6.853 Technical Report

Angelos Assos, Angelos Pelecanos, Aristofanis Rontogiannis

May 28, 2021

Contents

1 Introduction **2**

2 Estimating Bidder Distribution **2**
- 2.1 Estimation Algorithm . . . 3
- 2.2 Analysis . . . 3

3 A $(6 + \epsilon)$-Approximate Mechanism for One Bidder **5**

4 Extending to Multiple Bidders **6**
- 4.1 Estimating Bidder Valuations . . . 6
- 4.2 A $(8 + \epsilon)$-Approximate Mechanism for Multiple Bidders . . . 6

5 Acknowledgements **8**

A Contextual Mechanism Design **10**
- A.1 Classification Problem . . . 10

B Extending [BDHS15] to Multiple Bidders **11**
- B.1 Semi-Independent Setting . . . 11
- B.2 Remaining Reduction . . . 11

C Known Base Improvements **12**
- C.1 Better Than 12-Approximation . . . 12
- C.2 Using SDPs . . . 13

1 Introduction

Maximizing revenue is a fundamental problem in auction design. Auctioneers often run auctions to optimally sell many items to many bidders. For this to be possible, we assume that prior information about the bidders' values is known to the auctioneer. Ideally, the auctioneer knows the distributions from which the bidder's valuations for specific items are drawn, however in most cases the auctioneer only has samples from these distributions. The auctioneer has two tasks: first, approximate the bidders' distributions from the samples provided, and second, use those approximate distributions to design an auction that has revenue close to that of the optimal auction. A long line of research has tackled each of these two tasks separately, and recently there have been efforts to solve the two problems when the bidders' valuations are independent. In this project, we are exploring the case when items' values have a slight correlation in the form of a common shift to the values. Such an assumption seems realistic, since it can be used to model auctions when a specific type of item is sold (e.g. livestock, works of painting, cars). In that case, a bidders' value for the items is affected by how much they value the general category of items being sold.

Before we dive into the technical aspect of our research, we introduce some notation. Each bidder has a binary p-dimensional feature vector x that is known to the auctioneer. This can be thought of as features of the bidder that can be observed. We also have θ, a common p-dimensional vector for all items, which is unknown. We are tackling the high-dimensional case, when p is large. To make this problem more structured, we assume that θ is s-sparse, i.e. the number of non-zero components of θ is at most s. Finally, each item has its own valuation distribution F_j supported on the interval $[0, H]$. We assume that the values drawn from the F distributions are independent for each bidder and item.

Since we are tackling learning, we assume that we are given a history of q bidders in the form of pairs (x_i, v_{ij}). That is, we know that q bidders with feature vector x_i had value for item j equal to v_{ij}. Knowing this history, we are asked to sell the same m items to a new (or multiple new) bidder with feature vector x (which might have not been observed before). The formal model for the value of a bidder with feature vector x_i for item j is, where $\epsilon_{ij} \sim F_j$:

$$v_{ij} = \langle \theta, x_i \rangle + \epsilon_{ij}$$

The remaining of this paper is organized as follows. Section 2 contains our estimation algorithm and the proof that our empirical distribution is close to the true distribution. Sections 3 and 4 show how to use our empirical estimation to design simple and almost-optimal mechanisms for a single and multiple bidders respectively. In the appendices, we have included multiple other ideas that we have explored throughout the semester. We would like to emphasize our efforts to extend the result of [BDHS15] to multiple bidders. This result, although not directly related to our project, is another contribution of our work.

2 Estimating Bidder Distribution

From now on, we will consider the single bidder case. Extending to multiple bidders is possible and will be covered in Section 4.

Our framework assumption is that we are selling a set of m items and the bidders valuations for these items are not independent. In particular, we are considering that the bidder's value for the i^{th} item is equal to a 'base' term that equals the preference of the bidder for the item plus some 'error' that can be thought to cover other factors that affect the bidder's valuation.

$$v_j = \langle \theta, x \rangle + \epsilon_j$$

As explained in the Introduction, θ is a common parameters for all the items and x is the feature vector of the bidder. The error term ϵ_j is sampled from the error distribution F_j, which is considered to be independent for each item.

Normalizing the Errors. Thus, we can see that our problem boils down to estimating the error distributions and the shared θ vector. When estimating these parameters, we assume that the F_j distributions have a mean value of 0.

To justify this assumption, consider the case when the error distributions are not 0-meaned. This will create an offset which can be recovered by extending the θ and x vectors. The construction will be as follows. Let $\mu \in \mathbb{R}^m$ be the mean of the error distributions F_j. Let $F'_j = F_j - \mu_j$ be the centered error distribution that has a zero mean. We can also define a new vector θ_j for each item, equal to θ plus an extra coordinate equal to μ_j. Then, if we extend the feature vector of each bidder x' to have an extra coordinate with a value of 1, we can verify that

$$\langle \theta_j, x' \rangle + \epsilon'_j = \langle \theta, x \rangle + \epsilon'_j + \mu_j = \langle \theta, x \rangle + \epsilon_j$$

Using this assumption and by assuming that the samples we are given along with error distributions satisfy some favorable conditions given in [Wai09], we can assume that using Sparse High-Dimensional Linear Regression will recover our feature vector θ with some error.

Formally, we can recover an s-sparse vector $\hat{\theta}$ that satisfies

$$\|\theta - \hat{\theta}\|_2 \leq \sqrt{\frac{s \log p}{q}}$$

$$\Rightarrow \|\theta - \hat{\theta}\|_1 \leq \sqrt{s}\|\theta - \hat{\theta}\|_2 \leq s\sqrt{\frac{\log p}{q}}$$

2.1 Estimation Algorithm

Our estimation algorithm is simple. We will use Sparse High Dimensional Regression to estimate the shared θ vector. Then, we will use our empirical θ vector to obtain empirical error estimates and we will use these to learn the error distribution.

Algorithm 1 Learning Optimal Multi-Item Auctions with Side Information

```
Given samples {(v_ij, x_i)}
Estimate θ̂ using Sparse High-Dim Regression on v_ij ∼ x_i
Compute residual errors ε̂_ij = v_ij − ⟨θ̂, x_i⟩
Estimate discretized empirical distributions F̂_i from ε̂_ij
For a new bidder with feature vector x, compute their value for item j from ⟨θ̂, x⟩ + F̂_j
```

2.2 Analysis

In this subsection we show how to bound the Prokhorov distance between the empirical value distribution for item j and the true value distribution for the same item. Intuitively, we want to bound the difference between the two distributions in order to show that a good mechanism for the empirical distribution will also behave approximately as well in the true distribution. More details are in presented in Section 3.

We emphasize that we want to bound the Prokhorov distance for each item j. Since we have m items, we will try to estimate the distribution to an $\frac{\epsilon}{m}$ error.

Theorem 1. *Let $D_j = \langle \theta, x \rangle + F_j$ and $\hat{D}_j = \langle \hat{\theta}, x \rangle + \hat{F}_j$ be the true and discretized empirical distribution of the values of bidder x for item j. Then given at least $q = O\left(\frac{m^2 s^2 \log p}{\epsilon^2} + \frac{m^3 H(\log 1/\delta + \log m)}{\epsilon^2}\right)$ samples*

$$\|D_j - \hat{D}_j\|_P \leq \frac{\epsilon}{m}, \quad w.p. \ \geq 1 - \frac{\delta}{m}$$

Proof. We use a hybrid argument. We will define an intermediate distribution $\tilde{D}_j$. Then we will bound the Prokhorov distance between two adjacent distributions using Lemmas 1 and 2, which will bound the distance between D_j and $\hat{D}_j$ using the triangle inequality. The intermediate distribution is

$$\tilde{D}_j = \langle \hat{\theta}, x \rangle + F_j$$

We will show that $\|D_j - \tilde{D}_j\|_P \leq \frac{\epsilon}{2m}$ and $\|\tilde{D}_j - \hat{D}_j\|_P \leq \frac{\epsilon}{2m}$. Thus, by the triangle inequality we can conclude that

$$\|D_j - \hat{D}_j\|_P \leq \|D_j - \tilde{D}_j\|_P + \|\tilde{D}_j - \hat{D}_j\|_P$$
$$\leq \frac{\epsilon}{2m} + \frac{\epsilon}{2m} = \frac{\epsilon}{m}$$

□

Lemma 1. *Let* $D_j = \langle \theta, x \rangle + F_j$ *and* $\tilde{D}_j = \langle \hat{\theta}, x \rangle + F_j$. *Then given at least* q *samples guarantees that*

$$\|D_j - \tilde{D}_j\|_P \leq \frac{\epsilon}{2m}$$

Proof. The difference between the two distributions is their base terms $\langle \theta, x \rangle$ and $\langle \hat{\theta}, x \rangle$. Their ℓ_1 difference is from our use of LASSO from [Wai09]

$$\begin{aligned} |\langle \theta, x \rangle - \langle \hat{\theta}, x \rangle| &= |\langle \theta - \hat{\theta}, x \rangle| \\ &\leq \|\theta - \hat{\theta}\|_1 \\ &\leq s\sqrt{\frac{\log p}{q}} \\ &\leq s\sqrt{\frac{\log p}{\frac{4m^2 s^2 \log p}{\epsilon^2}}} = \frac{\epsilon}{2m} \end{aligned}$$

Hence the Prokhorov distance of the two distributions is at most $\frac{\epsilon}{2m}$, since we can always couple their samples such that their ℓ_1 distance is at most $\frac{\epsilon}{2m}$. □

Lemma 2. *Let tsse34ruj*$\tilde{D}_j = \langle \hat{\theta}, x \rangle + F_j$ *and* $\hat{D}_j = \langle \hat{\theta}, x \rangle + \hat{F}_j$. *Then*

$$\|\tilde{D}_j - \hat{D}_j\|_P \leq \frac{\epsilon}{2m}, \quad \textit{w.p.} \ \geq 1 - \frac{\delta}{m}$$

Proof. We extend the proof of Lemma 10 from [BCD20]. We will define the η-*rounding* procedure, where we replace each element x of a distribution by $\left\lfloor \frac{x}{\eta} \right\rfloor \cdot \eta$.

Using this rounding procedure, we will have the following distributions

- F_j: The true distribution of the valuation errors for item j
- $\tilde{F}_j$: The distribution obtained after $\frac{\epsilon}{m}$-rounding each element of the support of F_j
- $\hat{F}_j$: The uniform distribution over the $\frac{\epsilon}{m}$-rounded samples $\hat{\epsilon}_{ij}$.

Since the support of $\tilde{F}_j$ has size $\frac{mH}{\epsilon}$, we know that with $N = O\left(\frac{m^3 H(\log 1/\delta + \log m)}{\epsilon^3}\right)$ samples we will achieve

$$\|\tilde{F}_j - \hat{F}_j\|_{TV} \leq \frac{\epsilon}{2m}, \quad \text{w.p.} \ \geq 1 - \frac{\delta}{m}$$

Now, we will show that with the same number of samples we can create a Prokhorov coupling between F_j and $\hat{F}_j$. In particular, we can create a coupling γ_1 between F_j and $\tilde{F}_j$ such that their samples are at most $\frac{\epsilon}{2m}$ away in absolute value. Additionally, there exists a coupling γ_2 between $\tilde{F}_j$ and $\hat{F}_j$ such that their samples are equal with probability $\geq 1 - \frac{\epsilon}{2m}$ with probability $\geq 1 - \frac{\delta}{m}$. Hence, by composing the two couplings together, we know that there exists a coupling γ between F_j and $\hat{F}_j$ such that samples from the two distributions are within $\frac{\epsilon}{2m}$ absolute value with probability at least $1 - \frac{\epsilon}{2m}$. Thus, this means that with probability $\geq 1 - \frac{\delta}{m}$, $\|F_j - \hat{F}_j\|_P \leq \frac{\epsilon}{2m}$. □

3 A $(6 + \epsilon)$-Approximate Mechanism for One Bidder

The celebrated work of [BILW14] has provided a simple and optimal mechanism for selling multiple items to a single bidder with independent valuations. This mechanism is choosing the best of selling each item separately (denoted as $SRev$) or bundling them all together (denoted as $BRev$). Since there is a single bidder, to maintain incentive compatibility both mechanisms will post prices for the items or the bundle. Let $Rev(D)$ be the revenue achieved from such a simple mechanism when run for valuations drawn from distribution D. Then

$$Rev(D) \geq \frac{1}{6} OPT(D)$$

In this section we show that we can modify the above mechanism for the empirical valuation distribution to get an approximately-optimal one for the true distribution. To do this, we will use discounting.

Definition 1. *Let $\mathcal{M}$ be an $SRev$ mechanism with 1 bidder and m items, with r_j be the posted price for each item j. Define the δ-discounted $SRev$ mechanism $\mathcal{M}_\delta$ that sells the same m items by reducing all posted prices by δ. That is $r_j^\delta = r_j - \delta$.*

Definition 2. *Let $\mathcal{M}$ be $BRev$ mechanism with 1 bidder and m items, with R be the posted price for the bundle. Define the δ-discounted $BRev$ mechanism $\mathcal{M}_\delta$ for the same m items to be the mechanism with the bundle posted price reduced by δ. That is $R^\delta = R - \delta$.*

Theorem 2. *Let $D, \hat{D}$ be the true and empirical valuation distributions for all bidders respectively such that the Prokhorov distance between D_i and $\hat{D}_i$ is at most $\frac{\epsilon}{m}$. Consider the optimal Rev mechanism $\hat{\mathcal{M}}$ for $\hat{D}$. Let $\hat{\mathcal{M}}_\epsilon$ be the mechanism obtained when discounting*

- *The $SRev$ part of $\hat{\mathcal{M}}$ by $\frac{\epsilon}{m}$*
- *The $BREV$ part of $\hat{\mathcal{M}}$ by ϵ*

Then

$$Rev(D, \hat{\mathcal{M}}_\epsilon) \geq \frac{1}{6} OPT(D) - 2\epsilon$$

To prove the above theorem, we make use of the following key lemma.

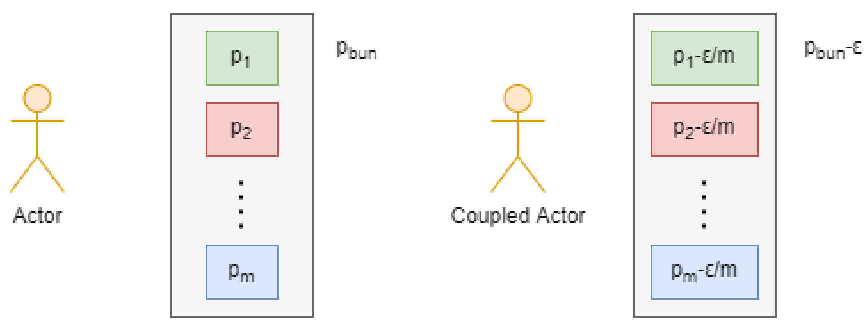

Figure 1: The constructive proof of Lemma 3. We couple the valuations of the bidders from P, Q and modify the mechanism for P (left) such that only ϵ from the revenue is lost for Q (right).

Lemma 3. *Let P, Q be two valuation distributions of a single bidder for m items, such that the valuation distribution for item j satisfies*

$$\|P_j - Q_j\|_P \leq \frac{\epsilon}{m}$$

Let $\mathcal{M}^P$ be the optimal Rev mechanism for P. If $\mathcal{M}_\epsilon^P$ is the discounted mechanism obtained as in Theorem 2, then

$$Rev(Q, \mathcal{M}_\epsilon^P) \geq Rev(P) - \epsilon$$

Proof. From the Prokhorov distance of the valuation distributions, we can couple the sampled bidders' values from P and Q such that they are all within $\frac{\epsilon}{m}$ with probability $\geq 1-\epsilon$. In that case, compare the execution of the mechanism $\mathcal{M}_\epsilon^P$ with values drawn from Q and the execution of $\mathcal{M}^P$ with values drawn from P.

We can see that in the $SRev$ mechanism, if a bidder buys a specific item in $\mathcal{M}^P$, then their coupled bidder will also buy that specific item in $\mathcal{M}_\epsilon^P$. This is because the price of the item has reduced by $\frac{\epsilon}{m}$, whereas the value of the bidder for that item has reduced less than that.

When it comes to the $BRev$ mechanism, the total value of the bidder for the bundle (assuming an additive valuation) has decreased by at most $m \cdot \frac{\epsilon}{m} = \epsilon$ for the bidder in $\mathcal{M}_\epsilon^P$. Thus, by reducing the posted price by ϵ we again guarantee that if the bundle is purchased in $\mathcal{M}^P$, it is also purchased in $\mathcal{M}_\epsilon^P$, for at most ϵ less.

Hence the required inequality holds, since changing the distribution can only cost us an ϵ quantity from the revenue. □

We now use Lemma 3 twice to prove Theorem 2.

Proof of Theorem 2. Let $\hat{\mathcal{M}}$ be the Rev mechanism constructed for the valuation distribution $\hat{D}$. From Lemma 3 we know that

$$Rev(D, \hat{\mathcal{M}}_\epsilon) \geq Rev(\hat{D}) - \epsilon$$

Additionally, let $\mathcal{M}$ be the optimal Rev mechanism constructed for the valuation distribution D. Lemma 4 implies that

$$Rev(\hat{D}) \geq Rev(\hat{D}, \mathcal{M}_\epsilon) \geq Rev(D) - \epsilon$$

Combining the two above inequalities we get

$$Rev(D, \hat{\mathcal{M}}_\epsilon) \geq Rev(\hat{D}) - \epsilon \geq Rev(D) - 2\epsilon \geq \frac{1}{6} OPT(D) - 2\epsilon$$

□

4 Extending to Multiple Bidders

4.1 Estimating Bidder Valuations

In this subsection we briefly argue that our current estimation algorithm as presented in Algorithm 1 can estimate with the same guarantees the valuation distributions for many bidders. Indeed, the error distributions F_j only depend on the number of items and the shared vector θ is common among all bidders. Thus, estimating F_j and θ using enough samples will provide us with an accurate empirical valuation distribution for each new bidder that arrives as presented in Theorem 1.

We will use this insight to extend the results of the previous section to the scenario when we are selling to multiple bidders that arrive, each of which has an individual feature vector. We do note that we in order to get a good approximate revenue, we need to estimate each bidder's valuation distribution to a Prokhorov distance of $\frac{\epsilon}{nm}$. This increases the number of required samples by a polynomial factor in the number of bidders.

4.2 A $(8+\epsilon)$-Approximate Mechanism for Multiple Bidders

The celebrated work of [CDW16] has provided a simple mechanism that obtains an 8-approximation of the optimal revenue when selling multiple independent items to many bidders. This mechanism has two components which we will denote as $SRev$ and $BVCG$, following standard notation introduced in [CDW16].

The components. $SRev$ is the mechanism that sells each item separately using second-price auctions with reserve prices. From Myerson's seminal work [Mye81] we know that such mechanism is optimal when selling a single item.

$BVCG$ is defined as a VCG auction with entry fees. In particular, each bidder is given a personalized entry fee that is only dependent on the bids of the other bidders. They then have to pay this entry fee to be eligible to participate in separate VCG auctions for each item.

Formally, Cai et al. [CDW16] showed that the best of $SRev$ and $BVCG$ mechanisms provide at least an eight of the optimal revenue when selling to independent items to multiple bidders. What is left is to decide the optimal entry fees and reserve prices for our learned valuation distribution. From now on, we will refer to this simple and approximately optimal mechanism as Rev and to the optimal mechanism as OPT. A similar line of reasoning as in Section 3 is presented, where we will discount the optimal Rev mechanism for $\hat{D}$ and show that it is almost optimal when selling to the true valuation distribution D.

We know that for any valuation distribution D that is a product of independent distributions for each bidder, we can obtain

$$Rev(D) \geq \frac{1}{8} OPT(D)$$

Let us first define the discounting procedures for $SRev$ and $BVCG$ mechanisms.

Definition 3. *Let $\mathcal{M}$ be an SRev mechanism with n bidders and m items, with r_j be the reserve price for each item j. Define the δ-discounted SRev mechanism $\mathcal{M}_\delta$ that sells m items to n bidders to be the mechanism with reserve prices reduced by δ. That is $r_j^\delta = r_j - \delta$.*

Definition 4. *Let $\mathcal{M}$ be a $BVCG$ mechanism with n bidders and m items, with e_i be the personalized entry fee computed after each bidder submits their bids. Define the δ-discounted $BVCG$ mechanism $\mathcal{M}_\delta$ that sells m items to n bidders to be the mechanism with entry fees reduced by δ. That is $e_i^\delta = e_i - \delta$.*

We are now ready to state our almost-8-approximate mechanism.

Theorem 3. *Let $D, \hat{D}$ be the true and empirical valuation distributions for all bidders respectively such that the Prokhorov distance between D_{ij} and $\hat{D}_{ij}$ is at most $\frac{\epsilon}{nm}$. Consider the optimal Rev mechanism $\hat{\mathcal{M}}$ for $\hat{D}$. Let $\hat{\mathcal{M}}_\epsilon$ be the mechanism obtained when discounting*

- *The SRev part of $\hat{\mathcal{M}}$ by $\frac{\epsilon}{nm}$*
- *The $BVCG$ part of $\hat{\mathcal{M}}$ by $\frac{\epsilon}{n}$*

Then

$$Rev(D, \hat{\mathcal{M}}_\epsilon) \geq \frac{1}{8} OPT(D) - 2\epsilon$$

The proof follows by extending section 3. The key lemma is the following.

Lemma 4. *Let P, Q be two valuation distributions of n bidders for m items, such that the valuation distribution of bidder i for item j satisfies*

$$\|P_{ij} - Q_{ij}\|_P \leq \frac{\epsilon}{nm}$$

Let $\mathcal{M}^P$ be the optimal Rev mechanism for P. If $\mathcal{M}_\epsilon^P$ is the discounted mechanism obtained as in Theorem 3, then

$$Rev(Q, \mathcal{M}_\epsilon^P) \geq Rev(P) - \epsilon$$

Proof. From the Prokhorov distance of the valuation distributions, we can couple the sampled bidders' values from P and Q such that they are all within $\frac{\epsilon}{nm}$ with probability $\geq 1 - \epsilon$. In that case, compare the execution of the mechanism $\mathcal{M}_\epsilon^P$ with values drawn from Q and the execution of $\mathcal{M}^P$ with values drawn from P.

We can see that in the $SRev$ mechanism, due to the IC property, the bidders will still report their true values. Thus, any item allocated in $\mathcal{M}^P$ is also allocated in $\mathcal{M}_\epsilon^P$. The only difference is that the reserve prices have decreased. This might decrease the revenue per item by at most $\frac{\epsilon}{nm}$. Thus, the total revenue loss is at most $\frac{\epsilon}{n}$.

When it comes to the $BVCG$ mechanism, the bidders will still report their true values. Additionally, if a bidder pays the entry fee in $\mathcal{M}^P$, then it is because their expected utility from the auction is higher than the entry fee. However, their expected utility can at most decrease by $\frac{\epsilon}{nm}$ per item, thus for a total of $\frac{\epsilon}{n}$. By decreasing the entry fee by this quantity, we are guaranteeing that any bidder that pays the entry fee in $\mathcal{M}^P$, their coupled bidder will also pay the entry fee in $\mathcal{M}_\epsilon^P$.

Then, the subset of bidders entering the $\mathcal{M}_\epsilon^P$ mechanism is larger than or equal to the bidders entering $\mathcal{M}^P$. Their bids have now at most decreased again by $\frac{\epsilon}{nm}$, thus the total revenue loss is at most $\frac{\epsilon}{n}$ as with $SRev$. However, the largest loss is from the entry fees, which can be up to $\frac{\epsilon}{n}$ per bidder, thus at most $\epsilon + o(\epsilon)$. □

We now proceed to proving Theorem 3 by applying Lemma 4 twice.

Proof of Theorem 3. Let $\hat{\mathcal{M}}$ be the *Rev* mechanism constructed for the valuation distribution $\hat{D}$. From Lemma 4 we know that

$$Rev(D, \hat{\mathcal{M}}_\epsilon) \geq Rev(\hat{D}) - \epsilon$$

Additionally, let $\mathcal{M}$ be the optimal *Rev* mechanism constructed for the valuation distribution D. Lemma 4 implies that

$$Rev(\hat{D}) \geq Rev(\hat{D}, \mathcal{M}_\epsilon) \geq Rev(D) - \epsilon$$

Combining the two above inequalities we get

$$Rev(D, \hat{\mathcal{M}}_\epsilon) \geq Rev(\hat{D}) - \epsilon \geq Rev(D) - 2\epsilon \geq \frac{1}{8} OPT(D) - 2\epsilon$$

□

5 Acknowledgements

We would like to thank Prof. Daskalakis and Dr. Syrgkanis for teaching 6.853 - Algorithmic Topics in Game Theory and Data Science. This project topic was their idea and it gave us a great opportunity to learn more about learning optimal auctions. We would also like to thank our TAs Noah Golowich and Maxwell Fishelson for providing their guidance and support along the course.

References

[BCD20] J. Brustle, Y. Cai, and C. Daskalakis. *Multi-Item Mechanisms without Item Independence: Learnability via Robustness.* EC'20.

[BDHS15] M.H. Bateni, S. Dehghani, M. T. Hajiaghayi, and S. Seddighin. *Revenue maximization for selling multiple correlated items.* ESA'15.

[BHP16] N. R. Devanur, Z. Huang, and C.-A. Psomas. *The sample complexity of auctions with side information.* STOC'16.

[BILW14] M. Babaioff, N. Immorlica, B. Lucier, and S. M. Weinberg. *A simple and approximately optimal mechanism for an additive buyer.* FOCS'14.

[CDW16] Y. Cai, N. R. Devanur, and S. M. Weinberg. *A duality based unified approach to bayesian mechanism design.* STOC'16.

[DZ20] C. Daskalakis and M. Zampetakis. *More Revenue from Two Samples via Factor Revealing SDPs.* EC'20.

[DRZ20] C. Daskalakis, D. Rohatgi and M. Zampetakis. *Truncated linear regression in high dimensions.* arXiv preprint arXiv:2007.14539, 2020.

[Mye81] R. B. Myerson. *Optimal Auction Design.* Mathematics of Operations Research, 1981.

[Wai09] M. J. Wainwright. *Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1- constrained quadratic programming.* IEEE Transactions on Information Theory, 2009

A Contextual Mechanism Design

A.1 Classification Problem

In this last section, we are trying to generalize the above results to arbitrary classes of auctions. In particular, given the context x of our bidder, we would like to employ the best mechanism among a particular class of mechanisms.

In class, we saw how to learn optimal auctions by using the Empirical Revenue Maximization rule. This framework would essentially choose the best mechanism h from a class of mechanisms $\mathcal{H}$ that obtained the highest empirical revenue. We cannot directly apply the ideas to our setting, since we have a different context for each bidder in the training set.

To take into account the extra information, we will use ideas from the *contextual bandit problem.* Formally, fix a space of mechanisms $\mathcal{H}$. What we are trying to find, is the best mechanism for each bidder context, namely a policy $\pi : [0,1]^p \to \mathcal{H}$. Given a p-dimensional vector $x \in [0,1]^p$, this policy should give the mechanism $h = \pi(x)$ that we should employ if we are selling to a bidder with feature vector x.

We now adopt the Empirical Revenue Maximization framework for our policy example. Define $R(\pi, x_i, v_i)$ to be the revenue of policy π when employed for a bidder with feature vector x_i. This revenue can be computed as we have access to the true valuations of the bidder.

Then, when a new bidder comes, we will choose the policy π^* that achieves the highest empirical revenue.

Algorithm 2 Choosing the Optimal Auction with Context

1: Given mechanisms $\mathcal{H}$ and policies $\pi : [0,1]^d \to \mathcal{H}$
2: Given samples $\{(v_{ij}, x_i)\}_{i\in[q]}$
3: For a new bidder with feature vector x choose policy π^*
4:
$$\pi^* = \arg\max_{\pi} \frac{1}{q} \sum_{i\in[q]} R(\pi, x_i)$$
5: Use mechanism $\pi^*(x)$

B Extending [BDHS15] to Multiple Bidders

This section was part of our effort to extend optimally pricing items from base-value distributions when multiple bidders are interested in them.

Even though this is not entirely relevant to our project, we initially thought that our valuation distribution is base-value and such results can be used to get approximately-optimal mechanisms.

B.1 Semi-Independent Setting

The way [BDHS15] proves that Simple vs Optimal mechanisms exist for base-value distributions was by first extending the [BILW14] result to semi-independent distributions. These are distributions when either the items are fully independent, or they are identical. Their extension was by following the original proof of [BILW14] and tweaking some terms in a summation to make it valid.

What we propose is a similar extension. We instead use the seminal work of [CDW16] that extended Simple vs Optimal mechanisms to many bidders. Then we will follow their proof but modify their summations such that they group identical items together. Although details have to be checked, we believe that such an extension is possible and thus we can prove that if D is a semi-independent valuation distribution then

$$Rev(D) \geq \frac{1}{8} OPT(D)$$

B.2 Remaining Reduction

In this subsection, we assume that we have proved the extension of the multi-bidder Simple vs Optimal mechanism from [CDW16] to the semi-independent setting. In particular, we have proved the following theorem for multiple additive bidders.

Theorem 4. *Let D be a semi-independent distribution of valuations for m items and n additive bidders. Then*

$$\max\{BVCG(D),\ SREV(D)\} \geq \frac{1}{8} REV(D)$$

In this subsection we will prove that if we have n additive bidders whose values for each item are drawn from independent base-value distributions of valuations, we have

$$\max\{BVCG(D),\ SREV(D)\} \geq \frac{1}{16} REV(D)$$

Note that we are assuming that the value of each bidder i for item j is equal to $b_i + e_{ij}$, where the b_i and e_{ij} are independent. This is not quite the case for our problem, since the $\langle \theta, x_i \rangle$ are not necessarily independent, due to the use of the common θ parameter.

Regardless, define a new distribution D' with $2m$ items: $B_1, \dots, B_m$ and $V_1, \dots, V_m$. Each bidder i will have value b_i for all items B and value e_{ij} for item V_j.

It is easy to see that $REV(D) \leq REV(D')$. This is because we can always define a mechanism on D' that bundles item B_i with V_i and then apply the optimal mechanism from D. This will give exactly $REV(D)$.

Note that we are now in the semi-independent setting. This is because items $B_1, \dots, B_m$ are similar, whereas the V items and any B item are independent. From Theorem 2 we conclude that

$$\frac{1}{8} REV(D) \leq \max\{BVCG(D'),\ SREV(D')\}$$

Take the $SREV(D')$ term. Selling item B_i and V_i separately will result in revenue equal to $SREV(B_i) + SREV(V_i)$. However, if we sell these two items as a bundle, we will always get at least half the revenue by selling the one with highest expected revenue and giving the other one 'for free'. Hence by summing over all pairs (B_i, V_i) we get the desired result.

$$SREV(B_i) + SREV(V_i) \leq 2BREV(B_i, V_i) = 2SREV(B_i + S_i)$$

$$\Rightarrow SREV(D') \leq 2SREV(D)$$

Take now the $BVCG(D')$ term. Consider the optimal BVCG mechanism, where each bidder i pays entry fee e_i and enters a VCG auction for each item independently. We know that the VCG

C Known Base Improvements

This section was an effort of ours to improve the 12-Approximation result of [BDHS15] if the base value is known instead of drawn from a distribution. This again turned out not to be very relevant to our project, since knowing the base value (the value of $\langle \theta, x \rangle$) makes the valuations for the items almost independent.

C.1 Better Than 12-Approximation

Looking back to the proof of [BDHS15] of the 12-approximation mechanism, we can see that the extra factor of 2 comes from the following inequality

$$SRev(B_i) + SRev(V_i) \leq 2BRev(B_i + V_i)$$

We provide a short proof of the above statement.

Claim. *When selling two items with independent values, optimally selling them together as a bundle obtains at least half the revenue of selling them separately.*

Proof. Let the two items be A, B. Selling them separately can give maximum revenue by setting prices α, β that maximize $\alpha \cdot p_A$ and $\beta \cdot p_B$. Here p_A is the probability that the bidder has value $\geq \alpha$ and similarly for p_B.

Thus, the total revenue of selling separately is $\alpha p_A + \beta p_B$. Intuitively, we will achieve the above inequality by selling the item with the most revenue and giving away the other item with it, 'for free'. Without loss of generality, assume that $\alpha p_A \geq \beta p_B$. Then one way to sell the bundle is to price both items at α. If the bidder buys A, then they will buy the bundle as well. Hence the probability that a bidder purchases the bundle is $\geq p_A$. As a result, the revenue of the optimal bundling mechanism is

$$BRev(A + B) \geq \alpha p_A \geq \frac{1}{2} \cdot 2\alpha p_A \geq \frac{1}{2}(\alpha p_A + \beta p_B) = \frac{1}{2}(SRev(A) + SRev(B))$$

□

In our case, one of the items has a very simple valuation distribution. Its value is equal to a scalar with certainty. The question here is whether we can obtain a better-than-2 approximation by bundling the items together. Since we cannot hope to improve the 6-approximation of the semi-independent items, this seems to be the only part that we can improve for our constant base-value distributions.

Formal Setup. We have two items to sell, item A with value always equal to $\alpha \geq 0$. Item B has value according to the CDF F. The optimal way of selling the two items separately is to sell A with price α and sell B at price β, the one that maximizes the expression $\beta(1 - F(\beta))$. Hence the maximum revenue achievable by selling separately is equal to

$$OPT := \alpha + \max_{\beta} \beta(1 - F(\beta))$$

When we bundle together, we are effectively selling a new item C whose value distribution has been shifted α upwards. Hence selling at a value of $\alpha + \beta$ will be purchased with probability $1 - F(\beta)$. Hence the optimal revenue of bundling together is

$$REV := \max_{\beta} (\alpha + \beta)(1 - F(\beta))$$

As a result, our goal is to verify whether or not min $\frac{REV}{OPT} \geq \frac{1}{2}$.

Non-Example. Consider the following example. The value of item A is $\alpha = 1$ and the value of B is drawn from the following distribution

$$\beta = \begin{cases} \epsilon & \text{w.p. } 1 - \epsilon \\ \frac{1}{\epsilon} & \text{w.p. } \epsilon \end{cases}$$

Here, you can think of ϵ as a positive constant very close to 0, certainly < 1. Hence $OPT = 1 + 1 = 2$ and

$$REV = \max\left\{(1+\epsilon), \left(1+\frac{1}{\epsilon}\right)\epsilon\right\} = 1+\epsilon$$

As a result, the ratio of selling separately and selling as a bundle is arbitrarily close to $\frac{1}{2}$. We believe that this is because the distribution of β is non-regular. If we restrict the valuations of β to regular distributions, we might be able to get something better than $\frac{1}{2}$. As an example, we offer a proof that if β comes from the exponential distribution, then the ratio is $\frac{1}{1+e}$.

Example. Consider the following example. The value of item A is α and the value of B is drawn from the exponential distribution with parameter λ, i.e. $\beta \le x$ with probability $1 - e^{-\lambda x}$.

$$OPT = \alpha + \max_{\beta} \beta \cdot e^{-\lambda\beta} \overset{\beta=1/\lambda}{=\!=\!=} \alpha + \frac{1}{e\lambda}$$

And for the other term

$$REV = \max_{\beta}(\alpha+\beta)e^{-\lambda\beta} \overset{\alpha+\beta=1/\lambda}{=\!=\!=\!=} \frac{1}{\lambda}e^{\alpha\lambda-1} = \frac{e^{\alpha\lambda}}{e\lambda}$$

The ratio of the two terms is equal to

$$\frac{REV}{OPT} = \frac{e^{\alpha\lambda}}{e\alpha\lambda+1} \ge e^{-1/e} \approx 0.692$$

C.2 Using SDPs

Now the problem boils down to finding the worst distribution of β that will minimize our ratio of REV/OPT. We adopt the framework of [DZ20] where they used a specific relaxation of SDPs to find the worst regular distribution for a specific pricing strategy. We would like to use a similar framework, where our program will have the following form.

$$\begin{aligned}
\alpha_\rho^* = & \min_{F,\, OPT,\, REV,\, p^*,\, p^{**},\, \alpha} \frac{REV}{OPT} \\
\text{s.t. } & F \text{ is regular} \\
& x(1-F(x)) \le OPT \\
& p^*(1-F(p^*)) \ge OPT \\
& (\alpha+x)(1-F(x)) \le REV \\
& (\alpha+p^{**})(1-F(p^{**})) \ge REV
\end{aligned}$$