Bandit Sequential Posted Pricing via Half-Concavity

SAHIL SINGLA*, Georgia Institute of Technology, USA
YIFAN WANG*, Georgia Institute of Technology, USA

Sequential posted pricing auctions are popular because of their simplicity in practice and their tractability in theory. A usual assumption in their study is that the Bayesian prior distributions of the buyers are known to the seller, while in reality these priors can only be accessed from historical data. To overcome this assumption, we study sequential posted pricing in the bandit learning model, where the seller interacts with n buyers over T rounds: In each round the seller posts n prices for the n buyers and the first buyer with a valuation higher than the price takes the item. The only feedback that the seller receives in each round is the revenue.

Our main results obtain nearly-optimal regret bounds for single-item sequential posted pricing in the bandit learning model. In particular, we achieve an $\widetilde{O}(\mathrm{poly}(n)\sqrt{T})$ regret for buyers with (Myerson's) regular distributions and an $\widetilde{O}(\mathrm{poly}(n)T^{2/3})$ regret for buyers with general distributions, both of which are tight in the number of rounds T. Our result for regular distributions was previously not known even for the single-buyer setting and relies on a new *half-concavity* property of the revenue function in the value space. For n sequential buyers, our technique is to run a generalized single-buyer algorithm for all the buyers and to carefully bound the regret from the sub-optimal pricing of the suffix buyers.

The full version of the paper is available at https://arxiv.org/abs/2312.12794.

CCS Concepts: • **Theory of computation → Computational pricing and auctions**; **Online learning theory**.

Additional Key Words and Phrases: bandit learning, sequential posted pricing, regular distributions

ACM Reference Format:
Sahil Singla and Yifan Wang. 2024. Bandit Sequential Posted Pricing via Half-Concavity. In *Conference on Economics and Computation (EC '24), July 8–11, 2024, New Haven, CT, USA.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3670865.3673495

*Supported in part by NSF award CCF-2327010.

Authors' Contact Information: Sahil Singla, ssingla@gatech.edu, School of Computer Science, Georgia Institute of Technology, Atlanta, GA, USA; Yifan Wang, ywang3782@gatech.edu, School of Computer Science, Georgia Institute of Technology, Atlanta, GA, USA.

EC '24, July 8–11, 2024, New Haven, CT, USA

ACM ISBN 979-8-4007-0704-9/24/07
https://doi.org/10.1145/3670865.3673495

1 Introduction

Sequential Posted Pricing (SPP) schemes are well-studied in mechanism design because of their simplicity/convenience in practice and their tractability in theory. In the basic setting, a seller wants to sell a single item to a group of buyers. The buyers arrive one-by-one and the seller presents a take-it-or-leave-it price. The first buyer with a value higher than the posted price takes the item by paying the price. The benefit of SPP is that it gives approximately-optimal revenue while being "simple". For instance, although it is known that Myerson's mechanism [Myerson, 1981] is optimal for selling a single item to n buyers with regular distributions, this mechanism is often impractical. This is because the winner's payment is defined via "complicated" virtual valuations and all buyer bids need to be simultaneously revealed, which is not possible in large markets. In contrast, SPP are known to give at least 1/2-fraction of the optimal revenue for regular buyers, while having fixed prices and buyers arriving one-by-one (see the books [Hartline, 2013, Roughgarden, 2017] for regular buyers, and [Yan, 2011] for a discussion on how ironing extends this result to general buyers). For multi-item (multi-parameters) setting, the difference between optimal and SPP mechanisms is even more stark. For instance, the optimal mechanism to sell n items to a single unit-demand bidder is known to be impossible (unless $P^{NP} = P^{\#P}$) [Chen et al., 2015], but we can use SPP to obtain 1/4-fraction of the optimal revenue [Chawla et al., 2007, 2010, 2015].

Sample Complexity. A common assumption among initial works on revenue-maximization was that the underlying distributions of the buyers are known to the seller. This is unrealistic in many applications since the distributions are unknown and need to be learnt from historical data. Inspired by this, a recent line of research studies the sample complexity for different types of mechanisms, i.e., how many samples are sufficient to learn an ϵ-optimal mechanism. For instance, starting with the pioneering work in [Balcan et al., 2008, Cole and Roughgarden, 2014], several papers studied the sample complexity of Myerson's mechanism [Devanur et al., 2016, Gonczarowski and Nisan, 2017, Huang et al., 2018, Roughgarden and Schrijvers, 2016], and finally [Guo et al., 2019] obtained the tight sample complexity bounds for regular buyers and for $[0, 1]$ bounded-support buyers. Sample complexity of many other classes of mechanisms have been studied, e.g., [Balcan et al., 2018, Jin et al., 2023] study second-price auctions, [Balcan et al., 2021, 2018] study posted-pricing mechanisms, [Guo et al., 2021] study strongly-monotone auctions (which includes Myerson and SPP), and [Cai and Daskalakis, 2017, Cai et al., 2022, Gonczarowski and Weinberg, 2021, Morgenstern and Roughgarden, 2016] study the sample complexity of multi-parameter auctions.

Bandit Learning. A stronger model of learning than the sample complexity model is the well-studied bandit learning model; see books [Bubeck and Cesa-Bianchi, 2012, Lattimore and Szepesvári, 2020]. In this model, the seller interacts with the buyers over T days. On each day, the seller proposes a parameterized mechanism to the buyer and sees only the revenue as feedback. The goal is to minimize the total regret, which is the difference between the total revenue achieved by the learning algorithm and the optimal mechanism over T days. It is stronger than sample complexity in two aspects: Firstly, it requires the learner to continually play near-optimal mechanisms, whereas in sample complexity the learner may lose a lot of the revenue while learning. Secondly, the feedback in the bandit model on any day is limited to only the proposed mechanism, and not every possible mechanism that could have been proposed.

The study of learning mechanisms with bandit feedback goes back to at least [Kleinberg and Leighton, 2003], where the authors provide an $\widetilde{O}(T^{2/3})$ regret bound for learning single-buyer posted pricing mechanism. Other examples of learning auctions in this model include bandits with knapsacks [Badanidiyuru et al., 2018, Immorlica et al., 2022], reserve price for i.i.d buyers [Cesa-Bianchi et al., 2015], and non-anonymous reserve prices [Niazadeh et al., 2021].

Despite a lot of work on learning SPP mechanisms and on learning auctions with bandit feedback, learning SPP in the bandit model was unknown. We answer this question and provide near-optimal regret bounds for SPP with n buyers having regular or general distributions.

1.1 Model and Results

In *Sequential Posted Pricing* (SPP), there are n sequential buyers with independent valuation distributions $\mathcal{D}_1, \cdots, \mathcal{D}_n$. We make the standard normalization assumption that each distribution $\mathcal{D}_i$ is supported in $[0, 1]$. The seller plays a group of prices $(p_1, \cdots, p_n) \in [0, 1]^n$ and then the buyers arrive one-by-one with valuation $v_i \sim \mathcal{D}_i$. The first buyer with $v_i \geq p_i$ takes the item and pays the price p_i as the revenue. The goal of the seller is to play prices to maximize the expected revenue. We define $R(p_1, \cdots, p_n)$ to be the expected revenue when playing prices $(p_1, \cdots, p_n)$, i.e.,

$$R(p_1, \cdots, p_n) \;:=\; \textstyle\sum_{i=1}^{n} p_i \cdot \mathbf{Pr}\left[i \text{ gets the item}\right] \;=\; \sum_{i=1}^{n} p_i \cdot (1 - F_i(p_i)) \prod_{j=1}^{i-1} F_j(p_j),$$

where $F_i(x)$ is the cumulative distribution function (CDF) of distribution $\mathcal{D}_i$. For SPP, the optimal prices are defined as

$$(p_1^*, \cdots, p_n^*) \;:=\; \operatorname{argmax}_{(p_1, \cdots, p_n)} R(p_1, \cdots, p_n).$$

These optimal prices can be easily calculated in polynomial time using a reverse dynamic program: price $p_n^* = \operatorname{argmax}_p p(1 - F_n(p))$, and with known optimal prices $p_{i+1}^*, \cdots, p_n^*$, we can calculate

$$p_i^* \;=\; \operatorname*{argmax}_p p(1 - F_i(p)) + F_i(p) \cdot \mathbf{E}\left[\text{revenue from } i+1, \cdots, n\right].$$

Bandit Sequential Posted Pricing (BSPP). In many practical applications of SPP, the valuation distributions $\mathcal{D}_i$ are unknown, so we will study them in the bandit learning model. Consider the following toy example to motivate the model: Suppose you want to sell an item each day for the entire next month using SPP. For simplicity, assume that each day exactly 3 bidders arrive: one each in the morning, afternoon, and evening. If you know the value distributions of these 3 bidders on each day, then you can find the optimal prices (p_1^*, p_2^*, p_3^*) as discussed above, and play them every day to maximize the total revenue. However, if the distributions are unknown then you need to learn the optimal prices during the month. A major challenge is that you don't even get to see the true valuations of the 3 bidders who show up each day, only whether they decide to buy the item, or not, at the played price.

Formally, BSPP is a T rounds/days repeated game where on each day $t \in [T]$ the seller proposes a group of n prices $(p_1^{(t)}, \cdots, p_n^{(t)})$. The buyers then arrive one-by-one with valuations $v_i^{(t)} \sim \mathcal{D}_i$ and the first buyer i with valuation $v_i^{(t)} \geq p_i^{(t)}$ takes the item. The seller receives revenue $\text{Rev}^{(t)} = p_i^{(t)}$ as the reward for this day and sees $p_i^{(t)}$ as the feedback, or equivalently sees the identity of the buyer that takes the item[1]. Note that none of the buyer valuations are ever revealed. If there is no buyer with a valuation higher than their price, the seller receives reward 0. The goal of the seller is to minimize in expectation the *total regret*:

$$\text{Regret} \;:=\; T \cdot R(p_1^*, \cdots, p_n^*) - \textstyle\sum_{t \in [T]} \text{Rev}^{(t)}.$$

Our first main result gives a near-optimal regret algorithm for BSPP with regular buyers.

Main Result 1 (Informal Theorem 3.1)**.** *For bandit sequential posted pricing with n buyers having regular distribution inside* $[0, 1]$*, there exists an algorithm with* $\widetilde{O}(\text{poly}(n)\sqrt{T})$ *regret.*

[1]By adding an arbitrarily small noise to each price, the seller can retrieve buyer's identity from the revenue.

It should be noted that this is the first bandit learning algorithm for regular distributions with optimal regret bound in T, even for the single-buyer or i.i.d. buyers setting. A very recent paper of [Leme et al., 2023b] studies "pricing query complexity" of this problem in the special case of a single-buyer. Their model is weaker than our bandit learning model since it only gives a *final regret* bound, i.e., one might incur a large regret during the T days of learning but the final learnt prices have a low 1-day regret. Hence, their $\Omega(1/\epsilon^2)$ lower bound for single-buyer query complexity immediately implies an $\Omega(\sqrt{T})$ lower bound for BSPP, showing tightness of the $\sqrt{T}$ factor in Main Result 1. However, their single-buyer pricing query complexity upper bounds do not apply to our stronger bandit learning model. Moreover, it's unclear how to extend their techniques beyond a single-buyer.

Our next result gives a near-optimal regret bound for BSPP with n buyers having general distributions. The authors of [Leme et al., 2023b] give an $\Omega(1/\epsilon^3)$ query complexity lower bound for single-buyer setting with general distributions, implying an $\Omega(T^{2/3})$ regret lower bound for BSPP with general distributions. Our result achieves a tight upper bound dependency on T.

Main Result 2 (Informal Theorem 4.1)**.** *For bandit sequential posted pricing problem with n buyers having value distribution inside* $[0, 1]$*, there exists an algorithm with* $\widetilde{O}(\text{poly}(n)T^{2/3})$ *regret.*

This result generalizes the single-buyer result of [Kleinberg and Leighton, 2003] to n sequential buyers. Interestingly, our techniques are very different from theirs since we are in the stochastic bandit model. Although [Kleinberg and Leighton, 2003] show an $\widetilde{O}(T^{2/3})$ regret algorithm even when the buyer valuations are chosen by an adversary, we observe in Section 5 that such a result is impossible for multiple buyers since already for 2 buyers with adversarial valuations, every online algorithm incurs $\Omega(T)$ regret.

1.2 Techniques

Single Regular Buyer via Half-Concavity. The proof of our first main result is based on a key observation that the revenue curve of a regular distribution is "half-concave". Recall that a distribution is regular if its revenue curve is concave in the quantile space. Simple examples show that regular distributions need not be concave in the value space (e.g., exponential distributions). Our half-concavity shows that regular distributions are still concave on one side of its maximum.

Main Result 3 (Informal Lemma 2.4)**.** *Let* $R(p)$ *be the revenue function of a regular distribution supported on* $[0, 1]$ *and let* $p^* = \text{argmax}_p R(p)$*. Then,* $R(p)$ *is concave in* $[0, p^*]$*.*

Concavity is a strong property that often allows efficient learning. We show that it's sufficient to learn a single-peaked function even if it is only half-concave. The high-level intuition for learning a half-concave follows from the standard recursive algorithm for learning concave functions: first, consider a (fully) concave function $R(x)$ defined on interval $[\ell, r]$. We set $a = \frac{2\ell+r}{3}$ and $b = \frac{\ell+2r}{3}$, and test $R(a)$ and $R(b)$ with sufficiently many samples. There can be three cases:

- Case 1: $R(a) < R(b)$. We can drop $[\ell, a]$ and recurse on $[a, r]$.
- Case 2: $R(a) > R(b)$. We can drop $[b, r]$ and recurse on $[\ell, b]$.
- Case 3: $R(a) \approx R(b)$. Concavity implies that $R(x)$ is nearly a constant for all $x \in [\ell, r]$, so we are done.

Now suppose $R(x)$ is only half-concave. We can still perform a similar algorithm: the first two cases use single-peakedness, so they remain the same. For the third case, half-concavity still guarantees that $R(x)$ can't be too high for $x \in [b, r]$, so we can drop $[b, r]$ and recurse on $[\ell, b]$.

Our final algorithm and proofs combine the above Case 2 and 3; see details in Section 2.

Generalizing Single-Buyer to Sequential Buyers. In the case of sequential buyers, our main idea is to run the single-buyer algorithm for all n buyers. A major difference compared to the single buyer setting is that the revenue function for this buyer becomes $R(p) = p \cdot (1 - F(p)) + C \cdot F(p)$, where $C \cdot F(p)$ represents the case that this buyer does not take the item and the seller receives an expected revenue C from the buyers after the current buyer. Therefore, the main for the generalization is to take care of this extra $C \cdot F(p)$ term. We show that for regular distributions the half-concavity still holds for the new function $R(p)$ in the interval $[C, 1]$, along with some other nice properties in the interval $[0, C]$. For general distributions, we show that the extra $C \cdot F(p)$ term does not make a huge difference.

1.3 Further Related Work

Bandit Learning and Single-Buyer Posted Pricing. The bandit learning model is well-established and we refer the readers to these books for classical results [Bubeck and Cesa-Bianchi, 2012, Cesa-Bianchi and Lugosi, 2006, Lattimore and Szepesvári, 2020]. The recent book of [Slivkins, 2019] is a great reference for work at the intersection of bandits and economics. In particular, the problem of learning the single-buyer posted pricing mechanism with bandit feedback has a long history, dating back to [Kleinberg and Leighton, 2003]. In their paper, the authors put forth an $\widetilde{O}(T^{2/3})$ regret bound for general buyer distributions and an $\widetilde{O}(\sqrt{T})$ regret bound under a non-standard assumption that the revenue curve is single-peaked, and its second derivative at this peak is a *strictly negative constant*, independent of the parameter T. This result is not directly comparable to our $\widetilde{O}(\sqrt{T})$ bound for regular distributions. This is primarily because the second derivative at the maximum of the revenue curve for a regular distribution can range from negligibly small to zero. Additionally, [Cesa-Bianchi et al., 2019] proposes an $\widetilde{O}(\sqrt{KT})$ regret bound, assuming the buyer's value resides within a discrete set with cardinality K. This result, however, is also not directly comparable to ours, as regular distributions inherently exhibit continuity, contrasting with the assumption of discrete value sets.

Threshold Query Model. Recently, the threshold query structure that underpins the bandit feedback model of the single-buyer posted pricing problem is extensively studied. In this model, the learner queries a threshold τ and only observes $\mathbf{1}[\tau < X \sim \mathcal{D}]$. The goal is to determine the minimum number of queries to learn a key parameter (e.g., median, mean, or CDF) of a distribution. We refer the readers to [Leme et al., 2023a,b, Meister and Nietert, 2021, Okoroafor et al., 2023] for learning complexity of the threshold query model.

Sequential Posted Pricing and Prophet Inequality. Sequential posted pricing (SPP) and its variants have been long popular, both in theory and in practice. One of the first results in their theoretical study is that posted prices obtain 78% of the optimal revenue for selling a single item in a large market [Blumrosen and Holenstein, 2008]. Subsequently, posted prices have been successfully analyzed in both single-dimensional and multi-dimensional settings; see this survey [Lucier, 2017] and book [Hartline, 2013]. A beautiful paper on the popularity of posted prices in practice is [Einav et al., 2018].

A closely related problem to SPP is the Prophet Inequality (PI) problem from optimal stopping theory. In both these problems a sequence of n independent buyers arrive one-by-one. However, in SPP we want to maximize the revenue and in PI we want to maximize the welfare. Interestingly, for known buyer value distributions, both these problems are equivalent [Correa et al., 2019], but this reduction requires virtual-value distributions and doesn't work for unknown distributions. In the bandit model, PI and SPP behave very differently. For instance, [Gatmiry et al., 2024] recently obtained $\widetilde{O}(\sqrt{T})$ regret algorithm for PI with general distributions, whereas an $\Omega(T^{2/3})$ regret lower

bound exists for SPP [Leme et al., 2023b]. Our $\widetilde{O}(\sqrt{T})$ regret results crucially rely on half-concavity of regular distributions.

Regular Distributions and Learning. Myerson's regularity of distributions has been greatly studied since it was introduced in [Myerson, 1981]. In particular, it is a standard assumption in learning theory for auctions [Cole and Roughgarden, 2014, Devanur et al., 2016, Dhangwatnotai et al., 2015, Guo et al., 2019, Huang et al., 2018, Roughgarden and Schrijvers, 2016]. For basic properties of regular distribution, and its important subclass of Monotone Hazard Rate distributions, see the books [Hartline, 2013, Roughgarden, 2017]. A recent paper of [Leme et al., 2023b], which studies pricing query complexity of the single-buyer single-item problem, proves an interesting "relative flatness" property of regular distributions in the value space. Roughly it says that if the revenue curve has nearly the same value at 4 different equidistant points then there cannot be a high revenue point in between. Comparing it to our idea of half-concavity, the two properties are in general incomparable as no one implies the other. However, we found our half-concavity to be more intuitive and convenient to work with, since proofs based on relative flatness often lead to a long case analysis.

2 A Single Buyer with Regular Distribution

In this section, we present an $\widetilde{O}(\sqrt{T})$ regret algorithm for BSPP in the special case of a single buyer, where we call the problem Bandit Posted Pricing. Moreover, we focus on the case when the buyer's value distribution is *regular*, which is a standard assumption in economics [Myerson, 1981].

Definition 2.1 (Regularity)**.** *Distribution $\mathcal{D}$ with CDF $F(x)$ and PDF $f(x)$ is called* regular *when $\phi(v) := v - \frac{1-F(v)}{f(v)}$ is monotone non-decreasing, or equivalently, its* revenue curve *$R_q(q)$ is concave in the quantile space, where*

$$R_q(q) := q \cdot F^{-1}(1-q).$$

In the *Bandit Posted Pricing* problem there is a single regular buyer with an unknown regular value distribution $\mathcal{D}$ having support $[0,1]$ and CDF $F(x)$. Our goal is to approach the optimal price p^* that maximizes the revenue function $R(p) := p \cdot (1 - F(p))$ in the the following bandit learning game over T days: On day $t \in \{1, \ldots, T\}$, we post a price $p_t \in [0,1]$ and the environment draws $v_t \sim \mathcal{D}$. Our reward is $p_t \cdot \mathbf{1}_{v_t \geq p_t}$, and the goal is to minimize in expectation the total *regret*: $T \cdot R(p^*) - \sum_{t\in[T]} p_t \cdot \mathbf{1}_{v_t \geq p_t}$. Our $\widetilde{O}(\sqrt{T})$ regret algorithm for this problem uses "half-concavity".

2.1 Half-Concavity

Our $\widetilde{O}(\sqrt{T})$ regret algorithm works beyond regular distributions, for the class of *half-concave* distributions. We first give the definition of half-concavity:

Definition 2.2 (Half-Concavity)**.** *A function $R(x) : \mathbb{R} \to \mathbb{R}$ is* half-concave *in interval $[\ell, r] \subseteq [0,1]$ if the following conditions hold:*

(i) $R(x)$ is single-peaked in $[\ell, r]$, i.e., $\exists p^* \in [\ell, r]$ satisfying that $R(x)$ is non-decreasing in $[\ell, p^*]$ and non-increasing in $[p^*, r]$.
(ii) $R(x)$ is 1-Lipschitz in $[\ell, p^*]$.
(iii) $R(x)$ is concave in $[\ell, p^*]$.

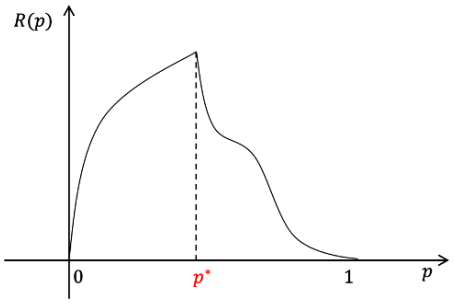

Fig. 1. An example of a half-concave function, where the function is Lipschitz and concave before p^*.

The concept of half-concavity implies that a function has nice properties properties on one side of the maximum (see Figure 1), rendering it learnable from this advantageous side. In the definition of half-concavity, both Lipschitzness and some kind of concavity are vital to ensure learnability. Without Lipschitzness, a function becomes unlearnable when p^* is exceedingly close to 0 due to the inability of accurately detecting the value of p^*. On the other hand, without concavity, simple examples show that a regret of $\Omega(T^{2/3})$ is inevitable. For instance, a revenue function characterized by multiple peaks where we can only ascertain the value of p^* when we examine each peak with a sufficient number of samples. We note that related works, such as [Combes and Proutière, 2014, Magureanu et al., 2014], address bandit problems with assumptions of unimodularity or Lipschitzness. However, these works only provide instance dependent regret bounds, which are at least $\omega(\sqrt{T})$ in the worst case, as they lack at least one of the key assumptions of Lipschitzness and concavity.

Given both Lipschitzness and (half) concavity, the following theorem gives $\widetilde{O}(\sqrt{T})$ regret.

Theorem 2.3. *For Bandit Posted Pricing with a single buyer, if the revenue function $R(p) := p \cdot (1 - F(p))$ is half-concave, then there exists an algorithm with $O(\sqrt{T} \log T)$ regret.*

The main application of Theorem 2.3 is for regular distributions due to the following lemma.

Lemma 2.4. *Let $F(x)$ be the CDF of a regular distribution with support in $[0, 1]$, and $R(p) := p \cdot (1 - F(p))$ be the expected revenue on playing price p. Then, $R(p)$ is half-concave in $[0, 1]$.*

Before proving this lemma, we observe the following immediate corollary.

Corollary 2.5 (Corollary of Theorem 2.3 and Lemma 2.4)**.** *For Bandit Posted Pricing with a single buyer having regular valuation distribution, there exists an algorithm with $O(\sqrt{T} \log T)$ regret.*

Proof of Lemma 2.4. We first provide a proof assuming the PDF $f(p)$ is non-zero and differentiable. We will prove the three properties of half-concavity one-by-one.
(i) *Single-peakedness*: By definition, revenue function of a regular distribution is concave in the quantile space. Hence, it's single-peaked in the quantile space. Suppose q^* is the quantile that achieves the maximum, then choosing price $p^* = F^{-1}(1 - q^*)$ proves single-peakedness of $R(p)$.

(ii) *Lipschitzness*: For $0 \le a \le b \le p^*$, we have

$$0 \;\le\; R(b) - R(a) \;=\; (b - a) - F(b) \cdot b + F(a) \cdot a \;\le\; b - a,$$

where the first inequality follows from single-peakedness and that $a \le b \le p^*$, and the second inequality uses $F(b) \ge F(a) \ge 0$ and that $b \ge a \ge 0$.

(iii) *Half-concavity*: We first provide a direct proof via the monotonicity of virtual valuation ϕ.

Recall that $\phi(p) := p - \frac{1-F(p)}{f(p)}$ is non-decreasing for regular distribution, so we have

$$\phi'(p) \;=\; 1 - \frac{-f^2(p) - (1-F(p)) \cdot f'(p)}{f^2(p)} \;\geq\; 0.$$

Rearranging the above inequality,

$$2f^2(p) + f'(p) \cdot (1 - F(p)) \;\geq\; 0. \tag{1}$$

Consider the derivative and the second order derivative of $R(p) = p \cdot (1 - F(p))$, i.e.,

$$R'(p) \;=\; 1 - F(p) - p \cdot f(p) \qquad \text{and} \qquad R''(p) \;=\; -2f(p) - p \cdot f'(p).$$

For half-concavity, we prove that $R''(p) \leq 0$ for every $p \in [0, p^*]$. Consider two cases:

- Case 1: $f'(p) > 0$. In this case, $R''(p) = -2f(p) - p \cdot f'(p) \leq 0$ immediately holds.
- Case 2: $f'(p) \leq 0$. In this case, we have

$$\begin{aligned} R''(p) \;&=\; -2f(p) - p \cdot f'(p) \\ &=\; \frac{1}{f(p)} \cdot \big(-2f^2(p) - f(p) \cdot p \cdot f'(p)\big) \\ &\leq\; \frac{1}{f(p)} \cdot (f'(p) \cdot (1 - F(p)) - f(p) \cdot p \cdot f'(p)) \\ &=\; \frac{f'(p)}{f(p)} \cdot \big(1 - F(p) - p \cdot f(p)\big) \;\leq\; 0, \end{aligned}$$

  where the first inequality is from (1) and the second inequality uses $f'(p) \leq 0$ along with the fact that $R'(p) \geq 0$ when $p \in [0, p^*]$.

The above proof requires the PDF $f(p)$ to be non-zero and differentiable. Below we provide an alternate proof via the concavity in the quantile space, which bypasses this assumption for $f(p)$. We keep both the proofs here since the reader may find one proof more instructive than the other.

An alternate proof for half-concavity. We prove half-concavity by contradiction. Assume there exist $0 \leq a < b < c \leq p^*$ satisfying $c = b + t(b - a)$ and $R(c) > R(b) + t \cdot \big(R(b) - R(a)\big)$.

Define $q(x) := 1 - F(x)$ and $\overline{q} := q(b) + t \cdot \big(q(b) - q(a)\big)$. We first show that $\overline{q} > q(c)$: Regularity implies that the revenue function is concave in the quantile space. Hence, for all $q' \geq \overline{q}$,

$$R(q^{-1}(q')) \;\leq\; R(b) + \frac{q(b) - q'}{q(a) - q(b)} \cdot (R(b) - R(a)) \;\leq\; R(b) + t \cdot (R(b) - R(a)) \;<\; R(c).$$

For every $q' \geq \bar{q}$, we have $q' \neq q(c)$. So there must be $q(c) < \bar{q}$. This gives the desired contradiction:

$$\begin{aligned} R(c) \;&>\; R(b) + t \cdot (R(b) - R(a)) \\ &=\; b \cdot q(b) + t \cdot \big(b \cdot q(b) - a \cdot q(a)\big) \\ &\geq\; b \cdot q(b) + t \cdot \big(b \cdot q(b) - a \cdot q(a)\big) - t(t+1)(b-a)(q(a) - q(b)) \\ &=\; \big((t+1)b - ta\big) \cdot \big((t+1)q(b) - tq(a)\big) \\ &=\; c \cdot \overline{q} \;\geq\; c \cdot q(c) \;=\; R(c). \end{aligned}$$

□

2.2 $\widetilde{O}(\sqrt{T})$ Regret for Half-Concave Functions: Proof of Theorem 2.3

Our algorithm relies on a subroutine that takes a parameter ϵ and a confidence interval $[\ell, r] \ni p^*$ as input, and then after $\widetilde{O}(\epsilon^{-2})$ rounds it generates with high probability a new confidence interval $[\ell', r']$ that contains the optimal price p^* and every price in $[\ell', r']$ has 1-day regret bounded by ϵ. Our final algorithm runs in $O(\log T)$ phases, where in each phase we call the sub-routine with the

parameter ϵ, and ϵ is halved in the next phase. The sub-routine is captured by the following lemma, which is the heart of the proof and will be proved in Section 2.3 using half-concavity of $R(p)$.

Lemma 2.6. *Let $R(p)$ be a half-concave revenue function defined in $[0,1]$. Given $\epsilon \geq 1/\sqrt{T}$ and $[\ell, r] \ni p^*$, there exists an algorithm that tests $O(\epsilon^{-2}\log^2 T)$ rounds with prices inside $[\ell, r - \frac{1}{T^{100}}]$ and outputs an interval $[\ell', r'] \subseteq [\ell, r]$ satisfying with probability $1 - T^{-5}$ that*

- *$p^* \in [\ell', r']$*
- *$R(x) \geq R(p^*) - \epsilon$ for any $x \in [\ell', r' - \frac{1}{T^{100}}]$.*

ALGORITHM 1: $O(\sqrt{T}\log T)$ Regret Algorithm

Input: Hidden Revenue Function $R(p)$, time horizon T
Let $\epsilon_1 \leftarrow 1$, $[\ell_1, r_1] \leftarrow [0,1]$, and $i \leftarrow 1$.
while $\epsilon_i > \frac{\log T}{\sqrt{T}}$ **do**
  Run the algorithm described in Lemma 2.6 with ϵ_i, $[\ell_i, r_i]$ as input, and get $[\ell', r']$.
  Let $[\ell_{i+1}, r_{i+1}] \leftarrow [\ell', r']$ and $\epsilon_{i+1} \leftarrow \frac{1}{2}\epsilon_i$.
  Let $i \leftarrow i+1$.
end
Finish remaining rounds with any price in $[\ell_i, r_i - T^{-100}]$.

Given Lemma 2.6, our algorithm for Theorem 2.3 is natural and has a simple proof.

Proof of Theorem 2.3. We claim that with probability at least $1 - T^{-4}$, the regret of Algorithm 1 is $O(\sqrt{T}\log T)$. Algorithm 1 uses Lemma 2.6 for multiple times with a halving error parameter. Assume Algorithm 1 ends with $i = k+1$, i.e., the while loop runs k times. Since we obtain $\epsilon_{i+1} = \epsilon_i/2$ and $\epsilon_k > \frac{\log T}{\sqrt{T}}$ holds, there must be $k = O(\log T)$.

For $i \in [k]$, let Alg_i represent the corresponding algorithm we call when using Lemma 2.6 with ϵ_i, $[\ell_i, r_i]$ as the input. To use Lemma 2.6 we need to verify that $p^* \in [\ell_i, r_i]$ for all $i \in [k]$. The condition $p^* \in [\ell_1, r_1] = [0,1]$ clearly holds. For $i = 2, 3, \cdots, k$, the condition $p^* \in [\ell_i, r_i]$ is guaranteed by Lemma 2.6 when calling Alg_{i-1}. The failing probability of Lemma 2.6 is T^{-5}. By the union bound, with probability $1 - k \cdot T^{-5} > 1 - T^{-4}$, Lemma 2.6 always holds in Algorithm 1.

Now we prove the regret of Algorithm 1. Lemma 2.6 guarantees that $R(p^*) - R(p) \leq \epsilon_i$ holds for all $p \in [\ell_{i+1}, r_{i+1} - T^{-100}]$, while when calling Alg_{i+1}, only the prices in $[\ell_{i+1}, r_{i+1} - T^{-100}]$ is tested. Therefore, the total regret of Algorithm 1 can be bounded by

$$1 \cdot O\big(\epsilon_1^{-2}\log^2 T\big) \;+\; \textstyle\sum_{i=2}^{k} \epsilon_{i-1} \cdot O(\epsilon_i^{-2}\log^2 T) \;+\; T \cdot \epsilon_k \;=\; O(\sqrt{T}\log T).$$

In this expression, the first term is the regret of Alg_1. The second term is the sum of the regret from $\mathsf{Alg}_2, \cdots, \mathsf{Alg}_k$. The third term is the regret for the remaining rounds.

Finally, we also need to check that Algorithm 1 uses no more than T rounds. Lemma 2.6 suggests that Alg_i runs $O(\epsilon_i^{-2}\log^2 T)$ rounds. So the total number of rounds in the while loop is $\sum_{i\in[k]} O(\epsilon_i^{-2}\log^2 T) = O(T)$. Therefore, Algorithm 1 is feasible. □

2.3 Main Sub-Routine: Proof of Lemma 2.6 via Half-Concavity

In this section, we present the main sub-routine of our bandit algorithm. It utilizes half-concavity to generate a confidence interval for the optimal price p^* in $\widetilde{O}(\frac{1}{\epsilon^2})$ rounds while ensuring that every price in the interval is ϵ-optimal.

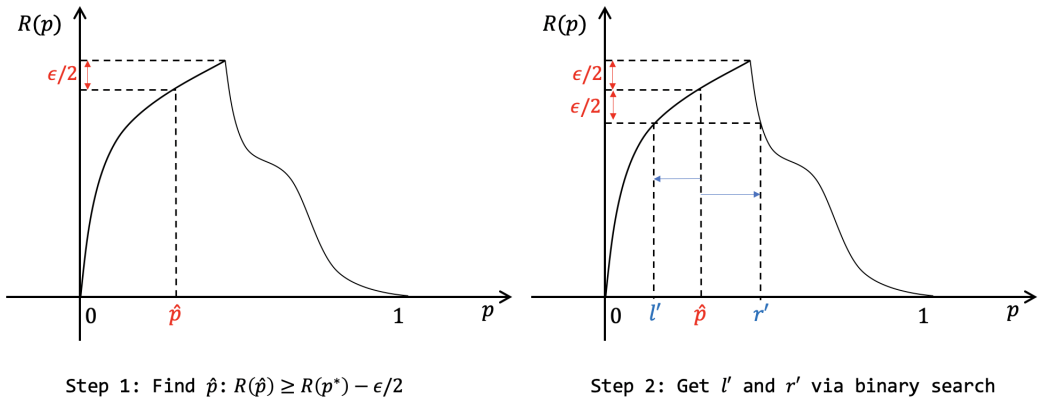

Fig. 2. The two main steps of Lemma 2.6

Algorithm Overview. The algorithm contains two major steps (also see Figure 2):

- Step 1: Find an approximation $\hat{p}$ such that $R(p^*) - R(\hat{p}) \le \frac{\epsilon}{2}$ via half-concavity. This step gives a sufficiently precise price estimate approximating p^*.
- Step 2: Given $\hat{p}$, construct a new confidence interval $[\ell', r']$ via binary search. As $R(p)$ exhibits a single-peak property, we can use $R(\hat{p})$ as a benchmark and implement a standard binary search algorithm to identify the leftmost point ℓ' and the rightmost point that are $\frac{\epsilon}{2}$-close to $R(\hat{p})$. These two endpoints describe the desired new confidence interval.

We note an important detail in Step 2: the Lipschitzness assumption in the range $[p^*, r]$ is absent. Consequently, the binary search algorithm fails to provide a satisfactory loss guarantee for a small tail of the new confidence interval. This introduces the non-standard error factor T^{-100} as discussed in Lemma 2.6.

2.3.1 Step 1: Find $\hat{p}$ to Approximate p^*. The first step of the sub-routine is to find $\hat{p}$ that approximates p^*. Specifically, we want to prove the following lemma:

Lemma 2.7. *Assume $R(p)$ is a half-concave revenue function defined in $[0, 1]$. Given interval $[\ell, r]$ and error parameter $\epsilon > 1/\sqrt{T}$, there exists an algorithm that tests $O(\epsilon^{-2}\log^2 T)$ rounds with prices inside $[\ell, r - T^{-100}]$ and outputs $\hat{p} \in [\ell, r - T^{-100}]$ satisfying with probability at least $1 - T^{-6}$ that $\max_{p\in[\ell,r]} R(p) - R(\hat{p}) < \frac{\epsilon}{2}$.*

When we have the assumption $p^* \in [\ell, r]$ from Lemma 2.6, Lemma 2.7 implies we are able to find $\hat{p}$ such that $R(p^*) - R(\hat{p}) \le \frac{\epsilon}{2}$ in the first step. We first give the pseudo-code of the algorithm required by Lemma 2.7 and its high-level idea.

Algorithm 2 is a recursive algorithm that runs $O(\log \frac{1}{\delta})$ rounds. In each round, the algorithm tests the two points a, b that divide the current interval $[\ell, r]$ into thirds. Each price is tested for $\widetilde{O}(\frac{1}{\delta^2})$ rounds, where $\delta = \epsilon/100$ is the scaled error parameter. A standard concentration inequality guarantee both $|R(a) - \hat{R}(a)| \le \delta$ and $|R(b) - \hat{R}(b)| \le \delta$ hold. Then, the algorithm drops one third of the interval according to the test results. There are two different cases (see Figure 3):
Case 1: $\hat{R}(a) < \hat{R}(b) - 2\delta$. The inequality is sufficient to show that $R(a) < R(b)$. Since function $R(p)$ is single-peaked, there must be $p^* \ge a$ and the sub-interval $[\ell, a]$ can be dropped.
Case 2: $\hat{R}(a) \ge \hat{R}(b) - 2\delta$. This inequality implies $R(b) - R(a)$ is sufficiently small. In this case, observe that if $p^* \in [b, r]$, the value of $R(p^*)$ can't be much better than $R(b)$, because the concavity

ALGORITHM 2: Finding $\hat{p}$

Input: Hidden Revenue Function $R(p)$, Interval $[\ell, r]$, Error Parameter ϵ
Let $\delta = \frac{\epsilon}{100}$ be the scaled error parameter, $\ell_1 \leftarrow \ell, r_1 \leftarrow r, i = 1$.
while $r_i - \ell_i > \delta$ **do**
  Let $a_i \leftarrow (2\ell_i + r_i)/3$ and $b_i \leftarrow (\ell_i + 2r_i)/3$.
  Test $C \cdot \delta^{-2}\log T$ rounds with price $p = a_i$. Let $\hat{R}(a_i)$ be the average reward.
  Test $C \cdot \delta^{-2}\log T$ rounds with price $p = b_i$. Let $\hat{R}(b_i)$ be the average reward.
  if $\hat{R}(a_i) < \hat{R}(b_i) - 2\delta$ **then**
    Let $\ell_{i+1} \leftarrow a_i, r_{i+1} \leftarrow r_i$
  else
    Let $\ell_{i+1} \leftarrow \ell_i, r_{i+1} \leftarrow b_i$
  end
  $i \leftarrow i + 1$
end
Test $C \cdot \delta^{-2}\log T$ rounds with price $p = \ell_i$. Let $\hat{R}(\ell_i)$ be the average reward.
Let $\hat{p} = \arg\max_{p \in P} \hat{R}(p)$, where $P = \{p : p \text{ is tested}\}$.
Output: $\hat{p}$.

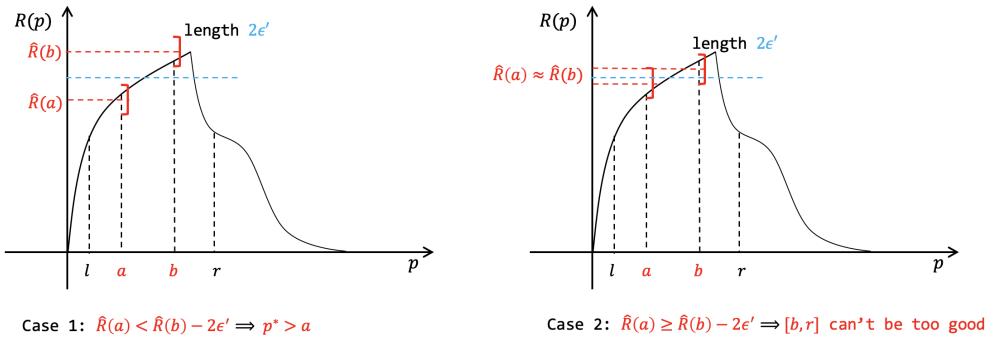

Fig. 3. Two cases of Step 1 as discussed in Algorithm 2.

in $[a, p^*]$ guarantees that $R(p^*) - R(b) \leq R(b) - R(a)$. Therefore, the sub-interval $[b, r]$ can be dropped because we don't need to further test it.

Finally, the recursive algorithm stops when the length of the remaining interval is bounded by δ, and the Lipschitzness guarantees the final interval is good.

Now, we give the formal proof of Lemma 2.7.

Proof of Lemma 2.7. We first show that Algorithm 2 tests $O(\frac{\log^2 T}{\epsilon^2})$ prices in $[\ell, r - T^{-100}]$.

The main body of Algorithm 2 is a while loop that maintains an interval to be tested. Assume the while loop stops when $i = k + 1$. Then, for $i \in [k]$, the length of $[\ell_{i+1}, r_{i+1}]$ is two-thirds of $[\ell_i, r_i]$, and we have $r_k - \ell_k > \delta$. Therefore, $k = O(\log \frac{1}{\delta}) = O(\log T)$. For interval $[\ell_i, r_i]$, two prices a_i, b_i are tested for $O(\delta^{-2}\log T)$ times, so in total Algorithm 2 tests $O(\delta^{-2}\log^2 T)$ rounds. To show every tested price is in $[\ell, r - T^{-100}]$, let p be a price tested in Algorithm 2. $p \in [\ell, r]$ directly follows from the algorithm. For the inequality $p \leq r - T^{-100}$, observe that the largest tested price must be $p = b_i$ for some $i \in [k]$, and $b_i \leq r - T^{-100}$ is guaranteed by the fact that $r_i - b_i \geq \frac{1}{3}\delta \gg T^{-100}$.

It only remains to show $\max_{p \in [\ell, r]} -R(\hat{p}) \leq \frac{\epsilon}{2}$. Define $p^*_{\ell,r} := \arg\max_{p \in [\ell, r]} R(p)$ to be the optimal price in $[\ell, r]$. We will show that $R(p^*_{\ell,r}) - R(\hat{p}) \leq 6\delta < \frac{\epsilon}{2}$ holds with probability $1 - T^{-6}$. We first need the following claim, which follows from standard concentration inequalities:

Claim 2.8. *In Algorithm 2, $|\hat{R}(p) - R(p)| \le \delta$ simultaneously holds for every tested price p with probability at least $1 - T^{-6}$.*

Proof. For a single tested price p, $\hat{R}(p)$ is estimated by calculating the average of $N = C \cdot \frac{\log T}{\delta^2}$ samples. By Hoeffding's Inequality,

$$\mathbf{Pr}\left[|\hat{R}(p) - R(p)| > \delta\right] \le 2\exp\left(-2N\delta^2\right) = 2T^{-2C} < T^{-7}.$$

The last inequality holds when C is a constant greater than 4. Then, we have $|\hat{R}(p) - R(p)| \le \delta$ holds with probability $1 - T^{-7}$ for a single tested price p.

Notice that Algorithm 2 can't test more than T prices. By the union bound, $|\hat{R}(p) - R(p)| \le \delta$ simultaneously holds for all tested prices with probability $1 - T^{-6}$. □

Next, we use Claim 2.8 to show $R(p^*_{\ell,r}) - R(\hat{p}) \le 6\delta$. We consider the following two cases:

The first case is that there exists $i \in [k]$, such that $p^*_{\ell,r}$ falls in $[\ell_i, r_i]$ but not in $[\ell_{i+1}, r_{i+1}]$. In this case, there must be $\hat{R}(a_i) \ge \hat{R}(b_i) - 2\delta$ and $[b_i, r_i]$ is dropped. If not, we have $\hat{R}(a_i) \ge \hat{R}(b_i) - 2\delta$ and $[\ell_i, a_i]$ is dropped. However, the inequality $\hat{R}(a_i) \ge \hat{R}(b_i) - 2\delta$ together with Claim 2.8 gives $R(a_i) < R(b_i)$. Then, the single-peakedness of $R(p)$ gives $p^*_{\ell,r} \ge a_i$, which is in contrast to the assumption that $p^*_{\ell,r} \notin [\ell_{i+1}, r_{i+1}]$. Therefore, we have $\hat{R}(a_i) \ge \hat{R}(b_i) - 2\delta$, and the assumption $p^*_{\ell,r} \notin [\ell_{i+1}, r_{i+1}]$ implies $p^*_{\ell,r} \in [b_i, r_i]$. Then, the concavity of $R(p)$ in $[a_i, p^*_{\ell,r}]$ guarantees

$$R(p^*_{\ell,r}) - R(b_i) \le R(b_i) - R(a_i) \le (\hat{R}(b_i) + \delta) - (\hat{R}(a_i) - \delta) \le 4\delta.$$

The second case is that the desired i in the first case doesn't exist. In this case, the only possibility is that $p^*_{\ell,r} \in [\ell_{k+1}, r_{k+1}]$. Then, the 1-Lipschitzness of function $R(p)$ guarantees that $R(p^*_{\ell,r}) - R(\ell_k) \le \delta$.

In both cases, we find a tested price p' satisfying $R(p') \ge R(p^*_{\ell,r}) - 4\delta$. Then,

$$R(\hat{p}) \ge \hat{R}(\hat{p}) - \delta \ge \hat{R}(p') - \delta \ge R(p') - 2\delta \ge R(p^*_{\ell,r}) - 6\delta.$$ □

2.3.2 Step 2: Generating New Confidence Interval. Upon executing Algorithm 2, we obtain an approximate optimal price $\hat{p}$. The subsequent stage of the algorithm uses $R(\hat{p})$ as a benchmark to establish the new upper and lower bounds of the confidence interval. Given that the function $R(p)$ is single-peaked, two independent binary search algorithms suffice to independently determine these new upper and lower bounds, leading to the following Lemma 2.9:

Lemma 2.9. *Assume $R(p)$ is a half-concave revenue function defined in $[0, 1]$. Given interval $[\ell, r] \ni p^*$, error parameter $\epsilon > \frac{1}{\sqrt{T}}$, near-optimal price $\hat{p} \in [\ell, r - T^{-100}]$ that satisfies $R(p^*) - R(\hat{p}) \le \frac{\epsilon}{2}$, there exists an algorithm that tests $O(\epsilon^{-2}\log^2 T)$ rounds with prices in $[\ell, r - T^{-100}]$ and outputs $[\ell', r'] \subseteq [\ell, r]$ satisfying with probability $1 - T^{-6}$ that:*

- *$p^* \in [\ell', r']$.*
- *For $p \in [\ell', r' - T^{-100}]$, we have $R(p^*) - R(p) \le \epsilon$.*

Combining Lemma 2.7 and Lemma 2.9 proves Lemma 2.6:

Proof of Lemma 2.6. We first run Algorithm 2 to get $\hat{p}$, and then run the algorithm described in Lemma 2.9 with $\hat{p}$ being part of the input. Combining Lemma 2.7 and Lemma 2.9 with union bound guarantees that we have the desired output with success probability at least $1 - 2T^{-6} > 1 - T^{-5}$. □

It only remains to prove Lemma 2.9. We first give the following Algorithm 3, and then show Algorithm 3 is the desired algorithm for Lemma 2.9.

We first give the following concentration bound for Algorithm 3:

We first give the following claim:

ALGORITHM 3: Getting $[\ell', r']$ via Binary Search

Input: Hidden Revenue Function $R(p)$, Interval $[\ell, r]$, Error Parameter ϵ, Near-Optimal price $\hat{p}$
Let $\delta = \frac{\epsilon}{100}$ be the scaled error parameter.
// First binary search to determine ℓ'
Test $C \cdot \frac{\log T}{\delta^2}$ rounds with price $p = \hat{p}$. Let $\hat{R}(\hat{p})$ be the average reward.
Let $\ell_{b1} \leftarrow \ell, r_{b1} \leftarrow \hat{p}$.
while $r_{b1} - \ell_{b1} \geq T^{-100}$ **do**
  Let $m = \frac{\ell_{b1} + r_{b1}}{2}$.
  Test $C \cdot \frac{\log T}{\delta^2}$ rounds with price $p = m$. Let $\hat{R}(m)$ be the average reward.
  if $\hat{R}(m) < \hat{R}(\hat{p}) - 2\delta$ **then** Update $r_{b1} \leftarrow m$ **else** Update $\ell_{b1} \leftarrow m$;
end
Let $\ell' \leftarrow \ell_{b1}$
// Second binary search to determine r'
Test $C \cdot \frac{\log T}{\delta^2}$ rounds with price $p = r - T^{-100}$. Let $\hat{R}(r - T^{-100})$ be the average reward.
if $\hat{R}(r - T^{-100}) \geq \hat{R}(\hat{p}) - 2\delta$ **then**
  // Special Case: original upper bound is good
  Let $r' \leftarrow r$
else
  Let $\ell_{b2} \leftarrow \hat{p}, r_{b2} \leftarrow r - T^{-100}$.
  while $r_{b2} - \ell_{b2} \geq T^{-100}$ **do**
    Let $m = \frac{\ell_{b2} + r_{b2}}{2}$.
    Test $C \cdot \frac{\log T}{\delta^2}$ rounds with price $p = m$. Let $\hat{R}(m)$ be the average reward.
    if $\hat{R}(m) < \hat{R}(\hat{p}) - 2\delta$ **then** Update $r_{b2} \leftarrow m$ **else** Update $\ell_{b2} \leftarrow m$;
  end
  Let $r' \leftarrow r_{b2}$.
end
Output: $[\ell', r']$.

Claim 2.10. *In Algorithm 3, $|\hat{R}(p) - R(p)| \leq \delta|$ simultaneously holds for every tested price p with probability at least $1 - T^{-6}$.*

Proof. For a single tested price p, $\hat{R}(p)$ is estimated by calculating the average of $N = C \cdot \frac{\log T}{\delta^2}$ samples. By Hoeffding's Inequality,

$$\mathbf{Pr}\left[|\hat{R}(p) - R(p)| > \delta\right] \leq 2\exp\left(-2N\delta^2\right) = 2T^{-2C} < T^{-7}.$$

The last inequality holds when C is a constant greater than 4. Then, we have $|\hat{R}(p) - R(p)| \leq \delta$ holds with probability $1 - T^{-7}$ for a single tested price p.

Notice that Algorithm 3 can't test more than T prices. By the union bound, $|\hat{R}(p) - R(p)| \leq \delta$ simultaneously holds for all tested prices with probability $1 - T^{-6}$. □

Now we prove Lemma 2.9 assuming Claim 2.10 holds.

Proof of Lemma 2.9. We show Algorithm 3 satisfies the statements in Lemma 2.9. Our proof starts from verifying Algorithm 3 runs $O(\frac{\log T}{\delta^2})$ rounds with prices in $[\ell, r - T^{-100}]$. For the number of tested rounds, Algorithm 3 runs two standard binary searches. In one binary search, the subroutine tests one price with $C \cdot \frac{\log T}{\delta^2}$ rounds, and the binary search stops when the length of the interval is less than T^{-100}. Therefore, The total number of rounds is $C \cdot \frac{\log T}{\delta^2} \cdot O(\log T^{100}) = O(\frac{\log^2 T}{\epsilon^2})$. For the

range constraint of all tested prices, Algorithm 3 directly guarantees that every tested price falls in $[\ell, r - T^{-100}]$.

Next, we show $p^* \in [\ell', r']$. We first prove $p^* \geq \ell'$. In Algorithm 3, the only way we update ℓ' is that we observe $\hat{R}(m) < \hat{R}(\hat{p}) - 2\delta$ in the first binary search, and ℓ' is updated to at least m. This is feasible because Claim 2.10 gives $R(m) < R(\hat{p}) \leq R(p^*)$, and the single-peakedness of $R(p)$ guarantees $m \leq p^*$. The proof of $p^* \leq r'$ is symmetric.

Finally, we show $R(p) \geq R(p^*) - \epsilon$ for all $p \in [\ell', r' - T^{-100}]$. Function $R(p)$ is single-peaked, so it is sufficient to prove that $R(\hat{p}) - R(\ell') \leq 5\delta < \frac{\epsilon}{2}$ and $R(\hat{p}) - R(r' - T^{-100}) \leq 5\delta < \frac{\epsilon}{2}$. Then, combining these two inequalities with the assumption $R(p^*) - R(\hat{p}) \leq \frac{\epsilon}{2}$ gives the desired statement.

We first prove that $R(\hat{p}) - R(\ell') \leq 5\delta$. The binary search subroutine guarantees that $\hat{R}(r_{b1}) \geq \hat{R}(\hat{p}) - 2\delta$ after the binary search loop is finished. Then, Claim 2.10 ensures that $R(\hat{p}) - R(r_{b1}) \leq 4\delta$. Since ℓ' is set to be $\ell_{b1} > r_{b1} - T^{-100}$, the 1-Lipschitzness of $R(p)$ in $[\ell_{b1}, \hat{p}]$ guarantees that

$$R(\hat{p}) - R(\ell') \;=\; R(\hat{p}) - R(r_{b1}) + R(r_{b1}) - R(\ell_{b1}) \;\leq\; 4\delta + T^{-100} \;\leq\; 5\delta.$$

For the inequality $R(\hat{p}) - R(r' - T^{-100}) \leq 5\delta$, a symmetric proof for the second binary search subroutine gives $R(\hat{p}) - R(\ell_{b2}) \leq 4\delta$, while r' is set to be $r_{b2} < \ell_{b2} + T^{-100}$. If $p^* < r' - T^{-100}$, the single-peakedness of $R(p)$ gives

$$R(r' - T^{-100}) \;\geq\; R(\ell_{b2}) \;\geq\; R(\hat{p}) - 4\delta.$$

Otherwise, the 1-Lipschitzness gives

$$R(r' - T^{-100}) \;\geq\; R(\ell_{b2}) - (\ell_{b2} - r_{b2} + T^{-100}) \;\geq\; R(\hat{p}) - 4\delta - T^{-100} > R(\hat{p}) - 5\delta. \qquad \square$$

3 n Buyers with Regular Distributions

In this section, we provide the $\widetilde{O}(\mathrm{poly}(n)\sqrt{T})$ regret algorithm for BSPP with n buyers having regular distributions. In this problem we have n buyers with independent regular value distributions $\mathcal{D}_1, \cdots, \mathcal{D}_n$. The seller posts prices $p_1, \cdots, p_n$, and the first buyer with value $v_i \sim \mathcal{D}_i \geq p_i$ gets the item and pays p_i. Our goal is to approach the optimal prices $p_1^*, \cdots, p_n^*$ that maximizes the expected revenue $R(p_1, \cdots, p_n) := p_i \cdot \mathbf{Pr}\,[i \text{ gets the item}]$ via a bandit learning game over T days: On day $t \in [T]$ we post a price vector $(p_1^{(t)}, \cdots, p_n^{(t)})$ and the environment draws $v_1^{(t)} \sim \mathcal{D}_1, \cdots, v_n^{(t)} \sim \mathcal{D}_n$. We only observe reward $\mathrm{Rev}_t = \sum_{i\in[n]} p_i^{(t)} \cdot \mathbf{1}[p_i^{(t)} \geq v_i^{(t)} \wedge \forall j < i, p_j^{(t)} < v_j^{(t)}]$. Our goal is to minimize the total regret $T \cdot R(p_1^*, \cdots, p_n^*) - \sum_{t\in[T]} \mathrm{Rev}_t$ in expectation. Our main result is:

Theorem 3.1. *For BSPP with n buyers having regular distributions, there exists an algorithm with $O(n^{2.5}\sqrt{T}\log T)$ regret.*

We defer the detailed proof of Theorem 3.1 to the full version of the paper, and provide a proof overview in the below Section 3.1.

3.1 Proof Overview for Theorem 3.1

The main structure of our algorithm follows the structure of the single-buyer algorithm. We maintain n confidence intervals $[\ell_1, r_1], \cdots, [\ell_n, r_n]$ for the optimal prices $p_1^*, \cdots, p_n^*$. The core of the algorithm is to design a sub-routine that uses $\widetilde{O}(\mathrm{poly}(n) \cdot \epsilon^{-2})$ rounds to update the confidence intervals to $[\ell_1', r_1'], \cdots, [\ell_n', r_n']$, such that playing any prices inside new confidence intervals has regret at most ϵ. Then, calling this sub-routine $O(\log T)$ times with halving error parameter ϵ suffices to obtain $\widetilde{O}(\mathrm{poly}(n)\sqrt{T})$ regret.

The core sub-routine contains two steps. The first step is to get a group of near-optimal prices $\hat{p}_1, \cdots, \hat{p}_n$ that approximate $p_1^*, \cdots, p_n^*$. The second step is to use prices $\hat{p}_1, \cdots, \hat{p}_n$ as a benchmark to find new upper- and lower- confidence bounds.

Step 1: Finding $\hat{p}_i$ with a Different Revenue Function. Compared to the single-buyer setting, the revenue function for sequential buyers is different. In particular, when we want to determine $\hat{p}_i$ for buyer i with CDF $F_i(p)$, the revenue function looks like

$$R_i(p) = p \cdot (1 - F_i(p)) + C \cdot F_i(p),$$

where the constant C represents the expected revenue from the buyers $i+1, \cdots, n$. Our goal is to find a near-optimal price $\hat{p}_i \approx \hat{p}_i^*$, where $\hat{p}_i^*$ is the price that maximizes $R_i(p)$.

The major challenge in this step is that function $R_i(p)$ is not always half-concave: it is half-concave in $[C, 1]$, while in $[0, C]$ it is increasing but not necessary concave. This prevents us from directly applying the single-buyer algorithm. Our main idea to fix this extra issue is a case discussion based on the value of $F_i(p)$:

Case 1: $F_i(r_i) > 0.1$. In this case, the probability of skipping buyer i is at least 0.1 when the price for buyer i is $p = r_i$. Therefore, it's sufficient to accurately learn $\hat{C} \approx C$ via posting $p = r_i$ for buyer i. Furthermore, observe that $R_i(p) \le R(C)$ for $p \le C$, i.e., C is a lower bound for $\hat{p}_i^*$. Therefore, running the single-buyer algorithm in the interval $[\hat{C}, 1] \cap [\ell_i, r_i]$ is sufficient to find the approximately optimal $\hat{p}_i$.

Case 2: $F_i(r_i) \le 0.5$. In this case, note that $R_i'(p) = 1 - F_i(p) + (C - p) \cdot f_i(p)$, where $f_i(p) = F_i'(p)$ is the PDF of buyer i. Therefore, we have $R_i'(p) \ge 1 - F_i(r_i) \ge 0.5$ when $p \in [\ell_i, r_i] \cap [0, C]$, i.e., $R_i(p)$ grows sufficiently fast in the interval $[0, C]$. Combining this observation with the Lipschitzness of $R_i(p)$ and the half-concavity in $[C, 1]$, for $a < b < \hat{p}_i^*$, it can be shown that

$$\frac{R_i(\hat{p}_i^*) - R_i(b)}{\hat{p}_i^* - b} \le 2 \cdot \frac{R_i(b) - R_i(a)}{b - a},$$

i.e., function $R_i(p)$ still preserves some kind of "concavity", with an extra factor of 2. We resolve this Case 2 by formally defining this variant of concavity as "generalized half-concavity" and give a corresponding algorithm via modifying the single-buyer algorithm.

Finally, observe that the two cases overlap when $F_i(r_i) \in (0.1, 0.5)$. This margin is used for including the error from learning the value of $F_i(r_i)$. Combining the two cases after learning $\hat{F}_i(r_i) \approx F_i(r_i)$ completes Step 1 of our algorithm.

Step 2: Binary Search with Prices $\hat{p}_1, \cdots, \hat{p}_n$. The second step of the sub-routine is to update confidence intervals via binary searches. In this step, the main new challenge is the error from $\hat{p}_1, \cdots \hat{p}_n$. For instance, while trying to update $[\ell_i, r_i]$ with revenue function $R_i(p) = p \cdot (1 - F_i(p)) + C \cdot F_i(p)$, the constant C comes from non-optimal prices $\hat{p}_{i+1}, \cdots, \hat{p}_n$. Therefore, the optimal $\hat{p}_i^*$ for function $R_i(p)$ does not equal to the global optimal price p_i^*. If we directly apply binary search algorithm to function $R_i(p)$, since p_i^* is not the peak of $R_i(p)$, it's possible that p_i^* falls outside the interval we retrieve from binary search.

To fix this issue, our main idea is to look at function $R_i^*(p) := p \cdot (1 - F_i(p)) + C^* \cdot F_i(p)$, where C^* represents the revenue from the optimal prices $p_{i+1}^*, \cdots p_n^*$. Then, p_i^* is now the optimal price for $R_i^*(p)$. Although function $R_i^*(p)$ is not directly accessible, we use the value of $R_i(p)$ to estimate $R^*(p)$ via bounding the difference between C and C^*. Running the binary search algorithm for this virtual function $R_i^*(p)$ completes Step 2 of our algorithm.

4 n Buyers with General Distributions

In this section, we provide the $\widetilde{O}(\text{poly}(n)T^{\frac{2}{3}})$ regret algorithm for BSPP problem with general distributions. The setting in this section is the same as Section 3, but with no extra regularity assumption. Our result generalizes the $\widetilde{O}(T^{2/3})$ single-buyer result studied in [Kleinberg and Leighton, 2003].

Theorem 4.1. *For BSPP problem with n buyers, there is an algorithm with $O(n^{5/3}T^{2/3}\log T)$ regret.*

We defer the detailed proof of Theorem 4.1 to the full version of the paper. An algorithm overview for Theorem 4.1 is provided below.

Algorithm Overview. Comparing to the regular buyers setting, one extra challenge for general buyers setting is that the revenue function in this case is no longer single-peaked, leading to the failure of the idea of confidence intervals. Therefore, the first step of our algorithm is to discretize the value space: it can be shown that rounding buyers' valuations to multiples of ϵ only brings an extra ϵ loss. The idea of discretization was first used in [Kleinberg and Leighton, 2003] for single buyer setting. We will show that the same idea also works for sequential buyers. For our algorithm, we take the discretization parameter $\epsilon = n^{5/3}T^{-1/3}$, and the accumulated regret from the discretization step would be $T \cdot O(\epsilon) = O(n^{5/3}T^{2/3})$.

After discretizing the value space, it suffices to solve BSPP for general distributions with discrete support. The main idea of our algorithm is almost identical to the algorithm for regular buyers - we design a subroutine that shrinks the possible candidates of p_i^* while keeping a low regret inside the new set of candidates. Calling this subroutine $O(\log T)$ times with a halving error parameter gives the desired algorithm. Comparing to the regular distributions setting, the idea of confidence intervals does not work for discrete distributions, because the candidates may be discontinuous. To fix this issue, we use n price sets to maintain possible candidate prices.

The detailed steps of our algorithm also follows the ideas for regular distributions: we first find a group of approximately prices $\hat{p}_1, \cdots, \hat{p}_n$, and then use these prices as a benchmark to update the candidate prices sets. Since the single-peakedness no longer holds for general distributions, instead of doing binary search, we simply enumerate all remaining prices to update the candidate sets.

5 Linear Regret Lower Bound for Adversarial Valuations

We show an $\Omega(T)$ regret lower bound for learning sequential posted pricing with adversarial buyer values. [Kleinberg and Leighton, 2003] gave an $\widetilde{O}(T^{2/3})$ regret upper bound for bandit sequential posted pricing with adversarial values when there is only a single buyer. The following result shows that it is not possible to generalize their result to multiple buyers.

Theorem 5.1. *For Online Sequential Posted Pricing problem with oblivious adversarial inputs, there exists an instance with $n = 2$ buyers such that the optimal fixed-threshold strategy has total revenue $\frac{3}{4}T$ but no online algorithm can obtain total value more than $\frac{1}{2}T$.*

Proof. The proof follows the hardness example for Online Prophet Inequality problem in [Gatmiry et al., 2024]. Here we restate the example for completeness.

Let s be a binary string in $\{0,1\}^T$. Define $Bin(s)$ to be the binary decimal corresponding to s. For example, $Bin(1000) = (0.1000)_2 = \frac{1}{2}$ and $Bin(0101) = (0.0101)_2 = \frac{5}{16}$.

At the beginning, the adversary chooses a T-bits string $s = s_1 s_2, \cdots, s_T$ uniformly at random, i.e., s_i is set to be 0 or 1 independently with probability $\frac{1}{2}$. In the i-th round, the value of the first buyer will be $v_1 = \frac{1}{2} + \varepsilon \cdot \alpha_i$, where α_i is set to be $Bin(s_1 s_2 ... s_{i-1} + 0 + 1^{T-i+1})$, and ε is an arbitrarily small constant that doesn't effect the value. The value of the second buyer only depends on s_i: v_2 is set to be 0 when $s_i = 1$, while $v_i = 1$ when $s_i = 0$.

The key idea of this example is that we have $Bin(s) > \alpha_i$ when s_i is 0, and $Bin(s) < \alpha_i$ when $s_i = 1$. Therefore, if we set $p_1 = Bin(s)$ and $p_2 = 1$, we can receive v_2 when v_2 is 1, and otherwise v_1. Since s is generated uniformly at random, the expected revenue is $\frac{3}{4}T$. However, for any online algorithm, it only knows that the value of v_2 is 0 or 1 with probability $\frac{1}{2}$. Therefore, it can only get revenue $\frac{1}{2}$ in expectation and the maximum total revenue is $\frac{1}{2}T$. □

Acknowledgements

We are thankful to Thomas Kesselheim as the project was initiated in discussions with him. We also thank the anonymous reviewers who helped greatly improve the presentation of the paper.

References

Ashwinkumar Badanidiyuru, Robert Kleinberg, and Aleksandrs Slivkins. 2018. Bandits with Knapsacks. *J. ACM* 65, 3 (2018), 13:1–13:55.

Maria-Florina Balcan, Avrim Blum, Jason D. Hartline, and Yishay Mansour. 2008. Reducing mechanism design to algorithm design via machine learning. *J. Comput. Syst. Sci.* 74, 8 (2008), 1245–1270.

Maria-Florina Balcan, Dan F. DeBlasio, Travis Dick, Carl Kingsford, Tuomas Sandholm, and Ellen Vitercik. 2021. How much data is sufficient to learn high-performing algorithms? generalization guarantees for data-driven algorithm design. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing*. ACM, 919–932.

Maria-Florina Balcan, Tuomas Sandholm, and Ellen Vitercik. 2018. A General Theory of Sample Complexity for Multi-Item Profit Maximization. In *Proceedings of the 2018 ACM Conference on Economics and Computation*. ACM, 173–174.

Liad Blumrosen and Thomas Holenstein. 2008. Posted prices vs. negotiations: an asymptotic analysis. *EC* 10 (2008), 1386790–1386801.

Sébastien Bubeck and Nicolò Cesa-Bianchi. 2012. Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems. *Foundations and Trends in Machine Learning* 5, 1 (2012), 1–122.

Yang Cai and Constantinos Daskalakis. 2017. Learning Multi-Item Auctions with (or without) Samples. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS*, Chris Umans (Ed.). IEEE Computer Society, 516–527.

Yang Cai, Argyris Oikonomou, and Mingfei Zhao. 2022. Computing simple mechanisms: Lift-and-round over marginal reduced forms. In *54th Annual ACM SIGACT Symposium on Theory of Computing, STOC*. 704–717.

Nicolo Cesa-Bianchi, Tommaso Cesari, and Vianney Perchet. 2019. Dynamic pricing with finitely many unknown valuations. In *Algorithmic Learning Theory*. PMLR, 247–273.

Nicolò Cesa-Bianchi, Claudio Gentile, and Yishay Mansour. 2015. Regret Minimization for Reserve Prices in Second-Price Auctions. *IEEE Trans. Inf. Theory* 61, 1 (2015), 549–564. https://doi.org/10.1109/TIT.2014.2365772

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

Shuchi Chawla, Jason D. Hartline, and Robert D. Kleinberg. 2007. Algorithmic Pricing via Virtual Valuations. In *the 8th ACM Conference on Electronic Commerce (EC)*.

Shuchi Chawla, Jason D. Hartline, David L. Malec, and Balasubramanian Sivan. 2010. Multi-parameter mechanism design and sequential posted pricing. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC*. 311–320.

Shuchi Chawla, David L. Malec, and Balasubramanian Sivan. 2015. The power of randomness in Bayesian optimal mechanism design. *Games and Economic Behavior* 91 (2015), 297–317.

Xi Chen, Ilias Diakonikolas, Anthi Orfanou, Dimitris Paparas, Xiaorui Sun, and Mihalis Yannakakis. 2015. On the Complexity of Optimal Lottery Pricing and Randomized Mechanisms. In *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS*. 1464–1479.

Richard Cole and Tim Roughgarden. 2014. The sample complexity of revenue maximization. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, David B. Shmoys (Ed.). ACM, 243–252. https://doi.org/10.1145/2591796.2591867

Richard Combes and Alexandre Proutière. 2014. Unimodal Bandits: Regret Lower Bounds and Optimal Algorithms. In *Proceedings of the 31th International Conference on Machine Learning, ICML (JMLR Workshop and Conference Proceedings, Vol. 32)*. JMLR.org, 521–529.

José R. Correa, Patricio Foncea, Dana Pizarro, and Victor Verdugo. 2019. From pricing to prophets, and back! *Oper. Res. Lett.* 47, 1 (2019), 25–29.

Nikhil R. Devanur, Zhiyi Huang, and Christos-Alexandros Psomas. 2016. The sample complexity of auctions with side information. In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC*. ACM, 426–439.

Peerapong Dhangwatnotai, Tim Roughgarden, and Qiqi Yan. 2015. Revenue maximization with a single sample. *Games Econ. Behav.* 91 (2015), 318–333.

Liran Einav, Chiara Farronato, Jonathan Levin, and Neel Sundaresan. 2018. Auctions versus posted prices in online markets. *Journal of Political Economy* 126, 1 (2018), 178–215.

Khashayar Gatmiry, Thomas Kesselheim, Sahil Singla, and Yifan Wang. 2024. Bandit Algorithms for Prophet Inequality and Pandora's Box. In *Proceedings of the thirty-sixth annual ACM-SIAM symposium on Discrete Algorithms, SODA*.

Yannai A. Gonczarowski and Noam Nisan. 2017. Efficient empirical revenue maximization in single-parameter auction environments. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC*. ACM, 856–868.

Yannai A. Gonczarowski and S. Matthew Weinberg. 2021. The Sample Complexity of Up-to-ϵ Multi-dimensional Revenue Maximization. *J. ACM* 68, 3 (2021), 15:1–15:28.

Chenghao Guo, Zhiyi Huang, Zhihao Gavin Tang, and Xinzhi Zhang. 2021. Generalizing Complex Hypotheses on Product Distributions: Auctions, Prophet Inequalities, and Pandora's Problem. In *Conference on Learning Theory, COLT*. 2248–2288.

Chenghao Guo, Zhiyi Huang, and Xinzhi Zhang. 2019. Settling the sample complexity of single-parameter revenue maximization. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC*. ACM, 662–673.

Jason D Hartline. 2013. Mechanism design and approximation. *Book draft. October* 122, 1 (2013).

Zhiyi Huang, Yishay Mansour, and Tim Roughgarden. 2018. Making the Most of Your Samples. *SIAM J. Comput.* 47, 3 (2018), 651–674.

Nicole Immorlica, Karthik Abinav Sankararaman, Robert E. Schapire, and Aleksandrs Slivkins. 2022. Adversarial Bandits with Knapsacks. *J. ACM* 69, 6 (2022), 40:1–40:47.

Yaonan Jin, Pinyan Lu, and Tao Xiao. 2023. Learning Reserve Prices in Second-Price Auctions. In *14th Innovations in Theoretical Computer Science Conference, ITCS*.

Robert D. Kleinberg and Frank Thomson Leighton. 2003. The Value of Knowing a Demand Curve: Bounds on Regret for Online Posted-Price Auctions. In *44th Symposium on Foundations of Computer Science, FOCS*. IEEE Computer Society, 594–605.

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

Renato Paes Leme, Balasubramanian Sivan, Yifeng Teng, and Pratik Worah. 2023a. Description Complexity of Regular Distributions. In *Proceedings of the 24th ACM Conference on Economics and Computation, EC*. ACM, 959.

Renato Paes Leme, Balasubramanian Sivan, Yifeng Teng, and Pratik Worah. 2023b. Pricing Query Complexity of Revenue Maximization. In *Proceedings of the thirty-fourth annual ACM-SIAM symposium on Discrete Algorithms*.

Brendan Lucier. 2017. An economic view of prophet inequalities. *SIGecom Exch.* 16, 1 (2017), 24–47.

Stefan Magureanu, Richard Combes, and Alexandre Proutière. 2014. Lipschitz Bandits: Regret Lower Bounds and Optimal Algorithms. *CoRR* abs/1405.4758 (2014). arXiv:1405.4758 http://arxiv.org/abs/1405.4758

Michela Meister and Sloan Nietert. 2021. Learning with Comparison Feedback: Online Estimation of Sample Statistics. In *Algorithmic Learning Theory*, Vol. 132. PMLR, 983–1001.

Jamie Morgenstern and Tim Roughgarden. 2016. Learning Simple Auctions. In *Proceedings of the 29th Conference on Learning Theory, COLT (JMLR Workshop and Conference Proceedings, Vol. 49)*. JMLR.org, 1298–1318.

Roger B Myerson. 1981. Optimal auction design. *Mathematics of operations research* 6, 1 (1981), 58–73.

Rad Niazadeh, Negin Golrezaei, Joshua R. Wang, Fransisca Susan, and Ashwinkumar Badanidiyuru. 2021. Online Learning via Offline Greedy Algorithms: Applications in Market Design and Optimization. In *EC '21: The 22nd ACM Conference on Economics and Computation*. ACM, 737–738.

Princewill Okoroafor, Vaishnavi Gupta, Robert Kleinberg, and Eleanor Goh. 2023. Non-Stochastic CDF Estimation Using Threshold Queries. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 3551–3572.

Tim Roughgarden. 2017. *Twenty lectures on algorithmic game theory*.

Tim Roughgarden and Okke Schrijvers. 2016. Ironing in the Dark. In *Proceedings of the 2016 ACM Conference on Economics and Computation, EC*. ACM, 1–18.

Aleksandrs Slivkins. 2019. Introduction to Multi-Armed Bandits. *Foundations and Trends in Machine Learning* 12, 1-2 (2019), 1–286.

Qiqi Yan. 2011. Mechanism Design via Correlation Gap. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2011, San Francisco, California, USA, January 23-25, 2011*, Dana Randall (Ed.). SIAM, 710–719. https://doi.org/10.1137/1.9781611973082.56