

Learning Multi-item Auctions with (or without) Samples

Yang Cai
McGill University, Canada
cai@cs.mcgill.ca

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

September 4, 2017

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 [28] and Morgenstern and Roughgarden [35], 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.

1 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 [36], 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, 42, 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 subadditive (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 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 led 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 [26, 17, 33, 31, 34, 22, 38, 29].

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 [35] 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 [35] 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 [35], 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 [42, 11], as described in Section 5.2. 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 [28] 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 [35, 28]. 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, if the number of bidders is at least a constant fraction of the number of items, the mechanism is a constant factor approximation; and

¹It 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 [24].

²We 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 [24].

- 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 6. 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 $i = 1, \dots, n$, 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 $i = 1, \dots, n$, 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. See Algorithm 3.

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 [30, 37, 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 $\hat{D}_1, \dots, \hat{D}_n$ over bidder valuations, we are looking to identify a mechanism \mathcal{M} satisfying the following *max-min style objective*:

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

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, \dots, 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 [25] 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

³Indeed, 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 2 for our formal model.

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 4, 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 3 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 4, in Section 5, we derive uniform convergence bounds for the revenue of a class of mechanisms shown to achieve a constant fraction of optimal revenue when all bidders have valuations that are constrained additive over independent items. These mechanisms are called Sequential Posted Price with Entry Fee Mechanisms, a.k.a. SPEMs,⁴. 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 anonymous sequential posted price with entry fees (ASPE) mechanism 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 5.1 to obtain computationally efficient solutions for both max-min and sample-based models. Similarly, Section 5.2 contains our results for additive bidders. We also generalize our sample-based results for constrained additive bidders to XOS bidders in Section 6. Finally, we provide computationally efficient solutions for symmetric XOS and even symmetric sub-additive bidders in Section 7. These results are based on showing that (i) the right parameters of SPEMs 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

⁴Note that any RSPM or ASPE is an SPEM.

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 [42]. Our sample-based results for XOS bidders show how to use samples to approximate the parameters of the RSPM and ASPE mechanisms 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.

2 Preliminaries

We focus on revenue maximization in the combinatorial auction with n independent bidders and m heterogeneous 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. [40] 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 $U, V \subseteq [m]$, $v_i(t_i, U \cup V) \leq v_i(t_i, U) + v_i(t_i, V)$.

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. [40] 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 $\sup_{x \in \mathbb{R}} |\Pr_{X \sim P}[X \leq x] - \Pr_{X \sim Q}[X \leq x]|$. 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 $\sup_{E \in \mathcal{F}} |P(E) - Q(E)|$.

3 Summary of Our Results

We summarize our results in the following two tables. Table 1 contains all sample-based results and Table 2 contains all results under the max-min learning model.

Valuations	# bidders	Distributions	Approximation	Sample Complexity
additive [28]	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 [35]	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 [35]	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	$\Omega\left(\frac{n}{\max\{n, m\}}\right) \cdot \text{OPT}$	<i>prior-independent</i>

Table 1: Summary of Our Sample-based Results.

4 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 3. Indeed, we obtain low sample complexity bounds for indicators over *single-intersecting* sets (see

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 2: Summary of Our Max-min Learning Results.

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.

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 $\mathcal{P} : \mathcal{X}^{s(\epsilon, \delta)} \rightarrow \Delta(\mathcal{X})$ such that for any distribution $\mathcal{D} \in \Delta(\mathcal{X})$ when $k = s(\epsilon, \delta)$:*

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

When \mathcal{X} is the Cartesian product of a collection of sets $\mathcal{X}_1, \dots, \mathcal{X}_k$, 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 $\mathcal{X}_1, \dots, \mathcal{X}_d$ 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 $i \in [d]$,*

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

then

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

Proof. Let \mathcal{F}_i and $\hat{\mathcal{F}}_i$ be the probability measure function for \mathcal{D}_i and $\hat{\mathcal{D}}_i$ respectively. We will prove the statement using a hybrid argument. We create a sequence of product distributions $\{\mathcal{D}^{(j)}\}_{j \leq d}$, where $\mathcal{D}^{(j)} = \hat{\mathcal{D}}_1 \times \dots \times \hat{\mathcal{D}}_j \times \mathcal{D}_{j+1} \times \dots \times \mathcal{D}_d$, and $\mathcal{D}^{(0)} = \mathcal{D}$, $\mathcal{D}^{(d)} = \hat{\mathcal{D}}$. To prove our claim, it suffices to show that for any integer $j \in [d]$,

$$\left| \mathbb{E}_{\mathbf{x} \sim \mathcal{D}^{(j-1)}} [f(\mathbf{x})] - \mathbb{E}_{\mathbf{x} \sim \mathcal{D}^{(j)}} [f(\mathbf{x})] \right| \leq \epsilon.$$

Next, we show how to derive this inequality.

$$\begin{aligned}
& \left| \mathbb{E}_{\mathbf{x} \sim \mathcal{D}^{(j-1)}} [f(\mathbf{x})] - \mathbb{E}_{\mathbf{x} \sim \mathcal{D}^{(j)}} [f(\mathbf{x})] \right| \\
&= \left| \int_{\times_{i \neq j} \mathcal{X}_i} \left(\int_{\mathcal{X}_j} f(x_j, x_{-j}) d\mathcal{F}_j(x_j) \right) d\hat{\mathcal{F}}_1(x_1) \cdots d\hat{\mathcal{F}}_{j-1}(x_{j-1}) d\mathcal{F}_{j+1}(x_{j+1}) \cdots d\mathcal{F}_d(x_d) \right. \\
&\quad \left. - \int_{\times_{i \neq j} \mathcal{X}_i} \left(\int_{\mathcal{X}_j} f(x_j, x_{-j}) d\hat{\mathcal{F}}_j(x_j) \right) d\hat{\mathcal{F}}_1(x_1) \cdots d\hat{\mathcal{F}}_{j-1}(x_{j-1}) d\mathcal{F}_{j+1}(x_{j+1}) \cdots d\mathcal{F}_d(x_d) \right| \\
&= \left| \int_{\times_{i \neq j} \mathcal{X}_i} \left(\mathbb{E}_{x_j \sim \mathcal{D}_j} [f(x_j, x_{-j})] - \mathbb{E}_{x_j \sim \hat{\mathcal{D}}_j} [f(x_j, x_{-j})] \right) d\hat{\mathcal{F}}_1(x_1) \cdots d\hat{\mathcal{F}}_{j-1}(x_{j-1}) d\mathcal{F}_{j+1}(x_{j+1}) \cdots d\mathcal{F}_d(x_d) \right| \\
&\leq \epsilon \cdot \int_{\times_{i \neq j} \mathcal{X}_i} d\hat{\mathcal{F}}_1(x_1) \cdots d\hat{\mathcal{F}}_{j-1}(x_{j-1}) d\mathcal{F}_{j+1}(x_{j+1}) \cdots d\mathcal{F}_d(x_d) \\
&= \epsilon
\end{aligned}$$

□

Theorem 1. Let $\mathcal{X}_1, \dots, \mathcal{X}_d$ 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 $\mathcal{H}_i = \{g \mid \exists f \in \mathcal{H} \exists a_{-i} \in \times_{j \neq i} \mathcal{X}_j \forall x_i \in \mathcal{X}_i, g(x_i) = f(x_i, a_{-i})\}$.

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

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

Proof of Theorem 1: Since $\ell \geq s_i(\epsilon/d, \delta/d)$, $\Pr \left[\sup_{g \in \mathcal{H}_i} \left| \mathbb{E}_{z \sim \hat{\mathcal{D}}_i} [g(z)] - \mathbb{E}_{z \sim \mathcal{D}_i} [g(z)] \right| \leq \epsilon/d \right] \geq 1 - \delta/d$ for all $i \in [d]$. By the union bound, with probability at least $1 - \delta$, $\sup_{g \in \mathcal{H}_i} \left| \mathbb{E}_{z \sim \hat{\mathcal{D}}_i} [g(z)] - \mathbb{E}_{z \sim \mathcal{D}_i} [g(z)] \right| \leq \epsilon/d$ for all $i \in [d]$. According to Lemma 1, $\sup_{f \in \mathcal{H}} \left| \mathbb{E}_{\mathbf{z} \sim \times_{i \in [d]} \hat{\mathcal{D}}_i} [f(\mathbf{z})] - \mathbb{E}_{\mathbf{z} \sim \times_{i \in [d]} \mathcal{D}_i} [f(\mathbf{z})] \right| \leq \epsilon$ with probability at least $1 - \delta$. □

Theorem 2. Let $\mathcal{X}_1, \dots, \mathcal{X}_d$ 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

$$s(\epsilon, \delta) = \min_{k, \text{partitions}} \max_{i=1, \dots, k} s_{T_i}(\epsilon/k, \delta/k). \quad (2)$$

$$T_1 \sqcup T_2 \sqcup \dots \sqcup T_k = [d]$$

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

In particular, let $\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(\ell)}$ 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 $\tilde{T}_1 \sqcup \tilde{T}_2 \sqcup \dots \sqcup \tilde{T}_{\tilde{k}} = [d]$. Define $\hat{\mathcal{D}}_{\tilde{T}_i} =$

$\mathcal{P}_{\tilde{T}_i}(\mathbf{z}_{\tilde{T}_i}^{(1)}, \dots, \mathbf{z}_{\tilde{T}_i}^{(\ell)})$, for all $i \in [d]$, where $\mathbf{z}_{\tilde{T}_i}^{(j)}$ contains the entries of sample $\mathbf{z}^{(j)}$ 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

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

Proof of Theorem 2: For every possible partition use Theorem 1. \square

Next, 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*

$$V_{\max} = \min_{k, \text{partitions}} \left\{ k^2 \cdot \max_{i=1, \dots, k} V_{T_i} \right\}. \quad (3)$$

$$T_1 \sqcup T_2 \sqcup \dots \sqcup T_k = [d]$$

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

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,

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

where for a given sample $\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(\ell)}$ 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 $\mathbf{z}_{\tilde{T}_i}^{(1)}, \dots, \mathbf{z}_{\tilde{T}_i}^{(\ell)}$, where $\mathbf{z}_{\tilde{T}_i}^{(j)}$ contains the entries of sample $\mathbf{z}^{(j)}$ in coordinates \tilde{T}_i .

Table 3 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 $\ell = O\left(\frac{d^2}{\epsilon^2} \cdot (\log \frac{d}{\delta} + \log \frac{d}{\epsilon})\right)$ 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 [25].

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}^ℓ , \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 $L_i = \{x \in \mathbb{R}^\ell \mid x_{-i} = a_{-i}\}$, where $a_{-i} \in \mathbb{R}^{\ell-1}$, the intersection of L_i and \mathcal{E} is of the form $\{x \in \mathbb{R}^\ell \mid x_{-i} = a_{-i}, x_i \in [a, \bar{a}]\}$ where $a \leq \bar{a}$. In particular, we allow 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}^ℓ . Then \mathcal{H} has (ϵ, δ) -p.m. uniform convergence with sample complexity $O\left(\frac{\ell^2}{\epsilon^2} \cdot \log \frac{\ell}{\delta}\right)$.

Proof. As the projected hypothesis class for the i -th coordinate simply contains all intervals in \mathbb{R} , the sample complexity for (ϵ, δ) -uniform convergence is $O\left(\frac{1}{\epsilon^2} \cdot \log \frac{1}{\delta}\right)$ due to the DKW inequality. The claim follows from Theorem 1. \square

Next, we show a slightly 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 $\|\mathcal{D}_i - \hat{\mathcal{D}}_i\|_K \leq \xi$, $|\Pr_{\mathcal{D}}[\mathcal{E}] - \Pr_{\hat{\mathcal{D}}}[\mathcal{E}]| \leq 2\xi \cdot \ell$ for any single-intersecting event \mathcal{E} .

Proof. Let $\mathcal{H} = \{\mathbb{1}_{x \in \mathcal{E}} : \mathcal{E} \text{ is single-intersecting}\}$. By the definition of single-intersecting events, \mathcal{H}_i is the set of the indicator functions of all intervals in \mathbb{R} for any $i \in [\ell]$. Since $\|\mathcal{D}_i - \hat{\mathcal{D}}_i\|_K \leq \xi$,

$$\sup_{g \in \mathcal{H}_i} \left| \mathbb{E}_{x \sim \mathcal{D}_i}[g(x)] - \mathbb{E}_{x \sim \hat{\mathcal{D}}_i}[g(x)] \right| \leq 2\xi.$$

By Lemma 1,

$$\sup_{f \in \mathcal{H}} \left| \mathbb{E}_{\mathbf{x} \sim \mathcal{D}}[f(\mathbf{x})] - \mathbb{E}_{\mathbf{x} \sim \hat{\mathcal{D}}}[f(\mathbf{x})] \right| \leq 2\xi \cdot \ell.$$

\square

The following table (Table 3) summarizes some uniform convergence bounds implied by our results in this section.

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 3: Number of samples required for $(\epsilon, \Theta(1))$ -p.m. uniform convergence for different \mathcal{H} 's.

5 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. SPEMs; 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.⁵ 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, 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.

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.

- 1: $S \leftarrow [m]$
- 2: **for** $i \in [n]$ **do**
- 3: Show bidder i the set of available items S and set the entry fee for bidder i to be $\delta_i(S)$.
- 4: **if** Bidder i pays the entry fee $\delta_i(S)$ **then**
- 5: i receives her favorite bundle S_i^* and pays $\sum_{j \in S_i^*} p_{ij}$.
- 6: $S \leftarrow S \setminus S_i^*$.
- 7: **else**
- 8: i gets nothing and pays 0.
- 9: **end if**
- 10: **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 $\|\mathcal{D}_j - \hat{\mathcal{D}}_j\|_K \leq \xi$ 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 $\|\mathcal{D}_j - \hat{\mathcal{D}}_j\|_K \leq \xi$ for all item j , $\|\mathcal{L} - \hat{\mathcal{L}}\|_{TV} \leq 2m\xi$.*

Proof. For any set $R \subseteq S$, let \mathcal{E}_R be the event that the bidder purchases set R . Proving that the total variation distance between \mathcal{L} and $\hat{\mathcal{L}}$ is no more than $2m \cdot \xi$ is the same as proving that for any $K \leq 2^{|S|}$, $\left| \Pr_{\mathcal{D}} \left[t \in \bigcup_{\ell=1}^K \mathcal{E}_{R_\ell} \right] - \Pr_{\hat{\mathcal{D}}} \left[t \in \bigcup_{\ell=1}^K \mathcal{E}_{R_\ell} \right] \right| \leq 2m \cdot \xi$ where R_1, \dots, R_K are arbitrary distinct subsets of

⁵Cai 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.

S . Since the dimension of the bidder's type space is m , if we can prove that $\bigcup_{\ell=1}^K \mathcal{E}_{R_\ell}$ is always single-intersecting, our claim follows from Lemma 3.

For any $j \in [m]$ and $a_{-j} \in \mathbb{R}_{\geq 0}^{m-1}$, let $L_j(a_{-j}) = \{(t_j, a_{-j}) \mid t_j \in \mathbb{R}_{\geq 0}\}$. We claim that $L_j(a_{-j})$ intersects with at most two different \mathcal{E}_U and \mathcal{E}_V where U and V are subsets of S . WLOG, we assume that $(0, a_{-j}) \in \mathcal{E}_U$.

- If $U = \emptyset$, that means the utility of the favorite set for type $(0, a_{-j})$ is smaller than the entry fee $\delta(S)$. If we increase the value of t_j , two cases could happen: (1) the utility of the favorite set is still lower than the entry fee; (2) the utility of the favorite set is higher than the entry fee. In case (1), $(t_j, a_{-j}) \in \mathcal{E}_\emptyset$. In case (2), the bidder pays the entry fee and purchases her favorite set V . Then item j must be in V , because otherwise the utility for set V does not change from type $(0, a_{-j})$ to type (t_j, a_{-j}) . If we keep increasing t_j , bidder i 's favorite set remains to be V and she keeps accepting the entry fee and purchasing V . Hence, $L_j(a_{-j})$ can intersect with at most one event \mathcal{E}_R where R is non-empty.
- If $U \neq \emptyset$, that means U is the favorite set of type $(0, a_{-j})$ and the utility for winning set U is higher than the entry fee. If we increase the value of t_j , two cases could happen: (1) U remains the favorite set; (2) a different set V becomes the new favorite set. In case (1), $(t_j, a_{-j}) \in \mathcal{E}_U$. In case (2), item j must lie in V but not in U , otherwise how could U be better than V for type $(0, a_{-j})$ but worse for type (t_j, a_{-j}) . If we keep increasing t_j , the bidder's favorite set remains to be V and she keeps accepting the entry fee and purchasing V . Hence, $L_j(a_{-j})$ can intersect with at most two different events.

It is not hard to see that any event \mathcal{E}_R is an intersection of halfspaces, so the intersection of $L_j(a_{-j})$ with any event \mathcal{E}_R is an interval. Also, notice that any type $t \in \mathbb{R}_{\geq 0}^m$ must lie in an event \mathcal{E}_R for some set $R \subseteq S$. If $L_j(a_{-j})$ intersects with two different events \mathcal{E}_U and \mathcal{E}_V , the two intersected intervals must lie back to back on $L_j(a_{-j})$. Otherwise, $L_j(a_{-j})$ intersects with at least three different events. Contradiction. Since $L_j(a_{-j})$ intersects with at most two different events, no matter which of these events are in $\{\mathcal{E}_{R_\ell}\}_{\ell \in [K]}$, the intersection of $L_j(a_{-j})$ and $\bigcup_{\ell=1}^K \mathcal{E}_{R_\ell}$ is either empty or an interval meaning $\bigcup_{\ell=1}^K \mathcal{E}_{R_\ell}$ is single-intersecting. Now our claim simply follows from Lemma 3. \square

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

$$\left| \text{REV} - \widehat{\text{REV}} \right| \leq 2nm\xi \cdot (mH + \text{OPT}).$$

Proof. We use a hybrid argument. Consider a sequence of distributions $\{D^{(i)}\}_{i \leq n}$, where $D^{(i)} = \widehat{D}_1 \times \dots \times \widehat{D}_i \times D_{i+1} \times \dots \times D_n$, and $D^{(0)} = D$, $D^{(n)} = \widehat{D}$. We use $\text{REV}^{(i)}$ to denote the expected revenue of the SPEM under $D^{(i)}$. To prove our claim, it suffices to argue that $\left| \text{REV}^{(i-1)} - \text{REV}^{(i)} \right| \leq 2\xi m \cdot (m \cdot H + \text{OPT})$. We denote by \mathcal{S}_k and \mathcal{S}'_k the random set of items that remain available after visiting the first k bidders under $D^{(i-1)}$ and $D^{(i)}$. Clearly, for $k \leq i-1$, $\|\mathcal{S}_k - \mathcal{S}'_k\|_{TV} = 0$, so the expected revenue collected from the first $i-1$ bidders under $D^{(i-1)}$ and $D^{(i)}$ is the same. According to Lemma 4, $\|\mathcal{S}_i - \mathcal{S}'_i\|_{TV} \leq 2m \cdot \xi$. The total amount of money bidder i spends can never be higher than her value for receiving all the items which is at most $m \cdot H$. So the difference in the expected revenue collected from bidder i under $D^{(i-1)}$ and $D^{(i)}$ is at most $2\xi \cdot m^2 H$. Suppose R is the set of remaining items after visiting the first i bidders, then the expected revenue collected from the last $n-i$ bidders is the same under $D^{(i-1)}$ and $D^{(i)}$, as these bidders have the same distributions. Moreover, this expected revenue is no more than OPT , since the optimal mechanism can simply just sell R to the last $n-i$ bidders using the same prices and entry fee as in the SPEM we consider. Of course, for any fixed R , the probabilities that $\mathcal{S}_i = R$ and $\mathcal{S}'_i = R$ are different, but since for any R the expected revenue from the last $n-i$ bidders is at most OPT , the difference in the expected revenue from the last $n-i$ bidders under

$D^{(i-1)}$ and $D^{(i)}$ is at most $\|\mathcal{S}_i - \mathcal{S}'_i\|_{TV} \cdot \text{OPT} \leq 2\xi \cdot m\text{OPT}$. Hence, the total difference between $\text{REV}^{(i-1)}$ and $\text{REV}^{(i)}$ is at most $2\xi m \cdot (mH + \text{OPT})$. \square

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 $\|D_{ij} - \hat{D}_{ij}\|_K \leq \epsilon$ 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 $\frac{\text{OPT}}{c} - \epsilon \cdot O(m^2 nH)$, 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 5.1 and 5.2 for details.

5.1 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, 32, 11] that there exists a sequential posted price mechanism (SPM see Algorithm 2 for details) 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 2 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 Appendix B.

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 $\frac{\text{OPT}}{144} - \epsilon H$ with probability $1 - \delta$ given $O\left(\left(\frac{1}{\epsilon}\right)^2 (m^2 n \log \frac{n}{\epsilon} + \log \frac{1}{\delta})\right)$ 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 $\frac{\text{OPT}}{33}$ 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 $(\frac{1}{4} - (n + m) \cdot \epsilon) \cdot (\frac{\text{OPT}}{8} - 2\epsilon \cdot mnH)$ ⁶.*

Sample Access to Bounded Distributions: the result is due to Morgenstern and Roughgarden [35].

Direct Access to Approximate Distributions: we first consider a convex program based on D (see Figure 1) 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} (see Figure 2) 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

⁶If we set ϵ to be $O(\frac{1}{m+n})$, this is the max-min guarantee we want to achieve.

we are given \hat{D} , we can solve the second convex program and convert its optimal solution into a randomized SPM in polynomial time. See Theorem 10 in Appendix B.1 for further 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. See Theorem 12 in Appendix B.3 for further details.

5.2 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 [42, 11]:

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

VCG with Entry Fee: the mechanism solicits bids $\mathbf{b} = (b_1, \dots, b_n)$ 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 $t_i \sim D_i$ ⁷. 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 [42, 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 [28] 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 $O\left(\left(\frac{m}{\epsilon}\right)^2 \cdot (n \log n \log \frac{1}{\epsilon} + \log \frac{1}{\delta})\right)$ 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 $\epsilon \leq \frac{1}{16 \max\{m, n\}}$.

Sample Access to Bounded Distributions: Goldner and Karlin’s proof [28] 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 [35] to approximate it. See Theorem 13 in Appendix C.1 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

⁷The entry fee function defined in [42, 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.

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. See Theorem 14 in Appendix C.2 for further details.

6 XOS Valuations

In this section we go beyond contained 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 almost the same as SPM in Algorithm 2, 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. See Algorithm 3 for details.

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 $G_{ij} := \sup_x \left\{ \Pr_{t_{ij} \sim D_{ij}} [V_i(t_{ij}) \geq x] \geq \frac{1}{5 \max\{m,n\}} \right\}$.*

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 5.1, 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 6.1. 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. We postpone the proof to Appendix D.

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 $\Pr_{t_i \sim D_i} [u'_i(t_i, S) \geq \delta'_i(S)] \in [1/2 - \mu, 1/2 + \mu]$ 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', δ') or the best RSPM achieves revenue at least $\frac{OPT}{C_1(\mu)} - C_2(\mu) \cdot (m + n) \cdot \epsilon$ when bidders' valuations are XOS over independent items. Both $C_1(\cdot)$ and $C_2(\cdot)$ 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 $\frac{OPT}{C_1(\mu)} - C_2(\mu) \cdot (m + n) \cdot \epsilon$.

6.1 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 $O\left(\left(\frac{mn}{\xi}\right)^2 \cdot (m \cdot \log \frac{m+n}{\xi} + \log \frac{1}{\delta})\right)$ 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 $O(\max\{m, n\}^2 m^2 n^2 (m \log(m+n) + \log \frac{1}{\delta}))$ samples from D .

The bounded case is proved as Theorem 15 in Appendix D.1. The regular case is proved as Theorem 16 in Appendix D.1.

7 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 6 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 Appendix E.

- **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. See Appendix E.2 for details.

- **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 5.1, 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. See Appendix E.3 for details.

Appendix

A Our Mechanisms

Here are the detailed description of the two major mechanisms we use: Sequential Posted Price Mechanism (SPM) and Anonymous Sequential Posted Price with Entry Fee Mechanism (ASPE). We also use the Rationed Sequential Posted Price Mechanism (RSPM) when bidders are not unit-demand. RSPM is almost identical to SPM except that there is an extra constraint saying that no bidder can purchase more than one item.

Algorithm 2 Sequential Posted Price Mechanism (SPM)

Require: P_{ij} is the price for bidder i to purchase item j .

- 1: $S \leftarrow [m]$
 - 2: **for** $i \in [n]$ **do**
 - 3: Show bidder i the set of available items S .
 - 4: i purchases her favorite bundle $S_i^* \in \max_{S' \subseteq S} v_i(t_i, S') - \sum_{j \in S'} P_{ij}$ and pays $\sum_{j \in S_i^*} P_{ij}$.
 - 5: $S \leftarrow S \setminus S_i^*$.
 - 6: **end for**
-

Algorithm 3 Anonymous Sequential Posted Price with Entry Fee Mechanism (ASPE)

Require: A collection of prices $\{p_j\}_{j \in [m]}$ 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.

- 1: $S \leftarrow [m]$
 - 2: **for** $i \in [n]$ **do**
 - 3: Show bidder i the set of available items S and set the entry fee for bidder i to be $\delta_i(S)$.
 - 4: **if** Bidder i pays the entry fee $\delta_i(S)$ **then**
 - 5: i receives her favorite bundle S_i^* and pays $\sum_{j \in S_i^*} p_j$.
 - 6: $S \leftarrow S \setminus S_i^*$.
 - 7: **else**
 - 8: i gets nothing and pays 0.
 - 9: **end if**
 - 10: **end for**
-

B Missing Details from Section 5.1

B.1 Unit-demand Valuations: direct access to approximate distributions

We first consider the model where we only have access to an approximate distribution \hat{D} . The following definition is crucial for proving our result.

Definition 6. For any single dimensional distribution \mathcal{D} with cdf F , we define its revenue curve $R_{\mathcal{D}} : [0, 1] \mapsto \mathbb{R}_{\geq 0}$ as

$$\begin{aligned} R_{\mathcal{D}}(q) &= \max x \cdot \underline{q} \cdot F^{-1}(1 - q) + (1 - x) \cdot \bar{q} \cdot F^{-1}(1 - \bar{q}) \\ \text{s.t. } &x \cdot \underline{q} + (1 - x) \cdot \bar{q} = q \\ &x, \underline{q}, \bar{q} \in [0, 1] \end{aligned}$$

where $F^{-1}(1 - p) = \sup\{x \in R : \Pr_{v \sim \mathcal{D}}[v \geq x] \geq p\}$.

Lemma 6 (Folklore). Let $\varphi_{ij}(\cdot)$ and $\hat{\varphi}_{ij}(\cdot)$ be the ironed virtual value function for distribution D_{ij} and \hat{D}_{ij} respectively, then for any $q \in [0, 1]$, $R_{D_{ij}}(q) = \int_{F_{ij}^{-1}(1-q)}^H \varphi(x) dF(x)$ and $R_{\hat{D}_{ij}}(q) = \int_{\hat{F}_{ij}^{-1}(1-q)}^H \hat{\varphi}(x) dF(x)$. Since the ironed virtual value function is monotonically non-decreasing, $R_{D_{ij}}(\cdot)$ and $R_{\hat{D}_{ij}}(\cdot)$ are concave functions.

We provide an upper bound of the optimal revenue using $R_{D_{ij}}$ in the next Lemma. To do that, we first need the definition of the *Single-Dimensional Copies Setting*.

Single-Dimensional Copies Setting: In the analysis for unit-demand bidders in [15, 11], the optimal revenue is upper bounded by the optimal revenue in the single-dimensional copies setting defined in [15]. We use the same technique. We construct nm agents, where agent (i, j) has value $V_i(t_{ij})$ of being served with $t_{ij} \sim D_{ij}$, and we are only allow to use matchings, that is, for each i at most one agent (i, k) is served and for each j at most one agent (k, j) is served⁸. Notice that this is a single-dimensional setting, as each agent's type is specified by a single number. Let $\text{OPT}^{\text{COPIES-UD}}$ be the optimal BIC revenue in this copies setting.

Lemma 7. For unit-demand bidders, there exists a collection of non-negative numbers $\{q_{ij}\}_{i \in [n], j \in [m]}$ satisfying $\sum_i q_{ij} \leq 1$ for all $j \in [m]$ and $\sum_j q_{ij} \leq 1$ for all $i \in [n]$, such that the optimal revenue

$$\text{OPT} \leq 4 \cdot \sum_{i,j} R_{D_{ij}}(q_{ij}).$$

Proof. As shown in [11], $\text{OPT} \leq 4\text{OPT}^{\text{COPIES-UD}}$. Let q_{ij} be the ex-ante probability that agent (i, j) is served in the optimal mechanism for the copies setting. Chawla et al. [15] showed that $\text{OPT}^{\text{COPIES-UD}} \leq \sum_{i,j} R_{D_{ij}}(q_{ij})$. Our statement follows from the two inequalities above. \square

Next, we consider a convex program (Figure 1) and argue that the value of the optimal solution of this program is at least $\frac{1}{8}$ of the optimal revenue.

Lemma 8. The optimal solution of convex program in Figure 1 is at least $\frac{\text{OPT}}{8}$.

⁸This is exactly the copies setting used in [15], if every bidder i is unit-demand and has value $V_i(t_{ij})$ with type t_i . Notice that this unit-demand multi-dimensional setting is equivalent as adding an extra constraint, each buyer can purchase at most one item, to the original setting with subadditive bidders.

$$\begin{aligned}
& \max \sum_{i,j} R_{D_{ij}}(q_{ij}) \\
\text{s.t. } & \sum_i q_{ij} \leq \frac{1}{2} \quad \text{for all } j \in [m] \\
& \sum_j q_{ij} \leq \frac{1}{2} \quad \text{for all } i \in [n] \\
& q_{ij} \geq 0 \quad \text{for all } i \in [n] \text{ and } j \in [m]
\end{aligned}$$

Figure 1: A Convex Program for Unit-demand Bidders with Exact Distributions.

Proof. Let $\{q'_{ij}\}$ be the collection of nonnegative numbers in Lemma 7. Clearly, $\left\{\frac{q'_{ij}}{2}\right\}$ is a set of feasible solution for the convex program. Since $R_{D_{ij}}(\cdot)$ is concave, $R_{D_{ij}}\left(\frac{q'_{ij}}{2}\right) \geq \frac{R_{D_{ij}}(q'_{ij})}{2} + \frac{R_{D_{ij}}(0)}{2} = \frac{R_{D_{ij}}(q'_{ij})}{2}$. Therefore,

$$\sum_{i,j} R_{D_{ij}}\left(\frac{q'_{ij}}{2}\right) \geq \frac{1}{2} \cdot \sum_{i,j} R_{D_{ij}}(q'_{ij}) \geq \frac{\text{OPT}}{8}.$$

□

If we know all F_{ij} exactly, we can solve the convex program (Figure 1) and use the optimal solution to construct an SPM via an approach provided in [15, 11]. The constructed sequential posted mechanism has revenue at least $\frac{1}{4}$ of the optimal value of the convex program, which is at least $\frac{\text{OPT}}{32}$. Next, we show that with only access to \hat{F}_{ij} , we can essentially carry out the same approach. Consider a different convex program (Figure 2).

$$\begin{aligned}
& \max \sum_{i,j} R_{\hat{D}_{ij}}(q_{ij}) \\
\text{s.t. } & \sum_i q_{ij} \leq \frac{1}{2} + n \cdot \epsilon \quad \text{for all } j \in [m] \\
& \sum_j q_{ij} \leq \frac{1}{2} + m \cdot \epsilon \quad \text{for all } i \in [n] \\
& q_{ij} \geq 0 \quad \text{for all } i \in [n] \text{ and } j \in [m]
\end{aligned}$$

Figure 2: A Convex Program for Unit-demand Bidders with Approximate Distributions.

Not that if the support size for all \hat{D}_{ij} is upper bounded by some finite number s , the convex program above can be rewritten as a linear program with size $\text{poly}(n, m, s)$. In the following Lemma, we prove that the optimal values of the two convex programs above are close.

Lemma 9. *Let $\{q_{ij}^*\}_{i \in [n], j \in [m]}$ and $\{\hat{q}_{ij}\}_{i \in [n], j \in [m]}$ be the optimal solution of the convex program in Figure 1 and 2 respectively.*

$$\sum_{i,j} R_{\hat{D}_{ij}}(\hat{q}_{ij}) \geq \sum_{i,j} R_{D_{ij}}(q_{ij}^*) - \epsilon \cdot mnH.$$

Proof. We first fix some notations. For any bidder i and item j , let q_{ij}^*, \bar{q}_{ij}^* and $x_{ij} \in [0, 1]$ be the numbers satisfy that $x_{ij} \cdot q_{ij}^* \cdot F_{ij}^{-1}(1 - q_{ij}^*) + (1 - x_{ij}) \cdot \bar{q}_{ij}^* \cdot F_{ij}^{-1}(1 - \bar{q}_{ij}^*) = R_{D_{ij}}(q_{ij}^*)$ and $x_{ij} \cdot q_{ij}^* + (1 - x_{ij}) \cdot \bar{q}_{ij}^* = q_{ij}^*$.

Let $p_{ij} = F_{ij}^{-1}(1 - q_{ij}^*)$, $\bar{p}_{ij} = F_{ij}^{-1}(1 - \bar{q}_{ij}^*)$, and $q'_{ij} = x_{ij} \cdot (1 - \hat{F}_{ij}(p_{ij})) + (1 - x_{ij}) \cdot (1 - \hat{F}_{ij}(\bar{p}_{ij}))$. By the definition of $R_{\hat{D}_{ij}}(\cdot)$,

$$R_{\hat{D}_{ij}}(q'_{ij}) \geq x_{ij} \cdot (1 - \hat{F}_{ij}(p_{ij})) \cdot p_{ij} + (1 - x_{ij}) \cdot (1 - \hat{F}_{ij}(\bar{p}_{ij})) \cdot \bar{p}_{ij} \quad (4)$$

Since $\|\hat{D}_{ij} - D_{ij}\|_K \leq \epsilon$, $\hat{F}_{ij}(p_{ij}) \in [1 - q_{ij}^* - \epsilon, 1 - q_{ij}^* + \epsilon]$ and $\hat{F}_{ij}(\bar{p}_{ij}) \in [1 - \bar{q}_{ij}^* - \epsilon, 1 - \bar{q}_{ij}^* + \epsilon]$. Hence, the RHS of inequality (4) is greater than $R_{D_{ij}}(q_{ij}^*) - \epsilon \cdot H$. Therefore, $R_{\hat{D}_{ij}}(q'_{ij}) \geq R_{D_{ij}}(q_{ij}^*) - \epsilon \cdot H$.

Next, we argue that $\{q'_{ij}\}_{i \in [n], j \in [m]}$ is a feasible solution for the convex program in Figure 2. Since $1 - \hat{F}_{ij}(p_{ij}) \leq q_{ij}^* + \epsilon$ and $1 - \hat{F}_{ij}(\bar{p}_{ij}) \leq \bar{q}_{ij}^* + \epsilon$, $q'_{ij} \leq q_{ij}^* + \epsilon$. Thus, $\sum_i q'_{ij} \leq \sum_i q_{ij}^* + n \cdot \epsilon \leq \frac{1}{2} + n \cdot \epsilon$ for all $j \in [m]$. Similarly, we can prove $\sum_j q'_{ij} \leq \frac{1}{2} + m \cdot \epsilon$ for all $i \in [n]$. As $\{\hat{q}_{ij}\}_{i \in [n], j \in [m]}$ is the optimal solution for the second convex program, $\sum_{i,j} R_{\hat{D}_{ij}}(\hat{q}_{ij}) \geq \sum_{i,j} R_{\hat{D}_{ij}}(q'_{ij}) \geq \sum_{i,j} R_{D_{ij}}(q_{ij}^*) - \epsilon \cdot mnH$. \square

Finally, we show how to use the optimal solution of the convex program in Figure 2 to construct an SPM that approximates the optimal revenue well. We first provide a general transformation that turns any approximately feasible solution of convex program in Figure 1 to an SPM mechanism.

Lemma 10. *For any distribution $\mathcal{D} = \times_{i \in [n], j \in [m]} \mathcal{D}_{ij}$, given a collection of independent random variables $\{p_{ij}\}_{i \in [n], j \in [m]}$ such that*

$$\sum_{i \in [n]} \Pr_{p_{ij}, t_{ij} \sim \mathcal{D}_{ij}} [t_{ij} \geq p_{ij}] \leq 1 - \eta_1, \quad \text{for all } j \in [m]$$

and

$$\sum_{j \in [m]} \Pr_{p_{ij}, t_{ij} \sim \mathcal{D}_{ij}} [t_{ij} \geq p_{ij}] \leq 1 - \eta_2, \quad \text{for all } i \in [n],$$

we can construct in polynomial time a randomized SPM such that the revenue under \mathcal{D} is at least

$$\eta_1 \eta_2 \cdot \sum_{i,j} \mathbb{E}_{p_{ij}} \left[p_{ij} \cdot \Pr_{t_{ij} \sim \mathcal{D}_{ij}} [t_{ij} \geq p_{ij}] \right].$$

Proof. Consider a randomized SPM that sells item j to bidder i at price p_{ij} . Notice that bidder i purchases exactly item j if all of the following three conditions hold: (i) for all bidders $\ell \neq i$, $t_{\ell j}$ is smaller than the corresponding price $p_{\ell j}$, (ii) for all items $k \neq j$, t_{ik} is smaller than the corresponding price p_{ik} , and (iii) t_{ij} is greater than the corresponding price p_{ij} . These three conditions are independent from each other. The first condition holds with probability at least $1 - \sum_{\ell \neq i} \Pr_{p_{\ell j}, t_{\ell j} \sim \mathcal{D}_{\ell j}} [t_{\ell j} \geq p_{\ell j}] \geq \eta_1$. The second condition holds with probability at least $1 - \sum_{k \neq j} \Pr_{p_{ik}, t_{ik} \sim \mathcal{D}_{ik}} [t_{ik} \geq p_{ik}] \geq \eta_2$. When the first two conditions hold, bidder i purchases item j whenever she can afford it. Her expected payment is $\mathbb{E}_{p_{ij}} [p_{ij} \cdot \Pr_{t_{ij} \sim \mathcal{D}_{ij}} [t_{ij} \geq p_{ij}]]$. Hence, the expected revenue for selling item j to bidder i is at least $\eta_1 \eta_2 \cdot \mathbb{E}_{p_{ij}} [p_{ij} \cdot \Pr_{t_{ij} \sim \mathcal{D}_{ij}} [t_{ij} \geq p_{ij}]]$ and the total expected revenue is at least $\eta_1 \eta_2 \cdot \sum_{i,j} \mathbb{E}_{p_{ij}} [p_{ij} \cdot \Pr_{t_{ij} \sim \mathcal{D}_{ij}} [t_{ij} \geq p_{ij}]]$. \square

Lemma 11. *Given any feasible solution $\{q_{ij}\}_{i \in [n], j \in [m]}$ of the convex program in Figure 2, we can construct a (randomized) SPM in polynomial time such that its revenue under D is at least $(\frac{1}{4} - (n + m) \cdot \epsilon) \cdot (\sum_{i,j} R_{\hat{D}_{ij}}(q_{ij}) - \epsilon \cdot nmH)$.*

Proof. We first fix some notations. For any bidder i and item j , let $\underline{q}_{ij}, \bar{q}_{ij}$ and $x_{ij} \in [0, 1]$ be the numbers satisfying $x_{ij} \cdot \underline{q}_{ij} \cdot \hat{F}_{ij}^{-1}(1 - \underline{q}_{ij}) + (1 - x_{ij}) \cdot \bar{q}_{ij} \cdot \hat{F}_{ij}^{-1}(1 - \bar{q}_{ij}) = R_{\hat{D}_{ij}}(q_{ij})$ and $x_{ij} \cdot \underline{q}_{ij} + (1 - x_{ij}) \cdot \bar{q}_{ij} = q_{ij}$.

We use p_{ij} to denote a random variable that is $p_{ij} = \hat{F}_{ij}^{-1}(1 - q_{ij})$ with probability x_{ij} and $\bar{p}_{ij} = \hat{F}_{ij}^{-1}(1 - \bar{q}_{ij})$ with probability $1 - x_{ij}$.

Next, we construct a randomized SPM based on $\{p_{ij}\}_{i \in [n], j \in [m]}$ according to Lemma 10. Note that

$$\sum_{i \in [n]} \Pr_{p_{ij}, t_{ij} \sim D_{ij}} [t_{ij} \geq p_{ij}] \leq \sum_{i \in [n]} \left(\Pr_{p_{ij}, t_{ij} \sim \hat{D}_{ij}} [t_{ij} \geq p_{ij}] + \epsilon \right) = \sum_{i \in [n]} q_{ij} + n\epsilon \leq \frac{1}{2} + 2n\epsilon$$

for all item j , and

$$\sum_{j \in [m]} \Pr_{p_{ij}, t_{ij} \sim D_{ij}} [t_{ij} \geq p_{ij}] \leq \sum_{j \in [m]} \left(\Pr_{p_{ij}, t_{ij} \sim \hat{D}_{ij}} [t_{ij} \geq p_{ij}] + \epsilon \right) = \sum_{i \in [m]} q_{ij} + m\epsilon \leq \frac{1}{2} + 2m\epsilon$$

for all bidder i . Hence, we can construct in polynomial time a randomized SPM with revenue at least

$$\begin{aligned} & \left(\frac{1}{2} - 2n\epsilon \right) \left(\frac{1}{2} - 2m\epsilon \right) \cdot \sum_{i,j} \mathbb{E}_{p_{ij}} \left[p_{ij} \cdot \Pr_{t_{ij} \sim D_{ij}} [t_{ij} \geq p_{ij}] \right] \\ & \geq \left(\frac{1}{4} - (n+m)\epsilon \right) \sum_{i,j} \mathbb{E}_{p_{ij}} \left[p_{ij} \cdot \left(\Pr_{t_{ij} \sim \hat{D}_{ij}} [t_{ij} \geq p_{ij}] - \epsilon \right) \right] \\ & \geq \left(\frac{1}{4} - (n+m)\epsilon \right) \sum_{i,j} \left(R_{\hat{D}_{ij}}(q_{ij}) - \epsilon \cdot nmH \right) \end{aligned}$$

The first inequality is because $\left\| D_{ij} - \hat{D}_{ij} \right\|_K \leq \epsilon$, and the second inequality is because p_{ij} is upper bounded by H and $\mathbb{E}_{p_{ij}} \left[p_{ij} \cdot \Pr_{t_{ij} \sim \hat{D}_{ij}} [t_{ij} \geq p_{ij}] \right] = R_{\hat{D}_{ij}}(q_{ij})$ by the definition of p_{ij} . \square

Theorem 10. *For unit-demand bidders, given distributions \hat{D}_{ij} where $\left\| \hat{D}_{ij} - D_{ij} \right\|_K \leq \epsilon$ for all $i \in [n]$ and $j \in [m]$, there is a polynomial time algorithm that constructs a randomized SPM whose revenue under D is at least $\left(\frac{1}{4} - (n+m)\epsilon \right) \cdot \left(\frac{OPT}{8} - 2\epsilon \cdot mnH \right)$.*

Proof. Our algorithm first computes the optimal solution $\{\hat{q}_{ij}\}_{i \in [n], j \in [m]}$ for the convex program in Figure 2, then constructs a randomized SPM based on $\{\hat{q}_{ij}\}_{i \in [n], j \in [m]}$ using Lemma 11. It is not hard to see that our algorithm runs in polynomial time. By chaining the inequalities in Lemma 8, 9 and 11, we can argue that the revenue of our mechanism is at least $\left(\frac{1}{4} - (n+m)\epsilon \right) \cdot \left(\frac{OPT}{8} - 2\epsilon \cdot mnH \right)$. \square

B.2 Unit-demand Valuations: sample access to bounded distributions

When the distributions D_{ij} are all bounded, the following theorem provides the sample complexity.

Theorem 11. [35] *When D_{ij} is supported on $[0, H]$ for all bidder i and item j , the sample complexity for (ϵ, δ) -uniformly learning the revenue of SPMs for unit-demand bidders is $O\left(\left(\frac{1}{\epsilon}\right)^2 \left(m^2 n \log n \log \frac{1}{\epsilon} + \log \frac{1}{\delta}\right)\right)$. That is, with probability $1 - \delta$, the empirical revenue based on the samples for any SPM is within $\epsilon \cdot H$ of its true expected revenue. Moreover, with the same number of samples, there is a polynomial time algorithm that learns an SPM whose revenue is at least $\frac{OPT}{144} - \epsilon H$ with probability $1 - \delta$.*

B.3 Unit-demand Valuations: sample access to regular distributions

In this section, we show there exists a polynomial time algorithm that learns an SPM whose revenue is at least a constant fraction of the optimal revenue with polynomial in n and m samples. Note that unlike in the previous two models, the error of our learning algorithm is only multiplicative when the distributions are regular. First, we present a Lemma regarding the revenue curve function for regular distributions.

Lemma 12. [7] *For any regular distribution F , let $R_F(\cdot)$ be the corresponding revenue curve. For any $0 < q' \leq q \leq p < 1$,*

$$(1 - p) \cdot R_F(q') \leq R_F(q).$$

Throughout this section, we use Z to denote $\max\{m, n\}$ and C to be a constant that will be specified later. Using Lemma 12, we show in the next Lemma that restricting q_{ij} to be at least $\frac{1}{CZ}$ does not affect the objective value of the convex program in Figure 1 by too much.

Lemma 13. *Suppose $\{q_{ij}^*\}_{i \in [n], j \in [m]}$ is the optimal solution of the convex program in Figure 1. Let $q'_{ij} = \max\{\frac{1}{CZ}, q_{ij}^*\}$, then $\sum_{i,j} R_{D_{ij}}(q'_{ij}) \geq (1 - \frac{1}{CZ}) \cdot \sum_{i,j} R_{D_{ij}}(q_{ij}^*) \geq (1 - \frac{1}{CZ}) \cdot \frac{OPT}{8}$.*

Proof. According to Lemma 8, $\sum_{i,j} R_{D_{ij}}(q_{ij}^*) \geq \frac{OPT}{8}$. So to prove the statement, it suffices to argue that for any i and j , $R_{D_{ij}}(q'_{ij}) \geq (1 - \frac{1}{CZ}) \cdot R_{D_{ij}}(q_{ij}^*)$. If $q_{ij}^* = q'_{ij}$, this inequality clearly holds. If $q_{ij}^* \neq q'_{ij}$, $q_{ij}^* \leq q'_{ij} = \frac{1}{CZ}$. Since F_{ij} is regular, we can apply Lemma 12 to q'_{ij} and q_{ij}^* and obtain inequality $R_{D_{ij}}(q'_{ij}) \geq (1 - \frac{1}{CZ}) \cdot R_{D_{ij}}(q_{ij}^*)$. \square

Using Lemma 13, we argue how to compute in polynomial time an approximately optimal SPM. Suppose D'_{ij} is the distribution that we obtain after truncating D_{ij} at a threshold H_{ij} ⁹, and we have direct access to a discrete distribution \hat{D}'_{ij} such that $\|\hat{D}'_{ij} - D'_{ij}\|_K \leq \epsilon$ for all i and j . We show in the following Lemma that the optimal solution of a convex program similar to the one in Figure 2 but for $\{\hat{D}'_{ij}\}_{i \in [n], j \in [m]}$ can guide us to design an approximately optimal SPM under D in polynomial time. As we have sample access to D , we will argue later that a polynomial number of samples suffices to generate $\{\hat{D}'_{ij}\}_{i \in [n], j \in [m]}$.

Lemma 14. *Let $\{H_{ij}\}_{i \in [n], j \in [m]}$ be a collection of positive numbers satisfying $F_{ij}(H_{ij}) \in [1 - \frac{1}{C \cdot Z}, 1 - \frac{1}{3C \cdot Z}]$ for all $i \in [n]$ and $j \in [m]$. Let D'_{ij} be the distribution of the random variable $\min\{t_{ij}, H_{ij}\}$ where $t_{ij} \sim D_{ij}$, and \hat{D}'_{ij} be a discrete distribution such that $\|\hat{D}'_{ij} - D'_{ij}\|_K \leq \epsilon$ for all $i \in [n]$ and $j \in [m]$. Suppose s is an upper bound of the support size for any distribution \hat{D}'_{ij} , then given direct access to \hat{D}'_{ij} , we can compute in time polynomial in n , m and s a randomized SPM that achieves revenue at least $(\frac{1}{2} - \frac{1}{C} - 2n\epsilon) \cdot (\frac{1}{2} - \frac{1}{C} - 2m\epsilon) \cdot ((1 - \frac{1}{CZ}) \cdot \frac{OPT}{8} - 2\epsilon \cdot nmH)$ under D , where $H = \max_{i,j} H_{ij}$.*

Proof. Consider the following convex program:

$$\begin{aligned} & \max \sum_{i,j} R_{\hat{D}'_{ij}}(q_{ij}) \\ \text{s.t.} \quad & \sum_i q_{ij} \leq \frac{1}{2} + \frac{1}{C} + n \cdot \epsilon \quad \text{for all } j \in [m] \\ & \sum_j q_{ij} \leq \frac{1}{2} + \frac{1}{C} + m \cdot \epsilon \quad \text{for all } i \in [n] \\ & q_{ij} \geq 0 \quad \text{for all } i \in [n] \text{ and } j \in [m] \end{aligned}$$

⁹Let $t_{ij} \sim D_{ij}$, then $\min\{t_{ij}, H_{ij}\}$ is the corresponding truncated random variable drawn from D'_{ij} .

Let $\{q_{ij}^*\}_{i \in [n], j \in [m]}$ be the optimal solution of the convex program in Figure 1 and $q'_{ij} = \max\{\frac{1}{C \cdot Z}, q_{ij}^*\}$. For every i and j , let $p'_{ij} = F_{ij}^{-1}(1 - q'_{ij})$ and $\tilde{q}_{ij} = \Pr_{t_{ij} \sim \hat{D}'_{ij}} [t_{ij} \geq p'_{ij}]$. By the definition of H_{ij} , $p'_{ij} \leq H_{ij}$, so

$$p'_{ij} \tilde{q}_{ij} \geq p'_{ij} \left(\Pr_{t_{ij} \sim \hat{D}'_{ij}} [t_{ij} \geq p'_{ij}] - \epsilon \right) \geq p'_{ij} q'_{ij} - \epsilon \cdot H_{ij} = R_{D_{ij}}(q'_{ij}) - \epsilon \cdot H_{ij}. \quad (5)$$

$p'_{ij} q'_{ij}$ equals to $R_{D_{ij}}(q'_{ij})$ because D_{ij} is a regular distribution.

Next, we argue that $\{\tilde{q}_{ij}\}_{i \in [n], j \in [m]}$ is a feasible solution of the convex program above. Observe that

$$\sum_i \tilde{q}_{ij} \leq \sum_i q'_{ij} + n\epsilon \leq \sum_i \left(q_{ij}^* + \frac{1}{CZ} \right) + n\epsilon \leq \frac{1}{2} + \frac{1}{C} + n\epsilon$$

for all item $j \in [m]$ and

$$\sum_j \tilde{q}_{ij} \leq \sum_j q'_{ij} + n\epsilon \leq \sum_j \left(q_{ij}^* + \frac{1}{CZ} \right) + m\epsilon \leq \frac{1}{2} + \frac{1}{C} + m\epsilon$$

for all bidder $i \in [n]$.

Let $\widehat{\text{OPT}}$ be the optimal solution of the convex program above. As $\{\tilde{q}_{ij}\}_{i \in [n], j \in [m]}$ is a feasible solution,

$$\widehat{\text{OPT}} \geq \sum_{i,j} R_{\hat{D}'_{ij}}(\tilde{q}_{ij}) \geq \sum_{i,j} p'_{ij} \tilde{q}_{ij} \geq \sum_{i,j} R_{D_{ij}}(q'_{ij}) - \epsilon \cdot nmH \geq \left(1 - \frac{1}{CZ}\right) \cdot \frac{\text{OPT}}{8} - \epsilon \cdot nmH.$$

The second last inequality is due to inequality (5) and the last inequality is due to Lemma 13.

So far, we have argued that the optimal solution of our convex program has value close to the OPT. We will show in the second part of the proof that using the optimal solution of our convex program, we can construct an SPM whose revenue under D is close to $\widehat{\text{OPT}}$. Let \hat{q}_{ij} be the optimal solution of the convex program above and \hat{p}_{ij} be the corresponding random price, that is, $\Pr_{\hat{p}_{ij}, t_{ij} \sim \hat{D}'_{ij}} [t_{ij} \geq \hat{p}_{ij}] = \hat{q}_{ij}$ and $R_{\hat{D}'_{ij}}(\hat{q}_{ij}) = \mathbb{E}_{\hat{p}_{ij}} \left[\hat{p}_{ij} \cdot \Pr_{t_{ij} \sim \hat{D}'_{ij}} [t_{ij} \geq \hat{p}_{ij}] \right]$. As $\hat{p}_{ij} \leq H_{ij}$,

$$\Pr_{\hat{p}_{ij}, t_{ij} \sim D_{ij}} [t_{ij} \geq \hat{p}_{ij}] = \Pr_{\hat{p}_{ij}, t_{ij} \sim \hat{D}'_{ij}} [t_{ij} \geq \hat{p}_{ij}] \in [\hat{q}_{ij} - \epsilon, \hat{q}_{ij} + \epsilon].$$

Therefore, for all item j

$$\sum_i \Pr_{\hat{p}_{ij}, t_{ij} \sim D_{ij}} [t_{ij} \geq \hat{p}_{ij}] \leq \sum_i \Pr_{\hat{p}_{ij}, t_{ij} \sim \hat{D}'_{ij}} [t_{ij} \geq \hat{p}_{ij}] + n\epsilon = \sum_i \hat{q}_{ij} + n\epsilon \leq \frac{1}{2} + \frac{1}{C} + 2n\epsilon$$

and for all bidder i

$$\sum_j \Pr_{\hat{p}_{ij}, t_{ij} \sim D_{ij}} [t_{ij} \geq \hat{p}_{ij}] \leq \sum_j \Pr_{\hat{p}_{ij}, t_{ij} \sim \hat{D}'_{ij}} [t_{ij} \geq \hat{p}_{ij}] + m\epsilon = \sum_j \hat{q}_{ij} + m\epsilon \leq \frac{1}{2} + \frac{1}{C} + 2m\epsilon.$$

According to Lemma 10, we can construct a randomized SPM with $\{\hat{p}_{ij}\}_{i \in [n], j \in [m]}$ whose revenue is at least $(\frac{1}{2} - \frac{1}{C} - 2n\epsilon) \cdot (\frac{1}{2} - \frac{1}{C} - 2m\epsilon) \cdot \sum_{i,j} \mathbb{E}_{\hat{p}_{ij}} \left[\hat{p}_{ij} \cdot \Pr_{t_{ij} \sim D_{ij}} [t_{ij} \geq \hat{p}_{ij}] \right]$ under D . Clearly,

$$\mathbb{E}_{\hat{p}_{ij}} \left[\hat{p}_{ij} \cdot \Pr_{t_{ij} \sim D_{ij}} [t_{ij} \geq \hat{p}_{ij}] \right] \geq \mathbb{E}_{\hat{p}_{ij}} \left[\hat{p}_{ij} \cdot \left(\Pr_{t_{ij} \sim \hat{D}'_{ij}} [t_{ij} \geq \hat{p}_{ij}] - \epsilon \right) \right] \geq R_{\hat{D}'_{ij}}(\hat{q}_{ij}) - \epsilon \cdot H_{ij}.$$

Therefore, the revenue of the constructed randomized SPM under D is at least

$$\begin{aligned} & \left(\frac{1}{2} - \frac{1}{C} - 2n\epsilon\right) \cdot \left(\frac{1}{2} - \frac{1}{C} - 2m\epsilon\right) \cdot \left(\widehat{\text{OPT}} - \epsilon \cdot nmH\right) \\ & \geq \left(\frac{1}{2} - \frac{1}{C} - 2n\epsilon\right) \cdot \left(\frac{1}{2} - \frac{1}{C} - 2m\epsilon\right) \cdot \left(\left(1 - \frac{1}{CZ}\right) \cdot \frac{\text{OPT}}{8} - 2\epsilon \cdot nmH\right). \end{aligned}$$

It is not hard to see that both $\{\hat{q}_{ij}\}_{i \in [n], j \in [m]}$ and $\{\hat{p}_{ij}\}_{i \in [n], j \in [m]}$ can be computed in time polynomial in n , m and s . \square

When ϵ is small enough, the additive error in Lemma 14 can be converted into a multiplicative error. Next, we argue that with a polynomial number of samples, we can learn $\{H_{ij}\}_{i \in [n], j \in [m]}$ and $\{\hat{D}'_{ij}\}_{i \in [n], j \in [m]}$ with enough accuracy.

Theorem 12. *If for all bidder i and item j , D_{ij} is a regular distribution, we can learn in polynomial time with probability $1 - \delta$ a randomized SPM whose revenue is at least $\frac{\text{OPT}}{33}$ with $O\left(Z^2 m^2 n^2 \cdot \log \frac{nm}{\delta}\right)$ ($Z = \max\{m, n\}$) samples.*

Proof. First, if we take $O\left(C^2 \cdot Z^2 \cdot \log \frac{nm}{\delta}\right)$ samples from each D_{ij} , we can find an H_{ij} such that $F_{ij}(H_{ij})$ lies in $\left[1 - \frac{1}{CZ}, 1 - \frac{1}{3CZ}\right]$ with probability $1 - \frac{\delta}{2nm}$. By the union bound, the probability that all H_{ij} satisfy the requirement is at least $1 - \frac{\delta}{2}$. From now on, we assume $F_{ij}(H_{ij}) \in \left[1 - \frac{1}{CZ}, 1 - \frac{1}{3CZ}\right]$ for all i and j . Observe that $\text{OPT} \geq \max_{i,j} H_{ij} \cdot \frac{1}{3CZ}$, as the expected revenue for selling item j to bidder i at price H_{ij} is at least $\frac{H_{ij}}{3CZ}$. Therefore, there exists sufficiently large constant d and C , if $\epsilon = \frac{1}{d \cdot Znm}$ the randomized SPM learned in Lemma 14 has revenue at least $\frac{\text{OPT}}{33}$. According to the Dvoretzky-Kiefer-Wolfowitz (DKW) inequality [25], if we take $O\left(d^2 Z^2 n^2 m^2 \cdot \log \frac{nm}{\delta}\right)$ samples from D'_{ij} (we can take samples from D_{ij} then cap the samples at H_{ij}) and let \hat{D}'_{ij} be the uniform distribution over the samples, $\left\|D'_{ij} - \hat{D}'_{ij}\right\|_K \leq \frac{1}{d \cdot Znm}$ with probability $1 - \frac{\delta}{2nm}$. By the union bound, $\left\|D'_{ij} - \hat{D}'_{ij}\right\|_K \leq \frac{1}{d \cdot Znm}$ for all $i \in [n]$ and $j \in [m]$ with probability at least $1 - \delta/2$. Finally, by another union bound, the H_{ij} and \hat{D}'_{ij} we learned from $O\left(Z^2 n^2 m^2 \cdot \log \frac{nm}{\delta}\right)$ samples satisfy $F_{ij}(H_{ij}) \in \left[1 - \frac{1}{CZ}, 1 - \frac{1}{3CZ}\right]$ and $\left\|D'_{ij} - \hat{D}'_{ij}\right\|_K \leq \frac{1}{d \cdot Znm}$ for all i and j with probability at least $1 - \delta$. In other words, we can learn a randomized SPM whose revenue is at least $\frac{\text{OPT}}{33}$ with probability at least $1 - \delta$ using $O\left(Z^2 n^2 m^2 \cdot \log \frac{nm}{\delta}\right)$ samples. Furthermore, the support size of any \hat{D}'_{ij} is at most $O\left(Z^2 n^2 m^2 \cdot \log \frac{nm}{\delta}\right)$ samples, so our learning algorithm runs in time polynomial in n and m . \square

C Missing Details from Section 5.2

C.1 Additive Valuations: sample access to bounded distributions

As shown by Goldner and Karlin [28], one sample suffices to design a mechanism that approximates BREV. The idea is to use the VCG with entry fee mechanism but replace the entry fee $e_i(b_{-i}, D_i)$ for bidder i with $e_i(b_{-i}, s_i) = \sum_{j \in [m]} (s_{ij} - \max_{k \neq i} b_{kj})^+$, where s_i is a sample drawn from D_i . It is easy to argue that for any b_{-i} , over the randomness of the sample s_i and bidder i 's real type t_i , the event that $e_i(b_{-i}, s_i) \geq e_i(b_{-i}, D_i)$ and bidder i accepts the entry fee $e_i(b_{-i}, s_i)$ happens with probability at least $\frac{1}{8}$. As $\frac{1}{2} \cdot \sum_{i \in [n]} \mathbb{E}_t[e_i(t_{-i}, D_i)] = \text{BREV}$, the expected revenue (over the randomness of the types and the samples) from their mechanism is at least $\frac{1}{8} \cdot \sum_{i \in [n]} \mathbb{E}_t[e_i(t_{-i}, D_i)] = \frac{\text{BREV}}{4}$. Next, we show how to learn a mechanism that approximates SREV.

Lemma 15. *When D_{ij} is supported on $[0, H]$ for all bidder i and item j , the sample complexity for (ϵ, δ) -uniformly learning the revenue of SPMs for additive bidders is $O\left(\left(\frac{1}{\epsilon}\right)^2 \left(m^2 n \log n \log \frac{1}{\epsilon} + \log \frac{1}{\delta}\right)\right)$. Moreover, we can learn in polynomial time an SPM whose revenue is at least $\frac{\text{SREV}}{4} - \frac{3\epsilon}{2} \cdot H$ with probability $1 - \delta$ given the same number of samples.*

Proof. The first half of the Lemma was proved by Morgenstern and Roughgarden [35]. We show how to prove the second half of the claim. Let OPT_j be the optimal revenue for selling item j . By the prophet inequality [41], there exists an SPM for selling item j with a collection of prices $\{p_{ij}\}_{i \in [n]}$ that achieves revenue at least $\text{OPT}_j/2$. As the bidders are additive, if we run the SPMs for selling each item simultaneously, the expected revenue is exactly the sum of the revenue of the SPM mechanisms for auctioning a single item. Note that the simultaneous SPM is indeed a SPM for selling all items. Hence, there exists an SPM that achieves revenue at least $\text{OPT}/2$. Since the sample complexity for (ϵ, δ) -uniformly learning the revenue of SPMs is $O\left(\left(\frac{m}{\epsilon}\right)^2 \left(n \log n \log \frac{1}{\epsilon} + \log \frac{1}{\delta}\right)\right)$, the empirical revenue induced by the samples is within $\epsilon \cdot H$ of the true expected revenue with probability $1 - \delta$ for any SPM.

We use ER_{opt} to denote the optimal empirical revenue obtained by any SPM. If we apply the prophet inequality to the empirical distribution, we can construct an SPM whose empirical revenue ER is at least $ER_{opt}/2$. Notice that ER_{opt} is at most $\epsilon \cdot H$ less than the optimal true expected revenue obtained by any SPM, which is at least $\text{OPT}/2$. Combining the two inequalities above, we have $ER \geq \text{OPT}/4 - \epsilon/2 \cdot H$ with probability $1 - \delta$. Also, the true expected revenue of our SPM is at least $ER - \epsilon \cdot H$, so our SPM achieves expected revenue at least $\frac{\text{OPT}}{4} - \frac{3\epsilon}{2} \cdot H$ with probability $1 - \delta$. \square

Now we are ready to prove our Theorem for additive bidders when their valuations are bounded.

Theorem 13. *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 $\frac{\text{OPT}}{32} - \epsilon \cdot H$ 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 .

Proof. According to Lemma 15, we can learn a mechanism whose revenue is at least $\frac{\text{SREV}}{4} - \frac{\epsilon}{24} \cdot H$ 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. As we explained in the beginning of this section, with one sample from the distribution we can construct a randomized mechanism whose expected revenue is at least $\frac{\text{BREV}}{4}$. Therefore, the better of our two mechanisms has expected revenue at least $\frac{\text{OPT}}{32} - \epsilon \cdot H$ with probability $1 - \delta$. \square

C.2 Additive Valuations: direct access to approximate distributions

In this section, we discuss how to learn an approximately optimal mechanism for additive bidders when we are given direct access to approximate value distributions. Again, we first show how to learn a mechanism whose revenue approximates SREV then we provide another mechanism whose revenue approximates BREV.

Lemma 16. *For additive bidders, given distributions \hat{D}_{ij} where $\left\|\hat{D}_{ij} - D_{ij}\right\|_K \leq \epsilon$ for all $i \in [n]$ and $j \in [m]$, there is a polynomial time algorithm that constructs a randomized SPM whose revenue under D is at least $\left(\frac{1}{4} - \epsilon \cdot n\right) \cdot \left(\frac{\text{SREV}}{8} - 2\epsilon \cdot mnH\right)$.*

Proof. Let OPT_j be the optimal revenue for selling item j . As the bidders are additive, if we can construct a randomized SPM M_j for every item j such that its expected revenue under D is at least $\left(\frac{1}{4} - \epsilon \cdot n\right) \cdot$

$\left(\frac{\text{OPT}_j}{8} - 2\epsilon \cdot nH\right)$, running these m randomized SPMs in parallel generates expected revenue at least

$$\sum_{j \in [m]} \left(\frac{1}{4} - \epsilon \cdot n\right) \cdot \left(\frac{\text{OPT}_j}{8} - 2\epsilon \cdot nH\right) = \left(\frac{1}{4} - \epsilon \cdot n\right) \cdot \left(\frac{\text{SREV}}{8} - 2\epsilon \cdot mnH\right)$$

under D . Due to Theorem 10, we can construct in polynomial time such a randomized SPM M_j for each item j based on $\times_{i \in [n]} \hat{D}_{ij}$. \square

Next, we show how to choose the entry fee based on $\hat{D} = \times_{i,j} \hat{D}_{ij}$, so that the VCG with entry fee mechanism has revenue that approximates BREV under the true distribution D . More specifically, we use the median of i 's utility under $\hat{D}_i = \times_{j \in [m]} \hat{D}_{ij}$ as bidder i 's entry fee. We prove the result in two steps. We first show that if we can use an entry fee function such that every bidder i accepts her entry fee with probability between $[1/2 - \eta, 1/2]$ for any possible bid profiles b_{-i} of the other bidders, the expected revenue is at least $(1/2 - \eta) \cdot \text{BREV}$. Second, we show how to compute in polynomial time such entry fee functions with $\eta = O(m\epsilon)$ based on \hat{D} .

Lemma 17. *Suppose for every bidder i , $d_i(\cdot) : T_{-i} \mapsto R$ is a randomized entry fee function such that for any bid profile $b_{-i} \in T_{-i}$ of the other bidders*

$$\Pr_{t_i \sim D_i} \left[\sum_{j \in [m]} \left(t_{ij} - \max_{k \neq i} b_{kj} \right)^+ \geq d_i(b_{-i}) \right] \in \left[\frac{1}{2} - \eta, \frac{1}{2} \right]$$

with probability at least $1 - \delta$. Then if we use $d_i(\cdot)$ as the entry fee function in the VCG with entry fee mechanism, the expected revenue is at least $(1 - \delta - 2\eta) \cdot \text{BREV}$.

Proof. When $\Pr_{t_i \sim D_i} \left[\sum_{j \in [m]} (t_{ij} - \max_{k \neq i} b_{kj})^+ \geq d_i(b_{-i}) \right] \in \left[\frac{1}{2} - \eta, \frac{1}{2} \right]$, $d_i(b_{-i})$ is no less than the original entry fee $e_i(b_{-i}, D_i)$ for any bid profile b_{-i} of the other bidders. The expected revenue under the new entry fee functions is at least $((1 - \delta) \cdot (\frac{1}{2} - \eta) \cdot \sum_{i \in [n]} \mathbb{E}_{b_{-i} \sim D_{-i}} [e_i(b_{-i}, D_i)]) \geq (1 - \delta - 2\eta) \cdot \text{BREV}$. \square

Lemma 18. *For any bidder i and any bid profile b_{-i} from the other bidders, let $\mathcal{F}_{i,b_{-i}}$ and $\hat{\mathcal{F}}_{i,b_{-i}}$ be the distributions for the random variable $\sum_{j \in [m]} (t_{ij} - \max_{k \neq i} b_{kj})^+$ when t_i is drawn from D_i and \hat{D}_i respectively. If $\left\| D_{ij} - \hat{D}_{ij} \right\|_K \leq \epsilon$ for all bidder i and item j , $\left\| \mathcal{F}_{i,b_{-i}} - \hat{\mathcal{F}}_{i,b_{-i}} \right\|_K \leq 2m\epsilon$ for all i and b_{-i} . Moreover, when $m\epsilon \leq 1/16$, we can compute a randomized mechanism whose expected revenue is at least $\frac{\text{BREV}}{5}$.*

Proof. For any real number x , consider event $\mathcal{E}_{i,b_{-i},x} = \left\{ t_i \mid \sum_{j \in [m]} (t_{ij} - \max_{k \neq i} b_{kj})^+ \geq x \right\}$. It is easy to see that $\mathcal{E}_{i,b_{-i},x}$ is single-intersecting for any any i, b_{-i} and x . According to Lemma 3,

$$\left| \Pr_{t_i \sim D_i} [\mathcal{E}_{i,b_{-i},x}] - \Pr_{t_i \sim \hat{D}_i} [\mathcal{E}_{i,b_{-i},x}] \right| \leq 2m\epsilon$$

for any i, b_{-i} and x . Hence, $\left\| \mathcal{F}_{i,b_{-i}} - \hat{\mathcal{F}}_{i,b_{-i}} \right\|_K \leq 2m\epsilon$.

Next, we argue how to construct a randomized entry fee $d_i(b_{-i})$ in polynomial time with only sample access of $\hat{\mathcal{F}}_{i,b_{-i}}$. Suppose we take k samples from $\hat{\mathcal{F}}_{i,b_{-i}}$ and sort them in descending order $s_1 \geq s_2 \geq \dots \geq s_k$. Let the entry fee $d_i(b_{-i})$ to be $s_{\lceil \frac{5k}{16} \rceil}$. By the Chernoff bound, with probability at least $1 - \exp(-k/128)$ (over the randomness of the samples) $\Pr_{t_i \sim \hat{D}_i} \left[\sum_{j \in [m]} (t_{ij} - \max_{k \neq i} b_{kj})^+ \geq d_i(b_{-i}) \right] = \Pr_{t_i \sim \hat{D}_i} [\mathcal{E}_{i,b_{-i},d_i(b_{-i})}]$ lies in $\left[\frac{1}{4}, \frac{3}{8} \right]$. Since $\Pr_{t_i \sim D_i} [\mathcal{E}_{i,b_{-i},d_i(b_{-i})}] = \Pr_{t_i \sim \hat{D}_i} [\mathcal{E}_{i,b_{-i},d_i(b_{-i})}] \pm 2m\epsilon$,

$$\Pr_{t_i \sim D_i} [\mathcal{E}_{i,b_{-i},d_i(b_{-i})}] \in \left[\frac{1}{8}, \frac{1}{2} \right],$$

if $m\epsilon \leq 1/16$. According to Lemma 17, the expected revenue under our entry fee $d_i(b_{-i})$ is at least $(\frac{1}{4} - \exp(-k/128)) \cdot \text{BREV} \geq \frac{\text{BREV}}{5}$ if we choose k to be larger than some absolute constant. Clearly, the procedure above can be completed in polynomial time with access to \hat{D} . \square

Combining Lemma 16 and 18, we are ready to prove our main result of this section.

Theorem 14. *If all bidders have additive valuations, given distributions \hat{D}_{ij} where $\|\hat{D}_{ij} - D_{ij}\|_K \leq \epsilon$ for all $i \in [n]$ and $j \in [m]$, there is a polynomial time algorithm that constructs a mechanism whose expected revenue under D is at least $\frac{\text{OPT}}{266} - 96\epsilon \cdot mnH$ when $\epsilon \leq \frac{1}{16 \max\{m,n\}}$.*

Proof. Since $\epsilon \leq \frac{1}{16 \max\{m,n\}}$, we can learn in polynomial time a randomized SPM whose revenue is at least $\frac{3}{16} \cdot (\frac{\text{SREV}}{8} - 2\epsilon \cdot mnH)$ and a VCG with entry fee mechanism whose revenue is at least $\text{BREV}/5$. As $\text{OPT} \leq 6 \cdot \text{SREV} + 2\text{BREV}$ (Theorem 6), the better of the two mechanisms we can learn in polynomial time has revenue at least $\frac{\text{OPT}}{266} - 96\epsilon \cdot mnH$. \square

D Missing Details from Section 6

Proof of Lemma 5: We only sketch the proof here. Let POSTREV denote the highest revenue obtainable by any RSPM. In [13], Cai and Zhao constructed an upper bound of the optimal revenue using duality and separated the upper bound into three components: SINGLE , TAIL and CORE . Both SINGLE and TAIL are within constant times the POSTREV , and the $\text{ASPE}(p^*, \delta^*)$ is used to bound the CORE . It turns out one can use essentially the same proof as in [13] to prove that the mechanism $\text{ASPE}(p', \delta')$ has revenue at least $a_1(\mu) \cdot \text{CORE} - a_2(\mu) \cdot \text{POSTREV} - a_3(\mu) \cdot (n+m) \cdot \epsilon$ where $a_1(\mu)$, $a_2(\mu)$ and $a_3(\mu)$ are functions that map μ to positive numbers. In other words, we can replace $\text{ASPE}(p^*, \delta^*)$ with $\text{ASPE}(p', \delta')$ and still obtain a constant factor approximation. \square

D.1 Missing Proofs from Section 6.1

We formalize the first step of our algorithm in the following lemma.

Lemma 19. *For any $B > 0$, $\epsilon > 0$, $\eta \in [0, 1]$ and $\mu \in [0, \frac{1}{4}]$, suppose we take $K = O\left(\frac{\log \frac{1}{\eta} + \log n + m \log \frac{B}{\epsilon}}{\mu^2}\right)$ samples $t^{(1)}, \dots, t^{(K)}$ from D . For any collection of prices $\{p_j\}_{j \in [m]}$ in the B -bounded ϵ -net, define the entry fee $\delta_i^{(p)}(S)$ of bidder i for set S under $\{p_j\}_{j \in [m]}$ to be the median of $u_i(t_i^{(1)}, S), \dots, u_i(t_i^{(K)}, S)$, where $u_i(t_i, S) = \max_{S^* \subseteq S} v_i(t_i, S^*) - \sum_{j \in S^*} p_j$. Then with probability $1 - \eta$, for any collection of prices $\{p_j\}_{j \in [m]}$ in the B -bounded ϵ -net, $\{\delta_i^{(p)}(\cdot)\}_{i \in [n]}$ is a collection of μ -balanced entry fee functions.*

Proof. For any fixed $\{p_j\}_{j \in [m]}$, fixed bidder i and fixed set S , it is easy to argue that the probability for $\Pr_{t_i \sim D_i}[u_i(t_i, S) \geq \delta_i^{(p)}(S)]$ to be larger than $\frac{1}{2} + \mu$ or smaller than $\frac{1}{2} - \mu$ is at most $2\exp(-2K\mu^2)$ due to the Chernoff bound. Next, we take a union bound over all $\{p_j\}_{j \in [m]}$ in the ϵ -net, all bidders and all possible subsets of $[m]$, so the probability that for any collection of prices $\{p_j\}_{j \in [m]}$ in the ϵ -net $\{\delta_i^{(p)}(\cdot)\}_{i \in [n]}$ is a collection of μ -balanced entry fee functions is at least $1 - 2\exp(-2K\mu^2) \cdot (\frac{B}{\epsilon})^m \cdot 2^m \cdot n$. If we take K to be at least $\frac{\log \frac{1}{\eta} + \log n + m \log \frac{B}{\epsilon}}{\mu^2}$, the success probability is at least $1 - \eta$. \square

Next, we formalize the second step of our learning algorithm.

Lemma 20. For any $B \geq 2G$, $\epsilon, \epsilon' > 0$, $\eta \in [0, 1]$ and $\mu \in [0, \frac{1}{4}]$, suppose for every collection of prices $\{p_j\}_{j \in [m]}$ in the B -bounded ϵ -net, $\{\delta_i^{(p)}(\cdot)\}_{i \in [n]}$ is a collection of μ -balanced entry fee functions. We use \mathcal{S} to denote the set that contains $\text{ASPE}(p, \delta^{(p)})$ for every p in the B -bounded ϵ -net. If we take $K = O\left(\frac{\log \frac{1}{\eta} + m \log \frac{B}{\epsilon}}{\epsilon'^2}\right)$ samples $t^{(1)}, \dots, t^{(K)}$ from D and let $\text{ASPE}(p', \delta^{(p')})$ be the mechanism that has the highest revenue in \mathcal{S} . Then with probability at least $1 - \eta$, the better of $\text{ASPE}(p', \delta^{(p')})$ and the best RSPM achieves revenue at least $\frac{\text{OPT}}{\mathcal{C}_1(\mu)} - \mathcal{C}_2(\mu) \cdot (m + n) \cdot \epsilon - 2mnB \cdot \epsilon'$.

Proof. For any $\{p_j\}_{j \in [m]}$ in the ϵ -net, define $\text{REV}(p)$ to be the expected revenue of $\text{ASPE}(p, \delta^{(p)})$ and $\widehat{\text{REV}}(p)$ be the average revenue of $\text{ASPE}(p, \delta^{(p)})$ among the K samples. First, we argue that $\widehat{\text{REV}}(p)$ is a random variable that lies between $[0, mnB]$. The revenue from selling the items can be at most mB as there are only m items and $p_j \leq B$ for all $j \in [m]$. How about the entry fee? For any bidder i ,

$$\Pr_{t_i \sim D_i} [v_i(t_i, [m]) \geq mG] \leq \sum_{j \in [m]} \Pr_{t_{ij} \sim D_{ij}} [V_i(t_{ij}) \geq G] \leq \frac{m}{5 \max\{m, n\}} \leq \frac{1}{5}.$$

The first inequality is because $v_i(t_i, \cdot)$ is a subadditive function for every type $t_i \in T_i$, so for $v_i(t_i, [m])$ to be greater than mG , there must exist a item j such that $V_i(t_{ij}) \geq G$. The second inequality follows from the definition of G in Theorem 8.

If there exists a set $S \subseteq [m]$ such that $\delta_i^{(p)}(S) > mG$, we have

$$\Pr_{t_i \sim D_i} [v_i(t_i, [m]) \geq mG] \geq \Pr_{t_i \sim D_i} [v_i(t_i, [m]) \geq \delta_i^{(p)}(S)] \geq \frac{1}{2} - \mu \geq \frac{1}{4}.$$

Contradiction. Note that the second inequality is because $\delta_i^{(p)}(\cdot)$ is μ -balanced. Hence, the entry fee is always upper bounded by mG and $\widehat{\text{REV}}(p)$ is at most $mnG + mB \leq mnB$. Also, notice that the expectation of $\widehat{\text{REV}}(p)$ is exactly $\text{REV}(p)$. By the Chernoff bound,

$$\Pr \left[\left| \text{REV}(p) - \widehat{\text{REV}}(p) \right| \leq mnB \cdot \epsilon' \right] \geq 1 - 2 \exp(-2K \cdot \epsilon'^2)$$

for any fixed $\{p_j\}_{j \in [m]}$. By the union bound, the probability that for all $\{p_j\}_{j \in [m]}$ in the ϵ -net

$$\left| \text{REV}(p) - \widehat{\text{REV}}(p) \right| \leq mnB \cdot \epsilon'$$

is at least $1 - 2 \exp(-2K \cdot \epsilon'^2) \cdot \left(\frac{B}{\epsilon}\right)^m$, which is lower bounded by $1 - \eta$ due to our choice of K . When this happens, the expected revenue of $\text{ASPE}(p', \delta^{(p')})$ is at most $2mnB \cdot \epsilon'$ less than the highest expected revenue achievable by any of these mechanisms, because

$$\text{REV}(p') \geq \widehat{\text{REV}}(p') - mnB \cdot \epsilon' \geq \widehat{\text{REV}}(p) - mnB \cdot \epsilon' \geq \text{REV}(p) - 2mnB \cdot \epsilon'$$

for any p in the ϵ -net. Combining this inequality with Corollary 2 completes our proof. \square

Note that Lemma 19 and 20 hold for all distributions D . The reason we require D to be bounded or regular is because without these restrictions, we do not know how to approximate the best RSPM. In the following Theorem, we combine Lemma 19, 20 and Theorem 11 to obtain the sample complexity of our learning algorithm for bounded distributions.

Theorem 15. When all bidders' valuations are XOS over independent items and the random variable $V_i(t_{ij})$ is supported on $[0, H]$ for any bidder i and any item j , with $O\left(\left(\frac{mn}{\xi}\right)^2 \cdot \left(m \cdot \log \frac{m+n}{\xi} + \log \frac{1}{\delta}\right)\right)$ samples from D , 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 $\frac{\text{OPT}}{c} - \xi \cdot H$ for some absolute constant $c > 1$.

Proof. With $O\left(\left(\frac{1}{\xi}\right)^2 \left(m^2 n \log n \log \frac{1}{\xi} + \log \frac{1}{\delta}\right)\right)$ samples, we can obtain an RSPM whose revenue is at least $\frac{1}{24}$ of the revenue of the best RSPM minus $\frac{\xi}{2} \cdot H$ with probability $1 - \delta/2$ according to Theorem 11. Let μ be some fixed constant in $[0, \frac{1}{4}]$, $B = 2H$, $\epsilon = \frac{\xi \cdot H}{6\mathcal{C}_2(\mu)(m+n)}$ and $\epsilon' = \frac{\xi}{12mn}$. According to Lemma 19, given $O\left(\log \frac{1}{\delta} + \log n + m \log \frac{m+n}{\xi}\right)$ samples, we can construct an entry fee function for each price vector in the B -bounded ϵ -net, such that all these entry fee functions are μ -balanced with probability at least $1 - \delta/4$. According to Lemma 20, we can learn an ASPE with $O\left(\left(\frac{mn}{\xi}\right)^2 \cdot \left(m \cdot \log \frac{m+n}{\xi} + \log \frac{1}{\delta}\right)\right)$ fresh samples from D , such that the better of the ASPE we learned and the best RSPM has revenue of at least $\frac{\text{OPT}}{\mathcal{C}_1(\mu)} - \frac{\xi}{2} \cdot H$ with probability $1 - \delta/4$. Combining the statements above, we can learn with probability $1 - \delta$ a mechanism whose revenue is at least $\frac{\text{OPT}}{c} - \xi \cdot H$ with $O\left(\left(\frac{mn}{\xi}\right)^2 \cdot \left(m \cdot \log \frac{m+n}{\xi} + \log \frac{1}{\delta}\right)\right)$ samples. \square

In the next Theorem, we combine Lemma 19, 20 and Theorem 12 to obtain the sample complexity of our learning algorithm for regular distributions.

Theorem 16. *When all bidders' valuations are XOS over independent items and the random variable $V_i(t_{ij})$ is regular for each item $j \in [m]$ and bidder $i \in [n]$, with $O\left(Z^2 m^2 n^2 \cdot \left(m \cdot \log(m+n) + \log \frac{1}{\delta}\right)\right)$ ($Z = \max\{m, n\}$) samples from D , 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 $\frac{\text{OPT}}{c}$ for some absolute constant $c > 1$.*

Proof. According to Theorem 12, we can learn with probability $1 - \delta/2$ a randomized RSPM whose revenue is at least $\frac{1}{33}$ of the optimal RSPM with $O\left(Z^2 m^2 n^2 \cdot \log \frac{nm}{\delta}\right)$ samples. Next, we learn an ASPE with high revenue. With $O\left(Z^2 \cdot \log \frac{nm}{\delta}\right)$ samples from each D_{ij} , we can estimate W_{ij} such that

$$\Pr_{t_{ij} \sim D_{ij}} [V_i(t_{ij}) \geq W_{ij}] \in \left[\frac{1}{6Z}, \frac{1}{5Z}\right]$$

with probability $1 - \frac{\delta}{4nm}$. By the union bound, the probability that all W_{ij} satisfy the requirement is at least $1 - \frac{\delta}{4}$. So with probability at least $1 - \frac{\delta}{4}$, $W_{ij} \geq G_{ij}$ for all $i \in [n]$ and $j \in [m]$.

Let $B = 2 \cdot \max_{i,j} W_{ij}$, μ be some fixed constant in $[0, \frac{1}{4}]$, $\epsilon = \frac{\xi \cdot B}{\mathcal{C}_2(\mu)Z(m+n)}$ and $\epsilon' = \frac{\xi}{2mnZ}$ for some small constant ξ , which will be specified later. We know that given $O\left(\log \frac{1}{\delta} + \log n + m \log(m+n)\right)$ samples, we can construct μ -balanced entry fee functions for all price vectors in the B -bounded ϵ -net with probability $1 - \delta/8$ due to Lemma 19. According to Lemma 20, we can learn an ASPE with

$$O\left(Z^2 m^2 n^2 \cdot \left(m \cdot \log(m+n) + \log \frac{1}{\delta}\right)\right)$$

fresh samples from D , such that the better of the ASPE we learned and the best RSPM has revenue of at least $\frac{\text{OPT}}{\mathcal{C}_1(\mu)} - \frac{2\xi \cdot B}{Z}$ with probability $1 - \delta/8$. Note that there exists a bidder i and an item j such that $W_{ij} = B/2$, so $\text{OPT} \geq \frac{B}{2} \cdot \frac{1}{6Z}$ and for sufficiently small ξ , $\frac{\text{OPT}}{\mathcal{C}_1(\mu)} - \frac{2\xi \cdot B}{Z} \geq \frac{\text{OPT}}{2\mathcal{C}_1(\mu)}$. Combining the statements above, we can learn with probability $1 - \delta$ a mechanism whose revenue is at least $\frac{\text{OPT}}{c}$ for some absolute constant c with $O\left(Z^2 m^2 n^2 \cdot \left(m \cdot \log(m+n) + \log \frac{1}{\delta}\right)\right)$ samples. \square

E Learning Algorithms for Symmetric Bidders

E.1 An Upper Bound of the Optimal Revenue for Symmetric Bidders

In this section, we introduce an upper bound to OPT based on duality [13], which is crucial for us to prove the approximation ratios of our learning algorithms. We first fix some notation. Let POSTREV be the highest revenue obtainable by any RSPM. As the bidders are symmetric, we drop the subscript i when there is no confusion. In particular, we use $V(t_{ij})$ to denote bidder i 's value for winning item j if her private information for item j is t_{ij} , and $v(t_i, S)$ to denote bidder i 's value for set S when her type is t_i . We use D_j to denote the distribution of the private information about item j . Let $\sigma_{iS}(t)$ be the interim probability for bidder i to receive exactly set $S \subseteq [m]$ when her type is t .

In [13], an upper bound of the optimal revenue is derived using duality theory. Their upper bound applies to asymmetric bidders with valuations that are subadditive over independent items. When the bidders are symmetric, we can simplify their upper bound. First, we need the definition of b -balanced thresholds.

Definition 7 (b -balanced Thresholds). *For any constant $b \in (0, 1)$, a collection of positive real numbers $\{\beta_j\}_{j \in [m]}$ is b -balanced if for all $i \in [n]$ and $j \in [m]$, $\Pr_{t_{ij} \sim D_j} [V(t_{ij}) \geq \beta_j] \in [\frac{b}{n}, \frac{b}{n-1}]$.*

Note that when bidders are asymmetric, b -balanced thresholds are not guaranteed to exist, as there may not exist any β_j that satisfies $\Pr_{t_{ij} \sim D_j} [V(t_{ij}) \geq \beta_j] \in [\frac{b}{n}, \frac{b}{n-1}]$ for all bidder i simultaneously. Next, we define the $\text{CORE}_\eta(\beta)$ which will be crucial for upper bounding the optimal revenue.¹⁰

Definition 8 (CORE). *Given any collection of thresholds $\{\beta_j\}_{j \in [n]}$ and a nonnegative constant $\eta \leq \frac{1}{4}$,*

- *if $\sum_{j \in [m]} \Pr_{t_j \sim D_j} [V(t_j) \geq \beta_j] \leq \frac{1}{2} - \eta$, let $c_\eta(\beta)$ be 0;*
- *otherwise, let $c_\eta(\beta)$ be a nonnegative number such that $\sum_{j \in [m]} \Pr_{t_j \sim D_j} [V(t_j) \geq \beta_j + c_\eta(\beta)] \in [\frac{1}{2} - \eta, \frac{1}{2}]$.*

For every type t , let $\mathcal{C}_\eta(t) = \{j \mid V(t_j) < \beta_j + c_\eta(\beta)\}$. Then,

$$\text{CORE}_\eta(\beta) = \max_{\sigma \in P(D)} \sum_{i \in [n]} \sum_{t_i \in T_i} f(t_i) \cdot \sum_{S \subseteq [m]} \sigma_{iS}(t_i) \cdot v(t_i, S \cap \mathcal{C}_\eta(t_i)),$$

where $P(D)$ is the set of all feasible interim allocation rules. That is, $\text{CORE}_\eta(\beta)$ is the maximum welfare a mechanism can extract out of the allocation of items whose individual value for the bidder they are allocated to is lower than the adjusted thresholds.

It was shown in [13] that every collection of thresholds induces an upper bound to the optimal revenue. In particular, for any choice of thresholds $\{\beta_j\}_{j \in [m]}$ and η ¹¹, the revenue $\text{REV}(M)$ of any BIC mechanism M is upper bounded by

$$2 \cdot \text{SINGLE}(M, \beta) + 4 \cdot \text{TAIL}_\eta(M, \beta) + 4 \cdot \text{CORE}_\eta(M, \beta) \quad (\text{Adapted from Theorem 2 in [13]}).$$

These terms depend on the choice of $\{\beta_j\}_{j \in [m]}$, η as well as the mechanism M . We refer interested readers to [13] for the definitions of these terms. To obtain a benchmark/upper bound of the optimal revenue, one can simply replace the above expression with

$$2 \cdot \max_M \text{SINGLE}(M, \beta) + 4 \cdot \max_M \text{TAIL}_\eta(M, \beta) + 4 \cdot \max_M \text{CORE}_\eta(M, \beta).$$

¹⁰For readers that are familiar with the definition of the CORE in [13], $\text{CORE}_\eta(\beta)$ is essentially the same term but adapted for symmetric bidders.

¹¹In [13], the thresholds are allowed to depend on the identity of the bidder. More specifically, for any $i \in [n]$ and $j \in [m]$, there is an associated threshold β_{ij} . Their upper bound applies to asymmetric thresholds as well. Indeed, when the bidders are asymmetric, their upper bound is induced by a set of asymmetric thresholds. As we only discuss symmetric bidders in this section, we focus on symmetric thresholds for simplicity. Regarding η , Cai and Zhao only considered the case when $\eta = 0$, but their analysis can be easily modified to accommodate any $\eta \leq 1/4$. See Theorem 17 for the modified upper bound.

It is not hard to see that this benchmark may be impossible to approximate for certain choices of the thresholds. Just imagine the case when the thresholds are extremely high, then $\max_M \text{CORE}_\eta(M, \beta)$ becomes the optimal social welfare which can be arbitrarily large comparing to the optimal revenue. What Cai and Zhao [13] showed was that when the thresholds are b -balanced, this upper bound can indeed be approximated by the revenue of an RSPM and an ASPE. From now on, we only consider b -balanced thresholds.

Using results in [13], we can further simplify the benchmark. In particular, $\max_M \text{SINGLE}(M, \beta)$ is less than $6 \cdot \text{POSTREV}$ for all choices of $\{\beta_j\}_{j \in [n]}$ and $\max_M \text{TAIL}_\eta(M, \beta)$ is less than $\frac{2}{1-b} \cdot \text{POSTREV}$ for any choice of η and b -balanced thresholds $\{\beta_j\}_{j \in [n]}$. Moreover, $\max_M \text{CORE}_\eta(M, \beta) \leq \text{CORE}_\eta(\beta)$. Combining the inequalities above, we obtain the following Theorem.

Theorem 17 (Adapted from [13]). *When the bidders are symmetric and have valuations that are subadditive over independent items, for any constant $b \in (0, 1)$, $\eta \leq \frac{1}{4}$ and a collection of b -balanced thresholds $\{\beta_j\}_{j \in [m]}$,*

$$\text{OPT} \leq \left(12 + \frac{8}{1-b}\right) \cdot \text{POSTREV} + 4 \cdot \text{CORE}_\eta(\beta).$$

E.2 Symmetric Bidders with XOS Valuations

In this section, we show how to learn in polynomial time an approximately optimal mechanism for symmetric bidders with XOS valuations given sample access to the distributions. According to Theorem 17, we only need to learn a mechanism that approximates POSTREV and $\text{CORE}_\eta(\beta)$. From Section 5.1, we know how to approximate POSTREV in polynomial time, so we focus on learning a mechanism whose revenue approximates $\text{CORE}_\eta(\beta)$.

First, we need a crucial property about XOS valuations.

Lemma 21 (Supporting Prices [23]). *If $v(t, \cdot)$ is an XOS function, for any subset $S \subseteq [m]$ there exists a collection of supporting prices $\{\theta_j^S(t)\}_{j \in S}$ for $v(t, S)$ such that*

1. $v(t, S') \geq \sum_{j \in S'} \theta_j^S(t)$ for all $S' \subseteq S$ and
2. $\sum_{j \in S} \theta_j^S(t) = v(t, S)$.

Let $v'(t_i, S) = v(t_i, S \cap \mathcal{C}_\eta(t_i))$ and \mathcal{F}_i be the distribution of the valuation $v'(t_i, S)$. As the bidders are symmetric, $\mathcal{F}_i = \mathcal{F}_{i'}$ for any i and i' . The $\text{CORE}_\eta(\beta)$ is exactly the maximum expected social welfare if every bidder i 's valuation is drawn independently from \mathcal{F}_i . Cai and Zhao [13] showed how to use an ASPE to approximate this term. In the next Lemma, we construct the prices used in their ASPE and show its relation to $\text{CORE}_\eta(\beta)$.

Lemma 22. (Adapted from [13]) *Let every bidder i 's valuation be $v'(t_i, S) = v(t_i, S \cap \mathcal{C}_\eta(t_i))$ when her type is t_i and σ^* be a symmetric allocation that achieves α -fraction of the optimal social welfare with respect to $v'(\cdot, \cdot)$. For every item $j \in [m]$, let*

$$Q_{\eta,j} = \frac{1}{2} \cdot \sum_{i \in [n]} \sum_{t_i \in T_i} f(t_i) \cdot \sum_{S: j \in S} \sigma_{iS}^*(t_i) \cdot \theta_j^{S \cap \mathcal{C}_\eta(t_i)}(t_i),$$

where $\{\theta_j^{S \cap \mathcal{C}_\eta(t_i)}(t_i)\}_{j \in S \cap \mathcal{C}_\eta(t_i)}$ is the supporting prices for $v(t_i, S \cap \mathcal{C}_\eta(t_i))$. Let

$$u^*(t, S) = \max_{S^* \subseteq S} v(t, S^*) - \sum_{j \in S^*} Q_{\eta,j}$$

be a bidder's utility for the set of items S when her type is t . We define $\delta^*(S)$ to be the median of the random variable $u^*(t, S)$ (with $t \sim \times_{j \in [m]} D_j$) for any set $S \subseteq [m]$. The revenue of ASPE($\{Q_{\eta,j}\}_{j \in [m]}, \delta^*$) is at least

$$\frac{\alpha \cdot \text{CORE}_\eta(\beta)}{2} - \mathcal{C}(b, \eta) \cdot \text{POSTREV},$$

where $\mathcal{C}(b, \eta)$ is a function that only depends on b and η .

Proof. We can essentially use the same proof in [13] to prove that the expected revenue of the ASPE is at least

$$\sum_{j \in [m]} Q_{\eta,j} - \mathcal{C}(b, \eta) \cdot \text{POSTREV}.$$

For readers that are familiar with that proof, the only thing we need to make sure is that our choice of σ^* and $\{\beta_j\}_{j \in [m]}$ satisfy Lemma 5 in [13]. Since σ^* is symmetric and $\{\beta_j\}_{j \in [m]}$ is b -balanced, for all bidder i and item j

$$\sum_{k \neq i} \Pr_{t_{kj} \sim D_j} [V(t_{kj}) \geq \beta_j] \leq \frac{b}{n-1} \cdot (n-1) = b,$$

and

$$\Pr_{t_{ij} \sim D_j} [V(t_{ij}) \geq \beta_j] / b \geq 1/n \geq \cdot \sum_{t_i \in T_i} f_i(t_i) \cdot \sum_{S: j \in S} \sigma_{iS}^*(t_i).$$

Next, we argue $\sum_{j \in [m]} Q_{\eta,j} \geq \frac{\alpha \cdot \text{CORE}_\eta(\beta)}{2}$. Observe that

$$\sum_{j \in [m]} Q_{\eta,j} = \frac{1}{2} \cdot \sum_{i \in [n]} \sum_{t_i \in T_i} f(t_i) \cdot \sum_S \sigma_{iS}^*(t_i) \cdot v'(t_i, S) \geq \alpha \cdot \text{CORE}_\eta(\beta).$$

The last inequality is because $\text{CORE}_\eta(\beta)$ is the maximum social welfare under $v'(\cdot, \cdot)$ and σ^* achieves α fraction of that. \square

Lemma 23. For any $\epsilon > 0$ and $\mu \in [0, \frac{1}{4}]$, let $\{Q_j\}_{j \in [m]}$ be a collection of prices such that $|Q_j - Q_{\eta,j}| \leq \epsilon$ for all $j \in [m]$. Let $\delta(S)$ be the entry fee function such that $\Pr_{t \sim \times_{j \in [m]} D_j} [u(t, S) \geq \delta(S)] \in [1/2 - \mu, 1/2 + \mu]$ for any set $S \subseteq [m]$, where $u(t, S) = \max_{S^* \subseteq S} v(t, S^*) - \sum_{j \in S^*} Q_j$. Then, the ASPE(Q, δ) achieves at least $\frac{\alpha \cdot \text{CORE}_\eta(\beta)}{\mathcal{B}_1(\mu)} - \mathcal{B}_2(b, \eta, \mu) \cdot \text{POSTREV} - \mathcal{B}_3(\mu) \cdot (m+n) \cdot \epsilon$ revenue when bidders' valuations are XOS over independent item. Both $\mathcal{B}_1(\mu)$ and $\mathcal{B}_3(\mu)$ are functions that only depend on μ and $\mathcal{B}_2(b, \eta, \mu)$ is a function that only depends on μ, b and η .

Proof. It turns out the proof in [13] is robust enough to accommodate the error ϵ and μ . We can prove the claim by following essentially the same analysis as in [13]. We do not include the details here. \square

E.2.1 Leaning the ASPE in Polynomial Time

We first show how to learn a collection of b -balanced thresholds and the corresponding $c_\eta(\beta)$.

Lemma 24. For any positive constant $b < 1$ and $\eta \leq \frac{1}{4}$, there is a polynomial time algorithm that computes a collection of b -balanced thresholds $\{\beta_j\}_{j \in [m]}$ and $c_\eta(\beta)$ with probability $1 - \delta$ using $O(m^2 n^4 \log \frac{m}{\delta})$ samples from distribution $\times_{j \in [m]} D_j$.

Proof. Given $K = O(m^2 n^4 (\log m + \log \frac{1}{\delta}))$ samples $t_j^{(1)}, \dots, t_j^{(K)}$ from distribution D_j , we construct \mathcal{F}_j as the uniform distribution over $V(t_j^{(1)}), \dots, V(t_j^{(K)})$. According to the DKW Theorem [25], with probability at least $1 - \delta/m$,

$$\left| \Pr_{t_j \sim D_j} [V(t_j) \geq x] - \Pr_{v_j \sim \mathcal{F}_j} [v_j \geq x] \right| \leq \frac{1}{c \cdot mn^2} \text{ for all } x \quad (6)$$

where c is a constant that will be specified later. From now on, we assume that Inequality (6) holds for every j , which happens with probability $1 - \delta$.

As $1/K \leq \frac{b}{3n^2} \leq \frac{b}{n-1} - \frac{b}{3n^2} - \frac{b}{n} - \frac{b}{3n^2}$, there must exist a sample $t_j^{(\ell)}$ such that $\Pr_{v_j \sim \mathcal{F}_j} [v_j \geq V(t_j^{(\ell)})] \in [\frac{b}{n} + \frac{b}{3n^2}, \frac{b}{n-1} - \frac{b}{3n^2}]$. Let $\beta_j = V(t_j^{(\ell)})$. Note that

$$\Pr_{t_j \sim D_j} [V(t_j) \geq \beta_j] \in \left[\Pr_{v_j \sim \mathcal{F}_j} [v_j \geq \beta_j] - \frac{1}{c \cdot mn^2}, \Pr_{v_j \sim \mathcal{F}_j} [v_j \geq \beta_j] + \frac{1}{c \cdot mn^2} \right].$$

If c is less than $\frac{b}{3}$, $\Pr_{t_j \sim D_j} [V(t_j) \geq \beta_j] \in [\frac{b}{n}, \frac{b}{n-1}]$. Thus, β_j is b -balanced for all item j .

Next, we argue how to learn $c_\eta(\beta)$. If $\sum_{j \in [m]} \Pr_{v_j \sim \mathcal{F}_j} [v_j \geq \beta_j] \leq \frac{1}{2} - \frac{\eta}{2}$, let $c_\eta(\beta) = 0$. This is a valid choice, as $\sum_j \Pr_{t_j \sim D_j} [V(t_j) \geq \beta_j]$ is at most $\frac{1}{2} - \frac{\eta}{2} + \frac{1}{cn^2} \leq \frac{1}{2}$ according to inequality (6). Suppose $\sum_{j \in [m]} \Pr_{v_j \sim \mathcal{F}_j} [v_j \geq \beta_j] > \frac{1}{2} - \frac{\eta}{2}$, as $m/K < \frac{\eta}{4}$, there must exist some item $k \in [m]$ and a sample $V(t_k^{(\ell)}) \geq \beta_k$ such that $\sum_{j \in [m]} \Pr_{v_j \sim \mathcal{F}_j} [v_j \geq \beta_j + V(t_k^{(\ell)}) - \beta_k] \in [\frac{1}{2} - \frac{\eta}{4}, \frac{1}{2} - \frac{\eta}{2}]$. Let $c_\eta(\beta) = V(t_k^{(\ell)}) - \beta_k$. According to inequality (6),

$$\sum_{j \in [m]} \Pr_{t_j \sim D_j} [V(t_j) \geq \beta_j + c_\eta(\beta)] \in \left[\frac{1}{2} - \frac{\eta}{4} - \frac{1}{cn^2}, \frac{1}{2} - \frac{\eta}{2} + \frac{1}{cn^2} \right].$$

For sufficiently large c , $\sum_j \Pr_{t_j \sim D_j} [V(t_j) \geq \beta_j + c_\eta(\beta)] \in [\frac{1}{2} - \eta, \frac{1}{2}]$.

Finding each β_j takes $O(K \log K)$ time and finding the $c_\eta(\beta)$ takes $O(mK)$ time. So we can learn in polynomial time a collection of b -balanced thresholds $\{\beta_j\}_{j \in [m]}$ and $c_\eta(\beta)$ with probability $1 - \delta$ using $O(m^2 n^4 \log \frac{m}{\delta})$ samples. \square

Next, we show how to learn the prices of the ASPE. As showed by Feige [27], there exists a polynomial time algorithm that achieves $1 - \frac{1}{e}$ fraction of the optimal social welfare when bidders have XOS valuations. We let σ^* be the interim allocation rule induced by Feige's algorithm and estimate the prices by running Feige's algorithm on sampled valuation profiles. To run Feige's algorithm, we need a demand oracle for bidder's valuations. In the following Lemma, we argue that $v'(t, \cdot)$ is an XOS function for any type t , and given a value (or demand, XOS) oracle for $v(t, \cdot)$, we can construct in polynomial time the corresponding oracle for $v'(t, \cdot)$. First, we define these oracles formally.

Definition 9. We consider the following three oracles for a bidder's valuation function $v(t, \cdot)$:

- **Value oracle:** takes a set $S \subseteq [m]$ as the input and returns $v(t, S)$.
- **Demand oracle:** takes a collection of prices $\{p_j\}_{j \in [m]}$ as an input and returns the favorite set under these prices, that is, $S^* \in \operatorname{argmax}_{S \subseteq [m]} v(t, S) - \sum_{j \in S} p_j$.
- **XOS oracle** (only when $v(t, \cdot)$ is XOS): takes a set $S \subseteq [m]$ as the input and returns the supporting prices $\{\theta_j^S(t)\}_{j \in S}$ for $v(t, S)$.

Lemma 25. *Given a collection of thresholds $\{\beta_j\}_{j \in [m]}$ and $c_\eta(\beta)$. For any set $S \subseteq [m]$, let $v'(t, S) = v(t, S \cap \mathcal{C}_\eta(t))$. If $v(t, \cdot)$ is an XOS function, $v'(t, \cdot)$ is also an XOS function. Given a value (or demand, XOS) oracle for $v(t, \cdot)$, we can construct in polynomial time a value (or demand, XOS) oracle for $v'(t, \cdot)$.*

Proof. If $v(t, \cdot)$ is an XOS function, $v(t, \cdot)$ can be represented as the max of a collection of additive functions. Observe that if we change the values for items in $\mathcal{C}_\eta(t)$ to 0 in each of these additive functions, $v'(t, \cdot)$ equals to the max of this new collection of additive functions. Hence, $v'(t, \cdot)$ is also an XOS function.

If we are given a value oracle for $v(t, \cdot)$, it is straightforward to construct a value oracle for $v'(t, \cdot)$. If we are given a demand oracle for $v(t, \cdot)$, here is how to construct a demand oracle for $v'(t, \cdot)$. For every queried price vector $\{p_j\}_{j \in [m]}$, we change the price for each item outside $\mathcal{C}_\eta(t)$ to $2v(t, [m])$ and keep the prices for the items in $\mathcal{C}_\eta(t)$. Let this new price vector be p' . We query the demand oracle of $v(t, \cdot)$ on p' . The output set should also be the demand set for $v'(t, \cdot)$ under prices p , as the bidder can only afford items in $\mathcal{C}_\eta(t)$ and $v'(t, S) = v(t, S)$ for any set $S \subseteq \mathcal{C}_\eta(t)$. Finally, we consider the XOS oracle. For any set S , let $\{\theta_j^{S \cap \mathcal{C}_\eta(t)}(t)\}_{j \in S \cap \mathcal{C}_\eta(t)}$ be the supporting prices for $v(t, S \cap \mathcal{C}_\eta(t))$. Let $\gamma_j^S(t) = \theta_j^{S \cap \mathcal{C}_\eta(t)}$ for all item j in $\mathcal{C}_\eta(t) \cap S$ and $\gamma_j^S(t) = 0$ for all item j in $S - \mathcal{C}_\eta(t)$. According to the definition of $v'(t, \cdot)$, $\{\gamma_j^S(t)\}_{j \in S}$ is the supporting price for $v'(t, S)$. So given an XOS oracle for $v(t, \cdot)$, we can compute the supporting price of any set S for $v'(t, \cdot)$ in polynomial time. \square

Lemma 25 shows that $v'(t, \cdot)$ is also an XOS function for any type t and with access to a demand oracle for $v(t, \cdot)$ we can construct a demand oracle for $v'(t, \cdot)$ in polynomial time. So we can indeed run Feige's algorithm on v' . In the next Lemma, we show how to learn a collection of prices $\{Q_j\}_{j \in [m]}$ and entry fee function $\delta(\cdot, \cdot)$ such that the corresponding ASPE has high revenue.

Lemma 26. *Given a collection of b -balanced thresholds $\{\beta_j\}_{j \in [m]}$ and $c_\eta(\beta)$, and access to value, demand and XOS oracles for valuation $v(t, \cdot)$ for every type t , there is a polynomial time algorithm that learns an ASPE($\{Q_j\}_{j \in [m]}, \delta$) whose revenue is at least $\frac{\text{CORE}_\eta(\beta)}{\mathcal{K}_1} - g(b, \eta) \cdot \text{POSTREV} - \mathcal{K}_2 \cdot \xi \cdot \text{OPT}$ with probability at least $1 - \zeta$ using $O\left(n^3(m+n)^2 \log \frac{m}{\zeta}\right)$ samples from $\times_{j \in [m]} D_j$, where \mathcal{K}_1 and \mathcal{K}_2 are positive absolute constants, and $g(b, \eta)$ is a function that only depends on b and η .*

Proof. According to Lemma 25, we can construct value, demand and XOS oracles for valuation $v'(t, \cdot)$ given access to the corresponding oracles for $v(t, \cdot)$. We use $\{\gamma_j^S(t)\}_{j \in S}$ to denote the output of the XOS oracle for $v'(t, \cdot)$ on set S . In particular, $\gamma_j^S(t) = 0$ for all $j \in S - \mathcal{C}_\eta(t)$ and $\gamma_j^S(t) = \theta_j^{S \cap \mathcal{C}_\eta(t)}(t)$ for all $j \in S \cap \mathcal{C}_\eta(t)$, where $\{\theta_j^{S \cap \mathcal{C}_\eta(t)}(t)\}_{j \in S \cap \mathcal{C}_\eta(t)}$ is the supporting prices for $v(t, S \cap \mathcal{C}_\eta(t))$. Let $\mathcal{A}(t)$ be the allocation computed by Feige's algorithm on the valuation profile $(v'(t_1, \cdot), \dots, v'(t_n, \cdot))$, where $\mathcal{A}_i(t)$ denotes the set of items that bidder i receives. Let σ^* be the interim allocation rule induced by $\mathcal{A}(\cdot)$ when bidders types are all drawn from $\times_{j \in [m]} D_j$ independently. That is, $\sigma_{iS}^*(t_i) = \Pr_{t_{-i}}[\mathcal{A}_i(t) = S]$. We use the same definition for $Q_{\eta, j}$ as in Lemma 22. In other words, $Q_{\eta, j}$ is the contribution of item j to the social welfare under allocation rule σ^* , so we can rewrite it as

$$\frac{1}{2} \cdot \mathbb{E}_t \left[\sum_{i \in [n]} \mathbb{1}[j \in \mathcal{A}_i(t)] \cdot \gamma_j^{\mathcal{A}_i(t)}(t_i) \right].$$

Let $\mathbf{t}^{(1)}, \dots, \mathbf{t}^{(K)}$ be K sampled type profiles, and $q^{(\ell)} = \frac{1}{2} \sum_{i \in [n]} \mathbb{1}[j \in \mathcal{A}_i(\mathbf{t}^{(\ell)})] \cdot \gamma_j^{\mathcal{A}_i(\mathbf{t}^{(\ell)})}(t_i^{(\ell)})$. We set Q_j to be $\frac{1}{K} \cdot \sum_{\ell \in [K]} q^{(\ell)}$. Since $\gamma_j^S(t) \leq \beta_j + c_\eta(\beta)$ for any j, S and t , $Q_j \leq \beta_j + c_\eta(\beta)$. By the Chernoff bound,

$$\Pr[|Q_j - Q_{\eta, j}| \leq \epsilon \cdot (\beta_j + c_\eta(\beta))] \geq 1 - 2 \exp(-2K \cdot \epsilon^2).$$

As $\{\beta_j\}_{j \in [m]}$ is a collection of b -balanced thresholds, we can obtain revenue $\beta_j \cdot \frac{b}{n}$ by only selling item j to one bidder at price β_j . Hence, $\beta_j \leq \frac{n \cdot \text{OPT}}{b}$. Now, consider a posted price mechanism that sells item j at price $\beta_j + c_\eta(\beta)$. A single bidder will purchase at least one item with probability at least $\sum_j \Pr_{t_j \sim D_j} [V(t_j) \geq \beta_j + c_\eta(\beta)]$ which is no less than $\frac{1}{2} - \eta$ if $c_\eta(\beta) > 0$. Hence, the revenue of this mechanism is at least $c_\eta(\beta) \cdot (\frac{1}{2} - \eta)$. As $\eta \leq \frac{1}{4}$, $c_\eta(\beta) \leq 4\text{OPT}$. If we let $\epsilon = \frac{\xi}{(m+n) \cdot (n/b+4)}$ for some small constant ξ which will be specified later and $K = \frac{\log \frac{4m}{\xi}}{2\epsilon^2}$, we have $\Pr \left[|Q_j - Q_{\eta,j}| \leq \frac{\xi}{m+n} \cdot \text{OPT} \right] \geq 1 - \frac{\zeta}{2m}$. In other words, with $O \left(n^3(m+n)^2 \log \frac{m}{\xi} \right)$ samples from $\times_{j \in [m]} D_j$ (as each t^ℓ costs n samples), we can learn in polynomial time a collection of prices $\{Q_j\}_{j \in [m]}$ such that $|Q_j - Q_{\eta,j}| \leq \frac{\xi}{m+n} \cdot \text{OPT}$ for all item j with probability $1 - \zeta/2$.

Next, we consider the entry fee function. We use essentially the same argument as in Lemma 19. Suppose we take L samples $t^{(1)}, \dots, t^{(L)}$ from $\times_{j \in [m]} D_j$. Define the entry fee $\delta(S)$ for set S under $\{Q_j\}_{j \in [m]}$ to be the median of $u(t^{(1)}, S), \dots, u(t^{(L)}, S)$, where $u(t, S) = \max_{S^* \subseteq S} v(t, S^*) - \sum_{j \in S^*} p_j$. Given any constant $\mu \in [0, 1/4]$, for any fixed set S , it is easy to argue that the probability for $\Pr_{t \sim \times_{j \in [m]} D_j} [u(t, S) \geq \delta(S)]$ to be larger than $\frac{1}{2} + \mu$ or less than $\frac{1}{2} - \mu$ is at most $2 \exp(-2L\mu^2)$ due to the Chernoff bound. If we let L to be $a \cdot \frac{m + \log 1/\zeta}{\mu^2}$ for a sufficiently large constant a , the probability that $\delta(\cdot)$ is a μ -balanced entry fee function is at least $1 - \zeta/2$ by the union bound.

Hence, with $O \left(n^3(m+n)^2 \log \frac{m}{\xi} \right)$ samples from $\times_{j \in [m]} D_j$, we can compute in polynomial time a collection of prices $\{Q_j\}_{j \in [m]}$ and a entry fee function $\delta(\cdot)$ such that the revenue of the ASPE($\{Q_j\}_{j \in [m]}, \delta(\cdot)$) is at least $\frac{(1-1/e) \cdot \text{CORE}_\eta(\beta)}{\mathcal{B}_1(\mu)} - \mathcal{B}_2(b, \eta, \mu) \cdot \text{POSTREV} - \xi \cdot \mathcal{B}_3(\mu) \cdot \text{OPT}$ with probability $1 - \zeta$ due to Lemma 23. Our claim follows by fixing the value of μ to be some constant. \square

Theorem 18. *For symmetric bidders with valuations that are XOS over independent items,*

1. *when $V(t_j)$ is upper bounded by H for any $j \in [m]$ and any t_j , with*

$$O \left((n^5 + m^2 n^4) \cdot \log \frac{m}{\delta} + \left(\frac{1}{\epsilon} \right)^2 \left(m^2 n \log n \log \frac{1}{\epsilon} + \log \frac{1}{\delta} \right) \right)$$

samples from $\times_{j \in [m]} D_j$, we can learn in polynomial time with probability $1 - \delta$ a mechanism whose revenue is at least $c_1 \cdot \text{OPT} - \epsilon \cdot H$ for some absolute constant c_1 ;

2. *when the distribution of random variable $V(t_j)$ with $t_j \sim D_j$ is regular for all item $j \in [m]$, with*

$$O \left(n^5 \cdot \log \frac{m}{\delta} + \max\{m, n\}^2 m^2 n^2 \cdot \log \frac{nm}{\delta} \right)$$

samples from $\times_{j \in [m]} D_j$, we can learn in polynomial time with probability $1 - \delta$ a mechanism whose revenue is at least $c_2 \cdot \text{OPT}$ for some absolute constant c_2 .

Proof of Theorem 18: Combining Lemma 24, Lemma 26 and Theorem 17, we know how to compute in polynomial time an ASPE whose revenue is at least $a_1 \cdot \text{OPT} - a_2 \cdot \text{POSTREV}$ with probability $1 - \delta/2$ for some absolute constant a_1, a_2 , and we only need $O \left((n^5 + m^2 n^4) \cdot \log \frac{m}{\delta} \right)$ samples from $\times_{j \in [m]} D_j$. When the distributions are bounded, we can learn in polynomial time an RSPM whose revenue is at least $\frac{\text{POSTREV}}{144} - \xi H$ with probability $1 - \delta/2$ using $O \left(\left(\frac{1}{\xi} \right)^2 \left(m^2 n \log n \log \frac{1}{\epsilon} + \log \frac{1}{\delta} \right) \right)$ samples (Theorem 11). By choosing the ratio between ξ and ϵ to be the right constant, we can show the first part of our claim. When $V(t_j)$ is a regular random variable for every item j , we can learn in polynomial time an RSPM whose revenue is at least $\frac{\text{POSTREV}}{33}$

with probability $1 - \delta/2$ using $O(\max\{m, n\}^2 m^2 n^2 \cdot \log \frac{nm}{\delta})$ samples (Theorem 12). Therefore, we can learn a mechanism in polynomial time such that with probability $1 - \delta$ whose revenue is at least a constant fraction of the OPT. This proves the second part of our claim. \square

E.3 Symmetric Bidders with Subadditive Valuations

In this section, we argue that if the bidders are symmetric and $m = O(n)$, there exists a collection of b -balanced thresholds $\{\beta_j\}_{j \in [m]}$ for a fixed constant b , such that POSTREV is within a constant fraction of the benchmark. Note that this argument only applies to symmetric bidders, as b -balanced thresholds may not even exist for asymmetric bidders.

We set $\eta = 0$ for this section and drop the subscript η when there is no confusion. We show how to upper bound $\text{CORE}(\beta)$ with POSTREV by choosing a particular collection of b -balanced thresholds. Let $Z = \max\{m, n\}$ and $b = \frac{n}{3Z}$. It is not hard to see that $\frac{n}{3Z}$ -balanced thresholds exist, as we can choose β_j such that $\Pr_{t_j \sim D_j}[V(t_j) \geq \beta_j] = \frac{1}{3Z}$.

Lemma 27. *Let $\{\beta_j\}_{j \in [m]}$ be a collection of $\frac{n}{3Z}$ -balanced thresholds, then $\text{CORE}(\beta) \leq \sum_{j \in [m]} \beta_j$.*

Proof. As $\{\beta_j\}_{j \in [m]}$ are $\frac{n}{3Z}$ -balanced, $\Pr_{t_j \sim D_j}[V(t_j) \geq \beta_j] \leq \frac{n}{(n-1) \cdot 3Z} \leq \frac{1}{2Z}$. Therefore,

$$\sum_{j \in [m]} \Pr_{t_j \sim D_j}[V(t_j) \geq \beta_j] \leq \frac{1}{2},$$

so $c(\beta) = 0$. Next, we upper bound $\text{CORE}(\beta)$ by $\sum_{j \in [m]} \beta_j$.

$$\begin{aligned} \text{CORE}(\beta) &= \max_{\sigma \in P(D)} \sum_{i \in [n]} \sum_{t_i \in T_i} f(t_i) \cdot \sum_{S \subseteq [m]} \sigma_{iS}(t_i) \cdot v(t_i, S \cap \mathcal{C}(t_i)) \\ &\leq \max_{\sigma \in P(D)} \sum_{i \in [n]} \sum_{t_i \in T_i} f(t_i) \cdot \left(\sum_{S \subseteq [m]} \sigma_{iS}(t_i) \cdot \sum_{j \in S} \beta_j \right) \\ &= \max_{\sigma \in P(D)} \sum_{i \in [n]} \sum_{j \in [m]} \beta_j \cdot \left(\sum_{t_i \in T_i} f(t_i) \cdot \sum_{S: j \in S} \sigma_{iS}(t_i) \right) \\ &= \max_{\sigma \in P(D)} \sum_j \beta_j \cdot \left(\sum_{i \in [n]} \sum_{t_i \in T_i} f_i(t_i) \cdot \sum_{S: j \in S} \sigma_{iS}(t_i) \right) \\ &\leq \sum_j \beta_j \end{aligned}$$

The first inequality is because $v(t_i, \cdot)$ is a subadditive function, so

$$v(t_i, S \cap \mathcal{C}_i(t_i)) \leq \sum_{j \in S \cap \mathcal{C}_i(t_i)} V(t_{ij}) \leq \sum_{j \in S \cap \mathcal{C}_i(t_i)} \beta_j \leq \sum_{j \in S} \beta_j.$$

The last inequality is because $\sum_{i \in [n]} \sum_{t_i \in T_i} f_i(t_i) \cdot \sum_{S: j \in S} \sigma_{iS}(t_i) \leq 1$ is the ex-ante probability for bidder i to receive item j , and for any feasible interim allocation σ , the sum of all bidders' ex-ante probabilities for receiving item j should not exceed 1. \square

In the following Lemma, we demonstrate that $\sum_{j \in [m]} \beta_j$ is upper bounded by $\frac{9Z}{n} \cdot \text{POSTREV}$.

Lemma 28. Let $\{\beta_j\}_{j \in [m]}$ be a collection of $\frac{n}{3Z}$ -balanced thresholds, $\text{POSTREV} \geq \frac{n}{9Z} \cdot \sum_{j \in [m]} \beta_j$.

Proof. Let us consider an RSPM where the price for selling item j to bidder i is β_j . Bidder i purchases item j if that is the only item she can afford and no one else can afford item j . As $\{\beta_j\}_{j \in [m]}$ are $\frac{n}{3Z}$ -balanced, the probability that no one else can afford item j is at least

$$\left(1 - \sum_{k \neq i} \Pr_{t_{kj} \sim D_j} [V(t_{kj}) \geq \beta_j]\right) \geq \left(1 - \frac{n}{3Z}\right) \geq \frac{2}{3}.$$

Also, the probability that i cannot afford any item other than j is at least

$$\left(1 - \sum_{\ell \neq j} \Pr_{t_{i\ell} \sim D_\ell} [V(t_{i\ell}) \geq \beta_\ell]\right) \geq 1 - \frac{n(m-1)}{3Z(n-1)} \geq \frac{1}{2}.$$

Therefore, bidder i purchases item j with probability at least $\frac{1}{3} \Pr_{t_{ij} \sim D_j} [V(t_{ij}) \geq \beta_j] \geq \frac{1}{9Z}$. Whenever this event happens, it contributes β_j to the revenue. So the total revenue is at least $\sum_j \sum_i \frac{\beta_j}{9Z} = \frac{n}{9Z} \cdot \sum_j \beta_j$. \square

Combining Theorem 17, Lemma 27 and 28, we obtain the following Theorem.

Theorem 19. For symmetric bidders with valuations that are subadditive over independent items,

$$\text{OPT} \leq \left(24 + \frac{36 \max\{n, m\}}{n}\right) \cdot \text{POSTREV}.$$

Proof. Combining Lemma 27 and 28, we have $\text{POSTREV} \geq \frac{n}{9 \max\{n, m\}} \cdot \text{CORE}(\beta)$ if $\{\beta_j\}_{j \in [m]}$ is a collection of $\frac{n}{3 \max\{n, m\}}$ -balanced thresholds. By setting b to be $\frac{n}{3 \max\{n, m\}}$ and replacing $\text{CORE}(\beta)$ with $\frac{9 \max\{n, m\}}{n}$ · POSTREV in Theorem 17, we have

$$\text{OPT} \leq \left(12 + \frac{8}{1 - \frac{n}{3 \max\{n, m\}}} + \frac{36 \max\{n, m\}}{n}\right) \cdot \text{POSTREV}.$$

As $\frac{n}{3 \max\{n, m\}} \leq 1/3$,

$$\text{OPT} \leq \left(24 + \frac{36 \max\{n, m\}}{n}\right) \cdot \text{POSTREV}.$$

\square

E.3.1 Learning an Approximately Optimal Mechanism for Symmetric Subadditive Bidders

With Theorem 19, we only need to learn a mechanism that approximates the optimal revenue obtainable by any RSPM. The next Lemma connects RSPMs with SPMs in an induced unit-demand setting.

Lemma 29. Consider n symmetric bidders whose types are drawn independently from $\times_{j=1}^m D_j$. Let \mathcal{F}_j be the distribution for random variable $V(t_j)$ where $t_j \sim D_j$. We define an induced unit-demand setting with n symmetric unit-demand bidders whose values for item j are drawn independently from \mathcal{F}_j . For any collection of prices $\{p_{ij}\}_{i \in [n], j \in [m]}$, the revenue of the RSPM with these prices in the original setting is exactly the same as the revenue of the SPM with these prices in the induced unit-demand setting.

Proof. As in an RSPM bidders can purchase at most one item, bidders behave exactly the same as in the induced unit-demand setting. Since the prices in the SPM and RSPM are the same, bidders purchase exactly the same items. Hence, the revenue is the same. \square

Corollary 3. For symmetric bidders with valuations that are subadditive over independent items,

$$OPT^{UD} \geq \Omega\left(\frac{n}{Z}\right) \cdot OPT,$$

where $Z = \max\{m, n\}$ and OPT^{UD} is the optimal revenue for the induced unit-demand setting.

Proof. Combine Theorem 19 and Lemma 29. □

Lemma 29 implies that learning an approximately optimal RSPM is equivalent as learning an approximately optimal SPM in the induced unit-demand setting. Next, we apply our results in Section 5.1 to the induced unit-demand setting to learn an RSPM that approximates the optimal revenue in the original setting.

In the next Theorem, we show that even though the bidders' valuations could be complex set functions, e.g., submodular, XOS and subadditive, as long as $m = O(n)$, the approximate distributions for the bidders' values for winning any *single item* provides sufficient information to learn an approximately optimal mechanism.

Theorem 20. For symmetric bidders with valuations that are subadditive over independent items, let \mathcal{F}_j be the distribution of $V(t_j)$ where $t_j \sim D_j$. If \mathcal{F}_j is supported on $[0, H]$ for all $j \in [m]$, given distributions $\hat{\mathcal{F}}_j$ where $\|\hat{\mathcal{F}}_j - \mathcal{F}_j\|_K \leq \epsilon$ for all $j \in [m]$, there is a polynomial time algorithm that constructs a randomized RSPM whose revenue under the true distribution D is at least

$$\left(\frac{1}{4} - (n + m) \cdot \epsilon\right) \cdot \left(\Omega\left(\frac{n}{\max\{m, n\}}\right) \cdot OPT - 2\epsilon \cdot mnH\right).$$

Proof. Let $Z = \max\{m, n\}$. According to Corollary 3, $OPT^{UD} = \Omega\left(\frac{n}{Z}\right) \cdot OPT$. Since $\|\hat{\mathcal{F}}_j - \mathcal{F}_j\|_K \leq \epsilon$ for all $j \in [m]$, we can learn a randomized SPM in the induced unit-demand setting whose revenue under the true distribution is at least $\left(\frac{1}{4} - (n + m) \cdot \epsilon\right) \cdot \left(\frac{OPT^{UD}}{8} - 2\epsilon \cdot mnH\right)$ based on Theorem 10. By Lemma 29, we can construct an RSPM with the same collection of (randomized) prices and achieve revenue

$$\left(\frac{1}{4} - (n + m) \cdot \epsilon\right) \cdot \left(\Omega\left(\frac{n}{Z}\right) \cdot OPT - 2\epsilon \cdot mnH\right)$$

in the original setting. □

If we are given sample access to bounded distributions, we show in the following Theorem that a polynomial number of samples suffices to learn an approximately optimal mechanism, when $m = O(n)$.

Theorem 21. For symmetric bidders with valuations that are subadditive over independent items, let \mathcal{F}_j be the distribution of $V(t_j)$ where $t_j \sim D_j$. If \mathcal{F}_j is supported on $[0, H]$ for all $j \in [m]$, there is a polynomial time algorithm that learns an RSPM whose revenue is $\Omega\left(\frac{n}{\max\{m, n\}}\right) \cdot OPT - \epsilon H$ with probability $1 - \delta$ using

$$O\left(\left(\frac{1}{\epsilon}\right)^2 \cdot \left(m^2 n \log n \log \frac{1}{\epsilon} + \log \frac{1}{\delta}\right)\right)$$

samples.

Proof. According to Corollary 3, $OPT^{UD} = \Omega\left(\frac{n}{\max\{m, n\}}\right) \cdot OPT$. Due to Theorem 11,

$$O\left(\left(\frac{1}{\epsilon}\right)^2 \cdot \left(m^2 n \log n \log \frac{1}{\epsilon} + \log \frac{1}{\delta}\right)\right)$$

samples suffices to learn in polynomial time with probability $1 - \delta$ an SPM with revenue at least $\Omega(OPT^{UD}) - \epsilon \cdot H$ for the induced unit-demand setting. By Lemma 29, we can construct an RSPM with the same collection of prices and achieve revenue $\Omega\left(\frac{n}{\max\{m, n\}}\right) \cdot OPT - \epsilon H$ in the original setting. □

Finally, if the distribution of the random variable $V(t_j)$ with $t_j \sim D_j$ is regular for all item $j \in [m]$, we prove in the next theorem that there exists a prior-independent mechanism that achieves a constant fraction of the optimal revenue if $m = O(n)$. Note that approximately optimal prior-independent mechanisms for symmetric unit-demand bidders are known due to the work by Devanur et al. [21] and Roughgarden et al. [39]. Our result is obtained by combining Theorem 19 and the afore-mentioned prior independent mechanisms.

Theorem 22. *For symmetric bidders with valuations that are subadditive over independent items, let \mathcal{F}_j be the distribution of $V(t_j)$ where $t_j \sim D_j$. If \mathcal{F}_j is regular for all $j \in [m]$, there is a prior-independent mechanism with revenue at least $\Omega\left(\frac{n}{\max\{m,n\}}\right) \cdot \text{OPT}$. Moreover, the mechanism can be implemented efficiently.*

Proof. The mechanism in [21] or [39] provides an approximately optimal prior-independent mechanism in the induced unit-demand setting. Let us use M to denote this mechanism. Suppose we restrict every bidder to purchase at most one item in the original setting and then run mechanism M . The expected revenue is the same as M 's expected revenue in the induced setting. Since M 's expected revenue is $\Omega(\text{OPT}^{UD})$ and $\text{OPT}^{UD} = \Omega\left(\frac{n}{\max\{m,n\}}\right) \cdot \text{OPT}$, the mechanism we constructed has revenue $\Omega\left(\frac{n}{\max\{m,n\}}\right) \cdot \text{OPT}$. Since M can be implemented efficiently for unit-demand bidders, our mechanism can also be implemented efficiently. \square

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. [1](#), [1](#), [5.1](#)
- [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. [1](#)
- [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. [1](#)
- [4] Susan Athey and Philip Haile. Nonparametric approaches to auctions. *Handbook of econometrics*, 6(Part A):3847–3965, 2007. [1](#)
- [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. [1](#)
- [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. [1](#)
- [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. [1](#), [1](#), [5.1](#), [12](#)
- [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. [1](#)
- [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. [1](#)
- [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. [1](#)
- [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. [1](#), [1](#), [5.1](#), [5.2](#), [6](#), [7](#), [B.1](#), [B.1](#), [B.1](#)
- [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. [1](#)
- [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. [1](#), [1](#), [5](#), [5](#), [6](#), [8](#), [7](#), [D](#), [E.1](#), [E.1](#), [10](#), [11](#), [17](#), [E.2](#), [22](#), [E.2](#), [E.2](#)
- [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. [1](#)

- [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. [1](#), [1](#), [5.1](#), [5.1](#), [B.1](#), [B.1](#), [8](#), [B.1](#)
- [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. [1](#)
- [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. [1](#)
- [18] Constantinos Daskalakis. Multi-item auctions defying intuition? *ACM SIGecom Exchanges*, 14(1):41–75, 2015. [1](#)
- [19] Constantinos Daskalakis, Alan Deckelbaum, and Christos Tzamos. Strong duality for a multiple-good monopolist. *Econometrica*, 2017. [1](#)
- [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. [1](#)
- [21] Nikhil R. Devanur, Jason D. Hartline, Anna R. Karlin, and C. Thach Nguyen. Prior-independent multi-parameter mechanism design. In *Internet and Network Economics - 7th International Workshop, WINE 2011, Singapore, December 11-14, 2011. Proceedings*, pages 122–133, 2011. [E.3.1](#), [E.3.1](#)
- [22] 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. [1](#)
- [23] Shahar Dobzinski, Noam Nisan, and Michael Schapira. Approximation algorithms for combinatorial auctions with complement-free bidders. In *STOC*, pages 610–618, 2005. [21](#)
- [24] Shaddin Dughmi, Li Han, and Noam Nisan. Sampling and representation complexity of revenue maximization. In *International Conference on Web and Internet Economics (WINE)*, 2014. [1](#), [2](#)
- [25] 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. [1](#), [4](#), [B.3](#), [E.2.1](#)
- [26] 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. [1](#)
- [27] Uriel Feige. On maximizing welfare when utility functions are subadditive. *SIAM Journal on Computing*, 39(1):122–142, 2009. [E.2.1](#)
- [28] 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. [\(document\)](#), [1](#), [1](#), [3](#), [5.2](#), [5.2](#), [C.1](#)
- [29] Yannai A. Gonczarowski and Noam Nisan. Efficient empirical revenue maximization in single-parameter auction environments. *CoRR*, abs/1610.09976, 2016. [1](#)
- [30] Emmanuel Guerre, Isabelle Perrigne, and Quang Vuong. Optimal nonparametric estimation of first-price auctions. *Econometrica*, 68(3):525–574, 2000. [1](#)

- [31] 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. [1](#)
- [32] Robert Kleinberg and S. Matthew Weinberg. Matroid Prophet Inequalities. In *the 44th Annual ACM Symposium on Theory of Computing (STOC)*, 2012. [5.1](#)
- [33] 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. [1](#)
- [34] Jamie Morgenstern and Tim Roughgarden. The pseudo-dimension of near-optimal auctions. *arXiv preprint arXiv:1506.03684*, 2015. [1](#)
- [35] 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. [\(document\)](#), [1](#), [1](#), [3](#), [5.1](#), [5.2](#), [11](#), [C.1](#)
- [36] Roger B. Myerson. Optimal Auction Design. *Mathematics of Operations Research*, 6(1):58–73, 1981. [1](#)
- [37] Harry J Paarsch and Han Hong. An introduction to the structural econometrics of auction data. *MIT Press Books*, 1, 2006. [1](#)
- [38] 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. [1](#)
- [39] Tim Roughgarden, Inbal Talgam-Cohen, and Qiqi Yan. Supply-limiting mechanisms. In *13th ACM Conference on Electronic Commerce (EC)*, 2012. [E.3.1](#), [E.3.1](#)
- [40] Aviad Rubinfeld 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. [1](#), [1](#)
- [41] Ester Samuel-Cahn. Comparison of threshold stop rules and maximum for independent nonnegative random variables. *Ann. Probab.*, 12(4):1213–1216, 11 1984. [C.1](#)
- [42] Andrew Chi-Chih Yao. An n-to-1 bidder reduction for multi-item auctions and its applications. In *SODA*, 2015. [1](#), [1](#), [5.2](#), [7](#)