University of Cincinnati

Date: 11/2/2021

I, Gonagala Mudiyanselage Nilupika K Herath, hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Mathematical Sciences.

It is entitled:

Some Aspects of Bayesian Multiple Testing

Student's name: **Gonagala Mudiyanselage Nilupika K Herath**

This work and its defense approved by:

Committee chair: Siva Sivaganesan, Ph.D.

Committee member: Hang Joon Kim, Ph.D.

Committee member: Seongho Song, Ph.D.

Committee member: Xia Wang, Ph.D.

40532

Some Aspects of Bayesian Multiple Testing

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Department of Mathematical Sciences
College of Arts and Sciences
University of Cincinnati, November 2021

Author: Gonagala Mudiyanselage Nilupika Kumari Herath
Degrees: B.S. Industrial Mathematics, 2013,
Rajarata University of Sri Lanka

Chair: Siva Sivaganesan, Ph.D.
Committee: Seongho Song, Ph.D.
Xia Wang, Ph.D.
Hang Joon Kim, Ph.D.

Abstract

Multiple testing in statistics refers to carrying out several statistical tests simultaneously. As the number of tests increases, the probability of incorrectly rejecting the null hypothesis also increases (multiplicity problem). Therefore, some multiplicity adjustment should always be considered to control the error rate. Making decisions without multiplicity adjustment can lead to error rates that are higher than the nominal rate. While several approaches to multiplicity adjustment are available, the Bayesian method is the only approach that inherently adjusts for multiplicity. This thesis considers the Bayesian approach to the multiple testing problem for different types of data: continuous and discrete data.

The first chapter explains the multiplicity problem and provides background information about the two popular approaches to multiplicity adjustment: frequentist and Bayesian approaches.

Chapter 2 investigates the sensitivity of the normality assumption of continuous response variables when the response variables instead follow a t-distribution. We focus on Bayesian multiple testing of means, or location parameters, when the response variable follows a t-distributions, determine suitable priors for the parameters, develop a computational strategy for computing the posterior probabilities of the hypotheses, and use it to study the sensitivity of the results under normality assumption using simulation studies. There is sensitivity to both the sampling model and the prior distribution of the location parameters for a small sample size. For a larger sample size, while the sensitivity to the sampling model is disappearing, the sensitivity to prior distribution is still present. However, as the degrees of freedom and the number of noises increase, sensitivity to prior distribution is also disappearing.

Chapter 3 and chapter 4 introduce two objective Bayesian multiple testing approaches for discrete data. In particular, while chapter 3 is about testing equality of two binomial proportions (two-sided alternatives), chapter 4 considers testing equality of two or more order-restricted binomial proportions (one-sided alternatives). The proposed work in chapters 3 and 4 has two novel contributions: providing a formal objective Bayesian approach to multiple testing of binomial proportions and developing a non-local prior approach to binomial testing.

Chapter 3 introduces two prior choices, Local and Threshold (non-local), to model the unknown proportions in testing the equality of two binomial proportions. While under the alternative hypothesis, both priors give equally fast convergence towards the true

alternative, under the null hypothesis, Threshold prior gives faster convergence towards the true null than the Local prior.

Chapter 4 first extends the use of Local and Threshold prior approaches for testing one-sided two binomial proportions. Chapter 4 gives similar results to chapter 3, i.e., the Threshold prior approach has faster convergence towards the true null hypothesis than the Local prior, and under the alternative hypothesis, both priors have equally fast convergence toward the true alternative. Later in chapter 4, we generalize the proposed Local prior approach for order-restricted testing of more than two binomial proportions.

Acknowledgments

First and foremost, I would like to convey my deepest gratitude to my supervisor Dr. Siva Sivaganesan for his endless support and encouragement towards completing my degree. Without his guidance and support, my Ph.D. study wouldn't have been a success. More than a supervisor, he was like a second father who listened to me, understood me, was patient with me, and backed me in every possible way. I was extremely fortunate to have such a wonderful advisor like him.

I am very grateful to my committee members, Dr. Seongho Song, Dr. Xia Wang and Dr. Hang Joon Kim for serving in my dissertation committee, giving fruitful suggestions to improve my dissertation, and the invaluable lectures they offered in the earlier years of my Ph.D.

I would like to thank the Department of Mathematical Sciences at the University of Cincinnati for financial support and for giving me this fantastic opportunity to pursue my graduate degree in the USA. I would like to extend my thanks to all the professors and other staff members in the department for helping me solidify my statistical foundation and making me comfortable during my tenure as a graduate student. Also, I thank Medpace Inc. for financial support for my last year and Ohio Supercomputer Center for computer time.

Moreover, I would like to thank my undergraduate advisor Dr. Kaushalya Premachandra and Mr. Ranjan Dissanayake for encouraging and supporting me in pursuing my graduate degree. I also would like to express my gratitude to all the teachers who have taught me so far.

Furthermore, I would like to express a very special thank to my family for their endless support and love, especially my dear cousin Samantha Kandearachchi for being a pillar of support in countless ways. Finally, I would like to thank fiance Yasan Fonseka for encouraging me and standing by me every step of the way throughout my tenures as an undergraduate student and a graduate student.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Multiple testing (MT) refers to the simultaneous testing of several hypotheses. The need for multiple testing arises commonly in many areas such as genomics, medicine, social science, etc., where many factors are investigated simultaneously. If one does not consider the multiplicity of tests, then the probability that some of the true null hypotheses are rejected will be larger than the assumed nominal level. For example, assume we test a single null hypothesis at significance level $\alpha = 0.05$. The maximum type I error rate (incorrect rejection of a true null) is 0.05. Now, if we have two null hypotheses and perform two independent tests, each at level $\alpha = 0.05$, the probability of rejecting at least one true null hypothesis is

$$P(\text{reject at least one true null hypothesis}) = 1 - 0.95^2 = 0.0975 (> 0.05).$$

Type I error rate is almost doubled here. In general, as the number of tests increases, the probability of incorrectly rejecting the null hypothesis is also increased (multiplicity problem). Therefore, statisticians describe the multiplicity problem as "the curse of dimensionality," and some multiplicity adjustment should always be considered to control the error rate.

In statistics, two main approaches can be used for addressing the multiplicity problem: the frequentist approach and the Bayesian approach. These two approaches adjust the multiplicity in different ways. The frequentist approach usually first gets the p-values for each test and then adjust them for multiplicity, "post-hoc adjustment." That is, in frequentist methods, multiple testing is not a part of the model. Nevertheless, in the Bayesian approach, multiplicity adjustment starts right at the beginning setting up priors for the parameters, including the proportion of true null, and then it goes through. In the Bayesian method, multiple testing is a part of the model.

1.1. Frequentist Approach to Multiplicity Adjustment

In the frequentist approach, one way of accounting for the multiplicity of individual tests is to control the familywise error rate ($FWER$) by keeping the probability of $FWER \leq \alpha$. The classical method to control the $FWER$ is the Bonferroni method, which rejects the null hypothesis(H_{0i}) if and only if the marginal p-value for testing $H_{0i} \leq \alpha/M$, where M is the number of tests. But in this procedure as M increases it is hard to reject the null hypothesis, leading to extremely low power. Hence, later in the literature researchers have suggested some other statistically more powerful methods such as controlling the false discovery rate (FDR) than controlling the FWER for adjusting multiplicity [17].

The most popular method of adjusting multiplicity in the frequentist point of view was proposed by Benjamini and Hochberg (BH) [5] for controlling FDR, which gains greater power and is less conservative than the Bonferroni correction. Therefore, many modifications of FDR and a wide range of its applications in different disciplines have been suggested in the literature [6, 22, 32, 1, 29].

1.2. Bayesian Approach to Multiplicity Adjustment

The Bayesian approach jointly uses available prior information (prior distribution) about the unknown quantity of interest and the observed data (likelihood function) to make inferences about the unknown quantity of interest using the posterior distribution. Typically, prior information may be obtained either from experts in the field, historical data, or both. The way of adjusting for multiplicity in the Bayesian approach is by reducing the posterior probabilities as the number of tests grows. Since the marginal likelihood does not change for fixed number tests, Bayesian multiplicity adjustment operates only through the prior probabilities assigned to hypotheses.

There is no need to introduce an extra penalty term for performing thousands of multiple testing; Bayesian testing has a built-in penalty (Ockham's razor effect [8]). As the number of multiple comparisons increases, the multiplicity adjustment is automatically made by reducing the posterior probabilities. This is one of the attractions of the Bayesian approach to become more prevalent in multiple testing recently. Scott and Berger [31, 4] discussed details about automatic handling of multiplicity adjustment for multiple testing with a mixture model and variable selection in linear models by choosing appropriate priors. Other than inherited adjustment for multiplicity, since the Bayesian method is a model-based approach, it allows incorporating more complex models as required by the specific application. When the data has a dependent structure, unlike in some of the commonly used frequentist methods in the Bayesian method, since the multiplicity is adjusted by assigning a reasonable prior probability, the Bayesian multiple testing methods do not depend on the dependence structure of the data. However, the ways to achieve the multiplicity adjustment and its implications in the Bayesian approach are still in development.

1.3. Organization of The Dissertation

This dissertation focuses on adapting the Bayesian approach for MT of both continuous and discrete data and is organized as follows.

Chapter 2 focuses on the Bayesian approach to multiple testing of continuous data and investigates the sensitivity to the normality assumption when the data are actually from a t-distribution. We consider multiple testing of means or location parameters when the response variables follow a t-distribution. First, we determine suitable priors for the parameters and develop a computational strategy for computing the posterior probabilities of the hypotheses. Then, use it to study the sensitivity of the results under normality assumption using a simulation study.

Chapter 3 presents the background of Bayesian and frequentist procedures for simultaneous testing discrete data and our novel formal Bayesian approach for testing the equality of two binomial proportions. We develop two priors, Local prior and Threshold (non-local) prior, for binomial testing and improve the convergence rates of Bayes factors in testing true null hypotheses with our novel Threshold prior approach.

Chapter 4 will extend our proposed method for testing equality of two binomial proportions to testing ordered binomial proportions. First, we modify the proposed Local and Threshold prior approaches in chapter 3 for testing one-sided alternatives and compare the results from two priors for convergence of Bayes factors under true null hypotheses. As expected, the Threshold approach improves convergence rates of Bayes factors in favor of true null hypotheses than the Local prior approach. Later on, we extend our formal Bayesian approach under Local prior for testing more than two order-restricted binomial proportions and apply it for simulated and real data examples.

Chapter 2

Bayesian multiple testing of means and its sensitivity to normality assumption

2.1. Introduction

Multiple testing has appeared as a critical problem in statistical inference as of its applicability in understanding extensive data involving many parameters [11]. One typical example is analyzing DNA microarray data where several thousand genes are measured simultaneously to detect the genes that are activated by a specific stimulus [11, 18, 21, 31]. Another example is genetics; researchers would like to see which genetic markers make a difference in a treatment effect among thousands of them [12]. Making decisions without adjusting for multiplicity can lead to error rates that are higher than the nominal rate [23]. Therefore, some multiplicity adjustment should always be taken to control the overall error rate.

Although various ways of performing multiple tests have been proposed in the literature over the past few decades, the Bayesian approach is the only available method that adjusts for multiplicity inherently, without involving any post hoc multiplicity adjustment. In the Bayesian approach, multiplicity adjustment starts right at the beginning setting up priors for the parameters, including the proportion of true null, and then it goes through. Therefore, multiple testing is a part of the model in the Bayesian approach.

In this chapter, we considered the Bayesian multiple testing of means when the underlying data distribution is not normal, specifically when the data is t-distributed. Our work is a theoretical development of Bayesian multiple testing, which extends the work of Scott and Berger [31] for a two-group model when the data is independent and Normally distributed.

Most of the work involving continuous random variables, assume Normal distribution assumptions in the literature. Nevertheless, in some situations, Normality assumption may be too restrictive. Therefore, one way to address this issue is to assume a different distribution like student's t-distribution. It is necessary to consider different possible distribution assumptions, as the final decision to reject or accept the null hypothesis may vary depending on our distribution assumption. For a simple example consider the situation $x = 2$ and we are testing one-sided hypothesis for population mean μ, $H_0 : \mu \leq 0\, vs. H_1 : \mu > 0$ with significance level $\alpha = 0.05$. On one hand, when $x \sim N(0, 1)$ the p-value is $0.0228 (< .05)$, which leads to reject the null hypothesis. On the other hand, when $x \sim t_5$ the p-value is $0.051 (> .05)$, which fails to reject the null hypothesis. Here, in two situations, the answer is different depending on the distribution assumption. So, it is important to develop methods for multiple testing when data is possible from different distributions rather than blindly assuming Normal assumptions for each situation.

Other than the Normal distribution, the student's t-distribution is another more commonly plausible distribution for modeling continuous random variables. For instance, even though almost everywhere in the literature Normality assumption has been using in analyzing DNA microarrays, it has been shown that 5% to 15 % of the samples deviate from Normality and are very close to t-distribution [21]. In that sense, it is probably worth looking at if the data is coming from t-distribution rather than Normal distribution. Since it is difficult to find the test statistic in the frequentist approach, t-distribution in the frequentist framework is even more difficult, especially when the sample size is small or moderate. So, our proposed work in this chapter is to implement a method to perform multiple testing when data is actually from a t-distribution using the Bayesian framework.

We first define our MT problem, discuss some criteria for choosing a prior in objective Bayesian approach, and derive the distributions for the priors and posteriors for the defined problem. Then in the next section, we discuss and compare the performance of two possible approaches for computation, the importance sampling approach and Monte Carlo Markov Chain(MCMC) approach. In the next section, we simulate data and compare the posteriors to assess the sensitivity to the assumption of the sampling and the prior distributions of non-zero means. In the last section, we present the overall conclusion of this chapter.

2.2. The Statistical Model Specifications

Assume we observe data $\boldsymbol{X} = (x_{11}, \cdots, x_{1n}; \cdots; x_{m1}, \cdots, x_{mn})$ of m different tests each with n observations, where x_{ik} arises independently from a particular distribution with unknown location μ_i and unknown scale σ^2, $x_{ik} \overset{iid}{\sim} f(location\ = \mu_i,\ scale\ = \sigma^2)$. We are interested in finding which of μ_is are differentially expressed. In multiple testing settings, the problem is to simultaneously assess whether each of μ_i is zero or nonzero and estimate the magnitudes of their effect. Based on this information, we define the multiple testing problem as below.

$$H_{0i} : \mu_i = 0 \ \text{ versus } \ H_{1i} : \mu_i \neq 0 \ \ i = 1, 2, \cdots, m.$$

Most of the literature for multiple testing with continuous response variables, conventionally, Normal distribution has been used for modeling the data. So each observation x_{ik} assumes coming from a Normal distribution with an unknown mean μ_i and variance σ^2, $x_{ik} \overset{iid}{\sim} N(\mu_i, \sigma^2)$.

However, since there are some situations where the Normality assumption might be too restrictive and data more resemble the t-distribution, we here assess the sensitivity to the choice of Normal sampling distributions when data is actually from a t distribution. Therefore, in our work, we assume both Normal and t sampling distributions and compare the results, $x_{ik} \overset{iid}{\sim} N(\mu_i, \sigma^2)$ or $x_{ik} \overset{iid}{\sim} t_v(\mu_i, \sigma^2)$. Here, while μ_i and σ^2 are assuming unknown and v is assuming either known or unknown.

Since the Bayesian approach jointly uses available prior information of the unknown parameters and the observed data to make inferences about the unknown quantity of interest, selecting suitable prior probabilities is very important. Next, we focus on some critical conditions for a prior to be satisfied in Bayesian hypothesis testing.

With the Normal sampling model, nonzero μ_is are also modeled as arising from a Normal density with mean zero, and unknown variance τ^2, $N(\mu_i|0, \tau^2)$ [31]. Nevertheless, we are interested in not only the Normal sampling model but also the t sampling model. So the question is, when assuming t sampling distribution, whether we can still use a Normal prior as in the Normal sampling model or should we use a t prior to model nonzero μ_is? Therefore, about choosing a prior for nonzero μ_is, $\pi(\mu_i|H_{1i})$, in section 2.2.1, we talk about some facts that may or may not be well known, called "information consistency."

2.2.1. Information Consistency

Many criteria have been proposed in objective Bayesian analysis to define objective prior distributions, and no single dominant criterion has emerged. Among those various conditions said for a prior to satisfying in the objective Bayesian approach, information consistency is one condition that Bayarri et al. [16] proposed in 2012. The information consistency criteria was originally proposed for objective Bayesian model selection as defining below.

Let $\boldsymbol{y}$ be a data vector of size n from one of the models, $M_0 : f_0(\boldsymbol{y}|\boldsymbol{\alpha})$ and $M_i : f_i(\boldsymbol{y}|\boldsymbol{\alpha}, \boldsymbol{\beta_i})$ for $i = 1, 2, \cdots, N-1$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta_i}$ are unknown model parameters. For any model M_i, if $\{\boldsymbol{y}_m, m = 1, 2, \cdots\}$ is a sequence of data vectors of fixed size such that, as $m \to \infty$,

$$\Lambda_{i0}(y_m) = \frac{sup_{\alpha,\beta} f_i(y_m|\alpha, \beta_i)}{sup_{\alpha} f_0(y_m|\alpha)} \to \infty \quad \text{then} \quad B_{i0}(y_m) \to \infty$$

That is, priors must be chosen to satisfy the condition; as the evidence favoring M_i increases, B_{i0} also increases. Therefore, we adopt the information consistency criteria to choice priors in terms of hypothesis testing context.

Information (In-)Consistency in Hypothesis Testing

Definition 2.2.1. *(Information consistency in single testing)*

Assume we observe data $\boldsymbol{x} = (x_1, ..., x_n) \overset{iid}{\sim} N(\mu, \sigma^2)$. Consider testing, $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$. Then, information consistency in terms of testing is Bayes factor, $B_{10}(\boldsymbol{x}) \to \infty$ as $\boldsymbol{x} \to \infty$. As a result,

$$P(H_o|\boldsymbol{x}) = \left[1 + \frac{(1-p)}{p} B_{10}\right]^{-1} \to 0 \text{ as } \boldsymbol{x} \to \infty.$$

Proposition 2.2.2. *(Normal sampling model with Normal prior)*

(a). Let $\boldsymbol{x} = (x_1, .., x_n) \overset{iid}{\sim} N(\mu, \sigma^2)$ with known σ. Assume a normal prior $\pi(\mu|H_1) = N(0, \tau^2)$, with a known τ. Then,

$$P(H_0|\boldsymbol{x}) \to 0 \ \textit{as} \ \bar{x} \to \infty$$

(b). Let $\boldsymbol{x} = (x_1, .., x_n) \overset{iid}{\sim} N(\mu, \sigma^2)$ where σ is unknown and assigned the prior $\pi(\sigma^2) = 1/\sigma^2$. Assume a normal prior $\pi(\mu|H_1) = N(0, \tau^2)$, with a known τ. Then,

$$P(H_0|\boldsymbol{x}) \not\to 0 \ \textit{as} \ \bar{x} \to \infty \ \ (\textit{with } s^2 \textit{ fixed})$$

Remark: Replacing $\pi(\sigma^2) = 1/\sigma^2$ by $\pi(\sigma^2) = IG(\alpha_0/2, \beta_0/2)$, has same limiting result(inconsistency).

Proposition 2.2.3. *(t sampling model with normal prior)*

Let $\boldsymbol{x} = \{x_1, ..., x_n\}$ with x_i's iid $t_\upsilon(\mu, \sigma^2.)$, with known σ. Assume a prior $\pi(\mu|H_1, \tau^2) : N(0, \tau^2)$.

(a). Assume τ is known. Then,

$$P(H_0|\boldsymbol{x}) \not\to 0 \ \ \textit{as} \ x \to \infty$$

(b). Assume τ is unknown and assign a prior $\pi(\tau^2)$. If the prior $\pi(\tau^2)$ has finite moments up to order $n(\upsilon + 1)$, then

$$P(H_0|x) \not\to 0; \ \ \textit{as} \ x \to \infty.$$

Note: The prior $\pi(\tau^2) = \sigma^2/(\sigma^2 + \tau^2)^2$ does not satisfy the above condition and hence may have information consistency.

Proposition 2.2.4. *(t sampling model with t prior)*

(a). Let $x \sim t_\upsilon(\mu, \sigma^2)$ with known σ. Assume a prior $\pi(\mu|H_1, \tau^2) : t_\nu(0, \tau^2)$ with fixed and known ν. Assume τ known. Then,

$$P(H_0|\boldsymbol{x}) \to 0 \ \ \textit{as} \ x \to \infty \ \textit{if} \ \upsilon > \nu$$

Remark: Same result holds if τ is assigned a proper prior.

(b). Let $\boldsymbol{x} = (x_1, .., x_n) \overset{iid}{\sim} t_\upsilon(\mu, \sigma^2)$ *where* σ *is known and* $\pi(\mu|H_1) = t_\nu(0, \tau^2)$ *with* $\pi(\tau^2)$ *proper. Then,*

$$P(H_0|\boldsymbol{x}) \to 0 \quad \text{as } \boldsymbol{x} \to \infty \text{ if } n\upsilon > \nu \text{ (with } s^2 \text{ fixed)}$$

Remark: We conjecture that the result in 2.2.4(b) is true when σ *is unknown and assign a prior.*

Proofs in Appendix A.1.

2.2.2. Prior Distributions

According to Westfall et al. [34], in the Bayesian setting, choosing the prior probabilities of hypotheses in multiple testing has a significant effect on the posterior probabilities. Using a common prior probability, $p = P(\mu_i = 0)$, for all the null hypotheses is natural in Bayesian MT except when a known covariate that could influence the prior probability of H_{0i} is involved. In other words, if the hypotheses are exchangeable [28], then there is a common prior probability, p, for all the null hypotheses. In our work, we assume the hypotheses are exchangeable, and there is a common prior probability p, which is drawn from $Beta(1, 1)$ for each $\mu_i = 0$, $\pi(p) \sim Beta(1, 1)$. Note that when $\alpha = \beta = 1$, Beta prior reduces to Uniform prior. Note that if the hypotheses are not exchangeable, much more complicated prior probabilities should be considered [12].

To complete the model specification, next, we define priors on unknown parameters, μ_i, and σ^2. Let prior for non-zero μ_is is given by the distribution function f_1,

$$\mu_i|H_{1i} \sim f_1(0, \tau^2)\,], \text{ where } \tau \text{ is unknown.}$$

Then, for either case sampling distribution, f is Normal, or t, the prior combination $f_1 = N(0, \tau^2)$ and $\pi(\tau^2, \sigma^2) = (\sigma^2 + \tau^2)^{-2}$ satisfy the findings in section 2.2.1, like information consistency and posterior existence. So that, we consider $N(\mu_i|0, \tau^2)$ as one suitable prior for nonzero μ_is in our model. Other than Normal prior, in the literature, t prior has been used to model location parameters of both Normal [19] and t sampling distributions [3]. Therefore we use t prior, $f_1 = t_\nu(0, \tau)$, for nonzero μ_is as an alternative and assess the sensitivity of f when the prior distribution, f_1 is Normal, and t separately using Bayesian hierarchical model. So we are considering following four model-prior combinations.

Case I: Normal sampling model and Normal prior (NN)

$$x_{ij} \sim N(\mu_i, \sigma^2) \text{ and } \mu_i|H_{1i} \sim N(0, \tau^2) \quad \text{i.e. } f \equiv N \text{ and } f_1 \equiv N$$

Case II: t sampling model (true model) and Normal prior (TN)

$$x_{ij} \sim t_\upsilon(\mu_i, \sigma^2) \text{ and } \mu_i|H_{1i} \sim N(0, \tau^2) \quad \text{i.e. } f \equiv T \text{ and } f_1 \equiv N$$

Case III: Normal sampling model and t prior (NT)

$$x_{ij} \sim N(\mu_i, \sigma^2) \text{ and } \mu_i|H_{1i} \sim t_\nu(0, \tau^2) \quad \text{i.e. } f \equiv N \text{ and } f_1 \equiv T$$

Case IV: t sampling model and t prior (TT)

$$x_{ij} \sim t_\upsilon(\mu_i, \sigma^2) \text{ and } \mu_i|H_{1i} \sim t_\nu(0, \tau^2) \quad \text{i.e. } f \equiv T \text{ and } f_1 \equiv T$$

As mentioned above, for unknown σ^2 and τ^2, we use the joint prior,

$$\pi(\tau^2, \sigma^2) = (\sigma^2 + \tau^2)^{-2},$$

which Scott and Berger [31] used to model unknown σ^2 and τ^2. Scott and Berger [31] have suggested this as a possible joint prior in the absence of vital information about σ^2 and τ^2. Although this joint prior for σ^2 and τ^2 is improper, the conditional prior for $\tau^2|\sigma^2$ is proper. Also, this prior was motivated by Berger and Strawderman [10] for admissibility considerations in hierarchical models.

2.2.3. The Model

The model is specified by defining an index parameter γ_i, which can take values either 0 or 1 depending on which hypothesis is true.

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

When $\mu_i = 0$, the corresponding observation is classified as a "noise" and otherwise as a "signal." Then, the original multiple testing problem of each μ_i equals zero or not can be rewritten as

$$H_{0i} : \gamma_i = 0 \text{ versus } H_{1i} : \gamma_i \neq 0, \quad \text{for } i = 1, 2, \cdots, m.$$

Given f is N or t_υ, the marginal distributions of x_{ij} has the form

$$x_{ij} \sim p \cdot f(0, \sigma^2) + (1 - p) \cdot f(\mu_i, \sigma^2).$$

Then the full likelihood functions under N and t_υ can be written as

$$f(\boldsymbol{X}|\boldsymbol{\mu}, \boldsymbol{\gamma}, \sigma^2) = \prod_{i=1}^{m} \prod_{k=1}^{n} \left[\frac{1}{\sqrt{2\pi\sigma^2}} exp\left(\frac{-1}{2\sigma^2} (x_{ik} - \gamma_i \mu_i)^2 \right) \right]$$

$$f(\boldsymbol{X}|\boldsymbol{\mu}, \boldsymbol{\gamma}, \sigma^2) = \prod_{i=1}^{m} \prod_{k=1}^{n} \frac{\Gamma[(\upsilon + 1)/2]}{\Gamma(\upsilon/2)\sqrt{\upsilon\pi}\sigma} \cdot \left[1 + \frac{1}{\upsilon} \frac{(x_{ik} - \gamma_i \mu_i)^2}{\sigma^2} \right]^{-(\upsilon+1)/2} \tag{2.2.1}$$

Under the above modeling assumptions, the posterior distributions of $p, \boldsymbol{\mu}, \boldsymbol{\gamma}, \sigma^2$, and τ^2 has the form

$$\pi(\Theta|\boldsymbol{X}) = C^{-1} \cdot \prod_{i=1}^{m} \left[\prod_{k=1}^{n} f(x_{ik}|\mu_i, \gamma_i, \sigma^2) \cdot \pi(\mu_i|0, \tau^2) p^{1-\gamma_i} (1 - p)^{\gamma_i} \right] \cdot \pi(\sigma^2, \tau^2) \pi(p) \tag{2.2.2}$$

where $\Theta \equiv (p, \boldsymbol{\mu}, \boldsymbol{\gamma}, \sigma^2, \tau^2)$ and f is either Normal or t likelihood function define in equation (2.2.1) and C is the normalization constant.

For case-I (NN), the posterior distribution 2.2.2 has a more straightforward form due to the conjugacy of the sampling, and prior distributions [31]. However, when the sampling and/or prior distributions are from a t distribution (case II to case IV), such simplification no longer holds, presenting a challenge in computing p_i. For case I to case IV, proof of posterior existence is given in section 2.2.1.

2.2.4. Posterior Existence

Having specified the sampling distribution and priors, the priors are not all proper prior. The joint prior, $\pi(\sigma^2, \tau^2) = (\sigma^2 + \tau^2)^{-2}$, is improper. Therefore, posterior begin proper is important to consider.

Lemma 2.2.5. *The posterior distribution in (5) is proper (i.e., C is finite)*

Proof of Lemma 2.2.5:

Case I: Normal sampling model and Normal prior (NN)

This proof has given in Scott and Berger [31].

Case II: t sampling model (true model) and Normal prior (TN)

Case III: Normal sampling model and t prior (NT)

Proofs of cases II and III are similar to case IV.

Case IV: t sampling model and t prior (TT)

We below provide the proof of the posterior existence of the TT model for a particular case of $m = 2$ to keep the proof simple and easy to follow. Nevertheless, this proof can be more general and extended to any number of tests, $m > 2$. Also, this proof can easily be modified and used to show the posterior existence of the other two models TN and NT.

Consider we observe two data values, $x_i \sim t_\upsilon(\mu_i, \sigma^2)$ for $i = 1, 2$ independently with means $\mu_1 = 0$ and $\mu_2 \neq 0$, respectively. For each i, we are going to set two hypotheses, $H_{0i} : \mu_i = 0$ versus $H_{1i} : \mu_i \neq 0$. Now we assume that $\mu_2 \sim t_\nu(0, \tau^2)$ and $\pi(\sigma^2, \tau^2) = (\sigma^2 + \tau^2)^{-2}$.

By introducing three new latent variables w_i (for $i = 1, 2$) and w, we can rewrite the above t variables, x_i and μ_2, in terms of Normal variables as, $x_i|w_i \sim N(\mu_i, \sigma^2 w_i)$ and $\mu_2|w \sim N(0, \tau^2 w)$ where; $w_i \sim IG(\upsilon/2, \upsilon/2)$, and $w \sim IG(\nu/2, \nu/2)$. Then, the posterior distribution has the form

$$\pi(\boldsymbol{\mu}, \sigma^2, \tau^2, \boldsymbol{w}, w|\boldsymbol{X}) \propto \frac{1}{\sqrt{2\pi\sigma^2 w_1}} e^{-\dfrac{x_1^2}{2\sigma^2 w_1}} \frac{1}{\sqrt{2\pi\sigma^2 w_2}} e^{-\dfrac{(x_2-\mu_2)^2}{2\sigma^2 w_2}} \cdot \frac{1}{\sqrt{2\pi\tau^2 w}} e^{-\dfrac{\mu_2^2}{2\tau^2 w}} \frac{1}{(\sigma^2+\tau^2)^2} \pi(\boldsymbol{w})\pi(w) \tag{2.2.3}$$

where, $\boldsymbol{\mu} = (\mu_1, \mu_2)$, $\boldsymbol{w} = (w_1, w_2)$, $\boldsymbol{X} = (x_1, x_2)$, $\pi(\boldsymbol{w}) = IG(\upsilon/2, \upsilon/2)$, and $\pi(w) = IG(\nu/2, \nu/2)$.

Now, let $\pi(\boldsymbol{w})\pi(w) = f(w_1, w_2, w)$, $A_1 = \dfrac{1}{\sqrt{2\pi\sigma^2 w_2}} \dfrac{1}{\sqrt{2\pi\tau^2 w}}$, $A_2 = \dfrac{1}{\sqrt{2\pi\sigma^2 w_1}} e^{-\dfrac{x_1^2}{2\sigma^2 w_1}}$, and $A_3 = \dfrac{1}{(\sigma^2+\tau^2)^2} f(w_1, w_2, w)$ and find the marginal posterior of $\Theta_1 \equiv (\sigma^2, \tau^2, w_1, w_2, w)$ by integrating above formula 2.2.3 over μ_2.

$$\begin{aligned}\pi(\Theta_1|\boldsymbol{X}) &\propto \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2 w_2}} e^{-\dfrac{(x_2-\mu_2)^2}{2\sigma^2 w_2}} \cdot \frac{1}{\sqrt{2\pi\tau^2 w}} e^{-\dfrac{\mu_2^2}{2\tau^2 w}} d\mu_2 \\ &\quad \cdot \frac{1}{\sqrt{2\pi\sigma^2 w_1}} e^{-\dfrac{x_1^2}{2\sigma^2 w_1}} \frac{1}{(\sigma^2+\tau^2)^2} f(w_1, w_2, w) \\ &\propto \int_{-\infty}^{\infty} A \cdot exp\Big\{ -\frac{x_2^2\tau^2 w - 2x_2\mu_2\tau^2 w + \mu_2^2\tau^2 w + \mu_2^2\sigma^2 w_2}{2\sigma^2\tau^2 w_2 w} \Big\} d\mu_2 \cdot A_2 \cdot A_3\end{aligned}$$

$$\propto \int_{-\infty}^{\infty} exp\left\{ -\frac{(\sigma^2 w_2 + \tau^2 w)}{2\sigma^2 w_2 \tau^2 w}\left[\left(\mu_2 - \frac{x_2 \tau^2 w}{(\sigma^2 w_2 + \tau^2 w)}\right)^2 + \frac{\tau^2 w x_2^2}{(\sigma^2 w_2 + \tau^2 w)} - \left(\frac{x_2^2 \tau^2 w}{(\sigma^2 w_2 + \tau^2 w)}\right)^2\right]\right\} d\mu_2 \cdot A_1 \cdot A_2 \cdot A_3$$

$$\propto \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi \frac{\sigma^2 w_2 \tau^2 w}{(\sigma^2 w_2 + \tau^2 w)}}} exp\left\{ -\frac{\left(\mu_2 - \frac{x_2 \tau^2 w}{(\sigma^2 w_2 + \tau^2 w)}\right)^2}{2\pi \sigma^2 w_2 \tau^2 w/(\sigma^2 w_2 + \tau^2 w)}\right\} d\mu_2$$

$$\cdot \sqrt{2\pi \frac{\sigma^2 w_2 \tau^2 w}{(\sigma^2 w_2 + \tau^2 w)}} exp\left\{ -\frac{\tau^2 w x_2^2}{2\sigma^2 w_2 \tau^2 w} + \frac{x_2^2 \tau^4 w^2}{2(\sigma^2 w_2 + \tau^2 w)\sigma^2 w_2 \tau^2 w}\right\}$$

$$\cdot A_1 \cdot A_2 \cdot A_3$$

$$\propto \frac{1}{\sqrt{2\pi(\sigma^2 w_2 + \tau^2 w)}} exp\left\{ \frac{-\tau^2 w x_2^2 (\sigma^2 w_2 + \tau^2 w) + x_2^2 \tau^4 w^2}{2\sigma^2 w_2 \tau^2 w (\sigma^2 w_2 + \tau^2 w)}\right\} \cdot A_2 \cdot A_3$$

$$\propto \frac{1}{\sqrt{2\pi\sigma^2 w_1}} \frac{1}{\sqrt{2\pi(\sigma^2 w_2 + \tau^2 w)}} e^{-\frac{x_1^2}{2\sigma^2 w_1}} e^{\frac{-x_2^2}{2(\sigma^2 w_2 + \tau^2 w)}} \cdot \frac{1}{(\sigma^2 + \tau^2)^2} f(w_1, w_2, w)$$

Therefore the marginal posterior of Θ_1 has the form,

$$\pi(\Theta_1|\boldsymbol{X}) \propto \frac{1}{\sqrt{\sigma^2 w_1 (\sigma^2 w_2 + \tau^2 w)}} e^{-\frac{x_1^2}{2\sigma^2 w_1}} e^{\frac{-x_2^2}{2(\sigma^2 w_2 + \tau^2 w)}} \frac{1}{(\sigma^2 + \tau^2)^2} f(w_1, w_2, w) \tag{2.2.4}$$

Then, by proving the marginal posterior, $\pi(\Theta_1|\boldsymbol{X})$ is finite, we can show that the posterior, $\pi(\boldsymbol{\mu}, \sigma^2, \tau^2, \boldsymbol{w}, w|\boldsymbol{X})$, does exist. In the expression, 2.2.4 for the marginal posterior of $\Theta_1 \equiv (\sigma^2, \tau^2, w_1, w_2, w)$, since two exponential components are always finite and bounded by $(0, 1]$, in order to show 2.2.3 is finite, we only need to consider the integration of the rest of expression 2.2.4 over Θ_1 and show it is finite.

Let $\Theta_1 \in [(0,\delta),\cdots,(0,\delta)]^C$ and first we want to show that $\int_{([0,\delta]^5)^C} \pi(\Theta_1) < \infty$. Since $e^{-x} \in (0,1)$for all $x > 0$ consider,

$$
\begin{aligned}
&\int_{([0,\delta]^5)^C} \frac{1}{\sqrt{\sigma^2 w_1(\sigma^2 w_2 + \tau^2 w)}} \frac{1}{(\sigma^2+\tau^2)^2} f(w_1,w_2,w)\, d\sigma^2\, d\tau^2\, dw\, dw_1\, dw_2 \\
&< \int_{([0,\delta]^5)^C} \frac{1}{\sqrt{\sigma^2 \delta(\delta^2 + \delta^2)}} \frac{1}{(\sigma^2+\tau^2)^2} f(w_1,w_2,w)\, d\sigma^2\, d\tau^2\, dw\, dw_1\, dw_2 \\
&= \int_{([0,\delta]^5)^C} \frac{1}{(\sigma^2)^{1/2}} \frac{1}{(2\delta^3)^{1/2}} \frac{1}{(\sigma^2+\tau^2)^2} f(w_1,w_2,w)\, d\sigma^2\, d\tau^2\, dw\, dw_1\, dw_2 \\
&< \int_w \int_{w_1} \int_{w_2} \left\{ \int_\delta^\infty \int_0^\infty \frac{1}{(\sigma^2)^{1/2}} \frac{1}{(2\delta^3)^{1/2}} \frac{1}{(\sigma^2+\tau^2)^2}\, d\tau^2\, d\sigma^2 \right\} f(w_1,w_2,w)\, dw\, dw_1\, dw_2 \\
&= \int_w \int_{w_1} \int_{w_2} \left\{ \int_\delta^\infty \frac{1}{(\sigma^2)^{1/2}} \frac{1}{(2\delta^3)^{1/2}} \frac{1}{(\sigma^2)}\, d\sigma^2 \right\} f(w_1,w_2,w)\, dw\, dw_1\, dw_2 \\
&= \int_w \int_{w_1} \int_{w_2} \frac{1}{(2\delta^3)^{1/2}} \left\{ \int_\delta^\infty \frac{1}{(\sigma^2)^{3/2}}\, d\sigma^2 \right\} f(w_1,w_2,w)\, dw\, dw_1\, dw_2 \\
&= \int_w \int_{w_1} \int_{w_2} \frac{1}{(2\delta^3)^{1/2}} \frac{2}{(\delta)^{1/2}} f(w_1,w_2,w)\, dw\, dw_1\, dw_2 \\
&= \frac{\sqrt{2}}{\delta^2} \int \pi(w) dw \int \pi(w_1) dw_1 \int \pi(w_2) dw_2 \\
&= \frac{\sqrt{2}}{\delta^2} < \infty
\end{aligned}
$$

Now we need to show that the integrand I is bounded in $\Theta_1 \in (0,\delta)^5$, i.e., given $\xi > 0$, there is a finite value δ s.t. $I = \frac{e^{-x_1^2/2\sigma^2 w_1}}{\sqrt{\sigma^2 w_1}} \frac{e^{-x_2^2/2(\sigma^2 w_2+\tau^2 w)}}{\sqrt{\sigma^2 w_2 + \tau^2 w}} \frac{1}{(\sigma^2+\tau^2)^2} < \xi$ for $|\Theta_1| < \delta$.

Consider,

$$
I < \frac{e^{-x_1^2/2\sigma^2 w_1}}{(\sigma^2)^{2.5}\sqrt{w_1}} \frac{e^{-x_2^2/2(\sigma^2 w_2+\tau^2 w)}}{\sqrt{\sigma^2 w_2 + \tau^2 w}} \quad = \frac{e^{-x_1^2/2\sigma^2 w_1}}{(\sigma^2 w_1)^{2.5}} \cdot (w_1)^2 \cdot \frac{e^{-x_2^2/2(\sigma^2 w_2+\tau^2 w)}}{\sqrt{\sigma^2 w_2 + \tau^2 w}}
$$

We know that for some $k > 0$,

$$\frac{e^{-c/y}}{y^k} \;\rightarrow\; 0 \text{ as } y \;\rightarrow\; 0 \text{ and } \left|\frac{e^{-c/y}}{y^k}\right| < \xi \text{ when } |y| < \delta.$$

Therefore, $\dfrac{e^{-x_1^2/2\sigma^2 w_1}}{(\sigma^2 w_1)^{2.5}} < \xi^{1/3}$ where $|\sigma^2 w_1| < \delta'$,

$(w_1)^2 < \xi^{1/3}$ where $|w_1| < \delta'$, and

$\dfrac{e^{-x_2^2/2(\sigma^2 w_2 + \tau^2 w)}}{\sqrt{\sigma^2 w_2 + \tau^2 w}} < \xi^{1/3}$ where $|\sigma^2 w_2 + \tau^2 w| < \delta'$; here $\delta' < 2\delta^2$.

Then given $\xi > 0$, $I < \xi$ when $|\Theta_1| < \delta$. So that,

$$\int\cdots\int_{[0,\delta]^5} \frac{1}{\sqrt{\sigma^2 w_1(\sigma^2 w_2 + \tau^2 w)}} e^{-\dfrac{x_1^2}{2\sigma^2 w_1}} e^{\dfrac{-x_2^2}{2(\sigma^2 w_2 + \tau^2 w)}} \frac{1}{(\sigma^2 + \tau^2)^2} f(w_1, w_2, w)\, d\Theta_1 < \xi$$

i.e., the posterior distribution is bounded.

2.3. Computations

In Bayesian inference, we are interested in computing the full posterior joint distribution of data over a set of random variables to find various summaries. Often posterior distribution of parameters of interest is not available in closed-form, and even if it is, it is not possible to do closed-form evaluation of required integrals. In such cases, we may proceed with sampling algorithms based on Monte Carlo Markov Chain (MCMC) techniques.

As the magnitude of m gets large, the posterior computation's complexity becomes more challenging due to the increase in the number of integrals. In this kind of large m situation, importance sampling is an efficient tool for computing posterior and its functions. Scott and Berger [31] have shown that the posterior computations are straightforward using an importance sampling scheme for NN model. Table 2.1 reports posterior probabilities of alternative hypotheses for ten signals using importance sampling and MCMC approaches under the NN model with forty tests.

$\bar{x}_i$		-3.11	-1.7	-1.23	-0.61	0.44	0.89	1.03	1.41	2.54	3.06
$P(H_{1i}\|X)$	**Imp**	1	0.99	0.78	0.24	0.18	0.44	0.59	0.91	1	1
	mcmc	1	0.99	0.75	0.21	0.16	0.40	0.54	0.90	1	1

Table 2.1.: Comparison of the posterior probabilities of H_{1i} for 10 signals using importance sampling (Imp) and mcmc approaches. Here each of 10 signals: $x_{ik} \sim N(\mu_i, 1)$, 30 noises: $x_{ik} \sim N(0, 1)$, and $\mu_i \sim N(0, \tau^2)$. Importance functions: $p \sim beta(9.80, 2.84)$ and $ln(\tau^2) = \zeta, \zeta \sim t_3(1.03, 0.31^2)$

Although importance sampling and MCMC approaches give very close results, the importance sampling approach is more time-consuming, even on this small-scale data set. However, it does not seem easy to implement an importance sampling approach to get a good accuracy for models which involve t-distribution. For instance, in TN model, when calculating p_i using the equation 2.3.1, there is a triple integral outside, and there is a marginal distribution with multiple integrals inside. So that makes it difficult, and we found that the importance sampling approach is inefficient. Since the MCMC approach seems to be working all right in the scope of the problem we try here, we consider the MCMC approach as a better alternative and implement for all four models in our simulations.

Importance Sampling Approach

Before implementing importance sampling approach, we use the transformations $\eta = ln(\sigma^2)$ and $\zeta = ln(\tau^2)$ to make the things more convenient. Then, as importance functions we use, $t_v(\mu_j : \bar{x}_i, s_i)$, $t_v(\eta_j : lns_i^2, s_i)$, and $t_v(\zeta_j : 0, \acute{s})$ where, $\bar{x}_i = \sum_{k=1}^{n} x_{ik}$, $s_i^2 = \sum_{k=1}^{n}(x_{ik} - \bar{x}_i)^2/(n-1)$, and $\acute{s}$ is known.

MCMC techniques are used to obtain samples from the desired posterior distribution of the parameters. Specifically, we use Gibbs sampling (GS) and Metropolis-Hashing (MH), well-known MCMC sampling methods for attaining such samples. To apply GS and MH requires deriving full-conditional distributions for each parameter involved in the posterior distribution. Then, to apply MH for specific parameters, we have to select appropriate proposal distributions for those parameters. The proposal distributions are,

$$\mu_{j+1}|\mu_j \sim cauchy\big(x_{median}, max\big(\frac{x_{range}}{2}, \tau\big)\big)$$
$$\sigma^2_{j+1}|\sigma^2_j \sim t_v\Big(\sigma^2_j, \Big(\frac{x_{range}}{2}\Big)^2\Big)$$
$$\tau_{j+1}|\tau_j \sim lognormal\big(\acute{\mu}, \acute{\sigma^2}\big)$$

where x_{median} and x_{range} are the median and range of observed data and $\acute{\mu}$ and $\acute{\sigma^2}$ are assumed to be know.

Result 2.3.1. $P(H_{0i}|\boldsymbol{X})$ from importance sampling approach for TN model (proof in Appendix A.2)

$$p_i = \frac{\int\int\int p\prod_k t_v(x_i|\mu_i = 0, \sigma^2)\prod_{j\neq i}\Big\{p\prod_k t_v(x_j|\mu_j, \sigma^2) + (1-p)m(x_j|\sigma^2, \tau^2)\Big\}\pi(p)\pi(\sigma^2, \tau^2)dpd\sigma^2 d\tau^2}{\int\int\int\prod_j\Big\{p\prod_k t_v(x_j|\mu_j, \sigma^2) + (1-p)\, m(x_j|\sigma^2, \tau^2)\Big\}\pi(p)\,\pi(\sigma^2, \tau^2)\, dp\, d\sigma^2 d\tau^2} \tag{2.3.1}$$

MCMC Approach

In order to perform the MCMC approach, we start by assigning initial values for unknown parameters μ_i, σ^2, τ^2, and p. We need to calculate the fully conditional posterior of null being true given by equation 2.3.2 for each test i. Now for each i, generate γ_i from

Bernoulli distribution with success probability, $P(\gamma_i = 0|\boldsymbol{X}, \boldsymbol{\mu}, \sigma^2, p)$, and update the parameter value of μ_i. Then update the rest of the unknown parameter values σ^2, τ^2, p and repeat the process. Here, to address the issue of varying dimensions in MCMC when H_{0i} is true, we used latent $\mu_i's$ drawn from their prior distribution, $\pi(\mu_i : 0, \tau^2)$, with an updated variance parameter.

$$P(\gamma_i = 0|\boldsymbol{X}, \boldsymbol{\mu}, \sigma^2, p) = \frac{\prod_{k=1}^{n} f(x_{ik}|\mu_i = 0, \sigma^2) \cdot p}{\prod_{k=1}^{n} f(x_{ik}|\mu_i = 0, \sigma^2) \cdot p + \prod_{k=1}^{n} f(x_{ik}|\mu_i, \sigma^2) \cdot (1-p)} \tag{2.3.2}$$

2.4. Simulations and Results

The above ideas are illustrated on simulated samples using R and WINBUGS software. The idea is to see if the data is from t distribution and we use a Normal distribution, how that affects the posterior probability of each hypothesis. As mentioned in section 1.2, nonzero μ_is are modeled with two prior distributions under each sampling distribution assumption, $N(\mu_i|0, \tau)$ and $t_\nu(\mu_i|0, \tau)$.

Data is generated from $t_\upsilon(\mu_i, 1)$ under two different degrees of freedom, $\upsilon = 3$ or 7. Each test either has a single observation ($n = 1$) or multiple observations ($n = 5$). The number of tests (m) is either 40 or 110. Here the number of the test consists of 10 signals, and the rest are noises. For noise, set mean, μ_i equals zero. For signal, generate mean μ_i from $t_3(0, 2)$. Then fit NN and TN models for these data.

2.4.1. Sensitivity When Using Normal Prior

We separate the sensitivity analysis into two parts assuming:

Part 1- data is coming from a t distribution with unknown degrees of freedom

Part 2- data is coming from a t distribution with known degrees of freedom

Part I - Data come from a t-dist with unknown df

When the df is unknown, we must assign a suitable prior for df parameter υ and use the uniform prior, $Uni(3, \cdots, 15, 50)$. Using this specific range is that we wanted to allow lower df and larger df to account for Normality. We tried some other range of values close to this range, but all those priors give similar posterior probabilities for $\mu_i \neq 0$. Table 2.2 shows how well the df has been estimated.

sample size	**df 3**		**df 7**	
	30 noise	**100 noise**	**30 noise**	**100 noise**
	Est/Std	**Est/Std**	**Est/Std**	**Est/Std**
1	10/10.0	8/4.1	12/11.4	10/3.2
5	3/0.5	3/0.5	10/3.1	10/2.8

Table 2.2.: Estimates and std of the estimates of υ for different data sets. Number of signals 10 and number of noises 30 and 100. Data generated from two df: 3 and 7, and sample size 1 and 5.

When the sample size is one, df is not estimated well; when the sample size is five, df is estimated quite well. However, estimating df is not very easy, and also, it is not our goal here. Our goal is to see how well the posterior probability is estimated and how sensitive it is to the choice of Normal versus t sampling distributions. Even if df is not estimated well precisely here, the sensitivity of the posterior probability of μ_i equals zero to Normal, or t sampling distributions is may or may not be affected by how well the df is estimated, we are not going to address this issue in the scope of our work.

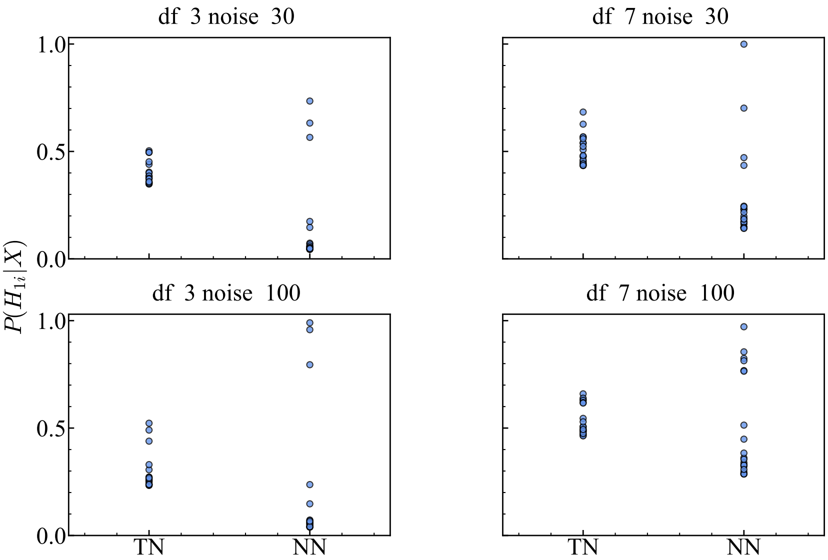


Figure 2.1.: The posterior probability of H_{1i} of some selected observations (max 5, middle 10, and min 5). Number of signals 10 and number of noises 30 and 100. Sample size is 1 and υ is unknown.

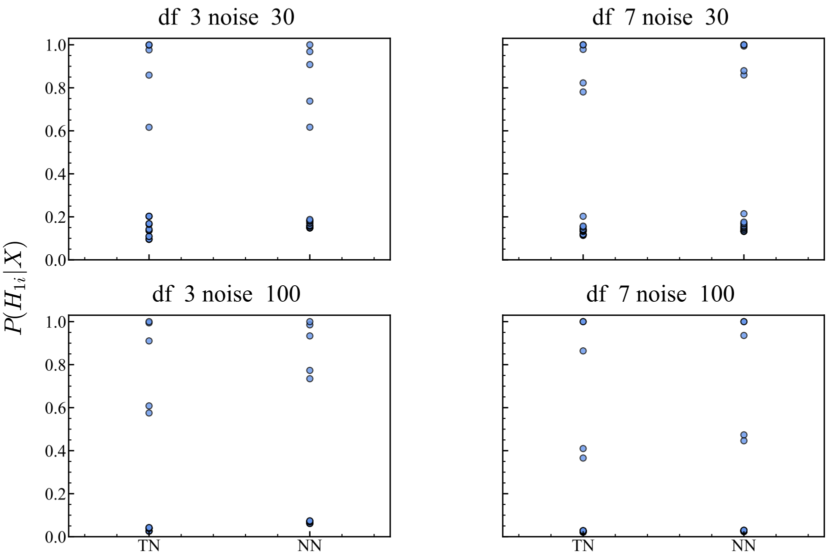


Figure 2.2.: The posterior probability of H_{1i} of some selected observations (max 5, middle 10, and min 5). Number of signals 10 and number of noises 30 and 100. Sample size is 5 and υ is unknown.

Part II - Data come from a t-dist with known df

Now we assume df is known. Because according to the Table 2.2 results, depending on sample size, df is estimated sometimes well and sometimes not. Therefore we want to see whether the sensitivity to the sampling model showed by TN and NN models may be due to unknown df; if df is known, does the sensitivity remains the same as before.

Figures 2.3 and 2.4 show posterior probabilities of alternative hypotheses of some selected observations (minimum five maximum five and middle ten observations) under TN and NN models at different simulation settings. Results from the known df assumption are the same as unknown df assumptions results. With sample size one, there is sensitivity to the sampling model, and posterior probabilities tend to be tight under the t distribution assumption than the Normal distribution assumption. With sample size five, sensitivity to the sampling model is disappearing.

Figures 2.5 and 2.6 represent another way to look at sensitivity to the sampling distribution. Figure 2.5 with sample size one and figure 2.6 with sample size five. We get these two figures by increasing the value of one particular test sample and keeping other samples fixed. The purpose of these two figures is also to compare Normal and t sampling models by checking how the posterior probability of one selected observation changes as its value(s) increase(s) in each model TN and NN. These two figures give somewhat similar results to figures 2.3 and 2.4. If we are trying to study the sensitivity figures 2.3 and 2.4 are more useful than figures 2.5 and 2.6. However, in figures 2.5 and 2.6, there is a difference between TN and NN models when the sample size is one, and if we increase it to five, that difference tends to disappear somewhat.

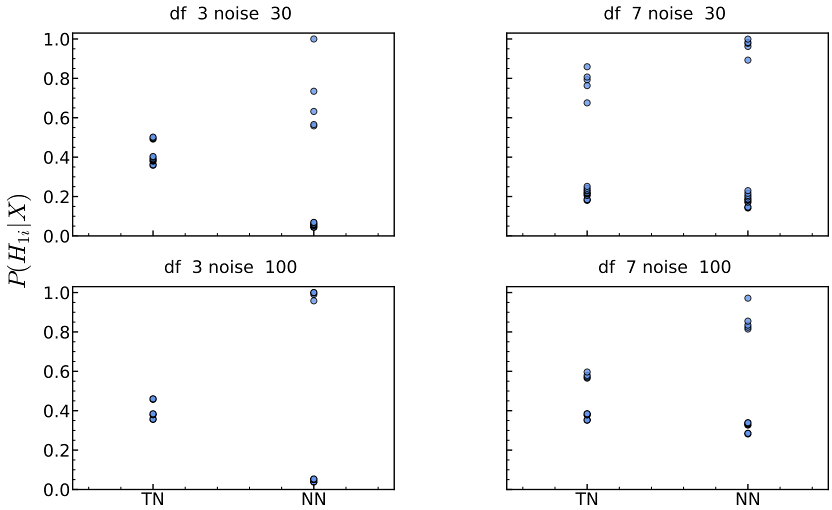


Figure 2.3.: The posterior probability of H_{1i} of some selected observations (max 5, middle 10, and min 5). Number of signals 10 and number of noises 30 and 100. Sample size is 1 and υ is known.

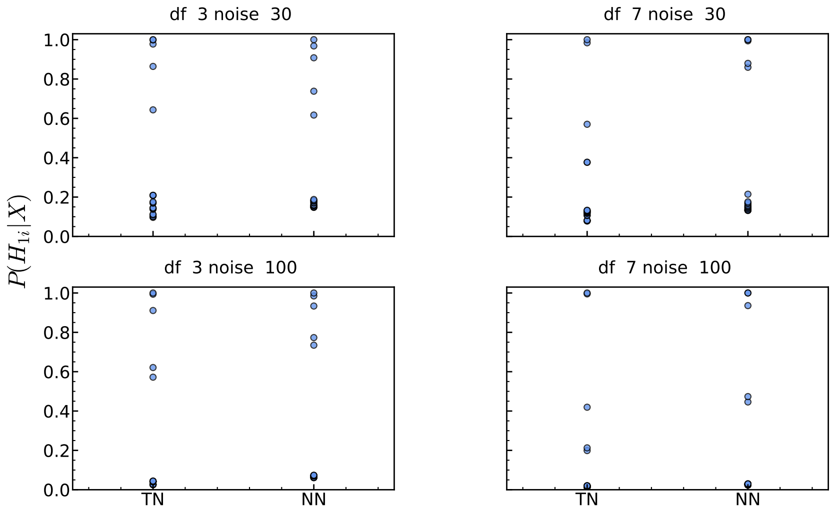


Figure 2.4.: The posterior probability of H_{1i} of some selected observations (max 5, middle 10, and min 5). Number of signals 10 and number of noises 30 and 100. Sample size is 5 and υ is known.

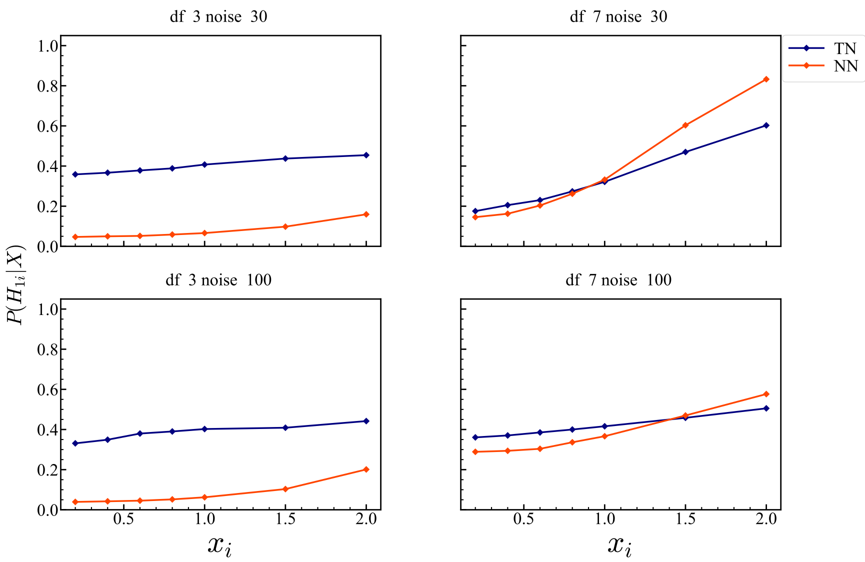


Figure 2.5.: The posterior probability of H_{1i} of one selected signal as it's value x_i increases. Number of signals 10 and number of noises 30 and 100. Sample size is 1 and υ is known (equate to original value).

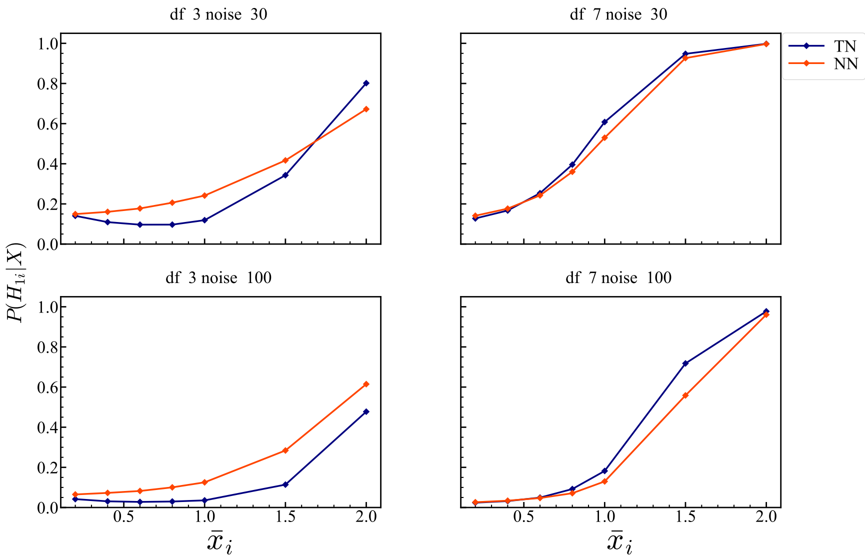


Figure 2.6.: The posterior probability of H_{1i} of one selected signal as it's sample mean $\bar{x}_i$ increases. Number of signals 10 and number of noises 30 and 100. Sample size is 5 υ is known (equate to original value).

2.4.2. Sensitivity When Using t Prior

So far, we discussed sensitivity to the choice of sampling distribution with a typical Normal prior. In the literature, t prior has been used to model location parameters of both Normal [19] and t [3] sampling distribution for robustness purposes. Therefore it is also of interest to see the sensitivity of sampling distribution when using a t prior. Degrees of freedom of t prior, ν is always assumed to be known and set to be equal to its original value $\nu = 3$.

As in figures 2.1-2.4 , we also compare posterior probabilities of $\mu_i \neq 0$ of some selected observations (minimum 5, maximum 5, and middle 10). While figure 2.7 represents results with sampling size one, figure 2.8 represents results with sampling size five. t prior results are similar to Normal prior results, with sample size one, there is a sensitivity to the sampling model, and this sensitivity is disappearing as the sample size increases to five.

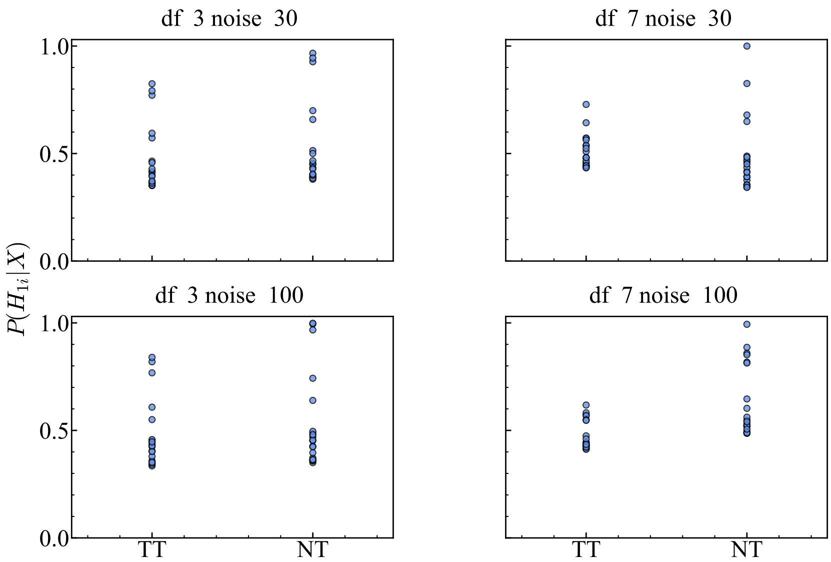


Figure 2.7.: The posterior probability of H_{1i} of some selected observations (max 5, middle 10, and min 5). Number of signals 10 and number of noises 30 and 100. Sample size is 1 υ is unknown.

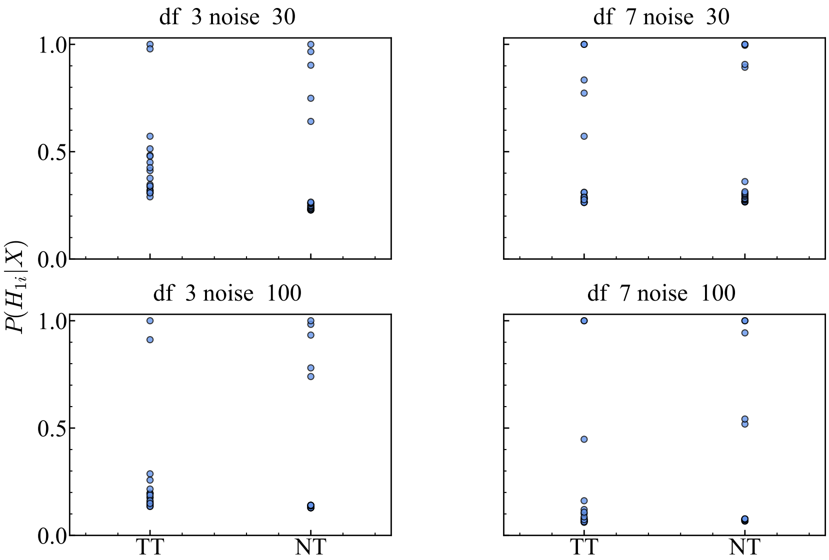


Figure 2.8.: The posterior probability of H_{1i} of some selected observations (max 5, middle 10, and min 5). Number of signals 10 and number of noises 30 and 100. Sample size is 5 and υ is unknown.

2.4.3. Compare Normal and t Priors With Normal Sampling Distribution

Topically, Normal prior is used with Normal sampling distribution in multiple testing. Some researchers are saying Normal sampling distribution with t prior is more robust [3, 19]. We want to do one more sensitivity study to the choice of prior Normal vs. t with Normal sampling distribution, i.e., we compare NN and NT models.

Figures 2.9 and 2.10 show posterior probabilities of H_{1i} being true of some selected observations (minimum 5, maximum 5, and middle 10) when the sample size is one (fig. 2.9) and five (fig. 2.10). There is a sensitivity to the choice of prior when the sample size is one: posterior probabilities under t and Normal priors are different, posterior probabilities tend to be tight under t prior than Normal prior, and t prior based posterior probabilities of H_{1i} tend to be higher than the corresponding Normal based values. When the sample size is increased to five, there is a slight sensitivity to the choice of prior and t prior based posterior probabilities tend to be a little higher than the corresponding Normal based values.

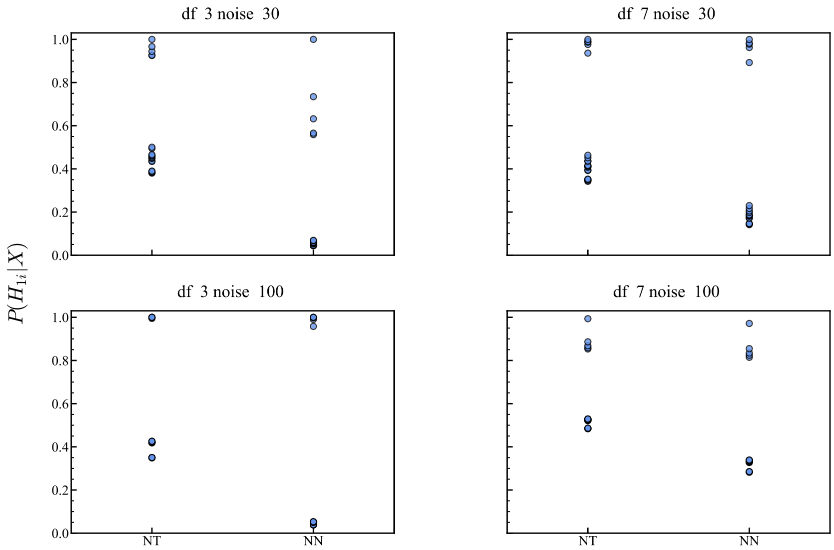


Figure 2.9.: The posterior probability of H_{1i} of some selected observations (max 5, middle 10, and min 5). Number of signals 10 and number of noises 30 and 100. Sample size is 1 and υ is unknown.

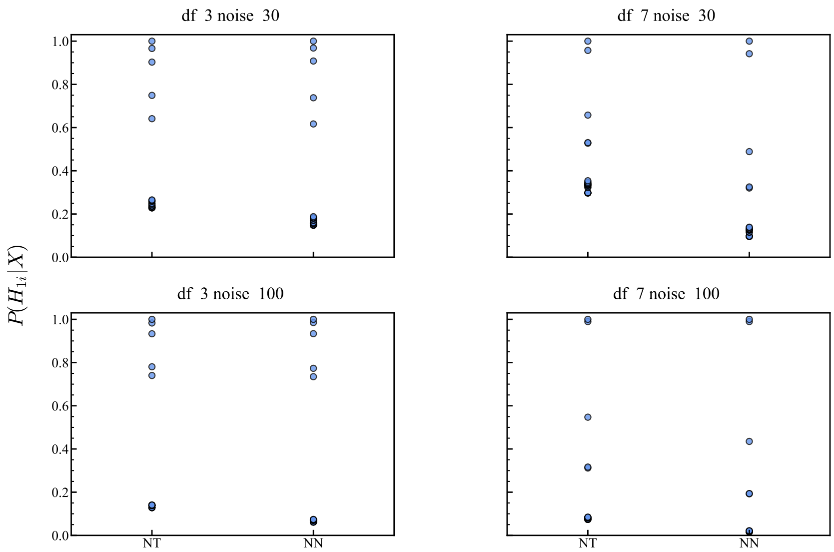


Figure 2.10.: The posterior probability of H_{1i} of some selected observations (max 5, middle 10, and min 5). Number of signals 10 and number of noises 30 and 100. Sample size is 5 and υ is unknown.

2.5. Conclusion

Literature on multiple testing and multiplicity adjustment with continuous response variables mostly focus on normally distributed responses. The normality assumption may be too restrictive in some cases, e.g. it has been shown that 5 to 15 % of the DNA samples deviate from normality and very close to t-distribution [21]. It would be of interest to study the sensitivity of normality assumption when the response variables instead follow a t-distribution. We focus on Bayesian multiple testing of means, or location parameters, when the response variable follows a t-distributions, determine suitable priors for the parameters, develop a computational strategy for computing the posterior probabilities of the hypotheses, and use it study the sensitivity of the results under normality assumption using simulation study.

In conclusion, we observe that with sample size one, the posterior probabilities are sensitive to the sampling model under both Normal and t priors for nonzero μ_is , and as the sample size increase to five, sensitivity to the sampling model is disappearing under both Normal and t priors for nonzero μ_is . So, in general, there is a sensitivity to the sampling distribution when the sample size is small, and this sensitivity is gradually decreased as the sample size increases.

When comparing for t and Normal priors for nonzero μ_is with the Normal sampling model, there is a sensitivity to the prior distribution with both sample sizes; one and five, but the sensitivity is getting smaller as the degrees of freedom and the number of noises increases.We discover that t-prior based posterior probabilities of H_1 tend to be higher than the corresponding Normal based values. This discrepancy in the posterior probabilities appears to be due to smaller degree of shrinkage when t-prior is used.

Chapter 3

Multiple testing of equality of two binomial proportions

3.1. Introduction

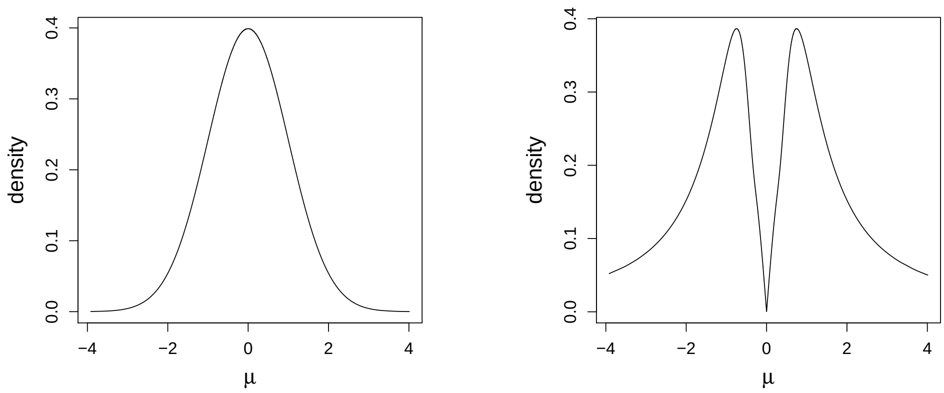
Even though there are many multiple testing (MT) scenarios where the data are discrete, for example, clinical studies, genetics, next-generation sequencing technology, psychological applications, etc., most of the research in multiple testing has developed for continuous data, the literature on multiple testing methods for discrete data is relatively scarce.

Over the past few decades, a significant amount of MT methods have been proposed on discrete data (based on false discovery rate (FDR) control) in the frequentist approach. However, most of these procedures were initially developed for continuous data, such as Benjamini-Hochberg (BH) procedure [5], and Storey's procedure [32].

As in the frequentist MT approach, the literature on the Bayesian MT approach has also focused chiefly on continuous data; not many formal Objective Bayesian approaches are yet available for comparing two binomial proportions. There is one in the literature which is done by Gecili [20]. However, this approach has a bit of asymmetry since one of the two proportions is considered control and the other as treatment. The proposed method in this chapter considers the two proportions symmetrically, and two proportions are not necessarily coming from control and treatment groups. We use some hierarchical exchangeable prior to model two proportions and develop a formal objective Bayesian method for comparing two proportions when the two proportions are exchangeable.

In Bayesian hypothesis testing, uncertainty about the unknown parameters are modeled with prior distributions under each hypothesis. Local priors and Non-local priors are two types of priors that can be used to model the unknown parameters under the alternative hypothesis.

Assume we observe a random sample of size n, $\boldsymbol{X} = (x_1, \cdots, x_n) \overset{iid}{\sim} N(\mu, \sigma^2)$ and test a point null, $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$. Let $\pi_1(\mu)$ be the prior density for μ under H_1, classified as either Local or Non-local. A Local prior is prior in which its density under H_1 peaks at the null value of the parameter, i.e., in our example, $\pi_1(\mu)$ has its maximum at $\mu = 0$. A prior is Non-local if its density under H_1 goes to zero near the null value of the parameter, $\pi_1(\mu) \to 0$ as $\mu \to 0$.



(a) (b)

Figure 3.1.: (a) Local and (b) Non-local priors for $\mu|\mu \neq 0$.

A widely accepted condition that should be required of a prior for testing a hypotheses is that the prior distribution under the alternative hypothesis should be centered and maximized at the null value of the parameter [7, 9, 13, 14], i.e., Local alternative prior. Hence, under a Local prior, if the summary statistic is very close to the null value of the parameter, then data not only strongly support H_0 but also strongly support H_1. Accordingly, under certain regularity conditions, for a true alternative hypothesis, the Bayes factor in favor of the null hypothesis decreases exponentially fast. In contrast, for a true null hypothesis, the Bayes factor favoring the alternative hypothesis decreases only at rate $O(n^{-1/2})$. These contrasting rates of convergence of Local priors imply that data is more likely to provide evidence in favor of a true alternative hypothesis than for a true null hypothesis [24].

However, a faster convergence rate of the Bayes factor favoring the true null hypothesis is desirable, especially when sparsity is desired, e.g., in MT with few signals. In the literature, it has been shown that Non-local prior has faster convergence towards the true null hypothesis compared to Local priors for continuous data. Using specific Non-local priors, Johnson and Rossell [24] have shown that a summary statistic near the null value of the parameter only strongly supports the null hypothesis, not both hypotheses; improved the discrepancy of convergence rates of the Bayes factors. Therefore, we consider Non-local priors as a good alternative to Local prior, especially when the data is sparse, and extend the use of Non-local prior for comparing two binomial proportions. The Threshold prior described in this chapter is a Non-local prior that improves on the discrepancy of convergence rates of Bayes factors between in favor of true null and true alternate hypotheses.

We adopt the formal objective Bayesian approach for testing equality of two binomial proportions under Local and Threshold priors. First, we consider testing a single hypothesis of a single proportion to select a suitable form for Threshold prior and investigate the convergence properties of the posterior disribution under each selected prior. Then, we extend to single testing of the equality of two binomial proportions and later to MT of the equality of two binomial proportions with the appropriate prior choices selected from sections 3.2 and 3.3.

3.2. Single Testing of a Binomial Proportion

Although testing a single proportion is an old problem, there are still recent developments like using Threshold and other priors about testing proportions [24]. Also, testing a binomial proportion is an illustration of two sample proportions. Therefore, we first consider a single test of a binomial proportion. We assume that the data comes from a binomial distribution with an unknown proportion, p_1; $x \sim Bin(n, p_1)$. Given that p_0 is a known constant, we are interested in testing the hypotheses,

$$H_0 : p_1 = p_0 \quad vs \quad H_1 : p_1 \neq p_0. \tag{3.2.1}$$

In the Bayesian approach, the uncertainty about the unknown parameters is expressed by assigning suitable prior distributions. We consider adopting Local and Non-local priors to model the unknown proportion p_1 under the alternative hypothesis. Let $\pi_1(p_1)$ is the prior for p_1 under H_1.

First, we adopt a Local prior for $\pi_1(p_1)$, which concentrates around p_0 and can move away from p_0 sufficiently. A mode-base Beta prior is one such candidate prior that can model the uncertainty of p_1 under the alternative hypothesis. We use the mode-base Beta prior, $\pi(p_1|r) \sim Beta(rp_0 + 1, r(1 - p_0) + 1)$, where $r = 1/w$ and $w \sim exp(1)$, used in Gecili 2018 [20]. Under this setting, p_1 has a mode value of p_0, and variance $Var(p_1) = \frac{(rp_0+1)(r(1-p_0)+1)}{(r+2)^2(r+3)}$. Here, r plays a vital role as it controls the variance of p1. As r increases, $Var(p_1)$ decreases so that p_1 is increasingly centered at its mode, p_0. Hence, as illustrate in figure 3.2, the prior distribution for p_1 would concentrate around the null hypothesis.

Later on, we propose two Non-local priors for $\pi_1(p_1)$ whose densities go to zero as p_1 goes to p_0. Local priors and can be turned into Non-local priors using a threshold. The term "Threshold prior" is used to refer to "Non-local prior" from now on.

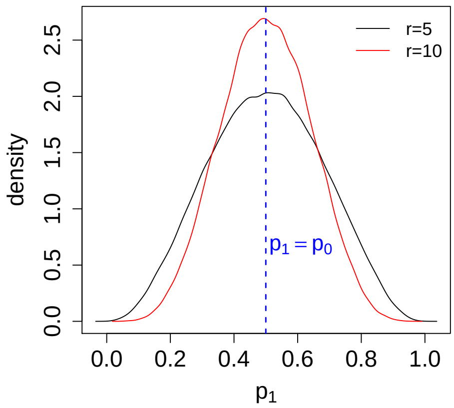


Figure 3.2.: Distribution of p_1 at two different choices of $r, (r = 5 \,\text{and}\, r = 10)$ when $p_0 = 0.5$ and $P(H_0) = 0.5$.

3.2.1. Prior Distributions for p_1

Local Prior

Here, we specify the prior distributions in two parts.

$$\text{Under } H_0 : p_1 = p_0 \tag{3.2.2}$$

$$\text{Under } H_1 : p_1|r \sim Beta(rp_0 + 1, r(1 - p_0) + 1),\ w \sim exp(1),\ \text{and } w = 1/r$$

In order to give equal priorities to both hypotheses before the data is collected, we set the prior probability of the null hypothesis, $P(H_0) = p$, equals to 0.5. Note that under H_1, the prior for p_1 is centered and has the maximum density at $p_1 = p_0$.

Then the density of p_1 under H_1 is,

$$\pi_1(p_1) = \int_0^\infty \pi(p_1|w) \cdot \pi(w)dw \tag{3.2.3}$$

Threshold Prior

Definition 3.2.1. *(Threshold prior)*

Here, we first consider a parameter p_1^*, a continuous version of p_1, with prior density given by $\pi(p_1^*|r) \equiv Beta(rp_0 + 1, r(1-p_0) + 1)$ as before. Then, we define p_1 as follows

$$p_1 = \begin{cases} p_0 & \text{if } |LOR| < \varepsilon \\ p_1^* & \text{otherwise} \end{cases}$$

where $LOR = log\left[\dfrac{p_1^*/(1-p_1^*)}{p_0/(1-p_0)}\right]$ and ε is the threshold.

To complete the model specification under the Threshold prior method, we investigate two Uniform priors for threshold parameter ε, $U(0, Kw)$ and $U(0, K)$.

Threshold prior-1

Unless a context involves substantive information to suggest the degrees of expected sparsity, a uniform prior is the natural default [27]. Hence we assume that threshold parameter ε has the Threshold prior-1, $U(0, Kw)$, adopted from Nakajima and West [27]. Here, $w \sim exp(1)$, $w = 1/r$, and K is a known constant.

Define p to be the prior probability that H_0 is true or, in this case, the probability that

$|LOR| < \varepsilon$. Given w, p has the form 3.2.4.

$$\begin{aligned} P(H_0|w) &= P(p_1 = p_0|w) \\ &= \int_0^1 \int_0^{Kw} I\big(\varepsilon > |LOR|\big)\, \pi(p_1|w)\, \pi(\varepsilon) d\varepsilon dp_1 \\ &= \int_0^1 \left\{ \int_0^{Kw} I\big(\varepsilon > |LOR|\big)\, d\varepsilon \right\} \left[\frac{1}{Kw} \right] \pi(p_1|w)\, dp_1 \\ &= \int_0^1 \left[1 - \frac{|LOR|}{Kw} \right] \pi(p_1|w)\, dp_1 \end{aligned} \tag{3.2.4}$$

By integrating 3.2.4 over w, p is a deterministic function of K, $g(K)$.

$$\begin{aligned} g(K) &= \int_0^1 \int_0^{\infty} \left[1 - \frac{|LOR|}{Kw} \right] \pi(p_1|w)\, \pi(w) I(|LOR| < Kw)\, dw\; dp_1 \\ &= \int_0^1 \int_{|LOR|/K}^{\infty} \left[1 - \frac{|LOR|}{Kw} \right] \pi(p_1|w)\, \pi(w)\, dw\, dp_1 \end{aligned} \tag{3.2.5}$$

Since $g(K)$ is monontinically incresing in K when p_0 is fixed, we can assign whatever the value we want for p by choosing an appropriate value of K. As assumed under the Local prior, here too we assume the prior probability of the null hypothesis, p, is equal to 0.5. Now for a given p_0, we want to find the value for K which satisfies the condition, $p = 0.5$. Figure 3.3 below represents the value of p at different values of K when $p_0 = 0.5$. At the value $p_0 = 0.5$, p is 0.5 when K equals to 2.607727.

Next, we need to find the prior distribution for p_1 under the alternative hypothesis, $\pi_1(p_1)$. Induced by the prior specification in 3.2.1, we have to consider two situations, $|LOR| < Kw$ and $|LOR| > Kw$, separately.

$$\pi_1(p_1|w) = \begin{cases} \displaystyle\int \pi(p_1|w)\, I\big(\varepsilon < |LOR| < Kw\big)\, \pi(\varepsilon)\, d\varepsilon & \text{if } |LOR| < Kw \\ \pi(p_1|w) & \text{if } |LOR| > Kw \end{cases}$$

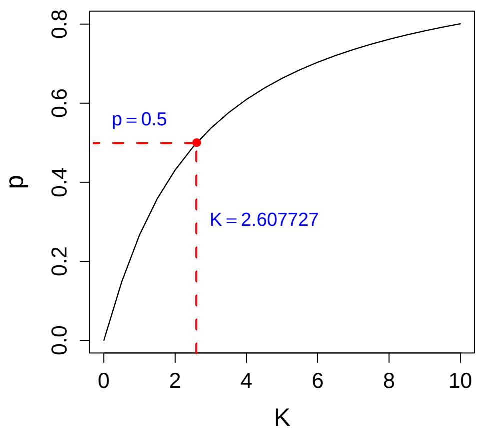


Figure 3.3.: p as a function of K when $p_0 = 0.5$.

This can be simplified as

$$
\begin{aligned}
\pi_1(p_1|w) &= \pi(p_1|w)\left[I\big(|LOR| > Kw\big) + \int_0^{Kw} I\big(\varepsilon < |LOR| < Kw\big)\pi(\varepsilon)\,d\varepsilon\right] \\
&= \pi(p_1|w)\left[B_1 + B_2\int_0^{|LOR|}\pi(\varepsilon)\,d\varepsilon\right] \\
&= \pi(p_1|w)\left[B_1 + B_2\left(\frac{|LOR|}{Kw}\right)\right] \qquad (3.2.6)
\end{aligned}
$$

where $B_1 = I\big(|LOR| > Kw\big)$ and $B_2 = I\big(|LOR| < Kw\big)$. Therefore the pdf of prior for $p_1|p_1 \neq p_0$ can be defined as

$$
\begin{aligned}
\pi_1(p_1) &= \frac{1}{(1-p)}\int_0^{\infty}\pi_1(p_1|w)\quad\pi(w)\,dw \\
&= \frac{1}{(1-p)}\quad\left[\int_0^{|LOR|/K}\pi(p_1|w)\quad\pi(w)\,dw\right. \qquad (3.2.7) \\
&\quad\left.+\int_{|LOR|/K}^{\infty}\pi(p_1|w)\left(\frac{|LOR|}{Kw}\right)\pi(w)\,dw\right]
\end{aligned}
$$

In order for $\pi_1(p_1)$ to be a Non-local prior, $\pi_1(p_1)$ must go to zero as $p_1 \to p_0$. The first term of $\pi_1(p_1)$ will be zero due to the upper limit of the integral $|LOR|/K$ being zero.

In the second term of $\pi_1(p_1)$, $\left(\dfrac{|LOR|}{Kw}\right)$ goes to zero as $p_1 \to p_0$. In both terms, the remaining factors are bounded or constants. So,

$$\lim_{p_1 \to p_0} \pi_1(p_1) = 0 \ \forall \text{ K and } p_0.$$

Hence, $\pi_1(p_1)$ is, in fact, a Non-local alternative prior and thus may have some of the desirable properties of a Non-local prior defined by Johnson and Rossell [24]. Figure 3.4 below displays the prior distribution of p_1 under H_1 for specific values of K and p_0 when densities are calculated from equation 3.2.7.

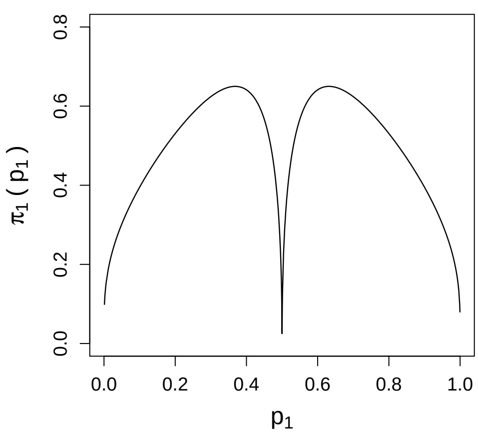


Figure 3.4.: Prior distribution of $p_1|p_1 \neq p_0$ under Threshold prior-1 when $K = 2.607727$ corresponding to $p_0 = 0.5$.

Threshold prior-2

Threshold prior-1 is influenced by the parameter w, which controls the variance of p_1. Now we want to try a prior, independent of the hyperparameters, to decide the values thresholded to zero. So, we propose a slightly modified version of Threshold prior-1 for ε, which is $\varepsilon \sim U(0, K)$.

Under this specification, the prior probability of H_0 is true can be written as

$$\begin{aligned} P(H_0|w) &= \int_0^1 \int_0^K I\big(\varepsilon > |LOR|\big)\, \pi(p_1|w)\, \pi(\varepsilon)\, d\varepsilon\, dp_1 \\ &= \int_0^1 \left[\int_0^K I\big(\varepsilon > |LOR|\big)\, d\varepsilon\right] \left[\frac{1}{K}\right] \pi(p_1|w)\, dp_1 \end{aligned}$$

and further simplified into

$$P(H_0|w) = \int_0^1 \left[1 - \frac{|LOR|}{K}\right] \pi(p_1|w)\, dp_1.$$

Therefore p can be written as a function of K

$$g(K) = \int_0^1 \int_0^\infty \left[1 - \frac{|LOR|}{K}\right] I\big(|LOR| < K\big)\, \pi(p_1|w)\, \pi(w)\, dw\, dp_1 \tag{3.2.8}$$

We can modify equation 3.2.6 and get the formula of the prior density for $\pi_1(p_1|w)$ according to the Threshold prior-2.

$$\pi_1(p_1|w) = \pi(p_1|w) \left[I\big(|LOR| > K\big) + I\big(|LOR| < K\big)\, \frac{|LOR|}{K}\right]$$

Then the pdf of prior $\pi_1(p_1)$ has the form

$$\begin{aligned} \pi_1(p_1) &= \frac{1}{(1-p)} \int_0^\infty \pi_1(p_1|w)\, \pi(w)\, dw \\ &= \frac{1}{(1-p)} \int_0^\infty \pi(p_1|w) \left[I\big(|LOR| > K\big) + I\big(|LOR| < K\big) \frac{|LOR|}{K}\right] \pi(w) dw \end{aligned} \tag{3.2.9}$$

3.2.2. Posterior Probability of H_0

Let $P(H_0|\boldsymbol{X})$ be the posterior probability of H_0 or the probability of $p_1 = p_0$.

$$
\begin{aligned}
P(H_0|\boldsymbol{X}) &= \frac{f(x|p_0)\cdot p}{f(x|p_0)\cdot p + \int_0^1 f(x|p_1)\,\pi_1(p_1)\,dp_1\ \cdot (1-p)} \\
&= \frac{Bin(x|n,p_0)\cdot p}{Bin(x|n,p_0)\cdot p + \int_0^1 Bin(x|n,p_1)\,\pi_1(p_1)\,dp_1\ \cdot (1-p)}
\end{aligned} \tag{3.2.10}
$$

In the expression 3.2.10, under each of the prior distribution assumption: Local, Threshold prior-1, and Threshold prior-2, $\pi_1(p_1)$ has the forms defined in 3.2.3, 3.2.7, and 3.2.9, respectively.

3.2.3. Bayes Factor

According to Johnson and Rossell [24], a point null hypothesis test $H_0 : \theta = \theta_0$ with a Local alternative prior density has convergence limitations of the Bayes Factor in favour of the true null hypothesis. Further, they discover that Non-local priors overcome this limitation and improve the convergence rate of the Bayes factor in favor of the true null hypothesis. So, other than comparing the posterior probability of H_0, we pay attention to use the Bayes factor as an alternative way of comparing convergence rates of true hypotheses under two prior choices, Local and Threshold priors.
The Bayes factor in favour of H_1 is defined as

$$BF(1|0) = \frac{m_1(\boldsymbol{X})}{m_0(\boldsymbol{X})}$$

where $m_i(\boldsymbol{X})$ is the marginal likelihood of data under the hypothesis H_i for $i = 0, 1$. So, $m_0(\boldsymbol{X}) = f(x|p_0)$ and $m_1(\boldsymbol{X}) = \int f(x|p_1)\,\pi_1(p_1)\,dp_1$. Under Local prior, Threshold prior-1 and Threshold prior-2, $\pi_1(p_1)$ has the forms given by equations 3.2.3, 3.2.7, and 3.2.9, respectively.

3.2.4. Computations

Since we are working on single testing here, posterior computations can be achieved efficiently using R integration. We apply this single testing procedure to a couple of simulated datasets to compare and evaluate the performances of proposed Threshold priors with the Local prior method. Under each prior distribution, Threshold prior-1, Threshold prior-2, and Local prior, we consider two sample sizes ($n = 10$ and 20) and two p_0 values (0.3 and 0.5).

We first compute K values for Threshold priors 1 and 2 so that p would be equal to 0.5 using equations 3.2.5 and 3.2.8, respectively. Then, we defined separate functions for $\pi_1(p_1)$ corresponding to equations 3.2.3, 3.2.7, and 3.2.9 relating to Local and two Threshold priors: Threshold prior-1, and Threshold prior-2. Next, compute the posterior probability of the null hypothesis for each case according to the equations 3.2.10.

Later we compute the average posterior probability of H_0 and average log Bayes factor in favour of H_1 over the data as follow.

- Average posterior probability of H_0 over x under hypothesis H_i

  $Avg_x\, P(H_0|X) = \Sigma_x\, P(H_0|X) \cdot P(X = x|H_i)$

- Average log Bayes factor in favour of H_1

  $Avg_x\, log_{10}[BF(1|0)] = \Sigma_x\, log_{10}[BF(1|0)] \cdot P(X = x|H_i)$

where $P(X = x|H_i) = Bin(x|n, p_i)$ and $i = 0, 1$.

3.2.5. Simulations and Results

In this part, we present single testing simulation results for different settings. First, we consider that the case p_0 is fixed at 0.5, and the sample size n is 10. We obtain the posterior probability of null, $P(H_0|\boldsymbol{X})$, at every value of x for $x = 1, 2, \cdots, 10$ under each prior distribution: Local prior, Threshold prior-1, and Threshold prior-2.

We repeat simulations for few other settings with $p_0 = 0.5, n = 20$; $p_0 = 0.3, n = 10$; and $p_0 = 0.3, n = 20$ and below provide some of those results.

When comparing the results from Local prior with two Threshold priors (see Table 3.1), both Non-local priors provide more substantial evidence in favor of true null hypothesis than Local prior. When comparing two Threshold priors, Threshold prior-2 provides relatively strong evidence than Threshold prior-1. Taking these results into account, we select Threshold prior-2 for further comparisons.

$\boldsymbol{x}$	$\boldsymbol{p_0 = 0.5}$			$\boldsymbol{p_0 = 0.3}$		
	$\boldsymbol{P(H_0\|X)_L}$	$\boldsymbol{P(H_0\|X)_{T_1}}$	$\boldsymbol{P(H_0\|X)_{T_2}}$	$\boldsymbol{P(H_0\|X)_L}$	$\boldsymbol{P(H_0\|X)_{T_1}}$	$\boldsymbol{P(H_0\|X)_{T_2}}$
0	0.0218	0.0157	0.01183	0.2621	0.2256	0.1998
1	0.1257	0.1046	0.0836	0.5170	0.5201	0.5408
2	0.3284	0.3117	0.2896	0.6378	0.6761	0.7456
3	0.5168	0.5324	0.5572	0.6608	0.7083	0.7896
4	0.6231	0.6643	0.7293	0.6094	0.6478	0.7150
5	0.6553	0.7045	0.7815	0.4789	0.4925	0.5176
6	0.6231	0.6643	0.7293	0.2824	0.2707	0.2526
7	0.5168	0.5324	0.5572	0.1066	0.0932	0.0746
8	0.3284	0.3117	0.2896	0.0240	0.0193	0.0140
9	0.1257	0.1046	0.0836	0.00312	0.0024	0.0017
10	0.0218	0.0157	0.0118	0.0002	0.0002	0.0001
$\boldsymbol{K}$		**2.607727**	**1.5378808**		**2.87295**	**1.7069674**

Table 3.1.: Posterior probability of H_0 under Local prior: $P(H_0|X)_L$, Threshold prior-1: $P(H_0|X)_{T_1}$, and Threshold prior-2: $P(H_0|X)_{T_2}$ at two values of p_0 : 0.5 and 0.3 and $n = 10$.

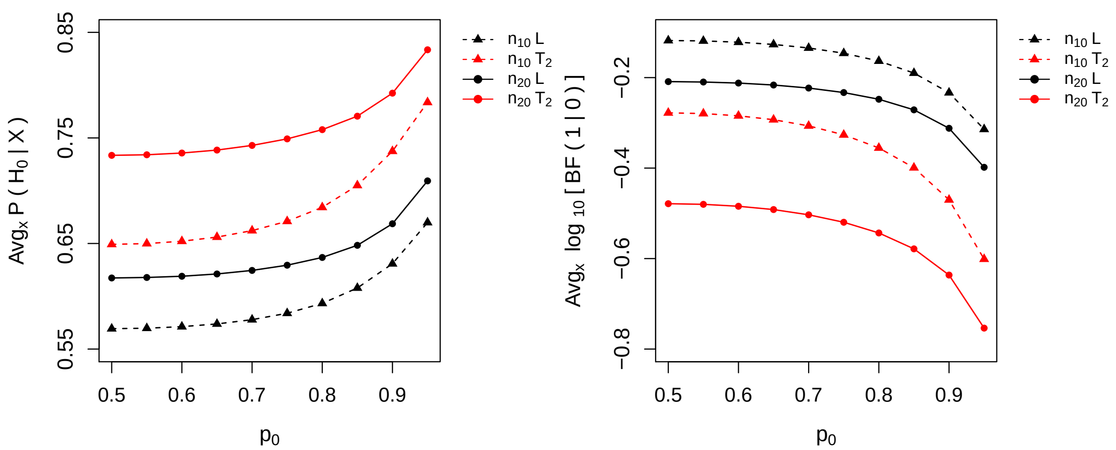
We calculate and compare the average posterior probability of the null hypothesis over x, $Avg_x\, P(H_0|X)$, and the average log Bayes factor in favor of the alternative hypothesis, $Avg_x\, log_{10}[BF(1|0)]$, under Local prior and Threshold prior-2 for two cases: true null and true alternative. Figures 3.5 and 3.6 represent the result for true null, and Figures

3.7 - 3.12 represent results for a true alternative. Under each case, we plot $Avg_x\, P(H_0|X)$ and $Avg_x\, log_{10}[BF(1|0)]$ versus either p_0 or the sample size (n).

I. When the null hypothesis is true:

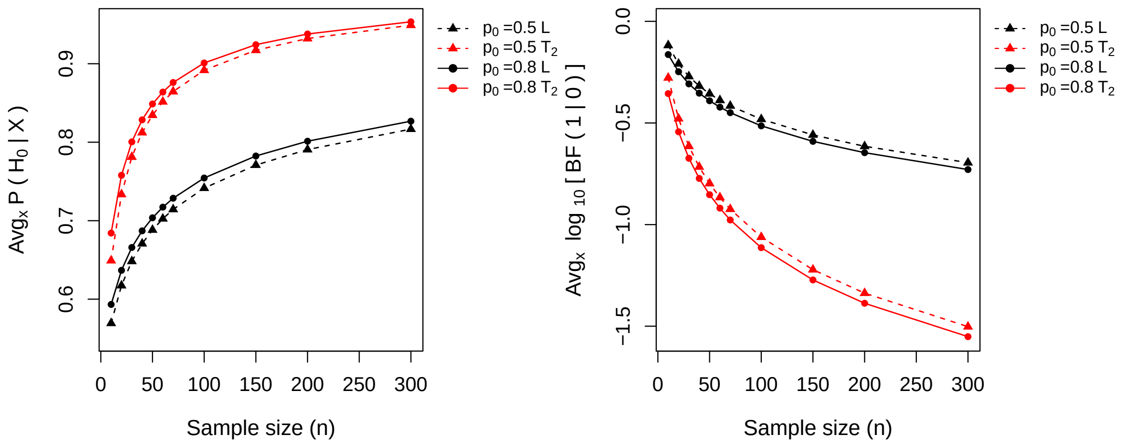
Figures 3.5 and 3.6 depict the performance of the Threshold prior versus the Local prior under the null hypothesis when the null and alternative hypotheses are treated equally. Each curve on figures 3.5 and3.6 (a) represents the average posterior probability of the null hypothesis, and 3.5 and 3.6 (b) represents the average log Bayes factor in favor of the alternative hypothesis when the null hypothesis is true and indicated the sample size.

As these figures illustrate, the Threshold prior provides strong support in favor of the null hypothesis than the Local prior. Also, the Threshold prior strongly supports the null hypothesis quickly as the sample size increases, while the Local prior requires more than 250 samples to achieve even the 80% of evidence in favor of the true null hypothesis.



(a) (b)

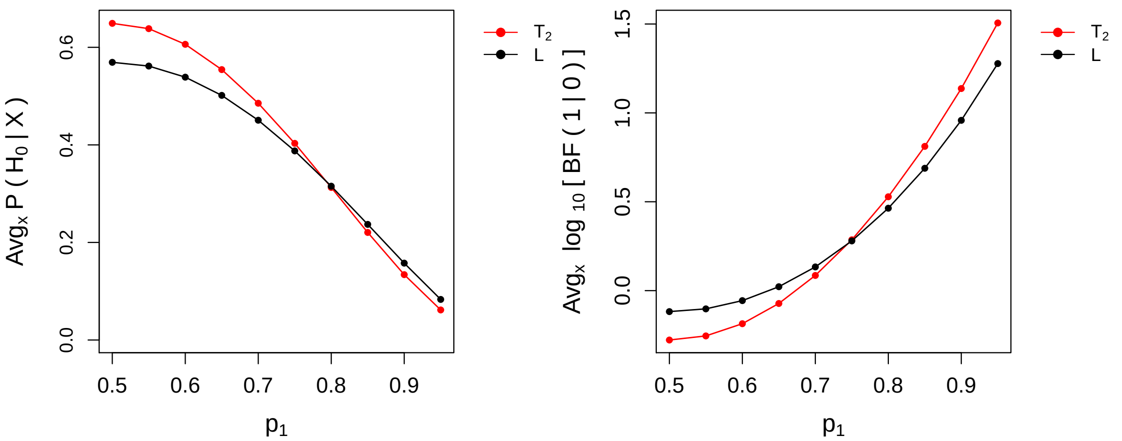
Figure 3.5.: (a) Average posterior probability of H_0 (b) average log Bayes factor in favour of H_1, over data x, as p_0 increases under Threshold prior-2 (red) and Local prior (black) when the null hypothesis is true and sample size n= 10 (dashline) and n=20 (solidline).



(a) (b)

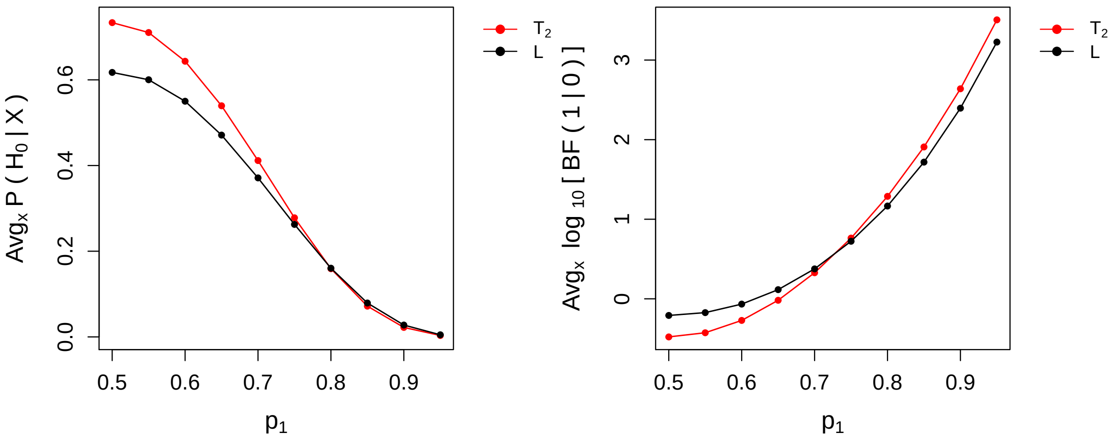
Figure 3.6.: (a) Average posterior probabilityof H_0 (b) average log Bayes factor in favour of H_1, over data x, as n increases under Threshold prior-2 (red) and Local prior (black) when the null hypothesis is true; $p_0 = 0.5$ with $K = 1.5378808$ (dashline) and $p_0 = 0.8$ with $K = 2.0002$(solidline).

II. When alternative hypothesis is true:



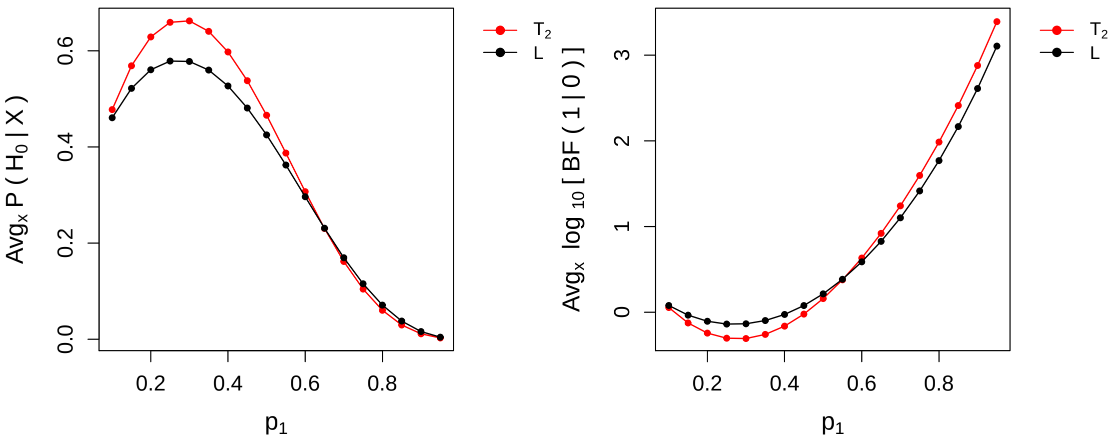
(a) (b)

Figure 3.7.: (a) Average posterior probabilityof H_0 (b) average log Bayes factor in favour of H_1, over data x, as p_1 increases and H_1 is true, under Threshold prior-2 with $K = 1.5378808$ (red) and Local prior (black) when $p_0 = 0.5$ and sample size n= 10.



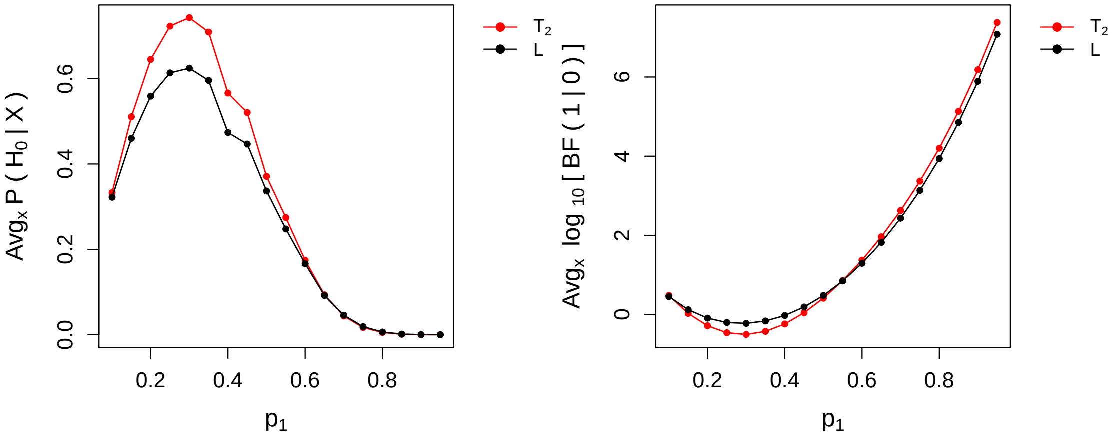
(a) (b)

Figure 3.8.: (a)Average posterior probabilityof H_0 (b)average log Bayes factor in favour of H_1, over data x, as p_1 increases and H_1 is true, under Threshold prior-2 with $K = 1.5378808$ (red) and Localprior (black) when $p_0 = 0.5$ and sample size n= 20.



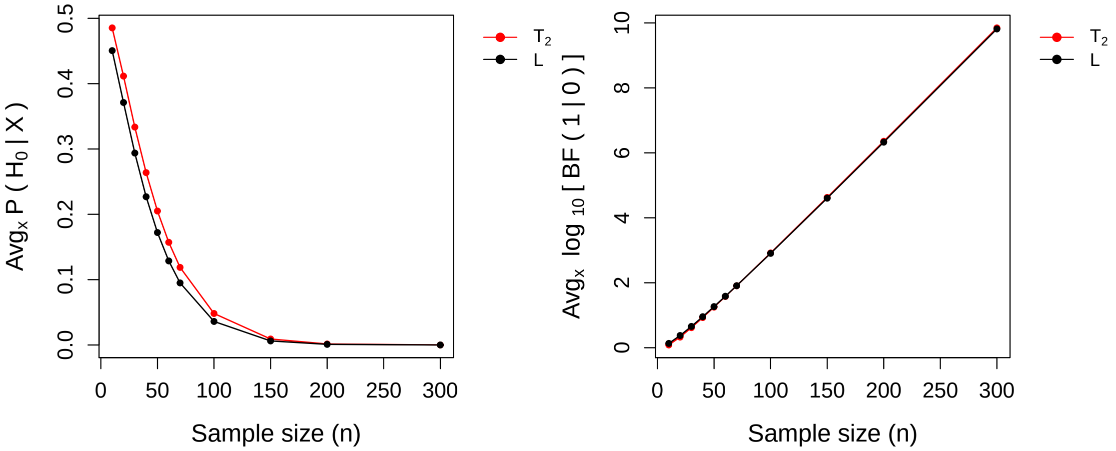
(a) (b)

Figure 3.9.: (a) Average posterior probabilityof H_0 (b) average log Bayes factor in favour of H_1, as p_1 increases and H_1 is true, under Threshold prior-2 with $K = 1.7069674$ (red) and Local prior (black) when $p_0 = 0.3$ and sample size n= 10.



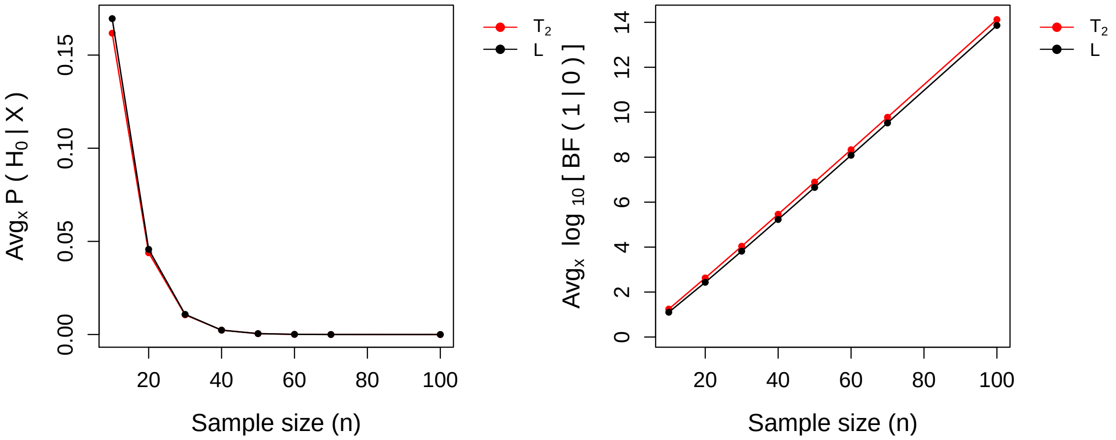
(a) (b)

Figure 3.10.: (a) Average posterior probabilityof H_0 (b) average log Bayes factor in favour of H_1, as p_1 increases and H_1 is true, under Threshold prior-2 with $K = 1.7069674$ (red) and Local prior (black) when $p_0 = 0.3$ and sample size n= 20.



(a) (b)

Figure 3.11.: (a) Average posterior probabilityof H_0 (b) average log Bayes factor in favour of H_1, over data x, as n increases under Threshold prior-2 with $K = 1.5378808$ (red) and Local prior (black) when the alternative hypothesis is true ($p_0 = 0.5$ and $p_1 = 0.3$).



(a) (b)

Figure 3.12.: (a) Average posterior probabilityof H_0 (b) average log Bayes factor in favour of H_1, over data x, as n increases under Threshold prior-2 with $K = 1.7069674$ (red) and Local prior (black) when the alternative hypothesis is true ($p_0 = 0.3$ and $p_1 = 0.7$).

Figures 3.7 - 3.12(a) illustrate the average posterior probability of H_0 and 3.7 - 3.12 (b) illustrate the average log Bayes factor in favor of H_1, under Local and Threshold prior-2 when the alternative hypothesis is true at different settings. Here, both Threshold and Local priors provide pretty similar results, and the average posterior probabilities of the null hypothesis are decreasing exponentially as the sample size increases (fig. 3.11 (a) and 3.12 (a)).

3.2.6. Conclusion for Single Testing of a Binomial Proportion

In all above results, while the Threshold prior provides substantially more evidence in favor of true null than the Local prior, for true alternative hypotheses, Threshold prior provide quite similar evidence to the Local prior. So given these results, for a true null hypothesis, under Local prior, the Bayes factor in favor of the alternative hypothesis decreases at a low rate as $n^{-1/2}$, and under the proposed Threshold prior, we improve this rate to n^{-1} for single testing (proof in Appendix B.1).

3.3. Single Testing of two Binormial Proportions

In this section, our interest is in testing equality of two unknown proportions against a two-sided alternative defined by

$$H_0 : p_1 = p_2 \quad vs \quad H_1 : p_1 \neq p_2 \tag{3.3.1}$$

Data are observed from two independent binomial distributions, $x_1 \sim Bin(n_1, p_1)$ and $x_2 \sim Bin(n_2, p_2)$. Assume that under H_0, $p_1 = p_2 = p_0$ and the prior probability of H_0 is $p = 0.5$. As discussed in section 3.2, model specification is completed by considering two priors, Local prior and Threshold prior-2.

3.3.1. Local Prior

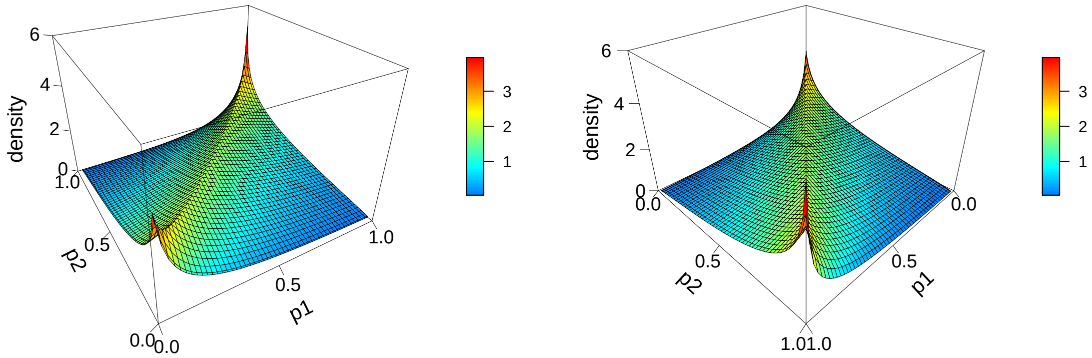
We can adopt the prior specifications in 3.2.2 to testing two binomial proportions as follows.

$$\begin{aligned} &\text{Under } H_0 : p_1 = p_2 = p_0 \text{ and } p_0 \sim U(0,1) \qquad (3.3.2)\\ &\text{Under } H_1 : p_j | p_0, r \overset{iid}{\sim} Beta(rp_0 + 1, r(1-p_0)+1) \text{ for } j = 1,2\\ &\qquad p_0 \sim U(0,1)\\ &\qquad w \sim exp(1) \text{ and } w = 1/r \end{aligned}$$

Denoting the joint prior for p_1 and p_2 under H_1 by $\pi_1(p_1, p_2)$, the density of $\pi_1(p_1, p_2)$ can be written as

$$\pi_1(p_1, p_2) = \int_0^1 \int_0^\infty \prod_{j=1}^{2} \pi(p_j | p_0, w)\, \pi(w) dw\, dp_0. \tag{3.3.3}$$

Figure 3.13 illustrates the 3D surfaces of the density of $\pi_1(p_1, p_2)$ under the Local prior in two perspective views.



(a) (b)

Figure 3.13.: $\pi_1(p_1, p_2)$ under Local prior at two different angles when $p = 0.5$.

3.3.2. Threshold Prior

Threshold prior definition 3.2.1 for testing a single binomial proportion can be modified for testing two binomial proportions.

$$p_j^*|p_0, r \overset{iid}{\sim} Beta(rp_0 + 1, r(1 - p_0) + 1) \tag{3.3.4}$$

$$p_1 = p_1^* \quad \text{and} \quad p_2 = \begin{cases} p_1 & \text{if } |LOR| < \varepsilon \\ p_2^* & \text{otherwise} \end{cases}$$

where $LOR = log\left[\dfrac{p_2^*/(1-p_2^*)}{p_1/(1-p_1)}\right], \varepsilon \sim U(0, K), p_0 \sim U(0, 1), w \sim exp(1)$, and $w = 1/r$.

We can get the expression for the prior probability of p as a function of K, $g(K)$, by modifing the equation 3.2.8 from section 3.2.1 as below. The value for K corresponding to $p = 0.5$ is 2.151137.

$$g(K) = \int_0^1 \int_0^1 \int_0^1 \int_0^\infty \left[1 - \frac{|LOR|}{K}\right] I\big(|LOR| < K\big) \prod_{j=1}^{2} \pi(p_j|p_0, w)\, \pi(w)\, dw\, dp_0\, dp_1\, dp_2 \tag{3.3.5}$$

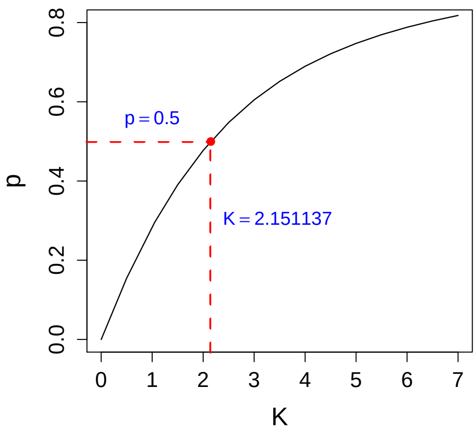
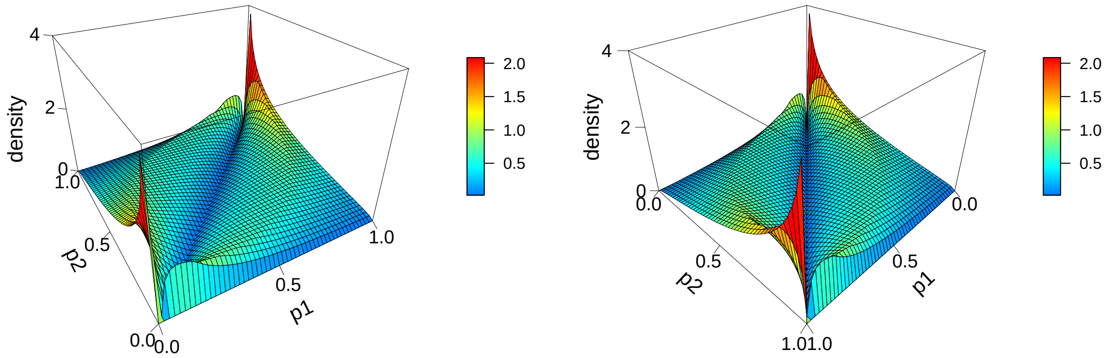


Figure 3.14.: p as a function of K

Now let $\pi_1(p_1, p_2)$ be the pdf of joint prior for p_1 and p_2 under H_1 and has the form given by equation 3.3.6, which we can derive from equation 3.2.9 in section 3.2.1.

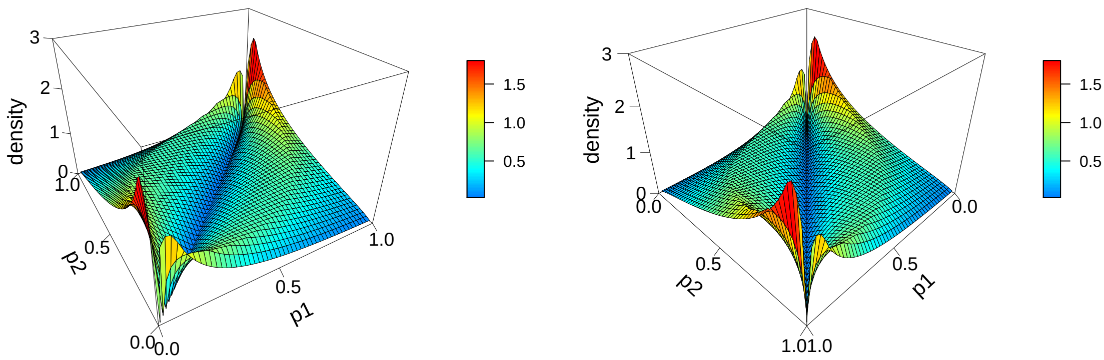
$$\pi_1(p_1, p_2)$$
$$= \frac{1}{(1-p)} \int_0^1 \int_0^\infty \prod_{j=1}^{2} \pi(p_j|p_0, w) \left[I\big(|LOR| > K\big) + I\big(|LOR| < K\big) \frac{|LOR|}{K} \right] \pi(w) dw\, dp_0 \tag{3.3.6}$$

Figures 3.15 and 3.16 display two 3D surface views of the density of $\pi_1(p_1, p_2)$ under Threshold prior-2. While Figure 3.15 displays density values over the full range of p_1 and p_2, figure 3.16 displays density values after removing points so close to the corners (0,0), (0,1), (1,0), and (1,1).



(a) (b)

Figure 3.15.: $\pi_1(p_1, p_2)$ under Threshold prior-2 at two different angles when $p = 0.5$ and $K = 2.151137$.



(a) (b)

Figure 3.16.: $\pi_1(p_1, p_2)$ under Threshold prior-2 at two different angles when $p = 0.5$ and $K = 2.151137$, after removing points so close to the corners $(0, 0)$,(0,1),(1,0), and $(1, 1)$

3.3.3. Posterior Probability of H_0

Let $L(H_0) = f(x_1|p_1)f(x_2|p_1)$ and $L(H_1) = f(x_1|p_1)f(x_2|p_2)$ then, the posterior probability of H_0 has the form

$$P(H_0|\boldsymbol{X}) = \frac{\int_0^1 L(H_0)\,\pi(p_0)\,dp_0 \cdot p}{\int_0^1 L(H_0)\,\pi(p_0)\,dp_0 \cdot p + \int_0^1 \int_0^1 L(H_1)\,\pi_1(p_1, p_2)\,dp_1\,dp_2 \cdot (1-p)} \tag{3.3.7}$$

3.3.4. Computations

Posterior computations can achieve by using R integration efficiently as in the fixed p_0 scenario. Under Threshold priors- 2, we first compute K using equation 3.3.5 in a way that p equals 0.5. Then we define corresponding functions and compute the posterior probability of the null hypothesis for each case according to equations 3.3.7.

To calculate the average and standard deviation of the posterior probability of H_0 from an MCMC sample, for example, in figure 3.19 (a), we first set $p_1 = p_2 = 0.5$ and $n_1 = n_2 = n$ then, select a value for n such that $n = 20, 40, \cdots, 200$. At these selected values of n, draw x_{i1} and x_{i2} from $Bin(n, 0.5)$ and calculate $P(H_0|x_{i1}, x_{i2})$. Here i represents MCMC iteration and $i = 1, 2, \cdots, M'$. Finally, calculate the average $P(H_0|X)$ and 95%CIs using this MCMC sample of size M' and repeat the same steps for all n.

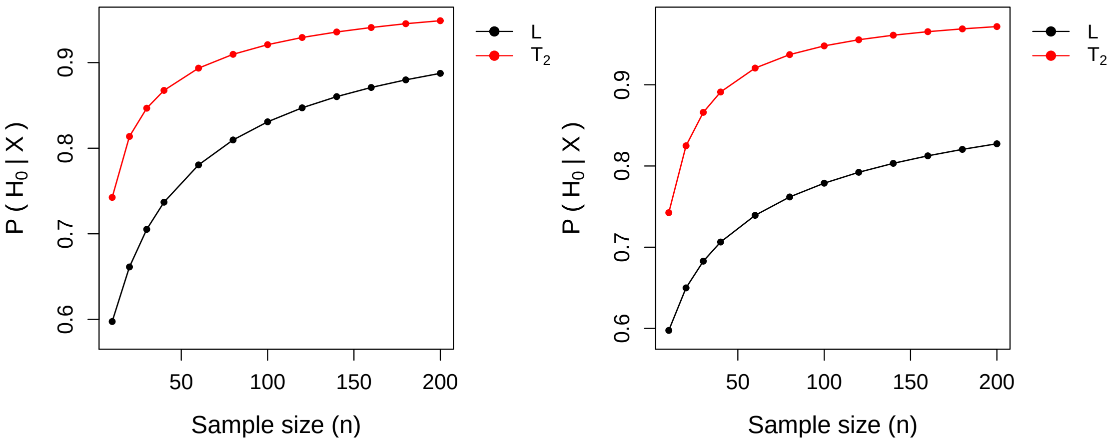
$$Avg\, P(H_0|X) = \frac{\Sigma_{i=1}^{M'} P(H_0|x_{i1}, x_{i2})}{M'} \quad SE = \frac{std}{\sqrt{M'}} \quad CI = Avg\, P(H_0|X) \pm 1.96 * SE$$

3.3.5. Results and Conclusion

The above figures in section 3.3.4 summarize the posterior and average posterior probabilities of the null hypothesis under Local prior and Threshold prior-2 when the null and alternative hypotheses are treated equally.

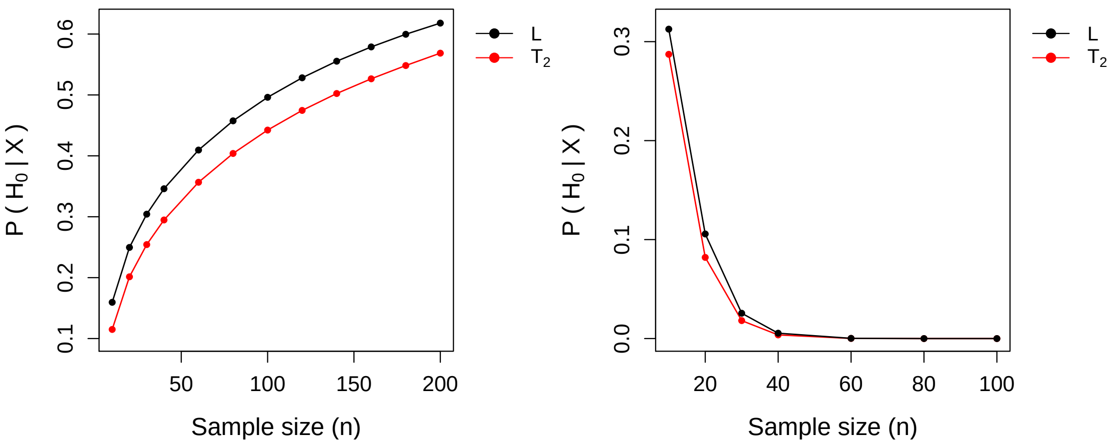
When the null hypothesis is true, Threshold prior-2 provides solid support favoring H_0 than the Local prior. Threshold prior-2 achieves strong support more quickly (with less than 20 samples) as the sample size increases, while the Local prior require larger samples to achieve even 80% of evidence favoring the true null.

In general, when comparing all these results, Threshold prior-2 provides substantially more evidence in favor of true null and similar or quite more evidence in favor of true alternative hypotheses than Local prior.



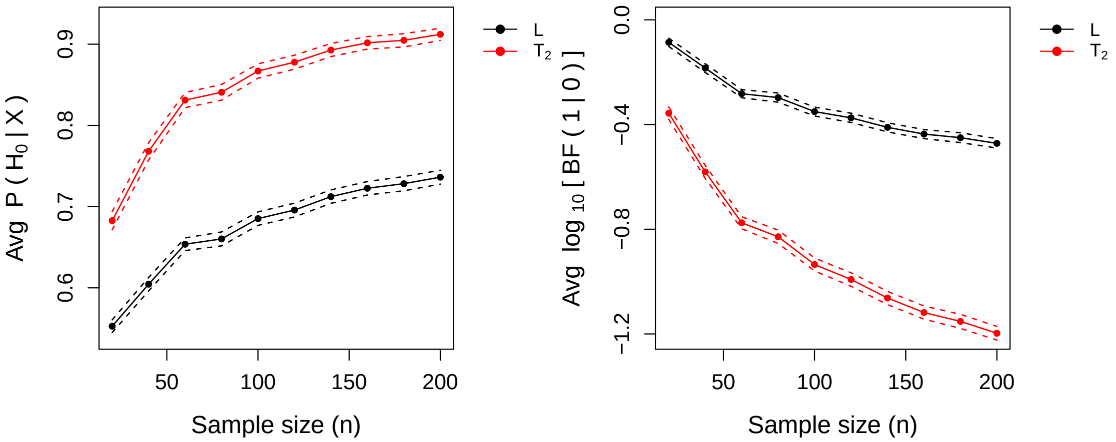
(a) (b)

Figure 3.17.: Posterior probability of H_0 when $p = 0.5$ under Local prior (black) and Threshold prior-2 with $K = 2.151137$ (red) when (a) $x_1 = x_2 = 5$ (b) $x_1 = x_2 = 0.5n$, as sample size (n) increases.



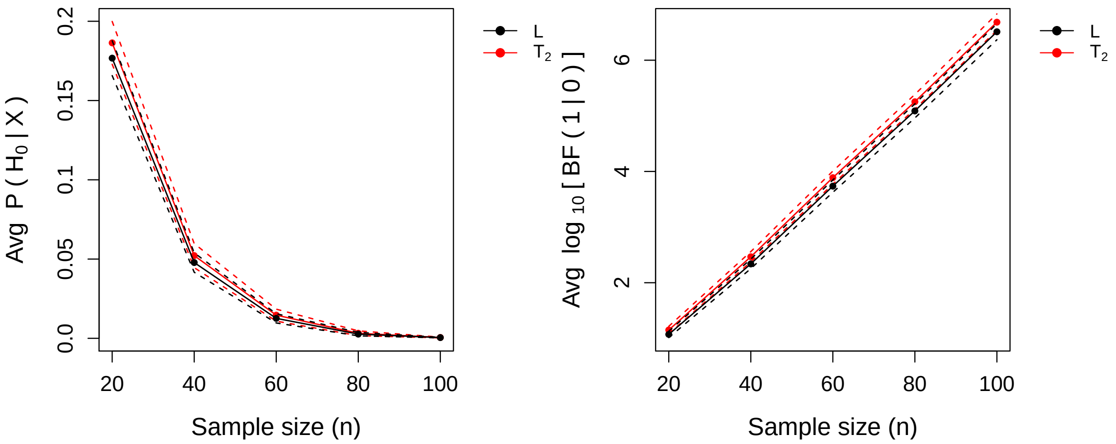
(a) (b)

Figure 3.18.: Posterior probability of H_0 when $p = 0.5$ under Local prior (black) and Threshold prior-2 with $K = 2.151137$ (red) when (a) $x_1 = 1$; $x_2 = 6$ (b) $x_1 = 0.2n$; $x_2 = 0.6n$, as sample size (n) increases.



(a) (b)

Figure 3.19.: (a) Average posterior probability of H_0 (b) average log Bayes factor in favour of H_1, as a function of sample size (n)under Local prior (black) and Threshold prior-2 with $K = 2.151137$ (red), when $p = 0.5$; $x_{i1}, x_{i2} \sim Bin(n, 0.5)$; and number of replicates $i = 1, \cdots, 1000$. Dash lines represent 95% CIs.



(a) (b)

Figure 3.20.: (a) Average posterior probability of H_0 (b) average log Bayes factor in favour of H_1, as a function of sample size (n) increases under Local prior (black) and Threshold prior-2 with $K = 2.151137$ (red), when $p = 0.5$; $x_{i1} \sim Bin(n, 0.3)$; $x_{i2} \sim Bin(n, 0.7)$; and number of replicates $i = 1, \cdots, 1000$. Dash lines represent 95% CIs.

3.4. Multiple Testing of Equality of Two Binormial Proportions

We extend the above-discussed methods to test multiple hypotheses for equality of two binomial proportions simultaneously. The goal is simultaneously testing the hypotheses given in equation 3.4.1 under the assumption that for each test i, two binomial counts are observed independently from $x_{i1} \sim Bin(n_{i1}, p_{i1})$ and $x_{21} \sim Bin(n_{i2}, p_{i2})$, where $i = 1, \cdots, M$ denotes the number of tests.

$$H_{0i} : p_{i1} = p_{i2} \quad vs \quad H_{1i} : p_{i1} \neq p_{i2} \tag{3.4.1}$$

For each i, we further assume that under the null hypothesis p_{i1} and p_{i2} equal to a common value p_{i0}. The Local and Threshold prior specifications in 3.3.2 and 3.3.4 can be modified for multiple testing as given below in 3.4.2 and 3.4.4.

3.4.1. Local Prior

Under multiple testing, for each test i, two parts of the Local prior distribution are

$$\begin{aligned} &\text{Under } H_{0i} : p_{i1} = p_{i2} = p_{i0} \text{ and } p_{i0} \sim U(0,1) \qquad (3.4.2)\\ &\text{Under } H_{1i} : p_{ij}|p_{i0}, r \overset{iid}{\sim} Beta(rp_{i0} + 1, r(1 - p_{i0}) + 1) \text{ for } j = 1, 2\\ &\qquad p_{i0} \sim U(0,1)\\ &\qquad w \sim exp(1) \text{ and} w = 1/r \end{aligned}$$

Let p be the proportion of true null hypotheses in M tests and assign the Uniform prior $U(0,1)$, i.e., $P(H_{0i}) = p \sim U(0,1)$. Denote $\pi(p_{ij}|p_{i0}, w)$ for $j = 1, 2$ is the prior distribution for each p_{ij} under H_{1i}. Then the joint prior for (p_{i1}, p_{i2}) is $\pi_{1i}(p_{i1}, p_{i2})$, given by equation 3.4.3.

$$\pi_{1i}(p_{i1}, p_{i2}) = \int_0^1 \int_0^\infty \prod_{j=1}^{2} \pi(p_{ij}|p_{i0}, w)\, \pi(w) dw\, dp_{i0} \tag{3.4.3}$$

3.4.2. Threshold Prior Method

Given $p_{i0} \sim U(0,1)$, $w \sim exp(1)$, and $w = 1/r$, for multiple testing, we define the Threshold prior as

$$p^*_{ij}|p_{i0}, r \overset{iid}{\sim} Beta(rp_{i0} + 1, r(1 - p_{i0}) + 1) \tag{3.4.4}$$
$$p_{i1} = p^*_{i1} \quad \text{and} \quad p_{i2} = \begin{cases} p_{i1} & \text{if } |LOR_i| < \varepsilon \\ p^*_{i2} & \text{otherwise} \end{cases}$$

where $LOR_i = log\left[\frac{p^*_{i2}/(1-p^*_{i2})}{p_{i1}/(1-p_{i1})}\right]$, $\varepsilon \sim U(0, K)$, and $K \sim \pi(K)$.

As in the Local prior method, we assume p be the prior probability of null hypotheses. From previous sections, we know that given K, p is a fixed quantity. So, p is a deterministic function of K. For a test i, when K is given, p has the form below derived from equation 3.3.5 in section 3.3.2.

$$\begin{aligned} g(K) &= P(p_{i1} = p_{i2}|K) \\ &= \int_0^1 \int_0^1 \int_0^1 \int_0^\infty \left[1 - \frac{|LOR_i|}{K}\right] I\big(|LOR_i| < K\big) \prod_{j=1}^{2} \pi(p_{ij}|p_{i0}, w)\pi(w)\, dw dp_{i0} dp_{i1} dp_{i2} \end{aligned}$$

By assigning derived exponential prior under the section *'Choose a Prior for K,'* $Exp(0.3)$, for K and integrating the above expression for $g(K)$ over K, p can be written as

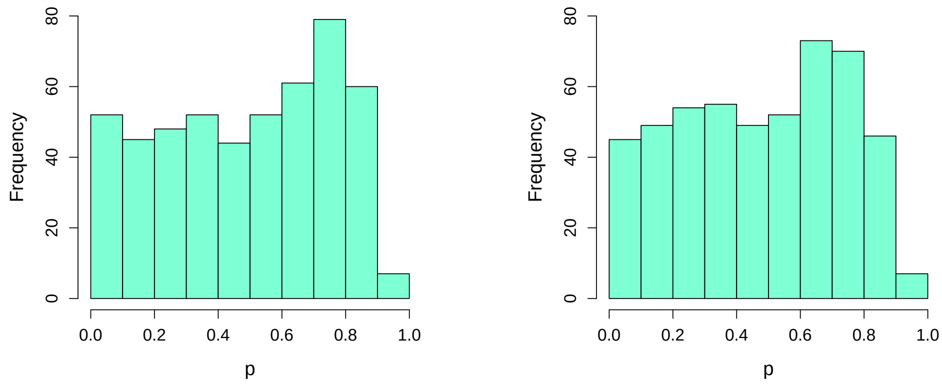
$$p = \int_{|LOR_i|}^\infty \int_0^1 \int_0^1 \int_0^1 \int_0^\infty \left[1 - \frac{|LOR_i|}{K}\right] \prod_{j=1}^{2} \pi(p_{ij}|p_{i0}, w)\pi(w)\pi(K)\, dw dp_{i0} dp_{i1} dp_{i2} dK. \tag{3.4.5}$$

Define $\pi_{1i}(p_{i1}, p_{i2})$ as the joint prior for p_{i1} and p_{i2} under H_{1i}. Then, we can get the below expression for $\pi_{1i}(p_{i1}, p_{i2})$ by upgrading corresponding notations and integrating equation 3.3.3 in section 3.3.2 over K.

$$\begin{aligned}
\pi_{1i}(p_{i1}, p_{i2}) = \frac{1}{(1-p)} \int_0^\infty \int_0^1 \int_0^\infty \Bigg[\prod_{j=1}^{2} \pi(p_{ij}|p_{i0}, w) I\big(|LOR_i| > K\big) \\
+ \prod_{j=1}^{2} \pi(p_{ij}|p_{i0}, w) I\big(|LOR_i| < K\big) \frac{|LOR_i|}{K} \Bigg] \pi(w)\pi(K) dw dK dp_{i0} \\
= \frac{1}{(1-p)} \int_0^{|LOR_i|} \int_0^1 \int_0^\infty \prod_{j=1}^{2} \pi(p_{ij}|p_{i0}, w)\, \pi(w)\, \pi(K) dw\, dK\, dp_{i0} \\
+ \frac{1}{(1-p)} \int_{|LOR_i|}^\infty \int_0^1 \int_0^\infty \prod_{j=1}^{2} \pi(p_{ij}|p_{i0}, w) \frac{|LOR_i|}{K}\, \pi(w)\, \pi(K) dw\, dK\, dp_{i0}
\end{aligned} \tag{3.4.6}$$

Choosing a Prior for K

Now we need to find a prior for K such that $p \sim U(0, 1)$. From section 3.3.2, we know that when $K = 0$, $p = 0$; $K = 2.15$, $p \approxeq 0.5$; and $K \to \infty$, $p \to 1$. Based on this information, we will find an exponential function that fits "best" to the curve of K vs. p in figure 3.14 in section 3.3.2. Among the few different functions we tried, $p = 1 - e^{-aK}$ with both $a = 0.3$ and $a = 0.32$ best fits the curve K vs. p in figure 3.14. Knowing a, we use variable transformation with $p \sim U(0, 1)$ to get the distribution for K, giving $K \sim Exp(a)$. Finally, based on two histograms in figure 3.21, we pick $Exp(0.3)$ as prior for K to use in our future works.



(a) (b)

Figure 3.21.: Histogram of the prior probabilities of H_{0i} at 500 random values of K generated from (a) $exp(0.3)$, with $p_{mean} = 0.49$ and $p_{std} = 0.268$. (b) $exp(0.32)$, with $p_{mean} = 0.48$ and $p_{std} = 0.256$.

3.4.3. Posterior Distributions

For each i, p_i is the posterior probability of H_{0i} given the data $\boldsymbol{X} = (x_{i1}, x_{i2})$. Let $\boldsymbol{p_1} = [p_{11}, \cdots, p_{M1}]$, $\boldsymbol{p_2} = [p_{12}, \cdots, p_{M2}]$, $L(H_{0i}) = f(x_{i1}|p_{i0})f(x_{i2}|p_{i0})$ and $L(H_{1i}) = f(x_{i1}|p_{i1})f(x_{i2}|p_{i2})$. Then,

$$
\begin{aligned}
p_i &= \frac{f(\boldsymbol{X}|p_{i1} = p_{i2}, \boldsymbol{p_1}_{-i}, \boldsymbol{p_2}_{-i}, p)\, \pi(\boldsymbol{p_1}, \boldsymbol{p_2}) \cdot \pi(p_{i1} = p_{i2})}{f_{\boldsymbol{X}}(\boldsymbol{x})} \\
&= \frac{f(x_{i1}|p_{i1}) f(x_{i2}|p_{i2})\, f(\boldsymbol{X_{-i}}|\boldsymbol{p_1}_{-i}, \boldsymbol{p_2}_{-i}, p)\, \pi(\boldsymbol{p_1}, \boldsymbol{p_2}) \cdot p}{f_{\boldsymbol{X}}(\boldsymbol{x})} \\
&= \frac{\displaystyle\int_0^1 L(H_{0i})\pi(p_{i0}) dp_{i0} \cdot p \times \prod_{q=1 (q \neq i)}^{M} \left[\int_0^1 L(H_{0q}) \pi(p_{q0})\, dp_{q0} \cdot p + \int_0^1 \int_0^1 L(H_{1q}) \pi_{1q}(p_{q1}, p_{q2})\, dp_{q1} dp_{q2} \cdot (1-p) \right]}{\displaystyle\prod_{q=1}^{M} \left[\int_0^1 L(H_{0q}) \pi(p_{q0})\, dp_{q0} \cdot p + \int_0^1 \int_0^1 L(H_{1q}) \pi_{1q}(p_{q1}, p_{q2})\, dp_{q1} dp_{q2} \cdot (1-p) \right]}
\end{aligned}
\tag{3.4.7}
$$

3.4.4. Implement Computations Via MCMC Approach

In sections 3.2.5 and 3.3.4 under single testing, we used R-integration to generate results from the Local and Threshold priors. Since the complexity of the posterior is given by 3.4.7, we use the MCMC approach to compare the two methods for multiple testing. We suggest two technical modifications to make MCMC computations better.

When using the Threshold prior defined in terms of LOR in section 3.4.2, having an issue that affects MCMC for data with zero, for example, the posterior probability of null hypothesis H_{0i} from the MCMC approach is less than that from the numerical integration approach when both x_{i1} and x_{i2} are zero. Therefore, we add the following adjustment to the original LOR method to make MCMC run better. Let ζ be a correction, a small value. If $p^*_{ij} \leq \zeta$ in 3.4.4, then we correct the value of p^*_{ij} as $p^*_{ij} + \zeta$. Only for $j = 1$ when $p^*_{i1} \geq 1 - \zeta$, then p^*_{i1} is corrected as $p^*_{i1} - \zeta$. Otherwise, we keep the value of p^*_{ij} as it is.

Since the posterior probability of the null hypothesis depends on the correction ζ, we need to find a reasonable value for it. As the numerical integration approach workes fine with the original LOR for single testing, and $\zeta = 0.025$ gives similar results using MCMC as the integration result, we use this choice in our future works. Also, another reason for this choice is, it is expected that the Threshold prior has at least a slightly higher posterior probability of null than Local prior, and for single testing, that happens for $\zeta = 0.025$.

Modification 3.4.1. To help justify MCMC for varying dimensions when H_{0i} is true in the Local prior approach, rather than assuming the same value for p_{i1} and p_{i2} under H_{0i}, we assume that the model will have two slightly different proportions under H_{0i}. That is, we assume that the model will have two proportions under each of H_{0i} and H_{1i}. So, the proportion corresponds to x_{i1} and x_{i2} under null and alternative hypotheses are as below.

$$p'_{ij} = \begin{cases} p_{ij0} & \text{under } H_{0i} \\ p_{ij} & \text{under } H_{1i} \end{cases}$$

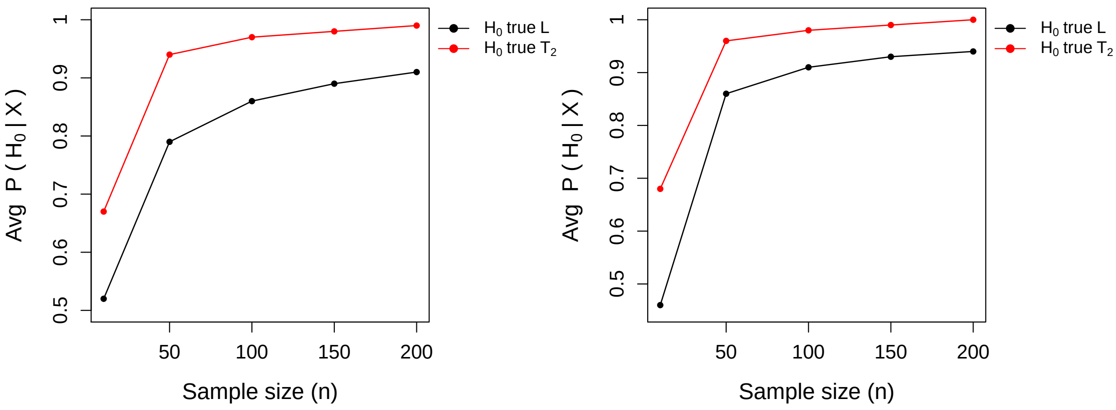
where $i = 1, \cdots, M$; $j = 1, 2$; and $p_{ij}|p_{i0}, r \overset{iid}{\sim} Beta(rp_{i0}, r(1-p_{i0}))$ as defined in 3.4.2.

When selecting a prior for p_{ij0}, we need a distribution that concentrates p_{ij0} very close to p_{i0}. This can be done by choosing a conditional distribution for p_{ij0} (under H_{0i}) similar to the prior distribution for p_{ij} under H_{1i}. Assume $p_{ij0}|p_{i0}, r_0 \sim Beta(r_0 p_{i0}, r_0(1-p_{i0}))$, and r_0 is a fixed large value. A larger value of r_0 will make the variance small and keep p_{ij0} very close to p_{i0} and yet different. Since $p_{i0} \sim U(0,1)$ and r_0 are fixed, the joint prior for $\pi(p_{i0}, p_{ij0}) \equiv \pi(p_{ij0}|p_{i0})$.

3.4.5. Simulations and Results

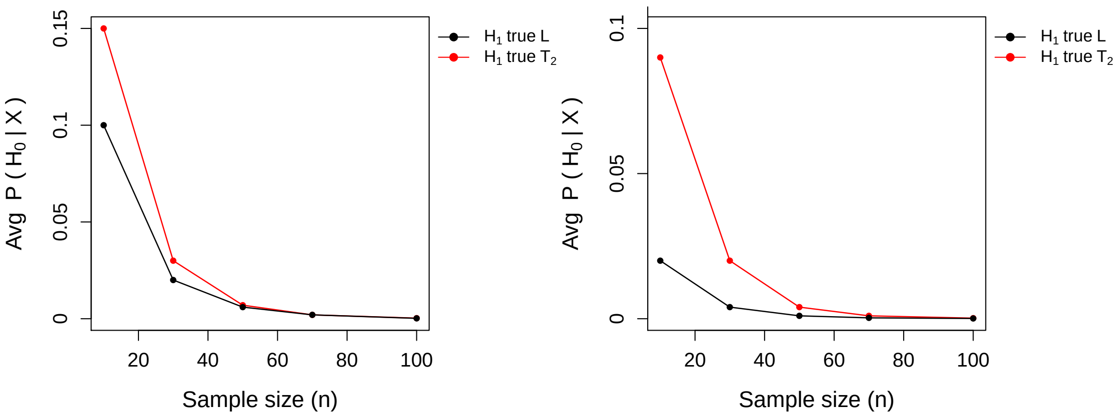
We illustrate the above two proposed Bayesian multiple testing procedures for different settings and calculate the posterior probabilities of the null hypotheses for each test. We first use synthetic datasets to evaluate the performances of two priors, and later we use a real data example to compare the results from two priors.

Similar to the setting in 3.3.4, in MT, we perform repeated simulations to calculate the average posterior probability of H_{0i} for each test i from the MCMC approach. We here consider three different cases for simulations; all the tests are null true, all the tests are alt true, and mixed case. As depicted in figure 3.22 for Threshold prior under the true null, the average posterior probability of H_{0i} goes to 1 faster than the Local prior. Under the true alternative (fig.3.23), the average posterior probability of H_{0i} for both priors goes to 0 exponentially fast. For mixed cases, out of m tests, k_0 of the number of tests are alternative true, and the rest are null true. We calculate the average posterior probability of H_{0i} for true null and true alternative hypotheses separately. Under the mixed case, for true nulls, for Threshold prior, the average posterior probability of H_{0i} goes to 1 faster than the Local prior; for true alternatives, the average posterior probability of H_{0i} from both priors goes to 0 equally faster (fig.3.24).



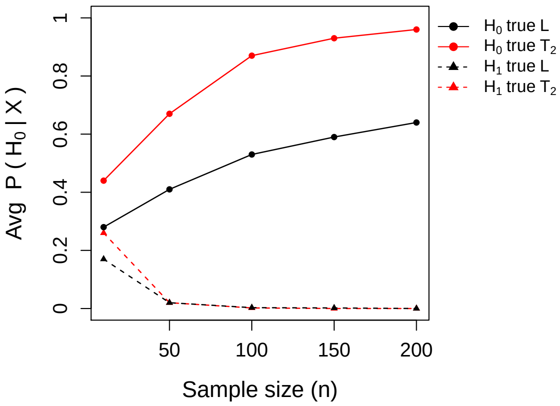
(a) (b)

Figure 3.22.: Average posterior probability of H_{0i} as a function of sample size for (a) $M = 10$ and (b) $M = 50$ tests when all the tests are null true: $x_{i1}, x_{i2} \sim Bin(n, 0.5)$, under Local prior (black) and Threshold prior (red), for 500 replicates.

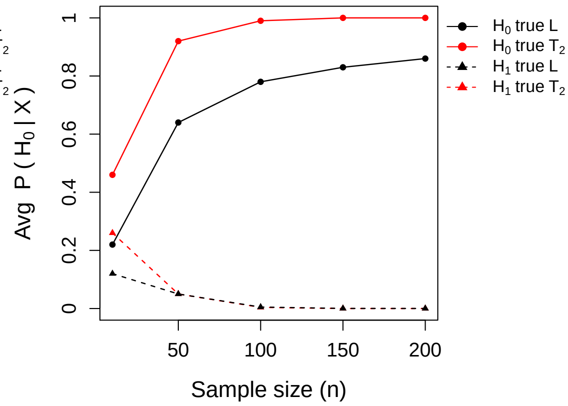


(a) (b)

Figure 3.23.: Average posterior probability of H_{0i} as a function of sample size for (a) $M = 10$ and (b) $M = 50$ tests when all the tests are alternative true: $x_{i1} \sim Bin(n, 0.3), x_{i2} \sim Bin(n, 0.7)$, under Local prior (black) and Threshold prior (red), for 500 replicates.



(a)



(b)

Figure 3.24.: Average posterior probability of H_{0i} as a function of sample size for (a) $M = 10, k_0 = 4$ and (b) $M = 50, k_0 = 10$ with k_0 number of test are alternative true: $x_{i1} \sim Bin(n, 0.3), x_{i2} \sim Bin(n, 0.7)$ and $M - k_0$ number of test are null true: $x_{i1}, x_{i2} \sim Bin(n, 0.5)$, under Local prior (black) and Threshold prior (red), for 500 replicates. While solid lines represent the results for true null hypotheses, dash lines represent true alternative hypotheses.

In figures 3.25 to 3.28, at each setting, we consider M number of tests. Out of these M tests, the k_0 number of tests are alternative true, and the rest of the $M - k_0$ tests are null true. Posterior probabilities of the null hypotheses are calculated for each of the M tests under the two priors for different settings.

We present results in two parts,

Part I - When the true proportions are far part: For true alternative hypotheses, we consider data are coming from two binomial distributions with proportions that are far apart from each other (e.g., $p_{i1} = 0.1$ and $p_{i2} = 0.9$).

Part II - When the separation between true proportions is intermediate: For true alternative hypotheses, data are coming from two binomial distributions with proportions that are close to each other (e.g., $p_{i1} = 0.3$ and $p_{i2} = 0.5$).

In figures 3.25 and 3.26, while (a) and (b) give the results for case true proportions are far apart, (c) and (d) give results for case the separation between true proportions is intermediate. All these results are based on 600K mcmc samples with a 200K burning phase.

Part I - When true proportions are far part

Data for true null hypotheses are simulated from the binomial distribution with success proportion 0.1 such that $x_{i1} \sim Bin(n, 0.1)$ and $x_{i2} \sim Bin(n, 0.1)$. For true alternative hypotheses, data are simulated from two binomial distributions with success proportions 0.1 and 0.9 such that $x_{i1} \sim Bin(n, 0.1)$ and $x_{i2} \sim Bin(n, 0.9)$. Further, we consider the sample sizes, $n = 10, 15$; the number of true alternatives, $k_0 = 5, 10$; and $r_0 = 50$.

Threshold prior-2 always gives the higher posterior probability of H_{0i} than the Local prior for true null hypotheses. When the alternative hypothesis is true, Threshold prior-2 gives either similar or a little higher result to Local prior.

As the sample size increases from 10 to 15, the posterior probability of H_{0i} for true nulls are increased; the posterior probability of H_{0i} for true alternatives are decreased, and the separation of the posterior probabilities of H_{0i} between true null and alternative hypotheses are very clearly visible ((a) and (b) in fig.3.25 and 3.27).

When the number of true alternatives, k_0, increases from 5 to 10 posterior probability of H_{0i} decreases for both true null and alternative hypotheses(fig.3.26 (a) and (b)).

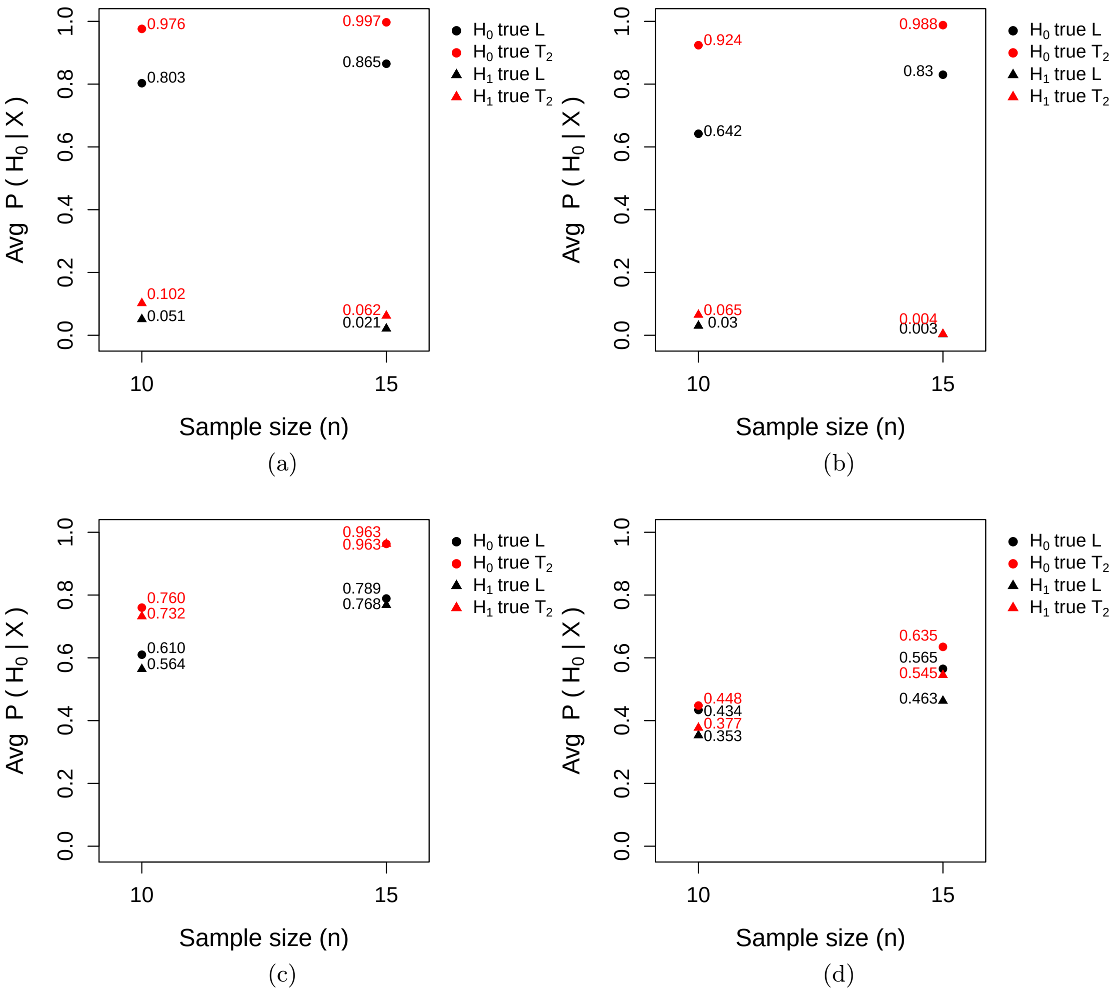


Figure 3.25.: Mean of the posterior probabilities of H_{0i}'s under Local prior (black) and Threshold prior-2 (red) as the sample size, n increases from 10 to 15 and $M = 50$. In (a) $k_0 = 5$ and (b) $k_0 = 10$: for true null $x_{i1}, x_{i2} \sim Bin(n, 0.1)$ and for true alternative $x_{i1} \sim Bin(n, 0.1), x_{i2} \sim Bin(n, 0.9)$. In (c) $k_0 = 5$ and (d) $k_0 = 10$: for true null $x_{i1}, x_{i2} \sim Bin(n, 0.3)$ and for true alternative $x_{i1} \sim Bin(n, 0.3), x_{i2} \sim Bin(n, 0.5)$. Circle-mean of $P(H_{0i}|X)$ of true null hypotheses and triangle-mean of $P(H_{0i}|X)$ of alternative hypotheses.

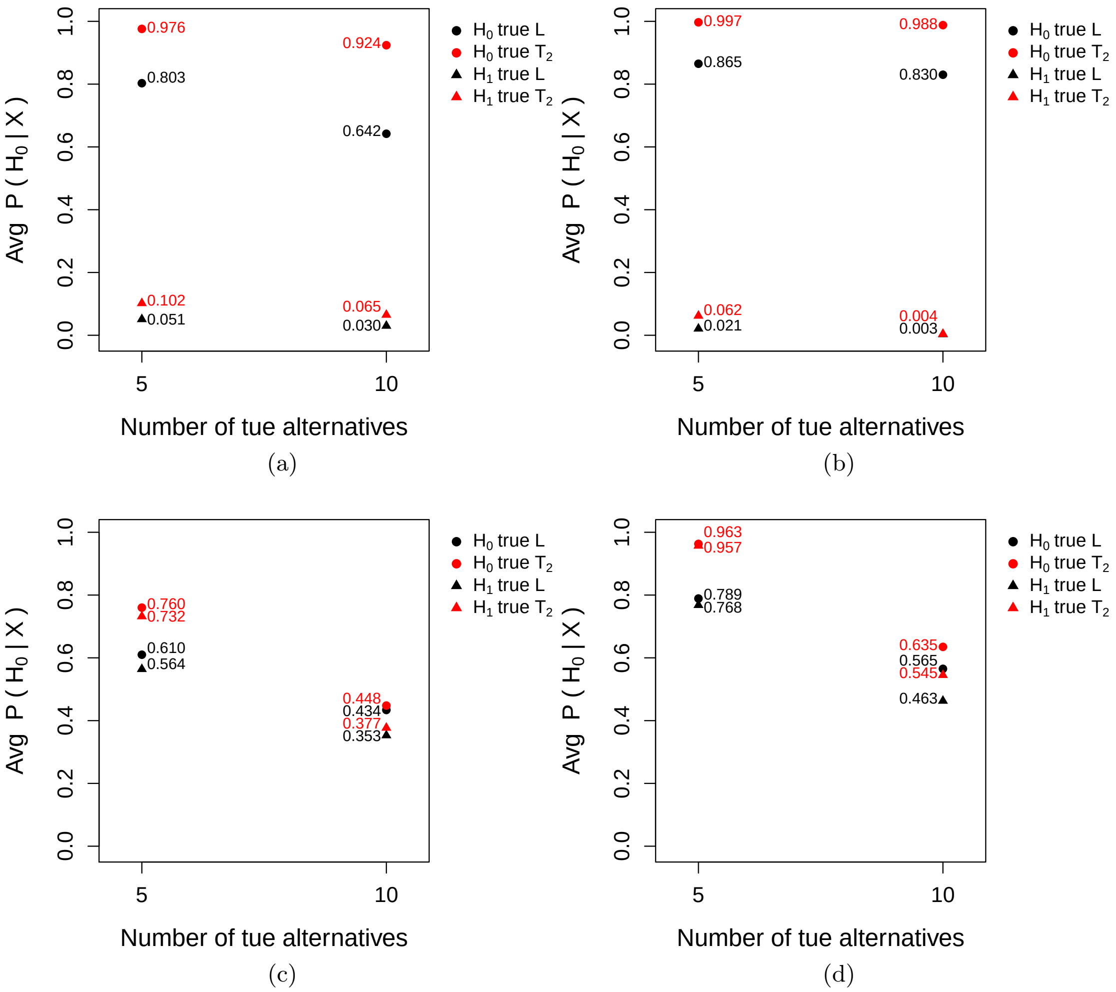


Figure 3.26.: Mean of the posterior probabilities of H_{0i}'s under Local prior (black) and Threshold prior-2 (red) as the number of true alternatives, k_0 increases from 5 to 10 and $M = 50$. In (a) $n = 10$ and (b) $n = 15$: for true null $x_{i1}, x_{i2} \sim Bin(n, 0.1)$ and for true alternative $x_{i1} \sim Bin(n, 0.1), x_{i2} \sim Bin(n, 0.9)$. In (c) $n = 10$ and (d) $n = 15$: for true null $x_{i1}, x_{i2} \sim Bin(n, 0.3)$ and for true alternative $x_{i1} \sim Bin(n, 0.3), x_{i2} \sim Bin(n, 0.5)$. Circle-mean of $P(H_{0i}|X)$ of true null hypotheses and triangle-mean of $P(H_{0i}|X)$ of alternative hypotheses.

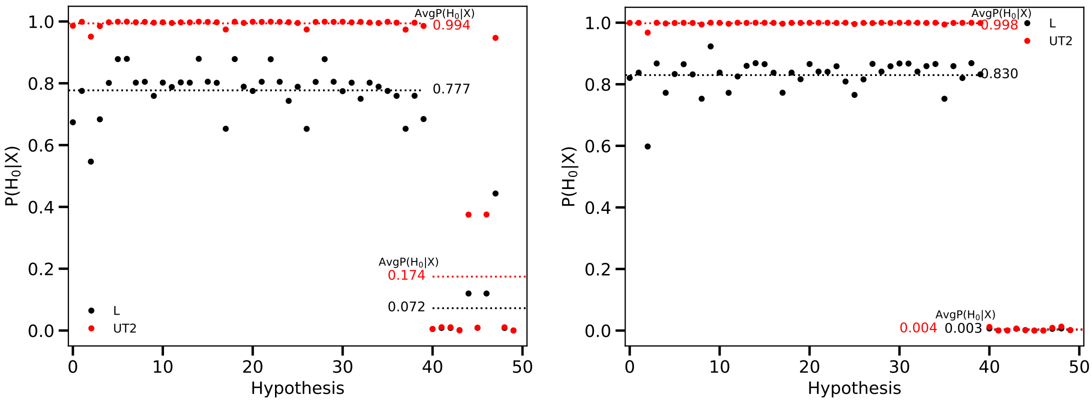
Part II - When the separation between true proportions is intermediate

Data for true null hypotheses are simulated from the binomial distribution with success proportion 0.3; $x_{i1} \sim Bin(n, 0.3)$ and $x_{i2} \sim Bin(n, 0.3)$, and true alternative hypotheses data are simulated from two binomial distributions with success proportions 0.3 and0.5; $x_{i1} \sim Bin(n, 0.3)$ and $x_{i2} \sim Bin(n, 0.5)$. Further, we consider the sample sizes, $n = 10, 15$; the number of true alternative, $k_0 = 5, 10$; and $r_0 = 50$.

When the sample size is small (e.g., $n = 10$) under the true null, for some observations Threshold prior-2 gives somewhat larger posterior probabilities for H_{0i} and some other observations, somewhat lower posterior probabilities for H_{0i} than Local prior (fig.3.28(a)). As the sample size increases from 10 to 15, Threshold prior-2 always gives larger posterior probabilities than local prior(fig.3.28(b)). However, in both cases, sample size is 10 and 15, on average, under both true nulls and alternatives, Threshold prior-2 gives a higher average posterior probability of H_{0i}, i.e., $Avg\, P(H_0|X)$, than the Local prior ((c) and (d) in fig.3.25 and fig.3.28). The separation of the posterior probabilities under the Local and Threshold priors is getting visible as the sample size increases.

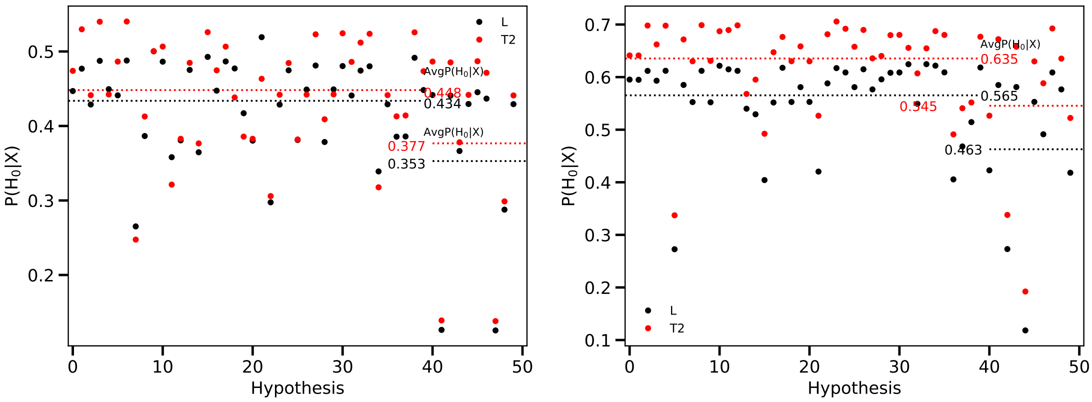
As the sample size increases from 10 to 15, the posterior probability of H_{0i} for true null and alternative hypotheses are increased. The separation of the posterior probabilities of H_{0i} between true null and alternative hypotheses are not visible ((c) and (d) in fig.3.25 and fig.3.28).

When the number of true alternatives, k_0, increases from 5 to 10, the posterior probability of true H_{0i} decreases. The separation of the posterior probability of H_{0i} between Threshold prior-2 and Local prior decrease(fig.3.26 (c) and (d)).



(a) (b)

Figure 3.27.: Posterior probability of H_{0i} under Local (black) and Threshold prior-2 (red) when $M = 50$, $k_0 = 10$ and (a) $n = 10$ (b)$n = 15$. $x_{i1} \sim Bin(n, 0.1)$ for $i = 1, \cdots, 50$; $x_{i2} \sim Bin(n, 0.1)$ for $i = 1, \cdots, 40$ and $x_{i2} \sim Bin(n, 0.9)$ for $i = 41, \cdots, 50$. Dash lines represent the average of $P(H_{0i}|X)$.



(a) (b)

Figure 3.28.: Posterior probability of H_{0i} under Local (black) and Threshold prior-2 (red) when $M = 50$, $k_0 = 10$ and (a) $n = 10$ (b)$n = 15$. $x_{i1} \sim Bin(n, 0.3)$ for $i = 1, \cdots, 50$; $x_{i2} \sim Bin(n, 0.3)$ for $i = 1, \cdots, 40$ and $x_{i2} \sim Bin(n, 0.5)$ for $i = 41, \cdots, 50$. Dash lines represent the average of $P(H_{0i}|X)$.

3.4.6. Real Data Applications

DNA Sequence Data in Taron (1990)

We apply our two proposed methods (Local prior and Threshold prior) to the DNA sequence data in Taron [33]. This data set compares the frequencies of nucleotide changes in the transcript from the control and study cells to determine if the transcribed RNA in the study cells differs from that in the control cells. The known sequence may be several hundred nucleotides in length, so the multiple comparisons problem must be addressed. Table 3.2 gives the frequencies of the nucleotide change observed at nine nucleotide sites in such an experiment. Table 3.2 summarizes the data and observed significance level with Bonferroni correction reported in Taron, along with the results of our two proposed methods. According to Tarone [33], out of 9 tests, only one rejects the null hypothesis at the nominal level $\alpha = 0.05$ using Fisher's exact test. Our two proposed Bayesian methods also result in rejecting only one null hypothesis with the cutoff of 0.5.

Control x_{i1}/n_{i1}	**Treated** x_{i2}/n_{i2}	P_i	$P(H_{0i}\|X)_L$	$P(H_{0i}\|X)_{T_2}$
1/11	3/9	0.217	0.511	0.507
2/11	4/10	0.268	0.526	0.586
2/11	2/10	0.669	0.626	0.666
1/10	8/11	0.006	0.113	0.128
1/10	3/10	0.291	0.557	0.555
2/9	2/9	0.712	0.604	0.668
2/9	2/9	0.712	0.610	0.668
2/9	2/8	0.665	0.608	0.663
3/8	2/7	0.818	0.563	0.669

Table 3.2.: The observed significance level (Bonferroni)-P_i and posterior probability of H_{0i} under: Local prior -$P(H_{0i}|X)_L$ and Threshold prior-2 -$P(H_{0i}|X)_{T_2}$. The number of tests $M = 9$ and $r_0 = 200\sqrt{n_{i1}}$.

Adverse Event Data in Heller and Gur (2011)

Now we consider the dataset relating treatment to an adverse event for ten studies considered by Gecili [20]. This dataset has also considered by Heller and Gur [23] and developed one-sided testing. The table 3.3 presents the data for the number of occurrences and non-occurrences for treatment and control groups, results from Gecili, and results from our Local prior and Threshold prior methods. Gecili investigates an objective Bayesian multiple testing procedure for testing equality of two binomial proportions under different prior specifications under the alternative hypothesis and reveals that "mode-based" Beta prior permits desirable characteristics and flexibility. In the table 3.3, MB1 represents the results from "mode-based" prior discussed by Gecili, similar to the prior we use in our work, $\pi(p_{ij}|p_{i0}, r) \equiv Beta(rp_{i0} + 1, r(1 - p_{i0} + 1))$ where $p_{i0} \sim U(0, 1)$, $w = 1/r$, and $w \sim exp(1)$. Prosoed Local and Threshold prior methods give similar results to Gecili, rejecting all the null hypotheses at the cutoff of 0.5 and for relatively close proportions Threshold prior method reports a somewhat larger posterior probability of H_{0i} than Local prior.

Control x_{0i}/n_{0i}	**Treated** x_{1i}/n_{1i}	**MB1**	$P(H_{0i}\|X)_L$	$P(H_{0i}\|X)_{T_2}$
5/25	7/21	0.234	0.229	0.328
7/22	3/14	0.249	0.266	0.384
12/32	2/38	0.003	0.005	0.004
12/20	10/40	0.025	0.026	0.034
13/16	1/16	0 .000	0.000	0.000
7/13	1/15	0.016	0.019	0.021
5/23	0/20	0.061	0.102	0.065
7/9	2/7	0.076	0.071	0.089
15/27	8/24	0.139	0.140	0.231
5/15	5/17	0.269	0.270	0.423

Table 3.3.: Posterior probability of H_{0i} from: Gecili-MB1, Local prior-$P(H_{0i}|X)_L$, and Threshold prior-2-$P(H_{0i}|X)_{T_2}$. The number of tests $M = 10$ and $r_0 = 200\sqrt{n_{i1}}$.

Clinical safety data in Mehrotra and Heyse (2004)

Now we consider applying our proposed methods to the clinical safety study data in Mehrotra and Heyse [25]. This study consist of a safety trial of a candidate quadrivalent vaccine against measles, mumps, rubella, and varicella (MMRV) conducted in 296 healthy toddlers. Participants are randomly assigned to receive either the quadrivalent MMRV on day 0 (Group 1) or the trivalent MMR on day 0, followed by varicella (V) on day 42 (Group 2). A comparison of adverse experiments between Group 1 ($n_1 = 148$), days 0-42, and Group 2 ($n_2 = 132$), days 42-84 are considered to compare the safety profile of MMRV to that of varicella component. Table1 in Mehrotra and Heyse [25] reports the number of reported cases for each of 40 adverse events. The goal is to test the null hypothesis that "variacella is not associated with the adverse event" for each of 40 adverse events. Therefore, for each adverse event, we test if the probabilities of the adverse event are the same between the two groups by assuming that the numbers of reported cases are counted from two Binomial distributions. For each i, posterior probability of H_{0i} calculated from Local and Threshold priors are larger than 0.5. Hence, with the cutoff of 0.5, we can conclude that varicella is not associated with any of the 40 adverse events. Also, the Threshold prior method reports a larger posterior probability of H_{0i} for all 40 tests than the Local prior.

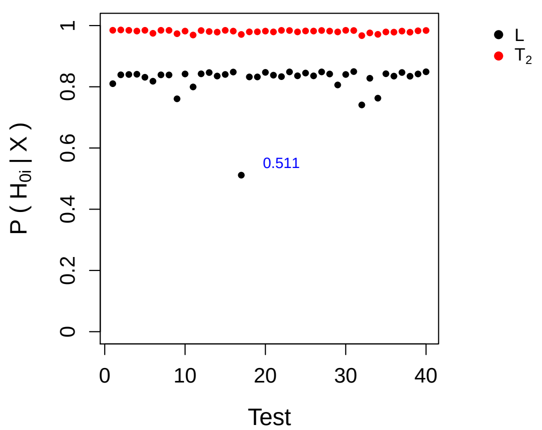


Figure 3.29.: $P(H_{0i}|X)$ for 40 tests under Local (black) and Threshold prior (red).

3.5. Conclusion

The literature on the Bayesian MT approach has focused chiefly on continuous data; a relatively small formal objective Bayesian approach is yet available in the literature for testing discrete data, especially on testing binomial proportions. Gecilli [20] proposed a formal objective Bayesian method to test the equality of two binomial proportions. However, this method has a bit of asymmetry since the two proportions are not exchangeable. In our proposed work, we consider that the two proportions are symmetric so that they are exchangeable. We propose two formal objective Bayesian approaches for testing equality of two binomial proportions based on the choice of prior under the alternative hypothesis, the Local prior approach and the Threshold(Non-local) prior approach.

This proposed work has two novel contributions; introducing a formal objective Bayesian method to test the equality of two binomial proportions when the two proportions are exchangeable and developing a Threshold Prior approach for MT of equality of two binomial proportions.

In a single testing problem, under certain regularity conditions, under a Local prior, while for a true alternative hypothesis, the Bayes factor in favor of H_0 decreases exponentially fast, for a true null hypothesis, the Bayes factor in favor of H_1 decreases only at a rate of $O(n^{-1/2})$. With the proposed Threshold prior approach, we prove that this convergence rate for a true null hypothesis can improve to $O(n^{-1})$.

Under both single and multiple testing, using repeated simulations, we show that under the null, while Threshold prior has faster convergence toward the true null than Local prior for true alternative, both Local and Threshold priors equally faster convergence towards the true alternative.

In the MT problem, for the true null hypothesis, if the true proportions are far apart Threshold prior approach always gives the higher posterior probability of H_{0i} than the Local prior approach; if the separation of true proportions is moderate, for some observations Threshold prior gives some larger posterior probabilities of H_{0i} and for some other observations little lower posterior probabilities of H_{0i} than Local prior. However, in general, Threshold prior gives a higher average posterior probability of H_{0i} (i.e. $Avg\, P(H_0|X)$) than the Local prior. When considering true alternative hypotheses in the MT problem, Threshold prior gives either similar or a little higher posterior probabilities of H_{0i} than Local prior. The posterior probabilities of the null hypotheses are impacted by the number of true alternative hypotheses and when the sample size increased such impact is diminished.

Chapter 4

Order-restricted tests for binomial proportions

4.1. Introduction

Simultaneous testing of multiple null hypotheses comparing binary response rates against order-restricted alternatives is often encountered in various research areas, including pharmaceutical research. While a significant number of frequentist work has been done for testing discrete data, there is little literature on Bayesian. Especially for multiple testing (MT) with order-restricted binomial proportions, there is only one paper available proposed by Sarkar and Chen [30]. They have proposed a Bayesian step-down approach to simultaneous testing of multiple points null hypotheses against one-sided alternatives where k treatments are compared with a control group in terms of some binary response rates. However, this approach uses approximations by converting the problem of testing binary proportions to the testing problem of normal means applying arcsin transformation, and no formal Bayesian approach is yet available.

Traditionally, a prior density under the alternative hypothesis is centered and is a maximum at the test value, indicating that the testing parameter's most likely value is its null value under the alternative hypothesis [7, 9, 13, 14]. These priors are called Local priors. Non-local priors are another type of priors that can be used under the alternative hypothesis to model the unknown parameters whose density goes to zero near the null value of the parameter under the alternative hypothesis [24, 27]. In this chapter, we consider adopting Local and Non-local priors for Bayesian testing of order-restricted binomial proportions.

This chapter extends the formal Bayesian approach discussed in chapter 3 to test the equality of binomial proportions to test order-restricted binomial proportions. We present this chapter in two parts; while in the first part, we consider only two binomial proportions and perform one-sided testing in the second part, we generalize our method for testing two or more ordered binomial proportions. This chapter is organized as follows.

First, we consider a single one-sided testing problem of two binomial proportions and discuss details of adopting our two prior choices, Local and Threshold priors. The Threshold prior we use in this chapter is a Non-local prior. We compare the convergence rates of the Bayes Factor under Local and Threshold prior approaches and find that Threshold prior provides a faster convergence rate in favoring true null hypothesis than Local prior in single testing. Next, we extend our methods for multiple testing of one-sided alternatives and provide synthetic and real data examples.

In the second part, we generalize our Local prior approach to accommodate testing two or more order-restricted binomial proportions. We give several actual data examples and compare our results with other methods in the literature.

As mentioned before, no formal Bayesian method is yet available for testing ordered binomial proportions in the literature. By doing the work presented in this chapter, we expect to provide a novel formal Bayesian approach for testing order-restricted binomial proportions.

4.2. One-sided Testing of Two Binomial Proportions

4.2.1. Single Testing

Consider testing the hypotheses

$$H_0 : p_1 = p_2 \quad vs \quad H_1 : p_1 < p_2 \tag{4.2.1}$$

based on observed data x_j independently from $Bin(n_j, p_j)$, where n_j is the known sample size, and p_j is the unknown binomial proportion for two groups $j = 1, 2$.

We adopt two Bayesian models based on the prior choice for the unknown binomial proportions p_j under the alternative hypothesis (Local and Threshold priors) and we assume that

- the prior probability of the null hypothesis (p) been true is 0.5
- under H_0, p_1 and p_2 equal to a common value p_0

Local Prior

In the literature for Bayesian MT, among various conditions required to satisfy for an objective prior under the alternative hypothesis, two some what related conditions are prior under the alternative hypothesis should concentrate around the null value, and the range of $p_j|p_0$ should be large enough to allow p_j to move away from p_0 [9, 7, 14, 16, 26]. Beta priors specified in 4.2.3 satisfy these conditions by concentrating their densities around the mode, p_0, and decreasing away from the p_0. In these Beta prior, r acts as the conditional variance of $p_j|p_0$. As r increases, the value of p_j concentrates more around p_0, and as r decreases, the value of p_j falls away from p_0. Gecili [20] has discussed details about the characteristics of mode-based beta priors.

For Local prior approach we define the model as below.

$$\text{Under } H_0 : p_1 = p_2 = p_0 \text{ where } p_0 \sim U(0,1) \tag{4.2.2}$$

$$\text{Under } H_1 : p_j^*|p_0, r \overset{iid}{\sim} Beta(rp_0 + 1, r(1-p_0) + 1) \text{ for } j = 1, 2 \tag{4.2.3}$$
$$(p_1, p_2) = (p_{(1)}^*, p_{(2)}^*) \text{ ordered smallest to largest}$$

Let $\pi(p_1, p_2|p_0, r)$ denotes the prior for (p_1, p_2) given p_0 and r, which has the form defined in 4.2.3 above. In order to complete the prior specification under the Local prior approach, we further assume that $p_0 \sim U(0,1), w \sim exp(1)$, and $w = 1/r$.

For MCMC calculation purposes, we modify the specification 4.2.2 and assume that the model will have two different proportions under the null hypotheses. Therefore, under H_0, we now assume that the model has the specification defined in 4.2.4.

$$\begin{aligned} \text{Under } H_0 : p_j|p_0, r_0 &\overset{iid}{\sim} Beta(r_0 p_0 + 1, r_0(1-p_0) + 1) \\ p_0 &\sim U(0,1) \\ w_0 &= 1/r_0 = 1/(50\sqrt{n_j}) \end{aligned} \tag{4.2.4}$$

The joint distributions of (p_1, p_2) under H_0 and H_1 have forms 4.2.5 and 4.2.6 below.

$$\pi_0(p_1, p_2) = \int_0^1 \prod_{j=1}^{2} \pi(p_j|p_0, w_0) dp_0 \tag{4.2.5}$$

$$\pi_1(p_1, p_2) = \int_0^1 \int_0^\infty \pi(p_1, p_2|p_0, w)\, \pi(w)\, dw\, dp_0 \tag{4.2.6}$$

Threshold Prior

Although most of the time in Bayesian literature, Local priors have been using to model the uncertainty of the unknown parameter under the alternative hypothesis, Johnson and Rossell [24] show that under a valid null hypothesis, data not only strongly support for the true null but also strongly support for the alternative hypothesis. Therefore, in single testing under certain regularity conditions, while for a true alternative hypothesis, the Bayes factor in favor of the null hypothesis decreases exponentially; for a true null hypothesis, the Bayes factor favoring the alternative hypothesis decreases only at rate $O(n^{-1/2})$.

Using specific Non-local priors, Johnson and Rossell [24] show that the above discrepancy of convergence rates of the Bayes factors can be improved. So that, Non-local priors show higher power than Local priors in single testing. We now consider using a Threshold prior (Non-local) for testing one-sided binomial proportions and comparing the convergence rates of the Bayes factors with those of the above proposed Local prior.

For Threshold prior we define the model as below.

$$p_j^*|p_0, r \overset{iid}{\sim} Beta(rp_0 + 1, r(1 - p_0) + 1) \tag{4.2.7}$$

$$(\tilde{p}_1, \tilde{p}_2) = (p^*_{(1)}, p^*_{(2)}) \text{ ordered smallest to largest}$$

Given p_0 and r, let $\pi(\tilde{p}_1, \tilde{p}_2|p_0, r)$ denotes the prior for $(\tilde{p}_1, \tilde{p}_2)$ defined in 4.2.7, where $p_0 \sim U(0, 1), w \sim exp(1)$, and $w = 1/r$. Then define

$$p_1 = \tilde{p}_1 \;\text{ and }\; p_2 = \begin{cases} p_1 & \text{if } 0 < LOR < \varepsilon \\ \tilde{p}_2 & \text{otherwise} \end{cases} \tag{4.2.8}$$

Here $LOR = log\left[\dfrac{\tilde{p}_2/(1-\tilde{p}_2)}{p_1/(1-p_1)}\right]$, $\varepsilon \sim U(0, K)$, and K is a known value to be determined.

Motivation to use the Uniform prior $U(0, K)$ for the threshold parameter ε is from Nakajima and West [27]. They have recommended using a Uniform prior as the natural default in the absence of a context that involves substantive information to suggest the degrees of expected data sparsity.

Let $\pi(p_1, p_2|p_0, w)$ be the prior for (p_1, p_2) given p_0 and w, then the prior probability of H_0, p, can be written as a function of K denoted by $g(K)$.

$$\begin{aligned} g(K) &= P(0 < LOR < \varepsilon) \\ &= \int \int_0^K I\big(LOR < \varepsilon\big)\, \pi(\varepsilon)\, d\varepsilon\ \pi(p_1, p_2|p_0, w)\, \pi(w)\, dw\, dp_0\, dp_1\, dp_2 \\ &= \int_0^1 \int_0^1 \int_0^1 \int_0^\infty \int_{LOR}^K \frac{1}{K}\, d\varepsilon\ \pi(p_1, p_2|p_0, w)\, \pi(w)\, dw\, dp_0\, dp_1\, dp_2 \\ &= \int_0^1 \int_0^1 \int_0^1 \int_0^\infty \left[1 - \frac{LOR}{K}\right]\ \pi(p_1, p_2|p_0, w)\, \pi(w)\, dw\, dp_0\, dp_1\, dp_2 \qquad (4.2.9) \end{aligned}$$

To give equal priority to both null and alternative hypotheses like in the Local prior, we pick the value $K = 2.365$ so that $g(K) = 0.5$.

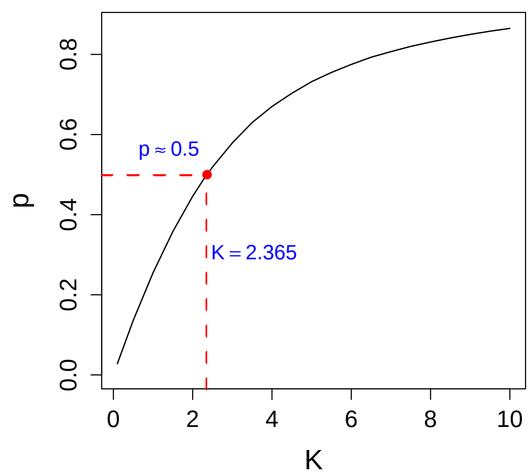


Figure 4.1.: Prior probability of H_0, p, as a function of K

Then, under the Threshold prior method, the joint prior for (p_1, p_2) under H_1 can be written as

$$\pi_1(p_1,p_2) = \frac{1}{(1-p)}\Big[\int\int_0^K I\big(\varepsilon < LOR < K\big)\,\pi(\varepsilon)\,d\varepsilon\;\pi(p_1,p_2|p_0,w)\,\pi(w)\,dw\,dp_0 + \int I\big(LOR > K\big)\;\pi(p_1,p_2|p_0,w)\,\pi(w)\,dw\,dp_0\Big]$$

$$= \frac{1}{(1-p)}\Big[\int I\big(LOR < K\big)\Big(\frac{LOR}{K}\Big)\;\pi(p_1,p_2|p_0,w)\,\pi(w)\,dw\,dp_0 + \int I\big(LOR > K\big)\;\pi(p_1,p_2|p_0,w)\,\pi(w)\,dw\,dp_0\Big] \tag{4.2.10}$$

Posterior Probability of H_0

Finally, given the data $\boldsymbol{X} = (x_1, x_2)$, the posterior probability of H_0 under the Local and Threshold priors are given by equations

For Local Prior:

$$P(H_0|\boldsymbol{X}) = \frac{\int_0^1\int_0^1 \prod_{j=1}^{2} f(x_j|p_j)\pi_0(p_1,p_2)\,dp_1\,dp_2}{\int_0^1\int_0^1 \prod_{j=1}^{2} f(x_j|p_j)\Big[\pi_0(p_1,p_2) + \pi_1(p_1,p_2)\Big]dp_1\,dp_2} \tag{4.2.11}$$

where $\pi_0(p_1, p_2)$ and $\pi_1(p_1, p_2)$ are given by equations 4.2.5 and 4.2.6.

For Threshold Prior:

$$P(H_0|\boldsymbol{X}) = \frac{\int_0^1 \prod_{j=1}^{2} f(x_j|p_0)\,dp_0}{\int_0^1 \prod_{j=1}^{2} f(x_j|p_0)\,dp_0 + \int_0^1\int_0^1 \prod_{j=1}^{2} f(x_j|p_j)\pi_1(p_1,p_2)\,dp_1\,dp_2}. \tag{4.2.12}$$

where $\pi_1(p_1, p_2)$ is given by the equation 4.2.10.

Simulation

Even though we can achieve the posterior computations for single testing by direct integration, to extend the single testing to multiple testing in the coming section, we use MCMC techniques for posterior calculations.

Let $n_1 = n_2 = n$, $p_1 = p_2 = 0.5$ (under true null), and $p_1 = 0.3, p_2 = 0.7$ (under true alternative). For a given n, for a true null, generate data from $x_1, x_2 \sim Bin(n, 0.5)$ independently. For a true alternative hypotheses, generate data from $x_1 \sim Bin(n, 0.3), x_2 \sim Bin(n, 0.7)$, for a given n. Then calculate the posterior probability of H_0 and log Bayes factor (log BF) under Local and Threshold priors. Finally, replicate 1000 simulations and calculate the average posterior probability of H_0, average log Bayes factor over the data, and 95% confidence intervals for simulated data. For these simulations we use R-jags software.

- Average posterior probability of H_0 over x under hypothesis H_i

  $Avg_x\, P(H_0|X) = \Sigma_x\, P(H_0|X) \cdot P(X = x|H_i)$

- Average log Bayes factor in favour of H_1

  $Avg_x\, log_{10}[BF(1|0)] = \Sigma_x\, log_{10}[BF(1|0)] \cdot P(X = x|H_i)$

where $P(X = x|H_i) = Bin(x|n, p_i)$ and $i = 0, 1$.

Below are some of the results illustrating the convergence rate of the Local and Threshold priors under true null and alternative hypotheses as the sample size(n) increases.

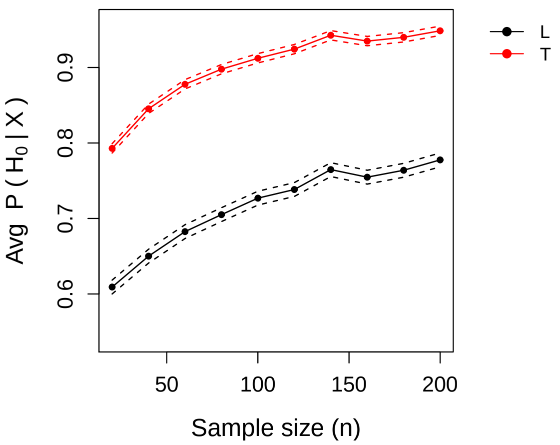
Result and Conclusion

When comparing the convergence of Local and Threshold priors (Fig.4.2-Fig.4.3), under the true null hypothesis, a faster convergence rate of the Threshold prior than the Local prior is visible. For instance, in figure 4.2 (a), under the true null hypothesis, average log

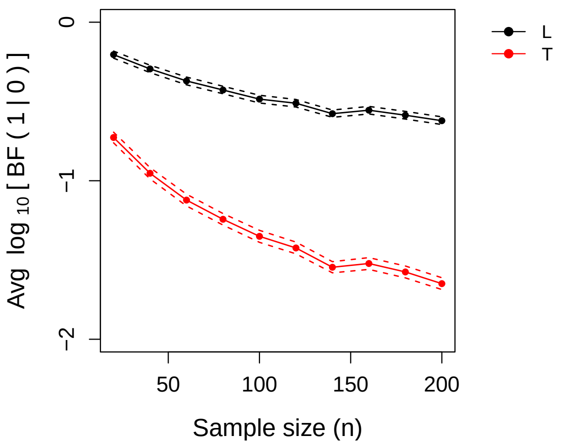
BF favoring H_1 from the Threshold prior is always less than -0.5. After about $n = 40$, average log BF is between -1 and -2, showing strong evidence against the alternative hypothesis. But for the Local prior, even with a larger sample size, average log BF in favor of H_1 doesn't deccrease beyond -0.5, which means insufficient evidence against H_1 for true null under Local prior.

For a true alternative, average log BF in favor of H_0 decreases faster as n increases for both Local and Threshold priors. For example, in figure 4.3 (b), when about $n > 20$, average log BF in favor of H_0 is less than -2 for both priors, which gives strong evidence against the null hypothesis.

Given figures 4.2 and 4.3, we see that for a true null hypothesis, under local prior, average log BF in favor of the alternative hypothesis decreases at a low rate and with the proposed Threshold prior, we can improve this convergence rate for single testing.



(a)



(b)

Figure 4.2.: (a) Average posterior probability of H_0 (b) average log Bayes factor in favour of H_1, as the sample size(n) increases under Local prior (black) and Threshold prior with $K = 2.336$ (red) when $x_{1j}, x_{2j} \sim Bin(n, 0.5)$ for $j = 1, \cdots, 1000$ number of replicates. Dash lines represent 95% CIs.

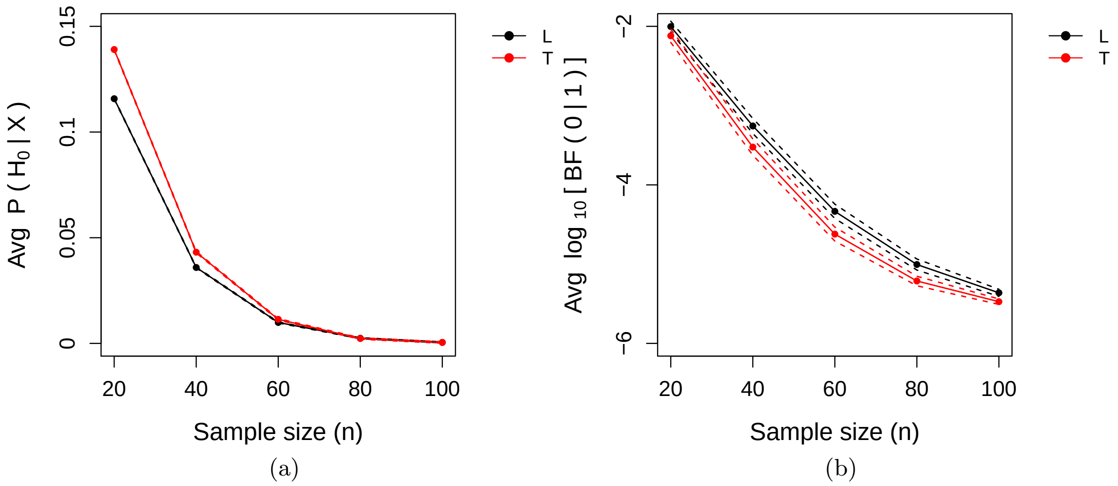


Figure 4.3.: (a) Average posterior probability of H_0 (b) average log Bayes factor in favour of H_0, as sample size (n) increases, under Local prior (black) and Threshold prior with $K = 2.336$ (red) when $x_{1j} \sim Bin(n, 0.3), x_{2j} \sim Bin(n, 0.7)$ for $j = 1, \cdots, 1000$ number of replicates. Dash lines represent 95% CIs.

4.2.2. Multiple Testing

We now consider MT of equality of two binomial proportions against one-sided alternatives. For a given test i, we assume the data for two groups comes from Binomial distributions, $x_{ij} \sim Bin(n_{ij}, p_{ij})$, with known sample size n_{ij} and unknown proportion p_{ij} for $i = 1, 2, \cdots, M$ and $j = 1, 2$. For each test i, we want to test the hypotheses

$$H_{0i} : p_{i1} = p_{i2} \quad vs \quad H_{1i} : p_{i1} < p_{i2}. \tag{4.2.13}$$

Assuming that all the hypotheses are exchangeable, we use a common p as the prior probability of H_{0i} for each test i and has $U(0, 1)$ distribution. To model the uncertainly of the unknown proportions, p_{ij}, we modify our proposed Local and Threshold prior approaches defined in section 4.2.1 as follows.

Local Prior

For each test i, under H_{0i}, we define

$$p_{ij}|p_{i0}, r_0 \stackrel{iid}{\sim} Beta(r_0 p_{i0} + 1, r_0(1 - p_{i0}) + 1) \tag{4.2.14}$$
$$p_{i0}|p_{00}, r_{00} \sim Beta(r_{00}p_{00} + 1, r_{00}(1 - p_{00}) + 1)$$
$$w_0 = 1/r_0 = 1/(200\sqrt{n_{ij}})$$
$$p_{00} \sim U(0, 1)$$
$$w_{00} \sim exp(1) \text{ and } w_{00} = 1/r_{00}$$

and under H_{1i} define

$$p^*_{ij}|p_{i0}, r \stackrel{iid}{\sim} Beta(rp_{i0} + 1, r(1 - p_{i0}) + 1) \tag{4.2.15}$$
$$(p_{i1}, p_{i2}) = (p^*_{(i1)}, p^*_{(2)}) \text{ ordered smallest to largest}$$

For each i, let $\pi(p_{i1}, p_{i2}|p_{i0}, r)$ denotes the prior for (p_{i1}, p_{i2}) given p_{i0} and r defined in 4.2.15. Here, p_{i0} and r have prior distributions given below in 4.2.16.

$$p_{i0}|p_{00}, r_{00} \sim Beta(r_{00}p_{00} + 1, r_{00}(1 - p_{00}) + 1) \tag{4.2.16}$$
$$w \sim exp(1) \text{ and } w = 1/r$$
$$p_{00} \sim U(0, 1)$$
$$w_{00} \sim exp(1) \text{ and } w_{00} = 1/r_{00}$$

Let $\pi_{0i}(p_{i1}, p_{i2})$ and $\pi_{1i}(p_{i1}, p_{i2})$ be the joint priors of (p_{i1}, p_{i2}) under H_{0i} and H_{1i} respectively. Now, by writting $\pi(\Theta_3) = \pi(p_{i0}|p_{00}, w_{00})\,\pi(w_{00})\,\pi(p_{00})$ and $\pi(\Theta_4) = \pi(p_{i0}|p_{00}, w_{00})\,\pi(w)\,\pi(w_{00})\,\pi(p_{00})$, where $\Theta_3 = (w_{00}, p_{00}, p_{i0})$ and $\Theta_4 = (w_{00}, p_{00}, w, p_{i0})$,

$\pi_{0i}(p_{i1}, p_{i2})$ and $\pi_{1i}(p_{i1}, p_{i2})$ has the forms given by equations 4.2.17 and 4.2.18.

$$\pi_{0i}(p_{i1}, p_{i2}) = \int_0^1 \int_0^1 \int_0^\infty \prod_{j=1}^{2} \pi(p_{ij}|p_{i0}, w_0)\, \pi(\Theta_3)\, dw_{00}\, dp_{00}\, dp_{i0} \tag{4.2.17}$$

$$\pi_{1i}(p_{i1}, p_{i2}) = \int_0^1 \int_0^\infty \int_0^1 \int_0^\infty \pi(p_{i1}, p_{i2}|p_{i0}, w)\pi(\Theta_4)\, dw_{00}\, dp_{00}\, dw\, dp_{i0} \tag{4.2.18}$$

Threshold Prior

Under the Threshold prior approach, for each i, define

$$p^*_{ij}|p_{i0}, r \overset{iid}{\sim} Beta(rp_{i0} + 1, r(1 - p_{i0}) + 1) \tag{4.2.19}$$
$$(\tilde{p}_{i1}, \tilde{p}_{i2}) = (p^*_{(i1)}, p^*_{(i2)}) \text{ ordered smallest to largest}$$

Now given p_{i0} and r, let $\pi(\tilde{p}_{i1}, \tilde{p}_{i2}|p_{i0}, r)$ denotes the prior for $(\tilde{p}_{i1}, \tilde{p}_{i2})$ defined in 4.2.19, where p_{i0} and r are defined as under 4.2.21.

Then define

$$p_{i1} = \tilde{p}_{i1} \quad \text{and} \quad p_{i2} = \begin{cases} p_{i1} & \text{if } 0 < LOR_i < \varepsilon \\ \tilde{p}_{i2} & \text{otherwise} \end{cases} \tag{4.2.20}$$

Here $LOR_i = log\left[\frac{\tilde{p}_{i2}/(1 - \tilde{p}_{i2})}{p_{i1}/(1 - p_{i1})}\right]$, $\varepsilon \sim U(0, K)$, and K is to be determined.

Prior distributions for other hyperparameters are

$$\begin{aligned} p_{i0}|p_{00}, r_{00} &\sim Beta(r_{00}p_{00} + 1, r_{00}(1 - p_{00}) + 1) \qquad (4.2.21) \\ w &\sim exp(1) \text{ and } w = 1/r \\ p_{00} &\sim U(0, 1) \\ w_{00} &\sim exp(1) \text{ and } w_{00} = 1/r_{00} \end{aligned}$$

Given the above prior specifications, we can modify equation 4.2.9 and write p as a deterministic function of K, which has the form

$$g(K) = \int \left[1 - \frac{LOR_i}{K}\right] I\big(0 < LOR_i < K\big)\pi(p_{i1}, p_{i2}|p_{i0}, r)\pi(\Theta_4)\,\Theta_4\, dp_{i1}\, dp_{i2}$$

where $\pi(\Theta_4) = \pi(p_{i0}|p_{00}, w_{00})\pi(w)\pi(w_{00})\,\pi(p_{00})$ is the prior density of $\Theta_4 = (w_{00}, p_{00}, w, p_{i0})$.

Choose a prior for K

Now we need to find a prior for K such that, $p \sim U(0, 1)$. From figure 4.4(a) we know that when $K = 0$, $p = 0$; $K = 2.365$, $p \approxeq 0.5$; and $K \to \infty$, $p \to 1$. Based on this information, we will find an exponential function that fits "best" to the curve p vs K in Figure 4.4 (a). Among the few different functions we tried, $p = 1 - e^{-aK}$ with $a = 0.3$ is the best fit for curve p vs. K in figure 3. Knowing $a = 0.3$, we use a variable transformation with $p \sim U(0, 1)$ to get the distribution for K and finally get $K \sim Exp(0.3)$.

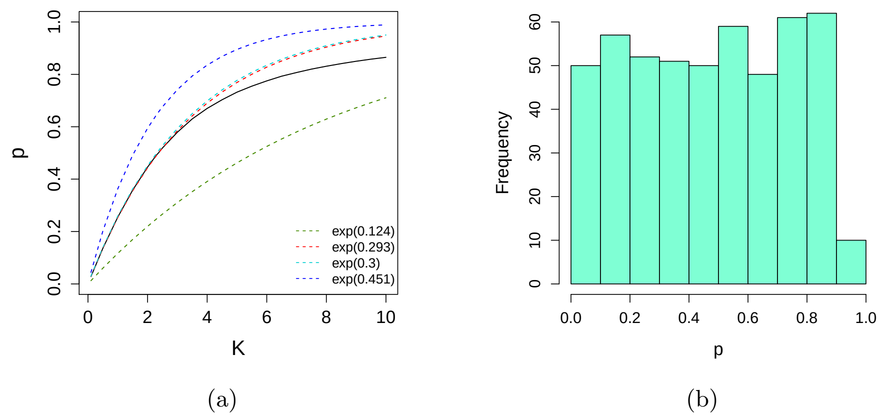


Figure 4.4.: (a) Different exponential fuctions to select best fit to p vs. K curve (b) Histogram of the prior probability of H_0 at 500 random values of K generated from $exp(0.3)$.

By assigning the prior $Exp(0.3)$ for K and integrating the above expression of $g(K)$ over K, we get the expression for p as

$$\begin{aligned} p &= \int_0^\infty g(K) \cdot \pi(K)\, dK \\ &= \int_{LOR_i}^\infty \int \left[1 - \frac{LOR_i}{K}\right] \pi(p_{i1}, p_{i2}|p_{i0}, w)\pi(\Theta_4)\, d\Theta_4\, p_{i1}\, dp_{i2} \cdot \pi(K)\, dK \end{aligned} \tag{4.2.22}$$

Under the Threshold prior approach, by modifying 4.2.10, the joint distribution of (p_{i1}, p_{i2}) under H_{1i} for MT can be written as

$$\begin{aligned} \pi_{1i}(p_{i1}, p_{i2}) &= \frac{1}{(1-p)}\Bigg[\int_0^\infty \int I\big(LOR_i > K\big)\, \pi(p_{i1}, p_{i2}|p_{i0}, w)\, \pi(\Theta_4)\, d\Theta_4 \cdot \pi(K)\, dK \\ &\quad + \int_0^\infty \int I\big(LOR_i < K\big) \left(\frac{LOR_i}{K}\right) \pi(p_{i1}, p_{i2}|p_{i0}, w)\, \pi(\Theta_4) d\Theta_4 \cdot \pi(K) dK\Bigg] \\ &= \frac{1}{(1-p)}\Bigg[\int_0^{LOR_i} \int \pi(p_{i1}, p_{i2}|p_{i0}, w)\, \pi(\Theta_4)\, d\Theta_4 \cdot \pi(K)\, dK \\ &\quad + \int_{LOR_i}^\infty \int \left(\frac{LOR_i}{K}\right) \pi(p_{i1}, p_{i2}|p_{i0}, w)\, \pi(\Theta_4)\, d\Theta_4 \cdot \pi(K)\, dK\Bigg] \end{aligned} \tag{4.2.23}$$

Posterior Probability of H_0

Posterior probability H_{0i} can be written as $P(H_{0i}|\boldsymbol{X}) = \dfrac{\boldsymbol{N_1} \times \boldsymbol{N_2}}{\boldsymbol{D}}$ where for Local prior

$$\boldsymbol{N_1} = \int_0^1 \int_0^1 \prod_{j=1}^{2} f(x_{ij}|p_{ij})\pi_{0i}(p_{i1}, p_{i2})\, dp_{i1}\, dp_{i2} \cdot p$$

$$\boldsymbol{N_2} = \prod_{q=1(q\neq i)}^{M} \Big[\int_0^1 \int_0^1 \prod_{j=1}^{2} f(x_{qj}|p_{qj})\, \pi_{0q}(p_{q1}, p_{q2})\, dp_{q1}\, dp_{q2} \cdot p + \int_0^1 \int_0^1 \prod_{j=1}^{2} f(x_{qj}|p_{qj})\pi_{1q}(p_{q1}, p_{q2})\, dp_{q1} dp_{q2} \cdot (1-p) \Big]$$

$$\boldsymbol{D} = \prod_{q=1}^{M} \Big[\int_0^1 \int_0^1 \prod_{j=1}^{2} f(x_{qj}|p_{q1})\, \pi_{0q}(p_{q1}, p_{q2})\, dp_{q1}\, dp_{q2} \cdot p + \int_0^1 \int_0^1 \prod_{j=1}^{2} f(x_{qj}|p_{qj})\pi_{1q}(p_{q1}, p_{q2})\, dp_{q1} dp_{q2} \cdot (1-p) \Big]$$

and for Threshold prior

$$\boldsymbol{N_1} = \int_0^1 \prod_{j=1}^{2} f(x_{ij}|p_{i0})\pi(p_{i1})\, dp_{i0} \cdot p$$

$$\boldsymbol{N_2} = \prod_{q=1(q\neq i)}^{M} \Big[\int_0^1 \prod_{j=1}^{2} f(x_{qj}|p_{q0})\pi(p_{q0})\, dp_{q0} \cdot p + \int_0^1 \int_0^1 \prod_{j=1}^{2} f(x_{qj}|p_{qj})\pi_{1q}(p_{q1}, p_{q2})\, dp_{q1} dp_{q2} \cdot (1-p) \Big]$$

$$\boldsymbol{D} = \prod_{q=1}^{M} \Big[\int_0^1 \prod_{j=1}^{2} f(x_{qj}|p_{q0})\pi(p_{q0})\, dp_{q0} \cdot p + \int_0^1 \int_0^1 \prod_{j=1}^{2} f(x_{qj}|p_{qj})\pi_{1q}(p_{q1}, p_{q2})\, dp_{q1} dp_{q2} \cdot (1-p) \Big]$$

For each i, while under Local prior $\pi_{0i}(p_{i1}, p_{i2})$ and $\pi_{1i}(p_{i1}, p_{i2})$ are given by 4.2.17 and 4.2.18; under Threshold prior $\pi_{1i}(p_{i1}, p_{i2})$ is given by 4.2.23 and $\pi(p_{i0}) \sim Beta(r_{00}p_{00} + 1, r_{00}(1-p_{00}) + 1)$.

Simulations

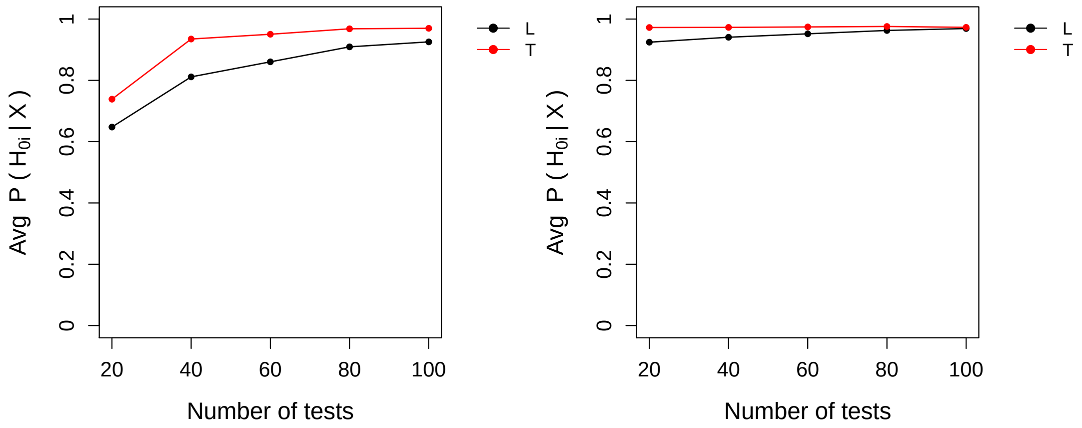
We perform simulations to study the convergence of proposed Local and Threshold priors for MT of one-sided two binomial proportions and present the results under three cases (all the tests are null true, the number of true alternatives is fixed, and the number of true alternatives is increasing with M). In all these settings, we assume two groups have equal sample sizes. We generate x_{ij} for $i = 1, \cdots, M; j = 1, 2$ from a Binomial distribution with size n and success probability p_{ij} for a given sample size n. We calculate the posterior probability of the null hypothesis for each test i using 500 thousand mcmc samples with a 50 thousand burning phase and report the average posterior probabilities of null hypotheses under different settings.

Results and Conclusions

Case 1: All null true

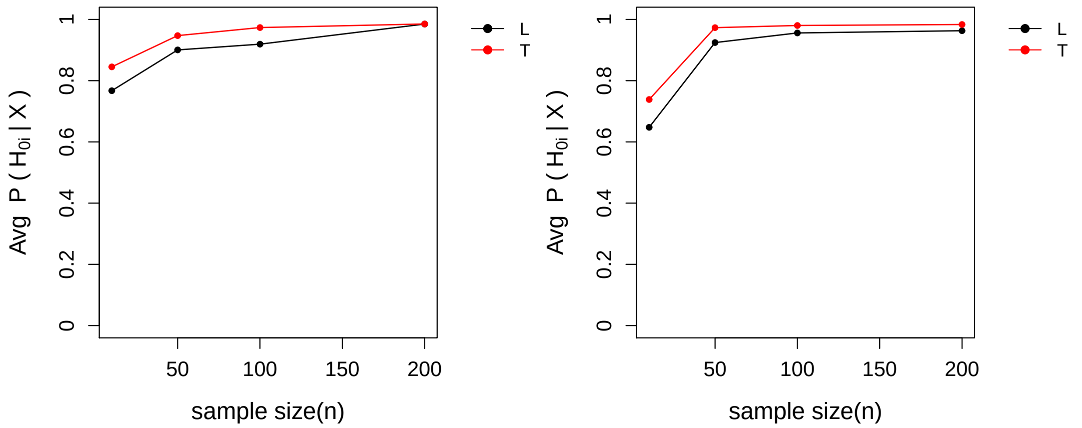
Figure 4.5 illustrates the average posterior probability of true M null hypotheses, $Avg\, P(H_{0i}|X)$, as M increases at two different sample sizes, (a) $n = 10$ and (b) $n = 50$. As M increases $Avg\, P(H_{0i}|X)$ increases for both priors. Threshold prior always has a higher $Avg\, P(H_{0i}|X)$ than Local prior. As n increases from 10 to 50, $Avg\, P(H_{0i}|X)$ increases for both Threshold and Local priors.

Figure 4.6 illustrates the average posterior probability of null hypotheses, $Avg\, P(H_{0i}|X)$, as n increases when all the tests ($M = 20$) are null true: fig.4.6 (a) $x_{i1}, x_{i2} \sim Bin(n, 0.05)$ and (b) $x_{i1}, x_{i2} \sim Bin(n, 0.2)$ independently. Threshold prior gives a higher $Avg\, P(H_{0i}|X)$ than the Local Prior. However, Local prior reports improved posterior probabilities of true nulls in multiple testing than in single testing.



(a) (b)

Figure 4.5.: Average of the posterior probabilities of H_{0i}s under Local (black) and Threshold (red) priors as the number of tesets(M) increases when all null hypotheses are true: (a) $x_{i1}, x_{i2} \sim Bin(10, 0.2)$ and (b) $x_{i1}, x_{i2} \sim Bin(50, 0.2)$.



(a) (b)

Figure 4.6.: Average of the posterior probabilities of H_{0i}s under Local (black) and Threshold (red)priors as the sample size(n) increases for $M = 20$ testes for data (a)$x_{i1}, x_{i2} \sim Bin(n, 0.05)$ (b) $x_{i1}, x_{i2} \sim Bin(n, 0.2)$.

Case 2: Mixed- number of true alternative fixed

Figure 4.7 depicts the average posterior probability of H_{0i}s, $Avg\, P(H_{0i}|X)$, of 10 true null and 10 true alternative hypotheses as sample size n increases when $M = 20$.

Figures 4.8 and 4.9 show the average posterior probability of H_{0i}s for true null and true alternative hypotheses of M tests as M increases at two different sample sizes, (a) $n = 10$ and (b) $n = 50$. Here the number of true alternative hypotheses is fixed at 10, and the number of true null hypotheses increases as M increases. For both priors, $Avg\, P(H_{0i}|X)$ increases as M increases when the number of true alternatives is fixed. Even though $Avg\, P(H_{0i}|X)$ is small for true alternatives, it also increases as M increases.

Except when the two proportions are close, and both n and M small (fig.4.7 and fig.4.8(a)), generally, Threshold prior gives higher posterior probabilities for true the null hypotheses than Local prior. Under both priors, as n increases, the average posterior probability for an actual null increases, while the mean of the posterior probabilities decreases for an actual alternative.

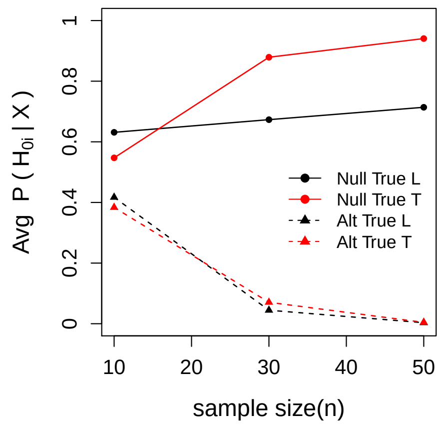
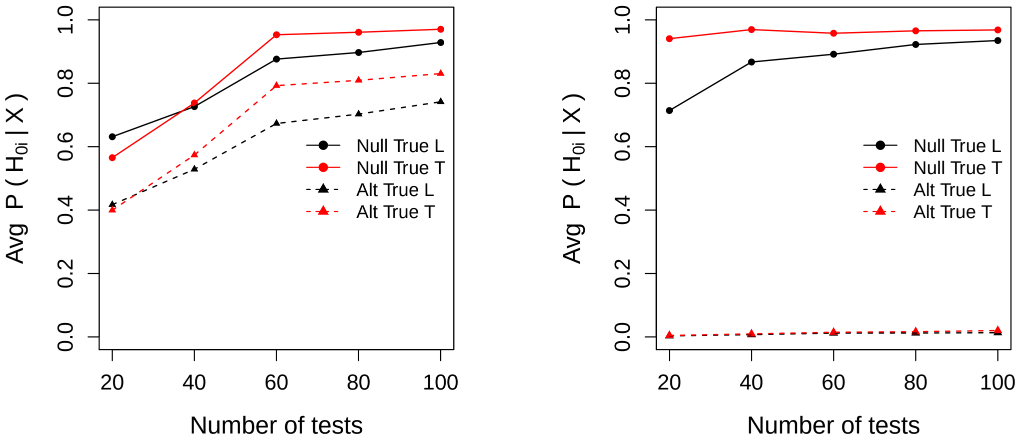
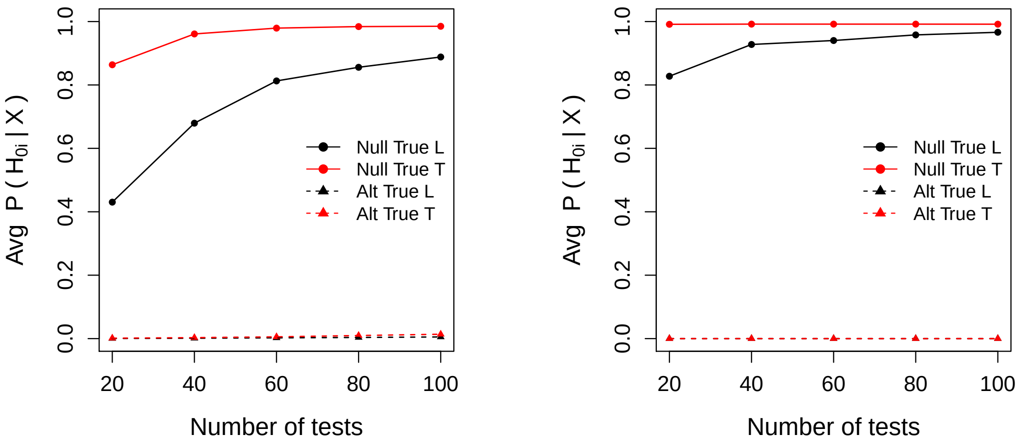


Figure 4.7.: Average posterior probability of H_{0i}s under Local (black) and Threshold (red) priors as sample size(n) increases when $M = 20$ (first 10 null true and last 10 alternative true): for true null $x_{i1}, x_{i2} \sim Bin(n, 0.2)$ (circle-solidline)and for true alternative $x_{i1} \sim Bin(n, 0.2), x_{i2} \sim Bin(n, 0.5)$ (triangle-dashline).



(a) (b)

Figure 4.8.: Average posterior probability of H_{0i}s under Local (black) and Threshold (red) priors as M increases (number of true null increases (circle) and number of true alternatives (triangle) is fixed at 10); for true null $x_{i1}, x_{i2} \sim Bin(n, 0.2)$ and for true alternative $x_{i1} \sim Bin(n, 0.2), x_{i2} \sim Bin(n, 0.5)$ where (a) $n = 10$ and (b) $n = 50$.



(a) (b)

Figure 4.9.: Average posterior probability of H_{0i}s under Local (black) and Threshold (red)priors as M increases (number of true null increases (circle) and number of true alternatives (triangle) is fixed at 10): for true null $x_{i1}, x_{i2} \sim Bin(n, 0.2)$ and for true alternative $x_{i1} \sim Bin(n, 0.2), x_{i2} \sim Bin(n, 0.9)$ where (a) $n = 10$ and (b) $n = 50$.

Case 3: Mixed- number of true alternative increases (10% to 20% of M)

Figures 4.10 and 4.11 illustrate the average of the posterior probabilities of H_{0i}, $Avg\, P(H_{0i}|X)$, for true null and true alternative hypotheses out of M tests as the number of true alternatives increases 10% to 20% of M at different combinations of sample sizes and the number of tests (a) $n = 10; M = 20$ (b) $n = 50; M = 20$ (c) $n = 10; M = 40$ (b) $n = 50; M = 40$.

With Local prior $Avg\, P(H_{0i}|X)$ of true null, and alternative hypotheses decrease as the number of true alternatives increases. With Threshold prior $Avg\, P(H_{0i}|X)$ of true alternative hypotheses decreases as the number of true alternatives increases.

When comparing the $Avg\, P(H_{0i}|X)$ as the number of tests, M, increases for a fixed sample size n (in both figures 4.10 and 4.11 compare (a) with (c) and (b) with (d)), for the true null $Avg\, P(H_{0i}|X)$ increase as M increases.

Similarly, when compare $Avg\, P(H_{0i}|X)$ as the sample size n increases for a fixed number of tests M(in both figures 4.10 and 4.11 compare (a) with (b) and (c) with (d)), for the true null $Avg\, P(H_{0i}|X)$ increases as n increases.

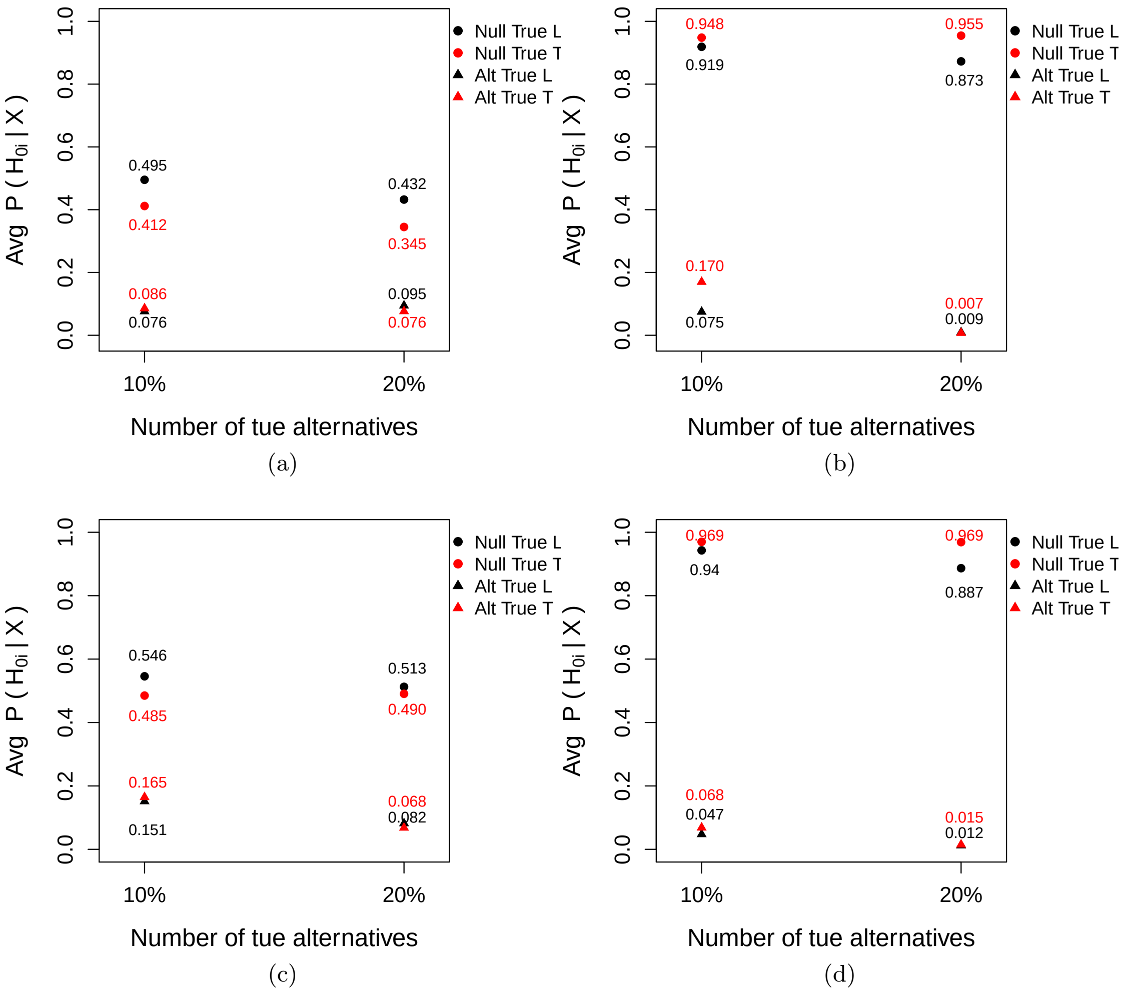


Figure 4.10.: Mean posterior probability of H_{0i}s under Local (black) and Threshold (red) priors as the number of the number of true alternatives increases from 10% to 20% of M: for true null $x_{i1}, x_{i2} \sim Bin(n, 0.2)$ and for true alternative $x_{i1} \sim Bin(n, 0.2), x_{i2} \sim Bin(n, 0.5)$ where (a) $M = 20; n = 10$ and (b) $M = 20; n = 50$ (c) $M = 40; n = 10$ (d) $M = 40; n = 50$.

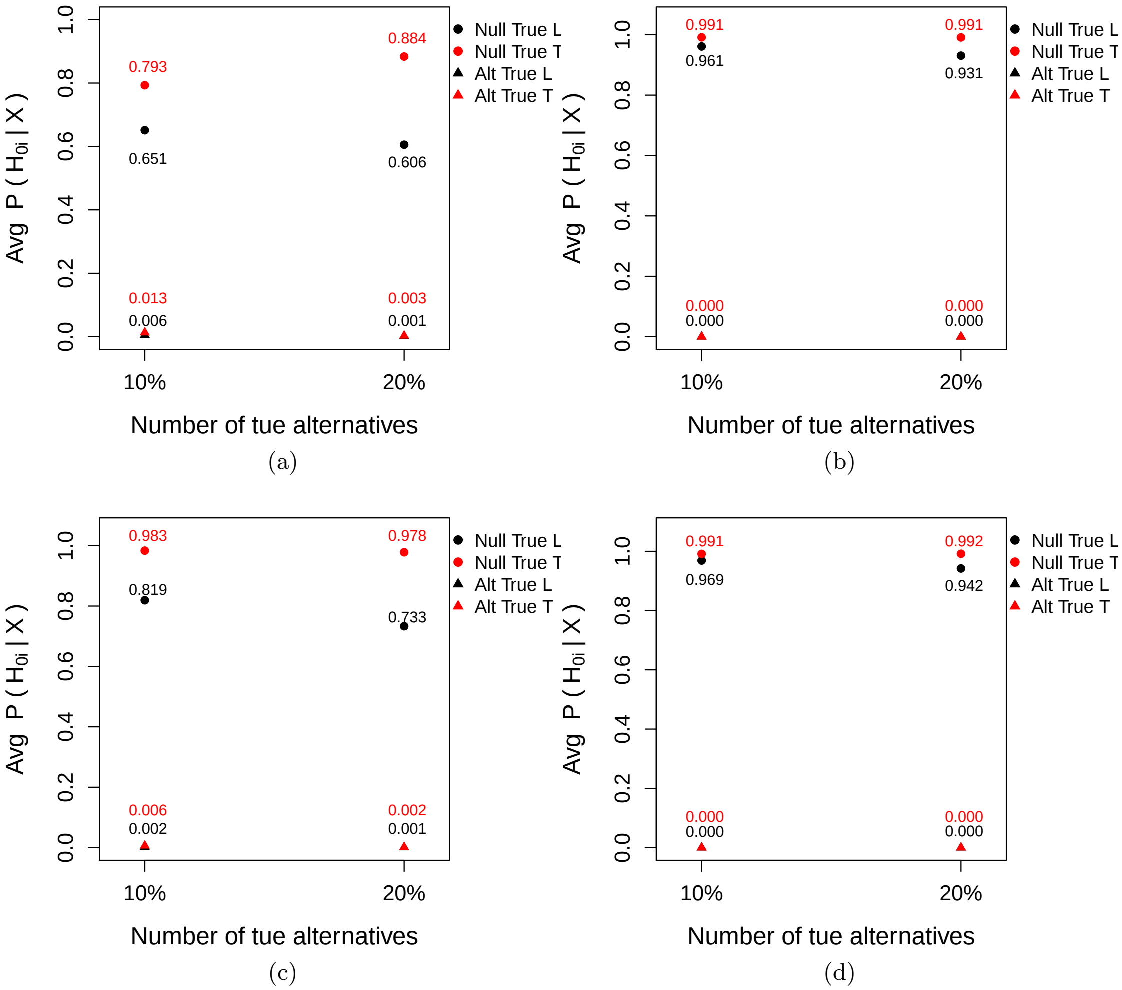


Figure 4.11.: Mean posterior probability of H_{0i}s under Local (black) and Threshold (red)priors as the number of the number of true alternatives increases from 10% to 20% of M: for true null $x_{i1}, x_{i2} \sim Bin(n, 0.2)$ and for true alternative $x_{i1} \sim Bin(n, 0.2), x_{i2} \sim Bin(n, 0.9)$ where (a) $M = 20; n = 10$ and (b) $M = 20; n = 50$ (c) $M = 40; n = 10$ (d) $M = 40; n = 50$.

4.3. Order-Restricted Testing of Q Binomial Proportions

This section aims to generalize the above proposed Local prior method for one-sided testing of two binomial proportions in section 4.2 to testing ordered restricted $Q(\geq 2)$ binomial proportions. First, consider single testing of equality of binomial proportions in Q groups and later extend the approach to MT.

4.3.1. Single Testing of Q Order-Restricted Binomial Proportions

Suppose for each group $j = 1, \cdots, Q$ we observe x_j from $Bin(n_j, p_j)$ independently; n_js may be the same or different. Our interest is testing the hypotheses,

$$H_0 : p_1 = p_2 = \cdots = p_Q \quad vs \quad H_1 : p_1 < p_2 < \cdots < p_Q. \tag{4.3.1}$$

We generalize the model proposed in section 4.2.1 for single one-sided testing of two proportions to test Q order-restricted proportions by simply changing the number of proportions from two to $Q(\geq 2)$ as follows. Note that in testing Q order-restricted proportions, we assume p_0 has different values under H_0 and H_1 in 4.3.2 and 4.3.3.

$$\begin{aligned} \text{Under } H_0 : p_1, \cdots, p_Q | p_0, r_0 &\overset{iid}{\sim} Beta(r_0 p_0 + 1, r_0(1 - p_0) + 1) \\ p_0 &\sim U(0, 1) \\ w_0 &= 1/r_0 = 1/(50\sqrt{n_j}) \end{aligned} \tag{4.3.2}$$

$$\begin{aligned} \text{Under } H_1 : p_1^*, \cdots, p_Q^* | p_0, r &\overset{iid}{\sim} Beta(r p_0 + 1, r(1 - p_0) + 1) \\ (p_1, \cdots, p_Q) &= (p_{(1)}^*, \cdots, p_{(Q)}^*) \text{ ordered smallest to largest} \end{aligned} \tag{4.3.3}$$

Let $\pi(p_1, \cdots, p_Q | p_0, r)$ be the prior for $(p_1, \cdots, p_Q)$ given p_0 and r which has the form 4.3.3. Further, under H_1 we assume that $p_0 \sim U(0, 1)$, $w \sim exp(1)$, and $w = 1/r$.

Then the joint prior for $(p_1, \cdots, p_Q)$ under H_0 and H_1 have the forms 4.3.4 and 4.3.5 respectively.

$$\pi_0(p_1, \cdots, p_Q) = \int_0^1 \prod_{j=1}^{Q} \pi(p_j|p_0, w_0) dp_0 \tag{4.3.4}$$

$$\pi_1(p_1, \cdots, p_Q) = \int_0^1 \int_0^\infty \pi(p_1, \cdots, p_Q|p_0, w)\, \pi(w)\, dw\, dp_0 \tag{4.3.5}$$

Posterior Probability of H_0

Let p be the prior probability of the null hypothesis and equals 0.5. Then given the data $\boldsymbol{X} = \{x_1, \cdots, x_Q\}$, the posterior probability of the null hypothesis can be written as

$$P(H_0|\boldsymbol{X}) = \frac{\int_0^1 \cdots \int_0^1 \prod_{j=1}^{Q} f(x_j|p_j) \pi_0(p_1, \cdots, p_Q)\, dp_1 \cdots dp_Q}{\int_0^1 \cdots \int_0^1 \prod_{j=1}^{Q} f(x_j|p_j) \Big[\pi_0(p_1, \cdots, p_Q) + \pi_1(p_1, \cdots, p_Q)\Big] dp_1 \cdots dp_Q} \tag{4.3.6}$$

Simulation, Result, and Conclusion

We illustrate the specified formal Bayesian hierarchical model using R-jags. We consider several simulated datasets and calculate the posterior probability of H_0 using equation 4.3.6. Table 4.1 displays the results for several simulated datasets at different values of $Q(= 3, 5)$ as the sample size increases $(n = 10, 20, 50, 100, 200)$ under true null and alternative hypotheses.

Table 4.1 results show that as the sample size and the number of binomial proportions increase, the posterior probability of the true null hypothesis increases. Under the true alternative, when the sample proportions are fixed, the posterior probability of the null hypothesis increases as the sample size increases. When the sample proportions vary with the sample size, then the posterior probability of the null hypothesis goes to zero fast by favoring the true alternative. Based on the results in table 4.1, we can conclude that our proposed formal Bayesian approach is working fine, and it is worth expanding to multiple testing.

Q	X	$P(H_0 \mid X)$				
		$n = 10$	$n = 20$	$n = 50$	$n = 100$	$n = 200$
			Null True			
3	5,5,5	0.691	0.792	0.911	0.944	0.995
	0.5n,0.5n,0.5n	0.704	0.7861	0.872	0.917	0.941
5	5,5,$\cdots$,5	0.859	0.930	0.963	0.999	0.999
	0.5n,$\cdots$,0.5n	0.860	0.923	0.973	0.980	0.998
			Alternative True			
3	1,5,10	4.437e-04	0.015	0.097	0.269	0.508
	0.1n,0.5,0.9n	2.848e-03	3.630e-06	0	0	0
5	1,3,5,7,9	3.838e-04	0.015	0.120	0.286	0.658
	0.1n,0.3n, $\cdots$,0.9n	5.690e-04	0	0	0	0

Table 4.1.: The posterior probability of H_0 for different data, $\boldsymbol{X}$, at different values of Q and n under true null and alternative hypotheses calculated from 700 thousand mcmc with 150 thousand burning phase.

4.3.2. Multiple Testing of Q Order-Restricted Binomial Proportions

Suppose we want to simultaneously test the equality of $Q(\geq 2)$ binomial proportions stratified by M levels of a control variable. Each x_{ij} is observed independently from $Bin(n_{ij}, p_{ij})$; n_{ij}s may be all the same or different, here $i = 1, 2, \cdots, M$ and $j = 1, 2, \cdots, Q$.

$$H_{0i} : p_{i1} = p_{i2} = \cdots = p_{iQ} \quad vs \quad H_{1i} : p_{i1} < p_{i2} < \cdots < p_{iQ} \tag{4.3.7}$$

Next, we want to define a suitable prior for unknown binomial proportions, $p_{i1}, \cdots, p_{iQ}$. Similar to the Local prior approach discussed in section 4.2.2, for each test i, we assume that the model has Q proportions under both H_{0i} and H_{1i}; and use mode-based Beta priors to model the uncertainty of p_{ij}s under each hypothesis. So that, we can modify the models 4.2.14 and 4.2.15 under the null and alternative hypotheses as follows to accommodate MT of Q order-restricted binomial proportions.

Prior Specifications

Define the joint prior for $p_{i1}, p_{i2}, \cdots, p_{iQ}$ under H_{0i} and H_{1i} as

$$\text{Under } H_{0i} : p_{i1}, \cdots, p_{iQ} | p_{i0}, r_0 \overset{iid}{\sim} Beta(r_0 p_{i0} + 1, r_0(1 - p_{i0}) + 1) \tag{4.3.8}$$

$$\text{Under } H_{1i} : p^*_{i1}, \cdots, p^*_{iQ} | p_{i0}, r \overset{iid}{\sim} Beta(r p_{0i} + 1, r(1 - p_{0i}) + 1) \tag{4.3.9}$$
$$(p_{i1}, \cdots, p_{iQ}) = (p^*_{(i1)}, \cdots, p^*_{(iQ)}) \text{ ordered smallest to largest}$$

For each i, let $\pi(p_{i1}, \cdots, p_{iQ} | p_{i0}, r)$ denotes the prior for $(p_{i1}, \cdots, p_{iQ})$ given p_{i0} and r defined in 4.3.9.The hyperparameters in 4.3.8 and 4.3.9 have prior distributions given in 4.2.14 and 4.2.16, respectively.

Joint Prior Distributions of $(p_{i1}, \cdots, p_{iQ})$

Let $\boldsymbol{\pi_{0i}} \equiv \pi_{0i}(p_{i1}, \cdots, p_{iQ})$ and $\boldsymbol{\pi_{1i}} \equiv \pi_{1i}(p_{i1}, \cdots, p_{iQ})$ be the joint prior densities for $(p_{i1}, \cdots, p_{iQ})$ under the null and alternative hypotheses for each test i.

$$\boldsymbol{\pi_{0i}} = \int_0^1 \int_0^1 \int_0^\infty \prod_{j=1}^{Q} \pi(p_{ij}|p_{i0}, w_0)\, \pi(\Theta_3)\, dw_{00}\, dp_{00}\, dp_{i0} \tag{4.3.10}$$

$$\boldsymbol{\pi_{1i}} = \int_0^1 \int_0^\infty \int_0^1 \int_0^\infty \pi(p_{i1}, \cdots, p_{iQ}|p_{i0}, w)\pi(\Theta_4)\, dw_{00}\, dp_{00}\, dw\, dp_{i0} \tag{4.3.11}$$

where $\pi(\Theta_3) = \pi(p_{i0}|p_{00}, w_{00})\, \pi(w_{00})\, \pi(p_{00})$ and $\pi(\Theta_4) = \pi(p_{i0}|p_{00}, w_{00})\, \pi(w)\, \pi(w_{00})\, \pi(p_{00})$.

Posterior Probability of H_{0i}

Given $p \sim U(0, 1)$ is the proportion of true nulls in M tests, the posterior probability of each test i, H_{0i}, can be written as

$$P(H_{0i}|\boldsymbol{X}) = \frac{\boldsymbol{N_1} \times \boldsymbol{N_2}}{\boldsymbol{D}} \tag{4.3.12}$$

where,

$$\boldsymbol{N_1} = \int_0^1 \cdots \int_0^1 \prod_{j=1}^{Q} f(x_{ij}|p_{ij})\, \boldsymbol{\pi_{0i}}\, dp_{i1} \cdots dp_{iQ} \cdot p$$

$$\boldsymbol{N_2} = \prod_{q=1 (q \neq i)}^{M} \left[\int_0^1 \cdots \int_0^1 \prod_{j=1}^{Q} f(x_{qj}|p_{qj}) \Big(\boldsymbol{\pi_{0q}} \cdot p + \boldsymbol{\pi_{1q}} \cdot (1-p) \Big)\, dp_{q1} \cdots dp_{qQ} \right]$$

$$\boldsymbol{D} = \prod_{q=1}^{M} \left[\int_0^1 \cdots \int_0^1 \prod_{j=1}^{Q} f(x_{qj}|p_{qj}) \Big(\boldsymbol{\pi_{0q}} \cdot p + \boldsymbol{\pi_{1q}} \cdot (1-p) \Big)\, dp_{q1} \cdots dp_{qQ} \right]$$

Simulation, Result, and Conclusion

We implement the proposed method using synthetic data to evaluate the Bayesian MT procedure and later use real data examples to illustrate and compare the results using this method with some other approaches in the literature. We use several different settings for synthetic data under three cases; all the tests are null true, all the tests are alternative true, and mixed cases. In all of the settings, we assume equal sample size (i.e. $n_{ij} = n$) for all $i = 1, \cdots, M$ and $j = 1, \cdots, Q$. We first calculate the posterior probability of the null hypothesis for each test i using 500 thousand mcmc samples with a 50 thousand burning phase and then the average of posterior probabilities of null hypotheses. Below are the results for some of the simulation settings.

Case 1: All null true

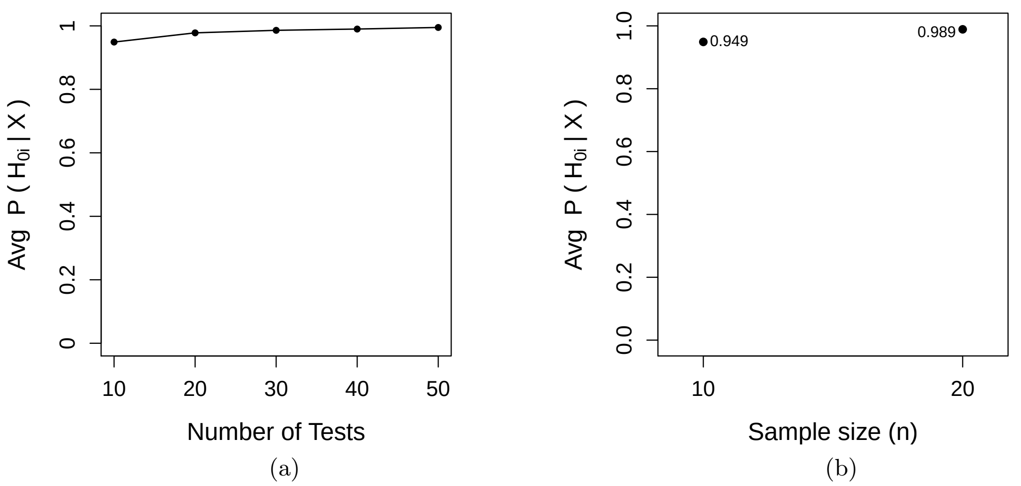
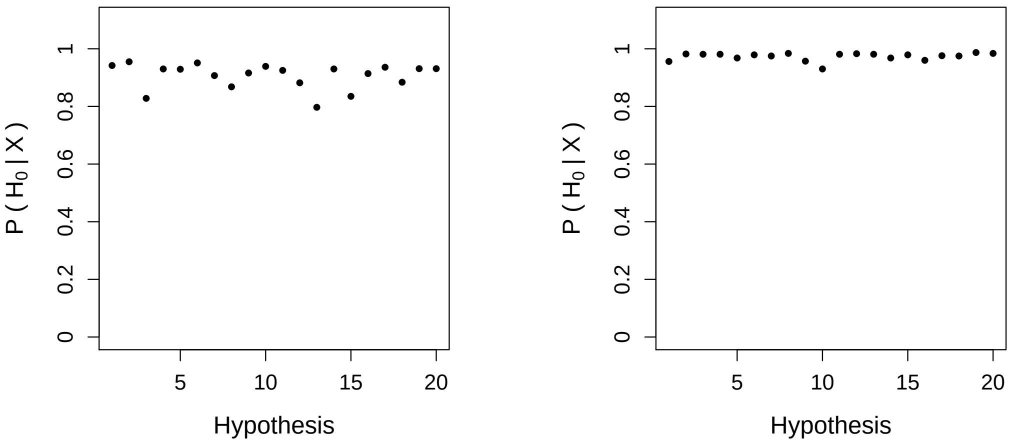


Figure 4.12.: Average of the posterior probabilities of H_{0i}s for $M = 20$ tests and $Q = 3$ proportions (a) as the number of test increases ($M = 10, 20, 30, 40, 50$) (b) as the sample size increases from $n = 10$ to $n = 20$ for data $x_{i1} = x_{i2} = x_{i3} = 5$.

Suppose we are interested in testing M independent hypotheses with Q binomial proportions; all are null true. Figure 4.13 (a) depicts the average of the posterior probabilities of H_{0i}s as n increases when $Q = 3$ with $x_{i1} = x_{i2} = x_{i3} = 5$. The average of posterior probabilities of the null hypotheses increases as M increases in the full null case. Figure 4.13(b) illustrates the average of posterior probabilities of H_{0i} as n increases from 10 to 20 for the same data. Figure 4.13 shows the posterior probabilities of each individual true null hypotheses for simulation settings $M = 20, Q = 3$, $x_{ij} \overset{iid}{\sim} Bin(n, 0.2)$ when (a) $n = 10$ and (b) $n = 15$. From figures 4.12 (b) and 4.13, we see that as n increases, posterior probabilities of true null hypotheses increase, and posterior probabilities are less sparse.



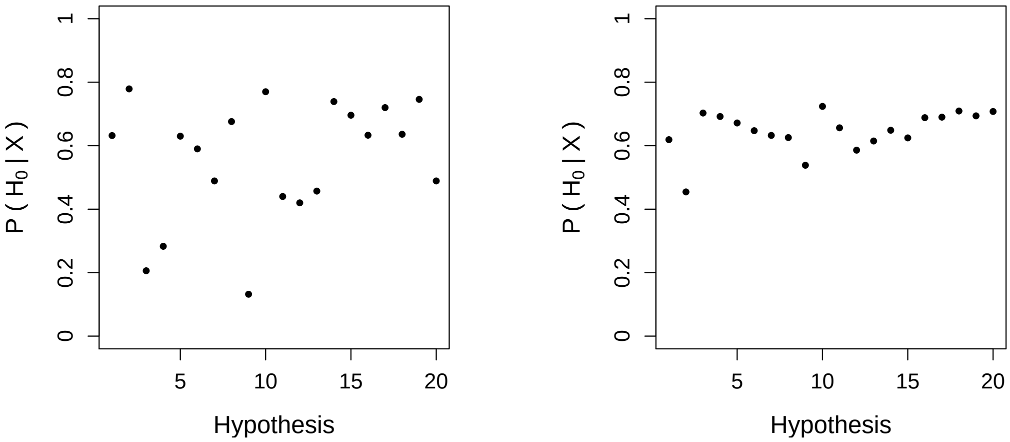
(a) (b)

Figure 4.13.: Posterior probability of H_{0i} for $M = 20$ tests with $Q = 3$ as the sample size increases from (a) $n = 10$ to (b) $n = 15$ when all the tests are null true: $x_{ij} \sim Bin(n, 0.2)$ independently for $i = 1, \cdots, 20$ and $j = 1, 2, 3$.

Case 2: All alternative true

Next we consider the situation, all the tests are alternative true. As in case 1, here also we set $M = 20$, $Q = 3$ and $n = 10, 15$. Although we present the result for two settings (fig.4.14), we consider few other different settings such as (i) $x_{i1} \sim Bin(n, 0.2)$, $x_{i2} \sim Bin(n, 0.5)$, $x_{i3} \sim Bin(n, 0.7)$; (ii) $x_{i1} \sim Bin(n, 0.2)$, $x_{i2} \sim Bin(n, 0.3)$, $x_{i3} \sim Bin(n, 0.7)$; (iii) $x_{i1} \sim Bin(n, 0.2)$, $x_{i2} \sim Bin(n, 0.3)$, $x_{i3} \sim Bin(n, 0.4)$; (iv) $x_{i1} \sim Bin(n, 0.2)$, $x_{i2} \sim Bin(n, 0.25)$, $x_{i3} \sim Bin(n, 0.3)$; (v) $x_{i1} \sim Bin(n, 0.2)$, $x_{i2} \sim Bin(n, 0.25)$, $x_{i3} \sim Bin(n, 0.3)$ for $i = 1, \cdots, 10$ and $x_{i1} \sim Bin(n, 0.2)$, $x_{i2} \sim Bin(n, 0.5)$, $x_{i3} \sim$

$Bin(n, 0.7)$ for $i = 11, \cdots, 20$. All of the settings give similar results as figure 4.14, as the number of tests increases posterior probabilities are less sparse.



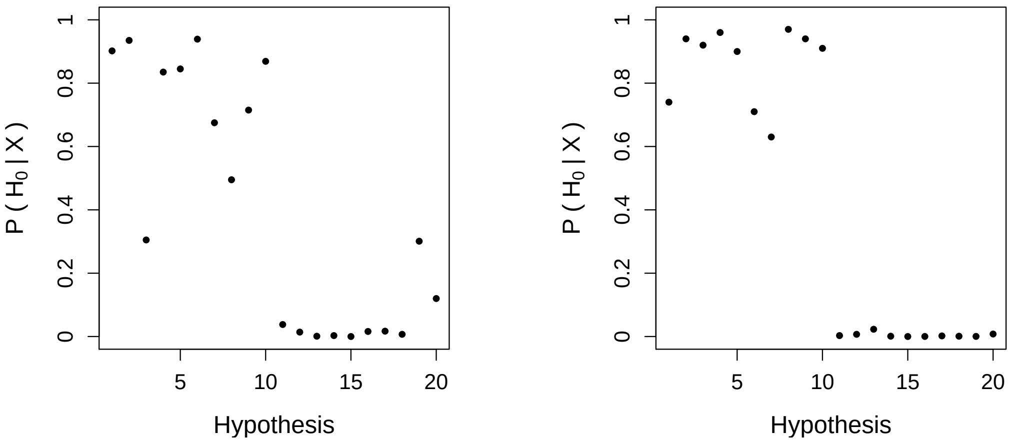
(a) (b)

Figure 4.14.: Posterior probability of H_{0i}s for $M = 20$ tests with $Q = 3$ as the sample size increases from (a) $n = 10$ to (b) $n = 15$ when all the tests are alternative true: $x_{i1} \sim Bin(n, 0.2)$, $x_{i2} \sim Bin(n, 0.25)$, and $x_{i3} \sim Bin(n, 0.3)$, for $i = 1, \cdots, 20$.

Case 3: Mixed case

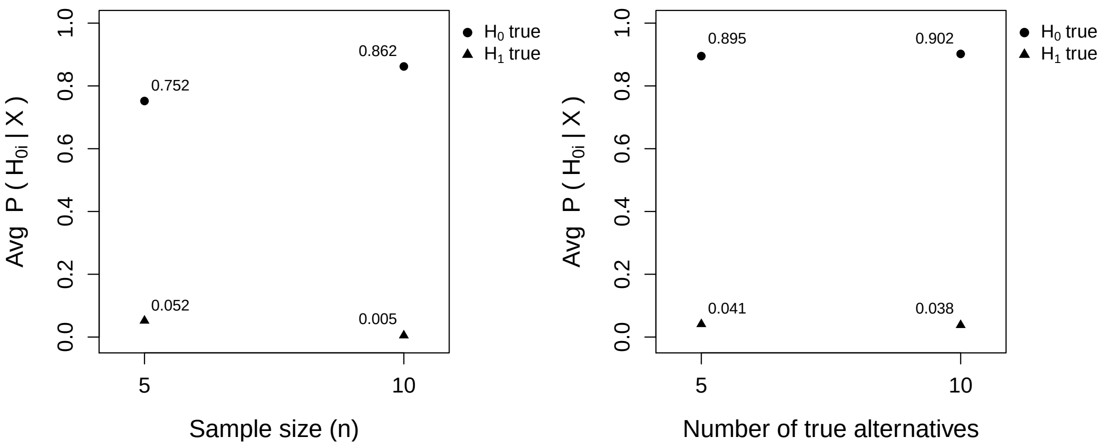
Now we consider the setting while some of the tests are null true, rest of them are alternative true. Figure 4.15 shows results when $M = 20$, $Q = 3$, and the number of true alternatives, k_0, increases 5 to 10 for data $x_{ij} \overset{iid}{\sim} Bin(10, 0.2)$ for $i = 1, \cdots, 20$; $j = 1, 2$ and $x_{i3} \overset{iid}{\sim} Bin(10, 0.2)$ for $i = 1, \cdots, k_0$; $x_{i3} \overset{iid}{\sim} Bin(10, 0.7)$ for $i = k_0 + 1, \cdots, 20$. As n increases, separation of the posterior probabilities of H_{0i}s for true null and true alternative hypotheses is more apparent, and posterior probabilities are less sparse.

Figure 4.16 displays the average posterior probability H_{0i}s, $Avg\, P(H_{0i}|X)$, as the sample size increases (fig.4.16 (a)) and as the number of true alternatives,k_0, increases (fig.4.16 (b)). As n increases, the average posterior probability of H_{0i} for true null increases, and true alternatives decrease. As the number of true alternatives increases, $Avg\, P(H_{0i}|X)$ of true null increases; $Avg\, P(H_{0i}|X)$ of true alternatives decreases.



(a) (b)

Figure 4.15.: Posterior probability of H_{0i}'s for $M = 20$ tests with $Q = 3$ as the number of true alternatives increases (a) $k_0 = 5$ (b) $k_0 = 10$ when $x_{i1}, x_{i2}, \sim Bin(10, 0.2)$ for $i = 1, \cdots, 20$ and $x_{i3} \sim Bin(10, 0.2)$ for $i = 1, \cdots, k_0$ and $x_{i3} \sim Bin(10, 0.7)$ for $i = k_0 + 1, \cdots, 20$.



(a) (b)

Figure 4.16.: Average of $P(H_{0i}|X)$ for $M = 20$ tests with $Q = 3$ proportions (a) as n increases for data: $x_{i1}, x_{i2}, \sim Bin(n, 0.2)$ for $i = 1, \cdots, 20$ and $x_{i3} \sim Bin(n, 0.2)$ for $i = 1, \cdots, 10$ and $x_{i3} \sim Bin(n, 0.7)$ for $i = 11, \cdots, 20$. (b) as the number of true alternative increases from $k_0 = 5$ to $k_0 = 10$ when $x_{i1}, x_{i2}, \sim Bin(10, 0.2)$ for $i = 1, \cdots, 20$ and $x_{i3} \sim Bin(10, 0.2)$ for $i = 1, \cdots, k_0$ and $x_{i3} \sim Bin(10, 0.7)$ for $i = k_0 + 1, \cdots, 20$.

4.4. Real Data Applications

We provide three applications of the proposed formal Bayesian procedures in this chapter for multiple testing of ordered binomial proportions.

4.4.1. Adverse event data from Heller and Gur (2011)

Consider applying proposed one-sided testing procedures, Local prior and Threshold prior methods, in section 4.2.2 to the data set in Heller and Gur [23] that relates treatment to an adverse event for ten studies. This data set reports the occurrences and nonoccurrences of the adverse event among the treated and controls. Heller and Gur investigate the association between reporting the adverse event and the studies by conducting Fisher's exact test (one-sided) for each study and adjusting for multiplicity by using several discrete FDR procedures. In table 4.2 we give adjusted p-values from two discreate multiple testing procedures reported in Heller and Gur [23], discreate Benjamini-Hochberg procedure (DBH) and discreate Benjamini and Liu procedure(DBL).

Given $x_{iC} \sim Bin(n_{iC}, p_{iC})$ and $x_{iT} \sim Bin(n_{iT}, p_{iT})$ are data for controls and treated groups for $i = 1, \cdots, 10$, we apply Local and Threshold prior methods provided in section 4.2.2 to test the hypotheses

$$H_{0i} : p_{iT} = p_{iC} \quad vs \quad H_{1i} : p_{iT} < p_{iC}.$$

Control x_{iC}/n_{iC}	Treated x_{iT}/n_{iT}	DBH	DBL	$P(H_{0i}\|X)_L$	$P(H_{0i}\|X)_T$
13/16	1/16	0.000	0.000	0.000	0.000
12/32	2/38	0.001	0.002	0.004	0.002
7/13	1/15	0.012	0.023	0.011	0.009
12/20	10/40	0.012	0.023	0.012	0.016
5/23	0/20	0.038	0.077	0.097	0.029
7/9	2/7	0.062	0.078	0.021	0.041
15/27	8/24	0.082	0.107	0.059	0.110
7/22	3/14	0.351	0.200	0.218	0.212
5/15	5/17	0.442	0.200	0.263	0.238
5/25	7/21	0.846	0.200	0.467	0.212

Table 4.2.: Table relating treatment to adverse event, for 10 studies. Number of occurrences and total number of observations for controls and treated are labled as *Control* and *Treated* repectively. Adjusted p-values from the multiple testing procedures: *DBH* and *DBL*; psterior probability of H_{0i} under Local prior-$P(H_{0i}|X)_L$, and Threshold prior-$P(H_{0i}|X)_T$ with $r_0 = 200\sqrt{n_{0i}}$ are reported in the table.

The DBH and DBL procedures with the nominal level $\alpha = 0.1$ lead to reject 3 and 4 null hypotheses, respectively. When considering the posterior probabilities from our proposed methods with 0.5 cutoff, both Local and Threshold prior methods reject all the null hypotheses. However, for the last test, posterior probability from the Threshold prior is smaller than Local prior.

4.4.2. Clinical trial data from Chen and Sarkar (2004)

We now modify and apply our proposed Local and Threshold prior methods in section 4.2.2 for multiple testing of one-sided two binomial proportions to the data set in Chen and Sarkar [15]. This data set considers a clinical trial that compares four formulations (A, B, C, and D) of a pharmaceutical compound with a placebo control in terms of some response rate. The goal here is to check if the additives used in the four formulations help produce more responses. During the follow-up period after the product administration, the number of patients who develop the event of interest in each group is recorded (Table 4.3).

Given that $i = 1, 2, 3, 4$ represent testing four formulations A, B, C, D against the Control, $x_0 \sim Bin(n_0, p_0)$ data for control, and $x_{i1} \sim Bin(n_{i1}, p_{i1})$ data for each formulation, the interest is testing the hypotheses

$$H_{0i} : p_{i1} = p_0 \quad vs \quad H_{1i} : p_{i1} < p_0.$$

Based on the given details of the data set in table 4.3, model specifications discussed in section 4.2.2 can be modified as follows.

For Local prior define

Under H_{0i} :

$p_{i1}|p_0, r_0 \sim Beta(r_0 p_0 + 1, r_0(1 - p_0) + 1)$

$p_0 \sim U(0, 1)$

$w_0 = 1/r_0 = 1/(100\sqrt{n_{i1}})$

Under H_{1i} :

$p^*_{i1}|p_0, r \sim Beta(rp_0 + 1, r(1 - p_0) + 1)$

$p_{i1} = p^*_{i1} \cdot I(p_{i1} > p_0)$

$p_0 \sim U(0, 1)$

$w \sim exp(1)$ and $w = 1/r$

and $p = P(H_{0i})$ has $U(0, 1)$ distribution.

For Threshold prior define

$$p^*_{i1}|p_0, r \sim Beta(rp_0 + 1, r(1 - p_0) + 1)$$
$$p_0 \sim U(0, 1)$$
$$w \sim exp(1) \text{ and } w = 1/r$$

$$p_{i1} = \begin{cases} p_0 & \text{if } 0 < LOR_i < \varepsilon \\ p^*_{i1} & \text{otherwise} \end{cases}$$

and p is calculated as $P(0 < LOR_i < \varepsilon)$ where $\varepsilon \sim U(0, K)$ and $K \sim Exp(0.3)$.

Formulation	# of patients	# of responders	Response rate	$B^{(r)}$	$P(H_{0i}\|X)$ Local	$P(H_{0i}\|X)$ Threshold
Control	200	93	0.465			
A	150	116	0.773	0.008	0.0000	0.0000
D	150	88	0.587	0.393	0.2347	0.2017
B	150	85	0.567	0.717	0.3239	0.2422
C	150	74	0.493	2.186	0.6296	0.5203

Table 4.3.: Summary statistics, stepwise Bayes factor-$B^{(r)}$ and posterior probability of H_{0i}- $P(H_{0i}|X)$, under Local and Threshold priors.

Chen and Sarkar has applied the Bayesian step-down approach to a general multiple testing problem proposed by Sarkar et al. [30] to the data set in table 4.3. This application of the Bayesian stepwise method in Chen and Sarkar indicates that formulations A, B, and D differ from the control in terms of the response rate. While our proposed Local and Threshold prior methods with cutoff 0.5 give the same results as the Bayesian stepwise method, i.e., formulations A, B, and D differ from the control, the Threshold prior method reports somewhat lower posterior probabilities than the Local prior method.

4.4.3. Clinical trial data from Agresti and Coull (1996)

So far in our work, we only assume an order restriction in each test; we did not make any further assumptions about how patterns of association vary across the tests. That is all our work is based on the assumption that tests are heterogeneous. However, there are situations where the patterns of association are not permitted to vary across the tests; this approach is called homogeneous testing. For example, Agresti and Coull [2] assume that the odds ratio between the response and a pair of treatment levels is identical in each test in one of their models.

Even though our original model in section 4.3.2 is more general and does not assume homogeneity, we can modify this model to get homogeneous tests under H_{1i} by making the odds ratio between the proportion and a pair of groups is identical in each test.

Test of Homogeneity

Step 1. Among the tests $i = 1, 2, \cdots, M$ select a single test, say $i = 1$.

Step 2. Then under H_{11}, generate $p^*_{1j}|p_{10}, r \overset{iid}{\sim} Beta\big(rp_{10} + 1, r(1 - p_{10}) + 1\big)$ for $j = 1, \cdots, Q$ and set $(p_{11}, \cdots, p_{1Q}) = (p^*_{(11)}, \cdots, p^*_{(1Q)})$ where $(p^*_{(11)}, \cdots, p^*_{(1Q)})$ is ordered from smallest to largest.

Step 3. Now calculate the odds ratio for treatments $j = 2, \cdots, Q$ for the test $i = 1$ as $OR_{1j} = \dfrac{p_{11} \cdot (1 - p_{1j})}{p_{1j} \cdot (1 - p_{11})}$.

Step 4. Next for rest of the tests $i = 2, \cdots, M$; for treatment $j = 1$; generate $p_{i1} \sim Beta(rp_{i0} + 1, r[1 - p_{i0}] + 1)$.

Step 5. Finally for $i = 2, \cdots, M$ and $j = 2, \cdots, Q$ set $p_{ij} = \dfrac{p_{i1}}{[p_{i1} + (1 - p_{i1}) \cdot OR_{1j}]}$. The expression for p_{ij} is obtained by equating corresponding odd ratios from tests 1 and i, and then solve for p_{ij}.

$$\frac{p_{i1} \cdot (1 - p_{ij})}{p_{ij} \cdot (1 - p_{i1})} = \frac{p_{11} \cdot (1 - p_{1j})}{p_{1j} \cdot (1 - p_{11})} => \frac{p_{i1} \cdot (1 - p_{ij})}{p_{ij} \cdot (1 - p_{i1})} = OR_{1j}$$

Agresti and Coull[2] presents the clinical trial data set in Table 4.4, which relates a binomial response to ordered levels of an explanatory variable representing drug doses with data collected at several centers. The study is conducted in 13 centers on 119 subjects with a certain medical condition. At each center, subjects are randomly assigned to three dose levels of a drug; the number of observations and the number of 'success' responses at each dose level is recorded.

One of the study goals is to test the hypothesis of no treatment effect. The probability of all the tests are null true, give the test at each center i, all the proportions $p_{i1}, \cdots, p_{iQ}$ are equal versus proportions are monotonically increasing function of dosage level.

Test the hypothesis of no treatment effect,

$$H_0: \text{ all the tests are null } (H_{0i}) \text{ true}$$

given the test at each center i

$$H_{0i}: p_{i1} = \cdots = p_{iQ} \quad \text{vs} \quad H_{1i}: p_{i1} < \cdots < p_{iQ}$$

for $i = 1, \cdots, 13$ and Q is either $1, 2,$ or 3.

We apply the order-restricted Local prior multiple testing procedure introduced in section 4.3.2 to the data set in Agresti and Coull under both homogeneous and heterogeneous assumptions. Based on this given data set, in our simulation setting, $M = 13$, and Q is either 1, 2, or 3, depending on the number of dosage levels used at each center. r_0 is fixed at 200 to avoid the sensitivity of posterior probabilities of each null hypothesis to the choice of r_0.

For each test i, we calculate the posterior probability of each null hypothesis H_{0i} is true and the estimate of proportions under homogeneous and heterogeneous assumptions.

Finally, we also calculate the posterior probability that all null hypotheses are true under both homogeneous and heterogeneous assumptions.

In Agresti and Coull, they introduce two likelihood-ratio tests, Homogeneous fit and Heterogeneous fit, which are sensitive to order-restricted alternatives and find estimates of exact p-values of the test statistics and success proportions under these two different settings. These two approaches differ in terms of whether they permit patterns of association to varying among centers. While the "Homogeneous fit" approach assumed that the odds ratio between the response and a pair of treatment levels is identical in each center, "Heterogeneous fit" is a more general approach allowing the odds ratio to vary in an unrestricted manner across centers.

We compare the results from our proposed order-restricted multiple testing procedures under the homogeneous and heterogeneous assumptions with those from order-restricted conditional independence tests in Agresti and Coull. Estimates of binomial proportions from our proposed order-restricted multiple testing procedures (table 4.4) clearly show that the binomial parameter is a monotonically increasing function of dosage level than Agresti and Coull (table 4.5). The posterior probabilities of all null hypotheses being true under homogeneous and heterogeneous assumptions are 0.026 and 0.049. Therefore, with a cutoff of 0.5, our proposed order-restricted multiple testing procedures under homogeneous and heterogeneous assumptions reject the null hypothesis that all the tests are null true. In comparison, the estimated exact p-values of the conditional independence test under homogeneous and heterogeneous fits in Agresti and Coull are 0.070 and 0.128. Our proposed order-restricted multiple testing procedure provides more substantial evidence of an association than the order-restricted conditional independence test in Agresti and Coull.

Center	Dose	# of trials	# of success	Success proportion	$\hat{p}_{ij}$ Hom.	$\hat{p}_{ij}$ Het.	$P(H_{0i}\|X)$ Hom.	$P(H_{0i}\|X)$ Het.
1	1	1	0	0.000	0.469	0.463	0.321	0.398
	2	1	1	1.000	0.548	0.588		
	3	4	3	0.750	0.674	0.678		
2	1	1	1	1.000	0.473	0.477	0.606	0.609
	3	1	0	0.000	0.563	0.531		
3	1	7	4	0.571	0.393	0.397	0.631	0.648
	2	6	1	0.167	0.420	0.421		
	3	2	1	0.500	0.496	0.477		
4	1	1	0	0.000	0.161	0.171	0.605	0.649
	2	2	0	0.000	0.178	0.203		
	3	2	0	0.000	0.223	0.246		
5	1	2	2	1.000	0.705	0.698	0.466	0.646
	2	1	1	1.000	0.747	0.739		
	3	4	3	0.750	0.804	0.773	0.374	
6	1	12	9	0.750	0.753	0.747	0.203	0.458
	2	10	8	0.800	0.818	0.808		
	3	9	9	1.000	0.892	0.868		
7	1	6	5	0.833	0.788	0.785	0.423	0.675
	2	5	5	1.000	0.832	0.824		
	3	6	5	0.833	0.877	0.846		
8	2	1	0	0.000	0.377	0.363	0.457	0.480
	3	2	1	0.500	0.474	0.477		
9	1	2	0	0.000	0.281	0.248	0.374	0.359
	2	2	0	0.000	0.333	0.337		
	3	3	2	0.667	0.457	0.483		
10	1	2	0	0.000	0.289	0.301	0.456	0.527
11	1	2	1	0.500	0.491	0.460	0.359	0.412
	2	3	1	0.333	0.558	0.543		
	3	2	2	1.000	0.677	0.661		
12	1	4	3	0.750	0.713	0.713	0.271	0.497
	2	5	4	0.800	0.799	0.784		
	3	5	5	1.000	0.872	0.845		
13	1	1	0	0.000	0.220	0.225	0.576	0.582
	2	1	0	0.000	0.247	0.274		
	3	1	0	0.000	0.309	0.334		

Table 4.4.: Estimates of binomial proportions, $\hat{p}_{ij}$Hom. and $\hat{p}_{ij}$Het., and posterior probabilities of H_{0i}s, $P(H_{0i}|X)$Hom. and $P(H_{0i}|X)$Het., from our proposed model in section 4.3.2. While $\hat{p}_{ij}$Hom. and $P(H_{0i}|X)$Hom. represent results under the homogeneity assumption, $\hat{p}_{ij}$Het. and $P(H_{0i}|X)$Het. represent results under the heterogeneous assumption.

Center	Dose	# of trials	# of success	Success proportion	Hom. fit	Het. fit
1	1	1	0	0.000	0.486	0.000
	2	1	1	1.000	0.486	0.800
	3	4	3	0.750	0.757	0.800
2	1	1	1	1.000	0.355	0.500
	3	1	0	0.000	0.645	0.500
3	1	7	4	0.571	0361	0.385
	2	6	1	0.167	0361	0.385
	3	2	1	0.500	0.651	0.500
4	1	1	0	0.000	0.000	0.000
	2	2	0	0.000	0.000	0.000
	3	2	0	0.000	0.000	0.000
5	1	2	2	1.000	0.775	0.857
	2	1	1	1.000	0.775	0.857
	3	4	3	0.750	0.919	0.857
6	1	12	9	0.750	0.801	0.750
	2	10	8	0.800	0.801	0.800
	3	9	9	1.000	0.930	1.000
7	1	6	5	0.833	0.847	0.833
	2	5	5	1.000	0.847	0.909
	3	6	5	0.833	0.948	0.909
8	2	1	0	0.000	0.176	0.000
	3	2	1	0.500	0.412	0.500
9	1	2	0	0.000	0.182	0.000
	2	2	0	0.000	0.182	0.000
	3	3	2	0.667	0.423	0.667
10	1	2	0	0.000	0.000	0.000
11	1	2	1	0.500	0.495	0.400
	2	3	1	0.333	0.495	0.400
	3	2	2	1.000	0.763	1.000
12	1	4	3	0.750	0.736	0.750
	2	5	4	0.800	0.814	0.800
	3	5	5	1.000	0.935	1.000
13	1	1	0	0.000	0.000	0.000
	2	1	0	0.000	0.000	0.000
	3	1	0	0.000	0.000	0.000

Table 4.5.: Estimates of binomial proportions from Agresti and Coull: *Hom.fit* - results under the homogeneity assumption and *Het.fit* - results under the heterogeneous assumption.

4.5. Conclusion

Though much frequentist work has been done for testing discrete data, there is little literature on Bayesian. Especially for MT with order-restricted binomial proportions, there is only one paper from Sarkar et al. [30], but this paper uses approximations, not the formal Bayesian approach. Since there is no formal Bayesian approach yet available for testing order-restricted binomial proportions, by doing the work in this chapter, our primary expectation is to provide a formal Bayesian approach for testing order-restricted binomial proportions. We develop two different Bayesian approaches for testing order-restricted binomial proportions based on the prior choice for modeling the unknown proportions under the alternative hypotheses.

Our proposed work has two novel contributions. We provide a formal Bayesian approach for testing order-restricted binomial proportions,"Local prior approach". The second contribution is we developed a non-local prior to one-sided testing of two binomial proportions," Threshold prior approach".

In single testing of one-sided two binomial proportions, when using the Local prior approach, we see a discrepancy between the convergence rates of the Bayes factors under true null and alternative hypotheses. With the proposed Threshold prior approach, we improve the convergence rate of the Bayes factor in favoring the true null. Also, when comparing the Local and Threshold priors results for MT of one-sided two binomial proportions, the Threshold prior approach is an acceptable alternative to the formal Local prior approach.

Generally, when consider testing $Q(\geq 2)$ order-restricted binomial proportions, based on the given results in this chapter, we can conclude that our proposed Bayesian approaches are working well and give similar or improved results to the Frequentist and Bayesian approaches relating to the studies in reference papers use in section 4.4.

Bibliography

[1] F. Abramovich, Y. Benjamini, D. L. Donoho, and I. M. Johnstone. "Adapting to unknown sparsity by controlling the false discovery rate". In: *Annals of Statistics* 34 (2 Apr. 2006), pp. 584–653. ISSN: 00905364. DOI: 10.1214/009053606000000074.

[2] A. Agresti and B. A. Coull. *Order-Restricted Tests for Stratified Comparisons of Binomial Proportions.* 1996, pp. 1103–1111. URL: https://www.jstor.org/stable/2533072.

[3] J.-F. Anger. "Bayesian inference for the location of a Student-t density". In: (2000).

[4] "Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem". In: *Annals of Statistics* 38 (5 Oct. 2010), pp. 2587–2619. ISSN: 00905364. DOI: 10.1214/10-AOS792.

[5] Y. Benjamini and Y. Hochberg. *Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.* 1995, pp. 289–300.

[6] Y. Benjamini and D. Yekutieli. *The Control of the False Discovery Rate in Multiple Testing under Dependency.* 2001, pp. 1165–1188.

[7] J. O. Berger. "An overview of robust Bayesian analysis (with discussion)". In: *Test* (3 1994), pp. 5–124.

[8] J. O. Berger and W. H. Jefferys. *THE APPLICATION OF ROBUST BAYESIAN ANALYSIS TO HYPOTHESIS TESTING AND OCCAM'S RAZOR.* 1992, pp. 17–32.

[9] J. O. Berger, T. Sellke, and J. Berger. *Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence.* 1987, pp. 112–122.

[10] J. O. Berger and W. E. Strawderman. "Choice of hierarchical priors: admissibility in estimation of normal means". In: *Annals of Statistics* 24 (3 1996), pp. 931–951.

[11] M. Bogdan, A. Chakrabarti, F. Frommlet, and J. K. Ghosh. "Asymptotic bayes-optimality under sparsity of some multiple testing procedures". In: *Annals of Statistics* 39 (3 June 2011), pp. 1551–1579. ISSN: 00905364. DOI: 10.1214/10-AOS869.

[12] M. Bogdan, J. K. Ghosh, A. Ochman, and S. T. Tokdar. "On the empirical bayes approach to the problem of multiple testing". In: vol. 23. Oct. 2007, pp. 727–739. DOI: 10.1002/qre.876.

[13] G. Casella and R. L. Berger. *Reconciling Bayesian and Frequentist Evidence in the One-Sided Testing Problem.* 1987, pp. 106–111.

[14] G. Casella and E. Moreno. *Objective Bayesian Analysis of Contingency Tables.* 2002. URL: www.cytel.com.

[15] J. Chen and S. K. Sarkar. "Multiple testing of response rates with a control: A Bayesian stepwise approach". In: *Journal of Statistical Planning and Inference* 125 (1-2 Oct. 2004), pp. 3–16. ISSN: 03783758. DOI: 10.1016/j.jspi.2003.05.001.

[16] "Criteria for bayesian model choice with application to variable selection". In: *Annals of Statistics* 40 (3 June 2012), pp. 1550–1577. ISSN: 00905364. DOI: 10.1214/12-AOS1013.

[17] S. Dudoit, J. P. Shaffer, and J. C. Boldrick. *Multiple Hypothesis Testing in Microarray Experiments.* 2003, pp. 71–103.

[18] B. Efron, R. Tibshirani, J. D. Storey, and V. Tusher. "Empirical bayes analysis of a microarray experiment". In: *Journal of the American Statistical Association* 96 (456 Dec. 2001), pp. 1151–1160. ISSN: 1537274X. DOI: 10.1198/016214501753382129.

[19] T.-H. Fan and J. O. Berger. "Behaviour of the posterior distribution and inference for a normal mean with t prior distributions". In: *Statistics and Decisions* 10 (1992), pp. 99–120.

[20] E. Gecili. *On Bayesian Multiplicity Adjustment in Multiple Testing.* 2018.

[21] "Generalization of DNA microarray dispersion properties: Microarray equivalent of t-distribution". In: *Biology Direct* 1 (Sept. 2006). ISSN: 17456150. DOI: 10.1186/1745-6150-1-27.

[22] C. Genovese and L. Wasserman. *Operating characteristics and extensions of the false discovery rate procedure.* 2002, pp. 499–517.

[23] R. Heller and H. Gur. "False discovery rate controlling procedures for discrete tests". In: (Dec. 2011). URL: http://arxiv.org/abs/1112.4627.

[24] V. E. Johnson and D. Rossell. *On the use of non-local prior densities in Bayesian hypothesis tests.* 2010, pp. 143–170. URL: http://www3.inter.

[25] D. V. Mehrotra and J. F. Heyse. "Use of the false discovery rate for evaluating clinical safety data". In: *Statistical Methods in Medical Research* 13 (3 2004), pp. 227–238. ISSN: 09622802. DOI: 10.1191/0962280204sm363ra.

[26] E. Moreno, F. J. Vazquez, and M. A. Negrin. *Objective Bayesian meta-analysis for sparse discrete data.* 2014, pp. 3676–3692. URL: www.cytel.com.

[27] J. Nakajima and M. West. "Bayesian Analysis of Latent Threshold Dynamic Models". In: *Journal of Business and Economic Statistics* 31 (2 Apr. 2013), pp. 151–164. ISSN: 07350015. DOI: 10.1080/07350015.2012.747847.

[28] B. O'neill. *Exchangeability, Correlation, and Bayes' Effect.* 2009, pp. 241–250.

[29] A. K. Ramdas, R. F. Barber, M. J. Wainwright, and M. I. Jordan. "A unified treatment of multiple testing with prior knowledge using the P-filter". In: *Annals of Statistics* 47 (5 2019), pp. 2790–2821. ISSN: 21688966. DOI: 10.1214/18-AOS1765.

[30] S. Sarkar, S. K. Sarkar, and J. Chen. *A Bayesian stepwise multiple testing procedure.* URL: https://www.researchgate.net/publication/228385611.

[31] J. G. Scott and J. O. Berger. "An exploration of aspects of Bayesian multiple testing". In: *Journal of Statistical Planning and Inference* 136 (7 SPEC. ISS. 2006), pp. 2144–2162. ISSN: 03783758. DOI: 10.1016/j.jspi.2005.08.031.

[32] J. D. Storey. *A direct approach to false discovery rates.* 2002, pp. 479–498.

[33] R. E. Tarone. *A Modified Bonferroni Method for Discrete Data.* 1990, pp. 515–522. URL: https://www.jstor.org/stable/2531456?seq=1&cid=pdf-reference#references_tab_contents.

[34] P. H. Westfall, W. O. Johnson, and J. M. Utts. *A Bayesian perspective on the Bonferroni adjustment.* 1997, pp. 419–427. URL: https://academic.oup.com/biomet/article/84/2/419/234015.

Appendix **A**

Appendix for Chapter 2

A.1. Information (In-)Consistency

Proof of Preposition 2.2.2:

(a). Let $\boldsymbol{x} = (x_1, .., x_n) \overset{iid}{\sim} N(\mu, \sigma^2)$ with known σ. Assume a normal prior $\pi(\mu|H_1) = N(0, \tau^2)$ with known τ.

First, consider the situation $n = 1$, i.e., $x \sim N(\mu, \sigma^2)$, then the marginal distribution of x under H_1 has the form

$$\begin{aligned}
m_1(x) &= \int f(x|\mu \neq 0, \sigma^2) \cdot \pi(\mu|0, \tau^2) d\mu \\
&= \int \frac{1}{\sqrt{2\pi\sigma^2}} . \frac{1}{\sqrt{2\pi\tau^2}} exp\Big\{ - \frac{(x-\mu)^2}{2\sigma^2} - \frac{\mu^2}{2\tau^2} \Big\} d\mu \\
&= \int \frac{1}{\sqrt{2\pi\sigma^2}} . \frac{1}{\sqrt{2\pi\tau^2}} exp\Big\{ - \frac{\big(\tau^2 x^2 - 2\tau^2 x\mu + \tau^2\mu^2 + \sigma^2\mu^2\big)}{2\sigma^2\tau^2} \Big\} d\mu \\
&= \int \frac{1}{\sqrt{2\pi\sigma^2}} . \frac{1}{\sqrt{2\pi\tau^2}} exp\Big\{ - \frac{(\sigma^2 + \tau^2)}{2\sigma^2\tau^2} \Big[\Big(\mu - \frac{x\tau^2}{(\sigma^2 + \tau^2)} \Big)^2 \\
&\qquad + \frac{\tau^2 x^2}{(\sigma^2 + \tau^2)} - \frac{\tau^4 x^2}{(\sigma^2 + \tau^2)^2} \Big] \Big\} d\mu
\end{aligned}$$

The above expression for $m_1(x)$ can be further simplified as below.

$$m_1(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{\sqrt{2\pi\tau^2}} \cdot \sqrt{2\pi \frac{\sigma^2\tau^2}{(\sigma^2+\tau^2)}}\, exp\left\{ \frac{-\left(x^2 - \frac{x^2\tau^2}{(\sigma^2+\tau^2)}\right)}{2\sigma^2} \right\} \cdot$$

$$\int \frac{1}{\sqrt{2\pi \frac{\sigma^2 v}{(\sigma^2+v)}}} exp\left\{ \frac{\left(\mu - \frac{x\tau^2}{(\sigma^2+\tau^2)}\right)^2}{\frac{2\sigma^2\tau^2}{(\sigma^2+\tau^2)}} \right\} d\mu$$

$$= \frac{1}{\sqrt{2\pi(\sigma^2+\tau^2)}} exp\left\{ \frac{-x^2}{2(\sigma^2+\tau^2)} \right\}$$

$$\text{Then, } B_{10} = \frac{m_1(x)}{f(x|0,\sigma^2)}$$

$$= \sqrt{\frac{\sigma^2}{(\sigma^2+\tau^2)}}\, exp\left\{ \frac{x^2\tau^2}{2\sigma^2(\sigma^2+\tau^2)} \right\} \to \infty \text{ as } x \to \infty$$

$$\therefore P(H_0|x) = \left[1 + \frac{(1-p)}{p} B_{10} \right]^{-1} \to 0 \text{ as } x \to \infty$$

From the above proof for $n = 1$, we found that the marginal distribution of x under H_1 has the form, $m_1(x) \sim N(0, \sigma^2 + \tau^2)$. Then, according to " The sum of Normal random variables follows Normal distribution," the random sample $\boldsymbol{x} = (x_1, .., x_n)$ also has the same marginal distribution $m_1(\boldsymbol{x}) \sim N(0, \sigma^2 + \tau^2)$.

$$\therefore P(H_0|\boldsymbol{x}) \to \infty \text{ as } \bar{x} \to \infty$$

(b). Let $\boldsymbol{x} = (x_1, .., x_n) \overset{iid}{\sim} N(\mu, \sigma^2)$ where σ is unknown and assigned the prior $\pi(\sigma^2) = 1/\sigma^2$. Assume a normal prior $\pi(\mu|H_1) = N(0, \tau^2)$, with a known τ. Then,

$$P(H_0|\boldsymbol{x}) = \left[1 + \frac{(1-p)}{p} B_{10}\right]^{-1}$$

Assuming that $p = 0.5$, then $P(H_0|\boldsymbol{x}) = \left[1 + B_{10}\right]^{-1}$. Here, $B_{10} = \left[1 + \frac{B_1}{A_1}\right]^{-1}$ where

$$A_1 = \int f(\boldsymbol{x}|\mu = 0, \sigma^2).1/\sigma^2 \, d\sigma^2 \quad \text{and}$$

$$B_1 = \int \left[\int f(\boldsymbol{x}|\mu, \sigma^2).1/\sigma^2 \, d\sigma^2\right].f(\mu|0, \tau^2) \, d\mu.$$

Defining $C_1 = \int f(\boldsymbol{x}|\mu, \sigma^2).\frac{1}{\sigma^2} \, d\sigma^2$, B_1 can be written as $B_1 = \int C_1 \,.\, f(\mu|0, \tau^2) \, d\mu$ and C_1 can be simplified as below.

$$\begin{aligned} C_1 &= \int f(\boldsymbol{x}|\mu, \sigma^2).\frac{1}{\sigma^2} \, d\sigma^2 \\ &= \frac{1}{[2\pi]^{n/2}} \int \frac{e^{\frac{-\Sigma(x_i-\mu)^2}{2\sigma^2}}}{[\sigma^2]^{\frac{n}{2}+1}} \\ &= \frac{1}{[2\pi]^{n/2}} \int \frac{e^{-\frac{S^2(\mu)}{2\sigma^2}}}{[\sigma^2]^{\frac{n}{2}+1}} d\sigma^2 \quad \text{where, } S^2(\mu) = s^2(n-1) + n(\bar{x} - \mu)^2 \\ &= \frac{\Gamma(n/2)}{[2\pi]^{n/2}[S^2(\mu)/2]^{n/2}} \int IG\left(\sigma^2|\alpha = \frac{n}{2}, \, \beta = \frac{S^2(\mu)}{2}\right) d\sigma^2 \qquad (*) \end{aligned}$$

Then, B_1 has the form

$$B_1 = \frac{\Gamma(n/2)}{[2\pi]^{n/2}} \, [2]^{n/2} \int \frac{e^{-\mu^2/2\tau^2}}{\sqrt{2\pi\tau^2} \, \left[s^2(n-1) + n(\bar{x} - \mu)^2\right]^{n/2}} \, d\mu.$$

Next consider A_1

$$
\begin{aligned}
A_1 &= \int \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\dfrac{x_i^2}{2\sigma^2}} \, d\sigma^2 \\
&= \frac{1}{[2\pi]^{n/2}} \int \frac{e^{-\frac{s^2(n-1)+n\bar{x}^2}{2\sigma^2}}}{[\sigma^2]^{\frac{n}{2}+1}} \, d\sigma^2 \\
&= \frac{\Gamma(n/2)}{[2\pi]^{n/2}\Big[\big(s^2(n-1)+n\bar{x}^2\big)/2\Big]^{n/2}} \int IG\Bigg(\sigma^2 | \alpha = \frac{n}{2}, \, \beta = \frac{\big(s^2(n-1)+n\bar{x}^2\big)}{2}\Bigg) \, d\sigma^2 \\
&= \frac{\Gamma(n/2)}{[2\pi]^{n/2}} [2]^{n/2} \frac{1}{\Big[\big(s^2(n-1)+n\bar{x}^2\big)\Big]^{n/2}}
\end{aligned}
$$

Now $P(H_0|\boldsymbol{x})$ has the form

$$
\begin{aligned}
P(H_0|\boldsymbol{x}) &= \left[1 + \int \left[\frac{s^2(n-1)+n\bar{x}^2}{s^2(n-1)+n(\bar{x}-\mu)^2}\right]^{n/2} \cdot \frac{1}{\sqrt{2\pi\tau^2}} e^{-\dfrac{\mu^2}{2\tau^2}} \, d\mu\right]^{-1} \\
&= \left[1 + \int C_2(\mu, \bar{x}) \cdot \frac{1}{\sqrt{2\pi\tau^2}} e^{-\dfrac{\mu^2}{4\tau^2}} \, d\mu\right]^{-1}
\end{aligned}
$$

where $C_2(\mu, \bar{x}) = \left[\dfrac{s^2(n-1)+n\bar{x}^2}{s^2(n-1)+n(\bar{x}-\mu)^2}\right]^{n/2} \cdot e^{-\dfrac{\mu^2}{4\tau^2}}$.

With s^2 a constant, $C_2(\mu, \bar{x})$ can be shown to be bounded. i.e., there exists K, $0 < K < \infty$ such that $C_2(\mu, \bar{x}) < K$ for all μ and $\bar{x}$. Hence,

$$
\begin{aligned}
P(H_0|\boldsymbol{x}) &> \left[1 + \int K \frac{e^{-\frac{\mu^2}{4\tau^2}}}{\sqrt{2\pi\tau^2}} \, d\mu\right]^{-1} \\
&= \left[1 + \sqrt{2}K\right]^{-1} > 0 \quad \text{i.e., } P(H_0|\boldsymbol{x}) \not\to 0 \text{ as } \bar{x} \to \infty
\end{aligned}
$$

Note:

(b)(i). Using a proper prior $\pi(\tau^2)$ for τ, still can get the same results.

(b)(ii). Replacing $\pi(\sigma^2) = 1/\sigma^2$ by $\pi(\sigma^2) = IG(\alpha_0/2, \beta_0/2)$, gives the same results. The change in the derivation is that, in the equation (*), parameters of the inverse gamma distribution will have the forms, $\alpha \to (n + \alpha_0)/2$ and $\beta \to (S^2(\mu) + \beta_0)/2$. Consequently, $\Gamma(n/2) \to \Gamma(\frac{n+\alpha_0}{2})$ and $[S^2(\mu)/2]^{n/2} \to \Big[\big(S^2(\mu) + \beta_0\big)/2\Big]^{n/2}$ etc.

Proof of Preposition 2.2.4:

(a). Let $x \sim t_\upsilon(\mu, \sigma^2)$ with known σ, $\pi(\mu|H_1, \tau^2) : t_\nu(0, \tau^2)$ with fixed and known ν, and τ is known. Assuming $p = P(H_0) = 0.5$, the posterior probability of H_0 can be written as

$$P(H_0|x) = \frac{t_\upsilon(0,\sigma)}{t_\upsilon(0,\sigma) + \int t_\upsilon(\mu,\sigma) t_\nu(0,\tau)\, d\mu}$$

Define $P(H_0|x) = \dfrac{1}{1+m}$ where

$$\begin{aligned}
m &= A_4 \int_{-\infty}^{\infty} \frac{\left(1 + \frac{1}{\upsilon}\frac{x^2}{\sigma^2}\right)^{(\upsilon+1)/2}}{\left(1 + \frac{1}{\upsilon}\frac{(x-\mu)^2}{\sigma^2}\right)^{(\upsilon+1)/2}} \cdot \frac{1}{\left(1 + \frac{1}{\nu}\frac{\mu^2}{\tau^2}\right)} \, d\mu \\
&> A_4 \int_{|x-\mu|<\delta}^{\infty} \left[\frac{1 + \frac{1}{\upsilon}\frac{x^2}{\sigma^2}}{1 + \frac{1}{\upsilon}\frac{(x-\mu)^2}{\sigma^2}}\right]^{(1+\upsilon)/2} \cdot \frac{1}{\left(1 + \frac{1}{\nu}\frac{\mu^2}{\tau^2}\right)} \, d\mu
\end{aligned}$$

Consider $|x - \mu| < \delta$ and this implies that $1 < 1 + \dfrac{(x-\mu)^2}{\sigma^2} < 1 + \dfrac{\delta^2}{\sigma^2}$. So that

$$\left[\frac{1 + \frac{1}{\upsilon}\frac{x^2}{\sigma^2}}{1 + \frac{1}{\upsilon}\frac{(x-\mu)^2}{\sigma^2}}\right]^{(1+\upsilon)/2} > \left[\frac{1 + \frac{1}{\upsilon}\frac{x^2}{\sigma^2}}{1 + \frac{1}{\upsilon}\frac{\delta^2}{\sigma^2}}\right]^{(1+\upsilon)/2}$$

$$\begin{aligned}
\therefore m &> A_4 \cdot \left[\frac{1 + \frac{1}{\upsilon}\frac{x^2}{\sigma^2}}{1 + \frac{1}{\upsilon}\frac{\delta^2}{\sigma^2}}\right]^{(1+\upsilon)/2} \int_{|x-\mu|<\delta}^{\infty} \frac{1}{\left(1 + \frac{1}{\nu}\frac{\mu^2}{\tau^2}\right)^{(1+\nu)/2}} \, d\mu \\
&= A_4 \cdot A_5 \cdot A_6
\end{aligned}$$

As $x \to \infty$, $A_5 \to \infty$ at a rate $(\upsilon + 1)$, $A_6 \to 0$ at rate $(\nu + 1)$. Hence $m \to \infty$ if $\upsilon > \nu$. i.e., $P(H_0|x) \to 0$ as $x \to \infty$ if $\upsilon > \nu$.

(a). Let $\boldsymbol{x} = (x_1, .., x_n) \overset{iid}{\sim} t_\upsilon(\mu, \sigma^2)$ where σ is known and $\pi(\mu|H_1) = t_\nu(0, \tau^2)$ with $\pi(\tau^2)$ proper.

As before let $P(H_0|X) = \dfrac{1}{1+m}$ where

$$
\begin{aligned}
m &= A_7 \int \prod_{i=1}^{n} \left[\frac{1 + \frac{1}{\upsilon}\frac{x_i^2}{\sigma^2}}{1 + \frac{1}{\upsilon}\frac{(x_i-\mu)^2}{\sigma^2}} \right]^{(1+\upsilon)/2} \pi(\mu|\tau^2)\,\pi(\tau^2)\,d\mu\,d\tau^2 \\
&> \int_{R=max|x_i-\mu|<\delta}^{\infty} \prod_{i=1}^{n} \left[\frac{1 + \frac{1}{\upsilon}\frac{x_i^2}{\sigma^2}}{1 + \frac{1}{\upsilon}\frac{(x_i-\mu)^2}{\sigma^2}} \right]^{(1+\upsilon)/2} \pi(\mu|\tau^2)\,\pi(\tau^2)\,d\mu\,d\tau^2
\end{aligned}
$$

Here R implies that, $\prod_{i=1}^{n} \left[1 + \dfrac{(x_i-\mu)^2}{\sigma^2} \right] < \left[1 + \dfrac{\delta^2}{\sigma^2} \right]^n$. Then,

$$
\begin{aligned}
m &> A_7 \cdot \prod_{i=1}^{n} \left[\frac{1 + \frac{1}{\upsilon}\frac{x_i^2}{\sigma^2}}{1 + \frac{1}{\upsilon}\frac{\delta^2}{\sigma^2}} \right]^{(1+\upsilon)/2} \cdot \int_{R}^{\infty} \frac{1}{\left(1 + \frac{1}{\nu}\frac{\mu^2}{\tau^2}\right)^{(1+\nu)/2}}\,d\mu\,d\tau^2 \\
&= A_7 \cdot A_8 \cdot A_9 \qquad \text{(A.1.1)}
\end{aligned}
$$

As $\boldsymbol{x} \to \infty$, $A_8 \to \infty$ at a rate $n(\upsilon + 1)$, $A_9 \to 0$ at rate $(\nu + 1)$. Hence $m \to \infty$ if $n\upsilon > \nu$. i.e., $P(H_0|\boldsymbol{x}) \to 0$ as $x \to \infty$ if $n\upsilon > \nu$.

A.2. $P(H_{0i}|X)$ for TN Model from Importance Sampling Approach

Proof of Result 2.3.1:

$$p_i = P(\mu_i = 0|\boldsymbol{X}) = \frac{P(\boldsymbol{X}, \mu_i = 0)}{P(\boldsymbol{X})}$$

where,

$$\begin{aligned}
P(\boldsymbol{X}, \mu_i = 0) &= \int ... \int P\big(\boldsymbol{X}, \mu_i = 0|p, \mu_{j_{(j\neq i)}}, \tau^2, \sigma^2\big)\, \pi\big(p, \mu_{j_{(j\neq i)}}, \sigma^2, \tau^2\big)\, dp\, d\mu_j's\, d\sigma^2 d\tau^2 \\
&= \int ... \int P\big(\boldsymbol{X}|p, \mu_i = 0, \mu_{j_{(j\neq i)}}, \tau^2, \sigma^2\big)\, P\big(\mu_i = 0|p\big)\, \pi\big(p, \mu_{j_{(j\neq i)}}, \sigma^2, \tau^2\big) \\
&\qquad d\mu_j's\, dp\, d\sigma^2 d\tau^2 \\
&= \int\int\int \Big\{ \int ... \int P\big(\boldsymbol{X}|p, \mu_i = 0, \mu_{j_{(j\neq i)}}, \sigma^2\big)\pi\big(\mu_{j_{(j\neq i)}}|\tau^2, p\big)\, d\mu_j's \Big\} \cdot p \\
&\qquad \pi(p)\, \pi(\sigma^2, \tau^2)\, dp\, d\sigma^2 d\tau^2 \\
&= \int\int\int pf(X_i|\mu_i = 0, \sigma^2) \Big\{ \int ... \int \prod_{j\neq i} P\big(X_j|p, \mu_{j_{(j\neq i)}}, \sigma^2\big)\, \pi\big(\mu_{j_{(j\neq i)}}|\tau^2, p\big)\, d\mu_j \Big\} \\
&\qquad \pi(p)\, \pi(\sigma^2, \tau^2)\, dp\, d\sigma^2 d\tau^2 \\
&= \int\int\int pf(X_i|\mu_i = 0, \sigma^2) \Big\{ \prod_{j\neq i} \int P\big(X_j|p, \mu_{j_{(j\neq i)}}, \sigma^2\big)\pi\big(\mu_{j_{(j\neq i)}}|\tau^2, p\big)\, d\mu_j \Big\} \\
&\qquad \pi(p)\, \pi(\sigma^2, \tau^2)\, dp\, d\sigma^2 d\tau^2 \\
&= \int\int\int p \prod_k t_\upsilon(x_i|\mu_i = 0, \sigma^2) \prod_{j\neq i} \Big\{ p \prod_k t_\upsilon(x_j|\mu_j, \sigma^2) \\
&\qquad + (1-p) \int \prod_k t_\upsilon(x_j|\mu_j, \sigma^2) N(\mu_j|0, \tau^2)\, d\mu_j \Big\} \pi(p)\, \pi(\sigma^2, \tau^2)\, dp\, d\sigma^2 d\tau^2
\end{aligned}$$

$$P(\boldsymbol{X}) = \int\int\int \prod_j \left\{ p \prod_k t_\upsilon(x_j|\mu_j, \sigma^2) + (1-p) \int \prod_k t_\upsilon(x_j|\mu_j, \sigma^2) N(\mu_j|0, \tau^2)\, d\mu_j \right\}$$
$$\pi(p)\, \pi(\sigma^2, \tau^2)\, dp\, d\sigma^2 d\tau^2$$

Letting $m(x_j|\sigma^2, \tau^2) = \int \prod_k t_\upsilon(x_j|\mu_j, \sigma^2) N(\mu_j|0, \tau^2)\, d\mu_j$; $P(\boldsymbol{X}, \mu_i = 0)$ and $P(\boldsymbol{X})$ can be rewrite as

$$P(\boldsymbol{X}, \mu_i = 0) = \int\int\int p \prod_k t_\upsilon(x_i|\mu_i = 0, \sigma^2) \prod_{j \neq i} \left\{ p \prod_k t_\upsilon(x_j|\mu_j, \sigma^2) + (1-p) m(x_j|\sigma^2, \tau^2) \right\}$$
$$\pi(p)\pi(\sigma^2, \tau^2) dp d\sigma^2 d\tau^2$$

$$P(\boldsymbol{X}) = \int\int\int \prod_j \left\{ p \prod_k t_\upsilon(x_j|\mu_j, \sigma^2) + (1-p)\, m(x_j|\sigma^2, \tau^2) \right\} \pi(p)\, \pi(\sigma^2, \tau^2)\, dp\, d\sigma^2 d\tau^2$$

Appendix B

Appendix for Chapter 3

B.1. Convergence of Bayes factor with Threshold prior - single proportion

Suppose we observe $X^{(n)} \equiv (x_1, \cdots, x_n)$ a random sample from a binomial distribution, $Bin(n, p_1)$, with density function $f(x|p_1)$ and for a known p_0 want to test

$$H_0 : p_1 = p_0 \quad \text{vs} \quad H_1 : p_1 \neq p_0 \tag{B.1.1}$$

Let, $P_n\big(X^{(n)}|p_1\big)$ denote the joint sampling density of the data, $L_n(p_1)$ denote the log-likelihood function, and $\hat{p_1}$ denote a maximum likelihood estimate of p_1, where

$$P_n\big(X^{(n)}|p_1\big) = \prod_{i=1}^{n} f(x_i|p_1)$$

$$L_n(p_1) = log\Big\{P_n\big(X^{(n)}|p_1\big)\Big\}$$

Given the prior for p_1 under H_1, $\pi_1(p_1)$, be a continuous density function with respect to Lebesgue measure, Bayes factor based on a sample size n is defined as

$$BF_n\big(1|0\big) = \frac{m_1\big(X^{(n)}\big)}{m_0\big(X^{(n)}\big)} = \frac{\displaystyle\int P_n\big(X^{(n)}|p_1\big)\pi_1(p_1)\,dp_1}{P_n\big(X^{(n)}|p_0\big)} \tag{B.1.2}$$

Now consider the first three terms of Taylar's expansion of $L_n(p_1)$ around $\hat{p}_1$.

$$L_n(p_1) \approx L_n(\hat{p}_1) + L_n'(\hat{p}_1)(p_1 - \hat{p}_1) + \frac{1}{2}\,L_n''(\hat{p}_1)(p_1 - \hat{p}_1)^2$$

Since $L_n(p_1)$ has maximum at $\hat{p}_1$ $L_n'(\hat{p}_1) = 0$. $\therefore$ $L_n(p_1) \approx L_n(\hat{p}_1) + \frac{1}{2}\,L_n''(\hat{p}_1)(p_1 - \hat{p}_1)^2$. By taking exponent on both sides of the expression for $L_n(p_1)$

$$\begin{aligned} P_n\big(X^{(n)}|p_1\big) &\approx P_n\big(X^{(n)}|\hat{p}_1\big) \cdot exp\left\{\frac{1}{2}\,L_n''(\hat{p}_1)(p_1 - \hat{p}_1)^2\right\} \\ &\approx P_n\big(X^{(n)}|\hat{p}_1\big) \cdot exp\left\{-\frac{1}{2}\,n\left[\frac{-L_n''(\hat{p}_1)}{n}\right](p_1 - \hat{p}_1)^2\right\} \\ &\approx P_n\big(X^{(n)}|\hat{p}_1\big) \cdot exp\left\{-\frac{1}{2}\,\frac{(p_1 - \hat{p}_1)^2}{\sigma^2/n}\right\} \end{aligned} \tag{B.1.3}$$

where $\sigma^2 = \left[\dfrac{-L_n''(\hat{p}_1)}{n}\right]^{-1}$. Then, multiply the formula B.1.3 by $\pi_1(p_1)$ and integrate with respect to p_1

$$\begin{aligned} \int P_n\big(X^{(n)}|p_1\big)\,\pi_1(p_1)\,dp_1 &\approx P_n\big(X^{(n)}|\hat{p}_1\big) \cdot \pi_1(\hat{p}_1) \cdot \sqrt{2\pi\sigma^2/n}\int \frac{1}{\sqrt{2\pi\sigma^2/n}} exp\left\{-\frac{1}{2}\,\frac{(p_1 - \hat{p}_1)^2}{\sigma^2/n}\right\} dp_1 \\ &\approx P_n\big(X^{(n)}|\hat{p}_1\big) \cdot \pi_1(\hat{p}_1) \cdot \sqrt{2\pi} \cdot \sqrt{\sigma^2/n} \cdot 1 \end{aligned}$$

Now by writing $\sigma_n = \sqrt{\sigma^2/n}$, $m_1\big(X^{(n)}\big)$ has the form below.

$$m_1\big(X^{(n)}\big) = \sqrt{2\pi}\,P_n\big(X^{(n)}|\hat{p}_1\big)\,\pi_1(\hat{p}_1)\,\sigma_n\big(1 + O(1/n)\big) \tag{B.1.4}$$

As $n \to \infty$ under H_0, $\hat{p}_1 \to p_0$ and $P_n\big(X^{(n)}|\hat{p}_1\big) \to P_n\big(X^{(n)}|p_0\big)$, hence

$$\begin{aligned}\frac{m_1\big(X^{(n)}\big)}{m_0\big(X^{(n)}\big)} &\approx \sqrt{2\pi}\,\frac{P_n\big(X^{(n)}|\hat{p}_1\big)}{P_n\big(X^{(n)}|\hat{p}_0\big)}\,\pi_1(\hat{p}_1)\,\sigma_n\big(1+O(1/n)\big)\\ &\approx \sqrt{2\pi}\,\pi_1(\hat{p}_1)\,\sigma_n\big(1+O(1/n)\big)\end{aligned} \tag{B.1.5}$$

In the formula B.1.5,

(a). for a Local prior, $\pi_1(\hat{p}_1)$ is a fixed constant as $\hat{p}_1 \to p_0$, i.e., $\pi_1(\hat{p}_1) \to \pi_1(p_0) \neq 0$. Therefore, for a Local prior $\dfrac{m_1\big(X^{(n)}\big)}{m_0\big(X^{(n)}\big)} \to 0$ at rate $n^{-1/2}$.

(b). for a Non-local prior $\pi_1(p_0) = 0$ and as $\hat{p}_1 \to p_0$ $\pi_1(\hat{p}_1) \to \pi_1(p_0) = 0$ at some rate.

Using the formula for $\pi_1(p_1)$ given in equation 3.2.9 under the chapter 3.2.1, for a large n under H_0

$$\begin{aligned}\pi_1(\hat{p}_1) &= \frac{1}{(1-p)}\int \pi(\hat{p}_1|w)\left[I\big(|LOR|>K\big)+I\big(|LOR|<K\big)\frac{|LOR|}{K}\right]\pi(w)dw\\ &= \frac{1}{(1-p)}\int \pi(\hat{p}_1|w)\left[0+1\cdot\frac{|LOR(\hat{p}_1,p_0)|}{K}\right]\pi(w)dw\\ &\to \frac{1}{(1-p)}\int \pi(\hat{p}_1|w)e^{-w}dw \times \frac{1}{K}\lim_{n\to\infty}|LOR\big(\hat{p}_1,p_0\big)|\end{aligned}$$

Consider $\Big[LOR\big(\hat{p}_1,p_0\big)\Big]^2$ and use Taylor series expansion around p_0 as a function of $\hat{p}_1$.

$$\begin{aligned}\Big[LOR\big(p_1,p_0\big)\Big]^2 &= 0+\left[2\,LOR\big(\hat{p}_1,p_0\big)\frac{\partial LOR\big(\hat{p}_1,p_0\big)}{\partial\hat{p}_1}\right]_{\hat{p}_1=p_0}\cdot\big(\hat{p}_1-p_0\big)\\ &\quad+\left[2\left(\frac{\partial LOR\big(\hat{p}_1,p_0\big)}{\partial\hat{p}_1}\right)^2+2LOR\big(\hat{p}_1,p_0\big)\frac{\partial^2 LOR\big(\hat{p}_1,p_0\big)}{\partial\hat{p}_1^2}\right]_{\hat{p}_1=p_0}\cdot\big(\hat{p}_1-p_0\big)^2\\ &= 2\left[\frac{\partial LOR\big(\hat{p}_1,p_0\big)}{\partial\hat{p}_1}\right]^2_{\hat{p}_1=p_0}\cdot\big(\hat{p}_1-p_0\big)^2\end{aligned}$$

We know that as $\hat{p}_1 \to p_0$, $\dfrac{\sqrt{n}(\hat{p}_1 - p_0)}{\sqrt{p_0(1-p_0)}} \to N(0,1)$. So that, $n(\hat{p}_1 - p_0)^2$ goes to a constant in probability and $(\hat{p}_1 - p_0)^2 \approx O(1/n)$ under H_0 as $n \to \infty$.

$$\therefore \Big[LOR\big(\hat{p}_1, p_0\big)\Big]^2 \to O(1/n) \text{ and } LOR\big(\hat{p}_1, p_0\big) \to O(1/\sqrt{n})$$

Hence, $\pi_1(\hat{p}_1) \to 0$ at a rate of $O(1/\sqrt{n})$. Therefore, under the Thrshold prior, from formula B.1.5, $\dfrac{m_1\big(X^{(n)}\big)}{m_0\big(X^{(n)}\big)} \to 0$ at rate n^{-1}.