A new multiple testing procedure under dependence

Prasenjit Ghosh and Arijit Chakrabarti

Applied Statistics Unit, Indian Statistical Institute, Kolkata, India

April 1, 2015

Abstract

In this article, we consider the problem of simultaneous testing of hypotheses when the individual test statistics are not necessarily independent. Specifically, we consider the problem of simultaneous testing of point null hypotheses against two-sided alternatives about the mean parameters of normally distributed random variables. We assume that conditionally given the vector means, these random variables jointly follow a multivariate normal distribution with a known but arbitrary covariance matrix. We consider a Bayesian framework where each unknown mean is modeled via a two component point mass mixture prior, whereby unconditionally the test statistics jointly have a mixture of multivariate normal distributions. A new testing procedure is developed that uses the dependence among the test statistics and works in a step down like manner. This procedure is general enough to be applied to even non-normal data. A decision theoretic justification in favor of the proposed testing procedure has been provided by showing that unlike the traditional p-value based stepwise procedures, this new method possesses a certain convexity property which is essential for the admissibility of a multiple testing procedure with respect to the vector risk function. Consistent estimation of the unknown proportion of alternative hypotheses and variance of the distribution of the non-zero means is theoretically investigated. An alternative representation of the proposed test statistics has also been established resulting in a great reduction in computational complexity. It is demonstrated through extensive simulations that for various forms of dependence and a wide range of sparsity levels, the proposed testing procedure compares quite favourably with several existing multiple testing procedures available in the literature in terms of overall misclassification probability.

1 Introduction

During the past two decades, multiple hypothesis testing has been one of the most significant areas of research, particularly for its wide applicability in the analysis of high throughput data coming from various scientific fields. For example, in microarray experiments, one tests thousands of hypotheses simultaneously to decide which genes are differentially expressed. These are genes which are associated with some biological trait of interest. Many different multiple testing procedures have been proposed in the literature so far, mostly with the aim of controlling some overall measure of type I error at a predetermined level $\alpha$. The familywise error rate (FWER), defined as the probability of making at least one false rejection, is a type I error measure which has been known and used for a very long time, with the Bonferroni correction being the most widely used FWER controlling procedure. Among the various type I error measures, the false discovery rate (FDR), due to Benjamini and Hochberg [1995], has received most attention from researchers in recent times. It is defined as the expected proportion of erroneously rejected null hypotheses among all the hypotheses that are rejected. In their seminal work, Benjamini and Hochberg [1995] show that when test statistics corresponding to the true null hypotheses are independent, their method, henceforth referred to as the BH method, controls the FDR at a prespecified level $\alpha$. A related FDR controlling procedure that also works under independence is in Benjamini and Liu [1999]. In an important paper, Storey [2002] introduced an approach where one estimates the FDR corresponding
to a fixed critical region and using this an FDR controlling procedure was proposed subsequently in 
Storey, Taylor and Siegmund (2004). An essential part of this approach is plugging in an estimate 
of the proportion of true nulls. Storey, Taylor and Siegmund (2004) show that their procedure is 
provides control over FDR at a prespecified level under the assumption of independence of the null 
\( p \)-values.

The theoretical properties of the procedures rely on the assumption that the underlying test 
statistics are independent. But in practice test statistics are often dependent and it has been 
observed in simulations and theoretical investigations that performance of BH may not remain sat-
sisfactory in such cases. This issue has received significant attention from the researchers in the last 
decade and a half. Firstly, there have been attempts to prove that BH continues to have the FDR 
controlling properties under certain form of dependence. For example, Benjamini and Yekutieli 
(2001) show that the BH method can be applied with the desired FDR control when the test 
statistics exhibit a form of dependence known as “positive regression dependence”. Secondly, there 
have been attempts to come up with new procedures suitable for the dependent case. We first 
mention Benjamini and Yekutieli (2001), who propose a multiple testing procedure, referred to as 
the Benjamini-Yekutieli (BY) procedure, which controls the FDR under any form of dependence. 
Sarkar (2002) extends their work by showing that some generalized step-up and step-down proce-
dures control the FDR at the desired level under general dependence. Storey, Taylor and Siegmund 
(2004) show that their approach has a desirable property in that it estimates FDR conservatively 
in the asymptotic sense under certain forms of weak dependence. Romano, Shaikh and Wolf (2008) 
consider a bootstrap approach for achieving an asymptotic control over the FDR under depen-
dence. These methods are based on \( p \)-values which depend only on the marginal distributions of 
the corresponding test statistics and do not take into account the correlations between them. Efron 
(2007) propose a novel testing procedure which depends on estimating a certain dispersion variate 
that acts as a measure of the overall correlation between the data. Leek and Storey (2008) and 
Friguet, Kloareg and Causeur (2009) propose multiple testing procedures based on factor model 
approaches. Fan, Han and Gu (2012) propose a new multiple testing procedure referred to as the 
Principal Factor Approximation (PFA) approach. They develop a novel testing procedure based on 
a set of dependence adjusted \( p \)-values, when the underlying test statistics jointly follow a multivari-
ate normal distribution with an arbitrary but known covariance matrix.

The above approaches study multiple testing under dependence from a frequentist point of view. 
The Bayesian approaches available in the literature in this problem are relatively few. Sun and Cal 
(2009) propose an oracle method combined with an asymptotically optimal data adaptive testing 
procedure within a Bayesian decision theoretic framework. They assume the underlying model 
parameters to be generated according to a homogeneous and irreducible two-state hidden Markov 
chain. Brown, Lazar and Datta (2011) introduce a conditional autoregressive model to account 
for the spatial dependence in the data for an fMRI analysis using a Bayesian approach. One of 
the most simplest and natural approach to model dependent data is to use a multivariate normal 
model. It is probably the most widely used model for capturing dependence. Xie et. al (2011) 
propose a Bayesian testing procedure for the two-sided multivariate normal mean problem, assum-
ing a short range dependence covariance structure. Their method is based on the central idea 
of approximating the joint oracle statistics by the corresponding marginal statistics arising out of 
a two-component Gaussian mixture model and estimation of these statistics as if the data were 
independent. Their method, though, performs well for the short range dependence case, is not 
expected to have similar performance under other forms of dependence. As in Xie et. al (2011), 
we consider in this article the problem of simultaneous testing of the individual components of a 
multivariate normal mean using a Bayesian approach. The difference in our approach is we develop 
a procedure that is applicable under arbitrary dependence, provided the covariance matrix is known.

Suppose we observe data \( X = (X_1, \ldots, X_m) \) such that \( X | \mu \sim N_m(\mu, \Sigma) \), where \( \mu = (\mu_1, \ldots, \mu_m) \) 
is an \( m \times 1 \) vector of unknown means and \( \Sigma \) is an \( m \times m \) positive definite covariance matrix and 
is assumed to be known. Note that we do not impose here any restriction on whether \( X_1, \ldots, X_m \)
are strongly or weakly correlated. We are interested in testing the following:

\[ H_0: \mu_i = 0 \text{ against } H_{A_i}: \mu_i \neq 0, \text{ for } i = 1, \ldots, m. \]

Since \( \Sigma \) is assumed to be known, without any loss of generality, one may assume \( \Sigma \) to be the correlation matrix of the \( X_i \)'s so that \( X_i \sim N(\mu_i, 1) \) for \( i = 1, \ldots, m \). Note that this transformation leaves the testing problem unchanged. We consider a Bayesian approach by introducing a set of latent variables that determines a two-component point mass mixture prior for each of the \( \mu_i \)'s. Such two groups formulation is considered a very natural way of formulating a multiple testing problem from a Bayesian viewpoint. See Efron, Tibshirani, Storey and Tusher (2001), Efron (2004), Scott and Berger (2004), Bogdan, Ghosh and Tokdar (2008), Jin and Cai (2007), Scott and Berger (2010), Bogdan et al. (2011) and Xie et. al. (2011) in this context. We develop a new Bayesian testing procedure using this two-groups framework, that works in a step-down manner which we call the Bayesian Step Down (BSD) procedure.

The BSD procedure has several desirable properties as a multiple testing procedures under dependence. Firstly, the BSD method fully utilizes the dependence between the test statistics \( X \) at every stage and applicable under arbitrary known covariance matrices. Most of the existing multiple testing procedures, such as the \( p \)-value based stepwise methods or Bayesian methods like Xie et. al. (2011), do not incorporate such information. Secondly, as explained below, unlike the typical \( p \)-value based stepwise procedures, the BSD procedure has a very important admissibility property, in that it is not possible to find another multiple testing procedure such that the individual tests induced by that procedure are uniformly better than those induced by the BSD procedure in terms of the risk corresponding to the usual 0-1 loss. Thirdly, it is easily implementable and can avoid Markov Chain Monte Carlo type computations which can often be very demanding from a computational point of view, specially when \( m \) is large. It can be applied to all situations when \( \Sigma \) is known as against some of the well known multiple testing procedures which are only valid for some special form of dependence, for example, positive regression dependence, among the test statistics. We observe in our numerical studies that, for any arbitrary choice of the covariance matrix \( \Sigma \) and a wide range of sparse situations, the Bayes misclassification risk of the proposed BSD method compares favorably with several existing methods, with a very high power. We define power is defined as the expected proportion of correctly identified signals. Moreover, it gives us a generic multiple testing algorithm which is applicable even for non-normal models.

We observe that there exists a close connection between the BSD method and the Maximum Residual Down (MRD) method, a step-down type method developed by Cohen, Sackrowitz and Xu (2009) for this multiple testing problem. We can show a functional relationship between the BSD statistics and the corresponding MRD statistics. This helps us prove an important theoretical justification concerning the use of the BSD procedure from a frequentist view point. We show that when \( X \) is generated according to a fixed, but unknown mean vector \( \mu \) with an arbitrary known positive definite covariance matrix, the BSD method possesses a certain convexity property which is both necessary and sufficient for the admissibility of a multiple testing procedure. For this purpose, we call a multiple testing procedure admissible if each of the induced testing rules is admissible with respect to the usual loss function for the corresponding individual testing problem. In a series of important research articles, Cohen and Sackrowitz (2005), Cohen and Sackrowitz (2007), Cohen and Sackrowitz (2008) and Cohen, Kolassa and Sackrowitz (2007) show that in many circumstances, given a typical \( p \)-value based step-up or step-down method, it is always possible to construct another multiple testing procedure having a smaller expected number of type I and type II errors. Moreover, Cohen and Sackrowitz (2008) demonstrate that in the context of a general linear regression problem or in a treatment versus control study, for testing point null hypotheses against the two-sided alternatives, there exist procedures whose individual tests have smaller expected number of type I and type II errors compared to the traditional \( p \)-value based stepwise methods. Thus the typical \( p \)-value based stepwise procedures become inadmissible whenever the risk is an increasing function of the expected number of type I and type II errors. Cohen, Sackrowitz and Xu
(2009) show that their proposed MRD method possesses such aforesaid admissibility property. Our proof of admissibility of the BSD method borrows the basic architecture of the corresponding proof for the MRD procedure in Cohen, Sackrowitz and Xu (2003). But the proof does not follow as a direct consequence of that of the MRD procedure. As will be evident later in this paper, we need a general technique invoking new arguments, which also makes proofs of some results in Cohen, Sackrowitz and Xu (2009) more explicit. Our argument also shows that within the aforesaid frequentist framework, any multiple testing procedure based on a set of test statistics which are strictly increasing functions of the absolute values of the corresponding MRD test statistics, will be admissible, thus extending the result of Cohen, Sackrowitz and Xu (2009), which is another important contribution of this paper.

The MRD procedure depends on a set of decreasing sequence of critical constants $C_1 \geq \cdots \geq C_m > 0$, choice of which are somewhat ad hoc and vary with $\Sigma$. Performance of the MRD procedure thus critically depends on the appropriate choice of $C_1 \geq \cdots \geq C_m$, and utmost care needs to be taken while deciding over the choice of these $C_i$’s. As opposed to this, we choose the thresholding constant $\delta$ used in the BSD method (see Section 2) to be equal to 1. Thus we have an automatic default choice of the corresponding significance thresholding constant for the BSD procedure that works for all choices of $\Sigma$. It can be shown that rejection of a hypothesis under the BSD method is possible only when the hypothesis is also rejected by the MRD test criterion used with a certain sequence $C_1(X) \geq \cdots \geq C_m(X) > 0$ of data dependent critical constants.

Practical implementations of the BSD method requires the knowledge of the proportion of true alternatives and the variance of the distribution of the non-null $\mu_i$’s along with certain posterior probabilities used in the definition of the BSD statistics in Section 2. When this knowledge is unavailable, one can use a full Bayes approach by placing hyperpriors on the underlying model parameters and then implement an MCMC method to estimate each of these quantities. However, as already mentioned before that it can be computationally very expensive, specially when the total number of null hypotheses ($m$) under consideration is large. So we prefer an empirical Bayes approach where one needs to estimate each of these underlying model parameters from the data and plug them into the BSD procedure. Estimation of the theoretical proportion of true alternatives has been so far an important problem in the domain of multiple testing. Several methods has been proposed in the literature. See, for example, Efron (2004), Meinshausen and Rice (2006), Jin (2006), Jin and Cai (2007), Cai, Jin and Low (2007), Jin (2008) and Cai and Jin (2010) and references therein. Cai and Jin (2010) propose an estimator that is both consistent and attains the corresponding minimax optimal rate of convergence. Their method, however, is proposed when the test statistics are i.i.d observations from a mixture of Gaussian distributions. In this article, we consider their estimator and show that it is consistent in a broad range of weak dependent situations. For estimating the remaining parameter, we consider a natural moment based estimator of the non-null variance which we show to be consistent under weak and some stronger form of dependence, given that we already have a consistent estimator of the proportion of true alternatives.

We conclude the introduction by explaining a very useful contribution of this work from a computational point of view. Both the BSD method (or its empirical version as explained above) and the MRD method, require inversion of $(m-t+1)$ many submatrices of $\Sigma$ and $(m-t+1)$ many ratios of determinants at each step $t$, $t = 1, \cdots, m$, which may be computationally very costly, specially, when $m$ is large. We derive an important alternative representation of the BSD test statistics due to which at each step we now need to compute the inverse of one submatrix only and can also avoid the need for computing ratios of $(m-t+1)$ many determinants as mentioned already. This reduces the overall computational complexity of the BSD method by a large extent. Such a representation works for any form of the covariance matrix $\Sigma$. It not only enables the BSD procedure to become computationally much faster compared to that in its original form, but does the same for the MRD method also. This is would be particularly very useful when $\Sigma$ corresponds to an intraclass correlation and a block (clumpy) dependence matrix. In particular, for the intraclass correlation model, we do not even require inversion of any matrix and thus the BSD method can be applied for any
arbitrarily large $m$.

The organisation of the paper is as follows. Section 2 gives our prior specification and the motivation and description of the proposed Bayesian Step Down procedure. Section 3 shows the connection between the BSD procedure and the MRD method and the admissibility property possessed by the BSD method. Consistent estimation of the proportion of true alternatives and the variance of the non-null distribution of the $\mu_i$’s are shown in Section 4. Equivalent representation of the BSD statistics and associated results are given in Section 5. Performance of the BSD method based on simulation studies will be demonstrated in Section 6. Proofs of most of the theoretical results are given in Appendix (Section 7) followed by some concluding remarks in Section 8.

2 Statistical Model and The Bayesian Step Down Method

For each $i = 1, \cdots, m$, let us define an indicator variable $\nu_i$ which takes the value 1 if $H_{Ai}$ is true and 0 if $H_{Ai}$ is false. Here $\nu_1, \cdots, \nu_m$ are unobservable. It is assumed that $\nu_i \overset{i.i.d.}{\sim} \text{Bernoulli}(p)$ for some $p \in (0,1)$. The parameter $p$ is often interpreted as the theoretical proportion of true alternatives. Given $\nu_i = 0$, $\mu_i$ is assumed to follow the distribution $\delta_0$ degenerated at the point 0, while it is assumed to follow some absolutely continuous density $g(\mu)$ given $\nu_i = 1$. Thus $\mu_i$’s are considered to be modelled as independent and identically distributed (i.i.d.) random observations coming from the following two-component mixture distribution, often referred to as a spike-and-slab prior:

$$\mu_i \overset{i.i.d.}{\sim} (1-p) \cdot \delta_0 + p \cdot g(\mu), \ i = 1, \cdots, m. \tag{2.1}$$

so that the common marginal distributions of $X_i$’s are given by the following Gaussian mixing density,

$$X_i \sim (1-p) \cdot f_0(x) + p \cdot f_1(x), \ i = 1, \cdots, m. \tag{2.2}$$

where $f_0 = \phi$ and $f_1(x) = \int_{\mu \neq 0} \phi(x - \mu)g(\mu)d\mu$ is the convolution of $g$ with $\phi(\cdot)$, where $\phi(\cdot)$ denotes the standard normal density over $\mathbb{R}$.

As mentioned already in the Introduction that the above two groups formulation in (2.2) is considered to be a natural way for formulating a problem of this kind. In a large number of practical problems, one can model the data through a mixture of Gaussian distributions such as in (2.2) above. Moreover, it is now well known that the set of all Gaussian mixing densities is dense in the set of all density functions under the $L_1$ metric.

Usually $g$ is assumed to be an absolutely continuous density over $\mathbb{R}$ with a flat tail. A natural choice for $g$ is an univariate normal density with some large variance. So, let us assume $g$ to be the density corresponding to a $N(0,V)$ distribution, where $V$ is assumed to be large. In such a case, the marginal joint distribution of $X = (X_1, \cdots, X_m)$ is given by Theorem 2.1 below.

**Theorem 2.1.** Under the above model assumptions, the conditional distribution of $X = (X_1, \cdots, X_m)$ given $\nu = (\nu_1, \cdots, \nu_m)$ is given by,

$$X | \nu \sim N_m(0, \Sigma + VB_\nu) \tag{2.3}$$

and the marginal joint distribution of $X = (X_1, \cdots, X_m)$ is given by,

$$X \sim \sum_{\nu \in \{0,1\}^m} \pi(\nu)N_m(0, \Sigma + VB_\nu) \tag{2.4}$$

where $\pi(\nu) = \prod_{i=1}^m p^{\nu_i}(1-p)^{1-\nu_i}$ denotes the prior distribution of $(\nu_1, \cdots, \nu_m)$ and $B_\nu$ is a diagonal matrix with diagonal elements $\nu_1, \cdots, \nu_m$ respectively.

**Proof.** See Appendix. 

---

5
To compare the global null hypothesis, we confine our attention to the sub-space of models where the global null hypothesis be true?”. A natural approach to answer this question is as follows. “Can at least one null hypothesis be rejected?” which is equivalent to answering the question “Can the median probability model be one of the cases where the global null hypothesis is true or equivalently with an alternative model being true to that of the global null hypothesis. If the maximum of these ratios of the posterior probability of the corresponding null hypothesis to that of the global null hypothesis exceeds some pre-specified threshold, say, \( \delta > 0 \), we conclude that the global null hypothesis cannot be true and is rejected along with the null hypothesis corresponding to the coordinate at which this maximum occurs and remove the corresponding \( X_i \) and move on to the next stage. Otherwise, we accept all of them and stop. We continue in this fashion till an acceptance occurs or we are

\[
\pi(\nu_i = 1|X) = \frac{\sum_{(\nu_1, \ldots, \nu_m) \in \{0,1\}^m} \pi((\nu_1, \ldots, \nu_m)) f(x|\nu_1, \ldots, \nu_m)}{\sum_{(\nu_1, \ldots, \nu_m) \in \{0,1\}^m} \pi((\nu_1, \ldots, \nu_m)) f(x|\nu_1, \ldots, \nu_m)}
\]

where \( \pi((\nu_1, \ldots, \nu_m)) \) and \( f(x|\nu_1, \ldots, \nu_m) \) denote the prior distribution of \( (\nu_1, \ldots, \nu_m) \) and the density corresponding to the conditional distribution of \( X|\nu_1, \ldots, \nu_m \), respectively. It therefore becomes evident from (2.6) that one needs to sum over \( 2^m \) many terms in order to obtain \( \pi(\nu_i = 1|X) \), for each \( i = 1, \ldots, m \), resulting in an overall computational complexity of an order of \( O(m2^m) \). Clearly, by taking either of these two aforesaid approaches, one faces a daunting computational task even when the conditional density \( f(x|\nu_1, \ldots, \nu_m) \) is completely known and the number of hypotheses \( m \) is moderately large. Chen and Sarkar (2004) propose a novel Bayesian step down method based on a set of conditional Bayes factors. Their method, though originally proposed and implemented when \( X_i \)'s are conditionally independent, is applicable in more general situations also. However, their method, though a natural Bayesian analogue to frequentist-type step-down procedures, crucially hinges upon an initial ordering of the marginal Bayes factors. For the present multiple testing problem, each of these marginal Bayes factors has an one-to-one correspondence to the posterior probability of the corresponding null hypothesis \( H_0 \) being true or equivalently with \( \pi(\nu_i = 1|X) \) in (2.6), resulting in the same computational issue as with the median probability model.

Recall that in a step-down method, we try to answer the following question at the first step: “Can at least one null hypothesis be rejected?” which is equivalent to answering the question “Can the global null hypothesis be true?” A natural approach to answer this question is as follows. To compare the global null hypothesis, we confine our attention to the sub-space \( \{(\nu_1, \ldots, \nu_m) \in \{0,1\}^m : \sum_{i=1}^m \nu_i = 1\} \) of the original model space, as the class of plausible alternatives to the global null. That is, we are considering only those models as plausible alternatives to the global null, each of which is a permutation of \( (m-1) \) many true null and exactly one false null hypothesis. For each of these models in this restricted sub-space, we enumerate the ratio of the posterior probability of an alternative model being true to that of the global null hypothesis. If the maximum of these ratios exceeds some pre-specified threshold, say, \( \delta > 0 \), we conclude that the global null hypothesis cannot be true and is rejected along with the null hypothesis corresponding to the coordinate at which this maximum occurs and remove the corresponding \( X_i \) and move on to the next stage. Otherwise, we accept all of them and stop. We continue in this fashion till an acceptance occurs or we are
exhausted with all the null hypotheses.

We now formally introduce the proposed Bayesian step-down procedure as follows:

We adopt similar convention of notations as used in Cohen, Sackrowitz and Xu (2009). Let \( \mathbf{X}^{(i_1, \ldots, i_t)} \) be an \((m - t) \times 1\) vector consisting of those components of \( \mathbf{X} = (X_1, \ldots, X_m) \) with \((X_{i_1}, \ldots, X_{i_t})\) left out. Suppose \( \Sigma_{(i_1, \ldots, i_t)} \) is the \((m - t) \times (m - t)\) submatrix obtained after eliminating the \(i_1, \ldots, i_t\)-th rows and the corresponding columns of \( \Sigma \). Let \( \sigma_{(j)}^{(i_1, \ldots, i_t)} \) be the \((m - t - 1) \times 1\) vector obtained by eliminating the \(i_1, \ldots, i_t\)-th and \(j\)-th elements of the \(j\)-th column vector of \( \Sigma \). Further suppose that

\[
\sigma_{j(i_1, \ldots, i_t)} = \sigma_{jj} - \sigma_{(j)}^{(i_1, \ldots, i_t)} \Sigma_{(i_1, \ldots, i_t),j} \sigma_{(j)}^{(i_1, \ldots, i_t)}.
\]

Let us define

\[
S_{ij}^{(i_1, \ldots, i_{t-1})}(X) = \frac{\pi(\nu_j = 1, \nu_{(i_1, \ldots, i_{t-1}, j)}) = 0|\mathbf{X}^{(i_1, \ldots, i_{t-1})}}{\pi(\nu_j = 0, \nu_{(i_1, \ldots, i_{t-1}, j)}) = 0|\mathbf{X}^{(i_1, \ldots, i_{t-1})}}
\]

for \(t, j = 1, \ldots, m, 1 \leq i_1 \neq \ldots \neq i_{t-1} \leq m\) and \(i_t \neq j\) for all \(l = 1, \ldots, t - 1\).

For \(t = 1, \ldots, m\), let us define the indices \(j_t(X)\) as,

\[
j_t(X) = \arg\max_{j \in \{1, \ldots, m\} \setminus \{j_1(X), \ldots, j_{t-1}(X)\}} S_{ij_t}^{(i_1, \ldots, i_{t-1})}(X).
\]

Given a pre-determined threshold \(\delta > 0\), the proposed Bayesian Step Down (BSD) procedure is described below:

1. At stage 1, let us consider the functions \(S_{1j}(X), j \in \{1, \ldots, m\}\). If \(S_{1j_1}(X) \leq \delta\), we stop and accept all \(H_{0j}\)’s. Otherwise we reject \(H_{0j_1}(X)\) and continue to stage 2.

2. At stage 2, we consider the functions \(S_{2j}(X), j \in \{1, \ldots, m\} \setminus \{j_1(X)\}\). If \(S_{2j_2}(X) \leq \delta\), we stop and accept all the remaining \(H_{0j}\)’s. Otherwise we reject \(H_{0j_2}(X)\) and continue to the next stage.

3. In general, at stage \(t\), we consider the \((m-t+1)\) many functions \(S_{ij_t}^{(i_1, \ldots, i_{t-1})(X)}\), where \(j \in \{1, \ldots, m\} \setminus \{j_1(X), \ldots, j_{t-1}(X)\}\). If \(S_{ij_t}(X) \leq \delta\), we stop and accept all the remaining \(H_{0j}\)’s. Otherwise we reject \(H_{0j_t}(X)\) and move on to the stage \((t+1)\).

4. We continue in this way till one hypothesis is accepted or there are no more null hypothesis to be tested (that is \(t = m\)), in which case we must stop.

Here the subscript \(t\) denotes the stage of the BSD procedure. The above description defines a class of Bayesian testing procedures for various choices of \(\delta > 0\). In this paper we will work with the choice of 1.

Observe that the statistic \(S_{ij_t}^{(i_1, \ldots, i_{t-1})(X)}\) may alternatively be written as

\[
S_{ij_t}^{(i_1, \ldots, i_{t-1})(X)} = \frac{\pi(\nu_j = 1|\nu_{(i_1, \ldots, i_{t-1}, j)}) = 0, \mathbf{X}^{(i_1, \ldots, i_{t-1})}}{\pi(\nu_j = 0|\nu_{(i_1, \ldots, i_{t-1}, j)}) = 0, \mathbf{X}^{(i_1, \ldots, i_{t-1})}}
\]

for \(t, j = 1, \ldots, m, 1 \leq i_1 \neq \ldots \neq i_{t-1} \leq m\) and \(i_t \neq j\) for all \(l = 1, \ldots, t - 1\). Thus the BSD test statistics above at stage \(t\) may be interpreted as the ratio of posterior probabilities of \(H_{ij}\) being true and \(H_{0j}\) being true given that that the rest of the null hypotheses still under consideration at stage \(t\) are true.
3 A Decision Theoretic Justification to the BSD Procedure

In this section, we shall give an important decision theoretic justification for the BSD method from a frequentist viewpoint. In particular, we show that the proposed BSD method based on the statistics $S_{ij}$'s in (2.4), will be admissible when $X$ is assumed to follow a multivariate normal distribution with a fixed, but unknown mean vector $\mu$ and a known covariance matrix $\Sigma$. Note that any multiple testing procedure $\Phi(x) = (\phi_1(x), \cdots, \phi_m(x))$ induces an individual test function $\phi_j(x)$ for testing $H_{0j}$ against $H_{Aj}$, where $\phi_j(x)$ denotes the probability of rejecting the $i$-th null hypothesis $H_{0j}$ when the data point $X = x$ is observed. The risk function corresponding to $\phi_j$ is

$$R_j(\phi_j(X), \mu) = (1 - I\{\mu_j = 0\})E_{\mu_{\mu_j=0}}(\phi_j(X)) + I\{\mu_j \neq 0\}E_{\mu_{\mu_j\neq0}}(\phi_j(X)) \quad (3.1)$$

We consider the risk function for the procedure $\Phi(X)$ to be defined as the vector risk function

$$R(\Phi(X), \mu) = (R_1(\phi_1(X), \mu), \cdots, R_m(\phi_m(X), \mu)) \quad (3.2)$$

A multiple testing procedure $\Phi(X)$ is said to be inadmissible with respect to the vector risk function (3.2) if there exists another multiple testing procedure $\Phi^*(X)$ such that $R_j(\phi_j^*(X), \mu) \leq R_j(\phi_j(X), \mu)$ for all $j = 1, \cdots, m$, with strict inequality holding for at least one $j$ and some $\mu \in \mathbb{R}^m$. A multiple testing procedure will be admissible if it is not inadmissible.

As mentioned in the introduction, in many frequently occurring situations, the typical $p$-value based stepwise multiple testing procedures, are inadmissible with respect to the above vector risk function (3.2) and given such a stepwise testing procedure, it is always possible to construct other multiple testing procedures whose individual tests have smaller risk compared to those for the given procedure. Moreover, an inadmissible procedure with respect to the vector risk function (3.2) will necessarily become inadmissible whenever the overall risk is an increasing function of the expected number of type I and type II errors, such as, when the risk is the expected number of misclassified hypotheses. One would certainly not expect a multiple testing procedure to be inadmissible. However, as we shall see later in this section, that, our proposed BSD method possesses such desired admissibility property.

3.1 Connection to the MRD method

Before we move further, we shall first establish an important connection between the proposed BSD procedure and the Maximum Residual Down (MRD) method, introduced by Cohen, Sackrowitz and Xu (2009), by showing that there exists a functional relationship between the proposed BSD statistics with those of the MRD statistics. This result would be essential for showing that our proposed multiple testing procedure based on the BSD statistics will be admissible for the present testing problem. Recall that, the MRD method due to Cohen, Sackrowitz and Xu (2009) is based on the set of adaptively formed residuals defined as:

$$U_{ij}(X) = \{X_j - \frac{\sigma_j}{\sigma_{i,j}}\sum_{i=1,i\neq j}^{j-1} \Sigma_{i,j} X_{(i+1)} \} \bigg/ \sigma_j \bigg( \frac{1}{2} \right)$$

for $t, j = 1, \cdots, m, 1 \leq i_1 \neq \cdots \neq i_{t-1} \leq m$ and $i_l \neq j$ for all $l = 1, \cdots, t - 1$. For $1 \leq t \leq m$, we define the index $j_t'(X)$ as

$$j_t'(X) = \arg\max_{j \in \{1, \cdots, m\} \setminus \{j_1'(X), \cdots, j_{t-1}'(X)\}} |U_{ij_t}(X)|. \quad (3.3)$$

Given a set of positive constants $C_1 \geq C_2 \geq \cdots \geq C_m$, the MRD method works in a step-down manner as follows:

1. At stage 1, let us consider the functions $|U_{ij}(X)|, \ j \in \{1, \cdots, m\}$. If $|U_{ij}(X)| \leq C_1$, we stop and accept all $H_{0j}$'s. Otherwise we reject $H_{0j}$ and continue to stage 2.
2. At stage 2, we consider the functions $|U_{2j}^{(X)}(X)|$, $j \in \{1, \cdots, m\} \setminus \{j_1(X)\}$. If $|U_{2j}^{(X)}(X)| \leq C_2$, we stop and accept all the remaining $H_0i$’s. Otherwise we reject $H_0j(X)$ and continue to the next stage.

3. In general, at stage $t$, we consider the $(m - t + 1)$ functions $|U_{tb}^{(X)}(X)|$, $j \in \{1, \cdots, m\} \setminus \{j_1(X), \cdots, j_{t-1}(X)\}$. If $|U_{tb}^{(X)}(X)| \leq C_t$, we stop and accept all the remaining $H_0i$’s. Otherwise we reject $H_0j(X)$ and move on to the stage $(t + 1)$.

4. We continue in this fashion until an acceptance occurs or there are no more null hypothesis to be tested, in which case we must stop.

**Remark 3.1.** Note that the indices $j_i(X)$ and $j'_i(X)$ defined in (3.8) and (3.9), respectively, need not necessarily be the same.

The next theorem characterises the relationship between the proposed BSD method and the MRD method due to Cohen, Sackrowitz and Xu (2009).

**Theorem 3.1.** Under the present set-up, the BSD statistics and the MRD statistics are associated through the following functional relationship:

$$S_{ij}^{(i_1, \cdots, i_{t-1})}(X) = \frac{p}{1-p} \times \frac{\sigma_{ji}^{(i_1, \cdots, i_{t-1})}}{\sqrt{V + \sigma_{ji}^{(i_1, \cdots, i_{t-1})}}} \times \exp \left\{ \frac{V}{2(V + \sigma_{ji}^{(i_1, \cdots, i_{t-1})})} |U_{ij}^{(i_1, \cdots, i_{t-1})}(X)|^2 \right\} \quad (3.4)$$

**Proof.** Observe that one can write each test statistic $S_{ij}$ as

$$S_{ij}^{(i_1, \cdots, i_{t-1})}(X) = \frac{\pi(\nu_j = 1, \nu^{(i_1, \cdots, i_{t-1})} = 0)f(X^{(i_1, \cdots, i_{t-1})}\nu_j = 1, \nu^{(i_1, \cdots, i_{t-1})} = 0)}{\pi(\nu_j = 0, \nu^{(i_1, \cdots, i_{t-1})} = 0)f(X^{(i_1, \cdots, i_{t-1})}\nu_j = 0, \nu^{(i_1, \cdots, i_{t-1})} = 0)} = \frac{p}{1-p} \times \frac{f(X^{(i_1, \cdots, i_{t-1})}\nu_j = 1, \nu^{(i_1, \cdots, i_{t-1})} = 0)}{f(X^{(i_1, \cdots, i_{t-1})}\nu_j = 0, \nu^{(i_1, \cdots, i_{t-1})} = 0)} \quad (3.5)$$

Let us write

$$\Sigma_{0i,j}^{(i_1, \cdots, i_{t-1})} = \Sigma_{(i_1, \cdots, i_{t-1})} + V\text{diag}(\nu_j = 0, \nu^{(i_1, \cdots, i_{t-1})} = 0)$$
and

$$\Sigma_{1i,j}^{(i_1, \cdots, i_{t-1})} = \Sigma_{(i_1, \cdots, i_{t-1})} + V\text{diag}(\nu_j = 1, \nu^{(i_1, \cdots, i_{t-1})} = 0).$$

It is important to note that for any $t = 1, \cdots, m$ and for any $j = 1, \cdots, m$, where $i_t \neq j$ for all $l$, we have the following:

$$\Sigma_{0i,j}^{(i_1, \cdots, i_{t-1})} = (\Sigma_{(i_1, \cdots, i_{t-1})} + V\text{diag}(\nu_j = 0, \nu^{(i_1, \cdots, i_{t-1})} = 0))_{(-j,-j)}$$

$$= \left(\Sigma_{(i_1, \cdots, i_{t-1})}_{(-j,-j)} + \Sigma_{(i_1, \cdots, i_{t-1})}\right)_{(-j,-j)}$$

$$= \Sigma_{(-j,-j)} \quad (3.6)$$

and

$$\Sigma_{1i,j}^{(i_1, \cdots, i_{t-1})} = (\Sigma_{(i_1, \cdots, i_{t-1})} + V\text{diag}(\nu_j = 1, \nu^{(i_1, \cdots, i_{t-1})} = 0))_{(-j,-j)}$$

$$= \left(\Sigma_{(i_1, \cdots, i_{t-1})}_{(-j,-j)} + \Sigma_{(i_1, \cdots, i_{t-1})}\right)_{(-j,-j)}$$

$$= \Sigma_{(-j,-j)} \quad (3.7)$$

which implies that for any $t = 1, \cdots, m$ and for any $j = 1, \cdots, m$, with $i_t \neq j$ for all $l$, we have

$$\Sigma_{0i,j}^{(i_1, \cdots, i_{t-1})} = \Sigma_{1i,j}^{(i_1, \cdots, i_{t-1})} \quad (3.8)$$
where $A_{[-j,-j]}$ denotes the submatrix of a matrix $A$ obtained after removing its $j^{th}$ row and $j^{th}$ column. Now, $f(X^{(i_1,\ldots,i_{t-1})}|\mu_j = 0, \nu^{(i_1,\ldots,i_{t-1}-1)} = 0)$ corresponds to the probability density function of a $N(0, \Sigma_{0,j})$ distribution. Therefore using equation (2.4) one can write,

$$
f(X^{(i_1,\ldots,i_{t-1})}|\mu_j = 0, \nu^{(i_1,\ldots,i_{t-1}-1)} = 0) = N(u_j^{(i_1,\ldots,i_{t-1})}(X),\sigma_j^{(i_1,\ldots,i_{t-1}-1)}(X_j))
\times N(0, \Sigma_{(i_1,\ldots,i_{t-1},j)})(X^{(i_1,\ldots,i_{t-1}-1)})
$$

(3.9)

where $u_j^{(i_1,\ldots,i_{t-1})}(X) = \sigma_j^{(i_1,\ldots,i_{t-1})}T \Sigma_{(i_1,\ldots,i_{t-1},j)} X^{(i_1,\ldots,i_{t-1},j)}$ and the term on the right hand side of equation (2.7) denote the probability densities of the corresponding normal distributions evaluated at appropriate points.

In a similar way one can write the following:

$$
f(X^{(i_1,\ldots,i_{t-1})}|\mu_j = 1, \nu^{(i_1,\ldots,i_{t-1}-1)} = 0) = N(u_j^{(i_1,\ldots,i_{t-1})}(X),V + \sigma_j^{(i_1,\ldots,i_{t-1}-1)}(X_j))
\times N(0, \Sigma_{(i_1,\ldots,i_{t-1},j)})(X^{(i_1,\ldots,i_{t-1}-1)})
$$

(3.10)

Equations (3.1) and (3.2) coupled with equations (3.5) and (3.6), complete the proof of the present theorem.

3.2 Admissibility of the BSD Procedure

In this subsection, we show that the proposed BSD method would become admissible when $X$ is assumed to follow a multivariate normal distribution with a fixed, but unknown mean vector $\mu$ and a known covariance matrix $\Sigma$. As in Cohen, Sackrowitz and Xu [2009], we shall use a result due to Matthes and Truax [1967] which gives a necessary and sufficient condition for the admissibility of a test of $H_{01}$ vs $H_{A1}$ when the joint distribution of $X$ belongs to an exponential family. We emphasise in this context that although the BSD test statistics can be expressed as functions of the corresponding MRD statistics, admissibility of the BSD method does not follow as a direct consequence of that of the MRD procedure. We adopt the broad architecture of the proof due to Cohen, Sackrowitz and Xu [2009]. However, it should be carefully noted that for each $t$, the functional relationship between the proposed BSD statistics and the corresponding MRD statistics in (3.4) involves the terms $\sigma_j^{(j_1(x),\ldots,j_{t-1}(x))}$ which depend on the set of indices $j_1(x),\ldots,j_{t-1}(x)$. Each of these indices $j_1(x),\ldots,j_{t-1}(x)$ is a function of the observed vector $x$ and they indicate the null hypotheses those have already been rejected before the $t$-th stage. It therefore becomes necessary to understand certain behaviour of these terms $\sigma_j^{(j_1(x),\ldots,j_{t-1}(x))}$ as a function of $x$ in the decision making process when the data $x$ is observed. Such behaviour would be extremely crucial for proving the admissibility of the proposed BSD method as we shall see later in this section. In this process, we establish that any step down procedure based on a set of statistics $S_{tj}$, which are strictly increasing functions of the corresponding $|U_{tj}|'$s, would be admissible, thus extending the results of Cohen, Sackrowitz and Xu [2009] for the present multiple testing problem.

Let $\phi_j(x)$ denotes the test function induced by the BSD procedure for testing $H_{01}$ vs $H_{A1}$ when we observe the data point $x$. The following result is due to Matthes and Truax [1967] which provides a necessary and sufficient condition for the admissibility of a testing procedure for testing $H_{01}$ vs $H_{A1}$ when $\Sigma$ is known.

Let $Y = \Sigma^{-1}X$.

**Lemma 3.1.** A necessary and sufficient condition for a test $\phi(y)$ of $H_{01}$ vs $H_{A1}$ to be admissible is that, for almost every fixed $y_2,\ldots,y_m$, the acceptance region of the test is an interval in $y_1$.

**Proof.** See Matthes and Truax [1967].
Note that, for fixed \((y_2, \cdots, y_m)\), to study the test function \(\phi(y) = \phi_1(x)\) as \(y_1\) varies, consider sample points \(x + rg\) where \(g\) is the first column of \(\Sigma\) and \(r\) varies. This is true, since \(y\) is a function of \(x\), and so \(y\) evaluated at \(x + rg\) is \(\Sigma^{-1}(x + rg) = y + (r, 0, \cdots, 0) = (y_1 + r, y_2, \cdots, y_m)\).

**Lemma 3.2.** The functions \(U_{ij}\) as given in equation have the following properties.

For \(t \in \{1, \cdots, m\}\) and for \(j_1, \cdots, j_{i-1} \in \{2, \cdots, m\}\) with \(j_i \neq j_{i'}\) for \(i \neq i'\),

\[
U_{ij}^{(j_1, \cdots, j_{i-1})}(x + rg) = U_{ij}^{(j_1, \cdots, j_{i-1})}(x) + r\sigma_i^{j_1}(j_1, \cdots, j_{i-1})
\]

For \(t \in \{1, \cdots, m\}\) and for \(j \in \{2, \cdots, m\} \setminus \{j_1, \cdots, j_{i-1}, j_1 \neq 1, \cdots, j_m \neq 1\}\),

\[
U_{ij}^{(j_1, \cdots, j_{i-1})}(x + rg) = U_{ij}^{(j_1, \cdots, j_{i-1})}(x)
\]

**Proof.** See the proof of Lemma (3.2) of Cohen, Sackrowitz and Xu (2009).

**Corollary 3.1.** For any \(r \in \mathbb{R}\), we have

\[
U_{ij}(x + rg) = U_{ij}(x) \text{ for all } j \in \{2, \cdots, m\},
\]

which, in turn, implies the following:

\[
S_{ij}(x + rg) = S_{ij}(x) \text{ for all } j \in \{2, \cdots, m\}.
\]

**Remark 3.2.** Since \(\sigma_1^{(j_1, \cdots, j_{i-1})} > 0\), it follows from Lemma 5.2 that for any fixed \(x \in \mathbb{R}^m\) and given any \((j_1, \cdots, j_{i-1})\), \(|U_{ij}^{(j_1, \cdots, j_{i-1})}(x + rg)|\) initially decreases with \(r\) and then increases as \(r\) increases. Also for each fixed \(x \in \mathbb{R}^m\) and given any \((j_1, \cdots, j_{i-1})\), \(U_{ij}^{(j_1, \cdots, j_{i-1})}(x + rg)\) is a strictly increasing function of \(r\). Therefore, when \(|U_{ij}^{(j_1, \cdots, j_{i-1})}(x + rg)|\) decreases in \(r\), \(U_{ij}^{(j_1, \cdots, j_{i-1})}(x + rg)\) is negative, while when \(|U_{ij}^{(j_1, \cdots, j_{i-1})}(x + rg)|\) is increasing in \(r\), \(U_{ij}^{(j_1, \cdots, j_{i-1})}(x + rg)\) is positive.

**Remark 3.3.** In Lemma 3.2 of Cohen, Sackrowitz and Xu (2004), the term \(\sigma_1^{(j_1, \cdots, j_{i-1})}/2\) was dropped, most likely, due to some typographical error. However, the presence of this term require some extra attention in our situation. As already mentioned in the beginning of this section that the term \(\sigma_1^{(j_1(x), \cdots, j_{i-1}(x))}\) depends on a set of indices \(j_1(x), \cdots, j_{i-1}(x)\), each of which is a function of the observed data vector \(x\). It, therefore, becomes necessary to know how this term \(\sigma_1^{(j_1(x), \cdots, j_{i-1}(x))}\) behaves as \(x\) varies. This will become evident through Lemma 3.3 given below.

Suppose \(\phi_1(x^*) = 0\), when \(x^*\) is observed, that is, \(x^*\) is an acceptance point of \(H_{01}\). Then the process must stop before \(H_{01}\) gets rejected. Suppose the testing procedure stops at some stage \(t\) without rejecting \(H_{01}\). Let \(x^* + r_0g\) be a point of rejection of \(H_{01}\), that is, \(\phi_1(x^* + r_0g) = 1\). Let the testing procedure reject \(H_{01}\) at some stage \(t_0\) when \(x^* + r_0g\) is observed. The next lemma establish an important identity between the set of indices \(j_l(x^* + r_0g)\) and \(j_l(x^*)\), \(1 \leq l \leq t_0 - 1\), which shows that these indices will remain invariant when \(\min\{t, t_0\} > 1\).

**Lemma 3.3.** Under the conditions \(\phi_1(x^*) = 0\) and \(\phi_1(x^* + r_0g) = 1\) the following holds when \(t > 1\) and \(t_0 > 1\):

\[
\hat{j}_{l}(x^* + r_0g) = \hat{j}_{l}(x^*) \text{ for all } l = 1, \cdots, t_0 - 1.
\]

**Proof.** See Appendix.

Lemma 3.3, coupled with Corollary 3.1, leads us to the following important result on the relation between \(t_0\) and \(t\) defined before.

**Lemma 3.4.** Under the conditions \(\phi_1(x^*) = 0\) and \(\phi_1(x^* + r_0g) = 1\), the BSD procedure must reject \(H_{01}\) within \(t\) steps when \(x^* + r_0g\) is observed, that is, \(t_0 \leq t\), where \(t_0\) and \(t\) are defined as before.
Proof. See Appendix. ■

Lemma 3.5. Suppose that for some \( x^* \) and \( r_0 > 0 \), \( \phi_1(x^*) = 0 \) and \( \phi_1(x^* + r_0 g) = 1 \). Then \( \phi_1(x^* + r_0 g) = 1 \) for all \( r > r_0 \).

Proof. See Appendix. ■

Theorem 3.2. For the given two sided multiple testing problem, the BSD procedure based on the statistics \( S_{ij} \)’s is admissible.

Proof. Recall that for the present testing problem admissibility of a multiple testing procedure implies that it would be admissible for each individual testing problem. Now using Lemma 3.1 and Lemma 3.5, it follows that for testing \( H_{01} \) vs \( H_{A1} \) the test \( \phi_1(x) \), induced by the BSD method, would be admissible. Proof that the other tests induced by the BSD method for the remaining individual testing problems will be admissible would follow analogously. ■

Remark 3.4. A careful inspection of the proof of Lemma 3.5 in Appendix reveals that one does not need the functional form of the statistics \( S_{ij} \)’s for proving the aforementioned convexity property stated in that lemma. What was only needed there was the fact that the \( S_{ij} \)’s are non-decreasing functions of the corresponding \( |U_{ij}| \)’s. This observation leads us to the following theorem from which admissibility of the MRD procedure due to Cohen, Sackrowitz and Xu (2009) follows immediately.

Theorem 3.3. For the present two sided multiple testing problem, any multiple testing procedure based on a set of statistics \( S_{ij} \), where the \( S_{ij} \)’s are non-decreasing functions of the absolute values of the corresponding MRD statistics \( |U_{ij}| \)’s, will be admissible with respect to the vector risk function.

4 Estimation of the proportion of non-nulls \( p \) and the variance \( V \) of the non-zero means

As already mentioned in the introduction that the BSD method proposed in Section 2, involves quantities like the proportion of true alternatives \( p \), the variance \( V \) of the distribution of the non-null \( \mu_i \)’s and the posterior probabilities of the form \( \pi(\nu_j = i, \nu^{(j_1, \cdots, j_{l-1}, j)}) = 0, X^{(j_1, \cdots, j_{l-1})} \), \( i = 0, 1 \), \( t, j = 1, \cdots, m \), \( 1 \leq j_1 \neq \cdots \neq j_{l-1} \leq m \) with \( j_l \neq j \) for all \( l \), which are usually not known in practice and are required to be estimated from the data. One natural approach of estimating them is to use appropriate hyperpriors for \( p \) and \( V \) and then finding their full Bayes estimates through employing an appropriate MCMC algorithm. Moreover, using the full Bayes estimates of \( p \) and \( V \) one can obtain an empirical Bayes version of the BSD procedure by plugging those full Bayes estimates into the functional relation \( (3.3) \) of Theorem 3.1. However, finding an efficient MCMC algorithm that works well in large dimensions may not be an easy as traditional Gibbs sampling and Metropolis-Hastings algorithms are known to break down in such situations. On the other hand, using \( (3.3) \) of Theorem 3.1, it follows that if we can obtain some good estimates \( p \) and \( V \) from the data otherwise, then by plugging those estimates into \( (3.3) \), one can directly enumerate the BSD statistics. In this paper, we prefer using the latter approach and thus avoid the need for MCMC-type computations.

We now turn our attention to the estimation of \( p \) and \( V \). Consistent estimation of the theoretical proportion of true alternatives has been so far one of the most important and challenging tasks in multiple testing and related inferential problems. Recall that the classical Bonferroni procedure and the BH method are both conservative by a factor of \((1 - p) \) at their respective FWER and FDR significance levels. Hence by exploiting information through an estimate \( \hat{p} \) of the theoretical proportion of true alternatives \( p \) and replacing \( \alpha/m \) by \( \alpha/(m(1 - \hat{p})) \) in the corresponding critical constants, one can enhance the power of these multiple testing procedure substantially. Moreover, there are situations where one may be more interested in the proportion of truly active signals, rather than detecting the true signals. See Jin (2006) and Jin (2008) in this context. A number of efforts have been made towards estimation of this proportion. Example of some early works includes
Benjamini and Hochberg (2000), Efron et al (2001), Storey (2002), Storey et al (2004) and Genovese and Wasserman (2004). Performances of these methods are, however, limited to situations when \( p \) is very small and they are found to be inconsistent, in general. In a pioneering work, Efron (2004) considers a Gaussian mixture model approach as in Section 2 and proposes a natural estimate of the proportion of true nulls \((1 - p)\) based on the empirical distribution of the test statistics \(X_i\)’s. His method particularly works well when \( p \leq 0.1 \). Efron’s estimate, however, is found to be inconsistent in general and tend to underestimate \( p \), specially when \( p \) is moderately large. In an important article, Meinshausen and Rice (2006) propose a 100\( (1 - \alpha) \) percent lower confidence bound for \( p \) based on the empirical distribution of the underlying \( p \)-values. Their method, however, is conservative and inconsistent, in general.

A major theoretical breakthrough in this direction has been made in Jin (2006). He proposes an estimator of \( p \) by exploiting certain concepts from Fourier analysis when the underlying test statistics are independent and identically distributed according to a Gaussian mixture model as in Section 2. His proposed estimator is based on the central idea of approximating, what he called the underlying characteristic function, by the corresponding empirical characteristic function when the null parameter values are identical or homogeneous and is shown to be uniformly consistent over a large parameter space. Detailed discussion on the construction of such estimators can be found in Jin (2006) and Jin (2008). In another important paper, Jin and Cai (2007) extend these works by consistently estimating the null parameters values along with the proportion of non-nulls in case the null distributions are assumed to be unknown and the null parameters are heterogeneous. Their estimators are shown to be consistent over a large parameter space and also in situations when the test statistics exhibit \( \alpha \)-mixing and short range dependence. These estimators, though consistent, fail to attain any optimal rate of convergence. In a more recent work, Cai and Jin (2010) consider the problem of finding consistent estimators of the null parameters and the proportion of non-nulls which attain the corresponding minimax optimal rates of convergence under an i.i.d. Gaussian mixture framework. For any fixed \( \gamma \in (0, 1/2) \), they propose the following estimate of \( p \), given by,

\[
\hat{p}(\gamma) = 1 - \frac{1}{m^{1-\gamma}} \sum_{j=1}^{m} \cos \left( \sqrt{2\gamma \log m} X_j \right)
\]

and show that the above estimator of \( p \) attains the corresponding minimax rate in situations when the parameter \( p \) is not too small compared to the total number of tests \( m \) and “vanishes asymptotically” as \( m \) grows to infinity. Cai and Jin (2010) conjecture that the above estimator of \( p \) in (4.1) will remain consistent under certain forms of weakly correlated structures. In this paper, we use this estimator \( \hat{p}(\gamma) \), defined in (4.1) above, for estimating the proportion \( p \) of non-null hypotheses. We show that the conjecture of Cai and Jin (2010) is indeed affirmative in the sense that their estimator remains consistent under certain weakly correlated structures, such as, finite block dependence, short range dependence, certain intraclass correlation model where the common correlation coefficient goes to zero at an appropriate rate as the number of tests grows to infinity, and also in situations when \( p \) is moderately sparse. It should be emphasised in this context that estimation of the theoretical proportion of true alternatives \( p \) under stronger form of dependence is indeed a very difficult problem to solve and is beyond the scope of the present study. We hope to address this problem somewhere else in future. We further assume that the variance parameter \( V \equiv V_m \) of the alternative distribution of the \( \mu_i \)’s varies with \( m \). It is natural to assume \( V_m \) to be large when \( m \) is large so that large signals can occur with a positive probability and thus can be detected easily. So, \( V_m \) is assumed to go to infinity at an appropriate rate as \( m \) grows to infinity. We consider the following asymptotic framework under which the aforesaid consistency results hold:

**Assumption 4.1.**

1. \( p \equiv p_m \to 0 \) as \( m \to \infty \) (asymptotically vanishing sparsity)

2. \( V \equiv V_m \to \infty \) as \( m \to \infty \)

3. \( p_m V_m = O(1) \) as \( m \to \infty \)
The following theorem shows that under Assumption (4.1), \( \hat{p}(\gamma) \) in (4.1) consistently estimates the proportion \( p_m \) of true alternatives for certain weakly correlated structure \( \Sigma \).

**Theorem 4.1.** Consider the mixture model (2.4) in Theorem 2.1, where \( \Sigma = (\sigma_{ij}) \) denotes the correlation matrix associated with the random vector \( X \). Let \( \hat{p}(\gamma) \) be an estimator of \( p \equiv p_m \) as defined in (4.1). Then under Assumption (4.1), if \( 0 < \gamma < 1/2 \) be such that \( m^2 \gamma p^2 \to \infty \) as \( m \to \infty \), and

\[
\lim_{m \to \infty} \frac{1}{m^2 p^2} \sum_{1 \leq i,j \leq m, i \neq j} \left( \frac{1}{m \sigma_{ij}} - \frac{1}{m^{-\gamma} \sigma_{ij}} \right)^2 = 0 \tag{4.2}
\]

then the following holds:

\[
\frac{\hat{p}(\gamma)}{p} \to 1 \text{ as } m \to \infty, \tag{4.3}
\]

where the above probability convergence is taken with respect to the joint distribution of \( X_1, X_2, \cdots \), given by (2.4).

**Proof.** See Appendix. ■

We now turn our attention for estimation of the variance \( V \equiv V_m \) of the alternative distribution of \( \mu_i \)'s. For that we observe that since \( \Sigma \) is assumed to be a correlation matrix, using Theorem 2.1, the common marginal distribution of the \( X_i \)'s is given by,

\[
X_i \sim (1 - p) N(0, 1) + pN(0, 1 + V) \tag{4.4}
\]

whence we have,

\[
E(X_i^2) = (1 - p) + p(1 + V) = 1 + pV, \text{ for each } i = 1, \cdots, m. \tag{4.5}
\]

Based on (4.5), we consider the following moment-based estimator of \( V \), given by,

\[
\hat{V} = \frac{1}{\hat{p}} \left( \frac{1}{m} \sum_{i=1}^{m} X_i^2 - 1 \right) \tag{4.6}
\]

where \( \hat{p} \) is an estimate of the proportion of true non-nulls \( p \).

The next theorem shows that whenever some consistent estimator of \( p \) is available, the above estimator \( \hat{V} \) will consistently estimate \( V \) under pretty general conditions.

**Theorem 4.2.** Consider the mixture model (2.4) in Theorem 2.1, where \( \Sigma = (\sigma_{ij}) \) denotes the correlation matrix associated with the random vector \( X \). Suppose \( \hat{p} \) is a consistent estimator of \( p \). Then under Assumption (4.1), if \( mp^2 V^2 \to \infty \) as \( m \to \infty \) and

\[
\lim_{m \to \infty} \left[ \frac{1}{m^2 p^2 V^2} \sum_{1 \leq i,j \leq m, i \neq j} \sigma_{ij}^2 \right] = 0 \tag{4.7}
\]

then

\[
\frac{\hat{V}}{V} \to 1 \text{ as } m \to \infty, \tag{4.8}
\]

where the above probability convergence is taken with respect to the joint distribution of \( X_1, X_2, \cdots \) in (2.4).

**Proof.** See Appendix. ■
5 Alternative Representation of BSD Statistics

Observe that in order to perform the BSD procedure based on the statistics $S_{ij}$’s as given by equation (2), at each stage $t$ we need to find the inverses of $m-t+1$ many submatrices of $\Sigma$ which can be a troublesome issue for the computation of the BSD and the MRD procedure as well, specially when $m$ is large. In this section we derive two important algebraic identities that would lead us to an alternative representation of the BSD statistics $S_{ij}$’s resulting in a huge amount of computational savings and facilitate the computation of both the procedures to a great extent.

**Lemma 5.1.** For any arbitrary variance-covariance matrix $\Sigma$ and for any $\nu \in \{0,1\}^m$, we have the following identity:

$$(\Sigma + VB_{\nu,\nu_i=1})^{-1} = (\Sigma + VB_{\nu,\nu_i=0})^{-1} - \frac{V}{1 + V b_{ii}(\nu)} b_i(\nu)b_i(\nu)^T$$

where $b_i(\nu)$ denotes the $i$-th column vector of the matrix $(\Sigma + VB_{\nu,\nu_i=1})^{-1}$ and $b_{ii}(\nu)$ is the $i$-th element of $b_i(\nu)$, that is, the $i$-th diagonal element of $(\Sigma + VB_{\nu,\nu_i=0})^{-1}$.

Proof. See Appendix. \[\blacksquare\]

As a consequence of Lemma 5.1, it immediately follows that, for any $x \in \mathbb{R}^m$ and for any $\nu \in \{0,1\}^m$,

$$x^T(\Sigma + VB_{\nu,\nu_i=1})^{-1}x = x^T(\Sigma + VB_{\nu,\nu_i=0})^{-1}x - \frac{V}{1 + V b_{ii}(\nu)} x^T b_i(\nu)b_i(\nu)^T x$$

$$= x^T(\Sigma + VB_{\nu,\nu_i=0})^{-1}x - \frac{V b_{ii}(\nu)}{1 + V b_{ii}(\nu)} (\sum_{i=1}^m b_{ij}(\nu)x_j)^2,$$

where $b_{ij}(\nu)$ denotes the $j$-th element of $b_i(\nu)$ already defined in Lemma 5.1.

**Lemma 5.2.** For any arbitrary positive definite covariance matrix $\Sigma$ and for any $\nu \in \{0,1\}^m$, we have the following identity:

$$\frac{|\Sigma + VB_{\nu,\nu_i=1}|}{|\Sigma + VB_{\nu,\nu_i=0}|} = 1 + V b_{ii}(\nu)$$

where $b_{ii}(\nu)$ denotes the $i$-th diagonal element of the matrix $(\Sigma + VB_{\nu,\nu_i=0})^{-1}$.

Proof. See Appendix. \[\blacksquare\]

**Lemma 5.3.** For any arbitrary positive definite covariance matrix $\Sigma$ and for any $\nu \in \{0,1\}^m$, we have the following identity:

$$b_{ii}(\nu) = \begin{bmatrix} \sigma_{ii} - \sigma_{(-i)}(\Sigma + VB_{\nu,\nu_i=0})(-i,-i)^{-1}\sigma_{(-i)} \end{bmatrix}^{-1}$$

and

$$(b(\nu))_{(-i)} = -b_{ii}(\nu)(\Sigma + VB_{\nu,\nu_i=0})(-i,-i)^{-1}\sigma_{(-i)}$$

where $b_{ii}(\nu)$ is same as in Lemma 5.2 and $(b(\nu))_{(-i)}$ denotes the vector obtained from $b_i(\nu)$ after removing its $i$-th coordinate and $(\Sigma + VB_{\nu,\nu_i=0})(-i,-i)$ is the submatrix obtained by removing the $i$-th row and the $i$-th column of $\Sigma + VB_{\nu,\nu_i=0}$.

Proof. See Appendix. \[\blacksquare\]

**Lemma 5.4.** For any arbitrary positive definite covariance matrix $\Sigma$ and for each $i = 1, \cdots, m$ we have the following:

$$f(x|\nu_i=1, \nu_{(-i)} = 0) = \sqrt{\frac{1}{1 + V b_{ii}}} \times \exp\left\{\frac{V}{2(1 + V b_{ii})} \left(\sum_{j=1}^m b_{ij}x_j\right)^2\right\}$$

where $b_i = (b_{ii}, \cdots, b_{im})^T$ denotes the $i$-th column vector of the precision matrix $\Sigma^{-1}$. \[\blacksquare\]
such computational issues. We only need to compute the inverse of $\Sigma$ when $m$ is large. However, the representation given by Theorem (5.1) above helps us to avoid all $(SMW)$ identity.

\[ m \]

\[ n \]

is reported to be zero, and thus producing erroneous results. This might happen for some $\Sigma$ equation (5.1) above, the denominator might individually be so small, that the computer might point even for moderately large $m$. Moreover, while computing the ratio of determinants given in Proposition 7.1.

7.1 Some auxiliary results

An immediate consequence of Theorem 5.1 is that at the $t$-th stage of the BSD procedure, we do not need to find the $(m-t+1)$ many submatrices $\Sigma_{(j_1, \ldots, j_{t-1}(x), j)}$, nor do we need to find the $(m-t+1)$ many ratios of determinants

\[ \frac{|\Sigma_{(j_1(x), \ldots, j_{t-1}(x))} + VB_{\beta_{j_1(x), \ldots, j_{t-1}(x)}}|_{\beta_{j_1(x), \ldots, j_{t-1}(x)} = 0}}{|\Sigma_{(j_1(x), \ldots, j_{t-1}(x))} + VB_{\beta_{j_1(x), \ldots, j_{t-1}(x)}}|_{\beta_{j_1(x), \ldots, j_{t-1}(x)} = 1}} \]

(5.1)

at each step $t$, for $t = 1, \ldots, m$, which might be a troublesome issue from computational viewpoint even for moderately large $m$. Moreover, while computing the ratio of determinants given in equation (5.1) above, the denominator might individually be so small, that the computer might report it to be zero, and thus producing erroneous results. This might happen for some $\Sigma$, specially when $m$ is large. However, the representation given by Theorem (5.1) above helps us to avoid all such computational issues. We only need to compute the inverse of $\Sigma_{(j_1(x), \ldots, j_{t-1}(x))}$ whose column vectors will be used for computing the $(m-t+1)$ statistics $S_{tj}^{(j_1(x), \ldots, j_{t-1}(x))}(x)$. Thus the overall BSD procedure becomes computationally very fast compared to its original formulation.

6 Simulations

We shall update the present version of this article very soon after including the simulation results in this section.

7 Appendix

7.1 Some auxiliary results

Proposition 7.1. For any $n \times 1$ vector $v$ and any $n \times n$ symmetric positive definite matrix $A$,

\[ \int_{\mathbb{R}^n} \exp(-\frac{1}{2}w^TA^{-1}w + v^Tw)dw = (2\pi)^{\frac{n}{2}}|A|^\frac{1}{2} \exp(\frac{1}{2}v^TAv). \]

Proof. See Lemma B.1.1 of Santner, Williams and Notz (2003).

Proposition 7.2. Suppose that $B$ is any $n \times n$ nonsingular matrix, $C$ is a $r \times r$ non-singular matrix, and $A$ is an arbitrary $n \times r$ matrix such that $(A^TB^{-1}A+C)^{-1}$ is nonsingular. Then $(B + A^TC^{-1}A)$ is $n \times n$ non-singular with inverse given by,

\[ (B + A^TC^{-1}A)^{-1} = B^{-1} - B^{-1}A(A^TB^{-1}A + C)^{-1}A^TB^{-1} \]

The above result from matrix algebra is popularly known as The Sherman-Morrison-Woodbury (SMW) identity.
Proof. See Lemma B.3.2 of Santner, Williams and Notz (2003). □

Proposition 7.3. Suppose

\[
\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N_2 \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix} \right) \quad (7.1)
\]

Then

\[
E \left[ \cos \left( \sqrt{2 \gamma \log m} Z_1 \right) \cos \left( \sqrt{2 \gamma \log m} Z_2 \right) \right] = \frac{1}{2} \left[ \exp \left\{ - (\sigma_1^2 + \sigma_2^2 + 2 \rho \sigma_1 \sigma_2) \gamma \log m \right\} + \exp \left\{ - (\sigma_1^2 + \sigma_2^2 - 2 \rho \sigma_1 \sigma_2) \gamma \log m \right\} \right] \quad (7.2)
\]

Proof. Using Euler’s formula we have for all \( x \in \mathbb{R} \),

\[
\cos(x) = \frac{1}{2} (e^{ix} + e^{-ix}). \quad (7.3)
\]

where \( i = \sqrt{-1} \). Therefore,

\[
E \left[ \cos \left( \sqrt{2 \gamma \log m} Z_1 \right) \cos \left( \sqrt{2 \gamma \log m} Z_2 \right) \right] = \frac{1}{4} E \left[ (e^{i \sqrt{2 \gamma \log m} Z_1} + e^{-i \sqrt{2 \gamma \log m} Z_1}) (e^{i \sqrt{2 \gamma \log m} Z_2} + e^{-i \sqrt{2 \gamma \log m} Z_2}) \right] \]

\[
= \frac{1}{4} E \left[ e^{i \sqrt{2 \gamma \log m} (Z_1 + Z_2)} + e^{i \sqrt{2 \gamma \log m} (Z_1 - Z_2)} + e^{-i \sqrt{2 \gamma \log m} (Z_1 + Z_2)} + e^{-i \sqrt{2 \gamma \log m} (Z_1 - Z_2)} \right]
\]

Recall that if \( Z \sim N_d(\theta, \Lambda) \) then the characteristic function of \( Z \) is given by,

\[
\phi_Z(t) = e^{i \theta^T t - \frac{1}{2} t^T \Lambda t}, \text{ for } t \in \mathbb{R}^d.
\]

Using this we obtain,

\[
E \left[ e^{i \sqrt{2 \gamma \log m} (Z_1 + Z_2)} \right] = e^{- (\sigma_1^2 + \sigma_2^2 + 2 \rho \sigma_1 \sigma_2) \gamma \log m} = E \left[ e^{i \sqrt{2 \gamma \log m} (-Z_1 - Z_2)} \right] \quad (7.5)
\]

and

\[
E \left[ e^{i \sqrt{2 \gamma \log m} (-Z_1 + Z_2)} \right] = e^{- (\sigma_1^2 + \sigma_2^2 - 2 \rho \sigma_1 \sigma_2) \gamma \log m} = E \left[ e^{i \sqrt{2 \gamma \log m} (Z_1 - Z_2)} \right] \quad (7.6)
\]

whence we have

\[
E \left[ \cos \left( \sqrt{2 \gamma \log m} Z_1 \right) \cos \left( \sqrt{2 \gamma \log m} Z_2 \right) \right] = \frac{1}{2} \left[ e^{- (\sigma_1^2 + \sigma_2^2 + 2 \rho \sigma_1 \sigma_2) \gamma \log m} + e^{- (\sigma_1^2 + \sigma_2^2 - 2 \rho \sigma_1 \sigma_2) \gamma \log m} \right]
\]

This completes the proof of Proposition 7.3. □

7.2 Proof of Theorem 2.1

Proof. Let us denote the vectors \((x_1, \cdots, x_m), (\mu_1, \cdots, \mu_m)\) and \((\nu_1, \cdots, \nu_m)\) by \(x, \mu\) and \(\nu\) respectively. The corresponding likelihood function is given by,

\[
f(x | \mu, \nu) = \frac{\exp\left\{-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right\}}{(2\pi)^{m/2} |\Sigma|^{1/2}}
\]

17
and the prior distribution of $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_m)$ is given by,
\[
\pi(\boldsymbol{\nu}) = \prod_{i=1}^{m} p^{\nu_i} (1-p)^{1-\nu_i}.
\]

Therefore, the marginal distribution of $X$ is given by,
\[
f(x) = \sum_{\boldsymbol{\nu} \in \{0,1\}^m} \pi(\boldsymbol{\nu}) f(x | \boldsymbol{\nu})
\]
\[
= \sum_{\boldsymbol{\nu} \in \{0,1\}^m} \pi(\boldsymbol{\nu}) \int_{\mathbb{R}^{|\boldsymbol{\nu}|}} f(x | \mu, \boldsymbol{\nu}) \prod_{i: \nu_i \neq 0} \frac{\exp(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu))}{(2\pi)^{m/2}|\Sigma|^{1/2}} \prod_{i: \nu_i = 1} \frac{\exp(-\frac{1}{2}\mu_i^2)}{\sqrt{2\pi} V^2} d\mu_i
\]
\[
= \sum_{\boldsymbol{\nu} \in \{0,1\}^m} \pi(\boldsymbol{\nu}) \int_{\mathbb{R}^{|\boldsymbol{\nu}|}} \frac{\exp(-\frac{1}{2}(\mu - x)^T \Sigma^{-1} (\mu - x))}{(2\pi)^{m/2}|\Sigma|^{1/2}} \prod_{i: \nu_i \neq 0} \frac{\exp(-\frac{1}{2}\mu_i^2)}{\sqrt{2\pi} V^2} \prod_{i: \nu_i = 1} \frac{\exp(-\frac{1}{2}\mu_i^2)}{\sqrt{2\pi} V^2} d\mu_i
\]
where $|\boldsymbol{\nu}| = \sum_{i=1}^{m} \nu_i = \#\{\nu_i : \nu_i = 1\}$ and the sum being taken over all possible choices of $\boldsymbol{\nu} \in \{0,1\}^m$.

Fix $\boldsymbol{\nu} \in \{0,1\}^m$.

Observe that $(\mu - x)^T \Sigma^{-1} (\mu - x)$ can also be written as
\[
(\mu - x)^T \Sigma^{-1} (\mu - x) = (\mu - x)^T P_\nu \Sigma^{-1} P_\nu^T (\mu - x) = (P_\nu (\mu - x))^T P_\nu \Sigma^{-1} P_\nu^T (P_\nu (\mu - x))
\]
where $P_\nu$ is a permutation matrix such that the first $(m - |\boldsymbol{\nu}|)$ components of $P_\nu (\mu - x)$ are precisely those components of $(\mu - x)$ for which $\mu_i = 0$ and the remaining $|\boldsymbol{\nu}|$ components of $P_\nu (\mu - x)$ correspond to those components of $(\mu - x)$ such that $\mu_i \neq 0$.

For notational convenience, let us write $P_\nu (\mu - x)$ as,
\[
P_\nu (\mu - x) = \begin{pmatrix} \mu_{1, \nu} - x_{1, \nu} \\ \mu_{2, \nu} - x_{2, \nu} \end{pmatrix}.
\]
Observe that, $\mu_{1, \nu} = 0_{(m - |\nu|) \times 1}$. Then
\[
\prod_{i: \nu_i \neq 0} \frac{\exp(-\frac{1}{2}\mu_i^2)}{\sqrt{2\pi} V^2} = \frac{\exp(-\frac{1}{2}\mu_{2, \nu}^T \mu_{2, \nu})}{(2\pi)^{|\nu|/2}}
\]
\[
= \frac{\exp(-\frac{1}{2}\mu_{2, \nu}^T (V I_{|\nu|})^{-1} \mu_{2, \nu})}{(2\pi)^{|\nu|/2}}
\]
Therefore,
\[
f(x | \nu) = \int_{\mathbb{R}^{|\nu|}} \exp\left\{ -\frac{1}{2} \begin{pmatrix} \mu_{1, \nu} - x_{1, \nu} \\ \mu_{2, \nu} - x_{2, \nu} \end{pmatrix}^T \Sigma^{-1}_{\nu} \begin{pmatrix} \mu_{1, \nu} - x_{1, \nu} \\ \mu_{2, \nu} - x_{2, \nu} \end{pmatrix} \right\} \frac{d\mu_{2, \nu}}{(2\pi)^{|\nu|/2}}
\]
\[
\times \frac{\exp\left\{ -\frac{1}{2}\mu_{2, \nu}^T (V I_{|\nu|})^{-1} \mu_{2, \nu} \right\}}{(2\pi)^{|\nu|/2}}
\]
where $\Sigma_\nu$ is a non-singular matrix such that $\Sigma_\nu^{-1} = P_\nu \Sigma^{-1} P_\nu^T$. Since $P_\nu$ is a permutation matrix, and hence orthogonal, we have,

\[
\Sigma_\nu = (P_\nu \Sigma^{-1} P_\nu^T)^{-1} = (P_\nu^T)^{-1}(\Sigma^{-1})^{-1}(P_\nu)^{-1} = (P_\nu^{-1})^T \Sigma P_\nu^T = P_\nu \Sigma P_\nu^T
\]

Consequently, $|\Sigma_\nu| = |P_\nu \Sigma P_\nu^T| = |\Sigma|$. Therefore,

\[
f(x|\nu) = \int_{\mathbb{R}^{|\nu|}} \exp\left\{ -\frac{1}{2} \left( \begin{array}{c} \mu_{1,\nu} - x_{1,\nu} \\ \mu_{2,\nu} - x_{2,\nu} \end{array} \right)^T \Sigma_\nu^{-1} \left( \begin{array}{c} \mu_{1,\nu} - x_{1,\nu} \\ \mu_{2,\nu} - x_{2,\nu} \end{array} \right) \right\} \\
\cdot \exp\left\{ -\frac{1}{2} \mu_{2,\nu}^T (VI_{|\nu|})^{-1} \mu_{2,\nu} \right\} \frac{1}{(2\pi)^{|\nu|/2}|\Sigma_\nu|^{|\nu|/2}} d\mu_{2,\nu}
\]

Let us partition the matrix $\Sigma_\nu$ as

\[
\Sigma_\nu = \begin{pmatrix} \Sigma_{11,\nu} & \Sigma_{12,\nu} \\ \Sigma_{21,\nu} & \Sigma_{22,\nu} \end{pmatrix}.
\]

Observe that

\[
\exp\left\{ -\frac{1}{2} \left( \begin{array}{c} \mu_{1,\nu} - x_{1,\nu} \\ \mu_{2,\nu} - x_{2,\nu} \end{array} \right)^T \Sigma_\nu^{-1} \left( \begin{array}{c} \mu_{1,\nu} - x_{1,\nu} \\ \mu_{2,\nu} - x_{2,\nu} \end{array} \right) \right\}
\]

is nothing but the probability density function of a $N_{m-|\nu|}(x_{1,\nu}, x_{2,\nu}, \Sigma_\nu)$ distribution, that can be written as the product of the probability densities of $N_{|\nu|}(x_{1,\nu}, \Sigma_{11,\nu})$ and $N_{m-|\nu|}(w_{\nu}, \Sigma_{2|1,\nu})$ distributions, where $w_{\nu} = x_{2,\nu} - \Sigma_{21,\nu}^{-1} \Sigma_{11,\nu} x_{1,\nu}$ and $\Sigma_{2|1,\nu} = \Sigma_{22,\nu} - \Sigma_{21,\nu} \Sigma_{11,\nu}^{-1} \Sigma_{12,\nu}$.

Hence

\[
\exp\left\{ -\frac{1}{2} \left( \begin{array}{c} \mu_{1,\nu} - x_{1,\nu} \\ \mu_{2,\nu} - x_{2,\nu} \end{array} \right)^T \Sigma_\nu^{-1} \left( \begin{array}{c} \mu_{1,\nu} - x_{1,\nu} \\ \mu_{2,\nu} - x_{2,\nu} \end{array} \right) \right\}
\]

\[
= \exp\left\{ -\frac{1}{2} \left( \begin{array}{c} \mu_{1,\nu} - x_{1,\nu} \\ \mu_{2,\nu} - w_{\nu} \end{array} \right)^T \Sigma_{2|1,\nu}^{-1} \left( \begin{array}{c} \mu_{1,\nu} - x_{1,\nu} \\ \mu_{2,\nu} - w_{\nu} \end{array} \right) \right\}
\]

\[
\times \exp\left\{ -\frac{1}{2} \left( \begin{array}{c} \mu_{2,\nu} - w_{\nu} \\ \mu_{2,\nu} - w_{\nu} \end{array} \right)^T \Sigma_{2|2,\nu}^{-1} \left( \begin{array}{c} \mu_{2,\nu} - w_{\nu} \\ \mu_{2,\nu} - w_{\nu} \end{array} \right) \right\}
\]

Therefore,

\[
f(x|\nu) = \frac{\exp\left\{ -\frac{1}{2} \left( \begin{array}{c} \mu_{1,\nu} - x_{1,\nu} \\ \mu_{2,\nu} - w_{\nu} \end{array} \right)^T \Sigma_{11,\nu}^{-1} x_{1,\nu} \right\}}{(2\pi)^{m-|\nu|/2}|\Sigma_{11,\nu}|^{m-|\nu|/2}}
\]

\[
\times \int_{\mathbb{R}^{|\nu|}} \frac{\exp\left\{ -\frac{1}{2} \left( \begin{array}{c} \mu_{2,\nu} - w_{\nu} \\ \mu_{2,\nu} - w_{\nu} \end{array} \right)^T \Sigma_{2|1,\nu}^{-1} \left( \begin{array}{c} \mu_{2,\nu} - w_{\nu} \\ \mu_{2,\nu} - w_{\nu} \end{array} \right) \right\}}{(2\pi)^{|\nu|/2}|\Sigma_{2|1,\nu}|^{|\nu|/2}} d\mu_{2,\nu}
\]

Now,

\[
(\mu_{2,\nu} - w_{\nu})^T \Sigma_{2|1,\nu}^{-1} (\mu_{2,\nu} - w_{\nu}) + \mu_{2,\nu}^T (VI_{|\nu|})^{-1} \mu_{2,\nu}
\]

\[
= \mu_{2,\nu}^T \Sigma_{2|1,\nu}^{-1} + (VI_{|\nu|})^{-1} \mu_{2,\nu}^2 - 2(\Sigma_{2|1,\nu}^{-1} w_{\nu})^T \mu_{2,\nu} + w_{\nu}^T \Sigma_{2|1,\nu}^{-1} w_{\nu}
\]

\[
= \mu_{2,\nu}^T A_{\nu} \mu_{2,\nu} - 2(\Sigma_{2|1,\nu}^{-1} w_{\nu})^T \mu_{2,\nu} + w_{\nu}^T \Sigma_{2|1,\nu}^{-1} w_{\nu}
\]

19
where $A_\nu = \Sigma_{2|1,\nu}^{-1} + (VI_\nu)^{-1}$. Therefore, using Proposition 6.1 it follows that,

$$f(x|\nu) = \frac{\exp\left\{ -\frac{1}{2} x^T \Sigma_{1|\nu}^{-1} x_{1,\nu} - \frac{1}{2} w^T \Sigma_{2|1,\nu}^{-1} w_\nu \right\}}{(2\pi)^{m/2} |\Sigma_{1|\nu}|^{1/2} \times (2\pi)^{|\nu|/2} |\Sigma_{2|1,\nu}|^{1/2}} \times \int_{\mathbb{R}^{|\nu|}} \exp\left\{ -\frac{1}{2} \mu_{2,\nu}^T A_\nu \mu_{2,\nu} + (\Sigma_{2|1,\nu}^{-1} w_\nu)^T \mu_{2,\nu} \right\} d\mu_{2,\nu}$$

$$= \frac{\exp\left\{ -\frac{1}{2} x^T \Sigma_{1|\nu}^{-1} x_{1,\nu} - \frac{1}{2} w^T \Sigma_{2|1,\nu}^{-1} w_\nu \right\}}{(2\pi)^{m/2} |\Sigma_{1|\nu}|^{1/2} \times (2\pi)^{|\nu|/2} |\Sigma_{2|1,\nu}|^{1/2}} \times \exp\left\{ \frac{1}{2} (\Sigma_{2|1,\nu}^{-1} w_\nu)^T A_\nu^{-1} (\Sigma_{2|1,\nu}^{-1} w_\nu) \right\}$$

$$= \frac{|A_\nu^{-1}|^{|\nu|/2}}{(2\pi)^{m/2} |\Sigma_{1|\nu}|^{1/2}} \exp\left\{ -\frac{1}{2} x^T \Sigma_{1|\nu}^{-1} x_{1,\nu} - \frac{1}{2} w^T (\Sigma_{2|1,\nu}^{-1} - \Sigma_{2|1,\nu}^{-1} A_\nu^{-1} \Sigma_{2|1,\nu}^{-1}) w_\nu \right\}$$

Next applying Proposition 6.2 we have,

$$\Sigma_{2|1,\nu} - \Sigma_{2|1,\nu}^{-1} A_\nu^{-1} \Sigma_{2|1,\nu}^{-1} = \Sigma_{2|1,\nu}^{-1} - \Sigma_{2|1,\nu}^{-1} (\Sigma_{2|1,\nu}^{-1} + (VI_\nu)^{-1})^{-1} \Sigma_{2|1,\nu}^{-1}$$

$$= \frac{1}{|A_\nu|} \Sigma_{1|\nu} A_\nu |V A_\nu| \Sigma_{2|1,\nu}^{-1}$$

Again note that $|A_\nu^{-1}| = \frac{1}{|A_\nu|}$ and $|\Sigma_\nu| = |\Sigma_{1|\nu}| |\Sigma_{2|1,\nu}|$. Therefore,

$$\frac{|A_\nu^{-1}|^{|\nu|/2}}{|V^{|\nu|/2}|\Sigma_\nu|^{1/2}} = \frac{1}{\sqrt{|V^{|\nu|/2}|\Sigma_{1|\nu}| |\Sigma_{2|1,\nu}|}}$$

Thus we have,

$$f(x|\nu) = \frac{\exp\left\{ -\frac{1}{2} x^T \Sigma_{1|\nu}^{-1} x_{1,\nu} \right\}}{(2\pi)^{m/2} |\Sigma_{1|\nu}|^{1/2}} \times \frac{\exp\left\{ -\frac{1}{2} (x_{2,\nu} - \Sigma_{21,\nu} \Sigma_{11,\nu}^{-1} x_{1,\nu})^T (\Sigma_{2|1,\nu} + V I_\nu)^{-1} (x_{2,\nu} - \Sigma_{21,\nu} \Sigma_{11,\nu}^{-1} x_{1,\nu}) \right\}}{(2\pi)^{|\nu|/2} |\Sigma_{2|1,\nu} + V I_\nu|^{1/2}}$$

which can be rewritten as

$$\frac{\exp\left\{ -\frac{1}{2} \left( \begin{array}{c} x_{1,\nu} \\ x_{2,\nu} \end{array} \right)^T \left( \begin{array}{cc} \Sigma_{11,\nu} & \Sigma_{12,\nu} \\ \Sigma_{21,\nu} & \Sigma_{22,\nu} + V I_\nu \end{array} \right)^{-1} \left( \begin{array}{c} x_{1,\nu} \\ x_{2,\nu} \end{array} \right) \right\}}{(2\pi)^{m/2} \left| \begin{array}{cc} \Sigma_{11,\nu} & \Sigma_{12,\nu} \\ \Sigma_{21,\nu} & \Sigma_{22,\nu} + V I_\nu \end{array} \right|^{1/2}}$$
Now,
\[
\begin{pmatrix} x_{1,\nu} \\ x_{2,\nu} \end{pmatrix}^T \begin{pmatrix} \Sigma_{11,\nu} & \Sigma_{12,\nu} \\ \Sigma_{21,\nu} & \Sigma_{22,\nu} + V I_{|\nu|} \end{pmatrix}^{-1} \begin{pmatrix} x_{1,\nu} \\ x_{2,\nu} \end{pmatrix} = x_{\nu}^T \begin{pmatrix} \Sigma_{11,\nu} & \Sigma_{12,\nu} \\ \Sigma_{21,\nu} & \Sigma_{22,\nu} + V I_{|\nu|} \end{pmatrix}^{-1} x_{\nu}
\]
\[
= x_{\nu}^T \left( (\Sigma_{11,\nu} \Sigma_{12,\nu}) + V \begin{pmatrix} 0 & 0 \\ 0 & I_{|\nu|} \end{pmatrix} \right)^{-1} x_{\nu}
\]
\[
= x_{\nu}^T \left( (P_{\nu}^T)^{-1} \Sigma P_{\nu}^{-1} + V \begin{pmatrix} 0 & 0 \\ 0 & I_{|\nu|} \end{pmatrix} \right)^{-1} x_{\nu}
\]
\[
= (P_{\nu}^T x_{\nu})^T \left( \Sigma + VP_{\nu} \begin{pmatrix} 0 & 0 \\ 0 & I_{|\nu|} \end{pmatrix} P_{\nu} \right)^{-1} (P_{\nu}^T x_{\nu})
\]
\[
= x_{\nu}^T (\Sigma + VB_{\nu})^{-1} x_{\nu}
\]

where $B_{\nu} = (\beta_{1,\nu}, \cdots, \beta_{m,\nu})$ and $\beta_{i,\nu} = \nu_i e_i, i = 1, \cdots, m$ and $e_i$ denotes a $m \times 1$ vector whose $i$-th element is 1 and rest are 0’s. That is, $B_{\nu} = \text{diag}(\nu_1, \cdots, \nu_m)$. Again from the preceding paragraph it immediately follows that
\[
\begin{pmatrix} \Sigma_{11,\nu} & \Sigma_{12,\nu} \\ \Sigma_{21,\nu} & \Sigma_{22,\nu} + V I_{|\nu|} \end{pmatrix} = P_{\nu} (\Sigma + VB_{\nu}) P_{\nu}^T
\]

which in turn implies that
\[
\left| \begin{pmatrix} \Sigma_{11,\nu} & \Sigma_{12,\nu} \\ \Sigma_{21,\nu} & \Sigma_{22,\nu} + V I_{|\nu|} \end{pmatrix} \right| = |\Sigma + VB_{\nu}|
\]

Hence we finally have
\[
f(x|\nu) = \frac{\exp\left\{-\frac{1}{2} x^T (\Sigma + VB_{\nu})^{-1} x \right\}}{(2\pi)^{m/2} |\Sigma + VB_{\nu}|^{1/2}}
\]

which means
\[
X = (X_1, \cdots, X_m)|\nu \sim N_m(0, \Sigma + VB_{\nu}).
\]

Hence the marginal distribution of $X$ is given by,
\[
f(x) = \sum_{\nu \in \{0, 1\}^m} \pi(\nu) f(x|\nu)
\]
\[
= \sum_{\nu \in \{0, 1\}^m} \pi(\nu) \frac{\exp\left\{-\frac{1}{2} x^T (\Sigma + VB_{\nu})^{-1} x \right\}}{(2\pi)^{m/2} |\Sigma + VB_{\nu}|^{1/2}}
\]

7.3 Proof of Lemma 3.3

Proof. First observe that since both $t > 1$ and $t_0 > 1$, one must have $j_1(x^* + r_0 g) \neq 1$ and $j_1(x^*) \neq 1$. Then using the observation made in Remark 3.3 we obtain,
\[
j_1(x^* + r_0 g) = \arg \max_{j \in \{2, \cdots, m\}} S_{1j}(x^* + r_0 g)
\]
\[
= \arg \max_{j \in \{2, \cdots, m\}} S_{1j}(x^*)
\]
\[
= j_1(x^*).
\] (7.7)

Now using Lemma 3.2 it follows that, for all $l = 1, \cdots, t_0 - 1$ with $j_l(x^* + r_0 g) \neq 1$, and for all $j \in \{1, \cdots, m\} \setminus \{j_1(x^* + r_0 g), \cdots, j_{l-1}(x^* + r_0 g)\}$,\n\[
U_{lj}^{(j_1(x^* + r_0 g), \cdots, j_{l-1}(x^* + r_0 g))}(x^* + r_0 g) = U_{lj}^{(j_1(x^* + r_0 g), \cdots, j_{l-1}(x^* + r_0 g))}(x^*),
\] (7.8)
whence we obtain for all \(l = 1, \cdots, t_0 - 1\) with \(j_l(x^* + r_0g) \neq 1\), and for all \(j \in \{1, \cdots, m\} \setminus \{j_1(x^* + r_0g), \cdots, j_{t_0-1}(x^* + r_0g)\}\),
\[
S_{lj}(j_1(x^* + r_0g), \cdots, j_{t_0-1}(x^* + r_0g))(x^* + r_0g) = S_{lj}(j_1(x^* + r_0g), \cdots, j_{t_0-1}(x^* + r_0g))(x^*).
\]
(7.9)

In particular, for all \(l < t_0\),
\[
S_{lj}(j_1(x^* + r_0g), \cdots, j_{t_0-1}(x^* + r_0g))(x^* + r_0g) = S_{lj}(j_1(x^* + r_0g), \cdots, j_{t_0-1}(x^* + r_0g))(x^*).
\]
(7.10)

Again using Lemma 3.2 we have for all \(l = 1, \cdots, t_0 - 1\) with \(j_l(x^* + r_0g) \neq 1\), and for all \(j \in \{1, \cdots, m\} \setminus \{j_1(x^* + r_0g), \cdots, j_{t_0-1}(x^* + r_0g)\}\) the following:
\[
U_{lj}(j_1(x^* + r_0g), \cdots, j_{t_0-1}(x^* + r_0g))(x^* + r_0g) = U_{lj}(j_1(x^* + r_0g), \cdots, j_{t_0-1}(x^* + r_0g))(x^*)
\]
\[\quad \quad + \quad \quad r_0, \quad \frac{\delta}{l} \frac{S_{lj}(j_1(x^* + r_0g), \cdots, j_{t_0-1}(x^* + r_0g))}{g(x^*)} \]
which means that only the values of \(S_{lj}(j_1(x^* + r_0g), \cdots, j_{t_0-1}(x^* + r_0g))(x^* + r_0g)\) can change for \(l = 1, \cdots, t_0 - 1\). Again since \(H_{01}\) is rejected at the \(t_0\)-th stage when \(x^* + r_0g\) is observed, for each \(l = 1, \cdots, t_0 - 1\), \(S_{lj}(j_1(x^* + r_0g), \cdots, j_{t_0-1}(x^* + r_0g))(x^* + r_0g)\) cannot be the maximum of the corresponding \(S_{lj}(j_1(x^* + r_0g), \cdots, j_{t_0-1}(x^* + r_0g))\)'s, since in that case \(H_{01}\) would have been rejected before the \(t_0\)-th step which would be a contradiction. Using this observation and equations (7.7), (7.9) and (7.10) it therefore follows that for any \(1 \leq l \leq t_0 - 1\),
\[
S_{lj}(j_1(x^* + r_0g), \cdots, j_{t_0-1}(x^* + r_0g))(x^* + r_0g) = \arg\max_{j \in \{2, \cdots, m\} \setminus \{j_1(x^* + r_0g), \cdots, j_{t_0-1}(x^* + r_0g)\}} S_{lj}(j_1(x^* + r_0g), \cdots, j_{t_0-1}(x^* + r_0g))(x^*)
\]
\[\quad \quad = \quad \quad \arg\max_{j \in \{2, \cdots, m\} \setminus \{j_1(x^*), \cdots, j_{t_0-1}(x^*)\}} S_{lj}(j_1(x^*), \cdots, j_{t_0-1}(x^*))(x^*)
\]
\[\quad \quad = \quad \quad j_l(x^*)
\]
This completes the proof of Lemma 3.3.

7.4 Proof of Lemma 3.4

Proof. First observe that when \(t_0 = 1\), since \(\phi_1(x^*) = 0\) and \(\phi_1(x^* + r_0g) = 1\), one cannot have \(t = 1\) due to (7.7). Therefore we must have \(t > 1\) when \(t_0 = 1\). Thus the result is true when \(t_0 = 1\). However, the proof for the case when both \(t > 1\) and \(t_0 > 1\), is non-trivial and requires a contrapositive argument and Lemma 3.3.

Since \(t > 1\), we have \(S_{lj}(j_1(x^*), \cdots, j_{t-1}(x^*))(x^*) > \delta\) for all \(l = 1, \cdots, t-1\), with \(j_l(x^*) \neq 1\) for each \(l\) and
\[
S_{lj}(j_1(x^*), \cdots, j_{t-1}(x^*))(x^*) \leq \delta \quad \Rightarrow \quad S_{lj}(j_1(x^*), \cdots, j_{t-1}(x^*))(x^*) \leq \delta
\]
for all \(j \in \{1, \cdots, m\} \setminus \{j_1(x^*), \cdots, j_{t-1}(x^*)\}\), with \(j_l(x^*) \neq 1\) for all \(l \in \{1, \cdots, t-1\}\).

On contrary, let us now assume that \(t_0 > t\). Then
\[
S_{lj}(j_1(x^* + r_0g), \cdots, j_{t-1}(x^* + r_0g))(x^* + r_0g) > \delta,
\]
otherwise the process would have stopped at stage \(t\) without rejecting \(H_{01}\) when \(x^* + r_0g\) is observed, which would be a contradiction.
and thereby we have

\[ S_{t_j}(x^*) = S_{t_j}(x^* + r_0 g) > \delta. \]

This means that when \( x^* \) is observed, the testing procedure cannot stop at stage \( t \) and consequently \( \phi_1(x^*) \neq 0 \), which is a contradiction. This completes the proof of Lemma 3.4.

### 7.5 Proof of Lemma 3.5

**Proof.** Let us first consider the situation when \( t_0 > 1 \).

Observe that when \( t_0 > 1 \) we have,

\[
S_{t_{10}}^{(j_1(x^*), \cdots, j_{10-1}(x^*))}(x^*) \leq S_{t_{10}}^{(j_1(x^*), \cdots, j_{10-1}(x^*))}(x^* + r_0 g)
\]

\[
= S_{t_{10}}^{(j_1(x^*), \cdots, j_{10-1}(x^*))}(x^* + r_0 g)
\]

\[
= S_{t_{10}}^{(j_1(x^* + r_0 g), \cdots, j_{10-1}(x^* + r_0 g))}(x^* + r_0 g)
\]

\[
\leq S_{t_{10}}^{(j_1(x^* + r_0 g), \cdots, j_{10-1}(x^* + r_0 g))}(x^* + r_0 g)
\]

\[
= S_{t_{10}}^{(j_1(x^*), \cdots, j_{10-1}(x^*))}(x^* + r_0 g)
\]

Since for given \( j_1, \cdots, j_{10-1}, S_{t_{10}}^{(j_1, \cdots, j_{10-1})} \) is a strictly increasing function of \( |U_{t_{10}}^{(j_1, \cdots, j_{10-1})}| \), it follows that

\[
|U_{t_{10}}^{(j_1(x^*), \cdots, j_{10-1}(x^*))}(x^*)| \leq |U_{t_{10}}^{(j_1(x^*), \cdots, j_{10-1}(x^*))}(x^* + r_0 g)|
\]

whence it follows from correction 3.1 that

\[
U_{t_{10}}^{(j_1(x^*), \cdots, j_{10-1}(x^*))}(x^* + r_0 g) > 0.
\]

But for given \( (j_1, \cdots, j_{10-1}) \), \( U_{t_{10}}^{(j_1, \cdots, j_{10-1})}(x^* + r_0 g) \) is strictly increasing in \( r \). Hence for all \( r > r_0 \), we have

\[
U_{t_{10}}^{(j_1(x^*), \cdots, j_{10-1}(x^*))}(x^* + r g) > U_{t_{10}}^{(j_1(x^*), \cdots, j_{10-1}(x^*))}(x^* + r_0 g) > 0
\]

We shall complete the proof now based on a contrapositive argument. Recall that, we need to show \( \phi_1(x^* + r g) = 1 \) for all \( r > r_0 \). On contrary, suppose this is not true. Then there exists some \( r_1 > r_0 \) such that \( \phi_1(x^* + r_1 g) = 0 \). Let \( t_1 \) denote the step at which the testing procedure must stop without rejecting \( H_{01} \) when \( x^* + r_1 g \) is observed. Then using Lemma 3.4 we have \( t_0 \leq t_1 \). Since \( t_0 > 1 \), using Lemma 3.3 it follows

\[
\bar{j}_1(x^* + r_1 g) = \bar{j}_1(x^* + r_0 g)
\]

\[
= \bar{j}_1(x^*).
\]

for all \( l = 1, \cdots, t_0 - 1 \). Again, replacing \( x^* \) by \( x^* + r_1 g \), and applying the preceding arguments, from (7.11) we obtain

\[
|U_{t_{10}}^{(j_1(x^*), \cdots, j_{10-1}(x^*))}(x^* + r_1 g)| \leq |U_{t_{10}}^{(j_1(x^*), \cdots, j_{10-1}(x^*))}(x^* + r_0 g)|
\]

23
Then for all \( r > r_0 \), therefore one must have \( \phi_1(x^* + rg) = 1 \) for all \( r > r_0 \), when \( t_0 > 1 \).

Next observe that when \( t_0 = 1 \), since \( S_{ij}(x^* + r_0g) = S_{ij}(x^*) \) for all \( j \in \{2, \cdots, m\} \), one must have \( S_{11}(x^*) < S_{11}(x^* + r_0g) \). Therefore, using exactly the same arguments as before we have,

\[
U_{11}(x^* + r_0g) > U_{11}(x^* + r_0g) > 0 \quad \text{for all} \quad r > r_0.
\]

Then for all \( r > r_0 \) we have,

\[
S_{11}(x^* + r_0g) > S_{11}(x^* + r_0g) \\
\geq S_{1j}(x^* + r_0g) \quad \text{for all} \quad j \in \{2, \cdots, m\} \quad \text{[since} \quad t_0 = 1]\]

\[
= S_{1j}(x^*) \quad \text{for all} \quad j \in \{2, \cdots, m\} \quad \text{[using Corollary 3.1]} \]

\[
= S_{1j}(x^* + rg) \quad \text{for all} \quad j \in \{2, \cdots, m\} \quad \text{[using Corollary 3.1]}
\]

which implies that every \( x^* + rg \) will be a point of rejection for \( H_0 \) for all \( r > r_0 \), that is, \( \phi_1(x^* + rg) = 1 \) for all \( r > r_0 \) when \( t_0 = 1 \). This completes the proof of Lemma 3.5.

\[\Box\]

### 7.6 Proof of Theorem 4.1

Proof. We first split \( E(\tilde{E}(\gamma)/p - 1)^2 \) into two parts as

\[
E\left(\frac{\tilde{E}(\gamma)}{p} - 1\right)^2 = Var\left(\frac{\tilde{E}(\gamma)}{p} - 1\right) + E\left(\frac{\tilde{E}(\gamma)}{p} - 1\right)^2, \tag{7.13}
\]

where

\[
Var\left(\frac{\tilde{E}(\gamma)}{p} - 1\right) = \frac{1}{m^2(1-\gamma)p^2}\left[\sum_{j=1}^{m} Var\left(\cos(\sqrt{2\gamma\log m}X_j)\right) + \sum_{j \neq j'} Cov\left(\cos(\sqrt{2\gamma\log m}X_j), \cos(\sqrt{2\gamma\log m}X_{j'})\right)\right] \tag{7.14}
\]

Now observe that, using (7.3), for each \( j = 1, \cdots, m \), we have,

\[
E\left[\cos(\sqrt{2\gamma\log m}X_j)\right] = \frac{1}{2}\left[E(e^{i\sqrt{2\gamma\log m}X_j}) + E(e^{-i\sqrt{2\gamma\log m}X_j})\right]
\]

\[
= E\left(e^{i\sqrt{2\gamma\log m}X_j}\right) \quad \text{[since} \quad X_j \overset{d}{=} -X_j]\]

\[
= (1-p)E\left(e^{i\sqrt{2\gamma\log m}X_j}|\nu_j = 0\right) + pE\left(e^{i\sqrt{2\gamma\log m}X_j}|\nu_j = 1\right)
\]

\[
= (1-p)e^{-\gamma\log m} + pe^{-(1+V)\gamma\log m} \tag{7.15}
\]

where \( i = \sqrt{-1} \), whence it follows that

\[
\frac{|E(\tilde{E}(\gamma)) - p\right|}{p} = \frac{1}{p}|1 - m\gamma|E\left[\cos(\sqrt{2\gamma\log m}X_j)\right] - p\right| \quad \text{[since} \quad X_j \overset{d}{=} X_j \text{for all} \ j]\]

\[
= \frac{1}{p}\left|1 - p - m^2(1-p) + pe^{-V\gamma\log m}\right| \tag{7.16}
\]

Clearly, from Assumption (A) we have,

\[
\lim_{m \to \infty} \frac{|E(\tilde{E}(\gamma)) - p\right|}{p} = 0.
\]

Let us now fix any \( 1 \leq j, j' \leq m \) with \( j \neq j' \). Then using Proposition 7.3 we obtain the following:

\[
E\left[\cos(\sqrt{2\gamma\log m}X_j)\cos(\sqrt{2\gamma\log m}X_{j'})\right] = E_1 + E_2 + E_3 + E_4 \tag{7.17}
\]
Hence by the condition (4.2) in the statement of Theorem 4.1, we have,

\[ E_1 = (1 - p)^2 E[ \cos(\sqrt{2\gamma \log mX_j}) \cos(\sqrt{2\gamma \log mX_{j'}}) | \nu_j = 0, \nu_{j'} = 0] \]

\[ = \frac{1}{2} (1 - p)^2 \left[ e^{-2(1+\sigma_{j,j'})\gamma \log m} + e^{-2(1-\sigma_{j,j'})\gamma \log m} \right] \]  

(7.18)

\[ E_2 = p(1 - p) E[ \cos(\sqrt{2\gamma \log mX_j}) \cos(\sqrt{2\gamma \log mX_{j'}}) | \nu_j = 1, \nu_{j'} = 0] \]

\[ = \frac{1}{2} p(1 - p) \left[ e^{-(2+2\sigma_{j,j'})\gamma \log m} + e^{-(2-2\sigma_{j,j'})\gamma \log m} \right] \]  

(7.19)

\[ E_3 = p(1 - p) E[ \cos(\sqrt{2\gamma \log mX_j}) \cos(\sqrt{2\gamma \log mX_{j'}}) | \nu_j = 0, \nu_{j'} = 1] \]

\[ = \frac{1}{2} p(1 - p) \left[ e^{-(2+2\sigma_{j,j'})\gamma \log m} + e^{-(2-2\sigma_{j,j'})\gamma \log m} \right] \]

\[ = E_2 \]  

(7.20)

where

and

\[ E_4 = p^2 E[ \cos(\sqrt{2\gamma \log mX_j}) \cos(\sqrt{2\gamma \log mX_{j'}}) | \nu_j = 1, \nu_{j'} = 1] \]

\[ = \frac{1}{2} p^2 \left[ e^{-(2+2\sigma_{j,j'})\gamma \log m} + e^{-(2-2\sigma_{j,j'})\gamma \log m} \right] \]

\[ \times \left[ (1 - p)e^{-\gamma \log m} + pe^{-(1+V)\gamma \log m} \right]^2. \]

(7.21)

Combining (7.17)–(7.21) and doing some simple algebraic manipulations thereafter, we obtain,

\[ E[ \cos(\sqrt{2\gamma \log mX_j}) \cos(\sqrt{2\gamma \log mX_{j'}})] = \frac{1}{2} \left[ e^{-(2+2\sigma_{j,j'})\gamma \log m} + e^{-(2-2\sigma_{j,j'})\gamma \log m} \right] \times \left[ (1 - p)e^{-\gamma \log m} + pe^{-(1+V)\gamma \log m} \right]^2. \]

Therefore for each pair of indices \( 1 \leq j, j' \leq m \) with \( j \neq j' \), we have,

\[ Cov(\cos(\sqrt{2\gamma \log mX_j}), \cos(\sqrt{2\gamma \log mX_{j'}})) \]

\[ = (m^{-2\gamma} / 2) \left( \frac{1}{m^{\gamma \sigma_{j,j'}}} - \frac{1}{m^{-\gamma \sigma_{j,j'}}} \right)^2 \left[ (1 - p) + pe^{-V \gamma \log m} \right]^2. \]

Hence by the condition (4.2) in the statement of Theorem 4.1, we have,

\[ \lim_{m \to \infty} \frac{1}{m^{2-2\gamma}p^2} \sum_{1 \leq j, j' \leq m, j \neq j'} Cov(\cos(\sqrt{2\gamma \log mX_j}), \cos(\sqrt{2\gamma \log mX_{j'}})) \]

\[ \leq \lim_{m \to \infty} \frac{1}{m^{2-2\gamma}p^2} \sum_{1 \leq j, j' \leq m, j \neq j'} \left( \frac{1}{m^{\gamma \sigma_{j,j'}}} - \frac{1}{m^{-\gamma \sigma_{j,j'}}} \right)^2 \]

\[ = 0. \]

(7.22)

Therefore, from (7.14) and (7.22) we obtain,

\[ \limsup_{m \to \infty} Var \left( \frac{\hat{P}(\gamma)}{p} - 1 \right) \leq \limsup_{m \to \infty} \frac{1}{m^{2(1-\gamma)p^2}} \sum_{j=1}^{m} Var(\cos(\sqrt{2\gamma \log mX_j})) \]

\[ \leq \limsup_{m \to \infty} \frac{1}{m^{2(1-\gamma)p^2}} \sum_{j=1}^{m} E(\cos^2(\sqrt{2\gamma \log mX_j})) \]

\[ \leq \limsup_{m \to \infty} \frac{1}{m^{1-2\gamma}p^2} \sum_{j=1}^{m} \left[ 1 \right] \]

\[ = \limsup_{m \to \infty} \frac{1}{m^{1-2\gamma}p^2} \]

\[ = 0 \]
whence it follows that
\[
\lim_{m \to \infty} Var \left( \frac{\hat{\beta}(\gamma)}{p} - 1 \right) = 0. \tag{7.23}
\]
Rest of the proof follows immediately by combining equations (7.13), (7.16) and (7.28) together. ■

7.7 Proof of Theorem 4.2

Proof. To prove Theorem 4.2, it will be enough to show
\[
\frac{\hat{\beta} \cdot \hat{V}}{p \cdot V} - 1 = \frac{1}{m} \sum_{i=1}^{m} X_i^2 - (1 + pV) \xrightarrow{L_2} 0 \text{ as } m \to \infty. \tag{7.24}
\]
For that we first observe that for each \( j = 1, \ldots, m \), \( E(X_i^2) = 1 + pV \) and hence \( E(\frac{\hat{\beta} \cdot \hat{V}}{p \cdot V} - 1) = 0 \) for all \( m \), so that
\[
E \left( \frac{\hat{\beta} \cdot \hat{V}}{p \cdot V} - 1 \right)^2 = Var \left( \frac{\hat{\beta} \cdot \hat{V}}{p \cdot V} \right)
= \frac{Var(X_i^2)}{mp^2V^2} + \frac{1}{m^2p^2V^2} \sum_{1 \leq i,j \leq m, i \neq j} cov(X_i^2, X_j^2)
= \frac{3(1 - p) + 3p(1 + V)^2}{mp^2V^2} + \frac{1}{m^2p^2V^2} \sum_{1 \leq i,j \leq m, i \neq j} cov(X_i^2, X_j^2)
\]

Next we observe the following:
\[
cov(X_i^2, X_j^2) = 2\sigma_{ij}^2 \text{ for all } i \neq j, \tag{7.25}
\]
under the mixture model (2.33). To prove (7.25), let us fix any \( i, j \leq m \) with \( i \neq j \). Observe that under (2.33), the joint distribution of \((X_i, X_j)^T\) is a four component mixture of bivariate normal distributions given by,
\[
\begin{align*}
f(x_i, x_j) &= (1 - p)^2f(x_i, x_j | \nu_i = 0, \nu_j = 0) + (1 - p)pf(x_i, x_j | \nu_i = 1, \nu_j = 0) \\
&\quad + p(1 - p)f(x_i, x_j | \nu_i = 0, \nu_j = 1) + (1 - p)^2f(x_i, x_j | \nu_i = 1, \nu_j = 1)
\end{align*}
\]
where
\[
\begin{pmatrix} X_i \\ X_j \end{pmatrix} \mid \nu_i = r, \nu_j = s \sim N_2 \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 + rV & \sigma_{ij} \\ \sigma_{ij} & 1 + sV \end{pmatrix} \right) \tag{7.27}
\]
for \( (r, s) \in \{0, 1\} \times \{0, 1\} \). Now applying the law of iterated variance and then using (7.27), we obtain,
\[
E(X_i^2 X_j^2) = \sum_{(r, s) \in \{0, 1\} \times \{0, 1\}} p^r(1 - p)^s E(X_i^2 X_j^2 | \nu_i = r, \nu_j = s)
= \sum_{(r, s) \in \{0, 1\} \times \{0, 1\}} p^r(1 - p)^s E(X_i^2 | X_j, \nu_i = r, \nu_j = s | \nu_i = r, \nu_j = s)
= (1 - p)^2 \left[ 1 + 2\sigma_{ij}^2 \right] + 2p(1 - p) \left[ (1 + V) + 2\sigma_{ij}^2 \right] + p^2 \left[ (1 + V)^2 + 2\sigma_{ij}^2 \right]
= 2\sigma_{ij}^2 \left[ (1 - p)^2 + 2p(1 - p) + p^2 \right] + \left[ (1 - p)^2 + 2p(1 - p)(1 + V) + p^2(1 + V)^2 \right]
= 2\sigma_{ij}^2 + (1 + pV)^2
\]
whence it follows that
\[
cov(X_i^2, X_j^2) = 2\sigma_{ij}^2 \text{ [since } E(X_i^2) = 1 + pV \text{ for all } i].
\]
Then by the given condition of Theorem 4.1, we have,
\[
\lim_{m \to \infty} \left[ \frac{1}{m^2p^2V^2} \sum_{1 \leq i,j \leq m, i \neq j} cov(X_i^2, X_j^2) \right] = 0. \tag{7.28}
\]
Rest of the proof now follows trivially. ■
7.8 Proof of Lemma 5.1

Proof. Let us write $B_{1,\nu} = \Sigma + VB_{\nu,i=1}$ and $B_{0,\nu} = \Sigma + VB_{\nu,i=0}$. Then
\[
B_{1,\nu} = (\Sigma + VB_{\nu,i=0}) + Vdiag(e_i) = B_{0,\nu} + Vdiag(e_i) = B_{0,\nu} + A\nu^{-1}A^T
\]
where $A = \sqrt{Vdiag(e_i)} = A^T$. Using Proposition 7.2 it follows that,
\[
B_{1,\nu}^{-1} = (B_{0,\nu} + A\nu^{-1}A^T)^{-1} = B_{0,\nu}^{-1} - B_{0,\nu}^{-1}A(A^TB_{0,\nu}^{-1}A + I)^{-1}A^TB_{0,\nu}^{-1}
\]
Let $B_{0,\nu}^{-1} = ((b_{ij}))_{m \times m}$. Since $B_{0,\nu}$ is positive definite, $B_{0,\nu}^{-1}$ is positive definite. So, $b_{ii} > 0$, for all $i = 1, \cdots, m$.

Now observe that,
\[
A^TB_{0,\nu}^{-1}A = V \begin{pmatrix}
0 & 0 & \cdots & 0 \\
\vdots & & & \\
0 & 0 & \cdots & 0 \\
\vdots & & & \\
0 & 0 & \cdots & 0
\end{pmatrix} b_{ii} b_{2i} \cdots b_{ni} (0, \cdots, 0, e_i, 0, \cdots, 0) = Vdiag(0, \cdots, 0, b_{ii}, 0, \cdots, 0) = Vb_{ii}diag(e_i)
\]
Therefore,
\[
I + A^TB_{0,\nu}^{-1}A = diag(1, \cdots, 1, 1 + Vb_{ii}, 1, \cdots, 1).
\]
which implies that
\[
(I + A^TB_{0,\nu}^{-1}A)^{-1} = diag(1, \cdots, 1, 1 + Vb_{ii}, 1, \cdots, 1)
\]
Noting that $A = A^T$ and using the preceding fact, it follows that,
\[
A(I + A^TB_{0,\nu}^{-1}A)^{-1}A^T = A^T(1 + Vb_{ii}diag(e_i)) = \frac{1}{1 + Vb_{ii}}AA^T.
\]
Therefore we have
\[
B_{0,\nu}^{-1}A(I + A^TB_{0,\nu}^{-1}A)^{-1}A^TB_{0,\nu}^{-1} = \frac{1}{1 + Vb_{ii}}B_{0,\nu}^{-1}AA^TB_{0,\nu}^{-1}
\]
\[
= \frac{V}{1 + Vb_{ii}}B_{0,\nu}^{-1}diag(e_i)diag(e_i)^T B_{0,\nu}^{-1}
\]
Now observe that
\[
B_{0,\nu}^{-1}diag(e_i) = (0, \cdots, 0, b_i, 0, \cdots, 0)
\]
whence it follows that
\[
B_{0,\nu}^{-1}A(I + A^TB_{0,\nu}^{-1}A)^{-1}A^TB_{0,\nu}^{-1} = \frac{V}{1 + Vb_{ii}}b_i b_i^T
\]
Thus for any $\nu \in \{0,1\}^m$ we have the following identity:
\[
(\Sigma + VB_{\nu,i=1})^{-1} = (\Sigma + VB_{\nu,i=0})^{-1} - \frac{V}{1 + Vb_{ii}}b_i b_i^T.
\]
7.9 Proof of Lemma 5.2

Proof. Let us write \( B_{1,\nu} = \Sigma + VB_{\nu,\nu=1} \) and \( B_{0,\nu} = \Sigma + VB_{\nu,\nu=0} \). Then

\[
\frac{\Sigma + VB_{\nu,\nu=1}}{\Sigma + VB_{\nu,\nu=0}} = \frac{B_{1,\nu}}{B_{0,\nu}} = B_{1,\nu}B_{0,\nu}^{-1}
\]

Observe that, \( B_{1,\nu} = B_{0,\nu} + V \text{diag}(\epsilon_i) \). Then using the facts derived in the proof of Lemma 5.1 it follows that

\[
B_{1,\nu}B_{0,\nu}^{-1} = I + VB_{0,\nu}^{-1}\text{diag}(\epsilon_i)
\]

\[
= I + V \text{diag}(0, \cdots, 0, b_i(\nu), 0, \cdots, 0)
\]

\[
\implies |B_{1,\nu}B_{0,\nu}^{-1}| = 1 + Vb_i(\nu).
\]

7.10 Proof of Lemma 5.3

Proof. We can write \( |\Sigma + VB_{\nu,\nu=0}| \) and \( |\Sigma + VB_{\nu,\nu=1}| \) as follows:

\[
|\Sigma + VB_{\nu,\nu=0}| = |(\Sigma + VB_{\nu,\nu=0})(-i,-i)| |\sigma_{ii} - \sigma^{T}_{(-i)}((\Sigma + VB_{\nu,\nu=0})(-i,-i))^{-1}\sigma_{(-i)}| \text{ and}
\]

\[
|\Sigma + VB_{\nu,\nu=1}| = |(\Sigma + VB_{\nu,\nu=1})(-i,-i)| |\sigma_{ii} + V - \sigma^{T}_{(-i)}((\Sigma + VB_{\nu,\nu=1})(-i,-i))^{-1}\sigma_{(-i)}|
\]

where \( A_{(-i,-i)} \) denotes the \((i,i)\)-th cofactor of a matrix \( A \) and \( \sigma_{(-i)} \) is the covariance vector between \( X_i \) and rest of the \( X_j \)'s.

Next observe that

\[
(\Sigma + VB_{\nu,\nu=0})(-i,-i) = (\Sigma + VB_{\nu,\nu=1})(-i,-i).
\]

Therefore,

\[
\frac{|\Sigma + VB_{\nu,\nu=1}|}{|\Sigma + VB_{\nu,\nu=0}|} = \frac{\sigma_{ii} + V - \sigma^{T}_{(-i)}((\Sigma + VB_{\nu,\nu=0})(-i,-i))^{-1}\sigma_{(-i)}}{\sigma_{ii} - \sigma^{T}_{(-i)}((\Sigma + VB_{\nu,\nu=0})(-i,-i))^{-1}\sigma_{(-i)}}
\]

(7.29)

The using equation (3) and Lemma (5.2) it follows that

\[
b_i(\nu) = \left| \sigma_{ii} - \sigma^{T}_{(-i)}((\Sigma + VB_{\nu,\nu=0})(-i,-i))^{-1}\sigma_{(-i)} \right|^{-1}
\]

Proof of the remaining part now follows trivially.

8 Discussion

We consider in this article the problem of simultaneous testing of the individual components of a multivariate normal mean vector when the underlying covariance matrix is assumed to be known. We propose a stepwise Bayesian testing procedure assuming a two component point mass mixture prior over the unknown means. The proposed Bayesian stepdown procedure is generic and can also be applied for non-normal models also. A decision theoretic justification for the newly developed testing procedure is established by showing that it possesses a certain desirable convexity property essential for the admissibility of a multiple testing procedure. Consistent estimation of the proportion of true alternatives and the variance of the non-zero means is established under certain weak correlation structures. An alternative representation of the proposed test statistics has also
been established that makes the computation faster. We hope that the present Bayesian stepwise procedure can be very useful in many practical situations.

For the present multiple testing problem, one can also use the optimal Bayes rule assuming an additive loss function and report the posterior inclusion probabilities $\pi(\theta_i = 1|X)$, $i = 1, \cdots, m$. However, as already mentioned, evaluating the posterior inclusion probabilities can often be computationally very demanding, specially when one wish to numerically evaluate the optimal Bayes risk by replicating the experiment a large number of times, say, 5000 times, when the number of hypotheses $m$ is large. It should be stressed at this point that finding an analytic expression of the optimal Bayes risk, at least asymptotically, is indeed a very hard problem to solve, even under specific forms of dependence. In a recent work [Bogdan et al (2011)] show that when $X_i$’s are i.i.d. observations from a two component Gaussian mixture distribution, the popular BH method and the Bonferroni procedure become asymptotic Bayes optimal under sparsity (ABOS). They also find conditions under which a multiple testing procedure will become ABOS when test statistics are independent. A natural question to ask then is under what conditions a multiple testing procedure, such as the BH method or the BSD method, will become ABOS for the present multivariate normal mean problem. This is definitely one of the most important and challenging open problems in this domain so far. Another very important and interesting problem is to investigate more general conditions under which consistent estimation of the proportion of non-nulls is possible. It would be of immense theoretical importance to investigate whether such an estimator attains any optimal rate of convergence, such as, the minimax rate of convergence, for the present multiple testing problem. We hope to address these problems elsewhere in future.

References

Benjamini, Y. and Hochberg, Y. (1995). “Controlling the false discovery rate: a practical and powerful approach to multiple testing.” Journal of the Royal Statistical Society, Series B, Methodological, 57(1): 289-300.

Benjamini, Y. and Liu, W. (1999). “A step-down multiple hypotheses that controls the false discovery rate under independence.” Journal of Statistical Planning and Inference, 82: 163-170.

Benjamini, Y. and Yekutieli, D. (2001). “The control of the false discovery rate in multiple testing under dependency.” The Annals of Statistics, 29(4): 1165-1188.

Brown D. A., Lazar, N. A. and Datta, G. S. (2011). “Bayesian Multiple Testing Under Dependence with Application to Magnetic Resonance Imaging.” Section on Bayesian Statistical Science JSM 2011.

Bogdan, M., Chakrabarti, A., Frommlet, F., and Ghosh, J. K. (2011). “Asymptotic Bayes-Optimality under Sparsity of Some Multiple Testing Procedures.” The Annals of Statistics, 39(3): 1551–1579.

Bogdan, M., Ghosh, J. K., and Tokdar, S. T. (2008). “A comparison of the Benjamini-Hochberg procedure with some Bayesian rules for multiple testing.” In Balakrishnan, N., Pena, E., and Silvapulle, M. J. (eds.), Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen, IMS Collections, volume 1, 211–230. IMS, Beachwood, OH.

Cai, T. T., Jin, J. and Low, M. G. (2007). “Estimation and Confidence Sets for Sparse Normal Mixtures.” The Annals of Statistics, 35 (6): 2421-2449.

Cai, T. T. and Jin, J. (2010). “Optimal Rates of Convergence for Estimating the Null Density and Proportion of Nonnull Effects in Large-Scale Multiple Comparisons.” The Annals of Statistics, 38 (1): 100-145.
Chen, C., Cohen, A. and Sackrowitz, H. B. (2009). “Admissible, consistent multiple testing with applications including variable selection.” The Electronic Journal of Statistics, 3: 633-650.

Chen, J. and Sarkar, S. K. (2004). “Multiple testing of response rates with a control: a Bayesian stepwise approach.” Journal of Statistical Planning and Inference, 125 (1-2): 3-16.

Cohen, A., Kolassa, J. and Sackrowitz, H. B. (2007). “A smooth version of the step-up procedure for multiple tests of hypotheses.” Journal of Statistical Planning and Inference, 137 (11): 3352-3360.

Cohen, A. and Sackrowitz, H. B. (2005). “Multiple testing of two-sided alternatives with dependent data.” The Annals of Statistics, 33 (1): 145-158.

Cohen, A. and Sackrowitz, H. B. (2007). “More on the inadmissibility of step-up.” Journal of Multivariate Analysis, 98 (3): 481-492.

Cohen, A. and Sackrowitz, H. B. (2008). “Multiple testing of two-sided alternatives with dependent data.” Statistica Sinica, 18 (4): 1593-1602.

Cohen, A., Sackrowitz, H. B. and Xu, M. (2009). “A new multiple testing method in the dependent case.” The Annals of Statistics, 37 (3): 1518-1544.

Efron, B., Tibshirani, R., Storey, J. D. and Tusher, V. (2001). “Empirical Bayes Analysis of a Microarray Experiment.” Journal of the American Statistical Association, 96 (456): 1151-1160.

Efron, B. (2004). “Large-scale simultaneous hypothesis testing: The choice of a null hypothesis.” Journal of the American Statistical Association, 99 (456): 96-104.

Efron, B. (2007). “Correlation and Large-Scale Simultaneous Significance Testing.” Journal of the American Statistical Association, 102 (477): 93-103.

Efron, B. (2008). “Microarrays, Empirical Bayes and the two-groups Model.” Statistical Science, 23(1): 1-22.

Fan, J., Han, X. and Gu, W. (2012). “Estimating false discovery proportion under arbitrary covariance dependence.” Journal of the American Statistical Association, 107 (499): 1019-1035.

Friguet, C., Kloareg, M. and Causeur, D. (2009). “A factor model approach to multiple testing under dependence.” Journal of the American Statistical Association, 104(488): 1406-1415.

Genovese, C., and Wasserman, L. (2004). “A Stochastic Process Approach to False Discovery Control” The Annals of Statistics, 32: 1035–1061.

Gopalan, R. and Berry, D. A. (1998). “Bayesian multiple comparisons using Dirichlet process priors.” Journal of the American Statistical Association, 93 (443): 1130-1139.

Guo, W., He, L. and Sarkar, S. K. (2014). “Further results on controlling the false discovery proportion.” The Annals of Statistics, 42 (3): 1070-1101.

Guo, W. and Sarkar, S. K. (2010). “Procedures controlling the k-FDR using bivariate distributions of the null p-values.” Statistica Sinica, 20 (3): 1227-1238.

Leek, J.T. and Storey, J. D. (2008). “A general framework for multiple testing dependence.” Proceedings of the National Academy of Sciences, 105 (48): 18718-18723.

Jin, J. (2006). “Proportion of Nonzero Normal Means: Universal Oracle Equivalences and Uniformly Consistent Estimations.” Technical report, Purdue University, Department of Statistics.

Jin, J. (2008). “Proportion of non-zero normal means: universal oracle equivalences and uniformly consistent estimators.” Journal of Royal Statistical Society Series B, 70 (3): 461-493.
Jin, J. and Cai, T. T. (2007). “Estimating the Null and the Proportion of Nonnull Effects in Large-Scale Multiple Comparisons.” Journal of the American Statistical Association, 102 (478): 495-506.

Meinshausen, N. and Rice, J. (2006). “Estimating the Proportion of False Null Hypotheses Among A Large Number Of Independently Tested Hypotheses.” The Annals of Statistics, 34 (1): 373-393.

Matthes, T. K. and Truax, D. R. (1967). “Tests of composite hypotheses for the multivariate exponential family.” The Annals of Mathematical Statistics, 38: 681-697.

Romano, J. P. and Shaikh, A. M. and Wolf, M. (2008). “Control of the false discovery rate under dependence using the bootstrap and subsampling.” TEST, 17(3): 417-442.

Sarkar, S. K. (2002). “Some results on false discovery rate in stepwise multiple testing procedures.” The Annals of Statistics, 30(1): 239-257.

Sarkar, S. K and Guo, W. (2009). “On a generalized false discovery rate.” The Annals of Statistics, 37 (3): 1545-1565.

Santner, T. J., Williams, B. J. and Notz, W. I. (2003). “The Design and Analysis of Computer Experiments.” Springer Series in Statistics, Appendix B: 211-214.

Scott, J. G. and Berger, J. O. (2006). “An exploration of aspects of Bayesian multiple testing.” Journal of Statistical Planning and Inference, 136 (7): 2144-2162.

Scott, J. G. and Berger, J. O. (2010). “Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem.” The Annals of Statistics, 38 (5): 2587-2619.

Storey, J. D. (2002). “A direct approach to false discovery rates.” Journal of Royal Statistical Society Series B, 64 (3): 479-498.

Storey, J. D., Taylor, J. E. and Siegmund, D. (2004). “Strong Control, Conservative Point Estimation and Simultaneous Conservative Consistency of False Discovery Rates: A Unified Approach.” Journal of Royal Statistical Society Series B, 66 (1): 187-205.

Sun, W. and Cai, T. T. (2009). “Large-scale multiple testing under dependence.” Journal of the Royal Statistical Society. Series B. Statistical Methodology, 71(2): 393-424.

Swannepoe, J. W. H. (1999). “The Limiting Behavior of a Modified Maximal Symmetric 2s-Spacing With Applications.” The Annals of Statistics, 27: 24-35.

Waller, R. A. and Duncan, D. B. (1969). “A Bayes rule for the symmetric multiple comparisons problem (I).” Journal of the American Statistical Association, 64: 1484-1503.

Waller, R. A. and Duncan, D. B. (1974). “A Bayes rule for the symmetric multiple comparisons problem (II).” Annals of the Institute of Statistical Mathematics, 26: 247–264.

Xie, J., Cai, T. T., Maris, J. and Li, H. (2011). “Optimal false discovery rate control for dependent data.” Statistics and its Interface, 4: 417-430.