Statistics and Its Interface Volume 16 (2023) 517–529

A default Bayesian multiple comparison of two binomial proportions

Emrah Gecili* and Siva Sivaganesan

We consider a default Bayesian approach to multiple testing of equality of two binomial proportions. While our approach is motivated by a scenario where one proportion corresponds to an experimental condition and the other to a control, we find it is also reasonable for comparing two proportions in general. We consider a selection of priors under the alternative(s) including the intrinsic prior and a newly proposed "mode-based" Beta prior, and investigate their properties in terms of certain desirable characteristics that we specify for default priors. We also develop priors for the hyperparameters based on the conventional hyperprior used for normal means multiple testing. We also consider a computationally more efficient empirical Bayes approach using the intrinsic prior and the proposed Beta prior. We use repeated simulation and real data sets to evaluate and illustrate the approach, and compare certain frequentist characteristics of the results based on intrinsic and mode based Beta prior using full Bayes and empirical Bayes approaches. Additionally, the results from the Bayesian approach are compared with a commonly used frequentist procedure using conventional thresholds in the respective settings. Overall, we find that the proposed mode-based Beta prior is a suitable default prior for multiple testing of equality of two proportions.

AMS 2000 subject classifications: Primary 62F15, 62C10.

Keywords and phrases: Binomial proportions, Multiple comparison, Objective Bayesian, Prior distribution.

1. INTRODUCTION

Bayesian approach to testing the equality of a single pair of proportions has received considerable attention for long, and until recently. But, the Bayesian approach to multiple testing of several pairs of proportions has received relatively little attention. Here, we consider a default Bayesian approach to multiple testing of equality of two binomial proportions under several similar conditions. We first focus on the choice of prior(s) under the alternative hypothesis in the context of single testing of equality of two proportions, conditional on a hyperparameter which controls the closeness of the prior to the null hypothesis. We seek priors that satisfy certain desirable characteristics which we specify for priors in a single testing context. We then extend the prior specification to the context of multiple testing of several pairs of proportions by using an exchangeable prior for the pairs of proportions under the alternatives along with the specification of a prior for the hyperparameter.

One motivating example for this work is the DNA sequence data example of Tarone (1990). Here, the author compared the proportions of nucleotide changes in transcripts from the control and the study cells to determine if the transcribed RNA in the study cells differs from that in the control cells, for different nucleotide sites of a gene of interest. Although our approach is particularly suitable for a setting where the interest is in comparing proportions under treatment and control, we also argue that the approach is also suitable for comparing two proportions in more general settings.

There are numerous articles in the literature on applications of Bayesian multiple testing, many involving Bayesian variable selection, including, Müller, Parmigiani, & Rice (2007), Bayarri et al. (2012), and Rockova & George (2014). On Bayesian multiple testing of two binomial proportions, Chen & Sarkar (2004) used a step-down approach involving Bayes factors for families of hypotheses. They used normal approximation to the binomial distribution and used constant prior probabilities for each null hypotheses to calculate Bayes factors, arguing that the step-down procedure itself would correct for multiplicity. Here, we use a more formal approach to multiple testing by accounting for multiplicity through the use of a prior for the common unknown probability of null hypotheses, and also use the exact likelihood. We also focus on two-sided alternatives as in Tarone (1990). Bayesian approaches to single testing of binomial proportions has a rich literature. These include, among others, Howard (1998) who used normal priors for the log of odds ratio, and Hsiao, Lee, & Kass (2005) who used a Beta prior. There have been several interesting more recent developments on Bayesian approach to testing about binomial proportion and contingency tables using intrinsic prior approach, starting with Casella & Moreno (2009). Later, Consonni & Rocca (2008) used intrinsic prior approach to testing equality of correlated proportions arising in pre- and post- polls, Consonni, Foster, & Rocca (2013) used intrinsic prior approach to non-local priors in the context of testing

*Corresponding author.

equality of two proportions, and Moreno, Vazquez, & Negrin (2014) used intrinsic prior in the context of meta analysis of binomial data. Other papers include Sverdlov, Ryeznik, & Wu (2014) and Pham-Gia, Thin, & Doan (2017). Finally, for a discussion of default (or, objective) Bayesian analysis including motivation and a historical perspective, we refer the interested readers to Berger (2006).

The paper is organized as follows. In Section 2, we describe the model and the form of the joint prior for two proportions for a single test conditional on a hyperparameter, and list certain desirable properties and evaluate them in terms of these properties. In Section 3, we choose priors for the hyperparameter for the priors deemed desirable from Section 2. In Section 4, we provide details of computation of posterior probabilities of the null hypotheses and an outline of an empirical Bayes approach for faster computing. In Section 5, we give examples based on simulated and real data, and end with a conclusion.

2. MODEL AND PRIORS FOR BAYESIAN MULTIPLE TESTING

Analogous to the setting described in the DNA sequence data example of Tarone (1990), we let x_{0i} and x_{1i} be the sample counts at position i for groups 0 and 1, respectively, and assume $x_{0i} \sim Binomial(n_{0i}, p_{0i})$ and $x_{1i} \sim Binomial(n_{1i}, p_{1i})$, $i = 1, ..., k$, and are independent given p_{0i} and p_{1i}. Our interest is in simultaneous testing of the hypotheses

$$(1) \qquad H_{0i} : p_{0i} = p_{1i} \quad vs \quad H_{1i} : p_{0i} \neq p_{1i}, \ i = 1, ..., k.$$

Multiplicity adjustment in Bayesian multiple testing is achieved by choosing an appropriate prior for the prior probability of the null hypotheses, and for the parameters in the models specified by the hypotheses. We let $p = P(H_{0i})$, for $i = 1, ..., k$, be the common probability of the null hypotheses, and use $Beta(\alpha, \beta)$ prior for p, with the default choice of $\alpha = \beta = 1$. We use the following notation and general form of priors for the rest of the parameters;

- Under H_{0i}: $p_{0i} \sim Uniform(0, 1)$, independently of each other, with $p_{1i} = p_{0i}$, for all i,
- Under H_{1i}: Conditional on hyperparameter τ, prior for (p_{0i}, p_{1i}) is $\pi_1(p_{0i}, p_{1i}|\tau)$, and for τ is $\pi(\tau)$.

The model specification will be completed by choosing an appropriate choice for $\pi_1(p_{0i}, p_{1i}|\tau)$ and $\pi(\tau)$.

3. CHOICE OF PRIORS FOR (p_{0i}, p_{1i}) UNDER THE ALTERNATIVE HYPOTHESES

We first focus on the specification of the prior $\pi_1(p_{0i}, p_{1i}|\tau)$ for (p_{0i}, p_{1i}) under H_{1i} for single test of equality of two proportions (for a fixed i), conditional on hyperparameter τ. For convenience, without loss of generality, we drop the suffix i in the rest of this and next subsection. Most of the priors used for estimation are not appropriate for testing hypotheses, because the null hypothesis is not taken into account in the formulation of the prior.

3.1 Desirable properties

Here we seek certain desirable properties for a default prior $\pi_1(p_0, p_1|\tau)$ to satisfy. The first one is that the prior be centered around the null. This is a widely accepted characteristic in the literature for objective Bayesian testing starting with the recommendation in Jeffreys (1961). For instance, consider testing a point null hypothesis $\theta = 0$ versus the alternative hypothesis that $\theta \neq 0$. In this context, 0 is a special value for θ among the all possible values of θ, and in the absence of any other prior information including the absence of any other special values for θ, it makes sense that the values of θ closer to 0 should a priori be considered more likely than those away from 0. This heuristic argument leads to the conclusion that prior for θ under the alternative hypothesis ought to be centered around 0, and decreasing away from 0. This notion is also borne out, naturally, by the intrinsic prior method of deriving an objective prior under the alternative hypothesis essentially with the only knowledge of the point null hypothesis under consideration, Berger and Pericchi (1996). The testing or model selection priors recommended by many authors all satisfy this property. These include the family of g-priors recommended in Zellner and Siow (1980), and later by Liang et al. (2008) and Bayarri et al. (2012) among others. The well-known intrinsic prior methodology used to construct reasonable objective priors for testing hypotheses are popular partly because it provides a way of converting a estimation priors into proper testing priors that are centered around the null, e.g., Casella & Moreno (2009) developed intrinsic priors that are centered around the null in the context comparing contingency tables.

On the other hand, it is also important that the prior is not overly concentrated around the null and allows flexibility for p_1 to deviate from p_0 by allowing a wide range of values for the correlation between p_0 and p_1. For instance, correlation only in a high range (near 1) would indicate that the prior would inherently constrain both parameters to be very close to each other rather than have the flexibility for them to differ based on the likelihood. (We note priors satisfying the first property would typically allow only positive correlation.) This property is also used in Moreno, Vazquez, & Negrin (2014) for motivation of a prior they used for proportions in the context of meta analysis. Similarly, we also seek that the prior allows flexibility near the boundary by allowing p_1 to deviate from p_0 when p_0 is close to either 0 (or 1),i.e., a priori, $|p_1 - p_0|$ does not go to 0 as p_0 goes to 0. This would thus allow the same flexibility a posteriori, rather than constraining p_1 also to be close to 0 (or, 1) when p_0 is, regardless of the likelihood. These properties can be formally stated as follows;

(i) The prior is centered around the null region $p_1 = p_0$, and decreasing away from the null region,
(ii) The prior allows a wide range of values between 0 and to 1 for correlation between p_1 and p_0, and
(iii) $Pr(|p_1 - p_0| > \varepsilon | p_0) > 0$ for some $\varepsilon > 0$, as $p_0 \to c \in \{0, 1\}$.

3.2 Choice of priors

We consider four choices for $\pi_1(p_0, p_1|\tau)$, three of which have been used in the literature in the context of single testing, and another that we propose anew. They are listed below along with whether they satisfy the three desirable properties.

Intrinsic Prior:

Intrinsic prior approach to testing, developed by Berger & Pericchi (1996) and Moreno, Bertolino, & Racugno (1998), is an attractive and natural way to derive an objective prior for testing hypotheses. In the context of testing hypotheses about proportions. Casella & Moreno (2009) used the approach to develop objective priors for testing independence in two way contingency tables (single testing of proportions). This was followed by a series of papers on the use of intrinsic priors in the context of testing single hypotheses about equality of proportions and contingency tables. They include Consonni & La Rocca (2008), Moreno, Vazquez, & Negrin (2014) and Consonni, Foster, & Rocca (2013).

The intrinsic prior for (p_0, p_1) using independent uniform priors as default prior for p_0 and p_1 can be written as, see Casella & Moreno (2009),

(2)

$$\pi_1(p_0, p_1|m) = \sum_{s=0}^{m}\sum_{t=0}^{m} \binom{m}{s}\binom{m}{t} \frac{\Gamma(s+t+1)\Gamma(2m-s-t+1)}{\Gamma(2m+2)} \times Be(p_0|s+1,\, m-s+1)\, Be(p_1|t+1,\, m-t+1), \quad m \geq 1.$$

where $Be(x|a, b)$ is Beta distribution with parameters a, b, and m is an integer representing the (imaginary) training sample size used in the construction of the prior. We can also show that (see Appendix A),

(3)

$$E(p_1|p_0) = \frac{2 + 2m + p_0 m^2}{(2+m)^2} \quad \text{and} \quad corr(p_0, p_1) = \frac{m^2}{(2+m)^2}.$$

Figure A1 in Appendix A displays the intrinsic prior for two different choices of m. This prior satisfies the first property in Section 3.1. It centered around the region of the null hypotheses, $p_0 = p_1$, and decreases away from the null. The degree of concentration is controlled by the training sample size m which we regard as a discrete hyperparameter. It is increasingly concentrated around the region of the null hypothesis with increasing m.

By (3), the correlation ρ, between p_0 and p_1 takes values in the range of (1/9,1), for $m \geq 1$, which is very close to (0,1), and hence satisfies the second property. It is also easy to see that the marginal distributions of p_0 and p_1 are both uniform over $(0, 1)$. Thus, the conditional distribution of p_1 given p_0, as seen from (2), is a mixture of Beta distributions. This can be used to show that the intrinsic prior also satisfies the third property in Section 3.1, and thus it satisfies all three properties.

Consonni, Foster, & Rocca (2013) stated that the results from the intrinsic prior and a suitable Beta prior centered at p_0 would be similar for single testing problem of binomial proportions. However, the probability mass transfer toward p_0 takes place more smoothly under the intrinsic prior than under a Beta prior with increasing precision, because the intrinsic prior is a mixture of Beta distributions.

Mode-Based Beta Prior: We propose a prior which is a mixture of Beta distributions given by

(4)

$$\pi_1(p_1|p_0, r) \sim Beta(rp_0 + 1, r(1-p_0) + 1), \qquad p_0 \sim Uni(0, 1)$$

where $r > 0$ is a hyperparameter. This prior is similar to a prior used by Kass & Hsiao (1993). With a prior placed on r, this prior is also a mixture of Beta distributions. For this prior, the mode of the conditional distribution of $p_1|p_0$ is p_0. As shown in Figure A3, Appendix B, the joint prior is centered and nearly symmetric around the the region of the null hypothesis for all values of r and thus satisfies the first property in Section 3.1.

The conditional variance of p_1 given p_0 for this prior is a decreasing function of r and thus prior becomes more concentrated around the null hypothesis as r becomes larger. We can also show the following expressions for the correlation and marginal variance for this prior distribution,

(5)

$$corr(p_0, p_1) = r\sqrt{\frac{r+3}{(r+2)(r^2+3r+6)}} \quad \text{and} \quad Var(p_1) = \frac{r^2+3r+6}{12(r+2)(r+3)}.$$

From the above it is clear that the prior allows the full range of (positive) values for the correlation satisfying the second property. Since the condional prior in (4) remains a valid Beta distribution in the limits as p_0 approaches 0 or 1, this prior satisfies the third property, and thus all three properties in Section 3.1.

In addition, we also considered three other priors. One is the mean-based Beta prior, similar to the mode-based Beta prior except that for this prior the mean of the conditional distribution of p_1 given p_0 is equal to p_0. The other is logit-normal prior which assigns a normal distribution for

the conditional distribution of the odds-ratio of p_1 to p_0, given p_0. The third one is a Farlie–Gumbel–Morgenstern (FGM) parametric class of probability distribution which has uniform marginals similar to the intrinsic prior, and was used for meta analysis for sparse discrete Binomial data by Moreno, Vazquez, & Negrin (2014). We show (in Appendix C) that the mean-based Beta prior does not satisfy the second and third properties, the logit-normal prior does not satisfy the third property (in Appendix D), and the FGM family prior does not satisfy the second property (in Appendix E). Moreno, Vazquez, & Negrin (2014) found this FGM type prior to have too large Type -I error probability and too small power compared to the intrinsic prior in the context of meta analysis.

We therefore focus on the the intrinsic prior and the mode-based Beta prior in the rest of the paper.

3.3 Specification of prior for the hyperparameter

We now turn to specifying a hyperprior for the parameter τ for the intrinsic prior and mode-based prior, namely the parameter m for the intrinsic prior and the parameter r for the mode-based prior.

Our choice of the hyperprior stems from the analogous prior used by Scott & Berger (2006) in the context of multiple testing of normal means, where $y_i|\mu_i, \sigma^2 \sim N(\mu_i, \tau^2)$ and the mean(s) μ_i was assigned $N(0, \tau^2)$ prior, and the hyperparameter τ^2 was assigned the prior

$$\pi(\tau^2|\sigma^2) = \frac{\sigma^2}{(\sigma^2+\tau^2)^2}, \qquad \tau^2 > 0. \tag{6}$$

The authors motivated this prior as a suitable approximation to the objective estimation-prior for the variance in the second level of a hierarchical normal model, Berger and Strawderman (1996). This prior has a peak at $\tau^2 = 0$ and decreases away from 0. Default priors for random effects variance recommended in the literature such as the shrinkage prior, Daniels (1999), and the half-t prior, Gelman (2006), also share the same feature.

To obtain a suitable default prior for the hyperparameter τ^2, we follow the form of the prior (6). For this, we let $\sigma^2 = E(Var(x_{1i} - x_{0i}))$, the variance of $(x_{1i} - x_{0i})$ marginalized over p_{1i} and p_{0i}. Also, we let $\tau^2 = Var(p_{1i} - p_{0i})$, conditional on the hyperparameter for a prior in Section 3.2. We use these specifications to construct a prior for τ^2 as in (6).

Prior for the Hyperparameter in the Mode-Based Beta Prior:

For this prior, $\sigma^2 = E(Var(x_{1i} - x_{0i})) \approx 1/3$, and $Var(p_{1i}|r) \approx 1/12$ where we use $\approx$ to mean approximate equality. We find it more convenient to use $w = 1/r$ henceforth instead of r. We can verify that, $\tau^2 = Var((p_{1i} - p_{0i})|r) \approx 1/(3(w^{-1}+2))$ which has a range of $(0, 1/6)$. Following the form of (6), we propose a prior for τ^2 given by

$$\pi(\tau^2) = \frac{3\sigma^2}{(\sigma^2+\tau^2)^2}, \quad 0 < \tau^2 < 1/6,$$

which leads to the following prior for w

$$\pi(w) = \frac{3}{(1+3w)^2}, \quad 0 < w < \infty. \tag{7}$$

Note that τ^2 is increasing in w and the prior $\pi(w)$ is monotonically decreasing with peak at $w = 0$ (or $\tau^2 = 0$).

Prior for the Hyperparameter in the Intrinsic Prior:

Using (3) and the fact that p_0 and p_1 each has uniform $U(0, 1)$ distribution as it's marginal, we have

$$Var(p_{1i} - p_{0i}|m) = \tau^2(m) = 2(m+1)/3(m+2)^2, \quad m = 1, 2, ...$$

which increases in m and approaches to zero as m approaches to ∞. Thus for a (discrete) prior for $\tau^2(m)$ to peak at 0, it would require a prior for m that is increasing as m approaches to ∞, an untenable condition since we also require the prior to be proper. As a way around, we could bound the range of m, but, we found that results using such priors are sensitive to the choice of the bound.

Alternatively, we may use an approach similar to the one in Moreno, Vazquez, and Negrin (2014) to derive a prior for m. First, we consider a continuous prior for τ^2, in the form of (6). With $\sigma^2 = 1/3$ as before, the resulting prior for τ^2 is

$$\pi(\tau^2) = (15/2)\frac{1}{(1+3\tau^2)^2}, \quad 0 < \tau^2 < 2/9. \tag{8}$$

Now, we construct a discrete prior for m by discretizing (8) and using the approximation $\tau^2(m) \approx 2/(3(m+2))$. This gives the prior

$$\pi(m) = \frac{5}{(m+4)(m+5)}, \quad m = 1, 2, ... \tag{9}$$

This prior induces a prior for τ^2 that has negligible mass near 0, and is increasing in τ^2 putting most of the mass at high values, (Figure A2 in Appendix A), unlike the priors referred to earlier. For the rest of the paper, we use the mode-based Beta prior with the hyperprior (7) for w and the intrinsic prior with hyperprior (9) for m in our simulation and real data examples, and compare the results.

More on the use of Mode-Based Beta Prior

While the mode-based Beta prior (4) was intended for use in the context for testing equality of the proportions for a treatment (p_1) and a control (p_0). It is also suitable for testing the equality of two proportions in a general context. In (4), marginal of p_0 is $U(0, 1)$ distribution the conditional distribution of p_1 given p_0 is a Beta distribution with mode

p_0. For this joint prior, we can verify that the marginal prior for p_1 is also unimodal and symmetric around 0.5 (Appendix B), with a variance of 0.268 (by numerical calculation) which is very close the $U(0,1)$ distribution. We also verified using numerical evaluation that the conditional distribution of p_0 given p_1 is unimodal with mode increasingly closer to p_0 as r increases. In addition, we plotted the conditional means of p_1 (given p_0) versus p_0 and the conditional mean of p_0 (given p_1) versus p_1 for the prior (4), Figure A4 in Appendix B. The plots show that these conditional expectations are equal to each other for a big range of given values of p_0 and p_1. These properties suggest an approximate symmetry between p_0 and p_1 in terms of the conditional distributions of one versus the other and symmetric distributions for the marginal of p_1.

Moreover, calculation with the prior (4) with the roles of p_0 and p_1 switched using simulated data gave nearly the same posterior probabilities in a variety of settings. As an illustration, we plotted the posterior probabilities of the null hypotheses H_0 using both priors and plotted them against each other for a simulated data with $m = 40$ tests, see Figure A5 in Appendix B.

The properties of the mode-based prior and the numerical evidence provided above suggest that this prior is to a certain degree symmetric between the two proportions and the choice of which of the two proportions we choose as p_0 has little difference in the answers. Hence, the mode-based Beta prior is also suitable for use in a wider context of testing equality of two proportions.

4. POSTERIOR PROBABILITY AND AN EMPIRICAL BAYES APPROACH

The general form of the multiplicity adjusted posterior probability of H_{0i} for mean-based Beta prior (it has similar form for the other priors) is given by

$$P(H_{0i}|x) = \frac{\int_0^1 p\, g(p, I_k \setminus \{i\})\, dp}{\int_0^1 g(p, I_k)\, dp} \tag{10}$$

where

$$g(p, I_k) = \int \prod_{j \in I_k} \left[p + K_j^* \int_0^1 p_{0j}^{x_{0j}} (1-p_{0j})^{n_{0j}-x_{0j}} h(r, p_{0j}) dp_{0j} \right] \pi(r) dr,$$

$$I_k = \{1, ..., k\}, h(r, p_{0j}) = \frac{Beta\,(x_{1j} + r p_{0j} + 1,\ r(1-p_{0j}) + n_{1j} - x_{1j} + 1)}{Beta\,(r p_{0j} + 1,\ r(1-p_{0j}) + 1)},$$

and

$$K_j^* = (1-p)\Gamma(n_{0j} + n_{1j} + 2) / \{\Gamma(x_{0j} + x_{1j} + 1)\Gamma(n_{0j} + n_{1j} - x_{0j} - x_{1j} + 1)\}.$$

We assume that a decision rule to choose between H_{0i} and H_{1i} is based on a $0-K$ loss and is of the form below for a pre-set threshold c,

$$\text{Choose } H_{1i} \text{ if } P(H_{0i}|x) < c, \text{ else, choose } H_{0i}. \tag{11}$$

Here, we use the conventional threshold of $c = 0.5$ when calculating error rates.

4.1 An empirical Bayes approach

The fully Bayesian (FB) approach discussed above is computationally considerably slower than the frequentist methods. A computationally more efficient alternative to FB is the empirical Bayes (EB) approach. Instead of assigning priors to some hyper-parameters and integrating out with respect to them, using their "consistent" estimates instead may simplify computation and reduce computation time to obtain posterior probabilities of the hypotheses. Hence, with the EB approach, one does not have to specify a prior for the hyperparameter w. In our numerical study, we also use the EB approach as well to see if it can provide a good approximation to the FB approach considered in the previous sections.

We considered a parametric EB approach by estimating p and w. The estimates of p and w are obtained by maximizing the logarithm of the marginal likelihood for (p, w), given by

$$\begin{aligned} & m\,(\underline{x}_0, \underline{x}_1 | p, w) \\ &= \prod_{i=1}^{k} P(\underline{x}_i | p, w) \\ &= \prod_{i=1}^{k} \Big[p \int P(\underline{x}_i | p_{0i}, w) \pi(p_{0i}) dp_{0i} \\ &\quad + (1-p) \int\int P(\underline{x}_i | p_{0i}, p_{1i}, w) \pi_i(p_{0i}, p_{1i} | w) dp_{0i} dp_{1i} \\ &= \prod_{i=1}^{k} \Big[p \int P(x_{0i} | p_{0i}) P(x_{1i} | p_{0i}) \pi(p_{0i}) dp_{0i} \\ &\quad + (1-p) \int\int P(x_{0i} | p_{0i}) P(x_{1i} | p_{1i}) \pi_i(p_{0i}, p_{1i} | w)\, dp_{0i} dp_{1i} \Big]. \end{aligned}$$

All computation in this paper was done using *R*. To compute the posterior probability of the null hypothesis using the FB approach and mode based prior, equation (10), we use the *integrate* function in *R* to do the integrals with respect to both p_{0j} and $r = 1/w$, and Monte Carlo (MC) integration was used to estimate the integration with respect to p.

For the EB approach, we use the *mle* function from the *stats4* package in *R* that utilizes the *optim* function to find the minimum of the negative log-likelihood and obtain the maximum likelihood estimators of p and w (m for the intrinsic prior).

Table 1. Average posterior probabilities of the null hypotheses for the Intrinsic prior (Intr) and mode based prior (MB). p_{0i}'s are generated from $Uni(0.2, 0.5)$, The number of tests k is 20, of which, for the first 15, the null hypotheses are true, and for the last 5 the alternative hypotheses are true. p_{0i}'s are generated from $Uni(0.2, 0.5)$ and p_{1i}'s are fixed so that the odds-ratio between the p_{0i} and p_{1i} is 2. The Binomial sample sizes are set at $n = 20$. The three dots indicate the omitted part of the table.

	$P(H_{0i}\|x)$ for true nulls											$P(H_{0i}\|x)$ for true alternatives				
i	1	2	3	4	5	...	11	12	13	14	15	**16**	**17**	**18**	**19**	**20**
Intr	0.69	0.71	0.70	0.69	0.70	...	0.70	0.71	0.69	0.69	0.70	**0.63**	**0.64**	**0.63**	**0.63**	**0.59**
MB	0.56	0.58	0.55	0.54	0.55	...	0.56	0.55	0.55	0.58	0.55	**0.49**	**0.49**	**0.49**	**0.48**	**0.46**

5. SIMULATION STUDIES AND REAL DATA EXAMPLES

In this section we report the results of two simulation studies and two real data examples. In the first simulation study, we compare the mode based Beta (MB) prior and the intrinsic (Intr) prior in terms of the posterior probabilities of null hypotheses in a specific multiple testing setting. In the second, we compare certain frequentist characteristics of the Bayesian approach(es) using the mode-based prior and the intrinsic prior and a frequentist approach, namely, the Discrete Benjamini-Hochberg (DBH) method of Heyse (2011). We also apply the empirical Bayes approach and use the results to assess its accuracy as an approximation to the fully Bayes approach. We end this section with two real data examples.

5.1 Repeated sampling and frequentist characteristics

Simulation Study 1

This simulation study is to compare the posterior probabilities of the null hypotheses based on the intrinsic and mode-based Beta priors in multiple testing of two proportions. We used a specific simulation setting described below and calculated the average posterior posterior probabilities of each of the null hypotheses over repeated sampling under the given setting.

The simulation setting considered involves $k = 20$ tests each of which for comparing two proportions p_{0i} and p_{1i}, $i = 1, .., k$. We assume that the null hypotheses are true for the the first 15 tests, i.e., $p_{1i} = p_{0i}$ for $i = 1, ..., 15$, and the alternative hypotheses are true for the last five tests, We set the values of p_{0i}'s, $i = 1, ..., k = 20$, by generating them from $Uni(0.2, 0.5)$, independently, and set $p_{1i} = p_{0i}$ for $i = 1, ..., 15$. For $i = 16, ..., 15$, we set the values of p_{1i} so that the the odds-ratio (OR) between p_{1i} and p_{0i} is 2. Using these parameter settings, we simulated x_{0i} and x_{1i}, independently, from $Binomial(n = 20, p_{0i})$ and $Binomial(n = 20, p_{1i})$ for $i = 1, .., k$ and calculated the posterior probabilities of the $k = 20$ null hypotheses. We did 500 replications of the above simulation and computation using the same fixed proportions and obtained the average posterior probabilities of the k null hypotheses.

The Table 1 presents the average posterior probabilities of the true null hypotheses ($i = 1, ..., 15$) and the false null hypotheses ($i = 16, ..., 20$) over the 500 repeated simulation. We see that they are consistently higher for the intrinsic prior than for the mode-based prior. The same phenomenon was observed in other similar simulations that we carried out. In Table 1, we omitted part the posterior probabilities of true nulls in the middle (indicated by the three dots) which were similar to what was observed for other true nulls.

Simulation Study 2

Here, we report on a simulation study with two settings similar to the two real data examples given later in this section. The first simulation setting is based on the DNA sequence data, Tarone (1990), which has a small number of tests with small sample size for each test, and the the second simulation setting is based on the HIV data, Gilbert (2005), which has a large number of tests and large sample size for each test. We used repeated sampling and calculated the following frequentist characteristics; average number of true discoveries (ATD), average number of false discoveries (AFD), false discovery rate (FDR), and false negative rate (FNR) using the default threshold of 0.5 for the posterior probability of null hypothesis (for rejecting the null hypothesis). In addition, we also calculated the frequentist characteristics for the frequentist DBH approach using the conventional cut-off 0.05 for the p-values. But, the posterior probabilities of null hypotheses and p-values are not directly comparable and hence neither are the use of these cut-offs. We merely use them to see how the use of the aforementioned conventional choices in each paradigm would compare in terms of their frequentist characteristics. and the results need only be interpreted in this limited context.

Simulation Setting (i) Here, we fix the numebr of tests $k = 9$ and Binomial sample size $n = 10$, as in the DNA sequence data example. We also set $p_{0i} = \hat{p_{0i}}$, i=1,...,9 where $\hat{p_{0i}} = x_{0i}/n_{0i}$, based on the DNA sequence data. For the first k_0 tests, we set the null hypotheses as true, ($p_{0i} = p_{1i}$, i=1, ... k_0), and for the last k_1 tests, ($k_1 = 9 - k_0$), we set the alternatives as true. For the last k_1 sets, we set the values of p_{1i}'s to make the odds-ratio of p_1 to p_0 equal to 5. We did this simulation for each of two choices of k_1, $k_1 = 3$ and 6, and for each case, we simulated 500 data sets, and

Table 2. The results of Simulation (i) with $k = 9$, $k_1 = 3$, 6 and $n = 10$. For the first k_0 tests, we set the null hypotheses as true, ($p_{0i} = p_{1i}$, i=1, ... k_0), and for the last k_1 tests, ($k_1 = 9 - k_0$), the alternatives are set as true. For the last k_1 sets, we set the values of p_{1i}'s to make the odds-ratio of p_1 to p_0 is 5. MB and EBMB stand for FB and EB methods based on the mode-based prior. Intr and EBIntr stand for FB and EB methods based on the intrinsic prior and DBH stands for the discrete Benjamini-Hochberg method

	$k_1 = 3$	**MB**	**EBMB**	**Intr**	**EBIntr**	**DBH**	$k_1 = 6$	**MB**	**EBMB**	**Intr**	**EBIntr**	**DBH**
ATD		2.25	2.24	1.81	1.99	0.66		5.52	5.62	4.92	4.8	1.09
AFD		0.7	0.92	0.43	0.56	0.01		1.25	1.84	2.02	1.8	0.005
FDR		0.24	0.29	0.19	0.22	0.015		0.18	0.25	0.29	0.27	0.005
FNR		0.25	0.25	0.4	0.34	0.78		0.08	0.06	0.18	0.2	0.81

Table 3. The results of Simulation (ii) with $k = 118$, k_1 =5, 10 and n =73. Set $p_{0i} = \hat{p_{0i}}$, i=1,. . . , 118 where $\hat{p_{0i}} = x_{0i}/n_{0i}$ are the estimates from the HIV data. For the first k_0 tests, let null hypotheses be true ($p_{0i} = p_{1i}$, i=1, ... k_0), while for the last k_1 tests, we let the alternatives be true, and generated the last k_1 p_{1i}'s by using an odds-ratio of 5. MB and EBMB stand for FB and EB methods based on the mode-based prior. Intr and EBIntr stand for FB and EB methods based on the intrinsic prior and DBH stands for the discrete Benjamini-Hochberg method

	$k_1 = 5$	**MB**	**EBMB**	**Intr**	**EBIntr**	**DBH**	$k_1 = 10$	**MB**	**EBMB**	**Intr**	**EBIntr**	**DBH**
ATD		4.68	4.64	4.44	4.51	4.82		8.15	8.14	8.03	8.12	8.60
AFD		0.06	0.05	0.047	0.052	0.26		0.07	0.06	0.075	0.08	0.35
FDR		0.0127	0.011	0.0107	0.011	0.051		0.0085	0.0073	0.009	0.01	0.039
FNR		0.064	0.072	0.11	0.098	0.036		0.185	0.18	0.197	0.188	0.14

applied the methods to each data set. The summary results are given in Table 2.

Simulation Setting (ii) Here, we used a setting similar to the HIV sequence data example, and fixed the number of tests $k = 108$ and the Binomial sample size $n = 73$. We set $p_{0i} = \hat{p_{0i}}$, i=1,. . . , 118 where $\hat{p_{0i}} = x_{0i}/n_{0i}$ are the estimates from the HIV data. For the first k_0 tests, we let null hypotheses be true ($p_{0i} = p_{1i}$, i=1, ... k_0), while for the last k_1 tests, we let the alternatives be true, and generated the last k_1 values of p_{1i}'s using an odds-ratio of 5, same as in the previous setting. We did the simulation for each of two values of k_1, $k_1 = 5$ and10, and for each case, did repeated simulation of 500 data sets and applied the methods to each data set. The summary results are given in Table 3.

The two simulation settings have very different sizes; the first one with a small number of tests, $k = 9$, and a small sample size, $n = 10$ (small sizes), and the second with a large number of tests, $k = 108$, and a large sample size, $n = 73$ (large sizes). For all approaches, the Bayes, empirical Bayes, and DBH, the error rates (FDR and FNR) are smaller in the second simulation with larger values for k and n than in the first. For the Bayes approach with either prior, when the number of true alternatives, k_1, is increased while the (total) number of tests is held fixed, FDR decreased and FNR increased in each setting. This latter phenomenon can be attributed to an overall decrease in the posterior probabilities of null hypotheses due to the multiplicity adjustment induced by an increase in overall evidence in favor of alternatives.

Based on the results for the Bayesian approach using the two priors, we find that in the first setting (small sizes); for the case with $k_1 = 3$), the mode-based Beta prior gave results indicating it is less conservative (with slightly larger FDR, AFD and ATD; and smaller FNR) than the intrinsic prior. For the same setting with $k_1 = 6$, FDR and FNR were smaller, and ATD was slightly larger for, indicating that the use of the mode-based Beta prior gives an overall more accurate results than the intrinsic prior. In the second setting (large sizes) use of both priors showed no difference giving approximately the same frequentist rates.

Comparing the fully Bayes (FB) and the empirical Bayes (EB) approaches, we can see that EB approach is a good approximation to FB approach for both priors when k and n are large. When k and n are small, EB approach does not provide a good approximation to FB. Specifically, EB approach using mode based Beta prior (EBMB) yields higher number of discoveries (true and false) than its FB counterpart (MB), while EB approach using intrinsic prior (EBIntr) yields lower number of discoveries than its FB counterpart (Intr) when k and n are small.

The results using the respective conventional cut-offs for the Bayes approaches and the DBH approach in the first setting (small sizes) indicate that the Bayes approaches have larger FDR, AFD, and ATD; and smaller FNR. In the second setting (large sizes), the results are much closer, but the Bayes approach gave smaller FDR and AFD, somewhat similar values for ATD and FNR. These properties indicate that Bayes approach with the conventional cut-off 0.5 is less conservative in the sense of finding more discoveries (true and false) when k and n are small, but has an overall better

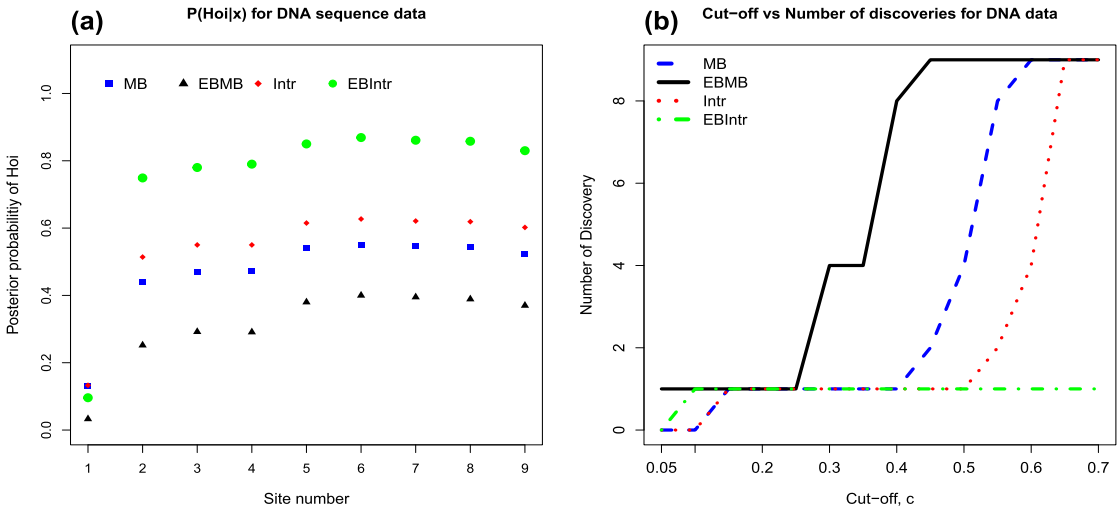

Figure 1. Results for DNA sequence data and comparison of proposed procedures. Figure (a) presents posterior probabilities of the null hypotheses. The blue square stands for MB, the black triangle is for EBMB, red diamond is for Intr, and the green filled circle is for EBIntr. Figure (b) shows the number of discovery as c changes, the blue dashed line is for the MB and the black solid line is for EBMB, red dotted line is for Intr and the dashed green line is for EBIntr.

accuracy when k and n are large, in comparison to the use of DBH.

5.2 Real data applications

We provide two applications for multiple testing problem of equality of two proportions; one using the DNA sequence data, and the other using an HIV sequence data.

(i) DNA sequence data (Tarone 1990)

The data set has 9 tests, corresponding to the nine nucleotide sites. The posterior probabilities of nulls for the 9 tests are plotted in Figure 1 (a). We note that the posterior probabilities of the null hypotheses for the mode-based Beta prior are smaller than those of the intrinsic prior, consistent with the results for the repeated simulations (i) with smaller k_1. For deciding between the null and the alternative hypotheses, we used a cut-off value c, and call it a "discovery" when the posterior probability of a null hypothesis is less than c. The number of discoveries for a range of values for c are displayed in Figure 1 (b). In particular, for the cut-off value $c = 0.5$ there are differences between the control and study cells at Sites 1, 2, 3 and 4 using the mode based prior, and at Sites 1 for the intrinsic prior, and for $c = 0.05$, there is only difference at Site 1 using the mode based prior while none using the intrinsic prior. For comparison, in Tarone (1990), the author used a modified Bonferroni procedure and found differences at Sites 3, 1, and 9 at 5% level. Additionally, DBH only yielded differences at Site 1 at 5% level.

We also note that EB approach has consistently smaller posterior probabilities for null hypotheses than FB approach for mode-based Beta prior, and larger probabilities for the intrinsic prior. This is similar to what we found earlier in the simulation study, where the EB approach did not provide a good approximation to FB when the number of tests was small.

(ii) HIV sequence data

This example is from an HIV sequence study considered by Gilbert (2005), where data were available on two sets of gag p24 amino-acid sequences, infected with subtype B HIV and subtype C HIV. For each sequence set, the degree of polymorphism at each position in the sequences can be measured by the frequency of non-consensus amino-acids at the position. The goal was to identify the positions at which the probability of a non-consensus amino-acid differs between the two sets of sequence sets; the consensus amino-acid is the modal amino-acid for the sequence set. The data consists of 146 gag p24 amino acid sequences, with 73 of them infected with subtype HIV C (Group 0), and the other 73 infected with HIV B (Group 1). Of the 231 positions in HIV gag p24, 113 have the modal amino acid in all 146 sequences, i.e. are perfectly conserved, and more details about this study can be found in Gilbert (2005). Thus, there are $k = 118$ positions that contribute information to the comparison of interest. We let p_{0i} and p_{1i} respectively be the probabilities of a non-consensus amino-acid at position i for group 0 and group 1 sequences, respectively. The goal is to test simultaneously $H_{0i} : p_{0i} = p_{1i}$ vs $H_{1i} : p_{0i} \neq p_{1i}$, $i = 1, \ldots, 118$. Here, the i^{th} null hypothesis (resp. alternative hypothesis) indicates that the probabilities of non-consensus amino-acid at i^{th} position are the same (resp. different) for Group 0 and Group 1. The results are summarized in Figure 2. We observe that the posterior probabilities of the null hypotheses for the mode based Beta prior are smaller than those for

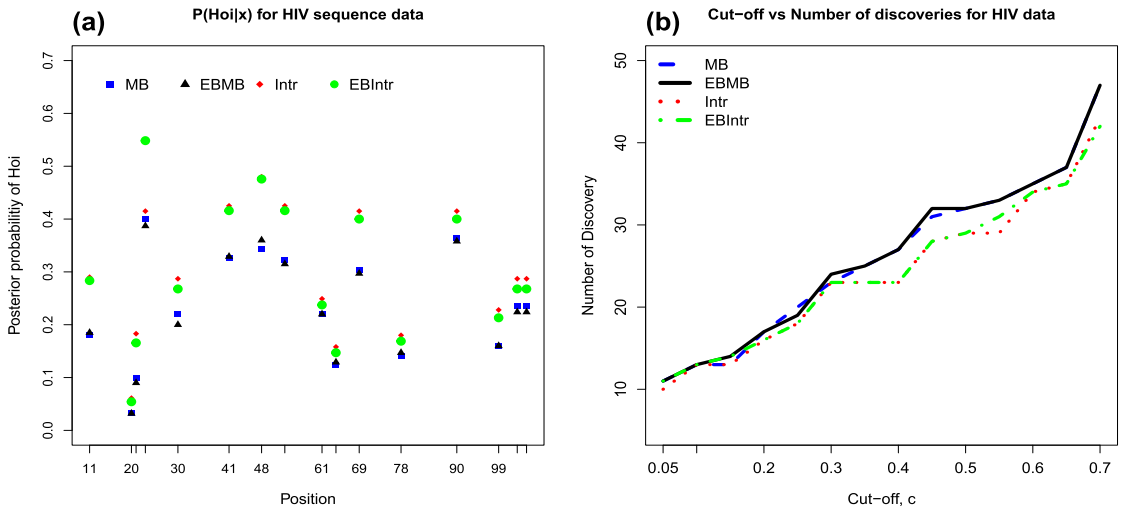

Figure 2. The posterior probabilities of nulls for the full Bayesian and empirical Bayesian methods based on the mode-based Beta prior for some of randomly selected tests (left), and the plot for the number of discoveries vs cut-off (right) for HIV data when k=118. Figure (a) presents posterior probabilities of the null hypotheses. The blue square stands for the MB, the black triangle is for EBMB, red diamond is for Intr, and the green filled circle is for EBIntr. Figure (b) shows the number of discovery as c changes, the blue dashed line is for the MB and the black solid line is for EBMB, red dotted line is for Intr and the dashed green line is for EBIntr.

the intrinsic prior, although no more than 0.1, and that the FB and EB based posterior probabilities, for each prior, are very close to each other. These results are consistent with the findings in Simulation (ii). From Figure (b), the numbers of discovery are equal to 32 for both MB and EBMB, and equal to 29 for both Intr and EBIntr for $c = 0.5$; and they are 11 and 10 when $c = 0.05$ for using mode based prior and intrinsic prior, respectively.

From these results, it appears that the EB approach can be a good approximation to the FB when the number of tests, k, is large. However, EB does not give a good approximation to FB when k is small, as observed in the simulation study and in the the DNA sequence data example.

6. CONCLUSION

We have provided a default Bayesian approach to multiple testing of equality of two proportions. We considered priors from the literature and introduced a new prior referred to as mode-based Beta prior and evaluated them in terms of three properties deemed desirable for a default prior in the context of test of hypothesis comparing two proportions. We found that the the intrinsic prior and the mode-based Beta prior satisfied the three properties and chose them as suitable default priors for multiple testing of the equality of two proportions. We also constructed default priors for the hyperparameters for the two priors. We found that prior for the hyperparameter for the mode-based Beta prior can be chosen to correspond to a prior for the variance of the differences in the proportions peak at zero and decreasing away from zero; a feature common for the default prior for variance for normal means multiple testing. But, a prior for the hyperparameter of the intrinsic prior does not correspond to a prior for the variance with the same feature. We used these two priors in two simulation studies involving repeated sampling and with two real data examples. We also used an Empirical Bayes approach as a computationally more efficient potential alternative to the fully Bayes approach and compared the results.

In the first simulation study we used repeated sampling in multiple testing scenarios by setting the null hypotheses as true for a (relatively large) number of tests and the alternative as true for the rest. In the first simulation study, we calculated the average posterior probability of the null hypothesis in each test using the two priors. The results reported here and results from other similar simulations showed that the intrinsic prior gave consistently higher posterior probabilities than the mode based prior regardless of whether the null or the alternative hypothesis is true in each test. This indicated that intrinsic prior is more conservative than the mode based Beta prior in the sense of favoring the null.

In the second simulation study we carried out repeated sampling simulation in two settings, large size and small size, and calculated certain frequentist characteristics including FDR and FNR of the Bayes approach using the two priors. For this, we used a conventional cut-off value of 0.5 for posterior probabilities of the null hypotheses. We also used the empirical Bayes approach that does not require specification of prior for the hyperparameter(s) and used it in the second repeated simulation study as well as with the

real data examples. In addition, we also calculated the same for the frequentist Discrete Benjamini-Hochberg (DBH) approach using the cut-off 0,05 for the p-values and compared the results with those of the Bayes approaches. But, the posterior probabilities of null hypotheses and p-values are not directly comparable and hence the above choice of cut-offs is only to see how the use of the conventional choices in each paradigm would compare in terms of the frequentist characteristics used, and need to be interpreted only in that limited context.

The frequentist characteristics for the two priors were nearly the same in the large size setting. In the small size setting the intrinsic prior was more conservative (favoring null more often) than the mode-based Beta prior for smaller number of true alternatives (the sparser case) and, for larger number of true alternatives there was an overall more accurate performance for the the mode-based Beta prior. In the real data examples, the intrinsic prior also gave larger posterior probabilities for the null hypotheses in most tests, consistent with findings from the first simulation study. We also found that the empirical Bayes approach did not provide good approximation to the fully Bayes approach either for the frequentist characteristics in the simulation study or for the posterior probabilities in the real data examples. This is consistent with the results of the second simulation study for this setting. However, it gave good approximations in the large size setting for both priors, both in the repeated simulation and in the real data example. This would be useful because the computation for the fully Bayes approach is slower for larger sample sizes. For instance, the empirical Bayes approach took only 5 seconds while the fully Bayes approach took more about 900 seconds for computation of the HIV sequence data example with 118 tests. Overall, based on the findings and the results we can conclude that the mode-based Beta prior is a suitable default prior for use in multiple testing of two proportions.

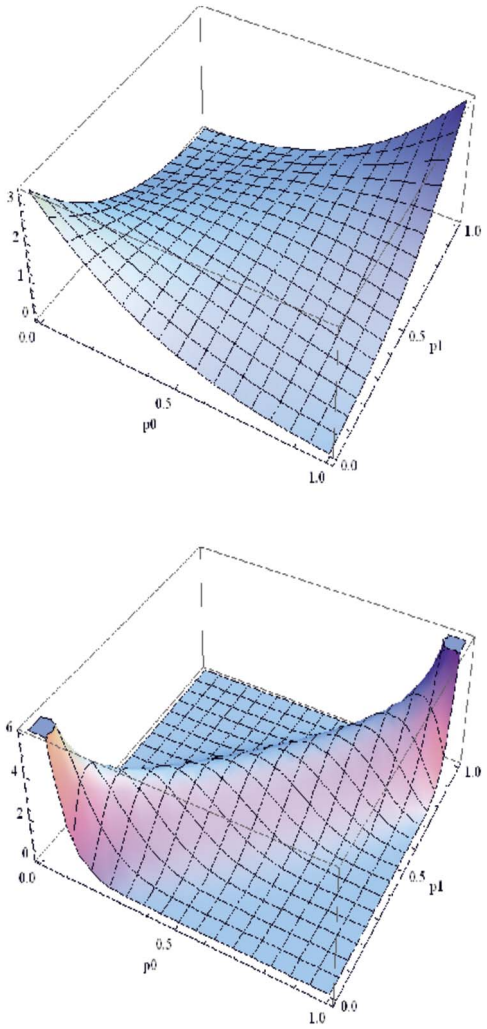

Figure A1. $\pi(p_0, p_1|m)$ *for* $m = 5$ *(top),* $m = 40$ *(bottom) for the intrinsic prior.*

APPENDIX A. INTRINSIC PRIOR

- Plots of intrinsic prior for two different choices of m.

Lemma A.1. *For the intrinsic prior in (2),*

$$E(p_1\,|p_0\,) = \frac{2+2m+m^2p_0}{(2+m)^2}. \tag{A1}$$

Proof. Given the joint probability density (2), the marginal dist. of p_0 is uniform over $(0,1)$ and hence the conditional probability density function of p_1 given p_0 has the same expression as the joint density function. Thus, letting $A_{ij} = \Gamma(i+j+1)\Gamma(2m-i-j+1)/\Gamma(2m+2))$, we have

$$\begin{aligned} &E(p_1|p_0) \\ &= \sum_{i=0}^{m}\sum_{j=0}^{m}\binom{m}{i}\binom{m}{j} A_{ij}\ Be(p_0\,|i+1,\,m-i+1)(\frac{j+1}{2+m}) \\ &= \frac{1}{2+m}\underbrace{\sum_{i=0}^{m}\sum_{j=0}^{m}\binom{m}{i}\binom{m}{j} A_{ij}\ Be(p_0\,|i+1,\,m-i+1)(j)}_{B} \\ &+ \frac{1}{2+m}\underbrace{\sum_{i=0}^{m}\sum_{j=0}^{m}\binom{m}{i}\binom{m}{j} A_{ij}\ Be(p_0\,|i+1,\,m-i+1)}_{B_1}. \end{aligned}$$

By integrating (2) over p_0, we see that the term indicated by B_1 in the above is equal to 1, and using the label B, for the expression as indicated in the first term on the right of the above equation, we have

$$E(p_1\,|p_0\,) = \frac{1}{(2+m)}\,[B+1] \tag{A2}$$

Now, we get

$$B = \sum_{i=0}^{m} \binom{m}{i} \frac{(m+1)!}{i!(m-i)!} p_0^i (1-p_0)^{m-i}.$$
$$\underbrace{\sum_{j=0}^{m} \frac{m!}{(m-j)!(j-1)!} \frac{(i+j)!(2m-i-j)!}{(2m+1)!}}_{C}$$

Using Vandermonde's identity and after some algebra, we have

$$C = \frac{m}{(m+2)\binom{m+1}{i+1}} \tag{A3}$$

Now, by plugging C in equation (A3), and after some more algebra which we skip here, we get

$$B = \frac{m}{(m+2)}(1+mp_0).$$

Finally, substituting B in equation (A2), yields (A1), completing the proof. □

Lemma A.2. *For the intrinsic prior in (2),*

$$corr(p_{0i}, p_{1i}) = \frac{m^2}{(2+m)^2}. \tag{A4}$$

Proof. This is easily derived using (A1) and the fact that each of p_{0i} and p_{1i} has its marginal as uniform distribution over $(0,1)$. □

• Prior distribution for the variance of the intrinsic prior induced by the prior on m.

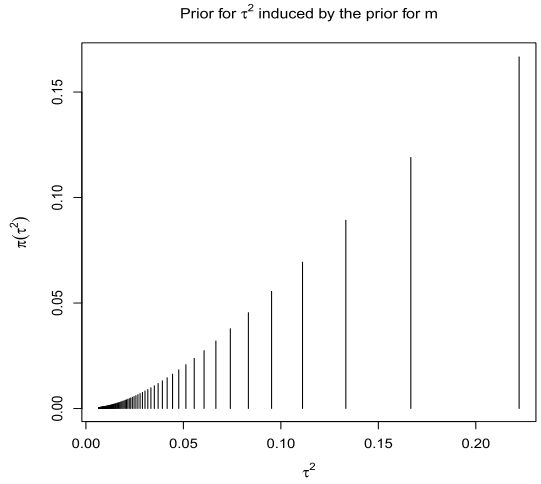

Figure A2. Prior for τ^2 induced by the prior (9) for the hyperparameter m of the intrinsic prior.

APPENDIX B. MODE-BASED PRIOR

• Symmetry of the marginal distribution of p_1:

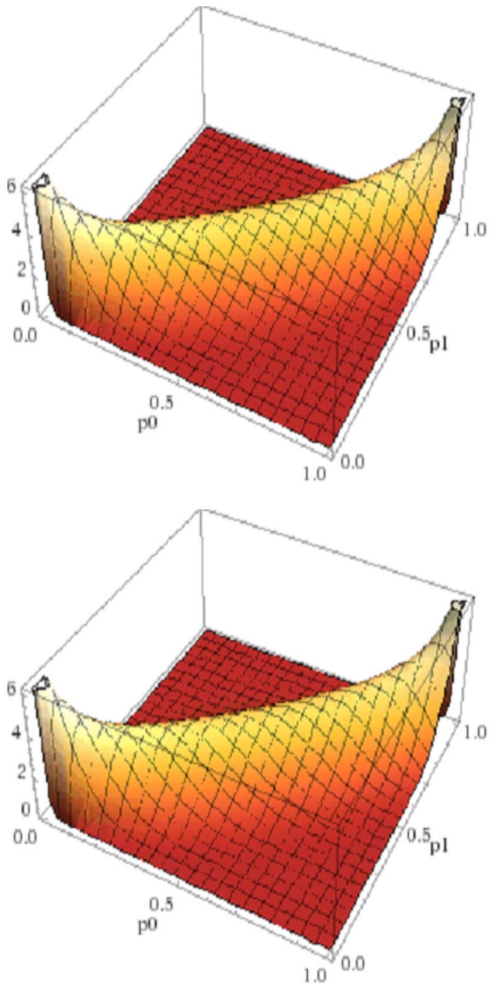

Figure A3. $\pi(p_0, p_1|r)$ for $r = 3$ (top), $r = 10$ (bottom) for the mode-based Beta prior.

Letting $Beta(x : a, b)$ denote the density function of the $Beta(a, b)$ distribution and $\pi(p_1|r)$ be the marginal distribution of p_1, we get from (4), using $B(x : a, b) = B(1 - x, b, a)$

$$\begin{aligned}\pi(p_1|r) &= \int Beta(1-p_1 : r(1-p_0)+1, rp_0+1)dp_0 \\ &= \int Beta(1-p_1 : rp_0'+1, r(1-p_0')+1)dp_0' \\ &= \pi(1-p_1|r).\end{aligned}$$

• Plots of the conditional mean of $p_{1i}|p_{0i}$ versus p_{0i} and the conditional mean of $p_{0i}|p_{1i}$ versus p_{1i}.

• Figure A5: The figure gives the plot of the posterior probability of the null hypothesis $p_{1i} = p_{0i}$ using the prior (4), and the prior obtained by switching p_0 and p_1 in (4) for a simulated data with 40 tests. Null hypothesis was assumed true for the first 30 tests, and a pair of observations

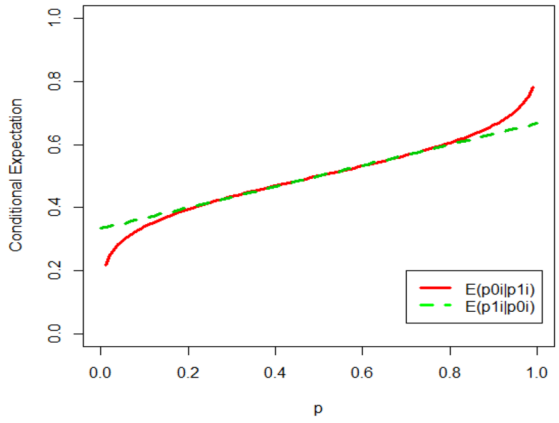

Figure A4. Plots of the conditional means of p_{0i} given p_{1i} vs. p_{1i}, and of p_{1i} given p_{0i} vs. p_{0i}, marginalized over r.

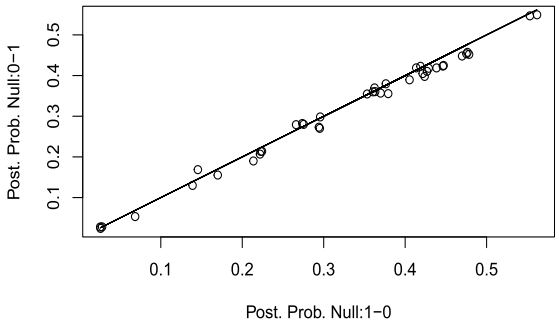

Figure A5. Posterior probabilities of the null hypothesis $p_{1i} = p_{0i}$ using the prior (4) indicated on the horizontal axis (1-0) and the corresponding posterior probabilities using the prior obtained by switching p_0 and p_1 in (4) on the vertical axis (0-1), for a simulated data with 40 tests.

for these were each simulated from Bin(10, 0.2) and the alternative hypotheses were assumed true for the last 10 tests and the observations were were generated from Bin(10, 0.2) and Bin(10,0.7).

APPENDIX C. LOGIT-NORMAL PRIOR

This prior is defined by letting $\theta_i = \log\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)$ and defining $\pi_1(p_0, p_1)$ indirectly by letting,

$$\text{(A5)} \qquad \theta_i|p_0, \tau^2 \sim N(0, \tau^2), \quad p_0 \sim Uni(0,1)$$

where τ^2 is the hyperparameter for this prior. By numerical calculation, we verified that this prior allows a range of $(0, 1)$ for the correlation between p_1 and p_0, and thus satisfies the second property in Section 3.1. However, it does not satisfy the third property, as shown below.

(A6)
$Pr(|p_1 - p_0| > \varepsilon|p_0) > 0$ for some $\varepsilon > 0$, as $p_0 \to c \in [0, 1]$.

is not satisfied by the logit-normal prior.

Proof. Given $\varepsilon > 0$, taking $0 < p_0 < \varepsilon$,

$$\begin{aligned} P(|p_1 - p_0| > \varepsilon|p_0) &= P(p_1 > p_0+\varepsilon|p_0)+P(p_1 < p_0 - \varepsilon|p_0), \\ &= P(p_1 > p_0 + \varepsilon|p_0), \\ &= P(logit(p_1) > logit(p_0 + \varepsilon)|p_0), \\ &= P(Z > (logit(p_0 + \varepsilon) - logit(p_0))/\tau), \end{aligned}$$

where, $Z \sim N(0, 1)$. Now, letting p_0 go to 0, $P(|p_1 - p_0| > \varepsilon|p_0) \to 0$. Similarly, we can also verify that $P(|p_1 - p_0| > \varepsilon|p_0) \to 0$ as $p_0 \to 1$. Hence the condition (A6) is not satisfied. □

APPENDIX D. MEAN BASED BETA PRIOR

This prior is a mixture of Beta distributions for $(p_0, p_1|r)$ and it is given by

$$\text{(A7)} \qquad p_1|p_0, r \sim Beta(rp_0, r(1 - p_0)), \qquad p_0 \sim Uni(0, 1)$$

where $r > 0$ is the hyperparameter. This prior was also used by Kass & Hsiao (1993), Kass & Raftery (1995), and Hsiao, Lee, & Kass (2005) in the context of testing the equality of a single pair of proportions (for $k = 1$). Kass & Hsiao (1993) used this prior with a fixed value for $w = 1/r$ in the single testing context. Under this prior, the conditional mean of p_1 given p_0 is p_0, hence the name mean-based Beta prior. Also,

$$Var(p_1|p_0) = \frac{p_0(1 - p_0)}{r + 1} \quad \text{and} \quad Corr(p_1, p_0) = \sqrt{\frac{(r + 1)}{(r + 3)}}.$$

By construction, this prior density is centered around the null hypothesis and decreases away from the null and thus satisfies the first property in Section 3.1.

Using the Markov inequality, for any $\varepsilon > 0$, we get

$$\text{(A8)} \qquad P(p_1 > \varepsilon|p_0, r) < \frac{E(p_1|p_0, r)}{\varepsilon} = \frac{p_0}{\varepsilon}.$$

Hence, $P(p_1 > \varepsilon|p_0, r) \to 0$, as $p_0 \to 0$, for any $\varepsilon > 0$. Furthermore, the correlation coefficient between p_1 and p_0 is in the range of $(0.5773, 1]$ as r varies in $(0, \infty)$. This means this prior is unable to model a correlation coefficient between p_1 and p_0 smaller than 0.5773. Thus, this prior does not satisfy the second and third properties in Section 3.1.

APPENDIX E. FARLIE GUMBEL MORGENSTERN PRIOR

A well known class of bivariate distributions that has uniform marginals is the FGM distribution which can be given by

$$\text{(A9)} \qquad \pi(p_1, p_0|\kappa) = 1 + \kappa(2p_1 - 1)(p_0 - 1)$$

where $0 \le \kappa \le 1$ is the hyperparameter. This prior was used for Bayesian meta analysis of sparse discrete Binomial

data by Moreno, Vazquez, & Negrin (2014). The conditional mean and variance are given by

$$\begin{aligned} E(p_1|p_0,\kappa) &= \frac{1}{2}+\frac{\kappa}{3}(p_0-\frac{1}{2}), \text{ and} \\ Var(p_1|p_0,\kappa) &= \frac{1}{3}+\frac{\kappa}{6}(2p_0-1)-(\frac{1}{2}+\frac{\kappa}{6}(2p_0-1))^2. \end{aligned}$$

The conditional mean is increasing (respectively, decreasing) in κ when $p_0 > 1/2$ (resp. when $p_0 < 1/2$), and the conditional variance is decreasing in κ, Also, $Corr(p_1, p_0) = \frac{\kappa}{3}$, and hence, the correlation coefficient between p_1 and p_0 is in the range of $[0, 1/3]$ as κ varies in $(0, 1)$. This prior is not able to model a correlation coefficient between p_1 and p_0 greater than 0.333. Thus, this prior does not satisfy the second property in Section 3.1.

Received 1 February 2022

REFERENCES

[1] Bayarri, M. J., Berger, J. O., Forte, A., & Garcia-Donato, G. (2012). Criteria for Bayesian model choice with application to variable selection. *The Annals of Statistics*, 40(3):1550–1577. MR3015035

[2] Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300. MR1325392

[3] Berger, J. O. & Pericchi, L. R. (1996). The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, 91(433), 109–122. MR1394065

[4] Berger, J. O. & Strawderman, W.E. (1996). Choice of hierarchical priors: admissibility in estimation of normal means. *The Annals of Statistics*, 24, 3, 931–951. MR1401831

[5] Berger, J. O. (2006). The Case for Objective Bayesian Analysis. *Bayesian Analysis*, 1, Number 3, 385–402. MR2221271

[6] Casella, G. & Moreno, E. (2009). Assessing robustness of intrinsic tests of independence in two-way contingency tables. *Journal of the American Statistical Association*, 104(487), 1261–1271. MR2750249

[7] Chen, J. & Sarkar, S. (2004). Multiple testing of response rates with a control: A Bayesian stepwise approach. *Journal of Statistical Planning and Inference*, 125, 3–16. 10.1016/j.jspi.2003.05.001. MR2086885

[8] Consonni, G. & La Rocca, L. (2008). Tests based on intrinsic priors for the equality of two correlated proportions. *Journal of the American Statistical Association*, 103, 1260–1269. MR2462897

[9] Consonni, G., Forster, J. J., La Rocca, L. (2013). The Whetstone and the Alum Block: balanced objective Bayesian comparison of nested models for discrete data. *Statistical Science*, 28, no. 3, 398–423. MR3135539

[10] Daniels, M. (1999). A prior for the variance in hierarchical models. *The Canadian Journal of Statistics / La Revue Canadienne De Statistique*, 27(3), 567–578. MR1745822

[11] Jeffreys, H. (1961). *Theory of Probability*. Clarendon Press, Oxford, 3rd edition. MR0187257

[12] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3): 515–534. 1, no. 3, 515–533. MR2221284

[13] Gilbert, P. B. (2005). A modified false discovery rate multiple-comparisons procedure for discrete data, applied to human immunodeficiency virus genetics. *Journal of the Royal Statistical Society: Series C* (Applied Statistics), 54(1): 143–158. MR2134603

[14] Heyse, J. F. (2011). A false discovery rate procedure for categorical data. *In: H. Zhang (Ed.) Recent advancements in biostatistics*. New Jersey: World Scientific Publishing Company, 43–58.

[15] Howard, J. V. (1998). The 2x2 table: a discussion from a Bayesian viewpoint. *Statistical Science*, 13 (1998), no. 4, 351–367. MR1705267

[16] Hsiao, C. K., Lee, M., & Kass, R. E. (2005). Bayesian tests of extra-Binomial variability. *Statistics in Medicine*, 24;49–64. MR2134495

[17] Kass, R. E. & Hsiao, C. K. (1993). A Bayesian approach to testing for extra-Binomial variability. *Technical Report*.

[18] Kass, R. E. & Raftery, A. E. (1995). Bayes factors, *Journal of the American Statistical Association*. 90, 773795. MR3363402

[19] Liang, K. (2016). False discovery rate estimation for large-scale homogeneous discrete p-values.*Biometrics*, 7(2), 639. MR3515790

[20] Moreno, E., Bertolino, F., & Racugno, W. (1998). An Intrinsic Limiting Procedure for Model Selection and Hypotheses Testing. *Journal of the American Statistical Association*,93, pp. 1451–1460. MR1666640

[21] Moreno, E., Vazquez, F.J., & Negrin, M.A. (2014). Objective Bayesian meta-analysis for sparse discrete data. *Statistics in Medicine*, 2014, 33, 3676–3692. MR3260653

[22] Müller, P., Parmigiani, G., & Rice, K. (2007). "FDR and Bayesian Multiple Comparisons Rules" in Bayesian Statistics 8, J.M. Bernardo, S. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, and M. West (eds.), Oxford University Press, pp. 349–370. MR2433200

[23] Pham-Gia, T., Thin, N., & Doan, P. (2017). Inferences on the difference of two proportions: a Bayesian approach. *Open Journal of Statistics*, 7, 1–15.

[24] Rockova, V. & George, E. (2014). EMVS: The EM Approach to Bayesian Variable Selection.*Journal of the American Statistical Association*, 109, 828–846. MR3223753

[25] Scott, J. G. & Berger, J. O. (2006). An exploration of aspects of Bayesian multiple testing.*Journal of Statistical Planning and Inference*, 136(7):2144–2162. MR2235051

[26] Sverdlov, O., Ryeznik, Y., & Wu, S. (2015). Exact Bayesian inference comparing Binomial proportions, with application to proof-of-concept clinical trials. *Therapeutic Innovation & Regulatory Science*, 2015;49(1):163–174.

[27] Tarone, R.E. (1990). A modified Bonferroni method for discrete data. *Biometrics*, 46:515–522.

Emrah Gecili
Cincinnati Children's Hospital Medical Center
Cincinnati, Ohio
USA
E-mail address: emrah.gecili@cchmc.org

Siva Sivaganesan
University of Cincinnati
Cincinnati, Ohio
USA
E-mail address: sivagas@ucmail.uc.edu