Lecture 15

Lecturer: Vasilis Syrgkanis *Scribe: Vasilis Syrgkanis*

1 Introduction

We look at the sample complexity of single-item, single-bidder optimal mechanisms. The bidder has a value function $v \in [0,1]$ drawn from a distribution D. We assume we are given a sample set $S = \{v_1, \ldots, v_m\}$, of m valuations, where each $v_t \sim D$.

We want to design a dominant strategy truthful mechanism that approximately maximizes expected revenue. Observe that if we knew the distribution D, then the optimal mechanism would be a posted price auction with a reserve price p, which maximizes the quantity $p \cdot (1 - F(p))$, where F is the CDF of distribution D.

For any posted price $p \in [0,1]$, let

$$\mathtt{r}(p, v) = p \cdot \mathbf{1}\{v \geq p\} \tag{1}$$

and

$$\mathrm{R}_D(p) = \mathbb{E}_{v\sim D}\left[\mathtt{r}(p, v)\right] = p \cdot (1 - F(p)) \tag{2}$$

be the expected revenue of a posted price p under the true distribution of values D.

Given a sample S of size m, we want to compute a reserve price p_S, such that:

$$\mathbb{E}_S\left[\mathrm{R}_D(p_S)\right] \geq \sup_{p\in[0,1]} \mathrm{R}_D(p) - \epsilon(m) \tag{3}$$

where $\epsilon(m) \to 0$ as $m \to \infty$. Equivalently, we want to have that for every ϵ, there exists $m_H(\epsilon)$, such that if $m \geq m_H(\epsilon)$:

$$\mathbb{E}_S\left[\mathrm{R}_D(p_S)\right] \geq \sup_{p\in[0,1]} \mathrm{R}_D(p) - \epsilon \tag{4}$$

The function $m_H(\epsilon)$ is the sample complexity of the problem.

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

$$\mathrm{R}_D(p_S) \geq \sup_{p\in[0,1]} \mathrm{R}_D(p) - \epsilon(m, \delta) \tag{5}$$

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

Observe, that this problem is an example of a PAC learning question, where the hypothesis space H is all posted price $p \in [0,1]$, the data space is the space of valuations $[0,1]$, the reward (rather than loss) function is the revenue $\mathtt{r}(p, v)$ and the distribution of data is D.

We will bound the sample complexity of this problem in four different ways, portraying different ways of using the PAC learning machinery.

2 Sample Complexity via Rademacher Complexity and ϵ-Covers

In the last lecture we show that the sample complexity is upper bounded by the Rademacher complexity of the problem:

$$\mathcal{R}(S, H) = \mathbb{E}_\sigma\left[\sup_{p\in[0,1]} \frac{2}{m}\sum_{t=1}^{m} \sigma_t \mathtt{r}(p, v_t)\right] \tag{6}$$

Then by the general PAC learning theorems we know that if p_S is the empirically optimal price, i.e.

$$p_S = \arg \sup_{p\in[0,1]} \mathrm{R}_S(p) \equiv \frac{1}{m}\sum_{t=1}^{m} \mathtt{r}(p, v_t) \tag{7}$$

then

$$\mathbb{E}_S\left[\mathrm{R}_D(p_S)\right] \geq \sup_{p\in[0,1]} \mathrm{R}_D(p) - \mathbb{E}_S\left[\mathcal{R}(S, H)\right] \tag{8}$$

and with probability $1-\delta$:

$$\mathrm{R}_D(p_S) \geq \sup_{p\in[0,1]} \mathrm{R}_D(p) - \mathbb{E}_S\left[\mathcal{R}(S, H)\right] - O\left(\sqrt{\frac{\log(1/\delta)}{m}}\right) \tag{9}$$

We first note the following: Let $H_\epsilon = \{0, \epsilon, 2\epsilon, \ldots, 1\} \cup S$ be an ϵ grid of $[0, 1]$, augmented with the actual values of the sample S. Then for any price $p \in [0, 1]$, there exists a price $p_\epsilon \in H_\epsilon$, such that for any $v_t \in S$:

$$|\mathtt{r}(p, v_t) - \mathtt{r}(p_\epsilon, v_t)| \leq \epsilon \tag{10}$$

We can construct such a v_t as follows: Assume that values v_t are ordered in increasing order and let $0 = v_0 \leq v_1 \leq \ldots \leq v_m \leq v_{m+1} = 1$. If $p \in [v_{t-1}, v_t)$, for some $t \in [m + 1]$, then consider two cases: if the closest multiple of ϵ below p is in the interval $[v_{t-1}, v_t)$, then let p_ϵ be that multiple, otherwise set $p_\epsilon = v_{t-1}$. By doing so, we have that for every $v_t \in S$, whether v_t gets allocated or not is the same under p and under p_ϵ, while the payment decreases by at most ϵ.

We then say that H_ϵ is an ϵ-cover of H. We then provide a general lemma on how to bound the Rademacher complexity via the existence of a cover:

Lemma 1 (Discretization Lemma)**.** *Let H any hypothesis and S a sample and suppose that H_ϵ is an ϵ-cover of S, i.e. for any $h \in H$, there exists $h_\epsilon \in H_\epsilon$ s.t.:*

$$\sup_{v\in S} |r(h, v) - r(h_\epsilon, v)| \leq \epsilon \tag{11}$$

Then:

$$\mathcal{R}(S, H) \leq \mathcal{R}(S, H_\epsilon) + 2\epsilon \tag{12}$$

Proof. For any h let $h_\epsilon \in H_\epsilon$ be the hypothesis that covers it (i.e. $\sup_{v\in S} |\mathtt{r}(h, v) - \mathtt{r}(h_\epsilon, v)| \leq \epsilon$). Then by the definition of the ϵ-cover, we have that:

$$\begin{aligned}
\mathcal{R}(S, H) =\ & \mathbb{E}_\sigma\left[\sup_{h\in H} \frac{2}{m}\sum_{t=1}^{m} \sigma_t \mathtt{r}(h, v_t)\right] \\
=\ & \mathbb{E}_\sigma\left[\sup_{h\in H}\left(\frac{2}{m}\sum_{t=1}^{m} \sigma_t \mathtt{r}(h_\epsilon, v_t) + \frac{2}{m}\sum_{t=1}^{m} \sigma_t\left(\mathtt{r}(h, v_t) - \mathtt{r}(h_\epsilon, v_t)\right)\right)\right] \\
\leq\ & \mathbb{E}_\sigma\left[\sup_{h\in H} \frac{2}{m}\sum_{t=1}^{m} \sigma_t \mathtt{r}(h_\epsilon, v_t) + \sup_{h\in H} \frac{2}{m}\sum_{t=1}^{m} \sigma_t\left(\mathtt{r}(h, v_t) - \mathtt{r}(h_\epsilon, v_t)\right)\right] \\
\leq\ & \mathbb{E}_\sigma\left[\sup_{h_\epsilon\in H_\epsilon} \frac{2}{m}\sum_{t=1}^{m} \sigma_t \mathtt{r}(h_\epsilon, v_t) + \sup_{h\in H} \frac{2}{m}\sum_{t=1}^{m} \sigma_t\left(\mathtt{r}(h, v_t) - \mathtt{r}(h_\epsilon, v_t)\right)\right] \\
=\ & \mathcal{R}(S, H_\epsilon) + \mathbb{E}_\sigma\left[\sup_{h\in H} \frac{2}{m}\sum_{t=1}^{m} \sigma_t\left(\mathtt{r}(h, v_t) - \mathtt{r}(h_\epsilon, v_t)\right)\right] \\
\leq\ & \mathcal{R}(S, H_\epsilon) + \mathbb{E}_\sigma\left[\sup_{h\in H} \frac{2}{m}\sum_{t=1}^{m} |\mathtt{r}(h, v_t) - \mathtt{r}(h_\epsilon, v_t)|\right] \\
\leq\ & \mathcal{R}(S, H_\epsilon) + 2\epsilon
\end{aligned}$$

□

Going back to the posted price problem, since H_ϵ is a cover for S of size $|H_\epsilon = m + \frac{1}{\epsilon}$, we can apply the discretization lemma and Massart's lemma for finite hypotheses spaces, to get:

$$\mathcal{R}(S,H) \leq \mathcal{R}(S,H_\epsilon) + 2\epsilon \leq \sqrt{\frac{2\log(|H_\epsilon|)}{m}} + 2\epsilon \leq \sqrt{\frac{2\log\left(m + \frac{1}{\epsilon}\right)}{m}} + 2\epsilon \tag{13}$$

By setting $\epsilon = 1/m$, we get: $\mathcal{R}(S,H) \leq O\left(\sqrt{\frac{\log(m)}{m}}\right)$. The latter implies that the sample complexity is $m_H(\epsilon) = O\left(\frac{\log(1/\epsilon)}{\epsilon^2}\right)$.

3 Sample Complexity via Rademacher Complexity Structure

Finally, we give another way of bounding the Rademacher complexity without the need for discretization and the discretization lemma. We remind that the Rademacher complexity is:

$$\mathcal{R}(S,H) = \mathbb{E}_\sigma\left[\sup_{p\in[0,1]} \frac{2}{m}\sum_{t=1}^m \sigma_t \mathbf{r}(p,v_t)\right] = \mathbb{E}_\sigma\left[\sup_{p\in[0,1]} \frac{2p}{m}\sum_{t=1}^m \sigma_t \mathbf{1}\{v_t \geq p\}\right] \tag{14}$$

Let $v_1 \leq \ldots \leq v_m$, be the values in the sample S. Observe that when p lies in the region between two values $[v_t, v_{t+1})$ the quantity:

$$\frac{2p}{m}\sum_{t=1}^m \sigma_t \mathbf{1}\{v_t \geq p\} \tag{15}$$

is linear in p. Thus dependent on the sign of the multiplier $\sum_{t=1}^m \sigma_t \mathbf{1}\{v_t \geq p\}$, the optimal p can only take either value $\inf_{\delta\geq 0} v_t + \delta$ (if the multiplier is negative) or v_{t+1} (if the multiplier is positive). Taking any $\delta > 0$, we have that:

$$\mathcal{R}(S,H) = \mathbb{E}_\sigma\left[\sup_{p\in[0,1]} \frac{2p}{m}\sum_{t=1}^m \sigma_t \mathbf{1}\{v_t \geq p\}\right] \leq \mathbb{E}_\sigma\left[\sup_{p\in\{v_1,\ldots,v_m\}\cup\{v_1+\delta,\ldots,v_m+\delta\}} \frac{2p}{m}\sum_{t=1}^m \sigma_t \mathbf{1}\{v_t \geq p\}\right] + \delta$$

By Massart's lemma we then get:

$$\mathcal{R}(S,H) \leq \sqrt{\frac{2\log(2m)}{m}} + \delta$$

Taking δ to 0, yields the same sample complexity as in the previous sections.

4 Sample Complexity via ERM on Discretized Space

The previous sections were arguing about the sample complexity via the ERM algorithm on the original class H of all prices $p \in [0,1]$. Here we will argue about sample complexity by changing the algorithm itself. We will look at running ERM on a discretized space (rather than arguing about Rademacher complexity via discretization).

Let $H_\epsilon = \{0, \epsilon, 2\epsilon, \ldots, 1\}$ be the ϵ-grid in $[0,1]$. Observe that for any price $p \in [0,1]$, there exist a price p_ϵ on the grid such that for any value v:

$$\mathtt{r}(p_\epsilon, v) \geq \mathtt{r}(p,v) - \epsilon \tag{16}$$

Simply round p down to the nearest multiple of ϵ. This can only allocate to more values, and the revenue to the values that p was allocating to, can only decrease by at most ϵ.

Suppose that we run ERM on this discrete price space:

$$p_S^\epsilon = \arg\sup_{p\in H_\epsilon} \mathrm{R}_S(p) \equiv \frac{1}{m}\sum_{t=1}^{m} \mathtt{r}(p, v_t) \tag{17}$$

H_ϵ is a finite hypothesis of size $1/\epsilon$. Thus by the sample complexity of finite hypothesis spaces we get:

$$\mathbb{E}_S\left[\mathrm{R}_D(p_S^\epsilon)\right] \geq \sup_{p\in H_\epsilon} \mathrm{R}_D(p) - \sqrt{\frac{\log(1/\epsilon)}{m}} \geq \sup_{p\in[0,1]} \mathrm{R}_D(p) - \sqrt{\frac{\log(1/\epsilon)}{m}} - \epsilon \tag{18}$$

Setting $\epsilon = 1/m$, yields the same sample complexity as we showed in the previous section.

5 Sample Complexity via Split-Sample Growth

In the next lecture we will also argue that the generalization error $\epsilon(m)$ is at most $O\left(\sqrt{\frac{\log(m)}{m}}\right)$, simply from the following fact: ERM on a sample S can only ever output a price that is equal to some value $v_t \in S$. Anything else is suboptimal. There are m such possible values. This will lead to the result.

Specifically, we will define the notion of a split-sample growth rate: for any set S of size m, how many different possible hypotheses can ERM output when run on any sub-sample T of S of size $m/2$. If that number is $\hat{\tau}_H(m)$, then we can upper bound the generalization error by:

$$\mathbb{E}_S\left[\mathrm{R}_D(p_S)\right] \geq \sup_{p\in[0,1]} \mathrm{R}_D(p) - \sqrt{\frac{2\log(\hat{\tau}_H(2m))}{m}} \tag{19}$$

Observe, that on any sub-sample of a sample S, ERM can only output a posted price that is equal to some value $v_t \in S$. Thus $\hat{\tau}_H(m) = m$, yielding the same sample complexity as before. Albeit when we are looking at high probability guarantees, this approach will lead to worst dependence on the confidence δ, than the previous approaches.

6 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, 6, 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, 6, 7, 8, 1]. The latter work, uses tools from supervised learning, such as pseudo-dimension [10, 7, 8] (a variant of VC dimension for real-valued functions), compression bounds [5, 8] and Rademacher complexity [10, 12, 6] 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] Nick Littlestone. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine learning*, 2(4):285–318, 1988.

[6] Mehryar Mohri and Andres Muñoz Medina. Learning theory and algorithms for revenue optimization in second-price auctions with reserve. In *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*, ICML'14, pages I–262–I–270. JMLR.org, 2014.

[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.