

---

# Chernoff Sampling for Active Testing and Extension to Active Regression

---

Subhojyoti Mukherjee\*  
UW-Madison

Ardhendu Tripathy\*  
Missouri S&T

Robert Nowak  
UW-Madison

## Abstract

Active learning can reduce the number of samples needed to perform a hypothesis test and to estimate the parameters of a model. In this paper, we revisit the work of Chernoff that described an asymptotically optimal algorithm for performing a hypothesis test. We obtain a novel sample complexity bound for Chernoff’s algorithm, with a non-asymptotic term that characterizes its performance at a fixed confidence level. We also develop an extension of Chernoff sampling that can be used to estimate the parameters of a wide variety of models and we obtain a non-asymptotic bound on the estimation error. We apply our extension of Chernoff sampling to actively learn neural network models and to estimate parameters in real-data linear and non-linear regression problems, where our approach performs favorably to state-of-the-art methods.

## 1 Introduction

In contrast to common machine learning algorithms that use independent and identically distributed (iid) samples for training, active learning promises to use fewer samples by allowing the algorithm to choose the samples it is trained on. While the benefit of active learning has been analyzed extensively for the problem of classification (Dasgupta, 2005; Hanneke, 2007; Dasgupta et al., 2008; Balcan et al., 2009; Balcan and Long, 2013; Zhang and Chaudhuri, 2014; Katz-Samuels et al., 2021), there are fewer works (Cai et al., 2016; Wu, 2018; Wu et al., 2019; Bu et al., 2019) that utilize active learning for regression. In this paper we extend an asymptotically optimal algorithm for active testing,

that was developed by Chernoff (Chernoff, 1959), to active regression. We empirically show that this resulted in more efficient estimation of parameters in regression models. In addition, we obtain non-asymptotic bounds on the sample complexity and estimation error for Chernoff’s algorithm in active testing and its extension in active regression, respectively. Non-asymptotic bounds characterize the performance of an algorithm when executed at a fixed confidence level, which is relevant for real-world applications. While theoretical results for active regression using maximum likelihood estimates were given by Chaudhuri and Mykland (1993) and Chaudhuri et al. (2015), our method is likelihood-free and is applicable to sub-Gaussian observations.

To frame our contributions, let us first establish some basic notation and a problem statement. Consider a sequential learning problem in which the learner may select one of  $n$  possible actions at each step. A sample resulting from action  $i$  is a realization of a sub-Gaussian random variable with mean  $\mu_i(\theta^*)$ , where the mean  $\mu_i(\theta)$  is a known function parameterized by  $\theta \in \Theta$ . The specific  $\theta^* \in \Theta$  that governs the observations is not known. Each action may be performed multiple times, resulting in i.i.d. observations, and observations from different actions are also statistically independent. This paper considers the problem of sequentially and adaptively choosing actions for the following goals:

**In Active Testing:**  $\Theta$  is finite and the goal is to correctly determine the true hypothesis  $\theta^*$ .

**In Active Regression:**  $\Theta$  is a compact (uncountable) space and the goal is to accurately estimate  $\theta^*$ .

Below, we detail our contributions to both problems.

- We revisit Chernoff’s sampling algorithm for the sequential design of experiments (Chernoff, 1959), which is equivalent to the situation where the parameter space  $\Theta$  is a finite set. The algorithm provably minimizes the number of samples used to identify  $\theta^* \in \Theta$  in the asymptotic high-confidence setting. We derive a non-asymptotic sample complexity bound for the algorithm in Chernoff (1959)

---

\* Equal contribution. Proceedings of the 25<sup>th</sup> International Conference on Artificial Intelligence and Statistics (AISTATS) 2022, Valencia, Spain. PMLR: Volume 151. Copyright 2022 by the author(s).

that characterizes its performance in low/medium confidence regimes. We also provide theoretical guarantees for three variations of the Chernoff sampling algorithm. We prove a minimax lower bound that shows that the algorithm in Chernoff (1959) can be optimal in the medium confidence regime. We also generalize it to handle sub-Gaussian distributions. Consequently, we replace the maximum likelihood criterion (which depends on the probability distribution) with the minimum sum-of-squared errors criterion (which depends only on the mean functions).

- We extend the algorithm in Chernoff (1959) to handle smoothly parameterized functions  $\mu_i(\boldsymbol{\theta})$  where  $\Theta \subseteq \mathbb{R}^d$ . A brute-force approach could involve using a finite, discrete covering of  $\Theta$ , but this is impractical. Instead, we prove that an optimal sampling distribution (according to Chernoff’s criterion) is generally sparse and may be obtained by solving a simple eigenvalue optimization problem closely related to the notion of E-optimality in experimental design (Dette and Studden, 1993). We provide a convergence guarantee for the smoothly parameterized setting that utilizes a new error metric. We demonstrate that the extension of Chernoff (1959) outperforms existing stage-based algorithms in benchmark real-life datasets and in a neural network experiment.

We derive our non-asymptotic sample complexity bound for Chernoff (1959) using the techniques in Naghshvar and Javidi (2013). The convergence proof for active regression extends the techniques of Frostig et al. (2015) and applies it to our extension of Chernoff (1959).

### 1.1 Related Work

The algorithm in Chernoff (1959) assumes that the probability distribution of an observation from any action  $i \in [n]$  under any hypothesis  $\boldsymbol{\theta} \in \Theta$  is known to the learner. Consider a partition of  $\Theta = \Theta_1 \cup \Theta_2$  and a hypothesis test between  $\boldsymbol{\theta}^* \in \Theta_1$  and  $\boldsymbol{\theta}^* \in \Theta_2$ . The objective is to choose actions such that the hypothesis test can be performed using as few samples as possible. Using past observations, a maximum likelihood estimate  $\hat{\boldsymbol{\theta}}$  is found and let  $\hat{\boldsymbol{\theta}} \in \Theta_1$ . The algorithm in Chernoff (1959) chooses the next action according to a probability mass function (pmf)  $\mathbf{p}$  over actions obtained by

$$\arg \max_{\mathbf{p}} \inf_{\boldsymbol{\theta}^* \in \Theta_2} \sum_{i=1}^n p(i) \text{KL}(\nu_i(x; \hat{\boldsymbol{\theta}}) \| \nu_i(x; \boldsymbol{\theta}^*)), \quad (1)$$

where  $\mathbf{p} = (p(1), \dots, p(n))$ ,  $\nu_i(\cdot; \boldsymbol{\theta})$  denotes the probability distribution of an observation from action  $i$  if  $\boldsymbol{\theta}$

were the true hypothesis, and KL denotes the Kullback-Leibler divergence. The optimization (1) is similar to those appearing in sample complexity lower bounds for best-arm identification in multi-armed bandits Garivier and Kaufmann (2016); Combes et al. (2017); Degenne et al. (2020). In those works the inf is taken over all  $\boldsymbol{\theta}'$  having an optimal action that is different from that under the true  $\boldsymbol{\theta}^*$ .

While the algorithm in Chernoff (1959) is asymptotically optimal under certain assumptions, subsequent works Blot and Meeter (1973), Naghshvar and Javidi (2013), Nitinawarat et al. (2013) have proposed modifications that work well outside the asymptotic limit, strengthen theoretical guarantees, and reduce the number of assumptions needed. Naghshvar and Javidi (2013) proposed a two-phase Bayesian policy **TP** which conducts forced exploration in the first phase and computes a posterior belief over the hypotheses. Then in the second phase, it switches to the optimal Chernoff sampling proportion in eq. (1) if the probability of one hypothesis crosses a threshold. **TP** can relax an assumption made in Chernoff (1959) which stated that sampling any action always provides some information about the true  $\boldsymbol{\theta}^*$ . If that assumption is true, then Chernoff Sampling (**CS**), which has no such separation of phases, empirically outperforms **TP**. It also enjoys both moderate and optimal asymptotic guarantees. Nitinawarat et al. (2013) have modified Chernoff (1959) by adding a small amount of uniform exploration to relax the previous assumption. In a different problem Vaidhiyan and Sundaresan (2017) have modified **CS** to quickly identify an odd Poisson point process having a different rate of arrival than others.

For estimating parameters of a regression model, efficient methods for selecting actions have been studied in the area of Optimum Experiment Design (Silvey, 1980; Pukelsheim, 2006; Pronzato and Pázman, 2013). However a major focus in these works has been on large-sample asymptotic properties of estimators obtained from a fixed sampling distribution. While adaptive sampling proportions have also been proposed (e.g. Section 8.5 in Pronzato and Pázman (2013)), there have been fewer works characterizing their theoretical properties. Most theoretical work on active learning has focused on learning binary classifiers that belong to a particular hypothesis class (Dasgupta, 2005; Hanneke, 2007; Dasgupta et al., 2008; Balcan et al., 2009; Balcan and Long, 2013; Zhang and Chaudhuri, 2014; Katz-Samuels et al., 2021). The works of Chaudhuri and Mykland (1993) and Chaudhuri et al. (2015) propose adaptive sampling methods for obtaining maximum likelihood estimates of the parameters in a regression model. Chaudhuri et al. (2015) propose a two-stage algorithm **ActiveS** that first samples uniformly at ran-

dom to obtain a preliminary estimate of the parameters, which is then used to find a sampling proportion for the second stage by solving an optimization problem. In contrast **CS** is a fully adaptive algorithm and enjoys a similar convergence under slightly stronger assumptions.

Sabato and Munos (2014) provides an active learning algorithm for linear regression problems under model mismatch. The same setting under heteroscedastic noise has been studied by Chaudhuri et al. (2017) where they propose a two-stage process adapted to the noise. Fontaine et al. (2019) also studies the linear regression setting under heteroscedastic noise but proposes a fully adaptive algorithm that is similar to A-optimal design. Bu et al. (2019) studies a different setting where  $\theta^*$  is changing with time. They modify the algorithm of Chaudhuri et al. (2015) to fully adaptive process where the optimization needs to be solved at every round. Wu (2018) studies the linear regression setting where the goal is to maximize the diversity of the samples. Cai et al. (2016) studies both the linear and non-linear regression setting and proposes the heuristic **EMCM** without any convergence guarantee. Similarly, Wu et al. (2019) also studies the active regression for noiseless setting but they provide no convergence proof. As opposed to these works **CS** has a convergence guarantee and performs well in real-world benchmark problems.

Another line of work is the non-parametric setup of Castro et al. (2005), where the objective is to estimate an unknown function over its entire domain. Here the error rates for learning are  $O(t^{-\gamma})$ , and the exponent  $\gamma$  decreases as the complexity of the hypothesis class of functions increases (i.e., there is a slower decrease in error when learning a more complicated function). For example, if the hypothesis class consists of Holder smooth functions defined on domain  $[0, 1]^d$  then  $\gamma = 1/(d - 1 + 1/d)$ . In contrast, our work is in the parametric setting, where we only want to estimate a single parameter  $\theta^*$  and  $\gamma = 1$ . We show that the **CS** algorithm has a smaller problem-dependent constant in the error bound. The work of Goetz et al. (2018) is also in a similar framework as that of Castro et al. (2005). Other forms of optimal experiment design have been explored in the context of active learning by Yu et al. (2006), and in different bandit problems by Soare et al. (2014); Fiez et al. (2019); Degenne et al. (2020). Note that our objective of identifying  $\theta^*$  is a strictly more difficult objective than best-arm identification in bandit problems.

## 2 Active Testing

A sequential policy  $\pi$  tasked to find  $\theta^*$  interacts with the environment in an iterative fashion. At time  $t$ , the policy samples action  $I_t$  and receives a random observation  $Y_t$  that follows the distribution  $\nu_{I_t}(\cdot; \theta^*)$ , where  $\theta^*$  is the true value of the unknown parameter that belongs to a set  $\Theta$ . In active testing,  $\Theta$  contains  $J$  discrete hypotheses. Let  $\mathcal{F}_t := \sigma(I_1, Y_1, I_2, Y_2, \dots, I_t, Y_t)$  denote the sigma-algebra generated by the sequence of actions and observations till time  $t$ . Then  $\pi$  is said to be  $\delta$ -PAC if: (1) at each  $t$  the sampling rule  $I_t$  is  $\mathcal{F}_{t-1}$  measurable, (2) it has a finite stopping time  $\tau_\delta$  with respect to  $\mathcal{F}_t$ , and (3) its final prediction  $\hat{\theta}(\tau_\delta)$  is based on  $\mathcal{F}_{\tau_\delta}$  and satisfies  $\mathbb{P}(\hat{\theta}(\tau_\delta) \neq \theta^*) \leq \delta$ . A table of notation is provided in Appendix A.10. Based on the observations  $(Y_1, Y_2, \dots, Y_t)$ , we define for every  $\theta \in \Theta$  the sum of squared errors and the difference between the sum of squared errors for  $\theta$  and  $\theta^*$  as follows:

$$L_t(\theta) := \sum_{s=1}^t (Y_s - \mu_{I_s}(\theta))^2, \quad (2)$$

$$\Delta_t(\theta) := L_t(\theta) - L_t(\theta^*). \quad (3)$$

**Assumption 1.** *An observation from any action under any hypothesis has bounded range, i.e.,  $Y_s \in [-\sqrt{\eta}/2, \sqrt{\eta}/2]$  almost surely at every round  $s$  for some fixed  $\eta > 0$ .*

Suppose  $\nu_i(Y; \theta)$  are Gaussian distributions with mean  $\mu_i(\theta)$  and variance  $1/2$ . Let  $\hat{\theta}(t)$  denote the estimate for  $\theta^*$  at time  $t$ . Using  $\Theta_1 = \{\hat{\theta}(t)\}$ ,  $\Theta_2 = \Theta \setminus \{\hat{\theta}(t)\}$  and expressions for KL divergence of Gaussian distributions in eq. (1), we obtain that **CS**, which is asymptotically optimal, samples the next action according to a probability mass function (pmf) that is a solution to the following max min optimization:

$$\mathbf{p}_{\hat{\theta}(t)} := \arg \max_{\mathbf{p}} \min_{\theta' \neq \hat{\theta}(t)} \sum_{i=1}^n p(i) (\mu_i(\theta') - \mu_i(\hat{\theta}(t)))^2. \quad (4)$$

We can solve (4) by formulating it as a linear program:

$$\max_{\mathbf{p}} z \text{ s.t. } \sum_{i=1}^n p(i) (\mu_i(\theta') - \mu_i(\hat{\theta}(t)))^2 \geq z \quad \forall \theta' \neq \hat{\theta}(t), \quad (5)$$

where the optimization variables are the scalar  $z$  and pmf  $\mathbf{p}$  satisfying the constraints  $p(i) \geq 0 \forall i$  and  $\sum_{i=1}^n p(i) = 1$ .

**Chernoff Sampling (CS):** Inspired by the sampling proportion in eq. (4), we use the same sampling strategy even though the distributions  $\{\nu_i(Y; \theta)\}_{i=1}^n$  are only assumed to be sub-Gaussian (Algorithm 1). Our

estimate of the most likely hypothesis (breaking ties at random) given the data is  $\hat{\theta}(t) := \arg \min_{\theta \in \Theta} L_t(\theta)$ . The action sampled at the next time  $t + 1$  is chosen by the randomized rule  $\mathbb{P}(I_{t+1} = i) = p_{\hat{\theta}(t)}(i), \forall i \in [n]$ . We stop sampling at  $\tau_\delta$  if the sum of squared errors for all competing hypothesis is greater than that of  $\hat{\theta}(\tau_\delta)$  by a threshold  $\beta(J, \delta)$  to be defined later.

---

**Algorithm 1** Chernoff Sampling for Active Testing
 

---

- 1: **Input:** Confidence parameter  $\delta$ , threshold  $\beta(J, \delta)$
  - 2: Sample  $I_1 \in [n]$  randomly, observe  $Y_1$  and find  $\hat{\theta}(1)$ .
  - 3: **for**  $t = 2, 3, \dots$  **do**
  - 4:   Sample  $I_t \sim \mathbf{p}_{\hat{\theta}(t-1)}$  from (4) and observe  $Y_t$ .
  - 5:   Calculate  $L_t(\theta)$  from (2)  $\forall \theta \in \Theta$ , find  $\hat{\theta}(t)$ .
  - 6:   **if**  $L_t(\theta^*) - L_t(\hat{\theta}(t)) > \beta(J, \delta) \forall \theta' \neq \hat{\theta}(t)$  **then**
  - 7:       **Return**  $\hat{\theta}(t)$  as the true hypothesis.
- 

In Chernoff (1959) they provide a sample complexity upper bound only for the asymptotic regime when  $\delta \rightarrow 0$ . We give a non-asymptotic fixed confidence sample complexity upper bound in Theorem 1. The following assumption, originally made by Chernoff (1959), is used to prove Theorem 1.

**Assumption 2.** *The mean of the observation from any action under  $\theta^*$  is different from its mean under any other hypothesis, i.e.,  $\min_{i \in [n]} \min_{\theta \neq \theta^*} |\mu_i(\theta) - \mu_i(\theta^*)| > 0$ .*

In subsequent works by Nitinawarat et al. (2013) and Naghshvar and Javidi (2013), it was shown that the above assumption can be relaxed if the algorithm is modified. We make the assumption since we give a non-asymptotic sample complexity bound for the original algorithm.

**Definition 1.** *Define the smallest squared difference of means between any two actions under any pair of hypotheses  $\theta, \theta'$  as  $\eta_0 := \min_{i \in [n]} \min_{\theta \neq \theta'} (\mu_i(\theta) - \mu_i(\theta'))^2$ . By Assumption 2 we have  $\eta_0 > 0$ .*

We define the threshold  $\beta(J, \delta) = \log(CJ/\delta)$  where  $C$  is a constant depending on  $\eta$ , and  $\eta_0$ . The values  $\eta$  (or an upper bound to it) and  $\eta_0$  are known to the learner.

**Theorem 1. (CS Sample Complexity)** *Let  $\tau_\delta$  denote the stopping time of CS in Algorithm 1. Let  $D_0$  be the objective value of the max min optimization in (4) when  $\theta = \theta^*$ , i.e.,*

$$D_0 := \max_{\mathbf{p}} \min_{\theta' \neq \theta^*} \sum_{i=1}^n p(i) (\mu_i(\theta') - \mu_i(\theta^*))^2.$$

Denote  $\mathbf{p}_\theta$  as the solution of (4) when  $\hat{\theta}(t)$  is replaced by any  $\theta \in \Theta$ , and  $D_1$  is the minimum possible objective

value over all  $\mathbf{p}_\theta$  when  $\hat{\theta}(t)$  is replaced by  $\theta^*$ , i.e.,

$$D_1 := \min_{\{\mathbf{p}_\theta: \theta \in \Theta\}} \min_{\theta' \neq \theta^*} \sum_{i=1}^n p_\theta(i) (\mu_i(\theta') - \mu_i(\theta^*))^2.$$

Assumption 2 ensures that  $D_1 > 0$ . The sample complexity of the  $\delta$ -PAC CS has the following upper bound, where  $J := |\Theta|$ ,  $C = O((\eta/\eta_0)^2)$  is a constant:

$$\mathbb{E}[\tau_\delta] \leq O\left(\frac{\eta \log(C) \log J}{D_1} + \frac{\log(J/\delta)}{D_0} + JC^{\frac{1}{\eta}} \delta^{\frac{D_0}{\eta^2}}\right).$$

*Proof. (sketch)* Algorithm 1 stops at  $\tau_\delta$  when the error for the returned hypothesis is smaller than the error for all the other hypotheses by an amount of  $\beta(J, \delta)$ . To obtain an upper bound to  $\mathbb{E}[\tau_\delta]$ , we instead look at a different random time  $\tau_{\theta^*} := \min\{t : \Delta_t(\theta) > \beta(J, \delta), \forall \theta \neq \theta^*\}$ , which is the first time when the error for the true hypothesis  $\theta^*$  is smaller than the error for all other hypotheses by  $\beta(J, \delta)$ . Either the hypothesis returned by the Algorithm 1 is  $\hat{\theta}(\tau_\delta) = \theta^*$ , in which case  $\tau_\delta = \tau_{\theta^*}$ , or  $\tau_{\theta^*}$  has not occurred yet and  $\tau_\delta < \tau_{\theta^*}$ . Hence we focus on bounding  $\mathbb{E}[\tau_{\theta^*}]$ . The key random quantity in the definition of  $\tau_{\theta^*}$  is  $\Delta_t(\theta)$ , and using Assumption 1 we can show that  $\Delta_t(\theta)$  concentrates to its expected value. The expected value  $\mathbb{E}[\Delta_t(\theta)]$  is increasing with  $t$  for each  $\theta \neq \theta^*$  and for large enough  $t$  it will be greater than  $\beta(J, \delta)$ . Since  $\Delta_t(\theta)$  concentrates to  $\mathbb{E}[\Delta_t(\theta)]$ , for large enough  $t$ ,  $\Delta_t(\theta)$  will also be greater than  $\beta(J, \delta)$  and  $\tau_{\theta^*}$  would occur. To quantify when  $\tau_{\theta^*}$  occurs, we lower bound  $\mathbb{E}[\Delta_t(\theta)]$  as follows:

$$\mathbb{E}[\Delta_t(\theta)] \geq \mathbb{E}[\tilde{\tau}_{\theta^*} D_1 + (t - \tilde{\tau}_{\theta^*}) D_0], \quad (6)$$

where  $\tilde{\tau}_{\theta^*}$  is the last time after which the error for the true hypothesis  $\theta^*$  is always smaller than the errors for all other hypotheses. Till the time  $\tilde{\tau}_{\theta^*}$ , the CS sampling proportion  $\mathbf{p}_{\hat{\theta}(t)}$  may not be  $\mathbf{p}_{\theta^*}$ , and  $\mathbb{E}[\Delta_t(\theta)]$  grows at the slower ‘‘exploration’’ rate  $D_1$  defined in the Theorem. After  $\tilde{\tau}_{\theta^*}$  the CS proportion is  $\mathbf{p}_{\theta^*}$ , and  $\mathbb{E}[\Delta_t(\theta)]$  increases at the optimal ‘‘verification’’ rate  $D_0$ . We finally bound the sample complexity by using  $\mathbb{E}[\tau_{\theta^*}] = \sum_t \mathbb{P}(\tau_{\theta^*} = t) \leq M + \mathbb{P}(\tau_{\theta^*} > M) \cap \tilde{\tau}_{\theta^*} \leq M + \mathbb{P}(\tau_{\theta^*} > M \cap \tilde{\tau}_{\theta^*} > M)$  where,  $M = \frac{\eta \log(C) \log J}{D_1} + \frac{\log(J/\delta)}{D_0}$ . The two tail events above is shown to be bounded by  $O(JC^{1/\eta} \delta^{D_0/\eta^2})$  in Lemma 5. The full proof is in Appendix A.2.5.  $\square$

In the result of Theorem 1 the first term  $\eta \log(C) \log(J)/D_1$  bounds the number of samples taken during the exploration phase when  $\hat{\theta}(t) \neq \theta^*$ . This is the non-asymptotic term that is not present in the analysis of Chernoff (1959). The second term  $\log(J/\delta)/D_0$  is the dominating term when  $\delta \rightarrow 0$ , and

it matches the asymptotic sample complexity expression of Chernoff (1959). Naghshvar and Javidi (2013) also derive a moderate confidence bound for their policy called **TP** but suffer from a worse non-asymptotic term  $\log(J/\delta)/D_{\text{NJ}}$  where  $D_{\text{NJ}} < D_1$  ( $D_{\text{NJ}}$  is denoted as  $I_1(M)$  in Naghshvar and Javidi (2013)). **TP** do not require the assumption that  $D_1 > 0$  as it conducts forced exploration in the first stage. The **TP** policy is asymptotically optimal but performs poorly in some instances (see Example 1 and Section 4) due to the fixed exploration in the first stage. We discuss further results in Appendix A.1.1.

The following example shows that the non-asymptotic term of **CS** may dominate.

**Example 1. (Non-asymptotes matter)** Consider an environment with two actions and  $\Theta = \{\theta^*, \theta', \theta''\}$ . The following table describes the values of  $\mu_1(\cdot), \mu_2(\cdot)$  under these three hypotheses.

$\theta$	$=$	$\theta^*$	$\theta'$	$\theta''$
$\mu_1(\theta)$	$=$	1	0.001	0
$\mu_2(\theta)$	$=$	1	1.002	0.998

For a choice of  $\delta = 0.1$ , we can evaluate that  $\log(J)/D_1 \approx 3 \times 10^5$  and  $\log(J/\delta)/D_0 \approx 3.4$ . While Theorem 1 is only an upper bound, empirically we do see that the non-asymptotic term dominates the sample complexity. The Figure 1a shows a box plot of the stopping times of four algorithms over 100 independent trials on the above environment. A box plot depicts a set of numerical data using their quartiles (Tukey, 1977). A uniform sampling baseline (**Unif**) performs much better than **CS**, as it samples action 1 half the time in expectation, and action 1 is the best choice to distinguish  $\theta^*$  from both  $\theta'$  and  $\theta''$ . **TP** also performs poorly compared to **Unif** and **CS** in this setting. The non-asymptotic term of **TP** scales as  $\log(J)/D_{\text{NJ}} \approx 4 \times 10^6$ .

**T2 Sampling:** Instead of sampling according to the optimal verification proportion in line 4 of Algorithm 1, we can use the following heuristic argument. At each time, consider the current most likely  $\hat{\theta}(t)$  and its “closest” competing hypothesis defined as  $\tilde{\theta}(t) := \arg \min_{\theta \neq \hat{\theta}(t)} L_t(\theta)$ . The  $\delta$ -PAC heuristic called Top-2 sampling (abbreviated as **T2**) samples an action that best discriminates between them, i.e.,

$$I_{t+1} := \arg \max_{i \in [n]} (\mu_i(\hat{\theta}(t)) - \mu_i(\tilde{\theta}(t)))^2. \quad (7)$$

This strategy requires lesser computation as we don’t need to compute  $p_{\hat{\theta}(t)}$  which could be useful when  $J$  or  $n$  is very large. It was proposed by Chernoff (1959) without any sample complexity proof.

**Proposition 1. (T2 Sample Complexity)** Let  $\tau_\delta$  denote the stopping time of **T2** following the sampling

strategy of (7). Consider the set  $\mathcal{I}(\theta, \theta') \subset [n]$  of actions that could be sampled following (7) when  $\hat{\theta}(t) = \theta$  and  $\tilde{\theta}(t) = \theta'$ , and let  $u_{\theta\theta'}$  denote a uniform pmf supported on  $\mathcal{I}(\theta, \theta')$ . Define

$$D'_0 := \min_{\theta, \theta' \neq \theta^*} \sum_{i=1}^n u_{\theta\theta'}(i) (\mu_i(\theta') - \mu_i(\theta^*))^2,$$

$$D'_1 := \min_{\theta \neq \theta', \theta' \neq \theta^*} \sum_{i=1}^n u_{\theta\theta'}(i) (\mu_i(\theta') - \mu_i(\theta^*))^2,$$

where we assume that  $D'_1 > 0$ . Then for a constant  $C > 0$  the sample complexity of **T2** has the following upper bound:

$$\mathbb{E}[\tau_\delta] \leq O\left(\frac{\eta \log(C) \log J}{D'_1} + \frac{\log(J/\delta)}{D'_0} + JC^{\frac{1}{\eta}} \delta^{\frac{D'_0}{\eta^2}}\right).$$

The bound above has a similar form as in Theorem 1 with three terms. The first term does not scale with error probability  $\delta$ , while the second term scales with  $\log(J/\delta)$ . The denominators of the two terms are different from  $D_0$  and  $D_1$  due to the different sampling rule of **T2** and hence **T2** is not asymptotically optimal.

**Batch Updates:** We can solve max min optimization for  $p_{\hat{\theta}}(t)$  every  $B$  rounds instead of at each round. This reduces computation while increasing the sample complexity by only an additive term as shown below.

**Proposition 2. (Batch-CS Sample Complexity)** Let  $\tau_\delta, D_0, D_1$  be defined as in Theorem 1 and  $B$  be the batch size. Then the sample complexity of  $\delta$ -PAC **Batch-CS** is

$$\mathbb{E}[\tau_\delta] \leq O\left(B + \frac{\eta \log(C) \log J}{D_1} + \frac{\log(J/\delta)}{D_0} + BJC^{\frac{1}{\eta}} \delta^{\frac{D_0}{\eta^2}}\right).$$

**CS with Exploration:** Recall that  $D_1 > 0$  (Assumption 2) is required to prove Theorem 1. We now relax this assumption with the policy **CSE** which follows the proportion  $p_{\hat{\theta}(t)}$  with probability  $1 - \epsilon_t$  and uniformly explores any other action  $i \in [n]$  with probability  $\epsilon_t$ . The exploration parameter  $\epsilon_t$  is chosen to reduce with time. Define  $D_e := \min_{\theta' \neq \theta^*} \sum_{i=1}^n \frac{1}{n} (\mu_i(\theta') - \mu_i(\theta^*))^2$  as the objective value for uniform sampling, then  $D_e > 0$ .

**Proposition 3. (CSE Sample Complexity)** Let  $\tau_\delta, D_0, C$  be defined as in Theorem 1,  $D_e$  be defined as above, and  $\epsilon_t := 1/\sqrt{t}$ . Then the sample complexity bound of  $\delta$ -PAC **CSE** with  $\epsilon_t$  exploration is given by

$$E[\tau_\delta] \leq O\left(\frac{\eta \log(C) \log J}{D_e} + \frac{\log(J/\delta)}{D_0} + JC^{\frac{1}{\eta}} \delta^{\frac{D_0}{\eta^2}}\right).$$

We can see that the non-asymptotic term does not depend on  $D_1$  and scales with  $D_e$ . Note that  $D_0 >$

$D_1$  and  $D_0 > D_e$  separately but  $D_e$  and  $D_1$  are not comparable because  $D_1$  is defined as the minimum over verification proportions for all the hypotheses while  $D_e$  is defined using a uniform sampling proportion. When  $\delta \rightarrow 0$  then the asymptotic term dominates and so **CSE** is asymptotically optimal. In Example 1 we can calculate that  $\log(J)/D_e \approx 2.1$  and  $\log(J/\delta)/D_0 = 3.4$ . So **CSE** performs similar to **Unif** (see Figure 1a) as well as theoretically enjoy asymptotic and moderate confidence guarantee similar to **CS**.

We now provide a brief proof sketch of the three propositions stated before. These proofs follows the technique of Theorem 1 with some key changes which we state now. For Proposition 1 observe that **T2** does not sample by  $\mathbf{p}_\theta$  but by the pmf  $\mathbf{u}_{\theta\theta'}$  defined in Proposition 1. This enables us to calculate  $\mathbb{E}[\Delta_t(\theta)] \geq \mathbb{E}[\tilde{\tau}_{\theta^*} D_1' + (t - \tilde{\tau}_{\theta^*}) D_0']$ , where  $\tilde{\tau}_{\theta^*}$  is defined in Theorem 1. After this we can follow a similar line of reasoning as Theorem 1 and bound **T2** sample complexity. The proof is in Appendix A.3.2. For Proposition 2 the key difference with Theorem 1 is that we calculate the  $\mathbf{p}_\theta$  after each batch of size  $B$ . We bound the number of total number of batches  $m_\delta$  instead of  $\tau_\delta$ . As the stopping condition is only checked at the end of every batch we divide the time  $\tau_\delta$  into batches of size  $B$  and use a similar argument as in Theorem 1 to bound  $m_\delta$ . The proof is in Appendix A.4. Finally, for Proposition 3 the key difference is the new exploration term  $D_e$ . By setting  $\epsilon_s := 1/\sqrt{s}$  we obtain  $\mathbb{E}[\Delta_t(\theta)] \geq \mathbb{E}[\tilde{\tau}_{\theta^*} D_e + (t - \tilde{\tau}_{\theta^*}) D_0]$ . Then following the same argument as in Theorem 1 we obtain the upper bound to  $\mathbb{E}[\tau_\delta]$ . The proof is given in Appendix A.5.

**Minimax lower bound:** While Chernoff (1959) had shown the policy to be optimal as  $\delta \rightarrow 0$ , we demonstrate an environment where **CS** has optimal sample complexity for any fixed value of  $\delta$ . Let  $\Gamma = \sqrt{\eta}/2$ . The following table depicts the values for  $\mu_1(\cdot), \mu_2(\cdot), \dots, \mu_n(\cdot)$  under  $J$  different hypotheses:

$$\begin{array}{rcccccc}
 \theta & = & \theta^* & \theta_2 & \theta_3 & \dots & \theta_J \\
 \hline
 \mu_1(\theta) & = & \Gamma & \Gamma - \frac{\Gamma}{J} & \Gamma - \frac{2\Gamma}{J} & \dots & \Gamma - \frac{(J-1)\Gamma}{J} \\
 \mu_2(\theta) & = & \iota_{21} & \iota_{22} & \iota_{23} & \dots & \iota_{2J} \\
 \vdots & & & & & & \\
 \mu_n(\theta) & = & \iota_{n1} & \iota_{n2} & \iota_{n3} & \dots & \iota_{nJ}
 \end{array} \quad (8)$$

Each  $\iota_{ij}$  is distinct and satisfies  $\iota_{ij} < \Gamma/4J$ .  $\mu_1(\cdot)$  is such that the difference of means across any pair of hypotheses is at least  $\Gamma/J$ . Theorem 2 is proved in Appendix A.6 by a change of measure argument. Note that action 1 is better than all others in discriminating between any pair of hypotheses, and any policy to identify  $\theta^*$  cannot do better than allocating all its samples to action 1.

**Theorem 2. (Lower Bound)** *Any  $\delta$ -PAC policy  $\pi$  that identifies  $\theta^*$  in (8) satisfies  $\mathbb{E}[\tau_\delta] \geq$*

$\Omega(J^2\Gamma^{-2}\log(1/\delta))$ . *Applying Theorem 1 to the same environment, the sample complexity of **CS** is  $O(J^2\Gamma^{-2}\log(J/\delta))$  which matches the lower bound upto log factors.*

### 3 Active Regression

In this section, we extend the Chernoff sampling policy to smoothly parameterized hypothesis spaces, such as  $\Theta \subseteq \mathbb{R}^d$ . The original sampling rule in (4) asks to solve a max min optimization, where the min is over all possible choices of the parameter that are not equal to the parameter  $\theta$  being verified. An extension of the rule for when  $\theta^*$  can take infinitely many values was first given by Albert (1961). In it, they want to identify which of two partitions  $\Theta_1 \cup \Theta_2 = \Theta$  does the true  $\theta^*$  belong to. For any given  $\theta_1 \in \Theta_1$ , their verification sampling rule (specialized to the case of Gaussian noise) is

$$\mathbf{p}_{\theta_1} = \arg \max_{\mathbf{p}} \inf_{\theta_2 \in \Theta_2} \sum_{i=1}^n p(i) (\mu_i(\theta_1) - \mu_i(\theta_2))^2. \quad (9)$$

Recall that  $\hat{\theta}(t) := \arg \min_{\theta \in \Theta} L_t(\theta)$ . Suppose we want to find the optimal verification proportion for testing  $\hat{\theta}(t)$ , the current best estimate of  $\theta^*$ . Let  $\mathcal{B}_r^c(\hat{\theta}(t)) := \{\theta \in \mathbb{R}^d : \|\theta - \hat{\theta}(t)\| > r\}$  denote the complement of a ball of radius  $r > 0$  centered at  $\hat{\theta}(t)$ . Instantiate (9) with  $\theta_1 = \hat{\theta}(t)$ ,  $\Theta_1 = \Theta \setminus \mathcal{B}_r^c(\hat{\theta}(t))$  and  $\Theta_2 = \mathcal{B}_r^c(\hat{\theta}(t))$ . Denote the solution of this optimization as  $\mathbf{p}_{\hat{\theta}(t), r}$  and let  $\mathbf{p}_{\hat{\theta}(t)} := \lim_{r \rightarrow 0} \mathbf{p}_{\hat{\theta}(t), r}$ . In case of multiple solutions, we let  $\mathbf{p}_{\hat{\theta}(t), r}$  denote the set of all possible maxima, and the limit is defined to be the limit of a sequence of sets. We show in Theorem 3 that  $\mathbf{p}_{\hat{\theta}(t)}$  (or an element from it) can be computed efficiently. For any  $i \in [n]$  the gradient of  $\mu_i(\cdot)$  evaluated at  $\theta$  is a column vector denoted as  $\nabla \mu_i(\theta)$ .

**Theorem 3.** *Assume that  $\mu_i(\theta)$  for all  $i \in [n]$  is a differentiable function, and the set  $\{\nabla \mu_i(\hat{\theta}(t)) : i \in [n]\}$  of gradients evaluated at  $\hat{\theta}(t)$  span  $\mathbb{R}^d$ . Consider a p.m.f.  $\mathbf{p}_{\hat{\theta}(t), r}$  from (9) for verifying  $\hat{\theta}(t)$  against all alternatives in  $\mathcal{B}_r^c(\hat{\theta}(t))$ . The limiting value of  $\mathbf{p}_{\hat{\theta}(t), r}$  as  $r \rightarrow 0$  is*

$$\mathbf{p}_{\hat{\theta}(t)} := \arg \max_{\mathbf{p}} \lambda_{\min} \left( \sum_{i=1}^n p(i) \nabla \mu_i(\hat{\theta}(t)) \nabla \mu_i(\hat{\theta}(t))^T \right).$$

*Proof. (sketch)* Define  $g_i(\theta) := (\mu_i(\theta) - \mu_i(\hat{\theta}(t)))^2$  for any  $\theta$ . Introducing a probability density function  $\mathbf{q}$  over  $\Theta$ , we can rewrite the optimization for  $\mathbf{p}_{\hat{\theta}(t), r}$  from (9) as

$$\max_{\mathbf{P}} \inf_{\mathbf{q}: \mathbf{q}(\theta)=0 \forall \theta \in \mathcal{B}_r(\hat{\theta}(t))} \int_{\Theta} \mathbf{q}(\theta) \sum_{i=1}^n p(i) g_i(\theta) d\theta. \quad (10)$$

We consider a family  $\mathcal{Q}_r$  of pdfs supported on the boundary of  $\mathcal{B}_r(\hat{\boldsymbol{\theta}}(t))$ . We show that the value of (10) in the limit as  $r \rightarrow 0$  is equal to

$$\lim_{r \rightarrow 0} \max_{\mathbf{p}} \inf_{\mathbf{q} \in \mathcal{Q}_r} \int_{\Theta} q(\boldsymbol{\theta}) \sum_{i=1}^n p(i) g_i(\boldsymbol{\theta}) d\boldsymbol{\theta}. \quad (11)$$

We use the Taylor series expansion for  $g_i(\boldsymbol{\theta})$  around  $\hat{\boldsymbol{\theta}}(t)$  in (11). Then  $\nabla g_i(\hat{\boldsymbol{\theta}}(t)) = 0$  and the second-order term in the Taylor series is  $0.5(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(t))^T \nabla^2 g_i(\hat{\boldsymbol{\theta}}(t)) (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(t)) = (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(t))^T \nabla \mu_i(\hat{\boldsymbol{\theta}}(t)) \nabla \mu_i(\hat{\boldsymbol{\theta}}(t))^T (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(t))$ . Using this in (11) along with the variational characterization of the minimum eigenvalue gives us the result. The full proof is given in Appendix A.7.  $\square$

To illustrate the use of Theorem 3, consider the problem of active learning in a hypothesis space of parametric functions  $\{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ . The target function is  $f_{\boldsymbol{\theta}^*}$  for an unknown  $\boldsymbol{\theta}^* \in \Theta$ . Assume that the learner may query the value of  $f_{\boldsymbol{\theta}^*}$  at points  $\mathbf{x}_1, \dots, \mathbf{x}_n$  in its domain. If point  $\mathbf{x}_i$  is queried, then the learner observes the value  $f_{\boldsymbol{\theta}^*}(\mathbf{x}_i)$  plus a realization of a zero-mean sub-Gaussian noise. This coincides with the setting above by setting  $\mu_i(\boldsymbol{\theta}) := f_{\boldsymbol{\theta}}(\mathbf{x}_i)$ .

---

#### Algorithm 2 Chernoff Sampling for Active Regression

---

- 1: **Input:** Parametric model  $\{\mu_i(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta, i \in [n]\}$ .
  - 2: Sample  $I_1 \in [n]$  randomly, observe  $Y_1$  and find  $\hat{\boldsymbol{\theta}}(1)$ .
  - 3: **for**  $t = 2, 3, \dots$  **do**
  - 4:     Sample  $I_t \sim \mathbf{p}_{\hat{\boldsymbol{\theta}}(t-1)}$  (Theorem 3), observe  $Y_t$ .
  - 5:     Compute  $\hat{\boldsymbol{\theta}}(t) = \arg \min_{\boldsymbol{\theta} \in \Theta} L_t(\boldsymbol{\theta})$ .
- 

Step 4 of Algorithm 2 requires solving the convex eigenvalue optimization in Theorem 3, which takes  $O((n^3 + n^2 d^2 + n d^3) \sqrt{n+d})$  operations ignoring log factors (Nesterov and Nemirovskii, 1994, Chap. 6). The results of Albert (1961) imply that for any  $r > 0$  the iterates in Algorithm 2 will converge to within  $r$  of  $\boldsymbol{\theta}^*$  using an optimal number of samples in the high confidence ( $\delta \rightarrow 0$ ) regime. In Theorem 4 we obtain a finite time bound on the expected loss of  $\hat{\boldsymbol{\theta}}(t)$ . Define  $\ell_s(\boldsymbol{\theta}) := (Y_s - \mu_{I_s}(\boldsymbol{\theta}))^2$  as the squared error for  $\boldsymbol{\theta}$  at round  $s$ . The average empirical loss is defined as  $\hat{P}_t(\boldsymbol{\theta}) := \frac{1}{t} \sum_{s=1}^t \ell_s(\boldsymbol{\theta})$ .

**Assumption 3.** We assume that  $\lambda_{\max}(\nabla^2 \mu_i(\boldsymbol{\theta})) \leq \lambda_1$  for each  $i \in [n]$  and all  $\boldsymbol{\theta} \in \Theta$ .

Assumption 3 is a mild assumption on the curvature of the mean function at any  $\boldsymbol{\theta} \in \Theta$ .

**Theorem 4. (Dense CS Sample Complexity)** Suppose  $\ell_1(\boldsymbol{\theta}), \ell_2(\boldsymbol{\theta}), \dots, \ell_t(\boldsymbol{\theta}) : \mathbb{R}^d \rightarrow \mathbb{R}$  are squared

loss functions from a distribution that satisfies Assumption 3 and Assumption 4 in Appendix A.8.2. Further define  $P_t(\boldsymbol{\theta}) = \frac{1}{t} \sum_{s=1}^t \mathbb{E}_{I_s \sim \mathbf{p}_{\hat{\boldsymbol{\theta}}(s-1)}} [\ell_s(\boldsymbol{\theta}) | \mathcal{F}^{s-1}]$  where,  $\hat{\boldsymbol{\theta}}_t = \arg \min_{\boldsymbol{\theta} \in \Theta} \sum_{s=1}^t \ell_s(\boldsymbol{\theta})$ . If  $t$  is large enough such that  $\frac{\gamma \log(dt)}{t} \leq c' \min \left\{ \frac{1}{C_1 C_2}, \frac{\text{diameter}(\mathcal{B})}{C_2} \right\}$  then for a constant  $\gamma \geq 2$  and universal constants  $C_1, C_2, c'$ , we show that

$$\begin{aligned} & (1 - \rho_t) \frac{\sigma_t^2}{t} - \frac{C_1^2}{t^{\gamma/2}} \\ & \leq \mathbb{E} \left[ P_t(\hat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*) \right] \\ & \leq (1 + \rho_t) \frac{\sigma_t^2}{t} + \frac{\max_{\boldsymbol{\theta} \in \Theta} (P_t(\boldsymbol{\theta}) - P_t(\boldsymbol{\theta}^*))}{t^\gamma}, \end{aligned}$$

where  $\sigma_t^2 := \mathbb{E} \left[ \frac{1}{2} \left\| \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}^2 \right]$ , and  $\rho_t := (C_1 C_2 + 2\eta^2 \lambda_1^2) \sqrt{\frac{\gamma \log(dt)}{t}}$ .

*Proof. (sketch)* The first step in the proof is to relate  $\nabla^2 \hat{P}_t(\boldsymbol{\theta})$  to  $\nabla^2 P_t(\boldsymbol{\theta}^*)$  for any  $\boldsymbol{\theta}$  in a ball  $\mathcal{B}$  around  $\boldsymbol{\theta}^*$ . The ball  $\mathcal{B}$  is assumed in Assumption 4 to be a neighborhood where  $\nabla^2 \ell_s(\boldsymbol{\theta})$  satisfies a Lipschitz property. Assumption 4 in Appendix A.8.2 are standard and have also been made by Frostig et al. (2015) and Chaudhuri et al. (2015). Using Assumption 3 and Assumption 4, we can show that for  $t$  as large as mentioned in the Theorem statement, (1)  $\nabla^2 P_t(\boldsymbol{\theta}^*)$  is sandwiched in the positive semidefinite order by scaled multiples of  $\nabla^2 \hat{P}_t(\boldsymbol{\theta})$  for any  $\boldsymbol{\theta} \in \mathcal{B}$ , and (2) the empirical error minimizing  $\hat{\boldsymbol{\theta}}(t)$  is in the ball  $\mathcal{B}$  with probability  $1 - 1/t^\gamma$ , which is the good event  $\mathcal{E}$ . Using a Taylor series expansion around  $\hat{\boldsymbol{\theta}}(t)$  and the fact that  $\nabla \hat{P}_t(\hat{\boldsymbol{\theta}}(t)) = 0$  along with the relation between  $\nabla^2 \hat{P}_t(\boldsymbol{\theta})$  and  $\nabla^2 P_t(\boldsymbol{\theta}^*)$ , we can obtain an upper bound to  $\|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}^*\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)}$  in terms of  $\|\nabla \hat{P}_t(\boldsymbol{\theta}^*)\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}$  that can be shown to be decreasing with  $t$ . Further,  $\|\hat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}^*\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)}$  can also be used to obtain an upper bound to  $P_t(\hat{\boldsymbol{\theta}}(t)) - P_t(\boldsymbol{\theta}^*)$  using a Taylor series expansion. Finally we can bound  $\mathbb{E}[P_t(\hat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*)] = \mathbb{E}[(P_t(\hat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*)) I(\mathcal{E})] + \mathbb{E}[(P_t(\hat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*)) I(\mathcal{E}^c)]$  where  $I(\cdot)$  is the indicator. Since  $\mathbb{P}(\mathcal{E}^c) \leq 1/t^\gamma$ , the second term can be bounded as  $\max_{\boldsymbol{\theta} \in \Theta} (P_t(\boldsymbol{\theta}) - P_t(\boldsymbol{\theta}^*)) / t^\gamma$ , while the first term simplifies to  $(1 + \rho_t) \sigma_t^2 / t$ . The full proof is in Appendix A.8.  $\square$

The quantity  $\mathbb{E}_{I_s \sim \mathbf{p}_{\boldsymbol{\theta}}} [\ell_s(\boldsymbol{\theta})]$  characterizes the worst-case loss we could suffer at time  $s$  due to estimation error. This is because  $\mathbb{E}[\ell_s(\boldsymbol{\theta})] = \mathbb{E}_{I_s} \mathbb{E}[(Y_s - \mu_{I_s}(\boldsymbol{\theta}))^2 | I_s] = \mathbb{E}_{I_s} [(\mu_{I_s}(\boldsymbol{\theta}^*) - \mu_{I_s}(\boldsymbol{\theta}))^2 + 1/2]$ , and the definition of  $\mathbf{p}_{\boldsymbol{\theta}}$  ensures that  $\sum_{i=1}^n \mathbf{p}_{\boldsymbol{\theta}}(i) (\mu_i(\boldsymbol{\theta}^*) - \mu_i(\boldsymbol{\theta}))^2$  is maximized for the most confusing  $\boldsymbol{\theta}' \notin \mathcal{B}_r(\boldsymbol{\theta})$  at small enough  $\epsilon$ . We contrast this with the average-case loss

$\mathbb{E}_{I_s \sim \text{Uniform}([n])}[(\mu_{I_s}(\theta') - \mu_{I_s}(\theta))^2]$ , which is not larger than the expected value under  $I_s \sim \mathbf{p}_\theta$  due to the arg max in (9). It can thus be seen that the Chernoff sampling allows us to bound a more stringent notion of risk than traditionally looked at in the literature. The theorem bounds the running average of the worst-case losses at each time step. The simplified bound of Theorem 4 scales as  $O(d\sqrt{\log(dt)}/t + 1/t^2)$  (See Table 2 in Appendix A.8).

The term  $\sigma_t^2$  includes a norm under  $(\nabla^2 P_t(\theta^*))^{-1}$ . It is bounded by a quantity proportional to the maximum eigenvalue of  $(\nabla^2 P_t(\theta^*))^{-1}$ , equivalently the inverse of the minimum eigenvalue of  $\nabla^2 P_t(\theta^*) = \frac{2}{t} \sum_{s=1}^t \sum_{i=1}^n p_{\hat{\theta}(s-1)}(i) \nabla \mu_i(\theta^*) \nabla \mu_i(\theta^*)^T$ . If  $\hat{\theta}(t) \approx \theta^*$ , which is true for large  $t$ , then  $\mathbf{p}_{\hat{\theta}(t)}$  is the optimization solution in Theorem 3 and it maximizes the minimum eigenvalue of  $\sum_{i=1}^n p_{\hat{\theta}(t)}(i) \nabla \mu_i(\hat{\theta}(t)) \nabla \mu_i(\hat{\theta}(t))^T$ . Thus CS approximately minimizes an upper bound to the estimation error.

## 4 Experiments

In this section we show numerical evaluations of CS against other algorithms. The confidence intervals (CI) that we plotted are just mean  $\pm 1$  standard deviation. A wider CI means more variability in performance across different trials. We run each experiment over 50 independent trials.

**Active testing experiment (3 Group setting):** Consider an environment of 50 actions and 6 parameters. The actions can be divided into three groups. The first group contains a single action that most effectively discriminates  $\theta^*$  from all other hypotheses. The second group consists of 5 actions, each of which can discriminate one hypothesis from the others. Finally, the third group of 44 actions are barely informative as their means are similar under all hypotheses. The mean values under all hypotheses are given in Appendix A.9.1. We show the empirical performance of different policies in the Figure 1b. Both CS and T2 outperforms TP which conducts a uniform exploration over the actions in the second group in its first phase. T2 performs well as it samples the action in the first group when  $\theta^*$  is either the most-likely or the second most-likely hypothesis. Unif performs the worst as it uniformly samples all actions, including the non-informative actions in the third group. CS outperforms CSE as it does not conduct forced exploration over non-informative actions. Batch-CS with  $B = 5, 10, 15$  has increasing sample complexity with larger batches.

**Non-Linear Model:** Consider a non-linear class of functions parameterized as  $\mu_i(\theta) := 1/(1 + \exp(-\mathbf{x}_i^T \theta))$ , where  $\|\theta\|_2 = 1$ , each action has an associated feature

vector  $\mathbf{x}_i \in \mathbb{R}^2$  and it returns a value whose expectation is  $\mu_i(\theta^*)$ . The goal is to choose actions such that the estimation error  $\|\hat{\theta}(t) - \theta^*\|_2$  reduces using as few samples as possible. The setting consist of three groups of actions: **a)** the optimal action, **b)** the informative action (orthogonal to optimal action) that maximally reduces the uncertainty of  $\hat{\theta}(t)$  and **c)** the 48 less-informative actions as shown in Figure 1c. Appendix A.9.2 contains more implementation details. We apply CS and CSE to this problem and compare it to baselines Unif, and ActiveS. Figure 1d shows that CS outperforms ActiveS and is able to find  $\theta^*$  quickly. CS performs similar to EMCM but note that EMCM has no convergence guarantees and requires hyper-parameter tuning.

**Neural Network:** Consider a collection of data points  $\{\mathbf{x}_i \in \mathbb{R}^2 : i \in [n]\}$ , each of which is assigned a ground truth scalar mean value by a non-linear function. The particular form for the mean function of action  $i$  is the following:  $\mu_i(\theta^*) = c_1 \sigma(\mathbf{w}_1^T \mathbf{x}_i + b_1) + c_2 \sigma(\mathbf{w}_2^T \mathbf{x}_i + b_2)$ , where  $\theta^* = (\mathbf{w}_1, b_1, \mathbf{w}_2, b_2, c_1, c_2)$  is the parameter characterizing the ground truth function, and  $\sigma(\cdot) := \max\{0, \cdot\}$  is the non-linear ReLU activation function. This is a single hidden-layer neural network with input layer weights  $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^2$ , biases  $b_1, b_2 \in \mathbb{R}$ , and output weights  $c_1, c_2 \in \{-1, 1\}$ . Our objective in the experiment is to learn the neural network from noisy observations:  $\{(\mathbf{x}_{I_s}, \mu_{I_s}(\theta^*) + \text{noise}) : s \in [t], \text{noise is i.i.d. zero mean Gaussian}\}$  collected by sampling  $(I_1, I_2, \dots, I_t)$  according to the Chernoff proportions defined in Algorithm 2. The architecture of the network is known, but the weights and biases must be learned. The data points are shown in a scatter plot in Figure 1e. More implementation details are in Appendix A.9.3. The non-uniformity of the data distribution increases the difficulty of the learning task. The performance of a learning algorithm is measured by tracking the estimation error  $\|\hat{\theta}(t) - \theta^*\|_2$  during the course of training. The plot in Figure 1f shows the average and standard deviation of the estimation error over 10 trials for CS, CSE, ActiveS and Unif baseline. Again both CS and CSE outperforms Unif and ActiveS. We note that other works (Cohn, 1996; Fukumizu, 2000) have also considered active training of neural networks. Other approaches for active sampling have been described in the survey by Settles (2009). Our neural network learning experiment shows the generality of our approach in active sampling.

**Real Dataset:** We consider two real world datasets from UCI called Red Wine (Cortez et al., 2009) (1600 actions) and Air Quality (De Vito et al., 2008) (1500 actions). The performance is shown in Figure 1g and Figure 1h respectively where CS and CSE outperforms ActiveS and Unif. Further experiment details are in Appendix A.9.4, including Figure 2 which shows that



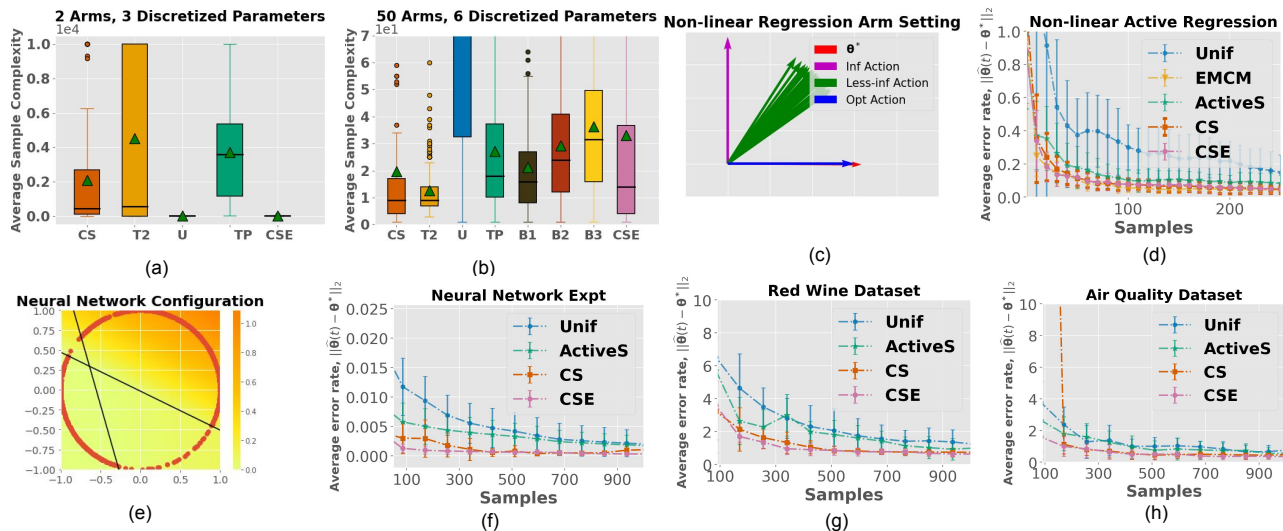


Figure 1: (a) Sample complexity over Example 1 with  $\delta = 0.1$ . (b) Sample complexity over 3 Group setting with  $\delta = 0.1$ . U = Unif, B1 = Batch-CS ( $B = 5$ ), B2 = Batch-CS ( $B = 10$ ), B3 = Batch-CS ( $B = 15$ ). The green triangle marker represents the mean stopping time. The horizontal line across each box represents the median. (c) The action feature vectors in  $\mathbb{R}^2$  for active regression experiment. Inf = Informative action, Opt = Optimal action, Less-inf = Less informative action. (d) Shows average error rate for Active Regression. (e) Training data for neural network example shown as red dots. Yellow to red colors indicate increasing values for the target mean function. Black lines denote “activation” boundaries of the two hidden layer neurons. (f) Learning a neural network model. (g), (h) Experiment with Red Wine and Air Quality Dataset.

in the real-world dataset CS proportion is sparse over the actions.

## 5 Conclusions, Limitations, & Future Work

This paper proposes a unifying approach to solve active testing and active regression problems. We obtain non-asymptotic guarantees on the performance of CS in active testing and extend it to the problem of active regression. CS has comparable performance to existing state-of-the-art methods and is a relatively easy algorithm to implement. Nevertheless, its sampling proportion is updated before collecting each sample, which increases the computational cost (one solution for this is Batch-CS which solves the optimization only after collecting a batch of samples). Further, Assumption 2 is a strong assumption, and other works have removed that in the context of active testing by modifying the CS strategy (one solution for this is CSE which incorporates a certain amount of random sampling).

CS can be excessively aggressive in the initial stages. This is because it chooses actions according to a sampling proportion that is optimal when  $\hat{\theta}(t) = \theta^*$ , which is not true initially. Other methods of exploration could be useful in the earlier stages.

The extension of CS to active regression requires to find the least squares estimate  $\hat{\theta}(t)$  which could be computationally expensive. Theoretical guarantees require several assumptions which may not always be satisfied. In addition to the regularity assumptions, we also need the smoothness of the mean function for Theorem 4. Future directions include obtaining a lower bound for the active testing in the moderate confidence regime and incorporating the geometry of the actions in the sampling strategy for the regression setting. Another direction is to obtain the sampling proportions when the mean function is not differentiable everywhere.

**Acknowledgements:** This work was partially supported by AFOSR grant FA9550-18-1-0166. The first author was supported by 2019-20 Chancellor’s Opportunity Fellowship by the University of Wisconsin-Madison.

## References

- Albert, A. E. (1961). The sequential design of experiments for infinitely many states of nature. *The Annals of Mathematical Statistics*, pages 774–799.
- Balcan, M.-F., Beygelzimer, A., and Langford, J. (2009). Agnostic active learning. *Journal of Computer and System Sciences*, 75(1):78–89.
- Balcan, M.-F. and Long, P. (2013). Active and passive learning of linear separators under log-concave distributions. In *Conference on Learning Theory*, pages 288–316. PMLR.
- Blot, W. J. and Meeter, D. A. (1973). Sequential experimental design procedures. *Journal of the American Statistical Association*, 68(343):586–593.
- Bu, Y., Lu, J., and Veeravalli, V. V. (2019). Active and adaptive sequential learning with per time-step excess risk guarantees. In *2019 53rd Asilomar Conference on Signals, Systems, and Computers*, pages 1606–1610. IEEE.
- Cai, W., Zhang, M., and Zhang, Y. (2016). Batch mode active learning for regression with expected model change. *IEEE transactions on neural networks and learning systems*, 28(7):1668–1681.
- Castro, R., Willett, R., and Nowak, R. (2005). Faster rates in regression via active learning. In *NIPS*, volume 18, pages 179–186.
- Chaudhuri, K., Jain, P., and Natarajan, N. (2017). Active heteroscedastic regression. In *International Conference on Machine Learning*, pages 694–702. PMLR.
- Chaudhuri, K., Kakade, S. M., Netrapalli, P., and Sanghavi, S. (2015). Convergence rates of active learning for maximum likelihood estimation. In *Advances in Neural Information Processing Systems*, pages 1090–1098.
- Chaudhuri, P. and Mykland, P. A. (1993). Nonlinear experiments: Optimal design and inference based on likelihood. *Journal of the American Statistical Association*, 88(422):538–546.
- Chernoff, H. (1959). Sequential design of experiments. *The Annals of Mathematical Statistics*, 30(3):755–770.
- Cohn, D. A. (1996). Neural network exploration using optimal experiment design. *Neural Networks*, 9(6):1071 – 1083.
- Combes, R., Magureanu, S., and Proutiere, A. (2017). Minimal exploration in structured stochastic bandits. In *Advances in Neural Information Processing Systems*, pages 1763–1771.
- Cortez, P., Cerdeira, A., Almeida, F., Matos, T., and Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decision support systems*, 47(4):547–553.
- Dasgupta, S. (2005). Coarse sample complexity bounds for active learning. In *NIPS*, volume 18, pages 235–242.
- Dasgupta, S., Hsu, D. J., and Monteleoni, C. (2008). A general agnostic active learning algorithm. In *International Symposium on Artificial Intelligence and Mathematics, ISAIM 2008, Fort Lauderdale, Florida, USA, January 2-4, 2008*.
- De Vito, S., Massera, E., Piga, M., Martinotto, L., and Di Francia, G. (2008). On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. *Sensors and Actuators B: Chemical*, 129(2):750–757.
- Degenne, R., Ménard, P., Shang, X., and Valko, M. (2020). Gamification of pure exploration for linear bandits. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 2432–2442. PMLR.
- Dette, H. and Studden, W. J. (1993). Geometry of e-optimality. *The Annals of Statistics*, 21(1):416–433.
- Fiez, T., Jain, L., Jamieson, K. G., and Ratliff, L. (2019). Sequential experimental design for transductive linear bandits. In Wallach, H., Larochelle, H., Beygelzimer, A., d’Álché Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 32, pages 10667–10677. Curran Associates, Inc.
- Fontaine, X., Perrault, P., Valko, M., and Perchet, V. (2019). Online a-optimal design and active linear regression. *arXiv preprint arXiv:1906.08509*.
- Frostig, R., Ge, R., Kakade, S. M., and Sidford, A. (2015). Competing with the empirical risk minimizer in a single pass. In Grünwald, P., Hazan, E., and Kale, S., editors, *Proceedings of The 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, pages 728–763, Paris, France. PMLR.
- Fukumizu, K. (2000). Statistical active learning in multilayer perceptrons. *IEEE Transactions on Neural Networks*, 11(1):17–26.
- Garivier, A. and Kaufmann, E. (2016). Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, pages 998–1027.

- Goetz, J., Tewari, A., and Zimmerman, P. (2018). Active learning for non-parametric regression using purely random trees. *Advances in Neural Information Processing Systems*, 31.
- Hanneke, S. (2007). A bound on the label complexity of agnostic active learning. In *Proceedings of the 24th international conference on Machine learning*, pages 353–360.
- Hsu, D., Kakade, S., Zhang, T., et al. (2012). A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17.
- Huang, R., Ajallooeian, M. M., Szepesvári, C., and Müller, M. (2017). Structured best arm identification with fixed confidence. In Hanneke, S. and Reyzin, L., editors, *International Conference on Algorithmic Learning Theory, ALT 2017, 15-17 October 2017, Kyoto University, Kyoto, Japan*, volume 76 of *Proceedings of Machine Learning Research*, pages 593–616. PMLR.
- Katz-Samuels, J., Zhang, J., Jain, L., and Jamieson, K. (2021). Improved algorithms for agnostic pool-based active classification. *arXiv preprint arXiv:2105.06499*.
- Lattimore, T. and Szepesvári, C. (2020). *Bandit algorithms*. Cambridge University Press.
- Naghshvar, M. and Javidi, T. (2013). Active sequential hypothesis testing. *The Annals of Statistics*, 41(6):2703–2738.
- Nesterov, Y. and Nemirovskii, A. (1994). *Interior-point polynomial algorithms in convex programming*. SIAM.
- Nitinawarat, S., Atia, G. K., and Veeravalli, V. V. (2013). Controlled sensing for multihypothesis testing. *IEEE Transactions on Automatic Control*, 58(10):2451–2464.
- Pronzato, L. and Pázman, A. (2013). Design of experiments in nonlinear models. *Lecture notes in statistics*, 212.
- Pukelsheim, F. (2006). *Optimal design of experiments*. SIAM.
- Sabato, S. and Munos, R. (2014). Active regression by stratification. *arXiv preprint arXiv:1410.5920*.
- Settles, B. (2009). Active learning literature survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison.
- Silvey, S. (1980). *Optimal design: an introduction to the theory for parameter estimation*. Monographs on applied probability and statistics. Chapman and Hall.
- Soare, M., Lazaric, A., and Munos, R. (2014). Best-arm identification in linear bandits. In *Advances in Neural Information Processing Systems*, pages 828–836.
- Tsybakov, A. B. (2008). *Introduction to nonparametric estimation*. Springer Science & Business Media.
- Tukey, J. W. (1977). *Exploratory data analysis*, volume 2. Reading, MA.
- Vaidhiyan, N. K. and Sundaresan, R. (2017). Learning to detect an oddball target. *IEEE Transactions on Information Theory*, 64(2):831–852.
- Wu, D. (2018). Pool-based sequential active learning for regression. *IEEE transactions on neural networks and learning systems*, 30(5):1348–1359.
- Wu, D., Lin, C.-T., and Huang, J. (2019). Active learning for regression using greedy sampling. *Information Sciences*, 474:90–105.
- Yu, K., Bi, J., and Tresp, V. (2006). Active learning via transductive experimental design. In *Proceedings of the 23rd international conference on Machine learning*, pages 1081–1088.
- Zhang, C. and Chaudhuri, K. (2014). Beyond disagreement-based agnostic active learning. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 442–450.

## Contents

1	Theoretical Comparison and Probability Tools .....	A.1
1.1	Theoretical Comparison .....	A.1.1
1.2	Probability Tools .....	A.1.2
2	Chernoff Sample Complexity Proof .....	A.2
2.1	Concentration Lemma .....	A.2.1
2.2	Concentration of $T_{Y^t}$ .....	A.2.2
2.3	Proof of correctness for General Sub-Gaussian Case .....	A.2.3
2.4	Stopping time Correctness Lemma for the Gaussian Case .....	A.2.4
2.5	Proof of <b>CS</b> Sample Complexity (Theorem 1) .....	A.2.5
3	Proof of <b>T2</b> Sample Complexity .....	A.3
3.1	Concentration Lemma .....	A.3.1
3.2	Proof of <b>T2</b> Sample Complexity (Proposition 1) .....	A.3.2
4	Proof of Sample Complexity of <b>Batch-CS</b> (Proposition 2) .....	A.4
5	Proof of <b>CSE</b> Sample Complexity (Proposition 3) .....	A.5
6	Minimax Optimality Proof (Theorem 2).....	A.6
7	Proof of Theorem 3 (Continuous hypotheses) .....	A.7
7.1	How to solve the optimization .....	A.7.1
8	<b>CS</b> Proof for Continuous Hypotheses .....	A.8
8.1	Theoretical Comparisons for Active Regression .....	A.8.1
8.2	Discussion on Definitions and Assumptions for Continuous Hypotheses .....	A.8.2
8.3	Concentration Lemma for Continuous Hypotheses .....	A.8.3
8.4	Support Lemma for Continuous Hypotheses.....	A.8.4
8.5	Proof of <b>CS</b> Convergence for Continuous Hypotheses (Theorem 4) .....	A.8.5
9	Additional Experiment Details .....	A.9
9.1	Hypothesis Testing Experiments .....	A.9.1
9.2	Active Regression Experiment for Non-linear Reward Model .....	A.9.2
9.3	Active Regression Experiment for Neural Networks .....	A.9.3
9.4	Active Regression for the UCI Datasets .....	A.9.4
10	Table of Notations .....	A.10

## A Appendix

### A.1 Theoretical Comparison of Active Testing and Probability Tools

In this section we compare theoretically our work against other existing works in active testing. We also state a few standard lemmas in Probability Tools that we use to prove our results.

### A.1.1 Theoretical Comparison of Active Testing

**Active Testing:** We compare our work against Chernoff (1959), Albert (1961) which extended Chernoff (1959) to continuous hypotheses space partitioned into two disjoint space, and the recent active testing algorithms of Naghshvar and Javidi (2013); Nitinawarat et al. (2013). We first recall the following problem complexity parameters as follows:

$$D_0 := \max_{\mathbf{p}} \min_{\theta' \neq \theta^*} \sum_{i=1}^n p(i) (\mu_i(\theta') - \mu_i(\theta^*))^2, \quad D_1 := \min_{\{\mathbf{p}_{\theta}: \theta \in \Theta\}} \min_{\theta' \neq \theta^*} \sum_{i=1}^n p_{\theta}(i) (\mu_i(\theta') - \mu_i(\theta^*))^2$$

$$D_{NJ} := \left( \min_{\theta \in \Theta} \min_{\theta' \neq \theta} \sum_{i=1}^n p_{\theta'}(i) (\mu_i(\theta) - \mu_i(\theta'))^2 \right)^2 \max_{\mathbf{p}} \min_{\theta \in \Theta} \min_{\theta' \neq \theta} \sum_{i=1}^n p(i) (\mu_i(\theta) - \mu_i(\theta'))^2 \quad (12)$$

$$D_e := \min_{\theta' \neq \theta^*} \sum_{i=1}^n \frac{1}{n} (\mu_i(\theta') - \mu_i(\theta^*))^2 \quad (13)$$

where, the quantity  $D_{NJ}$  is defined in Naghshvar and Javidi (2013) and is an artifact of the forced exploration conducted by their algorithm. Note that  $D_{NJ} < D_1$  by definition. Also,  $D_0 > D_e$  and  $D_0 > D_1$  and  $D_e > 0$  by definition. Note that due to the factor  $D_{NJ}$ , **TP** can perform worse than **CS** in certain instances (see Active testing experiment in Figure 1a, and Figure 1b). We summarize our result in context of other existing results in the following table:

Sample Complexity Bound	Comments
$\mathbb{E}[\tau_{\delta}] \leq \frac{C \log J/\delta}{D_0} + o(\log \frac{1}{\delta})$	Upper bound in Chernoff (1959). Optimal for $\delta \rightarrow 0$ .
$\mathbb{E}[\tau_{\delta}] \leq \frac{C \log J/\delta}{D_0} + o(\log \frac{1}{\delta})$	Upper bound in Albert (1961). Optimal for $\delta \rightarrow 0$ . Extension to compound hypotheses.
$\mathbb{E}[\tau_{\delta}] \leq \frac{C \log J/\delta}{D_0} + o(\log \frac{1}{\delta})$	Upper bound in (Nitinawarat et al., 2013). Bound valid for discrete hypotheses. It does not require Assumption 2.
$\mathbb{E}[\tau_{\delta}] \leq O\left(\frac{\log(J)}{D_{NJ}} + \frac{\log(J/\delta)}{D_0}\right)$	Upper bound of <b>TP</b> in Naghshvar and Javidi (2013). Valid for any $\delta \in (0, 1]$ . Asymptotically optimal for $\delta \rightarrow 0$ .
$\mathbb{E}[\tau_{\delta}] \leq O\left(\frac{\log(J)}{D_1} + \frac{\log(J/\delta)}{D_0}\right)$	<b>CS (Ours)</b> . Valid for any $\delta \in (0, 1]$ . Asymptotically optimal for $\delta \rightarrow 0$ .
$\mathbb{E}[\tau_{\delta}] \leq O\left(\frac{\log(J)}{D_e} + \frac{\log(J/\delta)}{D_0}\right)$	<b>CSE (Ours)</b> . Valid for any $\delta \in (0, 1]$ . Asymptotically optimal for $\delta \rightarrow 0$ . Does not require Assumption 2.

Table 1: Active Testing (top) and Regression (bottom) comparison.  $D_{NJ} < D_1$ ,  $D_1 < D_0$ , and  $D_e < D_0$ .

### A.1.2 Probability Tools

**Lemma 1. (Restatement of Lemma 15.1 in Lattimore and Szepesvári (2020), Divergence Decomposition)** Let  $B$  and  $B'$  be two bandit models having different optimal hypothesis  $\theta^*$  and  $\theta'^*$  respectively. Fix some policy  $\pi$  and round  $n$ . Let  $\mathbb{P}_{B,\pi}$  and  $\mathbb{P}_{B',\pi}$  be two probability measures induced by some  $n$ -round interaction of  $\pi$  with  $B$  and  $\pi$  with  $B'$  respectively. Then

$$\text{KL}(\mathbb{P}_{B,\pi} || \mathbb{P}_{B',\pi}) = \sum_{i=1}^n \mathbb{E}_{B,\pi}[Z_i(n)] \cdot \text{KL}(\mu_i(\theta) || \mu_i(\theta^*))$$

where,  $\text{KL}(\cdot|\cdot)$  denotes the Kullback-Leibler divergence between two probability measures and  $Z_i(n)$  denotes the number of times action  $i$  has been sampled till round  $n$ .

**Lemma 2. (Restatement of Lemma 2.6 in Tsybakov (2008))** Let  $\mathbb{P}, \mathbb{Q}$  be two probability measures on the same measurable space  $(\Omega, \mathcal{F})$  and let  $\xi \subset \mathcal{F}$  be any arbitrary event then

$$\mathbb{P}(\xi) + \mathbb{Q}(\xi^c) \geq \frac{1}{2} \exp(-\text{KL}(\mathbb{P}|\mathbb{Q}))$$

where  $\xi^c$  denotes the complement of event  $\xi$  and  $\text{KL}(\mathbb{P}|\mathbb{Q})$  denotes the Kullback-Leibler divergence between  $\mathbb{P}$  and  $\mathbb{Q}$ .

**Lemma 3. (Hoeffding's Lemma)** Let  $Y$  be a real-valued random variable with expected value  $\mathbb{E}[Y] = \mu$ , such that  $a \leq Y \leq b$  with probability one. Then, for all  $\lambda \in \mathbb{R}$

$$\mathbb{E}[e^{\lambda Y}] \leq \exp\left(\lambda\mu + \frac{\lambda^2(b-a)^2}{8}\right)$$

**Lemma 4. (Proposition 2 of (Hsu et al., 2012))** Let  $\mathbf{u}_1, \dots, \mathbf{u}_n$  be a martingale difference vector sequence (i.e.,  $\mathbb{E}[\mathbf{u}_i | \mathbf{u}_1, \dots, \mathbf{u}_{i-1}] = 0$  for all  $i = 1, \dots, n$ ) such that

$$\sum_{i=1}^n \mathbb{E}[\|\mathbf{u}_i\|^2 | \mathbf{u}_1, \dots, \mathbf{u}_{i-1}] \leq v \quad \text{and} \quad \|\mathbf{u}_i\| \leq b$$

for all  $i = 1, \dots, n$ , almost surely. For all  $t > 0$

$$\Pr\left[\left\|\sum_{i=1}^n \mathbf{u}_i\right\| > \sqrt{v} + \sqrt{8vt} + (4/3)bt\right] \leq e^{-t}$$

## A.2 Chernoff Sample Complexity Proof

### A.2.1 Concentration Lemma

**Lemma 5.** Define  $L_t(\boldsymbol{\theta}^*)$  as the sum squared errors for the hypothesis parameterized by  $\boldsymbol{\theta}^*$ . Let  $\tau_{\boldsymbol{\theta}^*} = \min\{t : L_t(\boldsymbol{\theta}') - L_t(\boldsymbol{\theta}^*) > \beta(J, \delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*\}$ . Then we can bound the probability that  $\tau_{\boldsymbol{\theta}^*}$  is larger than  $t$  as

$$\mathbb{P}(\tau_{\boldsymbol{\theta}^*} > t) \leq JC_1 \exp(-C_2 t)$$

where,  $J := |\Theta|$ ,  $C_1 := 110 + 55 \max\left\{1, \frac{\eta^2}{2D_1^2 c}\right\}$ ,  $C_2 := \frac{2D_1^2 \min\{(c-1)^2, c\}}{\eta^2}$ ,  $\eta > 0$  defined in Definition 1 and  $D_1 := \min_{\boldsymbol{\theta} \in \Theta, \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n p_{\boldsymbol{\theta}}(i) (\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2$ .

*Proof.* We consider the following events when the difference of squared errors is below certain values:

$$\begin{aligned} \xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t) &:= \{L_t(\boldsymbol{\theta}') - L_t(\boldsymbol{\theta}^*) < \beta(J, \delta)\}, \\ \tilde{\xi}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t) &:= \{L_t(\boldsymbol{\theta}') - L_t(\boldsymbol{\theta}^*) < \alpha(J)\}. \end{aligned}$$

Then we define the time  $\tau_{\boldsymbol{\theta}^*}$  as follows:

$$\tau_{\boldsymbol{\theta}^*} := \min\{t : L_t(\boldsymbol{\theta}') - L_t(\boldsymbol{\theta}^*) > \beta(J, \delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*\}$$

which is the first round when  $L_t(\boldsymbol{\theta}')$  crosses  $\beta(J, \delta)$  threshold against  $L_t(\boldsymbol{\theta}^*)$  for all  $\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*$ . We also define the time  $\tilde{\tau}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}$  as follows:

$$\tilde{\tau}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*} := \min\{t : L_{t'}(\boldsymbol{\theta}') - L_{t'}(\boldsymbol{\theta}^*) > \alpha(J), \forall t' > t\} \quad (14)$$

which is the first round when  $L_t(\boldsymbol{\theta}')$  crosses  $\alpha(J)$  threshold against  $L_t(\boldsymbol{\theta}^*)$ . We will be particularly interested in the time

$$\tilde{\tau}_{\boldsymbol{\theta}^*} := \max_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \{\tilde{\tau}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}\}. \quad (15)$$

Let  $\Delta_t(\boldsymbol{\theta}')$  denote the difference of squared errors between hypotheses  $\boldsymbol{\theta}'$  and  $\boldsymbol{\theta}^*$  (for Gaussian noise model, it is equal to the log-likelihood ratio between hypotheses parameterized by  $\boldsymbol{\theta}'$  and  $\boldsymbol{\theta}^*$ ) shown below.

$$\Delta_t(\boldsymbol{\theta}') = L_t(\boldsymbol{\theta}') - L_t(\boldsymbol{\theta}^*). \quad (16)$$

A key thing to note that for  $D_1 > 0$  (Assumption 2) the  $\mathbb{E}[\Delta_t(\boldsymbol{\theta}')] > tD_1$  which is shown as follows:

$$\begin{aligned} \mathbb{E}_{I^t, Y^t}[\Delta_t(\boldsymbol{\theta}')] &= \mathbb{E}_{I^t, Y^t}[L_t(\boldsymbol{\theta}') - L_t(\boldsymbol{\theta}^*)] = \mathbb{E}_{I^t, Y^t} \left[ \sum_{s=1}^t (Y_s - \mu_{I_s}(\boldsymbol{\theta}^*))^2 - \sum_{s=1}^t (Y_s - \mu_{I_s}(\boldsymbol{\theta}'))^2 \right] \\ &= \sum_{s=1}^t \mathbb{E}_{I_s} \mathbb{E}_{Y_s | I_s} \left[ (\mu_{I_s}(\boldsymbol{\theta}^*) - \mu_{I_s}(\boldsymbol{\theta}'))^2 | I_s \right] \\ &= \sum_{s=1}^t \sum_{i=1}^n \mathbb{P}(I_s = i) (\mu_i(\boldsymbol{\theta}^*) - \mu_i(\boldsymbol{\theta}'))^2 \stackrel{(a)}{\geq} tD_1 \end{aligned}$$

where, (a) follows from the definition of  $D_1$  in Theorem 1. Then it follows that,

$$\begin{aligned} \mathbb{P}(\tilde{\xi}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)) &= \mathbb{P}(\Delta_t(\boldsymbol{\theta}') < \alpha(J)) = \mathbb{P}(\Delta_t(\boldsymbol{\theta}') - \mathbb{E}[\Delta_t(\boldsymbol{\theta}')] < \alpha(J) - \mathbb{E}[\Delta_t(\boldsymbol{\theta}')] \\ &\stackrel{(a)}{\leq} \mathbb{P}(\Delta_t(\boldsymbol{\theta}') - \mathbb{E}[\Delta_t(\boldsymbol{\theta}')] < \alpha(J) - tD_1) \end{aligned}$$

where, in (a) the choice of  $tD_1$  follows as  $\mathbb{E}[\Delta_t(\boldsymbol{\theta}')] \geq tD_1$  for  $D_1 > 0$ . Similarly, we can show that  $\mathbb{E}[\Delta_s(\boldsymbol{\theta}')] \geq D_0$  for all rounds  $s \geq \tilde{\tau}_{\boldsymbol{\theta}^*}$  where  $\tilde{\tau}_{\boldsymbol{\theta}^*}$  is defined in (15). Then we can show that,

$$\begin{aligned} \mathbb{P}(\xi_{\boldsymbol{\theta}',\boldsymbol{\theta}^*}(t) | \tilde{\tau}_{\boldsymbol{\theta}^*}) &= \mathbb{P}(\Delta_t(\boldsymbol{\theta}') < \beta(J, \delta) | \tilde{\tau}_{\boldsymbol{\theta}^*}) = \mathbb{P}(\Delta_t(\boldsymbol{\theta}') - \mathbb{E}[\Delta_t(\boldsymbol{\theta}')] < \beta(J, \delta) - \mathbb{E}[\Delta_t(\boldsymbol{\theta}')] | \tilde{\tau}_{\boldsymbol{\theta}^*}) \\ &\stackrel{(a)}{\leq} \mathbb{P}(\Delta_t(\boldsymbol{\theta}') - \mathbb{E}[\Delta_t(\boldsymbol{\theta}')] < \beta(J, \delta) - (t - \tilde{\tau}_{\boldsymbol{\theta}^*})D_0 | \tilde{\tau}_{\boldsymbol{\theta}^*}) \\ &\stackrel{(b)}{=} \mathbb{P}\left(\Delta_t(\boldsymbol{\theta}') - \mathbb{E}[\Delta_t(\boldsymbol{\theta}')] < D_0 \left( \frac{\beta(J, \delta)}{D_0} - t + \tilde{\tau}_{\boldsymbol{\theta}^*} \right) | \tilde{\tau}_{\boldsymbol{\theta}^*}\right) \end{aligned}$$

where, in (a) the choice of  $(t - \tilde{\tau}_{\boldsymbol{\theta}^*})D_0$  follows as  $\mathbb{E}[\Delta_s(\boldsymbol{\theta}')] \geq D_0$  for any round  $s > \tilde{\tau}_{\boldsymbol{\theta}^*}$ , and finally in (b) the quantity  $D_0 \left( \frac{\beta(J, \delta)}{D_0} - t + \tilde{\tau}_{\boldsymbol{\theta}^*} \right)$  is negative for  $t \geq (1+c) \left( \frac{\alpha(J)}{D_1} + \frac{\beta(J, \delta)}{D_0} \right)$  and  $\tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D_1} + \frac{tc}{2}$  which allows us to apply the concentration inequality for conditionally independent random variables stated in Lemma 8. Then we can show that,

$$\begin{aligned} \mathbb{P}(\tau_{\boldsymbol{\theta}^*} > t) &\leq \mathbb{P}\left(\bigcup_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)\right) \\ &= \mathbb{P}\left(\left\{\bigcup_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)\right\} \cap \left\{\tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D_1} + \frac{tc}{2}\right\}\right) + \mathbb{P}\left(\bigcup_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \{\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)\} \cap \left\{\tilde{\tau}_{\boldsymbol{\theta}^*} \geq \frac{\alpha(J)}{D_1} + \frac{tc}{2}\right\}\right) \\ &\leq \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \mathbb{P}\left(\{\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)\} \cap \left\{\tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D_1} + \frac{tc}{2}\right\}\right) + \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \mathbb{P}\left(\tilde{\tau}_{\boldsymbol{\theta}^*} \geq \frac{\alpha(J)}{D_1} + \frac{tc}{2}\right) \\ &= \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \mathbb{P}\left(\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t) | \tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D_1} + \frac{tc}{2}\right) \mathbb{P}\left(\tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D_1} + \frac{tc}{2}\right) \\ &\quad + \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \mathbb{P}\left(\tilde{\xi}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t') \text{ is true for some } t' > \frac{\alpha(J)}{D_1} + \frac{tc}{2}\right) \\ &\leq \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \mathbb{P}\left(\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t) | \tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D_1} + \frac{tc}{2}\right) + \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t': t' \geq \frac{\alpha(J)}{D_1} + \frac{tc}{2}} \mathbb{P}\left(\tilde{\xi}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t')\right) \\ &\leq \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \mathbb{P}\left(\Delta_t(\boldsymbol{\theta}') - \mathbb{E}[\Delta_t(\boldsymbol{\theta}')] < D_0 \left( \frac{\beta(J, \delta)}{D_0} - t + \tilde{\tau}_{\boldsymbol{\theta}^*} \right) | \tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D_1} + \frac{tc}{2}\right) \\ &\quad + \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t': t' \geq \frac{\alpha(J)}{D_1} + \frac{tc}{2}} \mathbb{P}(\Delta_{t'}(\boldsymbol{\theta}') - \mathbb{E}[\Delta_{t'}(\boldsymbol{\theta}')] < \alpha(J) - t'D_1) \end{aligned}$$

$$\begin{aligned}
 &\leq \sum_{\theta' \neq \theta^*} \mathbb{P} \left( \Delta_t(\theta') - \mathbb{E}[\Delta_t(\theta')] < D_0 \left( \frac{\beta(J, \delta)}{D_0} + \frac{\alpha(J)}{D_1} - t + \frac{tc}{2} \right) \right) \\
 &\quad + \sum_{\theta' \neq \theta^*} \sum_{t': t' \geq \frac{\alpha(J)}{D_1} + \frac{tc}{2}} \mathbb{P}(\Delta_{t'}(\theta') - \mathbb{E}[\Delta_{t'}(\theta')] < \alpha(J) - t'D_1) \\
 &\stackrel{(a)}{\leq} \exp(4) \sum_{\theta' \neq \theta^*} \exp \left( -\frac{2D_1^2 t (c/2 - 1)^2}{\eta^2} \right) + \exp(4) \sum_{\theta' \neq \theta^*} \frac{\exp \left( -\frac{2D_1^2 tc}{\eta^2} \right)}{1 - \exp \left( -\frac{2D_1^2}{\eta^2} \right)} \\
 &\stackrel{(b)}{\leq} \exp(4) \sum_{\theta' \neq \theta^*} \exp \left( -\frac{2D_1^2 t (\frac{c}{2} - 1)^2}{\eta^2} \right) + \exp(4) \sum_{\theta' \neq \theta^*} \exp \left( -\frac{2D_1^2 tc}{\eta^2} \right) \left( 1 + \max \left\{ 1, \frac{\eta^2}{2D_1^2} \right\} \right) \\
 &\stackrel{(c)}{\leq} 55 \sum_{\theta' \neq \theta^*} \exp \left( -\frac{2D_1^2 t (\frac{c}{2} - 1)^2}{\eta^2} \right) + 55 \sum_{\theta' \neq \theta^*} \exp \left( -\frac{2D_1^2 tc}{\eta^2} \right) + 55 \sum_{\theta' \neq \theta^*} \exp \left( -\frac{2D_1^2 tc}{\eta^2} \right) \max \left\{ 1, \frac{\eta^2}{2D_1^2} \right\} \\
 &\leq \sum_{\theta' \neq \theta^*} \exp \left( -\frac{2D_1^2 t \min \{ (\frac{c}{2} - 1)^2, c \}}{\eta^2} \right) \left[ 110 + 55 \max \left\{ 1, \frac{\eta^2}{2D_1^2} \right\} \right] \stackrel{(d)}{\leq} JC_1 \exp(-C_2 t)
 \end{aligned}$$

where, (a) follows from Lemma 6 and Lemma 7, (b) follows from the identity that  $1/(1 - \exp(-x)) \leq 1 + \max\{1, 1/x\}$  for  $x > 0$ , (c) follows for  $0 < c < 1$ , and in (d) we substitute  $C_1 := 110 + 55 \max\left\{1, \frac{\eta^2}{2D_1^2}\right\}$ ,  $C_2 := \frac{2D_1^2 \min\{(c/2 - 1)^2, c\}}{\eta^2}$  and  $J := |\Theta|$ .  $\square$

**Lemma 6.** Let  $\Delta_t(\theta') := L_t(\theta') - L_t(\theta^*)$  from (16),  $D_1 := \min_{\theta \in \Theta, \theta' \neq \theta^*} \sum_{i=1}^n p_{\theta}(i) (\mu_i(\theta') - \mu_i(\theta^*))^2$ , and  $\alpha(J)$  and  $\beta(J, \delta)$  be the two thresholds. Then we can show that

$$\sum_{\theta' \neq \theta^*} \mathbb{P} \left( \Delta_t(\theta') - \mathbb{E}[\Delta_t(\theta')] < D_0 \left( \frac{\beta(J, \delta)}{D_0} + \frac{\alpha(J)}{D_1} - t + \frac{tc}{2} \right) \right) \leq \exp(4) \sum_{\theta' \neq \theta^*} \exp \left( -\frac{2D_1^2 t (c/2 - 1)^2}{\eta^2} \right).$$

for some constant  $c$  such that  $0 < c < 1$ .

*Proof.* Let us recall that the critical number of samples is given by  $(1 + c)M$  where

$$M := \frac{\beta(J, \delta)}{D_0} + \frac{\alpha(J)}{D_1}. \tag{17}$$

and  $c$  is a constant. Then we can show that for some  $0 < c < 1$ ,

$$\begin{aligned}
 &\sum_{\theta' \neq \theta^*} \mathbb{P} \left( \Delta_t(\theta') - \mathbb{E}[\Delta_t(\theta')] < D_0 \left( \frac{\beta(J, \delta)}{D_0} + \frac{\alpha(J)}{D_1} - t + \frac{tc}{2} \right) \right) \\
 &\stackrel{(a)}{\leq} \sum_{\theta' \neq \theta^*} \exp \left( -\frac{2D_0^2 \left( \frac{\beta(J, \delta)}{D_0} + \frac{\alpha(J)}{D_1} + t(c/2 - 1) \right)^2}{t\eta^2} \right) \stackrel{(b)}{\leq} \sum_{\theta' \neq \theta^*} \exp \left( -\frac{2D_1^2 (M + t(c/2 - 1))^2}{t\eta^2} \right) \\
 &\stackrel{(c)}{\leq} \sum_{\theta' \neq \theta^*} \exp \left( -\frac{2D_1^2 t^2 (c/2 - 1)^2 + 4D_1^2 t M (c/2 - 1)}{t\eta^2} \right) \stackrel{(d)}{\leq} \sum_{\theta' \neq \theta^*} \exp \left( -\frac{2D_1^2 t^2 (c/2 - 1)^2 + 4\eta_0^2 M t (c/2 - 1)}{t\eta^2} \right) \\
 &\stackrel{(e)}{\leq} \sum_{\theta' \neq \theta^*} \exp \left( -\frac{2D_1^2 t^2 (c/2 - 1)^2 + 4\eta_0^2 t (c/2 - 1)}{t\eta^2} \right) \stackrel{(f)}{=} \sum_{\theta' \neq \theta^*} \exp \left( \frac{4\eta_0^2 (1 - c/2)}{\eta^2} \right) \exp \left( -\frac{2D_1^2 t (c/2 - 1)^2}{\eta^2} \right) \\
 &\stackrel{(g)}{\leq} \sum_{\theta' \neq \theta^*} \exp(4(1 - c/2)) \exp \left( -\frac{2D_1^2 t (c/2 - 1)^2}{\eta^2} \right) \leq \exp(4) \sum_{\theta' \neq \theta^*} \exp \left( -\frac{2D_1^2 t (c/2 - 1)^2}{\eta^2} \right)
 \end{aligned}$$



where (a) follows from Lemma 8 and noting that  $D_0 \left( \frac{\beta(J, \delta)}{D_0} + \frac{\alpha(J)}{D_1} - t + \frac{tc}{2} \right) < 0$  for  $t > (1+c)M$ , the inequality (b) follows from definition of  $M$  and noting that  $D_0 \geq D_1$ , (c) follows as for  $M > 1$  we can show that  $M^2 + 2tM(c/2 - 1) + t^2(c/2 - 1)^2 \geq 2tM(c/2 - 1) + t^2(c/2 - 1)^2$ , (d) follows as  $D_1 \geq \eta_0$ , (e) follows as  $M > 1$ , (f) follows as  $0 < c < 1$ , and (g) follows as  $\eta_0 \leq \eta$ .  $\square$

**Lemma 7.** Let  $\Delta_{t'}(\boldsymbol{\theta}') := L_{t'}(\boldsymbol{\theta}') - L_{t'}(\boldsymbol{\theta}^*)$  from (16),  $D_1 = \min_{\boldsymbol{\theta} \in \Theta, \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n p_{\boldsymbol{\theta}}(i) (\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2$ , and  $\alpha(J)$  be the threshold depending only on  $J$ . Then for some constant  $0 < c < 1$  and  $\eta$  defined in Definition 1 we show that

$$\sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t': t' > \frac{\alpha(J)}{D_1} + \frac{tc}{2}} \mathbb{P}(\Delta_{t'}(\boldsymbol{\theta}') - \mathbb{E}[\Delta_{t'}(\boldsymbol{\theta}')] < \alpha(J) - t'D_1) \leq \exp(4) \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \frac{\exp\left(-\frac{2D_1^2 tc}{\eta^2}\right)}{1 - \exp\left(-\frac{2D_1^2}{\eta^2}\right)}.$$

*Proof.* Let us recall that

$$\begin{aligned} & \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t': t' > \frac{\alpha(J)}{D_1} + \frac{tc}{2}} \mathbb{P}(\Delta_{t'}(\boldsymbol{\theta}') - \mathbb{E}[\Delta_{t'}(\boldsymbol{\theta}')] < \alpha(J) - t'D_1) \stackrel{(a)}{\leq} \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t': t' > \frac{\alpha(J)}{D_1} + \frac{tc}{2}} \exp\left(-\frac{2(\alpha(J) - t'D_1)^2}{t'\eta^2}\right) \\ &= \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t': t' > \frac{\alpha(J)}{D_1} + \frac{tc}{2}} \exp\left(-\frac{2D_1^2(t')^2 + 2\alpha(J)^2 - 4D_1\alpha(J)t'}{t'\eta^2}\right) \\ &\stackrel{(b)}{\leq} \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t': t' > \frac{\alpha(J)}{D_1} + \frac{tc}{2}} \exp\left(-\frac{2D_1^2(t')^2}{t'\eta^2}\right) \exp\left(\frac{4\eta\alpha(J)t' - 2\alpha(J)^2}{t'\eta^2}\right) \\ &\stackrel{(c)}{\leq} \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t': t' > \frac{\alpha(J)}{D_1} + \frac{tc}{2}} \exp(4) \exp\left(-\frac{2D_1^2 t'}{\eta^2}\right) \stackrel{(d)}{\leq} \exp(4) \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \frac{\exp\left(-2D_1^2 \left(\frac{\alpha(J)}{D_1} + \frac{tc}{2}\right) / \eta^2\right)}{1 - \exp\left(-\frac{2D_1^2}{\eta^2}\right)} \\ &= \exp(4) \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \frac{\exp\left(-\frac{2D_1\alpha(J)}{\eta^2} - \frac{D_1^2 tc}{\eta^2}\right)}{1 - \exp\left(-\frac{2D_1^2}{\eta^2}\right)} \stackrel{(e)}{\leq} \exp(4) \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \frac{\exp\left(-\frac{2\eta_0\alpha(J)}{\eta^2}\right) \exp\left(-\frac{2D_1^2 tc}{\eta^2}\right)}{1 - \exp\left(-\frac{2D_1^2}{\eta^2}\right)} \\ &\stackrel{(f)}{\leq} \exp(4) \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \frac{\exp\left(-\frac{2D_1^2 tc}{\eta^2}\right)}{1 - \exp\left(-\frac{2D_1^2}{\eta^2}\right)} \end{aligned}$$

where (a) follows from Lemma 8 and noting that  $\alpha(J) - t'D_1 < 0$  for  $t' > \frac{\alpha(J)}{D_1} + \frac{tc}{2}$ , (b) follows as  $D_1 \leq \eta$ , the inequality (c) follows as for  $\alpha(J) > 1$  we can show that

$$\frac{4\eta\alpha(J)t' - 2\alpha(J)^2}{t'\eta^2} \leq 4 \implies 4\eta\alpha(J)t' - 2\alpha(J)^2 \leq 4t'\eta^2 \implies t' \geq \frac{2\alpha(J)^2}{4\eta\alpha - 4\eta^2} = \frac{\alpha(J)}{2\eta} \left( \frac{1}{1 - \frac{\eta}{\alpha(J)}} \right)$$

which is true in this lemma as  $t' > \frac{\alpha(J)}{D_1} + \frac{tc}{2}$ ,  $D_1 \leq \eta$  and  $\alpha(J) > \eta$ . Then (d) follows by applying the infinite geometric progression formula, (e) follows as  $D_1 \geq \eta_0$ , and (f) follows as  $\exp\left(-\frac{2\eta_0\alpha(J)}{\eta^2}\right) \leq 1$ .  $\square$

### A.2.2 Concentration of $\Delta_t$

**Lemma 8.** Define  $\Delta_s(\boldsymbol{\theta}) = \ell_s(\boldsymbol{\theta}) - \ell_s(\boldsymbol{\theta}^*)$ . Let  $\epsilon > 0$  be a constant and  $\eta > 0$  is the constant defined in Assumption 1. Then we can show that,

$$\mathbb{P}(\Delta_t(\boldsymbol{\theta}) - \mathbb{E}[\Delta_t(\boldsymbol{\theta})] \leq -\epsilon) \leq \exp\left(-\frac{2\epsilon^2}{t\eta^2}\right).$$

*Proof.* Recall that  $\Delta_s(\boldsymbol{\theta}) = \ell_s(\boldsymbol{\theta}) - \ell_s(\boldsymbol{\theta}^*) = (2Y_s - \mu_{I_s}(\boldsymbol{\theta}) - \mu_{I_s}(\boldsymbol{\theta}^*))(\mu_{I_s}(\boldsymbol{\theta}^*) - \mu_{I_s}(\boldsymbol{\theta}))$ . Define  $V_s = \Delta_s(\boldsymbol{\theta}) - \mathbb{E}[\Delta_s(\boldsymbol{\theta})]$ . Note that  $\mathbb{E}[V_s] = 0$  which can be shown as follows:

$$\mathbb{E}[V_s] = \mathbb{E}[\Delta_s(\boldsymbol{\theta}) - \mathbb{E}[\Delta_s(\boldsymbol{\theta})]] = \sum_{i=1}^n \mathbb{P}(I_s = i)(\mu_i(\boldsymbol{\theta}^*) - \mu_i(\boldsymbol{\theta}))^2 - \sum_{i=1}^n \mathbb{P}(I_s = i)(\mu_i(\boldsymbol{\theta}^*) - \mu_i(\boldsymbol{\theta}))^2 = 0.$$

Also note that  $\sum_{s=1}^t V_s = \Delta_t(\boldsymbol{\theta}) - \mathbb{E}[\Delta_t(\boldsymbol{\theta})]$ . Next, we show that the moment generating function of the random variable  $V_s$  is bounded. First note that the reward  $Y_s$  is bounded between  $-\sqrt{\eta}/2$  and  $\sqrt{\eta}/2$ . It then follows that:

$$\begin{aligned} V_s &= \Delta_s(\boldsymbol{\theta}) - \mathbb{E}[\Delta_s(\boldsymbol{\theta})] \\ &= (2Y_s - \mu_{I_s}(\boldsymbol{\theta}) - \mu_{I_s}(\boldsymbol{\theta}^*))(\mu_{I_s}(\boldsymbol{\theta}^*) - \mu_{I_s}(\boldsymbol{\theta})) - \sum_{i=1}^n \mathbb{P}(I_s = i)(\mu_i(\boldsymbol{\theta}^*) - \mu_i(\boldsymbol{\theta}))^2 \leq 2\eta. \end{aligned}$$

Similarly, it can be shown that  $V_s \geq -2\eta$ . Hence, for the bounded random variable  $V_s \in [-2\eta, 2\eta]$  we can show from Hoeffding's lemma in Lemma 3 that

$$\mathbb{E}[\exp(\lambda V_s)] \leq \exp\left(\frac{\lambda^2}{8} (2\eta - (-2\eta))^2\right) = \exp(2\lambda^2\eta^2)$$

for some  $\lambda \in \mathbb{R}$ . Now for any  $\epsilon > 0$  we can show that

$$\begin{aligned} \mathbb{P}(\Delta_t(\boldsymbol{\theta}) - \mathbb{E}[\Delta_t(\boldsymbol{\theta})] \leq -\epsilon) &= \mathbb{P}\left(\sum_{s=1}^t V_s \leq -\epsilon\right) = \mathbb{P}\left(-\sum_{s=1}^t V_s \geq \epsilon\right) \\ &= \mathbb{P}\left(e^{-\lambda \sum_{s=1}^t V_s} \geq e^{\lambda\epsilon}\right) \stackrel{(a)}{\leq} e^{-\lambda\epsilon} \mathbb{E}\left[e^{-\lambda \sum_{s=1}^t V_s}\right] \\ &= e^{-\lambda\epsilon} \mathbb{E}\left[\mathbb{E}\left[e^{-\lambda \sum_{s=1}^t V_s} \mid \widehat{\boldsymbol{\theta}}(t-1)\right]\right] \\ &\stackrel{(b)}{=} e^{-\lambda\epsilon} \mathbb{E}\left[\mathbb{E}\left[e^{-\lambda V_t} \mid \widehat{\boldsymbol{\theta}}(t-1)\right] \mathbb{E}\left[e^{-\lambda \sum_{s=1}^{t-1} V_s} \mid \widehat{\boldsymbol{\theta}}(t-1)\right]\right] \\ &\leq e^{-\lambda\epsilon} \mathbb{E}\left[\exp(2\lambda^2\eta^2) \mathbb{E}\left[e^{-\lambda \sum_{s=1}^{t-1} V_s} \mid \widehat{\boldsymbol{\theta}}(t-1)\right]\right] \\ &= e^{-\lambda\epsilon} e^{2\lambda^2\eta^2} \mathbb{E}\left[e^{-\lambda \sum_{s=1}^{t-1} V_s}\right] \\ &\vdots \\ &\stackrel{(c)}{\leq} e^{-\lambda\epsilon} e^{2\lambda^2 t\eta^2} \\ &\stackrel{(d)}{\leq} \exp\left(-\frac{2\epsilon^2}{t\eta^2}\right) \end{aligned}$$

where (a) follows by Markov's inequality, (b) follows as  $V_s$  is conditionally independent given  $\widehat{\boldsymbol{\theta}}(s-1)$ , (c) follows by unpacking the term for  $t$  times and (d) follows by taking  $\lambda = \epsilon/4t\eta^2$ .  $\square$

### A.2.3 Proof of correctness for General Sub-Gaussian Case

**Lemma 9.** Let  $L_t(\boldsymbol{\theta})$  be the sum of squared errors of the hypothesis parameterized by  $\boldsymbol{\theta}$  based on observation vector  $\mathbf{Y}^t$  from an underlying sub-Gaussian distribution. Let  $\tau_{\boldsymbol{\theta}^*, \boldsymbol{\theta}} := \min\{t : L_t(\boldsymbol{\theta}^*) - L_t(\boldsymbol{\theta}) > \beta(J, \delta)\}$ . Then we can show that

$$\mathbb{P}(L_{\tau_{\boldsymbol{\theta}^*, \boldsymbol{\theta}}}(\boldsymbol{\theta}^*) - L_{\tau_{\boldsymbol{\theta}^*, \boldsymbol{\theta}}}(\boldsymbol{\theta}) > \beta(J, \delta)) \leq \frac{\delta}{J}$$

where,  $\beta(J, \delta) := \log \left( \frac{(1 + \eta^2/\eta_0^2) J}{\delta} \right)$ .

*Proof.* Let  $-\Delta_t(\boldsymbol{\theta}) := L_t(\boldsymbol{\theta}^*) - L_t(\boldsymbol{\theta})$  be the difference of sum of squared errors between hypotheses parameterized by  $\boldsymbol{\theta}^*$  and  $\boldsymbol{\theta}$ . We again define  $\tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}}(Y^t) := \min\{t : -\Delta_t(\boldsymbol{\theta}) > \beta(J, \delta)\}$ . For brevity in the following proof we drop the  $Y^t$  in  $\tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}}(Y^t)$ . Then we can show that

$$\begin{aligned}
 \mathbb{P}(\exists t < \infty, -\Delta_t(\boldsymbol{\theta}) > \beta(J, \delta)) &= \sum_{t=1}^{\infty} \mathbb{P}(\tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}} = t, -\Delta_t(\boldsymbol{\theta}) > \beta(J, \delta)) \\
 &= \sum_{t=1}^{\infty} \mathbb{P}(\tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}} = t, -\Delta_t(\boldsymbol{\theta}) - \mathbb{E}[-\Delta_t(\boldsymbol{\theta})] > \beta(J, \delta) - \mathbb{E}[-\Delta_t(\boldsymbol{\theta})]) \\
 &\leq \sum_{t=1}^{\infty} \mathbb{P}(\tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}} = t, -\Delta_t(\boldsymbol{\theta}) - \mathbb{E}[-\Delta_t(\boldsymbol{\theta})] > \beta(J, \delta) + tD_1) \\
 &\stackrel{(a)}{\leq} \sum_{t=1}^{\infty} \exp\left(-\frac{2(\beta(J, \delta) + tD_1)^2}{t\eta^2}\right) \stackrel{(b)}{\leq} \sum_{t=1}^{\infty} \exp\left(-\left(\beta(J, \delta) + \frac{t^2 D_1^2}{t\eta^2}\right)\right) \\
 &= \sum_{t=1}^{\infty} \exp\left(-\left(\beta(J, \delta) + \frac{tD_1^2}{\eta^2}\right)\right) = \exp(-\beta(J, \delta)) [1 + \exp(-D_1^2/\eta^2) + \exp(-2D_1^2/\eta^2) + \exp(-3D_1^2/\eta^2) + \dots] \\
 &\stackrel{(c)}{=} \exp(-\beta(J, \delta)) \frac{1}{1 - \exp(-D_1^2/\eta^2)} \stackrel{(d)}{\leq} \exp(-\beta(J, \delta)) \left(1 + \frac{\eta^2}{D_1^2}\right) \stackrel{(e)}{\leq} \frac{\delta}{J}.
 \end{aligned}$$

where, (a) follows from Lemma 8 and noting that  $-\mathbb{E}[-\Delta_t(\boldsymbol{\theta})] \leq -tD_0$ , (b) follows as  $(\beta(J, \delta) + tD_0)^2 \geq 2\beta(J, \delta) + t^2 D_0^2$  for  $a, b > 0$ , (c) follows from the infinite geometric series sum formula, (d) follows as  $1/(1 - \exp(-x)) \leq 1 + 1/x$  for  $x > 0$ , and (e) follows as  $\beta(J, \delta) := \log \left( \frac{(1 + \eta^2/\eta_0^2) J}{\delta} \right)$  and noting that  $D_1 \geq \eta_0$ .  $\square$

#### A.2.4 Stopping time Correctness Lemma for the Gaussian Case

**Lemma 10.** *Let  $L_t(\boldsymbol{\theta})$  be the sum of squared errors of the hypothesis parameterized by  $\boldsymbol{\theta}$  based on observation vector  $\mathbf{Y}^t$  from an underlying Gaussian distribution. Let  $\tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}} := \min\{t : L_t(\boldsymbol{\theta}^*) - L_t(\boldsymbol{\theta}) > \beta(J, \delta)\}$ . Then we can show that*

$$\mathbb{P}(L_{\tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}}}(\boldsymbol{\theta}^*) - L_{\tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}}}(\boldsymbol{\theta}) > \beta(J, \delta)) \leq \frac{\delta}{J}$$

where we define the threshold function as,

$$\beta(J, \delta) := \log(J/\delta) \tag{18}$$

*Proof.* Let  $-\Delta_t(\boldsymbol{\theta}) := L_t(\boldsymbol{\theta}^*) - L_t(\boldsymbol{\theta})$  be the log-likelihood ratio between hypotheses parameterized by  $\boldsymbol{\theta}^*$  and  $\boldsymbol{\theta}$ . Define  $\tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}}(Y^t) := \min\{t : -\Delta_t(\boldsymbol{\theta}) > \beta(J, \delta)\}$ . For brevity in the following proof we drop the  $Y^t$  in  $\tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}}(Y^t)$ . Then we can show that at time  $t \geq \tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}}$  we have

$$\begin{aligned}
 -\Delta_t(\boldsymbol{\theta}) > \beta(J, \delta) &\implies \exp(-\Delta_t(\boldsymbol{\theta})) > \exp(\beta(J, \delta)) \\
 &\implies \left( \frac{\prod_{s=1}^t \mathbb{P}(Y_{I_s} = y_s | I_s, \boldsymbol{\theta})}{\prod_{s=1}^t \mathbb{P}(Y_{I_s} = y_s | I_s, \boldsymbol{\theta}^*)} \right) > \exp(\beta(J, \delta)) \\
 &\implies \exp(-\beta(J, \delta)) \left( \frac{\prod_{s=1}^t \mathbb{P}(Y_{I_s} = y_s | I_s, \boldsymbol{\theta})}{\prod_{s=1}^t \mathbb{P}(Y_{I_s} = y_s | I_s, \boldsymbol{\theta}^*)} \right) > 1.
 \end{aligned} \tag{19}$$

Following this we can show that the probability of the event  $\{-\Delta_t(\boldsymbol{\theta}) > \beta(J, \delta)\}$  is upper bounded by

$$\begin{aligned}
 \mathbb{P}(\exists t < \infty, -\Delta_t(\boldsymbol{\theta}) > \beta(J, \delta)) &= \sum_{t=1}^{\infty} \mathbb{P}(\tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}} = t) = \sum_{t=1}^{\infty} \mathbb{E}[\mathbb{I}\{\tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}} = t\}] \\
 &\stackrel{(a)}{\leq} \sum_{t=1}^{\infty} \mathbb{E} \left[ \mathbb{I}\{\tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}} = t\} \exp(-\beta(J, \delta)) \left( \frac{\prod_{s=1}^t \mathbb{P}(Y_{I_s} = y_s | I_s, \boldsymbol{\theta})}{\prod_{s=1}^t \mathbb{P}(Y_{I_s} = y_s | I_s, \boldsymbol{\theta}^*)} \right) \right] \\
 &= \exp(-\beta(J, \delta)) \sum_{t=1}^{\infty} \int_{\mathbb{R}^t} \mathbb{I}\{\tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}} = t\} \frac{\prod_{s=1}^t \mathbb{P}(Y_{I_s} = y_s | I_s, \boldsymbol{\theta})}{\prod_{s=1}^t \mathbb{P}(Y_{I_s} = y_s | I_s, \boldsymbol{\theta}^*)} \prod_{s=1}^t \mathbb{P}(Y_{I_s} = y_s | I_s, \boldsymbol{\theta}^*) dy_1 dy_2 \dots dy_t \\
 &= \exp(-\beta(J, \delta)) \sum_{t=1}^{\infty} \int_{\mathbb{R}^t} \mathbb{I}\{\tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}} = t\} \prod_{s=1}^t \mathbb{P}(Y_{I_s} = y_s | I_s, \boldsymbol{\theta}) dy_1 dy_2 \dots dy_t \\
 &= \exp(-\beta(J, \delta)) \sum_{t=1}^{\infty} \mathbb{P}(\tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}} = t | I^t, \boldsymbol{\theta}) \leq \exp(-\beta(J, \delta)) \stackrel{(b)}{\leq} \frac{\delta}{J}.
 \end{aligned}$$

where, (a) follows from (19), and (b) follows from (18). The claim of the lemma follows.  $\square$

### A.2.5 Proof of CS Sample Complexity (Theorem 1)

**Theorem 1. (Restatement)** Let  $\tau_\delta$  denote the stopping time of CS in Algorithm 1. Let  $D_0$  be the objective value of the max min optimization in (4) when  $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ , i.e.,

$$D_0 := \max_{\mathbf{p}} \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n p(i) (\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2.$$

Denoting  $\mathbf{p}_\boldsymbol{\theta}$  as the solution of (4) when  $\widehat{\boldsymbol{\theta}}(t)$  is replaced by any  $\boldsymbol{\theta} \in \Theta$ , let  $D_1$  be the minimum possible objective value of (4) over all  $\mathbf{p}_\boldsymbol{\theta}$  when  $\widehat{\boldsymbol{\theta}}(t)$  is replaced by  $\boldsymbol{\theta}^*$ , i.e.,

$$D_1 := \min_{\{\mathbf{p}_\boldsymbol{\theta} : \boldsymbol{\theta} \in \Theta\}} \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n p_{\boldsymbol{\theta}}(i) (\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2.$$

Assumption 2 ensures that  $D_1 > 0$ . The sample complexity of the  $\delta$ -PAC CS has the following upper bound, where  $J := |\Theta|$ ,  $C = O((\eta/\eta_0)^2)$  is a constant:

$$\mathbb{E}[\tau_\delta] \leq O \left( \frac{\eta \log(C) \log J}{D_1} + \frac{\log(J/\delta)}{D_0} + JC^{\frac{1}{\eta}} \delta^{\frac{D_0}{\eta^2}} \right).$$

*Proof. Step 1 (Definitions):* Define  $L_t(\boldsymbol{\theta})$  as the total sum of squared errors of hypothesis  $\boldsymbol{\theta}$  till round  $t$ . Let,  $-\Delta_t(\boldsymbol{\theta}) := L_t(\boldsymbol{\theta}^*) - L_t(\boldsymbol{\theta})$  be the difference of squared errors between  $\boldsymbol{\theta}^*$  and  $\boldsymbol{\theta}$ . Note that the p.m.f.  $\mathbf{p}_\boldsymbol{\theta}$  is the Chernoff verification proportion for verifying hypothesis  $\boldsymbol{\theta}$ .

**Step 2 (Define  $\tau_\delta$  and partition):** We define the stopping time  $\tau_\delta$  for the policy  $\pi$  as follows:

$$\tau_\delta := \min\{t : \exists \boldsymbol{\theta} \in \Theta, L_t(\boldsymbol{\theta}') - L_t(\boldsymbol{\theta}) > \beta(J, \delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}\} \quad (20)$$

where,  $\beta(J, \delta)$  is the threshold function.

**Step 3 (Define bad event):** We define the bad event  $\xi^\delta(\boldsymbol{\theta})$  for the sub-optimal hypothesis  $\boldsymbol{\theta} \neq \boldsymbol{\theta}^*$  as follows:

$$\xi^\delta(\boldsymbol{\theta}) = \{\widehat{\boldsymbol{\theta}}(\tau_\delta) = \boldsymbol{\theta}, L_{\tau_\delta}(\boldsymbol{\theta}') - L_{\tau_\delta}(\boldsymbol{\theta}) > \beta(J, \delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}\}. \quad (21)$$

The event  $\xi^\delta(\boldsymbol{\theta})$  denotes that a sub-optimal hypothesis  $\boldsymbol{\theta}$  is declared the optimal hypothesis when it has a smaller sum of squared errors than any other hypothesis  $\boldsymbol{\theta}'$  at  $\tau_\delta$ .

**Step 4 (Decomposition of bad event):** In this step we decompose the bad event to show that only comparing  $\boldsymbol{\theta}$  against  $\boldsymbol{\theta}^*$  is enough to guarantee a  $\delta$ -PAC policy. First we decompose the bad event  $\xi^\delta(\boldsymbol{\theta})$  as follows:

$$\begin{aligned} \xi^\delta(\boldsymbol{\theta}) &= \{\widehat{\boldsymbol{\theta}}(\tau_\delta) = \boldsymbol{\theta}, L_{\tau_\delta}(\boldsymbol{\theta}') - L_{\tau_\delta}(\boldsymbol{\theta}) > \beta(J, \delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}\} \\ &\stackrel{(a)}{=} \underbrace{\{\widehat{\boldsymbol{\theta}}(\tau_\delta) = \boldsymbol{\theta}, L_{\tau_\delta}(\boldsymbol{\theta}') - L_{\tau_\delta}(\boldsymbol{\theta}) > \beta(J, \delta), \forall \boldsymbol{\theta}' \in \Theta \setminus \{\boldsymbol{\theta}^*\}\}}_{\text{part A}} \\ &\quad \cap \underbrace{\{\widehat{\boldsymbol{\theta}}(\tau_\delta) = \boldsymbol{\theta}, L_{\tau_\delta}(\boldsymbol{\theta}^*) - L_{\tau_\delta}(\boldsymbol{\theta}) > \beta(J, \delta)\}}_{\text{part B}} \\ &\subseteq \{\widehat{\boldsymbol{\theta}}(\tau_\delta) = \boldsymbol{\theta}, L_{\tau_\delta}(\boldsymbol{\theta}^*) - L_{\tau_\delta}(\boldsymbol{\theta}) > \beta(J, \delta)\} \end{aligned} \quad (22)$$

where, (a) follows by decomposing the event in two parts containing  $\boldsymbol{\theta} \in \Theta \setminus \{\boldsymbol{\theta}^*\}$  and  $\{\boldsymbol{\theta}^*\}$ , (b) follows by noting that the intersection of events holds by taking into account only the event in part B.

**Step 5 (Proof of correctness):** In this step we want to show that based on the  $\tau_\delta$  definition and the bad event  $\xi^\delta(\boldsymbol{\theta})$  the CS stops and outputs the correct hypothesis  $\boldsymbol{\theta}^*$  with  $1 - \delta$  probability. As shown in Step 4, we can define the error event  $\xi^\delta(\boldsymbol{\theta})$  as follows:

$$\xi^\delta(\boldsymbol{\theta}) \subseteq \{\widehat{\boldsymbol{\theta}}(\tau_\delta) = \boldsymbol{\theta}, L_{\tau_\delta}(\boldsymbol{\theta}^*) - L_{\tau_\delta}(\boldsymbol{\theta}) > \beta(J, \delta)\}$$

Define  $\tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}} = \min\{t : -\Delta_t(\boldsymbol{\theta}) > \beta(J, \delta)\}$ . Then we can show for the stopping time  $\tau_\delta$ , the round  $\tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}}$  from Lemma 9 and the threshold  $\beta(J, \delta) := \log\left(\frac{(1 + \eta^2/\eta_0^2)J}{\delta}\right)$  we have

$$\begin{aligned} \mathbb{P}(\tau_\delta < \infty, \widehat{\boldsymbol{\theta}}(\tau_\delta) \neq \boldsymbol{\theta}^*) &\leq \mathbb{P}(\exists \boldsymbol{\theta} \in \Theta \setminus \{\boldsymbol{\theta}^*\}, \exists t \in \mathbb{N} : -\Delta_t(\boldsymbol{\theta}) > \beta(J, \delta)) \\ &\leq \sum_{\boldsymbol{\theta} \neq \boldsymbol{\theta}^*} \mathbb{P}(L_{\tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}}}(\boldsymbol{\theta}^*) - L_{\tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}}}(\boldsymbol{\theta}) > \beta(J, \delta), \tau_{\boldsymbol{\theta}^* \boldsymbol{\theta}} < \infty) \stackrel{(a)}{\leq} \sum_{\boldsymbol{\theta} \neq \boldsymbol{\theta}^*} \frac{\delta}{J} \leq \delta \end{aligned}$$

where, (a) follows from Lemma 9.

**Step 6 (Sample complexity analysis):** In this step we bound the total sample complexity satisfying the  $\delta$ -PAC criteria. We define the stopping time  $\tau_\delta$  as follows:

$$\tau_\delta := \min\left\{t : L_t(\boldsymbol{\theta}') - L_t(\widehat{\boldsymbol{\theta}}(t)) > \beta(J, \delta), \forall \boldsymbol{\theta}' \neq \widehat{\boldsymbol{\theta}}(t)\right\}$$

We further define the time  $\tau_{\boldsymbol{\theta}^*}$  for the hypothesis  $\boldsymbol{\theta}^*$  as follows:

$$\tau_{\boldsymbol{\theta}^*} := \min\{t : L_t(\boldsymbol{\theta}') - L_t(\boldsymbol{\theta}^*) > \beta(J, \delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*\}. \quad (23)$$

We also define the critical number of samples as  $(1 + c)M$  where  $M$  is defined as follows:

$$M := \left(\frac{\alpha(J)}{D_1} + \frac{C' + \log(J/\delta)}{D_0}\right) \quad (24)$$

where,  $C' = \log(1 + \eta^2/\eta_0^2)$ . Hence  $C' + \log(J/\delta)$  follows from the definition of  $\beta(J, \delta)$  in Lemma 9. We define the term  $D_1$  as follows:

$$D_1 := \min_{\{\mathbf{p}_\theta : \boldsymbol{\theta} \in \Theta\}} \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n p_{\boldsymbol{\theta}}(i) (\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2 \quad (25)$$

and the term  $D_0$  as follows:

$$D_0 := \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n p_{\boldsymbol{\theta}^*}(i) (\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2 \quad (26)$$

It then follows that

$$\begin{aligned}
 \mathbb{E}[\tau_\delta] &\leq \mathbb{E}[\tau_{\theta^*}] = \sum_{t=0}^{\infty} \mathbb{P}(\tau_{\theta^*} = t) \stackrel{(a)}{\leq} 1 + (1+c)M + \sum_{t:t > (1+c)M} \mathbb{P}(\tau_{\theta^*} > t) \\
 &\stackrel{(b)}{\leq} 1 + (1+c) \left( \frac{\alpha(J)}{D_1} + \frac{C' + 4\log(J/\delta)}{D_0} \right) + J \sum_{t:t > \left( \frac{\alpha(J)}{D_1} + \frac{\log(J/\delta)}{D_0} \right) (1+c)} C_1 \exp(-C_2 t) \\
 &\stackrel{(c)}{\leq} 1 + (1+c) \left( \frac{\alpha(J)}{D_1} + \frac{C' + 4\log(J/\delta)}{D_0} \right) + JC_1 \frac{\exp\left(-C_2(1+c) \left( \frac{\alpha(J)}{D_1} + \frac{4\log(J/\delta)}{D_0} \right)\right)}{1 - \exp(-C_2)} \\
 &\stackrel{(d)}{\leq} 1 + (1+c) \left( \frac{\alpha(J)}{D_1} + \frac{C' + 4\log(J/\delta)}{D_0} \right) + JC_1 \frac{\exp\left(-C_2(1+c) \left( \frac{\alpha(J)}{D_0} + \frac{4\log(J)}{D_0} + \frac{4\log(1/\delta)}{D_0} \right)\right)}{1 - \exp(-C_2)} \\
 &\stackrel{(e)}{\leq} 1 + (1+c) \left( \frac{\alpha(J)}{D_1} + \frac{C' + \log(J/\delta)}{D_0} \right) + JC_1 \frac{\exp\left(-C_2 \left( \frac{\alpha(J) + 4\log J}{D_0} \right) \exp\left(-4C_2 \frac{\log(1/\delta)}{D_0}\right)\right)}{1 - \exp(-C_2)} \\
 &\stackrel{(f)}{\leq} 1 + (1+c) \left( \frac{b\log J}{D_1} + \frac{C' + \log(J/\delta)}{D_0} \right) + JC_1 \frac{\exp\left(-C_2 \left( \frac{(b+4)\log J}{D_0} \right) \exp\left(-4C_2 \frac{\log(1/\delta)}{D_0}\right)\right)}{1 - \exp(-C_2)} \\
 &\leq 1 + (1+c) \left( \frac{b\log J}{D_1} + \frac{C' + \log(J/\delta)}{D_0} \right) + JC_1 \left( \frac{1}{J} \right)^{C_2(b+4)/D_0} \delta^{4C_2/D_0} \left( 1 + \max\{1, \frac{1}{C_2}\} \right) \\
 &\leq 1 + (1+c) \left( \frac{b\log J}{D_1} + \frac{C' + \log(J/\delta)}{D_0} \right) + C_1 \left( 2 + \frac{1}{C_2} \right) J^{1 - \frac{C_2(b+4)}{D_0}} \frac{4C_2}{\delta D_0} \\
 &\stackrel{(g)}{\leq} 1 + 2 \left( \frac{b\log J}{D_1} + \frac{C' + \log(J/\delta)}{D_0} \right) + \left( 165 + \frac{\eta^2}{D_1^2} \right) \left( 2 + \frac{\eta^2}{D_1^2} \right) J^{1 - \frac{D_1^2(b+4)}{2\eta^2 D_0}} \frac{D_1^2}{\delta \eta^2 D_0} \\
 &\leq 1 + 2 \left( \frac{b\log J}{D_1} + \frac{C' + \log(J/\delta)}{D_0} \right) + \left( 165 + \frac{\eta^2}{D_1^2} \right)^2 J^{1 - \frac{D_1^2(b+4)}{2\eta^2 D_0}} \frac{D_0}{\delta \eta^2} \\
 &\stackrel{(h)}{\leq} 1 + 2 \underbrace{\left( \frac{b\log J}{D_1} + \frac{C' + \log(J/\delta)}{D_0} \right)}_{\text{Term A}} + \underbrace{\left( 165 + \frac{\eta^2}{\eta_0^2} \right)^2 J^{1 - \frac{\eta_0^2(b+4)}{2\eta^3}}}_{\text{Term B}} \times \underbrace{\left( \frac{D_0}{\delta \eta^2} \right)}_{\text{Term C}} \tag{27}
 \end{aligned}$$

where, (a) follows from definition of  $M$  in (24), (b) follows from Lemma 5,  $C_1 := 110 + 55 \max\left\{1, \frac{\eta^2}{2D_1^2}\right\}$ ,  $C_2 := \frac{2D_1^2 \min\{(c/2 - 1)^2, c\}}{\eta^2}$ , (c) follows by applying the geometric progression formula, (d) follows as  $D_1 \leq D_0$ . The inequality (e) follows as  $c > 0$ , (f) follows by setting  $\alpha(J) = b\log J$  for some constant  $b > 1$ , (g) follows by setting  $c = \frac{1}{2}$  in  $C_1$  and  $C_2$ , and (h) follows as  $D_1 \geq \eta_0$ , and  $D_0 \leq \eta$ .

Now, note that in (27) the Term C  $\leq 1$  as  $\delta \in (0, 1)$ . Now for the Term B we need to find an  $b$  such that Term B  $\leq J^{1 - \frac{\eta_0^2(b+4)}{2\eta^3}} + \frac{b}{\eta \log J}$ . Hence,

$$\left( 165 + \frac{\eta^2}{\eta_0^2} \right)^2 J^{1 - \frac{\eta_0^2(b+4)}{2\eta^3}} \leq J^{1 - \frac{\eta_0^2(b+4)}{2\eta^3}} + \frac{b}{\eta \log J} \implies \left( 165 + \frac{\eta^2}{\eta_0^2} \right)^2 \leq J \frac{b}{\eta \log J} \implies \eta \log \left( 165 + \frac{\eta^2}{\eta_0^2} \right) \leq b$$

So for a constant  $b > 1$  such that if  $b$  satisfies the following condition

$$b = \eta \log \left( 165 + \frac{\eta^2}{\eta_0^2} \right) > \log \left( 1 + \frac{\eta^2}{\eta_0^2} \right) \quad (28)$$

then we have that Term B  $\leq J^{1 - \frac{\eta_0^2(b+4)}{2\eta^3} + \frac{b}{\eta \log J}}$ . Hence we set the value of  $C = 165 + \eta^2/\eta_0^2 > C'$  which shows up in our theorem statement. Plugging this in (27) we get that the expected sample complexity is upper bounded by

$$\begin{aligned} \mathbb{E}[\tau_\delta] &\leq 1 + 2 \left( \frac{\eta \log \left( 165 + \frac{\eta^2}{\eta_0^2} \right) \log J}{D_1} + \frac{\log \left( 1 + \frac{\eta^2}{\eta_0^2} \right) + \log(J/\delta)}{D_0} \right) + J^{1 - \frac{\eta_0^2(b+4)}{2\eta^3} + \frac{b}{\eta \log J}} \frac{D_0}{\delta \eta^2} \\ &\stackrel{(a)}{\leq} 1 + 2 \left( \frac{\eta \log \left( 165 + \frac{8\eta^2}{\eta_0^2} \right) \log J}{D_1} + \frac{\log \left( \left( 165 + \frac{\eta^2}{\eta_0^2} \right) \frac{J}{\delta} \right)}{D_0} \right) + J^{1 + \frac{\log \left( 165 + \frac{\eta^2}{\eta_0^2} \right)}{\eta \log J}} \frac{D_0}{\delta \eta^2} \\ &\stackrel{(b)}{\leq} 1 + 2 \left( \frac{\eta \log(C) \log J}{D_1} + \frac{\log(CJ/\delta)}{D_0} \right) + J^{1 + \frac{\log(C)}{\eta \log J}} \frac{D_0}{\delta \eta^2} \\ &= 1 + 2 \left( \frac{\eta \log(C) \log J}{D_1} + \frac{\log(CJ/\delta)}{D_0} \right) + J \cdot J^{\frac{\log(C)^{1/\eta}}{\log J}} \frac{D_0}{\delta \eta^2} \\ &= O \left( \frac{\eta \log(C) \log J}{D_1} + \frac{\log(J/\delta)}{D_0} + J(C)^{1/\eta} \delta^{D_0/\eta^2} \right) \end{aligned}$$

where, (a) follows as  $2 \log \left( 165 + \frac{\eta^2}{\eta_0^2} \right) \geq \log \left( 1 + \frac{\eta^2}{\eta_0^2} \right)$ , and in (b) we substitute  $C = \left( 165 + \frac{\eta^2}{\eta_0^2} \right)$ . The claim of the Theorem follows.  $\square$

### A.3 Proof of T2 Sample Complexity

#### A.3.1 Concentration Lemma

This section contains concentration lemma equivalent to the Lemma 5 of Appendix A.2.

**Lemma 11.** Define  $L_t(\boldsymbol{\theta}^*)$  as the sum of squared error of the hypothesis parameterized by  $\boldsymbol{\theta}^*$ . Let  $\tau_{\boldsymbol{\theta}^*} = \min\{t : L_t(\boldsymbol{\theta}') - L_t(\boldsymbol{\theta}^*) > \beta(J, \delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*\}$ . Then we can bound the probability of the event

$$\mathbb{P}(\tau_{\boldsymbol{\theta}^*} > t) \leq J C'_1 \exp(-C'_2 t)$$

where,  $J := |\Theta|$ ,  $C'_1 := 110 + 55 \max \left\{ 1, \frac{\eta^2}{2D_1'^2} \right\}$ ,  $C'_2 := \frac{2D_1'^2 \min\{(c-1)^2, c\}}{\eta^2}$ ,  $\eta > 0$  defined in Definition 1 and  $D_1' := \min_{\boldsymbol{\theta} \neq \boldsymbol{\theta}', \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n u_{\boldsymbol{\theta}\boldsymbol{\theta}'}(i) (\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2$ .

*Proof.* We define the event

$$\begin{aligned} \xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t) &= \{L_t(\boldsymbol{\theta}') - L_t(\boldsymbol{\theta}^*) < \beta(J, \delta)\} \\ \tilde{\xi}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t) &= \{L_t(\boldsymbol{\theta}') - L_t(\boldsymbol{\theta}^*) < \alpha(J)\} \end{aligned}$$

Then we define the time  $\tau_{\boldsymbol{\theta}^*}$  as follows:

$$\tau_{\boldsymbol{\theta}^*} = \min\{t : L_t(\boldsymbol{\theta}') - L_t(\boldsymbol{\theta}^*) > \beta(J, \delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*\}$$

which is the first round  $L_t(\boldsymbol{\theta}')$  crosses  $\beta(J, \delta)$  threshold against  $L_t(\boldsymbol{\theta}^*)$  for all  $\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*$ . We also define the time  $\tilde{\tau}_{\boldsymbol{\theta}^*}$  as follows:

$$\tilde{\tau}_{\boldsymbol{\theta}^*} = \min\{t : L_{t'}(\boldsymbol{\theta}') - L_{t'}(\boldsymbol{\theta}^*) > \alpha(J), \forall t' > t\}$$

which is the first round when  $L_t(\boldsymbol{\theta}^*)$  crosses  $\alpha(J)$  threshold against  $L_t(\boldsymbol{\theta}')$ . Then we define the time

$$\tilde{\tau}_{\boldsymbol{\theta}^*} = \max_k \{\tilde{\tau}_{\boldsymbol{\theta}^* \boldsymbol{\theta}'}\}$$

as the last time  $\tilde{\tau}_{\boldsymbol{\theta}^* \boldsymbol{\theta}'}$  happens. Define the term  $D'_0$  and  $D'_1$  as

$$D'_0 := \min_{\boldsymbol{\theta}, \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n u_{\boldsymbol{\theta}^* \boldsymbol{\theta}}(i) (\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2, \quad D'_1 := \min_{\boldsymbol{\theta} \neq \boldsymbol{\theta}', \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n u_{\boldsymbol{\theta} \boldsymbol{\theta}'}(i) (\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2$$

Let  $\Delta_t(\boldsymbol{\theta}') := L_t(\boldsymbol{\theta}') - L_t(\boldsymbol{\theta}^*)$  be the sum of squared errors between hypotheses parameterized by  $\boldsymbol{\theta}'$  and  $\boldsymbol{\theta}^*$ . Then it follows that,

$$\begin{aligned} \mathbb{P}(\tilde{\xi}_{\boldsymbol{\theta}' \boldsymbol{\theta}^*}(t)) &= \mathbb{P}(\Delta_t(\boldsymbol{\theta}') - tD'_1 < \alpha(J) - tD'_1) \\ &\leq \mathbb{P}(\Delta_t(\boldsymbol{\theta}') - \mathbb{E}[\Delta_t(\boldsymbol{\theta}')] < \alpha(J) - tD'_1) \end{aligned}$$

Similarly, we can show that,

$$\begin{aligned} \mathbb{P}(\xi_{\boldsymbol{\theta}' \boldsymbol{\theta}^*}(t)) &= \mathbb{P}(\Delta_t(\boldsymbol{\theta}') - \mathbb{E}[\Delta_t(\boldsymbol{\theta}')] < \beta(J, \delta) - \mathbb{E}[\Delta_t(\boldsymbol{\theta}')] \\ &\leq \mathbb{P}(\Delta_t(\boldsymbol{\theta}') - \mathbb{E}[\Delta_t(\boldsymbol{\theta}')] < \beta(J, \delta) - (t - \tilde{\tau}_{\boldsymbol{\theta}^*})D'_0) \\ &= \mathbb{P}\left(\Delta_t(\boldsymbol{\theta}') - \mathbb{E}[\Delta_t(\boldsymbol{\theta}')] < D'_0 \left(\frac{\beta(J, \delta)}{D'_0} - t + \tilde{\tau}_{\boldsymbol{\theta}^*}\right)\right) \end{aligned}$$

Then following the same approach as in Lemma 5 we can show that,

$$\begin{aligned} \mathbb{P}(\tau_{\boldsymbol{\theta}^*} > t) &\leq \mathbb{P}\left(\bigcup_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \xi_{\boldsymbol{\theta}' \boldsymbol{\theta}^*}(t)\right) \\ &\leq \mathbb{P}\left(\left\{\bigcup_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \xi_{\boldsymbol{\theta}' \boldsymbol{\theta}^*}(t)\right\} \cap \{\tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D'_1} + tc\}\right) + \mathbb{P}\left(\{\xi_{\boldsymbol{\theta}' \boldsymbol{\theta}^*}(t)\} \cap \{\tilde{\tau}_{\boldsymbol{\theta}^*} \geq \frac{\alpha(J)}{D'_1} + tc\}\right) \\ &\leq \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \mathbb{P}\left(\{\xi_{\boldsymbol{\theta}' \boldsymbol{\theta}^*}(t)\} \cap \{\tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D'_1} + tc\}\right) + \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t': t' \geq \frac{\alpha(J)}{D'_1} + tc} \mathbb{P}(\tilde{\xi}_{\boldsymbol{\theta}' \boldsymbol{\theta}^*}(t')) \\ &\leq \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \mathbb{P}\left(\{\xi_{\boldsymbol{\theta}' \boldsymbol{\theta}^*}\} \cap \{\tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D'_1} + tc\}\right) + \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t': t' \geq \frac{\alpha(J)}{D'_1} + tc} \mathbb{P}(\Delta_{t'}(\boldsymbol{\theta}') - t'D'_1 < \alpha(J) - t'D'_1) \\ &\leq \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \mathbb{P}\left(\Delta_{t'}(\boldsymbol{\theta}') - \mathbb{E}[\Delta_{t'}(\boldsymbol{\theta}')] < D'_0 \left(\frac{\beta(J, \delta)}{D'_0} + \frac{\alpha(J)}{D'_1} - t + tc\right)\right) \\ &\quad + \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t': t' \geq \frac{\alpha(J)}{D'_1} + tc} \mathbb{P}(\Delta_{t'}(\boldsymbol{\theta}') - t'D'_1 < \alpha(J) - t'D'_1) \\ &\stackrel{(a)}{\leq} \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp(4) \exp\left(-\frac{2D_1'^2 t (\frac{c}{2} - 1)^2}{\eta^2}\right) + \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp(4) \frac{\exp\left(-\frac{2D_1'^2 tc}{\eta^2}\right)}{1 - \exp\left(-\frac{2D_1'^2}{\eta^2}\right)} \\ &\stackrel{(b)}{\leq} 55 \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_1'^2 t (\frac{c}{2} - 1)^2}{\eta^2}\right) + 55 \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_1'^2 tc}{\eta^2}\right) + 55 \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \exp\left(-\frac{2D_1'^2 tc}{\eta^2}\right) \max\left\{1, \frac{\eta^2}{2D_1'^2}\right\} \\ &\stackrel{(c)}{\leq} JC'_1 \exp(-C'_2 t) \end{aligned}$$



where, (a) follows from Lemma 6 and Lemma 7 as their result holds for any  $D'_1, D'_0 > 0$ , (b) follows from the same steps as in Lemma 5 as  $D'_1 \leq D'_0$ ,  $D'_1 > 0$ ,  $D'_0 > 0$ , and in (c) we substitute  $C'_1 := 110 + 55 \max\left\{1, \frac{\eta^2}{2D_1^2}\right\}$ ,  $C'_2 := \frac{2D_1^2 \min\{(\frac{c}{2} - 1)^2, c\}}{\eta^2}$ , and  $J := |\Theta|$ .  $\square$

### A.3.2 Proof of T2 Sample Complexity (Proposition 1)

**Proposition 1. (Restatement)** Let  $\tau_\delta$  denote the stopping time of T2 stops sampling following the sampling strategy of (7). Consider the set  $\mathcal{I}(\theta, \theta') \subset [n]$  of actions that could be sampled following (7) when  $\hat{\theta}(t) = \theta$  and  $\tilde{\theta}(t) = \theta'$ , and let  $\mathbf{u}_{\theta\theta'}$  denote a uniform pmf supported on  $\mathcal{I}(\theta, \theta')$ . Define

$$D'_0 := \min_{\theta, \theta' \neq \theta^*} \sum_{i=1}^n u_{\theta^*\theta}(i) (\mu_i(\theta') - \mu_i(\theta^*))^2,$$

$$D'_1 := \min_{\theta \neq \theta', \theta' \neq \theta^*} \sum_{i=1}^n u_{\theta\theta'}(i) (\mu_i(\theta') - \mu_i(\theta^*))^2,$$

where we assume that  $D'_1 > 0$ . Then for a constant  $C > 0$  the sample complexity of T2 has the following upper bound:

$$\mathbb{E}[\tau_\delta] \leq O\left(\frac{\eta \log(C) \log J}{D'_1} + \frac{\log(J/\delta)}{D'_0} + JC^{1/\eta} \delta^{D'_0/\eta^2}\right).$$

*Proof. Step 1 (Definitions):* Let the action  $i_{\theta\theta'} := \arg \max_{i \in [n]} (\mu_i(\theta) - \mu_i(\theta'))^2$ . Let  $\hat{\theta}(t)$  denote the most likely hypothesis at round  $s$  and  $\tilde{\theta}(t)$  be the second most likely hypothesis at round  $s$ . Note that T2 only samples the action  $i_{\theta\theta'}$  at round  $t$  when  $\hat{\theta}(t) = \theta$  and  $\tilde{\theta}(t) = \theta'$ . Again, let  $L_t(\theta)$  denote the total sum of squared errors of hypothesis  $\theta$  till round  $t$ . We further define the set  $\mathcal{I} := \{i \in [n] : i = \arg \max_{i' \in [n]} (\mu_{i'}(\theta) - \mu_{i'}(\theta'))^2 \text{ for some } \theta, \theta' \in \Theta\}$ .

**Step 2 (Define stopping time  $\tau_\delta$ ):** We define the time  $\tau_\delta$  for the policy T2 as follows:

$$\tau_\delta := \min\{t : \exists \theta \in \Theta, L_t(\theta') - L_t(\theta) > \beta(J, \delta), \forall \theta' \neq \theta\} \quad (29)$$

where,  $\beta(J, \delta)$  is the threshold function.

**Step 3 (Define bad event):** We define the bad event  $\xi^\delta(\theta)$  for the sub-optimal hypothesis  $\theta$  as follows:

$$\xi^\delta(\theta) := \{L_{\tau_\delta}(\theta') - L_{\tau_\delta}(\theta) > \beta(J, \delta), \forall \theta' \neq \theta\}. \quad (30)$$

The event  $\xi^\delta(\theta)$  denotes that a sub-optimal hypothesis  $\theta$  has been declared optimal at time  $\theta$  and its sum of squared error is smaller than any other hypothesis  $\theta' \neq \theta$  at  $\tau_\delta$ .

**Step 4 (Decomposition of bad event):** Decomposing the bad event follows the same approach in Theorem 1. A crucial thing to note is that the stopping time  $\tau_\delta$  only depends on the threshold function  $\beta(J, \delta)$  and not on the sampling rule. Again we can decompose the bad event to show that only comparing  $\theta$  against  $\theta^*$  is enough to guarantee a  $\delta$ -PAC policy. Finally following (22) we can decompose the bad event  $\xi^\delta(\theta)$  as follows:

$$\xi^\delta(\theta) \subseteq \{\hat{\theta}(\tau_\delta) = \theta, L_{\tau_\delta}(\theta^*) - L_{\tau_\delta}(\theta) > \beta(J, \delta)\} \quad (31)$$

such that we compare the sub-optimal hypothesis  $\theta$  only with optimal hypothesis  $\theta^*$ .

**Step 5 (Control bad event):** The control of the bad event follows the same approach in Theorem 1. We want to show that based on the definition of  $\tau_\delta$  and the bad event  $\xi^\delta(\theta)$  the T2 stops and outputs the correct hypothesis  $\theta^*$  with  $1 - \delta$  probability. As shown in Step 4, we can define the error event  $\xi^\delta(\theta)$  as follows:

$$\xi^\delta(\theta) \subseteq \{\hat{\theta}(\tau_\delta) = \theta, L_{\tau_\delta}(\theta^*) - L_{\tau_\delta}(\theta) > \beta(J, \delta)\}$$

Again define  $\tau_{\theta^* \theta} := \min\{t : -\Delta_t(\theta) > \beta(J, \delta)\}$ . Then following the same steps as in Step 5 of Theorem 1 we can show that

$$\begin{aligned} \mathbb{P}\left(\tau_\delta < \infty, \widehat{\theta}(\tau_\delta) \neq \theta^*\right) &\leq \mathbb{P}(\exists \theta \in \Theta \setminus \{\theta^*\}, \exists t \in \mathbb{N} : -\Delta_t(\theta) > \beta(J, \delta)) \\ &\leq \sum_{\theta \neq \theta^*} \mathbb{P}(L_{\tau_{\theta^* \theta}}(\theta^*) - L_{\tau_{\theta^* \theta}}(\theta) > \beta(J, \delta), \tau_{\theta^* \theta} < \infty) \stackrel{(a)}{\leq} \sum_{\theta \neq \theta^*} \frac{\delta}{J} \leq \delta \end{aligned}$$

where (a) follows from Lemma 9 and the definition of  $\beta(J, \delta)$ .

**Step 6 (Sample complexity analysis):** In this step we bound the total sample complexity of **T2** satisfying the  $\delta$ -PAC criteria. Recall that the set  $\mathcal{I} := \{i \in [n] : i = \arg \max_{i' \in [n]} (\mu_{i'}(\theta) - \mu_{i'}(\theta'))^2 \text{ for some } \theta, \theta' \in \Theta\}$ . Note that **T2** does not sample by the Chernoff p.m.f.  $\mathbf{p}_\theta$ . Rather it samples by the p.m.f.

$$u_{\theta\theta'} := \frac{1}{|\mathcal{I}(\theta\theta')|} \quad (32)$$

where,  $\mathcal{I}(\theta\theta') := \{i \in \mathcal{I} : i = \arg \max_{i' \in [n]} (\mu_{i'}(\theta) - \mu_{i'}(\theta'))^2 \text{ for } \theta, \theta' \in \Theta\}$ . Hence  $u_{\theta\theta'}$  is a uniform random p.m.f between all the maximum mean squared difference actions between hypotheses  $\theta$  and  $\theta'$  which are  $\widehat{\theta}(t)$  and  $\widehat{\theta}'(t)$  respectively for some rounds  $s \in [\tau_\delta]$ . The rest of the analysis follows the same steps as in Step 6 of Theorem 1 as the proof does not rely on any specific type of sampling proportion. We define the stopping time  $\tau_\delta$  as follows:

$$\tau_\delta = \min \left\{ t : L_t(\theta') - L_t(\widehat{\theta}(t)) \geq \beta(J, \delta), \forall \theta' \neq \widehat{\theta}(t) \right\}.$$

We further define the time  $\tau_{\theta^*}$  for the hypothesis  $\theta^*$  as follows:

$$\tau_{\theta^*} := \min \{ t : L_t(\theta') - L_t(\theta^*) \geq \beta, \forall \theta' \neq \theta^* \} \quad (33)$$

We also define the critical number of samples as  $(1+c)M'$  where  $M'$  is defined as follows:

$$M := \left( \frac{\alpha(J)}{D'_1} + \frac{C' + \log(J/\delta)}{D'_0} \right) \quad (34)$$

where,  $C' = \log(1 + \eta^2/\eta_0^2)$ ,  $c > 0$  is a constant, and we define the term  $D'_1$  as follows:

$$D'_1 := \min_{\theta \neq \theta', \theta' \neq \theta^*} \sum_{i=1}^n u_{\theta\theta'}(i) (\mu_i(\theta') - \mu_i(\theta^*))^2$$

and the term  $D'_0$  as follows:

$$D'_0 := \min_{\theta, \theta' \neq \theta^*} \sum_{i=1}^n u_{\theta^* \theta}(i) (\mu_i(\theta') - \mu_i(\theta^*))^2.$$

It then follows that

$$\begin{aligned} \mathbb{E}[\tau_\delta] &\leq \mathbb{E}[\tau_{\theta^*}] = \sum_{t=0}^{\infty} \mathbb{P}(\tau_{\theta^*} > t) \stackrel{(a)}{\leq} 1 + (1+c)M + \sum_{t:t > (1+c)M} \mathbb{P}(\tau_{\theta^*} > t) \\ &\stackrel{(b)}{\leq} 1 + (1+c) \left( \frac{b \log J}{D'_1} + \frac{C' + \log(J/\delta)}{D'_0} \right) + \sum_{\theta' \neq \theta^*} \sum_{t:t > \left( \frac{\alpha(J)}{D'_1} + \frac{4 \log(J/\delta)}{D'_0} \right) (1+c)} C_1 \exp(-C_2 t) \\ &\leq 1 + (1+c) \left( \frac{b \log J}{D'_1} + \frac{C' + \log(J/\delta)}{D'_0} \right) + \sum_{\theta' \neq \theta^*} C_1 \frac{\exp\left(-C_2 \left( \frac{\alpha(J)}{D'_1} + \frac{4 \log(J/\delta)}{D'_0} \right)\right)}{1 - \exp(-C_2)} \\ &\stackrel{(c)}{\leq} \underbrace{1 + 2 \left( \frac{b \log J}{D'_1} + \frac{C' + \log(J/\delta)}{D'_0} \right)}_{\text{Term A}} + \underbrace{\left( 165 + \frac{2\eta^2}{\eta_0^2} \right)^2 J \left( 1 - \frac{\eta_0^2(b+4)}{2\eta^3} \right)}_{\text{Term B}} \times \underbrace{\left( \frac{D'_0}{\delta \eta^2} \right)}_{\text{Term C}} \end{aligned} \quad (35)$$

where, (a) follows from definition of  $M$  in (34), (b) follows from Lemma 11 where  $C_1 := 110 + 55 \max\left\{1, \frac{\eta^2}{2D_1^2}\right\}$ ,  $C_2 := \frac{2D_1^2 \min\{(\frac{c}{2} - 1)^2, c\}}{\eta^2}$ , and (c) follows the same steps in Theorem 1 by setting  $c = \frac{1}{2}$  in  $C_1$  and  $C_2$ .

Again, note that in (35) the Term  $C \leq 1$  as  $\delta \in (0, 1)$ . So for a constant  $b > 1$  such that if  $b$  satisfies the following condition

$$b = \eta \log\left(165 + \frac{\eta^2}{\eta_0^2}\right) > \log\left(1 + \frac{\eta^2}{\eta_0^2}\right) \quad (36)$$

we have that Term B  $\leq J^{1 - \frac{\eta_0^2(b+4)}{2\eta^3}} + \frac{b}{\eta \log J}$ . Hence, Plugging this in (35) we get that the expected sample complexity is of the order of

$$\mathbb{E}[\tau_\delta] \leq O\left(\frac{\eta \log(C) \log J}{D_1} + \frac{\log(J/\delta)}{D_0} + JC^{1/\eta} \delta^{D_0/\eta^2}\right).$$

where,  $C = 165 + \frac{\eta^2}{\eta_0^2}$ . The claim of the theorem follows.  $\square$

#### A.4 Proof of Proposition 2 (Batched Setting)

**Proposition 2. (Restatement)** *Let  $\tau_\delta$ ,  $D_0$ ,  $D_1$  be defined as in Theorem 1 and  $B$  be the batch size. Then the sample complexity of  $\delta$ -PAC Batch-CS is*

$$\mathbb{E}[\tau_\delta] \leq O\left(B + \frac{\eta \log(C) \log J}{D_1} + \frac{\log(J/\delta)}{D_0} + BJC^{\frac{1}{\eta}} \delta^{\frac{D_0}{\eta^2}}\right).$$

*Proof.* We follow the same proof technique as in Theorem 1. We define the last phase after which the algorithm stops as  $m_\delta$  defined as follows:

$$m_\delta = \min\{m : L_{mB}(\boldsymbol{\theta}') - L_{mB}(\widehat{\boldsymbol{\theta}}(t)) > \beta(J, \delta), \forall \boldsymbol{\theta}' \neq \widehat{\boldsymbol{\theta}}(t)\}.$$

We further define the phase  $m_{\boldsymbol{\theta}^*}$  as follows:

$$m_{\boldsymbol{\theta}^*} = \min\{m : L_{mB}(\boldsymbol{\theta}') - L_{mB}(\widehat{\boldsymbol{\theta}}(t)) > \beta(J, \delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*\}.$$

Then we can show that the expected last phase  $m_\delta$  is bounded as follows:

$$\mathbb{E}[m_\delta] \leq \mathbb{E}[m_{\boldsymbol{\theta}^*}] = \sum_{m=1}^{\infty} \mathbb{P}(m_{\boldsymbol{\theta}^*} > m) \leq 1 + \underbrace{(1+c)M_1}_{\text{Part A}} + \sum_{m': m' > (1+c)M_1} \underbrace{\mathbb{P}(m_{\boldsymbol{\theta}^*} > m')}_{\text{Part B}} \quad (37)$$

where, in Part A we define the critical number of phases

$$M_1 = \frac{\alpha(J)}{BD_1} + \frac{\beta(J, \delta)}{BD_0}.$$

and  $D_0$ ,  $D_1$  as defined in Theorem 1. Note that this definition of  $M_1$  is different that the critical number of samples defined in Theorem 1. Now we control the Part B. As like Lemma 5 we define the following bad events

$$\begin{aligned} \xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(m) &= \{L_{mB}(\boldsymbol{\theta}') - L_{mB}(\boldsymbol{\theta}^*) < \beta(J, \delta)\} \\ \tilde{\xi}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(m) &= \{L_{mB}(\boldsymbol{\theta}') - L_{mB}(\boldsymbol{\theta}^*) < \alpha(J)\} \end{aligned}$$

We further define the last good phase  $\tilde{m}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}$  as follows:

$$\begin{aligned} \tilde{m}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*} &= \min\{m : L_{m'B}(\boldsymbol{\theta}') - L_{m'B}(\boldsymbol{\theta}^*) > \alpha(J), \forall m' > m\} \\ \text{and } \tilde{m}_{\boldsymbol{\theta}^*} &= \max_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \{\tilde{m}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}\} \end{aligned}$$

denote the last phase after which in all subsequent phases we have  $\theta^*$  is  $\widehat{\theta}(t)$ . We further define  $\tilde{m}_{\theta^*}$  as

$$\tilde{m}_{\theta^*} = \frac{\alpha(J)}{BD_1} + \frac{mc}{2}.$$

Note that this definition of  $\tilde{m}_{\theta^*}$  is different that  $\tilde{\tau}_{\theta^*}$  in Lemma 5. Using Lemma 6 we can further show the event  $\xi_{\theta'\theta^*}$  is bounded as follows:

$$\begin{aligned} \mathbb{P}(\xi_{\theta'\theta^*}(m)) &= \mathbb{P}\left(\underbrace{L_{mB}(\theta') - L_{mB}(\theta^*)}_{:=\Delta_{mB}(\theta',\theta^*)} < \beta(J,\delta)\right) \\ &\leq \mathbb{P}\left(\Delta_{mB}(\theta',\theta^*) - \mathbb{E}[\Delta_{mB}(\theta',\theta^*)] < D_0 \left(\frac{\beta(J,\delta)}{D_0} - B(m - \tilde{m}_{\theta^*})\right)\right) \\ &\stackrel{(a)}{\leq} \exp(4B) \sum_{\theta' \neq \theta^*} \exp\left(-\frac{2D_1^2 B m \left(\frac{c}{2} - 1\right)^2}{\eta^2}\right) \end{aligned} \quad (38)$$

where, (a) follows usinf the same steps as in Lemma 6. Similarly we can show that,

$$\begin{aligned} \mathbb{P}\left(\tilde{\xi}_{\theta'\theta^*}(m)\right) &= \sum_{\theta' \neq \theta^*} \sum_{m': m' > \frac{\alpha(J)}{BD_1} + \frac{mc}{2}} \mathbb{P}(\Delta_{m'B} - \mathbb{E}[\Delta_{m'B}] < \alpha(J) - \mathbb{E}[\Delta_{m'B}]) \\ &\stackrel{(a)}{\leq} \sum_{\theta' \neq \theta^*} \sum_{m': m' > \frac{\alpha(J)}{BD_1} + \frac{mc}{2}} \mathbb{P}(\Delta_{m'B} - \mathbb{E}[\Delta_{m'B}] < \alpha(J) - m'BD_1) \stackrel{(b)}{\leq} \exp(4) \sum_{\theta' \neq \theta^*} \frac{\exp\left(-\frac{2D_1^2 mcB}{2}\right)}{1 - \exp\left(-\frac{2D_1^2 cB}{2}\right)} \end{aligned} \quad (39)$$

where, (a) follows as  $\mathbb{E}[\Delta_{m'B}] \geq m'BD_1$  for all  $m' > \frac{\alpha(J)}{BD_1} + \frac{mc}{2}$ , and (b) follows using the same steps as in Lemma 7.

Finally using eq. (38) and eq. (39) we can show that the Part B is bounded as follows:

$$\begin{aligned} \mathbb{P}(m_{\theta^*} > m) &\leq \mathbb{P}\left(\left\{\bigcup_{\theta' \neq \theta^*} \xi_{\theta'\theta^*}(m)\right\} \cap \left\{\tilde{m}_{\theta^*} < \frac{\alpha}{BD_1} + \frac{mc}{2}\right\}\right) + \sum_{\theta' \neq \theta^*} \sum_{m': m' > \frac{\alpha(J)}{BD_1} + \frac{mc}{2}} \mathbb{P}\left(\tilde{\xi}_{\theta'\theta^*}(m')\right) \\ &\stackrel{(a)}{\leq} \sum_{\theta' \neq \theta^*} \underbrace{\left[110 + 55 \max\left\{1, \frac{\eta^2}{2D_1^2 B}\right\}\right]}_{:=C_1} \exp\left(-\underbrace{\frac{2D_1^2 B \min\{(c/2 - 1)^2, c\}}{\eta^2} m}_{:=C_2}\right) \leq JC_1 \exp(-C_2 m) \end{aligned}$$

where, (a) follows from the same steps as in Lemma 5, and using eq. (38) and eq. (39). Plugging this back in eq. (37) we get that

$$\begin{aligned} \mathbb{E}[m_\delta] &\leq 1 + (1+c) \left(\frac{\alpha(J)}{BD_1} + \frac{\beta(J,\delta)}{BD_0}\right) + J \sum_{m: m > \frac{\alpha(J)}{BD_1} + \frac{\beta(J,\delta)}{BD_0}} C_1 \exp(-C_2 m) \\ &\stackrel{(b)}{\leq} 1 + 2 \left(\frac{\eta \log(165 + \eta^2/B\eta_0^2) \log J}{BD_1} + \frac{\log(1 + \eta^2/\eta_0^2 B) + \log\left(\frac{J}{\delta}\right)}{BD_0}\right) + J^{1+} \frac{\log(165 + \eta^2/B\eta_0^2)}{\log J} \delta^{D_0/\eta^2} \\ &\stackrel{(b)}{\leq} O\left(1 + \frac{\eta \log(C) \log J}{BD_1} + \frac{\log(J/\delta)}{BD_0} + JC^{1/\delta} \delta^{D_0/\eta^2}\right) \end{aligned}$$

where, (a) follows using the same steps as in Step 6 of Theorem 1, and in (b) we substitute  $C := (165 + \eta^2/\eta_0^2) > (165 + \eta^2/B\eta_0^2)$ . Finally, the expected total number of samples is given by

$$\mathbb{E}[\tau_\delta] \leq B\mathbb{E}[m_\delta] = O\left(B + \frac{\eta \log(C) \log J}{D_1} + \frac{\log(J/\delta)}{D_0} + BJ C^{1/\delta} \delta^{D_0/\eta^2}\right).$$

The claim of the proposition follows.  $\square$

### A.5 Sample Complexity Proof of CSE

**Proposition 3. (Restatement)** Let  $\tau_\delta$ ,  $D_0$ ,  $C$  be defined as in Theorem 1,  $D_e$  be defined as above, and  $\epsilon_t := 1/\sqrt{t}$ . Then the sample complexity bound of  $\delta$ -PAC CSE with  $\epsilon_t$  exploration is given by

$$E[\tau_\delta] \leq O\left(\frac{\eta \log(C) \log J}{D_e} + \frac{\log(J/\delta)}{D_0} + JC^{1/\eta} \delta^{D_0/\eta^2}\right).$$

*Proof.* Recall that

$$D_1 := \min_{\{\boldsymbol{\theta} : \boldsymbol{\theta} \in \Theta\}} \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n p_{\boldsymbol{\theta}}(i) (\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2.$$

Define  $D_e := \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n \frac{1}{n} (\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2$  as the objective value of uniform sampling optimization. Finally define the quantity at round  $s$  as

$$D_1^{\epsilon_s} := (1 - \epsilon_s)D_1 + \epsilon_s D_e \quad (40)$$

Let  $\Delta_t(\boldsymbol{\theta}') := L_t(\boldsymbol{\theta}') - L_t(\boldsymbol{\theta}^*)$ . Note that following Assumption 1 we can show that

$$\Delta_s(\boldsymbol{\theta}') - D_1^{\epsilon_s} \stackrel{(a)}{\leq} 4\eta, \quad \Delta_s(\boldsymbol{\theta}') - D_0 \leq 4\eta$$

where, (a) follows as  $D_1 \leq \eta$  and  $D_e \leq \eta$  which implies  $(1 - \epsilon)D_1 + \epsilon D_e \leq \eta$  as  $\epsilon \in (0, 1)$ . Now define the quantity

$$\alpha(J) := \frac{C' \log J}{D_e},$$

$$\beta(J, \delta) := \frac{C' + \log(J/\delta)}{D_0}$$

where, the constant  $C' := \log(1 + \eta^2/\eta_0^2)$ . Now define the failure events

$$\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t) := \{L_t(\boldsymbol{\theta}') - L_t(\boldsymbol{\theta}^*) < \beta(J, \delta)\},$$

$$\tilde{\xi}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t) := \{L_t(\boldsymbol{\theta}') - L_t(\boldsymbol{\theta}^*) < 2\alpha(J)\}$$

where,  $\alpha(J) := \frac{C' \log J}{D_e}$ , and  $\beta(J, \delta) = \frac{C' + \log(J/\delta)}{D_0}$ . Then we define the time  $\tau_{\boldsymbol{\theta}^*}$ ,  $\tilde{\tau}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}$  and  $\tilde{\tau}_{\boldsymbol{\theta}^*}$  as follows:

$$\tau_{\boldsymbol{\theta}^*} := \min\{t : L_t(\boldsymbol{\theta}') - L_t(\boldsymbol{\theta}^*) > \beta(J, \delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*\}$$

$$\tilde{\tau}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*} := \min\{t : L_{t'}(\boldsymbol{\theta}') - L_{t'}(\boldsymbol{\theta}^*) > 2\alpha(J), \forall t' > t\}$$

$$\tilde{\tau}_{\boldsymbol{\theta}^*} := \max_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \{\tilde{\tau}_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}\}.$$

Then we have that Assumption 2 is no longer required which can be shown as follows

$$\begin{aligned} \mathbb{E}_{I^t, Y^t}[\Delta_t(\boldsymbol{\theta}')] &= \mathbb{E}_{I^t, Y^t}[L_t(\boldsymbol{\theta}') - L_t(\boldsymbol{\theta}^*)] = \mathbb{E}_{I^t, Y^t}\left[\sum_{s=1}^t (Y_s - \mu_{I_s}(\boldsymbol{\theta}^*))^2 - \sum_{s=1}^t (Y_s - \mu_{I_s}(\boldsymbol{\theta}'))^2\right] \\ &= \sum_{s=1}^t \mathbb{E}_{I_s} \mathbb{E}_{Y_s | I_s} \left[ (\mu_{I_s}(\boldsymbol{\theta}^*) - \mu_{I_s}(\boldsymbol{\theta}'))^2 \mid I_s \right] \\ &= \sum_{s=1}^t \sum_{i=1}^n \mathbb{P}(I_s = i) (\mu_i(\boldsymbol{\theta}^*) - \mu_i(\boldsymbol{\theta}'))^2 \end{aligned}$$

In the case when  $D_1 \geq 0$ , we can show that

$$\begin{aligned}
 \mathbb{E}_{I^t, Y^t}[\Delta_t(\boldsymbol{\theta}')] &\stackrel{(a)}{\geq} \sum_{s=1}^t ((1 - \epsilon_s)D_1 + \epsilon_s D_e) \\
 &= \sum_{s=1}^{\tilde{\tau}_{\boldsymbol{\theta}^*}} (1 - \epsilon_s)D_1 + \sum_{s=1}^{\tilde{\tau}_{\boldsymbol{\theta}^*}} \epsilon_s D_e + \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*+1}}^t (1 - \epsilon_s)D_0 + \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*+1}}^t \epsilon_s D_e \\
 &\stackrel{(b)}{\geq} \sum_{s=1}^{\tilde{\tau}_{\boldsymbol{\theta}^*}} \epsilon_s D_e + \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*+1}}^t D_e - \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*+1}}^t \epsilon_s D_0 + \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*+1}}^t \epsilon_s D_e \\
 &\stackrel{(c)}{\geq} 2\sqrt{\tilde{\tau}_{\boldsymbol{\theta}^*}} D_e - 2D_e + (t - \tilde{\tau}_{\boldsymbol{\theta}^*} - 1)D_e - kD_0 \\
 &= (t - 1)D_e + 2\sqrt{\tilde{\tau}_{\boldsymbol{\theta}^*}} D_e - \tilde{\tau}_{\boldsymbol{\theta}^*} D_e - (kD_0 + 2D_e) \\
 &\stackrel{(d)}{\geq} (t - 1)D_e - \left( \frac{\alpha(J)}{D_e} + \frac{tc}{2} \right) D_e \\
 &= (t - 1)D_e - \alpha(J) - \frac{tc}{2} D_e \\
 &= \left( t - 1 - \frac{tc}{2} \right) D_e - \alpha(J) = tD_e \left( 1 - \frac{1}{t} - \frac{c}{2} \right) - \alpha(J) \\
 &\stackrel{(e)}{\geq} tD_e \underbrace{\left( \frac{1}{2} - \frac{c}{2} \right)}_{c_1} - \alpha(J) \\
 &= c_1 t D_e - \alpha(J)
 \end{aligned}$$

where, (a) follows due to the forced exploration definition, (b) follows by dropping  $D_1$ , (c) follows  $\epsilon_s > \frac{1}{\sqrt{s}}$  and  $\int_1^t \frac{1}{\sqrt{s}} ds = 2\sqrt{s} - 2$  and  $D_0 \geq D_e$ , and (d) follows by definition of  $\tilde{\tau}_{\boldsymbol{\theta}^*} = \frac{\alpha(J)}{D_e} + \frac{tc}{2}$  and trivially assuming that  $2\sqrt{\tilde{\tau}_{\boldsymbol{\theta}^*}} D_e - (kD_0 + 2D_e) > 0$  for large enough  $\tilde{\tau}_{\boldsymbol{\theta}^*}$ , and (e) follows as  $t \geq 2$ . Next we can show that for  $t \geq \tilde{\tau}_{\boldsymbol{\theta}^*}$

$$\begin{aligned}
 \mathbb{E}[\Delta_t(\boldsymbol{\theta}')] &\stackrel{(a)}{=} \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*}}^t (1 - \epsilon_s)D_0 + \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*}}^t \epsilon_s D_e \stackrel{(b)}{\geq} \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*}}^t (1 - \epsilon_s)D_0 + \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*}}^t \epsilon_s \frac{D_0}{n} \\
 &= (t - \tilde{\tau}_{\boldsymbol{\theta}^*})D_0 - \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*}}^t \epsilon_s D_0 + \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*}}^t \epsilon_s \frac{D_0}{n} \\
 &= (t - \tilde{\tau}_{\boldsymbol{\theta}^*})D_0 - D_0 \frac{n-1}{n} \sum_{s=\tilde{\tau}_{\boldsymbol{\theta}^*}}^t \epsilon_s \\
 &\geq (t - \tilde{\tau}_{\boldsymbol{\theta}^*})D_0 - kD_0 \\
 &= (t - \tilde{\tau}_{\boldsymbol{\theta}^*} - k)D_0
 \end{aligned}$$

where, (a) follows from the definition of exploration, and (b) follows from as  $D_0 \leq nD_e$ . It follows that

$$\begin{aligned}
 \mathbb{P}(\xi_{\boldsymbol{\theta}^*, \boldsymbol{\theta}^*}(t)) &= \mathbb{P}(\Delta_t(\boldsymbol{\theta}') < \beta(J, \delta)) \\
 &= \mathbb{P}(\Delta_t(\boldsymbol{\theta}') - \mathbb{E}[\Delta_t(\boldsymbol{\theta}')] < \beta(J, \delta) - \mathbb{E}[\Delta_t(\boldsymbol{\theta}')] ) \\
 &\stackrel{(a)}{\leq} \mathbb{P}(\Delta_t(\boldsymbol{\theta}') - \mathbb{E}[\Delta_t(\boldsymbol{\theta}')] < \beta(J, \delta) - (t - \tilde{\tau}_{\boldsymbol{\theta}^*} - k)D_0) \\
 &\stackrel{(b)}{=} \mathbb{P}\left( \Delta_t(\boldsymbol{\theta}') - \mathbb{E}[\Delta_t(\boldsymbol{\theta}')] < D_0 \left( \frac{\beta(J, \delta)}{D_0} - t + \tilde{\tau}_{\boldsymbol{\theta}^*} + k \right) \right)
 \end{aligned}$$

Once we define the failure events we can follow the same proof technique as Theorem 2 and show that

$$\mathbb{P}(\tilde{\xi}_{\boldsymbol{\theta}^*, \boldsymbol{\theta}^*}(t)) \stackrel{(a)}{\leq} \mathbb{P}(\Delta_t(\boldsymbol{\theta}') - \mathbb{E}[\Delta_t(\boldsymbol{\theta}')] < 2\alpha(J) - c_1 t D_e)$$

where, in (a) the choice of  $c_1 t D_1$  follows as  $\mathbb{E}[\Delta_t(\boldsymbol{\theta}')] \geq c_1 t D_e$  as  $D_e > 0$ . It follows from Lemma 5 since  $D_0 \geq D_e$  and  $D_e > 0$  that the probability of the failure event is bounded as follows:

$$\begin{aligned} \mathbb{P}(\tau_{\boldsymbol{\theta}^*} > t) &\leq \mathbb{P}\left(\bigcup_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)\right) \\ &= \mathbb{P}\left(\left\{\bigcup_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)\right\} \cap \{\tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D_e} + \frac{tc}{2}\}\right) + \mathbb{P}\left(\bigcup_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \{\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)\} \cap \{\tilde{\tau}_{\boldsymbol{\theta}^*} \geq \frac{\alpha(J)}{D_e} + \frac{tc}{2}\}\right) \\ &\leq \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \mathbb{P}\left(\{\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t)\} \cap \{\tilde{\tau}_{\boldsymbol{\theta}^*} < \frac{\alpha(J)}{D_e} + \frac{tc}{2}\}\right) + \sum_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{t': t' \geq \frac{\alpha(J)}{D_e} + \frac{tc}{2}} \mathbb{P}\left(\xi_{\boldsymbol{\theta}'\boldsymbol{\theta}^*}(t')\right) \stackrel{(a)}{\leq} J C_1 \exp(-C_2 t) \end{aligned}$$

where, in (a) we substitute  $C_1 := 110 + 55 \max\left\{1, \frac{\eta^2}{2(D_e)^2}\right\}$ ,  $C_2 := \frac{2(D_e)^2 \min\{(c/2 - 1)^2, c\}}{\eta^2}$ ,  $c > 0$  is a constant, and  $J := |\Theta|$ . Now we define the critical number of samples as follows:

$$M := \left(\frac{C' \log J}{D_e} + \frac{C' + \log(J/\delta)}{D_0}\right)$$

where,  $C' = \log(1 + \eta^2/\eta_0^2)$ . Then we can bound the sample complexity for some constant  $c > 0$  as follows:

$$\begin{aligned} \mathbb{E}[\tau_\delta] &\leq \mathbb{E}[\tau_{\boldsymbol{\theta}^*}] = \sum_{t=0}^{\infty} \mathbb{P}(\tau_{\boldsymbol{\theta}^*} > t) \leq 1 + (1+c)M + \sum_{t: t > (1+c)M} \mathbb{P}(\tau_{\boldsymbol{\theta}^*} > t) \\ &\stackrel{(a)}{\leq} O\left(\frac{\eta \log(C) \log J}{D_e} + \frac{\log(J/\delta)}{D_0} + J(C)^{1/\eta} \delta^{D_0/\eta^2}\right) \end{aligned}$$

where, in (a) we substitute  $\log(C) = \log(165 + \frac{\eta^2}{\eta_0^2}) > C'$  and the rest follows as  $\log(C)/D_0 < \log(C)/D_e$ .  $\square$

## A.6 Minimax Optimality Proof (Theorem 2)

**Example 2.** We define an environment model  $B_j$  consisting of  $N$  actions and  $J$  hypotheses with true hypothesis  $\boldsymbol{\theta}^* = \boldsymbol{\theta}_j$  ( $j$ -th column) as follows:

$$\begin{array}{rcccccc} \boldsymbol{\theta} & = & \boldsymbol{\theta}_1 & \boldsymbol{\theta}_2 & \boldsymbol{\theta}_3 & \dots & \boldsymbol{\theta}_J \\ \mu_1(\boldsymbol{\theta}) & = & \Gamma & \Gamma - \frac{\Gamma}{J} & \Gamma - \frac{2\Gamma}{J} & \dots & \Gamma - \frac{(J-1)\Gamma}{J} \\ \mu_2(\boldsymbol{\theta}) & = & \iota_{21} & \iota_{22} & \iota_{23} & \dots & \iota_{2J} \\ \vdots & & & & & & \\ \mu_n(\boldsymbol{\theta}) & = & \iota_{n1} & \iota_{n2} & \iota_{n3} & \dots & \iota_{nJ} \end{array}$$

where, each  $\iota_{ij}$  is distinct and satisfies  $\iota_{ij} < \Gamma/4J$ . Note that we introduce such  $\iota_{ij}$  for different hypotheses so as not to violate Assumption 2.  $\boldsymbol{\theta}_1$  is the optimal hypothesis in  $A_1$ ,  $\boldsymbol{\theta}_2$  is the optimal hypothesis in  $A_2$  and so on such that for each  $A_j$  and  $j \in [J]$  we have column  $j$  as the optimal hypothesis.

**Theorem 2. (Restatement)** Any  $\delta$ -PAC policy  $\pi$  that identifies  $\boldsymbol{\theta}^*$  in (8) satisfies  $\mathbb{E}[\tau_\delta] \geq \Omega(J^2 \Gamma^{-2} \log(1/\delta))$ . Applying Theorem 1 to the same environment, the sample complexity of CS is  $O(J^2 \Gamma^{-2} \log(J/\delta))$  which matches the lower bound upto log factors.

*Proof.* The proof follows the standard change of measure argument. We follow the proof technique in Theorem 1 of Huang et al. (2017). We first state a problem setup in Example 2.

Let,  $\Lambda_1$  be the set of alternate models having a different optimal hypothesis than  $\boldsymbol{\theta}^* = \boldsymbol{\theta}_1$  such that all models having different optimal hypothesis than  $\boldsymbol{\theta}_1$  such as  $A_2, A_3, \dots, A_J$  are in  $\Lambda_1$ . Let  $\tau_\delta$  be the stopping time for any  $\delta$ -PAC policy  $\pi$ . Let  $Z_i(t)$  denote the number of times the action  $i$  has been sampled till round  $t$ . Let  $\hat{\boldsymbol{\theta}}(t)$  be the predicted optimal hypothesis at round  $\tau_\delta$ . We first consider the model  $A_1$ . Define the event  $\xi = \{\hat{\boldsymbol{\theta}}(t) \neq \boldsymbol{\theta}^*\}$  as the error event in model  $A_1$ . Let the event  $\xi' = \{\hat{\boldsymbol{\theta}}(t) \neq \boldsymbol{\theta}'^*\}$  be the corresponding error event in model  $A_2$ . Note that  $\xi^c \subset \xi'$ . Now since  $\pi$  is  $\delta$ -PAC policy we have  $\mathbb{P}_{A_1, \pi}(\xi) \leq \delta$  and  $\mathbb{P}_{A_2, \pi}(\xi^c) \leq \delta$ . Hence we can show that,

$$\begin{aligned}
 2\delta &\geq \mathbb{P}_{A_1,\pi}(\xi) + \mathbb{P}_{A_2,\pi}(\xi^{\mathbb{G}}) \stackrel{(a)}{\geq} \frac{1}{2} \exp(-\text{KL}(P_{A_1,\pi}||P_{A_2,\pi})) \\
 &\quad \text{KL}(P_{A_1,\pi}||P_{A_2,\pi}) \geq \log\left(\frac{1}{4\delta}\right) \\
 &\quad \sum_{i=1}^n \mathbb{E}_{A_1,\pi}[Z_i(\tau_\delta)] \cdot \left(\mu_i(\boldsymbol{\theta}^*) - \mu_i(\boldsymbol{\theta}'^*)\right)^2 \stackrel{(b)}{\geq} \log\left(\frac{1}{4\delta}\right) \\
 \left(\Gamma - \Gamma + \frac{\Gamma}{J}\right)^2 \mathbb{E}_{A_1,\pi}[Z_1(\tau_\delta)] + \sum_{i=2}^n (\iota_{i1} - \iota_{i2})^2 \mathbb{E}_{A_1,\pi}[Z_i(\tau_\delta)] &\stackrel{(c)}{\geq} \log\left(\frac{1}{4\delta}\right) \\
 \left(\frac{1}{J}\right)^2 \Gamma^2 \mathbb{E}_{A_1,\pi}[Z_1(\tau_\delta)] + \sum_{i=2}^n (\iota_{i1} - \iota_{i2})^2 \mathbb{E}_{A_1,\pi}[Z_i(\tau_\delta)] &\geq \log\left(\frac{1}{4\delta}\right) \\
 \left(\frac{1}{J}\right)^2 \Gamma^2 \mathbb{E}_{A_1,\pi}[Z_1(\tau_\delta)] + \sum_{i=2}^n \frac{\Gamma^2}{4J^2} \mathbb{E}_{A_1,\pi}[Z_i(\tau_\delta)] &\stackrel{(d)}{\geq} \log\left(\frac{1}{4\delta}\right) \tag{41}
 \end{aligned}$$

where, (a) follows from Lemma 2, (b) follows from Lemma 1, (c) follows from the construction of the bandit environments, and (d) follows as  $(\iota_{ij} - \iota_{i'j})^2 \leq \frac{\Gamma^2}{4J^2}$  for any  $i$ -th action and  $j$ -th hypothesis pair.

Now, we consider the alternate model  $A_3$ . Again define the event  $\xi = \{\widehat{\boldsymbol{\theta}}(t) \neq \boldsymbol{\theta}^*\}$  as the error event in model  $A_1$  and the event  $\xi' = \{\widehat{\boldsymbol{\theta}}(t) \neq \boldsymbol{\theta}''^*\}$  be the corresponding error event in model  $A_3$ . Note that  $\xi^{\mathbb{G}} \subset \xi'$ . Now since  $\pi$  is  $\delta$ -PAC policy we have  $\mathbb{P}_{B_1,\pi}(\xi) \leq \delta$  and  $\mathbb{P}_{A_3,\pi}(\xi^{\mathbb{G}}) \leq \delta$ . Following the same way as before we can show that,

$$\left(\frac{2}{J}\right)^2 \Gamma^2 \mathbb{E}_{A_1,\pi}[Z_1(\tau_\delta)] + \sum_{i=2}^n \frac{\Gamma^2}{4J^2} \mathbb{E}_{A_1,\pi}[Z_i(\tau_\delta)] \stackrel{(d)}{\geq} \log\left(\frac{1}{4\delta}\right). \tag{42}$$

Similarly we get the equations for all the other  $(J-2)$  alternate models in  $\Lambda_1$ . Now consider an optimization problem

$$\begin{aligned}
 &\min_{x_i: i \in [n]} \sum x_i \\
 \text{s.t.} \quad &\left(\frac{1}{J}\right)^2 \Gamma^2 x_1 + \frac{\Gamma^2}{4J^2} \sum_{i=2}^n x_i \geq \log(1/4\delta) \\
 &\left(\frac{2}{J}\right)^2 \Gamma^2 x_1 + \frac{\Gamma^2}{4J^2} \sum_{i=2}^n x_i \geq \log(1/4\delta) \\
 &\vdots \\
 &\left(\frac{J-1}{J}\right)^2 \Gamma^2 x_1 + \frac{\Gamma^2}{4J^2} \sum_{i=2}^n x_i \geq \log(1/4\delta) \\
 &x_i \geq 0, \forall i \in [n]
 \end{aligned}$$

where the optimization variables are  $x_i$ . It can be seen that the optimum objective value is  $J^2\Gamma^{-2} \log(1/4\delta)$ . Interpreting  $x_i = \mathbb{E}_{A_1,\pi}[Z_i(\tau_\delta)]$  for all  $i$ , we get that  $\mathbb{E}_{A_1,\pi}[\tau_\delta] = \sum_i x_i$  which gives us the required lower bound. Let  $\mathbf{p}_{\boldsymbol{\theta}^*}$  be the sampling p.m.f for the environment  $A_1$  for verifying  $\boldsymbol{\theta}^*$ . We also know that the Chernoff verification



in (4) has a nice linear programming formulation as stated in (5). Using that for  $A_1$  we can show that,

$$\begin{aligned}
 & \max z \\
 \text{s.t. } & \mathbf{p}_{\theta^*}(1)\Gamma^2 \left(\frac{1}{J}\right)^2 + \sum_{i=2}^n \mathbf{p}_{\theta^*}(i)(\iota_{i\theta^*} - \iota_{i\theta'})^2 \geq z \\
 & \mathbf{p}_{\theta^*}(1)\Gamma^2 \left(\frac{2}{J}\right)^2 + \sum_{i=2}^n \mathbf{p}_{\theta^*}(i)(\iota_{i\theta^*} - \iota_{i\theta''})^2 \geq z \\
 & \vdots \\
 & \mathbf{p}_{\theta^*}(1)\Gamma^2 \left(\frac{J-1}{J}\right)^2 + \sum_{i=2}^n \mathbf{p}_{\theta^*}(i)(\iota_{i\theta^*} - \iota_{i\theta'''})^2 \geq z \\
 & \sum_{i=1}^n \mathbf{p}_{\theta^*}(i) = 1.
 \end{aligned}$$

We can relax this further by noting that  $(\iota_{i\theta^*} - \iota_{i\theta'})^2 \leq \frac{\Gamma^2}{4J^2}$  for all  $i \in [n]$  and  $\theta^*, \theta' \in \Theta$ . Hence,

$$\begin{aligned}
 \max z \quad \text{s.t. } & \mathbf{p}_{\theta^*}(1)\Gamma^2 \left(\frac{1}{J}\right)^2 + \sum_{i=2}^n \mathbf{p}_{\theta^*}(i) \frac{\Gamma^2}{4J^2} \geq z \\
 & \mathbf{p}_{\theta^*}(1)\Gamma^2 \left(\frac{2}{J}\right)^2 + \sum_{i=2}^n \mathbf{p}_{\theta^*}(i) \frac{\Gamma^2}{4J^2} \geq z \\
 & \vdots \\
 & \mathbf{p}_{\theta^*}(1)\Gamma^2 \left(\frac{J-1}{J}\right)^2 + \sum_{i=2}^n \mathbf{p}_{\theta^*}(i) \frac{\Gamma^2}{4J^2} \geq z \\
 & \sum_{i=1}^n \mathbf{p}_{\theta^*}(i) = 1.
 \end{aligned}$$

Solving this above optimization gives that  $\mathbf{p}_{\theta^*}(1) = 1$ , and  $\mathbf{p}_{\theta^*}(2) = \mathbf{p}_{\theta^*}(3) = \dots = \mathbf{p}_{\theta^*}(n) = 0$ . Similarly, for verifying any hypothesis  $\theta \in \Theta$  we can show that the verification proportion is given by  $\mathbf{p}_{\theta} = (1, \underbrace{0, 0, \dots, 0}_{(J-1) \text{ zeros}})$ . This

also shows that for Example 2,

$$\begin{aligned}
 D_0 &= \min_{\theta' \neq \theta^*} \sum_{i=1}^n p_{\theta^*}(i)(\mu_i(\theta') - \mu_i(\theta^*))^2 = p_{\theta^*}(1)\Gamma^2 \left(\frac{1}{J}\right)^2 + \sum_{i=2}^n p_{\theta^*}(i)(\iota_{i\theta^*} - \iota_{i\theta'})^2 = \frac{\Gamma^2}{J^2} \\
 D_1 &= \min_{\{\mathbf{p}_{\theta} : \theta \in \Theta\}} \min_{\theta' \neq \theta^*} \sum_{i=1}^n p_{\theta}(i)(\mu_i(\theta') - \mu_i(\theta^*))^2 \stackrel{(a)}{\geq} p_{\theta}(1)\Gamma^2 \left(\frac{1}{J}\right)^2 + \sum_{i=2}^n p_{\theta}(i)(\iota_{i\theta} - \iota_{i\theta'})^2 = \frac{\Gamma^2}{J^2}
 \end{aligned}$$

where, (a) follows as the verification of any hypothesis  $\theta$  is a one hot vector  $\mathbf{p}_{\theta} = (1, \underbrace{0, 0, \dots, 0}_{(J-1) \text{ zeros}})$ . Note that

$\eta/4 = \Gamma^2$ . Plugging this in Theorem 1 gives us that the upper bound of CS as

$$\begin{aligned}
 \mathbb{E}[\tau_{\delta}] &\leq O \left( \frac{J^2 \Gamma^2 \log(\Gamma^4/\eta_0^4) \Gamma \log J}{\Gamma^2} + \frac{J^2 \log(J/\delta)}{\Gamma^2} + J \log(\Gamma^4/\eta_0^4)^{1/\Gamma^2} \delta^{\Gamma^2/(J^2 \Gamma^4)} \right) \\
 &\leq O \left( J^2 \log(\Gamma^4/\eta_0^4) \Gamma \log J + \frac{J^2 \log(J/\delta)}{\Gamma^2} \right) \leq O \left( \frac{J^2 \log(J/\delta)}{\Gamma^2} \right).
 \end{aligned}$$

The claim of the theorem follows.  $\square$

### A.7 Proof of Theorem 3 (Continuous hypotheses)

**Theorem 3. (Restatement)** Assume that  $\mu_i(\theta)$  for all  $i \in [n]$  is a differentiable function, and the set  $\{\nabla \mu_i(\hat{\theta}(t)) : i \in [n]\}$  of gradients evaluated at  $\hat{\theta}(t)$  span  $\mathbb{R}^d$ . Consider a p.m.f.  $\mathbf{p}_{\hat{\theta}(t),r}$  from (9) for verifying  $\hat{\theta}(t)$  against all

alternatives in  $\mathcal{B}_r^{\mathbf{0}}(\widehat{\boldsymbol{\theta}}(t))$ . The limiting value of  $\mathbf{p}_{\widehat{\boldsymbol{\theta}}(t),r}$  as  $r \rightarrow 0$  is

$$\mathbf{p}_{\widehat{\boldsymbol{\theta}}(t)} := \arg \max_{\mathbf{p}} \lambda_{\min} \left( \sum_{i=1}^n p(i) \nabla \mu_i(\widehat{\boldsymbol{\theta}}(t)) \nabla \mu_i(\widehat{\boldsymbol{\theta}}(t))^T \right).$$

*Proof.* For brevity we drop the  $\widehat{\boldsymbol{\theta}}(t)$  argument from the notation for the closed ball  $\mathcal{B}_r$  and its complement  $\mathcal{B}_r^{\mathbf{0}}$  centered at  $\widehat{\boldsymbol{\theta}}(t)$ . Denoting  $g_i(\boldsymbol{\theta}) := (\mu_i(\boldsymbol{\theta}) - \mu_i(\widehat{\boldsymbol{\theta}}(t)))^2$  for any generic  $\boldsymbol{\theta}$ , we can rewrite the optimization using a probability density function (p.d.f.)  $\mathbf{q}$  over parameters in  $\Theta$ , as explained subsequently.

$$\begin{aligned} \mathbf{p}_{\widehat{\boldsymbol{\theta}}(t),r} &= \arg \max_{\mathbf{p}} \inf_{\boldsymbol{\theta} \in \mathcal{B}_r^{\mathbf{0}}} \sum_{i=1}^n p(i) g_i(\boldsymbol{\theta}) \\ &= \arg \max_{\mathbf{p}} \inf_{\mathbf{q}: q(\boldsymbol{\theta})=0 \forall \boldsymbol{\theta} \in \mathcal{B}_r} \int_{\Theta} q(\boldsymbol{\theta}) \sum_{i=1}^n p(i) g_i(\boldsymbol{\theta}) d\boldsymbol{\theta}. \end{aligned} \quad (43)$$

The above equality is true because if in the LHS, the inner infimum was attained at a  $\boldsymbol{\theta} \in \Theta$ , then the same value can be attained in the RHS by a degenerate p.m.f.  $\mathbf{q}$  that puts all its mass on that  $\boldsymbol{\theta}$ . Conversely, suppose the inner infimum in the RHS was attained at a p.d.f.  $\mathbf{q}^*$ . Since the objective function is a linear in  $\mathbf{q}$ , the objective value is the same for a degenerate pdf that puts all its mass on one of the support points of  $\mathbf{q}^*$ , and this value can also be attained by the LHS infimum.

Let  $\kappa_i = \sup_{\boldsymbol{\theta} \in \mathcal{B}_r} g_i(\boldsymbol{\theta})$ ,  $\kappa = \max_i \kappa_i$ , then  $\kappa \rightarrow 0$  as  $r \rightarrow 0$ . Since  $g_i(\widehat{\boldsymbol{\theta}}(t)) = 0$  for all  $i$  we have that  $\kappa > 0$ . Consider a family  $\mathcal{Q}_r$  of pdfs supported on the boundary of  $\mathcal{B}_r$ , i.e.,

$$\mathcal{Q}_r := \left\{ \mathbf{q} : \int_{\Theta} q(\boldsymbol{\theta}) d\boldsymbol{\theta} = 1, q(\boldsymbol{\theta}') = 0 \text{ if } \|\boldsymbol{\theta}' - \widehat{\boldsymbol{\theta}}(t)\| \neq \epsilon \right\}.$$

For any pmf  $\mathbf{p}$ , the suboptimality gap in (43) by restricting the infimum to be over the set  $\mathcal{Q}_r$  is non-negative, and is upper bounded by

$$\inf_{\mathbf{q} \in \mathcal{Q}_r} \sum_{i=1}^n p(i) \int_{\Theta} q(\boldsymbol{\theta}) g_i(\boldsymbol{\theta}) d\boldsymbol{\theta} \leq \sum_{i=1}^n p(i) \int_{\boldsymbol{\theta}: \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t)\| = r} q(\boldsymbol{\theta}) \kappa d\boldsymbol{\theta} \leq \kappa, \quad (44)$$

where the first inequality is true for any  $q(\boldsymbol{\theta}) \in \mathcal{Q}_r$  and the second inequality is true because  $\kappa$  is an upper bound to the integrand at any point on the surface of  $\mathcal{B}_r$ . For any  $r > 0$ , we have that

$$0 \leq \inf_{\mathbf{q}: q(\boldsymbol{\theta})=0 \forall \boldsymbol{\theta} \in \mathcal{B}_r} \int_{\Theta} q(\boldsymbol{\theta}) \sum_{i=1}^n p(i) g_i(\boldsymbol{\theta}) d\boldsymbol{\theta} \leq \inf_{\mathbf{q} \in \mathcal{Q}_r} \int_{\Theta} q(\boldsymbol{\theta}) \sum_{i=1}^n p(i) g_i(\boldsymbol{\theta}) d\boldsymbol{\theta} \leq \kappa,$$

where the first inequality is due to  $g_i(\boldsymbol{\theta}) \geq 0$ , the second inequality is because the domain of inf is reduced, and the third inequality is by (44). As  $r \rightarrow 0$ , the quantity  $\kappa \rightarrow 0$  and the suboptimality gap also tends to zero. Hence for any pmf  $\mathbf{p}$ ,

$$\lim_{r \rightarrow 0} \inf_{\mathbf{q}: q(\boldsymbol{\theta})=0 \forall \boldsymbol{\theta} \in \mathcal{B}_r} \int_{\Theta} q(\boldsymbol{\theta}) \sum_{i=1}^n p(i) g_i(\boldsymbol{\theta}) d\boldsymbol{\theta} = \lim_{r \rightarrow 0} \inf_{\mathbf{q} \in \mathcal{Q}_r} \int_{\Theta} q(\boldsymbol{\theta}) \sum_{i=1}^n p(i) g_i(\boldsymbol{\theta}) d\boldsymbol{\theta}.$$

Since the above is true for any  $\mathbf{p}$ , it also holds for the maximizer of the infimum at each value of  $r$  in the convergent series to 0. Hence, we have that

$$\begin{aligned} \lim_{r \rightarrow 0} \mathbf{p}_{\widehat{\boldsymbol{\theta}}(t),r} &= \lim_{r \rightarrow 0} \arg \max_{\mathbf{p}} \inf_{\mathbf{q}: q(\boldsymbol{\theta})=0 \forall \boldsymbol{\theta} \in \mathcal{B}_r} \int_{\Theta} q(\boldsymbol{\theta}) \sum_{i=1}^n p(i) g_i(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \lim_{r \rightarrow 0} \arg \max_{\mathbf{p}} \inf_{\mathbf{q} \in \mathcal{Q}_r} \int_{\Theta} q(\boldsymbol{\theta}) \sum_{i=1}^n p(i) g_i(\boldsymbol{\theta}) d\boldsymbol{\theta}. \end{aligned}$$

Consider the multivariable Taylor series of  $g_i$  around  $\widehat{\boldsymbol{\theta}}(t)$ , for a  $\boldsymbol{\theta} \in \mathcal{B}_r$ .

$$g_i(\boldsymbol{\theta}) = g_i(\widehat{\boldsymbol{\theta}}(t)) + (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t))^T \nabla g_i(\widehat{\boldsymbol{\theta}}(t)) + 0.5(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t))^T \nabla^2 g_i(\widehat{\boldsymbol{\theta}}(t))(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t)) + o(\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t)\|^3).$$

For indices  $j, k \in [d]$  we can evaluate

$$\begin{aligned} \nabla g_i(\boldsymbol{\theta}) &= \left[ \frac{\partial g_i}{\partial \boldsymbol{\theta}_j}(\boldsymbol{\theta}) \right]_j = \left[ 2(\mu_i(\boldsymbol{\theta}) - \mu_i(\widehat{\boldsymbol{\theta}}(t))) \frac{\partial \mu_i}{\partial \boldsymbol{\theta}_j}(\boldsymbol{\theta}) \right]_j, \text{ and} \\ \nabla^2 g_i(\boldsymbol{\theta}) &= \left[ \frac{\partial^2 g_i}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k}(\boldsymbol{\theta}) \right]_{j,k} \\ &= \left[ 2(\mu_i(\boldsymbol{\theta}) - \mu_i(\widehat{\boldsymbol{\theta}}(t))) \frac{\partial^2 \mu_i}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k}(\boldsymbol{\theta}) + 2 \frac{\partial \mu_i}{\partial \boldsymbol{\theta}_j}(\boldsymbol{\theta}) \frac{\partial \mu_i}{\partial \boldsymbol{\theta}_k}(\boldsymbol{\theta}) \right]_{j,k} \end{aligned}$$

giving that  $\nabla g_i(\widehat{\boldsymbol{\theta}}(t)) = 0$  and  $\nabla^2 g_i(\widehat{\boldsymbol{\theta}}(t)) = 2\nabla \mu_i(\widehat{\boldsymbol{\theta}}(t)) \nabla \mu_i(\widehat{\boldsymbol{\theta}}(t))^T$ . Then  $\mathbf{p}_{\widehat{\boldsymbol{\theta}}(t),r}$  is the solution to the following:

$$\begin{aligned} &\max_{\mathbf{p}} \inf_{\mathbf{q} \in \mathcal{Q}_r} \int_{\Theta} q(\boldsymbol{\theta}) \sum_{i=1}^n p(i) g_i(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &= \max_{\mathbf{p}} \inf_{\mathbf{q} \in \mathcal{Q}_r} \int_{\boldsymbol{\theta}: \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t)\|=r} q(\boldsymbol{\theta}) \sum_{i=1}^n p(i) \left( (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t))^T \nabla \mu_i(\widehat{\boldsymbol{\theta}}(t)) \nabla \mu_i(\widehat{\boldsymbol{\theta}}(t))^T (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t)) + o(\|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t)\|^3) \right) d\boldsymbol{\theta} \\ &= \max_{\mathbf{p}} \inf_{\mathbf{q} \in \mathcal{Q}_r} \int_{\boldsymbol{\theta}: \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t)\|=r} \frac{(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t))^T \sum_{i=1}^n p(i) \nabla \mu_i(\widehat{\boldsymbol{\theta}}(t)) \nabla \mu_i(\widehat{\boldsymbol{\theta}}(t))^T (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t))}{(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t))^T (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t))} q(\boldsymbol{\theta}) \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}(t)\|^2 d\boldsymbol{\theta} + o(r^3) \\ &= \max_{\mathbf{p}} \min \text{eigenvalue} \left( \sum_{i=1}^n p(i) \nabla \mu_i(\widehat{\boldsymbol{\theta}}(t)) \nabla \mu_i(\widehat{\boldsymbol{\theta}}(t))^T \right) r^2 + o(r^3). \end{aligned}$$

The last equality uses the variational characterization of the minimum eigenvalue of a matrix and the fact that the inf would put all its mass on the  $\boldsymbol{\theta}$  aligned with the corresponding eigenvector to attain the minimum value. In the limit  $r \rightarrow 0$ , the second term is insignificant compared to the first and we get the required result.  $\square$

### A.7.1 How to solve the optimization

The optimization in Theorem 3 can be solved using convex optimization software. This is because the objective function, i.e., the minimum eigenvalue function is a concave function of the matrix argument, and the domain of the optimization  $\{\mathbf{p} : \sum_{i=1}^n p(i) = 1, p(i) \geq 0 \forall i \in [n]\}$  is a convex set. Hence we can maximize the objective over the domain. The set of gradients  $\{\nabla \mu_i(\widehat{\boldsymbol{\theta}}(t)) : i \in [n]\}$  span  $\mathbb{R}^d$ . Hence the optimal objective value is positive. Note that the verification proportions are the solution to a convex optimization problem. So we can terminate it early to get an approximate solution. Ignoring accuracy factors, a solution can be obtained in  $O((n^3 + n^2 d^2 + n d^3) \sqrt{n+d})$  operations (Nesterov and Nemirovskii, 1994).

## A.8 CS Convergence Proof for Smooth Hypotheses Space

### A.8.1 Theoretical Comparisons for Active Regression

From the result of Chaudhuri et al. (2015) we can show that the **ActiveS** algorithm enjoys a convergence guarantee as follows:

$$\mathbb{E} \left[ P_U(\widehat{\boldsymbol{\theta}}(t)) - P_U(\boldsymbol{\theta}^*) \right] \leq O \left( \frac{\sigma_U^2 \sqrt{\log(dt)}}{t} + \frac{R}{t^2} \right) \quad (45)$$

where,  $P_U$  is the loss under uniform measure,  $R$  is the maximum loss under any measure for any  $\theta \in \Theta$ , and  $\sigma_U^2$  is defined as

$$\begin{aligned}
 \sigma_U^2 &\stackrel{(a)}{=} \frac{1}{t^2} \text{Trace} \left[ \left( \sum_{s=1}^t \sum_{s'=1}^t \mathbb{E}_{\ell_s \sim D} [\nabla \ell_s(\theta^*)] \mathbb{E}_{\ell_{s'} \sim D} [\nabla \ell_{s'}(\theta^*)^\top] \right) I_\Gamma(\theta^*)^{-1} I_U(\theta^*) I_\Gamma(\theta^*)^{-1} \right] \\
 &= \frac{1}{t^2} \text{Trace} \left[ \left( \sum_{s=1}^t \sum_{s'=1}^t \mathbb{E}_{\ell_s \sim D} [\nabla \ell_s(\theta^*)] \mathbb{E}_{\ell_{s'} \sim D} [\nabla \ell_{s'}(\theta^*)^\top] \right) I_\Gamma(\theta^*)^{-1} I_U(\theta^*) I_\Gamma(\theta^*)^{-1} \right] \\
 &\stackrel{(b)}{=} \frac{1}{t^2} \text{Trace} \left[ \left( \sum_{s=1}^t \sum_{s'=1}^t I_\Gamma(\theta^*) \right) I_\Gamma(\theta^*)^{-1} I_U(\theta^*) I_\Gamma(\theta^*)^{-1} \right] \\
 &= \text{Trace} \left[ I_U(\theta^*) I_\Gamma(\theta^*)^{-1} \right]
 \end{aligned} \tag{46}$$

where, in (a) the loss  $\ell_s$  and  $\ell_{s'}$  are i.i.d drawn from the same distribution  $D$ , and  $I_\Gamma$ ,  $I_U$  are the Fisher information matrix under the sampling distribution  $\Gamma$  and  $U$ , and (b) follows from Lemma 5 of Chaudhuri et al. (2015). Note that **ActiveS** is a two-stage process that samples according to the uniform distribution  $U$  to build an estimate of  $\theta^*$  and then solves an SDP to build the sampling proportion  $\Gamma$  that minimizes the quantity  $\sigma_U^2$  and follows that sampling proportion  $\Gamma$  for the second stage. It follows that

$$\begin{aligned}
 \sigma_U^2 &= \text{Trace} \left[ I_U(\theta^*) I_\Gamma(\theta^*)^{-1} \right] \\
 &\leq \lambda_{\max}(I_U(\theta^*)) \lambda_{\max}(I_\Gamma(\theta^*)^{-1}) d \leq \frac{\lambda_1}{\lambda_{\min, \text{ActiveS}}} d C_3 \eta.
 \end{aligned}$$

where  $\lambda_{\min, \text{ActiveS}} = \lambda_{\min}(I_\Gamma(\theta^*))$ . We also have that

$$\begin{aligned}
 I_U(\theta^*) &= \mathbb{E}_{I_s \sim U} \nabla_{\theta=\theta^*}^2 \ell_s(\theta) = \mathbb{E}_{I_s \sim U} \nabla_{\theta=\theta^*}^2 (Y_s - \mu_{I_s}(\theta))^2 \\
 &= \mathbb{E}_{I_s \sim U} 2 (Y_s - \mu_{I_s}(\theta^*)) \nabla^2 \mu_{I_s}(\theta^*) - 2 \nabla \mu_{I_s}(\theta^*) \nabla \mu_{I_s}(\theta^*)^\top \\
 &= 2 \sum_{i=1}^n p_{\text{unif}}(i) [(Y_s - \mu_i(\theta^*)) \nabla^2 \mu_{I_s}(\theta^*) - \nabla \mu_{I_s}(\theta^*) \nabla \mu_{I_s}(\theta^*)^\top]
 \end{aligned}$$

This leads to the following bound on the maximum eigenvalue of the matrix  $I_U(\theta^*)$

$$\lambda_{\max}[I_U(\theta^*)] \leq \lambda_{\max} \left[ \sum_{i=1}^n p_{\text{unif}}(i) (Y_s - \mu_i(\theta^*)) \nabla^2 \mu_{I_s}(\theta^*) \right] \leq \lambda_1 C_3 \eta.$$

Plugging this in the statement of the result in eq. (45) we get that

$$\mathbb{E} \left[ P_U(\hat{\theta}(t)) - P_U(\theta^*) \right] \leq O \left( \frac{d \sqrt{\log(dt)}}{t} + \frac{R}{t^2} \right)$$

where we are only concerned with the scaling with the dimension  $d$ . Comparing this to our result in Theorem 4 we have the following convergence rate

$$\mathbb{E} \left[ P_t(\hat{\theta}_t) - P_t(\theta^*) \right] \leq (1 + \rho_t) \frac{\sigma_t^2}{t} + \frac{R}{t^2} \tag{47}$$

where  $P_t$  is a worst case measure over the data points. Next, we can show that,

$$\begin{aligned}
 \sigma_t^2 &:= \mathbb{E} \left[ \left\| \nabla \widehat{P}_t(\boldsymbol{\theta}^*) \right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}^2 \right] = \mathbb{E} \left[ \nabla \widehat{P}_t(\boldsymbol{\theta}^*)^T \nabla^2 P_t(\boldsymbol{\theta}^*)^{-1} \nabla \widehat{P}_t(\boldsymbol{\theta}^*) \right] \\
 &\stackrel{(a)}{\leq} \mathbb{E} \left[ \lambda_{\max} \left( \nabla^2 P_t(\boldsymbol{\theta}^*)^{-1} \right) \left\| \nabla \widehat{P}_t(\boldsymbol{\theta}^*) \right\|^2 \right] \\
 &\stackrel{(b)}{\leq} \mathbb{E} \left[ \lambda_{\max} \left( \nabla^2 P_t(\boldsymbol{\theta}^*)^{-1} \right) dC_3\eta \right] \\
 &= \mathbb{E} \left[ \left( \lambda_{\min} \left( \nabla^2 P_t(\boldsymbol{\theta}^*) \right) \right)^{-1} dC_3\eta \right] \\
 &= \mathbb{E} \left[ \left( \lambda_{\min} \left( \frac{2}{t} \sum_{s=1}^t \sum_{i=1}^n p_{\widehat{\boldsymbol{\theta}}_s}(i) \nabla \mu_i(\boldsymbol{\theta}^*) \nabla \mu_i(\boldsymbol{\theta}^*)^T \right) \right)^{-1} dC_3\eta \right] \\
 &\leq \frac{2dC_3\eta}{\lambda_{\min,CS}}
 \end{aligned}$$

where, (a) follows from the min-max theorem (variational characterization of the maximum eigenvalue), in (b) the quantity  $\left\| \nabla \widehat{P}_t(\boldsymbol{\theta}^*) \right\|^2 \leq d^2 C_3 \eta$  almost surely by assumption and  $C_3$  is a constant, and (c) follows where  $\lambda_{\min,CS}$  is a lower bound to  $\mathbb{E} \left[ \left( \lambda_{\min} \left( \frac{1}{t} \sum_{s=1}^t \sum_{i=1}^n p_{\widehat{\boldsymbol{\theta}}_s}(i) \nabla \mu_i(\boldsymbol{\theta}^*) \nabla \mu_i(\boldsymbol{\theta}^*)^T \right) \right)^{-1} \right]$ . Plugging this in our result in eq. (47) we get

$$\mathbb{E} \left[ P_t(\widehat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*) \right] \leq O \left( \frac{d\sqrt{\log(dt)}}{t} + \frac{R}{t^2} \right),$$

where again we are only concerned with the scaling with the dimension  $d$ . The convergence result is summarized below in this table:

Sample Complexity Bound	Comments
$\mathbb{E} \left[ P_U(\widehat{\boldsymbol{\theta}}(t)) - P_U(\boldsymbol{\theta}^*) \right] \leq O \left( \frac{d\sqrt{\log(dt)}}{t} + \frac{R}{t^2} \right)$	Loss of <b>ActiveS</b> (Chaudhuri et al., 2015). The $P_U$ is loss under uniform measure over data points in pool.
$\mathbb{E} \left[ P_t(\widehat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*) \right] \leq O \left( \frac{d\sqrt{\log(dt)}}{t} + \frac{R}{t^2} \right)$	Loss for <b>CS (Ours)</b> . The $P_t$ is loss under a worst-case measure over data points.

Table 2: Active Regression comparison.

The two upper bounds have the same scaling, even though  $P_t$  is a different loss measure than  $P_U$ . The proof has steps similar to that of Chaudhuri et al. (2015), with some additional arguments to handle the fact that our loss measure varies with time.

### A.8.2 Discussion on Definitions and Assumptions for Continuous Hypotheses

**Definition 2.** We define the following star-norm quantity at round  $t$  as

$$\|A\|_* = \left\| (\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1/2} \cdot A \cdot (\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1/2} \right\|.$$

Now we state the two following assumptions required by the Theorem 4. Also note that we define the squared loss function  $\ell_s(\boldsymbol{\theta}) = (\mu_{I_s}(\boldsymbol{\theta}) - Y_s)^2$ , the cumulative loss function  $L_s(\boldsymbol{\theta}) = \sum_{s'=1}^s \ell_{s'}(\boldsymbol{\theta}^*) = \sum_{s'=1}^s (\mu_{I_{s'}}(\boldsymbol{\theta}) - Y_{s'})^2$ , and  $\ell_1(\boldsymbol{\theta}), \ell_2(\boldsymbol{\theta}), \dots, \ell_t(\boldsymbol{\theta})$  for any  $\boldsymbol{\theta} \in \Theta$  are not independent. In contrast Chaudhuri et al. (2015) assumes that the loss functions are independent for any time  $s \in [t]$ . Next we state the assumptions used for the proof of Theorem 4.

Assumption 3 in Section 3 is a mild assumption on the bounded nature of the eigenvalues of the Hessian matrix  $\nabla_{\theta=\theta'}^2 \mu_{I_s}(\theta)$  evaluated at any  $\theta' \in \Theta$ . Then following assumption states a few regularity assumptions required for Theorem 4. A similar set of assumptions has also been used by Chaudhuri et al. (2015); Bu et al. (2019).

**Assumption 4. (Assumptions for Theorem 4):** We assume the following assumptions hold with probability 1:

1. (**Convexity of  $\ell_s$** ): The loss function  $\ell_s$  is convex for all time  $s \in [t]$ .
2. (**Smoothness of  $\ell_s$** ): The  $\ell_s$  is smooth such that the first, second, and third derivatives exist at all interior points in  $\Theta$ .
3. (**Regularity Conditions**):
  - (a)  $\Theta$  is compact and  $\ell_s(\theta)$  is bounded for all  $\theta \in \Theta$  and for all  $s \in [t]$ .
  - (b)  $\theta^*$  is an interior point in  $\Theta$ .
  - (c)  $\nabla^2 \ell_s(\theta^*)$  is positive definite, for all  $s \in [t]$ .
  - (d) There exists a neighborhood  $\mathcal{B}$  of  $\theta^*$  and a constant  $C_1$ , such that  $\nabla^2 \ell_s(\theta)$  is  $C_1$ -Lipschitz. Hence, we have that  $\|\nabla^2 \ell_s(\theta) - \nabla^2 \ell_s(\theta')\|_* \leq C_1 \|\theta - \theta'\|_{\nabla^2 P_s(\theta^*)}$ , for  $\theta, \theta'$  in this neighborhood.
4. (**Concentration at  $\theta^*$** ): We further assume that  $\|\nabla \ell_s(\theta^*)\|_{(\nabla^2 P_s(\theta^*))^{-1}} \leq C_2$  hold with probability one.

Assumption 4 (c) is different from that of Chaudhuri et al. (2015), where they assumed that  $\nabla^2 \mathbb{E}[\psi_s](\theta^*)$  is positive definite, where  $\psi_s$  are i.i.d. loss functions from some distribution. In our case the loss functions are not i.i.d., which is why we make the assumption on the loss at every time  $s$ .

### A.8.3 Concentration Lemmas for Continuous Hypotheses

**Lemma 12.** The probability that  $\|\nabla \hat{P}_t(\theta^*)\|_{(\nabla^2 P_t(\theta^*))^{-1}}$  crosses the threshold  $\sqrt{\frac{c\gamma \log(dt)}{t}} > 0$  is bounded by

$$\mathbb{P}\left(\|\nabla \hat{P}_t(\theta^*)\|_{(\nabla^2 P_t(\theta^*))^{-1}} \geq C_2 \sqrt{\frac{c\gamma \log(dt)}{t}}\right) \leq \frac{1}{t^{c\gamma}}.$$

*Proof.* Define  $\mathbf{u}_s := \nabla(Y_s - \mu_{I_s}(\theta^*))^2$ . Then we have  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_t$  as random vectors such that

$$\mathbb{E}\left[\left\|\sum_{s=1}^t \mathbf{u}_s\right\|_{(\nabla^2 P_t(\theta^*))^{-1}}^2 \middle| \mathbf{u}_1, \dots, \mathbf{u}_{s-1}\right] = \mathbb{E}\left[\sum_{s=1}^t \mathbf{u}_s^\top (\nabla^2 P_t(\theta^*))^{-1} \mathbf{u}_s \middle| \mathbf{u}_1, \dots, \mathbf{u}_{s-1}\right] \leq tC_2^2$$

Also we have that  $\|\mathbf{u}_s\| \leq C_2$ . Finally we have that

$$\mathbb{E}[\nabla_{\theta=\theta^*} \mathbf{u}_s] = -2 \sum_{i=1}^n p_{\hat{\theta}_{s-1}}(\mu_i(\theta^*) - \mu_i(\theta^*)) \nabla_{\theta=\theta^*} \mu_i(\theta^*) = 0.$$

Then following Lemma 4 and by setting  $\epsilon = c\gamma \log(dt)$  we can show that

$$\begin{aligned} & \mathbb{P}\left(\left\|\frac{1}{t} \sum_{s=1}^t \mathbf{u}_s\right\|_{(\nabla^2 P_t(\theta^*))^{-1}}^2 - \mathbb{E}\left[\left\|\frac{1}{t} \sum_{s=1}^t \mathbf{u}_s\right\|_{(\nabla^2 P_t(\theta^*))^{-1}}^2\right] > \frac{1}{t} \sqrt{8tC_2^2\epsilon} + \frac{4C_2}{3\epsilon}\right) \\ &= \mathbb{P}\left(\left\|\frac{1}{t} \sum_{s=1}^t \mathbf{u}_s\right\|_{(\nabla^2 P_t(\theta^*))^{-1}}^2 > C_1^2 + C_2 \sqrt{\frac{8\epsilon}{t}} + \frac{4C_2}{3\epsilon}\right) \\ &\leq \mathbb{P}\left(\left\|\sum_{s=1}^t \mathbf{u}_s\right\|_{(\nabla^2 P_t(\theta^*))^{-1}}^2 > C_2 \sqrt{\frac{8\epsilon}{t}}\right) = \mathbb{P}\left(\left\|\sum_{s=1}^t \mathbf{u}_s\right\|_{(\nabla^2 P_t(\theta^*))^{-1}}^2 > 4C_2 \sqrt{\frac{c\gamma \log(dt)}{t}}\right) \\ &\leq \exp(-c\gamma \log(dt)) = \left(\frac{1}{dt}\right)^{c\gamma} \leq \frac{1}{t^{c\gamma}} \end{aligned}$$

The claim of the lemma follows.  $\square$

**Lemma 13.** Let  $\widehat{P}_t(\boldsymbol{\theta}^*) = \frac{1}{t} \sum_{s=1}^t \ell_s(\boldsymbol{\theta}^*)$  and  $\nabla^2 P_t(\boldsymbol{\theta}^*) = \frac{1}{t} \sum_{s=1}^t \nabla^2 \mathbb{E}[\ell_s(\boldsymbol{\theta}^*) | \mathcal{F}^{s-1}]$ . Then we can bound the

$$\mathbb{P} \left( \lambda_{\max}(\nabla^2 \widehat{P}_t(\boldsymbol{\theta}^*) - \nabla^2 P_t(\boldsymbol{\theta}^*)) > \sqrt{\frac{8\eta^2 \lambda_1^2 c \gamma \log(dt)}{t}} \right) \leq \frac{2}{(dt)^\gamma},$$

where  $c > 0$  is a constant.

*Proof.* Recall that  $\widehat{P}_t(\boldsymbol{\theta}^*) = \frac{1}{t} \sum_{s=1}^t \ell_s(\boldsymbol{\theta}^*)$  and  $\nabla^2 P_t(\boldsymbol{\theta}^*) = \nabla^2 \mathbb{E}[\ell_s(\boldsymbol{\theta}^*) | \mathcal{F}^{s-1}]$ . We define  $\nabla^2 P_t(\boldsymbol{\theta}^*) = \frac{1}{t} \sum_{s=1}^t \nabla^2 \mathbb{E}[\ell_s(\boldsymbol{\theta}^*) | \mathcal{F}^{s-1}]$ . Denote,  $\mathbf{V}_s = 2\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \mu_{I_s}(\boldsymbol{\theta}) \nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \mu_{I_s}(\boldsymbol{\theta})^\top - 2 \sum_{i=1}^n p_{\widehat{\boldsymbol{\theta}}_{s-1}}(i) \nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \mu_i(\boldsymbol{\theta}) \nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \mu_i(\boldsymbol{\theta})^\top$ . Then we can show that,

$$\begin{aligned} & \mathbb{P} \left( \lambda_{\max}(\nabla^2 \widehat{P}_t(\boldsymbol{\theta}^*) - \nabla^2 P_t(\boldsymbol{\theta}^*)) > \sqrt{\frac{8\eta^2 \lambda_1^2 c \gamma \log(dt)}{t}} \right) \\ &= \mathbb{P} \left( \lambda_{\max} \left( \nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}^2 \frac{1}{t} \sum_{s=1}^t \ell_s(\boldsymbol{\theta}) - \frac{1}{t} \sum_{s=1}^t \nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}^2 \mathbb{E}[\ell_s(\boldsymbol{\theta}) | \mathcal{F}^{s-1}] \right) > \sqrt{\frac{8\eta^2 \lambda_1^2 c \gamma \log(dt)}{t}} \right) \\ &= \mathbb{P} \left( \lambda_{\max} \left( \nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}^2 \frac{1}{t} \sum_{s=1}^t (\ell_s(\boldsymbol{\theta}) - \nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}^2 \mathbb{E}[\ell_s(\boldsymbol{\theta}) | \mathcal{F}^{s-1}]) \right) > \sqrt{\frac{8\eta^2 \lambda_1^2 c \gamma \log(dt)}{t}} \right) \\ &\stackrel{(a)}{=} \mathbb{P} \left( \lambda_{\max} \left( \frac{1}{t} \sum_{s=1}^t (Y_s - \mu_{I_s}(\boldsymbol{\theta}^*)) \nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}^2 \mu_{I_s}(\boldsymbol{\theta}^*) + \frac{1}{t} \sum_{s=1}^t \mathbf{V}_s \right) > \sqrt{\frac{8\eta^2 \lambda_1^2 c \gamma \log(dt)}{t}} \right) \\ &\leq \mathbb{P} \left( \lambda_{\max} \left( \frac{1}{t} \sum_{s=1}^t -2(Y_s - \mu_{I_s}(\boldsymbol{\theta}^*)) \nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}^2 \mu_{I_s}(\boldsymbol{\theta}^*) \right) > \frac{1}{2} \sqrt{\frac{8\eta^2 \lambda_1^2 c \gamma \log(dt)}{t}} \right) \\ &\quad + \mathbb{P} \left( \lambda_{\max} \left( \frac{1}{t} \sum_{s=1}^t \mathbf{V}_s \right) > \frac{1}{2} \sqrt{\frac{8\eta^2 \lambda_1^2 c \gamma \log(dt)}{t}} \right) \\ &\stackrel{(b)}{\leq} \mathbb{P} \left( \frac{1}{t} \sum_{s=1}^t -2(Y_s - \mu_{I_s}(\boldsymbol{\theta}^*)) \lambda_{\max}(\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}^2 \mu_{I_s}(\boldsymbol{\theta}^*)) > \frac{1}{2} \sqrt{\frac{8\eta^2 \lambda_1^2 c \gamma \log(dt)}{t}} \right) \\ &\quad + \mathbb{P} \left( \frac{1}{t} \sum_{s=1}^t \lambda_{\max}(\mathbf{V}_s) > \frac{1}{2} \sqrt{\frac{8\eta^2 \lambda_1^2 c \gamma \log(dt)}{t}} \right) \tag{48} \\ &\stackrel{(c)}{\leq} 2 \exp \left( -\frac{t^2 8\eta^2 \lambda_1^2 c \gamma \log(dt)}{4t} \cdot \frac{1}{2tc\eta^2 \lambda_1^2} \right) \stackrel{(d)}{\leq} 2 \left( \frac{1}{dt} \right)^\gamma. \end{aligned}$$

where, (a) follows from substituting the value of  $\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}^2 \ell_s(\boldsymbol{\theta}) - \nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}^2 \mathbb{E}[\ell_s(\boldsymbol{\theta}) | \mathcal{F}^{s-1}]$  from Lemma 16, and (b) follows by triangle inequality, (c) follows by using two concentration inequalities stated below, and (d) follows by simplifying the equations.

Denote  $Q_s = -2(Y_s - \mu_{I_s}(\boldsymbol{\theta}^*)) \lambda_{\max}(\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}^2 \mu_{I_s}(\boldsymbol{\theta}^*))$ . Also note that  $\lambda_{\max}(\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}^2 \mu_{I_s}(\boldsymbol{\theta}^*)) \leq \lambda_1$ .

$$\begin{aligned}
 \mathbb{P}\left(\sum_{s=1}^t -2(Y_s - \mu_{I_s}(\boldsymbol{\theta}^*)) \lambda_{\max}(\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}^2 \mu_{I_s}(\boldsymbol{\theta}^*)) \geq \epsilon\right) &= \mathbb{P}\left(-\sum_{s=1}^t Q_s \geq \epsilon\right) \\
 &= \mathbb{P}\left(e^{-\lambda \sum_{s=1}^t Q_s} \geq e^{\lambda \epsilon}\right) \stackrel{(a)}{\leq} e^{-\lambda \epsilon} \mathbb{E}\left[e^{-\lambda \sum_{s=1}^t Q_s}\right] = e^{-\lambda \epsilon} \mathbb{E}\left[\mathbb{E}\left[e^{-\lambda \sum_{s=1}^t Q_s} \mid \widehat{\boldsymbol{\theta}}(t-1)\right]\right] \\
 &\stackrel{(b)}{=} e^{-\lambda \epsilon} \mathbb{E}\left[\mathbb{E}\left[e^{-\lambda Q_t} \mid \widehat{\boldsymbol{\theta}}(t-1)\right] \mathbb{E}\left[e^{-\lambda \sum_{s=1}^{t-1} Q_s} \mid \widehat{\boldsymbol{\theta}}(t-1)\right]\right] \\
 &\leq e^{-\lambda \epsilon} \mathbb{E}\left[\exp(2\lambda^2 \lambda_1^2 \eta^2) \mathbb{E}\left[e^{-\lambda \sum_{s=1}^{t-1} Q_s} \mid \widehat{\boldsymbol{\theta}}(t-1)\right]\right] \\
 &= e^{-\lambda \epsilon} e^{2\lambda^2 \eta^2 \lambda_1^2} \mathbb{E}\left[e^{-\lambda \sum_{s=1}^{t-1} Q_s}\right] \\
 &\vdots \\
 &\stackrel{(c)}{\leq} e^{-\lambda \epsilon} e^{2\lambda^2 t \eta^2 \lambda_1^2} \stackrel{(d)}{\leq} \exp\left(-\frac{2\epsilon^2}{t \eta^2 \lambda_1^2}\right).
 \end{aligned}$$

where (a) follows by Markov's inequality, (b) follows as  $Q_s$  is conditionally independent given  $\widehat{\boldsymbol{\theta}}(s-1)$ , (c) follows by unpacking the term for  $t$  times and (d) follows by taking  $\lambda = \epsilon/4t\lambda_1^2\eta^2$ . Next we bound the second term of (48) below.

$$\begin{aligned}
 \mathbb{P}\left(\sum_{s=1}^t \lambda_{\max}(\mathbf{V}_s) \geq \epsilon\right) &= \mathbb{P}\left(\lambda \sum_{s=1}^t \lambda_{\max}(\mathbf{V}_s) \geq \lambda \epsilon\right) = \mathbb{P}\left(e^{\lambda \sum_{s=1}^t \lambda_{\max}(\mathbf{V}_s)} \geq e^{\lambda \epsilon}\right) \stackrel{(a)}{\leq} e^{-\lambda \epsilon} \mathbb{E}\left[e^{\lambda \sum_{s=1}^t \lambda_{\max}(\mathbf{V}_s)}\right] \\
 &= e^{-\lambda \epsilon} \mathbb{E}\left[\mathbb{E}\left[e^{\lambda \sum_{s=1}^t \lambda_{\max}(\mathbf{V}_s)} \mid \widehat{\boldsymbol{\theta}}(t-1)\right]\right] \\
 &\stackrel{(b)}{=} e^{-\lambda \epsilon} \mathbb{E}\left[\mathbb{E}\left[e^{\lambda \lambda_{\max}(\mathbf{V}_t)} \mid \widehat{\boldsymbol{\theta}}(t-1)\right] \mathbb{E}\left[e^{\lambda \sum_{s=1}^{t-1} \lambda_{\max}(\mathbf{V}_s)} \mid \widehat{\boldsymbol{\theta}}(t-1)\right]\right] \\
 &\stackrel{(c)}{\leq} e^{-\lambda \epsilon} \mathbb{E}\left[\exp(2c\lambda^2 \lambda_1^2 \eta^2) \mathbb{E}\left[e^{\lambda \sum_{s=1}^{t-1} \lambda_{\max}(\mathbf{V}_s)} \mid \widehat{\boldsymbol{\theta}}(t-1)\right]\right] \\
 &= e^{-\lambda \epsilon} e^{2c\lambda^2 \eta^2 \lambda_1^2} \mathbb{E}\left[e^{\lambda \sum_{s=1}^{t-1} \lambda_{\max}(\mathbf{V}_s)}\right] \\
 &\vdots \\
 &\stackrel{(d)}{\leq} e^{-\lambda \epsilon} e^{2c\lambda^2 t \eta^2 \lambda_1^2} \stackrel{(e)}{\leq} \exp\left(-\frac{2\epsilon^2}{tc\eta^2 \lambda_1^2}\right)
 \end{aligned}$$

where (a) follows by Markov's inequality, (b) follows as  $\lambda_{\max}(\mathbf{V}_s)$  is conditionally independent given  $\widehat{\boldsymbol{\theta}}(s-1)$ . In the inequality (c) using the always valid upper bound of  $2\lambda_1$ , we have that  $\mathbb{E}[\lambda_{\max}(\mathbf{V}_t)] \leq 2\lambda_1$ . So the term in inequality (c) will become  $e^{-\lambda \epsilon} e^{2\lambda^2 t \eta^2 \lambda_1^2 + 4t\lambda\lambda_1}$ . Hence, we can upper bound the inequality (c) by a constant  $c > 0$  such that we have  $\mathbb{E}[e^{\lambda \lambda_{\max}(\mathbf{V}_t)} \mid \widehat{\boldsymbol{\theta}}(t-1)] \leq e^{2\lambda^2 \lambda_1^2 \eta^2} e^{2\lambda \times 2\lambda_1} = \exp(2\lambda^2 \lambda_1^2 \eta^2 + 4\lambda\lambda_1) \leq \exp(2c\lambda^2 \lambda_1^2 \eta^2)$ . The inequality (d) follows by unpacking the term for  $t$  times and (e) follows by taking  $\lambda = \epsilon/4tc\lambda_1^2\eta^2$ .  $\square$

#### A.8.4 Support Lemma for Continuous Hypotheses

**Lemma 14.** *Let the  $j$ -th row and  $k$ -th column entry in the Hessian matrix  $\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2(\ell_s(\boldsymbol{\theta}))$  be denoted as  $[\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2(\ell_s(\boldsymbol{\theta}))]_{jk}$ . Then we have that*

$$[\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2(\ell_s(\boldsymbol{\theta}))]_{jk} = 2 \frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \theta_j} \frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \theta_k} + 2(\mu_{I_s}(\boldsymbol{\theta}) - Y_s) \frac{\partial^2 \mu_{I_s}(\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k}.$$

*Proof.* We want to evaluate the Hessian  $\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2(\ell_s(\boldsymbol{\theta}))$  at any  $\boldsymbol{\theta}' \in \Theta$ . We denote the  $j$ -th row and  $k$ -th column



entry in the Hessian matrix as  $[\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2(\ell_s(\boldsymbol{\theta}))]_{jk}$ . Then we can show that

$$\begin{aligned} [\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2(\ell_s(\boldsymbol{\theta}))]_{jk} &:= \frac{\partial}{\partial \boldsymbol{\theta}_j} \left[ \frac{\partial(\mu_{I_s}(\boldsymbol{\theta}) - Y_s)^2}{\partial \boldsymbol{\theta}_k} \right] = \frac{\partial}{\partial \boldsymbol{\theta}_j} \left[ 2(\mu_{I_s}(\boldsymbol{\theta}) - Y_s) \frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} \right] \\ &= \frac{\partial}{\partial \boldsymbol{\theta}_j} \left[ 2\mu_{I_s}(\boldsymbol{\theta}) \frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} - 2Y_s \frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} \right] \\ &= 2 \frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} + 2\mu_{I_s}(\boldsymbol{\theta}) \frac{\partial^2 \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k} - 2Y_s \frac{\partial^2 \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k} - 2 \frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \frac{\partial Y_s}{\partial \boldsymbol{\theta}_k} \\ &= 2 \frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} + 2(\mu_{I_s}(\boldsymbol{\theta}) - Y_s) \frac{\partial^2 \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k} \end{aligned}$$

The claim of the lemma follows.  $\square$

**Lemma 15.** *Let the  $j$ -th row and  $k$ -th column entry in the Hessian matrix  $\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2(\mathbb{E}[\ell_s(\boldsymbol{\theta})|\mathcal{F}^{s-1}])$  be denoted as  $[\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2(\mathbb{E}[\ell_s(\boldsymbol{\theta})|\mathcal{F}^{s-1}])]_{jk}$ . Then we have that*

$$[\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2 \mathbb{E}[\ell_s(\boldsymbol{\theta})|\mathcal{F}^{s-1}]]_{jk} = 2 \sum_{i=1}^n p_{\hat{\boldsymbol{\theta}}_{s-1}}(i) \left( \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} + 2(\mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}^*)) \frac{\partial^2 \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k} \right).$$

*Proof.* Now we want to evaluate the Hessian  $\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2(\mathbb{E}[\ell_s(\boldsymbol{\theta})|\mathcal{F}^{s-1}])$  at any  $\boldsymbol{\theta}' \in \Theta$ . We denote the  $j$ -th row and  $k$ -th column entry in the Hessian matrix as  $[\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2(\mathbb{E}[\ell_s(\boldsymbol{\theta})|\mathcal{F}^{s-1}])]_{jk}$ . Then we can show that

$$\begin{aligned} \nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2 \mathbb{E}[\ell_s(\boldsymbol{\theta})|\mathcal{F}^{s-1}] &= \nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2 (\mu_{I_s}^2(\boldsymbol{\theta}) + \mathbb{E}[Y_s^2|\mathcal{F}^{s-1}] - 2\mathbb{E}[Y_s|\mathcal{F}^{s-1}]\mu_{I_s}(\boldsymbol{\theta})) \\ &= \nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2 \sum_{i=1}^n p_{\hat{\boldsymbol{\theta}}_{s-1}}(i) \left( \mu_i^2(\boldsymbol{\theta}) + \mu_i^2(\boldsymbol{\theta}') + \frac{1}{2} - 2\mu_i(\boldsymbol{\theta}^*)\mu_i(\boldsymbol{\theta}) \right) \\ &= \nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2 \sum_{i=1}^n p_{\hat{\boldsymbol{\theta}}_{s-1}}(i) \left( (\mu_i(\boldsymbol{\theta}^*) - \mu_i(\boldsymbol{\theta}))^2 + \frac{1}{2} \right) \\ &= \nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2 \sum_{i=1}^n p_{\hat{\boldsymbol{\theta}}_{s-1}}(i) \left( (\mu_i(\boldsymbol{\theta}^*) - \mu_i(\boldsymbol{\theta}))^2 \right) \end{aligned} \quad (49)$$

We now denote the  $j$ -th row and  $k$ -th column entry of the Hessian Matrix  $\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2((\mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}^*))^2)$  as  $[\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2((\mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}^*))^2)]_{jk}$ . Then we can show that

$$\begin{aligned} [\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2((\mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}^*))^2)]_{jk} &:= \frac{\partial}{\partial \boldsymbol{\theta}_j} \left[ \frac{\partial(\mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}^*))^2}{\partial \boldsymbol{\theta}_k} \right] = \frac{\partial}{\partial \boldsymbol{\theta}_j} \left[ 2(\mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}^*)) \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} \right] \\ &= \frac{\partial}{\partial \boldsymbol{\theta}_j} \left[ 2\mu_i(\boldsymbol{\theta}) \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} - 2\mu_i(\boldsymbol{\theta}^*) \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} \right] \\ &= 2 \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} + 2\mu_i(\boldsymbol{\theta}) \frac{\partial^2 \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k} \\ &\quad - 2\mu_i(\boldsymbol{\theta}^*) \frac{\partial^2 \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k} - 2 \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \frac{\partial \mu_i(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}_k} \\ &= 2 \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} + 2(\mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}^*)) \frac{\partial^2 \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k} \end{aligned}$$

Plugging this back in eq. (49) we get that

$$[\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2 \mathbb{E}[\ell_s(\boldsymbol{\theta})|\mathcal{F}^{s-1}]]_{jk} = 2 \sum_{i=1}^n p_{\hat{\boldsymbol{\theta}}_{s-1}}(i) \left( \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} + 2(\mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}^*)) \frac{\partial^2 \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k} \right).$$

$\square$

**Lemma 16.** *The sum of the difference of the Hessians  $\sum_{s=1}^t \nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2 \ell_s(\boldsymbol{\theta}) - \mathbb{E}[\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2 \ell_s(\boldsymbol{\theta}) \mid \mathcal{F}^{s-1}]$  is given by*

$$\begin{aligned} \sum_{s=1}^t \nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2 \ell_s(\boldsymbol{\theta}) - \mathbb{E}[\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2 \ell_s(\boldsymbol{\theta}) \mid \mathcal{F}^{s-1}] &= \sum_{s=1}^t \left( -2(Y_s - \mu_{I_s}(\boldsymbol{\theta})) \frac{\partial^2 \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k} + 2 \frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} \right. \\ &\quad \left. - 2 \sum_{i=1}^n p_{\hat{\boldsymbol{\theta}}_{s-1}}(i) \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} \right). \end{aligned}$$

*Proof.* First note that the difference  $\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2 \ell_s(\boldsymbol{\theta}) - \mathbb{E}[\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2 \ell_s(\boldsymbol{\theta}) \mid \mathcal{F}^{s-1}]_{jk}$  is given by

$$\begin{aligned} \nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2 \ell_s(\boldsymbol{\theta}) - \mathbb{E}[\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2 \ell_s(\boldsymbol{\theta}) \mid \mathcal{F}^{s-1}] &\stackrel{(a)}{=} 2 \frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} + 2(\mu_{I_s}(\boldsymbol{\theta}) - Y_s) \frac{\partial^2 \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k} \\ &\quad - 2 \sum_{i=1}^n p_{\hat{\boldsymbol{\theta}}_{s-1}}(i) \left( \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} - (\mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}^*)) \frac{\partial^2 \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k} \right) \\ &= -2(Y_s - \mu_{I_s}(\boldsymbol{\theta})) \frac{\partial^2 \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k} + 2 \frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} \\ &\quad - 2 \sum_{i=1}^n p_{\hat{\boldsymbol{\theta}}_{s-1}}(i) \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} \end{aligned} \tag{50}$$

where, (a) follows from Lemma 14 and Lemma 15. Plugging this equality in Equation (50) below we get

$$\begin{aligned} \sum_{s=1}^t \nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2 \ell_s(\boldsymbol{\theta}) - \mathbb{E}[\nabla_{\boldsymbol{\theta}=\boldsymbol{\theta}'}^2 \ell_s(\boldsymbol{\theta}) \mid \mathcal{F}^{s-1}] &= \sum_{s=1}^t \left( -2(Y_s - \mu_{I_s}(\boldsymbol{\theta})) \frac{\partial^2 \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k} + 2 \frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \frac{\partial \mu_{I_s}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} \right. \\ &\quad \left. - 2 \sum_{i=1}^n p_{\hat{\boldsymbol{\theta}}_{s-1}}(i) \left( \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} \frac{\partial \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k} - 2(\mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}^*)) \frac{\partial^2 \mu_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}_k} \right) \right). \end{aligned}$$

The claim of the lemma follows.  $\square$

**Lemma 17.** *Let  $\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^* = \left( \nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}_t) \right)^{-1} \nabla \hat{P}_t(\boldsymbol{\theta}^*)$  where  $\tilde{\boldsymbol{\theta}}_t$  is between  $\hat{\boldsymbol{\theta}}_t$  and  $\boldsymbol{\theta}^*$ . Then we can show that*

$$\left\| \hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^* \right\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)} \leq \left\| \left( \nabla^2 P_t(\boldsymbol{\theta}^*) \right)^{1/2} \left( \nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}_t) \right)^{-1} \left( \nabla^2 P_t(\boldsymbol{\theta}^*) \right)^{1/2} \right\| \left\| \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right\|_{\left( \nabla^2 P_t(\boldsymbol{\theta}^*) \right)^{-1}}.$$

*Proof.* We begin with the definition of  $\left\| \hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^* \right\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)}$  as follows:

$$\begin{aligned} \left\| \hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^* \right\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)} &\stackrel{(a)}{=} \sqrt{(\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*)^T \nabla^2 P_t(\boldsymbol{\theta}^*) (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*)} \\ &\stackrel{(b)}{=} \sqrt{\left( \left( \nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}_t) \right)^{-1} \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right)^T \nabla^2 P_t(\boldsymbol{\theta}^*) \left( \left( \nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}_t) \right)^{-1} \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right)} \\ &\stackrel{(c)}{\leq} \left\| \nabla^2 P_t(\boldsymbol{\theta}^*)^{1/2} \left( \nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}_t) \right)^{-1} \nabla^2 P_t(\boldsymbol{\theta}^*)^{1/2} \right\| \sqrt{(\nabla \hat{P}_t(\boldsymbol{\theta}^*)^T (\nabla^2 P_t(\boldsymbol{\theta}^*)^{-1} \nabla \hat{P}_t(\boldsymbol{\theta}^*))} \\ &= \left\| \left( \nabla^2 P_t(\boldsymbol{\theta}^*) \right)^{1/2} \left( \nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}_t) \right)^{-1} \left( \nabla^2 P_t(\boldsymbol{\theta}^*) \right)^{1/2} \right\| \left\| \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right\|_{\left( \nabla^2 P_t(\boldsymbol{\theta}^*) \right)^{-1}}. \end{aligned}$$

where, (a) follows as  $\|x\|_M = \sqrt{x^T M x}$ , (b) follows as  $\left\| \hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^* \right\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)} = \left( \nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}_t) \right)^{-1} \nabla \hat{P}_t(\boldsymbol{\theta}^*)$ , and (c) follows from Cauchy Schwarz inequality.

The claim of the lemma follows.  $\square$

### A.8.5 Proof of Main Theorem 4

**Theorem 4. (Restatement)** Suppose  $\ell_1(\boldsymbol{\theta}), \ell_2(\boldsymbol{\theta}), \dots, \ell_t(\boldsymbol{\theta}) : \mathbb{R}^d \rightarrow \mathbb{R}$  are squared loss functions from a distribution that satisfies Assumption 3 and Assumption 4 in Appendix A.8.2. Further define  $P_t(\boldsymbol{\theta}) = \frac{1}{t} \sum_{s=1}^t \mathbb{E}_{I_s \sim \mathcal{P}_{\hat{\boldsymbol{\theta}}_{s-1}}} [\ell_s(\boldsymbol{\theta}) | \mathcal{F}^{s-1}]$  where,  $\hat{\boldsymbol{\theta}}_t = \arg \min_{\boldsymbol{\theta} \in \Theta} \sum_{s=1}^t \ell_s(\boldsymbol{\theta})$ . If  $t$  is large enough such that  $\frac{\gamma \log(dt)}{t} \leq c' \min \left\{ \frac{1}{C_1 C_2}, \frac{\text{diameter}(\mathcal{B})}{C_2} \right\}$  then for a constant  $\gamma \geq 2$  and universal constants  $C_1, C_2, c'$ , we can show that

$$(1 - \rho_t) \frac{\sigma_t^2}{t} - \frac{C_1^2}{t^{\gamma/2}} \leq \mathbb{E} \left[ P_t(\hat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*) \right] \leq (1 + \rho_t) \frac{\sigma_t^2}{t} + \frac{\max_{\boldsymbol{\theta} \in \Theta} (P_t(\boldsymbol{\theta}) - P_t(\boldsymbol{\theta}^*))}{t^\gamma},$$

where  $\sigma_t^2 := \mathbb{E} \left[ \frac{1}{2} \left\| \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}^2 \right]$ , and  $\rho_t := (C_1 C_2 + 2\eta^2 \lambda_1^2) \sqrt{\frac{\gamma \log(dt)}{t}}$ .

*Proof. Step 1:* We first bound the  $\left\| \nabla^2 \hat{P}_t(\boldsymbol{\theta}) - \nabla^2 P_t(\boldsymbol{\theta}^*) \right\|_*$  as follows

$$\begin{aligned} \left\| \nabla^2 \hat{P}_t(\boldsymbol{\theta}) - \nabla^2 P_t(\boldsymbol{\theta}^*) \right\|_* &\stackrel{(a)}{\leq} \left\| \nabla^2 \hat{P}_t(\boldsymbol{\theta}) - \nabla^2 \hat{P}_t(\boldsymbol{\theta}^*) \right\|_* + \left\| \nabla^2 \hat{P}_t(\boldsymbol{\theta}^*) - \nabla^2 P_t(\boldsymbol{\theta}^*) \right\|_* \\ &\stackrel{(b)}{\leq} C_1 \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)} + \sqrt{\frac{8\eta^2 \lambda_1^2 c' \gamma \log(dt)}{t}} \end{aligned} \quad (51)$$

where, (a) follows from triangle inequality, and (b) is due to Assumption 4.3.d and Lemma 13.

**Step 2 (Approximation of  $\nabla^2 P_t(\boldsymbol{\theta}^*)$ ):** By choosing a sufficiently smaller ball  $\mathcal{B}_1$  of radius of  $\min \{1/(10C_1), \text{diameter}(\mathcal{B})\}$ , the first term in (51) can be made small for  $\boldsymbol{\theta} \in \mathcal{B}_1$ . Also, for sufficiently large  $t$ , the second term in (51) can be made arbitrarily small (smaller than  $1/10$ ), which occurs if  $\sqrt{\frac{\gamma \log(dt)}{t}} \leq \frac{c'}{\sqrt{2\eta^2 \lambda_1^2}}$ . Hence for large  $t$  and  $\boldsymbol{\theta} \in \mathcal{B}_1$  we have

$$\frac{1}{2} \nabla^2 \hat{P}_t(\boldsymbol{\theta}) \preceq \nabla^2 P_t(\boldsymbol{\theta}^*) \preceq 2 \nabla^2 \hat{P}_t(\boldsymbol{\theta}) \quad (52)$$

**Step 3 (Show  $\hat{\boldsymbol{\theta}}_t$  in  $\mathcal{B}_1$ ):** Fix a  $\tilde{\boldsymbol{\theta}}$  between  $\boldsymbol{\theta}$  and  $\boldsymbol{\theta}^*$  in  $\mathcal{B}_1$ . Apply Taylor's series approximation

$$\hat{P}_t(\boldsymbol{\theta}) = \hat{P}_t(\boldsymbol{\theta}^*) + \nabla \hat{P}_t(\boldsymbol{\theta}^*)^\top (\boldsymbol{\theta} - \boldsymbol{\theta}^*) + \frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top \nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}) (\boldsymbol{\theta} - \boldsymbol{\theta}^*)$$

We can further reduce this as follows:

$$\begin{aligned} \hat{P}_t(\boldsymbol{\theta}) - \hat{P}_t(\boldsymbol{\theta}^*) &\stackrel{(a)}{=} \nabla \hat{P}_t(\boldsymbol{\theta}^*)^\top (\boldsymbol{\theta} - \boldsymbol{\theta}^*) + \frac{1}{2} \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_{\nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}})}^2 \\ &\stackrel{(b)}{\geq} \nabla \hat{P}_t(\boldsymbol{\theta}^*)^\top (\boldsymbol{\theta} - \boldsymbol{\theta}^*) + \frac{1}{4} \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)}^2 \\ &\geq -\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)} \left\| \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}} + \frac{1}{4} \left( \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)} \right)^\top \left( \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)} \right) \\ &= \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)} \left( -\left\| \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}} + \frac{1}{4} \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)} \right) \end{aligned} \quad (53)$$

where, (a) follows as  $\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_{\nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}})}^2 := (\boldsymbol{\theta} - \boldsymbol{\theta}^*)^\top \nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}) (\boldsymbol{\theta} - \boldsymbol{\theta}^*)$ , and (b) follows as  $\tilde{\boldsymbol{\theta}}$  is in between  $\boldsymbol{\theta}$  and  $\boldsymbol{\theta}^*$  and then using (52). Note that in (53) if the right hand side is positive for some  $\boldsymbol{\theta} \in \mathcal{B}_1$ , then  $\boldsymbol{\theta}$  is not a local minimum. Also, since  $\left\| \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right\| \rightarrow 0$ , for a sufficiently small value of  $\left\| \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right\|$ , all points on the boundary of  $\mathcal{B}_1$  will have values greater than that of  $\boldsymbol{\theta}^*$ . Hence, we must have a local minimum of  $\hat{P}_t(\boldsymbol{\theta})$  that is strictly inside  $\mathcal{B}_1$  (for  $t$  large enough). We can ensure this local minimum condition is achieved by choosing an  $t$  large enough so that  $\sqrt{\frac{\gamma \log(dt)}{t}} \leq c' \min \left\{ \frac{1}{C_1 C_2}, \frac{\text{diameter}(\mathcal{B})}{C_2} \right\}$ , using Lemma 12 (and our bound on the diameter of  $\mathcal{B}_1$ ). By convexity, we have that this is the global minimum,  $\hat{\boldsymbol{\theta}}_t$ , and so  $\hat{\boldsymbol{\theta}}_t \in \mathcal{B}_1$  for  $t$  large enough. We will assume now that  $t$  is this large from here on.

**Step 4 (Bound  $\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)}$ ):** For the  $\hat{\boldsymbol{\theta}}(t)$  that minimizes the sum of squared errors,  $0 = \nabla \hat{P}_t(\hat{\boldsymbol{\theta}}_t)$ . Again, using Taylor's theorem if  $\hat{\boldsymbol{\theta}}_t$  is an interior point, we have:

$$0 = \nabla \hat{P}_t(\hat{\boldsymbol{\theta}}_t) = \nabla \hat{P}_t(\boldsymbol{\theta}^*) + \nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}_t) \left( \hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^* \right) \quad (54)$$

for some  $\tilde{\boldsymbol{\theta}}_t$  between  $\boldsymbol{\theta}^*$  and  $\hat{\boldsymbol{\theta}}_t$ . Now observe that  $\tilde{\boldsymbol{\theta}}_t$  is in  $B_1$  (since, for  $t$  large enough,  $\hat{\boldsymbol{\theta}}_t \in \mathcal{B}_1$ ). Thus it follows from (54) that,

$$\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^* = \left( \nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}_t) \right)^{-1} \nabla \hat{P}_t(\boldsymbol{\theta}^*) \quad (55)$$

where the invertibility is guaranteed by (52) and the positive definiteness of  $\nabla^2 P_t(\boldsymbol{\theta}^*)$  (by Assumption 4 (3c)). We finally derive the upper bound to  $\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)}$  as follows

$$\begin{aligned} \|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)} &\stackrel{(a)}{\leq} \left\| \left( \nabla^2 P_t(\boldsymbol{\theta}^*) \right)^{1/2} \left( \nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}_t) \right)^{-1} \left( \nabla^2 P_t(\boldsymbol{\theta}^*) \right)^{1/2} \right\| \left\| \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}} \\ &\stackrel{(b)}{\leq} cC_2 \sqrt{\frac{\gamma \log(dt)}{t}} \end{aligned} \quad (56)$$

where (a) follows from Lemma 17, and (b) from Lemma 12, (53), and  $c$  is some universal constant.

**Step 5 (Introducing  $\tilde{\mathbf{z}}_t$ ):** Fix a  $\tilde{\mathbf{z}}_t$  between  $\boldsymbol{\theta}^*$  and  $\hat{\boldsymbol{\theta}}_t$ . Apply Taylor's series

$$P_t(\hat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*) = \frac{1}{2} \left( \hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^* \right)^\top \nabla^2 P_t(\tilde{\mathbf{z}}_t) \left( \hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^* \right) \quad (57)$$

Now note that both  $\tilde{\boldsymbol{\theta}}_t$  and  $\tilde{\mathbf{z}}_t$  are between  $\hat{\boldsymbol{\theta}}_t$  and  $\boldsymbol{\theta}^*$ , which implies  $\tilde{\boldsymbol{\theta}}_t \rightarrow \boldsymbol{\theta}^*$  and  $\tilde{\mathbf{z}}_t \rightarrow \boldsymbol{\theta}^*$  since  $\hat{\boldsymbol{\theta}}_t \rightarrow \boldsymbol{\theta}^*$ . By (51) and (56) and applying the concentration inequalities give us

$$\left\| \nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}_t) - \nabla^2 P_t(\boldsymbol{\theta}^*) \right\|_* \leq \rho_t \quad (58)$$

$$\left\| \nabla^2 P_t(\tilde{\mathbf{z}}_t) - \nabla^2 P_t(\boldsymbol{\theta}^*) \right\|_* \leq C_1 \|\tilde{\mathbf{z}}_t - \boldsymbol{\theta}^*\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)} \leq \rho_t \quad (59)$$

where  $\rho_t = c(C_1 C_2 + 2\eta^2 \lambda_1^2) \sqrt{\frac{\gamma \log(dt)}{t}}$ .

**Step 6 (Define  $\mathbf{M}_{1,t}$  and  $\mathbf{M}_{2,t}$ ):** It follows from the inequality (58) that

$$\begin{aligned} \nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}_t) &\preceq (1 + \rho_t) \nabla^2 P_t(\boldsymbol{\theta}^*) \implies \nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}_t) - \nabla^2 P_t(\boldsymbol{\theta}^*) \preceq \rho_t \nabla^2 P_t(\boldsymbol{\theta}^*) \\ &\implies \nabla^2 P_t(\boldsymbol{\theta}^*)^{-1/2} \left( \hat{P}_t(\tilde{\boldsymbol{\theta}}_t) - \nabla^2 P_t(\boldsymbol{\theta}^*) \right) \nabla^2 P_t(\boldsymbol{\theta}^*)^{-1/2} \preceq \rho_t I \\ &\implies \left\| \nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}_t) - \nabla^2 P_t(\boldsymbol{\theta}^*) \right\|_* \leq \rho_t. \end{aligned}$$

Then we can use the inequalities (58) and (59) to show that

$$\begin{aligned} (1 - \rho_t) \nabla^2 P_t(\boldsymbol{\theta}^*) &\preceq \nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}_t) \preceq (1 + \rho_t) \nabla^2 P_t(\boldsymbol{\theta}^*) \\ (1 - \rho_t) \nabla^2 P_t(\boldsymbol{\theta}^*) &\preceq \nabla^2 P_t(\tilde{\mathbf{z}}_t) \preceq (1 + \rho_t) \nabla^2 P_t(\boldsymbol{\theta}^*). \end{aligned}$$

Now we define the two quantities  $\mathbf{M}_{1,t}$  and  $\mathbf{M}_{2,t}$  as follows:

$$\begin{aligned} \mathbf{M}_{1,t} &:= \left( \nabla^2 P_t(\boldsymbol{\theta}^*) \right)^{1/2} \left( \nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}_t) \right)^{-1} \left( \nabla^2 P_t(\boldsymbol{\theta}^*) \right)^{1/2} \\ \mathbf{M}_{2,t} &:= \left( \nabla^2 P_t(\boldsymbol{\theta}^*) \right)^{-1/2} \nabla^2 P_t(\tilde{\mathbf{z}}_t) \left( \nabla^2 P_t(\boldsymbol{\theta}^*) \right)^{-1/2}. \end{aligned}$$

**Step 7 (Lower bound  $P_t(\hat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*)$ ):** Now for the lower bound it follows from Equation (57) that

$$\begin{aligned} P_t(\hat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*) &= \frac{1}{2} \left( \hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^* \right)^\top \nabla^2 P_t(\tilde{\mathbf{z}}_t) \left( \hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^* \right) \\ &= \frac{1}{2} \left( \hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^* \right)^\top \nabla^2 P_t(\boldsymbol{\theta}^*)^{\frac{1}{2}} \nabla^2 P_t(\boldsymbol{\theta}^*)^{-\frac{1}{2}} \nabla^2 P_t(\tilde{\mathbf{z}}_t) \nabla^2 P_t(\boldsymbol{\theta}^*)^{-\frac{1}{2}} \nabla^2 P_t(\boldsymbol{\theta}^*)^{\frac{1}{2}} \left( \hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^* \right) \\ &\stackrel{(a)}{=} \frac{1}{2} \mathbf{u}^\top \mathbf{M}_{2,t} \mathbf{u} \end{aligned}$$

where, in (a) we define the vector  $\mathbf{u} := (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*)^\top \nabla^2 P_t(\boldsymbol{\theta}^*)^{\frac{1}{2}}$ . Now observe from the definition of and then using the min-max theorem we can show that

$$\begin{aligned}
 P_t(\hat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*) &\geq \frac{1}{2} \lambda_{\min}(\mathbf{M}_{2,t}) \mathbf{u}^T \mathbf{u} \\
 &= \frac{1}{2} \lambda_{\min}(\mathbf{M}_{2,t}) \left\| \hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^* \right\|_{\nabla^2 P_t(\boldsymbol{\theta}^*)}^2 \\
 &= \frac{1}{2} \lambda_{\min}(\mathbf{M}_{2,t}) \left\| \nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}_t) (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*) \right\|_{(\nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}_t))^{-1} \nabla^2 P_t(\boldsymbol{\theta}^*) (\nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}_t))^{-1}}^2 \\
 &\geq \frac{1}{2} (\lambda_{\min}(\mathbf{M}_{1,t}))^2 \lambda_{\min}(\mathbf{M}_{2,t}) \left\| \nabla^2 \hat{P}_t(\tilde{\boldsymbol{\theta}}_t) (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*) \right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}^2 \\
 &\stackrel{(a)}{=} \frac{1}{2} (\lambda_{\min}(\mathbf{M}_{1,t}))^2 \lambda_{\min}(\mathbf{M}_{2,t}) \left\| \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}^2
 \end{aligned}$$

where, in (a) we use the eq. (55).

**Step 8:** Define  $I(\mathcal{E})$  as the indicator that the desired previous events hold, which we can ensure with probability greater than  $1 - 2 \left(\frac{1}{dt}\right)^\gamma$ . Then we can show that:

$$\begin{aligned}
 \mathbb{E} [P_t(\hat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*)] &\geq \mathbb{E} \left[ (P_t(\hat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*)) I(\mathcal{E}) \right] \\
 &\geq \frac{1}{2} \mathbb{E} \left[ (\lambda_{\min}(\mathbf{M}_{1,t}))^2 \lambda_{\min}(\mathbf{M}_{2,t}) \left\| \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}^2 I(\mathcal{E}) \right] \\
 &\geq (1 - c' \rho_t) \frac{1}{2} \mathbb{E} \left[ \left\| \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}^2 I(\mathcal{E}) \right] \\
 &= (1 - c' \rho_t) \frac{1}{2} \mathbb{E} \left[ \left\| \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}^2 (1 - I(\text{not } \mathcal{E})) \right] \\
 &\stackrel{(a)}{=} (1 - c' \rho_t) \left( \sigma_t^2 - \frac{1}{2} \mathbb{E} \left[ \left\| \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}^2 I(\text{not } \mathcal{E}) \right] \right) \\
 &\geq (1 - c' \rho_t) \sigma_t^2 - \mathbb{E} \left[ \left\| \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}^2 I(\text{not } \mathcal{E}) \right]
 \end{aligned}$$

where, in (a) we have  $\sigma_t^2 := \left\| \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}^2$ , and  $c'$  is an universal constant.

**Step 9:** Define the random variable  $Z = \left\| \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}$ . With a failure event probability of less than  $2 \left(\frac{1}{dt}\right)^\gamma$  for any  $z_0$ , we have:

$$\begin{aligned}
 \mathbb{E} [Z^2 I(\text{not } \mathcal{E})] &= \mathbb{E} [Z^2 I(\text{not } \mathcal{E}) I(Z^2 < z_0)] + \mathbb{E} [Z^2 I(\text{not } \mathcal{E}) I(Z^2 \geq z_0)] \\
 &\leq z_0 \mathbb{E}[I(\text{not } \mathcal{E})] + \mathbb{E} [Z^2 I(Z^2 \geq z_0)] \\
 &\leq \frac{z_0}{2t^\gamma} + \mathbb{E} \left[ Z^2 \frac{Z^2}{z_0} \right] \\
 &\leq \frac{z_0}{2t^\gamma} + \frac{\mathbb{E} [Z^4]}{z_0} \\
 &\leq \frac{\sqrt{\mathbb{E} [Z^4]}}{t^{\gamma/2}}
 \end{aligned}$$

where  $z_0 = t^{\gamma/2} \sqrt{\mathbb{E} [Z^4]}$ .

**Step 10 (Upper Bound):** For an upper bound we have that:

$$\begin{aligned}\mathbb{E} \left[ P_t(\widehat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*) \right] &= \mathbb{E} \left[ \left( P_t(\widehat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*) \right) I(\mathcal{E}) \right] + \mathbb{E} \left[ \left( P_t(\widehat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*) \right) I(\text{not } \mathcal{E}) \right] \\ &\leq \mathbb{E} \left[ \left( P_t(\widehat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*) \right) I(\mathcal{E}) \right] + \frac{\max_{\boldsymbol{\theta} \in \Theta} (P_t(\boldsymbol{\theta}) - P_t(\boldsymbol{\theta}^*))}{t^\gamma}\end{aligned}$$

since the probability of not  $\mathcal{E}$  is less than  $\frac{1}{t^\gamma}$ . Now for an upper bound of the first term, observe that

$$\begin{aligned}\mathbb{E} \left[ \left( P_t(\widehat{\boldsymbol{\theta}}_t) - P_t(\boldsymbol{\theta}^*) \right) I(\mathcal{E}) \right] &\leq \frac{1}{2} \mathbb{E} \left[ (\lambda_{\max}(\mathbf{M}_{1,t}))^2 \lambda_{\max}(\mathbf{M}_{2,t}) \left\| \nabla \widehat{P}_t(\boldsymbol{\theta}^*) \right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}^2 I(\mathcal{E}) \right] \\ &\leq (1 + c' \rho_t) \frac{1}{2} \mathbb{E} \left[ \left\| \nabla \widehat{P}_t(\boldsymbol{\theta}^*) \right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}^2 I(\mathcal{E}) \right] \\ &\leq (1 + c' \rho_t) \frac{1}{2} \mathbb{E} \left[ \left\| \nabla \widehat{P}_t(\boldsymbol{\theta}^*) \right\|_{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}^2 \right] \\ &= (1 + c' \rho_t) \frac{\sigma_t^2}{t}\end{aligned}$$

where,  $c'$  is another universal constant. □

## A.9 Additional Experiment Details

### A.9.1 Hypothesis Testing Experiments

In all the active testing experiments we use the threshold function for the Gaussian distribution as proved in Lemma 10. Hence the threshold function used is

$$\beta = \log(J/\delta).$$

Note that this threshold function is smaller than the general sub-Gaussian threshold function proved in Lemma 9.

**Example 1:** Recall that the Example 1 is given by the following table under the three different hypotheses  $\{\boldsymbol{\theta}^*, \boldsymbol{\theta}', \boldsymbol{\theta}''\}$

$\boldsymbol{\theta}$	=	$\boldsymbol{\theta}^*$	$\boldsymbol{\theta}'$	$\boldsymbol{\theta}''$
$\mu_1(\boldsymbol{\theta})$	=	1	0.001	0
$\mu_2(\boldsymbol{\theta})$	=	1	1.002	0.998

We can show that under  $p_{\boldsymbol{\theta}^*}$  we have the following optimization problem

$$\begin{aligned}\max z \\ \text{s.t. } p(1)0.999^2 + p(2)0.002^2 &\geq z \\ p(1)1^2 + p(2)0.002^2 &\geq z.\end{aligned}$$

The solution to the above optimization is given by  $p_{\boldsymbol{\theta}^*} = [p(1), p(2)] = [1, 0]$ . Similarly we can show that  $p_{\boldsymbol{\theta}'}$  we have the following optimization problem

$$\begin{aligned}\max z \\ \text{s.t. } p(1)0.001^2 + p(2)0.002^2 &\geq z \\ p(1)0.001^2 + p(2)0.004^2 &\geq z.\end{aligned}$$

The solution to the above optimization is given by  $p_{\boldsymbol{\theta}'} = [p(1), p(2)] = [0, 1]$ . Finally we can show that  $p_{\boldsymbol{\theta}''}$  we have the following optimization problem

$$\begin{aligned}\max z \\ \text{s.t. } p(1)0.001^2 + p(2)0.004^2 &\geq z \\ p(1)1^2 + p(2)0.002^2 &\geq z.\end{aligned}$$

The solution to the above optimization is given by  $p_{\theta''} = [p(1), p(2)] = [0, 1]$ . Hence,  $D_1 := \min_{p_{\theta} \in \Theta} \min_{\theta' \neq \theta^*} \sum_{i=1}^n p_{\theta}(i) (\mu_i(\theta') - \mu_i(\theta^*))^2 = 0.002^2 = 4 \times 10^{-6}$ . Similarly, we can compute that  $D_0 := \max_i \min_{\theta' \neq \theta^*} \sum_{i=1}^n (\mu_i(\theta') - \mu_i(\theta^*))^2 = 0.999^2$ . Hence the non-asymptotic term  $(\log J)/D_1 = 0.3 \times 10^6$  and the asymptotic term  $\log(J/\delta)/D_0 = 3.4$ .

**Active Testing environment (3 Group setting):** In this setting there are three groups of actions. In first group there is a single action that discriminates best between all pair of hypotheses. In second group there are 5 actions which can discriminate one hypotheses from others. Finally in the third group there are 44 actions which cannot discriminate between any pair of hypotheses. The following table describes the  $\mu_1(\cdot), \mu_2(\cdot), \dots, \mu_{50}(\cdot)$  under different hypotheses as follows:

$\theta$	$\theta^*$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_5$	$\theta_6$
$\mu_1(\theta)$	3	0	0	0	0	0
$\mu_2(\theta)$	2	3	2	2	2	2
$\mu_3(\theta)$	2	2	3	2	2	2
$\mu_4(\theta)$	2	2	2	3	2	2
$\mu_5(\theta)$	2	2	2	2	3	2
$\mu_6(\theta)$	2	2	2	2	2	3
$\mu_7(\theta)$	$1 + \iota_{7,1}$	$1 + \iota_{7,2}$	$1 + \iota_{7,3}$	$1 + \iota_{7,4}$	$1 + \iota_{7,5}$	$1 + \iota_{7,6}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\mu_{50}(\theta)$	$1 + \iota_{50,1}$	$1 + \iota_{50,2}$	$1 + \iota_{50,3}$	$1 + \iota_{50,4}$	$1 + \iota_{50,5}$	$1 + \iota_{50,6}$

In the above setting, we define  $\iota_{i,j}$  for the  $i$ -th action and  $j$ -th hypothesis as a small value close to 0 and  $\iota_{i,j} \neq \iota_{i',j'}$  for any pair of hypotheses  $j, j' \in [J]$  and actions  $i, i' \in [n]$ .

### A.9.2 Active Regression Experiment for Non-linear Reward Model

**Algorithmic Details:** We describe each of the algorithm used in this setting as follows:

1. **EMCM:** The EMCM algorithm of Cai et al. (2016) first quantifies the change as the difference between the current model parameters and the new model parameters learned from enlarged training data, and then chooses the data examples that result in the greatest change.
2. **CS:** The CS policy used is stated as in Section 3. To calculate the least square estimate  $\hat{\theta}(t)$  we use the python `scipy.optimize` least-square function which solves a nonlinear least-squares problem.
3. **Unif:** The Unif policy samples each action uniform randomly at every round.
4. **ActiveS:** The ActiveS policy in Chaudhuri et al. (2015) is a two-stage algorithm. It first samples all actions uniform randomly to build an initial estimate of  $\theta^*$ . It then solves an Semi-definite Programming (SDP) to obtain a new sampling distribution that minimizes the quantity  $\sigma_U^2$  as defined in Equation (46). In the second stage ActiveS follows this new sampling distribution to sample actions.

**Implementation Details:** This setting consist of 50 measurement actions divided into three groups. The first group consist of the optimal action  $\mathbf{x}_{i^*} := (1, 0)$  in the direction of  $\theta^* := (1, 0)$ . The second group consist of the informative action  $\mathbf{x}_2 := (0, 1)$  which is orthogonal to  $\mathbf{x}_{i^*}$  and selecting it maximally reduces the uncertainty of  $\hat{\theta}(t)$ . Finally the third group consist of 48 actions such that  $\mathbf{x}_i := (0.71 \pm \iota_i, 0.71 \mp \iota_i)$  for  $i \in [3, 50]$  where  $\iota_i$  is a small value close to 0 and  $\iota_i \neq \iota_{i'}$ . Note that these 48 actions are less informative in comparison to action 2. This is shown in Figure 1c.

### A.9.3 Active Regression Experiment for Neural Networks

**Implementation Details:** At every time step, we use the the least squares optimizer of `scipy` to find  $\hat{\theta}_t$ . Since  $c_1, c_2 \in \{-1, 1\}$ , we solve four different least squares problems at each step corresponding to all  $(c_1, c_2)$  choices, and use the values returned by the problem having the smallest sum of squares as our current estimate for  $(\mathbf{w}_1, \mathbf{w}_2, b_1, b_2)$ . The derivative with respect to any parameter is found by the backward pass of automatic differentiation.

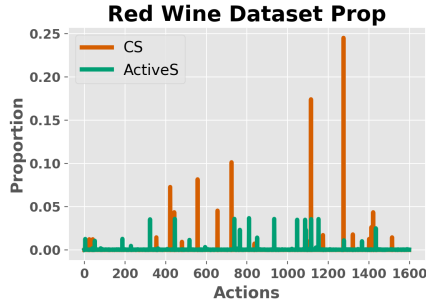


Figure 2: CS Proportions over 1600 actions in Red Wine Dataset. Note that CS Proportion is sparse.

#### A.9.4 Active Regression for the UCI Datasets

**Implementation Details:** The UCI Red Wine Quality dataset consist of 1600 samples of red wine with each sample  $i$  having feature  $\mathbf{x}_i \in \mathbb{R}^{11}$ . We first fit a least square estimate to the original dataset and get an estimate of  $\boldsymbol{\theta}^*$ . The reward model is linear and given by  $\mathbf{x}_{I_t}^T \boldsymbol{\theta}^* + \text{noise}$  where  $x_{I_t}$  is the observed action at round  $t$ , and the noise is a zero-mean additive noise. Note that we consider the 1600 samples as actions. Then we run each of our benchmark algorithms on this dataset and reward model and show the result in Figure 1g. We further show the CS proportion on this dataset in fig. 2 and show that it is indeed sparse with proportion concentrated on few actions. The Air quality dataset consist of 1500 samples each of which consist of 6 features. We again build an estimate of  $\boldsymbol{\theta}^*$  by fitting a least square regression on this dataset. We use a similar additive noise linear reward model as described before and run all the benchmark algorithms on this dataset.



## A.10 Table of Notations

Notation	Definition
$n$	Total number of actions
$J$	Total number of hypotheses
$\Theta$	Parameter Space
$\mu_i(\boldsymbol{\theta})$	Mean of action $i$ under hypothesis $\boldsymbol{\theta}$
$\pi$	Policy
$\delta$	Probability of error of $\delta$ -PAC policy
$\tau_\delta$	Stopping time of $\delta$ -PAC policy
$\beta(J, \delta)$	$\log(CJ/\delta)$ , $C$ is a constant depending on $\eta, \eta_0$
$\alpha(J)$	$b \log(J)$ , $b > 0$
$Y^t$	Vector of rewards observed till round $t$
$I^t$	Vector of actions sampled till round $t$
$I_s$	action sampled at round $s$
$Z_i(t)$	Number of time action $i$ is sampled till round $t$
$\eta$	Constant $> 0$ s.t. $Y_s \in [-\eta/2, \eta/2]$
$\eta_0$	$\min_{i \in [n]} \min_{\boldsymbol{\theta} \neq \boldsymbol{\theta}'} (\mu_i(\boldsymbol{\theta}) - \mu_i(\boldsymbol{\theta}'))^2$
$\mathbf{p}_\theta$	p.m.f. to verify hypothesis $\boldsymbol{\theta}$ (Solution to Chernoff optimization in (4))
$\text{KL}(\cdot    \cdot)$	Kullback-Leibler divergence
$\hat{\boldsymbol{\theta}}(t)$	Most likely hypothesis at round $t$
$\tilde{\boldsymbol{\theta}}(t)$	Second most likely hypothesis at round $t$
$L_t(\boldsymbol{\theta})$	Sum of squared errors till round $t$ under hypothesis $\boldsymbol{\theta}$
$\ell_s(\boldsymbol{\theta})$	Squared error at a specific round $s$ under hypothesis $\boldsymbol{\theta}$
$\Delta_t(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$	$L_t(\boldsymbol{\theta}) - L_t(\boldsymbol{\theta}^*)$
$\Delta_s(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$	$\ell_s(\boldsymbol{\theta}) - \ell_s(\boldsymbol{\theta}^*)$ at a specific round $s$
$\xi^\delta(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$	Event that $\{L_{\tau_\delta}(\boldsymbol{\theta}') - L_{\tau_\delta}(\boldsymbol{\theta}) > \beta(J, \delta), \forall \boldsymbol{\theta}' \neq \boldsymbol{\theta}\}$
$(1+c)M$	Critical number of samples $(1+c)O(\log J/D_1 + \log(J/\delta)/D_0)$ , for a constant $c > 0$
$\mathcal{I}$	$\{i \in [n] : i = \arg \max_{i' \in [n]} (\mu_{i'}(\boldsymbol{\theta}) - \mu_{i'}(\boldsymbol{\theta}'))^2 \text{ for some } \boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta\}$
$\gamma$	Constant $\geq 2$ , controlling the convergence rate
$d$	Dimension of the parameter space $\Theta$
$C$	$165 + \eta^2/\eta_0^2$
$D_0$	$\max_{\mathbf{p}} \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n p(i) (\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2$
$D_1$	$\min_{\{\mathbf{p}_\theta : \theta \in \Theta\}} \min_{\boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n p_\theta(i) (\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2$
$D'_0$	$\min_{\boldsymbol{\theta}, \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n u_{\boldsymbol{\theta}^* \boldsymbol{\theta}}(i) (\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2$
$D'_1$	$\min_{\boldsymbol{\theta} \neq \boldsymbol{\theta}', \boldsymbol{\theta}' \neq \boldsymbol{\theta}^*} \sum_{i=1}^n u_{\boldsymbol{\theta} \boldsymbol{\theta}'}(i) (\mu_i(\boldsymbol{\theta}') - \mu_i(\boldsymbol{\theta}^*))^2$
$P_t(\boldsymbol{\theta})$	$\frac{1}{t} \sum_{s=1}^t \mathbb{E}_{I_s \sim \mathbf{p}_{\hat{\boldsymbol{\theta}}_{s-1}}} [L_s(\boldsymbol{\theta})   \mathcal{F}^{s-1}]$
$\mathbf{P}_{\text{unif}}$	pmf of a uniform distribution over the actions.
$P_U(\boldsymbol{\theta})$	$\mathbb{E}_{I_s \sim \mathbf{P}_{\text{unif}}} [L_s(\boldsymbol{\theta})]$
$\sigma_t^2$	$\mathbb{E} \left\  \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right\ _{(\nabla^2 P_t(\boldsymbol{\theta}^*))^{-1}}^2$
$\overline{\sigma}_t^2$	$\mathbb{E} \left\  \nabla \hat{P}_t(\boldsymbol{\theta}^*) \right\ _{(\nabla^2 P_U(\boldsymbol{\theta}^*))^{-1}}^2$
$\sigma_U^2$	Trace $\left[ I_U(\boldsymbol{\theta}^*) I_\Gamma(\boldsymbol{\theta}^*)^{-1} \right]$ where $I_U$ and $I_\Gamma$ are Fisher Information matrices.

Table 3: Table of Notations