Statistical Applications in Genetics and Molecular Biology

Volume 6, *Issue* 1 2007 *Article* 3

A Generalized Sidak-Holm Procedure and Control of Generalized Error Rates under Independence

Wenge Guo* Joseph Romano†

*University of Cincinnati, guowg@email.uc.edu
†Stanford University, romano@stanford.edu

A Generalized Sidak-Holm Procedure and Control of Generalized Error Rates under Independence

Wenge Guo and Joseph Romano

Abstract

For testing multiple null hypotheses, the classical approach to dealing with the multiplicity problem is to restrict attention to procedures that control the familywise error rate (FWER), the probability of even one false rejection. In many applications, one might be willing to tolerate more than one false rejection provided the number of such cases is controlled, thereby increasing the ability of the procedure to detect false null hypotheses. This suggests replacing control of the FWER by controlling the probability of k or more false rejections, which is called the k-FWER. In Hommel and Hoffmann (1987) and Lehmann and Romano (2005a), single step and stepdown procedures are derived that control the k-FWER, without making any assumptions concerning the dependence structure of the p-values of the individual tests. However, if the p-values are mutually independent, one can improve the procedures. In fact, Sarkar (2005) provided such an improvement. However, we show other improvements are possible which appear to be generally much better, and are sometimes unimprovable. When k=1, the procedure reduces to the classical method of Sidak, and the stepdown procedure is unimprovable and strictly dominates that of Sarkar. Under a monotonicity condition, an unimprovable procedure is obtained. In the case k=2, the monotonicity condition is satisfied, and the condition can be checked numerically in general. We then develop a stepdown method that controls the false discovery proportion. Except for the case of k-FWER control with k=1, the gains are surprisingly dramatic, and theoretical and numerical evidence is given.

KEYWORDS: false discovery proportion, generalized familywise error rate, multiple testing, p-value, Sidak procedure, stepdown procedure

1 Introduction

Consider the general problem of simultaneously testing a finite number of null hypotheses H_i $(i = 1, \ldots, s)$. We shall assume that tests for the individual hypotheses are available and the problem is how to combine them into a simultaneous test procedure. More formally, suppose data X is available from some model $P \in \Omega$. A general hypothesis H can be viewed as a subset ω of Ω. For testing $H_i : P \in \omega_i$, $i = 1, \ldots, s$, let $I(P)$ denote the set of true null hypotheses when P is the true probability distribution; that is, $i \in I(P)$ if and only if $P \in \omega_i$.

The usual approach to dealing with this problem is to restrict attention to procedures that control the probability of one or more false rejections. This probability is called the familywise error rate (FWER). Here the term "family" refers to the collection of hypotheses $H_1, \ldots, H_s$ that is being considered for joint testing. Which tests are to be treated jointly as a family depends on the situation.

Once the family has been defined, control of the FWER (at joint level α) requires that

$$\text{FWER} \le \alpha \ , \tag{1}$$

for all possible distributions P of the data. A quite broad treatment of methods that control the FWER is presented in Hochberg and Tamhane (1987).

Safeguards against false rejections are of course not the only concern of multiple testing procedures. Corresponding to the power of a single test one must also consider the ability of a procedure to detect departures from the hypotheses when they do occur. When the number of tests is in the tens or hundreds of thousands, control of the FWER at conventional levels becomes so stringent that individual departures from the hypotheses have little chance of being detected. For this reason, we shall consider an alternative to the FWER that controls false rejections less severely and consequently provides better power.

Specifically, we shall consider the k-FWER, the probability of rejecting at least k true null hypotheses. Such an error rate with $k > 1$ is appropriate when one is willing to tolerate one or more false rejections, provided the number of false rejections is controlled.

More formally, the k-FWER, which depends on P is defined to be

$$k\!-\!\text{FWER} = P\{\text{reject at least } k \text{ hypotheses } H_i \text{ with } i \in I(P)\} \ . \tag{2}$$

Control of the k-FWER requires that k-FWER $\le \alpha$ for all P; that is,

$$P\{\text{reject at least } k \text{ hypotheses } H_i \text{ with } i \in I(P)\} \le \alpha \quad \text{for all } P \ . \tag{3}$$

Evidently, the case $k = 1$ reduces to control of the usual FWER.

The methods in this paper will assume that tests of individual hypotheses are based on p-values. So, before describing methods that provide control of the k-FWER, we first recall the notion of a p-value, since multiple testing methods are often described by the p-values of the individual tests. Consider a single null hypothesis $H : P \in \omega$. Assume a family of tests of H, indexed by α, with level α rejection regions S_α satisfying

$$P\{X \in S_\alpha\} \leq \alpha \quad \text{for all } 0 < \alpha < 1,\ P \in \omega \tag{4}$$

and

$$S_\alpha \subset S_{\alpha'} \quad \text{whenever } \alpha < \alpha' \ . \tag{5}$$

Then, the p-value is defined by

$$\hat{p} = \hat{p}(X) = \inf\{\alpha :\ X \in S_\alpha\} \ . \tag{6}$$

The important property of a p-value that will be used later is the following.

Lemma 1.1 *Assume $\hat{p}$ is defined as above.*
(i) If $P \in \omega$, then

$$P\{\hat{p} \leq u\} \leq u \ . \tag{7}$$

(ii) Furthermore,

$$P\{\hat{p} \leq u\} \geq P\{X \in S_u\} \ . \tag{8}$$

Therefore, if the S_α are such that equality holds in (4), then $\hat{p}$ is uniformly distributed on $(0,1)$ when $P \in \omega$.

A proof is given in Lehmann and Romano (2005a).

Two classic procedures that control the FWER are the Bonferroni procedure and the Holm procedure. The Bonferroni procedure rejects H_i if its corresponding p-value satisfies $\hat{p}_i \leq \alpha/s$. Assuming $\hat{p}_i$ satisfies

$$P\{\hat{p}_i \leq u\} \leq u \quad \text{for any } u \in (0,1) \quad \text{and any } P \in \omega_i \ , \tag{9}$$

the Bonferroni procedure provides strong control of the FWER. Unfortunately, the ability of the Bonferroni procedure to detect cases in which H_i is false, will typically be very low since H_i is tested at level α/s which - particularly if s is large - is orders smaller than the conventional α levels.

For this reason procedures are prized for which the levels of the individual tests are increased over α/s without an increase in the FWER. It turns out that such a procedure due to Holm (1979) is available under the present minimal assumptions.

The Holm procedure can conveniently be stated in terms of the p-values $\hat{p}_1, \ldots, \hat{p}_s$ of the s individual tests. Let the ordered p-values be denoted by

$$\hat{p}_{(1)} \leq \ldots \leq \hat{p}_{(s)} \ ,$$

and the associated hypotheses by $H_{(1)}, \ldots, H_{(s)}$. Then, the Holm procedure is as follows. Accept all hypotheses if $\hat{p}_{(1)} > \alpha/s$. Otherwise, reject $H_{(1)}, \ldots, H_{(j)}$ if $\hat{p}_{(i)} \leq \alpha/(s-i+1)$ for $i = 1, \ldots, j$ (and choose the largest such j).

The Bonferroni method is an example of a *single step* procedure, meaning any null hypothesis is rejected if its corresponding p-value is less than or equal to a common cutoff value (which in the Bonferroni case is α/s). The Holm procedure is a special case of a class of *stepdown* procedures, which we now briefly describe. Let

$$\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_s \tag{10}$$

be constants. If $\hat{p}_{(1)} > \alpha_1$, reject no null hypotheses. Otherwise, if

$$\hat{p}_{(1)} \leq \alpha_1, \ldots, \hat{p}_{(r)} \leq \alpha_r \ , \tag{11}$$

reject hypotheses $H_{(1)}, \ldots, H_{(r)}$ where the largest r satisfying (11) is used. That is, a stepdown procedure starts with the most significant p-value and continues rejecting hypotheses as long as their corresponding p-values are small. The Holm procedure uses $\alpha_i = \alpha/(s-i+1)$.

Recently, several new methods have been proposed which utilize error rates that are less stringent than the FWER. Control of the k-FWER as well as methods based on the FDP was first suggested in Victor (1982). An important initial paper is the now well-known method of Benjamini and Hochberg (1995), which controls the so-called false discovery rate (FDR). The false discovery rate is defined as the expected value of the false discovery proportion (FDP), where FDP is the ratio of the number of rejections of true null hypotheses to the total number of rejections (defined as 0 if there are no rejections). The original method of Benjamini and Hochberg (1995) was derived under independence of the individual p-values, but a modification was proved in Benjamini and Yekutieli (2001). Genovese and Wasserman (2004) study asymptotic procedures that control the FDP (and the FDR) in the framework of a random effects mixture model. These ideas are extended in Perone Pacifico, Genovese, Verdinelli and Wasserman (2004), where in the context of random fields, the number of null hypotheses is uncountable. In Hommel and Hoffmann (1987) and Lehmann and Romano (2005a), single step and stepdown methods for control of the k-FWER are derived under no dependence assumptions on the p-values. Stepdown and stepup improvements are provided in Romano and Shaikh (2006a, 2006b). Under independence assumptions on the p-values, further improvements are derived in Sarkar (2005). Korn, Troendle, McShane and Simon (2004) provide methods that control both the k-FWER and FDP; they provide some justification for their methods, but they are limited to a multivariate permutation model. Alternative methods of control of the k-FWER and FDP are given in van der Laan, Dudoit and Pollard (2004); they include both finite sample and asymptotic results. Further methods are discussed in Dudoit, van der Laan and Pollard (2004), van der Laan and Birnker (2004) and van der Laan, Birkner and Hubbard (2005). Like these papers, Romano and Wolf (2005) also provide methods for control of generalized error rates which account for the dependence structure of the individual test statistics or p-values. In this paper, the goal is modest. We make the strong assumption of independence, but it allows us to provide greatly improved methods. In fact, it was initially surprising that such dramatic improvements could be obtained (except for k-FWER control when $k = 1$). However, theoretical and numerical results support this claim of dramatic improvement. Even if the assumption of independence is unwarranted, these results point to the need for reliable methods that account for dependence.

The outline of the paper is as follows. In Section 2, we first consider single step control of the k-FWER and provide the optimal such procedure under independence. A stepdown improvement of this procedure is provided in Section 3 and can be viewed as a generalization of the classical Sidák procedure. In the case $k = 1$, this procedure is unimprovable among such stepdown procedures. In Section 4, we consider further improvements and, under a monotonicity condition, the optimal procedure is obtained. In the case $k = 2$, the monotonicity condition is satisfied. Control of the false discovery proportion, the ratio of false rejections to total number of rejections is considered in Section 5, and a new procedure is obtained. Throughout, improvements are provided under the assumption of independence. Such an assumption is obviously quite restrictive, but the results aid to deepen our understanding of the theory of multiple testing procedures.

2 Single Step Control of the k-FWER

The usual Bonferroni procedure compares each p-value $\hat{p}_i$ with α/s. Control of the k-FWER allows one to increase α/s to $k\alpha/s$, and thereby greatly increase the ability to detect false hypotheses. Part (i) of the following result is given in Hommel and Hoffmann (1987) and also in Lehmann and Romano (2005a), who derive (ii).

Theorem 2.1 *[Generalized Bonferroni] For testing $H_i : P \in \omega_i$, $i = 1, \ldots, s$, suppose $\hat{p}_i$ satisfies (9). Consider the procedure that rejects any H_i for which $\hat{p}_i \leq k\alpha/s$.*
(i) This procedure controls the k-FWER, so that (3) holds. Equivalently, if each of the hypotheses is tested at level $k\alpha/s$, then the k-FWER is controlled.
(ii) For this procedure, the inequality (3) is sharp in the sense that there exists a joint distribution for $(\hat{p}_1, \ldots, \hat{p}_s)$ for which equality is attained in (3).

Under independence, one can improve the constant $k\alpha/s$, but first some notation is useful. Suppose $\{y_i : i \in K\}$ is a collection of real numbers indexed by a finite set K having $|K|$ elements. Then, for $k \leq |K|$, the k-min$(y_i : i \in K)$ is used to denote the kth smallest value of the y_i with $i \in K$. So, if the elements y_i, $i \in K$, are ordered as $y_{(1)} \leq \cdots \leq y_{(|K|)}$, then

$$k\text{-min}(y_i : i \in K) = y_{(k)} \ .$$

Let $U_1, \ldots, U_s$ be i.i.d. uniform on (0,1). For $k \leq s$, let

$$H_{k,s}(u) = P\{k\text{-min}(U_1, \ldots, U_s) \leq u\}$$

be the distribution function of the kth order statistic. As is standard, we have

$$H_{k,s}(u) = \sum_{j=k}^{s} \binom{s}{j} u^j (1-u)^{s-j} \ . \tag{12}$$

Consider a single step procedure that rejects any H_i whose corresponding p-value $\hat{p}_i$ is $\leq C$. Then, in order to control the k-FWER at level α, the choice of $C = C_{k,s}(\alpha)$ satisfying

$$H_{k,s}(C_{k,s}(\alpha)) = \alpha \tag{13}$$

would work, because the chance that there are k or more false rejections is the chance that k-min$(\hat{p}_i : i \in I(P)) \leq C$. So, assuming all hypotheses are true and p-values are i.i.d. $U(0,1)$, we should take

$$C = H_{k,s}^{-1}(\alpha) \ .$$

Furthermore, if $D > C$, then using the critical value D instead of C would not control the k-FWER, because assuming all hypotheses are true, the k-FWER would be $H_{k,s}(D) > H_{k,s}(C) = \alpha$. The following establishes strong control (not just weak control when all hypotheses are true) of the k-FWER.

Theorem 2.2 *[Generalized Sidák] For testing $H_i : P \in \omega_i$, $i = 1, \ldots, s$, suppose $\hat{p}_i$ satisfies (9). Further assume the p-values are mutually independent. Consider the procedure that rejects any H_i for which $\hat{p}_i \leq C_{k,s}$, where $C_{k,s} = C_{k,s}(\alpha)$ satsifies $H_{k,s}(C_{k,s}(\alpha)) = \alpha$.*
(i) This procedure controls the k-FWER, so that (3) holds. Equivalently, if each of the hypotheses is tested at level $C_{k,s}(\alpha)$, then the k-FWER is controlled.
(ii) For this procedure, the inequality (3) is sharp in the sense that there exists a joint distribution of independent p-values for $(\hat{p}_1, \ldots, \hat{p}_s)$ for which equality is attained in (3).

Proof. (i) Fix any P and suppose H_i with $i \in I = I(P)$ are true and the remainder false, with $|I|$ denoting the cardinality of I. Order the $|I|$ p-values corresponding to true null hypotheses as

$$\hat{q}_{(1)} \leq \cdots \leq \hat{q}_{(|I|)} \ .$$

Assume $|I| \geq k$ or there is nothing to prove. Let N be the number of false rejections. Then,

$$P\{N \geq k\} = P\{\hat{q}_{(k)} \leq C_{k,s}(\alpha)\} \ .$$

But, using (7),

$$P\{\hat{q}_{(k)} \le C_{k,s}(\alpha)\} \le H_{k,|I|}(C_{k,s}(\alpha)) \ ,$$

with equality if and only if the p-values are all $U(0,1)$. Now, for any u,

$$H_{k,|I|}(u) \le H_{k,s}(u) \ ,$$

because the chance that, in a sample of i.i.d. $U(0,1)$ variables, k or more p-values are $\le u$ increases with the number of trials. Hence,

$$P\{N \ge k\} \le H_{k,s}(C_{k,s}(\alpha))$$

and the right hand side is α by definition (13).

To prove (ii), consider the following construction. Let all hypotheses be true, so that $|I| = s$ and assume the $\hat{p}_i$ are i.i.d. $U(0,1)$ variables. Then, the inequalities in the proof of (i) are all equalities. ■

In words, the single step procedure that rejects any hypothesis whose corresponding p-value is $\le C_{k,s}(\alpha)$ controls the k-FWER. Furthermore, the constant $C_{k,s}(\alpha)$ is tight in the sense that for any value $D > C_{k,s}(\alpha)$, the single step procedure using the cutoff value D could violate control of the k-FWER.

Having derived the best possible improvement under independence, a natural question is whether or not such an improvement is appreciable. In fact, for control of the k-FWER when $k = 1$, it is well-known that no big improvement over Bonferroni is possible. Indeed, Miller (1981) remarks that the Bonferroni method "is not as crude as one might think". For $k = 1$, the ratio of critical values satisfies:

$$\lim_{s\to\infty} \frac{C_{1,s}(\alpha)}{\alpha/s} \to \frac{-\log(1-\alpha)}{\alpha} \ ,$$

which equals 1.026 when $\alpha = 0.05$; see Lehmann and Romano (2005b), Problem 9.2. Thus, the gain is clearly negligible for large s. However, we now will argue this only holds when $k = 1$ and, for $k > 1$, the gain can be dramatic.

First, we perform some numerical comparisons of the critical constants of the new single step generalized Sidák method using the critical constant $C_{k,s}(\alpha)$ with Lehmann and Romano (2005a)'s generalized Bonferroni critical value $k\alpha/s$, which does not make any assumption concerning the dependence structure of the p-values. Some results are obtained in Table 1. We see that, for $s = 100$ and $\alpha = 0.05$, even for $k = 2$, the ratio of critical constants is substantial and is equal to 3.53. The effect increases with k and the ratio is over 10 when $k = 10$.

We now give a theoretical explanation of the above surprising results. For testing $H_i, i = 1, \ldots, s$, suppose H_i with $i \in I$ are true and the remainder false, with $|I|$ denoting the cardinality of I. Let $\hat{q}_i, i = 1, \ldots, |I|$ denote the $|I|$ p-values corresponding to true null hypotheses, and assume the $\hat{q}_i$ are mutually independent and uniformly distributed. If we use the generalized Bonferroni procedure with

Table 1: Single step constants for k-FWER control with $s = 100$ and $\alpha = 0.05$

k	$A = k\alpha/s$	$B = C_{k,s}(\alpha)$	B/A
1	0.0005	0.00051	1.026
2	0.0010	0.00353	3.530
3	0.0015	0.00806	5.376
4	0.0020	0.01337	6.686
5	0.0025	0.01913	7.653
6	0.0030	0.02518	8.392
7	0.0035	0.03140	8.972
8	0.0040	0.03774	9.436
9	0.0045	0.04416	9.814
10	0.0050	0.05062	10.124

critical constant $k\alpha/s$ to control k-FWER, then each true null hypothesis H_i with $i \in I$ will be rejected with probability $k\alpha/s$. Define $X_i = I\{\hat{q}_i \leq k\alpha/s\}$, where $I(\cdot)$ is indicator function. The X_i are i.i.d. Bernoulli random variables with success probability $k\alpha/s$. Let N be the number of false rejections, then $N = \sum_{i=1}^{|I|} X_i$, and so N is a Binomial random variable with parameters $|I|$ and $k\alpha/s$, i.e., $N \sim \text{Binomial}(|I|, k\alpha/s)$. Note that, the success probability $k\alpha/s$ is very small. So, when $|I|$ is large, N is approximately Poisson distributed with mean $\lambda = |I|k\alpha/s$, as is standard for fixed k with s and $|I|$ tending to ∞. For a Poisson distributed random variable with mean λ, we prove the following result.

Lemma 2.1 *If N is a Poisson distributed random variable with mean λ, then $P\{N \geq k\} \leq \lambda^k$.*

PROOF.

$$P\{N \geq k\} = \exp(-\lambda) \cdot \sum_{j=k}^{\infty} \frac{\lambda^j}{j!} = \exp(-\lambda)\lambda^k \sum_{i=0}^{\infty} \frac{\lambda^i}{(i+k)!}$$

$$\leq \exp(-\lambda)\lambda^k \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = \exp(-\lambda)\lambda^k \exp(\lambda) = \lambda^k \ . \blacksquare$$

Based on Lemma 2.1 and (2), we find that the k-FWER is bounded above (asymptotically for large s) by $\lambda^k = (|I|k\alpha/s)^k \leq (k\alpha)^k$; that is,

$$k\text{−FWER} = P\{N \geq k\} = O(\alpha^k)$$

as $\alpha \to 0$. Note that for the usual small nominal values of α, the k-FWER is approximately α in the case $k = 1$. But when $k > 1$, the k-FWER decreases dramatically below α. In other words, if all null hypotheses are true and p-values are mutually independent, using the critical value $C_{k,s}(\alpha)$ results in FWER equal to α, but using the critical value $k\alpha/s$ results in the FWER equal to $O(\alpha^k)$, which can be very much smaller than α, and hence is way too conservative. So, in this case, we can greatly increase the critical constant of the generalized Bonferroni procedure and still control the k-FWER at level α, even when $k = 2$. This is the reason why there is a big difference between the critical constants of the generalized Bonferroni and Sidák procedures.

3 Stepdown control of the k-FWER

As is the case for the Bonferroni method, the previous single step procedure can be strengthened by a Holm type of improvement. Consider the stepdown procedure described in (11), where now we specifically consider

$$\alpha_i = \begin{cases} \frac{k\alpha}{s} & i \le k \\ \frac{k\alpha}{s+k-i} & i > k \end{cases} \tag{14}$$

Of course, the α_i depend on s and k, but we suppress this dependence in the notation. The following result is stated in Hommel and Hoffmann (1987) and proved in Lehmann and Romano (2005a); it can be viewed as a generalization of Holm (1979).

Theorem 3.1 *[Generalized Bonferroni-Holm] For testing $H_i : P \in \omega_i$, $i = 1, \ldots, s$, suppose $\hat{p}_i$ satisfies (9).*
(i) The stepdown procedure described in (11) with α_i given by (14) controls the k-FWER; that is, (3) holds.
(ii) One cannot increase even one of the constants α_i for $i \ge k$ without violating control of the k-FWER.

Remark 3.1 Evidently, one can always reject the hypotheses corresponding to the smallest $k-1$ p-values without violating control of the k-FWER. That is, we could always apply a stepdown procedure with $\alpha_i = 1$ for $i < k$. However, it seems counterintuitive to consider a stepdown procedure whose corresponding α_i are not monotone nondecreasing. In addition, automatic rejection of $k-1$ hypotheses, regardless of the data, appears at the very least a little too optimistic. To ensure monotonicity, Lehmann and Romano (2005a) suggest using $\alpha_i = k\alpha/s$ for $i < k$.

Sarkar (2005) showed that, under the assumption of mutual independence of the p-values, the above procedure can be improved by using

$$\alpha_i' = \left(\alpha \prod_{j=1}^{k} \frac{j}{s-i+j} \right)^{1/k}, \quad i = k, \ldots, s \, . \tag{15}$$

For $i < k$, one can set $\alpha'_i = 1$ like in the above remark, but a more sensible choice is α'_k. It is easily checked that $\alpha'_i \geq \alpha_i$, at least if $1 \leq k \leq 1/\alpha$. However, under the independence assumption, an alternative procedure also provides an improvement, which we will see can be much greater.

Theorem 3.2 *[Generalized Sidák-Holm] Under the assumptions of Theorem 3.1, further assume the p-values are mutually independent. Consider the stepdown procedure using critical constants*

$$\alpha_1 \leq \cdots \leq \alpha_s \ ,$$

where

$$\alpha_1 = \cdots = \alpha_k = H^{-1}_{k,s}(\alpha) = C_{k,s}(\alpha)$$

and, for $j > 0$,

$$\alpha_{k+j} = H^{-1}_{k,s-j}(\alpha) = C_{k,s-j}(\alpha) \ . \tag{16}$$

(i) This procedure controls the k-FWER (as does the stepdown procedure replacing α_i with 1 if $i < k$).
(ii) For $k = 1$, it is not possible to increase even one of the α_i with $i \geq k$ without violating control of the k-FWER.

PROOF. Fix any P and let $I(P)$ be the indices of the true null hypotheses. Assume $|I(P)| \geq k$ or there is nothing to prove. Order the p-values corresponding to the $|I(P)|$ true null hypotheses; call them

$$\hat{q}_{(1)} \leq \cdots \leq \hat{q}_{(|I(P)|)} \ .$$

Let j be the smallest (random) index satisfying $\hat{p}_{(j)} = \hat{q}_{(k)}$, so

$$k \leq j \leq s - |I(P)| + k \tag{17}$$

because the largest possible index j occurs when all the smallest p-values correspond to the $s - |I(P)|$ false null hypotheses and the the next $|I(P)|$ p-values correspond to the true null hypotheses. Then, the stepdown procedure commits at least k false rejections if and only if

$$\hat{p}_{(1)} \leq \alpha_1, \hat{p}_{(2)} \leq \alpha_2, \ldots, \hat{p}_{(j)} \leq \alpha_j \ ,$$

which certainly implies that

$$\hat{q}_{(k)} = \hat{p}_{(j)} \leq \alpha_j = H^{-1}_{k,s-j+k}(\alpha) \ .$$

But by (17), since $j \geq k$ and the α_i are monotone,

$$\alpha_j \leq \alpha_{s-|I(P)|+k} = H^{-1}_{k,|I(P)|}(\alpha) \ .$$

So, the probability of at least k false rejections is bounded above by

$$P\left\{\hat{q}_{(k)} \le H^{-1}_{k,|I(P)|}(\alpha)\right\} \ .$$

By Theorem 2.2(i), the chance that the kth smallest among $I(P)$ p-values is $\le H^{-1}_{k,|I(P)|}(\alpha)$ is $\le \alpha$.

We now prove in the case $k = 1$ that one cannot improve even one of the critical values without violating error control. To see this with constant α_1, just take all hypotheses to be true, and it reduces to the single step result. For $i > 1$, let there be $s - i + 1$ true null hypotheses and the rest false. Furthermore, let the false ones have p-values which are identically zero and let the true ones be i.i.d. $U(0, 1)$ variables. Then, following the argument in (i) for this scenario, the false null hypotheses will be immediately rejected by the stepdown procedure and the event that the FWER is violated is identical to the event that $\hat{q}_{(1)} \le \alpha_i$. But,

$$\alpha_i = H^{-1}_{1,s-i+1}(\alpha) = H^{-1}_{1,|I(P)|}(\alpha) \ ,$$

and so the FWER becomes

$$P\left\{\hat{q}_{(1)} \le H^{-1}_{1,|I(P)|}(\alpha)\right\} = \alpha \ . \ \blacksquare$$

Remark 3.2 In the case $k = 1$, $H_{1,s}(u) = 1 - (1 - u)^s$ and

$$H^{-1}_{1,s}(\alpha) = 1 - (1 - \alpha)^{1/s} \ .$$

So, the critical values reduce to Sidák's method, where the α_i are given by

$$\alpha_i = H^{-1}_{1,s-i+1}(\alpha) = 1 - (1 - \alpha)^{1/(s-i+1)} \ .$$

Remark 3.3 The proofs of Theorem 2.2 and 3.2 show that the results remain true under the weaker condition that the p-values corresponding to true null hypotheses are mutually independent. The p-values corresponding to false null hypotheses can have an arbitrary dependence structure and can even be dependent on the p-values corresponding to true null hypotheses.

Remark 3.4 In the case $k = 1$, it is known that Sidák's method also controls the 1-FWER under certain types of positive dependence. For example, this holds if the p-values $\hat{p}_1, \ldots, \hat{p}_s$ are positively orthant dependent in the sense that

$$P\{\hat{p}_1 \le x_1, \ldots, \hat{p}_s \le x_s\} \ge \Pi_{i=1}^{s} P\{\hat{p}_i \le x_i\} \ ; \tag{18}$$

see Holland and Copenhaver (1987). Unfortunately, this result does not carry over in general when $k > 1$. The following construction was based on a personal communication with S. Sarkar. Assume $s = k$ and that all null hypotheses are true.

Then, $\alpha_s = \alpha_k = \alpha^{1/k}$. The p-values of the true null hypotheses are $\hat{q}_1, \ldots, \hat{q}_k$, assumed to be equal to a common variable U which is uniformly distributed on $(0, 1)$. Then, the k-FWER is given by

$$P\{k\text{-min}(\hat{q}_1, \ldots, \hat{q}_k) \le \alpha_k\} = \alpha^{1/k} > \alpha \ .$$

In general, if the $\hat{q}_i$s are positively dependent in the sense of (18), then

$$P\{k\text{-min}(\hat{q}_1, \ldots, \hat{q}_k) \le \alpha_s\} \ge \Pi_{i=1}^k P\{\hat{q}_i \le \alpha_s\}$$
$$= \alpha_s^k = [H_{k,k}^{-1}(\alpha)]^k = [\alpha^{1/k}]^k = \alpha \ .$$

Remark 3.5 The results in this paper can be viewed as special cases of a generalized closure method given in Romano and Wolf (2005). A variant of this principle is given in Guo and Rao (2006). In Romano and Wolf (2005), it is further shown how to construct tests that control generalized error rates under dependence, including some examples with exact finite sample control (such as parametric models and models where randomization or permutation tests apply), as well as general approximate methods based on bootstrap resampling or subsampling.

We now perform some numerical comparisons of the critical constants of the new generalized Sidák-Holm stepdown procedure in Theorem 3.2 and Lehmann and Romano (2005a)'s generalized Bonferroni-Holm stepdown procedure, which does not make any assumption concerning the dependence structure of the p-values. We plot in Figures 1 and 2 the two sequences of constants described in (14) and (16) for the cases in which $k = 10, s = 200, \alpha = 0.05$ and $k = 2, s = 50, \alpha = 0.05$, respectively. Panel (a) displays the critical constants based on (14), where panel (b) displays the critical constants based on (16). Panel (c) displays the ratio of the constants in panel (b) with the constants in panel (a). The dashed horizontal lines in panel (b) are of height 0.05. It is clear from panel (c) in Figures 1 and 2 that the constants in panel (b) are much larger than the constants of panel (a). Thus, if the assumption of independence of p-values is satisfied, the stepdown procedure based on the constants in panel (b) is clearly preferable to the one based on the constants in panel (a).

An interesting result we observe from Figures 1 and 2 is that there is a big difference if $k > 1$ between the critical constants of generalized Holm procedure and the new stepdown procedure presented here. It means that the assumption of independence gives us a large gain in power. For $k = 1$, the stepdown improvement over Holm is not dramatic, but it immediately becomes so when $k = 2$.

4 Further stepdown improvements

Consider the critical constants α_i defined in (16). Recall the argument in attempting to show optimality of the constant α_i. For $i > k$, let there be $s - i + k$ true

Figure 1: The critical constants of two k-FWER stepdown procedures based on (14) and (16) for $k = 10, s = 200,$ and $\alpha = 0.05$.

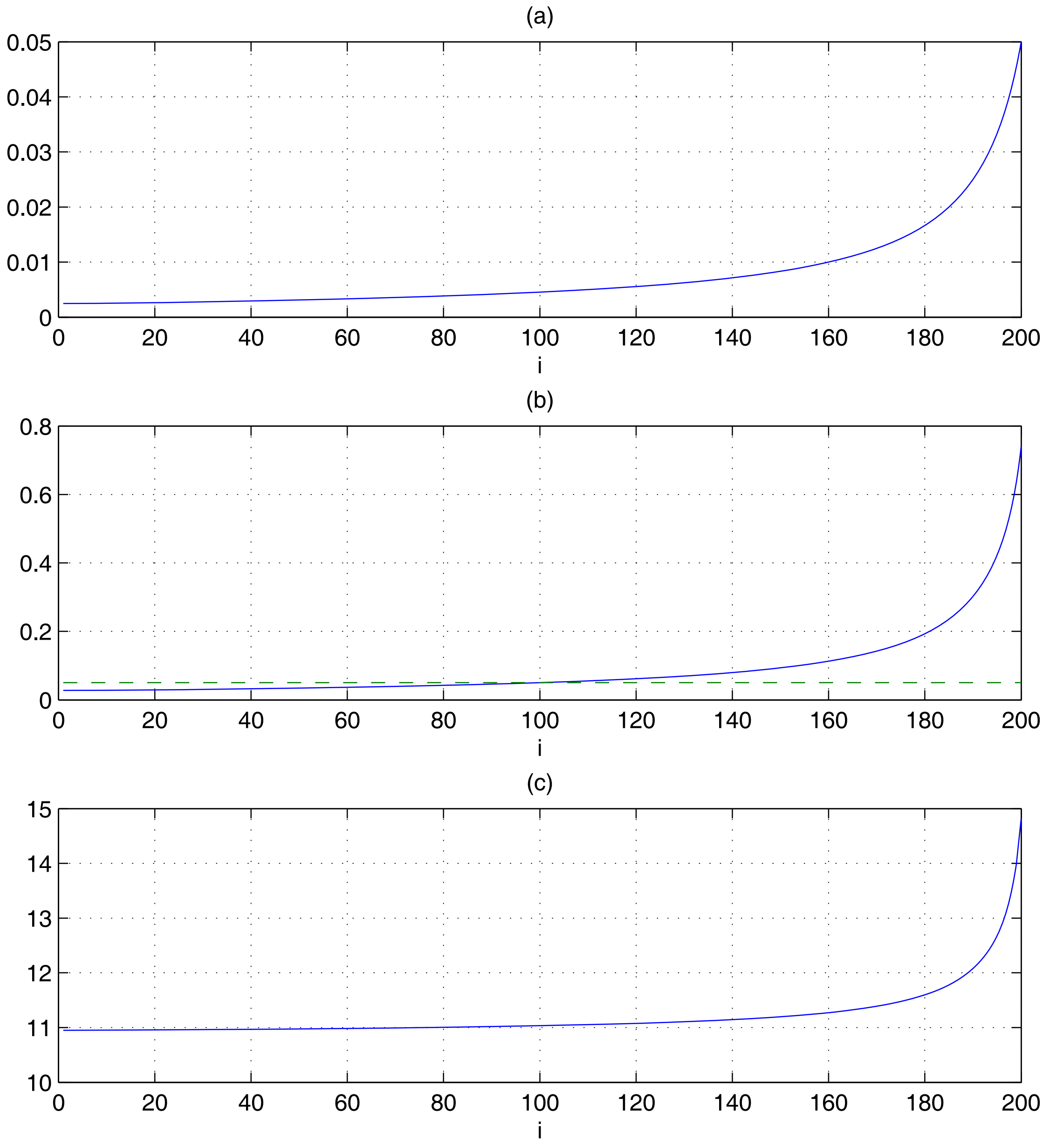
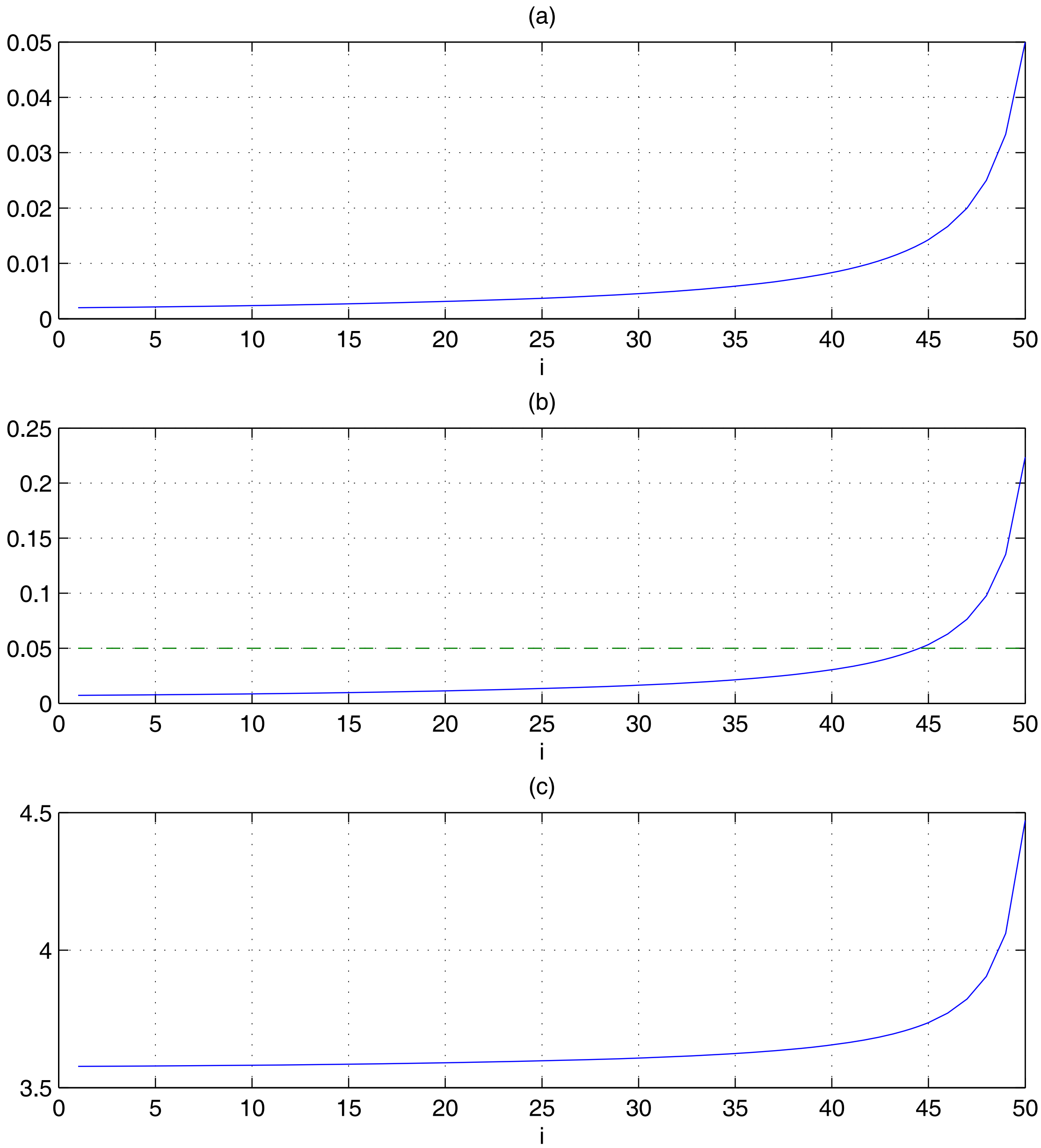


Figure 2: The critical constants of two k-FWER stepdown procedures based on (14) and (16) for $k = 2, s = 50,$ and $\alpha = 0.05$.



null hypotheses and the rest false. Furthermore, let the false ones have p-values which are identically zero. Then, the event that k-FWER is violated is identical to the event

$$\{\hat{q}_{1:s-i+k} \leq \alpha_{i-k+1}, \ldots, \hat{q}_{k:s-i+k} \leq \alpha_i\} \ ,$$

(which is not equivalent to the event $\{\hat{q}_{k:s-i+k} \leq \alpha_i\}$ unless $k = 1$), where

$$\hat{q}_{1:s-i+k} \leq \ldots \leq \hat{q}_{k:s-i+k}$$

are the first k ordered p-values corresponding to the $s-i+k$ true null hypotheses. So, it is possible to improve the constant α_i if $k > 1$.

A new sequence of critical constants $\{\alpha_i, i = 1, \ldots, s\}$, is defined as follows:

$$\alpha_i = \alpha_k \ , \quad \text{for } i < k \tag{19}$$

and

$$P\{U_{1:s-i+k} \leq \alpha_{i-k+1}, \ldots, U_{k:s-i+k} \leq \alpha_i\} = \alpha \ , \quad \text{for } i \geq k \ , \tag{20}$$

where $U_{1:v} \leq \ldots \leq U_{j:v}$ are the first j order statistics of v i.i.d. $U(0,1)$ variables. Note that, if $i = k$, then

$$P\{U_{1:s} \leq \alpha_1, \ldots, U_{k:s} \leq \alpha_k\} = P\{U_{k:s} \leq \alpha_k\} \ .$$

So, $\alpha_k = H^{-1}_{k,s}(\alpha)$.

We now prove that there exists a unique sequence of critical constants satisfying (19) and (20).

Lemma 4.1 *There exists a unique sequence of critical constants $\{\alpha_i, i = 1, \ldots, s\}$ satisfying (19) and (20).*

Proof. Let $\alpha_k = H^{-1}_{k,s}(\alpha)$ and $\alpha_i = \alpha_k$, for $i < k$. For $i > k$, suppose $\alpha_1, \ldots, \alpha_{i-1} \geq \alpha_k$ have been determined and satisfy (19) and (20). A function $G(u)$ is defined by

$$G(u) = P\{U_{1:s-i+k} \leq \alpha_{i-k+1}, \ldots, U_{k-1:s-i+k} \leq \alpha_{i-1}, U_{k:s-i+k} \leq u\} \tag{21}$$

where, $0 \leq u \leq 1$. Note that the event

$$\{U_{1:s-i+k+1} \leq \alpha_{i-k}, U_{2:s-i+k+1} \leq \alpha_{i-k+1}, \ldots, U_{k:s-i+k+1} \leq \alpha_{i-1}\}$$

implies the event

$$\{U_{1:s-i+k} \leq \alpha_{i-k+1}, U_{2:s-i+k} \leq \alpha_{i-k+2}, \ldots, U_{k-1:s-i+k} \leq \alpha_{i-1}\} \ .$$

So,

$$\begin{aligned}G(1) &= P\{U_{1:s-i+k} \le \alpha_{i-k+1}, U_{2:s-i+k} \le \alpha_{i-k+2}, \ldots, U_{k-1:s-i+k} \le \alpha_{i-1}\} \\ &\ge P\{U_{1:s-i+k+1} \le \alpha_{i-k}, U_{2:s-i+k+1} \le \alpha_{i-k+1}, \ldots, U_{k:s-i+k+1} \le \alpha_{i-1}\} \\ &= \alpha\end{aligned}$$

In addition, the event $\{U_{k:s-i+k} \le \alpha_k\}$ implies the event $\{U_{k:s} \le \alpha_k\}$. So,

$$\begin{aligned}G(\alpha_k) &= P\{U_{1:s-i+k} \le \alpha_{i-k+1}, \ldots, U_{k-1:s-i+k} \le \alpha_{i-1}, U_{k:s-i+k} \le \alpha_k\} \\ &\le P\{U_{k:s-i+k} \le \alpha_k\} \le P\{U_{k:s} \le \alpha_k\} \\ &= \alpha\end{aligned}$$

From the continuity of $G(u)$, there exists $u_0 \in [\alpha_k, 1]$ satisfying $G(u_0) = \alpha$; that is, a solution satisfying (20) exists for α_i.

Next, we show that the solution α_i is unique. Suppose there exist two solutions u_1 and u_2 satisfying (19), (20) and $u_1 \le u_2$; that is,

$$G(u_1) = G(u_2) = \alpha \tag{22}$$

Then,

$$G(u_2)-G(u_1) = P\{U_{1:s-i+k} \le \alpha_{i-k+1}, \ldots, \hat{q}_{k-1:s-i+k} \le \alpha_{i-1}, u_1 < U_{k:s-i+k} \le u_2\} = 0\,,$$

and note that $\alpha_j \ge \alpha_k > 0$, for $j = i-k+1, \ldots, i-1$. So, $u_1 = u_2$.

Hence, the result is proved by induction. ■

Later, through some numerical computations, we will find that, in some cases, the sequence $\{\alpha_i, i = 1, 2, \ldots, s\}$ satisfying (19) and (20) is not monotone increasing (see Figure 6). However, in the case $k = 2$, monotonicity of the sequence $\{\alpha_i, i = 1, 2, \ldots, s\}$ is proved to hold.

Lemma 4.2 *Suppose $\{\alpha_i, i = 1, \ldots, s\}$ be the sequence of critical constants satisfying (19) and (20). For $k = 2$, the sequence $\{\alpha_i, i = 1, \ldots, s\}$ is monotone increasing.*

For the proof, see Appendix A.

To construct a stepdown procedure using the sequence $\{\alpha_i, i = 1, 2, \ldots, s\}$ satisfying (19) and (20), the sequence must be modified to be monotone. We will consider the modified sequence $\{\alpha'_i, i = 1, 2, \ldots, s\}$, defined by

$$\alpha'_s = \alpha_s \quad \text{and} \quad \alpha'_i = \min(\alpha_i, \alpha'_{i+1}), \quad \text{for } i = 1, \ldots, s-1\,. \tag{23}$$

The modified constants satisfy the following property.

Lemma 4.3 *Let $\{\alpha_i, i = 1, \ldots, s\}$ be any sequence of critical constants, and $\{\alpha'_i, i = 1, \ldots, s\}$ be a modified sequence of $\{\alpha_i, i = 1, \ldots, s\}$, which is defined in (23).*
(i) The sequence $\{\alpha'_i, i = 1, \ldots, s\}$ is monotone increasing.
(ii) $\alpha'_i \leq \alpha_i$, for $1 \leq i \leq s$, and if the sequence $\{\alpha_i, i = 1, 2, \ldots, s\}$ is monotone increasing, then $\alpha'_i = \alpha_i$.

Proof. The proof is obvious. ■

We now prove the k-FWER controllability of the stepdown procedure using the modified sequence $\{\alpha'_i, i = 1, \ldots, s\}$ defined in (23).

Theorem 4.1 *Under the assumptions of Theorem 3.1, further assume the p-values are mutually independent. Consider the stepdown procedure using critical constants*

$$\alpha'_1 \leq \cdots \leq \alpha'_s \ ,$$

defined in (23), and based on the sequence $\{\alpha_i, i = 1, \ldots, s\}$ satisfying (19) and (20).
(i) This procedure controls the k-FWER at level α .
(ii) If the sequence $\{\alpha_i, i = 1, \ldots, s\}$ is monotone increasing, then it is not possible to increase even one of the α_i with $i \geq 1$ without violating control of the k-FWER.
(iii) For $k = 2$, it is not possible to increase even one of the α_i with $i \geq 1$ without violating control of the k-FWER.

Proof. Fix any P and let $I = I(P)$ be the indices of the true null hypotheses. Assume $|I| \geq k$ or there is nothing to prove. Order the p-values corresponding to the $|I|$ true null hypotheses; call them

$$\hat{q}_{1:|I|} \leq \cdots \leq \hat{q}_{|I|:|I|} \ .$$

Let j be the smallest (random) index satisfying $\hat{p}_{(j)} = \hat{q}_{k:|I|}$. Then, the stepdown procedure commits at least k false rejections if and only if the event

$$\{\hat{p}_{(1)} \leq \alpha'_1, \hat{p}_{(2)} \leq \alpha'_2, \ldots, \hat{p}_{(j)} \leq \alpha'_j\} \ ,$$

which implies the event

$$\{\hat{q}_{1:|I|} \leq \alpha'_{j-k+1}, \hat{q}_{2:|I|} \leq \alpha'_{j-k+2}, \ldots, \hat{q}_{k:|I|} \leq \alpha'_j\} \tag{24}$$

occurs. Note that, $|I| \leq s - j + k$. Then, the event (24) implies the event

$$\{\hat{q}_{1:s-j+k} \leq \alpha'_{j-k+1}, \hat{q}_{2:s-j+k} \leq \alpha'_{j-k+2}, \ldots, \hat{q}_{k:s-j+k} \leq \alpha'_j\} \ .$$

So, the probability of at least k false rejections is bounded above by

$$\begin{aligned}
&P\{\hat{q}_{1:s-j+k} \le \alpha'_{j-k+1}, \hat{q}_{2:s-j+k} \le \alpha'_{j-k+2}, \ldots, \hat{q}_{k:s-j+k} \le \alpha'_j\} \\
&\le P\{\hat{q}_{1:s-j+k} \le \alpha_{j-k+1}, \hat{q}_{2:s-j+k} \le \alpha_{j-k+2}, \ldots, \hat{q}_{k:s-j+k} \le \alpha_j\} \\
&\le P\{U_{1:s-j+k} \le \alpha_{j-k+1}, U_{2:s-j+k} \le \alpha_{j-k+2}, \ldots, U_{k:s-j+k} \le \alpha_j\} \\
&= \alpha
\end{aligned}$$

The second inequality follows from the assumption (9).

We now prove that one cannot improve even one of the critical values without violating error control. Since α_i is monotone increasing, $\alpha'_i = \alpha_i$, for $i = 1, \ldots, s$. For $i < k$, one cannot improve the constant $\alpha_i = \alpha_k$ since we demand the α_i to be monotone increasing. For $i = k$, the constant α_k cannot be improved, as seen by the argument for Theorem 2.2 (ii). For $i > k$, let there be $s - i + k$ true null hypotheses and the rest false. Furthermore, let the false ones have p-values which are identically zero and let the true ones be i.i.d. $U(0,1)$ variables. Then, the false null hypotheses will be immediately rejected by the stepdown procedure and the event that the k-FWER is violated is identical to the event

$$\{\hat{q}_{1:s-i+k} \le \alpha_{i-k+1}, \hat{q}_{2:s-i+k} \le \alpha_{i-k+2}, \ldots, \hat{q}_{k:s-i+k} \le \alpha_i\} \ .$$

So, the k-FWER is

$$P\{\hat{q}_{1:s-i+k} \le \alpha_{i-k+1}, \hat{q}_{2:s-i+k} \le \alpha_{i-k+2}, \ldots, \hat{q}_{k:s-i+k} \le \alpha_i\} = \alpha$$

Hence, it is impossible to increase α_i and the k-FWER is bounded above by α .

Part (iii) directly follows Lemma 4.2 and (ii). ■

We now discuss how to compute the critical constants $\{\alpha_i, i = 1, 2, \ldots, s\}$ satisfying (19) and (20).

If α_i is monotone increasing, one could apply a recursive formula for order statistics in Finner and Roters (1994). If $0 \le x_1 \le x_2 \le \ldots \le x_s \le 1$ is an increasing sequence, then the following formula holds:

$$P\{U_{1:s} \le x_1, \ldots, U_{k:s} \le x_k\} = 1 - \sum_{j=0}^{k-1} \binom{s}{j} P\{U_{1:j} \le x_1, \ldots, U_{j:j} \le x_j\}(1 - x_{j+1})^{s-j} \ . \tag{25}$$

For $i > k$, suppose that $\alpha_1, \ldots, \alpha_{i-1}$ have been computed. By using (25), α_i

can be obtained by solving the following equation,

$$\begin{aligned} &P\{U_{1:s-i+k} \le \alpha_{i-k+1}, U_{2:s-i+k} \le \alpha_{i-k+2}, \ldots, U_{k:s-i+k} \le \alpha_i\} \\ &= 1 - \sum_{j=0}^{k-2} \binom{s-i+k}{j} P\{U_{1:j} \le \alpha_{i-k+1}, \ldots, U_{j:j} \le \alpha_{i-k+j}\}(1-\alpha_{i-k+j+1})^{s-i+k-j} \\ &\quad - \binom{s-i+k}{k-1} P\{U_{1:k-1} \le \alpha_{i-k+1}, \ldots, U_{k-1:k-1} \le \alpha_{i-1}\}(1-\alpha_i)^{s-i+1} \\ &= \alpha \ , \end{aligned} \tag{26}$$

where $P\{U_{1:j} \le \alpha_{i-k+1}, \ldots, U_{j:j} \le \alpha_{i-k+j}\}$ is recursively computed by using (25),

$$\begin{aligned} &P\{U_{1:j} \le \alpha_{i-k+1}, U_{2:j} \le \alpha_{i-k+2}, \ldots, U_{j:j} \le \alpha_{i-k+j}\} \\ &= 1 - \sum_{l=0}^{j-1} \binom{j}{l} P\{U_{1:l} \le \alpha_{i-k+1}, \ldots, U_{l:l} \le \alpha_{i-k+l}\}(1-\alpha_{i-k+l+1})^{j-l} \ . \end{aligned} \tag{27}$$

If α_i is not monotone increasing, one could apply Newton's method in numerical analysis to approximate α_i. Let $H(u) = G(u) - \alpha$, where $G(u)$ is defined in (21). Then, $H(\alpha_i) = 0$. Note that, $U_{1:s-i+k}/U_{k:s-i+k}, \ldots, U_{k-1:s-i+k}/U_{k:s-i+k}$ are the order statistics of $k-1$ i.i.d. $U(0,1)$ on (0,1), independent of $U_{k:s-i+k}$. Then,

$$\begin{aligned} &H(u) = P\{U_{1:s-i+k} \le \alpha_{i-k+1}, \ldots, U_{k-1:s-i+k} \le \alpha_{i-1}, U_{k:s-i+k} \le u\} - \alpha \\ &= \int_0^u P\{U_{1:k-1} \le \frac{\alpha_{i-k+1}}{x}, \ldots, U_{k-1:k-1} \le \frac{\alpha_{i-1}}{x}\} \frac{(s-i+k)!}{(k-1)!(s-i)!} x^{k-1}(1-x)^{s-i} dx - \alpha \end{aligned}$$

and so

$$H'(u) = P\{U_{1:k-1} \le \frac{\alpha_{i-k+1}}{u}, \ldots, U_{k-1:k-1} \le \frac{\alpha_{i-1}}{u}\} \frac{(s-i+k)!}{(k-1)!(s-i)!} u^{k-1}(1-u)^{s-i} \ .$$

Newton's method can be described by

$$\begin{aligned} x_{n+1} &= x_n - H(x_n)/H'(x_n) \\ &= x_n - \frac{P\{U_{1:s-i+k} \le \alpha_{i-k+1}, \ldots, U_{k:s-i+k} \le x_n\} - \alpha}{P\{U_{1:k-1} \le \frac{\alpha_{i-k+1}}{x_n}, \ldots, U_{k-1:k-1} \le \frac{\alpha_{i-1}}{x_n}\} \frac{(s-i+k)!}{(k-1)!(s-i)!} x_n^{k-1}(1-x_n)^{s-i}} \ . \end{aligned} \tag{28}$$

Although the solution α_i^0 of equation (26) is not the correct value of the unknown α_i if α_i is less than α_{i-1} , α_i^0 is still close to α_i. In addition, α_{i-1} is also close to α_i, so one could choose α_i^0 or α_{i-1} as the initial point x_0 .

Also, Newton's iteration has a local convergence property since $H'(\alpha_i) > 0$. That is, the iteration will converge to α_i if the stating point α_i^0 (or α_{i-1}) is close enough to α_i (Kantrovich and Akilov 1964).

Remark 4.1 When the constants $c_1, \ldots, c_j$ are given, even though the constants are not monotone increasing, one could also use (25) to compute the probability $P\{U_{1:v} \leq c_1, \ldots, U_{j:v} \leq c_j\}$. The extra work just requires setting $c'_l = \min(c_l, 1)$, for $l = 1, \ldots, j$, and then modifying $c'_1, \ldots, c'_j$ according to (23).

Remark 4.2 Before computing α_i using (26), one does not know whether the critical constants are monotone increasing. If the solved α_i is monotone increasing, i.e., $\alpha_i \geq \alpha_{i-1}$, it shows that α_i is the correct solution. Otherwise, the solved α_i is not correct, and one could use Newton's method to compute the correct constant.

We now perform some numerical comparisons of the critical constants of two new stepdown procedures in Theorems 3.2 and 4.1 and Sarkar's procedure based on (15) with $\alpha'_i = \alpha'_k$. We first plot in Figure 3 the two sequences of constants described in (15) and (16) for the case in which $k = 10, s = 200$, and $\alpha = 0.05$. Panel (a) displays the critical constants based on (15), where panel (b) displays the critical constants based on (16). Panel (c) displays the ratio of the constants in panel (b) with the constants in panel (a). The dashed horizontal lines in panel (a) and (b) are of height 0.05. It is clear from panel (c) that the constants in panel (b) are much larger than the constants of panel (a). For example, more than 90% of the constants of panel (b) are at least 40% larger than the constants of panel (a). Thus, if the assumption of independence of p-values is satisfied, the stepdown procedure based on the constants in panel (b) is preferable to the one based on the constants in panel (a).

Next, we plot in Figure 4 the two sequences of constants described in (15), and (19), (20) and (23) for the case in which $k = 10, s = 200$, and $\alpha = 0.05$. Panel (a) displays the critical constants based on (15), where panel (b) displays the critical constants based on (19), (20) and (23). Panel (c) displays the ratio of the constants in panel (b) with the constants in panel (a). It is clear from panel (c) that the constants in panel (b) are uniformly larger and typically much larger than the constants of panel (a). Comparing panel (c) in Figure 3 and Figure 4, we find that their patterns are very similar except that there are a short sequence of irregular changes in Figure 4.

In Figure 5, we plot the two sequences of constants described in (16), and (19) and (20) for the case in which $k = 8, s = 200$, and $\alpha = 0.05$. Panel (a) displays the critical constants based on (16), where panel (b) displays the critical constants based on (19) and (20). Panel (c) displays the ratio of the constants in panel (b) with the constants in Panel (a), and panel (d) displays a short sequence of ratios in panel (c) from the first to the 160th one. It is clear from panel (c) that the constants in panel (b) are almost the same as the constants of panel (a), except for the last twenty constants. Further, we find from panel (d) that there is less than 0.2% difference between the first 160 constants in panels (a) and (b), and neither set of critical constants is uniformly larger than the another one. In addition,

Figure 3: The critical constants of two k-FWER stepdown procedures based on (15) and (16) for $k = 10, s = 200,$ and $\alpha = 0.05$.

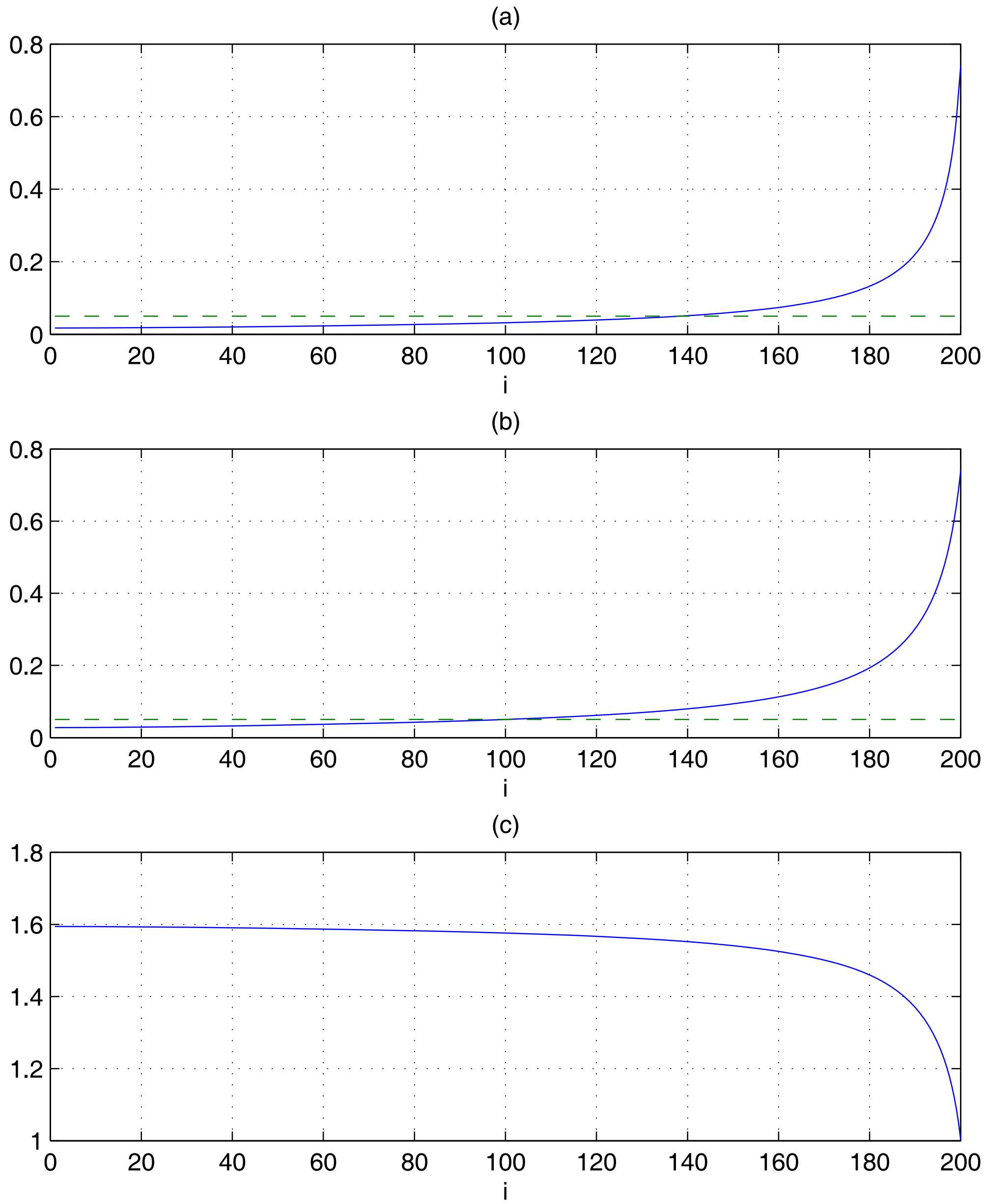


Figure 4: The critical constants of two k-FWER stepdown procedures based on (15) and (19), (20), and (23) for $k = 10, s = 200,$ and $\alpha = 0.05$.

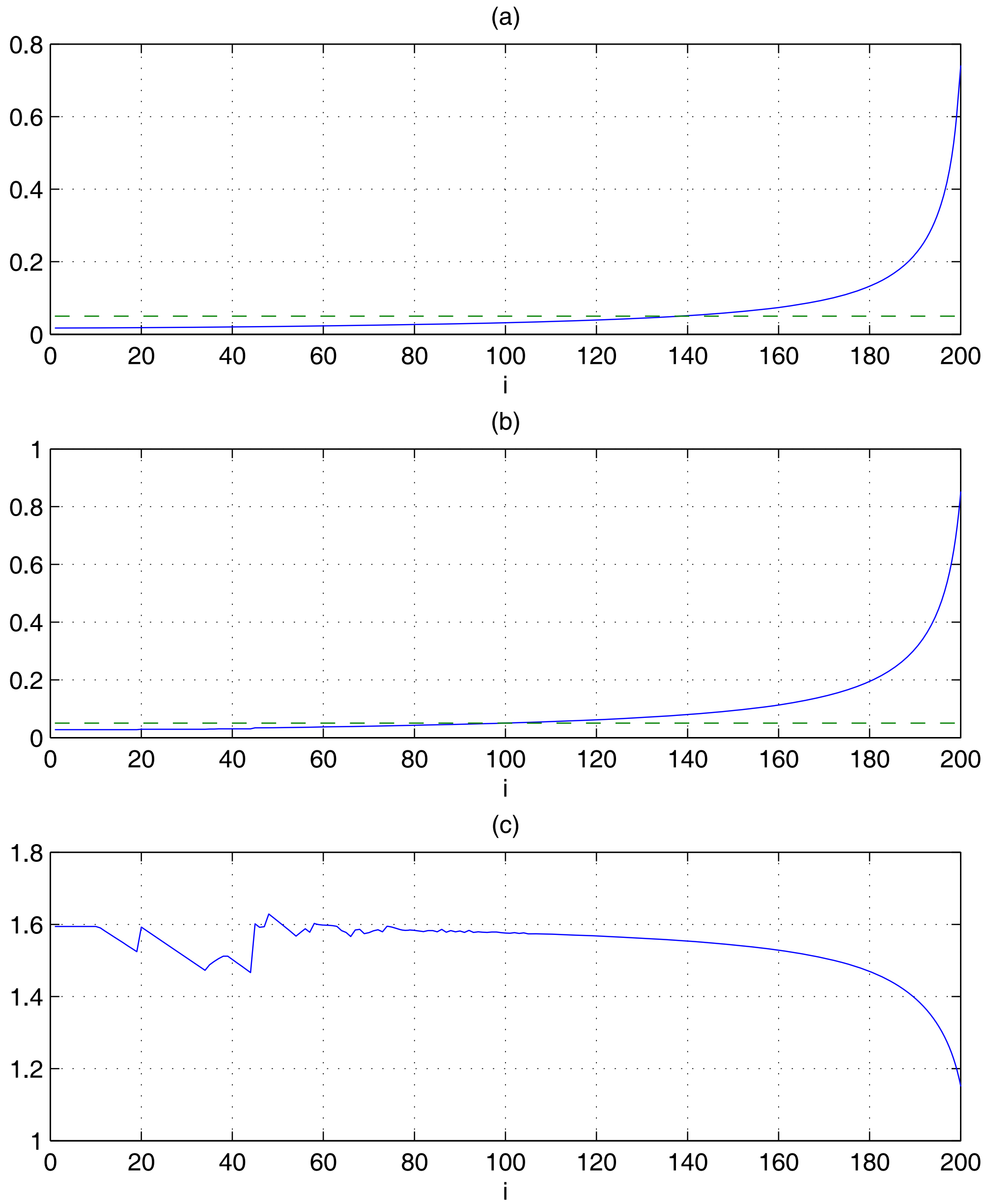


Figure 5: The critical constants of two k-FWER stepdown procedures based on (16), and (19) and (20) for $k = 8, s = 200,$ and $\alpha = 0.05$.

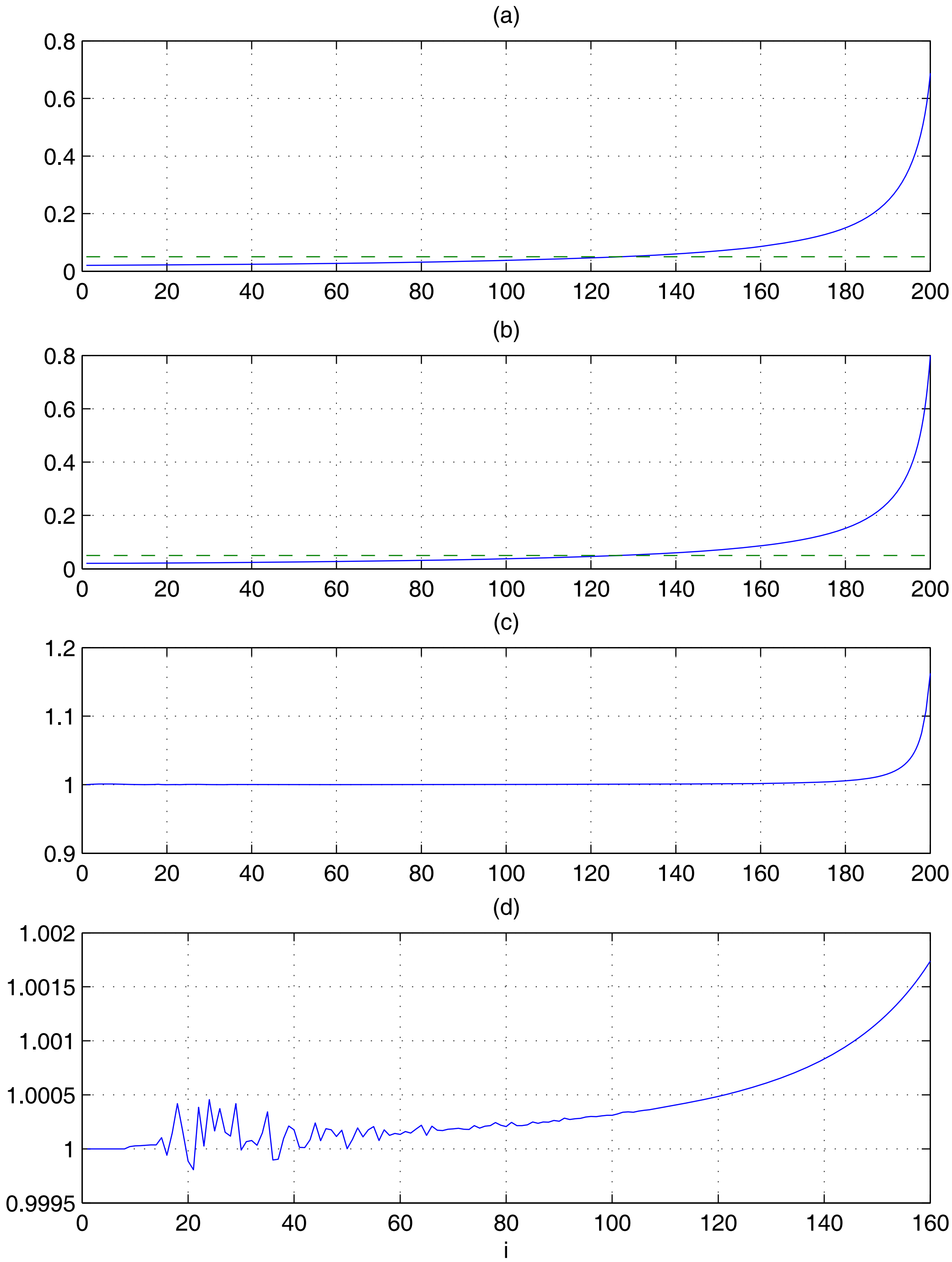
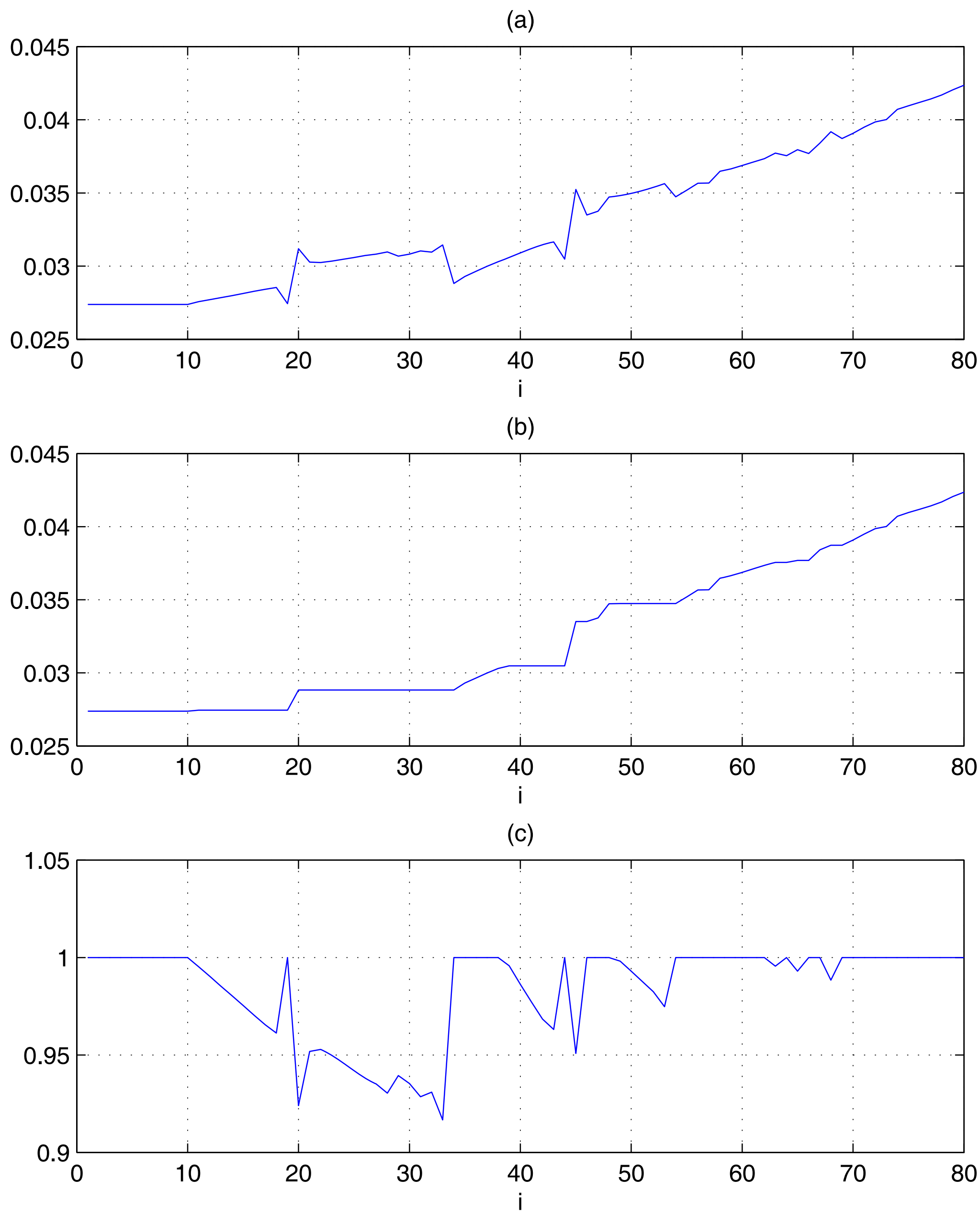


Figure 6: The critical constants based on (19) and (20), and (23) for $k = 10, s = 200$, and $\alpha = 0.05$.



note that the critical constants in panel (b) satisfying (19) and (20) are monotone increasing in this case. Thus, it seems that the stepdown procedure based on the constants in panel (a) is almost unimprovable.

Next, we plot in Figure 6 the two short sequences of constants described in (19) and (20), and (23) for the case in which $k = 10, s = 200$, and $\alpha = 0.05$. Panel (a) displays the critical constants based on (19) and (20), where Panel (b) displays the modified constants based on (23). Panel (c) displays the ratio of the constants in panel (b) with the constants in panel (a). Panel (a) shows that the critical constants satisfying (19) and (20) are not monotone increasing, and panel (c) shows that the modified constants in panel (b) are reduced by at most 7%. Thus, it seems that the stepdown procedure based on the constants in panel (b) is close to being unimprovable.

Remark 4.3 To appreciate why the stepdown method based on (16) is nearly unimprovable, consider a stepdown method using constants $\alpha_1, \ldots, \alpha_s$ which is assumed to control the k-FWER under independence. Then, we show that the critical constants α_i for $i \geq k$ must satisfy

$$\min(\alpha_i, \ldots, \alpha_{i+k-1}) \leq C_{k,s-i+1}(\alpha) \ . \tag{29}$$

In the case where the constants α_i are nondecreasing for $i \geq k$, this says

$$\alpha_i \leq C_{k,s-i+1}(\alpha) \ . \tag{30}$$

Here $C_{k,s-i+1}(\alpha)$ is the critical constant that the stepdown procedure given in (16) uses, not at step i, but at step $i+k-1$. Since, the constants $C_{k,s-j}(\alpha)$ do not vary much with j (except for very large j) for fixed small k, $C_{k,s-i+1}(\alpha)$ is not much bigger than $C_{k,s-i+k}(\alpha)$ anyway.

To prove (29), suppose $i-1$ hypotheses are false with p-values identically 0 and $s-i+1$ hypotheses are true, with p-values i.i.d. uniform. Let

$$\hat{q}_{(1)} \leq \cdots \leq \hat{q}_{(s-i+1)}$$

be the ordered p-values corresponding to the true hypotheses. Then, k or more false rejections occurs is the event

$$\{\hat{q}_{(1)} \leq \alpha_i, \hat{q}_{(2)} \leq \alpha_{i+1}, \ldots, \hat{q}_{(k)} \leq \alpha_{i+k-1}\} \ ,$$

and the k-FWER is probability of this event. But, then a lower bound to the k-FWER is

$$P\{\hat{q}_{(k)} \leq \min(\alpha_i, \ldots, \alpha_{i+k-1})\} = H_{k,s-i+1}(\min(\alpha_i, \ldots, \alpha_{i+k-1})) \ ,$$

and this much be $\leq \alpha$ since we are assuming the procedure controls the k-FWER. But, since

$$H_{k,s-i+1}(C_{k,s-i+1}(\alpha)) = \alpha \ ,$$

(29) follows.

5 Stepdown control of the FDP

In this section, we discuss the control of the false discovery proportion (FDP), which is defined by,

$$FDP = \begin{cases} \frac{V}{R} & R > 0 \\ 0 & R = 0 \end{cases} \tag{31}$$

where, V is the number of false rejections, and R is the total number of rejections. The FDP is the proportion of rejected hypotheses that are rejected erroneously.

For a given γ and α in (0,1), we require

$$P\{FDP > \gamma\} \leq \alpha \tag{32}$$

To develop a stepdown procedure satisfying (32), we want to control $FDP \leq \gamma$ at each step. That is, at step i, having rejected $i-1$ hypotheses, we want to guarantee $V/i \leq \gamma$, i.e., $V \leq \lfloor \gamma i \rfloor$, where $\lfloor x \rfloor$ is the greatest integer $\leq x$. So, if $k(i) = \lfloor \gamma i \rfloor + 1$, then $V \geq k(i)$ should have probability no greater than α; that is, we must control the number of false rejections to be $\leq k(i)$. Based on the above heuristics, Lehmann and Romano (2005a) constructed the constant α_i with this choice of $k(i)$ like in Theorem 3.1, i.e.,

$$\alpha_i = \frac{(\lfloor \gamma i \rfloor + 1)\alpha}{s + \lfloor \gamma i \rfloor + 1 - i} , \tag{33}$$

and proved that the stepdown procedure using the constants (33) controls the FDP at level α under mild conditions on the dependence structure of p-values. Under the independence assumption, this procedure can be improved, as we now show.

Theorem 5.1 *Under the assumptions of Theorem 3.1, further assume the p-values are mutually independent. Consider the stepdown procedure using critical constants*

$$\alpha_1 \leq \cdots \leq \alpha_s ,$$

where

$$\alpha_i = H^{-1}_{k(i),s-i+k(i)}(\alpha) \tag{34}$$

and,

$$k(i) = \lfloor \gamma i \rfloor + 1 , \quad \text{for } 1 \leq i \leq s .$$

Then, this procedure controls the FDP at level α.

Proof. Assume the number of true null hypotheses is $|I| > 0$ and the number of false null hypotheses is $f = s - |I|$. Let $\hat{q}_1, \ldots, \hat{q}_{|I|}$ denote the p-values corresponding to the $|I|$ true null hypotheses, and $\hat{r}_1, \ldots, \hat{r}_f$ denote the p-values of the false null hypotheses. Let $\alpha_0 = 0$ and R_i be the number of $\hat{r}_i$ in the interval $(\alpha_{i-1}, \alpha_i]$.

(Actually, assume R_1 includes the value 0 as well.) Given the values of $\hat{r}_1, \ldots, \hat{r}_f$, define $j = j(\hat{r}_1, \ldots, \hat{r}_f)$ as

$$j = \min\{m : m - \sum_{i=1}^{m} R_i > m\gamma\}. \quad (35)$$

Following the proof of Theorem 3.1 in Lehmann and Romano (2005a), observe that, conditional on the $\{\hat{r}_i\}$, the event that $FDP > \gamma$ is violated implies that the event that at least $k(j)$ of the $\hat{q}_i \leq \alpha_j$. That is,

$$P\{FDP > \gamma | \hat{r}_1, \ldots, \hat{r}_f\} \leq P\{\text{at least } k(j) \text{ of the } \hat{q}_i \leq \alpha_j | \hat{r}_1, \ldots, \hat{r}_f\}$$

Then, by Theorem 3.2 and the independence assumption, we have

$$P\{FDP > \gamma | \hat{r}_1, \ldots, \hat{r}_f\} \leq \alpha \quad (36)$$

For some values of $\hat{r}_1, \ldots, \hat{r}_f$, the set $\{m : m - \sum_{i=1}^{m} R_i > m\gamma\}$ is empty; that is, for each $m, m - \sum_{i=1}^{m} R_i \leq m\gamma$. It is equivalent to $FDP \leq \gamma$. So, we have

$$P\{FDP > \gamma | \hat{r}_1, \ldots, \hat{r}_f\} = 0 \quad (37)$$

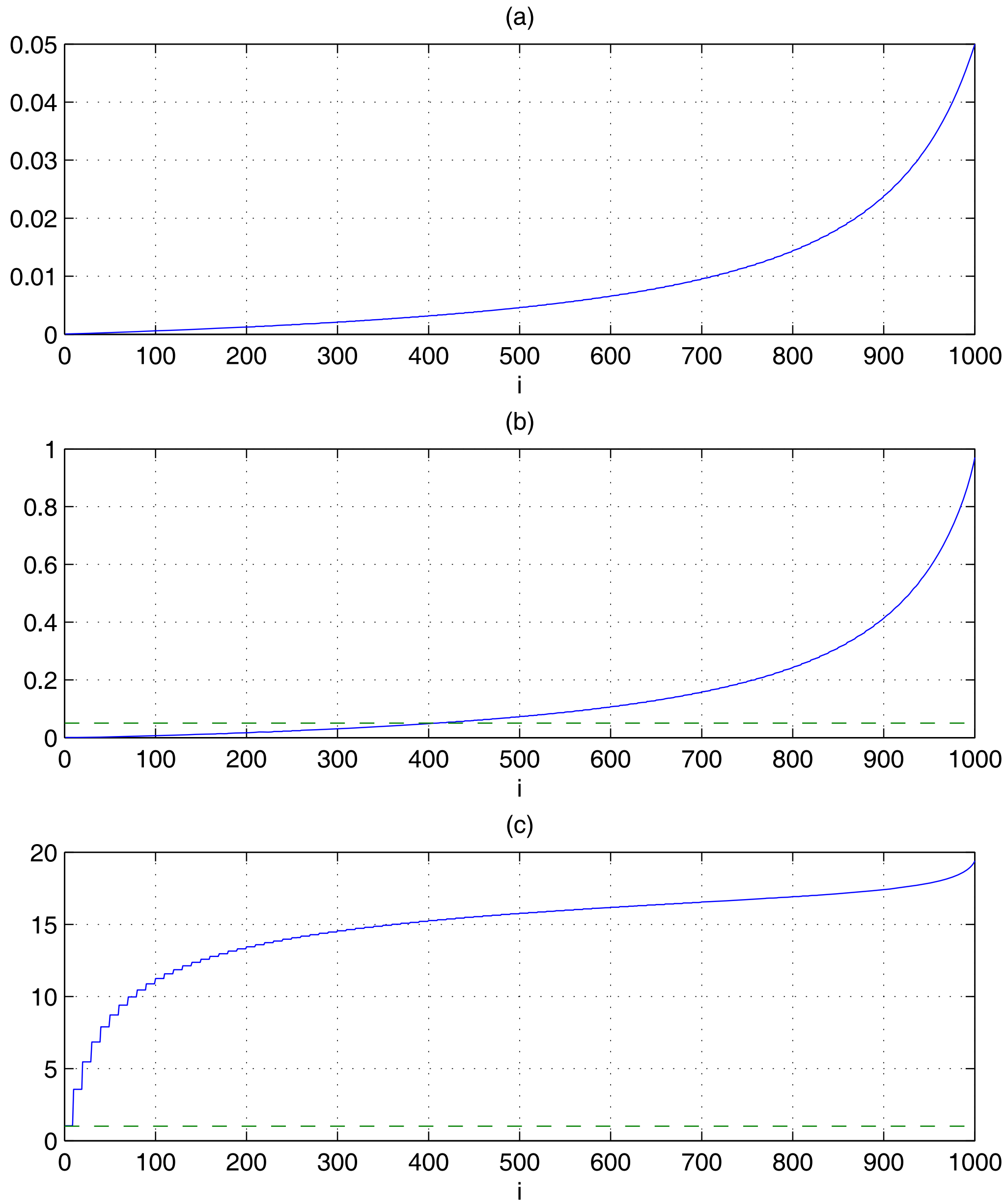
By (36) and (37), $P\{FDP > \gamma\} \leq \alpha$ is proved. ■

We now perform some numerical comparisons of the critical constants of the stepdown procedure in Theorem 5.1 with that of Lehmann and Romano's stepdown procedure based on (33). We plot in Figure 7 the two sequences of constants for the case in which $s = 1000, \gamma = 0.1$, and $\alpha = 0.05$. Panel (a) displays the critical constants based on (33), where panel (b) displays the critical constants based on (34). Panel (c) displays the ratio of the constants in panel (b) with the constants in panel (a). The dashed horizontal lines in panel (b) and (c) are of height 0.05 and 1, respectively. It is clear that the constants in panel (b) are dramatically larger than the constants of panel (a). For example, more than 90% of the constants of panel (b) are 10 times larger than the constants of panel (a). Thus, if the assumption of independence of p-values is satisfied, the stepdown procedure based on the constants in panel (b) is preferable to the one based on the constants in panel (a).

6 Conclusions

We have seen that very simple single step and stepdown procedures are available to control the k-FWER under the assumption of independence of the p-values. Therefore, the methods are an improvement over Hommel and Hoffmann (1987) and Lehmann and Romano (2005a) if independence holds, and can be viewed as a generalization of Sidák's method as a means of controlling the k-FWER. In

Figure 7: The critical constants of two FDP stepdown controlling procedures based on (33) and (34) for $s = 1000, \gamma = 0.1,$ and $\alpha = 0.05$.



the case $k = 1$, the method cannot be improved. We discuss the improvability in general and obtain an optimality result under a monotonicity condition. In the case $k = 2$, the monotonicity condition is satisfied. In other cases, it can be verified numerically. We also develop a method that controls the FDP. The perhaps surprising result of the wrok is that, for generalized error rates, the assumption of independence can yield dramatic gains. At the very least, if the independence is unwarranted, the results enforce the need for reliable methods that take into account the dependence structure, such as those in Dudoit, et. al. (2004), van der Laan, et. al. (2004, 2005) and Romano and Wolf (2005).

APPENDIX

A Proof of Lemma 4.2

For convenience of notation, let $\beta_1 = \beta_2$ and $\beta_i = H^{-1}_{2,s-i+2}(\alpha)$ for $i = 2, \ldots, s$. The β_i are monotone increasing and satisfy the following equality,

$$P\{U_{2:s-i+2} \le \beta_i\} = \alpha, \text{ for } i = 2, \ldots, s. \tag{A.1}$$

When $k = 2$, the critical constants $\alpha_i, i = 1, \ldots, s$ based on (19) and (20) satisfy the following equation,

$$P\{U_{1:s-i+2} \le \alpha_{i-1}, U_{2:s-i+2} \le \alpha_i\} = \alpha, \text{ for } i = 2, \ldots, s, \tag{A.2}$$

where $\alpha_1 = \alpha_2$. For $i = 2$, $\alpha_1 = \alpha_2 = \beta_2$ follows from (A.2). So, $\beta_2 \le \alpha_2 \le \beta_3$.

We now prove the following inequality using induction rule,

$$\beta_i \le \alpha_i \le \beta_{i+1}, \text{ for } i = 2, \ldots, s-1. \tag{A.3}$$

For $i = 2$, (A.3) holds. For $i > 2$, suppose $\beta_{i-1} \le \alpha_{i-1} \le \beta_i$ hold. Comparing (A.1) with (A.2), and using $\alpha_{i-1} \le \beta_i$, we have $\beta_i \le \alpha_i$. Similarly, comparing (A.2) with (A.21) in Lemma A.4, and using $\beta_{i-1} \le \alpha_{i-1}$, then $\alpha_i \le \beta_{i+1}$. So, $\beta_i \le \alpha_i \le \beta_{i+1}$, and then (A.3) is proved by induction.

Based on (A.3), the inequality $\alpha_s > \beta_s$ can be similarly obtained. So, the critical constants $\alpha_i, i = 1, 2, \ldots, s$ are monotone increasing when $k = 2$. Thus, Lemma 4.2 is proved. ■

In the following, we prove Lemma A.4. Before proving Lemma A.4, we present several equalities and lemmas.

Note that, (A.1) can be expressed as,

$$1 - (1 - \beta_i)^{s-i+2} - (s - i + 2)\beta_i(1 - \beta_i)^{s-i+1} = \alpha\ . \tag{A.4}$$

So,

$$(1 - \beta_i)^{s-i+1} = \frac{1 - \alpha}{1 + (s - i + 1)\beta_i}, \text{ for } i = 2, \ldots, s. \tag{A.5}$$

Lemma A.1 *For any $-1 < x < 1$, and positive integer m, the following inequality holds:*

$$(1+x)^m \geq 1 + mx \ . \tag{A.6}$$

Proof. The proof is obvious. ∎

Lemma A.2 *For any positive integer $s \geq 2$, the following inequality holds:*

$$(s-i)\beta_{i+1} \leq (s-i+1)\beta_i, \text{ for } i = 2, \ldots, s-1. \tag{A.7}$$

Proof. Note that,

$$\begin{aligned}(\frac{1-\beta_i}{1-\beta_{i+1}})^{s-i+1} &= (1 + \frac{\beta_{i+1}-\beta_i}{1-\beta_{i+1}})^{s-i+1} \\ &\geq 1 + (s-i+1)\frac{\beta_{i+1}-\beta_i}{1-\beta_{i+1}}\end{aligned} \tag{A.8}$$

The above inequality follows from Lemma A.1. From (A.5), the left hand side of (A.8) can be simplified as,

$$(\frac{1-\beta_i}{1-\beta_{i+1}})^{s-i+1} = \frac{1+(s-i)\beta_{i+1}}{(1-\beta_{i+1})[1+(s-i+1)\beta_i]} \tag{A.9}$$

Combining (A.8) and (A.9), (A.7) is proved. ∎

Lemma A.3 *For any positive integer $s \geq 2$, the following inequality holds:*

$$(s-i+1)\beta_{i+1} \geq (s-i+2)\beta_{i-1}(1-\beta_{i+1}), \text{ for } i = 2, \ldots, s-1. \tag{A.10}$$

Proof. Similar to the proof of Lemma A.2, the following inequality follows from Lemma A.1:

$$\begin{aligned}(\frac{1-\beta_{i+1}}{1-\beta_{i-1}})^{s-i+2} &= (1 - \frac{\beta_{i+1}-\beta_i}{1-\beta_{i-1}})^{s-i+2} \\ &\geq 1 - (s-i+2)\frac{\beta_{i+1}-\beta_{i-1}}{1-\beta_{i-1}} \ .\end{aligned} \tag{A.11}$$

From (A.5), the left hand side of (A.11) can be simplified as

$$(\frac{1-\beta_{i+1}}{1-\beta_{i-1}})^{s-i+2} = \frac{(1-\beta_{i+1})^2[1+(s-i+2)\beta_{i-1}]}{1+(s-i)\beta_{i+1}} \ . \tag{A.12}$$

Combining (A.11) and (A.12), we have

$$\begin{aligned}&(1-\beta_{i-1})(1-\beta_{i+1})^2[1+(s-i+2)\beta_{i-1}] \\ &\qquad \geq [1+(s-i)\beta_{i+1}][1-(s-i+2)\beta_{i+1}+(s-i+1)\beta_{i-1}] \ .\end{aligned} \tag{A.13}$$

Simplifying (A.13),

$$\begin{aligned}(s-i+1)^2\beta_{i+1}^2 \geq &(s-i+2)\beta_{i-1}^2 + (s-i+1)(s-i+2)\beta_{i-1}\beta_{i+1} \\ &- (s-i+1)\beta_{i-1}\beta_{i+1}^2 - 2(s-i+2)\beta_{i-1}^2\beta_{i+1} \\ &+ (s-i+2)\beta_{i-1}^2\beta_{i+1}^2 \ .\end{aligned} \tag{A.14}$$

Note that

$$\begin{aligned}(s-&i+2)\beta_{i-1}^2 + (s-i+1)(s-i+2)\beta_{i-1}\beta_{i+1} \\ &= (s-i+2)\beta_{i-1}[\beta_{i-1} + (s-i+1)\beta_{i+1}] \\ &\geq (s-i+2)^2\beta_{i-1}^2 \ ,\end{aligned} \tag{A.15}$$

and from Lemma A.2,

$$\begin{aligned}-(s-i+1)\beta_{i-1}\beta_{i+1}^2 &= -\frac{s-i+1}{s-i}\beta_{i-1}\beta_{i+1}[(s-i)\beta_{i+1}] \\ &\geq -\frac{(s-i+1)(s-i+2)}{s-i}\beta_{i-1}^2\beta_{i+1} \ .\end{aligned} \tag{A.16}$$

Then, the second part of the right hand side of (A.14) is expressed as

$$\begin{aligned}&- (s-i+1)\beta_{i-1}\beta_{i+1}^2 - 2(s-i+2)\beta_{i-1}^2\beta_{i+1} + (s-i+2)\beta_{i-1}^2\beta_{i+1}^2 \\ &\geq (s-i+2)\beta_{i-1}^2\beta_{i+1}[-(2+\frac{s-i+1}{s-i}) + \beta_{i+1}] \ .\end{aligned} \tag{A.17}$$

Note that

$$(s-i+1)(1-\beta_{i+1}) + [(s-i+1) - \frac{s-i+1}{s-i}] \geq 0 \ ; \tag{A.18}$$

that is,

$$-(2+\frac{s-i+1}{s-i}) + \beta_{i+1} \geq (s-i+2)(-2+\beta_{i+1}) \ . \tag{A.19}$$

Combining (A.14), (A.15), (A.17), and (A.19), we have,

$$\begin{aligned}(s-i+1)^2\beta_{i+1}^2 &\geq (s-i+2)^2(\beta_{i-1}^2 - 2\beta_{i-1}^2\beta_{i+1} + \beta_{i-1}^2\beta_{i+1}^2) \\ &= (s-i+2)^2[\beta_{i-1}(1-\beta_{i+1})]^2 \ .\end{aligned} \tag{A.20}$$

So, (A.10) is proved. ■

We now prove Lemma A.4.

Lemma A.4 *For any positive integer $s \geq 2$, the following inequality holds:*

$$P\{U_{1:s-i+2} \leq \beta_{i-1}, U_{2:s-i+2} \leq \beta_{i+1}\} \geq \alpha, \text{ for } i = 2, \ldots, s-1. \tag{A.21}$$

PROOF. Note that,

$$P\{U_{1:s-i+2} \le \beta_{i-1}, U_{2:s-i+2} \le \beta_{i+1}\}$$
$$= 1 - (1-\beta_{i-1})^{s-i+2} - (s-i+2)\beta_{i-1}(1-\beta_{i+1})^{s-i+1} \ . \qquad \text{(A.22)}$$

Comparing (A.4) with (A.22), we find that, (A.22) is $\ge \alpha$ if and only if

$$1 - (1-\beta_{i-1})^{s-i+2} - (s-i+2)\beta_{i-1}(1-\beta_{i+1})^{s-i+1}$$
$$\ge 1 - (1-\beta_{i-1})^{s-i+3} - (s-i+3)\beta_{i-1}(1-\beta_{i-1})^{s-i+2} \ ;$$

that is,

$$(1-\beta_{i-1})^{s-i+2} \ge (1-\beta_{i+1})^{s-i+1} \ . \qquad \text{(A.23)}$$

By (A.5) and Lemma A.3, (A.23) is obtained, and then (A.21) is proved. ■

References

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

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

Dudoit, S., van der Laan, M. and Birkner, M. (2004). Multiple testing procedures for controlling tail probability error rates. Working Paper 166, U.C. Berkeley Division of Biostatistics, Available at http://www.bepress.com/ucbbiostat/paper166/.

Dudoit, S., van der Laan, M. and Pollard, K. (2004). Multiple testing. Part I. Single-step procedures for control of general type I error rates. *Statistical Applications in Genetics and Molecular Biology*, 3(1): Article 13. Available at http://www.bepress.com/sagmb/vol3/iss1/art13.

Finner, H. and Roters, M. (1994). On the limit behaviour of the joint distribution function of order statistics. *Annals of the Institute of Statistical Mathematics*, **46**, 343-349.

Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control. *Annals of Statistics*, **32**, 1035–1061.

Guo, W. and Rao, B. (2006). On generalized closure principle for generalized familywise error rates. University of Cincinnati, Unpublished report.

Hochberg, Y. and Tamhane, A. (1987). *Multiple Comparison Procedures.* John Wiley, New York.

Holland, B. and Copenhaver, M. (1987). An improved sequentially rejective Bonferroni test procedure. *Biometrics*, **43**, 417–423.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**, 65-70.

Hommel, G. and Hoffmann, T. (1987). Controlled uncertainty. In *Multiple Hypothesis Testing* (P. Bauer, G. Hommel and E. Sonnemann, eds.) 154-161. Springer, Heidelberg.

Kantrovich, L. V. and Akilov, G. P. (1964). *Functional Analysis in Normed Spaces.* Pergamon Press, Elmford, New York.

Korn, E., Troendle, J., McShane, L. and Simon, R. (2004). Controlling the number of false discoveries: application to high-dimensional genomic data. *Journal of Statistical Planning and Inference.* **124**, 379–398.

Lehmann, E. L. and Romano, J. (2005a). Generalizations of the familywise error rate. *Annals of Statistics* **33**, 1138–1154.

Lehmann, E. L. and Romano, J. (2005b). *Testing Statistical Hypotheses*, 3rd edition, Springer-Verlag, New York.

Miller, R. (1981). *Simultaneous Statistical Inference*, 2nd edition, Springer-Verlag, New York.

Perone Pacifico, M., Genovese, C., Verdinelli, I. and Wasserman, L. (2004). False discovery rates for random fields. *Journal of the American Statistical Association.* **99**, 1002–1014.

Romano, J. and Shaikh, A. (2006a). On stepdown control of the false discovery proportion. In *The Second E. L. Lehmann Symposium – Optimality*, ed. Javier Rojo, Institute of Mathematical Statistics Lecture Notes-Monograph Series, Volume 49, 40–61.

Romano, J. and Shaikh, A. (2006b). Stepup procedures for control of generalizations of the familywise error rate. *Annals of Statistics*, to appear.

Romano, J. and Wolf, M. (2005). Control of generalized error rates in multiple testing. Technical Report 2005-12, Department of Statistics, Stanford University.

Sarkar, S. (2005). Generalizing Simes' test and Hochberg's stepup procedure. Unpublished report.

Sarkar, S. and Chang, C. (1997). The Simes method for multiple hypothesis testing with positively dependent test statistics. *Journal of the American Statistical Association.* **92**, 1601-1608.

Simes, R. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73**, 751–754.

van der Laan, M., Birkner, M., and Hubbard, A. (2005). Resampling based multiple testing procedure controlling tail probability of the proportion of false positives. Working Paper 172, U.C. Berkeley Division of Biostatistics. Available at `http://www.bepress.com/ucbbiostat/paper172`

van der Laan, M., Dudoit, S., and Pollard, K. (2004). Augmentation procedures for control of the generalized family-wise error rate and tail probabilities for the proportion of false positives. *Statistical Applications in Genetics and Molecular Biology*, **3**(1), Article 15. Available at `http://www.bepress.com/sagmb/vol3/iss1/art14`

Victor, N. (1982). Exploratory data analysis and clinical research. *Methods of Information in Medicine* **21**, 53-54.