Learning Multi-item Auctions with (or without) Samples

Yang Cai
School of Computer Science, McGill University
Montréal, Canada
Email: cai@cs.mcgill.ca

Constantinos Daskalakis
EECS and CSAIL, MIT
Cambridge, USA
Email: costis@csail.mit.edu

***Abstract*—We provide algorithms that learn simple auctions whose revenue is approximately optimal in multi-item multi-bidder settings, for a wide range of bidder valuations including unit-demand, additive, constrained additive, XOS, and subadditive. We obtain our learning results in two settings. The first is the commonly studied setting where sample access to the bidders' distributions over valuations is given, for both regular distributions and arbitrary distributions with bounded support. Here, our algorithms require polynomially many samples in the number of items and bidders. The second is a more general max-min learning setting that we introduce, where we are given "approximate distributions," and we seek to compute a mechanism whose revenue is approximately optimal simultaneously for all "true distributions" that are close to the ones we were given. These results are more general in that they imply the sample-based results, and are also applicable in settings where we have no sample access to the underlying distributions but have estimated them indirectly via market research or by observation of bidder behavior in previously run, potentially non-truthful auctions.**

All our results hold for valuation distributions satisfying the standard (and necessary) independence-across-items property. They also generalize and improve upon recent works of Goldner and Karlin [25] and Morgenstern and Roughgarden [32], which have provided algorithms that learn approximately optimal multi-item mechanisms in more restricted settings with additive, subadditive and unit-demand valuations using sample access to distributions. We generalize these results to the complete unit-demand, additive, and XOS setting, to i.i.d. subadditive bidders, and to the max-min setting.

Our results are enabled by new uniform convergence bounds for hypotheses classes under product measures. Our bounds result in exponential savings in sample complexity compared to bounds derived by bounding the VC dimension and are of independent interest.

I. INTRODUCTION

The design of revenue-optimal auctions is a central problem in Economics and Computer Science, which has found myriad applications in online and offline settings, ranging from sponsored search and online advertising to selling artwork by auction houses, and public goods such as drilling rights and radio spectrum by governments. The problem involves a seller who wants to sell one or several items to one or multiple strategic bidders with private valuation functions, mapping each bundle of items they may receive to how much value they derive from the bundle. As no meaningful revenue guarantee can possibly be achieved without any information about the valuations of the bidders, the problem has been classically studied under *Bayesian assumptions,* where a joint distribution from which all bidders' valuations are drawn is common knowledge, and the goal is to maximize revenue in expectation with respect to this distribution.

In the *single-item setting*, Bayesian assumptions have enabled beautiful and influential developments in auction theory. Already 36 years ago, a breakthrough result by Myerson identified the optimal single-item auction when bidder values are independent [33], and the ensuing decades saw a great deal of further understanding and practical applications of single-item auctions, importantly in online settings.

However, the quest for optimal *multi-item auctions* has been quite more challenging. It has been recognized that revenue-optimal multi-item auctions can be really complex, may exhibit counter-intuitive properties, and be fragile to changes in the underlying distributions; for a discussion and examples see survey [18]. As such, it is doubtful that there is a crisp characterization of the structure of optimal multi-item auctions, at least not beyond single-bidder settings [19]. On the other hand, there has been significant recent progress in efficient computation of revenue-optimal auctions [14], [15], [1], [7], [3], [8], [9], [12], [10], [2], [6], [20]. Importantly, this progress has enabled identifying *simple auctions* (mostly variations of sequential posted pricing mechanisms) that achieve constant factor approximations to the revenue of the optimum [5], [37], [11], [16], [13], under the *item-independence* assumption of Definition 1 and Example 1. These auctions are *way simpler* than the optimum, and have *strong incentive properties*: they are dominant strategy truthful, while still competing against the optimal Bayesian truthful mechanism. The current state-of-the-art is given as Theorem 8, which applies to bidders with valuation functions from the broad class of fractionally subbaditive (a.k.a. XOS) valuations, which contains submodular.

As our discussion illustrates, studying auctions assuming Bayesian priors has been quite fruitful, enabling us to identify guiding principles for how to structure auctions to achieve optimal (in single-item settings) or approximately optimal (in multi-item settings) revenue. To apply this theory to practice, however, one needs knowledge of the underlying

0272-5428/17 $31.00 © 2017 IEEE
DOI 10.1109/FOCS.2017.54

distributions. Typically, one would estimate these distributions via market research or by observations of bidder behavior in prior auctions, then use the estimated distributions to design a good auction. However, estimation involves approximation, and the performance of mechanisms can be quite fragile to errors in the distributions. This motivates studying whether optimal or approximately optimal auctions can be identified when one has imperfect knowledge of the true distributions.

With this motivation, recent work in Computer Science has studied whether approximately optimal mechanisms can be "learned" given sample access to the underlying distributions. This work has lead to an almost complete picture for the single-item (and the more general single-parameter) setting where Myerson's theory applies, showing how near-optimal mechanisms can be learned from polynomially many (in the approximation and the number of bidders) samples [24], [17], [30], [28], [31], [21], [35], [26].

On the multi-item front, however, where the analogue of Myerson's theory is elusive, and unlikely, our understanding is much sparser. Recent work of Morgenstern and Roughgarden [32] has taken a computational learning theory approach to identify the sample complexity required to optimize over classes of simple auctions. Combined with the afore-described results on the revenue guarantees of simple auctions, their work leads to algorithms that learn approximately optimal auctions in multi-item settings with multiple unit-demand bidders, or a single subadditive bidder, from polynomially many samples in the number of items and bidders. These results apply to distributions satisfying the *item-independence* assumption of Definition 1 and Example 1, under which the approximate optimality of simple auctions has been established.

While well-suited for identifying the sample complexity required to optimize over a class of simple mechanisms, which is a perfectly reasonable goal to have but not the one in this paper, the approach taken in [32] is arguably imperfect towards proving polynomial sample bounds for learning approximately optimal auctions in the settings where simple mechanisms are known to perform well in the first place. This is due to the following discordance: (i) On the one hand, simple and approximately optimal mechanisms in multi-item settings are mostly only known under item-independence. (ii) On the other hand, the computational learning techniques employed in [32], and in particular bounding the *pseudo-dimension* of a class of auctions, are not fine enough to discern the difference in sample complexity required to optimize under item-independence and without item-independence. As such, this technique can only obtain polynomial sample bounds for approximate revenue optimization if it so happens that a class of mechanisms is both learnable from polynomially-many samples under arbitrary distributions, and it guarantees approximately optimal revenue under item-independence, or for some other interesting class of distributions.[1]

In particular, bounding the pseudo-dimension of classes of auctions as a means to prove polynomial-sample bounds for approximate revenue optimization hits a barrier even for multiple additive bidders with independent values for items. In this setting, the approximately optimal auctions that are known are the best of selling the items separately or running a VCG mechanism with entry fees [37], [11], as described in Section V-B. Unfortunately, the latter can easily be seen to have pseudo-dimension that is exponential in the number of bidders, thus only implying a sufficient exponentially large sample size to optimize over these mechanisms. Is this exponential sample size really necessary or an artifact of the approach? Recent work of Goldner and Karlin [25] gives us hope that it is the latter. They show how to learn approximately optimal auctions in the multi-item multi-bidder setting with additive bidders using only one sample from each bidder's distribution, assuming that it is *regular* and independent across items.

Our results: We show that simple and approximately optimal mechanisms are learnable from polynomially-many samples for multi-item multi-bidder settings, whenever:

- the bidder valuations are fractionally subadditive (XOS), i.e. we can accommodate additive, unit-demand, constrained additive, and submodular valuations;
- the distributions over valuations satisfy the standard item-independence assumption of Definition 1 and Example 1, and their single-item marginals are arbitrary and bounded, or (have arbitrary supports but are) regular.[2]

In particular, our results constitute vast extensions of known results on the polynomial learnability of approximately optimal auctions in multi-item settings [32], [25]. Additionally we show that:

- whenever the valuations are additive and unit-demand, or whenever the bidders are symmetric and have XOS valuations, our approximately optimal mechanisms can be identified from polynomially many samples and in polynomial time;
- whenever the bidders are symmetric (i.e. their valuations are independent and identically distributed) and have *subadditive valuations*, we can compute from polynomially many samples and in polynomial-time a simple mechanism whose revenue is a $\Omega\left(\frac{n}{\max\{m,n\}}\right)$-fraction of the optimum, where m and n are respectively the number of items and bidders. In particular,

[1] It is known that some restriction *needs to be made* on the distribution to gain polynomial sample complexity, as otherwise exponential lower bounds are known for learning approximately optimal auctions even for a single unit-demand bidder [22].

[2] We note again that without the standard item-independence (or some other) restriction on the distributions, we cannot hope to learn approximately optimal auctions from sub-exponentially many samples, even for a single unit-demand bidder [22].

if the number of bidders is at least a constant fraction of the number of items, the mechanism is a constant factor approximation; and

- in the setting of the previous bullet, if the item marginals are regular, our mechanism is *prior-independent*, i.e. there is a single mechanism, identifiable without any samples from the distributions, providing the afore-described revenue guarantee.

Finally, the mechanisms learned by our algorithms for XOS bidders are either *rationed sequential posted price mechanisms* (RSPMs) or *anonymous sequential posted price mechanisms with entry fees* (ASPEs) as defined in Section VI. The mechanisms learned for symmetric subadditive bidders are RSPMs. RSPMs maintain a price p_{ij} for every bidder and item pair and, in some order over bidders $i = 1, \ldots, n$, they give one opportunity to bidder i to purchase *one* item j that has not been purchased yet at price p_{ij}. ASPEs maintain one price p_j for every item and, in some order over bidders $i = 1, \ldots, n$, they give one opportunity to bidder i to purchase *any subset* S' of the items S that have not been purchased yet as long as he also pays an "entry fee" that depends on S and the identity of the bidder. Please see Section VI-A for more details.

Learning without Samples: Thus far, our algorithms used *samples* from the valuation distributions to identify an approximately optimal and simple mechanism under item-independence. However, having sample access to the distributions may be impractical. Often we can observe the actions used by bidders in non-truthful auctions that were previously run, and use these observations to estimate the distributions over valuations using econometric methods [27], [34], [4]. In fact, it may likely be the case we have never sold all the items together in the past, and only have observations of bidder behavior in non-truthful auctions selling each item separately. Econometric methods would achieve better approximations in this case, but only for the item marginals. Finally, we may want to combine multiple sources of information about the distributions, combining past bidder behavior in several different auctions and with market research data.

With this motivation in mind, we would like to extend our learnability results beyond the setting where sample access to the valuation distributions is provided. We propose "learning" approximately optimal multi-item auctions given distributions that are close to the true distributions under some distribution distance $d(\cdot,\cdot)$. In particular, given approximate distributions $\hat{D}_1, \ldots, \hat{D}_n$ over bidder valuations, we are looking to identify a mechanism $\mathcal{M}$ satisfying the following *max-min style objective*:

$$\begin{aligned}&\forall D_1, \ldots, D_n \text{ s.t. } d(D_i, \hat{D}_i) \leq \epsilon, \forall i :\\&\text{Rev}_{\mathcal{M}}(D_1, \ldots, D_n) \geq\\&\qquad\qquad \Omega(\text{OPT}(D_1, \ldots, D_n)) - \text{poly}(\epsilon, m, n). \quad (1)\end{aligned}$$

That is, we want to find a mechanism $\mathcal{M}$ whose revenue is within a constant multiplicative and a $\text{poly}(\epsilon, m, n)$ additive error from optimum, simultaneously in all possible worlds $D_1, \ldots, D_n$, where $d(D_i, \hat{D}_i) \leq \epsilon, \forall i$. It is not a priori clear that such a "one-fits-all" mechanism actually exists.

There are several notions of distance $d(\cdot,\cdot)$ between distributions that we could study in the formulation of Goal (1), but we opt for an easy one to satisfy. We only require that we know every bidder's marginal distributions over single-item values to within ϵ in Kolmogorov distance;[3] see Definition 2. All that this requires is that the cumulative density functions of the approximating distributions over single-item values is within ϵ in infinity norm from the corresponding cumulative density functions of the corresponding true distributions. As such, it is an easy property to satisfy. For example, given sample access to any single-item marginal, the DKW inequality [23] implies that $O(\log(1/\delta)/\epsilon^2)$ samples suffice to learn it to within ϵ in Kolmogorov distance, with probability at least $1 - \delta$. So achieving Goal (1) directly also implies polynomial sample learnability of approximately optimal auctions. But a Kolmogorov approximation can also be arrived at by combining different sources of information about the single-item marginals such as the ones described above. Regardless of how the approximations were obtained, the max-min goal outlined above guarantees robustness of the revenue of the identified mechanism $\mathcal{M}$ with respect to all sources of error that came into the estimation of the single-item marginal distributions.

While Goal (1) is not a priori feasible, we show how to achieve it in multi-item multi-bidder settings with constrained additive bidders, or symmetric bidders with subadditive valuations, under the standard assumption of item-independence. Our results are polynomial-time in the same cases as our sample-based results discussed above.

Roadmap and Technical Ideas: In Section IV, we present a new approach for obtaining uniform convergence bounds for hypotheses classes under product distributions; see Theorem 2 and Corollary 1. We show that our approach can significantly improve the sample complexity bound obtained via traditional methods such as VC theory. In particular, Table III compares the sample complexity bounds obtained via our approach to those obtained by VC theory for different classes of hypotheses.

Our results for mechanisms make use of recent work on the revenue guarantees of simple mechanisms, which are mainly variants of sequential posted pricing mechanisms [11], [13]. Using our results from Section IV, in

[3]Indeed, Goal (1) is achievable only for bounded distributions even in the single-item single-bidder setting. Given any bounded distribution $\hat{D}$, create D by moving ϵ probability mass in $\hat{D}$ to $+\infty$. It is not hard to see that D and $\hat{D}$ are within ϵ in Kolmogorov distance, but no single mechanism can satisfy the approximation guarantee for both D and $\hat{D}$ simultaneously. Using a similar argument, we can argue that the additive error has to depend on H which is the upper bound on any bidder's value for a single item. See Section II for our formal model.

Section V, we derive uniform convergence bounds for *Sequential Posted Price with Entry Fee Mechanisms* (SPEMs), which essentially contain all RSPMs and ASPEs. As a corollary of the uniform convergence of SPEMs, we obtain our sample based results for constrained additive bidders. In fact, we obtain a slightly stronger statement than uniform convergence of the revenue of SPEMs, which also implies our max-min results for constrained-additive bidders; see Theorems 3 and 4. In particular, Theorem 4 and the DKW inequality imply the polynomial-sample learnability of approximately revenue-optimal auctions for constrained additive bidders.

Technically speaking, our sample based and max-min approximation results for constrained additive bidders provide a crisp illustration of how we leverage item-independence and our new uniform convergence bounds for product measures to sidestep the exponential pseudo-dimension of the class of mechanisms that we are optimizing over. Let us discuss our max-min results which are stronger. Suppose $D_i = \times_j D_{ij}$ is the true distribution over bidder i's valuation and $\hat{D}_i = \times_j \hat{D}_{ij}$ is the approximating distribution, where D_{ij} and $\hat{D}_{ij}$ are respectively the item j marginals. To argue that the revenue of some sequential posted price with entry fee mechanism (SPEM) is similar under $D = \times_i D_i$ and $\hat{D} = \times_i \hat{D}_i$, we need to couple in total variation distance the decisions of what sets all bidders buy in the execution of the mechanism under D and $\hat{D}$. The issue that we encounter is that there are exponentially many subsets each bidder may buy, hence the naive use of the Kolmogorov bound $||D_{ij} - \hat{D}_{ij}||_K \leq \epsilon$, on each single-item marginal results in an exponential blow-up in the total variation distance of what subset of items bidder i buys, invalidating our desired coupling. To circumvent this challenge, we argue in Lemma 4 that the events corresponding to which subset of items each buyer will buy are *single-intersecting*, according to Definition 4, when seen as events on the buyer's single-item values. Single-intersecting events may be non-convex and have infinite VC dimension. Nevertheless, because single-item values are independent, our new uniform convergence bounds for product measures (Lemma 3) imply that the difference in probabilities of any such event under D and $\hat{D}$ is only a factor of m, the number of items, larger than the bound ϵ on the Kolmogorov distance between single-item marginals.

We specialize our results to unit-demand bidders in Section V-A to obtain computationally efficient solutions for both max-min and sample-based models. Similarly, Section V-B contains our results for additive bidders. We also generalize our sample-based results for constrained additive bidders to XOS bidders in Section VI. Finally, we provide computationally efficient solutions for symmetric XOS and even symmetric subadditive bidders in Section VII. These results are based on showing that (i) the right parameters of RSPMs and ASPEs can be efficiently and approximately identified with sample or max-min access to the distributions; and (ii) that the revenue guarantees of simple mechanisms can be robustified to accommodate error in the setting of the parameters. In particular, our sample-based result for unit-demand bidders robustifies the ex-ante relaxation of the revenue maximization problem from [1] and its conversion to a sequential posted pricing mechanism from [15], and makes use of the extreme-value theorem for regular distributions from [7]. Our sample-based result for additive bidders shows how to use samples to design mechanisms that approximate the revenue of Yao's VCG with entry fees mechanism [37]. Our sample-based results for XOS bidders show how to use samples to approximate the parameters of the RSPMs and ASPEs of [13], and argue, by re-doing their duality proofs, that their revenue guarantees are robust to errors in the approximation. Finally, our sample-based result for symmetric subadditive bidders is based on a new, duality-based, approximation, showing how to eliminate the use of ASPEs from the result of [13]. This even allows us to obtain prior-independent mechanisms when the item marginals are regular.

II. Preliminaries

We focus on revenue maximization in the combinatorial auction with n independent bidders and m heterogenous items. Each bidder has a valuation that is **subadditive over independent items** (see Definition 1). We denote bidder i's type t_i as $\langle t_{ij} \rangle_{j=1}^m$, where t_{ij} is bidder i's private information about item j. For each i, j, we assume t_{ij} is drawn independently from the distribution D_{ij}. Let $D_i = \times_{j=1}^m D_{ij}$ be the distribution of bidder i's type and $D = \times_{i=1}^n D_i$ be the distribution of the type profile. We use T_{ij} (or T_i, T) and f_{ij} (or f_i, f) to denote the support and density function of D_{ij} (or D_i, D). For notational convenience, we let t_{-i} to be the types of all bidders except i. Similarly, we define D_{-i}, T_{-i} and f_{-i} for the corresponding distributions, support sets and density functions. When bidder i's type is t_i, her valuation for a set of items S is denoted by $v_i(t_i, S)$. Throughout the paper we use OPT to denote the optimal revenue obtainable by any randomized and Bayesian truthful mechanism.

Definition 1. *[36] For every bidder i, whose type t_i is drawn from a product distribution $F_i = \times_j F_{ij}$, her distribution, $\mathcal{V}_i$, over valuation functions $v_i(t_i, \cdot)$ is* ***subadditive over independent items*** *if:*

- $v_i(\cdot,\cdot)$ ***has no externalities***, *i.e., for each $t_i \in T_i$ and $S \subseteq [m]$, $v_i(t_i, S)$ only depends on $\langle t_{ij} \rangle_{j \in S}$, formally, for any $t_i' \in T_i$ such that $t_{ij}' = t_{ij}$ for all $j \in S$, $v_i(t_i', S) = v_i(t_i, S)$.*
- $v_i(\cdot,\cdot)$ ***is monotone***, *i.e., for all $t_i \in T_i$ and $U \subseteq V \subseteq [m]$, $v_i(t_i, U) \leq v_i(t_i, V)$.*
- $v_i(\cdot,\cdot)$ ***is subadditive***, *i.e., for all $t_i \in T_i$ and U, $V \subseteq [m]$,*

$v_i(t_i, U \cup V) \le v_i(t_i, U) + v_i(t_i, V)$.

We use $V_i(t_{ij})$ to denote $v_i(t_i, \{j\})$, as it only depends on t_{ij}. When $v_i(t_i, \cdot)$ is XOS (or constrained additive) for all i and $t_i \in T_i$, we say $\mathcal{V}_i$ is XOS (or constrained additive) over independent items.

Example 1. *[36] We may instantiate Definition 1 to define restricted families of subadditive valuations as follows. In all cases, suppose $t = \{t_j\}_{j\in[m]}$ is drawn from $\times_j D_j$. To define a valuation function that is:*

*- **unit-demand**, we can take t_j to be the value of item j, and set $v(t, S) = \max_{j\in S} t_j$.*

*- **additive**, we can take t_j to be the value of item j, and set $v(t, S) = \sum_{j\in S} t_j$.*

*- **constrained additive**, we can take t_j to be the value of item j, and set $v(t, S) = \max_{R\subseteq S, R\in\mathcal{I}} \sum_{j\in R} t_j$, for some downward closed set system $\mathcal{I} \subseteq 2^{[m]}$.*

*- **XOS (a.k.a. fractionally subadditive)**, we can take $t_j = \{t_j^{(k)}\}_{k\in[K]}$ to encode all possible values associated with item j, and take $v(t, S) = \max_{k\in[K]} \sum_{j\in S} t_j^{(k)}$.*

Note that constrained additive valuations contain additive and unit-demand valuations as special cases, and are contained in XOS valuations.

Distribution Access Models

We consider the following three different models to access the distributions.

- **Sample access to bounded distributions.** We assume that for any buyer i and any type $t_i \in T_i$, her value $V_i(t_{ij})$ for any single item j lies in $[0, H]$.
- **Sample access to regular distributions.** We assume that for any buyer i and any type $t_i \in T_i$, the distribution of her value $V_i(t_{ij})$ for any item j is regular.
- **Direct access to approximate distributions.** We assume that we have direct access to a distribution $\hat{D} = \times_{i\in[n], j\in[m]} \hat{D}_{ij}$, for example we can query the pdf, cdf of $\hat{D}$ and take samples from $\hat{D}$. Moreover, for any buyer i and any type $t_i \in T_i$, the distributions of the random variable $V_i(t_{ij})$ when t_{ij} is sampled from $\hat{D}_{ij}$ or D_{ij} are within ϵ in Kolmogorov distance, and both distributions are supported on $[0, H]$.

Definition 2. *The* Kolmogorov distance *between two distributions P and Q over $\mathbb{R}$, denoted $||P - Q||_K$, is defined as $\sup_{x\in\mathbb{R}} |\Pr_{X\sim P}[X \le x] - \Pr_{X\sim Q}[X \le x]|$. The* total variation distance *between two probability measures P and Q on a sigma-algebra $\mathcal{F}$ of subsets of some sample space Ω, denoted $||P - Q||_{TV}$, is defined as $\sup_{E\in\mathcal{F}} |P(E) - Q(E)|$.*

III. Summary of Our Results

We summarize our results in the following two tables. Table I contains all sample-based results and Table II contains all results under the max-min learning model.

IV. Uniform Convergence under Product Measures

In this section, we develop machinery for obtaining uniform convergence bounds for hypotheses over product measures. Our goal is to save on the sample complexity implied by VC dimension bounds, as summarized in Table III. Indeed, we obtain low sample complexity bounds for indicators over *single-intersecting* sets (see Definition 4), which play a key role in proving our results for learning approximately revenue-optimal auctions. Our main results of this section are Theorem 2 for general functions, and Corollary 1 for sets. Due to the space limit, we postpone all proofs to the full version.

We first define what type of uniform convergence bounds we seek to prove.

Definition 3 ((ϵ, δ)-uniform convergence with respect to proxy measure)**.** *A hypothesis class $\mathcal{H}$ of functions mapping domain set $\mathcal{X}$ to $\mathbb{R}$ has* (ϵ, δ)-uniform convergence with sample complexity $s(\epsilon, \delta)$ *iff, for all $\epsilon, \delta > 0$, there exists a* processing $\mathcal{P} : \mathcal{X}^{s(\epsilon,\delta)} \to \Delta(\mathcal{X})$ *such that for any distribution $\mathcal{D} \in \Delta(\mathcal{X})$ when $k = s(\epsilon, \delta)$:*

$$\Pr_{z_1,\cdots,z_k\sim\mathcal{D}}\left[\sup_{g\in\mathcal{H}}\left|\mathbb{E}_{z\sim\mathcal{P}(z_1,\cdots,z_k)}[g(z)] - \mathbb{E}_{z\sim\mathcal{D}}[g(z)]\right| \le \epsilon\right] \ge 1 - \delta.$$

When $\mathcal{X}$ is the Cartesian product of a collection of sets $\mathcal{X}_1, \ldots, \mathcal{X}_k$, i.e. $\mathcal{X} = \times_i \mathcal{X}_i$, we say that a hypothesis class $\mathcal{H}$ as above has (ϵ, δ)-p.m. uniform convergence with sample complexity $s(\epsilon, \delta)$ *if the above holds for all $\mathcal{D}$ that are product measures over $\mathcal{X}$.*

Next we provide a simple lemma, which leads to a simple version of our main result stated as Theorem 1. Our main result, stated as Theorem 2, follows.

Lemma 1. *Let $\mathcal{X}_1, \ldots, \mathcal{X}_d$ be d domain sets and $\mathcal{H}$ be a hypothesis class with functions mapping from the product space $\times_{i=1}^d \mathcal{X}_i$ to $\mathbb{R}$. For all $i \in [d]$, let $\mathcal{H}_i$ be the projected hypothesis class of $\mathcal{H}$ on $\mathcal{X}_i$, that is, $\mathcal{H}_i = \{g \mid \exists f \in \mathcal{H}\ \exists\ a_{-i} \in \times_{j\neq i}\mathcal{X}_j\ \forall\ x_i \in \mathcal{X}_i,\ g(x_i) = f(x_i, a_{-i})\}$. For every $i \in [d]$, let $\mathcal{D}_i$ and $\hat{\mathcal{D}}_i$ be two distributions supported on $\mathcal{X}_i$. Suppose for all $i \in [d]$,*

$$\sup_{g\in\mathcal{H}_i}\left|\mathbb{E}_{x\sim\mathcal{D}_i}[g(x)] - \mathbb{E}_{x\sim\hat{\mathcal{D}}_i}[g(x)]\right| \le \epsilon,$$

then

$$\sup_{f\in\mathcal{H}}\left|\mathbb{E}_{\boldsymbol{x}\sim\times_{i=1}^d\mathcal{D}_i}[f(\boldsymbol{x})] - \mathbb{E}_{\boldsymbol{x}\sim\times_{i=1}^d\hat{\mathcal{D}}_i}[f(\boldsymbol{x})]\right| \le d\cdot\epsilon.$$

Theorem 1. *Let $\mathcal{X}_1, \ldots, \mathcal{X}_d$ be d domain sets and $\mathcal{H}$ a hypothesis class of functions mapping from the product space $\times_{i=1}^d \mathcal{X}_i$ to $\mathbb{R}$. For all $i \in [d]$, let $\mathcal{H}_i$ be the projected hypothesis class of $\mathcal{H}$ on $\mathcal{X}_i$, that is $\mathcal{H}_i = \{g \mid \exists f \in \mathcal{H}\ \exists\ a_{-i} \in \times_{j\neq i}\mathcal{X}_j\ \forall\ x_i \in \mathcal{X}_i,\ g(x_i) = f(x_i, a_{-i})\}$.*

Valuations	# bidders	Distributions	Approximation	Sample Complexity
additive [25]	n	regular	$\Omega(\text{OPT})$	1
additive	n	arbitrary $[0, H]$	$\Omega(\text{OPT}) - \epsilon \cdot H$	$\text{poly}(n, m, 1/\epsilon)$
unit-demand [32]	n	arbitrary $[0, H]$	$\Omega(\text{OPT}) - \epsilon \cdot H$	$\text{poly}(n, m, 1/\epsilon)$
unit-demand	n	regular	$\Omega(\text{OPT})$	$\text{poly}(n, m)$
constrained additive	n	arbitrary $[0, H]$	$\Omega(\text{OPT}) - \epsilon \cdot H$	$\text{poly}(n, m, 1/\epsilon)$
constrained additive	n	regular	$\Omega(\text{OPT})$	$\text{poly}(n, m)$
XOS	n	arbitrary $[0, H]$	$\Omega(\text{OPT}) - \epsilon \cdot H$	$\text{poly}(n, m, 1/\epsilon)$
XOS	n	regular	$\Omega(\text{OPT})$	$\text{poly}(n, m)$
subadditive [32]	1	arbitrary $[0, H]$	$\Omega(\text{OPT}) - \epsilon \cdot H$	$\text{poly}(m, 1/\epsilon)$
subadditive	n i.i.d.	arbitrary $[0, H]$	$\Omega\left(\frac{n}{\max\{n,m\}}\right) \cdot \text{OPT} - \epsilon \cdot H$	$\text{poly}(n, m, 1/\epsilon)$
subadditive	n i.i.d.	regular	$\Omega\left(\frac{n}{\max\{n,m\}}\right) \cdot \text{OPT}$	*prior-independent*

Table I
SUMMARY OF OUR SAMPLE-BASED RESULTS.

Valuations	# bidders	Distributions	Approximation
additive	n	arbitrary $[0, H]$	$\Omega(\text{OPT}) - O(\epsilon \cdot n \cdot m \cdot H)$
unit-demand	n	arbitrary $[0, H]$	$\Omega(\text{OPT}) - O(\epsilon \cdot n \cdot m \cdot H)$
constrained additive	n	arbitrary $[0, H]$	$\Omega(\text{OPT}) - O(\epsilon \cdot n \cdot m^2 \cdot H)$
subadditive	n i.i.d.	arbitrary $[0, H]$	$\Omega\left(\frac{n}{\max\{n,m\}}\right) \cdot \text{OPT} - O(\epsilon \cdot n \cdot m \cdot H)$

Table II
SUMMARY OF OUR MAX-MIN LEARNING RESULTS.

Suppose that, for all $i \in [d]$, $\mathcal{H}_i$ has (ϵ, δ)-uniform convergence with sample complexity $s_i(\epsilon, \delta)$. Then $\mathcal{H}$ has (ϵ, δ)-p.m. uniform convergence with sample complexity $s(\epsilon, \delta) = \max_{i \in [d]} s_i(\epsilon/d, \delta/d)$.

In particular, let $\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(\ell)}$ be a sample of size $\ell = s(\epsilon, \delta)$ from a product measure $\times_{i \in [d]} \mathcal{D}_i$. Define $\hat{\mathcal{D}}_i = \mathcal{P}_i(z_i^{(1)}, \ldots, z_i^{(\ell)})$, for all $i \in [d]$, where $z_i^{(j)}$ is the i-th entry of sample $\boldsymbol{z}^{(j)}$ and $\mathcal{P}_i$ is the processing corresponding to $\mathcal{H}_i$'s uniform convergence. Then

$$\Pr_{\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(\ell)}} \left[\sup_{f \in \mathcal{H}} \left| \mathbb{E}_{\boldsymbol{z} \sim \times_i \hat{\mathcal{D}}_i} \left[f(\boldsymbol{z}) \right] - \mathbb{E}_{\boldsymbol{z} \sim \times_i \mathcal{D}_i} \left[f(\boldsymbol{z}) \right] \right| \leq \epsilon \right] \geq 1 - \delta.$$

Theorem 2. *Let $\mathcal{X}_1, \ldots, \mathcal{X}_d$ be d domain sets and $\mathcal{H}$ a hypothesis class of functions mapping from the product space $\times_{i=1}^d \mathcal{X}_i$ to $\mathbb{R}$. For all $T \subseteq [d]$, let $\mathcal{H}_T$ be the projected hypothesis class of $\mathcal{H}$ on $\mathcal{X}_T \equiv \times_{i \in T} \mathcal{X}_i$, that is, $\mathcal{H}_T = \{g \mid \exists f \in \mathcal{H}\ \exists\, a_{-T} \in \times_{j \notin T} \mathcal{X}_j\ \forall\, x_T \in \mathcal{X}_T,\ g(x_T) = f(x_T, a_{-T})\}$. Suppose that, for all $T \subseteq [d]$, $\mathcal{H}_T$ has (ϵ, δ)-p.m. uniform convergence with sample complexity $s_T(\epsilon, \delta)$, and define*

$$s(\epsilon, \delta) = \min_{\substack{k,\ \textit{partitions} \\ T_1 \sqcup T_2 \sqcup \ldots \sqcup T_k = [d]}} \max_{i = 1, \ldots, k} s_{T_i}(\epsilon/k, \delta/k). \tag{2}$$

Then $\mathcal{H}$ has (ϵ, δ)-p.m. uniform convergence with sample complexity $s(\epsilon, \delta)$.

In particular, let $\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(\ell)}$ be a sample of size $\ell = s(\epsilon, \delta)$ from a product measure $\times_{i \in [d]} \mathcal{D}_i$. Suppose that the optimum of (2) is attained at $k = \tilde{k}$ for partition $\tilde{T}_1 \sqcup \tilde{T}_2 \sqcup \ldots \sqcup \tilde{T}_{\tilde{k}} = [d]$. Define $\hat{\mathcal{D}}_{\tilde{T}_i} = \mathcal{P}_{\tilde{T}_i}(\boldsymbol{z}_{\tilde{T}_i}^{(1)}, \ldots, \boldsymbol{z}_{\tilde{T}_i}^{(\ell)})$, for all $i \in [d]$, where $\boldsymbol{z}_{\tilde{T}_i}^{(j)}$ contains the entries of sample $\boldsymbol{z}^{(j)}$ in coordinates $\tilde{T}_i$ and $\mathcal{P}_{\tilde{T}_i}$ is the processing corresponding to $\mathcal{H}_{\tilde{T}_i}$'s uniform convergence. Then

$$\Pr_{\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(\ell)}} [\sup_{f \in \mathcal{H}} |\mathbb{E}_{\boldsymbol{z} \sim \times_{i \in [\tilde{k}]} \hat{\mathcal{D}}_{\tilde{T}_i}} [f(\boldsymbol{z})] - \mathbb{E}_{\boldsymbol{z} \sim \times_{i \in [d]} \mathcal{D}_i} [f(\boldsymbol{z})]| \leq \epsilon] \geq 1 - \delta.$$

Nest, we specialize Theorem 2 to indicator functions over sets.

Corollary 1. *We use the same notation as in Theorem 2. Suppose that all functions in $\mathcal{H}$ map $\times_{i=1}^d \mathcal{X}_i$ to $\{0, 1\}$,* *i.e. they are indicators over sets. Suppose also that the VC dimension of $\mathcal{H}_T$ (viewed as a collection of sets) is V_T. Define*

$$V_{\max} = \min_{\substack{k,\ \text{partitions} \\ T_1 \sqcup T_2 \sqcup \ldots \sqcup T_k = [d]}} \left\{ k^2 \cdot \max_{i=1,\ldots,k} V_{T_i} \right\}. \quad (3)$$

Assume that the optimum of (3) *is attained at $k = \tilde{k}$ for partition $\tilde{T}_1 \sqcup \tilde{T}_2 \sqcup \ldots \sqcup \tilde{T}_{\tilde{k}} = [d]$.*

Then $\ell = O\left(\frac{V_{\max}}{\epsilon^2} \cdot \ln \frac{\tilde{k}}{\epsilon} + \frac{\tilde{k}^2}{\epsilon^2} \cdot \ln \frac{\tilde{k}}{\delta}\right)$ samples from $\times_{i \in [d]} \mathcal{D}_i$ suffice to obtain (ϵ, δ)-p.m. uniform convergence for $\mathcal{H}$. Formally,

$$\Pr_{\boldsymbol{z}^{(1)},\ldots,\boldsymbol{z}^{(\ell)}}[\sup_{f \in \mathcal{H}} |\mathbb{E}_{\boldsymbol{z} \sim \times_{i \in [\tilde{k}]} \hat{\mathcal{D}}_{\tilde{T}_i}}[f(\boldsymbol{z})] - \mathbb{E}_{\boldsymbol{z} \sim \times_{i \in [d]} \mathcal{D}_i}[f(\boldsymbol{z})]| \leq \epsilon] \geq 1 - \delta,$$

where for a given sample $\boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(\ell)}$ from a product distribution $\times_{i \in [d]} \mathcal{D}_i$ the distributions $\hat{\mathcal{D}}_{\tilde{T}_i}$ are defined to be uniform over $\boldsymbol{z}^{(1)}_{\tilde{T}_i}, \ldots, \boldsymbol{z}^{(\ell)}_{\tilde{T}_i}$, where $\boldsymbol{z}^{(j)}_{\tilde{T}_i}$ contains the entries of sample $\boldsymbol{z}^{(j)}$ in coordinates $\tilde{T}_i$.

Table III compares the sample complexity for uniform convergence implied by Theorem 2 and Corollary 1 to that implied by VC theory, when the underlying measures are product. Suppose $\mathcal{H}$ contains the indicator functions of all convex sets in $\mathbb{R}^d$. VC theory does not provide any finite sample bound for uniform convergence, as the VC dimension of $\mathcal{H}$ is ∞. Do our results provide a finite bound? Notice that, for all i, $\mathcal{H}_i$ simply contains all intervals in $\mathbb{R}$. Hence, $V_i = 2$ and Corollary 1 implies that $\ell = O(\frac{d^2}{\epsilon^2} \cdot \left(\log \frac{d}{\delta} + \log \frac{d}{\epsilon}\right))$ samples suffice to obtain (ϵ, δ)-p.m. uniform convergence for $\mathcal{H}$. In fact, our sample complexity bound can be improved to $O\left(\frac{d^2}{\epsilon^2} \cdot \log \frac{d}{\delta}\right)$, as $O\left(\frac{\log \frac{1}{\delta}}{\epsilon^2}\right)$ samples suffice to guarantee (ϵ, δ)-uniform convergence for all intervals in $\mathbb{R}$ due to the DKW inequality [23].

In the next a few sections, we apply our uniform convergence results to learn a mechanism with approximately optimal revenue. A type of events called *single-intersecting* (see Definition 4) plays a key role in our analysis. These events are defined based on the geometric shape of the corresponding sets. For example, balls, rectangles and all convex sets are single-intersecting, but this definition includes some non-convex sets as well, for example, "cross-shaped" sets. It turns out that being able to handle these non-convex sets is crucial for our results, as many events we care about are not convex but nonetheless are single-intersecting.

Definition 4 (Single-intersecting Events). *For any event $\mathcal{E}$ in $\mathbb{R}^\ell$, $\mathcal{E}$ is **single-intersecting** if the intersection of $\mathcal{E}$ and any line that is parallel to one of the axes is an interval. More formally, for any $i \in [\ell]$ and any line $L_i = \left\{x \in \mathbb{R}^\ell \mid x_{-i} = a_{-i}\right\}$, where $a_{-i} \in \mathbb{R}^{\ell-1}$, the intersection of L_i and $\mathcal{E}$ is of the form $\left\{x \in \mathbb{R}^\ell \mid x_{-i} = a_{-i}, x_i \in [\underline{a}, \bar{a}]\right\}$ where $\underline{a} \leq \bar{a}$. In particular, we allow $\underline{a}$ to be $-\infty$ and $\bar{a}$ to be $+\infty$.*

We establish a uniform convergence bound for single-intersecting events by combing the DKW inequality and Theorem 1.

Lemma 2. *For any integer ℓ, let $\mathcal{H}$ be the hypothesis class that contains all indicator functions for single-intersecting events in $\mathbb{R}^\ell$. Then $\mathcal{H}$ has (ϵ, δ)-p.m. uniform convergence with sample complexity $O\left(\frac{\ell^2}{\epsilon^2} \cdot \log \frac{\ell}{\delta}\right)$.*

Next, we state a stronger statement, which is a type of uniform convergence bound when access to approximate distributions is given. More specifically, we argue that for any single-intersecting event, the difference in the probability of this event under two product distributions $\mathcal{D} = \times_{i \in [\ell]} \mathcal{D}_i$ and $\hat{\mathcal{D}} = \times_{i \in [\ell]} \hat{\mathcal{D}}_i$ is at most $2\xi \cdot \ell$, if $||\mathcal{D}_i - \hat{\mathcal{D}}_i||_K \leq \xi$ for all i. It is not hard to see that Lemma 3 and the DKW inequality imply Lemma 2.

Lemma 3. *For any integer ℓ, let $\mathcal{D} = \times_{i=1}^{\ell} \mathcal{D}_i$ and $\hat{\mathcal{D}} = \times_{i=1}^{\ell} \hat{\mathcal{D}}_i$, where $\mathcal{D}_i$ and $\hat{\mathcal{D}}_i$ are both supported on $\mathbb{R}$ for any $i \in [\ell]$. If $||\mathcal{D}_i - \hat{\mathcal{D}}_i||_K \leq \xi$, $\left|\Pr_{\mathcal{D}}[\mathcal{E}] - \Pr_{\hat{\mathcal{D}}}[\mathcal{E}]\right| \leq 2\xi \cdot \ell$ for any single-intersecting event $\mathcal{E}$.*

The following table (Table III) summarizes some uniform convergence bounds implied by our results in this section.

V. Constrained Additive Bidders: Uniform Convergence of the Revenue of Sequential Posted Price with Entry Fee Mechanisms

We consider a specific class of mechanisms, namely Sequential Posted Price with Entry fee Mechanisms, a.k.a. **SPEM**s; see Algorithm 1 for details. Cai and Zhao [13] recently showed that if the bidders' valuations are XOS over independent items, the best SPEM achieves a constant fraction of the optimal revenue.[4] This section has two goals. The first is to show that, when bidders have constrained additive valuations over independent items, polynomially many samples suffice to guarantee uniform convergence for the revenue of all SPEMs, and hence our ability to select a near-optimal SPEM from polynomially many samples. This can be proven by applying our uniform convergence result for single-intersecting events (Lemma 2). The second (and stronger goal) is to show that we can learn a near-optimal SPEM under the max-min learning model (Theorem 4). We show that the revenue of any SPEM changes no more than $O(\epsilon \cdot m^2 \cdot n \cdot H)$ under the true and approximate valuation distributions (Theorem 3), where ϵ is an upper bound of the Kolmogorov distance between the true and approximate distributions for every item marginal of every bidder. It is,

[4]Cai and Zhao [13] showed that the best ASPE or RSPM achieves a constant fraction of the optimal revenue. Clearly, any ASPE is also a SPEM, and any RSPM is simply a SPEM if we force the bidders to be unit-demand by only allowing each of them to purchase at most one item.

Hypotheses Class	VC Bound	Bounds from Theorem 2 and Corollary 1
axis-aligned rectangles in $\mathbb{R}^d$	$\tilde{O}(d/\epsilon^2)$	$\tilde{O}(d/\epsilon^2)$
polytopes with k facets in $\mathbb{R}^d$	$\tilde{O}(dk/\epsilon^2)$	$\tilde{O}(d \cdot \min\{d, k\}/\epsilon^2)$
arbitrary convex sets in $\mathbb{R}^d$	∞	$\tilde{O}(d^2/\epsilon^2)$
single-intersecting sets in $\mathbb{R}^d$	∞	$\tilde{O}(d^2/\epsilon^2)$

Table III
NUMBER OF SAMPLES REQUIRED FOR $(\epsilon, \Theta(1))$-P.M. UNIFORM CONVERGENCE FOR DIFFERENT $\mathcal{H}$'S.

of course, not hard to see that Theorem 3 and the DKW inequality imply uniform convergence of the revenue of all SPEMS. To establish Theorem 3, we need to apply Lemma 3 instead of Lemma 2. All proofs can be found in the full version.

Algorithm 1 Sequential Posted Price with Entry Fee Mechanism (SPEM)

Require: A collection of prices $\{p_{ij}\}_{i\in[n],j\in[m]}$ and a collection of entry fee functions $\{\delta_i(\cdot)\}_{i\in[n]}$ where $\delta_i : 2^{[m]} \mapsto \mathbb{R}$ is bidder i's entry fee function.

$S \leftarrow [m]$
for $i \in [n]$ **do**
  Show bidder i the set of available items S and set the entry fee for bidder i to be $\delta_i(S)$.
  if Bidder i pays the entry fee $\delta_i(S)$ **then**
    i receives her favorite bundle S_i^* and pays $\sum_{j\in S_i^*} p_{ij}$.
    $S \leftarrow S \backslash S_i^*$.
  else
    i gets nothing and pays 0.
  end if
end for

We first establish a technical lemma, which states that, for any set of items S, any set of prices $\{p_j\}_{j\in[m]}$ and entry fee δ, the distribution over the *set of items* purchased by a constrained additive bidder whose valuation is drawn from $\mathcal{D} = \times_{j\in[m]}\mathcal{D}_j$ and $\hat{\mathcal{D}} = \times_{j\in[m]}\hat{\mathcal{D}}_j$ has total variation distance at most $2m\xi$, if $||\mathcal{D}_j - \hat{\mathcal{D}}_j||_K \le \xi$ for every item $j \in [m]$. This is quite surprising. Given that, for each set of items $S' \subseteq S$, the difference in the probability that the buyer will purchase this particular set S' under $\mathcal{D}$ and $\hat{\mathcal{D}}$ could already be as large as $\Theta(m\xi)$, and the distribution has an exponentially large support size, a trivial argument would give a bound of $2^m \cdot \Theta(m\xi)$. To overcome this analytical difficulty, we argue instead that for any collection of sets of items, the event that the buyer's favorite set lies in this collection is single-intersecting. Then our result follows from Lemma 3. Notice that it is crucial that Lemma 3 holds for all events that are single-intersecting, as the event we consider here is clearly non-convex in general.

Lemma 4. *For any set $S \subseteq [m]$, any prices $\{p_j\}_{j\in[m]}$ and entry fee $\delta(S)$, let $\mathcal{L}$ and $\hat{\mathcal{L}}$ be the distributions over the set of items purchased from S by a constrained additive bidder under prices $\{p_j\}_{j\in[m]}$ and entry fee δ when her type is drawn from $\mathcal{D} = \times_{j\in[m]}\mathcal{D}_j$ and $\hat{\mathcal{D}} = \times_{j\in[m]}\hat{\mathcal{D}}_j$ respectively. If $||\mathcal{D}_j - \hat{\mathcal{D}}_j||_K \le \xi$ for all item j, $||\mathcal{L} - \hat{\mathcal{L}}||_{TV} \le 2m\xi$.*

Theorem 3. *Suppose all bidders' valuations are constrained additive over independent items. For any SPEM, let* $\textsc{Rev}$ *and* $\widehat{\textsc{Rev}}$ *be its expected revenue under D and $\hat{D}$ respectively. If D_{ij} and $\hat{D}_{ij}$ are both supported on $[0, H]$, and $||D_{ij} - \hat{D}_{ij}||_K \le \xi$ for all $i \in [n]$ and $j \in [m]$,*

$$\left|\textsc{Rev} - \widehat{\textsc{Rev}}\right| \le 2nm\xi \cdot (mH + \mathit{OPT}).$$

With Theorem 3, we are ready to show the learnability of the approximately revenue-optimal mechanisms in the max-min learning model.

Theorem 4. *(Max-min Learning for Constrained Additive Bidders) When all bidders' valuations are constrained additive over independent items and for any bidder i and any item j, D_{ij} and $\hat{D}_{ij}$ are supported on $[0, H]$ and $||D_{ij} - \hat{D}_{ij}||_K \le \epsilon$ for some $\epsilon = O(\frac{1}{nm})$, then with only access to $\hat{D} = \times_{i,j}\hat{D}_{ij}$, our algorithm can learn an RSPM or ASPE whose revenue is at least $\frac{OPT}{c} - \epsilon \cdot O(m^2nH)$, where OPT is the optimal revenue by any BIC mechanism under $D = \times_{i,j}D_{ij}$. $c > 1$ is an absolute constant.*

Clearly, Theorem 4 also implies a polynomial sample complexity bound for learning an approximately revenue-optimal mechanism. A better sample complexity bound can be obtained directly, i.e. without invoking the uniform convergence of the revenue of SPEMs, and is stated as Theorem 9 for the broader class of XOS valuations. Similarly, when bidders have simpler valuations, i.e., additive or unit-demand valuations, we can sharpen our results and achieve polynomial-time learnability of the approximately optimal mechanism using more specialized techniques. See Sections V-A and V-B for details.

A. Unit-demand Valuations: Polynomial-Time Learning

In this section, we consider bidders with unit-demand valuations, sharpening our results to show how to learn approximately revenue-optimal mechanisms in polynomial time. It is shown in a sequence of works [15], [29], [11] that there exists a *sequential posted price mechanism* (SPM), a special case of SPEM with 0 entry fee for every bidder, that achieves at least $\frac{1}{24}$ of the optimal revenue when bidders are unit-demand. We show that under all three distribution access models of Section II there exists a polynomial-time algorithm that learns a sequential posted price mechanism whose revenue approximates the optimal revenue. We only sketch the proof here and postpone the details to the full version.

Theorem 5. *When all bidders have unit-demand valuations and*

- *D_{ij} is supported on $[0, H]$ for all bidder i and item j, there exists a polynomial time algorithm that learns an SPM whose revenue is at least $\frac{OPT}{144} - \epsilon H$ with probability $1-\delta$ given $O\left(\left(\frac{1}{\epsilon}\right)^2\left(m^2 n \log\frac{n}{\epsilon} + \log\frac{1}{\delta}\right)\right)$ samples from D; or*
- *D_{ij} is a regular distribution for all bidder i and item j, there exists a polynomial time algorithm that learns a randomized SPM whose revenue is at least $\frac{OPT}{33}$ with probability $1-\delta$ given $O(\max\{m,n\}^2 m^2 n^2 \cdot \log\frac{nm}{\delta})$ samples from D; or*
- *we are only given access to $\hat{D}_{ij}$ where $||\hat{D}_{ij} - D_{ij}||_K \leq \epsilon$ for all bidder i and item j, there is a polynomial time algorithm that constructs a randomized SPM whose revenue under D is at least $\left(\frac{1}{4} - (n+m)\cdot\epsilon\right)\cdot\left(\frac{OPT}{8} - 2\epsilon\cdot mnH\right)$[5].*

Sample Access to Bounded Distributions: the result is due to Morgenstern and Roughgarden [32].

Direct Access to Approximate Distributions: we first consider a convex program based on D which is usually referred to as the ex-ante relaxation of the revenue maximization problem [1], and use its optimum as a proxy for OPT. Next, we consider a similar convex program based on $\hat{D}$ and show that the optima of the two convex programs are close to each other. Finally, we use techniques developed by Chawla et al. [15] to convert the optimal solution of the second convex program into a randomized SPM. We can show that the constructed randomized SPM achieves a revenue that approximates the optimum of the second convex program under D, which implies that the mechanism's revenue also approximates the OPT. As we are given $\hat{D}$, we can solve the second convex program and convert its optimal solution into a randomized SPM in polynomial time. Please see the full version of the paper for more details.

Sample Access to Regular Distributions: we use a similar convex program relaxation based approach as in the previous case. The main difference is that regular distributions could be unbounded and thus ruin the approximation guarantee. We show how to use the Extreme Value theorem in [7] to truncate the distributions without hurting the revenue by much. Please see the full version for further details.

B. Additive Valuations: Polynomial-Time Learning

In this section, we consider bidders with additive valuations, again sharpening our results to show polynomial-time learnability. It is known that the better of the following two mechanisms achieves at least $\frac{1}{8}$ of the optimal revenue when all bidders have additive valuations [37], [11]:

Selling Separately: the mechanism sells each item separately using Myerson's optimal auction.

VCG with Entry Fee: the mechanism solicits bids $\mathbf{b} = (b_1, \cdots, b_n)$ from the bidders, then offers each bidder i the option to participate for an entry fee $e_i(b_{-i}, D_i)$, which is the median of the random variable $\sum_{j\in[m]}(t_{ij} - \max_{k\neq i} b_{kj})^+$, where $t_i \sim D_i$[6]. This random variable is exactly bidder i's utility when her type is t_i and the other bidders' are b_{-i}. If bidder i chooses to participate, she pays the entry fee and can take any item j at price $\max_{k\neq i} b_{kj}$. Notice that the mechanism never over allocate any item, as only the highest bidder for an item can afford it.

Indeed, only counting the revenue from the entry fee in the second mechanism and the optimal revenue from selling the items separately already suffices to provide an 8-approximation [37], [11].

Theorem 6 ([11])**.** *Let* SREV *be the optimal revenue for selling the items separately and* BREV *be the expected entry fee collected from the VCG with entry fee mechanism. Then* $OPT \leq 6 \cdot \text{SREV} + 2 \cdot \text{BREV}$.

Goldner and Karlin [25] showed that one sample suffices to learn a mechanism that achieves a constant fraction of the optimal revenue when D_{ij} is regular for all $i \in [n]$ and $j \in [m]$. We show how to learn an approximately optimal mechanism in the other two models.

Theorem 7. *When the bidders have additive valuations and*

- *D_{ij} is supported on $[0, H]$ for all bidder i and item j, we can learn in polynomial time a mechanism whose expected revenue is at least $\frac{OPT}{32} - \epsilon\cdot H$ with probability $1-\delta$ given $O\left(\left(\frac{m}{\epsilon}\right)^2\cdot\left(n\log n\log\frac{1}{\epsilon} + \log\frac{1}{\delta}\right)\right)$ samples from D; or*
- *we are only given access to distributions $\hat{D}_{ij}$ where $||\hat{D}_{ij} - D_{ij}||_K \leq \epsilon$ for all bidder i and item j, there is a polynomial time algorithm that constructs a mechanism whose expected revenue under D is at least $\frac{OPT}{266} - 96\epsilon\cdot mnH$ when $\epsilon \leq \frac{1}{16\max\{m,n\}}$.*

[5] Setting ϵ to be $O(\frac{1}{m+n})$ gives the desired max-min guarantee.

[6] The entry fee function defined in [37], [11] is slightly different. They showed that there exists an entry fee X_i, such that bidder i accepts the entry fee with probability at least $1/2$. Then they argued that extracting $X_i/2$ as the revenue in the VCG with entry fee mechanism is enough to obtain a factor 8 approximation. It is not hard to observe that our entry fee is accepted with probability exactly $1/2$, thus our entry fee is at least as large as X_i. So our mechanism also suffices to provide a factor 8 approximation.

Sample Access to Bounded Distributions: Goldner and Karlin's proof [25] can be directly applied to the bounded distributions to show a single sample suffices to learn a mechanism whose expected revenue approximates the BREV. Then as SREV is the revenue of m separate single-item auctions, we can use the result in [32] to approximate it. Please see the full version for further details.

Direct Access to Approximate Distributions: for each single item, we apply Theorem 5 to learn an individual auction, then run these learned auctions in parallel. Clearly, the combined auction's revenue approximates SREV. For BREV, we show that for every bidder i and every bid profile b_{-i} of the other bidders, the event that corresponds to bidder i accepting any entry fee is *single-intersecting* (see Definition 4). This implies that the probability for a bidder to accept an entry fee under $\hat{D}$ and D is close (Lemma 3). So we can essentially use the median of $\sum_{j\in[m]}(t_{ij} - \max_{k\neq i} b_{kj})^+$ with $t_i \sim \hat{D}_i$ as the entry fee. We postpone further details to the full version.

VI. XOS VALUATIONS

In this section we go beyond constained additive valuations to show learnability of approximately revenue-optimal auctions from polynomially many samples. The better of the following two mechanisms is known to achieve a constant fraction of the optimal revenue, when bidders have valuations that are XOS over independent items [13].

Rationed Sequential Posted Price Mechanism (RSPM): the mechanism is similar to SPM, except there is an extra constraint that every bidder can purchase at most one item.

Anonymous Sequential Posted Price with Entry Fee Mechanism (ASPE): every buyer faces the same collection of item prices $\{p_j\}_{j\in[m]}$. The seller visits the bidders sequentially. For every bidder, the seller shows her all the available items (i.e. items that have not yet been purchased) and the associated price for each item, then asks her to pay a personalized entry fee which depends on her type distribution and the set of available items. If the bidder accepts the entry fee, she can proceed to purchase any available item at the given price; if she rejects the entry fee, she neither receives nor pays anything.

Theorem 8. *[13] There exists a collection of prices $\{p_j^*\}_{j\in[m]}$, such that if we set the entry fee function $\delta_i^*(S)$ to be the median of bidder i's utility for set S, either the ASPE(p^*, δ^*) or the best RSPM achieves at least a constant fraction of the optimal revenue when bidders' valuations are XOS over independent items. More formally, let $u_i^*(t_i, S) = \max_{S^*\subseteq S} v_i(t_i, S^*) - \sum_{j\in S^*} p_j^*$ be bidder i's utility for the set of items S when her type is t_i. We define $\delta_i^*(S)$ to be the median of the random variable $u_i^*(t_i, S)$ (with $t_i \sim D_i$) for any set $S \subseteq [m]$. Moreover, the price p_j^* for any item j is no larger than $2G$, where $G = \max_{i,j} G_{ij}$ and $G_{ij} := \sup_x \left\{\Pr_{t_{ij}\sim D_{ij}}\left[V_i(t_{ij}) \geq x\right] \geq \frac{1}{5\max\{m,n\}}\right\}$.*

Our goal next is to bound the sample complexity for learning a near-optimal RSPM and the ASPE described in Theorem 8 under XOS valuations.

We consider first the task of learning a near-optimal RSPM. In a RSPM, all bidders are restricted to be unit-demand, so the revenue of the best RSPM is upper bounded by the optimal revenue in the corresponding unit-demand setting. In Section V-A, we have shown how to learn an approximately optimal mechanism for unit-demand bidders, and those algorithms can be used to approximate the best RSPM.

So, for the rest of this section, it suffices to focus on learning an ASPE whose revenue approximates the revenue of the ASPE described in Theorem 8. We will do this in Section VI-A. Before that, we need a robust version of Theorem 8. In the next Lemma, we argue that if we use a collection of prices $\{p'_j\}_{j\in[m]}$ sufficiently close to $\{p_j^*\}_{j\in[m]}$ and entry fee $\delta'_i(S)$ sufficiently close to the median of the utility for every bidder i and subset S, the better of the corresponding ASPE and the best RSPM still approximates the optimal revenue. See the full version for more details.

Lemma 5. *For any $\epsilon > 0$ and $\mu \in [0, \frac{1}{4}]$, let $\{p'_j\}_{j\in[m]}$ be a collection of prices such that $|p'_j - p_j^*| \leq \epsilon$ for all $j \in [m]$, where $\{p_j^*\}_{j\in[m]}$ is the collection of prices in Theorem 8. Let $\delta'_i(S)$ be bidder i's entry fee function such that $\Pr_{t_i\sim D_i}\left[u'_i(t_i, S) \geq \delta'_i(S)\right] \in [1/2-\mu, 1/2+\mu]$ for any set $S \subseteq [m]$, where $u'_i(t_i, S) = \max_{S*\subseteq S} v_i(t_i, S^*) - \sum_{j\in S^*} p'_j$. Then, either the ASPE$(p', \delta')$ or the best RSPM achieves revenue at least $\frac{OPT}{\mathcal{C}_1(\mu)} - \mathcal{C}_2(\mu)\cdot(m+n)\cdot\epsilon$ when bidders' valuations are XOS over independent items. Both $\mathcal{C}_1(\cdot)$ and $\mathcal{C}_2(\cdot)$ are monotonically increasing functions that only depend on μ.*

Definition 5. *We say a collection of prices $\{p_j\}_{j\in[m]}$ is in the B-bounded ϵ-net if p_j is a multiple of ϵ and no larger than B for any item j. For any collection of prices $\{p_j\}_{j\in[m]}$, we say the entry fee functions are μ-balanced if for every bidder i and every set $S \subseteq [m]$, her entry fee $\delta_i(S)$ satisfies $\Pr_{t_i\sim D_i}[u_i(t_i, S) \geq \delta_i(S)] \in [1/2-\mu, 1/2+\mu]$, where $u_i(t_i, S) = \max_{S*\subseteq S} v_i(t_i, S^*) - \sum_{j\in S^*} p_j$.*

Corollary 2. *For bidders with valuations that are XOS over independent items and any $\epsilon > 0$, there exists a collection of prices $\{p_j\}_{j\in[m]}$ in the $2G$-bounded ϵ-net such that for any μ-balanced entry fee functions $\{\delta_i(\cdot)\}_{i\in[n]}$ with $\mu \in [0, \frac{1}{4}]$, either the ASPE(p, δ) or the best RSPM achieves revenue at least $\frac{OPT}{\mathcal{C}_1(\mu)} - \mathcal{C}_2(\mu)\cdot(m+n)\cdot\epsilon$.*

A. XOS Valuations: sample access to bounded and regular distributions

In this section, we consider how to learn an ASPE with high revenue given sample access to D. Our learning algorithm is a two-step procedure. In the first step, we take a few samples from D and use these samples to set the entry fee for every collection of prices $\{p_j\}_{j\in[m]}$ in the ϵ-net. More specifically, to decide $\delta_i(S)$ we compute the utility of bidder i for set S under $\{p_j\}_{j\in[m]}$ over all the samples and take the empirical median among all these utilities to be $\delta_i(S)$. With a polynomial number of samples, we can guarantee that for any $\{p_j\}_{j\in[m]}$ in the ϵ-net the computed entry fee functions $\{\delta_i(\cdot)\}_{i\in[n]}$ are μ-balanced. Now, we have created an ASPE for every $\{p_j\}_{j\in[m]}$ in the ϵ-net. In the second step, we take some fresh samples from D and use them to estimate the revenue for each of the ASPEs we created in the first step, then pick the one that has the highest empirical revenue. It is not hard to argue that with a polynomial number of samples the mechanism we pick has high revenue with probability almost 1. Combining our algorithm with Theorem 5, we obtain the following theorem.

Theorem 9. *When all bidders' valuations are XOS over independent items and*

- *the random variable $V_i(t_{ij})$ is supported on $[0, H]$ for each bidder i and item j, we can learn an RSPM and an ASPE such that with probability at least $1-\delta$ the better of the two mechanisms has revenue at least $\frac{OPT}{c_1} - \xi \cdot H$ for some absolute constant $c_1 > 1$ given $O\left((\frac{mn}{\xi})^2 \cdot (m \cdot \log \frac{m+n}{\xi} + \log \frac{1}{\delta})\right)$ samples from D;*
- *the random variable $V_i(t_{ij})$ is regular for each bidder i and item j, we can learn an RSPM and an ASPE such that with probability at least $1-\delta$ the better of the two mechanisms has revenue at least $\frac{OPT}{c_2}$ for some absolute constant $c_2 > 1$ given $O\left(\max\{m,n\}^2 m^2 n^2 \left(m \log(m+n) + \log \frac{1}{\delta}\right)\right)$ samples from D.*

We postpone the proofs to the full version.

VII. Symmetric Bidders

In this section, we consider symmetric bidders ($D_i = D_{i'}$ for all i and $i' \in [n]$) with XOS and subadditive valuations. For XOS valuations, our goal is to improve our algorithms from Section VI to be computationally efficient under bidder symmetry. For subadditive valuations, our goal is to establish the learnability of approximately optimal mechanisms whose revenue improves as the number of bidders becomes comparable to the number of items. We only describe the results here and postpone the formal statements and proofs to the full version.

- **XOS valuations:** we can learn in polynomial time an approximately optimal mechanism with a polynomial number of samples when the valuations are XOS over independent items. Our algorithm essentially estimates all the parameters needed to run the RSPM and ASPE used in [13]. In general, it is not clear how to estimate these parameters efficiently. But when the bidders are symmetric, one only needs to consider "symmetric parameters" which greatly simplifies the search space and allows us to estimate all the parameters in polynomial time.
- **subadditive valuations:** when the valuations are subadditive over independent items, the optimal revenue is at most $O\left(\frac{n}{\max\{m,n\}}\right)$ times larger than the highest revenue obtainable by an RSPM. In other words, if the number of items is within a constant times the number of bidders, an RSPM suffices to extract a constant fraction of the optimal revenue. Applying our results for unit-demand bidders in Section V-A, we can learn a nearly-optimal RSPM, which is also a good approximation to OPT. In fact, when the distribution for random variable $V_i(t_{ij})$ is regular for every bidder i and item j, we can design a prior-independent mechanism that achieves a constant fraction of the optimal revenue.

References

[1] Saeed Alaei. Bayesian Combinatorial Auctions: Expanding Single Buyer Mechanisms to Many Buyers. In *the 52nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2011.

[2] Saeed Alaei, Hu Fu, Nima Haghpanah, and Jason Hartline. The Simple Economics of Approximately Optimal Auctions. In *the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2013.

[3] Saeed Alaei, Hu Fu, Nima Haghpanah, Jason Hartline, and Azarakhsh Malekian. Bayesian Optimal Auctions via Multi- to Single-agent Reduction. In *the 13th ACM Conference on Electronic Commerce (EC)*, 2012.

[4] Susan Athey and Philip Haile. Nonparametric approaches to auctions. *Handbook of econometrics*, 6(Part A):3847–3965, 2007.

[5] Moshe Babaioff, Nicole Immorlica, Brendan Lucier, and S. Matthew Weinberg. A Simple and Approximately Optimal Mechanism for an Additive Buyer. In *the 55th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2014.

[6] Anand Bhalgat, Sreenivas Gollapudi, and Kamesh Munagala. Optimal auctions via the multiplicative weight method. In *ACM Conference on Electronic Commerce, EC '13, Philadelphia, PA, USA, June 16-20, 2013*, pages 73–90, 2013.

[7] Yang Cai and Constantinos Daskalakis. Extreme-Value Theorems for Optimal Multidimensional Pricing. In *the 52nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2011.

[8] Yang Cai, Constantinos Daskalakis, and S. Matthew Weinberg. An Algorithmic Characterization of Multi-Dimensional Mechanisms. In *the 44th Annual ACM Symposium on Theory of Computing (STOC)*, 2012.

[9] Yang Cai, Constantinos Daskalakis, and S. Matthew Weinberg. Optimal Multi-Dimensional Mechanism Design: Reducing Revenue to Welfare Maximization. In *the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2012.

[10] Yang Cai, Constantinos Daskalakis, and S. Matthew Weinberg. Understanding Incentives: Mechanism Design becomes Algorithm Design. In *the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2013.

[11] Yang Cai, Nikhil R. Devanur, and S. Matthew Weinberg. A duality based unified approach to bayesian mechanism design. In *the 48th Annual ACM Symposium on Theory of Computing (STOC)*, 2016.

[12] Yang Cai and Zhiyi Huang. Simple and Nearly Optimal Multi-Item Auctions. In *the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2013.

[13] Yang Cai and Mingfei Zhao. Simple mechanisms for subadditive buyers via duality. In *the 49th Annual ACM Symposium on Theory of Computing (STOC)*, 2017.

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

[15] Shuchi Chawla, Jason D. Hartline, David L. Malec, and Balasubramanian Sivan. Multi-Parameter Mechanism Design and Sequential Posted Pricing. In *the 42nd ACM Symposium on Theory of Computing (STOC)*, 2010.

[16] Shuchi Chawla and J. Benjamin Miller. Mechanism design for subadditive agents via an ex-ante relaxation. In *Proceedings of the 2016 ACM Conference on Economics and Computation, EC '16, Maastricht, The Netherlands, July 24-28, 2016*, pages 579–596, 2016.

[17] Richard Cole and Tim Roughgarden. The sample complexity of revenue maximization. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, 2014.

[18] Constantinos Daskalakis. Multi-item auctions defying intuition? *ACM SIGecom Exchanges*, 14(1):41–75, 2015.

[19] Constantinos Daskalakis, Alan Deckelbaum, and Christos Tzamos. Strong duality for a multiple-good monopolist. *Econometrica*, 2017.

[20] Constantinos Daskalakis, Nikhil R. Devanur, and S. Matthew Weinberg. Revenue maximization and ex-post budget constraints. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation, EC '15, Portland, OR, USA, June 15-19, 2015*, pages 433–447, 2015.

[21] NR Devanur, Z Huang, and CA Psomas. The sample complexity of auctions with side information. In *48th Annual ACM SIGACT Symposium on Theory of Computing*, 2016.

[22] Shaddin Dughmi, Li Han, and Noam Nisan. Sampling and representation complexity of revenue maximization. In *International Conference on Web and Internet Economics (WINE)*, 2014.

[23] Aryeh Dvoretzky, Jack Kiefer, and Jacob Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, pages 642–669, 1956.

[24] Edith Elkind. Designing and learning optimal finite support auctions. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 736–745. Society for Industrial and Applied Mathematics, 2007.

[25] Kira Goldner and Anna R. Karlin. A prior-independent revenue-maximizing auction for multiple additive bidders. In *Web and Internet Economics - 12th International Conference, WINE 2016, Montreal, Canada, December 11-14, 2016, Proceedings*, pages 160–173, 2016.

[26] Yannai A. Gonczarowski and Noam Nisan. Efficient empirical revenue maximization in single-parameter auction environments. *CoRR*, abs/1610.09976, 2016.

[27] Emmanuel Guerre, Isabelle Perrigne, and Quang Vuong. Optimal nonparametric estimation of first-price auctions. *Econometrica*, 68(3):525–574, 2000.

[28] Zhiyi Huang, Yishay Mansour, and Tim Roughgarden. Making the most of your samples. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, 2015.

[29] Robert Kleinberg and S. Matthew Weinberg. Matroid Prophet Inequalities. In *the 44th Annual ACM Symposium on Theory of Computing (STOC)*, 2012.

[30] Mehryar Mohri and Andres Munoz Medina. Learning theory and algorithms for revenue optimization in second price auctions with reserve. In *ICML*, pages 262–270, 2014.

[31] Jamie Morgenstern and Tim Roughgarden. The pseudo-dimension of near-optimal auctions. *arXiv preprint arXiv:1506.03684*, 2015.

[32] Jamie Morgenstern and Tim Roughgarden. Learning simple auctions. In *Proceedings of the 29th Conference on Learning Theory, COLT 2016, New York, USA, June 23-26, 2016*, pages 1298–1318, 2016.

[33] Roger B. Myerson. Optimal Auction Design. *Mathematics of Operations Research*, 6(1):58–73, 1981.

[34] Harry J Paarsch and Han Hong. An introduction to the structural econometrics of auction data. *MIT Press Books*, 1, 2006.

[35] Tim Roughgarden and Okke Schrijvers. Ironing in the dark. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pages 1–18. ACM, 2016.

[36] Aviad Rubinstein and S. Matthew Weinberg. Simple mechanisms for a subadditive buyer and applications to revenue monotonicity. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation, EC '15, Portland, OR, USA, June 15-19, 2015*, pages 377–394, 2015.

[37] Andrew Chi-Chih Yao. An n-to-1 bidder reduction for multi-item auctions and its applications. In *SODA*, 2015.