MULTIPLE TESTING PROCEDURES FOR ONE- AND TWO-WAY CLASSIFIED HYPOTHESES

A Dissertation
Submitted to
the Temple University Graduate Board

---

in Partial Fulfillment
of the Requirements for the Degree of
DOCTOR OF PHILOSOPHY

---

by

Shinjini Nandi
August, 2019

Examining Committee Members:

Dr. Sanat K. Sarkar, Advisory Chair, Statistical Science
Dr. Edoardo M. Airoldi, Statistical Science
Dr. Zhigen Zhao, Statistical Science
Dr. Weijie Su, External Member, University of Pennsylvania

ABSTRACT

MULTIPLE TESTING PROCEDURES FOR ONE- AND TWO-WAY CLASSIFIED HYPOTHESES

by

Shinjini Nandi

DOCTOR OF PHILOSOPHY

Temple University, August, 2019

Prof. Sanat K. Sarkar, Dissertation Advisor

Multiple testing literature contains ample research on controlling false discoveries for hypotheses classified according to one criterion, which we refer to as 'one-way classified hypotheses'. However, one often encounters the scenario of 'two-way classified hypotheses' where hypotheses can be partitioned into two sets of groups via two different criteria. Associated multiple testing procedures that incorporate such structural information are potentially more effective than their one-way classified or non-classified counterparts. To the best of our knowledge, very little research has been pursued in this direction. This dissertation proposes two types of multiple testing procedures for two-way classified hypotheses. In the first part, we propose a general methodology for controlling the false discovery rate (FDR) using the Benjamini-Hochberg (BH) procedure based on weighted p-values. The weights can be appropriately chosen to reflect one- or two-way classified structure of hypotheses, producing novel multiple testing procedures for two-way classified hypotheses. Newer results for one-way classified hypotheses have been obtained in this process. Our proposed procedures control the false discovery rate (FDR) non-asymptotically in their oracle forms under positive regression dependence on subset of null p-values (PRDS) and in their data-adaptive forms for independent p-values. Simulation studies demonstrate that our proposed procedures can be considerably more powerful than some contemporary methods in many instances and that our data-adaptive procedures can non-asymptotically control the FDR under certain dependent scenarios. The proposed two-way adaptive procedure is applied to a data set from microbial abundance study, for which it makes more discoveries than an existing method. In the second part, we propose a Local false discovery rate (Lfdr) based multiple testing procedure for two-way classified hypotheses. The procedure has been developed in its oracle form under a model based framework that isolates the effects due to two-way grouping from the significance of an individual hypothesis. Simulation studies show that our proposed procedure successfully controls the average proportion of false discoveries, and is more powerful than existing methods.

ACKNOWLEDGEMENTS

My sincere thanks to Dr. Sanat Sarkar, my dissertation advisor, for his continuous guidance and support throughout my time as a student at Temple. I have been most fortunate to have him as my mentor, who cared so much about my overall growth as a researcher. I have been inspired by his wisdom, motivation and immense patience and will continue to follow his example throughout my professional career.

To Dr. Zhigen Zhao, for being a very supportive member of my dissertation committee, for his valuable comments and insight that helped me shape this dissertation and his invaluable advice that helped me make career choices.

To Dr. Edo Airoldi, for enthusiastically agreeing to be a member of my dissertation committee, for his valuable comments regarding my work and making research more interesting, if possible, within a very short period of time.

To Dr. Subhadeep Mukhopadhyay, for his support at the beginning of my career, and helping me write my first research article.

To Prince, Zeda, Scott, Qi, Nooreen, Gunes and all my fellow graduate students, past and present, at Temple, without whom, life would be extremely dull and lonely. To Uncle Tarun, and dear friend Rumi, for helping me settle down in the USA, for being home away from home. To Madhusree, my good old friend, for always being there to listen, support and encourage me during difficult times.

To my parents, who always encouraged me to follow my instincts and who taught me never to yield to difficulties that come in way. And finally, to my sister Shaoni, who believed in my success, long before I dreamed of it.

"…The Soul of the World is nourished by people's happiness. And also by unhappiness, envy and jealousy. To realize one's Personal Legend is a person's only real obligation. All things are one.

"And, when you want something, all the universe conspires in helping you achieve it."

From *The Alchemist*, by Paulo Coelho

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table

CHAPTER 1

INTRODUCTION

Large-scale multiple testing problems often involve classifying a set of hypotheses into several groups. In some cases, the families/groups might be formed naturally due to characteristics of the underlying scientific experiment. In other situations, a certain feature attributable to each hypothesis might serve as the basis of partition. Grouping of hypotheses due to any well defined argument, whether natural or artificial, benefit analyses. In all such instances, since the classification is due to a single factor, we refer to the setup as 'one-way classification of hypotheses'. Multiple testing procedures adapted to such arrangement of hypotheses incorporate the information of similar characteristics within each group. Consequently, they can address statistical issues specific to such structures, potentially yielding more power and better control over false discoveries than their counterparts that ignore the underlying group structures. One-way classified hypotheses have been widely investigated and several multiple testing procedures have been suggested in the literature to address varied research problems.

In many situations, a set of hypotheses might be classified according to more than one norm of classification. Just like one-way classified hypotheses, multiple interesting features or nature of the experiment may determine the norms. For example, brain imaging studies involving fMRI data (*Foygel Barber and Ramdas* (2015)), geographical studies involving data collected through satellite remote sensing (*Clements et al.* (2014)), studies in genetics involving microarray time course experiments (*Sun and Wei* (2011)), etc, comprise of spatio-temporal data. The multitude of hypotheses arising out of such data can be clustered into groups formed through aggregation of neighboring spatial units, and/or related time points. Other examples can be found in bioinformatics, studies involving association between genes and proteins, and genomewide association studies that involve analysis of association of SNPs with different regions of the brain (*Stein et al.* (2010)). Examples where more than two types of classification are imposed simultaneously on a set of hypotheses are very rare. If a set of hypotheses is classified in exactly two different ways, we call it a set of 'two-way classified hypotheses'.

In such cases, researchers are most interested in the hypotheses that emerge as significant when effects due to both classifications are factored in. The scope of existing multiple testing procedures is limited to one-way classified data and such methods are incapable to gauge the simultaneous effect of two-way classification. Some efforts made to study such structures in *Stein et al.* (2010), etc. involve repetitive application of one-way classification multiple testing procedures. Broadly speaking, in the first step, one of the two classifications is prioritized over the other. Considering the hypotheses as classified only due to this factor, significant groups and/or individual hypotheses are determined. In the second step, these significant elements are further tested for significance due to the second grouping criterion and finally the set of significant hypotheses is determined. *Foygel Barber and Ramdas* (2015), *Ramdas et al.* (2017) discuss multi-way classification and suggest an algorithm that recursively applies BH procedure to all partitions created and selects the set of hypotheses as significant which are rejected in all partitions.

The goal of this dissertation is to introduce a proper framework for two-way classified hypotheses and broaden the scope of multiple testing procedures to such structures. Two different types of multiple testing procedures are considered in this dissertation. One is designed to control the False Discovery Rate (FDR) as defined by *Benjamini and Hochberg* (1995), while the other is designed to control a Bayesian measure of false discoveries that is defined in terms of local FDR as in *Efron et al.* (2001).

The second chapter discusses the two-way classification structure and a generalized form of the weighted BH procedure that has been used to propose newer BH type FDR controlling procedures in their oracle forms for testing multiple hypotheses in such a setup, under the Positive Regression Dependency on Subset (PRDS) for the underlying test statistics or p-values. Also suggested in this chapter is a general method of estimating the weights appearing in the weighted BH procedure, producing a data-driven version of the above oracle procedure that is adaptable to one- or two-way classified hypotheses for certain choice of weights. Through a series of simulations we assert that our proposed methods have better control on the FDR and is more powerful than existing methods in various scenarios. We illustrate the utility of our procedure through an application on a microbiome dataset to analyze the abundance of microbes in different environments.

The third chapter is devoted to investigation of Local FDR based multiple testing methods applicable to two-way classified hypotheses. When hypotheses are sorted into groups due to one or more classification criteria, the state of each individual hypothesis is characterized by its parent groups and its own intrinsic properties. Several ideas of grouped hypotheses testing procedures in the literature are devoted to mine such properties of hypotheses and refine the effectiveness of testing procedures. We suggest extending the idea of multiple testing procedures for one way grouped hypotheses proposed in *Liu et al.* (2016) and *Sarkar and Zhao* (2017) to two-way classified hypotheses. Through utilization of a model that explicitly accounts for the states of the parent groups, our proposed procedure is able to measure the concurrent effects of the two classifications on each individual hypothesis and control a desired overall rate of false discoveries. We draw our conclusion in Chapter 4 with discussion on some possible future work.

CHAPTER 2

THE WEIGHTED BH PROCEDURE AND ITS DATA-ADAPTIVE VERSIONS TO TEST ONE- AND TWO-WAY CLASSIFIED HYPOTHESES

2.1 BH method adapted to One- and Two-way classified Hypotheses

2.1.1 Literature Review

In their seminal research work, *Benjamini and Hochberg* (1995) introduced the notion of False Discovery Rate (FDR) as an overall measure of type I errors (or false positives) in simultaneous testing of multiple hypotheses. For a set of N hypotheses, it is defined as

$$\text{FDR} = \text{E}\left[\frac{V}{R_N} I(R_N > 0)\right],$$

V being the number of false rejections and R_N being the total number of rejections. *Benjamini and Hochberg* (1995) also introduced a procedure, referred to as the BH procedure, which, in its simplest form is a step-up procedure, which when applied to such a set of hypotheses at level α rejects hypotheses corresponding to ordered p-values $P_{(1)}, \ldots, P_{(R)}$ where

$$R = \max\{1 \leq j \leq N : P_{(j)} \leq \frac{j\alpha}{N}\}.$$

If the hypotheses are independent, FDR is controlled at level $\pi_0\alpha$ where π_0 is the proportion of true nulls in the set. For positively dependent p-values satisfying the condition of Positive Regression Dependence on the subset of true nulls (PRDS), as defined later in Condition 1, the procedure conservatively controls FDR at the same level (*Benjamini and Yekutieli* (2001a), *Sarkar* (2002), *Sarkar* (2008)).

The conservative property of the BH procedure had encouraged researchers to correct it by suggesting adaptive versions of the method, which essentially involves using weighted p-values in the step-up procedure. Assuming that the proportion of true nulls is known, *Benjamini and Hochberg* (1997) suggested a generalized weighted FDR and corresponding weighted BH procedure. They showed that if the significant hypotheses are assigned zero weights, and the insignificant hypotheses assigned weights equal to the proportion of true nulls, the procedure controls FDR at level α for a set of independent hypotheses. *Genovese et al.* (2006) introduced the idea of using external information to define weights for the p-values. They showed that use of weights more often than not enhances power of the multiple testing procedure, unless the weights chosen are large and highly uninformative. Literature contains several instances of using prior data, expert knowledge and scientific insights to formulate weights. *Roeder et al.* (2006), *Roeder and Wasserman* (2009) utilize linkage data to define weights. *Benjamini and Cohen* (2017) build on the previously suggested weighted BH procedure using weights constructed using knowledge specific to clinical trials.

In turn, the concept of using weighted p-values was used to adapt the BH procedure to one-way classified hypotheses (*Pacifico et al.* (2004), *Benjamini and Heller* (2007), etc.). *Hu et al.* (2010) introduced a weighted BH procedure for one-way grouped hypotheses that calculates weights for each group proportional to the number of null hypotheses in it. Each member hypothesis of a particular group is assigned the same weight. The BH procedure is applied to these weighted hypotheses pooled together across all groups. *Ignatiadis et al.* (2016) suggested a data driven weighted procedure to test similarly classified hypotheses. Any set of weights that depend on external covariates and satisfy some simple constraints can be considered for a testing procedure similar to the method suggested in *Hu et al.* (2010). The optimum set of weights are chosen subject to maximization of power using data-based optimization techniques.

2.1.2 Weighted Multiple Testing Related to Hypotheses in a Single Group

The methods of multiple testing that we propose in this article are, in their oracle forms, applicable to positively dependent hypotheses that satisfy the PRDS condition. Suppose we have a set of N such positively dependent hypotheses $H_1, \ldots, H_N$. For the p-values corresponding to these hypotheses, we make the following assumption.

Assumption 1. *$P_i \sim U(0,1)$ for each $i \in I_0$, with I_0 being the set of indexes of null hypotheses.*

Regarding dependence among the p-values, we assume that they are positively regression dependent on subset (PRDS) of null p-values, as defined below generally for any set of random variable $X_1, \ldots, X_k$:

Condition 1. *A set of random variable $X_1, \ldots, X_k$ is said to be positively regression dependent on a particular subset S of these random variables if $E\left[\phi(X_1, \ldots, X_k)|X_i = x\right]$ is non-decreasing in x, for each $X_i \in S$ and for any (coordinatewise) non-decreasing function ϕ of $(X_1, \ldots, X_k)$.*

Clearly, independent p-values are PRDS. For examples of dependent p-values satisfying Condition 1, the readers are referred to *Benjamini and Yekutieli* (2001b), *Sarkar* (2002) and *Sarkar* (2008). A weaker form of positive dependence condition, with $E\left[\phi(X_1, \ldots, X_k)|X_i = x\right]$ replaced by $E\left[\phi(X_1, \ldots, X_k)|X_i \leq x\right]$, is often assumed in the literature in the context of BH type FDR controlling procedures (*Finner et al.*, 2009; *Sarkar*, 2008). This condition could have been used instead of Condition 1 in this paper without affecting our results relying on such a condition.

We next provide a general definition for the BH procedure applied to weighted p-values. This definition serves as our foundation and we systematically derive from it testing procedures suited to different classification structures levied on a set of hypotheses.

Definition 1. *For a set of N hypotheses, suppose that the ith p-value P_i is assigned a non-stochastic weight $w_i \geq 0$, for $i = 1, \ldots, N$. The weighted BH procedure at level α corresponding to these weights is a stepup procedure with the critical constants $i\alpha/N$, $i = 1, \ldots, N$, i.e., it orders the weighted p-values $P_i^w = w_i P_i, i = 1, \ldots, N$, in increasing order as $P_{(i)}^w$, $i = 1, \ldots, N$, and rejects the hypotheses $H_{(1)}, \ldots, H_{(R)}$ corresponding to $P_{(1)}^w, \ldots, P_{(R)}^w$ where*

$$R = \max\left\{1 \leq j \leq N : P_{(j)}^w \leq \frac{j\alpha}{N}\right\},$$

provided the maximum exists; otherwise, it rejects none.

Result 1. *The FDR of the weighted BH procedure based on p-values satisfying the PRDS condition is bounded above by $\frac{\alpha}{N} \sum\limits_{i \in I_0} \frac{1}{w_i}$.*

A proof of this result using techniques from *Sarkar* (2002) is provided in Appendix A.1.

Result 1 serves as our foundation. It leads to systematic development of our proposed procedures in their oracle forms through appropriate choice of weights suited to either one- or two-way classification structures levied on the set of hypotheses, before we construct their appropriate data-adaptive versions. More specifically, one can determine weights that satisfy

$$\sum_{i \in I_0} w_i^{-1} = N, \tag{2.1.1}$$

and appropriately capture the underlying classification structure to develop a weighted BH procedure in its oracle form that controls the FDR at α across all hypotheses, conservatively under PRDS, before constructing an appropriate data-adaptive version of it. For instance, the choice of weights, $w_i = \pi_0 = |I_0|/N$, for all $i = 1, \ldots, N$, satisfying this condition yields the single-group BH procedure in its oracle form. A data-adaptive version of it is the one that uses the existing data to estimate π_0. There are several such data-adaptive single-group BH procedures that have been put forward in the literature, for example, *Benjamini and Hochberg* (2000), *Storey et al.* (2004), *Sarkar* (2008) and *Blanchard and Roquain* (2009).

The same condition is also satisfied by the weights chosen by *Hu et al.* (2010) in their construction of a weighted BH procedure in its oracle form, referred to as the one-way grouped BH procedure, in the context of testing one-way classified hypotheses. Its control over the FDR under PRDS is also theoretically proved there, but now can be seen to follow from Result 1, which is more general. This procedure will be revisited in the next section where we introduce a data-adaptive version of it that is not only different from the one originally proposed in *Hu et al.* (2010), but seems more preferred since, unlike *Hu et al.* (2010), we have theoretically verified its non-asymptotic FDR control, at least under independence.

The next section will also contain FDR controlling procedures in their oracle as well as data-adaptive forms for testing two-way classified hypotheses, which this article introduces for the first time in the multiple testing literature, as far as we know. The non-asymptotic FDR control of all these new data-adaptive procedures for one- and two-way classified hypotheses under independence is established using the following result.

Result 2. *The FDR of a data-adaptive weighted BH procedure with (coordinate wise) non-decreasing estimated weight functions $\hat{w}_i(\mathbf{P}) > 0$, $i = 1, \ldots, N$, is bounded above* *by α under independence if*

$$E\left[\sum_{i\in I_0}\frac{1}{\hat{w}_i(\mathbf{P}^{(-i)},0)}\right]\leq N \tag{2.1.2}$$

where $\hat{w}_i(\mathbf{P}^{(-i)},0)$ represents $\hat{w}_i$ as a function of $\mathbf{P}^{(-i)}=\{P_1,\ldots,P_N\}\backslash P_i$ with $P_i=0$.

A proof of this result can be seen in *Sarkar* (2008).

We introduce below a newer class of estimates, expressing some of the existing ones in a more general form, which offers a wider scope of data-dependent adaptation of the BH procedure to both one- and two-way group structures of hypotheses with proven non-asymptotic FDR control under independence. The following lemma will be useful in checking the inequality in (2.1.2) for this larger class of estimates:

Lemma 1. *Let $R_N(\lambda)=\sum_{i=1}^N I(P_i\leq\lambda)$ and $R_{N-1}^{(-i)}(\lambda)=\sum_{j(\neq i)=1}^N I(P_j\leq\lambda)$, for a fixed $\lambda\in(0,1)$. Then, for any non-negative real valued function f of $R_N(\lambda)$, we have the following result:*

$$E\left\{\sum_{i\in I_0}\frac{(1-\lambda)f(R_{N-1}^{(-i)}(\lambda))}{N-R_{N-1}^{(-i)}(\lambda)}\right\}\leq E\left\{f(R_N(\lambda))\right\}. \tag{2.1.3}$$

Proof: Let $V_N(\lambda)=\sum_{i\in I_0}I(P_i\leq\lambda)$. Then, since $N_0-V_N(\lambda)\leq N-R_N(\lambda)$, $N_0-V_N(\lambda)=\sum_{i\in I_0}I(P_i>\lambda)$, and $R_N(\lambda)=R_{N-1}^{(-i)}(\lambda)+I(P_i\leq\lambda)$, for any $i\in I_0$, the inequality in (2.1.3) follows from the following:

$$\begin{aligned}
E\left\{f(R_N(\lambda))\right\} &\geq E\left\{\frac{[N_0-V_N(\lambda)]f(R_N(\lambda))}{[N-R_N(\lambda)]\vee 1}\right\}\\
&= E\left\{\sum_{i\in I_0}\frac{I(P_i>\lambda)f(R_{N-1}^{(-i)}(\lambda)+I(P_i\leq\lambda))}{[N-R_{N-1}^{(-i)}(\lambda)-I(P_i\leq\lambda)]\vee 1}\right\},
\end{aligned}$$

which reduces to the left-hand side of (2.1.3) since the P_i's are independent.

Remark 1. The lemma can be generalized with R_N and $R_{N-1}^{(-i)}$ re-defined, respectively, as $R_N = \max\left\{1 \le j \le N : P_{(j)} \le \lambda_j\right\}$ and

$$R_{N-1}^{(-i)} = \max\left\{1 \le j \le N-1 : P_{(j)}^{(-i)} \le \lambda_{j+1}\right\},$$

using some constants $0 \le \lambda_1 \le \cdots \le \lambda_N \le \lambda$, where $P_{(j)}, j = 1, \ldots, N$, are the increasingly ordered components of $(P_1, \ldots, P_N)$, $P_{(j)}^{(-i)}$, $j = 1, \ldots, N-1$ are the same for $(P_1, \ldots, P_N) \setminus \{P_i\}$, and R_N and $R_{N-1}^{(-i)}$ are zero if the maximums in their definitions don't exist. This generalized lemma can be proved as in Lemma 1 by noting from *Sarkar* (2008) that $R_N = \sum_{i=1}^N I(P_i \le \lambda_{R_{N-1}^{(-i)}+1}) = R_{N-1}^{(-i)} + I(P_i \le \lambda_{R_{N-1}^{(-i)}+1})$, and $N_0 - V_N = \sum_{i \in I_0} I(P_i > \lambda_{R_{N-1}^{(-i)}+1})$.

It is important to note, as we proceed to use Lemma 1 to develop procedures under more complex structures of hypotheses in the next section, that R will be subscripted differently under different structural settings since its definition should correctly reflect the number of hypotheses involved.

We generalize the concept of weights used by *Hu et al.* (2010) so the same procedure can be applied to any structure of grouped hypotheses by appropriately choosing the weights reflecting the structure. In the following sections, we consider three types of classifications commonly imposed on large sets of hypotheses, i.e., one-way, two-way with one hypothesis per cell, and two-way with multiple hypotheses per cell, and introduce the weights in their oracle forms and the corresponding weighted BH procedures controlling the FDR. We also introduce the data-adaptive versions of the procedures with proven control on FDR non-asymptotically under independence. Through detailed simulation studies, we compare the performances of our proposed methods with existing practices in terms of control on FDR and average power and finally, we explain our findings from application of one of our proposed methods on a microbiome dataset.

2.2 One-Way Grouped BH Procedure: Adapting the BH Procedure to One-Way Classified Hypotheses

Let us suppose that the N hypotheses to be simultaneously tested are being classified into m non-overlapping groups according to some pre-defined criterion, with $n_{g.}$ pairs of hypothesis and the corresponding p-value, (H_{gi}, P_{gi}), $i = 1, \ldots, n_{g.}$, falling in group g, and $N = \sum_{g=1}^{m} n_{g.}$. Let n_{g0} be the number of null hypotheses and $I_{g0} \subseteq \{1, \ldots, n_{g.}\}$ be the corresponding set of sub-indexes associated with i in group g. The set of indexes of all null hypotheses among all hypotheses can then be expressed as $I_0 = \bigcup_{g=1}^{m} I_{g0}$. Let $\pi_{g0} = n_{g0}/n_{g.}$ be the proportion of true nulls in group g, so that π_0, the proportion of true nulls in the entire set of N hypotheses, can be expressed as $\pi_0 = \sum_{g=1}^{m} n_{g.}\pi_{g0}/N$.

One-Way grouped BH, shortly One-Way GBH, is an oracle procedure. It is defined by *Hu et al.* (2010) as a weighted BH procedure with the weights being formulated in terms π_{g0}, for $g = 1, \ldots, m$, assuming they are known, in a way that allows the BH procedure to effectively adapt to the present structural setting of the hypotheses. We revisit it in the following sub-section, before developing our newly proposed data-adaptive version of it later in this section.

2.2.1 Oracle One-Way GBH Procedure

It is a weighted BH procedure with

$$w_g = \frac{\pi_{g0}(1 - \pi_0)}{1 - \pi_{g0}} \tag{2.2.1}$$

being assigned as weight to P_{gi}, for each $i = 1, \ldots, n_{g.}$, and $g = 1, \ldots, m$, assuming these proportions are all known. *Hu et al.* (2010) referred to it as simply grouped BH, shortly GBH procedure, but as said above, we will refer to it here as One-Way

GBH procedure. Since

$$\sum_{g=1}^{m} \sum_{i \in I_{g0}} w_g^{-1} = \frac{1}{1-\pi_0} \sum_{g=1}^{m} \frac{n_{g.} \pi_{g0}(1-\pi_{g0})}{\pi_{g0}} = \frac{1}{1-\pi_0} \sum_{g=1}^{m} n_{g.}(1-\pi_{g0}) = N, \tag{2.2.2}$$

the equality in (2.1.1) is satisfied by these weights, and so we have the following theorem, which of course was proved in *Hu et al.* (2010) using different arguments.

Theorem 2.1. *One-Way GBH procedure controls the overall FDR under PRDS and Assumption 1.*

There is a Bayesian justification behind the choice of these weights, as articulated by Hu et al. (2010). However, a look at these weights from a different point of view seems to provide further insight into the effectiveness of these weights under the current setting.

For a group with small proportion of true nulls, π_{g0} would be small. At the same time, it would have higher odds of being significant relative to other groups, as measured by $(1-\pi_{g0})/(1-\pi_0)$. Consequently, the weight associated with that group gets deflated, facilitating easier rejection of its members when the weighted BH procedure is applied to all the hypotheses.

This sort of interpretation for the weights guides us in understanding how to estimate them, differently from *Hu et al.* (2010), in constructing a data-adaptive version of One-Way GBH procedure that we will describe below.

2.2.2 Data-Adaptive One-Way GBH Procedure

We propose this procedure by considering the One-Way GBH and replacing the weight w_g in it by the following:

$$\hat{w}_g = \frac{n_{g.} - R_{n_{g.}} + 1}{N(1-\lambda)} \cdot \frac{R_N + m - 1}{R_{n_{g.}}}, g = 1, \ldots, m, \tag{2.2.3}$$

where, for some fixed $\lambda \in (0,1)$, $R_{n_{g.}} \equiv R_{n_{g.}}(\lambda) = \sum_{i=1}^{n_{g.}} I(P_{gi} \le \lambda)$, and $R_N = \sum_{g=1}^{m} R_{n_{g.}}$. We refer to this procedure as a data-adaptive One-Way GBH procedure.

The idea of using this type of estimate for each w_g came from its alternative interpretation noted above. We estimate π_{g0} by $\hat{\pi}_{g0} = (n_{g.} - R_{n_{g.}} + 1)/n_{g.}(1-\lambda)$, which is a slight adjustment (considered by *Storey et al.* (2004)), from $n_{g.} - R_{n_{g.}}$ to $n_{g.} - R_{n_{g.}} + 1$, made in the estimate originally used by *Storey* (2002) for the proportion of true nulls in the context of single-group multiple testing. To estimate $(1-\pi_0)/(1-\pi_{g0})$, we propose estimating $(1-\pi_{g0})/(1-\pi_0)$, which is the proportion of false nulls in group g among all false nulls, by $N(R_{n_{g.}} + m - 1)/n_g R_N$, having made a slight adjustment to its natural estimate $NR_{n_{g.}}/n_{g.}R_N$, and then inverting this estimate. When $m = 1$, the $\hat{w}_g$ in expression (2.2.3) reduces to that of *Storey et al.* (2004).

Theorem 2.2. *The above data adaptive One-Way GBH procedure controls the overall FDR under independence among all p-values and Assumption 1.*

Proof. The theorem will follow from Result 2 if we can show that the estimated weights, treated as functions of all the p-values, used in adaptive One-Way GBH satisfy the two conditions in Result 2 - (i) $\hat{w}_g$ is non-decreasing in P_{gi} for each g, and (ii) the inequality in (2.1.2) holds.

The first condition follows by noting that both $n_{g.} - R_{n_{g.}} + 1$ and $(R_N + m - 1)/R_{n_{g.}} = 1 + \sum_{g' \ne g} (R_{n_{g.'}} + 1)/R_{n_{g.}}$ are non-increasing in each $R_{n_{g.}}$, which itself is non-increasing in each P_{gi}.

To show that the second condition is also satisfied, let us first express $\hat{w}_g$ as a function of $\mathbf{P}$, the set of all p-values, i.e., as $\hat{w}_g(\mathbf{P})$, for each g. Then, note that if we set P_{gi} at 0 for a particular pair $(g, i \in I_{g0})$, we get

$$\hat{w}_g(\mathbf{P}^{-(g,i)}, 0) = \frac{n_{g.} - R^{(-i)}_{n_{g.}-1}}{N(1-\lambda)} \; \frac{R^{(-i)}_{n_{g.}-1} + \sum_{g' \ne g} R_{n_{g'_.}} + m}{R^{(-i)}_{n_{g.}-1} + 1}, \; g = 1, \ldots, m,$$

where $R^{(-i)}_{n_{g\cdot}-1} = \sum_{i' \neq i} I(P_{gi'} \leq \lambda)$. Thus, we have

$$\mathrm{E}\left\{\sum_{g=1}^{m}\sum_{i\in I_{g0}} \frac{1}{\hat{w}_g(\mathbf{P}^{-(g,i)},0)}\right\} = N\mathrm{E}\left\{\sum_{g=1}^{m}\sum_{i\in I_{g0}} \frac{(1-\lambda)f(R^{(-i)}_{n_{g\cdot}-1}, \sum_{g'\neq g} R_{n_{g'\cdot}})}{n_{g\cdot} - R^{(-i)}_{n_{g\cdot}-1}}\right\} \tag{2.2.4}$$

where

$$f\left(x, \sum_{g'\neq g} R_{n_{g'\cdot}}\right) = \frac{x+1}{x + \sum_{g'\neq g} R_{n_{g'\cdot}} + m}.$$

Applying Lemma 1 to the expectation in the right-hand side of equation (2.2.4) with respect to the p-values in the gth group, and completing the expectation with respect to all p-values, we see that the left-hand side of equation (2.2.4) is less than or equal to

$$N\mathrm{E}\left\{\sum_{g=1}^{m} \frac{R_{n_{g\cdot}}+1}{R_{n_{g\cdot}} + \sum_{g'\neq g} R_{n_{g'\cdot}} + m}\right\} = N\mathrm{E}\left\{\sum_{g=1}^{m} \frac{R_{n_{g\cdot}}+1}{R_N + m}\right\} = N\mathrm{E}\left\{\frac{R_N+m}{R_N+m}\right\} = N,$$

i.e., the second condition is also satisfied. Thus, the theorem is proved.

2.3 Two-Way Grouped BH Procedure: Adapting the BH method to Two-Way Classified Hypotheses

Suppose, the N hypotheses can be classified simultaneously according to two criteria and can be laid out in an $m \times n$ matrix with $n_{gh} \geq 1$ hypotheses at the intersection of the gth row and the hth column, i.e., in the (g,h)th cell of the matrix. We will consider the two different scenarios, one involving only one hypothesis per cell (i.e., $n_{gh} = 1$) and the other involving multiple hypotheses in each cell (i.e., $n_{gh} > 1$), separately in the following two sub-sections. This would give a clearer picture of how our proposed procedures extend from one- to multiple-hypotheses-per-cell. Also, in each of these scenarios, there is more than one choice of weights to define our proposed procedure in its oracle form, before constructing its data-adaptive version, that captures the underlying two-way structure. However, we will focus on one of them and formally present the corresponding oracle and data-adaptive procedures as our proposed ones to use for further evaluation, and simply point out the scope of deriving similar procedures using other weights.

2.3.1 One Hypothesis Per Cell: $n_{gh} = 1$

Let n_{g0} be the number of true nulls in the gth row, for $g = 1, \ldots m$, and m_{0h} be the number of true nulls in the hth column, for $h = 1, \ldots, n$. The subsets of indexes of true nulls associated with h in the gth row and g in the hth column are, respectively, $I_{g0} \subseteq \{1, \ldots, n\}$ and $I_{0h} \subseteq \{1, \ldots, m\}$. Consequently, the set of indexes of true nulls among the entire set of hypotheses can be expressed as $I_0 = \bigcup_{g=1}^{m} I_{g0} = \bigcup_{h=1}^{n} I_{0h}$.

The proportion of true nulls in the gth row is defined as $\pi_{g0} = n_{g0}/n$, and that in the hth column as $\pi_{0h} = m_{0h}/m$. The proportion of true nulls in the entire set of $N = mn$ hypotheses is $\pi_0 = \sum_{g=1}^{m} \pi_{g0}/m = \sum_{h=1}^{n} \pi_{0h}/n$.

2.3.1.1 Oracle Two-Way GBH Procedure With One Hypothesis Per Cell

The hypothesis H_{gh} at the intersection of the gth row and hth column is affected upon by its both parent row and column, which motivates us to consider assigning the following weight to P_{gh} corresponding to H_{gh}, assuming all these proportions are known:

$$w_{gh} = \left[\frac{1}{2}\left\{\frac{1}{\pi_{g0}}\frac{1-\pi_{g0}}{1-\pi_0} + \frac{1}{\pi_{0h}}\frac{1-\pi_{0h}}{1-\pi_0}\right\}\right]^{-1}, \; g = 1, \ldots, m, \; h = 1, \ldots, n, \qquad (2.3.1)$$

to simultaneously account for both row and column effects. The weighted BH procedure applied to the N hypotheses based on the weighted p-values $P_{gh}^{w} = w_{gh}P_{gh}$, for $g = 1, \ldots, m; h = 1, \ldots, n$, is one of our proposed procedures in its oracle form, which we refer to as an Oracle Two-Way GBH$_1$ procedure.

This weight is a simple extension of that from one- to two-way classification setting. If the parent row has a low proportion of true nulls π_{g0}, it subsequently has a large odds of being significant relative to other rows, as indicated by $(1-\pi_{g0})/(1-\pi_0)$. This reduces the weight w_{gh}, making H_{gh} more likely to be rejected. The weight is similarly affected by the parent column. We assume that both classifications have equal impacts on the individual hypothesis, and so the weight is a function of the simple mean of contributions from each of parent groups.

Noting that w_{gh} can be expressed as $w_{gh}^{-1} = \frac{1}{2}w_g^{-1} + \frac{1}{2}w_h^{-1}$, with w_g along the rows being defined in expression (2.2.1) and w_h being defined similarly along the columns as $w_h = \frac{\pi_{0h}(1-\pi_0)}{1-\pi_{0h}}$, $h = 1, \ldots, n$, one sees from (2.2.2) that

$$\sum_{g=1}^{m} \sum_{h \in I_{g0}} w_{gh}^{-1} = \frac{1}{2} \sum_{g=1}^{m} \sum_{h \in I_{g0}} w_g^{-1} + \frac{1}{2} \sum_{h=1}^{n} \sum_{g \in I_{0h}} w_h^{-1} = \frac{1}{2}N + \frac{1}{2}N = N, \tag{2.3.2}$$

since $\sum_{g=1}^{m} \sum_{h \in I_{g0}} = \sum_{h=1}^{n} \sum_{g \in I_{0h}}$. Thus, equality in (2.1.1) is satisfied for the weights in expression (2.3.1), and so we can state the following theorem from Result 1 without offering a proof.

Theorem 2.3. *Oracle Two-Way GBH$_1$ procedure based on the weights in (2.3.1) controls the overall FDR under PRDS and Assumption 1.*

Remark 2. The weight in (2.3.1) can be customized, still satisfying (2.3.2), to suit variable influence of the row and column classifications. As an example, the weight can be adapted to reflect the imbalance between the number of rows and columns as follows: :

$$w_{gh} = \left[\frac{1}{(m+n)} \left\{ \frac{m}{\pi_{g0}} \cdot \frac{1-\pi_{g0}}{1-\pi_0} + \frac{n}{\pi_{0h}} \cdot \frac{1-\pi_{0h}}{1-\pi_0} \right\} \right]^{-1}, \; g = 1, \ldots, m, \; h = 1, \ldots, n. \tag{2.3.3}$$

These weights additionally account for the proportion of groups along the rows and also along the columns out of the total number of groups. Clearly, if $m = n$, they reduce to those in (2.3.1). Such choice of weights in adapting BH procedure to two-way classification structure is not unique, and there remains a scope for other choices depending on variable factors or external information.

2.3.1.2 Data-Adaptive Two-Way GBH Procedure With One Hypothesis Per Cell

We consider the Oracle Two-Way GBH$_1$ procedure in Theorem 2.3 and estimate the weights in it by the following:

$$\hat{w}_{gh} = \left[\frac{N(1-\lambda)}{2} \left\{ \frac{1}{\{n - R_{n_g} + 1\}} \cdot \frac{R_{n_g}}{\{R_N + m - 1\}} + \frac{1}{\{m - R_{m_h} + 1\}} \cdot \frac{R_{m_h}}{\{R_N + n - 1\}} \right\} \right]^{-1},$$
$$g = 1, \ldots, m, \; h = 1, \ldots, n, \qquad (2.3.4)$$

where $R_{n_g} \equiv R_{n_g}(\lambda) = \sum_{h=1}^{n} I(P_{gh} \leq \lambda)$, $R_{m_h} \equiv R_{m_h}(\lambda) = \sum_{g=1}^{m} I(P_{gh} \leq \lambda)$, and $R_N \equiv R_N(\lambda) = \sum\limits_{g=1}^{m} R_{n_g}(\lambda) = \sum\limits_{h=1}^{n} R_{m_h}(\lambda)$, for some fixed $\lambda \in (0, 1)$. The estimated weight assigned to each p-value is similar to that for one-way classified hypotheses; however, it accounts for both parent row and column effects. We refer to this procedure as a Data-Adaptive Two-Way GBH$_1$.

We have the following theorem as an extension of Theorem 2.2 from one- to two-way classification setting.

Theorem 2.4. *The above Data-Adaptive Two-Way GBH$_1$ procedure controls the overall FDR under independence among all p-values and Assumption 1.*

Proof. This theorem can be proved based on the same arguments that were used to prove Theorem 2.2 using Result 2. First, note that $\hat{w}_{gh}^{-1} = \frac{1}{2}\tilde{w}_g^{-1} + \frac{1}{2}\tilde{w}_h^{-1}$, where

$$\tilde{w}_g = \frac{n - R_{n_g} + 1}{N(1-\lambda)} \frac{R_N + m - 1}{R_{n_g}},$$

and

$$\tilde{w}_h = \frac{m - R_{m_h} + 1}{N(1-\lambda)} \; \frac{R_N + n - 1}{R_{m_h}}.$$

From this we see that $\hat{w}_{gh}$ is non-decreasing in P_{gh}, since both $\tilde{w}_g$ and $\tilde{w}_h$ are non-decreasing in P_{gh}, which can be proved exactly the way it was proved for the $\hat{w}_g$ in Theorem 2.2.

Moreover, as proved in Theorem 2.2 for $\hat{w}_g$ using Lemma 1, we have the following inequalities for $\tilde{w}_g$ and $\tilde{w}_h$ under the assumption of independence among all p-values:

$$\mathrm{E}\left\{\sum_{g=1}^{m}\sum_{g\in I_{g0}} \frac{1}{\tilde{w}_g(\mathbf{P}^{-(g,h)},0)}\right\} \le N,$$

$$\mathrm{E}\left\{\sum_{h=1}^{n}\sum_{g\in I_{0h}} \frac{1}{\tilde{w}_h(\mathbf{P}^{-(g,h)},0)}\right\} \le N,$$

from which we see that

$$\begin{aligned}&\mathrm{E}\left\{\sum_{g=1}^{m}\sum_{h\in I_{g0}} \frac{1}{\hat{w}_{gh}(\mathbf{P}^{-(g,h)},0)}\right\}\\ =&\frac{1}{2}E\left\{\sum_{g=1}^{m}\sum_{h\in I_{g0}} \frac{1}{\tilde{w}_g(\mathbf{P}^{-(g,h)},0)}\right\} + \frac{1}{2}E\left\{\sum_{h=1}^{n}\sum_{g\in I_{0h}} \frac{1}{\tilde{w}_h(\mathbf{P}^{-(g,h)},0)}\right\} \le N,\end{aligned}$$

i.e., inequality (2.1.2) in Result 2 holds. Thus, the theorem is proved.

Remark 3. The estimate of weight considered above is stated in its simplest form. Like its oracle counterpart, it can also be modified as

$$\hat{w}_{gh} = \left[\frac{N(1-\lambda)}{m+n}\left\{\frac{1}{n - R_{n_g} + 1}\cdot\frac{mR_{n_g}}{R_N + m - 1} + \frac{1}{m - R_{m_h} + 1}\cdot\frac{nR_{m_h}}{R_N + n - 1}\right\}\right]^{-1},$$
$$g = 1,\ldots,m,\; h = 1,\ldots,n \qquad (2.3.5)$$

In addition to accounting for the row and column effects, this expression also accounts for the difference in numbers of rows and columns and accordingly emphasizes the corresponding effects on the individual hypothesis. Theorem 2.4 can also be stated in terms of adaptive two-way GBH with one hypothesis per cell in terms of these alternative weights.

2.3.2 Multiple Hypothesis Per Cell: $n_{gh} > 1$

Let $n_{g\bullet} = \sum_{h=1}^{n} n_{gh}$ and $n_{\bullet h} = \sum_{g=1}^{m} n_{gh}$ be the total numbers of hypotheses, respectively, in the gth row and hth column, so that

$$N = \sum_{g=1}^{m} n_{g\bullet} = \sum_{h=1}^{n} n_{\bullet h} = \sum_{g=1}^{m} \sum_{h=1}^{n} n_{gh}.$$

Let n_{gh0} be the number of true nulls in the (g, h)th cell and $I_{gh0} \subseteq \{1, \ldots, n_{gh}\}$ be the corresponding subset of indexes of true nulls. The overall set of indexes of the true nulls is

$$I_0 = \bigcup_{g=1}^{m} \bigcup_{h=1}^{n} I_{gh0}.$$

The proportion of true nulls in the (g, h)th cell of the $m \times n$ matrix is $\pi_{gh0} = n_{gh0}/n_{gh}$. This helps to define $\pi_{g00} = \sum_{h=1}^{n} n_{gh}\pi_{gh0} / \sum_{h=1}^{n} n_{gh}$, $\pi_{0h0} = \sum_{g=1}^{m} n_{gh}\pi_{gh0} / \sum_{g=1}^{m} n_{gh}$, and

$$N\pi_0 = \sum_{g=1}^{m} n_{g\bullet}\pi_{g00} = \sum_{h=1}^{n} n_{\bullet h}\pi_{0h0} = \sum_{g=1}^{m} \sum_{h=1}^{n} n_{gh}\pi_{gh0}$$

.

2.3.2.1 Oracle Two-Way GBH Procedure With Multiple Hypotheses Per Cell

Suppose P_{ghk} is the kth p-value in the (g, h)th cell, and H_{ghk} is the corresponding hypothesis. Assuming that all the proportions mentioned above are known, we consider assigning the following weights to P_{ghk}, for each $k = 1, \ldots, n_{gh}$, to capture the underlying two-way classification structure of the hypotheses, and refer to the resulting weighted BH procedure as an Oracle Two-Way GBH$_{>1}$.

$$w_{gh} = \left[\frac{1}{4}\left\{\frac{1}{\pi_{gh0}}\left(\frac{1-\pi_{gh0}}{1-\pi_{g00}} + \frac{1-\pi_{gh0}}{1-\pi_{0h0}}\right) + \left(\frac{1}{\pi_{g00}} \cdot \frac{1-\pi_{g00}}{1-\pi_0} + \frac{1}{\pi_{0h0}} \cdot \frac{1-\pi_{0h0}}{1-\pi_0}\right)\right\}\right]^{-1} \tag{2.3.6}$$

Expressing w_{gh} as $w_{gh}^{-1} = \frac{1}{4}w_{1,gh}^{-1} + \frac{1}{4}w_{2,gh}^{-1} + \frac{1}{4}w_g^{-1} + \frac{1}{4}w_h^{-1}$, where

$$\begin{aligned} &w_{1,gh} = \frac{\pi_{gh0}(1-\pi_{g00})}{1-\pi_{gh0}}, \; w_{2,gh} = \frac{\pi_{gh0}(1-\pi_{0h0})}{1-\pi_{gh0}}, \; w_g = \frac{\pi_{g00}(1-\pi_0)}{1-\pi_{g00}}, \\ &\text{and} \quad w_h = \frac{\pi_{0h0}(1-\pi_0)}{1-\pi_{0h0}}, \end{aligned}$$

one can see that

$$\sum_{g=1}^{m}\sum_{h=1}^{n}\sum_{k \in I_{gh0}} \left[w_{1,gh}^{-1} + w_{2,gh}^{-1} + w_g^{-1} + w_h^{-1}\right] = 4N,$$

that is, the equality in (2.1.1) is satisfied by the weights in (2.3.6). Therefore, we can state the following from Result 1 without a proof

Theorem 2.5. *Two-way GBH with multiple hypotheses per cell based on the weights in (2.3.6) controls the overall FDR conservatively under PRDS and Assumption 1.*

Remark 4. Of course, one can consider defining two-way GBH procedure with multiple hypotheses per cell based on other types of weight subject to the equality in (2.1.1). For instance, following the preceding case of one hypothesis per cell in the two-way classification setup, a natural choice of weight assigned to hypotheses in the (g, h)th cell would be

$$w_{gh} = \left[\frac{1}{2}\left\{\frac{1}{\pi_{g00}} \cdot \frac{1-\pi_{g00}}{1-\pi_0} + \frac{1}{\pi_{0h0}} \cdot \frac{1-\pi_{0h0}}{1-\pi_0}\right\}\right]^{-1} \tag{2.3.7}$$

The choice of weight for the two-way GBH procedure in Theorem 2.5 consists of an additional term that depends on the ratio of the proportion of signals in each cell to the same proportions in the parent row and column. Owing to unequal number of members at the intersections in the two-way layout, further modifications of the weights would be complicated. However, if there are an equal number, say $p > 0$, hypotheses at each cell, we can further edit these weights in (2.3.7) as

$$w_{gh} = \left[\frac{1}{p(m+n)}\left\{\frac{mp}{\pi_{g00}} \cdot \frac{1-\pi_{g00}}{1-\pi_0} + \frac{np}{\pi_{0h0}} \cdot \frac{1-\pi_{0h0}}{1-\pi_0}\right\}\right]^{-1}, \tag{2.3.8a}$$

and those in the procedure in Theorem 2.5 as

$$\begin{aligned} w_{gh} = \Bigg[\frac{1}{p(m+n)}\Bigg\{&\frac{p}{\pi_{gh0}}\left(\frac{1-\pi_{gh0}}{1-\pi_{g00}} + \frac{1-\pi_{gh0}}{1-\pi_{0h0}}\right) \\ &+ \left(\frac{p(m-1)}{\pi_{g00}} \cdot \frac{1-\pi_{g00}}{1-\pi_0} + \frac{p(n-1)}{\pi_{0h0}} \cdot \frac{1-\pi_{0h0}}{1-\pi_0}\right)\Bigg\}\Bigg]^{-1}. \end{aligned} \tag{2.3.8b}$$

2.3.2.2 Data-Adaptive Two-Way GBH Procedure With Multiple Hypotheses Per Cell

Consider the Two-Way $\mathrm{GBH}_{>1}$ in Theorem 2.5 to replace its weight w_{gh} by

$$\begin{aligned} \hat{w}_{gh} =& \Bigg[\frac{1}{4}\Bigg\{\frac{1-\lambda}{n_{gh} - R_{n_{gh}} + 1}\left(\frac{n_{g\cdot}R_{n_{gh}}}{R_{n_{g\cdot}} + n - 1} + \frac{n_{\cdot h}R_{n_{gh}}}{R_{n_{\cdot h}} + m - 1}\right) \\ &+ N(1-\lambda)\left(\frac{R_{n_{g\cdot}}}{\{n_{g\cdot} - R_{n_{g\cdot}} + 1\}\{R_N + m - 1\}} + \frac{R_{n_{\cdot h}}}{\{n_{\cdot h} - R_{n_{\cdot h}} + 1\}\{R_N + n - 1\}}\right)\Bigg\}\Bigg]^{-1} \\ & g = 1\ldots, m,\ h = 1, \ldots, n, \end{aligned} \tag{2.3.9}$$

where, for some fixed λ, $R_{n_{gh}} \equiv R_{n_{gh}}(\lambda) = \sum_{k=1}^{n_{gh}} I(P_{ghk} \leq \lambda)$, $R_{n_{g.}} \equiv R_{n_{g.}}(\lambda) = \sum_{h=1}^{n} R_{n_{gh}}(\lambda)$, $R_{n_{.h}} \equiv R_{n_{.h}}(\lambda) = \sum_{g=1}^{m} R_{n_{gh}}(\lambda)$.

$$R_N = \sum_{g=1}^{m} R_{n_{g.}} = \sum_{h=1}^{n} R_{n_{.h}} = \sum_{g=1}^{m} \sum_{h=1}^{n} R_{n_{gh}}.$$

It is referred to as a Data-Adaptive Two-Way GBH$_{>1}$.

Theorem 2.6. *The above Data-Adaptive Two-Way GBH$_{>1}$ controls the overall FDR under independence among all p-values and Assumption 1.*

Proof. Again, this theorem will be proved based on the same arguments that were used to prove Theorem 2.4 by verifying that the conditions in Result 2 are satisfied by the weight functions $\hat{w}_{gh}$; i.e., it is increasing in each P_{ghk} and that the following inequality holds with P_{ghk} being set to 0 in it: $\sum_{g=1}^{m} \sum_{h=1}^{n} \sum_{k \in I_{gh0}} \hat{w}_{gh}^{-1}(\mathbf{P}^{-(g,h,k)}, 0) \leq N$.

As in proving Theorem 2.4, let us consider $\hat{w}_{gh}$ in terms of the following representation: $\hat{w}_{gh}^{-1} = \frac{1}{4}\tilde{w}_{1,gh}^{-1} + \frac{1}{4}\tilde{w}_{2,gh}^{-1} + \frac{1}{4}\tilde{w}_{g}^{-1} + \frac{1}{4}\tilde{w}_{h}^{-1}$, where

$$\begin{aligned} \tilde{w}_{1,gh} &= \frac{n_{gh} - R_{n_{gh}} + 1}{n_{g.}(1-\lambda)} \cdot \frac{R_{n_{g.}} + n - 1}{R_{n_{gh}}}, \ \tilde{w}_{2,gh} = \frac{n_{gh} - R_{n_{gh}+1}}{n_{.h}(1-\lambda)} \cdot \frac{R_{n_{.h}} + m - 1}{R_{n_{gh}}}, \\ \tilde{w}_{g} &= \frac{n_{g.} - R_{n_{g.}} + 1}{N(1-\lambda)} \cdot \frac{R_N + m - 1}{R_{n_{g.}}}, \ \tilde{w}_{h} = \frac{n_{.h} - R_{n_{.h}} + 1}{N(1-\lambda)} \cdot \frac{R_N + n - 1}{R_{n_{.h}}}, \end{aligned} \tag{2.3.10}$$

As argued before in proving Theorems 2.2 and 2.4, each of the four weights in (2.3.10) can be shown to satisfy the same two properties that we intend to show for $\hat{w}_{gh}$. In other words, $\hat{w}_{gh}$ satisfies the desired two conditions in Result 2, and hence the theorem is proved.

Remark 5. Other choices of weights can be suggested as

$$\hat{w}_{gh} = \left[\frac{N(1-\lambda)}{2}\left\{\frac{1}{n_{g\bullet} - R_{g.} + 1}\frac{R_{g.}}{R_N + m - 1} + \frac{1}{n_{\bullet h} - R_{.h} + 1}\frac{R_{.h}}{R_N + n - 1}\right\}\right]^{-1} \tag{2.3.11}$$

As in the oracle case, these weights can be further modified to be more informative if there are an equal number of hypotheses ($p > 0$) at each cell. The modified choice corresponding to expression (2.3.9) would be

$$\begin{aligned}\hat{w}_{gh} =& \Bigg[\frac{1}{(m+n)p}\Bigg\{\frac{p(1-\lambda)}{n_{gh.} - R_{n_{gh}} + 1}\left(\frac{n_{g\bullet}R_{n_{gh}}}{R_{n_{g\bullet}} + n - 1} + \frac{n_{\bullet h}R_{n_{gh}}}{R_{n_{\bullet h}} + m - 1}\right) \\ &+ N(1-\lambda)\left(\frac{p(m-1)R_{n_{g\bullet}}}{\{n_{g\bullet} - R_{n_{g\bullet}} + 1\}\{R_N + m - 1\}} + \frac{p(n-1)R_{n_{\bullet h}}}{\{n_{\bullet h} - R_{n_{\bullet h}} + 1\}\{R_N + n - 1\}}\right)\Bigg\}\Bigg]^{-1} \\ &\forall g = 1\ldots, m,\ h = 1, \ldots, n \end{aligned} \tag{2.3.12}$$

and the modified choice corresponding to expression (2.3.11) is

$$\hat{w}_{gh} = \left[\frac{N(1-\lambda)}{(m+n)p}\left\{\frac{mp}{n_{g\bullet} - R_{g.} + 1}\frac{R_{g.}}{R_N + m - 1} + \frac{np}{n_{\bullet h} - R_{.h} + 1}\frac{R_{.h}}{R_N + n - 1}\right\}\right]^{-1} \tag{2.3.13}$$

2.4 Results on Simulated Data

We carried out extensive simulation studies to investigate the performances of our proposed procedures in Theorems 2-6 in terms FDR control and power (expected proportion of correctly rejected false nulls among all false nulls) against their relevant competitors. This section discusses these results.

2.4.1 One-Way Classified Hypotheses

Here, our study was designed to compare the performance of the Data Adaptive One-Way GBH procedure in Theorem 2.2 with its following three relevant competitors, the first two of which were considered in *Hu et al.* (2010) as extensions to one-way classification setting of the single-group data-adaptive BH procedures proposed, respectively, in *Benjamini and Hochberg* (2000) and *Benjamini et al.* (2006).

LSL (Least-Slope) Grouped BH: One-Way GBH procedure with π_{g0} in equation (2.2.1) being estimated by the following, for each g:

$$\hat{\pi}_{g0}^{\mathrm{LSL}} = \min\left(\frac{\lfloor l_{g,i} \rfloor + 1}{n}, 1\right), \; l_{g,i} = \frac{n - i + 1}{1 - P_{g,(i)}},$$

such that $l_{g,i} > l_{g,i-1}$, with $P_{g,(i)}$ being the ith minimum ordered p-value in the gth group.

TST (Two-Stage) Grouped BH: One-Way GBH procedure with π_{g0} in equation (2.2.1) being estimated by the following, for each g:

$$\hat{\pi}_{g0}^{\mathrm{TST}} = \frac{n - r_g}{n},$$

with r_g being the number of rejections obtained by applying the non-adaptive BH procedure to the p-values in the gth group at level $\alpha/(1+\alpha)$.

Naive Adaptive BH: The usual data-adaptive BH with the following estimate of π_0:

$$\hat{\pi}_0 = \frac{N - R_N + 1}{N(1-\lambda)}, \text{ with } R_N = \sum_{g=1}^{m} \sum_{i=1}^{n_g} I(P_{g,i} \leq \lambda), \tag{2.4.1}$$

for some fixed $\lambda \in (0, 1)$, applied to all hypotheses.

Simulation Setting. The following steps were taken to simulate values of FDR and power for the aforementioned procedures.

1. Generate $\theta_{g\cdot}$, for $g = 1, \ldots, m$, as a random sample from $\mathrm{Ber}(1 - \pi_{\cdot})$;

2. For each g such that $\theta_{g\cdot} = 1$, generate $\boldsymbol{\theta}_{\cdot|g} = (\theta_{1|g}, \ldots, \theta_{n|g})$ as a random vector of n i.i.d. $\mathrm{Ber}(1 - \pi)$;

3. Given $(\theta_{g\cdot}, \boldsymbol{\theta}_{\cdot|g})$, $g = 1, \ldots, m$, generate m independent n-dimensional random vectors $\mathbf{X}_g = (X_{g,1} \ldots, X_{g,n})$, $g = 1, \ldots, m$, as follows:

$$\mathbf{X}_g = \mu\theta_{g\cdot}\boldsymbol{\theta}_{\cdot|g} + \sqrt{(1-\rho_g)}\mathbf{Z}_g + \sqrt{\rho_g}Z_{g0},$$

   for some $0 \leq \rho_g < 1$, having generated $\{Z_{g0}, \mathbf{Z}_g = (Z_{g1}, \ldots, Z_{gn})^T\}$ as a random vector of $n+1$ i.i.d. $N(0,1)$ samples, for $g = 1, \ldots, m$.

4. Apply each procedure at FDR level $\alpha = 0.05$ for testing $H_{g,i} : \mathrm{E}(X_{g,i}) = 0$ against $K_{g,i} : \mathrm{E}(X_{g,i}) > 0$, simultaneously for all $g = 1, \ldots, m, i = 1, \ldots, n$, in terms of the corresponding p-values, and note the proportions of false rejections among all rejections and correct rejections among all false nulls.

5. Repeat Steps 1-4 200 times to simulate the values of FDR and power for each procedure by averaging out the corresponding proportions obtained in Step 4.

Remark 6. Our modeling of $\mathrm{E}(\mathbf{X}_g)$ in term of $(\theta_{g\cdot}, \boldsymbol{\theta}_{\cdot|g})$ allowed us to split the state of each hypothesis at two levels, group and individual, enabling us to regulate the density of signals in the entire set of hypotheses and disperse it evenly into the significant groups by using the following representation of true nulls among all hypotheses:

$$\pi_0 = 1 - (1 - \pi_\cdot)(1 - \pi) \ . \tag{2.4.2}$$

Simulation Findings. In each of the simulation scenario created using the above simulation settings, we considered the hypotheses to be arranged in $m = 50$ groups, each group containing $n_g = 100$ member hypotheses. Our two main objectives regarding the performance of our proposed Data-Adaptive One-Way GBH are - (i) to investigate how well it performs among all four procedures under independence when all of them are theoretically known to control FDR, and (ii) to investigate if it can possibly control FDR under PRDS in view of the fact that such control is yet to be theoretically proved.

Figures 2.1-2.2 display the findings of the first type of investigation, with $\lambda = 0.5$. Figure 2.1 considers situations where signals are distributed evenly across all groups. As seen from this figure, our proposed procedure performs better than the LSL and TST GBH procedures. It controls FDR less conservatively and is more powerful than its counterparts at all levels of density of true signals. However, its performance is quite similar to the adaptive BH procedure owing to the signals being uniformly distributed across all groups. Since signals may potentially be non-uniformly distributed across the groups, we considered a scenario where only half the groups may contain significant members; see Figure 2.2. Our proposed procedure is seen to be remarkably more powerful in this case than the other methods, and this has an edge over the others on capturing the structural information.

Figure 2.3 displays the findings of the second type of investigation. As seen from it, our proposed procedure can potentially control FDR in scenarios where concentration of signals is high and for certain choices of λ, preferably $< \alpha$. A few such scenarios with varying density of signals uniformly distributed in all groups and choices of $\lambda < \alpha$ have been shown in this figure.

Some additional simulations in Appendix A.2 discuss the effect of the parameters $\pi_{\cdot}$, π and λ on the variations caused in FDR and power of the data-adaptive One-way GBH and the adaptive BH procedure.

2.4.2 Two-Way Classified Hypotheses - One Hypothesis Per Cell

This section presents results associated with our simulation study that focused on investigating the performances of our proposed (i) Oracle Two-Way GBH_1 procedure (Theorem 2.3) against the usual single-group BH procedure and the p-filter algorithm (*Foygel Barber and Ramdas* (2015), *Ramdas et al.* (2017)) in their oracle forms, and (ii) Data-Adaptive Two-Way GBH_1 procedure (Theorem 2.4) against the Naive Data-

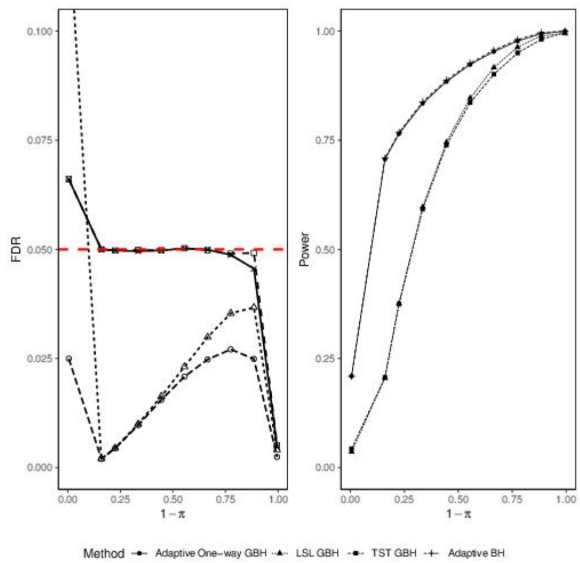

Figure 2.1: FDR and Power comparisons of the Data-Adaptive One-Way GBH proposed with other methods, under independence, $(m = 50, n = 100, \rho = 0, \pi_{.} = 0, \pi)$

.

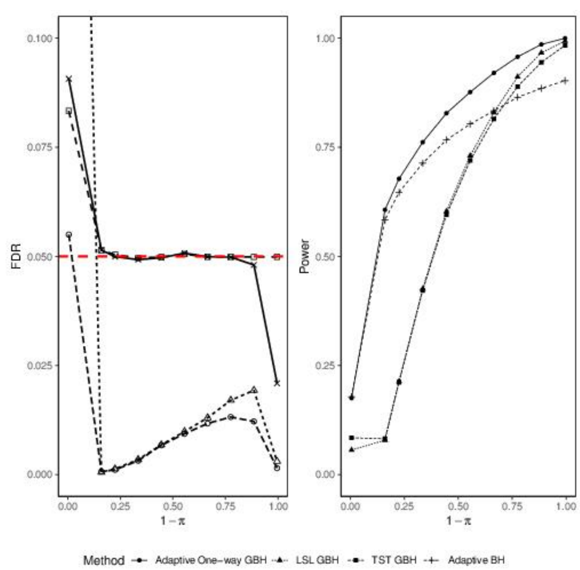

Figure 2.2: FDR and Power comparisons of the Data-Adaptive One-Way GBH with other methods applied to independent one-way classified hypotheses when true signals are unevenly distributed $(m = 50, n = 100, \rho = 0, \pi_{.} = 0.5, \pi)$

.

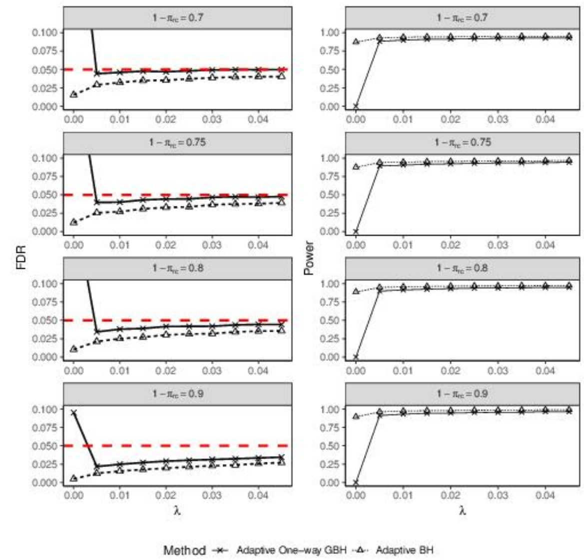

Figure 2.3: Comparison of the Data-Adaptive One-Way GBH with the naive Adaptive BH method under PRDS condition for varying choices of $0 < \lambda < 0.05 (= \alpha)$ $(m = 50, n = 100, \rho = 0.3, \pi_{\cdot} = 0, \pi)$

.

Adaptive BH in terms of FDR control and power under normal distributional settings.

Simulation Setting. The simulation setting here is a natural extension of that in the above section. More specifically, it consists of the following steps:

1. Generate $\boldsymbol{\Theta}_{mn}$ as an $m \times n$ random matrix of i.i.d. Ber$(1-\pi_{rc})$, $\boldsymbol{\theta}_m$ as a random vector of m i.i.d. Ber$(1-\pi_r)$, and $\boldsymbol{\theta}_n$ as a random vector of n i.i.d. Ber$(1-\pi_c)$;

2. Obtain

$$\boldsymbol{\Theta} = \boldsymbol{\Theta}_{mn} \star \boldsymbol{\theta}_m \mathbf{1}_n^T \star \mathbf{1}_m \boldsymbol{\theta}_n^T, \tag{2.4.3}$$

(with $A \star B$ denoting the Hadamard product between matrices A and B, and $\mathbf{1}_a$ representing the a-dimensional vector of 1's),

3. Given $\mathbf{\Theta}$, generate a random $m \times n$ matrix $\mathbf{X} = ((X_{gh}))$ as follows:

$$\begin{aligned}\mathbf{X} = \quad & \mu\mathbf{\Theta} + \sqrt{(1-\rho_r)(1-\rho_c)}\mathbf{Z}_{mn} + \sqrt{(1-\rho_r)\rho_c}\mathbf{Z}_m\mathbf{1}_n^T + \\ & \sqrt{\rho_r(1-\rho_c)}\mathbf{1}_m\mathbf{Z}_n^T + \sqrt{\rho_r\rho_c}Z_0\mathbf{1}_m\mathbf{1}_n^T,\end{aligned}$$

   having generated $\mathbf{Z}_{mn}$ as $m \times n$ random matrix, $\mathbf{Z}_m$ as m-dimensional random vector, and $\mathbf{Z}_n$ as n-dimensional random vector, each comprising i.i.d. $N(0,1)$, and Z_0 as an additional $N(0,1)$ random variable.

4. Apply each procedure at FDR level $\alpha = 0.05$ for testing $H_{gh} : \mathrm{E}(X_{gh}) = 0$ against $K_{gh} : \mathrm{E}(X_{gh}) > 0$, simultaneously for all $g = 1, \ldots, m, h = 1, \ldots, n$, in terms of the corresponding p-values, and note the proportions of false rejections among all rejections and correct rejections among all false nulls.

5. Repeat Steps 1-4 200 times to simulate the values of FDR and power for each procedure by averaging out the corresponding proportions obtained in Step 4.

Remark 7. Note that

$$\mathbf{vec}(\mathbf{X}) \sim \mathrm{N}_{mn}(\mathbf{vec}(\mu\mathbf{\Theta}), \Sigma_c \otimes \Sigma_r),$$

where $\Sigma_r = (1-\rho_r)I_n + \rho_r\mathbf{1}_n\mathbf{1}_n^T,\ \rho_r \in [0,1)$, and $\Sigma_c = (1-\rho_c)I_m + \rho_c\mathbf{1}_m\mathbf{1}_m^T,\ \rho_c \in [0,1)$. Thus, the test statistics are allowed to have different types of dependence structure by appropriately setting the value of ρ_r and/or ρ_c at 0.

Also, as seen from equation (2.4.3), the hidden state of each row and each column in terms of being significant or not has been factored into that of the hypothesis lying at their intersection. This enables us to incorporate the true effect of the underlying two-way classification structure into our simulation study. Specifically, we can regulate the density of signals in the entire matrix using the following

$$\pi_0 = 1 - (1 - \pi_{rc})(1 - \pi_r)(1 - \pi_c),$$

representing the proportion of true nulls in the entire set of mn hypotheses in terms of the proportions of rows $(1 - \pi_r)$ and columns $(1 - \pi_c)$ containing signals.

Simulation Findings. We fixed $m = 50$, $n = 100$, $\mu = 0$ for true null hypotheses, and $= 3$ for true signals.

Comparison of Oracle Procedures: We investigated the performance of the Oracle Two-Way GBH_1 under independence as well as under PRDS compared to the other oracle procedures being considered - one in terms of identifying signals and the other in terms of FDR control and power. The findings of these are displayed in Figures 2.4-2.6.

For the first type of investigation, in the 50×100 matrix, we arranged the significant hypotheses in two 15×15 blocks and along the diagonal of another 15×15 block, as shown in Figure 2.4(a). This arrangement helps to analyze the performance of a multiple testing procedure when the signals are dense (in the two blocks) as well as when they are sparse (along the diagonal). The performance of each method based on one trial is shown in the remaining plots in Figure 2.4, with that being shown in Figures 2.4(b)-2.4(d) for the independence case and in Figures 2.4(e) -2.4(g) in the PRDS case. In either case, the proposed Oracle Two-Way GBH_1 procedure is seen to be successful in identifying maximum number of clustered signals, and almost equally efficient as the BH procedure when the signals are sparse. The performances of the p-filter algorithm and the BH procedure are comparable. The BH better identifies sparse signals, although under independence. It makes marginally higher number of false rejections than the p-filter process.

For the second type of investigation, we varied the density of true signals in

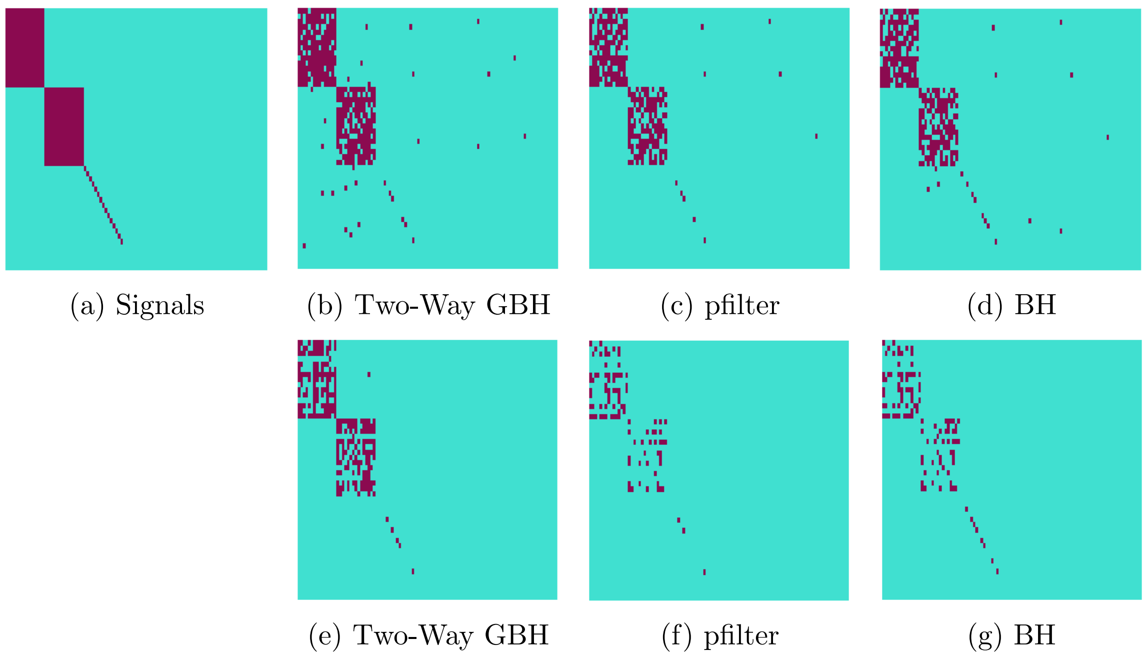

Figure 2.4: Comparison of the oracle Two-Way GBH$_1$, with other methods for one trial. (2.4a) shows the layout of the significant hypotheses. (2.4b), (2.4c) and (2.4d) show the performances of the proposed Two-Way GBH$_1$, the p-filter process and the BH procedure if the hypotheses are independent. (2.4e), (2.4f) and (2.4g) show the performances of these methods for the same setup, respectively, if there is positive dependence among the hypotheses. We choose $\rho_r = 0.3$ and $\rho_c = 0.4$ for the case of dependent hypotheses

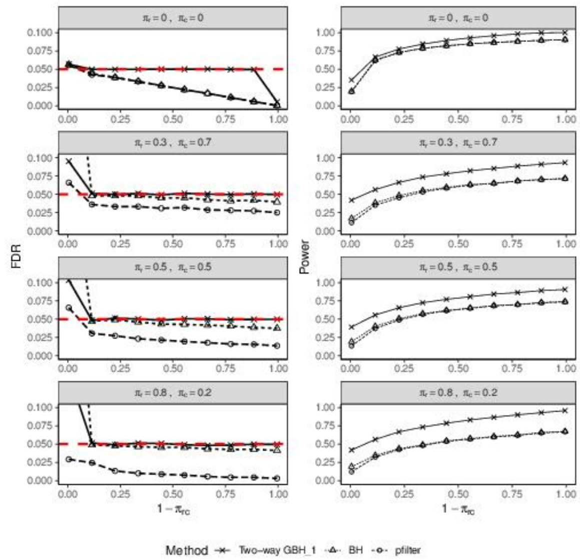

Figure 2.5: Comparison of the oracle Two-Way GBH_1 procedure with other methods, under independence. Set of parameters used is $(m = 50, n = 100, \rho_r = 0, \rho_c = 0, \pi_r, \pi_c, \pi_{rc})$.

the 50×100 blocks. The results are displayed in Figure 2.5 for the independent case and in Figure 2.6 for the PRDS case. Our proposed Oracle Two-Way GBH_1 procedure performs better than either p-filter algorithm or the BH procedure in terms of both FDR control and power. Performances of the p-filter process and the BH procedure are comparable. As the density of true signals increases, the proposed method maintains control on FDR at level α and is more powerful than the other two procedures.

Comparison of Data-Adaptive Procedures: Here, our focus had been to investigate the following two questions regarding performance of our proposed Data Adaptive Two-Way GBH_1 in Theorem 2.4 compared to its natural competitor, which is Naive Data-Adaptive BH: (i) How well it performs under independence when both are theoretically known to control FDR? (ii) If it can possibly control FDR under PRDS in view of the fact that such control is yet to be theoretically proved for both of these procedures.

Figures 2.7 and 2.8 display the findings of these investigation. Figure 2.7, which

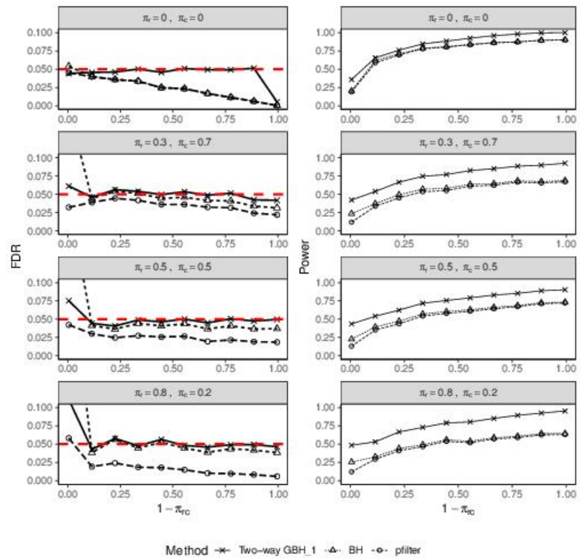

Figure 2.6: Comparison of the oracle Two-Way GBH_1 procedure for hypotheses with PRDS property, with other methods. Set of parameters used is $(m = 50, n = 100, \rho_r = 0.3, \rho_c = 0.4, \pi_r, \pi_c, \pi_{rc})$.

summarizes the results associated with answering question (i) (with $\lambda = 0.5$), indicates that, though both these methods have comparable power when the signals are uniformly distributed in all rows and columns (i.e., when $\pi_r = \pi_c = 0$), our proposed method seems significantly more powerful in these situations. We chose $\rho_r = 0.3$ and $\rho_c = 0.4$ to answer question (ii), with the related findings being summarized in Figure 2.8. It shows that the proposed Data-Adaptive Two-Way GBH_1 possibly can control FDR under PRDS when there is a high density of signals across all row and column groups; however, the choice of λ seems crucial in such situations, and its values should be chosen in the range $(0, \alpha)$.

Some additional simulations, comparing the determining effects of the parameters in terms of FDR and power for the data-adaptive Two-Way GBH_1 and the BH procedure are provided in the Appendix A.2.

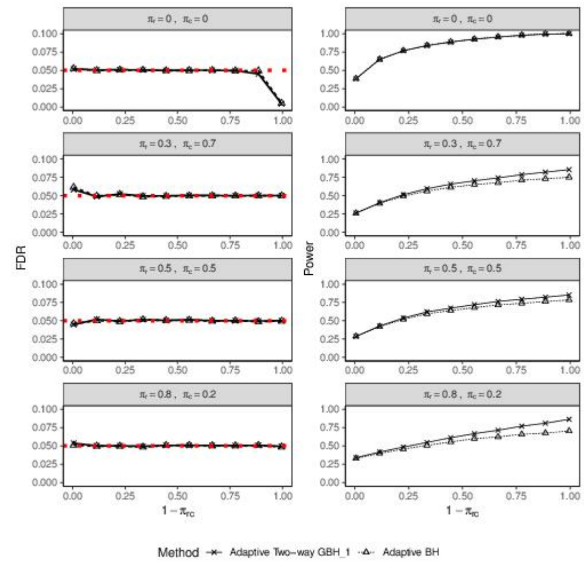

Figure 2.7: Comparison of the data-adaptive Two-Way GBH_1 procedure with the naive Adaptive BH method, under independence. Set of parameters used is $(m = 50, n = 100, \rho_r = 0, \rho_c = 0, \pi_r, \pi_c, \pi_{rc})$

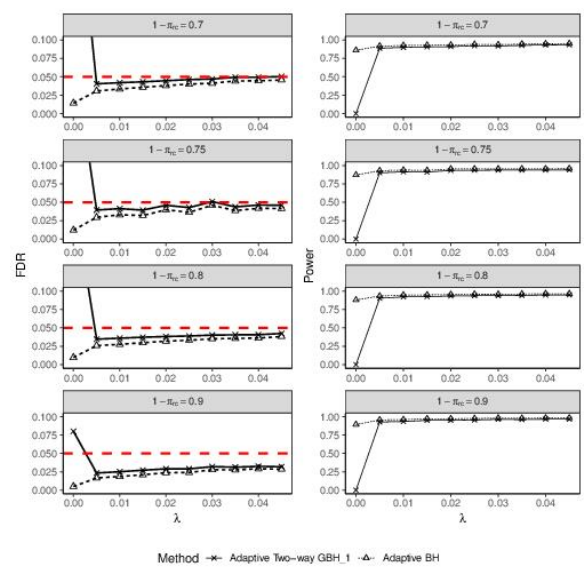

Figure 2.8: Comparison of the data-adaptive Two-Way GBH_1 procedure for hypotheses with PRDS property, with the naive Adaptive BH procedure, for varying choices of $0 < \lambda < 0.05 (= \alpha)$. Set of parameters used is $(m = 50, n = 100, \rho_r = 0.3, \rho_c = 0.4, \pi_r = 0, \pi_c = 0, \pi_{rc})$

2.4.3 Two-way Classified Hypotheses - Multiple hypotheses at each intersection

We compare the performances of (i) Oracle Two-Way GBH$_{>1}$ procedure (Theorem 2.5) with the single-group BH procedure in its oracle form and (ii) Data-Adaptive Two-Way GBH$_{>1}$ (Theorem 2.6) against the naive Adaptive BH in terms of FDR control and power under normal distributional settings.

Simulation Setting. We considered the case where $n_{gh} = p$ for all (g, h), so that our data generating process had to be designed to produce a random pair of third order tensors of dimension $m \times n \times p$, $(\mathbf{X}, \mathbf{\Theta})$, consisting of normally distributed test statistics and the Bernoulli hidden states of the corresponding hypotheses, respectively. The following were the steps in that process:

1. Generate $\mathbf{\Theta}_{mnp}$ as an $m \times n \times p$ dimensional random tensor of i.i.d. Ber$(1-\pi_{rc})$, $\boldsymbol{\theta}_m$ as a random vector of m i.i.d. Ber$(1-\pi_r)$, and $\boldsymbol{\theta}_n$ as a random vector of n i.i.d. Ber$(1-\pi_c)$.

2. Obtain

$$\boldsymbol{\theta} = \boldsymbol{\theta}_{mnp} \star (\boldsymbol{\theta}_m \circ \mathbf{1}_n \circ \mathbf{1}_p) \star (\mathbf{1}_m \circ \boldsymbol{\theta}_n \circ \mathbf{1}_p)$$

   (with $\boldsymbol{a} \circ \boldsymbol{b}$ denoting the outer product between the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$);

3. Given $\mathbf{\Theta}$, generate $\mathbf{X}$ as an $m \times n \times p$ dimensional tensor having a tensor normal distribution given below using its vectored form:

$$\mathbf{vec}(\mathbf{X}) \sim \mathrm{N}_{mnp}(\mathbf{vec}(\mu\boldsymbol{\theta}), \Sigma_p \otimes \Sigma_c \otimes \Sigma_r),$$

   where $\Sigma_r = (1-\rho_r)I_n + \rho_r \mathbf{1}_n \mathbf{1}_n^T,\ \rho_r \in [0,1)$, $\Sigma_c = (1-\rho_c)I_m + \rho_c \mathbf{1}_m \mathbf{1}_m^T,\ \rho_c \in [0,1)$, and $\Sigma_p = (1-\rho_p)I_p + \rho_p \mathbf{1}_p \mathbf{1}_p^T,\ \rho_p \in [0,1)$.

More specifically, the sample matrix is drawn from the following mixture distribution.

$$\begin{aligned}\mathbf{X} =& \mu\boldsymbol{\theta} + \sqrt{(1-\rho_r)(1-\rho_c)(1-\rho_p)}\mathbf{Z}_{m\times n\times p} + \sqrt{\rho_p(1-\rho_r)(1-\rho_c)}\mathbf{Z}_{m\times n\times 1} \times_3 \mathbf{1}_p \\ &+ \sqrt{\rho_c(1-\rho_r)(1-\rho_p)}\mathbf{Z}_{m\times 1\times p} \times_2 \mathbf{1}_n + \sqrt{\rho_r(1-\rho_c)(1-\rho_p)}\mathbf{Z}_{1\times n\times p} \times_1 \mathbf{1}_m \\ &+ \sqrt{\rho_c\rho_p(1-\rho_r)}(Z_{m\times 1\times 1} \times_2 \mathbf{1}_n) \times_3 \mathbf{1}_p + \sqrt{\rho_r\rho_p(1-\rho_c)}(Z_{1\times n\times 1} \times_1 \mathbf{1}_m) \times_3 \mathbf{1}_p \\ &+ \sqrt{\rho_r\rho_c(1-\rho_p)}(Z_{1\times 1\times p} \times_1 \mathbf{1}_m) \times_2 \mathbf{1}_n + \sqrt{\rho_r\rho_c\rho_p}Z_0\mathbf{1}_{m\times n\times p} \end{aligned} \tag{2.4.4}$$

Here $\mathbf{Z}_{m\times n\times p}$ is an $m \times n \times p$-dimensional array(tensor) consisting of mnp i.i.d. standard normal samples. $\mathbf{Z}_{m\times n\times 1}$, $\mathbf{Z}_{m\times 1\times p}$ and $\mathbf{Z}_{1\times n\times p}$ are second order tensors (matrices) comprising of mn, mp and np i.i.d standard normal samples respectively. $\mathbf{Z}_{m\times 1\times 1}$, $\mathbf{Z}_{1\times n\times 1}$ and $\mathbf{Z}_{1\times 1\times p}$ are first order tensors, or vectors, each comprising of m, n and p i.i.d. standard normal samples and Z_0 is an additional standard normal sample, which we multiply to a third order tensor $\mathbf{1}_{m\times n\times p}$ of dimensions $m \times n \times p$ with each element 1.

We use several instances of N-mode product of tensors with matrices (definition in *Kolda and Bader* (2009)), symbolized by $\mathbf{Z}_{m\times n\times p} \times_N \mathbf{A}_{a\times m}$ to replicate the generated samples over one or two dimensions. For example, the $1-$mode product $\mathbf{Z}_{m\times n\times p} \times_1 \mathbf{A}_{a\times m}$ results in a tensor of dimensions $a \times n \times p$.

Let X_{ghk} be the kth layer test statistic in the (g, h) cell. They can have different types of positive dependence structures determined through appropriate choices of the correlation coefficients ρ_r ρ_c and ρ_p. If there is independence along any dimension of the tensor $\mathbf{X}$, the corresponding correlation coefficient is set to 0. We considered the problem of testing $H_{gh} : \mathrm{E}(X_{ghk}) = 0$ against $K_{ghk} : \mathrm{E}(X_{ghk}) > 0$, simultaneously for all $g = 1, \ldots, m, h = 1, \ldots, n, k = 1, \ldots, p$. So, the next two steps in our simulation study were the following:

4. Apply each of the aforementioned procedures at FDR level $\alpha = 0.05$, and note

down each of the the proportions of false rejections among all rejections and correct rejections among all false nulls.

5. Repeat Steps 1-4 200 times to simulate the values of FDR and power for each procedure by averaging out the corresponding proportions noted in Step 4.

Simulation Findings. We considered fixed $m = 50$, $n = 100$, $p = 10$, and set μ at 0 for true null hypotheses and at $= 3$ for all true signals. The rest of the parameters are regulated to generate different situations and analyze the performance of our method in those settings. The combination of parameters (π_r, π_c) chosen are similar to those in the case of two-way classification with one hypothesis per cell. For each combination of values for (π_r, π_c), $1 - \pi_{rc}$ was varied between 0 and 1. Signals are sparse for smaller values of $1 - \pi_{rc}$ and the density increases with its value.

Comparison of Oracle Procedures: We wanted to make two types of investigation for Oracle Two-Way $\text{GBH}_{>1}$ in Theorem 2.5 under both independence and PRDS condition against the usual single group BH - how does it perform in terms of FDR control and power? The findings of these are displayed in Figures 2.9 (for the independent case) and 2.10 (for the PRDS case corresponding to $\rho_r = 0.3$, $\rho_c = 0.4$, and $\rho_p = 0.2$). In either case, the proposed method controls FDR, as expected, and seems to be powerful than the BH.

Comparison of Data-Adaptive Procedures: As before, our focus in this case was to investigate the following two questions regarding the performance of our proposed Data Adaptive Two-Way $\text{GBH}_{>1}$ in Theorem 2.6 compared to naive Adaptive BH: (i) How well it performs under independence when both are theoretically known to control FDR? (ii) Can it can possibly control FDR under PRDS in view of the fact that such control is yet to be theoretically proved for both of these procedures? Figures 2.11 and 2.12 display the findings of these investigation. Figure 2.11, which summarizes the results associated with answering question 1 (with $\lambda = 0.5$), indicates that both

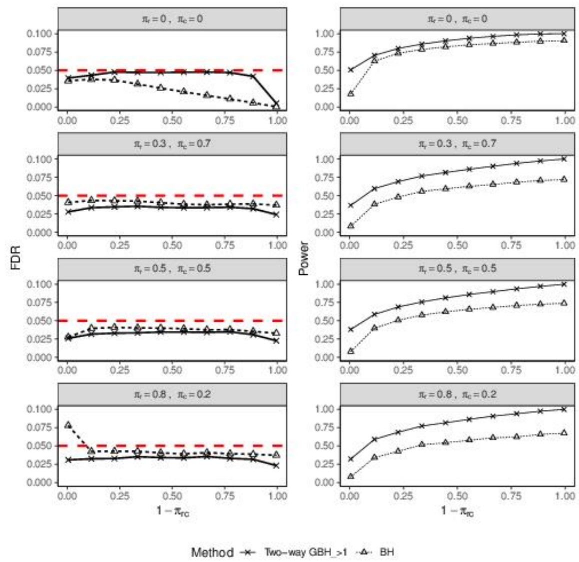

Figure 2.9: Comparison of the oracle Two-Way GBH$_{>1}$ procedure with the BH method, applied to independent hypotheses. Set of parameters used is $(m = 50, n = 100, p = 10, \rho_r = 0, \rho_c = 0, \rho_p = 0, \pi_r, \pi_c, \pi_{rc})$

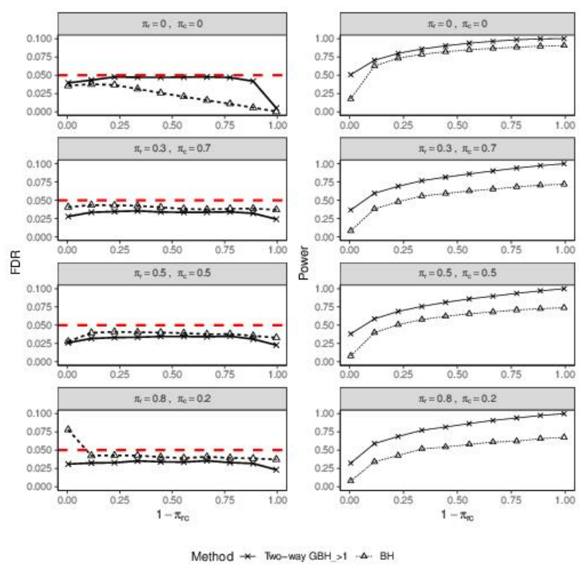

Figure 2.10: Comparison of the oracle Two-Way GBH$_{>1}$ procedure with the BH method, when the hypotheses satisfy PRDS condition. Set of parameters used is $(m = 50, n = 100, p = 10, \rho_r = 0.3, \rho_c = 0.4, \rho_p = 0.2, \pi_r, \pi_c, \pi_{rc})$

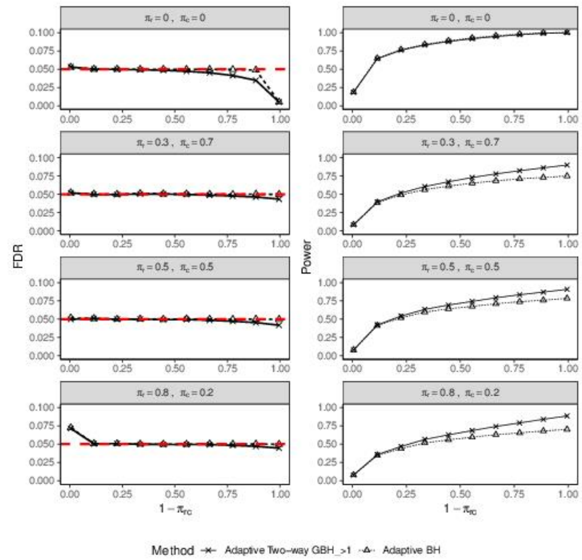

Figure 2.11: Comparison of the data-adaptive Two-Way $\text{GBH}_{>1}$ procedure with the naive Adaptive BH method, when the hypotheses are independent. Set of parameters used is $(m = 50, n = 100, p = 10, \rho_r = 0, \rho_c = 0, \rho_p = 0, \pi_r, \pi_c, \pi_{rc})$

methods have similar performance when the signals are uniformly distributed over the $m \times n$ grid (which occurs when $\pi_r = \pi_c = 0$). However, our proposed method is more powerful when the signals are not uniformly distributed, which is displayed for the other combinations of the (π_r, π_c) values. Figure 2.12 says, as in the case of two-way classification with one hypothesis per cell, our proposed Data-Adaptive Two-Way $\text{GBH}_{>1}$ can possibly control FDR under PRDS with choices of $\lambda < \alpha$, with a few instances being shown in the figure, when there is a high density of signals across all row and column groups.

2.5 Application to Microbiome Data

We apply our two-way classified method to a microbial abundance dataset to illustrate its application in real scientific problems. We consider the `GlobalPatterns` dataset available through the Bioconductor package `phyloseq`. The data was first studied in *Caporaso et al.* (2011) to analyze prevalence of microbial communities in different environments. The data consists of 19216 microbes identified by

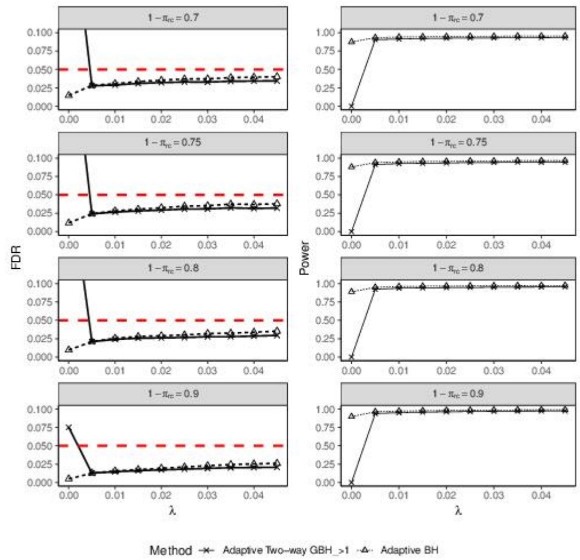

Figure 2.12: Comparison of the data-adaptive Two-Way $\mathrm{GBH}_{>1}$ procedure and the naive Adaptive BH procedure, applied to hypotheses with PRDS property, for varying choices of $0 < \lambda < 0.05(= \alpha)$. Set of parameters used is $(m = 50, n = 100, p = 10, \rho_r = 0.3, \rho_c = 0.4, \rho_p = 0.2, \pi_r = 0, \pi_c = 0, \pi_{rc})$

their Operational Taxonomic (OTU) Numbers obtained from 26 samples of 9 different environments, which includes a mock environment. The environments are characterized by 7 variables. Classification of the microbes according to their 7 taxonomic ranks is provided along-with the phylogenetic tree describing the relationships among the microbes. The data records abundance patterns of each microbe across the nine sample environments. Since microbes closely related at the tips of the phylogenetic tree have similar characteristics, it is quite likely that they have similar abundance patterns which renders a positive dependence in the data. A smaller subset of this dataset, consisting of data on only microbes specific to the Chlamydiae bacteria taxon, was studied by *Sankaran and Holmes* (2014). For their analysis, they classified 21 microbes into four groups formed according to their taxonomic families and invoked the Grouped BH procedure as suggested by *Hu et al.* (2010) to find which particular microbes are significantly abundant across the environments.

We perform a similar analysis on a larger scale on the entire GlobalPatterns dataset. In contrast to the analysis provided in *Sankaran and Holmes* (2014), we consider the p-values to be in a two-way classified structure. Each p-value corresponds to a particular microbe and an environment. The p-value P_{ij}, corresponding to the ith microbe and jth environment answers the question "Is the ith microbe abundantly present in the jth environment?" Considering the microbes as individual groups furnishes $m = 19216$ groups and together with $n = 9$ environments we obtain a two way structure of dimensions 19216×9.

Instead of considering the microbes by their individual species, we classify them according to their taxonomic families. While higher taxonomic ranks such as taxonomic class, phyla, etc. can also be utilized for classification of the microbes, groups formed as such are larger and members have wider variety of characteristics rendering the effect due to grouping vaguer. After adjusting for missing values and removing hypotheses with missing family labels, we obtain $N = 120942$ hypotheses classified into a grid of $m = 334$ families along rows and $n = 9$ environments along columns. Since there are unequal number of members (min: 1 and max: 1658) in each family, we use the data-adaptive method for two-way classified hypotheses with unequal number of hypotheses in each cell with weights as mentioned in expression (2.3.9). The method identified 7584 hypotheses as significant. In comparison, the adaptive BH procedure, applied to the entire set of hypotheses identified 7377 hypotheses as significant.

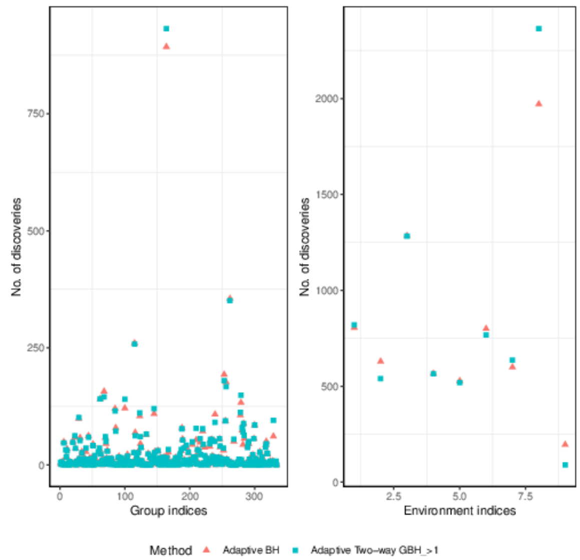

Figure 2.13: Comparison of the data-adaptive Two-way $\mathrm{GBH}_{>1}$ with the adaptive BH procedure, when applied to the microbiome dataset.

CHAPTER 3

Lfdr BASED TWO-WAY GATE PROCEDURE FOR TESTING TWO-WAY CLASSIFIED HYPOTHESES

3.1 Literature Review

Efron et al. (2001) introduced the concept of Local FDR as a Bayesian counterpart of the p-value. Consider a hypothesis $H_{\cdot}$, the state of which is represented by the binary random variable $\theta_{\cdot} = 0$ (or $= 1$) according as $H_{\cdot}$ is insignificant or otherwise. With X being the corresponding test statistic or p-value, and $\theta_{\cdot} \sim \text{Ber}(\pi)$, let X follow the two class mixture model given as

$$f(x) = (1 - \pi)f_0(x) + \pi f_1(x),$$

where f_0 and f_1 are the distributions of X under the true and false hypothesis respectively. The local FDR is defined as the posterior probability that $\theta_{\cdot} = 0$, given $X = x$, and is given by

$$\text{Lfdr}(x, \pi) = \Pr(\theta_{\cdot} = 0 | X = x) = \frac{(1 - \pi)f_0(x)}{f(x)}.$$

Lfdr based multiple testing methods that integrate Bayesian and decision theoretic concepts to test a single group of hypotheses have been suggested by *Efron* (2008), *Sun and Cai* (2007), etc. The optimal testing procedure in such cases target control of marginal FDR and minimize a form of marginal False Negative Rate (FNR). Use of external information have also been used in some research to influence the performance of multiple testing procedures. This is usually achieved either through formulation of weights utilized in multiple testing procedures, or through modification of the Lfdr formula and subsequent testing procedures to incorporate the additional information. We can find such instances in *Ferkingstad et al.* (2008), *Zablocki et al.* (2014), *Basu et al.* (2018), etc. *Ferkingstad et al.* (2008) used the covariates to classify the p-values into groups before invoking Lfdr based testing. This approach enabled the test to be more powerful. *Cai and Sun* (2009) extended their previous work in *Sun and Cai* (2007) to the one-way classification framework. They considered the two-class mixture model to be different in each group and suggested two methods to control the overall proportion of false discoveries made. First is the Pooled Lfdr method, that combines the null and non-null proportions across all hypotheses and compute the Lfdr values which are then pooled together to control the global fdr. The second method utilizes the Lfdr values (refered to as Conditional Lfdr (CLfdr)) for each hypothesis taking into account the two-class mixture model unique to the group it belongs to. The methods are optimal (Clfdr based method being asymptotically optimal) as they minimize some form of marginal False Negative Rate across hypotheses in all groups. Their data-driven versions were also proposed, that asymptotically satisfy all properties of the oracle procedures. Similar ideas were used alongwith external covariates to address different problems of microarray time course experiments in *Sun and Wei* (2011), spatio-temporal data in *Sun et al.* (2015), etc.

To the best of our knowledge, ours is the first attempt to use the Lfdr based approach for simultaneous testing of two way classified hypotheses. As mentioned in the introduction, we extend the idea of Lfdr based testing presented in *Liu et al.* (2016) and *Sarkar and Zhao* (2017) from one-way classified hypotheses to two-way classified hypotheses. In order to do the same, we first provide a brief review of the one-way classification model and the testing problem presented in these articles.

3.2 Lfdr Based Multiple Testing of One-way Classified Hypotheses: Review

3.2.1 Model Description

Suppose N hypotheses are classified into m non-overlapping groups formed due to one classification criterion. Considering n_g hypotheses $H_{g1}, \ldots, H_{gn_g}$ to be members of the gth group, the group is considered significant only if it contains at least one significant member. Consequently, we obtain a hypothesis $H_g = \cap_{h=1}^{n_g} H_{gh}$ for the gth group, $g = 1, \ldots, m$. The state of the group level hypotheses are denoted by the binary variable $\theta_{g.} = 0(\text{or} = 1)$ depending on truth or falsity of the hypothesis. The state of each individual hypothesis θ_{gh} can then be expressed as the product of the state of its parent group and its own hidden state, i.e.,

$$\theta_{gh} = \theta_{g.}\theta_{h|g},$$

where $\theta_{h|g} = \theta_{gh}/\max\{\theta_{g.}, 1\}$. This representation yields the following group-adjusted two class mixture model: hidden states are determined by the state of the parent group. The model for the setup is defined as follows.

Definition 2. *Consider X_{gh} is the random observation corresponding to H_{gh}. Let*

$(X_{gh}, \theta_{g\cdot}, \theta_{h|g}, h = 1, \ldots, n_g)$ *be the set of random variables for the gth group. Then,*

$$X_{gh}|\theta_{g\cdot}, \theta_{h|g} = (1-\theta_{g\cdot})\theta_{h|g} f_0(x_{gh}) + \theta_{g\cdot}\theta_{h|g} f_1(x_{gh}), f_0 \text{ and } f_1$$

being the distributions under null and alternate respectively.

$$\{\theta_{h|g}, \forall h|\theta_{g\cdot} = 0\} = 0 \text{ with probability 1.}$$

$$\{\theta_{h|g}, \forall h = 1, \ldots, n_g|\theta_{g\cdot} = 1\} \sim TBern_{n_g}(\pi_{2|1})$$

$$\theta_{g\cdot} \sim \mathrm{Ber}(\pi_{\cdot})$$

with $\mathrm{TBern}_n(\pi)$, for any $0 < \pi < 1$ being defined as follows:

Definition 3. *A set of independent and identical random variables $Z_1, \ldots, Z_n$ such that $Z_i \sim Ber(\pi), \forall i$, then they are jointly said to follow a Truncated Product Bernoulli Distribution $TBern_n(\pi)$ if their joint distribution is given as*

$$f(z_1, \ldots, z_n) \sim \frac{1}{1-(1-\pi)^n} \prod_{i=1}^{n} \pi^{z_i}(1-\pi)^{(1-z_i)} I\left(\sum_i z_i > 0.\right)$$

This distribution is an adjustment of that assumed for the hidden states of the hypotheses in the aforementioned two class mixture model, given that at least one of these hypotheses is significant.

3.2.2 Error Rates and Testing Procedures

Let

$$\mathrm{Lfdr}_{gh}(\boldsymbol{\pi}) = \mathrm{Lfdr}_{gh}(\mathbf{X}, \boldsymbol{\pi}) = \Pr(\theta_{gh} = 0|\mathbf{X})$$

$$\mathrm{Lfdr}_{g\cdot}(\boldsymbol{\pi}) = \mathrm{Lfdr}_{g\cdot}(\mathbf{X}, \boldsymbol{\pi}) = \Pr(\theta_{g\cdot} = 0|\mathbf{X})$$

$$\mathrm{Lfdr}_{h|g}(\boldsymbol{\pi}) = \mathrm{Lfdr}_{h|g}(\mathbf{X}, \boldsymbol{\pi}) = \Pr(\theta_{h|g} = 0|\theta_{g\cdot} = 1, \mathbf{X})$$

The error rate of interest is the Posterior Total FDR as given by

$$\text{PFDR}_T = E\left[\left.\frac{\sum_{g=1}^{m}\sum_{h=1}^{n_g}(1-\theta_{gh})\delta_{gh}(\mathbf{X})}{\sum_{g=1}^{m}\sum_{h=1}^{n_g}\delta_{gh}(\mathbf{X}) \vee 1}\right| \mathbf{X}\right] = \frac{\sum_{g=1}^{m}\sum_{h=1}^{n_g}\delta_{gh}(\mathbf{X})\text{Lfdr}_{gh}(\boldsymbol{\pi})}{\sum_{g=1}^{m}\sum_{h=1}^{n_g}\delta_{gh}(\mathbf{X}) \vee 1},$$

where $\boldsymbol{\pi} = (\pi_{\cdot}, \pi_{2|1})$ and δ_{gh} is the decision rule related to the hth hypothesis in the gth group. Like θ_{gh}, δ_{gh} can also be split into product of decision rules corresponding to parent group ($\delta_{g\cdot}$) and the hypothesis, conditional on the decision made for its parent, ($\delta_{h|g}$), i.e.,

$$\delta_{gh} = \delta_{g\cdot} \cdot \delta_{h|g} \;\forall\; g, h,$$

where $\delta_{g|h} = \delta_{gh}/\max(\delta_{g\cdot}, 1)$.

It can be shown that

$$\text{Lfdr}_{gh}(\boldsymbol{\pi}) = 1 - (1 - \text{Lfdr}_{g\cdot}(\boldsymbol{\pi}))(1 - \text{PFDR}_{W_g}(\boldsymbol{\pi})) \;\forall\; g, h,$$

where $\text{PFDR}_{W_g} = \sum_{h=1}^{n_g}\delta_{h|g}\text{Lfdr}_{h|g}/\max(\delta_{g\cdot}\sum_{h=1}^{n_g}\delta_{h|g}, 1)$, i.e., the posterior FDR within group g. The expressions of PFDR_T can be modified accordingly. This expression of $\text{Lfdr}_{gh}(\boldsymbol{\pi})$ facilitates designing two-stage multiple testing procedures that can identify the significant groups and individual hypotheses in two separate steps and can thus maintain control over between-group and within-group posterior FDR (PFDR_B and $\text{PFDR}_{\text{W}_\text{g}}$ respectively). Contextually, PFDR_B is given as $\sum_{g=1}^{m}\delta_{g\cdot}\text{Lfdr}_{g\cdot}/\max\{\sum_{g=1}^{m}\delta_{g\cdot}, 1\}$. The proposed testing methods to control PFDR_T at level α, when designed with decision rules $\delta_{gh}(\mathbf{X}) = I(\text{Lfdr}_{gh} \le c), c \in (0, 1)$, is optimal in the sense of possessing minimum posterior marginal FNR among all other tests satisfying $\text{PFDR}_\text{T} \le \alpha$.

With this background we are now ready to develop the model for two-way classified hypotheses in the following sections.

3.3 Two-way Classified Hypotheses: Model Description

As seen in Chapter 2, for brevity and clarity, two-way classified hypotheses can be laid out in a matrix, with rows and columns pertaining to groups formed by each of the two classification criteria. As such, the state of each hypothesis in this setup is influenced by its parent row and column. Suppose one of the grouping criteria has m groups, and the second grouping criterion has n groups, we can lay out the $N = mn$ hypotheses in an $m \times n$ matrix. The state of the parent row or column are dependent on the states of their member hypotheses. A row(column) is considered to be significant if it has at least one significant member. Thus mathematically, if H_{gh} represents the hypothesis at the cell (g, h), the hypothesis corresponding to row g is $H_{g\bullet} = \bigcap_{h=1}^{n} H_{gh}$ and that corresponding to column h is $H_{\bullet h} = \bigcap_{g=1}^{m} H_{gh}$. We can express the state of each hypothesis as $\theta_{gh} = \theta_{g\bullet} \cdot \theta_{\bullet h} \cdot \gamma_{gh}, \forall g = 1, \ldots, m,\ h = 1, \ldots, n$, where $\theta_{g\bullet} = 0(1)$ represents the state of the parent row and $\theta_{\bullet h} = 0(1)$ that of the parent column. The $\gamma_{gh} = \theta_{gh} / \max\{\theta_{g\bullet} \cdot \theta_{\bullet h}, 1\}$, denotes the state of the hypothesis H_{gh}, conditional on the states of the parent row and column. If $H_{g\bullet} \cap H_{\bullet h}$, i.e., the parent row and column are both significant, i.e., symbolically $\theta_{g\bullet} \cdot \theta_{\bullet h} = 1$, then $\gamma_{gh} = 1$ or 0 (i.e., significant or not), depending on the intrinsic state of the hypothesis H_{gh}. If at least either of the parent row or column are non-significant, i.e., $\theta_{g\bullet} \cdot \theta_{\bullet h} = 0$, logically, H_{gh} is insignificant and $\gamma_{gh} = 0$ with probability 1.

In terms of matrix notation, we write $\boldsymbol{\Theta} = ((\theta_{gh}))$ as follows:

$$\boldsymbol{\Theta} = \mathrm{Diag}(\theta_{1\bullet}, \ldots, \theta_{m\bullet})\, \boldsymbol{\Gamma}\, \mathrm{Diag}(\theta_{\bullet 1}, \ldots, \theta_{\bullet n}),$$

where $\boldsymbol{\Gamma} = ((\gamma_{gh}))$. This brings out the underlying two-way group structure of the hypotheses in terms of their binary hidden states and the states of the rows and columns to which they simultaneous belong. We describe the joint distribution of the hidden states using the Two-way Truncated Bernoulli distribution defined next.

3.3.1 Two-way Truncated Bernoulli Distribution

Let $\boldsymbol{Z} = ((Z_{gh}))$ be an $m \times n$ binary matrix, i.e., a matrix of binary random variables each taking value 0 or 1. Such a matrix is said to be significant if at least one entry in each of its rows and columns is 1.

The matrix $\boldsymbol{Z} = ((Z_{gh}))$ or its distribution is referred to as $\text{Bern}_{m\times n}(\pi)$ if Z_{gh}, for $g = 1, \ldots, m; h = 1, \ldots, n$, are i.i.d. Bernoulli($\pi$). The following two lemmas associated with $\text{Bern}_{m\times n}(\pi)$, will be helpful in understanding our model to be proposed following them.

Lemma 2. *$\text{Bern}_{r\times c}(\pi)$ is significant with the following probability:*

$$\begin{aligned}
\mathcal{P}_\pi(r, c) &= \sum_{j=0}^{c-1} (-1)^j \binom{c}{j} \{(1-\pi)^j - (1-\pi)^c\}^r \\
&= \sum_{i=0}^{r-1} (-1)^i \binom{r}{i} \{(1-\pi)^i - (1-\pi)^r\}^c \\
&= \sum_{i=0}^{r} \sum_{j=0}^{c} (-1)^{i+j} \binom{r}{i} \binom{c}{j} (1-\pi)^{rj+ci-ij}. \qquad (3.3.1)
\end{aligned}$$

Definition 4. [Two-Way Truncated Bernoulli Matrix (Two-Way $\text{TBern}_{r\times c}(\pi)$)] An $r \times c$ binary matrix $\boldsymbol{Z} = ((Z_{gh}))$ or its distribution is said to be Two-Way $\text{TBern}_{r\times c}(\pi)$ if its distribution is given by the conditional distribution of $\text{Bern}_{r\times c}(\pi)$ given that it is significant, which is

$$\frac{1}{\mathcal{P}_\pi(r, c)} \prod_{g=1}^{r} \prod_{h=1}^{c} \left\{ \pi^{Z_{gh}} (1-\pi)^{1-Z_{gh}} I\left(Z_{g\bullet} Z_{\bullet h} > 0\right) \right\}, \qquad (3.3.2)$$

with $Z_{g\bullet} = I(\sum_{h=1}^{c} Z_{gh} > 0)$ and $Z_{\bullet h} = I(\sum_{g=1}^{r} Z_{gh} > 0)$.

Lemma 3. *Let $\boldsymbol{Z}_{row} = (Z_{1\bullet}, \ldots, Z_{m\bullet})'$, $\boldsymbol{Z}_{col} = (Z_{\bullet 1}, \ldots, Z_{\bullet n})'$. Then, for $\boldsymbol{Z} = ((Z_{gh})) \sim \text{Bern}_{m\times n}(\pi)$, we have*

(i) $\Pr(\boldsymbol{Z} = \boldsymbol{0} \mid \boldsymbol{Z}_{row}, \boldsymbol{Z}_{col}) = 1$ *if* $\sum_{g=1}^{m} \sum_{h=1}^{n} Z_{g\bullet} Z_{\bullet h} = 0,$ *and*

$$\boldsymbol{Z} \mid \boldsymbol{Z}_{row}, \boldsymbol{Z}_{col} \sim \textit{Two-Way TBern}_{(\sum_{g=1}^{m} Z_{g\cdot})\times(\sum_{h=1}^{n} Z_{\cdot h})}(\pi) \textit{ if } \sum_{g=1}^{m}\sum_{h=1}^{n} Z_{g\cdot}Z_{\cdot h} > 0,$$

$$\text{(ii) } (\boldsymbol{Z}_{row}, \boldsymbol{Z}_{col}) \sim \mathcal{P}_{\pi}(\sum_{g=1}^{m} Z_{g\cdot}, \sum_{h=1}^{n} Z_{\cdot h})(1-\pi)^{mn-\sum_{g=1}^{m}\sum_{h=1}^{n} Z_{g\cdot}Z_{\cdot h}}.$$

Remark 8. *As shown in Appendix A.3, we can see that the marginal distribution of* $\boldsymbol{Z}_{row}$ *(or* $\boldsymbol{Z}_{col}$*) obtained from the joint distribution of* $(\boldsymbol{Z}_{row}, \boldsymbol{Z}_{col})$ *in Lemma 3 (ii) is indeed same as that of i.i.d.* $\text{TBern}_n(\pi)$ *(or* $\text{TBern}_m(\pi)$*).*

We are now ready to propose our model providing a framework for adapting an Lfdr based method of controlling false discoveries under the standard two-class mixture model to a two-way group structure of the hypotheses. As said before, this is brought about by considering the status of the binary hidden states of the hypotheses in terms of those of the rows and columns to which they simultaneously belong. More specifically, we consider a situation where one is more confident about making statement about the hidden states of the rows (or columns) than making such statements about the individual hypotheses and believes that the rows (or columns) are truly significant with a common probability π_1 (or π_2); i.e., our model is built under the assumption $Z_{1\cdot}, \ldots, Z_{m\cdot}$ are iid Bern(π_1) (or $Z_{\cdot 1}, \ldots, Z_{\cdot n}$ are iid Bern(π_2)), with π_1 and π_2 being such that $(1-\pi_1)^n = (1-\pi_2)^m$. It considers the standard two-class mixture model, with $\pi = 1-(1-\pi_1)^{\frac{1}{n}} = 1-(1-\pi_2)^{\frac{1}{m}}$ as the common success probability of the hidden states of the individual hypotheses, and adjusts itself as follows: The row and column hidden states are assumed to be distributed as in the standard two-class mixture model with this π (as in Lemma 3(ii)); however, the conditional distribution of the hidden states of the individual hypotheses given that they are in a significant sub-matrix of size $R \times C$, for any (R, C) such that $0 < RC \leq mn$, is Two-Way $\text{TBern}_{R\times C}(\pi^*)$, instead of Two-Way $\text{TBern}_{R\times C}(\pi)$.

Definition 5. [Row and Column Adjusted Two-Class Mixture Model for Two-Way Classified Hypotheses (Two-Way GAMM(π, π^*, m, n))]

Suppose X_{gh} be the test statistic associated to the hypothesis H_{gh}, at the intersection of gth row and hth column. This Two-Class Mixture Model for X_{gh} is defined as

(i) $X_{gh} \mid \boldsymbol{\theta}_{row}, \boldsymbol{\theta}_{col}, \boldsymbol{\Gamma} \overset{ind}{\sim} (1 - \theta_{g\boldsymbol{.}}\theta_{\boldsymbol{.}h}\gamma_{gh})f_0(x_{gh}) + \theta_{g\boldsymbol{.}}\theta_{\boldsymbol{.}h}\gamma_{gh}f_1(x_{gh})$, for some given densities f_0 and f_1,

(ii) $\Pr(\boldsymbol{\Gamma} = \mathbf{0} \mid \boldsymbol{\theta}_{row}, \boldsymbol{\theta}_{col}) = 1$ if $\sum\limits_{g=1}^{m}\sum\limits_{h=1}^{n} \theta_{g\boldsymbol{.}}\theta_{\boldsymbol{.}h} = 0,$ *and*
$\boldsymbol{\Gamma} \mid \boldsymbol{\theta}_{row}, \boldsymbol{\theta}_{col} \sim \text{Two-Way TBern}_{(\sum_{g=1}^{m} \theta_{g\boldsymbol{.}})\times(\sum_{h=1}^{n} \theta_{\boldsymbol{.}h})}(\pi^*)$ if $\sum\limits_{g=1}^{m}\sum\limits_{h=1}^{n} \theta_{g\boldsymbol{.}}\theta_{\boldsymbol{.}h} > 0,$

(iii) $(\boldsymbol{\theta}_{row}, \boldsymbol{\theta}_{col}) \sim \mathcal{P}_{\pi}(\sum\limits_{g=1}^{m} \theta_{g\boldsymbol{.}}, \sum\limits_{h=1}^{n} \theta_{\boldsymbol{.}h})(1-\pi)^{mn - \sum_{g=1}^{m}\sum_{h=1}^{n} \theta_{g\boldsymbol{.}}\theta_{\boldsymbol{.}h}}.$

Remark 9. The above model reduces to the standard two-class mixture model when $\pi^* = \pi$, since in this case the joint distribution of $(\boldsymbol{\theta}_{row}, \boldsymbol{\theta}_{col}, \boldsymbol{\Gamma})$, as seen from Lemma 3, is same as that of $\text{Bern}_{m\times n}(\pi)$. A difference between π and π^* determined through a new set of parameters $\boldsymbol{\lambda} = \{\lambda_r, \lambda_c, \lambda_{rc}\}$ can represent the underlying two-way grouping effect.

Let us define the following quantities, similar to that in the one-way classification setup.

$$\begin{aligned}
\text{Lfdr}_{gh}(\boldsymbol{\pi}) &= \text{Lfdr}_{gh}(\mathbf{x}, \boldsymbol{\pi}) = \Pr(\theta_{gh} = 0 | \mathbf{X}) \\
\text{Lfdr}_{g\boldsymbol{.},\boldsymbol{.}h}(\boldsymbol{\pi}) &= \text{Lfdr}_{g\boldsymbol{.},\boldsymbol{.}h}(\mathbf{x}, \boldsymbol{\pi}) = \Pr(\theta_{g\boldsymbol{.}} \cdot \theta_{\boldsymbol{.}h} = 0 | \mathbf{X}) \\
\text{Lfdr}_{gh|g\boldsymbol{.},\boldsymbol{.}h}(\boldsymbol{\pi}) &= \text{Lfdr}_{gh|g\boldsymbol{.},\boldsymbol{.}h}(\mathbf{x}, \boldsymbol{\pi}) = \Pr(\gamma_{gh} = 0 | \theta_{g\boldsymbol{.}} \cdot \theta_{\boldsymbol{.}h} = 1, \mathbf{X})
\end{aligned}$$

where $\boldsymbol{\pi} = (\pi, \pi^*)$.

It is easy to see that

$$\text{Lfdr}_{gh}(\boldsymbol{\pi}) = 1 - [1 - \text{Lfdr}_{gh|g\boldsymbol{.},\boldsymbol{.}h}(\boldsymbol{\pi})][1 - \text{Lfdr}_{g\boldsymbol{.},\boldsymbol{.}h}(\boldsymbol{\pi})]. \tag{3.3.3}$$

Furthermore, as shown in Appendix A.4, with

$$\text{Lfdr}_{gh}(\pi^*) \equiv \text{Lfdr}_{gh}(\pi^*, x_{gh}) = [(1-\pi^*)f_0(x_{gh})]/m(x_{gh};\pi^*), \tag{3.3.4}$$

where $m(x;\pi^*) = (1-\pi^*)f_0(x) + \pi^* f_1(x)$, and

$$\text{Lfdr}_{g\cdot}(\pi^*) = \prod_{h=1}^{n} \text{Lfdr}_{gh}(\pi^*), \quad \text{Lfdr}_{\cdot h}(\pi^*) = \prod_{g=1}^{m} \text{Lfdr}_{gh}(\pi^*),$$

and $\text{L}_{gh}(\pi^*) = \frac{\text{Lfdr}_{g\cdot}(\pi^*)\text{Lfdr}_{\cdot h}(\pi^*)}{\text{Lfdr}_{gh}(\pi^*)}$, we have

$$\text{Lfdr}_{gh|g\cdot,\cdot h}(\pi^*) = \frac{\text{Lfdr}_{gh}(\pi^*) - \text{Lfdr}_{g\cdot,\cdot h}(\pi^*)}{1 - \text{Lfdr}_{g\cdot,\cdot h}(\pi^*)}, \tag{3.3.5}$$

where

$$\text{Lfdr}_{g\cdot,\cdot h}(\pi^*) = \text{Lfdr}_{g\cdot}(\pi^*) + \text{Lfdr}_{\cdot h}(\pi^*) - \frac{\text{Lfdr}_{g\cdot}(\pi^*)\text{Lfdr}_{\cdot h}(\pi^*)}{\text{Lfdr}_{gh}(\pi^*)}, \tag{3.3.6}$$

Corresponding to the gth row and hth column, we can obtain

$$\text{Lfdr}_{g\cdot}(\boldsymbol{\pi}) = \frac{\text{Lfdr}_{g\cdot}(\pi^*)}{\text{Lfdr}_{g\cdot}(\pi^*) + \lambda_r(1 - \text{Lfdr}_{g\cdot}(\pi^*))} \tag{3.3.7}$$

$$\text{Lfdr}_{\cdot h}(\boldsymbol{\pi}) = \frac{\text{Lfdr}_{\cdot h}(\pi^*)}{\text{Lfdr}_{\cdot h}(\pi^*) + \lambda_c(1 - \text{Lfdr}_{\cdot h}(\pi^*))} \tag{3.3.8}$$

where $\lambda_r = \frac{1-(1-\pi)^n}{(1-\pi)^n} \cdot \frac{(1-\pi^*)^n}{1-(1-\pi^*)^n}$ and $\lambda_c = \frac{1-(1-\pi)^m}{(1-\pi)^m} \cdot \frac{(1-\pi^*)^m}{1-(1-\pi^*)^m}$

and

$$\text{L}_{gh}(\boldsymbol{\pi}) = \frac{\text{L}_{gh}(\pi^*)}{\lambda_{rc}} \cdot \frac{1}{1 - (1 - \frac{\lambda_r'}{\lambda_{rc}})\text{Lfdr}_{g\cdot}(\pi^*) - (1 - \frac{\lambda_c'}{\lambda_{rc}})\text{Lfdr}_{\cdot h}(\pi^*) + (1 - \frac{\lambda_r'+\lambda_c'-1}{\lambda_{rc}})L(\pi^*)} \tag{3.3.9}$$

where $\lambda_{rc} = \left(\frac{1-\pi^*}{1-\pi}\right)^{m+n-1} \cdot \frac{1-(1-\pi)^m-(1-\pi)^n+(1-\pi)^{m+n-1}}{1-(1-\pi^*)^m-(1-\pi^*)^n+(1-\pi^*)^{m+n-1}}$. Nature of λ_r' and λ_c' are respectively similar to λ_c and λ_r and hence their discussion is deferred to the Appendix.

Consequently, we have

$$\mathrm{Lfdr}_{g\bullet,\bullet h}(\boldsymbol{\pi}) = \mathrm{Lfdr}_{g\bullet}(\boldsymbol{\pi}) + \mathrm{Lfdr}_{\bullet h}(\boldsymbol{\pi}) - \mathrm{L}_{gh}(\boldsymbol{\pi}) \tag{3.3.10}$$

It is to be noted that a difference between probabilities π and π^* is made distinct by the constant ratios λ_r, λ_c and λ_{rc}. In case $\pi = \pi^*$, these ratios clearly reduce to 1. These results can be summarized in the following:

Proposition 1. *Let* $\mathrm{Lfdr}_{gh}(\pi^*)$ *and* $\mathrm{Lfdr}_{gh}(\pi, \pi^*)$ *be the local FDRs associated with* H_{gh} *in the* (g, h)*th cell under the standard single-group two-class mixture model with* π^* *as the probability of a hypotheses in that cell to be significant and Two-Way GAMM(*π, π^*, m, n*), respectively. Then,* $\mathrm{Lfdr}_{gh}(\pi, \pi^*)$ *can be expressed in terms of* $\mathrm{Lfdr}_{gh|g\bullet,\bullet h}(\pi^*)$ *and* $\boldsymbol{\lambda} \equiv \boldsymbol{\lambda}(\pi, \pi^*) = \{\lambda_r, \lambda_c, \lambda_{rc}\}$ *utilizing expressions (3.3.3), (3.3.5) and (3.3.10).*

This proposition implies that an Lfdr-based multiple testing procedure designed for testing hypotheses in this structure should be able to separately account for the Lfdr of the parent groups containing a particular hypothesis before accounting for its individual significance.

We define $\delta_{gh} \in \{0, 1\}$ to be the decision rule associated to θ_{gh}. The focus is on controlling the Posterior Total FDR, which has been previously defined, but we re-state it here in the context of the two-way classification setup.

$$\mathrm{PFDR}_T = E\left[\left.\frac{\sum_{g=1}^{m}\sum_{h=1}^{n}(1-\theta_{gh})\delta_{gh}(\mathbf{X})}{\sum_{g=1}^{m}\sum_{h=1}^{n}\delta_{gh}(\mathbf{X}) \vee 1}\right| \mathbf{X}\right] = \frac{\sum_{g=1}^{m}\sum_{h=1}^{n}\delta_{gh}(\mathbf{X})\mathrm{Lfdr}_{gh}(\boldsymbol{\pi})}{\sum_{g=1}^{m}\sum_{h=1}^{n}\delta_{gh}(\mathbf{X}) \vee 1}$$

In the remaining portion of the dissertation, we reduce the notation $\mathrm{Lfdr}_{gh}(\boldsymbol{\pi})$ to Lfdr_{gh} for convenience, though it continues to hold the same meaning. The following algorithm, stated in its oracle form proposes a method to identify the significant hypotheses in the two-way classification setup, factoring in the structural information, hence providing a Group Adjusted Testing (GATE) framework.

Algorithm 1 Two-way GATE Procedure

Step 1. Compute Lfdr_{gh} values as in equation (3.3.3) for all $g = 1, \ldots, m$, $h = 1, \ldots, n$.

Step 2. Pool these Lfdr values and arrange them in increasing order as $\mathrm{Lfdr}_{(1)} \leq \mathrm{Lfdr}_{(2)} \leq \ldots \leq \mathrm{Lfdr}_{(N)}$.

Step 3. Reject the hypotheses corresponding to $\mathrm{Lfdr}_{(k)}, k = 1, \ldots, K$ where

$$K = \max\left\{ l : \frac{1}{l} \sum_{k=1}^{l} \mathrm{Lfdr}_{(k)} \leq \alpha \right\}.$$

Theorem 3.1. *The above Two-Way GATE procedure controls $PFDR_T$ at level α.*

The proof of this theorem is immediate, from the definition of PFDR_T and Lfdr_{gh} values.

Remark 10. We next provide some arguments to show that the method also chooses the optimum number of rows and columns as significant with an upper bound on the number of rows and columns falsely discovered. These claims are supported by the simulation studies we performed in the following section.

The state of significant hypotheses in the matrix is given by

$$\theta_{gh} = \theta_{g.} \cdot \theta_{.h} \cdot \gamma_{gh} = 1.$$

Each of these hypotheses, render the corresponding parent row and column significant. The union of all the parent rows and columns spanned by these hypotheses constitute the significant block.

Suppose, R is the number of significant rows and S_R is the corresponding set of indices. Likewise, we can define C and S_C for the significant columns. For a particular combination of row and column (g, h),

$$\begin{aligned}\{g \in S_R, h \in S_C \mid \mathbf{X}\} \equiv &\{\theta_{gh} = 1 \mid \mathbf{X}\} \cup \\ &\{\theta_{gh} = 0, \theta_{gh'} = 1, \theta_{g'h} = 1, \text{for at least one } g' \neq g, h' \neq h \mid \mathbf{X}\}\end{aligned} \tag{3.3.11}$$

Thus, $\Pr(g \in S_R, h \in S_C \mid \mathbf{X}) \geq \Pr(\theta_{gh} = 1 \mid \mathbf{X}) = 1 - \text{Lfdr}_{gh}(\pi, \pi^*)$.
Here $\text{Lfdr}_{gh}(\pi, \pi^*)$ is the posterior probability that a particular hypothesis is null, under the assumption that it belongs to a significant row and column, i.e., $\theta_{g\boldsymbol{.}} \cdot \theta_{\boldsymbol{.}h} = 1$, and importantly, independent of the effect of the other rows and columns. We know

$$\text{Lfdr}_{gh}(\pi, \pi^*) = \Pr(\gamma_{gh} = 0 \mid \theta_{g\boldsymbol{.}} \theta_{\boldsymbol{.}h} = 1, \mathbf{X}) \Pr(\theta_{g\boldsymbol{.}} \theta_{\boldsymbol{.}h} = 1 \mid \mathbf{X})$$

In the Two-way GATE procedure, note that

$$K = \max \left\{ l : \frac{1}{l} \sum_{k=1}^{l} \text{Lfdr}_{(k)}(\pi, \pi^*) \leq t \right\} \in \{1, \ldots, mn\}. \tag{3.3.12}$$

Let $S_{K_{row}}$ being the subset of indices for rows containing these significantly small Lfdr_{gh} values and $S_{K_{col}}$ being similarly defined for columns,

$$\begin{aligned}\frac{1}{K} \sum_{k=1}^{K} \text{Lfdr}_{(k)}(\pi, \pi^*) &\geq 1 - \sum_{g \in S_{K_{row}}, h \in S_{K_{col}}} \Pr(g \text{ constitutes } S_R,\, h \text{ constitutes } S_C) \\ &= 1 - \Pr(S_R \subseteq S_{K_{row}}, S_C \subseteq S_{K_{col}})\end{aligned} \tag{3.3.13}$$

In fact, the estimates of S_R and S_C are $\widehat{S}_R = S_{K_{row}}, \widehat{S}_C = S_{K_{col}}$.

We thus provide an upper bound on the probability that '$\widehat{S}_R$ and $\widehat{S}_C$ are not the optimum set of significant rows and columns'. Inside the $R \times C$ significant block we thus determine, there is at least one significant hypothesis in each row and column.

3.4 Simulation Studies

In this section we considered a few simulation scenarios involving two-way classified hypotheses, to compare the performance of our proposed Two-way GATE procedure with the Two-way GBH_1 procedure and the classical BH procedure. We consider $N = 5000$ hypotheses arranged in $m = 50$ rows and $n = 100$ columns and apply the following steps.

1. Generate $\boldsymbol{\theta_{g.}}$ as a random vector of m i.i.d. $\text{Ber}(\pi_1)$, and $\boldsymbol{\theta_{.h}}$ as a random vector of n i.i.d. $\text{Ber}(\pi_2)$ samples, with π_1 and π_2 satisfying $(1 - \pi_1)^n = (1 - \pi_2)^m = 1 - \pi$. Mark the samples that are significant from $\boldsymbol{\theta_{g.}}$ as S_R and that from $\boldsymbol{\theta_{.h}}$ as S_C.

2. Define $\mathbf{\Gamma}$ as a null matrix of order $m \times n$. Further, in the significant sub-matrix of $\mathbf{\Gamma}$ formed at the intersection of the R significant rows and C columns found in step 1, draw a sample from $\text{Bern}_{R\times C}(\pi^*)$ distribution. Obtain

$$\mathbf{\Theta} = \text{Diag}(\boldsymbol{\theta_{g.}})\ \mathbf{\Gamma}\ \text{Diag}(\boldsymbol{\theta_{.h}}).$$

3. Given $\mathbf{\Theta}$, generate a random $m \times n$ matrix $X = ((X_{gh}))$ as follows:

$$X = \mu\mathbf{\Theta} + Z_{mn},$$

   Z_{mn} comprising of i.i.d. $N(0, 1)$ random samples.

4. Apply each procedure at FDR level $\alpha = 0.05$ for testing $H_{gh} : E(X_{gh}) = 0$ versus $K_{gh} : E(X_{gh}) > 0$, simultaneously $\forall g = 1, \ldots, m$, $h = 1, \ldots, n$ and note the proportions of false rejections among all rejections and correct rejections among all false nulls.

5. Repeat steps 1-4 100 times to simulate values of FDR and power for each pro-

cedure by averaging out the the corresponding proportions obtained in step 4.

We set $\mu = 0$ for true nulls and $= 3$ for the true signals.

In the first set of comparisons, we fixed $\pi^* = 0.65$ and monitored the effect varying π on on each of the methods in terms of FDR control and power. The results are shown in figure 3.1.

Two-way GATE maintains control on FDR less conservatively than the BH with increasing valuesof π, i.e., increasing proportion of significant rows and columns. At the same time, it is the most powerful among the three competing methods, though difference with Two-way GBH_1 becomes very fine with increasing π.

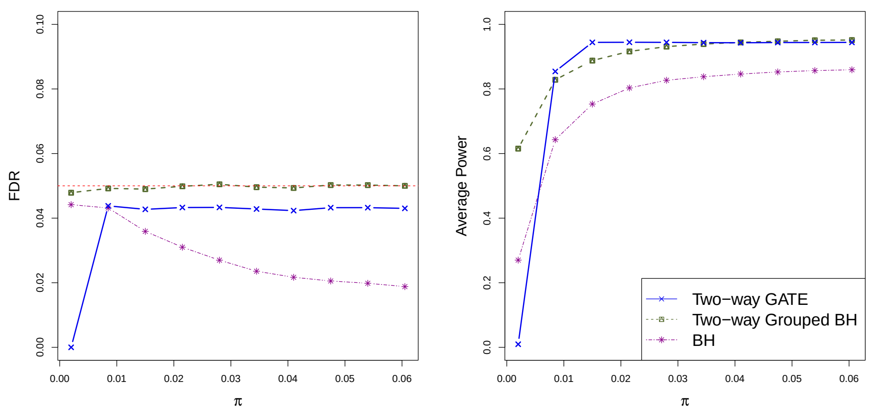

Figure 3.1: FDR and Power comparisons of the Two-way GATE procedure proposed with other methods, under independence, with increasing probabilities of row and column significance, increasing with π.

It is interesting to note that the Two-way GATE procedure maintains a good control on the average proportion of rows and columns wrongly identified as significant, and has very high power for almost all values of π. This corroborates with the theoretical findings in Remark 10. This is shown in figure 3.2.

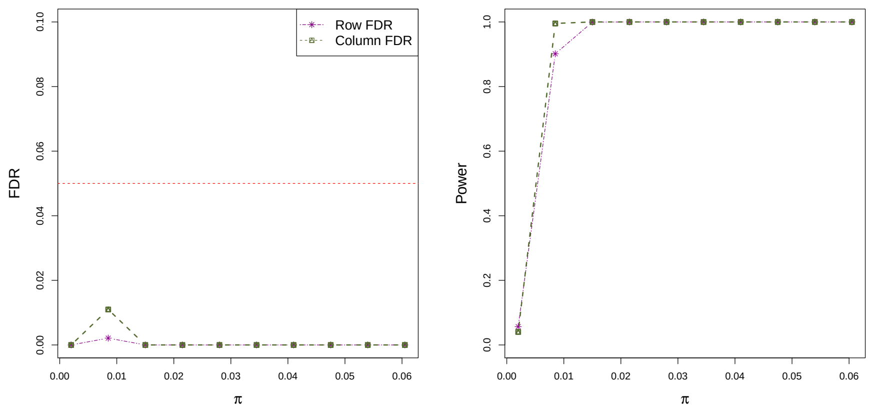

Figure 3.2: Comparisons of the average proportion of rows and columns falsely discovered (FDR), and the proportion of significant groups truly discovered (Power) by the Two-way GATE procedure, with varying probabilities of row and column significance, i.e., π.

In the second set of comparisons, we fixed $\pi = 0.4$ and monitored the effect varying π^* on on each of the methods in terms of FDR control and power. The results are shown in figure 3.3.

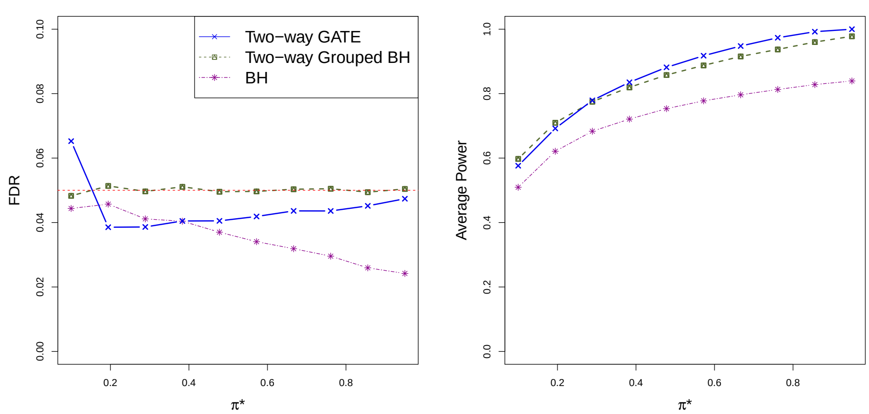

Figure 3.3: FDR and Power comparisons of the Two-way GATE procedure proposed with other methods, under independence, with varying probabilities of hidden state of a hypothesis being significant, i.e., π^*.

In this case too, we see that with increasing density of signals, the proposed Two-way GATE procedure is more powerful and less conservative than the BH procedure.

In terms of average power, it performs even marginally better than the Two-way GBH_1 procedure.

As in the previous case, the Two-way GATE procedure maintains a very conservative control on the average proportions of rows and columns falsely discovered and identifies the significant rows and columns with very high average power, as exhibited in figure 3.4.

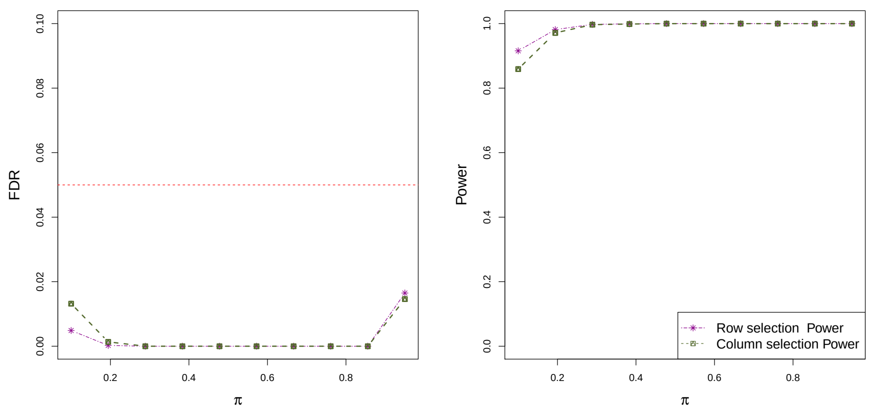

Figure 3.4: Comparisons of the average proportion of rows and columns falsely discovered (FDR), and the proportion of significant groups truly discovered (Power) by the Two-way GATE procedure, with varying π^*.

CHAPTER 4

CONCLUSION AND FUTURE WORK

In essence, this dissertation explores an underdeveloped area of multiple testing, which is adapting standard procedures to structures of hypotheses more complex than what these procedures were initially designed for. Occurrence of hypotheses exhibiting complex structures, especially in the form of being classified according to multiple criteria, is becoming more and more prevalent in modern statistical investigations. However, research focused on developing methods efficiently accommodating such structural information has been taking place at a much slower pace than expected. It is worth noting that no research has been put forward yet in the existing published literature, as far as we know, in the direction of adapting methods to two-way classification setting and beyond either in the BH type FDR framework or using Lfdr-based approaches, although many scientific experiments do give rise to such settings. This dissertation makes a significant contribution toward that direction, with motivation from the work of *Hu et al.* (2010), *Liu et al.* (2016) and *Sarkar and Zhao* (2017).

Our revisit of *Hu et al.* (2010) in the second chapter has led us to our research agenda aimed at complementing their work with new and improved results and extending it to broader domain. With our focus on testing hypotheses exhibiting one- or two-way classification structure, we have presented a unified framework founded upon the idea of adapting the FDR controlling BH procedure to such structures through p-value weighting without loosing the ultimate FDR control. The framework allows us to develop methods that adapt themselves to the underlying one- or two-way classification structure with the weights being appropriately chosen to reflect the structure. Our method for one-way classified hypotheses in its oracle form with known weights is same as the one in *Hu et al.* (2010). However, its data adaptive version here is different, with the weights being estimated using a different approach, and is seen to work much better in certain simulated scenarios where signals are not evenly dispersed across the significant groups. The FDR control for our adaptive methods for two-way classified hypotheses have been established, theoretically under a non-asymptotic setting and independent p-values and numerically through extensive simulations for certain dependence scenarios. Future work would address establishing that our proposed data-adaptive methods are able to control FDR under asymptotic settings.

Some open issues to be resolved in future research are as follows. In extending the Oracle One-Way GBH from one- to two-way classification setting before constructing a data-adaptive version of it, we have proposed using certain specific combinations of the row and column weights (see, (2.3.1), (2.3.3), (2.3.6), (2.3.7), (2.3.8a) and (2.3.8b)). However, it would be worthwhile to investigate if these weights can be combined in an optimal manner. The data-adaptive procedures here have been proposed by estimating weights using *Storey et al.* (2004) type estimates of proportions of true nulls. Developing alternative data-adaptive procedures using other types of estimates of these proportions would be an important undertaking.

Since two-way classification of hypotheses is a commonly occurring phenomenon in scientific studies, there is a vast scope for our proposed methods to be efficiently used in interesting research problems. Currently in our consideration is applying our two-way GBH testing framework to identify different brain regions influenced by alcohol consumption from Electroencephalographic (EEG) data. EEG being spatio-temporal data, provides a natural two-way classification structure. Challenge lies in effectively formulating the multiple hypotheses testing problem from the complex data-structure and extending the current testing framework so as to draw effective conclusions from the study.

In Chapter 3, we have laid out a framework for multiple testing procedures for two-way classified hypotheses founded on Bayesian concepts using the Lfdr approach. Our proposed model for the two-way classification setup has its roots in the model described in *Liu et al.* (2016) and *Sarkar and Zhao* (2017) for one-way classified hypotheses. It spells out the effects of the two simultaneous grouping imposed on each individual hypothesis. This enables us to account for the significance of each hypothesis, and the significance of each type of grouping as well as their intersecting effect. Future research would include strengthening this model with deeper analyses of its features. Our proposed Two-way GATE is a very powerful multiple testing method as depicted in the simulated scenarios. It is able to control the desired error rate at a pre-defined level of significance. At the same time, it is also capable to conservatively control the average proportion of groups falsely discovered, along either classifications. Though this suggests that the method is optimal in some sense, there remains scope to propose better arguments for optimality of the procedure. Development of data-driven versions of the test, would be an important but challenging extension, which would eventually lead to application of the procedure on real datasets.

BIBLIOGRAPHY

Basu, P., T. T. Cai, K. Das, and W. Sun (2018), Weighted false discovery rate control in large-scale multiple testing, *Journal of the American Statistical Association*, *113*(523), 1172–1183, 10.1080/01621459.2017.1336443.

Benjamini, Y., and R. Cohen (2017), Weighted false discovery rate controlling procedures for clinical trials, *Biostatistics*, *18*(1), 91–104, 10.1093/biostatistics/kxw030.

Benjamini, Y., and R. Heller (2007), False discovery rates for spatial signals, *Journal of the American Statistical Association*, *102*(480), 1272–1281, 10.1198/016214507000000941.

Benjamini, Y., and Y. Hochberg (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 Y. Hochberg (1997), Multiple hypotheses testing with weights, *24*(3), 407–418.

Benjamini, Y., and Y. Hochberg (2000), On the adaptive control of the false discovery rate in multiple testing with independent statistics, *Journal of Educational and Behavioral Statistics*, *25*(1), 60–83, 10.3102/10769986025001060.

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

Benjamini, Y., and D. Yekutieli (2001b), The control of the false discovery rate in multiple testing under dependency, *Ann. Statist.*, *29*(4), 1165–1188, 10.1214/aos/1013699998.

Benjamini, Y., A. M. Krieger, and D. Yekutieli (2006), Adaptive linear step-up procedures that control the false discovery rate, *Biometrika*, *93*(3), 491–507, 10.1093/biomet/93.3.491.

Blanchard, G., and E. Roquain (2009), Adaptive false discovery rate control under independence and dependence, *J. Mach. Learn. Res.*, *10*, 2837–2871.

Cai, T. T., and W. Sun (2009), Simultaneous testing of grouped hypotheses: Finding needles in multiple haystacks, *Journal of the American Statistical Association*, *104*(488), 1467–1481, 10.1198/jasa.2009.tm08415.

Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, C. A. Lozupone, P. J. Turnbaugh, N. Fierer, and R. Knight (2011), Global patterns of 16s rrna diversity at a depth of millions of sequences per sample, *Proc Natl Acad Sci U S A*, *108*(Suppl 1), 4516–4522, 10.1073/pnas.1000080107, 201000080[PII].

Clements, N., S. K. Sarkar, Z. Zhao, and D.-Y. Kim (2014), Applying multiple testing procedures to detect change in east african vegetation, *Ann. Appl. Stat.*, *8*(1), 286–308, 10.1214/13-AOAS686.

Efron, B. (2008), Microarrays, empirical bayes and the two-groups model, *Statist. Sci.*, *23*(1), 1–22, 10.1214/07-STS236.

Efron, B., R. Tibshirani, J. D. Storey, and V. Tusher (2001), Empirical bayes analysis of a microarray experiment, *Journal of the American Statistical Association*, *96*(456), 1151–1160, 10.1198/016214501753382129.

Ferkingstad, E., A. Frigessi, H. Rue, G. Thorleifsson, and A. Kong (2008), Unsupervised empirical bayesian multiple testing with external covariates, *Ann. Appl. Stat.*, *2*(2), 714–735, 10.1214/08-AOAS158.

Finner, H., T. Dickhaus, and M. Roters (2009), On the false discovery rate and an asymptotically optimal rejection curve, *Ann. Statist.*, *37*(2), 596–618, 10.1214/07-AOS569.

Foygel Barber, R., and A. Ramdas (2015), The p-filter: multi-layer FDR control for grouped hypotheses, *ArXiv e-prints*.

Genovese, C. R., K. Roeder, and L. Wasserman (2006), False discovery control with p-value weighting, *Biometrika*, *93*(3), 509–524, 10.1093/biomet/93.3.509.

Hu, J. X., H. Zhao, and H. H. Zhou (2010), False discovery rate control with groups, *Journal of the American Statistical Association*, *105*(491), 1215–1227, 10.1198/jasa.2010.tm09329.

Ignatiadis, N., B. Klaus, J. B. Zaugg, and W. Huber (2016), Data-driven hypothesis weighting increases detection power in genome-scale multiple testing, *Nature Methods*, *13*, 577 EP.

Kolda, T., and B. Bader (2009), Tensor decompositions and applications, *SIAM Review*, *51*(3), 455–500, 10.1137/07070111X.

Liu, Y., S. K. Sarkar, and Z. Zhao (2016), A new approach to multiple testing of grouped hypotheses, *Journal of Statistical Planning and Inference*, *179*, 1 – 14, https://doi.org/10.1016/j.jspi.2016.07.004.

Pacifico, M. P., C. Genovese, I. Verdinelli, and L. Wasserman (2004), False discovery control for random fields, *Journal of the American Statistical Association*, *99*(468), 1002–1014, 10.1198/0162145000001655.

Ramdas, A., R. Foygel Barber, M. J. Wainwright, and M. I. Jordan (2017), A unified treatment of multiple testing with prior knowledge using the p-filter, *ArXiv e-prints*.

Roeder, K., and L. Wasserman (2009), Genome-wide significance levels and weighted hypothesis testing, *Statist. Sci.*, *24*(4), 398–413, 10.1214/09-STS289.

Roeder, K., S.-A. Bacanu, L. Wasserman, and B. Devlin (2006), Using linkage genome scans to improve power of association in genome scans, *The American Journal of Human Genetics*, *78*(2), 243–252, 10.1086/500026.

Sankaran, K., and S. Holmes (2014), structssi: Simultaneous and selective inference for grouped or hierarchically structured data, *Journal of Statistical Software, Articles*, *59*(13), 1–21, 10.18637/jss.v059.i13.

Sarkar, S. K. (2002), Some results on false discovery rate in stepwise multiple testing procedures, *Ann. Statist.*, *30*(1), 239–257, 10.1214/aos/1015362192.

Sarkar, S. K. (2008), On methods controlling the false discovery rate, *Sankhyā: The Indian Journal of Statistics, Series A (2008-)*, *70*(2), 135–168.

Sarkar, S. K., and Z. Zhao (2017), Local False Discovery Rate Based Methods for Multiple Testing of One-Way Classified Hypotheses, *ArXiv e-prints*.

Stein, J. L., et al. (2010), Voxelwise genome-wide association study (vgwas), *NeuroImage*, *53*(3), 1160 – 1174, https://doi.org/10.1016/j.neuroimage.2010.02.032, imaging Genetics.

Storey, J. D. (2002), A direct approach to false discovery rates, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(3), 479–498, 10.1111/1467-9868.00346.

Storey, J. D., J. E. Taylor, and D. Siegmund (2004), Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach, *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *66*(1), 187–205.

Sun, W., and T. T. Cai (2007), Oracle and adaptive compound decision rules for false discovery rate control, *Journal of the American Statistical Association*, *102*(479), 901–912, 10.1198/016214507000000545.

Sun, W., and Z. Wei (2011), Multiple testing for pattern identification, with applications to microarray time-course experiments, *Journal of the American Statistical Association*, *106*(493), 73–88, 10.1198/jasa.2011.ap09587.

Sun, W., B. J. Reich, T. T. Cai, M. Guindani, and A. Schwartzman (2015), False discovery control in large-scale spatial multiple testing, *J R Stat Soc Series B Stat Methodol*, *77*(1), 59–83, 10.1111/rssb.12064, 25642138[pmid].

Zablocki, R. W., A. J. Schork, R. A. Levine, O. A. Andreassen, A. M. Dale, and W. K. Thompson (2014), Covariate-modulated local false discovery rate for genome-wide association studies, *Bioinformatics*, *30*(15), 2098–2104.

APPENDIX

A.1 Proof of Result 1

The FDR of a stepup procedure based on the weighted p-values and any set of critical constants $c_1 \leq \cdots \leq c_N$ can be expressed as follows [see, e.g., Sarkar (2002)]:

$$\begin{aligned} \text{FDR} &= \sum_{r=1}^{N} \sum_{i \in I_0} \frac{1}{r} \Pr\left(P_i^w \leq c_r, R_{(-i)}^w = r-1\right) \\ &= \sum_{r=0}^{N-1} \sum_{i \in I_0} \text{E}\left[\Pr\left(R_{(-i)}^w \geq r \mid P_i^w\right) \left\{ \frac{I(P_i^w \leq c_{r+1})}{r+1} - \frac{I(P_i^w \leq c_r)}{r} \right\}\right], \end{aligned} \tag{A.1}$$

(assuming $c_0 = 0$ and $0/0 = 0$), with $R_{(-i)}^w$ representing the number of rejections in the stepup procedure based on the weighted p-values $(P_1^w, \ldots, P_N^w) \setminus \{P_i^w\}$ and the critical constants c_i, $i = 2, \ldots, N$. With $c_i = ic_1$, $i = 1, \ldots, N$, it is bounded above by $\sum_{i \in I_0} \Pr(P_i^w \leq c_1)$ under Condition 1, which can be shown by making use of the following observations for each $i \in I_0$:

- For each $r = 1, \ldots, N-1$,

$$\begin{aligned} & \text{E}\left[\Pr\left(R_{(-i)}^w \geq r \mid P_i^w\right) \left\{ \frac{I(P_i^w \leq c_{r+1})}{r+1} - \frac{I(P_i^w \leq c_r)}{r} \right\}\right] \\ \leq\ & \Pr\left(R_{(-i)}^w \geq r \mid P_i^w = c_r\right) \left\{ \frac{\Pr(P_i^w \leq c_{r+1})}{r+1} - \frac{\Pr(P_i^w \leq c_r)}{r} \right\} \\ \leq\ & 0. \end{aligned} \tag{A.2}$$

The first inequality in (A.2) follows from the following two results:

(i) $\Pr\left(R^w_{(-i)} \geq r \mid P^w_i\right) = \mathrm{E}\left\{I(R^w_{(-i)} \geq r) \mid P^w_i\right\}$ is non-increasing in P^w_i, since $I(R^w_{(-i)} \geq r)$ is a non-increasing function of the weighted p-values, and the PRDS condition on the p-values translates into that for the weighted p-values, and

(ii) $\frac{I(P^w_i \leq c_{r+1})}{r+1} - \frac{I(P^w_i \leq c_r)}{r}$ changes sign from $-$ to $+$ at $P^w_i = c_r$ as P^w_i increases.

And, the second inequality follows from the fact that

$$\frac{\Pr(P^w_i \leq c_{r+1})}{r+1} - \frac{\Pr(P^w_i \leq c_r)}{r} = \frac{\min\left\{\frac{(r+1)c_1}{w_i}, 1\right\}}{r+1} - \frac{\min\left\{\frac{rc_1}{w_i}, 1\right\}}{r} \leq 0.$$

- For $r = 0$, the expectation in (A.2) equals $\Pr(P^w_i \leq c_1)$.

With $c_1 = \alpha/N$, $\sum_{i \in I_0} \Pr(P^w_i \leq c_1) = \sum_{i \in I_0} \min\{\frac{\alpha}{N w_i}, 1\} \leq \frac{\alpha}{N} \sum_{i \in I_0} \frac{1}{w_i}$.

Thus, Result 1 is proved.

Remark 11. The weaker form of the PRDS condition, as stated in Section 2.1.2, also suffices for this result. This is because the inequality in (A.2) can be proved using the following alternative arguments in this case:

The expectation in (A.2), which equals

$$\frac{\Pr\left(R^w_{(-i)} \geq r \mid P^w_i \leq c_{r+1}\right) \Pr(P^w_i \leq c_{r+1})}{r+1} - \frac{\Pr\left(R^w_{(-i)} \geq r \mid P^w_i \leq c_r\right) \Pr(P^w_i \leq c_r)}{r},$$

is less than or equal to the following under the weaker condition:

$$\Pr\left(R^w_{(-i)} \geq r \mid P^w_i \leq c_r\right) \left\{\frac{\Pr(P^w_i \leq c_{r+1})}{r+1} - \frac{\Pr(P^w_i \leq c_r)}{r}\right\} \leq 0.$$

A.2 Further comparison of the proposed methods and their competitors

The performance of our proposed methods were compared with relevant competitors in different simulated scenarios. These situations were regulated through the use of different parameters. In order to produce a comprehensive study of the effect of all the parameters on False Discovery Rate(FDR) and Power, we use an Analysis of Variance type table in each case.

- **Data-Adaptive One-way GBH vs. the Naive Adaptive BH.** The one-way layout consists of m groups, and each consists of n hypotheses. The state of gth group is $\theta_g \sim \text{Ber}(1-\pi_{\bullet})$, i.i.d., $\forall g = 1, \ldots, m$. Within each significant group, the state of the jth member hypothesis is $\theta_{gi} \sim \text{Ber}(1-\pi)$, i.i.d. $i = 1, \ldots, n$. Let $\theta_{\bullet|g} = \{\theta_{gi}, i = 1, \ldots, n\}$. The within group correlation is measured by $\rho \in [0, 1)$, in each group. Either adaptive procedures depend on an initial one-step procedure performed at level $\lambda \in (0, 1)$.

In ANOVA analyses based on the set of four variables $(\pi_{\bullet}, \pi, \rho, \lambda)$, the interaction effect of ρ - λ, and the interaction of ρ with $\pi_{\bullet}$ and π are of interest. The analyses yields the following table of sum of squares.

		Adaptive GBH		Adaptive BH	
	Df	Sum Sq	Mean Sq	Sum Sq	Mean Sq
$1-\pi_{\bullet}$	4	302.7222	75.6806	434.4386	108.6096
$1-\pi$	9	16.1465	1.7941	0.0033	0.0004
ρ	5	29.7121	5.9424	0.0003	0.0001
λ	9	14.4786	1.6087	0.0220	0.0024
ρ:λ	45	7.6129	0.1692	0.0027	0.0001
$1-\pi_{\bullet}$:$1-\pi$:ρ	281	24.4746	0.0871	0.0198	0.0001
Residuals	2646	13.4775	0.0051	0.0511	0.0000

Table A.1: ANOVA on FDR values obtained from each of the Adaptive GBH and Adaptive BH procedures applied to one-way classified hypotheses

	Df	Adaptive GBH		Adaptive BH	
		Sum Sq	Mean Sq	Sum Sq	Mean Sq
$1-\pi_{\bullet}$	4	313.6355	78.4089	290.3154	72.5789
$1-\pi$	9	35.2517	3.9169	41.4301	4.6033
ρ	5	0.0137	0.0027	0.0018	0.0004
λ	9	0.4658	0.0518	0.0535	0.0059
ρ:λ	45	0.5176	0.0115	0.0014	0.0000
$1-\pi_{\bullet}$:$1-\pi$:ρ	281	11.0364	0.0393	10.4292	0.0371
Residuals	2646	2.6276	0.0010	0.1076	0.0000

Table A.2: ANOVA on average Power obtained from each of the Adaptive GBH and Adaptive BH procedures applied to one-way classified hypotheses

In general, comparing the column of sum of squares for the two methods, the one-way Adaptive GBH generally has higher sum-of squares values than the adaptive BH procedure for each parameter and interaction effects. The pattern is same when the response is either FDR or power. This is responsible for the one-way adaptive GBH to capture more information. While this leads to its better performance in terms of identifying sparse signals, it is also more sensitive to the amount of positive dependence ρ and the choice of the tuning parameter λ.

- **Oracle Two-way GBH$_1$ vs. the BH procedure**

For the simulations in section 2.5.2 pertaining to the Oracle Two-way GBH$_1$ and the BH procedure, we apply each procedure at FDR level $\alpha = 0.05$ for testing $H_{g,i} : \mathrm{E}(X_{r,c}) = 0$ against $K_{r,c} : \mathrm{E}(X_{r,c}) > 0$. We set $m = 50, n = 100$, $\mathrm{E}(X_{r,c}) = \mu = 0$ for true nulls and $\mathrm{E}(X_{r,c}) = \mu = 3$ for true signals in our simulations. The variations in the simulated scenarios were governed by the set of parameters in the set $(\pi_r, \pi_c, \pi_{rc}, \rho_r, \rho_c)$. In order to calculate the ANOVA tables A.3 and A.4, we consider values of each of these parameters in the range $(0, 1)$. We consider 5 values for each of π_r, π_c and 10 values for π_{rc}. For ρ_r and ρ_c we consider 5 values each, including 0 for the case of independence. For each of the 6250 combinations of the parameters, we calculate the average FDR and average power obtained from 200 repetitions of both methods.

In ANOVA analyses based on the set of five variables $(\pi_r, \pi_c, \pi_{rc}, \rho_r, \rho_c)$, the interaction effect of ρ - λ, and the interaction of $1 - \pi_r{:}\rho_c$ and $1 - \pi_c{:}\rho_r$ are of interest, as they present the interaction of density of signals and positive dependence within each row(column) . The analyses yield the following tables of sum of squares.

		Two-way GBH_1		BH	
	Df	Sum Sq	Mean Sq	Sum Sq	Mean Sq
$1 - \pi_{rc}$	9	32.6340	3.6260	34.4242	3.8249
$1 - \pi_r$	4	0.1372	0.0343	0.3067	0.0767
$1 - \pi_c$	4	6.3184	1.5796	6.8649	1.7162
ρ_r	4	0.0014	0.0003	0.0056	0.0014
ρ_c	4	0.0007	0.0002	0.0029	0.0007
$1 - \pi_r{:}\rho_c$	16	0.0003	0.0000	0.0015	0.0001
$1 - \pi_c{:}\rho_r$	16	0.0006	0.0000	0.0034	0.0002
Residuals	99942	1263.1398	0.0126	1332.1496	0.0133

Table A.3: ANOVA on FDR values obtained from the Two-way GBH_1 and BH procedures applied to two-way classified hypotheses

For most parameters, the BH procedure has marginally higher sum-of squares values. This may imply each of the parameters have less impact on the Two-way GBH_1 procedure or more conclusive results may be obtained through simulation of larger data-sets. However, this is corroborated by the graphical results we obtained before, where for different choices of parameters, the methods did not show much deviation and very similar results were obtained across all the graphs.

		Two-way GBH_1		BH	
	Df	Sum Sq	Mean Sq	Sum Sq	Mean Sq
$1-\pi_{rc}$	9	25.3049	2.8117	21.8638	2.4293
$1-\pi_r$	4	1.6794	0.4199	1.2420	0.3105
$1-\pi_c$	4	4.4918	1.1230	3.9752	0.9938
ρ_r	4	0.0062	0.0015	0.0021	0.0005
ρ_c	4	0.0007	0.0002	0.0006	0.0001
$1-\pi_r{:}\rho_c$	16	0.0006	0.0000	0.0012	0.0001
$1-\pi_c{:}\rho_r$	16	0.0007	0.0000	0.0024	0.0002
Residuals	99942	2408.3724	0.0241	1476.3624	0.0148

Table A.4: ANOVA on average Power obtained from the Two-way GBH_1 and BH procedures applied to two-way classified hypotheses

On the other hand, when the response is power, the Two-way GBH_1 has marginally higher sum-of squares values. Conclusions are same as that of FDR.

A.3 Expressions of the marginal distributions of Z_{row} and Z_{col} obtained from their joint distribution in Lemma 3 (ii)

With $\mathbf{Z}$ being a binary matrix of dimensions $m\times, n$, i.e., a $\mathrm{Bern}_{\mathrm{m\times n}}(\pi)$ matrix, let $\mathbf{Z}_{row}$ and $\mathbf{Z}_{col}$ being as defined in lemma 3. Suppose $\sum_{g=1}^{m} Z_{g\boldsymbol{.}} = R$ and $\sum_{h=1}^{n} Z_{\boldsymbol{.}h} = C$. From lemma 3, we can see that the joint distribution of $(\mathbf{Z}_{row}, \mathbf{Z}_{col}) = P_\pi(R, C)(1-\pi)^{mn-RC}$. Note that C can be chosen in $\binom{n}{C}$ ways. We obtain the marginal distribution of $\mathbf{Z}_{row}$ as follows

$$\begin{aligned}
&\sum_{C=0}^{n}\binom{n}{C}P_\pi(R,C)(1-\pi)^{mn-RC}\\
&=\sum_{C=0}^{n}\binom{n}{C}\sum_{i=0}^{R}\sum_{j=0}^{C}(-1)^{i+j}\binom{R}{i}\binom{C}{j}(1-\pi)^{mn-(R-i)(C-j)}\\
&=(1-\pi)^{mn}\sum_{C=0}^{n}\binom{n}{C}\sum_{i=0}^{R}(-1)^{i}\binom{R}{i}\sum_{j=0}^{C}(-1)^{j}\binom{C}{j}\{(1-\pi)^{-(R-i)}\}^{C-j}\\
&=(1-\pi)^{mn}\sum_{i=0}^{R}(-1)^{i}\binom{R}{i}\sum_{C=0}^{n}\binom{n}{C}\{(1-\pi)^{-(R-i)}-1\}^{C}\\
&=(1-\pi)^{mn}\sum_{i=0}^{R}(-1)^{i}\binom{R}{i}(1-\pi)^{-n(R-i)}\\
&=(1-\pi)^{n(m-R)}\{1-(1-\pi)^{n}\}^{R}
\end{aligned}$$

Hence the marginal distribution of $\mathbf{Z}_{row}$ is indeed the joint distribution of m i.i.d., $\mathrm{Ber}(1-(1-\pi)^n)$ random variables, R of which are significant. Likewise, we can show that the marginal distribution of $\mathbf{Z}_{col}$ is given by $(1-\pi)^{m(n-C)}\{1-(1-\pi)^m\}^C$.

A.4 Derivation of expressions of $\mathbf{Lfdr}_{gh|g\boldsymbol{\cdot},\boldsymbol{\cdot}h}(\boldsymbol{\pi})$ in equation 3.3.5 and $\mathbf{Lfdr}_{g\boldsymbol{\cdot},\boldsymbol{\cdot}h}(\boldsymbol{\pi})$ in equation 3.3.10

Lemma 4. *From the Two-way GAMM(π,π^*,m,n) described in definition 5 and from the definitions*

$$\begin{aligned}
&Lfdr_{gh}(\pi^*)\equiv \mathrm{Lfdr}_{gh}(\pi^*,x_{gh})=[(1-\pi^*)f_0(x_{gh})]/m(x_{gh};\pi^*),\\
&\textit{ where } m(x;\pi^*)=(1-\pi^*)f_0(x)+\pi^* f_1(x),\\
&Lfdr_{g\boldsymbol{\cdot}}(\pi^*)=\prod_{h=1}^{n}\mathrm{Lfdr}_{gh}(\pi^*),\quad Lfdr_{\boldsymbol{\cdot}h}(\pi^*)=\prod_{g=1}^{m}\mathrm{Lfdr}_{gh}(\pi^*),\\
&\textit{and } L_{gh}(\pi^*)=\frac{\mathrm{Lfdr}_{g\boldsymbol{\cdot}}(\pi^*)\mathrm{Lfdr}_{\boldsymbol{\cdot}h}(\pi^*)}{\mathrm{Lfdr}_{gh}(\pi^*)}
\end{aligned}$$

we can show

(a).
$$\mathrm{Lfdr}_{gh|g\cdot,\cdot h}(\pi^*) = \frac{\mathrm{Lfdr}_{gh}(\pi^*) - \mathrm{Lfdr}_{g\cdot,\cdot h}(\pi^*)}{1 - \mathrm{Lfdr}_{g\cdot,\cdot h}(\pi^*)},$$

where

$$Lfdr_{g\cdot,\cdot h}(\pi^*) = \mathrm{Lfdr}_{g\cdot}(\pi^*) + \mathrm{Lfdr}_{\cdot h}(\pi^*) - \frac{\mathrm{Lfdr}_{g\cdot}(\pi^*)\mathrm{Lfdr}_{\cdot h}(\pi^*)}{\mathrm{Lfdr}_{gh}(\pi^*)},$$

(b). Corresponding to the gth row and hth column, we can obtain

(i).
$$Lfdr_{g\cdot}(\boldsymbol{\pi}) = \frac{\mathrm{Lfdr}_{g\cdot}(\pi^*)}{\mathrm{Lfdr}_{g\cdot}(\pi^*) + \lambda_r(1 - \mathrm{Lfdr}_{g\cdot}(\pi^*))} \text{ where}$$
$$\lambda_r = \frac{1-(1-\pi)^n}{(1-\pi)^n} \cdot \frac{(1-\pi^*)^n}{1-(1-\pi^*)^n}$$

(ii).
$$Lfdr_{\cdot h}(\boldsymbol{\pi}) = \frac{\mathrm{Lfdr}_{\cdot h}(\pi^*)}{\mathrm{Lfdr}_{\cdot h}(\pi^*) + \lambda_c(1 - \mathrm{Lfdr}_{\cdot h}(\pi^*))} \text{ with}$$
$$\lambda_c = \frac{1-(1-\pi)^m}{(1-\pi)^m} \cdot \frac{(1-\pi^*)^m}{1-(1-\pi^*)^m}$$

(iii).
$$L_{gh}(\boldsymbol{\pi}) = \frac{1}{\lambda_{rc}} \cdot \frac{\mathrm{L}_{gh}(\pi^*)}{1 - (1 - \frac{\lambda_r'}{\lambda_{rc}})Lfdr_{g\cdot}(\pi^*) - (1 - \frac{\lambda_c'}{\lambda_{rc}})Lfdr_{\cdot h}(\pi^*) + (1 - \frac{\lambda_r' + \lambda_c' - 1}{\lambda_{rc}})L(\pi^*)}$$

$$\text{where } \lambda_{rc} = \left(\frac{1-\pi^*}{1-\pi}\right)^{m+n-1} \cdot \frac{1-(1-\pi)^m - (1-\pi)^n + (1-\pi)^{m+n-1}}{1-(1-\pi^*)^m - (1-\pi^*)^n + (1-\pi^*)^{m+n-1}}$$

$$\lambda_c' = \frac{1-(1-\pi)^{n-1}}{(1-\pi)^{n-1}} \cdot \frac{(1-\pi^*)^{n-1}}{1-(1-\pi^*)^{n-1}} \text{ and } \lambda_r' = \frac{1-(1-\pi)^{m-1}}{(1-\pi)^{m-1}} \cdot \frac{(1-\pi^*)^{m-1}}{1-(1-\pi^*)^{m-1}}$$

(iv). $$Lfdr_{g.,.h}(\boldsymbol{\pi}) = Lfdr_{g.}(\boldsymbol{\pi}) + Lfdr_{.h}(\boldsymbol{\pi}) - \mathrm{L}_{gh}(\boldsymbol{\pi})$$

Proof of (a): Define $K(\pi^*) = (1-\pi^*)^m + (1-\pi^*)^n - (1-\pi^*)^{m+n-1}$, the joint probability of the hidden states γ_{gh} when at least either of the gth row or the hth column are insignificant, i.e., $\theta_{g.} \cdot \theta_{.h} = 0$.

$$\begin{aligned}
&\Pr(\mathbf{X}, \gamma_{gh} = 0 | \theta_{g.}\theta_{.h} = 1) \\
&= \frac{(1-\pi^*) f_0(x_{gh})}{1-K(\pi^*)} \prod_{h' \neq h}^{n} m(x_{gh'}) \left\{ 1 - \prod_{h' \neq h}^{n} \mathrm{Lfdr}_{gh'}(\pi^*) \right\} \prod_{g' \neq g}^{m} m(x_{g'h}) \left\{ 1 - \prod_{g' \neq g}^{m} \mathrm{Lfdr}_{g'h}(\pi^*) \right\} \\
&= \frac{M}{1-K(\pi^*)} \cdot \mathrm{Lfdr}_{gh}(\pi^*) \left\{ 1 - \frac{\mathrm{Lfdr}_{g.}(\pi^*)}{\mathrm{Lfdr}_{gh}(\pi^*)} \right\} \left\{ 1 - \frac{\mathrm{Lfdr}_{.h}(\pi^*)}{\mathrm{Lfdr}_{gh}(\pi^*)} \right\} \\
&= \frac{M}{1-K(\pi^*)} \left(\mathrm{Lfdr}_{gh}(\pi^*) - \mathrm{Lfdr}_{g.,.h}(\pi^*) \right) \qquad \text{(A.1)}
\end{aligned}$$

$$\Pr(\mathbf{X}, \gamma_{gh} = 1 | \theta_{g.}\theta_{.h} = 1) = \frac{\pi^* f_1(x_{gh})}{1-K(\pi^*)} \prod_{h' \neq h}^{n} m(x_{gh'}) \prod_{g' \neq g}^{m} m(x_{g'h}) = \frac{M}{1-K(\pi^*)} \left(1 - \mathrm{Lfdr}_{gh}(\pi^*)\right) \qquad \text{(A.2)}$$

(A.1) + (A.2) implies

$$\Pr(\mathbf{X} | \theta_{g.}\theta_{.h} = 1) = \frac{M}{1-K(\pi^*)} (1 - \mathrm{Lfdr}_{g.,.h}(\pi^*)) \qquad \text{(A.3)}$$

(A.1) / (A.3) implies

$\Pr(\gamma_{ij} = 0 | \theta_{g.}\theta_{.h} = 1, \mathbf{X}) = \mathrm{Lfdr}_{gh|g.,.h}(\pi^*) = \frac{\mathrm{Lfdr}_{ij}(\pi^*) - \mathrm{Lfdr}_{g.,.h}(\pi^*)}{1 - \mathrm{Lfdr}_{g.,.h}(\pi^*)}$.

Proof of (b): Note that in the gth row,

$$\mathbf{X}_{g\cdot}|\theta_{g\cdot} = 0 \sim \prod_{h'=1}^{n} f_0(x_{gh'}) = \frac{\prod_{h'=1}^{n} m(x_{gh'})}{(1-\pi^*)^n}\text{Lfdr}_{g\cdot}(\pi^*) \tag{A.4}$$

$$\mathbf{X}_{g\cdot}|\theta_{g\cdot} = 1 \sim \frac{\prod_{h'=1}^{n} m(x_{gh'})}{1-(1-\pi^*)^n}\left\{1-\text{Lfdr}_{g\cdot}(\pi^*)\right\} \tag{A.5}$$

Hence combining these two equations,

$$\mathbf{X}_{g\cdot} \sim \prod_{h'=1}^{n} m(x_{gh'}) \left\{ \left(\frac{1-\pi}{1-\pi^*}\right)^n \text{Lfdr}_{g\cdot}(\pi^*) + \frac{1-(1-\pi)^n}{1-(1-\pi^*)^n}\left(1-\text{Lfdr}_{g\cdot}(\pi^*)\right) \right\} \tag{A.6}$$

Dividing (A.4) by (A.6) proves (b)(i), i.e.,

$$\Pr(\theta_{g\cdot} = 0|\mathbf{X}) = \text{Lfdr}_{g\cdot}(\boldsymbol{\pi}) = \frac{\text{Lfdr}_{g\cdot}(\pi^*)}{\text{Lfdr}_{g\cdot}(\pi^*) + \lambda_r\left(1-\text{Lfdr}_{g\cdot}(\pi^*)\right)} \tag{A.7}$$

$$\text{where } \lambda_r = \frac{1-(1-\pi)^n}{(1-\pi)^n} \cdot \frac{(1-\pi^*)^n}{1-(1-\pi^*)^n}$$

Likewise, we can derive (b)(ii), i.e.,

$$\Pr(\theta_{\cdot h} = 0|\mathbf{X}) = \text{Lfdr}_{\cdot h}(\boldsymbol{\pi}) = \frac{\text{Lfdr}_{\cdot h}(\pi^*)}{\text{Lfdr}_{\cdot h}(\pi^*) + \lambda_c\left(1-\text{Lfdr}_{\cdot h}(\pi^*)\right)} \tag{A.8}$$

$$\text{where } \lambda_c = \frac{1-(1-\pi)^m}{(1-\pi)^m} \cdot \frac{(1-\pi^*)^m}{1-(1-\pi^*)^m}$$

At the intersection of the gth row and hth column, if either are insignificant,

$$\begin{aligned}\Pr(\theta_{g\cdot}=0,\theta_{\cdot h}=0,\mathbf{X}) &= \frac{1}{(1-\pi)f_0(x_{gh})}\cdot\prod_{g'=1}^{m}(1-\pi)f_0(x_{g'h})\prod_{h'=1}^{n}(1-\pi)f_0(x_{gh'}) \\ &= M\cdot\left(\frac{1-\pi}{1-\pi^*}\right)^{m+n-1}\frac{\mathrm{Lfdr}_{g\cdot}(\pi^*)\mathrm{Lfdr}_{\cdot h}(\pi^*)}{\mathrm{Lfdr}_{gh}(\pi^*)} \\ &= M\cdot\left(\frac{1-\pi}{1-\pi^*}\right)^{m+n-1}L(\pi^*) \end{aligned} \tag{A.9}$$

In order to compute $\Pr(\theta_{g\cdot}=0,\theta_{\cdot h}=1,\mathbf{X})$, consider the hidden states of the member hypotheses of the gth row and hth column. Since $\theta_{g\cdot}=0$, the joint distribution of the random observations is $\prod_{h'=1}^{n} f_0(x_{gh'})$. Specifically $\gamma_{gh}=0$ with probability 1. Besides, since $\theta_{\cdot h}=1$, it implies that at least one of the remaining $(m-1)$ hidden states in the hth column is significant. Hence the joint distribution of the random observations in the hth column is given as $\frac{\prod_{g'\neq g}^{m} m(x_{g'h})}{1-(1-\pi^*)^{m-1}}\left\{1-\frac{\mathrm{Lfdr}_{\cdot h}(\pi^*)}{\mathrm{Lfdr}_{gh}(\pi^*)}\right\}$.

$$\begin{aligned}\Pr(\mathbf{X}|\theta_{g\cdot}=0,\theta_{\cdot h}=1) &= \prod_{h=1}^{n} f_0(x_{gh})\cdot\frac{\prod_{g'\neq g}^{m} m(x_{g'h})}{1-(1-\pi^*)^{m-1}}\left\{1-\frac{\mathrm{Lfdr}_{\cdot h}(\pi^*)}{\mathrm{Lfdr}_{gh}(\pi^*)}\right\} \\ &= \frac{M}{(1-\pi^*)^n}\{\mathrm{Lfdr}_{g\cdot}(\pi^*)-L(\pi^*)\}\end{aligned}$$

$$\text{Hence } \Pr(\theta_{g\cdot}=0,\theta_{\cdot h}=1,\mathbf{X}) = M\left(\frac{1-\pi}{1-\pi^*}\right)^{n}\frac{1-(1-\pi)^{m-1}}{1-(1-\pi^*)^{m-1}}\{\mathrm{Lfdr}_{g\cdot}(\pi^*)-L(\pi^*)\} \tag{A.10}$$

Similarly,

$$\Pr(\theta_{g\cdot}=1,\theta_{\cdot h}=0,\mathbf{X}) = M\left(\frac{1-\pi}{1-\pi^*}\right)^{m}\frac{1-(1-\pi)^{n-1}}{1-(1-\pi^*)^{n-1}}\{\mathrm{Lfdr}_{\cdot h}(\pi^*)-L(\pi^*)\} \tag{A.11}$$

We can define $K(\pi)=(1-\pi)^m+(1-\pi)^n-(1-\pi)^{m+n-1}$, i.e., the probability that states of either or both of a particular row and column is/are insignificant similar to $K(\pi^*)$. From (A.3), we can derive that

$$\Pr(\theta_{g\cdot} \cdot \theta_{\cdot h} = 1, \mathbf{X}) = M\frac{1-K(\pi)}{1-K(\pi^*)}(1 - \mathrm{Lfdr}_{g\cdot,\cdot h}(\pi^*)) \tag{A.12}$$

Adding equations (A.9) through (A.12) gives

$$\begin{aligned}\Pr(\mathbf{X}) = M\left(\frac{1-\pi}{1-\pi^*}\right)^{m+n-1} & \big[L(\pi^*) + \lambda_r^{'}\{\mathrm{Lfdr}_{g\cdot}(\pi^*) - L(\pi^*)\} \\ & + \lambda_c^{'}\{\mathrm{Lfdr}_{\cdot h}(\pi^*) - L(\pi^*)\} + \lambda_{rc}(1 - \mathrm{Lfdr}_{g\cdot,\cdot h}(\pi^*))\big]\end{aligned} \tag{A.13}$$

(A.9)/(A.13) proves $\Pr(\theta_{g\cdot} = 0, \theta_{\cdot h} = 0|\mathbf{X}) = \mathrm{L}_{gh}(\boldsymbol{\pi})$

$$= \frac{L(\pi^*)}{L(\pi^*) + \lambda_r^{'}\{\mathrm{Lfdr}_{g\cdot}(\pi^*) - L(\pi^*)\} + \lambda_c^{'}\{\mathrm{Lfdr}_{\cdot h}(\pi^*) - L(\pi^*)\} + \lambda_{rc}(1 - \mathrm{Lfdr}_{g\cdot,\cdot h}(\pi^*))} \tag{A.14}$$

and on simplification of the denominator, the claim (b)(iii), where $\lambda_c^{'} = \frac{1-(1-\pi)^{n-1}}{(1-\pi)^{n-1}} \cdot \frac{(1-\pi^*)^{n-1}}{1-(1-\pi^*)^{n-1}}$ and $\lambda_r^{'} = \frac{1-(1-\pi)^{m-1}}{(1-\pi)^{m-1}} \cdot \frac{(1-\pi^*)^{m-1}}{1-(1-\pi^*)^{m-1}}$.

It is to be noted that $\lambda_c^{'}$ and $\lambda_{rc}^{'}$ resemble in structure λ_c and λ_r respectively.

Further, since

$$\begin{aligned}\mathrm{Lfdr}_{g\cdot,\cdot h}(\boldsymbol{\pi}) &= \Pr(\theta_{g\cdot} \cdot \theta_{\cdot h} = 0|\mathbf{X}) \\ &= \Pr(\theta_{g\cdot} = 0|\mathbf{X}) + \Pr(\theta_{\cdot h} = 0|\mathbf{X}) - \Pr(\theta_{g\cdot} = \theta_{\cdot h} = 0|\mathbf{X}),\end{aligned}$$

the expression in (b)(iv) automatically follows.