Journal of Statistical Planning and Inference 142 (2012) 65–78

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

On adaptive procedures controlling the familywise error rate

Sanat K. Sarkar [a,*], Wenge Guo [b], Helmut Finner [c]

[a] Temple University, Philadelphia, PA 19122, USA
[b] New Jersey Institute of Technology, Newark, NJ 07102, USA
[c] German Diabetes Center, Düsseldorf, Germany

ARTICLE INFO

Article history:
Received 28 December 2010
Received in revised form
20 June 2011
Accepted 22 June 2011
Available online 7 July 2011

Keywords:
Adaptive methods
Familywise error rate
Multiple testing
Stepdown method
Stepup method

ABSTRACT

The idea of modifying, and potentially improving, classical multiple testing methods controlling the familywise error rate (FWER) via an estimate of the unknown number of true null hypotheses has been around for a long time without a formal answer to the question whether or not such adaptive methods ultimately maintain the strong control of FWER, until Finner and Gontscharuk (2009) and Guo (2009) have offered some answers. A class of adaptive Bonferroni and Šidàk methods larger than considered in those papers is introduced, with the FWER control now proved under a weaker distributional setup. Numerical results show that there are versions of adaptive Bonferroni and Šidàk methods that can perform better under certain positive dependence situations than those previously considered. A different adaptive Holm method and its stepup analog, referred to as an adaptive Hochberg method, are also introduced, and their FWER control is proved asymptotically, as in those papers. These adaptive Holm and Hochberg methods are numerically seen to often outperform the previously considered adaptive Holm method.

1. Introduction

The familywise error rate (FWER), which is the probability of at least one false rejection, is frequently chosen as an overall measure of Type I error to control while testing multiple hypotheses. Several multiple testing methods controlling (strongly) the FWER are available in the literature, such as the Bonferroni, Holm (1979), and Hochberg (1988) methods (see, Hochberg and Tamhane, 1987; Hsu, 1996). Among these, the Bonferroni is one of the most popular, since its FWER control is guaranteed without any requirement for dependence structure of the p-values due to the Bonferroni inequality. It is a single-step method. The Holm is a stepdown method that sequentially applies the Bonferroni method. The Hochberg is the stepup version of the Holm method; that is, it is the stepup method with the same set of critical values as Holm's, although its FWER control is validated by the Simes inequality, not the Bonferroni inequality, of course requiring certain types of dependence structure of the p-values (Simes, 1986; Sarkar, 1998, 2008a; Sarkar and Chang, 1997).

Let H_i, $i=1,\ldots,n$, be the n null hypotheses to be tested based on their respective p-values P_i, $i=1,\ldots,n$. Then, the Bonferroni method rejects H_i if $P_i \le \alpha/n$. However, if n_0, the number of true null hypotheses, were known, one would have considered the cut-off point α/n_0, instead of α/n. This would provide a more powerful method of controlling the FWER than the original Bonferroni method. Since n_0 is typically unknown, a suitable estimate of it could be used in α/n_0, and thus the cut-off point of the Bonferroni method could be modified. This is the general idea towards developing a potentially

* Corresponding author.
E-mail addresses: sanat@temple.edu (S.K. Sarkar), wenge.guo@njit.edu (W. Guo), finner@ddz.uni-duesseldorf.de (H. Finner).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2011.06.022

improved version of the Bonferroni method, which was initially presented by Schweder and Spjøtvoll (1982). Hochberg and Benjamini (1990) referred to such a modification of the Bonferroni method as an adaptive Bonferroni method when the estimate of n_0 is obtained from the existing set of p-values. They proposed an adaptive Bonferroni method using an estimate of n_0 that is different from Schweder and Spjøtvoll (1982). They also proposed an adaptive Holm method, which is a stepdown implementation of their adaptive Bonferroni method, and an adaptive Hochberg method, which is the stepup analog of the adaptive Holm method.

Whether or not these adaptive procedures ultimately control the FWER has long been an open problem until Finner and Gontscharuk (2009) and Guo (2009) have recently offered some answers. With $R(\lambda)=\sum_{i=1}^{n} I(P_i \le \lambda)$, the number of p-values less than or equal to λ, Finner and Gontscharuk (2009) have considered the following estimate of n_0:

$$\hat{n}_0 = \frac{n-R(\lambda)+\kappa}{1-\lambda}, \tag{1}$$

with $\lambda \in (0,1)$ and $\kappa \in \mathcal{R}$ satisfying some restrictions depending on α and λ, and referred to the resulting modification of the Bonferroni method as a Bonferroni plug-in (BPI) test. Assuming that the p-values are independent and identically distributed (iid) as U(0,1) when the corresponding null hypotheses are true, they theoretically established the strong FWER control of BPI for $\kappa \ge 1$ and numerically investigated this control for $\kappa < 1$. They eventually concentrated on the BPI corresponding to

$$\hat{n}_0 = \frac{n-R(\lambda)+1}{1-\lambda}, \tag{2}$$

where $\lambda \in (0,1)$, having checked numerically that the least conservative BPI that exists for a κ slightly less than 1 is not appreciably different from the BPI with $\kappa = 1$. They also presented an appropriate stepdown implementation of the BPI, a stepdown method based on the (random) critical values $\hat{c}_i = \alpha/\min\{\hat{n}_0, n-i+1\}$, $i=1,\ldots,n$, with $\hat{n}_0$ in (1). They numerically noted that there are values of $\kappa > 1$ for which this stepdown method can control the FWER, but such methods may not have much power compared to just BPI, and proved that this stepdown method controls the FWER as $n_0 \to \infty$ when $\kappa = 1$.

Like Hochberg and Benjamini (1990), Guo (2009) referred to a modification of the Bonferroni method using an estimate of n_0 obtained from the existing data as an adaptive Bonferroni method. He considered only the estimates of n_0 as in (2), and a specific dependence model for the p-values that implies the iid U(0,1) distributional structure for the p-values when the corresponding null hypotheses are true. For this adaptive Bonferroni method, he proved using an argument different from Finner and Gontscharuk (2009) that the FWER is strongly controlled. He also presented a stepdown implementation of this adaptive Bonferroni method, referring to it as an adaptive Holm method, which is slightly different from that of Finner and Gontscharuk (2009), and a proof of its strong FWER control asymptotically as $n_0 \to \infty$. His simulation studies comparing the performances of adaptive Bonferroni method with adaptive Holm method and non-adaptive Bonferroni and Holm methods provide similar conclusions as obtained by Finner and Gontscharuk (2009); that is, an adaptive Bonferroni method performs better than its non-adaptive version, and even the usual Holm method, but its stepdown implementation does not result in any significant power improvement.

The rationale behind adapting the Bonferroni method to data through an estimate of n_0 is akin to that in modifying, and hopefully improving, the false discovery rate (FDR) controlling method of Benjamini and Hochberg (1995), the so called BH method. Several different estimates of n_0, including (2), have been considered while constructing adaptive BH methods (Benjamini et al., 2006; Blanchard and Roquain, 2009; Gavrilov et al., 2009; Sarkar, 2006, 2008b; Storey et al., 2004). These adaptive BH methods have been shown to ultimately control the FDR under independence of the p-values. However, there is numerical evidence that the method of Storey et al. (2004) that is based on (2) may often fail to provide a control of the FDR under dependence (Benjamini et al., 2006; Romano et al., 2008). Interestingly, the same disturbing feature of the estimate (2) holds when it is used to construct an adaptive Bonferroni method controlling the FWER, as noted by Finner and Gontscharuk (2009, Fig. 6) with $\lambda = 0.5$. On the other hand, there is numerical evidence that the adaptive BH method of Benjamini et al. (2006) that is based on an estimate of n_0 obtained by applying the BH method at an appropriate level can control the FDR under positive dependence, at least under equicorrelated normal distributional setting (Benjamini et al., 2006; Romano et al., 2008). This motivates our present work, to look into adaptive Bonferroni methods with FWER control based on estimates other than (1) or (2), and including the one in Benjamini et al. (2006).

So, we consider in this paper a larger class of estimates of n_0 than (1) by characterizing this class by a property shared not only by (1), and hence (2), but also estimates close to that in Benjamini et al. (2006) and some other. We establish the strong FWER control of adaptive Bonferroni methods corresponding to this larger class of n_0 estimates under a distributional setting that is slightly weaker than in Finner and Gontscharuk (2009) and Guo (2009), and thus generalize the work in those papers on adaptive Bonferroni methods. More importantly, our numerical investigations show that newer versions of adaptive Bonferroni methods proposed here do indeed perform better than the one based on (2) under positive dependence.

Finner and Gontscharuk (2009) have also considered the Šidàk single-step method rejecting an H_i if $P_i \le 1-(1-\alpha)^{1/n}$, and proved that an adaptive version of it obtained by replacing n by an estimate of n_0 in (1), which they referred to as a variant of BPI, strongly controls the FWER if the null p-values are iid as U(0,1). We also strengthen this result by proving strong FWER control for the larger class of adaptive Šidàk methods that correspond to the present class of n_0 estimates under our distributional setup weaker than in Finner and Gontscharuk (2009). As seen numerically, this class of adaptive Šidàk methods is practically not much different in terms of FWER control and power from its Bonferroni counterparts. Thus, like the Bonferroni method, we now have other versions of adaptive Šidàk method often performing better than the one considered by Finner and Gontscharuk (2009) under positive dependence. We also strengthen the work of both Finner and Gontscharuk (2009) and Guo (2009) related to adaptive Holm method. We consider an estimate of n_0 other than (2) and establish the strong FWER control when $n_0 \to \infty$ not only for the corresponding adaptive version of the Holm method but also for its stepup counterpart, calling it an adaptive Hochberg method. Our numerical calculations show that these newer adaptive Holm and Hochberg methods often have better performance.

The rest of the paper is organized in the following seven sections, including an Appendix. Section 2 presents some preliminaries; Section 3 gives the present class of n_0 estimates; Sections 4 and 5 contain the proposed adaptive FWER methods; Section 6 has the results of simulation studies; Section 7 provides some concluding remarks; and proofs of some supporting lemmas are put in Appendix.

2. Preliminaries

In this section we present some background results. First, let us recall the definitions of stepdown and stepup methods. Let $P_{(1)} \le \cdots \le P_{(n)}$ be the ordered versions of all the p-values, with $H_{(1)}, \ldots, H_{(n)}$ being their corresponding null hypotheses. Then, given a non-decreasing set of constants $0 < c_1 \le \cdots \le c_n < 1$, a stepdown method with these constants as critical values rejects the set of null hypotheses $\{H_{(i)}, i \le i^*_{SD}\}$, where $i^*_{SD} = \max\{1 \le i \le n : P_{(j)} \le c_j \forall j \le i\}$, if the maximum exists, otherwise accepts all the null hypotheses. A stepup method rejects the set of null hypotheses $\{H_{(i)}, i \le i^*_{SU}\}$, where $i^*_{SU} = \max\{1 \le i \le n : P_{(i)} \le c_i\}$, if the maximum exists, otherwise accepts all the null hypotheses. A stepdown or stepup method with a common critical constant is referred to as a single-step method.

Since FWER$=0$, and hence trivially controlled, if $n_0 = 0$, we will assume that $n_0 \ge 1$, and for notational convenience, the p-values corresponding to the true null hypotheses will often be identified by $\tilde{P}_i$, $i = 1, \ldots, n_0$, and their ordered versions by $\tilde{P}_{(1)} \le \cdots \le \tilde{P}_{(n_0)}$. The adaptive methods in this paper are like the aforementioned stepwise methods with critical constants replaced by some random quantities. For a single-step method with a common random rejection threshold $\hat{c}$ for each p-value, the FWER $= \mathrm{pr}(\tilde{P}_{(1)} \le \hat{c})$. For a stepwise method with random critical values $0 < \hat{c}_1 \le \cdots \le \hat{c}_n < 1$, the FWER satisfies, under any distributional setup, the inequality given in the following lemma.

Lemma 1. *For stepdown method*:

$$\mathrm{FWER} \le \mathrm{pr}(\tilde{P}_{(1)} \le \hat{c}_{n-n_0+1}),$$

For stepup method:

$$\mathrm{FWER} \le \mathrm{pr}\left(\bigcup_{i=1}^{n_0} \{\tilde{P}_{(i)} \le \hat{c}_{n-n_0+i}\}\right).$$

Proof. Let V^{SD} and V^{SU} denote, respectively, the number of false rejections in the stepdown and stepup methods. Then, this lemma follows easily by noting that

$$\{V^{SD} = 0\} \supseteq \{\tilde{P}_{(1)} > \hat{c}_{n-n_0+1}\},$$

$$\{V^{SU} = 0\} \supseteq \bigcap_{i=1}^{n_0} \{\tilde{P}_{(i)} > \hat{c}_{n-n_0+i}\}. \qquad \square$$

The only distributional assumption we make about the p-values in this paper is that they are independently distributed when the corresponding null hypotheses are true, with each of these null p-values having a stochastically larger than uniform distribution on [0,1]. In other words, we make the following assumption:

Assumption 1. $\tilde{P}_i$, $i = 1, \ldots, n_0$, are independently distributed with $\mathrm{pr}(\tilde{P}_i \le u) \le u$ for $u \in [0,1]$ and $i = 1, \ldots, n_0$.

This assumption is slightly less restrictive than the condition on the p-values in Finner and Gontscharuk (2009). They considered exactly a uniform distribution on [0,1] for each of the independent null p-values. Also, ours is a more general assumption than the conditional independence model assumed by Guo (2009).

The following proposition provides a foundation for constructing FWER controlling adaptive versions of single-step and stepwise methods that will be discussed in the following sections.

Proposition 1. *Consider a stepwise method with critical values $0 < \hat{c}_1 \le \cdots \le \hat{c}_n < 1$ that are non-increasing functions of the p-values. Then, under Assumption* 1, *the FWER of this methods satisfies the following inequality*:

$$\mathrm{FWER} \le \mathrm{FWER}_{DU(n_0)},$$

where $\mathrm{FWER}_{DU(n_0)}$ is the FWER of this method under the Dirac-uniform configuration, that is, when the p-values corresponding to the false null hypotheses are set to 0 and the remaining ones are iid uniforms on [0,1].

Proof. Replace each of the p-values corresponding to the false null hypotheses in $\hat{c}_i$ by 0 and denote the resulting critical value by $\tilde{c}_i$. Then, since $\hat{c}_i$ is non-increasing in each p-value, $\hat{c}_i \le \tilde{c}_i$ for each i, and hence we have from Lemma 1 that

$$\{V^{SD}=0\} \supseteq \{\tilde{P}_{(1)} > \tilde{c}_{n-n_0+1}\},$$

$$\{V^{SU}=0\} \supseteq \bigcap_{i=1}^{n_0} \{\tilde{P}_{(i)} > \tilde{c}_{n-n_0+i}\}. \tag{3}$$

Now, since the events in the right-hand side of (3) are non-decreasing in $\tilde{P}_i$, $i=1,\ldots,n_0$, we can apply the following results (see, for example, Tong, 1980, p. 121) to see that $1-\text{FWER} \ge 1-\text{FWER}_{DU(n_0)}$, for both stepdown and stepup methods, thus completing the proof. □

Result 1. *Let $X_i \sim F_i$, $i=1,\ldots,m$, be independent random variables, and X_i under F_i be stochastically larger than under G_i in the sense that $\text{pr}_{F_i}(X_i \ge x) \ge \text{pr}_{G_i}(X_i \ge x)$ for every x. Then, for any function $\phi(X_1,\ldots,X_m)$ that is non-decreasing in each X_i, we have*

$$E_{(F_1,\ldots,F_m)}\{\phi(X_1,\ldots,X_m)\} \ge E_{(G_1,\ldots,G_m)}\{\phi(X_1,\ldots,X_m)\}.$$

Remark 1. Proposition 1 generalizes what Finner and Gontscharuk (2009) obtained for some special plug-in tests. It basically says that the Dirac-uniform is the least favorable configuration under Assumption 1 for the FWER of a stepwise method with critical values that are non-increasing in the p-values. In other words, to prove the (strong) FWER control of such a method, it would be enough to establish the (weak) FWER control of the method under this configuration.

Remark 2. It is easy to see that Proportion 1 is actually true for any procedure such that the rejection set grows as the p-values decrease, because setting the false p-values to 0 will only increase the rejection set. Some examples of such procedures include Hommel procedure, fixed-sequence procedure, fallback procedure, and gatekeeping strategies for clinical trials (Dmitrienko et al., 2009). By using Proportion 1, the proofs of the FWER control can be greatly simplified for these procedures.

3. Estimates of n_0

We will introduce in this section a class of estimates of n_0 satisfying certain common conditions before developing our proposed adaptive methods with the FWER control in the following sections. Let us denote the vector $(P_1,\ldots,P_n)$ by $(\mathbf{P}^{(-i)},P_i)$. Also let $H_i=0$ or 1 according to it is true or false. We consider the class of estimates $\hat{n}_0(\mathbf{P}^{(-i)},P_i)$ of n_0 that satisfy the following:

Property 1. *$\hat{n}_0(\mathbf{P}^{(-i)},P_i)$ is non-decreasing in each P_i and, for any $1 \le n_0 \le n$:*

$$\sum_{i=1}^{n} I(H_i=0) E_{DU(n_0)}\left\{\frac{1}{\hat{n}_0(\mathbf{P}^{(-i)},0)}\right\} \le 1, \tag{4}$$

where $E_{DU(n_0)}$ is the expectation under the Dirac-uniform configuration of the p-values.

Remark 3. Property 1 is introduced to ensure $\hat{n}_0$ to be a relatively conservative estimate of n_0 such that approximately, $E(n_0/\hat{n}_0) \le 1$, which plays an important role in establishing the FWER control of adaptive multiple testing methods. This condition is similar to condition (16) given by Finner and Gontscharuk (2009). However, this condition (16) is given in the asymptotic settings, whereas, Property 1 is defined in the finite sample case. In the asymptotic settings, the condition (16) generally implies Property 1.

We will now give a class of estimates of n_0 containing (2) that satisfy Property 1.

Example 1. Consider the estimate

$$\hat{n}_0 = \frac{n-k+1}{1-P_{(k)}} \tag{5}$$

for any fixed $1 \le k \le n$. It satisfies Property 1, since it is clearly non-decreasing in each P_i, and, as seen in the following, it also satisfies condition (4).

Let $U_{k-n+n_0-1:n_0-1}$ be the $(k-n+n_0-1)$th ordered statistic based on a random sample of size n_0-1 from $U(0,1)$. Then, for this estimate

$$\sum_{i=1}^{n} I(H_i=0) E_{DU(n_0)}\left\{\frac{1}{\hat{n}_0(\mathbf{P}^{(-i)},0)}\right\} = \begin{cases} \dfrac{n_0}{n-k+1} & \text{for } k \le n-n_0+1 \\ \dfrac{n_0}{n-k+1} E\{1-\text{U}_{k-n+n_0-1:n_0-1}\} & \text{for } k > n-n_0+1, \end{cases} = \begin{cases} \dfrac{n_0}{n-k+1} & \text{for } k \le n-n_0+1 \\ 1 & \text{for } k > n-n_0+1, \end{cases}$$

which is less than or equal to one.

Example 2. Let $R_{n,SU}(\lambda_1,\ldots,\lambda_n)$ be the number of rejections observed while testing the n null hypotheses using a stepup method with any set of critical constants $0 \le \lambda_1 \le \cdots \le \lambda_n \le 1$. Consider the estimate

$$\hat{n}_0 = \frac{n - R_{n,SU}(\lambda_1,\ldots,\lambda_n)+1}{1-\lambda_n}. \tag{6}$$

It satisfies Property 1, since it is non-decreasing, as $R_n(\lambda_1,\ldots,\lambda_n)$ is non-increasing, in each P_i. It also satisfies condition (4), as explained in the following.

For this estimate, the following holds

$$\sum_{i=1}^{n} I(H_i=0)E_{DU(n_0)}\left\{\frac{1}{\hat{n}_0(\mathbf{P}^{(-i)},0)}\right\} = n_0 E\left\{\frac{1-\lambda_n}{n_0-\tilde{R}_{n_0-1,SU}(\lambda_{n-n_0+2},\ldots,\lambda_n)}\right\}, \tag{7}$$

where $\tilde{R}_{n_0-1,SU}(\lambda_{n-n_0+2},\ldots,\lambda_n)$ is the number of significant p-values observed when a stepup test with the critical values $\lambda_{n-n_0+2} \le \cdots \le \lambda_n$ is applied to n_0-1 p-values and the expectation is taken assuming that these p-values are iid $U(0,1)$. Let $\tilde{R}_{n_0-1}(\lambda_n)$ be the value of $\tilde{R}_{n_0-1,SU}(\lambda_{n-n_0+2},\ldots,\lambda_n)$ when $\lambda_{n-n_0+2} = \cdots = \lambda_n$. Since $\tilde{R}_{n_0-1,SU}(\lambda_{n-n_0+2},\ldots,\lambda_n) \overset{st}{\prec} \tilde{R}_{n_0-1}(\lambda_n) \sim \text{Bin}(n_0-1,\lambda_n)$, the expression in (7) is less than or equal to

$$n_0 E\left\{\frac{1-\lambda_n}{n_0-X}\middle| X \sim \text{Bin}(n_0-1,\lambda_n)\right\} = 1-\lambda_n^{n_0} \le 1,$$

as desired.

Example 3. Let $R_{n,SD}(\lambda_1,\ldots,\lambda_n)$ be the number of rejections in testing the n null hypotheses using a stepdown method with the critical values $\lambda_i = i\lambda_n/n$, $i=1,\ldots,n$, for any $0 < \lambda_n \le n/(n+2)$. With a λ_{n+1} such that $\lambda_n \le \lambda_{n+1} \le (1+\lambda_n)/2$, consider the estimate

$$\hat{n}_0 = \frac{n-R_{n,SD}(\lambda_1,\ldots,\lambda_n)+1}{1-\lambda_{R_{n,SD}(\lambda_1,\ldots,\lambda_n)+1}}. \tag{8}$$

Since $R_{n,SD}(\lambda_1,\ldots,\lambda_n)$ is non-increasing in each P_i, and as a function of $R \equiv R_{n,SD}(\lambda_1,\ldots,\lambda_n)$,

$$\hat{n}_0 = \begin{cases} \dfrac{n(n-R+1)}{n-(R+1)\lambda_n} & \text{if } R=0,1,\ldots,n-1, \\ \dfrac{1}{1-\lambda_{n+1}} & \text{if } R=n \end{cases}$$

is non-increasing under the assumed restrictions on λ_n and λ_{n+1}, the estimate in (8) is non-decreasing in each P_i. It also satisfies condition (4), as explained in the following.

As in Example 2, we see that for this estimate

$$\sum_{i=1}^{n} I(H_i=0)E_{DU(n_0)}\left\{\frac{1}{\hat{n}_0(\mathbf{P}^{(-i)},0)}\right\} = n_0 E\left\{\frac{1-\lambda_{\tilde{R}_{n_0-1,SD}(\lambda_{n-n_0+2},\ldots,\lambda_n)+1}}{n_0-\tilde{R}_{n_0-1,SD}(\lambda_{n-n_0+2},\ldots,\lambda_n)}\right\},$$

where $\tilde{R}_{n_0-1,SD}(\lambda_{n-n_0+2},\ldots,\lambda_n)$ is the number of significant p-values in a stepdown test with the critical values $\lambda_{n-n_0+2} \le \cdots \le \lambda_n$ applied to n_0-1 p-values and the expectation is taken assuming that these p-values are iid $U(0,1)$. The expectation in this last expression is seen to be less than or equal to $1/n_0$ by setting $m=n_0-1$ and $\gamma_i = \lambda_{n-n_0+1+i}, i=1,\ldots,n_0-1$, in the following lemma to be proved in Appendix. Thus, the desired condition is satisfied.

Lemma 2. *Let $R_{m,SD}(\gamma_1,\ldots,\gamma_m)$ be the number of rejections observed while testing m null hypotheses based on their p-values and using a stepdown test with critical values $0 < \gamma_1 \le \cdots \le \gamma_m < 1$. Then,*

$$E\left\{\frac{1-\lambda_{R_{m,SD}(\gamma_1,\ldots,\gamma_m)+1}}{m-R_{m,SD}(\gamma_1,\ldots,\gamma_m)+}\right\} \le \frac{1}{m+1},$$

when the underlying p-values are iid U(0,1).

Remark 4. It is easy to see that the estimate (2) is a special case of that in Example 2 or 3.

Remark 5. The estimates defined in (2), (5), (6) and (8) have been, respectively, proved to satisfy Property 1 by Benjamini et al. (2006), Blanchard and Roquain (2009), and Liu and Sarkar (2011), see Blanchard and Roquain's Corollary 13 and Liu and Sarkar's Lemma 5.2 for instance. In this paper, alternative proofs are provided for such results.

4. Adaptive Bonferroni and Šidàk methods

An adaptive Bonferroni method rejects H_i if $P_i \le \alpha/\hat{n}_0$, while an adaptive Šidàk method rejects H_i if $P_i \le 1-(1-\alpha)^{1/\hat{n}_0}$, with a suitable estimate $\hat{n}_0$ of n_0 obtained from the available p-values; see, for instance, Finner and Gontscharuk (2009).

We now propose our classes of adaptive Bonferroni and adaptive Šidàk methods based on the class of n_0 estimates introduced in the above section.

Definition 1 (*Level α adaptive Bonferroni method*).

1. Define an estimate $\hat{n}_0(\mathbf{P}^{(-i)},P_i)$ satisfying Property 1.
2. Reject H_i if $P_i \le \alpha/\hat{n}_0(\mathbf{P}^{(-i)},P_i)$.

Theorem 1. *Under Assumption* 1, *the FWER of an adaptive Bonferroni method is strongly controlled at α.*

Proof. As said in Remark 1, we will prove that $\mathrm{FWER}_{DU(n_0)} \le \alpha$. With the probabilities evaluated under the Dirac-uniform configuration, we have

$$\begin{aligned}\mathrm{FWER}_{DU(n_0)} &= \mathrm{pr}\left\{\tilde{P}_{(1)} \le \frac{\alpha}{\hat{n}_0(\mathbf{P}^{(-i)},P_i}\right\} \le \sum_{i=1}^{n} I(H_i=0)\mathrm{pr}\left\{P_i \le \frac{\alpha}{\hat{n}_0(\mathbf{P}^{(-i)},P_i)}\right\} \le \sum_{i=1}^{n} I(H_i=0)\mathrm{pr}\left\{P_i \le \frac{\alpha}{\hat{n}_0(\mathbf{P}^{(-i)},0)}\right\}\\ &= \alpha\sum_{i=1}^{n} I(H_i=0)E_{DU(n_0)}\left\{\frac{1}{\hat{n}_0(\mathbf{P}^{(-i)},0)}\right\} \le \alpha, \end{aligned} \tag{9}$$

In (9), the first inequality follows from the Bonferroni inequality, the second follows from the non-decreasing property of $\hat{n}_0$, and the third follows from condition (4) satisfied by $\hat{n}_0$. Thus, the desired result is proved. □

Remark 6. It is easy to see that there is an interesting connection between adaptive FWER control and adaptive FDR control. Specifically, considering (9) of the paper and Theorem 11 of Blanchard and Roquain (2009) (which essentially already appeared in Benjamini et al. (2006)), the FWER bound for an adaptive Bonferroni procedure is the same as the FDR bound for an adaptive BH procedure with the same estimate of n_0. Therefore, for a given estimate $\hat{n}_0$ of n_0, the condition providing that $\hat{n}_0$ can be plugged into the BH procedure is the same as the condition providing that it can be plugged into Bonferroni procedure to control the FWER. This corresponds to Property 1 in our paper.

Definition 2 (*Level α adaptive Šidàk method*).

1. Define an estimate $\hat{n}_0(\mathbf{P}^{(-i)},P_i)$ satisfying Property 1.
2. Reject H_i if $P_i \le 1-(1-\alpha)^{1/\hat{n}_0(\mathbf{P}^{(-i)},P_i)}$.

Theorem 2. *Under Assumption* 1, *the FWER of an adaptive Šidàk method is strongly controlled at α.*

The following result related to the positive association property of independent random variables (see, for instance, Esary et al., 1967) is an important tool towards proving this theorem.

Result 2. *Let $X_1,\ldots,X_m$ be independent random variables. Then, for any non-negative functions $g_j(X_1,\ldots,X_m)$, $j=1,\ldots,k$, which are either all non-decreasing or non-increasing in each X_i, we have*

$$E\prod_{j=1}^{k}\{g_j(X_1,\ldots,X_m)\} \ge \prod_{j=1}^{k} E\{g_j(X_1,\ldots,X_m)\}.$$

Proof of Theorem. As in Theorem 1, we will prove that $\mathrm{FWER}_{DU(n_0)} \le \alpha$. First, we note that

$$1-\mathrm{FWER}_{DU(n_0)} = \mathrm{pr}_{DU(n_0)}\left\{\tilde{P}_{(1)} > 1-(1-\alpha)^{1/\hat{n}_0(\mathbf{P}^{(-i)},P_i)}\right\}. \tag{10}$$

The right-hand side probability in (10) is equal to

$$\begin{aligned} &E_{DU(n_0)}\left\{\prod_{i=1}^{n_0} I(\tilde{P}_i > 1-[1-\alpha]^{I(H_i=0)/\hat{n}_0(\mathbf{P}^{(-i)},P_i)})\right\}\\ &\ge E_{DU(n_0)}\left\{\prod_{i=1}^{n_0} I(\tilde{P}_i > 1-[1-\alpha]^{I(H_i=0)/\hat{n}_0(\mathbf{P}^{(-i)},0)})\right\}\\ &\ge \prod_{i=1}^{n_0} E_{DU(n_0)}\{I(\tilde{P}_i > 1-[1-\alpha]^{I(H_i=0)/\hat{n}_0(\mathbf{P}^{(-i)},0)})\}\\ &= \prod_{i=1}^{n_0} E_{DU(n_0)}\{[1-\alpha]^{I(H_i=0)/\hat{n}_0(\mathbf{P}^{(-i)},0)}\}\\ &\ge \prod_{i=1}^{n_0}(1-\alpha)^{E_{DU(n_0)}\{I(H_i=0)/\hat{n}_0(\mathbf{P}^{(-i)},0)\}}\\ &= (1-\alpha)^{\sum_{i=1}^{n} E_{DU(n_0)}\{I(H_i=0)/\hat{n}_0(\mathbf{P}^{(-i)},0)\}} \ge 1-\alpha. \end{aligned} \tag{11}$$

In (11), the first inequality follows from the non-decreasing property of $\hat{n}_0$, the second follows from Result 2, since, for each $i=1,\ldots,n$, $I(\tilde{P}_i > 1-[1-\alpha]^{I(H_i=0)/\hat{n}_0(\mathbf{P}^{(-i)},0)})$ is a non-decreasing function of $(\mathbf{P}^{(-i)},P_i)$, the third follows from Jensen's inequality, and the final inequality follows from condition (4) satisfied by $\hat{n}_0$. Thus, the desired result is proved. □

Remark 7. The estimates in Example 1 are of the form (2) with some data-dependent choices for λ, since $\lambda = P_{(k)}$, for any $1 \le k \le n$, so we now have a proof of FWER control of adaptive Bonferroni method for such an estimate, which has not been available yet as noted in Finner and Gontscharuk (2009) in their concluding remarks. The choice of estimates in Examples 2 and 3 has been motivated by our attempt to generalize (2) from using the number of significant p-values based on a single-step test with a critical constant λ to that based on a stepup or stepdown test with a sequence of critical constants $0 \le \lambda_1 \le \cdots \le \lambda_n \le 1$. It should be noted that $\{n-R_{n,SU}(\lambda_1,\ldots,\lambda_n)+1\}/\{1-\lambda_{R_{n,SU}(\lambda_1,\ldots,\lambda_n)}\}$ and $\{n-R_{n,SD}(\lambda_1,\ldots,\lambda_n)+1\}/\{1-\lambda_{R_{n,SD}(\lambda_1,\ldots,\lambda_n)+1}\}$ are, respectively, direct stepup and stepdown generalizations of (2), but we have not been able to show that they satisfy the desired Property 1, unless we make some adjustments or choose the λ_i's appropriately, which we have done in these examples. Also, we should point out that in Example 2, choosing $\lambda_n = \lambda$ will not result in an improvement of the corresponding adaptive Bonferroni or Šidàk method over that based on (2), since the estimate (2) in this case is going to be smaller, making the corresponding adaptive Bonferroni or Šidàk method less conservative. Adaptive BH methods based on these classes of estimates of n_0 control the FDR under independence of the p-values (Sarkar, 2008b; Liu and Sarkar, 2011).

Theorems 1 and 2 extend the classes of adaptive Bonferroni methods given by Finner and Gontscharuk (2009) and Guo (2009) and adaptive Šidàk methods given by Finner and Gontscharuk (2009) with theoretically proven strong FWER control. We could consider $+\kappa$, where $\kappa \in \mathcal{R}$, as in Finner and Gontscharuk (2009), instead of $+1$ in the numerators of the estimates in Examples 1– 3. For each of these examples, such estimates are increasing in κ, and so it is easy to see that if $\kappa \ge 1$ the corresponding larger class of adaptive Bonferroni and Sidak methods would continue to strongly control the FWER.

5. Adaptive Holm and Hochberg methods

The Holm method at level α is the stepdown method with critical constants $c_i = \alpha/(n-i+1)$, $i=1,\ldots,n$; whereas, the Hochberg method at the same level is the stepup method with these same critical values. If n_0 were known, an ideal, more powerful version of the Holm method would be the stepdown method with the critical values $c_i = \alpha/\min\{n_0, n-i+1\}$, $i=1,\ldots,n$, as can be seen from Lemma 1. So, with unknown n_0, an appropriate adaptive Holm method is the stepdown method with the (random) critical values $\hat{c}_i = \alpha/\min\{\hat{n}_0, n-i+1\}$, $i=1,\ldots,n$, for some suitable estimate $\hat{n}_0$ of n_0 obtained from the available data. Finner and Gontscharuk (2009) and Guo (2009) have both considered such an adaptive Holm method using the estimate (2). We propose a different version of adaptive Holm method in this section using the estimate in Example 1. The stepup analog of this adaptive Holm method is our adaptive Hochberg method. As in Finner and Gontscharuk (2009) and Guo (2009), we establish the strong FWER control of these adaptive methods asymptotically as $n_0 \to \infty$, using different arguments.

Definition 3 (*Level α adaptive Holm method*).

1. Define
$$\hat{n}_0(k) = \frac{n-k+1}{1-P_{(k)}} \quad \text{for any fixed } 1 \le k \le n.$$
2. Reject H_i if $P_i \le P_{(\hat{r})}$, where
$$\hat{r} = \max\left\{ i : P_{(j)} \le \frac{\alpha}{\min\{\hat{n}_0(k), n-j+1\}} \quad \text{for all } j \le i \right\}.$$

Definition 4 (*Level α adaptive Hochberg method*).

1. Define $\hat{n}_0(k)$ as in Definition 3.
2. Reject H_i if $P_i \le P_{(\hat{s})}$, where
$$\hat{s} = \max\left\{ i : P_{(i)} \le \frac{\alpha}{\min\{\hat{n}_0(k), n-i+1\}} \right\}.$$

The following result holds for these adaptive Holm and Hochberg methods.

Theorem 3. *Let $n_0/n \to \pi_0$ and $k/n \to \gamma$, as $n \to \infty$, for some fixed $\pi_0, \gamma \in (0,1)$. Then, under Assumption 1, the FWERs of these adaptive Holm and Hochberg methods are strongly controlled at α asymptotically as $n \to \infty$.*

Proof. We prove this result only for the adaptive Hochberg method, which will imply the same result for the adaptive Holm method, since between stepdown and stepup methods with the same set of critical values, the FWER of the stepdown method is less than or equal to that of the stepup method.

As said in Remark 1, we will prove that $\limsup_{n_0\to\infty}\text{FWER}_{DU(n_0)} \le \alpha$. Under the Dirac-uniform configuration, $\hat{n}_0(k)$ reduces to $\tilde{n}_0(k)$, where $\tilde{n}_0(k) = n-k+1$ for $k \le n-n_0$, and $\tilde{n}_0(k) = (n-k+1)/[1-\tilde{P}_{(k-n+n_0)}]$ for $k > n-n_0$. Thus, we have,

$$\text{FWER}_{DU(n_0)} = \text{pr}\left(\bigcup_{i=1}^{n_0}\left\{\tilde{P}_{(i)} \le \frac{\alpha}{\min\{\tilde{n}_0(k), n_0-i+1\}}\right\}\right),$$

where $\tilde{P}_i$, $i=1,\ldots,n_0$, are iid U(0,1).

Consider the two different cases: (i) $k \le n-n_0$ and (ii) $k > n-n_0$. In case (i),

$$\text{FWER}_{DU(n_0)} = \text{pr}\left(\bigcup_{i=1}^{n_0}\left\{\tilde{P}_{(i)} \le \frac{\alpha}{n_0-i+1}\right\}\right) \le \text{pr}\left(\bigcup_{i=1}^{n_0}\left\{\tilde{P}_{(i)} \le \frac{i\alpha}{n_0}\right\}\right) = \alpha,$$

because of the Simes test (Simes, 1986; Sarkar, 1998, 2008a; Sarkar and Chang, 1997). In case (ii), since $i(n_0-i+1) \ge n_0$ for $1 \le i \le n_0$, we have

$$\text{FWER}_{DU(n_0)} \le \text{pr}\left(\bigcup_{i=1}^{n_0}\left\{\tilde{P}_{(i)} \le \max\left[\frac{\alpha(1-\tilde{P}_{(k-n+n_0)})}{n-k+1}, \frac{i\alpha}{n_0}\right]\right\}\right) = \text{pr}\left(\bigcup_{i=1}^{n_0}\left\{\tilde{P}_{(i)} \le \max\left[\frac{\alpha(1-\tilde{P}_{(j)})}{n_0-j+1}, \frac{i\alpha}{n_0}\right]\right\}\right), \tag{12}$$

with $1 \le j = k-n+n_0 \le n_0$. Note that, as $n\to\infty$, $n_0\to\infty$ and $j/n_0 \to \eta = (\pi_0+\gamma-1)/\pi_0 \in (0,1)$. So, the FWER in (12) is less than or equal to α asymptotically as $n\to\infty$, which follows from the following lemma. This lemma will be proved in Appendix. □

Lemma 3. *Consider testing the intersection of m null hypotheses H_i, $i=1,\ldots,m$, based on their respective p-values P_i, $i=1,\ldots,m$. Define $m^*(j) = (m-j+1)/(1-P_{(j)})$, for some fixed $1 \le j \le m$. Use the modified version of Simes' test rejecting the intersection null hypothesis if $P_{(i)} \le \hat{\alpha}_i$ for at least one $i=1,\ldots,m$, where $\hat{\alpha}_i = \max\{\alpha/m^*(j), i\alpha/m\}$. Let $j/m\to\eta$, for some fixed $\eta \in (0,1)$, as $m\to\infty$. The type I error rate is controlled at α asymptotically as $m\to\infty$ when these p-values are iid* U(0,1).

Remark 8. Through (B.2) in the proof of Lemma 3, it is easy to see that the upper bounds of the FWER for the above adaptive Holm and Hochberg procedures are only slightly larger than the pre-specified level α as the number of true nulls n_0 is moderately large. Therefore, these two adaptive procedures are slightly liberal at the most for finite samples.

Remark 9. In Theorem 3, π_0 is assumed to be positive. In fact, this theorem still holds when $\pi_0 = 0$. If $\pi_0 = 0$, then $k \le n-n_0$ always holds for large n. Thus, we only need to consider case (i) in the proof of Theorem 3. In case (i),

$$\text{FWER}_{DU(n_0)} = \text{pr}\left(\bigcup_{i=1}^{n_0}\left\{\tilde{P}_{(i)} \le \frac{i\alpha}{n_0}\right\}\right) = \alpha$$

holds because of the Simes test, no matter what value the π_0 is equal to. Therefore, by Proportion 1, the FWER control of the adaptive Holm and Hochberg procedures is still maintained as $\pi_0 = 0$.

Remark 10. We need to indicate that among all existing and newly suggested adaptive procedures controlling the FWER, the adaptive single-step procedures including adaptive Bonferroni and Šidàk methods always work for fixed n_0, whereas, for the adaptive stepdown and stepup procedures including adaptive Holm and Hochberg procedures, they only work as $n_0\to\infty$.

6. Numerical findings

We performed simulation studies to investigate the following questions:

Q1. How do the newly suggested adaptive Bonferroni and Šidàk methods perform in terms of the FWER control and power for dependent p-values compared to the adaptive Bonferroni method in Finner and Gontscharuk (2009) and Guo (2009) and the adaptive Šidàk method in Finner and Gontscharuk (2009)?

Q2. How well do the newly suggested adaptive Holm and Hochberg methods control the FWER compared to the original adaptive Holm method in Finner and Gontscharuk (2009) and Guo (2009) for dependent p-values?

In the study related to Q1, two different settings for dependent p-values were simulated using multivariate normal test statistics, one with a compound symmetric covariance matrix and the other with a block dependence covariance matrix. The dependence setting in the study related to Q2 was based on multivariate normal test statistics with a compound symmetric covariance matrix. The statistics have a common non-negative correlation ρ in case of compound symmetry and are broken up into g independent groups with a common non-negative correlation ρ within each group in case of block dependence.

The simulated FWER and (average) power, the expected proportion of false nulls that are rejected, were obtained for each adaptive method by (i) generating 100 $(=n)$ dependent normal random variables $N(\mu_i,1), i=1,\ldots,n$, with 50 $=(n_0)$ of the 100 μ_i's being equal to 0 and the rest $d=\sqrt{10}$, (ii) applying the method to the generated data to test $H_i:\mu_i=0$ against $K_i:\mu_i\neq 0$ simultaneously for $i=1,\ldots,100$ at level $\alpha=0.05$, and (iii) by repeating steps (i) and (ii) 2000 times.

We considered four different versions of estimate (2); they correspond to $\lambda = 0.2, 0.4, 0.6$ and 0.8. For each of these, we chose the estimate (5) with $k = [n\lambda]$, the largest integer contained in $n\lambda$, because it is compatible with the estimate (2).

The λ_i's in the estimates (6) and (8), considered while answering Q1, were chosen as follows: $\lambda_i = i\beta/n$, $i = 1, \ldots, n$, in both (6) and (8), and $\lambda_{n+1} = \lambda_n$ in (8). The β was fixed at $\alpha/(1+\alpha)$, the same significance level Benjamini et al. (2006) successfully used in estimating n_0 using a BH type method while constructing an adaptive version of the BH method controlling the FDR at α; see also Romano et al. (2008) and Sarkar and Heller (2008). The value of g was chosen to be 20.

The simulated FWERs and powers for adaptive Šidàk methods corresponding to all four estimates were noted to be slightly larger than the corresponding values for adaptive Bonferroni methods. So, we decided not to ultimately include the adaptive Bonferroni methods in our comparisons while answering Q1, keeping in mind that any conclusion we will reach for an adaptive Šidàk method can also be made for the corresponding adaptive Bonferroni method.

Fig. 1 provides answer to Q1 for adaptive Šidàk methods. The four adaptive Šidàk methods based on (2), (5), (6) and (8) are labeled adSid1–adSid4, respectively. Fig. 2 answers Q2, with the adaptive Holm method in Finner and Gontscharuk (2009) and Guo (2009), and the two new adaptive Holm and Hochberg methods based on (5) being labeled adHol1, adHol2 and adHoch, respectively.

Fig. 1 reveals an interesting feature about the performance of the adaptive Šidàk method in Finner and Gontscharuk (2009) with respect to the choice of λ when the p-values are computed from equicorrelated normal test statistics. If λ is chosen to be small, this method does not seem to lose its control over the FWER, contrary to what Finner and Gontscharuk (2009) observed when $\lambda = 0.5$ and what we also see in the first two rows of panels in Fig. 1. In fact, for small values of λ,

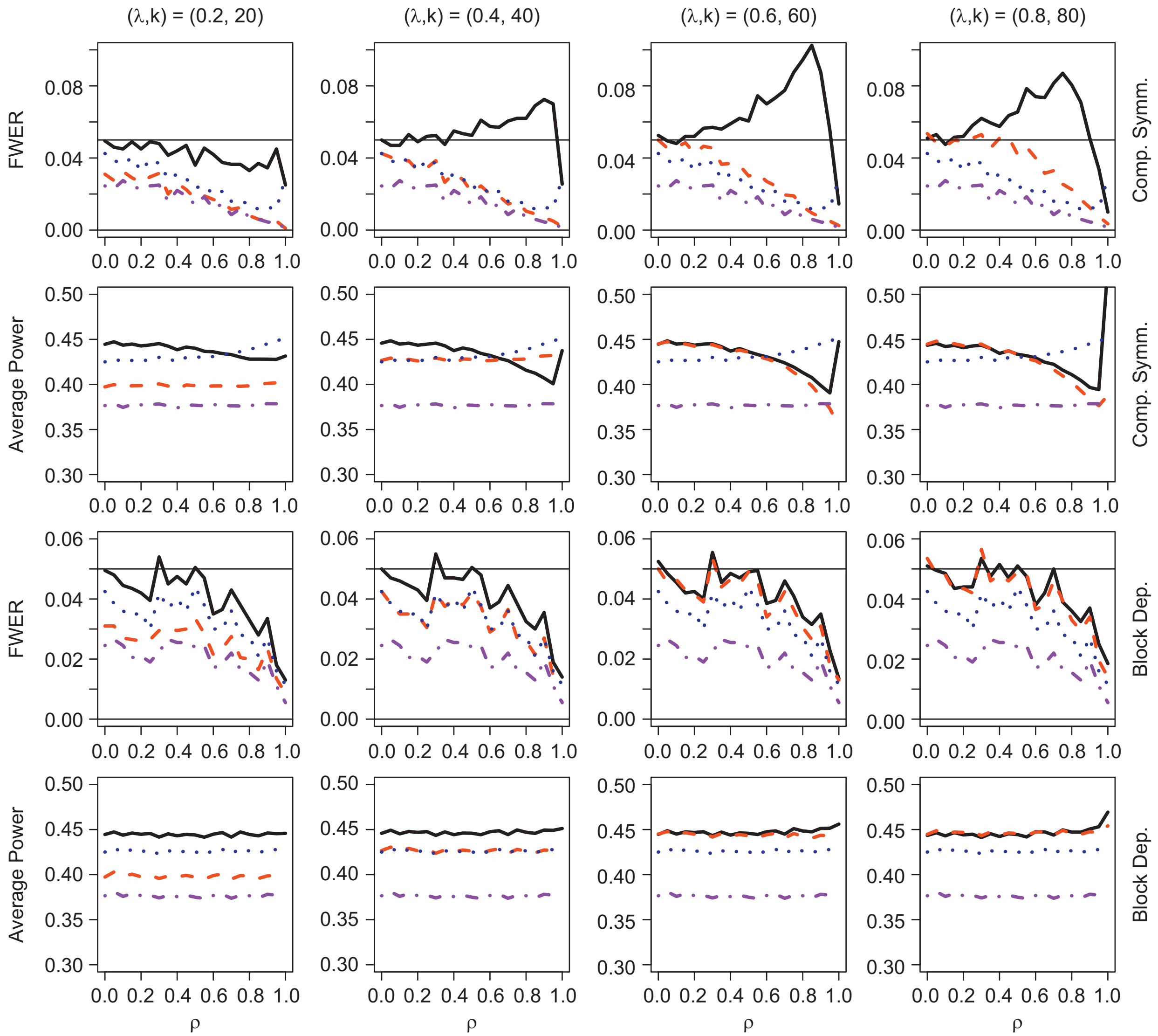

Fig. 1. Simulated FWERs and average powers of four adaptive Šidàk methods (adSid1, solid; adSid2, dashes; adSid3, dots; adSid4, dot dashes) with 100 dependent p-values generated from normal test statistics with compound symmetry or block dependence structure.

this version of adaptive Šidàk method appears to be the least conservative and more powerful, except when ρ is very close to one. If λ is not small, each of the three newly suggested adaptive Šidàk methods has better performance in terms of controlling the FWER. The simulated FWER for each of these new methods decreases and remains controlled at α with increasing ρ, whereas, for the adaptive method in Finner and Gontscharuk (2009), it increases with ρ, except for very large values of ρ, exceeding α for most of the ρ values, which was also observed by Finner and Gontscharuk (2009). Considering also the powers of these new adaptive methods it seems that among these methods, the one based on (6) works the best for large ρ or small λ, while the one based on (5) performs the best for small ρ and large λ.

Fig. 1 also suggests that, when the p-values are locally dependent, the two adaptive Šidàk methods based on (6) and (8) seem to control the FWER fairly well, whereas, other two adaptive Sĭdăk methods based on (2) and (5) seem to slightly lose the FWER control for all values of λ and some larger values of k. For larger values of (λ,k), the method of Finner and Gontscharuk (2009) and the new method based on (5) are very close to each other and more powerful than the other two; whereas, for smaller values of (λ,k), the new method based on (6) is slightly less powerful than the one in Finner and Gontscharuk (2009) but more powerful than the other two.

From Fig. 2, we see that the adaptive Holm method in Finner and Gontscharuk (2009) and Guo (2009) seems to be the least conservative if λ is chosen to be small under a normal distributional setting with equal positive correlation for the test statistics, like the adaptive Šidàk method. For larger values of λ, however, the simulated FWER of this method increases with ρ, except for very large values of ρ, often exceeding α by a significant margin; whereas, for each of the two new methods proposed here, the FWER decreases and remains controlled at α with increasing ρ.

Finally, following one referee's suggestions, we add some simulations to study the FWER control of the four adaptive Šidàk methods under some other settings. First, we do some simulations for the worst case of $n=n_0$ with 100 dependent p-values generated from standard normal test statistics with compound symmetry structure. Second, we do some simulations for $n=n_0$ with 10 dependent p-values generated from standard normal test statistics with block dependence structure. The 10 dependent p-values are grouped as five blocks and within each block, the p-values are negatively dependent with common correlation coefficient $\rho<0$. The simulation studies are used to evaluate the effect of negative

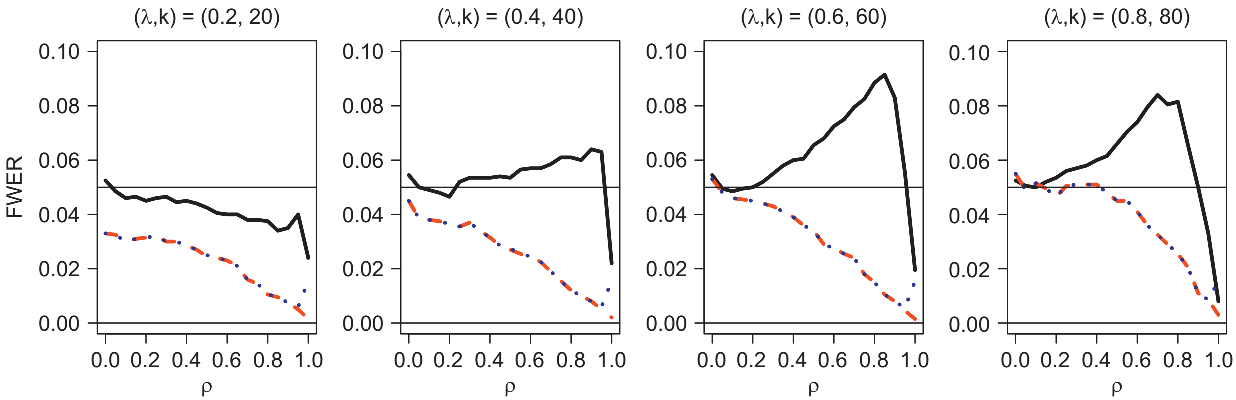

Fig. 2. Simulated FWERs of adaptive Holm and Hochberg methods (adHol1, solid; adHol2, dashes; adHoch, dots) with 100 dependent p-values generated from normal test statistics with compound symmetry structure.

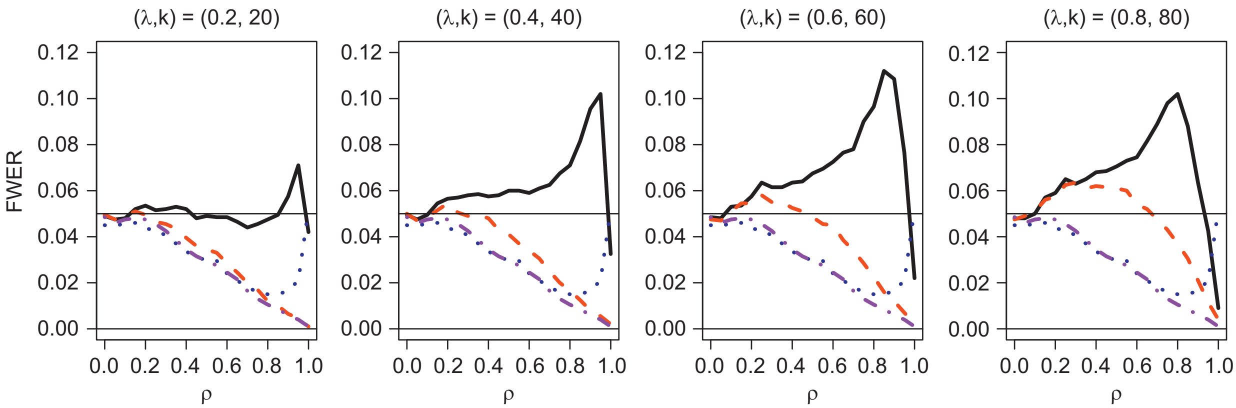

Fig. 3. Simulated FWERs of four adaptive Šidàk methods (adSid1, solid; adSid2, dashes; adSid3, dots; adSid4, dot-dashes) with 100 dependent p-values generated from standard normal test statistics with compound symmetry structure.

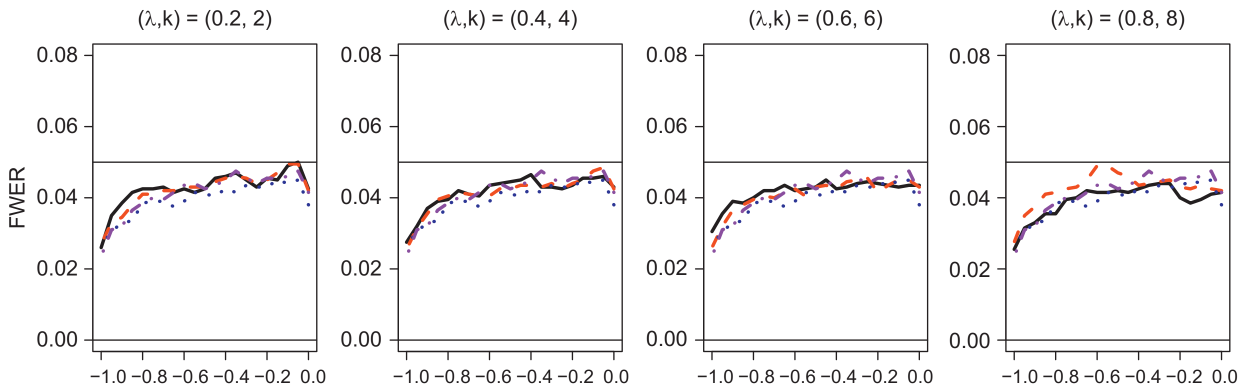

Fig. 4. Simulated FWERs of four adaptive Šidàk methods (adSid1, solid; adSid2, dashes; adSid3, dots; adSid4, dot-dashes) with 10 dependent p-values generated from standard normal test statistics with negative block dependence structure with block size = 2.

dependence on the FWER control of these adaptive methods. In these simulation scenarios, the simulated FWERs are graphically displayed in Figs. 3 and 4, respectively.

From Fig. 3, we see that when the value of k is chosen to be relatively large, the adaptive Šidàk method based on (5) loses the FWER control for slightly or moderately correlated p-values in the case of $n=n_0$, whereas, the FWER control of this adaptive method is still maintained in the case of $n_0=n/2$, as seen from the first row of Fig. 1. From Fig. 4, we see that under negative block dependence, the four adaptive Šidàk methods control the FWER fairly well, whereas, under positive block dependence, the adaptive Šidàk methods based on (2) and (5) slightly lose the FWER control for some values of λ and k, as seen from the third row of Fig. 1.

7. Concluding remarks

The primary focus of this paper has been to advance the theory of adaptive FWER controlling methods from what is recently known in the literature. We have given newer adaptive methods with proven FWER control under slightly weaker distributional assumptions, with numerical evidence that they often have better performance under certain type of dependent p-values. With regard to adaptive Bonferroni and Šidàk methods, we have given a unified theory of constructing them. Our proofs are different and simpler, not requiring explicit formulas for FWER under the Dirac-uniform configuration as in Finner and Gontscharuk (2009).

There is, however, a scope of doing further investigations, at least numerically, to see if our classes could be further extended. For instance, with $+\kappa$ instead of $+1$ in the numerators of the estimates in each of the classes of estimates in Examples 1–3, we could consider solving the equation:

$$\max_{1\le n_0\le n} \text{FWER}_{DU(n_0)} = \alpha$$

for κ, like what Finner and Gontscharuk (2009) did for BPI, having derived $\text{FWER}_{DU(n_0)}$ explicitly using the distribution of the order statistics of n_0 iid U(0,1) (Finner and Roters, 2002). Obviously, such a κ would be optimal within the corresponding class of estimates in the sense of providing adaptive Bonferroni or Šidàk method with the least conservative strong control of the FWER. Nevertheless, finding this κ needs to be carried out numerically as in Finner and Gontscharuk (2009), since doing so theoretically would be extremely difficult. For BPI, Finner and Gontscharuk (2009) noted that the optimum κ is slightly less than 1, but it does not offer an appreciable improvement of the corresponding BPI over the one with $\kappa=1$. It would be interesting to see if that happens with our estimates too, though we are going to do that in a different communication.

It is important to note that the expectation of

$$\sum_{i=1}^{n} I(H_i=0)\left\{\frac{1}{\hat{n}_0(\mathbf{P}^{(-i)},0)}\right\}$$

under the Dirac-uniform configuration provides information on how conservative an adaptive method based on the estimate $\hat{n}_0$ is, at least in the present independence setup. The smaller (or larger) this expectation is compared to 1, the more conservative (or liberal) the corresponding adaptive method is. For the estimate in (2), this expectation is $1-\lambda^{n_0}$; whereas, for the estimate in (5), it is $\min\{1,n_0/(n-k+1)\}$ (as seen in Section 3.1). So, in the sparse case, for instance, where a high proportion of null hypotheses are believed to be true, the estimate (5) with a moderately large k is going to provide a less conservative adaptive method than the one based on any estimate in (2). However, care should be taken in choosing k when there is dependence among the p-values. We simulated FWER's of adaptive Bonferroni methods based on (5) with different values of k in the least favorable setting of equicorrelated standardized normal test statistics. The graphs over

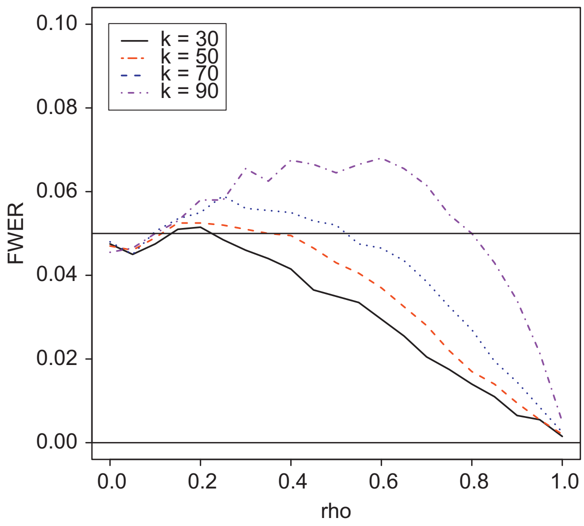

Fig. 5. Simulated FWERs of adaptive Bonferroni method based on estimate (5) with dependent p-values generated from 100 standardized normal test statistics with compound symmetry structure.

different values of correlation are displayed in Fig. 5. As seen from this figure, a relatively high value of k might cause the corresponding adaptive Bonferroni method to lose its control over the FWER when the dependence is not too high or too low.

Acknowledgements

The research of the first author is supported by the NSF Grants DMS-0603868 and DMS-1006344 and the research of the second author is supported by the NSF Grant DMS-1006021. We sincerely thank two referees for giving helpful and insightful comments which have led to an improved presentation of the paper.

Appendix A. Proof of Lemma 2

Consider testing $m+1$ null hypotheses using their p-values P_i, $i=1,\dots,m+1$, based on a stepdown method with the critical values $\gamma_1 \le \cdots \le \gamma_m \le \gamma_{m+1}$, where γ_{m+1} is chosen satisfying $\gamma_m \le \gamma_{m+1} < 1$. Let $R_{m+1,SD}(\gamma_1,\dots,\gamma_{m+1})$ be the number of significant p-values. Note that, we have for $r=0,1,\dots,m+1$,

$$\begin{aligned}
&[m+1-R_{m+1,SD}(\gamma_1,\dots,\gamma_{m+1})]I(R_{m+1,SD}(\gamma_1,\dots,\gamma_{m+1})=r)\\
&\quad=\sum_{i=1}^{m+1} I(P_i>\gamma_{r+1},R_{m+1,SD}(\gamma_1,\dots,\gamma_{m+1})=r)\\
&\quad=\sum_{i=1}^{m+1}\{P_i>\gamma_{r+1},P_{(1)}\le\gamma_1,\dots,P_{(r)}\le\gamma_r,P_{(r+1)}>\gamma_{r+1}\}\\
&\quad=\sum_{i=1}^{m+1}\{P_i>\gamma_{r+1},P_{(1)}^{(-i)}\le\gamma_1,\dots,P_{(r)}^{(-i)}\le\gamma_r,P_{(r+1)}^{(-i)}>\gamma_{r+1}\}\\
&\quad=\sum_{i=1}^{m+1}\{P_i>\gamma_{r+1},R_{m,SD}^{(-i)}(\gamma_1,\dots,\gamma_m)=r\},
\end{aligned}$$

where $\gamma_{m+2}=1$, $P_{(1)}^{(-i)}\le\cdots\le P_{(m)}^{(-i)}$ are the ordered versions of the m p-values except the P_i, and $R_{m,SD}^{(-i)}(\gamma_1,\dots,\gamma_m)$ is the number of significant p-values in the stepdown test with the critical values $\gamma_1\le\cdots\le\gamma_m$ based on these m p-values. So,

$$\begin{aligned}
&I(R_{m+1,SD}(\gamma_1,\dots,\gamma_{m+1})<m+1)=\frac{m+1-R_{m+1,SD}(\gamma_1,\dots,\gamma_{m+1})}{m+1-R_{m+1,SD}(\gamma_1,\dots,\gamma_{m+1})}I(R_{m+1,SD}(\gamma_1,\dots,\gamma_{m+1})<m+1)\\
&\quad=\sum_{r=0}^{m}\sum_{i=1}^{m+1}\frac{I(P_i>\gamma_{r+1},R_{m+1,SD}(\gamma_1,\dots,\gamma_{m+1})=r)}{m+1-r}=\sum_{r=0}^{m}\sum_{i=1}^{m+1}\frac{I(P_i>\gamma_{r+1},R_{m,SD}^{(-i)}(\gamma_1,\dots,\gamma_m)=r)}{m+1-r}.
\end{aligned}$$

Now, if the p-values are iid U(0,1), taking expectations on both sides in the above equalities, we see that

$$\sum_{i=1}^{m+1} E\left\{\frac{1-\gamma_{R_{m,SD}^{(-i)}(\gamma_1,\ldots,\gamma_m)+1}}{m+1-R_{m,SD}^{(-i)}(\gamma_1,\ldots,\gamma_m)}\right\} \le \text{pr}(R_{m+1,SD}(\gamma_1,\ldots,\gamma_{m+1})<m+1)\le 1.$$

Thus, the lemma is proved.

Appendix B. Proof of Lemma 3

This result can be proved borrowing ideas from Sarkar (2008b) and using the following additional lemma:

Lemma 4 (*Blanchard and Roquain, 2009*). *Given a random variable U satisfying* $\text{pr}(U\le u)\le u$, $u\in(0,1)$, *any positive valued non-increasing function* $g(\cdot)$, *and a fixed constant* c, *we have*

$$E\left\{\frac{I(U\le cg(U))}{g(U)}\right\}\le c.$$

Let $R_{m-1}^{(-i)}$ be the number of rejections in the stepup test based on the $m-1$ p-values $\{P_1,\ldots,P_m\}\setminus\{P_i\}$ and the critical values $\hat{\alpha}_i=\max\{\alpha/m^*(j), i\alpha/m\}$, $i=2,\ldots,m$. Then, from Sarkar (2008b), the type I error rate is equal to

$$\sum_{i=1}^{m} E\left\{\frac{I(P_i\le\hat{\alpha}_{R_{m-1}^{(-i)}+1})}{R_{m-1}^{(-i)}+1}\right\}, \tag{B.1}$$

where

$$\hat{\alpha}_{R_{m-1}^{(-i)}+1}=\max\left\{\frac{\alpha}{m^*(j)},\frac{[R_{m-1}^{(-i)}+1]\alpha}{m}\right\}.$$

Since $m^*(j)\ge(m-j+1)/[1-P_{(j-1)}^{(-i)}]$, we have

$$\hat{\alpha}_{R_{m-1}^{(-i)}+1}\le\max\left\{\frac{[1-P_{(j-1)}^{(-i)}]\alpha}{m-j+1},\frac{[R_{m-1}^{(-i)}+1]\alpha}{m}\right\}\le[R_{m-1}^{(-i)}+1]\frac{\alpha}{m}\max\left\{\frac{m[1-P_{(j-1)}^{(-i)}]}{m-j+1},1\right\},$$

and so the expectation inside the summation in (B.1) is less than or equal to

$$E\left\{\frac{I\left(P_i\le[R_{m-1}^{(-i)}+1]\frac{\alpha}{m}\max\left\{\frac{m[1-P_{(j-1)}^{(-i)}]}{m-j+1},1\right\}\right)}{R_{m-1}^{(-i)}+1}\right\}.$$

Notice that, given $\{P_1,\ldots,P_m\}\setminus\{P_i\}$, $(\alpha/m)\max\{m[1-P_{(j-1)}^{(-i)}]/(m-j+1),1\}$ is constant; whereas, $R_{m-1}^{(-i)}$ is a non-increasing function of P_i, since the critical values $\hat{\alpha}_i$, $i=2,\ldots,m$, are so. Thus, applying Lemma 4 to this expectation conditional on $\{P_1,\ldots,P_m\}\setminus\{P_i\}$, we see that the expression in (B.1) is less than or equal to

$$\alpha E\left(\max\left\{\frac{m(1-P_{(j-1)}^{(-i)})}{m-j+1},1\right\}\right)=\alpha E\left(\max\left\{\frac{m(1-U_{j-1:m-1})}{m-j+1},1\right\}\right)=\alpha\left[\text{pr}\left(U_{j-1:m}\le\frac{j-1}{m}\right)+\text{pr}\left(U_{j-1:m-1}\ge\frac{j-1}{m}\right)\right]\to\alpha. \tag{B.2}$$

In (B.2), the convergence to α follows from the fact that as $j/m\to\eta$, for some fixed $\eta\in(0,1)$, and $m\to\infty$, we have

$$\text{pr}\left(U_{j:m}\le\frac{j}{m}\right)=\text{pr}\left\{\text{Bin}\left(m,\frac{j}{m}\right)\ge j\right\}\approx\text{pr}\left\{\frac{1}{m}\text{Bin}(m,\eta)\ge\eta\right\}\to 1/2,$$

from the central limit theorem applied to the binomial distribution.

References

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57, 289–300.

Benjamini, Y., Krieger, K., Yekutieli, D., 2006. Adaptive linear step-up procedures that control the false discovery rate. Biometrika 93, 491–507.

Blanchard, G., Roquain, E., 2009. Adaptive FDR control under independence and dependence. J. Mach. Learn. 10, 2837–2871.

Dmitrienko, A., Tamhance, A.C., Bretz, F., 2009. In: Multiple Testing Problems in Pharmaceutical Statistics. Chapman & Hall/CRC.

Esary, J.D., Proscahn, F., Walkup, D.W., 1967. Association of random variables with applications. Ann. Math. Statist. 38, 1466–1474.

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.

Finner, H., Roters, M., 2002. Multiple hypotheses testing and expected number of type I errors. Ann. Statist. 30, 220–238.

Gavrilov, Y., Benjamini, Y., Sarkar, S.K., 2009. An adaptive step-down procedure with proven FDR control. Ann. Statist. 37, 619–629.

Guo, W., 2009. A note on adaptive Bonferroni and Holm procedures under dependence. Biometrika 96, 1012–1018.

Hochberg, Y., 1988. A sharper Bonferroni procedure for multiple tests of significance. Biometrika 75, 800–802.

Hochberg, Y., Benjamini, Y., 1990. More powerful procedures for multiple significance testing. Statist. Med. 9, 811–818.

Hochberg, Y., Tamhane, A.C., 1987. In: Multiple Comparison Procedures. Wiley.
Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scand. J. Statist. 6, 65–70.
Hsu, J., 1996. In: Multiple Comparisons: Theory and Methods. Chapman & Hall, New York.
Liu, F., Sarkar, S.K., 2011. A new adaptive method to control the false discovery rate. In: Bhattacharjee Dhar, Subramanian (Eds.), Recent Advances in Biostatistics: False Discovery, Survival Analysis and Related Topics, Series in Biostatistics, vol. 4. World Scientific, pp. 3–26.
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.
Sarkar, S.K., 1998. Some probability inequalities for ordered MTP_2 random variables: a proof of the Simes conjecture. Ann. Statist. 26, 494–504.
Sarkar, S.K., 2006. False discovery and false nondiscovery rates in single-step multiple testing procedures. Ann. Statist. 34, 394–415.
Sarkar, S.K., 2008a. On the Simes inequality and its generalization. In: Balakrishnan, Peña, Silvapulle (Eds.), Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen. IMS Collections, pp. 231–242.
Sarkar, S.K., 2008b. On methods controlling the false discovery rate (with discussions). Sankhya 70, 135–168.
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.
Sarkar, S.K., Heller, R., 2008. Comments on: control of the false discovery rate under dependence using the bootstrap and subsampling. TEST 17, 450–455.
Schweder, T., Spjøtvoll, E., 1982. Plots of p-values to evaluate many tests simulataneously. Biometrika 69, 493–502.
Simes, R.J., 1986. An improved Bonferroni procedure for multiple tests of significance. Biometrika 73, 751–754.
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.
Tong, Y.L., 1980. Probability Inequalities in Multivariate Distributions. Academic Press.