CSCI699: Topics in Learning and Game Theory

Lecture 12

Lecturer: Ilias Diakonikolas *Scribes: Cheng Cheng, Anastasia Voloshinov*

1 Review

Up-to-now, we have showed that learning a discrete distribution over $[n]$ requires $O(n)$ samples when total variation distance (d_{TV}) is used as the measurement metric. However, the problem arises when the distribution is continuous, since the above sample complexity no longer stays true. There are two directions to solve it:

1. We can use a weaker measurement metric, e.g. the Kolmogorov-Smirnov distance.
2. We can impose assumptions on the distribution. Then, we can use d_{TV}, but need to use kernels in this case.

1.1 Kernal Density Esitmation

Kernal Density Estimation provides a way for us to estimate the probability density function f of a random variable X using finite data samples. We convolute the discrete data with a continous kernal function $K(x)$ and the resulting continous function will give us a smooth estimation for f.

2 Settings

This lecture studies the paper "Optimal Nonparametric Estimation of First-Price Auctions" (Econometrica 2000), by Guerre, Perrigne, and Vuong. The paper shows a kernel-based estimator to learn the distribution of bidder's valuation using the actual bids under First Price Auction. The problem is set up as follows:

- Data from m First Price Auctions
- n bidders
- the valuation v_i of bidder $i = 1, ..., n$ are i.i.d samples from cumulative distribution function (CDF) F with probability density function (pdf) f

- First Price Auction
  - n bidders has valuations v_i for $i = 1, ..., n$ drawn i.i.d from distribution F, which is known to all bidders
  - Call that the first price auction is not truthfull. So, bidders give b_i for $i = 1, ..., n$ not truthfully, which means b_i's are different from v_i's
  - the seller can only observe b_i's, but wants to learn F and f

3 Formal Proof

Assumption 1. *Bidders play according to Bayesian-Nash Equilibrium, s.t.* $b_i = b(v_i)$

Assumption 2. *The CDF F is strictly increasing, continuous, and differentiable in the interval $[0, \bar{v}]$ (CDF: $F_x(u) = Pr(X \leq u)$). Hence, the pdf is always bigger than or equal to some positive constant.*

Lemma 3. *Under Assumption 1 and 2, the equilibrium of the First Price Auction is unique, symmetric, and*

$$b(v) = v - \frac{1}{F(v)^{n-1}} \int_0^v F(z)^{n-1} dz$$

for $0 < v < \bar{v}$, $b(0) = 0$, *and* $b(\bar{v}) = \bar{b} \leq \bar{v}$. *$n$ is the number of players.*

From Lemma 3 we can prove that b is strictly increasing, continuous and differentiable. And our following goal is to learn $F(v)$ by observing the distribution of the bids.

3.1 Identification

We want to show that we can solve the problem in principle. In other words, suppose that an infinite number of bids are provided. We want to show that in this case, we can reconstruct the distribution of the values.

We will do this by looking at the relation between bids and values by using the conditions of a symmetric BNE.

Let $b(v)$, where $v \sim F$, be the bidding distribution. Let G be the CDF of $b(v)$ and let g be the pdf.

Since a player is playing according to an equilibrium, we can connect his bid with his value by the best response condition. Recall that the best response condition says that if other people's actions our fixed, you cannot increase your utility by changing actions, and this must hold true for all players.

If a player submits a bid of b, what is the probability that he wins is $G(b)^{n-1}$. Since this is a first price auction, he only wins if everyone else bids lower than him. So, each of the $n-1$ players have a $G(b)$ probability of bidding less than him.

The expected utility of a player who submits a bid of b is $u(b,v) = (v-b)G(b)^{n-1}$. The maximum utility is thus attained at $b = b(v)$. So, the derivative of this function should be 0 at $b(v)$.

Since $b(v)$ maximizes the utility, we must have

$$\left.\frac{\delta u(b,v)}{\delta b}\right|_{b=b(v)} = 0$$

.

The derivative is

$$\frac{\delta u(b,v)}{\delta b} = -G(b)^{n-1} + (v-b)(n-1)G(b)^{n-2}g(b)$$

Then if we set it equal to 0, plug in $b(v)$ for b and rearrange, we get that

$$v - b(v) = \frac{G(b(v))}{(n-1)g(b(v))}$$

.

We want to reverse engineer the distribution of v from the distribution of b, so we want to have a function of b instead of a function of v. Since b is monotone and strictly increasing, it is therefore invertible, so we can do a change of variable.

Let $\xi(x) = b^{-1}(x)$, so then we get

$$v = \xi(b) = b + \frac{G(b)}{(n-1)g(b)}$$

where $b \sim G$.

Now, we have the value written as a function of b. Note that $\xi(b)$ is a function that is strictly increasing, continuous, and differentiable.

Claim 4. *We can approximate G and g by samples, since we have access to samples from the distribution, the bids. If we had G and g exactly, we could construct samples from the values, by drawing bids and applying the formula, which would give us samples from the values distribution. Then, we could get an infinite amount of samples from the values.*

If we know G and g, then given the bid b of a player, we can reverse engineer his value. Let $v = \xi(b)$. Therefore, we can identify F, the distribution of the values. Thus, we get

$$F(v) = Pr(V \leq v) = Pr(b(V) \leq b(v)) = G(b(v)) = G(\xi^{-1}(v))$$

where the second equality follows because b is strictly increasing and invertible, so the events are the same.

Thus, we have explicitly defined the distribution of the values.

4 Finite Sample Case

We have seen that we can find the distribution of the values if we have an infinite amount of samples of bids. Now, what if we have a finite number of samples?

In this case, we will take samples and estimate $\hat{G}$ and $\hat{g}$, so that $\hat{G} \approx G$ and $\hat{g} \approx g$.

Three Stage Algorithm:

1. Compute estimates $\hat{G}$ and $\hat{g}$ of the true bid distribution. Thus will be done by taking a finite number of samples of b.

2. Invert each bid to compute a value, using

$$\hat{v} = \hat{\xi}(b) = b + \frac{\hat{G}(b)}{(n-1)\hat{g}(b)}$$

   As a result, we will be able to find $\hat{F}$ and $\hat{f}$, which will be close to the true distribution F and f.

3. Learn the distribution of the pseudo-values ($\hat{v}$).

We will estimate $\hat{G}$ and $\hat{F}$ in the following way, which works by DKW.

$$\hat{G}(b) = \frac{1}{m} \sum_{t=1}^{m} \mathbb{1}_{\{b_t \leq b\}}$$

and

$$\hat{F}(v) = \frac{1}{m} \sum_{t=1}^{m} \mathbb{1}_{\{\hat{v}_t \leq v\}}$$

To estimate $\hat{g}$ and $\hat{f}$, we will use kernel density estimation. We get that

$$\hat{g}(b) = \frac{1}{m(h_g)} \sum_{t=1}^{m} k_g(\frac{b_t - b}{h_g})$$

and

$$\hat{f}(v) = \frac{1}{m(h_f)} \sum_{t=1}^{m} k_f(\frac{\hat{v}_t - v}{h_f})$$

Assumption 5. *f is λ-Lipschitz. This means that for all x, y, $|f(x) - f(y)| \leq \lambda |x - y|$. Intuitively, this means that the function doesn't change very rapidly, which we need to estimate f.*

Theorem 6. *Under assumptions 2 and 5, when k_f and k_g are uniform kernels, m is the number of samples, $h_g = \mathcal{O}(\frac{1}{m^{1/4}})$, $h_f = \mathcal{O}(\frac{1}{m^{1/8}})$, then for any interval $C_v(\epsilon) - [\epsilon, \hat{v} - \epsilon]$ if $m = \Omega(1)$, then with probability greater than or equal to $1 - delta$,*

$$sup_{v \in C_v(\epsilon)} |\hat{f}(v) - f(v)| \leq \mathcal{O}\Big(\Big(\frac{\log 1/\delta}{m}\Big)^{1/\delta}\Big) \leq \gamma$$

Thus, will small enough ϵ and large enough m, we can learn f with under total variation distance.

The analysis for this requires DKW and kernels, which we will do next class.

5 Kernel Density Estimation

Given samples $x_1, ..., x_m$ from unknown f, we estimate the density by

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\Big(\frac{x_i - x}{h}\Big)$$

where h is the bandwidth and $K : \mathbb{R} \to \mathbb{R}_+$ is the kernel function, and $\int_{\mathbb{R}} K(u) du = 1$. Examples:

1. Uniform
2. Epanechnikov, $K_1(u) = \frac{3}{4}(1 - u^2)\mathbb{1}_{\{|u| \leq 1\}}$
3. $K_{[\psi]}(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$

Definition 7. *The order of a kernel is the smallest non-zero moment. Examples 1 and 2 both have order 2.*