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
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

5 Reconciling frequentist and Bayesian methods in sequential endpoint testing

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Introduction</td>
<td>79</td>
</tr>
<tr>
<td>5.2 Conditional frequentist testing</td>
<td>81</td>
</tr>
<tr>
<td>5.2.1 Testing of two simple hypotheses</td>
<td>81</td>
</tr>
<tr>
<td>5.2.2 Conditioning partitions for the sequential endpoint problem</td>
<td>82</td>
</tr>
<tr>
<td>5.2.3 Conditional frequentist error probabilities for the random partition</td>
<td>84</td>
</tr>
<tr>
<td>5.3 Assessing the possibility of Bayesian/frequentist agreement</td>
<td>86</td>
</tr>
<tr>
<td>5.3.1 Testing two simple hypotheses</td>
<td>87</td>
</tr>
<tr>
<td>5.3.2 Testing composite hypotheses</td>
<td>91</td>
</tr>
<tr>
<td>5.4 Appendix</td>
<td>95</td>
</tr>
<tr>
<td>5.4.1 Conditional frequentist analysis for the fixed partition</td>
<td>95</td>
</tr>
<tr>
<td>5.4.2 Conditional frequentist test</td>
<td>96</td>
</tr>
</tbody>
</table>

Bibliography

Biography
## 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 $\tau, r$, different $\rho$. 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 $\tau$ under the null model, for fixed $n = 2000, \rho = r = 0.5$. 25

2.3 Convergence of $P(M_0 \mid x)$ 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 $\frac{1}{k^p} - \frac{1}{2}$ (y-axis) versus $\frac{p}{(1-p)(1-r)}$ (x-axis). ................. 35

2.7 $\hat{\tau}^2$ versus $\theta_i$ 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 $\theta_i$ 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 $\theta$ in the x-axis range. ......................... 37

2.9 $\theta_i$ 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 $k_{\text{max}} = 4$, $\tau = 2$, 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 $k_{\text{max}} = 2$, 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 $P(\mu_i \neq 0 \mid X)$ versus signal size $\theta_i$. $m = 100$, $k_{\text{max}} = 3$, $\tau = 1$. For each $\theta_i$, 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, $k_{\text{max}}$ is indicated in the subtitle, and simulate data $x$ 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 \( \mathbf{X} \).
\( x \) a vector of observations.
\( \gamma_i \) an indicator for nonzero means.
\( \gamma \) \( (\gamma_1, \ldots, \gamma_n) \)
\( M_0^i \) \( i^{th} \) 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.
\( \alpha \) Type-I error.
\( \rho \) 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.
1

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 $\alpha$ 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>
<thead>
<tr>
<th>Fail to Reject Null</th>
<th>Null True</th>
<th>Alternative True</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{Reject Null} )</td>
<td>( U )</td>
<td>( T )</td>
<td>( n - R )</td>
</tr>
<tr>
<td>( \text{Reject Null} )</td>
<td>( V )</td>
<td>( S )</td>
<td>( R )</td>
</tr>
<tr>
<td>( \text{Fail to Reject Null} )</td>
<td>( N_0 )</td>
<td>( n - N_0 )</td>
<td>( n )</td>
</tr>
</tbody>
</table>

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

\[
PCER \leq FDR \leq pFDR \leq FWER \leq PFER, \tag{1.1}
\]

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 $\alpha/n$, 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 Sidak’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 $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(n)}$ as the ordered raw $p$-values, and rejects a hypothesis if:

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

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

\[
\tilde{p}(i) \leq \min_{k \in \{1, \ldots, n\}} \{\min(n p(k)/k, 1)\}.
\]

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

\[
F(z) = p_0 F_0(z) + (1 - p_0) F_1(z), \tag{1.2}
\]

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 $X = (X_1, \ldots X_n)'$ arises from the multivariate normal distribution

$$X \sim \text{multinorm} \left( \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{pmatrix}, \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} \right), \quad (2.1)$$
where $\rho$ is the correlation between the observations. Consider testing the $n$ hypotheses $M_i^0 : \theta_i = 0$ versus $M_i^1 : \theta_i \neq 0$, 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 $\rho$ 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 $n \to \infty$, 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 $\theta_i$.

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

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

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

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

\[\square\]

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 \( \rho \to 1 \), \( \Phi(c) \to 1 - \frac{\alpha}{2} \), 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

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

If \( \rho \to 1 \), by Lemma 2.21,

\[
\lim_{\rho \to 1} P\left( \max_{1 \leq j \leq n} |X_j| > c \mid \theta_i = 0 \ \forall i \right)
\]

\[
= 1 - \lim_{\rho \to 1} \mathbb{E}_{Z_1, \ldots, Z_n} \left\{ P\left( |\sqrt{n}Z + \sqrt{1-\rho}Z_j| < c \mid Z_1, \ldots, Z_n \right) \right\}
\]

\[
= 1 - (\Phi(c) - \Phi(-c))
\]

\[
= 2 \left(1 - \Phi(c)\right).
\]

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

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

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

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

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

Computation yields, defining \( u_i = \sum_{j \neq i} x_j \),

\[
\hat{\theta}_i = \arg \max_{\theta_i} -\frac{1}{2} \tilde{x}_i^T \Sigma_0^{-1} \tilde{x}_i = x_i + \frac{b}{a} u_i,
\]
from which it is immediate that

$$LR = \min_i \frac{\exp \left\{ -\frac{1}{2} \left( (a - b)(\sum_i^n x_i^2) + b(\sum_j^n x_j)^2 \right) \right\}}{\exp \left\{ -\frac{1}{2} \left( (a - b)(\frac{b^2}{a^2} u_i^2 + \sum_{j \neq i}^n x_j^2) + b(-\frac{b}{a} u_i + \sum_{j \neq i}^n x_j^2) \right) \right\}}$$

$$= \min_i \exp \left\{ -\frac{1}{2} \left[ (a - b)(x_i^2 - \frac{b^2}{a^2} u_i^2) + b((u_i + x_i)^2 - u_i^2(\frac{b}{a} - 1)^2) \right] \right\}$$

(since $\sum_i^n x_j = u_i + x_i$)

$$= \min_i \exp \left\{ -\frac{1}{2} \left( ax_i^2 + 2bu_i x_i + \frac{b^2}{a} u_i^2 \right) \right\}$$

$$= \min_i \exp \left\{ -\frac{1}{2a} (ax_i + bu_i)^2 \right\}.$$ 

Noting that

$$\frac{1}{a} (ax_j + bu_j)^2$$

$$= \frac{1}{(1 + (n - 1)\rho)(1 + (n - 2)\rho)(1 - \rho)} (1 + (n - 2)\rho)x_j - \rho \sum_{k \neq j} x_k)^2$$

$$= \frac{1}{(1 + (n - 1)\rho)(1 + (n - 2)\rho)(1 - \rho)} [(1 - \rho)x_j + n\rho(x_j - \bar{x})]^2$$

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

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 $\rho \to 1$, $T \approx \max_i n^2 \left( \frac{x_i - \bar{x}}{\sqrt{1 - \rho}} \right)^2$, 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

$$M_0 : \theta_1 = \ldots = \theta_n = 0 \quad \text{(null model)}$$

$$M_i : \theta_i \neq 0, \theta_{(-i)} = 0$$

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

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

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

initially we will assume $\tau^2$ to be known, but later will consider it to be unknown. Then under model $M_i$, the marginal likelihood of model $M_i$ is

$$m_0(x) \sim N(0, \Sigma_0)$$

$$m_i(x) = \int f(x \mid \theta) \pi(\theta) d\theta \sim N(0, \Sigma_i) ,$$

(2.5)

where

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

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

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

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:

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

15
and the posterior probability of an alternative model $M_i$ is

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

Proof. The posterior probability of $M_i$ is

$$
P(M_i | x) = \frac{m_i(x) P(M_i)}{\sum_{j=0}^{n} m_j(x) P(M_j)} = \frac{\frac{1 - r}{\tau_{i}} |\Sigma_i|^\frac{1}{2} \exp \left\{ \frac{-1}{2} x' \Sigma_i^{-1} x \right\}}{r |\Sigma_0|^\frac{1}{2} \exp \left\{ \frac{-1}{2} x' \Sigma_0^{-1} x \right\} + \sum_{j=1}^{n} \frac{1 - r}{\tau_{j}} |\Sigma_j|^\frac{1}{2} \exp \left\{ \frac{-1}{2} x' \Sigma_j^{-1} x \right\}}
$$

$$
= \left\{ \begin{array}{l} \left( \frac{n r}{1 - r} \right) |\Sigma_i|^\frac{-1}{2} \exp \left\{ \frac{-1}{2} x' (\Sigma_0^{-1} - \Sigma_i^{-1}) x \right\} \\
+ 1 + \sum_{j \neq i}^{n} \exp \left\{ \frac{-1}{2} x' (\Sigma_j^{-1} - \Sigma_i^{-1}) x \right\} \end{array} \right\}^{-1}
$$

(2.6)

The expression can be simplified by further computing $\Sigma_i^{-1}, (\Sigma_i^{-1} - \Sigma_k^{-1})$ and $\text{det}(\Sigma_i)$. First notice that by the Cholesky decomposition

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

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

$$
\Sigma_i^{-1} = (\Sigma_0 + \tau_i \tau_i^T)^{-1} = \Sigma_0^{-1} - \Sigma_0^{-1} \tau_i (1 + \tau_i^T \Sigma_0^{-1} \tau_i)^{-1} \tau_i^T \Sigma_0^{-1}
$$

$$
= \Sigma_0^{-1} \left( 1 + \frac{-\tau^2}{1 + \tau^2 a} \Sigma_0^{-1} \right) \begin{pmatrix} b & \cdots & b \\
  b & a & b & \cdots & b \\
  \cdots & \cdots & \cdots & \cdots & \cdots \\
  b & b & \cdots & a & b \\
  0 & 0 & \cdots & 0 & 0 \end{pmatrix}
$$

where $\tau_i = (0, \cdots, \tau, \cdots, 0)^T$ (the $i^{th}$ element is $\tau^2$).
Therefore,

\[ x'(\Sigma_0^{-1} - \Sigma_i^{-1})x = \frac{\tau^2}{1 + \tau^2a} \left( x_i(a - b) + bn\bar{x} \right)^2 = \frac{\tau^2}{1 + \tau^2a} \left( \frac{x_i}{1 - \rho} + bn\bar{x} \right)^2 \]

\[ x'(\Sigma_k^{-1} - \Sigma_i^{-1})x = \frac{\tau^2}{1 + \tau^2a} (x_i^2 - x_k^2)(a - b)^2 + 2b(a - b)(n\bar{x})(x_i - x_k). \]

Also the ratio of two determinants is

\[ \frac{\det(\Sigma)}{\det(\Sigma_0)} = \frac{\det(\Sigma_1)}{\det(\Sigma_0)} = \det(I + \Sigma_0^{-1} \tau_1 \tau_1^T) \]

\[ = \det(I + LL^T \tau_1 \tau_1^T) \]

\[ = \det(I + \tau_1^T LL^T \tau_1) \]

\[ = (1 + \tau^2 L_{11}) \]

\[ = (1 + \tau^2a). \]

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

Remark 2.3.2. (2.24) gives the full expression for \( P(M_i \mid x) \), 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:

\[ P(M_0 \mid x) = \]

\[ \left\{ 1 + \frac{1 - r}{2r} \sqrt{\frac{1 - \rho^2}{1 - \rho^2 + \tau^2}} \sum_{i \in \{1,2\}} \exp \left\{ \frac{\tau^2}{2(1 - \rho^2 + \tau^2)} \left( \frac{x_i - \rho x_{(-i)}}{\sqrt{1 - \rho^2}} \right)^2 \right\} \right\}^{-1}, \]

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

\[ P(M_i \mid x) = \left\{ \sqrt{\frac{1 - \rho^2 + \tau^2}{1 - \rho^2}} \left( \frac{2r}{1 - r} \exp \left\{ \frac{-\tau^2}{2(1 - \rho^2 + \tau^2)} \left( \frac{x_{(-i)} - \rho x_i}{1 - \rho^2} \right)^2 \right\} + 1 \right\} \right\}^{-1}. \]

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 \( n = 2, \ i \in \{1, 2\} \), \( \rho \to 1 \), then:

\[
P(M_0 \mid X) \to 1 \quad \text{under null model},
\]

\[
P(M_i \mid X) \to \left(1 + \exp \left\{ \frac{-1}{2} \left( X^2_i - X^2_{(-i)} \right) \right\} \right)^{-1} \quad \text{under } M_1 \text{ or } M_2.
\]

When \( n > 2, \ i, j \in \{0, 1, \ldots, n\} \), \( \rho \to 1 \), under model \( M_j \):

\[
P(M_i \mid X) \to \hat{\delta}_i = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{else}. \end{cases}
\]

**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:**

\[
\frac{x_i - \rho x_{(-i)}}{\sqrt{1 - \rho^2}} = \frac{(\theta_i - \theta_{(-i)}) + \sqrt{1 - \rho}(z_i - \rho z_{(-i)}) + z \sqrt{\rho}(1 - \rho)}{\sqrt{1 - \rho^2}}
= \frac{\theta_i - \theta_{(-i)}}{\sqrt{1 - \rho^2}} + \frac{z_i - \rho z_{(-i)}}{\sqrt{1 + \rho}} + \sqrt{1 - \rho} \frac{\sqrt{\rho} z}{\sqrt{1 + \rho}}.
\] (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:

\[
P(M_0 \mid x) = (1 + O(\sqrt{1 - \rho^2})) = 1 + o(1).
\]

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

\[
P(M_i \mid x) = \left\{ \sqrt{\frac{\sigma^2}{2(1 - \rho^2)}} \left( \frac{2x}{1 - \sigma^2} \right) \exp \left\{ \frac{-1}{2} \frac{\theta_i^2}{2(1 - \rho^2)} + O\left( \frac{1}{\sqrt{1 - \rho^2}} \right) \right\} \right\}^{-1}
= \left( \frac{1}{1 + \exp\left\{ \frac{-1}{2} \theta_i(\theta_i + 2z) \right\}} \right)(1 + o(1))
= \frac{1}{1 + \exp\left\{ \frac{-1}{2} \left( x_i^2 - x_{(-i)}^2 \right) \right\}}(1 + o(1)),
\]

18
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 $z = (z_1, ..., z_n)$, $\bar{z} = 1/n \sum_{i} z_i$;

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

$$P(M_i \mid x)$$

$$= \frac{1}{\sqrt{c(1-\rho) \left( \frac{n}{1-r} \right)^\ast}} \exp \left\{ \frac{-\tau^2}{2} c \left( z_i - \bar{z} \right) + \frac{\sqrt{1-\rho}}{1+(n-1)\rho} \left( \sqrt{\rho z} + \sqrt{1-\rho} \bar{z} \right)^2 \right\} + 1 + \frac{O(\sqrt{1-\rho})}{1+(n-1)\rho}$$

$$\sum_{k \neq i} \exp \left\{ \frac{-\tau^2}{2} c \left( \frac{(z_i+z_k-2\bar{z})(z_i-z_k)}{1-\rho} + \frac{2(1-\rho)\bar{z}(z_i-z_k) + \sqrt{\rho(1-\rho)} z(z_i-z_k)}{1+(n-1)\rho} \right) \right\}$$

$$= \left\{ \frac{\sqrt{\tau^2(n-1)/n}}{1-\rho} \left( \frac{n}{1-r} \right) \exp \left\{ \frac{-n}{2(n-1)} (z_i - \bar{z})^2 \right\} + 1 + (1+o(1)) \right\}^{-1}$$

$$= \left\{ \frac{\sqrt{\tau^2(n-1)/n}}{1-\rho} \left( \frac{n}{1-r} \right) \exp \left\{ \frac{-n}{2(n-1)} (z_i - \bar{z})^2 \right\} \right\}^{-1} \left( 1 + o(1) \right)$$

$$= \sqrt{1-\rho} \sqrt{\frac{1-r}{r \tau^2}} (1 + o(1)) \to 0.$$
If $M_j$ is true ($j > 0$): Take $\theta_k = 0 \forall k \neq j$ in (2.24):

$$P(M_j \mid \mathbf{x})$$

$$= \left\{ \sqrt{\frac{1}{c(1-\rho)}} \left( \frac{\nu r}{n-1} \right) \exp \left\{ \frac{-\nu^2 c}{2(1-\rho)} \left( (1-2/n)\theta_j + \sqrt{1-\rho}(z_i - \bar{z}) + O(1-\rho) \right)^2 \right\} \right\}^{-1}$$

$$+ 1 + \sum_{k \neq j} \exp \left\{ \frac{-\nu^2 c}{2(1-\rho)} \left( (1-2/n)\theta_j + \sqrt{1-\rho}(z_i + z_k - 2\bar{z}) \right) + O(1) \right\}$$

$$= \left\{ \sqrt{\frac{n-1+\nu^2 c}{n-1-\rho}} \left( \frac{\nu r}{n-1} \right) \exp \left\{ \frac{-\nu^2 c}{2(n-1)} \left( (1-2/n)\theta_j^2 + O(1) \right) \right\} + 1 \right\}^{-1}$$

$$\rightarrow 1 \text{ since } \lim_{\rho \to 1} \sqrt{1/(1-\rho)} \exp(-1/(1-\rho)) = 0.$$  

\[\square\]

**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 $\rho \to 1$ and $n > 2$, but (as with the Bayesian test) is not fully powered when $n = 2$.

**Proof.** By Lemma 2.22:

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

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

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

$$= 2 \lim_{\rho \to 1} \max_{j \in \{1,2\}} \rho \left( \frac{x_j - x_{(-j)}}{2\sqrt{1-\rho}} \right)^2,$$

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

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

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

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 $n = 2$, $\rho \to 1$:

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

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 $n > 2$, $\rho \to 1$: 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 $\theta_i$ or $\theta_j$ is nonzero.

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

- still has probability $\alpha$ 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 \( n \to \infty \) under the null model,

\[
P(M_i \mid x) = \left( 1 + \frac{n}{1-\tau} \sqrt{\frac{1-\rho + \tau^2}{1-\rho}} \exp \left\{ \frac{-\tau^2}{2(1-\rho + \tau^2)} \left( \frac{x_i - \bar{x}}{\sqrt{1-\rho}} \right)^2 \right\} \right)^{-1} (1 + o(1))
\]  

(2.8)

almost surely.

**Proof.** Take \( \theta = 0 \) in (2.24):

\[
P(M_i \mid x)
\]

\[
= \left\{ \sqrt{\frac{1}{c(1-\rho)} \left( \frac{n \tau}{1-\tau} \right)} \exp \left\{ \frac{-\tau^2}{2} \right\} \left( z_i - \bar{z} + \frac{\sqrt{1-\rho}}{1+(n-1)\rho} (\sqrt{\rho z + \sqrt{1-\rho} \bar{z}}) \right)^2 \right\}^{-1} + \left. \sum_{k \neq i} \exp \left\{ \frac{-\tau^2}{2} \left( z_i + z_k - 2\bar{z}(z_i - z_k) + 2 \left( \sqrt{\rho z + \sqrt{1-\rho} \bar{z}} \right) \left( \frac{\sqrt{1-\rho}(z_i - z_k)}{1+(n-1)\rho} \right) \right) \right\} \right| I
\]

(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:

\[
I = \left( 1 - 1/n \right)^2 \frac{z_i^2}{O(n^{-1})} + \frac{\bar{z}_{(-i)}^2}{O(n^{-\epsilon})} - 2 \left( 1 - 1/n \right) \frac{z_i \bar{z}_{(-i)}}{O(n^{-\epsilon})} \left( \text{where } \bar{z}_{(-i)} = 1/n \sum_{k \neq i} z_k \right)
\]

\[
+ 2 \left( \frac{\sqrt{1-\rho}}{1+(n-1)\rho} \right) \left( 1 - 1/n \right) \frac{z_i - \bar{z}_{(-i)}}{O(n^{-1})} \frac{\sqrt{\rho z + \sqrt{1-\rho} \bar{z}}}{O(n^{-1/2-\epsilon})} \frac{(\sqrt{1-\rho}(z_i - z_k))}{O(n^{-1})}
\]

\[
+ \frac{\sqrt{1-\rho}}{1+(n-1)\rho} \right)^2 \frac{\sqrt{\rho z + \sqrt{1-\rho} \bar{z}}}{O(n^{-2})}
\]

\[
= (1 - 1/n)^2 z_i^2 + O(n^{-\epsilon}) .
\]
Therefore,
\[
\sqrt{\frac{1}{c(1-\rho)} \left( \frac{n \tau}{1-r} \right)} \exp \left\{ -\frac{\tau^2}{2} c I \right\} \\
= \sqrt{\frac{1}{c(1-\rho)} \left( \frac{n \tau}{1-r} \right)} \exp \left\{ -\frac{\tau^2}{2} c (1-1/n)^2 z_i^2 \right\} \left( 1 + O(n^{-\epsilon}) \right) \tag{2.10}
\]
\[
= \sqrt{\frac{1-\rho+\tau^2}{1-\rho} \left( \frac{n \tau}{1-r} \right)} \exp \left\{ -\frac{\tau^2}{2(1-\rho+\tau^2)} z_i^2 \right\} \left( 1 + o(1) \right).
\]

\[ II = (z_i^2 - z_k^2) (1 - 2/n) - \left( \frac{2}{n} \sum_{i \neq k} \frac{z_i}{O(n^{1/2-\epsilon})} \right) (z_i - z_k) \left( \frac{1}{O(n^{1/2-\epsilon})} \right) \]
\[
+ \left( \frac{2}{1 + (n-1)\rho} \right) \left\{ \frac{\sqrt{\rho(1-\rho)} z_i z_k}{O(n^{1/2-\epsilon})} + (1-\rho) \frac{z_i/n(-z_k)}{O(n^{-1/2-\epsilon})} \right\}
+ \left( 1 - \rho \right) \frac{z_i^2/n}{O(n^{1/2-\epsilon})} \left( \frac{\sqrt{\rho z} + \sqrt{1-\rho} z_i}{O(n^{1/2-\epsilon})} \right) \left( \frac{z_k}{O(n^{1/2-\epsilon})} \right)
\]
\[
= (z_i^2 - z_k^2) (1 - 2/n) + O(n^{-\epsilon}).
\]

The summation term in (2.9) becomes:
\[
\sum_{k \neq i}^n \exp \left\{ -\frac{\tau^2}{2} c II \right\}
\]
\[
= \exp \left\{ -c \frac{\tau^2}{2} (1 - 2/n) z_i^2 \right\} (1 + O(n^{-\epsilon})) \sum_{k=1}^n \exp \left\{ c \frac{\tau^2}{2} (1 - 2/n) z_k^2 \right\}
\]
\[
= \exp \left\{ -\frac{\tau^2}{2(1-\rho+\tau^2)} z_i^2 \right\} n \left( E^Z \left[ \exp \left\{ \frac{\tau^2}{2(1-\rho+\tau^2)} Z^2 \right\} \right] + o(1) \right) \left( 1 + o(1) \right) \tag{2.11}
\]
by Law of Large Number and \( Z \sim N(0, 1) \)
\[
= \exp \left\{ -\frac{\tau^2}{2(1-\rho+\tau^2)} z_i^2 \right\} n \sqrt{\frac{1-\rho+\tau^2}{1-\rho}} (1 + o(1)).
\]

The proof is completed by summing (2.10), (2.11).
Remark 2.5.2. Note that, by Lemma 2.22,

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

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

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

Remark 2.5.3. Figure 2.1 shows the ratio of the estimated \( P(M_1 \mid x) \) (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 \( \tau \). 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 \( \tau, r \), different \( \rho \). 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 $\tau$. Notice that, for fixed $n$, the estimated probability is closer to the true probability when $\tau$ is small but is worse for larger $\tau$, indicating that larger $n$ is required for obtaining the same precision as $\tau$ grows.

![Figure 2.2](image)

**Figure 2.2**: Estimated (red line) and true posterior probability (blue line) of $M_1$ for different $\tau$ 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 \( n \to \infty \) and \( \rho \in [0, 1) \), under the null model,

\[
P(M_0 \mid X) \to P(M_0)
\]

**Proof.** First note that

\[
\begin{align*}
a_n &= \frac{1}{r-\rho} + \frac{-\rho}{(1-\rho)(1+(n-1)\rho)} = \frac{1}{r-\rho} + O(1/n), \\
b_n &= \frac{1}{r-\rho} + \frac{-\rho}{(1-\rho)(1+(n-1)\rho)} = \frac{1}{r-\rho} + O(1/n).
\end{align*}
\]

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

\[
\left( \frac{1-r}{nr} \right) \frac{1}{\sqrt{1 + \tau^2/(1-\rho)}} \sum_1^n \exp \left\{ \frac{\tau^2}{2(1 + \tau^2/(1-\rho))} \left[ \frac{x_i - \bar{x}}{1-\rho} \right]^2 \right\} (1 + o(1))
\]

\[
= \left( \frac{1-r}{r} \right) 1/n \sqrt{1 - \rho + \tau^2} \sum_1^n \exp \left\{ \frac{\tau^2}{2(1 - \rho + \tau^2)} z_i^2 \right\} (1 + o(1)) \quad \text{(by Lemma 2.22)}
\]

\[
\to \frac{1-r}{r} \quad \text{(by the Strong Law of Large Numbers)}.
\]

Therefore, \( P(M_0 \mid X) \to (1 + (1-r)/r)^{-1} = r = P(M_0) \).

\[\square\]

**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.
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 $P(M_i \mid \mathbf{x})$, 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 $n \to \infty$,

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

Figure 2.3: Convergence of $P(M_0 \mid \mathbf{x})$ to the prior probability (0.5) under the null model. Each subplot has a different correlation and contains 50 simulations.
Proof. Under the null model, by (2.8), \( P(M_i \mid x) \gg p \) is equivalent to

\[
\begin{align*}
z_i^2 & \geq 2 \left( \frac{1 - \rho + \tau^2}{\tau^2} \right) \ln \left( \frac{n}{1 - r} \frac{p}{1 - p} \sqrt{\frac{1 - \rho + \tau^2}{1 - \rho}} \right) (1 + o(1)) \\
& = 2 \left( \frac{1 - \rho + \tau^2}{\tau^2} \right) \ln \left( \frac{n}{1 - r} \frac{p}{1 - p} \sqrt{\frac{1 - \rho + \tau^2}{1 - \rho}} \right) + o(1).
\end{align*}
\]

(2.14)

By Fact 5.4.1:

\[
P\left( |Z_i| > \gamma_n \right) \geq \frac{2 \left( \frac{1 - \rho + \tau^2}{\tau^2} \right) \ln \left( \frac{n}{1 - r} \frac{p}{1 - p} \sqrt{\frac{1 - \rho + \tau^2}{1 - \rho}} \right) + o(1)}{\gamma_n}
\]

\[
= \frac{1}{n} \left( \frac{2 n^{1 - \rho} \frac{1 - r}{r} p \sqrt{\frac{1 - \rho + \tau^2}{1 - \rho}}}{n} \left( 1 + o(1) \right) \right) + O\left( \frac{1}{n (\log n)^2} \right).
\]

\[\text{P(any false positive \mid M_0) = 1 - \prod_{i=1}^{n} P(|Z_i| < \gamma_n) = 1 - \prod_{i=1}^{n} \left( 1 - \frac{d_n^2}{n} \right) = 1 - (1 - d_n) + O(d_n^2) = 1 - \left( 1 - \frac{\log \frac{n}{1 - \tau} + \log \left( \frac{p}{1 - p} \sqrt{\frac{1 - \rho + \tau^2}{1 - \rho}} \right) }{\frac{1}{\tau}} \right) (1 + o(1)) = O(n^{-1} (\log n)^{-1}) \]
surely indication that assuming we know $\tau^2$ 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 $\tau^2$ with that prior variance which maximizes the marginal likelihood over all possible $\tau^2$; see Berger (1985) for discussion of this approach.

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

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

from which it follows that

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

**Proof.**

$$
\tilde{L}_n = r |\Sigma_0|^{-\frac{1}{2}} \exp \left\{ \frac{1}{2} x' \Sigma^{-1} x \right\} + \frac{1 - r}{n} |\Sigma_1|^{-\frac{1}{2}} \sum_{i=1}^{n} \exp \left\{ \frac{1}{2} x' \Sigma^{-1} x \right\}
$$

$$
= r |\Sigma_0|^{-\frac{1}{2}} \exp \left\{ \frac{1}{2} x' \Sigma_0^{-1} x \right\} + \frac{1 - r}{n} |\Sigma_0|^{-\frac{1}{2}} (1 + \tau^2 a)^{-\frac{1}{2}}
$$

$$
\times \exp \left\{ \frac{0}{2} x' \Sigma_0^{-1} x \right\} \sum_{i=1}^{n} \exp \left\{ \frac{\tau^2}{2(1 + \tau^2 a)} \left( x_i(a - b) + bn \bar{x} \right)^2 \right\}
$$

$$
= |\Sigma_0|^{-\frac{1}{2}} \exp \left\{ \frac{1}{2} x' \Sigma_0^{-1} x \right\} \left\{ r + \frac{1 - r}{n} \left( 1 + \tau^2 a \right)^{-\frac{1}{2}} \right\}
$$

$$
\times \left\{ \sum_{i=1}^{n} \exp \left\{ \frac{\tau^2}{2(1 + \tau^2 a)} \left( x_i(a - b) + bn \bar{x} \right)^2 \right\} \right\}.
$$

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

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

$$
\hat{\tau}_n^2 = \arg \max_{\tau^2} \tilde{L}_n(\tau^2) = \arg \max_{\tau^2} L_n(\tau^2)
$$

$$
= \arg \max_{\tau^2} \sqrt{\frac{1 - \rho}{1 - \rho + \tau_n^2}} \frac{1}{n} \sum_{i=1}^{n} \exp \left\{ \frac{\hat{\tau}_n^2 x_i^2}{2(1 - \rho + \tau_n^2)} \right\} (1 + o(1)) \text{ by (2.25)}.
$$
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 \( \rho \), and decision threshold \( p \), as \( n \to \infty \):

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

where \( k^* \) satisfies:

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

**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:

\[
z_1^2 > 2 \left( 1 + \frac{1 - \rho}{\hat{\tau}^2_n} \right) \log \left( n \frac{p}{(1 - p)(1 - r)} \sqrt{\frac{1 - \rho + \hat{\tau}^2_n}{1 - \rho}} \right) + o(1)
\]

\[
= 2 \log n + \log \left( \frac{1 - \rho + \hat{\tau}^2_n}{1 - \rho} \right) + 2 \log \left( \frac{p}{(1 - p)(1 - r)} \right) \tag{2.16}
\]

\[
+ 2(1 - \rho) \log \frac{1 - \rho + \hat{\tau}^2_n}{\hat{\tau}^2_n} + \frac{1 - \rho}{\hat{\tau}^2_n} \log \left( \frac{1 - \rho + \hat{\tau}^2_n}{1 - \rho} \right) + \frac{(1 - \rho)}{\hat{\tau}^2_n} u + o(1),
\]

\[
\lim_{n \to \infty} L_n(\hat{\tau}^2_n) = \lim_{n \to \infty} \left( \sqrt{\frac{1 - \rho}{1 - \rho + \hat{\tau}^2_n}} \right)^{1/n} \exp \left\{ \frac{\hat{\tau}^2_n z_1^2}{2(1 - \rho + \hat{\tau}^2_n)} \right\} +
\]

\[
\sqrt{\frac{1 - \rho}{1 - \rho + \hat{\tau}^2_n}} \sum_{i=2}^n \exp \left\{ \frac{\hat{\tau}^2_n z_i^2}{2(1 - \rho + \hat{\tau}^2_n)} \right\} (1 + o(1)).
\]

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

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

30
\[ I = \frac{1}{n \sqrt{n \log n}} \exp \left\{ \frac{\hat{\kappa}^2}{n \log n} + \frac{\hat{\kappa}^2}{n \log n} \log n + \frac{\hat{\kappa}^2}{n \log n} \log n + \frac{\hat{\kappa}^2}{n \log n} \log n + \frac{\hat{\kappa}^2 u}{n \log n} \right\} (1 + o(1)) \]

\[ = O((\hat{\kappa}^2 n)^{-1/(1 - \rho/2 + \hat{\kappa}^2 n)}) = o(1) \text{ since} \]

\[ \log(\hat{\kappa}^2 n)^{-1/(1 - \rho/2 + \hat{\kappa}^2 n)} = \frac{-(1 - \rho)}{2(1 - \rho + \hat{\kappa}^2 n)} \log \hat{\kappa}^2 n \to 0. \]

\( II \to 1 \) by Corollary 2.10.7.

If \( \frac{\hat{\kappa}^2}{\log n} \to \infty, L_n(\hat{\kappa}^2 n) \to e^{u/2} \) since

\[ z_1^2 = 2 \log n + \log \hat{\kappa}^2 n + u + o(1), \]

\[ I = (\hat{\kappa}^2 n)^{-1/2/n} \exp \left\{ \frac{\hat{\kappa}^2}{n \log n} \log n + \frac{\hat{\kappa}^2}{n \log n} \log n + \frac{\hat{\kappa}^2}{n \log n} \log n + \frac{\hat{\kappa}^2 u}{n \log n} \right\} (1 + o(1)) \]

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

\[ \begin{cases} \log(\hat{\kappa}^2 n)^{-1/(1 - \rho/2 + \hat{\kappa}^2 n)} = \frac{-(1 - \rho)}{2(1 - \rho + \hat{\kappa}^2 n)} \log \hat{\kappa}^2 n \to 0, \\ \log((n^{(1 - \rho)/2} + \hat{\kappa}^2 n^{(1 - \rho)/2})) = -\frac{1 - \rho}{1 - \rho + \hat{\kappa}^2 n} \log n \to 0. \end{cases} \]

\( II \to 0 \) by Corollary 2.10.7.

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

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

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

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

\[ \to e^{u/2} \text{ since} \]

\[ \begin{cases} \log((k \log n)^{-1/(2 + k \log n)}) = -\frac{\log \log n}{2 \log n} \to 0, \\ \log((n^{-1/(1 + k \log n)}) = -\frac{\log n}{1 + k \log n} \to -1/k. \end{cases} \]

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

Choose \( \epsilon = 0.5 \min\{1, e^{u/2}\} \). Because (1) \( I + II \to e^{u/2} + 2\Phi(\sqrt{2/k}) - 1, \) (2) \( \lim_{k \to 0} 2\Phi(\sqrt{2/k}) - 1 = 1, \) 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,

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

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

\[ c^* = \log k^* + 2 \log u + 2/k^* \]

\[ = -\log \left( \frac{1}{2} + \frac{\exp \left\{ -c^*/2 \right\}}{\sqrt{\pi}} \right) + 2 \log u + 1 + \frac{2 \exp \left\{ -c^*/2 \right\}}{\sqrt{\pi}}, \tag{2.17} \]

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

\[ c^*(k^*) = -2 \log \left( \sqrt{\pi} (1/k^* - 1/2) \right), \]

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

\[
= \frac{1}{\sqrt{2\pi}} \exp \left\{ \frac{-1}{2} \left( 1 + \frac{1}{\hat{\tau}^2_n} \right) \log \left( \frac{n}{1-p} \frac{p}{1-p} \right) \right\} + o(1) \]

\[
= 1 - \frac{1}{2} \left( 1 + \frac{1}{\hat{\tau}^2_n} \right) \log \left( \frac{n}{1-p} \frac{p}{1-p} \right) + o(1) \]

\[
= 1 - \frac{1}{\sqrt{2\pi}} \exp \left\{ \frac{-1}{2} \left( 1 + \frac{1}{\hat{\tau}^2_n} \right) \log \left( \frac{n}{1-p} \frac{p}{1-p} \right) \right\} + o(1) \]

\[
= d_n(1 + o(1)) + O\left( \frac{1}{(\log n)^2} \right) \]

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 \( \tau^2 \). In addition, the asymptotic FPP does not depend on \( \rho \).
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.

\[
\frac{1}{k^* - \frac{1}{2}} \log n
\]

\[
d_n = 1 - (1 - d_n)^n
\]

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 $\frac{p}{(1-p)(1-r)}$ (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 $\theta_i$ is nonzero. Figure 2.7 shows the numerical solution for the Type II ML $\hat{\tau}^2$, as a function of $\theta$, for the specified scenario.

![Figure 2.7: $\hat{\tau}^2$ versus $\theta_i$ 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 $\tau^2 > 1$.](image)

Figure 2.7 demonstrates the how the detection power (computed numerically) varies when the signal size increases, for the specified situation.
Figure 2.8: Power versus $\theta_i$ 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 $\theta$ in the x-axis range.

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

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

(2.18)
Hence, $m$ can be seen as the precision of $\bar{X}$. More generally, we will replace $1/m$ by a general function $\sigma_n^2$, where $\sigma_n^2$ 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:

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

(2.19)

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

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

Under $M_0$: $P(M_0 \mid X) \to (1 + \frac{1-r}{r} [2\Phi(\frac{1-r}{\tau^2}) - 1])^{-1}$.

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

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

$$P(M_0 \mid X) \to P(M_0).$$

**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 $\theta_i^*$ has prior $N(0, \tau^2/\sigma_n^2)$ and the generalized model 2.19 becomes:

$$X^* \sim \text{multinorm} \left( \begin{pmatrix} \theta_1^* \\ \theta_2^* \\ \vdots \\ \theta_n^* \end{pmatrix}, \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} \right).$$
Under the null model: by Theorem 2.3.1,

\[ P(M_0 \mid x)^{-1} = 1 + \left( \frac{1 - r}{nr} \right) \sqrt{\frac{\sigma_n^2}{\sigma_n^2 + \tau^2 a_n^*}} \sum_{i=1}^{n} \exp \left\{ \frac{\tau^2}{2(\sigma_n^2 + \tau^2 a_n)} \left( \frac{z_i}{\sqrt{1 - \rho}} + \sqrt{\rho z} \right)^2 + \frac{1}{(1 - \rho)} \right\} \]

\[ = 1 + \left( \frac{1 - r}{nr} \right) \sqrt{1 - cn} \sum_{i=1}^{n} \exp \left\{ \frac{c_n}{2} z_i^2 \right\} \left( 1 + O(1/\sqrt{n}) \right) \]

(2.20)

where

\[
\begin{aligned}
  a_n &= \frac{1}{1 - \rho} + \frac{1 - \rho}{(1 - \rho)(1 + (n-1)\rho)} = \frac{1}{1 - \rho} + O(1/n), \\
  nb_n &= \frac{1}{1 - \rho} + \frac{1 - \rho}{(1 - \rho)(1 + (n-1)\rho)} = \frac{1}{1 - \rho} + O(1/n), \\
  c_n &= \frac{(1 - \rho)\sigma_n^2}{\tau^2(1 - \rho)\sigma_n^2 + \tau^2} = \frac{1}{1 - \rho} \rho n + \tau^2.
\end{aligned}
\]

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):

\[
2\Phi \left( \frac{2(1 - c_n)}{c_n} \log \frac{n}{\sqrt{1 - c_n}} \right) - 1 = \frac{(1 - \rho)\sigma_n^2}{\tau^2} \log \left( n \sqrt{\frac{(1 - \rho)\sigma_n^2 + \tau^2}{(1 - \rho)\sigma_n^2}} \right)
\]

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

\[ = \frac{1 - \rho}{\tau^2} \left[ d_n + \frac{1}{2} \left( d_n \log \left( \frac{n}{d_n} \right) \right) (1 + O(1)) \right]
\]

\[
\rightarrow \begin{cases}
\infty & \text{if } d_n \to \infty, \\
\frac{1 - \rho}{\tau^2} d & \text{if } d_n \to d, \\
0 & \text{if } d_n \to 0.
\end{cases}
\]
Hence, under the null hypothesis,

\[
P(M_0 \mid X) = \begin{cases} 
P(M_0) & \text{if } d_n \to \infty, \\
(1 + \frac{1 - \tau}{\tau} (2\Phi(\frac{1 - \rho}{\tau}d) - 1))^{-1} & \text{if } d_n \to d, \\
1 & \text{if } d_n \to 0.
\end{cases}
\]

**Under the alternative model** \( M_j \): by Theorem 2.3.1,

\[
P(M_j \mid x)^{-1} = \sqrt{1 + \frac{a_n \tau^2}{\sigma_n^2}} \left( \frac{n r}{1 - r} \right) \exp \left\{ \frac{-\tau^2}{2(\sigma_n^2 + a_n \tau^2)} \left[ \frac{\theta_j^2}{(1 - \rho)^2 \sigma_n^2} + O\left( \frac{1}{\sigma_n} \right) \right] \right\} \\
+ 1 + \sum_{k \neq j}^n \exp \left\{ \frac{-\tau^2}{2(\sigma_n^2 + a_n \tau^2)} \left[ \frac{\theta_j^2}{(1 - \rho)^2 \sigma_n^2} + O\left( \frac{1}{\sigma_n} \right) \right] \right\}.
\]

the first term:

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

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

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

And the last term:

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

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

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

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

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 \( \rho \to 1 \), 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 \( n \to \infty \), the Bayes with fixed \( \tau^2 \) 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

\[
X_i = \theta_i + \sqrt{\rho Z} + \sqrt{1 - \rho} Z_i \quad \forall i \in \{1, 2, ..., n\},
\]

(2.21)

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

\[
\begin{cases}
\frac{\bar{\mathbf{X}}}{\sqrt{\rho}} = z + O\left(\frac{1}{\sqrt{n}}\right) \\
\frac{x_i - \bar{\mathbf{X}}}{\sqrt{1 - \rho}} = z_i + O\left(\frac{1}{\sqrt{n}}\right)
\end{cases}
\]

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

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:

\[
P(M_i | x) = \frac{1}{\sqrt{(1 - \rho + \tau^2)(1 + (n-1)\rho) - \tau^2|\theta|}} \left( \frac{n + 1}{1 - \tau^2} \right)^{\frac{1}{4}} \exp \left\{ \frac{-\tau^2}{2} \left[ \left(1 - \rho + \tau^2\right)(1 + (n-1)\rho) - \tau^2|\theta| \right] \left( \frac{(x_i - \bar{x}) + (1 - \rho)\bar{x}}{1 + (n-1)\rho} \right)^2 \right\}^{-1}
\]

\[ (2.23) \]

And in terms of \( z(x), z_i = z_i(x) \):

\[
P(M_i | x) = \frac{1}{\sqrt{(1 - \rho + \tau^2)(1 + (n-1)\rho) - \tau^2|\theta|}} \left( \frac{n + 1}{1 - \tau^2} \right)^{\frac{1}{4}} \exp \left\{ \frac{-\tau^2}{2} \left[ \left(1 - \rho + \tau^2\right)(1 + (n-1)\rho) - \tau^2|\theta| \right] \left( \theta_i - \theta + \sqrt{1 - \rho}|z_i - \bar{z}| + (\theta + \sqrt{1 - \rho}|z_i - \bar{z}|) |z_i - \bar{z}| \right)^2 \right\}^{-1}
\]

\[ (2.24) \]
Lemma 2.10.4. Suppose \( i \in \{1, 2, ..., n\}, Z \), 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:

\[
P( \text{for all } i, |Z_i| \leq n^{1/2-\epsilon}) = \left( 1 - P(|Z_1| \geq n^{1/2-\epsilon}) \right)^n
\]

\[
= \left( 1 - \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}n^{1-2\epsilon}\right) + O\left(\exp\left(-\frac{n^{1-2\epsilon}}{2}\right)\right) \right)^n
\]

\[
= \left( 1 - \frac{2^{n^{1/2-\epsilon}}}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}n^{1-2\epsilon}\right) + o(n^{-2}) \right)^n
\]

\[
= 1 - O\left(2^{n^{1/2+\epsilon}} \exp\left(-\frac{1}{2}n^{1-2\epsilon}\right)\right)
\]

\[
= 1 + o(1).
\]

\[\square\]

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 \( \beta_n \to \infty \) and let \( \bar{x}_{n,k} = X_{n,k}1_{|X_{n,k}|\leq \beta_n} \). Suppose that as \( n \to \infty \):

\[
\sum_{k=1}^{n} P(|X_{n,k}| > \beta_n) \to 0 \quad \text{and} \quad 1/\beta_n^2 \sum_{k=1}^{n} E\bar{x}_{n,k}^2 \to 0.
\]

Then

\[
\frac{(S_n - \alpha_n)}{\beta_n} \to 0 \text{ in probability}
\]

where \( S_n = X_{n,1} + ... + X_{n,n} \) 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

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

in probability.

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

The first assumption:

\[
P(|X_{n,i}| > \beta_n) = P \left( |z_i| > \sqrt{\frac{2}{c_n} \log \frac{n}{\sqrt{1 - c_n}}} \right)
\]

\[
= 2 \frac{1}{2\pi} \exp \left\{ -\frac{1}{2} \frac{2}{c_n} \log \frac{n}{\sqrt{1 - c_n}} \right\} + O \left( \frac{\left( \frac{n}{\sqrt{1 - c_n}} \right)^{-\frac{1}{2}}}{\left( \frac{1}{c_n \log \frac{n}{\sqrt{1 - c_n}}} \right)^3} \right)
\]

\[
= \frac{1}{\sqrt{\pi}} \frac{1}{n^{\frac{1}{2}} c_n^{-\frac{1}{2}}} \log \frac{n}{\sqrt{1 - c_n}} (1 + o(1))
\]

\[
< \frac{1}{\sqrt{\pi}} \frac{1}{n^{\frac{1}{2}} c_n^{-\frac{1}{2}}} \log \frac{n}{\sqrt{1 - c_n}} (1 + o(1)).
\]

Therefore,

\[
\sum_{i=1}^{n} P(|X_{n,k}| > \beta_n) = n P(|X_{n,k}| > \beta_n)
\]

\[
< n \frac{1}{c_n} \left( 1 - c_n \right)^{\frac{1}{2}} \frac{1}{\sqrt{\log n}}
\]

\[
= n \frac{1}{c_n} \left( 1 - c_n \right)^{\frac{1}{2}} \frac{1}{\sqrt{\log n}} \to 0
\]
The second assumption: since \( \lim_{n \to \infty} c_n = 1 \), without loss of generality, assume \( c_n > 3/4 \).

\[
\frac{1}{\beta^2_n} \sum_{k=1}^{n} E X_{n,k}^2 = \frac{1 - c_n}{n^2} \int_{|z| < \sqrt{\frac{2}{c_n} \log \frac{n}{\sqrt{1 - c_n}}} \frac{1}{\sqrt{2\pi}} \exp \left\{ \frac{1}{2} \exp \left\{ \frac{-1}{2} z^2 \right\} \right\} dz
\]

\[
\leq \frac{1 - c_n}{n} \frac{1}{\sqrt{2\pi}} \left\{ \int_{1 < |z| < \sqrt{\frac{2}{c_n} \log \frac{n}{\sqrt{1 - c_n}}} \frac{1}{\sqrt{2\pi}} \exp \left\{ (c_n - \frac{1}{2})z^2 \right\} dz + \int_{|z| < 1} \exp \left\{ (c_n - \frac{1}{2})z^2 \right\} dz \right\}
\]

\[
\leq \frac{1 - c_n}{n} \frac{1}{\sqrt{2\pi}} \left\{ \frac{1}{2(c_n - \frac{1}{2})} \exp \left\{ (c_n - \frac{1}{2}) \left( \frac{2}{c_n} \log \frac{n}{\sqrt{1 - c_n}} \right) \right\} + d' \right\}
\]

\[
= \frac{1}{\sqrt{2\pi}} \left( \frac{1}{2c_n - 1} \right) \frac{1 - c_n}{n} \left( \frac{n}{\sqrt{1 - c_n}} \right)^{2 - \frac{1}{c_n}} + o(1)
\]

\[
= \frac{1}{\sqrt{2\pi}} \left( \frac{1}{2c_n - 1} \right) n^{1 - \frac{1}{c_n}} (1 - c_n)^{\frac{1}{c_n}} + o(1)
\]

\[
\leq \frac{2}{\sqrt{2\pi}} n^{\frac{1 - c_n}{c_n}} (1 - c_n)^{\frac{1}{c_n}} + o(1)
\]

\[
= o(1).
\]
Therefore, both assumptions hold. The limit is:

\[
\frac{\sqrt{1 - c_n}}{n} \alpha_n = \frac{1 - c_n}{n} \sum_{i=1}^{n} E X_{n,i} \\
= \left(1 - c_n\right) \int_{|z|<\sqrt{\frac{2}{c_n} \log \frac{n}{\sqrt{1-c_n}}}} e^{\frac{c_n z^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, dz \\
= 2\left(\Phi\left(\sqrt{\frac{2}{c_n} \log \frac{n}{\sqrt{1-c_n}}} \right) - \frac{1}{2}\right),
\]

by Fact 3.5.2:

\[
S_n - \alpha_n \overset{\text{P}}{=} \frac{\sum_{i=1}^{n} e^{c_n z_i^2} - \alpha_n}{\sqrt{1 - c_n}} = \frac{\sqrt{1 - c_n} \sum_{i=1}^{n} e^{c_n z_i^2}}{n} - \frac{\sqrt{1 - c_n}}{n} \alpha_n \rightarrow 0.
\]

in probability. And the result follows.

**Corollary 2.10.7.** Let \( c_n = \frac{s_n^2}{1 - \rho + \sigma_n^2} \), then

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

in probability.

**Proof.** By Theorem 2.10.6:

**Case I:** \( \log \frac{n}{\sigma_n^2} \rightarrow \infty \)

\[
\frac{\sqrt{1 - c_n}}{n} \alpha_n \rightarrow 2\Phi\left(\sqrt{\frac{2(1-\rho)}{\sigma_n^2} \log \left(n \sqrt{\frac{1 - \rho + \sigma_n^2}{1 - \rho}}\right)}\right) - 1 \rightarrow 1.
\]

**Case II:** \( \log \frac{n}{\sigma_n^2} \rightarrow \frac{1}{(1-\rho)k} \)

\[
\frac{\sqrt{1 - c_n}}{n} \alpha_n \rightarrow 2\Phi\left(\sqrt{\frac{2(1-\rho)}{\sigma_n^2} \log \left(n \sqrt{\frac{1 - \rho + \sigma_n^2}{1 - \rho}}\right)}\right) - 1 \rightarrow 2\Phi\left(\sqrt{\frac{2}{k}}\right) - 1.
\]
Case III: \( \frac{\log n}{\tau_n} \to 0 \)

\[
\sqrt{1 - c_n} \alpha_n \to 2 \Phi \left( \frac{2(1 - \rho)}{\tau_n^2} \log \left( n \sqrt{\frac{1 - \rho + \tau_n^2}{1 - \rho}} \right) \right) - 1 \to 0 .
\]

Lemma 2.10.8.

\[
\lim_{n \to \infty} \frac{1}{n \sqrt{1 + \tau_n^2}} \sum_{i=1}^{n} \exp \left\{ \frac{\tau_n^2}{2(1 + \tau_n^2)} \left( \frac{x_i}{1 - \rho} + b n \bar{x} \right)^2 \right\} = \lim_{n \to \infty} \frac{1}{n \sqrt{1 + \tau_n^2}} \sum_{i=1}^{n} \exp \left\{ \frac{\tau_n x_i^2}{2(1 - \rho + \tau_n^2)} \right\} \text{ a.s.}
\]

(2.25)

**Proof.** Expand the coefficients:

\[
\frac{1}{1 + \tau_n^2} = \left( 1 + \frac{\tau_n^2(n - 2 \rho)}{(1 + (n - 1) \rho)(1 - \rho)} \right)^{-1}
\]

\[
= \frac{1 - \rho}{1 - \rho + \tau_n^2 \left( 1 + \frac{-\rho}{1 + (n - 1) \rho} \right)} = \frac{1 - \rho}{1 - \rho + \tau_n^2(1 + o(1))},
\]

and

\[
\left( \frac{x_i}{1 - \rho} + b n \bar{x} \right)^2 = \left( 1 - \rho \right)^2 \left( x_i + \frac{-\rho n \bar{x}}{1 - \rho} \right)^2
\]

\[
= \frac{1}{(1 - \rho)^2} \left( x_i - \bar{x} \left( 1 - \frac{1 - \rho}{1 + (n - 1) \rho} \right) \right)^2
\]

\[
= \frac{1}{(1 - \rho)^2} \left( \sqrt{1 - \rho} z_i + \sqrt{\rho} \left( \frac{1 - \rho}{1 - \rho + \rho n} \right) + \sqrt{1 - \rho} \frac{\bar{z}}{O(1/\sqrt{n})} \right)^2
\]

\[
= \frac{z_i^2}{1 - \rho} + O \left( \frac{\log n}{\sqrt{n}} \right).
\]

Therefore,

\[
\frac{1}{\sqrt{1 + \tau_n^2}} \sum_{i=1}^{n} \exp \left\{ \frac{\tau_n^2}{2(1 + \tau_n^2)} \left( \frac{x_i}{1 - \rho} + b n \bar{x} \right)^2 \right\}
\]

\[
= \sqrt{\frac{1 - \rho}{1 - \rho + \tau_n^2 + o(1)}} \sum_{i=1}^{n} \exp \left\{ \frac{\tau_n x_i^2}{2 \left[ 1 - \rho + \tau_n^2 + o(1) \right]} \right\} .
\]

48
Lemma 2.10.9 (Type II ML estimator). Under null model, suppose

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

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 $k(c) = (\frac{1}{2} + \frac{1}{\sqrt{e}} \exp\{ -\frac{c}{2} \})^{-1}$.

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

$$\max_{\tau_n^2} L(\tau_n^2) = \max_{\tau_n^2} \sqrt{\frac{1 - \rho}{1 - \rho + \tau_n^2}} \frac{1}{n} \sum \exp \left\{ \frac{\tau_n^2 z_i^2}{2(1 - \rho + \tau_n^2)} \right\} (1 + o(1))$$

$$= \max_{c_n} \frac{\sqrt{1 - c_n}}{n} \sum \exp \left\{ \frac{c_n z_i^2}{2} \right\} (1 + o(1))$$

$$= \max_{c_n} \left\{ \frac{1}{n} \sqrt{1 - c_n} \exp \left\{ \frac{c_n z_1^2}{2} \right\} + 2\Phi \left( \sqrt{2 \left( \frac{1}{c_n} - 1 \right) \log \left( \frac{n}{\sqrt{1 - c_n}} \right)} \right) - 1 \right\} (1 + o(1)),$$

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

$$f(c_n) = \frac{1}{n} \sqrt{1 - c_n} \exp \left\{ \frac{c_n z_1^2}{2} \right\} + 2\Phi \left( \sqrt{2 \left( \frac{1}{c_n} - 1 \right) \log \left( \frac{n}{\sqrt{1 - c_n}} \right)} \right) - 1,$$

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 $f(c_n)$:
\[
\sqrt{1 - c_n} \frac{dg}{dc_n} = \frac{1}{2} \left( 2 \frac{1 - \rho}{k} - 1 \right) \left( \log n \right)^{\frac{1}{2}} \exp \left\{ \frac{c}{2} \right\} \exp \left\{ - \frac{1 - \rho}{k} \right\} (1 + o(1)) \\
\sqrt{1 - c_n} \frac{dh}{dc_n} = -\frac{1}{\sqrt{\pi}} \exp \left\{ - \frac{1 - \rho}{k} \log \left( \frac{1 - \rho + k \log n}{1 - \rho} \right) \right\} \\
\left( \frac{1 - \rho}{k} - 1 \right) = \frac{2}{\sqrt{\pi}} \exp \left\{ - \frac{c}{2} \right\} (1 + o(1)).
\]

Therefore,

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

\[ X_i \sim N(\mu_i, 1) \quad \text{for} \quad i \in \{1, \ldots, m\}, \quad (3.1) \]

where \( \mu = (\mu_1, \ldots, \mu_m) \) are unknown a priori with the spike-slab prior:

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

where \( \delta_0 \) is a point mass at zero, \( p = P(\mu_i = 0) \). We focus on the testing question of which of the \( \mu_i \) are zero, and approach this through the Bayesian model selection approach. Following the usual convention, we use \( \gamma \) to index models:

\[ \gamma = (\gamma_1, \ldots, \gamma_m) \in \{0, 1\}^m, \]

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

thus \( M_\gamma \) represents the model with the indicated zero and nonzero means, i.e.

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

with \( I_n \) being the \( n \) by \( n \) identity matrix.

Model size or model complexity of \( M_\gamma \) is defined as \( |\gamma| = \sum_{i=1}^m 1_{\gamma_i \neq 0} \). 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:

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

Given \(p(0), p(1), \ldots, p(k_{\text{max}})\) satisfying \(p(k) \geq 0 \forall k\) and \(\sum_{k=0}^{k_{\text{max}}} p(k) = 1\), 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:

\[ p = P(\gamma_i \neq 0) \sim \text{Be}(1, 1) \quad \forall 1 \leq i \leq m, \]

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

which is the special case of (3.4) when \(p(0) = p(1) = \ldots = p(k_{\text{max}}) = \frac{1}{k_{\text{max}} + 1}\).

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 \(M_\gamma\) is:

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

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

Under the null model, the Bayes factor of \(M_\gamma\) to the null model is:

$$B_{\gamma 0}(x) = \frac{m_\gamma(x)}{m_0(x)} = \prod_{i: \gamma_i \neq 0} \frac{1}{\sqrt{1 + \tau^2}} \exp \left\{ -\frac{\tau^2}{2(1 + \tau^2)} x_i^2 \right\}. \quad (3.6)$$

3.1.2 Posterior model probabilities

The posterior model probabilities are

$$P(M_\gamma \mid x) = \frac{P(M_\gamma) B_{\gamma 0}(x)}{\sum_{\gamma \in \mathcal{M}_{k_{\text{max}}}} P(M_\gamma) B_{\gamma 0}(x)} = \frac{P(M_\gamma) B_{\gamma 0}(x)}{\sum_{k=0}^{k_{\text{max}}} \left( p(k) \frac{1}{\binom{m}{k}} \sum_{\eta: |\eta| = k} B_\eta \right)}. \quad (3.7)$$
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 \( k_{\text{max}} = O(m^r) \) where \( r \in [0, \frac{1}{1+\tau^2}] \) and \( \tau^2 > 1/3 \). Then, under the null model,

\[
P(M_0 \mid X) \to P(M_0) \text{ when } m \to \infty .
\]

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

\[
P(M_0 \mid X) = \frac{P(0)}{\sum_{k=0}^{k_{\text{max}}} p(k) \frac{1}{\binom{m}{k}} \sum_{|\gamma|=k} B_{\gamma 0}(X)} . \tag{3.8}
\]

The key step is to show

\[
\frac{1}{\binom{m}{k}} \sum_{|\gamma|=k} B_{\gamma 0} = 1 + o(1) \quad \forall k = \{1, 2, \ldots, k_{\text{max}}\}, \tag{3.9}
\]

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

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

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 \( \left(\frac{1}{m} \sum_1^m Y_j\right)^k \):

\[
\left[\frac{\sum_1^m Y_j}{m}\right]^k = \sum_{k_1+\ldots+k_m=k} \binom{k}{k_1, k_2, \ldots, k_m} \prod_{1\leq j\leq m} Y_j^{k_j} \frac{k!}{m^k} \prod_{1\leq j\leq m} Y_j^{k_j} . \tag{3.10}
\]

Notice that
\[ I = \frac{(m-1)\ldots(m-k+1)}{m^{k-1}} \frac{1}{\binom{m}{k}} \sum_{k_1 + \ldots + k_m = k, \max|k| = 1} \prod_{1 \leq j \leq m} \exp\left\{ \frac{\tau^2}{2(1 + \tau^2)} k_j X_j^2 \right\} \]

\[ = \frac{(m-1)\ldots(m-k+1)}{m^{k-1}} \frac{1}{\binom{m}{k}} \sum_{|\gamma| = k} B_{\gamma 0} \]

\[ = \left[ \frac{1}{\binom{m}{k}} \sum_{|\gamma| = k} B_{\gamma 0}(x) \right] (1 + o(1)) ; \]

\[ II = \frac{1}{m^k} \sum_{k_1 + \ldots + k_m = k, \max|k| \geq 2} \frac{k!}{k_1! \ldots k_m!} \prod_{1 \leq j \leq m} Y_{j}^{k_j} \]

\[ \leq \frac{k(k-1)}{m^k} \left( \sum_{i=1}^{m} y_i^2 \right) \sum_{k_1 + \ldots + k_m = k-2} \frac{(k-2)!}{k_1! \ldots k_m!} Y_{k_1} Y_{k_2} \ldots Y_{k_m} \]

\[ = \left( \frac{k(k-1)}{m^{2/(1+\tau^2)}} \right) \left( \frac{\sum_{i=1}^{m} y_i^2}{m^{2\tau^2/(1+\tau^2)}} \right) \left( \frac{\sum_{i=1}^{m} Y_i}{m} \right)^{k-2} . \]

To conclude that $II \to 0$ when $m \to \infty$, 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 $\tau^2 > 1/3$, \( \sum_{i=1}^{m} y_i^2 / m^{2\tau^2/(1+\tau^2)} = o(1) $;

By Lemma 3.5.5, \( \left( \frac{\sum_{i=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,

\[ \lim_{m \to \infty} \frac{1}{\binom{m}{k}} \sum_{|\gamma| = k} B_{\gamma 0} = 1 . \]

\[ \square \]

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 $p(0) = 0.5$; $p(1) = p(2) = \ldots = p(k_{\max}) = \frac{1-p(0)}{k_{\max}}$. 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

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

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 $m \to \infty$,

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

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

(3.12)

From the explicit expression for Bayes factors (3.6):

\[
B_{\gamma} = \frac{1}{\sqrt{1+\tau^2}} \exp \left\{ \frac{\tau^2 X_i^2}{2(1+\tau^2)} \right\} \sum_{|\gamma|=k-1} B_{\gamma}(\tilde{X}_i),
\]  

(3.13)

where

\[
\tilde{\gamma}_i = (\gamma_1, ..., \gamma_{i-1}, \gamma_{i+1}, ..., \gamma_m),
\]

\[
\tilde{X}_i = (X_1, ..., X_{i-1}, X_{i+1}, ..., X_m).
\]

By (3.9),

\[
\frac{1}{\binom{m}{k-1}} \sum_{|\gamma|=k-1} B_{\gamma}(\tilde{X}_i) \to 1 \forall \ i.
\]

Therefore,

\[
\frac{1}{\binom{m}{k}} \sum_{|\gamma|=k} B_{\gamma} = \frac{\binom{m-1}{k-1}}{\binom{m}{k}} \frac{1}{\sqrt{1+\tau^2}} \exp \left\{ \frac{\tau^2 X_i^2}{2(1+\tau^2)} \right\} (1 + o(1))
\]

\[
= \frac{k}{m} \frac{1}{\sqrt{1+\tau^2}} \exp \left\{ \frac{\tau^2 X_i^2}{2(1+\tau^2)} \right\} (1 + o(1)).
\]

Similarly, one can show

\[
\frac{1}{\binom{m}{k}} \sum_{|\gamma|=k} B_{\gamma} = \frac{\binom{m-1}{k-1}}{\binom{m}{k}} (1 + o(1)) = \frac{m-k}{m} (1 + o(1)).
\]

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

\[
P(\gamma_i = 1 \mid X) = \frac{\frac{1}{\sqrt{1+\tau^2}} \exp \left\{ \frac{\tau^2 X_i^2}{2(1+\tau^2)} \right\} \left( \frac{1}{m} \sum_{l=1}^{k_{\text{max}}} kp(l) \right)}{1 + \left( \frac{1}{m} \sum_{l=1}^{k_{\text{max}}} kp(l) \right) \left( \frac{1}{\sqrt{1+\tau^2}} \exp \left\{ \frac{\tau^2 X_i^2}{2(1+\tau^2)} \right\} - 1 \right)} (1 + o(1)).
\]  

(3.14)

\[\square\]
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.

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 $k_{\text{max}} = O(m^r)$ 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:

$$\frac{\left( \frac{1}{m^{\frac{1}{\tau^2}}} \sum_{k=1}^{k_{\text{max}}} k P_k \right)^{\frac{1+\tau^2}{\tau^2}}}{1-q} \sqrt{\log \left( \frac{q}{1-q} \sqrt{1+\tau^2} \left( \frac{m}{\sum_{k=1}^{k_{\text{max}}} k P_k} \right) \right)} \left( 1+\tau^2 \right)^{\frac{1+\tau^2}{2+\tau^2}}.$$
Proof. A false positive occurs when this probability is greater than \( p \), from Theorem 3.3.2:

\[
P\left( |X_i| > \sqrt{2 \left( \frac{1 + \tau^2}{\tau^2} \right) \log \left[ \frac{q}{1-q} \sqrt{1 + \tau^2 \left( \frac{m}{\sum_{k=1}^{k_{\text{max}}} k p(k)} \right)} \right] (1 + o(1))} \right)
\]

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

Hence, the FPP is:

\( P(\text{any false positive} \mid M_0) \)

\[
= 1 - P(\text{no false positive} \mid M_0) = 1 - \prod_{i=1}^{m} P\left( |X_i| \leq \gamma_m \right)
\]

\[
= 1 - \prod_{i=1}^{m} \left[ 1 - P(|X_i| > \gamma_m) \right] = 1 - \left[ 1 - P(|X_i| > \gamma_m) \right]^m
\]

\[
= m P(|X_1| > \gamma_m)(1 + o(1))
\]

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

\( \square \)

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.

![FPP versus m under null model](image)

**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.

\[
\mathbb{E} \left[ \sum_{i=1}^{m} \mathbb{1}_{P(\gamma_i=1|X)>1/2} \mid M_0 \right] \\
= \sum_{i=1}^{m} P \left( (P(\gamma_i = 1 \mid X) > 1/2) \mid M_0 \right) \\
= m P \left( |X_i| > \sqrt{2 \left( \frac{1 + \tau^2}{\tau^2} \right)} \log \left[ \frac{q}{1 - q} \sqrt{1 + \tau^2 \left( \frac{\frac{m}{\sum_{k=1}^{k_{max}} k p_k}}{\sum_{k=1}^{k_{max}} k p_k} \right)} \right] (1 + o(1)) \right) \\
= O \left[ \left( \frac{\sum_{k=1}^{k_{max}} k p_k}{m} \right)^{1+\frac{\tau^2}{\tau^2}} \frac{1}{\sqrt{\log \left( \frac{m}{\sum_{k=1}^{k_{max}} k p_k} \right)}} \right].
\]

(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 \( \sum_{k=1}^{k_{max}} k p(k) \) 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 \( \sum_{k=1}^{k_{max}} k p(k) \) can be obtained straightforwardly by Cauchy’s inequality:
Fact 3.3.8. When \( p(k) = \frac{2^k}{k_{max}(k_{max}+1)}(1 - p(0)) \), \( \sum_{k=1}^{k_{max}} k p(k) \) attains its maximum at

\[
\frac{2k_{max} + 1}{3}(1 - p(0)).
\]

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

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

\[
\sum_{k=1}^{k_{max}} k p(k) = \frac{k_{max}}{2}.
\]

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

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

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

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

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

$$mP \left\{ |X_i| > \sqrt{2 \left( \frac{1 + \tau^2}{\tau^2} \right) \log \left[ \frac{q(m, \alpha)}{1 - q(m, \alpha)} \sqrt{1 + \tau^2 \left( \sum_{k=1}^{k_{\text{max}}} k p(k) \right)} \right]} (1 + o(1)) \right\}$$

$$= mP \left\{ |X_i| > \sqrt{2 \log \left( m \left( \frac{1}{\sqrt{\pi \alpha^2 \log m}} \right) \right)} (1 + o(1)) \right\}$$

$$= mP \left\{ |X_i| > \sqrt{2 \log m - \log \left( \pi \alpha^2 \log m \right)} (1 + o(1)) \right\}$$

$$= m \frac{2 \frac{1}{m} \sqrt{\pi \alpha^2 \log m}}{\sqrt{2 \pi} \sqrt{2 \log m - \log \log m - \log(\pi \alpha^2)}} (1 + o(1)) \quad \text{by Fact 5.4.1}$$

$$= \alpha(1 + o(1)).$$

$\square$

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 $\alpha$ 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 $\alpha$) 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 \( 1 - (1 - \frac{0.05}{m})^m \) 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 \( \theta \). For each \( \theta \), 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 $\theta_i$, $m = 100, k_{\text{max}} = 3, \tau = 1$. For each $\theta_i$, 100 iterations were generated to get the box plot.

Remark 3.3.14. Figure 3.6 demonstrates how $\pi(p \mid x)$ varies with respect to different expected and true noise ratios. The model noise ratio is $k_{\text{max}}/m$, 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 ($k_{\text{max}} = m$), the posterior of $p$ concentrates at the true noise ratio. On the other hand, $\pi(p \mid x)$ tends to not surpass the model noise ratio $k_{\text{max}}/m$ 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 $X$. 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

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

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

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 $\beta_m \to \infty$ and let $\bar{X}_{m,k} = X_{m,k}1_{\{|X_{m,k}| \leq \beta_m\}}$.

Suppose that as $m \to \infty$

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

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

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

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

See Durrett (2010) for proof.

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

$$m^{-2\tau^2/(1+\tau^2)} \sum_{j=1}^{m} Y_j^2 \to 0 \text{ when } m \to \infty.$$
Proof. First we need the following order estimate before applying Fact 3.5.2: If \( \lim_{m \to \infty} a_m \in (0, \infty] \), \( \lim_{m \to \infty} b_m = \infty \), then

\[
\int_0^{b_m} e^{\frac{am x^2}{2}} dx = O\left( \frac{e^{a_m b_m^2/2}}{a_m b_m} \right). \tag{3.18}
\]

To show (3.18) is true, notice that

\[
\int_0^{b_m} e^{\frac{am x^2}{2}} dx = \frac{1}{\sqrt{a_m}} \int_0^{\sqrt{a_m} b_m} e^{\frac{v^2}{2}} dv
\]

and by L'Hôpital's rule:

\[
\lim_{m \to \infty} \frac{\frac{1}{\sqrt{a_m}} \int_0^{\sqrt{a_m} b_m} e^{\frac{v^2}{2}} dv}{e^{-\frac{am b_m^2}{2}} / (a_m b_m)} = \lim_{m \to \infty} \frac{\int_0^{a_m b_m} e^{\frac{v^2}{2}} dv}{e^{-\frac{am b_m^2}{2}} / (a_m b_m)} - \frac{1}{\sqrt{a_m} b_m}
\]

where \( v = \sqrt{a_m} b_m \). Then take \( X_{m,j} = \exp\left\{ \frac{\tau^2}{1 + \tau^2} x_j^2 \right\} \), \( \beta_m = m^{2\tau^2/(1 + \tau^2)} \) in Fact 3.5.2, both assumptions in Fact 3.5.2 hold since:

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

and

\[
\frac{1}{\beta_m^2} \sum_{j=1}^{m} \mathbb{E} X_{m,j}^2 = \frac{m}{m^{4\tau^2/(1 + \tau^2)}} \int_{|x| \leq \sqrt{\log m}} \frac{1}{\sqrt{2\pi}} e^{\left( \frac{2\tau^2}{1 + \tau^2} \right) x^2} dx
\]

\[
= \begin{cases} 
O\left( m^{\frac{2\tau^2}{1 + \tau^2}} \frac{1}{\sqrt{\log m}} \right) & \text{if } \tau^2 > 1, \text{ by (3.18)} \\
O\left( m^{\frac{2\tau^2}{1 + \tau^2}} \right) & \text{else,}
\end{cases}
\]

\[
= o(1) \quad \text{if } \tau^2 > 1/3.
\]

The limit is:
\[
\frac{1}{\beta_m} \sum_{j=1}^{m} E \bar{X}_{m,j} = m^{-2r^2/(1+r^2)} \frac{m}{\sqrt{2\pi}} \int_{|x|<\sqrt{2\log m}} \exp \left\{ \frac{x^2}{2} \left( \frac{2r^2}{1+r^2} - 1 \right) \right\}
\]

\[
= m^{(1-r^2)/(1+r^2)} O \left( m^{(r^2-1)/(1+r^2)} \frac{1}{\sqrt{\log m}} \right) \text{ by (3.18)}
\]

\[
= O \left( \frac{1}{\sqrt{\log m}} \right) \to 0.
\]

\[\square\]

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

\[\frac{S_m}{m^{1/p}} \to 0 \text{ a.s.}\]

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

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

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

**Proof.**

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

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

\[
\frac{\sum_{i=1}^{m} Y_j - m}{m^{1/p}} = \left( \frac{\sum_{i=1}^{m} Y_j}{m^{1/p}} - m^{1-1/p} \right) = o(1),
\]

\[
\frac{\sum_{i=1}^{m} Y_j}{m} - 1 = \frac{1}{m^{1-1/p}} \left( \frac{\sum Y_j}{m^{1/p}} - m^{1-1/p} \right) = o \left( \frac{1}{m^{1-1/p}} \right), \quad (3.19)
\]

69
the lemma follows since:

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

\[
\square
\]

**Lemma 3.5.6.**

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

**Proof.** By definition,

\[
B_{\gamma 0}(x_1, x_2, ..., x_m) = \frac{1}{(1 + \tau^2)^{\frac{|\gamma|}{2}}} \exp \left\{ \frac{\tau^2}{2(1 + \tau^2)} \sum_{i \in \gamma, i \neq 0} x_i^2 \right\}.
\]

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

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

\[
\sum_{\gamma:|\gamma|=k} B_{\gamma 0}(x) = \sum_{C^m_k} h^k(x_{i_1}, x_{i_2}, ..., x_{i_k}).
\]

Therefore,

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

\[
\square
\]

**Definition 3.5.7.** For \( c \in \{0, 1, ..., k\}, \)

\[
h^k_c(x_1, ..., x_c) = \mathbb{E}(h^k(x_1, x_2, ..., x_c, X_{c+1}, ..., X_k))
\]

\[
(\sigma^2_c) = \text{Var}(h^k_c(X_1, ..., X_c)).
\]

Note that \( h^k_k = h^k, (\sigma^2_k) = \text{Var}(h^k) = \sigma^2_k. \)
Fact 3.5.8. Let $U$ be the U-statistics with kernel $h^k$. If $\sigma_k^2 < \infty$, then
\[
\sqrt{m}\left[U - \mathbb{E}(h^k(X_1, \ldots, X_k))\right] \rightarrow N(0, k^2(\sigma_k^2)^2).
\]

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,
\[
\sqrt{m}\left[\frac{1}{m} \sum_{|\gamma|=k} B_{\gamma0}(x) - 1\right] \rightarrow N\left(0, k^2\left[\frac{1}{\sqrt{1-\tau^2}} - 1\right]\right).
\]

Proof. By Lemma 3.5.6, $1/(m) \sum_{|\gamma|=k} B_{\gamma0}$ is a U-statistics, then evaluate the expectation and variance in Definition 3.5.7, one can get under the null model:
\[
\mathbb{E}(B_{\gamma0}(X)) = \mathbb{E}(h^k_1(X)) = 1,
\]
\[
\mathbb{E}(B_{\gamma0}(X)^2) = \begin{cases} 
\frac{1}{\left[(1+\tau^2)(1-\tau^2)^{\frac{1}{2}}\right]^2} & \text{if } \tau^2 < 1, \\
\infty & \text{else}, 
\end{cases}
\]
\[
(\sigma_1^k)^2 = \begin{cases} 
\frac{1}{\sqrt{1-\tau^2}} - 1 & \text{if } \tau^2 < 1, \\
\infty & \text{else}, 
\end{cases}
\]
\[
(\sigma_k^k)^2 = \begin{cases} 
\frac{1}{\left[(1+\tau^2)(1-\tau^2)^{\frac{1}{2}}\right]^2} - 1 & \text{if } \tau^2 < 1, \\
\infty & \text{else}. 
\end{cases}
\]
then apply Fact 3.5.8. \qed
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 \( M_i^0 \) 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 \( \pi(p) = 1_{[0,1]}(p) \) in the following.

Denoting the marginal likelihood of the \( i^{th} \) trial under model \( M_i^j \) by \( m(x_i \mid M_i^j) \), the Bayes factor of the null hypothesis to the alternative hypothesis in the \( i^{th} \) trial is

\[
B_{01}^i = \frac{m(x_i \mid M_i^0)}{m(x_i \mid M_i^1)}.
\]

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

\[
P(M_{0}^{n+1} \mid x_1, \ldots, x_{n+1}) = \left(1 + \frac{1 - p_n}{p_n} \frac{1}{B_{01}^{n+1}}\right)^{-1},
\]

where \( p_n = \mathbb{E}(p \mid x_1, \ldots, x_n) \).

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

$$
\pi(p \mid x_1, \ldots, x_n) = \frac{(n + 1)}{(n + 2)} \left( \sum_{j=0}^{n} \frac{(1 - p)^j}{n-j} \sum_{1 \leq i_1 < \cdots < i_{n-j} \leq n} B_{01}^{i_1} \cdots B_{01}^{i_{n-j}} \right),
$$

(4.1)

$$
\mathbb{E}(p \mid x_1, \ldots, x_n) = \frac{(n + 1)}{(n + 2)} \left( \sum_{j=0}^{n} \frac{(1 + j)!}{j!} \sum_{1 \leq i_1 < \cdots < i_{n-j} \leq n} B_{01}^{i_1} \cdots B_{01}^{i_{n-j}} \right).
$$

Proof. By definition, the posterior of $p$ is

$$
\pi(p \mid x_1, x_2, \ldots, x_n) = \frac{\prod_{i=1}^{n} \left( pB_{01}^{i} + (1 - p) \right)}{\int_{0}^{1} \prod_{i=1}^{n} \left( pB_{01}^{i} + (1 - p) \right) dp}.
$$

(4.2)

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

$$
\int_{0}^{1} p^j (1 - p)^i dp = \frac{\Gamma(i + 1) \Gamma(j + 1)}{\Gamma(i + j + 2)} = \frac{i! j!}{(i + j + 1)!} = \frac{1}{\binom{i+j}{i}(i + j + 1)}.
$$

□

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

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

$$
p_1 = \mathbb{E}(p \mid x_1) = \int_{0}^{1} p \pi(p \mid x_1) dp = \frac{2B_{01}^{1} + 1}{3(B_{01}^{1} + 1)}.
$$

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

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

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

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

\[
m_i = m(x_i) = P(M_0^i)m(x_i \mid M_0^i) + P(M_1^i)m(x_i \mid M_1^i),
\]

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

**Corollary 4.2.4.**

\[
\pi(p \mid x_1, \ldots, x_n) = \pi(p \mid x_1, \ldots, x_{n-1}) \left( pB_0^n + (1 - p) \right) \frac{m_{n-1}}{m_n},
\]

where

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

Another possibility is dynamic programming:

**Corollary 4.2.5 (Recursive formula).**

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

where

\[
a_{ij} = B_{01}^i a_{(i-1)j} + a_{(i-1)(j-1)} \quad \forall i, j \in \{1, \ldots, n\}
\]

\[
a_{i0} = a_{0i} = 0 \quad \forall i \in \{1, \ldots, n\}
\]

\[
a_{n(n+1)} = 1
\]

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

\[
\begin{align*}
M_0^i : & \quad p_{i0} \leq p_{i1} \\
M_1^i : & \quad p_{i0} > p_{i1} .
\end{align*}
\]

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

\[
m(x_i, n_i \mid M_0^i) = \frac{1}{(\text{Beta}(a, b))^2} \int_0^1 \left( \int_0^{p_{i1}} \binom{n_{i1}}{x_{i1}} p_{i0}^{x_{i0}} (1 - p_{i0})^{n_{i0} - x_{i0}} p_{i1}^{a-1} (1 - p_{i1})^{b-1} dp_{i0} \right) \left( \binom{n_{i1}}{x_{i1}} p_{i1}^{x_{i1}} (1 - p_{i1})^{n_{i1} - x_{i1}} p_{i1}^{a-1} (1 - p_{i1})^{b-1} dp_{i1} ,
\right.
\]

\[
m(x_i, n_i \mid M_1^i) = \frac{1}{(\text{Beta}(a, b))^2} \int_0^1 \left( \int_{p_{i1}}^1 \binom{n_{i0}}{x_{i0}} p_{i0}^{x_{i0}} (1 - p_{i0})^{n_{i0} - x_{i0}} p_{i1}^{a-1} (1 - p_{i1})^{b-1} dp_{i0} \right) \left( \binom{n_{i1}}{x_{i1}} p_{i1}^{x_{i1}} (1 - p_{i1})^{n_{i1} - x_{i1}} p_{i1}^{a-1} (1 - p_{i1})^{b-1} dp_{i1} .
\right.
\]

76
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 3$^{rd}$ trial, without multiplicity adjustment, is $P(M_1^3 \mid x_3, n_3) = (1 + B_{01}^3)^{-1} = 0.872$. With multiplicity adjustment, the posterior probability is

$$P(M_1^3 \mid x_{1:3}, n_{1:3}) = \left(1 + \frac{p_2}{p_2 - B_{01}^3}\right)^{-1} = 0.839.$$ 

Thus the evidence for effectiveness of the 3$^{rd}$ 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 $B_{01}^i \geq -\epsilon p \log p$ 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

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

**Proof.**
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.**

\[p_n = \mathbb{E}(p \mid x_1, ..., x_n) = \mathbb{E}(p \mid B_{01}^1, ..., B_{01}^n) = \int_0^1 p \prod_{j=1}^n \left( pB_{01}^j + 1 - p \right) dp,
\]

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

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

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

\[
\frac{1}{n + 1} = \lim_{B_{01}^i \to \infty} \left( \frac{1 - p_n}{p_n} \right) \leq \left( \frac{1 - p_n}{p_n} \right) \leq \lim_{B_{01}^i \to 0} \left( \frac{1 - p_n}{p_n} \right) = n + 1.
\]
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 Alosh (2008)). In the original version of this procedure, endpoints are ordered a priori and tested sequentially at nominal level $\alpha$, 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, $\{M^i_0, M^i_1\}, i \in \{1, ..., n\}$, ordered in terms of importance (most important first).

- In the $i^{th}$ test ($i < n$), test $\{M^i_0, M^i_1\}$ at level $\alpha$, 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 $\alpha$. 

79
If the testing terminates at the last trial, report overall type-I error of $\alpha$ 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 $\alpha$, and report that the overall error probability (the chance that any rejection is erroneous) is $\alpha$.

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

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

where $P(M_i^1 | x_i)$ 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 $\alpha$ 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

\[
\begin{align*}
M_0^i : X_i & \text{ has density } f_0(x_i) \\
M_1^i : X_i & \text{ has density } f_1(x_i).
\end{align*}
\]

Then \(\alpha\)-level tests are as follows.

- The likelihood ratio test (LRT):

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

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

- Bayesian test: defining \(s_i = B_{01}^i(x_i)\), 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

\[
\chi_{s_1} = \{x_1 \in \chi_1 \mid B_{01}^i(x_1) \in \{s_1, \psi(s_1)\} \}
\]

where \(\psi(s_1) = (F_0^1)^{-1}(1 - F_1(s_1))\),

and then defining frequentist error probabilities conditioned on being in a partition.
Theorem 5.2.1 (Berger, Brown, Wolpert, 1994). For $0 < s_1 < 1$,

$$\alpha(s_1) = P_0(\text{rejecting } M_0^1 \mid \chi_{s_1}) = \frac{s_1}{1 + s_1}$$

= posterior probability of $M_0^1$ when $B_{01}^1 < 1$

$$\beta(s_1) = P_1(\text{accepting } M_0^1 \mid \chi_{s_1}) = \frac{1}{1 + \psi(s_1)}$$

= posterior probability of $M_1^1$ when $B_{01}^1 > 1$.

Two obvious consequences of the above theorem are that

Corollary 5.2.2.

$$P_0(\text{accepting } M_0^1 \mid \chi_s) = 1 - \frac{s}{1 + s}$$

= posterior probability of $M_1^1$ when $B_{01} < 1$;

$$P_1(\text{rejecting } M_0^1 \mid \chi_s) = 1 - \frac{1}{1 + \psi(s)}$$

= posterior probability of $M_1^1$ when $B_{01} > 1$.

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:

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

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 $\psi(s_i) = (F_{0i}^0)^{-1}(1 - F_{1i}^0(s_i))$, for any $0 < s_i \leq 1$, and then lets the partition be

$$\chi_{s_1, s_2} = \{ x \in \chi : B_{01}^i \in \{s_i, \psi(s_i)\}, \ i \in \{1, 2\} \}$$

$$= \{ B(x_1) = s_1, B(x_2) = s_2 \} \cup \{ B(x_1) = s_1, B(x_2) = \psi(s_2) \} \cup \{ B(x_1) = \psi(s_1) \}.$$
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 \( f_0^p(s_2) + f_1^p(s_2) \). The ensuing overall fixed partition of the sample space is

\[
\chi^{F}_{s_1, s_2} = \{ x \in \chi : B_{0i} \in \{ s_i, \psi(s_i) \}, \ i \in \{ 1, 2 \} \} \\
= \{ B(x_1) = s_1, B(x_2) = s_2 \} \cup \{ B(x_1) = s_1, B(x_2) = \psi(s_2) \} \\
\cup \{ B(x_1) = \psi(s_1) \}_{s_2} \\
= I \cup II \cup III. \tag{5.1}
\]

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 \( f_0^p(s_2) \) and \( f_1^p(\psi(s_2)) \), respectively, where \( f_T^p(\cdot) \) refers to the unknown true density. The resulting overall random partition is

\[
\chi^R_{s_1, s_2} = \{ x \in \chi : B_{0i} \in \{ s_i, \psi(s_i) \}, \ i \in \{ 1, 2 \} \} \\
= \{ B(x_1) = s_1, B(x_2) = s_2 \} \cup \{ B(x_1) = s_1, B(x_2) = \psi(s_2) \} \\
\cup \{ B(x_1) = \psi(s_1) \}_{s_2} \cup \{ B(x_1) = \psi(s_1) \}_{\psi(s_2)} \\
= I \cup II \cup IIIA \cup IIIB. \tag{5.2}
\]

The reason this is called a random partition is that we do not know \( f_0^p(s_2) \) and \( f_1^p(\psi(s_2)) \) 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 frequenist 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 \( M_1^1 \) and \( M_2^2 \) are true, we compute \( P_{ij}(\text{decision} \mid \chi_{s_1,s_2}) \), the error probability for the decision conditioned on \( \chi_{s_1,s_2} \), 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.

<table>
<thead>
<tr>
<th></th>
<th>((s_1, s_2))</th>
<th>((s_1, \psi(s_2)))</th>
<th>((\psi(s_1), s_2 \cup \psi(s_2)))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(P_{00}(\cdot \mid \chi_{s_1,s_2}))</td>
<td>(\frac{s_1}{s_1+1} \frac{s_2}{s_2+1})</td>
<td>(\frac{s_1}{s_1+1} \frac{1}{s_2+1})</td>
<td>(\frac{1}{s_1+1})</td>
</tr>
<tr>
<td>(P_{01}(\cdot \mid \chi_{s_1,s_2}))</td>
<td>(\frac{s_1}{s_1+1} \frac{\psi(s_2)}{\psi(s_2)+1})</td>
<td>(\frac{s_1}{s_1+1} \frac{1}{\psi(s_2)+1})</td>
<td>(\frac{1}{s_1+1})</td>
</tr>
<tr>
<td>(P_{10}(\cdot \mid \chi_{s_1,s_2}))</td>
<td>(\frac{\psi(s_1)}{\psi(s_1)+1} \frac{s_2}{s_2+1})</td>
<td>(\frac{\psi(s_1)}{\psi(s_1)+1} \frac{1}{s_2+1})</td>
<td>(\frac{1}{\psi(s_1)+1})</td>
</tr>
<tr>
<td>(P_{11}(\cdot \mid \chi_{s_1,s_2}))</td>
<td>(\frac{\psi(s_1)}{\psi(s_1)+1} \frac{\psi(s_2)}{\psi(s_2)+1})</td>
<td>(\frac{\psi(s_1)}{\psi(s_1)+1} \frac{1}{\psi(s_2)+1})</td>
<td>(\frac{1}{\psi(s_1)+1})</td>
</tr>
</tbody>
</table>
Theorem 5.2.3 (Type I error).

\[
P_{00}(\text{rejecting } M^1_0 \mid \chi_{s_1, s_2}) = P_{00}(\text{rejecting } M^1_0 \lor M^2_0 \mid \chi_{s_1, s_2})
\]

\[
= \frac{B^1_{01}}{B^1_{01} + 1} \quad \text{if } B^1_{01} < 1
\]

\[
P_{00}(\text{rejecting } M^2_0 \mid \chi_{s_1, s_2}) = P_{00}(\text{rejecting } M^1_0 \land M^2_0 \mid \chi_{s_1, s_2})
\]

\[
= \left( \frac{B^1_{01}}{B^1_{01} + 1} \right) \left( \frac{B^2_{01}}{B^2_{01} + 1} \right) \quad \text{if } B^1_{01} < 1; B^2_{01} < 1
\]

\[
P_{10}(\text{rejecting } M^2_0 \mid \chi_{s_1, s_2}) = P_{10}(\text{rejecting } M^1_0 \land M^2_0 \mid \chi_{s_1, s_2})
\]

\[
= \left( \frac{\psi(s_1)}{1 + \psi(s_1)} \right) \left( \frac{B^2_{01}}{1 + B^2_{01}} \right) \quad \text{if } B^2_{01} < 1
\]

\[
P_{01}(\text{rejecting } M^1_0 \mid \chi_{s_1, s_2}) = P_{01}(\text{rejecting } M^1_0 \land M^2_0 \mid \chi_{s_1, s_2})
\]

\[
= \left( \frac{B^1_{01}}{B^1_{01} + 1} \right) \quad \text{if } B^1_{01} < 1
\]

\[
P_{11}(\text{rejecting } M^1_0 \mid \chi_{s_1, s_2}) = \frac{\psi(s_1)}{1 + \psi(s_1)}
\]

\[
P_{11}(\text{rejecting } M^2_0 \mid \chi_{s_1, s_2}) = \left( \frac{\psi(s_1)}{1 + \psi(s_1)} \right) \left( \frac{\psi(s_2)}{1 + \psi(s_2)} \right)
\]

\[
P_{10}(\text{rejecting } M^2_0 \mid \text{rejecting } M^1_0, \chi_{s_1, s_2}) = \left( \frac{B^2_{01}}{B^2_{01} + 1} \right) \quad \text{if } B^2_{01} < 1
\]

85
Theorem 5.2.4 (Type II error).

\[ P_{01}(\text{accepting } M_0^2 \mid \chi_{s_1,s_2}) = P_{01}(\text{rejecting } M_0^1 \cap \text{accepting } M_0^2 \mid \chi_{s_1,s_2}) \]

\[ = \left( \frac{B_{01}^1}{B_{01}^1 + 1} \right) \left( \frac{1}{B_{01}^2 + 1} \right) \quad \text{if } B_{01}^1 < 1; B_{01}^2 > 1 \]

\[ P_{11}(\text{accepting } M_0^2 \mid \chi_{s_1,s_2}) = P_{11}(\text{rejecting } M_0^1 \cap \text{accepting } M_0^2 \mid \chi_{s_1,s_2}) \]

\[ = \left( \frac{\psi(s_1)}{\psi(s_1) + 1} \right) \left( \frac{1}{B_{01}^2 + 1} \right) \quad \text{if } B_{01}^2 > 1 \]

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

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

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

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

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 \( \alpha \) so that, if one can be found to match the standard Bayesian test, then the standard Bayesian test should also have expected error probability \( \alpha \) (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

<table>
<thead>
<tr>
<th>subset of ( \chi )</th>
<th>decision</th>
<th>( \alpha_f(\chi_s) )</th>
<th>( \alpha_B(\chi_s) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1 )</td>
<td>Accept ( M_0^1 )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( (R_1, A_2) )</td>
<td>Reject ( M_0^1 ), accept ( M_0^2 )</td>
<td>( \alpha )</td>
<td>( P(M_0^1 \mid x_1) )</td>
</tr>
<tr>
<td>( (R_1, R_2, A_3) )</td>
<td>Reject ( M_0^1), ( M_0^2 ), accept ( M_0^3 )</td>
<td>( \alpha )</td>
<td>( 1 - \prod_{i=1}^{2} P(M_i^1 \mid x_i) )</td>
</tr>
<tr>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
</tr>
<tr>
<td>( (R_1, ..., R_{n-1}, A_n) )</td>
<td>Reject ( M_0^{1} ) to ( M_0^{n-1} ), accept ( M_0^{n} )</td>
<td>( \alpha )</td>
<td>( 1 - \prod_{i=1}^{n-1} P(M_i^1 \mid x_i) )</td>
</tr>
<tr>
<td>( (R_1, ..., R_{n-1}, R_n) )</td>
<td>Reject all null models</td>
<td>( \alpha )</td>
<td>( 1 - \prod_{i=1}^{n} P(M_i^1 \mid x_i) )</td>
</tr>
</tbody>
</table>

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 \( \alpha \). For comparison with the Bayesian procedure, we formally compute this unconditional Type I error as the expected value of the frequentist reported error (\( \alpha \)), conditioned on rejecting.

**Example 5.3.1** (two endpoints, frequentist).

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

<table>
<thead>
<tr>
<th>partition ( \chi_s )</th>
<th>decision</th>
<th>frequentist error</th>
<th>( P(\chi_s) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1 )</td>
<td>Accept ( M_0^1 )</td>
<td></td>
<td>( 1 - \alpha )</td>
</tr>
<tr>
<td>( (R_1, A_2) )</td>
<td>Reject ( M_0^1 ), accept ( M_0^2 )</td>
<td>( \alpha )</td>
<td>( \alpha(1 - \alpha) )</td>
</tr>
<tr>
<td>( (R_1, R_2) )</td>
<td>Reject ( M_0^1 ), reject ( M_0^2 )</td>
<td>( \alpha )</td>
<td>( \alpha^2 )</td>
</tr>
</tbody>
</table>

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

\[
\alpha_f = \mathbb{E}(\alpha \mid \text{rejection})
\]

\[
= \frac{1}{\alpha} \left[ \int_{A_2} \int_{R_1} \alpha f_0(x_1) f_0(x_2) dx_1 dx_2 + \int_{R_2} \int_{R_1} \alpha f_0(x_1) f_0(x_2) dx_1 dx_2 \right]
\]

\[
= \frac{1}{\alpha} \left[ (1 - \alpha) \alpha^2 + \alpha^3 \right] = \alpha .
\]

87
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*

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

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

<table>
<thead>
<tr>
<th>partition $\chi_s$</th>
<th>decision</th>
<th>Bayesian error</th>
<th>$P(\chi_s)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>Accept $M_0^1$</td>
<td>$P(M_0^1 \mid x_1)$</td>
<td>$1 - \alpha$</td>
</tr>
<tr>
<td>$(R_1, A_2)$</td>
<td>Reject $M_0^1$, accept $M_0^2$</td>
<td>$P(M_0^1 \mid x_1)$</td>
<td>$\alpha(1 - \alpha)$</td>
</tr>
<tr>
<td>$(R_1, R_2)$</td>
<td>Reject $M_0^1$, reject $M_0^2$</td>
<td>$1 - P(M_0^1 \mid x_1)P(M_0^2 \mid x_2)$</td>
<td>$\alpha^2$</td>
</tr>
</tbody>
</table>

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

\[
\alpha_B = \mathbb{E}(\text{Bayesian error} \mid \text{rejection})
\]

\[
= \frac{1}{\alpha} \int_{A_2} \int_{R_1} P(M_0^1 \mid x_1)f_0(x_1)f_0(x_2)dx_1dx_2
\]

\[
+ \frac{1}{\alpha} \int_{R_2} \int_{R_1} \left(1 - P(M_0^1 \mid x_1)P(M_0^2 \mid x_2)\right)f_0(x_1)f_0(x_2)dx_1dx_2
\]

\[
= \frac{1}{\alpha} \left[ (1 - \alpha) \int_{R_1} \left(1 - P(M_0^1 \mid x_1)\right)f_0(x_1)dx_1 \right]
\]

\[
+ \frac{1}{\alpha^2} \left( \int_{R_1} P(M_0^1 \mid x_1)f_0(x_1)dx_1 \right) \left( \int_{R_2} P(M_0^2 \mid x_2)f_0(x_2)dx_2 \right)
\]

\[
= \frac{1}{\alpha} \left[ (1 - \alpha)v_1\alpha + \alpha^2 - (1 - v_1)(1 - v_2)\alpha^2 \right]
\]

\[
= (1 - v_1)v_2\alpha + v_1.
\]

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

88
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

\[
R_1(\alpha) = \{x_1 : x_1 > 1 - \alpha\}; P(R_1 \mid M_0^1) = \alpha.
\]

Since the posterior probability is

\[
P(M_0^1 \mid x_1) = \frac{f_0(x_1)}{f_0(x_1) + f_1(x_1)},
\]

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

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

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

\[
R_2(\alpha) = \{x_2 : x_2 < \sqrt{\alpha}\}; P(R_2 \mid M_0^2) = \alpha.
\]

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

\[
\begin{align*}
\lim_{\alpha \to 0} v_1 &= 1/3 \\
\lim_{\alpha \to 0} v_2 &= 0.
\end{align*}
\]

Since \(\lim_{\alpha \to 0} v_1 \to 0\), from (5.3) it is clear that \(\alpha_B(\alpha) > \alpha\) for small \(\alpha\).
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 $\lim_{\alpha \to 0} v_i = 0$, it can still be the case that $\alpha_B > \alpha$ for small $\alpha$, 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 $\alpha$:

\[
v_i = \frac{1}{\alpha} \int_{R_i(\alpha)} \frac{f_0(x_i)}{f_0(x_i) + f_1(x_i)} f_0 dx_i
\]

\[
= 1 - \frac{1}{\alpha} \int_{R_i(\alpha)} \frac{1}{1 + f_0/f_1} f_0 dx_i
\]

\[
= 1 - \frac{1}{\alpha} \int_{R_i(\alpha)} f_0 dx + \frac{1}{\alpha} \int_{R_i(\alpha)} \frac{f_0^2}{f_1} dx + O\left(\frac{1}{\alpha} \int_{R_i(\alpha)} \frac{f_0^3}{f_1^2} dx\right)
\]

\[
= \frac{1}{\alpha} \int_{R_i(\alpha)} f_0^2 dx + O\left(\frac{1}{\alpha} \int_{R_i(\alpha)} \frac{f_0^3}{f_1} dx\right).
\]

Also,

\[
\frac{1}{\alpha} \int_{R_i(\alpha)} f_0^2/f_1 dx = \frac{1}{\alpha} \int_{\sigma z_{1-\alpha} + \mu_0}^\infty \exp\{-1/(2\sigma^2)[2(x - \mu_0)^2 - (x - \mu_1)^2]\} dx
\]

\[
= \frac{1}{\alpha} \exp\{(\mu_0 - \mu_1)^2/\sigma^2\} \left(1 - \Phi(z_{1-\alpha} - \mu_0 + \mu_1)\right).
\]
By Fact 5.4.1 and Lemma 5.4.2,

\[ 1 - \Phi(z_{1-\alpha} + \mu_1 - \mu_0) \]

\[ = \frac{1}{\sqrt{2\pi}} e^{-\frac{\mu_1 - \mu_0}{2\sqrt{2}\pi}} \sqrt{-2 \log(\alpha\sqrt{2\pi})} \sqrt{\mu_1 - \mu_0 - 2\log(\alpha\sqrt{2\pi}) - \log(-2 \log(\alpha\sqrt{2\pi}))} (1 + o(1)) \]

\[ = \frac{e^{-\frac{\mu_1 - \mu_0}{2\sqrt{2}\pi}}}{\sqrt{1 + \frac{\mu_1 - \mu_0}{-2\log(\alpha\sqrt{2\pi})} - \frac{\log(-2 \log(\alpha\sqrt{2\pi}))}{\log(\alpha\sqrt{2\pi})}}} (1 + o(1)) \]

\[ = O(\alpha) \]

Thus, (5.5) is greater than \( \alpha \) when \( \alpha \) 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 \( \alpha \). 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:

\[ \begin{cases} M_0^i : x_i \sim f_0(x_i \mid \theta_0); \theta \in \Theta_0 \\ M_1^i : x_i \sim f_1(x_i \mid \theta_1); \theta \in \Theta_1. \end{cases} \tag{5.6} \]

Given priors on the unknown parameters \( \theta_0, \theta_1 \) and equal prior probability on both models \( (=1/2) \), the posterior probability of the null hypothesis is

\[ P(M_0^i \mid x_i, \pi_0, \pi_1) = \frac{m_0(x_i)}{m_1(x_i) + m_0(x_i)}, \]

where the marginal likelihood is

\[ m_j(x_i) = \int_{\Theta_j} f_j(x_i \mid \theta) \pi_j(\theta) d\theta \quad j \in \{0, 1\}. \]
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*

\[
\begin{align*}
M_0^i : f(x_i - \theta), \theta \leq 0, \\
M_1^i : f(x_i - \theta), \theta > 0,
\end{align*}
\]

*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*

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

*Proof.* By Casella and Berger (1987),

\[
P(M_0^i \mid x_i) = p(x_i),
\]

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

\[
\int_{R_i(\alpha)} p(x_i) f_0(x_i) dx_i = \int_c^{\infty} \left( \int_x^{\infty} f_0(t) dt \right) f_0(x_i) dx_i
\]

\[
= \int_c^{\infty} (1 - F_0(x_i)) f_0(x_i) dx_i
\]

\[
= \alpha - \int_{F_0^{-1}(c)}^{F_0^{-1}(\infty)} F_0(x_i) dF_0(x_i)
\]

\[
= \alpha - (1 - (1 - \alpha)^2)/2 = \alpha^2/2.
\]

The theorem follows by noticing that

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

\[\square\]
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

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

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

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

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

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

Proof. By Theorem 5.3.5 and (5.3),

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

\[\square\]

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 $\alpha_B$ is greater than $\alpha$ 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
\[
\begin{align*}
m_0(x_i) &= \frac{1}{\sqrt{2\pi \sigma^2}} \exp\{-x_i^2 / (2\sigma^2)\} \\
m_1(x_i) &= \frac{1}{\sqrt{2\pi(\sigma^2 + \tau^2)}} \exp\{-x_i^2 / (2\sigma^2 + 2\tau^2)\},
\end{align*}
\]

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

By Lemma 5.4.2,
\[
1 - \Phi(x_i) = \frac{\exp\left\{ \left( \frac{\tau^2}{\sigma^2 + \tau^2} + 1 \right) \log(\alpha \sqrt{\frac{\tau^2}{2}}) \right\} \exp\left\{ \left( \frac{\tau^2}{\tau^2 + \sigma^2} + 1 \right) \log\left( \frac{-2 \log(\alpha \sqrt{\frac{\tau^2}{2}})}{\tau^2 + \sigma^2} \right) \right\}}{\sqrt{2\pi} \sqrt{-2\log(\sqrt{\frac{\tau^2}{2}}) - \log(-2 \log(\alpha \sqrt{\frac{\tau^2}{2}}))}} \left( 1 + o(1) \right)
\]
\[
= \left( \frac{\alpha \sqrt{\frac{\tau^2}{2}}}{\sqrt{2\pi}} \right)^{\tau^2/(\sigma^2 + \tau^2) + 1} \left( \frac{-2 \log(\alpha \sqrt{\frac{\tau^2}{2}})}{\tau^2 + \sigma^2} \right)^{\tau^2/(\sigma^2 + \tau^2)} \left( 1 + o(1) \right)
\]
\[
= O(\alpha^{\tau^2/(\sigma^2 + \tau^2) + 1} (- \log \alpha)^{\tau^2/(\sigma^2 + \tau^2)} (1 + o(1))).
\]

Therefore, \( v_i \) is greater than \( \alpha \) for small \( \alpha \), showing that \( \alpha_B \) 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 \( \alpha \), 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

<table>
<thead>
<tr>
<th>region</th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>((B_{01}(x_1), B_{01}'(x_2)))</td>
<td>((s_1, s_2))</td>
<td>((s_1, \psi(s_2)))</td>
<td>((\psi(s_1), s_2))</td>
</tr>
<tr>
<td>decision</td>
<td>reject (M_0^1) &amp; (M_0^2)</td>
<td>reject (M_0^1), accept (M_0^2)</td>
<td>accept (M_0^1)</td>
</tr>
</tbody>
</table>

Table 5.6: Conditional probabilities up to a normalizing constant

<table>
<thead>
<tr>
<th>(P_{00}(\cdot \mid \chi_{s_1,s_2}))</th>
<th>I</th>
<th>II</th>
<th>III</th>
</tr>
</thead>
<tbody>
<tr>
<td>(P_{00}(\cdot \mid \chi_{s_1,s_2}))</td>
<td>(f_0^<em>(s_1)f_0^</em>(s_2))</td>
<td>(f_0^<em>(s_1)f_1^</em>(s_2))</td>
<td>(f_1^<em>(s_1)(f_0^</em>(s_2) + f_1^*(s_2)))</td>
</tr>
<tr>
<td>(P_{01}(\cdot \mid \chi_{s_1,s_2}))</td>
<td>(f_0^<em>(s_1)f_1^</em>(s_2))</td>
<td>(f_0^<em>(s_1)\frac{1}{\psi(s_2)}f_1^</em>(s_2))</td>
<td>(f_1^<em>(s_1)(f_0^</em>(s_2) + f_1^*(s_2)))</td>
</tr>
<tr>
<td>(P_{10}(\cdot \mid \chi_{s_1,s_2}))</td>
<td>(f_1^<em>(s_1)f_0^</em>(s_2))</td>
<td>(f_1^<em>(s_1)f_1^</em>(s_2))</td>
<td>(f_1^<em>(s_1)\frac{1}{\psi(s_2)}(f_0^</em>(s_2) + f_1^*(s_2)))</td>
</tr>
<tr>
<td>(P_{11}(\cdot \mid \chi_{s_1,s_2}))</td>
<td>(f_1^<em>(s_1)f_1^</em>(s_2))</td>
<td>(f_1^<em>(s_1)\frac{1}{\psi(s_2)}f_1^</em>(s_2))</td>
<td>(f_1^<em>(s_1)\frac{1}{\psi(s_2)}(f_0^</em>(s_2) + f_1^*(s_2)))</td>
</tr>
</tbody>
</table>

Error probabilities

- \(P_{00}(\text{reject } M_0^1 \lor M_0^2 \mid \chi_{s_1,s_2}) = P_{00}(\text{reject } M_0^1 \mid \chi_{s_1,s_2}) = \frac{I+II}{I+II+III} = \frac{f_0^*(s_1)}{f_0^*(s_1) + f_1^*(s_1)} = \frac{B_{00}}{B_{01} + 1}\)

- \(P_{01}(\text{reject } M_0^1 \mid \chi_{s_1,s_2}) = \frac{I+II}{I+II+III} = \frac{f_0^*(s_1)}{f_0^*(s_1) + f_1^*(s_1)(\frac{1+II}{I+II})}\)

- \(P_{10}(\text{reject } M_0^2 \mid \chi_{s_1,s_2}) = \frac{I}{I+II+III} = \frac{f_0^*(s_2)}{(1+\frac{1}{\psi(s_2)})(f_0^*(s_2) + f_1^*(s_2))} \leq \frac{f_0^*(s_2)}{f_0^*(s_2) + f_1^*(s_2)}\)

- \(P_{11}(\text{falsely reject } M_0^1 \lor M_0^2 \mid \chi_{s_1,s_2} \text{ serve}) = 0\)
Power

1. Power: Make all correct rejection

- \( P_{00} \) (correctly reject \( M_0^1, M_0^2 \mid \chi_{s_1, s_2} \)) = 0

- \( P_{01} \) (reject \( M_0^2 \mid \chi_{s_1, s_2} \)) = \frac{I + II + III}{I + II + III} = \frac{1}{\psi(s_1)} + \frac{1}{\psi(s_2)} + \frac{1}{\psi(s_1) + 1} = \frac{B_0^1}{B_0^1 + 1}

- \( P_{10} \) (reject \( M_1^1 \mid \chi_{s_1, s_2} \)) = \frac{I + II + III}{I + II + III} = \frac{1}{\psi(s_1)} + \frac{1}{\psi(s_1) + 1} = \frac{B_1^1}{B_1^1 + 1}

- \( P_{11} \) (reject \( M_1^1 \wedge M_0^2 \mid \chi_{s_1, s_2} \)) = \frac{I + II + III}{I + II + III} = \frac{1}{\psi(s_1) + \psi(s_2) + 1} = \frac{\psi(s_1) + \psi(s_2) + 1}{1 + \psi(s_1) + \psi(s_2)}

2. Marginalized power: \( P(\text{reject } M_0^1 \mid M_i^j) \)

5.4.2 Conditional frequentist test

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

\[
\frac{t, \frac{1}{\sqrt{2\pi}} e^{-t^2/2}}{t^2 + 1} \leq 1 - \Phi(t) \leq \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \cdot \frac{1}{t} \cdot \frac{1}{t^3}
\]

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

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 $\alpha$ is:

$$
\frac{1 - \Phi(z_{1-\alpha})}{\alpha} = \frac{\phi(z_{1-\alpha})}{z_{1-\alpha}} \frac{1}{\alpha}
$$

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

$$
= \left(1 + \frac{\log(-2 \log(\sqrt{2\pi}\alpha))}{2 \log(\sqrt{2\pi}\alpha)}\right)^{-1} (1+)
$$

$$
= 1 + o(1)
$$

\(\square\)

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

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

then

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

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

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

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

To prove $\lim_{\alpha \to 0} k(\alpha) = 0$, 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) $\alpha = G(k(\alpha))$. 

97
Since
\[ \lim_{\alpha \to 0} \alpha = \lim_{\alpha \to 0} G(k(\alpha)) = 0 \]

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

The exponential distribution satisfies Theorem 5.4.3.
Bibliography


Ioannidis, J. P. (2005), “Why most published research findings are false,” *Chance*, 18, 40–47.


101


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.