Lecture 10 Notes

Agenda: more nonparametric estimation
- recap from end of last time
- convergence rate
- localizing functionals
- implementation

Recap:

graphon function: \( \omega \)
node locations \( U_i \sim \text{Uniform}[0,1] \)
Ball \( B(x, y, h) := \{ (a, b) \in [0,1]^2 : |x - a| < h \text{ and } |y - b| < h \} \)

Method for approximating \( \omega \)
1. Pretend for now we observe all \( U \)'s
2. Take all \((u_i, u_j) \in B(x, y, h)\)
3. Average \( A_{ij} \) for those dyads
4. Return as \( \hat{\omega}(x, y) \)

recall:

\[
P(A_{ij} = 1 \mid u_i = x, u_j = y) = \omega(x, y) = \mathbb{E}[A_{ij} \mid u_i = x, u_j = y] \quad (1)
\]

If \( w \) is smooth, then \( \omega(x + \epsilon, y + \nu) \) is close to \( (x, y) \).
So averaging \( A_{ij} \) from points near \((x, y)\) should approximate \( \omega(x, y) \)

\[
\langle A; B(x, y, z) \rangle = \frac{1}{n^2|B|} \sum_{(i,j) \in B(x,y,h)} A_{ij} \quad (2)
\]
\[
\mathbb{E}[A; B] = \frac{1}{n^2|B|} \sum_{(i,j) \in B} \mathbb{E}[A_{ij}] = \frac{1}{n^2|B|} \sum_{(i,j) \in B} \omega(u_i, u_j) \quad (3)
\]

Assume that \( \omega \) is a smooth function - specifically that

\[
\frac{1}{|B|} \int_{B(x,y,h)} |\omega(u, v) - \omega(x, y)| \, du \, dv \leq kh^\gamma \quad (4)
\]

for some \( k, \gamma \in \mathbb{R}^+ \)

\[
\frac{1}{n^2|B|} \sum_{(i,j) \in B} \omega(u_i, u_j) \overset{\text{almost surely}}{\to} \frac{1}{|B|} \int_{B(x,y,h)} \omega(u, v) \, du \, dv = \mathbb{E}(x, y, h) \quad (5)
\]
As a consequence of this assumption
\[ |\varphi(x, y, h) - \omega(x, y)| \leq O(h^\gamma) \] (6)

What about variance?
\[ \text{Var}[\langle A; B \rangle] = \text{Var}[\langle \omega; B \rangle + \langle \epsilon; B \rangle] \] (7)
where:
\[ \epsilon_{ij} = A_{ij} - \omega(u_i, u_j) \] (8)

\( \epsilon_{ij} \) is dependent on \( \omega(u_i, u_j) \) but also has conditional mean 0
\( A_{ij} \in \{0, 1\} \)
\( \omega(u_i, u_j) \in (0, 1) \)

Therefore \( \epsilon_{ij} \) is uncorrelated with \( \omega(u_i, u_j) \)

Recall: for random variables \( X \) and \( Y \)... 
\[ \mathbb{C}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \]
\[ = \mathbb{E}[XE[Y|X]] - \mathbb{E}[X]\mathbb{E}[Y|X]] \] (9)

Also:
\[ \text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y] \] (10)

So:
\[ \text{Var}[\langle A; B \rangle] = \text{Var}[\langle \omega; B \rangle] + \text{Var}[\langle \epsilon; B \rangle] \] (11)

Now we take a closer look at (11).
Since \( \epsilon_{ij} \) is uncorrelated with \( \epsilon_{kl} \), it follows that
\[ \text{Var}[\langle \epsilon_{ij}; B \rangle] = O \left( \frac{1}{n^2 h^2} \right) \] (12)

And (since \( \text{Var}[\text{Bern}] = p(1 - p) \)) we have
\[ \text{Var}[\epsilon_{ij}] \leq \frac{1}{4} \] (13)

\( \text{Var}[\langle \omega; B \rangle] \) is much more annoying
\[ \text{Var}[\frac{1}{n^2 |B|} \sum_{(i,j) \in B} \omega(u_i, u_j)] \]
\( \omega(u_i, u_j) \) is correlated with \( \omega(u_i, u_k) \)

There are two approaches for us to work with
**Approach 1:** bound \( \text{Cov}[\omega(u_i, u_j), \omega(u_i, u_k)] \) using smoothness and finite size \( h \)
**Approach 2:** steal result on generalized U-statistics
U-Statistics

Given independent random variables $X_1, X_2, \ldots, X_n$ and a symmetric function $\psi$ of two args, a U-Statistic is

$$\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi(X_i, X_j) = U_\psi$$  \hspace{1cm} (14)

All terms in here are dependent on one another (for terms that share the same $X_i$)

general results on variance of $U_\psi$ based on:

$$\text{Var}[\psi(X_1, X_2)]$$  \hspace{1cm} (15)

$$\text{Var}[\mathbb{E}[\psi(X_1, X_2)|X_1]]$$  \hspace{1cm} (16)

(15) Variance of individual summands
(16) Covariance between summands

Generalized U-statistic: Given:

$$X_1, \ldots, X_n \overset{i.i.d.}{\sim} Y_1, \ldots, Y_n \overset{i.i.d.}{\sim} U$$

$$U_\psi = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \psi(X_i, Y_j)$$  \hspace{1cm} (17)

(note that Ys must have different distribution than Xs)

Generalized results on $\text{Var}[U_\psi]$ in terms of:

$\text{Var}[\psi(X_1, X_2)] \quad Y \sim \text{vertical coordinates}$

$\text{Var}[\mathbb{E}[\psi(X,Y)|X]] \quad X \sim \text{U-coordinates with horizontal limit of box}$

$\text{Var}[\mathbb{E}[\psi(X,Y)|Y]] \quad \psi \sim \omega$

Use smoothness of $\omega$ function

$$\text{Var}[\langle \omega; B \rangle] = O\left(\frac{1}{nh}\right)$$  \hspace{1cm} (18)

$$\text{Var}[\langle \epsilon; B \rangle] = \text{Var}[\langle \epsilon; B \rangle] + \text{Var}[\langle \omega; B \rangle]$$

$$= O \left( \frac{1}{n^2 h^2} \right) + O \left( \frac{1}{nh} \right)$$  \hspace{1cm} (19)

$$= O \left( \frac{1}{nh} \right)$$

$$\mathbb{E}[\langle A; B \rangle] = \omega(x, y, h)$$

$$= \omega(x, y) + O(h^\gamma)$$  \hspace{1cm} (20)
So we have the *Mean-Squared Error (MSE)* as $[\text{bias}]^2 + [\text{variance}]$

$$\mathbb{E}[(\langle A; B \rangle - \omega(x, y))^2] = O(h^{2\gamma}) + O\left(\frac{1}{nh}\right)$$

(21)

Pick $h$ that minimizes this.

note: there is a trade-off between bias & variance

$$O \left( h^{2\gamma-1} \right) + O \left( \frac{-1}{n^2 h^2} \right) = O$$

$$h^{2\gamma-1} = \frac{1}{nh^2}$$

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

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

(22)

And thus:

$$MSE = O \left( n^{-\frac{1}{2\gamma+1}} \right) + O \left( \frac{1}{n \cdot n^{-\frac{1}{2\gamma+1}}} \right)$$

$$= O \left( n^{-\frac{2\gamma}{2\gamma+1}} \right)$$

(23)

Recall that for parametric estimates we only get $O(n^{-1})$

**Topology Fact:**

$\dim(X) = \dim(Y) \iff \exists \phi : X \to Y \text{ s.t. } \phi \text{ is continuous and } \exists \phi^{-1} \text{ s.t. } \phi^{-1} \text{ is also continuous.}$

**Graphon Fact:** Any CID model is equivalent to a $\omega$-function $[0, 1]^2 \to [0, 1]$ i.e. $P(x_i, x_j) = \omega(\phi(u_i), \phi(u_j))$ for some $\phi : X \to [0, 1]$

Suppose $X$ is $\mathbb{R}^2$ or $\mathbb{R}^3$ as in Continuous Latent Space Models (CLSM). There cannot exist a homeomorphism between $\mathbb{R}^2$ and $[0, 1]$.

$\therefore \phi$ in $P(x_i, x_j) = \omega(\phi(u_i), \phi(u_j))$ must not be smooth.

$\therefore$ if $P$ is smooth, $\phi$ is not smooth, $\omega$ must not be smooth.

**Localizing Functionals and Implementation**

These topics will be covered in Lecture 11.