Lecture 7

§1 U-statistics 的应用: Network data

1. Network data 的相关定义

对于 network data (V, E), $N = |V|$, 有以下相关定义:

① adjacency matrix: $D = (D_{ij})_{|V| \times |V|}$

② homogeneous network: $1(D_{ij} = 1)$ 与位置 (i,j) 无关

③ network density: $\rho_N := P(D_{ij} = 1)$

④ network density 的一个 estimator: $\hat{\rho}_N := \binom{N}{2}^{-1} \sum_{i<j} D_{ij}$

⑤ expectation of average degree: $\lambda_N := (N-1)\rho_N$

⑥ expectation of average degree 的一个 estimator: $\hat{\lambda}_N := (N-1)\hat{\rho}_N$

⑦ dense: 若 $\lambda_N = \Omega(N)$ $(\rho_N = \Omega(1))$, 则 network 为 dense

⑧ sparse: 若 $\lambda_N = o(N)$ $(\rho_N = o(1))$, 则 network 为 sparse $(\lambda_N \to \lambda \text{ as } N \to \infty,\ 0 < \lambda < \infty)$

⑨ probability of certain structures:

$$P(\bullet\!\!-\!\!\bullet) := E[D_{ij}] \quad (= \rho_N)$$

$$P(\wedge) := E[D_{ij} D_{ik}]$$

⑩ standardized probability of certain structures:

$$\tilde{P}(\bullet\!\!-\!\!\bullet) := P(\bullet\!\!-\!\!\bullet) / \rho_N = 1$$

$$\tilde{P}(\wedge) := \tilde{P}(\wedge) / \rho_N^2$$

e.g. 对于 trading network (dense), 若需要判断某 node 是否必不可少(缺失后是否会产生 risk), 则可以移除该 node, 判断 network 是否变为 sparse.

2. Nyakatoke risk-sharing network

① Nyakatoke risk-sharing network 的特征

(1) 每个 node 有自己的 "feature variable":

$\{A_i, 1 \le i \le n\}$ 为 i.i.d. random variable

(2) Nodes 间建立联系的可能性由关于 features 的某个 kernel function 决定:

$h_N(A_i, A_j)$ 为某个 kernel function

(3) Nodes 间是否建立联系服从 Bernoulli distribution:

$$D_{ij} \mid A_i, A_j \overset{ind}{\sim} Ber(h_N(A_i, A_j))$$

② 对 ρ_N 的 inference

(Step 1: 对 $\hat{\rho}_N$ 的拆解)

$\hat{\rho}_N := \binom{N}{2}^{-1} \sum_{i<j} D_{ij}$ 可被写作 $\hat{\rho}_N = U_N + V_N$, 其中

$$U_N = \binom{N}{2}^{-1} \sum_{i<j} h_N(A_i, A_j) \quad \text{(是一个 U-statistic)}$$

$$V_N = \binom{N}{2}^{-1} \sum_{i<j} \{D_{ij} - h_N(A_i, A_j)\}$$

(Step 2: 证明 $V_N \to 0$)

对于 V_N，可证明其为 sum of non-correlated terms：

$$\begin{aligned}
&Cov(D_{ij}-h_N(A_i,A_j),\ D_{ik}-h_N(A_i,A_k)) \\
&= E[[D_{ij}-h_N(A_i,A_j)]\cdot[D_{ik}-h_N(A_i,A_k)]] - \underbrace{E[D_{ij}-h_N(A_i,A_j)]\cdot E[D_{ik}-h_N(A_i,A_k)]}_{=0} \\
&= E[\underbrace{E[[D_{ij}-h_N(A_i,A_j)]\cdot[D_{ik}-h_N(A_i,A_k)]\,|\,(A_i,A_j,A_k)]}_{=0}] \\
&= 0
\end{aligned}$$

因此由 Chebyshev's inequality：

$$\begin{aligned}
P(|V_n-\underbrace{E[V_n]}_{=0}|>\varepsilon) &\le \frac{1}{\varepsilon^2}Var(V_n) \\
&= \frac{1}{\varepsilon^2}\cdot\underbrace{\left(\frac{2}{n(n-1)}\right)^2}_{O(n^{-4})}\underbrace{\sum_{i<j}Var(D_{ij}-h_N(A_i,A_j))}_{O(n^2)} \\
&\to 0
\end{aligned}$$

$$\Rightarrow V_n \to 0$$

注：① 这是一个较 intuitive 的结论，因为 $D_{ij}|A_i,A_j \sim Ber(p)$ 而 $h_N(A_i,A_j)=p$

② $V_n \to 0$ 并不代表可以直接抛弃 V_n 项，V_n 仍有可能影响 $Var(\hat{p}_N)$

(Step 3: 拆分 U_N 项)

显然，可利用 Hoeffding theorem 求出 U_N 的 asymptotic distribution，但是会面对以下问题：

(1) δ_1 可能 $=0$

(2) 没有考虑 V_n 与 U_n 的 covariance / V_n 对 $Var(\hat{p}_N)$ 的影响

为了解决这些问题，我们利用 Hayek projection 对 U_n 进行分解：

$$U_N = p_N + \underbrace{U_{1N}}_{\text{projection}} + \underbrace{U_{2N}}_{\text{projection error}}$$

其中，

$$U_{1N} = \frac{2}{N}\sum_{i=1}^{N}\{h_{1N}(A_i)-p_N\}, \quad h_{1N}(A_i)=E[h_N(A_i,A_j)|A_i] \quad \text{(见 Lecture 5/§3/4)}$$

$$U_{2N} = U_N - p - U_{1N} = \frac{2}{N(N-1)}\sum_{i<j}\{h_N(A_i,A_j)-h_{1N}(A_i)-h_{1N}(A_j)+p_N\}$$

注：易证 U_{2N} 为 sum of uncorrelated terms，因此可以利用 Chebyshev inequality 证明 $U_{2N} \to 0$

(Step 4: 分析 $\hat{p}_N$ 的 variance 与 asymptotic distribution)

可以证明：

$$\Omega_{1N} := Var\,U_{1N} = p_N^2\cdot\{\tilde{Q}(\wedge)-\tilde{P}(\text{—})\cdot\tilde{P}(\text{—})\} = O\left(\frac{p_N^2}{N}\right)$$

$$\Omega_{2N} := Var\,U_{2N} = \binom{N}{2}^{-1}[O(p_N^2)-2\cdot O(p_N^2)] = O\left(\frac{p_N^2}{N^2}\right)$$

$$\Omega_{3N} := Var\,V_N = \binom{N}{2}^{-1}O(p_N) = O\left(\frac{p_N}{N^2}\right)$$

同时可以证明：

U_{1N}, U_{2N}, V_N 为 uncorrelated

因此，$\hat{p}_N = p_N + U_{1N} + U_{2N} + V_N$ 的 variance 为：

$$\underbrace{O\left(\frac{p_N^2}{N}\right)}_{U_{1N}} + \underbrace{O\left(\frac{p_N^2}{N^2}\right)}_{U_{2N}} + \underbrace{O\left(\frac{p_N}{N^2}\right)}_{V_N}$$

① 若 ρ_N 为 constant ($\Omega(1)$, dense network)，则 $\hat{\rho}_N$ 的 variance 为集

$$\xi_1^2 = O\left(\frac{1}{N}\right) \quad (U_{1N} \text{ 的 variance 为 dominated term})$$

$$\Rightarrow \sqrt{N}\cdot(\hat{\rho}_N - \rho_N) \longrightarrow N(0, N\xi_1^2)$$

② 若 $(N-1)\rho_N \to$ constant ($O(1)$, sparse network)，则 $\hat{\rho}_N$ 的 variance 为集

$$\xi_2^2 = O\left(\frac{1}{N^3}\right) \quad (U_{1N} \,\&\, V_N \text{ 的 variance 为 dominated term})$$

$$\Rightarrow \sqrt{N^3}\cdot(\hat{\rho}_N - \rho_N) \longrightarrow N(0, N^3\xi_1^2)$$

3. Dyadic regression

在 Dyadic regression 中，

① 研究 directional graph

② 每个 node 的 feature 不再是 i.i.d. 而是 exchangeable： $[X_{\sigma_x(i)}] \overset{d}{=} [X_i]$

③ edges 的 weight Y_{ij} 是 exchangeable： $[Y_{\sigma_y(i)\,\sigma_y(j)}] \overset{d}{=} [Y_{ij}]$

④ 对 edges 的 weight 进行 modeling，underlying model 可能为：

$$Y_{ij}^* = h(\underbrace{\alpha}_{\text{const.}}, X_i, X_j, D_{ij}, \underbrace{U_i, U_j}_{\text{node 的 degree}})$$

考虑以下简单 Poisson model：

$$\begin{cases} Y_{ij} \mid X_i, X_j \sim \text{Poisson}(e^{W_{ij}}) \\ W_{ij} = X_i\beta_1 + X_j\beta_2 \end{cases}$$

由于 Y_{ij} 间存在 covariance，无法直接写出 likelihood，因此考虑使用 composite likelihood (见 STA3020 Lecture 8)：

$$Cl := \sum_{i\neq j}\{(X_i\beta_1 + X_j\beta_2)\cdot y_{ij} - \exp(X_i\beta_1 + X_j\beta_2)\}$$

$$\Rightarrow \hat{\beta} = \underset{\beta=(\beta_1,\beta_2)}{\arg\max}\; Cl$$

可以得到以下有用的结论：

① $\sqrt{\cdots}\,(Cl - ECl) \xrightarrow{d} N(0, \cdots)$

② Cl concave，进而若 $\beta_0 := \underset{\beta}{\arg\max}\; ECl$，则有

(1) $\hat{\beta} \longrightarrow \beta_0$

(2) $\sqrt{\cdots}\,(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \cdots)$