Bayesian Analysis (2021) **16**, Number 1, pp. 111–128

Comparison of Bayesian and Frequentist Multiplicity Correction for Testing Mutually Exclusive Hypotheses Under Data Dependence*

Sean Chang† and James O. Berger‡

Abstract. The problem of testing mutually exclusive hypotheses with dependent test statistics is considered. Bayesian and frequentist approaches to multiplicity control are studied and compared to help gain understanding as to the effect of test statistic 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, when there is considerable test statistic dependence.

Keywords: multiple hypothesis testing, multiplicity correction, false positive probability, Bayesian inference.

MSC2020 subject classifications: Primary 62C10; secondary 62F05.

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. The asymptotic behavior of false discovery rate has 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), Berger et al. (2014) for a two-groups model, variable-selection in linear models, and subset analysis, respectively.

*Supported in part by NSF grants DMS-1007773 and DMS-1407775

†Department of Statistical Science Box 90251 Duke University Durham, NC 27705, U.S.A., sean.chang@duke.edu

‡Department of Statistical Science Box 90251 Duke University Durham, NC 27705, U.S.A., berger@stat.duke.edu

 https://doi.org/10.1214/20-BA1196

One of the appeals of the Bayesian approach to multiplicity control is that it does not depend on the dependence structure of the test statistics; 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 dependence. In contrast, frequentist approaches to multiplicity control are usually highly affected by test statistic 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 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. (2008), 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 test statistic dependence. The data $\boldsymbol{X} = (X_1, \ldots X_n)'$ arises from the multivariate normal distribution

$$\boldsymbol{X} \sim multinorm\left(\begin{pmatrix}\theta_1\\ \theta_2\\ \vdots\\ \theta_n\end{pmatrix}, \begin{pmatrix}1 & \rho & \cdots & \rho\\ \rho & 1 & \cdots & \rho\\ \vdots & \vdots & \ddots & \vdots\\ \rho & \rho & \cdots & 1\end{pmatrix}\right), \tag{1.1}$$

where ρ is the correlation between the observations; and $\boldsymbol{x}$ is a realization of $\boldsymbol{X}$. Consider testing the n hypotheses $M_0^i : \theta_i = 0$ versus $M_1^i : \theta_i \neq 0$, but under the assumption that *at most one* alternative hypothesis could be true. (It is possible that no alternative is true.) Although our study of this problem is pedagogical in nature, such testing problems can arise in signal detection, when a signal could arise in one and only one of n channels, and there is common background noise in all channels, leading to the equal correlation structure. More generalized results may be obtained by carefully investigate signals and noises among exponential number of models. We will, for convenience in exposition, use this language in referring to the situation.

In Section 2 we introduce two natural frequentist procedures for multiplicity control in this problem and, in Section 3, we introduce a natural Bayesian procedure. Section 4 explores a highly curious phenomenon that is encountered when ρ is near 1; when $n > 2$, the Bayesian procedure finds the true alternative hypothesis with near certainty, while an ad hoc frequentist procedure fails to do so. Sections 5 and 6 study the frequentist properties of the original Bayesian procedure and a Type-II maximum likelihood estimation (MLE) approach, showing that, as $n \to \infty$, the Bayesian procedures have strong frequentist control of error. Section 7 considers the situation in which there is a data sample of growing size m for each θ_i. Most omitted proofs can be found in the supplementary (cf. Chang and Berger, 2020); an added prefix 'S-' refers to the theorem or equation in the supplementary.

2 Frequentist Multiplicity Control

Two natural frequentist procedures are considered, and motivated by Donoho-Johnstone type hard thresholding in Donoho and Johnstone (1994).

2.1 An Ad Hoc Procedure

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

$$\alpha = P\left(\max_{1\le j\le n}|X_j| > c \;\middle|\; \theta_i = 0\ \forall i\right). \tag{2.1}$$

Lemma 2.1. *(2.1) 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 S.1, under the null model, X_i can be written as $X_i = \sqrt{\rho}Z + \sqrt{1-\rho}Z_i$, where the Z and the Z_i are independent standard normal random variables. Thus

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

□

Corollary 2.2. *The α and the cutoff c in the ad hoc procedure (2.1) have these properties:*

- *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.*

The extreme effect of dependence 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 Likelihood Ratio Test

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

Theorem 2.3. *The test statistic arising from the likelihood ratio test is*

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

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

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 $\rho \to 1$, $T \asymp n^2(1-\rho)^{-1} \max_i (x_i - \bar{x})^2$, which exhibits a quite different behavior that will be discussed later. Notice that the null distribution of T is well-behaved in ρ, i.e., the cutoff c in the likelihood ratio test is bounded as $\rho \to 1$ (Lemma S.2).

3 A Bayesian Test

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

$$\begin{aligned} &M_0 : \theta_1 = \ldots = \theta_n = 0 \,(\textit{null model})\,, \\ &M_i : \theta_i \neq 0,\ \theta_{(-i)} = \mathbf{0}\,, \end{aligned} \tag{3.1}$$

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

A simple prior assumption for an nonzero θ_i (if any) is:

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

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

$$\begin{aligned} m_0(\boldsymbol{x}) &\sim N(\mathbf{0}, \Sigma_0)\,, \\ m_i(\boldsymbol{x}) = \textstyle\int f(\boldsymbol{x} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} &\sim N\left(\mathbf{0}, \Sigma_i\right)\,, \end{aligned} \tag{3.2}$$

where

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

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

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

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

We will consider the prior $P(M_0) = r$ for some $r \in (0,1)$ and $P(M_i) = (1-r)/n$ for alternative models in the rest of the article.

Theorem 3.1. *For any $\rho \in [0,1)$ and positive integer $n > 1$, then the null posterior probability is:*

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

and the posterior probability of an alternative model M_i is

$$P(M_i \mid \boldsymbol{x}) = \left\{ \begin{array}{l} \sqrt{1+a\tau^2}\left(\frac{n\,r}{1-r}\right) \exp\left\{ \frac{-\tau^2}{2(1+\tau^2 a)} \left(\frac{x_i}{1-\rho} + n\bar{\boldsymbol{x}} b \right)^2 \right\} \\ + \sum_{k=1}^{n} \exp\left\{ \frac{-\tau^2}{2(1+\tau^2 a)} \left(\frac{x_i^2 - x_k^2}{(1-\rho)^2} + \frac{2b}{1-\rho}(n\bar{\boldsymbol{x}})(x_i - x_k) \right) \right\} \end{array} \right\}^{-1}, \quad (3.3)$$

where $a = a_n = \frac{1+(n-2)\rho}{(1+(n-1)\rho)(1-\rho)}$, $b = b_n = \frac{-\rho}{(1+(n-1)\rho)(1-\rho)}$.

Corollary 3.2. *In particular, when $n = 2$, the null posterior probability is:*

$$P(M_0 \mid \boldsymbol{x}) = \left\{ 1 + \frac{1-r}{2r} \sqrt{\frac{1-\rho^2}{1-\rho^2+\tau^2}} \sum_{i \in \{1,2\}} \exp\left\{ \frac{\tau^2}{2(1-\rho^2+\tau^2)} \left(\frac{x_i - \rho x_{(-i)}}{\sqrt{1-\rho^2}} \right)^2 \right\} \right\}^{-1}, \quad (3.4)$$

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

$$P(M_i \mid \boldsymbol{x}) = \left\{ \begin{array}{l} \sqrt{\frac{1-\rho^2+\tau^2}{1-\rho^2}} \left(\frac{2r}{1-r} \right) \exp\left\{ \frac{-\tau^2}{2(1-\rho^2+\tau^2)} \frac{(x_i - \rho x_{(-i)})^2}{1-\rho^2} \right\} \\ +1 + \exp\left\{ \frac{-\tau^2}{2(1-\rho^2+\tau^2)} (x_i^2 - x_{(-i)}^2) \right\} \end{array} \right\}^{-1}.$$

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 4.1. *If $n = 2$, $i \in \{1,2\}$ and $\rho \to 1$, then:*

$$P(M_0 \mid \boldsymbol{X}) \to 1 \quad \textit{under the null model},$$
$$P(M_i \mid \boldsymbol{X}) \to \left(1 + \exp\left\{ \frac{-1}{2} \left(X_i^2 - X_{(-i)}^2 \right) \right\} \right)^{-1} \textit{ under } M_1 \textit{ or } M_2\,.$$

If $n > 2$, $i, j \in \{0, 1, \ldots, n\}$ and $\rho \to 1$, under model M_j:

$$P(M_i \mid \boldsymbol{X}) \to \delta_i^j = \begin{cases} 1 \textit{ if } i = j, \\ 0 \textit{ else}\,. \end{cases}$$

Theorem 4.2. *The likelihood ratio test (Theorem 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 S.3:

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

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

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

For both $j = i$ or $j = (-i)$, the corresponding likelihood ratios go to infinity at the same asymptotic rate since $\left[\theta_i/(2\sqrt{1-\rho}) + (z_i - z_{(-i)})/2\right]^2 = \left[-\theta_i/(2\sqrt{1-\rho}) - (z_i - z_{(-i)})/2\right]^2$. When $n > 2$, under M_j, when $\rho \to 1$:

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

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

From Theorem 4.1 and 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 using (S.2).

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

$$x_i - x_j = [\theta_i + \sqrt{\rho}z + \sqrt{1-\rho}z_i] - [\theta_j + \sqrt{\rho}z + \sqrt{1-\rho}z_j] = \begin{cases} 0 \text{ under } M_0\,, \\ \theta_i \text{ or } -\theta_j \text{ else}\,. \end{cases}$$

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 θ_i or θ_j is nonzero.

Note that the ad hoc frequentist test does not have this behavior. As $\rho \to 1$, the test still has probability α of incorrectly rejecting a true M_0; and has positive probability of not detecting a signal when M_i is true.

This highlights the danger (in terms of lack of power) of using 'intuitive' procedures for multiplicity control.

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.

5.1 Posterior Probabilities

Lemma 5.1. *As $n \to \infty$ under the null model,*

$$P(M_i \mid \boldsymbol{x}) = \left(1 + \frac{n}{1-r}\sqrt{\frac{1-\rho+\tau^2}{1-\rho}}\exp\left\{\frac{-\tau^2}{2(1-\rho+\tau^2)}\left(\frac{x_i-\bar{\boldsymbol{x}}}{\sqrt{1-\rho}}\right)^2\right\}\right)^{-1}(1+o(1)) \tag{5.1}$$

almost surely.

Remark 5.2. Figure 1 shows the ratio of the estimated $P(M_1 \mid \boldsymbol{x})$ (from Lemma 5.1) and the true probability (from Theorem 3.1), as n grows. Each plot contains 200 different ratio curves based on independent simulations with fixed $\rho, P(M_0)$ and τ. As can be seen, the ratio goes to 1 when n grows and the convergence rate indeed depends on the correlation.

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 5.3. *As $n \to \infty$ and $\rho \in [0,1)$, under the null model,*

$$P(M_0 \mid \boldsymbol{X}) \to P(M_0)\,.$$

Proof. First note that

$$\begin{cases} a_n = \frac{1}{1-\rho} + \frac{-\rho}{(1-\rho)(1+(n-1)\rho)} = \frac{1}{1-\rho} + O(1/n)\,, \\ nb_n = \frac{-1}{1-\rho} + \frac{1-\rho}{(1-\rho)(1+(n-1)\rho)} = \frac{-1}{1-\rho} + O(1/n)\,. \end{cases} \tag{5.2}$$

The summation term in the null posterior (Theorem 3.1) becomes

$$\begin{aligned} &\left(\frac{1-r}{nr}\right)\frac{1}{\sqrt{1+\tau^2/(1-\rho)}}\sum_1^n \exp\left\{\frac{\tau^2}{2(1+\tau^2/(1-\rho))}\left[\frac{x_i-\boldsymbol{x}}{1-\rho}\right]^2\right\}(1+o(1)) \\ &= \left(\frac{1-r}{r}\right)1/n\sqrt{\frac{1-\rho}{1-\rho+\tau^2}}\sum_1^n \exp\left\{\frac{\tau^2}{2(1-\rho+\tau^2)}z_i^2\right\}(1+o(1)) \text{ (by Lemma 5.1)} \\ &\to \frac{1-r}{r} \quad \text{(by the Strong Law of Large Numbers).} \end{aligned}$$

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

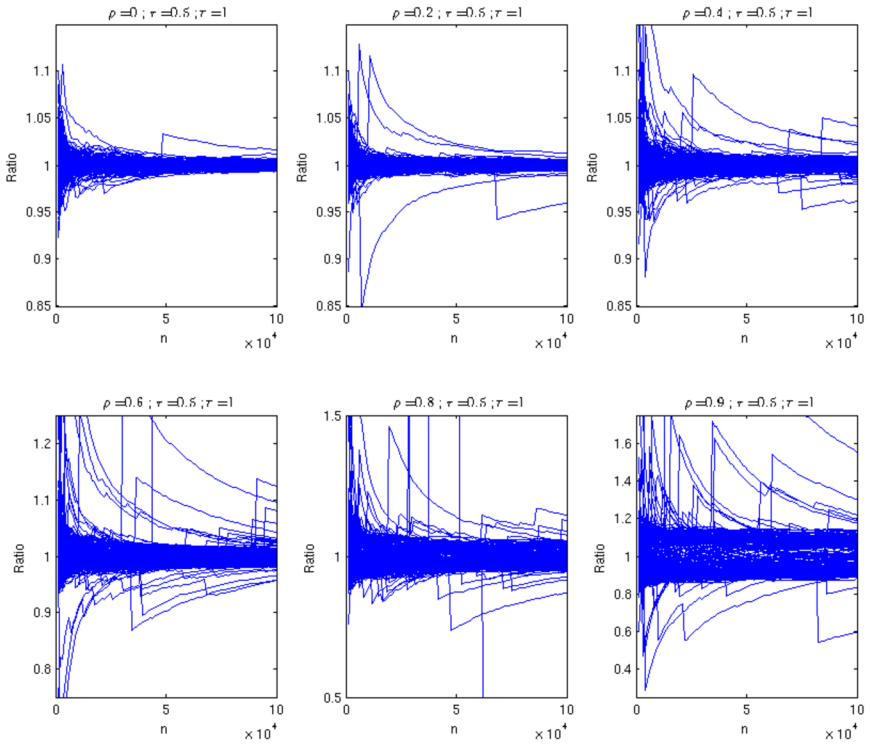

Figure 1: Ratio of estimated and true posterior probability of M_1 as n grows under the null model and fixed τ, r, different ρ. Each subplot is for different correlations and contains 200 simulations.

Remark 5.4. Figure 2 shows simulations of the null posterior probability for different numbers of hypotheses and different correlations. Interestingly, by Theorem 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 5.3, this posterior probability converges to its prior probability regardless of the correlation. This convergence can be seen in Figure 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 5.5 (Bayesian detection criterion)**.** Reject the null model M_0 if any alternative model has posterior probability greater than a specified threshold $p \in (0, 1)$, i.e. $\max_{1 \leq i \leq n} P(M_i \mid \boldsymbol{x}) > p$. Otherwise, accept M_0.

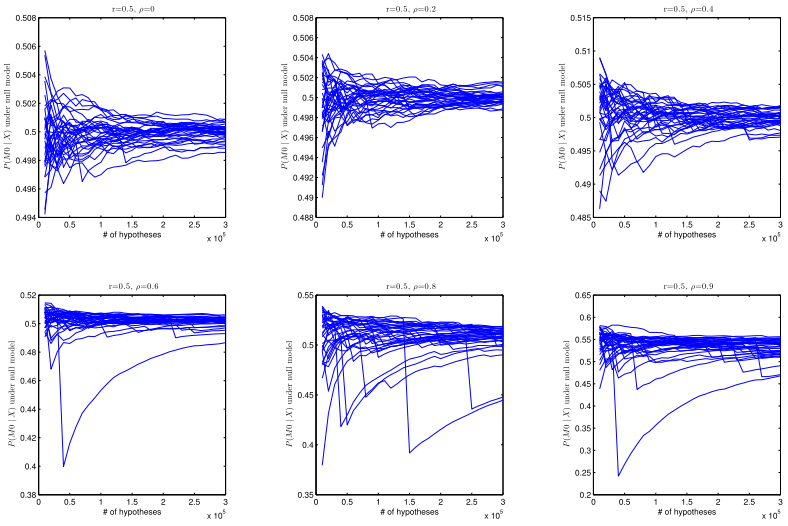

Figure 2: Convergence of $P(M_0 \mid \boldsymbol{x})$ to the prior probability (0.5) under the null model. Each subplot has a different correlation and contains 50 simulations.

Definition 5.6 (False positive probability, FPP)**.** Under the null model, the FPP is the frequentist probability of accepting a non-null model.

Theorem 5.7 (False positive probability)**.** *Under the null model, as* $n \to \infty$*,*

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

In the situations when users have a good sense of prior distribution, variance of the prior τ^2 can be assigned as a fixed number, which yields a false positive probability that goes to zero at a polynomial rate. On the other hand, if there is no good a priori knowledge of the variance, empirical Bayes offers an estimation of τ^2. The following section discusses empirical Bayes approach of the problem.

6 Adaptive Choice of τ^2

To increase the frequentist power of the Bayes test, we consider adaptive choices of τ^2. First, we consider the choice that maximizes the false positive probability. Then we consider a Type II maximum likelihood approach based on estimating τ^2.

6.1 The Adaptive τ^2 Which Maximizes FPP

Theorem 6.1. *Given null model prior probability* r*, correlation* ρ*, and decision threshold* p*, as* $n \to \infty$*, the choice of* τ^2 *that maximizes FPP is*

$$\tau_n^2 = (1-\rho)[2\log n + \log\log n + 2\log\frac{p}{(1-p)(1-r)} + \log 2]\,. \tag{6.1}$$

The resulting FPP is

$$\begin{aligned} &P(\textit{false positive} \mid \textit{null model}\,, \tau_n^2) \\ &= \left(e^{-1/2}\sqrt{\frac{2}{\pi}}\right)\left(\frac{(1-p)(1-r)}{p}\right)\left(2\log\, n + \log\log\, n + c_\tau\right)^{-1}(1+o(1))\,, \end{aligned} \tag{6.2}$$

where $c_\tau = 2\log\frac{p}{(1-p)(1-r)} + \log 2 + 1$.

Proof. Without loss of generality, assume $\max_i z_i^2 = z_1^2$. By the model selection criteria (S.20), z_1 is a false positive if:

$$z_1^2 \geq 2\left(1+\frac{1-\rho}{\tau_n^2}\right)\log\left(\frac{n\,p}{(1-p)(1-r)}\sqrt{\frac{1-\rho+\tau_n^2}{1-\rho}}\right) + o(1)\,. \tag{6.3}$$

Lemma S.21 establishes that (6.1) maximizes the FPP and, with this choice of τ_n^2, the rejection region becomes

$$z_1^2 > 2\,\log n + \log\log\, n + \underbrace{2\log\frac{p}{(1-p)(1-r)} + 1 + \log 2}_{c(p,r)} + o(1)\,. \qquad z \tag{6.4}$$

Finally,

$$\begin{aligned} &P(\text{false positive} \mid \hat\tau_n^2, p, r) \\ &= 1 - \left\{1 - 2\left\{\frac{\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{1}{2}2\left(1+\frac{1-\rho}{\hat\tau_n^2}\right)\log\left(\frac{n}{1-r}\frac{p}{1-p}\sqrt{\frac{1-\rho+\hat\tau_n^2}{1-\rho}}\right)+o(1)\right\}}{\sqrt{2\left(1+\frac{1-\rho}{\hat\tau_n^2}\right)log\left(\frac{np}{(1-r)(1-p)}\sqrt{\frac{1-\rho+\hat\tau_n^2}{1-\rho}}\right)+o(1)}} + o\left(\frac{1}{n\,(\log\, n)^2}\right)\right\}\right\}^n \\ &= 1 - \left\{1 - \sqrt{\frac{2}{\pi}}\frac{\left(\frac{np}{(1-p)(1-r)}\sqrt{1+c_{n,\tau}}\right)^{-\left(1+c_{n,\tau}^{-1}\right)}(1+o(1))}{\sqrt{\left(1+c_{n,\tau}^{-1}\right)2\log\left(\frac{np}{(1-r)(1-p)}\sqrt{1+c_{n,\tau}}\right)}}\right\}^n \\ &\qquad\qquad \text{where } c_{n,\tau} = 2\log\, n + \log\log\, n + c_\tau \\ &= 1 - \left\{1 - \frac{1}{n}\underbrace{\frac{\sqrt{2}\frac{(1-p)(1-r)}{p}\left(\frac{np}{(1-p)(1-r)}\sqrt{1+c_{n,\tau}}\right)^{-c_{n,\tau}^{-1}}}{\sqrt{\pi\left(1+c_{n,\tau}\right)\left(2\log n + 2c + \log\log\, n + \log 2 + 1\right)}}}_{d_n}(1+o(1))\right\}^n \end{aligned}$$

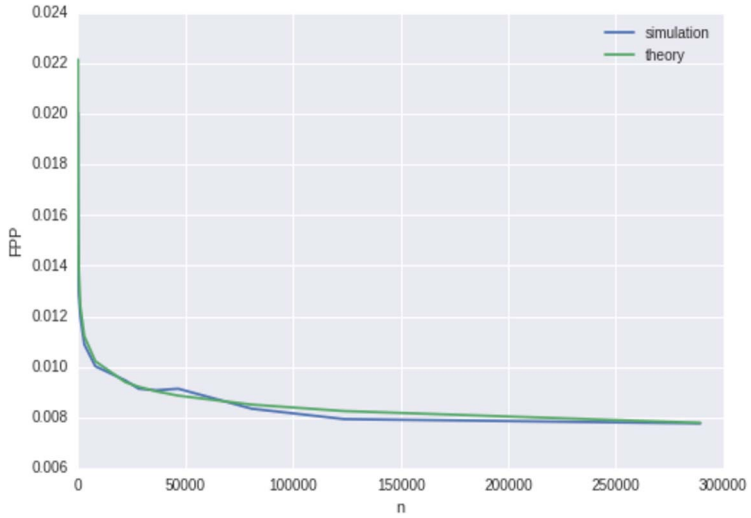

Figure 3: Comparison of the simulated FPP and its asymptotic approximation when $p = r = 0.5$, $\rho = 0$ as n varies from 10^1 to $3e^5$, τ^2 is the adaptive choice.

$$= d_n(1 + o(1)) + O\left(\frac{1}{(\log\, n + \log\log\, n)^2}\right)$$
$$= \left(e^{-1/2}\sqrt{\frac{2}{\pi}}\right)\left(\frac{(1-p)(1-r)}{p}\right)\left(2\log\, n + \log\log\, n + c_\tau\right)^{-1}(1 + o(1))\,. \qquad \square$$

So, with this adaptive choice of τ^2, the FPP only goes to zero at a $1/(\log n + \log\log n)$ rate, much slower than the polynomial rate achieved for fixed τ^2.

Remark 6.2. Figure 3 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 $1/(2\log n + \log\log n)$. Note that the FPP does not become extremely small even for very large n.

6.2 Type II Maximum Likelihood Estimation of τ^2

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

Lemma 6.3. *Let $\tilde{L}_n(\tau^2)$ be the marginal likelihood of τ^2 given $(x_1, \ldots, x_n)$, namely*

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

Defining

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

the Type II MLE, $\hat{\tau}_n^2$, can be found as

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

Proof.

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

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

Theorem 6.4 (Type II MLE false positive probability)**.** *Given null prior probability $P(M_0) = r$, correlation ρ, and decision threshold p, as $n \to \infty$*

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

where k^* satisfies:

$$-2\,\log\left(\sqrt{\pi}\left(\frac{1}{k^*} - \frac{1}{2}\right)\right) = \log\, k^* + 2\,\log\left(\frac{p}{(1-p)(1-r)}\right) + 2\left(\frac{1}{k^*}\right). \tag{6.5}$$

Proof. First, Lemma S.21 shows that (6.4) provides the absolute lower bound for z_1^2 to be in the rejection region; namely $2\log n + \log\log n + c(p,r)$. So the rejection region, denote it by Ω, corresponding to the Type II MLE choice of τ^2 must be a subset of $(2\log n + \log\log n + c(p,r), \infty)$. Divide this interval into

$$\begin{aligned}
\Omega_1 &= (2\log n + \log\log n + c(p,r), 2\log n + \log\log n + c(p,r) + K)\,, \\
\Omega_2 &= (2\log n + \log\log n + c(p,r) + K, \infty)\,,
\end{aligned}$$

where K will be chosen large, but fixed. We first determine $\Omega \cap \Omega_1$.

For any $z_1^2 = 2\log n + \log\log n + c \in \Omega_1$, Lemma S.27, shows that the Type-II MLE estimate is

$$\hat{\tau}_n^2 = (1-\rho)k(c)\ \log\ n(1+o(1)) \text{ where } k(c) = \left(1/2 + \exp\left\{-c/2\right\}/\sqrt{\pi}\right)^{-1}.$$

Thus, letting $z_1^{*2} = 2\log n + \log\log n + c^*$ denote the smallest value in $\Omega \cap \Omega_1$ (if it exists) and letting $\hat{\tau}_n^{*2} = (1-\rho)k(c^*)\ \log\ n(1+o(1))$ denote the corresponding Type-II MLE estimate, the smallest value must satisfy, by (6.3),

$$\begin{aligned}
z_1^{*2} &= 2\left(1 + \frac{1-\rho}{\hat{\tau}_n^{*2}}\right) ln\left(\frac{n}{1-r}\frac{p}{1-p}\sqrt{\frac{1-\rho+\hat{\tau}_n^{*2}}{1-\rho}}\right) + o(1) \\
&= 2\left(1 + \frac{1}{k(c^*)\ \log\ n(1+o(1))}\right) \\
&\qquad * \left(\log\ n + \log\frac{p}{(1-p)(1-r)} + \frac{1}{2}\log(1 + k(c^*)\ \log\ n\,(1+o(1)))\right) + o(1) \\
&= 2\ \log\ n + \log\ \log\ n + \underbrace{\left[\log\ k(c^*) + 2\ \log\left(\frac{p}{(1-p)(1-r)}\right) + 2\left(\frac{1}{k(c^*)}\right)\right]}_{l^*} + o(1)\Bigg).
\end{aligned}$$

This is equivalent to

$$\begin{aligned}
c^* &= \log\ k(c^*) + 2\ \log\left(\frac{p}{(1-p)(1-r)}\right) + 2\left(\frac{1}{k(c^*)}\right) \\
&= -\log\left(\frac{1}{2} + \frac{\exp\left\{-c^*/2\right\}}{\sqrt{\pi}}\right) + 2\ \log\ \left(\frac{p}{(1-p)(1-r)}\right) + 1 + 2\frac{\exp\left\{-c^*/2\right\}}{\sqrt{\pi}},
\end{aligned} \tag{6.6}$$

which, using Lemma S.28 (which shows that $l^* > c(p,r)$), can easily be shown to have a unique solution in $\Omega \cap \Omega_1$ (assuming K is larger than, say, $4\ \log\left(\frac{p}{(1-p)(1-r)}\right)$). It is also then easy to show that

$$\Omega \cap \Omega_1 = (2\ \log\ n + \log\ \log\ n + l^* + o(1), 2\ \log\ n + \log\ \log\ n + c(p,r) + K).$$

By S.4,

$$\begin{aligned}
&\frac{P(\Omega_2)}{P(\Omega \cap \Omega_1)} \\
&\leq \frac{\exp\left(-[2\log n + \log\log n + c(p,r) + K]/2\right)/\sqrt{2\log n + \log\log n + c(p,r) + K}}{\frac{\sqrt{2\log n + \log\log n + l^*}\exp\left(-[2\log n + \log\log n + l^*]/2\right)}{2\log n + \log\log n + c(p,r) + l^*} - \frac{\exp\left(-[2\log n + \log\log n + c(p,r) + K]/2\right)}{\sqrt{2\log n + \log\log n + c(p,r) + K}}} \\
&= \left(\exp\left([c(p,r) + K - l^*]/2\right) - 1\right)^{-1}(1 + o(1)).
\end{aligned}$$

$c(p,r)$ and l^* are fixed, we can clearly choose K large enough to make this smaller than any specified ϵ. Hence the region Ω_2 can be ignored in the computation of the FPP. (It is almost certainly part of the rejection region, but we do not know what $\hat{\tau}_n^2$ is for observations in that region and, hence can't say for sure.)

We can also use the same argument to say that

$$P(\Omega \cap \Omega_1) = P((2\log n + \log\log n + l^*, \infty))(1+\epsilon)\,.$$

Writing k^* for $k(c^*)$, it follows that the FPP is

$$\begin{aligned} FPP &= 1 - \left\{1 - \frac{\frac{2}{\sqrt{2\pi}}\exp\left\{\frac{-1}{2}(2\log n + \log\log n + l^*)\right\}(1+o(1))(1+\epsilon)}{\sqrt{2\log n + \log\log n + l^*}}\right\}^n \\ &= 1 - \left\{1 - \frac{1}{n}\sqrt{\frac{2}{\pi}}\frac{(1-p)(1-r)\exp(-1/k^*)}{p\sqrt{k^*}\sqrt{(\log n)(2\log n + \log\log n)}}(1+o(1))(1+\epsilon)\right\}^n \\ &= \sqrt{\frac{2}{\pi k^*}}\exp\left\{\frac{-1}{k^*}\right\}\left(\frac{(1-p)(1-r)}{p}\right)\Big[(\log n)(2\log n + \log\log n)\Big]^{-1/2} \\ &\qquad\qquad (1+o(1))(1+\epsilon) \\ &= \left(\frac{1}{k^*} - \frac{1}{2}\right)\frac{1}{\log n}(1+o(1))(1+\epsilon) \quad \text{by (6.6)}\,. \end{aligned}$$

Since ϵ can be made arbitrarily small, the result follows. □

Note that (6.5) can be solved numerically. For instance, when $p = r = 0.5$, $k^* \approx 1.6142$. The solution of $\frac{1}{k^*} - \frac{1}{2}$ with respect to $\frac{p}{(1-r)(1-p)}$ is, indeed, given in Figure 5. The Type II MLE FPP converges to 0 at a logarithmic rate in n, as did the maximal Bayesian FPP. Thus both are far less conservative than the Bayesian procedures with specified τ^2. Finally, it is interesting that neither of the adaptive asymptotic FPP's depend on ρ.

Remark 6.5. Figure 4 demonstrates how the threshold p (Definition 5.5) can be chosen to achieve a fixed FPP of 0.05. Because, for a fixed p, the FPP goes to zero as a function of n, smaller p are needed to achieve a fixed FPP as n grows. Note that the variation in p is actually quite small over the very large range of n considered in the figure.

Remark 6.6. Figure 5 gives the value of $\frac{1}{k^*} - \frac{1}{2}$ for different $\frac{p}{(1-r)(1-p)}$.

Remark 6.7. Figure 6 demonstrates how the detection power varies when the signal size increases.

7 Analysis as the Information Grows

In this section, we generalize model (3.1) to the scenario where each channel has m i.i.d. observations. Then the sample mean satisfies

$$\bar{\boldsymbol{X}} \sim multinorm\left(\begin{pmatrix}\theta_1\\ \theta_2\\ \vdots\\ \theta_n\end{pmatrix}, \frac{1}{m}\begin{pmatrix}1 & \rho & \cdots & \rho\\ \rho & 1 & \cdots & \rho\\ \vdots & \vdots & \ddots & \vdots\\ \rho & \rho & \cdots & 1\end{pmatrix}\right)\,. \tag{7.1}$$

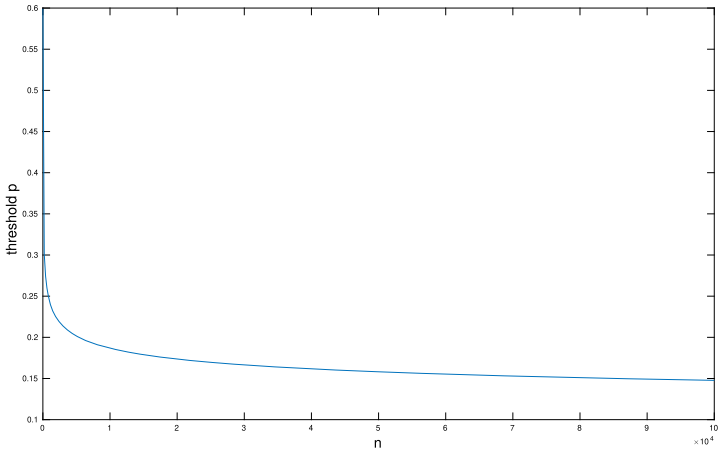

Figure 4: For fixed prior probability of 0.5 for the null model, this gives, as the number of hypotheses n increases, the Bayesian threshold probability p that would achieve an FPP of $= 0.05$.

Hence, m can be seen as the precision of $\bar{\boldsymbol{X}}$. More generally, we will replace $1/m$ by a function σ_n^2, where σ_n^2 decreases to zero as n grows.

The theorem below gives the rate of decrease of σ_n^2 which guarantees consistency. For the i.i.d. case, consistency of all models is only guaranteed if m grows faster than $\log n$; consistency fails if m grows slower than $\log n$; and consistency depends on the parameter value if m is $O(\log n)$.

Theorem 7.1. *Consider model (3.1), with the altered covariance matrix below:*

$$\boldsymbol{X} \sim multinorm\left(\begin{pmatrix}\theta_1\\ \theta_2\\ \vdots\\ \theta_n\end{pmatrix}, \sigma_n^2\begin{pmatrix}1 & \rho & \cdots & \rho\\ \rho & 1 & \cdots & \rho\\ \vdots & \vdots & \ddots & \vdots\\ \rho & \rho & \cdots & 1\end{pmatrix}\right), \tag{7.2}$$

1. *When $\sigma_n^2 \log n \to 0$, consistency holds for both the null and alternative models.*
2. *When $\sigma_n^2 \log n \to d \in (0, \infty)$,*
   - *Under M_0: $P(M_0 \mid \boldsymbol{X}) \to \left(1 + \frac{1-r}{r}\left[2\Phi\left(\frac{(1-\rho)d}{\tau^2}\right) - 1\right]\right)^{-1}$, failing to be consistent.*
   - *Under an alternative model M_j, if $d \in (0, \frac{\theta_j^2}{2(1-\rho)})$, consistency holds for M_j, whereas consistency does not hold otherwise.*
3. *When $\sigma_n^2 \log n \to \infty$ and $\sigma_n^2 \log n = o(\log n)$, consistency does not hold for any model. In addition, when the null hypothesis is true,*

$$P(M_0 \mid \boldsymbol{X}) \to P(M_0)\,.$$

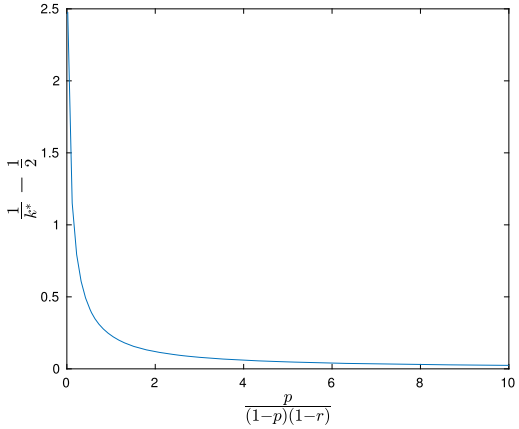

Figure 5: Solution of $\frac{1}{k^*} - \frac{1}{2}$ (y-axis) with respect to different $\frac{p}{(1-p)(1-r)}$ (x-axis).

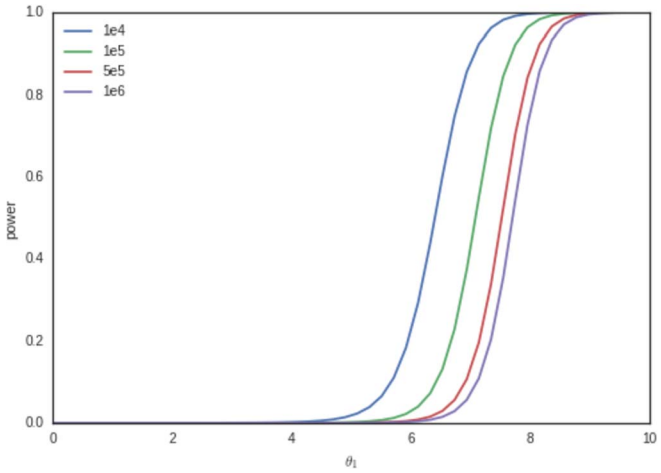

Figure 6: Power versus θ_i for fixed $p = r = 0.5$, $\rho = 0$, and different n (see the color legend on the top left). Each point is the average acceptance rate of the true non-null model when θ_i is as specified on the x-axis.

8 Conclusions

The main purpose of this work was to gain understanding of the behavior of Bayesian procedures that control for multiple testing, under a scenario of high dependence among test statistics, where frequentist methods for multiplicity control become more difficult to implement when trying to maintain high power. In Section 4, the Bayesian procedure was shown to have unexpectedly high power as the correlation gets large, providing an illustration of the gains that can be had by approaching multiplicity control from the

Bayesian side. (Bayes theorem often produces things that we could not have produced through our intuition alone.)

The other main issue concerning the behavior of the Bayesian procedure is the extent to which it also exhibits desirable frequentist control. Surprising to us was that the Bayesian procedure exhibited too-strong frequentist control, with the FPP (false positive probability under the null model) going to zero at a polynomial rate, as the number n of tests grows. To a Bayesian who believed in the prior distribution that was utilized this would not be viewed as a problem, but we tend to prefer procedures that have a dual Bayesian/frequentist interpretation. To this end, adaptive versions of the Bayesian procedure were considered, and found to have FPP's going to 0 at the much slower $1/\log n$ rate; indeed, unless n is huge, the resulting FPPs were reasonably moderate.

A number of other surprises were also encountered, such as the fact that, as the number of tests n grows, the posterior probability of the null model converges to its prior probability. (This is actually a very general phenomenon that will be reported elsewhere.) The situation of having i.i.d. replicate observations m was also considered, and it was shown that one needs m to grow faster than $\log n$ to achieve consistency, under both the null and alternative models.

Methodologically, if a frequentist were to encounter this particular multiple testing problem and desired a procedure that is fully powered and achieves an FPP of α, we would suggest using the adaptive Bayesian procedure in Section 6.1. One solves (6.2) for the Bayesian rejection threshold p (with, say, the default choice of $r = 1/2$ for the prior probability of the null model), and then rejects the null and accepts M_i if $P(M_i \mid \boldsymbol{x}) > p$, where $P(M_i \mid \boldsymbol{x})$ is as in (3.1) with τ^2 chosen as in (6.1). This Bayesian procedure will have the unusual power benefits outlined in Section 4 when the correlation is high, while achieving the desired frequentist FPP (at least asymptotically) and likely having the greatest power against alternatives, since the τ^2 in (6.1) was chosen, in essence, to maximize the power.

Supplementary Material

Supplementary Material to "Comparison of Bayesian and Frequentist Multiplicity Correction for Testing Mutually Exclusive Hypotheses Under Data Dependence" (DOI: 10.1214/20-BA1196SUPP; .pdf).

References

Abramovich, F. and Angelini, C. (2006). "Bayesian maximum a posteriori multiple testing procedure." *Sankhyā: The Indian Journal of Statistics*, 436–460. MR2322194. 112

Abramovich, F., Benjamini, Y., Donoho, D. L., Johnstone, I. M., et al. (2006). "Adapting to unknown sparsity by controlling the false discovery rate." *The Annals of Statistics*,

34(2): 584–653. MR2281879. doi: https://doi.org/10.1214/009053606000000074. 111

Benjamini, Y. and Hochberg, Y. (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing." *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300. MR1325392. 111

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer Science & Business Media. MR0804611. doi: https://doi.org/10.1007/978-1-4757-4286-2. 121

Berger, J. O., Wang, X., and Shen, L. (2014). "A Bayesian approach to subgroup identification." *Journal of Biopharmaceutical Statistics*, 24(1): 110–129. MR3196130. doi: https://doi.org/10.1080/10543406.2013.856026. 111

Bogdan, M., Ghosh, J. K., Tokdar, S. T., et al. (2008). "A comparison of the Benjamini-Hochberg procedure with some Bayesian rules for multiple testing." In *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen*, 211–230. Institute of Mathematical Statistics. MR2462208. doi: https://doi.org/10.1214/193940307000000158. 112

Chang, S. and Berger, J. O. (2020). "Supplementary Material to "Comparison of Bayesian and Frequentist Multiplicity Correction for Testing Mutually Exclusive Hypotheses Under Data Dependence"." *Bayesian Analysis*. doi: https://doi.org/10.1214/20-BA1196SUPP. 112

Donoho, D. L. and Johnstone, J. M. (1994). "Ideal spatial adaptation by wavelet shrinkage." *Biometrika*, 81(3): 425–455. MR1311089. doi: https://doi.org/10.1093/biomet/81.3.425. 113

Efron, B. (2004). "Large-scale simultaneous hypothesis testing." *Journal of the American Statistical Association*, 99(465). MR2054289. doi: https://doi.org/10.1198/016214504000000089. 111

Guo, M. and Heitjan, D. F. (2010). "Multiplicity-calibrated Bayesian hypothesis tests." *Biostatistics*, 11(3): 473–483. 112

Noble, W. S. (2009). "How does multiple testing correction work?" *Nature Biotechnology*, 27(12): 1135–1137. 111

Scott, J. G. and Berger, J. O. (2006). "An exploration of aspects of Bayesian multiple testing." *Journal of Statistical Planning and Inference*, 136(7): 2144–2162. MR2235051. doi: https://doi.org/10.1016/j.jspi.2005.08.031. 111

Scott, J. G. and Berger, J. O. (2010). "Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem." *The Annals of Statistics*, 38(5): 2587–2619. MR2722450. doi: https://doi.org/10.1214/10-AOS792. 111

Storey, J. D. (2003). "The positive false discovery rate: A Bayesian interpretation and the q-value." *Annals of Statistics*, 2013–2035. MR2036398. doi: https://doi.org/10.1214/aos/1074290335. 111