Optimal Rates and Tradeoffs in Multiple Testing

Maxim Rabinovich†, Aaditya Ramdas*,†
Michael I. Jordan*,†, Martin J. Wainwright*,†

{rabinovich,aramdas,jordan,wainwrig}@berkeley.edu

Departments of Statistics* and EECS†, University of California, Berkeley

May 12, 2017

Abstract

Multiple hypothesis testing is a central topic in statistics, but despite abundant work on the false discovery rate (FDR) and the corresponding Type-II error concept known as the false non-discovery rate (FNR), a fine-grained understanding of the fundamental limits of multiple testing has not been developed. Our main contribution is to derive a precise non-asymptotic tradeoff between FNR and FDR for a variant of the generalized Gaussian sequence model. Our analysis is flexible enough to permit analyses of settings where the problem parameters vary with the number of hypotheses n, including various sparse and dense regimes (with $o(n)$ and $\mathcal{O}(n)$ signals). Moreover, we prove that the Benjamini-Hochberg algorithm as well as the Barber-Candès algorithm are both rate-optimal up to constants across these regimes.

1 Introduction

The problem of multiple comparisons has been a central topic in statistics ever since Tukey's influential 1953 book [27]. In broad terms, suppose that one observes a sequence of n independent random variables $X_1, \ldots, X_n$, of which some unknown subset are drawn from a null distribution, corresponding to the absence of a signal or effect, whereas the remainder are drawn from a non-null distribution, corresponding to signals or effects. Within this framework, one can pose three problems of increasing hardness: the *detection* problem of testing whether or not there is at least one signal; the *localization* problem of identifying the positions of the nulls and signals; and the *estimation* problem of returning estimates of the means and/or distributions of the observations. Note that these problems form a hierarchy of difficulty: identifying the signals implies that we know whether there is at least one of them, and estimating each mean implies we know which are zero and which are not. The focus of this paper is on the problem of localization.

There are a variety of ways of measuring type I errors for the localization problem, including the *family-wise error rate*, which is the probability of incorrectly rejecting at least one null, and the *false discovery rate* (FDR), which is the expected ratio of incorrect rejections to total rejections. An extensive literature has developed around both of these metrics, resulting in algorithms geared towards controlling one or the other. Our focus is the FDR metric, which has been widely studied, but for which relatively little is known about the behavior of existing algorithms in terms of the corresponding Type-II error concept, namely the *false non-discovery rate* (FNR).[1] Indeed, it is only very recently that Arias-Castro and Chen [2], working within a version of the sparse generalized Gaussian sequence model, established asymptotic consistency for the FDR-FNR localization problem. Informally, in this framework, we receive n independent observations $X_1, \ldots, X_n$, out of which

[1]We follow Arias-Castro and Chen [2] in defining the FNR as the ratio of undiscovered to total non-nulls, which differs from the definition of Genovese and Wasserman [15].

$n - n^{1-\beta_n}$ are nulls, and the remainder are non-nulls. The null variables are drawn from a centered distribution with tails decaying as $\exp\left(-\frac{|x|^\gamma}{\gamma}\right)$, whereas the remaining $n^{1-\beta_n}$ non-nulls are drawn from the same distribution shifted by $(\gamma r_n \log n)^{1/\gamma}$. Using this notation, Arias-Castro and Chen [2] considered the setting with fixed problem parameters $r_n = r$ and $\beta_n = \beta$, and showed that when $r < \beta < 1$, all procedures must have risk FDR + FNR $\to 1$. They also showed that in the achievable regime $r > \beta > 0$, the Benjamini-Hochberg (BH) is consistent, meaning that FDR + FNR $\to 0$. Finally, they proposed a new "distribution-free" method inspired by the knockoff procedure by Barber and Candès [13], and they showed that the resulting procedure is also consistent in the achievable regime.

These existing consistency results are asymptotic. To date there has been no study of the important non-asymptotic questions that are of interest in comparing procedures. For instance, for a given FDR level, what is the best possible achievable FNR? What is the best-possible non-asymptotic behavior of the risk FDR + FNR attainable in finite samples? And, perhaps most importantly, non-asymptotic questions regarding whether or not procedures such as BC and BH are *rate-optimal* for the FDR+FNR risk—remain unanswered. The main contributions of this paper are to develop techniques for addressing such questions, and to essentially resolve them in the context of the sparse generalized Gaussians model.

Specifically, we establish the tradeoff between FDR and FNR in finite samples (and hence also asymptotically), and we use the tradeoff to determine the best attainable rate for the FDR + FNR risk. Our theory is sufficiently general to accommodate sequences of parameters (r_n, β_n), and thereby to reveal new phenomena that arise when $r_n - \beta_n = o(1)$. For a fixed pair of parameters (r, β) in the achievable regime $r > \beta$, our theory leads to an explicit expression for the optimal rate at which FDR+FNR can decay. In particular, defining the γ-"distance" $D_\gamma(a, b) := \left|a^{1/\gamma} - b^{1/\gamma}\right|^\gamma$ between pairs of positive numbers, we show that the equation

$$\kappa = D_\gamma(\beta + \kappa, r)$$

has a unique solution κ_*, and moreover that the combined risk of any threshold-based multiple testing procedure $\mathcal{I}$ is lower bounded as $\mathcal{R}_n(\mathcal{I}) \gtrsim n^{-\kappa_*}$. Moreover, by direct analysis, we are able to prove that both the Benjamini-Hochberg (BH) and the Barber-Candès (BC) algorithms attain this optimal rate.

At the core of our analysis is a simple comparison principle, and the flexibility of the resulting proof strategy allows us to identify a new critical regime in which $r_n - \beta_n = o(1)$, but the problem is infeasible, meaning that if the FDR is driven to zero, then the FNR must remain bounded away from zero. Moreover, we are able to study some challenging settings in which the fraction of signals is a constant $\pi_1 \in (0, 1)$ and not asymptotically vanishing, which corresponds to the setting $\beta_n = \frac{\log(1/\pi_1)}{\log n}$, so that $\beta_n \to 0$. Perhaps surprisingly, even in these regimes, the BH and BC algorithms continue to be optimal, though the best rate can weaken from polynomial to subpolynomial in the number of hypotheses n.

1.1 Related work

As noted above, our work provides a non-asymptotic generalization of recent work by Arias-Castro and Chen [2] on asymptotic consistency in localization, using FDR + FNR as the notion of risk. It should be noted that this notion of risk is distinct from the asymptotic Bayes optimality under sparsity (ABOS) studied in past work by Bogdan et al. [7] for Gaussian sequences, and more recently by Neuvial and Roquain [23] for binary classification with extreme class imbalance. The ABOS results concern a risk derived from the probability of incorrectly rejecting a single null sample (false positive, or FP for short) and the probability of incorrectly failing to reject a single non-null sample (false negative, or FN for short). Concretely, one has $\mathcal{R}_n^{\text{ABOS}} = w_1 \cdot \text{FP} + w_2 \cdot \text{FN}$ for some pair of positive weights (w_1, w_2) that need not be equal. As this risk is based on the error probability for a single sample, it is much closer to misclassification risk or single-testing risk than to the ratio-based FDR + FNR risk studied in this paper.

Using the notation of this paper, the work of Neuvial and Roquain [23] can be understood as focusing on the particular setting $r = \beta$, a regime referred to as the "verge of detectability" by these authors, and with performance metric given by the Bayes classification risk, rather than the combination of FDR and FNR studied here. In comparison, our results provide additional insight into models that are close to the verge of detectability, in that even when $\beta_n = \beta$ is fixed, we can provide quantitative lower and upper bounds on the FDR/FNR ratio as $r_n \to \beta$ from above; moreover, these bounds depend on how quickly r_n approaches β. These conclusions actually make it clear that a further transition in rates occurs in the case where $r = \beta$ exactly for all n, though we do not explore the latter case in depth. We suspect that the methods developed in this paper may have sufficient precision to answer the non-asymptotic minimaxity questions posed by Neuvial and Roquain [23] as to whether any threshold-based procedure can match the Bayes optimal classification error rate up to an additive error $\ll \frac{1}{\log n}$.

The above line of work is complementary to the well-known asymptotic results by Donoho and Jin [9, 11] on phase transitions in detectability using Tukey's higher-criticism statistic, employing the standard type-I and type-II errors for testing of the single global null hypothesis. Note that Donoho and Jin use the generalized Gaussian assumption directly on the PDFs, while our assumption (5) is on the survival function. Just as in Arias-Castro and Chen [2], Donoho and Jin also consider the asymptotic setting[2] with $r_n = r$ and $\beta_n = \beta$, which they sometimes call the ARW (asymptotic, rare and weak) model.

Our paper is also complementary to work on estimation, the most notable result being the asymptotic minimax optimality of BH-derived thresholding for denoising an approximately-sparse high-dimensional vector [1, 10]. The relevance of our results on the minimaxity of BH for approximately-sparse denoising problems lies primarily in the use of deterministic thresholds as a useful proxy for BH and other procedures that determine their threshold in a manner that has complex dependence on the input data [10]. Unlike the strategy of Donoho and Jin [10], which depends on establishing concentration of the empirical threshold around the population-level value, we use a more flexible comparison principle. Deterministic approximations to optimal FDR thresholds are also studied by Chi [8] and Genovese et al. [16]. Other related papers are discussed in Section 5, when discussing directions for future work.

The remainder of this paper is organized as follows. In Section 2, we provide background on the multiple testing problem, as well as the particular model we consider. In Section 3, we provide an overview of our main results: namely, optimal tradeoffs between FDR and FNR, which imply lower bounds on the FDR+FNR risk, and optimality guarantees for the BH and BC algorithms. In Section 4, we prove our main results, focusing first on the lower bounds and then using the ideas we have developed to provide matching upper bounds for the well-known and popular Benjamini-Hochberg (BH) procedure and the recent Barber-Candès (BC) algorithm for multiple testing with FDR control. Proofs of some technical lemmas are given in the apepndices.

2 Problem formulation

In this section, we provide background and a precise formulation of the problem under study.

[2] We are not aware of any non-asymptotic results for detection akin to the results that the current paper provides for localization.

2.1 Multiple testing and false discovery rate

Suppose that we observe a real-valued sequence $X_1^n := \{X_1, \ldots, X_n\}$ of n independent random variables. When the null hypothesis is true, X_i is assumed to have zero mean; otherwise, it is assumed that the mean of X_i is some unknown number $\mu_n > 0$. We introduce the sequence of binary labels $\{H_1, \ldots, H_n\}$ to encode whether or not the null hypothesis holds for each observation; the setting $H_i = 0$ indicates that the null hypothesis holds. We define

$$\mathcal{H}_0 := \{i \in [n] \mid H_i = 0\}, \quad \text{and} \quad \mathcal{H}_1 := \{i \in [n] \mid H_i = 1\}, \tag{1}$$

corresponding to the *nulls* and *signals*, respectively. Our task is to identify a subset of indices that contains as many signals as possible, while not containing too many nulls.

More formally, a testing rule $\mathcal{I} : \mathbb{R}^n \to 2^{[n]}$ is a measurable mapping of the observation sequence X_1^n to a set $\mathcal{I}(X_1^n) \subseteq [n]$ of *discoveries*, where the subset $\mathcal{I}(X_1^n)$ contains those indices for which the procedure rejects the null hypothesis. There is no single unique measure of performance for a testing rule for the localization problem. In this paper, we study the notion of the *false discovery rate* (FDR), paired with the *false non-discovery rate* (FNR). These can be viewed as generalizations of the type-I and type-II errors for single hypothesis testing.

We begin by defining the false discovery proportion (FDP), and false non-discovery proportion (FNP), respectively, as

$$\mathrm{FDP}_n(\mathcal{I}) := \frac{\mathrm{card}(\mathcal{I}(X_1^n) \cap \mathcal{H}_0)}{\mathrm{card}(\mathcal{I}(X_1^n)) \vee 1}, \quad \text{and} \quad \mathrm{FNP}_n(\mathcal{I}) := \frac{\mathrm{card}(\mathcal{I}(X_1^n) \cap \mathcal{H}_1)}{\mathrm{card}(\mathcal{H}_1)}. \tag{2}$$

Since the output $\mathcal{I}(X_1^n)$ of the testing procedure is random, both quantities are random variables. The FDR and FNR are given by taking the expectations of these random quantities—that is

$$\mathrm{FDR}_n(\mathcal{I}) := \mathbb{E}\Big[\frac{\mathrm{card}(\mathcal{I}(X_1^n) \cap \mathcal{H}_0)}{\mathrm{card}(\mathcal{I}(X_1^n)) \vee 1}\Big], \quad \text{and} \quad \mathrm{FNR}_n(\mathcal{I}) := \mathbb{E}\Big[\frac{\mathrm{card}(\mathcal{I}(X_1^n) \cap \mathcal{H}_1)}{\mathrm{card}(\mathcal{H}_1)}\Big], \tag{3}$$

where the expectation is taken over the random samples X_1^n. In this paper, we measure the overall performance of a given procedure in terms of its *combined risk*

$$\mathcal{R}_n(\mathcal{I}) := \mathrm{FDR}_n(\mathcal{I}) + \mathrm{FNR}_n(\mathcal{I}). \tag{4}$$

Finally, when the testing rule $\mathcal{I}$ under discussion is clear from the context, we frequently omit explicit reference to this dependence from all of these quantities.

2.2 Tail generalized Gaussians model

In this paper, we describe the distribution of the observations for both nulls and non-nulls in terms of a *tail generalized Gaussians model*. Our model is a variant of the generalized Gaussian sequence model studied in past work [2, 9]; the only difference is that whereas a γ-generalized Gaussian has a density proportional to $\exp\big(-\frac{|x|^\gamma}{\gamma}\big)$, we focus on distributions whose tails are proportional to $\exp\big(-\frac{|x|^\gamma}{\gamma}\big)$. This alteration is in line with the asymptotically generalized Gaussian (AGG) distributions studied by Arias-Castro and Chen [2], with the important caveat that our assumptions are imposed in a non-asymptotic fashion.

For a given degree $\gamma \geq 1$, a γ-tail generalized Gaussian random variable with mean 0, written as $G \sim \mathrm{tGG}_\gamma(0)$, has a survival function $\Psi(t) := \mathbb{P}\big(G \geq t\big)$ that satisfies the bounds

$$\frac{e^{\frac{-|t|^\gamma}{\gamma}}}{Z_\ell} \;\leq\; \min\{\Psi(t), 1 - \Psi(t)\} \;\leq\; \frac{e^{\frac{-|t|^\gamma}{\gamma}}}{Z_u}, \tag{5}$$

for some constants $Z_\ell > Z_u > 0$. (Note that $t \mapsto \Psi(t)$ is a decreasing function, and becomes smaller than $1-\Psi(t)$ at the origin.) As a concrete example, a γ-tail generalized Gaussian with $Z_\ell = Z_u = 1$ can be generated by sampling a standard exponential random variable E and a Rademacher random variable ε and putting $G = \varepsilon\big(\gamma E\big)^{1/\gamma}$. We use the terminology "tail generalized Gaussian" because of the following connection: the survival function of a 2-tail Gaussian random variable is on the order of $\exp(-|x|^2/2)$, whereas that of a Gaussian is on the order of $\frac{1}{\mathrm{poly}(x)}\exp(-x^2/2)$. In particular, this observation implies a tGG_2 random variable has tails that are equivalent to a Gaussian in terms of their exponential decay rates.

In terms of this notation, we assume that each observation X_i is distributed as

$$X_i \sim \begin{cases} \mathrm{tGG}_\gamma(0) & \text{if } i \in \mathcal{H}_0 \\ \mathrm{tGG}_\gamma(0) + \mu_n & \text{if } i \in \mathcal{H}_1, \end{cases} \tag{6}$$

where our notation reflects the fact that the mean shift μ_n is permitted to vary with the number of observations n. See Section 3.1 for further discussion of the scaling of the mean shift.

2.3 Threshold-based procedures

Following prior work [2, 9], we restrict attention to testing procedures of the form

$$\mathcal{I}(X_1^n) = \big\{i \in [n] \mid X_i \geq T_n(X_1^n)\big\}, \tag{7}$$

where $T_n(X_1^n) \in \mathbb{R}_+$ is a data-dependent threshold. We refer to such methods as *threshold-based procedures*. The BH and BC procedures both belong to this class. Moreover, from an intuitive standpoint, the observations are exchangeable in the absence of prior information, and we are considering testing between a single unimodal null distribution and a single positive shift of that distribution. In this setting, it is hard to imagine that an optimal procedure would ever reject the hypothesis corresponding to one observation while rejecting a hypothesis with a smaller observation value. Threshold-based procedures therefore appear to be a very reasonable class to focus on.

It will be convenient to reason about the performance metrics associated with rules of the form

$$\mathcal{I}_t(X_1^n) = \big\{i \in [n] \mid X_i \geq t\big\}, \tag{8}$$

where $t > 0$ is a pre-specified (fixed, non-random) threshold. In this case, we adopt the notation $\mathrm{FDR}_n(t)$, $\mathrm{FNR}_n(t)$ and $\mathcal{R}_n(t)$ to denote the metrics associated with the rule $X_1^n \mapsto \mathcal{I}_t(X_1^n)$.

2.4 Benjamini-Hochberg (BH) and Barber-Candès (BC) procedures

Arguably the most popular threshold-based procedure that provably controls FDR at a user-specified level q_n is the *Benjamini-Hochberg* (BH) procedure. More recently, Arias-Castro and Chen [2] proposed a method that we refer to as the *Barber-Candès* (BC) procedure. Both algorithms are based on estimating the FDP_n that would be incurred at a range of possible thresholds and choosing one that is as large as possible (maximizing discoveries) while satisfying an upper bound linked to q_n (controlling FDR_n). Further, they both only consider thresholds that coincide with one of the values X_1^n, which we denote as a set by $\mathcal{X}_n = \big\{X_1, \ldots, X_n\big\}$. The data-dependent threshold for both can be written as

$$t_n\big(X_1, \ldots, X_n\big) = \max\big\{t \in \mathcal{X}_n \colon \widehat{\mathrm{FDP}}_n\big(t\big) \leq q_n\big\}. \tag{9}$$

The two algorithms differ in the estimator $\widehat{\mathrm{FDP}}_n(t)$ they use. The BH procedure assumes access to the true null distribution through its survival function Ψ and sets

$$\widehat{\mathrm{FDP}}_n^{\mathrm{BH}}(t) = \frac{\Psi(t)}{\#(X_i \geq t)/n}, \quad \text{for } t \in \mathcal{X}_n. \tag{10}$$

The BC procedure instead estimates the survival function $\Psi(t)$ from the data and therefore does not even need to know the null distribution. This approach is viable when $\#(X_i \leq -t)/n$ is a good proxy for $\Psi(t)$, which our upper and lower tail bounds guarantee; more typically, the BC procedure is applicable when the null distribution is (nearly) symmetric, and the signals are shifted by a positive amount (as they are in our case). Then, the BC estimator is given by

$$\widehat{\mathrm{FDP}}_n^{\mathrm{BC}}(t) = \frac{[\#(X_i \leq -t) + 1]/n}{\#(X_i \geq t)/n}, \quad \text{for } t \in \mathcal{X}_n. \tag{11}$$

With these definitions in place, are now ready to describe our main results.

3 Main results

We now turn to a statement of our main results, along with some illustrations of their consequences. Our first main result (Theorem 1) characterizes the optimal tradeoff between FDR and FNR for any testing procedure. By optimizing this tradeoff, we obtain a lower bound on the combined FDR and FNR of any testing procedure (Corollary 1). Our second main result (Theorem 2), shows that BH achieves the optimal FDR-FNR tradeoff up to constants and that BC almost achieves it. In particular, our result implies that with the proper choice of target FDR, both BH and BC can achieve the optimal combined FDR-FNR rate (Corollary 2).

3.1 Scaling of sparsity and mean shifts

We study a sparse instance of the multiple testing problem in which the number of signals is assumed to be small relative to the total number of hypotheses. In particular, motivated by related work in multiple hypothesis testing [2, 9, 11, 21], we assume that the number of signals scales as

$$\mathrm{card}(\mathcal{H}_1) = m_n = n^{1-\beta_n} \quad \text{for some } \beta_n \in (0, 1). \tag{12}$$

Note that to the best of our knowledge, all previous results in the literature assume that $\beta_n = \beta$ is actually independent of n. In this case, the sparsity assumption (12) implies that all but a polynomially vanishing fraction of the hypotheses are null. In contrast, as indicated by our choice of notation, the set-up in this paper allows for a sequence of parameters β_n that can vary with the number of hypotheses n. In this way, our framework is flexible enough to handle relatively dense regimes (e.g., those with $\frac{n}{\log n}$ or even $\mathcal{O}(n)$ signals).

The non-null hypotheses are distinguished by a positively shifted mean $\mu_n > 0$. It is natural to parameterize this mean shift in terms of a quantity $r_n > 0$ via the relation

$$\mu_n = \left(\gamma r_n \log n\right)^{1/\gamma}. \tag{13}$$

As shown by Arias-Castro and Chen [2], when the pair (β, r) are fixed such that $r < \beta$, the problem is asymptotically infeasible, meaning that there is no procedure such that $\mathcal{R}_n(\mathcal{I}) \to 0$ as $n \to \infty$. Accordingly, we focus on sequences (β_n, r_n) for which $r_n > \beta_n$. Further, even though the asymptotic consistency boundary of $r < \beta$ versus $r > \beta$ is apparently independent of γ, we will see that the rate at which the risk decays to zero is determined jointly by r, β and γ.

3.2 Lower bound on any threshold-based procedure

In this section, we assume :

$$\beta_n \overset{(i)}{\geq} \frac{\log 2}{\log n} \iff n^{1-\beta_n} \leq n/2, \quad \text{and} \tag{14a}$$

$$\max\{\beta_n, \frac{1}{\log^{\frac{\gamma-1/2}{\gamma}} n}\} \overset{(ii)}{<} r_n \overset{(iii)}{<} r_{\max} \quad \text{for some constant } r_{\max} < 1. \tag{14b}$$

Condition (i) requires that the proportion π_1 of non-nulls is at most $1/2$. Condition (ii) asserts that the natural requirement of $r_n > \beta_n$ is not enough, but further insists that r_n cannot approach zero too fast. The constants $\log 2$ and $\frac{\gamma-1/2}{\gamma}$ are somewhat arbitrary and can be replaced, respectively, by $\log \frac{1}{\pi_{\max}}$ for any $0 < \pi_{\max} < 1$ and $\frac{\gamma-1+\rho}{\gamma}$ for any $\rho > 0$, but we fix their values in order not to introduce unnecessary extra parameters. As for condition (iii), although the assumption $r_n < 1$ is imposed because the problem becomes qualitatively easy for $r_n \geq 1$, the assumption that it is bounded away from one is a technical convenience that simplifies some of our proofs.

Our analysis shows that the FNR behaves differently depending on the closeness of the parameter r_n to the boundary of feasiblity given by β_n. In order to characterize this closeness, we define

$$r_{\min} = r_{\min}(\kappa_n) := \begin{cases} \beta_n + \kappa_n + \frac{\log \frac{1}{6Z_\ell}}{\log n} & \text{if } \kappa_n \leq 1 - \beta_n - \frac{\log \frac{3}{\log 16}}{\log n}, \\ 1 + \frac{\log \frac{1}{24Z_\ell}}{\log n} & \text{otherwise.} \end{cases} \tag{15}$$

Here κ_n is to be interpreted as the "exponent" of a target FDR rate q_n, in the sense that $q_n = n^{-\kappa_n}$. The rate q_n may differ from the actual achieved FDR_n, but it is nonetheless useful for parameterizing the quantities that enter into our analysis. When we need to move between q_n and κ_n, we shall write $\kappa_n = \kappa_n(q_n) = \frac{\log(1/q_n)}{\log n}$ and $q_n = q_n(\kappa_n) = n^{-\kappa_n}$. For mathematical convenience, we wish to have the target FDR q_n to be bounded away from one, and we therefore impose one further technical but inessential assumption in this section:

$$q_n \leq \min\left\{\frac{1}{24}, \frac{1}{6Z_\ell}\right\} \iff \kappa_n \geq \frac{\log \max\left\{24, 6Z_\ell\right\}}{\log n}. \tag{16}$$

The theorem that follows will apply to all sample sizes $n > n_{\min,\ell}$ (subscript ℓ for lower), where

$$n_{\min,\ell} := \min\left\{n \in \mathbb{N} : \exp\left(-\frac{n^{1-r_{\max}}}{24(Z_\ell \vee 1)}\right) \leq \frac{1}{4}\right\} = \left\lfloor [24(Z_\ell \vee 1)\log 4]^{\frac{1}{1-r_{\max}}} \right\rfloor. \tag{17}$$

Finally, for $\gamma \in [1, \infty)$ and non-negative numbers $a, b > 0$, let us define the associated γ-"distance":

$$D_\gamma(a, b) := \left|a^{1/\gamma} - b^{1/\gamma}\right|^\gamma. \tag{18}$$

Our first main theorem states that for $r_n > r_{\min}(\kappa_n)$, the FNR decays as a power of $1/n$, with exponent specified by the γ-distance.

Theorem 1. *Consider the γ-tail generalized Gaussians testing problem with sparsity β_n and signal level r_n satisfying conditions (14a), and (14b), and with sample size $n > n_{\min,\ell}$ from definition (17). Then, for any choice of exponent $\kappa_n \in (0,1)$ satisfying condition (16), there exists a minimum signal strength $r_{\min}(\kappa_n)$ from definition (15), such that any threshold-based procedure $\mathcal{I}$ that satisfies* $\mathrm{FDR}_n(\mathcal{I}) \leq n^{-\kappa_n}$ *must have its* FNR *lower bounded as*

$$\mathrm{FNR}_n(\mathcal{I}) \geq \begin{cases} \frac{1}{32} & \text{if } r_n \in \left[\beta_n, r_{\min}\right] \\ c(\beta_n, \gamma)\, n^{-D_\gamma(\beta_n+\kappa_n, r_n)} & \text{otherwise,} \end{cases} \tag{19}$$

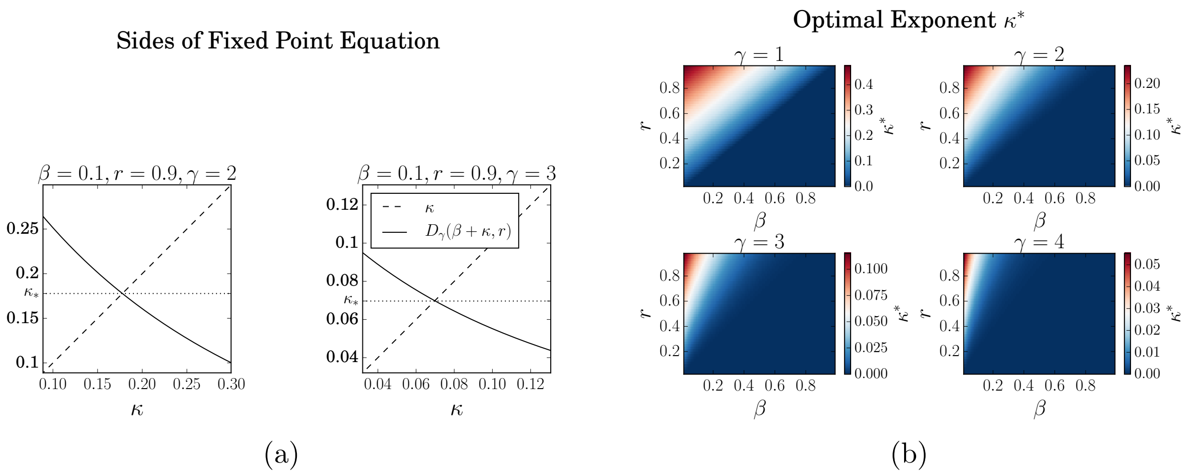


Figure 1. Visualizations of the fixed-point equation (20). (a) Plots comparing the left- and right-hand sides of the fixed-point equation. (b) The optimal exponent κ^* as a function of r and β.

where $c(\beta_n, \gamma) := c_0 \exp\left(c_1 \beta_n^{\frac{1-\gamma}{\gamma}}\right)$*, with* (c_0, c_1) *being positive constants depending only on* (Z_ℓ, Z_u, γ)*.*

The proof of this theorem is provided in Section 4.1. Note that the theorem holds for any choice of $\kappa_n \in (0, 1)$. In the special case of constant pairs (β, r), this choice can be optimized to achieve the best possible lower bound on the risk $\mathcal{R}_n(\mathcal{I}) = \mathrm{FDR}_n(\mathcal{I}) + \mathrm{FNR}_n(\mathcal{I})$, as summarized below.

Corollary 1. *When* $r > \beta$*, let* $\kappa_* = \kappa_*(\beta, r, \gamma) > 0$ *be the unique solution to the equation*

$$\kappa = D_\gamma\left(\beta + \kappa, r\right). \tag{20}$$

Then the combined risk of any threshold-based multiple testing procedure $\mathcal{I}$ *is lower bounded as*

$$\mathcal{R}_n(\mathcal{I}) \gtrsim n^{-\kappa_*}, \tag{21}$$

where $\gtrsim$ *denotes inequality up to a pre-factor independent of* n*.*

The proof of this corollary is provided in Section 4.2. Figure 1 provides an illustration of the predictions in Corollary 1. In particular, panel (a) shows how the unique solution κ_* to equation (20) is determined for varying settings of the triple (r, β, γ). Panel (b) shows how κ_* varies over the interval $(0, 0.5)$, again for different settings of the triple (r, β, γ). As would be expected, the fixed point κ_* increases as a function of the difference $r - \beta > 0$.

3.3 Upper bounds for some specific procedures

Thus far, we have provided general lower bounds applicable to any threshold procedure. We now turn to the complementary question—how do these lower bounds compare to the results achievable by the BH and BC algorithms introduced in Section 2.3? Remarkably, we find that up to the constants defining the prefactor, both the BH and BC procedures achieve the minimax lower bound of Theorem 1.

We state these achievable results in terms of the fixed point κ_* from equation (20). Moreover, they apply to all problems with sample size $n > n_{\min,u}$ (subscript u for upper), where

$$\begin{aligned} n_{\min,u} \;&:=\; \min\left\{n \in \mathbb{N} : \exp\left(-\frac{n^{1-r_{\max}}}{24}\right) \leq \frac{1}{Z_u n}\right\} \\ &=\; \min\left\{n \in \mathbb{N} : n \;\geq\; \left[24 \log(Z_u n)\right]^{\frac{1}{1-r_{\max}}}\right\}. \end{aligned} \tag{22}$$

In order to state our results cleanly, let us introduce the constants

$$c_{\mathrm{BH}} \;:=\; \frac{Z_u}{36 Z_\ell}, \quad c_{\mathrm{BC}} \;:=\; \frac{Z_u}{48 Z_\ell}, \quad \text{and} \quad \zeta \;:=\; \max\left\{6 Z_\ell,\; \frac{1}{6 Z_\ell}\right\}, \tag{23}$$

and require in particular that $r_n \geq r_{\min}\left(\kappa_n\left(c_{\mathrm{A}} q_n\right)\right)$ for algorithm $\mathrm{A} \in \{\mathrm{BH}, \mathrm{BC}\}$. Note that $c_{\mathrm{A}} < 1$ since $Z_\ell \geq Z_u$ by definition, and that the introduction of c_{A} into the argument of $r_{\min}$ only changes the minimum allowed value of r_n by a conceptually negligible amount of $\mathcal{O}\big(\frac{1}{\log n}\big)$.

Lastly, we note that BC requires an additional mild condition that the number of non-nulls $n^{1-\beta_n}$ is large relative to the target FDR $q_n = n^{-\kappa_n}$ (otherwise, in some sense, the problem is too hard if there are too few non-nulls and a very strict target FDR). Specifically, we need that both quantities cannot simultaneously be too small, formalized by the assumption:

$$\exists n_{\min,\mathrm{BC}} \;\text{ such that for all }\; n \geq n_{\min,\mathrm{BC}} \;\text{ we have }\; \frac{3 c_{\mathrm{BC}}}{4} \cdot \frac{q_n}{\log \frac{1}{q_n}} \cdot n^{1-\beta_n} \;\geq\; 1. \tag{24}$$

We note that when $r_n = r$ and $\beta_n = \beta$ are constants, this decay condition is satisfied by $q_n = n^{-\kappa_*}$.

Our second main theorem delivers an optimality result for the BH and BC procedures, showing that under some regularity conditions, their performance achieves the lower bounds in Theorem 1 up to constant factors.

Theorem 2. *Consider the β_n-sparse γ-tail generalized Gaussians testing problem with target FDR level q_n upper bounded as in condition (16).*

(a) Guarantee for BH procedure: Given a signal strength $r_n \geq r_{\min}(\kappa_n(c_{\mathrm{BH}} q_n))$ and sample size $n > n_{\min,u}$ as in condition (22), the BH procedure satisfies the bounds

$$\mathrm{FDR}_n \leq q_n \quad \textit{and} \quad \mathrm{FNR}_n \leq \frac{2 \zeta_{\mathrm{BH}}^{2\beta_n^{\frac{1-\gamma}{\gamma}}}}{Z_u} \cdot n^{-D_\gamma(\beta_n + \kappa_n, r_n)}, \qquad \textit{where } \zeta_{\mathrm{BH}} := \tfrac{\zeta}{c_{\mathrm{BH}}}. \tag{25}$$

(b) Guarantee for BC procedure: Given a signal strength $r_n \geq r_{\min}(\kappa_n(c_{\mathrm{BC}} q_n))$ and sample size $n > \max\{n_{\min,\mathrm{BC}}, n_{\min,u}\}$ as in condition (24), the BC procedure satisfies the bounds

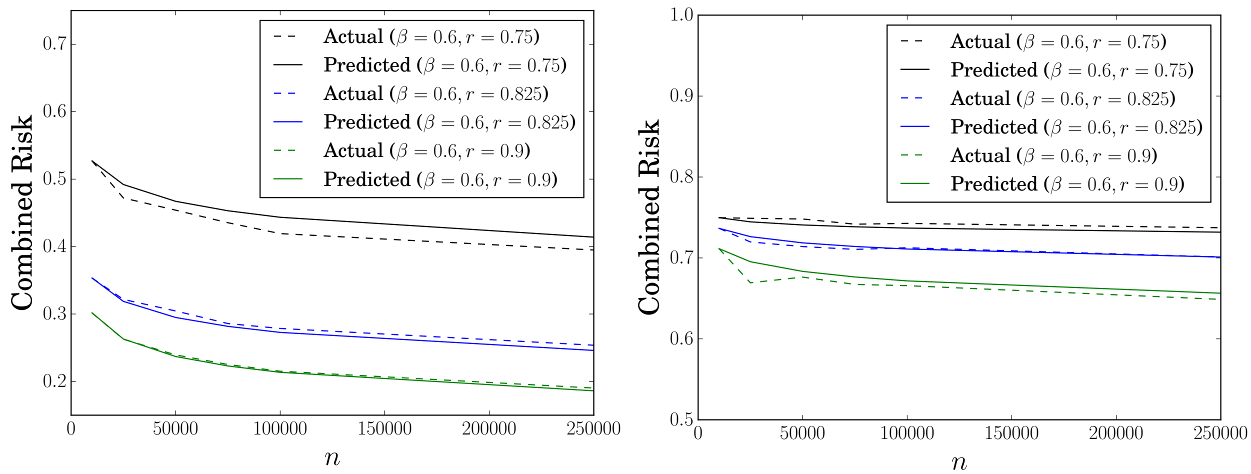
$$\mathrm{FDR}_n \leq q_n \quad \textit{and} \quad \mathrm{FNR}_n \leq \frac{2 \zeta_{\mathrm{BC}}^{2\beta_n^{\frac{1-\gamma}{\gamma}}}}{Z_u} \cdot n^{-D_\gamma(\beta_n + \kappa_n, r_n)} + q_n, \qquad \textit{where } \zeta_{\mathrm{BC}} := \tfrac{\zeta}{c_{\mathrm{BC}}}. \tag{26}$$

The proof of the theorem can be found in Section 4.3. For constant pairs (r, β), Theorem 2 can be applied with a target FDR proportional to $n^{-\kappa_*}$ to show that both BH and BC achieve the optimal decay of the combined FDR-FNR up to constant factors, as stated formally below.

Corollary 2. *For $\beta < r$ and $q_* = c_* n^{-\kappa_*}$ with $0 < c_* \leq \min\left\{\frac{1}{24}, \frac{1}{6 Z_\ell}\right\}$, the BH and BC procedures with target FDR q_* satisfy*

$$\mathcal{R}_n \lesssim n^{-\kappa_*}. \tag{27}$$

To help visualize the result of Corollary 2, Figure 2 displays the results of some simulations of the BH procedure that show correspondence between its performance and the theoretically predicted rate of $n^{-\kappa_*}$. Despite the optimality, Figures 1 and 2 paint a fairly dark picture from a practical point of view: while asymptotic consistency can be achieved when $r > \beta$, the convergence of the risk to zero can be extremely slow, exhibiting nonparametric rates far slower than $n^{-1/2}$. Figure 2



(a) $\gamma = 1$ (b) $\gamma = 2$

Figure 2. Results of simulations comparing the predicted combined risk with the actual experimentally-observed risk for the BH procedure. Agreement is good across the board and improves as the gap $(r - \beta)$ increases. We believe the latter phenomenon arises because the sampling error is a smaller fraction of the risk as the separation increases.

shows in particular that the decay to zero may be barely evident even for sample sizes as large as $n = 250,000$, even with comparatively strong signals.

The nonparametric nature may arise because the dimensionality of the decision space increases linearly with sample size, and asymptotically, the upside of having increasing data seems to *just* overcome the downside of having to make an increasing number of decisions. However, non-asymptotically, one cannot hope to drive both FDR and FNR to zero at any practical sample size in this general setting, at least when the mean signal lies below the maximum of the nulls (i.e., $r_n < 1$).

Regime of linear sparsity: We turn to the regime of *linear sparsity*—that is, when the number of signals scales as $\pi_1 n$ for some scalar $\pi_1 \in (0, 1)$. Recalling that we have parameterized the number of signals as $n^{1-\beta_n}$, some algebra leads to $\beta_n = \frac{\log \frac{1}{\pi_1}}{\log n}$, so both Theorem 1 and Theorem 2 predict an upper and lower bound on the risk of the form

$$c_0 \exp\left(c_1 \left[\frac{\log n}{\log \frac{1}{\pi_1}}\right]^{\frac{\gamma-1}{\gamma}}\right) \cdot n^{-\kappa_*}. \tag{28}$$

Note that here we overload the exponent κ_* to the case when it is nonconstant. In order to interpret this result, observe that if $r_n = r$ is constant, then $\kappa_* = \frac{r}{2\gamma} - o(1)$, so the rate is $n^{-r/2^\gamma}$ up to subpolynomial factors in n. On the other hand, if $r_n = \frac{1}{\log^{\frac{\gamma-1/2}{\gamma}} n}$ is at the extreme lower limit permitted by the lower bound (ii) in (14b), then it is not hard to see that $\kappa_* \approx \log^{-\frac{\gamma-1/2}{\gamma}} n$, which ensures that $n^{\kappa_*} \gg \exp\left(\log^{\frac{\gamma-1}{\gamma}} n\right)$, so that the risk (28) still approaches zero asymptotically, albeit subpolynomially in n.

4 Proofs

We now turn to the proofs of our main results, namely Theorems 1 and 2, along with Corollary 1.

4.1 Proof of Theorem 1

The main idea of the proof is to reduce the problem of lower bounding the FNR_n of threshold-based procedures that use random, data-dependent thresholds T_n, to the easier problem of lower bounding the FNR_n of threshold-based procedures that use a deterministic, data-independent threshold t_n. We refer to the latter class of procedures as *fixed threshold procedures*, and we parameterize them by their target FDR $q_n = n^{-\kappa_n}$. Concretely, we define the *critical threshold*, derived from the critical regime boundary $r_{\min}$ from equation (15), by

$$\tau_{\min}(\kappa_n) := \big(\gamma r_{\min}(\kappa_n) \log n\big)^{1/\gamma} \quad \equiv \quad \tau_{\min}(q_n) := \left(\gamma r_{\min}\left(\frac{\log(1/q_n)}{\log n}\right) \log n\right)^{1/\gamma}. \tag{29}$$

Here and throughout the proof, we express $\tau_{\min}$ and $r_{\min}$ as functions of q_n rather than κ_n; this formulation turns out to make certain calculations in the proof simpler to express.

From data-dependent threshold to fixed threshold: Our first step is to reduce the analysis from data-dependent to fixed threshold procedures. In particular, consider a threshold procedure, using a possibly random threshold T_n, that satisfies the FDR uppper bound $\mathrm{FDR}_n(T_n) \leq q_n$. We claim that the FNR of any such procedure must be lower bounded as

$$\mathbb{E}[\mathrm{FNP}_n(T_n)] \;\geq\; \frac{\mathrm{FNR}_n\big(\tau_{\min}(4q_n)\big)}{16}. \tag{30}$$

This lower bound is crucial, as it reduces the study of random threshold procedures (LHS) to study of fixed threshold procedures (RHS). In order to establish the claim (30), define the events

$$\mathcal{E}_1 := \big\{T_n \geq \tau_{\min}(4q_n)\big\}, \quad \text{and} \quad \mathcal{E}_2 := \left\{\mathrm{FNP}_n\big(\tau_{\min}(4q_n)\big) \geq \frac{\mathrm{FNR}_n\big(\tau_{\min}(4q_n)\big)}{2}\right\}.$$

The following lemma guarantees that both of these events have a non-vanishing probability:

Lemma 1. *For any threshold T_n such that $\mathrm{FDR}_n(T_n) \leq q_n$, we have*

$$\mathbb{P}[\mathcal{E}_1] \overset{(a)}{\geq} 3/8, \quad \textit{and} \quad \mathbb{P}[\mathcal{E}_2] \overset{(b)}{\geq} 3/4. \tag{31a}$$

The proof of this lemma can be found in Appendix A. Using this result, we now complete the proof of claim (30). Define the event

$$\mathcal{E} := \left\{\mathrm{FNP}_n(T_n) \geq \frac{\mathrm{FNR}_n\big(\tau_{\min}(4q_n)\big)}{2}\right\}.$$

The monotonicity of the function $t \mapsto \mathrm{FNP}_n(t)$ ensures that the inclusion $\mathcal{E} \supseteq \mathcal{E}_1 \cap \mathcal{E}_2$ must hold. Consequently, we have

$$\mathbb{P}[\mathcal{E}] \geq \mathbb{P}[\mathcal{E}_1 \cap \mathcal{E}_2] \;\geq\; \mathbb{P}[\mathcal{E}_2] - \mathbb{P}[\mathcal{E}_1^c] \overset{(i)}{\geq} \frac{3}{4} - \frac{5}{8} \;=\; 1/8,$$

where step (i) follows by applying the probability bounds from Lemma 1.

Finally, by Markov's inequality, we have

$$\mathrm{FNR}_n(T_n) \;=\; \mathbb{E}[\mathrm{FNP}_n(T_n)] \geq \mathbb{P}[\mathcal{E}]\, \frac{\mathrm{FNR}_n\big(\tau_{\min}(4q_n)\big)}{2} \;\geq\; \frac{\mathrm{FNR}_n\big(\tau_{\min}(4q_n)\big)}{16},$$

which establishes the claim (30).

Our next step is to lower bound the FNR for choices of the threshold $t \geq \tau_{\min}(q_n)$:

Lemma 2. *For any $t \geq \tau_{\min}(q_n)$, we have*

$$\mathrm{FNR}_n(t) \geq \begin{cases} \frac{\zeta^{2\beta_n^{\frac{1-\gamma}{\gamma}}}}{Z_\ell} \cdot n^{-D_\gamma(\beta_n+\kappa_n, r)} & \text{if } r > r_{\min}\big(\kappa_n(q_n)\big), \\ \frac{1}{2} & \text{otherwise,} \end{cases} \tag{32}$$

where ζ was previously defined (23).

The proof of this lemma can be found in Appendix B. Armed with Lemma 2 and the lower bound (30), we can now complete the proof of Theorem 1. We split the argument into two cases:

Case 1: First, suppose that $r \leq r_{\min}(\kappa_n(4q_n))$. In this case, we have

$$\mathrm{FNR}_n(T_n) \overset{(i)}{\geq} \frac{\mathrm{FNR}_n\big(\tau_{\min}(4q_n)\big)}{16} \overset{(ii)}{\geq} \frac{1}{32},$$

where step (i) follows from the lower bound (30), and step (ii) follows by lower bounding the FNR by 1/2, as is guaranteed by Lemma 2 in the regime $r \leq r_{\min}(\kappa_n(4q_n))$.

Case 2: Otherwise, we may assume that $r > r_{\min}(4q_n)$. In this case, we have

$$\mathrm{FNR}_n(T_n) \overset{(i)}{\geq} \frac{\mathrm{FNR}_n\big(\tau_{\min}(4q_n)\big)}{16} \overset{(ii)}{\geq} \frac{(4\zeta)^{2\beta_n^{\frac{1-\gamma}{\gamma}}}}{Z_\ell} \cdot n^{-D_\gamma(\beta_n+\kappa_n, r)}.$$

Here step (i) follows from the lower bound (30), whereas step (ii) follows from applying Lemma 2 in the regime $r > r_{\min}(\kappa_n(4q_n))$. With some further algebra, we find that

$$\mathrm{FNR}_n(T_n) \geq \frac{1}{Z_\ell} \exp\left(2\log\left(4\zeta\right) \cdot \beta_n^{\frac{1-\gamma}{\gamma}}\right) n^{-D_\gamma(\beta+\kappa_n, r)} = c_0 \exp\left(c_1 \beta_n^{\frac{1-\gamma}{\gamma}}\right) n^{-D_\gamma(\beta+\kappa_n, r)},$$

where $c_0 := \frac{1}{Z_\ell}$ and $c_1 := 2\log\big(4\zeta\big)$. Note that since $Z_\ell > 0$ and $\zeta \geq 1$, both of the constants c_0 and c_1 are positive, as claimed in the theorem statement.

4.2 Proof of Corollary 1

We now turn the proof of Corollary 1. Although it can be proved from the statement of Theorem 1, we instead prove it more directly, as this allows us to reuse parts of the proof of Lemma 2, thereby saving some additional messy calculations.

First, we verify that there is indeed a unique solution κ_* to the fixed point equation (20). Define the function as $g(\kappa) := D_\gamma\left(\beta + \kappa, r\right)^{1/\gamma} - \kappa^{1/\gamma}$. Clearly the solutions to (20) are the roots of g. We would like to argue that any such root must occur in $[0,\ r - \beta)$ and that in fact g has a unique root in this interval. For the first claim, note that $g(r - \beta) = -(r - \beta)^{1/\gamma} < 0$. On the other hand, we have

$$g'(\kappa) = \begin{cases} -\frac{1}{\gamma}\left[(\beta+\kappa)^{-\frac{\gamma-1}{\gamma}} + \kappa^{-\frac{\gamma-1}{\gamma}}\right] & \text{if } 0 \leq \kappa < r - \beta, \\ \frac{1}{\gamma}\left[(\beta+\kappa)^{-\frac{\gamma-1}{\gamma}} - \kappa^{-\frac{\gamma-1}{\gamma}}\right] & \text{if } \kappa > r - \beta. \end{cases}$$

It is immediately clear that $g'(\kappa) < 0$ for $0 \leq \kappa < r - \beta$ and, since $\beta + \kappa > \kappa$, we may also deduce that $g'(\kappa) < 0$ for $\kappa > r - \beta$, so g is decreasing on its domain. Therefore, $g(\kappa) < g(r - \beta) < 0$ for all $\kappa > r - \beta$.[3] We conclude that any root of g must occur on $[0,\ r-\beta)$. To finish the argument, note that $g(0) > 0 > g(r-\beta)$, so that g does indeed have a root on $[0,\ r-\beta)$.

Turning now to the proof of the lower bound (21), let $\mathcal{I}$ be an arbitrary threshold-based multiple testing procedure. We may assume without loss of generality that

$$\mathrm{FDR}_n(\mathcal{I}) \leq \min\big\{n^{-\kappa_*}, \frac{1}{24}\big\} \leq c(\beta,\gamma) n^{-\kappa_*}, \tag{33}$$

where the quantity $c(\beta,\gamma) \geq 1$ was defined in the statement of Theorem 1 (otherwise, the claimed lower bound (21) follows immediately).

Applying the second part of Lemma 2 and defining $\tilde{c} = 4c(\beta,\gamma)$, we conclude that

$$\begin{aligned}
\mathrm{FNR}_n(T_n) &\geq \frac{\mathrm{FNR}_n\big(\tau_{\min}(\tilde{c}n^{-\kappa_*})\big)}{16} \\
&\geq \frac{(\tilde{c}\zeta)^{2\beta^{\frac{1-\gamma}{\gamma}}}}{Z_\ell} \cdot n^{-D_\gamma(\beta+\kappa^*, r)} \\
&= \frac{(\tilde{c}\zeta)^{2\beta^{\frac{1-\gamma}{\gamma}}}}{Z_\ell} \cdot n^{-\kappa_*} \\
&= c' n^{-\kappa_*}.
\end{aligned}$$

4.3 Proof of Theorem 2

We now aim to show that the Benjamini-Hochberg (BH) and Barber-Candès (BC) algorithms achieve the minimax rate (19) when $r_n > r_{\min}(\kappa_n(c_{\mathrm{A}} q_n))$, where $\mathrm{A} \in \{\mathrm{BH},\ \mathrm{BC}\}$ and c_{A} is the algorithm-dependent constant defined in (23). The proof strategy for both algorithms is essentially the same. Given a target FDR rate q_n, we apply each algorithm q_n as the target FDR level and prove that the resulting threshold satisfies $t_{\mathrm{A}} \leq \tau_{\min}(c_{\mathrm{A}} q_n)$ with high probability. The known properties of the algorithms guarantee the required FDR bounds [as studied by the authors of 2, 13, 3], while the following converse to Lemma 2 provides the requisite upper bounds on the FNR.

Lemma 3. *If $r_n > r_{\min}(cq_n)$ and $t \leq \tau_{\min}(cq_n)$ for some $c > 0$, then we have*

$$\mathrm{FNR}_n(t) \leq \frac{\big(\max\{c,\ 1/c\}\cdot\zeta\big)^{2\beta_n^{\frac{1-\gamma}{\gamma}}}}{Z_u} \cdot n^{-D_\gamma(\beta_n+\kappa_n, r)},$$

where constant ζ is defined in (23).

4.3.1 Achievability result for the BH procedure

In this section, we prove that BH achieves the lower bound whenever $r_n > r_{\min}(c_{\mathrm{BH}} q_n)$. Specifically, we prove the claim (25) stated in Theorem 2.

As described in the previous section, the key step is to compare t_{BH}, the random threshold set by BH with target FDR of q_n, to the critical value $\tau_{\min,\mathrm{BH}} := \tau_{\min}(c_{\mathrm{BH}} q_n)$. To this end, we argue that

$$\mathbb{P}(t_{\mathrm{BH}} > \tau_{\min,\mathrm{BH}}) \leq \exp\big(-\frac{n^{1-r_{\max}}}{24}\big). \tag{34}$$

[3]Note that we have suppressed the issue of non-differentiability of g at $\kappa = r-\beta$. We may do so because it is left- and right-differentiable at this point, and we argue separately for the intervals $[0,\ r-\beta)$ and $[r-\beta,\ \infty)$.

We begin by showing how to derive the upper bound (25) from the probability bound (34). Note that since BH is a valid FDR control procedure, we necessarily have $\mathrm{FDR}_n(t_{\mathrm{BH}}) \le q_n$. To bound the FNR, first let $\mathcal{E} = \{t_{\mathrm{BH}} \le \tau_{\mathrm{min,BH}}\}$ and let $\mathrm{FNR}_n(\cdot \mid \mathcal{E})$ and $\mathrm{FNR}_n(\cdot \mid \mathcal{E}^c)$ denote the FNR_n conditional on the event and its complement, respectively. In this notation, the bound (34), together with Lemma 3, implies that

$$\begin{aligned}
\mathrm{FNR}_n(t_{\mathrm{BH}}) &\le \mathbb{P}\big(\mathcal{E}\big) \cdot \mathrm{FNR}_n\big(\tau_{\mathrm{min,BH}} \mid \mathcal{E}\big) + \mathbb{P}\big(\mathcal{E}^c\big) \\
&\le \mathrm{FNR}_n\big(\tau_{\mathrm{min,BH}}\big) + \mathbb{P}\big(\mathcal{E}^c\big) \\
&\le \mathrm{FNR}_n\big(\tau_{\mathrm{min,BH}}\big) + \exp\big(-\frac{n^{1-r_{\max}}}{24}\big) \\
&\le \frac{\zeta_{\mathrm{BH}}^{2\beta_n\frac{1-\gamma}{\gamma}}}{Z_u} \cdot n^{-D_\gamma(\beta_n+\kappa_n, r_n)} + \exp\big(-\frac{n^{1-r_{\max}}}{24}\big) \\
&\le \frac{2\zeta_{\mathrm{BH}}^{2\beta_n\frac{1-\gamma}{\gamma}}}{Z_u} \cdot n^{-D_\gamma(\beta_n+\kappa_n, r_n)},
\end{aligned}$$

where the final step uses the definition (22) of $n_{\min,u}$, and the fact that $\frac{1}{Z_u n} \le \frac{\zeta_{\mathrm{BH}}^{2\beta_n\frac{1-\gamma}{\gamma}}}{Z_u} \cdot n^{-D_\gamma(\beta_n+\kappa_n,r_n)}$, which is easily verified by noting that $\zeta_{\mathrm{BH}}^{2\beta_n\frac{1-\gamma}{\gamma}} \ge 1$ and $D_\gamma(\beta_n+\kappa_n, r_n) \le 1$.

We now prove the bound (34) with an argument using p-values and survival functions that parallels that of Arias-Castro and Chen [2] but sidesteps CDF asymptotics. We study the relationship between the population survival function Ψ and the empirical survival function $\hat{\Psi}$, defined by

$$\hat{\Psi}\big(t\big) = \big(1 - \frac{1}{n^{\beta_n}}\big) \cdot \hat{\Psi}_0\big(t\big) + \frac{1}{n^{\beta_n}} \cdot \hat{\Psi}_1\big(t\big), \tag{35}$$

$$\text{where} \qquad \hat{\Psi}_0\big(t\big) = \frac{1}{n - n^{1-\beta_n}} \sum_{i \in \mathcal{H}_0} \mathbf{1}\big(X_i \ge t\big) \quad \text{and} \quad \hat{\Psi}_1\big(t\big) = \frac{1}{n^{1-\beta_n}} \sum_{i \notin \mathcal{H}_0} \mathbf{1}\big(X_i \ge t\big).$$

Now, sort the observations in *decreasing* order, so that $X_{(1)} \ge X_{(2)} \ge \cdots \ge X_{(n)}$, and define p-values

$$p_{(i)} = \Psi\left(X_{(i)}\right) \quad \text{and} \quad \hat{\Psi}\left(X_{(i)}\right) = \frac{i}{n}, \tag{36}$$

so that $p_{(1)} \le p_{(2)} \le \cdots \le p_{(n)}$ are in *increasing* order. Then, we may characterize the indices rejected by BH as those satisfying $X_i \ge X_{(i_{\mathrm{BH}})}$, where

$$i_{\mathrm{BH}} = \max\big\{1 \le i \le n \colon \Psi\left(X_{(i)}\right) \le q_n \hat{\Psi}\left(X_{(i)}\right)\big\}. \tag{37}$$

Moving t_{BH} within $\big(X_{(i_{\mathrm{BH}}+1)},\ X_{(i_{\mathrm{BH}})}\big]$ if necessary, we may therefore assume $\Psi(t_{\mathrm{BH}}) = q_n\hat{\Psi}(t)$ whenever $t < t_{\mathrm{BH}}$, and combining this knowledge with (35), we obtain the chain of inclusions

$$\begin{aligned}
\mathcal{E}^c = \{t_{\mathrm{BH}} \le \tau_{\mathrm{min,BH}}\} &\supset \Big\{\Psi\big(\tau_{\mathrm{min,BH}}\big) \le q_n\hat{\Psi}\big(\tau_{\mathrm{min,BH}}\big)\Big\} \\
&\supset \Big\{\Psi\big(\tau_{\mathrm{min,BH}}\big) \le \frac{q_n}{n^{\beta_n}} \cdot \frac{W_n}{n^{1-\beta_n}}\Big\} =: \widetilde{\mathcal{E}},
\end{aligned} \tag{38}$$

where $W_n = \sum_{i \notin \mathcal{H}_0} \mathbf{1}\big(X_i \ge \tau_{\mathrm{min,BH}}\big) \sim \mathrm{Bin}\big(\Psi\big(\tau_{\mathrm{min,BH}} - \mu_n\big),\ n^{1-\beta_n}\big)$.

We now argue that $\Psi\big(\tau_{\min,\mathrm{BH}}\big) \le \frac{q_n}{4n^{\beta_n}}$, so that $\mathbb{P}(\mathcal{E}) \ge \mathbb{P}\left(W_n > \frac{n^{1-\beta_n}}{4}\right)$. For this, observe that by the definition of $r_{\min}$ in (15) and the upper tail bound (5), we have

$$\begin{aligned}
\log \Psi\big(\tau_{\min,\mathrm{BH}}\big) &\le -r_{\min}\left(c_{\mathrm{BH}} q_n\right)\log n + \log\frac{1}{Z_u} \\
&\le -\beta_n \log n + \log(c_{\mathrm{BH}} q_n) - \log\frac{1}{6Z_\ell} + \log\frac{1}{Z_u} \\
&= -\beta_n \log n + \log q_n + \log\frac{6c_{\mathrm{BH}} Z_\ell}{Z_u} \\
&= \log\frac{q_n}{6n^{\beta_n}} \;<\; \log\frac{q_n}{4n^{\beta_n}}.
\end{aligned}$$

We conclude

$$\mathbb{P}\left(t_{\mathrm{BH}} > \tau_{\min,\mathrm{BH}}\right) \;\le\; 1 - \mathbb{P}\left(\widetilde{\mathcal{E}}\right) \;\le\; 1 - \mathbb{P}\left(W_n > \frac{n^{1-\beta_n}}{4}\right) \;=\; \mathbb{P}\left(W_n \le \frac{n^{1-\beta_n}}{4}\right).$$

Finally, by a Bernstein bound, we find

$$\begin{aligned}
\mathbb{P}\left(W_n \le \frac{n^{1-\beta_n}}{4}\right) &\le \mathbb{P}\left(W_n \le \frac{\mathbb{E}\left[W_n\right]}{2}\right) \\
&\le \exp\Big(-\frac{\mathbb{E}\left[W_n\right]}{12}\Big) \\
&\le \exp\Big(-\frac{n^{1-\beta_n}}{24}\Big) \\
&\le \exp\Big(-\frac{n^{1-r_{\max}}}{24}\Big),
\end{aligned}$$

where we have used the fact that $\tau_{\min,\mathrm{BH}} \le \mu_n$ to conclude that $\Psi\left(\tau_{\min,\mathrm{BH}} - \mu_n\right) \ge \frac{1}{2}$ and therefore $\mathbb{E}[W_n] \ge \frac{n^{1-\beta_n}}{2}$. We have therefore established the required claim (34), concluding the proof of optimality of the BH procedure.

4.3.2 Achievability result for the BC procedure

Our overall strategy for analyzing BC procedure resembles the one we used for the BH procedure. As with our analysis of the BH procedure, we define $\tau_{\min,\mathrm{BC}} := \tau_{\min}(c_{\mathrm{BC}} q_n)$ and derive the bound (25) by controlling the algorithm's threshold as

$$\mathbb{P}\left(t_{\mathrm{BC}} > \tau_{\min,\mathrm{BC}}\right) \le q_n + \exp\Big(-\frac{n^{1-r_{\max}}}{24}\Big). \tag{39}$$

Since the proof of equation (26) from the bound (39) is essentially identical to the corresponding derivation for the BH procedure, we omit it. We now prove the bound (39) by an argument somewhat different than that used in analyzing the BH procedure. Define the integers

$$N_+(t) = \sum_{i=1}^{n} \mathbf{1}\big(X_i \ge t\big) \quad\text{and}\quad N_-(t) = \sum_{i=1}^{n} \mathbf{1}\big(X_i \le -t\big).$$

Then, the definition of the BC procedure gives

$$t_{\mathrm{BC}} = \inf\Big\{t \in \mathbb{R} \colon \frac{1 + N_-(t)}{1 \vee N_+(t)} \le q_n\Big\}.$$

To prove (39), it therefore suffices to show that

$$\mathbb{P}\left(\frac{1+N_-(\tau_{\min,\mathrm{BC}})}{1\vee N_+(\tau_{\min,\mathrm{BC}})} > q_n\right) \le q_n + \exp\big(-\frac{n^{1-r_{\max}}}{24}\big). \tag{40}$$

We prove the bound (40) in two parts:

$$\mathbb{P}\big(1\vee N_+(\tau_{\min,\mathrm{BC}}) < \frac{n^{1-\beta_n}}{4}\big) \le \exp\big(-\frac{n^{1-r_{\max}}}{24}\big), \tag{41a}$$

$$\mathbb{P}\left(1+N_-(\tau_{\min,\mathrm{BC}}) > q_n\cdot\frac{n^{1-\beta_n}}{4}\right) \le q_n. \tag{41b}$$

These bounds are a straightforward consequence of elementary Bernstein bounds, and together they imply the claim (40). We explain them below.

The lower bound (41a) follows because $1\vee N_+(\tau_{\min,\mathrm{BC}}) \ge N_+(\tau_{\min,\mathrm{BC}})$ and $N_+(\tau_{\min,\mathrm{BC}})$ is the sum of two binomial random variables, corresponding to nulls and signals, respectively, and the latter has a $\Psi\big(\tau_{\min,\mathrm{BC}}-\mu\big) \ge \frac{1}{2}$ probability of success. More precisely, we may write $N_+\left(\tau_{\min,\mathrm{BC}}\right) = N_+^{\mathrm{null}} + N_+^{\mathrm{signal}}$, with

$$N_+^{\mathrm{null}} \sim \mathrm{Bin}\left(\Psi\left(\tau_{\min,\mathrm{BC}}\right),\ n-n^{1-\beta_n}\right) \quad\text{and}\quad N_+^{\mathrm{signal}} \sim \mathrm{Bin}\left(\Psi\left(\tau_{\min,\mathrm{BC}}-\mu_n\right),\ n^{1-\beta_n}\right),$$

implying $N_+\left(\tau_{\min,\mathrm{BC}}\right) \ge N_+^{\mathrm{signal}}$, whence

$$\mathbb{E}\big[N_+(\tau_{\min,\mathrm{BC}})\big] \ge \mathbb{E}\big[N_+^{\mathrm{signal}}\big] = n^{1-\beta_n}\cdot\Psi\big(\tau_{\min,\mathrm{BC}}-\mu_n\big) \ge \frac{n^{1-\beta_n}}{2},$$

where we have used the fact that $\tau_{\min,\mathrm{BC}} \le \mu_n$. With this bound in hand, a Bernstein bound yields

$$\begin{aligned}\mathbb{P}\big(N_+(\tau_{\min,\mathrm{BC}}) < \frac{n^{1-\beta_n}}{4}\big) &\le \mathbb{P}\big(N_+(\tau_{\min,\mathrm{BC}}) \le \frac{\mathbb{E}\left[N_+(\tau_{\min,\mathrm{BC}})\right]}{2}\big)\\ &\le \exp\big(-\frac{n^{1-\beta_n}}{24}\big) \le \exp\big(-\frac{n^{1-r_{\max}}}{24}\big),\end{aligned}$$

as required to prove equation (41a). The proof of equation (41b) follows a similar pattern. Here, we note that $N_-(\tau_{\min,\mathrm{BC}})$ is a sum of two binomial random variables, with a total of n trials, such that—using the definition (15) of $r_{\min}$ and the upper bound on the tail (5)—each one has probability of success upper bounded by $1-\Psi\big(-\tau_{\min,\mathrm{BC}}\big) \le \frac{6Z_\ell}{Z_u}\cdot c_{\mathrm{BC}}q_n n^{-\beta_n} = \frac{1}{8}\cdot q_n n^{-\beta_n}$. Formally, we may write $N_-(\tau_{\min,\mathrm{BC}}) = N_-^{\mathrm{null}} + N_-^{\mathrm{signal}}$, with

$$N_-^{\mathrm{null}} \sim \mathrm{Bin}\left(1-\Psi\left(-\tau_{\min,\mathrm{BC}}\right),\ n-n^{1-\beta_n}\right) \quad\text{and}\quad N_-^{\mathrm{signal}} \sim \mathrm{Bin}\left(1-\Psi\left(-\tau_{\min,\mathrm{BC}}-\mu\right),\ n^{1-\beta_n}\right).$$

Since $1-\Psi\left(-\tau_{\min,\mathrm{BC}}-\mu\right) \le 1-\Psi\left(-\tau_{\min,\mathrm{BC}}\right)$, we deduce

$$\mathbb{E}\big[N_-(\tau_{\min,\mathrm{BC}})\big] \le \big[1-\Psi\left(-\tau_{\min,\mathrm{BC}}\right)\big]\cdot n \le \frac{q_n}{2}\cdot\frac{n^{1-\beta_n}}{4}.$$

On the other hand, using the lower bound in (5), we find $1-\Psi\big(-\tau_{\min,\mathrm{BC}}\big) \ge 6c_{\mathrm{BC}}q_n n^{-\beta_n}$. Using the additional fact that $n-n^{1-\beta_n} \ge \frac{n}{2}$ by (14a), we may conclude that

$$\begin{aligned}\mathbb{E}\big[N_-(\tau_{\min,\mathrm{BC}})\big] &\ge \mathbb{E}\big[N_-^{\mathrm{null}}\big]\\ &= \big(n-n^{1-\beta_n}\big)\cdot\big[1-\Psi\big(-\tau_{\min,\mathrm{BC}}\big)\big]\\ &\ge \frac{n}{2}\cdot 6c_{\mathrm{BC}}q_n n^{-\beta_n}\\ &\ge 3c_{\mathrm{BC}}q_n n^{1-\beta_n}.\end{aligned}$$

By a Bernstein bound, it follows that

$$\begin{aligned}\mathbb{P}\Bigg(N_{-}(\tau_{\min,\mathrm{BC}}) \geq q_n \cdot \frac{n^{1-\beta_n}}{4}\Bigg) &\leq \mathbb{P}\Bigg(N_{-}(\tau_{\min,\mathrm{BC}}) \geq 2\mathbb{E}\big[N_{-}(\tau_{\min,\mathrm{BC}})\big]\Bigg)\\ &\leq \exp\Bigg(-\frac{\mathbb{E}\big[N_{-}(\tau_{\min,\mathrm{BC}})\big]}{4}\Bigg)\\ &\leq \exp\big(-\frac{3c_{\mathrm{BC}}}{4}\cdot q_n n^{1-\beta_n}\big)\\ &\leq q_n,\end{aligned}$$

where we have invoked the decay condition (24) for the last step.

4.4 Proof of Corollary 2

The corollary is a nearly immediate consequence of Theorem 2. We will prove it for both algorithms simultaneously. Observe that

$$r_{\min}(\kappa_n(c_{\mathrm{A}}q_*)) = \beta + \kappa_* + \frac{\log \frac{1}{6c_* c_{\mathrm{A}} Z_\ell}}{\log n}. \tag{42}$$

Suppose for now that the decay condition (24) holds for q_* and some choice of $n_{\min,\mathrm{BC}}$. Then, using (42) and the fact that $r > \beta + \kappa_*$, we may choose $n'_{\min} \geq n_{\min,\mathrm{BC}}$ large enough so that $r > r_{\min}(\kappa_n(c_{\mathrm{A}}q_*))$ for all $n \geq n'_{\min}$ and $\mathrm{A} \in \{\mathrm{BH},\ \mathrm{BC}\}$. From Theorem 2, we conclude that there exists a constant c' such that both algorithms satisfy

$$n \geq n'_{\min} \Longrightarrow \mathcal{R}_n \leq c' n^{-\kappa_*}.$$

By replacing c' by $\tilde{c} = \max\{c',\ (n'_{\min})^{\kappa_*}\}$ (and recalling $\mathcal{R}_n \leq 1$ always), we obtain $\mathcal{R}_n \leq \tilde{c} n^{-\kappa_*}$ for all $n \geq 1$, obtaining the claimed result.

In order to check the decay condition (24), note that, as $\kappa_* \leq r - \beta \leq 1 - \beta$, we have for sufficiently large n that

$$\frac{q_n}{\log \frac{1}{q_n}} = \frac{n^{-\kappa_*}}{\kappa_* \log n} \geq \frac{4}{3c_{\mathrm{BC}}} \cdot n^{-(1-\beta)},$$

which completes the proof.

5 Discussion

Despite considerable interest in multiple testing with false discovery rate (FDR) control, there has been relatively little understanding of the non-asymptotic trade-off between controlling FDR and the analogous measure of power known as the false non-discovery rate (FNR). In this paper, we explored this issue in the context of the sparse generalized Gaussians model, and derived the first non-asymptotic lower bounds on the sum of FDR and FNR. We complemented these lower bounds by establishing the non-asymptotic minimaxity of both the Benjamini-Hochberg (BH) and Barber-Candès (BC) procedures for FDR control. The theoretical predictions are validated in simple simulations, and our results recover recent asymptotic results [2] as special cases. Our work introduces a simple proof strategy based on reduction to deterministic and data-oblivious procedures. We suspect this core idea may apply to other multiple testing settings: in particular, since our arguments do not depend on CDF asymptotics in the way that many classical analyses of both global null testing and FDR control procedures do, we hope they will be possible to adapt for other problems described below.

As mentioned after the statement of Theorem 2, the practical implications of our results are somewhat pessimistic. Even for rather simple problems having $r - \beta$ of constant order, the resulting rate at which the risk tends to zero can be far slower than $n^{-1/2}$. (Indeed, it seems like such a parametric rate is only achievable when $\gamma = 1, r_n \to 1, \beta_n \to 0$.) Hence, in practice, one must carefully consider whether good FDR or good FNR is more important, as achieving both may not be possible unless most of the signals to be identified are rather large.

Future work

A large part of the multiple testing literature focuses on the development of valid FDR control procedures that can gain power or precision by explicitly using prior knowledge such as null-proportion adaptivity [25, 26], groups or partitions of hypotheses [14, 19], prior or penalty weights [4, 16], or other forms of structure [22, 24], and it would be of great interest to extend our techniques to such structured settings. It is also important to handle the cases of positive or arbitrary dependence [5, 6, 24], hence this is another natural direction in which to extend our work. (Such extensions were explored by [17, 21, 18] for the higher criticism statistic for the detection problem.) Lastly, all of the discussed literature is in the offline setting, and it could be of interest to develop tight lower and upper bounds for online FDR procedures [12, 20].

A general proof technique for establishing non-asymptotic lower bounds in multiple testing remains an important direction for future work. As our arguments are based on analytical calculations, they are sensitive to the particular observation model under consideration. It would be desirable to build on our approach to identify the key properties of multiple testing problems that make them difficult or easy, and establish lower bounds in terms of these properties. We hope our work will help lay the foundation for progress in understanding the limits of multiple testing more generally.

Acknowledgements

This work was partially supported by Office of Naval Research Grant DOD-ONR-N00014, Air Force Office of Scientific Research Grant AFOSR-FA9550-14-1-0016, and Office of Naval Research grant W911NF-16-1-0368NSF. In addition, MR was supported by an NSF Graduate Research Fellowship and a Fannie and John Hertz Foundation Google Fellowship.

A Proof of Lemma 1

This appendix is devoted to the proof of Lemma 1, which was involved in the proof of Theorem 1. Our proof makes use of the following auxiliary lemma:

Lemma 4. *For $q_n \in (0, 1/24)$, we have*

$$\mathbb{P}\Big[\mathrm{FDP}_n(t) \geq 8q_n \quad \textit{for all } t \in [0, \tau_{\min}(4q_n)]\Big] \geq \frac{1}{2}. \tag{43}$$

We return to prove this claim in Appendix A.1. For the moment, we take it as given and complete the proof of Lemma 1.

Control of $\mathcal{E}_1$: Let us now prove the first bound in Lemma 1, namely that $\mathbb{P}[\mathcal{E}_1] \geq \frac{3}{8}$ where $\mathcal{E}_1 := \{T_n \geq \tau_{\min}(4q_n)\}$. So as to simplify notation, let us define the event

$$\mathcal{D} := \Big\{ \text{FDP}_n(t) \geq 8q_n \quad \text{for all } t \in [0, \tau_{\min}(4q_n)] \Big\}. \tag{44a}$$

Now observe that

$$\mathbb{P}\Big[T_n \geq \tau_{\min}(4q_n)\Big] \;\geq\; \mathbb{P}\Big[T_n \geq \tau_{\min}(4q_n) \text{ and } \mathcal{D}\Big] \;=\; \mathbb{P}[\mathcal{D}] - \mathbb{P}\big[T_n \leq \tau_{\min}(4q_n) \text{ and } \mathcal{D}\big]. \tag{44b}$$

Now by the definition (44a) of the event $\mathcal{D}$, we have the inclusion

$$\Big\{T_n \leq \tau_{\min}(4q_n) \text{ and } \mathcal{D}\Big\} \subseteq \Big\{\text{FDP}_n(T_n) \geq 8q_n\Big\}.$$

Combining with our earlier bound (44b), we see that

$$\mathbb{P}\Big[T_n \geq \tau_{\min}(4q_n)\Big] \geq \mathbb{P}[\mathcal{D}] - \mathbb{P}\big[\text{FDP}_n(T_n) \geq 8q_n\big].$$

It remains to control the two probabilities on the right-hand side of this bound. Applying Lemma 4 guarantees that $\mathbb{P}[\mathcal{D}] \geq \frac{1}{2}$. On the other hand, by Markov's inequality, the assumed lower bound $\text{FDR}_n(T_n) \leq q_n$ implies that $\mathbb{P}\Big[\text{FDP}_n(T_n) \geq 8q_n\Big] \leq \frac{1}{8}$. Putting together the pieces, we conclude that

$$\mathbb{P}[\mathcal{E}_1] \;=\; \mathbb{P}\Big[T_n \geq \tau_{\min}(4q_n)\Big] \;\geq\; \frac{1}{2} - \frac{1}{8} \;=\; \frac{3}{8},$$

as claimed.

Control of $\mathcal{E}_2$: Let us now prove the lower bound $\mathbb{P}[\mathcal{E}_2] \geq 3/4$. We split our analysis into two cases.

Case 1: First, suppose that $r_n > r_{\min}$. In this case, we can write

$$\text{FNP}_n(t) = \frac{F_n(t)}{n^{1-\beta_n}}, \qquad \text{where } F_n(t) \sim \text{Bin}\big(1 - \Psi(t-\mu),\; n^{1-\beta_n}\big).$$

Since $r_{\min} > \beta_n$, we have

$$\mu - \tau_{\min} \geq \big(\gamma \log n\big)^{1/\gamma} \big[r_n^{1/\gamma} - \beta_n^{1/\gamma}\big] \geq \big(\gamma \log n\big)^{1/\gamma} \cdot \big(r_n - \beta_n\big)^{1/\gamma},$$

from which it follows that

$$\mathbb{E}[F_n] \;=\; 1 - \Psi\big(t - \mu\big) \geq \frac{n^{\beta_n - r_n}}{Z_\ell}. \tag{45}$$

Now by applying the Bernstein bound to the binomial random variable F_n, we have

$$\begin{aligned} 1 - \mathbb{P}[\mathcal{E}_2] = \mathbb{P}\left[F_n \leq \frac{\mathbb{E}[F_n]}{2}\right] &\leq \exp\Big(-\frac{\mathbb{E}[F_n]}{12}\Big) \\ &\stackrel{(i)}{\leq} \exp\Big(-\frac{n^{1-r_n}}{12 Z_\ell}\Big) \\ &\stackrel{(ii)}{\leq} \exp\Big(-\frac{n^{1-r_{\max}}}{12 Z_\ell}\Big), \end{aligned} \tag{46}$$

where step (i) follows from the lower bound (45), and step (ii) follows since $r_n < r_{\max}$ by assumption.

Case 2: Otherwise, we may assume that $r_n \in \big(\beta_n, r_n^{-1}\big)$. In this regime, we have the lower bound $\tau_{\min} - \mu \geq 0$, so that the binomial random variable F_n stochastically dominates a second binomial distributed as $\widetilde{F}_n \sim \text{Bin}\big(\frac{1}{2},\ n^{1-\beta_n}\big)$. By this stochastic domination condition, it follows that

$$1 - \mathbb{P}\big[\mathcal{E}_2\big] \ \leq\ \mathbb{P}\Big[F_n \leq \frac{n^{1-\beta_n}}{4}\Big] \leq \mathbb{P}\Big[\widetilde{F}_n \leq \frac{\mathbb{E}\big[\widetilde{F}_n\big]}{2}\Big].$$

By applying the Bernstein bound to $\widetilde{F}_n$, we find that

$$1 - \mathbb{P}\big[\mathcal{E}_2\big] \leq \exp\big(-\frac{n^{1-\beta_n}}{24}\big)\ \leq \exp\big(-\frac{n^{1-r_{\max}}}{24}\big), \tag{47}$$

where the final step follows since $r_{\max} > \beta_n$.

Putting together the two bounds (46) and (47), we conclude that $\mathbb{P}[\mathcal{E}_2] \geq \frac{3}{4}$ for all sample sizes n large enough to ensure that

$$\max\big\{\exp\big(-\frac{n^{1-r_{\max}}}{24 Z_\ell}\big),\ \exp\big(-\frac{n^{1-r_{\max}}}{24}\big)\big\} \leq \frac{1}{4}, \tag{48}$$

as was claimed. Note that condition (48) is identical to condition (17), so that our definition of $n_{\min}$ guarantees that (48) is satisfied. This completes the proof.

A.1 Proof of Lemma 4

It remains to prove our auxiliary result stated in Lemma 4. For notational economy, let $\tau = \tau_{\min}(s)$ and let $\beta = \beta_n$. The FDP at a threshold t can be expressed in terms of two binomial random variables

$$L_n(t) = \sum_{i \in \mathcal{H}_0} \mathbf{1}\big(X_i \geq t\big) \quad \text{and} \quad W_n(t) = \sum_{i \notin \mathcal{H}_0} \mathbf{1}\big(X_i \geq t\big) \leq n^{1-\beta}.$$

Here $L_n(t)$ and $W_n(t)$ correspond (respectively) to the number of nulls, and the number of signals that exceed the threshold t. In terms of these two binomial random variables, we have the expression

$$\text{FDP}_n(t) = \frac{L_n(t)}{L_n(t) + W_n(t)} \geq \frac{L_n(t)}{L_n(t) + n^{1-\beta}}.$$

Note that the inequality here follows by replacing $W_n(t)$ by the potentially very loose upper bound $n^{1-\beta}$; doing so allows us to reduce the problem of bounding the FDP to control of $L_n(t)$ uniformly for $t \in [0, \tau]$. By definition of $L_n(t)$, we have the lower bound

$$\frac{L_n(t)}{L_n(t) + n^{1-\beta}} \geq \frac{L_n(\tau)}{L_n(\tau) + n^{1-\beta}} \qquad \text{for all } t \in [0, \tau].$$

Moreover, observe that

$$\frac{3s}{1+3s} \geq \frac{12}{5} s \geq 2s \qquad \text{for all } s \in (0, 1/6).$$

Combining these bounds, we find that

$$\begin{aligned}
\mathbb{P}\Big[\text{FDP}_n(t) \geq 2s \quad \text{for all } t \in [0, \tau]\Big] &\geq \mathbb{P}\Big[\frac{L_n(\tau)}{L_n(\tau) + n^{1-\beta}} \geq \frac{3s}{1+3s}\Big] \\
&= \mathbb{P}\big[L_n(\tau) \geq 3s n^{1-\beta}\big].
\end{aligned}$$

Consequently, the remainder of our proof is devoted to proving that

$$\mathbb{P}\big[L_n(\tau) \geq 3sn^{1-\beta}\big] \geq 1/2. \tag{49}$$

We split our analysis into two cases:

Case 1: First, suppose that $q_n \geq \frac{2\log 4}{3n^{1-\beta}}$ In this case, we have

$$\alpha :\ = \Psi(\tau) \geq \frac{6s}{n^{\beta}} > \frac{16\log 4}{n}. \tag{50}$$

A simple calculation based on this inequality yields

$$\alpha n - 3sn^{1-\beta} \geq \frac{\alpha n}{2} \geq \sqrt{(4\log 4)\alpha(1-\alpha)n} := a\sigma, \tag{51}$$

where $a = \sqrt{4\log 4}$ and $\sigma = \sqrt{\alpha(1-\alpha)n}$. Notice that $\sigma^2 = \mathrm{Var}\,[L_n(\tau)]$.

We now apply the Bernstein inequality to $L_n(\tau)$ to obtain

$$\begin{aligned}
\mathbb{P}\big(L_n \leq 3sn^{1-\beta}\big) &\leq \mathbb{P}\big(L_n \leq \pi_1 n - a\sigma\big)\\
&\leq 2\cdot\exp\big(-\frac{a^2\sigma^2}{2[\sigma^2 + a\sigma]}\big)\\
&\leq 2\cdot\exp\big(-\frac{a^2}{2(1+\frac{a}{\sigma})}\big).\\
&\leq \exp\big(-\frac{a^2}{4}\big) = \frac{1}{4},
\end{aligned}$$

where we have used the fact that $a < \sigma$. We conclude that

$$\mathbb{P}\big(L_n \geq 3sn^{1-\beta}\big) \geq \frac{1}{2},$$

as desired.

Case 2: Otherwise, we may assume that $q_n < \frac{2\log 4}{3n^{1-\beta}}$. The definition of τ implies that

$$\alpha \geq \frac{24}{n} \quad \text{and} \quad 3sn^{1-\beta} \leq 8\log 4.$$

It follows that $\mathbb{E}[L_n(\tau)] \geq 24$. On the other hand, given that $8\log 4 < 12$, it suffices to prove that

$$P\big[L_n(\tau) \leq 12\big] \leq \frac{1}{2}.$$

This is straightforward, however, since Bernstein's inequality gives

$$\begin{aligned}
\mathbb{P}\big[L_n(\tau) \leq 12\big] \;=\; \mathbb{P}\big[L_n(\tau) \leq \frac{\mathbb{E}[L_n(\tau)]}{2}\big] &\leq \exp\big(-\frac{24}{12}\big)\\
&= e^{-2}\\
&< \frac{1}{2},
\end{aligned}$$

which completes the proof.

B Proof of Lemmas 2 and 3

This appendix is devoted to the proofs of Lemmas 2 and 3. We combine the proofs, since these two lemmas provide lower and upper bounds, respectively, on the because they are matching lower and upper bounds, respectively, on the FNR for a fixed threshold procedure, and their proofs involve extremely similar calculations.

So as to simplify notation, we make use of the convenient shorthands let $\tau = \tau_{\min}(q_n)$, $\beta = \beta_n$, and $\mu = \mu_n$ throughout the proof. Recall that the FNP can be written as the ratio $\mathrm{FNP}_n(t) = \frac{F_n(t)}{n^{1-\beta}}$, where

$$F_n(t) = \sum_{i \notin \mathcal{H}_0} \mathbf{1}\big(X_i \leq t\big) \sim \mathrm{Bin}\Big(1 - \Psi\big(t - \mu\big),\ n^{1-\beta}\Big) \tag{52}$$

is a binomial random variable. We split the remainder of the analysis into two cases.

Case 1: First, suppose that $\tau \geq \mu$. In this case, we only seek to prove a lower bound. For this, observe that $\Psi\left(\tau - \mu\right) \leq \Psi\left(0\right) = \frac{1}{2}$, so $1 - \Psi\left(\tau - \mu\right) \geq \frac{1}{2}$. Thus,

$$\mathrm{FDR}_n\left(\tau\right) = \frac{\mathbb{E}\left[F_n\right]}{n^{1-\beta}} = 1 - \Psi\left(\tau - \mu\right) \geq \frac{1}{2},$$

as claimed.

Case 2: Otherwise, we may assume that $\mu > \tau$. Recall the parameterization (13) of μ in terms of r, the definition (15) of $r_{\min}$, and the definition (18) of the D_γ distance. In terms of these quantities, we have

$$\begin{aligned}
\mu - \tau &= \big(\gamma \log n\big)^{1/\gamma} \left\{r^{1/\gamma} - r_{\min}(\kappa_n)\right\}^{1/\gamma} \\
&= \Big\{\gamma D_\gamma\big(r_{\min}(\kappa_n), r\big)\ \log n\Big\}^{1/\gamma} \\
&= \Big[\gamma D_\gamma\Big(\beta + \kappa_n + \frac{\log \frac{1}{6Z_\ell}}{\log n}, r\Big)\ \log n\Big]^{1/\gamma},
\end{aligned}$$

which shows how the quantity D_γ determines the rate. In order to complete the proof, we need to show that the additional order of $\frac{1}{\log n}$ term inside D_γ can be removed.

More precisely, it suffices to establish the sandwich relation

$$\frac{\zeta^{2\beta^{\frac{1-\gamma}{\gamma}}}}{Z_u} \cdot n^{-D_\gamma(\beta+\kappa_n, r)} \geq 1 - \Psi\big(\tau - \mu\big) \geq \frac{\zeta^{2\beta^{\frac{1-\gamma}{\gamma}}}}{Z_\ell} \cdot n^{-D_\gamma(\beta+\kappa_n, r)},$$

where $\zeta = \max\left\{6Z_\ell,\ \frac{1}{6Z_\ell}\right\}$ as in (23). But now note that

$$\tau - \mu = \big(\gamma \log n\big)^{1/\gamma}\big[\big(r_{\min}\big)^{1/\gamma} - r\big] = -\big[\gamma D_\gamma\left(r_{\min}, r\right) \log n\big]^{1/\gamma},$$

allowing us to deduce that

$$\frac{1}{Z_u} \cdot n^{-D_\gamma(r_{\min}, r)} \geq 1 - \Psi\big(\tau - \mu\big) \geq \frac{1}{Z_\ell} \cdot n^{D_\gamma(r_{\min}, r)},$$

so we need only show

$$\left|D_\gamma\left(\beta+\kappa_n, r\right)-D_\gamma\left(r_{\min}, r\right)\right| \leq \frac{\beta^{\frac{1-\gamma}{\gamma}} \log \zeta}{\log n}.$$

To prove this, we let

$$\tilde{r} \ :\ = \min\left(\beta+\kappa_n,\ r_{\min}\right)$$

and note that by (16), we must have $\tilde{r} \in [\beta,\ r]$. Under this definition, we consider the function $f(x)=D_\gamma\left(\tilde{r}+x, r\right)$. A simple calculation shows that for $x \geq 0$, we have

$$f'(x)=\begin{cases}-\left(\tilde{r}+x\right)^{\frac{1-\gamma}{\gamma}} D_\gamma\left(\tilde{r}+x, r\right)^{\frac{\gamma-1}{\gamma}} & \text{if } \tilde{r}+x \leq r, \\ \left(\tilde{r}+x\right)^{\frac{1-\gamma}{\gamma}} D_\gamma\left(\tilde{r}+x, r\right)^{\frac{\gamma-1}{\gamma}} & \text{o.w.}\end{cases}$$

We observe that we only need to allow $0 \leq x \leq \max\left(\beta+\kappa_n,\ r_{\min}\right)-\tilde{r}=: \tilde{R}-\tilde{r}$, so in particular, we will always have $\tilde{r}+x \leq \tilde{R} \leq 2$. This, together with the lower bound $\tilde{r} \geq \beta$, yields

$$\sup_{0 \leq x \leq \tilde{R}-\tilde{r}}\left|f'(x)\right| \leq 2 \beta^{\frac{1-\gamma}{\gamma}}.$$

Applying this result, we find

$$\begin{aligned}\left|D_\gamma\left(\beta+\kappa_n, r\right)-D_\gamma\left(r_{\min}, r\right)\right| &= \left|D_\gamma\left(\tilde{R}, r\right)-D_\gamma\left(\tilde{r}, r\right)\right| \\ &\leq 2 \beta^{\frac{1-\gamma}{\gamma}} \cdot\left(\tilde{R}-\tilde{r}\right) \\ &= 2 \beta^{\frac{1-\gamma}{\gamma}} \cdot \frac{\log \zeta}{\log n}.\end{aligned}$$

If we now consider $q_n'=c q_n$, we can recover the more refined statements in Lemmas 2 and 3, simply by noting that the same reasoning as above shows

$$\left|D_\gamma\left(\beta+\kappa_n, r\right)-D_\gamma\left(\beta+\kappa_n', r\right)\right| \leq 2 \beta^{\frac{1-\gamma}{\gamma}} \cdot \frac{|\log c|}{\log n},$$

concluding the argument.

References

[1] F. Abramovich, Y. Benjamini, D. Donoho, and I.M. Johnstone. Adapting to unknown sparsity by controlling the false discovery rate. *Annals of Statistics*, 34(2):584–653, 2006.

[2] E. Arias-Castro and S. Chen. Distribution-free multiple testing. *arXiv preprint arXiv:1604.07520*, 2016.

[3] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 57(1):289–300, 1995.

[4] Y. Benjamini and Y. Hochberg. Multiple hypotheses testing with weights. *Scandinavian Journal of Statistics*, 24(3):407–418, 1997.

[5] Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, 29(4):1165–1188, 2001.

[6] G. Blanchard and E. Roquain. Two simple sufficient conditions for FDR control. *Electronic Journal of Statistics*, 2:963–992, 2008.

[7] M. Bogdan, A. Chakrabarti, F. Frommlet, and J.K. Ghosh. Asymptotic Bayes-optimality under sparsity of some multiple testing procedures. *Annals of Statistics*, 39(3):1551–1579, 2011.

[8] Z. Chi. On the performance of FDR control: Constraints and a partial solution. *Annals of Statistics*, 35(4):1409–1431, 2007.

[9] D. Donoho and J. Jin. Higher criticism for detecting sparse heterogeneous mixtures. *Annals of Statistics*, 32(3):962–994, 2004.

[10] D. Donoho and J. Jin. Asymptotic minimaxity of false discovery rate thresholding for sparse exponential data. *Annals of Statistics*, 34(6):2980–3018, 2006.

[11] D. Donoho and J. Jin. Higher criticism for large-scale inference, especially for rare and weak effects. *Statistical Science*, 30(1):1–25, 2015.

[12] D.P. Foster and R.A. Stine. α-investing: A procedure for sequential control of expected false discoveries. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(2): 429–444, 2008.

[13] R. Foygel Barber and E.J. Candès. Controlling the false discovery rate via knockoffs. *Annals of Statistics*, 43(5):2055–2085, 2015.

[14] R. Foygel Barber and A. Ramdas. The p-filter: Multi-layer FDR control for grouped hypotheses. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2016.

[15] C.R. Genovese and L. Wasserman. Operating characteristics and extensions of the false discovery rate procedure. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):499–517, 2002.

[16] C.R. Genovese, K. Roeder, and L. Wasserman. False discovery control with p-value weighting. *Biometrika*, 93(3):509–524, 2006.

[17] P. Hall and J. Jin. Properties of higher criticism under strong dependence. *Annals of Statistics*, 36(1):381–402, 2008.

[18] P. Hall and J. Jin. Innovated higher criticism for detecting sparse signals in correlated noise. *Annals of Statistics*, 38(3):1686–1732, 2010.

[19] J. Hu, H. Zhao, and H. Zhou. False discovery rate control with groups. *Journal of the American Statistical Association*, 105(491):1215–1227, 2010.

[20] A. Javanmard and A. Montanari. On online control of false discovery rate. *arXiv preprint arXiv:1502.06197*, 2015.

[21] J. Jin and T. Ke. Rare and weak effects in large-scale inference: Methods and phase diagrams. *arXiv preprint arXiv:1410.4578*, 2014.

[22] A. Li and R. Foygel Barber. Multiple testing with the structure adaptive Benjamini-Hochberg algorithm. *arXiv preprint arXiv:1606.07926*, 2016.

[23] Pierre Neuvial and Etienne Roquain. On false discovery rate thresholding for classification under sparsity. *Annals of Statistics*, 40(5):2572–2600, 2012.

[24] A. Ramdas, R. Foygel Barber, M.J. Wainwright, and M.I. Jordan. A unified treatment of multiple testing with prior knowledge. *arXiv preprint arXiv:1703.06222*, 2017.

[25] J. Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):479–498, 2002.

[26] J. Storey, J. Taylor, and D. Siegmund. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: A unified approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(1):187–205, 2004.

[27] J.W. Tukey. *The Problem of Multiple Comparisons: Introduction and Parts A, B, and C.* Princeton University, 1953.