Positive false discovery proportions: intrinsic bounds and adaptive control

Zhiyi Chi and Zhiqiang Tan

University of Connecticut and The Johns Hopkins University

Running title: Bounds and control of pFDR and pFDEP

December 22, 2006

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 first show that, unlike FDR, these two positive error rates may not be controllable at an arbitrarily low level. Given a multiple testing problem, there can exist positive intrinsic lower bounds, such that no procedures can ever attain a pFDR or pFDEP level below the corresponding bound. To reduce misinterpretations of testing results, we then 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 also 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, Benjamin and Hochberg (1995) introduced 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 & Hochberg 2000, Genovese & Wasserman 2002, 2004, 2006, Lehmann & Romano 2005, Storey 2002, 2003, Storey et al. 2004, van der Laan et al. 2004, and references therein).

FDR is defined as the expectation of false discovery proportion (FDP), which is the proportion of falsely rejected nulls among all the rejected ones 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 & 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 & Wasserman 2006), and both built upon 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 & 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 previous procedures that can 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 & Wasserman (2002) and Chi (2006) 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, say, $\beta_* = 0.9$, then, 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 propose that a desirable procedure should be adaptive, that is, it automatically achieves a specified control level whenever the level is attainable, and avoid making any rejections otherwise.

To develop a methodology of adaptive control, we shall consider two scenarios. In the first one, the fraction of false nulls π is known. We shall propose procedures that are adaptive in the above sense 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 being replaced by 0. The procedures are still adaptive, but become conservative. On the other hand, they can achieve adaptive control even when false nulls become sparse in the sense that 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 so-called "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 & Wasserman 2002, 2004, 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 $\xi_i = \xi_j$. In other words, the decision on each null is completely determined by its p-value regardless of the indices of the nulls.

By the meaning of p-value, it is often required for 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 the number 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

– False Discovery Rate: $\text{FDR} = E[V/(R \vee 1)]$,

– Positive false Discovery Rate: $\text{pFDR} = E[V/R \,|\, R > 0]$,

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

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

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

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

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

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

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 "local FDR" (Efron et al. 2001, Broberg 2005). Therefore, the lowest attainable FDR and local FDR are

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

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 that $\alpha_* \le \beta_*$ 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

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

(b) for $1 \le k \le n$, conditioning on $R_t = k$, V_t follows $\mathrm{Bin}(k, \alpha_t)$, i.e., the binomial distribution with k trials and success probability α_t per trial.

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 our proposed procedures in Sections 4 and 5.

Notations. 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 $p = 0$, then $\mathrm{Bin}(n, p)$ is defined to be the singular measure concentrated at 0. Denote by Φ the distribution function of $N(0, 1)$. Finally, adopt the convention that $\max \emptyset = 0$.

As we shall 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, $(\xi_1^{(n)}, H_1^{(n)})$, …, $(\xi_n^{(n)}, H_n^{(n)})$ 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$_\alpha$. 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,

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

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

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

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

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* $\mathrm{pFDEP}_\alpha \geq 1 - \max_{k \geq 1} \mathrm{pbin}(\alpha k; k, \beta_*) > 0$.

Note that the lower bounds in Proposition 3.1 are intrinsic to a multiple testing problem, regardless of the procedure applied. The lower bounds reveals 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 translated into results on pFDR directly. 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 to below β_* (Genovese & Wasserman 2002, Chi 2006). 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 $R \sim \epsilon n$ at the same time, 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 shall distinguish between the two cases. We call the case $\alpha > \beta_*$ subcritical and the case $\alpha < \beta_*$ supercritical. The critical case where $\alpha = \beta_*$ rarely occurs in practice and will not be 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

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

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 Uniform$(0,1)$ and Q_1 that of 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 $\sup_{0\le t\le 1} g(t) = \lim_{t\to 0} g(t) = b$, $\beta_* > 0$.

(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

$$Q_0(x) = \frac{1}{2} + \frac{\arctan x}{\pi}, \quad Q_1(x) = Q_0(x/c), \qquad -\infty < x < \infty,$$

and

$$g(t) = \frac{c}{1 + (c^2 - 1)\sin^2(\pi t)}, \qquad 0 \le t \le 1,$$

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. Now 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 is, 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 design of any multiple testing procedure, but only due 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 in general. Under the assumption, (1) $\alpha_t \le \beta_t$, and both are increasing on [0, 1]; and (2) $\alpha_* = \beta_*$, and both are equal to $\alpha_0 = (1 - \pi)/(1 - \pi + \pi g(0))$.

4 Procedure: fixed known fraction of false nulls

In this section, we consider the scenario where $\pi_n \equiv \pi \in (0, 1)$ is known, and $G_n \equiv G$ is unknown and has 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 & Hochberg 2000, Langaas et al. 2005, 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), $E(V_t/R_t \mid R_t)$ and $P(V_t/R_t > \alpha \mid R_t)$ can be estimated respectively by

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

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

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

Consider the following idea to control (p)FDR based on the estimate (4.1): reject the R smallest p-values, where

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

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

$$R = \max\left\{k \geq 1 : n\xi_{n:k} \leq \alpha k\right\}. \tag{4.4}$$

Benjamini & 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

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

This procedure is structurally similar to procedure (4.3), except that quantiles of binomial distributions are used rather than the 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 is adaptive to both subcritical and supercritical conditions. That is, if $\alpha > \alpha_*$, the procedure attains pFDR $\leq \alpha$; 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$_\alpha \leq \gamma$; and if $\alpha < \alpha_*$ and $\gamma < \gamma_*$, it almost never makes rejections.

In order to modify (4.3) and (4.5) to achieve the 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

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

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 $\boldsymbol{\pi}$: *Reject the R smallest p-values, where*

$$R = \max\left\{k \geq 1 : \frac{(1-\pi)n\xi_{n:(k \vee k_n)}}{k \vee k_n} \leq \alpha\right\}. \tag{4.6}$$

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 know $\boldsymbol{\pi}$: *Reject the R smallest p-values, where*

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

with

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

Note that any $k_n \to \infty$ of order $o(n)$ can be in (4.7) to yield the same asymptotic behavior of the procedure. In practice, we have used $k_n = c \log n$, with $c > 0$ being a constant.

Overall, procedure (4.7) accommodates both sub- and supercritical cases automatically. When $\alpha > \alpha_*$, R converges to a finite random variable, and (4.7) is asymptotically identical to

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

where $\Gamma(t) = \Phi[\sqrt{(1-(1-\pi)t)/(1-\alpha)}\,\Phi^*(1-\gamma)]$ (cf. Theorems 4.2). On the other hand, when $\alpha < \alpha_*$, R grows roughly linearly in n, and (4.7) is asymptotically identical to (cf. Theorem 4.3)

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

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 $P(V \le \alpha R) \approx 1-\gamma$. Let θ be the correct replacement. By the definition of R, $\operatorname{qbin}(\theta; R, \zeta n/R) \approx \alpha R$, where $\zeta = (1-\pi)\xi_{n:R}$. Now

$$\{V \le \alpha R\} = \left\{\frac{V-\zeta n}{\sigma_R} \le \frac{\operatorname{qbin}(\theta; R, \zeta n/R) - \zeta n}{\sigma_R}\right\},$$

where for each k, σ_k is the standard deviation of $\operatorname{Bin}(n, \zeta n/k)$. By 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, $V \sim \operatorname{Bin}(n, \zeta)$, whose standard deviation is $\sigma' = \sqrt{\zeta(1-\zeta)n}$. Because $\sigma_R \approx \sqrt{\zeta n(1-\zeta n/R)} \approx \sqrt{\zeta n(1-\alpha)}$,

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

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 $\alpha \geq 1-\pi$ is trivial since all the null hypotheses can be rejected with pFDR $= 1-\pi \leq \alpha$. 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 & Wasserman 2002, Chi 2006).

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 procedure* (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 in pFDR control. Indeed, pFDR is asymptotically controlled exactly at the target control level under the subcritical condition, whereas the number of rejections tends to 0 under the supercritical condition.

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 identical to 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. In contrast, for the BH procedure (4.3), by Proposition 4.1 (a), $\mathrm{pFDEP}_\alpha \to 1/2$. 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. In contrast, 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) $R/n \xrightarrow{\mathrm{P}} F(u^*)$, $\mathrm{pFDR} \to \alpha$, and $\mathrm{pFDEP}_\alpha \to \gamma$.

(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 $\ell_0 = \max\{k \ge 1 : \mathrm{qbin}(1-\gamma; k, \alpha_*) \le \alpha k\}$ and $\ell_1 = \max\{k \ge 1 : \mathrm{qbin}(1-\gamma; k, \alpha_*) + 1 \le \alpha k\}$. The following statements hold for procedure* (4.7) *as $n \to \infty$.*

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

(b) For $\ell = \ell_0,\ \ell_1$, $V \mid R = \ell \xrightarrow{d} \mathrm{Bin}(\ell, \alpha_*)$.

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

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

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

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

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 maintains the same power as the BH procedure (4.3), but achieves a stricter control in terms of pFDR and pFDEP.

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

5 Procedure: unknown fraction of false nulls and increasingly sparse false nulls

We now consider the scenario where both π_n and G_n are unknown, with G_n having a continuous and strictly decreasing density. We will restrict our discussion to pFDEP control. Similarly, pFDR control can be treated.

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 $\boldsymbol{\pi_n}$: *Reject the R smallest p-values, where*

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

with $k_n \sim c \log n$ *as* $n \to \infty$, *where* $c > 0$ *is a constant, and*

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

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 $c > 0$ being a 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 & Hochberg 2000, Storey et al. 2004). For this reason, the sub- and supercritical cases need to be written as

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

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

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

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

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

First, consider the scenario where $\pi_n \equiv \pi > 0$ 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 & Wasserman (2002) and Chi (2006).

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) (see Benjamini & Hochberg 2000, Finner & Roters 2001, Storey 2002, 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 n is relatively not large enough. It is also of interest to address the same question for the BH procedure (4.4) and 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, Donoho & 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 as well.

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

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

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

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

and denote the limit by α_*. Let $\ell_0 = \max\{k \ge 1 : \mathrm{qbin}(1-\gamma; k, \alpha_*) \le \alpha k\}$ and $\ell_1 = \max\{k \ge 1 : \mathrm{qbin}(1-\gamma; k, \alpha_*) + 1 \le \alpha k\}$. 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, $V \mid R = \ell \xrightarrow{d} \mathrm{Bin}(\ell, \alpha_*)$.

(c) If $\gamma > \gamma_*$, then $\varlimsup \mathrm{pFDEP}_\alpha \le \gamma$. 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 a set of real 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 & 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

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

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

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

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

6.1 Simulation study

Throughout, $\alpha = .2$ and $\gamma = .05$. 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 the quantities are computed as Monte Carlo averages from 10000 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 4 subcritical configurations, with $(\pi, b) = (.1, 100)$, $(.05, 199)$, $(.02, 496)$, and $(.01, 991)$, respectively. In these configurations, the alternative distribution G is increasingly concentrated near 0, but the fraction π of false nulls is

Table 1: Simulation results for procedure (5.1): subcritical case, $\alpha = .2$, $\gamma = .05$. For each pair (n, k_n), results are obtained for 4 different (π, b), with π the fraction of the alternative 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	.1	.05	.02	.01	.1	.05	.02	.01
	$n = 2000,\ k_n = 7$				$n = 20000,\ k_n = 9$			
$P(R > 0)$	.9960	.8502	.4588	.2795	1	1	1	.9951
pFDR	.13	.13	.12	.12	.17	.17	.16	.15
FDR	.13	.11	.054	.034	.17	.17	.16	.15
pFDEP_α	.011	.046	.10	.14	0	.01	.034	.045
FDEP_α	.011	.039	.047	.038	0	.01	.034	.045
Power	.70	.48	.19	.12	.85	.83	.78	.70
	$n = 2000,\ k_n = 14$				$n = 20000,\ k_n = 18$			
$P(R > 0)$	.9950	.8289	.3545	.1091	1	1	1	.9952
pFDR	.13	.13	.13	.15	.17	.17	.16	.15
FDR	.13	.11	.045	.017	.17	.17	.16	.15
pFDEP_α	.011	.046	.11	.20	0	.01	.034	.045
FDEP_α	.011	.038	.039	.022	0	.01	.034	.045
Power	.70	.48	.18	.067	.85	.83	.78	.70

decreasing to 0, so that $\alpha_*/(1 - \pi)$ is fixed at $1/10.9 = .09$. Recall that procedure (5.1) involves a sequence $k_n \sim c \log n$, with $c > 0$ a 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 $\text{pFDEP}_{0.2}$ at level $\gamma = .05$ for all the four configurations when $n = 20000$, but not for $\pi = .01$ or .02 when $n = 2000$. In the latter cases, π is considerably 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 of $R = 0$ is .54 or greater for $\pi = .01$ or .02 and $n = 2000$.

Table 2 summarizes the simulation results for procedure (5.1) under 4

Table 2: Simulation results for procedure (5.1): supercritical case, $\alpha = .2$, $\gamma = .05$.

π	.1	.05	.02	.01	.1	.05	.02	.01
		$n = 2000,\ k_n = 7$				$n = 20000,\ k_n = 9$		
$P(R > 0)$	0	.0002	0	0	0	0	.0002	0
pFDR	NA	.30	NA	NA	NA	NA	.55	NA
FDR	0	<.0001	0	0	0	0	.0001	0
pFDEP$_\alpha$	NA	.50	NA	NA	NA	NA	1	NA
FDEP$_\alpha$	0	.0001	0	0	0	0	.0002	0
Power	0	<.0001	0	0	0	0	<.0001	0
		$n = 2000,\ k_n = 14$				$n = 20000,\ k_n = 18$		
$P(R > 0)$	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$_\alpha$	NA	NA	NA	NA	NA	NA	NA	NA
FDEP$_\alpha$	0	0	0	0	0	0	0	0
Power	0	0	0	0	0	0	0	0

supercritical configurations, with $(\pi, b) = (.1, 10)$, $(.05, 19)$, $(.02, 46)$, and $(.01, 91)$, 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 $1/1.9 = .53$. In this case, it is not possible for any procedure to control pFDEP$_{0.1}$ at level $\gamma = .05$. 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 10000.

Finally, Table 3 summarizes the simulation results for procedure (5.1), BH, VDP, and LR with $(\pi, b) = (.05, 199)$ and $(.05, 19)$. The results are qualitatively similar under the other configurations studied in the previous two simulations. Note that the four procedures are not strictly comparable as they are designed for different purposes: procedure (5.1) for pFDEP con-

Table 3: Comparison of simulation results for procedures: $\alpha = .2$, $\gamma = .05$.

$\pi = .05$	CT	BH	VDP	LR	CT	BH	VDP	LR
Subcritical		$n = 2000,\ k_n = 7$				$n = 20000,\ k_n = 9$		
$P(R > 0)$	.8502	1	.4260	.0768	1	1	.4168	.0615
pFDR	.13	.19	.084	.10	.17	.19	.091	.082
FDR	.11	.19	.036	.0077	.17	.19	.038	.0051
pFDEP_α	.046	.39	.10	.10	.01	.22	.12	.086
FDEP_α	.039	.39	.044	.0078	.01	.22	.048	.0053
Power	.48	.89	.0051	.0007	.83	.89	.0005	.0001
Supercritical		$n = 2000,\ k_n = 7$				$n = 20000,\ k_n = 9$		
$P(R > 0)$	.0002	.3749	.0892	.0151	0	.3862	.0953	.0114
pFDR	.30	.49	.48	.52	NA	.50	.51	.49
FDR	$<.0001$	.18	.043	.0079	0	.19	.048	.0056
pFDEP_α	.50	.69	.49	.52	NA	.70	.52	.50
FDEP_α	.0001	.26	.044	.0079	0	.27	.050	.0057
Power	$<.0001$	.0049	.0005	$<.0001$	0	.0005	$<.0001$	$<.0001$

trol, 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, 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.

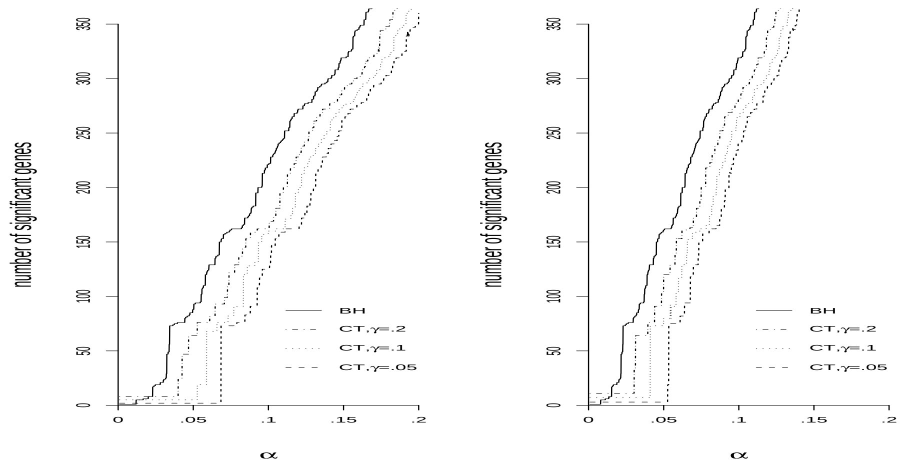
6.2 Application to gene expression

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 3226 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 3170 genes on the $\log_2$ scale.

First, following Storey & Tibshirani (2003), we use 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 apply procedure (5.1) as well as the BH (4.4), VDP, and LR procedures to the resulting p-values. For this example, Storey & Tibshirani (2003) estimated that 67% of the genes are not differentially expressed. Based on this estimate, we also apply procedures (4.7) and (4.3) with $1 - \pi \approx .67$. For all the adaptive procedures, we use $k_n = \lfloor \log n \rfloor$ 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 $\alpha \leq .2$ and $\gamma \leq .2$. 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

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 $1 - \pi \approx .67$. For $\gamma = .2, .1$ and .05, the number of significant genes are 8–10, 6–7 and 2 with VDP, and 1–5, 1–2, and 1 with LR.

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 $\mathrm{pFDEP}_{0.1} \le \gamma = .05$. That is, in order to have false discovery proportion below .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 $\alpha \le .02$ and $\gamma \le .2$, the VDP procedure rejects at most 10, and the LR procedure rejects at most 5 genes. Finally, by using the estimate .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 Conclusion

In many fields of science and engineering, such as genomics and imaging, multiple hypothesis testing is becoming increasingly important for exploratory data analysis. The error rates pFDR and pFDEP are particularly relevant to follow-up studies conducted once positive findings are obtained. In this article, we investigate the controllability of pFDR and pFDEP, and propose several adaptive procedures to control them respectively.

The work can be extended in several directions. First, our results are obtained under a mixture model where p-values are independent. This simple setting 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 is interesting to study the variations of the point estimates, and investigate how to incorporate interval estimates in developing more reliable procedures (Storey 2002).

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

Acknowledgement

This work was partially supported by NIH Grants MH68028-01 and DC007206-01A1 to Z. Chi and NIH Grants to Z. Tan.

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), ..., (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 $R > 0$,

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

Take expectation over $t_1, \ldots, t_n$ 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, 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_*), the Bernoulli distribution with probability of 1 equal to β_*. Therefore,

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

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

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

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

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}, we get $P(V/R \leq \alpha \,|\, R = r) \leq e^{-rI}$. Therefore, given $c > 0$, for any n,

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

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

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 $\text{Bin}(R, \alpha)$ 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 $\text{qbin}(\theta; R, \beta_*) \le \alpha R$, where $\theta > 0$ is a constant. However, because $\beta_* > \alpha$, by the WLLN, $\text{qbin}(\theta; k\beta_*) = (1 + o(1))\beta_* k > \alpha k$ 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 $\text{qbin}(1-\gamma; k, \beta_*) \le \alpha k$. The remaining 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.

References

Abramovich, F., Benjamini, Y., Donoho, D. & Johnstone, I. (2006), 'Adapting to unknown sparsity by controlling the false discovery rate', *Ann. Statist.*

34(2), 584–653.

Benjamini, Y. & Hochberg, Y. (1995), 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *J. R. Stat. Soc. Ser. B* **57**(1), 289–300.

Benjamini, Y. & Hochberg, Y. (2000), 'On the adaptive control of the false discovery rate in multiple testing with independent statistics', *J. Educ. Behav. Statist.* **25**(1), 60–83.

Broberg, P. (2005), 'A comparative review of estimates of the proportion unchanged genes and the false discovery rate', *BMC Bioinformatics* **6**(1), 199.

Chi, Z. (2006), 'On the performance of FDR control: constraints and a partial solution', *Ann. Statist.* Accepted.

Donoho, D. & Jin, J. (2004), 'Higher criticism for detecting sparse heterogeneous mixtures exponential data', *Ann. Statist.* **32**(3), 962–994.

Donoho, D. & Jin, J. (2005), 'Asymptotic minimaxity of FDR for sparse exponential model', *Ann. Statist.* **32**(3), 962–994.

Efron, B., Tibshirani, R., Storey, J. D. & Tusher, V. G. (2001), 'Empirical Bayes analysis of a microarray experiment', *J. Amer. Statist. Assoc.* **96**(456), 1151–1160.

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

Genovese, C. & Wasserman, L. (2002), 'Operating characteristics and extensions of the false discovery rate procedure', *J. R. Stat. Soc. Ser. B* **64**(3), 499–517.

Genovese, C. & Wasserman, L. (2004), 'A stochastic process approach to false discovery control', *Ann. Statist.* **32**(3), 1035–1061.

Genovese, C. & Wasserman, L. (2006), 'Exceedance control of the false discovery proportion', *J. Amer. Statist. Assoc.* **101**(476), 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. & Trent, J. (2001), 'Gene-expression profiles in hereditary breast cancer', *N. Engl. J. Medicine* **344**(8), 539–548.

Langaas, M., Lindqvist, B. H. & Ferkingstad, E. (2005), 'Estimating the proportion of true null hypotheses, with application to DNA microarray data', *J. R. Stat. Soc. Ser. B* **67**(4), 555–572.

Lehmann, E. L. & Romano, J. P. (2005), 'Generalizations of the familywise error rate', *Ann. Statist.* **33**(3), 1138–1154.

Shorack, G. R. & Wellner, J. A. (1986), *Empirical processes with applications to statistics*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York.

Simes, R. J. (1986), 'An improved Bonferroni procedure for multiple tests of significance', *Biometrika* **73**(3), 751–754.

Storey, J. D. (2002), 'A direct approach to false discovery rates', *J. R. Stat. Soc. Ser. B* **64**(3), 479–498.

Storey, J. D. (2003), 'The positive false discovery rate: a Bayesian interpretation and the q-value', *Ann. Statist.* **31**(6), 2012–2035.

Storey, J. D. & Tibshirani, R. (2003), 'Statistical significance for genomewide studies', *Proc. Natl. Acad. Sci. USA* **100**(16), 9440–9445.

Storey, J. D., Taylor, J. E. & 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* **66**(1), 187–205.

van der Laan, M. J., Dudoit, S. & 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, UC Berkeley, Division of Biostatistics.

Department of Statistics, University of Connecticut, Storrs, CT 06269.
E-mail: zchi@stat.uconn.edu.

Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205.
E-mail: ztan@jhsph.edu.

Supplemental Materials

for "Positive false discovery proportions: intrinsic bounds and adaptive control" by Z. Chi and Z. Tan

S1 Notations

In this Supplemental Materials, we prove the main theoretical results in the article, i.e., Theorems 4.2, 4.3, 5.1 and 5.2. The proofs for the first two theorems are similar to those for the other two. Since the latter ones are more of interest to applications, they will be demonstrated in detail. The proof of Theorems 4.2 and 4.3 will only be outlined afterwards.

The following notations will be used. If X_n and Y_n are random variables, then $X_n \ge_p Y_n$, $X_n \le_p Y_n$, and $X_n \sim_p Y_n$ denote $P(X_n \ge Y_n) \to 1$, $P(X_n \le Y_n) \to 1$, and $X_n/Y_n \xrightarrow{\text{P}} 1$, respectively. The notation $X_n = o_p(Y_n)$ means "$|X_n| \le_p \epsilon |Y_n|$ for any $\epsilon > 0$", whereas $X_n = O_p(Y_n)$ means "for any $\epsilon > 0$, there are $M > 0$ and $n_0 > 0$, such that $P(|X_n| \ge M|Y_n|) < \epsilon$ for all $n \ge n_0$". Finally, when necessary, to make explicit the dependence on the number n of tested hypotheses, we use a superscript to index a random variable. For example, denote by $R^{(n)}$ the number of rejections when there are n null hypotheses.

It will be easier to work with continuous time to prove the theorems. For procedure (5.1), given p-values $\xi_1^{(n)}, \ldots, \xi_n^{(n)}$, the $R_{\tau_n}^{(n)}$ smallest ones are rejected, where

$$\tau_n = \sup_{t \in [0,1]} \left\{ \operatorname{qbin}\left(\Gamma_*(t); R_t^{(n)}, \frac{n(t \vee \xi_{n:k_n})}{R_t^{(n)} \vee k_n} \wedge 1 \right) \le \alpha(R_t^{(n)} \vee 1) \right\}. \tag{S1.1}$$

Henceforth denote

$$q_n(t; z) := \operatorname{qbin}\left(z;\, R_t^{(n)},\, (nt/R_t^{(n)}) \wedge 1\right).$$

For brevity, write $\tau = \tau_n$, $R_t = R_t^{(n)}$, and $V_t = V_t^{(n)}$. By the notations in Sections 4 and 5, $R = R_\tau$ and $V = V_\tau$. The same relationship holds for the BH procedure (4.4), except that

$$\tau = \tau_n = \sup\{t \in [0,1] : nt \le R_t\} \tag{S1.2}$$

S2 Subcritical case with increasingly sparse false nulls

Define $\eta_t = \eta^{(n)} = nt/R_t$, $\theta_t = \theta_t^{(n)} = t/F_n(t)$, and $\rho = 1/\alpha - 1 > 0$. Then $q_n(z;t) = \mathrm{qbin}(z; R_t,\ \eta_t \wedge 1)$ and by $u_n = \alpha F_n(u_n)$,

$$\theta_{u_n} = \alpha, \quad (\pi_n + \rho)u_n = \pi_n G_n(u_n). \tag{S2.1}$$

Suppose the subcritical condition (5.4) is satisfied. Then we have the following lemmas.

Lemma S2.1 *For procedure* (5.1) *and the BH procedure* (4.4), $\tau/u_n \xrightarrow{\mathrm{P}} 1$ *and*

$$(a)\ \frac{G_n(\tau)}{G_n(u_n)} \xrightarrow{\mathrm{P}} 1, \quad (b)\ \theta_\tau \xrightarrow{\mathrm{P}} \alpha, \quad (c)\ \eta_\tau \xrightarrow{\mathrm{P}} \alpha, \quad (d)\ \frac{R_\tau}{(\log n)^4} \xrightarrow{\mathrm{P}} \infty.$$

Lemma S2.2 *For both procedures,* $\theta_\tau R_\tau - n\tau = \alpha R_{u_n} - nu_n + o_p(\sqrt{nu_n})$.

Lemma S2.3 *For procedure* (5.1), $\alpha R_\tau = q_n(\tau; \Gamma(\tau)) + O_p(1)$, *and for the BH procedure* (4.4), $\alpha R_\tau = n\tau$.

Let $R_{t-} = R_{t-}^{(n)}$ and $V_{t-} = V_{t-}^{(n)}$ be the numbers of rejected nulls and rejected true nulls, respectively, whose p-values are strictly less than t.

Lemma S2.4 *Given* $t \in (0,1)$ *and* $k > 0$, *for procedure* (5.1), *conditioning on* $\tau = t$ *and* $R_{\tau-} = k$, $V_{\tau-} \sim \mathrm{Bin}(k, t/F_n(t))$. *The statement holds as well for the BH procedure* (4.4).

Recall that if $p_n \in (0,1)$ satisfies $np_n(1-p_n) \to \infty$, then by Lindeberg's CLT, for $S_n \sim \mathrm{Bin}(n, p_n)$ and $z \in (0,1)$,

$$\frac{S_n - np_n}{\sqrt{np_n(1-p_n)}} \xrightarrow{d} N(0,1), \quad \frac{\mathrm{qbin}(z;\, n,\, p_n) - np_n}{\sqrt{np_n(1-p_n)}} \to \Phi^*(z). \tag{S2.2}$$

Proof of Theorem 5.1. Assume the Lemmas are true for now. We show (a)–(d) in sequel.

(a) By Lemma S2.1, $\eta_\tau \xrightarrow{\mathrm{P}} \alpha$ and

$$n\tau(1-\eta_\tau) = R_\tau \eta_\tau (1-\eta_\tau) \sim_p \alpha(1-\alpha) R_\tau \xrightarrow{\mathrm{P}} \infty.$$

Then by Lemma S2.3,

$$\begin{aligned} P(V_\tau \le \alpha R_\tau) &= P\left(V_\tau \le q_n(\tau; \Gamma(\tau)) + O_p(1)\right) + o(1) \\ &= P\left(\frac{V_\tau - n\tau}{\sqrt{n\tau(1-\eta_\tau)}} \le \frac{q_n(\tau;\Gamma(\tau)) - n\tau}{\sqrt{n\tau(1-\eta_\tau)}} + o_p(1) \right) + o(1). \end{aligned}$$

Since $q_n(\tau; \Gamma(\tau)) = \mathrm{qbin}(\Gamma(\tau); R_\tau, \eta_\tau)$ and $n\tau = R_\tau \eta_\tau$, Lindeberg's CLT yields

$$\frac{q_n(\tau;\Gamma(\tau)) - n\tau}{\sqrt{n\tau(1-\eta_\tau)}} \sim_p \Phi^*(\Gamma(\tau)) = \sqrt{\frac{1-\tau}{1-\alpha}} \Phi^*(1-\gamma)\,. \tag{S2.3}$$

Write $(V_\tau - n\tau)/\sqrt{n\tau(1-\eta_\tau)} = Z_1 Z + Z_2$, where

$$Z_1 = \frac{V_\tau - \theta_\tau R_\tau}{\sqrt{\theta_\tau(1-\theta_\tau)R_\tau}}, \quad Z = \sqrt{\frac{\theta_\tau(1-\theta_\tau)R_\tau}{n\tau(1-\eta_\tau)}}, \quad Z_2 = \frac{\theta_\tau R_\tau - n\tau}{\sqrt{n\tau(1-\eta_\tau)}}\,.$$

By Lemma S2.4, conditioning on $\tau = t$ and $R_{\tau-} = k$, $V_{\tau-} \sim \mathrm{Bin}(k, \theta_t)$. Since $R_\tau - R_{\tau-}$, $V_\tau - V_{\tau-} \in \{0,1\}$, and $\theta_\tau R_\tau \to \infty$, it follows that $Z_1 \xrightarrow{d} N(0,1)$. By Lemma S2.1, $Z \xrightarrow{\mathrm{P}} 1$. By Lemma S2.2, $Z_2 = Z_2' + o_p(1)$, where $Z_2' = (\alpha R_{u_n} - nu_n)/\sqrt{n\tau(1-\eta_\tau)}$. From $F_n(u_n) = u_n/\alpha \to 0$, $R_{u_n} \sim$

$\text{Bin}(n, F_n(u_n))$, $\tau/u_n \xrightarrow{\text{P}} 1$, and $\eta_\tau \xrightarrow{\text{P}} \alpha$, it follows that

$$Z_2' = \frac{\alpha(R_{u_n} - nF_n(u_n))}{\sqrt{nF_n(u_n)(1 - F_n(u_n))}} \sqrt{\frac{u_n(1 - F_n(u_n))}{\alpha\tau\,(1 - \eta_\tau)}} \xrightarrow{d} \sqrt{\frac{\alpha}{1 - \alpha}} N(0, 1)$$

and hence $Z_2 \xrightarrow{d} \sqrt{\alpha/(1-\alpha)} N(0, 1)$.

Let U_1, U_2 be iid $\sim N(0, 1)$. We next show

$$(Z_1, Z_2) \xrightarrow{d} (U_1, \sqrt{\alpha/(1-\alpha)}\, U_2).$$

Let $f(x, y) = E(e^{ixZ_1 + iyZ_2})$. Then by Lemma S2.4 and CLT, for any $a_n \to \infty$ and $t_n \in (0, 1)$, as long as $a_n\theta_{t_n}(1 - \theta_{t_n}) \to \infty$,

$$\lim_{n\to\infty} E(e^{ixZ_1} \mid R_\tau = a_n,\ \tau = t_n) = e^{-x^2/2}.$$

Since Z_2 is a deterministic function of τ and R_τ, by $R_\tau\theta_\tau(1 - \theta_\tau) \xrightarrow{\text{P}} \infty$ and dominated convergence,

$$\begin{aligned} E(e^{ixZ_1 + iyZ_2}) &= E(E(e^{ixZ_1 + iyZ_2} \mid R_\tau, \tau)) \\ &\sim e^{-x^2/2} E(e^{iyZ_2}) \to \exp\left\{-\frac{x^2}{2} - \frac{\alpha y^2}{2(1 - \alpha)}\right\}. \end{aligned}$$

Combining all the above results, it follows that

$$\frac{V_\tau - n\tau}{\sqrt{n\tau(1 - \eta_\tau)}} \xrightarrow{d} \frac{U}{\sqrt{1 - \alpha}} \quad \text{with } \ U \sim N(0, 1)\,,$$

which, together with (S2.3) and $\tau \xrightarrow{\text{P}} 0$, implies

$$P(V_\tau \le \alpha R_\tau) \sim P\left(U \le \sqrt{1 - \alpha}\Phi^*(\Gamma(\tau))\right) \to 1 - \gamma.$$

This completes the proof of part (a).

(b) This directly follows from Lemma S2.1(d).

(c) For the BH procedure (4.4), from (S1.2), $0 \le \alpha R_\tau - n\tau \le \alpha$. By Lemma S2.2,

$$P(V_\tau \le \alpha R_\tau) = P(V_\tau \le n\tau + O(1)) = P(Z_1 \le Z_2 + o(1))$$

with

$$Z_1 = (V_\tau - \theta_\tau R_\tau)/\sqrt{\theta_\tau(1-\theta_\tau)R_\tau},$$
$$Z_2 = (n\tau - \theta_\tau R_\tau)/\sqrt{\theta_\tau(1-\theta_\tau)R_\tau}.$$

Following the argument for part (a), $(Z_1, Z_2) \xrightarrow{d} (U_1, \sqrt{\alpha/(1-\alpha)}\, U_2)$, where U_1 and U_2 are iid $\sim N(0,1)$. As a result, $P(V_\tau \le \alpha R_\tau) \to 1/2$.

(d) From (5.4), $n\pi_n \to \infty$. Since $n - N_0 \sim \text{Bin}(n, \pi_n)$, then by the weak law of large numbers (WLLN), $(n - N_0)/(n\pi_n) \xrightarrow{\text{P}} 1$. By Lemmas S2.1, S2.2, and S2.4,

$$\frac{R_\tau}{nF_n(u_n)} \xrightarrow{\text{P}} 1, \qquad \frac{R_\tau - V_\tau}{R_\tau} \xrightarrow{\text{P}} 1-\alpha.$$

Therefore, by (S2.1) and $\pi_n \to 0$,

$$\frac{\psi_n}{G_n(u_n)} = \frac{R_\tau - V_\tau}{G_n(u_n)(n - N_0)} \sim_p \frac{(1-\alpha)nF_n(u_n)}{G_n(u_n)n\pi_n} = \frac{\rho u_n}{\pi_n G_n(u_n)} \xrightarrow{\text{P}} 1.$$

The proof for the BH procedure (4.4) is similar and hence is omitted. $\square$

To show the lemmas, the following representation of the p-values $\xi_1^{(n)}$, $\ldots$, $\xi_n^{(n)}$ will be used. Let $\zeta_k^{(n)} = F_n(\xi_k^{(n)})$. Then $\zeta_1^{(n)}, \ldots, \zeta_n^{(n)}$ are iid $\sim U(0,1)$. Let

$$W_t := W_t^{(n)} = \#\left\{k \ge 1 : \zeta_k^{(n)} \le t\right\}, \tag{S2.4}$$

so that $R_t = W_{F_n(t)}$. Recall the following result (Shorack & Wellner 1986, p 600). Let $b_n = \sqrt{2\log\log n}$, $c_n = 2\log\log n + \log\sqrt{\log\log n} - \log\sqrt{4\pi}$, and

$Z_t = (W_t - nt)/\sqrt{nt(1-t)}$. Then for any $x \in (-\infty, \infty)$, as $n \to \infty$

$$P\left(b_n \sup_{t\in[0,1]} |Z_t| \ge c_n + x\right) \to e^{-4e^x}. \tag{S2.5}$$

From (5.4), it can be seen that as $n \to \infty$, $nu_n/(\log n)^4 \to \infty$ and

$$\frac{\sqrt{n}}{(\log n)^2} \frac{\pi_n^2 [\lambda G_n(u_n) - G_n(\lambda u_n)]}{\sqrt{u_n}} \to \begin{cases} \infty, & \text{if } \lambda > 1 \\ -\infty, & \text{if } 0 < \lambda < 1. \end{cases} \tag{S2.6}$$

Indeed, because G_n is strictly concave, if $\lambda > 1$, (5.4) implies that

$$\frac{\sqrt{nu_n}}{(\log n)^2} \pi_n \left(\lambda - \frac{G_n(\lambda u_n)}{G_n(u_n)}\right) \to \infty.$$

Then the first limit in (S2.6) follows by (S2.1). The second limit similarly holds.

Proof of Lemma S2.1 We will only show the Lemma for procedure (5.1). The proof for the BH procedure (4.4) is similar.

The main part of the proof is devoted to $\tau/u_n \xrightarrow{\text{P}} 1$. Denote

$$d_n = \sqrt{n \log\log n}, \quad f_n(t) := \pi_n G_n(t) - (\pi_n + \rho)t.$$

Because G_n is strictly concave, so is f_n. By (S2.1), $f_n(u_n) = 0$. Also, $f_n(t) >(<)\ 0$ for $t <(>)\ u_n$. Given $\lambda > 1$, let $v_n = u_n/\lambda$. On the one hand, by (S2.2), $q_n(v_n; \Gamma(v_n)) \le nv_n + \sqrt{nv_n} A_n$, with $A_n \to \Phi^*(\Gamma(v_n))$. On the other, by (S2.5), for large n, $R_{v_n} = W_{F_n(v_n)} \ge_p nF_n(v_n) - 2d_n\sqrt{F_n(v_n)}$. Therefore, by $F_n(v_n) \le \alpha v_n$ and $F_n(v_n) \le F_n(u_n) = u_n/\alpha$,

$$\begin{aligned} &R_{v_n} - \frac{1}{\alpha} q_n(v_n; \Gamma(v_n)) \\ &\ge_p n\left(F_n(v_n) - (1+\rho)v_n\right) - \sqrt{nF_n(v_n)/\alpha}\, A_n - 2d_n\sqrt{F_n(v_n)} \\ &\ge_p nf_n(v_n) - 3d_n\sqrt{u_n/\alpha}. \end{aligned}$$

By $\pi_n + \rho = \pi_n G_n(u_n)/u_n$,

$$nf_n(v_n) = n\pi_n \left(G_n(v_n) - \frac{\pi_n G_n(u_n)}{u_n} v_n \right) \ge n\pi_n \left(G_n(u_n/\lambda) - \frac{G_n(u_n)}{\lambda} \right).$$

Then by (S2.6), $nf_n(v_n) - 3d_n\sqrt{u_n/\alpha} \xrightarrow{\mathrm{P}} \infty$, yielding

$$\alpha R_{v_n} - q_n(v_n; \Gamma(v_n)) \xrightarrow{\mathrm{P}} \infty$$

and hence $P(\tau > v_n) \to 1$.

Now let $w_n = \lambda u_n$. Then for all $t \ge w_n$, $F_n(t) < t/\alpha$. Similar to the above argument, the probability that

$$R_t - \frac{1}{\alpha} q_n(t; \Gamma(t)) \le f_n(t) + 3d_n\sqrt{F_n(t)} \le f_n(t) + 3d_n\sqrt{t/\alpha}, \quad \text{all } t \ge w_n$$

tends to 1. Because $f_n(t)$ is concave, it is upper bounded by $f_n(w_n)t/w_n$. Note $f_n(w_n) < 0$. Therefore, the probability that

$$\begin{aligned} R_t - \frac{1}{\alpha} q_n(t; \Gamma(t)) &\le f_n(w_n)t/w_n + 3d_n\sqrt{t/\alpha} \\ &\le \sqrt{t/w_n}\, \underbrace{\left(f_n(w_n) + 3d_n\sqrt{w_n/\alpha} \right)}_{x_n} \end{aligned}$$

tends to 1. Similar to the above argument, $x_n \to -\infty$ by (S2.6). Then $P(\tau \le w_n) \to 1$. Together with $P(\tau > v_n) \to 1$ and $\lambda > 1$ being arbitrary, $\tau/u_n \xrightarrow{\mathrm{P}} 1$. Now we can show parts (a)–(d) in sequel.

(a) Because G_n is concave,

$$G_n(u_n/\lambda) > G_n(u_n)/\lambda, \quad G_n(\lambda u_n) < \lambda G_n(u_n).$$

Then from $\tau/u_n \xrightarrow{\mathrm{P}} 1$, $P(G_n(u_n)/\lambda \le G_n(\tau) < \lambda G_n(u_n)) \to 1$ and hence $G_n(\tau)/G_n(u_n) \xrightarrow{\mathrm{P}} 1$.

(b) Since $F_n(t) = (1 - \pi_n)t + \pi_n G_n(t)$, by the above results,

$$F_n(\tau)/F_n(u_n) \xrightarrow{\mathrm{P}} 1.$$

By $F_n(u_n) = u_n/\alpha$, $\theta_\tau = \tau/F_n(\tau) \xrightarrow{\mathrm{P}} \alpha$.

(c) Given $\lambda > 1$, define v_n and w_n as above. By $P(R_\tau > R_{v_n}) \to 1$ and $P(\tau < w_n) \to 1$, $P(\eta_\tau < nw_n/R_{v_n}) \to 1$. On the other hand, since $nv_n \to \infty$, by the CLT in (S2.2), $R_{v_n} = W_{F_n(v_n)} \sim nF_n(v_n)$ and $w_n/F_n(v_n) = \lambda^2 v_n/F_n(v_n) \le \alpha\lambda^2$. Then $P(\eta_\tau < \alpha\lambda^2) \to 1$. Similarly, $P(\eta_\tau > \alpha/\lambda^2) \to 1$. Hence $\eta_\tau \xrightarrow{\mathrm{P}} \alpha$.

(d) This follows from $R_\tau \ge R_{v_n} \sim nF_n(v_n) \ge nv_n$ and $nv_n/(\log n)^4 \to \infty$.
$\square$

By the weak version of Hungarian construction (Shorack & Wellner 1986, p.494), for each n, there exist a Brownian bridge $B_t^{(n)} \overset{d}{=} Z_t - tZ_1$ and a stochastic process $r_t^{(n)}$ defined on the same probability space as $\zeta_1^{(n)}, \ldots, \zeta_n^{(n)}$, where Z_t is a standard Brownian motion, such that $\sup_{t\in[0,1]} |r_t^{(n)}| = O_p(1)$ and $W_t = nt + \sqrt{n}B_t^{(n)} + r_t^{(n)}(\log n)^2$.

Proof of Lemma S2.2 By $R_t = W_{F_n(t)}$ and the Hungarian construction,

$$\theta_t R_t - nt = \sqrt{n}\theta_t B_{F_n(t)}^{(n)} + (\log n)^2 \theta_t r_{F_n(t)}^{(n)}.$$

Note $\theta_{u_n} = \alpha$ and by Lemma S2.1, $\theta_\tau = O_p(1)$. Since $(\log n)^2/\sqrt{nu_n} \to 0$ and $r_t^{(n)}$ is bounded, in order to show Lemma S2.2 for τ, it is enough to show that

$$\frac{1}{\sqrt{u_n}}\left[\theta_\tau B_{F_n(\tau)}^{(n)} - \alpha B_{F_n(u_n)}^{(n)}\right] \xrightarrow{\mathrm{P}} 0.$$

Write the left hand side as $I_1 + \alpha I_2$, where $I_1 = (\theta_\tau - \alpha) B^{(n)}_{F_n(\tau)}/\sqrt{u_n}$ and $I_2 = (B^{(n)}_{F_n(\tau)} - B^{(n)}_{F_n(u_n)})/\sqrt{u_n}$. Given $\lambda > 1$, since $u_n \to 0$, $P(\tau < \lambda u_n) \to 1$ and $F_n(\lambda u_n) < \lambda F_n(u_n) = \lambda u_n/\alpha$, it is seen that $|B^{(n)}_{F_n(\tau)}|$ asymptotically is dominated by $\sup_{t \le \lambda u_n/\alpha} |B^{(n)}_t|$, hence stochastically dominated by $\sup_{t\le \lambda u_n/\alpha} |Z_t| + (\lambda u_n/\alpha)|Z_1|$. Then $B^{(n)}_{F_n(\tau)}/\sqrt{u_n} = O_p(1)$. By $\theta_\tau \xrightarrow{\text{P}} \alpha$, $I_1 \xrightarrow{\text{P}} 0$.

Similarly, letting $D_n = \lambda u_n - u_n/\lambda$, $B^{(n)}_{F_n(\tau)} - B^{(n)}_{F_n(u_n)}$ asymptotically is stochastically dominated by $\sup_{t\in[0,D_n]} |Z_t| + D_n|Z_1|$. Therefore, I_2 asymptotically is stochastically dominated by $\sqrt{\lambda - 1/\lambda}\sup_{t\in[0,1]} |Z_t| + o_p(1)$. Because λ is arbitrary, $I_2 \xrightarrow{\text{P}} 0$. $\square$

Recall that $\text{qbin}(z; n, p)$ is increasing and left-continuous in z and p respectively; for $z, p \in (0, 1)$,

$$\lim_{x\downarrow z} \text{qbin}(x; n, p), \quad \lim_{x\downarrow p} \text{qbin}(z; n, x) \in \{\text{qbin}(z; n, p), \text{qbin}(z; n, p) + 1\};$$

and $\text{qbin}(z; n, p) \le \text{qbin}(z; n - 1, p) + 1$.

Proof of Lemma S2.3 By the definition of τ, when $\tau > 0$ and $R_\tau > 0$, for all $t > \tau$, $q_n(t; \Gamma(t)) > \alpha R_t$. If $t - \tau > 0$ is small enough, $R_t = R_\tau$. Since $\Gamma(t)$ is decreasing in t, $\text{qbin}(\Gamma(\tau); R_\tau, nt/R_\tau) \ge \alpha R_\tau$. Letting $t \downarrow \tau$ then yields $q_n(\tau, \Gamma(\tau)) \ge \alpha R_\tau - 1$.

On the other hand, there is a sequence $t_j \uparrow \tau$, such that $q_n(t_j; \Gamma(t_j)) \le \alpha R_{t_j}$. If R_t is continuous at τ, then for large j, $R_{t_j} = R_\tau$ and letting $j \to \infty$ yields $q_n(\tau, \Gamma(\tau)) \le \alpha R_\tau$. If R_t has a jump at τ, then for large j, $R_{t_j} = R_\tau - 1$ and letting $j \to \infty$ yields

$$\begin{aligned} &\text{qbin}\left(\Gamma(\tau); R_\tau - 1, \frac{n\tau}{R_\tau - 1}\right) \le \alpha R_\tau + 1 \\ \Longrightarrow\ &\text{qbin}\left(\Gamma(\tau); R_\tau, \frac{n\tau}{R_\tau - 1}\right) \le \alpha R_\tau + 2. \end{aligned}$$

Then $q_n(\tau, \Gamma(\tau)) - \alpha R_\tau \le \mathrm{qbin}\left(\Gamma(\tau); R_\tau, n\tau/(R_\tau - 1)\right) - \alpha R_\tau \le 2$.

The proof that for the BH procedure (4.4) is standard so is omitted. $\square$

Proof of Lemma S2.4 Let $\mathcal{F}_t = \sigma(\mathbf{1}\{\xi_i \le s\}, s \in [t, 1], i = 1, \ldots, n)$. Then for t running backward from 1 to 0, $\mathcal{F}_t$ consist a filtration and for both procedure (5.1) and the BH procedure (4.4), τ is a stopping time with respect to the filtration. In particular, $\{\tau \ge t\} \cap \{R_{\tau-} = k\} \in \mathcal{F}_t$. Let $i_1, \ldots, i_{R_{t-}}$ be the random indices of those ξ_i that are strictly less than t. By the independence of $(\xi_1, H_1), \ldots, (\xi_n, H_n)$ and $V_{t-} = H_{i_1} + \ldots + H_{R_{t-}}$, it is not difficult to see that for any t, $A \in \mathcal{F}_t$, $k \ge 0$, and $n_1 < \ldots < n_k$, conditioning on $E = \{R_{t-} = k, i_1 = n_1, \ldots i_k = n_k\}$, V_{t-} and A are independent, i.e. $P(\{V_{t-} = v\} \cap A \,|\, E) = P(\{V_{t-} = v\} \,|\, E) \times P(A \,|\, E)$. Consequently, $P(V_{t-} = v \,|\, \tau = t, E) = P(V_{t-} = v \,|\, E)$. By Proposition 2.1, the right end is $P(S = v)$, with $S \sim \mathrm{Bin}(k, nt/G_n(t))$. Since the conditional probability does not involve $n_1, \ldots, n_k$, then $P(V_{t-} = v \,|\, \tau = t, R_{t-} = k) = P(S = v)$. $\square$

S3 Supercritical case with increasing sparsity of false nulls

Let $\zeta_k^{(n)}$ be defined as in (S2.4). We need two lemmas in order to prove Theorem 5.2.

Lemma S3.1 *Given $p_0 \in (0, 1)$, for any $\epsilon > 0$,*

$$\lim_{n \to \infty} \sup_{p \in [p_0, 1]} P\left(|X_{1,p} + \cdots + X_{n,p} - np| \ge \epsilon n\right) = 0.$$

where for each p, $X_{1,p}, X_{2,p}, \ldots$ are iid $\sim \mathrm{Bernoulli}(p)$.

Lemma S3.2 *If $k_n \le n$ satisfies $k_n \to \infty$, then*

$$\sup_{k_n \le k \le n} \left| \frac{\zeta_{n:k}}{k/n} - 1 \right| \xrightarrow{\mathrm{P}} 0.$$

Proof of Theorem 5.2 Assume the lemmas are true for now. Fix $\epsilon > 0$ such that $(1-\epsilon)^2 \underline{\lim}_{n\to\infty} \alpha_*^{(n)} > \alpha$. Then by condition (5.5), for n large enough, $(1-\epsilon)^2/F_n'(0) = (1-\epsilon)^2\alpha_*^{(n)}/(1-\pi_n) > \alpha$.

First, we show that for some $K_0 > 0$, $P(R_\tau < K_0) \to 1$. Let $m_n(t) = R_t \vee k_n$. Then

$$\frac{n(t \vee \xi_{n:k_n})}{R_t \vee k_n} \ge \frac{nF_n^*\left(\zeta_{m_n(t)}\right)}{m_n(t)}. \tag{S3.1}$$

By the selection of k_n, $m_n(t) \to \infty$. Then by the convexity of $F_n^*(x)$ (because F_n is concave) and Lemma S3.2,

$$\begin{aligned} \frac{nF_n^*\left(\zeta_{m_n(t)}\right)}{m_n(t)} &\ge_p \frac{F_n^*\left((1-\epsilon)m_n(t)/n\right)}{m_n(t)/n} \\ &\ge (1-\epsilon)(F_n^*)'(0) = \frac{1-\epsilon}{F_n'(0)} \ge \frac{\alpha}{1-\epsilon} \end{aligned} \tag{S3.2}$$

and hence by Lemma S3.1, there is $K_0 > 0$, such that for all $K \ge K_0$,

$$\mathrm{qbin}\left(1-\gamma;\, K,\, \frac{1-\epsilon}{F_n'(0)}\right) > \frac{(1-\epsilon)^2 K}{F_n'(0)} > \alpha K.$$

Combined with (S3.1) and (S3.2), this implies

$$P\left(\bigcap_{t: R_t \ge K_0} \left\{\mathrm{qbin}\left(1-\gamma;\, R_t,\, \frac{n(t\vee \xi_{n:k_n})}{R_t \vee k_n}\right) > \alpha R_t\right\}\right) \to 1.$$

As a result, $P\left(R_\tau < K_0\right) \to 1$.

Now suppose condition (5.6) is satisfied. We show parts (a)–(d) in sequel.

(a) Note that

$$R_\tau = \max\left\{k \ge 1 : \text{qbin}\left(1-\gamma; k, \frac{n\xi_{n:(k\vee k_n)}}{k\vee k_n} \wedge 1\right) \le \alpha k\right\}$$

Because $P(R_\tau < K_0) \to 1$ and $k_n \to \infty$, $P(R_\tau = R_n') \to 1$, where

$$R_n' = \max\left\{k \ge 1 : \text{qbin}\left(1-\gamma; k,\, p_n \wedge 1\right) \le \alpha k\right\}$$

with $p_n = n\xi_{n:k_n}/k_n$. Since $p_n \xrightarrow{\text{P}} \alpha_*$, by the properties of qbin as listed before the proof of Lemma S2.3, part (a) then follows from

$$P\left(\text{qbin}(1-\gamma; k, p_n) \in \{\text{qbin}(1-\gamma; k, \alpha_*),\, \text{qbin}(1-\gamma; k, \alpha_*) + 1\}\right) \to 1.$$

(b) Since F_n is concave, given $R_\tau = \ell > 0$ and $\tau = t$, V_τ is stochastically dominated by $\text{Bin}(\ell,\, (1-\pi_n)t/F_n(t))$, but stochastically dominates $\text{Bin}(\ell,\, (1-\pi_n)/F_n'(0))$. Because $\tau \xrightarrow{\text{P}} 0$, $(1-\pi_n)\tau/F_n(\tau) \xrightarrow{\text{P}} \alpha_*$. Part (b) therefore follows.

(c) Let $Z_1, Z_2, \ldots$ be iid $\sim$ Bernoulli(α_*). If $\gamma < \gamma_*$, then $\ell_0 = 0$, or else there were $k > 0$ such that $P(Z_1 + \ldots + Z_k \le \alpha k) \ge 1-\gamma$. Then $\gamma \ge 1 - P(Z_1 + \ldots + Z_k \le \alpha_k) \ge \gamma_*$, which is a contradiction. It is clear that $\ell_1 = 0$. Therefore, by part (a), $R_\tau \xrightarrow{\text{P}} 0$. On the other hand, if $\gamma > \gamma_*$, then $\ell_0 > 0$. By part (b), $P(V_\tau \le \alpha R_\tau \,|\, R_\tau = \ell_0) \to P(Z_1 + \ldots + Z_{\ell_0} \le \alpha\ell_0) \ge 1-\gamma$, and hence $\overline{\lim}_n P(V_\tau/R_\tau > \alpha \,|\, R_\tau = \ell_0) \le \gamma$. The case $R_\tau = \ell_1$ can be similarly shown as long as $\ell_1 > 0$. This completes the proof of (c).

(d) For both procedure (5.1) and the BH procedure (4.4), in order to show that their respective powers tend to 0, by $P(R \le K_0) \to 1$, it is enough to show $n - N_0 \xrightarrow{\text{P}} \infty$. Denote $s_n = F_n^*(k_n/n)$. Since $M_n := \#\{i \le n : \xi_i^{(n)} \le s_n,\ H_i^{(n)} = 1\} \sim \text{Bernoulli}(n, \pi_n G_n(s_n))$ and $M_n \le n - N_0$, it is enough to show $n\pi_n G_n(s_n) \to \infty$. Since $k_n/n = F_n(s_n) = (1-\pi_n)s_n + \pi_n G_n(s_n)$ and $s_n/F_n(s_n) \to \alpha_* < 1$,

$$\frac{\pi_n G_n(s_n)}{k_n/n} = 1 - \frac{(1-\pi_n)s_n}{F_n(s_n)} \to 1 - \alpha_* > 0,$$

yielding $n\pi_n G_n(s_n) \sim (1-\alpha_*)k_n \to \infty$. □

Proof of Lemma S3.1 It is enough to show

$$\lim_{n\to\infty} \sup_{p\in[p_0,1]} P\left(X_{1,p} + \cdots + X_{n,p} > (p+\epsilon)n\right) = 0, \quad \text{and}$$

$$\lim_{n\to\infty} \sup_{p\in[p_0,1]} P\left(X_{1,p} + \cdots + X_{n,p} < (p-\epsilon)n\right) = 0.$$

We will only show the first limit. The second one can be shown similarly.

Clearly, when $p \geq 1-\epsilon$, $P(X_{1,p} + \cdots + X_{n,p} > (p+\epsilon)) = 0$. If $p < 1-\epsilon$, then by Chernoff's inequality, $P\left(X_{1,p} + \cdots + X_{n,p} > (p+\epsilon)n\right) \leq e^{-nI(p)}$, where $I(p) = \sup_{t>0}\left((p+\epsilon)t - \Lambda_p(t)\right)$, with $\Lambda_p(t) = \log(1-p+pe^t)$. Since $\Lambda_p(t)$ is convex and $\Lambda_p'(0) = p$, $I(p) > 0$. It can be verified that $I(p)$ is continuous on $[0, 1-\epsilon)$. Letting $I(p) = \infty$ for $p \geq 1-\epsilon$, it follows that $\inf_{p\geq p_0} I(p) > 0$, which implies the limit. □

Proof of Lemma S3.2 $\xi_{n:1}, \ldots, \xi_{n:n}$ have the same joint distribution as

$$\left(\frac{S_1}{S_{n+1}}, \ldots, \frac{S_n}{S_{n+1}}\right)$$

where $S_k = U_1 + \cdots + U_k$ and $U_1, U_2, \ldots$ are iid $\sim$ Exp(1). By the WLLN, $S_{n+1}/n \xrightarrow{\text{P}} 1$. Therefore, it is enough to show $\sup_{k\geq k_n} |S_k/k - 1| \xrightarrow{\text{P}} 0$, which follows from the strong law of large numbers (SLLN). □

S4 Nonsparse case

Proof of Theorem 4.1 The proof of part (a) is omitted because it follows closely Genovese & Wasserman (2002). For part (b), let R' be the number of projections in (4.3). Then by Proposition 4.1, $P(R' > k_n) \to 0$. Since $P(R > 0) \le P(R' > k_n)$, part (b) follows. $\square$

To prove Theorem 4.2, we need the following standard result for empirical processes.

Lemma S4.1 *Suppose τ_n is a sequence of random variables taking values in $[0, 1]$, such that for some $u \in (0, 1)$, $\tau_n \xrightarrow{d} u$ as $n \to \infty$. Then, letting $\pi_0 = 1 - \pi$,*

$$\frac{V_{\tau_n} - n\pi_0\tau_n}{\sqrt{n\pi_0 u(1 - \pi_0 u)}} \xrightarrow{d} N(0, 1).$$

Following the proof for the sparse case, for procedure (4.7), define

$$\tau_n = \sup\left\{t \in [0, 1] : \mathrm{qbin}\left(\Gamma_*(t);\, R_t,\, \frac{\pi_0 n(t \vee \xi_{n:k_n})}{R_t \vee k_n} \wedge 1\right) \le \alpha(R_t \vee 1)\right\}$$

and for the BH procedure (4.3), define $\tau_n = \sup\{t \in [0, 1] : \pi_0 nt \le R_t\}$. Then following the same notations, $R = R_\tau$ and $V = V_\tau$.

The proof of Theorem 4.2 follows closely that of Theorem 5.1, so we only give its sketch.

Proof of Theorem 4.2 (a) Following Genovese & Wasserman (2002), $\tau \xrightarrow{\mathrm{P}} u^*$, with $u^* \in (0, 1)$ the only positive solution to $\pi_0 u = \alpha F(u)$. By the definition of R and $P(\xi_i \le u^*,\ H_i = 1) = F(u^*)$, from WLLN, it follows that $R/n \xrightarrow{\mathrm{P}} F(u^*) > 0$ and hence pFDR $\sim$ FDR $= \alpha$. Furthermore, $\Gamma_*(t)$ can be replaced with $\Gamma(t)$ and by Lemma S2.3,

$$P(V_\tau \le \alpha R_\tau) = P\left(V_\tau \le \mathrm{qbin}\left(\Gamma(\tau); R_\tau, \frac{\pi_0 n\tau}{R_\tau} \wedge 1\right) + O_p(1)\right)$$

Denote the binomial quantile on the right hand side by K. Applying the CLT to the binomial distributions, from $\tau \xrightarrow{d} u^*$, it follows that

$$\frac{K - \pi_0 n\tau}{\sqrt{n\pi_0 u^*(1-\alpha)}} \sim_p \frac{K - \pi_0 n\tau}{\sqrt{\pi_0 n\tau(1 - \pi_0 n\tau / R_\tau)}} \xrightarrow{\mathrm{P}} \Phi^*(\Gamma(u^*)) \, .$$

Combining this with Lemma S4.1 yields

$$P(V_\tau \leq \alpha R_\tau) \to P\left(\sqrt{\frac{1-\pi_0 u^*}{1-\alpha}} Z \leq \Phi^*(\Gamma(u^*))\right) = 1 - \gamma.$$

(b) Since $k_n/n \to 0$ whereas $R/n \xrightarrow{d} F(u^*) > 0$, part (b) easily follows. $\square$

Proof of Proposition 4.1 (a) Following the proof of Theorem 4.2 (a),

$$P(V_\tau \leq \alpha R_\tau) \to P\left(\sqrt{\frac{1-\pi_0 u^*}{1-\alpha}} Z \leq 0\right) = \frac{1}{2}\,,$$

where, for the BH procedure (4.3),

$$\tau = \tau_n = \sup\left\{t \in [0,1] : \pi(1-\pi)nt \leq R_t\right\}.$$

(b) See Chi (2006). $\square$

The proof of Theorem 4.3 is almost identical to that of Theorem 5.2 and so is omitted.

Proof of Proposition 4.2 When $\alpha \in (\alpha_*, 1-\pi)$, then $\tau \xrightarrow{\mathrm{P}} u^*$. Because $R - V = \#\left\{k \leq n : H_k = 1,\, \xi_k^{(n)} \leq \tau\right\}$ and $n - N_0 = \#\left\{k \leq n : H_k = 1\right\}$, by the WLLN, $\psi_n \xrightarrow{\mathrm{P}} P(\xi \leq u^* \,|\, H = 1) = G(u^*)$. On the other hand, when $\alpha < \alpha_*$, then for procedures (4.6) and (4.7), by Theorem 4.1 and 4.2, it is apparent that $\psi_n = O_p(1/n)$, and for the BH procedure (4.3), from Chi (2006), $\psi_n \xrightarrow{\mathrm{P}} 0$ as well. $\square$