6.853 Algorithmic Game Theory and Data Science	April 7, 2017
Lecture 16	
Lecturer: Vasilis Syrgkanis	*Scribe: Vasilis Syrgkanis*

1 Introduction

We look at the sample complexity of single-item optimal auctions among n bidders. Each bidder has a value function v_i drawn independently from a distribution D_i and we denote with D the joint distribution.

We assume we are given a sample set $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$, of m valuation vectors, where each $\mathbf{v}_t \sim D$. Let H denote the class of all dominant strategy truthful single item auctions (i.e. auctions where no player has incentive to report anything else other than his true value to the auction, independent of what other players do). Moreover, let

$$\mathtt{r}(h, \mathbf{v}) = \sum_{i=1}^{n} p_i^h(\mathbf{v}) \tag{1}$$

where $p_i^h(\cdot)$ is the payment function of mechanism h, and $\mathtt{r}(h, \mathbf{v})$ is the revenue of mechanism h on valuation vector $\mathbf{v}$. Finally, let

$$\mathrm{R}_D(h) = \mathbb{E}_{\mathbf{v}\sim D}\left[\mathtt{r}(h, \mathbf{v})\right] \tag{2}$$

be the expected revenue of mechanism h under the true distribution of values D.

Given a sample S of size m, we want to compute a dominant strategy truthful mechanism h_S, such that:

$$\mathbb{E}_S\left[\mathrm{R}_D(h_S)\right] \geq \sup_{h\in H} \mathrm{R}_D(h) - \epsilon(m) \tag{3}$$

where $\epsilon(m) \to 0$ as $m \to \infty$. More formally, we can define the sample complexity of an auction class as:

Definition 1 (Sample Complexity of Auction Class)**.** *The (additive error) sample complexity of an auction class H and a class of distributions D, for an accuracy target ϵ is defined as the smallest number of samples $m(\epsilon)$, such that for any $m \geq m(\epsilon)$:*

$$\mathbb{E}_S\left[\mathrm{R}_D(h_S)\right] \geq \sup_{h\in H} \mathrm{R}_D(h) - \epsilon \tag{4}$$

We might also be interested in a multiplcative error sample complexity, i.e.

$$\mathbb{E}_S\left[\mathrm{R}_D(h_S)\right] \geq (1-\epsilon) \sup_{h\in H} \mathrm{R}_D(h) \tag{5}$$

If one assumes that the optimal revenue on the distribution is lower bounded by some constant quantity, then an additive error implies a multiplicative error. For instance, if one assumes that players values are bounded away from zero with significant probability, then that implies a lower bound on revenue. Such assumptions for instance, are made in the work of [7]. We will mostly focus on additive error in this work.

We might also be interested in proving high probability guarantees, i.e. with probability $1 - \delta$:

$$\mathrm{R}_D(h_S) \geq \sup_{h\in H} \mathrm{R}_D(h) - \epsilon(m, \delta) \tag{6}$$

where for any δ, $\epsilon(m, \delta) \to 0$ as $m \to \infty$.

2 Generalization Error via the Split-Sample Growth Rate

We turn to the general PAC learning framework, and we give generalization guarantees in terms of a new notion of complexity of a hypothesis space H, which we denote as split-sample growth rate.

Consider an arbitrary hypothesis space H and an arbitrary data space Z, and suppose we are given a set S of m samples $\{z_1, \ldots, z_m\}$, where each z_t is drawn i.i.d. from some distribution D on Z. We are interested in maximizing some reward function $\mathtt{r} : H \times Z \to [0,1]$, in expectation over distribution D. In particular, denote with $\mathrm{R}_D(h) = \mathbb{E}_{z\sim D}[\mathtt{r}(h,z)]$.

We will look at the Expected Reward Maximization algorithm on S, with some fixed tie-breaking rule. Specifically, if we let

$$\mathrm{R}_S(h) = \frac{1}{m}\sum_{t=1}^{m} \mathtt{r}(h, z_t) \tag{7}$$

then ERM is defined as:

$$h_S = \arg\sup_{h\in H} \mathrm{R}_S(h) \tag{8}$$

where ties are broken based on some pre-defined manner.

We define the notion of a split-sample hypothesis space:

Definition 2 (Split-Sample Hypothesis Space). *For any sample S, let $\hat{H}_S$, denote the set of all hypothesis h_T output by the ERM algorithm (with the pre-defined tie-breaking rule), on any subset $T \subset S$, of size $\lceil |S|/2 \rceil$, i.e.:*

$$\hat{H}_S = \{h_T : T \subset S, |T| = \lceil |S|/2 \rceil\} \tag{9}$$

Based on the split-sample hypothesis space, we also define the split-sample growth rate of a hypothesis space H at value m, as the largest possible size of $\hat{H}_S$ for any set S of size m.

Definition 3 (Split-Sample Growth Rate). *The split-sample growth rate of a hypothesis H and an ERM process for H, is defined as:*

$$\hat{\tau}_H(m) = \sup_{S:|S|=m} |\hat{H}_S| \tag{10}$$

We first show that the generalization error is upper bounded by the Rademacher complexity evaluated on the split-sample hypothesis space of the union of two samples of size m. The Rademacher complexity $\mathcal{R}(S,H)$ of a sample S of size m and a hypothesis space H is defined as:

$$\mathcal{R}(S,H) = \mathbb{E}_\sigma\left[\sup_{h\in H} \frac{2}{m}\sum_{z_t\in S} \sigma_t \cdot \mathtt{r}(h,z_t)\right] \tag{11}$$

where $\sigma = (\sigma_1, \ldots, \sigma_m)$ and each σ_t is an independent binary random variable taking values $\{-1, 1\}$, each with equal probability.

Lemma 1. *For any hypothesis space H, and any fixed ERM process, we have:*

$$\mathbb{E}_S[\mathrm{R}_D(h_S)] \geq \sup_{h\in H} \mathrm{R}_D(h) - \mathbb{E}_{S,S'}\left[\mathcal{R}(S, \hat{H}_{S\cup S'})\right], \tag{12}$$

where S and S' are two independent samples of some size m.

Proof. Let h_* be the optimal hypothesis for distribution D. First we re-write the left hand side, by adding and subtracting the expected empirical reward:

$$\begin{aligned}
\mathbb{E}_S[\mathrm{R}_D(h_S)] &= \mathbb{E}_S[\mathrm{R}_S(h_S)] - \mathbb{E}_S[\mathrm{R}_S(h_S) - \mathrm{R}_D(h_S)] \\
&\geq \mathbb{E}_S[\mathrm{R}_S(h_*)] - \mathbb{E}_S[\mathrm{R}_S(h_S) - \mathrm{R}_D(h_S)] && (h_S \text{ maximizes empirical reward}) \\
&= \mathrm{R}_D(h_*) - \mathbb{E}_S[\mathrm{R}_S(h_S) - \mathrm{R}_D(h_S)] && (h_* \text{ is independent of } S)
\end{aligned}$$

Thus it suffices to upper bound the second quantity in the above equation.

Since $\mathrm{R}_D(h) = \mathbb{E}_{S'}[\mathrm{R}_{S'}(h)]$ for a fresh sample S' of size m, we have:

$$\begin{aligned}\mathbb{E}_S[\mathrm{R}_S(h_S) - \mathrm{R}_D(h_S)] =&\ \mathbb{E}_S[\mathrm{R}_S(h_S) - \mathbb{E}_{S'}[\mathrm{R}_{S'}(h_S)]]\\ =&\ \mathbb{E}_{S,S'}[\mathrm{R}_S(h_S) - \mathrm{R}_{S'}(h_S)]\end{aligned}$$

Now, consider the set $\hat{H}_{S\cup S'}$. Since S is a subset of $S \cup S'$ of size $|S \cup S'|/2$, we have by the definition of the split-sample hypothesis space that $h_S \in \hat{H}_{S\cup S'}$. Thus we can upper bound the latter quantity by taking a supremum over $h \in \hat{H}_{S\cup S'}$:

$$\begin{aligned}\mathbb{E}_S[\mathrm{R}_S(h_S) - \mathrm{R}_D(h_S)] \leq&\ \mathbb{E}_{S,S'}\left[\sup_{h\in\hat{H}_{S\cup S'}} \mathrm{R}_S(h) - \mathrm{R}_{S'}(h)\right]\\ =&\ \mathbb{E}_{S,S'}\left[\sup_{h\in\hat{H}_{S\cup S'}} \frac{1}{m}\sum_{t=1}^m (\mathtt{r}(h,z_t) - \mathtt{r}(h,z'_t))\right]\end{aligned}$$

Now observe, that we can rename any sample $z_t \in S$ to z'_t and sample $z'_t \in S'$ to z_t. By doing show we do not change the distribution. Moreover, we do not change the quantity $H_{S\cup S'}$, since $S \cup S'$ is invariant to such swaps. Finally, we only change the sign of the quantity $(\mathtt{r}(h,z_t) - \mathtt{r}(h,z'_t))$. Thus if we denote with $\sigma_t \in \{-1,1\}$, a Rademacher variable, we get the above quantity is equal to:

$$\mathbb{E}_{S,S'}\left[\sup_{h\in\hat{H}_{S\cup S'}} \frac{1}{m}\sum_{t=1}^m (\mathtt{r}(h,z_t) - \mathtt{r}(h,z'_t))\right] = \mathbb{E}_{S,S'}\left[\sup_{h\in\hat{H}_{S\cup S'}} \frac{1}{m}\sum_{t=1}^m \sigma_t (\mathtt{r}(h,z_t) - \mathtt{r}(h,z'_t))\right] \tag{13}$$

for any vector $\sigma = (\sigma_1, \ldots, \sigma_m) \in \{-1,1\}^m$. The latter also holds in expectation over σ, where σ_t is randomly drawn between $\{-1,1\}$ with equal probability. Hence:

$$\mathbb{E}_S[\mathrm{R}_S(h_S) - \mathrm{R}_D(h_S)] \leq \ \mathbb{E}_{S,S',\sigma}\left[\sup_{h\in\hat{H}_{S\cup S'}} \frac{1}{m}\sum_{t=1}^m \sigma_t (\mathtt{r}(h,z_t) - \mathtt{r}(h,z'_t))\right]$$

By splitting the supremma into a positive and negative part and observing that the two expected quantities are identical, we get:

$$\begin{aligned}\mathbb{E}_S[\mathrm{R}_S(h_S) - \mathrm{R}_D(h_S)] \leq&\ 2\mathbb{E}_{S,S',\sigma}\left[\sup_{h\in\hat{H}_{S\cup S'}} \frac{1}{m}\sum_{t=1}^m \sigma_t \mathtt{r}(h,z_t)\right]\\ =&\ \mathbb{E}_{S,S'}\left[\mathcal{R}(S, \hat{H}_{S\cup S'})\right]\end{aligned}$$

where $\mathcal{R}(S,H)$ denotes the Rademacher complexity of a sample S and hypothesis H. □

Observe, that the latter theorem is a strengthening of the fact that the Rademacher complexity upper bounds the generalization error, simply because:

$$\mathbb{E}_{S,S'}\left[\mathcal{R}(S, \hat{H}_{S\cup S'})\right] \leq \mathbb{E}_{S,S'}[\mathcal{R}(S,H)] = \mathbb{E}_S[\mathcal{R}(S,H)] \tag{14}$$

Thus if we can bound the Rademacher complexity of H, then the latter lemma gives a bound on the generalization error. However, the reverse might not be true. Finally, we show our main theorem, which shows that if the split-sample hypothesis space has small size, then we immediately get a generalization bound, without the need to further analyze the Rademacher complexity of H.

Theorem 1 (Main Theorem)**.** *For any hypothesis space H, and any fixed ERM process, we have:*

$$\mathbb{E}_S[\mathrm{R}_D(h_S)] \geq \sup_{h\in H} \mathrm{R}_D(h) - \sqrt{\frac{2\log(\hat{\tau}_H(2m))}{m}} \tag{15}$$

Moreover, with probability $1-\delta$:

$$\mathrm{R}_D(h_S) \geq \sup_{h\in H} \mathrm{R}_D(h) - \frac{1}{\delta}\sqrt{\frac{2\log(\hat{\tau}_H(2m))}{m}} \tag{16}$$

Proof. By applying Massart's lemma (see previous lecture) we have that:

$$\mathcal{R}(S, \hat{H}_{S\cup S'}) \leq \sqrt{\frac{2\log(|\hat{H}_{S\cup S'}|)}{m}} \leq \sqrt{\frac{2\log(\hat{\tau}_H(2m))}{m}} \tag{17}$$

Combining the above with Lemma 1, yields the first part of the theorem.

Finally, the high probability statement follows from observing that the random variable $\sup_{h\in H} R_D(h) - R_D(h_S)$ is non-negative and by applying Markov's inequality: with probability $1-\delta$

$$\sup_{h\in H} R_D(h) - R_D(h_S) \leq \frac{1}{\delta}\mathbb{E}_S\left[\sup_{h\in H} R_D(h) - R_D(h_S)\right] \leq \frac{1}{\delta}\sqrt{\frac{2\log(\hat{\tau}_H(2m))}{m}} \tag{18}$$

□

The latter theorem can be trivially extended to the case when $\mathtt{r} : H \times Z \to [\alpha, \beta]$, leading to a bound of the form:

$$\mathbb{E}_S\left[\mathrm{R}_D(h_S)\right] \geq \sup_{h\in H} \mathrm{R}_D(h) - (\beta-\alpha)\sqrt{\frac{2\log(\hat{\tau}_H(2m))}{m}} \tag{19}$$

We note that unlike the standard Rademacher complexity, which is defined as $\mathcal{R}(S, H)$, our bound, which is based on bounding $\mathcal{R}(S, \hat{H}_{S\cup S'})$ for any two datasets S, S' of equal size, does not imply a high probability bound via McDiarmid's inequality (see e.g. Chapter 26 of [12] of how this is done for Rademacher complexity analysis), but only via Markov's inequality. The latter yields a worse dependence on the confidence δ on the high probability bound of $1/\delta$, rather than $\log(1/\delta)$. The reason for the latter is that the quantity $\mathcal{R}(S, \hat{H}_{S\cup S'})$, depends on the sample S, not only in terms of on which points to evaluate the hypothesis, but also on determining the hypothesis space $\hat{H}_{S\cup S'}$. Hence, the function:

$$f(z_1, \ldots, z_m) = \mathbb{E}_{S'}\left[\sup_{h\in \hat{H}_{\{z_1,\ldots,z_m\}\cup S'}} \frac{1}{m}\sum_{t=1}^{m} \sigma_t \left(\mathtt{r}(h, z_t) - \mathtt{r}(h, z_t')\right)\right] \tag{20}$$

does not satisfy the stability property that $|f(\mathbf{z}) - f(z_i'', \mathbf{z}_{-i})| \leq \frac{1}{m}$. The reason being that the supremum is taken over a different hypothesis space in the two inputs. This is unlike the case of the function:

$$f(z_1, \ldots, z_m) = \mathbb{E}_{S'}\left[\sup_{h\in H} \frac{1}{m}\sum_{t=1}^{m} \sigma_t \left(\mathtt{r}(h, z_t) - \mathtt{r}(h, z_t')\right)\right] \tag{21}$$

which is used in the standard Rademacher complexity bound analysis, which satisfies the latter stability property.

3 Sample Complexity of Auctions via Split-Sample Growth

We now present the application of the latter measure of complexity to the analysis of the sample complexity of revenue optimal auctions. Thoughout this section we assume that the revenue of any auction lies in the range $[0, 1]$. The results can be easily adapted to any other range $[\alpha, \beta]$, by re-scaling the equations, which will lead to blow-ups in the sample complexity of the order of an extra $(\beta - \alpha)$ multiplicative factor. This limits the results here to bounded distributions of values. However, as was shown in [3], one can always cap the distribution of values up to some upper bound, for the case of regular distributions, by losing only an ϵ fraction of the revenue. So one can apply the results below on this capped distribution.

3.1 A Few Simple Applications

Single bidder and single item. Consider the case of a single bidder and single item auction. In this setting, the space of hypothesis is $H = \{\text{post a reserve price } r \text{ for } r \in [0, 1]\}$. We consider, the ERM rule, which for any set S, in the case of ties, it favors reserve prices that are equal to some valuation $v_t \in S$. Wlog assume that samples $v_1, \ldots, v_m$ are ordered in increasing order. Observe, that for any set S, this ERM rule on any subset T of S, will post a reserve price that is equal to some value $v_t \in T$. Any other reserve price in between two values $[v_t, v_{t+1}]$ is weakly dominated by posting $r = v_{t+1}$, as it does not change which samples are allocated and we can only increase revenue. Thus the space $\hat{H}_S$ is a subset of {post a reserve price $r \in \{v_1, \ldots, v_m\}$}. The latter is of size m. Thus the split-sample growth of H is $\hat{\tau}_H(m) \leq m$. This yields:

$$\mathbb{E}_S \left[\mathrm{R}_D(h_S)\right] \geq \sup_{h \in H} \mathrm{R}_D(h) - \sqrt{\frac{2\log(2m)}{m}} \tag{22}$$

Equivalently, the sample complexity is $m_H(\epsilon) = O\left(\frac{\log(1/\epsilon)}{\epsilon^2}\right)$.

Multiple i.i.d. regular bidders and single item. In this case the space of hypotheses are the space of second price auctions with some reserve $r \in [0, 1]$. Again if we consider ERM which in case of ties favors a reserve that equals to a value in the sample (assuming that is part of the tied set, or outputs any other value otherwise), then observe that for any subset T of a sample S, ERM on that subset will pick a reserve price that is equal to one of the values in the samples S. Thus $\hat{\tau}_H(m) \leq n \cdot m$. This yields:

$$\mathbb{E}_S \left[\mathrm{R}_D(h_S)\right] \geq \sup_{h \in H} \mathrm{R}_D(h) - \sqrt{\frac{2\log(2 \cdot n \cdot m)}{m}} \tag{23}$$

Equivalently, the sample complexity is $m_H(\epsilon) = O\left(\frac{\log(n/\epsilon^2)}{\epsilon^2}\right)$.

Non-i.i.d. regular bidders, single item, second price with player specific reserves. In this case the space of hypotheses H_{SP} are the space of second price auctions with some reserve $r_i \in [0, 1]$ for each player i. Again if we consider ERM which in case of ties favors a reserve that equals to a value in the sample (assuming that is part of the tied set, or outputs any other value otherwise), then observe that for any subset T of a sample S, ERM on that subset will pick a reserve price r_i that is equal to one of the values v_t^i of player i in the sample S. There are m such possible choices for each player, thus m^n possible choices of reserves in total. Thus $\hat{\tau}_H(m) \leq m^n$. This yields:

$$\mathbb{E}_S \left[\mathrm{R}_D(h_S)\right] \geq \sup_{h \in H_{SP}} \mathrm{R}_D(h) - \sqrt{\frac{2n\log(2m)}{m}} \tag{24}$$

If H is the space of all dominant strategy truthful mechanisms, then by prophet inequalities (see [5]), we know that $\sup_{h \in H_{SP}} \mathrm{R}_D(h) \geq \frac{1}{2} \sup_{h \in H} \mathrm{R}_D(h)$. Thus:

$$\mathbb{E}_S \left[\mathrm{R}_D(h_S)\right] \geq \frac{1}{2} \sup_{h \in H} \mathrm{R}_D(h) - \sqrt{\frac{2n\log(2m)}{m}} \tag{25}$$

3.2 Non-i.i.d. regular bidders single item

In this case the space of hypotheses are the space of all virtual welfare maximizing auctions: For each player i, pick a monotone function $\hat{\phi}_i(v_i) \in [-1, 1]$ and allocate to the player with the highest non-negative virtual value, charging him the lowest value he could have bid and still win the item. If no player has non-negative virtual value, then don't allocate the item.

A first failed attempt. Observe, that for any such auction, ERM on a sample S of size m, will output a function $\hat{\phi}_i(\cdot)$ for each player that is a step function on the observed valuations (see Figure 1. The reasoning is the following: at every realization of the valuations in the sample, each player is paying the smallest value he could have submitted and still win the item. Now consider a single player i and let $v_i^1 \leq \ldots \leq v_i^m$ be the valuations for this player that appear in the sample S. Consider a virtual value

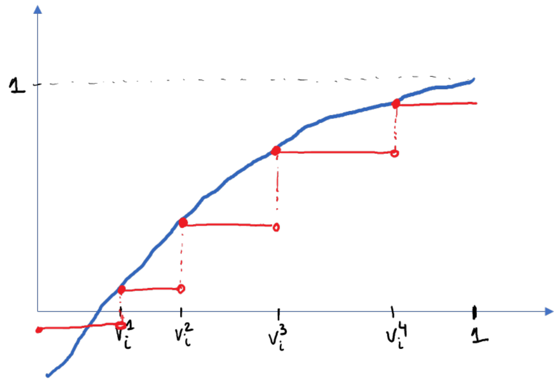

Figure 1: Any non-step function (blue line), can be transformed to a step function (red line), which yields point-wise higher revenue, for any valuation vector in the sample S.

function $\hat{\phi}_i(v_i)$ that is not a step function, i.e. the function is not constant in the region between two values $[v_i^t, v_i^{t+1})$. By setting the value of the function in that region, equal to $\hat{\phi}_i(v_i^t)$, then we are not changing the allocation function for any of the realized valuation vectors in the sample. Moreover, the new function that we created is pointwise lower than the previous virtual value function. Thus any time that player i wins for some realization of the value vector, the minimum value he needs to report to still win, is weakly larger. Thus we are collecting more revenue, for any realization $\mathbf{v}^t$ in S, by doing this modification.

Moreover, for any subset T of S, the ERM on T also outputs for each player i a step virtual value function $\hat{\phi}_i$ which increases only at points $v_i^t \in T$. The latter is also obviously included in the class of step functions on values $v_i^t \in S$, since these are more valuation points. Thus:

$$\hat{H}_S \subseteq \{\text{Virtual value maximizers, w. each virtual value function being a step function on points } v_i^t \in S\}$$

An alternative way of expressing $\hat{H}_S$ is the following: for each player i and for each $v_i^t \in S$, assign a rank $\sigma(i, v_i^t) \in [n \cdot m]$. Moreover, this ranking is monotone for each bidder. Then the mechanism for each realization $\mathbf{v}^t$ allocates to the player with the highest rank, charging him the smallest $v_i^{t'}$ for which he remains the winner. We can count the number of all such mechanisms as follows: for each player i and for each rank $t \in [n \cdot m]$, we can report the smallest value $v_i^t \in S$, for which the rank of the player i goes above t. This uniquely defines a monotone ranking for player i. The number of all such possible rankings for a player is $m^{n \cdot m}$, leading to the bound: $|\hat{H}_S| \leq m^{n^2 \cdot m}$. Thus: $\hat{\tau}_H(m) \leq m^{n^2 \cdot m}$.

However, the latter is still very large and as the number of samples m is in the exponent. Thus the split-sample growth results will not give a bound which decays to zero. This is not a problem of the split-sample growth approach, but rather the space of all virtual welfare maximizers is too large to run ERM on top of it. To deal with this will consider a coarsening H^ϵ of our hypothesis space H, and run ERM on H^ϵ, rather than H. Then we will argue that the best mechanism in H^ϵ is ϵ-close in terms of expected revenue to the best mechanism in H.

Coarsening the hypothesis space. We can do the coarsening of the space of virtual value functions in two ways: either group values v_i^t in buckets of size ϵ, rounding each value to the closest multiple of ϵ or group virtual values $\hat{\phi}_i(v_i^t)$ in buckets of size ϵ, allowing the virtual value function to take values only on multiples of ϵ. If we run ERM on any of these sub-spaces H^ϵ, then the number of possible virtual value functions that can appear in $\hat{H}_S^\epsilon$ (i.e. returned by ERM over H^ϵ applied to a subset T of S), is at most $O\left((1/\epsilon)^{n^2/\epsilon}\right)$, which does not grow exponentially with the sample size m. The first approach was taken in the paper of [3], while the second approach was taken in the paper of [7]. We will take the second approach and limit the space of values of the virtual value function and argue more formally it's sample complexity and it's error from the overall optimal mechanism.

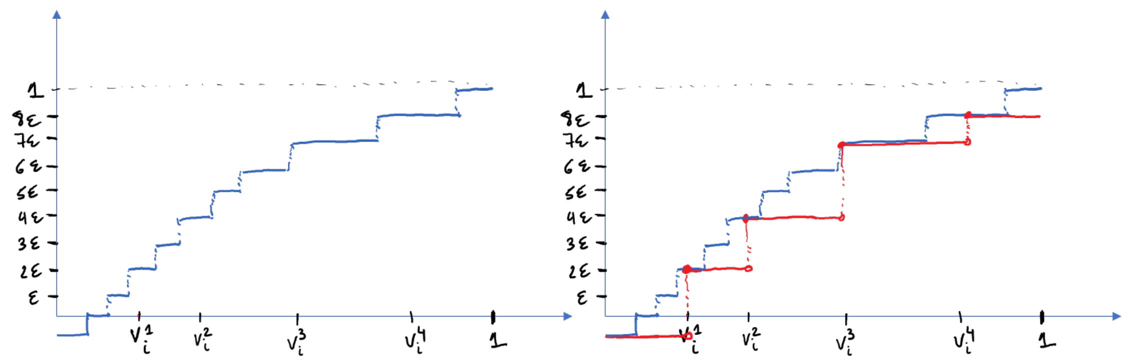

Figure 2: An example of a virtual value function in H^ϵ (blue line on left and right figure) and an example of a coarse step value function in H^ϵ_S, that only changes at points $v_i^t \in S$ (red line on right figure).

Coarse virtual value functions. We consider the following hypothesis space (see also Figure 2)

$$H^\epsilon = \{\text{Virtual value maximizers, with virtual value functions taking values in multiples of } \epsilon\}$$

We first argue about the sample complexity of H^ϵ and then we show that: $\sup_{h\in H^\epsilon} \mathrm{R}_D(h) \geq \sup_{h\in H} \mathrm{R}_D(h) - \epsilon$.

First observe, that by the reasoning that we did above, we have that the split-sample hypothesis space is a subset of the following set:

$$\hat{H}^\epsilon_S \subseteq \left\{ \begin{array}{l} \text{Virtual value maximizers, w. each virtual value function being} \\ \text{a monotone step function on points } v_i^t \in S \text{ taking values in multiples of } \epsilon \end{array} \right\}$$

The number of such functions for each player is $m^{1/\epsilon}$, since every such monotone function $\hat{\phi}_i$ can be uniquely defined by defining for each multiple of ϵ, $\kappa\epsilon$, the smallest value v_i^t for which $\hat{\phi}_i(v_i^t) \geq \kappa\epsilon$. Thus the number of possible such monotone step functions is: $m^{1/\epsilon}$. Therefore the set of all possible combinations of such functions, one for each player, is $m^{n/\epsilon}$. Thus $\hat{\tau}_{H^\epsilon}(m) \leq m^{n/\epsilon}$. Hence, by applying Theorem 1, if h^ϵ_S is the output of ERM over H^ϵ on set S, we get:

$$\mathbb{E}_S \left[\mathrm{R}_D(h^\epsilon_S)\right] \geq \sup_{h\in H_\epsilon} \mathrm{R}_D(h) - \sqrt{\frac{2n\log(2m)}{\epsilon \cdot m}} \tag{26}$$

Error of H^ϵ on the distribution. We now turn to comparing the optimal over H^ϵ on the true distribution D, with the optimal over H. We will show that for any product distribution D, we do not lose more than ϵ.

Lemma 2. *If $D = D_1 \times \ldots \times D_n$ and each D_i is a regular distribution, then:*

$$\sup_{h\in H^\epsilon} \mathrm{R}_D(h) \geq \sup_{h\in H} \mathrm{R}_D(h) - \epsilon \tag{27}$$

Proof. Consider, the optimal mechanism for a product of regular distributions $D = D_1 \times \ldots \times D_n$. By Myerson's theorem this is equal to the mechanism that maximizes the virtual welfare for virtual value functions:

$$\phi_i(v_i) = v_i - \frac{1 - F_i(v_i)}{f_i(v_i)} \tag{28}$$

where $F_i(\cdot)$ is the CDF and $f_i(v_i)$ is the pdf of distribution D_i. By regularity, the later function is monotone.

The optimal mechanism h_* for D, allocates to the player with the highest virtual value, as long as this virtual value is non-negative and to no player otherwise. Let $i_* = \arg\max_{i\in[n]} \phi_i(v_i)$, denote the random variable that indicates who is the winner in the optimal auction.

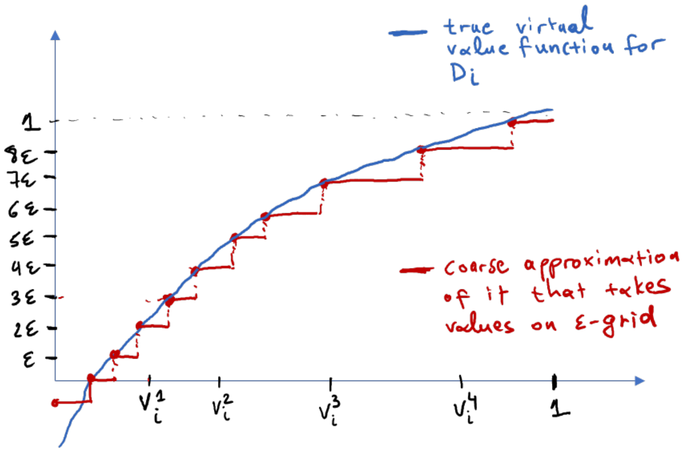

Figure 3: True virtual value function and coarse approximation of it, that achieves ϵ approximate expected revenue.

Let h^ϵ, be the mechanism defined by the following virtual value functions (see also Figure 3): for any number x, let $\lfloor x \rfloor_\epsilon$ be the closest multiple of ϵ below x, then:

$$\phi_i^\epsilon(v_i) = \begin{cases} \lfloor \phi_i(v_i) \rfloor_\epsilon & \text{if } \phi_i(v_i) \geq 0 \\ -\epsilon & \text{if } \phi_i(v_i) < 0 \end{cases} \tag{29}$$

Moreover, let $i_\epsilon = \arg\max_{i \in n} \phi_i^\epsilon(v_i)$, denote the random variable that indicates who is the winner in the auction h^ϵ.

From Myerson's characterization we know that the expected revenue of any mechanism, when valuations are independent, is equal to its expected virtual welfare. Thus:

$$\mathrm{R}_D(h^\epsilon) = \mathbb{E}_{\mathbf{v}\sim D}\left[[\phi_{i_\epsilon}(v_{i_\epsilon})]^+\right] \tag{30}$$

Since, ϕ_i^ϵ is a rounded down version of ϕ_i, we have:

$$\mathrm{R}_D(h^\epsilon) \geq \mathbb{E}_{\mathbf{v}\sim D}\left[[\phi_{i_\epsilon}^\epsilon(v_{i_\epsilon})]^+\right] \tag{31}$$

Since i_ϵ maximizes ϕ_i^ϵ, we have:

$$\mathrm{R}_D(h^\epsilon) \geq \mathbb{E}_{\mathbf{v}\sim D}\left[\max_i [\phi_i^\epsilon(v_i)]^+\right] \tag{32}$$

Since, $\phi_i^\epsilon(v_i)$ is also at least $\phi_i(v_i) - \epsilon$:

$$\mathrm{R}_D(h^\epsilon) \geq \mathbb{E}_{\mathbf{v}\sim D}\left[\max_i [\phi_i(v_i))]^+ - \epsilon\right] = \mathbb{E}_{\mathbf{v}\sim D}\left[[\phi_{i_*}(v_{i_*})]^+\right] - \epsilon = \mathrm{R}_D(h_*) - \epsilon \tag{33}$$

Hence we can conclude that:

$$\sup_{h \in H^\epsilon} \mathrm{R}_D(h) \geq \mathrm{R}_D(h^\epsilon) \geq \mathrm{R}_D(h_*) - \epsilon \tag{34}$$

□

Combining Equation (26) and Lemma 2, we get:

$$\mathbb{E}_S\left[\mathrm{R}_D(h_S^\epsilon)\right] \geq \sup_{h \in H} \mathrm{R}_D(h) - \sqrt{\frac{2n\log(2m)}{\epsilon \cdot m}} - \epsilon \tag{35}$$

Picking, $\epsilon = \left(\frac{2n\log(2m)}{m}\right)^{1/3}$, we get:

$$\mathbb{E}_S\left[\mathrm{R}_D(h_S)\right] \geq \sup_{h\in H} \mathrm{R}_D(h) - 2\left(\frac{2n\log(2m)}{m}\right)^{1/3} \tag{36}$$

Equivalently, the sample complexity is $m_H(\epsilon) = O\left(\frac{n\log(1/\epsilon)}{\epsilon^3}\right)$.

The latter rate was obtained by [7] via the use of the notion of the pseudo-dimension of a hypothesis class. The pseudo-dimension based proof is more involved, but offers better high probability bounds, in terms of dependence on δ. In particular, the pseudo dimension upper bounds the classical Rademacher complexity, and not only the split-sample version that we presented here. Hence, subsequently we can invoke the McDiarmid based proof we mentioned in the previous lecture to show that with probability $1-\delta$ over the sample S:

$$\mathrm{R}_D(h_S) \geq \sup_{h\in H} \mathrm{R}_D(h) - 2\left(\frac{2n\log(2m)}{m}\right)^{1/3} - O\left(\sqrt{\frac{\log(1/\delta)}{m}}\right) \tag{37}$$

which has a dependence of $\log(1/\delta)$ rather than $1/\delta$ which is implied by Theorem 1.

4 Historical Remarks

The seminal work of [9] gave a recipe for designing the optimal truthful auction when the distribution over bidder valuations is completely known to the auctioneer. Recent work, starting from [2], addresses the question of how to design optimal auctions when having access only to samples of values from the bidders. We refer the reader to [3] for an overview of the existing results in the literature. [2, 7, 8, 1] give bounds on the sample complexity of optimal auctions without computational efficiency, while recent work has also focused on getting computationally efficient learning bounds [3, 11, 4].

The results we presented here focus solely on sample complexity and not computational efficiency and thus is more related to [2, 7, 8, 1]. The latter work, uses tools from supervised learning, such as pseudo-dimension [10] (a variant of VC dimension for real-valued functions), compression bounds [6] and Rademacher complexity [10, 12] to bound the sample complexity of simple auction classes. We presented an analysis via a new measure of sample complexity, which is a strengthening the Rademacher complexity analysis.

References

[1] Maria-Florina F Balcan, Tuomas Sandholm, and Ellen Vitercik. Sample complexity of automated mechanism design. In *Advances in Neural Information Processing Systems*, pages 2083–2091, 2016.

[2] Richard Cole and Tim Roughgarden. The sample complexity of revenue maximization. In *46th*, pages 243–252. ACM, 2014.

[3] Nikhil R. Devanur, Zhiyi Huang, and Christos-Alexandros Psomas. The sample complexity of auctions with side information. In *Proceedings of the Forty-eighth Annual ACM Symposium on Theory of Computing*, STOC '16, pages 426–439, New York, NY, USA, 2016. ACM.

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

[5] Jason D. Hartline and Tim Roughgarden. Simple versus optimal mechanisms. In *Proceedings of the 10th ACM Conference on Electronic Commerce*, EC '09, pages 225–234, New York, NY, USA, 2009. ACM.

[6] Nick Littlestone. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine learning*, 2(4):285–318, 1988.

[7] Jamie Morgenstern and Tim Roughgarden. The pseudo-dimension of near-optimal auctions. In *Proceedings of the 28th International Conference on Neural Information Processing Systems*, NIPS'15, pages 136–144, Cambridge, MA, USA, 2015. MIT Press.

[8] Jamie Morgenstern and Tim Roughgarden. Learning simple auctions. In *COLT 2016*, 2016.

[9] Roger B Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.

[10] D. Pollard. *Convergence of Stochastic Processes*. Springer Series in Statistics. 2011.

[11] Tim Roughgarden and Okke Schrijvers. Ironing in the dark. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, EC '16, pages 1–18, New York, NY, USA, 2016. ACM.

[12] S. Shalev-Shwartz and S. Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press, 2014.