Statistica Sinica **18**(2008), 837-860

POSITIVE FALSE DISCOVERY PROPORTIONS: INTRINSIC BOUNDS AND ADAPTIVE CONTROL

Zhiyi Chi and Zhiqiang Tan

University of Connecticut and Rutgers University

Abstract: A useful paradigm for multiple testing is to control error rates derived from the false discovery proportion (FDP). The false discovery rate (FDR) is the expectation of the FDP, which is defined to be zero if no rejection is made. However, since follow-up studies are based on hypotheses that are actually rejected, it is important to control the positive FDR (pFDR) or the positive false discovery excessive probability (pFDEP), i.e., the conditional expectation of the FDP or the conditional probability of the FDP exceeding a specified level, given that at least one rejection is made. We show that, unlike FDR, these two positive error rates may not be controllable at a desired level. Given a multiple testing problem, there can exist positive intrinsic lower bounds, such that no procedures can attain a pFDR or pFDEP level below the corresponding bound. To reduce misinterpretations of testing results, we propose several procedures that are adaptive, i.e., they achieve pFDR or pFDEP control when the target control level is attainable, and make no rejections otherwise. The adaptive control is established under a sparsity condition where the fraction of false nulls is increasingly close to zero as well as under the condition where the fraction of false nulls is a positive constant. We demonstrate that the power of the proposed procedures is comparable to the Benjamini-Hochberg FDR controlling procedure.

Key words and phrases: False discovery excessive probability, false discovery rate, multiple testing, positive false discovery proportion, p-value, sparsity.

1. Introduction

Traditionally, multiple hypothesis testing aims to control familywise error rate (FWER), i.e., the probability of falsely rejecting one or more null hypotheses. To balance between error rate control and power, Benjamini and Hochberg (1995) introduced the false discovery rate (FDR), and established that FDR can be controlled at any specified level by a procedure originally due to Simes (1986), henceforth referred to as the BH procedure. Since then, there have been considerable researches on both the theory and applications of FDR control (cf., Benjamini and Hochberg (2000), Genovese and Wasserman (2002, 2004, 2006), Lehmann and Romano (2005), Storey (2002, 2003), Storey, Taylor and Siegmund (2004) and van der Laan, Dudoit and Pollard (2004) and references therein).

FDR is defined as the expectation of the false discovery proportion (FDP), which is the proportion of falsely rejected nulls among those rejected if there are any, and 0 otherwise. Two aspects of FDP are of interest. First, control of FDP can be considered in terms of the false discovery excessive probability (FDEP), which is the probability that FDP exceeds a specified level. Several procedures have been proposed for FDEP control. For example, Genovese and Wasserman (2006) suggested an inversion-based procedure, and van der Laan et al. (2004) proposed an augmentation-based procedure. These two procedures are equivalent under mild conditions (Genovese and Wasserman, 2006), and both are built on procedures that control FWER or k-FWER (i.e., the probability of falsely rejecting at least k nulls) without making assumptions on statistical dependency among p-values. On the other hand, Lehmann and Romano (2005) derived step-down procedures to control FDEP and k-FWER.

Second, FDR combines two factors: the probability of making no discovery, and the conditional expectation of FDP given that at least one discovery is made. Storey (2002, 2003) referred to the latter as positive FDR (pFDR), and argued that it is a more suitable error rate than FDR. By definition, pFDR is more relevant than FDR to follow-up studies conducted once positive findings are obtained. For the same reason, it is useful to consider positive FDEP (pFDEP), i.e., the conditional probability that FDP exceeds a specified level given that at least one discovery is made. However, to our knowledge, there are no procedures that realize control of pFDR or pFDEP when it is feasible. Storey (2002) proposed estimates of FDR and pFDR for fixed rejection regions, and showed that they are pointwise conservative. Storey et al. (2004) proved that these estimates are simultaneously conservative for fixed rejection regions with thresholds bounded away from 0, and that the procedure of Storey (2002) can achieve control of FDR (but not pFDR) at any specified level.

The objective of this article is twofold: theoretically, to understand the controllability of pFDR and pFDEP and methodologically, to develop suitable procedures to control them. First we establish that, given a multiple testing problem, there exists a possibly positive lower bound β_* on pFDR and, if the exceedance level for *FDP* is specified below β_*, there also exists a positive lower bound on the pFDEP. Genovese and Wasserman (2002) and Chi (2007) showed a dichotomous effect of β_* on the BH procedure: the number of rejections grows to ∞ or converges to a finite random variable as the number of tested hypotheses increases, depending on whether the FDR control level is above or below $\beta_*/(1-\pi)$, where π is the fraction of *false* nulls being tested. As a result, the asymptotic power is positive or zero.

Given a multiple testing problem, the above lower bounds are intrinsic, determined solely by the data-generating distribution. Therefore, no procedure can

ever attain a pFDR or pFDEP below the corresponding bound. The existence of the bounds has serious implications. For example, the lower bound β_* can be arbitrarily close to 1. If $\beta_* = 0.9$, say, given a *nonempty* set of rejected nulls, on average 90% of them are false rejections whatever multiple testing procedure is used. In this situation, it seems reasonable to require *no* rejections be made at all, in order to avoid grossly mistaken interpretations about the results.

Because the intrinsic lower bounds are beyond control at the stage of data analysis, and generally unknown, it is futile to seek a procedure that can control pFDR or pFDEP at any specified level. From this perspective, we suggest that a desirable procedure be adaptive, i.e., that it automatically achieves a specified control level whenever the level is attainable, and avoids making any rejections otherwise.

To develop a methodology of adaptive control, we consider two scenarios. In the first, the fraction π of false nulls is known. We propose procedures that are adaptive to control pFDR and pFDEP, respectively. In the second scenario, π is unknown. The proposed procedures are similar to the previous ones, but with π replaced by 0. The procedures are still adaptive, but are conservative. On the other hand, they can achieve adaptive control even when π tends to 0.

The rest of the article is organized as follows. Section 2 describes the setup. Section 3 studies the intrinsic lower bounds on pFDR and pFDEP and the resulting "subcritical" and "supercritical" cases. Sections 4 and 5 present several adaptive pFDR or pFDEP controlling procedures and related asymptotic results. Section 6 reports a simulation study and an application to gene expression data. Section 7 gives concluding remarks. The Appendix collects selected technical details. Proofs of major theorems can be found in the Supplemental Materials.

2. Setup

Suppose that there are $n\,(\geq 1)$ null hypotheses to be tested. For $1 \leq i \leq n$, let ξ_i be the p-value associated with the ith null, and let $H_i = 0$ (resp. 1) if the ith null is true (resp. false). Consider the following mixture model (Efron et al. (2001), Genovese and Wasserman (2002, 2004) and Storey (2003)):

$$(\xi_1, H_1), \ldots, (\xi_n, H_n) \text{ are iid, such that}$$
$$H_i \sim \text{ Bernoulli}(\pi), \quad \xi_i \,|\, H_i = \theta \sim \begin{cases} \text{Uniform}(0,1), & \text{if } \theta = 0, \\ G \text{ with density } g, & \text{otherwise.} \end{cases}$$

Under this model, each p-value ξ_i has the (marginal) distribution function

$$F(t) = (1-\pi)t + \pi G(t), \quad t \in [0,1].$$

Formally, a multiple testing procedure is defined through a mapping

$$\begin{gathered}\delta = (\delta_1, \ldots, \delta_n) : [0,1]^n \to \{0,1\}^n, \\ \text{such that the } i\text{th null is accepted } \iff \delta_i(\xi_1, \ldots, \xi_n) = 0.\end{gathered} \tag{2.1}$$

It follows that the set of rejected nulls is completely determined by the p-values $\xi_1, \ldots, \xi_n$. As far as we know, all multiple testing procedures in the literature are strictly based on p-values in the sense that $\delta_i = \delta_j$ whenever $\xi_i = \xi_j$.

By the meaning of p-value, it is often required of a multiple testing procedure that whenever a null is rejected, all those with smaller or equal p-values be rejected as well. Equivalently, such a procedure can be identified with a "threshold" function $\tau : [0,1]^n \to [0,1]$, such that $\delta_i = \mathbf{1}\{\xi_i \le \tau(\xi_1, \ldots, \xi_n)\}$ for each $1 \le i \le n$. Section 3 will consider the controllability of pFDR and pFDEP under the general form (2.1). On the other hand, the proposed procedures in Sections 4 and 5 all involve threshold functions.

Given a multiple testing procedure, denote by R the number of rejected nulls and V that of rejected true nulls. A procedure is called trivial if it makes no rejection, i.e., $P(R = 0) = 1$. For a nontrivial procedure, define

The False Discovery Rate as $\mathrm{FDR} = E[V/(R \vee 1)]$,

The Positive false Discovery Rate as $\mathrm{pFDR} = E[V/R \,|\, R > 0]$,

The False Discovery Excessive Probability at FDP exceedance level $\alpha \in (0,1)$ as $\mathrm{FDEP}_\alpha = P[V/(R \vee 1) > \alpha]$,

The Positive False Discovery Excessive Probability at FDP exceedence level $\alpha \in (0,1)$ as $\mathrm{pFDEP}_\alpha = P[V/R > \alpha \,|\, R > 0]$,

where $a \vee b$ denotes the larger one between a and b. Apparently,

$$\mathrm{FDR} = \mathrm{pFDR} \times P(R > 0)\,, \quad \mathrm{FDEP}_\alpha = \mathrm{pFDEP}_\alpha \times P(R > 0).$$

Therefore, FDR (resp. FDEP) consists of two conceptually distinct factors: $P(R = 0)$, i.e., the probability of rejecting no null and pFDR (resp. pFDEP), as a measure of error conditional on rejecting at least one null.

A simple but important class of multiple testing procedures is to reject all the nulls with p-values up to a fixed threshold. Let

$$R_t = \#\{i : \xi_i \le t\}\,, \quad V_t = \#\{i : H_i = 0,\ \xi_i \le t\}\,, \qquad 0 \le t \le 1. \tag{2.2}$$

Note that $E(R_t/n) = (1-\pi)t + \pi G(t)$ and $E(V_t/n) = (1-\pi)t$. Define

$$\alpha_t = \frac{(1-\pi)t}{(1-\pi)t + \pi G(t)}, \quad \beta_t = \frac{1-\pi}{1-\pi+\pi g(t)}, \tag{2.3}$$

where α_0 is taken to be β_0 by continuous extension, and β_t is called the "local FDR" (Efron et al. (2001) and Broberg (2005)). Therefore, the lowest attainable

FDR and local FDR are

$$\begin{aligned}\alpha_* &= \inf_{0\le t\le 1} \alpha_t = \frac{1-\pi}{1-\pi+\sup\limits_{0\le t\le 1}\dfrac{G(t)}{t}},\\ \beta_* &= \inf_{0\le t\le 1} \beta_t = \frac{1-\pi}{1-\pi+\sup\limits_{0\le t\le 1} g(t)}.\end{aligned} \tag{2.4}$$

In general, $\sup_t G(t)/t \le \sup_t g(t)$ and $\alpha_* \ge \beta_*$, because $G(t)/t = g(s)$ for some $s \in [0,t]$ by the Mean Value Theorem. On the other hand, if G is concave, then $\alpha_t \le \beta_t$ and both are increasing in $[0,1]$ (Broberg (2005)), so $\alpha_* \le \beta_*$ and thus $\alpha_* = \beta_*$.

The following proposition is straightforward but fundamental.

Proposition 2.1. *Let $0 < t \le 1$. Under the mixture model,*
(a) *$H_1, \ldots, H_n$ are independent given $\xi_1, \ldots, \xi_n$, and*

$$P(H_i = 0 \,|\, \xi_i \le t) = \alpha_t, \quad P(H_i = 0 \,|\, \xi_i = t) = \beta_t, \qquad 1 \le i \le n;$$

(b) *for $1 \le k \le n$, conditioning on $R_t = k$, V_t is binomial on k trials and success probability α_t per trial* ($\mathrm{Bin}(k, \alpha_t)$).

Result (a) implies that given all the observed p-values, the probability that an individual null is true is completely determined by its own p-value, regardless of the others. Result (b) provides the conditional distribution of the number of false rejections given the total number of rejections. It is the basis for the procedures in Sections 4 and 5.

Notation. For a distribution function F, let $F^*(t) = \inf\{x : F(x) \ge t\}$, $0 < t < 1$, be the corresponding quantile function. Denote by $\mathrm{pbin}(\gamma; n, p)$ the distribution function of $\mathrm{Bin}(n,p)$, and by $\mathrm{qbin}(\gamma; n, p)$ the corresponding quantile function. If $n = 0$ or $p = 0$, $\mathrm{Bin}(n,p)$ is concentrated at 0. Denote by Φ the distribution function of $N(0,1)$. Finally, adopt the convention that $\max \emptyset = 0$.

As we only consider in-probability asymptotics of multiple testing procedures, the sets of nulls for different n need not be nested. Henceforth, assume that for each n, $(\xi_1^{(n)}, H_1^{(n)}), \ldots, (\xi_n^{(n)}, H_n^{(n)})$ are iid from a mixture model, where $\pi = \pi_n$ and $G = G_n$ may depend on n. Denote by $\xi_{n:1} \le \cdots \le \xi_{n:n}$ the order statistics of the p-values.

3. Subcritical vs Supercritical Conditions

Given $\alpha, \gamma \in (0,1)$, we say that (p)FDR is controlled at level α if (p)FDR $\le \alpha$, and (p)FDEP$_\alpha$ at level γ if (p)FDEP$_\alpha \le \gamma$. The BH procedure is useful in that it can control FDR at any desired level α. However, several important

issues remain. To what degree can the BH procedure control pFDR? Is there a procedure that can control pFDR at any level α? Similar questions can be raised for pFDEP_α. As seen below, the answers depend critically on how large the level α is.

To start, consider a procedure that rejects nulls with p-values no greater than a fixed $t \in (0,1)$. Then $R = R_t$, $V = V_t$ and, by Proposition 2.1,

$$E\Big(\frac{V_t}{R_t} \,|\, R_t = k\Big) = \alpha_t, \tag{3.1}$$

$$P\Big(\frac{V_t}{R_t} > \alpha \,|\, R_t = k\Big) = 1 - \mathrm{pbin}(\alpha k; k, \alpha_t) \tag{3.2}$$

for any $k \geq 1$. Therefore, $\mathrm{pFDR} = \alpha_t$ is lower bounded by $\alpha_* \geq \beta_*$ defined in (2.4). Likewise, pFDEP_α is lower bounded by γ_*, where

$$\gamma_* = \gamma_*(\alpha) = 1 - \sup_{k \geq 1} \mathrm{pbin}(\alpha k; k, \alpha_*). \tag{3.3}$$

Note that $\gamma_* = 0$ if $\alpha > \alpha_*$, and $\gamma_* > 0$ if $\alpha < \alpha_*$. As a result, no procedure with a fixed rejection threshold can attain pFDR below α_* and, if $\alpha < \alpha_*$, no such procedure can attain pFDEP_α below γ_*. In general, similar results can be established for nontrivial multiple testing procedures.

Proposition 3.1. *Under the mixture model, the following statements hold for* any *nontrivial multiple testing procedure* (2.1).
(a) $\mathrm{pFDR} \geq \beta_*$.
(b) *If* $\alpha < \beta_*$, *then* $\mathrm{pFDEP}_\alpha \geq 1 - \sup_{k \geq 1} \mathrm{pbin}(\alpha k; k, \beta_*) > 0$.

Note that the lower bounds in Proposition 3.1 are intrinsic to a multiple testing problem, regardless of the procedure applied. The lower bounds reveal an important difference between pFDR (resp. pFDEP) and FDR (resp. FDEP): the latter can be made arbitrarily small since $P(R > 0)$ can be arbitrarily close to 0. This difference seems not yet well appreciated in the literature. In the context of FDR control using a fixed rejection region, Storey et al. (2004) noted that pFDR and FDR are asymptotically equivalent, and any asymptotic results on FDR can essentially be directly translated into results on pFDR. Nevertheless, this perspective of asymptotic equivalence cannot generally be extended to data-dependent random rejection regions, because the presumption that $P(R > 0)$ tends to 1 may no longer hold. Indeed, by Proposition 3.1, any procedure that controls FDR at level $\alpha < \beta_*$ necessarily makes no rejection with a positive probability,

$$P(R = 0) = 1 - \frac{\mathrm{FDR}}{\mathrm{pFDR}} \geq 1 - \frac{\alpha}{\beta_*} > 0.$$

As an example, consider the BH procedure when the fraction π of false nulls is known and incorporated (cf. (4.3)). The behavior of the procedure is categorically changed when α is decreased below β_* (Genovese and Wasserman (2002) and Chi (2007)). When $\alpha > \beta_*$, the number of rejections grows approximately linearly with the number of tested nulls; the pFDR is approximately equivalent to FDR and hence is controlled at level α. When $\alpha < \beta_*$, the number of rejections converges in distribution to a finite random variable that has a positive probability of being zero; meanwhile the pFDR approaches β_*. When π is unknown, the BH procedure (4.4) has a similar "phase transition" in its behavior, but with a higher critical value $\beta_*/(1-\pi)$ for α, due to the conservative estimation of an unknown π by 0.

The discussion so far has only involved the marginal distributions of V/R and R. The next result concerns their joint distribution. It implies that, when the number n of tested nulls is large, it is essentially impossible to have both $V/R \le \alpha < \beta_*$ and $R \sim \epsilon n$, no matter how close α is to β_* and how small $\epsilon > 0$ is.

Proposition 3.2. *Under the mixture model, if $\alpha < \beta_*$, then there exists a constant $c > 0$ such that, with probability one, $V \le \alpha R$ implies $R \le c \log n$ for all n large enough.*

Since setting the FDR or FDP exceedence level α above or below β_* has critical consequences for the control of false discovery proportions, we distinguish between the two cases. We call the case $\alpha > \beta_*$ subcritical and the case $\alpha < \beta_*$ supercritical. The critical case $\alpha = \beta_*$ rarely occurs in practice and is not considered.

In principle, when $\beta_* = 0$, i.e., $\sup_{0 \le t \le 1} g(t) = \infty$, any FDR or FDP exceedence level leads to a subcritical case. However, situations where $\beta_* > 0$ can arise rather naturally.

Example 3.1. For $1 \le i \le n$, let X_i be a test statistic with continuously differentiable distribution function Q_0 under $H_i = 0$, or Q_1 under $H_i = 1$. Suppose that for each null, rejection is made on the *left* tail of X_i, and the associated p-value is $\xi_i = Q_0(X_i)$. Then ξ_i has distribution function

$$G(t) = P(\xi_i \le t \,|\, H_i = 1) = P(X_i \le Q_0^*(t) \,|\, H_i = 1) = Q_1(Q_0^*(t)),$$

with density function

$$g(t) = \frac{Q_1'(Q_0^*(t))}{Q_0'(Q_0^*(t))} = \text{likelihood ratio of } Q_0^*(t). \tag{3.4}$$

Therefore, g is bounded on $[0,1]$ if and only if $Q_1'(x)/Q_0'(x)$ is bounded on $(-\infty,\infty)$. By Proposition 3.1, we obtain

$$\begin{aligned}\beta_* &= \text{infimum of pFDR} \\ &= \frac{1-\pi}{1-\pi+\pi\times(\text{supremum of likelihood ratio})}.\end{aligned} \tag{3.5}$$

Consider the following examples of Q_0 and Q_1.

(1) Let Q_0 be the distribution function of $N(0,1)$ and Q_1 that of $N(-a,1)$, with $a>0$. Then

$$g(t)=\frac{\exp\Big\{-\dfrac{(Q_0^*(t)+a)^2}{2}\Big\}}{\exp\Big\{-\dfrac{Q_0^*(t)^2}{2}\Big\}}=\exp\Big\{-aQ_0^*(t)-\frac{a^2}{2}\Big\},\quad 0\le t\le 1,$$

is strictly decreasing. It is easy to see that $\sup_{0\le t\le 1} g(t)=\lim_{t\to 0} g(t)=\infty$. Therefore, $\beta_*=0$.

(2) Let Q_0 be the distribution function of a Uniform$(0,1)$ and Q_1 that of a Beta$(1,b)$, with $b>1$, i.e., $Q_1(x)=1-(1-x)^b$, $x\in[0,1]$. Then

$$g(t)=b[1-Q_0^*(t)]^{b-1}=b(1-t)^{b-1}\qquad 0\le t\le 1,$$

is strictly decreasing. Because $\sup_{0\le t\le 1} g(t)=\lim_{t\to 0}(t)=b$, $\beta_*>0$.

(3) Let Q_0 be the standard Cauchy distribution function and Q_1 a scaled version of Q_0 with scaling factor $c>1$. Then

$$\begin{aligned}Q_0(x)=\frac{1}{2}+\frac{\arctan x}{\pi},\quad Q_1(x)=Q_0\Big(\frac{x}{c}\Big),\quad -\infty<x<\infty,\\ g(t)=\frac{c}{1+(c^2-1)\sin^2(\pi t)},\qquad 0\le t\le 1,\end{aligned}$$

so g is strictly decreasing if $t<1/2$ and strictly increasing otherwise. Because $\sup_{0\le t\le 1} g(t)=\lim_{t\to 0} g(t)=\lim_{t\to 1} g(t)=c$, $\beta_*>0$.

As noted in the Introduction, the lowest attainable pFDR level β_* can be arbitrarily close to 1. This can be seen from (3.5). Indeed, if the likelihood ratios associated with the test statistics are uniformly bounded, then the smaller the fraction π of false nulls, the closer β_* is to 1. As a result, it becomes increasingly difficult to pick true discoveries out of any *nonempty* set of rejections. It is worth pointing out again that this difficulty is not due to the multiple testing procedure but to the problem itself.

In what follows, we assume G is concave on [0, 1]. By (3.4), the assumption means that for each null, the smaller the associated test statistic is, the stronger

the evidence against the null. The global concavity assumption simplifies technicalities but is not essential for our results. Thus (1) $\alpha_t \le \beta_t$ and both are increasing on [0, 1]; and (2) $\alpha_* = \beta_* = \alpha_0 = (1-\pi)/(1-\pi+\pi g(0))$.

4. Procedure: Fixed Known Fraction of False Nulls

In this section, we take $\pi_n \equiv \pi \in (0,1)$ known, and $G_n \equiv G$ unknown, with a continuous and strictly decreasing density g. The purpose is twofold: to illustrate the basic ideas underlying the proposed procedures, and to accommodate the possibility that the fraction of false nulls can be found either from prior knowledge or by estimation (cf. Benjamini and Hochberg (2000), Langaas, Lindqvist and Ferkingstad (2005), and Storey (2002)).

4.1. Motivation

Our procedures are motivated by the idea that (p)FDR or (p)FDEP control can be realized by using the estimated conditional distribution of the number of false rejections given the total number of rejections; see Proposition 2.1. Given $0 < t < 1$, if $R_t > 0$, then $\alpha_t = (1-\pi)nt/E(R_t)$ can be estimated by $\hat{\alpha}_t = (1-\pi)nt/R_t$, with R_t in place of $E(R_t)$. Then by (3.1) and (3.2), $E(V_t/R_t \,|\, R_t)$ and $P(V_t/R_t > \alpha \,|\, R_t)$ can be estimated, respectively, by

$$\widehat{E}\Big(\frac{V_t}{R_t} \,|\, R_t\Big) = \hat{\alpha}_t = \frac{(1-\pi)nt}{R_t}, \tag{4.1}$$

$$\widehat{P}\Big(\frac{V_t}{R_t} > \alpha \,|\, R_t\Big) = 1 - \mathrm{pbin}(\alpha R_t; R_t, \hat{\alpha}_t). \tag{4.2}$$

In Storey (2002), $\hat{\alpha}_t$ was used as an estimate of $E(V_t/(R_t \vee 1))$ and $\hat{\alpha}_t/[1-(1-t)^n]$ as an estimate of $E(V_t/R_t \,|\, R_t > 0)$. The factor $1-(1-t)^n$ is asymptotically 0 for fixed $t > 0$, and has no effect on our proposed procedures.

Consider controlling (p)FDR based on the estimate (4.1): reject the R smallest p-values, where

$$\begin{aligned} R &= \max\Big\{k \ge 1 : \text{for } t = \xi_{n:k},\ \hat{E}\Big(\frac{V_t}{R_t} \,|\, R_t\Big) \le \alpha\Big\} \\ &= \max\Big\{k \ge 1 : \frac{(1-\pi)n\xi_{n:k}}{k} \le \alpha\Big\} \\ &= \max\{k \ge 1 : (1-\pi)n\xi_{n:k} \le \alpha k\}, \end{aligned} \tag{4.3}$$

with $\max\emptyset$ defined to be 0. The procedure is a BH procedure with π being known. If $1-\pi$ is replaced by 1, it becomes the original BH procedure, which rejects the R smallest p-values with

$$R = \max\{k \ge 1 : n\xi_{n:k} \le \alpha k\}. \tag{4.4}$$

Benjamini and Hochberg (2000) and Storey et al. (2004) showed that procedure (4.3) has FDR $= \alpha$, whereas procedure (4.4) has FDR $= (1-\pi)\alpha$.

Similar to (4.3), one possible way to control (p)FDEP$_\alpha$ at level γ is as follows: reject the R smallest p-values, where

$$\begin{aligned} R &= \max\Big\{k \ge 1 : \text{for } t = \xi_{n:k},\ \hat{P}\Big(\frac{V_t}{R_t} > \alpha \,|\, R_t\Big) \le \gamma\Big\} \\ &= \max\Big\{k \ge 1 : \text{pbin}\left(\alpha k;\ k,\ \frac{(1-\pi)n\xi_{n:k}}{k}\right) \ge 1-\gamma\Big\} \\ &= \max\Big\{k \ge 1 : \text{qbin}\left(1-\gamma;\ k,\ \frac{(1-\pi)n\xi_{n:k}}{k}\right) \le \alpha k\Big\}. \end{aligned} \tag{4.5}$$

This procedure is structurally similar to procedure (4.3), except that quantiles of binomial distributions are used rather than expected values.

4.2. Modification

Because G is concave, $\alpha_* = \beta_*$. By Proposition 3.1, no procedure can attain pFDR $< \alpha_*$, and no procedure with FDP exceedence level $\alpha < \alpha_*$ can attain pFDEP$_\alpha < \gamma_*$. Since α_* is unknown, the best possibility for a pFDR controlling procedure is that it be adaptive to both subcritical and supercritical conditions. That is, if $\alpha > \alpha_*$, the procedure attains pFDR $\le \alpha$; and if $\alpha < \alpha_*$, it almost never makes rejections, thus indicating that the pFDR cannot be controlled at level α. Likewise, the best possibility for a pFDEP controlling procedure is as follows: if $\alpha > \alpha_*$ or $\alpha < \alpha_*$ but $\gamma > \gamma_*$, the procedure attains pFDEP$_\alpha \le \gamma$; and if $\alpha < \alpha_*$ and $\gamma < \gamma_*$, it almost never makes rejections.

In order to modify (4.3) and (4.5) to achieve adaptive control, we first need to deal with the fluctuation in $\hat{\alpha}_t$ if $t \to 0$ as $n \to \infty$. Although $\hat{\alpha}_t$ converges to α_t for each fixed $0 < t < 1$, the process $(\hat{\alpha}_t)_{0<t<1}$ does not converge uniformly to $(\alpha_t)_{0<t<1}$. For example, $\hat{\alpha}_{\xi_{n:1}}$ converges in distribution to an exponentially distributed random variable with mean α_* rather than to the constant α_*. To avoid such instability, we replace $\hat{\alpha}_t$ with

$$\tilde{\alpha}_t = \frac{(1-\pi)n(t \vee \xi_{n:k_n})}{R_t \vee k_n}, \qquad \text{where } k_n \to \infty,\ \frac{k_n}{n} \to 0.$$

It follows that $\hat{\alpha}_{\xi_{n:k_n}}$ converges to α_*, and hence the process $(\tilde{\alpha}_t)_{0<t<1}$ converges uniformly to $(\alpha_t)_{0<t<1}$, i.e. $\sup_{0<t<1} |\tilde{\alpha}_t - \alpha_t| \to 0$.

By substituting $\tilde{\alpha}_t$ for $\hat{\alpha}_t$ in (4.3), we get the following adaptive pFDR controlling procedure at target pFDR control level α.

PFDR control with known π: *Reject the R smallest p-values, where*

$$R = \max\Big\{k \ge 1 : \frac{(1-\pi)n\xi_{n:(k\vee k_n)}}{k \vee k_n} \le \alpha\Big\}. \tag{4.6}$$

To modify (4.5) in order to control pFDEP_α at level γ, in addition to the fluctuation in α_t, we also need to deal with the fluctuation in the number of nulls. We first present a modification that correctly incorporates the fluctuations, and then give a heuristic argument.

PFDEP control with known π: *Reject the R smallest p-values, where*

$$R = \max\left\{k \geq 1 : \mathrm{qbin}\Big(\Gamma_*(\xi_{n:k});\, k,\, \frac{(1-\pi)n\xi_{n:(k\vee k_n)}}{k \vee k_n}\Big) \leq \alpha k\right\} \tag{4.7}$$

with

$$\Gamma_*(t) = \Phi\Big(\sqrt{1 + \frac{\alpha - (1-\pi)t}{1-\alpha}\mathbf{1}\left\{t > \xi_{n:k_n}\right\}}\Phi^*(1-\gamma)\Big).$$

Note that any $k_n \to \infty$ of order $o(n)$ can be used in (4.7) to yield the same asymptotic behavior of the procedure. In practice, we have taken $k_n = c\log n$, with c a positive constant.

Overall, procedure (4.7) accommodates both sub- and supercritical cases automatically. When $\alpha > \alpha_*$, R grows roughly linearly in n and (4.7) is asymptotically

$$R = \max\left\{k : \mathrm{qbin}\Big(\Gamma(\xi_{n:k});\, k,\, \frac{(1-\pi)n\xi_{n:k}}{k}\Big) \leq \alpha k\right\}, \tag{4.7a}$$

where $\Gamma(t) = \Phi[\sqrt{(1-(1-\pi)t)/(1-\alpha)}\,\Phi^*(1-\gamma)]$ (cf. Theorem 4.2). On the other hand, when $\alpha < \alpha_*$, R converges to a finite random variable and (4.7) is asymptotically (cf. Theorem 4.3)

$$R = \max\left\{k : \mathrm{qbin}\Big(1-\gamma;\, k, \frac{(1-\pi)n\xi_{n:k_n}}{k_n}\Big) \leq \alpha k\right\}. \tag{4.7b}$$

Heuristics. The supercritical case (4.7b) is straightforward, following the same idea as (4.6). For the subcritical case (4.7a), it remains to be seen why $1-\gamma$ in (4.5) should be replaced with $\Gamma(\xi_{n:k})$ so that $P(V \leq \alpha R) \approx 1-\gamma$. Let θ be the correct replacement. By the definition of R, $\mathrm{qbin}(\theta; R, \zeta n/R) \approx \alpha R$, where $\zeta = (1-\pi)\xi_{n:R}$. Now

$$\{V \leq \alpha R\} = \left\{\frac{V - \zeta n}{\sigma_R} \leq \frac{\mathrm{qbin}\Big(\theta; R, \frac{\zeta n}{R}\Big) - \zeta n}{\sigma_R}\right\}$$

where, for each k, σ_k is the standard deviation of $\mathrm{Bin}(n, \zeta n/k)$. By the normal approximation, the second fraction on the right side converges to $\Phi^*(\theta)$. On

the other hand, V is the number of true nulls with p-values no greater than $\xi_{n:R}$. Loosely speaking, under the mixture model, the probability that a p-value is no greater than $\xi_{n:R}$ and associated with a true null is $(1-\pi)\xi_{n:R} = \zeta$. Therefore $V \sim \mathrm{Bin}(n, \zeta)$, with standard deviation $\sigma' = \sqrt{\zeta(1-\zeta)n}$. Because $\sigma_R \approx \sqrt{\zeta n(1-\zeta n/R)} \approx \sqrt{\zeta n(1-\alpha)}$,

$$P(V \le \alpha R) \approx P\Big(\frac{V-\zeta n}{\sigma'} \le \sqrt{\frac{1-\alpha}{1-\zeta}}\Phi^*(\theta)\Big) \approx \Phi\Big(\sqrt{\frac{1-\alpha}{1-\zeta}}\Phi^*(\theta)\Big).$$

Therefore, if $\theta = \Gamma(\xi_{n:R}) = \Gamma(\zeta/(1-\pi))$, then $P(V \le \alpha R) \le 1-\gamma$.

4.3. Asymptotic results

By assumption, $F(u) = (1-\pi)u + \pi G(u)$ is concave on $[0,1]$. The case $\alpha \ge 1-\pi$ is trivial since all the null hypotheses can be rejected with pFDR $= 1-\pi \le \alpha$. For $\alpha \in (\alpha_*, 1-\pi)$, define

$$u^* = u^*(\alpha) = \text{the unique } u \in (0,1) \text{ with } (1-\pi)u = \alpha F(u)\,,$$

which is a counterpart of the solution to $u = \alpha F(u)$ for the original BH procedure (4.4) (Genovese and Wasserman (2002) and Chi (2007)).

For comparison with our procedures, Proposition 4.1 summarizes the asymptotic behavior of the BH procedure (4.3) under the subcritical and the supercritical conditions, respectively.

Proposition 4.1. *The following are true for the procedure at* (4.3).

(a) *If* $\alpha \in (\alpha_*, 1-\pi)$ *then, as* $n \to \infty$, $R/n \xrightarrow{\mathrm{P}} F(u^*)$, $\mathrm{pFDR} \to \alpha$, *and* $\mathrm{pFDEP}_\alpha \to 1/2$.

(b) *If* $\alpha \in (0, \alpha_*)$ *then, as* $n \to \infty$, $R \xrightarrow{d} \kappa$, $\mathrm{pFDR} \to \alpha_*$, *and*

$$\mathrm{pFDEP}_\alpha \to 1 - \sum_{k=1}^{\infty} \mathrm{pbin}(\alpha k; k, \alpha_*) q_k$$

where, letting $c = \alpha/\alpha_*$, $q_k = k^k(1-c)c^k e^{-kc}/k!$.

Our first result states that procedure (4.6) is adaptive for pFDR control. Specifically, in the subcritical case, the pFDR is asymptotically controlled at the target level, whereas in the supercritical case, the number of rejections tends to 0.

Theorem 4.1. (pFDR control with known $\boldsymbol{\pi}$) *The following are true for procedure* (4.6) *as* $n \to \infty$.

(a) *If* $\alpha \in (\alpha_*, 1-\pi)$, *the procedure is asymptotically the BH procedure* (4.3), *so that* $R/n \xrightarrow{\mathrm{P}} F(u^*)$ *and* $\mathrm{pFDR} \to \alpha$.

(b) *If* $\alpha \in (0, \alpha_*)$, *the procedure is asymptotically trivial:* $P(R = 0) \to 1$.

The adaptability of the pFDEP controlling procedure (4.7) is established next. First, in the subcritical case, the procedure can asymptotically control pFDEP_α at any specified level. For the BH procedure (4.3), by Proposition 4.1(a), $\text{pFDEP}_\alpha \to 1/2$. Second, in the supercritical case for procedure (4.7), R asymptotically can take at most two values, and pFDEP_α is asymptotically controlled if the specified level is attainable. For the BH procedure (4.3), by Proposition 4.1(b), R can take a large value with a positive probability, and pFDEP_α tends to a constant level.

Theorem 4.2. (Subcritical pFDEP control with known $\boldsymbol{\pi}$) *Let* $\alpha_* < \alpha < 1 - \pi$ *and* $0 < \gamma < 1$. *The following are true for procedure* (4.7) *as* $n \to \infty$.

(a) $R/n \xrightarrow{\text{P}} F(u^*)$, $\text{pFDR} \to \alpha$, *and* $\text{pFDEP}_\alpha \to \gamma$.

(b) *The probability that* (4.7) *and* (4.7a) *are identical tends to* 1.

Theorem 4.3. (Supercritical pFDEP control with known $\boldsymbol{\pi}$) *Let* $0 < \alpha < \alpha_*$. *Define* $\ell_0 = \max\{k \geq 1 : \text{qbin}(1 - \gamma; k, \alpha_*) \leq \alpha k\}$ *and* $\ell_1 = \max\{k \geq 1 : \text{qbin}(1 - \gamma; k, \alpha_*) + 1 \leq \alpha k\}$. *The following statements hold for procedure* (4.7) *as* $n \to \infty$.

(a) $P(R \in \{\ell_0, \ell_1\}) \to 1$.

(b) *For* $\ell = \ell_0$, ℓ_1, $V \mid R = \ell \xrightarrow{d} \text{Bin}(\ell, \alpha_*)$.

(c) *If* $\gamma > \gamma_*$, *then* $\overline{\lim}\, \text{pFDEP}_\alpha \leq \gamma$; *if* $\gamma < \gamma_*$, *then* $P(R = 0) \to 1$.

(d) *The probability that* (4.7) *and* (4.7b) *are identical tends to* 1.

We next consider power of the proposed adaptive procedures. Let N_0 be the total number of true nulls. Then the realized power is

$$\psi_n = \frac{(R - V)}{(n - N_0)}. \tag{4.8}$$

The result below shows that the powers of procedures (4.6), (4.7) and (4.3) are asymptotically the same. Consequently, procedures (4.6) and (4.7) asymptotically maintain the same power as the BH procedure (4.3), but achieve a stricter control in terms of pFDR and pFDEP.

Proposition 4.2. *If* $\alpha \in (\alpha_*, 1 - \pi)$, *then* $\psi_n \xrightarrow{\text{P}} G(u_*)$ *for procedures* (4.3), (4.6), and (4.7). *If* $\alpha \in (0, \alpha_*)$, *then* $\psi_n \xrightarrow{\text{P}} 0$ *for the three procedures.*

5. Procedure: Unknown Fraction of False Nulls and Increasingly Sparse False Nulls

We now consider that both π_n and G_n are unknown, but that G_n has a continuous and strictly decreasing density. We restrict our discussion to pFDEP control. pFDR control can be treated similarly.

5.1. Description

Our approach is to modify procedure (4.7) for pFDEP control. Because π_n is unknown, we replace it with the most conservative estimate for the fraction of false nulls, i.e., 0. Then we obtain the following procedure.

PFDEP control with unknown π_n: *Reject the R smallest p-values, where*

$$R = \max\left\{k : \mathrm{qbin}\Big(\Gamma_*(\xi_{n:k});\ k, \frac{n\xi_{n:(k\vee k_n)}}{k\vee k_n}\Big) \le \alpha k\right\} \tag{5.1}$$

with $k_n \sim c\log n$ as $n\to\infty$, where c is a positive constant and

$$\Gamma_*(t) = \Phi\Big(\sqrt{1+\frac{\alpha-t}{1-\alpha}\mathbf{1}\{t>\xi_{n:k_n}\}}\,\Phi^*(1-\gamma)\Big).$$

Under the conditions of Theorem 5.1, any $k_n\to\infty$ of order $o((\log n)^4)$ can be used in (5.1) to yield the same asymptotic behavior of the procedure. In practice, we have used $k_n \sim c\log n$ with c a positive constant.

Similar to (4.7), procedure (5.1) adapts to both sub- and supercritical cases. Note that, because $1-\pi_n$ is not incorporated in (5.1), given the value of α, the FDR and FDP exceedance level actually realized by the procedure is $(1-\pi_n)\alpha$ (cf. Benjamini and Hochberg (2000) and Storey et al. (2004)). For this reason, the sub- and supercritical cases need to be written as

$$\begin{aligned} &\text{subcritical:} && \overline{\lim_n}\,\frac{\alpha_*^{(n)}}{(1-\pi_n)} < \alpha,\\ &\text{supercritical:} && \underline{\lim_n}\,\frac{\alpha_*^{(n)}}{(1-\pi_n)} > \alpha. \end{aligned} \tag{5.2}$$

Recall $\alpha_*^{(n)} = 1/F_n'(0)$. In the subcritical case, asymptotically,

$$R = \max\left\{k : \mathrm{qbin}\Big(\Gamma(\xi_{n:k});\ k, \frac{n\xi_{n:k}}{k}\Big) \le \alpha k\right\}, \tag{5.1a}$$

where $\Gamma(t) = \Phi(\sqrt{(1-t)/(1-\alpha)}\,\Phi^*(1-\gamma))$. In the supercritical case, asymptotically,

$$R = \max\left\{k : \mathrm{qbin}\Big(1-\gamma;\ k, \frac{n\xi_{n:k_n}}{k_n}\Big) \le \alpha k\right\}. \tag{5.1b}$$

First, suppose $\pi_n \equiv \pi > 0$ and $G_n \equiv G$. For procedure (5.1), the critical value of the level α that divides the subcritical and the supercritical cases is

$\alpha_*/(1-\pi)$ due to the conservative estimation of π by 0. This follows from a similar argument for the BH procedure (4.4) in Genovese and Wasserman (2002) and Chi (2007).

In the supercritical case $\alpha < \alpha_*/(1-\pi)$, it is straightforward to extend Theorem 4.3: procedure (5.1) asymptotically controls pFDEP_α if $\gamma > \gamma_*$, and makes no rejection if $\gamma < \gamma_*$. In the subcritical case $\alpha > \alpha_*/(1-\pi)$, both the BH procedure (4.4) and procedure (5.1) asymptotically control pFDEP_α, and in fact become more conservative than the target pFDR control level α and pFDEP_α control level γ:

$$\text{pFDR} \to (1-\pi)\alpha, \qquad \text{pFDEP}_\alpha \to 0.$$

Such "over-control" is known for the BH procedure (4.4) (cf. Benjamini and Hochberg (2000), Finner and Roters (2001), Storey (2002) and Storey et al. (2004)), and can be similarly demonstrated for procedure (5.1).

Nevertheless, the over-control of pFDEP is an asymptotic behavior of procedure (5.1), and is evident only when n is sufficiently large. In fact, the smaller π is, the larger n has to be for the asymptotic behavior to take effect; see Section 6. In this situation, it seems more relevant to characterize the performance of procedure (5.1) when π is close to 0 but relatively, n is not large enough. It is also of interest to address the same question for the BH procedure (4.4) and to compare the two procedures. The approach we take is to investigate the asymptotic behaviors when false null hypotheses become increasingly sparse, i.e., $\pi_n \to 0$ as $n \to \infty$.

5.2. Asymptotic results

The presence of sparsity raises some interesting questions (Abramovich et al. (2006) and Donoho and Jin (2004, 2005)). Previous studies showed that, when false nulls become increasingly sparse, the disparity between the null and alternative distributions must increase accordingly in order to achieve good estimation. The same point applies to pFDEP control.

First, consider the subcritical case. Under the increasing sparsity condition, the subcritical case defined in (5.2) can be rewritten as

$$\pi_n \to 0, \quad \overline{\lim_{n\to\infty}}\, \alpha_*^{(n)} < \alpha. \tag{5.3}$$

In order to achieve pFDEP control, some constraints are necessary on how fast the fraction of false nulls can decrease and, at the same time, how fast the disparity between the null and alternative distributions should increase. In Theorem 5.1 below, the constraints are specified by condition (5.4). The result reveals a significant difference between procedure (5.1) and the BH procedure (4.4): the

former asymptotically achieves exact control of pFDEP whereas the latter gradually fails to control pFDEP.

For each large n, let

$$u_n = \text{the unique point } u \in (0,1) \text{ such that } u = \alpha F_n(u).$$

By the concavity of F_n and condition (5.3), u_n is well-defined for large n.

Theorem 5.1. (Subcritical pFDEP control with vanishing π_n) *Under condition* (5.3) *suppose that, for any* $\lambda \neq 1$,

$$\frac{n u_n \pi_n^2}{(\log n)^4} \left[\lambda - \frac{G_n(\lambda u_n)}{G_n(u_n)}\right]^2 \to \infty. \tag{5.4}$$

Then the following statements hold as $n \to \infty$.

(a) *For procedure* (5.1), $R \xrightarrow{\mathrm{P}} \infty$ *and* $\mathrm{pFDEP}_\alpha \to \gamma$. *That is, both* FDEP_α *and* pFDEP_α *are asymptotically controlled exactly at* γ.

(b) *The probability that* (5.1) *and* (5.1a) *are identical tends to* 1.

(c) *In contrast to* (a), *for the BH procedure* (4.4), $\mathrm{pFDEP}_\alpha \to 1/2$.

(d) *For both procedures* (4.4) *and* (5.1), *the power* ψ_n *as defined in* (4.8) *satisfies* $\psi_n / G_n(u_n) \xrightarrow{\mathrm{P}} 1$.

As an example of condition (5.4), let $G_n(u) = u^{\theta_n}$ with $\theta_n \downarrow 0$. Because $G_n'(0) = \infty$, the procedure is always subcritical. Let $c = 1/\alpha - 1$. Then $u_n = \alpha F_n(u_n)$ implies $u_n = [\pi_n/(c+\pi_n)]^{1/(1-\theta_n)}$. From $G_n(\lambda u_n)/G_n(u_n) = \lambda^{\theta_n} \to 1$, $\theta_n \to 0$, and $\pi_n \to 0$, it follows that $G_n(u_n) \sim \pi_n^{\theta_n/(1-\theta_n)}$ and (5.4) is equivalent to $n\pi_n^{2+1/(1-\theta_n)}/(\log n)^4 \to \infty$. Therefore, if $\pi_n \sim n^{-1/3+\epsilon}$ with $\epsilon > 0$, then (5.4) is satisfied.

Next consider the supercritical case where

$$\pi_n \to 0, \quad \varliminf_{n\to\infty} \alpha_*^{(n)} > \alpha. \tag{5.5}$$

Theorem 5.2. (Supercritical pFDEP control with vanishing π_n) *Under condition* (5.5), *the probability that procedures* (5.1) *and* (5.1b) *are identical tends to* 1, *and there is a constant* K_0 *such that* $P(R < K_0) \to 1$.

Furthermore, suppose that

$$\lim_{n\to\infty} \alpha_*^{(n)} = \lim_{n\to\infty} \frac{F_n^*\left(\frac{k_n}{n}\right)}{\frac{k_n}{n}} < 1, \tag{5.6}$$

and denote the limit by α_*. *Let* $\ell_0 = \max\{k \geq 1 : \mathrm{qbin}(1-\gamma; k, \alpha_*) \leq \alpha k\}$ *and* $\ell_1 = \max\{k \geq 1 : \mathrm{qbin}(1-\gamma; k, \alpha_*) + 1 \leq \alpha k\}$. *Then the following statements hold for procedure* (5.1) *as* $n \to \infty$.

(a) $P(R \in \{\ell_0, \ell_1\}) \to 1$.
(b) *For* $\ell = \ell_0$, ℓ_1, $V \,|\, R = \ell \xrightarrow{d} \mathrm{Bin}(\ell, \alpha_*)$.
(c) *If* $\gamma > \gamma_*$, *then* $\overline{\lim}\, \mathrm{pFDEP}_\alpha \le \gamma$; *if* $\gamma < \gamma_*$, *then* $P(R = 0) \to 1$.
(d) *For both procedures* (5.1) *and* (4.4), *the power satisfies* $\psi_n \xrightarrow{\mathrm{P}} 0$.

6. Numerical Studies

In this section, we report numerical studies based on simulated data and on a set of gene expression data to assess the proposed procedures. We label the procedures as CT and compare them with the BH procedure (4.4) and the procedures proposed by van der Laan et al. (2004, VDP) and Lehmann and Romano (2005, LR). The VDP procedure rejects the smallest $R_{\mathrm{VDP}} = \lfloor R_{\mathrm{Hommel}}/(1-\alpha) \rfloor$ p-values, where $\lfloor \cdot \rfloor$ denotes the floor function and

$$R_{\mathrm{Hommel}} = \max\left\{k : \xi_{1:n} \le \frac{\gamma}{n}, \cdots, \xi_{k:n} \le \frac{\gamma}{(n+1-k)}\right\}.$$

The LR procedure rejects the smallest R_{LR} p-values, where

$$R_{\mathrm{LR}} = \max\left\{k : \xi_{1:n} \le \frac{\gamma\alpha_1}{C_n}, \cdots, \xi_{k:n} \le \frac{\gamma\alpha_k}{C_n}\right\}$$

with $\alpha_k = (\lfloor \alpha k \rfloor + 1)/(\lfloor \alpha k \rfloor + n + 1 - k)$ and $C_n = \sum_{j=1}^{\lfloor \alpha n \rfloor + 1} 1/j$. The VDP and LR procedures yield $\mathrm{FDEP}_\alpha \le \gamma$ under arbitrary statistical dependency among the p-values.

6.1. Simulation study

Throughout, $\alpha = 0.2$ and $\gamma = 0.05$. We examine the performances of the procedures in terms of several quantities, including $P(R > 0)$, $\mathrm{pFDEP}_\alpha = P(V/R > \alpha \,|\, R > 0)$, and power $= E[(R-V)/(n-N_0)]$, where N_0 is the total number of true nulls. All quantities are computed as Monte Carlo averages from $10,000$ repeated simulations.

In each simulation, the parameter π is the fraction of false nulls under the mixture model, while the alternative distribution G is a Beta distribution with density $b(1-x)^{b-1}$. As a result, the distribution of p-values has density $(1-\pi)x + \pi b(1-x)^{b-1}$, and hence $\alpha_*/(1-\pi) = 1/[1+(b-1)\pi]$.

Table 1 summarizes the simulation results for procedure (5.1) under four subcritical configurations, with $(\pi, b) = (0.1, 100)$, $(0.05, 199)$, $(0.02, 496)$, and $(0.01, 991)$, respectively. In these configurations, the alternative distribution G is increasingly concentrated near 0, but the fraction π of false nulls is decreasing to 0, so that $\alpha_*/(1-\pi)$ is fixed at $1/10.9 = 0.09$. Recall that procedure (5.1) involves a sequence $k_n \sim c \log n$, with c a positive constant. We apply the procedure for

$k_n = \lfloor \log n \rfloor$ and for $2\lfloor \log n \rfloor$. Table 1 shows that the results are similar. It also shows that procedure (5.1) controls $\mathrm{pFDEP}_{0.2}$ at level $\gamma = 0.05$ for all the four configurations when $n = 20,000$, but not for $\pi = 0.01$ or 0.02 when $n = 2,000$. In the latter cases, π is close to 0 but n is not sufficiently large for the asymptotic control to take effect. In fact, although $R \to \infty$ in probability as $n \to \infty$, the probability that $R = 0$ is 0.54 or greater for $\pi = 0.01$ or 0.02 and $n = 2,000$.

Table 2 summarizes the simulation results for procedure (5.1) under four supercritical configurations, with $(\pi, b) = (0.1, 10)$, $(0.05, 19)$, $(0.02, 46)$, and $(0.01, 91)$, respectively. The distribution G is increasingly concentrated near 0 and the fraction π is decreasing to 0 as in Table 1, but $\alpha_*/(1-\pi)$ is now fixed at $1/1.9 = 0.53$. In this case, it is not possible for any procedure to control $\mathrm{pFDEP}_{0.1}$ at level $\gamma = 0.05$. Procedure (5.1) responds to this fact by almost never making rejections. In fact, for each configuration of (π, b), rejections only occur in 0−2 simulations out of 10,000.

Finally, Table 3 summarizes the simulation results for procedure (5.1), BH, VDP, and LR with $(\pi, b) = (0.05, 199)$ and $(0.05, 19)$. The results are qualitatively similar. Note that the four procedures are not strictly comparable as they are designed for different purposes: procedure (5.1) for pFDEP control, the BH procedure for FDR control, and the VDP and LR procedures for FDEP control. Nevertheless, three observations are worth mentioning. First, procedure (5.1) is adaptive, making rejections appropriately under the subcritical configuration,

Table 1. Simulation results for procedure (5.1): subcritical case, $\alpha = 0.2$, $\gamma = 0.05$. For each pair (n, k_n), results are obtained for 4 different (π, b), with π the fraction of the alternative $\mathrm{Beta}(1, b)$ distribution, $b = 100$, 199, 496, and 991, respectively. *Top* $k_n = \lfloor \log n \rfloor$. *Bottom*: $k_n = 2\lfloor \log n \rfloor$.

π	0.1	0.05	0.02	0.01	0.1	0.05	0.02	0.01
	$n = 2000,\ k_n = 7$				$n = 20,000,\ k_n = 9$			
$P(R > 0)$	0.9960	0.8502	0.4588	0.2795	1	1	1	0.9951
pFDR	0.13	0.13	0.12	0.12	0.17	0.17	0.16	0.15
FDR	0.13	0.11	0.054	0.034	0.17	0.17	0.16	0.15
pFDEP_α	0.011	0.046	0.10	0.14	0	0.01	0.034	0.045
FDEP_α	0.011	0.039	0.047	0.038	0	0.01	0.034	0.045
Power	0.70	0.48	0.19	0.12	0.85	0.83	0.78	0.70
	$n = 2000,\ k_n = 14$				$n = 20,000,\ k_n = 18$			
$P(R > 0)$	0.9950	0.8289	0.3545	0.1091	1	1	1	0.9952
pFDR	0.13	0.13	0.13	0.15	0.17	0.17	0.16	0.15
FDR	0.13	0.11	0.045	0.017	0.17	0.17	0.16	0.15
pFDEP_α	0.011	0.046	0.11	0.20	0	0.01	0.034	0.045
FDEP_α	0.011	0.038	0.039	0.022	0	0.01	0.034	0.045
Power	0.70	0.48	0.18	0.067	0.85	0.83	0.78	0.70

Table 2. Simulation results for procedure (5.1): supercritical case, $\alpha = 0.2$, $\gamma = 0.05$.

π	0.1	0.05	0.02	0.01	0.1	0.05	0.02	0.01
	$n = 2,000,\ k_n = 7$				$n = 20,000,\ k_n = 9$			
$P(R > 0)$	0	0.0002	0	0	0	0	0.0002	0
pFDR	NA	0.30	NA	NA	NA	NA	0.55	NA
FDR	0	<0.0001	0	0	0	0	0.0001	0
pFDEP_α	NA	0.50	NA	NA	NA	NA	1	NA
FDEP_α	0	0.0001	0	0	0	0	0.0002	0
Power	0	<0.0001	0	0	0	0	<0.0001	0
	$n = 2,000,\ k_n = 14$				$n = 20,000,\ k_n = 18$			
$P(R > 0)$	0	0	0	0	0	0	0	0
pFDR	NA	NA	NA	NA	NA	NA	NA	NA
FDR	0	0	0	0	0	0	0	0
pFDEP_α	NA	NA	NA	NA	NA	NA	NA	NA
FDEP_α	0	0	0	0	0	0	0	0
Power	0	0	0	0	0	0	0	0

and almost never under the supercritical configuration. Second, the BH procedure controls the FDR at the specified level $\alpha = 0.1$ but, unlike procedure (5.1), fails to control $\text{FDEP}_{0.1}$ or $\text{pFDEP}_{0.1}$ under the subcritical configuration. Third, although the VDP and LR procedures are able to control FDEP for any dependency structure of the p-values, they appear substantially less powerful than procedure (5.1) and the BH procedure, especially in the subcritical case.

6.2. Application to gene expression

We analyze the data reported in the study of Hedenfalk et al. (2001), who sought to identify differentially expressed genes between breast cancer tumors in patients who were BRCA1- and BRCA2-mutation-positive (cf. `http://research.nhgri.nih.gov/microarray/NEJM_Supplement/`). The raw data consist of 3,226 genes on 7 BRCA1 arrays and 8 BRCA2 arrays. For ease of comparison, we remove the genes with measurements exceeding 20 and analyze the data for the remaining 3,170 genes on the $\log_2$ scale.

First, following Storey and Tibshirani (2003), we used a two-sample t-statistic and compute its p-value based on permutations of array labels to test each gene for differential expression between BRCA1 and BRCA2 arrays. Next we applied procedure (5.1), as well as the BH (4.4), VDP, and LR procedures, to the resulting p-values. For this example, Storey and Tibshirani (2003) estimated that 67% of the genes are not differentially expressed. Based on this estimate, we also applied procedures (4.7) and (4.3) with $1 - \pi \approx 0.67$. For all the adaptive

Table 3. Comparison of simulation results for procedures: $\alpha = 0.2$, $\gamma = 0.05$.

$\pi = .05$	CT	BH	VDP	LR	CT	BH	VDP	LR
Subcritical	$n = 2,000,\ k_n = 7$				$n = 20,000,\ k_n = 9$			
$P(R > 0)$	0.8502	1	0.4260	0.0768	1	1	0.4168	0.0615
pFDR	0.13	0.19	0.084	0.10	0.17	0.19	0.091	0.082
FDR	0.11	0.19	0.036	0.0077	0.17	0.19	0.038	0.0051
pFDEP_α	0.046	0.39	0.10	0.10	0.01	0.22	0.12	0.086
FDEP_α	0.039	0.39	0.044	0.0078	0.01	0.22	0.048	0.0053
Power	0.48	0.89	0.0051	0.0007	0.83	0.89	0.0005	0.0001
Supercritical	$n = 2,000,\ k_n = 7$				$n = 20,000,\ k_n = 9$			
$P(R > 0)$	0.0002	0.3749	0.0892	0.0151	0	0.3862	0.0953	0.0114
pFDR	0.30	0.49	0.48	0.52	NA	0.50	0.51	0.49
FDR	<0.0001	0.18	0.043	0.0079	0	0.19	0.048	0.0056
pFDEP_α	0.50	0.69	0.49	0.52	NA	0.70	0.52	0.50
FDEP_α	0.0001	0.26	0.044	0.0079	0	0.27	0.050	0.0057
Power	<0.0001	0.0049	0.0005	<0.0001	0	0.0005	<.0001	<.0001

procedures, we used $k_n = \lfloor \log n \rfloor$ and $2 \lfloor \log n \rfloor$. We only report the results obtained with $k_n = \lfloor \log n \rfloor$. The results obtained with $k_n = 2 \lfloor \log n \rfloor$ are similar.

Figure 1 shows the number of rejections (i.e., significant genes) by the tested procedures across a range of values of $\alpha \leq 0.2$ and $\gamma \leq 0.2$. Each procedure declares a gene significant if the associated p-value is below a threshold. Therefore, the sets of significant genes generated by the procedures are nested within each other. Note that the procedures are based on different criteria of controlling false discoveries and therefore are not strictly comparable. Compared with the BH procedures (4.4) and (4.3), the proposed procedures (4.7) and (5.1) control the FDP in terms of excessive probability rather than expectation, and therefore are stricter when it comes to labelling genes as significant. For example, at control level $\alpha = 0.1$, 221 genes are rejected by the BH procedure (4.4), but only 125 of them are rejected by the procedure (5.1) with $\text{pFDEP}_{0.1} \leq \gamma = 0.05$. That is, in order to have false discovery proportion below 0.1 with 95% of chance, only about 1/2 the genes can be rejected by procedure (5.1). The nonparametric VDP and LR procedures yield much more conservative results. Across all the range of values of $\alpha \leq 0.02$ and $\gamma \leq 0.2$, the VDP procedure rejects at most 10, and the LR procedure rejects at most 5 genes. Finally, by using the estimate 0.67 of $1 - \pi$ instead of 1, each procedure yields more genes declared significant at the same level of α and γ, leading to improved power.

7. Remarks

The work can be extended in several directions. First, our results are obtained under a mixture model where p-values are independent. This simple set-

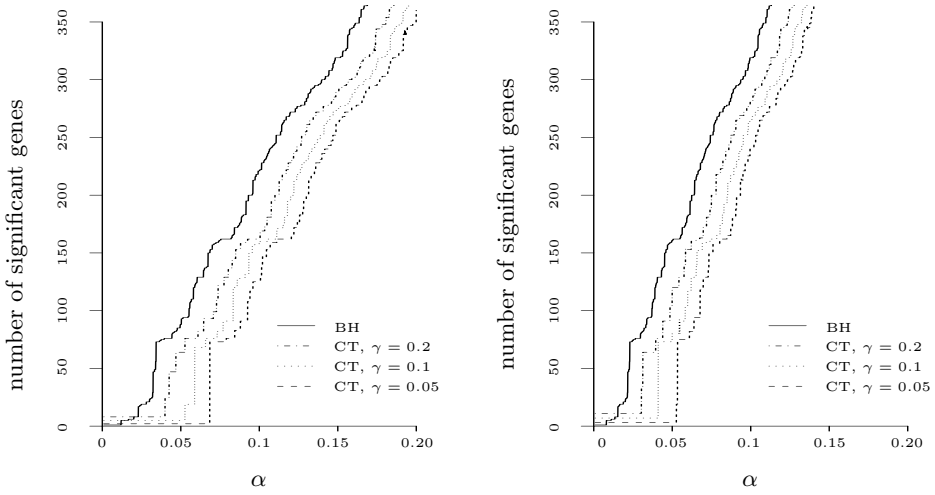

Figure 1. Comparison of procedures applied to gene expression data.

Left: BH and CT stand for procedures (4.4) and (5.1), respectively. *Right*: BH and CT stand for procedures (4.3) and (4.7) with $k_n = 8$, respectively, both using $1 - \pi \approx 0.67$. For $\gamma = 0.2, 0.1$ and 0.05, the number of significant genes are 8−10, 6−7 and 2 with VDP, 1−5, 1−2, and 1 with LR.

ting helps us in understanding the intrinsic nature of pFDR and pFDEP and the mechanisms that can be exploited to achieve adaptive control. The insights gained here are valuable for further investigation of the control of pFDR or pFDEP in more general settings. A potentially important idea is to estimate the distribution or the mean and variance of the number of false rejections given the total number of rejections. Resampling techniques can be employed for this purpose in multiple testing problems with dependent p-values.

Second, we have used point estimates of pFDR and pFDEP for fixed rejection regions to construct procedures to control pFDR and pFDEP. It would be interesting to study the variations of the point estimates, and to investigate how to incorporate interval estimates (Storey (2002)).

Third, the fraction of false nulls, if unknown, is underestimated by 0 in the proposed procedures. This fraction may be estimated from data, and a less biased estimate may yield higher power given the same level of control; see Benjamini and Hochberg (2000) and Storey (2002), and our example in Section 6.2. It would be interesting to investigate how to estimate this fraction, and important to evaluate how uncertainty in the estimate might affect the proposed procedures.

Appendix: Selected Theoretical Details

A1. Proof of Proposition 3.1

(a) According to (2.1), let $\delta = (\delta_1, \ldots, \delta_n)$ be a multiple testing procedure. Then, given $\xi_k = t_k$, $k = 1, \ldots, n$, the set of rejected nulls is uniquely determined. Thus R is uniquely determined as well. Because (ξ_1, H_1), $\ldots$, (ξ_n, H_n) are independent, by Proposition 2.1(b), for each k,

$$P(H_k = 0 \,|\, \xi_j = t_j,\, j = 1, \ldots, n) = P(H_k = 0 \,|\, \xi_k = t_k) = \beta_{t_k} \geq \beta_*.$$

Then for $t_1, \ldots, t_n$ with $R > 0$,

$$\begin{aligned} &E\left(V/R \,\middle|\, \xi_j = t_j,\, t = 1, \ldots, n\right) \\ =&E\left(\frac{1}{R}\sum_{k=1}^{n}(1 - H_k) \,\middle|\, \xi_j = t_j,\, t = 1, \ldots, n\right) \\ =&\frac{1}{R}\sum_{k=1}^{n} E(1 - H_k \,|\, \xi_j = t_j, t = 1, \ldots, n) \geq \beta_*. \end{aligned}$$

Take the expectation over those $t_1, \ldots, t_n$ for which $R > 0$. Then pFDR $\geq \beta_*$. If $G(t)$ is concave, then $\alpha_* = \beta_*$ and hence pFDR $\geq \alpha_*$.

(b) Given $R = r > 0$, let $\xi_{i_1}, \ldots, \xi_{i_r}$ be the rejected p-values. By Proposition 2.1, the H_{i_k} are independent of each other and $P(H_{i_k} = 0) \geq \beta_*$. As a result, $V = \sum_{k=1}^{r}(1 - H_{i_k})$ dominates $Z_1 + \ldots + Z_r$, where $Z_1, \ldots, Z_r$ are iid $\sim$ Bernoulli(β_*). Therefore,

$$P(V > \alpha r \,|\, R = r) \geq P(Z_1 + \ldots + Z_r > \alpha r) \geq 1 - \sup_{k \geq 1} \mathrm{pbin}(\alpha k; k, \alpha).$$

Because the procedure is nontrivial, i.e., $P(R > 0) > 0$, taking expectation over $r > 0$, we get $P(V > \alpha R \,|\, R > 0) \geq 1 - \sup_{k \geq 1} \mathrm{pbin}(\alpha k; k, \alpha)$.

A2. Proof of Proposition 3.2

Given $R = r > 1$, let $\xi_{i_1}, \ldots, \xi_{i_r}$ be the p-values associated with rejected nulls. As in the proof of Proposition 3.1, V stochastically dominates $Z_1 + \ldots + Z_r$, where $Z_1, \ldots, Z_r$ are iid $\sim$ Bernoulli(β_*). Then

$$\begin{aligned} &P\Big(\frac{V}{R} \leq \alpha \,|\, R = r,\ \xi_{i_k},\ k = 1, \ldots, r, \ \text{ are rejected } p\text{-values}\Big) \\ \leq &P(Z_1 + \ldots + Z_r \leq \alpha n). \end{aligned}$$

Since $\alpha < \beta_*$, $I = \sup_{t<0}[\alpha t - \log E(e^{tZ_1})] > 0$. On the other hand, by Chernoff's inequality, $P(Z_1 + \ldots + Z_r \leq \alpha r) \leq e^{-rI}$. Because the bound is independent of ξ_{i_k}, $P(V/R \leq \alpha \,|\, R = r) \leq e^{-rI}$. Fix $c > 0$. For any n,

$$P\Big(\frac{V}{R} \leq \alpha, R \geq c \log n\Big) \leq \max_{r \geq c \log n} e^{-rI} \leq n^{-cI}.$$

If $c > 1/I$, then $P_n := P(V/R \le \alpha, R \ge c \log n)$ has a finite sum over n and hence, by the Borel-Cantelli lemma, with probability 1, for all large n, the events that $V/R \le \alpha$ and $R \ge c \log n$ cannot happen at the same time. This completes the proof.

A3. Sketch proofs of main theorems

Theorems 4.2 and 5.1 deal with the subcritical case. The proof of Theorem 4.2 follows closely the heuristics given in Section 4.2. The proof of Theorem 5.1 follows the same idea. The only subtle point is that the fraction of false nulls is increasingly smaller. In order for the Central Limit Theorem (CLT) to still apply, we need to show (1) although the number of rejections R is $o(n)$, it converges to ∞, and (2) the number of false rejections, V, closely follows $\mathrm{Bin}(R, \alpha)$ as $n \to \infty$. These two facts guarantee that the argument based on the CLT in the heuristics still holds, hence leading to the desired convergence. Condition (5.4) will be used to establish the two facts.

Theorems 4.3 and 5.2 deal with the supercritical case. For Theorem 4.3, first, R is bounded (in probability) as $n \to \infty$. Indeed, if $R \to \infty$, then by the weak law of large numbers (WLLN), it can be shown that $(1-\pi)n\xi_{n:(k\vee k_n)}/(k \vee k_n) \to (1-\pi)/F'(0) = \beta_*$. Then by (4.7), one would have $\mathrm{qbin}(\theta; R, \beta_*) \le \alpha R$, where θ is a positive constant. However, because $\beta_* > \alpha$, by the WLLN, $\mathrm{qbin}(\theta; k\beta_*) = (1 + o(1))\beta_* k > \alpha k$ as $k \to \infty$. This contradiction implies that R is finite. This is the main step of the proof. Then, because $k_n \to \infty$, (4.7) is asymptotically the same as (4.7b), which implies that R must be the largest k satisfying $\mathrm{qbin}(1-\gamma; k, \beta_*) \le \alpha k$. The remainder of the proof of Theorem 4.3 follows from this observation. Theorem 5.2 can be proved in a similar way.

For more details of the proofs, see Supplemental Materials.

Acknowledgement

This work was partially supported by NIH Grants MH68028 and DC007206 to Z. Chi and NIH Grants to Z. Tan.

References

Abramovich, F., Benjamini, Y., Donoho, D. and Johnstone, I. (2006). Adapting to unknown sparsity by controlling the false discovery rate. *Ann. Statist.* **34**, 584-653.

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.

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.

Broberg, P. (2005). A comparative review of estimates of the proportion unchanged genes and the false discovery rate. *BMC Bioinformatics* **6**, 199.

Chi, Z. (2007). On the performance of FDR control: constraints and a partial solution. *Ann. Statist.* **35**, 1409-1431.

Donoho, D. and Jin, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32**, 962-994.

Donoho, D. and Jin, J. (2005). Asymptotic minimaxity of false discovery rate thresholding for sparse exponential data. *Ann. Statist.* **34**, 2980-3018.

Efron, B., Tibshirani, R., Storey, J. D. and Tusher, V. G. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.* **96**, 1151-1160.

Finner, H. and Roters, M. (2001). On the false discovery rate and expected type I errors. *Biometrical J.* **43**, 985-1005.

Genovese, C. and Wasserman, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *J. Roy. Statist. Soc. Ser. B* **64**, 499-517.

Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.* **32**, 1035-1061.

Genovese, C. and Wasserman, L. (2006). Exceedance control of the false discovery proportion. *J. Amer. Statist. Assoc.* **101**, 1408-1417.

Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Kallioniemi, O. P., Wilfond, B., Borg, A. and Trent, J. (2001). Gene-expression profiles in hereditary breast cancer. *N. Engl. J. Medicine* **344**, 539-548.

Langaas, M., Lindqvist, B. H. and Ferkingstad, E. (2005). Estimating the proportion of true null hypotheses, with application to DNA microarray data. *J. Roy. Statist. Soc. Ser. B* **67**, 555-572.

Lehmann, E. L. and Romano, J. P. (2005). Generalizations of the familywise error rate. *Ann. Statist.* **33**, 1138-1154.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*. John Wiley, New York.

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

Storey, J. D. (2002). A direct approach to false discovery rates. *J. Roy. Statist. Soc. Ser. B* **64**, 479-498.

Storey, J. D. (2003). The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann. Statist.* **31**, 2012-2035.

Storey, J. D. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. USA* **100**, 9440-9445.

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. Roy. Statist. Soc. Ser. B* **66**, 187-205.

van der Laan, M. J., Dudoit, S. and Pollard, K. S. (2004). Multiple testing. Part III: Procedures for control of the generalized family-wise error rate and proportion of false positives. Technical Report 141, Division of Biostatistics, UC Berkeley.

Department of Statistics, University of Connecticut, 215 Glenbrook Road, U-4120, Storrs, CT 06269, U.S.A.

E-mail: zchi@stat.uconn.edu.

Deaprtment of Statistics, Rutgers University, 110 Frelinghuysen Road, Piscataway, NJ 08854, U.S.A.

E-mail: ztan@stat.rutgers.edu

(Received June 2006; accepted February 2007)