m}$ and $\alpha_m = \sum_{k=1}^m E \bar{X}_{m,k}$, then

$$\frac{(S_m - \alpha_m)}{\beta_m} \rightarrow 0 \text{ in probability.}$$

See Durrett (2010) for proof.

Lemma 3.5.3. *If $Y_j = \frac{1}{\sqrt{1+\tau^2}} \exp\{\frac{\tau^2}{2(1+\tau^2)} X_j^2\}$ and $\tau^2 > 1/3$, then*

$$m^{-2\tau^2/(1+\tau^2)} \sum_{j=1}^m Y_j^2 \rightarrow 0 \quad \text{when } m \rightarrow \infty.$$

Proof. First we need the following order estimate before applying Fact 3.5.2: If $\lim_{m \rightarrow \infty} a_m \in (0, \infty]$, $\lim_{m \rightarrow \infty} b_m = \infty$, then

$$\int_0^{b_m} e^{-\frac{a_m x^2}{2}} dx = O\left(\frac{e^{a_m b_m^2/2}}{a_m b_m}\right). \quad (3.18)$$

To show (3.18) is true, notice that $\int_0^{b_m} e^{-\frac{a_m x^2}{2}} dx = \frac{1}{\sqrt{a_m}} \int_0^{\sqrt{a_m} b_m} e^{-\frac{x^2}{2}} dx$ and by L'Hôpital's rule:

$$\begin{aligned} \lim_{m \rightarrow \infty} \frac{\frac{1}{\sqrt{a_m}} \int_0^{\sqrt{a_m} b_m} e^{-\frac{x^2}{2}} dx}{e^{-\frac{a_m b_m^2}{2}} / (a_m b_m)} &= \lim_{m \rightarrow \infty} \frac{\int_0^{\sqrt{a_m} b_m} e^{-\frac{x^2}{2}} dx}{e^{-\frac{a_m b_m^2}{2}} / (\sqrt{a_m} b_m)} \\ &= \lim_{v \rightarrow \infty} \frac{\int_0^v e^{-\frac{x^2}{2}} dx}{e^{-\frac{v^2}{2}} / v} \quad \text{where } v = \sqrt{a_m} b_m \\ &= \lim_{v \rightarrow \infty} \frac{\frac{d}{dv} \int_0^v e^{-x^2/2} dx}{\frac{d}{dv} \frac{e^{-v^2/2}}{v}} = \lim_{v \rightarrow \infty} \frac{1}{1 - \frac{1}{v^2}} = 1. \end{aligned}$$

Then take $X_{m,j} = \exp\{\frac{\tau^2}{1+\tau^2} x_j^2\}$, $\beta_m = m^{2\tau^2/(1+\tau^2)}$ in Fact 3.5.2, both assumptions in Fact 3.5.2 hold since:

$$\sum_{j=1}^m P(|X_{m,j}| > \beta_m) = O\left(\frac{1}{\sqrt{\log m}}\right) \rightarrow 0 \quad \text{by Fact 5.4.1,}$$

and

$$\begin{aligned} \frac{1}{\beta_m^2} \sum_{j=1}^m E \bar{X}_{m,j}^2 &= \frac{m}{m^{4\tau^2/(1+\tau^2)}} \int_{|x| \leq \sqrt{2 \log m}} \frac{1}{\sqrt{2\pi}} e^{(\frac{2\tau^2}{1+\tau^2}-1)\frac{x^2}{2}} dx \\ &= \begin{cases} O\left(m^{-\frac{2\tau^2}{1+\tau^2}} \frac{1}{\sqrt{\log m}}\right) & \text{if } \tau^2 > 1, \text{ by (3.18)} \\ O\left(m^{\frac{1-3\tau^2}{1+\tau^2}}\right) & \text{else,} \end{cases} \\ &= o(1) \quad \text{if } \tau^2 > 1/3. \end{aligned}$$

The limit is:

$$\begin{aligned}
\frac{1}{\beta_m} \sum_{j=1}^m E\bar{X}_{m,j} &= m^{-2\tau^2/(1+\tau^2)} \frac{m}{\sqrt{2\pi}} \int_{|x| < \sqrt{2 \log m}} \exp \left\{ \frac{x^2}{2} \left(\frac{2\tau^2}{1+\tau^2} - 1 \right) \right\} \\
&= m^{(1-\tau^2)/(1+\tau^2)} O \left(m^{(\tau^2-1)/(1+\tau^2)} \frac{1}{\sqrt{\log m}} \right) \text{ by (3.18)} \\
&= O \left(\frac{1}{\sqrt{\log m}} \right) \rightarrow 0.
\end{aligned}$$

□

Fact 3.5.4 (Marcinkiewicz and Zygmund). Let X_1, X_2, \dots be i.i.d. with $EX_1 = 0$ and $E|X_1|^p < \infty$ where $1 < p < 2$. If $S_m = X_1 + \dots + X_m$ then

$$\frac{S_m}{m^{1/p}} \rightarrow 0 \text{ a.s.}$$

Proof of the claim can be found in Durrett (2010).

Lemma 3.5.5. Let $Y_j = \frac{1}{\sqrt{1+\tau^2}} \exp\{\frac{\tau^2}{2(1+\tau^2)} X_j^2\}$, $k = O(m^r)$ where $r \in [0, \frac{1}{1+\tau^2}]$ then:

$$\left(\frac{1}{m} \sum_{j=1}^m Y_j \right)^k = 1 + o(1).$$

Proof.

When $r = 0$, note that $\mathbb{E}(Y_i | M_0) = 1 \forall i$, then by Strong Law of Large number, $1/m \sum_{j=1}^m Y_j \rightarrow 1$, this case is proved.

Now consider $r \in (0, \frac{1}{1+\tau^2}]$: take $X_i = Y_i - 1$, $p = \frac{1+\tau^2}{\tau^2}$ in Fact 3.5.4

$$\begin{aligned}
\frac{\sum_{j=1}^m Y_j - m}{m^{1/p}} &= \left(\frac{\sum_{j=1}^m Y_j}{m^{1/p}} - m^{1-1/p} \right) = o(1), \\
\frac{\sum_{j=1}^m Y_j}{m} - 1 &= \frac{1}{m^{1-1/p}} \left(\frac{\sum_{j=1}^m Y_j}{m^{1/p}} - m^{1-1/p} \right) = o \left(\frac{1}{m^{1-1/p}} \right), \tag{3.19}
\end{aligned}$$

the lemma follows since:

$$\log \left(\left[\frac{1}{m} \sum_{j=1}^m Y_j \right]^{k(m)} \right) = k(m) \log \left(\frac{1}{m} \sum_{j=1}^m Y_j \right) = O(m^r) o \left(m^{-\frac{1}{1+\tau^2}} \right) = o(1).$$

□

Lemma 3.5.6.

$$\frac{1}{\binom{m}{k}} \sum_{|\gamma|=k} B_{\gamma 0}(x) \text{ is a U-statistic.}$$

Proof. By definition,

$$B_{\gamma 0}(x_1, x_2, \dots, x_m) = \frac{1}{(1 + \tau^2)^{\frac{|\gamma|}{2}}} \exp \left\{ \frac{\tau^2}{2(1 + \tau^2)} \sum_{i:\gamma_i \neq 0} x_i^2 \right\}$$

$$h^k(x_1, x_2, \dots, x_k) = \frac{1}{(1 + \tau^2)^{\frac{k}{2}}} \exp \left\{ \frac{\tau^2}{2(1 + \tau^2)} \sum_{i=1}^k x_i^2 \right\}.$$

Notice that an iteration over models is the same as an iteration through inputs:

$$\sum_{\gamma:|\gamma|=k} B_{\gamma 0}(x) = \sum_{C_k^m} h^k(x_{i_1}, x_{i_2}, \dots, x_{i_k}).$$

Therefore,

$$\frac{1}{\binom{m}{k}} \sum_{|\gamma|=k} B_{\gamma 0}(x) \text{ is a U-statistics with kernel } h^k.$$

□

Definition 3.5.7. For $c \in \{0, 1, \dots, k\}$,

$$h_c^k(x_1, \dots, x_c) = \mathbb{E}(h^k(x_1, x_2, \dots, x_c, X_{c+1}, \dots, X_k))$$

$$(\sigma_c^k)^2 = \text{Var}(h_c^k(X_1, \dots, X_c)).$$

Note that $h_k^k = h^k$, $(\sigma_k^k)^2 = \text{Var}(h^k) = \sigma_k^2$.

Fact 3.5.8. Let U be the U-statistics with kernel h^k . If $\sigma_k^2 < \infty$, then

$$\sqrt{m}[U - \mathbb{E}(h^k(X_1, \dots, X_k))] \rightarrow N(0, k^2(\sigma_1^k)^2).$$

The asymptotic theory of U-statistics can be found in Ferguson (2003).

Theorem 3.5.9. If $\tau^2 < 1$, and $k_{max} = o(m)$, then under the null model,

$$\sqrt{m} \left[\frac{1}{\binom{m}{k}} \sum_{|\gamma|=k} B_{\gamma 0}(x) - 1 \right] \rightarrow N \left(0, k^2 \left[\frac{1}{\sqrt{1-\tau^4}} - 1 \right] \right).$$

Proof. By Lemma 3.5.6, $1/\binom{m}{k} \sum_{|\gamma|=k} B_{\gamma 0}$ is a U-statistics, then evaluate the expectation and variance in Definition 3.5.7, one can get under the null model:

$$\mathbb{E}(B_{\gamma 0}(X)) = \mathbb{E}(h_1^k(X)) = 1,$$

$$\mathbb{E}(B_{\gamma 0}(X)^2) = \begin{cases} \frac{1}{[(1+\tau^2)(1-\tau^2)]^{\frac{|\gamma|}{2}}} & \text{if } \tau^2 < 1, \\ \infty & \text{else,} \end{cases}$$

$$(\sigma_1^k)^2 = \begin{cases} \frac{1}{\sqrt{1-\tau^4}} - 1 & \text{if } \tau^2 < 1, \\ \infty & \text{else,} \end{cases}$$

$$(\sigma_k^k)^2 = \begin{cases} \frac{1}{[(1+\tau^2)(1-\tau^2)]^{\frac{k}{2}}} - 1 & \text{if } \tau^2 < 1, \\ \infty & \text{else.} \end{cases}$$

then apply Fact 3.5.8. □

Bayesian multiple testing for a sequence of trials

4.1 Introduction

Much progress has been made regarding multiplicity issues in drug development and clinical trials (Alosh et al. (2014), Streiner (2015)). These methods usually involve experimental design on multiple endpoints and subgroups, such as fixed sequence procedures (Holm (1979), Westfall and Krishen (2001)), group sequential procedures (O'Brien and Fleming (1979), Whitehead (1997). Siegmund (1985)), and graphical approaches (Bretz et al. (2009)). Here we consider a somewhat different scenario, imagining a sequence of completely separate trials designed to investigate a particular condition. For instance, consider the (by now large) series of trials that have been conducted in an effort to find a treatment for Alzheimer's disease. As these trials have been conducted by different companies and organizations, it is not common to think of the situation as one requiring multiplicity adjustment. But, from the perspective of society, this is a multiple testing problem and society should make a multiplicity adjustment to understand the evidence resulting from the sequence of trials.

In each of the trials, we presume that there is a null model M_0^i that reflects no treatment effect, and an alternative model M_1^i that indicates a treatment effect. We presume that

the data x_i from trial i is independent of the data from the other trials, and that that any parameters of the models in each trial are also apriori independent.

4.2 Bayesian sequence multiplicity control

The key idea is to just treat the sequence of trials as an ordinary multiple testing problem, letting p denote the (assumed common) prior probability that the the null hypothesis M_0^i is true in each trial, and allowing p to be updated with each new trial. Thus, if a majority of the trials fail to reject, p will move towards 1 and the next trial will need stronger evidence for a rejection. While this could be treated exactly as in Chapter 3, with a new multiple testing analysis at the end of each trial (taking the trials up to that point as the collection of multiple tests), it is of interest to examine how the formulas update in a sequential fashion. For simplicity in exposition, we will utilize the objective prior $\pi(p) = \mathbb{1}_{[0,1]}(p)$ in the following.

Denoting the marginal likelihood of the i^{th} trial under model M_j^i by $m(x_i | M_j^i)$, the Bayes factor of the null hypothesis to the alternative hypothesis in the i^{th} trial is

$$B_{01}^i = \frac{m(x_i | M_0^i)}{m(x_i | M_1^i)}.$$

Because of the sequential updating aspect of Bayes theorem, the posterior probability of the $n + 1$ null hypothesis M_0^{n+1} , in the sequence of trials, is given by

$$P(M_0^{n+1} | x_1, \dots, x_{n+1}) = \left(1 + \frac{1 - p_n}{p_n} \frac{1}{B_{01}^{n+1}}\right)^{-1},$$

where $p_n = \mathbb{E}(p | x_1, \dots, x_n)$.

The following theorem gives the explicit form of the posterior probability of p and its expectation.

Theorem 4.2.1. *The probability density function and expectation of the posterior proba-*

bility of p are

$$\pi(p \mid x_1, \dots, x_n) = \binom{n+1}{n+2} \frac{\sum_{j=0}^n p^{n-j}(1-p)^j \sum_{1 \leq i_1 < \dots < i_{n-j} \leq n} B_{01}^{i_1} \dots B_{01}^{i_{n-j}}}{\sum_{j=0}^n 1/\binom{n}{n-j} \sum_{1 \leq i_1 < \dots < i_{n-j} \leq n} B_{01}^{i_1} \dots B_{01}^{i_{n-j}}}, \quad (4.1)$$

$$\mathbb{E}(p \mid x_1, \dots, x_n) = \binom{n+1}{n+2} \frac{\sum_{j=0}^n 1/\binom{n+1}{n+1-j} \sum_{1 \leq i_1 < \dots < i_{n-j} \leq n} B_{01}^{i_1} \dots B_{01}^{i_{n-j}}}{\sum_{j=0}^n 1/\binom{n}{n-j} \sum_{1 \leq i_1 < \dots < i_{n-j} \leq n} B_{01}^{i_1} \dots B_{01}^{i_{n-j}}}.$$

Proof. By definition, the posterior of p is

$$\pi(p \mid x_1, x_2, \dots, x_n) = \frac{\prod_{i=1}^n \left(p B_{01}^i + (1-p) \right)}{\int_0^1 \prod_{i=1}^n \left(p B_{01}^i + (1-p) \right) dp}. \quad (4.2)$$

The result follows by expanding the products in (4.2) and using the following identity to evaluate the denominator:

$$\int_0^1 p^i (1-p)^j dp = \frac{\Gamma(i+1)\Gamma(j+1)}{\Gamma(i+j+2)} = \frac{i!j!}{(i+j+1)!} = \frac{1}{\binom{i+j}{i}(i+j+1)}.$$

□

Example 4.2.2. When $n = 1$, the posterior of p is

$$\pi(p \mid x_1) = \frac{pm(x_1 \mid M_0^1) + (1-p)m(x_1 \mid M_1^1)}{\frac{1}{2}m(x_1 \mid M_0^1) + \frac{1}{2}m(x_1 \mid M_1^1)} = 2 \frac{pB_{01}^1 + (1-p)}{B_{01}^1 + 1},$$

$$p_1 = \mathbb{E}(p \mid x_1) = \int_0^1 p\pi(p \mid x_1)dp = \frac{2B_{01}^1 + 1}{3(B_{01}^1 + 1)}.$$

Example 4.2.3. When $n = 2$, the posterior density of p , given x_1, x_2 , is

$$\pi(p \mid x_1, x_2) = \frac{f(x_1, x_2 \mid p)\pi(p)}{m(x_1, x_2)} = 6 \frac{p^2 B_{01}^1 B_{01}^2 + p(1-p)(B_{01}^1 + B_{01}^2) + (1-p)^2}{2B_{01}^1 B_{01}^2 + B_{01}^1 + B_{01}^2 + 2}.$$

$$p_2 = \mathbb{E}(p \mid x_1, x_2) = \frac{3B_{01}^1 B_{01}^2 + B_{01}^1 + B_{01}^2 + 1}{2(2B_{01}^1 B_{01}^2 + B_{01}^1 + B_{01}^2 + 2)}.$$

4.2.1 Recursive formula

Letting $m(x_i)$ denote the overall marginal likelihood of the i^{th} trial,

$$m_i = m(x_i) = P(M_0^i)m(x_i | M_0^i) + P(M_1^i)m(x_i | M_1^i),$$

a recursive formula for the posterior density of p can be obtained, as follows:

Corollary 4.2.4.

$$\pi(p | x_1, \dots, x_n) = \pi(p | x_1, \dots, x_{n-1}) \left(pB_{01}^n + (1-p) \right) \frac{m_{n-1}}{m_n},$$

where

$$m_n = \frac{1}{n+1} \left[\frac{\prod_{i=1}^n B_{01}^i}{\binom{n}{n}} + \frac{\sum_{j=1}^n \prod_{i \neq j} B_{01}^i}{\binom{n}{n-1}} + \frac{\sum_{j < k} \prod_{i \neq j, k} B_{01}^i}{\binom{n}{n-2}} + \dots + \frac{1}{\binom{n}{0}} \right].$$

Another possibility is dynamic programming:

Corollary 4.2.5 (Recursive formula).

$$m_n = \frac{1}{n+1} \left[\frac{a_{n1}}{\binom{n}{n}} + \frac{a_{n2}}{\binom{n}{n-1}} + \frac{a_{n3}}{\binom{n}{n-2}} + \dots + \frac{1}{\binom{n}{0}} \right],$$

$$\text{where } \begin{cases} a_{ij} = B_{01}^i a_{(i-1)j} + a_{(i-1)(j-1)} & \forall i, j \in \{1, \dots, n\} \\ a_{i0} = a_{0i} = 0 & \forall i \in \{1, \dots, n\} \\ a_{n(n+1)} = 1 \end{cases}$$

By caching the n coefficients of m_n (i.e. $a_{ni}, 1 \leq i \leq n+1$), m_{n+1} can be computed in linear time when the new data or Bayes factor B_{01}^{n+1} is provided.

4.3 HIV vaccines

In this section, we take the HIV vaccine trial study (Gilbert et al. (2011)) as an example to demonstrate how multiplicity control works in a sequence of trials. Here, a sequence of three vaccines were tested, yielding the following results:

Vax004:

127 out of 1805 infected in the placebo group

241 out of 3598 infected in the vaccine group

Vax003:

105 out of 1260 infected in the placebo group

106 out of 1267 infected in the vaccine group

Lee et al. analysis

7 out of 240 infected in the placebo group.

24 out of 1257 infected in the vaccine group

For convenience, use $i \in \{1, 2, 3\}$ for these trials and $j = 0, 1$ for the placebo and vaccine group, respectively; n_{ij} is the number of subjects in vaccine i and group j ; x_{ij} is the number of infected cases in each group; $\mathbf{x}_i = \{x_{i0}, x_{i1}\}$; and $\mathbf{n}_i = \{n_{i0}, n_{i1}\}$. We model $x_{ij} \sim \text{Binomial}(p_{ij}, n_{ij})$ with a $\text{Beta}(a, b)$ prior on p_{ij} .

To investigate the efficacy of each vaccine, test

$$\begin{cases} M_0^i : p_{i0} \leq p_{i1} \\ M_1^i : p_{i0} > p_{i1} . \end{cases}$$

The marginal likelihood for the null model and the alternative model are

$$\begin{aligned} m(\mathbf{x}_i, \mathbf{n}_i \mid M_0^i) &= \frac{1}{(\text{Beta}(a, b))^2} \int_0^1 \left(\int_0^{p_{i1}} \binom{n_{i1}}{x_{i0}} p_{i0}^{x_{i0}} (1 - p_{i0})^{n_{i0} - x_{i0}} p_{i0}^{a-1} (1 - p_{i0})^{b-1} dp_{i0} \right) \\ &\quad \binom{n_{i1}}{x_{i1}} p_{i1}^{x_{i1}} (1 - p_{i1})^{n_{i1} - x_{i1}} p_{i1}^{a-1} (1 - p_{i1})^{b-1} dp_{i1} , \end{aligned}$$

$$\begin{aligned} m(\mathbf{x}_i, \mathbf{n}_i \mid M_1^i) &= \frac{1}{(\text{Beta}(a, b))^2} \int_0^1 \left(\int_{p_{i1}}^1 \binom{n_{i0}}{x_{i0}} p_{i0}^{x_{i0}} (1 - p_{i0})^{n_{i0} - x_{i0}} p_{i0}^{a-1} (1 - p_{i0})^{b-1} dp_{i0} \right) \\ &\quad \binom{n_{i1}}{x_{i1}} p_{i1}^{x_{i1}} (1 - p_{i1})^{n_{i1} - x_{i1}} p_{i1}^{a-1} (1 - p_{i1})^{b-1} dp_{i1} . \end{aligned}$$

One can compute the resulting Bayes factors, B_{01}^i , numerically. For the uniform prior ($a = b = 1$) on the p_{ij} , these are 2.185, 1.048, 0.147, respectively. The posterior probability of a treatment effect in the 3rd trial, without multiplicity adjustment, is $P(M_1^3 | \mathbf{x}_3, \mathbf{n}_3) = (1 + B_{01}^3)^{-1} = 0.872$. With multiplicity adjustment, the posterior probability is

$$P(M_1^3 | \mathbf{x}_{1:3}, \mathbf{n}_{1:3}) = \left(1 + \frac{p_2}{1 - p_2} B_{01}^3\right)^{-1} = 0.839.$$

Thus the evidence for effectiveness of the 3rd vaccine is somewhat weakened due to failure of first two trials.

4.4 Analysis when only p -values are available

As we are considering a sequence of completely different trials, it may be that the data from each trial is unavailable. Indeed, we may only have the p -value from each trial. Lower bounds on the Bayes factors in each trial can still be obtained, based on Sellke et al. (2001). Indeed, they showed that $B_{01}^i \geq -ep \log p$ when the corresponding p -value is less than $1/e$.

Lemma 4.4.1 (lower bound on p_n). *If $B_{01}^i \geq c_i \forall i$, then*

$$p_n \geq \left(\frac{n+1}{n+2}\right) \left(\prod_{i=1}^n c_i\right) \left(\frac{\prod_{i=1}^n c_i}{\binom{n}{n}} + \frac{\sum_{j=1}^n \prod_{i \neq j} c_i}{\binom{n}{n-1}} + \frac{\sum_{j < k} \prod_{i \neq j, k} c_i}{\binom{n}{n-2}} + \dots + \frac{1}{\binom{n}{0}}\right)^{-1}.$$

Proof.

$$\begin{aligned} p_n &= \frac{\int_0^1 p \prod_1^n (p B_{01}^i + (1-p)) dp}{\int_0^1 \prod_1^n (p B_{01}^i + 1-p) dp} = \frac{\int_0^1 p \prod_1^n \left(p + \frac{1-p}{B_{01}^i}\right) dp}{\int_0^1 \prod_1^n \left(p + \frac{1-p}{B_{01}^i}\right) dp} \\ &\geq \frac{\int_0^1 p \prod_1^n (p+0) dp}{\int_0^1 \prod_1^n \left(p + \frac{1-p}{c_i}\right) dp} = \frac{\frac{1}{n+2} \prod_1^n c_i}{\int_0^1 \prod_1^n (p c_i + 1-p) dp} \\ &= \left(\frac{n+1}{n+2}\right) \left(\prod_i c_i\right) \left(\frac{\prod_{i=1}^n c_i}{\binom{n}{n}} + \frac{\sum_{j=1}^n \prod_{i \neq j} c_i}{\binom{n}{n-1}} + \frac{\sum_{j < k} \prod_{i \neq j, k} c_i}{\binom{n}{n-2}} + \dots + \frac{1}{\binom{n}{0}}\right)^{-1}. \end{aligned}$$

□

We finish with general, but not very useful, bounds on $(1 - p_n)/p_n$.

Corollary 4.4.2 (generic bounds).

$$\frac{1}{n+1} \leq \frac{1-p_n}{p_n} \leq n+1.$$

Proof.

$$p_n = \mathbb{E}(p \mid x_1, \dots, x_n) = \mathbb{E}(p \mid B_{01}^1, \dots, B_{01}^n) = \int_0^1 p \prod_{j=1}^n (pB_{01}^j + 1 - p) dp,$$

$$\frac{\partial p_n}{\partial B_{01}^i} = \int_0^1 p^2 \prod_{j \neq i} (pB_{01}^j + 1 - p) dp,$$

$$\frac{\partial}{\partial B_{01}^i} \left(\frac{1-p_n}{p_n} \right) = \frac{-\frac{\partial p_n}{\partial B_{01}^i}}{p_n^2} = \frac{-\int_0^1 p^2 \prod_{j \neq i} (pB_{01}^j + 1 - p) dp}{p_n^2} \leq 0.$$

Hence, $(1 - p_n)/p_n$ is monotone decreasing with respect to B_{01}^i ,

$$\frac{1}{n+1} = \lim_{B_{01}^i \rightarrow \infty \forall i} \left(\frac{1-p_n}{p_n} \right) \leq \left(\frac{1-p_n}{p_n} \right) \leq \lim_{B_{01}^i \rightarrow 0 \forall i} \left(\frac{1-p_n}{p_n} \right) = n+1.$$

□

Reconciling frequentist and Bayesian methods in sequential endpoint testing

5.1 Introduction

Sequential endpoint testing (or fixed-sequence testing) is about testing a sequence of possible different treatment effects. It was considered by Maurer et al. (1995), Westfall and Krishen (2001), Wiens (2003), and is widely used in clinical trials (Huque and Alosch (2008)). In the original version of this procedure, endpoints are ordered *a priori* and tested sequentially at nominal level α , until a nonsignificant endpoint is found. This procedure controls FWER in the strong sense, but may lose power when stopping too early. In this chapter, we will be discussing the implied multiplicity issue and the possibility of reconciling Bayesian and frequentist methods in sequential endpoint testing.

More formally, the *frequentist sequential endpoint test* proceeds as follows:

- Pre-specify n null and alternative hypotheses pairs, $\{M_0^i, M_1^i\}$, $i \in \{1, \dots, n\}$, ordered in terms of importance (most important first).
- In the i^{th} test ($i < n$), test $\{M_0^i, M_1^i\}$ at level α , continuing to the next test if the i^{th} null model is rejected; if the i^{th} test is not a rejection, stop testing and report overall type-I error (the probability that any rejection is an error) as α .

- If the testing terminates at the last trial, report overall type-I error of α for the rejections that were made.

The startling feature of this test is that there is apparently no penalty for conducting multiple tests. One could conceivably reject all n hypotheses, each at nominal level α , and report that the overall error probability (the chance that any rejection is erroneous) is α .

This is in stark contrast to Bayesian testing. If one has followed the above procedure and has rejected the first m null hypotheses, the Bayesian probability of at least one of the rejections being in error is

$$1 - \prod_{i=1}^m P(M_1^i | x_i),$$

where $P(M_1^i | x_i)$ is the posterior probability of the i^{th} alternative given the data. As m grows, this error probability will go to one. Note that this is true even without introducing a multiplicity adjustment in the Bayesian testing.

The major difference between frequentist and Bayesian testing here is that of conditioning. The frequentist procedure is unconditional, and has overall error α because it is very unlikely to observe a long series of rejections. In contrast the Bayesian procedure is fully conditional, and assesses the error given that a long series of rejections has happened.

This type of frequentist/Bayesian conflict has been addressed in other scenarios by utilizing the conditional frequentist approach (Berger et al. (1994) and Dass and Berger (2003)), showing that, if conditioning is introduced into the frequentist paradigm, then the conflict disappears. The purpose of this chapter is to investigate whether the conditional frequentist paradigm can successfully remove the conflict in sequential endpoint testing. The answer appears to be no, in general.

Section 5.2.2 covers a partition construction on the sequential endpoint sample space when testing two simple hypotheses. Section 5.3 collects several examples explaining when frequentist and Bayesian error are not equivalent.

5.2 Conditional frequentist testing

5.2.1 Testing of two simple hypotheses

Consider testing two simple hypotheses

$$\begin{cases} M_0^i : X_i \text{ has density } f_0(x_i) \\ M_1^i : X_i \text{ has density } f_1(x_i). \end{cases}$$

Then α -level tests are as follows.

- The likelihood ratio test (LRT):

$$\text{reject } M_0^i \text{ if } B_{01}^i(x_i) = \frac{f_0(x_i)}{f_1(x_i)} < k_i,$$

$$\text{where } k_i \text{ is such that } \alpha = P\left(x_i : B_{01}^i(x_i) < k_i \mid M_0^i\right).$$

- Bayesian test: defining $s_i \equiv B_{01}^i(x_i)$, the Bayesian test can be summarized as
 - if $s_i \leq 1$, reject M_0^i and report the conditional error probability $\alpha = \frac{B_{01}^i}{1+B_{01}^i}$;
 - if $s_i \geq 1$, accept M_0^i and report the conditional error probability $\beta = \frac{1}{1+B_{01}^i}$.

Note that the frequentist and Bayesian test statistics are equal in this case. In the following, denote the cumulative density function (cdf) of the null and alternative models by $F_0^i = F_0(x_i)$, $F_1^i = F_1(x_i)$, respectively.

For a single test, Berger et al. (1994) showed the equivalence of the Bayesian procedure and a conditional frequentist procedure. In particular, they defined the conditional frequentist procedure by considering a partition of the sample space, with partitions

$$\chi_{s_1} = \{x_1 \in \chi_1 \mid B_{01}^1(x_1) \in \{s_1, \psi(s_1)\}\} \quad \text{where} \quad \psi(s_1) = (F_0^1)^{-1}(1 - F_1(s_1)),$$

and then defining frequentist error probabilities conditioned on being in a partition.

Theorem 5.2.1 (Berger, Brown, Wolpert, 1994). For $0 < s_1 < 1$,

$$\begin{aligned}\alpha(s_1) = P_0(\text{rejecting } M_0^1 \mid \chi_{s_1}) &= \frac{s_1}{1 + s_1} \\ &= \text{posterior probability of } M_0^1 \text{ when } B_{01}^1 < 1 \\ \beta(s_1) = P_1(\text{accepting } M_0^1 \mid \chi_{s_1}) &= \frac{1}{1 + \psi(s_1)} \\ &= \text{posterior probability of } M_1^1 \text{ when } B_{01}^1 > 1.\end{aligned}$$

Two obvious consequences of the above theorem are that

Corollary 5.2.2.

$$\begin{aligned}P_0(\text{accepting } M_0^1 \mid \chi_s) &= 1 - \frac{s}{1 + s} \\ &= \text{posterior probability of } M_1^1 \text{ when } B_{01} < 1; \\ P_1(\text{rejecting } M_0^1 \mid \chi_s) &= 1 - \frac{1}{1 + \psi(s)} \\ &= \text{posterior probability of } M_0^1 \text{ when } B_{01} > 1.\end{aligned}$$

We first generalize this construction to the sequential endpoint testing scenario.

5.2.2 Conditioning partitions for the sequential endpoint problem

Similar to Theorem 5.2.1, we attempt to establish a conditioning partition for the sequential endpoint sample space:

$$\chi = (X_{1A}) \cup (X_{1R}, X_{2A}) \cup (X_{1R}, X_{2R}, X_{3A}) \cup \cdots \cup (X_{1R}, X_{2R}, \dots, X_{(n-1)R}, X_n),$$

where the X_{iA} , X_{iR} represent the acceptance and rejection regions for test i , respectively.

We will consider two different choices of partitions when $n = 2$.

A fixed partition: A seemingly natural partition to consider begins by defining $\psi(s_i) = (F_0^i)^{-1}(1 - F_1^i(s_i))$, for any $0 < s_i \leq 1$, and then lets the partition be

$$\begin{aligned}\chi_{s_1, s_2} &= \{x \in \chi : B_{01}^i \in \{s_i, \psi(s_i)\}, i \in \{1, 2\}\} \\ &= \{B(x_1) = s_1, B(x_2) = s_2\} \cup \{B(x_1) = s_1, B(x_2) = \psi(s_2)\} \cup \{B(x_1) = \psi(s_1)\}.\end{aligned}$$

Unfortunately, this does not work because $\{B(x_1) = \psi(s_1)\}$ is of a different dimension than the other two parts of the partition. To overcome this, we must partition $\{B(x_1) = \psi(s_1)\}$ further, and an interesting way to do this is to define its subpartitions so that they ‘match up’ with $\{B(x_1) = s_1, B(x_2) = s_2\} \cup \{B(x_1) = s_1, B(x_2) = \psi(s_2)\}$ in terms of probability under the null hypothesis. Specifically, we will partition $\{B(x_1) = \psi(s_1)\}$ into subsets indexed by s_2 , with the conditional mass of the subsets being $f_0^*(s_2) + f_1^*(s_2)$. The ensuing overall fixed partition of the sample space is

$$\begin{aligned}
\chi_{s_1, s_2}^F &= \{x \in \mathcal{X} : B_{01}^i \in \{s_i, \psi(s_i)\}, i \in \{1, 2\}\} \\
&= \{B(x_1) = s_1, B(x_2) = s_2\} \cup \{B(x_1) = s_1, B(x_2) = \psi(s_2)\} \\
&\quad \cup \{B(x_1) = \psi(s_1)\}_{s_2} \\
&= I \cup II \cup III.
\end{aligned} \tag{5.1}$$

This will be considered in Appendix 5.4.1.

A random partition: The fixed partition does not yield results as simple as we had hoped, and so we also consider a random partition, defined by partitioning $\{B(x_1) = \psi(s_1)\}$ into subsets indexed by s_2 , as follows:

$$\{B(x_1) = \psi(s_1)\} = \cup_{0 \leq s_2 \leq 1} \left[\{B(x_1) = \psi(s_1)\}_{s_2} \cup \{B(x_1) = \psi(s_1)\}_{\psi(s_2)} \right],$$

where the $\{B(x_1) = \psi(s_1)\}_{s_2}$ and $\{B(x_1) = \psi(s_1)\}_{\psi(s_2)}$ are chosen to have conditional probabilities proportional to $f_T^*(s_2)$ and $f_T^*(\psi(s_2))$, respectively, where $f_T^*(\cdot)$ refers to the unknown true density. The resulting overall random partition is

$$\begin{aligned}
\chi_{s_1, s_2}^R &= \{x \in \mathcal{X} : B_{01}^i \in \{s_i, \psi(s_i)\}, i \in \{1, 2\}\} \\
&= \{B(x_1) = s_1, B(x_2) = s_2\} \cup \{B(x_1) = s_1, B(x_2) = \psi(s_2)\} \\
&\quad \cup \{B(x_1) = \psi(s_1)\}_{s_2} \cup \{B(x_1) = \psi(s_1)\}_{\psi(s_2)} \\
&= I \cup II \cup IIIA \cup IIIB.
\end{aligned} \tag{5.2}$$

The reason this is called a random partition is that we do not know $f_T^*(s_2)$ and $f_T^*(\psi(s_2))$ and, hence, the partition will change with the true distribution.

Use of a random partition might seem strange, but it does satisfy the key requirements of conditional frequentist inference. First, conditional frequentist error probabilities are still computable; these probabilities are always calculated separately for each given possible true model. Hence, in their computation, the partition is known.

Second, the resulting conditional error measures still satisfy the overall frequentist long run error property, since again that property is defined with respect to a given model.

As a final comment on the use of this random partition, note that the randomness only arises in the way that $\{B(x_1) = \psi(s_1)\}$ is partitioned; this partitioning is rather artificial (it is only x_1 that arises from the sequential sample here), so setting up the partition so that the answer does not depend on the supposedly irrelevant x_2 seems quite appealing.

5.2.3 Conditional frequentist error probabilities for the random partition

In this section, the conditional frequentist error probabilities, given the random partition, are provided. Assuming M_i^1 and M_j^2 are true, we compute $P_{ij}(\text{decision} \mid \chi_{s_1, s_2})$, the error probability for the decision conditioned on χ_{s_1, s_2} , in the following two theorems. This results are also summarized in the following table.

Table 5.1: Frequentist error probabilities, conditional on being in the random partition.

	(s_1, s_2)	$(s_1, \psi(s_2))$	$(\psi(s_1), s_2 \cup \psi(s_2))$
	reject M_0^1 , reject M_0^2	reject M_0^1 , accept M_0^2	accept M_0^1
$P_{00}(\cdot \mid \chi_{s_1, s_2})$	$\left(\frac{s_1}{s_1+1}\right) \left(\frac{s_2}{s_2+1}\right)$	$\left(\frac{s_1}{s_1+1}\right) \left(\frac{1}{s_2+1}\right)$	$\frac{1}{s_1+1}$
$P_{01}(\cdot \mid \chi_{s_1, s_2})$	$\left(\frac{s_1}{s_1+1}\right) \left(\frac{\psi(s_2)}{\psi(s_2)+1}\right)$	$\left(\frac{s_1}{s_1+1}\right) \left(\frac{1}{\psi(s_2)+1}\right)$	$\frac{1}{s_1+1}$
$P_{10}(\cdot \mid \chi_{s_1, s_2})$	$\left(\frac{\psi(s_1)}{\psi(s_1)+1}\right) \left(\frac{s_2}{s_2+1}\right)$	$\left(\frac{\psi(s_1)}{\psi(s_1)+1}\right) \left(\frac{1}{s_2+1}\right)$	$\frac{1}{\psi(s_1)+1}$
$P_{11}(\cdot \mid \chi_{s_1, s_2})$	$\left(\frac{\psi(s_1)}{\psi(s_1)+1}\right) \left(\frac{\psi(s_2)}{\psi(s_2)+1}\right)$	$\left(\frac{\psi(s_1)}{\psi(s_1)+1}\right) \left(\frac{1}{\psi(s_2)+1}\right)$	$\frac{1}{\psi(s_1)+1}$

Theorem 5.2.3 (Type I error).

$$\begin{aligned} P_{00}(\text{rejecting } M_0^1 \mid \chi_{s_1, s_2}) &= P_{00}(\text{rejecting } M_0^1 \vee M_0^2 \mid \chi_{s_1, s_2}) \\ &= \frac{B_{01}^1}{B_{01}^1 + 1} \quad \text{if } B_{01}^1 < 1 \end{aligned}$$

$$\begin{aligned} P_{00}(\text{rejecting } M_0^2 \mid \chi_{s_1, s_2}) &= P_{00}(\text{rejecting } M_0^1 \wedge M_0^2 \mid \chi_{s_1, s_2}) \\ &= \left(\frac{B_{01}^1}{B_{01}^1 + 1} \right) \left(\frac{B_{01}^2}{B_{01}^2 + 1} \right) \quad \text{if } B_{01}^1 < 1; B_{01}^2 < 1 \end{aligned}$$

$$\begin{aligned} P_{10}(\text{rejecting } M_0^2 \mid \chi_{s_1, s_2}) &= P_{10}(\text{rejecting } M_0^1 \wedge M_0^2 \mid \chi_{s_1, s_2}) \\ &= \left(\frac{\psi(s_1)}{1 + \psi(s_1)} \right) \left(\frac{B_{01}^2}{1 + B_{01}^2} \right) \quad \text{if } B_{01}^2 < 1 \end{aligned}$$

$$\begin{aligned} P_{01}(\text{rejecting } M_0^1 \mid \chi_{s_1, s_2}) &= P_{01}(\text{rejecting } M_0^1 \wedge M_0^2 \mid \chi_{s_1, s_2}) \\ &= \left(\frac{B_{01}^1}{B_{01}^1 + 1} \right) \quad \text{if } B_{01}^1 < 1 \end{aligned}$$

$$P_{11}(\text{rejecting } M_0^1 \mid \chi_{s_1, s_2}) = \frac{\psi(s_1)}{1 + \psi(s_1)}$$

$$P_{11}(\text{rejecting } M_0^2 \mid \chi_{s_1, s_2}) = \left(\frac{\psi(s_1)}{1 + \psi(s_1)} \right) \left(\frac{\psi(s_2)}{1 + \psi(s_2)} \right)$$

$$P_{10}(\text{rejecting } M_0^2 \mid \text{rejecting } M_0^1, \chi_{s_1, s_2}) = \left(\frac{B_{01}^2}{B_{01}^2 + 1} \right) \quad \text{if } B_{01}^2 < 1$$

Theorem 5.2.4 (Type II error).

$$\begin{aligned} P_{01}(\textit{accepting } M_0^2 \mid \chi_{s_1, s_2}) &= P_{01}(\textit{rejecting } M_0^1 \cap \textit{accepting } M_0^2 \mid \chi_{s_1, s_2}) \\ &= \left(\frac{B_{01}^1}{B_{01}^1 + 1} \right) \left(\frac{1}{B_{01}^2 + 1} \right) \quad \textit{if } B_{01}^1 < 1; B_{01}^2 > 1 \end{aligned}$$

$$\begin{aligned} P_{11}(\textit{accepting } M_0^2 \mid \chi_{s_1, s_2}) &= P_{11}(\textit{rejecting } M_0^1 \cap \textit{accepting } M_0^2 \mid \chi_{s_1, s_2}) \\ &= \left(\frac{\psi(s_1)}{\psi(s_1) + 1} \right) \left(\frac{1}{B_{01}^2 + 1} \right) \quad \textit{if } B_{01}^2 > 1 \end{aligned}$$

$$P_{10}(\textit{accepting } M_0^1 \mid \chi_{s_1, s_2}) = \left(\frac{1}{B_{01}^1 + 1} \right) \quad \textit{if } B_{01}^1 > 1$$

$$P_{11}(\textit{accepting } M_0^1 \mid \chi_{s_1, s_2}) = \left(\frac{1}{B_{01}^1 + 1} \right) \quad \textit{if } B_{01}^1 > 1$$

$$P_{11}(\textit{accepting } M_0^1 \vee M_0^2 \mid \chi_{s_1, s_2}) = 1 - \left(\frac{B_{01}^1}{B_{01}^1 + 1} \right) \left(\frac{B_{01}^2}{B_{01}^2 + 1} \right) \quad \textit{if } B_{01}^1, B_{01}^2 > 1$$

$$P_{01}(\textit{accepting } M_0^2 \mid \textit{rejecting } M_0^1, \chi_{s_1, s_2}) = \left(\frac{1}{B_{01}^2 + 1} \right) \quad \textit{if } B_{01}^2 > 1$$

Unfortunately, these frequentist errors, conditioned on this random partition, can be quite different from the Bayesian reported errors. In general, constructing appropriate partitions on the sequential endpoint space seems very hard. In the next section, we investigate whether it is possible to construct partitions that lead to Bayesian/frequentist agreement.

5.3 Assessing the possibility of Bayesian/frequentist agreement

One cannot conclude from the failure in the previous section to find a conditional frequentist procedure that agrees with the Bayesian procedure that no such procedure exists. But note that conditional frequentist procedures still have unconditional error probability α so that, if one can be found to match the standard Bayesian test, then the standard Bayesian test should also have expected error probability α (at least approximately), when averaged over the rejection regions. In this section we study if this is so, for a variety of examples.

To begin, the following table gives the decisions and reported errors, for the different

parts of the sample space, in sequential endpoint testing. Here $\alpha_f(\chi_s)$ and $\alpha_B(\chi_s)$ are the reported frequentist and Bayesian error probabilities, respectively.

Table 5.2: Decisions and reported errors in sequential endpoint tests

subset of χ	decision	$\alpha_f(\chi_s)$	$\alpha_B(\chi_s)$
A_1	Accept M_0^1		
(R_1, A_2)	Reject M_0^1 , accept M_0^2	α	$P(M_0^1 x_1)$
(R_1, R_2, A_3)	Reject M_0^1, M_0^2 , accept M_0^3	α	$1 - \prod_{i=1}^2 P(M_1^i x_i)$
\vdots	\vdots	\vdots	\vdots
$(R_1, \dots, R_{n-1}, A_n)$	Reject M_0^1 to M_0^{n-1} , accept M_0^n	α	$1 - \prod_{i=1}^{n-1} P(M_1^i x_i)$
$(R_1, \dots, R_{n-1}, R_n)$	Reject all null models	α	$1 - \prod_{i=1}^n P(M_1^i x_i)$

5.3.1 Testing two simple hypotheses

We first consider the case when $n = 2$. For completeness, we give the frequentist computation showing that the overall error probability of the sequential endpoint test is α . For comparison with the Bayesian procedure, we formally compute this unconditional Type I error as the expected value of the frequentist reported error (α), conditioned on rejecting.

Example 5.3.1 (two endpoints, frequentist).

Table 5.3: Decisions and reported frequentist errors for two endpoints.

partition χ_s	decision	frequentist error	$P(\chi_s)$
A_1	Accept M_0^1		$(1 - \alpha)$
(R_1, A_2)	Reject M_0^1 , accept M_0^2	α	$\alpha(1 - \alpha)$
(R_1, R_2)	Reject M_0^1 , reject M_0^2	α	α^2

Thus the expected frequentist report, conditioned on being in the rejection region, is

$$\begin{aligned}
 \alpha_f &= \mathbb{E}(\alpha \mid \text{rejection}) \\
 &= \frac{1}{\alpha} \left[\int_{A_2} \int_{R_1} \alpha f_0(x_1) f_0(x_2) dx_1 dx_2 + \int_{R_2} \int_{R_1} \alpha f_0(x_1) f_0(x_2) dx_1 dx_2 \right] \\
 &= \frac{1}{\alpha} \left[(1 - \alpha)\alpha^2 + \alpha^3 \right] = \alpha.
 \end{aligned}$$

Next we give the frequentist expectation, conditioned on rejecting, of the reported Bayesian error for two endpoints.

Lemma 5.3.2 (two endpoints, Bayesian). *The expected Bayesian reported error, conditional on the rejection region, is*

$$\alpha_B = (1 - v_1)v_2\alpha + v_1, \quad \text{where} \quad v_i = \frac{1}{\alpha} \int_{R_i} P(M_0^i | x_i) f_0(x_i) dx_i. \quad (5.3)$$

Proof. The decisions and Bayesian reported error probabilities in two dimension are

Table 5.4: Decisions and reported errors

partition χ_s	decision	Bayesian error	$P(\chi_s)$
A_1	Accept M_0^1		$(1 - \alpha)$
(R_1, A_2)	Reject M_0^1 , accept M_0^2	$P(M_0^1 x_1)$	$\alpha(1 - \alpha)$
(R_1, R_2)	Reject M_0^1 , reject M_0^2	$1 - P(M_1^1 x_1)P(M_1^2 x_2)$	α^2

The expectation of this Bayesian error probability, conditioned on the rejection region, is

$$\begin{aligned} \alpha_B &= \mathbb{E}(\text{Bayesian error} \mid \text{rejection}) \\ &= \frac{1}{\alpha} \int_{A_2} \int_{R_1} P(M_0^1 | x_1) f_0(x_1) f_0(x_2) dx_1 dx_2 \\ &\quad + \frac{1}{\alpha} \int_{R_2} \int_{R_1} \left(1 - P(M_1^1 | x_1) P(M_1^2 | x_2) \right) f_0(x_1) f_0(x_2) dx_1 dx_2 \\ &= \frac{1}{\alpha} \left[(1 - \alpha) \int_{R_1} \left(1 - P(M_1^1 | x_1) \right) f_0(x_1) dx_1 \right] \\ &\quad + \frac{1}{\alpha} \left[\alpha^2 - \left(\int_{R_1} P(M_1^1 | x_1) f_0(x_1) dx_1 \right) \left(\int_{R_2} P(M_1^2 | x_2) f_0(x_2) dx_2 \right) \right] \\ &= \frac{1}{\alpha} \left[(1 - \alpha)v_1\alpha + \alpha^2 - (1 - v_1)(1 - v_2)\alpha^2 \right] \\ &= (1 - v_1)v_2\alpha + v_1. \end{aligned}$$

□

In practice, we are interested in small α , e.g. $\alpha = 0.05$. An immediate consequence of (5.3) is that, if v_1 does not converge to zero when α goes to zero, then α_B cannot possibly equal α . Here is an example.

Example 5.3.3 (bounded distributions). Suppose the primary endpoint is testing:

$$\begin{cases} M_0^1 : f_0(x_1) = 1 \text{ where } x_1 \in [0, 1] \\ M_1^1 : f_1(x_1) = 2x \text{ where } x_1 \in [0, 1]. \end{cases}$$

The likelihood ratio is

$$\frac{f_0(x_1)}{f_1(x_1)} = \frac{1}{2x_1}.$$

Therefore, the rejection region is

$$R_1(\alpha) = \{x_1 : x_1 > 1 - \alpha\}; P(R_1 | M_0^1) = \alpha.$$

Since the posterior probability is

$$P(M_0^1 | x_1) = \frac{f_0(x_1)}{f_0(x_1) + f_1(x_1)},$$

$$v_1 = \frac{1}{\alpha} \int_{1-\alpha}^1 \frac{dx}{2x+1} = \frac{1}{2\alpha} \log\left(\frac{3}{3-2\alpha}\right).$$

Now consider the secondary endpoint:

$$\begin{cases} M_0^2 : x_2 \sim Uniform(0, 1) \\ M_1^2 : x_2 \sim Beta(1/2, 1). \end{cases}$$

The likelihood ratio and rejection region are

$$\frac{f_0(x_2)}{f_1(x_2)} = \frac{1}{(2\sqrt{x_2})^{-1}} = 2\sqrt{x_2},$$

$$R_2(\alpha) = \{x_2 : x_2 < \sqrt{\alpha}\}; P(R_2 | M_0^2) = \alpha.$$

Therefore,

$$v_2 = \frac{1}{\alpha} \int_0^{\alpha} \frac{2\sqrt{x}dx}{2\sqrt{x}+1} = 1 - \frac{1}{\alpha} \left[\sqrt{\alpha} - \frac{1}{2} \log(2\sqrt{\alpha} + 1) \right].$$

Notice that

$$\begin{cases} \lim_{\alpha \rightarrow 0} v_1 = 1/3 \\ \lim_{\alpha \rightarrow 0} v_2 = 0. \end{cases}$$

Since $\lim_{\alpha \rightarrow 0} v_1 \rightarrow 0$, from (5.3) it is clear that $\alpha_B(\alpha) > \alpha$ for small α .

This example shows that it is not possible to find a partition on the sequential endpoint testing sample space having the same conditional frequentist and Bayesian errors at every endpoint. Nonetheless, since $\alpha_B > \alpha$, the reported Bayesian error probability is greater than the frequentist error (on average), so that this Bayesian procedure could still be considered a valid frequentist procedure (though overly conservative).

Even if $\lim_{\alpha \rightarrow 0} v_i = 0$, it can still be the case that $\alpha_B > \alpha$ for small α , as the following example shows.

Example 5.3.4 (Gaussian distribution).

$$\begin{cases} M_0^i : f_0(x_i) \sim N(\mu_0, \sigma^2) \\ M_1^i : f_1(x_i) \sim N(\mu_1, \sigma^2) \end{cases} \quad \text{where } \mu_0 < \mu_1,$$

$$R_i(\alpha) = \{x : x > \sigma z_{1-\alpha} + \mu_0\}.$$

Since v_i does not have closed form here, we use the following approximation instead, which is valid for small α :

$$\begin{aligned} v_i &= \frac{1}{\alpha} \int_{R_i(\alpha)} \frac{f_0(x_i)}{f_0(x_i) + f_1(x_i)} f_0 dx_i \\ &= 1 - \frac{1}{\alpha} \int_{R_i(\alpha)} \frac{1}{1 + f_0/f_1} f_0 dx_i \\ &= 1 - \frac{1}{\alpha} \int_{R_i(\alpha)} f_0 dx + \frac{1}{\alpha} \int_{R_i(\alpha)} \frac{f_0^2}{f_1} dx + O\left(\frac{1}{\alpha} \int_{R_i(\alpha)} \frac{f_0^3}{f_1^2} dx\right) \\ &= \frac{1}{\alpha} \int_{R_i(\alpha)} \frac{f_0^2}{f_1} dx + O\left(\frac{1}{\alpha} \int_{R_i(\alpha)} \frac{f_0^3}{f_1^2} dx\right). \end{aligned} \tag{5.4}$$

Also,

$$\begin{aligned} \frac{1}{\alpha} \int_{R_i(\alpha)} f_0^2/f_1 dx &= \frac{1}{\alpha} \int_{\sigma z_{1-\alpha} + \mu_0}^{\infty} \exp\{-1/(2\sigma^2)[2(x - \mu_0)^2 - (x - \mu_1)^2]\} dx \\ &= \frac{1}{\alpha} \exp\{(\mu_0 - \mu_1)^2/\sigma^2\} \left(1 - \Phi(z_{1-\alpha} - \mu_0 + \mu_1)\right). \end{aligned} \tag{5.5}$$

By Fact 5.4.1 and Lemma 5.4.2,

$$\begin{aligned}
& 1 - \Phi(z_{1-\alpha} + \mu_1 - \mu_0) \\
&= \frac{\frac{1}{\sqrt{2\pi}} e^{-(\mu_1 - \mu_0)/2} \alpha \sqrt{2\pi} \sqrt{-2 \log(\alpha \sqrt{2\pi})}}{\sqrt{\mu_1 - \mu_0 - 2 \log(\alpha \sqrt{2\pi}) - \log(-2 \log(\alpha \sqrt{2\pi}))}} (1 + o(1)) \\
&= \frac{e^{-(\mu_1 - \mu_0)/2} \alpha \sqrt{2\pi}}{\sqrt{1 + \frac{\mu_1 - \mu_0}{-2 \log(\alpha \sqrt{2\pi})} - \frac{\log(-2 \log(\alpha \sqrt{2\pi}))}{\log(\alpha \sqrt{2\pi})}}} (1 + o(1)) \\
&= O(\alpha)
\end{aligned}$$

Thus, (5.5) is greater than α when α is small $\Rightarrow v_i = O(1) \Rightarrow \alpha_B > \alpha$.

Similarly, with the exponential distribution (testing $f_0(x_i) = \lambda_0 \exp\{-\lambda_0 x_i\}$ versus $f_1(x_i) = \lambda_1 \exp\{-\lambda_1 x_i\}$, $\lambda_1 > \lambda_0$), it can be shown that $\alpha_B > \alpha$ for small α . Thus we have several examples involving testing simple hypotheses where the Bayesian and conditional frequentist cannot possibly report the same error probabilities in sequential endpoint testing. In the next section, we investigate composite hypotheses.

5.3.2 Testing composite hypotheses

For composite hypothesis testing:

$$\begin{cases} M_0^i : x_i \sim f_0(x_i | \theta_0); \theta \in \Theta_0 \\ M_1^i : x_i \sim f_1(x_i | \theta_1); \theta \in \Theta_1. \end{cases} \quad (5.6)$$

Given priors on the unknown parameters θ_0, θ_1 and equal prior probability on both models ($=1/2$), the posterior probability of the null hypothesis is

$$P(M_0^i | x_i, \pi_0, \pi_1) = \frac{m_0(x_i)}{m_1(x_i) + m_0(x_i)},$$

where the marginal likelihood is

$$m_j(x_i) = \int_{\Theta_j} f_j(x_i | \theta) \pi_j(\theta) d\theta_j \quad j \in \{0, 1\}.$$

In particular, when $f_0 = f_1$; $\pi_0 = \pi_1$; $\Theta_0 = (-\infty, 0)$; $\Theta_1 = (0, \infty)$, (5.6) is a one-sided test. The following theorem shows that, using the standard constant prior, the average Bayesian null posterior probability over the rejection region is equal to $\alpha/2$, which is anti-conservative.

Theorem 5.3.5 (one-sided hypothesis testing). *Suppose*

$$\begin{cases} M_0^i : f(x_i - \theta), \theta \leq 0, \\ M_1^i : f(x_i - \theta), \theta > 0, \end{cases}$$

where f is symmetric about zero and has monotone likelihood ratio (MLR) with constant prior $\pi(\theta) = 1$. Then the expected Bayesian error probability, conditional on rejection, satisfies

$$\frac{1}{\alpha} \int_{R_i(\alpha)} P(M_0^i | x_i) f(x_i) dx_i = \alpha/2.$$

Proof. By Casella and Berger (1987),

$$P(M_0^i | x_i) = p(x_i),$$

where $p(x_i)$ is the p -value. Since f has MLR, without loss of generality, let $R_i(\alpha) = \{x : x > c\}$, so that

$$\begin{aligned} \int_{R_i(\alpha)} p(x_i) f_0(x_i) dx_i &= \int_c^\infty \left(\int_x^\infty f_0(t) dt \right) f_0(x_i) dx_i \\ &= \int_c^\infty (1 - F_0(x_i)) f_0(x_i) dx_i \\ &= \alpha - \int_{F_0^{-1}(c)}^{F_0^{-1}(\infty)} F_0(x_i) dF_0(x_i) \\ &= \alpha - (1 - (1 - \alpha)^2)/2 = \alpha^2/2. \end{aligned}$$

The theorem follows by noticing that

$$\frac{1}{\alpha} \int_{R_i(\alpha)} P(M_0^i | x_i) f_0(x_i) dx_i = \frac{1}{\alpha} \int_{R_i(\alpha)} p(x_i) f_0(x_i) dx_i = \alpha/2.$$

□

While the Bayesian report here is the same as the p -value, neither are valid frequentist reports, both having expected value, conditional on the rejection region, of $\alpha/2$; i.e., both are anti-conservative.

Remark 5.3.6. Casella and Berger also showed that

$$\inf_{\pi \in \Gamma_s} P(M_0^i | x_i, \pi) = p(x_i)$$

where $\Gamma_s = \{\text{all distributions symmetric about zero}\}$ and the infimum is attained when the prior is constant. Hence, under the same assumptions,

$$\inf_{\pi \in \Gamma_s} \frac{1}{\alpha} \int_{R_i(\alpha)} P(M_0^i | x_i, \pi) f(x_i) dx_i = \alpha/2.$$

Corollary 5.3.7. *If the primary endpoint satisfies the assumptions in Theorem 5.3.5, then the average Bayesian error, conditional on rejecting, is a proper frequentist procedure if and only if the secondary experiment has*

$$v_2 > \frac{1/2}{1 - \alpha/2}. \quad (5.7)$$

Proof. By Theorem 5.3.5 and (5.3),

$$\alpha_B = \frac{\alpha/2}{\alpha/2 + (1 - \alpha/2)(1 - v_2)} > \frac{\alpha/2}{\alpha/2 + (1/2 - \alpha/2)} = \alpha.$$

□

When testing two-sided hypotheses, the p -value is much smaller than the Bayesian error probability (Sellke et al. (2001)), so that it might well be the case that α_B is greater than α for two-sided tests. Thus we finish with the example of a two-sided test with the Gaussian distribution. This example is similar to that in Example 5.3.4.

Example 5.3.8 (two-sided hypothesis testing, Gaussian distribution).

$$\begin{cases} M_0^i : f_0(x_i) \sim N(0, \sigma^2) \\ M_1^i : f_1(x_i) \sim N(\mu, \sigma^2); \mu \neq 0 \end{cases}$$

Suppose $\mu \sim N(0, \tau^2)$, so that the marginal likelihoods are

$$\begin{aligned} & \begin{cases} m_0(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-x_i^2/(2\sigma^2)\} \\ m_1(x_i) = \frac{1}{\sqrt{2\pi(\sigma^2+\tau^2)}} \exp\{-x_i^2/(2\sigma^2 + 2\tau^2)\}, \end{cases} \\ & R_i(\alpha) = \{x_i : |x_i| > \sigma z_{1-\alpha/2}\}, \\ v_i &= \frac{1}{\alpha} \int_{R_i(\alpha)} \frac{1}{1 + m_0/m_1} f_0(x_i) dx_i \\ &= \frac{1}{\alpha\sqrt{2\pi}} \int_{|x_i| > \sigma z_{1-\alpha/2}} \sqrt{\frac{\sigma^2 + \tau^2}{\sigma^2}} \exp\left\{\frac{-x_i^2}{2\sigma^2} \left[\frac{\tau^2}{\sigma^2 + \tau^2} + 1\right]\right\} dx_i \\ &= \frac{2(\sigma^2 + \tau^2)}{\alpha\sqrt{2\tau^2 + \sigma^2}} \left[1 - \Phi\left(z_{1-\alpha/2} \sqrt{\frac{\tau^2}{\sigma^2 + \tau^2} + 1}\right)\right]. \end{aligned}$$

By Lemma 5.4.2,

$$\begin{aligned} & 1 - \Phi(x_i) = \\ & \frac{\exp\left\{\left(\frac{\tau^2}{\sigma^2 + \tau^2} + 1\right) \log(\alpha\sqrt{\frac{\pi}{2}})\right\} \exp\left\{\left(\frac{\tau^2}{\tau^2 + \sigma^2} + 1\right) \log\left(-2 \log(\alpha\sqrt{\frac{\pi}{2}})\right)\right\}}{\sqrt{2\pi} \sqrt{-2 \log(\sqrt{\frac{\pi}{2}}\alpha) - \log(-2 \log(\alpha\sqrt{\frac{\pi}{2}}))}} (1 + o(1)) \\ &= \frac{\left(\alpha\sqrt{\frac{\pi}{2}}\right)^{\tau^2/(\tau^2 + \sigma^2) + 1} \left(-2 \log(\alpha\sqrt{\frac{\pi}{2}})\right)^{\tau^2/(\sigma^2 + \tau^2)}}{\sqrt{2\pi} \sqrt{1 + \frac{\log(-2 \log(\alpha\sqrt{\frac{\pi}{2}}))}{-\log(\alpha\sqrt{\frac{\pi}{2}})}}} (1 + o(1)) \\ &= O(\alpha^{\tau^2/(\sigma^2 + \tau^2) + 1} (-\log \alpha)^{\tau^2/(\sigma^2 + \tau^2)}) (1 + o(1)). \end{aligned}$$

Therefore, v_i is greater than α for small α , showing that α_B is a conservative frequentist report.

These examples show that virtually anything is possible in sequential endpoint testing; the Bayesian answers can be conservative frequentist reports or anti-conservative frequentist reports. Obtaining exact agreement appears almost impossible: first one would need to find a (peculiar) situation in which the Bayesian procedure has exact average reported error of α , and then one would need to find a partitioning of the sample space so that the conditional frequentist error probabilities equal the Bayesian error probabilities.

5.4 Appendix

5.4.1 Conditional frequentist analysis for the fixed partition

In this section, we derive conditional frequentist error probabilities using the fixed partition (5.1).

Table 5.5: Decision rules

region	I	II	III
$(B_{01}^1(x_1), B_{01}^2(x_2))$	(s_1, s_2)	$(s_1, \psi(s_2))$	$(\psi(s_1), s_2)$
decision	reject M_0^1 & M_0^2	reject M_0^1 , accept M_0^2	accept M_0^1

Table 5.6: Conditional probabilities up to a normalizing constant

	I	II	III
$P_{00}(\cdot \chi_{s_1, s_2}) \propto$	$f_0^*(s_1)f_0^*(s_2)$	$f_0^*(s_1)f_1^*(s_2)$	$f_1^*(s_1)(f_0^*(s_2) + f_1^*(s_2))$
$P_{01}(\cdot \chi_{s_1, s_2}) \propto$	$f_0^*(s_1)f_1^*(s_2)$	$f_0^*(s_1)\frac{1}{\psi(s_2)}f_1^*(s_2)$	$f_1^*(s_1)(f_0^*(s_2) + f_1^*(s_2))$
$P_{10}(\cdot \chi_{s_1, s_2}) \propto$	$f_1^*(s_1)f_0^*(s_2)$	$f_1^*(s_1)f_1^*(s_2)$	$f_1^*(s_1)\frac{1}{\psi(s_1)}(f_0^*(s_2) + f_1^*(s_2))$
$P_{11}(\cdot \chi_{s_1, s_2}) \propto$	$f_1^*(s_1)f_1^*(s_2)$	$f_1^*(s_1)\frac{1}{\psi(s_2)}f_1^*(s_2)$	$f_1^*(s_1)\frac{1}{\psi(s_1)}(f_0^*(s_2) + f_1^*(s_2))$

Error probabilities

- $$P_{00}(\text{reject } M_0^1 \vee M_0^2 | \chi_{s_1, s_2}) = P_{00}(\text{reject } M_0^1 | \chi_{s_1, s_2}) = \frac{I+II}{I+II+III} = \frac{f_0^*(s_1)}{f_0^*(s_1)+f_1^*(s_1)} = \frac{B_{01}^1}{B_{01}^1+1}$$
- $$P_{01}(\text{reject } M_0^1 | \chi_{s_1, s_2}) = \frac{I+II}{I+II+III} = \frac{f_0^*(s_1)}{f_0^*(s_1)+f_1^*(s_1)\left(\frac{1+s_2}{1+1/\psi(s_2)}\right)}$$
- $$P_{10}(\text{reject } M_0^2 | \chi_{s_1, s_2}) = \frac{I}{I+II+III} = \frac{f_0^*(s_2)}{\left(1+\frac{1}{\psi(s_1)}\right)(f_0^*(s_2)+f_1^*(s_2))} \leq \frac{f_0^*(s_2)}{f_0^*(s_2)+f_1^*(s_2)}$$
- $$P_{11}(\text{falsely reject } M_0^1 \vee M_0^2 | \chi_{s_1, s_2} \text{ serve}) = 0$$

Power

1. Power: Make all correct rejection

- $P_{00}(\text{correctly reject } M_0^1, M_0^2 \mid \chi_{s_1, s_2}) = 0$
- $P_{01}(\text{reject } M_0^2 \mid \chi_{s_1, s_2}) = \frac{II}{I+II+III} = \frac{\frac{1}{\psi(s_2)}}{1 + \frac{1}{\psi(s_2)} + \frac{1+s_2}{s_1}}$
- $P_{10}(\text{reject } M_0^1 \mid \chi_{s_1, s_2}) = \frac{I+II}{I+II+III} = \frac{1}{1 + \frac{1}{\psi(s_1)}} = \frac{\psi(s_1)}{\psi(s_1)+1} = \frac{B_{01}^1}{B_{01}^1+1}$
- $P_{11}(\text{reject } M_0^1 \wedge M_0^2 \mid \chi_{s_1, s_2}) = \frac{I}{I+II+III} = \frac{1}{1 + \frac{1}{\psi(s_2)} + \frac{1}{\psi(s_1)}(s_2+1)} = \frac{1}{1 + \frac{1}{\psi(s_2)} + \frac{1}{\psi(s_1)} + \frac{s_2}{\psi(s_1)}}$

2. Marginalized power: $P(\text{reject } M_0^i \mid M_1^i)$

5.4.2 *Conditional frequentist test*

Fact 5.4.1 (Normal tail probability). Let $\Phi(t)$ be cumulative distribution function of standard normal, then

$$\frac{t \frac{1}{\sqrt{2\pi}} e^{-t^2/2}}{t^2 + 1} \leq 1 - \Phi(t) \leq \frac{\frac{1}{\sqrt{2\pi}} e^{-t^2/2}}{t}$$

$$1 - \Phi(t) = \frac{\frac{1}{\sqrt{2\pi}} e^{-t^2/2}}{t} + O\left(\frac{e^{-t^2/2}}{t^3}\right)$$

Proof can be found in Durrett (2010).

Lemma 5.4.2 (Asymptote of Gaussian quantile). Let z_p be the p -quantile of the standard Gaussian distribution i.e.

$$p = \Phi(z_p) \text{ where } \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-x^2/2} dx$$

then

$$z_{1-\alpha} = \sqrt{-2 \log(\sqrt{2\pi}\alpha) - \log(-2 \log(\sqrt{2\pi}\alpha))} (1 + o(1))$$

Proof. By Fact 5.4.1, the ratio of normal tail probability and α is:

$$\begin{aligned}
\frac{1 - \Phi(z_{1-\alpha})}{\alpha} &= \frac{\phi(z_{1-\alpha})}{z_{1-\alpha}} \frac{1}{\alpha} \\
&= \frac{\frac{1}{\sqrt{2\pi}} \exp\{\log\sqrt{2\pi\alpha}\} \exp\{\frac{1}{2}\log\sqrt{2\pi\alpha}\}}{\alpha\sqrt{-2\log(\sqrt{2\pi\alpha}) - \log(-2\log(\sqrt{2\pi\alpha}))}(1 + o(1))} \\
&= \left(1 + \frac{\log(-2\log\sqrt{2\pi\alpha})}{2\log\sqrt{2\pi\alpha}}\right)^{-1} (1 + o(1)) \\
&= 1 + o(1)
\end{aligned}$$

□

Theorem 5.4.3. *Suppose there is no constant $c > 0$ such that*

$$P\left(x_i \mid f_0(x_i) < cf_1(x_i), M_0^i\right) = 1$$

then

$$\lim_{\alpha \rightarrow 0} v_i(\alpha) = 0$$

Proof. Under the null model, v_i , the average of null posterior over the rejection region at i^{th} endpoint, is bounded above by:

$$v_i = \left(\frac{1}{\alpha} \int_{x_i: f_0 < k_i(\alpha)f_1} \frac{1}{1 + f_1/f_0} f_0 dx_i\right) < \left(\frac{1}{\alpha} \frac{k_i(\alpha)}{1 + k_i(\alpha)} \int_{x_i: f_0 < k_i(\alpha)f_1} f_0 dx_i\right) = \frac{k_i(\alpha)}{1 + k_i(\alpha)}.$$

Note that $v_i > 0$ and the upper v_i from Remark 5.4.2, if $\lim_{\alpha \rightarrow 0} k(\alpha) = 0$, then the bounds force $v_i(\alpha)$ to 0, i.e. $\lim_{\alpha \rightarrow 0} v_i(\alpha) = 0$.

To prove $\lim_{\alpha \rightarrow 0} k(\alpha) = 0$, let

$$G(c) = \int_{x_i: f_0 < cf_1} f_0 dx_i$$

which has several properties by definition: (1) $G(0) = 0$ (2) G is monotonic increasing. (3) $\alpha = G(k(\alpha))$.

Since

$$\lim_{\alpha \rightarrow 0} \alpha = \lim_{\alpha \rightarrow 0} G(k(\alpha)) = 0$$

And by assumption, $G(c)$ can be zero only when $c = 0$. Thus, $\lim_{\alpha \rightarrow 0} k(\alpha) = 0$. \square

The exponential distribution satisfies Theorem 5.4.3.

Bibliography

- Abramovich, F. and Angelini, C. (2006), “Bayesian maximum a posteriori multiple testing procedure,” *Sankhyā: The Indian Journal of Statistics*, pp. 436–460.
- Abramovich, F., Benjamini, Y., Donoho, D. L., Johnstone, I. M., et al. (2006), “Adapting to unknown sparsity by controlling the false discovery rate,” *The Annals of Statistics*, 34, 584–653.
- Alosh, M., Bretz, F., and Huque, M. (2014), “Advanced multiplicity adjustment methods in clinical trials,” *Statistics in medicine*, 33, 693–713.
- Banks, D. et al. (2011), “Reproducible research: A range of response,” *Statistics, Politics, and Policy*, 2, 4.
- Benjamini, Y. and Hochberg, Y. (1995), “Controlling the false discovery rate: a practical and powerful approach to multiple testing,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 289–300.
- Berger, J. O. (1985), *Statistical Decision Theory and Bayesian Analysis*, Springer Science & Business Media.
- Berger, J. O., Brown, L. D., and Wolpert, R. L. (1994), “A unified conditional frequentist and Bayesian test for fixed and sequential simple hypothesis testing,” *The Annals of Statistics*, pp. 1787–1807.
- Berger, J. O., Wang, X., and Shen, L. (2014), “A Bayesian approach to subgroup identification,” *Journal of biopharmaceutical statistics*, 24, 110–129.
- Bogdan, M., Ghosh, J. K., Ochman, A., and Tokdar, S. T. (2007), “On the Empirical Bayes approach to the problem of multiple testing,” *Quality and Reliability Engineering International*, 23, 727–739.
- Bogdan, M., Ghosh, J. K., Tokdar, S. T., et al. (2008a), “A comparison of the Benjamini-Hochberg procedure with some Bayesian rules for multiple testing,” in *Beyond parameters in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen*, pp. 211–230, Institute of Mathematical Statistics.
- Bogdan, M., Frommlet, F., Biecek, P., Cheng, R., Ghosh, J. K., and Doerge, R. (2008b), “Extending the Modified Bayesian Information Criterion (mBIC) to dense markers and multiple interval mapping,” *Biometrics*, 64, 1162–1169.

- Bogdan, M., Chakrabarti, A., Frommlet, F., and Ghosh, J. K. (2011), “Asymptotic Bayes-optimality under sparsity of some multiple testing procedures,” *The Annals of Statistics*, pp. 1551–1579.
- Bretz, F., Maurer, W., Brannath, W., and Posch, M. (2009), “A graphical approach to sequentially rejective multiple test procedures,” *Statistics in medicine*, 28, 586.
- Cai, T., Liu, W., and Xia, Y. (2013), “Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings,” *Journal of the American Statistical Association*, 108, 265–277.
- Cai, T. T. and Xia, Y. (2014), “High-dimensional sparse MANOVA,” *Journal of Multivariate Analysis*, 131, 174–196.
- Casella, G. and Berger, R. L. (1987), “Reconciling Bayesian and frequentist evidence in the one-sided testing problem,” *Journal of the American Statistical Association*, 82, 106–111.
- Chipman, H., George, E. I., McCulloch, R. E., Clyde, M., Foster, D. P., and Stine, R. A. (2001), “The practical implementation of Bayesian model selection,” *Lecture Notes-Monograph Series*, pp. 65–134.
- Clyde, M. A., Ghosh, J., and Littman, M. L. (2011), “Bayesian adaptive sampling for variable selection and model averaging,” *Journal of Computational and Graphical Statistics*, 20.
- Cui, W. and George, E. I. (2008), “Empirical Bayes vs. fully Bayes variable selection,” *Journal of Statistical Planning and Inference*, 138, 888–900.
- Dass, S. C. and Berger, J. O. (2003), “Unified conditional frequentist and Bayesian testing of composite hypotheses,” *Scandinavian Journal of Statistics*, 30, 193–210.
- Datta, J., Ghosh, J. K., et al. (2013), “Asymptotic properties of Bayes risk for the horseshoe prior,” *Bayesian Analysis*, 8, 111–132.
- Dmitrienko, A., Tamhane, A. C., and Bretz, F. (2009), *Multiple testing problems in pharmaceutical statistics*, CRC Press.
- Dudoit, S. and Van Der Laan, M. J. (2007), *Multiple testing procedures with applications to genomics*, Springer Science & Business Media.
- Durrett, R. (2010), *Probability: theory and examples*, vol. 3, Cambridge university press.
- Dutta, R., Bogdan, M., and Ghosh, J. K. (2015), “Model Selection and Multiple Testing- A Bayesian and Empirical Bayes Overview and some New Results,” *arXiv preprint arXiv:1510.00547*.
- Efron, B. (2004), “Large-scale simultaneous hypothesis testing,” *Journal of the American Statistical Association*, 99.

- Efron, B., Tibshirani, R., Storey, J. D., and Tusher, V. (2001), “Empirical Bayes analysis of a microarray experiment,” *Journal of the American statistical association*, 96, 1151–1160.
- Ferguson, T. S. (2003), “U-statistics,” *Notes for Statistics*.
- Ge, Y., Dudoit, S., and Speed, T. P. (2003), “Resampling-based multiple testing for microarray data analysis,” *Test*, 12, 1–77.
- Genovese, C. R., Lazar, N. A., and Nichols, T. (2002), “Thresholding of statistical maps in functional neuroimaging using the false discovery rate,” *Neuroimage*, 15, 870–878.
- George, E. I. and McCulloch, R. E. (1993), “Variable selection via Gibbs sampling,” *Journal of the American Statistical Association*, 88, 881–889.
- Gilbert, P. B., Berger, J. O., Stablein, D., Becker, S., Essex, M., Hammer, S. M., Kim, J. H., and DeGruttola, V. G. (2011), “Statistical interpretation of the RV144 HIV vaccine efficacy trial in Thailand: a case study for statistical issues in efficacy trials,” *Journal of Infectious Diseases*, 203, 969–975.
- Goeman, J. J. and Solari, A. (2014), “Multiple hypothesis testing in genomics,” *Statistics in medicine*, 33, 1946–1978.
- Guindani, M., Müller, P., and Zhang, S. (2009), “A Bayesian discovery procedure,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71, 905–925.
- Guo, M. and Heitjan, D. F. (2010), “Multiplicity-calibrated Bayesian hypothesis tests,” *Biostatistics*, 11, 473–483.
- Heller, R., Yekutieli, D., et al. (2014), “Replicability analysis for genome-wide association studies,” *The Annals of Applied Statistics*, 8, 481–498.
- Holm, S. (1979), “A simple sequentially rejective multiple test procedure,” *Scandinavian journal of statistics*, pp. 65–70.
- Huque, M. F. and Alosch, M. (2008), “A flexible fixed-sequence testing method for hierarchically ordered correlated multiple endpoints in clinical trials,” *Journal of Statistical Planning and Inference*, 138, 321–335.
- Ioannidis, J. P. (2005), “Why most published research findings are false,” *Chance*, 18, 40–47.
- Karr, A. F., Sanil, A. P., and Banks, D. L. (2006), “Data quality: A statistical perspective,” *Statistical Methodology*, 3, 137–173.
- Malloy, M. and Nowak, R. (2011), “Sequential analysis in high-dimensional multiple testing and sparse recovery,” in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pp. 2661–2665, IEEE.

- Maurer, W., Hothorn, L., and Lehmacher, W. (1995), “Multiple comparisons in drug clinical trials and preclinical assays: a-priori ordered hypotheses,” *Biometrie in der chemisch-pharmazeutischen Industrie*, 6, 3–18.
- Miller, C. J., Genovese, C., Nichol, R. C., Wasserman, L., Connolly, A., Reichart, D., Hopkins, A., Schneider, J., and Moore, A. (2001), “Controlling the false-discovery rate in astrophysical data analysis,” *The Astronomical Journal*, 122, 3492.
- Muller, P., Parmigiani, G., and Rice, K. (2006), “FDR and Bayesian multiple comparisons rules,” .
- Noble, W. S. (2009), “How does multiple testing correction work?” *Nature biotechnology*, 27, 1135–1137.
- O’Brien, P. C. and Fleming, T. R. (1979), “A multiple testing procedure for clinical trials,” *Biometrics*, pp. 549–556.
- Scott, J. G. (2009), “Nonparametric Bayesian multiple testing for longitudinal performance stratification,” *The Annals of Applied Statistics*, pp. 1655–1674.
- Scott, J. G. and Berger, J. O. (2006), “An exploration of aspects of Bayesian multiple testing,” *Journal of Statistical Planning and Inference*, 136, 2144–2162.
- Scott, J. G. and Berger, J. O. (2010), “Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem,” *The Annals of Statistics*, 38, 2587–2619.
- Sellke, T., Bayarri, M., and Berger, J. O. (2001), “Calibration of ρ values for testing precise null hypotheses,” *The American Statistician*, 55, 62–71.
- Siegmund, D. (1985), *Sequential Analysis: Tests and Confidence Intervals*, Springer Science & Business Media.
- Storey, J. D. (2003), “The positive false discovery rate: A Bayesian interpretation and the q-value,” *Annals of Statistics*, pp. 2013–2035.
- Storey, J. D. and Tibshirani, R. (2003), “Statistical significance for genomewide studies,” *Proceedings of the National Academy of Sciences*, 100, 9440–9445.
- Streiner, D. L. (2015), “Best (but oft-forgotten) practices: the multiple problems of multiplicity whether and how to correct for many statistical tests,” *The American Journal of Clinical Nutrition*, 102, 721–728.
- Tony Cai, T., Liu, W., and Xia, Y. (2014), “Two-sample test of high dimensional means under dependence,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76, 349–372.
- Wang, R., Lagakos, S. W., Ware, J. H., Hunter, D. J., and Drazen, J. M. (2007), “Statistics in medicine reporting of subgroup analyses in clinical trials,” *New England Journal of Medicine*, 357, 2189–2194.

- Westfall, P. H. and Krishen, A. (2001), “Optimally weighted, fixed sequence and gatekeeper multiple testing procedures,” *Journal of Statistical Planning and Inference*, 99, 25–40.
- Westfall, P. H. and Young, S. S. (1993), *Resampling-based multiple testing: Examples and methods for p-value adjustment*, vol. 279, John Wiley & Sons.
- White, J. T. and Ghosal, S. (2014), “Multiple testing approaches for removing background noise from images,” in *Topics in Nonparametric Statistics*, pp. 95–104, Springer.
- Whitehead, J. (1997), *The design and analysis of sequential clinical trials*, John Wiley & Sons.
- Wiens, B. L. (2003), “A fixed sequence Bonferroni procedure for testing multiple endpoints,” *Pharmaceutical Statistics*, 2, 211–215.

Biography

Shih-Han Chang was born August 31, 1987 in Taipei, Taiwan. He married Dr. Chien-Kuang Ding in 2015. Chang received his B.Sc.in Mathematics from National Taiwan University in 2010, M.A. in Mathematics from Duke University in 2013, and Ph.D. in 2015 under the supervision of professor James Berger. Under his advisors mentorship, he established Bayesian procedures on false positive probability of mutually exclusive hypotheses under the scenario of data dependence and investigated Asymptotic behavior in Bayesian multiple testing and sequential endpoint tests.

During summers at the graduate school, Chang worked at Facebook (Menlo Park, CA), Goldman Sachs (London, UK), and Verisk Analytics (San Francisco, CA). His roles included Data Scientist and Risk Management Associate. He is interested in Bayesian statistics and machine learning and motivated to pursue a career as a data scientist. He was on the Dean's List with Honors at National Taiwan University, awarded the Study Abroad Scholarship from the Ministry of Education in Taiwan. During his graduate studies at Duke, he represented Duke for 2015 National College Table Tennis Championships and participated in the Triangle Healthcare Challenge (2015), where his team won both the Grand Prize and the Validic mHealth prize.