Learning Revenue Maximization Using Posted Prices for Stochastic Strategic Patient Buyers

Eitan-Hai Mashiah[†1], **Idan Attias**[†2], **Yishay Mansour**[1,3]
[1]Tel Aviv University, Israel
[2]Ben-Gurion University, Israel
[3]Google Research, Israel
eitanhaimashiah@gmail.com, idanatti@post.bgu.ac.il, mansour.yishay@gmail.com

Abstract

We consider a seller faced with buyers which have the ability to delay their decision, which we call patience. Each buyer's type is composed of value and patience, and it is sampled i.i.d. from a distribution. The seller, using posted prices, would like to maximize her revenue from selling to the buyer. In this paper, we formalize this setting and characterize the resulting Stackelberg equilibrium, where the seller first commits to her strategy, and then the buyers best respond. Following this, we show how to compute both the optimal pure and mixed strategies. We then consider a learning setting, where the seller does not have access to the distribution over buyer's types. Our main results are the following. We derive a sample complexity bound for the learning of an approximate optimal pure strategy, by computing the fat-shattering dimension of this setting. Moreover, we provide a general sample complexity bound for the approximate optimal mixed strategy. We also consider an online setting and derive a vanishing regret bound with respect to both the optimal pure strategy and the optimal mixed strategy.

1 Introduction

Pricing is ubiquitous, and it is the primary means by which sellers and buyers interact. It is no surprise that revenue maximization pricing is the topic of a vast amount of literature in economic theory and algorithmic game theory (see (Hartline 2021; Nisan et al. 2007)). In most of the literature the seller and buyer interact instantaneously, and either a transaction occurs (the buyer purchases an item) or not. We are interested in this work in the case where the buyer can potentially delay the purchase decision, depending on his type. We call such buyers *patient buyers*.

There are many examples of patient buyers in the real world. One example is shipping cost, where there are different costs depending on the duration of the shipping. Normally, same day delivery is more expensive than next day delivery, which is more expensive than two-day delivery, and so on. Another example is an online merchant whose production cost varies with the delivery date. Items that have to be shipped immediately cost more to produce than items that need to be shipped after 10 business days. Another scenario is regarding online merchant who observe that a buyer has a shopping bag that was not purchased. The merchant sometimes offers the buyer a limited time discount on the items in the buyer's shopping bag. The buyer has uncertainty regarding future prices after the discount terminates, the prices might return to the original ones or there may be a new offer with an even larger discount. Our model of patient buyers abstracts this phenomena from the buyer perspective, the ability to prolong the time to receive of the desired item.

Patient buyers were introduced in Feldman et al. (2016) and later studied in (Koren, Livni, and Mansour 2017a,b). They presented an adversarial online model where the buyers have a valuation and duration for the purchase, namely the sequence of arrivals is controlled by an adversary. They studied the regret compared to the best fixed price. In this work we consider a stochastic setting, and we study the expected revenue from an optimal sequence of prices (rather than a single fixed price).

Our model of patient buyers can be intuitively described as follows. We have a seller that has an unlimited supply from a single item, and would like to maximize her expected revenue. Each buyer has a type (v, w) where v is his value for the item and $w \in [\hat{w}]$ is his patience, where $\hat{w}$ is the maximum patience. A buyer of type (v, w) has a value v for the item if it is purchased in the first w time steps from his arrival. The types of the buyers are sampled i.i.d. from a distribution $\mathcal{D}$. The seller proposes a price for the buyer at each time step, and observes whether the buyer bought the item or continued to the next time step. Initially, the seller commits to her pricing strategy and the buyer best respond to it, i.e., this is a *Stackelberg game* where the seller is the leader and the buyer is the follower. We consider both the case where the distribution $\mathcal{D}$ is known and the case where $\mathcal{D}$ is unknown.

1.1 Our Contributions

We initially assume that the buyer's type distribution $\mathcal{D}$ is known and derive the following results.

- We show a separation between the best fixed price, the best pure strategy, which is a fixed sequence of prices, and the best mixed strategy, which is a distribution over price sequences.
- We characterize the optimal pure strategy of the seller

[†]These authors contributed equally.

and show that the sequence of prices are non-increasing and that the buyers will always buy at the end of them patience, if they decide to buy.

- For mixed strategies we characterize the buyer's best response strategy.
- We show how to compute efficiently the optimal pure strategy. For the optimal mixed strategy, we give an algorithm which is exponential in the maximum patience and polynomial in the support of the distribution.

We then consider a learning setting, where the seller does not know the distribution $\mathcal{D}$ over buyer's types, but can learn it from samples.

- We derive a sample complexity bound for the learning of an approximate optimal pure strategy, by computing the fat-shattering dimension of the setting and showing that it is linear in the maximum patience of a buyer, i.e., $\hat{w}$. Using the bound on the fat shattering dimension, we derive an upper bound on the sample complexity of
$$\mathcal{O}\left(\min\left\{\frac{\hat{w}}{\varepsilon^2}\,,\,\frac{\log(\hat{w})}{\varepsilon^3}\right\}+\frac{1}{\varepsilon^2}\log\frac{1}{\delta}\right),$$
and a lower bound of
$$\Omega\left(\frac{\hat{w}}{\varepsilon}+\frac{1}{\varepsilon^2}\log\frac{1}{\delta}\right).$$
- We give a general sample complexity bound for the approximate optimal mixed strategy. Our sample bound is
$$\mathcal{O}\left(\frac{\hat{w}^4}{\varepsilon^3}+\frac{\hat{w}^2}{\varepsilon^2}\log\frac{1}{\delta}\right).$$
- We consider an online setting with T buyers whose type is drawn i.i.d. from an unknown distribution $\mathcal{D}$. We derive a regret bound with respect to the optimal pure strategy of $\widetilde{\mathcal{O}}\left(\sqrt{T}\sqrt{\hat{w}}\right)$. We derive a regret bound with respect to the optimal mixed strategy of $\widetilde{\mathcal{O}}\left(T^{2/3}\hat{w}^{4/3}\right)$.

1.2 Related Work

The FedEx problem. The FedEx problem was presented in (Fiat et al. 2016) and later studied in (Saxena, Schvartzman, and Weinberg 2018; Devanur et al. 2020). In the FedEx problem the seller is faced with a buyer which has a varying patience for the duration of the delivery date of a package. The seller offers the buyer a menu, with a lottery for each possible duration. The buyer selects one of those lotteries and later pays the realized price of the lottery. The main issue is that this mechanism maximizes the revenue of the seller, over all incentive-compatible mechanisms.

While the two models are clearly related, there are a few important differences between the two models. The main difference is regarding *what* the buyer observes and *when* it observes it. In the FedEx problem the buyer observes only the menu. Our setting is more interactive. In each day the buyer first observes the realized price (which is potentially drawn from a distribution) and only then decides if to buy or wait. This implies that the buyer has more information, observing the sequence of prices until the current day, before deciding whether to buy or wait. In contrast, in the FedEx the buyer never observes any realization of prices, except for the lottery it selected. For example, if the FedEx problem has two lotteries, both uniform $[0, 1]$, then the buyer will pay an expected revenue of $1/2$ regardless which lottery he picks. In contrast, in our setting if the seller offers in the first two days a uniform price $[0, 1]$, the buyer can decide to buy in the first day if the price is less than $1/2$ and otherwise buy in the second day. This would give an expected revenue of $3/8$. A minor issue is that we focus on posted prices while the FedEx allow for an arbitrary mechanism.

Revenue maximization. The seminal work of Myerson (1981) derives the optimal mechanism for revenue maximization, and shows that for many distributions it coincides with a sealed bid second price auction with a reserve price. That model allows for single parameter buyers, and does not allow to incorporate the dimension of patience. The main focus of this paper is on pricing strategies for patient buyers which falls outside that framework.

The work of Kleinberg and Leighton (2003) derives regret bounds for a seller faced with multiple stochastic buyers. The regret is with respect to the best fixed price. In contrast, we compete with the optimal pure and mixed strategies over sequences of prices, due to our patient buyers.

Repeated interaction between single seller and single buyer. The works of (Mohri and Medina 2015; Mohri and Munoz 2014; Amin, Rostamizadeh, and Syed 2013, 2014; Vanunts and Drutsa 2019) consider a model of repeated interaction between a single buyer and a single seller. The main issue is that due to the repeated interaction, the buyer has an incentive to lower future prices at the cost of sacrificing current utility. They define strategic regret and derive near optimal strategic regret bounds for various valuation models, using the fact that the buyer's utility is discounted. First, the buyer has no patience, at each step he needs to decide if to buy or not. Second, they consider a single fixed buyer while we consider a distribution over buyer's valuation and patience. Third, they use discounting to decay the buyer's utility over time, while in our model the buyer's utility depends only on the paid price. Lastly, they compare to the best fixed price while we compare to either a pure or mixed price sequence.

Patient buyers. As mentioned in the introduction, patient buyers were introduced in Feldman et al. (2016) and later studied in (Koren, Livni, and Mansour 2017a,b). The focus of those works is on regret minimization with respect to the best fixed price. They consider an adversarial online model where the buyers have a valuation and duration for the purchase, namely the sequence of arrivals is controlled by an adversary. In this work we consider a stochastic setting, namely, the buyers types are sampled i.i.d. from a distribution. We compare the seller expected revenue to the optimal expected revenue from a pure strategy (a fixed sequence of prices) or mixed strategy (a distribution over price sequences). Clearly, our benchmarks allows for a much higher expected revenue.

Learning approximate revenue-maximizing mechanisms was initiated by Balcan et al. (2008), using samples to design near optimal revenue-maximizing mechanism. Huang, Mansour, and Roughgarden (2018) use i.i.d. samples to derive the optimal sell price. The works of (Morgenstern and Roughgarden 2015, 2016; Gonczarowski and Nisan 2017) study the complexity of learning a near optimal revenue maximizing mechanism. We differ from all this literature due to the patience of our buyers.

Stackelberg games. Blum et al. (2019) derive hardness results for large action Stackelberg games.

2 Model

We consider a setting of a single seller and multiple buyers, where the seller has unlimited supply of a single item to sell. The seller observes a sequence of T buyers, and with each buyer she interacts for $\hat{w}$ steps, in each she offers the buyer a (potentially different) price. Each buyer appears only once, and can purchase the item at most once.

The seller's pricing strategy may be either deterministic, $\mathbf{p} = (p_1, \ldots, p_{\hat{w}}) \in [0,1]^{\hat{w}}$, or randomized $\mathcal{P} \in \Delta([0,1]^{\hat{w}})$. We refer to it as *pure* and *mixed* strategies, respectively. When a pure strategy uses only a single price, we refer to it as a *fixed price*. We assume no price discrimination, the seller plays the same strategy against each of the buyers. We denote a shorthand of a pricing vector by $\mathbf{p}_{1:i} = (p_1, \ldots, p_i)$.

Denote by $\mathbf{e}_1, \ldots, \mathbf{e}_{\hat{w}}$ the unit vectors of size $\hat{w}$ and by $\mathbf{e}_0$ the zero vector of size $\hat{w}$. Namely, $\mathbf{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ has 1 in the i-th location, and $\mathbf{e}_0 = (0, \ldots, 0)$. Define the buyer's decision whether to purchase the item at step i while observing prices $p_1, \ldots, p_i$ by $\pi^i(\mathbf{p}_{1:i}) \mapsto \{\mathbf{e}_i, \textit{continue}\}$, for $i \in [\hat{w}]$. The buyer's strategy $\pi_{v,w} = (\pi^1_{v,w}, \ldots, \pi^{\hat{w}}_{v,w})$ of a buyer with value v and patience w, is online and defined as $\pi_{v,w}(\mathbf{p}) = \mathbf{e}_i$ if $i \leqslant w$ is the first step where $\pi^i_{v,w}(\mathbf{p}_{1:i}) = \mathbf{e}_i$ or $\pi_{v,w}(\mathbf{p}) = \mathbf{e}_0$ if no such index $i \leqslant w$ exists, i.e., the buyer does not purchase the item. The *utility* function of a buyer type (v, w) given pricing $\mathbf{p}$ and decision $\mathbf{e}_i$ is defined as $\text{util}_{v,w}(\mathbf{p}, \mathbf{e}_i) = (v - \mathbf{p} \cdot \mathbf{e}_i)\,\mathbb{I}\{0 < i \leqslant w\}$ where $\mathbf{x} \cdot \mathbf{y}$ denotes the scalar product of $\mathbf{x}$ and $\mathbf{y}$. Note that $\text{util}_{v,w}(\mathbf{p}, \mathbf{e}_0) = 0$. The seller's *revenue* for pricing $\mathbf{p}$ and decision $\mathbf{e}_i$ is defined as $\text{rev}(\mathbf{p}, \mathbf{e}_i) = \mathbf{p} \cdot \mathbf{e}_i = p_i$.

Define the utility of a buying strategy $\pi_{v,w}$ for a buyer type (v, w), given a selling strategy $\mathcal{P}$ by

$$u_{v,w,\mathcal{P}}(\pi_{v,w}) = \mathop{\mathbb{E}}_{\mathbf{p} \sim \mathcal{P}} \left[\text{util}_{v,w}(\mathbf{p}, \pi_{v,w}(\mathbf{p}))\right].$$

The buyer would like to maximize his utility, and select $\pi^\star_{v,w,\mathcal{P}} = \text{argmax}_{\pi_{v,w}} u_{v,w,\mathcal{P}}(\pi_{v,w})$. Define the total revenue of a selling strategy $\mathcal{P}$ for distribution $\mathcal{D}$ by

$$r(\mathcal{P}; \mathcal{D}) = \mathop{\mathbb{E}}_{\mathbf{p} \sim \mathcal{P}} \mathop{\mathbb{E}}_{(v,w) \sim \mathcal{D}} \left[\text{rev}(\mathbf{p}, \pi^\star_{v,w,\mathcal{P}}(\mathbf{p}))\right].$$

The seller would like to maximize her total revenue, and select $\mathcal{P}^\star = \text{argmax}_{\mathcal{P}}\, r(\mathcal{P}; \mathcal{D})$.

More precisely, we consider a *Stackelberg* equilibrium. The seller fixes her pricing strategy and the buyer best responds (at each time step, decides whether to buy at this step or waits for the next step). Clearly, the seller selects a pricing strategy that maximizes her expected revenue given that the buyer best responds.

Learning. In the learning setting the seller does not know the distribution $\mathcal{D}$ over buyer types $\mathcal{Z} = [0,1] \times [\hat{w}]$, instead she receives an i.i.d. samples from $\mathcal{D}$ in order to learn a selling strategy which maximize her expected revenue. We define the formal model of revenue learning with patient buyers.

Definition 2.1 (Revenue PAC-learning) For any $(\varepsilon, \delta) \in (0,1)$, the sample complexity of (ε, δ)-PAC revenue learning with respect to a set of strategies Ω, denoted by $\mathcal{M}(\varepsilon, \delta, \Omega)$, is defined as the smallest $m \in \mathbb{N} \cup \{0\}$, for which there exists an algorithm $\mathcal{A} : \mathcal{Z}^* \to \Delta([0,1]^{\hat{w}})$, such that for any distribution $\mathcal{D}$ over $\mathcal{Z}$, upon receiving a random sample $S \sim \mathcal{D}^m$, with probability $1 - \delta$ it holds that,

$$r(\mathcal{A}(S); \mathcal{D}) \geqslant \max_{\mathcal{P} \in \Omega} r(\mathcal{P}; \mathcal{D}) - \varepsilon.$$

We consider Ω to be the set of pure strategies, i.e., $\Omega = [0,1]^{\hat{w}}$, or mixed strategies, i.e., $\Omega = \Delta([0,1]^{\hat{w}})$.

Our second learning model is in the online setting, where the seller gets to see the sample sequentially instead of receiving the whole sample at once. The seller is facing a sequence of T buyers of types $z_1, \ldots, z_T$ such that each buyer type z_t is drawn i.i.d. from the unknown distribution $\mathcal{D}$. We assume that the seller interacts with a single buyer at a time, that is, each round of the learning consists of one interaction between the seller and a buyer.

We denote by $r(\mathcal{P}_t; z_t)$, the revenue of an online learner $\mathcal{A} : (\mathcal{P}_{1:t-1}; z_{1:t-1}) \mapsto \mathcal{P}_t$, at round t, given a buyer z_t. The regret compared to a set of strategies Ω of a seller $\mathcal{A}$ for playing strategies $\mathcal{P}_1, \ldots, \mathcal{P}_T$, given a sequence of buyer types $z_1, \ldots, z_T$, defined by

$$\text{Regret}^{\Omega}_T(\mathcal{A}; \mathcal{D}) = \max_{\mathcal{P}^\star \in \Omega} \sum_{t=1}^{T} \mathop{\mathbb{E}}_{z_t \sim \mathcal{D}} \left[r(\mathcal{P}^\star; z_t) - r(\mathcal{P}_t; z_t)\right].$$

Similar to the offline setting, we consider both the case that Ω is the set of pure strategies, and the case that Ω is the set of mixed strategies.

Notation. Vectors are denoted by bold lower case letters, e.g., $\mathbf{p}$; we denote historical prices until step i by $\mathbf{p}_{1:i} = (p_1, \ldots, p_i)$, where $\mathbf{p}_{1:0}$ denotes the null vector; $\mathcal{D}_v$ and $\mathcal{D}_w$ denote the marginal distributions of the buyer's value and patience of distribution $\mathcal{D}$, respectively; $\mathcal{V}$ and $\mathcal{W}$ denote the support of $\mathcal{D}_v$ and $\mathcal{D}_w$, respectively; $S_{\mathcal{D}} \subseteq \mathcal{V} \times \mathcal{W}$ denotes the support of $\mathcal{D}$; $[n]$ denotes the set $\{1, \ldots, n\}$; $\lesssim$ and $\gtrsim$ denote inequalities up to an absolute constant factor.

3 Optimal Pricing: Characterization and Planning

In this section we derive basic properties of our model and show how to compute both the optimal pure and mixed strategies given that the distribution $\mathcal{D}$ is known.

Product distribution. We start by showing that when the distribution $\mathcal{D}$ is a product distribution over values and patience, and the distribution over buyers' values is regular, then the seller cannot outperform the single fixed price, as in Myerson (1981) (see Appendix A.1). For this reason we focus on a joint distribution $\mathcal{D}$, and show that there is a separation between the best fixed price and best pure strategy, and also between the best pure strategy and the best mixed strategy (see Appendix A.2).

Optimal pure selling strategy. In this section, we characterize the optimal pure selling strategies, and use it to compute efficiently an optimal pure strategy.

Theorem 3.1 *Assume the support of the marginal distribution of the buyer's value, $\mathcal{V}$, is contained in $[\underline{v}, \overline{v}] \subseteq [0,1]$. Then, there exists an optimal non-increasing pure selling strategy using only prices from $[\underline{v}, \overline{v}]$. Moreover, if $\mathcal{V}$ is a finite set, there exists an optimal non-increasing pure selling strategy using only prices from $\mathcal{V}$.*

Intuitively, the existence of an optimal non-increasing pure selling strategy follows since each time the price increases, no buyer would buy at the higher price (since he can buy at the lower price). This implies that we can "replace" the higher price by the lower price. Notice that when faced with a sequence of non-increasing prices, the buyer is better off waiting for the last step in his patience window where the price is the lowest. (See a complete proof in Appendix A.3.) Based on the characterization result, we obtain:

Theorem 3.2 *There exists an algorithm which produces an optimal pure selling strategy for distributions $\mathcal{D}$ over $\mathcal{V} \times [\hat{w}]$ where $\mathcal{V}$ is a finite set of values, with running time $\mathcal{O}(|\mathcal{V}|^2 \hat{w})$.*

Our algorithm uses a dynamic programming approach, which generates a pricing with non-increasing prices. For non-increasing pricing, the strategic buyer buys at the last step in his patience window, as long as the price is lower than his value. The algorithm takes advantage of this in order to simplify the computation of the seller's revenue (see Appendix A.4).

Optimal mixed selling strategy. In this section, we characterize the buyer's best-response strategy against a given mixed strategy, and use it to find an optimal mixed selling strategy. We present a simple class of buying strategies, which we call *threshold strategies*. We show that for any mixed selling strategy, there exists a buyer's best response strategy which is a threshold strategy.

Definition 3.3 A buying strategy π is a *threshold strategy* if, for any history of prices, there is a threshold $\theta^i(\mathbf{p}_{1:i})$, such that the buyer buys at the first step $i \leqslant w$ in which the price is at most the corresponding threshold, i.e., $\theta^i(\mathbf{p}_{1:i}) \geqslant p_i$.

Theorem 3.4 *For any mixed selling strategy $\mathcal{P}$, there is a threshold buying strategy which is a best response.*

Intuitively, the threshold at step i is set to the price which makes the buyer "indifferent" between buying at step i and continuing to step $i + 1$. If the offered price is lower, the buyer makes the purchase and if the price is higher the buyer waits (see Appendix A.5). Using the fact that there always exists a buyer's best response strategy which is a threshold strategy, the following holds:

Theorem 3.5 *There exists an algorithm which produces an optimal mixed selling strategy for distributions $\mathcal{D}$ over $\mathcal{V} \times [\hat{w}]$ where $\mathcal{V}$ is a finite set of values, and given a finite set of prices P, with running time $\mathcal{O}(\mathrm{poly}(|P|^{\hat{w}})|V|^{2\hat{w}}\hat{w}^2)$.*

Similar to the algorithm for finding an optimal pure strategy, we compute the optimal mixed selling strategy from the last time step backwards until the first time step. The main challenge is that the set of buyer's types that reach time step i and potentially buy there is not anymore simply the buyers with patience i. This creates an intricate dependence between the strategy in steps up to step i and the strategy from step i onward. This is the main reason that the resulting algorithm has a running time exponential in the maximum patience $\hat{w}$. More precisely, we make an exponential number of calls to a linear programming (see Appendix A.5).

4 Learning Selling Strategies

Up to now we assumed that the type distribution $\mathcal{D}$ is common knowledge, i.e., both the seller and buyer know it. We showed how the seller can compute an optimal pure strategy and an optimal mixed strategy, given the distribution $\mathcal{D}$. This section would focus on learning, namely, the seller does not have any a priori information about the distribution $\mathcal{D}$. First, we consider an offline learning model (see Definition 2.1), where the seller observes a random sample $S \sim \mathcal{D}^m$ of the buyer's type. We would like to understand the sample complexity for the seller to learn an approximate optimal pure strategy and approximate optimal mixed strategy. In Section 4.2, we derive upper and lower bounds for the class of pure strategies. In order to derive the upper bound we compute the fat-shattering of pure strategies, and show that it is linear in the maximum patience of a buyer. In Section 4.3, we study the sample complexity of mixed strategies, and derive upper bounds, via learning discrete distributions. Furthermore, in Section 4.4, we consider an online setting with T buyers whose type is drawn i.i.d. from an unknown distribution $\mathcal{D}$. We derive regret bounds with respect to the optimal pure strategy and optimal mixed strategy.

4.1 Background on Learning and Notation

We use the shorthand of $r(\mathcal{P}; z)$ for the revenue of a selling strategy $\mathcal{P}$ from a buyer z, i.e., $r(\mathcal{P}; z) = \mathbb{E}_{\mathbf{p}\sim\mathcal{P}}\left[\mathrm{rev}(\mathbf{p}, \pi^\star_{z,\mathcal{P}}(\mathbf{p}))\right]$, where $\pi^\star_{z,\mathcal{P}}(\cdot)$ is the buyer's best response when having type z. For a sample of buyers' types $S = \{z_1, \ldots, z_m\}$ define the *empirical revenue* of a selling strategy $\mathcal{P} \in \Delta([0,1]^{\hat{w}})$ with respect to S, as

$$\widehat{r}(\mathcal{P}; S) = \frac{1}{m}\sum_{z\in S} r(\mathcal{P}; z).$$

The *empirical revenue maximization* learning algorithm ERM on sample S with respect to a set of strategies Ω, is defined as

$$\mathrm{ERM}_\Omega(S) \in \underset{\mathcal{P}\in\Omega}{\operatorname{argmax}}\ \widehat{r}(\mathcal{P}; S).$$

In this section we consider the following set of strategies. The set of pure strategies, which is denoted by $\Omega^{\text{pure}} = [0,1]^{\hat{w}}$. The set of mixed strategies, denoted by $\Omega^{\text{mixed}} = \Delta([0,1]^{\hat{w}})$, and the set of mixed strategies that offers prices from a set $\mathcal{V}$, denoted by $\Omega^{\text{mixed}}_{\mathcal{V}} = \Delta(\mathcal{V}^{\hat{w}})$.

Recall the definition of the fat-shattering dimension (Kearns and Schapire 1994; Alon et al. 1997).

Definition 4.1 (Fat-shattering dimension) Let $\mathcal{F}$ be a class of real-valued functions from input space $\mathcal{Z}$ and $\gamma > 0$. We say that $S = \{z_1, \dots, z_m\} \subseteq \mathcal{Z}$ is γ-shattered by $\mathcal{F}$ if there exists a witness $\mathbf{c} = (c_1, \dots, c_m) \in \mathbb{R}^m$ such that for each $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_m) \in \{-1, 1\}^m$ there is a function $f_{\boldsymbol{\sigma}} \in \mathcal{F}$ such that

$$\forall i \in [m] \ \begin{cases} f_{\boldsymbol{\sigma}}(z_i) \geqslant c_i + \gamma, & \text{if } \sigma_i = 1 \\ f_{\boldsymbol{\sigma}}(z_i) \leqslant c_i - \gamma, & \text{if } \sigma_i = -1. \end{cases}$$

The fat-shattering dimension of $\mathcal{F}$ at scale γ is the cardinality of the largest set of points in $\mathcal{Z}$ that can be γ-shattered by $\mathcal{F}$.

The Pseudo-dimension of a function class $\mathcal{F}$ (Pollard 1990; Haussler 1992) can be defined as $\operatorname{Pdim}(\mathcal{F}) = \lim_{\gamma \searrow 0} \operatorname{fat}_\gamma(\mathcal{F})$. From the monotonicity of the fat-shattering, it holds that $\operatorname{Pdim}(\mathcal{F}) \leqslant \operatorname{fat}_\gamma(\mathcal{F})$, for any $\gamma > 0$.

The following is a well known uniform convergence theorem for classes with finite fat-shattering for all $\gamma > 0$. (The proofs for this section appear in Appendix B.1).

Theorem 4.2 *Let $\mathcal{F}$ be a function class of real-valued functions mapping from $\mathcal{Z}$ to $[0,1]$. For an i.i.d. sample $S = \{z_1, \dots, z_m\}$ from a distribution $\mathcal{D}$ over $\mathcal{Z}$, with probability $1 - \delta$ it holds that,*

$$\sup_{f \in \mathcal{F}} \left| \frac{1}{m} \sum_{i=1}^{m} f(z_i) - \mathop{\mathbb{E}}_{z \sim \mathcal{D}} [f(z)] \right| \lesssim \frac{1}{\sqrt{m}} \int_0^\infty \sqrt{\operatorname{fat}_\gamma(\mathcal{F})}\, d\gamma + \sqrt{\frac{\log \frac{1}{\delta}}{m}},$$

where $\lesssim$ means up to an absolute constant factor.

4.2 Sample Complexity of Pure Selling Strategies

Consider the pure selling strategies with non-increasing prices $\mathbf{p} = (p_1, \dots, p_{\hat{w}}) \in [0,1]^{\hat{w}}$, such that $p_1 \geqslant \dots \geqslant p_{\hat{w}}$, where $\hat{w}$ is the maximal patience window of all buyers. Recall that by Theorem 3.1, there exist an optimal non-increasing pure strategy, so it suffices to find an approximation of the optimal pure selling strategy in this set of strategies.

Our main result for this section is upper and lower bounds on the sample complexity (Definition 2.1) for learning an approximate optimal pure selling strategy. (The proofs for this section are in Appendix B.2).

Theorem 4.3 *The sample complexity for learning pure selling strategies is*

$$\mathcal{M}(\varepsilon, \delta, \Omega^{pure}) = \mathcal{O}\left(\min\left\{ \frac{\hat{w}}{\varepsilon^2}, \frac{\log(\hat{w})}{\varepsilon^3} \right\} + \frac{1}{\varepsilon^2} \log \frac{1}{\delta} \right),$$

$$\mathcal{M}(\varepsilon, \delta, \Omega^{pure}) = \Omega\left(\frac{\hat{w}}{\varepsilon} + \frac{1}{\varepsilon^2} \log \frac{1}{\delta} \right).$$

Remark 4.4 Note that for a sample of size m, when $m \lesssim \hat{w}^3$ (up to log factors), the error scales roughly as $1/m^{1/3}$ and for $m \gtrsim \hat{w}^3$ it scales as $\sqrt{\hat{w}/m}$. Moreover, for $\hat{w} = \mathcal{O}(1)$, our sample complexity bound is tight.

The work of (Guo et al. 2021) proved an improved sample complexity for product distributions. Our results apply for any distribution.

For obtaining the first upper bound, we compute the fat-shattering dimension for the class of revenues with respect to non-increasing pure strategies, and the claim follows from a uniform convergence argument. Define the class,

$$\mathcal{R}^{\text{pure}}_{\geqslant} = \Big\{ z \mapsto r(\mathbf{p}, z) : \mathbf{p} = (p_1, \dots, p_{\hat{w}}) \in [0,1]^{\hat{w}}, p_1 \geqslant \dots \geqslant p_{\hat{w}} \Big\}.$$

Lemma 4.5 *For any $\gamma \in (0, 1/4)$, we have $\operatorname{fat}_\gamma\left(\mathcal{R}^{pure}_{\geqslant}\right) \in [\hat{w}, 2\hat{w}]$. For any $\gamma \in [1/4, 1/2]$, we have $\operatorname{fat}_\gamma\left(\mathcal{R}^{pure}_{\geqslant}\right) \leqslant \hat{w}$. For any $\gamma > 1/2$, we have $\operatorname{fat}_\gamma\left(\mathcal{R}^{pure}_{\geqslant}\right) = 0$.*

We briefly explain how we compute the fat-shattering dimension. For the upper bound, we first "project" the class $\mathcal{R}^{\text{pure}}_{\geqslant}$ on each patience $w \in [\hat{w}]$ to obtain $\hat{w}$ classes with fixed patience. We show that the fat-shattering of any such projected class is exactly 2 for $\gamma \in (0, 1/4)$, 1 for $\gamma \in [1/4, 1/2]$, and 0 for $\gamma > 1/2$. We show that this implies the appropriate upper bound on the fat-shattering dimension of $\mathcal{R}^{\text{pure}}_{\geqslant}$. As for the lower bound, we present a set of $\hat{w}$ buyer types (v_i, i), with values v_i decreasing with patience i. To show that the set is γ-shattered, we take the witness $(c_1, \dots, c_{\hat{w}})$, defined by $c_i = v_i - \gamma$, and prove that for each sequence $(\sigma_1, \dots, \sigma_{\hat{w}}) \in \{-1, +1\}^{\hat{w}}$, there exits a corresponding shattering pricing. The proof is in Appendix B.2, with additional more refined bounds.

By plugging in the fat-shattering dimension to the uniform convergence bound, we obtain a bound on the sample complexity.

Lemma 4.6 *The sample complexity for learning pure selling strategies is*

$$\mathcal{M}(\varepsilon, \delta, \Omega^{pure}) = \mathcal{O}\left(\frac{\hat{w} + \log \frac{1}{\delta}}{\varepsilon^2} \right).$$

We proceed to the second upper bound with a better dependence on $\hat{w}$, albeit a worse dependence on ε. We are doing so by discretizing the set of prices from which the non-increasing strategy is choosing from. The discretization would cause that there are only few changes in prices in the price sequence. This discretization implicitly implies a shorter horizon $\hat{w}$, which can be used (implicitly) to derive the improved bound. The generalization follows from learning a finite class.

Lemma 4.7 *The sample complexity for learning pure selling strategies is*

$$\mathcal{M}(\varepsilon, \delta, \Omega^{pure}) = \mathcal{O}\left(\frac{\log(\hat{w})}{\varepsilon^3} + \frac{1}{\varepsilon^2} \log \frac{1}{\delta} \right).$$

For the regime where $\hat{w} \gtrsim 1/\varepsilon$ (up to log factors), the upper bound in Lemma 4.7 is better, and when $\hat{w} \lesssim 1/\varepsilon$ the bound in Lemma 4.6 gives the sample complexity upper bound.

Concerning the lower bounds, we start with the following simple lower bound, for a distribution where all buyer types have the same patience. This bound follows from a standard claim on the number of samples needed in order to distinguish between a Bernoulli random variable with parameter $1/2$ and a Bernoulli random variable with parameter $1/2+\varepsilon$ (e.g., Slivkins (2019, Lemma 2.7)).

Lemma 4.8 *The sample complexity for learning pure selling strategies is*

$$\mathcal{M}\left(\varepsilon, \delta, \Omega^{pure}\right) = \Omega\left(\frac{1}{\varepsilon^2}\log\frac{1}{\delta}\right).$$

We prove a second lower bound for a set of natural distributions, where buyer types with larger patience window have strictly lower values, i.e., for any (v, w) and (v', w') in the support of $\mathcal{D}$, if $w' > w$ then $v' < v$. In order to prove this lower bound, we first claim that for such distributions, the Bayes optimal is a pure selling strategy which offers prices at each step i, only from values of buyer types with patience i. Note that for such distributions, it suffices to choose the optimal price for each step, independently from the other steps. We then define a family of distributions, such that in order to find the optimal price at step i, the learner should see at least $\approx \frac{1}{\varepsilon}$ samples from patience i. This eventually leads to the following bound.

Lemma 4.9 *Let $\hat{w} \geqslant 2$. The sample complexity for learning pure selling strategies is*

$$\mathcal{M}\left(\varepsilon, \delta, \Omega^{pure}\right) = \Omega\left(\frac{\hat{w}}{\varepsilon}\right).$$

Theorem 4.3 follows immediately by combining Lemmas 4.6 and 4.7 for the upper bound, and Lemmas 4.8 and 4.9 for the lower bound.

4.3 Sample Complexity of Mixed Selling Strategies

In this section, we address the challenging case of learning an approximate optimal mixed strategy from samples. Initially we will assume that the support of $\mathcal{D}$ is finite and use the empirical distribution to approximate it. Later we generalize the result to arbitrary support, using a discretization of the support. (The proofs for this Section are in Appendix B.3).

Let $\mathcal{D}$ be a distribution over buyer types $[0,1] \times [\hat{w}]$, where the size of the set of distinct values $\mathcal{V}$ is at most k, hence, the support of $\mathcal{D}$ is at most $k\hat{w}$. The sample complexity of learning discrete distributions over a known domain of size $k\hat{w}$, with respect to the total variation distance, is $\Theta\left(\frac{k\hat{w}+\log\frac{1}{\delta}}{\varepsilon^2}\right)$.[1]

[1]This is also known as Bretagnolle Huber-Carol inequality, for a proof, see Canonne (2020, Theorem 1).

Using the sample, we learn an approximation $\hat{\mathcal{D}}$ to $\mathcal{D}$ with a total variation distance of at most ε. We then use the distribution $\hat{\mathcal{D}}$ to derive an approximate optimal mixed strategy. We conclude an upper bound on the sample complexity of the set $\Omega_{\mathcal{V}}^{\text{mixed}}$.

Theorem 4.10 *The sample complexity for learning mixed selling strategies $\Omega_{\mathcal{V}}^{mixed}$ is*

$$\mathcal{M}\left(\varepsilon, \delta, \Omega_{\mathcal{V}}^{mixed}\right) = \mathcal{O}\left(\frac{k\hat{w}}{\varepsilon^2} + \frac{1}{\varepsilon^2}\log\frac{1}{\delta}\right).$$

When we have large set $\mathcal{V}$ we can discretize it using the parameter ε. Namely, given $\varepsilon > 0$, we discretize the set of values to multiples of ε, resulting in $k = 1/\varepsilon$ distinct values. We have shown that the error incurred in the discretization process is at most $\varepsilon\hat{w}$ (Lemma A.19). Now, using Theorem 4.10, with an accuracy parameter $\varepsilon/\hat{w}$, we derive the following theorem.

Theorem 4.11 *The sample complexity for learning mixed selling strategies Ω^{mixed} is*

$$\mathcal{M}\left(\varepsilon, \delta, \Omega^{mixed}\right) = \mathcal{O}\left(\frac{\hat{w}^4}{\varepsilon^3} + \frac{\hat{w}^2}{\varepsilon^2}\log\frac{1}{\delta}\right).$$

4.4 Regret Minimization

In this section, we address the online setting, where the buyers arrive in an online fashion, and the seller needs to adjust her strategy. This is a stochastic online setting, where the buyers' types are sampled from an unknown distribution $\mathcal{D}$. The goal of the seller is to minimize the regret w.r.t. a given set of strategies. We naturally consider both the case of pure strategies and the case of mixed strategies.

Online model. In the online setting, the seller is facing a sequence of T buyers $z_1, \dots, z_T$ such that each buyer type z_t is drawn i.i.d. from an unknown distribution $\mathcal{D}$. The seller interacts with only a single buyer at a time, that is, each round of the learning consists of one interaction between the seller and a buyer. At the end of the interaction the seller observes the buyer type (regardless of the outcome). We prove regret minimization results with respect to the selling strategies: (1) pure selling strategies, i.e., Ω^{pure}, (2) mixed selling strategies with prices in $\mathcal{V}$, i.e., $\Omega_{\mathcal{V}}^{\text{mixed}}$, and (3) general mixed strategies, i.e., Ω^{mixed}. (The proofs for this section are in Appendix B.4).

We start with the case of comparing to pure strategies. The main idea is to keep the observed buyers' types. When buyer i arrives, there are already $i-1$ observed buyer types. The seller would use the historical observed buyers' types to implicitly learn the distribution $\mathcal{D}$ to a certain accuracy. More explicitly, the seller would invoke an ERM oracle that would select the best empirical strategy on the historical observations. In order to minimize the number of calls to the ERM oracle, we invoke the ERM oracle only for buyers t which are a power of two, i.e., $t = 2^i$ for some integer i. The difference between the three setting comes from the different convergence rates that we derived for each setting in previous sections.

For the pure strategy setting we use Theorem 4.3 to derive the following regret bound.

Theorem 4.12 *There exists an algorithm $\mathcal{A}$, such that for any distribution D, with probability $1-\delta$,*

$$\textit{Regret}_T^{\Omega^{pure}}(\mathcal{A};\mathcal{D}) = \begin{cases} \mathcal{O}\left(\sqrt{T}\sqrt{\hat{w}} + \sqrt{T}\sqrt{\log \frac{\log T}{\delta}}\right), & T > \frac{\hat{w}^3}{\log^2 \hat{w}} \\ \mathcal{O}\left(T^{2/3}\log^{1/3}(\hat{w}) + \sqrt{T}\sqrt{\log \frac{\log T}{\delta}}\right), & T \leqslant \frac{\hat{w}^3}{\log^2 \hat{w}}, \end{cases}$$

where $\mathcal{A}$ makes $\log T$ calls to an ERM *oracle.*

The bound follows from Theorem 4.3 by setting the accuracy ε_m as a function of the sample size m. For small sample size m we have that $\varepsilon_m \approx \mathcal{O}((\log \hat{w}/m)^{1/3})$, and for large sample size we have that $\varepsilon_m \approx \mathcal{O}(\sqrt{\hat{w}/m})$. The regret is now $\sum_{m=1}^{T} \varepsilon_m$.

For the mixed strategy setting with a limited price set, we use Theorem 4.10 to derive the following regret bound.

Theorem 4.13 *There exists an algorithm $\mathcal{A}$, such that for any distribution D, with probability $1-\delta$,*

$$\textit{Regret}_T^{\Omega_{\mathcal{V}}^{mixed}}(\mathcal{A};\mathcal{D}) = \mathcal{O}\left(\sqrt{T}\sqrt{|\mathcal{V}|\cdot\hat{w}} + \sqrt{T}\sqrt{\log \frac{\log T}{\delta}}\right),$$

where $\mathcal{A}$ makes $\log T$ calls to an ERM *oracle.*

The bound follows from Theorem 4.10 by setting the accuracy ε_m as a function of the sample size m. Here we have that $\varepsilon_m \approx \mathcal{O}(\sqrt{|\mathcal{V}|\hat{w}/m})$. Again, the regret is now $\sum_{m=1}^{T} \varepsilon_m \approx \mathcal{O}(\sqrt{\hat{w}|\mathcal{V}|T})$.

For the setting with general mixed strategies we use Theorem 4.11 to derive the following regret bound.

Theorem 4.14 *There exists an algorithm $\mathcal{A}$, such that for any distribution D, with probability $1-\delta$,*

$$\textit{Regret}_T^{\Omega^{mixed}}(\mathcal{A};\mathcal{D}) = \mathcal{O}\left(T^{2/3}\hat{w}^{4/3} + \sqrt{T}\hat{w}\sqrt{\log \frac{\log T}{\delta}}\right),$$

where $\mathcal{A}$ makes $\log T$ calls to an ERM *oracle.*

We now have the freedom to select the set $\mathcal{V}$ as to minimize the regret. We use a distritization of roughly $T^{1/3}$, which results in the regret bound of order $T^{2/3}$. (The discretization is implicit in the proof of Theorem 4.11.) The bound follows from Theorem 4.11 by setting the accuracy ε_m as a function of the sample size m. Here we have that $\varepsilon_m \approx \mathcal{O}(\hat{w}^{4/3}/m^{1/3})$. Again, the regret is now $\sum_{m=1}^{T} \varepsilon_m \approx \mathcal{O}(T^{2/3}\hat{w}^{4/3})$.

5 Conclusion, Discussion and Open Problems

The main focus of this work is on patient buyers, which can delay their purchasing decision. We presented a new stochastic model where each buyer's type is composed from a value and a patience, and a seller who posts prices and would like to maximize her revenue. We formalize this setting as a Stackelberg game between a leader (the seller) and a follower (the buyer).

Unlike much of the previous works, our focus is on a sequence of prices rather than a single fixed price. For this end, we show a separation between the best fixed price, the best pure strategy, which is a fixed sequence of prices, and the best mixed strategy, which is a distribution over price sequences.

We characterize the optimal pure strategy of the seller and show that the sequence of prices are non-increasing and that the buyers will always buy at the end of their patience, if they decide to buy. We also give an efficient algorithm to compute the optimal pure selling strategy. We derive a sample complexity bound for the learning of an approximate optimal pure strategy which is polynomial in $\hat{w}, 1/\varepsilon$ and $\log(1/\delta)$. We derive our sample bound by computing the fat-shattering dimension of the setting and showing that it is linear in the maximum patience of a buyer, i.e., $\hat{w}$. We also consider an online setting and bound the regret with respect to the optimal pure strategy by $\widetilde{\mathcal{O}}\left(\sqrt{T}\sqrt{\hat{w}}\right)$.

For mixed strategies, we characterize the buyer's best response strategy as a threshold strategy, and show that the expected buyer partial utility decreases with the time steps. We give an algorithm to compute the optimal mixed selling strategy which is exponential in the maximum patience and polynomial in the support of the distribution. We give a general sample complexity bound for the approximate optimal mixed strategy which is polynomial in $\hat{w}, 1/\varepsilon, \log(1/\delta)$. We also consider an online setting and bound the regret with respect to the optimal mixed strategy by $\widetilde{\mathcal{O}}\left(T^{2/3}\hat{w}^{4/3}\right)$.

Our work leaves many interesting open problems.

- **Computational.** It is unclear whether one can compute the optimal mixed selling strategy in time polynomial in $\hat{w}$, or maybe there is a hardness result as in other Stackelberg games with large action spaces (Blum et al. 2019).
- **Sample complexity.** There is a gap between our upper and lower bounds for the sample complexity both for the pure and mixed strategies. Resolving those gaps would be highly interesting.
- **Online learning.** First, it is unclear whether one can get a regret bound of $\sqrt{T}$ with respect to the optimal mixed strategy, or whether there is a lower bound of $T^{2/3}$. Another interesting challenge for the online learning is to consider more limited feedback models, for example, when we observes only whether a purchase was made.
- **Interaction between buyers.** In our model the seller interacts with each buyer separately. It would be interesting to consider a model where there are potentially multiple buyers per interaction. At each step there are some buyers that arrive and other buyers that did not purchase at the previous step, and can still benefit from buying. Such a model will introduce many new and intriguing research challenges.

Acknowledgments

We deeply thank Aryeh Kontorovich and Yuval Dagan for very helpful discussions. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 882396), by the Israel Science Foundation (grants 993/17, 1602/19), Tel Aviv University

Center for AI and Data Science (TAD), and the Yandex Initiative for Machine Learning at Tel Aviv University. I.A. is supported by the Vatat Scholarship from the Israeli Council for Higher Education, Kreitman School of Advanced Graduate Studies, and the Lynn and William Frankel Center for Computer Science at Ben-Gurion University.

References

Alon, N.; Ben-David, S.; Cesa-Bianchi, N.; and Haussler, D. 1997. Scale-sensitive dimensions, uniform convergence, and learnability. *Journal of the ACM (JACM)*, 44(4): 615–631.

Amin, K.; Rostamizadeh, A.; and Syed, U. 2013. Learning Prices for Repeated Auctions with Strategic Buyers. In Burges, C. J. C.; Bottou, L.; Ghahramani, Z.; and Weinberger, K. Q., eds., *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, 1169–1177.

Amin, K.; Rostamizadeh, A.; and Syed, U. 2014. Repeated Contextual Auctions with Strategic Buyers. In Ghahramani, Z.; Welling, M.; Cortes, C.; Lawrence, N. D.; and Weinberger, K. Q., eds., *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, 622–630.

Balcan, M.; Blum, A.; Hartline, J. D.; and Mansour, Y. 2008. Reducing mechanism design to algorithm design via machine learning. *J. Comput. Syst. Sci.*, 74(8): 1245–1270.

Block, A.; Dagan, Y.; and Rakhlin, A. 2021. Majorizing Measures, Sequential Complexities, and Online Learning. In *Conference on Learning Theory*, 587–590. PMLR.

Blum, A.; Haghtalab, N.; Hajiaghayi, M.; and Seddighin, S. 2019. Computing Stackelberg Equilibria of Large General-Sum Games. In *International Symposium on Algorithmic Game Theory*, 168–182. Springer.

Canonne, C. L. 2020. A short note on learning discrete distributions. *arXiv preprint arXiv:2002.11457*.

Devanur, N. R.; Goldner, K.; Saxena, R. R.; Schvartzman, A.; and Weinberg, S. M. 2020. Optimal mechanism design for single-minded agents. In *Proceedings of the 21st ACM Conference on Economics and Computation*, 193–256.

Dudley, R. M. 1967. The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *Journal of Functional Analysis*, 1(3): 290–330.

Feldman, M.; Koren, T.; Livni, R.; Mansour, Y.; and Zohar, A. 2016. Online pricing with strategic and patient buyers. In *Advances in Neural Information Processing Systems*, 3864–3872.

Fiat, A.; Goldner, K.; Karlin, A. R.; and Koutsoupias, E. 2016. The fedex problem. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, 21–22.

Gonczarowski, Y. A.; and Nisan, N. 2017. Efficient empirical revenue maximization in single-parameter auction environments. In Hatami, H.; McKenzie, P.; and King, V., eds., *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, 856–868. ACM.

Guo, C.; Huang, Z.; Tang, Z. G.; and Zhang, X. 2021. Generalizing complex hypotheses on product distributions: Auctions, prophet inequalities, and pandora's problem. In *Conference on Learning Theory*, 2248–2288. PMLR.

Hartline, J. D. 2021. Mechanism design and approximation. http://jasonhartline.com/MDnA/. Accessed: 2022-02-11.

Haussler, D. 1992. Decision theoretic generalizations of the PAC model for neural net and other learning applications. *Information and computation*, 100(1): 78–150.

Huang, Z.; Mansour, Y.; and Roughgarden, T. 2018. Making the Most of Your Samples. *SIAM J. Comput.*, 47(3): 651–674.

Kearns, M. J.; and Schapire, R. E. 1994. Efficient distribution-free learning of probabilistic concepts. *Journal of Computer and System Sciences*, 48(3): 464–497.

Kleinberg, R.; and Leighton, T. 2003. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.*, 594–605. IEEE.

Koren, T.; Livni, R.; and Mansour, Y. 2017a. Bandits with movement costs and adaptive pricing. In *Conference on Learning Theory*, 1242–1268. PMLR.

Koren, T.; Livni, R.; and Mansour, Y. 2017b. Multi-Armed Bandits with Metric Movement Costs. In Guyon, I.; von Luxburg, U.; Bengio, S.; Wallach, H. M.; Fergus, R.; Vishwanathan, S. V. N.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, 4119–4128.

Mendelson, S.; and Vershynin, R. 2003. Entropy and the combinatorial dimension. *Inventiones mathematicae*, 152(1): 37–55.

Mohri, M.; and Medina, A. M. 2015. Revenue Optimization against Strategic Buyers. In *NIPS*, 2530–2538. Citeseer.

Mohri, M.; and Munoz, A. 2014. Optimal Regret Minimization in Posted-Price Auctions with Strategic Buyers. In Ghahramani, Z.; Welling, M.; Cortes, C.; Lawrence, N.; and Weinberger, K. Q., eds., *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc.

Mohri, M.; Rostamizadeh, A.; and Talwalkar, A. 2018. *Foundations of machine learning*. MIT press.

Morgenstern, J.; and Roughgarden, T. 2015. On the Pseudo-Dimension of Nearly Optimal Auctions. In Cortes, C.; Lawrence, N. D.; Lee, D. D.; Sugiyama, M.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, 136–144.

Morgenstern, J.; and Roughgarden, T. 2016. Learning Simple Auctions. In Feldman, V.; Rakhlin, A.; and Shamir, O., eds., *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, 1298–1318. Columbia University, New York, New York, USA: PMLR.

Myerson, R. B. 1979. Incentive compatibility and the bargaining problem. *Econometrica: journal of the Econometric Society*, 61–73.

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

Nisan, N.; Roughgarden, T.; Tardos, E.; and Vazirani, V. V. 2007. *Algorithmic Game Theory*. New York, NY, USA: Cambridge University Press.

Pollard, D. 1990. Empirical processes: theory and applications. In *NSF-CBMS regional conference series in probability and statistics*, i–86. JSTOR.

Rudelson, M.; and Vershynin, R. 2006. Combinatorics of random processes and sections of convex bodies. *Annals of Mathematics*, 603–648.

Saxena, R. R.; Schvartzman, A.; and Weinberg, S. M. 2018. The menu complexity of "one-and-a-half-dimensional" mechanism design. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2026–2035. SIAM.

Slivkins, A. 2019. Introduction to multi-armed bandits. *arXiv preprint arXiv:1904.07272*.

Vanunts, A.; and Drutsa, A. 2019. Optimal Pricing in Repeated Posted-Price Auctions with Different Patience of the Seller and the Buyer. In Wallach, H.; Larochelle, H.; Beygelzimer, A.; d'Alché-Buc, F.; Fox, E.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

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