Expert Advice with Costly Observations

Lev Reyzin[1] **Aadirupa Saha**[2] **Shuo Wu**[3]

[1]Department of Mathematics, Statistics, and Computer Science, University of Illinois Chicago
[2]Department of Computer Science, University of Illinois Chicago
[3]Department of Electrical and Computer Engineering, University of Illinois Chicago

Abstract

Querying expert predictions or human feedback incurs explicit, often heterogeneous costs—a challenge central to crowdsourcing and LLM training. We study online learning with expert advice under observation costs, where the learner must adaptively select which experts to query (p_i is expert-dependent) to minimize cumulative cost (losses plus query costs). While classical bandits and recent paid observation models have advanced the field, they fail to capture the rich cost-adaptive selection landscape and do not provide tight, problem-dependent performance guarantees for heterogeneous costs. This paper provides the first comprehensive characterization. For adversarial losses, we prove tight minimax regret $\Theta\left(\min_{m\in\{1,\dots,K\}}\left\{\sqrt{\frac{K}{m}T\log K}+(m-1)\bar{p}\,T\right\}\right)$, interpolating between bandit and full-information regimes and revealing the optimal observation budget, where $\bar{p}$ is a tight upper bound of p_is. For stochastic losses, our novel m-ELIM algorithm achieves instance-dependent regret of $O\left(\frac{\log T}{m}\sum_{i\neq i^*}\frac{1}{\Delta_i}+(m-1)\bar{p}T\right)$, Δ_i being the suboptimality gap of the i-th arm, showing how gap structure and costs interact. Matching lower bounds (adversarial and stochastic) establishes minimax optimality. These results provide the first rigorous framework for optimally allocating annotation budgets across heterogeneous workers in crowdsourcing systems, directly informing cost-effective strategies for collecting human feedback in adaptive learning applications.

1 INTRODUCTION

In many applications, observing expert predictions or gathering annotator feedback incurs explicit costs. Consider crowdsourcing platforms such as Amazon Mechanical Turk, where requesting a label from a worker requires monetary payment. Similarly, in large language model (LLM)-based training systems, acquiring human feedback through worker judgments for preference learning, reward modeling, or RLHF (Reinforcement Learning with Human Feedback) involves costs per judgment. In these settings, a learner must balance the benefits of querying multiple experts' predictions or workers' outputs against the associated observation costs.

This paper studies the online learning problem with expert advice under costly observations, where the learner must strategically decide which experts to query and at what cost. Formally, we consider K experts. At each round $t = 1, \dots, T$: Each expert $i \in [K]$ produces a prediction and receives a loss $\ell_t(i) \in [0, 1]$ (either stochastic or adversarial). The learner selects a subset $S_t \subseteq [K]$ of experts to observe, incurring observation cost $\sum_{i\in S_t} p_i$, where $p_i \in [0, 1]$ is the fixed cost of observing expert i. After selecting S_t, the learner chooses an expert $I_t \in S_t$ to follow, observing the losses $\{\ell_t(i)\}_{i\in S_t}$. The learner incurs instantaneous cost $\ell_t(I_t) + \sum_{i\in S_t} p_i$.

Performance is measured by the *regret*, which compares the learner's cumulative cost (losses plus observation costs) against the best fixed expert in hindsight, who pays both its loss and observation cost each round. In particular, the learner compares against the best fixed expert $i^* = \arg\min_i(\mu_i + p_i)$ in the stochastic case.

In crowdsourcing systems for data annotation, selecting which workers to query for each task is crucial. Workers vary in both expertise (and hence accuracy/loss) and cost (e.g., expert annotators are more expensive but more reliable). Our framework naturally captures the problem of actively choosing which workers to query for annotations while minimizing total labeling cost, a fundamental problem in

Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence (UAI 2026), PMLR 337:5700–5715.

crowdsourcing systems. When training language models with human feedback, practitioners often collect preference judgments from multiple human raters or use multiple LLM-as-judge models. Each "expert" represents a rater or judge with different accuracy (and hence loss) and cost (e.g., API fees, latency, human worker wages). A learning system must decide how many raters to query per instance and which specific raters to involve, balancing judgment quality against expenses. This is particularly relevant for training reward models in RLHF pipelines.

1.1 RELATED WORK

Expert Advice and Bandits. Prediction with expert advice under full information achieves logarithmic regret via exponential-weighting algorithms. The bandit feedback variant is harder: Auer et al. [2002] introduced EXP3, achieving $\widetilde{O}(\sqrt{KT})$ regret using importance-weighted loss estimates. Our setting interpolates between full-information and bandit regimes through selective expert queries.

Limited Feedback and Paid Observations. Seldin et al. [2014] studied prediction with limited advice (query at most M of N experts) and multiarmed bandits with uniform observation cost c, proving regret bounds interpolating between $\Theta(\sqrt{KT})$ and $\Theta(\log K)$. Our work advances this by: (i) allowing expert-dependent costs p_i (realistic for crowdsourcing with heterogeneous wages and LLM-based systems with varying API costs), (ii) providing the first instance-dependent gap-dependent regret bounds in the stochastic setting, and (iii) characterizing the fundamental cost-quality tradeoff. Building on Seldin et al. [2014], Thune and Seldin [2018] showed that one additional observation per round enables exploiting small loss ranges; while we study the regime with heterogeneous costs.

Costly Observations. Tucker et al. [2023] developed Bandits with Costly Reward Observations with uniform cost c per observation, proving regret $\Omega(c^{1/3}T^{2/3})$. Our setting generalizes this to non-uniform expert-dependent costs with simultaneous subset observation.

Constrained Bandits. Bandits with knapsacks Agrawal and Devanur [2016], Kesselheim and Singla [2020] and resourceful contextual bandits Badanidiyuru et al. [2014] consider budget-constrained exploration. Slivkins [2013] studies budget allocation across arms. These works focus on cumulative regret under budget constraints, while we characterize both worst-case and instance-dependent bounds under per-round observation costs in the expert advice setting.

Crowdsourcing and Active Learning. The framework of active learning from crowds Yan et al. [2011], Whitehill et al. [2009], Raykar et al. [2010] selects informative samples and reliable annotators under labeling budgets. Herde et al. [2024] surveys cost-aware strategies (instance-, annotator-, and query-dependent costs). In contrast, we provide rigorous minimax regret guarantees for heterogeneous costs in the online expert advice setting.

Main Contributions. Compared to the existing works, the contributions of our works is as follows:

- **Novel Problem Formulation & Cost Adaptivity.** We introduce the first comprehensive study of expert advice with heterogeneous, expert-dependent observation costs. Unlike prior work [Seldin et al., 2014, Tucker et al., 2023], our framework captures realistic settings such as crowdsourcing with variable worker wages and LLM systems with heterogeneous API costs, and supports both fixed and adaptive observation set selection (Algorithms 1 and 2).
- **Worst-Case Bounds (Adversarial Losses).** We prove tight minimax regret bounds of $\Theta(\sqrt{(K/m)T} + (m-1)\bar{p}T)$ (Theorem 6, Corollary 7), identifying three regimes: expensive observations with bandit-like regret, an intermediate regime interpolating cost and variance effects, and cheap observations approaching full-information performance. This is the first characterization showing how observation costs fundamentally reshape the classical bandit–full-information tradeoff and determine the optimal observation budget.
- **Instance-Dependent Bounds (Stochastic Losses).** We derive gap-dependent regret bounds of $O((\log T/m)\sum_{i\neq i^*} 1/\Delta_i + (m-1)\bar{p}T)$ via `m-ELIM` (Theorem 4, Corollary 5), explicitly quantifying the interaction between loss gaps and observation costs.
- **Matching Information-Theoretic Lower Bounds.** We establish minimax-optimal lower bounds in both adversarial and stochastic settings (Theorems 6, 8; Corollaries 7, 9), matching our upper bounds up to universal constants and revealing three phase transitions that confirm the necessity of the fundamental cost–quality–variance tradeoff.
- **Unified Framework Bridging Classical Regimes.** Our results seamlessly interpolate between classical MAB (expensive experts), full-information learning (free observations), and cost-efficient learning (cheap experts), providing tight worst-case and instance-dependent guarantees. This framework unifies and extends multiple threads in the online learning literature.

2 PROBLEM SETTING

We consider an online learning problem with a finite set of experts/actions and explicit expert-dependent observation costs. Time proceeds in rounds $t = 1, \ldots, T$. There are K experts. At each round t, expert $i \in [K]$ produces a prediction $y_{i,t}$, and receives a loss $\ell_t(i)$ (either stochastic or adversarially chosen) based on the prediction at round t. At each round t, the learner selects a subset $S_t \subseteq [K]$ of experts to observe. Observing expert i incurs a cost $p_i \in [0,1]$, and the learner pays the total observation cost $\sum_{i\in S_t} p_i$. The costs p_i are fixed and known to the learner.

After selecting S_t, the learner must choose an expert $I_t \in S_t$ to play (hence S_t cannot be $\emptyset$), and the losses $\{\ell_t(i)\}_{i \in S_t}$ are revealed. Let $\ell_t(i) \in [0,1]$ denote the loss incurred at round t if expert i is followed. At round t, the learner incurs the instantaneous cost $\ell_t(I_t) + \sum_{i \in S_t} p_i$.

For clarity, we collect here the standing assumptions of our model, which are assumed throughout the paper.

Assumption 1 (Bounded losses; fixed and known costs)**.** *For all rounds $t \in [T]$ and experts $i \in [K]$: (i) the losses are bounded, $\ell_t(i) \in [0,1]$; (ii) the observation costs $p_i \in [0,1]$ are fixed across rounds and known to the learner. We denote by $\bar{p} := \max_{i \in [K]} p_i$ the (tight) upper bound on the observation costs, which is thus also known to the learner.*

All upper and lower bounds below are stated under Assumption 1; individual theorem statements may repeat parts of these conditions for the reader's convenience. The regret after T rounds is defined as

$$R_T := \sum_{t=1}^{T} \Big(\ell_t(I_t) + \sum_{i \in S_t} p_i \Big) - \min_{i \in [K]} \sum_{t=1}^{T} \big(\ell_t(i) + p_i \big).$$

That is, the learner is compared against the best fixed expert in hindsight, who pays its loss and observation cost once per round. In Section 3, the losses $\{\ell_t(i)\}$ are chosen adversarially to obtain worst-case bounds. In Section 4, we consider a stochastic setting in which, for each expert i, the losses $\ell_t(i)$ are i.i.d. over time with mean $\mu_i := \mathbb{E}[\ell_t(i)]$. Section 5 analyzes the worst-case and instance-dependent lower bound of the problem of cost-adaptive expert advice, respectively, for adversarial and stochastic loss sequences.

3 MINIMAX REGRET

The minimax regret in this model depends on the scale of the observation costs. When observation is expensive, the learner effectively operates in a bandit regime and the minimax regret scales as $\Theta(\sqrt{KT})$. When observation is cheap, the learner may observe multiple experts per round and interpolate toward full-information behavior. The results below characterize this tradeoff.

3.1 WARM-UP: VANILLA EXP3 FOR ADVERSARIAL COST-ADAPTIVE EXPERTS

We begin with a simple baseline that ignores the ability to observe multiple experts per round. The learner treats each expert as a bandit arm with loss equal to the sum of prediction loss and observation cost, and applies standard EXP3 using only the loss of the played expert.

Theorem 1 (Regret Upper Bound with Vanilla EXP3)**.** *Assume $\ell_t(i) \in [0,1]$ and $p_i \in [0,1]$ for all t, i. Run EXP3 on the combined losses*

$$L_t(i) = \ell_t(i) + p_i \in [0,2],$$

using bandit feedback $L_t(I_t)$. Then there exists a learning rate η such that

$$\mathbb{E}[R_T] \le 4\sqrt{TK \log K}.$$

Proof. Apply EXP3 to the rescaled losses $L_t(i)/2 \in [0,1]$ and multiply the standard bound by 2. □

This baseline achieves the standard adversarial bandit regret rate and does not exploit cheap observations. This motivates later algorithms that adaptively observe multiple experts when costs permit.

3.2 ALGORITHM FOR ADVERSARIAL LOSSES

At each round t, it first samples a subset $S_t \subseteq [K]$ of size m uniformly at random, representing the experts whose losses will be observed. An expert I_t is then selected from S_t proportionally to its current weight, and the losses of all experts in S_t are revealed.

Algorithm 1 m-EXP3

Require: experts $[K]$, horizon T, costs $\{p_i\}$, observation budget $m \in \{1, \dots, K\}$

Initialize weights $w_i(1) = 1$ for all $i \in [K]$
for $t = 1$ **to** T **do**
  Define $q_t(i) = \frac{w_i(t)}{\sum_{j=1}^{K} w_j(t)}$ for all i
  Sample $I_t \sim q_t$
  Sample $U_t \subseteq [K] \setminus \{I_t\}$ uniformly at random with $|U_t| = m - 1$
  Set $S_t \leftarrow \{I_t\} \cup U_t$
  Observe $\{\ell_t(i)\}_{i \in S_t}$
  For each $i \in [K]$, let $o_t(i) := \Pr(i \in S_t \mid q_t) = q_t(i) + (1 - q_t(i)) \frac{m-1}{K-1}$.
  Define the unbiased loss estimator $\widehat{\ell}_t(i) := \frac{\ell_t(i) \mathbf{1}\{i \in S_t\}}{o_t(i)}$.
  Define $\widehat{L}_t(i) \leftarrow \widehat{\ell}_t(i) + p_i$ for all i
  Update $w_i(t+1) \leftarrow w_i(t) \exp(-\eta \, \widehat{L}_t(i))$ for all i
end for

To account for partial observability, the algorithm constructs an importance-weighted loss estimate for each expert by scaling the observed losses by a factor of K/m, and setting the estimates to zero for unobserved experts. A per-expert observation cost is incorporated by adding a fixed penalty p_i to the estimated loss. The weights are then updated using the standard EXP3 multiplicative update with respect to these estimated total losses. This procedure balances exploration through random subset selection, exploitation through

weighted sampling within the subset, and the cost of acquiring additional observations. The complete pseudocode is provided in Algorithm 1.

Regret Analysis of Algorithm 1. We now show that by observing up to m experts per round and using importance-weighted feedback, the learner can interpolate smoothly between the bandit and full-information regimes. The resulting regret bound decomposes cleanly into a statistical term that decreases with m and a linear term reflecting the cumulative observation cost.

Theorem 2 (Adversarial regret bound for Algorithm 1)**.** *Assume $\ell_t(i) \in [0,1]$ and $p_i \le \bar{p}$ for all t, i. Fix $m \in \{1, \dots, K\}$ and run Algorithm 1. Then for a suitable learning rate η, the expected regret satisfies*

$$\mathbb{E}[R_T] \le 2\sqrt{\frac{K-1}{m-1}\, T \log K} + (m-1)\bar{p}\, T \qquad (m \ge 2),$$

and for $m = 1$ it recovers the standard $\mathbb{E}[R_T] \le 2\sqrt{KT \log K}$ bound (up to constants).

The proof is given in Appendix Section A.

The linear term $(m-1)\bar{p}\,T$ does not make the regret linear. While $(m-1)\bar{p}\,T$ grows linearly in T for any fixed $m \ge 2$, the observation budget is a design choice: when observations are expensive, Corollary 3 prescribes $m = 1$, for which the term vanishes and the bound reduces to the standard bandit rate, while cheap observations allow larger m, approaching the full-information rate. Under this cost-adaptive choice of m, the regret in Corollary 3 is sublinear in T in all three regimes. The same consideration applies to the instance-dependent bound of Theorem 4 below, whose linear term likewise vanishes for $m = 1$ and whose cost-adaptive choice of m (Corollary 5) again yields sublinear regret in every regime.

Our setting advances the work of Seldin et al. [2014] along two critical dimensions. First, we allow expert-dependent observation costs p_i, rather than assuming a uniform cost, thereby capturing realistic heterogeneity in applications such as crowdsourcing platforms with workers paid at different rates and LLM-based evaluation systems with varying API costs. Second, our regret bounds are tight and explicitly characterize how observation costs interact with variance reduction. In particular, when costs are sufficiently small ($\bar{p} \ll \sqrt{KT}$), observing multiple experts per round can reduce regret below the standard bandit rate $\sqrt{KT}$, approaching full-information performance, whereas expensive observations necessarily induce a bandit-like regime. The resulting cost–variance tradeoff and the optimal observation budget m^* are fully characterized in Corollary 3, providing principled guidance for practitioners. In contrast, Seldin et al. [2014] interpolate between regimes by restricting to single-observation settings (observing at most M of N experts per round) without explicit cost modeling, while our framework enables simultaneous observation of adaptive subsets with explicit cost–quality tradeoffs.

Corollary 3. *Assume $\ell_t(i) \in [0,1]$ and $p_i \le \bar{p}$ for all t, i. Let $a := C\sqrt{(K-1)T \log K}, \quad b := \bar{p}\,T$, where $C > 0$ is the universal constant from Theorem 2. Then there exists an algorithm (namely Algorithm 1 with an appropriate choice of m and η) such that*

$$\mathbb{E}[R_T] \le \begin{cases} a, & b \ge \frac{1}{2}a, \\ \dfrac{3}{2^{2/3}}\, a^{2/3} b^{1/3} + b, & \dfrac{a}{2(K-1)^{3/2}} \le b \le \frac{1}{2}a, \\ \dfrac{a}{\sqrt{K-1}} + b(K-1), & b \le \dfrac{a}{2(K-1)^{3/2}}. \end{cases}$$

In particular, in the intermediate regime $\frac{a}{2(K-1)^{3/2}} \le b \le \frac{1}{2}a$, choosing $m = 1 + \left\lceil \left(\frac{a}{2b}\right)^{2/3} \right\rceil$ yields

$$\mathbb{E}[R_T] \le \frac{3}{2^{2/3}}\, a^{2/3} b^{1/3} + b = O\Big((K \log K)^{1/3}\, \bar{p}^{1/3}\, T^{2/3} \Big).$$

Proof is given in Appendix A.

4 PROBLEM-DEPENDENT REGRET

In this section, we specialize to a stochastic setting (i.i.d. losses over time), and derive an instance-dependent regret bound that exhibits an explicit tradeoff between observing many experts per round and paying observation costs.

Some Useful Notations. Before diving into the technical details, let us define the following quantities for ease of exposition: For each expert $i \in [K]$, we define: (i) $\theta_i := \mu_i + p_i$, (ii) $\theta_\star := \min_{i \in [K]} \theta_i$, (iii) $i^\star \in \arg\min_{i \in [K]} \theta_i$, (iv) $\Delta_i := \theta_i - \theta_\star \quad (i \ne i^\star)$. (v) Let us also denote by $\bar{p} := \max_{i \in [K]} p_i$.

4.1 A SUCCESSIVE-ELIMINATION ALGORITHM

We start with a simple elimination-style algorithm for the problem (Algorithm 2), which queries a fixed set of m experts every round and eliminates 'weaker-experts' through a UCB-based elimination rule.

For each expert i, O_i is number of times i has been observed up to (and including) the current round t, and let $\widehat{\mu}_i$ be its empirical mean based on those O_i samples. The `m-ELIM` maintains an active set A (initially $A = [K]$). Proceed in stages $r = 1, 2, \dots$ with target accuracy $\varepsilon_r := 2^{-r}$ and $n_r := \left\lceil \frac{2L}{\varepsilon_r^2} \right\rceil$. In stage r, it partitions A into groups $G_1, \dots, G_g$ of size at most m. For each group G, it runs n_r rounds in which it observes exactly the experts in G (so $S_t = G$) and plays experts in G in round-robin order. After all groups are processed, each $i \in A$ has received n_r fresh observations in this stage and at most $\lceil n_r/m \rceil$ plays.

Algorithm 2 m-ELIM

Require: experts $[K]$, time T, costs $\{p_i\}$, observations m

$A \leftarrow [K]$; $O_i \leftarrow 0$, $\widehat{\mu}_i \leftarrow 0$ for all i
$L \leftarrow 2\log(2KT)$; $t \leftarrow 1$
for $r = 1, 2, \ldots$ **while** $|A| > 1$ **and** $t \le T$ **do**
  $\varepsilon_r \leftarrow 2^{-r}$; $n_r \leftarrow \lceil 2L/\varepsilon_r^2 \rceil$
  Partition A into groups of size at most m
  for all groups G **do**
    for $s = 1$ **to** n_r **do**
      $S_t \leftarrow G$; choose $I_t \in G$ round-robin
      Observe $\{\ell_t(i)\}_{i \in S_t}$. Update $O_i \leftarrow O_i + 1$ and $\widehat{\mu}_i \leftarrow \frac{(O_i - 1)\widehat{\mu}_i + \ell_t(i)}{O_i}$ for all $i \in S_t$
      $t \leftarrow t + 1$; **if** $t > T$ **break**
    end for
  end for
  $\widehat{\theta}_i \leftarrow \widehat{\mu}_i + p_i$; $\mathrm{rad}_i \leftarrow \sqrt{L/(2O_i)}$, for all $i \in A$
  $b \in \arg\min_{i \in A}(\widehat{\theta}_i + \mathrm{rad}_i)$, $A \leftarrow \{i \in A : \widehat{\theta}_i - \mathrm{rad}_i \le \widehat{\theta}_b + \mathrm{rad}_b\}$
end for
Play the remaining expert for all remaining rounds

At the end of stage r, for each $i \in A$, the algorithm computes the confidence radius $\mathrm{rad}_i := \sqrt{\frac{L}{2O_i}}$ and $\widehat{\theta}_i := \widehat{\mu}_i + p_i$, where recall that $\widehat{\mu}_i$ is the empirical estimate of μ_i.

Let $b \in \arg\min_{j \in A} \widehat{\theta}_j + \mathrm{rad}_j$ and eliminate all $i \in A$ such that $\widehat{\theta}_i - \mathrm{rad}_i > \widehat{\theta}_b + \mathrm{rad}_b$. If $|A| = 1$, observe only the remaining expert for all remaining rounds.

Theorem 4 (Regret Analysis of Algorithm 2)**.** *Assume the stochastic i.i.d. model above. For any $m \in \{1, \ldots, K\}$, the algorithm m-ELIM satisfies*

$$\mathbb{E}[R_T] \le C \frac{\log(KT)}{m} \sum_{i \ne i^\star} \frac{1}{\Delta_i} + (m-1)\bar{p}T + 2,$$

for a universal constant $C > 0$.

Proof of Theorem 4. Recall $\theta_i := \mu_i + p_i$, $\theta_\star := \min_i \theta_i$, fix $i^\star \in \arg\min_i \theta_i$, and for $i \ne i^\star$ let $\Delta_i := \theta_i - \theta_\star > 0$.

We first decompose the regret:

$$\begin{aligned} R_T &= \sum_{t=1}^{T} \Big(\ell_t(I_t) + \sum_{j \in S_t} p_j \Big) - \sum_{t=1}^{T} (\ell_t(i^\star) + p_{i^\star}) \\ &= \sum_{t=1}^{T} \big(\ell_t(I_t) + p_{I_t} - (\ell_t(i^\star) + p_{i^\star}) \big) + \sum_{t=1}^{T} \sum_{j \in S_t \setminus \{I_t\}} p_j. \end{aligned}$$

Taking expectations and using $\mathbb{E}[\ell_t(I_t) \mid I_t] = \mu_{I_t}$ gives

$$\mathbb{E}[R_T] = \mathbb{E}\Big[\sum_{t=1}^{T} (\theta_{I_t} - \theta_\star)\Big] + \mathbb{E}\Big[\sum_{t=1}^{T} \sum_{j \in S_t \setminus \{I_t\}} p_j\Big].$$

As $|S_t| \le m$ always and $p_j \le \bar{p}$, we have deterministically

$$\sum_{t=1}^{T} \sum_{j \in S_t \setminus \{I_t\}} p_j \le (m-1)\bar{p}\,T,$$

so it remains to bound $\mathbb{E}[\sum_{t=1}^{T} (\theta_{I_t} - \theta_\star)]$.

Let $O_i(t)$ be the number of observations of expert i up to time t, and $\widehat{\mu}_i(t)$ be its empirical mean. Fix $L := 2\log(2KT)$ and

$$\mathcal{E} := \Big\{ \forall i \in [K],\ \forall n \in [T] :\ |\widehat{\mu}_{i,n} - \mu_i| \le \sqrt{L/(2n)} \Big\}.$$

By Hoeffding's inequality and a union bound over i and $n \le T$, $\mathbb{P}(\mathcal{E}) \ge 1 - \frac{1}{KT}$.

We condition on $\mathcal{E}$. At any elimination step the algorithm forms, for each active i,

$$\widehat{\theta}_i := \widehat{\mu}_i + p_i, \qquad \mathrm{rad}_i := \sqrt{L/(2O_i)}.$$

On $\mathcal{E}$, for all active i we have $\theta_i \in [\widehat{\theta}_i - \mathrm{rad}_i,\ \widehat{\theta}_i + \mathrm{rad}_i]$. Let $b \in \arg\min_{j \in A}(\widehat{\theta}_j + \mathrm{rad}_j)$ and update

$$A \leftarrow \{i \in A :\ \widehat{\theta}_i - \mathrm{rad}_i \le \widehat{\theta}_b + \mathrm{rad}_b\}.$$

Then $i^\star$ is never eliminated on $\mathcal{E}$, because

$$\widehat{\theta}_{i^\star} - \mathrm{rad}_{i^\star} \ \le\ \theta_\star \ \le\ \theta_b \ \le\ \widehat{\theta}_b + \mathrm{rad}_b,$$

where the first inequality holds on $\mathcal{E}$ (as $\theta_{i^\star} = \theta_\star$ lies in the confidence interval of $i^\star$), the second holds since $\theta_\star = \min_{i \in [K]} \theta_i$, and the third holds on $\mathcal{E}$ (as $b \in A$). Hence the elimination condition $\widehat{\theta}_{i^\star} - \mathrm{rad}_{i^\star} > \widehat{\theta}_b + \mathrm{rad}_b$ is never triggered for $i^\star$.

Now fix any $i \ne i^\star$. Suppose at some elimination step $\mathrm{rad}_i \le \Delta_i/8$ and $\mathrm{rad}_{i^\star} \le \Delta_i/8$. On $\mathcal{E}$,

$$\widehat{\theta}_i - \mathrm{rad}_i \ge \theta_i - 2\mathrm{rad}_i \ge \theta_\star + \Delta_i - \frac{\Delta_i}{4} = \theta_\star + \frac{3\Delta_i}{4},$$

$$\text{and } \widehat{\theta}_{i^\star} + \mathrm{rad}_{i^\star} \le \theta_\star + 2\mathrm{rad}_{i^\star} \le \theta_\star + \frac{\Delta_i}{4}.$$

Since b minimizes $(\widehat{\theta}_j + \mathrm{rad}_j)$ over $j \in A$ and $i^\star \in A$,

$$\widehat{\theta}_b + \mathrm{rad}_b \le \widehat{\theta}_{i^\star} + \mathrm{rad}_{i^\star} \le \theta_\star + \frac{\Delta_i}{4},$$

hence $\widehat{\theta}_i - \mathrm{rad}_i > \widehat{\theta}_b + \mathrm{rad}_b$ and i is eliminated.

Let $r_i := \min\{r :\ \varepsilon_r \le \Delta_i/4\}$, where $\varepsilon_r = 2^{-r}$ is the stage accuracy parameter and $n_r = \lceil 2L/\varepsilon_r^2 \rceil$ is the number of rounds per group in stage r. In each stage, an active expert is observed exactly n_r times (it belongs to exactly one processed group and is observed in all n_r rounds for that group). Therefore, after completing stage r_i we have $O_i \ge n_{r_i}$ and $O_{i^\star} \ge n_{r_i}$, so

$$\mathrm{rad}_i \le \sqrt{\frac{L}{2n_{r_i}}} \le \sqrt{\frac{L}{2 \cdot (2L/\varepsilon_{r_i}^2)}} = \frac{\varepsilon_{r_i}}{2} \le \frac{\Delta_i}{8},$$

and likewise $\mathrm{rad}_{i^\star} \le \Delta_i/8$. Hence i is eliminated at (or immediately after) the elimination step following stage r_i. Consequently, on $\mathcal{E}$, expert i is played only during stages $1, 2, \ldots, r_i$.

Within any group G of size at most m, the algorithm plays round-robin for n_r rounds, so each expert in G is played at most $\lceil n_r/|G| \rceil \le \lceil n_r/m \rceil$ times in that group-processing block. Thus, letting N_i be the total number of plays of expert i over the horizon,

$$N_i \le \sum_{r=1}^{r_i} \left\lceil \frac{n_r}{m} \right\rceil \le \frac{1}{m} \sum_{r=1}^{r_i} n_r + r_i.$$

Using $n_r \le 2L/\varepsilon_r^2 + 1$ and $\varepsilon_r = 2^{-r}$,

$$\sum_{r=1}^{r_i} n_r \le \sum_{r=1}^{r_i} \left(\frac{2L}{\varepsilon_r^2} + 1 \right) \le 2L \sum_{r=1}^{r_i} 4^r + r_i \le C_1 L \cdot 4^{r_i} + r_i$$

for a universal constant C_1. Since $\varepsilon_{r_i} \le \Delta_i/4$, we have $4^{r_i} = \varepsilon_{r_i}^{-2} \le 16/\Delta_i^2$, hence

$$N_i \le \frac{C_2 L}{m \Delta_i^2} + C_3 \log\Big(\frac{1}{\Delta_i}\Big)$$

for universal constants C_2, C_3 (using $r_i = O(\log(1/\Delta_i))$).

On any round s.t. $I_t = i$, the expected excess (loss+played-cost) is $\theta_i - \theta_\star = \Delta_i$, so conditioning on $\mathcal{E}$,

$$\mathbb{E}\Big[\sum_{t=1}^{T} (\theta_{I_t} - \theta_\star) \,\Big|\, \mathcal{E}\Big] = \sum_{i \ne i^\star} \Delta_i \, \mathbb{E}[N_i \mid \mathcal{E}] \le \sum_{i \ne i^\star} \left(\frac{C_2 L}{m \Delta_i} + C_3 \Delta_i \log\Big(\frac{1}{\Delta_i}\Big) \right).$$

Since $\Delta \log(1/\Delta) \le 1$ for $\Delta \in (0, 1]$, the second term contributes at most $C_3(K-1)$, which contributes at most $O(K)$ and can be absorbed into the final additive constant term for sufficiently large T. Thus

$$\mathbb{E}\Big[\sum_{t=1}^{T} (\theta_{I_t} - \theta_\star) \,\Big|\, \mathcal{E}\Big] \le \frac{C_4 L}{m} \sum_{i \ne i^\star} \frac{1}{\Delta_i}$$

for a universal constant C_4.

On $\mathcal{E}^c$, we use the crude bound $0 \le \theta_{I_t} - \theta_\star \le 2$, hence

$$\mathbb{E}\Big[\sum_{t=1}^{T} (\theta_{I_t} - \theta_\star) \mathbf{1}_{\mathcal{E}^c}\Big] \le 2T \mathbb{P}(\mathcal{E}^c) \le 2T \cdot \frac{1}{KT} \le 2.$$

Combining everything and recalling $L = 2\log(2KT)$,

$$\mathbb{E}[R_T] \le \frac{C_4 \cdot 2\log(2KT)}{m} \sum_{i \ne i^\star} \frac{1}{\Delta_i} \;+\; (m-1)\bar{p}T \;+\; 2.$$

Renaming constants yields the stated bound. □

Corollary 5 (Implications of Theorem 4)**.** *Under the assumptions of Theorem 4, let* $H := \sum_{i \ne i^\star} \frac{1}{\Delta_i}$, $a := C \log(KT)\, H$, *and* $b := \bar{p}T$. *Then the choice* $m = \min\{K, \max\{1, \lceil \sqrt{a/b} \rceil\}\}$ *yields*

$$\mathbb{E}[R_T] \;\le\; \begin{cases} a + 2, & b \ge a, \\ 2\sqrt{ab} + 2, & \frac{a}{K^2} \le b \le a, \\ \frac{a}{K} + (K-1)b + 2, & b \le \frac{a}{K^2}. \end{cases}$$

In particular, if $\sqrt{a/b} \le 1$ *(i.e.* $a \le b$*), then* $m = 1$, *and* $\mathbb{E}[R_T] \le a + 2 = O\left(\log(KT) \sum_{i \ne i^\star} \frac{1}{\Delta_i}\right)$.

Proof is given in Section B.

Corollary 5 translates the instance-dependent regret bound of Theorem 4 into concrete guidance for selecting the observation budget m. The key insight is that the optimal choice of m depends on the intrinsic difficulty of the problem, represented by the term $H := \sum_{i \ne i^\star} \frac{1}{\Delta_i}$, and the cost scale $\bar{p}T$. Let $a = C \log(KT)\, H$ and $b = \bar{p}T$. When observation costs dominate problem difficulty ($b \ge a$), the optimal strategy is to observe a single expert per round ($m = 1$), recovering the classical MAB regret $O(H \log T)$. In the intermediate regime ($a/K^2 \le b \le a$), the optimal choice $m \approx \sqrt{a/b}$ interpolates between $m = 1$ and $m = K$, yielding regret $O(\sqrt{\bar{p}T \cdot H \log T})$. Finally, when observations are cheap ($b \le a/K^2$), observing all experts ($m = K$) is optimal, achieving accelerated regret $O(H \log T/K + (K-1)\bar{p}T)$ due to parallelization. These regimes illustrate a fundamental tradeoff: cheap observations enable aggressive parallel exploration, whereas expensive observations necessitate conservative sequential selection.

5 LOWER BOUND ANALYSIS

We now complement our upper bounds with matching lower bounds, showing that the dependence on the number of observations and observation costs is unavoidable. We present both worst-case adversarial lower bounds and instance-dependent lower bounds for stochastic losses.

5.1 A LOWER BOUND ON MINIMAX REGRET

We now establish the minimax optimality of the regret bounds derived in Section 3 by proving matching lower bounds in the adversarial setting. In particular, we show that for any algorithm, there exists an adversarial loss sequence under which the expected regret is at least $\Omega\Big(\sqrt{\frac{K}{\hat{m}}T} + (\hat{m}-1)\bar{p}\,T\Big)$, where $\hat{m}$ denotes the effective number of experts observed per round. This lower bound matches the functional form of our upper bound in Theorem 2, demonstrating that the cost-dependent interpolation

between bandit and full-information regimes induced by heterogeneous observation costs is minimax-optimal and cannot be improved in general.

Theorem 6 (Worst Case Lower Bound with Adversarial Losses)**.** *Assume $K \geq 2$ and $T \geq K$. Fix a constant $p \in (0, 1]$ and set $p_i \equiv p$ for all i. Let the optimal number of observations*

$$m^\star := \left(\frac{1}{2p}\sqrt{\frac{K}{T}}\right)^{2/3}, \ \widehat{m} := \min\{K, \max\{1, \lfloor m^\star \rceil\}\},$$

where $\lfloor x \rceil$ denotes rounding x to the nearest integer (ties broken arbitrarily). Then there exists an oblivious loss sequence distribution with $\ell_t(i) \in \{0, 1\}$ such that every (possibly randomized) algorithm satisfies

$$\mathbb{E}[R_T] \ \geq \ c\left(\sqrt{\frac{K}{\widehat{m}}T} + pT(\widehat{m} - 1)\right), \tag{1}$$

for a universal constant $c > 0$.

Proof of Theorem 6. Choose J uniformly in $[K]$ and fix $\varepsilon \in (0, 1/4]$. Generate losses independently with

$$\ell_t(J) \sim \text{Bernoulli}\Big(\tfrac{1}{2} - \varepsilon\Big), \ \ell_t(i) \sim \text{Bernoulli}\Big(\tfrac{1}{2}\Big) \ (i \neq J).$$

Let $N_i = \sum_{t=1}^T \mathbf{1}\{I_t = i\}$ and let $M = \sum_{t=1}^T (|S_t| - 1)$. Since the comparator pays observation cost p each round,

$$R_T = \Big(\sum_{t=1}^T \ell_t(I_t) - \min_i \sum_{t=1}^T \ell_t(i)\Big) + pM.$$

Also $\min_i \sum_t \ell_t(i) \leq \sum_t \ell_t(J)$, hence for each $j \in [K]$,

$$\mathbb{E}_j[R_T] \geq \varepsilon \mathbb{E}_j[T - N_j] + p\mathbb{E}_j[M].$$

Averaging over J and letting $\mathbb{P}_j$ denote the induced distribution under instance j and $\mathbb{P}_0$ the null instance where all arms are Bernoulli($\frac{1}{2}$), Pinsker's inequality (see Lemma 10 in the Appendix) implies

$$\mathbb{E}_j[T - N_j] \geq T - T\sqrt{\tfrac{1}{2} D_{\mathrm{KL}}(\mathbb{P}_j \| \mathbb{P}_0)}.$$

By the KL chain rule (see Lemma 11 in the Appendix), only rounds in which expert j is observed contribute, and

$$D_{\mathrm{KL}}(\mathbb{P}_j \| \mathbb{P}_0) \leq 8\varepsilon^2 \mathbb{E}_j\left[\sum_{t=1}^T \mathbf{1}\{j \in S_t\}\right].$$

Averaging over j and using $\sum_{j=1}^K \sum_{t=1}^T \mathbf{1}\{j \in S_t\} = T + M$ yields

$$\frac{1}{K}\sum_{j=1}^K \mathbb{E}_j[T - N_j] \geq c\sqrt{\frac{K}{m_{\text{eff}}}T},$$

for a universal constant $c > 0$, where $m_{\text{eff}} := 1 + \frac{\mathbb{E}[M]}{T} \in [1, K]$. Now consider the function

$$f(m) := \sqrt{\frac{K}{m}T} + pT(m - 1), \qquad m \in [1, K].$$

A direct calculation shows that the unique minimizer of f over $m \geq 1$ is $m^\star = \left(\frac{1}{2p}\sqrt{\frac{K}{T}}\right)^{2/3}$. Let $m_{\text{clip}} := \min\{K, \max\{1, m^\star\}\}$ and let $\widehat{m} := \lfloor m_{\text{clip}} \rceil$. Since f is convex on $(0, \infty)$ (being the sum of $m \mapsto cm^{-1/2}$ and a linear term), rounding and clipping affect the value of f by at most a universal constant factor; in particular,

$$f(m_{\text{eff}}) \ \geq \ \min_{m \in [1,K]} f(m) \ = \ f(m_{\text{clip}}) \ \geq \ c_2 f(\widehat{m})$$

for a universal constant $c_2 \in (0, 1]$. Combining the previous display with the information-theoretic lower bound above gives

$$\frac{1}{K}\sum_{j=1}^K \mathbb{E}_j[R_T] \geq c_1 f(m_{\text{eff}}) \geq c_1 c_2 \left(\sqrt{\frac{K}{\widehat{m}}T} + pT(\widehat{m} - 1)\right).$$

Therefore, there exists an index j such that $\mathbb{E}_j[R_T]$ is at least the right-hand side, and since the adversary is oblivious, this implies the stated minimax lower bound with $c := c_1 c_2$. □

Corollary 7 remarks the implications of the lower bound of Theorem 6 and the tradeoff between p vs other problem-dependent parameters.

Corollary 7. *Under the assumptions of Theorem 6, there exists an oblivious loss sequence distribution with $\ell_t(i) \in \{0, 1\}$ s.t. every (possibly randomized) algorithm satisfies*

$$\mathbb{E}[R_T] \ \geq \ c \cdot \begin{cases} \sqrt{KT}, & p \geq \frac{1}{2}\sqrt{\frac{K}{T}}, \\ (Kp)^{1/3}T^{2/3}, & \frac{1}{2K\sqrt{T}} \leq p \leq \frac{1}{2}\sqrt{\frac{K}{T}}, \\ \sqrt{T} + pT(K-1), & p \leq \frac{1}{2K\sqrt{T}}, \end{cases}$$

for a universal constant $c > 0$.

Proof is given in Appendix C.

5.2 AN INSTANCE-DEPENDENT LOWER BOUND

For the stochastic i.i.d. setting, we complement the instance-dependent upper bound of Theorem 4 (and Corollary 5) with matching lower bounds establishing minimax optimality. In particular, we show that any algorithm that observes exactly m experts per round must incur regret at least $\Omega\Big(\frac{\log T}{m}\sum_{i \neq i^*} \frac{1}{\Delta_i} + (m-1)\bar{p}\,T\Big)$ on appropriately constructed Bernoulli instances with heterogeneous gaps $\{\Delta_i\}$.

Theorem 8 (Instance-Dependent Lower Bound with Stochastic Losses)**.** *Fix $K \geq 2$, $T \geq 3$, and uniform costs $p_i = p \in (0, 1]$. Fix $m \in \{1, \dots, K\}$ and consider any (possibly randomized) algorithm that observes exactly m experts per round, i.e. $|S_t| = m$ almost surely for all t. Then there exists a stochastic i.i.d. Bernoulli instance s.t., with*

$$\theta_i := \mu_i + p, \quad i^\star \in \arg\min_i \theta_i, \quad \Delta_i := \theta_i - \theta_{i^\star} \ (i \neq i^\star),$$

the expected regret satisfies

$$\mathbb{E}[R_T] \ = \Omega\left(\frac{\log T}{m} \sum_{i \neq i^\star} \frac{1}{\Delta_i} \ + \ pT(m-1)\right).$$

Proof of Theorem 8. Since $p_i \equiv p$ and $|S_t| = m$ a.s.,

$$\begin{aligned} R_T &= \sum_{t=1}^{T} \ell_t(I_t) - \min_i \sum_{t=1}^{T} \ell_t(i) \ + \ p \sum_{t=1}^{T} (|S_t| - 1) \\ &= \sum_{t=1}^{T} \ell_t(I_t) - \min_i \sum_{t=1}^{T} \ell_t(i) \ + \ pT(m-1). \end{aligned}$$

It remains to lower bound the loss-only term.

Fix gaps $\{\Delta_i\}_{i=2}^K \subset (0, 1/8]$ and set $\Delta_1 = 0$. Let $\Delta_{\max} := \max_{i \geq 2} \Delta_i$ and define means

$$\mu_1 := \tfrac{1}{2} - 2\Delta_{\max}, \qquad \mu_i := \mu_1 + \Delta_i \ (i \geq 2),$$

so $\mu_i \in [1/4, 3/4]$ and $i^\star = 1$. For each $i \geq 2$, define an alternative instance that only changes arm i:

$$\mu_i^{(i)} := \mu_1 - \Delta_i, \qquad \mu_j^{(i)} := \mu_j \ (j \neq i),$$

under which i is the unique minimizer of $\theta_j = \mu_j + p$.

Let $\mathbb{P}$ and $\mathbb{P}^{(i)}$ be the induced interaction distributions under the base and i-th alternative instances, and let $O_i := \sum_{t=1}^{T} \mathbf{1}\{i \in S_t\}$. By the KL chain rule (Lemma 11), only rounds where i is observed contribute, hence

$$D_{\mathrm{KL}}(\mathbb{P} \| \mathbb{P}^{(i)}) \ \leq \ \mathbb{E}[O_i] \ D_{\mathrm{kl}}(\mu_i, \mu_i^{(i)}).$$

Since $\mu_i, \mu_i^{(i)} \in [1/4, 3/4]$ and $|\mu_i - \mu_i^{(i)}| = 2\Delta_i$, we have $D_{\mathrm{kl}}(\mu_i, \mu_i^{(i)}) \leq C_0 \Delta_i^2$ for a universal C_0. A standard two-point testing argument using Pinsker's inequality (Lemma 10) gives $D_{\mathrm{KL}}(\mathbb{P} \| \mathbb{P}^{(i)}) \geq c_1 \log T$, and therefore

$$\mathbb{E}[O_i] \ \geq \ \frac{c_2 \log T}{\Delta_i^2} \tag{2}$$

for a universal $c_2 > 0$.

Finally, $\sum_{i=1}^K O_i = \sum_{t=1}^T |S_t| = mT$. Combining this identity with (2) and the usual Lai–Robbins counting step (which turns the $\Omega(\log T / \Delta_i^2)$ observation requirement into $\Omega(\log T / \Delta_i)$ loss regret, here sped up by a factor m because m losses are revealed per round) yields

$$\mathbb{E}\left[\sum_{t=1}^{T} \ell_t(I_t) - \min_i \sum_{t=1}^{T} \ell_t(i)\right] \ \geq \ c_3 \frac{\log T}{m} \sum_{i \neq i^\star} \frac{1}{\Delta_i},$$

for a universal $c_3 > 0$. Adding the deterministic term $pT(m-1)$ completes the proof. □

Corollary 9 remarks the implications of the lower bound of Theorem 8 through the tradeoff between p vs instance-dependent parameters Δ_i and the optimal size of the observation set.

Corollary 9. *Under the assumptions of Theorem 8, define the instance-dependent difficulty*

$$S := \sum_{i \neq i^\star} \frac{1}{\Delta_i}, m^\star := \sqrt{\frac{(\log T)\, S}{pT}},$$
$$\widehat{m} := \min\{K, \max\{1, \lfloor m^\star \rceil\}\}.$$

Then for every (possibly randomized) algorithm that observes exactly m experts per round (i.e. $|S_t| = m$ a.s.), there exists a stochastic i.i.d. Bernoulli instance for which

$$\mathbb{E}[R_T] \ \geq \ c\left(\frac{(\log T)\, S}{m} + pT(m-1)\right)$$

for a universal constant $c > 0$. In particular, taking $m = \widehat{m}$ yields the following parameter regimes:

$$a := \frac{(\log T)\, S}{T}, \qquad b := \frac{(\log T)\, S}{TK^2}.$$

$$\mathbb{E}[R_T] \ \geq \ c' \cdot \begin{cases} (\log T)\, S & \text{if } p \geq a \ (m^\star \leq 1), \\ \sqrt{pT\,(\log T)\, S} & \text{if } b \leq p \leq a \ (1 \leq m^\star \leq K), \\ \frac{(\log T)\, S}{K} + pT(K-1) & \text{if } p \leq b \ (m^\star \geq K). \end{cases}$$

for a universal constant $c' > 0$. Moreover, the corresponding minimizing choices are $m = 1$ in the first regime, $m \approx m^\star$ in the middle regime, and $m = K$ in the third regime.

Proof is given in Appendix C.

6 EMPIRICAL VALIDATION

We compare `m-EXP3` and `m-ELIM` against a UCB baseline on three synthetic instances ($T = 10{,}000$, $R = 40$ independent runs), with the observation budget m set per Corollaries 3 and 5. UCB serves as a natural baseline that ignores the parallel observation structure and is included to quantify the cost of not exploiting simultaneous queries.

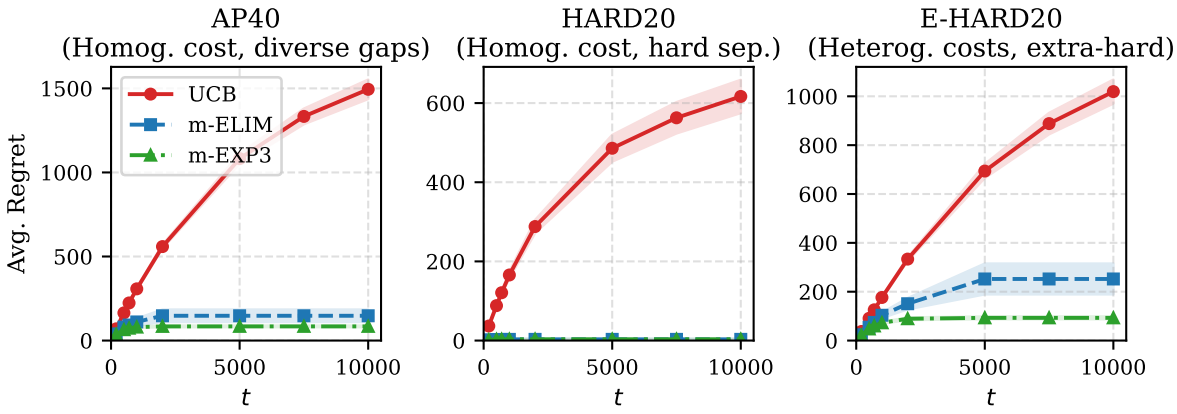

Figure 1: Average cumulative regret vs. t ($R = 40$ runs; shaded bands = ±std). `m-EXP3` and `m-ELIM` quickly identify i^* and plateau in all three instances, while UCB's regret continues to grow with T. `m-EXP3` slightly outperforms `m-ELIM`, consistent with the advantage of gradient-based over elimination-based methods.

Instances. **AP40** ($K = 40$, $p_i = 0.001$, $\mu \in \{0.7^{\times 11}, 0.5^{\times 15}, 0.3^{\times 13}, 0.1^{\times 1}\}$) has very cheap per-arm costs and a unique best arm i^* with $\mu^* = 0.1$, separated from the rest by gaps $\Delta_i \in \{0.1, 0.2, 0.4, 0.6\}$. **HARD20** ($K = 20$, $p_i = 0.05$, $\mu \in \{0.2^{\times 19}, 0.1^{\times 1}\}$) is designed to be hard: all suboptimal arms share the same loss 0.2, creating a uniform small gap $\Delta = 0.1$ that makes identification difficult. **E-HARD20** ($K = 20$, $p_1 = 0.2$, $p_i = 0.1$ for $i \in [2, 10]$, $p_i = 0.2$ for $i \in [11, 20]$, $\mu_1 = 0.05$, $\mu_i = 0.25$ for $i \neq 1$) stresses cost heterogeneity: the optimal arm $i^* = 1$ has a higher observation cost than half the suboptimal arms, testing whether our algorithms correctly account for the cost-quality tradeoff. Full numerical results appear in Table 1 (Appendix E).

Results. Figure 1 reveals a consistent pattern across all instances. UCB accumulates growing regret throughout, failing to exploit parallel observations and effectively treating the problem as a standard K-armed bandit. In contrast, both `m-EXP3` and `m-ELIM` rapidly identify i^* and plateau to near-constant cumulative regret, consistent with our theoretical prediction of $O(\log T)$ instance-dependent regret.

The effect is most dramatic on **HARD20**: despite the small uniform gap $\Delta = 0.1$, both algorithms converge to regret ≈ 3 by $t \approx 2{,}000$ and grow negligibly thereafter—a roughly $200\times$ improvement over UCB at $T = 10{,}000$. On **AP40**, cheap observations enable a large m, leading to fast identification and convergence by $t \approx 2{,}000$ across the diverse gap structure. On **E-HARD20**, convergence is slower ($t \approx 5{,}000$) due to heterogeneous costs complicating the identification of i^*, but both algorithms still plateau far below UCB.

7 CONCLUSION

We introduced the cost-adaptive expert advice problem, providing the first unified theoretical treatment of online learning with heterogeneous, expert-dependent observation costs. This setting captures the economic realities of crowdsourcing, human feedback, and LLM-based evaluation systems, extending classical expert advice and bandit models that assume free or uniform-cost observations. We developed a sequence of algorithms—ranging from fixed-budget sampling to fully adaptive cost-aware strategies—and established tight minimax and instance-dependent regret bounds in both adversarial and stochastic settings. Our results reveal three fundamental regimes (expensive, balanced, and cheap observations), characterize the optimal observation budget, and provide matching information-theoretic lower bounds, demonstrating that our guarantees are optimal up to constants.

This work opens several natural directions for future research. Important extensions include contextual and instance-dependent costs and losses, learning with unknown or estimated costs, heterogeneous loss scales, and correlated experts. Another compelling direction is to move beyond our additive loss-plus-cost objective to non-additive (e.g., submodular or supermodular) aggregate losses over the queried set of experts, capturing settings where the combined output of several experts or workers behaves as a general set function; recent advances in multi-agent contract design [Castiglioni et al., 2023, Dütting et al., 2023] provide a natural starting point for this extension. More broadly, our framework provides a principled foundation for cost-aware learning from heterogeneous information sources, with immediate implications for active learning and RLHF pipelines, where intelligent allocation of expensive human and model-based feedback is increasingly critical.

Acknowledgements

This work was supported by NSF grant #2217023.

References

Shipra Agrawal and Nikhil R Devanur. Linear contextual bandits with knapsacks. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 2208–2216, 2016.

Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.

Ashwinkumar Badanidiyuru, John Langford, and Aleksandrs Slivkins. Resourceful contextual bandits. In *Conference on Learning Theory (COLT)*, pages 1109–1134, 2014.

Matteo Castiglioni, Alberto Marchesi, and Nicola Gatti. Multi-agent contract design: How to commission multiple agents with individual outcomes. In *Proceedings of the 24th ACM Conference on Economics and Computation (EC)*, pages 412–448, 2023.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications and Signal Processing. John Wiley & Sons, Hoboken, NJ, 2nd edition, 2006.

Paul Dütting, Tomer Ezra, Michal Feldman, and Thomas Kesselheim. Multi-agent contracts. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing (STOC)*, pages 1311–1324, 2023.

Marek Herde, David Husseljic, Bernhard Sick, and Alejandro Calma. A survey on cost types, interaction schemes, and learning strategies for active learning in classification. *IEEE Transactions on Knowledge and Data Engineering*, 2024.

Thomas Kesselheim and Sahil Singla. Online learning with vector costs and bandits with knapsacks. In *Conference on Learning Theory (COLT)*, volume 125 of *Proceedings of Machine Learning Research*, pages 2286–2305, 2020.

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

Vikas C Raykar, Shipeng Yu, Linda H Zhao, Gerardo H Valadez, Charles Florin, Luca Bogoni, and Linda Moy. Learning from crowds. *Journal of Machine Learning Research*, 12:1297–1322, 2010.

Yevgeny Seldin, Peter Bartlett, Koby Crammer, and Yasin Abbasi-Yadkori. Prediction with limited advice and multiarmed bandits with paid observations. In *International Conference on Machine Learning (ICML)*, pages 280–287. PMLR, 2014.

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

Tobias Sommer Thune and Yevgeny Seldin. Adaptation to easy data in prediction with limited advice. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 5096–5106, 2018.

Aaron D Tucker, Caleb Biddulph, Helen Zhang, and Md Jalal Uddin Rahman. Bandits with costly reward observations. In *International Conference on Machine Learning (ICML)*, pages 34826–34851. PMLR, 2023.

Jacob Whitehill, Paul Ruvolo, Tingfeng Wu, Jacob Bergsma, and Robert Harpaz. Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 2035–2043, 2009.

Yan Yan, Rómer Rosales, Glenn Fung, Mark W Schmidt, Gerardo Hermosillo, Luca Bogoni, and Linda Moy. Active learning from crowds. In *International Conference on Machine Learning (ICML)*, pages 1161–1168, 2011.

SUPPLEMENTARY: EXPERT ADVICE WITH COSTLY OBSERVATIONS

A APPENDIX FOR SECTION 3

A.1 PROOF OF THEOREM 2

Proof of Theorem 2. Let $L_t(i) := \ell_t(i) + p_i \in [0, 2]$ denote the combined per-round loss of expert i. Decompose regret as

$$\begin{aligned} R_T &= \sum_{t=1}^{T} \Big(\ell_t(I_t) + \sum_{i \in S_t} p_i \Big) - \min_i \sum_{t=1}^{T} (\ell_t(i) + p_i) \\ &= \Big(\sum_{t=1}^{T} L_t(I_t) - \min_i \sum_{t=1}^{T} L_t(i) \Big) + \sum_{t=1}^{T} \sum_{i \in S_t \setminus \{I_t\}} p_i. \end{aligned}$$

We bound the two terms separately.

The extra observation costs can be bound as follows. Since $|S_t| = m$ and $p_i \le \bar{p}$,

$$\sum_{t=1}^{T} \sum_{i \in S_t \setminus \{I_t\}} p_i \;\le\; (m-1)\bar{p}\, T.$$

For regret with respect to the loss, we fix t and define the observation probability

$$o_t(i) = \Pr(i \in S_t \mid q_t) = q_t(i) + (1 - q_t(i)) \frac{m-1}{K-1}.$$

By construction, for each i the estimator

$$\widehat{\ell}_t(i) = \frac{\ell_t(i) \mathbf{1}\{i \in S_t\}}{o_t(i)}$$

is unbiased:

$$\mathbb{E}[\widehat{\ell}_t(i) \mid q_t] = \ell_t(i).$$

Define $\widehat{L}_t(i) := \widehat{\ell}_t(i) + p_i$, so $\mathbb{E}[\widehat{L}_t(i) \mid q_t] = L_t(i)$.

We now apply the standard multiplicative-weights (EXP3-style) potential argument with unbiased estimated losses. Let $W_t = \sum_{i=1}^{K} w_i(t)$ and $q_t(i) = w_i(t)/W_t$. Using $\exp(-x) \le 1 - x + x^2/2$ for $x \ge 0$ and the weight update,

$$\begin{aligned} \frac{W_{t+1}}{W_t} &= \sum_{i=1}^{K} q_t(i) \exp(-\eta\, \widehat{L}_t(i)) \\ &\le \sum_{i=1}^{K} q_t(i) \Big(1 - \eta \widehat{L}_t(i) + \frac{\eta^2}{2} \widehat{L}_t(i)^2 \Big) \\ &= 1 - \eta \sum_{i=1}^{K} q_t(i) \widehat{L}_t(i) + \frac{\eta^2}{2} \sum_{i=1}^{K} q_t(i) \widehat{L}_t(i)^2. \end{aligned}$$

Taking logs and summing over t gives

$$\log \frac{W_{T+1}}{W_1} \le -\eta \sum_{t=1}^{T} \sum_{i=1}^{K} q_t(i) \widehat{L}_t(i) + \frac{\eta^2}{2} \sum_{t=1}^{T} \sum_{i=1}^{K} q_t(i) \widehat{L}_t(i)^2. \tag{3}$$

On the other hand, for any fixed expert i,

$$W_{T+1} \ge w_i(T+1) = \exp\Big(-\eta \sum_{t=1}^{T} \widehat{L}_t(i) \Big), \qquad W_1 = K,$$

so

$$\log \frac{W_{T+1}}{W_1} \geq -\eta \sum_{t=1}^{T} \widehat{L}_t(i) - \log K. \tag{4}$$

Combining (3) and (4), rearranging, and taking expectations yields

$$\begin{gathered} \mathbb{E}\Big[\sum_{t=1}^{T}\sum_{i=1}^{K} q_t(i)\widehat{L}_t(i)\Big] - \mathbb{E}\Big[\sum_{t=1}^{T}\widehat{L}_t(i)\Big] \\ \leq \\ \frac{\log K}{\eta} + \frac{\eta}{2}\mathbb{E}\Big[\sum_{t=1}^{T}\sum_{i=1}^{K} q_t(i)\widehat{L}_t(i)^2\Big]. \end{gathered} \tag{5}$$

Since $\mathbb{E}[\widehat{L}_t(i) \mid q_t] = L_t(i)$, the left-hand side of (5) equals

$$\mathbb{E}\Big[\sum_{t=1}^{T} L_t(I_t)\Big] - \sum_{t=1}^{T} L_t(i),$$

because $I_t \sim q_t$.

It remains to bound the second-moment term. Using $(x+y)^2 \leq 2x^2 + 2y^2$, we have for each i,

$$\widehat{L}_t(i)^2 = (\widehat{\ell}_t(i) + p_i)^2 \leq 2\widehat{\ell}_t(i)^2 + 2p_i^2.$$

Since $\ell_t(i) \in [0,1]$, we have $\widehat{\ell}_t(i)^2 \leq \mathbf{1}\{i \in S_t\}/o_t(i)^2$. Therefore,

$$\begin{aligned} \sum_{i=1}^{K} q_t(i)\widehat{L}_t(i)^2 &\leq 2\sum_{i=1}^{K} q_t(i)\widehat{\ell}_t(i)^2 + 2\sum_{i=1}^{K} q_t(i)p_i^2 \\ &\leq 2\sum_{i=1}^{K} q_t(i)\frac{\mathbf{1}\{i \in S_t\}}{o_t(i)^2} + 2\bar{p}^2. \end{aligned}$$

Taking expectation conditional on q_t and using $\Pr(i \in S_t \mid q_t) = o_t(i)$ yields

$$\begin{aligned} &\mathbb{E}\left[\sum_{i=1}^{K} q_t(i)\widehat{L}_t(i)^2 \;\middle|\; q_t\right] \\ &\leq 2\sum_{i=1}^{K} q_t(i)\frac{o_t(i)}{o_t(i)^2} + 2\bar{p}^2 = 2\sum_{i=1}^{K}\frac{q_t(i)}{o_t(i)} + 2\bar{p}^2. \end{aligned}$$

Finally, for all i we have $o_t(i) \geq \frac{m-1}{K-1}$, hence

$$\sum_{i=1}^{K}\frac{q_t(i)}{o_t(i)} \leq \frac{K-1}{m-1}\sum_{i=1}^{K} q_t(i) = \frac{K-1}{m-1}.$$

Therefore,

$$\mathbb{E}\Big[\sum_{t=1}^{T}\sum_{i=1}^{K} q_t(i)\widehat{L}_t(i)^2\Big] \leq T\left(2\cdot\frac{K-1}{m-1} + 2\bar{p}^2\right).$$

Plugging into (5) and optimizing $\eta = \sqrt{\frac{(m-1)\log K}{(K-1)T}}$ yields

$$\mathbb{E}\Big[\sum_{t=1}^{T} L_t(I_t) - \min_i \sum_{t=1}^{T} L_t(i)\Big] \leq 2\sqrt{\frac{K-1}{m-1}\,T\log K} \;+\; \eta T\bar{p}^2.$$

Since $\bar{p} \leq 1$, we have $\eta T\bar{p}^2 \leq \eta T\bar{p}$. For the chosen learning rate, $\eta = \sqrt{\frac{(m-1)\log K}{(K-1)T}} \leq 1$, whenever $T \geq \frac{(m-1)\log K}{K-1}$, and otherwise the bound is trivial up to constants. Hence $\eta T\bar{p}^2 \leq T\bar{p}$. For $m \geq 2$, this term is of order $\eta T\bar{p}^2 \leq \bar{p}T$ and hence is dominated by the $(m-1)\bar{p}\,T$ term up to universal constants whenever $m \geq 2$. Adding the extra observation cost term $(m-1)\bar{p}\,T$ completes the proof. □

A.2 PROOF OF COROLLARY 3

Proof of Corollary Corollary 3. From Theorem 2, for each integer $m \in \{2, \dots, K\}$ there exists a choice of η such that

$$\mathbb{E}[R_T] \;\le\; a(m-1)^{-1/2} + b(m-1),$$

where $a = C\sqrt{(K-1)T\log K}$ and $b = \bar{p}T$. For $m = 1$ we have the bandit bound $\mathbb{E}[R_T] \le a$ up to constants.

Define $x := m - 1 \in \{0, \dots, K-1\}$ and consider

$$f(x) := ax^{-1/2} + bx, \qquad x \in [1, K-1].$$

If $b = 0$, take $m = K$ to get $\mathbb{E}[R_T] \le a/\sqrt{K-1}$.

Assume $b > 0$ and treat x as real. Then

$$f'(x) = -\tfrac{1}{2}ax^{-3/2} + b.$$

The stationary point is $x^\star = \left(\frac{a}{2b}\right)^{2/3}$.

We split into three cases.

Case 1: $x^\star \le 1$. This is equivalent to $b \ge \frac{1}{2}a$. Then $m = 1$ is optimal and $\mathbb{E}[R_T] \le a$.

Case 2: $1 \le x^\star \le K-1$. Let $x = \lceil x^\star \rceil$. Then

$$f(x) \le f(x^\star) + b.$$

Direct calculation gives $f(x^\star) = \frac{3}{2^{2/3}}a^{2/3}b^{1/3}$. Hence

$$\mathbb{E}[R_T] \le \frac{3}{2^{2/3}}a^{2/3}b^{1/3} + b.$$

Case 3: $x^\star \ge K-1$. Then $m = K$ minimizes regret and

$$\mathbb{E}[R_T] \le \frac{a}{\sqrt{K-1}} + b(K-1).$$

This completes the proof. □

B APPENDIX FOR SECTION 4

B.1 PROOF OF COROLLARY 5

Proof of Corollary 5. Apply Theorem 4 and write its bound:

$$\mathbb{E}[R_T] \;\le\; \frac{a}{m} + (m-1)b + 2,$$

Minimizing over $m \in [K]$ yields the first inequality. Since $f(m) = a/m + (m-1)b$ is convex over $m \ge 1$, rounding the unconstrained minimizer to the nearest integer changes the value by at most an additive b term, which is absorbed into constants.

If $b = 0$, then $\min_{1\le m\le K}\{a/m + (m-1)b\} = a/K$, giving the third case (with $b = 0$). Now assume $b > 0$ and consider $f(m) = a/m + (m-1)b$ over real $m \ge 1$. Its unconstrained minimizer is $m^\star = \sqrt{a/b}$, with $f(m^\star) = 2\sqrt{ab} - b$.

Case 1: $m^\star \le 1$ ($b \ge a$). Then f is minimized at $m = 1$, giving $\min_{1\le m\le K} f(m) = f(1) = a$.

Case 2: $1 \le m^\star \le K$ with $\frac{a}{K^2} \le b \le a$. Let $m := \lceil m^\star \rceil$. Since $m \ge m^\star$, the term a/m is at most $a/m^\star = \sqrt{ab}$, and since $m \le m^\star + 1$, the linear term increases by at most b. Thus $f(m) \le f(m^\star) + b \le 2\sqrt{ab}$.

Case 3: $m^\star \ge K$ ($b \le a/K^2$). Then the minimum over $m \in [1, K]$ is attained at $m = K$, giving $\min_{1\le m\le K} f(m) = f(K) = a/K + (K-1)b$. Adding the $+2$ term completes the proof. □

Remark (Choosing m without knowledge of the gaps: online estimation of Δ_i). The cost-adaptive choice of m in Corollary 5 depends on the problem difficulty $H = \sum_{i \neq i^\star} 1/\Delta_i$, which is unknown a priori. The gaps can, however, be estimated online from the same confidence bounds already maintained by m−ELIM, at no additional cost in the regret guarantee. For each expert i and round t, Hoeffding's inequality yields the confidence bounds

$$\mathrm{UCB}(\theta_i, t) := \widehat{\theta}_i(t) + \mathrm{rad}_i(t), \qquad \mathrm{LCB}(\theta_i, t) := \widehat{\theta}_i(t) - \mathrm{rad}_i(t),$$

with $\mathrm{rad}_i(t) = \sqrt{L/(2O_i(t))}$ as in Algorithm 2. From these one can construct optimistic and pessimistic gap estimates

$$\mathrm{UCB}(\Delta_i, t) := \mathrm{UCB}(\theta_i, t) - \min_{j \in [K]} \mathrm{LCB}(\theta_j, t), \qquad \mathrm{LCB}(\Delta_i, t) := \mathrm{LCB}(\theta_i, t) - \min_{j \in [K]} \mathrm{UCB}(\theta_j, t).$$

On the event $\mathcal{E}$ of the proof of Theorem 4, every θ_j lies in $[\mathrm{LCB}(\theta_j, t), \mathrm{UCB}(\theta_j, t)]$, so that $\min_j \mathrm{LCB}(\theta_j, t) \le \theta_\star \le \min_j \mathrm{UCB}(\theta_j, t)$ and hence

$$\mathrm{LCB}(\Delta_i, t) \;\le\; \Delta_i \;\le\; \mathrm{UCB}(\Delta_i, t) \qquad \text{for all } i \neq i^\star \text{ and all } t.$$

Consequently, the case analysis of Corollary 5 can be made fully data-driven by substituting $\mathrm{LCB}(\Delta_i, t)$ for Δ_i in conditions of the form $(\cdot) \ge a$ and $\mathrm{UCB}(\Delta_i, t)$ in conditions of the form $(\cdot) \le a$, which accounts for the estimation uncertainty while preserving the validity of each case. This, in turn, naturally leads to an adaptive, per-round choice of the observation budget m_t based on the current gap estimates. Since the widths $\mathrm{UCB}(\Delta_i, t) - \mathrm{LCB}(\Delta_i, t) \le 2\,\mathrm{rad}_i(t) + 2\max_j \mathrm{rad}_j(t)$ shrink at the fast Hoeffding rate $O\big(\sqrt{\log(KT)/O_i(t)}\big)$ under the elimination schedule of Algorithm 2, accounting for the estimation error affects the bound of Theorem 4 only by universal constant factors; i.e., the adaptive choice of m_t incurs no additional regret cost.

C APPENDIX FOR SECTION 5

C.1 PROOF OF COROLLARY 7

Proof of Corollary 7. Using Theorem 6 (and its notation), we split into cases according to the value of $m^\star$.

Case 1: $m^\star \le 1$. The condition $m^\star \le 1$ is equivalent to $p \ge \frac{1}{2}\sqrt{K/T}$. In this case $\widehat{m} = 1$, and (1) gives

$$\mathbb{E}[R_T] \;\ge\; c\sqrt{KT}.$$

Case 2: $1 \le m^\star \le K$. This is equivalent to $\frac{1}{2K\sqrt{T}} \le p \le \frac{1}{2}\sqrt{K/T}$. In this case $\widehat{m} = \lfloor m^\star \rceil$ and in particular

$$\frac{1}{2} m^\star \;\le\; \widehat{m} \;\le\; 2m^\star, \tag{6}$$

since rounding to the nearest integer changes a positive quantity by at most a factor 2. Using (6) in (1) yields, for a universal constant $c_2 > 0$,

$$\mathbb{E}[R_T] \;\ge\; c_2 \left(\sqrt{\frac{K}{m^\star} T} + pT(m^\star - 1) \right).$$

Now compute $\sqrt{\frac{K}{m^\star} T} = (KT)^{1/2} (m^\star)^{-1/2} = (KT)^{1/2} \left(\frac{1}{2p} \sqrt{\frac{K}{T}} \right)^{-1/3} = 2^{1/3} (Kp)^{1/3} T^{2/3}$, and

$$pTm^\star = pT \left(\frac{1}{2p} \sqrt{\frac{K}{T}} \right)^{2/3} = 2^{-2/3} (Kp)^{1/3} T^{2/3}.$$

Therefore, $\sqrt{\frac{K}{m^\star} T} + pT(m^\star - 1) = \frac{3}{2^{2/3}} (Kp)^{1/3} T^{2/3} - pT$. Dropping the negative $-pT$ term and absorbing constants into c' gives

$$\mathbb{E}[R_T] \;\ge\; c' (Kp)^{1/3} T^{2/3}.$$

Case 3: $m^\star \ge K$. The condition $m^\star \ge K$ is equivalent to $p \le \frac{1}{2K\sqrt{T}}$. In this case $\widehat{m} = K$, and (1) gives

$$\mathbb{E}[R_T] \;\ge\; c'' \left(\sqrt{T} + pT(K - 1) \right).$$

Finally, renaming constants completes the proof. □

C.2 PROOF OF COROLLARY 9

Proof of Corollary 9. By Theorem 8, there exists an instance such that for any fixed $m \in \{1, \dots, K\}$,

$$\mathbb{E}[R_T] \;\geq\; c\left(\frac{(\log T)\, S}{m} + pT(m-1)\right).$$

Consider the function $f(m) \coloneqq \frac{(\log T)\, S}{m} + pT(m-1)$ for $m \in [1, K]$. Ignoring the additive constant $-pT$, the convex function $\frac{(\log T)\, S}{m} + pTm$ has unique minimizer

$$m^\star = \sqrt{\frac{(\log T)\, S}{pT}}.$$

We split into cases according to the value of $m^\star$.

Case 1: $m^\star \leq 1$. This is equivalent to $pT \geq (\log T)\, S$, i.e. $p \geq \frac{(\log T)\, S}{T}$. Then the clipped minimizer is $m = 1$ and

$$\mathbb{E}[R_T] \;\geq\; c\,((\log T)\, S + pT(0)) = c(\log T)\, S.$$

Case 2: $1 \leq m^\star \leq K$. This is equivalent to $\frac{(\log T)\, S}{TK^2} \leq p \leq \frac{(\log T)\, S}{T}$. Rounding to $\widehat{m} = \lfloor m^\star \rceil$ changes $m^\star$ by at most a constant factor, hence $f(\widehat{m}) \geq c_0 f(m^\star)$ for a universal $c_0 \in (0, 1]$. Evaluating at $m^\star$ gives

$$f(m^\star) = \frac{(\log T)\, S}{m^\star} + pT(m^\star - 1) = 2\sqrt{pT\,(\log T)\, S} - pT,$$

and we can drop the negative term and absorb constants.

Case 3: $m^\star \geq K$. This is equivalent to $(\log T)\, S \geq pTK^2$, i.e. $p \leq \frac{(\log T)\, S}{TK^2}$. Then the clipped minimizer is $m = K$ and

$$\mathbb{E}[R_T] \;\geq\; c\left(\frac{(\log T)\, S}{K} + pT(K-1)\right).$$

Renaming constants completes the proof. □

D INFORMATION-THEORETIC TOOLS

We recall two standard results used in the proofs in Section 5. Both of these results are classical [Cover and Thomas, 2006, Lattimore and Szepesvári, 2020].

Lemma 10 (Pinsker's inequality)**.** *For any two probability measures P and Q,*

$$\|P - Q\|_{\mathrm{TV}} \leq \sqrt{\tfrac{1}{2} D_{\mathrm{KL}}(P\|Q)}.$$

Lemma 11 (KL chain rule)**.** *Let P and Q be distributions over a sequence of random variables $(Z_1, \dots, Z_T)$ with respective conditional distributions $P_t(\cdot \mid Z_{<t})$ and $Q_t(\cdot \mid Z_{<t})$. Then*

$$D_{\mathrm{KL}}(P\|Q) = \sum_{t=1}^{T} \mathbb{E}_P[D_{\mathrm{KL}}(P_t(\cdot \mid Z_{<t})\|Q_t(\cdot \mid Z_{<t}))]\,.$$

E EXPERIMENTAL DETAILS

Table 1 reports the full numerical results for the empirical validation in Section 6 (mean $\pm$ std over $R = 40$ independent runs).

Table 1: Average cumulative regret (mean $\pm$ std, $R = 40$ runs). `m-EXP3` and `m-ELIM` plateau after identifying i^*, while UCB grows linearly throughout.

Instance	t	UCB	`m-ELIM`	`m-EXP3`
AP40	1,000	307.56 ± 6.35	110.77 ± 19.75	80.01 ± 4.05
	5,000	1082.25 ± 29.75	147.46 ± 36.56	84.33 ± 4.47
	10,000	1494.44 ± 57.51	147.46 ± 36.56	84.33 ± 4.47
HARD20	1,000	165.67 ± 8.09	2.97 ± 1.60	3.10 ± 0.79
	5,000	485.86 ± 33.49	3.01 ± 1.79	3.21 ± 0.80
	10,000	616.76 ± 41.78	3.01 ± 1.79	3.21 ± 0.80
E-HARD20	1,000	176.40 ± 2.67	103.56 ± 10.70	72.65 ± 3.53
	5,000	693.97 ± 28.40	252.15 ± 62.74	92.96 ± 5.35
	10,000	1019.26 ± 48.78	252.15 ± 62.74	92.96 ± 5.35