Sample Complexity of Single-parameter Revenue Maximization

CHENGHAO GUO
IIIS, Tsinghua University
and
ZHIYI HUANG
The University of Hong Kong
and
XINZHI ZHANG
IIIS, Tsinghua University

The sample complexity of learning Myerson's optimal auction from i.i.d. samples of bidders' values has received much attention since its introduction by Cole and Roughgarden (STOC 2014). This letter gives a brief introduction of a recent work that settles the sample complexity by showing matching upper and lower bounds, up to a poly-logarithmic factor, for all families of value distributions that have been considered in the literature. The upper bounds are unified under a novel framework, which builds on the strong revenue monotonicity by Devanur, Huang, and Psomas (STOC 2016), and an information theoretic argument. This is fundamentally different from the previous approaches that rely on either constructing an ϵ-net of the mechanism space, either explicitly, or implicitly via statistical learning theory, or learning an approximately accurate version of the virtual values. To our knowledge, it is the first time information theoretical arguments are used to show sample complexity upper bounds, instead of lower bounds. The lower bounds are also unified under a meta construction of hard instances.

Categories and Subject Descriptors: J.4 [**Social and Behavioral Sciences**]: Economics

General Terms: Algorithms, Economics, Theory

Additional Key Words and Phrases: Sample Complexity, Revenue Maximization, Myerson Auction

1. INTRODUCTION

Alice is a college student majoring in Economics. In an effort to raise fund for her road trip in the upcoming summer, she decides to sell her old smart phone on the internet. After an hour of research, Alice finds two options. The first one is eBay, which supports an auction format that is essentially the second price auction with a reserve. Having collected the bids from different bidders, it gives the phone to the bidder with the highest bid so long as it is at least the reserve price, and charges a price that is either the reserve price or the second highest bid, whichever is higher. The second option is Yabe, a new startup platform that supports arbitrary auction formats. As a student in Economics, Alice feels obliged to take the second option and to put the theory that she learns into practice.

Alice recalls that the optimal auction by [Myerson 1981] would maximize the

Authors' addresses: guoch16@mails.tsinghua.edu.cn, zhiyi@cs.hku.hk, and zhang-xz16@mails.tsinghua.edu.cn

revenue, if she knows who the bidders are and if further *the distributions from which their values for the phone are drawn are explicitly given.* In a nutshell, the optimal auction works as follows. For each bidder i, suppose F_i and f_i are the cumulative distribution function (cdf) and probability density function (pdf) of her value distributions, and v_i is the realized value. Then, it maps v_i to a virtual value:

$$\phi_i(v_i) = v_i - \frac{1 - F_i(v_i)}{f_i(v_i)} \ , \tag{1}$$

and gives the phone to the bidder with the highest nonnegative virtual value;[1] the winner pays the minimum bid that would still make her win.

Bayesian Model. More formally, suppose there is a single type of item for sale. Let there be n bidders. Each bidder i has a value for an item, $v_i \geq 0$, that is independently drawn from a Bayesian prior D_i. The prior $\vec{D} = D_1 \times D_2 \times \cdots \times D_n$ is publicly known to the seller and all bidders; the realization of value v_i, however, is private information of bidder i.

For simplicity, most parts of this letter consider the single-item setting, where the seller has only one copy of the item and, therefore, can allocate to at most one bidder. Nonetheless, we will briefly explain at the end of the letter how the techniques and results generalize to problems with multiple copies of the item, and even more generally to the matroid setting.

Digging deeper into Yabe, Alice finds that it does provide certain information about the bidders, but not in the forms of explicitly given priors from which their values are drawn, let alone the cdf's and pdf's. Instead, the types of the bidders, e.g., according to the demographic information, browsing history, etc., and certain data about each type of bidders, e.g., past bids by the same type of bidders for second-hand smart phones, market research, etc., are given by Yabe. How can Alice design an auction based on the data such that the revenue is close to the optimal revenue by Myerson's auction?

Sample Complexity Model. Under an idealized assumption that the data are independent and identically distributed (i.i.d.) samples from the prior, it becomes the sample complexity model by [Cole and Roughgarden 2014]. Let A be an algorithm that takes i.i.d. samples as input and outputs a truthful auction. For any family of distributions, e.g., those that are regular, the minimum number of samples such that the auction returned by algorithm A is a $(1-\epsilon)$-approximation in revenue, so long as the prior comes from the family, is called the sample complexity of the algorithm, with respect to the family of distributions. The minimum sample complexity achievable by any algorithm is called the sample complexity of the family of distributions.

Of late, the sample complexity of revenue maximizing auctions has received much attention in the Algorithmic Game Theory community. This letter presents a brief

[1] This is true only for regular distributions, for which the virtual value $\phi_i(v_i)$ is nondecreasing in the value v_i; in general, an additional procedure known as ironing is needed to make it monotone. We omit it for brevity as it plays a minor role in the sample complexity problem in this letter.

introduction of a recent result of ours in [Guo et al. 2019] that pins down the sample complexity of essentially all families of distributions that have been considered in the literature, under a unifying algorithm and analysis framework.

2. EXISTING APPROACHES AND OBSTACLES

We first present a brief overview on the previous approaches for analyzing the sample complexity of revenue maximization, which can be categorized into two groups, and explain their limitations.

Statistical Learning Theory. The first approach relies on constructing an ϵ-net of the mechanism space, namely, a subset of mechanisms such that for any distribution in the family, there always exists an approximately optimal mechanism in the subset. Then, it remains to identify such an approximately optimal mechanism in the ϵ-net. This can be done via a a standard combination of concentration plus union bounds. Informally, the resulting sample complexity will be:[2]

$$\frac{\log\big(\text{size of the } \epsilon\text{-net}\big)}{\epsilon^2} .$$

The construction of the ϵ-net can be either explicit (e.g., [Devanur et al. 2016], [Gonczarowski and Nisan 2017], [Gonczarowski and Weinberg 2018]), or implicit via various learning dimensions from statistical learning theory (e.g., [Morgenstern and Roughgarden 2015], [Syrgkanis 2017]).

The main limitation of this approach is that the size of the ϵ-net seems to have an unavoidable exponential dependence in ϵ^{-1} (see below for an example). Recall that the sample complexity upper bound will be $\log(\text{size of the } \epsilon - \text{net})/\epsilon^2$, this exponential dependence leads to an *at least cubic dependence in* ϵ^{-1} in the sample complexity upper bounds. For example, we sketch below an explicit construction of the ϵ-net by [Devanur et al. 2016]. With an appropriate discretization, it suffices to consider ϵ^{-1} distinct values. Further, since the optimal auction chooses the winner to maximize virtual value, it suffices to know the ordering of 0 and $\phi_i(v)$'s for all n bidders and all ϵ^{-1} values. Hence, the number of auctions that we need to consider is no more than the number of orderings over the $n\epsilon^{-1}$ virtual values $\phi_i(v)$'s and 0, which equals $(n\epsilon^{-1} + 1)!$ and is singly exponential in both n and ϵ^{-1}. Getting rid of the exponential dependence in ϵ^{-1} intuitively means that it suffices to consider a constant number of distinct values, which seems implausible.

Learning the Virtual Values. An alternative approach (e.g., [Cole and Roughgarden 2014], [Roughgarden and Schrijvers 2016]) is to learn the individual value distributions well enough to obtain enough approximately accurate information about the virtual values, which induces a mechanism. Then, we analyze the revenue approximation using the connections between expected revenue and virtual values. Importantly, this approach does not need to take a union bound over exponentially many candidate mechanisms, circumventing the bottleneck that introduces the undesirable cubic dependence in ϵ^{-1} in the learning theory approach. Indeed, for the

[2] This form relies on the assumption that $O(\epsilon^{-2})$ samples are sufficient for estimating the expected revenue of a mechanism up to an ϵ error, which need not be true in general especially with unbounded value distributions.

Setting	Lower Bound	Upper Bound
Regular	$\Omega(\max\{n\epsilon^{-1}, \epsilon^{-3}\})$ [ad]	$\tilde{O}(n\epsilon^{-4})$ [b]
MHR	$\Omega(\max\{n\epsilon^{-1/2}, \epsilon^{-3/2}\})$ [ad]	$\tilde{O}(n\epsilon^{-3})$ [be]
$[1, H]$	$\Omega(H\epsilon^{-2})$ [d]	$\tilde{O}(nH\epsilon^{-3})$ [b]
$[0, 1]$-additive	$\Omega(\epsilon^{-2})$ [d]	$\tilde{O}(n\epsilon^{-3})$ [bc]

[a] [Cole and Roughgarden 2014] [b] [Devanur et al. 2016] [c] [Gonczarowski and Nisan 2017]
[d] [Huang et al. 2015] [e] [Morgenstern and Roughgarden 2015]

Table I. Best known sample complexity bounds prior to [Guo et al. 2019]

special case of independently and identically distributed (i.i.d.) bidders with $[0, 1]$-bounded distributions and additive approximation, [Roughgarden and Schrijvers 2016] showed a sample complexity upper bound of $\tilde{O}(n^2\epsilon^{-2})$, which is the only previous example, to our knowledge, with a sub-cubic dependence in ϵ^{-1}.

The main limitation of this approach roots in the form of the virtual value as defined in Eqn. (1). It involves three components, the value v, the complementary cumulative distribution function $1 - F_i(v)$, a.k.a., the quantile, and the pdf $f_i(v)$. Here, the value v is given as input; the quantile $1-F_i(v)$ is relatively easy to estimate accurately via standard concentration inequalities. It is, however, impossible to get an accurate estimation of the density function f_i in general. As a result, it is infeasible to learn the virtual values accurately point-wise. This is a major technical hurdle that prevents existing works using this approach from getting tight sample complexity upper bounds; in particular, they all have *super-linear dependence in* n. Even for the special case of i.i.d. bidders, the bound is quadratic in n [Roughgarden and Schrijvers 2016]; the dependence is at least n^7 for the general case [Cole and Roughgarden 2014]. Note that a linear dependence in n follows almost trivially from the learning theory approach (e.g., [Devanur et al. 2016]).

Prior Knowledge of the Distribution Family. Another limitation of the existing approaches is that they generally rely on knowing the family of distributions upfront. Even for the special case of a single bidder, the best known algorithms are different for regular, MHR, and bound-support distributions (e.g., [Huang et al. 2015]). For MHR distributions, we may simply pick the optimal price w.r.t. the empirical distribution, i.e., the uniform distribution over the samples. For regular and $[1, H]$ bounded support distributions, however, we need to introduce a threshold $\delta > 0$ and to choose the optimal price subject to having a sale probability at least δ. Further, the threshold is chosen differently for regular and $[1, H]$ bounded support distributions. If we fail to introduce a threshold when it is an arbitrary regular distribution, the expected revenue may not converge to the optimal at all [Dhangwatnotai et al. 2015]. If we set the threshold under the belief that the distribution has a $[1, H]$ bounded support while it is in fact an arbitrary regular distribution, the convergence rate will be far from optimal. It would definitely be nice to have a more robust algorithm.

Other Related Works. Prior to [Cole and Roughgarden 2014], works such as [Elkind 2007] [Dhangwatnotai et al. 2015] had the flavor of learning the optimal price/auctions from samples, but not yet fitting into the language of sample complexity.

Setting	Lower Bound	Upper Bound
Regular	$\Omega(n\epsilon^{-3})$	$\tilde{O}(n\epsilon^{-3})$
MHR	$\tilde{\Omega}(n\epsilon^{-2})$	$\tilde{O}(n\epsilon^{-2})$
$[1,H]$	$\Omega(nH\epsilon^{-2})$	$\tilde{O}(nH\epsilon^{-2})$
$[0,1]$-additive	$\Omega(n\epsilon^{-2})$	$\tilde{O}(n\epsilon^{-2})$

Table II. Sample complexity bounds in [Guo et al. 2019]

The learning theory approach has also been used to learn approximately optimal auction among a restricted family of simple auctions, both for single-parameter problems (e.g., [Morgenstern and Roughgarden 2016]), and for multi-parameter problems (e.g., [Morgenstern and Roughgarden 2015], [Cai and Daskalakis 2017], [Balcan et al. 2016], [Balcan et al. 2018], [Syrgkanis 2017]).

For the more general multi-parameter setting, [Dughmi et al. 2014] showed that learning an approximately optimal auction needed exponentially many samples in general; [Gonczarowski and Weinberg 2018] proved a polynomial sample complexity upper bound if the bidders' valuations have polynomially many dimensions (note that it could be exponential in general, e.g., in combinatorial auctions), and if we allowed approximate truthfulness. Their approach achieved exact truthfulness in the special case of single-parameter problems considered in this letter.

The online learning version has also been considered, both in the full information setting, i.e., the seller runs a direction revelation auction and observes the bidder's valuation, and in the bandit setting, i.e., the seller runs a posted price auction and only observes if the bidder buys the item. [Blum and Hartline 2005] introduced the optimal algorithm in terms of a regret bound that scaled with H, the upper bound on bidders' values. [Bubeck et al. 2017] further improved the regret bound to scale with the optimal price instead of H, and their algorithm matched the optimal sample complexity bounds when the bidder's values in different rounds were i.i.d. from a prior distribution.

Intriguingly, [Hart and Reny 2015] showed that even the weaker notion of revenue monotonicity ceased to hold in multi-parameter problems. On the other hand, approximate versions of revenue monotonicity were known for restricted families of valuations (e.g., [Rubinstein and Weinberg 2015], [Yao 2018]).

3. RESULTS AND TECHNIQUES IN A NUTSHELL

We introduce an algorithm that achieves the optimal sample complexity, up to a poly-logarithmic factor, simultaneously for all families of distributions that have been considered in the literature. Our upper and lower bounds, summarized in Table II, improve the best known bounds in all cases.

Our Algorithm. The algorithm constructs from the samples a dominated empirical distribution, denoted as $\tilde{\mathbf{E}} = \tilde{E}_1 \times \tilde{E}_2 \times \cdots \times \tilde{E}_n$, which is dominated by the true value distribution $\mathbf{D}$ in the sense of first-order stochastic dominance, but is as close to $\mathbf{D}$ as possible. Then, it chooses the optimal mechanism w.r.t. $\tilde{\mathbf{E}}$. We call it the dominated empirical Myerson auction.

To construct the dominated empirical distribution, we first look at the estimation error by the empirical distribution, in terms of the difference between the empirical quantiles and the true quantiles. This can be bounded using standard concentration inequalities. For example, suppose a value v has quantile q. Then, Bernstein's inequality gives that, with high probability, its quantile in the empirical distribution is approximately equal to q, up to an additive error of:

$$\tilde{O}\left(\sqrt{\frac{q(1-q)}{m}}\right) . \tag{2}$$

To ensure that the error bound holds for all values, one can simply take a union bound at the cost of an extra logarithmic factor inside the square root. Intuitively, the dominated empirical distribution is obtained by subtracting this term from the quantile of each value v in the empirical distribution.

Next we explain the main difference between our algorithm and those in previous works, with the exception of [Roughgarden and Schrijvers 2016]. Previous works generally pick the optimal auction w.r.t. the empirical distribution, with a distribution-family-dependent preprocessing on the sample values, in the form of truncating large but rare values and/or a discretization of the values. The preprocessing is to avoid choosing the auction based on some rare but high values in the samples. In contrast, our algorithm picks the optimal auction w.r.t. the dominated empirical distribution, without any preprocessing or any knowledge of the underlying family of distributions. The conservative estimates of quantiles by the dominated empirical distribution implicitly tune down the impact of rare but high values, simultaneously for all families of distributions.

The algorithm by [Roughgarden and Schrijvers 2016] is the most similar one to ours. They also constructed a dominated empirical distribution and picked the corresponding optimal auction. A subtle difference is that they used the Dvoretzky-Kiefer-Wolfowitz (DKW) inequality [Dvoretzky et al. 1956] to bound the estimation error of the empirical distribution and to construct the dominated empirical distribution, which on one hand avoided losing a logarithmic factor from the union bound, but on the other hand did not get the better bounds for values with quantiles close to 0 or 1 as in Eqn. 2. The latter property is crucial for our analysis. We leave as an interesting open question whether there is a strengthened version of the DKW inequality with quantile-dependent bounds. Such an inequality will improve the logarithmic factor in the upper bounds of this paper. We stress that while the algorithms are similar in spirit, our analysis is fundamentally different, as we will explain next. Importantly, our sample complexity upper bounds hold for the general non-i.i.d. case while the upper bound of [Roughgarden and Schrijvers 2016] holds only for the special case of i.i.d. bidders.

Analysis via Revenue Monotonicity. Our analysis consists of two components. The first one is two inequalities that lower bound the expected revenue of the dominated empirical Myerson auction on the true distribution, where the inequalities are enabled by the strong revenue monotonicity of single-parameter problems by [Devanur et al. 2016]. The strong revenue monotonicity states that the optimal auction w.r.t. a distribution that is dominated by the true distribution gets at least the optimal revenue of the dominated distribution. In particular, running the dominated empirical Myerson on the true value distribution $\mathbf{D}$ gets at least the optimal revenue of the dominated empirical distribution $\tilde{\mathbf{E}}$. Further, consider a doubly shaded version of the true distribution, denoted as $\tilde{\mathbf{D}}$, which intuitively is obtained by subtracting twice the error term in Eqn. 2 from the quantiles of the true distribution. Then, $\tilde{\mathbf{D}}$ is dominated by $\tilde{\mathbf{E}}$ and, thus, its optimal revenue is at most that of $\tilde{\mathbf{E}}$. This weaker notion of revenue monotonicity is folklore in the literature and follows as a direct corollary of the stronger notion. Therefore, we conclude that the expected revenue of the dominated empirical Myerson auction is at least the optimal revenue of the doubly shaded distribution $\tilde{\mathbf{D}}$. It remains to compare the optimal revenue of $\mathbf{D}$ and $\tilde{\mathbf{D}}$.

This idea is quite powerful on its own. The key observation is that $\tilde{\mathbf{D}}$ approximately preserves the probability density/mass of $\mathbf{D}$ almost *point-wise*, except for a small subset of values that have little impact on the optimal revenue. Intuitively, this is because it consistently underestimates the quantiles; in contrast, the empirical distribution has fluctuations in its estimations. Hence, $\tilde{\mathbf{D}}$ approximately preserves the virtual values of $\mathbf{D}$ almost point-wise, circumventing the technical hurdle faced by the second previous approach discussed in Section 2. By this idea and standard accounting arguments for the expected revenue, we can get the optimal sample complexity upper bound for regular distributions in Table II, and match the best previous upper bounds for the other three families of distributions in Table I.

Analysis via Information Theory. To get the optimal sample complexity upper bounds for all families of distributions under a unified framework, we need the second idea, namely, to bound the difference between the optimal revenues of $\mathbf{D}$ and $\tilde{\mathbf{D}}$ with an information theoretic argument. The argument consists of two claims: 1) the distributions $\mathbf{D}$ and $\tilde{\mathbf{D}}$ are similar in the information theoretic sense so that it takes many samples to distinguish them, and 2) we can estimate the expected revenue of any given mechanism on $\mathbf{D}$ and $\tilde{\mathbf{D}}$ with a small number of samples. Concretely, we will show that the Kullback-Leibler (KL) divergence between $\mathbf{D}$ and $\tilde{\mathbf{D}}$ is at most $\tilde{O}(\frac{n}{m})$, omitting some caveats which are explained in details in [Guo et al. 2019]. By standard information theoretic arguments, it implies that one needs at least $\tilde{\Omega}(\frac{m}{n})$ samples to distinguish these two distributions. For example, consider a $[0, 1]$-bounded distribution $\mathbf{D}$ and an additive ϵ approximation. Suppose m is at least $\tilde{O}(n\epsilon^{-2})$ as in Table II. Then, we get that it takes at least $C \cdot \epsilon^{-2}$ samples to distinguish $\mathbf{D}$ and $\tilde{\mathbf{D}}$ for some sufficiently large constant $C > 0$. On the other hand, it takes less than $C \cdot \epsilon^{-2}$ samples to estimate the expected revenue of any mechanism on both $\mathbf{D}$ and $\tilde{\mathbf{D}}$ up to an additive ϵ factor. Thus, the expected revenue of any mechanism differs by at most ϵ on the two distributions; otherwise, we can distinguish them with less than $C \cdot \epsilon^{-2}$ samples by estimating the expected revenue of the mechanism. As a result, the optimal revenues of $\mathbf{D}$ and $\tilde{\mathbf{D}}$ differ by at most ϵ.

To our knowledge, this is the first time information theory is used to show sample complexity upper bounds for revenue maximization. Previously, it was used only for lower bounds (e.g., [Huang et al. 2015]). We believe it will find further applications in studying the sample complexity of multi-parameter revenue maximization and other learning problems. We stress that our algorithm is constructive and, in fact, can be implemented in quasi-linear time;[3] both the doubly shaded distribution $\tilde{\mathbf{D}}$ and the information theoretic arguments are used only in the analysis.

Lower Bound Constructions. Our lower bounds are unified under a meta construction, with some components chosen based on the family of distributions. We briefly sketch the construction below. Let the first bidder's value distribution be a point mass. She will serve as the default winner in the optimal auction. The value distribution of each of the other $n-1$ bidders will be either D^h or D^ℓ. These two distributions satisfy that there is a value interval such that for any value in it, the corresponding virtual value wins over bidder 1 if and only if the distribution is D^h. Both D^h and D^ℓ will have an $O(\frac{1}{n})$ chance of realizing a value in this interval. Intuitively, to find a near optimal mechanism we must be able to distinguish the bidders with distribution D^h from those with distribution D^ℓ. Finally, we will construct D^h and D^ℓ to be similar so that it takes many samples to distinguish them. The meta construction, inspired by the hard instances by [Cole and Roughgarden 2014], can be viewed as a non-trivial generalization of the lower bound framework by [Huang et al. 2015] for the special case of single bidder.

4. BEYOND SINGLE-ITEM AUCTIONS

The aforementioned techniques generalize to the case when the seller has multiple copies of the item and therefore can allocate to multiple bidders. Suppose there are k copies, and bidders are unit-demand in the sense that giving a bidder multiple copies does not make her more happy than having just one copy. Further, suppose the bidders' values have support in $[0, 1]$ and the goal is achieving an ϵ additive approximation. Then, our techniques show that the same algorithm, which returns Myeron's optimal auction w.r.t. the dominated empirical distributions, achieves the sample complexity $\tilde{\Theta}(nk\epsilon^{-2})$, which is optimal up to a poly-logarithmic factor. Comparing with the bounds for the single-item setting, the only difference is a linear dependence in the number of copies available to the seller.

Finally, this sample complexity lower and upper bounds generalize to the matroid setting with rank k. The above setting with k copies of the item is a special case known as the k-uniform matroid.

REFERENCES

Balcan, M.-F., Sandholm, T., and Vitercik, E. 2018. A general theory of sample complexity for multi-item profit maximization. In *Proceedings of the 19th ACM Conference on Economics and Computation*. ACM, 173–174.

Balcan, M.-F. F., Sandholm, T., and Vitercik, E. 2016. Sample complexity of automated mechanism design. In *Advances in Neural Information Processing Systems*. 2083–2091.

Blum, A. and Hartline, J. D. 2005. Near-optimal online auctions. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 1156–1163.

Bubeck, S., Devanur, N. R., Huang, Z., and Niazadeh, R. 2017. Online auctions and multi-scale online learning. In *Proceedings of the 18th ACM Conference on Economics and Computation*. ACM, 497–514.

[3]For each bidder, it takes $O(m \log m)$ time to sort the samples and to compute the quantiles of the empirical distribution, and $O(m)$ time to compute the quantiles of the dominated empirical distribution, and $O(m \log m)$ times to compute the convex hull of the corresponding revenue curve, which characterizes the optimal auction.

CAI, Y. AND DASKALAKIS, C. 2017. Learning multi-item auctions with (or without) samples. In *Proceedings of the 58th Annual IEEE Symposium on Foundations of Computer Science*.

COLE, R. AND ROUGHGARDEN, T. 2014. The sample complexity of revenue maximization. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*. ACM, 243–252.

DEVANUR, N. R., HUANG, Z., AND PSOMAS, C.-A. 2016. The sample complexity of auctions with side information. In *Proceedings of the 48th annual ACM symposium on Theory of Computing*. ACM, 426–439.

DHANGWATNOTAI, P., ROUGHGARDEN, T., AND YAN, Q. 2015. Revenue maximization with a single sample. *Games and Economic Behavior 91*, 318–333.

DUGHMI, S., HAN, L., AND NISAN, N. 2014. Sampling and representation complexity of revenue maximization. In *International Conference on Web and Internet Economics*. Springer, 277–291.

DVORETZKY, A., KIEFER, J., AND WOLFOWITZ, J. 1956. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, 642–669.

ELKIND, E. 2007. Designing and learning optimal finite support auctions. In *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 736–745.

GONCZAROWSKI, Y. A. AND NISAN, N. 2017. Efficient empirical revenue maximization in single-parameter auction environments. In *Proceedings of the 49th Annual ACM Symposium on Theory of Computing*. ACM, 856–868.

GONCZAROWSKI, Y. A. AND WEINBERG, S. M. 2018. The sample complexity of up-to-*varepsilon* multi-dimensional revenue maximization. In *Proceedings of the 59th Annual IEEE Symposium on Foundations of Computer Science*.

GUO, C., HUANG, Z., AND ZHANG, X. 2019. Settling the sample complexity of single-parameter revenue maximization. In *Proceedings of the 51st Annual ACM Symposium on Theory of Computing*. ACM.

HART, S. AND RENY, P. J. 2015. Maximal revenue with multiple goods: Nonmonotonicity and other observations. *Theoretical Economics 10,* 3, 893–922.

HUANG, Z., MANSOUR, Y., AND ROUGHGARDEN, T. 2015. Making the most of your samples. In *Proceedings of the 16th ACM Conference on Economics and Computation*. ACM, 45–60.

MORGENSTERN, J. AND ROUGHGARDEN, T. 2016. Learning simple auctions. In *Conference on Learning Theory*.

MORGENSTERN, J. H. AND ROUGHGARDEN, T. 2015. The pseudo-dimension of nearly optimal auctions. In *Advances in Neural Information Processing Systems*. 136–144.

MYERSON, R. B. 1981. Optimal auction design. *Mathematics of Operations Research 6,* 1, 58–73.

ROUGHGARDEN, T. AND SCHRIJVERS, O. 2016. Ironing in the dark. In *Proceedings of the 17th ACM Conference on Economics and Computation*. ACM, 1–18.

RUBINSTEIN, A. AND WEINBERG, S. M. 2015. Simple mechanisms for a subadditive buyer and applications to revenue monotonicity. In *Proceedings of the 16th ACM Conference on Economics and Computation*. ACM, 377–394.

SYRGKANIS, V. 2017. A sample complexity measure with applications to learning optimal auctions. In *Advances in Neural Information Processing Systems*. 5352–5359.

YAO, A. C.-C. 2018. On revenue monotonicity in combinatorial auctions. In *Proceedings of the 11th International Symposium on Algorithmic Game Theory*. Springer, 1–11.