The edge of discovery: Controlling the local false discovery rate at the margin

Jake A. Soloff[1], Daniel Xiang[2], and William Fithian[1]

[1]*Department of Statistics, University of California, Berkeley*
[2]*Department of Statistics, University of Chicago*

March 31, 2022

Abstract

Despite the popularity of the false discovery rate (FDR) as an error control metric for large-scale multiple testing, its close Bayesian counterpart the local false discovery rate (lfdr), defined as the posterior probability that a particular null hypothesis is false, is a more directly relevant standard for justifying and interpreting individual rejections. However, the lfdr is difficult to work with in small samples, as the prior distribution is typically unknown. We propose a simple multiple testing procedure and prove that it controls the expectation of the maximum lfdr across all rejections; equivalently, it controls the probability that the rejection with the largest p-value is a false discovery. Our method operates without knowledge of the prior, assuming only that the p-value density is uniform under the null and decreasing under the alternative. We also show that our method asymptotically implements the oracle Bayes procedure for a weighted classification risk, optimally trading off between false positives and false negatives. We derive the limiting distribution of the attained maximum lfdr over the rejections, and the limiting empirical Bayes regret relative to the oracle procedure.

1 Introduction

A common goal in applications of multiple hypothesis testing is to identify a relatively short list of candidate "discoveries" that are sufficiently promising to undertake some costly further action. In scientific applications, for example, each discovery may be the focus of a follow-up experiment, which wastes resources if the apparent discovery was only a mirage. The *false discovery rate* (FDR, Benjamini and Hochberg, 1995) has become a cornerstone of modern large-scale multiple testing because it directly measures the rate of this wastage:

> [T]he proportion of errors in the pool of candidates is of great economical significance since follow-up studies are costly, and thus avoiding multiplicity control is costly. Indeed, the FDR criterion is economically interpretable; when considering

a potential threshold, the adjusted FDR gives the proportion of the investment that is about to be wasted on false leads. (Reiner et al., 2003)

An analyst who controls FDR at level $q = 5\%$, then, is willing to waste resources following up on one false discovery in exchange for every nineteen real discoveries.

Carrying this reasoning further, however, we can apply the same cost-benefit analysis to each individual rejection, not only to the list of rejections taken as a whole. In economic terminology, we should consider not only the *average utility* of our entire rejection set, but also the *marginal utility* of each rejection we make, since we always have the option to exclude any rejection that is not individually promising. For example, in Section 4 we reproduce the simulations of Benjamini and Hochberg (1995) and find in some settings that, even while the Benjamini–Hochberg (BH) procedure controls FDR at level $q = 5\%$, the *last discovery* (i.e. the discovery with the largest p-value) is false more than 30% of the time. In such settings, unless we are willing to suffer one false discovery for every two true discoveries, we would be better served by excluding the last rejection from the BH rejection set. More generally, to decide where to set our rejection threshold, we should ask about the proportion of false leads among the incremental rejections that we would add or remove by raising or lowering it.

The likelihood that an individual discovery is a false lead is called its *local false discovery rate* (lfdr, Efron et al., 2001). For $i = 1, \ldots, m$, let $H_i = 0$ if the ith hypothesis is null and $H_i = 1$ otherwise, and consider the simple *Bayesian two-groups model*

$$p_i \mid H_i = h \overset{\text{ind}}{\sim} f_h, \qquad \text{with} \qquad H_i \overset{\text{iid}}{\sim} \text{Bern}(1 - \pi_0), \qquad \text{for } i = 1, \ldots, m, \tag{1}$$

where $f_0 := 1_{[0,1]}$ and f_1 are densities (null and alternative, respectively) supported on the unit interval $[0, 1]$, and the null proportion is $\pi_0 \in [0, 1]$. Let $f := \pi_0 + (1 - \pi_0) f_1$ denote the common mixture density of the p-values in model (1), and let $F(t) := \int_0^t f(u)\, du$ denote the corresponding cumulative distribution function (cdf). The lfdr is then defined as the posterior probability that $H_i = 0$, conditional on the observed p-value p_i:

$$\text{lfdr}(t) := \mathsf{P}\left(H_i = 0 \mid p_i = t\right) = \frac{\pi_0}{f(t)}. \tag{2}$$

If we knew the problem parameters π_0 and f_1, then the definition (2) would neatly solve the problem posed above: we should reject only those hypotheses whose lfdr is below the break-even threshold of our cost-benefit tradeoff. Concretely, let $\lambda > 0$ define the ratio between the cost of each false discovery and the benefit of each true discovery. Then the utility of making R rejections, of which V are false discoveries, is proportional to $(R - V) - \lambda V$, and a simple calculation shows that we should reject the ith hypothesis if and only if $\text{lfdr}(p_i) \le \alpha := \frac{1}{1+\lambda}$.

We will usually work under the additional assumption that $f_1(t)$ is non-increasing in t, or equivalently that $\text{lfdr}(t)$ is non-decreasing, so that smaller p-values represent stronger evidence against the null. This assumption, common in multiple testing (see, e.g., Genovese and Wasserman, 2004; Langaas et al., 2005; Strimmer, 2008), lets us restrict our attention to procedures that reject all p-values below a given threshold: if f_1 is non-increasing then rejecting when $\text{lfdr}(p_i) \le \alpha$ is equivalent to rejecting when p_i is sufficiently small.

In practice, π_0 and f_1 are typically unknown and must be estimated from the data, and many estimators have been proposed; see e.g. Efron et al. (2001); Pounds and Morris (2003); Scheid and Spang (2004); Aubert et al. (2004); Efron (2004, 2008); Liao et al. (2004); Pounds and Cheng (2004); Robin et al. (2007); Strimmer (2008); Muralidharan (2010); Patra and Sen (2016); Stephens (2017). To the best of our knowledge, however, there are no known finite-sample lfdr control guarantees for multiple testing procedures based on these methods. By contrast, simple, robust, and well-known methods like the Benjamini–Hochberg (BH) procedure of Benjamini and Hochberg (1995) enjoy finite-sample FDR control without requiring the analyst to model the p-value distribution.

In this work, we introduce a new error control metric that measures the lfdr of a multiple testing procedure's least promising rejection. We represent a generic multiple testing method as a function $\mathcal{R}(p_1, \ldots, p_m)$ returning an index set $\mathcal{R} \subseteq \{1, \ldots, m\}$, where hypothesis i is rejected if and only if $i \in \mathcal{R}$. We say the procedure's *max-lfdr* is

$$\text{max-lfdr}(\mathcal{R}) := \mathbb{E}\left[\max_{i \in \mathcal{R}} \text{lfdr}(p_i)\right], \tag{3}$$

defining the maximum as zero if no rejections are made. If f_1 is non-increasing, then the max-lfdr of $\mathcal{R}$ coincides with the probability that the last rejection is a false discovery.

We also introduce a simple multiple testing procedure, which we call the *support line* (SL) procedure, that provably controls the max-lfdr under mild assumptions. Define the p-value order statistics $p_{(1)} \leq \cdots \leq p_{(m)}$, and let $p_{(0)} = 0$ by convention. Then our procedure rejects p-values up to the last (and a.s. unique) minimizer

$$R_q := \operatorname*{argmin}_{k=0,\ldots,m} \left\{ p_{(k)} - \frac{qk}{m} \right\}. \tag{4}$$

That is, we reject $\mathcal{R}_q := \{i : p_i \leq \tau_q\}$, for the threshold $\tau_q = p_{(R_q)}$. Under the two-groups model (1), with non-increasing f_1, we show in Theorem 1 that

$$\text{max-lfdr}(\mathcal{R}_q) = \pi_0 q.$$

Our method can be implemented without knowing π_0 or f_1, apart from the shape constraint, and bears a close relationship to the BH procedure, which replaces R_q in (4) with

$$R_q^{\text{BH}} := \max\left\{ k \in \{0, \ldots, m\} : p_{(k)} \leq \frac{qk}{m} \right\},$$

rejecting $\mathcal{R}_q^{\text{BH}} := \{i : p_i \leq \tau_q^{\text{BH}}\}$, for $\tau_q^{\text{BH}} = qR_q^{\text{BH}}/m \geq p_{(R_q^{\text{BH}})}$. Because $R_q \leq R_q^{\text{BH}}$, the BH method makes at least as many rejections as the SL method, and both methods make at least one rejection if and only if $p_{(k)} \leq \frac{qk}{m}$ for some $k \geq 1$; however, as we will argue, in general, the SL method should be run with a strictly larger q than we would use for BH. The left panel of Figure 1 illustrates the relationship between the two methods by reproducing the familiar plot of the BH procedure as an operation on the order statistics $p_{(1)}, \ldots, p_{(m)}$.

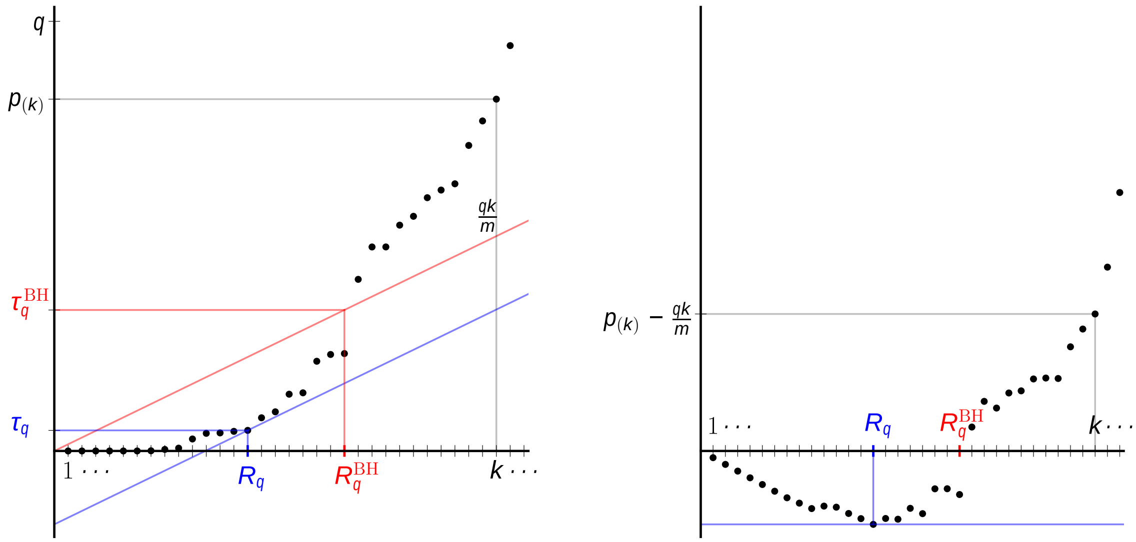

Figure 1: Left: The order statistics $p_{(k)}$ of the p-values as a function of the index k, shown in black. The BH procedure, in red, finds the largest index R_q^{BH} such that $p_{(R_q^{\mathsf{BH}})}$ falls below the ray of slope q/m; by contrast, our procedure finds the (last and almost surely unique) boundary point $(R_q, p_{(R_q)})$ of the supporting line of slope q/m. Right: The same plot with the ray through the origin of slope q/m subtracted off. The black dots represent a running estimate (7) of the weighted classification loss (5), which our procedure minimizes. BH(q) finds the largest threshold where the estimated loss is negative.

1.1 Multiple testing and the weighted classification loss

To formalize our analysis above, define the per-instance *weighted classification loss*:

$$L_\lambda(H, \mathcal{R}) := \frac{(1+\lambda)V - R}{m}. \tag{5}$$

This loss can be derived, up to additive and multiplicative constants, by viewing each of the m hypotheses as a binary classification problem, where we incur a cost c_1 for each type I error or false discovery ($i \in \mathcal{R}$, but $H_i = 0$), and cost c_2 from each type II error or false non-discovery ($i \notin \mathcal{R}$, but $H_i = 1$). If the total number of non-nulls is $m_1 = \sum_i H_i$, then there are $m_1 - (R - V)$ false non-discoveries, so the total loss over all m instances is

$$c_1 V + c_2(m_1 - (R - V)) = c_2 m \cdot L_\lambda(H, \mathcal{R}) + c_2 m_1,$$

where $\lambda = c_1/c_2$ is the ratio between the two misclassification costs. L_λ as defined in (5) is normalized so that rejecting nothing incurs zero loss, and each true discovery has value $1/m$.

Under the two-groups model (1), Sun and Cai (2007, Theorem 2) show that the corresponding Bayes risk $\mathbb{E}L_\lambda(H, \mathcal{R})$ is minimized by the oracle procedure

$$\mathcal{R}^* := \{i : \mathrm{lfdr}(p_i) \le \alpha\}, \quad \text{where} \quad \alpha = \frac{1}{1+\lambda}. \tag{6}$$

The ratio λ specifies the "break-even exchange rate" at which we are willing to trade true discoveries for false leads; e.g., if $\lambda = 19$ then we are willing to suffer a single false discovery for exactly 19 true discoveries, and we should reject a hypothesis only if its lfdr falls below the break-even tolerance $\alpha = 0.05$. If f_1 is non-increasing, then the oracle procedure reduces to thresholding p-values at a fixed threshold

$$R^* = \{i : p_i \leq \tau^*\}, \quad \text{for} \quad \tau^* := \max\{t \in [0,1] : \text{lfdr}(t) \leq \alpha\},$$

with $\tau^* = 0$ if no such threshold exists.

Our method can be directly interpreted as minimizing an empirical proxy of the weighted classification loss. For a candidate threshold $t \in [0,1]$, the expected number of null p-values below the threshold is $m\pi_0 t$. If π_0 is known, we obtain a running estimator of the loss:

$$\hat{L}^{\lambda}(t; \pi_0) = \frac{(1+\lambda)m\pi_0 t - mF_m(t)}{m} = (1+\lambda)\left(\pi_0 t - \alpha F_m(t)\right), \tag{7}$$

where $F_m(t)$ represents the empirical cumulative distribution function (ecdf) of the p-values:

$$F_m(t) = \frac{1}{m}\sum_{i=1}^{m} 1\{p_i \leq t\}.$$

Because $\hat{L}^{\lambda}(t; \pi_0)$ is increasing between successive order statistics, it is minimized at one of the order statistics, or at $p_{(0)} = 0$:

$$\underset{k=0,1,\ldots,m}{\text{argmin}}\ \hat{L}^{\lambda}(p_{(k)}; \pi_0) \;=\; \underset{k=0,1,\ldots,m}{\text{argmin}}\ \pi_0 p_{(k)} - \frac{\alpha k}{m}.$$

Comparing the last expression to the definition of our procedure in (4), we see that $\hat{L}^{\lambda}(t; \pi_0)$ is minimized at $t = \tau_q$ for $q = \alpha/\pi_0$. By Theorem 1, we then have exactly max-lfdr$(R_q) = \alpha$.

By contrast, τ_q^{BH} for $q = \alpha/\pi_0$ is the largest value of t that gives $\hat{L}^{\lambda}(t; \pi_0) = 0$, the same loss we would achieve by rejecting nothing at all. In other words, the BH procedure at level α/π_0 only aims to break even; to do better, we should run BH at a strictly smaller level $q < \alpha/\pi_0$, viewing q as a tuning parameter as in Neuvial and Roquain (2012).

To select q for our SL procedure when π_0 is unknown, we can either conservatively bound $\pi_0 \leq 1$ and run the procedure at $q = \alpha$, or estimate π_0 and use $q = \alpha/\hat{\pi}_0$. To avoid confusion, we will always use the notation q to represent our method's tuning parameter, and reserve $\alpha = \frac{1}{1+\lambda}$ to represent the true target lfdr, defined in terms of the cost ratio λ.

Our procedure can alternatively be derived as a plug-in maximum likelihood estimator (MLE) of the oracle procedure R^*, where we estimate $f(t)$ using Grenander's nonparametric MLE for a non-increasing density (Grenander, 1956):

$$\hat{f}_m := \underset{\substack{g:[0,1] \to \mathbb{R}_+ \\ \text{non-increasing density}}}{\text{argmax}}\ \frac{1}{m}\sum_{i=1}^{m} \log g(p_i). \tag{8}$$

As we will see in Section 3.2, τ_q is also the largest value $t \in [0,1]$ for which $\hat{f}_m(t) \geq q^{-1}$. Thus, if we run our procedure at $q = \alpha/\pi_0$, we have

$$R_{\alpha/\pi_0} = \left\{ i : \hat{f}_m(p_i) \geq (\alpha/\pi_0)^{-1} \right\} = \left\{ i : \frac{\pi_0}{\hat{f}_m(p_i)} \leq \alpha \right\}.$$

As above, if π_0 is unknown, we can either estimate it or conservatively bound $\pi_0 \leq 1$.

The relationship between our method and the Grenander estimator is convenient for asymptotic analysis because the latter is very well studied; see the book by Groeneboom and Jongbloed (2014) for a thorough treatment. The Grenander estimator has previously been considered for estimating the lfdr (Strimmer, 2008) as well as for estimating the null proportion π_0 (Langaas et al., 2005). While $\hat{f}_m$ may be efficiently computed via the pool adjacent violators algorithm (Robertson et al., 1988), the definition in (4) is usually preferred for computational purposes.

1.2 The max-lfdr and the FDR

The max-lfdr in (3) and the FDR are two different error criteria that both appeal to the logic of trading off true and false discoveries. The key difference is that the FDR, defined as

$$\mathrm{FDR}(R) := \mathbb{E}\left[\frac{V}{R} 1\{R > 0\} \right],$$

measures the likelihood that a *randomly selected* rejection is null, whereas the max-lfdr instead measures the likelihood that the *least promising* rejection is null. In both cases the event in question is deemed not to have occurred if $R = 0$, so that under the global null (all $H_i = 0$, almost surely), both criteria reduce to the probability of making a single rejection.

Throughout this section, we will restrict our attention to procedures that reject the R hypotheses with the smallest p-values. That is, we assume a procedure R rejects $H_{(1)}, \ldots, H_{(R)}$, where $H_{(k)}$ represents the hypothesis corresponding to $p_{(k)}$. If f_1 is non-increasing, then the procedure's *last rejection* $H_{(R)}$ is the least promising, and the max-lfdr can be equivalently characterized as the probability that the last rejection is a false discovery:

$$\text{max-lfdr}(R) = \mathbb{E}\left[\text{lfdr}\left(p_{(R)}\right) 1\{R > 0\} \right] = \mathbb{P}\left(H_{(R)} = 0, R > 0 \right). \tag{9}$$

If $\text{max-lfdr}(R) > \alpha = \frac{1}{1+\lambda}$, then we can improve R by excluding its last discovery.[1] Let R^{-1} denote the procedure that makes one fewer rejection than R, meaning it rejects $H_{(1)}, \ldots, H_{(R-1)}$ if $R > 0$, and makes no rejections if $R = 0$. Then we have

$$\begin{aligned}
\mathbb{E}[L_\lambda(H, R) - L_\lambda(H, R^{-1})] &= \frac{1}{m}\mathbb{E}\left[(1+\lambda) 1\{H_{(R)} = 0, R > 0\} - 1\{R > 0\} \right] \\
&= \frac{1+\lambda}{m}\left(\text{max-lfdr}(R) - \alpha \mathbb{P}\{R > 0\} \right),
\end{aligned}$$

[1] Without the shape constraint on f_1, max-lfdr $> \alpha$ still implies that the analyst could improve the procedure by removing the least promising rejection, which may not be the same as the last rejection. However, this improvement is only feasible if the analyst can recognize which rejection is least promising.

which is positive if max-lfdr($\mathcal{R}$) $> \alpha$. The converse, that dropping the last rejection does not improve the risk if max-lfdr($\mathcal{R}$) $\leq \alpha$, is almost true if $\mathsf{P}\{R > 0\} \approx 1$. Under the global null, however, any procedure is improved by making fewer rejections.

This thought experiment — what if we dropped the last rejection? — is at the heart of our motivation for proposing the max-lfdr as an error criterion. Even when a rejection set's average quality is high, the rejections near the threshold may be recognizably bad bets. In that case, we are better off "trimming the fat" from our rejection set until all of the rejections that remain are individually worth following up on.

Because max-lfdr($\mathcal{R}$) $\geq$ FDR($\mathcal{R}$), controlling the max-lfdr is more conservative than controlling FDR at the same level q, in most cases considerably so. From this, it is tempting to conclude that max-lfdr control is an inherently more conservative goal than FDR control, but this conclusion would be mistaken. An analyst whose break-even exchange rate is $\lambda = 9$ and break-even tolerance is $\alpha = 0.1$, for example, would never choose a method with a 10% FDR; the resulting rejection set would be no better on average than rejecting nothing at all, so there would be no point in collecting the data in the first place. Thus, an analyst who is satisfied with a 10% FDR must have a larger break-even tolerance, say $\alpha = 0.2$ or 0.3.

By the same token, it would be unfair to evaluate the risk under L_λ of the BH procedure at level $q = \alpha = \frac{1}{1+\lambda}$, since an analyst whose break-even tolerance is α would want to control FDR at a strictly smaller level q, like $\alpha/2$ or $\alpha/10$. However, as we show in Section 3.1, the performance of BH(q) with such *a priori* choices of q can depend sensitively on the unknown alternative density f_1.

1.3 Outline and contributions

In Section 2, we state and prove our main result, that max-lfdr($\mathcal{R}_q$) $= \pi_0 q$ under the Bayesian two-groups model with non-increasing f_1, applying a result of Takács (1967). Even without monotonicity of f_1, we have $\mathsf{P}\{H_{(R_q)} = 0, R_q > 0\} = \pi_0 q$, but monotonicity ensures that the lfdr is not out of control for rejections in the interior of the rejection region. We also prove max-lfdr control for an adaptive method that estimates π_0 from the data in the same way as the procedure of Storey (2002).

In Section 3, we investigate our method's asymptotic performance relative to the oracle procedure $\mathcal{R}^*$. Extending asymptotic results for the Grenander estimator, we show that our method's attained lfdr threshold, lfdr(τ_q), concentrates at a rate $m^{-1/3}$ around its expectation $\pi_0 q$, giving an explicit formula for its asymptotic distribution. We also show that our method's asymptotic regret relative to the oracle shrinks at the rate $m^{-2/3}$. Section 4 illustrates our results with selected simulations, and Section 5 concludes.

2 Finite-sample max-lfdr control

2.1 Main result

Our main result is that our procedure $\mathcal{R}_q$ controls the max-lfdr at exactly $\pi_0 q$.

Theorem 1. Suppose $p_1, \ldots, p_m$ follow the Bayesian two-groups model (1), with $f_0 = 1_{[0,1]}$. For the procedure defined in (4), we have

$$\mathbb{E}\left[\mathrm{lfdr}\left(p_{(R_q)}\right) 1\{R_q > 0\}\right] = \mathbb{P}\left(H_{(R_q)} = 0, R_q > 0\right) = \pi_0 q. \qquad (10)$$

If f_1 is non-increasing, then we have

$$\text{max-lfdr}(\mathcal{R}_q) = \pi_0 q.$$

The familiar optional-stopping arguments from the FDR control literature, introduced by Storey et al. (2004), do not seem to apply to our procedure, since the minimizer R_q of the sequence $p_{(k)} - qk/m$ for $k = 0, \ldots, m$ is not a stopping time. We instead prove Theorem 1 via a conditioning argument, which crucially relies on the fact that each null p-value has exactly a q/m chance of being the last rejection $p_{(R_q)}$.

Lemma 2. Fix $p_1, \ldots, p_{m-1} \in [0,1]$ and let $p_m \sim \mathrm{Unif}(0,1)$. Then $\mathbb{P}\{p_{(R_q)} = p_m\} = q/m$.

Given Lemma 2, the proof of Theorem 1 is straightforward:

Proof of Theorem 1. Because the (H_i, p_i) pairs are independent and identically distributed, we can decompose the probability in (10) as

$$\begin{aligned}
\mathbb{P}\left(H_{(R_q)} = 0, R_q > 0\right) &= \sum_{i=1}^{m} \mathbb{P}\left(H_i = 0, p_{(R_q)} = p_i\right) \\
&= m\,\mathbb{P}\left(H_m = 0, p_{(R_q)} = p_m\right) \\
&= \pi_0 m\,\mathbb{P}\left(p_{(R_q)} = p_m \mid H_m = 0\right) \\
&= \pi_0 q,
\end{aligned}$$

where the last step comes from conditioning on $p_1, \ldots, p_{m-1}$ and applying Lemma 2. If $f_1(t)$ is non-increasing, then $\mathrm{lfdr}(t)$ is non-decreasing, so that $\max_{i \in \mathcal{R}_q} \mathrm{lfdr}(p_i) = \mathrm{lfdr}(p_{(R_q)})$ almost surely, completing the argument. □

We now turn to proving Lemma 2. Because p_m is uniform, the probability statement is equivalent to a showing that, for any fixed $p_1, \ldots, p_{m-1} \in [0,1]$, the set of "winning values" $p_m \in [0,1]$, for which $\tau_q(p_1, \ldots, p_m) = p_m$, has Lebesgue measure q/m. To prove this fact, we rely on a useful result of Takács (1967), which we state next:

Lemma 3. (Takács, 1967, Theorem 1) Let $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ denote a non-decreasing step function with $\varphi(0) = 0$. Assume that, for some positive q, we have $\varphi(u + q) = \varphi(u) + \varphi(q)$ for all $u \geq 0$, and define

$$\delta(u) = 1\{v - \varphi(v) \geq u - \varphi(u) \text{ for all } v \geq u\}.$$

Then we have

$$\int_0^q \delta(u)\,du = (q - \varphi(q))_+ .$$

Lemma 2 is proved by designing a function φ for which the corresponding indicator $\psi(p_m)$ in Lemma 3 checks whether p_m is the last rejection when we run our method on $(p_i)_{i=1}^m$.

Proof of Lemma 2. Let $F_{m-1}(t) = \frac{1}{m-1}\sum_{i=1}^{m-1} 1\{p_i \le t\}$ denote the ecdf of $p_1, \ldots, p_{m-1}$, and define a new function φ on $[0, q]$

$$\varphi(v) := \begin{cases} qF_{m-1}(v)\frac{m-1}{m} & \text{if } v < q \\ q\frac{m-1}{m} & \text{if } v = q. \end{cases}$$

Next, extend φ to a non-decreasing step function on all of $\mathbb{R}_+$ by $\varphi(kq + v) = k\varphi(q) + \varphi(v)$ for all positive integers k and $v \in [0, q]$.

Now let $F_m(t) = \frac{1}{m}\sum_{i=1}^{m} 1\{p_i \le t\}$. If $\tau_q = p_{(R_q)} = p_m$ then we have $p_m \le \frac{qR_q}{m}$ $p_{(0)} \le q\frac{0}{m} = 0$, so we may restrict our attention to $p_m \le q$. On the range $v \in [p_m, q)$ we have

$$mF_m(v) = 1 + (m-1)F_{m-1}(v), \quad \text{so} \quad \varphi(v) = qF_m(v) - \frac{q}{m}.$$

On the range $v \in [q, q + p_m)$, we have

$$mF_m(v - q) = (m-1)F_{m-1}(v - q), \quad \text{so} \quad \varphi(v) = qF_m(v - q) - \frac{q}{m} + q.$$

Letting $\psi(p_m) := 1\{p_m = \tau_q(p_1, \ldots, p_m)\}$,

$$\begin{aligned} \psi(p_m) &= 1\{v - qF_m(v) \le p_m - qF_m(p_m) \text{ for all } v \in [0, 1]\} \\ &= 1\{v - \varphi(v) \le p_m - \varphi(p_m) \text{ for all } v \in [p_m, q + p_m)\} \\ &= 1\{v - \varphi(v) \le p_m - \varphi(p_m) \text{ for all } v \ge p_m\}; \end{aligned}$$

where the last step follows from the fact that $\varphi(v) > \varphi(v - q)$ for all $v \ge q + p_m$. We have checked the conditions of Lemma 3 (Takács, 1967, Theorem 1), from which we conclude

$$P(\tau_q = p_m) = \int_0^q \psi(p_m)\, dp_m = (q - \varphi(q))_+ = \frac{q}{m}. \qquad \square$$

To convey some intuition for our result, Figure 2 depicts an illustrative example, highlighting in green the "winning values" of p_m such that $\hat{\tau}_q = p_m$.

Remark 4. Because the set of "winning values" in Lemma 2 is a subset of $[0, q]$ with Lebesgue measure q/m, we can trivially extend the result to conclude $P\{p_{(R_q)} = p_m\} \le q/m$, if p_m is drawn from any density f_0 with $f_0(t) \le 1$ for all $t \in [0, q]$. Likewise, we can extend Theorem 1 to show that $\text{max-lfdr}(R_q) \le \pi_0 q$ with a more general null density f_0, as long as $\text{lfdr}(t)$ is non-decreasing and $f_0(t) \le 1$ for all $t \in [0, q]$.

Figure 2: Intuition for Lemma 2. Black points represent the empirical cdf (scaled by $\frac{m-1}{m}$) of $p_1, \ldots, p_{m-1}$; red points represent how the empirical cdf gets shifted after adding a point p_m to its left. Adding a point p_m can shift the supporting line by at most $\frac{q}{m}$, and each possible shift in $[0, q/m]$ corresponds to precisely one p_m where p_m becomes the new support point.

2.2 Estimating π_0

Theorem 1 parallels the exact FDR guarantee $\mathrm{FDR}(R_q^{BH}) = \pi_0 q$ for the BH procedure. If we bound $\pi_0 \le 1$, we can run our method at level $q = \alpha$ and ensure that we conservatively control max-lfdr at $\pi_0 \alpha$, but our method will be overly conservative. In this section, we consider estimating π_0 using the Storey (2002) estimator of the null proportion, defined as

$$\hat{\pi}_0 := \frac{1 + \#\{i : p_i > \lambda\}}{(1-\lambda)m}, \tag{11}$$

modifying an estimator originally proposed by Schweder and Spjøtvoll (1982).

Our next result shows that plugging in $\hat{\pi}_0$ and running a modification of our procedure at level $q = \alpha/\hat{\pi}_0$ controls max-lfdr at level α in finite samples:

Theorem 5. Suppose $p_1, \ldots, p_m$ follow the Bayesian two-groups model (1), with $f_0 = 1_{[0,1]}$ and f_1 non-increasing. Fix $\lambda \in (0,1)$, and define a modified version of our SL procedure that only examines order statistics below λ:

$$R_q := \operatorname*{argmin}_{k \ge 0:\, p_{(k)} \le \lambda} \; \hat{\pi}_0 p_{(k)} - \frac{qk}{m}, \tag{12}$$

and $R_q = \{i : p_i \le p_{(R_q)}\}$. Then we have

$$\text{max-lfdr}\left(R_q\right) = q \, \frac{(1-\tau)\,\pi_0}{1-F(\tau)}\left(1 - F(\tau)^m\right) \le q.$$

The proof of Theorem 5 is deferred to the Appendix. The method R^* coincides with $R_{\alpha=\hat\pi_0}$, our original procedure applied at the corrected level $\hat q = \alpha/\hat\pi_0$, whenever $\hat q \le \ldots$. Since we usually have $\hat q \le 0.5$, the two methods are identical for all practical purposes.

In the next section, we will investigate the asymptotic regret of methods that estimate π_0. In particular, we will show that this estimation error is asymptotically negligible if it shrinks at a faster rate than $m^{-1/3}$. We can indeed achieve this with $\hat\pi_0$ if f_1 has two continuous derivatives in a neighborhood of 1, with $f_1'(1) = f_1(1) = 0$. By Taylor's theorem, we have

$$1 - F(\lambda) = (1-\lambda)\,\pi_0 + \frac{(1-\pi_0)f_1''(\xi)}{6}(1-\lambda)^3;$$

for some $\xi \in [\lambda; 1]$. Assuming $\pi_0 \in (0;1)$ and taking $\lambda = 1 - m^{-1/5}$, we then have

$$m^{2/5}\left(\hat\pi_0 - \pi_0\right) \approx m^{2/5}\left(\frac{1 + \text{Binom}(m;\, 1-F(\lambda))}{(1-\lambda)m} - \pi_0\right) \xrightarrow{d} N\left(\frac{(1-\pi_0)f_1''(1)}{6};\, \pi_0\right); \tag{13}$$

with subgaussian errors for finite m, so the results in Section 3.3 generally apply. See Genovese and Wasserman (2004) and Patra and Sen (2016) for a discussion of estimators for π_0.

3 Asymptotic regret analysis

In this section, we study our procedure's empirical Bayes regret under the weighted classification risk $\mathbb{E}[L_\lambda(H; R)]$, where the expectation is taken over $H_1; \ldots; H_m$ and $p_1; \ldots; p_m$ according to (1), and L_λ is defined as in (5). Throughout this section we will be considering a sequence of problems with $m \to \infty$.

A fundamental result of Sun and Cai (2007) is that the oracle (6) minimizes the weighted classification risk over all procedures, thus representing a benchmark against which we can compare methods that are feasible without *a priori* knowledge the lfdr. In the empirical Bayes literature (see, e.g., Efron, 2019), the price of our ignorance of the model parameters is measured by the *regret*, or average excess risk, given by the optimality gap

$$\text{Regret}_m(R) := \mathbb{E}\left[L_\lambda(H; R) - L_\lambda(H; R^*)\right]. \tag{14}$$

3.1 Population regret

Before tackling the more delicate problem of calculating the regret for procedures with data-dependent p-value rejection thresholds, we first investigate the regret of fixed-threshold methods. For $t \in [0;1]$, let $R_t^{\text{Fix}} := \{i : p_i \le t\}$, and note that the oracle method is $R^* = R_{\tau}^{\text{Fix}}$. We introduce the function $\psi(t)$ to represent the regret of this method, which is free of m:

$$\psi(t) := \text{Regret}_m(R_t^{\text{Fix}}) = F(\tau) - F(t) - \frac{\pi_0}{\lambda}(\tau - t). \tag{15}$$

Figure 3: Left: The xed-threshold regret (t) (15) with Beta alternatives $f_1(t) =\ t$ as a function of $\ 2\ [0; :5]$. Right: a normalized version $(t)= (0)$, such that BH at level $=\ _0$ has unit normalized regret, identical to the regret of the procedure that rejects nothing. The null proportion is $_0 = 0:8$ and the cost-bene t ratio is $= 4$.

If lfdr( ) = , then we also have $f(\ ) =\ _0 =$, and (t) is simply the error of the rst-order Taylor expansion of F around , also known as the Bregman divergence associated with F. If f is continuously di erentiable between t and , then

$$(t) = \frac{f^0(\ _t)}{2}(t\ \ )^2; \quad \text{for some } _t \text{ between } t \text{ and } : \tag{16}$$

Since F is concave, (t) 0. Finally, we can also rewrite (15) as an integral

$$(t) = \int\ \ 1\ \ ^1 \text{lfdr}(t)\ dF(t): \tag{17}$$

This form for the regret underscores the relationship between the lfdr and the regret, and will prove useful for analyzing the regret with data-dependent thresholds.

We can evaluate to investigate the regret of population versions of our procedure and the BH procedure, i.e. versions of the procedures with rejection thresholds chosen using the true cdf F in place of the empirical cdf F_m. The population BH threshold at an arbitrary level $q\ 2\ (0; 1)$ is found by intersecting F with the ray of slope $q\ ^1$, i.e.

$$t_q^{\text{BH-POP}} := \max f\, t\ 2\ [0; 1] : F(t)\ \ t=q = 0\, g:$$

By comparison, the population version of our procedure $_q$ is

$$t_q := \max\ \ t\ 2\ [0; 1] : f(t)\ \ q\ ^1\ \ ;$$

which coincides with the oracle threshold when $q = \; = \;_0$. Note that t_q is equivalent to the population BH threshold $t^{\text{BH-POP}}_{q^0}$ at the lower level

$$q^0 = \frac{t_q}{F(t_q)}: \tag{18}$$

Thus, there is always some value q^0 for which the BH procedure approximately reproduces the oracle, namely $t_{=\;_0} = F(t_{=\;_0})$, but generally we cannot use it unless we know f_1 and $_0$.

To illustrate the population regret in a concrete example, we consider a parametric alternative distribution

$$f_1(t;\;) \coloneqq t^{\;\;\;1} \quad \text{for some} \;\; 2\;(0;1);$$

which is a Beta(; 1) density. This form is called a Lehmann alternative in the multiple testing literature (see, e.g., Pounds and Morris, 2003). In this case, the population procedures at level $q\;2\;(0;1)$ use rejection thresholds

$$t_q = \left(\frac{q^{\;1}\;\;\;_0}{(1\;\;\;_0)}\right)^{\frac{1}{1}}; \quad \text{and} \quad t^{\text{BH-POP}}_q = \left(\frac{q^{\;1}\;\;\;_0}{1\;\;\;_0}\right)^{\frac{1}{1}}:$$

Furthermore, the threshold equivalence (18) gives

$$q^0 = \frac{q}{1\;\;(1\;\;\;)\;_0 q}\qquad q;$$

where the approximation holds for small values of q. Thus, the correspondence between q and q^0 depends on the parameter , which controls the signal strength under the alternative. For small values of , the signal is very strong, and the \correct" choice of q^0 is much smaller than the desired max-lfdr level , but for weaker signals (larger), we should choose q^0 closer to . Without knowing the signal strength in advance, it is di cult to know at what values of q^0 the BH method will perform well.

In Figure 3 we plot the population regret for various choices of the level of the procedure, $_0 = 0{:}8$ and $= 4$ and varying the parameter . The population version of our procedure at level $\frac{}{_0}$ with $= \frac{1}{1+} = 0{:}2$ is the oracle (6), so it achieves zero regret, while the conservative version of our procedure with $q =$ performs quite well for all values of the alternative parameter . In this example, the asymptotic error incurred from conservatively bounding $_0$ by one in the procedure is small compared to the error incurred by using BH(q^0) at an ad hoc value. The BH procedure at levels $\frac{}{_0}$ or incurs substantial asymptotic regret by comparison. In particular, note that the BH($=\;_0$) procedure incurs the same asymptotic regret as the procedure that rejects nothing; i.e. $(t^{\text{BH-POP}}_{=\;_0}) = (0)$. If we run BH at a lower level like $=2$, $=10$, or $=100$, we can do well for some range of values, but struggle at other parts of the parameter space. No single level for BH dominates in terms of regret, so for the classi cation risk it is more appropriate to view the BH level as a tuning parameter (Neuvial and Roquain, 2012).

3.2 Relationship of our method to the Grenander estimator

Since the marginal density f appears in the denominator of the lfdr, bounding $\pi_0 \le 1$ and plugging in Grenander's estimator $\hat{f}_m$ (defined in (8)) gives the conservative estimate

$$\widehat{\mathrm{lfdr}}(t) := \frac{1}{\hat{f}_m(t)}, \quad t \in [0,1].$$

Similar to how the BH procedure chooses an interval $[0,t]$ as large as possible subject to a constraint on an estimate of the FDP, the rejection threshold of the SL procedure can be equivalently expressed as

$$\tau_q = \operatorname*{argmax}_{p_{(0)},\ldots,p_{(m)}} \left\{ \frac{qk}{m} - p_{(k)} \right\} = \sup \left\{ t \in [0,1] : \widehat{\mathrm{lfdr}}(t) \le q \right\}, \tag{19}$$

taking the convention that $\sup \emptyset = 0$. The equivalence in (19) is illustrated in Figure 4. Let $\hat{F}_m$ denote the least concave majorant of the empirical cdf $\mathbb{F}_m$, plotted as a dotted blue line in the left panel of Figure 4. By definition of $\widehat{\mathrm{lfdr}}(t)$, the supremum on the right hand side is equal to the largest t for which $\frac{d}{dt}(q\hat{F}_m(t) - t) = q\hat{f}_m(t) - 1 \ge 0$, which corresponds to the maximizer of the function $q\hat{F}_m(t) - t$, illustrated for example in the right panel of Figure 4. $\hat{F}_m \ge \mathbb{F}_m$ implies

$$q\hat{F}_m(t) - t \ge q\mathbb{F}_m(t) - t, \quad t \in [0,1],$$

with equality at the knots of $\hat{F}_m$, and since the maximizer of the left hand side occurs at a knot of $\hat{F}_m$, it is also the maximizer of the right hand side, i.e. the argmax of $\frac{qk}{m} - p_{(k)}$.

We can again compare this result with the BH(q) threshold, given by

$$\tau_q^{BH} = \max_{k=0,\ldots,m} \left\{ p_{(k)} : \frac{qk}{m} - p_{(k)} \ge 0 \right\} = \sup \left\{ t \in [0,1] : \mathbb{F}_m(t) \ge q^{-1}t \right\},$$

which is the largest t for which the ray $q^{-1}t$ lies below the ecdf $\mathbb{F}_m(t)$. Our procedure instead finds the last intersection of the graph of $\mathbb{F}_m$ with a support line of slope q^{-1}, since

$$\widehat{\mathrm{lfdr}}(t) \le q \iff \hat{f}_m(t) \ge q^{-1}.$$

This relationship is illustrated in the left panel of Figure 4.

3.3 Asymptotic behavior of our procedure

Equation (16) suggests that, when f is sufficiently regular near τ^*, the regret is closely related to the squared error of the rejection threshold. Our main result in this section establishes cube-root asymptotics for the behavior of our procedure R_q with $q = \alpha/\hat{\pi}_0$, where $\hat{\pi}_0$ consistently estimates π_0; if π_0 is known, then the results apply directly with $\hat{\pi}_0 = \pi_0$.

We derive limiting distributions for the threshold τ_q, the lfdr at the threshold, and the regret of R_q. All three are given in terms of Chernoff's distribution (Chernoff, 1964), which is

Figure 4: Left: empirical cdf F_m and its least concave majorant $\hat{F}_m$. The support line of slope q^{-1} touches both curves at the decision threshold τ_q. Right: the same plot with the line t/q subtracted off.

defined as the distribution of the maximizer Z of a standard two-sided Brownian motion $W = (W(t))_{t \in \mathbb{R}}$ with parabolic drift:

$$Z = \operatorname*{argmax}_{t \in \mathbb{R}} \, W(t) - t^2. \tag{20}$$

The random variable Z has a density with respect to the Lebesgue measure on $\mathbb{R}$ that is symmetric about zero. Dykstra and Carolan (1999) suggest approximating the density and cdf of Z by those of $N\left(0, (.52)^2\right)$. This approximation can be somewhat crude but gives a rough sense for the distribution of Z. Groeneboom and Wellner (2001) provide much more accurate numerical methods to compute the density, cdf, quantiles and moments of Z.

Theorem 6. Suppose $p_1, \ldots, p_m$ follow the Bayesian two-groups model (1), with $\pi_0 \in (0, 1)$, $f_0 = 1_{[0,1]}$, and f_1 non-increasing. For $q \in (0, \pi_0^{-1})$, assume additionally that

(i) there is a unique value $t_q \in (0, 1)$ for which $f(t_q) = q^{-1}$,

(ii) f is continuously differentiable in a neighborhood of t_q with $f'(t_q) < 0$, and

(iii) $\hat{q}$ is any random variable with $m^{1/3}(\hat{q} - q) \xrightarrow{p} 0$ as $m \to \infty$.

Then we have, as $m \to \infty$,

$$m^{1/3}(\tau_{\hat{q}} - t_q) \xrightarrow{d} \left(\frac{q}{4} f'(t_q)^2\right)^{-1/3} Z; \quad \text{and} \tag{21}$$

$$m^{1/3} \left(\frac{\mathrm{lfdr}(\tau_{\hat{q}}) - \pi_0 q}{\pi_0 q}\right) \xrightarrow{d} \left(4q^2 \, |f'(t_q)|\right)^{1/3} Z. \tag{22}$$

where Z follows Chernoff's distribution defined in (20). Further, suppose that

$$P\{m^{1/3}|\hat{\tau}_q - \tau_q| > \epsilon\} = o\left(m^{-2/3}\right), \quad \text{for all } \epsilon > 0. \tag{23}$$

Then we also have $m^{1/3}E[\hat{\tau}_q] \to t_q$. In addition,

$$m^{2/3}\mathrm{Var}(\hat{\tau}_q) \to \left(\frac{q}{4} f'(t_q)^2\right)^{-2/3}\mathrm{Var}(Z); \quad \text{and} \tag{24}$$

$$m^{2/3}\mathrm{Var}\left(\frac{\mathrm{lfdr}(\hat{\tau}_q) - \pi_0 q}{\pi_0 q}\right) \to \left(4q^2 \, |f'(t_q)|\right)^{2/3}\mathrm{Var}(Z); \tag{25}$$

where $\mathrm{Var}(Z) \approx 0.26$.

The proof of Theorem 6 is deferred to Appendix B. It is well-known that the Grenander estimator $\hat{f}_m$ estimates f at a cube root rate pointwise, away from zero, but this result, due to (Rao, 1969), is too weak to describe the behavior of our procedure. We rely on a stronger version of this result due to Dümbgen et al. (2016) that approximates the local behavior of the Grenander estimator near t_q.

The distributional result (22) complements our result from Theorem 1, by showing that $\mathrm{lfdr}(\tau_q) = \max_{i \in R_q} \mathrm{lfdr}(p_i)$ is not only controlled in expectation, but also concentrates at rate $m^{-1/3}$ around its expectation. In particular, because $P\{Z \geq 1\} \approx 0.05$, we have

$$\frac{\mathrm{lfdr}(\tau_q) - \pi_0 q}{\pi_0 q} \leq m^{-1/3}\left(4q^2 \, |f'(t_q)|\right)^{1/3},$$

with roughly 95% probability in large samples. For example, suppose we use $q = 0.2$, so $f(t_q) = 5$, and suppose that $f'(t_q) = -50$. Then, whereas Theorem 1 guarantees $E[\mathrm{lfdr}(\tau_q)] \leq 0.2$ exactly, the asymptotic estimate from Theorem 6 bounds the 95th percentile of $\mathrm{lfdr}(\tau_q)$ at 0.24 if $m = 1000$, or at 0.21 if $m = 64{,}000$.

To understand why the error is of order $m^{-1/3}$, consider fixed q and recall that the threshold τ_q maximizes the stochastic process

$$U(t) := F_m(t) - F_m(t_q) - \frac{t - t_q}{q}.$$

Because $f(t_q) = q^{-1}$, we have for t near t_q,

$$F(t) - F(t_q) \approx \frac{t - t_q}{q} + \frac{f'(t_q)}{2}(t - t_q)^2.$$

Introducing the local parameterization $t = t_q + m^{-a}h$ for $a > 0$ leads to

$$U(t_q + m^{-a}h) \approx -\frac{|f'(t_q)|}{2} \cdot \frac{h^2}{m^{2a}} + N\left(0, \frac{h}{qm^{a+1}}\right).$$

Setting $a = 1/3$ balances the mean and variance, giving

$$m^{2/3}U(t_q + m^{-1/3}h) \xrightarrow{d} -\frac{|f'(t_q)|}{2}h^2 + N\left(0, \frac{h}{q}\right).$$

Under this local scaling, $U(t)$ converges to a Brownian motion with parabolic drift, and its maximizer τ_q converges to Chernoff's distribution. Theorem 6 applies a more careful version of this argument, replacing $F_m(t)$ with its LCM $\hat{F}_m(t)$ and using a result of Dümbgen et al. (2016) to characterize the process $\hat{f}_m(t)$ under the same local scaling. The corresponding results for $\mathrm{lfdr}(\tau_q)$ follow from first-order Taylor expansion of $\mathrm{lfdr}(t) = \pi_0/f(t)$ around t_q.

By specializing Theorem 6 to $q = \alpha = \pi_0$ and $\hat{q} = \hat{\alpha} = \hat{\pi}_0$, we obtain the limiting regret for our procedure with a known or accurately estimated null proportion.

Theorem 7. Suppose $p_1, \ldots, p_m$ follow the Bayesian two-groups model (1), with $\pi_0 \in (0,1)$, $f_0 = 1_{[0,1]}$, and f_1 non-increasing. Assume additionally that

(i) there is a unique value $\tau^* \in (0,1)$ for which $\mathrm{lfdr}(\tau^*) = \frac{\pi_0}{f(\tau^*)} = \lambda$,

(ii) f is continuously differentiable in a neighborhood of τ^* with $f'(\tau^*) < 0$, and

(iii) $\hat{\pi}_0$ is any estimator of π_0 with $P\left(m^{1/3}(\hat{\pi}_0 - \pi_0) > \varepsilon\right) = o\left(m^{-2/3}\right)$ for all $\varepsilon > 0$.

Then we have, as $m \to \infty$,

$$m^{2/3}\,\mathrm{Regret}_m(R_{\alpha = \hat{\pi}_0}) \to \frac{\lambda^2}{2\pi_0^2}\,|f'(\tau^*)|^{1/3}\,\mathrm{Var}(Z); \tag{26}$$

where Z follows Chernoff's distribution defined in (20), and $\mathrm{Var}(Z) \approx 0.26$.

Theorems 6–7 deal with the regret for $\pi_0 \in (0,1)$. Under the global null, represented in the Bayesian model by $\pi_0 = 1$, the behavior is different and the regret is simply $\alpha\,\mathrm{EV}$, which is $O(m^{-1})$, as we see next.

Proposition 8. Suppose $(p_i)_{i=1}^m$ follow a two-groups model (1) with $f_0 = 1_{[0,1]}$ and $\pi_0 = 1$, i.e. $H_i = 0$ for all i and $p_i \overset{iid}{\sim} \mathrm{Unif}(0,1)$. Then as $m \to \infty$, we have

$$m\,\mathrm{Regret}_m(R_q) \to \sum_{k=1}^{\infty} P\{U_k \le q\}; \qquad \text{for } U_k \sim \mathrm{Gamma}(k,k);$$

which is finite for every $q \in [0,1)$.

Proposition 8 is closely related to results derived in Finner and Roters (2001).

4 Numerical results

This section highlights our main results on simulation experiments. We adapt a simulation setting of Benjamini and Hochberg (1995) to the two-groups model (1). Specifically, define the alternative density

$$f_1(t) = \frac{\frac{1}{4}\sum_{i=1}^{4} \phi\left(\Phi^{-1}(t) - \frac{5i}{4}\right)}{\phi\left(\Phi^{-1}(t)\right)} \qquad \text{for } 0 \le t \le 1; \tag{27}$$

Figure 5: Above: Mixture density f (left) and lfdr (right), with alternative density f_1 defined in (27) and null proportion $\pi_0 = 0.75$. Note f_1 diverges as $t \downarrow 0$. Below: Comparison of FDR control (left) and max-lfdr control (right) on simulated data. The estimate of the null proportion is (11) with $\lambda = 0.5$.

where ϕ and $\bar{\Phi}$ denote the density and survival function of the standard Gaussian distribution. Concretely, a non-null p-value $p_i \sim f_1$ can be constructed by first taking $Y_i \sim N(\mu_i; 1)$ where μ_i is drawn uniformly at random from the set $\{5\frac{i}{4} : i = 1; 2; 3; 4\}$; then, $p_i = \bar{\Phi}(Y_i)$ is a one sided p-value for the null-hypothesis that $\mu_i = 0$. We use a null proportion of $\pi_0 = 0.75$. Figure 5 shows the mixture density and corresponding lfdr.

We repeatedly sampled from the above two-groups model with $m = 64$ hypotheses. Figure 5 shows the FDR (left panel) and max-lfdr (right panel) for both our procedure and the BH procedure, at conservative level $q =$ and estimated level $\hat{q} = \; = \hat{}_0$. The BH procedure, shown in red, achieves FDR exactly $_0 q$, whereas the max-lfdr can be much larger. By contrast, our procedure, shown in blue, conservatively controls FDR substantially below the level $_0 q$ but has max-lfdr equal to $_0 q$.

Figure 6 highlights some features of the regret of our procedure present in Theorem 7. The left panel shows a log-log plot of the regret as a function of the sample size m. The red curve shows the regret of our uncorrected procedure $\hat{R}$ for $= 0{:}05$, which asymptotically tends to (t) and hence asymptotically incurs some non-vanishing regret described in Section 3.1. The blue curve shows the regret of the corrected procedure $\hat{R}_{= \; 0}$ with known $_0$. For larger samples, the simulated regret closely matches the asymptotic prediction $m^{2=3} \frac{2}{2 \; _0^2} \; jf^0(\;)j^{1=3} EZ^2$ from (26), shown in black. The green curve (which is nearly indistinguishable from the blue curve) shows the corrected procedure with an estimated null proportion $\hat{}_0$ based on (11) with $= 1 \; m^{1=5}$. The right panel of Figure 6 confirms that the asymptotic distribution of the conditional expectation $E\left[L(H; \hat{R}_{= \hat{}_0}) \; L(H; R) \; j \; F_m\right]$ closely matches the theoretical prediction.

Figure 6: Simulation results demonstrating some features of Theorem ?? with the alternative density f_1 defined in (27), cost-benefit ratio $= 19$ and null proportion $_0 = 0{:}75$. Left: a log-log plot of the regret (14) as a function of the sample size. The black line shows the asymptotic prediction (26). Right: a comparison of the empirical quantiles of the conditional expectation of $E\left[L(H; \hat{R}_{= \hat{}_0(\;_m)}) \; L(H; R) \; j \; F_m\right]$, scaled so the quantity tends to Chernoff's distribution, for $m = 10^6$. Groeneboom and Wellner (2001) provide quantiles of Chernoff's distribution.

5 Discussion

In this work we have introduced a new error criterion, the max-lfdr, which modifies the FDR by redirecting attention away from the average quality of the rejection set and toward the rejections that are close to the rejection boundary. Despite the seeming difficulty of measuring the quality of a single rejection, we also introduce a simple new multiple testing procedure that controls the max-lfdr at level $\pi_0 q$ in finite samples, where q is a tuning parameter and π_0 is the null proportion. We assume only that the data follow a Bayesian two-groups model in which smaller p-values reflect stronger evidence against the null. We find that our method is better able than the BH method to adapt to the unknown problem structure, and to perform well without knowledge of the true underlying distribution.

The BH procedure owes its enduring utility for FDR control in part to its versatility beyond this basic setting, however. It is known to still control FDR, for instance, when the null p-values are super-uniform and under certain forms of positive dependence, two of many possible extensions that we leave open for our procedure.

Another seeming advantage of the FDR criterion is that it requires no Bayesian assumptions, whereas the max-lfdr is only defined with reference to a Bayesian model. A possible avenue for generalizing the max-lfdr to frequentist settings is to work with its characterization as the probability that the last rejection is a false discovery. Indeed, our proof of Theorem 1 implies that max-lfdr is controlled even conditional on $H_1, \ldots, H_m$. This is initially puzzling: if each H_i is fixed, then how can we speak of the probability that the last rejection is a false discovery? The answer is that $H_{(R)}$ is random even if $H_1, \ldots, H_m$ are fixed, since its index is random. We leave further development of the frequentist connection to the max-lfdr to future work.

Acknowledgements

We are indebted to Stephen Bates, Aditya Guntuboyina, Michael I. Jordan, Peter McCullagh and Jim Pitman for helpful discussions. J. A. Soloff was supported by NSF Grant DMS-2023505 and by the Office of Naval Research under the Vannevar Bush Fellowship. William Fithian was supported by the NSF DMS-1916220 and a Hellman Fellowship from Berkeley.

References

Aubert, J., Bar-Hen, A., Daudin, J.-J. and Robin, S. (2004). Determination of the differentially expressed genes in microarray experiments using local fdr, BMC bioinformatics 5(1): 1–9.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing, Journal of the Royal statistical society: series B (Methodological) 57(1): 289–300.

Cherno , H. (1964). Estimation of the mode, Annals of the Institute of Statistical Mathematics 16(1): 31{41.

Dümbgen, L., Wellner, J. A. and Wol , M. (2016). A law of the iterated logarithm for Grenander's estimator, Stochastic processes and their applications 126(12): 3854{3864.

Dykstra, R. and Carolan, C. (1999). The distribution of the argmax of two-sided brownian motion with quadratic drift, Journal of Statistical Computation and Simulation 63(1): 47{58.

Efron, B. (2004). Large-scale simultaneous hypothesis testing: the choice of a null hypothesis, Journal of the American Statistical Association 99(465): 96{104.

Efron, B. (2008). Microarrays, empirical Bayes and the two-groups model, Statistical science pp. 1{22.

Efron, B. (2019). Bayes, oracle Bayes and empirical Bayes, Statistical Science 34(2): 177{201.

Efron, B., Tibshirani, R., Storey, J. D. and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment, Journal of the American statistical association 96(456): 1151{1160.

Finner, H. and Roters, M. (2001). On the false discovery rate and expected type I errors, Biometrical Journal 43(8): 985{1005.

Genovese, C. and Wasserman, L. (2004). A stochastic process approach to false discovery control, The annals of statistics 32(3): 1035{1061.

Grenander, U. (1956). On the theory of mortality measurement: Part II, Scandinavian Actuarial Journal 1956(2): 125{153.

Groeneboom, P. and Jongbloed, G. (2014). Nonparametric estimation under shape constraints, Vol. 38, Cambridge University Press.

Groeneboom, P. and Wellner, J. A. (2001). Computing Cherno 's distribution, Journal of Computational and Graphical Statistics 10(2): 388{400.

Kallenberg, O. (2002). Foundations of modern probability, 2 edn, Springer.

Langaas, M., Lindqvist, B. H. and Ferkingstad, E. (2005). Estimating the proportion of true null hypotheses, with application to DNA microarray data, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67(4): 555{572.

Liao, J., Lin, Y., Selvanayagam, Z. E. and Shih, W. J. (2004). A mixture model for estimating the local false discovery rate in DNA microarray analysis, Bioinformatics 20(16): 2694{2701.

Muralidharan, O. (2010). An empirical Bayes mixture method for e ect size and false discovery rate estimation, The Annals of Applied Statistics pp. 422{438.

Neuvial, P. and Roquain, E. (2012). On false discovery rate thresholding for classi cation under sparsity, The Annals of Statistics 40(5): 2572{2600.

Patra, R. K. and Sen, B. (2016). Estimation of a two-component mixture model with applications to multiple testing, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78(4): 869{893.

Pounds, S. and Cheng, C. (2004). Improving false discovery rate estimation, Bioinformatics 20(11): 1737{1745.

Pounds, S. and Morris, S. W. (2003). Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values, Bioinformatics 19(10): 1236{1242.

Rao, B. P. (1969). Estimation of a unimodal density, Sankhya: The Indian Journal of Statistics, Series A pp. 23{36.

Reiner, A., Yekutieli, D. and Benjamini, Y. (2003). Identifying di erentially expressed genes using false discovery rate controlling procedures, Bioinformatics 19(3): 368{375.

Robbins, H. (1951). Asymptotically subminimax solutions of compound statistical decision problems, Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, The Regents of the University of California.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). Order restricted statistical inference, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Ltd., Chichester.

Robin, S., Bar-Hen, A., Daudin, J.-J. and Pierre, L. (2007). A semi-parametric approach for mixture models: Application to local false discovery rate estimation, Computational statistics & data analysis 51(12): 5483{5493.

Scheid, S. and Spang, R. (2004). A stochastic downhill search algorithm for estimating the local false discovery rate, IEEE/ACM Transactions on Computational Biology and Bioinformatics 1(3): 98{108.

Schweder, T. and Spj tvoll, E. (1982). Plots of p-values to evaluate many tests simultaneously, Biometrika 69(3): 493{502.

Shorack, G. R. and Wellner, J. A. (2009). Empirical processes with applications to statistics, SIAM.

Stephens, M. (2017). False discovery rates: a new deal, Biostatistics 18(2): 275{294.

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

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

Strimmer, K. (2008). A unified approach to false discovery rate estimation, BMC bioinformatics 9(1): 1–14.

Sun, W. and Cai, T. T. (2007). Oracle and adaptive compound decision rules for false discovery rate control, Journal of the American Statistical Association 102(479): 901–912.

Takacs, L. (1967). On combinatorial methods in the theory of stochastic processes, in L. M. Le Cam and J. Neyman (eds), Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 3, University of California Press, pp. 431–447.

A Proofs of results from Section 2

Proof of Theorem 5. As in the proof of Theorem 1, we have

$$\text{max-lfdr}\left(R_q\right) = P\left\{H_{(R_q)} = 0;\, R_q > 0\right\} = m\,P\left\{H_m = 0;\, p_{(R_q)} = p_m\right\}.$$

Define the σ-field $\mathcal{F} = \sigma(p_1, \ldots, p_{m-1}; H_m; 1\{p_m \le \alpha\})$. We restrict our attention to the event $A = \{H_m = 0; p_m \le \alpha\}$, since the event $\{H_m = 0;\, p_{(R_q)} = p_m\}$ cannot occur except on A. On A, which is $\mathcal{F}$-measurable, we have $p_m/\alpha \mid \mathcal{F} \sim U[0,1]$.

Let $m_\alpha = \#\{i : p_i \le \alpha\}$, which is also $\mathcal{F}$-measurable. If $j_1 \le \cdots \le j_{m_\alpha} = m$ are the indices of the p-values that are below α, define the modified p-values $\tilde{p}_i = p_{j_i}/\alpha$, for $i = 1, \ldots, m_\alpha$. Because the order statistics of $\tilde{p}_1, \ldots, \tilde{p}_{m_\alpha}$ are also the first m_α order statistics of $p_1, \ldots, p_m$, the quantity R_q defined in (12) can be rewritten as

$$\begin{aligned} R_q &= \underset{k=0,\ldots,m}{\operatorname{argmin}}\; p_{(k)} \le \frac{q}{\hat{\pi}_0}\,\frac{k}{m} \\ &= \underset{k=0,\ldots,m_\alpha}{\operatorname{argmin}}\; \tilde{p}_{(k)} \le \frac{\tilde{q}\,k}{m_\alpha}; \quad \text{for } \tilde{q} = \frac{q\,m_\alpha}{\alpha\,\hat{\pi}_0 m}. \end{aligned}$$

Applying Lemma 2, we have

$$P\left\{H_m = 0;\, p_{(R_q)} = p_m \mid \mathcal{F}\right\} = \frac{\tilde{q}}{m_\alpha}\,1_A = \frac{q}{\alpha\,\hat{\pi}_0 m}\,1_A$$

Marginalizing over F, and noting that $P(A) = \pi_0$, we obtain

$$
\begin{aligned}
P\left\{ H_m = 0;\ p_{(R_q)} = p_m \right\} &= \frac{q}{m} E\left[\frac{\pi_0}{\hat{\pi}_0} \,\middle|\, A \right] \\
&= \frac{q}{m} \frac{(1-\gamma)\pi_0}{1-F(\gamma)} E\left[\frac{(1-F(\gamma))m}{1 + \#\{i < m : p_i > \gamma\}} \right] \\
&= \frac{q}{m} \frac{(1-\gamma)\pi_0}{1-F(\gamma)} (1 - F(\gamma)^m) \\
&\le \frac{q}{m},
\end{aligned}
$$

completing the proof. The final inequality is a standard binomial identity:

$$
\begin{aligned}
E\left[\frac{m}{1 + \mathrm{Binom}(m-1, \theta)} \right] &= \sum_{k=0}^{m-1} \frac{m}{1+k} \binom{m-1}{k} \theta^k (1-\theta)^{m-1-k} \\
&= \sum_{k=0}^{m-1} \binom{m}{k+1} \theta^{k+1} (1-\theta)^{m-(k+1)} \\
&= \sum_{j=1}^{m} \binom{m}{j} \theta^j (1-\theta)^{m-j} \\
&= P\{\mathrm{Binom}(m, \theta) \ge 1\} \\
&= 1 - (1-\theta)^m. \qquad \square
\end{aligned}
$$

B Proofs of results from Section 3

Proof of Theorem 6. Our proof will use the switching relation that states, for any $t \in (0, 1)$, we have almost surely

$$\hat{\tau}_q \le t \iff \hat{f}_m(t) \le \hat{q}^{-1}.$$

We will work with a local expansion of $\hat{f}_m(t)$ around t_q using the local parameterization $t = t_q + m^{-1/3}h$. Using $f(t_q) = q^{-1}$, the switching relation becomes

$$m^{1/3}(\hat{\tau}_q - t_q) \le h \iff \hat{f}_m\left(t_q + m^{-1/3}h\right) - f(t_q) \le \hat{q}^{-1} - q^{-1}.$$

Now let W denote a standard two-sided Brownian motion, and let $S_{a,b}$ denote the process of left derivatives of the least concave majorant of $X_{a,b}(t) = aW(t) - bt^2$, where $a = \sqrt{f(t_q)}$ and $b = |f'(t_q)|/2$. Under our regularity assumptions, Dümbgen et al. (2016) show

$$m^{1/3}\left(\hat{f}_m\left(t_q + m^{-1/3}h\right) - f(t_q)\right) \Rightarrow S_{a,b}(h)$$

in the Skorokhod topology on $D[-K, K]$ for every finite $K > 0$. Since $m^{1/3}(\hat{q}^{-1} - q^{-1}) \xrightarrow{p} 0$ by assumption, we have

$$P\left\{ m^{1/3}(\hat{\tau}_q - t_q) \le h \right\} \to P\{S_{a,b}(h) \le 0\}.$$

Observe that $\mathsf{S}_{a,b}(h) \geq 0$ iff $t_{a,b} \leq h$, where $t_{a,b}$ is the (a.s. unique) maximizer of $X_{a,b}$ (note the maximizer $t_{a,b}$ is always a knot in the concave majorant since the horizontal line with intercept $X_{a,b}(t_{a,b})$ is a supporting line intersecting $(t_{a,b}, X_{a,b}(t_{a,b}))$). Combining this observation with the previous display, we have

$$m^{1/3}(\tau_{\hat{q}} - t_q) \xrightarrow{d} t_{a,b} \overset{d}{=} (b/a)^{-2/3} Z = \left(\frac{q}{4} f'(t_q)^2\right)^{-1/3} Z,$$

proving (21). Next we turn to the lfdr asymptotics. By Taylor's theorem,

$$m^{1/3}(\mathrm{lfdr}(\tau_{\hat{q}}) - \pi_0 q) = \mathrm{lfdr}'(\omega)\, m^{1/3}(\tau_{\hat{q}} - t_q)$$

for some ω between $\tau_{\hat{q}}$ and t_q. Using

$$\mathrm{lfdr}'(t_q) = -\frac{\pi_0 f'(t_q)}{f(t_q)^2} = \pi_0 q^2 |f'(t_q)|,$$

and applying the continuous mapping theorem and Slutsky's theorem, we obtain

$$\mathrm{lfdr}'(\omega)\, m^{1/3}(\tau_{\hat{q}} - t_q) \xrightarrow{d} \mathrm{lfdr}'(t_q) \left(\frac{q}{4} f'(t_q)^2\right)^{-1/3} Z = \pi_0 q \left(4q^2 |f'(t_q)|\right)^{1/3} Z,$$

proving (22). Next, under the strengthened assumption (23), fix $\varepsilon > 0$ and define the event

$$A_\varepsilon = \left\{ |\hat{q} - q| \leq m^{-1/3}\varepsilon,\ |\tau_{\hat{q}} - t_q| \leq m^{-2/9} \right\}, \tag{28}$$

and the truncated random variable

$$Y_m = m^{1/3}(\tau_{\hat{q}} - t_q)\, 1_{A_\varepsilon},$$

We will show that $\mathsf{P}(A_\varepsilon^c) = o\left(m^{-2/3}\right)$. As a result, Y_m has the same limit in distribution as $m^{1/3}(\tau_{\hat{q}} - t_q)$. If we can show that the sequence Y_m^2 is uniformly integrable, we will have convergence of its mean and variance to the mean and variance of its limiting distribution. Then, because

$$\mathsf{E}\left[\left(m^{1/3}(\tau_{\hat{q}} - t_q) - Y_m\right)^2\right] \leq m^{2/3}\mathsf{P}(A_\varepsilon^c) \to 0,$$

we will have the same limiting mean and variance for $m^{1/3}(\tau_{\hat{q}} - t_q)$.

To show that $\mathsf{P}(A_\varepsilon^c) = o\left(m^{-2/3}\right)$, let $q_1 = q - m^{-1/3}\varepsilon$ and $q_2 = q + m^{-1/3}\varepsilon$ and assume that m is sufficiently large that $m^{-1/3}\varepsilon \leq m^{-2/9}/2$, and

$$f'(t) \leq f'(t_q)/2, \quad \text{for all } t \in [t_q - m^{-2/9},\ t_q + m^{-2/9}].$$

As a result, for all $t \geq t_{q_2} + m^{-2/9}/2$, we have

$$\begin{aligned} F(t) - F(t_{q_2}) - \frac{t - t_{q_2}}{q_2} &\leq F(t_{q_2} + m^{-2/9}/2) - F(t_{q_2}) - \frac{m^{-2/9}}{2q_2} \\ &\leq \frac{f'(t_q)}{16} m^{-4/9} \end{aligned}$$

Then, since $\tau_{\hat q} \ge \tau_{q_2}$ a.s. on A_ε, we have

$$\begin{aligned}
\mathsf{P}\left\{\tau_{\hat q} > t_q + m^{-2/9},\ A_\varepsilon\right\} &\le \mathsf{P}\left\{\tau_{q_2} > t_{q_2} + m^{-2/9}/2\right\} \\
&\le \mathsf{P}\left(\sup_{t \ge t_{q_2} + m^{-2/9}/2} F_m(t) - F_m(t_{q_2}) - \frac{t - t_{q_2}}{q_2} \ge 0 \right) \\
&\le \mathsf{P}\left(\sup_{t \ge t_{q_2} + m^{-2/9}/2} F_m(t) - F(t) - (F_m(t_{q_2}) - F(t_{q_2})) \ge \frac{|f'(t_q)|}{16} m^{-4/9} \right) \\
&\le \mathsf{P}\left(\sup_{t \in [0,1]} |F_m(t) - F(t)| \ge \frac{|f'(t_q)|}{32} m^{-4/9} \right) \\
&\le C_{\mathrm{DKW}} \exp\left(-\frac{f'(t_q)^2}{512} m^{1/9} \right),
\end{aligned}$$

where C_{DKW} is the constant for the Dvoretzky–Kiefer–Wolfowitz inequality. An analogous argument yields the same bound for $\mathsf{P}\{\tau_{\hat q} < t_q - m^{-2/9}\}$. $\square$

Proof of Theorem 7. Define $q = \alpha/\pi_0$ and $\hat q = \alpha/\hat\pi_0$, and let $\Delta \subseteq \{1, \ldots, m\}$ denote the symmetric difference between the two rejection sets:

$$\Delta = \begin{cases} \{R_{\hat q} + 1, \ldots, R^*\} & \text{if } R_{\hat q} < R^* \\ \{R^* + 1, \ldots, R_{\hat q}\} & \text{if } R_{\hat q} > R^* \\ \emptyset & \text{if } R_{\hat q} = R^* \end{cases}.$$

Then we have

$$L^*(H, R_{\hat q}) - L^*(H, R^*) = \frac{1}{m}\left(R^* - R_{\hat q} + \frac{\mathrm{sgn}(R_{\hat q} - R^*)}{\alpha} \sum_{i \in \Delta} (1 - H_i) \right).$$

Conditional on F_m, we have $H_i \stackrel{\text{ind}}{\sim} \mathrm{Bern}(1 - \mathrm{lfdr}(p_{(i)}))$, giving conditional expectation

$$\begin{aligned}
\Gamma_m &:= \mathsf{E}\left[L^*(H, R_{\hat q}) - L^*(H, R^*) \mid F_m \right] \\
&= \frac{1}{m}\left(R^* - R_{\hat q} + \frac{\mathrm{sgn}(R_{\hat q} - R^*)}{\alpha} \sum_{i \in \Delta} \mathrm{lfdr}(p_{(i)}) \right) \\
&= \int_{\tau_{\hat q}}^{\tau^*} 1 - \alpha^{-1}\mathrm{lfdr}(t) \, dF_m(t) \\
&= \rho(\tau_{\hat q}) + \alpha^{-1} \int_{\tau_{\hat q}}^{\tau^*} (\alpha - \mathrm{lfdr}(u)) \left(dF_m(u) - dF(u) \right)
\end{aligned}$$

Define the same truncation event A_ε as in (28).

$$A_\varepsilon = \left\{ |\hat q - q^*| \le m^{-1/3}\varepsilon,\ |\tau_{\hat q} - \tau^*| \le m^{-2/9} \right\}.$$

Then, because $|\Gamma_m| \le \alpha^{-1}$ we have

$$\begin{aligned} \text{Regret}_m(R_{\hat q}) &\le \mathsf{E}\left[\rho(\tau_{\hat q}) 1_{A_\varepsilon}\right] \\ &\quad - \alpha^{-1}\mathsf{E}\left[\int_{\hat q} (\alpha - \text{lfdr}(u))\,(dF_m(u) - dF(u))\, 1_{A_\varepsilon} \right] + \alpha^{-1}\mathsf{P}\left(A_\varepsilon^c\right). \end{aligned} \tag{29}$$

We showed in the proof of Theorem 6 that $\mathsf{P}\left(A_\varepsilon^c\right) = o\left(m^{-2/3}\right)$. Furthermore,

$$\begin{aligned} m^{2/3}\mathsf{E}\left[\rho(\tau_{\hat q})1_{A_\varepsilon}\right] &= \mathsf{E}\left[\frac{f'(\xi_{\hat q})}{2}\, m^{2/3}(\tau_{\hat q} - \tau^*)^2\, 1_{A_\varepsilon} \right] \\ &\to \frac{f'(\tau^*)}{2} \left(\frac{\alpha}{4\pi_0} f'(\tau^*)^2 \right)^{-2/3} \text{Var}(Z) \\ &= \left(\frac{\alpha^2}{2\pi_0^2} |f'(\tau^*)| \right)^{-1/3} \text{Var}(Z), \end{aligned}$$

where we have used the fact that $f'(\xi_{\hat q})$ is uniformly close to $f'(\tau^*)$ on A_ε. It remains only to show that the first term on the right-hand side of (29) is $o\left(m^{-2/3}\right)$.

□

Proof of Proposition 8. Since $H_i = 0$ for all i

$$L_\lambda(H, \hat R_\alpha) - L_\lambda(H, R^{\text{OPT}}) = \frac{\lambda \hat R_\alpha}{m}.$$

Recall $\hat R_\alpha$ is the argmax of the random walk $k \mapsto \alpha\frac{k}{m} - p_{(k)}$, which has exchangeable increments. We will use Corollary 11.14 of Kallenberg (2002), due to Sparre-Andersen, that, by exchangeability, the number of rejections $\hat R_\alpha$ is equal in distribution to the time the walk stays positive:

$$\hat R_\alpha \overset{d}{=} P_\alpha := \sum_{k=1}^{m} 1\left\{ p_{(k)} \le \alpha \frac{k}{m} \right\}.$$

Under the global null, the regret thus has mean

$$\begin{aligned} m\mathsf{E}\left[L_\lambda(H, \hat R_\alpha) - L_\lambda(H, R^{\text{OPT}}) \right] = \lambda \mathsf{E}\hat R_\alpha &= \lambda \sum_{k=1}^{m} \mathsf{P}\left\{ p_{(k)} \le \alpha \frac{k}{m} \right\} \\ &\to \lambda \sum_{k=1}^{\infty} \mathsf{P}_{U_k \sim \text{Gamma}(k,k)}\left\{ U_k \le \alpha \right\}, \end{aligned}$$

where the last step follows from the law of rare events.

□