Supplemental Materials

The supplementary material is organized as follows. First we provide additional empirical results and discussions in Supp. Section 1 and Supp. Section 2 respectively. Next we present the technical proofs. Specifically, the correctness result Theorem 1 is proved in Supp. Section 3. The instance-wise upper bound Theorem 2 is proved in Supp. Section 4 while the $\tilde{O}(\sqrt{n}m)$ upper bound Proposition 1 is proved in Supp. Section 5. The lower bound Theorem 3 is proved in Supp. Section 6. Finally, the auxiliary lemmas are in Supp. Section 7.

1. Additional Results

Table 1. Small GWAS on chromosome 4.

dbSNP ID	Original p-value	fMC p-value	Rej. at α=0.1	Rej. at α=0.05
rs2242330	1.7e-6	8.0e-6	$\surd$	$\surd$
rs6826751	2.1e-6	1.6e-5	$\surd$	$\surd$
rs4862792	3.5e-5	4.0e-6	$\surd$	$\surd$
rs3775866	4.6e-5	3.6e-5	$\surd$	$\surd$
rs355477	7.9e-5	8.0e-5	$\surd$	$\times$
rs355461	8.3e-5	8.0e-5	$\surd$	$\times$
rs355506	8.3e-5	8.0e-6	$\surd$	$\times$
rs355464	8.9e-5	1.3e-4	$\surd$	$\times$
rs1497430	9.7e-5	5.2e-5	$\surd$	$\surd$
rs11946612	9.7e-5	4.8e-5	$\surd$	$\surd$

Table 2. Small GWAS on chromosome 1-3 (There is no discovery reported on chromosomes 2-3 from the orignal paper).

dbSNP ID	Chromosome	Original p-value	fMC p-value	Rej. at α=0.1
rs988421	1	4.9e-5	3.6e-5	$\surd$
rs1887279	1	5.7e-5	4.4e-5	$\surd$
rs2986574	1	6.3e-5	4.4e-5	$\surd$
rs3010040	1	8.0e-5	6.0e-5	$\surd$
rs2296713	1	8.0e-5	6.0e-5	$\surd$

2. Additional Discussions

2.1. Choosing the parameter for sMC

For sMC the parameter s need to be chosen *a priori*. A back-of-the-envelope calculation shows that for a hypothesis test with the ideal p-value p^∞, the sMC p-value is around $p^\infty \pm \frac{p^\infty}{\sqrt{s}}$ while the fMC p-value is around $p^\infty \pm \sqrt{\frac{p^\infty}{n}}$. Suppose the BH threshold on the ideal p-values is τ^∞. Since it is desirable for the BH result on the MC p-values (sMC, fMC) to be close to the BH result on the ideal p-values, the accuracy of the MC p-values with corresponding ideal p-values close to τ^∞ can be thought of as the accuracy of the entire multiple testing problem. Matching such accuracy for sMC and fMC gives that $s = \tau^\infty n = \frac{r^\infty}{m}\alpha n$. When n=10m and α=0.1, we have that s=r^∞. That is, s should be at least 100 if there are more than 100 discoveries on the ideal p-values. However, since we do not know r^∞ before running the experiment, a larger value is preferred. It is noted that values s=30-120 are recommended in a recent work (Thulin et al., 2014).

2.2. Comparison to bandit FDR

In the bandit FDR setting (Jamieson & Jain, 2018), each arm has a parameter μ_i with $\mu_i = \mu_0$ for null arms and $\mu_i > \mu_0 + \Delta$ for alternative arms, for some μ_0 and $\Delta > 0$ given before the experiment. For arm i, i.i.d. observations are available that are bounded and have expected value μ_i. The goal is to select a subset of arms and the selected set should control FDR while achieving a certain level of power.

Both bandit FDR and AMT aim to select a subset of "good arms" as defined by comparing the arm parameters to a threshold. In bandit FDR this threshold is given as μ_0. In AMT, however, this is the BH threshold that is not known ahead of time and needs to be learned from the observed data. The two frameworks also differ in the error criterion. Bandit FDR considers FDR and power for the selected set, a novel criterion in MAB literature. AMT, on the other hand, adopts the traditional PAC-learning criterion of recovering the fMC discoveries with high probability. These distinctions lead to different algorithms: bandit FDR uses an algorithm similar to thresholding MAB (Locatelli et al., 2016) but with carefully designed confidence bounds to control FDR; AMT devises a new LUCB (lower and upper confidence bound) algorithm that adaptively estimates two things simultaneously: the BH threshold and how each arm compares to the threshold.

2.3. Future works

We have shown that AMT improves the computational efficiency of the fMC workflow, i.e., applying BH on the fMC p-values. A direct extension is to the workflow of applying the Storey-BH procedure (Storey et al., 2004) on the fMC p-values. In addition, in many cases, especially in genetic research, additional covariate information is available for each null hypothesis, e.g., functional annotations of the SNPs in GWAS, where a covariate-dependent rejection threshold can be used to increase testing power (Xia et al., 2017; Zhang et al., 2018). Extending AMT to such cases would allow both efficient computation of MC p-values and increased power via covariate-adaptive thresholding. Last but not least, MC sampling is an important building block in some modern multiple testing approaches like the model-X knockoff (Candes et al., 2018) or the conditional permutation test (Berrett et al., 2018), where ideas in the present paper may be used to improve the computational efficiency.

3. Proof of Theorem 1

Proof. (Proof of Theorem 1) To show (9), it suffices to show that conditional on any set of fMC p-values $\{P_i^{\text{fMC}}\} = \{p_i\}$,

$$\mathbb{P}\left(\mathcal{R}^{\text{AMT}} = \mathcal{R}^{\text{fMC}} \middle| \{P_i^{\text{fMC}}\} = \{p_i\}\right) \geq 1 - \delta. \quad (13)$$

Let $\mathcal{E}$ denote the event that all CBs hold. Since the number of CBs is at most $2mL$ and each of them holds with probability at least $1 - \frac{\delta}{2mL}$ conditional on the fMC p-values, by union bound,

$$\mathbb{P}\left(\mathcal{E} \middle| \{P_i^{\text{fMC}}\} = \{p_i\}\right) \geq 1 - \delta.$$

Next we show that $\mathcal{E}$ implies $\mathcal{R}^{\text{AMT}} = \mathcal{R}^{\text{fMC}}$, which further gives (13). Let T be the total number of rounds, which is finite since at most mn MC samples will be computed. For any round t, let "(t)" represent the corresponding values before the MC sampling of the round, e.g., $\hat{r}(t)$, $\hat{\tau}(t)$, $\mathcal{C}_{\text{g}}(t), \mathcal{C}_{\text{l}}(t), \mathcal{U}(t)$. Also, let $(T+1)$ represent the values at termination. For any $t \in [T+1]$,

1. if $\hat{r}(t) > r^*$, by (5) more than $m - \hat{r}(t)$ fMC p-values are greater than $\hat{\tau}(t)$ whereas $|\mathcal{C}_{\text{g}}(t)| = m - \hat{r}(t)$. Thus, there is at least one hypothesis that has fMC p-value greater than $\hat{\tau}(t)$ and is not in $\mathcal{C}_{\text{g}}(t)$. On $\mathcal{E}$, it cannot be in $\mathcal{C}_{\text{l}}(t)$. Hence, it is in $\mathcal{U}(t)$, giving that $\mathcal{U}(t) \neq \emptyset$. Thus, $t \neq T+1$ and the algorithm will not terminate.
2. if $\hat{r}(t) = r^*$, there are $m - r^*$ hypotheses in $\mathcal{C}_{\text{g}}(t)$ corresponding to those with fMC p-values greater than τ^*. Other hypotheses all have fMC p-values less than τ^* and hence, on $\mathcal{E}$, will not enter $\mathcal{C}_{\text{g}}$ after further sampling. Therefore, $\hat{r}(t)$ will not further decrease.

Therefore, $\hat{r}(T+1) = r^*$. Since $\mathcal{U}(T+1) = \emptyset$, on $\mathcal{E}$, $\mathcal{C}_{\text{l}}(T+1)$ contains all hypotheses with fMC p-values less than τ^*, i.e., $\mathcal{C}_{\text{l}}(T+1) = \mathcal{R}^{\text{fMC}}$. Hence, we have shown (13).

Next we prove FDR control. Let $\text{FDP}(\mathcal{R}^{\text{fMC}})$ and $\text{FDR}(\mathcal{R}^{\text{fMC}})$ denote the false discovery proportion and FDR of the set $\mathcal{R}^{\text{fMC}}$, respectively. It is noted that $\text{FDR}(\mathcal{R}^{\text{fMC}}) = \mathbb{E}[\text{FDP}(\mathcal{R}^{\text{fMC}})]$. Let $\mathcal{E}_1$ denote the event that $\mathcal{R}^{\text{AMT}} = \mathcal{R}^{\text{fMC}}$ and $\mathcal{E}_1^c$ be the complement of $\mathcal{E}_1$. Then $\mathbb{P}(\mathcal{E}_1^c) \leq \delta$ due to (9) that we have just proved. For AMT,

$$\text{FDR}(\mathcal{R}^{\text{AMT}}) = \mathbb{E}[\text{FDP}(\mathcal{R}^{\text{AMT}})] \quad (14)$$
$$= \mathbb{E}[\text{FDP}(\mathcal{R}^{\text{AMT}})|\mathcal{E}_1]\mathbb{P}(\mathcal{E}_1) + \mathbb{E}[\text{FDP}(\mathcal{R}^{\text{AMT}})|\mathcal{E}_1^c]\mathbb{P}(\mathcal{E}_1^c). \quad (15)$$

The first term of (15)

$$\mathbb{E}[\text{FDP}(\mathcal{R}^{\text{AMT}})|\mathcal{E}_1]\mathbb{P}(\mathcal{E}_1) = \mathbb{E}[\text{FDP}(\mathcal{R}^{\text{fMC}})|\mathcal{E}_1]\mathbb{P}(\mathcal{E}_1)$$
$$\leq \mathbb{E}[\text{FDP}(\mathcal{R}^{\text{fMC}})] = \text{FDR}(\mathcal{R}^{\text{fMC}}) \leq \pi_0\alpha,$$

where the last inequality is because the fMC p-values are stochastically greater than the uniform distribution under the null hypothesis, and hence, applying BH on them controls FDR at level $\pi_0\alpha$.

The second term of (15) is upper bounded by δ as FDP is always no greater than 1. Therefore,

$$\text{FDR}(\mathcal{R}^{\text{AMT}}) \leq \pi_0\alpha + \delta.$$

□

4. Proof of Theorem 2

Proof. (Proof of Theorem 2) The entire analysis is conditional on the fMC p-values $\{P_i^{\text{fMC}}\} = \{p_i\}$. Without loss of generality assume $p_1 \leq p_2 \leq \cdots \leq p_m$. Let T be the total number of rounds, which is finite since at most mn MC samples will be computed. For any round t, let "(t)" represent the corresponding values before the MC sampling of the round. Note that "$(T+1)$" represent the values at termination. The quantities useful to the analysis include

1. $N_i(t)$: number of MC samples for arm i.
2. $p_i^{\text{lb}}(t), p_i^{\text{ub}}(t)$: lower and upper CBs for arm i.
3. Empirical mean $\hat{p}_i(t) = \frac{1}{N_i(t)}\left(1 \vee \sum_{j=1}^{N_i(t)} B_{i,j}\right)$.
4. $\mathcal{C}_{\text{g}}(t), \mathcal{C}_{\text{l}}(t), \mathcal{U}(t)$: hypothesis sets as defined in (6).
5. $\hat{r}(t), \hat{\tau}(t)$: critical rank estimate and the corresponding BH threshold estimate.

Let $\mathcal{E}$ denote the event that all CBs hold. Since the number of CBs is at most $2mL$ and each of them holds with probability at least $1 - \frac{\delta}{2mL}$ conditional on the fMC p-values, by union bound,

$$\mathbb{P}\left(\mathcal{E} \middle| \{P_i^{\text{fMC}}\} = \{p_i\}\right) \geq 1 - \delta.$$

Conditional on $\mathcal{E}$, when the algorithm terminates, $\mathcal{U}(T+1) = \emptyset$. There are $m - r^*$ hypotheses in $\mathcal{C}_{\text{g}}(T+1)$ and r^* hypotheses in $\mathcal{C}_{\text{l}}(T+1)$. We next upper the number of MC samples for hypotheses in these two sets separately.

Step 1. Hypotheses in $\mathcal{C}_{\text{g}}(T+1)$. On $\mathcal{E}$, there are $m - r^*$ hypotheses in $\mathcal{C}_{\text{g}}(T+1)$. For any $i \in [m - r^*]$, let g_i be the ith hypothesis entering $\mathcal{C}_{\text{g}}$. For two hypotheses entering $\mathcal{C}_{\text{g}}$ in the same round, the one is considered entering earlier if it has a larger upper CB p^{ub} before the MC sampling in the entering round.

Consider any g_i that enters after MC sampling in round t_i and let g_j be the first hypothesis entering $\mathcal{C}_{\text{g}}$ in the same round. Here, we note that $t_i = t_j$ and the number of MC samples $N_{g_i}(T+1) = N_{g_j}(T+1)$. In addition,

$$N_{g_j}(T+1) = N_{g_j}(t_j+1) \le (1+\gamma)N_{g_j}(t_j), \quad (16)$$

since the batch sizes is a geometric sequence with ratio γ. Now we focus on $N_{g_j}(t_j)$.

Since g_j is sampled in round t_j, we have that $g_j \notin \mathcal{C}_\text{g}(t_j)$. This indicates that in round t_j, the lower CB of g_j should be no greater than the estimated threshold $\hat{\tau}(t_j)$ before MC sampling; otherwise g_j would have entered $\mathcal{C}_\text{g}$ before round t_j. Hence,

$$p^{\text{lb}}_{g_j}(t_j) \le \hat{\tau}(t_j). \quad (17)$$

Also, being the first to enter $\mathcal{C}_\text{g}$ in round t_j, its upper CB is the largest among all elements in $\mathcal{U}(t_j)$, i.e.,

$$p^{\text{ub}}_{g_j}(t_j) = \max_{k\in\mathcal{U}(t_j)} p^{\text{ub}}_k(t_j). \quad (18)$$

Subtracting (17) from (18) to have the width of the confidence interval

$$\begin{aligned} p^{\text{ub}}_{g_j}(t_j) - p^{\text{lb}}_{g_j}(t_j) &\ge \max_{k\in\mathcal{U}(t_j)} p^{\text{ub}}_k(t_j) - \hat{\tau}(t_j) \\ &\ge \max_{k\in\mathcal{U}(t_j)} p_k - \hat{\tau}(t_j), \end{aligned} \quad (19)$$

where the last inequality is conditional on $\mathcal{E}$. Since $|\mathcal{C}_\text{g}(t_j)| = j-1$, we have that $\max_{k\in\mathcal{U}(t_j)} p_k \ge p_{m-j+1}$. Therefore (19) can be further written as

$$p^{\text{ub}}_{g_j}(t_j) - p^{\text{lb}}_{g_j}(t_j) \ge p_{m-j+1} - \hat{\tau}(t_j) = \Delta_{m-j+1}. \quad (20)$$

Since the CBs satisfy (7), equations (17) and (20) can be rewritten as

$$\begin{aligned} &\hat{p}_{g_j}(t_j) - \sqrt{\frac{c\left(\frac{\delta}{2mL}\right)\hat{p}_{g_j}(t_j)}{T_{g_j}(t_j)}} \le \hat{\tau}(t_j), \\ &2\sqrt{\frac{c\left(\frac{\delta}{2mL}\right)\hat{p}_{g_j}(t_j)}{T_{g_j}(t_j)}} \ge \Delta_{m-j+1}. \end{aligned} \quad (21)$$

Note that $\hat{\tau}(t_j) = \frac{m-j+1}{m}\alpha$. By Lemma 1,

$$\begin{aligned} N_{g_j}(t_j) &\le \frac{4c\left(\frac{\delta}{2mL}\right)\left(\frac{m-j+1}{m}\alpha + \frac{\Delta_{m-j+1}}{2}\right)}{\Delta^2_{m-j+1}} \quad (22) \\ &\le \frac{4c\left(\frac{\delta}{2mL}\right)p_{m-j+1}}{\Delta^2_{m-j+1}}. \quad (23) \end{aligned}$$

Since $i \ge j$, we have that $m-j+1 \ge m-i+1$. Therefore.

$$\begin{aligned} \mathbb{E}[N_{g_i}(T+1)|\mathcal{E}] &\le (1+\gamma)\mathbb{E}[N_{g_i}(t_i)|\mathcal{E}] \\ &\le (1+\gamma)\frac{4c\left(\frac{\delta}{2mL}\right)p_{m-j+1}}{\Delta^2_{m-j+1}} \\ &\le \max_{k\ge m-i+1} \frac{4(1+\gamma)c\left(\frac{\delta}{2mL}\right)p_k}{\Delta^2_k}. \end{aligned} \quad (24)$$

Step 2. Hypotheses in $\mathcal{C}_\text{l}(T+1)$. On $\mathcal{E}, \mathcal{C}_\text{l}(T+1) = \mathcal{R}^{\text{fMC}}$ and $\hat{\tau}(T+1) = \tau^*$. Consider any hypothesis $i \in \mathcal{C}_\text{l}(T+1)$ whose fMC p-value is $p_i \le \tau^*$. It will be sampled until its upper CB is no greater than τ^*. Let its last sample round be t_i. Then,

$$p^{\text{ub}}_{g_i}(t_i) > \tau^*, \quad p^{\text{ub}}_{g_i}(t_i+1) \le \tau^*, \quad p^{\text{lb}}_{g_i}(t_i) \le p_i. \quad (25)$$

Subtracting the third term from the first term yields

$$p^{\text{ub}}_{g_i}(t_i) - p^{\text{lb}}_{g_i}(t_i) > \Delta_i. \quad (26)$$

Since the CBs satisfy (7), the second term in (25) along with (26) can be rewritten as

$$\begin{aligned} &\hat{p}_i(t_i+1) + \sqrt{\frac{c\left(\frac{\delta}{2mL}\right)\hat{p}_i(t_i+1)}{N_i(t_i+1)}} \le \tau^*, \\ &2\sqrt{\frac{c\left(\frac{\delta}{2mL}\right)\hat{p}_i(t_i)}{N_i(t_i)}} > \Delta_i. \end{aligned} \quad (27)$$

Note that $N_i(t_i+1) \le (1+\gamma)N_i(t_i)$ and $\hat{p}_i(t_i+1) \ge \frac{1}{1+\gamma}\hat{p}_i(t_i)$, (27) can be further written as

$$\begin{aligned} &\hat{p}_i(t_i) + \sqrt{\frac{c\left(\frac{\delta}{2mL}\right)\hat{p}_i(t_i)}{N_i(t_i)}} \le (1+\gamma)\tau^* \\ &2\sqrt{\frac{c\left(\frac{\delta}{2mL}\right)\hat{p}_i(t_i)}{N_i(t_i)}} > \Delta_i. \end{aligned} \quad (28)$$

Furthermore,

$$N_i(t_i) \le \frac{4(1+\gamma)c\left(\frac{\delta}{2mL}\right)\tau^*}{\Delta^2_i}. \quad (29)$$

and the number of MC samples for hypothesis i

$$\begin{aligned} \mathbb{E}[N_i(T+1)|\mathcal{E}] &\le (1+\gamma)\mathbb{E}[N_i(t_i)|\mathcal{E}] \\ &\le \frac{4(1+\gamma)^2 c\left(\frac{\delta}{2mL}\right)\tau^*}{\Delta^2_i}. \end{aligned} \quad (30)$$

Step 3. Combine the result. Finally, noting that a hypothesis can be at most sampled n times, the total expected MC samples

$$\begin{aligned} \mathbb{E}[N] &\le \mathbb{E}\left[\sum_{i=1}^m N_i(T+1)\Big|\mathcal{E}\right] + \delta mn \quad (31) \\ &\le \sum_{i=1}^{r^*} n \wedge \left(\frac{4(1+\gamma)^2 c\left(\frac{\delta}{2mL}\right)\tau^*}{\Delta^2_i}\right) \quad (32) \\ &\sum_{i=r^*+1}^{m} n \wedge \left(\max_{k\ge i} \frac{4(1+\gamma)c\left(\frac{\delta}{2mL}\right)p_k}{\Delta^2_k}\right) + \delta mn. \quad (33) \end{aligned}$$

□

5. Proof of Proposition 1

Proof. (Proof of Proposition 1) First let us consider the case where $f(p)$ is continuous and monotonically decreasing. The case where $f(p)=1$ is easy and is dealt with at the end.

Step 0. Notations. Since this proof is an asymptotic analysis, we use subscript "n, m" to denote the quantities for the fMC p-values with n MC samples and m hypotheses. We are interested in the regime where $m \to \infty$ while $n = \Omega(m)$.

For an instance with m hypotheses and n MC samples for each hypothesis, let $\tilde{\tau}_{n,m}$ be the BH threshold and $\tilde{F}_{n,m}$ be the empirical distribution of the fMC p-values $\tilde{F}_{n,m}(x) = \frac{1}{m}\sum_{i=1}^m \mathbb{I}\{P_i^{\text{fMC}} \leq x\}$. Also let $\tilde{f}_{n,m}$ be the probability mass function $\tilde{f}_{n,m}(x) = \frac{1}{m}\sum_{i=1}^m \mathbb{I}\{P_i^{\text{fMC}} = x\}$.

For the distribution of the ideal p-values F, define $g(x) = x - F(x)\alpha$ and let $\tau^* = \sup_{[0,1]}\{\tau : g(\tau) \leq 0\}$. τ^* is actually the BH threshold in the limiting case, as will be shown in Step 2 below. There are a few properties we would like to point out. By definition $g(\tau^*) = 0$. As a result, $F(\tau^*) = \frac{\tau^*}{\alpha}$. Since $f(p)$ is monotonically decreasing, $f(\tau^*) < \frac{F(\tau^*)}{\tau^*} = \frac{1}{\alpha}$. Furthermore, $g'(\tau^*) = 1 - f(\tau^*)\alpha > 0$.

Step 1. $\tilde{F}_{n,m}$ converges uniformly to F. Let F_n be the distribution of the fMC p-values with n MC samples. Then F_n converges uniformly to F. Furthermore, by Glivenko-Cantelli theorem $\tilde{F}_{n,m}$ converges uniformly to F_n. Therefore, $\tilde{F}_{n,m}$ converges uniformly to F.

Step 2. $\tilde{\tau}_{n,m}$ converges in probability to τ^*. For an instance with m hypotheses and n MC samples for each hypothesis, let $\tilde{g}_{n,m}(x) = x - \tilde{F}_{n,m}(x)\alpha$. Then $\tilde{\tau}_{n,m} = \sup_{[0,1]}\{\tau : \tilde{g}_{n,m}(\tau) \leq 0\}$. Since $\tilde{F}_{n,m}$ converges uniformly to F, $\tilde{g}_{n,m}$ converges uniformly to g. Since $g'(\tau^*) > 0$ and is continuous at τ^*, $\exists \epsilon_0 > 0$ such that $g(x)$ is monotonically increasing on $[\tau^* - \epsilon_0, \tau^* + \epsilon_0]$. Since $\tilde{g}_{n,m}$ converges uniformly to g on this interval, for any $0 < \epsilon' < \epsilon$, $\mathbb{P}(|\tilde{\tau}_{n,m} - \tau^*| > \epsilon') \to 0$. Thus, $\tilde{\tau}_{n,m} \xrightarrow{p} \tau^*$.

Step 3. Upper bound $\mathbb{E}[N]$. Let $\delta = \frac{1}{mn}$ and let $\tilde{c}$ denote any log factor (in both m and n) in general. Then for the fMC p-values with n MC samples and m hypotheses, by Theorem 1, and omitting additive constants,

$$\mathbb{E}[N] \leq \tilde{c}\mathbb{E}\left[\sum_{i=1}^{r^*} n \wedge \frac{\tilde{\tau}_{n,m}}{\Delta_{(i)}^2} + \sum_{i=r^*+1}^{m} n \wedge \max_{k \geq i} \frac{P_{(k)}^{\text{fMC}}}{\Delta_{(k)}^2}\right]$$
$$\leq \tilde{c}\mathbb{E}\left[\sum_{i=1}^{r^*} n \wedge \frac{1}{\Delta_{(i)}^2} + \sum_{i=r^*+1}^{m} n \wedge \max_{k \geq i} \frac{1}{\Delta_{(k)}^2}\right]. \tag{34}$$

Notice that $\tilde{F}_{n,m}(P_{(k)}^{\text{fMC}}) \geq \frac{k}{m}$ where the inequality is because there might be several hypotheses with the same value. Therefore for any $P_{(k)}^{\text{fMC}} > \tilde{\tau}_{n,m}$,

$$\frac{1}{\Delta_{(k)}^2} = \frac{1}{\left(P_{(k)}^{\text{fMC}} - \frac{k}{m}\alpha\right)^2}$$
$$\leq \frac{1}{\left(P_{(k)}^{\text{fMC}} - \tilde{F}_{n,m}(P_{(k)}^{\text{fMC}})\alpha\right)^2} = \frac{1}{\tilde{g}_{n,m}(P_{(k)}^{\text{fMC}})^2}.$$

Hence, summing over all possible values of the empirical distribution of the fMC p-values, i.e., $P^{\text{fMC}} = \frac{1}{n+1}, \frac{2}{n+1}, \cdots, 1$ (note the definition of the fMC p-values in (4)), to further write (34) as

$$\mathbb{E}[N] \leq$$
$$\tilde{c}m\mathbb{E}\left[\sum_{i=1}^{\lfloor (n+1)\tilde{\tau}_{n,m}\rfloor} \left(n \wedge \frac{1}{\left(\frac{i}{n+1} - \tilde{\tau}_{n,m}\right)^2}\right)\tilde{f}_{n,m}\left(\frac{i}{n+1}\right)\right.$$
$$\left. + \sum_{i=\lceil (n+1)\tilde{\tau}_{n,m}\rceil}^{n+1} \left(n \wedge \max_{k \geq i} \frac{1}{\tilde{g}_{n,m}(\frac{k}{n+1})^2}\right)\tilde{f}_{n,m}\left(\frac{i}{n+1}\right)\right]. \tag{35}$$

Since $f(x)$ is continuous, $g'(x)$ is also continuous. Recall that $g'(\tau^*) > 0$. Hence, $\exists \epsilon, c_0 > 0$ such that $\forall x \in [\tau^* - \epsilon, 1]$, $g'(x) > c_0$. Recall that $\tilde{g}_{m,n}$ converges uniformly to g and $\tilde{\tau}_{n,m} \xrightarrow{p} \tau^*$. Note that by definition $\tilde{g}_{n,m}(\tilde{\tau}_{n,m}) = 0$. Therefore, $\exists c_1 > 0$ such that for large enough n, m, for any $k \geq \lceil (n+1)\tilde{\tau}_{n,m}\rceil$,

$$\tilde{g}_{n,m}\left(\frac{k}{n+1}\right) = \tilde{g}_{n,m}\left(\frac{k}{n+1}\right) - \tilde{g}_{n,m}(\tilde{\tau}_{n,m}) \tag{36}$$
$$\geq c_1\left(\frac{k}{n+1} - \tilde{\tau}_{n,m}\right). \tag{37}$$

Hence, (35) can be further rewritten as

$$\mathbb{E}[N] \leq$$
$$\tilde{c}m\mathbb{E}\left[\sum_{i=1}^{\lfloor (n+1)\tilde{\tau}_{n,m}\rfloor} \left(n \wedge \frac{1}{\left(\frac{i}{n+1} - \tilde{\tau}_{n,m}\right)^2}\right)\tilde{f}_{n,m}\left(\frac{i}{n+1}\right)\right.$$
$$\left. + \sum_{i=\lceil (n+1)\tilde{\tau}_{n,m}\rceil}^{n+1} \left(n \wedge \frac{1}{c_1^2\left(\frac{i}{n+1} - \tilde{\tau}_{n,m}\right)^2}\right)\tilde{f}_{n,m}\left(\frac{i}{n+1}\right)\right]$$
$$\leq \frac{\tilde{c}}{c_1^2}m\mathbb{E}\left[\sum_{i=1}^{n+1} n \wedge \frac{1}{(\frac{i}{n} - \tilde{\tau}_{n,m})^2}\tilde{f}_{n,m}\left(\frac{i}{n+1}\right)\right]$$
$$= \frac{\tilde{c}}{c_1^2}m\mathbb{E}\left[n \wedge \frac{1}{(P_i^{\text{fMC}} - \tilde{\tau}_{n,m})^2}\right].$$

Since F_n converges uniformly to F and $\tilde{\tau}_{n,m} \xrightarrow{p} \tau^*$, by Slutsky's theorem and the continuous mapping theorem, the RHS will converge to

$$\frac{\tilde{c}}{c_1^2} m \mathbb{E}\left[n \wedge \frac{1}{(P_i^\infty - \tau^*)^2}\right]. \tag{38}$$

Last we evaluation the expectation:

$$\mathbb{E}\left[n \wedge \frac{1}{(P_i^\infty - \tau^*)^2}\right] = \int_0^{\tau^* - \frac{1}{\sqrt{n}}} \frac{1}{(p-\tau^*)^2} dF(p) + \int_{\tau^* - \frac{1}{\sqrt{n}}}^{\tau^* + \frac{1}{\sqrt{n}}} n dF(p) + \int_{\tau^* + \frac{1}{\sqrt{n}}}^{1} \frac{1}{(p-\tau^*)^2} dF(p).$$

By noting that $f(\tau^*) < \frac{1}{\alpha}$ and $f(p)$ is monotonically decreasing it is clear that all three terms are $\tilde{O}(\sqrt{n})$, which concludes the proof of this case.

When $f(p) = 1$, the limiting BH threshold $\tau^* = 0$. Furthermore, $g(x) = (1-\alpha)x$ and $g'(x) = 1 - \alpha > 0$. Therefore, $g(\frac{k}{n+1}) \geq (1-\alpha)(\frac{k}{n+1} - \tilde{\tau}_{n,m})$. Then, similarly we have the total number of MC samples

$$\mathbb{E}[N] \leq \frac{\tilde{c}}{(1-\alpha)^2} m \mathbb{E}\left[n \wedge \frac{1}{(P_i^{\text{fMC}} - \tilde{\tau}_{n,m})^2}\right], \tag{39}$$

which converges to

$$\frac{\tilde{c}}{(1-\alpha)^2} m \mathbb{E}\left[n \wedge \frac{1}{(P_i^\infty)^2}\right] \tag{40}$$

that is $\tilde{O}(\sqrt{n}m)$. □

6. Proof of Theorem 3

Proof. (Proof of Theorem 3) Let F_n be the distribution of the fMC p-values with n MC samples. By Lemma 2, conditional on the fMC p-values $\{P_i^{\text{fMC}}\} = \{p_i\}$, $\exists \delta_0 > 0$, $c_0 > 0$, $c_1 > 0$, s.t. $\forall \delta < \delta_0$, a δ-correct algorithm satisfies

$$\mathbb{E}\left[N \middle| \{P_i^{\text{fMC}}\} = \{p_i\}\right] \geq c_0 n \sum_{i=1}^m \mathbb{I}\{\tau^* < p_i \leq \tau^* + \frac{c_1}{\sqrt{n}}\}. \tag{41}$$

Taking expectation with respect to the fMC p-values to have

$$\mathbb{E}[N] \geq c_0 n m \mathbb{P}\left[\tau^* < P_i^{\text{fMC}} \leq \tau^* + \frac{c_1}{\sqrt{n}}\right]. \tag{42}$$

Since the null fMC p-values follow a uniform distribution,

$$\mathbb{E}[N] \geq c_0 \pi_0 n m \frac{c_1}{\sqrt{n}} = c_0 c_1 \pi_0 \sqrt{n} m, \tag{43}$$

which completes the proof.

□

7. Auxiliary Lemmas

Lemma 1. *For $c > 0$, $\hat{p} > 0$, $\Delta > 0$, $\tau > 0$, if*

$$\hat{p} - \sqrt{\frac{c\hat{p}}{n}} \leq \tau, \quad 2\sqrt{\frac{c\hat{p}}{n}} \geq \Delta, \tag{44}$$

then

$$n \leq \frac{4c(\tau + \frac{\Delta}{2})}{\Delta^2}. \tag{45}$$

Proof. (Proof of Lemma 1) Rearranging the first inequality in (44) and taking square of both sides to have

$$\hat{p}^2 - 2\tau\hat{p} + \tau^2 \leq \frac{c\hat{p}}{n}.$$

This further gives that

$$\hat{p} \leq \tau + \frac{c}{2n} + \sqrt{\frac{c}{n}\tau + \frac{c^2}{4n^2}}.$$

Combining the above with the second inequality in (44) to have

$$\frac{\Delta^2}{4c} n \leq \hat{p} \leq \tau + \frac{c}{2n} + \sqrt{\frac{c}{n}\tau + \frac{c^2}{4n^2}},$$

which can be rearranged as

$$\frac{\Delta^2}{4c} n - \tau - \frac{c}{2n} \leq \sqrt{\frac{c}{n}\tau + \frac{c^2}{4n^2}}.$$

Taking square of both sides and cancel the repeated terms to have

$$\left(\frac{\Delta^2}{4c} n\right)^2 - \frac{\Delta^2 \tau}{2c} n + \tau^2 - \frac{\Delta^2}{4} \leq 0,$$

which is equivalent to

$$\left(\frac{\Delta^2}{4c} n - \tau\right)^2 \leq \frac{\Delta^2}{4}.$$

Taking square root of both sides and we completed the proof. □

Lemma 2. *Given the fMC p-values $\{P_i^{fMC}\} = \{p_i\}$ with BH threshold τ^*, $\exists \delta_0 \in (0, 0.5)$, $c_0 > 0$, $c_1 > 0$, s.t. $\forall \delta < \delta_0$, a δ-correct algorithm satisfies*

$$\mathbb{E}\left[N \middle| \{P_i^{fMC}\} = \{p_i\}\right] \geq c_0 n \sum_{i=1}^m \mathbb{I}\{\tau^* < p_i \leq \tau^* + \frac{c_1}{\sqrt{n}}\}.$$

Proof. (Proof of Lemma 2) Consider any δ-correct algorithm and let us denote the true (unknown) fMC p-values by $\{q_i\}$. For any null hypothesis l with fMC p-value $\tau^* < p_l \le \tau^* + \frac{c_1}{\sqrt{n}}$, consider the following settings:

$$H_0 : q_i = p_i, \quad \text{for } i \in [m], \tag{46}$$

$$H_l : q_l = \tau^*, \quad q_i = p_i, \text{ for } i \neq l. \tag{47}$$

The δ-correct algorithm should accept the lth null hypothesis under H_0 and reject it under H_l, both with probability at least $1-\delta$. For $x \in \{0, l\}$, we use $\mathbb{E}_x$ and $\mathbb{P}_x$ to denote the expectation and probability, respectively, conditional on the fMC p-values $\{P_i^{\text{fMC}}\} = \{q_i\}$, under the algorithm being considered and under setting H_x. Let N_l be the total number of MC samples computed for null hypothesis l. In order to show Lemma 2, it suffices to show that $\mathbb{E}_0[N_l] \ge c_0 n$. We prove by contradiction that if $\mathbb{E}_0[N_l] < c_0 n$ and if the algorithm is correct under H_0 with probability at least 0.5, the probability that it makes a mistake under H_l is bounded away from 0.

Notations. Let $S_{l,t}$ to be the number of ones when t MC samples are collected for the lth null hypothesis. We also let S_l be the number of ones when all N_l MC samples are collected. Let $k_0 = (n+1)p_l - 1$ and $k_l = (n+1)\tau^* - 1$. Given N_l, S_l follows hypergeometric distribution with parameters (N_l, k_0, n) and (N_l, k_l, n) under H_0 and H_l, respectively. Let $\Delta_k = k_0 - k_l$. We note that

$$\Delta_k = (n+1)(p_l - \tau^*) \in (0, \frac{c_1(n+1)}{\sqrt{n}}]. \tag{48}$$

Define key events. Let $c_0 = 1/8$ and define the event

$$\mathcal{A}_l = \{N_l \le 0.5n\}. \tag{49}$$

Then by Markov's inequality, $\mathbb{P}_0(\mathcal{A}_l) \ge \frac{3}{4}$.

Let $\mathcal{B}_l$ be the event that the lth null hypothesis is accepted. Then $\mathbb{P}_0(\mathcal{B}_l) \ge 1 - \delta > 1/2$.

Let $\mathcal{C}_l$ be the event defined by

$$\mathcal{C}_l = \left\{ \max_{1 \le t \le 0.5n} |S_{l,t} - tk_0/n| < 2\sqrt{n} \right\}. \tag{50}$$

By Lemma 3 $\mathbb{P}_0(\mathcal{C}_l) \ge 7/8$.

Finally, define the event $\mathcal{S}_l$ by $\mathcal{S}_l = \mathcal{A}_l \cap \mathcal{B}_l \cap \mathcal{C}_l$. Then $\mathbb{P}_0(\mathcal{S}_l) > 1/8$.

Lower bound the likelihood ratio. We let W be the history of the process (the sequence of null hypotheses chosen to sample at each round, and the sequence of observed MC samples) until the algorithm terminates. We define the likelihood function L_l by letting

$$L_l(w) = \mathbb{P}_l(W = w), \tag{51}$$

for every possible history w. Note that this function can be used to define a random variable $L_l(W)$.

Given the history up to round $t-1$, the null hypotheses to sample at round t has the same probability distribution under either setting H_0 and H_l; similarly, the MC sample at round t has the same probability setting, under either hypothesis, except for the lth null hypothesis. For this reason, the likelihood ratio

$$\begin{aligned} \frac{L_l(W)}{L_0(W)} &= \frac{\binom{k_l}{S_l}\binom{n-k_l}{N_l-S_l}}{\binom{k_0}{S_l}\binom{n-k_0}{N_l-S_l}} \\ &= \prod_{r=0}^{S_l-1} \frac{k_l - r}{k_0 - r} \prod_{r=0}^{N_l-S_l-1} \frac{n-k_l-r}{n-k_0-r} \\ &= \prod_{r=0}^{S_l-1} \left(1 - \frac{\Delta_k}{k_0-r}\right) \prod_{r=0}^{N_l-S_l-1} \left(1 + \frac{\Delta_k}{n-k_0-r}\right) \end{aligned} \tag{52}$$

Next we show that on the event $\mathcal{S}_l$, the likelihood ratio is bounded away from 0.

If $S_l \le 100\sqrt{n}$, then the likelihood ratio

$$\frac{L_l(W)}{L_0(W)} \ge \left(1 - \frac{\Delta_k}{k_0 - S_l}\right)^{S_l} \ge \left(1 - \frac{c_2}{\sqrt{n}}\right)^{100\sqrt{n}} > c_3, \tag{53}$$

for some constants $c_2 > 0$, $c_3 > 0$.

If $S_l > 100\sqrt{n}$, further write (52) as

$$\frac{L_l(W)}{L_0(W)} = \prod_{r=0}^{S_l-1} \left\{ \left[1 - \left(\frac{\Delta_k}{k_0-r}\right)^2\right] \left(1 + \frac{\Delta_k}{k_0-r}\right)^{-1} \right\} \prod_{r=0}^{N_l-S_l-1} \left(1 + \frac{\Delta_k}{n-k_0-r}\right). \tag{54}$$

Since $S_l > 100\sqrt{n}$, on $\mathcal{C}_l$, $\frac{N_l - S_l}{S_l} > 1$. Note that if $a \ge 1$, then the mapping $x \mapsto (1+x)^a$ is convex for $x > -1$. Thus, $(1+x)^a \ge 1 + ax$, which implies that for any $0 \le r \le k_0$,

$$\left(1 + \frac{\Delta_k}{\frac{N_l-S_l}{S_l}(k_0-r)}\right)^{\frac{N_l-S_l}{S_l}} \overset{\mathcal{C}_l}{\ge} \left(1 + \frac{\Delta_k}{k_0-r}\right). \tag{55}$$

Then, (54) can be further written as

$$\begin{aligned} \frac{L_l(W)}{L_0(W)} \overset{(55)}{\ge} &\prod_{r=0}^{S_l-1} \left[1 - \left(\frac{\Delta_k}{k_0-r}\right)^2\right] \\ &\prod_{r=0}^{S_l-1} \left(1 + \frac{\Delta_k}{\frac{N_l-S_l}{S_l}(k_0-r)}\right)^{-\frac{N_l-S_l}{S_l}} \\ &\prod_{r=0}^{N_l-S_l-1} \left(1 + \frac{\Delta_k}{n-k_0-r}\right). \end{aligned} \tag{56}$$

Note that the 2nd term is no less than

$$\prod_{r=0}^{N_l-S_l-1}\left(1+\frac{\Delta_k}{\frac{N_l-S_l}{S_l}k_0-r}\right)^{-1}. \tag{57}$$

Eq. (56) can be further written as

$$\frac{L_l(W)}{L_0(W)} \geq \prod_{r=0}^{S_l-1}\left[1-\left(\frac{\Delta_k}{k_0-r}\right)^2\right]$$
$$\prod_{r=0}^{N_l-S_l-1}\left[\left(1+\frac{\Delta_k}{\frac{N_l-S_l}{S_l}k_0-r}\right)^{-1}\left(1+\frac{\Delta_k}{n-k_0-r}\right)\right] \tag{58}$$

Next we show that both terms in (58) are bounded away from 0.

First term in (58)

$$\prod_{r=0}^{S_l-1}\left[1-\left(\frac{\Delta_k}{k_0-r}\right)^2\right] \geq \left[1-\left(\frac{\Delta_k}{k_0-S_l}\right)^2\right]^{S_l} \tag{59}$$

$$\overset{\mathcal{A}_l,\mathcal{C}_l}{\geq} \left(1-\frac{c_4}{n}\right)^n \geq c_5 > 0, \tag{60}$$

for some constants $c_4>0$, $c_5>0$.

Second term in (55)

$$\prod_{r=0}^{N_l-S_l-1}\left[\left(1+\frac{\Delta_k}{\frac{N_l-S_l}{S_l}k_0-r}\right)^{-1}\left(1+\frac{\Delta_k}{n-k_0-r}\right)\right]$$
$$=\prod_{r=0}^{N_l-S_l-1}\left(1+\frac{\frac{\Delta_k}{n-k_0-r}-\frac{\Delta_k}{\frac{N_l-S_l}{S_l}k_0-r}}{1+\frac{\Delta_k}{\frac{N_l-S_l}{S_l}k_0-r}}\right)$$
$$=\prod_{r=0}^{N_l-S_l-1}\left(1+\frac{\Delta_k\frac{N_l}{S_l}\left(k_0-\frac{S_l}{N_l}n\right)}{\left(1+\frac{\Delta_k}{\frac{N_l-S_l}{S_l}k_0-r}\right)(n-k_0-r)(\frac{N_l-S_l}{S_l}k_0-r)}\right)$$
$$\overset{\mathcal{A}_l,\mathcal{C}_l,S_l>100\sqrt{n}}{\geq}\left(1-\frac{c_6}{N_l\sqrt{n}}\right)^{N_l} \geq c_7, \tag{61}$$

for some constants $c_4>0$, $c_5>0$.

Hence $\exists c_8>0$, such that on $\mathcal{S}_l$ the likelihood ratio

$$\frac{L_l(W)}{L_0(W)} \geq c_8 > 0. \tag{62}$$

Therefore, the probability of making an error under H_l

$$\begin{aligned}&\mathbb{P}_l(\text{error}) \geq \mathbb{P}_l(\mathcal{S}_l) = \mathbb{E}_l[\mathbb{I}\{S_l\}]\\ &= \mathbb{E}_0\left[\mathbb{I}\{S_l\}\frac{L_l(W)}{L_0(W)}\right] \geq c_8\mathbb{P}_0(\mathcal{S}_l) \geq \frac{c_8}{8}.\end{aligned} \tag{63}$$

Hence, there does not exist a δ-correct algorithm for any $\delta \leq \frac{c_8}{8}$, completing the proof. □

Lemma 3. *Let $X_1,\cdots,X_n$ be random variables sampled without replacement from the set $\{x_1,\cdots,x_N\}$, where $n \leq N$ and $x_i \in \{0,1\}$. Let $\mu = \frac{1}{N}\sum_{i=1}^N x_i$ and for $k \in [N]$, let $S_k = \sum_{i=1}^k X_i$. Then for any $\theta > 0$,*

$$\mathbb{P}\left(\max_{1\leq k\leq n}|S_k-\mu k| \geq \sqrt{n\theta}\right) \leq \frac{1}{\theta}. \tag{64}$$

This is a direct consequence of Corollary 1.2 in the paper (Serfling, 1974).