



Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Asymptotically optimal sequential FDR and pFDR control with (or without) prior information on the number of signals

Xinrui He ^a, Jay Bartroff ^{b,*}^a Acumen LLC and the SPHERE Institute, Burlingame, CA, USA^b Department of Mathematics, University of Southern California, Los Angeles, CA, USA

ARTICLE INFO

Article history:

Received 9 December 2019

Received in revised form 1 May 2020

Accepted 6 May 2020

Available online 15 May 2020

Keywords:

False discovery rate

Positive false discovery rate

Sequential analysis

Streaming data

ABSTRACT

We investigate asymptotically optimal multiple testing procedures for streams of sequential data in the context of prior information on the number of false null hypotheses (“signals”). We show that the “gap” and “gap-intersection” procedures, recently proposed and shown by Song and Fellouris (2017) to be asymptotically optimal for controlling type 1 and 2 familywise error rates (FWEs), are also asymptotically optimal for controlling FDR/FNR when their critical values are appropriately adjusted. Generalizing this result, we show that these procedures, again with appropriately adjusted critical values, are asymptotically optimal for controlling *any* multiple testing error metric that is bounded between multiples of FWE in a certain sense. This class of metrics includes FDR/FNR but also pFDR/pFNR, the per-comparison and per-family error rates, and the false positive rate. Our analysis includes asymptotic regimes in which the number of null hypotheses approaches ∞ as the type 1 and 2 error metrics approach 0.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

For decades, the problem of how to efficiently and powerfully test multiple statistical hypotheses while controlling some notion of type 1 error frequency has been fundamental and active in the statistics methodology literature. The majority of this research has concerned testing procedures which operate on fixed-sample data, typically in the form of a collection of p -values, one for each null hypothesis, which are combined in some way to reach reject/accept decisions for each null. Recently, driven by applications where data is streaming or arrives sequentially, multiple testing procedures that can handle sequential data have been proposed and studied. Applications with data of this type include the analysis of streaming internet data (Wegman and Marchette, 2003), multiple channel signal detection in sensor networks (Dragalin et al., 1999; Mei, 2008), high throughput sequencing technology (Jiang and Salzman, 2012), and multi-arm and multiple endpoint clinical trials (Bartroff and Lai, 2010).

Existing multiple testing procedures for sequential data occur in essentially two forms. In one, the individual data streams can be terminated at different times (e.g., Bartroff and Song, 2014, 2015; Bartroff, 2018; Malloy and Nowak, 2014). In the other form, termination of sampling must occur at the same time for all streams (e.g., De and Baron, 2012a,b; Song and Fellouris, 2017, 2019). The applications mentioned in the previous paragraph span both of these forms. Recently, a third form of “sequential” multiple testing procedure has been studied in which not the data, but rather the hypotheses themselves arrive sequentially in time, each with its own fixed-sample p -value. Javanmard and Montanari (2018) proposed

* Corresponding author.

E-mail address: bartroff@usc.edu (J. Bartroff).

an FDR-controlling procedure in this setup, which [Chen and Arias-Castro \(2017\)](#) showed to be optimal under certain distributional assumptions. A recent manuscript by [Zrnic et al. \(2018\)](#) studies optimality more generally in this setup.

The current paper adopts the second form described above in which all the data streams are terminated at the same time, and we investigate the optimal choice of that stopping rule subject to desired bounds on the type 1 and 2 error metrics, asymptotically as these bounds approach zero at arbitrary rates. This is done under the condition of prior information on the number of false null hypotheses (“signals”) in the form of a known number of signals, or known bounds on this number. The latter case of known bounds on the number of signals includes the non-informative setting with lower bound 0 and upper bound equal to the total number of null hypotheses.¹ Our work springs from and generalizes the results of [Song and Fellouris \(2017\)](#) who found asymptotically optimal procedures in these settings when the error metrics are type 1 and 2 familywise error rates (FWEs). By modifying the procedures of Song and Fellouris, we find the corresponding asymptotically optimal procedures for controlling the false discovery rate (FDR) and its type 2 analog, the false non-discovery rate (FNR). Further, we are able to find the asymptotically optimal procedures for controlling *any* multiple testing error metric that is bounded between multiples of FWE in a certain sense, which includes FDR/FNR, the positive false discovery and non-discovery rates (pFDR and pFNR), the per comparison error rate, and other metrics. Further, we are able to consider asymptotic regimes in which the number J of null hypotheses approaches ∞ .

After introducing notation and describing our approach in Section 2, the case of the number of signals known exactly is addressed in Section 3 where first the general result for arbitrary error metrics is stated, followed by its application to FDR/FNR and pFDR/pFNR. The case of bounds on the number of signals is addressed in Section 4 where, again, the general result for arbitrary error metrics followed by its application to FDR/FNR and pFDR/pFNR. Simulation studies of procedures for FDR/FNR control in finite-sample settings are presented in Section 5, and we conclude with a discussion of related issues in Section 6.

2. Set up and summary of our approach

2.1. Notation and assumptions

We consider $J \geq 2$ independent data streams

$$X^j = \{X_1^j, X_2^j, \dots\}, \quad j \in [J],$$

where $[J]$ denotes $\{1, 2, \dots, J\}$ throughout. We assume throughout the paper that the streams are independent of each other, but not always that the elements of a given stream X_1^j, X_2^j, \dots are independent. In particular, we will prove error control of the proposed procedures only under independence of the streams, but make the additional assumption that each stream is made up of i.i.d. observations in order to prove asymptotic optimality.

Letting P^j denote the probability distribution of X^j , we consider simultaneously testing J null-alternative simple hypothesis pairs

$$H_0^j : P^j = P_0^j \quad \text{vs.} \quad H_1^j : P^j = P_1^j, \quad j \in [J], \quad (1)$$

where, for each j , the P_0^j and P_1^j are known and distinct distributions, assumed to be mutually absolutely continuous when restricted to $\sigma_n = \sigma(\{X_i^j : i \in [n], j \in [J]\})$, the σ -field generated by the observations

$$X_1^j, X_2^j, \dots, X_n^j, \quad j \in [J]. \quad (2)$$

We utilize the informal but illustrative terminology that refers to the j th stream as *noise* (resp. *signal*) if H_0^j (resp. H_1^j) is true. We shall refer to the subscript n in X_n^j as *time* and assume that at time n , the observations (2) (and *only* these observations) have been observed. We define a *sequential test* (or *procedure*) of (1) as a pair of random variables (T, D) such that T is a $\{\sigma_n\}$ -stopping time (i.e., the event $\{T = n\} \in \sigma_n$) and $D = (D^1, \dots, D^J)$ is a σ_T -measurable decision rule taking values in $\{0, 1\}^J$. The interpretation of (T, D) is that sampling of all streams is terminated at time T and the procedure classifies the j th stream as noise (resp. signal) if $D^j = 0$ (resp. $D^j = 1$), for each $j \in [J]$.

For each pair of distributions P_0^j and P_1^j in (1), let f_0^j and f_1^j denote the respective density functions with respect to some common measure μ^j . Denote the log-likelihood ratio for the j th stream at time n by

$$\lambda^j(n) := \log \frac{f_1^j(X_1^j, \dots, X_n^j)}{f_0^j(X_1^j, \dots, X_n^j)}$$

and the Kullback–Leibler information numbers by

$$I_0^j := \int (-\lambda^j(1)) f_0^j(X_1^j) d\mu^j \quad \text{and} \quad I_1^j := \int \lambda^j(1) f_1^j(X_1^j) d\mu^j. \quad (3)$$

¹ This case is the reason for the word “without” in the paper’s title.

Define

$$v_0^j := \int \left(-\lambda^j(1) - l_0^j\right)^2 f_0^j(X_1^j) d\mu^j \quad \text{and} \quad v_1^j := \int \left(\lambda^j(1) - l_1^j\right)^2 f_1^j(X_1^j) d\mu^j. \tag{4}$$

We assume the likelihood ratio statistics satisfy

$$P_0^j \left(\lim_{n \rightarrow \infty} \lambda^j(n) = -\infty\right) = P_1^j \left(\lim_{n \rightarrow \infty} \lambda^j(n) = \infty\right) = 1 \tag{5}$$

for all $j \in [J]$. For example, if the elements of the stream X_1^j, X_2^j, \dots are i.i.d., then (5) follows from the strong law of large numbers. Since, as mentioned in the first paragraph of this section, we assume the streams are independent throughout but only make the i.i.d. assumption for some of our results, the assumption (5) guarantees that the proposed sequential tests will terminate a.s. even when the i.i.d. assumption is not made. It is under the i.i.d. assumption that the Kullback–Leibler information numbers (3) are utilized.

In what follows we will make frequent use of the log-likelihood ratios' order statistics which we denote by

$$\lambda^{(1)}(n) \geq \lambda^{(2)}(n) \geq \dots \geq \lambda^{(J)}(n),$$

with ties broken arbitrarily. To refer to a particular order statistic let $i_j(n), j \in [J]$, be such that $\lambda^{i_j(n)}(n) = \lambda^{(j)}(n)$. Also, letting $|\cdot|$ denote set cardinality throughout, define $p(n) := |\{j \in [J] : \lambda^j(n) > 0\}|$ to be the number of positive λ^j at time n .

We will describe states of nature by subsets $A \subseteq [J]$ which we call *signal sets*, describing the situation where H_0^j is false (signal) for all $j \in A$, and H_0^j is true (noise) for all $j \in A^c := [J] \setminus A$. For a signal set A , let P_A denote the probability measure determined by A , i.e.,

$$P_A := \bigotimes_{j=1}^J P_{\mathbf{1}_{\{j \in A\}}}, \tag{6}$$

where $\mathbf{1}\{\cdot\}$ denotes an indicator function throughout. Let E_A denote expectation with respect to P_A . For a given signal set A , let

$$\eta_0^A := \min_{j \in A^c} l_0^j \quad \text{and} \quad \eta_1^A := \min_{j \in A} l_1^j,$$

which can be thought of as the “worst case” information numbers for signal set A , and appear in the expected sample sizes of optimal procedures.

For any multiple testing procedure under consideration, let V denote the number of true null hypotheses rejected (i.e., the number of false positives), W the number of false null hypotheses accepted (i.e., the number of false negatives), and R the number of null hypotheses rejected. The number of null hypotheses accepted is therefore $J - R$. Under signal set A the type 1 and 2 familywise error rates (FWEs) are

$$\text{FWE}_{1,A} = P_A(V \geq 1), \quad \text{FWE}_{2,A} = P_A(W \geq 1)$$

and the false discovery and non-discovery rates (FDR, FNR) are

$$\text{FDR}_A = E_A \left(\frac{V}{R \vee 1} \right), \quad \text{FNR}_A = E_A \left(\frac{W}{(J - R) \vee 1} \right),$$

where $x \vee y = \max\{x, y\}$. In these and other multiple testing metrics, we will include the procedure being evaluated as an argument (e.g., $\text{FWE}_{1,A}(T, D)$) when needed but omit it when it is clear from the context or when a statement holds for any procedure. Likewise, we will omit the subscript A in expressions that hold for arbitrary signal sets, such as in the next paragraph.

2.2. Summary of our approach

For the asymptotic theory governing optimal FDR and FNR control it will suffice to consider the following elementary bounds between these metrics and the type 1 and 2 FWEs. On one hand,

$$\text{FDR} = E \left(\frac{V}{R \vee 1} \right) = E \left(\frac{V}{R \vee 1} \cdot \mathbf{1}\{V \geq 1\} \right) \leq E(1 \cdot \mathbf{1}\{V \geq 1\}) = P(V \geq 1) = \text{FWE}_1. \tag{7}$$

On the other hand,

$$\text{FDR} = E \left(\frac{V}{R \vee 1} \cdot \mathbf{1}\{V \geq 1\} \right) \geq E \left(\frac{1}{J} \cdot \mathbf{1}\{V \geq 1\} \right) = \frac{1}{J} \cdot P(V \geq 1) = \frac{1}{J} \cdot \text{FWE}_1. \tag{8}$$

Similar arguments show that

$$\frac{1}{J} \cdot \text{FWE}_2 \leq \text{FNR} \leq \text{FWE}_2. \tag{9}$$

Although crude, it will turn out that bounds like (7)–(9) suffice to show that sequential procedures that are asymptotically optimal for FWE control are also asymptotically optimal for FDR/FNR control. More generally, we will show that the sequential procedures that are asymptotically optimal for FWE control are also asymptotically optimal (with slightly modified critical values) for control of any multiple testing error metric whose type 1 and 2 versions are bounded above by some constant multiple of the corresponding FWE_i when evaluated on the optimal procedure, and are bounded below by some constant multiple of the corresponding FWE_i when evaluated on any procedure; these statements are made precise in conditions (i)–(ii) of Theorems 3.1 and 4.1. Furthermore, it will turn out that these “bounds” between the error metrics need not be constant and the lower bounds can be allowed to approach 0 and the upper bounds ∞ , allowing asymptotic regimes in which the number J of null hypotheses approaches ∞ . See part (3) of Theorems 3.1 and 4.1.

To state these more general results, we denote the type 1 and 2 versions of a generic multiple testing error metric by $MTE = (MTE_1, MTE_2)$, which is any pair of functions mapping multiple testing procedures into $[0, 1]$. As above, we will add a signal set A as a subscript and a procedure (T, D) as an argument when needed. Our more general results in Theorems 3.1 and 4.1 will produce asymptotic optimality results for FDR/FNR in Corollaries 3.1 and 4.1 and for pFDR/pFNR in Corollaries 3.2 and 4.2, after verifying that pFDR/pFNR satisfy similar bounds.

We will use \mathcal{P} to denote a signal class which is a collection of signal sets; more precisely, \mathcal{P} is a subset of the power set of $[J]$. We will consider the classes of sequential tests (T, D) controlling the type 1 and 2 versions of various multiple testing error metrics at specific levels² $\alpha, \beta > 0$. For a generic metric MTE, let

$$\Delta_{\mathcal{P}}^{MTE}(\alpha, \beta) = \{(T, D) : MTE_{1,A} \leq \alpha \text{ and } MTE_{2,A} \leq \beta \text{ for all } A \in \mathcal{P}\}. \tag{10}$$

When the MTE is FWE we have

$$\Delta_{\mathcal{P}}^{FWE}(\alpha, \beta) = \{(T, D) : FWE_{1,A} \leq \alpha \text{ and } FWE_{2,A} \leq \beta \text{ for all } A \in \mathcal{P}\}$$

and, by a slight but obvious abuse of notation, for FDR/FNR control we have

$$\Delta_{\mathcal{P}}^{FDR}(\alpha, \beta) = \{(T, D) : FDR_A \leq \alpha \text{ and } FNR_A \leq \beta \text{ for all } A \in \mathcal{P}\}.$$

3. Number of signals known exactly

For $m \in [J - 1]$ let $\mathcal{P}_m = \{A \subseteq [J] : |A| = m\}$. Thus \mathcal{P}_m is the class of signal sets with exactly m signals. For a given such m , Song and Fellouris (2017) defined the following gap rule $(T_G(c), D_G(c))$ with threshold $c > 0$, which is the sequential procedure that stops sampling as soon as the “gap” between the m th and $(m + 1)$ st ordered log-likelihood ratio statistics is at least c . The decision taken at that time is that the null hypotheses with the largest m statistics contain signal. That is,

$$T_G(c) := \inf\{n \geq 1 : \lambda^{(m)}(n) - \lambda^{(m+1)}(n) \geq c\}, \tag{11}$$

$$D_G(c) := \{i_1(T_G(c)), \dots, i_m(T_G(c))\}. \tag{12}$$

Theorem 3.1 establishes admissibility and asymptotic optimality of the gap rule for control of any metric that is bounded above by some constant multiple of FWE_i when evaluated on the gap rule in the sense of (13), and is bounded below by some constant multiple of FWE_i when evaluated on any procedure in the sense of (14).

3.1. Main result and its application to FDR/FNR and pFDR/pFNR control

Theorem 3.1. Fix $m \in [J - 1]$ and let $(T_G(c), D_G(c))$ denote the gap rule with number of signals m and threshold $c > 0$. Let MTE be a multiple testing error metric such that:

(i) there is a constant C_1 such that

$$MTE_{i,A}(T_G(c), D_G(c)) \leq C_1 \cdot FWE_{i,A}(T_G(c), D_G(c)) \tag{13}$$

for $i = 1$ and 2, for all $A \in \mathcal{P}_m$, and for all $c > 0$, and

(ii) there is a constant C_2 such that

$$MTE_{i,A}(T, D) \geq C_2 \cdot FWE_{i,A}(T, D) \tag{14}$$

for $i = 1$ and 2, for all $A \in \mathcal{P}_m$, and for all procedures (T, D) .

Given $\alpha, \beta \in (0, 1)$ let (T_G^*, D_G^*) denote the gap rule with number of signals m and threshold

$$c = \lceil \log((\alpha/C_1) \wedge (\beta/C_1)) \rceil + \log(m(J - m)).$$

Then the following hold.

² Here we do not explicitly require that $\alpha, \beta < 1$ in the definitions of the classes that follow. This is because below we will multiply and divide inputs α, β to the class (10) by various positive constants. This does not present a problem because, for example, if $\alpha \geq 1$ then the class requirement in $\Delta_{\mathcal{P}}^{MTE}(\alpha, \beta)$ that $MTE_{1,A} \leq \alpha$ will automatically be satisfied since $MTE_i \in [0, 1]$ by assumption. Similar statements apply if $\beta \geq 1$.

(1) Under the assumption that the streams X^1, \dots, X^J are independent, (T_G^*, D_G^*) is admissible for MTE control. That is,

$$(T_G^*, D_G^*) \in \Delta_{\mathcal{P}_m}^{\text{MTE}}(\alpha, \beta). \tag{15}$$

(2) Under the additional assumption that each stream $X^j = (X_1^j, X_2^j, \dots)$ is made up of i.i.d. random variables X_1^j, X_2^j, \dots , the procedure (T_G^*, D_G^*) is asymptotically optimal for MTE control with respect to class \mathcal{P}_m . That is, for all $A \in \mathcal{P}_m$,

$$E_A(T_G^*) \sim \frac{|\log(\alpha \wedge \beta)|}{\eta_1^A + \eta_0^A} \sim \inf_{(T,D) \in \Delta_{\mathcal{P}_m}^{\text{MTE}}(\alpha, \beta)} E_A(T) \tag{16}$$

as $\alpha, \beta \rightarrow 0$.

(3) Parts (1) and (2) still hold if J, m, C_1 , and C_2 are allowed to vary with α, β as long as

$$J = o(|\log(\alpha \wedge \beta)|^{1/4}), \quad \log C_1 = o(|\log(\alpha \wedge \beta)|), \quad \text{and} \quad \log C_2^{-1} = o(|\log(\alpha \wedge \beta)|), \tag{17}$$

and

$$\text{the information numbers } \{I_0^j, I_1^j : j \in [J]\} \text{ in (3) are bounded above } 0, \text{ and} \tag{18}$$

$$\text{the variances } \{v_0^j, v_1^j : j \in [J]\} \text{ in (4) are bounded below } \infty. \tag{19}$$

In particular, the asymptotic optimality in part (2) still holds if $J, C_1 \rightarrow \infty$ and $C_2 \rightarrow 0$ as $\alpha, \beta \rightarrow 0$ as long as (17)–(19) hold, with $m \in [J - 1]$ allowed to vary arbitrarily.

Proof. For part (1), fix arbitrary $A \in \mathcal{P}_m$. Song and Fellouris (2017, Theorem 3.1) establish that

$$\text{FWE}_{1,A}(T_G^*, D_G^*) \leq \alpha/C_1 \quad \text{and} \quad \text{FWE}_{2,A}(T_G^*, D_G^*) \leq \beta/C_1.$$

Applying (13) yields

$$\text{MTE}_{1,A}(T_G^*, D_G^*) \leq C_1 \cdot \text{FWE}_{1,A}(T_G^*, D_G^*) \leq C_1(\alpha/C_1) = \alpha.$$

A similar argument shows that $\text{MTE}_{2,A}(T_G^*, D_G^*) \leq \beta$, establishing (15).

For part (2), fix arbitrary $A \in \mathcal{P}_m$ and consider $\alpha, \beta \rightarrow 0$. Let κ denote $|\log(\alpha \wedge \beta)| \rightarrow \infty$. The first inequality in the following is established by Song and Fellouris (2017, Lemma 5.2), and the rest are straightforward calculations:

$$E_A(T_G^*) \leq \frac{c}{\eta_1^A + \eta_0^A} + O(m(J - m)\sqrt{c}) = \frac{\kappa}{\eta_1^A + \eta_0^A} + O(\log C_1) + O(J^2\sqrt{\kappa}) \sim \frac{\kappa}{\eta_1^A + \eta_0^A}. \tag{20}$$

To establish that the last expression in (20) is also a lower bound for any procedure in $\Delta_{\mathcal{P}_m}^{\text{MTE}}(\alpha, \beta)$, it follows from (14) that

$$\Delta_{\mathcal{P}_m}^{\text{MTE}}(\alpha, \beta) \subseteq \Delta_{\mathcal{P}_m}^{\text{FWE}}(\alpha/C_2, \beta/C_2),$$

thus

$$\inf_{(T,D) \in \Delta_{\mathcal{P}_m}^{\text{MTE}}(\alpha, \beta)} E_A(T) \geq \inf_{(T,D) \in \Delta_{\mathcal{P}_m}^{\text{FWE}}(\alpha/C_2, \beta/C_2)} E_A(T).$$

Song and Fellouris (2017, Theorem 5.3) establish that the latter is of order

$$\frac{|\log((\alpha/C_2) \wedge (\beta/C_2))|}{\eta_1^A + \eta_0^A} = \frac{\kappa}{\eta_1^A + \eta_0^A} - O(\log C_2^{-1}) \sim \frac{\kappa}{\eta_1^A + \eta_0^A}. \tag{21}$$

Thus

$$\inf_{(T,D) \in \Delta_{\mathcal{P}_m}^{\text{MTE}}(\alpha, \beta)} E_A(T) \geq \frac{\kappa}{\eta_1^A + \eta_0^A} (1 + o(1)) \tag{22}$$

and combining this with (20) gives the desired result.

For part (3), letting J, m, C_1 , and C_2 possibly vary with α, β such that (17)–(19) hold does not affect the calculations for part (1) since they are non-asymptotic. The condition (18) guarantees that the term $1/(\eta_1^A + \eta_0^A)$ in (16) is bounded away from ∞ , and it is bounded away from 0 since the η_i^A are minima of sets of finite numbers. A more explicit form of the $O(\cdot)$ term after the first inequality in (20) is

$$O\left(m(J - m)\sqrt{c \max_{i,j} v_i^j / (I_i^j)^3}\right), \tag{23}$$

which follows from results in renewal theory; see Gut (2009, Chapter 3.9) and Mei (2008, Theorem 2). The conditions (18)–(19) thus guarantee that this maximum is $O(1)$, thus (20) still holds when J and C_1 satisfy (17), hence the same asymptotic upper bound on $E_A(T_G^*)$ holds. Similarly, (21) still holds because C_2 satisfies (17), hence the same asymptotic lower bound (22) on $E_A(T)$ holds. \square

3.1.1. FDR/FNR control under known number of signals

To apply **Theorem 3.1** to FDR/FNR, we see that the upper bounds on FDR/FNR in (7) and (9) suffice³ for condition (i) of the theorem with $C_1 = 1$, and the lower bounds in (8) and (9) suffice for condition (ii) with $C_2 = 1/J$, yielding the following corollary. And letting $J \rightarrow \infty$ such that $J = o(|\log(\alpha \wedge \beta)|^{1/4})$ satisfies the requirements in (17) of both J and $C_2 = 1/J$.

Corollary 3.1. For $m \in [J - 1]$ and $\alpha, \beta \in (0, 1)$, let (T_G^*, D_G^*) denote the gap rule with number of signals m and threshold $c = |\log(\alpha \wedge \beta)| + \log(m(J - m))$.

(1) Under the assumption that the streams X^1, \dots, X^J are independent, (T_G^*, D_G^*) is admissible for FDR/FNR control. That is,

$$(T_G^*, D_G^*) \in \Delta_{\mathcal{P}_m}^{\text{FDR}}(\alpha, \beta).$$

(2) Under the additional assumption that each stream $X^j = (X_1^j, X_2^j, \dots)$ is made up of i.i.d. random variables X_1^j, X_2^j, \dots , the procedure (T_G^*, D_G^*) is asymptotically optimal for FDR/FNR control with respect to class \mathcal{P}_m . That is, for all $A \in \mathcal{P}_m$,

$$E_A(T_G^*) \sim \frac{|\log(\alpha \wedge \beta)|}{\eta_1^A + \eta_0^A} \sim \inf_{(T,D) \in \Delta_{\mathcal{P}_m}^{\text{FDR}}(\alpha, \beta)} E_A(T)$$

as $\alpha, \beta \rightarrow 0$.

(3) Parts (1) and (2) still hold if $J \rightarrow \infty$ as $\alpha, \beta \rightarrow 0$ such that $J = o(|\log(\alpha \wedge \beta)|^{1/4})$, (18)–(19) hold, and $m \in [J - 1]$ varies arbitrarily with α, β .

3.1.2. pFDR/pFNR control under known number of signals

Continuing to use the notation of Section 2, pFDR and its type 2 analog pFNR are defined as

$$\text{pFDR} = E\left(\frac{V}{R} \mid R \geq 1\right) \quad \text{and} \quad \text{pFNR} = E\left(\frac{W}{J - R} \mid J - R \geq 1\right). \tag{24}$$

As above, we will add the signal set A in the subscript and a procedure as an argument when needed. For $\alpha, \beta > 0$ and signal class \mathcal{P} , define

$$\Delta_{\mathcal{P}}^{\text{pFDR}}(\alpha, \beta) = \{(T, D) : \text{pFDR}_A \leq \alpha \text{ and } \text{pFNR}_A \leq \beta \text{ for all } A \in \mathcal{P}\}.$$

To apply **Theorem 3.1** to pFDR/pFNR, unlike with FDR/FNR it is not possible to provide universal upper bounds like (7) and (9). Proceeding similarly, one obtains

$$\begin{aligned} \text{pFDR} &= E\left(\frac{V}{R} \mid R \geq 1\right) = E\left(\frac{V}{R} \mid R \geq 1, V \geq 1\right) P(V \geq 1 \mid R \geq 1) \leq 1 \cdot \frac{P(V \geq 1)}{P(R \geq 1)} \\ &= \frac{\text{FWE}_1}{P(R \geq 1)}. \end{aligned} \tag{25}$$

Similarly, $\text{pFNR} \leq P(W \geq 1) / P(J - R \geq 1)$, and because $P(R \geq 1)$ and $P(J - R \geq 1)$ cannot be bounded below in general, universal upper bounds like (7) and (9) do not hold for pFDR/pFNR. However, condition (i) in **Theorem 3.1** merely requires such upper bounds to hold for the gap rule (T_G, D_G) with m signals, which always rejects exactly m nulls, and recalling that $1 \leq m \leq J - 1$ by assumption (i.e., $m \in [J - 1]$), the gap rule satisfies

$$P(R \geq 1) = P(J - R \geq 1) = 1.$$

Thus, (25) yields $\text{pFDR}(T_G, D_G) \leq \text{FWE}_1$ and $\text{pFNR}(T_G, D_G) \leq \text{FWE}_2$, so condition (i) of the theorem holds with $C_1 = 1$. The lower bounds are more straightforward:

$$\begin{aligned} \text{pFDR} &= E\left(\frac{V}{R} \mid R \geq 1, V \geq 1\right) P(V \geq 1 \mid R \geq 1) \geq \frac{1}{J} \cdot \frac{P(V \geq 1)}{P(R \geq 1)} \geq \frac{1}{J} \cdot \frac{\text{FWE}_1}{1} \\ &= \frac{1}{J} \cdot \text{FWE}_1, \end{aligned} \tag{26}$$

and similarly

$$\text{pFNR} \geq (1/J) \cdot \text{FWE}_2. \tag{27}$$

Thus, condition (ii) of the theorem is satisfied with $C_2 = 1/J$, yielding the following corollary.

Corollary 3.2. For $m \in [J - 1]$ and $\alpha, \beta \in (0, 1)$, let (T_G^*, D_G^*) denote the gap rule with number of signals m and threshold $c = |\log(\alpha \wedge \beta)| + \log(m(J - m))$.

³ In fact, (7) and (9) are stronger than what is needed since they hold for all procedures, whereas condition (i) just requires this to hold for the gap rule.

(1) Under the assumption that the streams X^1, \dots, X^J are independent, (T_G^*, D_G^*) is admissible for pFDR/pFNR control. That is,

$$(T_G^*, D_G^*) \in \Delta_{\mathcal{P}_m}^{\text{pFDR}}(\alpha, \beta).$$

(2) Under the additional assumption that each stream $X^j = (X_{1j}^j, X_{2j}^j, \dots)$ is made up of i.i.d. random variables $X_{1j}^j, X_{2j}^j, \dots$, the procedure (T_G^*, D_G^*) is asymptotically optimal for pFDR/pFNR control with respect to class \mathcal{P}_m . That is, for all $A \in \mathcal{P}_m$,

$$E_A(T_G^*) \sim \frac{|\log(\alpha \wedge \beta)|}{\eta_1^A + \eta_0^A} \sim \inf_{(T, D) \in \Delta_{\mathcal{P}_m}^{\text{pFDR}}(\alpha, \beta)} E_A(T)$$

as $\alpha, \beta \rightarrow 0$.

(3) Parts (1) and (2) still hold if $J \rightarrow \infty$ as $\alpha, \beta \rightarrow 0$ such that $J = o(|\log(\alpha \wedge \beta)|^{1/4})$, (18)–(19) hold, and $m \in [J - 1]$ varies arbitrarily with α, β .

3.2. Other multiple testing error metrics

Theorem 3.1 applies to any multiple testing error metrics which satisfy the theorem's conditions. In addition to FDR/FNR and pFDR/pFNR these include the *per-comparison error rate (PCER)* $E(V/J)$ as defined by Benjamini and Hochberg (1995), the *false positive rate* $E(V/m)$ (e.g., Burke et al., 1988), the *per-family error rate*⁴ (PFER) EV which appears in social and behavioral science research (e.g., Frane, 2015; Keselman, 2015), and their type 2 analogs. These type 1 metrics are all proportional to EV and for the gap rule,

$$EV = E(V\mathbf{1}\{V \geq 1\}) \leq m \cdot P(V \geq 1) = m \cdot \text{FWE}_1, \tag{28}$$

and for any procedure,

$$EV = E(V\mathbf{1}\{V \geq 1\}) \geq 1 \cdot P(V \geq 1) = \text{FWE}_1. \tag{29}$$

Similar statements apply to the type 2 versions.

4. Bounds on the number of signals

In this section we consider asymptotically optimal procedures for scenarios in which the number of signals is known to be between two given values $0 \leq \ell < u \leq J$, the $\ell = u$ case having already been considered in Section 3. That is, we consider asymptotic optimality in signal classes of the form $\mathcal{P}_{\ell, u} := \{A \subseteq [J] : \ell \leq |A| \leq u\}$ for such ℓ, u . Mirroring our results for when the number of signals is known exactly, we show that the sequential procedures that are asymptotically optimal for FWE control are also asymptotically optimal (with modified critical values) for control of any multiple testing error metric whose type 1 and 2 versions are bounded between constant multiples of the corresponding FWE_i in certain situations, made precise in conditions (i)–(ii) of Theorem 4.1.

Given such bounds ℓ and u , Song and Fellouris (2017) defined the following *gap-intersection rule* (T_{GI}, D_{GI}) , depending on four positive thresholds a, b, c, d , in terms of the stopping times

$$\tau_1 := \inf\{n \geq 1 : \lambda^{(\ell+1)}(n) \leq -a, \lambda^{(\ell)}(n) - \lambda^{(\ell+1)}(n) \geq c\}, \tag{30}$$

$$\tau_2 := \inf\{n \geq 1 : \ell \leq p(n) \leq u \text{ and } \lambda^{(j)}(n) \notin (-a, b) \text{ for all } j \in [J]\}, \tag{31}$$

$$\tau_3 := \inf\{n \geq 1 : \lambda^{(u)}(n) \geq b, \lambda^{(u)}(n) - \lambda^{(u+1)}(n) \geq d\}. \tag{32}$$

In (30)–(32) we set $\lambda^{(0)}(n) = -\infty$ and $\lambda^{(J)}(n) = \infty$ for all n to handle the cases $\ell = 0$ or $u = J$. Finally, the gap-intersection rule (T_{GI}, D_{GI}) is defined as

$$T_{GI} := \min\{\tau_1, \tau_2, \tau_3\} \text{ and } D_{GI} := \{i_1(T_{GI}), \dots, i_{p'}(T_{GI})\},$$

where $p' := (p(T_{GI}) \vee \ell) \wedge u$ is the value in $\{\ell, \ell + 1, \dots, u\}$ closest to $p(T_{GI})$. Song and Fellouris (2017) describe τ_2 as similar to De and Baron's (2012a) "intersection rule", which requires only the second condition in (31), but modified to incorporate the prior information on the number of signals by requiring $\ell \leq p(\tau_2) \leq u$. The stopping times τ_1 and τ_3 provide needed additional efficiency when the number of signals equals ℓ or u .

⁴ This metric is not constrained to take values in $[0, 1]$ as we assumed above, but the theory there can be modified to allow metrics to take any nonnegative values with only notational changes. Alternatively, one could think of standardizing the false positive rate by dividing by m or J , leading to the other two metrics mentioned.

4.1. Main result and its application to FDR/FNR and pFDR/pFNR control

Theorem 4.1. Fix integers $0 \leq \ell < u \leq J$ and let (T_{GI}, D_{GI}) denote the gap-intersection rule with bounds ℓ, u on the number of signals and thresholds $a, b, c, d > 0$. Let MTE be multiple testing error metric such that:

(i) there is a constant C_1 such that

$$MTE_{i,A}(T_{GI}, D_{GI}) \leq C_1 \cdot FWE_{i,A}(T_{GI}, D_{GI}) \tag{33}$$

for $i = 1$ and 2 , for all $A \in \mathcal{P}_{\ell,u}$, and for all thresholds $a, b, c, d > 0$, and

(ii) there is a constant C_2 such that

$$MTE_{i,A}(T, D) \geq C_2 \cdot FWE_{i,A}(T, D) \tag{34}$$

for $i = 1$ and 2 , for all $A \in \mathcal{P}_{\ell,u}$, and for all procedures (T, D) .

Given $\alpha, \beta \in (0, 1)$ let (T_{GI}^*, D_{GI}^*) denote the gap rule with bounds ℓ, u on the number of signals, and thresholds

$$\begin{aligned} a &= \lceil \log(\beta/C_1) \rceil + \log J, & b &= \lceil \log(\alpha/C_1) \rceil + \log J, \\ c &= \lceil \log(\alpha/C_1) \rceil + \log(J - \ell), & d &= \lceil \log(\beta/C_1) \rceil + \log(u). \end{aligned}$$

Then the following hold.

(1) Under the assumption that the streams X^1, \dots, X^J are independent, (T_{GI}^*, D_{GI}^*) is admissible for MTE control. That is,

$$(T_{GI}^*, D_{GI}^*) \in \Delta_{\mathcal{P}_{\ell,u}}^{MTE}(\alpha, \beta). \tag{35}$$

(2) Under the additional assumption that each stream $X^j = (X_1^j, X_2^j, \dots)$ is made up of i.i.d. random variables X_1^j, X_2^j, \dots , the procedure (T_{GI}^*, D_{GI}^*) is asymptotically optimal for MTE control with respect to class $\mathcal{P}_{\ell,u}$. That is, for all $A \in \mathcal{P}_{\ell,u}$,

$$\begin{aligned} E_A(T_{GI}^*) &\sim \inf_{(T,D) \in \Delta_{\mathcal{P}_{\ell,u}}^{MTE}(\alpha, \beta)} E_A(T) \\ &\sim \begin{cases} \max\{\lceil \log \beta \rceil / \eta_0^A, \lceil \log \alpha \rceil / (\eta_0^A + \eta_1^A)\}, & \text{if } |A| = \ell \\ \max\{\lceil \log \beta \rceil / \eta_0^A, \lceil \log \alpha \rceil / \eta_1^A\}, & \text{if } \ell < |A| < u \\ \max\{\lceil \log \alpha \rceil / \eta_1^A, \lceil \log \beta \rceil / (\eta_0^A + \eta_1^A)\}, & \text{if } |A| = u \end{cases} \end{aligned} \tag{36}$$

as $\alpha, \beta \rightarrow 0$.

(3) Parts (1) and (2) still hold if J, m, ℓ, u, C_1 , and C_2 are allowed to vary with α, β as long as (17)–(19) hold. In particular, the asymptotic optimality in part (2) still holds if $J, C_1 \rightarrow \infty$ and $C_2 \rightarrow 0$ as $\alpha, \beta \rightarrow 0$ as long as (17)–(19) hold, with $m \in [J - 1]$ and $0 \leq \ell < u \leq J$ allowed to vary arbitrarily.

Proof. For part (1), fix arbitrary $A \in \mathcal{P}_{\ell,u}$. Song and Fellouris (2017, Theorem 3.2) establish that

$$FWE_{1,A}(T_{GI}^*, D_{GI}^*) \leq \alpha/C_1 \quad \text{and} \quad FWE_{2,A}(T_{GI}^*, D_{GI}^*) \leq \beta/C_1.$$

Applying (33) yields

$$MTE_{1,A}(T_{GI}^*, D_{GI}^*) \leq C_1 \cdot FWE_{1,A}(T_{GI}^*, D_{GI}^*) \leq C_1(\alpha/C_1) = \alpha,$$

and a similar argument shows that $MTE_{2,A}(T_{GI}^*, D_{GI}^*) \leq \beta$, establishing (35).

For part (2), fix arbitrary $A \in \mathcal{P}_{\ell,u}$ and consider $\alpha, \beta \rightarrow 0$. Letting $\kappa(\alpha, \beta) \rightarrow \infty$ denote the corresponding expression in (36), it is not hard to see that

$$\kappa(\alpha/C_1, \beta/C_1) = \kappa(\alpha, \beta) + O(\log C_1), \tag{37}$$

which can be verified for each of the three cases. For the second case of (36), Song and Fellouris (2017, Theorem 5.5) establish that

$$E_A(T_{GI}^*) \leq \kappa(\alpha/C_1, \beta/C_1) + O(\sqrt{a \vee b}) = \kappa(\alpha, \beta) + O(\log C_1) + O(\sqrt{\kappa(\alpha, \beta)}) \sim \kappa(\alpha, \beta). \tag{38}$$

And similar arguments show that in the first case of (36),

$$E_A(T_{GI}^*) \leq \kappa(\alpha/C_1, \beta/C_1) + O(\sqrt{a \vee c}) = \kappa(\alpha, \beta) + O(\log C_1) + O(\sqrt{\kappa(\alpha, \beta)}) \sim \kappa(\alpha, \beta), \tag{39}$$

and in the third case of (36),

$$E_A(T_{GI}^*) \leq \kappa(\alpha/C_1, \beta/C_1) + O(\sqrt{b \vee d}) = \kappa(\alpha, \beta) + O(\log C_1) + O(\sqrt{\kappa(\alpha, \beta)}) \sim \kappa(\alpha, \beta). \tag{40}$$

To establish that $\kappa(\alpha, \beta)$ is also an asymptotic lower bound in $\Delta_{\mathcal{P}_{\ell,u}}^{\text{MTE}}(\alpha, \beta)$, it follows from (34) that $\Delta_{\mathcal{P}_{\ell,u}}^{\text{MTE}}(\alpha, \beta) \subseteq \Delta_{\mathcal{P}_{\ell,u}}^{\text{FWE}}(\alpha/C_2, \beta/C_2)$, thus

$$\inf_{(T,D) \in \Delta_{\mathcal{P}_{\ell,u}}^{\text{MTE}}(\alpha, \beta)} E_A(T) \geq \inf_{(T,D) \in \Delta_{\mathcal{P}_{\ell,u}}^{\text{FWE}}(\alpha/C_2, \beta/C_2)} E_A(T).$$

Song and Fellouris (2017, Theorem 5.5) establish that the latter is greater than or equal to

$$\kappa(\alpha/C_2, \beta/C_2)(1 + o(1)) = \kappa(\alpha, \beta) + O(|\log C_2|),$$

by arguments similar to (37). Thus

$$\inf_{(T,D) \in \Delta_{\mathcal{P}_{\ell,u}}^{\text{MTE}}(\alpha, \beta)} E_A(T) \geq \kappa(\alpha, \beta) + O(|\log C_2|) \sim \kappa(\alpha, \beta). \tag{41}$$

and combining this with (38) gives the desired result.

The proof of part (3) is similar to the corresponding part in Theorem 3.1 since the equivalences in (38)–(41) still hold when the parameters and constants are allowed to vary with α, β in the way described. \square

4.1.1. FDR/FNR control under bounds on the number of signals

The application of Theorem 4.1 to FDR/FNR is again immediate from the bounds (7)–(9), from which we see that condition (i) of the theorem holds with $C_1 = 1$ and condition (ii) holds with $C_2 = 1/J$, yielding the following corollary. And letting $J \rightarrow \infty$ such that $J = o(|\log(\alpha \wedge \beta)|^{1/4})$ satisfies the requirements in (17) of both J and $C_2 = 1/J$.

Corollary 4.1. Given $\alpha, \beta \in (0, 1)$ and integers $0 \leq \ell < u \leq J$, let (T_{GI}^*, D_{GI}^*) denote the gap-intersection rule with bounds ℓ, u on the number of signals, and thresholds

$$\begin{aligned} a &= |\log \beta| + \log J, & b &= |\log \alpha| + \log J, \\ c &= |\log \alpha| + \log(J - \ell)J, & d &= |\log \beta| + \log(uJ). \end{aligned}$$

Then the following hold.

(1) Under the assumption that the streams X^1, \dots, X^J are independent, (T_{GI}^*, D_{GI}^*) is admissible for FDR/FNR control. That is,

$$(T_{GI}^*, D_{GI}^*) \in \Delta_{\mathcal{P}_{\ell,u}}^{\text{FDR}}(\alpha, \beta).$$

(2) Under the additional assumption that each stream $X^j = (X_1^j, X_2^j, \dots)$ is made up of i.i.d. random variables X_1^j, X_2^j, \dots , the procedure (T_{GI}^*, D_{GI}^*) is asymptotically optimal for FDR/FNR control with respect to class $\mathcal{P}_{\ell,u}$. That is, for all $A \in \mathcal{P}_{\ell,u}$,

$$\begin{aligned} E_A(T_{GI}^*) &\sim \inf_{(T,D) \in \Delta_{\mathcal{P}_{\ell,u}}^{\text{FDR}}(\alpha, \beta)} E_A(T) \\ &\sim \begin{cases} \max\{|\log \beta|/\eta_0^A, |\log \alpha|/(\eta_0^A + \eta_1^A)\}, & \text{if } |A| = \ell \\ \max\{|\log \beta|/\eta_0^A, |\log \alpha|/\eta_1^A\}, & \text{if } \ell < |A| < u \\ \max\{|\log \alpha|/\eta_1^A, |\log \beta|/(\eta_0^A + \eta_1^A)\}, & \text{if } |A| = u \end{cases} \end{aligned}$$

as $\alpha, \beta \rightarrow 0$.

(3) Parts (1) and (2) still hold if $J \rightarrow \infty$ as $\alpha, \beta \rightarrow 0$ such that $J = o(|\log(\alpha \wedge \beta)|^{1/4})$, (18)–(19) hold, and $m \in [J - 1]$ and $0 \leq \ell < u \leq J$ vary arbitrarily with α, β .

4.1.2. pFDR/pFNR control under bounds on the number of signals

In applying Theorem 4.1 to pFDR/pFNR, some care must be taken when considering condition (i) of the theorem since pFDR and pFNR cannot be universally bounded above by a constant multiple of FWE₁ and FWE₂, respectively, as in (25). Although condition (i) only requires such bounds to hold for the gap-intersection rule, these may still fail when the lower bound is $\ell = 0$ and $P(R \geq 1)$ is close to zero, or when the upper bound is $u = J$ and $P(J - R \geq 1)$ is close to zero. For this reason, in the following corollary which applies Theorem 4.1 to pFDR/pFNR control, we restrict attention to prior bounds $\ell \geq 1$ and $u \leq J - 1$, under which the gap-intersection rule satisfies $P(R = 0) = P(R = J) = 0$ and thus condition (i) holds with $C_1 = 1$ by (25). We note that problems with pFDR and pFNR when $P(R \geq 1)$ and $P(J - R \geq 1)$, respectively, are close to zero are not unique to our setup here, and these quantities are of course undefined when these probabilities equal zero; see Section 6 for more discussion of this topic. The needed lower bounds in condition (ii) of the theorem are again provided by the universal bounds (26)–(27) with $C_2 = 1/J$.

Corollary 4.2. Given $\alpha, \beta \in (0, 1)$ and integers $1 \leq \ell < u \leq J - 1$, let (T_{GI}^*, D_{GI}^*) denote the gap-intersection rule with bounds ℓ, u on the number of signals, and thresholds

$$\begin{aligned} a &= |\log \beta| + \log J, & b &= |\log \alpha| + \log J, \\ c &= |\log \alpha| + \log(J - \ell)J, & d &= |\log \beta| + \log(uJ). \end{aligned}$$

Then the following hold.

Table 1

Operating characteristics of sequential gap rule and fixed-sample BH and BH_m procedures in simulation study as described in Section 5 with J = 10 data streams.

m	Gap rule				BH			BH _m		
	c	ET	FDR (%)	FNR (%)	n (Savings)	FDR (%)	FNR (%)	n (Savings)	FDR (%)	FNR (%)
1	3.5	29.0 (0.15)	4.30 (0.20)	0.48 (0.02)	70 (59%)	4.49 (0.15)	0.61 (0.02)	50 (42%)	4.54 (0.21)	0.50 (0.02)
2	2.9	31.6 (0.14)	4.60 (0.15)	1.15 (0.04)	60 (47%)	3.97 (0.12)	1.50 (0.04)	46 (31%)	4.74 (0.15)	1.18 (0.04)
3	2.6	31.7 (0.14)	4.75 (0.12)	2.04 (0.05)	59 (46%)	3.55 (0.09)	2.05 (0.05)	45 (30%)	4.40 (0.11)	1.88 (0.05)
4	2.3	30.2 (0.13)	4.59 (0.10)	3.06 (0.07)	54 (44%)	2.84 (0.08)	3.24 (0.07)	40 (25%)	4.80 (0.10)	3.20 (0.06)
5	2.1	28.7 (0.12)	4.66 (0.09)	4.66 (0.09)	52 (45%)	2.55 (0.07)	4.67 (0.08)	37 (22%)	4.75 (0.09)	4.75 (0.09)
6	2.3	30.5 (0.13)	3.18 (0.07)	4.77 (0.10)	54 (44%)	2.06 (0.05)	4.53 (0.09)	40 (24%)	3.32 (0.07)	4.99 (0.10)
7	2.5	30.8 (0.13)	2.14 (0.05)	4.90 (0.12)	56 (45%)	1.50 (0.04)	4.91 (0.11)	43 (28%)	2.10 (0.05)	4.90 (0.12)
8	2.8	30.7 (0.14)	1.22 (0.04)	4.89 (0.15)	60 (49%)	0.96 (0.03)	4.92 (0.13)	45 (32%)	1.31 (0.04)	5.24 (0.15)
9	3.4	28.5 (0.15)	0.49 (0.02)	4.39 (0.20)	65 (56%)	0.50 (0.02)	4.77 (0.15)	50 (43%)	0.48 (0.02)	4.33 (0.20)

(1) Under the assumption that the streams X^1, \dots, X^J are independent, (T_{GI}^*, D_{GI}^*) is admissible for pFDR/pFNR control. That is,

$$(T_{GI}^*, D_{GI}^*) \in \Delta_{\mathcal{P}_{\ell,u}}^{\text{pFDR}}(\alpha, \beta).$$

(2) Under the additional assumption that each stream $X^j = (X_1^j, X_2^j, \dots)$ is made up of i.i.d. random variables X_1^j, X_2^j, \dots , the procedure (T_{GI}^*, D_{GI}^*) is asymptotically optimal for pFDR/pFNR control with respect to class $\mathcal{P}_{\ell,u}$. That is, for all $A \in \mathcal{P}_{\ell,u}$,

$$E_A(T_{GI}^*) \sim \inf_{(T,D) \in \Delta_{\mathcal{P}_{\ell,u}}^{\text{pFDR}}(\alpha, \beta)} E_A(T) \sim \begin{cases} \max\{|\log \beta|/\eta_0^A, |\log \alpha|/(\eta_0^A + \eta_1^A)\}, & \text{if } |A| = \ell \\ \max\{|\log \beta|/\eta_0^A, |\log \alpha|/\eta_1^A\}, & \text{if } \ell < |A| < u \\ \max\{|\log \alpha|/\eta_1^A, |\log \beta|/(\eta_0^A + \eta_1^A)\}, & \text{if } |A| = u \end{cases}$$

as $\alpha, \beta \rightarrow 0$.

(3) Parts (1) and (2) still hold if $J \rightarrow \infty$ as $\alpha, \beta \rightarrow 0$ such that $J = o(|\log(\alpha \wedge \beta)|^{1/4})$, (18)–(19) hold, and $m \in [J - 1]$ and $0 < \ell < u < J$ vary arbitrarily with α, β .

4.2. Other multiple testing error metrics

In addition to FDR/FNR and pFDR/pFNR, Theorem 4.1 applies to any multiple testing error metrics which satisfy the theorem’s conditions including the metrics mentioned in Section 3.2. The argument is similar, with the upper bound (28) being replaced by $u \cdot \text{FWE}_1$ for the gap-intersection rule.

5. Simulation study

In this section we study the non-asymptotic performance of the sequential gap rule for FDR/FNR control as well as comparable fixed-sample procedures based on the Benjamini and Hochberg (1995) procedure (BH) through a simulation study in the setting of Section 3 where the number m of signals is known exactly. We consider J independent streams $X_1^j, X_2^j, \dots, j \in [J]$, of i.i.d. $N(\mu, 1)$ data with hypotheses $H_0^j : \mu = 0$ vs. $H_1^j : \mu = 1/2, j \in [J]$. Table 1 contains the operating characteristics of three procedures under this setup with $J = 10$ and all possible values of $m \in [J - 1]$ in the table’s rows. The columns under the heading Gap Rule in Table 1 describe the performance of the gap rule as defined in (11)–(12) with values of the threshold c listed in the table. In order to study the non-asymptotic performance of this procedure, the asymptotic value of c in Corollary 3.1 was dispensed with and for each m , a value of c was determined by Monte Carlo simulation such that both the achieved FDR and FNR are close to, but no larger than, the nominal values $\alpha = \beta = .05$. We note that this same approach is available to users in practice since m and the simple hypotheses H_0^j, H_1^j are all known and thus the achieved FDR and FNR can be estimated to arbitrary accuracy via Monte Carlo before gathering any data. For the gap rule with c determined in this way, the expected stopping time ET and achieved FDR and FNR, estimated from 10,000 Monte Carlo replications, are given in the table with their standard errors given in parentheses.

A natural competitor for the gap rule is the fixed-sample BH procedure. In implementing the BH procedure two parameters must be chosen: the procedure’s nominal FDR control level α and its fixed sample size, denoted by n in Table 1. In order to have an even-handed comparison, the same nominal level $\alpha = .05$ as the gap rule was used, and for each m the BH procedure’s sample size n was then chosen so that its achieved FNR was as close as possible to that of the gap rule. The columns under the heading BH in Table 1 contain the operating characteristics of the BH procedure with parameters chosen in this way, where again the FDR and FNR are estimated from 10,000 Monte Carlo replications with standard errors given in parentheses. The first of these columns contains the fixed sample size n as well as the expected savings $1 - ET/n$ in sample size of the sequential gap rule over the fixed-sample rule.

Table 2

Operating characteristics of sequential gap rule and fixed-sample BH and BH_m procedures in simulation study as described in Section 5 with $J = 100$ data streams.

m	Gap rule				BH			BH_m		
	c	ET	FDR (%)	FNR (%)	n (Savings)	FDR (%)	FNR (%)	n (Savings)	FDR (%)	FNR (%)
1	3.9	48.8 (0.21)	4.43 (0.21)	0.04 (0.00)	90 (46%)	4.77 (0.16)	0.08 (0.00)	77 (37%)	4.74 (0.21)	0.05 (0.00)
10	1.9	61.0 (0.16)	4.65 (0.06)	0.52 (0.01)	70 (13%)	4.52 (0.06)	0.63 (0.01)	68 (10%)	4.73 (0.06)	0.53 (0.01)
20	1.3	57.3 (0.14)	4.76 (0.04)	1.19 (0.01)	65 (12%)	3.94 (0.04)	1.17 (0.01)	62 (8%)	4.57 (0.04)	1.14 (0.01)
30	1.0	52.8 (0.12)	4.70 (0.03)	2.01 (0.01)	60 (12%)	3.50 (0.03)	2.01 (0.02)	57 (7%)	4.81 (0.02)	3.21 (0.02)
40	0.8	48.0 (0.11)	4.74 (0.03)	3.16 (0.02)	56 (14%)	3.00 (0.03)	3.22 (0.02)	50 (3%)	4.81 (0.02)	3.20 (0.02)
50	0.7	45.1 (0.10)	4.47 (0.03)	4.47 (0.03)	53 (15%)	2.53 (0.02)	4.90 (0.03)	47 (4%)	4.39 (0.02)	4.39 (0.02)
60	0.8	48.2 (0.11)	3.19 (0.02)	4.79 (0.03)	56 (14%)	2.02 (0.02)	4.94 (0.03)	50 (3%)	3.17 (0.02)	4.76 (0.02)
70	1.0	52.8 (0.12)	2.03 (0.01)	4.74 (0.03)	60 (12%)	1.52 (0.01)	4.74 (0.04)	57 (7%)	2.00 (0.01)	4.59 (0.03)
80	1.3	57.0 (0.13)	1.20 (0.01)	4.78 (0.04)	64 (11%)	1.00 (0.01)	5.00 (0.05)	63 (10%)	1.09 (0.02)	4.38 (0.04)
90	1.9	61.8 (0.16)	0.51 (0.01)	4.63 (0.06)	72 (14%)	0.50 (0.01)	4.87 (0.06)	71 (13%)	0.47 (0.01)	4.24 (0.06)
99	3.9	48.7 (0.21)	0.04 (0.00)	4.10 (0.20)	90 (46%)	0.05 (0.00)	5.62 (0.17)	79 (38%)	0.04 (0.00)	4.13 (0.20)

Although the BH procedure as implemented in the previous paragraph does make some use of knowledge of the number m of signals through the choice of its fixed sample size n for each m , the decision rule itself does not make explicit use of m and may not necessarily reject exactly m null hypotheses. In order to compare with a fixed-sample procedure that does make explicit use of m , like the gap rule, we also implemented a fixed-sample procedure (denoted BH_m) that always rejects exactly m null hypotheses by sampling n observations from each stream and rejecting the m null hypotheses with the smallest p -values. Such a procedure makes no use of a nominal level α , so we implemented BH_m by choosing n so that its achieved FDR and FNR are close to, but no larger than, the nominal levels $\alpha = \beta = .05$, similar to how the gap rule's threshold c was determined. The last three columns of Table 1, under the heading BH_m , give the operating characteristics of this procedure, including the savings in sample size achieved by the gap rule.

From the Savings columns in Table 1 we see that the sequential gap rule provides dramatic efficiency gains, in terms of average sample size, relative to the BH and BH_m procedures in this setting. The savings are in the 40%–60% range versus BH, and less so but still substantial in the 20%–45% range versus BH_m , which makes more explicit use of the true number of signals m than the BH procedure. The savings in sample size are largest for values of m near 0 and J , and decreases for m near $J/2$. The achieved FDR and FNR values are comparable among the three procedures.

Table 2 contains the operating characteristics of these procedures with parameters chosen in the analogous way for the $J = 100$ data stream version of the same setup. The three procedures have a similar relationship to each other as in the $J = 10$ setting, however the savings in sample size are less pronounced for this larger value of J .

6. Discussion

Summary and extensions

We have shown that sequential procedures proposed and shown to be asymptotically optimal by Song and Fellouris (2017) for FWE control can be made by modification of their critical values to be asymptotically optimal for control of other metrics, including FDR/FNR and pFDR/pFNR, in the setting of a known number of signals, or bounds on the true number of signals. One interpretation of these results is that first order optimality is not fine grain enough to distinguish between FWE, FDR, pFDR, and other metrics satisfying the main theorems, 3.1 and 4.1, however it remains an open question whether optimal procedures for control of these metrics *must* look different asymptotically.

The boundedness conditions (i)–(ii) in Theorems 3.1 and 4.1 are required to hold for all values of the nominal error probabilities α and β , but this is just to obtain admissibility of the gap and gap-intersection rules (Part 1 of Theorems 3.1 and 4.1, respectively) for all α and β , and is not required for asymptotic optimality (Part 2 of the theorems) which only need to consider α, β near 0. Thus, the optimality results can be further extended to multiple testing error metrics that only satisfy the boundedness conditions (i)–(ii) for small α, β .

Exclusion of the $\ell = 0, u = J$ cases for pFDR/pFNR control

In Section 4.1.2 we have ruled out the cases of the number of signals equal to 0 or J for pFDR/pFNR control based on the reasoning that either of these metrics are not defined for procedures or scenarios in which $P(R = 0) = 1$ or $P(R = J) = 1$, which are plausible when no, or all, streams contain signals. Focusing on the case of no signals (with similar remarks applying to the other case), the event $R = 0$ of no rejections has previously been recognized as a difficulty with utilizing pFDR in fixed-sample analyses, and so is not unique to the sequential sampling considered here. Even in Storey's (2002) original proposal for estimating pFDR, R must be replaced by $R \vee 1$ in the denominator of his estimator to avoid a singularity on the event $R = 0$, and the probability of this event must be bounded below, relying on independence and exact uniform distribution of the associated p -values. Black (2004, Section 4.3) points out that this approach introduces a bias into the

estimate of pFDR and argues that an improved estimator should indeed be left undefined on $R = 0$. These arguments further support our choice to leave out the cases $\ell = 0$ and $u = J$ in [Corollary 4.2](#).

The other metrics mentioned in [Section 4.2](#), to which the general theorem applies, do not have this problem because of elementary bounds like [\(28\)](#) (replacing m by u) and [\(29\)](#).

A referee suggested an alternative approach to pFDR/pFNR control that would include the number of signals being 0 or J by introducing a hybrid error metric that equals FDR/FNR on these cases and pFDR/pFNR otherwise. Since the bounds $C_1 = 1$ and $C_2 = 1/J$ are the same for both FDR/FNR control in [Corollary 4.1](#) and pFDR/pFNR control in [Corollary 4.2](#), it immediately follows that the common gap intersection rule would be asymptotically optimal for controlling this hybrid metric.

Relationship to classification problems

The problem addressed here may naturally be considered a classification problem, in which J objects are each classified into one of two respective classes, i.e., the corresponding null and alternative hypotheses. This is essentially the same setting for which classification via FDR thresholding has been investigated for fixed-sample data by many authors including [Abramovich et al. \(2006\)](#), [Bogdan et al. \(2008\)](#), [Donoho and Jin \(2006\)](#), [Genovese and Wasserman \(2002\)](#), and [Neuville and Roquain \(2012\)](#). FDR and related metrics have previously been considered as classifiers based on sequential data (e.g., [Bartroff, 2018](#); [Bartroff and Song, 2020](#)) but the current paper appears to be the first time asymptotic optimality has been considered and achieved for these metrics.

The setup considered here accommodates more specialized classification problems as well. For example, *slippage problems* (see [Dragalin et al., 1999](#); [Ferguson, 1967](#); [Mosteller, 1948](#)) consider J independent populations, of which at most one is in a non-null state, represented by the alternative hypothesis. In the fixed-sample setting, [Ferguson \(1967\)](#) considered a slippage problem in which both the null and alternative are known. [Tartakovsky \(1997\)](#) found minimax solutions for more general hypotheses, as did [Dragalin et al. \(1999\)](#) in a sequential Bayesian setting. The current paper can address slippage problems by taking the bounds in [Section 4](#) on the number of signals to be $\ell = 0$ and $u = 1$, since at most one stream has a signal, and allows the asymptotically optimal application of FDR/FNR or any other metric satisfying [Theorem 4.1](#). However, we note that the utility of FDR beyond FWE_1 is inherently limited in slippage problems because the two metrics will coincide for any procedure which chooses either $\ell = 0$ or $u = 1$ signals. This is because, using the notation in the sections above, both the number V of falsely rejected nulls and the number R of rejected nulls can only take the values 0 or 1, hence

$$\text{FDR} = E\left(\frac{V}{R \vee 1}\right) = E\left(\frac{V}{1}\right) = P(V = 1) = P(V \geq 1) = \text{FWE}_1. \quad (42)$$

However, the same does not hold for FNR and FWE_2 in this case hence, in spite of [\(42\)](#), the asymptotically optimal FDR/FNR control in the current paper is distinct from the FWE control of [Song and Fellouris \(2017\)](#) in slippage problems and its application here is novel, as well as any other metrics satisfying [Theorem 4.1](#).

CRedit authorship contribution statement

Xinrui He: Conceptualization, Formal analysis, Investigation, Methodology, Software, Writing - original draft, Writing - review & editing. **Jay Bartroff:** Conceptualization, Formal analysis, Investigation, Methodology, Supervision, Writing - original draft, Writing - review & editing.

Acknowledgments

The authors are grateful for the comments of two reviewers which improved the paper. The majority of He's work was completed while she was a Ph.D. student in the Department of Mathematics at the University of Southern California.

References

- [Abramovich, F., Benjamini, Y., Donoho, D.L., Johnstone, I.M., 2006. Adapting to unknown sparsity by controlling the false discovery rate. *Ann. Statist.* 34 \(2\), 584–653.](#)
- [Bartroff, J., 2018. Multiple hypothesis tests controlling generalized error rates for sequential data. *Statist. Sinica* 28, 363–398.](#)
- [Bartroff, J., Lai, T.L., 2010. Multistage tests of multiple hypotheses. *Comm. Statist. – Theory Methods* 39, 1597–1607, \(Special Issue Honoring M. Akahira, M. Aoshima, ed.\).](#)
- [Bartroff, J., Song, J., 2014. Sequential tests of multiple hypotheses controlling type I and II familywise error rates. *J. Statist. Plann. Inference* 153, 100–114.](#)
- [Bartroff, J., Song, J., 2015. A rejection principle for sequential tests of multiple hypotheses controlling familywise error rates. *Scand. J. Statist.* 43, 3–19.](#)
- [Bartroff, J., Song, J., 2020. Sequential tests of multiple hypotheses controlling false discovery and nondiscovery rates. *Sequential Anal.* 39, 65–91.](#)
- [Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300.](#)
- [Black, M., 2004. A note on the adaptive control of false discovery rates. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 66 \(2\), 297–304.](#)

- Bogdan, M., Ghosh, J.K., Tokdar, S.T., 2008. A comparison of the Benjamini-Hochberg procedure with some Bayesian rules for multiple testing. In: Balakrishnan, N., Peña, E.A., Silvapulle, M.J. (Eds.), *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, Vol. 1. Institute of Mathematical Statistics, Beachwood, Ohio, USA, pp. 211–230.
- Burke, D.S., Brundage, J.F., Redfield, R.R., Damato, J.J., Schable, C.A., Putman, P., Visintine, R., Kim, H.I., 1988. Measurement of the false positive rate in a screening program for human immunodeficiency virus infections. *New Engl. J. Med.* 319 (15), 961–964.
- Chen, S., Arias-Castro, E., 2017. *Sequential multiple testing*. ArXiv preprint, <http://arxiv.org/abs/1705.10190>.
- De, S., Baron, M., 2012a. Sequential Bonferroni methods for multiple hypothesis testing with strong control of family-wise error rates I and II. *Sequential Anal.* 31 (2), 238–262.
- De, S., Baron, M., 2012b. Step-up and step-down methods for testing multiple hypotheses in sequential experiments. *J. Statist. Plann. Inference* 142, 2059–2070.
- Donoho, D., Jin, J., 2006. Asymptotic minimaxity of false discovery rate thresholding for sparse exponential data. *Ann. Statist.* 34 (6), 2980–3018.
- Dragalin, V.P., Tartakovsky, A.G., Veeravalli, V.V., 1999. Multihypothesis sequential probability ratio tests I: Asymptotic optimality. *IEEE Trans. Inform. Theory* 45 (7), 2448–2461.
- Ferguson, T.S., 1967. *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, New York.
- Frane, A.V., 2015. Are per-family type I error rates relevant in social and behavioral science? *J. Modern Appl. Statist. Methods* 14 (1).
- Genovese, C., Wasserman, L., 2002. Operating characteristics and extensions of the false discovery rate procedure. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 64 (3), 499–517.
- Gut, A., 2009. *Stopped Random Walks*. Springer, New York.
- Javanmard, A., Montanari, A., 2018. Online rules for control of false discovery rate and false discovery exceedance. *Ann. Statist.* 46 (2), 526–554.
- Jiang, H., Salzman, J., 2012. Statistical properties of an early stopping rule for resampling-based multiple testing. *Biometrika* 99 (4), 973–980.
- Keselman, H.J., 2015. Per family or familywise type I error control: Eether, eyether, neether, nyther, let's call the whole thing off!. *J. Modern Appl. Statist. Methods* 14 (1), 24–37.
- Malloy, M.L., Nowak, R.D., 2014. Sequential testing for sparse recovery. *IEEE Trans. Inform. Theory* 60 (12), 7862–7873.
- Mei, Y., 2008. Asymptotic optimality theory for decentralized sequential hypothesis testing in sensor networks. *IEEE Trans. Inform. Theory* 54 (5), 2072–2089.
- Mosteller, F., 1948. A k -sample slippage test for an extreme population. *Ann. Math. Stat.* 19 (1), 58–65.
- Neuviel, P., Roquain, E., 2012. On false discovery rate thresholding for classification under sparsity. *Ann. Statist.* 40 (5), 2572–2600.
- Song, Y., Fellouris, G., 2017. Asymptotically optimal, sequential, multiple testing procedures with prior information on the number of signals. *Electron. J. Stat.* 11 (1), 338–363.
- Song, Y., Fellouris, G., 2019. Sequential multiple testing with generalized error control: An asymptotic optimality theory. *Ann. Statist.* 47 (3), 1776–1803.
- Storey, J.D., 2002. A direct approach to false discovery rates. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 64 (3), 479–498.
- Tartakovsky, A.G., 1997. Minimax invariant regret solution to the N -sample slippage problem. *Math. Methods Statist.* 6 (4), 491–508.
- Wegman, E.J., Marchette, D.J., 2003. On some techniques for streaming data: a case study of internet packet headers. *J. Comput. Graph. Statist.* 12 (4), 893–914.
- Zrnic, T., Ramdas, A., Jordan, M.I., 2018. Asynchronous online testing of multiple hypotheses. ArXiv preprint, <http://arxiv.org/abs/1812.05068>.