Appendix to "The Power of Batching in Multiple Hypothesis Testing"

Tijana Zrnic
UC Berkeley

Daniel L. Jiang
Amazon

Aaditya Ramdas
CMU

Michael I. Jordan
UC Berkeley

1 ONLINE BATCH FDR CONTROL UNDER POSITIVE DEPENDENCE

The guarantees of $\mathrm{Batch}_{\mathrm{BH}}$ and $\mathrm{Batch}_{\mathrm{St\text{-}BH}}$ presented thus far relied on independence between p-values. In this section we generalize $\mathrm{Batch}_{\mathrm{BH}}$ to one natural form of dependence, namely *positive dependence* (Benjamini and Yekutieli, 2001). We call this modification $\mathrm{Batch}_{\mathrm{BH}}^{\mathrm{PRDS}}$, and it controls FDR when the p-values in one batch are positively dependent, and independent across batches. Such a setting might occur in multi-armed clinical trials where different treatments are tested against a common control arm (Robertson and Wason, 2018).

First we establish the definition of positive dependence we consider.

Definition 3. Let $\mathcal{D} \subseteq [0,1]^n$ be any non-decreasing set, meaning that $x \in \mathcal{D}$ implies $y \in \mathcal{D}$, for all y such that $y_i \geq x_i$, for all $i \in [n]$. We say that a vector of p-values $\mathbf{P} = (P_1, \ldots, P_n)$ satisfies positive regression dependency on a subset (PRDS), or positive dependence for short, if for any null index $i \in \mathcal{H}^0$ and arbitrary non-decreasing set $\mathcal{D} \subseteq [0,1]^n$, the function $t \mapsto \mathbb{P}\{P \in \mathcal{D} \mid P_i \leq t\}$ is non-decreasing over $t \in (0,1]$.

This definition has been a common formulation of positive dependence in prior FDR works, e.g. (Benjamini and Yekutieli, 2001; Blanchard and Roquain, 2008; Ramdas et al., 2019). Clearly, independent p-values satisfy PRDS. A non-trivial example is given for Gaussian observations. Suppose $\mathbf{P} = (\Phi(Z_1), \ldots, \Phi(Z_n))$, where $(Z_1, \ldots, Z_n)$ is a multivariate Gaussian vector with covariance matrix Σ. Then, $\mathbf{P}$ satisfies PRDS if and only if $\Sigma_{i,j} \geq 0$ for all $i \in \mathcal{H}^0$ and $j \in [n]$.

Now we are ready to define the FDP estimate of $\mathrm{Batch}_{\mathrm{BH}}^{\mathrm{PRDS}}$.

Definition 4. The $\mathrm{Batch}_{\mathrm{BH}}^{\mathrm{PRDS}}$ procedure is any rule for assigning test levels α_t such that

$$\widehat{\mathrm{FDP}}_{\mathrm{Batch}_{\mathrm{BH}}^{\mathrm{PRDS}}}(t) = \sum_{s \leq t} \alpha_s \frac{n_s}{n_s + \sum_{r<s} R_r}$$

is controlled under α for all $t \in \mathbb{N}$.

Below is an example update rule that satisfies Definition 4.

Algorithm 3 The $\mathrm{Batch}_{\mathrm{BH}}^{\mathrm{PRDS}}$ algorithm

Input: FDR level α, non-negative sequence $\{\gamma_s\}_{s=1}^{\infty}$ such that $\sum_{s=1}^{\infty} \gamma_s = 1$.
Set $\alpha_1 = \gamma_1 \alpha$;
for $t = 1, 2, \ldots$ **do**
  Run the BH procedure under level α_t on batch $\mathbf{P}_t$;
  Set $\alpha_{t+1} = \alpha \frac{\gamma_{t+1}}{n_{t+1}} (n_{t+1} + \sum_{s=1}^{t} R_s)$;
end

We state our main FDR guarantees for $\mathrm{Batch}_{\mathrm{BH}}^{\mathrm{PRDS}}$ below. Our proof relies on a "super-uniformity lemma", similar to several lemmas in prior work that consider PRDS p-values (Blanchard and Roquain, 2008; Benjamini and Yekutieli, 2001; Ramdas et al., 2019). We prove both this lemma and Theorem 3 later in the Appendix.

Theorem 3. *Suppose that every batch of p-values $\mathbf{P}_t$ satisfies PRDS, and additionally that $P_{t,i}$ and $\{\mathbf{P}_s : s \in \mathcal{I}\}$ are independent whenever $t \notin \mathcal{I}$, for all $i \in \mathcal{H}_t^0$. Then, the $\mathrm{Batch}_{BH}^{PRDS}$ procedure provides anytime* FDR *control: for every $t \in \mathbb{N}$, $\mathrm{FDR}(t) \leq \alpha$.*

In other words, $\mathrm{Batch}_{\mathrm{BH}}^{\mathrm{PRDS}}$ ensures FDR control when p-values are independent across different batches, and positively dependent within each batch. Theorem 3 is a generalization of an earlier result which states that the BH algorithm controls FDR under PRDS Benjamini and Yekutieli (2001).

In online FDR control, handling dependence has generally proved challenging. Javanmard and Montanari (2018) have proposed procedures which control the FDR under arbitrary dependence, however their updates imply an essentially alpha-spending (online Bonferroni) type correction which controls a more stringent criterion called the family-wise error rate (Gordon Lan and DeMets, 1983). Their earlier algorithm called LOND Javanmard and Montanari (2015) was recently proved to control the FDR under PRDS Zrnic et al. (2018), and is a more powerful alternative for the fully online setting than the arbitrary depenence procedure. Indeed, $\mathrm{Batch}_{\mathrm{BH}}^{\mathrm{PRDS}}$ is a minibatch generalization of the LOND algorithm. Finally, it is worth pointing out that the notion of positive dependence we consider in this paper resembles local dependence proposed by Zrnic et al. (2018), although their solutions only control modified FDR (mFDR).

2 EMPIRICAL FDP ESTIMATES IN PRIOR WORK

We give a brief overview of BH and Storey-BH, and we do so in the FDP estimation spirit of Section 2. These derivations were first stated by Storey (2002). Let $\mathbf{P} = \{P_1, \ldots, P_n\}$ be a set of tested p-values, and α be the target FDR level. For any threshold c, Storey defined

$$\widehat{\mathrm{FDP}}_{\mathrm{BH}} := \frac{nc}{\sum_{i=1}^n \mathbf{1}\{P_i \leq c\}}.$$

Picking the maximum c such that $\widehat{\mathrm{FDP}}_{\mathrm{BH}} \leq \alpha$, and rejecting all p-values less than such c, is a succinct statement of the BH procedure. This is a rederivation of the equivalent rule given by Benjamini and Hochberg, who suggested finding

$$k^* = \max\left\{i \in [n] : P_{(i)} \leq \frac{\alpha}{n} i\right\},$$

where $P_{(i)}$ denotes the i-th order statistic of $\mathbf{P}$ in non-decreasing order, and rejecting $P_{(1)}, \ldots, P_{(k^*)}$. This interpretation inspired Storey to improve upon the BH procedure by defining

$$\widehat{\mathrm{FDP}}_{\text{St-BH}} := \frac{ns\widehat{\pi}_0}{\sum_{i=1}^n \mathbf{1}\{P_i \leq c\}},$$

where $\widehat{\pi}_0 = \frac{1+\sum_{i=1}^n \mathbf{1}\{P_i > \lambda\}}{n(1-\lambda)}$, for a user-chosen parameter $\lambda \in (0,1)$. Storey-BH finds the maximum c such that $\widehat{\mathrm{FDP}}_{\text{St-BH}} \leq \alpha$, and rejects all p-values less than such c. The motivation for using Storey-BH is the observation that BH might be overly conservative when there are many non-nulls with a strong signal, because it essentially assumes that $\widehat{\pi}_0 \approx 1$, where $\widehat{\pi}_0$ acts as an estimate of the proportion of nulls in the p-value set.

The FDP estimate approach was also taken in more recent, online FDR work (Ramdas et al., 2017, 2018; Zrnic et al., 2018; Tian and Ramdas, 2019). It started with Ramdas et al. (2017) who rederived and improved upon the LORD algorithm (Javanmard and Montanari, 2018) by noticing that it implicitly controls

$$\widehat{\mathrm{FDP}}_{\mathrm{LORD}}(t) = \frac{\sum_{j=1}^t \alpha_j}{\sum_{i=1}^t \mathbf{1}\{P_i \leq \alpha_i\}},$$

where P_i is a single p-value observed at time i, and α_i is its corresponding test level. Inspired by Storey's idea of making the BH procedure less conservative, the SAFFRON algorithm was derived as a rule for controlling the estimate

$$\widehat{\mathrm{FDP}}_{\mathrm{SAFFRON}}(t) = \frac{\sum_{j=1}^t \frac{\alpha_j}{1-\lambda_j} \mathbf{1}\{P_j > \lambda_j\}}{\sum_{i=1}^t \mathbf{1}\{P_i \leq \alpha_i\}},$$

for some sequence of user-chosen parameters $\{\lambda_t\}$. Several different update rules for α_t have been proposed for LORD and SAFFRON, all of which control the respective FDP estimates under the target FDR level α; for more details, see the respective papers (Javanmard and Montanari, 2018; Ramdas et al., 2018).

Since the original SAFFRON FDP estimate, as stated above, is written in a slightly different, albeit equivalent form to that of Section 4, we point out a subtle difference in the meaning of "α_s" for Storey-BH and SAFFRON. For SAFFRON, α_s denotes the decision threshold for $P_{s,1}$, while in the batch setting, α_s is the Storey-BH level. If Storey-BH is applied to a single p-value under level α_s, then it is rejected if and only if $P_{s,1} \leq (1-\lambda_s)\alpha_s$. This difference should be kept in mind when comparing $\widehat{\text{FDP}}_{\text{Batch}_{\text{St-BH}}}(t)$ to the usual form of $\widehat{\text{FDP}}_{\text{SAFFRON}}(t)$.

3 PROOF OF Theorem 1

First we introduce some additional notation necessary to state the proof. Let $P_{s,1}^{(-i)}, \ldots, P_{s,n_s}^{(-i)}$ be a sequence of p-values that is identical to $P_{s,1}, \ldots, P_{s,n_s}$, *except* for $P_{s,i}^{(-i)}$, which is set to 0. Let also $R_s^{(-i)}$ denote the number of rejections had BH under level α_s been run on $P_{s,1}^{(-i)}, \ldots, P_{s,n_s}^{(-i)}$.

Fix the number of tested batches t, and suppose $\widehat{\text{FDP}}_{\text{Batch}_{\text{BH}}}(t) \leq \alpha$. We prove that this implies $\text{FDR}(t) \leq \alpha$. Starting by definition,

$$\begin{aligned}
\text{FDR}(t) &= \mathbb{E}\left[\frac{\sum_{r\leq t}|\mathcal{R}_r \cap \mathcal{H}_r^0|}{1 \vee \sum_{s\leq t} R_s}\right] \\
&= \sum_{r\leq t}\sum_{i\in\mathcal{H}_r^0}\mathbb{E}\left[\frac{\mathbf{1}\left\{P_{r,i} \leq \frac{\alpha_r}{n_r}R_r\right\}}{1\vee\sum_{s\leq t}R_s}\right] \\
&= \sum_{r\leq t}\sum_{i\in\mathcal{H}_r^0}\mathbb{E}\left[\frac{\mathbf{1}\left\{P_{r,i} \leq \frac{\alpha_r}{n_r}R_r\right\}}{R_r^{(-i)} + \sum_{s\leq t, s\neq r}R_s}\right],
\end{aligned}$$

where the second equality follows by definition of the BH procedure and the third equality follows by observing that, on the event $\{P_{r,i} \leq \frac{\alpha_r}{n_r}R_r\}$, $R_r = R_r^{(-i)}$.

Now we focus on a fixed index $i \in \mathcal{H}_r^0$, for a fixed batch r. Imagine a sequence of batches of p-values identical to the original one, only with $P_{r,i}$ deterministically set to 0. Denote the set of rejections in batch $s \in \mathbb{N}$ in this slightly modified sequence by $\tilde{R}_s^{(-r,i)}$. Notice that $R_s = \tilde{R}_s^{(-r,i)}$ for all $s < r$, and for $s \geq r$ we have $R_s = \tilde{R}_s^{(-r,i)}$ if $P_{r,i}$ from the original sequence is rejected. Therefore, on the event $\{P_{r,i} \leq \frac{\alpha_r}{n_r}R_r\}$, $\tilde{R}_s^{(-r,i)} = R_s$ for all $s \in \mathbb{N}$. This implies

$$\begin{aligned}
\text{FDR}(t) &= \sum_{r\leq t}\sum_{i\in\mathcal{H}_r^0}\mathbb{E}\left[\frac{\mathbf{1}\left\{P_{r,i} \leq \frac{\alpha_r}{n_r}R_r\right\}}{R_r^{(-i)} + \sum_{s\leq t, s\neq r}\tilde{R}_s^{(-r,i)}}\right] \\
&\leq \sum_{r\leq t}\sum_{i\in\mathcal{H}_r^0}\mathbb{E}\left[\frac{\mathbf{1}\left\{P_{r,i} \leq \frac{\alpha_r}{n_r}R_r^{(-i)}\right\}}{R_r^{(-i)} + \sum_{s\leq t, s\neq r}\tilde{R}_s^{(-r,i)}}\right],
\end{aligned}$$

where the final inequality uses the fact that $R_r^{(-i)} \geq R_r$. Conditional on $\mathcal{F}^{r-1}$, $P_{r,i}$ is independent of all other random variables in the final term, namely α_r, $R_r^{(-i)}$ and $\tilde{R}_s^{(-r,i)}, s \in [t], s \neq r$. This allows us to exploit the super-uniformity of $P_{r,i}$ to obtain

$$\begin{aligned}
\text{FDR}(t) &\leq \sum_{r\leq t}\sum_{i\in\mathcal{H}_r^0}\mathbb{E}\left[\mathbb{E}\left[\frac{\mathbf{1}\left\{P_{r,i} \leq \frac{\alpha_r}{n_r}R_r^{(-i)}\right\}}{R_r^{(-i)} + \sum_{s\leq t, s\neq r}\tilde{R}_s^{(-r,i)}}\;\middle|\; \mathcal{F}^{r-1}, R_r^{(-i)}\right]\right] \\
&= \sum_{r\leq t}\sum_{i\in\mathcal{H}_r^0}\mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\left\{P_{r,i} \leq \frac{\alpha_r}{n_r}R_r^{(-i)}\right\}\;\middle|\; \mathcal{F}^{r-1}, R_r^{(-i)}\right]\mathbb{E}\left[\frac{1}{R_r^{(-i)} + \sum_{s\leq t, s\neq r}\tilde{R}_s^{(-r,i)}}\;\middle|\; \mathcal{F}^{r-1}, R_r^{(-i)}\right]\right] \\
&= \sum_{r\leq t}\sum_{i\in\mathcal{H}_r^0}\mathbb{E}\left[\frac{\alpha_r}{n_r}\frac{R_r^{(-i)}}{R_r^{(-i)} + \sum_{s\leq t, s\neq r}\tilde{R}_s^{(-r,i)}}\right].
\end{aligned}$$

Since the update rule for α_r is monotone by assumption, we have

$$\mathrm{FDR}(t) \leq \sum_{r\leq t}\sum_{i\in\mathcal{H}_r^0}\mathbb{E}\left[\frac{\alpha_r}{n_r}\frac{R_r^{(-i)}}{R_r^{(-i)}+\sum_{s\leq t,s\neq r}R_s}\right].$$

Finally, we use the fact that the function $f(x)=\frac{x}{x+a}$ is a non-decreasing function for $a\geq 0$ to conclude

$$\begin{aligned}
\mathrm{FDR}(t) &\leq \sum_{r\leq t}\sum_{i\in\mathcal{H}_r^0}\mathbb{E}\left[\frac{\alpha_r}{n_r}\frac{R_r^+}{R_r^+ +\sum_{s\leq t,s\neq r}R_s}\right]\\
&\leq \sum_{r\leq t}\mathbb{E}\left[\alpha_r\frac{R_r^+}{R_r^+ +\sum_{s\leq t,s\neq r}R_s}\right]\\
&= \mathbb{E}\left[\widehat{\mathrm{FDP}}_{\mathrm{Batch}_{\mathrm{BH}}}\right]\\
&\leq \alpha,
\end{aligned}$$

where the last inequality is deterministic, by design of the algorithm. This concludes the proof.

4 PROOF OF Theorem 2

As in the proof of Theorem 1, we introduce some additional notation necessary to state the proof. Recall that the Storey-BH procedure uses a null proportion estimate of the form

$$\widehat{\pi}_{0,s}=\frac{1+\sum_{j=1}^{n_s}P_{s,j}}{n_s(1-\lambda_s)}.$$

Let $P_{s,1}^{(-i)},\dots,P_{s,n_s}^{(-i)}$ be a sequence of p-values that is identical to $P_{s,1},\dots,P_{s,n_s}$, *except* for $P_{s,i}^{(-i)}$, which is set to 0. Let also $R_s^{(-i)}$ denote the number of rejections had Storey-BH under level α_s been run on $P_{s,1}^{(-i)},\dots,P_{s,n_s}^{(-i)}$. With this, define the "hallucinated" null proportion as

$$\widehat{\pi}_{0,s}^{(-i)}=\frac{1+\sum_{j=1}^{n_s}P_{s,j}^{(-i)}}{n_s(1-\lambda_s)}.$$

Fix the number of tested batches t, and suppose $\widehat{\mathrm{FDP}}_{\mathrm{Batch}_{\text{St-BH}}}(t)\leq\alpha$. We prove that this condition implies $\mathrm{FDR}(t)\leq\alpha$. Starting by definition,

$$\begin{aligned}
\mathrm{FDR}(t) &= \mathbb{E}\left[\frac{\sum_{r\leq t}|\mathcal{R}_r\cap\mathcal{H}_r^0|}{1\vee\sum_{s\leq t}R_s}\right]\\
&= \sum_{r\leq t}\sum_{i\in\mathcal{H}_r^0}\mathbb{E}\left[\frac{\mathbf{1}\left\{P_{r,i}\leq\frac{\alpha_r}{\widehat{\pi}_{0,r}n_r}R_r\right\}}{1\vee\sum_{s\leq t}R_s}\right]\\
&= \sum_{r\leq t}\sum_{i\in\mathcal{H}_r^0}\mathbb{E}\left[\frac{\mathbf{1}\left\{P_{r,i}\leq\frac{\alpha_r}{\widehat{\pi}_{0,r}n_r}R_r\right\}}{R_r^{(-i)}+\sum_{s\leq t,s\neq r}R_s}\right],
\end{aligned}$$

where the second equality follows by definition of the Storey-BH procedure and the third equality follows by observing that, on the event $\{P_{r,i}\leq\frac{\alpha_r}{n_r\widehat{\pi}_{0,r}}R_r\}$, $R_r=R_r^{(-i)}$.

Now we focus on a fixed $i\in\mathcal{H}_r^0$, for a fixed batch r. Imagine a sequence of batches of p-values identical to the original one, only with $P_{r,i}$ deterministically set to 0. Denote the set of rejections in batch $s\in\mathbb{N}$ in this slightly modified sequence by $\tilde{R}_s^{(-r,i)}$. Notice that $R_s=\tilde{R}_s^{(-r,i)}$ for all $s<r$, and for $s\geq r$ we have $R_s=\tilde{R}_s^{(-r,i)}$ if $P_{r,i}$

from the original sequence in rejected. Therefore, on the event $\{P_{r,i} \leq \frac{\alpha_r}{n_r \widehat{\pi}_{0,r}} R_r\}$, $\tilde{R}_s^{(-r,i)} = R_s$ for all $s \in \mathbb{N}$. This implies

$$\begin{aligned} \mathrm{FDR}(t) &= \sum_{r \leq t} \sum_{i \in \mathcal{H}_r^0} \mathbb{E}\left[\frac{\mathbf{1}\left\{ P_{r,i} \leq \frac{\alpha_r}{n_r \widehat{\pi}_{0,r}} R_r \right\}}{R_r^{(-i)} + \sum_{s \leq t, s \neq r} \tilde{R}_s^{(-r,i)}} \right] \\ &\leq \sum_{r \leq t} \sum_{i \in \mathcal{H}_r^0} \mathbb{E}\left[\frac{\mathbf{1}\left\{ P_{r,i} \leq \frac{\alpha_r}{n_r \widehat{\pi}_{0,r}^{(-i)}} R_r^{(-i)} \right\}}{R_r^{(-i)} + \sum_{s \leq t, s \neq r} \tilde{R}_s^{(-r,i)}} \right], \end{aligned}$$

where the final inequality uses the fact that $R_r^{(-i)} \geq R_r$ and $\widehat{\pi}_{0,r} \geq \widehat{\pi}_{0,r}^{(-i)}$. Note that $P_{r,i}$ is independent of all other random variables in the final term, namely $\alpha_r, R_r^{(-i)}$, $\widehat{\pi}_{0,r}^{(-i)}$ and $\tilde{R}_s, s \in [t], s \neq r$. This allows us to exploit the super-uniformity of $P_{r,i}$ to obtain

$$\begin{aligned} \mathrm{FDR}(t) &\leq \sum_{r \leq t} \sum_{i \in \mathcal{H}_r^0} \mathbb{E}\left[\mathbb{E}\left[\frac{\mathbf{1}\left\{ P_{r,i} \leq \frac{\alpha_r}{n_r \widehat{\pi}_{0,r}^{(-i)}} R_r^{(-i)} \right\}}{R_r^{(-i)} + \sum_{s \leq t, s \neq r} \tilde{R}_s^{(-r,i)}} \;\middle|\; \mathcal{F}^{r-1}, R_r^{(i)}, \widehat{\pi}_{0,r}^{(-i)} \right] \right] \\ &= \sum_{r \leq t} \sum_{i \in \mathcal{H}_r^0} \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}\left\{ P_{r,i} \leq \frac{\alpha_r}{n_r \widehat{\pi}_{0,r}^{(-i)}} R_r^{(-i)} \right\} \;\middle|\; \mathcal{F}^{r-1}, R_r^{(i)}, \widehat{\pi}_{0,r}^{(-i)} \right] \mathbb{E}\left[\frac{1}{R_r^{(i)} + \sum_{s \leq t, s \neq r} \tilde{R}_s^{(-r,i)}} \;\middle|\; \mathcal{F}^{r-1}, R_r^{(-i)}, \widehat{\pi}_{0,r}^{(-i)} \right] \right] \\ &= \sum_{r \leq t} \sum_{i \in \mathcal{H}_r^0} \mathbb{E}\left[\frac{\alpha_r}{n_r \widehat{\pi}_{0,r}^{(-i)}} \frac{R_r^{(i)}}{R_r^{(-i)} + \sum_{s \leq t, s \neq r} \tilde{R}_s^{(-r,i)}} \right]. \end{aligned}$$

Since the update for α_r is monotone, and since setting a p-value to 0 can only increase the number of rejections in a given batch, we have

$$\mathrm{FDR}(t) \leq \sum_{r \leq t} \sum_{i \in \mathcal{H}_r^0} \mathbb{E}\left[\frac{\alpha_r}{n_r \widehat{\pi}_{0,r}^{(-i)}} \frac{R_r^{(-i)}}{R_r^{(-i)} + \sum_{s \leq t, s \neq r} R_s} \right].$$

Now we use a similar trick of ignoring one p-value as given above. Imagine a sequence of p-values identical to the original one, however with $P_{r,i}$ deterministically set to 1. Denote the set of rejections in batch $s \in \mathbb{N}$ in this modified sequence by $\tilde{R}_s^{(+r,i)}$. We have $R_s = R_s^{(+r,i)}$ for $s < r$, and the same holds for $s \geq r$ on the event $\{P_{r,i} > \lambda_r\}$. From this, we can conclude the following

$$\begin{aligned} \mathbb{E}\left[\frac{\mathbf{1}\{P_{r,i} > \lambda_r\}}{(1-\lambda_r)} \frac{\alpha_r}{n_r \widehat{\pi}_{0,r}^{(-i)}} \frac{R_r^{(-i)}}{R_r^{(-i)} + \sum_{s \leq t, s \neq l} R_s} \right] &= \mathbb{E}\left[\frac{\mathbf{1}\{P_{r,i} > \lambda_r\}}{(1-\lambda_r)} \frac{\alpha_r}{n_r \widehat{\pi}_{0,r}^{(-i)}} \frac{R_r^{(-i)}}{R_r^{(-i)} + \sum_{s \leq t, s \neq r} \tilde{R}_s^{(+r,i)}} \right] \\ &= \mathbb{E}\left[\frac{\mathbf{1}\{P_{r,i} > \lambda_r\}}{(1-\lambda_r)} \right] \mathbb{E}\left[\frac{\alpha_r}{n_r \widehat{\pi}_{0,r}^{(-i)}} \frac{R_r^{(-i)}}{R_r^{(-i)} + \sum_{s \leq t, s \neq r} \tilde{R}_s^{(+r,i)}} \right] \\ &\geq \mathbb{E}\left[\frac{\alpha_r}{n_r \widehat{\pi}_{0,r}^{(-i)}} \frac{R_r^{(-i)}}{R_r^{(-i)} + \sum_{s \leq t, s \neq r} \tilde{R}_s^{(+r,i)}} \right] \\ &\geq \mathbb{E}\left[\frac{\alpha_r}{n_r \widehat{\pi}_{0,r}^{(-i)}} \frac{R_r^{(-i)}}{R_r^{(-i)} + \sum_{s \leq t, s \neq r} R_s} \right], \end{aligned}$$

where the first inequality uses super-uniformity of null p-values, and the second inequality uses monotonicity of the test level update rule. Therefore, we can write

$$\mathrm{FDR}(t) \leq \sum_{r \leq t} \sum_{i \in \mathcal{H}_r^0} \mathbb{E}\left[\frac{\mathbf{1}\{P_{r,i} > \lambda_r\}}{(1-\lambda_r)} \frac{\alpha_r}{n_r \widehat{\pi}_{0,r}^{(-i)}} \frac{R_r^{(-i)}}{R_r^{(-i)} + \sum_{s \leq t, s \neq r} R_s} \right].$$

Finally, we use the fact that the function $f(x) = \frac{x}{x+a}$ is a non-decreasing function for $a \geq 0$ to conclude

$$\begin{aligned}
\mathrm{FDR}(t) &\leq \sum_{r \leq t} \sum_{i \in \mathcal{H}_r^0} \mathbb{E}\left[\frac{\mathbf{1}\{P_{r,i} > \lambda_r\}}{(1-\lambda_r)} \frac{\alpha_r}{n_r \min_j \widehat{\pi}_{0,r}^{(-j)}} \frac{R_r^+}{R_r^+ + \sum_{s \leq t, s \neq r} R_s}\right] \\
&\leq \sum_{r \leq t} \mathbb{E}\left[\frac{\alpha_r k_r R_r^+}{R_r^+ + \sum_{s \leq t, s \neq r} R_s}\right] \\
&= \mathbb{E}\left[\widehat{\mathrm{FDP}}_{\mathrm{Batch}_{\text{St-BH}}}\right] \\
&\leq \alpha,
\end{aligned}$$

where once again the last inequality is deterministic by design of the algorithm, thus completing the proof of the theorem.

5 $\mathbf{Batch}_{\mathbf{BH}}^{\mathbf{PRDS}}$ PROOFS

To facilitate the proof of FDR control, we prove a a "super-uniformity lemma", similar to several lemmas in prior work that consider PRDS p-values (Blanchard and Roquain, 2008; Benjamini and Yekutieli, 2001; Ramdas et al., 2019).

Lemma 1. *Let $U \in [0,1]$ and $V \in \mathbb{N} \cup \{0\}$ be random variables that satisfy the following:*

- *U is super-uniform, i.e. $\mathbb{P}\{U \leq u\} \leq u$ for $u \in [0,1]$.*
- *$\mathbb{P}\{V \leq r \mid U \leq u\}$ is non-decreasing in u, for every fixed $r > 0$.*
- *$V \leq n$ almost surely.*

Then, for every $a \geq 0, c > 0$, $\mathbb{E}\left[\frac{\mathbf{1}\{U \leq cV\}}{V+a}\right] \leq \frac{cn}{n+a}$.

Proof. The proof when $a = 0$ is given by Blanchard and Roquain (2008) (Lemma 3.2), so in what follows we assume $a > 0$.

We expand the expectation as follows:

$$\begin{aligned}
\mathbb{E}\left[\frac{\mathbf{1}\{U \leq cV\}}{V+a}\right] &= \sum_{i=0}^{n} \frac{1}{i+a} \mathbb{P}\{U \leq ci, V = i\} \\
&= \sum_{i=0}^{n} \frac{\mathbb{P}\{U \leq ci\}}{i+a} \frac{\mathbb{P}\{U \leq ci, V = i\}}{\mathbb{P}\{U \leq ci\}} \\
&\leq \sum_{i=0}^{n} \frac{ci}{i+a} \mathbb{P}\{V = i \mid U \leq ci\} \\
&= \sum_{i=0}^{n} \frac{ci}{i+a} \left(\mathbb{P}\{V \leq i \mid U \leq ci\} - \mathbb{P}\{V \leq i-1 \mid U \leq ci\}\right) \\
&= \frac{cn}{n+a} \sum_{i=0}^{n} \left(\mathbb{P}\{V \leq i \mid U \leq ci\} - \mathbb{P}\{V \leq i-1 \mid U \leq ci\}\right),
\end{aligned}$$

where the inequality follows by the super-uniformity assumption on U. By the second assumption of the lemma,

$\mathbb{P}\{V \leq i-1 \mid U \leq ci\} \geq \mathbb{P}\{V \leq i-1 \mid U \leq c(i-1)\}$, hence

$$\begin{aligned}
&\frac{cn}{n+a}\sum_{i=0}^{n}\left(\mathbb{P}\{V \leq i \mid U \leq ci\} - \mathbb{P}\{V \leq i-1 \mid U \leq ci\}\right) \\
\leq\ &\frac{cn}{n+a}\sum_{i=0}^{n}\left(\mathbb{P}\{V \leq i \mid U \leq ci\} - \mathbb{P}\{V \leq i-1 \mid U \leq c(i-1)\}\right) \\
\leq\ &\frac{cn}{n+a},
\end{aligned}$$

which follows by a telescoping sum argument. □

5.1 Proof of Theorem 3

Fix the number of tested batches t, and suppose $\widehat{\mathrm{FDP}}_{\mathrm{Batch}_{\mathrm{BH}}^{\mathrm{PRDS}}}(t) \leq \alpha$. We prove that this implies $\mathrm{FDR}(t) \leq \alpha$. Starting by definition,

$$\begin{aligned}
\mathrm{FDR}(t) &= \mathbb{E}\left[\frac{\sum_{r \leq t}|\mathcal{R}_r \cap \mathcal{H}_r^0|}{1 \vee \sum_{s \leq t} R_s}\right] \\
&= \sum_{r \leq t}\sum_{i \in \mathcal{H}_r^0}\mathbb{E}\left[\frac{\mathbf{1}\left\{P_{r,i} \leq \frac{\alpha_r}{n_r}R_r\right\}}{1 \vee \sum_{s \leq t} R_s}\right] \\
&\leq \sum_{r \leq t}\sum_{i \in \mathcal{H}_r^0}\mathbb{E}\left[\frac{\mathbf{1}\left\{P_{r,i} \leq \frac{\alpha_r}{n_r}R_r\right\}}{1 \vee (R_r + \sum_{s < r} R_s)}\right],
\end{aligned}$$

where the second equality follows by definition of the BH procedure and the inequality follows by ignoring all rejections in the denominator after the r-th batch.

We now condition on $\mathcal{F}^{r-1}$ to obtain

$$\begin{aligned}
\sum_{r \leq t}\sum_{i \in \mathcal{H}_r^0}\mathbb{E}\left[\frac{\mathbf{1}\left\{P_{r,i} \leq \frac{\alpha_r}{n_r}R_r\right\}}{1 \vee (R_r + \sum_{s < r} R_s)}\right] &= \sum_{r \leq t}\sum_{i \in \mathcal{H}_r^0}\mathbb{E}\left[\mathbb{E}\left[\frac{\mathbf{1}\left\{P_{r,i} \leq \frac{\alpha_r}{n_r}R_r\right\}}{1 \vee (R_r + \sum_{s < r} R_s)} \,\middle|\, \mathcal{F}^{r-1}\right]\right] \\
&\leq \sum_{r \leq t}\sum_{i \in \mathcal{H}_r^0}\mathbb{E}\left[\frac{\frac{\alpha_r}{n_r}n_r}{1 \vee (n_r + \sum_{s < r} R_s)}\right] \\
&= \sum_{r \leq t}\mathbb{E}\left[\frac{n_r\alpha_r}{1 \vee (n_r + \sum_{s < r} R_s)}\right],
\end{aligned}$$

where the inequality applies Lemma 1 and the fact that α_r and $\{R_s, s < r\}$ are measurable with respect to the conditioning, and the final equality uses the fact that $|\mathcal{H}_r^0| \leq n_r$.

Since the final expression is equal to $\mathbb{E}\left[\widehat{\mathrm{FDP}}_{\mathrm{Batch}_{\mathrm{BH}}^{\mathrm{PRDS}}}\right]$, we can conclude that

$$\mathrm{FDR}(t) \leq \mathbb{E}\left[\widehat{\mathrm{FDP}}_{\mathrm{Batch}_{\mathrm{BH}}^{\mathrm{PRDS}}}\right] \leq \alpha,$$

as desired.

6 PROOF OF Fact 1

The control of the estimate follows by observing

$$\begin{aligned}\widehat{\mathrm{FDP}}_{\mathrm{Batch_{BH}}}(t+1) &\leq \sum_{s\leq t}\alpha_s \frac{R_s^+}{R_s^+ + \sum_{r\leq t+1, r\neq s} R_r} + \alpha_{t+1}\frac{n_{t+1}}{n_{t+1}+\sum_{r\leq t} R_r}\\ &= \sum_{s\leq t}\gamma_s\alpha - \sum_{s\leq t}\alpha_s\frac{R_s^+}{R_s^+ + \sum_{r\leq t, r\neq s} R_r} + \sum_{s\leq t}\alpha_s\frac{R_s^+}{R_s^+ + \sum_{r\leq t+1, r\neq s} R_r}\\ &\leq \alpha,\end{aligned}$$

where the second step follows by replacing α_{t+1} with the update rule from Algorithm 1, and the final inequality follows by the assumption that $\sum_{j=1}^{\infty}\gamma_j = 1$.

7 DETECTING CREDIT CARD FRAUD

We apply our algorithms to real credit card transaction data. Credit card companies test for whether transactions are fraudulent; if the transactions are deemed to be fraudulent, they are denied. However, it is important to control the proportion of transactions that are falsely identified as fraudulent, as these false identifications inconvenience users by declining legitimate transactions. Note that major credit card companies generally have thousands of transactions per second, so testing for fraud in batches would incur no delay in their processing.

We use a dataset released by the Machine Learning Group of Université Libre de Bruxelles for a Kaggle competition[1](Dal Pozzolo et al., 2015). The dataset comprises 492 fraudulent transactions and 284,315 legitimate transactions. For each transaction, the null hypothesis is that the transaction is not fraudulent, which means that the proportion of non-nulls π_1 is approximately 0.173%. Such asymmetry between the proportion of nulls and non-nulls is typical in applications of FDR methods.

Each transaction in the dataset has 28 principal component analysis (PCA) features, the monetary value of the transaction, and a binary label indicating whether the transaction is fraudulent. The PCA features are provided instead of the original features for confidentiality. For each transaction i, let $y_i \in \{0,1\}$ denote whether the transaction is fraudulent ($y_i = 1$ denotes a fraudulent transaction) and let $x_i \in \mathbb{R}^{29}$ denote the vector of the transaction's PCA features and monetary value.

In a similar fashion to Javanmard and Montanari (2018), we randomly partition the transactions into the subsets Train1 (60% of the transactions), Train2 (20% of the transactions), and Test (20% of the transactions). We use the training set to learn the null distribution for the purpose of generating p-values, and compare different hypothesis testing procedures on the p-values of the test subset. We fit a logistic regression model to Train1. In particular, for i in Train1, we model the probability that transaction i is fraudulent as

$$\mathbb{P}\{Y_i = 1 | X_i = x_i\} = \sigma(\theta^T x_i),$$

where $\sigma(x) = \frac{1}{1+\mathrm{e}^{-x}}$.

For each i in Train2 and each j in Test, we compute $q_i = \sigma(\theta^T x_i)$ and $q_j^{\mathrm{Test}} = \sigma(\theta^T x_j)$. Let T_0 denote the subset of Train2 that are non-fraudulent transactions. We construct the p-value P_j as

$$P_j = \frac{1}{n_0}|\{i\in T_0 : q_i > q_j^{\mathrm{Test}}\}|.$$

We set $\alpha = 0.1$ and we set all other hyperparameters the same way as in previous experiments. We use 100 random splits of the transactions into Train1, Train2, and Test in order to compute the average and one standard deviation around the average of power and FDR.

For both the non-adaptive and the adaptive methods, we observe higher power for online batch procedures than for standard online procedures, across several batch sizes of different orders of magnitude. Our findings are summarized in Table 1 and Table 2. However, as observed in experiments on synthetic data as well, we do not observe a monotone relationship between batch size and power.

[1] https://www.kaggle.com/mlg-ulb/creditcardfraud

Table 1: Non-Adaptive Algorithms on Real Data

	Power	FDR
Batch_{BH} (10^1-size)	0.242 ± 0.053	0.126 ± 0.075
Batch_{BH} (10^2-size)	0.299 ± 0.100	0.102 ± 0.067
Batch_{BH} (10^3-size)	0.260 ± 0.086	0.082 ± 0.064
LORD	0.231 ± 0.051	0.082 ± 0.067

Table 2: Adaptive Algorithms on Real Data

	Power	FDR
$\text{Batch}_{\text{St-BH}}$ (10^1-size)	0.240 ± 0.052	0.125 ± 0.081
$\text{Batch}_{\text{St-BH}}$ (10^2-size)	0.291 ± 0.098	0.096 ± 0.065
$\text{Batch}_{\text{St-BH}}$ (10^3-size)	0.246 ± 0.075	0.074 ± 0.063
SAFFRON	0.211 ± 0.041	0.137 ± 0.086

8 ADDITIONAL MONOTONE UPDATE RULES

In this section we provide several monotone updates for Batch_{BH} and $\text{Batch}_{\text{St-BH}}$, which control the FDR for arbitrary, possibly adversarially chosen p-value distributions.

8.1 Batch_{BH} Rules

Algorithm 4 One version of the Batch_{BH} algorithm

input: FDR level α, non-increasing sequence $\{\gamma_t\}_{t=1}^\infty$ summing to 1, initial wealth $W_0 \leq \alpha$
Set $\alpha_1 = \gamma_1 \frac{W_0}{n_1}$
for $t = 1, 2, \dots$ **do**
  Run the BH procedure under level α_t on batch $\mathbf{P}_t$
  Set $\alpha_{t+1} = \gamma_{t+1} \frac{W_0}{n_{t+1}} + \frac{\alpha}{n_{t+1}} \sum_{s=1}^t \gamma_{t+1-s} R_s - \frac{W_0}{n_{t+1}} \sum_{s=1}^t \gamma_{t+1-s} \mathbf{1}\{s = \tau_1\}$,
  where $\tau_1 = \min\{s \geq 1 : R_s > 0\}$
end

Fact 2. The algorithm given in Algorithm 4 is monotone and guarantees $\widehat{\text{FDP}}_{\text{Batch}_{\text{BH}}}(t) \leq \alpha$.

Proof. First we prove that the algorithm guarantees $\widehat{\text{FDP}}_{\text{Batch}_{\text{BH}}}(t) \leq \alpha$. Starting by definition, we have

$$
\begin{aligned}
\widehat{\text{FDP}}_{\text{Batch}_{\text{BH}}}(t) &= \sum_{j=1}^t \alpha_j \frac{R_j^+}{R_j^+ + \sum_{k \leq t, k \neq j} R_k} \\
&\leq \sum_{j=1}^t \alpha_j \frac{n_j}{n_j + \sum_{k \leq t, k \neq j} R_k} \\
&\leq \frac{\sum_{j=1}^t \alpha_j n_j}{1 \vee \sum_{k=1}^t R_k} \\
&= \frac{W_0 \sum_{j=1}^t \gamma_j + \alpha \sum_{j=1}^t \sum_{l=1}^{j-1} \gamma_{j-l} R_l - W_0 \sum_{j=1}^t \sum_{l=1}^{j-1} \gamma_{j-l} \mathbf{1}\{l = \tau_1\}}{1 \vee \sum_{k=1}^t R_k} \\
&= \frac{W_0 \sum_{j=1}^t \gamma_j + \alpha \sum_{j=1}^t \sum_{l=1}^{j-1} \gamma_{j-l} R_l - W_0 \sum_{j=\tau_1+1}^t \gamma_{j-\tau_1}}{1 \vee \sum_{k=1}^t R_k},
\end{aligned}
$$

where the first inequality follows because $R_j^+ \leq n_j$, the second inequality follows because $n_j \geq R_j \vee 1$, the second equality follows by the definition of α_j, as given in Algorithm 4, and the third equality is obtained by removing the summation terms where $l \neq \tau_1$.

If $t < \tau_1$, then $R_1 = R_2 = \cdots = R_t = 0$ by the definition of τ_1, so the bound above evaluates to $W_0 \sum_{j=1}^t \gamma_j \leq W_0 \leq \alpha$, which is the desired conclusion. Thus, for the remainder of the proof, we assume that $t \geq \tau_1$.

Since $R_i = 0$ for $i < \tau_1$, we can remove such terms from consideration, leaving us with

$$\begin{aligned}\widehat{\mathrm{FDP}}_{\mathrm{Batch_{BH}}}(t) &\leq \frac{W_0 \sum_{j=1}^t \gamma_j + \alpha \sum_{j=\tau_1+1}^t \sum_{l=\tau_1}^{j-1} \gamma_{j-l} R_l - W_0 \sum_{j=\tau_1+1}^t \gamma_{j-\tau_1}}{\sum_{k=\tau_1}^t R_k} \\ &= \frac{W_0 \sum_{j=1}^t \gamma_j + \alpha \sum_{j=\tau_1+2}^t \sum_{l=\tau_1+1}^{j-1} \gamma_{j-l} R_l + (\alpha R_{\tau_1} - W_0) \sum_{j=\tau_1+1}^t \gamma_{j-\tau_1}}{\sum_{k=\tau_1}^t R_k}.\end{aligned}$$

Since $\{\gamma_t\}_{t=1}^\infty$ is defined to be a non-negative sequence summing to 1, then $W_0 \sum_{j=1}^t \gamma_j \leq W_0$ and $(\alpha R_{\tau_1} - W_0) \sum_{j=\tau_1+1}^t \gamma_{j-\tau_1} \leq \alpha R_{\tau_1} - W_0$. We apply this observation to obtain

$$\begin{aligned}\widehat{\mathrm{FDP}}_{\mathrm{Batch_{BH}}}(t) &\leq \alpha \frac{R_{\tau_1} + \sum_{j=\tau_1+2}^t \sum_{l=\tau_1+1}^{j-1} \gamma_{j-l} R_l}{\sum_{k=\tau_1}^t R_k} \\ &\leq \alpha \frac{R_{\tau_1} + \sum_{l=\tau_1+1}^{t-1} R_l}{\sum_{k=\tau_1}^t R_k} \leq \alpha,\end{aligned}$$

where we again use the fact that the sequence $\{\gamma_t\}_{t=1}^\infty$ sums to one. The final inequality concludes the proof that $\widehat{\mathrm{FDP}}_{\mathrm{Batch_{BH}}}(t)$ is controlled.

We now prove that the update rule is monotone. We restate the test level update rule in a more convenient form:

$$\alpha_{t+1} = \gamma_{t+1} \frac{W_0}{n_{t+1}} + \frac{1}{n_{t+1}} \sum_{j=1}^t \gamma_{t+1-j} (\alpha R_j - W_0 \mathbf{1}\{j = \tau_1\}). \tag{1}$$

Suppose we have two sequences of p-values $(P_{1,1}, P_{1,2}, \ldots, P_{t,n_t})$ and $(\tilde{P}_{1,1}, \tilde{P}_{1,2}, \ldots, \tilde{P}_{t,n_t})$ such that $(P_{1,1}, P_{1,2}, \ldots, P_{t,n_t}) \leq (\tilde{P}_{1,1}, \tilde{P}_{1,2}, \ldots, \tilde{P}_{t,n_t})$ coordinate-wise. Denote all relevant $\mathrm{Batch_{BH}}$ quantities on these two sequences using a similar notation.

If $\alpha_t \geq \tilde{\alpha}_t$, then $R_t \geq \tilde{R}_t$ by the definition of the BH procedure. The final observation is that the above update is monotonically increasing in $(R_1, R_2, \ldots, R_t)$, which concludes the proof. □

If we know that all batches are of size at least M, we can also derive the following rule which is expected to be more powerful than the one above, when the batch sizes do not vary too much. Moreover, when all batches are of the same size $n_j \equiv M$, the rule is strictly more powerful.

Algorithm 5 One version of the $\mathrm{Batch_{BH}}$ algorithm when $n_s \geq M$ for all s

input: FDR level α, non-increasing sequence $\{\gamma_t\}_{t=1}^\infty$ summing to 1, initial wealth $W_0 \leq \alpha$
Set $\alpha_1 = \gamma_1 \frac{M}{n_1} \alpha$
for $t = 1, 2, \ldots$ **do**
  Run the BH procedure under level α_t on batch $\mathbf{P}_t$
  Set $\alpha_{t+1} = \gamma_{t+1} \frac{M}{n_{t+1}} \alpha + \frac{\alpha}{n_{t+1}} \sum_{s=1}^t \gamma_{t+1-s} R_s^{\mathrm{add}}$, where $R_s^{\mathrm{add}} = \min\{R_s, \max\{R_r : r < s\}\}$
end

Fact 3. If $n_s \geq M$ for all $s \in \mathbb{N}$, the algorithm given in Algorithm 5 is monotone and guarantees $\widehat{\mathrm{FDP}}_{\mathrm{Batch_{BH}}}(t) \leq \alpha$.

Proof. We first prove that the algorithm guarantees $\widehat{\mathrm{FDP}}_{\mathrm{Batch}_{\mathrm{BH}}}(t) \leq \alpha$. Starting by definition, we have

$$\begin{aligned}\widehat{\mathrm{FDP}}_{\mathrm{Batch}_{\mathrm{BH}}}(t) &= \sum_{j=1}^{t} \alpha_j \frac{R_j^+}{R_j^+ + \sum_{k\leq t, k\neq j} R_k} \leq \sum_{j=1}^{t} \alpha_j \frac{n_j}{n_j + \sum_{k\leq t, k\neq j} R_k} \\ &\leq \sum_{j=1}^{t} \alpha_j \frac{n_j}{M + \sum_{k\leq t, k\neq j} R_k} \leq \frac{\sum_{j=1}^{t} \alpha_j n_j}{M + \min_i \sum_{k\leq t, k\neq i} R_k},\end{aligned}$$

where the first inequality follows because $R_j^+ \leq n_j$ and second inequality follows by the assumption that $n_j \geq M$ for all j. Substituting in the update rule from Algorithm 5, we obtain

$$\widehat{\mathrm{FDP}}_{\mathrm{Batch}_{\mathrm{BH}}}(t) \leq \alpha \frac{M\sum_{j=1}^{t}\gamma_j + \sum_{j=1}^{t}\sum_{l=1}^{j-1}\gamma_{j-l}R_l^{\mathrm{add}}}{M + \min_i \sum_{k\leq t, k\neq i} R_k} \leq \alpha \frac{M + \sum_{l=1}^{t-1} R_l^{\mathrm{add}}}{M + \min_i \sum_{k\leq t, k\neq i} R_k},$$

where we use the fact that the sequence $\{\gamma_t\}_{t=1}^{\infty}$ is defined to be non-negative and summing to one. Since $\sum_{l=1}^{t-1} R_l^{\mathrm{add}} = \min_i \sum_{k<t, k\neq i} R_k$ by the definition of R_l^{add}, we can conclude $\widehat{\mathrm{FDP}}_{\mathrm{Batch}_{\mathrm{BH}}}(t) \leq \alpha$, as desired.

Monotonicity follows by the same steps as in the proof of Fact 2, thus completing the proof.

□

Below we give one more monotone $\mathrm{Batch}_{\mathrm{BH}}$ update, based on a different idea.

Algorithm 6 One version of the $\mathrm{Batch}_{\mathrm{BH}}$ algorithm

input: FDR level α, sequence $\{\gamma_t\}_{t=1}^{\infty}$ summing to 1 such that $\gamma_2 \geq \gamma_1$
Set $\alpha_1 = \gamma_1\alpha$
  Run the BH procedure under level α_1 on batch $\mathbf{P}_1$
  Set $\alpha_2 = (\gamma_2\alpha - \alpha_1)\frac{R_1+n_2}{n_2}$
  for $t = 2, 3, \ldots$ **do**
    Run the BH procedure under level α_t on batch $\mathbf{P}_t$
    Set $\alpha_{t+1} = \left(\frac{R_t\left(\sum_{i=1}^{t-1}\alpha_i n_i\right)}{(\sum_{j=1}^{t-1} n_j)(\sum_{k=1}^{t-1} n_k + R_t)} + \gamma_{t+1}\alpha\right)\frac{\sum_{l=1}^{t} R_l + n_{t+1}}{n_{t+1}}$
end

Fact 4. The update given in Algorithm 6 controls $\widehat{\mathrm{FDP}}_{\mathrm{Batch}_{\mathrm{BH}}}$ and is monotone.

Proof. First we use induction to prove that for every $t \in \mathbb{N}$ the update controls $\sum_{i=1}^{t}\alpha_i\frac{n_i}{n_i+\sum_{j<t,j\neq i} R_j}$ under $\sum_{i=1}^{t}\gamma_i\alpha$. Then, since $\sum_{i=1}^{t}\alpha_i\frac{n_i}{n_i+\sum_{j<t,j\neq i} R_j} \geq \sum_{i=1}^{t}\alpha_i\frac{R_i^+}{R_i^+ + \sum_{j<t,j\neq i} R_j} \geq \widehat{\mathrm{FDP}}_{\mathrm{Batch}_{\mathrm{BH}}}(t)$, the first claim in the fact immediately follows.

This statement is clearly true for the two special cases when $t \in \{1, 2\}$, and now assume $\sum_{i=1}^{t}\alpha_i\frac{n_i}{n_i+\sum_{j<t,j\neq i} R_j} \leq \sum_{i=1}^{t}\gamma_i\alpha$.

We can write

$$\frac{R_t\left(\sum_{i=1}^{t-1}\alpha_i n_i\right)}{(\sum_{j=1}^{t-1} n_j)(\sum_{k=1}^{t-1} n_k + R_t)} = \sum_{i=1}^{t-1}\alpha_i\left(\frac{n_i}{\sum_{j=1}^{t-1} n_j} - \frac{n_i}{\sum_{k=1}^{t-1} n_k + R_t}\right). \tag{2}$$

We use this to rewrite $\sum_{i=1}^{t+1}\alpha_i\frac{n_i}{n_i+\sum_{j<t+1,j\neq i} R_j}$ as

$$\sum_{i=1}^{t+1}\alpha_i\frac{n_i}{n_i+\sum_{j<t+1,j\neq i} R_j} = \sum_{i=1}^{t}\alpha_i\frac{n_i}{n_i+\sum_{j\leq t,j\neq i} R_j} + \sum_{i=1}^{t-1}\alpha_i\left(\frac{n_i}{\sum_{j=1}^{t-1} n_j} - \frac{n_i}{\sum_{k=1}^{t-1} n_k + R_t}\right) + \gamma_{t+1}\alpha,$$

where we apply the test level update given in Algorithm 6. Further, we have

$$\begin{aligned}\sum_{i=1}^{t+1}\alpha_i\frac{n_i}{n_i+\sum_{j<t+1,j\neq i}R_j} &\leq \sum_{i=1}^{t}\alpha_i\frac{n_i}{n_i+\sum_{j\leq t,j\neq i}R_j} + \sum_{i=1}^{t-1}\alpha_i\left(\frac{n_i}{n_i+\sum_{j\leq t-1,j\neq i}R_j} - \frac{n_i}{n_i+\sum_{k\leq t,k\neq i}R_k}\right)\\ &\quad + \gamma_{t+1}\alpha\\ &= \sum_{i=1}^{t}\alpha_i\frac{n_i}{n_i+\sum_{j<t,j\neq i}R_j} + \gamma_{t+1}\alpha\\ &\leq \sum_{i=1}^{t+1}\gamma_i\alpha,\end{aligned}$$

where the last step follows by the induction hypothesis. This completes the proof that $\widehat{\mathrm{FDP}}_{\mathrm{Batch_{BH}}}$ is controlled.

Monotonicity now follows by observing that the test levels updates are increasing in the rejection counts R_i, as well as previous test levels. The only exception is α_2 which is decreasing in α_1, however because α_1 is non-random, this does not hurt monotonicity. ☐

8.2 Batch$_{\text{St-BH}}$ Rules

Algorithm 7 One version of the Batch$_{\text{St-BH}}$ algorithm

input: FDR level α, sequence of constants $\{\lambda_t\}_{t=1}^\infty$, non-increasing sequences $\{\gamma_t\}_{t=1}^\infty$ and $\{\gamma'_t\}_{t=1}^\infty$ summing to 1, initial wealth $W_0 \leq \alpha$

Set $\alpha_1 = \gamma_1 \frac{W_0}{n_1}$

for $t = 1, 2, \ldots$ **do**

Run the Storey-BH procedure under level α_t with parameter λ_t on batch $\mathbf{P}_t$

Set $\alpha_{t+1} = \frac{1}{n_{t+1}}\Big(\gamma_{t+1}W_0 + \alpha\sum_{s=1}^t \gamma_{t+1-s}R_s - W_0\sum_{s=1}^t\gamma_{t+1-s}\mathbf{1}\{s=\tau_1\} + \sum_{s=1}^t\gamma'_{t+1-s}(1-k_s)n_s\alpha_s\Big)$, where $\tau_1 = \min\{s\geq 1 : R_s > 0\}$

end

Below we prove that the update rule of Algorithm 7 is monotone and satisfies Definition 2.

Fact 5. The algorithm given in Algorithm 7 is monotone and guarantees $\widehat{\mathrm{FDP}}_{\mathrm{Batch_{St\text{-}BH}}}(t) \leq \alpha$.

Proof. First we prove that the algorithm guarantees $\widehat{\mathrm{FDP}}_{\mathrm{Batch_{St\text{-}BH}}}(t) \leq \alpha$.

It is not hard to see that $\widehat{\mathrm{FDP}}_{\mathrm{Batch_{St\text{-}BH}}}(t) \leq \frac{\sum_{j=1}^t n_j\alpha_j k_j}{1\vee\sum_{j=1}^t R_j}$. Therefore, it suffices to prove $\sum_{j=1}^t n_j\alpha_jk_j \leq \alpha(1\vee \sum_{j=1}^t R_j)$ for all t.

For all t, define $s(t) := \gamma_t W_0 + \alpha\sum_{j=1}^{t-1}\gamma_{t-j}R_j - W_0\sum_{j=1}^{t-1}\gamma_{t-j}\mathbf{1}\{j=\tau_1\}$. With this, the test levels are equal to $n_t\alpha_t = s(t) + \sum_{j=1}^{t-1}\gamma'_{t-j}(1-k_j)n_j\alpha_j$. In Fact 2 we have proved that $\sum_{j=1}^t s(j) \leq \alpha(1\vee\sum_{j=1}^t R_j)$, so it suffices to prove $\sum_{j=1}^t n_j\alpha_jk_j \leq \sum_{j=1}^t s(j)$. We do so by peeling terms off one by one:

$$\begin{aligned}\sum_{j=1}^t n_j\alpha_jk_j &\leq \sum_{j=1}^{t-1} n_j\alpha_jk_j + n_t\alpha_t\\ &= \sum_{j=1}^{t-1} n_j\alpha_jk_j + s(t) + \sum_{j=1}^{t-1}\gamma'_{t-j}(1-k_j)n_j\alpha_j\\ &\leq \sum_{j=1}^{t-2} n_j\alpha_jk_j + s(t) + \sum_{j=1}^{t-2}\gamma'_{t-j}(1-k_j)n_j\alpha_j + n_{t-1}\alpha_{t-1}.\end{aligned}$$

By repeating a similar argument recursively we obtain

$$\sum_{j=1}^{t} n_j \alpha_j k_j \leq \sum_{j=1}^{t} s(j).$$

Invoking Fact 2 now completes the proof that $\widehat{\mathrm{FDP}}_{\mathrm{Batch}_{\text{St-BH}}}(t) \leq \alpha$ for all $t \in \mathbb{N}$.

Now we prove that the update rule is additionally monotone. Take two sequences of p-values such that $(P_{1,1}, P_{1,2}, \ldots, P_{t,n_t}) \geq (\tilde{P}_{1,1}, \tilde{P}_{1,2}, \ldots, \tilde{P}_{t,n_t})$ coordinate-wise. Denote all relevant $\mathrm{Batch}_{\text{St-BH}}$ quantities on these two sequences using a similar notation, for example we distinguish between k_i on $(P_{1,1}, P_{1,2}, \ldots, P_{t,n_t})$ and $\tilde{k}_i$ on $(\tilde{P}_{1,1}, \tilde{P}_{1,2}, \ldots, \tilde{P}_{t,n_t})$.

The first observation is that $\tilde{k}_i \leq k_i$, for all $i \in [t]$. This follows because $\mathbf{1}\left\{\tilde{P}_{i,j} > \lambda_i\right\} \leq \mathbf{1}\left\{P_{i,j} > \lambda_i\right\}$ and $\mathbf{1}\left\{\tilde{P}_{i,\tilde{\max}_i} > \lambda_i\right\} \leq \mathbf{1}\left\{P_{i,\max_i} > \lambda_i\right\}$, so

$$\tilde{k}_i = \frac{\sum_{j=1}^{n_i} \mathbf{1}\left\{\tilde{P}_{i,j} > \lambda_i\right\}}{1 + \sum_{j=1}^{n_i} \mathbf{1}\left\{\tilde{P}_{i,j} > \lambda_i\right\} - \mathbf{1}\left\{\tilde{P}_{i,\tilde{\max}_i} > \lambda_i\right\}} \leq \frac{\sum_{j=1}^{n_i} \mathbf{1}\left\{P_{i,j} > \lambda_i\right\}}{1 + \sum_{j=1}^{n_i} \mathbf{1}\left\{P_{i,j} > \lambda_i\right\} - \mathbf{1}\left\{P_{i,\max_i} > \lambda_i\right\}} = k_i.$$

The rest of the proof follows by combining the monotonicity proof in Fact 2 and the fact that the update for α_t is non-increasing in $(k_1, \ldots, k_{t-1})$. □

As for the $\mathrm{Batch}_{\mathrm{BH}}$ family of algorithms, we also propose a rule with provable guarantees when $n_j \geq M$ for all j, which is expected to be more powerful when the batch sizes are roughly of the same order; moreover, if $n_j \equiv M$, the rule is strictly more powerful than the one above.

Algorithm 8 One version of the $\mathrm{Batch}_{\text{St-BH}}$ algorithm when $n_s \geq M$ for all s

input: FDR level α, non-increasing sequences $\{\gamma_t\}_{t=1}^{\infty}$ and $\{\gamma'_t\}_{t=1}^{\infty}$ summing to one
Set $\alpha_1 = \gamma_1 \frac{M}{n_1} \alpha$
for $t = 1, 2, \ldots$ **do**
  Run the Storey-BH procedure under level α_t with parameter λ_t on batch $\mathbf{P}_t$
  Set $\alpha_{t+1} = \frac{M}{n_{t+1}} \gamma_{t+1} \alpha + \frac{\alpha}{n_{t+1}} \sum_{s=1}^{t} \gamma_{t+1-s} R_s^{\mathrm{add}} + \frac{1}{n_{t+1}} \sum_{s=1}^{t} \gamma'_{t+1-s} (1 - k_s) n_s \alpha_s$,
  where $R_s^{\mathrm{add}} = \min\{R_s, \max\{R_r : r < s\}\}$
end

Fact 6. The algorithm given in Algorithm 8 is monotone and guarantees $\widehat{\mathrm{FDP}}_{\mathrm{Batch}_{\text{St-BH}}}(t) \leq \alpha$.

Proof. First we prove that it guarantees $\widehat{\mathrm{FDP}}_{\mathrm{Batch}_{\text{St-BH}}}(t) \leq \alpha$.

It is not hard to see that $\widehat{\mathrm{FDP}}_{\mathrm{Batch}_{\text{St-BH}}}(t) \leq \frac{\sum_{j=1}^{t} n_j \alpha_j k_j}{M + \sum_{j=1}^{t} R_j^{\mathrm{add}}}$ (recall that $R_1^{\mathrm{add}} = 0$ by definition). Therefore, it suffices to prove $\sum_{j=1}^{t} n_j \alpha_j k_j \leq \alpha(M + \sum_{j=1}^{t} R_j^{\mathrm{add}})$.

If we denote $s(t) := M\gamma_t \alpha + \alpha \sum_{j=1}^{t-1} \gamma_{t-j} R_j^{\mathrm{add}}$, by the same argument as in Fact 5 we can conclude that

$$\sum_{j=1}^{t} n_j \alpha_j k_j \leq \sum_{j=1}^{t} s(j).$$

Following the steps of Fact 3, we can also show that $\sum_{j=1}^{t} s(j) \leq \alpha(M + \sum_{j=1}^{t} R_j^{\mathrm{add}})$, which completes the proof that $\widehat{\mathrm{FDP}}_{\mathrm{Batch}_{\text{St-BH}}}(t) \leq \alpha$ for all $t \in \mathbb{N}$.

The proof that the update rule is additionally monotone combines the monotonicity proofs of Fact 3 and Fact 5. □

9 POWER AND FDR IN NUMERICAL EXPERIMENTS AGAINST TIME

For the experiments presented in Section 5, we plot the power and FDR as functions of time, for $\pi_1 \in \{0.1, 0.5\}$. We observe both power and FDR to be stable across time for our default algorithms.

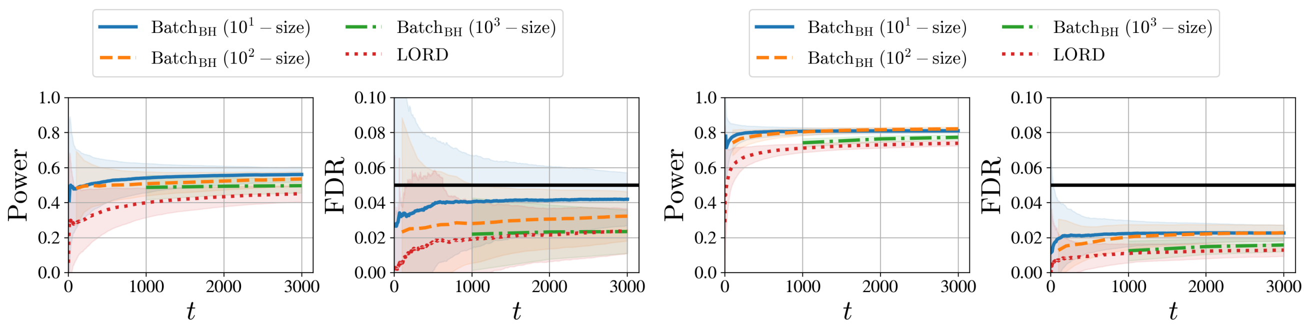

Figure 10: Statistical power and FDR versus number of hypotheses seen t for Batch_{BH} (at batch sizes 10, 100, and 1000) and LORD. We choose the probability of a non-null hypothesis to be $\pi_1 = 0.1$ (left) and $\pi_1 = 0.5$ (right). The observations under the null are $N(0, 1)$, and the observations under the alternative are $N(3, 1)$.

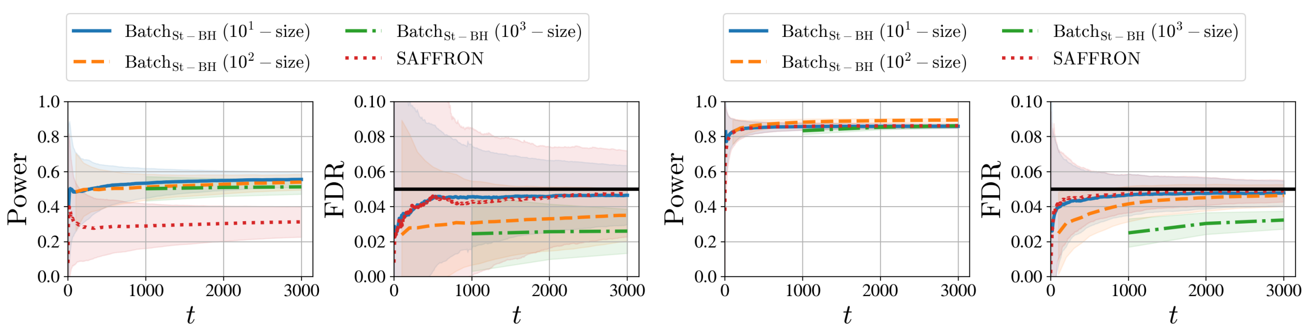

Figure 11: Statistical power and FDR versus number of hypotheses seen t for $\text{Batch}_{\text{St-BH}}$ (at batch sizes 10, 100, and 1000) and SAFFRON. We choose the probability of a non-null hypothesis to be $\pi_1 = 0.1$ (left) and $\pi_1 = 0.5$ (right). The observations under the null are $N(0, 1)$, and the observations under the alternative are $N(3, 1)$.

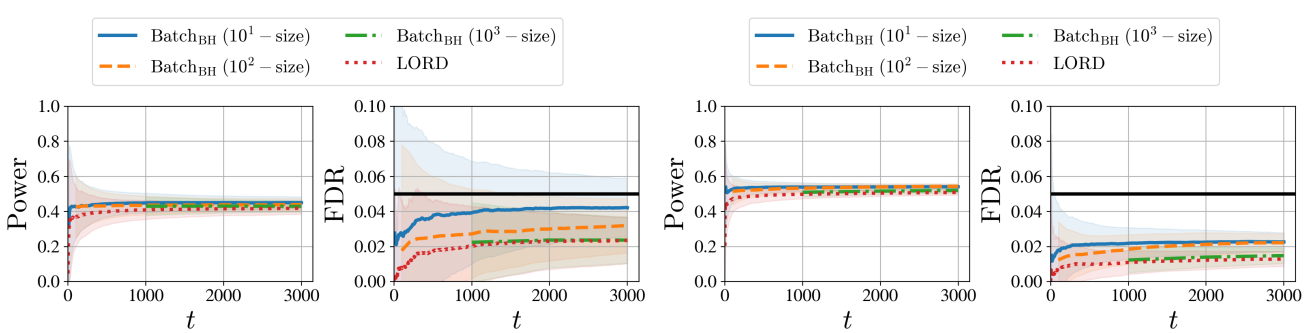

Figure 12: Statistical power and FDR versus number of hypotheses seen t for Batch_{BH} (at batch sizes 10, 100, and 1000) and LORD. We choose the probability of a non-null hypothesis to be $\pi_1 = 0.1$ (left) and $\pi_1 = 0.5$ (right). The observations under the null are $N(0, 1)$, and the observations under the alternative are $N(\mu_1, 1)$ where $\mu_1 \sim N(0, 2 \log T)$.

Additionally, in Figure 14 we plot $R_t^+ - R_t$ for a single trial of Batch_{BH} and the first experimental setting of constant Gaussian means, at $\pi_1 = 0.1$. We observe similar behavior for $\text{Batch}_{\text{St-BH}}$ and other problem parameters

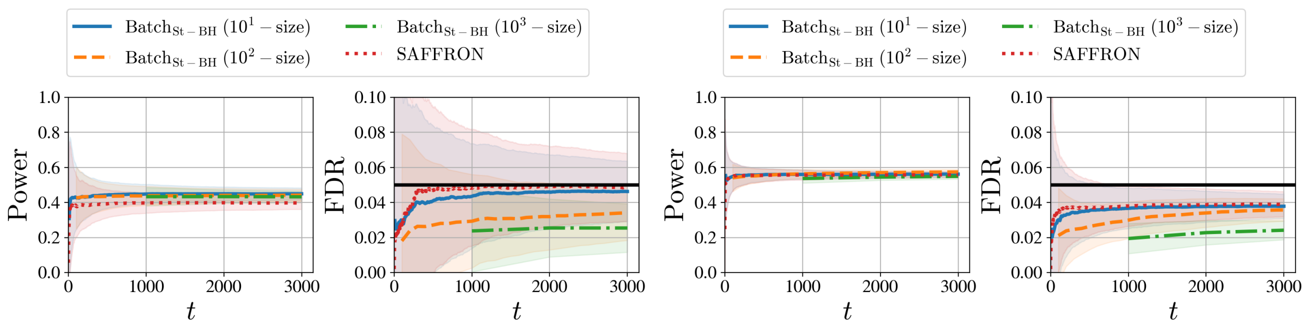

Figure 13: Statistical power and FDR versus number of hypotheses seen t for Batch$_{\text{St-BH}}$ (at batch sizes 10, 100, and 1000) and SAFFRON. We choose the probability of a non-null hypothesis to be $\pi_1 = 0.1$ (left) and $\pi_1 = 0.5$ (right). The observations under the null are $N(0, 1)$, and the observations under the alternative are $N(\mu_1, 1)$ where $\mu_1 \sim N(0, 2 \log T)$.

as well. This experiment shows that $R_t^+ - R_t$ highly concentrates around the value 1, and in our experiments is no larger than 4. Hence, when designing new practical algorithms, it is a reasonable heuristic to assume $R_t^+ = R_t + 1$.

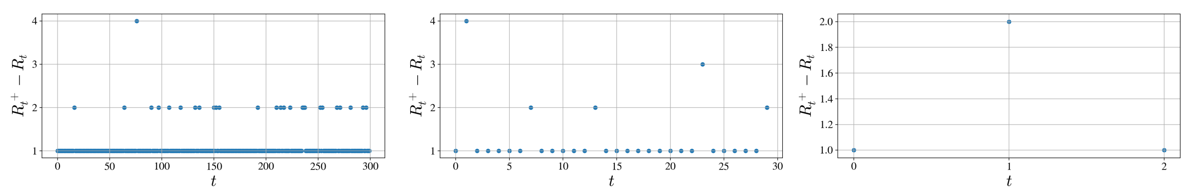

Figure 14: $R_t^+ - R_t$ versus batch index t for Batch$_{\text{BH}}$, at batch sizes 10 (left), 100 (middle) and 1000 (right). We choose the probability of a non-null hypothesis to be $\pi_1 = 0.1$. The observations under the null are $N(0, 1)$, and the observations under the alternative are $N(3, 1)$.

10 MONOTONICITY IN NUMERICAL EXPERIMENTS

We verify numerically that Batch$_{\text{BH}}$ and Batch$_{\text{St-BH}}$ are monotone with high probability, as required by Theorem 1 and Theorem 2. Although this is a heuristic way to justify the FDR control of our procedures, we found that both Batch$_{\text{BH}}$ and Batch$_{\text{St-BH}}$ exhibit monotonicity with high probability, as well as FDR control, across various problem settings.

For a given p-value sequence, we first run either Batch$_{\text{BH}}$ (or Batch$_{\text{St-BH}}$) as usual. We then randomly pick a batch i and set a random p-value in that batch to 0. Finally, we run Batch$_{\text{BH}}$ (or Batch$_{\text{St-BH}}$) again on the modified p-value sequence and check whether the condition $\sum_{j=i+1}^{t} R_j \leq \sum_{j=i+1}^{t} \tilde{R}_j$ holds, where $\tilde{R}_j$ is the number of rejections in the j-th batch of the sequence in which the fixed p-value is set to 0. If we find that the condition holds, then we deem Batch$_{\text{BH}}$ (or Batch$_{\text{St-BH}}$) to be monotone on the given p-value sequence.

We do this for every p-value sequence created in Section 5.1 and Section 5.2. This means that for each of the experimental settings, we perform this monotonicity check on 500 p-value sequences for each π_1 in $\{0.01, 0.02, \ldots, 0.09\} \cup \{0.1, 0.2, \ldots, 0.5\}$. For the experimental setting in Section 5.1, Figure 15 shows that Batch$_{\text{BH}}$ is monotone on at least 97.4% of the sequences, and that Batch$_{\text{St-BH}}$ is monotone on at least 96.6% of the sequences. For the experimental setting in Section 5.2, Figure 16 shows that Batch$_{\text{BH}}$ is monotone on at least 99.0% of the sequences, and that Batch$_{\text{St-BH}}$ is monotone on at least 98.2% of the sequences.

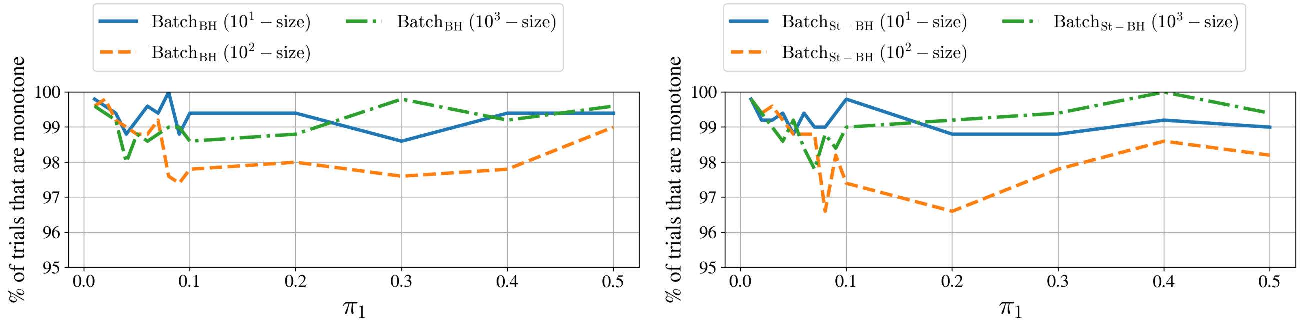

Figure 15: For Batch$_{\text{BH}}$, the minimum is 97.4%. For Batch$_{\text{St-BH}}$, the minimum is 96.6%.

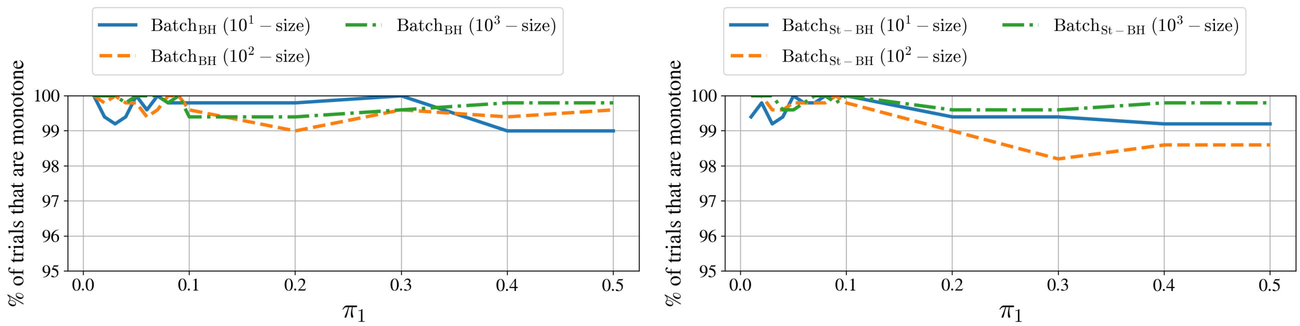

Figure 16: For Batch$_{\text{BH}}$, the minimum is 99.0%. For Batch$_{\text{St-BH}}$, the minimum is 98.2%.

11 mFDR CONTROL OF THE BH PROCEDURE

Recall the definition of *modified*, or *marginal*, false discovery rate up to time t:

$$\mathrm{mFDR}(t) = \frac{\mathbb{E}\left[\sum_{i=1}^t |\mathcal{H}_i^0 \cap \mathcal{R}_i|\right]}{\mathbb{E}\left[\sum_{i=1}^t |\mathcal{R}_i|\right]}.$$

As discussed in Section 6, mFDR is a desirable false discovery metric due to its composition properties; ensuring mFDR control under level α in two disjoint batches of hypotheses guarantees mFDR at most α when the two batches of results are merged. It is thus natural to analyze mFDR control of the BH procedure. Unfortunately, it is not difficult to see that the BH algorithm does not imply mFDR control, although it does provide control asymptotically (Genovese and Wasserman, 2002). Below we present a result of possibly independent interest, which shows that mFDR can be upper bounded in terms of the stability of the number of rejections. Our result implies favorable properties as the batch size tends to infinity, however it has been noted that the rejection set might be highly unstable for finite batch sizes (Gordon et al., 2007).

Proposition 1. Let the p-values $P_1, \dots, P_n$ be independent. Denote by $\mathcal{R}$ the set of indices corresponding to the discoveries in the batch, and let $R = |\mathcal{R}|$. Then, the Benjamini-Hochberg procedure at level α satisfies

$$\mathrm{mFDR} \leq \max\{1, \delta\}\alpha,$$

where $\delta := \sup_{i \in \mathcal{H}^0} \frac{\mathbb{E}[R \mid P_i \in \mathcal{R}]}{\mathbb{E}[R \mid P_i \notin \mathcal{R}]}$.

Proof. Let $\mathcal{H}^0$ denote the nulls in $[n]$. Let the order statistic corresponding to $\mathbf{P} := \{P_1, \dots, P_n\}$ be $P_{(1)}, \dots, P_{(n)}$. Denote by $\mathbf{P}^{(-i)}$ the set $\mathbf{P} \setminus P_i$, and let $P_{(j)}^{(-i)}$ be the j-th order statistic in $\mathbf{P}^{(-i)}$. Define $R^{(-i)}$ to be the number of rejections when running *modified* BH on $\mathbf{P}^{(-i)}$, which rejects the smallest $R^{(-i)}$ p-values in $\mathbf{P}^{(-i)}$, where

$R^{(-i)} = \max\{1 \leq j \leq n-1 : P^{(-i)}_{(j)} \leq \frac{\alpha}{n}(j+1)\}$. For any $i, r \in [n]$, we have:

$$\begin{aligned}
\mathbf{1}\left\{P_i \leq \frac{\alpha}{n} r, R = r\right\} &= \mathbf{1}\left\{P_i \leq \frac{\alpha}{n} r, P_{(r)} \leq \frac{\alpha}{n} r, P_{(r+1)} > \frac{\alpha}{n}(r+1), \ldots, P_{(n)} > \frac{\alpha}{n} n\right\} \\
&= \mathbf{1}\left\{P_i \leq \frac{\alpha}{n} r, P^{(-i)}_{(r-1)} \leq \frac{\alpha}{n} r, P^{(-i)}_{(r)} > \frac{\alpha}{n}(r+1), \ldots, P^{(-i)}_{(n-1)} > \frac{\alpha}{n} n\right\} \\
&= \mathbf{1}\left\{P_i \leq \frac{\alpha}{n} r, R^{(-i)} = r-1\right\}.
\end{aligned}$$

In words, if BH makes r discoveries and a p-value P_i is in the rejected set, then the modified BH ran on the set that drops P_i will make *exactly* $r-1$ discoveries. Denote by V the number of of false discoveries in $\mathcal{R}$. We can express it as:

$$\begin{aligned}
V &= \sum_{i \in \mathcal{H}^0} \mathbf{1}\left\{P_i \leq \frac{\alpha}{n} R, R > 0\right\} \\
&= \sum_{i \in \mathcal{H}^0} \sum_{r=1}^{n} \mathbf{1}\left\{P_i \leq \frac{\alpha}{n} r, R = r\right\} \\
&= \sum_{i \in \mathcal{H}^0} \sum_{r=1}^{n} \mathbf{1}\left\{P_i \leq \frac{\alpha}{n} r, R = r, P_i \in \mathcal{R}\right\} \\
&= \sum_{i \in \mathcal{H}^0} \sum_{r=1}^{n} \mathbf{1}\left\{P_i \leq \frac{\alpha}{n} r, R^{(-i)} = r-1, P_i \in \mathcal{R}\right\}.
\end{aligned}$$

The third equality follows because the event $\{P_i \in \mathcal{R}\}$ is implied by the event $\{P_i \leq \frac{\alpha}{n} r, R = r\}$, and the last equality just uses the first derivation in this proof. By the super-uniformity of null p-values, we have

$$\mathbb{P}\left\{P_i \leq \frac{\alpha}{n} r, R^{(-i)} = r-1, P_i \in \mathcal{R}\right\} \leq \frac{\alpha}{n} r \mathbb{P}\left\{R^{(-i)} = r-1 \;\middle|\; P_i \leq \frac{\alpha}{n} r, P_i \in \mathcal{R}\right\},$$

where we use the trivial bound $\mathbb{P}\left\{P_i \in \mathcal{R} \;\middle|\; P_i \leq \frac{\alpha}{n} r\right\} \leq 1$. If the p-values are independent, then

$$\mathbb{P}\left\{R^{(-i)} = r-1 \;\middle|\; P_i \leq \frac{\alpha}{n} r, P_i \in \mathcal{R}\right\} = \mathbb{P}\left\{R^{(-i)} = r-1 \;\middle|\; P_i \in \mathcal{R}\right\}.$$

Combining the previous steps, we conclude

$$\begin{aligned}
\mathbb{E}[V] &= \sum_{i \in \mathcal{H}^0} \sum_{r=1}^{n} \mathbb{P}\left\{P_i \leq \frac{\alpha}{n} r, R^{(-i)} = r-1, P_i \in \mathcal{R}\right\} \\
&\leq \sum_{i \in \mathcal{H}^0} \sum_{r=1}^{n} \frac{\alpha}{n} r \mathbb{P}\left\{R^{(-i)} = r-1 \;\middle|\; P_i \in \mathcal{R}\right\} \\
&= \frac{\alpha}{n} \sum_{i \in \mathcal{H}^0} \sum_{r=1}^{n} (r-1+1) \mathbb{P}\left\{R^{(-i)} = r-1 \;\middle|\; P_i \in \mathcal{R}\right\} \\
&= \frac{\alpha}{n} \sum_{i \in \mathcal{H}^0} \left(\mathbb{E}\left[R^{(-i)} \;\middle|\; P_i \in \mathcal{R}\right] + \sum_{r=1}^{n} \mathbb{P}\left\{R^{(-i)} = r-1 \;\middle|\; P_i \in \mathcal{R}\right\}\right).
\end{aligned}$$

By the tower property and the first derivation in this proof,

$$\mathbb{E}\left[R^{(-i)} \;\middle|\; P_i \in \mathcal{R}\right] = \mathbb{E}\left[\mathbb{E}\left[R^{(-i)} \;\middle|\; R, P_i \in \mathcal{R}\right]\right] = \mathbb{E}\left[R-1 \mid P_i \in \mathcal{R}\right].$$

Also, due to $\sum_{r=1}^{n} \mathbb{P}\left\{R^{(-i)} = r-1 \;\middle|\; P_i \in \mathcal{R}\right\} = 1$:

$$\mathbb{E}[V] \leq \frac{\alpha}{n} \sum_{i \in \mathcal{H}^0} \mathbb{E}\left[R \mid P_i \in \mathcal{R}\right].$$

Denote by $\epsilon_i := \max\{\mathbb{E}[R \mid P_i \in \mathcal{R}] - \mathbb{E}[R \mid P_i \notin \mathcal{R}], 0\}$. Then

$$\begin{aligned}\mathbb{E}[R] &= \mathbb{P}\{P_i \in \mathcal{R}\}\,\mathbb{E}[R \mid P_i \in \mathcal{R}] + \mathbb{P}\{P_i \notin \mathcal{R}\}\,\mathbb{E}[R \mid P_i \notin \mathcal{R}] \\ &\geq \mathbb{E}[R \mid P_i \in \mathcal{R}] - \epsilon_i \mathbb{P}\{P_i \notin \mathcal{R}\} \\ &\geq \mathbb{E}[R \mid P_i \in \mathcal{R}] - \epsilon_i.\end{aligned}$$

Therefore, $\mathbb{E}[R \mid P_i \in \mathcal{R}] \leq \mathbb{E}[R] + \epsilon_i$, and with this we can conclude

$$\mathbb{E}[V] \leq \frac{\alpha}{n} \sum_{i \in \mathcal{H}^0} (\mathbb{E}[R] + \epsilon_i) \leq \alpha(\mathbb{E}[R] + \max_i \epsilon_i).$$

Define $i^* := \arg\max_i \epsilon_i$. Rearranging the terms in the previous expression, we have

$$\frac{\mathbb{E}[V]}{\mathbb{E}[R] + \epsilon_{i^*}} = \frac{\mathbb{E}[V]}{\mathbb{E}[R]} \left(\frac{1}{1 + \frac{\epsilon_{i^*}}{\mathbb{E}[R]}} \right) \leq \alpha. \tag{3}$$

Now consider the term $\frac{\epsilon_{i^*}}{\mathbb{E}[R]}$. It is strictly positive if and only if $\mathbb{E}[R \mid P_{i^*} \in \mathcal{R}] > \mathbb{E}[R \mid P_{i^*} \notin \mathcal{R}]$, and so the maximizer in $\epsilon_{i^*} = \max\{\mathbb{E}[R \mid P_{i^*} \in \mathcal{R}] - \mathbb{E}[R \mid P_{i^*} \notin \mathcal{R}], 0\}$ is the first term if and only if $\mathbb{E}[R \mid P_{i^*} \in \mathcal{R}] > \mathbb{E}[R \mid P_{i^*} \notin \mathcal{R}]$. Now suppose this indeed holds; then, since $\mathbb{E}[R]$ is a convex combination of $\mathbb{E}[R \mid P_{i^*} \in \mathcal{R}]$ and $\mathbb{E}[R \mid P_{i^*} \notin \mathcal{R}]$, we have

$$\frac{\epsilon_{i^*}}{\mathbb{E}[R]} \leq \frac{\mathbb{E}[R \mid P_{i^*} \in \mathcal{R}] - \mathbb{E}[R \mid P_{i^*} \notin \mathcal{R}]}{\mathbb{E}[R \mid P_{i^*} \notin \mathcal{R}]} = \frac{\mathbb{E}[R \mid P_{i^*} \in \mathcal{R}]}{\mathbb{E}[R \mid P_{i^*} \notin \mathcal{R}]} - 1.$$

All previous observations combined, we can conclude that

$$1 + \frac{\epsilon_{i^*}}{\mathbb{E}[R]} \leq 1 + \max\left\{0, \frac{\mathbb{E}[R \mid P_{i^*} \in \mathcal{R}]}{\mathbb{E}[R \mid P_{i^*} \notin \mathcal{R}]} - 1\right\} = \max\left\{1, \frac{\mathbb{E}[R \mid P_{i^*} \in \mathcal{R}]}{\mathbb{E}[R \mid P_{i^*} \notin \mathcal{R}]}\right\} := \max\{1, \delta\},$$

where we define $\delta := \sup_{i \in \mathcal{H}^0} \frac{\mathbb{E}[R \mid P_i \in \mathcal{R}]}{\mathbb{E}[R \mid P_i \notin \mathcal{R}]}$. Going back to equation (3), this implies

$$\frac{\mathbb{E}[V]}{\mathbb{E}[R]} \leq \max\{1, \delta\}\alpha,$$

as desired. □

References

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

Gilles Blanchard and Etienne Roquain. Two simple sufficient conditions for FDR control. *Electronic journal of Statistics*, 2:963–992, 2008.

Andrea Dal Pozzolo, Olivier Caelen, Reid A Johnson, and Gianluca Bontempi. Calibrating probability with undersampling for unbalanced classification. In *2015 IEEE Symposium Series on Computational Intelligence*, pages 159–166, 2015.

Christopher Genovese and Larry 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.

Alexander Gordon, Galina Glazko, Xing Qiu, Andrei Yakovlev, et al. Control of the mean number of false discoveries, bonferroni and stability of multiple testing. *The Annals of Applied Statistics*, 1(1):179–190, 2007.

KK Gordon Lan and David L DeMets. Discrete sequential boundaries for clinical trials. *Biometrika*, 70(3): 659–663, 1983.

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

Adel Javanmard and Andrea Montanari. Online rules for control of false discovery rate and false discovery exceedance. *The Annals of Statistics*, 46(2):526–554, 2018.

Aaditya Ramdas, Fanny Yang, Martin Wainwright, and Michael Jordan. Online control of the false discovery rate with decaying memory. In *Advances In Neural Information Processing Systems*, pages 5655–5664, 2017.

Aaditya Ramdas, Tijana Zrnic, Martin Wainwright, and Michael Jordan. SAFFRON: an adaptive algorithm for online control of the false discovery rate. In *Proceedings of the 35th International Conference on Machine Learning*, pages 4286–4294, 2018.

Aaditya K Ramdas, Rina F Barber, Martin J Wainwright, and Michael I Jordan. A unified treatment of multiple testing with prior knowledge using the p-filter. *The Annals of Statistics*, 47(5):2790–2821, 2019.

David S Robertson and James Wason. Online control of the false discovery rate in biomedical research. *arXiv preprint arXiv:1809.07292v1*, 2018.

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

Jinjin Tian and Aaditya Ramdas. ADDIS: adaptive algorithms for online FDR control with conservative nulls. *Advances in Neural Information Processing Systems*, 2019.

Tijana Zrnic, Aaditya Ramdas, and Michael I Jordan. Asynchronous online testing of multiple hypotheses. *arXiv preprint arXiv:1812.05068*, 2018.