CSCI699: Topics in Learning and Game Theory

Lecture 13

Lecturer: Ilias *Scribes: Reem Alfayez, Sarah Al-Hussaini*

Last 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. We also had this formula:

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

$$b(v) = v - \frac{1}{F(v)^{n-1}} \int_0^{v'} F(z)^{(n-1)} dz \tag{2}$$

Where:

b=bid.

F=CDF.

f=PDF. $\hat{v}$ = maximum possible v.

b(v)= the bid when the value is v. We want to learn (F,f) on values, but we have (G,g) distribution on bids, by pretending that we know them exactly where G=CDF and g=PDF then we take a single sample from that distribution of bids then use formula 1 to generate a sample.

$$\hat{\xi}(b) = b + \frac{\hat{G}(b)}{(n-1)\hat{g}(b)} \tag{3}$$

We will explain what are kernels and why do we need them. Kernels were used to learn continuous distributions by smoothing out discrete approximations. A kernel has two important parameters:

- Kernel function.
- Bandwidth which is not automated and need to be changed if the problem changed.

Kernel density estimation

Given n samples:

$$x_1, x_2, x_3, ..., x_n$$

idd from unknown distribution with PDF f. The estimation:

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

$$K : R \rightarrow R \int_{-\infty}^{\infty} K(u) du = 1$$

where h is the bandwidth and K is the kernel function. Example of Kernels are:

- uniform Kernel.
- Epanechnikov Kernel.
- Gaussian Kernel.

Very important property of a kernel is its order. The order of a kernel is the smallest j where

$$M_j(k) = \int u^j k(u) du \neq 0$$

where M is a moment. The higher order of a kernel is the better the approximation, convergence, smoothing and learning become. **What can we do with kernels?** Assumptions:

- f is r-times differentiable.
- K has order $\geq$ r

$$\hat{f}(x) \overset{n \to \infty}{\rightarrow} f(x)$$

$$Mean\, Square\, Error\, MSE(\hat{f}(x)) = \mathbb{E}\left[(\hat{f}(x) - f(x))^2\right]$$

$$= Bias(\hat{f}(x))^2 + Variance[\hat{f}(x)]$$

$$\simeq (\frac{1}{r!} f^r(x) M_r(k).h^r) + \frac{f(x)R(k)}{nh} \int_{-\infty}^{\infty} k^2 du \quad [\because Bias = \mathbb{E}\left[(\hat{f}(x)\right] - f(x)]$$

The two sides are competing with each other. Our goal is to find a value h that makes both sides equal.

$$\mathbb{E}\left[\hat{f}(x)\right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\frac{1}{n} k \frac{x_i - x}{h}\right] = \int_{-\infty}^{\infty} k(u) f(x + hu) du$$

By Taylor's theorem

$$\mathbb{E}\left[\hat{f}(x)\right] = f(x) + \frac{1}{r!}f^r(x)M_r(k)h^r + o(h^r)$$

$$h^{2r} \simeq \frac{1}{nh}$$

$$h^{2r+1} = \frac{1}{n}$$

There are formal justification based on the central limit theorem:

$$d_k(\hat{f}(x) - f(x)) \overset{n\to\infty}{\to} 0$$

We are interested in a result that holds for a finite sample.

Finite sample approximation

Suppose that :

$$f \in \Delta([0,1]), \delta - Lipschitz : |f(x) - f(y)| \leqslant \delta\,|x-y|\,\forall x, y \in [0,1]$$

Use kernel density estimation of uniform kernel that has order 2. Look at TV distance between $f(x), \hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} k(\frac{x_i - x}{h})$
Where: n is the number of samples and h is the bandwidth.

$$2d_{TV}(\hat{f}, f) = \int_0^1 \left|\hat{f}(x) - f(x)\right| dx = \int_0^1 \left| f(x) - \frac{1}{nh}\sum_i k\frac{\hat{x}_i - x}{h}\right| dx$$

$$\int_0^1 \left| f(x) - \frac{1}{2nh}\sum |\{i : x_i \; \in [x-h, x+h]\}|\right| dx$$

$$= \int \left| f(x) - \frac{1}{2h}\frac{\hat{f}([x-h, x+h])}{\hat{f}(x+h) - \hat{f}(x-h)}\right| dx$$

$$[w.p \geq 1-\delta] \overset{DKW}{=} \int_0^1 \left| f(x) - \frac{1}{2h}F([x-h, x+h]) \pm O(\frac{\sqrt{\lg\frac{1}{\delta}}}{\sqrt{n}})\right| dx$$

$$\leqslant \int_0^1 \left| f(x) - \frac{1}{2h}F\left([x-h, x+h]\right)\right| dx + O(\frac{\sqrt{\lg\frac{1}{\delta}}}{2h\sqrt{n}})$$

$$F\left([x-h,x+h]\right)=\int_{x-h}^{x+h}f(t)dt$$
$$=\int_0^1\left|f(X)-\frac{1}{2h}\int_{x-h}^{x+h}f(t)dt\right|$$
$$=\int_0^1\left|\frac{1}{2h}\int_{x-h}^{x+h}f(X)-f(t)dt\right|dx$$
$$\leq\int_0^1\frac{1}{2h}\int_{x-h}^{x+h}\left|f(x)-f(t)\right)|\,dtdx$$
$$f(x)-f(t)\leq\delta(x-t)$$
$$\leq\int_0^1\frac{1}{2h}\int_{x-h}^{x+h}\delta\left|(x-t)\right|dtdx$$
$$\leq\delta h+O(\frac{\sqrt{\lg\frac{1}{\delta}}}{2h\sqrt{n}})$$

Choose h:

$$\delta h^2\simeq(\frac{\sqrt{\lg\frac{1}{\delta}}}{\sqrt{n}})\Rightarrow h=\frac{1}{\sqrt{\delta}}\left(\frac{\lg\frac{1}{\delta}}{n}\right)^{\frac{1}{4}}$$
$$Final\,error=\sqrt{\delta}\left(\frac{\lg\frac{1}{\delta}}{n}\right)^{\frac{1}{4}}$$

Que. We have just bounded d_{TV}. But what if we want to bound $\sup_x|\hat{f}(x)-f(x)|$?

Ans. We can bound this using the same calculation without integration. However, we will get a different bound with that approach.

Algorithm

We know,

$$b(v)=v-\frac{1}{F(v)^{n-1}}\int_0^{\bar{v}}F(z)^{n-1}dz\quad\text{is Nash Equilibrium}$$

$$\xi(b)=b+\frac{G(b)}{(n-1)g(b)}\quad\text{is value } v \text{ for bid } b$$

The algorithm has two stages:

1. Compute $\hat{G}, \hat{g}$

2. Invert each bid using:

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

Estimate $\hat{F}$, $\hat{f}$ using pseudo samples $\hat{v}_t$ s

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

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

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

$$\hat{f}(b) = \frac{1}{m}\sum_{t=1}^{m} \frac{1}{h_f} K_g(\frac{v_t - v}{h_f})$$

$$K_f = K_g = K_u(z) = \frac{1}{2} 1_{\{|z|<1\}}$$

A1. Assumption 1. CDF F is continuous, differentiable; the r.v. is in $[0, \bar{v}]$
A2. Assumption 2. f is λ-Lipschitz

Theorem

Under A1, A2 and $K_g = K_f = K_u, h_g = O\Big(\frac{1}{m^{\frac{1}{4}}}\Big), h_f = O\Big(\frac{1}{m^{\frac{1}{8}}}\Big)$
for any $C_v(\epsilon) \in [\epsilon, \bar{v} - \epsilon]$, w.p. $\geq 1 - \delta$,

$$\sup_{v \in C_v(\epsilon)} |\hat{f}(v) - f(v)| \leq O\Bigg(\bigg(\frac{log\frac{1}{\delta}}{m}\bigg)^{\frac{1}{8}}\Bigg)$$

-For this, what is m?
-Since the algorithm has two stages, we need to bound the error in two stages

First Stage

Lemma 2: According to DKW, w.p. $\geq 1 - \lambda$,

$$\sup |\hat{G}(b) - G(b)| \leq O\Big(\frac{\sqrt{log\frac{1}{\delta}}}{\sqrt{m}}\Big)$$

The next lemma is to bound $f(b), \hat{f}(b)$

Lemma 3: w.p. $\geq 1 - \delta$,

$$\sup |g(b) - \hat{g}(b)| \leq \lambda_g h_g + \frac{1}{h_g} O\Big(\frac{\sqrt{log\frac{1}{\delta}}}{\sqrt{m}}\Big)$$

How to prove? Under A1 and A2, g is λ-Lipschitz (can be proven with calculus) with a different constant λ_g. Then use the result from first part of the lecture to show this.

$$h_g = O\Big(\Big(\frac{1}{m}\Big)^{\frac{1}{4}}\Big)$$

$$Error = O\Big(\Big(\frac{log\frac{1}{\delta}}{m}\Big)^{\frac{1}{4}}\Big)$$

Note: Earlier, we defined F for $[0, \bar{v}]$ values. But later we took $[0, 1]$. That is not a problem. This is just requires larger number of samples that is dependent on $\bar{v}$ with a weird relation.

Inversion Error

Lemma 4: For any interior set of the bid dis. domain,

$$\sup |\hat{\xi}(b) - \xi(b)| \leq O\Big(\Big(\frac{log\frac{1}{\delta}}{m}\Big)^{\frac{1}{4}}\Big)$$

Proof:

$$
\begin{aligned}
|\hat{\xi}(b)-\xi(b)| &= |b+\frac{\hat{G}(b)}{(n-1)\hat{g}(b)}-b-\frac{G(b)}{(n-1)g(b)}| \\
&= \frac{1}{(n-1)\ g(b)\ \hat{g}(b)}|\hat{G}(b)g(b)-G(b)\hat{g}(b)| \\
&= \frac{1}{(n-1)\ g(b)\ \hat{g}(b)}|(\hat{G}(b)-G(b))\ g(b)+G(b)\ (g(b)-\hat{g}(b))| \\
&\leq \frac{1}{(n-1)\ g(b)\ \hat{g}(b)}\Big(|\hat{G}(b)-G(b)|\ g(b)+G(b)\ |g(b)-\hat{g}(b)|\Big)
\end{aligned} \tag{4}
$$

From Lemma 1 & 2, $|\hat{G}(b)-G(b)| \to 0$ and $|g(b)-\hat{g}(b)| \to 0$
Now if $g(b), \hat{g}(b)$ are close to zero, the total term can be large.
But according to A1, A2,
$g(b)$ is not close to 0
$\therefore \hat{g}(b)$ is also not close to 0 since $\hat{g}(b) \to g(b)$

$$\therefore |\hat{\xi}(b)-\xi(b)| \to 0$$

Errors due to Estimated Inverted Values

We will show that the estimated pdf ($\hat{f}$) from m samples is close to the ideal pdf (f), i.e., $\hat{f} \to f$

Suppose, if we had infinitely many samples, the estimated pdf would be $\tilde{f}$

Since $\hat{f} \to \tilde{f}$, we just need to prove $\tilde{f} \to f$

Let

$$\tilde{f}_h(v) = \frac{1}{m}\sum_{t=1}^{m}\frac{1}{h}K_u\Big(\frac{v-v_t}{h}\Big)$$

Note: for $\hat{f}_h(v)$, v_t would be replaced by $\hat{v}_t$.

Lemma 5: Under A2,

$$\sup |\tilde{f}_h(v) - f(v)| \leq \lambda_f h + \frac{1}{h} O\Big(\frac{log\frac{1}{\delta}}{\sqrt{m}}\Big)$$

Lemma 6: For $h_f = O\Big(\frac{1}{m^{\frac{1}{8}}}\Big), h_g = O\Big(\frac{1}{m^{\frac{1}{4}}}\Big)$, w.p. $\geq 1-\delta$,

$$\sup |\tilde{f}_h(v) - f(v)| \leq O\Bigg(\Big(\frac{log\frac{1}{\delta}}{m}\Big)^{\frac{1}{8}}\Bigg)$$

We need both upper bounds and lower bounds

Upper bounds: Let,

$$\Delta = O\Bigg(\Big(\frac{log\frac{1}{\delta}}{m}\Big)^{\frac{1}{4}}\Bigg)$$

Now,

$$\begin{aligned}
\hat{f}(v) &= \frac{1}{m}\sum_{t=1}^{m}\frac{1}{h_f}1_{\{|v-\hat{v}_t|\leq h_f\}} \\
&\leq \frac{1}{m}\sum_{t=1}^{m}\frac{1}{h_f}1_{\{|v-v_t|\leq h_f+\Delta\}} \ [\because \hat{v}_t \text{ and } v_t \text{ are } \Delta-\text{close}, Lemma-4] \\
&\leq \frac{h_f+\Delta}{h_f}\frac{1}{m}\sum_{f=1}^{m}\frac{1}{h_f+\Delta}1_{\{|v-v_t|\leq h_f+Delta\}} \\
&= \frac{h_f+\Delta}{h_f}\tilde{f}_{h_f+\Delta}(v) \ [\text{from the equation of } \tilde{f}_h(v) \text{ (before Lemma 5)}] \\
&\leq \frac{h_f+\Delta}{h_f}(f(v)+\lambda_f(h_f+\Delta)+\frac{1}{h_f+\Delta}O\Big(\frac{log\frac{1}{\delta}}{\sqrt{m}}\Big) \ [\text{Lemma 5}] \\
&= \Big(1+\frac{\Delta}{h_f}\Big)(f(v)+\lambda_f(h_f+\Delta)+\frac{1}{h_f+\Delta}O\Big(\frac{log\frac{1}{\delta}}{\sqrt{m}}\Big))
\end{aligned} \tag{5}$$

By picking $h_f = O\Big(\frac{1}{m^{\frac{1}{8}}}\Big)$,
$\hat{f}(v)$ and $f(v)$ are close to each other point-wise.