Applying the Benjamini–Hochberg procedure to a set of generalized $p$-values

Fredrik Jonsson
Applying the Benjamini–Hochberg procedure to a set of generalized $p$-values

Fredrik Jonsson, Uppsala University

Abstract

Let $\delta_1, \ldots, \delta_n$ be decision functions with associated loss functions $L_1, \ldots, L_n$, with respect to some statistical model $\{P_\theta : \theta \in \Theta\}$. We assume non-negative random variables $\hat{R}_1, \ldots, \hat{R}_n$ (referred to as generalized $p$-values), satisfying

$$E_\theta(L_k(\delta_k, \theta)I\{\hat{R}_k \leq \alpha\}) \leq \alpha,$$

for $\alpha > 0$, $\theta \in \Theta$ and $k = 1, \ldots, n$.

It is shown, assuming independence between the pairs $(\delta_1, \hat{R}_1), \ldots, (\delta_1, \hat{R}_1)$, that the Benjamini–Hochberg procedure [1], determining the number of confirmed decisions through $N = \max\{i : \hat{R}(i) \leq i\alpha/n\}$, with respect to an ordering $\hat{R}(1) \leq \cdots \leq \hat{R}(n)$, controls the expected average loss among confirmed decisions at level $\alpha$. We also verify, without any assumption on dependence, that the expected average loss among confirmed decisions is bounded by $\alpha \sum_{k=1}^{n} 1/k$, thus extending the results of Benjamini and Yekutieli [5] and Hommel [13]. The generalized perspective on $p$-values is motivated with reference to the treatment of directional conclusions in [14] and [15]. Finally, we demonstrate in a simulation study that several related procedures (e.g. the Holm procedure [12]) can not be extended in a similar manner.

1 Introduction

The well-known Benjamini–Hochberg procedure [1] (also known as Simes’ procedure [21]) assumes $p$-values $\hat{p}_1, \ldots, \hat{p}_n$ and a significance level $\alpha$. The procedure then rejects hypotheses corresponding to the $N$ smallest values in an ordering $\hat{p}(1) \leq \cdots \leq \hat{p}(n)$, with $0 \leq N \leq n$ determined by $N = \max\{i : \hat{p}(i) \leq i\alpha/n\}$. Simes’ proved [21], under the assumption that $\hat{p}_1, \ldots, \hat{p}_n$ are independent random variables, that the probability of at least one rejection is bounded by $\alpha$ in the intersection of the null hypotheses. Simes’ result was later extended by Benjamini and Hochberg [1], proving that the procedure generally — under the assumption of independence — controls the false discovery rate (FDR, the expected proportion of false rejections among the total number of rejections) at level $\alpha$.

We consider a generalization of the logic of $p$-values in Section 2 below. Thus, starting with a loss function $L$ (corresponding to some decision function $\delta$), we assume


Keywords and phrases. Multiple testing; false discovery rate; Simes’ procedure; significance testing; decision theory.

Date. November 15, 2011
that its expected loss can be controlled at level \( \alpha \) through an indicator random variable \( I \), in the sense that

\[
E_\theta (L(\delta, \theta) I) \leq \alpha, \quad \text{for all } \theta \in \Theta.
\]

A loss function \( L = I\{\theta \in \Theta_0\} \) here corresponds to level-\( \alpha \) testing of \( \Theta_0 \), with a confirmation \( I = 1 \) referred to as a rejection of \( \Theta_0 \). Following this path of generalization, we refer to a non-negative random variable \( \hat{R} \) as a generalized \( p \)-value, with respect to \( L \) and \( \delta \), in case the following condition is fulfilled

\[
E_\theta (L(\delta, \theta) I\{\hat{R} \leq \alpha\}) \leq \alpha, \quad \text{for all } \alpha > 0 \text{ and } \theta \in \Theta.
\]

In the setting of loss functions \( L_1, \ldots, L_n \) and confirmative indicators \( I_1, \ldots, I_n \), note that the false discovery proportion (the proportion of false rejections among the total number of rejections), naturally generalizes into the following average loss among confirmed decisions,

\[
L = \frac{L_1 I_1 + \cdots + L_n I_n}{I_1 + \cdots + I_n},
\]

(1)

(interpreted as zero if \( I_1 = \cdots = I_n = 0 \), that is, if no decisions are confirmed). We show (Section 3) that the Benjamini–Hochberg procedure applied to a set of generalized \( p \)-values \( \hat{R}_1, \ldots, \hat{R}_n \) controls the expected average loss among confirmed decisions, in the sense that

\[
E_\theta (L) \leq \alpha m/n, \quad \text{for all } \theta \in \Theta,
\]

assuming pairwise independence for \( (\delta_1, \hat{R}_1), \ldots, (\delta_n, \hat{R}_n) \), among which \( n - m \) decisions are “free of risk”.

We were motivated to investigate the generalized setting in view of the proposal by Jones and Tukey in [14] (see also [15]). These papers consider a slightly different perspective compared to the conventional “one-sided” and “two-sided” significance testing. More precisely, with three-decision procedures for directional conclusions, control was imposed on “reversal rates” instead of probabilities of false rejections. A connection between the present context of generalized \( p \)-values and three-decision procedures for directional conclusions is presented in Section 2.1.

Interestingly, many procedures in the tradition of the Benjamini–Hochberg procedure do not have similar extensions. For instance, with \( L_1, \ldots, L_n \) being zero-one loss functions (reporting the errors of the decision functions \( \delta_1, \ldots, \delta_n \)), one could guess that the well-known Holm procedure [12] applied to \( \hat{R}_1, \ldots, \hat{R}_n \) would control the probability of at least one error among confirmed decisions (so-called FWER, familywise error rate). This is however not the case, as demonstrated in Section 4. Nevertheless, the fundamental Bonferroni argument is valid in this context (cf. Section 2.2). We also consider more recent “adaptive” FDR-controlling procedures in Section 4, showing by numerical results that their control of expected average loss among confirmed decisions does not extend to the generalized setting.

### 1.1 General dependence

Several authors have considered the conservativeness of the Benjamini–Hochberg procedure with respect to control of FDR in cases of dependent \( p \)-values \( \hat{p}_1, \ldots, \hat{p}_n \). Simes
[21] included a simulation study concerning correlated test statistics and the intersection of the null hypotheses (where the notions of FDR and FWER coincide). Later, Samuel-Cahn [18] and Hochberg and Rom [11] gave examples with negatively dependent test statistics where conservativeness is violated (albeit only to a small extent) in the intersection of the null hypotheses. Hommel [13] proved (work which was later extended by Falk [6]) that the probability of a false rejection is bounded by

$$\alpha^* = \min \left\{ 1, \alpha \sum_{k=1}^{n} k^{-1} \right\}.$$  \hspace{1cm} (2)

Moreover, examples demonstrating the sharpness of the bound were given. Sarkar [19] proved that Simes' original result on conservativeness for independent test statistics can be extended to a class of positively dependent test statistics.

Regarding the complement of the intersection of the null hypotheses, Benjamini and Yekutieli [5] generalized Hommel’s bound (2), proving that

$$\text{FDR} \leq \frac{m}{n} \sum_{k=1}^{n} k^{-1},$$  \hspace{1cm} (3)

with $m$ referring to the number of true null hypotheses. Also, conservativeness was extended in [5] to a class of positively dependent test statistics (with generalizations by Sarkar [20] and Finner, Dickhaus and Roters [8, Section 4]). With an asymptotic perspective (regarding the number of null hypotheses), Finner, Dickhaus and Roters [7] investigated, theoretically and numerically, the impact of dependence on FDR for some commonly encountered test statistics.

In the following (Section 3) we generalize the bounds of Hommel (2) and Benjamini–Yekutieli (3) by showing that, without any assumption of dependence,

$$\mathbb{E}_{\theta}(\mathcal{L}) \leq \frac{m}{n} \sum_{k=1}^{n} k^{-1},$$

with $\mathcal{L}$ referring to the average loss among confirmed decisions (as in (1)), and $n - m$ referring to the number of risk-free decisions.

## 2 Risk-controlled decisions

Following the basic set-up in statistical decision theory, consider a random element $X$ and a set of probability distributions $\{P_{\theta} : \theta \in \Theta\}$, where each element is a possible law of $X$ (cf. for instance [16, Section 1.1] or [17, Chapter 1]). Also, consider a non-negative loss function $L$ with respect to some decision function $\delta$ and the parameter space $\Theta$ (cf. [17, Chapter 3]).

In this setting, assume that an indicator random variable $I$ (a zero-one function of $X$) is given, with the purpose of confirming $\delta$ at maximal risk $\alpha$. Thus, $\delta$ is confirmed whenever $I = 1$. Moreover, it is assumed that the following condition is fulfilled:

$$\mathbb{E}_{\theta}(L(\delta, \theta)I) \leq \alpha, \quad \text{for all} \quad \theta \in \Theta.$$  \hspace{1cm} (4)

Note that condition (4) generalizes the concept of a statistical test. Indeed, given some null hypothesis $\Theta_0 \subset \Theta$, and the following zero-one loss function,

$$L = I\{\theta \in \Theta_0\},$$  \hspace{1cm} (5)
condition (4) characterizes that the indicator $I$ serves as a level-$\alpha$ test of $\Theta_0$ (with $I = 1$ referred to as a rejection of $\Theta_0$). Thus, condition (4) generalizes statistical testing in two senses: First, losses are allowed to assume other values than 0 and 1. Second, losses are allowed to depend on some decision function $\delta$, i.e. on the random data $X$.

Next, generalizing the treatment of $p$-values in [16, Section 3.3], assume that some family $\{I(\alpha)\}_{0 \leq \alpha < \infty}$ of indicator random variables is given. Let the family be stochastically non-decreasing, in the sense that $I(\alpha) = 1$ implies $I(\alpha') = 1$ for $\alpha < \alpha'$, and assume that

$$E_\theta(L(\delta, \theta) I(\alpha)) \leq \alpha, \quad \text{for all } \theta \in \Theta \text{ and } 0 \leq \alpha < \infty. \quad (6)$$

Then, by defining a random variable

$$\hat{R} = \inf \{\alpha : I(\alpha) = 1\},$$

condition (6) transfers into the following condition (cf. [16, Lemma 3.3.1]):

$$E_\theta(II(\hat{R} \leq \alpha)) \leq \alpha, \quad \text{for all } \theta \in \Theta \text{ and } 0 \leq \alpha < \infty. \quad (7)$$

For brevity, we refer to $\hat{R}$ in (7) as a generalized $p$-value with respect to $L$ and $\delta$.

### 2.1 Directional conclusions

Following the treatment of three-decision procedures for directional conclusions in [14] and [15], let a partitioning $\{\Theta_1, \Theta_2\}$ of the parameter space $\Theta$ be given. For instance, $\Theta_1$ and $\Theta_2$ may refer to the two possible “directions” of a treatment effect. Then, being interested in rational conclusions of whether $\theta \in \Theta_1$ or $\theta \in \Theta_2$ holds, consider decision functions $\delta$ into the binary decision space $\{1, 2\}$. A natural loss function (distinguishing between true and false conclusions) is then given by

$$L(\delta, \theta) = I(\theta \in \Theta_1)I(\delta = 2) + I(\theta \in \Theta_2)I(\delta = 1). \quad (8)$$

Introducing an indicator random variable $I$ fulfilling condition (4) in this context, we obtain a three-decision procedure by replacing $\delta$ with an indefinite decision in case $I = 0$. Moreover, the reversal rate is then controlled at level $\alpha$ (in the terminology of [14]).

In fact, there is a close correspondence between three-decision procedures and “non-colliding” simultaneous tests of $\Theta_1$ and $\Theta_2$ in this context (cf. [15]). Thus, if $\hat{p}_1$ is a $p$-value for testing $\Theta_1$ and $\hat{p}_2$ is a $p$-value for testing $\Theta_2$, consider the directional decision function

$$\delta = 2 \cdot I(\hat{p}_1 \leq \hat{p}_2) + 1 \cdot I(\hat{p}_1 > \hat{p}_2).$$

In other words, $\delta$ concludes $\theta \in \Theta_1$ if it appears to be more profitable to reject $\Theta_2$, and vice versa. Moreover, consider

$$\hat{R} = \min \{\hat{p}_1, \hat{p}_2\}.$$

Then, with $L$ given by (8),

$$E_\theta(II(\hat{R} \leq \alpha)) = E_\theta(I(\hat{R} \leq \alpha)(I(\theta \in \Theta_1)I(\hat{p}_1 \leq \hat{p}_2) + I(\theta \in \Theta_2)I(\hat{p}_1 > \hat{p}_2)))$$

$$= I(\theta \in \Theta_1)P_\theta(\hat{p}_1 \leq \alpha) + I(\theta \in \Theta_2)P_\theta(\hat{p}_2 \leq \alpha) \leq \alpha,$$
which proves that \( \hat{R} \) is an generalized \( p \)-value with respect to \( L \) and \( \delta \).

As a final remark in this context, comparing the two loss functions in (5) and (8), note that both are zero-one loss functions. However, \( L \) in (8) depends on given data (through the directional decision function \( \delta \)).

2.2 Zero-one losses and the Bonferroni argument

Now, let loss functions \( L_1, \ldots, L_n \) be given, with respect to one and the same statistical model \( \{P_\theta : \theta \in \Theta\} \). Moreover, assume that corresponding generalized \( p \)-values \( \hat{R}_1, \ldots, \hat{R}_n \) are given. Then, applying the indicator random variables \( I_k = I\{R_k \leq \alpha\} \) simultaneously in order to confirm subsets of \( \delta_1, \ldots, \delta_n \), with respect to some level of significance \( \alpha \), may lead to some undesirable consequences.

For instance, assuming that \( L_1, \ldots, L_n \) are zero-one loss functions (reporting the errors of \( \delta_1, \ldots, \delta_n \) respectively), with \( \theta \) given such that each pair \( (L_k(\delta_k, \theta), \hat{R}_k) \) is independent of any other pair \( (L_{k'}(\delta_{k'}, \theta), \hat{R}_{k'}) \), and \( \alpha \) given such that the controlled risks satisfy

\[
E_\theta(L_1 I_1) = \cdots = E_\theta(L_n I_n) = \alpha,
\]
then, the probability of no error among confirmed decisions is given by,

\[
P_\theta\left(\bigcap_{k=1}^n \{ \min (L_k, I_k) = 0 \}\right) = \prod_{k=1}^n \left(1 - E_\theta(L_k I_k)\right) = (1 - \alpha)^n,
\]
which is considerably smaller than \( 1 - \alpha \) if \( n \) is large.

As an alternative, consider replacing \( I_k \) by the more restrictive “Bonferroni-indicators” \( I'_k = I\{R_k \leq \alpha/n\} \). Then, once more assuming zero-one loss functions \( L_1, \ldots, L_n \), the probabilities of at least one error among confirmed decisions satisfy the following bound:

\[
P_\theta\left(\bigcup_{k=1}^n \{ L_k I'_k = 1 \}\right) \leq \sum_{k=1}^n E_\theta(L_k I'_k) \leq n \cdot \alpha/n = \alpha. \tag{9}
\]

2.3 Maximal and average loss among confirmed decisions

Note that condition (9) refers to the fact that the expected maximal loss among confirmed decisions is controlled at level \( \alpha \). In other words, \( E_\theta(\mathcal{L}) \leq \alpha \) for all \( \theta \in \Theta \), with respect to

\[
\mathcal{L} = \max \{L_1 I_1, \ldots, L_n I_n\}. \tag{10}
\]

Here, \( \mathcal{L} \) may be considered as a combined measure of loss among confirmed decisions. An alternative measure is given by the following average loss among confirmed decisions,

\[
\mathcal{L} = (L_1 I_1 + \cdots + L_n I_n)/(I_1 + \cdots + I_n), \tag{11}
\]
with \( \mathcal{L} \) interpreted as zero if no decisions are confirmed \( (I_1 = \cdots = I_n = 0) \).

In multiple testing, obtaining \( E_\theta(\mathcal{L}) \leq \alpha \) with respect to (10) is referred to as controlling the FWER (family-wise error rate, cf. [16, p. 349]) at level \( \alpha \). Similarly, obtaining \( E_\theta(\mathcal{L}) \leq \alpha \) with respect to (11) is referred to as controlling the FDR (false discovery rate, cf. [1]) at level \( \alpha \).
2.4 The Benjamini–Hochberg procedure

Consider non-random, zero-one loss functions $L_1, \ldots, L_n$ and corresponding $p$-values $\hat{R}_1, \ldots, \hat{R}_n$, with respect to some statistical model $\{P_\theta : \theta \in \Theta\}$. Let $\hat{R}(1), \ldots, \hat{R}(n)$ refer to $\hat{R}_1, \ldots, \hat{R}_n$ in a given ordering $\hat{R}(1) \leq \cdots \leq \hat{R}(n)$. Benjamini and Hochberg proved [1], assuming $\hat{R}_1, \ldots, \hat{R}_n$ to be independent random variables and a level $\alpha$, that the following indicator random variables (with $\hat{R}(\alpha+1) := \infty$ for notational convenience),

$$I_k = \sum_{i=1}^{n} I\{n\hat{R}_k \leq \alpha i\} \prod_{j=i+1}^{n+1} I\{n\hat{R}(j) > \alpha j\} I\{n\hat{R}(i) \leq \alpha i\}, \quad k = 1, \ldots, n,$$  

controls the expected average loss among confirmed decisions at level $\alpha$.

Note that Definition (12) is usually interpreted in a stepwise manner. Thus, starting at the bottom of the hierarchy of ordered $p$-values, $\hat{R}(n)$ is first compared with $\alpha$. If smaller than $\alpha$, all decisions are confirmed. Otherwise, $n\hat{R}(n-1)$ is compared with $\alpha(n-1)$, confirming all decisions corresponding to $\hat{R}(1), \ldots, \hat{R}(n-1)$ in case $\alpha(n-1)$ exceeds $n\hat{R}(n-1)$. Otherwise, $n\hat{R}(n-2)$ is compared with $(n-2)\alpha$, etcetera. Thus, the procedure confirms a random number $N$, $0 \leq N \leq n$, of the given decisions, corresponding to the $N$ smallest among $\hat{R}_1, \ldots, \hat{R}_n$, with $N$ determined by

$$N = \max\{i : \hat{R}(i) \leq i\alpha/n\},$$

(setting $\hat{R}(0) := 0$ for notational convenience).

Applying the definition (11) of $\mathcal{L}$ as the average loss among confirmed decisions to the indicators (12), one obtains

$$\mathcal{L} = \sum_{i=1}^{n} \frac{1}{i} I\{N = i\} \sum_{k=1}^{n} L_k I\{n\hat{R}_k \leq \alpha i\}$$

$$= \sum_{k=1}^{n} \sum_{i=1}^{n} \frac{1}{i} L_k I\{n\hat{R}_k \leq \alpha i\} \prod_{j=i+1}^{n+1} I\{n\hat{R}(j) > \alpha j\} I\{n\hat{R}(i) \leq \alpha i\}.$$ 

The fact that $E_\theta(\mathcal{L}) \leq \alpha$ can thus be expressed as

$$\sum_{k=1}^{n} \sum_{i=1}^{n} \frac{1}{i} E_\theta \left( L_k I\{n\hat{R}_k \leq \alpha i\} \prod_{j=i+1}^{n+1} I\{n\hat{R}(j) > \alpha j\} I\{n\hat{R}(i) \leq \alpha i\} \right) \leq \alpha. \quad (13)$$

3 Main results

Theorem 3.1 generalizes the validity of the bound (13) to the setting of non-negative loss functions $L_1, \ldots, L_n$ and corresponding generalized $p$-values $\hat{R}_1, \ldots, \hat{R}_n$. Part (i) generalize the original result [1]. Here, the pairs $(\hat{R}_1, \hat{R}_1), \ldots, (\hat{R}_n, \hat{R}_n)$ are assumed to be independent. However, nothing is assumed about the dependence between $\hat{R}_k$ and $\hat{R}_k$ within a given pair $(\hat{R}_k, \hat{R}_k)$. Part (ii) generalizes Theorem 1.3 in [5]. In this case, an increased bound on the expected average loss among confirmed decisions is given, valid without any assumption on dependence.
Theorem 3.1. Let \( \delta_1, \ldots, \delta_n \) be decision functions, with associated loss functions \( L_1, \ldots, L_n \) on some statistical model \( \{ P_\theta : \theta \in \Theta \} \). Moreover, let non-negative random variables \( \hat{R}_1, \ldots, \hat{R}_n \) and \( \theta \in \Theta \) be given such that, for an integer \( m, 0 \leq m \leq n \), relations (14)-(15) hold, for any \( \beta \leq \alpha \),
\[
\begin{align*}
 E_\theta(L_k I\{\hat{R}_k \leq \beta\}) & \leq \beta, \quad \text{for} \quad 1 \leq k \leq m, \quad (14) \\
 E_\theta(L_k I\{\hat{R}_k \leq \beta\}) & = 0, \quad \text{for} \quad m + 1 \leq k \leq n. \quad (15)
\end{align*}
\]

For a given \( \alpha > 0 \), consider the corresponding level-\( \alpha \) Benjamini–Hochberg procedure, and let \( \mathcal{L} \) refer to the average loss among confirmed decisions.

(i) Assume that the pairs \( (\delta_1, \hat{R}_1), \ldots, (\delta_n, \hat{R}_n) \) are independent. Then
\[
 E_\theta(\mathcal{L}) \leq \alpha m/n. \quad (16)
\]

Moreover, if equality holds in (14), for all \( 1 \leq k \leq m \) and all \( \beta \leq \alpha \), then equality also holds in (16).

(ii) Without any additional assumptions,
\[
 E_\theta(\mathcal{L}) \leq \alpha m/n \sum_{i=1}^{n} \frac{1}{i}. \]

Proof. To begin with, recall from (12)-(13) that \( E_\theta(\mathcal{L}) \) is given by (with the convention \( \hat{R}_{(n+1)} = \infty \)):
\[
 E_\theta(\mathcal{L}) = \sum_{k=1}^{m} \sum_{i=1}^{n} \frac{1}{i} E_\theta(L_k I\{n\hat{R}_k \leq \alpha i\} \prod_{j=i+1}^{n+1} I\{n\hat{R}_{(j)} > \alpha j\} I\{n\hat{R}_{(i)} \leq \alpha i\}).
\]

Here, applying assumption (15), it follows that
\[
 E_\theta(\mathcal{L}) = \sum_{k=1}^{m} \sum_{i=1}^{n} \frac{1}{i} E_\theta(L_k I\{n\hat{R}_k \leq \alpha i\} \prod_{j=i+1}^{n+1} I\{n\hat{R}_{(j)} > \alpha j\} I\{n\hat{R}_{(i)} \leq \alpha i\}). \quad (17)
\]

For each integer \( k, 1 \leq k \leq m \), let \( X_{2,k} \leq \cdots \leq X_{n,k} \) denote the ordered random variables \( \hat{R}_{(1)}, \ldots, \hat{R}_{(n)} \) when \( \hat{R}_k \) is removed. Moreover, set \( X_{1,k} = 0 \) and \( X_{n+1,k} = \infty \), for convenience. Thus, rewriting (17) we obtain
\[
 E_\theta(\mathcal{L}) = \sum_{k=1}^{m} \sum_{i=1}^{n} \frac{1}{i} E_\theta(L_k I\{n\hat{R}_k \leq \alpha i\} \prod_{j=i+1}^{n+1} I\{nX_{j,k} > \alpha j\} I\{nX_{i,k} \leq \alpha i\}). \quad (18)
\]

Regarding part (i), note that \( X_{2,k} \leq \cdots \leq X_{n,k} \) are independent of the two random variables \( L_k(\delta_k, \theta) \) and \( \hat{R}_k \). Hence, (18) simplifies to
\[
 E_\theta(\mathcal{L}) = \sum_{k=1}^{m} \sum_{i=1}^{n} \frac{1}{i} E_\theta(L_k I\{n\hat{R}_k \leq \alpha i\}) E\left( \prod_{j=i+1}^{n+1} I\{nX_{j,k} > \alpha j\} I\{nX_{i,k} \leq \alpha i\} \right). \quad (19)
\]

\[7\]
Here, note that, for any \(1 \leq k \leq m\),
\[
\sum_{i=1}^{n} \prod_{j=i+1}^{n+1} I\{nX_{j,k} > \alpha j\} I\{nX_{i,k} \leq \alpha i\} = 1,
\]
(20)
since \(0 = X_{1,k} \leq X_{2,k} \leq \cdots \leq X_{n+1,k} = \infty\). Thus, applying assumption (14) in (19), it follows from (20) that \(E_{\theta}(L) \leq \alpha m/n\), with equality in case of equality in (14), which completes the proof of part (i).

Regarding part (ii), we note in view of (20) that, for any \(1 \leq k \leq m\),
\[
\sum_{i=1}^{n} \prod_{j=i+1}^{n} I\{n\hat{R}_{k} \leq \alpha i\} \sum_{i=1}^{n} \prod_{j=i+1}^{n+1} I\{nX_{j,k} > \alpha j\} I\{nX_{i,k} \leq \alpha i\}
\]
\[
\leq \max_{1 \leq i \leq n} \left\{ \frac{1}{i} I\{n\hat{R}_{k} \leq \alpha i\} \right\} \sum_{i=1}^{n} \prod_{j=i+1}^{n+1} I\{nX_{j,k} > \alpha j\} I\{nX_{i,k} \leq \alpha i\}
\]
\[
= \max_{1 \leq i \leq n} \left\{ \frac{1}{i} I\{n\hat{R}_{k} \leq \alpha i\} \right\}.
\]
(21)
We also note that
\[
\max_{1 \leq i \leq n} \left\{ \frac{1}{i} I\{n\hat{R}_{k} \leq \alpha i\} \right\} = \sum_{i=1}^{n} \frac{1}{i} I\{\alpha(i-1) < n\hat{R}_{k} \leq \alpha i\}
\]
\[
= \frac{1}{n} I\{\hat{R}_{k} \leq \alpha\} + \sum_{i=1}^{n-1} \left( \frac{1}{i} - \frac{1}{i+1} \right) I\{n\hat{R}_{k} \leq \alpha i\}
\]
\[
= \frac{1}{n} I\{\hat{R}_{k} \leq \alpha\} + \sum_{i=1}^{n-1} \frac{1}{i(i+1)} I\{n\hat{R}_{k} \leq \alpha i\}.
\]
(22)
Applying (21)-(22) in (18), it follows that
\[
E_{\theta}(L) \leq \sum_{k=1}^{m} E_{\theta}\left(\frac{1}{n} I\{\hat{R}_{k} \leq \alpha\} + \sum_{i=1}^{n-1} \frac{1}{i(i+1)} I\{n\hat{R}_{k} \leq \alpha i\}\right).
\]
(23)
Finally, applying assumption (14) to (23), we deduce that
\[
E_{\theta}(L) \leq \sum_{k=1}^{m} \left( \frac{\alpha}{n} + \frac{\alpha}{n} \sum_{i=1}^{n-1} \frac{1}{i+1} \right) = \alpha \frac{m}{n} \sum_{i=1}^{n} \frac{1}{i},
\]
proving part (ii) of the theorem.

\[\square\]

4 A simulation study

As a simple statistical model for evaluation of multiple testing procedures, we now consider so-called Dirac-uniform configurations (referred to as “U1” in Table 1). More precisely, let \(0 \leq m \leq n\) be given, with \(m\) determining the number of true null hypotheses. Then, let \(\hat{p}_1, \ldots, \hat{p}_m\) refer to i.i.d. random variables uniformly distributed...
on the interval $[0, 1]$ and assume that remaining $p$-values are constantly equal to zero, 
\[ \hat{p}_{m+1} = \cdots = \hat{p}_n = 0 \]

With the generalized perspective (cf. Section 2) we also consider a related model “$U_\frac{1}{2}$.” Similarly, with $0 \leq m \leq n$ given, let $\hat{R}_1, \ldots, \hat{R}_m$ be i.i.d. random variables uniformly distributed on the interval $[0, 1/2]$, with associated decisions $\delta_1, \ldots, \delta_m$ i.i.d. uniformly distributed on the binary set $\{0, 1\}$. The remaining generalized $p$-values and decision functions are assumed to be constantly equal to zero, $\hat{R}_{m+1} = \cdots = \hat{R}_n = 0 = \delta_{m+1} = \cdots = \delta_n$. All decisions are evaluated with respect to a loss function $L(\delta) = I\{\delta = 1\}$.

As a motivation of the $U_\frac{1}{2}$-model in this situation, consider the treatment of directional conclusions in Section 2.1. A simple example in that context is given by the statistical model $\{N(\mu, 1) : \mu \in \mathbb{R}\}$ of unit variance, univariate normal distributions, divided into $H_1 : \mu \leq 0$ and $H_2 : \mu > 0$. Let the naive directional decision function be given by

\[ \delta = 2 \cdot I\{X > 0\} + 1 \cdot I\{X \leq 0\} \]

Then, $\hat{R} = \Phi(-|X|)$ serves as a generalized $p$-value with respect to $\delta$ and the directional loss function,

\[ L = I\{\mu \leq 0\}I\{X > 0\} + I\{\mu > 0\}I\{X \leq 0\} \]

Note that $\hat{R}$ is uniformly distributed on $[0, 1/2]$, and that $\delta$ is uniformly distributed on $\{0, 1\}$, with $\hat{R}$ and $\delta$ independent, in case $\mu = 0$. Thus, the $U_\frac{1}{2}$-model corresponds to directional conclusions with $\mu_1 = \cdots = \mu_m = 0, \mu_{m+1} = \cdots = \mu_n = \infty$, and independent observations.

Table 1 reports on estimated FDR and FWER regarding six multiple testing procedures applied to $\hat{R}_1, \ldots, \hat{R}_n$ in the above models $U_1$ and $U_\frac{1}{2}$. All procedures are performed at level $\alpha = 0.05$, with respect to $n = 100$ and $m = 2, 5, 20, 99$. Moreover, all procedures are known to control the FDR/FWER (respectively) at level $\alpha = 0.05$ for $U_1$.

<table>
<thead>
<tr>
<th></th>
<th>$m = 2$</th>
<th></th>
<th>$m = 5$</th>
<th></th>
<th>$m = 20$</th>
<th></th>
<th>$m = 99$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$U_1$</td>
<td>$U_\frac{1}{2}$</td>
<td>$U_1$</td>
<td>$U_\frac{1}{2}$</td>
<td>$U_1$</td>
<td>$U_\frac{1}{2}$</td>
<td>$U_1$</td>
<td>$U_\frac{1}{2}$</td>
</tr>
<tr>
<td>BH</td>
<td>0.001</td>
<td>0.001</td>
<td>0.0025</td>
<td>0.0025</td>
<td>0.010</td>
<td>0.010</td>
<td>0.050</td>
<td>0.050</td>
</tr>
<tr>
<td>HM</td>
<td>0.049</td>
<td>0.051</td>
<td>0.049</td>
<td>0.050</td>
<td>0.049</td>
<td>0.049</td>
<td>0.049</td>
<td>0.049</td>
</tr>
<tr>
<td>HG</td>
<td>0.050</td>
<td>0.053</td>
<td>0.049</td>
<td>0.050</td>
<td>0.049</td>
<td>0.049</td>
<td>0.049</td>
<td>0.049</td>
</tr>
<tr>
<td>BL</td>
<td>0.020</td>
<td>0.010</td>
<td>0.050</td>
<td>0.025</td>
<td>0.004</td>
<td>0.004</td>
<td>0.026</td>
<td>0.025</td>
</tr>
<tr>
<td>GBS</td>
<td>0.016</td>
<td>0.010</td>
<td>0.034</td>
<td>0.025</td>
<td>0.050</td>
<td>0.083</td>
<td>0.050</td>
<td>0.049</td>
</tr>
<tr>
<td>$S_{\frac{1}{2}}$</td>
<td>0.010</td>
<td>0.010</td>
<td>0.025</td>
<td>0.025</td>
<td>0.050</td>
<td>0.100</td>
<td>0.050</td>
<td>0.495</td>
</tr>
</tbody>
</table>

Table 1: Estimated FDR and FWER for the two models $U_1$ and $U_\frac{1}{2}$ for $n = 100$ and selected values of $m$. BH refers to the Benjamini–Hochberg FDR-procedure, HM to the Holm FWER-procedure [12], HG to the Hochberg FWER-procedure [10], BL to the Benjamini–Liu FDR-procedure [4], GBS to the Gavrilov–Benjamini–Sarkar FDR-procedure [9], and $S_{\frac{1}{2}}$ to the Storey FDR-procedure [23] with $\lambda = 1/2$.

Note to begin with that the error rate of the Benjamini–Hochberg procedure is constant when comparing the two models $U_1$ and $U_\frac{1}{2}$ with $m$ fixed. Indeed, the
expected proportion of false conclusions is given by \( \alpha m/n \), as explained by Theorem 3.1 (i) above.

Next, the probability of a false conclusion with the Holm procedure [12] applied to the \( U_{1/2} \) model with \( m = 2 \) is slightly larger than 5\%. Indeed, an error is committed with probability 1/2 in case exactly one of \( \delta_1 \) and \( \delta_2 \) is confirmed, which occurs in case

\[
\hat{R}_1 \leq \alpha/2, \hat{R}_2 > \alpha \quad \text{or} \quad \hat{R}_2 \leq \alpha/2, \hat{R}_1 > \alpha.
\]

Thus, the probability of confirming exactly one of \( \delta_1 \) and \( \delta_2 \) equals \( 2\alpha(1 - 2\alpha) \). Moreover, an error is committed with probability 3/4 in case both \( \delta_1 \) and \( \delta_2 \) are confirmed, that is, whenever

\[
\hat{R}_1 \leq \alpha/2, \hat{R}_2 \leq \alpha \quad \text{or} \quad \hat{R}_2 \leq \alpha/2, \hat{R}_1 \leq \alpha.
\]

Here, the probability of confirming both \( \delta_1 \) and \( \delta_2 \) equals \( 2 \cdot \alpha \cdot 2\alpha - \alpha^2 = 3\alpha^2 \). Thus, the total probability of an error is given by

\[
2\alpha(1 - 2\alpha)/2 + 9\alpha^2/4 = \alpha + \alpha^2/4 > \alpha.
\]

In the case of \( \alpha = 0.05 \),

\[
\alpha + \alpha^2/4 \approx 0.0506 \approx 0.051,
\]

as indicated by Table 1.

As a consequence of the behaviour of the Holm procedure, the Hochberg FWER-procedure [10] applied to the \( U_{1/2} \) model with \( m = 2 \) also yields an error larger than 5\%, since it always confirms more decisions than the Holm procedure. Indeed, the procedure HG is the “step-up” version of the “step-down” procedure HM.

The step-down FDR-controlling procedure of Benjamini–Liu [4] is known to be more powerful compared to the Benjamini-Hochberg procedure when \( m \) is small (cf. [4]). This is in agreement with Table 1, where it can be seen that BL produces larger false discovery rates for \( m = 2 \) and 5 compared to BH. Also, it appears from Table 1 that BL is conservative with respect to the model \( U_{1/2} \). In fact, one may verify that the false discovery rates of BL applied to \( U_{1/2} \) is approximately bounded by 0.025, for any \( 0 \leq m \leq 100 \) in this case.

Finally, the two alternative FDR-procedures which we here consider (cf. [8], [9] for GBS, and [22], [23] for \( S_{1/2} \)) sometimes produce false discovery rates much larger than 5\% when applied to the \( U_{1/2} \)-model. This seems to be characteristic for “adaptive” FDR-controlling procedures (cf. also [3] and [2]). Indeed, the notion of adaptiveness often refers to the strategy of incorporating an estimated value of the number of true null hypotheses into an existing step-wise procedure (cf. e.g. [3, Section 3]). This strategy may thus be closely related to the assumption of data-invariant, zero-one loss functions \( L_1, \ldots, L_n \).

References


