Electronic Journal of Statistics
Vol. 13 (2019) 882–910
ISSN: 1935-7524
https://doi.org/10.1214/19-EJS1544

Variability and stability of the false discovery proportion

Marc Ditzhaus and Arnold Janssen

Heinrich-Heine University Düsseldorf
Mathematical Institute
Universitätststraße 1
40225 Düsseldorf, Germany
e-mail: marc.ditzhaus@hhu.de; janssena@uni-duesseldorf.de

Abstract: Much effort has been done to control the "false discovery rate" (FDR) when m hypotheses are tested simultaneously. The FDR is the expectation of the "false discovery proportion" FDP $= V/R$ given by the ratio of the number of false rejections V and all rejections R. In this paper, we have a closer look at the FDP for adaptive linear step-up multiple tests. These tests extend the well known Benjamini and Hochberg test by estimating the unknown amount m_0 of the true null hypotheses. We give exact finite sample formulas for higher moments of the FDP and, in particular, for its variance. Using these allows us a precise discussion about the stability of the FDP, i.e., when the FDP is asymptotically close to its mean. We present sufficient and necessary conditions for this stability. They include the presence of a stable estimator for the proportion m_0/m. We apply our results to convex combinations of generalized Storey type estimators with various tuning parameters and (possibly) data-driven weights. The corresponding step-up tests allow a flexible adaptation. Moreover, these tests control the FDR at finite sample size. We compare these tests to the classical Benjamini and Hochberg test and discuss the advantages of them.

MSC 2010 subject classifications: Primary 62G10; secondary 62G20.
Keywords and phrases: False discovery rate (FDR), moments of the false discovery proportion (FDP), multiple testing, adaptive Benjamini Hochberg methods, Storey estimator, p-values.

Received August 2018.

1. Introduction

Testing $m \geq 2$ hypotheses simultaneously is a frequent issue in statistical practice, e.g., in genomic research. A widely used criterion for deciding which of these hypotheses should be rejected is the so-called "false discovery rate" (FDR) promoted by Benjamini and Hochberg [3]. The FDR is the expectation of the "false discovery proportion" (FDP), the ratio

$$\mathrm{FDP}_m = \frac{V_m}{R_m}$$

of the number of false rejections V_m and the amount of all rejections R_m. We call a multiple test procedure (FDR-)α-controlling for a pre-specified level $\alpha \in (0,1)$

when $\mathrm{FDR}_m = \mathbb{E}(\mathrm{FDP}_m) \leq \alpha$. Under the so-called basic independence (BI) assumption, which will be introduced in more detail below, the classical Benjamini and Hochberg linear step-up test, in short BH test, is α-controlling. In fact, there is an exact formula for its FDR, namely $\mathrm{FDR}_m = (m_0/m)\alpha$ [3, 19], where m_0 is the unknown amount of true null hypotheses. Especially, if m_0/m is not close to 1 the BH test should be improved regarding a better exhaustion of the FDR level to achieve higher power. For this purpose, so-called adaptive procedure can be used. The basic idea is to estimate m_0 by an appropriate estimator $\widehat{m}_0$ in a first step and to apply the BH test for the data dependent level $\alpha' = (m/\widehat{m}_0)\alpha$ in the second step. We can expect a better FDR exhaustion for a good estimator $\widehat{m}_0 \approx m_0$ because, heuristically, $\mathrm{FDR}_m \approx (m_0/m)\alpha' \approx \alpha$. Various estimators are suggested in the literature [4, 5, 7, 8, 36, 38, 39, 40, 42]. Generalized Storey estimators with data dependent weights discussed by Heesen and Janssen [24] will be our prime example in later discussions of our general results. The latter lead to α-controlling procedures, whereas other approaches are often just asymptotically α-controlling, i.e., $\limsup_{m\to\infty} \mathrm{FDR}_m \leq \alpha$. Sufficient conditions for estimators leading to (finite sample) α-controlling procedures can be found in Sarkar [32] and Heesen and Janssen [23, 24]. Adaptive procedures are also used to get procedures controlling the family-wise error rate $\mathrm{FWER}_m = P(V_m > 0)$, another criterion for multiple tests, for details we refer to Finner and Gontscharuk [17] and Sarkar et al. [33] as well as the references therein.

Due to the additional estimation step, the variability of the FDP_m is higher for adaptive procedures. This is contrary to the actual idea of α-controlling methods, namely to ensure in a certain way that the proportion of false rejections FDP_m is small. In fact, methods are preferable when the inequality $\mathrm{FDP}_m \leq (\alpha+\varepsilon)$ holds with a high probability and small $\varepsilon > 0$. That is why we address the question for which adaptive procedures this property can be expected. Ferreira and Zwinderman [15] presented formulas for higher moments of FDP_m for the BH test and Roquain and Villers [31] did so for step-up and step-down tests with general (but data independent) critical values. We extend these formulas to adaptive procedures. In particular, we derive a finitely exact variance formula for FDP_m. Combining this and Chebyshev's inequality we obtain an upper bound for the undesired event of a relatively large FDP for α-controlling procedures:

$$P\Big(\frac{V_m}{R_m} > \alpha + \varepsilon\Big) \leq \frac{\mathrm{Var}(V_m/R_m)}{(\alpha - \mathrm{FDR}_m + \varepsilon)^2}. \tag{1.1}$$

Under mixture p-value models Chi and Tan [11] already derived bounds and asymptotic results for $P(V_m > \alpha R_m)$, see also Chi [10]. In the spirit of (1.1), for good procedures we expect that the variance of FDP_m is small or even vanishes in the asymptotic set-up, i.e., if the number of hypothesis m tends to infinity. In the latter case, we say that FDP_m is stable. To be mathematically more precise, FDP_m is (asymptotically) stable if

$$\frac{V_m}{R_m} - \mathbb{E}\Big(\frac{V_m}{R_m}\Big) \to 0 \text{ in probability, or equivalently } \mathrm{Var}\Big(\frac{V_m}{R_m}\Big) \to 0. \tag{1.2}$$

Note that $\mathbb{E}(V_m/R_m)$ is not convergent in general but for appropriate subsequences. At long these subsequences V_m/R_m converges in probability to a constant under (1.2). Using our exact variance formula for FDP_m we determine sufficient and necessary conditions for stability in the sense of (1.2). We also treat the more challenging case of sparsity in the sense that $m_0/m \to 1$ as $m \to \infty$. This situation can be compared to the one of Abramovich et al. [1], who derived an estimator of the (sparse) mean of a multivariate normal distribution using FDR procedures. In the asymptotic set-up, stochastic process methods were applied to study the asymptotic behavior of FDP_m and FWER_m, e.g. asymptotic confidence intervals were calculated [20, 27, 28, 29]. Since FDP_m is an unknown quantity in practice, estimation of it is a further interesting topic. For various correlated test statistics, mainly normal and χ^2-statistics, estimators of FDP_m and FDR_m were studied [13, 30, 35, 39, 41].

Outline of the results. In Section 2, we introduce the model as well as the adaptive step-up tests and, in particular, the generalized Storey estimators serving as our prime examples. Section 3 provides exact finite sample variance formulas for the FDP_m under the BI model. Extensions to higher moments can be found in the appendix, see Section 9. These results apply to the variability and the stability of FDP_m, see Section 4. Roughly speaking, we have stability if we have a stable estimator $\widehat{m}_0/m \approx C_0$ and the number of rejections tends to infinity. Section 5 is devoted to concrete adaptive step-up tests mainly based on the convex combinations of generalized Storey estimators with data dependent weights. We will see that stability cannot be achieved in general. Under mild assumptions the adaptive tests based on the estimators mentioned above are superior compared to the BH test: 1. The adaptive procedures lead to a more exhausted FDR while remaining α-controlling. 2. The corresponding FDP is stable whenever the FDP of the BH test is stable. In Section 6, we discuss least favorable configurations which serve as useful technical tools. For the reader's convenience we add a discussion and summary of the paper in Section 7. All proofs are collected in Section 8.

2. Preliminaries

2.1. *The model and general step-up tests*

Let us first describe the model and the procedures. A multiple testing problem consists of m null hypotheses $(H_{i,m}, p_{i,m})$ with associated p-values $0 \le p_{i,m} \le 1$ on a common probability space $(\Omega_m, \mathcal{A}_m, P_m)$. For all asymptotic consideration the limits are meant as $m \to \infty$, if not stated otherwise. From now on, we always suppose the basic independence **(BI)** assumption given by

(BI1) The set of hypotheses can be divided in the disjoint union $I_{0,m} \bigcup I_{1,m} = \{1, ..., m\}$ of the (unknown) true null portion $I_{0,m}$ and the (unknown) false null portion $I_{1,m}$. Denote by $m_j = \# I_{j,m}$ the cardinality of $I_{j,m}$ for $j = 0, 1$.

(BI2) The p-value vectors $(p_{i,m})_{i \in I_{0,m}}$ and $(p_{i,m})_{i \in I_{1,m}}$ are independent, where any dependence structure is allowed for the p-values $(p_{i,m})_{i \in I_1}$ corresponding to false hypotheses.

(BI3) The p-values $(p_{i,m})_{i\in I_{0,m}}$ corresponding to true hypotheses are independent and uniformly distributed on $[0,1]$, i.e., $P_m(p_{i,m} \le x) = x$ $(x \in [0,1])$.

Throughout the paper, let $m_0 \ge 1$ be nonrandom. Similarly to Heesen and Janssen [24], our results can be extended to more general models with random m_0 by conditioning under m_0. By using this modification our results can easily be transferred to familiar mixture models discussed, among others, by Abramovich et al. [1] and Genovese and Wassermann [20]. We study adaptive multiple step-up tests with estimated critical values extending the famous Benjamini and Hochberg [3] step-up test, in short BH test. In the following we recall the definition of these kinds of tests. Let

$$0 = \alpha_{0:m} < \alpha_{1:m} \le \alpha_{2:m} \le \ldots \le \alpha_{m:m} < 1 \tag{2.1}$$

denote possibly data dependent critical values. As an example for the critical values we recall the ones for the BH test, which do not depend on the data:

$$\alpha^{\mathrm{BH}}_{i:m} = \frac{i}{m}\alpha.$$

If $p_{1:m} \le p_{2:m} \le \ldots \le p_{m:m}$ denote the ordered p-values then the number of rejections is given by

$$R_m := \max\{i = 0, \ldots, m : p_{i:m} \le \alpha_{i:m}\}, \text{ where } p_{0:m} := 0,$$

and the rejection rule for the multiple procedure is as follows:

$$\text{reject } H_{i,m} \text{ iff } p_{i,m} \le \alpha_{R_m:m}.$$

Moreover, let

$$V_m := \#\{i \in I_{0,m} \cup \{0\} : p_{i:m} \le \alpha_{R_m:m}\} \tag{2.2}$$

be the number of falsely rejected null hypotheses. Then the false discovery rate FDR_m and the false discovery proportion FDP_m are given by

$$\mathrm{FDP}_m = \frac{V_m}{R_m} \text{ and } \mathrm{FDR}_m = \mathbb{E}\Big(\frac{V_m}{R_m}\Big) \text{ with } \frac{0}{0} = 0. \tag{2.3}$$

Good multiple tests like the BH test or the frequently applied adaptive test of Storey et al. [39] control the FDR at a pre-specified acceptance error bound α at least under the BI assumption, in short we say that they are α-controlling. In addition to this property, two further aspects for multiple procedures are of importance and discussed below:

(i) To make the test sensitive for signal detection the FDR should exhaust the level α as best as possible.
(ii) On the other hand the variability of FDP_m is of interest in order to judge the stability of FDP_m.

For a large class of adaptive tests exact FDR formulas were established in Heesen and Janssen [24]. These formulas are now completed by formulas for exact higher FDP moments and, in particular, for the variance. These results can be used to discuss conditions for (1.2). Throughout the paper, $\liminf_{m\to\infty} \mathrm{FDR}_m > 0$ is assumed. Then the following condition is necessary for the stability of FDP_m:

$$\lim_{m\to\infty} P_m(V_m > 0) = 1. \tag{2.4}$$

As already stated we can not expect stability in general. In the following we discuss two negative results concerning the stability for the BH test.

Example 2.1. (a) We explain in the following that (2.4) is never fulfilled for the BH test in the extreme case of $m_1 \geq 0$ being fixed. Hence, FDP_m is never stable for the BH test in this case. First, note that $P_m(V_m^{\mathrm{BH}} = 0)$ is minimal (implying that $P_m(V_m^{\mathrm{BH}} > 0)$ is maximal) for the so-called Dirac uniform configuration $\mathrm{DU}(m, m_1)$, where all entries of $(p_{i,m})_{i\in I_{1,m}}$ are equal to zero. Under this configuration $V_m^{\mathrm{BH}}(\alpha, m_1) \to V_{\mathrm{SU}}(\alpha, m_1)$ in distribution with

$$P\Big(V_{\mathrm{SU}}(\alpha, m_1) = 0\Big) = (1-\alpha)\exp(-m_1\alpha) > 0,$$

see Finner and Roters [19] and Theorem 4.8 of Scheer [34]. The limit variable belongs to the class of linear Poisson distributions [12, 18, 25].

(b) Let $(p_{i,m})_{i\in I_{1,m}}$ be i.i.d. uniformly distributed on $[0,\lambda]$ for $\lambda \in (\alpha, 1)$. Note that the distribution of $p_{i,m}$ given the event $\{p_{i,m} \leq \lambda\}$ is the same for all $i = 1, \dots, m$. In Theorem 5.1(b) below, we show that FDP_m is not stable for the BH test in this setting.

The requirement for stability will be somehow between $\mathrm{DU}(m, m_1)$ alternatives and the setting of Example 2.1(b), where the assumption $m_1 \to \infty$ will always be needed. More information about $\mathrm{DU}(m, m_1)$ and least favorable configurations can be found in Section 6.

2.2. Our step-up tests and underlying assumptions

In the following we introduce the adaptive step-up tests. Let $0 < \alpha < 1$ be a fixed level. A tuning parameter $\lambda \in [\alpha, 1)$ is chosen such that no null $\mathcal{H}_{i,m}$ with $p_{i,m} > \lambda$ should be rejected. For instance, it is uncommon to reject a null if the corresponding p-value is large than $\lambda_0 = 1/2$, even a rejection when $p_{i,m} > \lambda_1 = \alpha$ is rather unusual. In this spirit, we split the range $[0,1]$ of the p-values into a decision region $[0,\lambda]$, where we may reject the corresponding null hypotheses, and an estimation region $(\lambda, 1]$, where we use the p-values to estimate m_0, see Figure 1. To be more specific, we consider estimators

$$\widehat{m}_0 = \widehat{m}_0((\widehat{F}_m(t))_{t\geq\lambda}) > 0 \tag{2.5}$$

for m_0, which are measurable functions depending only on $(\widehat{F}_m(t))_{t\geq\lambda}$. As usual we denote by $\widehat{F}_m$ the empirical distribution function of the p-values

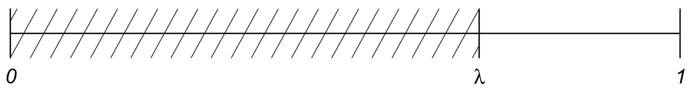

FIG 1. *Decision region* $[0,\lambda]$ *(dashed) and estimation region* $(\lambda,1]$.

$p_{1,m},\dots,p_{m,m}$. As motivated in the introduction, we now plug-in these estimators in the BH test. Doing this we obtain the data driven critical values

$$\widehat{\alpha}_{i:m} = \min\Big\{\Big(\frac{i}{\widehat{m}_0}\alpha\Big),\lambda\Big\}, \quad i=1,\dots,n, \tag{2.6}$$

where we promote to use the upper bound λ as Heesen and Janssen [24] already did. The following two quantities will be frequently used:

$$R_m(t) = m\widehat{F}_m(t) \quad \text{and} \quad V_m(t) = \sum_{i\in I_0} 1\{p_{i,m}\le t\},\ t\in[0,1]. \tag{2.7}$$

Throughout this paper, we investigate different mild assumptions. For our main results we fix the following two assumptions:

(A1) Suppose that

$$\frac{m_0}{m} \to \kappa_0 \in (0,1].$$

If only $0 < \liminf_{m\to\infty} m_0/m$ is valid then our results apply to appropriate subsequences. The most interesting case is $\kappa_0 > \alpha$ since otherwise (if $m_0/m \le \alpha$) the FDR can be controlled, i.e. $\mathrm{FDR}_m \le \alpha$, by rejecting everything.

(A2) Suppose that $\widehat{m}_0$ is always positive and

$$\frac{\lambda}{\alpha}\widehat{m}_0 \ge R_m(\lambda).$$

If (A2) is not fulfilled then consider the estimator $\max\{\widehat{m}_0,(\alpha/\lambda)R_m(\lambda)\}$ instead of $\widehat{m}_0$. Note that both estimators lead to the same critical values (2.6) and, thus, it is irrelevant which of these two estimators is used. Consequently, (A2) is not a restriction for the practical application but is improving the readability of our formulas significantly.

Remark 2.2. Under (A2) the FDR of the adaptive multiple test was obtained for the BI model by Heesen and Janssen [24]:

$$\mathrm{FDR}_m = \frac{\alpha}{\lambda}\mathbb{E}\Big(\frac{V_m(\lambda)}{\widehat{m}_0}\Big). \tag{2.8}$$

In particular, we obtain

$$\mathrm{FDR}_m \le \mathbb{E}\Big(\frac{V_m(\lambda)}{R_m(\lambda)}\Big) \le P_m(V_m(\lambda)>0),$$

where the upper bound is always strictly smaller than 1 for finite m.

A prominent α-controlling adaptive test is based on the Storey estimator

$$\widetilde{m}_0^{\mathrm{Stor}}(\lambda) = m\frac{1-\widehat{F}_m(\lambda)+\frac{1}{m}}{1-\lambda}. \tag{2.9}$$

To obtain an estimator fulfilling (A2), we consider

$$\widehat{m}_0^{\mathrm{Stor}}(\lambda) := \max\{\widetilde{m}_0(\lambda), \frac{\alpha}{\lambda}R_m(\lambda)\} \tag{2.10}$$

instead. A refinement was established by Heesen and Janssen [24]. They introduced a couple of inspection points $0 < \lambda = \lambda_0 < \lambda_1 < \ldots < \lambda_k = 1$, where m_0 is estimated by

$$\widetilde{m}_0(\lambda_{i-1},\lambda_i) := m\frac{\widehat{F}_m(\lambda_i)-\widehat{F}_m(\lambda_{i-1})+\frac{1}{m}}{\lambda_i-\lambda_{i-1}} \tag{2.11}$$

on each interval $(\lambda_{i-1},\lambda_i]$. Liang and Nettleton [26] already used the estimators (2.11) in another context. The Storey estimator can be rewritten as a linear convex combination of these estimators:

$$\widetilde{m}_0^{\mathrm{Stor}}(\lambda) = \sum_{i=1}^{k}\beta_i m\frac{\widehat{F}_m(\lambda_i)-\widehat{F}_m(\lambda_{i-1})+\frac{1}{m}}{\lambda_i-\lambda_{i-1}} = \sum_{i=1}^{k}\beta_i\widetilde{m}_0(\lambda_{i-1},\lambda_i)$$

with weights $\beta_i = (\lambda_i-\lambda_{i-1})/(1-\lambda)$ fulfilling the condition $\sum_{i=1}^k \beta_i = 1$. Heesen and Janssen [23] allowed the weights to be data dependent and proved that the corresponding adaptive test is α-controlling under the BI assumption, see Proposition 2.3 below. In Section 5, we discuss the stability of FDP_m for these procedures.

Proposition 2.3 (cf. Thm 10 in [24]). *Let $\widehat{\beta}_{i,m} = \widehat{\beta}_{i,m}((\widehat{F}_m(t))_{t\geq\lambda_i}) \geq 0$ be random weights for $i \leq k$ with $\sum_{i=1}^k \widehat{\beta}_{i,m} = 1$. The adaptive step-up test using the following estimator $\widetilde{m}_0$ is α-controlling:*

$$\widetilde{m}_0 := \sum_{i=1}^{k}\widehat{\beta}_{i,m}\widetilde{m}_0(\lambda_{i-1},\lambda_i). \tag{2.12}$$

Finally, we want to present a sufficient condition of asymptotic α-control.

Proposition 2.4 (cf. Thm 6.1 in [23]). *Suppose that (A1), (A2) holds. If*

$$P_m\Big(\frac{\widehat{m}_0}{m_0} \leq 1-\delta\Big) \to 0 \textit{ for all } \delta > 0$$

then we have asymptotic (FDR-)α-control, i.e., $\limsup_{m\to\infty} FDR_m \leq \alpha$.

It should be mentioned that Proposition 2.4 is even valid for reverse martingale models, a huge model class including, among others, the BI model. Finner and Gontscharuk [17] proved asymptotic FWER-control under the same conditions.

3. Moments

This section provides exact second moment formulas of $\mathrm{FDP}_m = V_m/R_m$ for our adaptive step-up tests for a fixed regime P_m. Our method of proof relies on conditioning with respect to the σ-algebra

$$\mathcal{F}_{\lambda,m} := \sigma\Big(\mathbf{1}\{p_{i,m} \le s\} : s \in [\lambda, 1],\, 1 \le i \le m\Big).$$

Conditionally under the (non-observable) σ-algebra $\mathcal{F}_{\lambda,m}$ the quantities $\widehat{m}_0$, $R_m(\lambda)$ and $V_m(\lambda)$ are fixed values. But only $R_m(\lambda) = m\widehat{F}_m(\lambda)$ and $\widehat{m}_0$ are observable. The FDR formula (2.8) is now completed by an exact variance formula. The proof offers also a rapid approach to the known moment formulas of Ferreira and Zwinderman [15] for the Benjamini and Hochberg test (with $\widehat{m}_0 = m$ and $\lambda = \alpha$). Without loss of generality we reorder the p-values such that

$$I_{0,m} = \{1, \dots, m_0\} \text{ and } I_{1,m} = \{m_0 + 1, \dots, m\}.$$

Now, we introduce a new p-value vector $p_m^{(1,\lambda)}$. If $V_m(\lambda) > 0$ then set $p_m^{(1,\lambda)}$ equal to the vector $p_m = (p_{1,m}, \dots, p_{m,m})$ while replacing one p-value $p_{i,m} \le \lambda$, $i \le m_0$, corresponding to a true null hypothesis by 0, for convenience take the smallest integer $i \le m_0$ with this property. If $V_m(\lambda) = 0$ then set $p_m^{(1,\lambda)} = p_m$. Let $R_m^{(1,\lambda)} = R_m^{(1,\lambda)}(p_m^{(1,\lambda)})$ be the number of rejections based on the substituted vector $p_m^{(1,\lambda)}$ instead of the original p-value vector p_m. Note that $\widehat{m}_0$ remains the same when considering $p_m^{(1,\lambda)}$ instead of p_m while $R_m^{(1,\lambda)} > R_m$ is possible.

Theorem 3.1. *Suppose that our assumptions (A2) are fulfilled.*

(a) The second moment of FDP_m is given by

$$\mathbb{E}\Big(\Big(\frac{V_m}{R_m}\Big)^2\Big) = \mathbb{E}\Big(\frac{\alpha^2 V_m(\lambda)(V_m(\lambda)-1)}{\lambda^2 \widehat{m}_0^2} + \frac{\alpha}{\lambda}\frac{V_m(\lambda)}{\widehat{m}_0}\mathbb{E}\Big(\frac{1}{R_m^{(1,\lambda)}}\Big|\mathcal{F}_{\lambda,m}\Big)\Big).$$

(b) The variance of FDP_m fulfills

$$Var\Big(\frac{V_m}{R_m}\Big) = \frac{\alpha^2}{\lambda^2}\Big[\frac{\lambda}{\alpha}\mathbb{E}\Big(\frac{V_m(\lambda)}{\widehat{m}_0}\mathbb{E}\Big(\frac{1}{R_m^{(1,\lambda)}}\Big|\mathcal{F}_{\lambda,m}\Big)\Big) + Var\Big(\frac{V_m(\lambda)}{\widehat{m}_0}\Big) - \mathbb{E}\Big(\frac{V_m(\lambda)}{\widehat{m}_0^2}\Big)\Big].$$

(c) We have

$$\mathbb{E}(V_m) = \frac{\alpha}{\lambda}\mathbb{E}\Big(\frac{V_m(\lambda)}{\widehat{m}_0}\mathbb{E}(R_m^{(1,\lambda)}|\mathcal{F}_{\lambda,m})\Big).$$

Exact higher moment formulas are established in Section 9.

4. The variability and stability of $\mathbf{FDP}_m$

In this section, we use the exact variance formula to study conditions for stability (1.2) of the FDP_m. For this purpose we need a further mild assumption:

(A3) There is some $K > 0$ such that $\widehat{m}_0 \le Km$ for all $m \in \mathbb{N}$.

We want to point out that any $K > 0$, not only $K = 1$, is allowed and, hence, Assumption (A3) is not a real restriction. Clearly, (A3) is fulfilled for all generalized weighted estimators of the form (2.12) with $K = 2\sum_{i=1}^{k}(\lambda_i - \lambda_{i-1})^{-1}$. Note that (A1) and (A3) imply $\liminf_{m\to\infty} \mathrm{FDR}_m > 0$ and, hence, (2.4) is a necessary condition for stability in this case. Below, we give boundaries for the variance of $\mathrm{FDP}_m = V_m/R_m$ depending on the leading term in the variance formula of Theorem 3.1:

$$C_{m,\lambda} := \frac{\alpha}{\lambda}\mathbb{E}\Big(\frac{V_m(\lambda)}{\widehat{m}_0}\mathbb{E}\Big(\frac{1}{R_m^{(1,\lambda)}}\Big|\mathcal{F}_{\lambda,m}\Big)\Big) + \Big(\frac{\alpha}{\lambda}\Big)^2 \mathrm{Var}\Big(\frac{V_m(\lambda)}{\widehat{m}_0}\Big).$$

Lemma 4.1. *Suppose that (A2) is fulfilled.*

(a) We have

$$\mathbb{E}\Big(\frac{V_m(\lambda)}{\widehat{m}_0^2}\Big) \le \Big(\frac{\lambda}{\alpha}\Big)^2 \frac{2}{\lambda(m_0+1)} \quad \textit{and} \tag{4.1}$$

$$C_{m,\lambda} \ \ge \ Var\Big(\frac{V_m}{R_m}\Big) \ \ge \ C_{m,\lambda} \ - \frac{2}{\lambda(m_0+1)}. \tag{4.2}$$

(b) Additionally, let (A3) be fulfilled. Then $\widehat{m}_0 \le m_0 K_m$ with $K_m := Km/m_0$ and for all $t > 0$

$$P_m\Big(\mathbb{E}(V_m|\mathcal{F}_{\lambda,m}) \le t\Big) \le P_m(V_m = 0) + tD_{m,\lambda}, \ \textit{where}$$
$$D_{m,\lambda} := \frac{4K_m^2}{\alpha^2}\Big[Var\Big(\frac{V_m}{R_m}\Big) + \frac{2}{\lambda(m_0+1)} - \frac{\alpha^2}{\lambda^2} Var\Big(\frac{V_m(\lambda)}{\widehat{m}_0}\Big)\Big] + \frac{\lambda m_0 K_m}{\alpha \exp(m_0\lambda/[8(1-\lambda)])}.$$

Since $m_0 \to \infty$ under (A1), we have stability iff $C_{m,\lambda} \to 0$.

Theorem 4.2. *Under (A1)-(A3) the following* (a) *and* (b) *are equivalent.*

(a) (Stability) The convergence in (1.2) *holds.*
(b) We have

$$\frac{\widehat{m}_0}{m} - \mathbb{E}\Big(\frac{\widehat{m}_0}{m}\Big) \to 0 \ \textit{in } P_m\textit{-probability and} \tag{4.3}$$

$$R_m^{(1,\lambda)} \to \infty \ \textit{in } P_m\textit{-probability.} \tag{4.4}$$

The conditions (4.3) and (4.4) are competing. The choice $\widehat{m}_0 = m$ (BH-test) always fulfills (4.3), whereas a random $\widehat{m}_0$ may lead to more rejections, which is preferable regarding (4.4). In Example 4.3 below, we present a situation, in which FDP_m is stable when using the Storey estimator but is not stable for the BH test. But first, we want to point out that (4.3) does not imply that the estimator $\widehat{m}_0$ is consistent for m_0, i.e., $\widehat{m}_0/m_0 \to 1$ in probability. Consistent estimators only exists under strong additional assumptions, see, e.g., Genovese and Wassermann [20]. Although being not consistent, the usual (random) estimators $\widehat{m}_0$ fulfill the stability condition (4.3), see Section 5.1 for a detailed discussion concerning generalized Storey estimators.

Example 4.3. Let $m_0 = m_1$ and $U_1, U_2, \dots, U_m$ be i.i.d. uniformly distributed on $(0, 1)$. Fix $1/2 < \lambda < 1$ and define $x_0 := 1/6$. Set $p_{i,m} := U_{i,m}$ for all $i \in I_{0,m}$ and $p_{i,m} := \min\{U_i, x_0\}$ for every $i \in I_{1,m}$. The stability of the FDP_m depends on the level α:

(a) For $\alpha = 1/4$ the FDP_m is stable for the Storey procedure using the estimator (2.10) but is not stable for the BH test.
(b) For $\alpha = 1/2$ we have stability for both procedures mentioned in (a).

The results concerning the BH test can be motivated by Figure 2. For the purpose, recall that the BH procedure can also be formulated by using the Simes line $(0, 1) \ni t \mapsto f_\alpha(t) = t/\alpha$. First, the empirical distribution function $\widehat{F}_m$ of the p-values $p_{1,m}, \dots, p_{m,m}$ and the Simes line f_α are compared. Let t^* be the largest intersection of them. Then all hypotheses with $p_{i,m} \le t^*$ are rejected. By the Glivenko-Cantelli Theorem $\widehat{F}_m$ converges uniformly to F given by $F(t) = t\mathbf{1}\{t < x_0\} + 1/2(t + 1)\mathbf{1}\{t \ge x_0\}$ $(t \in [0, 1])$. It is easy to check that F and $f_{1/4}$ have a non-trivial intersection point, namely $t^* = 1/3$, whereas F and $f_{1/2}$ just intersect at $t^* = 0$. By the first observation it follows that the number of rejections tends in probability to ∞ for the BH test with $\alpha = 1/2$. In Section 8, we give a rigorous proof that this is not the case when $\alpha = 1/4$.

The previous example, in particular the part concerning the BH tests, is in line with the results of Chi and Tan [11], see their Section 4.3. For a mixture p-value model they showed for the BH test that the number of rejections remains finite for all levels α smaller than some threshold $\alpha^* \in (0, 1)$ and the number of rejections tends to ∞ for all levels $\alpha > \alpha^*$.

In the following we want to discuss the condition (4.4) more closely. Although $R_m \to \infty$ implies $R_m^{1,\lambda} \to \infty$, both in probability, the reverse is not obvious and may be false. But it is easy to see that, at least, $\mathbb{E}(R_m) \to \infty$ holds under stability: First, observe that under (A1)–(A3) stability, i.e., $\mathrm{Var}(V_m/R_m) \to 0$, implies

$$\mathbb{E}\Big(\frac{1}{R_m^{(1,\lambda)}}\Big|\mathcal{F}_{\lambda,m}\Big) \to 0 \text{ in } P_m\text{-probability.} \tag{4.5}$$

Since $R_m^{(1,\lambda)}$ is positive (4.5) implies $\mathbb{E}(R_m^{(1,\lambda)}|\mathcal{F}_{\lambda,m}) \to \infty$ in P_m-probability. Finally, we can conclude from Theorem 3.1(c) that $\mathbb{E}(V_m) \to \infty$ and, hence, $\mathbb{E}(R_m) \to \infty$ holds under stability and (A1)–(A3). Convergence in probability of R_m can be obtained if FDP_m is stable for all levels from an interval (α_1, α_2).

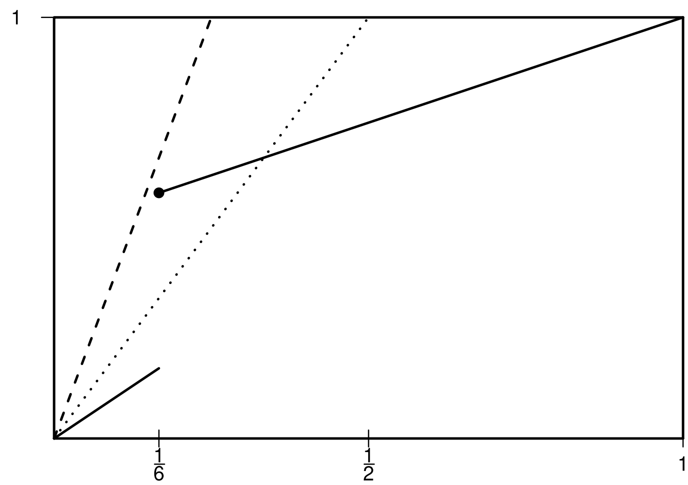

Fig 2. *Plot of the Simes lines $f_{1/4}$ (dashed) and $f_{1/2}$ (dotted) as well as of the joint distribution function F (solid) from Example 4.3.*

Before we present the corresponding theorem, we recall that V_m and R_m depend, of course, on the pre-specified level α. That is why we prefer (only) for this theorem the notation $V_{m,\alpha}$ and $R_{m,\alpha}$.

Theorem 4.4. *Suppose (A1)-(A3). Moreover, we assume that we have stability for all level $\alpha \in (\alpha_1, \alpha_2)$ and some $0 < \alpha_1 < \alpha_2 < 1$. Then we have in P_m-probability for all $\alpha \in (\alpha_1, \alpha_2)$ that*

$$V_{m,\alpha} \to \infty \text{ and, thus, } R_{m,\alpha} \to \infty.$$

5. Various stability and instability results

To avoid the ugly estimator $\widehat{m}_0 = (\alpha/\lambda) \max\{R_m(\lambda), 1\}$, which could lead to rejecting all hypotheses with $p_{i,m} \leq \lambda$, we introduce:

(A4) There exists a constant $C > 1$ with

$$\lim_{m\to\infty} P_m\Big(\widehat{m}_0 \geq \frac{C\alpha}{\lambda} \max\{R_m(\lambda), 1\}\Big) = 1.$$

Note that (A4) guarantees that (A2) holds at least with probability tending to one. The next theorem yields a necessary condition for stability.

Theorem 5.1. *Suppose that $\liminf_{m\to\infty} FDR_m > 0$ and (A4) holds.*

(a) If we have stability then $m_1 \to \infty$.
(b) Suppose that (A1) holds. If all $(p_{i,m})_{i\in I_{1,m}}$ are i.i.d. uniformly distributed on $[0, \lambda]$ then we have no stability.

By our Theorem 4.2 we already know that (4.3) and (4.4) are necessary and sufficient for stability. Turning to convergent subsequence we can assume without loss of generality under (A1), (A3) and (A4) that $\mathbb{E}(\widehat{m}_0/m)$ converges to a constant $C_0 \in [\kappa_0\alpha, K]$, say. In this case (4.3) is equivalent to

$$\frac{\widehat{m}_0}{m} \to C_0 \in [\kappa_0\alpha, K] \text{ in } P_m\text{-probability.} \tag{5.1}$$

Throughout the section's rest, we work with (5.1) instead of (4.3). If (5.1) is fulfilled it remains to verify (4.4) to obtain stability. If $R_m \to \infty$ in probability then (4.4) would follow immediately. For the BH test, for which (5.1) is obviously fulfilled with $C_0 = 1$, Ferreira and Zwinderman [15] already stated, see their Proposition 2.2, that $R_m^{BH} \to \infty$ in P_m-probability is sufficient for stability. For this purpose Ferreira and Zwinderman [15] found conditions such that $R_m/m \to \widetilde{C} > 0$ in P_m-probability. However, the sparse signal case $\kappa_0 = 1$ is more delicate since R_m/m always tends to 0 even for adaptive tests. Below, we discuss the convergence behavior of R_m more closely, in particular for the sparse case.

Lemma 5.2. *Suppose that (A1) with $\kappa_0 = 1$ and (A4) are fulfilled. Then $\widehat{\alpha}_{R_m:m} \to 0$ in P_m-probability. In particular, under (A3) we have $R_m/m \to 0$ in P_m-probability.*

As already mentioned in Example 4.3, the rejection rule for the BH test can be defined via the Simes line. The same can be done for adaptive procedures, where now the Simes line $t \mapsto f(t) =: (\widehat{m}_0/m)(t/\alpha)$ is random. Let t_m^* be the largest intersection point of $\widehat{F}_m$ and f then $\widehat{\alpha}_{R_m:m} \le t^* < \widehat{\alpha}_{R_m+1:m}$. From Lemma 5.2 we can conclude $t_m^* \to 0$ in probability when $\kappa_0 = 1$. Thus, in the sparse case the asymptotic behaviour of $\widehat{F}_m$ close to 0 is crucial.

Theorem 5.3. *Assume that (A1), (A3), (A4) and (5.1) hold. Let $\delta > 0$ and $(t_m)_{m\in\mathbb{N}}$ be some sequence in $(0, \lambda)$ such that $mt_m \to \infty$ and*

$$P_m\Big(\frac{m_1}{m}\frac{\widehat{F}_{1,m}(t_m)}{t_m} \ge \delta - \kappa_0 + \frac{1}{\alpha}C_0\Big) \to 1, \tag{5.2}$$

where $\widehat{F}_{j,m}(x) := m_j^{-1}\sum_{i\in I_{j,m}} \mathbf{1}\{p_{i,m} \le x\}$, $j \in \{1,2\}$, denotes the empirical distribution function of the p-values corresponding to the true and false null hypotheses, respectively. Then $V_m \to \infty$ in P_m-probability. In particular, stability follows from Theorem 4.2.

Remark 5.4. (a) Suppose that $(p_{i,m})_{i\in I_{1,m}}$ are i.i.d. with distribution function F_1. Then the statement of Theorem 5.3 remains valid if we replace (5.2) by

$$\frac{m_1}{m}\frac{F_1(t_m)}{t_m} \ge \delta - \kappa_0 + \frac{1}{\alpha}C_0 \text{ for all sufficiently large } m \in \mathbb{N}. \tag{5.3}$$

A detailed proof is given in Section 8.

(b) In the case $\kappa_0 = 1$ we need a sequence $(t_m)_{m\in\mathbb{N}}$ tending to 0.

(c) For the DU(m_1, m)-configuration the assumption (5.2) is fulfilled for $t_m = (m_1/m)(K+2)$ as long as the necessary condition $m_1 \to \infty$ holds.

As already stated, stability only holds under certain additional assumptions. In the following we compare stability of the classical BH test and of adaptive tests with appropriate estimators.

Lemma 5.5. *Suppose that (A1), (A3) and (A4) are fulfilled. Assume that (5.1) holds for some $C_0 \in [\alpha\kappa_0, 1]$. If $C_0 = 1$ then additionally suppose that*

$$P_m\Big(\frac{\widehat{m}_0}{m} \le 1\Big) \to 1. \tag{5.4}$$

Then stability of the BH test implies stability of the adaptive test for the same level α.

Under some mild assumptions Lemma 5.5 is applicable for the weighted estimator (2.12), see Corollary 5.6(c) for sufficient conditions.

5.1. Combination of generalized Storey estimators

In this section, we are more concrete by discussing the combined Storey estimators $\widetilde{m}_0$ introduced in (2.12). For this purpose we need the following assumption to ensure that (A4) is fulfilled.

(A5) Suppose that $\kappa_0 > \alpha(1-\kappa_0)/[\lambda(1-\alpha)]$.

Corollary 5.6. *Let (A1), (A5) and all assumptions of Theorem 2.3 be fulfilled. Consider the adaptive multiple test with $\widehat{m}_0 = \max\{\widetilde{m}_0, (\alpha/\lambda)R_m(\lambda)\}$.*

(a) Suppose that $\kappa_0 = 1$. Then (5.1) holds with $C_0 = 1$ and

$$\lim_{m\to\infty} FDR_m = \alpha = \lim_{m\to\infty} FDR_m^{BH}. \tag{5.5}$$

(b) Suppose that $\kappa_0 < 1$ and

$$(\lambda_1-\lambda_0)^{-1}\widehat{\beta}_{1,m} \le (\lambda_2-\lambda_1)^{-1}\widehat{\beta}_{2,m} \le \ldots \le (1-\lambda_{k-1})^{-1}\widehat{\beta}_{k,m} \tag{5.6}$$

with probability one for every $m \in \mathbb{N}$. Moreover, assume that

$$\liminf_{m\to\infty} \widehat{F}_m(\lambda_i) \ge \lambda_i + \varepsilon_i \text{ a.s. for some } \varepsilon_i \in [0, 1-\lambda_i] \tag{5.7}$$

and all $i = 1, \ldots, k$. If there is some $j \in \{1, \ldots, k\}$ and $\delta > 0$ such that

$$\liminf_{m\to\infty} \frac{\widehat{\beta}_{j,m}}{\lambda_j - \lambda_{j-1}} - \frac{\widehat{\beta}_{j-1,m}}{\lambda_{j-1} - \lambda_{j-2}} \ge \delta \text{ a.s. and } \varepsilon_j > 0, \tag{5.8}$$

where $\widehat{\beta}_0 := 0 =: \lambda_{-1}$, then we have an improvement of the FDR_m asymptotically compared to the Benjamini-Hochberg procedure, i.e.,

$$\liminf_{m\to\infty} FDR_m > \kappa_0\alpha = \lim_{m\to\infty} FDR_m^{BH}. \tag{5.9}$$

(c) *(Stability) Suppose that the weights are asymptotically constant, i.e., $\widehat{\beta}_{i,m} \to \beta_i$ a.s. for all $i \in \{1,\dots,k\}$, and fulfill (5.6). Assume that*

$$\lim_{m\to\infty} \widehat{F}_{1,m}(\lambda_i) = \lambda_i + \varepsilon_i \text{ a.s. for some } \varepsilon_i \in [0, 1-\lambda_i] \tag{5.10}$$

and for all $i = 1,\dots,k$. Moreover, suppose that

$$\gamma_j := \frac{\beta_j}{\lambda_j - \lambda_{j-1}} - \frac{\beta_{j-1}}{\lambda_{j-1} - \lambda_{j-2}} > 0 \text{ and } \varepsilon_j > 0 \tag{5.11}$$

for some $j \in \{1,\dots,k\}$, where $\beta_0 := \widehat{\beta}_0 := 0 =: \lambda_{-1}$. Additionally, assume $m_1/\sqrt{m} \to \infty$ if $\kappa_0 = 1$. Then (5.1) holds for some $C_0 \in [0,1]$ and stability of FDP_m for the BH test implies stability of FDP_m for the adaptive test. Moreover, if (5.2) holds for C_0 and a sequence $(t_m)_{m\in\mathbb{N}}$ with $mt_m \to \infty$ then we have always stability for the adaptive test.

It is easy to see that the assumptions of (c) imply the ones of (b). Typically, the p-values $p_{i,m}$, $i \in I_{1,m}$, from the false null are stochastically smaller than the uniform distribution, i.e., $P_m(p_{i,m} \le x) \ge x$ for all $x \in (0,1)$ (with strict inequality for some $x = \lambda_i$). This may lead to (5.7) or (5.10).

Remark 5.7. If $p_{m_0+1,m},\dots,p_{m,m}$ are i.i.d. with distribution function F_1 such that $F_1(\lambda_i) \ge \lambda_i$ for all $i = 1,\dots,k$. Then (5.7) and (5.10) are fulfilled. Moreover, if $\kappa_0 < 1$ and $F_1(\lambda_i) > \lambda_i$ then $\varepsilon_i > 0$ holds.

If the weights $\widehat{\beta}_i = \beta_i$ are deterministic then weights fulfilling (5.6) produce convex combinations of Storey estimators with different tuning parameters λ_i, compare to (2.10)-(2.11).

5.2. *Asymptotically optimal rejection curve*

Our results can be transferred to general deterministic critical values (2.1), which are not of the form (2.6) and do not use a plug-in estimator for m_0. Analogously to Section 4 and Section 9, we define $R_m^{(j)}$ for $j \in \{1,\dots,m_0\}$ by setting j p-values from true null hypotheses to 0. By the same arguments as in the proof of Theorem 3.1 we obtain

$$E\Big(\frac{V_m}{R_m}\Big) = m_0 \mathbb{E}\Big(\frac{\alpha_{R_m^{(1)}:m}}{R_m^{(1)}}\Big) \text{ and}$$
$$E\Big(\Big(\frac{V_m}{R_m}\Big)^2\Big) = m_0 \mathbb{E}\Big(\frac{\alpha_{R_m^{(1)}:m}}{(R_m^{(1)})^2}\Big) + m_0(m_0-1)\mathbb{E}\Big(\Big(\frac{\alpha_{R_m^{(2)}:m}}{(R_m^{(2)})}\Big)^2\Big).$$

The first formula can also be found in Benditkis et al. [2], see the proof of Theorem 2 therein. The proof of the second one is left to the reader. By these formulas we can now treat an important class of critical values given by

$$\alpha_{i:m} = \frac{i\alpha}{m+b-ai},\ i \le m,\ 0 \le \min(a,b). \tag{5.12}$$

A necessary condition for valid step-up tests is $\alpha_{m:m} < 1$. This condition holds for the critical values (5.12) if

$$b > 0 \text{ and } a \in [0, 1-\alpha] \text{ or } b = 0 \text{ and } a \in [0, 1-\alpha). \tag{5.13}$$

These critical values are closely related to

$$\alpha_{i:m}^{\mathrm{AORC}} = \frac{i\alpha}{m - i(1-\alpha)} = f_\alpha^{-1}\Big(\frac{i}{m}\Big),\ i < m, \tag{5.14}$$

where f_α defined by $f_\alpha(t) = t/(t(1-\alpha)+\alpha)$ is the asymptotically optimal rejection curve, which is the optimal curve in terms of asymptotic power, i.e., there is no other curve being (asymptotically) α-controlling and having a higher power, see Finner et al. [16] for more details. Note that the case $i = m$ is excluded on purpose because it would lead to $\alpha_{m:m}^{\mathrm{AORC}} = 1$. The remaining coefficient $\alpha_{m:m}^{\mathrm{AORC}}$ has to be defined separately such that $\alpha_{m-1:m}^{\mathrm{AORC}} \leq \alpha_{m:m}^{\mathrm{AORC}} < 1$, see Finner et al. [16] and Gontscharuk [21] for a detailed discussion. It is well-known that neither for (5.12) with $b = 0$ and $a > 0$ nor for (5.14) we have control of the FDR by α over all BI models simultaneously. This follows from Lemma 4.1 of Heesen and Janssen [23] since $\alpha_{1:m} > \alpha/m$. However, Heesen and Janssen [23] proved that for all fixed $b > 0$, $\alpha \in (0,1)$ and $m \in \mathbb{N}$ there exists a unique parameter $a_m \in (0, b)$ such that

$$\sup_{P_m}\{\mathrm{FDR}_{(b,a_m)}\} = \alpha,$$

where the supremum is taken over all BI models P_m at sample size m. The value a_m may be found under the least favorable configuration $\mathrm{DU}(m, m_1)$ using numerical methods.

By transferring our techniques to this type of critical values we get the following sufficient and necessary conditions for stability.

Lemma 5.8. *Suppose (A1). Let $(a_m)_{m\in\mathbb{N}}$ and $(b_m)_{m\in\mathbb{N}}$ be sequences in $\mathbb{R}$ such that $b_m/m \to 0$ and (a_m, b_m) fulfill (5.13). Now consider the step-up test with critical values given by (5.12) with $(a,b) = (a_m, b_m)$.*

(a) Then we have stability iff the following conditions (5.15)-(5.17) hold in P_m-probability:

$$R_m^{(1)} \to \infty, \tag{5.15}$$

$$\frac{a_m}{m}\Big(R_m^{(2)} - R_m^{(1)}\Big) \to 0, \tag{5.16}$$

$$\frac{m + b_m - a_m R_m^{(1)}}{m_0} - \mathbb{E}\Big(\frac{m + b_m - a_m R_m^{(1)}}{m_0}\Big) \to 0. \tag{5.17}$$

(b) If $\kappa_0 = 1$, $m_1 \to \infty$ and $\limsup_{m\to\infty} a_m < 1 - \alpha$ then (5.15) is sufficient for stability and, moreover, $FDR_m \to \alpha$ holds.

6. Least favorable configurations

Below, least favorable configurations (LFC) are derived for the p-values $(p_{i,m})_{i\in I_{1,m}}$ corresponding to false null hypotheses. Subsequently, we use "increasing" and "decreasing" in their weak form, i.e., equality is allowed, whereas other authors use "nondecreasing" and "nonincreasing" instead. When deterministic critical values $i \mapsto \alpha_{i:m}/i$ are increasing then the FDR is decreasing in each argument $p_{i,m}$, $i \in I_{1,m}$, for fixed m_1, see Benjamini and Yekutieli [6] or Benditkis et al. [2] for a short proof. In that case the Dirac uniform configuration $\mathrm{DU}(m, m_1)$, see Example 2.1, has maximum FDR, in other words it is LFC. Such least favorable configurations are useful tools for all kinds of proofs.

Remark 6.1. In contrast to (2.6), the original Storey adaptive test is based on $\widehat{\alpha}_{i:m}^{\mathrm{Stor}} = (i/\widetilde{m}_0^{\mathrm{Stor}})\alpha$ for the estimator $\widetilde{m}_0^{\mathrm{Stor}}$ from (2.9). It is known that in this situation $\mathrm{DU}(m, m_1)$ is not LFC for the FDR, see Blanchard et al. [9]. However, we will see that for our modification $\widehat{\alpha}_{i:m}^{\mathrm{Stor}} \wedge \lambda$ the $\mathrm{DU}(m, m_1)$-model is LFC.

Our exact moment formulas provide various LFC-results which are collected below. To formulate these we introduce a new assumption

(A6) For every $1 \le j \le m$ let $(p_{1,m}, \ldots, p_{m,m}) \mapsto \widehat{m}_0(p_{1,m}, \ldots, p_{m,m})$ be increasing in the coordinate $p_{j,m}$ while keeping the others fixed.

Below, we condition on $(p_{i,m})_{i\in I_{1,m}}$. Due to the independence assumption in (BI2) we may write $P_m = P_{0,m} \otimes P_{1,m}$, where $P_{j,m}$ represents the distribution of $(p_{i,m})_{i\in I_{j,m}}$, and $\mathbb{E}(X|((p_{i,m})_{i\in I_{1,m}})) = \int X((p_{i,m})_{i\le m}) \, \mathrm{d}P_{0,m}((p_{i,m})_{i\in I_{0,m}})$.

Theorem 6.2 (LFC for adaptive tests)**.** *Suppose that (A2) is fulfilled. Define the vector* $p^*_{\lambda,m} := (p_{i,m}\mathbf{1}\{p_{i,m} > \lambda\})_{i\in I_{1,m}}$.

(a) (Conditional LFC)

(i) The conditional FDR conditioned on $(p_{i,m})_{i\in I_{1,m}}$

$$\mathbb{E}\Big(\frac{V_m}{R_m}\Big|(p_{i,m})_{i\in I_{1,m}}\Big) = \mathbb{E}\Big(\frac{V_m}{R_m}\Big|p^*_{\lambda,m}\Big)$$

only depends on the portion $p_{i,m} > \lambda$, $i \in I_{1,m}$.

(ii) Conditioned on $p^*_{\lambda,m}$ *a configuration* $(p_{i,m})_{i\in I_{1,m}}$ *is conditionally Dirac uniform if* $p_{i,m} = 0$ *for all* $p_{i,m} \le \lambda$, $i \in I_{1,m}$*. The conditional variance of* V_m/R_m

$$\mathit{Var}\Big(\frac{V_m}{R_m}\Big|p^*_{\lambda,m}\Big) := \mathbb{E}\Big(\Big(\frac{V_m}{R_m}\Big)^2\Big|p^*_{\lambda,m}\Big) - \mathbb{E}\Big(\frac{V_m}{R_m}\Big|p^*_{\lambda,m}\Big)^2$$

is minimal under $DU_{cond}(m, M_{1,m}(\lambda))$*, where* $M_{1,m}(\lambda) := R_m(\lambda) - V_m(\lambda)$ *is fixed conditionally on* $p^*_{\lambda,m}$.

(b) (Comparison of different regimes $P_{1,m}$*) Under (A6) we have:*

(i) If $p_{i,m}$ *decreases for some* $i \in I_{1,m}$ *then* FDR_m *increases.*
If $p_{i,m} \le \lambda$, $i \in I_{1,m}$*, decreases then* $Var(FDP_m)$ *decreases.*

(ii) *For fixed m_1 the $DU(m, m_1)$ configuration is LFC with maximal FDR_m. Moreover, it has minimal $Var_m(FDP_m)$ for all models with $p_{i,m} \leq \lambda$ a.s. for all $i \in I_{1,m}$.*

While any deterministic convex combination of Storey estimators $\widetilde{m}_0^{\text{Stor}}(\lambda_i)$ fulfills (A6) it may fail for estimators of the form (2.11). But if the weights fulfill (5.6) then (A6) holds also for a convex combination (2.12) of these estimators. This follows from the other representation of the estimator (2.12) used in the proof of Corollary 5.6(c).

7. Discussion and summary

In this paper, we presented finite sample variance and higher moment formulas for the false discovery proportion (FDP) of adaptive step-up tests. These formulas allow a better understanding of FDP. Among others, the formulas can be used to discuss stability of FDP_m, which is preferable for application since the fluctuation and so the uncertainty vanishes. We determined a sufficient and necessary two-part condition for stability:

(i) We need a stable estimator in the sense that $\widehat{m}_0/m - \mathbb{E}(\widehat{m}_0/m)$ tends to 0 in probability.
(ii) The p-values corresponding to false null hypotheses have to be stochastically small "enough" compared to the uniform distribution such that the number of rejections tends to ∞ in probability.

Since the latter is more difficult to verify we gave a sufficient condition for it, see (5.2). This condition also applies to the sparse signal case $m_0/m \to \kappa_0 = 1$, which is more delicate than the usually studied case $\kappa_0 < 1$.

In addition to the general results we discussed data dependently weighted combinations of generalized Storey estimators. Tests based on these estimators were already discussed by Heesen and Janssen [24], who showed finite (FDR)-α-control. Heesen [22] and Heesen and Janssen [24] presented practical guidelines how to choose the data dependent weights. For our results, the additional condition (5.6) is required. We want to summarize briefly advantages of these tests in comparison to the classical BH test (see Corollary 5.6(c)):

- The adaptive tests attain (if $\kappa_0 = 1$) or even exhaust (if $\kappa_0 < 1$) the (asymptotic) FDR level $\kappa_0\alpha$ of the BH test.
- Under mild assumptions stability of FDP_m for the BH test always implies stability of FDP_m for the adaptive test.

In Section 5.2 we explained that our results can also be transferred to general deterministic critical values $\alpha_{i:m}$, which are not based on plug-in estimators of m_0. The same should be possible for general random critical values under appropriate conditions. Due to lack of space, we leave a discussion about other estimators for future research.

8. Proofs

8.1. Proof of Theorem 3.1

To improve the readability of the proof, all indices m are omitted, i.e., we write p_i instead of $p_{i,m}$ etc. First, we determine $\mathbb{E}(\mathrm{FDP}^2|\mathcal{F}_\lambda)$. Without loss of generality we can assume conditioned on $\mathcal{F}_\lambda$ that the first $V(\lambda)$ p-values correspond to the true null and $p_1,\ldots,p_{V(\lambda)}\leq\lambda$. In particular, we may consider $p^{(1)}=(0,p_2,p_3,\ldots,p_m)$ and $p^{(2)}=(0,0,p_3,\ldots,p_m)$ if $V(\lambda)\geq 1$ and $V(\lambda)\geq 2$, respectively. Since $\widehat{\alpha}_{R:m}\leq\lambda$ we deduce from (BI3) that

$$E\Big(\Big(\frac{V}{R}\Big)^2\Big|\mathcal{F}_\lambda\Big)=V(\lambda)\mathbb{E}\Big(\frac{\mathbf{1}\{p_1\leq\widehat{\alpha}_{R:m}\}}{R^2}\Big|\mathcal{F}_\lambda\Big)$$
$$+V(\lambda)(V(\lambda)-1)\mathbb{E}\Big(\frac{\mathbf{1}\{p_1\leq\widehat{\alpha}_{R:m},p_2\leq\widehat{\alpha}_{R:m}\}}{R^2}\Big|\mathcal{F}_\lambda\Big).$$

Note that $p_1,\ldots,p_{V(\lambda)}$ conditioned on $\mathcal{F}_\lambda$ are i.i.d. uniformly distributed on $(0,\lambda)$ if $V(\lambda)>0$. It is easy to see that $p_1\leq\widehat{\alpha}_{R:m}$ implies $R=R^{(1,\lambda)}$ and, thus, $P(p_1\in(\widehat{\alpha}_{R:m},\widehat{\alpha}_{R^{(1,\lambda)}:m}])=0$. Both were already known and used, for instance, in Heesen and Janssen [23, 24]. Since p_1 and $R^{(1,\lambda)}$ are independent conditionally on $\mathcal{F}_\lambda$ we obtain from Fubini's Theorem that

$$\mathbb{E}\Big(\frac{\mathbf{1}\{p_1\leq\widehat{\alpha}_{R:m}\}}{R^2}\Big|\mathcal{F}_\lambda\Big)=\mathbb{E}\Big(\frac{\mathbf{1}\{p_1\leq\widehat{\alpha}_{R^{(1,\lambda)}:m}\}}{(R^{(1,\lambda)})^2}\Big|\mathcal{F}_\lambda\Big)=\frac{\alpha}{\lambda\widehat{m}_0}\mathbb{E}\Big(\frac{1}{R^{(1,\lambda)}}\Big|\mathcal{F}_\lambda\Big).$$

Hence, we get the second summand of the right-hand side in (a). To obtain the first term, it is sufficient to consider $V(\lambda)\geq 2$. Since p_1, p_2 and $R^{(2,\lambda)}$ are independent conditionally on $\mathcal{F}_\lambda$ we get similarly to the previous calculation:

$$E\Big(\frac{\mathbf{1}\{p_1\leq\widehat{\alpha}_{R:m},p_2\leq\widehat{\alpha}_{R:m}\}}{R^2}\Big|\mathcal{F}_\lambda\Big)\tag{8.1}$$
$$=E\Big(\frac{\mathbf{1}\{p_1\leq\widehat{\alpha}_{R^{(2,\lambda)}:m},p_2\leq\widehat{\alpha}_{R^{(2,\lambda)}:m}\}}{(R^{(2,\lambda)})^2}\Big|\mathcal{F}_\lambda\Big)=\frac{\alpha^2}{\lambda^2}\frac{1}{\widehat{m}_0^2},\tag{8.2}$$

which completes the proof of (a). Combining (a), (2.8) and the variance formula $\mathrm{Var}(Z)=\mathbb{E}(Z^2)-\mathbb{E}(Z)^2$ yields (b). The proof of (c) is based on the same techniques as the one of (a), to be more specific:

$$\mathbb{E}(V|\mathcal{F}_{\lambda,m})=V(\lambda)P(p_1\leq\widehat{\alpha}_{R^{(1,\lambda)}:m}\mid\mathcal{F}_\lambda)=V(\lambda)\mathbb{E}\Big(\frac{R^{(1,\lambda)}\alpha}{\lambda\widehat{m}_0}\Big|\mathcal{F}_\lambda\Big).\tag{8.3}$$

8.2. Proof of Lemma 4.1

To improve the readability, all indices m are omitted except for K_m.

(a): By Theorem 3.1(b) and (A2) it remains to show that

$$\Big(\frac{\alpha}{\lambda}\Big)^2\mathbb{E}\Big(\frac{V(\lambda)}{\widehat{m}_0^2}\Big)\leq\mathbb{E}\Big(\frac{V(\lambda)}{R(\lambda)^2}\Big)\leq\mathbb{E}\Big(\frac{\mathbf{1}\{V(\lambda)>0\}}{V(\lambda)}\Big)$$

is smaller than $2/(\lambda(m_0+1))$. It is known and can easily be verified that $\mathbb{E}(\mathbf{1}\{X>0\}X^{-1}) \le 2\mathbb{E}((1+X)^{-1}) \le 2p^{-1}(n+1)^{-1}$ for any Binomial-distributed $X \sim B(n,p)$. From this and $V(\lambda) \sim B(m_0,\lambda)$ we obtain the desired upper bound, see also p. 47ff of Heesen and Janssen [24] for details.

(b): We can deduce from (8.3) that

$$Y_\lambda := \frac{\mathbf{1}\{V(\lambda)>0\}}{\mathbb{E}(V|\mathcal{F}_\lambda)} = \frac{\lambda}{\alpha}\frac{\widehat{m}_0}{V(\lambda)}\frac{\mathbf{1}\{V(\lambda)\ge 1\}}{\mathbb{E}(R^{(1,\lambda)}|\mathcal{F}_\lambda)}.$$

Note that

$$P\Big(\mathbb{E}(V|\mathcal{F}_\lambda) \le t\Big) \le P(V=0) + P\Big(Y_\lambda \ge \frac{1}{t}\Big).$$

Thus, by Markov's inequality it remains to verify $\mathbb{E}(Y_\lambda) \le D_\lambda$. We divide the discussion of $\mathbb{E}(Y_\lambda)$ into two parts. First, we use Hoeffding's inequality

$$P\Big(\frac{X-np}{\sqrt{n}} \le -t\Big) \le \exp\Big(-\frac{t^2}{2p(1-p)}\Big)$$

for $X \sim B(n,p)$ and all $t>0$ [37, p.440]. Since $V(\lambda) \sim B(m_0,\lambda)$ we obtain

$$P\Big(V(\lambda) \le \frac{m_0\lambda}{2}\Big) = P\Big(\frac{V(\lambda)-m_0\lambda}{\sqrt{m_0}} \le -\frac{\sqrt{m_0}\lambda}{2}\Big) \le \exp\Big(-\frac{m_0\lambda}{8(1-\lambda)}\Big).$$

Second, we can conclude from Jensen's inequality and Theorem 3.1(b) that

$$\begin{aligned}\mathbb{E}\Big(Y_\lambda \mathbf{1}\Big\{\frac{V(\lambda)}{m_0} \ge \frac{\lambda}{2}\Big\}\Big) &\le \frac{\lambda}{\alpha}\Big(\frac{2K_m}{\lambda}\Big)^2 \mathbb{E}\Big(\frac{V(\lambda)}{\widehat{m}_0}\mathbb{E}\Big(\frac{1}{R^{(1,\lambda)}}|\mathcal{F}_\lambda\Big)\Big)\\ &\le \Big(\frac{2K_m}{\lambda}\Big)^2\Big[\frac{\lambda^2}{\alpha^2}\mathrm{Var}\Big(\frac{V}{R}\Big) + \mathbb{E}\Big(\frac{V(\lambda)}{\widehat{m}_0^2}\Big) - \mathrm{Var}\Big(\frac{V(\lambda)}{\widehat{m}_0}\Big)\Big].\end{aligned}$$

Finally, combining this with (4.1) yields the statement.

8.3. Proof of Theorem 4.2

Since V_m/R_m is bounded by 1 the stability statement in (a) is equivalent to $\mathrm{Var}(V_m/R_m) \to 0$. Due to $V_m(\lambda)/m \to \kappa_0\lambda$ a.s. and $K \ge (\widehat{m}_0/m) \ge (\alpha/\lambda)(V_m(\lambda)/m) \to \alpha\kappa_0$ a.s. we deduce from Lemma 4.1 that $\mathrm{Var}(\mathrm{FDP}_m) \to 0$ is equivalent to $\mathrm{Var}(V_m(\lambda)/\widehat{m}_0) \to 0$ and $\mathbb{E}((R_m^{(1,\lambda)})^{-1}|\mathcal{F}_{\lambda,m}) \to 0$ in P_m-probability. Moreover, we can conclude that $\mathrm{Var}(V_m(\lambda)/\widehat{m}_0) \to 0$ is equivalent to (4.3). Since $R_m^{(1,\lambda)} \ge 1$ we have equivalence of $\mathbb{E}((R_m^{(1,\lambda)})^{-1}|\mathcal{F}_{\lambda,m}) \to 0$ in P_m-probability and (4.4).

8.4. Proof of Example 4.3

The statements concerning the Storey procedure follow immediately from Corollary 5.6(c). It remains to verify that FDP_m is not stable for the BH test when the underlying level is $\alpha := 1/4$. In this case the Simes line is given by $t \mapsto f_\alpha(t) = 4t$. Clearly, the Simes line lies strictly above F, the uniform limit of $\widehat{F}_m$, on $(0,1)$, see also Figure 2. Hence, $P_m(\sup_{\varepsilon \le x \le 1} \widehat{F}_m(x) - f_\alpha(x) < 0) \to 1$ for all $\varepsilon > 0$. Let $0 < \lambda_0 < x_0$. Then $P_m(\alpha_{R_m:m} \le \lambda_0) \to 1$ follows. That is why we can restrict our asymptotic considerations to the portion of p-values with $p_{i,m} \le \lambda_0$ and the instability follows analogously to the proof of Theorem 5.1(b).

8.5. Proof of Theorem 4.4

At long an appropriate subsequence $n(m) \to \infty$ we can always obtain

$$\lim_{m\to\infty} \mathbb{E}\Big(\frac{V_{n(m)}(\lambda)}{\widehat{m}_0}\Big) \to C \in \Big[\frac{1}{K}, \frac{\lambda}{\alpha}\Big].$$

We suppose, contrary to our claim, that $V_{m,\alpha}$ does not converge to ∞ in P_m-probability for some $\alpha \in (\alpha_1, \alpha_2)$. Since $\alpha \mapsto V_{m,\alpha}$ is increasing we can suppose without loss of generality that $\lambda^{-1}\alpha C \notin \mathbb{Q}$ (otherwise take a smaller $\alpha > \alpha_1$). By our contradiction assumption there is some $k \in \mathbb{N} \cup \{0\}$ and a subsequence of $\{n(m) : m \in \mathbb{N}\}$, which we denote by simplicity also by $n(m)$, with $n(m) \to \infty$ such that $P_{n(m)}(V_{n(m),\alpha} = k) \to \beta \in (0,1]$. We can deduce from (2.8) and the stability that

$$(V_{n(m),\alpha}/R_{n(m),\alpha})\mathbf{1}\{V_{n(m),\alpha} = k\} - (\alpha/\lambda)C\mathbf{1}\{V_{n(m),\alpha} = k\} \to 0$$

in $P_{n(m)}$-probability. In particular,

$$P_{n(m)}\Big(R_{n(m),\alpha} = \frac{k\lambda}{C\alpha}, V_{n(m),\alpha} = k\Big) \to \beta > 0,$$

which leads to a contradiction since $(\lambda k)/(C\alpha) \notin \mathbb{N} \cup \{0\}$.

8.6. Proof of Theorem 5.1

(a): We suppose contrary to our claim that $\liminf_{m\to\infty} m_1 = k \in \mathbb{N} \cup \{0\}$. Then $m_1 = k$ for infinitely many $m \in \mathbb{N}$. Turning to subsequences we can assume without a loss that $m_1 = k$ for all $m \in \mathbb{N}$. Note that (A1) holds for $\kappa_0 = 1$ in this case. Hence, it is easy to see that $\liminf_{m\to\infty} R_m(\lambda)/m \ge \lambda$ a.s. Combining this and (A4) yields

$$P_m\Big(\widehat{\alpha}_{i:m} \le \Big(\frac{i}{m}\widetilde{\alpha}\Big) \text{ for all } i = 1, \dots, m\Big) \to 1 \text{ with } \widetilde{\alpha} = \frac{1}{C} < 1.$$

Hence, we can deduce from Example 2.1(a) that

$$\liminf_{m\to\infty} P(V_m = 0) \ge \liminf_{m\to\infty} P_m(V_m^{\mathrm{BH}}(\widetilde{\alpha}, k) = 0) > 0.$$

But this contradicts the necessary condition (2.4) for stability.

(b): Suppose for a moment that we condition on $\mathcal{F}_{\lambda,m}$. Hence, $R_m(\lambda)$ and $\widehat{m}_0$ can be treated as fixed numbers. Without loss of generality we assume that $p_{1,m},\dots,p_{R_m(\lambda),m} \le \lambda$. Define new p-values $q_{1,R_m(\lambda)},\dots,q_{R_m(\lambda),R_m(\lambda)}$ by $q_{i,R_m(\lambda)} := p_{i,m}/\lambda$ for all $i = 1,\dots,R_m(\lambda)$. The values V_m and R_m are the same for the step-up test based on the p-values $(p_{i,m})_{i\le m}$ with critical values $\widehat{\alpha}_{i:m} = (i/\widehat{m}_0)\alpha$ as well as for the one based on $(q_{i,R_m(\lambda)})_{i\le R_m(\lambda)}$ with critical values $\widehat{\alpha}^{(q)}_{i:R_m(\lambda)} = (i/R_m(\lambda))\widetilde{\alpha}_m$, where $\widetilde{\alpha}_m = (R_m(\lambda)/\widehat{m}_0)(\alpha/\lambda)$. Here, $q_{1,R_m(\lambda)},\dots,q_{R_m(\lambda),R_m(\lambda)}$ are i.i.d. uniformly distributed on $(0,1)$ and, thus, they correspond to a $\mathrm{DU}(R_m(\lambda),0)$-configuration. That is why

$$P_m(V_m = 0|\mathcal{F}_{\lambda,m}) = P_m(V^{\mathrm{BH}}_{R_m(\lambda)}(\widetilde{\alpha}_m, 0) = 0|\mathcal{F}_{\lambda,m}). \tag{8.4}$$

By the strong law of large numbers and (A4) we have $R_m(\lambda) \to \infty$ a.s. and $P_m(\widetilde{\alpha}_m \le C^{-1}) \to 1$. Hence, we can conclude from (8.4) and Example 2.1(a)

$$\liminf_{m\to\infty} P_m(V_m = 0) \ge (1 - C^{-1})$$

and, consequently, the necessary condition (2.4) for stability is not fulfilled.

8.7. *Proof of Lemma 5.2*

Analogously to the proof of Theorem 5.1(b), we condition under $\mathcal{F}_{\lambda,m}$ and introduce the new p-value $q_{i,R_m(\lambda)}$ and the new critical value $\widehat{\alpha}^{(q)}_{i:R_m(\lambda)}$ for $i \le R_m(\lambda)$ as well as the new level $\widetilde{\alpha}_m$. The respective empirical distribution functions of the new p-values, $(q_{i,m})_{i\le R_m(\lambda)}$ are denoted by $\widehat{F}^{(q)}_{R_m(\lambda)}, \widehat{F}^{(q)}_{0,R_m(\lambda)}, \widehat{F}^{(q)}_{1,R_m(\lambda)}$, compare to the definition of $\widehat{F}_{j,m}$ in Theorem 5.3. Note that $R_m(\lambda) \ge m_0\widehat{F}_{0,m}(\lambda) \to \infty$ P_m-a.s. and, hence, by (A4) $P_m(\widetilde{\alpha}_m \le C^{-1}) \to 1$. From this and the Glivenko-Cantelli Theorem we obtain that for all $\varepsilon \in (0,1)$

$$P_m\Big(\sup_{t\in[\varepsilon,1]} \widehat{F}^{(q)}_{0,R_m(\lambda)}(t) - f_{\widetilde{\alpha}_m}(t) \le -\frac{1}{2}(C-1)\varepsilon \Big| \mathcal{F}_{\lambda,m}\Big) \to 1, \tag{8.5}$$

where $t \mapsto f_{\widetilde{\alpha}_m}(t) =: t/\widetilde{\alpha}_m$ is the corresponding Simes line, see also Example 4.3. Recall that $\widehat{\alpha}^{(q)}_{R_m:R_m(\lambda)}$ is smaller than the largest intersection point of $\widehat{F}^{(q)}_m$ and $f_{\widetilde{\alpha}_m}$. Combining this, (8.5) and $|\widehat{F}^{(q)}_{R_m(\lambda)}(t) - \widehat{F}^{(q)}_{0,R_m(\lambda)}(t)| \le 2m_1/m \to 0$ we can deduce $P_m(\widehat{\alpha}_{R_m:m} \le \lambda\varepsilon) \to 1$ for all $\varepsilon > 0$.

8.8. *Proof of Theorem 5.3*

Clearly, all $p_{i,m} \le t_m$ are rejected and, in particular, $V_m \ge V_m(t_m)$ if $p_{R_m(t_m):m} \le (R_m(t_m)/\widehat{m}_0)\alpha$. The latter is fulfilled if $(t_m/\alpha)\widehat{m}_0 \le R_m(t_m)$, or equivalently

$$\frac{\widehat{m}_0}{m\alpha} \le \frac{m_0}{m}\frac{\widehat{F}_{0,m}(t_m)}{t_m} + \frac{m_1}{m}\frac{\widehat{F}_{1,m}(t_m)}{t_m}. \tag{8.6}$$

By Chebyshev's inequality

$$P_m\Big(\frac{m_0}{m}\frac{\widehat{F}_{0,m}(t_m)}{t_m} \geq \kappa_0 - \frac{1}{2}\delta\Big) \geq 1 - \frac{1}{mt_m}\Big(\frac{1}{2}\delta + \frac{m_0}{m} - \kappa_0\Big)^{-2} \to 1.$$

Combining this, (5.1), (5.2) and (8.6) yields

$$P_m(V_m \geq V_m(t_m)) \to 1.$$

Finally, the statement follows from $V_m(t_m) \sim B(m_0, t_m)$ and $m_0 t_m \to \infty$.

8.9. *Proof of Remark 5.4*

By Theorem 5.3 it remains to show that

$$\begin{aligned}
&P_m\Big(\frac{m_1}{m}\frac{\widehat{F}_{1,m_1}(t_m)}{t_m} \geq \frac{1}{2}\delta - \kappa_0 + \frac{1}{\alpha}C_0\Big) \\
&= P_m\Big(\sqrt{m_1}\frac{\widehat{F}_{1,m_1}(t_m) - F_1(t_m)}{\sqrt{F_1(t_m)(1-F_1(t_m))}} \geq \sqrt{m_1}t_m \frac{\frac{m}{m_1}(\frac{\delta}{2} - \kappa_0 + \frac{1}{\alpha}C_0 - \frac{m_1}{m}\frac{F_1(t_m)}{t_m})}{\sqrt{F_1(t_m)(1-F_1(t_m))}}\Big)
\end{aligned}$$

converges to 1. Note that the left-hand side of the last row converges in distribution to $Z \sim N(0,1)$. Moreover, by straightforward calculations it can be concluded from (5.3) and $C_0 \geq \kappa_0\alpha$ that the right-hand side tends to $-\infty$, which completes the proof.

8.10. *Proof of Lemma 5.5*

It is easy to see that (5.4) always holds if $C_0 < 1$. From (5.4) we obtain immediately that

$$\begin{aligned}
&P_m\Big(\max_{i=1,\dots,m}\Big\{\widehat{\alpha}_{i,m}^{\mathrm{BH}} - \widehat{\alpha}_{i,m}\Big\} \leq 0\Big) \leq P_m\Big(\frac{1}{m} - \frac{1}{\widehat{m}_0} \leq 0\Big) \to 1 \\
&\text{and so } P_m\Big(R_m^{(1,\lambda)} \geq R_m^{(1,\lambda),\mathrm{BH}}\Big) \to 1,
\end{aligned}$$

where $R_m^{(1,\lambda),\mathrm{BH}}$ is the corresponding random variable for the BH test. Now, suppose that we have stability of FDP_m for the BH test. Then combining Theorem 4.2 with the above yields that $R_m^{(1,\lambda),\mathrm{BH}}$ and so $R_m^{(1,\lambda)}$ converges to infinity in P_m-probability. Finally, we deduce the stability for the adaptive test from Theorem 4.2.

8.11. *Proof of Corollary 5.6*

(a): Clearly, $\widetilde{m}_0(\lambda_{i-1}, \lambda_i)/m \to 1$ a.s. for all $i = 1, \dots, k$ and $R_m(\lambda)/m \to \lambda$ a.s. Thus, (5.1) holds for $C_0 = 1$. Finally, (5.5) follows from (2.8).

(b): First, we introduce new estimators $\widetilde{m}_{0,i}$ and new weights $\widehat{\gamma}_{i,m} \geq 0$:

$$\widetilde{m}_{0,i} := m\frac{1-\widehat{F}_m(\lambda_{i-1})-\frac{i}{m}}{1-\lambda_{i-1}} \text{ and } \widehat{\gamma}_i := \Big(\frac{\widehat{\beta}_{i,m}}{\lambda_i-\lambda_{i-1}} - \frac{\widehat{\beta}_{i-1,m}}{\lambda_{i-1}-\lambda_{i-2}}\Big)(1-\lambda_{i-1}),$$

where $\widehat{\beta}_{0,m} := 0$. It is easy to check $\widetilde{m}_0 = \sum_{i=1}^k \widehat{\gamma}_{i,m}\widetilde{m}_{0,i}$ and $\sum_{i=1}^k \widehat{\gamma}_{i,m} = 1$. From (5.7) and the strong law of large numbers it follows

$$\frac{V_m(\lambda)}{m} \to \kappa_0\lambda \text{ a.s. and } \limsup_{m\to\infty} \frac{\widetilde{m}_{0,i}}{m} \leq 1 - \frac{\varepsilon_i}{1-\lambda_i} \text{ a.s.} \tag{8.7}$$

In particular, by (5.8)

$$\limsup_{m\to\infty} \frac{\widetilde{m}_0}{m} \leq 1 - \frac{\delta(1-\lambda_{j-1})\varepsilon_j}{1-\lambda_j} \leq \frac{1}{1+\delta_0} \text{ a.s.}$$

for some $\delta_0 > 0$. Consequently,

$$\liminf_{m\to\infty} \frac{V_m(\lambda)}{\widetilde{m}_0} \geq \lambda\kappa_0(1+\delta_0) \text{ a.s.}$$

It is easy to verify that (A5) implies $P_m(\widetilde{m}_0(\lambda_{i-1},\lambda_i) > C_i(\alpha/\lambda)R_m(\lambda)) \to 1$ for appropriate $C_i > 1$ and for all i. Hence, (A4) is fulfilled and, in particular, $P_m(\widehat{m}_0 = \widetilde{m}_0) \to 1$. Finally, we obtain the statement from (2.8).

(c): Define $\widetilde{m}_{0,i}$ and $\widehat{\gamma}_{i,m}$ as in the proof of (b). Then

$$\widehat{\gamma}_i \to \frac{\beta_i}{\lambda_i-\lambda_{i-1}} - \frac{\beta_{i-1}}{\lambda_{i-1}-\lambda_{i-2}} =: \gamma_i \text{ a.s.}$$

for all $i = 1,\ldots,k$. Clearly, (A3) and (A4) are fulfilled, see for the latter the end of the proof of (b). Moreover, (5.1) holds for some $C_0 \in [0,1]$ since

$$\frac{\widehat{m}_0}{m} \to 1-(1-\kappa_0)\sum_{i=1}^k \frac{\gamma_i\varepsilon_i}{1-\lambda_i} \text{ a.s.}$$

Due to (5.11) we have $C_0 < 1$ iff $\kappa_0 < 1$. Consequently, by Theorem 5.3 and Lemma 5.5 it remains to verify (5.4) in the case of $\kappa_0 = 1$.

Consider $\kappa_0 = 1$. Keep in mind that $m_1/\sqrt{m} \to \infty$ in this case. For all $i = 1,\ldots,k$ we can deduce from the central limit theorem

$$Z_{i,m} := \sqrt{m}\frac{m_0}{m}\Big(\frac{1-\widehat{F}_{0,m}(\lambda_i)}{1-\lambda_i} - 1\Big) \xrightarrow{\text{d}} Z_i \sim N(0,\sigma_i^2) \tag{8.8}$$

for some $\sigma_i \in (0,\infty)$. Let $\xi := \varepsilon_j/(8(1-\lambda_j)) > 0$. By (5.10) and (5.11)

$$P_m\Big(\frac{1-\widehat{F}_{1,m}(\lambda_j)}{1-\lambda_j} \leq 1-4\xi\Big) \to 1 \tag{8.9}$$

$$\text{and } P_m\Big(\frac{1-\widehat{F}_{1,m}(\lambda_i)}{1-\lambda_i} \le 1+\frac{\xi}{2}\gamma_j\Big) \to 1. \tag{8.10}$$

for all $i \in \{1,\dots,k\} \setminus \{j\}$. Moreover, from (8.8) we get

$$\begin{aligned}
&P_m\Big(\frac{m_0}{m}\frac{1-\widehat{F}_{1,m}(\lambda_j)+\frac{1}{m}}{1-\lambda_j}+\frac{m_1}{m}\Big(1-4\xi+\frac{\frac{1}{m}}{1-\lambda_j}\Big) \le 1-\frac{m_1}{m}2\xi\Big)\\
&= P_m\Big(Z_{i,m} \le \frac{m_1}{\sqrt{m}}2\xi - \frac{1}{\sqrt{m}}\frac{1}{1-\lambda_i}\Big) \to 1.
\end{aligned}$$

By this and (8.9) $P_m(\widetilde{m}_{0,j} \le 1-(m_1/m)2\xi) \to 1$ follows. Analogously, we obtain from (8.10) that $P_m(\widetilde{m}_{0,i} \le 1+(m_1/m)\gamma_j\xi) \to 1$ for all $i \neq j$. Since $\sum_{i=1,i\neq j}^k \widehat{\gamma}_i \le 1$ and $P_m(2\widehat{\gamma}_j \ge \gamma_j) \to 1$ we can finally conclude (5.4).

8.12. Proof of Lemma 5.8

(a): First, we introduce for $j=1,2$:

$$\psi_{m,j} := \frac{m_0\alpha}{m+b_m-a_mR_m^{(j)}}. \tag{8.11}$$

Using the formulas presented at the beginning of Section 5.2 we obtain:

$$\mathrm{Var}\Big(\frac{V_m}{R_m}\Big) = \mathbb{E}\Big(\frac{1}{R_m^{(1)}}\psi_{m,1}\Big) + \mathrm{Var}(\psi_{m,1}) + \mathbb{E}\Big(\psi_{m,2}^2-\psi_{m,1}^2\Big) - \frac{1}{m_0}\mathbb{E}(\psi_{m,2})^2.$$

From $R_m^{(j)} \le m$, $0 \le b_m \le m$ for large m and (5.12) we get:

$$\kappa_0 \leftarrow \frac{m_0}{m+b_m} \le \psi_{m,j} \le \frac{m_0}{\alpha m} \to \frac{\kappa_0}{\alpha}. \tag{8.12}$$

Hence, the fourth summand $-\mathbb{E}(\psi_{m,2})^2/m_0$ in the formula for $\mathrm{Var}(V_m/R_m)$ tends always to 0. Since, clearly, the first three summands are non-negative it remains to show that each of these summands tends to 0 iff our conditions (5.15)-(5.17) are fulfilled. By (8.12) we have equivalence of (5.17) and $\mathrm{Var}(\psi_{m,1}) \to 0$, as well as of (5.15) and $\mathbb{E}(\psi_{m,1}/R_m^{(1)}) \to 0$. Observe that $\psi_{m,2}-\psi_{m,1} = Z_m\psi_{m,1}\psi_{m,2}$ with $Z_m := (a_m/m_0)(R_m^{(2)}-R_m^{(1)})$. From (8.12) and $0 \le Z_m \le 1-\alpha$ we obtain that $\mathbb{E}(\psi_{m,2}-\psi_{m,1}) \to 0$ iff (5.16) holds. Finally, combining this, (8.12), $\psi_{m,2}^2-\psi_{m,1}^2 = (\psi_{m,2}-\psi_{m,1})(\psi_{m,2}+\psi_{m,1})$ and $\psi_{m,2} \ge \psi_{m,1}$ yields that $\mathbb{E}(\psi_{m,2}^2-\psi_{m,1}^2) \to 0$ iff (5.16) is fulfilled.

(b): Similarly to Lemma 5.2, we obtain by considering the (least favorable) $\mathrm{DU}(m, m_1+j)$-configuration that $\alpha_{R_m^{(j)}:m} \to \infty$ and so $R_m^{(j)}/m \to \infty$ for $j \in \{1,2\}$, both in P_m-probability. Clearly, (5.16) follows. Moreover, we deduce from this, $a_m \le 1-\alpha$ and $b_m/m \to 0$ that $\psi_{m,j}$ defined by (8.11) converges to α in P_m-probability for $j=1,2$. This implies (5.17) and $\mathbb{E}(V_m/R_m) = \mathbb{E}(\psi_{m,1}) \to \alpha$. In particular, we have stability by (a).

8.13. Proof of Theorem 6.2

(ai): Let $p_i^* \in [0,1]$ be fixed for each $i \in I_{1,m}$. Let P_m^* be the distribution fulfilling BI, where $p_{i,m} \equiv p_i^*$ a.s. for all $i \in I_{1,m}$. From (2.8) we get

$$\int \frac{V_m}{R_m} \,\mathrm{d}P_m^* = \frac{\alpha}{\lambda} \mathbb{E}\Big(\frac{V_m(\lambda)}{\widehat{m}_0} \Big).$$

Moreover, we observe that the right-hand side only depends on p_i^*, $i \in I_{1,m}$, if $p_i^* > \lambda$. Consequently, we obtain the statement.

(aii): Due to (ai) it remains to show that the conditional second moment is minimal under $\mathrm{DU}_{\mathrm{cond}}(m, M_{1,m}(\lambda))$. Clearly, BI and (A2) are also fulfilled conditioned on $p_{\lambda,m}^*$. From Theorem 3.1(a) we obtain

$$\mathbb{E}\Big(\Big(\frac{V_m}{R_m} \Big)^2 \Big| p_{\lambda,m}^* \Big) = \mathbb{E}\Big(\frac{\alpha^2 V_m(\lambda)(V_m(\lambda)-1)}{\lambda^2 \widehat{m}_0^2} + \frac{\alpha}{\lambda} \frac{V_m(\lambda)}{\widehat{m}_0} \frac{1}{R_m^{(1,\lambda)}} \Big| p_{\lambda,m}^* \Big).$$

It is easy to see that $V_m(\lambda)$ and $\widehat{m}_0$ are not affected and $R_m^{(1,\lambda)}$ increase if we set all $M_{1,m}(\lambda)$ p-value $p_{i,m} \le \lambda$, $i \in I_{n,1}$, to 0.

(bi): Since $V_m(\lambda)$ is not affected by any $p_{i,m}$, $i \in I_{1,m}$, the first statement follows from (A6) and (2.8). If $p_{i,m} \le \lambda$, $i \in I_{1,m}$, decreases than $V_m(\lambda)$ and $\widehat{m}_0$ are not affected, and R_m as well as $R_m^{(1,\lambda)}$ increase. Hence, the second statement follows from Theorem 3.1(b).

(bii): The statement follows immediately from (bi).

9. Appendix: Higher moments

We extend the idea of the definition of $p_m^{(1,\lambda)}$ and $R_m^{(1,\lambda)}$ from Section 4. For every $1 < j \le m_0$ we introduce a new p-value vector $p_m^{(j)}$ as a modification of $p_m = (p_{1,m}, \ldots, p_{m,m})$ iteratively. If $V_m(\lambda) \ge j$ then we define $p_m^{(j,\lambda)}$ by setting $p_{i_k,m}$ equal to 0 for j different indices $i_1, \ldots, i_j \in I_{0,m}$ with $p_{i_k,m} \le \lambda$, for convenience take the smallest j indices with this property. Otherwise, if $V_m(\lambda) < j$ then set $p_m^{(j,\lambda)}$ equal to $p_m^{(j-1,\lambda)}$. Moreover, let $R_m^{(j,\lambda)} = R_m^{(j,\lambda)}(p_m^{(j,\lambda)})$ be the number of rejections of the adaptive test for the (new) p-value vector $p_m^{(j,\lambda)}$. Note that $\widehat{m}_0$ is not affected by these replacements.

Theorem 9.1. *Under (A2) we have for every $k \le m$*

$$\mathbb{E}\Big(\Big(\frac{V_m}{R_m} \Big)^k \Big) = \sum_{j=1}^{k} \alpha^j C_{j,k} \mathbb{E}\Big(\frac{V_m(\lambda) \ldots (V_m(\lambda)-j+1)}{(\widehat{m}_0)^j} \mathbb{E}\Big(\Big(R_m^{(j,\lambda)} \Big)^{j-k} \Big| \mathcal{F}_{\lambda,m} \Big) \Big),$$

where $C_{j,k} = \dfrac{1}{j!} \displaystyle\sum_{r=0}^{j-1} (-1)^r \binom{j}{r} (j-r)^k$.

Remark 9.2. (a) If we set $\widehat{m}_0 = m_0$ and $\lambda = 1$ then this formula coincide up to the factor $C_{j,k}$ with the result of Ferreira and Zwinderman [15]. By carefully reading their proof it can be seen that the coefficients $C_{j,k}$ have to be added. It is easy to check that $C_{1,k} = C_{k,k} = 1$ but $C_{j,k} > 1$ for all $1 < j < k$. In particular, the coefficients $C_{1,1}$, $C_{1,2}$, $C_{2,2}$, which are needed for the variance formula, are equal to 1.

(b) For treating one-sided null hypotheses the assumption (BI3) need to be extended to i.i.d. $(p_{i,m})_{i\in I_{0,m}}$ p-values of the true null hypothesis, which are stochastically larger than the uniform distribution, i.e., $P(p_{i,m} \leq x) \leq x$ for all $x \in [0,1]$. In this case the equality in Theorem 9.1 is not valid in general but the statement remains true if "=" is replaced by "≤", analogously to the results of Ferreira and Zwinderman [15].

Proof of Theorem 9.1. For the proof we extend the ideas of the proof of Theorem 3.1. In particular, we condition on $\mathcal{F}_{\lambda,m}$. First, observe that

$$\mathbb{E}\Big(\frac{V_m^k}{R_m^k}\Big|\mathcal{F}_{\lambda,m}\Big) = \sum_{i_1,\ldots,i_k=1}^{V_m(\lambda)} \mathbb{E}\Big(\frac{\mathbf{1}\{p_{i_s,m} \leq \widehat{\alpha}_{R_m:m},\, s \leq k\}}{R_m^k}\Big|\mathcal{F}_{\lambda,m}\Big). \tag{9.1}$$

Due to (BI3) it is easy to see that each summand only depends on the number $j = \#\{i_1,\ldots,i_k\}$ of different indices. At the end of the proof we determine these summands in dependence of j. But first we count the number of possibilities of choosing $(i_1,\ldots,i_k)$ leading to the same j. Let $j \in \{1,\ldots,V_m(\lambda)\wedge k\}\}$ be fixed. Clearly, there are $\binom{V_m(\lambda)}{j}$ possibilities to draw j different numbers $\{M_1,\ldots,M_j\}$ from the set $\{1,\ldots,V_m(\lambda)\}$. Moreover, by simple combinatorial considerations there are

$$\sum_{r=0}^{j-1}(-1)^r\binom{j}{r}(j-r)^k$$

possibilities of choosing indices $i_1,\ldots,i_k$ from $\{M_1,\ldots,M_j\}$ such that every M_s, $1 \leq s \leq j$, is picked at least once, see, e.g., (II.11.6) in Feller [14]. Consequently, we obtain from (BI3) that (9.1) equals

$$\sum_{j=1}^{V_m(\lambda)\wedge k} C_{j,k}V_m(\lambda)\ldots(V_m(\lambda)-j+1)\mathbb{E}\Big(\frac{\mathbf{1}\{p_{s,m} \leq \widehat{\alpha}_{R_m:m}, s \leq j\}}{R_m^k}\Big|\mathcal{F}_{\lambda,m}\Big).$$

Clearly, we can replace $V_m(\lambda)\wedge k$ by k since each additional summand is equal to 0. It remains to determine the summands. Let $j \leq V_m(\lambda)\wedge k$. Without loss of generality we can assume, conditioned on $\mathcal{F}_{\lambda,m}$, that the first $V_m(\lambda)$ p-values correspond to true null hypotheses and $p_{1,m},\ldots,p_{V_m(\lambda),m} \leq \lambda$. In particular, we may consider $p_m^{(j,\lambda)} = (0,\ldots,0,p_{j+1,m},\ldots,p_{m,m})$. We obtain analogously to the calculation in (8.1) and the one before it that

$$\begin{aligned}&\mathbb{E}\Big(\frac{\mathbf{1}\{p_{s,m} \leq \widehat{\alpha}_{R_m:m}, s \leq j\}}{R_m^k}\Big|\mathcal{F}_{\lambda,m}\Big)\\&= \mathbb{E}\Big(\frac{\mathbf{1}\{p_{s,m} \leq \widehat{\alpha}_{R_m^{(j,\lambda)}:m}, s \leq j\}}{(R_m^{(j,\lambda)})^k}\Big|\mathcal{F}_{\lambda,m}\Big) = \Big(\frac{\alpha}{\widehat{m}_0}\Big)^j\mathbb{E}\Big(\Big(R_m^{(j,\lambda)}\Big)^{j-k}\Big|\mathcal{F}_{\lambda,m}\Big). \quad \square\end{aligned}$$

Acknowledgement

The authors thank the editor, the associate editor and the three referees for pointing out relevant references and reading our manuscript carefully. Their comments improved the paper's quality substantially. This work was partially supported by the *Deutsche Forschungsgemeinschaft* (Grant no. JA 472/11-3).

References

[1] Abramovich, F., Benjamini, Y., Donoho, D. L. and Johnstone, I. M. (2006). Adapting to unknown sparsity by controlling the false discovery rate. *Ann. Statist.* **34**, 584–653. MR2281879

[2] Benditkis, J., Heesen, P. and Janssen, A. (2018). The false discovery rate (FDR) of multiple tests in a class room lecture. *Statist. Probab. Lett.* **134**, 29–35. MR3758578

[3] 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**, 289–300. MR1325392

[4] Benjamini, Y. and Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *J. Educ. Behav. Statist.* **25**, 60–83.

[5] Benjamini, Y., Krieger, A. M. and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93**, 491–507. MR2261438

[6] Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependence. *Ann. Statist.* **29**, 1165–1188. MR1869245

[7] Blanchard, G. and Roquain, E. (2008). Two simple sufficient conditions for FDR control. *Electron. J. Stat.* **2**, 963–992. MR2448601

[8] Blanchard, G. and Roquain, E. (2009). Adaptive false discovery rate control under independence and dependence. *J. Mach. Learn. Res.* **10**, 2837–2871. MR2579914

[9] Blanchard, G., Dickhaus, T., Roquain, E. and Villers, F. (2014). On least favorable configurations for step-up-down-tests. *Statist. Sinica* **24**, 1–23. MR3184590

[10] Chi, Z. (2007). On the Performance of FDR Control: Constraints and a Partial Solution.. *Ann. Stat.* **35**, 1409–1431. MR2351091

[11] Chi, Z. and Tan, Z. (2008). Positive false discovery proportions: intrinsic bounds and adaptive control. *Statist. Sinica* **18**, 837–860. MR2440397

[12] Consul, P.C. and Famoye, F. (2006). *Lagrangian probability distributions.* Birkhäuser Boston, Inc., Boston, MA. MR2209108

[13] Fan, J., Han, X. and Gu, W. (2012). Estimating false discovery proportion under arbitrary covariance dependence. *J. Amer. Statist. Assoc.* **107**, 1019–1035. MR3010887

[14] Feller, W. (1968). *An introduction to probability theory and its applications Vol. I.* Third edition. Wiley & Sons. MR0228020

[15] Ferreira, J. A. and Zwinderman, A. H. (2006). On the Benjamini-Hochberg method. *Ann. Statist.* **34**, 1827–1849. MR2283719

[16] Finner, H., Dickhaus, T. and Roters, M. (2009). On the false discovery rate and an asymptotically optimal rejection curve. *Ann. Statist.* **37**, 596–618. MR2502644

[17] Finner, H. and Gontscharuk, V. (2009). Controlling the familywise error rate with plug-in estimator for the proportion of true null hypotheses. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71**, 1031–1048. MR2750256

[18] Finner, H., Kern, P. and Scheer, M. (2015). On some compound distributions with Borel summands. *Insurance Math. Econom.* **62**, 234–244. MR3348873

[19] Finner, H. and Roters, M. (2001). On the false discovery rate and expected type I errors. *Biom. J.* **43**, 985–1005. MR1878272

[20] Genovese, C. and Wassermann, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.* **32**, 1035–1061. MR2065197

[21] Gontscharuk, V. (2010). Asymptotic and exact results on FWER and FDR in multiple hypothesis testing. Ph.D. thesis, Heinrich-Heine University Düsseldorf. `https://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=16990`

[22] Heesen, P. (2014). Adaptive step-up tests for the false discovery rate (FDR) under independence and dependence. Ph.D. thesis, Heinrich-Heine University Düsseldorf. `https://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=33047`

[23] Heesen, P. and Janssen, A. (2015). Inequalities for the false discovery rate (FDR) under dependence. *Electron. J. Stat.* **9**, 679–716. MR3331854

[24] Heesen, P. and Janssen, A. (2016). Dynamic adaptive multiple tests with finites sample FDR control. *J. Statist. Plann. Inference* **168**, 38–51. MR3412220

[25] Jain, G.C. (1975). A linear function Poisson distribution. *Biom. Z.* **17**, 501–506. MR0394970

[26] Liang, K. and Nettleton, D. (2012). Adaptive and dynamic adaptive procedures for false discovery rate control and estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74**, 163–182. MR2885844

[27] Meinshausen, N. and Bühlmann, P. (2005). Lower bounds for the number of false null hypotheses for multiple testing of associations under general dependence structures. *Biometrika* **92**, 893–907. MR2234193

[28] Meinshausen, N. and Rice, J. (2006). Estimating the proportion of false null hypotheses among a large number of independently tested hypotheses. *Ann. Statist.* **34**, 373–393. MR2275246

[29] Neuvial, P. (2008). Asymptotic properties of false discovery rate controlling procedures under independence. *Electron. J. Stat.* **2**, 1065–1110. Corrigendum **3**, 1083. MR2460858

[30] Owen, A. B. (2005). Variance of the number of false discoveries. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**, 411–426. MR2155346

[31] Roquain, E. and Villers, F. (2011). Exact calculations for false discovery proportion with application to least favorable configurations. *Ann.*

Statist. **39**, 584–612. MR2797857

[32] Sarkar, S. K. (2008). On methods controlling the false discovery rate. *Sankhyā* **70**, 135–168. MR2551809

[33] Sarkar, S. K., Guo, W. and Finner, H. (2012). On adaptive procedures controlling the familywise error rate. *J. Statist. Plann. Inference* **142**, 65–78. MR2827130

[34] Scheer, M. (2012). Controlling the number of false rejections in multiple hypotheses testing. Phd-thesis. Heinrich-Heine University Düsseldorf.

[35] Schwartzman, A. and X. Lin (2011). The effect of correlation in false discovery rate estimation. *Biometrika* **98**, 199–214. MR2804220

[36] Schweder, T. and Spjøtvoll, E. (1982). Plots of p-values to evaluate many tests simultaneously. *Biometrika* **69**, 493–502.

[37] Shorack, G.R. and Wellner, J.A. (2009). *Empirical Processes with Applications to Statistics.* Society for Industrial and Applied Mathematics, Philadelphia. MR3396731

[38] Storey, J. D. (2002). A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64**, 479–498. MR1924302

[39] Storey, J. D., Taylor, J. E. and Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **66**, 187–205. MR2035766

[40] Storey, J. D. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *PNAS* **100**, 9440–9445. MR1994856

[41] Yekutieli, D. and Benjamini, Y. (1999). Resampling-based false discovery rate controlling multiple test procedures for correlated test statistics. *J. Statist. Plann. Inference* **82**, 171–196. MR1736442

[42] Zeisel, A., Zuk, O. and Domany, E. (2011). FDR control with adaptive procedures and FDR monotonicity. *Ann. Appl. Stat.* **5**, 943–968. MR2840182