A Bandit Approach to Classification and False-Discovery Control

Lalit Jain[1] and Kevin Jamieson[1]

[1]Paul G. Allen School of Computer Science & Engineering
University of Washington

February 21, 2019

Abstract

In many scientific settings there is a need for adaptive experimental design to guide the process of identifying regions of the search space that contain as many true positives as possible subject to a low rate of false discoveries (i.e. false alarms). Such regions of the search space could differ drastically from a predicted set that minimizes $0/1$ error and moreover accurate identification could require very different sampling strategies. Like active learning for binary classification, this experimental design can not be optimally chosen a priori, but rather the data must be taken sequentially and adaptively in a closed loop. However, unlike active classification, finding a set with high true positive rate and low false discovery rate (FDR) is not as well understood. In this paper we provide the first provably sample efficient adaptive algorithm for this problem. Along the way we highlight connections between classification, combinatorial bandits, and FDR control making contributions to each.

1 Introduction

As machine learning has become ubiquitous in the biological, chemical, and material sciences, it has become irresistible to use machine learning not only for making inferences about *previously* collected data, but to guide data collection process itself [9, 34, 36, 35, 29, 27]. However, from a statistically rigorous perspective, many of these strategies appear very greedy and the impact of this biased data collection process is unknown. For example, it is not difficult to imagine situations where a practitioner collects some data, trains a data-poor model, collects data based on that model and discovers very little, nonethewiser to whether the experimental procedure failed, or the phenomenon was unobservable given the measurement budget constraints. Motivated by these recent works, this paper attempts to formalize a popular scientific objective and propose actionable algorithmic takeaways.

Consider a recent high-throughput protein synthesis experiment [29] in which tens of thousands of short <60 length amino acid sequences were evaluated with the goal of identifying and characterizing a set of sequences that are very likely to fold among all those possible. The first round of experiments $S_1 \subset [n]$ were uniformly sampled from a pre-defined subset of short amino acid sequences. Each sequence was then synthesized to observe whether it was in the set of sequences that will fold, $\mathcal{H}_1$, or in $\mathcal{H}_0 = \mathcal{H}_1^c$. Using physics-based features and the observed binary outcomes, a linear logistic

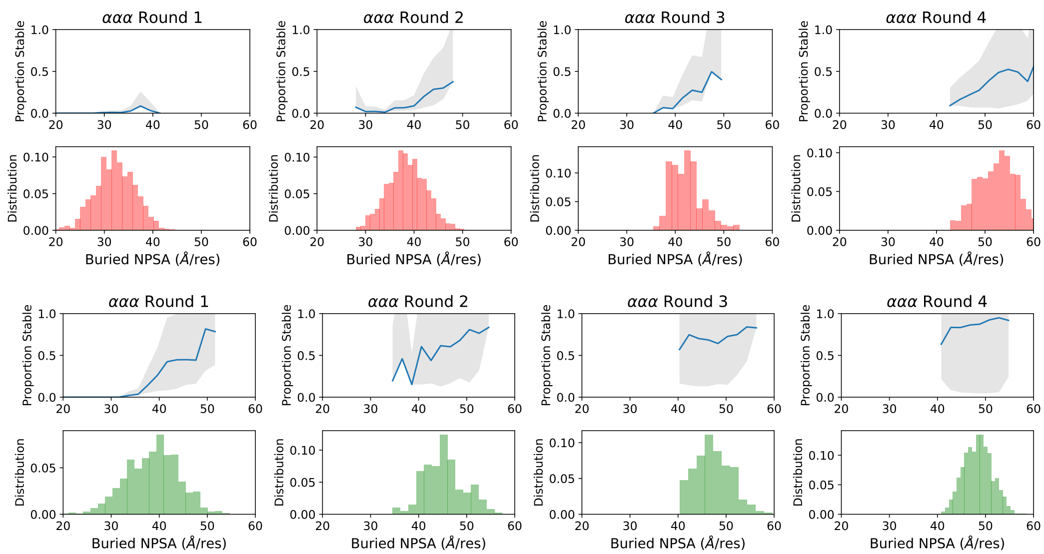

Figure 1: The distribution of a feature that was highly correlated with the fitted logistic model. One observes that the distribution of this feature for the sequences across the models moves to the right which could optimize just a portion of the the whole space.

regression classifier was trained and in the next round, a set of sequences $S_2 \subset [n]$, were chosen to maximize the probability of folding according to this empirically trained logistic model. This per-round procedure was repeated for two more rounds resulting in several tens of thousands total tested sequences. Though a logistic regression model was used to output a set of predictions $\pi \subset [n]$ of proteins likely to fold, it is not clear that classification was the correct objective–indeed, the proteins tested with secondary structure $\beta\alpha\beta\beta$ (see Figure 1) so infrequently folded (class-imbalance) and were poorly predicted by the features that even the most likely predictions are less than one-half. A near-optimal classifier within any reasonable function class would predict "not fold" for every sequence (i.e., the *classification* error $\frac{|\pi \cap \mathcal{H}_0| + |\pi^c \cap \mathcal{H}_1|}{n}$ is minimized by $\pi = \emptyset$ the empty set). Selecting S_{k+1} to maximize the likelihood of hits given past rounds' data is effectively using the logistic regression model to perform *optimization* (i.e., find the *most* probable sequence to fold, comparable to follow the leader). Such a method is likely to result in a high success rate of those evaluated, i.e., $\mathbb{E}[\frac{|\mathcal{H}_1 \cap S|}{|S|}]$, but provide little information towards learning a set π with large true positive rate (TPR) $\mathbb{E}[\frac{|\mathcal{H}_1 \cap \pi|}{|\mathcal{H}_1|}]$. Of course, one could maximize TPR by just taking $\pi = [n]$, the entire space, so we restrict ourselves to only those sets with a bounded false discovery rate (FDR) $\mathbb{E}[\frac{|\mathcal{H}_0 \cap \pi|}{|\pi|}]$ so that of the items we recommend, a fixed proportion of them are guaranteed to be true discoveries.

In the above example of $\beta\alpha\beta\beta$ where the best predictions for discoveries never exceed $1/2$, minimizing classification error would yield an empty set. However, if one were to identify a set π that maximized $\mathbb{E}[\frac{|\mathcal{H}_1 \cap \pi|}{|\mathcal{H}_1|}]$ subject to $\mathbb{E}[\frac{|\mathcal{H}_0 \cap \pi|}{|\pi|}] \leq 9/10$ then π would be non-empty with the guarantee that at least one in every 10 suggestions was a true-positive; not ideal, but making the best of a bad situation. In a different extreme, consider $\alpha\alpha\alpha$ of Figure 1 where the predictions are quite high. In this case, the set that minimizes classification error would contain nearly half the proteins on round 1 (i.e., where the probability of success passes through 0.5) and of those items recommended in the set, a little over 50% would be true positives. But with so many positive examples present, a practitioner may desire a set that is most *pure* which could be achieved by identifying a set π that maximizes $\mathbb{E}[\frac{|\mathcal{H}_1 \cap \pi|}{|\mathcal{H}_1|}]$ subject to $\mathbb{E}[\frac{|\mathcal{H}_0 \cap \pi|}{|\pi|}] \leq 1/10$ so that at least 90% of the recommendations are true positives. At a high-level, this discussion is just considering different points on the receiver operating characteristic (ROC) curve if we had access to all the data. But the reality is that *we don't have all the data*, and any dataset collected in an arbitrary way may be ill-equipped to optimize a particular point on the ROC curve. This paper addresses the following challenge: given an $\alpha \in (0, 1)$ such that it is desired to have false discovery rate bounded by α, how should an algorithm *collect data* in order to maximize the true positive rate.

1.1 Combinatorial Bandits with FDR Control

Let $\{(x_i, y_i)\}_{i=1}^n \subset \mathbb{R}^d \times \{0, 1\}$ be a fixed set of feature-label pairs where the x_i vectors are known a priori, but the y_i variables are unknown. At each round t the player chooses an $I_t \in [n]$ and immediately observes y_{I_t}. For a fixed $\alpha \in (0, 1)$, after T samples the player's goal is to output a π informed by the observed features and labels that maximizes the true positive rate (TPR) subject to a constraint on the false discovery rate (FDR):

$$\text{Maximize} \quad TPR(\pi) = \mathbb{E}\left[\frac{|\pi \cap \mathcal{H}_1|}{|\mathcal{H}_1|}\right] \quad \text{subject to} \quad FDR(\pi) = \mathbb{E}\left[\frac{|\pi \cap \mathcal{H}_0|}{|\pi|}\right] \leq \alpha$$

where the expectation is with respect to the possible external randomness used by the player. Instead of considering all possible sets $\pi \subset [n]$, we will restrict ourselves to a finite class $\Pi \subset 2^{[n]}$ which is induced by the feature vectors and a function class. With an abuse of notation, we will interpret $\pi \in \Pi$ as both a set $\pi \subset [n]$ and a function $\pi(x_i) \in \{0, 1\}$ for any x_i. With this dual interpretation, one can imagine very structured classes, such as Π being the space of linear separators over $\{x_i\}$, or more combinatorial sets, like Π being all subsets of $[n]$ of size k. We call this the *Combinatorial Bandits with FDR control* problem. In a different community, this problem may have been called *active learning for information retrieval*. In that terminology, the player sequentially requests labels and seeks a binary classifier that obtains the maximum recall with a precision of at least $1 - \alpha$.

1.2 Contributions

Our goal is to propose the first provably sample-efficient adaptive sampling algorithm for maximizing TPR subject to an FDR constraint. This problem has deep connections to active binary classification (e.g., active learning) and pure-exploration for combinatorial bandits. Along the way to deriving our FDR results, we improve upon state of the art sample complexity results for both of these related areas. In particular, for classification we provide novel sample complexity bounds that do not depend on the disagreement-coefficient for both the stochastic noise and persistent noise setting (i.e., sampling with or without replacement). Our bounds are more granular than those in the active learning literature in that they characterize the number of times an iid label is requested from a given example, if at all. These results follow directly from our improvements to the state of the art in combinatorial bandits including innovations in making explicit connections to VC dimension, extending methods to be near-optimal for non-matroid classes where one need not sample every arm at least once. In the next subsections we briefly review binary classification and combinatorial bandits for posterity. However, the main contribution of this work is the development and analysis of an adaptive sampling algorithm that minimizes the number of samples to identify the set that maximizes the true positive rate subject to a bound on the false discovery rate. To the best of our knowledge, this is the first work to demonstrate a sample complexity for this problem that is provably better than non-adaptive sampling.

1.3 Binary Classification

Here we describe what is known as the pool-based setting for active learning for binary classification with stochastic labels. Throughout the following we assume access to a finite feature space of items $[n] = \{1, \cdots, n\}$ with an associated label space $\{0, 1\}$. As in the previous setting the items can be fixed vectors $\{x_i\}_{i=1}^n \in \mathbb{R}^d$ but we do not restrict to this case. Associated to each $i \in [n]$ there is an associated Bernoulli distribution $\text{Ber}(\eta_i)$ with $\eta_i \in [0, 1]$. Let $\{Y_{i,j}\}_{j=1}^\infty$ be an infinite sequence of i.i.d. random variables where $Y_{i,j} \sim \text{Ber}(\eta_i)$ is the label obtained after the j draw from i's distribution. Such a model is reasonable in a crowd-sourcing environment where we can request multiple labels for a single example from independent (but not necessarily identically distributed) workers. We will refer to this level of generality as the **stochastic noise** setting. The case when $\eta_i \in \{0, 1\}$, i.e. each point $i \in [n]$ has a deterministic label $Y_{i,j} = \eta_i$ for all $j \geq 1$ will be referred to as the **persistent noise** setting. In this setting we can define $\mathcal{H}_1 = \{i : \eta_i = 1\}, \mathcal{H}_0 = [n] \setminus \mathcal{H}_1$. This is a natural setting if the experimental noise is negligible so that performing the same measurement multiple times gives the same result.

As described in the introduction a classifier is a decision rule that assigns each item $i \in [n]$ a fixed label. In general, we can drop the explicit dependence on the feature vectors x_i and consider only sets so that for every classifier $f : \mathbb{R}^d \to \{0, 1\}$ there exists a set $\pi = \{i : i \in [n], f(x_i) = 1\}$ with the convention $\pi(i) = 1$ iff $i \in \pi$. The *classification error*, or *risk* of a classifier is given by the expected number of incorrect labels, i.e.

$$R(\pi) = \mathbb{P}_{i \sim \text{Unif}([n]), Y_i \sim \text{Ber}(\eta_i)} (\pi(i) \neq Y_i) = \frac{\sum_{i \notin \pi} \eta_i + \sum_{i \in \pi}(1 - \eta_i)}{n}$$

for any $\pi \in \Pi$. In the case of persistent noise the above reduces to

$$R(\pi) = \frac{|\pi \cap \mathcal{H}_0| + |\pi^c \cap \mathcal{H}_1|}{n} = \frac{|\mathcal{H}_1 \Delta \pi|}{n}$$

where $A \Delta B = (A \cup B) - (A \cap B)$ for any sets A, B. Note that the Bayes classifier, i.e. $\text{argmin}_\pi R(\pi)$, is given by $\{i \in [n] : \eta_i \geq 1/2\}$. Often we have access to a hypothesis class that may or may not contain the Bayes classifier and we are interested in the lowest loss classifier in our class. In this work we consider the so-called *agnostic* setting in the sense that we do not assume the Bayes classifier is contained in our set Π of interest.

Problem 1:(Classification) Given a hypothesis class $\Pi \subseteq 2^{[n]}$ identify $\pi^* := \text{argmin}_{\pi \in \Pi} R(\pi)$ by requesting as few labels as possible.

In many situations we are not as interested in finding the lowest risk classifier, but instead returning $\pi \subset X$ that contains many *discoveries* $\pi \cap \mathcal{H}_1$ without too many false alarms $\pi \cap \mathcal{H}_0$. Define $\eta_\pi := \sum_{i \in \pi} \eta_x$. The *false discovery rate (FDR)* and *true positive rate (TPR)* of a set π in the stochastic noise setting are given by

$$FDR(\pi) := 1 - \frac{\eta_\pi}{|\pi|} \quad \text{and} \quad TPR(\pi) := \frac{\eta_\pi}{\eta_{[n]}}$$

In the case of persistent noise, $FDR(\pi) = \frac{|\mathcal{H}_0 \cap \pi|}{|\pi|} = 1 - \frac{|\mathcal{H}_1 \cap \pi|}{|\pi|}$ and $TPR(\pi) = \frac{|\mathcal{H}_1 \cap \pi|}{|\mathcal{H}_1|}$. A convenient quantity that we can use to reparametrize these quantities is the *true positives:* $TP(\pi) := \sum_{i \in \pi} \eta_\pi$. Throughout the following we let $\Pi_\alpha = \{\pi \in \Pi : FDR(\pi) \leq \alpha\}$.

Problem 2:(Combinatorial FDR Control) Given an $\alpha \in (0, 1)$ and hypothesis class $\Pi \subseteq 2^{[n]}$ identify $\pi^*_\alpha = \underset{\pi \in \Pi, FDR(\pi) \leq \alpha}{\text{argmax}} \; TPR(\pi)$ by requesting as few labels as possible.

1.4 Classification and FDR as Combinatorial Bandits

A closely related problem to classification is the *pure-exploration combinatorial bandit* problem. As above we have access to a set of arms $[n]$, and associated to each arm is an unknown distribution ν_i. We let $\{R_{i,j}\}_{j=1}^{\infty}$ be a sequence of random variables where $R_{i,j} \sim \nu_i$ is the jth draw from ν_i satisfying $\mathbb{E}[R_{i,j}] = \mu_i \in [-1, 1]$. In the persistent noise setting we assume that ν_i is a point mass at $\mu_i \in [-1, 1]$. Given a collection of sets $\Pi \subseteq 2^{[n]}$, for each $\pi \in \Pi$ we define $\mu_\pi := \sum_{i \in \pi} \mu_i$ the sum of means in π.

Problem 3: (Combinatorial Bandits) Given a hypothesis class $\Pi \subseteq 2^{[n]}$ identify $\pi^* = \underset{\pi \in \Pi}{\text{argmax}} \mu_\pi$ by requesting as few labels as possible.

The combinatorial bandit extends many problems considered in the multi-armed bandit literature. For example if $\Pi = \{\{i\} : i \in [n]\}$ then this is equivalent to the best-arm identification problem. In this work, we are most interested in the connection between combinatorial bandits and classification.

Returning to the classification setting for a moment, we make the connection between classification and combinatorial bandits explicit. For each i define $\mu_i := 2\eta_i - 1 \in [-1, 1]$ so $\eta_i = \frac{1+\mu_i}{2}$. By a simple manipulation of the definition of $R(\pi)$ above we have

$$R(\pi) = \frac{1}{n} \sum_{i=1}^{n} \eta_i + \frac{1}{n} \sum_{i \in \pi} (2\eta_i - 1) = \frac{1}{n} \sum_{i=1}^{n} \eta_i + \frac{1}{n} \sum_{i \in \pi} \mu_i$$

so that $\underset{\pi \in \Pi}{\text{argmin}} \quad R(\pi) = \underset{\pi \in \Pi}{\text{argmax}} \sum_{i \in \pi} \mu_i$. Hence, if for some $i \in [n]$ we map the jth draw of its label $Y_{i,j} \mapsto 2Y_{i,j} - 1$ then the $\mathbb{E}[2Y_{i,j} - 1] = \mu_i$ and returning an optimal classifier in the set is equivalent to returning a subset π with the largest μ_π.

The connection between FDR control and combinatorial bandits is more direct: we are seeking to find $\pi \in \Pi$ with maximum η_π subject to FDR-constraints. This already highlights a key difference between classification and FDR-control. In one we choose to sample to maximize η_π subject to FDR constraints where each $\eta_i \in [0, 1]$, whereas in classification we are trying to maximize μ_π where each $\mu_i \in [-1, 1]$. A major consequence of this difference is that $\eta_\pi \leq \eta_{\pi'}$ whenever $\pi \subseteq \pi'$, but such a condition does not hold for $\mu_\pi, \mu_{\pi'}$.

2 Related Work

The pure-exploration combinatorial bandit game has been studied for the case of all subsets of $[n]$ of size k known as the Top-K problem [19, 25, 26, 24, 33, 14] the bases of a rank-k matroid, for which Top-K is a particular instance [15, 20, 12], and in the general case [10, 13]. The combinatorial bandit component of our work is closest to [10] that was designed for the most general setting and demonstrates a sample complexity that performs no worse then these previous works up to a factor of $\log(n)$. Exploring precisely what $\log$ factors are necessary has been an active area. Indeed, [13] demonstrates a family of instances in which they show in the worst-case, the sample complexity must scale with $\log(|\Pi|)$. However, there are many classes like best-arm identification and matroids where sample complexity does *not* scale with $\log(|\Pi|)$ (see references above). Our own work provides some insight into what $\log$ factors are necessary by presenting our results in terms of VC dimension, and when a $\log(n)$ could potentially be avoided or not by appealing to Sauer's lemma. The algorithm of [10] uses a disagreement-based algorithm which is in the spirit of Successive Elimination for bandits [19], or the A^2 for binary classification [2].

Active learning for binary classification is a mature field (see surveys [32, 22] and references therein). The major theoretical results of the field can coarsely be partitioned into the streaming setting [2, 5, 17, 23] and the pool-based setting [16, 21, 28], noting that algorithms for the former can be used for the latter, and our algorithm, inspired by [2] is such an example. These results rely on different complexity measures like the splitting index, the teaching dimension, and arguably the most popular the disagreement coefficient. While there have been remarkable efforts to make some of these methods more computationally efficient, we believe even given infinite computation, many of these methods are fundamentally inefficient from a sample complexity perspective. This stems from the fact that when applied to common bandit problems like Top-K, these algorithms have sample complexities that are off by at least $\log(n)$ factors from the best algorithms cited above. Consequently, in our work we focus on sample complexity alone, and leave matters of computational efficiency for future work.

Though there is extensive work on the sample complexity of computing measures related to precision and recall such as AUC, and F-scores [31, 8, 1], there has been just a few works that consider the adaptive problem of maximizing precision with recall constraints [30, 4]. The most closely related work is that of [4]. Similar to our proposed algorithm, they sample in the union of all active sets and maintain statistics on the empirical FDR of each set, along the way removing sets that are not FDR-controlled or have lower TPR than an FDR-controlled set. However, they fail to sample in the symmetric difference, effectively missing an important link between FDR-control and combinatorial bandits that leads to lower sample complexities. They also only consider the case of persistent noise. Finally, their sample complexity results are no better than those achieved by the passive algorithm that samples each item uniformly. In [30], the problem of adaptively estimating the whole ROC curve for a threshold class is considered under a monotonicity assumption on the true positives–an assumption our algorithm is agnostic to.

3 Combinatorial Bandits and Active Classification

Input: δ, Confidence bound $C(\pi', \pi, t, \delta)$.
Let $\mathcal{A}_1 = \Pi$, $k = 1$, $\mathcal{A}_k$ will be the active sets in round k
for $t = 1, 2, \cdots$
  if $t == 2^k$**:**
    Set $\delta_k = .5\delta/k^2$. Let $t_k = 2^k$. For each π, π' let
    $$\widehat{\mu}_{\pi'\setminus\pi} - \widehat{\mu}_{\pi\setminus\pi'} = \frac{n}{t}\Big(\sum_{s=1}^{t} R_{I_s,s}\mathbf{1}\{I_s \in \pi' \setminus \pi\} - \sum_{s=1}^{t} R_{I_s,s}\mathbf{1}\{I_s \in \pi \setminus \pi'\}\Big)$$
    $$\mathcal{A}_{k+1} = \mathcal{A}_k - \big\{\pi \in \mathcal{A}_k : \exists \pi' \in \mathcal{A}_k \text{ with } \widehat{\mu}_{\pi'\setminus\pi} - \widehat{\mu}_{\pi\setminus\pi'} > C(\pi', \pi, t_k, \delta_k)\big\}.$$
    $$T_k = \Big(\bigcup_{\pi \in \mathcal{A}_k} \pi\Big) - \Big(\bigcap_{\pi \in \mathcal{A}_k} \pi\Big).$$
    $k \leftarrow k + 1$
  endif
  Stochastic Noise:
    If $|\mathcal{A}_k| = 1$, **Break**. Otherwise, draw I_t uniformly at random from $[n]$ and if $I_t \in T_k$ receive an associated reward $R_{I_t,t} \overset{iid}{\sim} \nu_i$.
  Persistent Noise:
    If $|\mathcal{A}_k| = 1$ or $t > n$, **Break**. Otherwise, draw I_t uniformly at random from $[n] \setminus \{I_s : 1 \leq s < t\}$ and if $I_t \in T_k$ receive associated reward $R_{I_t,t} = \mu_{I_t}$.
Output: $\pi' \in \mathcal{A}_k$ such that $\widehat{\mu}_{\pi'\setminus\pi} - \widehat{\mu}_{\pi\setminus\pi'} \geq 0$ for all $\pi \in \mathcal{A}_k \setminus \pi'$

Figure 2: Action Elimination for Pure-Exploration Combinatorial Bandits

We begin by giving a general algorithm for the combinatorial bandit problem in Figure 2. Recall that in the case of classification, the rewards are $R_{i,j} = 2Y_{i,j} - 1$ where $Y_{i,j} \sim \text{Ber}(\eta_i)$. The algorithm maintains an active set of sets $\mathcal{A}_k \subseteq \Pi$ and an active set of arms $T_k \subseteq [n]$ which is just the symmetric difference of all sets in $\mathcal{A}_k$. To see why we only sample in T_k, note that if $i \in \cap_{\pi \in \mathcal{A}_k} \pi$ then π and π' agree on the label of item i, and any contribution of arm i is cancelled in each difference $\widehat{\mu}_\pi - \widehat{\mu}_{\pi'} = \widehat{\mu}_{\pi \setminus \pi'} - \widehat{\mu}_{\pi' \setminus \pi}$ for all $\pi, \pi' \in \mathcal{A}_k$ so we should not pay to sample it. In each round policies π with lower total empirical means that fall outside of the confidence interval of higher total means are removed. Algorithm 2 is similar to previous action elimination algorithms for combinatorial bandits in the literature, e.g. algorithm 4 in [10]. However, unlike previous algorithms, we do not insist on pulling each arm once, an unrealistic requirement for classification settings. In the specific case of one dimensional thresholds highlighted in the next section, this will result in a sublinear sample complexity. The confidence interval taken as input by the algorithm will be defined shortly. For any $\mathcal{A} \subseteq 2^{[n]}$ define $V(\mathcal{A})$ as the VC-dimension of a collection of sets $\mathcal{A}$.

Definition 1. *Given a family of sets,* $\Pi \subseteq 2^{[n]}$*, define* $B_1(k) := \{\pi \in \Pi : |\pi| = k\}$, $B_2(k, \pi') := \{\pi \in \Pi : |\pi \Delta \pi'| = k\}$*. Also define the following complexity measures:*

$$V_\pi := \min\{V(B_1(|\pi|)), |\pi|\}$$
$$V_{\pi,\pi'} := \min\{\max\{V(B_2(|\pi\Delta\pi'|, \pi), V(B_2(|\pi\Delta\pi'|, \pi'))\}, |\pi\Delta\pi'|\}.$$

Note that in general $V_\pi, V_{\pi,\pi'} \leq V(\Pi)$.

Lemma 1. *Assume that for each arm* $i \leq n$ *there is an associated distribution* ν_i *with support* $[-1, 1]$*, mean* μ_i *and variance* $\sigma_i^2 \leq 1$*. Assume access to the observations* $(I_1, y_{I_1}) \cdots, (I_t, y_{I_t})$ *in two different but related settings.*

1. ***Stochastic Noise*** $I_k \sim$ *Unif*$([n])$ *and* $y_{I_k} \sim \nu_{I_k}$.
2. ***Persistent Noise*** $I_k \in [n]$ *are drawn without replacement,* $y_{I_k} = \mu_{I_k}$, $t \leq n$

For any $A \subseteq [n]$ *define* $\widehat{\mu}_A = \frac{n}{T}\sum_{k=1}^{T} y_k \mathbf{1}\{I_k \in A\}$*. Then in both settings*

1. *With probability greater than* $1 - \delta$ *for all* $\pi \in \Pi$

$$|\widehat{\mu}_\pi - \mu_\pi| \leq C_1(\pi, t, \delta) := \sqrt{\frac{4|\pi| n V_\pi \log\left(\frac{n}{\delta}\right)}{t}} + \frac{4nV_\pi \log\left(\frac{n}{\delta}\right)}{3t}$$

2. *Fix* $\pi' \in \Pi$*. With probability greater than* $1 - \delta$ *for all* $t > 0$ *and* $\pi \in \Pi$

$$\left|\widehat{\mu}_{\pi'\setminus\pi} - \widehat{\mu}_{\pi\setminus\pi'} - (\mu_{\pi'\setminus\pi} - \mu_{\pi\setminus\pi'})\right| \leq C_2(\pi, \pi', t, \delta) := \sqrt{\frac{8|\pi\Delta\pi'| n V_{\pi,\pi'} \log\left(\frac{n}{\delta}\right)}{t}} + \frac{4nV_{\pi,\pi'} \log\left(\frac{n}{\delta}\right)}{3t}$$

In the appendix we have stronger confidence bounds that hold for sampling without replacement. We have the following result on the sample complexity of algorithm of Figure 2.

Theorem 1 (Stochastic Noise)**.** *For each* $i \in [n]$ *let* $\mu_i \in [-1, 1]$ *be fixed but unknown and assume* $\{R_{i,j}\}_{j=1}^{\infty}$ *is an i.i.d sequence of random variables such that* $\mathbb{E}[R_{i,j}] = \mu_i$ *and* $R_{i,j} \in [-1, 1]$*. Define* $\pi^* = \operatorname{argmax}_{\pi \in \Pi} \mu_\pi$, $\widetilde{\Delta}_\pi = |\mu_\pi - \mu_{\pi^*}|/|\pi\Delta\pi^*|$*, and*

$$\tau_\pi = \frac{V_{\pi,\pi^*}}{|\pi^*\Delta\pi|} \frac{1}{\widetilde{\Delta}_\pi^2} \log\left(n \log(\widetilde{\Delta}_\pi^{-2})/\delta\right).$$

Using $C_2(\pi', \pi, t, \delta)$ *from Lemma 1, algorithm of Figure2 identifies* π_* *using no more than* $c \sum_{i=1}^{n} \max_{\pi \in \Pi : i \in \pi\Delta\pi^*} \tau_\pi$ *samples with probability at least* $1 - \delta$ *where* c *is an absolute constant.*

Theorem 2 (Persistent Noise)**.** *For each $i \in [n]$ let $\mu_i \in [-1, 1]$ be fixed but unknown. Let π^* and $\widetilde{\Delta}_\pi$ be as in the previous theorem. Then using $C_2(\pi', \pi, t, \delta)$ from* (1)*, algorithm of Figure2 identifies π_* using no more than $c \sum_{i=1}^{n} \min\left\{1, \max_{\pi \in \Pi : i \in \pi \Delta \pi^*} \tau_\pi\right\}$ samples with probability at least $1 - \delta$ where c is an absolute constant.*

Note one always has $\frac{1}{|\pi^* \Delta \pi|} \leq \frac{V_{\pi,\pi^*}}{|\pi^* \Delta \pi|} \leq 1$ and both bounds are achievable by different classes Π.

3.1 Comparison with previous Active Classification results.

One of the foundational works on active learning is the DHM algorithm of [17] and the A^2 algorithm that preceded it [2]. In their setting a set of points, $x_1, x_2, \cdots$ are streamed to a learner who chooses whether to label a point or not. Similar in spirit to our algorithm, DHM maintains a set of points that it is certain of the label of under π^* and then requests the labels of any points that it is uncertain about. A key quantity arising in the sample complexity of DHM (and many previous works on active classification) has been that of the disagreement coefficient of the set π^*: $\theta = \theta(\epsilon, \pi^*) := \sup_{r \geq n(\epsilon+\nu)} \left\{ \frac{|x : x \in \pi \Delta \pi^*, \pi \in \Pi \text{ and } |\pi \Delta \pi^*| \leq r|}{r} \right\}$ where $\nu = \mathbb{P}(\pi^*(x) \neq y)$ and ϵ is a bound on the excess error of the set $\widehat{\pi}$ returned by an active learning algorithm. After being streamed m points, DHM returns a classifier with error at most $O(\nu + d\log(m/\delta)/m + \sqrt{d\nu \log(m/\delta)/m})$ after labeling $O\left(\theta\left(\nu m + d\log^2(m) + \log\left(\frac{\log(m)}{\delta}\right)\right)\right)$ samples (provided $\epsilon \leq \nu$–the realistic setting in the non-realizable noisy case). Ignoring log factors, this roughly says that a classifier with error at most $\nu + \epsilon$ is returned after $\theta(\nu m + d) \approx \theta d\nu \max\{\epsilon^{-1}, \nu\epsilon^{-2}\}$ samples.

In general the analysis of the DHM algorithm can not characterize the contribution of each arm to the overall sample complexity. Consider the case of best-arm identification so that $d = 1$, $\pi_i = \{i\}$ and $\pi^* = \{i^*\}$ where $i^* = \text{argmax}_{i \leq n} \mu_i$. If we take $\mu_i \in [-1/2, 1/2]$ for all i then $\frac{1}{4} - \frac{1}{2n} \leq \nu \leq \frac{3}{4} + \frac{1}{2n}$ and for best-arm we necessarily have $\epsilon = \min_{j \neq i^*} \frac{1}{n}(\mu_{i^*} - \mu_j)$. One can show for this problem $\theta = \frac{1}{\nu+\epsilon}$ and so the bound of Theorem 1 of [17] scales like $\theta d\nu \max\{\epsilon^{-1}, \nu\epsilon^{-2}\} = \frac{\nu}{\nu+\epsilon} \max\{\epsilon^{-1}, \nu\epsilon^{-2}\} \approx \epsilon^{-2} = n^2 \max_{i \neq i_*} \Delta_i^{-2}$ for $\Delta_i = \mu_{i^*} - \mu_i$, which is substantially worse than our bound for this problem which scales like $\sum_{i \neq i_*} \Delta_i^{-2}$, describing the contribution from each individual arm. Similar arguments can be made for matroid-like classes.

3.2 Comparison with previous Combinatorial Bandits results.

The general combinatorial bandit problem is considered in [10]. There they present an algorithm with sample complexity,

$$C \sum_{i=1}^{n} \max_{\pi : i \in \pi \Delta \pi^*} \frac{1}{|\pi \Delta \pi^*|} \frac{1}{\widetilde{\Delta}_\pi^2} \log\left(\max(|B(|\pi\Delta\pi^*|, \pi)|, |B(|\pi\Delta\pi^*|, \pi^*)|) \frac{n}{\delta}\right)$$

where $B(r, \pi) = \{\pi \in \Pi : |\pi \Delta \pi^*| = r\}$. This complexity parameter is difficult to interpret directly so we compare it to one more familiar in statistical learning - the VC dimension. To see how this sample complexity relates to ours, note that $\log_2 |B(k, \pi^*)| \leq \log_2 \binom{n}{k} \lesssim k \log_2(n)$. Thus by the Sauer-Shelah lemma,

$$V(B(r, \pi^*)) \lesssim \log_2(|B(r, \pi^*)|) \lesssim \min\{V(B(r, \pi^*)), r\} \log_2(n) \tag{1}$$

where $\lesssim$ hides a constant. The proof of Lemma 1 effectively combines these two facts along with a union bound over all sets in $B(r, \pi^*)$.

Perhaps the more interesting question is whether we can acheive the bound on the left of the equation, i.e. whether the $\log(n)$ can be dropped. The answer lies in empirical process theory. In general given a class of sets $\mathcal{A}$, with $\sup_{\pi\in\mathcal{A}} |\pi| \leq s$, with the uniform measure, we have that the empirical process $\sup_{\pi\in\Pi} \frac{1}{\sqrt{T}} |\sum_{i=1}^{T} \mathbf{1}\{i \in \pi\} - |\pi|| \leq \sqrt{sV\log(n/s)}$ (see Theorem 13.7 in [6]). The $\log(n/s)$ term can be dropped in highly structured cases, but is needed in general. We give such an example where the $\log(n)$ is not necessary in Appendix C for the case of one-dimensional thresholds. Indeed, our general purpose algorithm that is agnostic to assumptions on the noise recovers the optimal sample complexity obtained by special purpose algorithms that explicitly assume the Bayes classifier is in Π [11].

Input: Confidence bounds $C_1(\pi, t, \delta), C_2(\pi, \pi', t, \delta)$
$\mathcal{A}_k \subset \Pi$ will be the set of active sets in round k. $\mathcal{C}_k \subset \Pi$ is the set of FDR-controlled policies in round k.
$\mathcal{A}_1 = \Pi,\ \mathcal{C}_1 = \emptyset,\ S_1 = \cup_{\pi\in\Pi}\pi,\ T_1 = \cup_{\pi\in\Pi}\pi - \cap_{\pi\in\Pi}\pi, k = 1.$
for $t = 1, 2, \cdots$
  if $t = 2^k$**:**
    Let $\delta_k = .25\delta/k^2, t_k = 2^k$
    For each set $\pi \in \mathcal{A}_k$, and each pair $\pi', \pi \in \mathcal{A}_k$ update the estimates

$$\widehat{FDR}(\pi) = 1 - \frac{n}{|\pi| t}\sum_{s=1}^{t} Y_{I_s,s}\mathbf{1}\{I_s \in \pi\}$$

$$\widehat{TP}(\pi'\backslash\pi) - \widehat{TP}(\pi\backslash\pi') = \frac{n}{t}\left(\sum_{s=1}^{t} Y'_{J_s,s}\mathbf{1}\{J_s \in \pi'\backslash\pi\} - \sum_{s=1}^{t} Y'_{J_s,s}\mathbf{1}\{J_s \in \pi\backslash\pi'\}\right)$$

    $\mathcal{C}_{k+1} = \mathcal{C}_k \cup \{\pi \in \mathcal{A}_k \setminus \mathcal{C}_k : \widehat{FDR}(\pi) + C_1(\pi, t_k, \delta_k)/|\pi| \leq \alpha\}$
    $\mathcal{A}_{k+1} = \mathcal{A}_k$
    Remove any π from $\mathcal{A}_{k+1}$ and $\mathcal{C}_{k+1}$ such that one of the conditions is true:

1. $\widehat{FDR}(\pi) - C_1(\pi, t_k, \delta_k)/|\pi| > \alpha$
2. $\exists\pi' \in \mathcal{C}_{k+1}$ with $\widehat{TP}(\pi'\backslash\pi) - \widehat{TP}(\pi\backslash\pi') > C_2(\pi, \pi', t_k, \delta_k)$ and add π to a set R

    Remove any π from $\mathcal{A}_{k+1}$ and $\mathcal{C}_{k+1}$ such that:

3. $\exists\pi' \in \mathcal{C}_{k+1} \cup R$, such that $\pi \subset \pi'$.

    Set $S_{k+1} := \bigcup_{\pi\in\mathcal{A}_{k+1}\setminus\mathcal{C}_{k+1}} \pi,$ and $T_{k+1} = \bigcup_{\pi\in\mathcal{A}_{k+1}} \pi - \bigcap_{\pi\in\mathcal{A}_{k+1}} \pi.$
    $k \leftarrow k + 1$
  endif

  Stochastic Noise:
    if $|\mathcal{A}_k| > 1$, **Break**. Otherwise:
    Sample $I_t \sim$ Unif($[n]$). If $I_t \in S_k$, then receive a label $Y_{I_t,t} \sim$ Ber(η_{I_t}).
    Sample $J_t \sim$ Unif($[n]$). If $J_t \in T_k$, then receive a label $Y'_{J_t,t} \sim$ Ber(η_{J_t}).
  Persistent Noise:
    If $|\mathcal{A}_k| = 1$ or $t > n$, **Break.** Otherwise:
    Sample $I_t \sim [n]\backslash\{I_s : 1 \leq s < t\}$. If $I_t \in S_k$, then receive a label $Y_{I_t,t} = \eta_{I_t}$.
    Sample $J_t \sim [n]\backslash\{J_s : 1 \leq s < t\}$. If $J_t \in T_k$, then receive a label $Y'_{J_t,t} = \eta_{J_t}$.
  Return $\max_{t\in\mathcal{C}_{k+1}} \widehat{TP}(\pi)$

Figure 3: Active FDR control in persistent and bounded noise settings.

4 Combinatorial FDR Control

In this section we focus on the question of finding a set FDR controlled at level α that has maximal TPR, i.e. π_α^*. The algorithm of Figure 3 provides an active sampling method for determining

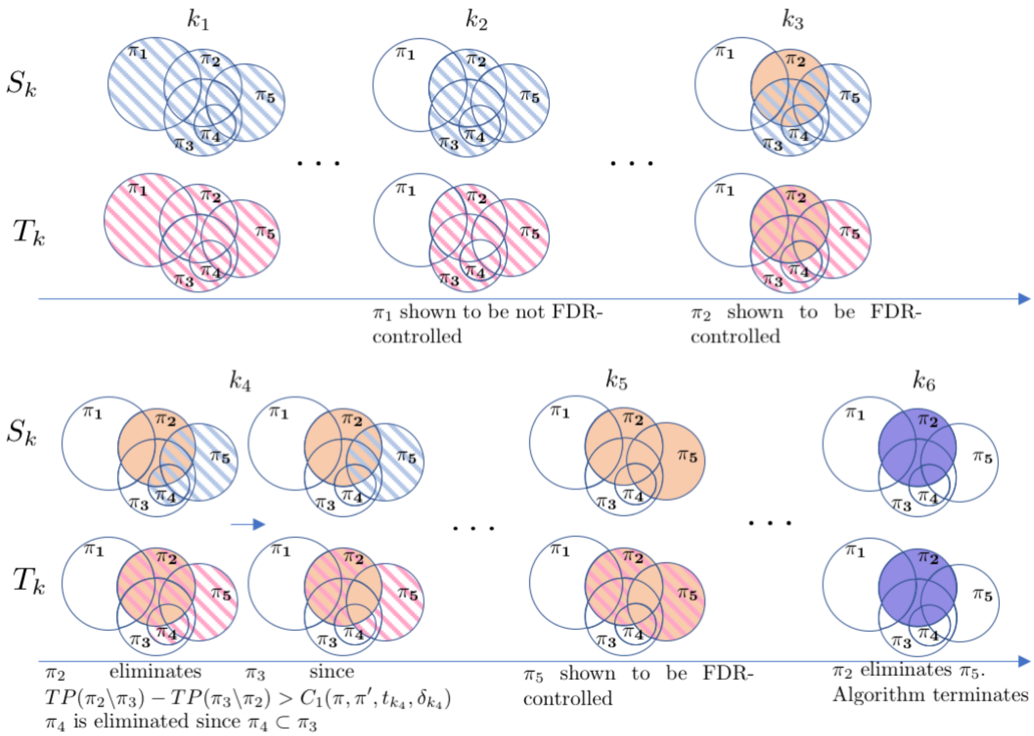

Figure 4: Example run of the algorithm of Figure 3. Example run of Algorithm 3, showing the evolution of sampling regions S_k (blue stripes), T_k (pink stripes) and FDR controlled set C_k (orange fill) at each time k_t.

$\pi \in \Pi$ with $FDR(\pi) \leq \alpha$ and maximal TPR. Since $TPR(\pi) = TP(\pi)/\eta_{[n]}$, we can ignore the denominator and so maximizing the TPR is the same as maximizing TP.

The algorithm proceeds in epochs that occur in dyadic powers. At all times a collection $\mathcal{A}_k \subseteq \Pi$ of active sets is maintained, and in addition a collection of FDR-controlled sets $\mathcal{C}_k \subseteq \mathcal{A}_k$ is also maintained. In each time step, random indexes I_t and J_t are sampled from the union $S_k = \cup_{\pi \in \mathcal{A}_k \backslash \mathcal{C}_k} \pi$ and the symmetric difference $T_k = \cup_{\pi \in \mathcal{A}_k} \pi - \cap_{\pi \in \mathcal{A}_k} \pi$ respectively. Associated random labels $Y_{I_t,t}, Y_{J_t,t} \in \{0, 1\}$ are then obtained from the underlying label distributions $\text{Ber}(\eta_{I_t})$ and $\text{Ber}(\eta_{J_t})$, respectively. At the start of each epoch, any set whose FDR is statistically known to be under α is added to $\mathcal{C}_k$, and likewise any sets whose FDR are greater than α are removed from $\mathcal{A}_k$ in condition 1. Similar to the combinatorial bandits algorithm of Figure 2 for combinatorial bandits, a set $\pi \in \mathcal{A}_k$ is removed in condition 2 if $TP(\pi)$ is shown to be statistically less than $TP(\pi')$ for some $\pi' \in \mathcal{C}_k$. However, crucially, we require the set doing the eliminaton, π' to be FDR controlled. Indeed, in general we expect many sets with higher TP than π^*_α that are not FDR-controlled. Finally in condition 3, we exploit an added advantage of the positivity of the η_i's: if $\pi \subset \pi'$ then deterministically $TP(\pi) \leq TP(\pi')$, so if π' is FDR controlled it can be used to knock out π.

The choice of T_k is motivated by combinatorial bandits, we only need to sample in the symmetric difference. Of course, our goal is not merely to determine if a set π has higher TP or not, also to ensure that π is FDR-controlled. To accomplish the latter, it is important that we sample in the entirety of the union of all $\pi \in \mathcal{A}_k \setminus \mathcal{C}_k$, not just the symmetric difference of the sets of $\mathcal{A}_k$ which motivates the choice of S_k.

Figure 4 demonstrates a model run of the algorithm in the case of five sets $\Pi = \{\pi_1, \dots, \pi_5\}$. At distinct times in the run of the algorithm $k_1, k_2, \dots$ the figure denotes S_k and T_k at time k by stripes inside the regions carved up by the sets of Π. Note that at time k_4, regardless of whether or not π_4 is FDR-controlled, $\pi_4 \subset \pi_3$ and because π_3 gets knocked out by $\pi_2 = \pi_\alpha^*$ (condition 2), there is no need to add it to $\mathcal{C}_k$ (this sequence demonstrates the elimination of π_4 by condition 3). By k_5, π_2, π_5 are the only remaining sets in $\mathcal{A}_k$. In general if $TP(\pi_2) \approx TP(\pi_5)$ it could take a large number of samples after round k_5 to knock out π_5, but samples are only taken in their symmetric difference and not their union, because they are both FDR-controlled. Eventually, it is determined that $\pi_\alpha^* = \pi_2$.

4.1 Sample Complexity for Active FDR Control

Recall that Π_α is the subset of Π that is FDR-controlled. The following gives a sample complexity result for the number of rounds before the algorithm terminates.

Theorem 3 (Stochastic Noise)**.** *Assume that for each $i \leq n$ there is an associated $\eta_i \in [0,1]$ and $\{Y_{i,j}\}_{j=1}^{\infty}$ is an i.i.d. sequence of random variables such that $Y_{i,j} \sim Ber(\eta_i)$. For any $\pi \in \Pi$ define $\Delta_{\pi,\alpha} = |FDR(\pi) - \alpha|$, and $\widetilde{\Delta}_\pi = |TP(\pi_\alpha^* \setminus \pi) - TP(\pi \setminus \pi_\alpha^*)|/|\pi \Delta \pi^*|$*

$$s_\pi^{FDR} = \frac{V_\pi n}{|\pi|} \frac{1}{\Delta_{\pi,\alpha}^2} \log\left(n \log(\Delta_{\pi,\alpha}^{-2})/\delta\right), \quad s_\pi^{TP} = \frac{V_{\pi,\pi_\alpha^*} n}{|\pi \Delta \pi_\alpha^*|} \frac{1}{\widetilde{\Delta}_\pi^2} \log\left(n \log(\widetilde{\Delta}_\pi^{-2})/\delta\right)$$

In addition define $T_\pi^{FDR} = \min\left\{s_\pi^{FDR},\ \max\{s_\pi^{TP},\ s_{\pi_\alpha^*}^{FDR}\},\ \min_{\substack{\pi' \in \Pi_\alpha \\ \pi \subset \pi'}} s_{\pi'}^{FDR}\right\}$ and $T_\pi^{TP} = \min\left\{\max\{s_\pi^{TP},\ s_{\pi_\alpha^*}^{FDR}\},\ \min_{\substack{\pi' \in \Pi_\alpha \\ \pi \subset \pi'}} s_{\pi'}^{FDR}\right\}$. For a fixed constant c, with probability greater than $1-\delta$, Algorithm 3 returns π_α^* after a number of samples no more than

$$c \sum_{i=1}^{n} \frac{1}{n}\Big(\max_{\pi \in \Pi : i \in \pi} T_\pi^{FDR} + \max_{\pi \notin \Pi_\alpha : i \in \pi \Delta \pi_\alpha^*} T_\pi^{FDR} + \max_{\pi \in \Pi_\alpha : i \in \pi \Delta \pi_\alpha^*} T_\pi^{TP} \Big) \tag{2}$$

In practical experiments persistent noise is not uncommon and also avoids the potential for unbounded sample complexities that potentially occur when $FDR(\pi) \approx \alpha$.

Theorem 4 (Persistent Noise)**.** *Assume that for each $i \leq n$ there is an associated $\eta_i \in \{0,1\}$. Adopting the notation from Theorem 3, for a fixed constant c, with probability greater than $1-\delta$, Algorithm 3 returns π_α^* after a number of samples no more than*

$$c \sum_{i=1}^{n} \min\left\{1, \frac{1}{n}\Big(\max_{\pi \in \Pi : i \in \pi} T_\pi^{FDR} + \max_{\pi \notin \Pi_\alpha : i \in \pi \Delta \pi_\alpha^*} T_\pi^{FDR} + \max_{\pi \in \Pi_\alpha : i \in \pi \Delta \pi_\alpha^*} T_\pi^{TP} \Big)\right\} \tag{3}$$

Note that an FDR-controlled set, i.e. $\pi \in \Pi_\alpha$ is only removed from $\mathcal{A}_k$ if it is eliminated by an FDR-controlled set with higher TP (condition 2 and condition 3). The time T_π^{TP} represents when this happens. Likewise, T_π^{FDR} represents the time at which we can conclude that π has left $\mathcal{A}_k \setminus \mathcal{C}_k$.

In every round, any given i can be sampled at most twice, once for if it is in S_k and once for if it is in T_k. Thus in expectation $i \in [n]$ contributes $2/n$ to the total sample complexity while there is either i) a set π such that $\pi \in \mathcal{A}_k \backslash \mathcal{C}_k$ with $i \in \pi$, i.e. $i \in S_k$, or ii) $i \in \pi \Delta \pi'$ for some $\pi' \in \mathcal{A}_k$, i.e. $i \in T_k$. The first case corresponds to the first term in (2) $\max_{\pi \in \Pi, i \in \pi} T_\pi^{FDR}$. The second and third terms correspond to the second case–note that the middle term is $\max_{i\in\pi\Delta\pi_\alpha^*:\pi\notin\Pi_\alpha} T_\pi^{FDR}$ since any set $\pi \notin \Pi_\alpha$ is removed at time T_π^{FDR}.

An important takeway of the result is that the sample complexity is at least the number of samples to show that π_α^* is FDR-controlled at α plus the number of samples needed to play the combinatorial bandit game of removing each set $\pi \in \Pi_\alpha$. In the next section we will highlight this effect in the case of one-dimensional thresholds. Finally, we highlight the active gains of our algorithm. A naive passive algorithm that continues to sample until both the FDR of every set is determined, and π_α^* is provably better than every other FDR controlled set gives a sample complexity of $O(n \max\{\max_{\pi\in\Pi_\alpha} T_\pi^{FDR}, \max_{\pi\notin\Pi_\alpha} T_\pi^{TP}\})$.

5 Contrasting Active Classification with Active FDR Control

A particularly insighful case of FDR-control is that of **Thresholds** where $\Pi = \{[t], t \leq n\}$. Let $\pi_\alpha^* = [t_\alpha^*]$, where $t_\alpha^* = \max\{t \in [n] : FDR([t]) \leq \alpha\}$. In this case, up to $\log\log$ factors, the sample complexity takes on a particularly nice form,

$$t_\alpha^* \max_{t\geq t_\alpha^*} \frac{\log(1/\delta)}{t\Delta_{[t],\alpha}^2} + \sum_{i>t^*} \max_{t'>i} \frac{\log(1/\delta)}{t'\Delta_{[t'],\alpha}^2}$$

Indeed, by definition $[t]$ is not FDR-controlled for any $t > t_\alpha^*$. Hence we will only stop sampling $i \leq t_\alpha^*$ once each $[t]$ for all $t > t^*$ is shown to not be FDR-controlled and $[t_\alpha^*]$ is shown to be FDR-controlled. If $i > t_\alpha^*$, we will stop sampling i once every $t \geq i$ is shown to no longer be FDR-controlled. At any time $T_k \subset [\max\{t, [t] \in \mathcal{A}_k\}]$, so we only incur a constant factor more samples by sampling in the symmetric difference.

We conclude that for this particular class, the algorithm in the end-game will be sampling uniformly on a set slightly larger than $[t_\alpha^*]$. Contrast this with classification that would be spending most of its samples on a neighborhood round the lowest risk threshold (if the Bayes classifier is well-approximated by the class) that may be significantly smaller or larger based on the value of α. In general, the end-game sampling will sample non-trivially in S_k and T_k as illustrated in Figure 4.

References

[1] Shivani Agarwal, Thore Graepel, Ralf Herbrich, Sariel Har-Peled, and Dan Roth. Generalization bounds for the area under the roc curve. *Journal of Machine Learning Research*, 6(Apr):393–425, 2005.

[2] Maria-Florina Balcan, Alina Beygelzimer, and John Langford. Agnostic active learning. *Journal of Computer and System Sciences*, 75(1):78–89, 2009.

[3] Rémi Bardenet, Odalric-Ambrym Maillard, et al. Concentration inequalities for sampling without replacement. *Bernoulli*, 21(3):1361–1385, 2015.

[4] Paul N Bennett, David M Chickering, Christopher Meek, and Xiaojin Zhu. Algorithms for active classifier selection: Maximizing recall with precision constraints. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 711–719. ACM, 2017.

[5] Alina Beygelzimer, Sanjoy Dasgupta, and John Langford. Importance weighted active learning. *arXiv preprint arXiv:0812.4952*, 2008.

[6] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[7] Olivier Bousquet. A bennett concentration inequality and its application to suprema of empirical processes. *Comptes Rendus Mathematique*, 334(6):495–500, 2002.

[8] Kendrick Boyd, Kevin H Eng, and C David Page. Area under the precision-recall curve: Point estimates and confidence intervals. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 451–466. Springer, 2013.

[9] Diogo M Camacho, Katherine M Collins, Rani K Powers, James C Costello, and James J Collins. Next-generation machine learning for biological networks. *Cell*, 2018.

[10] Tongyi Cao and Akshay Krishnamurthy. Disagreement-based combinatorial pure exploration: Efficient algorithms and an analysis with localization. *arXiv preprint arXiv:1711.08018*, 2017.

[11] Rui M Castro and Robert D Nowak. Minimax bounds for active learning. *IEEE Transactions on Information Theory*, 54(5):2339–2353, 2008.

[12] Lijie Chen, Anupam Gupta, and Jian Li. Pure exploration of multi-armed bandit under matroid constraints. In *Conference on Learning Theory*, pages 647–669, 2016.

[13] Lijie Chen, Anupam Gupta, Jian Li, Mingda Qiao, and Ruosong Wang. Nearly optimal sampling algorithms for combinatorial pure exploration. In *Conference on Learning Theory*, pages 482–534, 2017.

[14] Lijie Chen, Jian Li, and Mingda Qiao. Nearly instance optimal sample complexity bounds for top-k arm selection. In *Artificial Intelligence and Statistics*, pages 101–110, 2017.

[15] Shouyuan Chen, Tian Lin, Irwin King, Michael R Lyu, and Wei Chen. Combinatorial pure exploration of multi-armed bandits. In *Advances in Neural Information Processing Systems*, pages 379–387, 2014.

[16] Sanjoy Dasgupta. Coarse sample complexity bounds for active learning. In *Advances in neural information processing systems*, pages 235–242, 2006.

[17] Sanjoy Dasgupta, Daniel J Hsu, and Claire Monteleoni. A general agnostic active learning algorithm. In *Advances in neural information processing systems*, pages 353–360, 2008.

[18] Devdatt P Dubhashi and Alessandro Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.

[19] Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of machine learning research*, 7(Jun):1079–1105, 2006.

[20] Victor Gabillon, Alessandro Lazaric, Mohammad Ghavamzadeh, Ronald Ortner, and Peter Bartlett. Improved learning complexity in combinatorial pure exploration bandits. In *Artificial Intelligence and Statistics*, pages 1004–1012, 2016.

[21] Steve Hanneke. Teaching dimension and the complexity of active learning. In *International Conference on Computational Learning Theory*, pages 66–81. Springer, 2007.

[22] Steve Hanneke et al. Theory of disagreement-based active learning. *Foundations and Trends® in Machine Learning*, 7(2-3):131–309, 2014.

[23] Tzu-Kuo Huang, Alekh Agarwal, Daniel J Hsu, John Langford, and Robert E Schapire. Efficient and parsimonious agnostic active learning. In *Advances in Neural Information Processing Systems*, pages 2755–2763, 2015.

[24] Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil'ucb: An optimal exploration algorithm for multi-armed bandits. In *Conference on Learning Theory*, pages 423–439, 2014.

[25] Shivaram Kalyanakrishnan, Ambuj Tewari, Peter Auer, and Peter Stone. Pac subset selection in stochastic multi-armed bandits. In *ICML*, volume 12, pages 655–662, 2012.

[26] Zohar Karnin, Tomer Koren, and Oren Somekh. Almost optimal exploration in multi-armed bandits. In *International Conference on Machine Learning*, pages 1238–1246, 2013.

[27] Armaghan W Naik, Joshua D Kangas, Devin P Sullivan, and Robert F Murphy. Active machine learning-driven experimentation to determine compound effects on protein patterns. *Elife*, 5:e10047, 2016.

[28] Robert D Nowak. The geometry of generalized binary search. *IEEE Transactions on Information Theory*, 57(12):7893–7906, 2011.

[29] Gabriel J Rocklin, Tamuka M Chidyausiku, Inna Goreshnik, Alex Ford, Scott Houliston, Alexander Lemak, Lauren Carter, Rashmi Ravichandran, Vikram K Mulligan, Aaron Chevalier, et al. Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science*, 357(6347):168–175, 2017.

[30] Ashish Sabharwal and Yexiang Xue. Adaptive stratified sampling for precision-recall estimation. pages 825–834, 2018.

[31] Christoph Sawade, Niels Landwehr, and Tobias Scheffer. Active estimation of f-measures. In *Advances in Neural Information Processing Systems*, pages 2083–2091, 2010.

[32] Burr Settles. Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 6(1):1–114, 2012.

[33] Max Simchowitz, Kevin Jamieson, and Benjamin Recht. The simulator: Understanding adaptive sampling in the moderate-confidence regime. In *Conference on Learning Theory*, pages 1794–1834, 2017.

[34] Yuriy Sverchkov and Mark Craven. A review of active learning approaches to experimental design for uncovering biological networks. *PLoS computational biology*, 13(6):e1005466, 2017.

[35] Lorillee Tallorin, JiaLei Wang, Woojoo E Kim, Swagat Sahu, Nicolas M Kosa, Pu Yang, Matthew Thompson, Michael K Gilson, Peter I Frazier, Michael D Burkart, et al. Discovering de novo peptide substrates for enzymes using machine learning. *Nature communications*, 9(1):5253, 2018.

[36] Lu Zhang, Jianjun Tan, Dan Han, and Hao Zhu. From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug discovery today*, 22(11):1680–1685, 2017.

A Proofs

A.1 Confidence Bounds for Combinatorial Bandits

In this section, we build confidence intervals useful in our general combinatorial bandit setup. The union bounds presented are motivated by those in [10]. The constants used in the case without replacement are motivated by Corollary 3.6 in [3].

Lemma 2. *Assume that for each arm $i \leq n$ there is an associated distribution ν_i with support $[-1, 1]$, mean μ_i and variance $\sigma_i^2 \leq 1$. Assume access to the observations $(I_1, y_{I_1}) \cdots, (I_t, y_{I_t})$ in two different but related settings.*

1. ***Stochastic Noise*** *$I_k \sim$ Unif$([n])$ and $y_{I_k} \sim \nu_{I_k}$.*
2. ***Persistent Noise*** *$I_k \in [n]$ are drawn without replacement, $y_{I_k} = \mu_{I_k}$, $t \leq n$*

Let $\widehat{\mu}_\pi = \frac{n}{T}\sum_{k=1}^{T} y_k \mathbf{1}\{I_k \in \pi\}$. Then

1. *With probability greater than $1-\delta$ for all $\pi \in \Pi$*

$$|\widehat{\mu}_\pi - \mu_\pi| \leq C_1(\pi, t, \delta) := \sqrt{\frac{4\rho_t|\pi|nV_\pi \log\left(\frac{n}{\delta}\right)}{t}} + \frac{4n\kappa_t V_\pi \log\left(\frac{n}{\delta}\right)}{3t} \tag{4}$$

2. *Fix $\pi' \in \Pi$. With probability greater than $1-\delta$ for all $t > 0$ and $\pi \in \Pi$*

$$\left|\widehat{\mu}_{\pi'\setminus\pi} - \widehat{\mu}_{\pi\setminus\pi'} - (\mu_{\pi'\setminus\pi} - \mu_{\pi\setminus\pi'})\right| \leq C_2(\pi, \pi', t, \delta) := \sqrt{\frac{8\rho_t|\pi\Delta\pi'|nV_{\pi,\pi'} \log\left(\frac{n}{\delta}\right)}{t}} \tag{5}$$

$$+ \frac{4\kappa_t nV_{\pi,\pi'} \log\left(\frac{n}{\delta}\right)}{3t} \tag{6}$$

where $\rho_t, \kappa_t = 1$ in the stochastic case and in the persistent case

$$\rho_t = \begin{cases} 1 - \frac{t-1}{n} & t \leq n/2 \\ 1 - \frac{t}{n} & t \geq n/2 \end{cases} \qquad \kappa_t = \frac{4}{3} + \begin{cases} \sqrt{\frac{t(t-1)}{n(n-t+1)}} & t \leq n/2 \\ \sqrt{\frac{(n-t-1)(n-t)}{(t+1)n}} & t \geq n/2 \end{cases}$$

Note that by negative associativity the confidence bounds that hold in the case of sampling with replacement also hold when sampling without replacement.

Proof. Define the complexity measures

$$B_1(k) = \{\pi \in \mathcal{A} : |\pi| = k\}, B_2(k, \pi') = \{\pi \in \mathcal{A} : |\pi\Delta\pi'| = k\}.$$

Firstly note that for any $\pi \in \Pi$

$$\begin{aligned}
\text{var}(\widehat{\mu}_\pi) &= \frac{n^2}{T}\text{var}\left(y_1\mathbf{1}\{I_1 \in \pi\}\right) \\
&= \frac{n^2}{T}\left(\mathbb{E}[y_1^2\mathbf{1}\{I_1 \in \pi \setminus \pi'\}] - \left(\frac{1}{n}\sum_{i\in\pi\setminus\pi'}\mu_i\right)^2\right) \\
&\leq \frac{n^2}{T}\left(\frac{1}{n}\sum_{i\in\pi}(\sigma_i^2 + \mu_i^2)\right) \leq \frac{2|\pi|n}{T}
\end{aligned}$$

Thus by Bernstein's inequality and a union bound,

$$\mathbb{P}\left(\exists \pi \in \Pi : |\widehat{\mu}_\pi - \mu_\pi| > \sqrt{\frac{2|\pi| n \log(nB_1(|\pi|)/\delta)}{T}} + \frac{2n\log(nB_1(|\pi|)/\delta)}{3T}\right) \leq \sum_{\pi \in \Pi} \frac{\delta}{nB(k)}$$
$$\leq \sum_{k=1}^{n} B(k) \frac{\delta}{nB(k)} \leq \delta$$

For the second assertion, firstly note that for any π, π', $\widehat{\mu}_\pi - \widehat{\mu}_{\pi'} = \widehat{\mu}_{\pi\setminus\pi'} - \widehat{\mu}_{\pi'\setminus\pi}$ and so

$$\begin{aligned}
&\mathrm{var}(\widehat{\mu}_\pi - \widehat{\mu}_{\pi'}) \\
&\quad = \mathrm{var}(\widehat{\mu}_{\pi\setminus\pi'} - \widehat{\mu}_{\pi'\setminus\pi}) \\
&\quad = \mathrm{var}(\widehat{\mu}_{\pi\setminus\pi'}) + \mathrm{var}(\widehat{\mu}_{\pi'\setminus\pi})) \\
&\quad = \frac{n^2}{T}\mathrm{var}\left(y_1 \mathbf{1}\{I_1 \in \pi \setminus \pi'\}\right) + \frac{n^2}{T}\mathrm{var}\left(y_1 \mathbf{1}\{I_1 \in \pi \setminus \pi'\}\right) \\
&\quad = \frac{n^2}{T}\left(\mathbb{E}[y_1^2 \mathbf{1}\{I_1 \in \pi\setminus\pi'\}] - \left(\frac{1}{n}\sum_{i\in\pi\setminus\pi'} \mu_i\right)^2 + \mathbb{E}[y_1^2 \mathbf{1}\{I_1 \in \pi'\setminus\pi\}] - \left(\frac{1}{n}\sum_{i\in\pi'\setminus\pi} \mu_i\right)^2\right) \\
&\quad \leq \frac{n^2}{T}\left(\frac{1}{n}\sum_{i\in\pi\setminus\pi'}(\sigma_i^2 + \mu_i^2) + \frac{1}{n}\sum_{i\in\pi'\setminus\pi}(\sigma_i^2 + \mu_i^2)\right) \\
&\quad \leq \frac{4|\pi\Delta\pi'|n}{T}
\end{aligned}$$

Let $b_\pi = \max\{|B_2(|\pi\Delta\pi'|, \pi)|, |B_2(|\pi\Delta\pi'|, \pi')|\}$

$$\begin{aligned}
\mathbb{P}&\left(\exists \pi, \pi' \in \Pi : |\widehat{\mu}_\pi - \widehat{\mu}_{\pi'}| > \sqrt{\frac{8|\pi\Delta\pi'|\log(nb_\pi/\delta)}{T}} + \frac{2n\log(b_\pi/\delta)}{3T}\right) \\
&\leq \sum_{\pi\in\Pi} \frac{\delta}{nb_\pi} \\
&\leq \sum_{k=1}^{n}\sum_{\pi\in\Pi} \mathbf{1}\{|\pi\Delta\pi'| = k\}\frac{\delta}{nb_\pi} \\
&= \sum_{k=1}^{n}\sum_{\pi\in\Pi} \mathbf{1}\{|\pi\Delta\pi'| = k\}\frac{\delta}{n\max\{|B_2(|\pi\Delta\pi'|, \pi)|, |B_2(|\pi\Delta\pi'|, \pi')|\}} \\
&\leq \sum_{k=1}^{n}\sum_{\pi\in\Pi} \mathbf{1}\{|\pi\Delta\pi'| = k\}\frac{\delta}{n|B_2(|\pi\Delta\pi'|, \pi')|} \\
&\leq \sum_{k=1}^{n} \frac{\delta}{n} \leq \delta
\end{aligned}$$

Now by the Sauer-Shelah Lemma for any k

$$\log(B_1(k)) \leq V(B_1(k))\log(en/V(B_1(k))).$$

At the same time, $|B_1(k)| \leq |\{\pi \in \Pi : |\pi| = k\}| \leq n^k$. Hence

$$\begin{aligned}\log(n|B_1(k)|/\delta) &\leq \min\{V(B_1(k))\log(en/V(B_1(k))) + \log(n/\delta), (k+1)\log(n/\delta)\}\\ &\leq 4\min\{V(B_1(k)), k\}\log(en/\delta)\end{aligned}$$

Similarly for any k,

$$\log(B_2(k,\pi')) \leq V(B_2(k,\pi'))\log(en/V(B_2(k,\pi')))$$

and $|\{\pi \in \Pi : |\pi\Delta\pi^*| = k\}| = \binom{n}{k} \leq n^k$. In particular,

$$\begin{aligned}\log(n|B_2(k,\pi')|/\delta) &\leq \min\{V(B_2(k,\pi'))\log(en/V(B_2(k,\pi'))) + \log(n/\delta), (k+1)\log(n/\delta)\}\\ &\leq 4\min\{V(B_2(k,\pi')), k\}\log(en/\delta)\end{aligned}$$

So using identical logic

$$\begin{aligned}\log(nb_\pi/\delta) &\leq \log(n\max\{|B_2(|\pi\Delta\pi'|,\pi)|, |B_2(|\pi\Delta\pi'|,\pi')|\}/\delta)\\ &\leq \max\{\log(n|B_2(|\pi\Delta\pi'|,\pi)|/\delta), \log(n|B_2(|\pi\Delta\pi'|,\pi')|)/\delta)\}\\ &\leq 4\min\{\max\{V(B_2(|\pi\Delta\pi'|,\pi)), V(B_2(|\pi\Delta\pi'|,\pi'))\}, |\pi\Delta\pi'|\}\log(en/\delta)\end{aligned}$$

Finally, in the case of without replacement, we can use the confidence intervals from Theorem 3.6 of [3] and the result follows. □

A.2 Proof of Theorems 1 and 2

Proof. Throughout the following, let $\Delta_\pi := \mu_{\pi^*\setminus\pi} - \mu_{\pi\setminus\pi^*}$. Define

$$\mathcal{E} = \bigcap_{k\in\mathbb{N}}\bigcap_{\pi\in\Pi}\{|(\widehat{\mu}_{\pi_*,k} - \widehat{\mu}_{\pi,k}) - (\mu_{\pi_*} - \mu_\pi)| \leq C(\pi_*,\pi,t_k,\delta_k)\}$$

where we recall $C(\pi_*,\pi,t_k,\delta_k) = C(\pi,\pi_*,t_k,\delta_k)$. By Lemma 1 we have that $\mathbb{P}(\mathcal{E}) \geq 1 - \sum_{k=1}^\infty \delta_k \geq 1-\delta$ so assume $\mathcal{E}$ holds in what follows.

First we show $\pi_* \in \mathcal{A}_k$ for all k. Assume $\pi_* \in \mathcal{A}_k$. Then for any $\widehat{\pi} \in \mathcal{A}_k$ we have

$$\begin{aligned}\widehat{\mu}_{\widehat{\pi}\setminus\pi_*,k} - \widehat{\mu}_{\pi_*\setminus\widehat{\pi},k} &\overset{\mathcal{E}}{\leq} \mu_{\widehat{\pi}\setminus\pi_*} - \mu_{\pi_*\setminus\widehat{\pi}} + C(\widehat{\pi},\pi_*,t_k,\delta_k)\\ &\leq C(\widehat{\pi},\pi_*,t_k,\delta_k)\end{aligned}$$

which implies that $\pi_* \in \mathcal{A}_{k+1}$. The result follows by the fact that $\pi_* \in \mathcal{A}_0$.

Now we bound the expected number of samples taken. An arm i is sampled at time t if there are at least two policies $\pi,\pi' \in \mathcal{A}_t$ such that $i \in \pi\Delta\pi'$. Since we just showed that $\pi_* \in \mathcal{A}_t$ for all t, it follows that $\min\left\{k : \widehat{\mu}_{\pi_*\setminus\pi,k} - \widehat{\mu}_{\pi\setminus\pi_*,k} > C(\pi_*,\pi,t_k,\delta_k)\right\}$ is an upper bound on the number of rounds before π is removed from Π_t. Since $\mu_{\pi_*} > \mu_\pi$ for all $\pi \in \Pi$, for each $\pi \in \Pi$ there exists a random minimum round K_π such that

$$\widehat{\mu}_{\pi_*\setminus\pi,K_\pi} - \widehat{\mu}_{\pi\setminus\pi_*,K_\pi} \geq C(\pi_*,\pi,t_{K_\pi},\delta_{K_\pi}).$$

But for every $\pi \in \Pi$ and $k \in \mathbb{N}$ we have

$$\widehat{\mu}_{\pi_*\setminus\pi,k} - \widehat{\mu}_{\pi\setminus\pi_*,k} \overset{\mathcal{E}}{\geq} \Delta_\pi - C(\pi_*,\pi,t_k,\delta_k)sorry$$

so define

$$k_\pi := \min\{k : \Delta_\pi/2 \geq C(\pi_*, \pi, t_k, \delta_k)\}.$$

Also define $k_{\max} = \max_\pi k_\pi$ and note that $k_{\max}$ is finite since $C(\pi_*, \pi, t_k, \delta_k)$ is decreasing in k. Now we have that

$$\begin{aligned} S_k &= \{i \in [n] : \exists \pi \in \Pi : i \in \pi_* \Delta \pi, K_\pi \geq k\} \\ &\overset{\mathcal{E}}{\subseteq} \{i \in [n] : \exists \pi \in \Pi : i \in \pi_* \Delta \pi, k_\pi \geq k\} \\ &=: s_k \end{aligned}$$

Thus, we trivially have $\mathbf{1}\{I_s \in S_k\} \leq \mathbf{1}\{I_s \in s_k\}$ and whether or not I_s are drawn uniformly at random from $[n]$ (with replacement) or uniformly at random from $[n] \setminus \{i : I_s = i, 1 \leq s < t\}$ (without replacement for persistent noise), the I_s indices are negatively associated random variables [18]. Consequently, standard a multiplicative Chernoff bounds apply:

$$\begin{aligned} &\mathbb{P}\left(\sum_{k=1}^{k_{\max}} \sum_{s=t_{k-1}+1}^{t_k} \mathbf{1}\{I_s \in S_k\} \geq (1+r) \sum_{k=1}^{k_{\max}} t_k \frac{|s_k|}{n} \right) \\ &\qquad\qquad \leq \mathbb{P}\left(\sum_{k=1}^{k_{\max}} \sum_{s=t_{k-1}+1}^{t_k} \mathbf{1}\{I_s \in s_k\} \geq (1+r) \sum_{k=1}^{k_{\max}} t_k \frac{|s_k|}{n} \right) \\ &\qquad\qquad \leq \exp\left(-\tfrac{\min\{r, r^2\}}{3} \sum_{k=1}^{k_{\max}} t_k \frac{|s_k|}{n} \right) \end{aligned}$$

Taking $r = \max\left\{ \frac{3\log(1/\delta)}{\sum_{k=1}^{k_{\max}} t_k \frac{|s_k|}{n}}, \sqrt{\frac{3\log(1/\delta)}{\sum_{k=1}^{k_{\max}} t_k \frac{|s_k|}{n}}} \right\}$ we have with probability at least $1-\delta$ that

$$\begin{aligned} \sum_{k=1}^{k_{\max}} \sum_{s=t_{k-1}+1}^{t_k} \mathbf{1}\{I_s \in S_k\} &\leq \max\left\{ 3\log(1/\delta), \sqrt{3\log(1/\delta) \sum_{k=1}^{k_{\max}} t_k \frac{|s_k|}{n}} \right\} + \sum_{k=1}^{k_{\max}} t_k \frac{|s_k|}{n} \\ &\leq \frac{9}{2}\log(1/\delta) + \frac{3}{2} \sum_{k=1}^{\infty} t_k \frac{|s_k|}{n} \end{aligned}$$

where the last inequality follows by the arithmetic-geometric mean inequality. Now

$$\begin{aligned} \sum_{k=1}^{\infty} t_k \frac{|s_k|}{n} &= \sum_{k=1}^{\infty} t_k \sum_{i=1}^{n} \frac{1}{n} \mathbf{1}\{\exists \pi \in \Pi : i \in \pi_* \Delta \pi, k_\pi \geq k\} \\ &= \sum_{k=1}^{\infty} \sum_{i=1}^{n} \frac{t_k}{n} \mathbf{1}\{\exists \pi \in \Pi : i \in \pi_* \Delta \pi, k_\pi \geq k\} \\ &= \sum_{i=1}^{n} \sum_{k=1}^{\infty} \frac{2^k}{n} \mathbf{1}\{\exists \pi \in \Pi : i \in \pi_* \Delta \pi, 2^{k_\pi} \geq 2^k\} \\ &\leq \sum_{i=1}^{n} \max_{\pi \in \Pi : i \in \pi_* \Delta \pi} \frac{2^{k_\pi + 1}}{n} \end{aligned}$$

Now, using the specific confidence interval $C_2(\pi', \pi, t_k, \delta_k)$ from 1

$$\begin{aligned}
2^{k^{\pi}} &\le 2 \min\{t \in \mathbb{N} : \Delta_\pi/2 < C_2(\pi_*, \pi, t, \delta_{\lceil \log_2 t \rceil})\} \\
&\le c_1 n V_{\pi,\pi'} \left(\frac{|\pi^* \Delta \pi|}{\Delta_\pi^2} + \frac{1}{\Delta_\pi} \right) \log \left(\frac{n \log\left(\Delta_\pi^{-2}\right)}{\delta} \right) \\
&\le c_2 n V_{\pi,\pi'} \frac{|\pi^* \Delta \pi|}{\Delta_\pi^2} \log \left(\frac{n \log\left(\Delta_\pi^{-2}\right)}{\delta} \right) \\
&\le c_2 \frac{n V_{\pi,\pi'}}{|\pi^* \Delta \pi|} \frac{1}{\widetilde{\Delta}_\pi^2} \log \left(\frac{n \log\left(\widetilde{\Delta}_\pi^{-2}\right)}{\delta} \right)
\end{aligned}$$

where the second to last line follows from

$$\frac{|\pi^* \Delta \pi|}{\Delta_\pi^2} + \frac{1}{\Delta_\pi} \le \frac{1}{\Delta_\pi} \left(\frac{|\pi^* \Delta \pi|}{\Delta_\pi} + 1 \right) \le \frac{2|\pi^* \Delta \pi|}{\Delta_\pi^2}$$

since $\Delta_\pi \le |\pi_* \Delta \pi|$. But for the persistent noise case we have $k_\pi \le \log_2(n)$ which implies for any i, $\max_{\pi \in \Pi : i \in \pi_* \Delta \pi} \frac{2^{k_\pi + 1}}{n} \le 2$. The result now follows. □

B Proof of Theorem 3

Proof. **Step 1: Correctness** Let $t_k = 2^k$. Let $\mathcal{E}$ be the event that, for each k and for each $\pi \in \Pi$,

$$\left| \widehat{FDR}(\pi) - FDR(\pi) \right| < C_1(\pi_t, n, t_k, \delta_k)/|\pi|$$

and

$$|(\widehat{TP}(\pi^* \setminus \pi) - \widehat{TP}(\pi \setminus \pi^*)) - (TP(\pi^* \setminus \pi) - TP(\pi \setminus \pi^*))| \le C_2(\pi^*, \pi, t_k, \delta_k).$$

By Lemma 1 and a union bound,

$$\mathbb{P}(\mathcal{E}^c) \le \sum_{k \ge 1} 2 \frac{2\delta}{8k^2} \le \delta$$

First we argue that π^* is never eliminated on event $\mathcal{E}$. Note that since $FDR(\pi^*) < \alpha$

$$\begin{aligned}
\widehat{FDR}(\pi^*) - \alpha &\overset{\mathcal{E}}{\le} FDR(\pi^*) - \alpha + C_1(\pi, t_k, \delta_k)/|\pi| \\
&< C_1(\pi, t_k, \delta_k)/|\pi|
\end{aligned}$$

Also for any $\pi \in \Pi_\alpha$,

$$\begin{aligned}
\widehat{TP}(\pi \setminus \pi^*) - \widehat{TP}(\pi^* \setminus \pi) &\overset{\mathcal{E}}{\le} TP(\pi \setminus \pi^*) - TP(\pi^* \setminus \pi) + C_2(\pi, \pi^*, t_k, \delta_k) \\
&= TP(\pi) - TP(\pi_*) + C_2(\pi, \pi^*, t_k, \delta_k) \\
&\le C_2(\pi, \pi^*, t_k, \delta_k)
\end{aligned}$$

and by definition π^* is the maximal TP set in Π_α so π^* will never be removed by another π.

Finally note that on event $\mathcal{E}$, any π' (not just π_*) can knock out π using line 2 or 3 of the algorithm iff $TP(\pi') > TP(\pi)$ and $\pi' \in \Pi_\alpha$.

We define a few key random rounds

$$K_\pi := \max\{k : \pi \in \mathcal{A}_k\}$$
$$K_\pi^{FDR,1} := \max\{k : \pi \in \mathcal{A}_k \setminus \mathcal{C}_k\}$$
$$K_\pi^{FDR,2} := \min\{k : |\widehat{FDR}(\pi) - \alpha| > C_1(\pi, t_k, \delta_k)\}$$
$$K_\pi^{TP} := \min\{k : \exists \pi' \in \mathcal{C}_k \text{ such that } \widehat{TP}(\pi' \setminus \pi) - \widehat{TP}(\pi \setminus \pi') > C_2(\pi', \pi, t_k, \delta_k)\}$$
$$K_\pi^{<} := \min\{k : \exists \pi' \in \mathcal{C}_k \text{ with } \pi \subset \pi'\}$$

Our objective is to bound $\max_{\pi \in \Pi \setminus \pi_*} K_\pi$, which marks the termination of the algorithm.

Bound on $K_\pi^{FDR,1}$: We begin by establishing a deterministic bound on $K_\pi^{FDR,1}$ that holds when event $\mathcal{E}$ is true. Note that $K_\pi^{FDR,1}$ is immediately before the first k such that $\pi \notin \mathcal{A}_k \setminus \mathcal{C}_k$. There are three ways this can occur: i) if π becomes FDR-controlled or if π is determined to not be FDR-controlled, and ii) a $\pi' \in C_k$ knocks out π using statistics about TP (i.e., line 2 of the algorithm), or iii) a $\pi' \in C_k$ knocks out π deterministically by line 3 of the algorithm. These cases are reflected with the $\min$ respectively:

$$K_\pi^{FDR,1} = \min\{K_\pi^{FDR,2}, K_\pi^{TP}, K_\pi^{<}\}.$$

We provide a bound for each one of these terms under $\mathcal{E}$.

- Since $C_1(\pi, t_k, \delta_k)$ is a decreasing function of k, note that

  $$|FDR(\pi) - \alpha| > 2C_1(\pi, t_k, \delta_k)/|\pi| \quad \implies \quad |\widehat{FDR}(\pi) - \alpha| > C_1(\pi, t_k, \delta_k)/|\pi|$$

  so on event $\mathcal{E}$, $K_\pi^{FDR,2} < k_\pi^{FDR,2}$ where

  $$k_\pi^{FDR,2} := \min\{k : \Delta_{\pi,\alpha}/2 > C_1(\pi, t_k, \delta_k)/|\pi|\}.$$

- On event $\mathcal{E}$, only sets from Π_α will enter $\mathcal{C}_k$, so only they can be used to knock out other sets in Line 2 of the algorithm. Since π^* is never eliminated on event $\mathcal{E}$, we have that:

  $$K_\pi^{TP} \overset{\mathcal{E}}{\leq} \min\{k : \pi^* \in \mathcal{C}_k \text{ and } \widehat{TP}(\pi^* \setminus \pi) - \widehat{TP}(\pi \setminus \pi^*) > C_2(\pi^*, \pi, t_k, \delta_k)\}.$$

  Thus denoting $\Delta_\pi = TP(\pi^* \setminus \pi) - TP(\pi \setminus \pi^*)$let

  $$k_\pi^{TP} = \min\{k : \Delta_\pi/2 > C_2(\pi^*, \pi, t_k, \delta_k) \text{ and } \Delta_{\pi^*,\alpha}/2 > C_1(\pi^*, t_k, \delta_k)/|\pi^*|\}$$

  and note that $K_\pi^{TP} \overset{\mathcal{E}}{\leq} k_\pi^{TP}$(note that this is potentially infinite if $TP(\pi) > TP(\pi_*)$).

- Using similar logic, on event $\mathcal{E}$ a set π' will knock out a set π using Line 3 of the algorithm only if π' is in $\mathcal{C}_k \cup R$ and $\pi \subset \pi'$. If $\pi' \in \mathcal{C}_k$ then $TP(\pi') \geq TP(\pi)$ so we can remove π. If $\pi' \in R$ but $\pi' \notin \mathcal{C}_k$ yet, there exists a $\pi'' \in \mathcal{C}_k$ (in particular, the π'' that eliminated π' into R) with $TP(\pi'') > TP(\pi') > TP(\pi)$ so we can safely remove π. Either way this implies that the $K_\pi^{<}$ is bounded by the time it takes to guarantee that π' is FDR-controlled, hence

  $$K_\pi^{<} \overset{\mathcal{E}}{\leq} \min_{\substack{\pi' \in \Pi_\alpha \\ \pi \subset \pi'}} K_{\pi'}^{FDR,2} \overset{\mathcal{E}}{\leq} \min_{\substack{\pi' \in \Pi_\alpha \\ \pi \subset \pi'}} k_{\pi'}^{FDR,2}.$$

Putting all of this together we set

$$k_\pi^{FDR,1} := \min\{k_\pi^{FDR,2}, k_\pi^{TP}, \min_{\substack{\pi\in\Pi_\alpha \\ \pi\subset\pi'}} k_{\pi'}^{FDR,2}\} \tag{7}$$

This is necessarily finite since $k_\pi^{FDR,2}$ is finite.

Summarizing:, on event $\mathcal{E}$, $k_\pi^{FDR,1}$ is an upper bound on $K_\pi^{FDR,1}$, the minimal round where $\pi \notin \mathcal{A}_{k+1} \setminus \mathcal{C}_{k+1}$.

Part 2 Bound on K_π: If $\pi \in \Pi_\alpha$, on event $\mathcal{E}$, π will be removed from $\mathcal{A}_k$ only when it demonstrably has lower TP than some other set $\pi' \in \Pi_\alpha$ regardless of whether it is in $\mathcal{C}_k$ or not. If $\pi \notin \Pi_\alpha$, on event $\mathcal{E}$, $K_\pi^{FDR,1} = K_\pi$, since the moment it's FDR is confirmed to be greater than α it is removed. Hence using the exact same logic as above, we have $K_\pi \overset{\mathcal{E}}{\leq} k_\pi$ where

$$k_\pi := \begin{cases} \min\{k_\pi^{TP}, \min_{\substack{\pi'\in\Pi_\alpha \\ \pi\subset\pi'}} k_{\pi'}^{FDR,2}\} & \pi \in \Pi_\alpha \\ k_\pi^{FDR,1} & \pi \notin \Pi_\alpha \end{cases} \tag{8}$$

Summarizing: On event $\mathcal{E}$, k_π is an upper bound on K_π and thus the algorithm terminates at some random round $K \leq k_{\max} := \max_{\pi\in\Pi\setminus\pi_*} k_\pi$ and outputs π_*.

Part 3: Bound the contribution of each arm. By the last step, we clearly have that the total sample complexity is bounded by

$$\sum_{k=1}^{k_{\max}} \sum_{t=t_{k-1}+1}^{t_k} \mathbf{1}\{I_t \in S_k\} + \mathbf{1}\{J_t \in T_k\}.$$

Since I_t, J_t are uniformly distributed over $[n]$, we have $\mathbb{E}[\mathbf{1}\{I_t \in S_k\}|S_k] = \frac{|S_k|}{n}$ and $\mathbb{E}[\mathbf{1}\{J_t \in T_k\}|T_k] = \frac{|T_k|}{n}$. However, because $|S_k|$ and $|T_k|$ are random variables, we will upper bound them by deterministic quantities, and then show that the sample complexity concentrates.

For each $i \in [n]$, in round k, note that arm $i \in S_k$ if there is a set $\pi \in \mathcal{A}_k \setminus \mathcal{C}_k$ with $i \in \pi$. Hence

$$S_k = \{i \in [n] : \exists \pi \in \Pi : K_\pi^{FDR,1} > k\} \overset{\mathcal{E}}{\subset} \{i \in [n] : \exists \pi \in \Pi : k_\pi^{FDR,1} > k\} =: \psi_k$$

Similarly, $i \in T_k$ if there is $\pi, \pi' \in \mathcal{A}_k$ with $i \in \pi\Delta\pi'$. On event $\mathcal{E}$, $\pi^* \in \mathcal{A}_k$ for all k, thus $i \in T_k$ iff $i \in \pi\Delta\pi^*$ for some $\pi \in \mathcal{A}_k$. Thus

$$T_k = \{\pi \in \Pi : i \in \pi\Delta\pi^*, K_\pi > k\} \overset{\mathcal{E}}{\subset} \{\exists \pi \in \Pi : i \in \pi\Delta\pi^*, k_\pi > k\} =: \tau_k$$

We now follow an argument similar to that in the proof of Theorem 1. Thus $\mathbf{1}\{I_t \in S_k\} \leq \mathbf{1}\{I_t \in \psi_k\}$ and $\mathbf{1}\{J_t \in T_k\} \leq \mathbf{1}\{J_t \in \tau_k\}$ regardless of whether I_t, J_t are drawn uniformly at random from $[n]$ or uniformly at random from $[n] \setminus \{i : I_s = i, 1 \leq s \leq t\}$ respectively $[n] \setminus \{i : J_s = i, 1 \leq s \leq t\}$. In particular, I_t, J_t are negatively associated so we can apply standard multiplicative Chernoff Bounds. In particular,

$$\begin{aligned}
\mathbb{P}\Big(\sum_{k=1}^{k_{\max}}\sum_{t=t_{k-1}+1}^{t_k}\mathbf{1}\{I_t\in S_k\}\geq(1+r)\sum_{k=1}^{k_{\max}}t_k\frac{|\psi_k|}{n}\Big)&\\
\leq\mathbb{P}\left(\sum_{k=1}^{k_{\max}}\sum_{t=t_{k-1}+1}^{t_k}\mathbf{1}\{I_t\in\psi_k\}\geq(1+r)\sum_{k=1}^{k_{\max}}t_k\frac{|\psi_k|}{n}\right)&\\
\leq\exp\left(-\frac{\min\{r,r^2\}}{3}\sum_{k=1}^{k_{\max}}t_k\frac{|\psi_k|}{n}\right)&
\end{aligned}$$

with the appropriate choice of r, with probability greater than $1-\delta$,

$$\sum_{k=1}^{k_{\max}}\sum_{t=t_{k-1}+1}^{t_k}\mathbf{1}\{I_t\in S_k\}\leq\frac{9}{2}\log(2/\delta)+\frac{3}{2}\sum_{k=1}^{\infty}t_k\frac{|\psi_k|}{n}$$

An identical argument gives that with probability greater than $1-\delta$,

$$\sum_{k=1}^{k_{\max}}\sum_{t=t_{k-1}+1}^{t_k}\mathbf{1}\{J_t\in T_k\}\leq\frac{9}{2}\log(2/\delta)+\frac{3}{2}\sum_{k=1}^{\infty}t_k\frac{|\tau_k|}{n}.$$

While we have provided a bound on the sample complexity in terms of deterministic quantities ψ_k and τ_k, we now want to provide natural and interpretable upper bounds on these quantities for a final result.

Putting it all together we have that

$$\begin{aligned}
\sum_{k=1}^{\infty}t_k\frac{\psi_k+\tau_k}{n}&=\sum_{k=1}^{\infty}\frac{2^k}{n}(\psi_k+\tau_k)\\
&=\sum_{i=1}^{n}\sum_{k=1}^{\infty}\frac{2^k}{n}(\mathbf{1}\{\exists\pi\in\Pi:i\in\pi,k_\pi^{FDR,1}>k\}\\
&\qquad+\mathbf{1}\{\exists\pi\in\Pi:i\in\pi\Delta\pi^*,k_\pi>k\})\\
&\leq\sum_{i=1}^{n}\sum_{k=1}^{\infty}\frac{2^k}{n}(\mathbf{1}\{\exists\pi\in\Pi:i\in\pi,k_\pi^{FDR,1}>k\}\\
&\qquad+\mathbf{1}\{\exists\pi\in\Pi,\pi\in\Pi_\alpha:i\in\pi\Delta\pi^*,k_\pi>k\}\\
&\qquad+\mathbf{1}\{\exists\pi\in\Pi,\pi\notin\Pi_\alpha:i\in\pi\Delta\pi^*,k_\pi>k\})\\
&\leq\sum_{i=1}^{n}\max_{i\in\pi}\frac{2^{k_\pi^{FDR,1}+1}}{n}+\max_{\substack{\pi\in\Pi_\alpha\\ i\in\pi\Delta\pi^*}}\frac{2^{k_\pi^{FDR,1}+1}}{n}+\max_{\substack{\pi\notin\Pi_\alpha\\ i\in\pi}}\frac{2^{k_\pi+1}}{n}
\end{aligned}$$

Solving for k, shows that for some constant c_1

$$\begin{aligned}
2^{k_\pi^{FDR,2}}&\leq\min\left\{m:2C(\pi,n,m,\delta_{\lfloor\log_2(m)\rfloor})<|FDR(\pi)-\alpha|\right\}\\
&\leq c_1nV_\pi\frac{\log(n\log(\Delta_{\pi,\alpha}^{-2})}{|\pi|\Delta_{\pi,\alpha}^2}
\end{aligned}$$

An identical argument shows that for arbitrary π, π', there is a constant c_2 such that

$$2^{k_\pi^{TP}} \leq \max\left\{ c_2 n V_{\pi,\pi^*} \left(\frac{|\pi \Delta \pi^*|}{\Delta_\pi^2} + \frac{1}{\Delta_\pi} \right) \log\left(\frac{n \log(\Delta_\pi^{-2})}{\delta} \right), 2^{k_{\pi^*}^{FDR,2}} \right\}$$
$$= \max\left\{ c_2 \frac{n V_{\pi,\pi^*}}{|\pi \Delta \pi^*|} \frac{1}{\widetilde{\Delta}_\pi^2} \log\left(\frac{n \log(\widetilde{\Delta}_\pi^{-2})}{\delta} \right), 2^{k_{\pi^*}^{FDR,2}} \right\}$$

But for the persistent noise case we have $k_\pi, k^{FDR,2} \leq \log_2(n)$ which implies for any i, $\max_{\pi \in \Pi: i \in \pi_* \Delta \pi} \frac{2^{k_\pi+1}}{n} \leq 2$. The theorem now follows.

□

C One-dimensional thresholds

We can get tighter characterizations of Lemma and consequently, better sample complexity guarantees for particular VC classes. In particular, those classes that have sets with substantial overlap like thresholds. In the case of **Thresholds** we have the following improvement that manages to remove the extra $\log(n)$ terms in Lemma 1.

Lemma 3. *Assume that for each $i \in [n]$ there is an associated distribution ν_i with support $[-1,1]$, mean μ_i and variance $\sigma_i^2 \leq 1$. Assume access to the observations $(y_1, I_1) \cdots, (y_T, I_T)$ where $I_k \sim$ Unif($[n]$) and $y_k \sim \nu_{I_k}$. Let $\widehat{\mu}_t = \frac{1}{T}\sum_{k=1}^T y_k \mathbf{1}\{I_k \leq t\}$. Fix $t' \leq n$. Then with probability greater than $1-\delta$ for any $s \leq n$,*

$$|\widehat{\mu}_s - \widehat{\mu}_{t'} - (\mu_s - \mu_{t'})| \leq \sqrt{\frac{2|s-t'|}{nT} \left(43 + 2\sqrt{2}\log(2\log_2^2(4|s-t'|)/3\delta)\right)} + \frac{12+\log\left(2\log_2^2(4|s-t'|)/3\delta\right)}{3T}$$

An analogous result can be proven in the persistent noise case of sampling without replacement.

Active Classification for One-dimensional thresholds with Tsybakov Noise - Let $h \in (0,1]$, $\alpha \geq 0$, $z \in [0,1]$ for some $i \in [n-1]$ and assume that $X_{i,j} \in \{-1,1\}$ are Bernoulli with $\mathbb{P}(X_{i,j} = \text{SIGN}(z - i/n)) = \frac{1}{2} + \frac{1}{2}h|z-i/n|^\alpha$ so that $\mu_i = h|z-i/n|^\alpha \text{SIGN}(z-i/n)$. Let $\Pi = \{[k] : k \leq n\}$. In this case, inspecting the dominating term of the sum for $i \in \pi^*$ we have $\arg\max_{\pi \in \Pi: i \notin \pi} \frac{V_{\pi,\pi^*}}{|\pi \Delta \pi^*|} \frac{1}{\widetilde{\Delta}_\pi^2} = [i-1]$ and takes a value of $\left(\frac{1+\alpha}{h}\right)^2 n^{-1}(z-i/n)^{2\alpha+1}$. Trivially upper bounding the other terms and summing, the sample complexities can be calculated to be within a constant of

$$\text{if } \alpha = 0,\ \log(n)\log(\log(n)/\delta)/h^2 \qquad \text{if } \alpha > 0 \quad n^{2\alpha}\log(\log(n)/\delta)/h^2$$

These rates match the minimax lower bound rates given in [11] up to $\log\log$ factors. Note that unlike the algorithms given there, our algorithm works in the *agnostic* setting, i.e. it is making no assumptions about whether the Bayes classifier is in the class. In the case of non-adaptive sampling, the sum is replaced with the max times n yielding

$$\text{if } \alpha \geq 0\ n^{2\alpha+1}\log(\log(n)/\delta)/h^2$$

which is substantially worse than adaptive sampling.

We are now ready to prove the theorem.

Proof. Let

$$f_t(I_k, y_k) = \begin{cases} y_k \mathbf{1}\{I_k \in [t', t]\} & t \geq t' \\ -y_k \mathbf{1}\{I_k \in [t, t']\} & t \leq t' \end{cases}$$

In particular, $\widehat{\mu}_t - \widehat{\mu}_{t'} = \frac{1}{T}\sum_{k=1}^{T} f_t(I_k, y_k)$. Note that the random variables (y_s, I_s), for $s = 1, \cdots, n$ are by definition i.i.d. drawn from a distribution on $[n] \times \{0, 1\}$. Note

$$\mathbb{E}\left[\frac{1}{T}\sum_{k=1}^{n} f_t(I_k, y_k)\right] = \begin{cases} \frac{1}{n}\sum_{k=t'}^{t} \eta_i & t \geq t' \\ \frac{1}{n}\sum_{k=t}^{t'} -\eta_i & t \leq t' \end{cases}$$

and (assuming that $t \leq t'$, an identical computation applies when $t \geq t'$)

$$\begin{aligned} \text{var}(f_t) &= \text{var}(y_s \mathbf{1}\{I_s \in [t, t']\}) \\ &\leq \mathbb{E}[y_s^2 \mathbf{1}\{I_s \in [t, t']\}] \\ &= \frac{1}{n}\sum_{i=t}^{t'} (\sigma_i^2 + \eta_i^2) \leq \frac{2}{n}|t' - t|. \end{aligned}$$

By Theorem 2.3 in [7], given $\delta > 0$, for each $\{s : s \leq n, |s - t'| \leq \tau\}$ we have that

$$\mathbb{P}\Bigg(\left|\tfrac{1}{T}\sum_{k=1}^{T} f_s(I_k, y_k) - \mathbb{E}[f_s]\right| > 2\mathbb{E}\left[\sup_{|s-t'|\leq\tau}\left|\frac{1}{T}\sum_{k=1}^{T} f_s(I_k, y_k) - \mathbb{E}[f_s]\right|\right] + \sqrt{\frac{2\tau\log(1/\delta)}{nT}} + \tfrac{7\log(1/\delta)}{3T}\Bigg) \leq \delta$$

To obtain a bound over all time, we now face two major tasks. Firstly, we must apply a peeling argument to the set of t's. Secondly, and perhaps more immediate, we need bounds on the empirical process

$$\mathbb{E}\left[\sup_{|s-t'|\leq\tau}\left|\frac{1}{T}\sum_{k=1}^{T} f_s(I_k, y_k) - \mathbb{E}[f_s]\right|\right]$$

Let's start with the latter. Denote $Z_t = \frac{1}{T}\sum_{k=1}^{T} f_t(I_k, y_k) - \mathbb{E}[\frac{1}{T}\sum_{k=1}^{T} f_t(I_k, y_k)]$. Firstly note that,

$$|(f_s - f_t)(I_k, y_k)| = \begin{cases} y_k \mathbf{1}\{I_k \in [t, s]\} & s > t \\ -y_k \mathbf{1}\{I_k \in [s, t]\} & t > s \end{cases}$$

In particular the computation above shows,

$$\text{var}\left((f_s - f_t)(I_k, y_k)\right) \leq 2\frac{|t - s|}{n}.$$

Hence,

$$\begin{aligned} \text{var}\left(\frac{1}{T}\sum_{k=1}^{T} f_t(I_k, y_k) - \mathbb{E}[f_t] - \left(\frac{1}{T}\sum_{k=1}^{T} f_s(I_k, y_k) - \mathbb{E}[f_s]\right)\right) &= \frac{\text{var}(f_t(I_k, y_k) - f_s(I_k, y_k))}{T} \\ &\leq \frac{2|t - s|}{nT} \end{aligned}$$

In particular, since $|\frac{1}{T}f_t(I_s, y_s)| \leq \frac{1}{T}$, Bernstein's inequality implies,

$$\log(\mathbb{E}[e^{\lambda(Z_t - Z_s)})] \leq \frac{\lambda^2 \frac{2|t-s|}{nT}}{2(1 - \lambda/3T)}.$$

Let $d^2(t, s) = |\frac{t}{n} - \frac{s}{n}|$. Then, Lemma 13.1 of [6] with $\nu = 2/T$ and $c = 1/3T$ we have that,

$$\begin{aligned}
\mathbb{E}\left[\sup_{|s-t'|\leq\tau} |Z_s|\right] &\leq \frac{12\sqrt{2}}{\sqrt{T}} \int_0^{\sqrt{\tau/n}/2} \sqrt{\log(\tfrac{\sqrt{\tau/n}}{2u})}du + \frac{4}{T}\int_0^{\sqrt{\tau/n}/2} \log(\tfrac{\sqrt{\tau/n}}{2u})du \\
&\leq \frac{12\sqrt{2}}{\sqrt{T}} \int_0^{\infty} \sqrt{\frac{\tau}{n}} v^2 e^{-v^2} dv + \frac{4}{T}\int_0^{\infty} \frac{1}{2}\sqrt{\frac{\tau}{n}} v e^{-v} dv \\
&\leq \frac{12\sqrt{\pi}}{\sqrt{T}} \sqrt{\frac{\tau}{n}} + \frac{2}{T}\sqrt{\frac{\tau}{n}} \\
&\leq 12\sqrt{\pi}\sqrt{\frac{\tau}{nT}} + \frac{2}{T}
\end{aligned}$$

the third line follows from the second by doing the substitution, $v = \sqrt{\log(\sqrt{\tau/n}/u))}$ and similarly $u = \log(\sqrt{\tau/n}/u))$ on the second integral.

Hence for all $s : |s - t'| \leq \tau$, using the fact that $\sqrt{a} + \sqrt{b} \leq \sqrt{2(a+b)}$

$$\mathbb{P}\left(\left|\tfrac{1}{T}\textstyle\sum_{k=1}^T f_s(I_k, y_k) - \mathbb{E}[f_s]\right| > \sqrt{\tfrac{\tau}{nT}\left(43 + 2\sqrt{2}\log(\tfrac{1}{\delta})\right)} + \tfrac{12+\log(1/\delta)}{3T}\right) \leq \delta$$

At this point we need to apply a peeling argument. Let $S_r = \{s \leq n : 2^{r-1} \leq |s - t'| \leq 2^r\}$. Note that $r \leq \log_2(2|s-t'|+2) \leq \log_2(4|s-t'|)$. For each such s, since $2^r \leq 2|s-t'|$, with probability greater than $1 - \frac{2\delta}{3r^2}$,

$$\begin{aligned}
\left|\tfrac{1}{T}\textstyle\sum_{k=1}^T f_s(I_k, y_k) - \mathbb{E}[f_s]\right| &< \sqrt{\frac{2|s-t'|}{nT}\left(43 + 2\sqrt{2}\log(\tfrac{2r^2}{3\delta})\right)} + \frac{12+\log\left(\frac{2r^2}{3\delta}\right)}{3T} \\
&\leq \sqrt{\frac{2|s-t'|}{nT}\left(43 + 2\sqrt{2}\log(\tfrac{2\log_2^2(4|s-t'|)}{3\delta})\right)} + \frac{12+\log\left(\frac{2\log_2^2(4|s-t'|)}{3\delta}\right)}{3T}
\end{aligned}$$

Now union-bounding over each $r = 1, \cdots, \log_2(n - t')$, we have that

$$\left|\tfrac{1}{T}\textstyle\sum_{k=1}^T f_s(I_k, y_k) - \mathbb{E}[f_s]\right| \leq \sqrt{\tfrac{2|s-t'|}{nT}\left(43 + 2\sqrt{2}\log(\tfrac{2\log_2^2(4|s-t'|)}{3\delta})\right)} + \frac{12+\log\left(\frac{2\log_2^2(4|s-t'|)}{3\delta}\right)}{3T}$$

with probability greater than

$$\sum_{k=1}^{\log_2(n-t')} \frac{2\delta}{3k^2} \leq \sum_{k=1}^{\infty} \frac{2\delta}{3}k^2 \leq \delta$$

□