Mechanism Design via Machine Learning*

Maria-Florina Balcan †
Department of Computer Science
Carnegie Mellon University
ninamf@cs.cmu.edu

Avrim Blum †
Department of Computer Science
Carnegie Mellon University
avrim@cs.cmu.edu

Jason D. Hartline
Microsoft Research
Mountain View, CA 94043
hartline@microsoft.com

Yishay Mansour‡
School of Computer Science
Tel-Aviv University
mansour@cs.tau.ac.il

Abstract

We use techniques from sample-complexity in machine learning to reduce problems of incentive-compatible mechanism design to standard algorithmic questions, for a wide variety of revenue-maximizing pricing problems. Our reductions imply that for these problems, given an optimal (or β-approximation) algorithm for the standard algorithmic problem, we can convert it into a $(1+\epsilon)$-approximation (or $\beta(1+\epsilon)$-approximation) for the incentive-compatible mechanism design problem, so long as the number of bidders is sufficiently large as a function of an appropriate measure of complexity of the comparison class of solutions. We apply these results to the problem of auctioning a digital good, the attribute auction *problem, and to the problem of item-pricing in unlimited-supply combinatorial auctions. From a learning perspective, these settings present several challenges: in particular, the loss function is discontinuous and asymmetric, and the range of bidders' valuations may be large.*

1. Introduction

A common goal in the design of many pricing mechanisms is that of obtaining more profit than is possible from a single sale price. There are two prevalent practices in discriminatory pricing. The first is using public information about each consumer in the calculation of offer prices. Such pricing is the de facto standard, for example, in pricing automobile insurance. The second is to distinguish between the products for sale in a way that causes consumers to have different preferences for the products. A single price for each product, then, effectively charges consumers different prices when they choose different products. This is standard procedure in the sale of all sorts of commodities; common examples include computer software, computer hardware, and airline tickets. When either of these types of discriminatory pricing is possible, auctions that approximate the optimal single price sale, e.g., [11, 4], may no longer be near-optimal.

We consider the design of pricing mechanisms in a game theoretic setting where the consumers (a.k.a., agents or bidders) may choose to dishonestly report their preferences if it might benefit them. We will adopt the now standard paradigm of considering only *incentive compatible* mechanisms, i.e., ones explicitly designed so that each bidder has a dominant strategy of reporting their true preferences.

Our main result is to use techniques from sample-complexity in machine learning theory to reduce the design of revenue-maximizing incentive-compatible mechanisms to standard algorithmic questions. When the number of agents is sufficiently large as a function of an appropriate measure of complexity of the class of solutions being compared to, this reduction loses only a $1+\epsilon$ factor in solution quality; that is, an algorithm (or β-approximation) for the standard algorithmic problem can be converted to a $(1+\epsilon)$-approximation (or $\beta(1+\epsilon)$-approximation) for the incentive-compatible design problem. We do this in a fairly general setting that includes the following as special cases:

Auction of digital goods to indistinguishable bidders. In this problem, studied in [11, 7], we have a digital

*The full version of this paper is available as Technical Report CMU-CS-05-143.

†Research supported in part by NSF grants CCR-0105488, CCR-0122581, and IIS-0121678 and a BSF grant.

‡This work was supported in part by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778, by a grant no. 1079/04 from the Israel Science Foundation, by a grant from BSF and an IBM faculty award. This publication reflects only the authors' views.

good (a good of unlimited supply with zero marginal cost) and n bidders, where each bidder i has some valuation v_i between 1 and h. Our goal is to sell our good so as to make profit comparable to the best single price: the price p maximizing $p \times |\{i : v_i \geq p\}|$.

For this problem, Goldberg et al. [11] give a simple auction based on random sampling and show that it gives near 6-approximation so long as the optimal revenue is large compared to h.[1] We analyze a slight variant and show (Theorem 6) that it is a $(1+\epsilon)$-approximation so long as the optimal revenue is large compared to $\frac{h}{\epsilon^2}\log(1/\epsilon)$.

Attribute Auctions. In many generalizations of the digital-good auction, the bidders are not a priori indistinguishable; instead, publicly known information about bidders may allow (or even require) differential treatment. For example, the motion picture industry uses region encodings so that they can charge different prices for DVDs sold in different markets.

This introduces the natural question of how to use the distinguishing features of consumers to price-discriminate to the maximum benefit of the seller. We consider the following abstraction of these situations. The bidders in an *attribute auction* are not indistinguishable but instead have a set of publicly-known *attributes* and the goal is to achieve revenue comparable to the best pricing function over these attributes from some available class, $\mathcal{G}$, of pricing functions. For example, [3] considers the special case of 1-dimensional attributes and a comparison class $\mathcal{G}$ of piece-wise constant functions. Piece-wise constant functions divide the attribute space into contiguous regions (a.k.a., markets) and charge a single price in each. We give bounds for this setting more generally, including a generalization of the class of functions considered in [3] to higher dimensions.

Item-pricing in combinatorial auctions. This problem is a different generalization of the first problem above, and studied in [12, 15]. The setting here is we have m different items, each in unlimited supply (like a supermarket), and bidders have valuations on *subsets* of items. Our goal is to achieve revenue nearly as large as the best sale that uses item prices (assigns a separate price to each item), a natural comparison class. Our results imply that $\tilde{O}(hm^2/\epsilon^2)$ bidders are sufficient to achieve revenue close to the optimum item-pricing (assuming the algorithmic problem can be solved for the given bidders), no matter how complicated those bidders' valuations are. In the unit-demand case, when each bidder wants at most one item (such as in pricing different versions of the same software or pricing airline tickets), our bounds give a $(1+\epsilon)$-approximation when the optimal revenue is large compared to $\tilde{O}(hm/\epsilon^2)$ which improves by roughly a factor of m over the results of [12].

The basic reduction we apply to solve these auction problems is as follows. Given an algorithm $\mathcal{A}$ (exact or approximate) for the non-incentive-compatible pricing problem (finding the optimal pricing function in class $\mathcal{G}$ for a given set of bidders) and given a set of bidders S, we will split bidders randomly into two sets S_1 and S_2, run the algorithm separately on each set (perhaps adding an additional penalty term to the objective to penalize solutions that are too "complex" according to some measure), and then apply the solution found on S_1 to S_2 and the solution found on S_2 to S_1. Sample-complexity techniques from machine learning theory can then give a guarantee on the quality of the results if the number of bidders is sufficiently large compared to an appropriate measure of the complexity of the class of possible solutions. From a learning perspective, the mechanism-design setting presents a number of technical challenges: in particular, the loss function is discontinuous and asymmetric, and the range of bid values may be large.

In addition to the generic reduction, we also give specific analyses for several of the above problems, using their structure to yield better bounds on the number of bidders needed to achieve a desired approximation factor.

Related work: Several papers [4, 3] have applied machine learning techniques to mechanism design in the context of online auctions. The online setting is more difficult than the "batch" setting we consider, but the flip-side is that as a result, that work only applies to quite simple mechanism design settings where the class $\mathcal{G}$ of comparison functions has small size and can be easily listed.

Structure of this paper: We begin by defining our general setting (Section 2) and giving a basic reduction at this level of generality (Section 3). We then proceed to give a tighter analysis for the basic auction of a digital good (Section 4) and describe in Section 5 how the complexity measures of Section 3 can be instantiated for the case of attribute auctions. We consider item-pricing in combinatorial auctions in Section 6 and the multicast pricing problem in Section 7. We give our conclusions and some outstanding research directions in Section 8.

2. Definitions

We will be considering mechanism design problems of the following general form. We have a set S of n bidders,

[1] This problem has also been considered in a framework where the auction's performance is compared to the profit obtained from the optimal sale price that results in a sale of *at least two* items [7]. In this context the best known auction is $13/4$-competitive [14].

and we assume that each bidder i has some private information $priv_i$ (like how much they are willing to pay for a digital good), as well as public information pub_i (such as their location in a network). The game itself will be defined by an abstract space of legal *offers* (like an offer to sell a good at \$17) together with a mapping ρ that defines how much profit a given offer yields from a given bidder. For example, in the case of auctioning a digital good, $\rho(\text{"offer \$17"}, priv_i) = 17$ if $priv_i \geq 17$ and 0 otherwise. We can think of ρ as defining the assumption about how bidders behave as a function of their private values. The standard assumption in incentive compatible mechanism design is that bidders prefer the outcome that maximizes their *utility* which is defined as the difference between their *valuation* for the outcome (as specified by their preferences) and the payment they are required to make.

Definition 1 *A* comparison class*, $\mathcal{G}$, of pricing functions is a set of functions g that map the public information of a bidder to an offer. The* profit *of a function g is $\sum_i \rho(g(pub_i), priv_i)$. Note that we are implicitly considering only* unlimited supply *mechanism design problems, because the profit from bidder i does not depend on whether g received profit from other bidders.*

Given a comparison class, $\mathcal{G}$, the *algorithm design* problem is: given both the public and private information in S, find the $g \in \mathcal{G}$ of highest total profit $\mathrm{OPT}_{\mathcal{G}}$. Some of the problems we consider will also have costs for various functions g: for instance, in multicast pricing, a comparison function g consists of both a tree and a proposed price at each node, and its cost is the cost of the tree. In this case, we should think of ρ as a **revenue** function, and the algorithm design problem will be to find the g of highest revenue minus cost. In our reductions, we may also want to perform "structural risk minimization", which adds additional fake penalties to different functions g based on some measure of their complexity, in which case we will need to assume we have an algorithm that optimizes revenue minus penalty.

We now need to define what we mean by an incentive compatible mechanism. An incentive-compatible mechanism is a function that takes in the public information of all the bidders, plus the private information of all bidders *except* the given bidder i and outputs an offer. Our goal will be to design such a mechanism whose total profit is nearly as large as the profit of the best function in comparison class $\mathcal{G}$. Note that typically our mechanisms will not actually belong to $\mathcal{G}$, such as offering one price to some subset of bidders and another price to another even if our class $\mathcal{G}$ is the set of all single price functions.

One final point at this level of generality: we will assume that we are given an upper bound h on the value of ρ; that is, no individual bidder can influence profit by more than h. This term will come into our sample-complexity bounds.

2.1. Examples

Auction of digital goods to indistinguishable bidders. As described in the introduction, in this setting the bidders have no public information (equivalently, all the bidders have the *same* public information *pub*) and the private information of bidder i is exactly its valuation v_i for the digital good, which is a real number between 1 and h. Here, a natural comparison class $\mathcal{G} = \{g_p\}$ is the class of all functions that offer a single price p, and ρ is a function defined by $\rho(p, priv_i) = p$ if $p \leq priv_i$ and $\rho(p, priv_i) = 0$ otherwise.

Attribute Auctions. This is the same as the setting above except now each bidder i is associated a public **attribute** $pub_i \in \mathcal{X}$ where $\mathcal{X}$ is the **attribute space**. We view $\mathcal{X}$ as an abstract space, but one can envision it as $\mathbb{R}^d$, for example. $\mathcal{G}$ is then a class of pricing functions from $\mathcal{X}$ to $\mathbb{R}_+$, such as all linear functions or all functions that partition $\mathcal{X}$ into k markets (say based on distance to k cluster centers) and offer a different price in each. The mapping ρ is a function from $\mathbb{R}_+ \times [1, h]$ to $[0, h]$ defined (as in the case of indistinguishable bidders) by $\rho(p, priv_i) = p$ if $p \leq priv_i$ and $\rho(p, priv_i) = 0$ otherwise. We will give analyses of several interesting classes of comparison functions in Section 5.

Combinatorial Auctions. Here we have a set J of m distinct items, each in unlimited supply. Each consumer has a private valuation $v_i(s)$ for each bundle $s \subseteq J$ of items, which measures how much receiving bundle s would be worth to the consumer i. The private information of bidder i can be described by a vector of all its valuations on subsets of J (for simplicity, we assume bidders are indistinguishable, i.e., no public information). A natural class of comparison functions $\mathcal{G}$ (studied in [15]) is the class of functions that assign a separate price to each item, such that the price of a bundle is just the sum of the prices of the items in it (called item-pricing). The mapping ρ is then defined by assuming bidders will buy the bundle (if any) with largest positive gap between its value to them and its total cost.

3. Generic reductions

We are interested in reducing incentive-compatible mechanism design to the standard algorithm design problem. Our reductions will be based on random sampling. Let $\mathcal{A}$ be an algorithm for the (non incentive-compatible) problem of optimizing over $\mathcal{G}$. The simplest mechanism that we consider, which we call $\mathrm{RSOPF}_{(\mathcal{G},\mathcal{A})}$ (Random Sampling Optimal Pricing Function), is the following generalization of the random sampling digital-goods auction from [11]:

1. Randomly split the bidders into two groups S_1 and S_2, flipping a fair coin for each bidder.

2. Run $\mathcal{A}$ to determine the best (or approximately best) function $g_1 \in \mathcal{G}$ over S_1, and similarly the best (or approximately best) $g_2 \in \mathcal{G}$ over S_2.

3. Finally, apply g_1 to S_2 and g_2 to S_1.

We will also consider various more refined versions of RSOPF$_{(\mathcal{G},\mathcal{A})}$, that discretize $\mathcal{G}$ or perform some type of *structural risk minimization* (in which case we will need to assume $\mathcal{A}$ can optimize over the modifications made to $\mathcal{G}$).

3.1. The Basic Analysis

In order to simplify notation, for a given setting (defined by ρ and $\mathcal{G}$), define $g(i)$ for a pricing function g and bidder i to be the profit made by g on i; i.e., $g(i) = \rho(g(pub_i), priv_i)$. Similarly, for a set of bidders $S' \subseteq S$, let $g(S') = \sum_{i \in S'} g(i)$. So, $\mathrm{OPT}_{\mathcal{G}} = \max_{g \in \mathcal{G}} g(S)$. Note that if $g_1(i) = g_2(i)$ for all $i \in S$ then they are equivalent from the point of view of the auction; we will use $|\mathcal{G}|$ to denote the number of *different* such functions in $\mathcal{G}$.

The following lemma is key to our analysis. Note that using Hoeffding bounds would produce an h^2 term in the exponent; by applying McDiarmid's inequality instead we only need to lose a factor of $O(h)$.

Lemma 1 *Consider a pricing function g and a profit level p. If we randomly partition S into S_1 and S_2, then the probability that $|g(S_1) - g(S_2)| \geq \epsilon \max[g(S), p]$ is at most $2e^{-\epsilon^2 p/(2h)}$.*

Proof: Let $Y_1, \ldots, Y_n$ be i.i.d random variables such that Y_i is 1 with probability $1/2$ and Y_i is 2 with probability $1/2$, and that define the partition of S into S_1 and S_2. Let $t(y_1, ..., y_n) = \sum_{i: y_i = 1} g(i)$. So, as a random variable, $g(S_1) = t(Y_1, ..., Y_n)$ and clearly $\mathbf{E}[t(Y_1, ..., Y_n)] = g(S)/2$. Assume first that $g(S) \geq p$. From the McDiarmid concentration inequality (see Appendix A) we get: $\mathbf{Pr}\{|g(S_1) - g(S)/2| \geq \frac{\epsilon}{2} g(S)\} \leq 2e^{-\epsilon^2 g(S)/(2h)}$. This is true since by plugging $c_i = g(i)$ in Theorem 15, we get:

$$\mathbf{Pr}\left\{\left|g(S_1) - \frac{g(S)}{2}\right| \geq \frac{\epsilon}{2} g(S)\right\} \leq 2e^{-\left[\frac{\epsilon^2 g(S)^2}{2\sum_{i=1}^{n} g(i)^2}\right]}.$$

Since $\sum_{i=1}^n g(i)^2 \leq \max_i\{g(i)\} \sum_{i=1}^n g(i)$, we obtain:

$$\mathbf{Pr}\left\{\left|g(S_1) - \frac{g(S)}{2}\right| \geq \frac{\epsilon}{2} g(S)\right\} \leq 2e^{-\left[\frac{\epsilon^2 g(S)}{2h}\right]}.$$

Moreover, since $g(S_1) + g(S_2) = g(S)$ and $g(S) \geq p$, we get that $\mathbf{Pr}\{|g(S_1) - g(S_2)| \geq \epsilon g(S)\} \leq 2e^{-\epsilon^2 p/(2h)}$. Consider now that $g(S) < p$. Again, using the McDiarmid inequality we have

$$\mathbf{Pr}\{|g(S_1) - g(S_2)| \geq \epsilon p\} \leq 2e^{-\left[\frac{\epsilon^2 p^2}{2\sum_{i=1}^{n} g(i)^2}\right]}.$$

Since $\sum_{i=1}^n g(i)^2 \leq hg(S) \leq ph$ we obtain again that $\mathbf{Pr}\{|g(S_1) - g(S_2)| \geq \epsilon n\} \leq 2e^{-\epsilon^2 p/(2h)}$, which gives us the desired bound. ■

We can now give our simplest generic reduction, based on just the number of functions in $\mathcal{G}$. Note that in a number of settings (see Sections 3.3, 4, and 5.2) we will be able to get stronger guarantees by a more refined analysis.

Theorem 2 *Given comparison class $\mathcal{G}$ and a β-approximation algorithm $\mathcal{A}$ for optimizing over $\mathcal{G}$, then so long as $\mathrm{OPT}_{\mathcal{G}} \geq \beta n$ and the number of bidders n satisfies*

$$n \geq \frac{8h}{\epsilon^2} \ln(2|\mathcal{G}|/\delta),$$

then with probability at least $1-\delta$, the profit of RSOPF$_{(\mathcal{G},\mathcal{A})}$ is at least $(1-\epsilon)\,\mathrm{OPT}_{\mathcal{G}}/\beta$.

Proof: Let g_1 be the function in $\mathcal{G}$ produced by $\mathcal{A}$ over S_1 and g_2 be the function in $\mathcal{G}$ produced by $\mathcal{A}$ over S_2. Let g_{OPT} be the optimal function in $\mathcal{G}$ over S. Since the optimal function over S_1 is at least as good as g_{OPT} on S_1 (and likewise for S_2), the fact that $\mathcal{A}$ is a β-approximation implies that $g_1(S_1) \geq g_{\mathrm{OPT}}(S_1)/\beta$ and $g_2(S_2) \geq g_{\mathrm{OPT}}(S_2)/\beta$.

By Lemma 1 (using $p = n$) and plugging in our bound on n and applying a simple union bound, with probability $1-\delta$, *every* $g \in \mathcal{G}$ satisfies $|g(S_1) - g(S_2)| \leq \frac{\epsilon}{2}\max[g(S), n]$. In particular, $g_1(S_2) \geq g_1(S_1) - \frac{\epsilon}{2}\max[g_1(S), n]$, and $g_2(S_1) \geq g_2(S_2) - \frac{\epsilon}{2}\max[g_2(S), n]$.

Since $\mathrm{OPT}_{\mathcal{G}} \geq \beta n$, summing the above two inequalities and performing a simple case-analysis we get that the profit of RSOPF$_{(\mathcal{G},\mathcal{A})}$, namely the sum $g_1(S_2) + g_2(S_1)$, is at least $(1-\epsilon)\,\mathrm{OPT}_{\mathcal{G}}/\beta$. ■

3.2. Structural Risk Minimization

In many natural cases, $\mathcal{G}$ consists of functions at different "levels of complexity" k, such as partitioning bidders into k markets for different values of k. One natural approach to such a setting is to perform *structural risk minimization* (SRM), that is, to assign a penalty term to functions based on their complexity and then to run a version of RSOPF$_{(\mathcal{G},\mathcal{A})}$ in which $\mathcal{A}$ optimizes profit minus penalty. Specifically, let $\bar{\mathcal{G}}$ be a series of pricing function classes $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \ldots$, and let pen be a penalty function defined over these classes. Also for simplicity assume $\beta = 1$ (we have an optimal algorithm for the underlying problem). We then define the procedure RSOPF-SRM$_{(\bar{\mathcal{G}},\mathsf{pen})}$ as follows:

1. Randomly partition the bidders into two sets, S_1 and S_2, flipping fair coin for each bidder.

2. Compute g_1 to maximize $\max_k \max_{g \in \mathcal{G}_k} [g(S_1) - \mathsf{pen}(\mathcal{G}_k)]$ and similarly compute g_2 from S_2.

3. Use price function g_1 for bidders in S_2 and g_2 for bidders in S_1.

A straightforward extension of Theorem 2 to this case would introduce a quadratic dependence in h, but we will be able to reduce this to nearly linear. Define $\mathrm{OPT}_k = \mathrm{OPT}_{\mathcal{G}_k}$.

Corollary 3 *Suppose we randomly partition S into S_1 and S_2. With probability at least $1-\delta$, we obtain that for all functions g in $\mathcal{G}_k$ such that $g(S) \geq \frac{2h}{\epsilon^2}\left[\ln\left(\frac{2}{\delta}\right) + \ln(|\mathcal{G}_k|)\right]$ we have $|g(S_1) - g(S_2)| \leq \epsilon g(S)$.*

Proof: Follows from Lemma 1 by plugging in $p = g(S)$, for $g \in \mathcal{G}_k$ and then using a simple union bound. ■

Notice that a key difference between the above lemma and Theorem 2 is that the lemma requires only a lower bound on *profit* rather than on the number of bidders. Using Corollary 3 we can now prove the following lemma:

Lemma 4 *Assume that we have an algorithm $\mathcal{A}_k$ for optimizing over $\mathcal{G}_k$ and let g_i be the best function in $\mathcal{G}_k$ over S_i. For any given value of n, ϵ, and δ, with probability at least $1-\delta$ we have that if $\mathrm{OPT}_k \geq \frac{6}{3-\epsilon}\frac{72h}{\epsilon^2}\ln(2|\mathcal{G}_k|/\delta)$ then $g_i(S_j) \geq \frac{1-\epsilon}{2}\mathrm{OPT}_k$, for $i = 1, 2$, $i \neq j$. In particular, this implies the revenue of RSOPF$_{\mathcal{G}_k,\mathcal{A}_k}$ is at least:*

$$(1-\epsilon)\,\mathrm{OPT}_k - \mathsf{pen}(\mathcal{G}_k),$$

where $\mathsf{pen}(\mathcal{G}_k) = \frac{6}{3-\epsilon}\frac{72h}{\epsilon^2}\ln(2|\mathcal{G}_k|/\delta)$.

Proof: We prove that, with probability $1-\delta$, we have that if $\mathrm{OPT}_k \geq \frac{2}{1-\epsilon'}\frac{8h}{\epsilon'^2}\ln(2|\mathcal{G}_k|/\delta)$ then $g_i(S_j) \geq \frac{1}{2}\frac{(1-\epsilon')^2}{1+\epsilon'}\mathrm{OPT}_k$; this implies our desired result using $\epsilon' = \epsilon/3$. Notice that if $\mathrm{OPT}_k \geq \frac{2}{1-\epsilon'}\frac{8h}{\epsilon'^2}\ln(2|\mathcal{G}_k|/\delta)$, then from Corollary 3, we have $g_{\mathrm{OPT}}(S_i) \geq \frac{8h}{\epsilon'^2}\ln(2|\mathcal{G}_k|/\delta)$ for $i = 1, 2$, which implies that $g_i(S_i) \geq \frac{8h}{\epsilon'^2}\ln(2|\mathcal{G}_k|/\delta)$ (since $g_i(S_i) \geq g_{\mathrm{OPT}}(S_i)$) . Using again Corollary 3 we obtain that $g_i(S_j) \geq \frac{1-\epsilon'}{1+\epsilon'}g_i(S_i)$ for $j \neq i$, which then implies the desired result. ■

We can finally obtain a guaranteed for the RSOPF-SRM$_{(\bar{\mathcal{G}},\mathsf{pen})}$ mechanism as follows:

Theorem 5 *Assuming that we have an algorithm for solving the optimization problem required by RSOPF-SRM$_{(\bar{\mathcal{G}},\mathsf{pen})}$ then for any given value of n, ϵ, and δ, with probability at least $1-\delta$, the revenue of RSOPF-SRM$_{(\bar{\mathcal{G}},\mathsf{pen})}$ for $\mathsf{pen}(\mathcal{G}_k) = \frac{6}{(1-\epsilon)^2}\frac{72h}{\epsilon^2}\ln(8k^2|\mathcal{G}_k|/\delta)$ is*

$$\max_k\left((1-\epsilon)\,\mathrm{OPT}_k - \mathsf{pen}(\mathcal{G}_k)\right).$$

Proof Sketch: Follows from Lemma 4 using the union bound over values $\delta_k = \delta/(4k^2)$. ■

3.3. Better Bounds via Cover Arguments and Discretization

In a number of cases, $|\mathcal{G}|$ is overkill as a measure of the true complexity of the class $\mathcal{G}$. In this section, we discuss a number of methods that can produce better bounds. These include both *analysis* techniques, where we do not change the mechanism but instead provide a stronger guarantee, and *design* techniques, where we modify the mechanism to produce a better bound. Due to space restrictions, we only outline the methods here. The details can be found in our full version of the paper.

Discretizing. In many cases, we can greatly reduce $|\mathcal{G}|$ without much affecting $\mathrm{OPT}_{\mathcal{G}}$ by performing some type of discretization. For instance, for auctioning a digital good, there are infinitely many single-price functions but only $\log_{1+\epsilon} h \approx \frac{1}{\epsilon}\ln h$ prices at powers of $(1+\epsilon)$. Also, since rounding down the optimal price to the nearest power of $1+\epsilon$ can reduce revenue for this auction by at most a factor of $1+\epsilon$, the optimal function in the discretized class must be close to the optimal function in the original class. More generally, if we can find a smaller class $\mathcal{G}'$ such that $\mathrm{OPT}_{\mathcal{G}'}$ is guaranteed to be close to $\mathrm{OPT}_{\mathcal{G}}$, then we can instruct our algorithm $\mathcal{A}$ to optimize over $\mathcal{G}'$ and get better bounds. Note that we do not know whether the simple $(1+\epsilon)^i$ discretization is guaranteed to only minimally affect OPT in the case of *combinatorial auctions* (see Section 6).

Counting possible outputs. Suppose our algorithm $\mathcal{A}$, run on a subset of S, can only output pricing functions from a restricted set $\mathcal{G}_{\mathcal{A}} \subset \mathcal{G}$. Then, we can simply replace $|\mathcal{G}|$ with $|\mathcal{G}_{\mathcal{A}}|$ (or $|\mathcal{G}_{\mathcal{A}}|+1$ if the optimal function is not one of them) in all the above arguments. For example, if $\mathcal{A}$ picks the optimal single price over its input for auctioning a digital good, then this price must be one of the bids, so $|\mathcal{G}_{\mathcal{A}}| \leq n$.

Using cover size. Suppose $\mathcal{G}$ has the property that there exists a much smaller class $\mathcal{G}'$ that "covers" it, with respect to the given set of bidders S. In particular, for every $g \in \mathcal{G}$ there exists $g' \in \mathcal{G}'$ such that g' extracts the same revenue as g does from each bidder, up to a $1+\epsilon$ factor; that is, $|g(i) - g'(i)| \leq \epsilon g(i)$ for all i. In this case ($\mathcal{G}'$ is an L_∞ multiplicative ϵ-cover of $\mathcal{G}$), it is clear that if all functions in $\mathcal{G}'$ perform similarly on S_1 as they do on S_2, then this will be true for all functions in $\mathcal{G}$ as well. E.g., our results in Section 5.2 use this type of analysis to avoid discretization. In the full version of this paper, we show that similar bounds can be proved with L_1-covers, where we require only that $\sum_{i\in S}|g(i) - g'(i)| \leq \epsilon\sum_{i\in S} g(i)$. We also demonstrate the utility of L_1 covers by showing the existence of L_1 covers of size $o(n)$ for the digital good auction; this is not possible for the other cover notions above.

It is worth noting that a straightforward application of analogous ϵ-cover results in learning theory [1] (which would require an additive, rather than multiplicative gap of ϵ for every bidder) would add an extra factor of h into our sample-size bounds.

4. Auctioning Digital Goods to Indistinguishable Bidders

We now consider the simplest problem of auctioning a digital good to indistinguishable bidders, and competing against the best single price. We can apply the discretization technique by defining $\mathcal{G}$ to be the set of all constant-price functions whose price $p \in [1, h]$ is a power of $(1 + \epsilon/2)$: if we can get revenue at least $(1 - \epsilon/2)$ times the optimal in this class, we will be within $(1-\epsilon)$ of the optimal fixed price overall. Applying Theorem 2 ($\mathcal{A}$ can trivially find the best function in $\mathcal{G}$ by simply trying all of them), with probability $1 - \delta$ we get at least $(1 - \epsilon)$ times the optimal fixed price so long as the number of bidders n is at least $\frac{32h}{\epsilon^2} \ln(\frac{4 \ln h}{\epsilon\delta})$.

It is interesting to contrast these results with that of [11] which showed that RSOPF over the set of constant-price functions is near 6-competitive with the promise that $n \gg h$. A much more complicated analysis of RSOPF in a slightly different competitive framework is given in [10].

We now present a more refined analysis, which gives us even better guarantees.

Theorem 6 *Let $\mathcal{G}$ be the class of constant price functions, discretized at powers of $(1+\frac{\epsilon}{2})$, and let $\delta < 1/2$. Then with probability $1 - \delta$, RSOPF$_{(\mathcal{G},\mathcal{A})}$ obtains profit at least*

$$\mathrm{OPT}_{\mathcal{G}} - 8\sqrt{h \, \mathrm{OPT}_{\mathcal{G}} \log(2/(\epsilon\delta))}.$$

So, this implies that for $\mathrm{OPT}_{\mathcal{G}} \geq (\frac{16}{\epsilon})^2 h \log(2/(\epsilon\delta))$ we get profit at least $(1 - \epsilon/2)\,\mathrm{OPT}_{\mathcal{G}}$, which is at least $(1 - \epsilon)$ times the optimal non-discretized fixed price. So, even in the worst-case that the optimal single-price solution is at price 1 (so $\mathrm{OPT}_{\mathcal{G}} = n$) we get an $O(\log\log h)$ improvement over the generic bound, but if $\mathrm{OPT}_{\mathcal{G}}$ extracts substantially more profit on average per bidder, we can get an improvement of up to $O(h \log\log h)$.

To prove Theorem 6, let us for convenience define α to be the discretization parameter (which was $\epsilon/2$ above) and assume h is a power of $(1 + \alpha)$. For comparison function g_v offering price v, let n_v denote the number of winners (bidders whose value is at least v), and let $r_v = v \cdot n_v$ denote the profit of g_v on S. Denote by $\hat{r}_v$ the observed revenue of g_v on S_1 (and so $\hat{r}_v = v \cdot \hat{n}_v$, where $\hat{n}_v$ is the number of winners in S_1 for g_v). So, we have $\mathbf{E}[\hat{r}_v] = \frac{r_v}{2}$. We now begin with the following lemma.

Lemma 7 *Let $\epsilon < 1, \delta < 1/2$. With probability at least $1 - \delta$ we have that, for every $g_v \in \mathcal{G}$ the observed revenue on S_1 satisfies:*

$$\left|\hat{r}_v - \frac{r_v}{2}\right| \leq \max\left(\frac{h \log(1/(\alpha\delta))}{\epsilon}, \epsilon r_v\right).$$

Proof: First for a given price v let $a_{n,v}$ be $|\hat{n}_v - \frac{n_v}{2}|$. To prove our lemma we will use the consequence of Chernoff bound we present in Appendix A (see Theorem 16). For any v and $j \geq 1$ we consider $n' = \frac{(1+\alpha)^j \log(1/(\alpha\delta))}{\epsilon^2}$, and so we get $\mathbf{Pr}\left\{a_{n,v} \geq \epsilon \max\left(n_v, \frac{(1+\alpha)^j \log(1/(\alpha\delta))}{\epsilon^2}\right)\right\} \leq 2e^{-2(1+\alpha)^j \log(1/(\alpha\delta))}$. This further implies that we have $a_{n,v} \geq \epsilon \max\left(n_v, \frac{(1+\alpha)^j \log(1/(\alpha\delta))}{\epsilon^2}\right)$ with probability at most $2(\alpha\delta)^{2(1+\alpha)^j}$. Therefore for $v = h/(1+\alpha)^j$ we have $\mathbf{Pr}\left\{\left|\hat{r}_v - \frac{r_v}{2}\right| \geq \max\left(\frac{h \log(1/(\alpha\delta))}{\epsilon}, \epsilon r_v\right)\right\} \leq 2(\alpha\delta)^{2(1+\alpha)^j}$, and so the probability that there exists a $g_v \in \mathcal{G}$ such that $\left|\hat{r}_v - \frac{r_v}{2}\right| \geq \max\left(\frac{h}{\epsilon}, \epsilon r_v\right)$ is at most $2\sum_j (\alpha\delta)^{2(1+\alpha)^j} \leq 2\sum_{j'} \frac{1}{\alpha}(\alpha\delta)^{2 \cdot 2^{j'}} \leq \delta$. This implies that with high probability, at least $1 - \delta$, we have that *simultaneously*, for every $g_v \in \mathcal{G}$ the observed revenue on S_1 satisfies: $\left|\hat{r}_v - \frac{r_v}{2}\right| \leq \max\left(\frac{h \log(1/(\alpha\delta))}{\epsilon}, \epsilon r_v\right)$. ■

Proof of Theorem 6: Assume now that it is the case that for every $g_v \in \mathcal{G}$ we have $\left|\hat{r}_v - \frac{r_v}{2}\right| \leq \max\left(\frac{H}{\epsilon}, \epsilon r_v\right)$, where $H = h\log(2/(\alpha\delta))$. Let v^* be the optimal price level among prices in $\mathcal{G}$, and let $\tilde{v}^*$ be the price that looks best on S_1. Obviously, our gain on S_2 is $r_{\tilde{v}^*} - \hat{r}_{\tilde{v}^*}$. We have $\hat{r}_{v^*} \geq \frac{r_v^*}{2} - \frac{H}{\epsilon} - \epsilon r_{v^*} = r_{v^*}(1 - 2\epsilon)/2 - \frac{H}{\epsilon}$, $\hat{r}_{\tilde{v}^*} \geq \hat{r}_{v^*}$ and $\hat{r}_{\tilde{v}^*} \leq \frac{r_{\tilde{v}^*}}{2} + \frac{H}{\epsilon} + \epsilon r_{\tilde{v}^*} \leq \frac{r_{\tilde{v}^*}}{2} + \frac{H}{\epsilon} + \epsilon r_{v^*}$, and therefore $r_{\tilde{v}^*} - \hat{r}_{\tilde{v}^*} \geq \hat{r}_{\tilde{v}^*} - \frac{H}{\epsilon} - \epsilon r_{v^*}$, which finally implies that $r_{\tilde{v}^*} - \hat{r}_{\tilde{v}^*} \geq r_{v^*}\left(\frac{1}{2} - 2\epsilon\right) - 2\frac{H}{\epsilon}$. This implies that with probability at least $1 - \delta/2$ our gain on S_2 is at least $r_{v^*}\left(\frac{1}{2} - 2\epsilon\right) - 2\frac{H}{\epsilon}$, and similarly our gain on S_1 is at least $r_{v^*}\left(\frac{1}{2} - 2\epsilon\right) - 2\frac{H}{\epsilon}$. Therefore, with probability $1 - \delta$, our revenue is $\mathrm{OPT}_{\mathcal{G}}(1 - 4\epsilon) - 4\frac{h \log(1/(\alpha\delta))}{\epsilon}$. Optimizing the bound we set $\epsilon = \sqrt{h \log(1/(\alpha\delta))/OPT_{\mathcal{G}}}$ and get a revenue of $OPT_{\mathcal{G}} - 8\sqrt{h \; OPT_{\mathcal{G}} \; \log(1/(\alpha\delta))}$, which completes the proof. ■

5. Attribute Auctions

We begin by instantiating the results in Section 3 for market pricing auctions, and connecting to the notion of VC-dimension. We then give an analysis for general pricing functions over the attribute space that uses the notion of covers to avoid discretization.

5.1. Market Pricing

For attribute auctions, one natural class of comparison functions are those that partition bidders into *markets* in some simple way and then offer a single sale price in each market. For example, suppose we define $\mathcal{G}_k$ to be the set of functions that choose k bidders $b_1, \ldots, b_k$, use these as cluster centers to partition S into k markets based on distance to the nearest center in attribute space, and then offer a single price in each market. In that case, if we discretize prices to powers of $(1+\epsilon)$, then clearly the number of functions in $\mathcal{G}_k$ is at most $n^k(\log_{1+\epsilon} h)^k$, so Theorem 2 implies that so long as $n \geq \frac{8h}{\epsilon^2}\left[\ln(2/\delta) + k\ln n + k\ln\left(\log_{1+\epsilon} h\right)\right]$ and we can solve the algorithmic problem then with probability at least $1-\delta$, we can get profit at least $(1-\epsilon)\,\mathrm{OPT}_{\mathcal{G}_k}$.

Another interesting and general way to do market pricing is the following. Let C be a class of subsets of $\mathcal{X}$, which we will call *feasible markets*. For k a positive integer, we consider $F_{k+1}(C)$ to be the set of all pricing functions of the following form: pick k disjoint subsets s_1,...,s_k from C, and $k+1$ prices p_0,...,p_k discretized to powers of $1+\epsilon$. Assign price p_i to bidders in s_i, and price p_0 to bidders not in any of s_1,...,s_k. For example, if $\mathcal{X} = \mathbb{R}^d$ a natural C might be the set of axis-parallel rectangles in $\mathbb{R}^d$. The specific case of $d=1$ was studied in [3].

We can apply the results in Section 3 by using the machinery of VC-dimension to count the number of distinct such functions over any given set of bidders S. In particular, let $D = VCdim(C)$ be the VC-dimension of C and assume $D < \infty$. Define $C[S]$ to be the number of distinct subsets of S induced by C. Then, Sauer's Lemma [1] states that $C[S] \leq \left(\frac{en}{D}\right)^D$, and therefore the number of different pricing functions in $F_k(C)$ over S is at most $\left(\log_{1+\epsilon} h\right)^k \left(\frac{en}{D}\right)^{kD}$. Thus applying Theorem 2 here, and performing simple algebra (to remove the "$\ln n$" term from the right-hand-side) we get:

Corollary 8 *Given a β-approximation algorithm $\mathcal{A}$ for optimizing over $\mathcal{G} = F_k(C)$, then so long as $\mathrm{OPT}_{\mathcal{G}} \geq \beta n$ and the number of bidders n satisfies*

$$n \geq \frac{16h}{\epsilon^2}\left[\ln\left(\frac{2}{\delta}\right) + k\ln\left(\frac{1}{\epsilon}\ln h\right) + kD\ln\left(\frac{4kh}{\epsilon^2}\right)\right],$$

then with probability at least $1-\delta$, the profit of RSOPF$_{\mathcal{G},\mathcal{A}}$ is at least $(1-\epsilon)\,\mathrm{OPT}_{\mathcal{G}}/\beta$.

For certain classes C we can get better bounds. In the following, denote by C_k the concept class of unions of at most k sets from C. E.g., if C is the class of all axis parallel rectangles, then the VC-dimension of C_k is $O(kd)$ [8]. In these cases we can remove the $\log k$ term in our bounds, which is nice because it means we can interpret our results (e.g., Corollary 8) as charging OPT a penalty for each market it creates. However, we do not know how to remove this $\log k$ term in general, since in general the VC-dimension of C_k can be as large as $2Dk\log(2Dk)$ (see [2, 6]).

Corollary 8 gives a guarantee in the revenue of RSOPF$_{F_k(C),\mathcal{A}}$ so long as we have enough bidders n. In the following, for $k \geq 0$ let $\mathrm{OPT}_k = \mathrm{OPT}_{F_k(C)}$. We can also use Lemma 4 to show a bound that holds for all n, but with an additive loss term (we assume for simplicity here that $\beta = 1$):

Theorem 9 *For any given value of n, k, ϵ, and δ, with probability $1-\delta$, the revenue of RSOPF$_{F_k(C),\mathcal{A}}$ is*

$$(1-\epsilon)\,\mathrm{OPT}_k - h \cdot r_F(k, D, h, \epsilon, \delta),$$

where $r_F(k,D,h,\epsilon,\delta) = O\left(\frac{kD}{\epsilon^2}\ln\left(\frac{kDh}{\epsilon\delta}\right)\right)$.

Finally, using Theorem 5 we can extend our results to use SRM, where we want the algorithm to optimize over k, by viewing the additive loss term as a penalty function.

Theorem 10 *Let $\bar{\mathcal{G}}$ be the sequence of pricing function classes $F_1(C), F_2(C), \ldots, F_n(C)$, and let $\mathsf{pen}(F_k(C))$ be the additive-loss term below. Then for any value of n with probability $1-\delta$ the revenue of RSOPF-SRM$_{\bar{\mathcal{G}},\mathsf{pen}}$ is*

$$\max_k\left((1-\epsilon)\,\mathrm{OPT}_k - h \cdot r'_F(k, D, h, \epsilon, \delta)\right),$$

where $r'_F(k,D,h,\epsilon,\delta) = O\left(\frac{kD}{\epsilon^2}\ln\left(\frac{kDh}{\epsilon\delta}\right)\right)$.

To illustrate the relevance of Theorem 10, notice that for the special case of pricing using interval functions (the case of $d=1$ was studied in [3]), the following lower bound holds.

Theorem 11 *There is no randomized incentive compatible mechanism whose revenue is $\Omega\left(\max_k\left(\mathrm{OPT}_k - o(k)h\right)\right)$.*

A similar lower bound holds for most base classes; note also for the case of intervals on the line, an auction in [3] essentially matches this lower bound.

5.2 General Pricing Functions over the Attribute Space

In this section we generalize the results in Section 5.1 in two ways: to general classes of pricing functions (not just piecewise-constant functions defined over markets) and by removing the need for discretization by using the notion of covers. For example, we might want to consider a comparison class of linear functions over the attributes, or quadratic functions, or perhaps functions that divide the space into markets and are linear (rather than constant) in each market.

Assume that $\mathcal{X} \subseteq R^d$, and let $\mathcal{G}$ be fixed a class of pricing functions over the attribute space $\mathcal{X}$. For $g \in \mathcal{G}$ let $\rho_g : \mathcal{X} \times [1,h] \to R$ be its associated profit function. Let's denote by $\rho(\mathcal{G})$ be the class of the profit functions corresponding to $\mathcal{G}$. Consider $\mathrm{OPT}_{\mathcal{G}} = \mathrm{OPT}(S,\mathcal{G})$ to be the profit of the optimal pricing function in $\mathcal{G}$ over S. Now, let $\mathcal{G}_d$ be the class of decision surfaces (in $\mathbb{R}^{d+1}$) induced by $\mathcal{G}$: that is, to each $g \in \mathcal{G}$ we associate the set of all $(x,v) \in \mathcal{X} \times [1,h]$ such that $g(x) \leq v$. Finally, let $D = VCdim(\mathcal{G}_d)$. Assume in the following that $D < \infty$.

We now give our main lemma.

Lemma 12 *If we randomly partition S into S_1 and S_2, then $n \geq \frac{8h}{\epsilon^2}\left[\ln\left(\frac{2}{\delta}\right) + D\ln\left[\frac{ne}{D}\left(\frac{4}{\epsilon}\ln h + 1\right)\right]\right]$ bidders are sufficient so that with probability at least $1-\delta$ for all functions g in $\mathcal{G}$ we have $|g(S_1) - g(S_2)| \leq \epsilon \max[g(S), n]$.*

Proof Sketch: For each bidder (x, v) we conceptually introduce $O(\frac{1}{\alpha}\ln h)$ "phantom bidders" having the same attribute value x and bid values $1, (1+\alpha), (1+\alpha)^2, \cdots, h$ (we fix α shortly). Let S^* be the set S together with the set of all phantom bidders; let $n^* = |S^*|$. Let $Split$ be the set of possible splittings of S^* with surfaces from $\mathcal{G}_d$. We clearly have $|Split| \leq \mathcal{G}_d[n^*]$. For each element $s \in Split$ consider a representative function in $\mathcal{G}$ that induces splitting s in terms of its winning bidders, and let $Split_{\mathcal{G}}$ be the set of these representative functions. We then have that $\rho(Split_{\mathcal{G}})$ induces a multiplicative α-cover for $\rho(\mathcal{G})_{|S}$ with respect to the L_∞ norm. That is, for every function in $\mathcal{G}$ there is a function in $Split_{\mathcal{G}}$ that extracts nearly the same profit from every bidder. Moreover, by construction, for every function in $\rho_{g_1} \in \rho(\mathcal{H})$, there exists $\rho_g \in \rho(Split_{\mathcal{G}})$ such that for every $(x, v) \in S$, we have both $\rho_{g_1}((x,v)) \leq (1+\alpha)\rho_g((x,v))$ and $\rho_g((x,v)) \leq (1+\alpha)\rho_{g_1}((x,v))$. This implies that for for every function in $g_1 \in \mathcal{G}$, there exist $g \in Split_{\mathcal{G}}$ s.t.$|g_1(S_1) - g_1(S_2)| \leq \alpha g_1(S) + |g(S_1) - g(S_2)|$. Choosing $\alpha = \frac{\epsilon}{4}$, it follows that in order to prove the desired result it is enough to show that with probability at least $1-\delta$, for each function in $Split_{\mathcal{G}}$ we have $|g(S_1) - g(S_2)| \leq \frac{\epsilon}{2}\max[g(S), n]$. This is true since by Lemma 1 for a fixed $g \in Split_{\mathcal{G}}$ we have $\mathbf{Pr}\{|g(S_1) - g(S_2)| \geq \frac{\epsilon}{2}\max[g(S), n]\} \leq 2e^{\left[-\frac{\epsilon^2 n}{8h}\right]}$; we also have $|Split_{\mathcal{G}}| \leq \left(\frac{n^* e}{D}\right)^D$. ■

Simple algebra (to remove the "n" on the RHS) yields:

Corollary 13 *If we randomly partition S into S_1 and S_2, then $n \geq \frac{16h}{\epsilon^2}\left[\ln\left(\frac{2}{\delta}\right) + D\ln\left(\frac{16h}{\epsilon^2}\left(\frac{4}{\epsilon}\ln h + 1\right)\right)\right]$ bidders are sufficient so that with probability at least $1-\delta$ for all functions g in $\mathcal{G}$ we have $|g(S_1) - g(S_2)| \leq \epsilon \max[g(S), n]$.*

Corollary 13 together with an analysis similar with the one in Theorem 2 imply that:

Theorem 14 *Given comparison class $\mathcal{G}$ and a β-approximation algorithm $\mathcal{A}$ for optimizing over $\mathcal{G}$, then so long as $\mathrm{OPT}_{\mathcal{G}} \geq \beta n$ and the number of bidders n satisfies*

$$n \geq \frac{64h}{\epsilon^2}\left[\ln\left(\frac{2}{\delta}\right) + D\ln\left(\frac{64h}{\epsilon^2}\left(\frac{16}{\epsilon}\ln h + 1\right)\right)\right],$$

then with probability at least $1-\delta$, the profit of RSOPF$_{(\mathcal{G},\mathcal{A})}$ is at least $(1-\epsilon)\,\mathrm{OPT}_{\mathcal{G}}/\beta$.

6 Combinatorial Auctions

Combinatorial auctions have received much attention in recent years because of the difficulty of merging the complexity issue of computing an optimal outcome with the game-theoretic issue of incentive compatibility. To date almost exclusively the focus has been on socially optimal combinatorial auctions. Deviating from this literature, we consider the goal of profit maximization of the seller in the case where the items for sale are available in unlimited supply. In this section we consider the general version of the combinatorial auction problem as well as the special cases of *unit-demand* bidders (bidders desire only singleton bundles) and *single-minded* bidders (each bidder has a single desired bundle).

It is interesting to restrict our attention to the case of item-pricing, where the auctioneer intuitively is attempting to set a price for each of the distinct items and bidders then choose their favorite bundle given these prices. Item-pricing is without loss of generality for the unit-demand case, and the general bundle-pricing can be realized with an auction with $m' = 2^m$ "items", one for each of possible bundle of the original m items.[2]

For combinatorial auctions, the size of the class of all possible item-pricings, $|\mathcal{G}|$, is infinite. Nonetheless, we can use the technique of counting possible outputs (See Section 3.3) to get a bound on the performance of the random sampling auction. This approach calls for bounding $|\mathcal{G}_{\mathcal{A}}|$, the number of different pricings RSOPF$_{(\mathcal{G},\mathcal{A})}$ can possibly output. We restrict our analysis here to considering exact algorithms for computing the optimal item pricing; for a discussion of this approach for approximation algorithms, see the full version of the paper. Our results for this approach are summarized in the first row of Table 1 and proofs of these results are given in the full version of the paper.

We can obtain better bounds if we are willing to optomize over a smaller class of discretized item-pricings (again, see Section 3.3). In particular, if we can find a small class $\mathcal{G}'$ with the property that $\mathrm{OPT}_{\mathcal{G}'}$ is guaranteed to be close to $\mathrm{OPT}_{\mathcal{G}}$, we can argue that RSOPF$_{(\mathcal{G}',\mathcal{A})}$ performs well compared to $\mathrm{OPT}_{\mathcal{G}}$ using bounds on the size of $|\mathcal{G}'|$. No such small set $\mathcal{G}'$ is known to exist for item-pricing in general combinatorial auctions; however, for the unit-demand and single-minded special cases we can use the classes of discretized item-pricings constructed in [13]. Note that these constructions are not as simple as the discretization for digital-good auctions (Section 4). The discretization results from [13] are summarized in the second row of Table 1.

We can apply Theorem 2 to the sizes of the complexity classes in Table 1 to get good bounds on the profit of random sampling auctions for combinatorial item pricing. In particular, we get that $\tilde{O}(hm^2/\epsilon^2)$ bidders are sufficient to

[2] We make the assumption that all desired bundles contain at most one of each item. This assumption can be easily relaxed and our results applied given any bound on the number of copies of each item that are desired by any one consumer.

	general	unit-demand	single-minded
$\|\mathcal{G}_\mathcal{A}\|$	$n^m 2^{2m^2}$	$n^m(m+1)^{2m}$	n^m
$\|\mathcal{G}'\|$		$O(m^m \log^m_{1+\epsilon} \frac{n}{\epsilon})$	$O(\log^m_{1+\epsilon} \frac{n}{\epsilon})$

Table 1. Size of comparison classes for combinatorial auctions.

achieve revenue close to the optimum item-pricing in the general case, and $\tilde{O}(hm/\epsilon^2)$ bidders are sufficient for the unit-demand case. Also, by using Corollary 3 instead of Theorem 2 we can replace the condition on the number of bidders with a condition on $\mathrm{OPT}_\mathcal{G}$, which is factor of m improvement on the bound given by [12].

7 Multicast Pricing

In the multicast pricing problem, each bidder resides at some node of a tree, and in order to sell its service to some bidder, the service-provider must have purchased all edges on the path from the root to that vertex. Given a set of edge costs, our goal as service-provider is to determine a subtree together with prices at nodes of this tree that achieves highest revenue minus cost. A 4-approximation to this problem, under the assumption that the optimal solution has revenue at least 4 times its cost and that there is sufficient competition at each node is given in [7].

Using our generic results we can say that so long as the optimal solution has revenue at least $1/\epsilon$ times its cost, and we have on average $\tilde{O}(h/\epsilon^2)$ bidders at each node (using Theorem 2) or at least $\tilde{O}(h/\epsilon^2)$ revenue at each node (using Corollary 3) then we get a $(1+O(\epsilon))$-approximation.

Briefly, to apply the generic results, we define our algorithm $\mathcal{A}$ so that it finds the revenue-maximizing tree but *only over* the subset of trees whose revenue on the given subset of bidders is at least $(2+\epsilon)/\epsilon$ times its cost. By Corollary 3, with high probability the optimal tree has this property over both S_1 and S_2, and so the revenue achieved by $\mathcal{A}$ is nearly that of the optimal tree, and by design the cost of the tree produced by $\mathcal{A}$ is only an $O(\epsilon)$ factor of revenue.

We can also apply structural-risk-minimization in the case that the total number of bidders is not sufficient for the entire class of trees. In particular, one interesting case is the comparison-class of functions that choose some subtree and add fake "markups" between 0 and nh to the edges of that subtree, and then perform cost-sharing on the result (also add a "super-root" with a single zero-cost edge into the root). If we define $\mathcal{G}_k$ to be the set of such functions whose subtree has k edges, then $|\mathcal{G}_k| \leq (n \log_{1+\epsilon}(nh))^k$. We can then perform SRM using Theorem 5. An interesting special case to consider is a simple depth-1 multicast tree whose edges have cost 0 and with two bidders at each leaf: one with value 1 and one with value h. In this case, there is not sufficient competition at the leaves for the results of [7], but we can extract $\Omega(nh)$ using $\mathcal{G}_1$.

8. Conclusions

In this work we have made the connection between machine learning and mechanism design explicit. In doing so, we obtain a unified approach to considering a variety of profit maximizing mechanism design problems including many that have been previously considered in the literature.

Some of our techniques give suggestions for the *design* of mechanisms and others for their *analysis*. In terms of design, these include the use of discretization to produce smaller function classes, and the use of structural-risk-minimization to choose an appropriate level of complexity of the mechanism for a given set of bidders. In terms of analysis, these include both the use of basic sample-complexity arguments, and the notion of multiplicative covers for better bounding the true complexity of a given set of functions.

Our bounds on random sampling auctions for digital goods [11] not only show how the auction profit approaches the optimal profit, but also weaken the required assumptions by a constant factor. Similarly for random sampling auctions for multiple digital goods [12] our unified analysis gives a bound that approaches the optimal profit with assumptions weakened by a factor of more than m, the number of distinct items. This multiple digital good auction problem is a special case of the a more general unlimited supply combinatorial auction problem for which we obtain the first positive worst-case results by showing that it is possible to approximate the optimal profit with an incentive-compatible mechanism. Furthermore, unlike the case for combinatorial auctions for social welfare maximization, our incentive-compatible mechanisms can be based on approximation algorithms instead of exact ones.

We have also explored the attribute auction problem proposed in [3], a special case of general profit maximizing mechanism design, in a very general setting: the attribute values can be multi-dimensional and the target pricing functions considered can be arbitrarily complex. We bound the performance of random sampling auctions as a function of the complexity of the target pricing functions. Our attribute auction results can be used for more general problems such as multicast pricing, where there is a cost to be paid by the mechanism that is a function of its outcome.

Our random sampling auctions assume the existence of exact or approximate pricing algorithms. Solutions to these pricing problem have been proposed for several of our settings. In particular, optimal item-pricings for combinatorial auctions in the single-minded and unit-demand special cases have been considered in [15, 13]. On the other hand for attribute auctions, many of the clustering and market-segmenting pricing algorithms have yet to be considered at all.

Probably the most important direction for future work is in relaxing the assumption that the items for sale are available in unlimited supply. In the random sampling framework, we propose the following mechanism: randomly partition the bidders into two sets, evenly divide the items among the two sets, compute the optimal *envy-free*[3] pricing function for the two partitions, and applying the pricing function to the opposite partition. Of course, a pricing function g that is envy-free for S_1 may not necessarily be envy-free for S_2. There are several approaches that may work here. First, we could artificially deplete the supply by a constant factor and ask for an pricing function that is envy-free for the depleted supply. Then it may be possible to argue that it is envy-free for both S_1 and S_2 with high probability. Another option would be to take the bidders of S_1 in an arbitrary (or random) order and allow them to take an item if they desire one. When we run out of items, stop. The remaining bidders get none, whether they want one or not. It is easy to see that the technique outlined above results in an incentive compatible mechanism. Is it also close to optimal?

It is possible to further generalize the feasibility constraints imposed by limited supply to arrive at the general single-parameter agent auction problem (See e.g., [9] for a precise definition). This abstract problem can be viewed as auctioning a service to a number of agents where the service provider must pay a cost that is a function of the agents served. In its full generality, this cost function could be arbitrary. Note that the multicast pricing problem is a special case of this problem where the cost function is defined by a tree. The possibly asymmetric cost function can be viewed as endowing the agents with public attributes, or the agents could have additional attributes. A very interesting direction for future research is in determining for what classes of cost functions the general problem of profit maximization in this setting can be solved.

References

[1] M. Anthony and P. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 1999.

[2] P. Bartlett and W. Maass. Vapnik Chervonenkis Dimension of Neural Nets. In *The Handbook of Brain Theory and Neural Networks*. MIT Press, 2003.

[3] A. Blum and J. Hartline. Near-Optimal Online Auctions. In *Proc. 16th Symp. on Discrete Algorithms*, 2005.

[4] A. Blum, V. Kumar, A. Rudra, and F. Wu. Online Learning in Online Auctions. In *Proc. 14th Symp. on Discrete Algorithms*, 2003.

[5] L. Devroye, L. Gyorfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer-Verlag, 1996.

[6] L. Devroye and G. Lugosi. *Combinatorial Methods in Density Estimation*. Springer-Verlag, 2001.

[7] A. Fiat, A. Goldberg, J. Hartline, and A. Karlin. Competitive Generalized Auctions. In *Proc. 34th ACM Symposium on the Theory of Computing*. ACM Press, New York, 2002.

[8] P. Fische and S. Kwek. Minimizing Disagreement for Geometric Regions Using Dynamic Programming, with Applications to Machine Learning and Computer Graphics. 1996.

[9] A. Goldberg and J. Hartline. Collusion-Resistant Mechanisms for Single-Parameter Agents. In *Proc. 16th Symp. on Discrete Algorithms*, 2005.

[10] A. Goldberg, J. Hartline, A. Karlin, M. Saks, and A. Wright. Competitive auctions and digital goods. *Games and Economic Behavior*, 2002. Submitted for publication. An earlier version available as InterTrust Technical Report STAR-TR-99.09.01.

[11] A. Goldberg, J. Hartline, and A. Wright. Competitive Auctions and Digital Goods. In *Proc. 12th Symp. on Discrete Algorithms*, 2001.

[12] J. Hartline and A. Goldberg. Competitive auctions for multiple digital goods. In *ESA*, 2001.

[13] J. Hartline and Vladlen Koltrun. Near-Optimal Pricing in Near-Linear Time. In *Proc. of 9th Workshop on Algorithms and Data Structures*. Springer-Verlag, 2005.

[14] J. Hartline and R. McGrew. From optimal limited to unlimited supply auctions. In *EC*, 2005.

[15] V. Guruswami and J. Hartline and A. Karlin and D. Kempe and C. Kenyon, and F. McSherry. On Profit-Maximizing Envy-Free Pricing. In *Proc. 16th Symp. on Discrete Algorithms*, 2005.

[3] To generalize envy-freedom [15] to attribute auctions, declare a price function $g \in \mathcal{G}$ envy-free for bidders S if there are enough items such that all bidders that have strictly positive utility for an item under g can simultaneously be sold one.

A Concentration inequalities

Here is the McDiarmid inequality (see [5]) we use in our proofs:

Theorem 15 *Let $Y_1, ..., Yn$ be independent random variables taking values in some set A, and assume that $t : A \to R$ satisfies:*

$$\sup_{y_1,\ldots,y_n \in A, \overline{y}_i \in A} |t(y_1, ..., y_n) - t(y_1, ..., y_{i-1}, \overline{y}_i, y_{i+1}, y_n)| \leq c_i,$$

for all i, $1 \leq i \leq n$. Then for all $\gamma > 0$ we have:

$$\mathbf{Pr}\left\{|t(Y_1, ..., Y_n) - \mathbf{E}[t(Y_1, ..., Y_n)]| \geq \gamma\right\} \leq 2e^{\left[-\frac{2\gamma^2}{\sum_{i=1}^{N} c_i^2}\right]}$$

Here is also a consequence of the Chernoff bound that we used in Lemma 7.

Theorem 16 *Let $X_1, ..., X_n$ be independent Poisson trials such that, for $1 \leq i \leq n$, $\mathbf{Pr}[X_i = 1] = 1/2$ and let $X = \sum_{i=1}^{n} X_i$. Then any n' we have:*

$$\mathbf{Pr}\left\{\left|X - \frac{n}{2}\right| \geq \epsilon \max\{n, n'\}\right\} \leq 2e^{\left[-2n'\epsilon^2\right]}$$