

The Sample Complexity of Up-to- ϵ Multi-dimensional Revenue Maximization

YANNAI A. GONCZAROWSKI, Microsoft Research

S. MATTHEW WEINBERG, Princeton University

15

We consider the sample complexity of revenue maximization for multiple bidders in unrestricted multi-dimensional settings. Specifically, we study the standard model of n additive bidders whose values for m heterogeneous items are drawn independently. For any such instance and any $\epsilon > 0$, we show that it is possible to learn an ϵ -Bayesian Incentive Compatible auction whose expected revenue is within ϵ of the optimal ϵ -BIC auction from only polynomially many samples.

Our fully nonparametric approach is based on ideas that hold quite generally and completely sidestep the difficulty of characterizing optimal (or near-optimal) auctions for these settings. Therefore, our results easily extend to general multi-dimensional settings, including valuations that are not necessarily even *subadditive*, and arbitrary allocation constraints. For the cases of a single bidder and many goods, or a single parameter (good) and many bidders, our analysis yields exact incentive compatibility (and for the latter also computational efficiency). Although the single-parameter case is already well understood, our corollary for this case extends slightly the state of the art.

CCS Concepts: • **Theory of computation** → **Design and analysis of algorithms**; **Sample complexity and generalization bounds**; **Algorithmic game theory**; **Algorithmic mechanism design**; **Convergence and learning in games**; **Computational pricing and auctions**;

Additional Key Words and Phrases: Algorithmic game theory, algorithmic mechanism design, auctions, sample complexity, generalization bounds, PAC learning, approximate revenue maximization, multi-dimensional auctions

ACM Reference format:

Yannai A. Gonczarowski and S. Matthew Weinberg. 2021. The Sample Complexity of Up-to- ϵ Multi-dimensional Revenue Maximization. *J. ACM* 68, 3, Article 15 (March 2021), 28 pages.

<https://doi.org/10.1145/3439722>

An extended abstract of this work appeared in FOCS 2018 [Gonczarowski and Weinberg 2018].

This research was conducted while Gonczarowski was at the Hebrew University of Jerusalem and Microsoft Research. Gonczarowski was supported in part by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities. His work was supported in part by ISF grant 1435/14 administered by the Israeli Academy of Sciences, by Israel-USA Binational Science Foundation (BSF) grant number 2014389, and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 740282). Weinberg was supported by NSF CCF-1717899 and by NSF CAREER award CCF-1942497.

Authors' addresses: Y. A. Gonczarowski, Microsoft Research, One Memorial Drive, Cambridge, MA 02139; email: yannai@gonch.name; S. M. Weinberg, Princeton University, 35 Olden Street, Princeton, NJ 08540; email: smweinberg@princeton.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

0004-5411/2021/03-ART15 \$15.00

<https://doi.org/10.1145/3439722>

1 INTRODUCTION

A fundamental question at the heart of the literature on mechanism design is that of *revenue maximization* by a single seller who is offering for sale any number of goods to any number of (potential) bidders. In the classic economic literature, this problem is studied in a *Bayesian* setting: The seller has prior knowledge of (often, independent) distributions from which the valuation of each bidder for each good is drawn and wishes to devise a truthful mechanism that maximizes her revenue in expectation over these prior distributions. Over the past few years, numerous works at the interface of economics and computation are now studying a more demanding model: that of *mechanism design from samples*. In this model, rather than possessing complete knowledge of the distributions from which the bidders' values for the various items are drawn, the seller more realistically only has access to samples from these distributions (e.g., past data). The goal in this setting is to learn with high probability an auction with good revenue guarantees given polynomially many (in the parameters of the problem) samples.

Revenue maximization from samples is commonly thought of as a “next step” beyond Bayesian revenue maximization. That is, existing works so far in this context take settings for which simple auctions in the related Bayesian problem are already well understood and prove that these simple auctions can be learned efficiently via samples (up to an ϵ loss, which will always be lost when optimizing from samples). For example, in single-parameter settings, the seminal work of Myerson [1981] completely characterizes a simple and optimal auction in the Bayesian setting, and works such as Cole and Roughgarden [2014], Morgenstern and Roughgarden [2015], Devanur et al. [2016], Hartline and Taggart [2019], Roughgarden and Schrijvers [2016], and Gonczarowski and Nisan [2017] prove that these simple mechanisms or variants thereof can be learned with polynomially many samples. Similarly, in multi-parameter settings with independent items, works of Chawla et al. [2007], Chawla et al. [2010], Chawla et al. [2015], Hart and Nisan [2012], Babaioff et al. [2014], Rubinstein and Weinberg [2015], Yao [2015], Cai et al. [2016], Chawla and Miller [2016], and Cai and Zhao [2017] prove that simple mechanisms achieve constant-factor approximations in rich multi-dimensional settings, and works of Morgenstern and Roughgarden [2016], Balcan et al. [2016], Balcan et al. [2018], Cai and Daskalakis [2017], and Syrgkanis [2017] prove that simple mechanisms with these guarantees can be learned with polynomially many samples. These analyses rely on a delicate understanding of the structure and/or inherent dimensionality of auctions that give such revenue guarantees to show how to learn such an auction without overfitting the samples.

It is therefore unsurprising that the problem of learning an up-to- ϵ revenue-maximizing multi-item auction from samples has not been previously studied, since the structure/dimensionality of optimal (precisely or up-to- ϵ) multi-item auctions is not understood even when there is only one bidder, and even with independent items. Such auctions are known to be extremely complex, suffering from properties such as randomization [Thanassoulis 2004], uncountable menu complexity [Daskalakis et al. 2013], and non-monotonicity [Hart and Reny 2015]. These domains provably lack the natural starting point of all previous works: a structured/low-dimensional mechanism in the Bayesian setting to learn via samples.

In this article, we show that despite these challenges, up-to- ϵ optimal multi-item auctions can be learned from polynomially many samples from the underlying bidder-item distributions. More formally, in a setting with n bidders and m items where the value of each bidder i for each item j is drawn independently from a distribution $V_{i,j}$ supported on $[0, H]$ for some H that is known to the seller, we show that polynomially many samples suffice for learning, with probability at least $1 - \delta$, an m -item almost-truthful auction that maximizes the expected revenue among all possible m -item almost-truthful auctions up to an additive ϵ . An auction is Bayesian Incentive Compatible (BIC) if it is in every bidder's interest to bid truthfully, given that all other bidders do so as well.

THEOREM 1.1 (MAIN RESULT—INFORMAL VERSION OF THEOREM 3.1). *For n bidders with independent values for m items supported on $[0, H]$, for every $\epsilon, \delta > 0$ and for every $\eta \leq \text{poly}(n, m, H, \epsilon)$, the sample complexity of learning, w.p. $1 - \delta$, an η -BIC auction that maximizes revenue (among all η -BIC auctions) up to an additive ϵ is $\text{poly}(n, m, H, 1/\epsilon, 1/\eta, \log 1/\delta)$.*

The above theorem is informal mostly because we have not specified exactly how bidders value bundles of items. Essentially, the bidders may have arbitrary (i.e., not necessarily additive, not even necessarily subadditive) valuations subject to some Lipschitz condition (i.e., changing the value of bidder i for item j by ϵ only changes the bidder's value for any outcome by at most $L\epsilon$ for some absolute constant L).¹ We defer a formal definition to Section 2 but only note here that commonly studied classes of valuations such as additive, unit-demand or “additive subject to constraints” with independent items (as well as several natural subadditive and superadditive valuation classes and even certain valuation classes with externalities) all satisfy our definition with Lipschitz constant 1.

The main challenge in proving our result for $m > 1$ items is noted above: The structure of (up-to- ϵ) optimal mechanisms for such settings is not understood, even for additive valuations. In particular, there is no known low-dimensional class of mechanisms that is guaranteed to contain an (up-to- ϵ) optimal mechanism for any product distribution, thus barring the use of many learning-theoretic arguments. Our result relies on a succinct structured argument, allowing us to reduce revenue maximization from samples to related problems of revenue maximization from given discrete distributions.

As the corresponding Bayesian question remains open (i.e., whether one can find, given the distributions explicitly, an up-to- ϵ optimal mechanism in poly-time), our result is of course information-theoretic: It shows that polynomially many samples suffice for a computationally unbounded seller but provides no computationally efficient learning algorithm. Concretely, the algorithm that we give uses as a black box an oracle that can perform (optimal or almost-optimal) multi-item Bayesian revenue maximization given (the full description of) finite prior distributions.²

1.1 Brief Overview of Techniques

Most prior works (for single- as well as multi-dimensional settings) take the following approach: First, define a class C_ϵ of auctions as a function of ϵ . Second, prove that, for all possible distributions \mathcal{D} , the class C_ϵ contains an up-to- ϵ optimal mechanism for \mathcal{D} . Finally, prove that the best-in-class (up to ϵ) for C_ϵ can be learned with polynomially many samples. In prior works, ingenuity is required for both steps: C_ϵ is explicitly defined, proved to contain up-to- ϵ optimal auctions, and proved to have some low-dimensional structure allowing efficient learnability.

Our approach indeed follows this rough outline, with two notable simplifying exceptions. First is our approach to defining C_ϵ . Here, we first define $C_{\epsilon, S}$ be the space of all auctions that are optimal for an empirical distribution over S -many $\text{poly}(\epsilon)$ -rounded samples (that is, optimal for any discrete product distribution where each marginal is (a) only supported on multiples of $\text{poly}(\epsilon)$ and (b) uniform over a multiset of size S). This fully nonparametric approach stands in contrast with the popular existing approach of taking C_ϵ to be some parametric family of auctions. Unlike in such existing approaches where C_ϵ is fixed, in our approach the set $C_{\epsilon, S}$ grows with the number of samples S . Nonetheless, we show that the rate of its growth is moderate enough so that there exists a “sweet-spot” number of samples $S = \text{poly}(1/\epsilon)$ such that, on the one hand, $C_{\epsilon, S}$ contains an auction that is up-to- ϵ optimal for the “true distribution” \mathcal{D} and, on the other hand, the

¹Our results in fact hold even more generally: to arbitrary outcomes that do not even correspond to bundles of items. See Section 2 for the full details.

²Note, however, that if computationally efficient algorithms were to be developed for up-to- ϵ optimal mechanisms given an explicit prior, then our approach would immediately become computationally efficient as well.

best-in-class from $C_{\varepsilon, S}$ can be learned from S samples. So in the language of prior work, one could say that we set $C_\varepsilon = C_{\varepsilon, S}$ for this $S = \text{poly}(1/\varepsilon)$.

To show that C_ε does in fact contain, for all distributions \mathcal{D} , an auction that is up-to- ε optimal for \mathcal{D} , we simply take enough samples to guarantee uniform convergence (of the revenue) over C_ε and *additionally* the optimal auction for \mathcal{D} . It is far from obvious why this should suffice, as the optimal auction for \mathcal{D} is not an element of C_ε , nor even of the same format.³ Still existing tools (namely, the ε -BIC-to-BIC reduction of Daskalakis and Weinberg [2012] and Rubinstein and Weinberg [2015]), when applied correctly, suffice to complete the argument. This part of our proof is conceptually much simpler than prior works (despite making use of a big technical hammer), as this approach holds quite generally and is robust due to not requiring the analysis of any specific class of mechanisms.

Second, our argument that the best-in-class can be in fact learned (up to ε) with $S = \text{poly}(\varepsilon)$ samples is simply a counting argument, and does not require any notions of a learning dimension. This is indeed in the spirit of some recent *single-dimensional results*, however in those results the counting argument is highly dependent on the structure of auctions in C_ε . As discussed above, such dependence is damning for multi-dimensional settings where such structure provably does not exist. Again, the proof does require some hammers (notably, arguments originally developed for *reduced forms* via samples in Cai et al. [2012], and a concentration inequality of Babichenko et al. [2017], see also Devanur et al. [2016]), but they are applied in a fairly transparent manner.

The above approach should help explain how we are able to extend far beyond prior works, which relied on a detailed analysis of specific structured mechanisms: The key tools we use are applicable quite generally, whereas the specific mechanisms analyzed in prior work are only known to maintain guarantees only in restricted settings. For example, Theorem 1.1 already constitutes the first up-to- ε optimal-mechanism learning result for any multi-parameter setting *even if it held only for additive valuations (and one bidder)*. But the approach is so general that extending it to arbitrary Lipschitz valuations with independent items is simply a matter of updating notation.

1.2 Applications and Extensions

Specialized to a single-bidder setting, our construction in fact yields exact truthfulness (more on that in Section 6), showing that an ε -optimal mechanism can be found for a single bidder with independent item values (with Lipschitz valuations) using only polynomially many samples. This should be contrasted with a result of Dughmi et al. [2014], which shows that achieving this is *not* possible for correlated distributions, even for a bidder with additive valuations.

COROLLARY 1.2 (SINGLE BIDDER—INFORMAL VERSION OF THEOREM 3.2). *For one bidder with independent values for m items supported on $[0, H]$, for every $\varepsilon, \delta > 0$, the sample complexity of learning, w.p. $1 - \delta$, an IC auction that maximizes revenue (among all IC auctions) up to an additive ε is $\text{poly}(m, H, 1/\varepsilon, \log 1/\delta)$.*

Specialized to single-dimensional settings, our analysis once again yields a strengthened result, both in giving exact Dominant Strategy Incentive Compatibility (DSIC)⁴ and in providing a computationally efficient algorithm (due to known efficient solutions [Elkind 2007; Myerson 1981] to single-parameter revenue maximization from given discrete distributions):

³That is, the optimal auction for \mathcal{D} is a mapping from the support of \mathcal{D} to outcomes, whereas the elements of C_ε are mappings from a finite space to outcomes. Furthermore, notions of Bayesian incentive compatibility for \mathcal{D} do not imply nor are implied by these notions for the various discrete distributions defining C_ε .

⁴A mechanism is DSIC if it is a dominant strategy for each bidder to bid truthfully. For a single good, Myerson [1981] shows that the maximal revenue attainable by a BIC mechanism and by a DSIC mechanism is the same.

COROLLARY 1.3 (SINGLE-PARAMETER—INFORMAL VERSION OF THEOREM 3.3). *For n single-parameter bidders with independent values in $[0, H]$, for every $\epsilon, \delta > 0$, the sample complexity of **efficiently learning**, w.p. $1 - \delta$, a DSIC auction that maximizes revenue (among all DSIC auctions) up to an additive ϵ is $\text{poly}(n, H, 1/\epsilon, \log 1/\delta)$.*

Corollary 1.3 nicely complements the existing literature on single-parameter sample complexity in the following ways. First, our algorithm/analysis immediately follows as a special case of Theorem 1.1 (without referencing structural results about optimal single-parameter auctions), so it is in some sense more principled. Second, our analysis holds even for arbitrary constraints on the allocations (putting it in the same class as the state-of-the-art⁵ single-parameter results [Gonczarowski and Nisan 2017; Hartline and Taggart 2019], and even slightly beyond⁶).

Finally, portions of our approach are specific to Bayesian Incentive Compatible auctions (versus Dominant Strategy Incentive Compatible auctions), but portions are not. We are therefore able to use the same techniques to conclude similar, albeit qualitatively weaker, results for ϵ -DSIC auctions in Theorem A.2. See Appendix A for further details.

1.3 Related Work and Brief Discussion

Two active lines of work are directly related to the present article. First are papers that study rich multi-dimensional settings and aim to show that mechanisms with good approximation guarantees can be learned with few samples, such as Morgenstern and Roughgarden [2016], Balcan et al. [2016], Balcan et al. [2018], Cai and Daskalakis [2017], and Syrgkanis [2017]. The main approach in each of these works is to show that specific classes of structured mechanisms (e.g., classes that are known to allow for constant-factor revenue maximization) are inherently low-dimensional with respect to some notion of dimensionality. Our results are stronger than these in some regards and weaker in others. More specifically, our results are stronger in the sense that with comparably many samples, our mechanisms guarantee an up-to- ϵ approximation to the optimal mechanism instead of a constant-factor. Our results are weaker in the sense that our learning algorithms are information-theoretic (do not run in poly-time), and our mechanisms are not “as simple.” As discussed earlier, both weaknesses are necessary to possibly surpass the constant-factor barrier (at least, barring the resolution of major open questions, such as a computationally efficient up-to- ϵ approximation even when all distributions are explicitly known. Again, note that should this question be resolved affirmatively, our results would immediately become computationally efficient as well).

Most related to our work, at least in terms of techniques, is the rich line of works on *single-dimensional settings* [Cole and Roughgarden 2014; Devanur et al. 2016; Dhangwatnotai et al. 2015; Gonczarowski and Nisan 2017; Hartline and Taggart 2019; Huang et al. 2015; Morgenstern and Roughgarden 2015; Roughgarden and Schrijvers 2016]. These works show that up-to- ϵ optimal mechanisms can be learned in richer and richer settings. In comparison to these works, our single-dimensional results slightly extend the state of the art [Gonczarowski and Nisan 2017; Hartline and Taggart 2019] as a *corollary* of a more general theorem that applies to multi-dimensional settings. Even restricted to single-dimensional settings, our proof is perhaps more transparent.

⁵Here and throughout the article when we refer to “state-of-the-art” for single-parameter settings, we are specifically referring to allocation constraints that can be accommodated.

⁶Note that for the single-parameter setting, our algorithm in fact coincides with that of Devanur et al. [2016]. However our analysis, unlike theirs, extends to arbitrary allocation constraints. Our approach also transparently handles mild extensions of constraints *beyond* those considered in Hartline and Taggart [2019] and Gonczarowski and Nisan [2017]. Gonczarowski and Nisan [2017] explicitly state that their techniques cannot handle such extensions and leave this question (which we successfully resolve) open.

We conclude with a brief discussion and an open problem. Corollaries 1.2 and 1.3 are both deduced from Theorem 1.1 by use of an argument as to why the resulting ε -BIC auction is in fact BIC or by using an ε -BIC to BIC reduction that loses negligible revenue. Given that we have explicitly referenced the existence of a quite general ε -BIC-to-BIC reduction, the reader may be wondering why this reduction does not in fact allow our general results to be exactly BIC as well.

The main barrier is the following: To actually *run* the ε -BIC-to-BIC reduction as part of our auction for $n > 1$ bidders, one must take samples *exponential in the number of items* from each bidders' value distribution. This means that even though we can learn an ε -BIC mechanism with few samples, plugging it through the reduction to remove the ε would cost us exponentially many samples in addition. Note that our current use of these theorems is non-constructive: we only use them to claim that the revenues achievable by the optimal BIC and ε -BIC mechanisms are not far off. This conclusion does not actually require *running* the reduction, but rather simply observing that it could be run (more details in Section 5).

When bidder valuations are drawn from a product distribution, it seems conceivable (especially given our results), that sample complexity polynomial in the number of items should suffice. Indeed, if each bidders' values are drawn i.i.d., this is known due to exploitations of symmetry [Daskalakis and Weinberg 2012]. But subexponential sample complexity is not known to suffice for any other restricted class of distributions, despite remarkable recent progress in developing connections to combinatorial Bernoulli factories [Dughmi et al. 2017]. We state below what we consider to be the main open problem left by our work in the context of this article, but readers familiar with black-box reductions in Bayesian mechanism design will immediately recognize a corresponding open problem for the original welfare-maximization setting studied in Hartline et al. [2011] and Bei and Huang [2011] that is equally enticing.

OPEN PROBLEM 1. *Given an ε -BIC auction for some product distribution over $n > 1$ bidders and $m > 1$ items, even with additive bidders, is it possible to transform it into a (precisely) BIC auction with negligible ($\text{poly}(\varepsilon) \cdot \text{poly}(n, m, H)$) revenue loss using polynomially many samples from this product distribution?*

1.3.1 Subsequent Related Work. Since the presentation of a preliminary version of this work at FOCS 2018, progress has been made on three fronts that merits brief discussion. First is recent work of Kothari et al. [2019], which gives a QPTAS for a single unit-demand buyer over m independent items. As referenced in the previous section, our results now immediately extend to provide an up-to- ε approximation to the optimal revenue for a single unit-demand buyer using polynomially-many samples with quasi-polynomial runtime.

Second is recent work of Gergatsouli et al. [2019], which identifies barriers toward a problem related to Open Problem 1. Specifically, one can view Open Problem 1 as asking for an algorithm that takes as input a *fully known auction*, and has *sample access to D* , and outputs a BIC auction. Gergatsouli et al. [2019] prove an impossibility result in a setting where the algorithm is given as input a *fully known D* , but only has *query access to the auction* (there are also a few other additional differences). To best see why this difference is significant, consider a single bidder and m items. With a (fully-known) ε^2 -IC auction A , multiplying all prices by a multiplicative $(1 - \varepsilon)$ results in a IC auction that up-to- ε approximates the revenue of A , using no samples from D (i.e., this resolves the single-bidder version of Open Problem 1; see Section 6.1). However, the auction A may offer $\exp(m)$ different outcomes, so actually learning A via query access to figure out which outcome of A a given buyer might choose (even when D is explicitly known) might require $\exp(m)$ queries to A . Indeed, this reduction positively resolves the single-bidder version of Open Problem 1, but breaks down for the question studied by Gergatsouli et al. [2019], which they instead *negatively* resolve. Gergatsouli et al. [2019] therefore identify barriers to a positive resolution of a problem

related to Open Problem 1, but in a model that is different enough that one should not make too strong a conjecture regarding Open Problem 1 based on their results.

Finally, there is the recent work of Guo et al. [2019], which applies techniques similar to ours to study the precise sample complexity of several domains, including single-buyer revenue maximization (the scenario we discuss in Section 6.1). While we are most interested in establishing that the sample complexity is polynomial (rather than nailing it down precisely), the latter work establishes that certain parts of our analysis are in fact tight up to logarithmic factors. Still, it is not known whether our full results are tight, and it is a fascinating open problem to determine this (also posed explicitly in Guo et al. [2019]). Specifically, Guo et al. [2019] also establish a lower bound on learning an up-to- ϵ optimal hypothesis over a product distribution, as a function of the size of support of the marginals. Put another way, their lower bound implies that improving our sample complexity (beyond logarithmic factors) would require either (a) discretizing differently, and/or quantitatively improving state-of-the-art ϵ -IC to IC reductions or (b) exploiting some (currently unknown) structure of revenue-maximizing multi-item auctions.⁷ Both directions are important for future work.

The remainder of this article is structured as follows. In Section 2, we formally present the model and setting. In Section 3, we formally state our results, which are informally stated above as Theorem 1.1 and Corollaries 1.2 and 1.3. In Section 4, we overview the main ideas behind the proof of Theorem 1.1, a proof that we give in full detail in Section 5. In Section 6, we derive Corollaries 1.2 and 1.3. We present some extensions in Section 7. In Appendix A, we state and prove a result analogous to Theorem 1.1 for DSIC auctions, using similar proof techniques. Parts of certain proofs are relegated to Appendix B.

2 MODEL AND PRELIMINARIES

The Decision Maker (Seller), Bidders, and Outcomes. A single decision maker has the power to choose a social outcome, such as who gets which good that is for sale, or such as which pastime activities are offered in which of the weekends of the upcoming year. There are n bidders who have stakes in this outcome. (The decision maker will be able to charge the bidders and will wish to maximize her revenue.) The possible set of allowed outcomes is denoted by X and can be completely arbitrary. A central example is that of an m -parameter auction: The decision maker is a seller who has m items for sale, and the set of outcomes/allocation is $X \subseteq [0, 1]^{n \cdot m}$, where an allocation $\vec{x} = (x_{i,j})_{i \in [n], j \in [m]} \in X$ specifies for each bidder i and good j the amount of good j that bidder i wins. The traditional multi-item setting is the special case with $X = X_{\text{multi-item}} \triangleq \{(x_{i,j})_{i,j} \in \{0, 1\}^{n \cdot m} \mid \forall j : \sum_{i=1}^n x_{i,j} \leq 1\}$, while outcomes with fractional coordinates occur for example in the canonical model of position auctions, where smaller coordinates denote smaller click-through rates.

Values. Bidder $i \in [n]$ has a valuation function $v_i(\cdot)$ over the set of possible outcomes X . This function is parametrized by m values $v_{i,1}, \dots, v_{i,m}$ (we will not explicitly write $v_{i,\vec{v}_i}(\cdot)$, but refer to the parameters implicitly for ease of notation. Moreover, as $v_i(\cdot)$ is completely determined by $v_{i,1}, \dots, v_{i,m}$, we will sometimes simply refer to \vec{v}_i as bidder i 's value, and to $v_{i,j}$ as bidder i 's value for parameter j) and drawn from a given distribution such that:

- (Independent Parameters) The $v_{i,j}$ s are independent random variables, drawn from distributions $V_{i,j}$ that are all supported in $[0, H]$.

⁷ As already noted, understanding the structure of such auctions is a major open problem independently of auction learning.

- (Lipschitz) There exists an absolute constant L , such that if $v'_{i,1}, \dots, v'_{i,m}$ is obtained from $v_{i,1}, \dots, v_{i,m}$ by modifying one of the $v_{i,j}$ s by at most an additive ε , then $|v_i(x) - v'_i(x)| \leq L\varepsilon$ for all $x \in X$.

For example, in the multi-item setting described above, $v_i(x) = \sum_{j=1}^m x_{i,j} \cdot v_{i,j}$ (and $L = 1$).⁸ An additive-up-to- k -items setting may also be easily captured using this setting (again with $L = 1$), and so can even settings with complementarities, such as a setting in which good $2j$ is worth $v_{i,j}$ to bidder i iff bidder i also gets good $2j + 1$ (and is otherwise worthless to bidder i). Since X can be completely arbitrary (in particular does not have to be a subset of $[0, 1]^{n \cdot m}$), we can most generally think of the $v_{i,j}$ s as parameters that capture the “relevant attributes” of each bidder, such as affinity to action films, affinity to winter sports, willingness to spend a lot of time in a pastime activity, and so on. The only requirement is that these attributes are a “natural” parametrization in the sense that the utility of the bidder from any given outcome $x \in X$ smoothly depends on (i.e., is Lipschitz in) each of them,⁹ and that they are independently drawn.

We note that both properties above (independent items and Lipschitz) together imply that the valuation of each bidder for each outcome is bounded in $[0, mLH]$.

Payments, Priced Outcomes, and Mechanisms. A payment specification $p = (p_i)_{i \in [n]}$ specifies for each bidder i to be charged p_i . A priced outcome is a pair (x, p) of an allocation and a payment specification. The utility of bidder i with value $v_i(\cdot)$ from priced outcome (x, p) is $u_i(v_i, (x, p)) = v_i(x) - p_i$. An auction/mechanism is a function that maps each valuation profile $(v_{i,j})_{i \in [n], j \in [m]}$ to a distribution over priced outcomes. The seller’s expected revenue from a mechanism μ is¹⁰ $\mathbb{E}_{v \sim \times_{i,j} v_{i,j}} [\sum_{i \in [n]} p_i(v)]$, where $p(v)$ is the payment specification chosen by the mechanism for the valuation profile v .

Truthfulness. An auction μ is individually rational (IR) if the expected utility of a truthful bidder is nonnegative at any valuation profile, i.e.: $\mathbb{E}[u_k(v_k, \mu(v))] \geq 0$ for every $k \in [n]$ and $v \in [0, H]^{n \cdot m}$, where the expectation is over the randomness of the auction. For $\varepsilon > 0$, an auction μ is ε -dominant-strategy incentive compatible (ε -DSIC) if truthful bidding maximizes a bidder’s expected utility at any valuation profile up to an additive ε , i.e.:¹¹ $\mathbb{E}[u_k(v_k, \mu(v))] \geq \mathbb{E}[u_k(v_k, \mu(v'_k, v_{-k}))] - \varepsilon$ for every $k \in [n]$, $v \in [0, H]^{n \cdot m}$, and $v'_k \in [0, H]^m$, where the expectation is once again over the randomness of the auction. An auction is DSIC if it is 0-DSIC. An auction μ is ε -Bayesian incentive compatible (ε -BIC) if truthful bidding maximizes a bidder’s utility in expectation over all valuations of the other bidders, up to an additive ε , i.e.: $\mathbb{E}_{v_{-k} \sim \times_{i \neq k} v_{i,j}} [u_k(v_k, \mu(v_k, v_{-k}))] \geq \mathbb{E}_{v_{-k} \sim \times_{i \neq k} v_{i,j}} [u_k(v_k, \mu(v'_k, v_{-k}))] - \varepsilon$ for every $k \in [n]$ and $v_k, v'_k \in [0, H]^m$, where the

⁸This case is actually Lipschitz in a stronger sense: If $v'_{i,1}, \dots, v'_{i,m}$ is obtained from $v_{i,1}, \dots, v_{i,m}$ by modifying one of the $v_{i,j}$ s by at most an additive ε , then $|v_i(\vec{x}) - v'_i(\vec{x})| \leq |\vec{x}|_1 \cdot L\varepsilon$ for $L = 1$. We note that using this stronger property (as well as other properties of the multi-item setting such as monotonicity), our analysis (mutatis mutandis) can be used to quantitatively improve the polynomial dependency of our sample complexity on the parameters of the problems, however we do not follow this direction in this article. In general, in this article we always choose generality of results over tighter polynomials.

⁹This rules out such “tricks” as using bit-interleaving to condense the m parameters into a single parameter.

¹⁰This expectation is both over the draw of valuation profile v and over the draw of priced outcome from $\mu(v)$; to avoid clutter, we will not explicitly mention the latter in our notation throughout this article.

¹¹As is standard in the Game Theory literature, given an n -bidder valuation profile $v = (v_1, \dots, v_n)$, we use v_{-k} for any $k \in [n]$ to denote this valuation profile without the valuation of bidder k , and use (v'_k, v_{-k}) for any $v'_k \in [0, H]^m$ to denote the valuation profile obtained from v by replacing bidder k ’s valuation with v'_k .

expectation is both over the valuations of the bidders other than k and over the randomness of the auction. An auction is BIC if it is 0-BIC.

Additional Notation. We will use the following additional notation in our analysis, where $\varepsilon > 0$:

- For $v \in [0, H]$, we denote by $\lfloor v \rfloor_\varepsilon$ the value of v , rounded down to the nearest integer multiple of ε .
- We use $[0, H]_\varepsilon = \{\lfloor v \rfloor_\varepsilon \mid v \in [0, H]\}$ to denote the set of integer multiples of ε in $[0, H]$.
- For every i, j , we denote by $\lfloor V_{i,j} \rfloor_\varepsilon$ the distribution of $\lfloor v_{i,j} \rfloor_\varepsilon$ for $v_{i,j} \sim V_{i,j}$.

Existing Tools. In our analysis, we will make use of the following two theorems, which we state below in a way that is adapted to the notation of our article. The first shows the optimal revenue over all ε -BIC auctions and the optimal revenue over all BIC auctions are close (while this is stated in Rubinstein and Weinberg [2015] with respect to multi-parameter settings with allocations in $\{0, 1\}^{n \cdot m}$, the same proof holds verbatim for arbitrary outcome sets X):

THEOREM 2.1 (RUBINSTEIN AND WEINBERG [2015];¹² SEE ALSO DASKALAKIS AND WEINBERG [2012]). *Let \mathcal{D} be any joint distribution over arbitrary valuations, where the valuations of different bidders are independent. The maximum revenue attainable by any IR and ε -BIC auction for \mathcal{D} is at most $2n\sqrt{mLH\varepsilon}$ greater than the maximum revenue attainable by any IR and BIC auction for that distribution.*

The second is a Chernoff-style concentration inequality for product distributions:

THEOREM 2.2 (BABICHENKO ET AL. [2017]; SEE ALSO DEVANUR ET AL. [2016]). *Let D_1, \dots, D_ℓ be discrete distributions. Let $S \in \mathbb{N}$. For every i , draw S independent samples from D_i , and let $D_i^{(S)}$ be the uniform distribution over these samples. For every $\varepsilon > 0$ and $f : \times_{i=1}^\ell \text{supp } D_i \rightarrow [0, H]$, we have that $\Pr(|\mathbb{E}_{\times_{i=1}^\ell D_i^{(S)}}[f] - \mathbb{E}_{\times_{i=1}^\ell D_i}[f]| > \varepsilon) \leq \frac{4H}{\varepsilon} \exp(-\frac{\varepsilon^2 S}{8H^2})$.*

Unlike standard Chernoff bounds, which state that the expectation over the empirically sampled distribution (over ℓ -tuples) well approximates the expectation over the true distribution, this concentration inequality states that the expectation over the *product of the marginals of the empirically sampled distribution* well approximates the expectation over the true distribution. This difference is crucial to us, since, as we will see, there are far less such possible products of empirical marginals than there are empirical distributions, so a far smaller number of auctions can guarantee revenue optimization over the former, and this concentration inequality lets us generalize this revenue optimization to be over the true distribution.

3 MAIN RESULTS

In this section, we formally state our main results, which were informally presented as Theorem 1.1 and Corollaries 1.2 and 1.3 in the Introduction. We start with our main result.

THEOREM 3.1 (MAIN RESULT). *For every $\varepsilon, \delta > 0$ and for every $\eta \leq \text{poly}(n, m, L, H, \varepsilon)$, the sample complexity of learning an up-to- ε optimal IR and η -BIC auction is $\text{poly}(n, m, L, H, 1/\varepsilon, 1/\eta, \log 1/\delta)$.*

¹²If we denote by R_ε the maximum expected revenue attainable by any IR and ε -BIC auction for the bidders' product distribution, and by R the maximum revenue attainable by any IR and BIC auction for the same distribution, then the result of Rubinstein and Weinberg [2015] is that for any $\eta > 0$, it is the case that $R \geq (1 - \eta) \cdot (R_\varepsilon - \frac{n\varepsilon}{\eta})$. Choosing $\eta = \sqrt{\frac{\varepsilon}{mLH}}$ yields Theorem 2.1 as stated above, since R_ε is trivially bounded from above by the maximum possible sum of valuations, i.e., by $nmLH$.

That is, there exists a deterministic algorithm¹³ that given $\text{poly}(n, m, L, H, 1/\epsilon, 1/\eta, \log 1/\delta)$ samples from each $V_{i,j}$, with probability $1 - \delta$ outputs an IR and η -BIC auction that attains from $\times_{i,j} V_{i,j}$ expected revenue at most an additive ϵ smaller than any IR and η -BIC auction.

The following corollary of our main result should be contrasted with a result of Dughmi et al. [2014], which shows that finding an ϵ -optimal mechanism for a single additive bidder with correlated item distributions requires exponentially many samples.

THEOREM 3.2 (SINGLE BIDDER). *When there is $n = 1$ bidder, for every $\epsilon, \delta > 0$, the sample complexity of learning an up-to- ϵ optimal IR and IC¹⁴ auction is $\text{poly}(n, m, L, H, 1/\epsilon, \log 1/\delta)$. That is, there exists a deterministic algorithm¹⁵ that given $\text{poly}(n, m, L, H, 1/\epsilon, \log 1/\delta)$ samples from each $V_{i,j}$, with probability $1 - \delta$ outputs an IR and IC auction that attains from $\times_{i,j} V_{i,j}$ expected revenue at most an additive ϵ smaller than any IR and IC auction.*

The following corollary of our main result unifies and even somewhat extends the state-of-the-art results for single-parameter ($m = 1$) revenue maximization. To state it we restrict ourselves to the setting where revenue maximization has been solved by Myerson [1981]: Assume that $X \subseteq [0, 1]^n$ and that $v_i(x) = v_i \cdot x_i$.¹⁶

THEOREM 3.3 (SINGLE-PARAMETER). *In an $m = 1$ -parameter setting with $X \subseteq [0, 1]^n$ and $v_i(x) = v_i \cdot x_i$, for every $\epsilon, \delta > 0$, the sample complexity of **efficiently** learning an up-to- ϵ IR and DSIC auction is $\text{poly}(n, m, L, H, 1/\epsilon, \log 1/\delta)$. That is, there exists a deterministic algorithm with running time $\text{poly}(n, m, L, H, 1/\epsilon, \log 1/\delta)$ that given $\text{poly}(n, m, L, H, 1/\epsilon, \log 1/\delta)$ samples from each $V_{i,j}$, with probability $1 - \delta$ outputs an IR and DSIC auction that attains from $\times_{i,j} V_{i,j}$ expected revenue at most an additive ϵ smaller than any IR and BIC auction.*

4 MAIN RESULT PROOF OVERVIEW

In this section, we roughly sketch our learning algorithm and present each of the main ideas behind its analysis, by presenting a proof overview structured to present each of these ideas separately. The proof overview is given in this section only for an additive multi-item setting, and some elements of the proof are omitted or glossed over for readability. The full proof, which contains all omitted details and applies to a general arbitrary Lipschitz setting, and in which the main ideas that are surveyed in this section separately are quite intermingled, is given in Section 5.¹⁷

Our learning algorithm is similar in nature to the one presented in Devanur et al. [2016] for certain single-parameter environments; however, the analysis that we will use to show that it does not overfit the samples is completely different (even for single-parameter environments, where our analysis holds for arbitrary allocation constraints). Recall that our result is (necessarily)

¹³Recall that this result is information-theoretic and not computationally efficient (by necessity, without resolving major open problems), so our decision maker (seller) is computationally unbounded, and we allow the algorithm to make calls to any deterministic oracle that has no access to any $V_{i,j}$. In particular, we assume access to an oracle that can solve the revenue maximization problem on any precisely given $V'_{i,j}$ of finite support.

¹⁴Recall that for a single bidder, the notions of BIC and DSIC coincide.

¹⁵See Footnote 13.

¹⁶This setting is slightly more general than the state-of-the-art single-parameter results, which assume $X \subseteq \{0, 1\}^n$ and/or the special setting of position auctions [Gonczarowski and Nisan 2017; Hartline and Taggart 2019]. As noted in the Introduction, Gonczarowski and Nisan [2017] explicitly state that their techniques cannot accommodate arbitrary $X \subseteq [0, 1]^n$ and leave this question (which we successfully resolve) open.

¹⁷The somewhat less involved proof for DSIC auctions that is sketched in this section as an intermediary proof is given in Appendix A (in that appendix, though, the analysis is more general as it refers to general Lipschitz valuations and not only to additive valuations as in this section), both since we find that result interesting in its own right and to allow interested readers to familiarize themselves with that proof before diving into the more involved proof for BIC auctions in Section 5.

information-theoretic and not computationally efficient. Therefore, some steps in the algorithm perform operations that are not known to be performable in poly-time (but can certainly be performed without access to any $V_{i,j}$). In particular, our algorithm will solve an instance of a Bayesian revenue maximization problem for a precisely given input of finite support (step 2).

Algorithm. We start with S (to be determined later) independent samples from each $V_{i,j}$. Our algorithm roughly proceeds as follows:

- (1) For each bidder i and good j , round all samples from $V_{i,j}$ down to the nearest multiple of ϵ . Denote the uniform distribution over these rounded samples by $W_{i,j}$.
- (2) Find an IR and $O(\epsilon)$ -IC (see below) multi-item auction that maximizes the revenue from the product of the rounded empirical distributions $W_{i,j}$. Denote this auction by μ .
- (3) Return the auction μ^ϵ , which on input \vec{v} , rounds down all actual bids to the nearest multiple of ϵ , $\lfloor \vec{v} \rfloor_\epsilon$, and allocates and charges payments according to the output of $\mu(\lfloor \vec{v} \rfloor_\epsilon)$ when run on these rounded bids.

We start with a simpler scenario, namely that of learning DSIC auctions, and only toward the end of this section introduce the additional issues that arise when learning BIC auctions. We therefore start by showing that if in step 2 of our algorithm we interpret “IC” as “DSIC,” that is, that if in that step we take an IR and $O(\epsilon)$ -DSIC auction that maximizes the revenue from the product of the rounded empirical distributions, then there exists $S = \text{poly}(n, m, H, 1/\epsilon, \log 1/\delta)$ such that the auction μ^ϵ output by our algorithm is $O(\epsilon)$ -DSIC and its revenue from $\times_{i,j} V_{i,j}$ is, with probability at least $1 - \delta$, up-to- $O(\epsilon)$ -close to the maximum revenue attainable from $\times_{i,j} V_{i,j}$ by any DSIC auction. (The formal statement and full proof are given in Appendix A.) We note that the auction output by the algorithm is indeed $O(\epsilon)$ -DSIC, since the output μ in step 2 is $O(\epsilon)$ -DSIC, and the rounding of the actual bids as defined in step 3 only loses another $m\epsilon$.¹⁸

Uniform Convergence of the Revenue of All Possible Output Mechanisms. Note that for every i, j , each rounded sample from step 1 of the algorithm is independently distributed according to $\lfloor V_{i,j} \rfloor_\epsilon$. The main challenge is in showing that the resulting auction gives up-to- $O(\epsilon)$ optimal revenue not only on the rounded empirical distributions $\times_{i,j} W_{i,j}$, but also on the rounded true distributions $\times_{i,j} \lfloor V_{i,j} \rfloor_\epsilon$. That is, the main challenge is in showing that no overfitting occurs, in the absence of any structural properties that we can exploit for the mechanisms that are optimal (or up-to- $O(\epsilon)$ optimal) for $\times_{i,j} W_{i,j}$.

This is the point where our approach makes a sharp departure from prior works. Prior work deems this task to be hopeless, and proceeds by proving structural results on optimal mechanisms for restricted domains. We circumvent this by instead simply counting the number possible inputs we will ever query in step 2, and observing that the number of mechanisms over which we have to obtain uniform convergence is at most this number. A crucial observation is that while we do have to consider more and more mechanisms as the number of samples S grows, the number of mechanisms that we have to consider grows moderately enough so as to not eclipse our gains from increasing the number of samples that we take. For this argument to hold, it is *essential* that our distributions are product distributions.

Let \mathcal{V} be the set of all product distributions $\times_{i,j} W'_{i,j}$ where each $W'_{i,j}$ is the uniform distribution over some multiset of S values from $[0, H]_\epsilon$. Let M be the set of all mechanisms returned by step 2 of the algorithm for some distribution $\times_{i,j} W'_{i,j} \in \mathcal{V}$. *At the heart of our analysis, and of this part of our analysis in particular, is the observation that $|\mathcal{V}| \leq (S + 1)^{n \cdot m \cdot \lceil H/\epsilon \rceil}$. Crucially, this expression*

¹⁸The astute reader will notice that $m\epsilon \notin O(\epsilon)$. As all our bounds are polynomial in $m, 1/\epsilon$ anyway, this is immaterial, and one example of a detail that we glossed over in this section in the interest of cleanliness, as promised.

has S only in the base and not in the exponent. Indeed, for every $\times_{i,j} W'_{i,j} \in \mathcal{V}$, for every i, j , and for every integer multiple of ε in $[0, H]$ (there are $\lceil H/\varepsilon \rceil$ many such values), the probability of this value in $W'_{i,j}$ can be any of the $S + 1$ values $0, 1/S, \dots, 1$. Therefore, $|M| \leq (S + 1)^{n \cdot m \cdot \lceil H/\varepsilon \rceil}$.

We will choose S so that with probability at least $1 - \delta$, it simultaneously holds for all mechanisms $\mu \in M$ that

$$|\text{Rev}_{\times_{i,j} W_{i,j}}(\mu) - \text{Rev}_{\times_{i,j} [V_{i,j}]_\varepsilon}(\mu)| \leq \varepsilon. \quad (1)$$

To this end, we will use a Chernoff-style concentration bound (Theorem 2.2) for product distributions, which when applied to our setting shows that for each auction separately Equation (1) is violated with probability exponentially small in $\frac{\varepsilon^2 S}{m^2 H^2}$. So, to have Equation (1) hold with probability at most $1 - \delta$ for all auctions in M simultaneously, we choose S so that the violation probability for each auction separately is at most $\delta/|M|$, and use a union bound. Since $|M| \leq (S + 1)^{n \cdot m \cdot \lceil H/\varepsilon \rceil}$, we have that it is enough to take S such that $\frac{\varepsilon^2 S}{m^2 H^2}$ is of order of magnitude at least $\log |M|/\delta = \log \frac{(S+1)^{n \cdot m \cdot \lceil H/\varepsilon \rceil}}{\delta} \approx \log 1/\delta + n \cdot m \cdot H/\varepsilon \log S$, which is clearly possible by taking a suitable S that is polynomial in $n, m, H, 1/\varepsilon$, and $\log 1/\delta$. So, taking a number of sample of this magnitude gives that with probability at least $1 - \delta$, Equation (1) simultaneously holds for all mechanisms in M and so the mechanism output by step 2 of the algorithm gets up-to- $O(\varepsilon)$ the same revenue on the product of the rounded empirical distributions $\times_{i,j} W_{i,j}$ as it does on the product of the rounded true distributions $\times_{i,j} [V_{i,j}]_\varepsilon$. So, the revenue that the mechanism μ^ε output by (step 3 of) the algorithm attains from $\times_{i,j} V_{i,j}$ is identical to the revenue that the mechanism μ output by step 2 of the algorithm attains from $\times_{i,j} [V_{i,j}]_\varepsilon$, which is up-to- $O(\varepsilon)$ the optimal revenue attainable from $\times_{i,j} W_{i,j}$.

Revenue Close to Optimal. Our next task is to show that with high probability the optimal revenue attainable from $\times_{i,j} W_{i,j}$ by any $O(\varepsilon)$ -DSIC auction is up-to- $O(\varepsilon)$ the same as the optimal revenue attainable from $\times_{i,j} V_{i,j}$ by any DSIC auction, which would imply that the revenue that μ^ε attains from $\times_{i,j} V_{i,j}$ is close to optimal, as required. Let OPT be the DSIC auction that maximizes the revenue (among such auctions) in expectation over $\times_{i,j} V_{i,j}$. *At the heart of this part of our analysis is the fact that while our algorithm cannot hope to find OPT , we can nonetheless carefully reason about it in our analysis, as it is nonetheless fixed and well-defined (in particular, it does not depend on the drawn samples).* Let OPT_ε be the mechanism defined over $\times_{i,j} [V_{i,j}]_\varepsilon$ as follows: for each bidder i and item j , let $w_{i,j}$ be the input bid of bidder i for item j (a multiple of ε), and replace it by a bid $v_{i,j}$ independently drawn from the distribution $V_{i,j}$ conditioned upon being in the interval $[w_{i,j}, w_{i,j} + \varepsilon]$; the auction OPT_ε allocates and charges payments according to the output of OPT when run on these drawn replacement bids. Obviously, the auction OPT_ε is an $O(\varepsilon)$ -DSIC auction¹⁹ whose revenue from $\times_{i,j} [V_{i,j}]_\varepsilon$ is identical to that of the auction OPT from $\times_{i,j} V_{i,j}$, i.e., to the optimal revenue from $\times_{i,j} V_{i,j}$, so it is enough to show that the revenue of the auction OPT_ε from $\times_{i,j} [V_{i,j}]_\varepsilon$ and from $\times_{i,j} W_{i,j}$ is the same up-to- $O(\varepsilon)$ with high probability, that is, that Equation (1) also holds for the mechanism OPT_ε with high probability. To do so, we modify the definition of the set M to also include the (well-defined even prior to sampling, despite being unknown to our algorithm) mechanism OPT_ε —since the order of magnitude of $|M|$ does not change, the order of magnitude of the number of samples required to guarantee that Equation (1) holds for all auctions in M (including OPT_ε) does not change.

Bayesian Incentive Compatibility. We conclude our proof overview by adapting the proof to the more delicate BIC notion of incentive compatibility, thus showing that if in step 2 of our algorithm

¹⁹Or more accurately, $O(m\varepsilon)$ -DSIC; see Footnote 18.

we take an $O(\varepsilon)$ -BIC (rather than $O(\varepsilon)$ -DSIC) and IR auction that maximizes the revenue from the product of the rounded empirical distributions, then there exists $S = \text{poly}(n, m, H, 1/\varepsilon, \log 1/\delta)$ such that the auction μ^ε output by our algorithm is, with probability at least $1 - \delta$, an $O(\varepsilon)$ -BIC auction whose revenue from $\times_{i,j} V_{i,j}$ is up-to- $O(\varepsilon)$ -close to the maximum revenue attainable from $\times_{i,j} V_{i,j}$ by any BIC auction (and therefore, by Theorem 2.1, up-to- $O(\sqrt{\varepsilon})$ -close to the revenue attainable from this distribution by any $O(\varepsilon)$ -BIC auction).²⁰ The challenge here is that (approximate) BIC is a distribution-dependent property of a mechanism (as opposed to DSIC, which is a distribution-agnostic incentive compatibility notion). Indeed, examining our analysis above with (ε) -DSIC replaced by (ε) -BIC, we note that the resulting analysis falls short of carrying through in two points: it is unclear why OPT_ε is $O(\varepsilon)$ -BIC not only with respect to $\times_{i,j} V_{i,j}$ but also with respect to $\times_{i,j} W_{i,j}$, and it is unclear why any mechanism that can be output by step 2 of our algorithm is $O(\varepsilon)$ -BIC not only with respect to $\times_{i,j} W_{i,j}$ but only with respect to $\times_{i,j} V_{i,j}$. *At the heart of this part of our analysis is the observation that the set of all interim expected utilities, of all bidders' possible types, from all possible reported types, in all mechanisms,*²¹ *on the one hand, is comprised of a small-enough number of random variables to still enable uniform convergence, and, on the other hand, contains sufficient information to show that incentive constraints do not deteriorate much.* Concretely, we will choose S so that with probability at least $1 - \delta$, simultaneously for all mechanisms in M (including OPT_ε) not only does Equation (1) hold, but also the following holds for every bidder $k \in [n]$ and values $v_k, v'_k \in [0, H]_\varepsilon^m$:

$$\left| \mathbb{E}_{v_{-k} \sim \times_{i \neq k} W_{i,j}} [u_k(v_k, \mu(v'_k, v_{-k}))] - \mathbb{E}_{v_{-k} \sim \times_{i \neq k} [V_{i,j}]_\varepsilon} [u_k(v_k, \mu(v'_k, v_{-k}))] \right| \leq \varepsilon. \quad (2)$$

We note that for every mechanism μ , we require that Equation (2) hold for $n \cdot \lceil H/\varepsilon \rceil^{2m}$ distinct combinations of $k \in [n]$ and $v_k, v'_k \in [0, H]_\varepsilon^m$. Crucially, this number does not depend on S . So, the number of instances of Equation (2) that we would like to hold simultaneously with high probability is $|M| \cdot n \cdot \lceil H/\varepsilon \rceil^{2m}$, and so we have $|M| \cdot (1 + n \cdot \lceil H/\varepsilon \rceil^{2m}) \leq (1 + n \cdot \lceil H/\varepsilon \rceil^{2m}) \cdot (S + 1)^{n \cdot m \cdot \lceil H/\varepsilon \rceil}$ instances of either Equation (1) or Equation (2) that we would like to hold simultaneously with high probability.²² As this number still has S only in the base and not in the exponent, we can proceed as above to guarantee this with high probability using only a polynomial number of samples.

5 PROOF OF MAIN RESULT

In this section, we give the full details of the proof of our main result, Theorem 3.1. The proofs of supporting lemmas are relegated to Appendix B.

PROOF OF THEOREM 3.1. We assume that for every $i \in [n]$ and $j \in [m]$, we have S (to be determined later) independent samples $(v_{i,j}^s)_{s=1}^S$ from $V_{i,j}$. Algorithm 1 presents our learning algorithm, which is similar in nature to the one presented in Devanur et al. [2016] for certain single-parameter environments, however the analysis that we will use to show that it does not overfit the samples is completely different (even for single-parameter environments, where our analysis holds for arbitrary allocation constraints). We now analyze Algorithm 1. Note that for every i, j, s , we have that $w_{i,j}^s \sim [V_{i,j}]_\varepsilon$ independently. \square

²⁰Theorem 2.1 in fact allows us to also reduce to a cleaner oracle, which finds an optimal BIC auction rather than an optimal $O(\varepsilon)$ -BIC auction, in Section 5.

²¹The set of all such interim expected utilities for a single mechanism is sometimes referred to as the *reduced form* of the mechanism.

²²A somewhat similar idea appeared in Cai et al. [2012], albeit without exploiting independence across items.

ALGORITHM 1: Empirical Multi-Parameter Up-to- ε BIC Revenue Maximization.

Function EmpiricalOptimize($H, X, \varepsilon, \delta, (v_{i,j}^s)_{i \in [n], j \in [m]}^{s \in [S]}$):

Input: For every $i \in [n], j \in [m]$, $(v_{i,j}^s)_{s=1}$ is a sequence of $S = \tilde{O}(\frac{H^2}{\varepsilon^2} \cdot (\log 1/\delta + \frac{nmH}{\varepsilon}))$ samples from $V_{i,j}$

Output: With probability $1 - \delta$, an IR and $(4mL\varepsilon)$ -BIC mechanism for $\times_{i,j} V_{i,j}$, defined over $[0, H]^{n \cdot m}$ with allocations in X , whose expected revenue from $\times_{i,j} V_{i,j}$ is up to an additive $(4nmL(\varepsilon + \sqrt{2H\varepsilon}))$ smaller than that of any IR and $(4mL\varepsilon)$ -BIC mechanism for $\times_{i,j} V_{i,j}$ with allocations in X

for $i \in [n], j \in [m]$ **do**

for $s \in [S]$ **do**

$w_{i,j}^s \leftarrow \lfloor v_{i,j}^s \rfloor_\varepsilon$

end

$W_{i,j} \leftarrow$ the uniform distribution over $(w_{i,j}^s)_{s=1}^S$

end

$\mu \leftarrow$ OptimizationOracle($H, X, \varepsilon, (W_{i,j})_{i \in [n], j \in [m]}$) // See definition below

return The mechanism that for input $(v_{i,j})_{i \in [n], j \in [m]}$ outputs $\mu(\lfloor v_{i,j} \rfloor_\varepsilon)_{i \in [n], j \in [m]}$, modified to charge each bidder $mL\varepsilon$ less.

Function OptimizationOracle($H, X, \varepsilon, (W_{i,j})_{i \in [n], j \in [m]}$):

Input: For every $i \in [n], j \in [m]$, $W_{i,j}$ is a distribution over $[0, H]_\varepsilon$

Output: An IR and BIC mechanism for $\times_{i,j} W_{i,j}$, defined over $[0, H]_\varepsilon^{n \cdot m}$ with allocations in X , which maximizes the expected revenue from $\times_{i,j} W_{i,j}$ among all IR and BIC mechanisms for $\times_{i,j} W_{i,j}$ with allocations in X

$\mu \leftarrow$ an IR and BIC mechanism for $\times_{i,j} W_{i,j}$ (defined over $\times_{i,j} \text{supp } W_{i,j}$) with allocations in X , which maximizes the expected revenue from $\times_{i,j} W_{i,j}$ among all such mechanisms

return The mechanism obtained by extending μ to be defined over $(v_{i,j})_{i,j} \in [0, H]_\varepsilon^{n \cdot m}$ as follows: for every $k \in [n]$ s.t. $v_{k,j} \notin \text{supp } W_{k,j}$ for some $j \in [m]$, replace the entire bid vector $v_k = (v_{k,j})_{j \in [m]}$ of bidder k with a bid vector $v'_k = (v'_{k,j})_{j \in [m]} \in \text{supp } \times_j W_{k,j}$ that maximizes $\mathbb{E}_{v'_{-k} \sim \times_{i,j:i \neq k} W_{i,j}} [u_k(v_k, \mu(v'_k, v'_{-k}))]$

Let \mathcal{V} be the set of all product distributions $\times_{i,j} W'_{i,j}$ where each $W'_{i,j}$ is the uniform distribution over some multiset of S values from $[0, H]_\varepsilon$. Let M be the set of all mechanisms of the form OptimizationOracle($H, X, \varepsilon, \times_{i,j} W'_{i,j}$) for all $\times_{i,j} W'_{i,j} \in \mathcal{V}$. At the heart of our analysis is the observation that $|\mathcal{V}| < (S + 1)^{n \cdot m \cdot \lceil H/\varepsilon \rceil}$. (Crucially, this expression has S only in the base and not in the exponent!) Indeed, for every $\times_{i,j} W'_{i,j} \in \mathcal{V}$, for every i, j , and for every integer multiple of ε in $[0, H]$ (there are $\lceil H/\varepsilon \rceil$ many such values), the probability of this value in $W'_{i,j}$ can be any of the $S + 1$ values $0, 1/S, \dots, 1$. (The inequality on $|\mathcal{V}|$ is strict, since, for example, not all probabilities can be 0 simultaneously.) Therefore, $|M| < (S + 1)^{n \cdot m \cdot \lceil H/\varepsilon \rceil}$.

Let OPT be the IR and $(4mL\varepsilon)$ -BIC auction for $\times_{i,j} V_{i,j}$ that maximizes the expected revenue (among such auctions) from $\times_{i,j} V_{i,j}$ (our learning algorithm cannot hope to find OPT, but in our analysis we may carefully reason about it, as it is nonetheless well defined; in particular, it does not depend on S). Let OPT $_\varepsilon$ be the (randomized) mechanism defined over $\times_{i,j} \lfloor V_{i,j} \rfloor_\varepsilon$ as follows: Let $w = (w_{i,j})_{i \in [n], j \in [m]}$ be an input valuation; for each i, j , independently draw $v_{i,j} \sim V_{i,j} \lfloor \cdot \rfloor_{[w_{i,j}, w_{i,j} + \varepsilon]}$; let $(x, p) = \text{OPT}((v_{i,j})_{i \in [n], j \in [m]})$; the allocation of OPT $_\varepsilon(w)$ is x , and the payment of each bidder i is $p_i - mL\varepsilon$.

LEMMA 5.1. OPT_ε is an IR and $(6mL\varepsilon)$ -BIC mechanism for $\times_{i,j}[V_{i,j}]_\varepsilon$, whose expected revenue from $\times_{i,j}[V_{i,j}]_\varepsilon$ is $nmL\varepsilon$ smaller than the expected revenue of OPT from $\times_{i,j} V_{i,j}$.

We will choose S so that with probability at least $1 - \delta$, both of the following simultaneously hold for all mechanisms $\mu \in M \cup \{\text{OPT}_\varepsilon\}$:

- $|\text{Rev}_{\times_{i,j} W_{i,j}}(\mu) - \text{Rev}_{\times_{i,j}[V_{i,j}]_\varepsilon}(\mu)| \leq nmL\varepsilon,$ (3)
- For every agent $k \in [n]$ and types $v_k, v'_k \in [0, H]_\varepsilon^m$:

$$|\mathbb{E}_{v_{-k} \sim \times_{i,j:i \neq k} W_{i,j}} [u_k(v_k, \mu(v'_k, v_{-k}))] - \mathbb{E}_{v_{-k} \sim \times_{i,j:i \neq k}[V_{i,j}]_\varepsilon} [u_k(v_k, \mu(v'_k, v_{-k}))]| \leq mL\varepsilon. \quad (4)$$

We note that for every mechanism μ , we require that Equation (4) hold for $n \cdot \lceil H/\varepsilon \rceil^{2m}$ distinct combinations of $k \in [n]$ and $v_k, v'_k \in [0, H]_\varepsilon^m$. Crucially, this number does not depend on S .

By Theorem 2.2 (with $\ell \triangleq n \cdot m$, and note that any mechanism's revenue is bounded by $nmLH$), we have that for each mechanism $\mu \in M \cup \{\text{OPT}_\varepsilon\}$ separately Equation (3) holds with probability at least $1 - \frac{4He^{-\varepsilon^2 S/(8H^2)}}{\varepsilon}$, and for each combination of (μ, k, v_k, v'_k) separately Equation (4) holds with probability at least $1 - \frac{4He^{-\varepsilon^2 S/(8H^2)}}{\varepsilon}$.

Choosing S so that each of these probabilities is at least $1 - \frac{\delta}{(|M|+1) \cdot n \cdot \lceil H/\varepsilon \rceil^{2m}}$, we obtain that both Equation (3) holds simultaneously for all mechanisms $\mu \in M \cup \{\text{OPT}_\varepsilon\}$ and Equation (4) holds simultaneously for all combinations (μ, k, v_k, v'_k) with probability at least $1 - \delta$. We now estimate S . Since $|M \cup \{\text{OPT}_\varepsilon\}| \leq (S+1)^{n \cdot m \cdot \lceil H/\varepsilon \rceil}$, we have that it is enough to take S such that

$$S \geq \frac{8H^2}{\varepsilon^2} \cdot (\log \frac{4H}{\varepsilon} + \log 1/\delta + \log n + 2m \log \lceil H/\varepsilon \rceil + nm \lceil H/\varepsilon \rceil \log(S+1)).$$

Therefore,²³ there exists an appropriate

$$S = O\left(\frac{H^2}{\varepsilon^2} \log 1/\delta + \frac{nmH^3}{\varepsilon^3} \log \frac{nmH^3}{\varepsilon^3}\right) = \tilde{O}\left(\frac{H^2}{\varepsilon^2} \cdot (\log 1/\delta + \frac{nmH}{\varepsilon})\right).$$

Let μ be the output of the call to `OptimizationOracle` in Algorithm 1, and let μ^ε be the final output of the algorithm (the output of `EmpiricalOptimize`).

LEMMA 5.2. If Equation (4) holds for every mechanism in $M \cup \{\text{OPT}_\varepsilon\}$ and for every (k, v_k, v'_k) , then:

- OPT_ε is a $(8mL\varepsilon)$ -BIC mechanism for $\times_{i,j} W_{i,j}$.
- The expected revenue of μ from $\times_{i,j} W_{i,j}$ is at most $4nmL\sqrt{2H\varepsilon}$ smaller than the expected revenue of OPT_ε from $\times_{i,j} W_{i,j}$.
- μ is a $(2mL\varepsilon)$ -BIC mechanism for $\times_{i,j}[V_{i,j}]_\varepsilon$.

LEMMA 5.3. μ^ε is an IR mechanism whose expected revenue from $\times_{i,j} V_{i,j}$ is $nmL\varepsilon$ smaller than the expected revenue of μ from $\times_{i,j}[V_{i,j}]_\varepsilon$. Furthermore, if μ is $(2mL\varepsilon)$ -BIC for $\times_{i,j}[V_{i,j}]_\varepsilon$, then μ^ε is $(4mL\varepsilon)$ -BIC for $\times_{i,j} V_{i,j}$.

²³The requirement is of the form $S \geq a + b \log S$, so the tight solution is of order $a + b \log b$.

So, with probability at least $1 - \delta$, we have both that μ^ε is an IR and $(4mL\varepsilon)$ -BIC mechanism (by Lemmas 5.2 and 5.3) and that:

$$\begin{aligned}
& \text{Rev}_{\times_{i,j} V_{i,j}}(\mu^\varepsilon) = && \text{(by Lemma 5.3)} \\
& = \text{Rev}_{\times_{i,j} [V_{i,j}]_\varepsilon}(\mu) - nmL\varepsilon && \text{(by Equation (3) and since } \mu \in M) \\
& \geq \text{Rev}_{\times_{i,j} W_{i,j}}(\mu) - 2nmL\varepsilon && \text{(by Lemma 5.2)} \\
& \geq \text{Rev}_{\times_{i,j} W_{i,j}}(\text{OPT}_\varepsilon) - 4nmL\sqrt{2H\varepsilon} - 2nmL\varepsilon && \text{(by Equation (3) for } \text{OPT}_\varepsilon) \\
& \geq \text{Rev}_{\times_{i,j} [V_{i,j}]_\varepsilon}(\text{OPT}_\varepsilon) - 3nmL\varepsilon - 4nmL\sqrt{2H\varepsilon} && \text{(by Lemma 5.1)} \\
& = \text{Rev}_{\times_{i,j} V_{i,j}}(\text{OPT}) - 4nmL\varepsilon - 4nmL\sqrt{2H\varepsilon} \\
& = \text{Rev}_{\times_{i,j} V_{i,j}}(\text{OPT}) - 4nmL(\varepsilon + \sqrt{2H\varepsilon}). && \square
\end{aligned}$$

6 FROM APPROXIMATE TO EXACT INCENTIVE COMPATIBILITY

In this section, we derive sample complexity results for exact incentive compatibility for the special cases of a single bidder (Theorem 3.2) or a single good/single-parameter setting (Theorem 3.3). As mentioned in the Introduction, whether this can also be done for more general settings remains an open question.

6.1 One Bidder

In this section, we will prove Theorem 3.2. For a single bidder, the following theorem, which to the best of our knowledge first implicitly appeared in Balcan et al. [2005], where it is attributed to Nisan,²⁴ provides an ε -IC to IC reduction with negligible revenue loss.

THEOREM 6.1 (NISAN, CIRCA 2005). *Let μ be an IR and ε -IC²⁵ mechanism for a single bidder. Modifying each possible priced outcome by multiplying the payment in that priced outcome by $1 - \sqrt{\varepsilon}$ and letting the bidder choose the (modified) priced outcome that maximizes her utility yields an IR and IC mechanism μ' with expected revenue at least $(1 - \sqrt{\varepsilon}) \cdot (\text{Rev}(\mu) - \sqrt{\varepsilon})$.*

For completeness, we provide a proof of this theorem. The idea is that discounting more expensive priced outcomes more heavily makes sure that incentives do not drive the bidder toward a much cheaper priced outcome. More concretely, due to the auction being only ε -IC, the utility of a bidder from choosing a cheaper priced outcome can be higher by at most ε . Since for any priced outcome whose price is cheaper by more than a $\sqrt{\varepsilon}$ compared to the bidder's original priced outcome, the given discount is smaller by more than $\sqrt{\varepsilon}^2 = \varepsilon$, this smaller discount more than eliminates any potential utility gain due to choosing the cheaper priced outcome, so such a cheaper priced outcome would not become the most-preferred one.

²⁴It appears there and in following papers [Babaioff et al. 2017; Chawla et al. 2007; Dughmi et al. 2014; Gonczarowski 2018; Hart and Nisan 2013] as part of a two-step reduction sometimes called “nudge and round” (this is the “nudge” part), which reduces the menu size of a single-bidder auction with negligible revenue loss. To the best of our knowledge, the first reference to this argument as a general ε -IC to IC reduction rather than as part of a “nudge and round” operation (where it fixes IC issues resulting from rounding) is in Daskalakis and Weinberg [2012], who also attribute it to Nisan following Chawla et al. [2007], who in turn attribute it to Nisan following Balcan et al. [2005]. A similar technique appears in Madarász and Prat [2017] within the context of correcting for model misspecification. In this context, our rounding of the empirical distributions can be viewed as a deliberately introduced model misspecification that on the one hand novelly serves as a tool against overfitting and on the other hand is carefully controlled so that its adverse effect is limited to at most an ε loss in incentive compatibility. In the one bidder case, after guaranteeing that no overfitting occurs, we correct for this loss in incentive compatibility using this technique as in the model misspecification literature.

²⁵Recall once again that for a single bidder, the notions of BIC and DSIC coincide.

PROOF. Fix a type $v \in [0, H]^m$ for the bidder. Let e be the priced outcome (a distribution over priced outcomes, i.e., a random variable, if μ is randomized) according to μ when the bidder has type v . It is enough to show that the bidder pays at least $(1 - \sqrt{\varepsilon})(p_e - \sqrt{\varepsilon})$ in expectation according to μ' when he has type v . (We denote the expected price of, e.g., e by p_e .) Let f' be a possible priced outcome of μ' , and let f be the priced outcome of μ that corresponds to it. We will show that if $p_{f'} < (1 - \sqrt{\varepsilon})(p_e - \sqrt{\varepsilon})$, then the bidder strictly prefers the priced outcome e' of μ' that corresponds to e over f' (and so does not choose f' in μ' , completing the proof). Indeed, since in this case $p_f < p_e - \sqrt{\varepsilon}$, we have that

$$\begin{aligned} \mathbb{E}[u(v, e')] &= \mathbb{E}[u(v, e)] + \sqrt{\varepsilon} \cdot p_e \geq \mathbb{E}[u(v, f)] - \varepsilon + \sqrt{\varepsilon} \cdot p_e = \mathbb{E}[u(v, f')] - \sqrt{\varepsilon} \cdot p_f - \varepsilon + \sqrt{\varepsilon} \cdot p_e \\ &= \mathbb{E}[u(v, f')] - \varepsilon + \sqrt{\varepsilon} \cdot (p_e - p_f) > \mathbb{E}[u(v, f')] - \varepsilon + \sqrt{\varepsilon} \cdot \sqrt{\varepsilon} = \mathbb{E}[u(v, f')], \end{aligned}$$

as required. \square

Applying Theorem 6.1 to the auction output by²⁶ Algorithm 1 yields Theorem 3.2.

6.2 Single-Parameter Settings

In this section, we will prove Theorem 3.3. The algorithm presented in Section 5 constitutes a black-box reductions from ε -BIC revenue maximization from samples to BIC revenue maximization from given distributions. As noted in the Introduction, the latter are mostly unsolved for more than one good. For a single good, however, the problem of DSIC/BIC revenue maximization was completely resolved by the seminal work of Myerson [1981] (who, in particular, showed that the optimal BIC mechanism is DSIC), and the computation complexity of the solution for discrete distributions was shown by Elkind [2007] to be polynomial.

Definition 6.2 (Myersonian Auction, Myerson [1981]). An n -bidder Myersonian auction (for valuations in $[0, H]$) is a tuple $(\phi_i)_{i \in N}$, where for every $i \in [n]$, $\phi_i : [0, H] \rightarrow \mathbb{R}$ is a nondecreasing function called the *ironed virtual valuation* of bidder i . The chosen outcome is $x \in X$ that maximizes $\sum_{i \in [n]} x_i \cdot \phi_i(v_i)$, with ties broken in a consistent manner. The payment is defined by the payment identity of Myerson [1981], which guarantees that the auction is IR and DSIC.

THEOREM 6.3 (MYERSON [1981]). *For every product distribution $\times_{i=1}^n W_i$, there exists a (DSIC) Myersonian auction $(\phi_i)_{i=1}^n$ that attains maximum revenue from $\times_{i=1}^n W_i$ among all IR and BIC auctions. Moreover, for every $i \in [n]$, the ironed virtual valuation ϕ_i depends only on W_i .*

THEOREM 6.4 (ELKIND [2007]). *Let $S \in \mathbb{N}$. There exists an algorithm that runs in time $\text{poly}(S)$, such that given a discrete distribution W with support of size at most S , outputs a nondecreasing function $\phi : \text{supp } W \rightarrow \mathbb{R}$, such that for every product $\times_{i=1}^n W_i$ of discrete distributions each having support of size at most S , the (DSIC) Myersonian auction $(\phi_i)_{i \in N}$ (where ϕ_i is the output of the algorithm given W_i) maximizes the expected revenue from $\times_{i=1}^n W_i$ among all IR and BIC auctions.*

Plugging²⁷ Theorem 6.4 into Algorithm 1 brings us closer (by making the algorithm efficient) to proving Theorem 3.3; however, this seems to still result in an ε -DSIC (rather than precisely DSIC) auction. Indeed, in the notation of Section 5, while μ is DSIC, it seems that μ^ε is only $O(m\varepsilon)$ -DSIC. To complete the proof of Theorem 3.3, we note that the latter is in fact exactly DSIC in this case.

²⁶Or, for a somewhat simpler analysis, to the auction output by Algorithm 2 in Appendix A. For readers interested in precise polynomial dependencies, in this case the allowed revenue deviation in Equation (5) could also be loosened to $\sqrt{mL\varepsilon}$ to save on samples, since the revenue loss from running the reduction of Theorem 6.1 would be far greater anyway.

²⁷Extending the mechanism μ returned by the optimizer in OptimizationOracle to be defined over all of $[0, H]_\varepsilon^{n \cdot m}$ as follows, though [following Gonczarowski and Nisan 2017]: for every $i \in [n]$ and $j \in [m]$ s.t. $v_{i,j} \notin \text{supp } W_{i,j}$, replace $v_{i,j}$ with $\max\{w_{i,j} \in \text{supp } W_{i,j} \mid w_{i,j} \leq v_{i,j}\}$.

Indeed, its allocation rule is monotone, and it satisfies the payment identity of Myerson [1981] for every bidder.

7 EXTENSIONS

7.1 On Computational Efficiency

As mentioned above, it is currently not known how to efficiently implement the optimization oracle, outputting an up-to- $O(\varepsilon)$ -optimal auction, as used in our algorithm. Nonetheless, there has been quite some work on efficiently finding auctions with weaker revenue guarantees with respect to optimal, such as guaranteeing some constant fraction of the optimal revenue. As the structured argument of our analysis provides a black-box reduction from BIC-revenue-maximization given a full distributions description to ε -BIC-revenue-maximization from samples, we can plug in any such algorithm into our analysis (in lieu of the optimization oracle) to obtain a learning algorithm with matching computational complexity and revenue degradation:

THEOREM 7.1 (“META THEOREM”: BLACK-BOX REDUCTION FOR EFFICIENT UP-TO-CONSTANT GUARANTEES). *If there exists a polynomial-time algorithm for Bayesian revenue-maximization up to a constant factor C given an explicitly specified finite product distribution, then for every $\varepsilon, \delta > 0$ and for every $\eta \leq \text{poly}(n, m, L, H, \varepsilon)$, there exists a polynomial-time algorithm that given $\text{poly}(n, m, L, H, 1/\varepsilon, 1/\eta, \log 1/\delta)$ samples from each $V_{i,j}$, with probability at least $1 - \delta$ outputs an IR and η -BIC auction that attains from $\times_{i,j} V_{i,j}$ expected revenue at most an additive ε smaller than a C fraction of that of any IR and η -BIC mechanism.*

7.2 Partial Correlations

In some settings, there could be partial correlations between the distributions of the values of each bidder for the various goods. Our construction and analysis can also be modified to analyze such settings, to obtain sample bounds that are polynomial in the independent dimensions. To give a few examples:

- If there are correlations across values of different goods for the same bidder, but different bidders’ valuations are independent, then our upper bound for $|M|$ becomes $(S + 1)^{n \cdot (\lceil H/\varepsilon \rceil)^m}$, and so our analysis would yield sample complexity that is polynomial in the number of bidders (as our upper bound of $|M|$ in this case is still singly exponential in the number of bidders) and exponential in the number of goods (as our upper bound of $|M|$ in this case is doubly exponential in the number of goods).
- If for each bidder i the values of every two goods $2j, 2j+1 \in [m]$ are correlated, but are independent of the values of these goods for any other bidder, and of the values of any other good for any bidder, then our upper bound for $|M|$ becomes $(S + 1)^{n \cdot m/2 \cdot (\lceil H/\varepsilon \rceil)^2}$, and so our analysis still yields sample complexity that is polynomial in the both the number of bidders and the number of goods (as our upper bound of $|M|$ in this case is still singly exponential in both parameters). This is an example for a weaker form of correlation for which our analysis can still yield sample complexity that is polynomial in all parameters.

APPENDICES

A DOMINANT-STRATEGY INCENTIVE COMPATIBILITY

In this appendix, we demonstrate a somewhat less involved version of our analysis and prove the following result, which we find interesting in its own right—a version of our main result for dominant strategy (rather than Bayesian) incentive compatibility. The proofs of supporting lemmas

are relegated to Appendix B. We start by phrasing a weak technical assumption that this result additionally requires.

Definition A.1 (Weakly Downward Closed). We say that the set of allowed allocations X is *weakly downward closed* if for every outcome $x \in X$ and for every $k \in [n]$, there exists $x' \in X$ such that a) $v_k(x') = v_k(x)$ for every $v_k = (v_{k,j})_{j \in [m]}$, and b) $v_i(x') = 0$ for every $i \neq k$ and $v_i = (v_{i,j})_{j \in [m]}$.²⁸

THEOREM A.2 (MAIN RESULT—DSIC VARIANT). *If X is weakly downward closed, then for every $\varepsilon, \delta > 0$, the sample complexity of learning an IR and ε -DSIC auction whose revenue differs from that of the optimal IR and DSIC auction by less than an additive ε is at most $\text{poly}(n, m, L, H, 1/\varepsilon, \log 1/\delta)$. That is, there exists a deterministic algorithm²⁹ that given $\text{poly}(n, m, L, H, 1/\varepsilon, \log 1/\delta)$ samples from each $V_{i,j}$, with probability $1 - \delta$ outputs an IR and ε -DSIC auction that attains from $\times_{i,j} V_{i,j}$ expected revenue at most an additive ε smaller than any IR and DSIC auction.*

PROOF. We assume that for every $i \in [n]$ and $j \in [m]$, we have S (to be determined later) independent samples $(v_{i,j}^s)_{s=1}^S$ from $V_{i,j}$. Algorithm 2 presents our learning algorithm, which is similar in nature to the one presented in Devanur et al. [2016] for certain single-parameter environments, however the analysis that we will use to show that it does not overfit the samples is completely different (even for single-parameter environments, where our analysis holds for arbitrary weakly downward closed allocation constraints—note that this assumption guarantees that the modification to the allocation rule of the mechanism returned by `OptimizationOracleDSIC` still results in allocations in X). We now analyze Algorithm 2. Note that for every i, j, s , we have that $w_{i,j}^s \sim [V_{i,j}]_\varepsilon$ independently.

Let \mathcal{V} be the set of all product distributions $\times_{i,j} W'_{i,j}$ where each $W'_{i,j}$ is the uniform distribution over some multiset of S values from $[0, H]_\varepsilon$. Let \mathcal{M} be the set of all mechanisms of the form `OptimizationOracleDSIC`($H, X, \varepsilon, \times_{i,j} W'_{i,j}$) for all $\times_{i,j} W'_{i,j} \in \mathcal{V}$. At the heart of our analysis is the observation that $|\mathcal{V}| < (S + 1)^{n \cdot m \cdot \lceil H/\varepsilon \rceil}$. (Crucially, this expression has S only in the base and not in the exponent!) Indeed, for every $\times_{i,j} W'_{i,j} \in \mathcal{V}$, for every i, j , and for every integer multiples of ε in $[0, H]$ (there are $\lceil H/\varepsilon \rceil$ many such values), the probability of this value in $W'_{i,j}$ can be any of the $S + 1$ values $0, 1/S, \dots, 1$. (The inequality on $|\mathcal{V}|$ is strict, since, for example, not all probabilities can be 0 simultaneously.) Therefore, $|\mathcal{M}| < (S + 1)^{n \cdot m \cdot \lceil H/\varepsilon \rceil}$.

Let OPT be the IR and DSIC auction defined over $\times_{i,j} \text{supp } V_{i,j}$ that maximizes the expected revenue (among such auctions) from $\times_{i,j} V_{i,j}$ (our learning algorithm cannot hope to find OPT , but in our analysis we may carefully reason about it, as it is nonetheless well defined; in particular, it does not depend on S). Let OPT_ε be the (randomized) mechanism defined over $\times_{i,j} [V_{i,j}]_\varepsilon$ as follows: let $w = (w_{i,j})_{i \in [n], j \in [m]}$ be an input valuation; for each i, j , independently draw $v_{i,j} \sim V_{i,j}|_{[w_{i,j}, w_{i,j} + \varepsilon]}$; let $(x, p) = \text{OPT}((v_{i,j})_{i \in [n], j \in [m]})$; the allocation of $\text{OPT}_\varepsilon(w)$ is x , and the payment of each bidder i is $p_i - mL\varepsilon$.

LEMMA A.3. *OPT_ε is an IR and $(2mL\varepsilon)$ -DSIC mechanism whose expected revenue from $\times_{i,j} [V_{i,j}]_\varepsilon$ is $nL\varepsilon$ smaller than the expected revenue of OPT from $\times_{i,j} V_{i,j}$.*

²⁸For example, for the multi-item setting this is satisfied if for every $x = (x_{i,j})_{i \in [n], j \in [m]} \in X$ and for every $i \in [n]$, we have that $(x_i, 0_{-i}) \in X$ as well.

²⁹Once again, recall that this result is information-theoretic and not computationally efficient (by necessity, without resolving major open problems), so our decision maker (seller) is computationally unbounded, and we allow the algorithm to make calls to any deterministic oracle that has no access to any $V_{i,j}$. In particular, we assume access to an oracle that can solve the revenue maximization problem on any precisely given $V'_{i,j}$ of finite support.

ALGORITHM 2: Empirical Multi-Parameter Up-to- ε DSIC Revenue Maximization.

Function EmpiricalOptimizeDSIC($H, X, \varepsilon, \delta, (v_{i,j}^s)_{i \in [n], j \in [m]}^{s \in [S]}$):

Input: For every $i \in [n], j \in [m]$, $(v_{i,j}^s)_{s=1}^S$ is a sequence of $S = \tilde{O}(\frac{H^2}{\varepsilon^2} \cdot (\log 1/\delta + \frac{nmH}{\varepsilon}))$ samples from $V_{i,j}$

Output: With probability $1 - \delta$, an IR and $(4mL\varepsilon)$ -DSIC mechanism defined over $[0, H]^{n \cdot m}$ with allocations in X , whose expected revenue from $\times_{i,j} V_{i,j}$ is up to an additive $4nmL\varepsilon$ smaller than that of any IR and DSIC mechanism defined over $\times_{i,j} \text{supp } W_{i,j}$ with allocations in X

for $i \in [n], j \in [m]$ **do**

for $s \in [S]$ **do**

$w_{i,j}^s \leftarrow \lfloor v_{i,j}^s \rfloor_\varepsilon$

end

$W_{i,j} \leftarrow \text{the uniform distribution over } (w_{i,j}^s)_{s=1}^S$

end

$\mu \leftarrow \text{OptimizationOracleDSIC}(H, X, \varepsilon, (W_{i,j})_{i \in [n], j \in [m]})$ // See definition below

return The mechanism that for input $(v_{i,j})_{i \in [n], j \in [m]}$ outputs $\mu(\lfloor v_{i,j} \rfloor_\varepsilon)_{i \in [n], j \in [m]}$, modified to charge each bidder $mL\varepsilon$ less.

Function OptimizationOracleDSIC($H, X, \varepsilon, (W_{i,j})_{i \in [n], j \in [m]}$):

Input: For every $i \in [n], j \in [m]$, $W_{i,j}$ is a distribution over $[0, H]_\varepsilon$

Output: An IR and $(2mL\varepsilon)$ -DSIC mechanism defined over $[0, H]_\varepsilon^{n \cdot m}$ with allocations in X , which maximizes the expected revenue from $\times_{i,j} W_{i,j}$ among all IR and $(2mL\varepsilon)$ -DSIC mechanisms defined over $\times_{i,j} \text{supp } W_{i,j}$ with allocations in X

$\mu \leftarrow \text{an IR and } (2mL\varepsilon)\text{-DSIC mechanism defined over } \times_{i,j} \text{supp } W_{i,j} \text{ with allocations in } X, \text{ which maximizes the expected revenue from } \times_{i,j} W_{i,j} \text{ among all such mechanisms}$

return The mechanism obtained by extending μ to be defined over all $(v_{i,j})_{i,j} \in [0, H]_\varepsilon^{n \cdot m}$ as follows: if $v_{k,j} \notin \text{supp } W_{k,j}$ for precisely one bidder k (and one or more parameters j), then an outcome is chosen such that bidder k 's valuation and payment are the same as in $\mu(v'_k, v_{-k})$, for $v'_k = (v'_{k,j})_{j \in [m]} \in \text{supp } \times_j W_{k,j}$ that maximizes $\mathbb{E}[u_k(v_k, \mu(v'_k, v_{-k}))]$, and such that all other bidders' valuations of the outcome are zero irrespective of their valuations, and all other bidders' payments are zero. If $v_{i,j} \notin \text{supp } W_{i,j}$ for more than one bidder i , then an outcome is chosen such that all bidders' valuations are zero irrespective of their valuations and all payments are zero.

We will choose S so that with probability at least $1 - \delta$, it simultaneously holds for all mechanisms $\mu \in M \cup \{\text{OPT}_\varepsilon\}$ that

$$|\text{Rev}_{\times_{i,j} W_{i,j}}(\mu) - \text{Rev}_{\times_{i,j} \lfloor V_{i,j} \rfloor_\varepsilon}(\mu)| \leq nmL\varepsilon. \quad (5)$$

By Theorem 2.2 (with $\ell \triangleq n \cdot m$, and note that any mechanism's revenue is bounded by $nmLH$), we have that for each mechanism $\mu \in M \cup \{\text{OPT}_\varepsilon\}$ separately, Equation (5) holds with probability at least $1 - \frac{4He^{-\varepsilon^2 S/(8H^2)}}{\varepsilon}$. Choosing S so that this probability is at least $1 - \frac{\delta}{|M|+1}$, we obtain that Equation (5) holds simultaneously for all mechanisms $\mu \in M \cup \{\text{OPT}_\varepsilon\}$ with probability at least $1 - \delta$. We now estimate S . Since $|M \cup \{\text{OPT}_\varepsilon\}| \leq (S+1)^{n \cdot m \cdot \lceil H/\varepsilon \rceil}$, we have that it is enough to take S such that

$$S \geq \frac{8H^2}{\varepsilon^2} \cdot (\log \frac{4H}{\varepsilon} + \log 1/\delta + nm \lceil H/\varepsilon \rceil \log(S+1)).$$

Therefore,³⁰ there exists an appropriate

$$S = O\left(\frac{H^2}{\varepsilon^2} \log 1/\delta + \frac{nmH^3}{\varepsilon^3} \log \frac{nmH^3}{\varepsilon^3}\right) = \tilde{O}\left(\frac{H^2}{\varepsilon^2} \cdot (\log 1/\delta + \frac{nmH}{\varepsilon})\right).$$

Let μ be the output of the call to `OptimizationOracleDSIC` in Algorithm 2 (note that it is well defined, since X is weakly downward closed), and let μ^ε be the final output of the algorithm (the output of `EmpiricalOptimizeDSIC`).

LEMMA A.4. μ^ε is an IR and $(4mL\varepsilon)$ -DSIC mechanism whose expected revenue from $\times_{i,j} V_{i,j}$ is $nmL\varepsilon$ smaller than the expected revenue of μ from $\times_{i,j} [V_{i,j}]_\varepsilon$.

So, we have that μ^ε is an IR and $(4m\varepsilon)$ -DSIC mechanism (by Lemma A.4) and that with probability at least $1 - \delta$:

$$\begin{aligned} & \text{Rev}_{\times_{i,j} V_{i,j}}(\mu^\varepsilon) && \text{(by Lemma A.4)} \\ &= \text{Rev}_{\times_{i,j} [V_{i,j}]_\varepsilon}(\mu) - nmL\varepsilon && \text{(by Equation (5) and since } \mu \in M) \\ &\geq \text{Rev}_{\times_{i,j} W_{i,j}}(\mu) - 2nmL\varepsilon && \text{(since } \text{OPT}_\varepsilon \text{ is IR and } (2m\varepsilon)\text{-DSIC and by optimality of } \mu) \\ &\geq \text{Rev}_{\times_{i,j} W_{i,j}}(\text{OPT}_\varepsilon) - 2nmL\varepsilon && \text{(by Equation (5) for } \text{OPT}_\varepsilon) \\ &\geq \text{Rev}_{\times_{i,j} [V_{i,j}]_\varepsilon}(\text{OPT}_\varepsilon) - 3nmL\varepsilon && \text{(by Lemma A.3)} \\ &= \text{Rev}_{\times_{i,j} V_{i,j}}(\text{OPT}) - 4nmL\varepsilon. \end{aligned}$$

Two main differences between the statements of Theorem A.2 and of Theorem 3.1 stand out: the technical requirement that X be weakly downward closed, and the fact that the learned mechanism has weaker incentive properties than the benchmark. In Section 5, we have avoided the latter within the context of Bayesian incentive compatibility via Theorem 2.1. It would be interesting to understand whether and to what extent a similar result also applies within the context of dominant-strategy incentive compatibility.

OPEN PROBLEM 2. For $n > 1$ bidders and $m > 1$ items, even with additive bidders, is the maximum revenue attainable by any IR and ε -DSIC auction from a given product distribution at most negligibly ($\text{poly}(\varepsilon) \cdot \text{poly}(n, m, H)$) greater than the maximum revenue attainable by any IR and DSIC auction from the same distribution?

Since the ε -dominant strategy incentive compatibility implies ε -Bayesian incentive compatibility, then the question in Open Problem 2 holds whenever the optimal BIC auction is also DSIC. Recently, Yao [2017] has shown that in general the revenue obtained by the optimal IR and BIC auctions can be higher than that obtained by the optimal IR and DSIC auctions—it is such cases in which Open Problem 2 is open and interesting. An affirmative result for Open Problem 2 would of course immediately imply an analogue of Open Problem 1 for DSIC auctions, which, if true, would allow Theorem A.2 to be strengthened to allow for learning a precisely DSIC auction.

OPEN PROBLEM 3. Given an IR and ε -DSIC auction and given some product distribution over $n > 1$ bidders and $m > 1$ items, even with additive bidders, is it possible to transform the given auction into an IR and (precisely) DSIC auction with negligible ($\text{poly}(\varepsilon) \cdot \text{poly}(n, m, H)$) revenue loss from that distribution using polynomially many samples from this product distribution?

³⁰See Footnote 23.

B OMITTED PROOFS

B.1 Proofs Omitted from Section 5

PROOF OF LEMMA 5.1. Individual Rationality: Let $k \in [n]$, let $w = (w_{i,j})_{i \in [n], j \in [m]}$ be a (rounded) valuation profile, and let $v = (v_{i,j})_{i \in [n], j \in [m]}$ be drawn as defined above. Then:

$$\begin{aligned}
 & \mathbb{E} [u_k(w_k, \text{OPT}_\varepsilon(w))] \\
 &= \mathbb{E} [u_k(w_k, \text{OPT}(v)) + mL\varepsilon] \quad (\text{since valuations are Lipschitz}) \\
 &\geq \mathbb{E} \left[u_k(v_k, \text{OPT}(v)) - \sum_{j \in [m]} L(v_{k,j} - w_{k,j}) + mL\varepsilon \right] \quad (\text{since } v_{k,j} < w_{k,j} + \varepsilon) \\
 &\geq \mathbb{E} [u_k(v_k, \text{OPT}(v))] \geq 0. \quad (\text{since OPT is IR})
 \end{aligned}$$

Incentive Compatibility: Let $k \in [n]$, let $w_k = (w_{k,j})_{j \in [m]}$ and $w'_k = (w'_{k,j})_{j \in [m]}$ be (rounded) valuations, and let $v_k = (v_{k,j})_{j \in [m]}$ and $v'_k = (v'_{k,j})_{j \in [m]}$ be respectively drawn as defined above. Then:

$$\begin{aligned}
 & \mathbb{E}_{w_{-k} \sim \times_{i,j \neq k} [V_{i,j}]_\varepsilon} [u_k(w_k, \text{OPT}_\varepsilon(w_k, w_{-k}))] \\
 &= \mathbb{E}_{v_{-k} \sim \times_{i,j \neq k} V_{i,j}} [u_k(w_k, \text{OPT}(v_k, v_{-k})) + mL\varepsilon] \quad (\text{since valuations are Lipschitz}) \\
 &\geq \mathbb{E}_{v_{-k} \sim \times_{i,j \neq k} V_{i,j}} \left[u_k(v_k, \text{OPT}(v_k, v_{-k})) - \sum_{j \in [m]} L(v_{k,j} - w_{k,j}) + mL\varepsilon \right] \quad (\text{since } v_{k,j} < w_{k,j} + \varepsilon) \\
 &\geq \mathbb{E}_{v_{-k} \sim \times_{i,j \neq k} V_{i,j}} [u_k(v_k, \text{OPT}(v_k, v_{-k}))] \quad (\text{since OPT is } (4mL\varepsilon)\text{-BIC for } \times_{i,j} V_{i,j}) \\
 &\geq \mathbb{E}_{v_{-k} \sim \times_{i,j \neq k} V_{i,j}} [u_k(v_k, \text{OPT}(v'_k, v_{-k}))] - 4mL\varepsilon \quad (\text{since valuations are Lipschitz}) \\
 &\geq \mathbb{E}_{v_{-k} \sim \times_{i,j \neq k} V_{i,j}} \left[u_k(w_k, \text{OPT}(v'_k, v_{-k})) - \sum_{j \in [m]} L(v_{k,j} - w_{k,j}) \right] - 4mL\varepsilon \\
 &\quad (\text{since } v_{k,j} < w_{k,j} + \varepsilon) \\
 &\geq \mathbb{E}_{v_{-k} \sim \times_{i,j \neq k} V_{i,j}} [u_k(w_k, \text{OPT}(v'_k, v_{-k}))] - 5mL\varepsilon \\
 &= \mathbb{E}_{w_{-k} \sim \times_{i,j \neq k} [V_{i,j}]_\varepsilon} [u_k(w_k, \text{OPT}_\varepsilon(w'_k, w_{-k})) - mL\varepsilon] - 5mL\varepsilon \\
 &= \mathbb{E}_{w_{-k} \sim \times_{i,j \neq k} [V_{i,j}]_\varepsilon} [u_k(w_k, \text{OPT}_\varepsilon(w'_k, w_{-k}))] - 6mL\varepsilon.
 \end{aligned}$$

Revenue: Let $\text{OPT}_\varepsilon = (x_\varepsilon, p_\varepsilon)$ and let $\text{OPT} = (x, p)$. Then:

$$\begin{aligned}
 \text{Rev}_{\times_{i,j} [V_{i,j}]_\varepsilon}(\text{OPT}_\varepsilon) &= \mathbb{E}_{w \sim \times_{i,j} [V_{i,j}]_\varepsilon} [p_\varepsilon(w)] \\
 &= \mathbb{E}_{v \sim \times_{i,j} V_{i,j}} \left[p(v) - \sum_{i \in [n]} mL\varepsilon \right] \\
 &= \mathbb{E}_{v \sim \times_{i,j} V_{i,j}} [p(v)] - nmL\varepsilon \\
 &= \text{Rev}_{\times_{i,j} V_{i,j}}(\text{OPT}) - nmL\varepsilon.
 \end{aligned}$$

□

PROOF OF LEMMA 5.2. Incentive Compatibility of OPT_ε : Let $k \in [n]$, let $w_k = (w_{k,j})_{j \in [m]}$ and $w'_k = (w'_{k,j})_{j \in [m]}$ be (rounded) valuations. Then:

$$\begin{aligned}
& \mathbb{E}_{w_{-k} \sim \times_{i \neq k} W_{i,j}} [u_k(w_k, \text{OPT}_\varepsilon(w_k, w_{-k}))] && \text{(by Equation (4) for } \text{OPT}_\varepsilon) \\
& \geq \mathbb{E}_{w_{-k} \sim \times_{i \neq k} [V_{i,j}]_\varepsilon} [u_k(w_k, \text{OPT}_\varepsilon(w_k, w_{-k}))] - mL\varepsilon && \text{(since } \text{OPT}_\varepsilon \text{ is } (6mL\varepsilon)\text{-BIC for } \times_{i,j} [V_{i,j}]_\varepsilon) \\
& \geq \mathbb{E}_{w_{-k} \sim \times_{i \neq k} [V_{i,j}]_\varepsilon} [u_k(w_k, \text{OPT}_\varepsilon(w'_k, w_{-k}))] - 7mL\varepsilon && \text{(by Equation (4) for } \text{OPT}_\varepsilon) \\
& \geq \mathbb{E}_{w_{-k} \sim \times_{i \neq k} W_{i,j}} [u_k(w_k, \text{OPT}_\varepsilon(w'_k, w_{-k}))] - 8mL\varepsilon.
\end{aligned}$$

Revenue: by Theorem 2.1 (for $\times_{i,j} W_{i,j}$), since OPT_ε is IR and $(8mL\varepsilon)$ -BIC for $\times_{i,j} W_{i,j}$ and by optimality of μ .

Incentive Compatibility of μ : Let $k \in [n]$, let $w_k = (w_{k,j})_{j \in [m]}$ and $w'_k = (w'_{k,j})_{j \in [m]}$ be (rounded) valuations. Then:

$$\begin{aligned}
& \mathbb{E}_{w_{-k} \sim \times_{i \neq k} [V_{i,j}]_\varepsilon} [u_k(w_k, \mu(w_k, w_{-k}))] && \text{(by Equation (4) and since } \mu \in M) \\
& \geq \mathbb{E}_{w_{-k} \sim \times_{i \neq k} W_{i,j}} [u_k(w_k, \mu(w_k, w_{-k}))] - mL\varepsilon && \text{(since } \mu \text{ is BIC for } \times_{i,j} W_{i,j}) \\
& \geq \mathbb{E}_{w_{-k} \sim \times_{i \neq k} W_{i,j}} [u_k(w_k, \mu(w'_k, w_{-k}))] - mL\varepsilon && \text{(by Equation (4) since } \mu \in M) \\
& \geq \mathbb{E}_{w_{-k} \sim \times_{i \neq k} [V_{i,j}]_\varepsilon} [u_k(w_k, \mu(w'_k, w_{-k}))] - 2mL\varepsilon. && \square
\end{aligned}$$

PROOF OF LEMMA 5.3. Individual Rationality: let $k \in [n]$ and let $v = (v_{i,j})_{i \in [n], j \in [m]}$ be a valuation profile. Then:

$$\begin{aligned}
& \mathbb{E} [u_k(v_k, \mu^\varepsilon(v))] \\
& = \mathbb{E} [u_k(v_k, \mu((\lfloor v_{i,j} \rfloor_\varepsilon)_{i \in [n], j \in [m]})) + mL\varepsilon] && \text{(since valuations are Lipschitz)} \\
& \geq \mathbb{E} \left[u_k((\lfloor v_{k,j} \rfloor_\varepsilon)_{j \in [m]}, \mu((\lfloor v_{i,j} \rfloor_\varepsilon)_{i \in [n], j \in [m]})) - \sum_{j \in [m]} L(v_{k,j} - \lfloor v_{k,j} \rfloor_\varepsilon) + mL\varepsilon \right] \\
& \geq \mathbb{E} [u_k((\lfloor v_{k,j} \rfloor_\varepsilon)_{j \in [m]}, \mu((\lfloor v_{i,j} \rfloor_\varepsilon)_{i \in [n], j \in [m]}))] \geq 0. && \text{(since } \mu \text{ is IR)}
\end{aligned}$$

Revenue: Let $\mu^\varepsilon = (x^\varepsilon, p^\varepsilon)$ and let $\mu = (x, p)$. Then:

$$\begin{aligned}
\text{Rev}_{\times_{i,j} V_{i,j}}(\mu^\varepsilon) &= \mathbb{E}_{v \sim \times_{i,j} V_{i,j}} [p^\varepsilon(v)] \\
&= \mathbb{E}_{w \sim \times_{i,j} [V_{i,j}]_\varepsilon} \left[p(w) - \sum_{i \in [n]} mL\varepsilon \right] \\
&= \mathbb{E}_{w \sim \times_{i,j} [V_{i,j}]_\varepsilon} [p(w)] - nmL\varepsilon \\
&= \text{Rev}_{\times_{i,j} [V_{i,j}]_\varepsilon}(\mu) - nmL\varepsilon.
\end{aligned}$$

Incentive Compatibility: Let $k \in [n]$, let $v_k = (v_{k,j})_{j \in [m]}$ and $v'_k = (v'_{k,j})_{j \in [m]}$ be valuations, and denote $\lfloor v_k \rfloor_\varepsilon = (\lfloor v_{k,j} \rfloor_\varepsilon)_{j \in [m]}$ and $\lfloor v'_k \rfloor_\varepsilon = (\lfloor v'_{k,j} \rfloor_\varepsilon)_{j \in [m]}$. Then:

$$\begin{aligned}
& \mathbb{E}_{v_{-k} \sim \times_{i,j \neq k} V_{i,j}} [u_k(v_k, \mu^\varepsilon(v_k, v_{-k}))] \\
&= \mathbb{E}_{w_{-k} \sim \times_{i,j \neq k} \lfloor V_{i,j} \rfloor_\varepsilon} [u_k(v_k, \mu(\lfloor v_k \rfloor_\varepsilon, w_{-k})) + mL\varepsilon] \quad (\text{since valuations are Lipschitz}) \\
&\geq \mathbb{E}_{w_{-k} \sim \times_{i,j \neq k} \lfloor V_{i,j} \rfloor_\varepsilon} \left[u_k(\lfloor v_k \rfloor_\varepsilon, \mu(\lfloor v_k \rfloor_\varepsilon, w_{-k})) - \sum_{j \in [m]} L(v_{k,j} - \lfloor v_{k,j} \rfloor_\varepsilon) + mL\varepsilon \right] \\
&\geq \mathbb{E}_{w_{-k} \sim \times_{i,j \neq k} \lfloor V_{i,j} \rfloor_\varepsilon} [u_k(\lfloor v_k \rfloor_\varepsilon, \mu(\lfloor v_k \rfloor_\varepsilon, w_{-k}))] \quad (\text{since } \mu \text{ is } (2mL\varepsilon)\text{-BIC for } \times_{i,j \neq k} \lfloor V_{i,j} \rfloor_\varepsilon) \\
&\geq \mathbb{E}_{w_{-k} \sim \times_{i,j \neq k} \lfloor V_{i,j} \rfloor_\varepsilon} [u_k(\lfloor v_k \rfloor_\varepsilon, \mu(\lfloor v'_k \rfloor_\varepsilon, w_{-k}))] - 2mL\varepsilon \quad (\text{since valuations are Lipschitz}) \\
&\geq \mathbb{E}_{w_{-k} \sim \times_{i,j \neq k} \lfloor V_{i,j} \rfloor_\varepsilon} \left[u_k(v_k, \mu(\lfloor v'_k \rfloor_\varepsilon, w_{-k})) - \sum_{j \in [m]} L(v_{k,j} - \lfloor v_{k,j} \rfloor_\varepsilon) \right] - 2mL\varepsilon \\
&\geq \mathbb{E}_{w_{-k} \sim \times_{i,j \neq k} \lfloor V_{i,j} \rfloor_\varepsilon} [u_k(v_k, \mu(\lfloor v'_k \rfloor_\varepsilon, w_{-k}))] - 3mL\varepsilon \\
&= \mathbb{E}_{v_{-k} \sim \times_{i,j \neq k} V_{i,j}} [u_k(v_k, \mu^\varepsilon(v_k, v_{-k}))] - 4mL\varepsilon. \quad \square
\end{aligned}$$

B.2 Proofs Omitted from Appendix A

PROOF OF LEMMA A.3. Individual Rationality: Let $k \in [n]$, let $w = (w_{i,j})_{i \in [n], j \in [m]}$ be a (rounded) valuation profile, and let $v = (v_{i,j})_{i \in [n], j \in [m]}$ be drawn as defined above. Then:

$$\begin{aligned}
& \mathbb{E} [u_k(w_k, \text{OPT}_\varepsilon(w))] \\
&= \mathbb{E} [u_k(w_k, \text{OPT}(v)) + mL\varepsilon] \quad (\text{since valuations are Lipschitz}) \\
&\geq \mathbb{E} \left[u_k(v_k, \text{OPT}(v)) - \sum_{j \in [m]} L(v_{k,j} - w_{k,j}) + mL\varepsilon \right] \quad (\text{since } v_{k,j} < w_{k,j} + \varepsilon) \\
&\geq \mathbb{E} [u_k(v_k, \text{OPT}(v))] \geq 0. \quad (\text{since OPT is IR})
\end{aligned}$$

Incentive Compatibility: Let $k \in [n]$, let $w = (w_{i,j})_{i \in [n], j \in [m]}$ be a (rounded) valuation profile, let $w'_k = (w'_{k,j})_{j \in [m]}$ be a (rounded) valuation, and let $v = (v_{i,j})_{i \in [n], j \in [m]}$ and $v'_k = (v'_{k,j})_{j \in [m]}$ be

respectively drawn as defined above. Then:

$$\begin{aligned}
& \mathbb{E} [u_k(w_k, \text{OPT}_\varepsilon(w_k, w_{-k}))] \\
&= \mathbb{E} [u_k(w_k, \text{OPT}(v_k, v_{-k})) + mL\varepsilon] && \text{(since valuations are Lipschitz)} \\
&\geq \mathbb{E} \left[u_k(v_k, \text{OPT}(v_k, v_{-k})) - \sum_{j \in [m]} L(v_{k,j} - w_{k,j}) + mL\varepsilon \right] && \text{(since } v_{k,j} < w_{k,j} + \varepsilon \text{)} \\
&\geq \mathbb{E} [u_k(v_k, \text{OPT}(v_k, v_{-k}))] && \text{(since OPT is DSIC)} \\
&\geq \mathbb{E} [u_k(v_k, \text{OPT}(v'_k, v_{-k}))] && \text{(since valuations are Lipschitz)} \\
&\geq \mathbb{E} \left[u_k(w_k, \text{OPT}(v'_k, v_{-k})) - \sum_{j \in [m]} L(v_{k,j} - w_{k,j}) \right] && \text{(since } v_{k,j} < w_{k,j} + \varepsilon \text{)} \\
&\geq \mathbb{E} [u_k(w_k, \text{OPT}(v'_k, v_{-k}))] - mL\varepsilon \\
&= \mathbb{E} [u_k(w_k, \text{OPT}_\varepsilon(w'_k, w_{-k})) - mL\varepsilon] - mL\varepsilon \\
&= \mathbb{E} [u_k(w_k, \text{OPT}_\varepsilon(w'_k, w_{-k}))] - 2mL\varepsilon.
\end{aligned}$$

Revenue: Let $\text{OPT}_\varepsilon = (x_\varepsilon, p_\varepsilon)$ and let $\text{OPT} = (x, p)$. Then:

$$\begin{aligned}
\text{Rev}_{\times_{i,j} [v_{i,j}]_\varepsilon}(\text{OPT}_\varepsilon) &= \mathbb{E}_{w \sim \times_{i,j} [v_{i,j}]_\varepsilon} [p_\varepsilon(w)] \\
&= \mathbb{E}_{v \sim \times_{i,j} v_{i,j}} \left[p(v) - \sum_{i \in [n]} mL\varepsilon \right] \\
&= \mathbb{E}_{v \sim \times_{i,j} v_{i,j}} [p(v)] - nmL\varepsilon \\
&= \text{Rev}_{\times_{i,j} v_{i,j}}(\text{OPT}) - nmL\varepsilon. \quad \square
\end{aligned}$$

PROOF OF LEMMA A.4. Individual Rationality: let $k \in [n]$ and let $v = (v_{i,j})_{i \in [n], j \in [m]}$ be a valuation profile. Then:

$$\begin{aligned}
& \mathbb{E} [u_k(v_k, \mu^\varepsilon(v))] \\
&= \mathbb{E} [u_k(v_k, \mu((\lfloor v_{i,j} \rfloor_\varepsilon)_{i \in [n], j \in [m]})) + mL\varepsilon] && \text{(since valuations are Lipschitz)} \\
&\geq \mathbb{E} \left[u_k((\lfloor v_{k,j} \rfloor_\varepsilon)_{j \in [m]}, \mu((\lfloor v_{i,j} \rfloor_\varepsilon)_{i \in [n], j \in [m]})) - \sum_{j \in [m]} L(v_{k,j} - \lfloor v_{k,j} \rfloor_\varepsilon) + mL\varepsilon \right] \\
&\geq \mathbb{E} [u_k((\lfloor v_{k,j} \rfloor_\varepsilon)_{j \in [m]}, \mu((\lfloor v_{i,j} \rfloor_\varepsilon)_{i \in [n], j \in [m]}))] \geq 0. && \text{(since } \mu \text{ is IR)}
\end{aligned}$$

Incentive Compatibility: Let $k \in [n]$, let $v = (v_{i,j})_{i \in [n], j \in [m]}$ be a valuation profile, let $v'_k = (v'_{k,j})_{j \in [m]}$ be a valuation, and denote $\lfloor v_k \rfloor_\varepsilon = (\lfloor v_{k,j} \rfloor_\varepsilon)_{j \in [m]}$, $\lfloor v'_k \rfloor_\varepsilon = (\lfloor v'_{k,j} \rfloor_\varepsilon)_{j \in [m]}$ and $\lfloor v_{-k} \rfloor_\varepsilon =$

$(\lfloor v_{i,j} \rfloor_\varepsilon)_{i \in [n] \setminus \{k\}, j \in [m]}$. Then:

$$\begin{aligned}
& \mathbb{E} [u_k(v_k, \mu^\varepsilon(v_k, v_{-k}))] \\
&= \mathbb{E} [u_k(v_k, \mu(\lfloor v_k \rfloor_\varepsilon, \lfloor v_{-k} \rfloor_\varepsilon)) + mL\varepsilon] \quad (\text{since valuations are Lipschitz}) \\
&\geq \mathbb{E} \left[u_k(\lfloor v_k \rfloor_\varepsilon, \mu(\lfloor v_k \rfloor_\varepsilon, \lfloor v_{-k} \rfloor_\varepsilon)) - \sum_{j \in [m]} L(v_{k,j} - \lfloor v_{k,j} \rfloor_\varepsilon) + mL\varepsilon \right] \\
&\geq \mathbb{E} [u_k(\lfloor v_k \rfloor_\varepsilon, \mu(\lfloor v_k \rfloor_\varepsilon, \lfloor v_{-k} \rfloor_\varepsilon))] \quad (\text{since } \mu \text{ is } (2mL\varepsilon)\text{-DSIC}) \\
&\geq \mathbb{E} [u_k(\lfloor v_k \rfloor_\varepsilon, \mu(\lfloor v'_k \rfloor_\varepsilon, \lfloor v_{-k} \rfloor_\varepsilon))] - 2mL\varepsilon \quad (\text{since valuations are Lipschitz}) \\
&\geq \mathbb{E} \left[u_k(v_k, \mu(\lfloor v'_k \rfloor_\varepsilon, \lfloor v_{-k} \rfloor_\varepsilon)) - \sum_{j \in [m]} L(v_{k,j} - \lfloor v_{k,j} \rfloor_\varepsilon) \right] - 2mL\varepsilon \\
&\geq \mathbb{E} [u_k(v_k, \mu(\lfloor v'_k \rfloor_\varepsilon, \lfloor v_{-k} \rfloor_\varepsilon))] - 3mL\varepsilon \\
&= \mathbb{E} [u_k(v_k, \mu^\varepsilon(v'_k, v_{-k}))] - 4mL\varepsilon.
\end{aligned}$$

Revenue: Let $\mu^\varepsilon = (x^\varepsilon, p^\varepsilon)$ and let $\mu = (x, p)$. Then:

$$\begin{aligned}
\text{Rev}_{\times_{i,j} V_{i,j}}(\mu^\varepsilon) &= \mathbb{E}_{v \sim \times_{i,j} V_{i,j}} [p^\varepsilon(v)] \\
&= \mathbb{E}_{w \sim \times_{i,j} \lfloor V_{i,j} \rfloor_\varepsilon} \left[p(w) - \sum_{i \in [n]} mL\varepsilon \right] \\
&= \mathbb{E}_{w \sim \times_{i,j} \lfloor V_{i,j} \rfloor_\varepsilon} [p(w)] - nmL\varepsilon \\
&= \text{Rev}_{\times_{i,j} \lfloor V_{i,j} \rfloor_\varepsilon}(\mu) - nmL\varepsilon. \quad \square
\end{aligned}$$

ACKNOWLEDGMENTS

We thank Noam Nisan and Costis Daskalakis for helpful conversations and thank the anonymous referees for helpful feedback.

REFERENCES

- Moshe Babaioff, Yannai A. Gonczarowski, and Noam Nisan. 2017. The menu-size complexity of revenue approximation. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing (STOC'17)*. 869–877.
- Moshe Babaioff, Nicole Immorlica, Brendan Lucier, and S. Matthew Weinberg. 2014. A simple and approximately optimal mechanism for an additive buyer. In *Proceedings of the 55th Annual IEEE Symposium on Foundations of Computer Science (FOCS'14)*. 21–30.
- Yakov Babichenko, Siddharth Barman, and Ron Peretz. 2017. Empirical distribution of equilibrium play and its testing application. *Math. Operat. Res.* 42, 1 (2017), 15–29. Preliminary version (“Simple approximate equilibria in large games”) in *Proceedings of the 15th ACM Conference on Economics and Computation (EC)*. 2014.
- Maria-Florina Balcan, Avrim Blum, Jason D. Hartline, and Yishay Mansour. 2005. Mechanism design via machine learning. In *Proceedings of the 46th Annual IEEE Symposium on the Foundations of Computer Science (FOCS'05)*. 426–439. Full version: Technical Report CMU-CS-05-143.
- Maria-Florina Balcan, Tuomas Sandholm, and Ellen Vitercik. 2016. Sample complexity of automated mechanism design. In *Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS'16)*. 2083–2091.
- Maria-Florina Balcan, Tuomas Sandholm, and Ellen Vitercik. 2018. A general theory of sample complexity for multi-item profit maximization. In *Proceedings of the 19th ACM Conference on Economics and Computation (EC'18)*. 173–174.
- Xiaohui Bei and Zhiyi Huang. 2011. Bayesian incentive compatibility via fractional assignments. In *Proceedings of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'11)*. 720–733.
- Yang Cai and Constantinos Daskalakis. 2017. Learning multi-item auctions with (or without) samples. In *Proceedings of the 58th IEEE Annual Symposium on Foundations of Computer Science (FOCS'17)*. 516–527.

- Yang Cai, Constantinos Daskalakis, and S. Matthew Weinberg. 2012. Optimal multi-dimensional mechanism design: Reducing revenue to welfare maximization. In *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS'12)*. 130–139.
- Yang Cai, Nikhil Devanur, and S. Matthew Weinberg. 2016. A duality based unified approach to Bayesian mechanism design. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing (STOC'16)*. 926–939.
- Yang Cai and Mingfei Zhao. 2017. Simple mechanisms for subadditive buyers via duality. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing (STOC'17)*. 170–183.
- Shuchi Chawla, Jason D. Hartline, and Robert Kleinberg. 2007. Algorithmic pricing via virtual valuations. In *Proceedings of the 8th ACM Conference on Electronic Commerce (EC'07)*. 243–251.
- Shuchi Chawla, Jason D. Hartline, David L. Malec, and Balasubramanian Sivan. 2010. Multi-parameter mechanism design and sequential posted pricing. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing (STOC'10)*. 311–320.
- Shuchi Chawla, David L. Malec, and Balasubramanian Sivan. 2015. The power of randomness in Bayesian optimal mechanism design. *Games Econ. Behav.* 91 (2015), 297–317. Preliminary version in *Proceedings of the 11th ACM Conference on Electronic Commerce (EC)*. 2010.
- Shuchi Chawla and J. Benjamin Miller. 2016. Mechanism design for subadditive agents via an ex ante relaxation. In *Proceedings of the 17th ACM Conference on Economics and Computation (EC'16)*. 579–596.
- Richard Cole and Tim Roughgarden. 2014. The sample complexity of revenue maximization. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC'14)*. 243–252.
- Constantinos Daskalakis, Alan Deckelbaum, and Christos Tzamos. 2013. Mechanism design via optimal transport. In *Proceedings of the 14th ACM Conference on Electronic Commerce (EC'13)*. 269–286.
- Constantinos Daskalakis and S. Matthew Weinberg. 2012. Symmetries and optimal multi-dimensional mechanism design. In *Proceedings of the 13th ACM Conference on Electronic Commerce (EC'12)*. 370–387.
- Nikhil R. Devanur, Zhiyi Huang, and Christos-Alexandros Psomas. 2016. The sample complexity of auctions with side information. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing (STOC'16)*. 426–439.
- Peerapong Dhangwatnotai, Tim Roughgarden, and Qiqi Yan. 2015. Revenue maximization with a single sample. *Games Econ. Behav.* 91 (2015), 318–333.
- Shaddin Dughmi, Li Han, and Noam Nisan. 2014. Sampling and representation complexity of revenue maximization. In *Proceedings of the 10th Conference on Web and Internet Economics (WINE'14)*. 277–291.
- Shaddin Dughmi, Jason D. Hartline, Robert Kleinberg, and Rad Niazadeh. 2017. Bernoulli factories and black-box reductions in mechanism design. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing (STOC'17)*. 158–169.
- Edith Elkind. 2007. Designing and learning optimal finite support auctions. In *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'07)*. 736–745.
- Evangelia Gergatsouli, Brendan Lucier, and Christos Tzamos. 2019. The complexity of black-box mechanism design with priors. In *Proceedings of the 20th ACM Conference on Economics and Computation (EC'19)*. 869–883.
- Yannai A. Gonczarowski. 2018. Bounding the menu-size of approximately optimal auctions via optimal-transport duality. In *Proceedings of the 50th Annual ACM Symposium on Theory of Computing (STOC'18)*. 123–131.
- Yannai A. Gonczarowski and Noam Nisan. 2017. Efficient empirical revenue maximization in single-parameter auction environments. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing (STOC'17)*. 856–868.
- Yannai A. Gonczarowski and S. Matthew Weinberg. 2018. The sample complexity of up-to- ϵ multi-dimensional revenue maximization. In *Proceedings of the 59th Annual IEEE Symposium on Foundations of Computer Science (FOCS'18)*. 416–426.
- Chenghao Guo, Zhiyi Huang, Zhihao Gavin Tang, and Xinzhi Zhang. 2019. Generalizing Complex Hypotheses on Product Distributions: Auctions, Prophet Inequalities, and Pandora's Problem. arXiv:1911.11936. Retrieved from <https://arxiv.org/abs/1911.11936>.
- Sergiu Hart and Noam Nisan. 2012. Approximate revenue maximization with multiple items. In *Proceedings of the 13th ACM Conference on Electronic Commerce (EC'12)*. 656.
- Sergiu Hart and Noam Nisan. 2013. The menu-size complexity of auctions. In *Proceedings of the 14th ACM Conference on Electronic Commerce (EC'12)*. 565.
- Sergiu Hart and Philip J. Reny. 2015. Maximal revenue with multiple goods: Nonmonotonicity and other observations. *Theor. Econ.* 10, 3 (2015), 893–922.
- Jason Hartline and Samuel Taggart. 2019. Sample complexity for non-truthful mechanisms. In *Proceedings of the 20th ACM Conference on Economics and Computation (EC'19)*. 399–416.
- Jason D. Hartline, Robert Kleinberg, and Azarakhsh Malekian. 2011. Bayesian incentive compatibility via matchings. In *Proceedings of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'11)*. 734–747.
- Zhiyi Huang, Yishay Mansour, and Tim Roughgarden. 2015. Making the most of your samples. In *Proceedings of the 16th ACM Conference on Economics and Computation (EC'15)*. 45–60.

- Pravesh Kothari, Divyarthi Mohan, Ariel Schwartzman, Sahil Singla, and S. Matthew Weinberg. 2019. Approximation schemes for a buyer with independent items via symmetries. In *Proceedings of the 60th Annual IEEE Symposium on Foundations of Computer Science (FOCS'19)*.
- Kristóf Maderász and Andrea Prat. 2017. Sellers with misspecified models. *The Review of Economic Studies* 84, 2 (2017), 790–815.
- Jamie Morgenstern and Tim Roughgarden. 2015. On the pseudo-dimension of nearly optimal auctions. In *Proceedings of the 29th Annual Conference on Neural Information Processing Systems (NIPS'15)*. 136–144.
- Jamie Morgenstern and Tim Roughgarden. 2016. Learning simple auctions. In *Proceedings of the 29th Annual Conference on Learning Theory (COLT'16)*. 1298–1318.
- Roger Myerson. 1981. Optimal auction design. *Math. Oper. Res.* 6, 1 (1981), 58–73.
- Tim Roughgarden and Okke Schrijvers. 2016. Ironing in the dark. In *Proceedings of the 17th ACM Conference on Economics and Computation (EC'16)*. 1–18.
- Aviad Rubinfeld and S. Matthew Weinberg. 2015. Simple mechanisms for a combinatorial buyer and applications to revenue monotonicity. In *Proceedings of the 16th ACM Conference on Economics and Computation (EC'15)*. 377–394.
- Vasilis Syrgkanis. 2017. A sample complexity measure with applications to learning optimal auctions. In *Proceedings of the 31th Annual Conference on Neural Information Processing Systems (NIPS'17)*. 5352–5359.
- John Thanassoulis. 2004. Hagglng over substitutes. *J. Econ. Theory* 117, 2 (2004), 217–245.
- Andrew Chi-Chih Yao. 2015. An n -to-1 bidder reduction for multi-item auctions and its applications. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'15)*. 92–109.
- Andrew Chi-Chih Yao. 2017. Dominant-strategy versus Bayesian multi-item auctions: Maximum revenue determination and comparison. In *Proceedings of the 18th ACM Conference on Economics and Computation (EC'17)*. 3–20.

Received December 2019; revised September 2020; accepted November 2020