Chapter 1

Stepdown Procedures Controlling A Generalized False Discovery Rate

Wenge Guo and Sanat K. Sarkar*

Department of Mathematical Sciences
New Jersey Institute of Technology
Newark, NJ 07102, U.S.A.
wenge.guo@njit.edu
Department of Statistics, Temple University,
Philadelphia, PA 19122, U.S.A.
sanat@temple.edu

Often in practice when a large number of hypotheses are simultaneously tested, one is willing to allow a few false rejections, say at most $k-1$, for some fixed $k > 1$. In such a case, the ability of a procedure controlling an error rate measuring at least one false rejection can potentially be improved in terms of its ability to detect false null hypotheses by generalizing this error rate to one that measures at least k false rejections and using procedures that control it. The k-FDR which is the expected proportion of k or more false rejections and a natural generalization of the false discovery rate (FDR) is such a generalized notion of error rate that has recently been introduced and procedures controlling it have been proposed. Many of these procedures are stepup procedures. Some stepdown procedures controlling the k-FDR are presented in this article.

1.1. Introduction

For simultaneous testing of null hypotheses using tests that are available for each of them, procedures have traditionally been developed exercising a control over the familywise error rate (FWER), which is the probability of rejecting at least one true null hypothesis [Hochberg and Tamhane (1987)], until it is realized that this notion of error rate is too stringent while test-

*The research of Wenge Guo is supported by NSF Grants DMS-1006021 and DMS-1309162 and the research of Sanat Sarkar is supported by NSF Grant DMS-1006344 and DMS-1309273.

ing a large number of hypotheses, as it happens in many modern scientific investigations, allowing little chance to detect many false null hypotheses. Therefore, researchers have focused in the last decade or so on defining alternative less stringent error rates and developing multiple testing methods that control them.

The false discovery rate (FDR), which is the expected proportion of falsely rejected among all rejected null hypotheses, introduced by Benjamini and Hochberg (1995), is the first of these alternative error rates that has received the most attention [Benjamini, Krieger and Yekutieli (2006), Benjamini and Yekutieli (2001, 2005), Finner, Dickhaus and Roters (2007, 2009), Gavrilov, Benjmaini and Sarkar (2009), Genovese and Wasserman (2002, 2004), Sarkar (2002, 2004, 2006, 2008a), Storey (2002, 2003) and Storey, Taylor and Siegmund (2004)]. Recently, the ideas of controlling the probabilities of falsely rejecting at least k null hypotheses, which is the k-FWER, and the false discovery proportion (FDP) exceeding a certain threshold $\gamma \in [0, 1)$ have been introduced as alternatives to the FWER and methods controlling these new error rates have been suggested [Dudoit, van der Laan and Pollard (2004), Guo and Rao (2010), Guo and Romano (2007), Hommel and Hoffmann (1987), Lehmann and Romano (2005), Korn, Troendle, McShane and Simon (2004), Romano and Shaikh (2006a, b) and Romano and Wolf (2005, 2007), Sarkar (2007, 2008b) and van der Laan, Dudoit and Pollard (2004)].

Sarkar (2007) has advocated using the expected ratio of k or more false rejections to the total number of rejections, the k-FDR, which is a less conservative notion of error rate than the k-FWER. Several procedures controlling the k-FDR have been developed under different dependence assumptions on the p-values. Sarkar (2007) has utilized the kth order joint distributions of the null p-values and developed procedures under the MTP_2 positive dependence [due to Karlin and Rinott (1980)] and arbitrary dependence conditions on the p-values. Sarkar and Guo (2009) considered a mixture model involving independent p-values and provided a simple and intuitive upper bound to the k-FDR through which they developed newer procedures controlling the k-FDR. Sarkar and Guo (2010) relaxed the requirement of using the kth order joint distributions of the null p-values and also the MTP_2 condition used in Sarkar (2007). They utilized only the bivariate distributions of the null p-values and developed different k-FDR procedures, assuming a positive dependence condition that is weaker than the MTP_2, the same one under which the procedure of Benjmaini and Hochberg (1995) controls the FDR [Benjamini and Yekutieli (2001) and

Sarkar (2002)], and also an arbitrary dependence condition on the p-values.

Suppose that $H_1, \dots, H_n$ are the null hypotheses that we want to simultaneously test using their respective p-values $p_1, \dots, p_n$. Let $p_{(1)} \leq \cdots \leq p_{(n)}$ be the ordered p-values and $H_{(1)}, \dots, H_{(n)}$ the associated null hypotheses. Then, given a non-decreasing set of critical constants $0 \leq \alpha_1 \leq \cdots \leq \alpha_n \leq 1$, a stepup multiple testing procedure rejects the set of null hypotheses $\{H_{(i)}, i \leq i^*_{SU}\}$ and accepts the rest, where $i^*_{SU} = \max\{i : p_{(i)} \leq \alpha_i\}$, if the maximum exists, otherwise accepts all the null hypotheses. A stepdown procedure, on the other hand, rejects the set of null hypotheses $\{H_{(i)}, i \leq i^*_{SD}\}$ and accepts the rest, where $i^*_{SD} = \max\{i : p_{(j)} \leq \alpha_j \ \forall\, j \leq i\}$, if the maximum exists, otherwise accepts all the null hypotheses. The critical constants are determined subject to the control at a pre-specified level α of a suitable error rate which, in this case, is the k-FDR defined as follows. Let R be the total number of rejected null hypotheses, among which V are falsely rejected and S are correctly rejected. Then, the k-FDR is defined as

$$k\text{-FDR} = \mathrm{E}(k\text{-FDP}), \quad \text{where } k\text{-FDP} = \frac{VI(V \geq k)}{R \vee 1}, \tag{1.1}$$

with $R \vee 1 = \max\{R, 1\}$, which reduces to the original FDR when $k = 1$.

Most of the procedures developed so far for controlling the k-FDR are stepup procedures, except a few developed in Sarkar and Guo (2010) that are stepdown procedures developed for independent as well as dependent p-values. In this article, we will focus mainly on developing some more stepdown procedures controlling the k-FDR under the independence as well as some forms of dependence conditions on the p-values.

1.2. Preliminaries

In this section, we will present two lemmas related to a general stepdown procedure which will be useful in developing stepdown procedures controlling the k-FDR in the next section.

Let n_0 and $n_1 (= n - n_0)$ be respectively the numbers of true and false null hypotheses. Define $\hat{q}_1, \cdots, \hat{q}_{n_0}$ and $\hat{r}_1, \cdots, \hat{r}_{n_1}$ to be the p-values corresponding to the true and false null hypotheses respectively and let $\hat{q}_{(1)} \leq \cdots \leq \hat{q}_{(n_0)}$ and $\hat{r}_{(1)} \leq \cdots \leq \hat{r}_{(n_1)}$ be their ordered values.

First, we have the following lemma.

Lemma 1.1. *Consider a stepdown procedure with critical constants* $0 \leq \alpha_1 \leq \cdots \leq \alpha_n \leq 1$. *Let* R *be the total number of rejections, of which* V *are* *false and S are correct. For a fixed $k > 0$, let J be the (random) largest index such that $\hat{r}_{(1)} \leq \alpha_{k+1}, \cdots, \hat{r}_{(J)} \leq \alpha_{k+J}$ (with $J = 0$ if $\hat{r}_{(1)} > \alpha_k$). Then, given $J = j$, $V \geq k$ implies that (i) $S \geq j$ and (ii) $\hat{q}_{(k)} \leq \alpha_{k+j}$.*

Proof. If $j = 0$, the lemma (i) obviously holds. If $R = n$, then the lemma (i) also holds. Now suppose $j > 0$ and $R < n$. Then, $p_{(R+1)} > \alpha_{R+1}$. Let us assume $R < k+j$, then $R-k+1 \leq j$. Thus $\hat{r}_{(R-k+1)} \leq \alpha_{R+1}$. Noting that $V \geq k$, then $\hat{q}_{(k)} \leq \alpha_R$. Therefore, $p_{(R+1)} \leq \max\{\hat{q}_{(k)}, \hat{r}_{(R-k+1)}\} \leq \alpha_{R+1}$. It leads to a contradiction. Thus $R \geq k+j$. Observe that $\hat{r}_{(j)} \leq \alpha_{k+j} \leq \alpha_R$, then $S \geq j$. Thus, the lemma (i) follows.

To prove the lemma (ii), we use reverse proof. Assume $\hat{q}_{(k)} > \alpha_{k+j}$. Noting that $\hat{r}_{(j)} \leq \alpha_{k+j}$ and $\hat{r}_{(j+1)} > \alpha_{k+j+1}$ when $j < n_1$, thus $R < k+j$ and then $V < k$. Therefore, if $V \geq k$, then $\hat{q}_{(k)} \leq \alpha_{k+j}$. The lemma (ii) is proved. □

The following second lemma is taken from Sarkar and Guo (2010).

Lemma 1.2. *Given a stepdown procedure with critical constants $0 \leq \alpha_1 \leq \cdots \leq \alpha_n \leq 1$, consider the corresponding stepdown procedure in terms of the null p-values $\hat{q}_1, \ldots, \hat{q}_{n_0}$ and the critical values $\alpha_{n_1+1} \leq \cdots \leq \alpha_n$. Let V_n denote the number of false rejections in the original stepdown procedure and $\hat{R}_{n_0}$ denote the number of rejections in the stepdown procedure involving the null p-values. Then, we have for any fixed $k \leq n_0 \leq n$,*

$$\{V_n \geq k\} \subseteq \left\{\hat{R}_{n_0} \geq k\right\}. \tag{1.2}$$

1.3. Stepdown k-FDR Procedures

We will develop some new stepdown procedures in this section that control the k-FDR. Before we do that, we want to emphasize a few important points.

First, while seeking to control k or more false rejections, we are tolerating at most $k-1$ of the null hypotheses to be falsely rejected. In other words, we can allow the first $k-1$ critical values to be arbitrarily chosen to be as high as possible, even all equal to one. However, it is not only counterintuitive to have a stepwise procedure with critical values that are not monotonically non-decreasing but also it is unrealistic to allow the first $k-1$ most significant null hypotheses to be rejected without having any control over them. So, the best option is to keep these critical values at the same level as the kth one; see also Lehmann and Romano (2005) and Sarkar (2007, 2008b). The stepdown procedures that we are going to develop next will have their first k critical values same. Second, the k-FDR procedures developed here are all generalized versions of some stepdown FDR procedures available in the literature. So, by developing these procedures we are actually providing some general results related to FDR methodologies. Third, although an FDR procedure also controls the k-FDR, the k-FDR procedures that we develop here are all more powerful than the corresponding FDR procedures.

Theorem 1.1. *Assume that the p-values satisfy the following condition:*

$$Pr\left\{\hat{q}_i \le u \middle| \hat{r}_1, \cdots, \hat{r}_{n_1}\right\} \le Pr\left\{\hat{q}_i \le u\right\} \le u, \quad u \in (0,1), \tag{1.3}$$

for any $i = 1, \cdots, n_0$. Then, the stepdown procedure with the critical constants

$$\alpha_i = \begin{cases} \frac{k\alpha}{n} & \text{if} \quad i = 1, \ldots, k \\ \min\left\{\frac{kn\alpha}{(n-i+k)^2}, 1\right\} & \text{if} \quad i = k+1, \ldots, n \end{cases} \tag{1.4}$$

controls the k-FDR at α.

Proof. When $n_0 < k$, there is nothing to prove, as in this case the k-$FDR = 0$ and hence trivially controlled. So, we will assume $k \le n_0 \le n$ while proving this theorem.

Using Lemma 1.1 and noting $V \le n_0$, we have

$$\begin{aligned} E\left(k\text{-}FDP \middle| J = j\right) &= E\left(\frac{V}{V+S} \cdot I(V \ge k) \middle| J = j\right) \\ &\le \frac{n_0}{n_0 + j} Pr\left(V \ge k \middle| J = j\right) \\ &\le \frac{n_0}{n_0 + j} Pr\left(\hat{q}_{(k)} \le \alpha_{j+k} \middle| J = j\right), \end{aligned} \tag{1.5}$$

for any fixed $j = 0, 1, \ldots, n_1$. Let N be the number of p-values corresponding to true null hypotheses that are less than or equal to constant α_{j+k}. Then, using Markov's inequality and condition (1.3), we have

$$\begin{aligned} Pr\left(\hat{q}_{(k)} \le \alpha_{j+k} \middle| J = j\right) &= Pr\left\{N \ge k \middle| J = j\right\} \le \frac{E(N|J=j)}{k} \\ &= \frac{1}{k}\sum_{i=1}^{n_0} Pr\left(\hat{q}_i \le \alpha_{j+k} \middle| J = j\right) \le \frac{n_0 \alpha_{j+k}}{k}. \end{aligned} \tag{1.6}$$

Thus, from (1.4), (1.5) and (1.6), we have

$$E\left(k\text{-}FDP \middle| J = j\right) \le \frac{n_0^2 n \alpha}{(n_0 + j)(n - j)^2}, \tag{1.7}$$

which is less than or equal to α, since $n_0 \le n - j$ and $(n_0 + j)(n - j) = n_0 n + j(n - n_0 - j) \ge n_0 n$. This proves the theorem. □

Remark 1.1. Note that the critical constants in (1.4) satisfy the following inequality:

$$\alpha_i \ge \alpha_i^* = \begin{cases} \frac{k\alpha}{n} & \text{if} \quad i = 1, \ldots, k \\ \frac{k\alpha}{n-i+k} & \text{if} \quad i = k+1, \ldots, n, \end{cases} \tag{1.8}$$

where α_i^*'s are the critical constants of the stepdown k-FWER procedure in Lehmann and Romano (2005). In other words, the k-FDR procedure in Theorem 1.1 is more powerful than the k-FWER procedure in Lehmann and Romano (2005), as one would expect, although the latter does not require any particular assumption on the dependence structure of the p-values.

Romano and Sheikh (2006b) gave a stepdown FDR procedure under the same condition as in (1.3). This procedure is generalized in Theorem 1.1 to a k-FDR procedure. Condition (1.3) is slightly weaker than the independence assumption between the sets of true and false p-values. No other assumptions are made here regarding the dependence structure within each of these sets. If, however, the null p-values are independent among themselves with each being distributed as $U(0, 1)$, this procedure can be improved to the one given in the following theorem.

Theorem 1.2. *Let*

$$G_{k,s}(u) = P\left\{U_{(k)} \le u\right\} = \sum_{j=k}^{s} \binom{s}{j} u^j (1-u)^{s-j}, \tag{1.9}$$

the cdf of the kth order statistic based on s iid $U(0, 1)$. The stepdown procedure with the following critical constants

$$\alpha_i = \begin{cases} G_{k,n}^{-1}(\alpha) & \text{if} \quad i = 1, \ldots, k \\ G_{k,n-i+k}^{-1}\left(\frac{n\alpha}{n-i+k}\right) & \text{if} \quad i = k+1, \ldots, n \end{cases} \tag{1.10}$$

controls the k-FDR at α if the $\hat{q}_i$'s are iid $U(0, 1)$ and independent of $(\hat{r}_1, \ldots, \hat{r}_{n_1})$.

Proof. For any fixed $j = 0, 1, \ldots, n_1$, we have from (1.5) and (1.10)

$$
\begin{aligned}
E\left(k\text{-}FDP \middle| J = j\right) &\le \frac{n_0}{n_0 + j} G_{k,n_0}(\alpha_{j+k}) \\
&= \frac{n_0}{n_0 + j} G_{k,n_0}\left(G^{-1}_{k,n-j}\left(\frac{n\alpha}{n-j}\right)\right) \\
&\le \frac{n_0}{n_0 + j} G_{k,n-j}\left(G^{-1}_{k,n-j}\left(\frac{n\alpha}{n-j}\right)\right) \\
&\le \frac{n_0 n\alpha}{(n_0 + j)(n - j)} \le \alpha. \qquad (1.11)
\end{aligned}
$$

The second inequality follows from the fact that $n_0 \le n - j$ and the cdf $G_{k,s}$ is increasing in s. This proves the theorem. □

Remark 1.2. Benjamini and Liu (1999) obtained a stepdown procedure assuming complete independence of all the p-values. We generalize this procedure in Theorem 1.2 from an FDR to a k-FDR procedure, but under a slightly weaker assumption allowing the false p-values to have an arbitrary dependence structure.

We now go back to Sarkar and Guo (2010) and generalize a stepdown k-FDR procedure given there assuming independence of the p-values. More specifically, we have the following theorem.

Theorem 1.3. *The stepdown procedure with critical constants $\alpha_1 = \cdots = \alpha_k \le \cdots \le \alpha_n$, where α_i/i is decreasing in i and*

$$\frac{i\alpha_k}{k} G_{k-1,i-1}(\alpha_{n-i+k}) \le \alpha \text{ for all } k \le i \le n, \qquad (1.12)$$

controls the k-FDR at α when the p-values are positively dependent in the sense that $E\left\{\phi(p_1, \ldots, p_n) \mid \hat{q}_i \le u\right\}$ is nondecreasing in u for every $\hat{q}_i$ and any nondecreasing (coordinatewise) function ϕ, and the $\hat{q}_i$'s are iid $U(0,1)$.

Proof.

$$
\begin{aligned}
k\text{-}FDR &= E\left\{\frac{V}{R}\cdot I\left(V\geq k\right)\right\} = E\left\{\sum_{r=k}^{n}\frac{1}{r}\sum_{i=1}^{n_0} I\left(\hat{q}_i\leq\alpha_r, V\geq k, R=r\right)\right\}\\
&= \sum_{i=1}^{n_0}\sum_{r=k}^{n}\frac{1}{r}P\left\{p_{(k)}\leq\alpha_k,\cdots,p_{(r)}\leq\alpha_r,p_{(r+1)}>\alpha_{r+1},\hat{q}_i\leq\alpha_r,V\geq k\right\}\\
&= \sum_{i=1}^{n_0}\frac{1}{k}P\left\{p_{(k)}\leq\alpha_k,\hat{q}_i\leq\alpha_k,V\geq k\right\}+\sum_{i=1}^{n_0}\sum_{r=k+1}^{n}E\left[P\left\{p_{(k)}\leq\alpha_k,\cdots,\right.\right.\\
&\qquad \left.\left. p_{(r)}\leq\alpha_r,V\geq k\middle|\hat{q}_i\right\}\left\{\frac{I(\hat{q}_i\leq\alpha_r)}{r}-\frac{I(\hat{q}_i\leq\alpha_{r-1})}{r-1}\right\}\right] \qquad (1.13)\\
&= \sum_{i=1}^{n_0}\frac{1}{k}P\left\{\hat{q}_i\leq\alpha_k,V\geq k\right\}+\sum_{i=1}^{n_0}\sum_{r=k+1}^{n}\frac{1}{r}P\left\{p_{(k)}\leq\alpha_k,\cdots,p_{(r)}\leq\alpha_r,\right.\\
&\qquad \left.\alpha_{r-1}<\hat{q}_i\leq\alpha_r,V\geq k\right\}-\sum_{i=1}^{n_0}\sum_{r=k+1}^{n}\frac{1}{r(r-1)}P\left\{p_{(k)}\leq\alpha_k,\cdots,\right.\\
&\qquad \left. p_{(r)}\leq\alpha_r,\hat{q}_i\leq\alpha_{r-1},V\geq k\right\}
\end{aligned}
$$

(with $p_{(n+1)}=1$ and $\alpha_{n+1}=1$).

Let $R_{n-1}^{(-i)}$ and $V_{n-1}^{(-i)}$ denote respectively the total numbers of rejections and false rejections in the stepdown procedure based on the ordered values $p_{(1)}^{(-i)}\leq\cdots\leq p_{(n-1)}^{(-i)}$ of $\{p_1,\ldots,p_n\}\setminus\{p_i\}$ and the $n-1$ critical values $\alpha_2\leq\cdots\leq\alpha_n$. Note that

$$
P\left\{\hat{q}_i\leq\alpha_k,V\geq k\right\}=P\left\{\hat{q}_i\leq\alpha_k,V_{n-1}^{(-i)}\geq k-1\right\}, \qquad (1.14)
$$

$$
\begin{aligned}
&P\left\{P_{(k)}\leq\alpha_k,\cdots,P_{(r)}\leq\alpha_r,\alpha_{r-1}<\hat{q}_i\leq\alpha_r,V\geq k\right\}\\
&= P\left\{P_{(k)}^{(-i)}\leq\alpha_k,\cdots,P_{(r-1)}^{(-i)}\leq\alpha_{r-1},\alpha_{r-1}<\hat{q}_i\leq\alpha_r,V_{n-1}^{(-i)}\geq k-1\right\}\\
&= P\left\{R_{n-1}^{(-i)}\geq r-1,\alpha_{r-1}<\hat{q}_i\leq\alpha_r,V_{n-1}^{(-i)}\geq k-1\right\}\\
&= P\left\{R_{n-1}^{(-i)}\geq r-1,V_{n-1}^{(-i)}\geq k-1 \mid \hat{q}_i\leq\alpha_r\right\}\alpha_r\\
&\quad - P\left\{R_{n-1}^{(-i)}\geq r-1,V_{n-1}^{(-i)}\geq k-1 \mid \hat{q}_i\leq\alpha_{r-1}\right\}\alpha_{r-1}\\
&\leq P\left\{R_{n-1}^{(-i)}\geq r-1,V_{n-1}^{(-i)}\geq k-1 \mid \hat{q}_i\leq\alpha_r\right\}\left(\alpha_r-\alpha_{r-1}\right), \qquad (1.15)
\end{aligned}
$$

with the last inequality following from the positive dependence assumption made in the theorem and the fact that the set

$\left\{R_{n-1}^{(-i)} \geq r-1, V_{n-1}^{(-i)} \geq k-1\right\}$ is a decreasing set in the p-values, and

$$\begin{aligned}
&P\left\{P_{(k)} \leq \alpha_k, \cdots, P_{(r)} \leq \alpha_r, \hat{q}_i \leq \alpha_{r-1}, V \geq k\right\} \\
&\geq P\left\{P_{(k)}^{(-i)} \leq \alpha_k, \cdots, P_{(r-1)}^{(-i)} \leq \alpha_{r-1}, \hat{q}_i \leq \alpha_{r-1}, V^{(-i)} \geq k-1\right\} \\
&= P\left\{R_{n-1}^{(-i)} \geq r-1, \hat{q}_i \leq \alpha_{r-1}, V_{n-1}^{(-i)} \geq k-1\right\} \\
&= P\left\{R_{n-1}^{(-i)} \geq r-1, V_{n-1}^{(-i)} \geq k-1 \mid \hat{q}_i \leq \alpha_{r-1}\right\} \alpha_{r-1}, \qquad (1.16)
\end{aligned}$$

for $k+1 \leq r \leq m$. Therefore, using (1.14)-(1.16) in (1.13), we have

$$\begin{aligned}
k\text{-}FDR &\leq \sum_{i=1}^{n_0} \frac{1}{k} P\left\{\hat{q}_i \leq \alpha_k, V_{n-1}^{(-i)} \geq k-1\right\} \qquad (1.17) \\
&+ \sum_{i=1}^{n_0} \sum_{r=k+1}^{n} P\left\{R_{n-1}^{(-i)} \geq r-1, V_{n-1}^{(-i)} \geq k-1 \mid \hat{q}_i \leq \alpha_r\right\}\left\{\frac{\alpha_r}{r} - \frac{\alpha_{r-1}}{r-1}\right\} \\
&\leq \sum_{i=1}^{n_0} \frac{1}{k} P\{\hat{q}_i \leq \alpha_k, V_{n-1}^{(-i)} \geq k-1\} \leq \sum_{i=1}^{n_0} \frac{\alpha_k}{k} P\{\hat{R}_{n_0-1}^{(-i)} \geq k-1\} \\
&\leq \sum_{i=1}^{n_0} \frac{\alpha_k}{k} P\{\hat{q}_{(k-1)}^{(-i)} \leq \alpha_{n_1+k}\} = \frac{n_0 \alpha_k}{k} G_{k-1,n_0-1}(\alpha_{n_1+k}),
\end{aligned}$$

which is controlled at level α if the α_i's are chosen subject to (1.12). The third inequality in (1.17) follows from Lemma 1.2. Thus the theorem is proved. □

Remark 1.3. A variety of stepdown procedures can be obtained using critical values satisfying the conditions in Theorem 1.3 once the distributional assumptions in the theorem hold. For instance, one may choose the critical constants $\alpha_i = (i \vee k)\beta/n$ with the β determined subject to

$$\frac{\beta}{n} \max_{k \leq n_0 \leq n} \left\{ n_0 G_{k-1,n_0-1}\left(\frac{(n-n_0+k)\beta}{n}\right)\right\} = \alpha. \qquad (1.18)$$

This is what Sarkar and Guo (2010) proposed under the independence of the p-values. Similarly, we can consider the critical constants $\alpha_i = \left(\frac{i}{n}\right)^\gamma \beta$ or $\alpha_i = \frac{i+d}{n+d}\beta$, for some pre-specified constants $0 < \gamma < 1$ and $d > 0$, with the β chosen as large as possible subject to Condition (1.12).

Recently, Gavrilov, Benjamini and Sarkar (2009) obtained a stepdown procedure controlling the FDR with independent p-values. We now derive a generalized version of this procedure providing a control of the k-FDR in the following theorem.

Theorem 1.4. *The step-down procedure with the critical values* $\alpha_1 = \cdots = \alpha_k \leq \cdots \leq \alpha_n$ *satisfying* $\alpha_i/(1-\alpha_i) \leq i\beta/(n-i+1), i = 1, \ldots, n$, *controls the k-FDR at level* α *for any fixed* β *satisfying*

$$\beta G_{k-1,i}(\alpha_{n+k-i-1}) \leq \alpha \textit{ for all } k \leq i \leq n, \tag{1.19}$$

if the p-values are independent and $\hat{q}_i \sim U(0,1)$.

Proof. From the proof of Theorem 1.3, we notice that for independent p-values

$$\begin{aligned}
k\text{-}FDR &\leq \sum_{i=1}^{n_0} \frac{\alpha_k}{k} P\left\{V_{n-1}^{(-i)} \geq k-1\right\} + \sum_{i=1}^{n_0} \sum_{r=k+1}^{n} P\left\{P_{(k)}^{(-i)} \leq \alpha_k, \cdots, \right. \\
&\quad \left. P_{(r-1)}^{(-i)} \leq \alpha_{r-1}, V_{n-1}^{(-i)} \geq k-1\right\} \left\{\frac{\alpha_r}{r} - \frac{\alpha_{r-1}}{r-1}\right\} \\
&= \sum_{i=1}^{n_0} \sum_{r=k}^{n} \frac{\alpha_r}{r} P\left\{P_{(k)}^{(-i)} \leq \alpha_k, \cdots, P_{(r-1)}^{(-i)} \leq \alpha_{r-1}, P_{(r)}^{(-i)} > \alpha_r, \right. \\
&\quad \left. V_{n-1}^{(-i)} \geq k-1\right\} \\
&\leq \beta \sum_{i=1}^{n_0} \sum_{r=k}^{n} \frac{1-\alpha_r}{n-r+1} P\left\{P_{(k)}^{(-i)} \leq \alpha_k, \cdots, P_{(r-1)}^{(-i)} \leq \alpha_{r-1}, P_{(r)}^{(-i)} > \alpha_r, \right. \\
&\quad \left. V_{n-1}^{(-i)} \geq k-1\right\} \\
&= \beta \sum_{i=1}^{n_0} \sum_{r=k}^{n} \frac{1}{n-r+1} P\left\{P_{(k)}^{(-i)} \leq \alpha_k, \cdots, P_{(r-1)}^{(-i)} \leq \alpha_{r-1}, P_{(r)}^{(-i)} > \alpha_r, \right. \\
&\quad \left. \hat{q}_i > \alpha_r, V_{n-1}^{(-i)} \geq k-1\right\} \\
&= \beta \sum_{i=1}^{n_0} \sum_{r=k}^{n} \frac{1}{n-r+1} P\left\{P_{(k)} \leq \alpha_k, \cdots, P_{(r-1)} \leq \alpha_{r-1}, P_{(r)} > \alpha_r, \right. \\
&\quad \left. \hat{q}_i > \alpha_r, V \geq k-1\right\}.
\end{aligned} \tag{1.20}$$

With A and U denoting, respectively, the total numbers of accepted and correctly accepted null hypotheses, we note that the last expression in (1.3) is $\beta E\{UI(V \geq k-1)/A \vee 1\}$, which is less than or equal to

$$\begin{aligned}
\beta P\{V \geq k-1\} &\leq \beta P\{\hat{R}_{n_0} \geq k-1\} \quad \text{(from Lemma 1.2)} \\
&\leq \beta P\{\hat{q}_{(k-1)} \leq \alpha_{n_1+k-1}\} \\
&= \beta G_{k-1,n_0}(\alpha_{n_1+k-1})
\end{aligned}$$

This is controlled at α for any β satisfying (1.19). Thus, the theorem is proved. $\square$

Remark 1.4. Specifically, if we consider the critical constants $\alpha_i = i\beta/(n-i+1+i\beta)$, $i = k, \cdots, n$, as in Gavrilov, Benjamini and Sarkar (2009), where the constant β is determined subject to

$$\max_{k \le n_0 \le n} \left\{ \beta G_{k-1,n_0} \left(\frac{\beta(n - n_0 + k - 1)}{n_0 - k + 2 + \beta(n - n_0 + k - 1)} \right) \right\} = \alpha, \qquad (1.21)$$

the corresponding stepdown procedure will provide a control of the k-FDR at level α.

1.4. Numerical studies

In this section, we present the results of a numerical study comparing the four different stepdown k-FDR procedures developed in Theorems 1.1-1.4 in terms of their critical values to gain an insight into their relative performance with respect to the number of discoveries. Let us denote the four sets of critical constants as $\alpha_i^{(j)}$, $i = k, \cdots, n$, with j referring to the jth procedure. For the first two procedures, the critical constants are defined in (1.4) and (1.10), respectively, and for the last two, the critical constants are $\alpha_i^{(3)} = (i \vee k)\beta/n$ and $\alpha_i^{(4)} = i\beta/(n - i + 1 + i\beta)$, where β is determined subject to (1.18) and (1.21), respectively. As the baseline method for comparison, we choose the stepdown procedure with the critical constants $\gamma_i = i\alpha/n, i = 1, \cdots, n$. This is the stepdown analog of the Benjamini-Hochberg (BH) stepup FDR procedure that, as Sarkar (2002) has shown, controls the FDR and hence the k-FDR under the conditions considered in Theorems 1.2-1.4, and would have been used by researchers without the knowledge of other stepdown k-FDR procedures. Considering $\zeta_i^{(j)} = \log_{10}\left(\alpha_i^{(j)}/\gamma_i\right), i = 1, \cdots, n$, for $j = 1, \cdots, 4$, we plot in Figure 1.1 the four sequences $\zeta_i^{(j)}, j = 1, \cdots, 4$, with $n = 500$, $k = 8$, and $\alpha = 0.05$.

As seen from Figure 1.1, the critical constants of the procedure in Theorem 1.1 (labeled RS) are all much less than those of the stepdown analog of the BH procedure (referred to as the stepdown BH in this article). For the procedure in Theorem 1.2 (labeled BL), the first few of its critical values are seen to be larger than those of the stepdown BH (labeled BH). The critical constants of the procedures in Theorems 1.3 and 1.4 (labeled SG and GBS respectively) are all uniformly larger than the corresponding critical values of the BH. Thus, there is a numerical evidence that the procedures in Theorems 1.3 and 1.4 are both more powerful than the stepdown BH, but the procedure in Theorem 1.1 is not. Since for a stepdown procedure, the power is mostly determined by some of its first critical values, the procedure in Theorem 1.2 may sometimes be more powerful than the stepdown BH procedure.

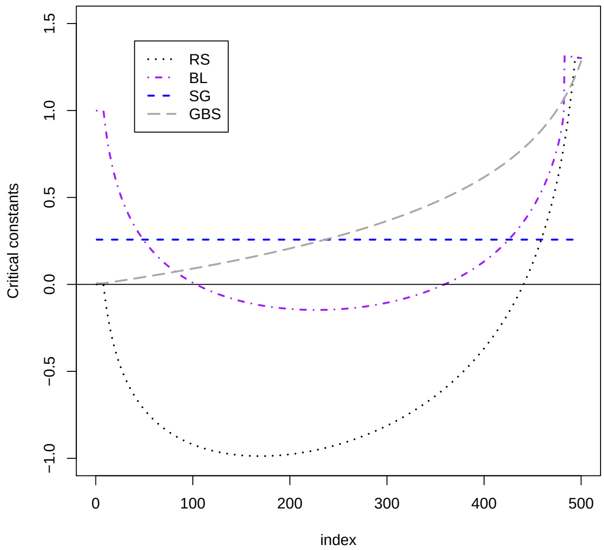

Fig. 1.1. The logarithms with base 10 of ratios of critical constants of four stepdown k-FDR procedures with respect to that of the BH procedure for $n = 500, k = 8,$ and $\alpha = 0.05$.

We also compared the four stepdown k-FDR procedures with the stepdown BH procedure in terms of their power. We simulated the average power, the expected proportion of false null hypotheses that are rejected, for each of these procedures. Figure 1.2 presents this power comparison. Each simulated power was obtained by (i) generating $n = 200$ independent normal random variables $N(\mu_i, 1),\ i = 1, \cdots, n$ with n_1 of the 200 μ_i's being equal to $d = 2$ and the rest 0, (ii) applying the stepdown BH procedure and the four stepdown k-FDR procedures with $k = 4$ to the generated data to test $H_i : \mu_i = 0$ against $K_i : \mu_i > 0$ simultaneously for $i = 1, \ldots, 200$ at $\alpha = 0.05$, and (iii) repeating steps (i) and (ii) 1,000 times before observing the proportion of the n_1 false H_i's that are correctly declared significant.

As seen from Figure 1.2, the SG procedure in Theorem 1.3 is uniformly more powerful than the stepdown BH, with the power difference getting

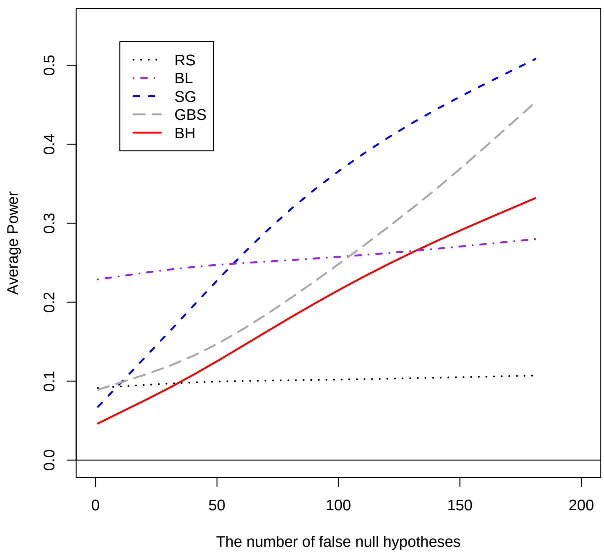

Fig. 1.2. Power of four stepdown k-FDR procedures in the case of independence with $n = 200, k = 4, d = 2$ and $\alpha = 0.05$.

significantly higher with increasing number of false null hypotheses, while the GBS procedure in Theorem 1.4 is marginally more powerful than the stepdown BH, with the power difference getting significantly higher only after the number of false null hypotheses becomes moderately large. The BL procedure in Theorem 1.2 is the most powerful among these four stepdown procedures when the proportion of false null hypotheses is small. Even when the false proportion is moderately large, this is also more powerful than the stepdown BH. However, it loses its advantage over the stepdown BH when the proportion of false null hypotheses is very large. Finally, the RS procedure in Theorem 1.1 is less powerful than the stepdown BH, as we expected from Figure 1.1 showing the numerical comparisons of the critical constants of these procedures.

1.5. An application to gene expression data

Hereditary breast cancer is known to be associated with mutations in BRCA1 and BRCA2 proteins. Hedenfalk et al. (2001) report that a group

Table 1.1. Numbers of differentially expressed genes for the data in Hedenfalk et al. (2001) using four stepdown k-FDR procedures.

procedure	level α	$k = 2$	5	8	10	15	20	30
RS	0.03	2	5	8	8	12	18	22
	0.05	3	8	11	16	20	24	32
	0.07	5	11	18	19	24	29	47
BL	0.03	8	34	73	82	120	150	191
	0.05	11	42	75	86	125	157	200
	0.07	11	47	76	91	129	159	203
SG	0.03	8	8	11	20	21	24	73
	0.05	74	75	76	76	76	82	94
	0.07	103	110	123	124	129	131	139
GBS	0.03	3	5	8	8	12	18	22
	0.05	73	73	73	73	73	73	73
	0.07	96	96	96	96	96	96	96

of genes are differentially expressed between tumors with BRCA1 mutations and tumors with BRCA2 mutations. The data, which are publicly available from the web site http://research.nhgri.nih.gov/microarray/NEJM_Supplement/, consist of 22 breast cancer samples, among which 7 are BRCA1 mutants, 8 are BRCA2 mutants, and 7 are sporadic (not used in this illustration). Expression levels in terms of florescent intensity ratios of a tumor sample to a common reference sample, are measured for $3,226$ genes using cDNA microarrays. If any gene has one ratio exceeding 20, then this gene is eliminated. Such preprocessing leaves $n = 3,170$ genes.

We tested each gene for differential expression between these two tumor types by using a two-sample t-test statistic. For each gene, the base 2 logarithmic transformation of the ratio was obtained before computing the two-sample t-test statistic based on the transformed data. A permutation method from Storey and Tibshirani (2003) with the permutation number $B = 2,000$ was then used to calculate the corresponding raw p-value. Finally, we applied to these raw p-values the stepdown BH and the four stepdown k-FDR procedures in Theorems 1.1-1.4.

At $\alpha = 0.03$, 0.05 and 0.07, the stepdown BH results in 3, 33 and 95 significant genes respectively, while those numbers for the present methods are presented in Table 1.1 for $k = 2, 5, 8, 10, 15, 20$ and 30, with the four procedures in Theorems 1.1-1.4 labeled RS, BL, SG and GBS respectively. As we can see from this table, the RS procedure in Theorem 1.1 generally detects less significant genes than the stepdown BH for moderate or large values of α. The BL procedure in Theorem 1.2 always detects more differentially expressed genes than the stepdown BH for slightly moderate values of k and small or moderate values of α. The SG procedure in Theorem 1.3 is seen to always detect more significant genes than the stepdown BH, while the GBS procedure in Theorem 1.4 detects almost the same number of differentially expressed genes as the stepdown BH except for moderate α.

1.6. Conclusions

We have presented a number of new stepdown k-FDR procedures in this article under different assumptions on the dependence structure of the p-values, generalizing some existing stepdown FDR procedures. These would be of use in situations where one is willing to tolerate at most $k-1$ false rejections and is looking for a stepdown procedure controlling a powerful notion of error rate than the k-FWER for exercising a control over at least k false rejections. Although any FDR stepdown procedure can also control the k-FDR, ours are powerful than the corresponding FDR versions. Moreover, we offer better k-FDR stepdown procedures than the stepdwon analog of the BH stepup procedure, with its first $k-1$ critical values equal to the kth one, which would have been commonly used by researchers without knowing the existence of any other stepdown k-FDR procedure.

References

1. Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing, *J. Roy. Statist. Soc. Ser. B* **57**, pp. 289–300.
2. Benjamini, Y., Krieger, A. M. and Yekutieli, D. (2006). Adaptive linear step-up false discovery rate controlling procedures, *Biometrika* **93**, pp. 491–507.
3. Benjamini, Y. and Liu, W. (1999). A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence, *Journal of Statistical Planning and Inference* **82**, pp. 163–170.
4. Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency, *Ann. Statist.* **29**, pp. 1165–1188.
5. Benjamini, Y. and Yekutieli, D. (2005). False discovery rate-adjusted multiple confidence intervals for selected parameters, *J Amer. Statist. Assoc.* **100**, pp. 71–93.
6. 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, *Statist. App. Gen. Mol. Bio.* **3**, 1, Article 13.
7. Finner, H., Dickhaus, T. and Roters, M. (2007). Dependency and false discovery rate: Asymptotics, *Ann. Statist.* **35**, pp. 1432–1455.
8. Finner, H., Dickhaus, T. and Roters, M. (2009). On the false discovery rate and an asymptotically optimal rejection curve, *Ann. Statist.* **37**, pp. 596–618.
9. Gavrilov, Y., Benjamini, Y. and Sarkar, S. K. (2009). An adaptive step-down procedures with proven FDR control under independence, *Ann. Statist.* **37**, pp. 619–629.
10. Genovese, C. and Wasserman, L. (2002). Operarting characteristics and extensions of the false discovery rate procedure, *J. Roy. Statist. Soc. Ser. B* **64**, pp. 499–517.
11. Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control, *Ann. Statist.* **32**, pp. 1035–1061.
12. Guo, W. and Rao, M. B. (2010). On stepwise control of the generalized familywise error rates, *Electronic Journal of Statistics* **4**, pp. 472-485.
13. Guo, W. and Romano, J. P. (2007). A generalized Sidak-Holm procedure and control of generalized error rates under independence, *Statist. App. Gen. Mol. Bio.* **6**, 1, Article 3.
14. Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Kallioniemi, OP, Wilfond, B., Borg, A. and Trent, J. (2001). Gene-expression profiles in hereditary breast cancer, *New Eng. J. Med.* **344**, pp. 539–548.
15. Hochberg, Y. and Tamhane, A. C. (1987). *Multiple Comparison Procedures*, Wiley, NY.
16. Hommel, G. and Hoffmann, T. (1987). Controlled uncertainty. In *Multiple Hypothesis Testing* P. Bauer, G. Hommel and E. Sonnemann, eds. pp. 154–161, Springer, Heidelberg.
17. Karlin, S. and Rinott, Y. (1980). Classes of orderings of measures and related correlation inequalities I: Multivariate totally positive distributions, *J. Mult. Anal.* **10**, pp. 467–498.
18. Korn, E., Troendle, T., McShane, L. and Simon, R. (2004). Controlling the number of false discoveries: Application to high-dimensional genomic data, *J. Statist. Plann. Inf.* **124**, pp. 279–398.
19. Lehmann, E. L. and Romano, J. P. (2005). Generalizations of the familywise error rate, *Ann. Statist.* **33**, pp. 1138–1154.
20. Romano, J. P. and Shaikh, A. M. (2006a). Stepup procedures for control of generalizations of the familywise error rate, *Ann. Statist.* **34**, pp. 1850–1873.
21. Romano, J. P. and Shaikh, A. M. (2006b). On stepdown control of the false discovery proportion. In *The Second E.L. Lehamann Symposium - Optimality* J. Rojo, eds. pp. 40–61, IMS Lecture Notes - Monograph Series **49**.
22. Romano, J. P. and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping, *Econometrica* **73**, pp. 1237–1282.
23. Romano, J. P. and Wolf, M. (2007). Control of generalized error rates in multiple testing, *Ann. Statist.* **35**, pp. 1378–1408.
24. Sarkar, S. K. (2002). Some results on false discovery rate in stepwise multiple

testing procedures, *Ann. Statist.* **30**, pp. 239–257.
25. Sarkar, S. K. (2004). FDR-controlling stepwise procedures and their false negatives rates, *J. Statist. Plann. Inf.* **125**, pp. 119–137.
26. Sarkar, S. K. (2006). False discovery and false non-discovery rates in single-step multiple testing procedures, *Ann. Statist.* **34**, pp. 394–415.
27. Sarkar, S. K. (2007). Stepup procedures controlling generalized FWER and generalized FDR, *Ann. Statist.* **35**, pp. 2405–2420.
28. Sarkar, S. K. (2008a). Two-stage stepup procedures controlling FDR, *J. Statist. Plann, Inf.* **138**, pp. 1072–1084.
29. Sarkar, S. K. (2008b). Generalizing Simes' test and Hochberg's stepup procedure, *Ann. Statist.* **36**, pp. 337–363.
30. Sarkar, S. K. and Guo, W. (2009). On a generalized false discovery rate, *Ann. Statist.*, **37**, pp. 337–363.
31. Sarkar, S. K. and Guo, W. (2010). Procedures controlling generalized false discovery rate, *Statist. Sinica*, **20**, pp. 1227–1238.
32. Storey, J. D. (2002). A direct approach to false discovery rates, *J. Roy. Statist. Soc. Ser. B* **64**, pp. 479–498.
33. Storey, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value, *Ann. Statist.* **31**, pp. 2013–2035.
34. Storey, J. D., Taylor, J. E. and Siegmund, D.(2004). Strong control, conservative point estimation and simultanaeous conservative consistency of false discovery rates: A unified approach, *J. Roy. Statist. Soc. Ser. B* **66**, pp. 187–205.
35. Storey, J. D. and Tibshirani, R. (2003). Statistical significance for genomewide studies, *Proc. Nat. Acad. Sci.* **100**, pp. 9440–9445.
36. van der Laan, M., Dodoit, 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, *Stat. App. Gen. Mol. Bio.* **3**, 1, Article 15.