Notes Class December 8 2016

Cristobal De La Maza 36-781 - Fall 2016

December 21, 2016

1 Recapitulate from last class

A graph is generated from w, such that:

$w : [0,1] \times [0,1] \rightarrow [0,1]$

by $u_i \overset{iid}{\sim} U(0,1)$ with $i \in 1..n$

and $\zeta_i \overset{iid}{\sim} U(0,1)$ with $i \in 1..n$

Then,

$A_{ij} = 1\{\zeta_{ij} \leq w(u_i, u_j)\}$ and $P(A_{ij} = 1|u) = w(u_i, u_j)$

Also let's define a box $B(x, y, h)$ with all points (x,y) $\pm h$ on both axes. Then:

$w(\hat{x}, y) = \langle A; B(x, y, h) \rangle = \frac{1}{n^2|B|} \sum_{i,j \in (u_i,u_j) \in B} A_{ij}$
Where,

$w(\hat{x}, y) \rightarrow w(x, y)$ if w is integrally smooth in h and $h_n \rightarrow 0$ slowly enough.

$E[(w(\hat{x}, y) - w(x, y))^2] = O(h^2) + O(\frac{1}{nh})$ so best $MSE = O(n^{\frac{-2\gamma}{(2\gamma+1)}})$

When,

$\frac{1}{|B|} \int_{|B|} |w(u, v) - w(x, y)| dudv \leq kh^\gamma$

This assumes we know u_i, but we don't. We can fix this assuming the existence of a good proxy for u_i as a function of the data.

a) A graphon function $\phi_u \rightarrow [0,1]$ is discriminating for w when:

$$w(u, .) \neq w(v, .) \implies \phi_u(w) \neq \phi_v(w) \tag{1}$$

b) Given a sequence of graphs G_n, all generated from w, ordered by inclusion, ϕ is projectively consistent when:

$$\lim_{x \to \infty} \phi(w_{G_n}, \frac{i}{n}) \to \phi(w, u_i) \tag{2}$$

c) ϕ is localizing when it is discriminating and uniformly projectively consistent. Example:

Suppose w is a stochastic block models, and all blocks have different expected degrees. Block probabilities $\rho_1, ..., \rho_k$ with $\rho_i \geq 0$ and $\sum_{i=1}^{k} \rho_i = 1$. Block affinity matrix is $1 \leq b_{k \times k} \geq 0$, we require that $\sum_{j=1}^{k} \rho_j b_{ij} \neq \sum_{j=1}^{k} \rho_j b_{ik}$ and $\forall j \neq k$. Then, a discriminating functional is:

$$\phi(w, u) = \int_0^1 w(u, v) dv \tag{3}$$

It is also projectively consistent by LLN:

$$\phi(G_n, \frac{i}{n}) = degree(i)/n \to \int_0^1 w(u, v) dv \tag{4}$$

Assumptions:

For a localizing functional $\phi(w, u)$ exists a w that is integrally smooth for some $\gamma > 0$. Then using $\hat{u_i} = \phi(w_{G_n}, \frac{i}{n})$ instead of u_i still gives consistency of the smoothing estimator $\hat{w}$.

This works because $\hat{u_i} \to u_i$ (as ϕ is uniformly projectively consistent and if $w(u_i, .) \neq w(u_j, .)$ then $\hat{u_i}$ and $\hat{u_j}$ are eventually separate. Using $\hat{u_i}$ instead u_i introduces two sorts of error into $\hat{w}$. Node pairs that should not be in the box and node pairs outside the box that should be included.

2 Exchangeability revisited

2.1 Summary for the course

A distribution over graphs is exchangeable when it is invariant under permutation of nodes, i.e., if and only if all isomorphic graphs are equally probable.

Theorem 1 (Aldous, Hoover) *An infinite exchangeable distribution is either a CID model or a mixture of CID models.*

Theorem 2 (A,H, Kallenberg) *Any CID model can be written as:*
$w : [0, 1] \times [0, 1] \to [0, 1]$

$u_i \overset{iid}{\sim} U(0,1)$ *with* $i \in 1..n$
$\zeta_i \overset{iid}{\sim} U(0,1)$ *with* $i \in 1..n$
Then,
$A_{ij} = 1\{\zeta_{ij} \leq w(u_i, u_j)\}$ *and* $P(A_{ij} = 1|u) = w(u_i, u_j)$

Theorem 3 (A,H, Lovaz et al) *If* G_n *are a sequence of w random graphs the* $G_n \xrightarrow{a.s.} w$ *in the sense of motif density convergence.*

This implies that inference is possible and is used for all sorts of elimination and testing.

2.2 Dense graphs (sequences) are weird

A sequence of graphs G_n, $n = 1, ..., \infty$ is dense when $|E(G_n)| = \Theta(|V_{G_n}|^2)$. A graph sequence is sparse when $|E(G_n)| = o(|V_{G_n}|^2)$.Strictly speaking, no graph is dense or sparse only a sequence of graphs. In a dense graph sequence, infinitely many nodes must have degree $\rightarrow$ inf:

$$|E(G_n)| = \frac{1}{2} \sum_{i \in V_{G_n}} degree(i) \tag{5}$$

$$= \Theta(|V_{G_n}|^2) \tag{6}$$

$$\frac{|E(G_n)|}{|V(G_n)|} = \frac{1}{2|V_{G_n}|} \sum_{i \in V_{G_n}} degree(i) \tag{7}$$

$$= \Theta(|V_{G_n}|) \tag{8}$$

But, $degree(i) \leq |V_{G_n}| - 1$. So in a dense graph sequence the average degree is $\Theta(|V_{G_n}|)$. Thus, w-random graph sequences are dense (or empty).

Define $\bar{d(u)} = \int_0^1 w(u,v)dv$. This is the average value of w on a slice v or the average probability that a node at u connects to any other node.

$$P(Aij = 1|u_i = u, u_j = v) = w(u, v) \tag{9}$$

$$= \int_0^1 P(A_{ij} = 1|u_i = u, u_j = v)\rho_{uv}(v)dv \tag{10}$$

$$= \int_0^1 w(u, v)dv \tag{11}$$

$$= \bar{d(u)} \tag{12}$$

Define:

$D_i = \frac{1}{n-1} \sum_{i \neq j} A_{ij}$.
Claim:

$D_i \xrightarrow{a.s.} d(\bar{u})$

Where D_i is a mean of independent random variables, with mean $d(\bar{u})$. Alternatively, use bounded difference inequality. Conditional on u_i, A_{ij} are independent, changing one A_{ij} changes D_i by a fraction $\leq \frac{1}{n-1}$.

So,

$P(|D_i - E[D_i]| > \epsilon) \leq 2exp(-2\epsilon^2(n-1))$

This is exponentially small on n, hence summable, or Borel-Cantelli. Hence,

$|D_i - E[D_i]| > \epsilon$ happens only finitely often with probability one and

$E[D_i] = d(\bar{u_i})$.

Since $\forall i$, $P(D_i \to d(\bar{u_i})) = 1$ because we have only countable many nodes,

$\implies P(\forall i, D_i \to d(\bar{u_i})) = 1$

But $D_i = \frac{degree(i)}{n-1}$, either $d(\bar{u_i}) = 0$ for all u or $P(\forall i, degree(i)) = \Theta(n)) = 1$.

Hence,

$P(\forall i, degree(i)) = \Theta(n) or d(\bar{u_i}) = 0) = 1$

Thus,

$$D_i = d(\bar{u}) + o(1) \tag{13}$$

$$\frac{degree(i)}{n-1} = d(\bar{u}) + o(1) \tag{14}$$

$$degree(i) = (n-1)d(\bar{u}) + o(n) = \Theta(n) \tag{15}$$

$$\forall \bar{d} > 0 \tag{16}$$

This implies that if $d(\bar{u})$ is not equal to zero for almost all u, then $degree(i) = \Theta(n)$ for infinitely many i and so $|E(G_n)| = \Theta(|V(G_n)|^2)$, unless $d(\bar{u}) = 0$ for almost all u. This will happen if and only if $w(u,v) = 0$ for almost all (u,v). Hence,$d(\bar{u}) = \int_0^1 w(u,v)dv \iff w(u,v) = 0$ for almost all v. So, unless $w(u,v) = 0$ for almost all (u,v), the w-random graph sequence is dense with probability 1 and dense graph sequences are weird.

3 Alternatives

3.1 Chayes, Borgs, Cohn and Zhao

Rescale the graphon function.

Ingredients: w function sequence $\rho_n \to 0$ where:

$P(A_{ij}^{(n)} = 1/u_i = u, uj = v) = \rho_n w(u, v)$

Problems: tends got lead to disconnected graphs in sparse case.

Solution: let w: $[0, 1] \times [0, 1] \to [0, \infty]$

Restriction: w $\in L_p$ i.e. $\int_{[0,1]^2} w^p(u, v) dudv < \infty$ for some p and set $P(A_{ij}^{(n)} = 1/u_i = u, uj = v) = min(\rho_n w(u, v), 1)$.

If w has a little "island of infinity", nodes there are always connected and nodes nearby are very likely to be connected.

Convergence: say that $G_n \xrightarrow{sparse} w$ when $w_{G_n} \rho_n \xrightarrow{dense} w/||w||$. $\rho_n = 1$, recovers usual w-random graphs.

Problems with this approach is that we need to know ρ_n which could not possibly be learned from the graph.

A bigger problem: the model is not projective. Remember that a model is projective when:

$P_m(X_m = a) = \sum_b P_{1:n}(X_m = a, X_{m+1:n} = b), \forall m < n$

Projectivity is automatic for IID models.

Projectivity is automatic for time series or spatial statistics:

$P_{1:n}(X_m, X_{m+1:n}) = P_m(X_m) P_{m+1:n/m}(X_{m+1:n}/X_n)$

Graphons are also projective. Therefore graphons first m nodes independent of edges to nodes $m+1 : n$ given $u_1, ..., u_m$. If the model is projective $P(A_{ij} = 1)$ does not change with sample size. If not, not. In L_p graphons, $P(A_{ij} = 1) \to 0$ as n grows.

3.2 Geometric random graphs

Points x_i lay down following a poisson process $exp(\lambda)$ on a (potentially) infinite manifold (say R^2 or S^2). $P(A_{ij} = 1 | x_i = x, x_j = y) = w(d(x, y))$. If the manifold is infinite, like R^2 this is different provided follows a poisson process.

Suppose $w = 0$ if $d(x, y) > k$. How does degree(i) grows as n grows? Number of possible neighbors is $O_p(\lambda k)$. So even if we observe at some n, degrees are say 10 or 100 not n.