Adaptive Controls of FWER and FDR Under Block Dependence

Wenge Guo
Department of Mathematical Sciences
New Jersey Institute of Technology
Newark, NJ 07102, U.S.A.
Email: wenge.guo@njit.edu

Sanat Sarkar
Department of Statistics, Temple University
Philadelphia, PA 19122, U.S.A.

November 6, 2016

Abstract

Often in multiple testing, the hypotheses appear in non-overlapping blocks with the associated p-values exhibiting dependence within but not between blocks. We consider adapting the Benjamini-Hochberg method for controlling the false discovery rate (FDR) and the Bonferroni method for controlling the familywise error rate (FWER) to such dependence structure without losing their ultimate controls over the FDR and FWER, respectively, in a non-asymptotic setting. We present variants of conventional adaptive Benjamini-Hochberg and Bonferroni methods with proofs of their respective controls over the FDR and FWER. Numerical evidence is presented to show that these new adaptive methods can capture the present dependence structure more effectively than the corresponding conventional adaptive methods. This paper offers a solution to the open problem of constructing adaptive FDR and

FWER controlling methods under dependence in a non-asymptotic setting and providing real improvements over the corresponding non-adaptive ones.

KEY WORDS: Adaptive Benjamini-Hochberg method, adaptive Bonferroni method, false discovery rate, familywise error rate, multiple testing.

1 Introduction

In many multiple hypothesis testing problems arising in modern scientific investigations, the hypotheses appear in non-overlapping blocks. Such block formation is often a natural phenomenon due to the underlying experimental process or can be created based on other considerations. For instance, the hypotheses corresponding to (i) the different time-points in a microarray time-course experiment (Guo, Sarkar and Peddada, 2010; Sun and Wei, 2011) for each gene; or (ii) the phenotypes (or the genetic models) with (or using) which each marker is tested in a genome-wide association study (Lei et al., 2006); or (iii) the conditions (or subjects) considered for each voxel in brain imaging (Heller et al. 2007), naturally form a block. While applying multiple testing in astronomical transient source detection from nightly telescopic image consisting of large number of pixels (each corresponding to a hypotheses), Clements, Sarkar and Guo (2012) considered grouping the pixels into blocks of equal size based on telescope 'point spread function.'

A special type of dependence, which we call block dependence, is the relevant dependence structure that one should take into account while constructing multiple testing procedures in presence of such blocks. This dependence can be simply described by saying that the hypotheses or the corresponding p-values are mostly dependent within but not between blocks. Also known as the clumpy dependence (Storey, 2003), this has been considered mainly in simulation studies to investigate how multiple testing procedures proposed under independence continue to perform under it (Benjamini, Krieger and Yekutieli, 2006; Finner, Dickhaus, and Roters, 2007; Sarkar, Guo and Finner, 2012, and Storey, Taylor and Siegmund, 2004), not in offering FDR or FWER controlling procedures precisely utilizing it. In this article, we focus on constructing procedures controlling the FDR and the FWER that incorporate the block dependence in a non-asymptotic setting in an attempt to improve the corresponding procedures that ignore this structure. More specifically, we consider the Benjamini-Hochberg (BH, 1995) method for the FDR control and the Bonferroni method for the FWER control and adapt them to the data in two ways - incorporating the block dependence and estimating the number of true null hypotheses capturing such dependence.

Adapting to unknown number of true nulls has been a popular way to improve the FDR and FWER controls of the BH and Bonferroni methods, respectively. However, construction of such adaptive methods with proven control of the ultimate FDR or FWER in a non-asymptotic setting and providing real improvements under dependence is an open problem (Benjamini, Krieger and Yekutieli, 2006; Blanchard and Roquaine, 2009). We offer some solutions to this open problem in this paper under a commonly encountered type of dependence, the block dependence.

2 Preliminaries

Suppose that H_{ij}, $i = 1, \ldots, b$; $j = 1, \ldots, s_i$, are the $n = \sum_{i=1}^{b} s_i$ null hypotheses appearing in b blocks of size s_i for the ith block that are to be simultaneously tested based on their respective p-values P_{ij}, $i = 1, \ldots, b$; $j = 1, \ldots, s_i$. Let n_0 of these null hypotheses be true, which for notational convenience will often be identified by $\hat{P}_{ij}$'s. We assume that $\hat{P}_{ij} \sim U(0,1)$ and make the following assumption regarding dependence of P_{ij}'s:

Assumption 1. (Block Dependence) *The rows of p-values $(P_{i1}, \ldots, P_{is_i})$, $i = 1, \ldots, b$, forming the b blocks are independent of each other.*

Under this assumption, the null p-values are independent between but not within blocks. Regarding dependence within blocks, our assumption will depend on whether we want to control the FDR or FWER. More specifically, we develop methods adapting to this block dependence structure and controlling the FDR under positive dependence of the p-values within each block or the FWER under arbitrary dependence of the p-values within each block. The positive dependence condition, when assumed for each i, will be of the type characterized by the following:

$$E\left\{\phi_i(P_{i1}, \ldots, P_{is_i}) \mid \hat{P}_{ij} \leq u\right\} \uparrow u \in (0,1), \tag{1}$$

for each $\hat{P}_{ij}$ and any (coordinatewise) non-decreasing function ϕ_i. This type of positive dependence is commonly encountered and used in multiple testing; see, for instance, Sarkar (2008) for references. We will sometimes refer to block dependence more specifically as *positive block dependence* in case when this dependence defined by (1) in each block or as *arbitrary block dependence* in case of any dependence within each block, to avoid any apparent double meaning.

We will be using two types of multiple testing procedure in this paper - stepup and single-step. Let (P_i, H_i), $i = 1, \ldots, n$, be the pairs of p-value and the corresponding null hypothesis, and $P_{(1)} \leq \cdots \leq P_{(n)}$ be the ordered p-values. Given a set of critical constants $0 \leq \alpha_1 \leq \cdots \leq \alpha_n \leq 1$, a stepup test rejects H_i for all i such that $P_i \leq P_{(R)}$, where $R = \max\{1 \leq i \leq n : P_{(i)} \leq \alpha_i\}$, provided this maximum exists, otherwise, it accepts all the null hypotheses. A single-step test rejects H_i if $P_i \leq c$ for some constant $c \in (0, 1)$.

Let V be the number of falsely rejected among all the R rejected null hypotheses in a multiple testing procedure. Then, the FDR or FWER of this procedure, defined respectively by $\text{FDR} = E(V/\max\{R, 1\})$ or $\text{FWER} = \text{pr}(V \geq 1)$, is said to be controlled at level α, strongly unless stated otherwise, if it is bounded above by α. That is, for for any configuration of true and false null hypotheses, the FDR or FWER of this procedure is less than or equal to α.

The BH method controlling the FDR at level α is a stepup test with the critical constants $\alpha_i = i\alpha/n$; whereas, the Bonferroni method controlling the FWER at level α is a single-step test with the critical constant α/n.

3 Adaptive FDR control under block dependence

The method we propose in this section is based on the idea of adapting the BH method to the block dependence structure without losing the ultimate control over the FDR in a non-asymptotic setting. Our adaptation is done in two steps. First, we adjust it to the block dependence structure and then develop its oracle version given the number of true nulls. Second, we consider the data-adaptive version of this oracle method by estimating n_0 using an estimate that also captures the block dependence.

Towards adjusting the BH method to the block structure, we note that it is natural to first identify blocks that are significant by applying the BH method to simultaneously test the intersection null hypotheses $\tilde{H}_i = \bigcap_{j=1}^{s_i} H_{ij}$, $i = 1, \ldots, b$, based on some block specific p-values, and then go back to each significant block to see which hypotheses in that block are significant. Let $\tilde{P}_i$, $i = 1, \ldots, b$, be the block p-values obtained by combining the p-values in each block through a combination function. Regarding the choice of this combination function, we note that the combination test for $\tilde{H}_i$ based on $\tilde{P}_i$ must allow simultaneous testing of the individual hypotheses H_{ij}, $j = 1, \ldots, s_i$, with a strong control of the FWER. This limits our choice to the Bonferroni adjusted minimum p-value; see also Guo, Sarkar and Peddada (2010). With these in mind, we consider adjusting the BH method as follows:

Definition 1 (Two-stage BH under block dependence)

1. *Choose* $\tilde{P}_i = \bar{s} \min_{1 \le j \le s_i} P_{ij}$ *as the* i*th block* p*-value, for* $i = 1, \ldots, b$, *with* $\bar{s} = \frac{1}{b} \sum_{i=1}^{b} s_i = n/b$ *being the average block size.*

2. *Order the block* p*-values as* $\tilde{P}_{(1)} \le \cdots \le \tilde{P}_{(b)}$, *and find* $B = \max\{1 \le i \le b : \tilde{P}_{(i)} \le i\alpha/b\}$.

3. *Reject* H_{ij} *for all* (i, j) *such that* $\tilde{P}_i \le \tilde{P}_{(B)}$ *and* $P_{ij} \le B\alpha/n$, *provided the above maximum exists, otherwise, accept all the null hypotheses.*

The number of false rejections in this two-stage BH method is given by

$$V = \sum_{i=1}^{b} \sum_{j=1}^{s_i} I(H_{ij} = 0, P_{ij} \le B\alpha/n),$$

where $H_{ij} = 0$ or 1 according to whether it is true or false. So, with R as the total number of rejections, the FDR of this method under block dependence is

$$\begin{aligned} \text{FDR} &= \sum_{i=1}^{b} \sum_{j=1}^{s_i} I(H_{ij} = 0) E\left(\frac{I(P_{ij} \le B\alpha/n)}{\max\{R, 1\}} \right) \\ &\le \sum_{i=1}^{b} \sum_{j=1}^{s_i} I(H_{ij} = 0) E\left(\frac{I(P_{ij} \le B\alpha/n)}{\max\{B, 1\}} \right), \end{aligned} \qquad (2)$$

since $R \geq B$. For each (i, j),

$$\frac{I(P_{ij} \leq B\alpha/n)}{\max\{B, 1\}} = \sum_{k=1}^{b} \frac{I(P_{ij} \leq k\alpha/n, B^{(-i)} = k-1)}{k}, \tag{3}$$

where $B^{(-i)}$ is the number of significant blocks detected by the adjusted BH method based on $\{\tilde{P}_1, \ldots, \tilde{P}_b\} \backslash \{\tilde{P}_i\}$, the $b-1$ block p-values other than the $\tilde{P}_i$, and the critical values $i\alpha/b$, $i = 2, \ldots, b$. Taking expectation in (3) under the block dependence and applying it to (2), we see that

$$\begin{aligned} \text{FDR} &\leq \frac{\alpha}{n} \sum_{i=1}^{b} \sum_{j=1}^{s_i} I(H_{ij} = 0) \sum_{k=1}^{b} \text{pr}(B^{(-i)} = k-1) \\ &= \frac{\alpha}{n} \sum_{i=1}^{b} \sum_{j=1}^{s_i} I(H_{ij} = 0) = \pi_0 \alpha, \end{aligned} \tag{4}$$

where $\pi_0 = n_0/n$. Thus, we have the following result holds:

Result 1. *The above defined two-stage BH method strongly controls the FDR at α under Assumption 1 of arbitrary block dependence.*

If n_0, and hence π_0, were known, the FDR control of this two-stage BH method could be made tighter, from $\pi_0\alpha$ to α, by shrinking each p-value from P_{ij} to $\pi_0 P_{ij}$. This would be the oracle form of the adjusted BH method. Since π_0 is unknown, one would consider using $\widehat{\pi}_0$ to estimate π_0 from the available p-values and then use the estimate $\widehat{\pi}_0$ to define the so-called shrunken or adaptive p-values $Q_{ij} = \widehat{\pi}_0 P_{ij}$ to be used in place of the original p-values in the adjusted BH method. This will be our proposed adaptive BH method.

For estimating n_0 capturing the block dependence structure before defining the adaptive p-values, we consider using an estimate of the form $\widehat{n}_0(\mathbf{P})$ that satisfies the following property. In this property, $\mathbf{P} = ((P_{ij}))$ denotes the set of p-values and $\mathbf{H} = ((H_{ij}))$.

Property 1. *Let $\widehat{n}_0(\mathbf{P})$ be a non-decreasing function of each P_{ij} such that*

$$\sum_{i=1}^{b} \sum_{j=1}^{s_i} I(H_{ij} = 0) E_{DU} \left\{ \frac{1}{\widehat{n}_0(\mathbf{P}^{(-i)}, \mathbf{0})} \right\} \leq 1, \tag{5}$$

where $\mathbf{P}^{(-i)}$ *is the subset of p-values obtained by deleting the ith row,* $\hat{n}_0(\mathbf{P}^{(-i)}, \mathbf{0})$ *is obtained from* $\hat{n}_0(\mathbf{P})$ *by replacing the entries in the ith row of* $\mathbf{P}$ *by zeros, and* E_{DU} *is the expectation under the Dirac-uniform configuration of* $\mathbf{P}^{(-i)}$*, that is, when the p-values in* $\mathbf{P}^{(-i)}$ *that correspond to the false null hypotheses are set to* 0 *and each of the remaining p-values are considered to be uniformly distributed on [0,1].*

We are now ready to define our proposed adaptive BH method in the following:

Definition 2 (Adaptive BH under block dependence)

1. *Consider an estimate* $\widehat{n}_0(\mathbf{P})$ *satisfying Property 1 and define the adaptive p-values* $Q_{ij} = \widehat{\pi}_0 P_{ij}$ *using* $\widehat{\pi}_0 = \widehat{n}_0/n$.

2. *Find* $B^* = \max\{1 \leq i \leq b : \tilde{Q}_{(i)} \leq i\alpha/b\}$*, where* $\tilde{Q}_{(i)} = \widehat{\pi}_0 \tilde{P}_{(i)}$.

3. *Reject* H_{ij} *for all* (i,j) *such that* $\tilde{Q}_i \leq \tilde{Q}_{(B^*)}$ *and* $Q_{ij} \leq B^*\alpha/n$*, provided the maximum in Step 2 exists, otherwise, accept all the null hypotheses.*

Theorem 1 *Consider the block dependence structure in which the p-values are positively dependent as in (1) within each block. The FDR of the above adaptive BH method is controlled at* α *under such positive block dependence.*

A proof of this theorem will be given in Appendix.

What is exactly an estimate satisfying Property 1 that one can use in this adaptive BH? The following result, which is again going to be proved in Appendix, provides an answer to this question.

RESULT 2. *Consider the estimate*

$$\widehat{n}_0^{(1)} = \frac{n - R(\lambda) + s_{\max}}{1 - \lambda}, \tag{6}$$

for any $(2b+3)^{-\frac{2}{b+2}} \leq \lambda < 1$*, where* $s_{\max} = \max_{1 \leq i \leq b} s_i$ *and* $R(\lambda) = \sum_{i=1}^{b} \sum_{j=1}^{s_i} I(P_{ij} \leq \lambda)$ *is the number of p-values in* $\mathbf{P}$ *not exceeding* λ*. It satisfies Property 1 under Assumption 1.*

Based on Theorem 1 and Result 2, we have the following result.

RESULT 3. *The adaptive BH method of the above type based on the estimates* $\widehat{n}_0^{(1)}$ *with* $(2b+3)^{-\frac{2}{b+2}} \leq \lambda < 1$ *strongly controls the FDR at* α *under the positive block dependence considered in Theorem 1.*

Remark 1 When $s_{\max} = 1$, the estimate $\widehat{n}_0^{(1)}$ reduces to

$$\widehat{n}_0^{(0)} = \frac{n - R(\lambda) + 1}{1 - \lambda},$$

the Storey at al.'s (2004) estimate, considered in the context of adaptive FDR control (by Benjamini, Krieger and Yekutieli, 2006; Blanchard and Roquain, 2009; Sarkar, 2008 and Storey, Taylor and Siegmund, 2004) without any block structure. Of course, Result 1 holds for any $\lambda \in (0, 1)$ when $s_{\max} = 1$. Also, in this case, we are basically assuming that the p-values are independent. Thus, as a special case, Result 3 provides the following known result available in the aforementioned papers:

NOTE 1. *The adaptive BH method in Definition 2 based on the estimate $\widehat{n}_0^{(0)}$ controls the FDR at α under independence of the p-values.*

Remark 2 Blanchard and Roquain (2009) presented an adaptive BH method that continues to control the FDR under the same dependence assumption of the p-values as made for the original BH method. Their idea is to estimate n_0 independently through an FWER controlling method before incorporating that into the original BH method. While this adaptive BH method would be applicable to our present context, it does not capture the group structure of the data. Moreover, their simulation studies only show an improvement of their adaptive BH method over the original BH method in very limited situations. Hu, Zhao and Zhou (2010) considered adjusting the BH method in presence of group structure by weighting the p-values according to the relative importance of each group before proposing its adaptive version by estimating these weights. However, this version of the adaptive BH method is known to control the FDR only in an asymptotic setting and under weak dependence.

4 Adaptive FWER control under block dependence

Our proposed method here is based on the idea of adapting the Bonferroni method to the block dependence structure with ultimate control of the FWER in a non-asymptotic setting. Given an estimate $\hat{n}_0$ of n_0 obtained from the available p-values, the Bonferroni method can be adapted to the data through $\hat{n}_0$ by rejecting H_{ij} if

$P_{ij} \leq \alpha/\hat{n}_0$; see, for instance, Finner and Gontscharuk (2009) and Guo (2009). Our method is such an adaptive version of the Bonferroni method, but based on an estimate of n_0 satisfying Property 1 that captures the block dependence.

Definition 3 (Adaptive Bonferroni under block dependence)

1. *Define an estimate $\widehat{n}_0(\mathbf{P})$ satisfying Property 1.*
2. *Reject H_{ij} if $P_{ij} \leq \alpha/\widehat{n}_0(\mathbf{P})$.*

Theorem 2 *Consider the block dependence structure in which the p-values within each block are arbitrarily dependent. The FWER of the above adaptive Bonferroni method is controlled at α under such arbitrary block dependence.*

This will be proved in Appendix. Based on Theorem 2 and Result 2, we have the following result.

RESULT 4. *The adaptive Bonferroni method of the above type based on the estimates $\widehat{n}_0^{(1)}$ with $(2b+3)^{-\frac{2}{b+2}} \leq \lambda < 1$ strongly controls the FWER at α under the arbitrary block dependence considered in Theorem 2.*

Remark 3 Similar to what we have said in Remark 1 on adaptive FDR control, the Storey at al.'s (2004) estimate $\widehat{n}_0^{(0)}$ corresponding to the case $s_{\max} = 1$ was also considered in the context of adaptive FWER control (by Finner and Gontscharuk, 2009, Guo, 2009, and Sarkar, Guo and Finner, 2012), of course without any block structure. Thus, as a special case of Result 4, we get the following result derived in these papers:

NOTE 2. *The adaptive Bonferroni method in Definition 3 based on the estimate $\widehat{n}_0^{(0)}$ controls the FWER at α under independence of the p-values.*

5 Simulation studies

We performed simulation studies to investigate the following questions:

Q1. How does the newly suggested adaptive BH method based on the estimate $\widehat{n}_0^{(1)}$ perform in terms of the FDR control and power with respect to the block size s, the parameter λ, and the strength of dependence among the p-values compared to the original BH method and the two existing adaptive BH methods in Storey et al. (2004) and Benjamini et al. (2006)?

Q2. How does the newly suggested adaptive Bonferroni method based on the estimate $\widehat{n}_0^{(1)}$ perform in terms of the FWER control and power with respect to the block size s, the parameter λ, and the strength of dependence among the p-values compared to the original Bonferroni method and the existing adaptive Bonferroni method based on the estimate $\widehat{n}_0^{(0)}$?

To simulate the values of FDR (or FWER) and average power, the expected proportion of false nulls that are rejected, for each of the methods referred to in Q1 and Q2, we first generated n block dependent normal random variables $N(\mu_i, 1), i = 1, \ldots, n$, with n_0 of the μ_i's being equal to 0 and the rest being equal to $d = \sqrt{10}$, and a correlation matrix $\Gamma = I_{\frac{n}{s}} \otimes [(1-\rho)I_s + \rho 1_s 1_s']$ with the block size s and non-negative correlation coefficient ρ within each block. We then applied each method to the generated data to test $H_i : \mu_i = 0$ against $K_i : \mu_i \neq 0$ simultaneously for $i = 1, \ldots, n$, at level $\alpha = 0.05$. We repeated the above two steps $2,000$ times.

In the simulations on adaptive BH methods, we set $n = 240, n_0 = 120, s = 2, 3, 4,$ or 6 and $\lambda = 0.2, 0.5,$ or 0.8. Thus, $n = 240$ block dependent normal random variables $N(\mu_i, 1), i = 1, \ldots, n$, are generated and grouped into $b = 120, 80, 60,$ or 40 blocks. When $s = 2, 4,$ or 6, half of the μ_i's in each block are 0 while the rest are $d = \sqrt{10}$. When $s = 3$, one μ_i is 0 while the rest are $d = \sqrt{10}$ in each of the first 40 blocks, and two μ_i's are 0 while the rest are $d = \sqrt{10}$ in each of the remaining 40 blocks. Similarly, in the simulations on adaptive Bonferroni methods, we set $n = 100, n_0 = 50$, $s = 2, 4, 10,$ or 20, and $\lambda = 0.2, 0.5,$ or 0.8. Thus, $n = 100$ block dependent normal random variables $N(\mu_i, 1), i = 1, \ldots, n$, are generated and grouped into $b = 50, 25, 10,$ or 5 blocks, with half of the μ_i's in each block being 0 while the rest $d = \sqrt{10}$.

The following are the observations from the above simulations:

From Figure 1 and 2: The simulated FDRs and average powers for the three adaptive BH methods remain unchanged with increasing ρ for different values of s and λ. For small s and different λ, all these three adaptive BH methods seem to be more powerful than the conventional BH method. However, when s is large, the

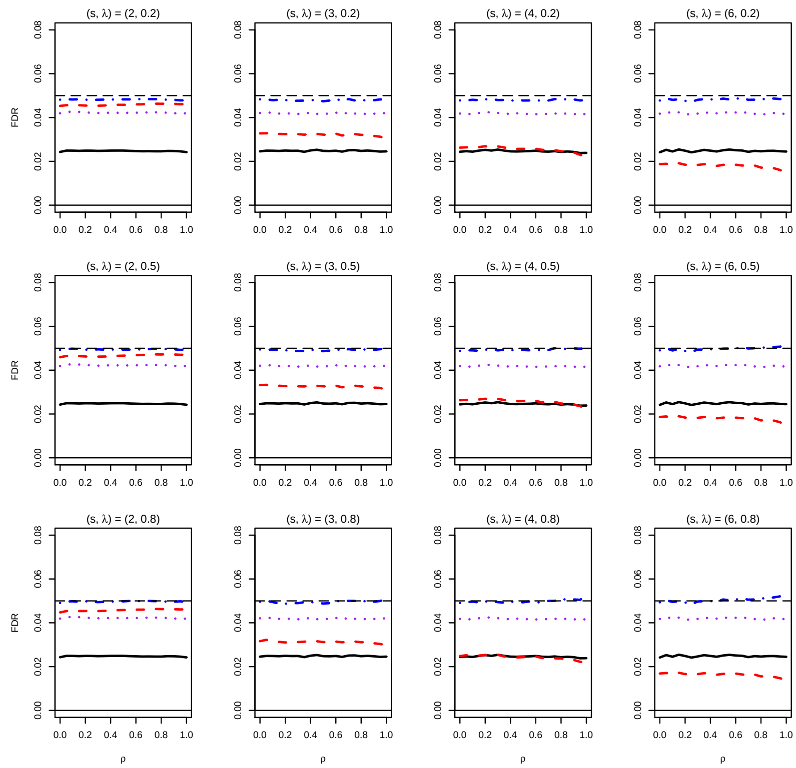

Figure 1: Simulated FDRs of the four multiple testing methods – the original BH and the three adaptive BH methods (adBH1, based on $\widehat{n}_0^{(0)}$; adBH2, based on $\widehat{n}_0^{(1)}$; adBH3, the adaptive BH method introduced in Benjamini et al, 2006) with $n = 240, n_0 = 120$, $s = 2, 3, 4$ or 6, and $\lambda = 0.2, 0.5$ or 0.8 at level $\alpha = 0.05$. [BH – solid; adBH1 – dot-dashes; adBH2 – long dashes; adBH3 – dotted.]

new adaptive method seems to lose its edge over the conventional BH method.

From Figure 3 and 4: When s and λ are both small, both adaptive Bonferroni methods slightly lose the control over the FWER for most values of ρ; however, when λ is chosen to be large, the FWER of the new adaptive method is controlled at α with increasing ρ, whereas the existing adaptive method still loses control of the FWER. When s is moderate or large, the new adaptive method maintains a control over the FWER whatever be the ρ, whereas the existing adaptive method can lose control over the FWER for some values of ρ. In addition, comparing the power performances of the two adaptive methods along with their FWER control,

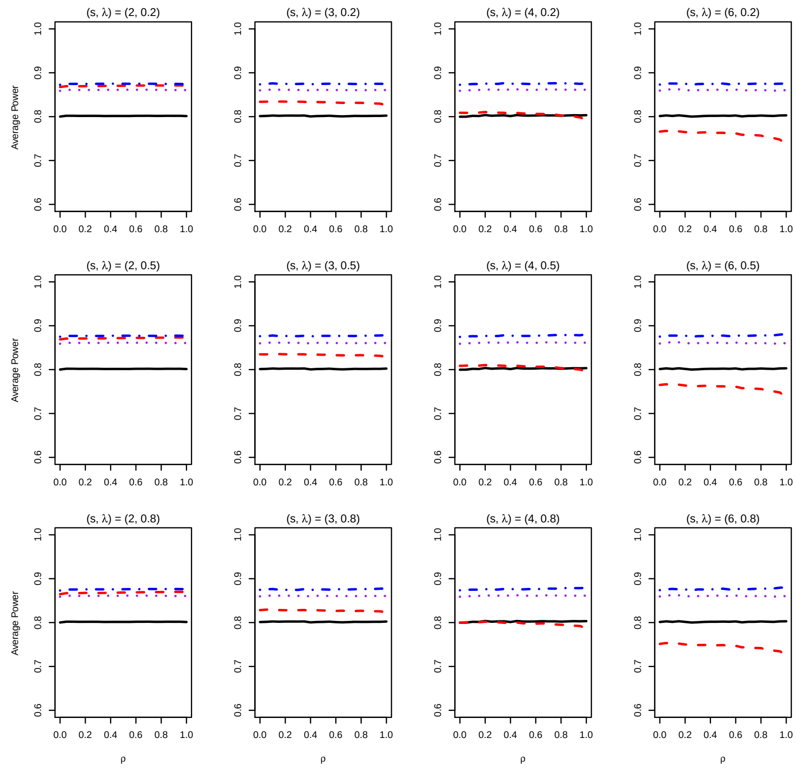

Figure 2: Average powers of the four multiple testing methods – the original BH and the three adaptive BH methods (adBH1, based on $\widehat{n}_0^{(0)}$; adBH2, based on $\widehat{n}_0^{(1)}$; adBH3, the adaptive BH method introduced in Benjamini et al, 2006) with $n = 240, n_0 = 120$, $s = 2, 3, 4$ or 6, and $\lambda = 0.2, 0.5$ or 0.8 at level $\alpha = 0.05$. [BH – solid; adBH1 – dot-dashes; adBH2 – long dashes; adBH3 – dotted.]

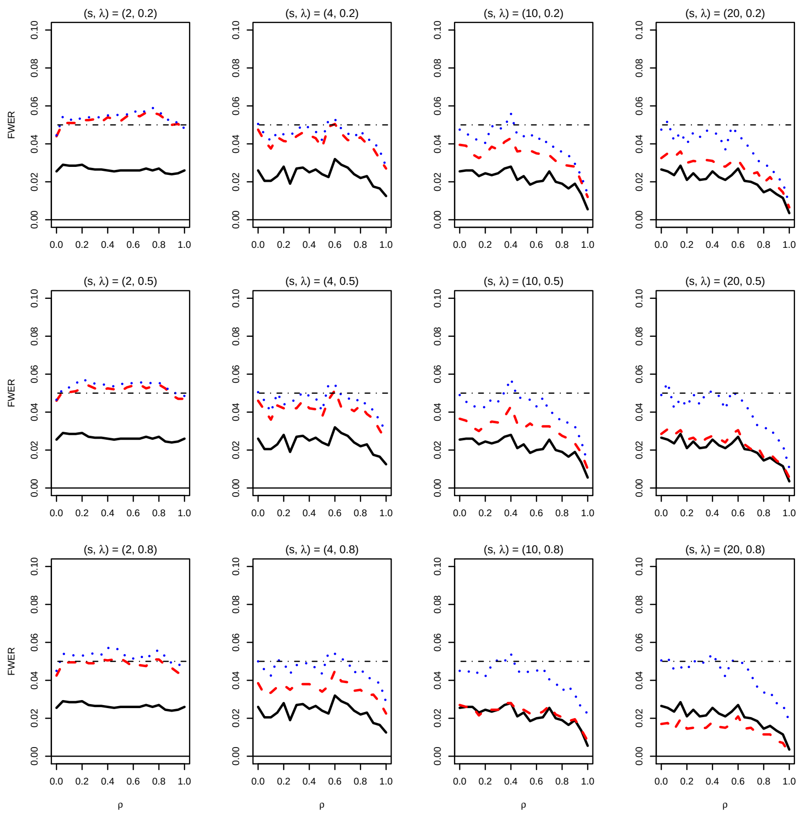

Figure 3: Simulated FWERs of the three multiple testing methods – the original Bonferroni method (Bonf.) and the two adaptive Bonferroni methods (adBon1, based on $\widehat{n}_0^{(0)}$; adBon2, based on $\widehat{n}_0^{(1)}$) with $n = 100, n_0 = 50$, $s = 2, 4, 10$ or 20, and $\lambda = 0.2, 0.5$ or 0.8 at level $\alpha = 0.05$. [Bonf – solid; adBon1 – dotted; adBon2 – long dashes]

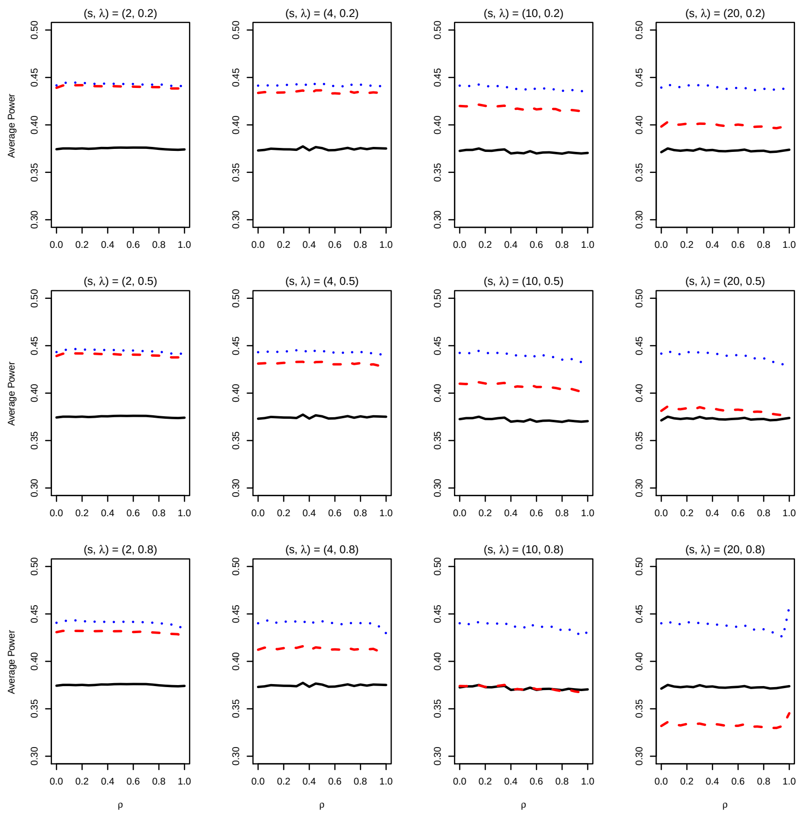

Figure 4: Average powers of the three multiple testing methods – the original Bonferroni method (Bonf.) and the two adaptive Bonferroni methods (adBon1, based on $\widehat{n}_0^{(0)}$; adBon2, based on $\widehat{n}_0^{(1)}$) with $n = 100, n_0 = 50$, $s = 2, 4, 10$ or 20, and $\lambda = 0.2, 0.5$ or 0.8 at level $\alpha = 0.05$. [Bonf – solid; adBon1 – dotted; adBon2 – long dashes]

it is clear that the new method is a better choice as an adaptive version of the Bonferroni method under block dependence than the existing one when s is not very large. However, when s is very large, the new method loses its edge over the existing one.

6 Concluding remarks

Construction of adaptive multiple testing methods with proven control of the ultimate FDR or FWER under dependence in non-asymptotic setting is an open problem. In this paper, we have offered a solution to this open problem under a commonly assumed form of dependence, the block dependence. We have developed new adaptive BH method with proven FDR control under positive block dependence and new adaptive Bonferroni method with proven FWER control under arbitrary block dependence. They often provide real improvements over the corresponding conventional BH and Bonferroni methods.

The type of block dependence structure we consider here is often seen in real applications. It perfectly fits in genetic research where the locations are independent on different chromosomes but dependent inside the same chromosome. It also arises in the context of simultaneous testing of multiple families of hypotheses, which is often considered in large scale data analysis in modern scientific investigations, such as DNA microarray and fMRI studies. Each family of null hypotheses here can be interpreted as a block.

Benjamini and Bogomolov (2014) recently discussed a related problem of testing multiple families of hypotheses and developed a related procedure: Use the BH procedure across families, and then use the Bonferroni procedure within the selected families, with the B/b adjustment, where B is the number of the selected families and b is the number of the tested families. However, in the aforementioned paper, the objective is to control a general average error rate over the selected families including average FDR and FWER instead of the overall FDR and FWER, which is different from ours. Also, there is no explicit discussions of adaptive procedures in that paper as in the methods suggested in this paper. It would be interesting to investigate the connection between the theory and methods developed in this paper and those in aforementioned paper.

Acknowledgements

The research of the first author is supported by the NSF Grants DMS-1006021, DMS-1309162 and the research of the second author is supported by the NSF Grant DMS-1006344, DMS-1309273.

Appendix

Proof of Theorem 1. Proceeding as in showing [in (4)] that the FDR of the adjusted BH method is bounded above by $\pi_0\alpha$, we first have

$$\mathrm{FDR} \;\leq\; \sum_{i=1}^{b}\sum_{j=1}^{s_i} I(H_{ij}=0)\sum_{k=1}^{b}\frac{\mathrm{pr}(Q_{ij}\leq k\alpha/n, B^{*(-i)}=k-1)}{k}, \tag{7}$$

where $B^{*(-i)}$ is the number of significant blocks detected by the BH method based on the $b-1$ block specific adaptive p-values $\{\tilde{Q}_1,\ldots,\tilde{Q}_b\}\setminus\{\tilde{Q}_i\}$ and the critical values $i\alpha/b$, $i=2,\ldots,b$. For each (i,j),

$$\begin{aligned}
& \frac{1}{k}I(H_{ij}=0)\sum_{k=1}^{b}\mathrm{pr}\left(Q_{ij}\leq k\alpha/n, B^{*(-i)}=k-1\right)\\
= \; & \frac{1}{k}I(H_{ij}=0)\sum_{k=1}^{b}\mathrm{pr}\left(P_{ij}\leq k\alpha/\widehat{n}_0(\mathbf{P}), B^{*(-i)}=k-1\right)\\
\leq \; & \frac{1}{k}I(H_{ij}=0)\sum_{k=1}^{b}\mathrm{pr}\left(P_{ij}\leq k\alpha/\widehat{n}_0(\mathbf{P}^{(-i)},\mathbf{0}), B^{*(-i)}=k-1\right)\\
\leq \; & \alpha E\left\{\frac{I(H_{ij}=0)}{\widehat{n}_0(\mathbf{P}^{(-i)},\mathbf{0})}\sum_{k=1}^{b}\mathrm{pr}\left(B^{*(-i)}=k-1 \;\middle|\; P_{ij}\leq k\alpha/\widehat{n}_0(\mathbf{P}^{(-i)},\mathbf{0}),\mathbf{P}^{(-i)}\right)\right\}.
\end{aligned} \tag{8}$$

Now,

$$
\begin{aligned}
& \sum_{k=1}^{b} \mathrm{pr}\left(B^{*(-i)} = k-1 \;\middle|\; P_{ij} \le k\alpha/\widehat{n}_0(\mathbf{P}^{(-i)}, \mathbf{0}), \mathbf{P}^{(-i)}\right) \\
= & \sum_{k=1}^{b} \mathrm{pr}\left(B^{*(-i)} \ge k-1 \;\middle|\; P_{ij} \le k\alpha/\widehat{n}_0(\mathbf{P}^{(-i)}, \mathbf{0}), \mathbf{P}^{(-i)}\right) - \\
& \qquad \sum_{k=1}^{b-1} \mathrm{pr}\left(B^{*(-i)} \ge k \;\middle|\; P_{ij} \le k\alpha/\widehat{n}_0(\mathbf{P}^{(-i)}, \mathbf{0}), \mathbf{P}^{(-i)}\right) \\
\le & \sum_{k=1}^{b} \mathrm{pr}\left(B^{*(-i)} \ge k-1 \;\middle|\; P_{ij} \le k\alpha/\widehat{n}_0(\mathbf{P}^{(-i)}, \mathbf{0}), \mathbf{P}^{(-i)}\right) - \\
& \qquad \sum_{k=1}^{b-1} \mathrm{pr}\left(B^{*(-i)} \ge k \;\middle|\; P_{ij} \le (k+1)\alpha/\widehat{n}_0(\mathbf{P}^{(-i)}, \mathbf{0}), \mathbf{P}^{(-i)}\right) \\
= & \; \mathrm{pr}\left(B^{*(-i)} \ge 0 \;\middle|\; P_{ij} \le \alpha/\widehat{n}_0(\mathbf{P}^{(-i)}, \mathbf{0}), \mathbf{P}^{(-i)}\right) = 1.
\end{aligned}
\tag{9}
$$

The validity of the inequality in (9) can be argued as follows: Since $(P_{i1}, \ldots, P_{is_i})$ is independent of $\mathbf{P}^{(-i)}$ and $I(B^{*(-i)} \ge k)$ is decreasing in P_{ij}'s, the conditional probability

$$\mathrm{pr}\left(B^{*(-i)} \ge k \;\middle|\; P_{ij} \le l\alpha/\widehat{n}_0(\mathbf{P}^{(-i)}, \mathbf{0}), \mathbf{P}^{(-i)}\right),$$

considered as a function of l, with k and $\mathbf{P}^{(-i)}$ being fixed, is of the form

$$g(l) = E\left\{\phi\left(P_{i1}, \ldots, P_{is_i}\right) \mid P_{ij} \le lu\right\},$$

for a decreasing function ϕ and a constant $u > 0$. From the positive dependence condition assumed in the theorem, we note that $g(l)$ is decreasing in l, and hence $g(k+1) \le g(k)$.

From (7)-(9), we finally get

$$\mathrm{FDR} \le \alpha \sum_{i=1}^{b} \sum_{j=1}^{s_i} E\left\{\frac{I(H_{ij}=0)}{\widehat{n}_0(\mathbf{P}^{(-i)}, \mathbf{0})}\right\} \le \alpha, \tag{10}$$

which proves the desired result. □

Proof of Result 2. Before we proceed to prove this result, we state two lemmas in the following that will facilitate our proof. These lemmas will be proved later after we finish proving the result.

Lemma 1 *Given a $p \times q$ matrix $\mathbf{A} = ((a_{ij}))$, where $a_{ij} = 0$ or 1 and $\sum_{i=1}^{p} \sum_{j=1}^{q} a_{ij} = m$, the entries of $\mathbf{A}$ can be always rearranged to form a new $p \times q$ matrix $\mathbf{B} = ((b_{ij}))$ in such a way that, for each $j = 1, \ldots, q$, the entries in the jth column of $\mathbf{B}$ are the entries of $\mathbf{A}$ in different rows, $\sum_{i=1}^{p} b_{ij} = \lfloor \frac{m}{q} \rfloor$ or $\lfloor \frac{m}{q} \rfloor + 1$, and $\sum_{i=1}^{p} \sum_{j=1}^{q} b_{ij} = m$.*

Lemma 2 *The function $f(x) = (2x+3)^{-\frac{2}{x+2}}$ is increasing in $x \geq 1$ and $f(x) \leq f(1)$ for all $0 \leq x \leq 1$.*

We are now ready to prove the result. First, note that the result is unaffected if we augment $\mathbf{P}$ to a complete $b \times s_{\max}$ matrix by adding $s_{\max} - s_i$ more cells in the ith row containing only 0's and assuming that the H_{ij}'s corresponding to these additional zero p-values are all equal to 1, for each $i = 1, \ldots, b$. In other words, we will assume without any loss of generality, while proving this result, that $\mathbf{P}$ is a $b \times s_{\max}$ matrix with $s_{\max} - s_i$ entries in the ith row being identically zero. Let $s_{\max} = s$ for notational convenience.

Consider the expectation

$$E_{DU} \left\{ \frac{1}{\widehat{n}_0^{(1)}(\mathbf{P}^{(-i)}, \mathbf{0})} \right\},$$

in terms of $\mathbf{P}^{(-i)}$. Let $\mathbf{H}^{(-i)}$ be the sub-matrix of $\mathbf{H}$ corresponding to $\mathbf{P}^{(-i)}$. Since this expectation remains unchanged under the type of rearrangements considered in Lemma 1 for $\mathbf{H}^{(-i)}$, we can assume without any loss of generality that the number of true null p-values in the jth column of $\mathbf{P}^{(-i)}$ is $n_{0j}^{(-i)} = \lfloor \frac{n_0 - m_i}{s} \rfloor$ or $\lfloor \frac{n_0 - m_i}{s} \rfloor + 1$ for each $j = 1, \ldots, s$, where $m_i = \sum_{j=1}^{s} I(H_{ij} = 0)$.

Let $\widehat{W}_j^{(-i)}(\lambda) = \sum_{i' (\neq i)=1}^{b} I(H_{i'j} = 0, P_{i'j} > \lambda)$, for $j = 1, \ldots, s$. Under Assumption 1 and the Dirac-uniform configuration of $\mathbf{P}^{(-i)}$, $\widehat{W}_j^{(-i)}(\lambda) \sim \text{Bin}(n_{0j}^{(-i)}, 1 - \lambda)$. So, we have

$$\begin{aligned} E_{DU} \left\{ \frac{1}{\widehat{n}_0^{(1)}(\mathbf{P}^{(-i)}, \mathbf{0})} \right\} &= E \left\{ \frac{1 - \lambda}{\sum_{j=1}^{s} \left[\widehat{W}_j^{(-i)}(\lambda) + 1 \right]} \right\} \\ \leq \frac{1}{s^2} \sum_{j=1}^{s} E \left\{ \frac{1 - \lambda}{\widehat{W}_j^{(-i)}(\lambda) + 1} \right\} &= \frac{1}{s^2} \sum_{j=1}^{s} \frac{1 - \lambda^{n_{0j}^{(-i)} + 1}}{n_{0j}^{(-i)} + 1}, \end{aligned} \tag{11}$$

with the first inequality following from the well-known inequality between the arithmetic and harmonic means or using the Jensen inequality and the second equality following from the result: $E\{(1+X)^{-1}\} = [1-(1-\theta)^{n+1}]/(n+1)\theta$, for $X \sim \mathrm{Bin}(n,\theta)$ (see, for instance, Liu and Sarkar, 2010).

Let $n_0 - m_i = (a_i + \beta_i)s$, for some non-negative integer a_i and $0 \le \beta_i < 1$. Note that

$$a_i s \le n_0 \le (a_i + \beta_i + 1)s. \tag{12}$$

Also, $(1-\beta_i)$ proportion of the s values $n_{0j}^{(-i)}$, $j = 1, \ldots, s$, are all equal to a_i and the remaining β_i proportion are all equal to $a_i + 1$. So, the right-hand side of (11) is equal to

$$\begin{aligned}
& \frac{1}{s}\left[\frac{1-\beta_i}{a_i+1}\left(1-\lambda^{a_i+1}\right) + \frac{\beta_i}{a_i+2}\left(1-\lambda^{a_i+2}\right)\right] \le \frac{1}{s}\left[\frac{1-\beta_i}{a_i+1} + \frac{\beta_i}{a_i+2}\right]\left(1-\lambda^{a_i+2}\right) \\
= \quad & \frac{(a_i+2-\beta_i)(1-\lambda^{a_i+2})}{s(a_i+1)(a_i+2)} \le \frac{(a_i+1+\beta_i)(a_i+2-\beta_i)(1-\lambda^{a_i+2})}{n_0(a_i+1)(a_i+2)} \\
= \quad & \frac{1}{n_0}\left[1 + \frac{\beta_i(1-\beta_i)}{(a_i+1)(a_i+2)}\right]\left(1-\lambda^{a_i+2}\right) \le \frac{1}{n_0}\left[1 + \frac{1}{4(a_i+1)(a_i+2)}\right]\left(1-\lambda^{a_i+2}\right).
\end{aligned}$$

Here, the second inequality follows from (12). The desired inequality (5) then holds for this estimate if

$$\left[1 + \frac{1}{4(a_i+1)(a_i+2)}\right]\left(1-\lambda^{a_i+2}\right) \le 1,$$

which is true if and only if

$$\lambda \ge [1 + 4(a_i+1)(a_i+2)]^{-\frac{1}{a_i+2}} = (2a_i+3)^{-\frac{2}{a_i+2}}. \tag{13}$$

Let $f(a_i) = (2a_i+3)^{-\frac{2}{a_i+2}}$. As seen from (12), $a_i \le n_0/s \le b$, thus, the inequality $f(b) \ge f(a_i)$ holds for all $a_i \ge 0$, since $f(b) \ge f(a_i)$ if $a_i \ge 1$ and $f(b) \ge f(1) \ge f(a_i)$ if $0 \le a_i \le 1$, due to Lemma 2. So, the inequality (13) holds if $\lambda \ge (2b+3)^{-2/(b+2)}$. This completes our proof of Result 2. □

Proof of Lemma 1. Let $s = (s_1, \ldots, s_q)$ be the column sum vector of $\mathbf{A}$, that is, $s_j = \sum_{i=1}^{p} a_{ij}$, $j = 1, \ldots, q$, and $\sum_{j=1}^{q} s_j = m$. Without any loss of generality, we can assume that $s_1 \ge \ldots \ge s_q$. Consider a given column sum vector $s^* = (s_1^*, \ldots, s_q^*)$

satisfying $s_1^* \geq \ldots \geq s_q^*$, where $s_j^* = \lfloor \frac{m}{q} \rfloor$ or $\lfloor \frac{m}{q} \rfloor + 1$ for $j = 1, \ldots, q$, and $\sum_{j=1}^q s_j^* = m$.

We prove that s^* is majorized by s; that is, for each $k = 1, \ldots, q$,

$$\sum_{j=k}^{q} s_j^* \geq \sum_{j=k}^{q} s_j. \tag{14}$$

Suppose the inequality (14) does not hold for some $k = 1, \ldots, q$. Let $k_1 = \max\{k : \sum_{j=k}^q s_j^* < \sum_{j=k}^q s_j\}$. Since $s_{k_1} > s_{k_1}^*$, thus for each $j = 1, \ldots, k_1 - 1$, $s_j \geq s_{k_1} \geq s_{k_1}^* + 1 \geq \lfloor \frac{m}{q} \rfloor + 1 \geq s_j^*$, implying that

$$\sum_{j=1}^{q} s_j = \sum_{j=1}^{k_1-1} s_j + \sum_{j=k_1}^{q} s_j > \sum_{j=1}^{k_1-1} s_j^* + \sum_{j=k_1}^{q} s_j^* = m,$$

which is a contradiction. So, s^* is majorized by s.

By Theorem 2.1 of Ryser (1957), one can rearrange the 1's in the rows of $\mathbf{A}$ to construct a new $p \times q$ matrix which has the column sum vector s^*. Thus, the desired result follows. □

Proof of Lemma 2. Let $g(x) = \ln f(x) = -\frac{2}{x+2} \ln(2x+3)$ for $x \geq 0$ and $\varphi(u) = \ln u - \frac{1}{u} - 1$ for $u \geq 3$. Thus,

$$g'(x) = \frac{1}{(x+2)^2}\left[2\ln(2x+3) - \frac{4x+8}{2x+3}\right] = \frac{2\varphi(2x+3)}{(x+2)^2}.$$

Note that $\varphi(u)$ is a strictly increasing continuous function in $[3, \infty)$ with $\varphi(3) < 0$ and $\varphi(5) > 0$, thus there exists a unique $u^* \in (3, 5)$ satisfying $\varphi(u^*) = 0$. Let $x^* = \frac{u^*-3}{2}$, then $x^* \in (0, 1)$ and $g'(x^*) = 0$. Thus, $g'(x) < 0$ for $x \in [0, x^*)$ and $g'(x) > 0$ for $x \in (x^*, \infty)$. Based on $x^* < 1$, we have that $g'(x) > 0$ for $x \geq 1$ and $g(x) \leq \max\{g(0), g(1)\} = \max\{-\ln 3, -2\ln 5/3\} = g(1)$ for $0 \leq x \leq 1$. Thus, the desired result follows. □

Proof of Theorem 2. The FWER of the method in this theorem is given by

$$\begin{aligned}
\text{FWER} &= \text{pr}\left\{\bigcup_{i=1}^{b}\bigcup_{j=1}^{s_i}\left(P_{ij} \leq \frac{\alpha I(H_{ij}=0)}{\widehat{n}_0(\mathbf{P})}\right)\right\} \leq \sum_{i=1}^{b}\sum_{j=1}^{s_i} \text{pr}\left\{P_{ij} \leq \frac{\alpha I(H_{ij}=0)}{\widehat{n}_0(\mathbf{P})}\right\} \\
&\leq \sum_{i=1}^{b}\sum_{j=1}^{s_i} \text{pr}\left\{P_{ij} \leq \frac{\alpha I(H_{ij}=0)}{\widehat{n}_0(\mathbf{P}^{(-i)}, \mathbf{0})}\right\} \leq \alpha \sum_{i=1}^{b}\sum_{j=1}^{s_i} E_{DU}\left\{\frac{I(H_{ij}=0)}{\widehat{n}_0(\mathbf{P}^{(-i)}, \mathbf{0})}\right\} \\
&\leq \alpha.
\end{aligned} \tag{15}$$

In (15), the first inequality follows from the Bonferroni inequality, the second and third follow from the non-decreasing property of $\widehat{n}_0$ and that $\hat{P}_{ij} \sim U(0,1)$ and the assumption of arbitrary block dependence, and the fourth follows from the condition (5) satisfied by $\widehat{n}_0$. Thus, the desired result is proved. $\square$

References

[1] Benjamini, Y. & Bogomolov, M. (2014). Selective inference on multiple families of hypotheses. *J. Roy. Statist. Soc. Ser. B* **76**, 297–318.

[2] Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57**, 289-300.

[3] Benjamini, Y, Krieger, K. & Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93**, 491-507.

[4] Benjamini, Y. & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29**, 1165–1188.

[5] Blanchard, G. & Roquain, E. (2009). Adaptive FDR control under independence and dependence. *J. Mach. Learn.* **10**, 2837–2871.

[6] Clements, N., Sarkar, S. & Guo, W. (2012). Astronomical transient detection using grouped p-values and controlling the false discovery rate. In *Statistical Challenges in Modern Astronomy*, edited by Eric D. Feigelson and G. Joseph Babu, Lecture Notes in Statistics, Vol. 209, Part 4, Springer-Verlag, 383-396.

[7] Finner, H., Dickhaus, T. & Roters, M. (2007). Dependency and false discovery rate: Asymptotics. *Ann. Statist.* **35**, 1432-1455.

[8] Finner, H. & Gontscharuk, V. (2009). Controlling the familywise error rate with plug-in estimator for the proportion of true null hypotheses. *J. Roy. Statist. Soc., Ser. B* **71**, 1031–1048.

[9] Gavrilov, Y., Benjamini, Y. & Sarkar, S. K. (2009). An adaptive step-down procedure with proven FDR control. *Ann. Statist.* **37**, 619–629.

[10] Guo, W. (2009). A note on adaptive Bonferroni and Holm procedures under dependence. *Biometrika*, **96**, 1012-1018.

[11] Guo, W., Sarkar, S. & Peddada, S. (2010). Controlling false discoveries in multidimensional directional decisions, with applications to gene expression data on ordered categories. *Biometrics* **66**, 485-492.

[12] Heller R., Golland Y., Malach R. & Benjamini Y. (2007). Conjunction Group Analysis: An alternative to mixed/random effect analysis. *NeuroImage* **37**, 1178-1185.

[13] Hochberg, Y. & Benjamini, Y. (1990). More powerful procedures for multiple significance testing. *Statist. Med.* **9**, 811-818.

[14] Hochberg, Y., & Tamhane, A. C. (1987). *Multiple Comparison Procedures.* Wiley: New York.

[15] Hu, J., Zhao, H. & Zhou, H. (2010). False Discovery Rate Control with Groups. *J. American Statistical Association* **105**, 1215-1227.

[16] Lei S., Chen Y., Xiong D., Li L. & Deng H. (2006) Ethnic difference in osteoporosis-related phenotypes and its potential underlying genetic determination. *Journal of Musculoskelet Neuronal Interact.* **6**, 36-46.

[17] Liu, F. & Sarkar, S. K. (2010). A note on estimating the false discovery rate under mixture model. *Journal of Statistical Planning and Inference* **140**, 1601-1609.

[18] Liu, F. & Sarkar, S. K. (2011). A new adaptive method to control the false discovery rate. *Series in Biostatistics* **4**, World Scientific, 3-26.

[19] Romano, J. P., Shaikh, A. M. & Wolf, M. (2008). Control of the false discovery rate under dependence using the bootstrap and subsampling. *TEST* **17**, 417–442.

[20] Ryser, H. J. (1957). Combinatorial properties of matrices of zeros and ones. *Canad. J. Math.* **9**, 371-377.

[21] Sarkar, S. K. (1998). Some probability inequalities for ordered MTP_2 random variables: a proof of the Simes conjecture. *Ann. Statist.* **26**, 494–504.

[22] Sarkar, S. K. (2002). Some results on false discovery rate in stepwise multiple testing procedures. *Ann. Statist.* **30**, 239–257.

[23] Sarkar, S. K. (2006). False discovery and false nondiscovery rates in single-step multiple testing procedures. *Ann. Statist.* **34**, 394 –415.

[24] Sarkar, S. K. (2008). On methods controlling the false discovery rate (with discussions). *Sankhya* **70**, 135–168.

[25] Sarkar, S. K. & Chang, C-K. (1997). The Simes method for multiple hypothesis testing with positively dependent test statistics. *J. Amer. Statist. Assoc.* **92**, 1601-1608.

[26] Sarkar, S., Guo, W. & Finner, H. (2012). On adaptive procedures controlling the familywise error rate. *Journal of Statistical Planning and Inference*, **142**, 65-78.

[27] Schweder, T. & Spjϕtvoll, E. (1982). Plots of p-values to evaluate many tests simulataneously. *Biometrika* **69**, 493-502.

[28] Storey J. (2003). Comment on 'Resampling-based multiple testing for DNA microarray data analysis' by Ge, Dudoit, and Speed. *Test* **12**, 52-60.

[29] Storey, J. D., Taylor, J. E. & Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. Roy. Statist. Soc., Ser. B* **66**, 187-205.

[30] Sun W & Wei Z. (2011). Multiple testing for pattern identification, with applications to microarray time course experiments. *Journal of the American Statistical Association* **106**, 73-88.