Optimal Rates and Tradeoffs in Multiple Testing

Maxim Rabinovich[†] Aaditya Ramdas[*,†]

Michael I. Jordan[*,†] Martin J. Wainwright[*,†]

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

Supplementary Material

These materials contain proofs that — for space reasons — we could not accommodate in the main body. These include parts of proofs of theorems, as well as proofs of corollaries and technical lemmas.

S1 Proof of (4.31)

In order to establish the claim (4.31), define the events

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

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

Lemma S1. For any threshold T_n such that $\mathrm{FDR}_n(T_n) \leq q_n$, we have

$$\mathbb{P}[\mathcal{E}_1] \overset{(a)}{\geq} 3/8, \quad \text{and} \quad \mathbb{P}[\mathcal{E}_2] \overset{(b)}{\geq} 3/4. \tag{S1.1a}$$

We prove this lemma below. Using it, we can complete the proof of claim (4.31).

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

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

where step (i) follows by applying the probability bounds from Lemma S1.

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 (4.31).

S1.1 Proof of Lemma S1

Our proof makes use of the following auxiliary lemma:

Lemma S2. For $q_n \in (0, 1/24)$, we have

$$\mathbb{P}\Big[\mathrm{FDP}_n\big(t\big) \geq 8q_n \quad \text{for all } t \in [0, \tau_{\min}(4q_n)]\Big] \geq \frac{1}{2}. \tag{S1.2}$$

We return to prove this claim in Appendix S1.2. For the moment, we take it as given and complete the proof of Lemma S1.

Control of $\mathcal{E}_1$: Let us now prove the first bound in Lemma S1, namely that $\mathbb{P}[\mathcal{E}_1] \geq \frac{3}{8}$ where $\mathcal{E}_1 := \big\{T_n \geq \tau_{\min}(4q_n)\big\}$. So as to simplify notation, let us define the event

$$\mathcal{D} := \Big\{\mathrm{FDP}_n\big(t\big) \geq 8q_n \quad \text{for all } t \in [0, \tau_{\min}(4q_n)]\Big\}. \tag{S1.3a}$$

Now observe that

$$\mathbb{P}\Big[T_n \geq \tau_{\min}\big(4q_n\big)\Big] \;\geq\; \mathbb{P}\Big[T_n \geq \tau_{\min}\big(4q_n\big) \text{ and } \mathcal{D}\Big] \;=\; \mathbb{P}[\mathcal{D}] - \mathbb{P}\big[T_n \leq \tau_{\min}\big(4q_n\big) \text{ and } \mathcal{D}\big]. \tag{S1.3b}$$

Now by the definition (S1.3a) of the event $\mathcal{D}$, we have the inclusion

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

Combining with our earlier bound (S1.3b), we see that

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

It remains to control the two probabilities on the right-hand side of this bound. Applying Lemma S2 guarantees that $\mathbb{P}[\mathcal{D}] \geq \frac{1}{2}$. On the other hand, by Markov's inequality, the assumed lower bound $\mathrm{FDR}_n\big(T_n\big) \leq q_n$ implies that $\mathbb{P}\Big[\mathrm{FDP}_n\big(T_n\big) \geq 8q_n\Big] \leq \frac{1}{8}$. Putting together the pieces, we conclude that

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

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

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

Since $r_{\min} > \beta_n$, we have $\big|t - \mu\big| = \mu - \tau_{\min}$ and

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

from which it follows that

$$\frac{\mathbb{E}[F_n]}{n^{1-\beta_n}} \;=\; 1 - \Psi\big(t - \mu\big) \geq \frac{n^{\beta_n - r_n}}{Z_\ell}. \tag{S1.4}$$

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

$$\begin{aligned} 1 - \mathbb{P}[\mathcal{E}_2] = \mathbb{P}\left[F_n \leq \frac{\mathbb{E}[F_n]}{2}\right] &\leq \exp\big(-\frac{\mathbb{E}[F_n]}{12}\big) \\ &\stackrel{(i)}{\leq} \exp\big(-\frac{n^{1-r_n}}{12 Z_\ell}\big) \\ &\stackrel{(ii)}{\leq} \exp\big(-\frac{n^{1-r_{\max}}}{12 Z_\ell}\big), \end{aligned} \tag{S1.5}$$

where step (i) follows from the lower bound (S1.4), and step (ii) follows since $r_n < r_{\max}$ by assumption.

Case 2: Otherwise, we may assume that $r_n \in \big(\beta_n, r_{\min}\big)$. 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

$$1 - \mathbb{P}\big[\mathcal{E}_2\big] \leq \exp\big(-\frac{n^{1-\beta_n}}{24}\big)\ \leq \exp\big(-\frac{n^{1-r_{\max}}}{24}\big), \tag{S1.6}$$

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

Putting together the two bounds (S1.5) and (S1.6), we conclude that $\mathbb{P}[\mathcal{E}_2] \geq \frac{3}{4}$ for all sample sizes n large enough to ensure that

$$\max\big\{\exp\big(-\frac{n^{1-r_{\max}}}{24 Z_\ell}\big),\ \ \exp\big(-\frac{n^{1-r_{\max}}}{24}\big)\big\} \leq \frac{1}{4}, \tag{S1.7}$$

as was claimed. Note that condition (S1.7) is identical to condition (3.18), so that our definition of $n_{\min}$ guarantees that (S1.7) is satisfied. This completes the proof.

S1.2 Proof of Lemma S2

It remains to prove our auxiliary result stated in Lemma S2. 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

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

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

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

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

$$\mathbb{P}\big[L_n(\tau) \geq 3sn^{1-\beta}\big] \geq 1/2, \tag{S1.8}$$

where $s = 4q_n \in \big(0,\ 1/6\big)$ by assumption. 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

$$\alpha \colon = \Psi\big(\tau\big) \geq \frac{6s}{n^\beta} > \frac{16\log 4}{n}. \tag{S1.9}$$

A simple calculation based on this inequality yields

$$\alpha n - 3sn^{1-\beta} \geq \frac{\alpha n}{2} \geq \sqrt{\big(4\log 4\big)\alpha\big(1-\alpha\big)n} := a\sigma, \tag{S1.10}$$

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

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

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

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.

S2 Proof of Lemmas 1 and 2

This appendix is devoted to the proofs of Lemmas 1 and 2 from the main paper. 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

$$F_n(t) = \sum_{i \notin \mathcal{H}_0} \mathbf{1}\big(X_i \leq t\big) \sim \mathrm{Bin}\Big(1 - \Psi\big(t - \mu\big),\ n^{1-\beta}\Big) \tag{S2.11}$$

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,

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

as claimed.

Case 2: Otherwise, we may assume that $\mu > \tau$. Recall the parameterization (3.13) of μ in terms of r, the definition (3.15) of $r_{\min}$, and the definition (3.19) of the D_γ distance. In terms of these quantities, we have

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

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 (3.24). 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

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

To prove this, we let

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

and note that by (3.16), 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

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

We observe that we only need to allow $0 \leq x \leq \max\left(\beta + \kappa_n,\ r_{\text{min}}\right) - \tilde{r} =: \tilde{R} - \tilde{r}$, 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

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

If we now consider $q_n' = c q_n$, we can recover the more refined statements in Lemmas 1 and 2, simply by noting that the same reasoning as above shows

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

concluding the argument.

S3 Proof of Corollary 1

Although Corollary 1 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 1, thereby saving some additional messy calculations.

First, we verify that there is indeed a unique solution κ_* to the fixed point equation (3.21). Define the function as $g(\kappa) := D_\gamma\left(\beta+\kappa, r\right)^{1/\gamma} - \kappa^{1/\gamma}$. Clearly the solutions to (3.21) 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 $\kappa > r-\beta$. 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 (3.22), let $\mathcal{I}$ be an arbitrary threshold-based multiple testing procedure. We may assume without loss of generality that

$$\mathrm{FDR}_n(\mathcal{I}) \leq \min\left\{n^{-\kappa_*}, \frac{1}{24}\right\} \leq c(\beta,\gamma)n^{-\kappa_*}, \tag{S3.12}$$

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 $[r-\beta,\ \infty)$.

where the quantity $c(\beta, \gamma) \geq 1$ was defined in the statement of Theorem 1 (otherwise, the claimed lower bound (3.22) follows immediately).

Applying the second part of Lemma 1 and defining $\tilde{c} = 4c(\beta, \gamma)$, we conclude that

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

S4 Proof details for Theorem 2

S4.1 Achievability for the BH procedure

In this section, we prove that BH achieves the lower bound whenever $r_n > r_{\min}\big(c_{\mathrm{BH}} q_n\big)$. Specifically, we prove the claim (3.26) stated in Theorem 2.

We first show how to derive the upper bound (3.26) from the probability bound (4.33) and then prove the probability bound itself. Note that since BH is a valid FDR control procedure, we necessarily have $\mathrm{FDR}_n(t_{\mathrm{BH}}) \leq q_n$. To bound the FNR, first let $\mathcal{E} = \{t_{\mathrm{BH}} \leq \tau_{\min,\mathrm{BH}}\}$ 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 (4.33), together with Lemma 2, implies that

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

where the final step uses the definition (3.23) 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 probability bound with an argument using p-values and survival functions that parallels that of Arias-Castro and Chen (2017) but that sidesteps CDF asymptotics. To carry out the analysis, we first study the relationship between the population survival function Ψ and the empirical survival function $\hat{\Psi}$, defined by

$$\hat{\Psi}(t) = \big(1 - \frac{1}{n^{\beta_n}}\big) \cdot \hat{\Psi}_0(t) + \frac{1}{n^{\beta_n}} \cdot \hat{\Psi}_1(t), \tag{S4.13}$$

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

Now, sort the observations in *decreasing* order, so that $X_{(1)} \ge X_{(2)} \ge \cdots \ge$

$X_{(n)}$, and define p-values

$$p_{(i)} = \Psi\left(X_{(i)}\right) \quad \text{and} \quad \hat{\Psi}\left(X_{(i)}\right) = \frac{i}{n}, \tag{S4.14}$$

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

$$i_{\mathrm{BH}} = \max\left\{1 \leq i \leq n \colon \Psi\left(X_{(i)}\right) \leq q_n \hat{\Psi}\left(X_{(i)}\right)\right\}. \tag{S4.15}$$

Moving t_{BH} within $\left(X_{(i_{\mathrm{BH}}+1)},\ X_{(i_{\mathrm{BH}})}\right]$ if necessary, we may therefore assume $\Psi(t) > q_n \hat{\Psi}(t)$ whenever $t < t_{\mathrm{BH}}$, and combining this knowledge with (S4.13), we obtain the chain of inclusions

$$\begin{aligned}\mathcal{E}^c = \{t_{\mathrm{BH}} > \tau_{\min,\mathrm{BH}}\} &\subset \left\{\Psi\left(\tau_{\min,\mathrm{BH}}\right) > q_n \hat{\Psi}\left(\tau_{\min,\mathrm{BH}}\right)\right\} \\ &\subset \left\{\Psi\left(\tau_{\min,\mathrm{BH}}\right) > \frac{q_n}{n^{\beta_n}} \cdot \frac{W_n}{n^{1-\beta_n}}\right\} =: \widetilde{\mathcal{E}}^c, \end{aligned} \tag{S4.16}$$

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

We now argue that $\Psi\left(\tau_{\min,\mathrm{BH}}\right) \leq \frac{q_n}{4n^{\beta_n}}$, so that $\mathbb{P}\left(\mathcal{E}^c\right) \leq \mathbb{P}\left(\widetilde{\mathcal{E}}^c\right) \leq \mathbb{P}\left(W_n \leq \frac{n^{1-\beta_n}}{4}\right)$. For this, observe that by the definition of $r_{\min}$ in (3.15) and the upper tail bound (2.5), we have

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

We conclude

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

Finally, by a Bernstein bound, we find

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

where we have used the fact that $\tau_{\min,\mathrm{BH}} \leq \mu_n$ to conclude that $\Psi\left(\tau_{\min,\mathrm{BH}} - \mu_n\right) \geq \frac{1}{2}$ and therefore $\mathbb{E}[W_n] \geq \frac{n^{1-\beta_n}}{2}$. We have therefore established the required claim (4.33), concluding the proof of optimality of the BH procedure.

S4.2 Achievability 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 (3.26) by controlling the algorithm's threshold as Since the proof of equation (3.27) from the bound (4.34) is essentially identical to the corresponding derivation for the BH procedure, we omit it. We now prove the bound (4.34) by an argument somewhat different than that used in analyzing the BH procedure. Define the integers

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

Then, the definition of the BC procedure gives

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

To prove (4.34), it therefore suffices to show that

$$\mathbb{P}\left(\frac{1 + N_-(\tau_{\min,\mathrm{BC}})}{1 \vee N_+(\tau_{\min,\mathrm{BC}})} > q_n\right) \leq q_n + \exp\big(-\frac{n^{1-r_{\max}}}{24}\big). \tag{S4.17}$$

We prove the bound (S4.17) in two parts:

$$\mathbb{P}\big(1 \vee N_+(\tau_{\min,\mathrm{BC}}) < \frac{n^{1-\beta_n}}{4}\big) \leq \exp\big(-\frac{n^{1-r_{\max}}}{24}\big), \tag{S4.18a}$$

$$\mathbb{P}\left(1 + N_-(\tau_{\min,\mathrm{BC}}) > q_n \cdot \frac{n^{1-\beta_n}}{4}\right) \leq q_n. \tag{S4.18b}$$

These bounds are a straightforward consequence of elementary Bernstein bounds, and together they imply the claim (S4.17). We explain them below.

The lower bound (S4.18a) follows because $1 \vee N_+(\tau_{\min,\mathrm{BC}}) \geq N_+(\tau_{\min,\mathrm{BC}})$ 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 $\Psi\big(\tau_{\min,\mathrm{BC}} - \mu\big) \geq \frac{1}{2}$ 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 $N_+\left(\tau_{\min,\mathrm{BC}}\right) \geq N_+^{\mathrm{signal}}$, whence

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

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

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

as required to prove equation (S4.18a). The proof of equation (S4.18b) 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 (3.15) of $r_{\min}$ and the upper bound on the tail (2.5)—each one has probability of success upper bounded by $1 - \Psi\big(-\tau_{\min,\mathrm{BC}}\big) \leq \frac{6Z_\ell}{Z_u} \cdot c_{\mathrm{BC}} q_n n^{-\beta_n} = \frac{1}{8} \cdot q_n n^{-\beta_n}$. Formally, we may write $N_-(\tau_{\min,\mathrm{BC}}) = N_-^{\mathrm{null}} +$

N_-^{signal}, with

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

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

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

On the other hand, using the lower bound in (2.5), we find $1 - \Psi\big(-\tau_{\text{min,BC}}\big) \geq 6c_{\text{BC}} q_n n^{-\beta_n}$. Using the additional fact that $n - n^{1-\beta_n} \geq \frac{n}{2}$ by (3.14a), we may conclude that

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

By a Bernstein bound, it follows that

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

where we have invoked the decay condition (3.25) for the last step.

S5 Proof of Corollary 2

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

$$r_{\min}(\kappa_n(c_{\mathrm{A}} q_*)) = \beta + \kappa_* + \frac{\log \frac{1}{6c_* c_{\mathrm{A}} Z_\ell}}{\log n}. \tag{S5.19}$$

Suppose for now that the decay condition (3.25) holds for q_* and some choice of $n_{\min,\mathrm{BC}}$. Then, using (S5.19) 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 $\tilde{c} = \max\left\{c',\ (n'_{\min})^{\kappa_*}\right\}$ (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 (3.25), 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.

Bibliography

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.

E. Arias-Castro and S. Chen. Distribution-free multiple testing. *Electronic Journal of Statistics*, 11(1):1983–2001, 2017.

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.

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

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.

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

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.

S. Chen and E. Arias-Castro. Sequential multiple testing. *arXiv preprint arXiv:1705.10190*, 2017.

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

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

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

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

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.

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

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.

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.

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

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

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

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

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

P. Ji and J. Jin. UPS delivers optimal phase diagram in high-dimensional variable selection. *Annals of Statistics*, 40(1):73–103, 2012.

P. Ji and Z. Zhao. Rate optimal multiple testing procedure in high-dimensional regression. *arXiv preprint arXiv:1404.2961*, 2014.

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

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

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

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.

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

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.

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