Online Algorithms for Repeated Optimal Stopping: Achieving Both Competitive Ratio and Regret Bounds

Tsubasa Harada[1], Yasushi Kawase[2], and Hanna Sumita[1]

[1]Institute of Science Tokyo
[2]University of Tokyo

Abstract

We study the *repeated optimal stopping problem*, which generalizes the classical optimal stopping problem with an unknown distribution to a setting where the same problem is solved repeatedly over T rounds. In this framework, we aim to design algorithms that guarantee a competitive ratio in each round while also achieving sublinear regret across all rounds. Our primary contribution is a general algorithmic framework that achieves these objectives simultaneously for a wide array of repeated optimal stopping problems. The core idea is to dynamically select an algorithm for each round, choosing between two candidates: (1) an empirically optimal algorithm derived from the history of observations, and (2) a sample-based algorithm with a proven competitive ratio guarantee. Based on this approach, we design an algorithm that performs no worse than the baseline sample-based algorithm in every round, while ensuring that the total regret is bounded by $\tilde{O}(\sqrt{T})$. We demonstrate the broad applicability of our framework to canonical problems, including the prophet inequality, the secretary problem, and their variants under adversarial, random, and i.i.d. input models. For example, for the repeated prophet inequality problem, our method achieves a $1/2$-competitive ratio from the second round on and an $\tilde{O}(\sqrt{T})$ regret. Furthermore, we establish a regret lower bound of $\Omega(\sqrt{T})$ even in the i.i.d. model, confirming that our algorithm's performance is almost optimal with respect to the number of rounds.

1 Introduction

Imagine a firm that repeatedly faces a sequence of purchase offers. In each period, it observes a sequence of prices and must decide when to accept an offer to sell. Early on, the firm operates with limited data and cannot yet determine effective stopping rules. As more rounds are played, it adapts its strategy using the experience it accumulates. Importantly, a single large loss in any period can severely damage the firm's financial health or reputation, making robust performance in individual rounds just as essential as overall success. Therefore, over time, the firm's goal is not only to ensure long-term performance, but also to avoid severe losses in any single period.

Two classical performance measures capture complementary viewpoints. Minimizing *regret* allows the algorithm to learn and gradually converge to the best strategy that could have been chosen in hindsight. However, it does not prevent large losses in individual rounds. In contrast, maintaining a strong *competitive ratio* in each round ensures reliable performance even in the worst situations, but it overlooks the potential for long-term learning and improvement. Neither objective

alone fully reflects the challenges of repeated online decision-making, where both reliability within each round and adaptability across rounds are essential. This raises the following question:

> *Can we design an online algorithm that simultaneously achieves a sublinear regret over all rounds and a strong competitive ratio in every single round?*

We investigate this question through the framework of *repeated optimal stopping*. In each round, the algorithm faces one optimal stopping instance: it observes a sequence of values drawn from some probability distributions and decides when to stop without knowing the future. Although the distributions of these values are initially unknown, they remain fixed across rounds, allowing the algorithm to learn from accumulated experience.

Our framework unifies and extends these two paradigms, revealing new algorithmic challenges at their intersection. In particular, the algorithm must adapt its stopping thresholds over time while ensuring that each round remains competitive. We provide nearly optimal guarantees in this setting, establishing the first theoretical foundation for achieving both per-round robustness and cross-round learning in sequential stopping problems.

1.1 Settings

In the repeated optimal stopping problem, we are asked to solve the same optimal stopping problem instance T times, aiming to maximize the total profit over T rounds. We begin by describing the optimal stopping problem in a general form. For a positive integer k, define $[k] \coloneqq \{1, 2, \ldots, k\}$. Let n be a positive integer. Let $\mathbf{D} = (D_1, \ldots, D_n; \Pi)$ denote n independent probability distributions $D_1, \ldots, D_n$ over $[0, 1]$, and a distribution Π over the set of permutations of $[n]$. For each $i \in [n]$, let Y_i be a random variable drawn from D_i. Let π be a random permutation sampled according to Π. We denote the random vector $\mathbf{X} = (X_1, \ldots, X_n)$ as $(Y_{\pi(1)}, \ldots, Y_{\pi(n)})$. In the optimal stopping problem, realizations $\boldsymbol{y} = (y_1, \ldots, y_n)$ and τ are drawn from $D_1 \times \cdots \times D_n$ and Π, respectively, and define $\boldsymbol{x} = (x_1, \ldots, x_n) = (y_{\tau(1)}, \ldots, y_{\tau(n)})$. At each step $i \in [n]$, the value x_i and the index $\tau(i)$ are revealed to the algorithm. The algorithm can know that the value x_i is drawn from $D_{\tau(i)}$. Based on the observed values $(x_1, \ldots, x_i)$ and $(\tau(1), \ldots, \tau(i))$, the algorithm must decide irrevocably whether to (a) accept x_i and stop (i.e., automatically reject the subsequent values), or (b) reject x_i and proceed to the next step. Accepting x_i yields a profit of $p(\boldsymbol{x}, i)$, determined by a known profit function $p\colon [0, 1]^n \times [n + 1] \to [0, B]$, where B is a fixed constant. If all values are rejected, the profit is $p(\boldsymbol{x}, n + 1)$. An instance of the optimal stopping problem is identified by the pair $(\mathbf{D}, p)$, where $\mathbf{D}$ is unknown in advance, but p is known.

We remark that if Π is a point-mass at a single permutation, then the problem setting corresponds to the *adversarial order* model, and if Π is a uniform distribution, then it corresponds to the *random order* model. Thus, the problem includes these major models. See Section 2 for details.

A *repeated optimal stopping problem* is specified by a profit function p, a family of possible distributions $\mathfrak{D}$, and the number of rounds T. A repeated optimal stopping problem instance is denoted as $(\mathbf{D}, p; T)$, where $\mathbf{D} \in \mathfrak{D}$ is an unknown distribution, p is a known profit function, and T is the known number of rounds. In each round $t \in [T]$, (1) a vector $\boldsymbol{x}^{(t)}$ and a permutation $\tau^{(t)}$ are determined according to $\mathbf{D}$, (2) the algorithm makes a decision step by step for $(\boldsymbol{x}^{(t)}, \tau^{(t)})$, and (3) the pair $(\boldsymbol{x}^{(t)}, \tau^{(t)})$ is given to the algorithm at the end of the round (i.e., full feedback setting). Just before round t, the algorithm selects an (possibly randomized) online algorithm h_t for round t based on the observations $\mathbb{x}_{t-1} = (\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(t-1)})$ and $\mathbb{\tau}_{t-1} = (\tau^{(1)}, \ldots, \tau^{(t-1)})$ in previous $t - 1$ rounds. The goal is to maximize the sum of profits obtained in the T rounds.

1.2 Competitive Ratio and Regret

Next, we introduce the performance measures of *competitive ratio* and *regret*, highlighting their differences. We denote by $\mathrm{Inj}([i],[n])$ the set of injective functions from $[i]$ to $[n]$. The set of permutations over $[n]$ is represented as $\mathrm{Inj}([n],[n])$. For $i \in [n]$, $\boldsymbol{x} \in [0,1]^n$, and $\tau \in \mathrm{Inj}([n],[n])$, we write $\boldsymbol{x}_{\le i} \coloneqq (x_1,\dots,x_i) \in [0,1]^i$ and $\tau_{\le i} \coloneqq (\tau(1),\dots,\tau(i)) \in \mathrm{Inj}([i],[n])$.

A *deterministic online algorithm* for the optimal stopping problem $(\mathfrak{D},p)$ is defined by a sequence of n functions $h = (f_i)_{i=1}^n$, where each $f_i \colon [0,1]^i \times \mathrm{Inj}([i],[n]) \to \{0,1\}$ takes as input the first i observed values $\boldsymbol{x}_{\le i}$ and $\tau_{\le i}$, and outputs 1 if the algorithm chooses to accept at step i, and 0 otherwise. The algorithm proceeds sequentially: it selects the first index $i \in [n]$ such that $f_i(\boldsymbol{x}_{\le i},\tau_{\le i}) = 1$, and stops; all subsequent variables are rejected. If $f_i(\boldsymbol{x}_{\le i},\tau_{\le i}) = 0$ for all $i \in [n]$, the algorithm rejects all the indices and obtains $p(\boldsymbol{x},n+1)$. Let $\mathcal{H}$ denote the set of all deterministic online algorithms. A *randomized online algorithm* is a probability distribution over $\mathcal{H}$.

For a profit function p, an algorithm $h = (f_i)_{i\in[n]}$, and a pair $(\boldsymbol{x},\tau) \in [0,1]^n \times \mathrm{Inj}([n],[n])$, we define the profit of h for $(\boldsymbol{x},\tau)$ as

$$h(\boldsymbol{x},\tau;p) \coloneqq p(\boldsymbol{x},\min\{i \in [n+1] \mid i = n+1 \text{ or } f_i(\boldsymbol{x}_{\le i},\tau_{\le i}) = 1\}).$$

For an optimal stopping problem instance $(\mathbf{D},p)$, the expected profit $h(\mathbf{D};p)$ of h is defined as

$$h(\mathbf{D};p) \coloneqq \mathbb{E}_{(\mathbf{X},\pi)\sim\mathbf{D}}[h(\mathbf{X},\pi;p)].$$

If h is a randomized algorithm, the expectation is also taken over the algorithm's internal randomness. When the profit function p is clear from context, we may abbreviate $h(\boldsymbol{x},\tau;p)$ and $h(\mathbf{D};p)$ as $h(\boldsymbol{x},\tau)$ and $h(\mathbf{D})$, respectively. We define the *optimal online profit* by

$$\mathrm{Opt}_{\mathrm{online}}(\mathbf{D};p) \coloneqq \sup_{h\in\mathcal{H}} h(\mathbf{D};p) = \sup_{h\in\mathcal{H}} \mathbb{E}_{(\mathbf{X},\pi)\sim\mathbf{D}}\left[h(\mathbf{X},\pi;p)\right],$$

where the supremum is taken over all online algorithm h. The *optimal offline profit* is given by

$$\mathrm{Opt}_{\mathrm{offline}}(\mathbf{D};p) \coloneqq \mathbb{E}_{(\mathbf{X},\pi)\sim\mathbf{D}}\left[\max_{i\in[n+1]} p(\mathbf{X},i)\right],$$

which corresponds to the expected profit achievable when the realizations are completely known in advance. By definition, we have $\mathrm{Opt}_{\mathrm{online}}(\mathbf{D};p) \le \mathrm{Opt}_{\mathrm{offline}}(\mathbf{D};p)$ for any $\mathbf{D}$ and p.

For an online algorithm h and an optimal stopping (profit maximization) problem instance $(\mathbf{D},p)$, the *competitive ratio* is defined as the ratio between the expected profit of h and the optimal offline profit, i.e.,

$$\frac{h(\mathbf{D};p)}{\mathrm{Opt}_{\mathrm{offline}}(\mathbf{D};p)}.$$

The ratio is at most $\mathrm{Opt}_{\mathrm{online}}(\mathbf{D};p)/\mathrm{Opt}_{\mathrm{offline}}(\mathbf{D};p)$ (≤ 1), and larger values indicate better performance. If $(\mathbf{D},p)$ is a cost minimization problem, then the ratio is at *least* 1, and smaller values indicate better performance. For an online algorithm h and an optimal stopping (profit maximization) problem $(\mathfrak{D},p)$, the competitive ratio is defined as the worst-case ratio $\inf_{\mathbf{D}\in\mathfrak{D}} h(\mathbf{D};p)/\mathrm{Opt}_{\mathrm{offline}}(\mathbf{D};p)$. Similarly, for a cost minimization problem $(\mathfrak{D},p)$, the competitive ratio of h is defined as the worst-case ratio $\sup_{\mathbf{D}\in\mathfrak{D}} h(\mathbf{D};p)/\mathrm{Opt}_{\mathrm{offline}}(\mathbf{D};p)$.

In the repeated optimal stopping problem $(\mathbf{D}, p; T)$, the algorithm selects an online algorithm for each round t based on the realizations from previous rounds, which we denote by $\mathbb{X}_{t-1} = (\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(t-1)})$ and $\mathbb{\tau}_{t-1} = (\tau^{(1)}, \dots, \tau^{(t-1)})$. We denote by $h[\mathbb{X}_{t-1}, \mathbb{\tau}_{t-1}]$ the online algorithm chosen in round t. Then, the competitive ratio of the algorithm in round t is defined as

$$\frac{\mathbb{E}_{(\mathbf{X}^{(1)},\tau^{(1)}),\dots,(\mathbf{X}^{(t)},\tau^{(t)}) \overset{\text{i.i.d.}}{\sim} \mathbf{D}} \left[h[\mathbb{X}_{t-1}, \mathbb{\tau}_{t-1}](\mathbf{X}^{(t)}, \tau^{(t)}; p) \right]}{\mathrm{Opt}_{\mathrm{offline}}(\mathbf{D}; p)}.$$

On the other hand, the regret is defined for the outcome over T rounds. The *regret* of the algorithm is defined as

$$\text{Regret} := \mathbb{E}_{(\mathbf{X}^{(1)},\tau^{(1)}),\dots,(\mathbf{X}^{(T)},\tau^{(T)}) \overset{\text{i.i.d.}}{\sim} \mathbf{D}} \left[\sum_{t=1}^{T} \left(\mathrm{Opt}_{\mathrm{online}}(\mathbf{D}; p) - h[\mathbb{X}_{t-1}, \mathbb{\tau}_{t-1}](\mathbf{X}^{(t)}, \tau^{(t)}; p) \right) \right], \quad (1)$$

where the expectation is taken over all randomness in the process (including both the draws from $\mathbf{D}$ and any internal randomness of the algorithms). For a cost minimization problem, the regret is defined as the negation of (1). The regret is trivially at most $B \cdot T$, and a regret bound that is sublinear in T is required.

The competitive ratio measures the difference from the offline optimal, i.e., the situation where the complete information of the future is available. This is a strong goal, and it is often criticized that the competitive analysis can be overly pessimistic. On the other hand, the regret measures the difference from the best online algorithm that knows the complete information of the distribution. While this is a weaker yet reasonable goal, the regret analysis has two main limitations: the regret analysis is suitable for the problems with repetitions, and it guarantees long-term and asymptotic performance, offering little insight into a single round, especially in the earlier rounds.

This motivates the need for a framework that can simultaneously evaluate both regret and competitive ratio. Specifically, our goal is to guarantee the overall performance by the regret as well as the round-wise performance by the competitive ratio.

1.3 Related Work

The optimal stopping problem is a central topic in online optimization, where the objective is to select the best option from a sequence of candidates that appear sequentially. This problem has been widely studied in various frameworks, most notably the *prophet inequality* and the *secretary problem*. The prophet inequality addresses the challenge of maximizing the expected reward by selecting one of n options, each with a random value drawn independently from known distributions. In contrast, the secretary problem focuses on maximizing the probability of choosing the best option from a sequence presented in random order, where decisions must be made immediately and irrevocably upon seeing each option. It is well-known that the best possible competitive ratio for the prophet inequality is $1/2$ [29], and that for the secretary problem is $1/e$ [15].

In many practical scenarios, the distribution $\mathbf{D}$ is unknown, and only revealed through samples. This leads to the study of the *optimal stopping problem with samples*, where the goal is to select the best option based on a limited number of samples drawn from the distribution. Rubinstein et al. [36] proposed a stopping policy with a competitive ratio of $1/2$ for the prophet inequality, using only a single sample from each distribution. Nuti and Vondrák [35] studied the single-sample secretary problem in both random and adversarial order settings. There are also various studies on the optimal stopping problem with samples [4, 6, 9–13, 25, 40].

The *metrical task system* (MTS) has long been a central subject in competitive analysis, and it is also applicable to regret analysis through the framework of *online convex optimization with switching costs*. As a result, some efforts have been made to design algorithms for MTS that perform well with respect to both competitive ratio and regret [2, 5, 14, 19]. These studies leverage the repetitive structure inherent in MTS, which makes it challenging to directly apply such approaches to the optimal stopping problem considered in this paper.

Research concerning repeated settings of optimal stopping problems has garnered significant attention in recent years, with particularly vigorous studies focused on the prophet inequality (and the Pandora's box problem)[1, 17, 18, 20, 21, 24]. Within the framework of PAC learning, Jin et al. [24] showed that $O(\frac{1}{\varepsilon^2} \cdot \log^2(1/\delta))$ samples are sufficient to find an ε-additive approximation of the optimal online algorithm with probability at least $1-\delta$ for the prophet inequality. Gatmiry et al. [18] proposed an algorithm for the repeated prophet inequality that achieves a regret of $O(n^3\sqrt{T}\log T)$ under partial feedback, where only the values of the selected variables are observed. They also established a regret lower bound of $\Omega(\sqrt{T})$ using two random variables drawn from different distributions. Agarwal et al. [1] introduced a general framework for stochastic optimization problems satisfying monotonicity under semi-bandit feedback, and they derived an upper bound on the regret of $O(Bn\sqrt{T\log T})$. Furthermore, Shah and Rajkumar [37] performed a regret analysis for the repeated ski rental problem. Note that these studies primarily focus on regret analysis and sample complexity, and do not address the per-round performance.

1.4 Our Results

Our main contribution is to present a general framework of algorithms for the repeated optimal stopping problem. Our general result (Theorem 5) is informally stated as follows.

Theorem 1 (informal version of Theorem 5)**.** *Let g_t be an online algorithm for $(\mathbf{D}, p)$ with $t-1$ samples for each $t \in [T]$. Suppose the profit function p satisfies mild assumptions. Then, we construct a sequence of algorithms $(h_t)_{t\in[T]}$ for $(\mathbf{D}, p; T)$ that achieves the following:*

- *For each round $t \in [T]$, the expected profit of h_t is at least that of g_t.*
- *The regret is $O(B\sqrt{n|\Pi|T\log T})$, where $|\Pi|$ is the support size of Π for $\mathbf{D} = (D_1, \dots, D_n; \Pi)$.*

This implies that the competitive ratio for the overall profit is asymptotically optimum, i.e.,

$$\frac{\mathbb{E}_{(\mathbf{X}^{(1)},\tau^{(1)}),\dots,(\mathbf{X}^{(T)},\tau^{(T)})\overset{\text{i.i.d.}}{\sim}\mathbf{D}}\left[\sum_{t=1}^{T} h[\mathbb{X}_{t-1}, \mathbb{w}_{t-1}](\mathbf{X}^{(t)}, \tau^{(t)}; p)\right]}{T \cdot \mathrm{Opt}_{\text{offline}}(\mathbf{D}; p)} = (1 - o(1)) \cdot \frac{\mathrm{Opt}_{\text{online}}(\mathbf{D}; p)}{\mathrm{Opt}_{\text{offline}}(\mathbf{D}; p)}.$$

Our framework is applicable to a repeated setting on broad classes of optimal stopping problems. We refer to a class of the repeated optimal stopping problem $(\mathfrak{D}, p; T)$ as "the repeated X" whenever the pair $(\mathfrak{D}, p)$ corresponds to problem X.[1] We consider the following profit functions, each satisfying the mild assumptions required by our analysis: reward $p^{\text{rwd}}(\boldsymbol{x}, i) = x_i$, best choice $p^{\text{best}}(\boldsymbol{x}, i) = \mathbb{1}_{\{x_i = \max_{j\in[n]} x_j\}}$, last success $p^{\text{ls}}(\boldsymbol{x}, i) = \mathbb{1}_{\{i = \max\{i' | x_{i'} = 1\}\}}$, and ski-rental $p^{\text{ski}}(\boldsymbol{x}, i) = \sum_{j=1}^{i-1} x_j + b$. Similarly, we define four distribution models: the adversarial-order model $\mathfrak{D}^{\text{adv}}$, the random-order model $\mathfrak{D}^{\text{r.o.}}$, the i.i.d. distribution model $\mathfrak{D}^{\text{i.i.d.}}$, and the general order model $\mathfrak{D}^{\text{all}}$. Our general theorem yields the following results, which are described in Section 5.

[1] In the literature, the term "secretary" is used in two different senses: (i) to indicate the *random-order* arrival model and (ii) to denote the *best-choice* objective of selecting the overall maximum.

- The repeated prophet inequality problem $(\mathfrak{D}^{\mathrm{adv}}, p^{\mathrm{rwd}}; T)$: Equipped with the single-sample algorithm [36], our algorithm attains a competitive ratio of $1/n$ for the first round and $1/2$ for rounds $t \geq 2$. The same competitive ratios hold even for the repeated general order prophet inequality problem $(\mathfrak{D}^{\mathrm{all}}, p^{\mathrm{rwd}}; T)$ with the identical algorithm.
- The repeated prophet secretary problem $(\mathfrak{D}^{\mathrm{r.o.}}, p^{\mathrm{rwd}}; T)$: Using the well-known classical secretary algorithm [16], the single-sample algorithm [36] for the prophet inequality problem, and the algorithm proposed by Cristi and Ziliotto [13], our algorithm has a competitive ratio of $1/e$ for the first round, $1/2$ for the second round, and $0.688 - O(t^{-\frac{1}{5}})$ for rounds $t \geq 3$.
- The repeated i.i.d. prophet inequality $(\mathfrak{D}^{\mathrm{i.i.d.}}, p^{\mathrm{rwd}}; T)$: our algorithm equipped with the fresh-looking samples algorithm [10] and Correa et al.'s algorithm [12] has a competitive ratio of $1/e$ for the first round, $1 - 1/e$ for the second round, and $0.745 - O(t^{-1})$ for rounds $t \geq 3$.
- The repeated adversarial order secretary problem $(\mathfrak{D}^{\mathrm{adv}}, p^{\mathrm{best}}; T)$: When combined with the single-sample algorithm [35], our algorithm achieves a competitive ratio of $1/n$ for the first round and $1/4$ for rounds $t \geq 2$. Using the same algorithm, these competitive ratios also hold for the repeated general order secretary problem $(\mathfrak{D}^{\mathrm{all}}, p^{\mathrm{best}}; T)$.
- The repeated random order secretary problem $(\mathfrak{D}^{\mathrm{r.o.}}, p^{\mathrm{best}}; T)$: With the single-sample algorithm [35], our algorithm has a competitive ratio of $1/e$ for the first round and of at least 0.5009 for rounds $t \geq 2$. This algorithm is also applicable to the repeated i.i.d. secretary problem.
- The repeated last success problem $(\mathfrak{D}^{\mathrm{all}}, p^{\mathrm{ls}}; T)$: With the single-sample algorithm [40], our algorithm attains a competitive ratio of $1/n$ for the first round and of $1/4$ for rounds $t \geq 2$.
- The repeated ski-rental problem $(\mathfrak{D}^{\mathrm{all}}, p^{\mathrm{ski}}; T)$: Paired with the randomized ski-rental algorithm [26], our algorithm has a competitive ratio of $e/(e-1)$ for every round.

For any setting, the regret of our algorithm with respect to T is bounded by $O(B\sqrt{\kappa_{n,\Pi} T \log T})$, where $\kappa_{n,\Pi}$ is a constant determined by the problem setting and independent of the number of rounds T. For instance, when the distribution is $\mathfrak{D}^{\mathrm{adv}}$, we have $\kappa_{n,\Pi} = n$ and the regret is bounded by $O\big(B\sqrt{nT \log T}\big)$.

We also establish a lower bound of $\Omega(\sqrt{T})$ on the regret when $n = 2$, the profit function is p^{rwd} or p^{best}, and the distribution includes $\mathfrak{D}^{\mathrm{i.i.d.}}$. Therefore, by ignoring $\kappa_{n,\Pi}$, our regret bound is tight with respect to T, up to a logarithmic factor.

Furthermore, for the repeated prophet inequality problem $(\mathfrak{D}^{\mathrm{adv}}, p^{\mathrm{rwd}}; T)$, we also obtain the following result, which improves the dependence on $\kappa_{n,\Pi}$.

Theorem 2. *For the repeated prophet inequality problem $(\mathfrak{D}^{\mathrm{adv}}, p^{\mathrm{rwd}}; T)$ and any $\mathbf{D} \in \mathfrak{D}^{\mathrm{adv}}$, there exists a sequence of algorithms $(h_t)_{t\in[T]}$ that achieves the following:*

1. *h_1 is $1/n$-competitive,*
2. *h_t is $1/2$-competitive for $t \geq 2$, and*
3. *the regret of $(h_t)_{t\in[T]}$ is $O(B\sqrt{T}\log T)$.*

By combining the lower bound $\Omega(\sqrt{T})$ with Theorem 2, our result for the repeated prophet inequality guarantees the best possible competitive ratio in each round and a near-optimal regret.

Technique Overview.

Our goal is to simultaneously guarantee both the competitive ratio and regret. If we only aim for a competitive ratio guarantee, we can simply use existing sample-based competitive algorithms in every round, leveraging prior results. However, achieving both guarantees at once is significantly more challenging. The goal of competitive analysis is to ensure performance in the worst-case instance, while sublinear regret requires that the average performance over T rounds converges to that of an optimal online algorithm with full knowledge of the distribution. For example, in the prophet inequality problem $(\mathfrak{D}^{\mathrm{adv}}, p^{\mathrm{rwd}})$, a single-sample 1/2-competitive algorithm is known [36], and no online algorithm achieves a better competitive ratio even when the distribution is known [29]. However, there can be a performance gap between the online optimal algorithm and the sample-based online algorithm in specific instances $(\mathbf{D}, p^{\mathrm{rwd}})$, which leads to a linear regret bound. We provide such an example of linear regret in Section B.

We establish a sublinear regret bound for the repeated optimal stopping problem. The main challenge is that the space of all online algorithms is infinite. Directly selecting the one that is optimal for the empirical distribution in each round can lead to overfitting, which means a poor performance in the next round. To address this, we focus on the class of *threshold algorithms*, which decide to accept a value based on whether it exceeds a certain threshold determined by the observed history.[2] In Section 2.1, we show that for many canonical optimal stopping problems, there exists an optimal online algorithm that can be represented as a threshold algorithm. Focusing on threshold algorithms reduces the dimensionality of the space of online algorithms and helps mitigate overfitting to empirical data. As a result, we can achieve sublinear regret in T by selecting an empirically optimal threshold algorithm in each round. However, this approach alone does not provide round-wise performance bounds, especially in the early rounds, as regret analysis primarily ensures long-term performance.

The next challenge in our work is to simultaneously guarantee both measures. One possible approach is to select, in each round t, the empirically optimal threshold algorithm only if its performance, evaluated as the expectation under the true distribution $\mathbf{D}$, exceeds that of the sample-based online algorithm. Since $\mathbf{D}$ is unknown, we instead evaluate performance as the expected value under the empirical distribution and apply the empirically optimal threshold algorithm only when it outperforms the sample-based online algorithm by a suitably chosen margin. Note that we need to select a margin that is sufficiently large to ensure the performance comparison under the learning errors, yet sufficiently small to guarantee that the regret bound remains sublinear.

2 Preliminaries

In this section, we describe the basic facts in the optimal stopping problem.

First, we describe the settings of the profit function p that appear in the literature. Throughout this paper, we assume that the profit function $p\colon [0,1]^n \times [n+1] \to [0, B]$ satisfies the following two natural properties:

[2]Existing approaches reduce the number of candidate algorithms by discretizing the value space with an ε-grid or heavily rely on the property (such as "monotonicity") of an individual problem [1, 21], which requires strong conditions such as continuity of the objective function. Consequently, these methods are not directly applicable to non-monotone problems like the secretary problem, where the goal is to select the single best option from a sequence. In contrast, we reduce the number of algorithms by restricting attention to threshold-based ones, which works even under much milder assumptions on the objective function.

(i) $p((x_1, \dots, x_i, \dots, x_n), i) \leq p((x_1, \dots, x'_i, \dots, x_n), i)$ for any $i \in [n]$ and $x_1, \dots, x_n, x'_i \in [0, 1]$ with $x_i \leq x'_i$, i.e., the profit is non-decreasing in the value of the accepted variable.

(ii) $p((x_1, \dots, x_i, \dots, x_n), j) \geq p((x_1, \dots, x'_i, \dots, x_n), j)$ for any $i, j \in [n]$ and $x_1, \dots, x_n, x'_i \in [0, 1]$ with $x_i \leq x'_i$ and $j > i$, i.e., the profit is non-increasing in the value of any previously rejected variable.

This paper mainly focuses on profit maximization, while our results can be easily extended to cost minimization by negating the profit function. For cost minimization, the above inequalities are reversed. The properties (i) and (ii) retain the generality, as all of the profit functions in major optimal stopping problems, such as the prophet inequality, the secretary problem, the last success problem, and the ski-rental problem, satisfy the properties. Here, we assume that $x_{n+1} = 0$.

Expected reward: $p(\boldsymbol{x}, i) = x_i$. The objective is to maximize the expected value of the selected variable. This is the most standard setting in the prophet inequality. We denote this profit function as p^{rwd}.

Best choice: $p(\boldsymbol{x}, i) = \mathbb{1}_{\{x_i = \max_{j \in [n]} x_j\}}$. The objective is to maximize the probability of selecting the maximum value. We denote this profit function as p^{best}.

Last success: $p(\boldsymbol{x}, i) = \mathbb{1}_{\{i = \max\{i' | x_{i'} = 1\}\}}$. The objective is to maximize the probability of selecting the last success index, where X_i is called *success* if $X_i = 1$. We denote this function as p^{ls}.

Ski-rental: $p(\boldsymbol{x}, i) = \sum_{j=1}^{i-1} x_j + b$ if $i \in [n]$ and $p(\boldsymbol{x}, i) = \sum_{j=1}^{n} x_j$ if $i = n + 1$, where b is a fixed positive real number. The objective is to minimize the total cost of renting ski equipment, where x_i denotes the rental cost at step i and b is the cost of purchasing the equipment. Once the equipment is purchased, there is no need to rent it in subsequent steps. We denote this profit function as p^{ski}.

Note that the maximum profit value B can be set to 1 for the expected reward, best choice, last success, and can be set to $n + b$ for the ski-rental setting. An example of a profit function that does not satisfy properties (i) and (ii) is the scenario in which one aims to maximize the probability of selecting the second-largest value [38].

Next, we describe classes of the distributions $\mathbf{D} = (D_1, \dots, D_n; \Pi)$ that arise in typical settings of the optimal stopping problem.

General Order: The family of distributions $\mathbf{D} = (D_1, \dots, D_n; \Pi)$ where Π is any distribution over the permutations of $[n]$. We denote this set by $\mathfrak{D}^{\mathrm{all}}$. This general setting has been studied, for example, in the context of secretary problems with non-uniform arrival orders [28].

Adversarial Order: The family of distributions $\mathbf{D} = (D_1, \dots, D_n; \Pi)$ where Π is the point-mass distribution at the identity permutation. We denote this set by $\mathfrak{D}^{\mathrm{adv}}$, and refer to the point-mass distribution itself as Π^{id}.

Random Order: The family of distributions $\mathbf{D} = (D_1, \dots, D_n; \Pi)$ where Π is the uniform distribution over the permutations of $[n]$. We denote this set by $\mathfrak{D}^{\mathrm{r.o.}}$.

IID: The set of distributions $\mathbf{D} = (D_1, \dots, D_n; \Pi^{\mathrm{id}})$ where $D_1 = \cdots = D_n$. We denote it by $\mathfrak{D}^{\mathrm{i.i.d.}}$.

Forward-backward: The set of distributions $\mathbf{D} = (D_1, \dots, D_n; \Pi)$ where Π is the uniform distribution over the identity permutation of $[n]$ and its reverse. This setting has been studied in the context of online rationing at food banks, where the forward-backward model captures scenarios with both forward and reverse arrival orders, such as driving along cities for food rations [32].

It is straightforward to see that $\mathfrak{D}^{\text{i.i.d.}} \subsetneq \mathfrak{D}^{\text{adv}} \subsetneq \mathfrak{D}^{\text{all}}$ and $\mathfrak{D}^{\text{i.i.d.}} \subsetneq \mathfrak{D}^{\text{r.o.}} \subsetneq \mathfrak{D}^{\text{all}}$.

2.1 Threshold Algorithm

In this subsection, we show that any optimal online algorithm can be represented as a threshold algorithm. A threshold algorithm is an algorithm $h = (f_i)_{i=1}^n$ specified by a sequence of functions $f_i \colon [0,1]^i \times \mathrm{Inj}([i],[n]) \to \{0,1\}$ for each $i \in [n]$, such that $f_i(\boldsymbol{x}_{\le i}, \tau_{\le i})$ is 1 if $x_i > \theta_i(\tau_{\le i})$ and 0 if $x_i < \theta_i(\tau_{\le i})$, where $\theta_i \colon \mathrm{Inj}([i],[n]) \to [0,1] \cup \{\infty\}$ is a threshold function that depends only on the order of first i values $\tau_{\le i}$. If $x_i = \theta_i(\tau_{\le i})$, the value of $f_i(\boldsymbol{x}_{\le i}, \tau_{\le i})$ can be either 0 or 1. Thus, the function f_i can be represented as

$$f_i(\boldsymbol{x}_{\le i}, \tau_{\le i}) = \mathbb{1}_{\{x_i \ge \theta_i(\tau_{\le i})\}} \quad \text{or} \quad \mathbb{1}_{\{x_i > \theta_i(\tau_{\le i})\}}.$$

Note that the threshold function θ_i can take the value ∞, which means that the algorithm never accepts ith variable regardless of its value.

Lemma 1. *Suppose that the profit function p satisfies properties (i) and (ii), and*

$$\begin{aligned} &\text{for each } i \in [n], \text{there exists a function } q_i \colon [0,1]^{n-i+1} \times \{i, i+1, \dots, n+1\} \to \mathbb{R} \text{ such that} \\ &\quad q_i(\boldsymbol{x}_{\ge i}, j) = p(\boldsymbol{x}, j) - p(\boldsymbol{x}, i) \text{ holds for any } j \in \{i, i+1, \dots, n+1\} \text{ and } \boldsymbol{x} \in [0,1]^n. \end{aligned} \tag{2}$$

Then, any optimal stopping problem instance $(\mathbf{D}, p)$ admits a threshold algorithm $h^* = (f_i^*)_{i=1}^n$ such that $\mathrm{Opt}_{\text{online}}(\mathbf{D}, p) = h(\mathbf{D}, p)$.

The proof is deferred to Section A.1. In the proof, we show a setting of a threshold $\theta_i(\tau_{\le i})$ using the assumption (2).

We remark that the threshold value $\theta_i(\tau_{\le i})$ generally depends not only on the set $\{\tau(1), \dots, \tau(i-1)\}$ but also on the specific order $\tau_{\le i-1}$ of the first $i-1$ variables. This is because the conditional distribution of the remaining permutations may depend on the observed order so far. However, when Π is the uniform distribution over all permutations (i.e., the random order model), the conditional distribution is independent of the observed order, and thus it suffices for the threshold to depend only on the set $\{\tau(1), \dots, \tau(i-1)\}$ and $\tau(i)$.

It is not difficult to see that the profit function p for expected reward, last success, and ski-rental satisfies the assumption (2).

Corollary 1. *If the profit function p is given by expected reward, last success, or ski-rental, then there exists a threshold algorithm that is an optimal online algorithm.*

Unfortunately, the profit function p for best choice $p(\boldsymbol{x}, i) = \mathbb{1}_{\{x_i = \max_{j \in [n]} x_j\}}$ does not satisfy the assumption (2). However, we can construct an optimal online algorithm that is almost a threshold algorithm. If $x_i \ge \max_{j \in [i-1]} x_j$, then $p(\boldsymbol{x}, j) - p(\boldsymbol{x}, i)$ depends on $\boldsymbol{x}_{\ge i}$ and j, and hence we can choose a function q_i with $q_i(\boldsymbol{x}_{\ge i}, j) = p(\boldsymbol{x}, j) - p(\boldsymbol{x}, i)$ for any $j \ge i$. We can set a threshold $\theta_i(\tau_{\le i})$ for such a case (see the proof of Lemma 1 in Section A.1 for details). If $x_i < \max_{j \in [i-1]} x_j$, then there is no loss in rejecting the ith variable. Thus, the following claim holds.

Corollary 2. *If the profit function p is given by best choice, then there exists an optimal online algorithm $h = (f_i)_{i=1}^n$ and a threshold algorithm $h' = (f'_i)_{i=1}^n$ such that*

$$f_i(\boldsymbol{x}_{\le i}, \tau_{\le i}) = \begin{cases} f'_i(\boldsymbol{x}_{\le i}, \tau_{\le i}) & \text{if } x_i \ge \max_{j\in[i-1]} x_j, \\ 0 & \text{if } x_i < \max_{j\in[i-1]} x_j. \end{cases}$$

We call the function defined in this corollary as an *essentially threshold* algorithm. In the following sections, we focus on (essentially) threshold algorithms in analysis of the regret.

3 Uniform Law of Large Numbers

In this section, we introduce key results on the *uniform law of large numbers*, which will be used to derive the regret bound for the repeated optimal stopping problem. The uniform law of large numbers provides a way to control the deviation of empirical averages from their expected values uniformly over a class of functions, which corresponds to online algorithms at each step in our context. We begin by defining the Rademacher complexity, which quantifies the richness of a function class. For t samples $\mathbb{x}_t = (\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(t)})$ of random vectors and a function class $\tilde{\mathcal{H}}$, we denote

$$\tilde{\mathcal{H}} \circ \mathbb{x}_t \coloneqq \{(h(\boldsymbol{x}^{(1)}), \dots, h(\boldsymbol{x}^{(t)})) \mid h \in \tilde{\mathcal{H}}\}.$$

Definition 1. *Let A be a subset of $\mathbb{R}^t$. The* Rademacher complexity *of A is defined as*

$$\mathrm{Rad}(A) \coloneqq \mathbb{E}_\sigma \left[\frac{1}{t} \cdot \sup_{(a_1,\dots,a_t)\in A} \left| \sum_{s=1}^t \sigma_s a_s \right| \right],$$

where $\sigma_1, \dots, \sigma_t$ are independent Rademacher random variables (i.e., each σ_s takes values $+1$ or -1 with probability $1/2$).

Let $\tilde{\mathcal{H}}$ be a class of functions. The empirical Rademacher complexity *of $\tilde{\mathcal{H}}$ with respect to t samples $\mathbb{x}_t = (\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(t)})$ of random vectors is defined as*

$$\mathrm{Rad}_{\mathbb{x}_t}(\tilde{\mathcal{H}}) \coloneqq \mathrm{Rad}(\tilde{\mathcal{H}} \circ \mathbb{x}_t) = \mathbb{E}_\sigma \left[\frac{1}{t} \sup_{h\in\tilde{\mathcal{H}}} \left| \sum_{s=1}^t \sigma_s h(\boldsymbol{x}^{(s)}) \right| \right].$$

Furthermore, the Rademacher complexity *of $\tilde{\mathcal{H}}$ with respect to t i.i.d. samples from distribution $\mathbf{D}$ is defined as*

$$\mathrm{Rad}_{t,\mathbf{D}}(\tilde{\mathcal{H}}) \coloneqq \mathbb{E}_{\mathbf{X}^{(1)},\dots,\mathbf{X}^{(t)} \overset{i.i.d.}{\sim} \mathbf{D}} \left[\mathrm{Rad}_{(\mathbf{X}^{(1)},\dots,\mathbf{X}^{(t)})}(\tilde{\mathcal{H}}) \right].$$

The following lemma is used to upper bound the Rademacher complexity. This is an immediate consequence of Massart's Lemma [33, Lemma 5.2] and the proof is given in Section A.2.

Lemma 2. *For any finite subset A of $\mathbb{R}^t$, letting $r \coloneqq \max_{a\in A} \|a\|_2$, we have*

$$\mathrm{Rad}(A) \le \frac{r\sqrt{2\ln(2|A|)}}{t}.$$

Theorem 3 (Uniform Law of Large Numbers [39, Theorem 4.10])**.** *Let $\tilde{\mathcal{H}}$ be a function class such that $\|h\|_\infty \le B$ for all $h \in \tilde{\mathcal{H}}$. For any $\eta \ge 0$ and positive integer t, we have*

$$\mathbf{Pr}_{\mathbf{X}^{(1)},\dots,\mathbf{X}^{(t)} \overset{\text{i.i.d.}}{\sim} \mathbf{D}} \left[\sup_{h \in \tilde{\mathcal{H}}} \left| \mathbb{E}_{\mathbf{X} \sim \mathbf{D}}[h(\mathbf{X})] - \frac{1}{t} \sum_{s=1}^{t} h(\mathbf{X}^{(s)}) \right| > 2\mathrm{Rad}_{t,\mathbf{D}}(\tilde{\mathcal{H}}) + \eta \right] \le 2 \exp\left(-\frac{t\eta^2}{2B^2} \right).$$

By this theorem, the gap between the expected profits of optimal online algorithms for $\mathbf{D}$ and the empirical distribution is bounded by using the Rademacher complexity and a constant η with high probability.

4 General Result

In this section, we present a general result on the repeated optimal stopping problem $(\mathfrak{D}, p; T)$. Our framework provides a unified method for constructing algorithms that guarantee a competitive ratio in each round, while also achieving sublinear regret over multiple rounds.

As a high-level idea, we start from algorithms g_t that make decisions based on the first $t-1$ samples and guarantee a competitive ratio, as studied in the context of online optimization with samples. However, applying g_t in each round t does not necessarily achieve a regret bound, so we need to construct another sequence of algorithms h_t that can ensure a bounded regret. We will construct such algorithms h_t by utilizing the uniform law of large numbers and, when a more effective algorithm is known for the specific problem setting, by employing it accordingly.

In each round t, if we could select the algorithm with a higher expected reward between g_t and h_t, we would get both the competitive ratio and regret guarantees. However, since the underlying distribution $\mathbf{D}$ is unknown, we cannot evaluate expectations directly. To overcome this, we compare their performance using the hold-out method, which splits the observed samples into training and test sets. Specifically, in round t, we use $\zeta(t) - 1$ samples for training and employ either $g_{\zeta(t)}$ or $h_{\zeta(t)}$. We will set the partition sizes to be basically the same, but use the first few constant samples entirely for training to guarantee a good competitive ratio.

The challenging part is that, to preserve the competitive ratio of $g_{\zeta(t)}$, we cannot allow switching rules that lead to worse performance than the true expectation of $g_{\zeta(t)}$. On the other hand, if we rely too much on $g_{\zeta(t)}$, we cannot bound the regret. Nevertheless, we show that with a proper design, it is possible to achieve both guarantees, depending on the complexity of $(\mathbf{D}, p)$.

Remark 1. *By the uniform law of large numbers (Theorem 3), it is possible to evaluate the performance of g_t and h_t even when all samples are used as training data. This allows us to achieve a competitive ratio no worse than that of g_t in every round t. However, to obtain a meaningful regret bound under this approach, g_t needs to belong to a small class algorithms, such as (essentially) threshold algorithms. In contrast, our method based on the hold-out approach does not require such restrictions on g_t.*

4.1 Switching rule

For each $t \in [T]$, assume we have an algorithm h_t such that there exist positive reals $\varepsilon_1(t)$ and $\delta_1(t)$ satisfying the following: h_t is an $\varepsilon_1(t)$-additive approximation of $\mathrm{Opt}_{\mathrm{online}}$ with probability at least $1 - \delta_1(t)$; that is,

$$\mathbf{Pr}_{(\mathbf{X}^{(1)},\pi^{(1)}),\dots,(\mathbf{X}^{(t-1)},\pi^{(t-1)}) \overset{\text{i.i.d.}}{\sim} \mathbf{D}} \left[h_t[\mathbb{X}_{t-1}, \mathbb{T}_{t-1}](\mathbf{D}) \ge \mathrm{Opt}_{\mathrm{online}}(\mathbf{D}) - \varepsilon_1(t) \right] \ge 1 - \delta_1(t). \tag{3}$$

We will construct such algorithms in the next subsection. The parameters $\varepsilon_1(t)$ and $\delta_1(t)$ are directly related to the *sample complexity* of $(\mathbf{D}, p)$. The sample complexity of $(\mathbf{D}, p)$ is defined as the minimum number of i.i.d. samples required to find an ε-additive approximation of $\mathrm{Opt}_{\mathrm{online}}$ with probability at least $1-\delta$. The parameters $\varepsilon_1(t)$ and $\delta_1(t)$ correspond to the accuracy and confidence levels achieved when the sample complexity of $(\mathcal{D}, p)$ is t.

When both $\varepsilon_1(\zeta(t))$ and $\delta_1(\zeta(t))$ are sufficiently small, applying algorithm $h_{\zeta(t)}$ in each round achieves sublinear regret. In each round $t \in [T]$, we choose between executing $h_{\zeta(t)}$ and $g_{\zeta(t)}$ based on their estimated performance.

When an algorithm h depends on samples drawn from $\mathbf{D}$ in the first $t-1$ rounds, we sometimes write $h[\mathbb{X}_{t-1}, \mathbb{w}_{t-1}]$ instead of h to make its dependence on $(\mathbb{X}_{t-1}, \mathbb{w}_{t-1})$ explicit. In each round t with $\zeta(t) < t$ (i.e., there is at least one test sample available), we estimate the expected reward $h[\mathbb{X}_{\zeta(t)-1}, \mathbb{w}_{\zeta(t)-1}](\mathbf{D})$ of algorithm h, which depends on the first $\zeta(t)-1$ samples, as

$$\hat{h}(\mathbb{X}_{t-1}, \mathbb{w}_{t-1}; \zeta(t)) \coloneqq \frac{1}{t-\zeta(t)} \sum_{s=\zeta(t)}^{t-1} h[\mathbb{X}_{\zeta(t)-1}, \mathbb{w}_{\zeta(t)-1}](\mathbf{X}^{(s)}, \pi^{(s)}; p). \tag{4}$$

This estimate is simply the arithmetic mean of the rewards obtained by h on the test samples. We emphasize that this partitioning ensures the independence between the algorithm parameters and the test samples, so that the random variables $h[\mathbb{X}_{\zeta(t)-1}, \mathbb{w}_{\zeta(t)-1}](\mathbf{X}^{(s)}, \pi^{(s)}; p)$ for $s = \zeta(t), \zeta(t)+1, \ldots, t-1$ are mutually independent. Consequently, Hoeffding's inequality can be applied. Since (4) is the mean of $(t-\zeta(t))$ independent random variables, each taking values in $[0, B]$, for any $\eta > 0$ we have

$$\mathbf{Pr}\left[\left|h[\mathbb{X}_{\zeta(t)-1}, \mathbb{w}_{\zeta(t)-1}](\mathbf{D}) - \hat{h}(\mathbb{X}_{t-1}, \mathbb{w}_{t-1}; \zeta(t))\right| > \eta\right] \leq \delta_0(t-\zeta(t), \eta) \tag{5}$$

where

$$\delta_0(t-\zeta(t), \eta) \coloneqq 2 \cdot \exp\left(-\frac{2(t-\zeta(t))\eta^2}{B^2}\right).$$

For each t, define $\mathcal{E}_t$ as the event where all of the following three conditions hold simultaneously:

$$\mathcal{E}_t : \quad \begin{aligned} |g_{\zeta(t)}[\mathbb{X}_{\zeta(t)-1}, \mathbb{w}_{\zeta(t)-1}](\mathbf{D}) - \hat{g}_{\zeta(t)}(\mathbb{X}_{t-1}, \mathbb{w}_{t-1}; \zeta(t))| &\leq \varepsilon(t), \\ |h_{\zeta(t)}[\mathbb{X}_{\zeta(t)-1}, \mathbb{w}_{\zeta(t)-1}](\mathbf{D}) - \hat{h}_{\zeta(t)}(\mathbb{X}_{t-1}, \mathbb{w}_{t-1}; \zeta(t))| &\leq \varepsilon(t), \\ |\mathrm{Opt}_{\mathrm{online}}(\mathbf{D}) - h_{\zeta(t)}[\mathbb{X}_{\zeta(t)-1}, \mathbb{w}_{\zeta(t)-1}](\mathbf{D})| &\leq \varepsilon(t), \end{aligned} \tag{6}$$

where

$$\varepsilon(t) \coloneqq \varepsilon_1(\zeta(t)).$$

The event $\mathcal{E}_t$ indicates the success in estimating the expected reward and approximating the optimal online algorithm. By (5) and (3), and the union bound, we can see that $\mathcal{E}_t$ occurs with probability of at least $1-\delta(t)$, where

$$\delta(t) \coloneqq 2\delta_0\big(t-\zeta(t), \varepsilon_1(\zeta(t))\big) + \delta_1\big(\zeta(t)\big).$$

Next, define another event $\mathcal{C}_t$ as

$$\mathcal{C}_t : \quad \hat{g}_{\zeta(t)}(\mathbb{X}_{t-1}, \mathbb{w}_{t-1}; \zeta(t)) + \varepsilon(t) + B\delta(t) \leq (1-\delta(t))\big(\hat{h}_{\zeta(t)}(\mathbb{X}_{t-1}, \mathbb{w}_{t-1}; \zeta(t)) - \varepsilon(t)\big).$$

The event $\mathcal{C}_t$ indicates that the estimated reward of $h_{\zeta(t)}$ is sufficiently greater than that of $g_{\zeta(t)}$. If $t = \zeta(t)$, the estimated values are not well-defined, but we regard the event $\mathcal{C}_t$ as not having occurred.

For each round t, we define h_t^* as the algorithm that uses $h_{\zeta(t)}[\mathbb{X}_{\zeta(t)-1}, \mathbb{w}_{\zeta(t)-1}]$ when the event $\mathcal{C}_t$ occurs and $g_{\zeta(t)}[\mathbb{X}_{\zeta(t)-1}, \mathbb{w}_{\zeta(t)-1}]$ otherwise. The next theorem shows that h_t^* achieves at least the expected performance of g_t and the expected regret is bounded.

Theorem 4. *For a round $t \in [T]$ with $\zeta(t) < t$, suppose that $h_{\zeta(t)}$ satisfies* (3) *and Then, the following statements hold:*

1. $\mathbb{E}\left[h_t^*(\mathbf{D})\right] \geq \mathbb{E}\left[g_{\zeta(t)}(\mathbf{D})\right]$ *(i.e., h_t^* achieves at least the expected performance of $g_{\zeta(t)}$),*
2. $\mathbb{E}\left[\mathrm{Opt}_{\mathrm{online}}(\mathbf{D}) - h_t^*(\mathbf{D})\right] \leq 5\varepsilon(t) + 4B\delta(t)$ *(i.e., the expected regret per round is bounded).*

Proof. By the definition of h_t^*, we have

$$\mathbb{E}\left[h_t^*(\mathbf{D})\right] = \mathbf{Pr}\left[\neg\mathcal{C}_t\right] \cdot \mathbb{E}\left[g_{\zeta(t)}(\mathbf{D}) \;\middle|\; \neg\mathcal{C}_t\right] + \mathbf{Pr}\left[\mathcal{C}_t\right] \cdot \mathbb{E}\left[h_{\zeta(t)}(\mathbf{D}) \;\middle|\; \mathcal{C}_t\right].$$

By the definition of $\mathcal{C}_t$ and $\mathcal{E}_t$, it follows that

$$\begin{aligned}
&\mathbb{E}\left[h_{\zeta(t)}(\mathbf{D}) \mid \mathcal{C}_t\right] \\
&\quad \geq \mathbf{Pr}\left[\mathcal{E}_t\right] \cdot \mathbb{E}\left[h_{\zeta(t)}(\mathbf{D}) \;\middle|\; \mathcal{C}_t, \mathcal{E}_t\right] \\
&\quad \geq (1-\delta(t)) \cdot \mathbb{E}\left[h_{\zeta(t)}(\mathbf{D}) \;\middle|\; \mathcal{C}_t, \mathcal{E}_t\right] && (\mathbf{Pr}\left[\mathcal{E}_t\right] \geq 1-\delta(t)) \\
&\quad \geq (1-\delta(t)) \cdot \mathbb{E}\left[\hat{h}_{\zeta(t)}(\mathbb{X}_{t-1}, \mathbb{w}_{t-1}; \zeta(t)) - \varepsilon(t) \;\middle|\; \mathcal{C}_t, \mathcal{E}_t\right] && (\text{by } \mathcal{E}_t) \\
&\quad \geq \mathbb{E}\left[\hat{g}_{\zeta(t)}(\mathbb{X}_{t-1}, \mathbb{w}_{t-1}; \zeta(t)) + \varepsilon(t) + B\delta(t) \;\middle|\; \mathcal{C}_t, \mathcal{E}_t\right] && (\text{by } \mathcal{C}_t) \\
&\quad \geq \mathbb{E}\left[g_{\zeta(t)}(\mathbf{D}) \;\middle|\; \mathcal{C}_t, \mathcal{E}_t\right] + B\delta(t) && (\text{by } \mathcal{E}_t) \\
&\quad \geq \mathbf{Pr}\left[\mathcal{E}_t\right] \cdot \mathbb{E}\left[g_{\zeta(t)}(\mathbf{D}) \;\middle|\; \mathcal{C}_t, \mathcal{E}_t\right] + \mathbf{Pr}\left[\neg\mathcal{E}_t\right] \cdot \mathbb{E}\left[g_{\zeta(t)}(\mathbf{D}) \;\middle|\; \mathcal{C}_t, \neg\mathcal{E}_t\right] \\
&\quad = \mathbb{E}\left[g_{\zeta(t)}(\mathbf{D}) \;\middle|\; \mathcal{C}_t\right],
\end{aligned}$$

where the last inequality holds because $\mathbf{Pr}\left[\mathcal{E}_t\right] \leq 1$, $\mathbf{Pr}\left[\neg\mathcal{E}_t\right] \leq \delta(t)$, and $g_{\zeta(t)}(\mathbf{D}) \leq B$. Thus, we have

$$\begin{aligned}
\mathbb{E}[h_t^*(\mathbf{D})] &= \mathbf{Pr}\left[\neg\mathcal{C}_t\right] \cdot \mathbb{E}\left[g_{\zeta(t)}(\mathbf{D}) \;\middle|\; \neg\mathcal{C}_t\right] + \mathbf{Pr}\left[\mathcal{C}_t\right] \cdot \mathbb{E}\left[h_{\zeta(t)}(\mathbf{D}) \;\middle|\; \mathcal{C}_t\right] \\
&\geq \mathbf{Pr}\left[\neg\mathcal{C}_t\right] \cdot \mathbb{E}\left[g_{\zeta(t)}(\mathbf{D}) \;\middle|\; \neg\mathcal{C}_t\right] + \mathbf{Pr}\left[\mathcal{C}_t\right] \cdot \mathbb{E}\left[g_{\zeta(t)}(\mathbf{D}) \;\middle|\; \mathcal{C}_t\right] = \mathbb{E}[g_{\zeta(t)}(\mathbf{D})].
\end{aligned}$$

Next, we bound the expected regret. We have

$$\begin{aligned}
&\mathrm{Opt}_{\mathrm{online}}(\mathbf{D}) - \mathbb{E}\left[h_t^*(\mathbf{D})\right] \\
&\quad = \mathrm{Opt}_{\mathrm{online}}(\mathbf{D}) - \mathbf{Pr}\left[\neg\mathcal{C}_t\right] \cdot \mathbb{E}\left[g_{\zeta(t)}(\mathbf{D}) \;\middle|\; \neg\mathcal{C}_t\right] - \mathbf{Pr}\left[\mathcal{C}_t\right] \cdot \mathbb{E}\left[h_{\zeta(t)}(\mathbf{D}) \;\middle|\; \mathcal{C}_t\right] \\
&\quad = \mathrm{Opt}_{\mathrm{online}}(\mathbf{D}) - \mathbf{Pr}\left[\neg\mathcal{C}_t\right] \cdot \mathbb{E}\left[g_{\zeta(t)}(\mathbf{D}) \;\middle|\; \neg\mathcal{C}_t\right] - \left(\mathbb{E}\left[h_{\zeta(t)}(\mathbf{D})\right] - \mathbf{Pr}\left[\neg\mathcal{C}_t\right] \cdot \mathbb{E}\left[h_{\zeta(t)}(\mathbf{D}) \;\middle|\; \neg\mathcal{C}_t\right]\right) \\
&\quad = \mathbf{Pr}\left[\neg\mathcal{C}_t\right] \cdot \mathbb{E}\left[h_{\zeta(t)}(\mathbf{D}) - g_{\zeta(t)}(\mathbf{D}) \;\middle|\; \neg\mathcal{C}_t\right] + \mathbb{E}\left[\mathrm{Opt}_{\mathrm{online}}(\mathbf{D}) - h_{\zeta(t)}(\mathbf{D})\right] \\
&\quad \leq \mathbb{E}\left[h_{\zeta(t)}(\mathbf{D}) - g_{\zeta(t)}(\mathbf{D}) \;\middle|\; \neg\mathcal{C}_t\right] + \mathbb{E}\left[\mathrm{Opt}_{\mathrm{online}}(\mathbf{D}) - h_{\zeta(t)}(\mathbf{D})\right].
\end{aligned}$$

Here, the term $\mathbb{E}\left[\mathrm{Opt}_{\mathrm{online}}(\mathbf{D}) - h_{\zeta(t)}(\mathbf{D})\right]$ can be bounded as follows:

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{Opt}_{\mathrm{online}}(\mathbf{D}) - h_{\zeta(t)}(\mathbf{D})\right] \\
&\quad = \mathbf{Pr}\left[\mathcal{E}_t\right] \cdot \mathbb{E}\left[\mathrm{Opt}_{\mathrm{online}}(\mathbf{D}) - h_{\zeta(t)}(\mathbf{D}) \;\middle|\; \mathcal{E}_t\right] + \mathbf{Pr}\left[\neg\mathcal{E}_t\right] \cdot \mathbb{E}\left[\mathrm{Opt}_{\mathrm{online}}(\mathbf{D}) - h_{\zeta(t)}(\mathbf{D}) \;\middle|\; \neg\mathcal{E}_t\right] \\
&\quad \leq 1 \cdot \varepsilon(t) + \delta(t) \cdot B = \varepsilon(t) + B\delta(t),
\end{aligned}$$

where the inequality holds because $\mathbf{Pr}\left[\mathcal{E}_t\right] \leq 1$, $\mathbb{E}\left[\mathrm{Opt}_{\mathrm{online}}(\mathbf{D}) - h_{\zeta(t)}(\mathbf{D}) \;\middle|\; \mathcal{E}_t\right] \leq \varepsilon(t)$ by (6), $\mathrm{Opt}_{\mathrm{online}}(\mathbf{D}) \leq B$, and $\mathbf{Pr}\left[\neg\mathcal{E}_t\right] \leq \delta(t)$. For readability, we abbreviate $\hat{g}_{\zeta(t)}(\mathbb{X}_{t-1}, \mathbb{\pi}_{t-1}; \zeta(t))$ and $\hat{h}_{\zeta(t)}(\mathbb{X}_{t-1}, \mathbb{\pi}_{t-1}; \zeta(t))$ as $\hat{g}_{\zeta(t)}$ and $\hat{h}_{\zeta(t)}$ respectively. Then, the term $\mathbb{E}\left[h_{\zeta(t)}(\mathbf{D}) - g_{\zeta(t)}(\mathbf{D}) \;\middle|\; \neg\mathcal{C}_t\right]$ is bounded as

$$\begin{aligned}
&\mathbb{E}\left[h_{\zeta(t)}(\mathbf{D}) - g_{\zeta(t)}(\mathbf{D}) \;\middle|\; \neg\mathcal{C}_t\right] \\
&\quad = \mathbf{Pr}\left[\mathcal{E}_t\right] \cdot \mathbb{E}\left[h_{\zeta(t)}(\mathbf{D}) - g_{\zeta(t)}(\mathbf{D}) \;\middle|\; \neg\mathcal{C}_t, \mathcal{E}_t\right] + \mathbf{Pr}\left[\neg\mathcal{E}_t\right] \cdot \mathbb{E}\left[h_{\zeta(t)}(\mathbf{D}) - g_{\zeta(t)}(\mathbf{D}) \;\middle|\; \neg\mathcal{C}_t, \neg\mathcal{E}_t\right] \\
&\quad \leq 1 \cdot \mathbb{E}\left[h_{\zeta(t)}(\mathbf{D}) - g_{\zeta(t)}(\mathbf{D}) \;\middle|\; \neg\mathcal{C}_t, \mathcal{E}_t\right] + \delta(t) \cdot B \\
&\quad \leq \mathbb{E}\left[\hat{h}_{\zeta(t)} - \hat{g}_{\zeta(t)} \;\middle|\; \neg\mathcal{C}_t, \mathcal{E}_t\right] + B\delta(t) + 2\varepsilon(t) \\
&\quad = \mathbb{E}\left[(\hat{h}_{\zeta(t)} - \varepsilon(t)) - (\hat{g}_{\zeta(t)} + \varepsilon(t) + B\delta(t)) \;\middle|\; \neg\mathcal{C}_t, \mathcal{E}_t\right] + 2B\delta(t) + 4\varepsilon(t) \\
&\quad \leq \mathbb{E}\left[(\hat{h}_{\zeta(t)} - \varepsilon(t)) - (1 - \delta(t))(\hat{h}_{\zeta(t)} - \varepsilon(t)) \;\middle|\; \neg\mathcal{C}_t, \mathcal{E}_t\right] + 2B\delta(t) + 4\varepsilon(t) \\
&\quad = \delta(t) \cdot \mathbb{E}\left[\hat{h}_{\zeta(t)} - \varepsilon(t) \;\middle|\; \neg\mathcal{C}_t, \mathcal{E}_t\right] + 2B\delta(t) + 4\varepsilon(t) \\
&\quad \leq \delta(t) \cdot B + 2B\delta(t) + 4\varepsilon(t) = 3B\delta(t) + 4\varepsilon(t).
\end{aligned}$$

By combining these, we obtain

$$\mathrm{Opt}_{\mathrm{online}}(\mathbf{D}) - \mathbb{E}\left[h_t^*(\mathbf{D})\right] \leq (3B\delta(t) + 4\varepsilon(t)) + (\varepsilon(t) + B\delta(t)) = 5\varepsilon(t) + 4B\delta(t).$$

This completes the proof. □

Theorem 4 does not provide a method for finding h_t that satisfies equation (3). Therefore, next we will provide a general method for finding h_t that satisfies (3) and a specific functional form for $\varepsilon_1(t)$ and $\delta_1(t)$.

4.2 Regret bound

Let $\tilde{\mathcal{H}}$ be a set of algorithms such that, for any $\mathbf{D} \in \mathfrak{D}$, $\mathrm{Opt}_{\mathrm{online}}(\mathbf{D}) = h(\mathbf{D}; p)$ for some $h \in \tilde{\mathcal{H}}$. For example, if p satisfies the assumption (2) in Lemma 1, then $\tilde{\mathcal{H}}$ can be chosen as the set of threshold algorithms. Also, if p is a profit function of best choice, then $\tilde{\mathcal{H}}$ can be chosen as the set of essentially threshold algorithms. Let $\hat{\mathbf{D}}^{(t)}$ be the empirical distribution where $(\mathbf{X}^{(s)}, \pi^{(s)})$ is sampled with probability $1/(t-1)$ for each $s \in [t-1]$. Thus, for any algorithm $h \in \tilde{\mathcal{H}}$, the expected profit of h for the empirical distribution $\hat{\mathbf{D}}^{(t)}$ at round t is defined as

$$h(\hat{\mathbf{D}}^{(t)}; p) = \frac{1}{t-1} \sum_{s=1}^{t-1} h(\mathbf{X}^{(s)}, \pi^{(s)}; p).$$

For a distribution $\mathbf{D}'$ that is either $\mathbf{D}$ or $\hat{\mathbf{D}}^{(t)}$, we denote by $h_{\mathbf{D}'}$ an optimal algorithm in $\tilde{\mathcal{H}}$ for $\mathbf{D}'$, i.e., $h_{\mathbf{D}'} \in \arg\max_{h \in \tilde{\mathcal{H}}} h(\mathbf{D}'; p)$. It should be noted that $h_{\mathbf{D}'}$ is not necessarily an optimal

online algorithm for $\mathbf{D}'$ because an empirical distribution may not be in $\mathfrak{D}$. Moreover, the set $\arg\max_{h\in\tilde{\mathcal{H}}} h(\mathbf{D}';p)$ is nonempty. This is because, if $\mathbf{D}'=\mathbf{D}$, then $\mathrm{Opt}_{\mathrm{online}}(\mathbf{D})=h(\mathbf{D};p)$ for some $h\in\tilde{\mathcal{H}}$ by the assumption. If $\mathbf{D}'=\hat{\mathbf{D}}^{(t)}$, then we can partition $\tilde{\mathcal{H}}$ into at most $(n+1)^{t-1}$ classes according to the acceptance index of each realization in the first $t-1$ rounds. Because this yields only finitely many possibilities, an optimal algorithm in $\tilde{\mathcal{H}}$ must exist. We will omit the parameter p when it is clear from the context.

For any positive integer t, let M_{t-1} be a positive integer such that

$$M_{t-1}=\sup_{(\boldsymbol{x}^{(1)},\tau^{(1)}),\dots,(\boldsymbol{x}^{(t-1)},\tau^{(t-1)})}\left|\left\{\big(h(\boldsymbol{x}^{(1)},\tau^{(1)};p),\dots,h(\boldsymbol{x}^{(t-1)},\tau^{(t-1)};p)\big)\;\middle|\;h\in\tilde{\mathcal{H}}\right\}\right| \tag{7}$$

where the supremum is taken over all choices of $\boldsymbol{x}^{(1)},\dots,\boldsymbol{x}^{(t-1)}$ in $[0,1]$ and $\tau^{(1)},\dots,\tau^{(t-1)}$ in the set of permutations that occur with positive probability. By utilizing the uniform law of large numbers (Theorem 3), we can select $\varepsilon_1(t)$ and $\delta_1(t)$ such that the assumption (3) holds as follows.

Lemma 3. *For each integer $t\geq 2$ and any $M\geq M_{t-1}$, the algorithm $h_{\hat{\mathbf{D}}^{(t)}}$ and the following settings of $\varepsilon_1(t)$ and $\delta_1(t)$ satisfy the assumption* (3)*:*

$$\varepsilon_1(t)\coloneqq 6B\sqrt{\frac{2\ln(2M)}{t-1}}\ \text{ and }\ \delta_1(t)\coloneqq\frac{1}{M}.$$

Proof. Since we assume that the range of the profit is bounded by $[0,B]$, we have

$$\left\|\big(h(\boldsymbol{x}^{(1)},\tau^{(1)};p),\dots,h(\boldsymbol{x}^{(t-1)},\tau^{(t-1)};p)\big)\right\|_2\leq B\sqrt{t-1}$$

for any $(\boldsymbol{x}^{(1)},\tau^{(1)}),\dots,(\boldsymbol{x}^{(t-1)},\tau^{(t-1)})$ in the support of $\mathbf{D}$. By setting $\eta=B\sqrt{\frac{2\ln(2M)}{t-1}}$, we have

$$2\mathrm{Rad}_{t-1,\mathbf{D}}(\tilde{\mathcal{H}})+\eta\leq 2\frac{B\sqrt{t-1}\cdot\sqrt{2\ln(2M)}}{t-1}+B\sqrt{\frac{2\ln(2M)}{t-1}}=3B\sqrt{\frac{2\ln(2M)}{t-1}}=\frac{1}{2}\varepsilon_1(t),$$

where we use Lemma 2 to obtain the inequality.

Then, by Theorem 3, we have

$$\begin{aligned}
&\mathbf{Pr}_{(\mathbf{X}^{(1)},\pi^{(1)}),\dots,(\mathbf{X}^{(t-1)},\pi^{(t-1)})\overset{\text{i.i.d.}}{\sim}\mathbf{D}}\left[\sup_{h\in\tilde{\mathcal{H}}}\left|h(\mathbf{D})-h(\hat{\mathbf{D}}^{(t)})\right|>\frac{1}{2}\varepsilon_1(t)\right]\\
&\quad\leq\mathbf{Pr}_{(\mathbf{X}^{(1)},\pi^{(1)}),\dots,(\mathbf{X}^{(t-1)},\pi^{(t-1)})\overset{\text{i.i.d.}}{\sim}\mathbf{D}}\left[\sup_{h\in\tilde{\mathcal{H}}}\left|h(\mathbf{D})-h(\hat{\mathbf{D}}^{(t)})\right|>2\mathrm{Rad}_{t-1,\mathbf{D}}(\tilde{\mathcal{H}})+\eta\right]\\
&\quad\leq 2\exp\left(-\frac{(t-1)\eta^2}{2B^2}\right)=2\exp\left(-\frac{(t-1)\cdot(B^2\cdot 2\ln(2M)/(t-1))}{2\cdot B^2}\right)\\
&\quad=2\exp\left(-\ln(2M)\right)=\frac{1}{M}=\delta_1(t).
\end{aligned}$$

Thus, with probability at least $1-\delta_1(t)$, we obtain $\sup_{h\in\tilde{\mathcal{H}}}\left|h(\mathbf{D})-h(\hat{\mathbf{D}}^{(t)})\right|\leq\frac{1}{2}\varepsilon_1(t)$ for each $h\in\tilde{\mathcal{H}}$. If this holds, then it follows that

$$\begin{aligned}
h_{\hat{\mathbf{D}}^{(t)}}(\mathbf{D})&\geq h_{\hat{\mathbf{D}}^{(t)}}(\hat{\mathbf{D}}^{(t)})-\varepsilon_1(t)/2\geq h_{\mathbf{D}}(\hat{\mathbf{D}}^{(t)})-\varepsilon_1(t)/2\\
&\geq h_{\mathbf{D}}(\mathbf{D})-\varepsilon_1(t)=\mathrm{Opt}_{\mathrm{online}}(\mathbf{D})-\varepsilon_1(t),
\end{aligned}$$

where we use the optimality of $h_{\mathbf{D}}(\hat{\mathbf{D}}^{(t)})$ in the second inequality. Therefore, $h_{\hat{\mathbf{D}}^{(t)}}$ and the above-defined $\varepsilon_1(t), \delta_1(t)$ satisfy the assumption (3). □

In general, M_{t-1} in (7) can be upper bounded by $(n+1)^{t-1}$, since each of the $t-1$ rounds can have $n+1$ different outcomes. However, combining this upper bound with Lemma 3 only yields a regret bound that is linear in the number of rounds T. Thus, we need a more refined upper bound for M_{t-1} to achieve a sublinear regret bound, but in general this is not possible. To handle this, we focus on a specific class of algorithms $\tilde{\mathcal{H}}$.

Lemma 4. *Suppose that $\tilde{\mathcal{H}}$ is the set of threshold algorithms or the set of essentially threshold algorithms. For each $t \in [T]$ and any constant $C \geq 1$, we have*

$$M_{t-1} \leq t^{\sum_{i=1}^{n} |\{\tau_{\leq i}^{(s)} | s \in [t-1]\}|} \leq t^{\min(n|\Pi|, 2n!)} \leq (Ct)^{\min(n|\Pi|, 2n!)},$$

where $|\Pi|$ be the number of distinct permutations that appear with positive probability in $\mathbf{D}$.

Proof. Recall that any threshold algorithm $h \in \tilde{\mathcal{H}}$ can be specified by a sequence of threshold functions $(\theta_1, \ldots, \theta_n)$, where $\theta_i \colon \mathrm{Inj}([i], [n]) \to [0,1] \cup \{\infty\}$ for each $i \in [n]$. For each $i \in [n]$ and $\tau_{\leq i} \in \mathrm{Inj}([i], [n])$, consider the set of observed values denoted by $\chi(\tau_{\leq i}) = \{x_i^{(s)} \mid \tau_{\leq i}^{(s)} = \tau_{\leq i}, s \in [t-1]\}$. The threshold $\theta_i(\tau_{\leq i})$ can be classified into at most $|\chi(\tau_{\leq i})|+1$ types, depending on which of the $|\chi(\tau_{\leq i})|+1$ subintervals of $[0, \infty]$, determined by partitioning it with the observed values, the threshold falls into.

Therefore, in round t, the total number of distinct output patterns over all n steps is at most

$$\begin{aligned}
\prod_{i=1}^{n} \prod_{\tau_{\leq i} \in \mathrm{Inj}([i],[n])} (|\chi(\tau_{\leq i})| + 1) &= \prod_{i=1}^{n} \prod_{\tau_{\leq i} \in \{\tau_{\leq i}^{(s)} | s \in [t-1]\}} (|\chi(\tau_{\leq i})| + 1) \\
&\leq \prod_{i=1}^{n} \prod_{\tau_{\leq i} \in \{\tau_{\leq i}^{(s)} | s \in [t-1]\}} t \\
&= t^{\sum_{i=1}^{n} |\{\tau_{\leq i}^{(s)} | s \in [t-1]\}|}.
\end{aligned}$$

Hence, $M_{t-1} \leq t^{\sum_{i=1}^{n} |\{\tau_{\leq i}^{(s)} | s \in [t-1]\}|}$ if $\tilde{\mathcal{H}}$ is the set of threshold algorithms.

If $\tilde{\mathcal{H}}$ is the set of essentially threshold algorithms, we can similarly show that the number of distinct output patterns is at most $t^{\sum_{i=1}^{n} |\{\tau_{\leq i}^{(s)} | s \in [t-1]\}|}$ since an essentially threshold algorithm is uniquely determined by a threshold algorithm. If $|\Pi|$ is $n!$ (e.g., the random order model and the general order model), we have $\sum_{i=1}^{n} |\{\tau_{\leq i}^{(s)} \mid s \in [t-1]\}| \leq \sum_{i=1}^{n} n!/i! \leq 2n!$. In addition, the number $\max_{i \in [n]} |\{\tau_{\leq i}^{(s)} \mid s \in [t-1]\}| = |\{\tau^{(s)} \mid s \in [t-1]\}|$ is at most $|\Pi|$ and we obtain $\sum_{i=1}^{n} |\{\tau_{\leq i}^{(s)} \mid s \in [t-1]\}| \leq n|\Pi|$. This completes the proof. □

Lemma 4 implies that if $|\Pi|$ is a constant (e.g., the adversarial order model, the i.i.d. model, and the forward-backward model), we can obtain a bound of $M_{t-1} \leq t^n$. If $|\Pi|$ is $n!$ (e.g., the random order model and the general order model), we have $M_{t-1} \leq t^{2n!}$.[3]

By combining Lemmas 3 and 4 with $C = 2$, we can set $\varepsilon_1(t)$ and $\delta_1(t)$ to satisfy the assumption (3).

[3]For the random order model, we may assume that the threshold θ_i only depends on the the set $\{\tau(1), \ldots, \tau(i-1)\}$ and $\tau(i)$. With this assumption, we can obtain a better bound of $M_{t-1} \leq t^{n \cdot 2^{n-1}}$, since $\sum_{i=1}^{n} n \cdot \binom{n-1}{i-1} = n \cdot 2^{n-1}$.

Corollary 3. *Suppose that $\tilde{\mathcal{H}}$ is the set of threshold algorithms or the set of essentially threshold algorithms. Then, $\varepsilon_1(t) = 6B\sqrt{(2\min(n|\Pi|, 2n!)\ln(4t))/(t-1)}$ and $\delta_1(t) = 1/(2t)^{\min(n|\Pi|,2n!)}$ satisfy the assumption (3) for any $t \in [T]$ with $t \geq 2$, where $C \geq 1$ is a constant.*

By combining Theorem 4 and Corollary 3, we obtain our main theorem which provides a general method of constructing a selection rule of algorithms with guarantees of both a competitive ratio in each round and a sublinear regret.

Theorem 5. *Let $(\mathbf{D}, p; T)$ be an instance of the repeated optimal stopping problem $(\mathfrak{D}, p; T)$, and let $\tilde{\mathcal{H}}$ be the set of (essentially) threshold algorithms. Suppose that $\mathrm{Opt}_{\mathrm{online}}(\mathbf{D}; p) = h(\mathbf{D}; p)$ for some $h \in \tilde{\mathcal{H}}$.*

For some constant $t_0 \geq 1$, let $\zeta(t) \coloneqq \min\left\{t, \max\{t_0+1, \lfloor (t+1)/2 \rfloor\right\}$. For each $t \in [T]$, let $g_t[\mathbb{X}_{t-1}, \mathbb{w}_{t-1}]$ be an algorithm for the optimal stopping problem $(\mathbf{D}, p)$ with $t-1$ samples and $h_t[\mathbb{X}_{t-1}, \mathbb{w}_{t-1}]$ be the optimal algorithm among $\tilde{\mathcal{H}}$ with respect to the empirical distribution $\hat{\mathbf{D}}^{(t)}$. Let h_t^* be the algorithm that uses $h_{\zeta(t)}$ when the event $\mathcal{C}_t$ occurs and $g_{\zeta(t)}$ otherwise, where we set

$$\varepsilon_1(t) \coloneqq 6B\sqrt{\frac{2\min(n|\Pi|, 2n!)\ln(4t)}{t-1}} \quad \text{and} \quad \delta_1(t) \coloneqq \frac{1}{2t^{\min(n|\Pi|,2n!)}}.$$

Then, the sequence of online algorithms $(h_t^*)_{t\in[T]}$ satisfies the following:

- *For each $t \in [T]$, the expected profit of h_t^* is at least that of $g_{\zeta(t)}$, i.e.,*

$$\mathbb{E}\left[h_t^*(\mathbf{D})\right] \geq \mathbb{E}\left[g_{\zeta(t)}[\mathbb{X}_{\zeta(t)-1}, \mathbb{w}_{\zeta(t)-1}](\mathbf{D})\right].$$

- *The regret of the algorithm $(h_t^*)_{t\in[T]}$ is $O(B\sqrt{\min(n|\Pi|, n!)T\log T})$.*

Proof. By Theorem 4 and Corollary 3, we have

$$\mathbb{E}\left[h_t^*(\mathbf{D})\right] \geq \mathbb{E}\left[g_{\zeta(t)}[\mathbb{X}_{\zeta(t)-1}, \mathbb{w}_{\zeta(t)-1}](\mathbf{D})\right].$$

Let $\kappa \coloneqq \min(n|\Pi|, 2n!)$ for short. We upper bound $\varepsilon(t)$ and $\delta(t)$ used in the definition of $\mathcal{E}_t$ and $\mathcal{C}_t$ when $t \geq 2t_0 + 1$. Since $\zeta(t) \geq t/2$ and $\ln(4x)/(x-1)$ is monotone decreasing for $x > 1$, we have

$$\varepsilon(t) = \varepsilon_1(\zeta(t)) \leq \varepsilon_1(t/2) = 6B\sqrt{\frac{2\kappa\ln(2t)}{(t-2)/2}} = 12B\sqrt{\frac{\kappa\ln(2t)}{t-2}}.$$

Moreover, by $t/2 \leq \zeta(t) \leq (t+1)/2$ and $\kappa \geq 1$, we have

$$\begin{aligned}
\delta(t) &= 2\delta_0(t-\zeta(t), \varepsilon_1(\zeta(t))) + \delta_1(\zeta(t)) \\
&= 4\exp\left(-\frac{2(t-\zeta(t))}{B^2}\cdot\frac{72B^2\kappa\ln(4\zeta(t))}{\zeta(t)-1}\right) + \frac{1}{(\zeta(t))^\kappa} \\
&\leq 4\exp\left(-\frac{t-1}{B^2}\cdot\frac{72B^2\kappa\ln(4\cdot t/2)}{(t-1)/2}\right) + \frac{1}{(t/2)^\kappa} = \frac{4}{(2t)^{144\kappa}} + \frac{1}{(t/2)^\kappa} \leq \frac{4}{t}.
\end{aligned}$$

By Theorem 4 and Corollary 3, the regret of h_t^* for $(\mathbf{D}, p; T)$ is bounded as follows:

$$\begin{aligned}
\text{Regret} &\le 2t_0 B + \sum_{t=2t_0+1}^{T} (5\varepsilon(t) + 4B\delta(t)) \le 2t_0 B + \sum_{t=2t_0+1}^{T} \left(60B\sqrt{\frac{\kappa \ln(2t)}{t-2}} + \frac{16B}{t} \right) \\
&\le 2t_0 B + 60B\sqrt{\kappa \ln(2T)} \cdot \sum_{t=1}^{T-2} \frac{1}{\sqrt{t}} + 16B \cdot \sum_{t=1}^{T} \frac{1}{t} \\
&\le 2t_0 B + 60B\sqrt{\kappa \ln(2T)} \cdot \left(1 + \int_{t=1}^{T-2} \frac{\mathrm{d}t}{\sqrt{t}} \right) + 16B \cdot \int_{t=1}^{T} \frac{\mathrm{d}t}{t} \\
&= 2t_0 B + 60B\sqrt{\kappa \ln(2T)} \cdot \left(1 + \left[2\sqrt{T-2} - 2 \right] \right) + 16B \cdot [\ln(T) - \ln(1)] \\
&= O(B\sqrt{\kappa T \log T}).
\end{aligned}$$

□

Corollary 4. *Suppose that $\tilde{\mathcal{H}}$ is the set of (essentially) threshold algorithms. If $\mathrm{Opt}_{\mathrm{online}}(\mathbf{D}; p) = h(\mathbf{D}; p)$ for some $h \in \tilde{\mathcal{H}}$ for each $t \in [T]$, then the sequence of algorithms $(h_t^*)_{t\in[T]}$ constructed in Theorem 5 satisfies the following:*

- *For $\mathfrak{D}^{\mathrm{adv}}$ and $\mathfrak{D}^{\mathrm{i.i.d.}}$, the regret bound is $O(B\sqrt{nT \log T})$.*
- *For $\mathfrak{D}^{\mathrm{r.o.}}$ and $\mathfrak{D}^{\mathrm{all}}$, the regret bound is $O(B\sqrt{n!T \log T})$.*

5 Applications

In this section, we demonstrate how Theorem 5 and Corollary 4 can be applied to specific repeated optimal stopping problems $(\mathfrak{D}, p; T)$. We treat repeated versions of the prophet inequality, the prophet secretary problem, the i.i.d. prophet inequality, the adversarial order secretary problem, the random order secretary problem, and the i.i.d. secretary problem.

5.1 Prophet Inequality

We first apply our general framework to the repeated prophet inequality $(\mathfrak{D}^{\mathrm{adv}}, p^{\mathrm{rwd}}; T)$, where the objective is to maximize the expected reward $p^{\mathrm{rwd}}(\boldsymbol{x}, i) = x_i$ and Π is the point-mass distribution on the identity permutation for any $\mathbf{D} = (D_1, \ldots, D_n; \Pi) \in \mathfrak{D}^{\mathrm{adv}}$. In this setting, we construct a sequence $(g_t^{\mathrm{PI}})_{t\in[T]}$ of randomized threshold algorithms as follows:

- Let g_1^{PI} be the randomized algorithm that selects one variable uniformly at random.
- For $t \ge 2$, let $g_t^{\mathrm{PI}}[\mathbb{X}_{t-1}, \mathbb{m}_{t-1}]$ be the single-sample algorithm for the prophet inequality [36], i.e., g_t^{PI} selects the first variable $X_i^{(t)}$ such that $X_i^{(t)} \ge \max_{i' \in [n]} X_{i'}^{(1)}$.

It is easy to see that g_1^{PI} achieves a competitive ratio of $1/n$. Rubinstein et al. [36] proved that the single-sample algorithm g_t^{PI} $(t \ge 2)$ achieves a competitive ratio of $1/2$ for the prophet inequality. Note that, since the algorithms guarantee competitive ratios for the adversarial order, the same competitive ratios hold even for the general order $\mathfrak{D}^{\mathrm{all}}$.

By applying Theorem 5 with $t_0 = 1$, we can construct a sequence of algorithms $(h_t^{\mathrm{PI}})_{t\in[T]}$ for the repeated prophet inequality problem with baseline algorithms $(g_t^{\mathrm{PI}})_{t\in[T]}$. By Corollary 4, these algorithms satisfy the following performance guarantees.

Corollary 5. *For the repeated prophet inequality problem* $(\mathfrak{D}^{\mathrm{adv}}, p^{\mathrm{rwd}}; T)$ *and any* $\mathbf{D} \in \mathfrak{D}^{\mathrm{adv}}$*, the sequence of algorithms* $(h_t^{\mathrm{PI}})_{t\in[T]}$ *achieves the following:*

1. h_1^{PI} *is* $1/n$*-competitive,*
2. h_t^{PI} *is* $1/2$*-competitive for* $t \geq 2$*,*
3. *the regret of* $(h_t^{\mathrm{PI}})_{t\in[T]}$ *is* $O(B\sqrt{nT\log T})$*.*

Additionally, the same algorithms work for the general order prophet inequality problem.

Corollary 6. *For the repeated general order prophet inequality problem* $(\mathfrak{D}^{\mathrm{all}}, p^{\mathrm{rwd}}; T)$ *and any* $\mathbf{D} \in \mathfrak{D}^{\mathrm{all}}$*, the sequence of algorithms* $(h_t^{\mathrm{PI}})_{t\in[T]}$ *achieves the following:*

1. h_1^{PI} *is* $1/n$*-competitive,*
2. h_t^{PI} *is* $1/2$*-competitive for* $t \geq 2$*, and*
3. *the regret of* $(h_t^{\mathrm{PI}})_{t\in[T]}$ *is* $O(B\sqrt{n!T\log T})$*.*

We can improve Corollary 5 by leveraging the result by Jin et al. [24]. They established that the sample complexity of the prophet inequality is at most $(25B^2\ln^2(2e/\delta))/\varepsilon^2$, which implies that, with $t-1$ samples of distribution $\mathbf{D} \in \mathfrak{D}^{\mathrm{adv}}$, we can construct an algorithm $h'_t[\mathbb{X}_{t-1}, \mathbb{m}_{t-1}]$ such that the assumption (3) holds with

$$\varepsilon_1(t) \coloneqq \frac{5B\ln(4et)}{\sqrt{t-1}} \text{ and } \delta_1(t) \coloneqq \frac{1}{2t}. \tag{8}$$

Define $\zeta(t) \coloneqq \min\left\{t, \max\{2, \lfloor (t+1)/2 \rfloor\}\right\}$. For $t \in [T]$, let $h_t^{\mathrm{PI}*}$ be the algorithm that uses $h'_{\zeta(t)}$ when $\mathcal{C}_t$ occurs and $g_{\zeta(t)}^{\mathrm{PI}}$ otherwise, where ε_1 and δ_1 are given by (8). Note that h_t^{PI} and $h_t^{\mathrm{PI}*}$ differ in the setting of ε_1 and δ_1, as well as in how they approximate the optimal online algorithm $\mathrm{Opt}_{\mathrm{online}}$ when the event $\mathcal{C}_t$ occurs.

Since the upper bound on the sample complexity $O((\ln^2(1/\delta))/\varepsilon^2)$ by Jin et al. [24] is near-optimal, the sequence of algorithms $(h_t^{\mathrm{PI}*})_{t\in[T]}$ achieves the best possible competitive ratio at each round and a near-optimal regret.

Theorem 6 (Restatement of Theorem 2)**.** *For the repeated prophet inequality problem* $(\mathfrak{D}^{\mathrm{adv}}, p^{\mathrm{rwd}}; T)$ *and any* $\mathbf{D} \in \mathfrak{D}^{\mathrm{adv}}$*, there exists a sequence of algorithms* $(h_t^{\mathrm{PI}*})_{t\in[T]}$ *that achieves the following:*

1. $h_1^{\mathrm{PI}*}$ *is* $1/n$*-competitive,*
2. $h_t^{\mathrm{PI}*}$ *is* $1/2$*-competitive for* $t \geq 2$*, and*
3. *the regret of* $(h_t^{\mathrm{PI}*})_{t\in[T]}$ *is* $O(B\sqrt{T}\log T)$*.*

Proof. By Theorem 4 and Corollary 3, we have

$$\mathbb{E}\left[h_t^*(\mathbf{D})\right] \geq \mathbb{E}\left[g_{\zeta(t)}[\mathbb{X}_{\zeta(t)-1}, \mathbb{m}_{\zeta(t)-1}](\mathbf{D})\right].$$

This implies that $h_1^{\mathrm{PI}*}$ is $1/n$-competitive and $h_t^{\mathrm{PI}*}$ is $1/2$-competitive for $t \geq 2$.

We now evaluate $\varepsilon(t)$ and $\delta(t)$ for $t \geq 3$. Since $\zeta(t) \geq t/2$, we have

$$\varepsilon(t) = \varepsilon_1(\zeta(t)) \leq \varepsilon_1(t/2) = \frac{5\sqrt{2}B\ln(2et)}{\sqrt{t-2}}.$$

Additionally, since $\zeta(t) \leq (t+1)/2$, we have

$$\begin{aligned}
\delta(t) &= 2\delta_0(t-\zeta(t), \varepsilon_1(\zeta(t))) + \delta_1(\zeta(t)) \\
&\leq 2\delta_0(t-\zeta(t), \varepsilon_1((t+1)/2)) + \delta_1((t+1)/2) \\
&= 4\exp\left(-\frac{2(t-\zeta(t))}{B^2} \cdot \frac{50B^2\ln^2(2e(t+1))}{t-1}\right) + \frac{1}{t+1} \\
&\leq 4\exp\left(-\frac{t-1}{B^2} \cdot \frac{50B^2\ln^2(2e(t+1))}{t-1}\right) + \frac{1}{t} \\
&= \frac{4}{(2e(t+1))^{50\ln(2e(t+1))}} + \frac{1}{t} \leq \frac{2}{t}.
\end{aligned}$$

Thus, we can obtain a regret bound $O(B\sqrt{T}\log T)$ by a similar evaluation as in the proof of Theorem 5. □

5.2 Adversarial Order Secretary Problem (Best Choice Prophet Inequality)

Next, we apply our framework to the repeated adversarial order secretary problem $(\mathfrak{D}^{\mathrm{adv}}, p^{\mathrm{best}}; T)$.

We define a sequence $(g_t^{\mathrm{AS}})_{t\in[T]}$ of randomized threshold algorithms for the adversarial order secretary problem as follows:

- Let g_1^{AS} be the randomized algorithm that selects one variable uniformly at random.
- For $2 \leq t \leq T$, let g_t^{AS} be the single-sample algorithm for the problem [35], i.e., g_t^{AS} selects the first variable X_i such that $X_i \geq \max_{i'\in[n]} X_{i'}^{(1)}$.

For this problem, Nuti and Vondrák [35] proved that the above simple threshold algorithm achieves a competitive ratio of $1/4$, and no algorithm can have a better competitive ratio in the single sample setting. We remark that the algorithm also has the same competitive ratio for the repeated general order secretary problem $(\mathfrak{D}^{\mathrm{all}}, p^{\mathrm{best}}; T)$.

Let $(h_t^{\mathrm{AS}})_{t\in[T]}$ be algorithms as constructed in Theorem 5 with baseline threshold algorithms $(g_t^{\mathrm{AS}})_{t\in[T]}$ when $t_0 = 1$. Then, we have the following corollary by applying Corollary 4.

Corollary 7. *For the repeated adversarial order secretary problem $(\mathfrak{D}^{\mathrm{adv}}, p^{\mathrm{best}}; T)$ and any $\mathbf{D} \in \mathfrak{D}^{\mathrm{adv}}$, the algorithms $(h_t^{\mathrm{AS}})_{t\in[T]}$ achieve the following:*

1. *h_1^{AS} is $1/n$-competitive,*
2. *h_t^{AS} is $1/4$-competitive for $t \geq 2$, and*
3. *the regret of $(h_t^{\mathrm{AS}})_{t\in[T]}$ is $O(\sqrt{nT\log T})$.*

Furthermore, the same algorithms work for the repeated general order secretary problem.

Corollary 8. *For the repeated general order secretary problem* $(\mathfrak{D}^{\mathrm{all}}, p^{\mathrm{best}}; T)$ *and any* $\mathbf{D} \in \mathfrak{D}^{\mathrm{all}}$*, the algorithms* $(h_t^{\mathrm{AS}})_{t\in[T]}$ *achieve*

1. h_1^{AS} *is* $1/n$*-competitive,*
2. h_t^{AS} *is* $1/4$*-competitive for* $t \geq 2$*, and*
3. *the regret of* $(h_t^{\mathrm{AS}})_{t\in[T]}$ *is* $O(\sqrt{n!T\log T})$.

5.3 Prophet Secretary Problem

We apply our general framework to the repeated prophet secretary problem $(\mathfrak{D}^{\mathrm{r.o.}}, p^{\mathrm{rwd}}; T)$. For $t = 1$, let g_1^{PS} be the classical secretary algorithm that rejects the first n/e variables and accepts the first remaining variable X_i $(i > n/e)$ such that $X_i \geq \max_{1\leq i'\leq n/e} X_{i'}$. It is well known that this algorithm can accept the maximum realized value $\max_{i'\in[n]} X_{i'}$ with probability at least $1/e$ [16]. This implies that g_1^{PS} achieves a competitive ratio of at least $1/e$.

For $t = 2$, let g_2^{PS} be the single-sample algorithm for the prophet secretary, i.e., g_t^{PS} selects the first variable X_i such that $X_i \geq \max_{i'\in[n]} X_{i'}^{(1)}$. This algorithm achieves a competitive ratio of $1/2$ even for the prophet secretary problem, because the threshold $\max_{i'\in[n]} X_{i'}^{(1)}$ is independent of the ordering of the first round and $1/2$ guarantee holds regardless of the permutation observed in round t.

For $t \geq 3$, let g_t^{PS} be the algorithm with $t-1$ samples presented by Cristi and Ziliotto [13]. They demonstrated that, with $\Omega(1/\varepsilon^5)$ samples, we can obtain an $(\alpha^{\mathrm{PS}} - \varepsilon)$-competitive threshold algorithm, where α^{PS} is the current best known competitive ratio for the prophet secretary problem. α^{PS} is now 0.688, which is established by Chen et al. [7]. Hence, g_t^{PS} $(t \geq 3)$ achieves a competitive ratio of at least $0.688 - O(t^{-\frac{1}{5}})$.

Thus, by Theorem 5 with $t_0 \geq 2$, we can construct a sequence of algorithms $(h_t^{\mathrm{PS}})_{t\in[T]}$ for the repeated prophet secretary problem with baseline algorithms $(g_t^{\mathrm{PS}})_{t\in[T]}$. By Corollary 4, these algorithms satisfy the following performance guarantees.

Corollary 9. *For any repeated prophet secretary problem instance* $(\mathbf{D}, p^{\mathrm{rwd}}; T)$ *with* $\mathbf{D} \in \mathfrak{D}^{\mathrm{r.o.}}$*, the algorithms* $(h_t^{\mathrm{PS}})_{t\in[T]}$ *achieve the following:*

1. h_1^{PS} *is* $1/e$*-competitive,*
2. h_2^{PS} *is* $1/2$*-competitive,*
3. h_t^{PS} *is* $(\alpha^{\mathrm{PS}} - O(t^{-\frac{1}{5}}))$*-competitive for* $t \geq 3$*, and*
4. *the regret of* $(h_t^{\mathrm{PS}})_{t\in[T]}$ *is* $O(\sqrt{n!T\log T})$*,*

where α^{PS} *is the current best known competitive ratio for the prophet secretary problem.*

5.4 IID Prophet Inequality

Now, we consider the repeated i.i.d. prophet inequality $(\mathfrak{D}^{\mathrm{i.i.d.}}, p^{\mathrm{rwd}}; T)$. For this problem, it is known that the best possible competitive ratio is $\alpha^{\mathrm{PIiid}} \simeq 0.745$ [8, 22, 27, 31].

For $t = 1$, we define g_1^{PIiid} to be the algorithm that rejects the first n/e variables and accepts the first remaining variable X_i $(i > n/e)$ such that $X_i \geq \max_{1\leq i'\leq n/e} X_{i'}$. This algorithm

achieves a competitive ratio of at least $1/e$, since this setting can be interpreted as a prophet i.i.d. secretary problem. Correa et al. [10] provided the *fresh-looking samples algorithm*, which is $(1-1/e)$-competitive for the i.i.d. prophet inequality when $n-1$ samples $X'_1, \dots, X'_{n-1}$ are available. The algorithm accepts the first variable X_i such that X_i is greater than or equal to the maximum of a random subset of size $n-1$ from the set $\{X'_1, \dots, X'_{n-1}, X_1, \dots, X_{i-1}\}$. This implies that we can construct a $(1-1/e)$-competitive algorithm g_2^{PIiid} for the i.i.d. prophet inequality by using the algorithm of Correa et al. [10].

In addition, Correa et al. [12] showed that, with $\Omega(1/\varepsilon)$ samples from $\mathbf{D} \in \mathfrak{D}^{\mathrm{i.i.d.}}$, there exists a threshold algorithm for the i.i.d. prophet inequality whose competitive ratio is $\alpha^{\mathrm{PIiid}} - \varepsilon$. By this result, for $t \geq 3$, we obtain an $(\alpha^{\mathrm{PIiid}} - O(t^{-1}))$-competitive algorithm g_t^{PIiid}.

By an argument similar to that of Theorem 2, we can construct a sequence of algorithms $(h_t^{\mathrm{PIiid}})_{t\in[T]}$ with baseline $(g_t^{\mathrm{PIiid}})_{t\in[T]}$ that satisfy the following performance guarantees:

Corollary 10. *For any repeated i.i.d. prophet inequality instance* $(\mathbf{D}, p^{\mathrm{rwd}}; T)$ *with* $\mathbf{D} \in \mathfrak{D}^{\mathrm{i.i.d.}}$*, the algorithms* $(h_t^{\mathrm{PIiid}})_{t\in[T]}$ *achieve the following:*

1. h_1^{PIiid} *is* $1/e$*-competitive,*
2. h_2^{PIiid} *is* $(1-1/e)$*-competitive,*
3. h_t^{PIiid} *is* $(\alpha^{\mathrm{PIiid}} - O(t^{-1}))$*-competitive for* $t \geq 3$*, and*
4. *the regret of* $(h_t^{\mathrm{PIiid}*})_{t\in[T]}$ *is* $O(B\sqrt{T}\log T)$*,*

where $\alpha^{\mathrm{PIiid}} \simeq 0.745$ *is the best possible competitive ratio for i.i.d. prophet inequality.*

5.5 Random Order Secretary Problem / IID Secretary Problem

We address the repeated random order secretary problem $(\mathfrak{D}^{\mathrm{r.o.}}, p^{\mathrm{best}}; T)$ and the repeated secretary problem with identical distributions $(\mathfrak{D}^{\mathrm{i.i.d.}}, p^{\mathrm{best}}; T)$.

For the random order secretary problem, we define baseline algorithms $(g_t^{\mathrm{RS}})_{t\in[T]}$ as follows:

- Let g_1^{RS} be an algorithm that rejects the first n/e variables and accepts the first remaining variable X_i $(i > n/e)$ such that $X_i \geq \max_{1\leq i'\leq n/e} X_{i'}$.
- For $t \geq 2$, let g_t^{RS} be the block-rank algorithm [35] for the random order secretary problem.

For the random order secretary problem with a single sample $\mathbf{X}^{(1)}$, Nuti and Vondrák [35] showed that the block-rank algorithm achieves a competitive ratio of 0.5009 for the random order secretary problem and $0.5024 - O(1/\log n)$ for the i.i.d. secretary problem. They also established that no algorithm can obtain a competitive ratio better than 0.5024.

Let $(h_t^{\mathrm{RS}})_{t\in[T]}$ be algorithms as constructed in Theorem 5 with baseline $(g_t^{\mathrm{RS}})_{t\in[T]}$. Thus, we have the following corollary.

Corollary 11. *For any repeated random order secretary problem instance* $(\mathbf{D}, p^{\mathrm{best}}; T)$ *with* $\mathbf{D} \in \mathfrak{D}^{\mathrm{r.o.}}$*, the algorithms* $(h_t^{\mathrm{RS}})_{t\in[T]}$ *with baseline* $(g_t^{\mathrm{RS}})_{t\in[T]}$ *achieve the following:*

1. h_1^{RS} *is* $1/e$*-competitive,*
2. h_t^{RS} *is* 0.5009*-competitive for* $t \geq 2$*, and*

3. *the regret of* $(h_t^{\mathrm{RS}})_{t\in[T]}$ *is* $O(\sqrt{n!T\log T})$,

As $\mathfrak{D}^{\text{i.i.d.}}$ is a subset of $\mathfrak{D}^{\text{r.o.}}$, these algorithms are also applicable to the repeated i.i.d. secretary problem.

Corollary 12. *For any repeated i.i.d. secretary problem instance* $(\mathbf{D}, p^{\mathrm{best}}; T)$ *with* $\mathbf{D} \in \mathfrak{D}^{\text{i.i.d.}}$, *the algorithms* $(h_t^{\mathrm{RS}})_{t\in[T]}$ *with baseline* $(g_t^{\mathrm{RS}})_{t\in[T]}$ *achieve the following:*

1. h_1^{RS} *is* $1/e$*-competitive,*
2. h_t^{RS} *is* $(0.5024 - O(1/\log n))$*-competitive for* $t \geq 2$, *and*
3. *the regret of* $(h_t^{\mathrm{RS}})_{t\in[T]}$ *is* $O(\sqrt{nT\log T})$,

5.6 Last Success Problem

We explore the repeated adversarial order last success problem $(\mathfrak{D}^{\mathrm{adv}}, p^{\mathrm{ls}}; T)$, where the objective is to maximize the probability of selecting the last success index, that is, $p^{\mathrm{ls}}(\boldsymbol{x}, i) = \mathbb{1}_{\{i=\max\{i'|x_{i'}=1\}\}}$.

Let g_1^{LS} be the randomized algorithm that selects one variable uniformly at random. For $2 \leq t \leq T$, let g_t^{LS} be the one-sample algorithm for the adversarial order last success problem that accepts the first success after the index of the second last successes in the sample sequence.

Yoshinaga and Kawase [40] proved that the algorithm accepts the last success with probability at least $1/4$ if $R \geq (\sqrt{3}-1)/2$ and $\frac{R(4+3R)}{4(1+R)^2}$ if $0 \leq R \leq (\sqrt{3}-1)$, where

$$R = \sum_{i=1}^{n} \mathbf{Pr}\left[X_i = 1\right] / \mathbf{Pr}\left[X_i < 1\right].$$

Combining this with the fact that

$$\mathrm{Opt}_{\mathrm{offline}}(\mathbf{D}; p^{\mathrm{ls}}) = 1 - \prod_{i=1}^{n} \mathbf{Pr}\left[X_i < 1\right] \leq 1 - e^R,$$

we can obtain that the competitive ratio of the algorithm is at least $1/4$.

Let $(h_t^{\mathrm{LS}})_{t\in[T]}$ be algorithms as constructed in Theorem 5 with baseline algorithms $(g_t^{\mathrm{LS}})_{t\in[T]}$. Then, we have the following corollary by applying Corollary 4.

Corollary 13. *For the repeated adversarial last success problem instance* $(\mathbf{D}, p^{\mathrm{ls}}; T)$ *with* $\mathbf{D} \in \mathfrak{D}^{\mathrm{adv}}$, *the algorithms* $(h_t^{\mathrm{LS}})_{t\in[T]}$ *satisfy the following:*

1. h_1^{LS} *is* $1/n$*-competitive,*
2. h_t^{LS} *is* $1/4$*-competitive for* $t \geq 2$, *and*
3. *the regret of* $(h_t^{\mathrm{LS}})_{t\in[T]}$ *is* $O(\sqrt{nT\log T})$.

Note that the one-sample algorithm is also applicable in the general order model $\mathfrak{D}^{\mathrm{all}}$, since we can count the number of successes in the sample sequence from the distributions D_i that have not yet been observed.

Corollary 14. *For the repeated general order last success problem instance* $(\mathbf{D}, p^{\mathrm{ls}}; T)$ *with* $\mathbf{D} \in \mathfrak{D}^{\mathrm{all}}$*, the algorithms* $(h_t^{\mathrm{LS}})_{t\in[T]}$ *satisfy the following:*

1. h_1^{LS} *is* $1/n$*-competitive,*
2. h_t^{LS} *is* $1/4$*-competitive for* $t \geq 2$*, and*
3. *the regret of* $(h_t^{\mathrm{LS}})_{t\in[T]}$ *is* $O(\sqrt{n!T\log T})$*.*

5.7 Ski-rental Problem

We analyze the repeated ski-rental problem $(\mathfrak{D}^{\mathrm{adv}}, p^{\mathrm{ski}}; T)$, where the objective is to minimize the total cost $p(\boldsymbol{x}, i) = \sum_{j=1}^{i-1} x_j + b$. Note that the maximum cost in each round is at most $n + b$.

It is known that there exists a randomized algorithm without samples that achieves a competitive ratio of $e/(e-1)$ for this problem [26].[4] This online algorithm accepts the ith variable if the cumulative rental cost $\sum_{j=1}^{i} x_j$ exceeds a randomly preselected value. Then, for each $t \in [T]$, let g_t^{SKI} be the above algorithm without samples and let $(h_t^{\mathrm{SKI}})_{t\in[T]}$ be algorithms as constructed in Theorem 5 with baseline algorithms $(g_t^{\mathrm{SKI}})_{t\in[T]}$. This setting and Corollary 4 yield the following corollary.

Corollary 15. *Let* $(\mathbf{D}, p^{\mathrm{ski}}; T)$ *be any repeated ski-rental problem instance with* $\mathbf{D} \in \mathfrak{D}^{\mathrm{adv}}$*. Then,* h_t^{SKI} *is* $e/(e-1)$*-competitive for any* $t \in [T]$ *and the regret of* $(h_t^{\mathrm{SKI}})_{t\in[T]}$ *is* $O((n+b)\sqrt{nT\log T})$*.*

6 Lower Bound

In this section, we establish a regret lower bound of $\Omega(\sqrt{T})$ for optimal stopping problems. We have shown an upper bound of $O(\sqrt{T\log T})$ in Theorem 5. Therefore, ignoring the dependency of n and $|\Pi|$, this lower bound is tight up to a logarithmic factor with respect to the number of rounds T. Specifically, by Theorem 2 and Corollary 10, the lower bound $\Omega(\sqrt{T})$ is nearly tight for the repeated (adversarial / i.i.d.) prophet inequality problem.

Theorem 7. *Even when* $n = 2$*, any algorithm for both* $(\mathfrak{D}^{\mathrm{i.i.d.}}, p^{\mathrm{rwd}}; T)$ *and* $(\mathfrak{D}^{\mathrm{i.i.d.}}, p^{\mathrm{best}}; T)$ *incurs a regret of* $\Omega(\sqrt{T})$*.*

Since $\mathfrak{D}^{\mathrm{i.i.d.}}$ is a special case of $\mathfrak{D}^{\mathrm{adv}}$, $\mathfrak{D}^{\mathrm{r.o.}}$, and $\mathfrak{D}^{\mathrm{all}}$, the same lower bound holds for the corresponding problems with these distribution families.

Our results strengthen and generalize the recent finding by Gatmiry et al. [18, Theorem 5.1], who established a lower bound $\Omega(\sqrt{T})$ for the repeated prophet inequality $(\mathfrak{D}^{\mathrm{adv}}, p^{\mathrm{rwd}}; T)$ with $n = 2$ *distinct* distributions. In contrast, we establish the same lower bound for the case of $n = 2$ *identical* distributions. Consequently, our results apply not only to the adversarial order model but also to the random order and i.i.d. models. Moreover, since Gatmiry et al. [18] did not consider the best-choice profit p^{best}, our result is the first to establish this lower bound in the context of optimal stopping problems involving p^{best}.

To prove the theorem, we first define three distributions D_0, D_+, D_-, supported on $\{0, 1/2, 1\}$. Let $\varepsilon \in (0, 1/6)$ be a parameter to be chosen appropriately later. The distribution D_0 assigns $1/3$ to each value. The distribution D_+ assigns probability $1/3+\varepsilon$ to 1, probability $1/3-\varepsilon$ to $1/2$, and

[4] This online algorithm works in our setting by regarding a rental cost x_i as infinitesimally small costs.

probability 1/3 to 0. D_- assigns probability $1/3-\varepsilon$ to 1, probability $1/3+\varepsilon$ to 1/2, and probability 1/3 to 0. Let $\mathbf{D}_0 = (D_0, D_0; \Pi^{\mathrm{id}})$, $\mathbf{D}_+ = (D_+, D_+; \Pi^{\mathrm{id}})$, and $\mathbf{D}_- = (D_-, D_-; \Pi^{\mathrm{id}})$.

For the distributions $\mathbf{D}_+$ and $\mathbf{D}_-$, we have the following lemma.

Lemma 5. *Let $\mathbf{D} \in \{\mathbf{D}_+, \mathbf{D}_-\}$, and $p \in \{p^{\mathrm{rwd}}, p^{\mathrm{best}}\}$. Suppose that h is a deterministic algorithm for the optimal stopping problem $(\mathbf{D}, p)$ that does not achieve the optimal online profit. Then, we have*

$$\mathrm{Opt}_{\mathrm{online}}(\mathbf{D}; p) - h(\mathbf{D}; p) \geq \varepsilon/12.$$

Proof. For any distribution supported on $\{0, 1/2, 1\}$, it suffices to consider algorithms that (i) always accept the first variable realizing 1 and reject the first variable realizing 0, and (ii) always accept the second variable if the first variable is rejected. Thus, with 2 variables X_1 and X_2, there are only two deterministic algorithms: h_+ and h_-. Here, h_+ rejects X_1 when $X_1 = 1/2$, while h_- accepts X_1 when $X_1 = 1/2$.

For the expected reward maximization problem (i.e., the profit function is p^{rwd}), the expected profit of h_+ and h_- under $\mathbf{D}_+$ and $\mathbf{D}_-$ can be calculated as follows:

$$\begin{aligned}
h_+(\mathbf{D}_+; p^{\mathrm{rwd}}) &= \left(\frac{1}{3}+\varepsilon\right)\cdot 1 + \left(\frac{1}{3}-\varepsilon\right)\cdot\frac{1+\varepsilon}{2} + \frac{1}{3}\cdot\frac{1+\varepsilon}{2},\\
h_-(\mathbf{D}_+; p^{\mathrm{rwd}}) &= \left(\frac{1}{3}+\varepsilon\right)\cdot 1 + \left(\frac{1}{3}-\varepsilon\right)\cdot\frac{1}{2} + \frac{1}{3}\cdot\frac{1+\varepsilon}{2},\\
h_+(\mathbf{D}_-; p^{\mathrm{rwd}}) &= \left(\frac{1}{3}-\varepsilon\right)\cdot 1 + \left(\frac{1}{3}+\varepsilon\right)\cdot\frac{1-\varepsilon}{2} + \frac{1}{3}\cdot\frac{1-\varepsilon}{2},\\
h_-(\mathbf{D}_-; p^{\mathrm{rwd}}) &= \left(\frac{1}{3}-\varepsilon\right)\cdot 1 + \left(\frac{1}{3}+\varepsilon\right)\cdot\frac{1}{2} + \frac{1}{3}\cdot\frac{1-\varepsilon}{2}.
\end{aligned}$$

Then, when the given distribution is $\mathbf{D}_+$, h_+ achieves the optimal online profit and h_- incurs a per-round regret of at least

$$h_+(\mathbf{D}_+; p^{\mathrm{rwd}}) - h_-(\mathbf{D}_+; p^{\mathrm{rwd}}) = \left(\frac{1}{3}-\varepsilon\right)\cdot\frac{\varepsilon}{2} \geq \frac{\varepsilon}{12},$$

where the inequality is due to $0 < \varepsilon \leq 1/6$. When the given distribution is $\mathbf{D}_-$, h_- achieves the optimal online profit and h_+ incurs a per-round regret of at least

$$h_-(\mathbf{D}_-; p^{\mathrm{rwd}}) - h_+(\mathbf{D}_-; p^{\mathrm{rwd}}) = \left(\frac{1}{3}+\varepsilon\right)\cdot\frac{\varepsilon}{2} \geq \frac{\varepsilon}{6}.$$

Similarly, for the best choice probability maximization problem (i.e., the profit function is p^{best}), the expected profit of h_+ and h_- under $\mathbf{D}_+$ and $\mathbf{D}_-$ can be calculated as follows:

$$\begin{aligned}
h_+(\mathbf{D}_+; p^{\mathrm{best}}) &= \left(\frac{1}{3}+\varepsilon\right)\cdot 1 + \left(\frac{1}{3}-\varepsilon\right)\cdot\frac{2}{3} + \frac{1}{3}\cdot 1,\\
h_-(\mathbf{D}_+; p^{\mathrm{best}}) &= \left(\frac{1}{3}+\varepsilon\right)\cdot 1 + \left(\frac{1}{3}-\varepsilon\right)\cdot\left(\frac{2}{3}-\varepsilon\right) + \frac{1}{3}\cdot 1,\\
h_+(\mathbf{D}_-; p^{\mathrm{best}}) &= \left(\frac{1}{3}-\varepsilon\right)\cdot 1 + \left(\frac{1}{3}+\varepsilon\right)\cdot\frac{2}{3} + \frac{1}{3}\cdot 1,\\
h_-(\mathbf{D}_-; p^{\mathrm{best}}) &= \left(\frac{1}{3}-\varepsilon\right)\cdot 1 + \left(\frac{1}{3}+\varepsilon\right)\cdot\left(\frac{2}{3}+\varepsilon\right) + \frac{1}{3}\cdot 1.
\end{aligned}$$

Then, when the given distribution is $\mathbf{D}_+$, h_+ achieves the optimal online profit and h_- incurs a per-round regret of at least

$$h_+(\mathbf{D}_+; p^{\mathrm{best}}) - h_-(\mathbf{D}_+; p^{\mathrm{best}}) = \left(\frac{1}{3} - \varepsilon\right) \cdot \varepsilon \geq \frac{\varepsilon}{6},$$

where the inequality is due to $0 < \varepsilon \leq 1/6$. When the given distribution is $\mathbf{D}_-$, h_- achieves the optimal online profit and h_+ incurs a per-round regret of at least

$$h_-(\mathbf{D}_-; p^{\mathrm{best}}) - h_+(\mathbf{D}_-; p^{\mathrm{best}}) = \left(\frac{1}{3} + \varepsilon\right) \cdot \varepsilon \geq \frac{\varepsilon}{3}.$$

□

Proof of Theorem 7. By Yao's principle, it suffices to show a regret lower bound for any deterministic selection rule against the uniform distribution over $\{\mathbf{D}_+, \mathbf{D}_-\}$. Fix an arbitrary deterministic selection rule, and let h_t be a deterministic algorithm chosen in round t. Since we consider a deterministic rule, h_t is determined by the previous samples $\mathbb{X}_{t-1} = (\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(t-1)})$ and does not have any other randomness, i.e., h_t is either h_+ or h_- for each round $t \in [T]$.

For any T samples $\mathbb{X}_T = (\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(T)}) \in \{0, 1/2, 1\}^{2 \times T}$, let $f_+(\mathbb{X}_T)$ (resp. $f_-(\mathbb{X}_T)$) denote the number of t such that $h_t = h_+$ (resp. $h_t = h_-$) when $\mathbb{X}_T$ is realized. Let $\mathbf{Pr}_+[\cdot]$, $\mathbf{Pr}_-[\cdot]$, and $\mathbf{Pr}_0[\cdot]$ denote the probability when the underlying distribution is $\mathbf{D}_+$, $\mathbf{D}_-$, and $\mathbf{D}_0$, respectively. Similarly, $\mathbb{E}_+[\cdot]$, $\mathbb{E}_-[\cdot]$, and $\mathbb{E}_0[\cdot]$ denote the expectation when the underlying distribution is $\mathbf{D}_+$, $\mathbf{D}_-$, and $\mathbf{D}_0$, respectively. We use the following claim.

Claim 1. *The following inequalities hold:*

$$\mathbb{E}_+[f_+(\mathbb{X}_T)] \leq \mathbb{E}_0[f_+(\mathbb{X}_T)] + 2\varepsilon T\sqrt{T}, \tag{9}$$

$$\mathbb{E}_-[f_-(\mathbb{X}_T)] \leq \mathbb{E}_0[f_-(\mathbb{X}_T)] + 2\varepsilon T\sqrt{T}. \tag{10}$$

Proof of Claim 1. We first prove (9). It follows that

$$\begin{aligned}
\mathbb{E}_+[f_+(\mathbb{X}_T)] - \mathbb{E}_0[f_+(\mathbb{X}_T)] &= \sum_{\mathbb{X}_T \in \{0,1/2,1\}^{2\times T}} f_+(\mathbb{X}_T)\Big(\mathbf{Pr}_+[\mathbb{X}_T] - \mathbf{Pr}_0[\mathbb{X}_T]\Big) \\
&\leq \sum_{\substack{\mathbb{X}_T \in \{0,1/2,1\}^{2\times T} \\ \mathbb{X}_T : \mathbf{Pr}_+[\mathbb{X}_T] \geq \mathbf{Pr}_0[\mathbb{X}_T]}} f_+(\mathbb{X}_T)\Big(\mathbf{Pr}_+[\mathbb{X}_T] - \mathbf{Pr}_0[\mathbb{X}_T]\Big) \\
&\leq T \cdot \sum_{\substack{\mathbb{X}_T \in \{0,1/2,1\}^{2\times T} \\ \mathbb{X}_T : \mathbf{Pr}_+[\mathbb{X}_T] \geq \mathbf{Pr}_0[\mathbb{X}_T]}} \Big(\mathbf{Pr}_+[\mathbb{X}_T] - \mathbf{Pr}_0[\mathbb{X}_T]\Big) \\
&= \frac{T}{2} \cdot \sum_{\mathbb{X}_T \in \{0,1/2,1\}^{2\times T}} \big|\mathbf{Pr}_+[\mathbb{X}_T] - \mathbf{Pr}_0[\mathbb{X}_T]\big| \\
&\leq \frac{T}{2} \cdot \sqrt{2\,\mathrm{KL}(\mathbf{Pr}_0 \parallel \mathbf{Pr}_+)},
\end{aligned}$$

where the last inequality is due to Pinsker's inequality and $\mathrm{KL}(\mathbf{Pr}_0 \parallel \mathbf{Pr}_+)$ denotes the KL-divergence between $\mathbf{Pr}_0$ and $\mathbf{Pr}_+$, i.e.,

$$\mathrm{KL}(\mathbf{Pr}_0 \parallel \mathbf{Pr}_+) \coloneqq \sum_{\mathbb{X}_T \in \{0,1/2,1\}^{2\times T}} \mathbf{Pr}_0[\mathbb{X}_T] \ln \frac{\mathbf{Pr}_0[\mathbb{X}_T]}{\mathbf{Pr}_+[\mathbb{X}_T]}.$$

Since $\mathbb{X}_T = ((X_1^{(1)}, X_2^{(1)}), \ldots, (X_1^{(T)}, X_2^{(T)}))$ follows product distribution D_0^{2T} or D_+^{2T}, we can evaluate $\mathrm{KL}(\mathbf{Pr}_0 \parallel \mathbf{Pr}_+)$ as follows:

$$\begin{aligned}
\mathrm{KL}(\mathbf{Pr}_0 \parallel \mathbf{Pr}_+) &= 2T \cdot \mathrm{KL}(D_0 \parallel D_+) \\
&= 2T \left(\frac{1}{3} \ln \frac{1/3}{1/3+\varepsilon} + \frac{1}{3} \ln \frac{1/3}{1/3-\varepsilon} + \frac{1}{3} \ln \frac{1/3}{1/3} \right) \\
&= -\frac{2T}{3} \ln(1 - 9\varepsilon^2) \\
&\leq \frac{2T}{3} \cdot 4 \ln(4/3) \cdot 9\varepsilon^2,
\end{aligned}$$

where the last inequality is due to the convexity of $-\ln(1-x)$ and $9\varepsilon^2 \leq 1/4$. Then, by letting $C = 2\sqrt{3\ln(4/3)} \approx 1.86 < 2$, we have (9) as follows:

$$\mathbb{E}_+[f_+(\mathbb{X}_T)] - \mathbb{E}_0[f_+(\mathbb{X}_T)] \leq \frac{T}{2} \sqrt{2 \cdot \frac{2T}{3} \cdot 4 \ln(4/3) \cdot 9\varepsilon^2} = C\varepsilon T\sqrt{T} < 2\varepsilon T\sqrt{T}.$$

The proof of (10) is analogous. □

We return to the proof of Theorem 7. By Claim 1, it holds that

$$\frac{1}{2}(\mathbb{E}_+[f_+(\mathbb{X}_T)] + \mathbb{E}_-[f_-(\mathbb{X}_T)]) \leq \frac{1}{2} \cdot \mathbb{E}_0[f_+(\mathbb{X}_T) + f_-(\mathbb{X}_T)] + 2\varepsilon T\sqrt{T} = \frac{T}{2} + 2\varepsilon T\sqrt{T}.$$

This implies that, against the uniform distribution over $\{\mathbf{D}_+, \mathbf{D}_-\}$, any deterministic rule can select the "correct algorithm" (h_+ for distribution $\mathbf{D}_+$ or h_- for distribution $\mathbf{D}_-$) only $T/2 + 2\varepsilon T\sqrt{T}$ times in expectation. By Lemma 5, any deterministic selection rule incurs cumulative regret at least

$$\frac{\varepsilon}{12} \cdot \Big(T - \big(T/2 + 2\varepsilon T\sqrt{T} \big) \Big) = \frac{\varepsilon}{12} \cdot \big(T/2 - 2\varepsilon T\sqrt{T} \big).$$

Tuning $\varepsilon = 1/(8\sqrt{T})$ yields the lower bound $\Omega(\sqrt{T})$. □

7 Discussion and Future Work

In this paper, we proposed a general framework for repeated optimal stopping problems that achieves both competitive ratio guarantees for each round and sublinear regret over all rounds. Our main contribution is an adaptive algorithm that, in each round, selects either a sample-based competitive algorithm or an empirically optimal algorithm. This approach leverages the structure of threshold algorithms and their bounded Rademacher complexity to ensure performance guarantees in both metrics.

One drawback of our framework is that it requires full feedback, meaning that the algorithm must observe the realized values of all variables in each round. This assumption is difficult to relax if we wish to achieve performance better than that of the baseline sample-based algorithm in every round. For example, consider the repeated prophet inequality, where we employ the single-sample algorithm that sets the maximum of the first round's samples as the threshold. In each round, if the first variable equals the threshold, the algorithm must accept it to achieve the competitive

ratio guarantee, and no information about the realized values of the other variables is available. Whether it is possible to handle partial feedback by incorporating an ε-greedy method together with slightly weakening the round-wise guarantee remains an open question for future work.

Our framework demonstrates significant potential for extension beyond the specific problems analyzed. For example, it can be naturally generalized to settings involving top-k-choice or multiple choices, as a threshold strategy is often optimal even in such cases. This allows us to bound the Rademacher complexity, as well as the competitive ratio and regret, in a similar manner. Moreover, our framework can be naturally extended to repeated settings of other online problems, such as online matching problems [23, 34] or the online dial-a-ride problem [3, 30]. Even when the underlying distribution lacks an explicit repetitive structure, introducing a periodic reset to a fixed initial state can create an artificial repeated structure suitable for our analysis. For each problem, the key is to design an appropriate sample-based algorithm and identify a class of online algorithms with manageable Rademacher complexity, which enables both competitive ratio and regret guarantees.

Acknowledgments

This work was partially supported by the joint project of Kyoto University and Toyota Motor Corporation, titled "Advanced Mathematical Science for Mobility Society", JST ERATO Grant Number JPMJER2301, JST ASPIRE Grant Number JPMJAP2302, and JSPS KAKENHI Grant Numbers JP21K17708, JP21H03397, JP25K00137.

A Omitted Proofs

A.1 Proof of Lemma 1

We construct a threshold algorithm $h^* = (f_i^*)_{i=1}^n$ iteratively from $i = n$ to 1.

For $i \in [n]$, $\boldsymbol{x}_{\le i} \in [0,1]^i$, and $\tau_{\le i} \in \mathrm{Inj}([i],[n])$, let $\beta(\boldsymbol{x}_{\le i}, \tau_{\le i})$ be the optimal expected profit (of an online algorithm) if the realization of $(\mathbf{X}_{\le i}, \pi_{\le i})$ is $(\boldsymbol{x}_{\le i}, \tau_{\le i})$, and the first i variables are rejected, that is,

$$\begin{aligned}\beta(\boldsymbol{x}_{\le i}, \tau_{\le i}) &= \sup_{h \in \mathcal{H}\colon \text{always reject first } i \text{ variables}} \mathbb{E}\left[h(\mathbf{X}, \pi; p) \;\middle|\; \begin{array}{l}\mathbf{X}_{\le i} = \boldsymbol{x}_{\le i},\\ \pi_{\le i} = \tau_{\le i}\end{array}\right] \\ &= \sup_{f_{i+1},\dots,f_n} \mathbb{E}\left[p(\mathbf{X}, \min\{j \in \{i+1,\dots,n\} \mid f_j(\boldsymbol{x}_{\le j}, \tau_{\le j}) = 1\}) \;\middle|\; \begin{array}{l}\mathbf{X}_{\le i} = \boldsymbol{x}_{\le i},\\ \pi_{\le i} = \tau_{\le i}\end{array}\right],\end{aligned}$$

where we denote $\min \emptyset = n+1$. By definition, we have $\beta(\boldsymbol{x}, \tau) = p(\boldsymbol{x}, n+1)$ for any $\boldsymbol{x} \in [0,1]^n$ and $\tau \in \mathrm{Inj}([n],[n])$. In addition, for $i \in [n]$, $\boldsymbol{x}_{\le i} \in [0,1]^i$, and $\tau_{\le i} \in \mathrm{Inj}[i],[n]$, let $\alpha(\boldsymbol{x}_{\le i}, \tau_{\le i})$ be the expected profit if the realization of $(\mathbf{X}_{\le i}, \pi_{\le i})$ is $(\boldsymbol{x}_{\le i}, \tau_{\le i})$, and the ith variables is accepted, i.e.,

$$\alpha(\boldsymbol{x}_{\le i}, \tau_{\le i}) \coloneqq \mathbb{E}\left[p(\mathbf{X}, i) \mid \mathbf{X}_{\le i} = \boldsymbol{x}_{\le i} \text{ and } \pi_{\le i} = \tau_{\le i}\right].$$

Then, when the realizations of the first steps are $\boldsymbol{x}_{\le i}$ and $\tau_{\le i}$ and the first $i-1$ variables are rejected, an optimal online algorithm accepts the ith variable if

$$\alpha(\boldsymbol{x}_{\le i}, \tau_{\le i}) \ge \beta(\boldsymbol{x}_{\le i}, \tau_{\le i}).$$

Suppose that

$$(f_{i+1}^*, \dots, f_n^*) \in \operatorname*{arg\,max}_{f_{i+1},\dots,f_n} \mathbb{E}\left[p(\mathbf{X}, \min\{j \in \{i+1,\dots,n\} \mid f_j(\boldsymbol{x}_{\le j}, \tau_{\le j}) = 1\}) \;\middle|\; \begin{array}{l} \mathbf{X}_{\le i} = \boldsymbol{x}_{\le i}, \\ \pi_{\le i} = \tau_{\le i} \end{array} \right].$$

For notational convenience, let $\gtrsim_j$ represents $\ge$ if $f_j^*(\boldsymbol{x}_{\le j}, \tau_{\le j}) = \mathbb{1}_{\{x_j \ge \theta_j(\tau_{\le j})\}}$ and $>$ otherwise, i.e., $f_j^*(\boldsymbol{x}_{\le j}, \tau_{\le j}) = \mathbb{1}_{\{x_j > \theta_j(\tau_{\le j})\}}$. Then, we have

$$\begin{aligned}
&\beta(\boldsymbol{x}_{\le i}, \tau_{\le i}) - \alpha(\boldsymbol{x}_{\le i}, \tau_{\le i}) \\
&\quad = \max_{f_{i+1},\dots,f_n} \mathbb{E}\left[p\big(\mathbf{X}, \min\big\{j \in \{i+1,\dots,n\} \mid f_j(\boldsymbol{x}_{\le j}, \tau_{\le j}) = 1\big\}\big) - p(\mathbf{X}, i) \;\middle|\; \begin{array}{l} \mathbf{X}_{\le i} = \boldsymbol{x}_{\le i}, \\ \pi_{\le i} = \tau_{\le i} \end{array} \right] \\
&\quad = \mathbb{E}\left[p\big(\mathbf{X}, \min\big\{j \in \{i+1,\dots,n\} \mid f_j^*(\boldsymbol{x}_{\le j}, \tau_{\le j}) = 1\big\}\big) - p(\mathbf{X}, i) \;\middle|\; \begin{array}{l} \mathbf{X}_{\le i} = \boldsymbol{x}_{\le i}, \\ \pi_{\le i} = \tau_{\le i} \end{array} \right] \\
&\quad = \mathbb{E}\left[p\big(\mathbf{X}, \min\big\{j \in \{i+1,\dots,n\} \mid X_j \gtrsim_j \theta_j(\tau_{\le j})\big\}\big) - p(\mathbf{X}, i) \;\middle|\; \begin{array}{l} \mathbf{X}_{\le i} = \boldsymbol{x}_{\le i}, \\ \pi_{\le i} = \tau_{\le i} \end{array} \right] \\
&\quad = \mathbb{E}\left[q_i\big(\mathbf{X}_{\ge i}, \min\big\{j \in \{i+1,\dots,n\} \mid X_j \gtrsim_j \theta_j(\tau_{\le j})\big\}\big) \;\middle|\; \begin{array}{l} \mathbf{X}_{\le i} = \boldsymbol{x}_{\le i}, \\ \pi_{\le i} = \tau_{\le i} \end{array} \right] \\
&\quad = \mathbb{E}\left[q_i\big(\mathbf{X}_{\ge i}, \min\big\{j \in \{i+1,\dots,n\} \mid X_j \gtrsim_j \theta_j(\tau_{\le j})\big\}\big) \;\middle|\; \begin{array}{l} \mathbf{X}_i = \boldsymbol{x}_i, \\ \pi_{\le i} = \tau_{\le i} \end{array} \right]. \qquad (11)
\end{aligned}$$

Thus, whether $\alpha(\boldsymbol{x}_{\le i}, \tau_{\le i}) \ge \beta(\boldsymbol{x}_{\le i}, \tau_{\le i})$ does not depend on the realization of the first $i-1$ variables.

Define a threshold as

$$\theta_i(\tau_{\le i}) \coloneqq \inf\left\{ x_i \in [0,1] \;\middle|\; \mathbb{E}\left[q_i\big(\mathbf{X}_{\ge i}, \min\big\{j \in \{i+1,\dots,n\} \mid X_j \gtrsim_j \theta_j(\tau_{\le j})\big\}\big) \;\middle|\; \begin{array}{l} X_i = x_i, \\ \pi_{\le i} = \tau_{\le i} \end{array} \right] \le 0 \right\}.$$

If $\alpha((\boldsymbol{x}_{\le i-1}, x), \tau_{\le i}) \ge \beta((\boldsymbol{x}_{\le i-1}, x), \tau_{\le i})$ for all $x \in [0,1]$, then we set $\theta_i(x_1, \dots, x_{i-1}) = \infty$. Note that the definition of θ_i does not depend on the choice of $\boldsymbol{x}_{\le i-1}$.

By the properties (i) and (ii), the value of $q_i(\boldsymbol{x}_{\ge i}, j) = p(\boldsymbol{x}, j) - p(\boldsymbol{x}, i)$ is monotone non-increasing with respect to x_i. This implies that

$$\mathbb{E}\left[q_i\big(\mathbf{X}_{\ge i}, \min\big\{j \in \{i+1,\dots,n\} \mid X_j \gtrsim_j \theta_j(\tau_{\le j})\big\}\big) \;\middle|\; X_i = x_i \text{ and } \pi_{\le i} = \tau_{\le i} \right]$$

is also monotone non-increasing with respect to x_i. Therefore, if $x_i > \theta_i(\tau_{\le i})$, we have $\alpha(\boldsymbol{x}_{\le i}, \tau_{\le i}) \ge \beta(\boldsymbol{x}_{\le i}, \tau_{\le i})$ by (11). Also, if $x_i < \theta_i(\tau_{\le i})$, we have $\alpha(\boldsymbol{x}_{\le i}, \tau_{\le i}) < \beta(\boldsymbol{x}_{\le i}, \tau_{\le i})$ by (11). Thus, $f_i^*(\boldsymbol{x}_{\le i}, \tau_{\le i}) = \mathbb{1}_{\{x_i \ge \theta_i(\tau_{\le i})\}}$ or $\mathbb{1}_{\{x_i > \theta_i(\tau_{\le i})\}}$ is an optimal choice of the function. Hence, by induction, we conclude that the threshold algorithm $h^* = (f_i^*)_{i=1}^n$ is an optimal online algorithm.

A.2 Proof of Lemma 2

To prove Lemma 2, we use the following lemma.

Lemma 6 (Massart's Lemma [33, Lemma 5.2])**.** *Let A be a finite subset of $\mathbb{R}^t$, and let $r \coloneqq \max_{a \in A} \|a\|_2$. Then,*

$$\mathbb{E}_\sigma\left[\frac{1}{t} \sup_{a \in A} \sum_{s=1}^t \sigma_s a_s \right] \le \frac{r\sqrt{2 \ln |A|}}{t},$$

where $\sigma_1, \dots, \sigma_t$ are independent Rademacher random variables.

Let $-A \coloneqq \{-a \mid a \in A\}$. Obviously, the following holds:

$$\sup_{a \in A \cup -A} \sum_{s=1}^{t} \sigma_s a_i = \sup_{a \in A} \left| \sum_{s=1}^{t} \sigma_s a_i \right|.$$

Then, we have

$$\mathrm{Rad}(A) = \mathbb{E}_\sigma \left[\frac{1}{t} \sup_{a \in A} \left| \sum_{s=1}^{t} \sigma_s a_i \right| \right] = \mathbb{E}_\sigma \left[\frac{1}{t} \sup_{a \in A \cup -A} \sum_{s=1}^{t} \sigma_s a_i \right] \leq \frac{r\sqrt{2\ln(2|A|)}}{t},$$

where the inequality is due to Lemma 6.

B Counterexample

In this section, we show an example in which simply adopting a single-sample algorithm in every round leads to a linear regret.

Let ε be a positive real that is less than $1/2$. Consider the repeated prophet inequality with two random variables and T rounds where:

- $X_1 = 1/2$ with probability 1,
- $X_2 = 1$ with probability $1/2 + \varepsilon$, $X_2 = 0$ with probability $1/2 - \varepsilon$.

For this instance, the single-sample online algorithm [36] for the prophet inequality operates as follows:

1. If the first sample is $(1/2, 1)$ (the probability is $1/2 + \varepsilon$), then the single-sample algorithm chooses the first value ≥ 1, and hence the expected reward in each round is $1/2 + \varepsilon$.
2. If the first sample is $(1/2, 0)$, then the single-sample algorithm always chooses the value $1/2$ of X_1.

We assume that, for the first round, it accepts X_2. Then, the expected total reward of the single-sample online algorithm over T rounds is

$$\begin{aligned}
&(1/2 + \varepsilon) \cdot (1 + (T-1)(1/2 + \varepsilon)) + (1/2 - \varepsilon) \cdot (0 + (T-1) \cdot 1/2) \\
&\quad = T \cdot (1/2 + \varepsilon) - (T-1) \cdot (\varepsilon/2 - \varepsilon^2)
\end{aligned}$$

On the other hand, the optimal online algorithm always chooses the value of X_2 because $\mathbb{E}[X_2] = 1/2 + \varepsilon > 1/2 = \mathbb{E}[X_1]$. Then the expected total reward of the optimal online algorithm is $T \cdot (1/2 + \varepsilon)$.

Therefore, the regret is $(T-1)(\varepsilon/2 - \varepsilon^2)$, which increases linearly with respect to T.

References

[1] A. Agarwal, R. Ghuge, and V. Nagarajan. Semi-bandit learning for monotone stochastic optimization. In *Proceedings of the 65th Annual Symposium on Foundations of Computer Science*, pages 1260–1274, 2024.

[2] L. Andrew, S. Barman, K. Ligett, M. Lin, A. Meyerson, A. Roytman, and A. Wierman. A tale of two metrics: Simultaneous bounds on competitiveness and regret. In *Proceedings of the 26th Annual Conference on Learning Theory*, pages 741–763, 2013.

[3] N. Ascheuer, S. O. Krumke, and J. Rambau. Online dial-a-ride problems: Minimizing the completion time. In *Proceedings of the 17th Annual Symposium on Theoretical Aspects of Computer Science*, pages 639–650, 2000.

[4] P. D. Azar, R. Kleinberg, and S. M. Weinberg. Prophet inequalities with limited information. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1358–1377, 2014.

[5] N. Buchbinder, S. Chen, J. S. Naor, and O. Shamir. Unified algorithms for online learning and competitive analysis. *Mathematics of Operations Research*, 41(2):612–625, 2016.

[6] C. Caramanis, P. Dütting, M. Faw, F. Fusco, P. Lazos, S. Leonardi, O. Papadigenopoulos, E. Pountourakis, and R. Reiffenhäuser. Single-sample prophet inequalities via greedy-ordered selection. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1298–1325, 2022.

[7] J. Chen, P. Dütting, E. Gergatsouli, R. Rezvan, and A. Tsigonias-Dimitriadis. Prophet secretary and matching: the significance of the largest item. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1371–1401, 2025.

[8] J. Correa, P. Foncea, R. Hoeksma, T. Oosterwijk, and T. Vredeveld. Posted price mechanisms for a random stream of customers. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 169–186, 2017.

[9] J. Correa, P. Dütting, F. Fischer, K. Schewior, and B. Ziliotto. Unknown i.i.d. prophets: Better bounds, streaming algorithms, and a new impossibility. In *Proceedings of the 12th Innovations in Theoretical Computer Science Conference*, page 86:1, 2021.

[10] J. R. Correa, P. Dütting, F. Fischer, and K. Schewior. Prophet inequalities for i.i.d. random variables from an unknown distribution. In *Proceedings of the 20th ACM Conference on Economics and Computation*, pages 3–17, 2019.

[11] J. R. Correa, A. Cristi, B. Epstein, and J. A. Soto. The two-sided game of googol and sample-based prophet inequalities. In *Proceedings of the 31st ACM-SIAM Symposium on Discrete Algorithms*, pages 2066–2081, 2020.

[12] J. R. Correa, A. Cristi, B. Epstein, and J. A. Soto. Sample-driven optimal stopping: From the secretary problem to the i.i.d. prophet inequality. *Mathematics of Operations Research*, 49(1): 441–475, 2024.

[13] A. Cristi and B. Ziliotto. Prophet inequalities require only a constant number of samples. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 491–502, 2024.

[14] A. Daniely and Y. Mansour. Competitive ratio vs regret minimization: Achieving the best of both worlds. In *Proceedings of the 30th International Conference on Algorithmic Learning Theory*, pages 333–368, 2019.

[15] E. B. Dynkin. The optimum choice of the instant for stopping a Markov process. *Soviet Mathematics*, 4:627–629, 1963.

[16] T. S. Ferguson. Who solved the secretary problem? *Statistical science*, 4(3):282–289, 1989.

[17] H. Fu and T. Lin. Learning utilities and equilibria in non-truthful auctions. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, pages 14231–14242, 2020.

[18] K. Gatmiry, T. Kesselheim, S. Singla, and Y. Wang. Bandit algorithms for prophet inequality and Pandora's box. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 462–500, 2024.

[19] G. Goel, N. Agarwal, K. Singh, and E. Hazan. Best of both worlds in online control: Competitive ratio and policy regret. In *Proceedings of the 5th Annual Learning for Dynamics and Control Conference*, pages 1345–1356, 2023.

[20] C. Guo, Z. Huang, and X. Zhang. Settling the sample complexity of single-parameter revenue maximization. In *Proceedings of the 51st Annual ACM Symposium on Theory of Computing*, pages 662–673, 2019.

[21] C. Guo, Z. Huang, Z. G. Tang, and X. Zhang. Generalizing complex hypotheses on product distributions: Auctions, prophet inequalities, and Pandora's problem. In *Proceedings of the 34th Conference on Learning Theory*, pages 2248–2288, 2021.

[22] T. P. Hill and R. P. Kertz. Comparisons of stop rule and supremum expectations of iid random variables. *The Annals of Probability*, pages 336–345, 1982.

[23] Z. Huang, Z. G. Tang, and D. Wajc. Online matching: A brief survey. *ACM SIGecom Exchanges*, 22(1):135–158, 2024.

[24] B. Jin, T. Kesselheim, W. Ma, and S. Singla. Sample complexity of posted pricing for a single item. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, pages 82296–82317, 2024.

[25] H. Kaplan, D. Naori, and D. Raz. Competitive analysis with a sample and the secretary problem. In *Proceedings of the 31st Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2082–2095, 2020.

[26] A. R. Karlin, M. S. Manasse, L. A. McGeoch, and S. Owicki. Competitive randomized algorithms for nonuniform problems. *Algorithmica*, 11(6):542–571, 1994.

[27] R. P. Kertz. Stop rule and supremum expectations of iid random variables: a complete comparison by conjugate duality. *Journal of multivariate analysis*, 19(1):88–112, 1986.

[28] T. Kesselheim, R. Kleinberg, and R. Niazadeh. Secretary problems with non-uniform arrival order. In *Proceedings of the 47th annual ACM symposium on Theory of computing*, pages 879–888, 2015.

[29] U. Krengel and L. Sucheston. Semiamarts and finite values. *Bulletin of the American Mathematical Society*, 83(4):745–747, 1977.

[30] S. O. Krumke. *Online optimization: Competitive analysis and beyond*. Zuse Institute Berlin, 2006.

[31] A. Liu, R. P. Leme, M. Pál, J. Schneider, and B. Sivan. Variable decomposition for prophet inequalities and optimal ordering. *arXiv preprint arXiv:2004.10163*, 2020.

[32] W. Ma, C. MacRury, and C. Stein. Forward-backward contention resolution schemes for fair rationing. In *Proceedings of the 26th ACM Conference on Economics and Computation*, page 945, 2025.

[33] P. Massart. Some applications of concentration inequalities to statistics. In *Annales de la Faculté des sciences de Toulouse: Mathématiques*, volume 9, pages 245–303, 2000.

[34] A. Mehta et al. Online matching and ad allocation. *Foundations and Trends® in Theoretical Computer Science*, 8(4):265–368, 2013.

[35] P. Nuti and J. Vondrák. Secretary problems: The power of a single sample. In *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms*, pages 2015–2029, 2023.

[36] A. Rubinstein, J. Z. Wang, and S. M. Weinberg. Optimal single-choice prophet inequalities from samples. In *Proceedings of the 11th Innovations in Theoretical Computer Science Conference*, pages 60:1–60:10, 2020.

[37] A. Shah and A. Rajkumar. Sequential ski rental problem. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1173–1181, 2021.

[38] R. J. Vanderbei. The postdoc variant of the secretary problem. *Mathematica Applicanda*, 49 (1), 2021.

[39] M. J. Wainwright. *Uniform laws of large numbers*, pages 98–120. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019.

[40] T. Yoshinaga and Y. Kawase. The last success problem with samples. In *Proceedings of the 32nd Annual European Symposium on Algorithms*, pages 105:1–105:15, 2024.