A Survey on Sample Complexity of Mechanism Design

Austen Z. Fan
zfan64@wisc.edu

Sihan Liu
sliu556@wisc.edu

Ziyang Liu
zliu695@wisc.edu

December 17, 2020

1 Introduction

Mechanism design can be seen as the study of algorithm design with inputs from strategic bidders. This field initially emerged as the combination of auction theory, game theory and computer science. There has been significant progress on this topic since 2000.

One important direction is to find out an optimal mechanism for selling items which maximizes the revenue earned. When perfect information of the distribution over the buyers' valuations are given, there are well-established theorems and methodologies which allow designers to fine-tune a mechanism for maximum expected revenue ([HL10]).

However, in practice, sellers rarely have perfect information over the consumer population. This then leads to the important question whether one can still implement efficient mechanisms given just sample access of past consumer data.

While there are some experimental works that investigate how pricing of mechanism can be adjusted with information of past data ([OS11]), this survey focuses on some recent theoretical progress on how good a mechanism based on empirical data could be compared to the optimal one with perfect information. In particular, it visits import results from [HMR18], [CR14], [MR16] and outlines the analysis practiced in these works.

The logical dependence is such starting from Single Item & Single Buyer, then Single Item & Multiple Buyers, and finally Multiple Items & Multiple Buyers. Our survey is mainly based on three papers and by no means inclusive. However, it could serve as a starting point for whom willing to delve more into this topic.

2 Single Item, Single Buyer

In real-life mechanism design, the distribution of a buyer's valuation is usually unknown. In their paper [HMR18], the authors consider the single-item, single-buyer setting, where the buyer's valuation is drawn from an unknown distribution D. The seller only has information about D in the form of m i.i.d. samples. The authors prove the number of samples that are necessary and sufficient to obtain a $(1-\epsilon)$-approximation to the optimal revenue. That is, given these m samples, we can determine a reserve price p that gives a revenue at least $(1-\epsilon)$ times the optimal revenue.

The main result of this paper is that, $\widetilde{O}(\epsilon^{-3/2})$ samples are sufficient for MHR distributions. It is worth mentioning that, this problem is different from estimating the distribution D itself, because the optimal reserve price from an estimated distribution might not give a near-optimal approximation to the true optimal revenue. Moreover, this problem is also different from estimating the optimal revenue directly. As a direct corollary of a result in information theory, we can show that $\widetilde{O}(\epsilon^{-2})$ samples are necessary to estimate the expected revenue by even a single fixed reserve price. In this sense, the paper shows that near-optimal revenue maximization is a strictly easier problem than learning even the very simple statistics of a distribution.

Distribution	Upper Bound	Lower Bound
MHR	$O(\epsilon^{-3/2}\log\epsilon^{-1})$	$\Omega(\epsilon^{-3/2})$
Regular	$O(\epsilon^{-3}\log\epsilon^{-1})$	$\Omega(\epsilon^{-3})$
General	$O(\delta^{-1}\epsilon^{-2}\log(\delta^{-1}\epsilon^{-1}))$	$\Omega(\delta^{-1}\epsilon^{-2})$
Bounded Support	$O(H\epsilon^{-2}\log(H\epsilon^{-1}))$	$\Omega(H\epsilon^{-2})$

The results of sample complexity bounds from this paper are summarized in the table above. For bounded-support distributions, the support is $[1, H]$. For general distributions, instead of considering the optimal revenue R^*, we consider R^*_δ, which is the optimal revenue from reserve prices with sale probability q at least some $\delta \in [0,1]$. In other words, we do not consider the low quantiles that correspond to high values. The reason is that, if the optimal reserve price has sale probability at least δ, then approximating R^*_δ is equivalent to approximating R^*. This is the case for sufficiently small δ and most of the typical "reasonable" distributions. Even if not all distributions satisfy this property, this approach still enables parameterized sample complexity bounds that do not require distributional assumptions such as regularity.

2.1 Asymptotic Upper Bounds

To obtain the asymptotic sample complexity upper bounds, we first need the following definition.

Definition 2.1. *Given m samples $v_1 \geq \cdots \geq v_m$, the empirical reserve is*

$$\arg\max_{v_i} i \cdot v_i.$$

If we only consider $i \geq cm$ for some $c \in [0,1]$, then it is called the c-guarded empirical reserve.

The main result of this paper gives the sample complexity upper bound for MHR distributions.

Theorem 2.2. *The empirical reserve with $m = \Theta(\epsilon^{-3/2}\log\epsilon^{-1})$ samples is $(1-\epsilon)$-approximate for all MHR distributions.*

To prove Theorem 2.2, we use the following two properties of MHR distributions. Lemma 2.3 states that the optimal quantile of an MHR distribution is at least $\frac{1}{e}$. Lemma 2.4 states that the revenue decreases quadratically in distance between the reserve price and the optimal one in quantile space.

Lemma 2.3. *For all MHR distributions,* $q^* \geq \frac{1}{e}$.

Lemma 2.4. *For all MHR distributions,* $R(q^*) - R(q) \geq \frac{1}{4}(q^* - q)^2 R(q^*)$ *for every* $q \in [0, 1]$.

Proof sketch for Theorem 2.2: First, for any two samples v_1, v_2 with quantiles q_1, q_2 such that either $q_1 < q_2 < q^*$ or $q^* < q_1 < q_2$, if the revenue of one of them is at least $(1 - \frac{\epsilon}{2})$ times smaller than that of the other one, then with probability at least $1 - o(\frac{1}{m^2})$, the algorithm would choose the latter sample. Furthermore, with high probability, there is at least one sample between both $[(1 - \frac{\epsilon}{2})q^*, q^*]$ and $[q^*, (1 + \frac{\epsilon}{2})q^*]$ in quantile space. These samples are $(1 - \frac{\epsilon}{2})$-optimal by concavity of the revenue curve. The theorem then follows from union bound.

The sample complexity upper bound for general distributions is given in the following theorem.

Theorem 2.5. *The* $\frac{\delta}{2}$*-guarded empirical reserve with* $m = \Theta(\delta^{-1}\epsilon^{-2}\log(\delta^{-1}\epsilon^{-1}))$ *gives revenue at least* $(1 - \epsilon)R^*_\delta$ *for all distributions.*

Proof sketch for Theorem 2.5: Let q^*_δ be the optimal sale probability at least δ. First, with high probability, there exists at least one sampled price with quantile between $[(1 - \frac{\epsilon}{3})q^*_\delta, q^*_\delta]$. This price has revenue at least $(1 - \frac{\epsilon}{3})R^*_\delta$. Then since $q^*_\delta \geq \delta$, with high probability, this price has rank at least $\frac{\delta}{2}$ among sampled prices. So it is considered by the empirical reserve algorithm. Furthermore, with high probability, any sampled price with rank at least $\frac{\delta}{2}$ has sale probability at least $\frac{\delta}{4}$. Lastly, with high probability, for prices with sale probability at least $\frac{\delta}{4}$, the algorithm estimates their sale probability up to a $(1 - \frac{\epsilon}{3})$ factor with $m = \Theta(\delta^{-1}\epsilon^{-2}\log(\delta^{-1}\epsilon^{-1}))$ samples.

For distributions with bounded support $[1, H]$, the optimal sale probability q^* is at least $1/H$. This is because the optimal reserve price p^* is at most H and the optimal revenue R^* is at least 1, as we can set the reserve price to be $p = 1$ and obtain a sale probability $q = 1$. Therefore we have the following theorem as a direct corollary of Theorem 2.5.

Theorem 2.6. *The empirical reserve with* $m = \Theta(H\epsilon^{-2}\log(H\epsilon^{-1}))$ *samples is* $(1-\epsilon)$*-approximate for all distributions with bounded support* $[1, H]$.

2.2 Asymptotic Lower Bounds

The authors borrow techniques from differential privacy and information theory to give the asymptotic sample complexity lower bounds. The high-level plan is to reduce the pricing problem to a classification problem. First, we need to introduce some information theory preliminaries. Consider two distributions P_1 and P_2 over a sample space Ω. Let p_1 and p_2 be the corresponding probability density functions. We have the following definitions and theorems.

- Statistical Distance: $\delta(P_1, P_2) = \frac{1}{2}\int_\Omega |p_1(\omega) - p_2(\omega)|\, d\omega$.
- KL Divergence: $\mathrm{KL}(P_1 \| P_2) = \mathbf{E}_{\omega \sim P_1}\left[\ln \frac{p_1(\omega)}{p_2(\omega)}\right]$.
- Pinsker's Inequality: $\delta^2(P_1, P_2) \leq \frac{1}{2}\mathrm{KL}(P_1 \| P_2)$.
- No classification algorithm can distinguish P_1 and P_2 with probability greater than $\frac{\delta(P_1, P_2)+1}{2}$.

One nice property of the KL divergence is that it is additive over samples: if P_1 and P_2 are the distributions over m samples of D_1 and D_2, then $\mathrm{KL}(P_1 \| P_2) = m \cdot \mathrm{KL}(D_1 \| D_2)$. In addition, to distinguish P_1 and P_2 correctly with probability at least $\frac{2}{3}$, the statistical distance between P_1 and P_2 should be at least $\frac{1}{3}$. Furthermore, Pinsker's inequality relates the KL divergence and the statistical distance. It implies that the statistical distance of m samples from D_1 and D_2 is upper bounded by $\frac{1}{2}\sqrt{m \cdot (\mathrm{KL}(D_1 \| D_2) + \mathrm{KL}(D_2 \| D_1))}$.

Using techniques in differential privacy, we can construct two distributions D_1 and D_2 with small KL divergence but disjoint approximately optimal price sets. Then we can show that, if there exists a pricing algorithm that is $(1-\epsilon)$-approximate for both D_1 and D_2, then there exists a classification algorithm that distinguishes P_1 and P_2 correctly with probability at least $\frac{2}{3}$, using the same number of samples as the pricing algorithm. By the arguments above, the classification algorithm requires the statistical distance between P_1 and P_2 to be at least $\frac{1}{3}$, and we need at least $m = \frac{4}{9} \cdot \frac{1}{\mathrm{KL}(D_1 \| D_2) + \mathrm{KL}(D_2 \| D_1)}$ samples for this. This is exactly the number of samples necessary for the pricing algorithm. Therefore, we obtain a sample complexity lower bound as intended.

3 Single Item, Multiple Buyers

In this section, we extended our discussion to the setting of single item and multiple buyers. This section is mainly based on the paper by Cole and Roughgarden in 2014. To be more concrete, we will assume that there are k bidders and that bidder i's distribution is drawn from F_i where $1 \le i \le k$ and F_i's are independent but not necessarily identical. Denote the joint distribution $F = F_1 \times F_2 \times \cdots \times F_k$. The main result is as follows: samples of the number polynomial in k and $1/\epsilon$ are *necessary and sufficient* to achieve a $(1-\epsilon)$ approximation of the optimal revenue.

3.1 m-sample auction strategy and α-strongly regular distribution

Suppose m independent and identically distributed samples $v_1, \ldots, v_m$ are drawn from F. In learning theory, the performance of an algorithm depends on the data. That is, the expected revenue of the auction resulting from samples should condition on samples. This motivates the definition m-sample auction strategy which will be used to measure the performance of the auction.

Definition 3.1. *An m-sample auction strategy is a map from the m samples to an auction. The expected revenue of a m-auction strategy is with respect to both the samples $v_1, v_2, \ldots, v_m$ and the testing input v_{m+1}.*

We remark that under this language, the expected revenue of an optimal auction is with respect to a single sample from F which is just the input of the auction.

To obtain useful results, we have to further restrict our attention to some special class of allowable distributions. This is motivated by considering the following distribution class:

$$f_M = \begin{cases} M^2 & \text{with probability } 1/M \\ 0 & \text{with probability } 1 - 1/M \end{cases}$$

The optimal auction will earn expected revenue of M by pricing at M^2. However, for any fixed m, there is no optimal m-sample auction strategy with near-optimal revenue for this class of distributions. This is because for sufficiently large M, all m samples are 0 with high probability and thus the auction strategy will have no information about what that M really is. The right notion seems to be what is called the α-strongly regular distribution.

Definition 3.2. *Let F be a distribution with positive density function f on its support $[a, b]$, where $0 \leq a < \infty$ and $a \leq b \leq \infty$. Let $\varphi(v) = v - \frac{1-F(v)}{f(v)}$ denote the corresponding virtual valuation function. We call F to be an α-strongly regular distribution if*

$$\varphi(y) - \varphi(x) \geq \alpha(y - x)$$

whenever $y > x \geq 0$.

We note that for a differentiable virtual valuation function ϕ, it is α-strongly regular if and only if its derivative is greater than or equal to α at every point. This is a generalization of regular distributions and the MHR distributions since they are the special cases when $\alpha = 0$ and $\alpha = 1$ respectively. The authors gave some evidence why the definition of α-strongly regular distribution seems to be the right notion in terms of the following quantile estimation.

Lemma 3.3. *Let F be an α-strongly regular distribution with $\alpha \in (0, 1)$ and monopoly price r. If $q(r)$ is the quantile of valuation r in the distribution F, then $q(r) \geq \alpha^{1/(1-\alpha)}$.*

Setting the limit for $\alpha = 1$ we recover the well-known property of MHR distributions as in Lemma 2.3.

3.2 The lower bound

Proving lower bound is notoriously known in computational complexity. In that paper, however, Cole and Roughgarden managed to show samples of number polynomial in k and $\frac{1}{\epsilon}$ are necessary for a $(1 - \epsilon)$ approximation.

Theorem 3.4. *For every auction strategy Σ, for every $k \geq 2$, for every sufficiently small $\epsilon > 0$, for every $\alpha \geq 0$ and m satisfying:*

- $\alpha = 1$ *and* $m \leq \left(\frac{1-\ln 2}{96e^3 \min\{1, \frac{k}{e}\} \ln \max\{e,k\}}\right)^{1/2} \frac{k}{\sqrt{\epsilon}}$;
- $0 < \alpha < 1, \alpha^{1/(1-\alpha)} \geq \frac{1}{k}$, *and* $m \leq \left(\frac{1-\alpha 2^{1-\alpha}}{96e^3}\right)^{1/(1+\alpha)} \frac{k}{\epsilon^{1/(1+\alpha)}}$;
- $0 < \alpha < 1, \frac{1}{2m} < \alpha^{1/(1-\alpha)} < \frac{1}{k}$, *and* $m \leq \left(\frac{1-\alpha 2^{1-\alpha}}{96e^3}\right)^{1/(1+\alpha)} \left(\frac{1}{k\alpha^{1/(1-\alpha)}}\right)^{\alpha/(1+\alpha)} \frac{k}{\epsilon^{1/(1+\alpha)}}$;
- $0 < \alpha < 1, \alpha^{1/(1-\alpha)} \leq \frac{1}{2m}$, *and* $m \leq \frac{(1-\alpha 2^{1-\alpha})2^\alpha}{96e^3} \frac{k}{\epsilon}$;
- $\alpha = 0$ *and* $m \leq \frac{1}{96e^3} \frac{k}{\epsilon}$,

there exists a set $F_1, \ldots, F_k$ of α-strongly regular valuation distributions such that the expected revenue of Σ (over the m samples and the input) is less than $1 - \epsilon$ times that of an optimal auction for $F_1, \ldots, F_k$.

This theorem essentially says that in most cases we need at least linear or near-linear in the number of bidders k that amount of samples in order to achieve a good approximation, even when we are restricted to α-strongly regular distributions. We note that at the two extreme cases when $\alpha = 0$ and $\alpha = 1$, namely for regular distributions or MHR distributions, the lower bounds are of the order $\frac{k}{\epsilon}$ and $\frac{k}{\sqrt{\epsilon ln k}}$ respectively.

This lower bound forms a contrast to other previous results. For example, in the paper by Dhangwatnotai, Roughgarden and Yan in 2010, they showed that for single item single buyer setting and if the distribution is regular, then samples of the number polynomial many in $1/\epsilon$ suffice for a $(1-\epsilon)$ approximation. Also, in the same paper, they showed that for single item, multi buyers but with identical distribution, and further if the distribution is regular, then samples of the number polynomial many in $1/\epsilon$ also suffice for a $(1-\epsilon)$ approximation. The lower bound result by Cole and Roughgarden result is in contrast to previous upper bounds because the polynomials there do not depend on k. This contrast shows that the interplay between bidder competition fundamentally changes the sampling complexity. Another result proved by Hartline and Roughgarden in 2009 says that in the setting of single item and multi buyers with not necessarily identical regular distributions, if only a quarter approximation of the optimal expected revenue is required, then literally a single sample suffices. Thus the requirement of a nearly optimal approximation also plays a fundamental role.

The proof to Theorem 3.4 is essentially by construction. The idea is to consider the most "tail-heavy" distributions for any fixed α. That is,

$$\begin{array}{ll} F^{\alpha}(v) = 1 - \left(\frac{1}{1+(1-\alpha)v}\right)^{\frac{1}{1-\alpha}} & \text{for } 0 \leq \alpha < 1 \\ F^{1}(v) = 1 - e^{-v} & \text{for } \alpha = 1 \end{array}$$

They showed that for any fixed α and any auction strategy Σ, there exists a set of $F_1, \ldots, F_k$ α-strongly regular distributions such that when the number of samples is below what is stated in Theorem 3.4, the expected revenue of that auction strategy Σ is not nearly optimal. The exact statement for the above sentence is technical. The idea is to define a distribution over α-strongly regular distributions as the following. Each bidder i is either type A or type B with equal possibility. If a bidder i is of type B, a number q is drawn uniformly from the interval $\left[\frac{\delta}{2k}, \frac{\delta}{k}\right]$ and the value $H_i = (F^{\alpha})^{-1}(1-q)$ is computed. The bidder i 's distribution F_i is then defined to be F^{α} on $[0, H_i)$ with a point mass with the remaining probability $1 - F^{\alpha}(H_i)$ at H_i. If a bidder j is of type A, then the distribution is exactly the same except its H_j is set to $(F^{\alpha})^{-1}\left(1 - \frac{\delta}{2k}\right)$.

3.3 The upper bound

The upper bound restricts back to regular distributions. This is probably due to simplicity consideration since the proof for regular distributions is already highly technical.

Theorem 3.5. *In a single-item auction with k bidders with independent regular valuation distributions, if $m = \Omega\left(\frac{k^{10}}{\epsilon^7} \ln^3 \frac{k}{\epsilon}\right)$, then there is an m -sample auction strategy with expected revenue at least $1-\epsilon$ times that of an optimal auction.*

This is proved by analyzing the so-called "empirical Myerson auction". Essentially, the empirical Myerson auction does what is natural to do, i.e. draw the "empirical revenue curve" for each bidder i using the samples and then run the optimal Myerson's auction on the empirical ironed virtual valuations. The algorithm for drawing the empirical ironed revenue curve for each bidder i is stated as follows:

- Suppose that the m independent samples drawn from F_i have values $v_{i,1} \geq v_{i,2} \geq \ldots \geq v_{i,m}$. Define the "empirical quantile" of $v_{i,j}$ as $\frac{2j-1}{2m}$.
- Discard the $\lfloor \hat{\xi} m \rfloor - 1$ largest samples, for a suitable $\hat{\xi} > 0$. Let S denote the remaining samples.
- For each remaining sample $v_{i,j} \in S$, plot a point $\left(\frac{2j-1}{2m}, \frac{2j-1}{2m} v_{ij}\right)$.
- Add points at (0,0) and (1,0) .
- Take the convex hull - the least concave upper bound - of this point set. Denote the resulting "ironed empirical revenue curve" by $\overline{\text{CR}}_i$.
- The empirical ironed virtual value for v is defined to be v itself if $v > v_{i,\hat{\xi}m}$ and to be the slope of the revenue curve $\overline{\text{CR}}_i$ in the interval defined by the empirical quantiles of $v_{i,j}$ and $v_{i,j+1}$ where $v_{i,j}$ and $v_{i,j+1}$ are two samples in S sandwiching the value v otherwise.

We remark that the closed formula for $\hat{\xi}$ in step 2 is not explicitly given in the paper. However, this turns out to be an easily computable value of the order $O(\frac{\epsilon}{k} + 1/2m)$.

We summarize here the high-level point of view for how those lemma pieces in the paper are organized. The first thing to do is to establish a relation between the empirical quantile and the true expected quantile. This argument is standard by the deviation bound, the Chernoff bound. With high probability and for almost all values, the empirical quantile is a good multiplicative approximations to the their true expectation.The next step is to show that with high probability and for almost all quantiles, the empirical virtual value $\bar{\phi}(q)$ is sandwiched between the true virtual values with small multiplicative parameters. Formally speaking, Cole and Roughgarden showed that we can always choose suitale δ_1 and δ_2 such that $\bar{\phi}(q)$ is sandwiched between $\phi(q(1+\delta_1))$ and $\phi(q/(1+\delta_2))$ with some additive factors which is a function of k and ϵ. This gives us a revenue bound when the empirical Myerson's auction turns out to allocate the item to the same bidder as the optimal Myerson's auction. The last part is to bound the revenue loss which is caused in the following scenario: the empirical virtual valuation of a different bidder j might be larger than its true value, which is the nature of sampling, and then it causes the empirical Myerson's auction to allocate to j instead of the rightful winner i. Cole and Roughgarden proved that the revenue reduction caused by this scenario is also under control.

The discussion for one item and multiple buyers setting is now complete.

4 Multiple Item, Multiple buyers, Distribution Independent

Most of the previous results rely on structural properties of the valuation distribution. In this section, we investigate the paper [MR16], where the authors explore possibilities of achieving guarantees in the distribution-independent setting.

To achieve that, we need to borrow tools from learning theory where we aim at searching for an optimal *hypothesis* in a class of candidate hypothesis. In general, the more complicated the search space, the harder it is to find solutions that enjoy good generalization properties. Similar tradeoff happens when we want to *learn* the optimal mechanism from sample data. In particular, the more complex the auction, more sample we need in order to ensure good generalization property. Hence, in this section, we mainly focus on simple auction class such as pricing and second-price bidding.

4.1 Revenue maximization and PAC learning

PAC learning (acronym for probably approximately correct learning) is a important model in learning theory. To formalize the problem, we are first given a class of functions $\mathcal{H} := \{h : \mathcal{X} \mapsto \mathcal{Y}\}$. Over there, there is a fixed, albeit unknown, hypothesis h_* often denoted as the target. The goal is to give an (ϵ, δ)-learner whose definition we give as follow.

Definition 4.1. *An algorithm $\mathcal{A}$ is denoted as an (ϵ, δ) learner if, with oracle access of the form $(x, h_*(x))$ sampled as $x \sim \mathcal{D}$, it can output a hypothesis $\tilde{h} \in \mathcal{H}$ satisfying*

$$err\left(h_*(x)\,, \tilde{h}(x)\right) \leq \epsilon$$

with probability at least $1 - \delta$.

Usually, given the samples, the algorithm $\mathcal{A}$ itself is just the *Empirical Risk Minimizer* (E.R.M). Namely, given m pairs $(x^{(i)}, h_*(x^{(i)}))$, we simply output $\text{argmin}_{h \in \mathcal{H}} \sum_{i \in [m]} \text{err}\left(h_*(x^{(i)}), h(x^{(i)})\right)$. The main result we will rely on from learning theory is a bound on the generalization error of the E.R.M hypothesis. At a high level, the theorem states that the average error of a hypothesis won't differ too much with its expected error over the entire population if the number of samples are large enough.

As we have briefly discussed, the bound crucially depends on the "complexity" of the inner structure of the class of hypothesis considered. When the hypothesis is a real-valued function, one commonly used metric is the *Pseudo-Dimension* ($\mathcal{PD}$).

Definition 4.2 (Pseudo-dimension)**.** *Let $\mathcal{F} \subseteq \mathbb{R}^{\mathcal{X}}$ and let $x_1^m = (x_1, \cdots, x_m) \in \mathcal{X}^m$. We say x_1^m is pseudo-shattered by $\mathcal{F}$ if there exists $\mathbf{r} = (r_1, \cdots, x_m) \in \mathbb{R}^m$ such that for any $\mathbf{b} = (b_1, \cdots, b_m) \in \{-1, 1\}^m$, there exsits $f_{\mathbf{b}} \in \mathcal{F}$ such that $sign(f_{\mathbf{b}}(x_i) - r_i) = b_i$ for all $i \in [m]$. The pseudo-dimension of $\mathcal{F}$ is the cardinality of the largest set of points in $\mathcal{X}$ that can be pseudo-shattered by $\mathcal{F}$:*

$$\mathcal{PD}(\mathcal{F}) = \max\left\{m \in \mathbb{N} \,\middle|\, \textit{there exists } x_1^m \in \mathcal{X}^m \textit{ such that } x_1^m \textit{ is pseudo-shattered by } \mathcal{F}\right\}.$$

Relying on the notion of pseudo-dimension, we state the following bound on generalization error.

Theorem 4.3. *Suppose $\mathcal{F}$ is a class of real-valued functions with range in $[0, H]$ and pseudo-dimension $\mathcal{PD}(\mathcal{F})$. For every $\epsilon > 0, \delta \in [0, 1]$, the sample complexity of (ϵ, δ)-uniformly learning the class $\mathcal{F}$ is*

$$n = O\left(\left(\frac{H}{\epsilon}\right)^2 \left(\mathcal{PD}(\mathcal{F}) \ln \frac{H}{\epsilon} + \ln \frac{1}{\delta}\right)\right).$$

So far, all the tools are stated in the learning theory context. We now need to connect it to revenue maximization of auction. The connection is not explicitly spelled out in the work of [MR16] but is touched in a previous work [MR15].

At a high level, an auction class can be regarded as a class of revenue functions mapping from the valuation tuples of customers $\mathcal{V}^n$ to the revenue i.e., $\mathcal{F} : \mathcal{V}^n \mapsto \mathbb{R}$.

Among all revenue functions, there is a "target hypothesis" f^* corresponding to the optimal auction. Given a set of past customer data $\mathbf{v}^{(1)}, \cdots, \mathbf{v}^{(m)}$, our goal is to output a hypothesis $\tilde{f} \in \mathcal{F}$ that tries to minimize $f^*(\mathbf{v}) - \tilde{f}(\mathbf{v})$ where $\mathbf{v}$ is a unforeseen future valuation tuple coming from the same underlying distribution.

Notice that, unlike traditional supervised learning task, the label of the target hypothesis f^* is not given. Hence, we could no longer compute the E.R.M solution with respect to the past samples. Nevertheless, we can see that *Empirical Revenue Maximizer* achieves a similar effect.

Conceptually, we can compare the revenue difference between both f^* and $\tilde{f}$ to a "base" revenue function f_0 that maps everything to 0 (the revenue function of an auction that does not sell anything). By Theorem 4.3, we obtain that the empirical revenue is a good estimation over the population revenue for any auction. Since the empirical revenue of $\tilde{f}$ is always at least that of f^* (since $\tilde{f}$ is the maximizer), the population revenue of $\tilde{f}$ with high probability will be ϵ-close to that of f^*.

4.2 Linear-Separability and (a, b)-Factorable Class

The remaining work is to bound the pseudo-dimension of a class of auction. This section offers a framework for doing so. As a slight digression, we first visit the concept of linear-separability (L.S) that is developed in the context of multi-classification, where the hypothesis takes the form $\mathcal{F} : \mathcal{X} \mapsto \mathcal{Q}$.

Definition 4.4. *A class $\mathcal{F}$ is d-dimensionally linearly separable if there exists a function $\psi : \mathcal{V} \times \mathcal{Y} \mapsto \mathbb{R}^d$ and for any $f \in \mathcal{F}$, there exists some $\mathbf{w}^f \in \mathbb{R}^d$ with $f(v) \in argmax_y \langle \mathbf{w}^f . \psi(v, y) \rangle$ and $|argmax_y \langle \mathbf{w}^f, \psi(v, y) \rangle| = 1$.*

Intuitively, the function ψ and the vector $\mathbf{w}^f$ offers ways to encode input and hypothesis as high dimensional vectors such that the classification rule becomes argmax over a simple linear expression with respect to the encoding vectors. Similar to the notion of pseudo-dimension, the higher the dimensions needed to perform the encoding, the more complex the hypothesis class and the poorer its generalization ability.

The reason we mention multi-classification and its related measure of class complexity is that the allocation rule of an auction is exactly a multi-classification task. In particular, the rule receives input from the space $\mathcal{V}^n$ and outputs an allocation in $\{0, 1\}^{n \cdot k}$, where specifies whether customer $i \in [n]$ is allocated item $j \in [k]$.

Moreover, the allocation rule offers an effective way to decompose the revenue function of a large class of auctions: whenever the auction can be expressed as some pricing rule, its revenue would be simply the sum of the price of all allocated items. With such decomposition in mind, the pseudo-dimension of the revenue (or the complexity of it) is essentially determined by the complexities of two rather loosely coupled components: the allocation rule and the pricing.

At a high level, the allocation rule offers a way to "bucket" the revenue functions in an auction class. Fixing the allocation outcome, the revenue function is usually reduced to nothing but a linear combination of some pricing values. Hence, bucketing makes it easier to bound the pseudo-dimension. To formalize the idea, the authors give the definition of (a, b)-factorability.

Definition 4.5 ((a, b)-factorable class)**.** *Consider some $\mathcal{F} = \{f : \mathcal{X} \mapsto \mathbb{R}\}$. Suppose, for each $f \in \mathcal{F}$, there exists $(f_1, f_2), f_1 : \mathcal{X} \mapsto Y, f_2 : \mathcal{Y} \times \mathcal{X} \mapsto \mathbb{R}$ such that $f_2(f_1(x), x) = f(x)$ for all $x \in \mathcal{X}$. Let*

$$\mathcal{F}_1 = \{f_1 : (f_1, f_2) \textit{ is a decomposition of some } f \in \mathcal{F}\}$$

and

$$\mathcal{F}_2 = \{f_2 : (f_1, f_2) \textit{ is a decomposition of some } f \in \mathcal{F}\}.$$

The set $\mathcal{F}$ is (a, b)-factors over $\mathcal{Q}$ if:

1. *$\mathcal{F}_1$ is a-dimensionally linealy separable over $\mathcal{Q} \in \mathcal{Y}$.*
2. *For every $f_1 \in \mathcal{F}_1$ and sample $S \in \mathcal{X}$ of size m, the set*

   $$\mathcal{F}_{2|f_1(S)} = \{f_2' : \mathcal{X} \mapsto \mathbb{R}, f_2'(x) = f_2(f_1'(x), x) | f_1(S) = f_1'(S) \textit{ and } (f_1', f_2) \textit{ realizes some } f \in \mathcal{F}\}$$

   has pseudo-dimension at most b.

If a hypothesis class admits an (a, b)-factoring over the set $\mathcal{Q}$, which the authors refer as the *Intermediate Label Space*, the overall pseudo dimension can be bounded by a, b and the cardinality of $\mathcal{Q}$.

Theorem 4.6. *Suppose $\mathcal{F}$ is (a, b)-factorable over $\mathcal{Q}$. Then,*

$$\mathcal{PD}(\mathcal{F}) = O(\max((a + b)\ln(a + b)), a \ln |Q|).$$

The size of the intermediate label space plays an important role in the bound. Hence, it is crucial to optimize it in order to obtain tight analysis. In particular, it may be possible to introduce a *relabeling* $q : \mathcal{Q} \mapsto Q'$ which reduces the size of the intermediate label space.

In terms of auction design, the intermediate labels contain exactly the information of the allocation outcome of an auction. However, the information is often redundantly detailed while used to compute the final revenue. In particular, for auction that does not discriminate among customers, it does not matter to whom an item is sold. Hence, we can introduce a relabeling which shrinks the intermediate space to just $\{0, 1\}^k$ (whether item $j \in [k]$ is sold or not). The following remark gives the interface for performing such optimization.

Remark 4.7. *Suppose $\mathcal{F}$ is d-dimensionally linearly separable over $\mathcal{Q}$. Fix some $q : \mathcal{Q} \mapsto \mathcal{Q}'$. Then, the set $q \circ \mathcal{F} = \{q \circ f | f \in \mathcal{F}\}$ is d-dimensionally linearly separable over $\mathcal{Q}'$.*

4.3 $\mathcal{PD}$ of simple pricing auction

To begin with, we apply the framework on an auction with 1 customer and 1 item (or in other word, a grand-bundle pricing auction). Then, any truthful mechanism is just a pricing. Namely,

$$\text{Alloc}^h(v) = \mathbb{1}\{v \geq p^h\} \tag{1}$$

$$\text{Rev}^h(v) = \text{Alloc}^h(v) \cdot p^h \tag{2}$$

Claim 4.8. *Let $\mathcal{F}$ be the class of anonymous grand bundle pricing. Then $\mathcal{PD}(\mathcal{F}) = O(1)$.*

Proof Sketch. For the allocation rule with the pricing rule p, we can use the encoding $\mathbf{w}^h = (1, p^h)$ and $\psi(v, b) = \mathbb{1}(b = 1) \cdot (v, -1)$. Hence, the function class is 2-dimensional linearly separable.

Fixing the allocation outcome $b \in \{0, 1\}$, the revenue function is reduced to $\text{Rev}^h(v) = b \cdot p^h$, which is just a constant function in 1-d. Obviously, the pseudo-dimension is 2.

Hence, the auction class is $(2, 2)$ factorable over $Q = \{0, 1\}$. Using Theorem 4.6, the auction class has pseudo dimension $O(1)$. □

The approach can be easily generalized to the k items, 1 customer setting. Next, we will see how auctions facing multiple customers can be similarly analyzed by treating them as sequential execution of a single-customer allocation rule to each customer in order.

Definition 4.9 (n-fold anonymous and non-anonymous sequential allocations)**.** *Let $\mathcal{H}$ be some class with $h : \mathcal{V} \times \{0, 1\}^k \mapsto \mathcal{Q}$ for all $h \in \mathcal{H}$ and some $\mathcal{Q} \in \{0, 1\}^k$. For some n functions $h_1, \cdots, h_n \in \mathcal{H}$ and every $\mathbf{v} \in \mathcal{V}^n$, inductively define $X_1(\mathbf{v}) = [k], X_i(\mathbf{v}) = X_{i-1}(\mathbf{v}) \backslash h_{i-1}(\mathbf{v}_{i-1}, X_{i-1}(\mathbf{v}))$. Then, define the n-wise product function $\prod_{(h_1, \cdots, h_n)}$ to be*

$$\prod_{(h_1, \cdots, h_n)} (\mathbf{v}) = \left((h_1(\mathbf{v}_1), X_1(\mathbf{v})), (h_2(\mathbf{v}_2), X_2(\mathbf{v})), \cdots (h_n(\mathbf{v}_n), X_n(\mathbf{v}))\right).$$

Then, we call any such function an n-wise non-anonymous sequential allocation. If $h_1 = h_2 \cdots = h_n$, it is called the n-wise anonymous sequential allocation drawn from $\mathcal{H}$.

We are now ready to present the analysis for pricing based auctions facing multiple customers and multiple items.

Claim 4.10. *Let $\mathcal{F}$ be the class of anonymous item pricings. Then $\mathcal{PD}(\mathcal{F}) = O(k^2)$. If $\mathcal{F}$ is the class of non-anonymous item pricings, then $\mathcal{PD}(\mathcal{F}) = O(nk^2 \ln(n))$.*

Proof Sketch. Following a similar argument as Claim 4.8, a unit-allocation which assigns a subset of k-item to one single customer is $(k + 1)$-separable. By definition, it means that

$$\text{Alloc}^h(\mathbf{v}) = \text{argmax}_{b \in \mathcal{Q}} \langle \psi(\mathbf{v}, b), \mathbf{w}^h \rangle,$$

where $\mathbf{w}^h$ and $\psi(\mathbf{v}, b)$ are vectors in $\mathbb{R}^{(k+1)}$.

We will show how we could use it to encode the rule of sequential allocations that performs $h^{(1)}, \cdots h^{(n)}$ in order, where each unit allocation is $(k + 1)$ linearly separable.

Following the definition of sequential allocation, the overall allocation can be written as

$$\text{Alloc}^{h^{(1)}, \cdots h^{(n)}}(\mathbf{v}) = \text{argmax}_{\mathbf{b} \in \mathcal{Q}} \sum_{i=1}^{n} \left(2^i \cdot C\right) \langle \psi^{(i)}(\mathbf{v}_i, \mathbf{b}_i), \mathbf{w}^{h_i} \rangle,$$

where C is a large-enough constant used to enforce the sequence of allocation. In particular, if C is large enough, $\mathbf{b}_i$, the allocation outcome to customer i, is always chosen in priority compared to

$\mathbf{b}_j$ where $j < i$.

It is easy to see that the whole expression is an inner product in the space of dimension $(k+1) \cdot n$. Moreover, if we further impose the constrain that the allocation is anonymous. We have that $\mathbf{w}^{h_1} = \cdots \mathbf{w}^{h_n}$. By linearity of the inner product, we can group the expression as

$$\left\langle \left(\sum_{i=1}^{n} \left(2^i \cdot C\right) \psi^{(i)}(\mathbf{v}_i, \mathbf{b}_i) \right), \mathbf{w}^h \right\rangle.$$

Then, it is easy to see the allocation rule is still $(k+1)$ dimensionally linearly separable.

It remains to bound the pseudo-dimension of the revenue function with fixed allocation outcome. For non-anonymous auction, the revenue function is reduced to $\sum_{i,j} \mathbf{b}_{i,j} \cdot \mathbf{p}_{i,j}$, where $b_{i,j}$ is the indicator variable for whether item i is sold to customer j and $\mathbf{p}_{i,j}$ is the pricing of item i shown to customer j. So the revenue is reduced to a constant function in space of dimension $n \cdot k$.

Overall, the non-anonymous auction class is $(O(nk), O(nk))$-factorable over the intermediate label space $\mathcal{Q} \subseteq [n]^k$. By Theorem 4.6, we have pseudo-dimension of $O(nk^2 \ln(n))$.

For the anonymous auction, we only need the information on whether an item is sold or not. Hence, we can use the optimization trick outlined in Remark 4.7. Then, the intermediate label space can be reduced to $\{0,1\}^k$. Further, the revenue function with fixed allocation outcome becomes a constant function in space of dimension k.

Overall, the anonymous auction class is $(k+1, k+1)$-factorable over the intermediate label space with cardinality 2^k. By Theorem 4.6, we have pseudo-dimension of $O(k^2)$. (so the generalization ability does not become poorer as we have more customers) □

Lastly, we present a rather surprising result from the authors showing that second-price auction with reserved price can also be analyzed as a sequential auction as long as the customers are additive-buyers.

Claim 4.11. *Suppose $\mathcal{V}$ is some set of additive valuations. Let $\mathcal{F}$ be the class of second-price item auctions with anonymous reserves. Then, $\mathcal{PD}(\mathcal{F}) = O(k^2)$. If $\mathcal{F}$ is the class of second-price item auctions with anonymous reserves, then $\mathcal{PD}(\mathcal{F}) = O(nk^2 \ln(n))$.*

Proof Sketch. It is not obvious why an auction based on bidding can be treated as sequential allocations since the outcome depends on the bids coming from all customers. However, it should be noticed that the property that "items are only allocated to highest bidders" is not specific to an instance of second-price auction (more specifically, it does not depend on the reserve price) but rather a global property satisfied by all second-price auctions. This makes it possible to impose the constrain on the intermediate label space rather than a specific allocation rule. The following remark offers the interface for implementing such constrain.

Remark 4.12. *Suppose for each $x \in \mathcal{X}$, there exists some $\mathcal{Q}_x \subseteq \mathcal{Q}$ such that $f_1(x) \in \mathcal{Q}_x$ for all $f_1 \in \mathcal{F}_1$, and that for each x, $\mathcal{F}_1$ is linearly separable in a dimensions for that x over $\mathcal{Q}_x$. Assume there is a subset of dimension $T^+ \subseteq [a]$ for which $\mathbf{w}^f_{t \in T^+} \geq 0$ and $\sum_{t \in T^+} \mathbf{w}^f_t > 0$ for all f. Suppose that for all $x \in \mathcal{X}$, $f \in \mathcal{F}_1$, $\max_{y \in \mathcal{Q}_x} \psi(x, y) \cdot \mathbf{w}^f \geq 0$. Then, $\mathcal{F}_1$ is a-linearly separable over $\mathcal{Q}$.*

Intuitively, the remark allows us to bound linear separability on a smaller intermediate space $\mathcal{Q}_x$ customized for an input. We can therefore enforce the highest-bidder constrain on the space directly rather than making it part of the allocation rule.

With such simplification, the unit allocation rule becomes clear: we could allocate item i to customer j if he/she is the highest bidder and his/her bid exceeds the reserved price. The situation hence becomes similar to the one in Claim 4.10. As a result, we get the allocation rule is $k+1$ dimensionally linearly separable for anonymous auction and $n(k+1)$ linearly separable for non-anonymous auction.

It remains to bound the pseudo-dimension of revenue with fixed allocation outcome. Here, we focus on anonymous reserve price as the two cases are similar.

Fixing the allocation outcome, the revenue function becomes

$$\mathrm{Rev}(\mathbf{v}) = \sum_j \max(\mathbf{p}_j, \max_{i' \neq i_j^*} \mathbf{v}_{i'}(\{j\})) \cdot f_1(\mathbf{v})_j.$$

For each item j, if the relative ordering of p_j^f and $\max_{i' \neq i_j^*} \mathbf{v}_{i'}(\{j\})$ were fixed, the revenue function would again be a constant function in a space of dimension k. Given m sample data, it means that $\mathcal{F}$ can induce at most m^{k+1} labelings (the labeling is the one used in Definition 4.2).

For each item j, p_j^f can have $m+1$ different ranking over the array of $\max_{i' \neq i_j^*} \mathbf{v}_{i'}(\{j\})$ in the m sampled data points. Since there are k items, it can induce at most $(m+1)^k$ different ordering. So in the worse case, the revenue function can induce $m^{k+1} \cdot (m+1)^k$ different labels. By Definition 4.2, the pseudo dimension is bounded by $O(k \ln k)$.

Overall, second-price auction with anonymous reserved price is $(O(k), O(k \ln k))$-factorable over $\mathcal{Q} \in \{0,1\}^k$. Similarly, second-price auction with anonymous reserved price is $(O(nk), O(nk \ln(nk)))$-factorable over $\mathcal{Q} \in [n]^k$. By Theorem 4.6, their pseudo-dimensions are bounded by $O(k^2)$ and $O(nk^2 \ln(n))$ respectively.

□

Acknowledgement

We sincerely thank Professor Shuchi Chawla for her wonderful lectures and kind support in the course COMP SCI 880 in Fall 2020.

References

[CR14] Richard Cole and Tim Roughgarden. The sample complexity of revenue maximization. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 243–252, 2014.

[HL10] Jason D Hartline and Brendan Lucier. Bayesian algorithmic mechanism design. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 301–310, 2010.

[HMR18] Zhiyi Huang, Yishay Mansour, and Tim Roughgarden. Making the most of your samples. *SIAM Journal on Computing*, 47(3):651–674, 2018.

[MR15] Jamie H Morgenstern and Tim Roughgarden. On the pseudo-dimension of nearly optimal auctions. *Advances in Neural Information Processing Systems*, 28:136–144, 2015.

[MR16] Jamie Morgenstern and Tim Roughgarden. Learning simple auctions. In *Conference on Learning Theory*, pages 1298–1318, 2016.

[OS11] Michael Ostrovsky and Michael Schwarz. Reserve prices in internet advertising auctions: A field experiment. In *Proceedings of the 12th ACM conference on Electronic commerce*, pages 59–60, 2011.