The Good, the Bad, and the Sampled: a No-Regret Approach to Safe Online Classification

Tavor Z. Baharav[1,2,*,†]
[1] Eric and Wendy Schmidt Center, Broad Institute
[2] Department of Data Science, Dana Farber Cancer Institute

Spyros Dragazis[3,*]
[3] Department of Computer Science, Boston University

Aldo Pacchiano[1,4,†]
[1] Eric and Wendy Schmidt Center, Broad Institute
[4] Faculty of Computing & Data Sciences, Boston University

Abstract

We study sequential testing for a binary disease outcome when risk follows an unknown logistic model. At each round, the decision maker may either pay for a test revealing the true label or predict the outcome based on patient features and past data. The goal is to minimize costly tests while ensuring the misclassification rate stays below α with probability at least $1-\delta$. We propose a method that jointly estimates the logistic parameter $\theta^\star$ and the feature distribution, using a conservative threshold on the logistic score to decide when to test. We prove our procedure achieves the target error with high probability and requires only $\widetilde{O}(\sqrt{T})$ more tests than an oracle with full knowledge. This is the first no-regret guarantee for error-constrained logistic testing, with direct applications to medical screening. Simulations corroborate our theoretical results, showing safe classification of patients and efficient estimation of $\theta^\star$ with few excess tests.

1 INTRODUCTION

Modern machine learning has recently provided solutions to real-world automated decision-making systems in various fields such as drug discovery Vamathevan et al. (2019); Dara et al. (2022), recommendation systems Afsar et al. (2022); Zhu and Van Roy (2023), online ad-allocation Slivkins (2013), and portfolio selection Pinelis and Ruppert (2022). Bandit algorithms Lattimore and Szepesvári (2020) and reinforcement learning Sutton et al. (1999) play a significant role in building interactive decision-making systems that collect feedback from users and improve their performance with each interaction. Two primary challenges exist in the aforementioned applications: the first is the learning challenge, estimating the problem parameters which are vital for decision-making; the second is the decision-making challenge, where effective performance is required concurrently with learning.

Although machine learning systems perform well in practice, safety is crucial in human-centric applications Gu et al. (2022); Giudici (2024). In sequential decision making, safety has been formalized in several ways, including (i) constraining actions to a safe set defined by a cost signal Pacchiano et al. (2021); Yao et al. (2021); Gangrade et al. (2024); Camilleri et al. (2022), (ii) conservative bandits, which require performance comparable to a baseline Kazerouni et al. (2017), and (iii) enforcing explicit performance threshold constraints on the learning process itself Bar-El Avidan and Bistritz (2025). Satisfying such criteria generally complicates reward acquisition, making the central challenge one of optimally balancing safety and learning.

Motivated by the COVID-19 pandemic, we study an online learning problem with a novel safety constraint. Patients arrive sequentially, each presenting with a feature vector (e.g., fever, fatigue, blood oxygen) and an unobserved disease state. The hospital aims to minimize costly tests while ensuring sick patients are properly quarantined. We assume an unknown logistic model links features to disease status, allowing the hospital to learn, for example, that low oxygen and high fever imply high risk without testing. Thus, the task is to simultaneously learn (a) the patient distribution, (b) the logistic model parameters, and (c) the testing threshold.

Related problems have been studied in the active learning and selective sampling literature Settles (2009);

[*]Equal contribution [†]Equal senior contribution.

Proceedings of the 29th International Conference on Artificial Intelligence and Statistics (AISTATS) 2026, Tangier, Morocco. PMLR: Volume 300. Copyright 2026 by the author(s).

Hanneke and Yang (2021); Orabona et al. (2011); Cesa-Bianchi et al. (2006); Freund et al. (1997); Seung et al. (1992); Dekel et al. (2012), which study a similar observation model and generalization error (regret) metric but without a safety constraint. These study settings where context information may be abundant but the labels are hard to come by Duan et al. (2023).

Focusing on the classification task and changing the objective from minimizing the generalization error to minimizing the cumulative pseudo regret (with respect to the optimal labeling policy), various algorithms have been developed in the online selective sampling literature, such as Orabona et al. (2011); Sekhari et al. (2023), considering both stochastic and adversarial contexts. The objective in these works is to achieve sublinear regret while minimizing the expected number of queries made. Similarly, online selective classification Gangrade et al. (2021a,b); Goel et al. (2023) allows the learner to abstain from predicting. The objective is to simultaneously minimize the expected number of abstentions and mistakes.

However, in real-world scenarios (Bastani et al., 2022), it makes sense to require that the prediction error rate remain under a safety threshold with high probability while minimizing the number of queries. For example in the streaming patient scenario we described above, where patients arrive one by one and the medical provider needs to classify them as sick or not. In this problem, due to the sensitive nature of making misclassification mistakes, the selective testing procedure must guarantee that the misclassification rate remains below a safety threshold $\alpha \in [0, 1]$. Testing every patient clearly attains this safety threshold, but can be prohibitively expensive. Our question is thus:

Can we design an adaptive algorithm that minimizes the expected number of tests while maintaining a misclassification rate below a given safety threshold?

We define a baseline testing policy, that is optimal when the α error rate is only required to hold in expectation, which tests $p^\star \triangleq p^\star(\alpha)$ fraction of the time. We develop an adaptive algorithm to ensure this α error rate with probability at least $1 - \delta$, which requires only a sublinear number of excess tests: $\mathcal{O}\left(\sqrt{\frac{dT}{p^\star(\alpha)\lambda_0}\log(T/\delta)}\right)$, where λ_0 is the minimum eigenvalue of the covariance matrix of the contexts observed under the baseline policy. In Lemma 1, we derive a lower bound on λ_0 in terms of the problem parameters (additional discussion in Appendix A.2). In the special case where the context distribution is uniform over the d-dimensional sphere, $\lambda_0 = \Theta(1/d)$, recovering the standard $\tilde{\mathcal{O}}(d\sqrt{T})$ regret rate for linear bandits (Lattimore and Szepesvári, 2020). We corroborate our theoretical results through comprehensive synthetic experiments.

2 RELATED WORK

The most related line of work to ours is *online learning with abstention* Cortes et al. (2018, 2016), where the learner may abstain from labeling a context $x \in \mathcal{X}$ by incurring an abstention cost c. In particular, Cortes et al. (2018) studies both the stochastic and adversarial settings, with the goal of minimizing regret; the resulting guarantees depend, in addition to the standard problem parameters such as the horizon T, on the abstention cost c. The key difference from our work is that, while we also aim to reduce the cost induced by abstention, we simultaneously control the misclassification rate so that it remains below a prescribed target level throughout the entire process.

Another relevant line of work is *online selective classification* Gangrade et al. (2021b,a); Sekhari et al. (2023), where the learner may abstain from releasing a prediction and instead observe the true outcome. The goal in this setting is to jointly minimize the number of mistakes and the number of label queries. Although this framework is the closest to ours, existing guarantees do not enforce uniform control of the cumulative misclassification rate throughout the learning process. This distinction is especially important in applications such as medicine, where maintaining the error rate below a prescribed threshold at all times is often essential.

Moving away from regret minimization, another related direction in the bandit literature is best-arm identification under selective sampling Camilleri et al. (2021, 2022). In Camilleri et al. (2021), the authors precisely characterize the trade-off between the number of observed labels and the stopping time, and provide a geometric interpretation of the optimal decision rule. In Camilleri et al. (2022), they study the problem of actively selecting the covariates or contexts whose labels are to be observed, while ensuring that the selected covariates lie within a safe set.

Finally, the recent literature on PAC-labeling Candès et al. (2025) addresses a closely related problem from a different perspective. In that setting, one assumes access to an AI model that predicts the labels of an unlabeled dataset. For each prediction Y_i, the "expert" model also provides an uncertainty score U_i. The algorithmic challenge is then to leverage these uncertainty scores to produce "PAC labels", or, in our terminology, to satisfy (α, δ)-safety.

Our work is also related to the literature on learning halfspaces under label noise. In that line of work, the objective is to learn a linear separator with small classification error from noisy labeled data, either in the batch or online setting. Recent results study efficient learning under Tsybakov noise Tsybakov (2004); Diakonikolas et al. (2021) and online learning

under Massart noise Diakonikolas et al. (2024). These works focus on statistical and computational guarantees for recovering an accurate classifier; by contrast, our goal is different. We study a sequential decision problem in which the learner must decide, for each context, whether to query the label or to predict, while maintaining a uniform high-probability bound on the cumulative misclassification rate and simultaneously minimizing the number of queries.

A common assumption in the halfspace literature is the Tsybakov noise condition. Writing $\eta(x) = \mathbb{P}(Y = 1 \mid X = x)$, this condition controls the probability mass of examples whose conditional label probability lies close to the ambiguous value $1/2$. Intuitively, it quantifies how much probability mass concentrates near the Bayes decision boundary, and therefore how hard it is to identify an accurate classifier. Under this perspective, the difficulty of learning is governed by the Tsybakov parameters. In our setting, however, points near the decision boundary are precisely the ones that must be tested in order to guarantee safety. As a result, the relevant notion of difficulty is not how sharply the distribution separates around the boundary, but rather the baseline testing rate $p^\star$, namely the mass of contexts on which even the oracle safe policy must abstain from autonomous prediction.

Another common assumption in active learning for halfspaces is that the unlabeled contexts follow a particularly structured distribution, such as the uniform distribution on the unit sphere Dasgupta et al. (2005). In contrast, our analysis only requires the context distribution to satisfy the regularity condition of Assumption 2. This allows a broader family of distributions, including smooth radial densities of the form $f(x) = g(\|x\|)$ and truncated Gaussian distributions. These assumptions are used to guarantee stability of the threshold policy and positivity of the covariance matrix of the queried contexts, rather than to enable efficient recovery of a separator from label noise alone.

3 PRELIMINARIES

Notation We adopt the following notation throughout the paper. The inner product between two vectors $x, y \in \mathbb{R}^d$ will be denoted either as $x^\top y$ or as $\langle x, y\rangle$. We denote the ℓ_2 norm of a vector $x \in \mathbb{R}^d$ as $\|x\|_2 = \sqrt{\langle x, x\rangle}$ and $\|x\|_A = \sqrt{x^\top A x}$ for any positive semi-definite matrix A. The minimum eigenvalue of a matrix A will be denoted as $\lambda_{\min}(A)$. The set $\{1, 2, \ldots, n\}$ is denoted as $[n]$. The logistic function is denoted as $\mu(z) = \frac{1}{1+\exp(-z)}$ and $\mathbb{1}(E)$ denotes the indicator function of an event E. For two functions f, g we say that $f(x) \preccurlyeq g(x)$ when there exists an absolute constant $c > 0$ such that $f(x) \le cg(x)$ for all $x > 0$. We use upper case letters for random variables and lower case for scalars. For any measurable set A we denote the set of all distributions on A as $\Delta(A)$. An $\mathcal{L}_2$ ball centered at $\mathbf{c} \in \mathbb{R}^d$ with radius $r > 0$ is symbolized as $\mathcal{B}(\mathbf{c}, r)$.

3.1 Problem Definition

We consider the following repeated interaction between a learner and the environment. At every round $t \in [T]$, the environment generates a context $X_t \in \mathbb{R}^d$ in the unit ball. These contexts are identically distributed, and are drawn independently from an unknown distribution with density P. Every patient-context has an unseen random label $Y_t \in \{0, 1\}$ that represents their disease status. We assume that $Y_t \sim \mathrm{Ber}(\mu(X_t^\top \theta^\star))$, independent from all other $X_{t'}$ and $Y_{t'}$. Here, $\theta^\star \in \mathbb{R}^d$ is some fixed parameter vector unknown to the learner, with $\|\theta^\star\|_2 = 1$.

At each round, the learner observes the patient's context X_t and must decide whether or not to test the patient, denoted by $Z_t \in \{0, 1\}$. Then, the learner must predict whether the patient is healthy or sick, denoted by $\hat{Y}_t \in \{0, 1\}$. If $Z_t = 1$, the patient is tested, and the learner observes the true label Y_t, and so can predict $\hat{Y}_t = Y_t$. The random variable Z_t can depend on information obtained prior to that decision, i.e., $\mathcal{H}_t = \{X_1, Z_1, Z_1Y_1, X_2, Z_2, Z_2Y_2, \ldots, X_t\}$ and possibly on internal randomization of the learner. Similarly, $\hat{Y}_t$ must be $\mathcal{F}_t = \sigma\{X_1, Z_1Y_1, X_2, Z_2Y_2, \cdots, X_t, Z_tY_t\}$ measurable. The goal of the learner is to minimize the expected number of tests applied, while guaranteeing that the misclassification rate is less than a desired threshold α, with probability at least $1 - \delta$. We define this constraint as (α, δ)-safety, where our objective is to minimize the expected number of tests required while retaining this (α, δ)-safety.

Definition 1. *An algorithm outputting $\{\hat{Y}_t\}$ satisfies (α, δ)-safety if*

$$\mathbb{P}\left(\bigcap_{\bar{T}=1}^{T}\left\{\frac{1}{\bar{T}}\sum_{t=1}^{\bar{T}}\mathbb{1}\{\hat{Y}_t \neq Y_t\} \le \alpha\right\}\right) \ge 1 - \delta.$$

where the probability is computed with respect to the randomness in $\{X_t\}, \{Y_t\}$, and any randomness internal to the algorithm.

3.2 Baseline policy

First, we characterize the baseline testing strategy satisfying (α, δ)-safety in the case where the feature distribution P and optimal discriminator $\theta^\star$ are known a priori to the learner. Although many decision rules Z_t are possible, we focus on threshold rules of the form below (Equation (1)).

$$Z_t = \mathbb{1}\{|\langle X_t, \theta^\star\rangle| \le \tau\}$$
$$\hat{Y}_t = \begin{cases} 0 & \text{if } \langle X_t, \theta^\star\rangle < -\tau, \\ Y_t & \text{if } |\langle X_t, \theta^\star\rangle| \le \tau, \\ 1 & \text{if } \langle X_t, \theta^\star\rangle > \tau. \end{cases} \tag{1}$$

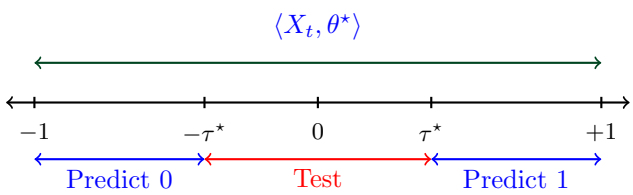


Figure 1: Threshold-based testing policy.

When P and $\theta^\star$ are known, a threshold decision rule is optimal when the safety constraint is imposed only in expectation, as we show in the following proposition.

Proposition 1. *Consider a variant of safe learning where the constraint is only required to hold in expectation, at the final time step:*

$$\min_{\{Z_t, \hat{Y}_t\}} \mathbb{E}\left[\sum_{t=1}^{T} Z_t\right] \quad s.t. \quad \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{\hat{Y}_t \neq Y_t\}\right] \le \alpha.$$

Then, an optimizing rule for $\hat{Y}_t, Z_t$ is the threshold policy Equation (1).

The proof of Proposition 1 follows by relating this to the fractional knapsack problem, which we detail in Appendix A. We provide additional discussion on how this does not naively yield (α, δ)-safety, but still motivates the use of a threshold policy as a baseline. As a consequence, we consider competing against the optimal threshold decision rule $\tau^\star$ that is a function of P, $\theta^\star$, and α, henceforth referred to as the baseline policy.

To identify the optimal threshold, we define the function $p_{\text{err}}(\theta, P, \tau)$ as the probability of misclassification incurred by the threshold τ, if θ was the underlying logistic parameter, and where the expectation is taken with respect to P:

$$p_{\text{err}}(\theta, P, \tau) \tag{2}$$
$$= \int (1 + \exp(|x^\top\theta|))^{-1} \mathbb{1}\left\{|x^\top\theta| > \tau\right\} P(dx).$$

The term inside the integral $(1 + \exp(|x^\top\theta|))^{-1}$ is the optimal misclassification error for a fixed x, θ pair. The term $\mathbb{1}\left\{|x^\top\theta| > \tau\right\}$ equals one only if we predict the label $\hat{y}$ without observing the real label y for context x, when using a threshold rule. Having defined the error probability for a given threshold τ, we can now easily define the optimal threshold. For any problem parameters $\theta \in \mathbb{R}^d, \alpha' \in [0, 1]$, and distribution $\rho \in \Delta(\mathcal{X})$, we define the optimal decision threshold $\tau^\star$ as the minimum value of $\tau \in [0, 1]$ that satisfies the α-fraction misclassification constraint:

$$\tau^\star(\theta, \rho, \alpha') \triangleq \min\{\tau : p_{\text{err}}(\theta, \rho, \tau) \le \alpha'\}. \tag{3}$$

When considering the in-expectation objective from Proposition 1 we conclude that any algorithm requires an expected number of tests $p^\star T$:

$$\tau^\star \triangleq \tau^\star(\theta^\star, P, \alpha), \; p^\star \triangleq \mathbb{P}\left(x : |x^\top\theta^\star| \le \tau^\star\right), \tag{4}$$

where $\theta^\star, P$, and α are the true parameters. Here, we have overloaded notation for $\tau^\star$ as both a function, and the evaluation of this function at the true problem parameters. Note that in practice, p_{err} must be estimated using $\hat{P}$, our observed samples from P, in addition to $\theta^\star$ being unknown.

Before introducing our regret objective, we examine the relationship between the safety parameter α, which serves as an input, and the baseline policy testing probability $p^\star$. When the misclassification rate threshold α approaches zero, the system must minimize error rates, necessitating testing of all cases. This constraint leads to increased values of $\tau^\star$ and, consequently, higher values of $p^\star$. Conversely, in the degenerate scenarios where α grows large, policies become indifferent to misclassification errors and conduct vanishing testing, yielding values of $p^\star$ that approach zero.

This lets us define the "safe regret" of an algorithm as the number of excess tests it takes over this oracle baseline, while satisfying (α, δ)-safety. Formally, the regret is $\texttt{Regret}(T) \triangleq \mathbb{E}[\sum_{t=1}^{T}(Z_t - p^\star)]$. An algorithm could trivially sample at each time step and satisfy the misclassification criterion; the question is, for a given misclassification rate α and error probability δ, can a learner achieve sublinear safe regret in T. To analyze this quantity, we make the following natural assumptions.

Assumption 1. *The optimal baseline tests a nonzero fraction of the time, i.e. $p^\star > 0$.*

Other works such as, Orabona et al. (2011), Sekhari et al. (2023), use the notation T_ε to describe the number of times the Bayes optimal classifier outputs a label with confidence less than a fixed parameter $\varepsilon > 0$. Our $p^\star$ is analogous to T_ε: it serves as a measure to quantify the inherent difficulty of the problem instance (how many patients are close to the decision boundary). We additionally assume that the density P is smooth, which is reasonable for patient data with continuous-valued features.

Assumption 2. *The density P is upper and lower bounded by constants: $0 < m \le P(x) \le M < \infty$, for all x such that $\|x\|_2 \le 1$.*

This is necessary for ensuring the stability of our estimates of $\tau^\star$ with respect to small perturbations in θ, $\hat{P}$, and α. Using Assumption 2 we derive the following result regarding the minimum eigenvalue of the covariance matrix of the baseline policy. This lemma ensures that $\theta^\star$ can be well estimated from the observed data. We refer the reader to Section 2 for a detailed discussion of analogous assumptions and problem formulations in the literature.

Lemma 1. *There is a constant $\lambda_0^{\min}(\tau^\star, d) > 0$:*

$$\lambda_{\min}\left(\mathbb{E}_P\left[XX^\top \mid |\langle X, \theta^\star\rangle| \le \tau^\star\right]\right) = \lambda_0 \ge \lambda_0^{\min}(\tau^\star, d) > 0.$$

As $\|\theta^\star\|_2 = 1$, a ball of radius $\tau^\star$ is a subset of the contexts tested by the baseline policy. The contexts drawn from this ball form a positive definite covariance matrix, which implies that the minimum eigenvalue of the overall covariance matrix is positive. We defer the proof to Appendix A.2.

Importantly, these assumptions are strictly for the *analysis* of our algorithm. Our algorithm does not require knowledge of any of these parameters m, M, λ_0, or $p^\star$ as input. We are able to learn and adapt to them on the fly, they simply requiring them to be strictly positive and finite.

3.3 Logistic Bandits tools

Our algorithm leverages existing confidence intervals for $\theta^\star$ Faury et al. (2020). We utilize their ellipsoidal confidence set to simplify our analysis, noting that tighter confidence intervals exist Lee et al. (2025). In our setting, the non-linearity of the logistic function over the decision set $(\mathcal{X}, \Theta)$ is bounded as $\kappa \le 6$. Borrowing notation Faury et al. (2020), we denote the set of labeled samples $\big((X_t, Y_t)$ pairs$\big)$ collected up to the beginning of round t which are used to estimate $\theta^\star$ by $\mathcal{S}_\theta^t$, and the nonoverlapping set of samples (only the context, X_t) used to estimate the distribution P by $\mathcal{S}_P^t$. We denote the cardinalities of these two sets by N_θ^t and N_P^t respectively. We define the regularized log-likelihood objective as:

$$\mathcal{L}_t(\theta) = \sum_{s \in \mathcal{S}_\theta^t} \ell(x_s, y_s, \theta) - \frac{1}{2}\|\theta\|_2^2,$$

where $\ell(x, y, \theta) = y \log \mu(x^\top\theta) + (1-y)\log(1-\mu(x^\top\theta))$, and its maximum (regularized) likelihood estimator as $\hat{\theta}_t = \operatorname{argmax}_{\theta \in \mathbb{R}^d} \mathcal{L}_t(\theta)$. We also denote the design matrix as $V_t = \sum_{s \in \mathcal{S}_\theta^t} X_s X_s^\top + \kappa \mathbf{I}_d$, and for technical reasons we consider a projection θ_t^L of $\hat{\theta}_t$ onto the feasible set Θ defined as follows,

$$\theta_t^L \triangleq \operatorname*{argmin}_{\theta \in \Theta} \left\| g_t(\theta) - g_t(\hat{\theta}_t) \right\|_{V_t^{-1}}, \text{ where}$$
$$g_t(\theta) = \sum_{s \in \mathcal{S}_\theta^t} \mu(\langle x_s, \theta\rangle) x_s + \theta. \tag{5}$$

These allow us to define the confidence ellipsoid $\mathcal{C}_t$ for $\theta^\star$, which is implicitly a function of a confidence parameter δ', and its radius $B_t(\delta')$:

$$\mathcal{C}_t \triangleq \left\{\theta \in \Theta, \left\|\theta - \theta_t^L\right\|_{V_t} \le B_t(\delta')\right\}, \tag{6}$$
$$B_t(\delta') \triangleq 2\kappa\left(1 + \sqrt{\log\left(\frac{1}{\delta'}\right) + 2d\log\left(1 + \frac{N_\theta^t}{\kappa d}\right)}\right).$$

We omit the dependence of quantities like B_t on the confidence level δ' when clear from context. In the end we will designate $\delta' = \delta/7$ to obtain the desired result via a union bound. These confidence intervals Faury et al. (2020) satisfy the following anytime guarantees:

Lemma 2. *[Lemma 12 of Faury et al. (2020).] For any fixed δ', let G_θ be the event that the confidence intervals defined in Equation* (6) *hold:*

$$\mathbb{P}(G_\theta) = \mathbb{P}\left(t \in \mathbb{N},\ \theta^\star \in \mathcal{C}_t\right) \ge 1 - \delta'.$$

Since the number of samples N_θ^t collected to estimate $\mathcal{C}_t$ is a random variable in our setting, we condition on its value in Lemma 2.

Before diving into our algorithm and its analysis, we discuss the role and behavior of key quantities that will arise. To begin, the number of samples collected N_θ^t used to build our confidence intervals grows linearly in t satisfying $N_\theta^t \succcurlyeq p^\star t$. As a consequence, the bound B_t used in $\mathcal{C}_t$ (which satisfies $B_t \le B_T$) grows extremely slowly in t, with $B_t \preccurlyeq \sqrt{d\log(1 + \frac{p^\star t}{d})}$. The other portion of the confidence interval involves upper bounding $\|x\|_{V_t^{-1}}$. The lower bound on N_θ^t and Lemma 1 yield that $\|x\|_{V_t^{-1}} \preccurlyeq 1/\sqrt{t\lambda_0}$. Note that $\lambda_{\min}^t$ is computable from the observed data, obviating knowledge of λ_0. This enables us to prove a regret upper bound without using the elliptical potential lemma, as is done in many prior works in Online Logistic Regression Cesa-Bianchi and Lugosi (2006) or in Linear Bandits Abbasi-Yadkori et al. (2011).

4 ALGORITHM DESIGN

The pseudo-code of our algorithm `SCOUT` (Safe Contextual Online Understanding with Thresholds) is presented in Algorithm 1. `SCOUT` tests a patient ($Z_t = 1$) if the inner product between their context X_t and the current estimate θ_t^L has a magnitude smaller than an

Algorithm 1 SCOUT

Input: Number of rounds T, target error rate α, confidence level δ
Initialize: $\mathcal{S}_P^{(1)} = \emptyset$, $\mathcal{S}_\theta^{(1)} = \emptyset$. Maintain $N_P^t = |\mathcal{S}_P^t|$, $N_\theta^t = |\mathcal{S}_\theta^t|$
for $t = 1, 2, \ldots, T$ **do**
  Observe context X_t
  if $t \leq 2$ **then**
    Set $Z_t = 1$
  else
    Compute θ_t^L from (5) and τ_t from (8)
    Set $Z_t = \mathbb{1}\{|\langle \theta_t^L, X_t \rangle| \leq \tau_t\}$
  end if
  if $Z_t = 1$ **then**
    Observe Y_t
    Predict $\hat{Y}_t = Y_t$
  else
    Predict $\hat{Y}_t = \mathbb{1}\{\langle X_t, \theta_t^L \rangle > 0\}$
  end if
  if $Z_t = 1$ and t is even **then**
    Set $\mathcal{S}_\theta^{t+1} = \mathcal{S}_\theta^t \cup \{(X_t, Y_t)\}$
  end if
  if t is odd **then**
    Set $\mathcal{S}_P^{t+1} = \mathcal{S}_P^t \cup \{X_t\}$
  end if
end for

estimator τ_t of the true threshold $\tau^\star$. To iteratively refine the estimates of $\theta^\star$ and $\tau^\star$, SCOUT employs a classical sample-splitting trick to avoid dependencies. The context distribution P is estimated as $\hat{P}_t$, the empirical distribution of contexts observed from odd samples, $\mathcal{S}_P^t$, enabling estimation of $\tau^\star$. $\theta^\star$ is estimated as θ_t^L, using labeled data from even samples where a test was performed, $\mathcal{S}_\theta^t$.

The testing condition $Z_t \triangleq \mathbb{1}\{|\langle X_t, \theta_t^L \rangle| \leq \tau_t\}$ is computed as follows (we defer additional discussion to Section 5). Recall that θ_t^L is the MLE defined in Equation (5), $\hat{P}_t$ is the empirical distribution of the contexts, and $\lambda_{\min}^t \triangleq \lambda_{\min}(V_t)$.

$$\zeta_t(\delta') \triangleq \sqrt{\frac{(d+1)\log(1/\varepsilon_Q) + \log\left(\frac{\pi^2 t^2}{\delta'}\right)}{4t}}, \tag{7}$$

$$\tau_t \triangleq \tau^\star\left(\theta_t^L, \hat{P}_t, \alpha_t - \zeta_t - 2B_t/\sqrt{\lambda_{\min}^t} - \varepsilon_Q\right) + 3B_t/\sqrt{\lambda_{\min}^t} + \varepsilon_Q \tag{8}$$

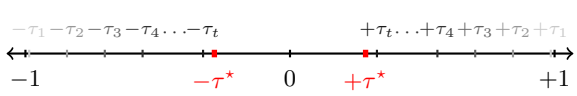


Figure 2: Pessimistic choice of $|\tau_t|$.

Our testing threshold τ_t is designed to be systematically pessimistic. We begin with a plug-in estimator of the optimal threshold as $\tau^\star(\theta_t^L, \hat{P}_t, \alpha)$. To guarantee safety, we inflate our threshold to account for estimation errors. First, we reduce α to $\alpha_t = \max(0, \alpha - \sqrt{\log(2t^2/\delta')/2t})$ (discussed in Appendix C.4) to guarantee (α, δ)-safety, if the true $\theta^\star$ and P were known. We set δ' in Theorem 1 as $\delta' = \delta/7$. Then, we reduce α further by ζ_t (implicitly, $\zeta_t(\delta')$) to account for the fact that P is unknown and we only have $\hat{P}_t$. Most critically, we add buffer terms proportional to $B_t/\sqrt{\lambda_{\min}^t}$, which tracks the fact that θ_t^L is not equal to $\theta^\star$, but is not too far away. Finally, ε_Q is a quantization parameter to ensure that all the estimators are simultaneously accurate, and is taken as $\varepsilon_Q \triangleq \varepsilon_Q(t) = 1/t^2$. Through our careful theoretical analysis we can avoid searching through the entire confidence ellipsoid $\mathcal{C}_t$ in (6), which is typically accomplished via expensive convex optimization subroutines, and can instead employ a computationally efficient threshold surrogate τ_t. This threshold τ_t provably tests whenever the optimal baseline threshold policy tests, and makes a vanishing number of excess tests.

5 THEORETICAL ANALYSIS

We begin by showing that SCOUT can accurately estimate p_{err}. The learner does not start with knowledge of P or $\theta^\star$, and by extension $\tau^\star$ but we show that as SCOUT improves its estimation of each of these, its estimate of p_{err} improves. We analyze this with a sequence of lemmas.

First, we show that, with high probability, our estimates $p_{\text{err}}(\theta, \hat{P}_t, \hat{\tau}_t)$ are close to the true error probability $p_{\text{err}}(\theta, P, \tau)$ (Lemma 7). To control this across *all* $\theta \in \mathcal{B}(0,1)$ and $\tau \in [0,1]$, we quantize the set of possible θ and τ (denoted $\mathcal{Q}_\theta$, and $\mathcal{Q}_\tau$ respectively), and use a union bound to ensure that our error estimates hold simultaneously for all quantized values. We define this good event as $G_{p_{\text{err}}}$ (Equation (15)), and show that it holds with probability at least $1 - \delta'$ in Lemma 8. Additionally, we define our quantized estimator of τ as $\tau_Q^\star$, which is close to $\tau^\star$:

$$\tau_Q^\star(\theta, \hat{P}, \alpha) \triangleq \min\{\tau_Q \in \mathcal{Q}_\tau : p_{\text{err}}(\theta, \hat{P}, \tau_Q) \leq \alpha\},$$
$$0 \leq \tau_Q^\star(\theta, \hat{P}, \alpha) - \tau^\star(\theta, \hat{P}, \alpha) \leq \varepsilon_Q. \tag{9}$$

Having established the stability of the optimal threshold to changes in P (Lemma 7), we now show that it is also stable under changes in the parameter θ. To state our results, for any $\theta_Q \in \mathcal{Q}_\theta \cap \mathcal{C}_t$ we define an estimator $\hat{\tau}$, which is lower bounded by $\tau^\star$ on $G_{p_{\text{err}}}$ and G_θ:

$$\hat{\tau}(\theta_Q, \hat{P}_t, \alpha) \triangleq \tau_Q^\star\left(\theta_Q, \hat{P}_t, \alpha - \zeta_t - 2B_t/\sqrt{\lambda_{\min}^t}\right) + 2B_t/\sqrt{\lambda_{\min}^t}, \tag{10}$$

$$\hat{\tau}(\theta_Q, \hat{P}_t, \alpha) \geq \tau^\star(\theta^\star, P, \alpha) \;\forall\; \theta_Q \in \mathcal{Q}_\theta \cap \mathcal{C}_t. \tag{11}$$

In other words, the empirical $\hat{\tau}$ estimator evaluated at the approximate values θ_Q and $\hat{P}_t$ provides us with an upper bound for the true threshold $\tau^\star$ evaluated at $\theta^\star$ and P. This enables our design of τ_t used in the algorithm. The last property we will need for our analysis is that $\tau^\star$ does not vary too quickly with respect to α. We show that for small γ, $\tau^\star(\theta^\star, P, \alpha - \gamma)$ is not much larger than $\tau^\star$ (Lemma 11). This necessitates bounding the context probability mass in a subsection of an annulus, and may be of independent interest. For more details on our stability analysis, we refer the reader to Appendix B.

5.1 Defining a good event

As is common practice in Multi-Armed Bandit analyses, we define a "good event" under which all concentration arguments hold, and condition on this event for the remainder of our analysis. Recall that $N_\theta^t = |\mathcal{S}_\Theta^t|$ denotes the number of samples (X_s, Y_s) collected to estimate $\theta^\star$ up to round t, and similarly $N_P^t = |\mathcal{S}_P^t|$ is the number of samples collected used in the context distribution estimation.

Definition 2. *The good event $G = G_\theta \cap G_{p_{err}} \cap G_N \cap G_\lambda$ is comprised of the following:*

1. *G_θ: The confidence sets $\mathcal{C}_t$ (defined in Lemma 2) are valid, in that $\theta^\star \in \mathcal{C}_t$ for all t.*
2. *$G_{p_{err}}$: The estimates of p_{err} on $\mathcal{Q}_\theta \times \mathcal{Q}_\tau$ are ζ_t accurate for all t (Lemma 7).*
3. *G_N: the confidence sets get enough samples. $G_N = \bigcap_{t=1}^T G_N^{(t)}$, where $G_N^{(t)}$ is the event that $N_\theta^{(t)} \geq p^\star t/2 - \sqrt{\frac{\ln(\pi t^2/(3\delta'))}{2}}$.*
4. *G_λ: The minimum eigenvalue of the empirical covariance matrix grows linearly in t. Concretely, $G_\lambda = \cap_{t=T_0}^T G_\lambda^{(t)}$, where $G_\lambda^{(t)}$ is the event that $\lambda_{\min}^t \geq p^\star t \lambda_0/12$.*

Detailed proofs are deferred to Appendix E. The first event G_θ satisfies $\mathbb{P}(G_\theta) \geq 1 - \delta'$ by Lemma 2. The second event $G_{p_{\text{err}}}$ satisfies $\mathbb{P}(G_{p_{\text{err}}}) \geq 1 - \delta'$ by Lemma 8. To prove that G_N holds with high probability, we utilize the fact that on G_θ and $G_{p_{\text{err}}}$, when the optimal policy tests then our policy tests as well, as proved in Lemma 19. Combining this fact with Hoeffding's inequality yields the desired result in Lemma 23. When G_N holds, we have $N_\theta^t \geq p^\star t/3$ for all $t \geq T_0$ where T_0 is a large constant. For the last event, $\mathbb{P}(G_\lambda) \geq 1 - 2\delta'$, which we show via a covering argument used to bound the minimum eigenvalue of the empirical covariance matrix V_t (Lemma 24), and G_N to lower bound the number of samples used. Thus,

Lemma 3. *The good event G holds with high probability: $\mathbb{P}(G) \geq 1 - 6\delta'$.*

5.2 Safety Analysis

Our testing rule is designed to be computationally efficient and pessimistic. Here, pessimism means that whenever the baseline policy tests, our policy does the same. To prove the (α, δ)-safety of `SCOUT` , we utilize two helper lemmas. In Lemma 19, we prove that when the baseline policy tests for $\tau^\star(\theta^\star, P, \alpha_t)$, our policy tests as well. In Lemma 20, we prove that when the baseline policy predicts, our policy outputs the same prediction. Combining these yields the desired result.

Lemma 4. *Under G, `SCOUT` is (α, δ')-safe.*

5.3 Regret Analysis

To derive a regret bound, we begin by proving a bound on the instantaneous regret during rounds $t > T_0$ (Lemma 22, proof in Appendix D). Summing this lemma over t yields the following Theorem, where we set $\delta' = \delta/7$.

Theorem 1. *`SCOUT` satisfies (α, δ)-safety and has regret bounded by*

$$T_0 + \tilde{C}\frac{M}{m}\sqrt{\frac{dT\log(T/\delta)}{p^\star \lambda_0}},$$

for an absolute constant $\tilde{C} > 0$, which is made explicit in the proof (Appendix D).

Note that the probability parameter δ can scale exponentially in T without changing the regret. While at first our algorithm may appear to beat the linear dimension dependence expected in linear bandits, this missing factor is hidden in λ_0. In Appendix D, we can apply a lower bound for λ_0 (see Lemma 1) and recover the $\tilde{\mathcal{O}}(d\sqrt{T})$ regret bound. For a detailed synopsis of our work and potential future extensions, see Section 7.

6 NUMERICAL RESULTS

We corroborate our theoretical guarantees with numerical simulations, showing that `SCOUT` is able to efficiently compute the testing rule and converge to the optimal error rate. The code for reproducing our experiments is available online: `https://github.com/TavorB/SCOUT`. We generate simulations varying the dimensionality and the target error rate α, highlighting the rapid convergence of our method when $p^\star$ is large. We discuss several algorithmic modifications in Appendix F, including batched parameter updates and omission of the projection step, which allow the

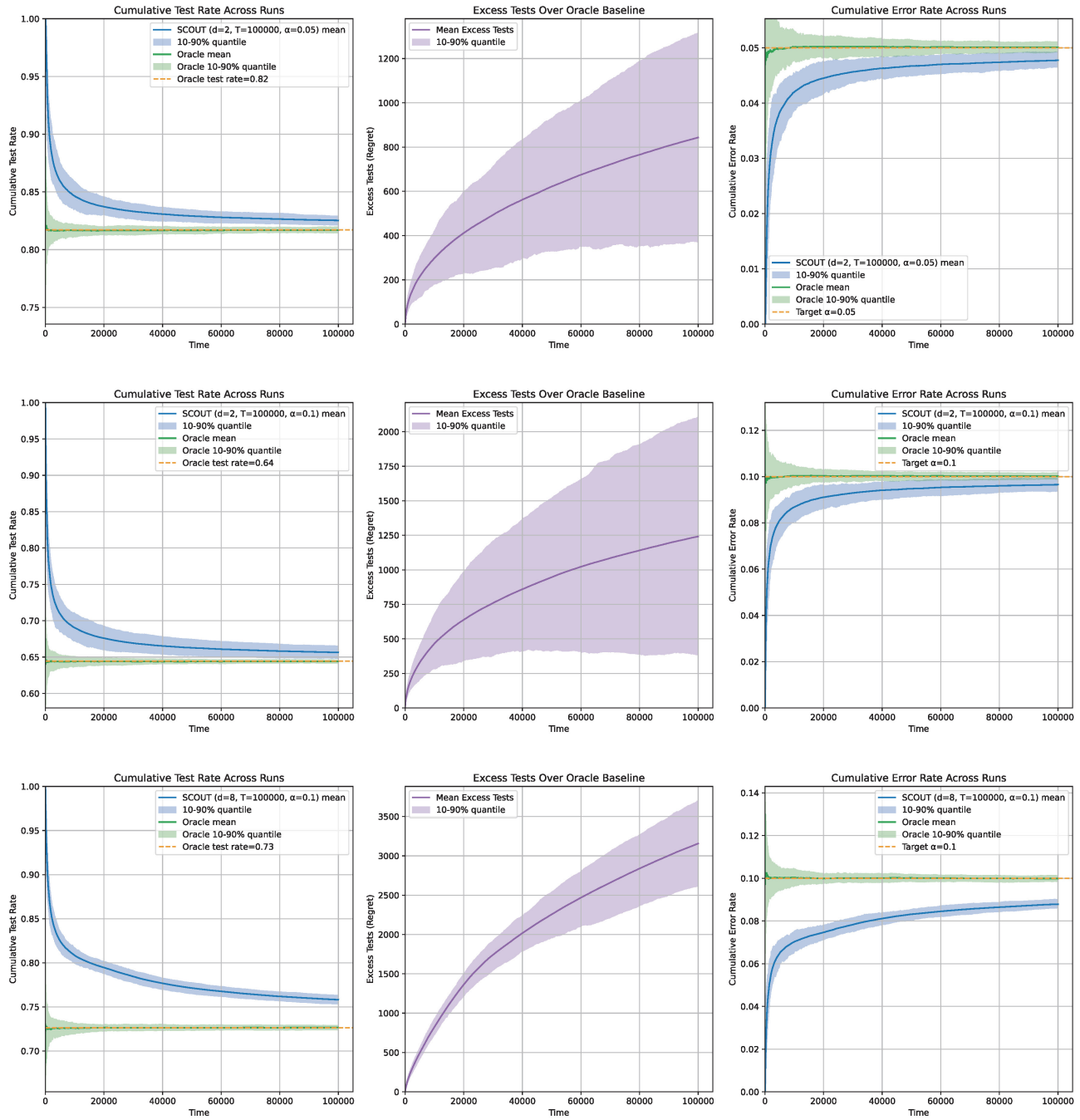


Figure 3: Simulation results. Rows 1 and 2 correspond to $d = 2$, where the first shows $\alpha = 0.05$, and the second $\alpha = 0.1$. Row 3 shows $d = 8, \alpha = 0.1$. x-axis corresponds to time t. Left plots show the cumulative test rate (10-90% quantiles shaded), where blue shows the performance of `SCOUT` , orange shows the oracle test rate at $p^\star$, and green shows the empirical test rate under $\tau^\star$. The middle plots show the excess number of tests, demonstrating the sublinear regret of `SCOUT` . The right plots show the misclassification rate of `SCOUT` . While the optimal baseline policy fluctuates around the desired threshold α, often exceeding it, `SCOUT` starts far below (very safe) then gradually learns to be more aggressive, approaching misclassification rate α but never exceeding it.

algorithm to run efficiently while retaining the core principles of `SCOUT` . The empirical results, which demonstrate sublinear regret and adherence to the safety constraint across all instances, validate that these practical simplifications do not compromise the algorithm's performance in our simulated environments.

7 DISCUSSION

In this work we introduced SCOUT , the first algorithm that provably balances **no-regret learning** with a **high-probability safety guarantee** on the empirical misclassification rate in logistic bandits. Our analysis shows that a simple, efficiently-computable testing rule suffices to achieve the order optimal $\widetilde{O}\big(\sqrt{dT/\lambda_0}\big)$ excess-test rate. Empirical results confirm that these bounds translate to practice on moderately large horizons.

In medical triage—our motivating use-case—SCOUT can be viewed as a "test-or-treat" policy that automatically calibrates how aggressively to screen as new evidence accrues. Because the policy is pessimistic by design, it never tests less than an oracle baseline that knows both the patient distribution and the ground-truth regression coefficients. This property is attractive in any high-stakes domain where misclassifications are costly (e.g. credit risk, fraud detection, or industrial quality control).

There are many interesting directions of future work. One simple extension is handling unequal Type-I and Type-II error control. The threshold-selection step can be split to handle false positives and false negatives separately by using two one-sided versions of p_{err}. Additionally, we can use improved confidence bounds from Lee et al. (2025) in Lemma 2 to remove the κ factor in B_t and generalize to larger context and Θ sets. Less straightforwardly, we have the setting where the optimal baseline does not need to test, i.e., $p^\star = 0$. If the optimal policy never tests, can one detect *fast enough* that screening is unnecessary while still retaining the high-probability safety constraint? Going beyond stochastic contexts, we plan to explore whether the ideas behind SCOUT can be combined with online calibration tools to handle non-stationary or even adversarial X_t.

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011.

M Mehdi Afsar, Trafford Crump, and Behrouz Far. Reinforcement learning based recommender systems: A survey. *ACM Computing Surveys*, 55(7):1–38, 2022.

Ella Bar-El Avidan and Ilai Bistritz. Juggler: Multi-task learning with task performance constraints. In *2025 IEEE 64th Conference on Decision and Control (CDC)*, pages 8009–8014. IEEE, 2025.

Hamsa Bastani, Kimon Drakopoulos, Vishal Gupta, Jon Vlachogiannis, Christos Hadjichristodoulou, Pagona Lagiou, Gkikas Magiorkinis, Dimitrios Paraskevis, and Sotirios Tsiodras. Interpretable operations research for high-stakes decisions: Designing the greek covid-19 testing system. *INFORMS Journal on Applied Analytics*, 52(5):398–411, 2022.

Romain Camilleri, Zhihan Xiong, Maryam Fazel, Lalit Jain, and Kevin G Jamieson. Selective sampling for online best-arm identification. *Advances in Neural Information Processing Systems*, 34:11071–11082, 2021.

Romain Camilleri, Andrew Wagenmaker, Jamie H Morgenstern, Lalit Jain, and Kevin G Jamieson. Active learning with safety constraints. *Advances in Neural Information Processing Systems*, 35:33201–33214, 2022.

Emmanuel J Candès, Andrew Ilyas, and Tijana Zrnic. Probably approximately correct labels. *arXiv preprint arXiv:2506.10908*, 2025.

Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

Nicolo Cesa-Bianchi, Claudio Gentile, Luca Zaniboni, and Manfred Warmuth. Worst-case analysis of selective sampling for linear classification. *Journal of Machine Learning Research*, 7(7), 2006.

Fan Chung and Linyuan Lu. Concentration inequalities and martingale inequalities: a survey. *Internet mathematics*, 3(1):79–127, 2006.

Corinna Cortes, Giulia DeSalvo, and Mehryar Mohri. Learning with rejection. In *Algorithmic Learning Theory: 27th International Conference, ALT 2016, Bari, Italy, October 19-21, 2016, Proceedings 27*, pages 67–82. Springer, 2016.

Corinna Cortes, Giulia DeSalvo, Claudio Gentile, Mehryar Mohri, and Scott Yang. Online learning with abstention. In *international conference on machine learning*, pages 1059–1067. PMLR, 2018.

Suresh Dara, Swetha Dhamercherla, Surender Singh Jadav, CH Madhu Babu, and Mohamed Jawed Ahsan. Machine learning in drug discovery: a review. *Artificial intelligence review*, 55(3):1947–1999, 2022.

Sanjoy Dasgupta, Adam Tauman Kalai, and Claire Monteleoni. Analysis of perceptron-based active learning. In *International conference on computational learning theory*, pages 249–263. Springer, 2005.

Ofer Dekel, Claudio Gentile, and Karthik Sridharan. Selective sampling and active learning from single and multiple teachers. *The Journal of Machine Learning Research*, 13(1):2655–2697, 2012.

Ilias Diakonikolas, Daniel M Kane, Vasilis Kontonis, Christos Tzamos, and Nikos Zarifis. Efficiently learning halfspaces with tsybakov noise. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 88–101, 2021.

Ilias Diakonikolas, Vasilis Kontonis, Christos Tzamos, and Nikos Zarifis. Online learning of halfspaces with massart noise. *arXiv preprint arXiv:2405.12958*, 2024.

Yue Duan, Zhen Zhao, Lei Qi, Luping Zhou, Lei Wang, and Yinghuan Shi. Towards semi-supervised learning with non-random missing labels. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16121–16131, 2023.

Vera Egorova, Amparo Gil, Javier Segura, NM Temme, et al. Computation of the regularized incomplete beta function. 2023.

Louis Faury, Marc Abeille, Clément Calauzènes, and Olivier Fercoq. Improved optimistic algorithms for logistic bandits. In *International Conference on Machine Learning*, pages 3052–3060. PMLR, 2020.

Gerald B Folland. *Real analysis: modern techniques and their applications*. John Wiley & Sons, 1999.

Yoav Freund, H Sebastian Seung, Eli Shamir, and Naftali Tishby. Selective sampling using the query by committee algorithm. *Machine learning*, 28:133–168, 1997.

Aditya Gangrade, Anil Kag, Ashok Cutkosky, and Venkatesh Saligrama. Online selective classification with limited feedback. *Advances in Neural Information Processing Systems*, 34:14529–14541, 2021a.

Aditya Gangrade, Anil Kag, and Venkatesh Saligrama. Selective classification via one-sided prediction. In *International Conference on Artificial Intelligence and Statistics*, pages 2179–2187. PMLR, 2021b.

Aditya Gangrade, Tianrui Chen, and Venkatesh Saligrama. Safe linear bandits over unknown polytopes. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 1755–1795. PMLR, 2024.

Paolo Giudici. Safe machine learning. *Statistics*, 58 (3):473–477, 2024.

Surbhi Goel, Steve Hanneke, Shay Moran, and Abhishek Shetty. Adversarial resilience in sequential prediction via abstention. *Advances in Neural Information Processing Systems*, 36:8027–8047, 2023.

Shangding Gu, Long Yang, Yali Du, Guang Chen, Florian Walter, Jun Wang, and Alois Knoll. A review of safe reinforcement learning: Methods, theory and applications. *arXiv preprint arXiv:2205.10330*, 2022.

Steve Hanneke and Liu Yang. Toward a general theory of online selective sampling: Trading off mistakes and queries. In *International Conference on Artificial Intelligence and Statistics*, pages 3997–4005. PMLR, 2021.

Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

Michael Jorgensen. Volumes of n-dimensional spheres and ellipsoids, 2014.

Abbas Kazerouni, Mohammad Ghavamzadeh, Yasin Abbasi Yadkori, and Benjamin Van Roy. Conservative contextual linear bandits. *Advances in Neural Information Processing Systems*, 30, 2017.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Junghyun Lee, Se-Young Yun, and Kwang-Sung Jun. A unified confidence sequence for generalized linear models, with applications to bandits. *Advances in Neural Information Processing Systems*, 37:124640–124685, 2025.

Shengqiao Li. Concise formulas for the area and volume of a hyperspherical cap. *Asian Journal of Mathematics & Statistics*, 4(1):66–70, 2010.

Francesco Orabona, Nicolo Cesa-Bianchi, et al. Better algorithms for selective sampling. In *Proceedings of the 28th international conference on machine learning: Bellevue, Washington, USA*, pages 433–440. Omnipress, 2011.

Aldo Pacchiano, Mohammad Ghavamzadeh, Peter Bartlett, and Heinrich Jiang. Stochastic bandits with linear constraints. In *International conference on artificial intelligence and statistics*, pages 2827–2835. PMLR, 2021.

Michael Pinelis and David Ruppert. Machine learning portfolio allocation. *The Journal of Finance and Data Science*, 8:35–54, 2022.

Ayush Sekhari, Karthik Sridharan, Wen Sun, and Runzhe Wu. Selective sampling and imitation learning via online regression. *Advances in Neural Information Processing Systems*, 36:67213–67268, 2023.

Burr Settles. Active learning literature survey. 2009.

H Sebastian Seung, Manfred Opper, and Haim Sompolinsky. Query by committee. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 287–294, 1992.

Aleksandrs Slivkins. Dynamic ad allocation: Bandits with budgets. *arXiv preprint arXiv:1306.0155*, 2013.

Richard S Sutton, Andrew G Barto, et al. Reinforcement learning. *Journal of Cognitive Neuroscience*, 11 (1):126–134, 1999.

Alexander B Tsybakov. Optimal aggregation of classifiers in statistical learning. *The Annals of Statistics*, 32(1):135–166, 2004.

Jessica Vamathevan, Dominic Clark, Paul Czodrowski, Ian Dunham, Edgardo Ferran, George Lee, Bin Li, Anant Madabhushi, Parantu Shah, Michaela Spitzer, et al. Applications of machine learning in drug discovery and development. *Nature reviews Drug discovery*, 18(6):463–477, 2019.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press, 2019.

Jiayu Yao, Emma Brunskill, Weiwei Pan, Susan Murphy, and Finale Doshi-Velez. Power constrained bandits. In *Machine Learning for Healthcare Conference*, pages 209–259. PMLR, 2021.

Zheqing Zhu and Benjamin Van Roy. Scalable neural contextual bandit for recommender systems. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 3636–3646, 2023.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes]

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   (b) Complete proofs of all theoretical results. [Yes]
   (c) Clear explanations of any assumptions. [Yes]

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Yes]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. [Not Applicable]
   (b) The license information of the assets, if applicable. [Not Applicable]
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

Appendix

A Baseline policy

Here we provide some discussion and proofs regarding the optimal baseline we compare to.

A.1 Proof of Proposition 1

Proof. When the value of the parameter $\theta^\star$ and the collection of the contexts $\{X_t\}_{t=1}^T$ are known, we can equivalently write the problem as follows. Let $p_t = \mu(X_t^\top \theta^\star)$, the labels $Y_t \sim Ber(p_t)$ independently across t.

To compute the expected error, that is $\mathbb{E}(E_t) \triangleq \mathbb{E}(\mathbb{1}\{\hat{Y}_t \neq Y_t\})$, we only need to examine the case where we do not test. When we do test, we observe the true label and incur zero error. For $Z_t = 0$ then, the expected error is

1. If $\hat{Y}_t = 1$ then $\mathbb{E}(\mathbb{1}\{\hat{Y}_t \neq Y_t\} \mid \hat{Y}_t = 1) = 1 - p_t$.
2. Else if $\hat{Y}_t = 0$ then $\mathbb{E}(\mathbb{1}\{\hat{Y}_t \neq Y_t\} \mid \hat{Y}_t = 0) = p_t$.

The optimal policy then is to output the prediction with the smallest error. The expected error then is equal to

$$\mathbb{E}(\mathbb{1}\{\hat{Y}_t \neq Y_t\}) \triangleq \min\{1 - p_t, p_t\}.$$

We denote $\mathbb{P}(Z_t = 0) = \eta_t$. The optimal policy choice is reduced to the following optimization problem.

$$\min_{\{\eta_t\}} \sum_{t=1}^T 1 - \eta_t \quad \text{s.t.} \quad \frac{1}{T}\sum_{t=1}^T \min\{1 - p_t, p_t\}\eta_t \leq \alpha, \quad 0 \leq \eta_t \leq 1. \tag{12}$$

Or equivalently can be written as:

$$\max_{\{\eta_t\}} \sum_{t=1}^T \eta_t \quad \text{s.t.} \quad \frac{1}{T}\sum_{t=1}^T \min\{1 - p_t, p_t\}\eta_t \leq \alpha, \quad 0 \leq \eta_t \leq 1. \tag{13}$$

The solution of this Linear Program is the solution of the *Fractional Knapsack* problem with budget α. This problem can be optimally solved with a greedy strategy, sorting the coefficients $\min\{1 - p_t, p_t\}$ in non-increasing order and assigning $\eta = 1$ to the lowest "error" contexts until we do not violate the budget constraint α. This strategy is clearly a threshold strategy that depends on a.

□

A.1.1 Conversion to (α, δ) safety

It is worth mentioning that solving the problem by satisfying the constraint in expectation does not provide any guarantees when we require the constraint to hold with high probability. Even if we apply the Markov's inequality to convert the constraint in expectation to a high probability one, we derive a very loose bound (need to target error rate α^2 to obtain a high probability bound of α).

$$\mathbb{P}\left(\frac{1}{T}\sum_{t=1}^T \mathbb{1}\{\hat{Y}_t \neq Y_t\} \geq \alpha\right) \leq \frac{\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^T \mathbb{1}\{\hat{Y}_t \neq Y_t\}\right]}{\alpha} \leq 1.$$

However, we show that we are still competitive with respect to this fixed baseline policy.

A.2 Proof of Lemma 1

We outline the proof as follows; as $\|\theta^\star\| = 1$, a ball of radius $\tau^\star$ is a subset of the contexts tested by the baseline policy. The contexts drawn from this ball form a positive definite covariance matrix, which implies that the minimum eigenvalue of the overall covariance matrix is positive.

Lemma 1. *There is a constant $\lambda_0^{\min}(\tau^\star, d) > 0$:*

$$\lambda_{\min}\left(\mathbb{E}_P\left[XX^\top \;\middle|\; |\langle X, \theta^\star\rangle| \le \tau^\star\right]\right) = \lambda_0 \ge \lambda_0^{\min}(\tau^\star, d) > 0.$$

Proof. By Cauchy-Schwarz, $|X^\top\theta^\star| \le \|X\|$, as $\|\theta^\star\| = 1$. As a result all contexts $X \in \mathcal{B}(0, \tau^\star)$ satisfy $|X^\top\theta^\star| \le \tau^\star$ and thus are tested by the baseline policy. We can split the set of contexts to be tested by the baseline policy, $\mathcal{T} = \{X \in \mathcal{B}(0,1) : |X^\top\theta^\star| \le \tau^\star\}$ into $\mathcal{B}(0,\tau^\star) \cup (\mathcal{T} \setminus \mathcal{B}(0,\tau^\star))$.

We begin by showing that the covariance matrix of the contexts tested by the baseline policy under a *uniform* context distribution has a positive minimum eigenvalue. Then, leveraging the assumption that P is lower bounded (Assumption 2), we prove our desired claim.

Let $V_d(1)$ be the volume of the *d-dimensional* unit ball. We begin by showing that the minimum eigenvalue of the uniform distribution on the *d-dimensional* unit ball is positive using standard arguments as in Vershynin (2018) (Version 2, Section 3.3.3).

Lemma 5. *The minimum eigenvalue of X drawn uniformly from the d-dimensional ball satisfies:*

$$\lambda_{\min}\left(\mathbb{E}_{\mathbf{x}\sim\mathrm{Unif}(\mathcal{B}(0,1))}\left[\mathbf{x}\mathbf{x}^\top\right]\right) = \frac{1}{d+2}.$$

Proof. The quantity $\mathbb{E}[\mathbf{x}\mathbf{x}^\top]$ is the covariance matrix of the uniform over the unit *d-dimensional* ball. For $\mathbf{x} \sim \mathrm{Unif}(\mathcal{B}(0,1))$, $\mathbb{E}[\mathbf{x}\mathbf{x}^\top]$ can be written as $a\mathbf{I}_d$ due to spherical symmetry.

By a change of variables, we can obtain that $\mathbb{E}[\mathbf{x}_i\mathbf{x}_j] = -\mathbb{E}[\mathbf{x}_i\mathbf{x}_j]$ for $i \ne j$, implying that $\mathbb{E}[\mathbf{x}_i\mathbf{x}_j] = 0$. To compute the diagonal entries:

$$\begin{aligned}
\mathbb{E}[x_i^2] &= \frac{1}{d}\mathbb{E}[\mathbf{x}^2] \\
&= \frac{1}{d}\int_{\|\mathbf{x}\|_2^2 \le 1} \frac{\mathbf{x}^2}{V_d(1)} d\mathbf{x} \\
&= \frac{1}{dV_d(1)}\int_{\mathcal{S}^{d-1}}\int_{0\le r\le 1} r^2 r^{d-1} dr d\sigma(\omega) \\
&= \frac{S_d(1)}{V_d(1)}\frac{1}{d(d+2)} \\
&= \frac{1}{d+2}
\end{aligned}$$

where $S_d(1)$ is the surface of the unit sphere and $d\sigma$ any surface measure. In the last line, we leverage the volume to surface area ratio of $\mathcal{B}(0,1)$:

$$\frac{V_d(1)}{S_d(1)} = \frac{\frac{\pi^{d/2}}{\Gamma(d/2+1)}}{\frac{d\pi^{d/2}}{\Gamma(d/2+1)}} = \frac{1}{d}.$$

Thus, all eigenvalues of this covariance matrix are equal to $1/(d+2)$. □

Now, as our density is smooth, we can use that for all $\mathbf{x}, \upsilon \in \mathcal{B}(0,1)$ it holds $(\mathbf{x}^\top\upsilon)^2 p(\mathbf{x}) \ge (\mathbf{x}^\top\upsilon)^2 m$ and so:

$$
\begin{aligned}
\lambda_0 &= \lambda_{\min}\left(\mathbb{E}_P\left[XX^\top \;\middle|\; |\langle X, \theta^\star\rangle| \le \tau^\star\right]\right) \\
&= \min_{\|v\|=1} v^\top \mathbb{E}_P\left[XX^\top \;\middle|\; |\langle X, \theta^\star\rangle| \le \tau^\star\right] v \\
&= \frac{1}{p^\star} \min_{\|v\|=1} \int_{|X^\top\theta^\star|\le\tau^\star} (\mathbf{x}^\top v)^2 p(\mathbf{x}) d\mathbf{x} \\
&\overset{(a)}{\ge} \frac{1}{p^\star} \min_{\|v\|=1} \int_{\mathcal{B}(0,\tau^\star)} (\mathbf{x}^\top v)^2 m d\mathbf{x} \\
&\overset{(b)}{=} \frac{m(\tau^\star)^{d+2} V_d(1)}{p^\star} \min_{\|v\|=1} \int_{\mathcal{B}(0,1)} (\mathbf{u}^\top v)^2 \frac{1}{V_d(1)} d\mathbf{u} \\
&\overset{(c)}{=} \frac{m(\tau^\star)^{d+2} V_d(1)}{p^\star} \frac{S_d(1)}{V_d(1)} \frac{1}{d(d+2)} \\
&\overset{(d)}{=} \frac{m(\tau^\star)^{d+2} V_d(1)}{p^\star(d+2)} \triangleq \lambda_0^{\min}(\tau^\star, d).
\end{aligned}
$$

(a) utilizes the fact that $p(x) \ge m$ from Assumption 2, $\mathcal{B}(0,\tau^\star) \subseteq \left\{| X^\top\theta^\star | \le \tau^\star\right\}$, and $(\mathbf{x}^\top v)^2 \ge 0$ for all $\mathbf{x}, v$. (b) comes from a change of variables, with $\mathbf{x} \mapsto \tau^\star \mathbf{u}$, with $d\mathbf{x} = \tau^\star d\mathbf{u}$. (c) utilizes Lemma 5, and (d) simplifies the volume to surface area ratio.

□

B Stability of error estimates

To analyze `SCOUT` , we first study the stability of p_{err}. Since the learner does not start with knowledge of P or $\theta^\star$, and by extension $\tau^\star$ we must show that, as time progresses `SCOUT` 's estimates of the error probabilities are not too far off.

Before analyzing the stability of the $\tau(\cdot)$ function, we present an auxiliary lemma that will be employed throughout the subsequent analysis.

Lemma 6. *For any $x > 0$ and any θ, P, it holds that*

$$\min\{\tau \in [0,1] : p_{err}(\theta, P, \tau - x) \le \alpha\} \le \min\{\tau \in [0,1] : p_{err}(\theta, P, \tau) \le \alpha\} + x.$$

Proof. Let

$$g(\tau) \triangleq p_{\text{err}}(\theta, P, \tau).$$

It holds that g is non-increasing on $\mathbb{R}$, $(p_{\text{err}}(\theta, P, \tau) = 1/2, \text{ for } \tau < 0, p_{\text{err}}(\theta, P, \tau) = 0, \text{ for } \tau > 1)$. Define

$$\tilde{\tau} \triangleq \min\{\tau \in [0,1] : g(\tau) \le \alpha\}.$$

We want to prove

$$\min\{\tau \in [0,1] : g(\tau - x) \le \alpha\} \;\le\; \tilde{\tau} + x.$$

Let $s \triangleq \tilde{\tau} + x$. We consider the following two cases.

First case; $s \le 1$. Then $s - x = \tilde{\tau}$. By definition of $\tilde{\tau}$ we have $g(\tilde{\tau}) \le \alpha$. Since g is non-increasing, it follows that

$$g(s - x) = g(\tilde{\tau}) \le \alpha,$$

so s belongs to the set $\{\tau \in [0,1] : g(\tau - x) \le \alpha\}$. Hence

$$\min\{\tau \in [0,1] : g(\tau - x) \le \alpha\} \le s = \tilde{\tau} + x.$$

Second case; $s > 1$. In this case,

$$\min\{\tau \in [0,1] : g(\tau - x) \le \alpha\} \le 1 < s = \tilde{\tau} + x.$$

In either case, we conclude that

$$\min\{\tau \in [0,1] : p_{\text{err}}(\theta, P, \tau - x) \leq \alpha\} \;\leq\; \min\{\tau \in [0,1] : p_{\text{err}}(\theta, P, \tau) \leq \alpha\} + x.$$

□

B.1 Smoothness of $\tau^\star$ with respect to $\hat{P}_t$

Since P is unknown, SCOUT estimates it via its empirical counterpart $\hat{P}_t$. In the following Lemma we bound the error between $p_{\text{err}}(\theta, \hat{P}_t, \tau)$ and $p_{\text{err}}(\theta, P, \tau)$.

Lemma 7. *Let $\hat{P}_t$ be the empirical distribution of constructed from $\lceil t/2 \rceil$ i.i.d. samples from P. Then, for any fixed θ and τ, with probability at least $1 - \delta'$ over the randomness in $\hat{P}_t$:*

$$\left| p_{err}(\theta, \hat{P}_t, \tau) - p_{err}(\theta, P, \tau) \right| \leq \sqrt{\frac{\log\left(\frac{\pi^2 t^2}{3\delta'}\right)}{4t}}$$

The proof of this result uses standard concentration bounds (Hoeffding's inequality Wainwright (2019)) using the fact that for any fixed θ and τ, (2) is the expectation of a $[0, 1/2]$ bounded random variable.

Proof of Lemma 7. First, we collect a context as a sample at every odd round, so at round t it holds that $|\mathcal{S}_P^t| = \lceil t/2 \rceil \geq t/2$. Indexing these samples as x_i, we can write the empirical error $p_{\text{err}}(\theta, \hat{P}_t, \tau)$ as follows:

$$\begin{aligned} p_{\text{err}}(\theta, \hat{P}_T, \tau) - p_{\text{err}}(\theta, P, \tau) &= \int (1 + \exp(|x^\top \theta|))^{-1} \mathbb{1}\left\{|x^\top \theta| > \tau\right\} \hat{P}_t(dx) - p_{\text{err}}(\theta, P, \tau) \\ &= \frac{1}{\lceil t/2 \rceil} \sum_{i=1}^{\lceil t/2 \rceil} \left(\xi_i - p_{\text{err}}(\theta, P, \tau)\right), \end{aligned} \tag{14}$$

where we define ξ_i as the i-th term in this sum:

$$\xi_i = (1 + \exp(|x_i^\top \theta|))^{-1} \mathbb{1}\left\{|x_i^\top \theta| > \tau\right\}.$$

As $0 \leq (1 + \exp(|x_i^\top \theta|))^{-1} \leq \frac{1}{2}$, the summands ξ_i are i.i.d. [0,1/2] random variables with mean $p_{\text{err}}(\theta, P, \tau)$, so we can apply Hoeffding's inequality Wainwright (2019):

$$\mathbb{P}\left(\left| \frac{1}{\lceil t/2 \rceil} \sum_{i=1}^{\lceil t/2 \rceil} \left(\xi_i - p_{\text{err}}(\theta, P, \tau)\right) \right| \geq \sqrt{\frac{\log(2/\delta'')}{4t}} \right) \leq \delta''.$$

By taking the union bound over all rounds $t \geq 1$ and setting $\delta'' \triangleq \frac{6\delta'}{\pi^2 t^2}$ we derive:

$$\mathbb{P}\left(\left| \frac{1}{\lceil t/2 \rceil} \sum_{i=1}^{\lceil t/2 \rceil} \left(\xi_i - p_{\text{err}}(\theta, P, \tau)\right) \right| \leq \sqrt{\frac{\log\left(\frac{\pi^2 t^2}{3\delta'}\right)}{4t}}, \forall t : t \geq 1 \right) \geq 1 - \delta'.$$

Here, we apply the well-known result for the Basel series: $\sum_{t=1}^{\infty} \frac{1}{t^2} = \frac{\pi^2}{6}$.

□

Since we require this bound to hold over all $\theta \in \Theta$ and $\tau \in [0, 1]$ and these sets are uncountable, we utilize an ϵ-net analysis for both $\tau \in [0, 1]$ and $\theta \in \Theta$. We detail this quantization analysis strategy in the following section.

B.1.1 Quantization to enable union bounding

We define quantized versions of τ and θ, to bound the failure probability of our estimators over a countable quantized set. We take progressively finer and finer quantizations, with our quantization accuracy scaling as $\varepsilon_Q = t^{-2}$ (t suppressed from notation). We consider an ε_Q covering of the unit interval for τ as $\mathcal{Q}_\tau \triangleq \mathcal{N}([0, 1], \varepsilon_Q)$, denoting

the quantized τ value as $\tau_Q \in \mathcal{Q}_\tau$ and an ε_Q cover of the d-dimensional unit sphere for θ as $\mathcal{Q}_\theta \triangleq \mathcal{N}(\mathcal{S}^{d-1}, \varepsilon_Q)$, denoting the quantized θ value as $\theta_Q \in \mathcal{Q}_\theta$. We can bound the size of these covering sets as $|\mathcal{Q}_\tau| \leq \varepsilon_Q^{-1}$ and $|\mathcal{Q}_\theta| \leq (3/\varepsilon)^d$ Vershynin (2018).

We are now able to define the "good" event $G_{p_{\text{err}}}$ where our error probability estimates are uniformly bounded by ζ_t on our quantized sets as:

$$G_{p_{\text{err}}} = \left\{ \left| p_{\text{err}}(\theta_Q, \hat{P}_t, \tau_Q) - p_{\text{err}}(\theta_Q, P, \tau_Q) \right| \leq \zeta_t \ : \ \forall t \in [T], \forall \theta_Q \in \mathcal{Q}_\theta, \forall \tau_Q \in \mathcal{Q}_\tau \right\}. \tag{15}$$

The following lemma shows that $G_{p_{\text{err}}}$ happens with high probability.

Lemma 8. *The good event $G_{p_{err}}$ satisfies $\mathbb{P}(G_{p_{err}}) \geq 1 - \delta'$.*

The proof of this result utilizes Lemma 7 and the union bound over the quantized sets $\mathcal{Q}_\theta$ and $\mathcal{Q}_\tau$.

Proof of Lemma 8. To extend Lemma 7 to hold simultaneously for all $\theta_Q \in \mathcal{Q}_\theta$ and $\tau_Q \in \mathcal{Q}_\tau$, we define an ε_Q-net for each, and union bound over their cartesian product. By Lemma 7 we know that for any fixed θ, τ, and $\delta'' > 0$:

$$\mathbb{P}\left(\left| p_{\text{err}}(\theta, \hat{P}_t, \tau) - p_{\text{err}}(\theta, P, \tau) \right| \leq \sqrt{\frac{\log(\frac{\pi^2 t^2}{3\delta''})}{4t}}, \forall t \geq 1 \right) \geq 1 - \delta''.$$

Let $\mathcal{Q}_\theta = \mathcal{N}(\mathcal{S}^{d-1}, \varepsilon_\theta)$ an ε_Q-cover of the unit ball $\mathcal{S}^{d-1}$. By **Corollary 4.2.13** of Vershynin (2018) we have that the covering number of $\mathcal{S}^{d-1}$ satisfies for any $\varepsilon_Q \in (0, 1]$;

$$\left(\frac{1}{\varepsilon_Q}\right)^d \leq |\mathcal{Q}_\theta| \leq \left(\frac{2}{\varepsilon_Q} + 1\right)^d < \left(\frac{3}{\varepsilon_Q}\right)^d.$$

As τ lives in $[0, 1]$, an ε-net of the unit segment in the real line is $\{\varepsilon, 2\varepsilon, \ldots, \lfloor \frac{1}{\varepsilon} \rfloor \varepsilon\}$, and so $|\mathcal{Q}_\tau| \leq \frac{1}{\varepsilon_\tau}$. By taking a union bound over all $\tau_Q \in \mathcal{Q}_\tau$ and all $\theta_Q \in \mathcal{Q}_\theta$, i.e. taking $\delta'' = \delta'/(|\mathcal{Q}_\theta| \cdot |\mathcal{Q}_\tau|)$, we have

$$\mathbb{P}(G_{p_{\text{err}}}) = \mathbb{P}\left(\left| p_{\text{err}}(\theta_Q, \hat{P}_t, \tau_Q) - p_{\text{err}}(\theta_Q, P, \tau_Q) \right| \leq \zeta_t, \forall t \geq 1, \theta_Q \in \mathcal{Q}_\theta, \tau_Q \in \mathcal{Q}_\tau \right) \geq 1 - \delta'.$$

Recall that ζ_t is defined in Equation (7) as

$$\zeta_t \triangleq \sqrt{\frac{(d+1)\log(1/\varepsilon_Q) + \log\left(\frac{\pi^2 t^2}{\delta'}\right)}{4t}}.$$

This stems from the union bound with $\delta'' = \delta'/(|\mathcal{Q}_\theta| \cdot |\mathcal{Q}_\tau|)$,

$$\begin{aligned}
\sqrt{\frac{\log(\frac{\pi^2 t^2}{3\delta''})}{4t}} &= \sqrt{\frac{\log\left(\frac{\pi^2 t^2 |\mathcal{Q}_\theta| \cdot |\mathcal{Q}_\tau|}{3\delta'}\right)}{4t}} \\
&\leq \sqrt{\frac{\log\left(\frac{\pi^2 t^2 \left(3\varepsilon_Q^{-d-1}\right)}{3\delta'}\right)}{4t}} \\
&= \sqrt{\frac{(d+1)\log(1/\varepsilon_Q) + \log\left(\frac{\pi^2 t^2}{\delta'}\right)}{4t}} \\
&= \zeta_t,
\end{aligned} \tag{16}$$

as claimed. As discussed, we utilize $\varepsilon_Q = 1/t^2$ to simplify the regret analysis in Theorem 1. □

Having established guarantees on the closeness of the p_{err} estimators to their true values over our quantized set, we turn our attention to the task of understanding - for a fixed θ - the closeness of the optimal estimated threshold $\tau_Q^\star(\theta, \hat{P}, \alpha)$ over the quantized set defined as

$$\tau_Q^\star(\theta, \hat{P}, \alpha) \triangleq \min\{\tau_Q \in \mathcal{Q}_\tau : p_{\text{err}}(\theta, \hat{P}, \tau_Q) \leq \alpha\}, \tag{17}$$

and the optimal estimated threshold $\tau^\star(\theta, \hat{P}, \alpha)$ over the entire domain of τ. The following "sandwich" relationship between $\tau^\star$ and $\tau^\star_Q$ holds:

$$\tau^\star(\theta, \hat{P}, \alpha) \overset{(i)}{\leq} \tau^\star_Q(\theta, \hat{P}, \alpha) \overset{(ii)}{\leq} \tau^\star(\theta, \hat{P}, \alpha) + \varepsilon_Q. \tag{18}$$

where (i) holds because $\tau^\star(\theta, \hat{P}, \alpha) = \min\{\tau \in [0,1] : p_{\text{err}}(\theta, \hat{P}, \tau) \leq \alpha\}$ and $\mathcal{Q}_\tau \subset [0,1]$ thus showing $\tau^\star(\theta, \hat{P}, \alpha)$ is the result of minimizing the same function p_{err} over a larger set than in the definition of $\tau^\star_Q(\theta, \hat{P}, \alpha)$. Inequality (ii) holds because by definition of the covering set $\mathcal{Q}_\tau$ the threshold in the cover closest to $\tau^\star(\theta, \hat{P}, \alpha)$ from above (say $\tilde{\tau} \in \mathcal{Q}_\tau$) must satisfy $\tau^\star(\theta, \hat{P}, \alpha) \leq \tilde{\tau} \leq \tau^\star_Q(\theta, \hat{P}, \alpha)$ and $|\tau^\star(\theta, \hat{P}, \alpha) - \tilde{\tau}| \leq \varepsilon_Q$. Since $p_{\text{err}}(\theta, \hat{P}, \tau^\star) \leq \alpha$ and p_{err} is monotonically decreasing in τ we see that $p_{\text{err}}(\theta, \hat{P}, \tilde{\tau}) \leq \alpha$ and therefore, due to the definition of $\tau^\star_Q(\theta, \hat{P}, \alpha)$ as the minimum threshold in $\mathcal{Q}_\tau$ satisfying $p_{\text{err}} \leq \alpha$, $\tilde{\tau} = \tau^\star_Q(\theta, \hat{P}, \alpha)$. Combining these observations we conclude that $|\tau^\star(\theta, \hat{P}, \alpha) - \tau^\star_Q(\theta, \hat{P}, \alpha)| \leq \varepsilon_Q$ and therefore the desired result.

B.2 Stability of $\tau^\star$ with respect to θ

Having established the stability of the optimal threshold to changes in P, we now show that it is also stable under changes in the parameter θ. To state our results, for any $\theta_Q \in \mathcal{Q}_\theta \cap \mathcal{C}_t$ we define an estimator $\hat{\tau}$ as (see Equation (10))

$$\hat{\tau}(\theta_Q, \hat{P}_t, \alpha) \triangleq \tau^\star_Q\left(\theta_Q, \hat{P}_t, \alpha - \zeta_t - 2B_t/\sqrt{\lambda^t_{\min}}\right) + 2B_t/\sqrt{\lambda^t_{\min}}. \tag{19}$$

This section's main result is that as long as $G_{p_{\text{err}}}, G_\theta$ hold then,

$$\hat{\tau}(\theta_Q, \hat{P}_t, \alpha) \geq \tau^\star(\theta^\star, P, \alpha) \text{ for all } \theta_Q \in \mathcal{Q}_\theta \cap \mathcal{C}_t. \tag{20}$$

In other words, the empirical $\hat{\tau}$ estimator evaluated at the estimated θ_Q and $\hat{P}_t$ provides us with an upper bound for the true threshold $\tau^\star$ evaluated at $\theta^\star$ and P. Eventually, for our regret bound, we require the reverse direction: that our estimated threshold $\hat{\tau}$ is not too much larger than $\tau^\star$, so that we do not perform too many excess tests. In order to show this we first establish a helper Lemma showing that our estimate $p_{\text{err}}(\theta, \hat{P}, \tau)$ is close to $p_{\text{err}}(\theta^\star, \hat{P}, \tau)$ when θ is close to $\theta^\star$, for any distribution ρ and threshold τ.

Lemma 9. *For all $\theta, \theta' \in \Theta$, $\tau \geq \frac{\|\theta - \theta'\|_{V_t}}{\sqrt{\lambda^t_{\min}}}$, and density $\rho(x)$ on $\mathcal{X}$:*

$$p_{err}(\theta, \rho, \tau) \leq \; p_{err}\left(\theta', \rho, \tau - \frac{\|\theta - \theta'\|_{V_t}}{\sqrt{\lambda^t_{\min}}}\right) + \frac{\|\theta - \theta'\|_{V_t}}{\sqrt{\lambda^t_{\min}}}.$$

To prove this we leverage algebraic properties of $p_{\text{err}}(\theta, \rho, \tau)$ and the Hölder inequality, a standard technique in Linear Bandits (see Lattimore and Szepesvári (2020), Part V).

Proof. Here, we use x as a dummy variable for integration:

$$
\begin{aligned}
p_{\text{err}}(\theta,\rho,\tau) &= \int (1+\exp(|x^\top\theta|))^{-1}\mathbb{1}\left\{|x^\top\theta|>\tau\right\}\rho(dx)\\
&= \int (1+\exp(|x^\top\theta'+x^\top(\theta-\theta')|))^{-1}\mathbb{1}\left\{|x^\top\theta'+x^\top(\theta-\theta')|>\tau\right\}\rho(dx)\\
&\le \int (1+\exp(|x^\top\theta'|-|x^\top(\theta-\theta')|))^{-1}\mathbb{1}\left\{|x^\top\theta'|>\tau-|x^\top(\theta-\theta')|\right\}\rho(dx)\\
&\le \int \left((1+\exp(|x^\top\theta'|))^{-1}+|x^\top(\theta-\theta')|\right)\mathbb{1}\left\{|x^\top\theta'|>\tau-|x^\top(\theta-\theta')|\right\}\rho(dx)\\
&\le \max_{x'\in\mathcal{X}} \int \left((1+\exp(|x^\top\theta'|))^{-1}+|x'^\top(\theta-\theta')|\right)\mathbb{1}\left\{|x^\top\theta'|>\tau-|x'^\top(\theta-\theta')|\right\}\rho(dx)\\
&= \max_{x'\in\mathcal{X}} p_{\text{err}}(\theta',\rho,\tau-|x'^\top(\theta-\theta')|)+\int |x^\top(\theta-\theta')|\mathbb{1}\left\{|x^\top\theta'|>\tau-|x'^\top(\theta-\theta')|\right\}\rho(dx)\\
&\le \max_{x'\in\mathcal{X}} p_{\text{err}}(\theta',\rho,\tau-\|\theta-\theta'\|_{V_t}\|x'\|_{V_t^{-1}})+\|\theta-\theta'\|_{V_t}\|x'\|_{V_t^{-1}}\mathbb{P}_\rho\left(|x^\top\theta'|>\tau-|x^\top(\theta-\theta')|\right)\\
&\le \max_{x'\in\mathcal{X}} p_{\text{err}}(\theta',\rho,\tau-\|\theta-\theta'\|_{V_t}\|x'\|_{V_t^{-1}})+\|\theta-\theta'\|_{V_t}\|x'\|_{V_t^{-1}}\\
&= p_{\text{err}}(\theta',\rho,\tau-\frac{\|\theta-\theta'\|_{V_t}}{\sqrt{\lambda^t_{\min}}})+\frac{\|\theta-\theta'\|_{V_t}}{\sqrt{\lambda^t_{\min}}}
\end{aligned}
$$

The first inequality follows from the triangle inequality, and the second inequality follows from the fact that $1/(1+\exp(z))$ is 1/4-Lipschitz (coarsely upper bounded as 1). The third bounds by looking at the worst case context x'. The fourth inequality utilizes Hölder's inequality, on the worst case context x', and that p_{err} is monotone in τ. The second to last inequality follows from the fact that a probability is always less than or equal to 1. Finally, we apply the following bound for any $x'\in\mathcal{X}$; $\|x'\|_{V_t^{-1}}\le\frac{1}{\sqrt{\lambda^t_{\min}}}$, where we have implicitly used that $\|x'\|\le 1,\forall x'\in\mathcal{X}$.

□

Lemma 9 indicates that as our ability to estimate θ improves, so will our error probability estimates. Now, conditioning on the good event $G_{p_{\text{err}}}$, we show that $\tau^\star_Q(\theta_Q,\hat{P}_t,\alpha)$ is close to $\tau^\star$ when θ_Q is close to $\theta^\star$.

Lemma 10. *Conditioning on $G_{p_{err}}$, for any $\theta_Q\in\mathcal{Q}_\theta\cap\mathcal{C}_t,\theta\in\mathcal{C}_t$ such that $\frac{\|\theta_Q-\theta\|_{V_t}}{\sqrt{\lambda^t_{\min}}}\le\tau^\star(\theta,P,\alpha-\zeta_t-\frac{\|\theta_Q-\theta^\star\|_{V_t}}{\sqrt{\lambda^t_{\min}}})$ it is true that:*

$$
\begin{aligned}
\tau^\star_Q(\theta_Q,\hat{P}_t,\alpha) &\le \tau^\star\left(\theta,P,\alpha-\zeta_t-\frac{\|\theta_Q-\theta\|_{V_t}}{\sqrt{\lambda^t_{\min}}}\right)+\frac{\|\theta_Q-\theta\|_{V_t}}{\sqrt{\lambda^t_{\min}}}+\varepsilon_Q,\\
\tau^\star_Q(\theta_Q,\hat{P}_t,\alpha) &\ge \tau^\star\left(\theta,P,\alpha+\zeta_t+\frac{\|\theta_Q-\theta\|_{V_t}}{\sqrt{\lambda^t_{\min}}}\right)-\frac{\|\theta_Q-\theta\|_{V_t}}{\sqrt{\lambda^t_{\min}}}
\end{aligned}
\tag{21}
$$

The proof of the above lemma relies on Equation (15) to relate $\tau^\star_Q(\cdot,\hat{P}_t,\cdot)$ to $\tau^\star_Q(\cdot,P,\cdot)$ and Lemma 9 to connect $\tau^\star_Q(\theta_Q,P,\cdot)$ to $\tau^\star_Q(\theta,P,\cdot)$.

Proof. Conditioning on the good event $G_{p_{\text{err}}}$, we have that

$$\begin{aligned}
\tau_Q^\star(\theta_Q, \hat{P}_t, \alpha) &= \min\{\tau_Q \in \mathcal{Q}_\tau : p_{\text{err}}(\theta_Q, \hat{P}_t, \tau_Q) \leq \alpha\} \\
&\overset{(a)}{\leq} \min\{\tau_Q \in \mathcal{Q}_\tau : p_{\text{err}}(\theta_Q, P, \tau_Q) \leq \alpha - \zeta_t\} \\
&\overset{(b)}{\leq} \min\left\{\tau_Q \in \mathcal{Q}_\tau : p_{\text{err}}(\theta, P, \tau_Q) \leq \alpha - \zeta_t - \frac{\|\theta_Q - \theta\|_{V_t}}{\sqrt{\lambda_{\min}^t}}\right\} + \frac{\|\theta_Q - \theta\|_{V_t}}{\sqrt{\lambda_{\min}^t}} \\
&\leq \min\left\{\tau \in [0,1] : p_{\text{err}}(\theta, P, \tau) \leq \alpha - \zeta_t - \frac{\|\theta_Q - \theta\|_{V_t}}{\sqrt{\lambda_{\min}^t}}\right\} + \frac{\|\theta_Q - \theta\|_{V_t}}{\sqrt{\lambda_{\min}^t}} + \varepsilon_Q \\
&= \tau^\star\left(\theta, P, \alpha - \zeta_t - \frac{\|\theta_Q - \theta\|_{V_t}}{\sqrt{\lambda_{\min}^t}}\right) + \frac{\|\theta_Q - \theta\|_{V_t}}{\sqrt{\lambda_{\min}^t}} + \varepsilon_Q
\end{aligned} \tag{22}$$

Where inequality (a) follows from conditioning on the good event $G_{p_{\text{err}}}$, and (b) follows from Lemma 9.

The lower bound for $\tau_Q^\star(\theta_Q, \hat{P}_t, \alpha)$ follows analogously:

$$\begin{aligned}
\tau_Q^\star(\theta_Q, \hat{P}_t, \alpha) &= \min\{\tau_Q \in \mathcal{Q}_\tau : p_{\text{err}}(\theta_Q, \hat{P}_t, \tau_Q) \leq \alpha\} \\
&\overset{(a)}{\geq} \min\{\tau_Q \in \mathcal{Q}_\tau : p_{\text{err}}(\theta_Q, P, \tau_Q) \leq \alpha + \zeta_t\} \\
&= \tau_Q^\star(\theta_Q, P, \alpha + \zeta_t) \\
&\geq \tau^\star(\theta_Q, P, \alpha + \zeta_t),
\end{aligned}$$

where (a) follows by the good event $G_{p_{\text{err}}}$, and the final inequality from the looseness of quantization. Now, we will lower bound $\tau^\star(\theta_Q, P, \alpha)$ in terms of $\tau^\star$ using Lemma 9.

$$\begin{aligned}
\tau^\star(\theta_Q, P, \alpha) &= \min\{\tau \in [0,1] : p_{\text{err}}(\theta_Q, P, \tau) \leq \alpha\} \\
&\overset{(a)}{\geq} \min\left\{\tau \in [0,1] : p_{\text{err}}\left(\theta, P, \tau + \frac{\|\theta_Q - \theta\|_{V_t}}{\sqrt{\lambda_{\min}^t}}\right) - \frac{\|\theta_Q - \theta\|_{V_t}}{\sqrt{\lambda_{\min}^t}} \leq \alpha\right\} \\
&\geq \min\left\{\tau \in [0,1] : p_{\text{err}}(\theta, P, \tau) \leq \alpha + \frac{\|\theta_Q - \theta\|_{V_t}}{\sqrt{\lambda_{\min}^t}}\right\} - \frac{\|\theta_Q - \theta\|_{V_t}}{\sqrt{\lambda_{\min}^t}} \\
&= \tau^\star\left(\theta, P, \alpha + \frac{\|\theta_Q - \theta\|_{V_t}}{\sqrt{\lambda_{\min}^t}}\right) - \frac{\|\theta_Q - \theta\|_{V_t}}{\sqrt{\lambda_{\min}^t}},
\end{aligned}$$

where (a) follows from Lemma 9.

□

Putting this all together we have that on $G_{p_{\text{err}}}, G_\theta$, evaluating at $\theta = \theta^\star$,

$$\begin{aligned}
\hat{\tau}(\theta_Q, \hat{P}_t, \alpha) &= \tau_Q^\star\left(\theta_Q, \hat{P}_t, \alpha - \zeta_t - 2B_t/\sqrt{\lambda_{\min}^t}\right) + 2B_t/\sqrt{\lambda_{\min}^t} \\
&\overset{(a)}{\geq} \tau^\star(\theta^\star, P, \alpha).
\end{aligned}$$

where (a) leverages Lemma 10. This uses the fact that when G_θ and $G_{p_{\text{err}}}$ hold,

$$\|\theta_Q - \theta^\star\|_{V_t} \leq 2B_t.$$

B.3 Smoothness of $\tau^\star$ with respect to α

The last property we will need for our analysis is that $\tau^\star$ does not vary too quickly with respect to α. We show that for small γ, $\tau^\star(\theta^\star, P, \alpha + \gamma)$ is not much smaller than $\tau^\star$. Note that while p_{err} is continuous with respect to

τ when evaluated at P the true distribution, it is discontinuous when evaluated at $\hat{P}$ because this is an empirical distribution.

However, by Assumption 2, the true distribution of contexts is upper and lower bounded by constants and so p_{err}, which integrates the distribution, will change at an upper and lower bounded rate. We leverage these properties to prove the following stability result.

Lemma 11. *Under Assumptions 1 and 2,*

$$\tau^\star(\theta^\star, P, \alpha - \gamma) \leq \tau^\star(\theta^\star, P, \alpha) + \frac{(1+e)\gamma}{m \cdot V_d(1) C_1(\tau^\star)}, \tag{23}$$

for $0 < \gamma < \min\{\frac{\tau^\star \cdot m \cdot V_d(1) \cdot C_1(\tau^\star)}{2(1+e)}, \frac{m \cdot V_d(1) \cdot f(\frac{1+\tau^\star}{2}) \cdot (1-\tau^\star)}{2(1+e)}\}$, *where* $f(\cdot)$ *is the PDF of* $Z \sim Beta(\frac{1}{2}, \frac{d+1}{2})$.

The proof proceeds as follows. First, we study the stability of $\tau^\star$ when the contexts follow the uniform distribution on the unit ball, characterizing the mass of contexts satisfying $||\langle X, \theta^\star, |\rangle \leq \tau^\star$ (Lemma 12). Then we use Assumption 2 to derive bounds for the unknown distribution P (Lemma 17). Finally, we leverage these upper and lower bounds to derive the stability of $\tau^\star$ with respect to α (Lemma 11).

Lemma 12. *For any* $0 \leq \tau \leq 1$ *intersection of* $\{\mathbf{x} : \|\mathbf{x}\| \leq 1\}$ *with* $\{\mathbf{x} : |\langle \mathbf{x}, \theta^\star \rangle| \leq \tau\}$ *is a spherical segment (see Figure 4) with volume equal to*

$$\int_{\|\mathbf{x}\| \leq 1} \mathbb{1}\{|\langle \mathbf{x}, \theta^\star \rangle| \leq \tau\} d\mathbf{x} = V_d \cdot I_{\tau^2}\left(\frac{1}{2}, \frac{d+1}{2}\right),$$

where V_d *is the volume of the d-dimensional unit ball, and* $I_x(a,b) = \frac{\int_0^x t^{a-1}(1-t)^{b-1} dt}{B(a,b)}$ *is the regularized Beta function Egorova et al. (2023), that is the cumulative distribution function of the Beta distribution.*

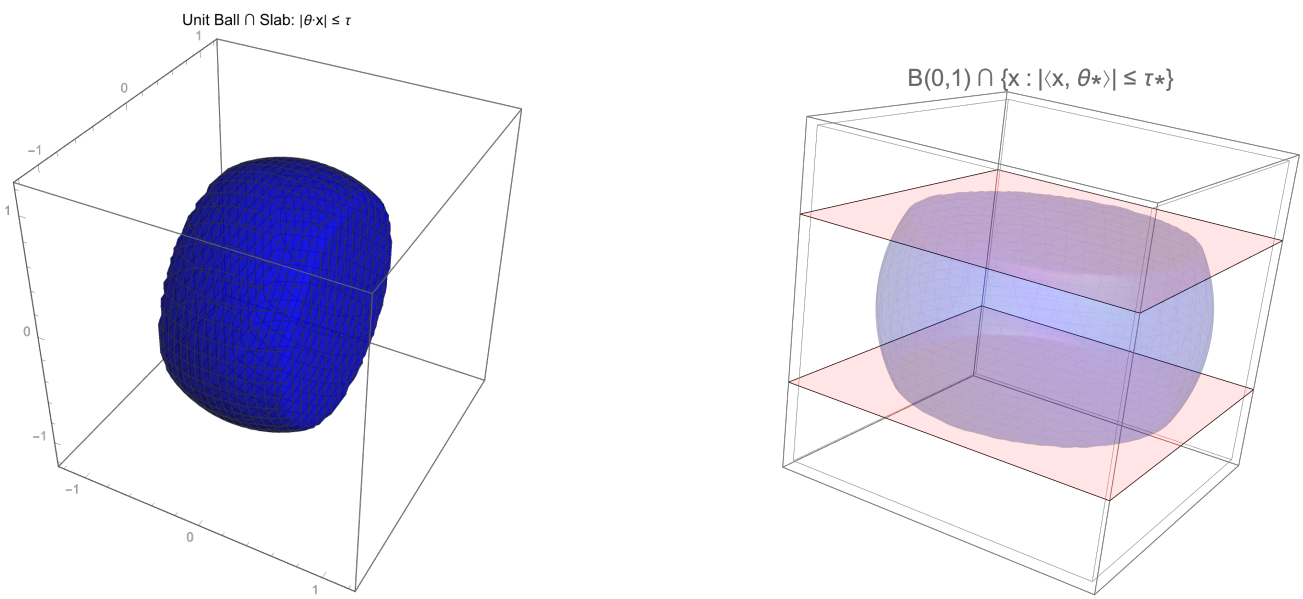


Figure 4: $\mathcal{B}(0,1) \cap \{x : |\langle x, \theta^\star \rangle| \leq \tau^\star\}$

Before proving Lemma 12 we will first prove an auxiliary lemma that allows us to work with a more convenient vector in the surface of the unit ball instead of $\theta^\star$. For more details about orthogonal transformations we refer the reader to Horn and Johnson (2012).

Lemma 13. *Let* $\theta, \theta' \in \mathbb{R}^d$ *be vectors on the unit sphere, i.e.* $\|\theta\| = \|\theta'\| = 1$. *Then there exists an orthogonal matrix* $S \in \mathbb{R}^{d \times d}$ *such that*

$$S\theta' = \theta.$$

Proof. If $\theta' = \theta$, the claim holds with $S = I$.

Otherwise, set

$$u \triangleq \frac{\theta' - \theta}{\|\theta' - \theta\|}, \qquad S \triangleq I - 2uu^T.$$

The matrix S is called a *Householder reflection*. It satisfies $S^\top S = I$, so it is orthogonal.

We compute

$$u^T\theta' = \frac{(\theta' - \theta)^T\theta'}{\|\theta' - \theta\|} = \frac{1 - \theta^T\theta'}{\|\theta' - \theta\|} = \frac{\|\theta' - \theta\|}{2},$$

since $\|\theta\| = \|\theta'\| = 1$ implies

$$\|\theta' - \theta\|^2 = \|\theta'\|^2 + \|\theta\|^2 - 2\theta'^T\theta = 2(1 - \theta'^T\theta).$$

Hence

$$S\theta' = \theta' - 2u(u^T\theta') = \theta' - u\|\theta' - \theta\| = \theta' - (\theta' - \theta) = \theta.$$

Thus S is an orthogonal matrix such that $S\theta' = \theta$. □

We will apply now this lemma for $\theta' = \theta^\star$ and $\theta = (0, 0, \ldots, 1)$ to compute the area of integration at Lemma 12.

Proof of Lemma 12. A similar proof, but for spherical caps, can be found in Li (2010). We follow similar steps to the didactic work of Jorgensen (2014).

For $\theta^\star \in \mathbb{R}^d$ with $\|\theta^\star\| = 1$, we have to integrate over all $\mathbf{x} \in \mathbb{R}^d$ such that

$$\{\mathbf{x}^\top\mathbf{x} \le 1\} \cap \{|\mathbf{x}^\top\theta^\star| \le \tau\}. \tag{24}$$

We apply Lemma 13 for $\theta' = \theta^\star$ and $\theta = (0, 0, \ldots, 1)^\top$. Then, let S be the orthogonal matrix such that

$$S\theta^\star = (0, 0, \ldots, 1)^\top.$$

We can use then Equation (24) to change the limits of integration;

$$\begin{aligned}\{\mathbf{x}^\top\mathbf{x} \le 1\} \cap \{|\mathbf{x}^\top\theta^\star| \le \tau\} &= \{\mathbf{x}^\top S^\top S\mathbf{x} \le 1\} \cap \{|\mathbf{x}^\top S^\top S\theta^\star| \le \tau\} \\ &= \{(S\mathbf{x})^\top(S\mathbf{x}) \le 1\} \cap \{|(S\mathbf{x})^\top(0, 0, \ldots, 1)| \le \tau\}\end{aligned}$$

Let $\tilde{\mathbf{x}} = S\mathbf{x}$ then the new integration domain is

$$\{\sum_{i=1}^{d-1} \tilde{x}_i^2 \le 1 - \tilde{x}_d^2\} \cap \{|\tilde{x}_d| \le \tau\}.$$

We define the volume of interest as

$$V_I = \int_{\|\mathbf{x}\| \le 1} \mathbb{1}\{|\langle \mathbf{x}, \theta^\star\rangle| \le \tau\} d\mathbf{x}. \tag{25}$$

By integrating first with respect to the first $n - 1$ dimensions and then to the last one we get

$$V_I = \int_{-\tau}^{\tau} \left(\int_{\{\mathbf{x} \in \mathbb{R}^{d-1} : \|\mathbf{x}\| \le \sqrt{1 - x_n^2}\}} dx_1 \ldots dx_{d-1} \right) dx_d.$$

Now, we can use that the volume of a sphere with radius r in d dimensions is equal to (Jorgensen, 2014)

$$V_d(r) = \frac{r^d \pi^{(d/2)}}{\Gamma(\frac{d}{2} + 1)},$$

and calculate the inner integral as

$$\int_{-\tau}^{\tau} \left(\int_{\{\mathbf{x} \in \mathbb{R}^{d-1} : \|\mathbf{x}\| \le 1 - x_n^2\}} dx_1 \ldots dx_{d-1} \right) dx_d = \frac{\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2} + 1)} \int_{-\tau}^{\tau} (1 - x_d^2)^{\frac{d-1}{2}} dx_d.$$

We use the fact that the function $1-x^2$ is even and the previous expression becomes

$$V_I = 2\frac{\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2}+1)}\int_0^{\tau}(1-x_d^2)^{\frac{d-1}{2}}dx_d.$$

We now make the change of variables, $x_d \triangleq \sqrt{t}$ and $dx_d = \frac{1}{2}t^{-\frac{1}{2}}dt$. The new limits of integration are; when $x_d = 0$ then $t = 0$ and when $x_d = \tau$, $t = \tau^2$.

$$\begin{aligned}
V_I &= \frac{\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2}+1)}\int_0^{\tau^2}(1-t)^{\frac{d-1}{2}}\cdot t^{-\frac{1}{2}}dt \\
&= \frac{\pi^{\frac{d-1}{2}}\Gamma(1/2)\Gamma(d/2+1)}{\Gamma(\frac{d-1}{2}+1)\Gamma(1/2)\Gamma(d/2+1)}\int_0^{\tau^2}(1-t)^{\frac{d-1}{2}}\cdot t^{-\frac{1}{2}}dt \\
&= \frac{\pi^{\frac{d}{2}}}{\Gamma(d/2+1)}\cdot\frac{\Gamma(d/2+1)}{\Gamma(1/2)\Gamma(\frac{d-1}{2}+1)}\cdot\int_0^{\tau^2}(1-t)^{\frac{d-1}{2}}\cdot t^{-\frac{1}{2}}dt,
\end{aligned}$$

where we used that $\Gamma(1/2) = \sqrt{\pi}$. We further use the definition of the Beta function $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ and that $V_d(1) = \frac{\pi^{\frac{d}{2}}}{\Gamma(d/2+1)}$ (see Jorgensen (2014)).

$$\begin{aligned}
V_I &= V_d(1)\cdot\frac{\int_0^{\tau^2}(1-t)^{\frac{d-1}{2}}\cdot t^{-\frac{1}{2}}dt}{B(\frac{1}{2},\frac{d+1}{2})} \\
&= V_d(1)I_{\tau^2}\left(\frac{1}{2},\frac{d+1}{2}\right).
\end{aligned}$$

□

We are interested in studying the stability of the previous quantity when we evaluate at $\tau-\lambda$, for $0<\lambda<\tau$ instead of at τ. This is the difference between the CDF of the Beta distribution evaluated at $(\tau-\lambda)^2$ and at τ^2, i.e. $I_{\tau^2}\left(\frac{1}{2},\frac{d+1}{2}\right) - I_{(\tau-\lambda)^2}\left(\frac{1}{2},\frac{d+1}{2}\right)$.

We will show that for the given parameters α,β for $Z\sim Beta(\alpha,\beta)$, the CDF $F(z) = \mathbb{P}(Z\le z)$ is a concave function. Then, we will bound the difference $F(1-\tau^2) - F(1-(\tau+\lambda)^2)$ by using standard arguments for increasing, concave functions that lie in $[0,1]$. These can be summarized in the following lemmata.

Lemma 14. *For $Z\sim Beta(\frac{1}{2},\frac{d+1}{2})$, $d\ge 1$, the CDF of Z is non-decreasing and concave over its support.*

Proof. Let $F(z) = \mathbb{P}(Z\le z)$. Then, $F'(z) =: f(z) > 0$ for all $z>0$, as f is a density, and so F is non-decreasing. We calculate the derivative of the density function by differentiating its logarithm.

$$\begin{aligned}
f(z) &= z^{-\frac{1}{2}}(1-z)^{\frac{d-1}{2}}, \\
\log(f(z)) &= -\frac{1}{2}\log z + \frac{d-1}{2}\log(1-z), \\
\log(f(z))' &= -\frac{1}{2z} - \frac{d-1}{2(1-z)} < 0.
\end{aligned}$$

Then, for all $0<z<1$, $f'(z) = \log(f(z))' f(z) < 0$, and so F is concave. Figure 5 illustrates the CDF across various values of parameter $d>1$.

□

To continue in our analysis, we will need to show that the τ for which we are evaluating stability is bounded away from one. Concretely, we wish to evaluate at $\tau < (1+\tau^\star)/2 < 1$ for stability purposes.

Lemma 15. *Under Assumption 2 $\tau^\star < 1$.*

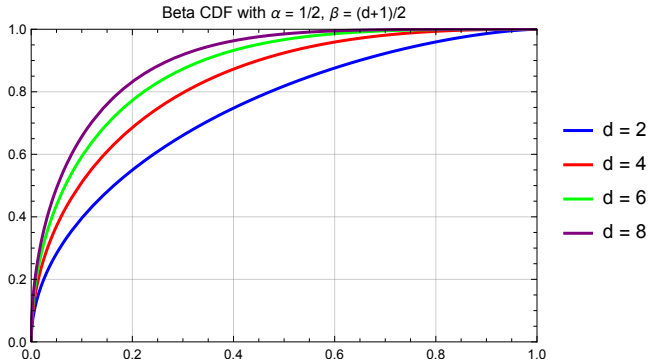


Figure 5: The CDF of $Beta(\frac{d+1}{2}, \frac{1}{2})$ for various values of d.

Proof. $\|\theta^\star\| = 1$, and $\|X\|_2 \leq 1$ a.s., and so $|\langle X, \theta^\star\rangle| \leq 1$.

Recall that p_{err} is defined as,

$$p_{\text{err}}(\theta, P, \tau) = \int (1 + \exp(|x^\top \theta|))^{-1} \mathbb{1}\left\{|x^\top \theta| > \tau\right\} P(dx).$$

We will show that $p_{\text{err}}(\cdot)$ is continuous in τ, that is for any $\tau_0 \in [0, 1]$

$$\lim_{\tau \to \tau_0} p_{\text{err}}(\theta, P, \tau) = p_{\text{err}}(\theta, P, \tau_0).$$

We will apply Lemma 13 to compute the integral

$$\int (1 + \exp(|x^\top \theta|))^{-1} \mathbb{1}\left\{|x^\top \theta| > \tau\right\} P(dx),$$

for $\theta, [0, 0, \ldots, 1]^\top$. Let S the orthogonal matrix such that $\theta = S \cdot [0, 0, \ldots, 1]^\top$.

For any x let $u = S \cdot x$, and u_i its i-th coordinate, we can write $x^\top \theta$ as

$$\begin{aligned} x^\top \theta &= x^\top S^\top S \theta \\ &= (Sx)^\top [0, 0, \ldots, 1]^\top \\ &= u_d. \end{aligned}$$

The inequality $|x^\top \theta| > \tau$ can be written as

$$\begin{aligned} |x^\top \theta| &> \tau \\ |x^\top S^\top S \theta| &> \tau \\ |u_d| &> \tau. \end{aligned}$$

By the change of variable $u \mapsto Sx$, we have that

$$\begin{aligned} \|Sx\|_2 = \|x\|_2 \leq 1 \iff \|u\|_2 \leq 1 \\ dx = |\det S^\top| du = du. \end{aligned}$$

Then, we have that

$$\int_{\|x\|_2\leq 1}(1+\exp(|x^\top\theta|))^{-1}\mathbb{1}\left\{|x^\top\theta|>\tau\right\}P(dx)=\int_{\|u\|_2\leq 1}(1+\exp(|u_d|))^{-1}\mathbb{1}\left\{|u_d|>\tau\right\}P(S^\top u)du$$

Now, to prove continuity we fix a sequence $\tau_n\to\tau$ for an arbitrary value of τ. We must prove now

$$\lim_{n\to\infty}\int_{\|u\|_2\leq 1}(1+\exp(|u_d|))^{-1}\mathbb{1}\left\{|u_d|>\tau_n\right\}P(S^\top u)du=\int_{\|u\|_2\leq 1}(1+\exp(|u_d|))^{-1}\mathbb{1}\left\{|u_d|>\tau\right\}P(S^\top u)du.$$

As $\tau_n\to\tau$ we know that for every $\varepsilon>0$ there exists $N(\varepsilon)\in\mathbb{N}$ such that for all $n\geq N(\varepsilon)$ it holds that $|\tau_n-\tau|<\varepsilon$. We will use the dominated convergence theorem (Theorem 2.24 Folland (1999)). Let

$$g_n(u)\triangleq(1+\exp(u_d))^{-1}\mathbb{1}\left\{|u_d|>\tau_n\right\}P(S^\top u),$$
$$g(u)\triangleq(1+\exp(u_d))^{-1}\mathbb{1}\left\{|u_d|>\tau\right\}P(S^\top u).$$

We will prove first that $g_n(u)\to g(u)$ almost everywhere. Equivalently we can prove that $\mathbb{1}\left\{|u_d|>\tau_n\right\}\to\mathbb{1}\left\{|u_d|>\tau\right\}$ almost everywhere. We will consider three cases for the range of values of u_d.

Consider three cases for the fixed real number $|u_d|$.

Case 1: $|u_d|>\tau$. Let $\varepsilon=\frac{1}{2}(|u_d|-\tau)>0$. For all $n\geq N(\varepsilon)$ such that $|\tau_n-\tau|<\varepsilon$ we have

$$\tau_n\leq\tau+\varepsilon<\tau+\tfrac{1}{2}(|u_d|-\tau)=\tfrac{1}{2}(\tau+|u_d|)<|u_d|,$$

so $|u_d|>\tau_n$ and therefore $g_n(u)=1$. Hence $g_n(u)=1=g(u),\forall n\geq N(\varepsilon)$.

Case 2: $|u_d|<\tau$. Let $\varepsilon=\frac{1}{2}(\tau-|u_d|)>0$. For all sufficiently large $n\geq N(\varepsilon)$ with $|\tau_n-\tau|<\varepsilon$ we get

$$\tau_n\geq\tau-\varepsilon>\tau-\tfrac{1}{2}(\tau-|u_d|)=\tfrac{1}{2}(\tau+|u_d|)>|u_d|,$$

so $|u_d|\leq\tau_n$ and $g_n(u)=0$. Hence $g_n(u)=0=g(u)$.

Case 3: $|u_d|=\tau$. For the third case an alternating sequence would not converge but it does not matter as the set $\{u\in\mathcal{B}(0,1):|u_d|=\tau\}$ has measure zero under P.

As a result now we proved that $g_n(u)\to g(u)$ almost everywhere. Moreover, $0\leq g_n(u)\leq\frac{1}{2}$ for all $n\in\mathbb{N}$ for every u. By applying the dominated convergence theorem, we get that $p_{\text{err}}(\cdot)$ is continuous at τ;

$$\lim_{n\to\infty}\int_{\|u\|_2\leq 1}g_n(u)du=\int_{\|u\|_2\leq 1}g(u)du$$
$$\lim_{n\to\infty}\int_{\|u\|_2\leq 1}(1+\exp(|u_d|))^{-1}\mathbb{1}\left\{|u_d|>\tau_n\right\}P(S^\top u)du=\int_{\|u\|_2\leq 1}(1+\exp(|u_d|))^{-1}\mathbb{1}\left\{|u_d|>\tau\right\}P(S^\top u)du.$$

Now, we will show that $p_{\text{err}}(\cdot)$ is strictly decreasing in τ. Let $P_d(S^\top u)$ the marginal distribution at the d-th coordinate. Then for $\tau_1<\tau_2$

$$\begin{aligned}p_{\text{err}}(\theta,P,\tau_1)-p_{\text{err}}(\theta,P,\tau_2)&=\int_{-1}^{1}(1+\exp(|u_d|))^{-1}\left(\mathbb{1}\left\{|u_d|>\tau_1\right\}-\mathbb{1}\left\{|u_d|>\tau_2\right\}\right)P_d(S^\top u)du_d\\&=\int_{-\tau_2}^{-\tau_1}(1+\exp(|u_d|))^{-1}P_d(S^\top u)du_d+\int_{\tau_1}^{\tau_2}(1+\exp(|u_d|))^{-1}P_d(S^\top u)du_d\\&>0\end{aligned}$$

where we have strict inequality as $P_d(S^\top u)>0$ by Assumption 2.

This concludes the proof that $p_{\text{err}}(\cdot)$ is strictly decreasing as a function of τ.

Since $x^\top\theta \leq 1$ for all x, as $\|x\|_2 \leq 1, \|\theta\|_2$=1. It follows that indicator$(|x^\top\theta| > \tau) = 0$ and therefore that $p_{\text{err}}(\theta^\star, P, 1) = 0$.

Finally, since $p_{\text{err}}(\theta^\star, P, 1) = 0 < \alpha$, and $p_{\text{err}}(\theta^\star, P, \tau)$ is a strictly monotone (decreasing) and continuous function of τ, we get that $\tau^\star < 1$.

□

Now, analyzing the Beta CDF by using concavity, monotonicity, and the fact that $F(0) = 0$ and $F(1) = 1$ (Lemma 14) we will derive upper and lower bounds for the difference $F(\tau^2) - F\left((\tau-\lambda)^2\right)$. As in our algorithm we design a sequence of threshold converging to the real one, one can imagine λ as part of a sequence $\{\lambda_t\}$ that converges to zero.

Lemma 16. *Under Assumption 1, for all* $0 < \lambda < \frac{\tau^\star}{2} < \tau < \frac{1+\tau^\star}{2} < 1$ *there exist functions* $C_1(\tau^\star), C_2(\tau^\star, \lambda)$ *such that it holds that;*

$$C_1(\tau^\star)\cdot\lambda \leq F\left(\tau^2\right) - F\left((\tau-\lambda)^2\right) \leq C_2(\tau^\star,\lambda)\cdot\lambda,$$

where $F(\cdot)$ *denotes the CDF of the random variable* $Z \sim Beta(\alpha,\beta)$ *and* f *its density.* $C_1(\tau^\star), C_2(\tau^\star,\lambda)$ *are defined as follows;*

$$C_1(\tau^\star) \triangleq \frac{\tau^\star}{2}\left(1 - F\left(\frac{(1+\tau^\star)^2}{4}\right)\right),$$
$$C_2(\tau^\star,\lambda) \triangleq 2\frac{1}{(\frac{\tau^\star}{2}-\lambda)^2}.$$

Proof. We apply the mean value theorem in the intervals $[0,(\tau-\lambda)^2], [(\tau-\lambda)^2,\tau^2], [\tau^2,1]$.

By applying the mean value theorem to these intervals there exists $\xi_1 \in (0,(\tau-\lambda)^2), \xi_2 \in ((\tau-\lambda)^2,\tau^2), \xi_3 \in (\tau^2,1)$ such that

$$\frac{F\left((\tau-\lambda)^2\right) - F(0)}{(\tau-\lambda)^2} = F'(\xi_1) = f(\xi_1),$$
$$\frac{F\left(\tau^2\right) - F\left((\tau-\lambda)^2\right)}{\tau^2-(\tau-\lambda)^2} = F'(\xi_2) = f(\xi_2),$$
$$\frac{F(1) - F\left(\tau^2\right)}{1-\tau^2} = F'(\xi_3) = f(\xi_3).$$

As $F(0) = 0, F(1) = 1$, $F'(x) = f(x)$ and $f'(x) < 0$ it holds that

$$f(\xi_1) \geq f(\xi_2) \geq f(\xi_3).$$

We replace the values of $f(\xi_1), f(\xi_2), f(\xi_3)$;

$$\frac{F(1) - F\left(\tau^2\right)}{1-\tau^2} \leq \frac{F\left(\tau^2\right) - F\left((\tau-\lambda)^2\right)}{\tau^2-(\tau-\lambda)^2} \leq \frac{F\left((\tau-\lambda)^2\right) - F(0)}{(\tau-\lambda)^2}. \tag{26}$$

Using that $0 < \lambda < \frac{\tau^\star}{2} < \tau$ and $0 < F(\cdot) \leq 1$ we can upper bound $\frac{F\left((\tau-\lambda)^2\right)-F(0)}{(\tau-\lambda)^2}$ as follows

$$\frac{F\left((\tau-\lambda)^2\right) - F(0)}{(\tau-\lambda)^2} \leq \frac{1}{(\frac{\tau^\star}{2}-\lambda)^2}. \tag{27}$$

As $F(\cdot)$ is increasing, $F(1) = 1$ and $F(\tau^2) \leq F((1+\tau^\star)^2/4)$ since by assumption $\tau < \frac{1+\tau^\star}{2}$, we also have that

$$\frac{F(1) - F\left(\tau^2\right)}{1-\tau^2} \geq 1 - F\left(\frac{(1+\tau^\star)^2}{4}\right). \tag{28}$$

In order to derive an upper and lower bound for the middle term of Equation (26), it remains to upper and lower bound its denominator; $\tau^2 - (\tau-\lambda)^2 = 2\lambda\tau - \lambda^2$ as

$$\lambda\frac{\tau^\star}{2} \overset{(i)}{\leq} \tau^2 - (\tau-\lambda)^2 \overset{(ii)}{\leq} 2\lambda. \tag{29}$$

For the lower bound (i) of Equation (29) we used the inequalities

$$2\lambda\tau - \lambda^2 = \lambda(2\tau-\lambda) \geq \lambda\tau \geq \lambda\frac{\tau^\star}{2},$$

where the inequalities hold because $\lambda < \frac{\tau^\star}{2} < \tau$. For the upper bound (ii) in Equation (29) we used that $\tau < 1$. By replacing Equations (27) to (29) into Equation (26) we have that;

$$\frac{\tau^\star}{2}\left(1 - F\left(\frac{(1+\tau^\star)^2}{4}\right)\right)\lambda \leq F\left(\tau^2\right) - F\left((\tau-\lambda)^2\right) \leq 2\frac{1}{(\frac{\tau^\star}{2}-\lambda)^2}\lambda.$$

Defining the functions $C_1(\tau^\star), C_2(\tau^\star,\lambda)$ as

$$C_1(\tau^\star) \triangleq \frac{\tau^\star}{2}\left(1 - F\left(\frac{(1+\tau^\star)^2}{4}\right)\right),$$
$$C_2(\tau^\star,\lambda) \triangleq 2\frac{1}{(\frac{\tau^\star}{2}-\lambda)^2}.$$

we obtain the desired result. □

With these results in place, we are able to upper and lower bound the volume in this spherical segment.

Lemma 17. *Under Assumption 2, for all* $0 < \lambda < \frac{\tau^\star}{2} < \tau < \frac{1+\tau^\star}{2} < 1$*, we have that*

$$m \cdot V_d(1) \cdot C_1(\tau^\star) \cdot \lambda \leq \mathbb{P}\left(\tau - \lambda < |X^\top\theta^\star| \leq \tau\right) \leq M \cdot V_d(1) \cdot C_2(\tau^\star,\lambda) \cdot \lambda.$$

Proof. We first use that

$$\mathbb{P}\left(\tau - \lambda < |X^\top\theta^\star| \leq \tau\right) = \int_{\|\mathbf{x}\|\leq 1} \left(\mathbb{1}\{|\langle\mathbf{x},\theta^\star\rangle| \leq \tau\} - \mathbb{1}\{|\langle\mathbf{x},\theta^\star\rangle| \leq \tau-\lambda\}\right) p(\mathbf{x})d\mathbf{x} \tag{30}$$

We can use the smoothness property of our distribution to sandwich Equation (30) as

$$\begin{aligned}
&m\int_{\|\mathbf{x}\|\leq 1} \left(\mathbb{1}\{|\langle\mathbf{x},\theta^\star\rangle| \leq \tau\} - \mathbb{1}\{|\langle\mathbf{x},\theta^\star\rangle| \leq \tau-\lambda\}\right) d\mathbf{x} \\
&\leq \mathbb{P}\left(\tau < |X^\top\theta^\star| \leq \tau+\lambda\right) \\
&\leq M\int_{\|\mathbf{x}\|\leq 1} \left(\mathbb{1}\{|\langle\mathbf{x},\theta^\star\rangle| \leq \tau\} - \mathbb{1}\{|\langle\mathbf{x},\theta^\star\rangle| \leq \tau-\lambda\}\right) d\mathbf{x}.
\end{aligned}$$

Now, let $Z \sim Beta(\frac{1}{2}, \frac{d+1}{2})$ and $F(\cdot)$ its CDF function, then, Lemma 12 allows us to write the integral as

$$\int_{\|\mathbf{x}\|\leq 1} \left(\mathbb{1}\{|\langle\mathbf{x},\theta^\star\rangle| \leq \tau\} - \mathbb{1}\{|\langle\mathbf{x},\theta^\star\rangle| \leq \tau-\lambda\}\right) d\mathbf{x} = V_d(1)\left(F\left((\tau)^2\right) - F\left((\tau-\lambda)^2\right)\right),$$

and the previous equation becomes

$$
\begin{aligned}
&mV_d(1)\left(F\left((\tau)^2\right) - F\left((\tau-\lambda)^2\right)\right)\\
&\leq \mathbb{P}\left(\tau < |X^\top\theta^\star| \leq \tau+\lambda\right)\\
&\leq MV_d(1)\left(F\left((\tau)^2\right) - F\left((\tau-\lambda)^2\right)\right).
\end{aligned}
$$

Finally, we apply Lemma 16 to lower and upper bound $V_d(1)\left(F\left((\tau)^2\right) - F\left((\tau-\lambda)^2\right)\right)$ and conclude the proof.

$$
m \cdot V_d(1) \cdot C_1(\tau^\star) \cdot \lambda \leq \mathbb{P}\left(\tau - \lambda < |X^\top\theta^\star| \leq \tau\right) \leq M \cdot V_d(1) \cdot C_2(\tau^\star, \lambda) \cdot \lambda.
$$

□

Before proving Lemma 11, we first prove an auxiliary lemma to derive a range of γ for which we can apply Lemma 17, i.e. $\frac{\tau^\star}{2} < \tau < \frac{1+\tau^\star}{2}$.

Lemma 18. *For any* $0 < \gamma < \frac{m \cdot V_d(1) \cdot f(\frac{1+\tau^\star}{2}) \cdot (1-\tau^\star)}{2(1+e)}$ *it holds that*[0]

$$
\min\left\{\tau \in \left[\frac{\tau^\star}{2}, 1\right] : p_{err}(\theta^\star, P, \tau) \leq \alpha - \gamma\right\} = \min\left\{\tau \in \left[\frac{\tau^\star}{2}, \frac{1+\tau^\star}{2}\right] : p_{err}(\theta^\star, P, \tau) \leq \alpha - \gamma\right\}.
$$

Proof. To prove this, we show that for these values of γ there exists a $\tau(\gamma) \in [\tau^\star, \frac{1+\tau^\star}{2}] \subset [\frac{\tau^\star}{2}, \frac{1+\tau^\star}{2}]$ such that $p_{\text{err}}(\theta^\star, P, \tau(\gamma)) = \alpha - \gamma$. Thus,

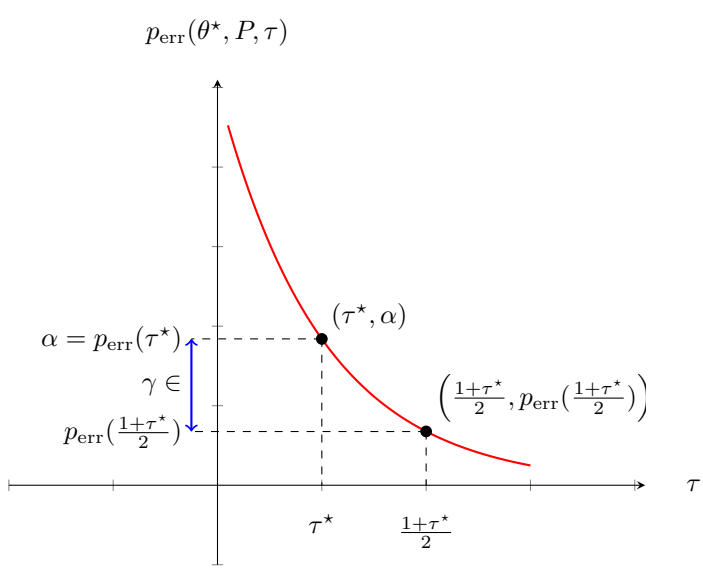
$$
\begin{aligned}
\gamma &\overset{(a)}{=} p_{\text{err}}(\theta^\star, P, \tau^\star) - p_{\text{err}}(\theta^\star, P, \tau(\gamma))\\
&\overset{(b)}{\leq} p_{\text{err}}(\theta^\star, P, \tau^\star) - p_{\text{err}}\left(\theta^\star, P, \frac{1+\tau^\star}{2}\right).
\end{aligned}
$$

(a) uses that $p_{\text{err}}(\theta^\star, P, \tau)$ is strictly decreasing and continuous with respect to its third argument τ (see the proof of Lemma 15), thus $p_{\text{err}}(\theta^\star, P, \tau^\star) = \alpha$, and in (b) the monotonicity of $p_{\text{err}}(\theta^\star, P, \tau)$. It now remains to find a lower bound for

$$
p_{\text{err}}(\theta^\star, P, \tau^\star) - p_{\text{err}}\left(\theta^\star, P, \frac{1+\tau^\star}{2}\right).
$$

We remind the reader that by definition of $p_{\text{err}}(\cdot)$

$$
p_{\text{err}}(\theta, P, \tau) = \int (1 + \exp(|x^\top\theta|))^{-1} \mathbb{1}\left\{|x^\top\theta| > \tau\right\} P(dx).
$$



[0] $f(\cdot)$ is the PDF of the random variable $Z \sim Beta(\alpha, \beta)$.

$$
\begin{aligned}
p_{\text{err}}(\theta^\star, P, \tau^\star) - p_{\text{err}}(\theta^\star, P, (1+\tau^\star)/2) &= \int (1+\exp(|x^\top\theta|))^{-1} \left(\mathbb{1}\left\{|x^\top\theta| > \tau^\star\right\} - \mathbb{1}\left\{|x^\top\theta| > \frac{(1+\tau^\star)}{2}\right\}\right) P(dx) \\
&\overset{(a)}{\geq} \frac{1}{1+e} \int \mathbb{1}\left\{\tau^\star \leq |x^\top\theta| \leq \frac{(1+\tau^\star)}{2}\right\} P(dx) \\
&\overset{(b)}{\geq} \frac{m \cdot V_d(1)}{1+e} \int \mathbb{1}\left\{\tau^\star \leq |x^\top\theta| \leq \frac{(1+\tau^\star)}{2}\right\} \frac{1}{V_d(1)} dx \\
&\overset{(c)}{=} \frac{m \cdot V_d(1)}{1+e} \left(F\left(\frac{(1+\tau^\star)}{2}\right) - F(\tau^\star)\right).
\end{aligned}
$$

where (a) comes from $|x^\top\theta| \leq 1$, (b) from Assumption 2 and (c) from Lemma 12 (recall that $F(\cdot)$ is the CDF of the random variable $Z \sim \text{Beta}(\alpha, \beta)$). To derive a lower bound for $F\left(\frac{(1+\tau^\star)}{2}\right) - F(\tau^\star)$ we will use the Mean Value Theorem as in Lemma 16 applied in $[\tau^\star, \frac{1+\tau^\star}{2}]$ for $F(\cdot)$. Then, there exists a $\xi \in (\tau^\star, \frac{1+\tau^\star}{2})$ such that

$$
\frac{F\left(\frac{(1+\tau^\star)}{2}\right) - F(\tau^\star)}{\frac{(1+\tau^\star)}{2} - \tau^\star} = f(\xi) \geq f\left(\frac{1+\tau^\star}{2}\right),
$$

where $f(\cdot)$ is a decreasing function as we proved in Lemma 16.

Combining the above we get

$$
p_{\text{err}}(\theta^\star, P, \tau^\star) - p_{\text{err}}(\theta^\star, P, (1+\tau^\star)/2) \geq \frac{m \cdot V_d(1) \cdot f(\frac{1+\tau^\star}{2}) \cdot (1-\tau^\star)}{2(1+e)}.
$$

As a consequence for all $\gamma \in [0, \frac{m \cdot V_d(1) \cdot f(\frac{1+\tau^\star}{2}) \cdot (1-\tau^\star)}{2(1+e)}]$ we know that

$$
p_{\text{err}}(\theta^\star, P, (1+\tau^\star)/2) \leq \alpha - \gamma,
$$

and

$$
\min\left\{\tau \in \left[\frac{\tau^\star}{2}, 1\right] : p_{\text{err}}(\theta^\star, P, \tau) \leq \alpha - \gamma\right\} = \min\left\{\tau \in \left[\frac{\tau^\star}{2}, \frac{1+\tau^\star}{2}\right] : p_{\text{err}}(\theta^\star, P, \tau) \leq \alpha - \gamma\right\}.
$$

□

B.3.1 Proof of Lemma 11

Lemma 11. *Under Assumptions 1 and 2,*

$$
\tau^\star(\theta^\star, P, \alpha - \gamma) \leq \tau^\star(\theta^\star, P, \alpha) + \frac{(1+e)\gamma}{m \cdot V_d(1) C_1(\tau^\star)}, \tag{23}
$$

for $0 < \gamma < \min\{\frac{\tau^\star \cdot m \cdot V_d(1) \cdot C_1(\tau^\star)}{2(1+e)}, \frac{m \cdot V_d(1) \cdot f(\frac{1+\tau^\star}{2}) \cdot (1-\tau^\star)}{2(1+e)}\}$, *where* $f(\cdot)$ *is the PDF of* $Z \sim Beta(\frac{1}{2}, \frac{d+1}{2})$.

Proof. For arbitrary $\tau < 1$, we begin by studying the difference between p_{err} evaluated at thresholds $\tau - \lambda$ and τ. By applying Lemma 17, for all $0 < \lambda < \frac{\tau^\star}{2} < \tau < \frac{1+\tau^\star}{2} < 1$ it is true that;

$$
\begin{aligned}
p_{\text{err}}(\theta^\star, P, \tau-\lambda) - p_{\text{err}}(\theta^\star, P, \tau) &= \int (1+\exp(|\langle x, \theta^\star\rangle|))^{-1} \mathbb{1}\{\tau-\lambda < |\langle x, \theta^\star\rangle| < \tau\} P(dx) \\
&\geq \int \frac{1}{1+e} \mathbb{1}\{\tau-\lambda < |\langle x, \theta^\star\rangle| < \tau\} P(dx) \\
&= \frac{1}{1+e} \mathbb{P}(\tau-\lambda < |\langle x, \theta^\star\rangle| < \tau) \\
&\geq \frac{m}{1+e} \cdot V_d(1) \cdot C_1(\tau^\star) \cdot \lambda, \qquad (31)
\end{aligned}
$$

$$
\begin{aligned}
\tau^\star(\theta^\star, P, \alpha-\gamma) &= \min\{\tau \in [0,1] : p_{\text{err}}(\theta^\star, P, \tau) \leq \alpha-\gamma\} \\
&\overset{(a)}{\leq} \min\left\{\tau \in \left[\frac{\tau^\star}{2}, \frac{1+\tau^\star}{2}\right] : p_{\text{err}}(\theta^\star, P, \tau) \leq \alpha-\gamma\right\} \\
&\overset{(b)}{\leq} \min\left\{\tau \in \left[\frac{\tau^\star}{2}, \frac{1+\tau^\star}{2}\right] : p_{\text{err}}(\theta^\star, P, \tau-\lambda) \leq \alpha-\gamma+\frac{m}{1+e} \cdot V_d(1) \cdot C_1(\tau^\star) \cdot \lambda\right\} \\
&\overset{(c)}{\leq} \min\left\{\tau \in \left[\frac{\tau^\star}{2}, \frac{1+\tau^\star}{2}\right] : p_{\text{err}}\left(\theta^\star, P, \tau-\frac{(1+e)\gamma}{m \cdot V_d(1) C_1(\tau^\star)}\right) \leq \alpha\right\} \\
&\overset{(d)}{\leq} \min\left\{\tau \in \left[\frac{\tau^\star}{2}, \frac{1+\tau^\star}{2}\right] : p_{\text{err}}(\theta^\star, P, \tau) \leq \alpha\right\} + \frac{(1+e)\gamma}{m \cdot V_d(1) C_1(\tau^\star)} \\
&= \tau^\star(\theta^\star, P, \alpha) + \frac{(1+e)\gamma}{m \cdot V_d(1) C_1(\tau^\star)}.
\end{aligned}
$$

In (a) we used Lemma 18, in (b) we leveraged the p_{err} difference bound derived in Equation (31), (c) follows from setting $\lambda = \frac{(1+e)\gamma}{m \cdot V_d(1) C_1(\tau^\star)}$, and (d) from Lemma 6 by setting $x \triangleq \frac{(1+e)\gamma}{m \cdot V_d(1)}$. We observe that for $0 < \gamma < \min\{\frac{\tau^\star \cdot m \cdot V_d(1) \cdot C_1(\tau^\star)}{2(1+e)}, \frac{m \cdot V_d(1) \cdot f(\frac{1+\tau^\star}{2}) \cdot (1-\tau^\star)}{2(1+e)}\}$, we satisfy the condition of Lemma 17. □

C Safety analysis

We begin by providing a sketch of the results proved in this section. First, in Appendix C.1 we prove Lemma 21, which is an analogue of Lemma 9 but with ℓ_2 error, to show that shifting from θ to $\theta_Q \in \mathcal{Q}_\theta$ doesn't change τ much. Then, we have the following two safety lemmas, which compare `SCOUT` 's performance with the optimal testing policy for confidence α_t, i.e. $Z_t^\star = \mathbb{1}\{|\langle X_t, \theta^\star\rangle| \leq \tau^\star(\theta^\star, P, \alpha_t\}$.

Lemma 19. *The testing rule Z_t defined in Algorithm 1 satisfies, conditioned on $G_{p_{err}}$ and G_θ, $Z_t^\star = 1 \implies Z_t = 1$, i.e. $Z_t \geq Z_t^\star$ a.s.*

This follows by the monotonicity of the threshold $\tau^\star$ with respect to α and by using a "safer" error tolerance α_t than α. We defer the proof to Appendix C.2.

Another property of our testing rule is that when $G_{p_{\text{err}}}$ holds it makes no more errors than the baseline policy. As formalized in the following lemma, `SCOUT` 's predictions are identical to those of the oracle policy when it does not test, ensuring its (α, δ)-safety.

Lemma 20. *Let $\hat{Y}_t$ be the prediction of our policy, where $Y_t^\star$ is the prediction of the oracle baseline policy. When $G_{p_{err}}$ and G_θ holds, and $Z_t = 0$ (which implies that $Z_t^\star = 0$) then $\hat{Y}_t = \hat{Y}_t^\star$.*

To show the previous lemma, we use the fact that, on the good event, when we do not test, all the inner products $\langle X_t, \theta^\star\rangle$ have the same sign. We defer the proof to Appendix C.3.

More formally, we define the Bernoulli random variable $\xi_t = \mathbb{1}\{\hat{Y}_t \neq Y_t\}$, that denotes whether the algorithm made a mistake at round t, and $\xi_t^\star = \mathbb{1}\{\hat{Y}_t^\star \neq Y_t\}$ respectively for the baseline policy. When the algorithm tests

(i.e. $Z_t = 1$) then we observe the label and it holds that $\xi_t = 0$. Conditioning on the good event G, the random variables ξ_t and $\xi_t^\star$ satisfy $\xi_t \leq \xi_t^\star$ (formalized in Appendix C.4). This implies a total error probability bound, stated in the following lemma.

C.1 τ stability lemma

The safety analysis requires the application of Lemma 10 for $\theta \triangleq \theta_t^L$. However, it is not guaranteed that $\theta_t^L \in \mathcal{Q}_\theta$. To surpass this technical detail, we use the stability of p_{err} in θ, similar to Lemma 9, but expressing the result in the ℓ_2 distance, the metric with respect to which the covering is defined.

Lemma 21. *For all $\theta, \theta' \in \Theta$, $\tau \geq \|\theta - \theta'\|_2$, and density $\rho(x)$ on $\mathcal{X}$:*

$$p_{err}(\theta, \rho, \tau) \leq \; p_{err}\left(\theta', \rho, \tau - \|\theta - \theta'\|_2\right) + \|\theta - \theta'\|_2.$$

Proof. Here, we use x as a dummy variable for integration:

$$\begin{aligned}
p_{\text{err}}(\theta, \rho, \tau) &= \int (1 + \exp(|x^\top \theta|))^{-1} \mathbb{1}\left\{|x^\top \theta| > \tau\right\} \rho(dx) \\
&= \int (1 + \exp(|x^\top \theta' + x^\top(\theta - \theta')|))^{-1} \mathbb{1}\left\{|x^\top \theta' + x^\top(\theta - \theta')| > \tau\right\} \rho(dx) \\
&\leq \int (1 + \exp(|x^\top \theta'| - |x^\top(\theta - \theta')|))^{-1} \mathbb{1}\left\{|x^\top \theta'| > \tau - |x^\top(\theta - \theta')|\right\} \rho(dx) \\
&\leq \int \left((1 + \exp(|x^\top \theta'|))^{-1} + |x^\top(\theta - \theta')|\right) \mathbb{1}\left\{|x^\top \theta'| > \tau - |x^\top(\theta - \theta')|\right\} \rho(dx) \\
&\leq \max_{x' \in \mathcal{X}} \int \left((1 + \exp(|x^\top \theta'|))^{-1} + |x'^\top(\theta - \theta')|\right) \mathbb{1}\left\{|x^\top \theta'| > \tau - |x'^\top(\theta - \theta')|\right\} \rho(dx) \\
&= \max_{x' \in \mathcal{X}} p_{\text{err}}(\theta', \rho, \tau - |x'^\top(\theta - \theta')|) + \int |x^\top(\theta - \theta')| \mathbb{1}\left\{|x^\top \theta'| > \tau - |x'^\top(\theta - \theta')|\right\} \rho(dx) \\
&\leq \max_{x' \in \mathcal{X}} p_{\text{err}}(\theta', \rho, \tau - \|\theta - \theta'\|_2 \|x'\|_2) + \|\theta - \theta'\|_2 \|x'\|_2 \mathbb{P}_\rho\left(|x^\top \theta'| > \tau - |x^\top(\theta - \theta')|\right) \\
&\leq \max_{x' \in \mathcal{X}} p_{\text{err}}(\theta', \rho, \tau - \|\theta - \theta'\|_2 \|x'\|_2) + \|\theta - \theta'\|_2 \|x'\|_2 \\
&= p_{\text{err}}(\theta', \rho, \tau - \|\theta - \theta'\|_2) + \|\theta - \theta'\|_2
\end{aligned}$$

The details of this proof are identical to those of Lemma 9. We also make use that our contexts lie in the unit ball, i.e. $\|x\|_2 \leq 1$.

□

Using the previous lemma we derive a similar expression to that of Lemma 10;

$$\tau_Q^\star(\theta_Q, \hat{P}_t, \alpha) \geq \tau_Q^\star\left(\theta, \hat{P}_t, \alpha + \|\theta - \theta_Q\|_2\right) - \|\theta - \theta_Q\|_2. \tag{32}$$

C.2 Proof of Lemma 19

Proof. Let $\tilde{\theta}_t^L \in Q_\theta$ such that $\left\|\tilde{\theta}_t^L - \theta_t^L\right\|_2 \leq \varepsilon_Q$, as θ_t^L lies in the interior of $\mathcal{C}_t$.

Leveraging Lemma 21, we relate θ_t^L to $\tilde{\theta}_t^L$ as (using the definition of Equation (8)), on the good events $G_{p_{\text{err}}}$ and G_θ:

$$\begin{aligned}
\tau_t &= \tau^\star\left(\theta_t^L, \hat{P}_t, \alpha_t - \zeta_t - 2B_t/\sqrt{\lambda_{\min}^t} - \varepsilon_Q\right) + 3B_t/\sqrt{\lambda_{\min}^t} + \varepsilon_Q \\
&= \hat{\tau}\left(\theta_t^L, \hat{P}_t, \alpha_t - \varepsilon_Q\right) + B_t/\sqrt{\lambda_{\min}^t} + \varepsilon_Q \\
&\geq \hat{\tau}\left(\tilde{\theta}_t^L, \hat{P}_t, \alpha_t\right) + B_t/\sqrt{\lambda_{\min}^t} \\
&\geq \tau^\star\left(\theta^\star, P, \alpha_t\right) + B_t/\sqrt{\lambda_{\min}^t}
\end{aligned} \tag{33}$$

Here, we used the monotonicity of $\tau^\star$ with respect to α, in addition to Lemma 10. Then, we upper bound the inner product:

$$|\langle X_t, \theta_t^L\rangle| \leq |\langle X_t, \theta^\star\rangle| + \|\theta_t^L - \theta^\star\|_{V_t}\|X_t\|_{V_t^{-1}} \leq |\langle X_t, \theta^\star\rangle| + B_t/\sqrt{\lambda_{\min}^t}$$

By Holder. Combining these together yields that, on $G_{p_{\text{err}}}$ and G_θ,

$$|\langle X_t, \theta^\star\rangle| \leq \tau^\star\left(\theta^\star, P, \alpha_t\right) \quad \Longrightarrow \quad |\langle X_t, \theta_t^L\rangle| \leq \tau_t. \tag{34}$$

i.e. $Z_t^\star = 1 \implies Z_t = 1$ □

C.3 Proof of Lemma 20

Proof. On $G_{p_{\text{err}}}$ and G_θ, we have that $Z_t = 0$ implies that $\langle \theta, X_t\rangle$ has the same sign for all $\theta \in \mathcal{C}_t$. This is because, $Z_t = 0$ only when:

$$|\langle \theta_t^L, X_t\rangle| \geq \tau_t = \tau^\star\left(\theta_t^L, \hat{P}_t, \alpha_t - \zeta_t - 2B_t/\sqrt{\lambda_{\min}^t} - \varepsilon_Q\right) + 3B_t/\sqrt{\lambda_{\min}^t} + \varepsilon_Q.$$

As before, we know that

$$\tau_t \geq \tau^\star\left(\theta^\star, P, \alpha_t\right) + B_t/\sqrt{\lambda_{\min}^t}$$

We also have that for all $\theta \in \mathcal{C}_t$:

$$|\langle \theta_t^L, X_t\rangle - \langle \theta, X_t\rangle| \leq B_t/\sqrt{\lambda_{\min}^t}.$$

Thus, if $|\langle \theta_t^L, X_t\rangle| \geq \tau_t$, and assuming without loss of generality that $\langle \theta_t^L, X_t\rangle > 0$, then for all $\theta \in \mathcal{C}_t$:

$$\begin{aligned}
0 &\leq \langle \theta_t^L, X_t\rangle - \tau_t \\
&\leq \left(\langle \theta, X_t\rangle + B_t/\sqrt{\lambda_{\min}^t}\right) - \left(\tau^\star\left(\theta^\star, P, \alpha_t\right) + B_t/\sqrt{\lambda_{\min}^t}\right) \\
&= \langle \theta, X_t\rangle - \tau^\star\left(\theta^\star, P, \alpha\right)
\end{aligned} \tag{35}$$

i.e. $\langle \theta, X_t\rangle \geq \tau^\star\left(\theta^\star, P, \alpha_t\right) > \tau^\star\left(\theta^\star, P, \alpha\right) > 0$ for all $\theta \in \mathcal{C}_t$ on $G_{p_{\text{err}}}$ and G_θ (as $\alpha_t < \alpha$). □

C.4 (α, δ) safety (proof of Lemma 4)

To prove this lemma, we define the Bernoulli random variable $\xi_t = \mathbb{1}\{\hat{Y}_t \neq Y_t\}$, that denotes whether the algorithm made a mistake at round t, and $\xi_t^\star = \mathbb{1}\{\hat{Y}_t^\star \neq Y_t\}$ respectively for the baseline policy. When the algorithm tests (i.e. $Z_t = 1$) then we observe the label and it holds that $\xi_t = 0$. Conditioning on the good event G, we show that the random variables ξ_t and $\xi_t^\star$ satisfy $\xi_t \leq \xi_t^\star$. This implies a total error probability bound.

Lemma 4. *Under G, SCOUT is (α, δ')-safe.*

Proof. We analyze the four possible outcomes of the binary random variables $(Z_t^\star, Z_t)$, under the good events G_θ and $G_{p_{\text{err}}}$. Recall that ξ_t is whether our algorithm makes a mistake at time t, and $\xi_t^\star$ is whether the optimal baseline which tests at threshold $\tau^\star$ makes an error at time t.

Case 1: $(Z_t^\star, Z_t) = (1, 1)$**.** In this case, both our policy and the oracle baseline observe the true label and $\xi_t = \xi_t^\star = 0$, i.e. neither method makes an error.

Case 2: $(Z_t^\star, Z_t) = (1, 0)$**.** Under the good event G, by Lemma 19 this cannot occur.

Case 3: $(Z_t^\star, Z_t) = (0, 1)$**.** When, $Z_t^\star = 0$ and $Z_t = 1$, our policy tests and observes the true label while the optimal baseline predicts $\hat{Y}_t^\star$, in which case $0 = \xi_t \leq \xi_t^\star$ a.s.

Case 4: $(Z_t^\star, Z_t) = (0, 0)$**.** When, $Z_t^\star = 0$ and $Z_t = 0$, from Lemma 20 it holds that $\hat{Y}_t = \hat{Y}_t^\star$ a.s., and so $\xi_t = \xi_t^\star$ a.s.

Combining these 4 cases together, we have shown that $\xi_t \leq \xi_t^\star$ a.s. Now, $\xi_t^\star$ are independent binary random variables with $\mathbb{E}(\xi_t^\star) \leq \alpha_t$, since the sequence α_t is decreasing. Then at any time $\bar{T} \leq T$:

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{\bar{T}}\sum_{t=1}^{\bar{T}} \xi_t \geq \alpha \,\middle|\, G\right) &\leq \mathbb{P}\left(\frac{1}{\bar{T}}\sum_{t=1}^{\bar{T}} \xi_t^\star \geq \alpha \,\middle|\, G\right) \\
&\leq \mathbb{P}\left(\frac{1}{\bar{T}}\sum_{t=1}^{\bar{T}} (\xi_t^\star - \mathbb{E}\xi_t^\star) \geq \alpha - \alpha_{\bar{T}} \,\middle|\, G\right) \\
&\leq \exp(-2\bar{T}(\alpha - \alpha_{\bar{T}})^2).
\end{aligned}$$

Recall that

$$\alpha_t = \alpha - \sqrt{\frac{\log(2t^2/\delta')}{2t}},$$

Thus:

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{\bar{T}=1}^{T}\left\{\frac{1}{\bar{T}}\sum_{t=1}^{\bar{T}} \xi_t \geq \alpha\right\} \,\middle|\, G\right) &\leq \sum_{\bar{T}=1}^{T} \mathbb{P}\left(\frac{1}{\bar{T}}\sum_{t=1}^{\bar{T}} \xi_t \geq \alpha \,\middle|\, G\right) \\
&\leq \sum_{\bar{T}=1}^{T} \exp(-2\bar{T}(\alpha - \alpha_{\bar{T}})^2) \\
&\leq \sum_{t=1}^{T} \frac{\delta'}{2t^2} \\
&\leq \delta'
\end{aligned} \tag{36}$$

□

D Regret analysis

We begin by bounding the instantaneous regret at time $t > T_0$.

Lemma 22. *For every round $t > T_0$, conditioned on the good event G, the regret is bounded as:*

$$\mathbb{E}[Z_t - Z_t^\star | G] \leq M \cdot V_d(1) \cdot C_2(\tau^\star)\left(\frac{12\left(\zeta_t + 8B_t/\sqrt{p^\star t\lambda_0}\right)}{m \cdot V_d(1) \cdot C_1(\tau^\star)} + 2\varepsilon_Q + 28B_t/\sqrt{p^\star t\lambda_0}\right).$$

Proof of Lemma 22. For $t \leq T_0$ we can bound each term of the regret by 1, i.e. $\mathbb{E}[Z_t - Z] \leq 1$. For $t > T_0$ this requires analyzing $\mathbb{E}[Z_t - Z]$, essentially upper bounding how often we test in excess of the optimal baseline. We test whenever $c_t = |\langle X_t, \theta_t^L\rangle| - \tau_t \leq 0$. Thus, we need to lower bound c_t to show that we do not perform too many excess tests.

$$
\begin{aligned}
c_t &= |\langle X_t, \theta_t^L\rangle| - \tau_t \\
&= |\langle X_t, \theta_t^L\rangle| - \tau^\star\left(\theta_t^L, \hat{P}_t, \alpha_t - \zeta_t - 2B_t/\sqrt{\lambda_{\min}^t} - \varepsilon_Q\right) - 3B_t/\sqrt{\lambda_{\min}^t} - \varepsilon_Q \\
&\overset{(a)}{\geq} |\langle X_t, \theta_t^L\rangle| - \tau_Q^\star\left(\theta_Q, \hat{P}_t, \alpha_t - 2\zeta_t - 4B_t/\sqrt{\lambda_{\min}^t}\right) - 5B_t/\sqrt{\lambda_{\min}^t} - \varepsilon_Q \\
&\overset{(b)}{\geq} |\langle X_t, \theta^\star\rangle| - \tau^\star\left(\theta^\star, P, \alpha - 3\zeta_t - 6B_t/\sqrt{\lambda_{\min}^t}\right) - 7B_t/\sqrt{\lambda_{\min}^t} - 2\varepsilon_Q \\
&\overset{(c)}{\geq} |\langle X_t, \theta^\star\rangle| - \tau^\star\left(\theta^\star, P, \alpha\right) - \frac{3(1+e)\left(\zeta_t + 2B_t/\sqrt{\lambda_{\min}^t}\right)}{m \cdot V_d(1) \cdot C_1(\tau^\star)} - 2\varepsilon_Q - 7B_t/\sqrt{\lambda_{\min}^t}
\end{aligned}
$$

a) comes from Lemmas 10 and 21 to analyze a quantized version of θ_t^L. Concretely, we utilize θ_Q as the projection of θ_t^L onto $\mathcal{C}_t \cap \Theta_Q$. (b) applies Lemma 10 in the reverse direction, to get $\tau^\star$ evaluated at $\theta^\star$. We also use the fact that $\alpha_t \geq \alpha - \zeta_t$. Additionally, $|\langle X_t, \theta_t^L\rangle| \geq |\langle X_t, \theta^\star\rangle| - B_t/\sqrt{\lambda_{\min}^t}$ on $G_{p_{\text{err}}}, G_\theta$. Then, in (c), we apply Lemma 11, where the condition is met for sufficiently large T_0 under G.

$$
\begin{aligned}
\mathbb{E}R_t &= \mathbb{E}[Z_t - Z|G] \\
&= \mathbb{P}\left(\{c_t \leq 0\} \cap \{|\langle X_t, \theta^\star\rangle| \geq \tau^\star\}\,|G\right) \\
&\overset{a}{\leq} \mathbb{P}\left(\tau^\star \leq |\langle X_t, \theta^\star\rangle| \leq \tau^\star + \frac{3(1+e)\left(\zeta_t + 2B_t/\sqrt{\lambda_{\min}^t}\right)}{m \cdot V_d(1) \cdot C_1(\tau^\star)} + 2\varepsilon_Q + 7B_t/\sqrt{\lambda_{\min}^t} \;\middle|\, G\right) \\
&\overset{b}{\leq} M \cdot V_d(1) \cdot C_2(\tau^\star)\left(\frac{12\left(\zeta_t + 2B_t/\sqrt{\lambda_{\min}^t}\right)}{m \cdot V_d(1) \cdot C_1(\tau^\star)} + 2\varepsilon_Q + 7B_t/\sqrt{\lambda_{\min}^t}\right) \\
&\overset{c}{\leq} M \cdot V_d(1) \cdot C_2(\tau^\star)\left(\frac{12\left(\zeta_t + 8B_t/\sqrt{p^\star t\lambda_0}\right)}{m \cdot V_d(1) \cdot C_1(\tau^\star)} + 2\varepsilon_Q + 28B_t/\sqrt{p^\star t\lambda_0}\right)
\end{aligned}
$$

a) follows by the upper bounding of the thresholding condition, and b) follows from Lemma 17, and c) from G that $\lambda_{\min}^t \geq p^\star t\lambda_0/12$.

An important technical detail in applying Lemma 17 is that the upper and lower bounds of our spherical segment are sufficiently close to $\tau^\star$. When we apply this lemma, the perturbation is a constant multiple of $\zeta_t + B_t/\sqrt{p^\star t\lambda_0}$ which are of order $\mathcal{O}(1/\sqrt{t})$ under G. Thus, for sufficiently large constant T_0, for all $t \geq T_0$, we are able to apply Lemma 17.

□

With this instantaneous regret, we are now able to sum across all time steps to compute our total regret. We are then also able to prove the (α, δ) safety of `SCOUT` .

Theorem 1. *`SCOUT` satisfies (α, δ)-safety and has regret bounded by*

$$
T_0 + \tilde{C}\frac{M}{m}\sqrt{\frac{dT\log(T/\delta)}{p^\star\lambda_0}},
$$

for an absolute constant $\tilde{C} > 0$, which is made explicit in the proof (Appendix D).

Proof of Theorem 1. We first show that `SCOUT` satisfies (α, δ) safety. Define A as the event where `SCOUT` is

(α, δ)-safe.

$$\begin{aligned}\mathbb{P}(\bar{A}) &= \mathbb{P}(\bar{A}|G)\mathbb{P}(G) + \mathbb{P}(\bar{A}|\bar{G})\mathbb{P}(\bar{G}) \\ &\leq \mathbb{P}(\bar{A}|G) + \mathbb{P}(\bar{G}) \\ &\leq \delta' + 6\delta' \\ &= \delta\end{aligned}$$

Here we used the law of total probability, and leveraged from Lemma 8 that the good event happens with probability at least $1 - 6\delta'$, and from Lemma 4 that conditioned on G, `SCOUT` is (α, δ')-safe. In the last line we plugged in that $\delta' = \delta/7$.

Analyzing the number of excess tests, we use Lemma 22 and condition on G, to find that with probability at least $1 - \delta$:

$$\begin{aligned}\texttt{Regret}(T) &\leq T_0 + \sum_{t=T_0}^{T} \mathbb{E}R_t \\ &= T_0 + 12\frac{M}{m}\frac{C_2(\tau^\star)}{C_1(\tau^\star)}\sum_{t=T_0}^{T}\left(\zeta_t + 8B_t/\sqrt{p^\star t\lambda_0}\right) \\ &+ 2M \cdot V_d(1) \cdot C_2(\tau^\star)\sum_{t=T_0}^{T}\varepsilon_Q + 28M \cdot V_d(1) \cdot C_2(\tau^\star)\sum_{t=T_0}^{T} B_t/\sqrt{p^\star t\lambda_0}\end{aligned}$$

Both $\varepsilon_Q = 1/t^2$ and the ζ_t (Equation (7)) terms are dominated by the term: $\sum_{t=T_0}^{T} B_t/\sqrt{p^\star t\lambda_0}$. Finally, for B_t (from Equation (6)), we can use from G that we get enough samples, i.e. N_θ^t grows linearly in t.

$$\begin{aligned}B_t &= 2\kappa\left(1 + \sqrt{\log\left(\frac{1}{\delta}\right) + 2d\log\left(1 + \frac{N_\theta^t}{\kappa d}\right)}\right) \\ &\leq 13\sqrt{2d\log\left(N_\theta^t/\delta\right)} \\ \sum_{t=T_0}^{T} B_t(p^\star t\lambda_0/12)^{-1/2} &\leq B_T\sum_{t=T_0}^{T}(p^\star t\lambda_0/12)^{-1/2} \\ &\leq 13\sqrt{2d\log\left(T/\delta\right)}\sum_{t=T_0}^{T}(p^\star t\lambda_0/12)^{-1/2} \\ &\leq 52\sqrt{\frac{dT\log\left(T/\delta\right)}{p^\star\lambda_0}}\end{aligned}$$

Combining this all together we have that:

$$\begin{aligned}\texttt{Regret}(T) &\leq T_0 + \sum_{t=T_0}^{T} \mathbb{E}R_t \\ &= T_0 + 12\frac{M}{m}\frac{C_2(\tau^\star)}{C_1(\tau^\star)}\sum_{t=T_0}^{T}\left(\zeta_t + 8B_T/\sqrt{p^\star t\lambda_0}\right) \\ &+ M \cdot V_d(1) \cdot C_2(\tau^\star)\sum_{t=T_0}^{T}\frac{1}{t^2} + 28M \cdot V_d(1) \cdot C_2(\tau^\star)B_T\sum_{t=T_0}^{T} 1/\sqrt{p^\star t\lambda_0} \\ &\preceq T_0 + 4992\frac{M}{m}\frac{C_2(\tau^\star)}{C_1(\tau^\star)}\sqrt{\frac{dT\log\left(T/\delta\right)}{p^\star\lambda_0}} \qquad (37)\end{aligned}$$

We can further bound the regret by using the lower bound for λ_0 from Lemma 1,

$$\lambda_0 \geq \frac{m(\tau^\star)^3 V_d(1)}{p^\star(d+2)}.$$

Using that,we derive the following asymptotic lower bound

$$\texttt{Regret}(T) = \mathcal{O}\left(d\sqrt{\frac{T\log(T/\delta)}{(\tau^\star)^{d+2}}}\right)$$

We note that our dependence in the number of dimensions is of order $\tilde{\mathcal{O}}(d\sqrt{T})$, same as in linear and logistic bandits (see Lattimore and Szepesvári (2020)). Then, we observe that the edge cases when $\tau^\star = 0$, that is equivalent to $p^\star = 0$ characterize the problem's difficulty. As we have already mentioned in the main text, for $\tau^\star \to 0$ implies that $p^\star = 0$, and we cannot collect enough samples to form our estimators. □

E Good event proof

E.1 Theta estimation set gets enough samples

Lemma 23. *On G_θ and $G_{p_{err}}$, $N_\theta^t \geq p^\star t/2 - \sqrt{\frac{\ln(\pi t^2/(3\delta'))}{2}}$ with probability at least $1-\delta'$.*

Proof of Lemma 23. In Lemma 19 we proved that, with high probability, our policy tests whenever the optimal one does, when G_θ and $G_{p_{\text{err}}}$ hold. This implies that $N_\theta^t \geq N_{OPT}^t$.

As we show, just considering the even time steps, the optimal baseline policy will collect at least $N_{OPT}^t \geq p^\star t/2 - \sqrt{\frac{\ln(\pi t^2/(3\delta'))}{2}}$ samples with high probability up to time t. Using $Z_t^\star$ as whether the optimal thresholding rule would test at time t, we have that, on $G_{p_{\text{err}}}$ and G_θ,

$$N_{OPT}^t \geq \sum_{t=1}^{T//2} Z_{2t}^\star.$$

This implies that:

$$\begin{aligned}\mathbb{P}\left(N_{OPT}^T \leq p^\star\lfloor T/2\rfloor - \nu_T\right) &\leq \mathbb{P}\left(\sum_{t=1}^{T//2}(Z_{2t}^\star - p^\star) \leq -\nu_T\right)\\ &\leq \exp\left(-2\nu_T^2/\lfloor T/2\rfloor\right)\\ &\leq \frac{\delta'\pi^2}{6t^2}\end{aligned}$$

by careful construction of ν_T.

Since δ' is a constant (we simply require that $\delta' = \Omega(T^2 e^{-T})$), then, for some T_0, we have that for all $t \geq T_0$ with probability at least $1-\delta'$;

$$N_\theta^t \geq N_{OPT}^t \geq p^\star t/3. \tag{38}$$

□

To show that $\mathbb{P}(G_\lambda) \geq 1-\delta$ we will use a covering argument to derive a lower bound for the minimum covariance matrix. Then, we will use Lemma 23 as a lower bound on the number of samples collected to construct the empirical covariance matrix. Finally, we will union bound these two events to complete the proof.

E.2 $\lambda_{\min}^t$ grows linearly in t

Lemma 24. *Let $\delta \in (0,1)$. Consider a random $d \times d$ dimensional matrix valued process $\{A_t\}_{t=0}^\infty$ adapted to a filtration $\mathcal{F}_t = \sigma(A_k \mid k \leq t)$, where each $A_t \in \mathbb{R}^{d\times d}$ is symmetric ($A_t = A_t^\top$), positive semi-definite, satisfies $\|A_t\|_{op} \leq 1$ almost surely and such that there is a constant $\lambda_0 > 0$ satisfying*

$$\mathbb{P}\left(\lambda_{\min}(\mathbb{E}[A_t|\mathcal{F}_{t-1}]) \geq \lambda_0 \ \forall t \in \mathbb{N}\right) \geq 1-\tilde{\delta}.$$

Let $\lambda_{\min}^t \triangleq \lambda_{\min}\left(\sum_{s=0}^t A_s\right)$. *Then, for* $\varepsilon > 0$, *the following holds:*

$$\mathbb{P}\left\{\lambda_{\min}^t \geq t(\lambda_0 - 2\varepsilon) - \sqrt{\frac{t}{2}\left(d\log\left(\frac{2}{\varepsilon}+1\right) + \log\left(\frac{4t^2}{\delta'}\right)\right)}\ \forall t \in \mathbb{N}\right\} \geq 1 - \delta'.$$

Proof of Lemma 24. Let the random variable $Z_t^\upsilon \triangleq \upsilon^\top A_t \upsilon - \mathbb{E}[\upsilon^\top A_t \upsilon \mid \mathcal{F}_{t-1}]$, such that $\upsilon \in \mathcal{S}^{d-1}$. Notice that Z_t^υ is a martingale difference sequence as;

1.

$$\begin{aligned}\mathbb{E}[|Z_t^\upsilon|] &\leq \mathbb{E}[|\upsilon^\top A_t \upsilon|] + \mathbb{E}|\mathbb{E}[\upsilon^\top A_t \upsilon \mid \mathcal{F}_{t-1}]| \\ &\leq \mathbb{E}[\upsilon^\top A_t \upsilon] + \mathbb{E}\mathbb{E}[\upsilon^\top A_t \upsilon \mid \mathcal{F}_{t-1}] \\ &\leq 1 + 1 = 2 < \infty.\end{aligned}$$

2.

$$\mathbb{E}[Z_t^\upsilon \mid \mathcal{F}_{t-1}] = \mathbb{E}[\upsilon^\top A_t \upsilon \mid \mathcal{F}_{t-1}] - \mathbb{E}[\upsilon^\top A_t \upsilon \mid \mathcal{F}_{t-1}] = 0.$$

By the Azuma-Hoeffding Inequality (Chung and Lu, 2006), as $Z_t^\upsilon \in [0,1]$ a.s., for a fixed $t \in [T]$ we have, $c \geq 0$;

$$\mathbb{P}\left\{\sum_{s=0}^t (\upsilon^\top A_s \upsilon - \mathbb{E}[\upsilon^\top A_s \upsilon \mid \mathcal{F}_{s-1}]) \leq -c\right\} \leq \exp\left(-\frac{2c^2}{t}\right).$$

Setting the error probability to δ_t,

$$\mathbb{P}\left\{\sum_{s=0}^t (\upsilon^\top A_s \upsilon - \mathbb{E}[\upsilon^\top A_s \upsilon \mid \mathcal{F}_{s-1}]) \leq -\sqrt{\frac{\log(\frac{1}{\delta_t})t}{2}}\right\} \leq \delta_t.$$

Thus, substituting $\delta_t = \frac{\tilde{\delta}}{2t^2}$ and using the union bound we get,

$$\mathbb{P}\left\{\sum_{s=0}^t (\upsilon^\top A_s \upsilon - \mathbb{E}[\upsilon^\top A_s \upsilon \mid \mathcal{F}_{s-1}]) \leq -\sqrt{\frac{\log(\frac{2t^2}{\tilde{\delta}})t}{2}}\ \ \forall t \in \mathbb{N}\right\} \leq \sum_{t=1}^\infty \delta_t \leq \tilde{\delta}.$$

Let $\mathcal{N}(\mathcal{S}^{d-1}, \varepsilon)$ an ε-cover of $\mathcal{S}^{d-1}$. By **Corollary 4.2.13** at Vershynin (2018) we have that the covering numbers of $\mathcal{S}^{d-1}$ satisfy for any $\varepsilon > 0$;

$$\left(\frac{1}{\varepsilon}\right)^d \leq \mathcal{N}(\mathcal{S}^{d-1}, \varepsilon) \leq \left(\frac{2}{\varepsilon}+1\right)^d.$$

For convenience, we define $\nu(t, \tilde{\delta}) \triangleq \sqrt{\frac{[d\log(2/\varepsilon+1)+\log(\frac{2t^2}{\tilde{\delta}})]t}{2}}$. By taking the union bound over all $\upsilon_i \in \mathcal{N}(\mathcal{S}^{d-1}, \varepsilon)$ we have

$$\mathbb{P}\left\{\exists \upsilon_i \in \mathcal{N}(\mathcal{S}^{d-1}, \varepsilon) : \sum_{s=0}^t (\upsilon_i^\top A_s \upsilon_i - \mathbb{E}[\upsilon_i^\top A_s \upsilon_i \mid \mathcal{F}_{s-1}]) \leq -\nu(t, \tilde{\delta})\ \ \forall t \in \mathbb{N}\right\} \leq \tilde{\delta} \tag{39}$$

Let $\upsilon_t^\star \triangleq \operatorname{argmin}_{\upsilon \in \mathcal{S}^{d-1}} \upsilon^\top \sum_{s=0}^t A_s \upsilon$, then there exists an $\upsilon_{i_t} \in \mathcal{N}(\mathcal{S}^{d-1}, \varepsilon)$ such that $\|\upsilon_{i_t} - \upsilon_t^\star\|_2 \leq \varepsilon$. We are

going to bound $|{v_t^\star}^\top \sum_{s=0}^t A_s v_t^\star - v_{i_t}^\top \sum_{s=0}^t A_s v_{i_t}|$ by a function of ε.

$$\begin{aligned}
|{v_t^\star}^\top \sum_{s=0}^t A_s v_t^\star - v_{i_t}^\top \sum_{s=0}^t A_s v_{i_t}| &= |{v_t^\star}^\top \sum_{s=0}^t A_s v_t^\star - {v_t^\star}^\top \sum_{s=0}^t A_s v_{i_t} + {v_t^\star}^\top \sum_{s=0}^t A_s v_{i_t} - v_{i_t}^\top \sum_{s=0}^t A_s v_{i_t}| \\
&= |{v_t^\star}^\top \sum_{s=0}^t A_s (v_t^\star - v_{i_t}) + (v_t^\star - v_{i_t})^\top \sum_{s=0}^t A_s v_{i_t}| \\
&= |(v_t^\star - v_{i_t})^\top \sum_{s=0}^t A_s (v_{i_t} + v_t^\star)| \\
&\leq \|v_t^\star - v_{i_t}\|_2 \left\| \sum_{s=0}^t A_s (v_{i_t} + v_t^\star) \right\|_2 \\
&\leq \varepsilon \sum_{s=0}^t \|A_s\|_{op} \left(\|v_{i_t}\|_2 + \|v_t^\star\|_2\right) \\
&= 2t\varepsilon. \qquad (40)
\end{aligned}$$

Using inequality 39 we have

$$\mathbb{P}\left\{ \sum_{s=0}^t v_{i_t}^\top A_s v_{i_t} \geq \sum_{s=0}^t \mathbb{E}[v_{i_t}^\top A_s v_{i_t} \mid \mathcal{F}_{s-1}] - \nu(t, \tilde{\delta}) \quad \forall t \in \mathbb{N} \right\} \geq 1 - \tilde{\delta}.$$

where i_t is a point in the cover $\mathcal{N}(\mathcal{S}^{d-1}, \varepsilon)$ such that $\|v_{i_t} - v_t^\star\|_2 \leq \varepsilon$. Equation 40 can be used to relate $\sum_{s=0}^t v_{i_t}^\top A_s v_{i_t}$ and $\lambda_{\min}^t$,

$$\mathbb{P}\left\{ \underbrace{\sum_{s=0}^t {v_t^\star}^\top A_s v_t^\star}_{\lambda_{\min}^t} + 2t\varepsilon \geq \sum_{s=0}^t \mathbb{E}[v_{i_t}^\top A_s v_{i_t} \mid \mathcal{F}_{s-1}] - \nu(t, \tilde{\delta}) \quad \forall t \in \mathbb{N} \right\} \geq 1 - \tilde{\delta}.$$

Using the fact that $\mathbb{E}[v_{i_t}^\top A_s v_{i_t} \mid \mathcal{F}_{s-1}] \geq \lambda_{\min}(\mathbb{E}[A_s \mid \mathcal{F}_{s-1}])$ we conclude that,

$$\mathbb{P}\left\{ \lambda_{\min}^t + 2t\varepsilon \geq \sum_{s=0}^t \lambda_{\min}(\mathbb{E}[A_s \mid \mathcal{F}_{s-1}]) - \nu(t, \tilde{\delta}) \quad \forall t \in \mathbb{N} \right\} \geq 1 - \tilde{\delta}.$$

Finally, the assumption that $\mathbb{P}\left(\lambda_{\min}(\mathbb{E}[A_t|\mathcal{F}_{t-1}]) \geq \lambda_0 \ \forall t \in \mathbb{N}\right) \geq 1 - \tilde{\delta}$ and a union bound allows us to conclude that,

$$\begin{aligned}
&\mathbb{P}\left\{ \lambda_{\min}^t \geq t(\lambda_0 - 2\varepsilon) - \nu(t, \tilde{\delta}) \quad \forall t \in \mathbb{N} \right\} \\
&\geq \mathbb{P}\left\{ \lambda_{\min}^t + 2t\varepsilon \geq \sum_{s=0}^t \lambda_{\min}(\mathbb{E}[A_s \mid \mathcal{F}_{s-1}]) - \nu(t, \tilde{\delta}) \cap \lambda_{\min}(\mathbb{E}[A_t|\mathcal{F}_{t-1}]) \geq \lambda_0 \quad \forall t \in \mathbb{N} \right\} \\
&\geq 1 - 2\tilde{\delta}.
\end{aligned}$$

This finalizes the result for $\delta' = 2\tilde{\delta}$.

□

We will apply this lemma for $A_t = X_t X_t^\top$. We use the fact that $\lambda_{\min}\left(\kappa \boldsymbol{I}_d + \sum_{s \in \mathcal{S}_\Theta^t} X_s X_s^\top\right) > \lambda_{\min}\left(\sum_{s \in \mathcal{S}_\Theta^t} X_s X_s^\top\right)$. It is true that $\left\|X_t X_t^\top\right\|_{op} = \|X_t\|_2 \leq 1$. We will make again the same observation, by choosing the covering parameter as $\varepsilon = \frac{\lambda_0}{5}$, then we have that for all $t \geq T_0$

$$\lambda_{\min}^t \geq N_\theta^t \cdot \frac{\lambda_0}{4}. \qquad (41)$$

In Lemma 23 we proved that with probability at least $1-\delta'$, it holds that $N_\theta^t \geq \frac{p^\star t}{3}$. By taking the union bound over the two events, we have that with probability at least $1-2\delta'$

$$\lambda_{\min}^t \geq p^\star t \cdot \frac{\lambda_0}{12}.$$

E.3 Combining all together

Proof of Lemma 3. By using the product rule we have that

$$\mathbb{P}(G_\theta \cap G_N \cap G_{p_{\text{err}}}) = \mathbb{P}(G_N \mid G_\theta \cap G_{p_{\text{err}}})\mathbb{P}(G_\theta \cap G_{p_{\text{err}}})$$

As $\mathbb{P}(G_\theta) \geq 1-\delta$ from Lemma 2 and $\mathbb{P}(G_{p_{\text{err}}}) \geq 1-\delta$ from Lemma 8, by using the union bound we have $\mathbb{P}(G_\theta \cap G_{p_{\text{err}}}) \geq 1-2\delta$. By using also Lemma 23 we have

$$\begin{aligned}\mathbb{P}(G_N \mid G_\theta \cap G_{p_{\text{err}}})\mathbb{P}(G_\theta \cap G_{p_{\text{err}}}) &\geq (1-2\delta')^2 \\ &\geq 1-4\delta'.\end{aligned}$$

As $\mathbb{P}(G_\lambda) \geq 1-2\delta'$ by Lemma 24, by taking the union bound again we have that

$$\mathbb{P}(G_\theta \cap G_{p_{\text{err}}} \cap G_N \cap G_\lambda) \geq 1-6\delta'.$$

□

F Modifications from written algorithm

For our numerical simulations, we implemented a version of `SCOUT` with a few minor modifications from Algorithm 1 to enable it to a) run computationally faster, and b) converge more quickly statistically. These changes are common in practical applications of online learning algorithms to balance theoretical rigor with empirical performance. Full details are available online: `https://github.com/TavorB/SCOUT`.

Batched Parameter Updates: as written, `SCOUT` updates the parameter estimate and the testing threshold at every time step t. In a setting with a large time horizon T, re-running the estimation procedures on ever-growing datasets at each step is computationally wasteful, as these will not change too much iteration to iteration. Instead, our implementation updates these estimates only periodically. Concretely, the estimates for θ and τ are cached and reused for a block of subsequent time steps. The frequency of these updates is decreased as the simulation progresses, reflecting the gradual convergence of the parameters.

Simplified Testing Condition: the testing condition of `SCOUT` is given by $\langle X_t, \theta_t^L\rangle| \leq \tau_t$. This incorporates several uncertainty terms derived from our theoretical analysis. While crucial for the regret bounds, computing these quantities at every step is not necessary in practice, and the same performance can be obtained by simply collapsing these terms into a) the τ estimate, and b) a bound on $B_t\|X_t\|_{V_t^{-1}}$ (note that in practice this second term may not be known, as it will depend on λ_0, which `SCOUT` will learn and adapt to). The testing decision becomes $Z_t = 1$ if $|\langle X_t, \theta_t^L\rangle|$ is less than the sum of these two terms.

Omission of the Projection Step: Our theoretical analysis utilizes two estimators. First, the regularized maximum likelihood estimator $\hat{\theta}_t = \text{argmax}_{\theta\in\mathbb{R}^d} \mathcal{L}_t(\theta)$, where $\mathcal{L}_t(\theta)$ is the regularized log-likelihood. Second, for analysis purposes, a projection of this estimator, θ_t^L, is defined in Equation (5). This projection is in practice unneeded, and so we simply utilize $\hat{\theta}_t$ as our θ estimate.

In addition, we empirically reduce the leading constants for confidence related quantities, for example in the B_t bound. These were modified once up front, retaining the empirical safety guarantees of our algorithm, and dramatically reducing time to convergence.