Lecture 2 — September 17th, 2014

Prof. Daniel M. Roy *Scribe: Jinlong Fu*

1 Overview

This lecture borrows heavily from Kolaczyk's book, in particular, section 6.5, and Hoff's course notes at UW, in particular, lectures 9–13. (See the end of the course syllabus for links).

2 Predecessors to ERGMs

RCE models: Row Column Effect

$$Y = (Y_{i,j}), \{Y_{i,j}\} \text{ independent}$$

$$\mathbb{P}\{Y_{i,j} = 1\} = \frac{e^{\mu + a_i + b_j}}{1 + e^{\mu + a_i + b_j}}$$

where a_i: sender effect; b_i receiver effect

NO explicit model of reciprocity

$$S(Y) = \sum_{i<j} y_{i,j} y_{j,i}$$

larger than expected under **RCE** in most data

3 P_1 - model

Introduced by Holland & Leinhardt (1981)

$$\text{dyads } \{(y_{i,j}, y_{j,i})\} \text{ are independent}$$

$$\begin{aligned}
\mathcal{P}\{Y_{i,j} = 1 | Y_{i,j} = 0\} &= \frac{e^{u_{i,j}}}{1 + e^{u_{i,j}}} \\
\mathcal{P}\{Y_{i,j} = 1 | Y_{i,j} = 1\} &= \frac{e^{u_{i,j} + \gamma}}{1 + e^{u_{i,j} + \gamma}} \\
\mathcal{P}\{Y_{i,j} = y_{i,j}, Y_{j,i} = y_{j,i}\} &\propto \\
&\exp\{\mu_{i,j} y_{i,j} + \mu_{j,i} y_{j,i} + \gamma y_{i,j} y_{j,i}\}
\end{aligned}$$

when $\mu_{i,j} = \mu + a_i + b_j$

4 Exponential Random Graph Models

4.1 Exponential Families

A family $(P_\theta : \theta \in \mathbb{R}^P)$ of distributions on a space $\mathcal{S}$ is an exponential family, when, for all $\theta \in \mathbb{R}^P$,

$$\frac{d\mathbb{P}_\theta}{dh}(z) = \exp\{\theta^T g(z) - \Psi(\theta)\}$$

- $h(\cdot)$ is a measure on $\mathcal{S}$
- $g : \mathcal{S} \to \mathbb{R}^P$
- Ψ satisfies $\int \exp\{\theta^T g(z)\} e^{-\Psi(\theta)} h(dz) = 1$
  i.e., $\Psi(\theta) = \log \int \exp\{\theta^T g(z)\} h(dz)$

4.2 $\mathcal{S}$ is discrete

$$\begin{aligned} h &= \# \text{ “counting measure”} \\ \#(A) &= \text{cardinality of } A \\ \frac{d\mathcal{P}_\theta}{d\#} & \text{ is a p.m.f.} \end{aligned}$$

4.3 $\mathcal{S} = \mathbb{R}^n$

- $h = \lambda =$ Lebesgue measure
- $\Rightarrow \dfrac{d\mathcal{P}}{d\lambda}$ is a p.d.f.

4.4 Consider a random graph $G = (V, E)$

Let $Y_{ij} = \begin{cases} 1 \text{ edge}(i,j) \in E \\ 0 \text{ otherwise} \end{cases}$

$$\frac{d\mathcal{P}_\theta}{d\#}(y) = \mathcal{P}_\theta\{Y = y\} = \frac{1}{K(\theta)} \exp\{\sum_H \theta_H g_H(y)\}$$

- H is a configuration, i.e., $H \subset V \times V$
- $g_H(y) = \prod_{(i,j)\in H} y_{i,j} = \begin{cases} 1, \text{ H “appears” in E} \\ 0, \text{ o/w} \end{cases}$
- informally, $\theta_H > 0 \Rightarrow \{y_{i,j} : (i,j) \in H\}$ are dependent
- $K = K(\theta)$ is a normalization constant,
  $K(\theta) = \sum_y \exp\{\sum_H \theta_H g_H(y)\}$

4.5 Example

Erdos-Renyi Random Grpah
Presence/absence of an edge independent of all others. This implies that $\theta_H = 0$ for all H with more than 2 vertices.

$$\mathcal{P}_\theta\{Y = y\} = \frac{1}{K(\theta)} \exp\{\sum_{i,j} \theta_{i,j} y_{i,j}\}$$

i.e., each edge (i, j) appears independently,

$$p_{i,j} = \frac{e^{\theta_{i,j}}}{1 + e^{\theta_{i,j}}}$$

Another common assumption is that

$$\mathcal{P}_\theta\{Y = y\} = \mathcal{P}_\theta\{Y = \Pi y\}$$

for all permutation matrices Π. (This is exchangeability.) This implies that,

$$\mathcal{P}_\theta\{Y_{i,j} = 1\} = \mathcal{P}_\theta\{Y_{1,2} = 1\}$$
$$\forall i, j \; \theta_{i,j} = \theta$$

4.6 Markov Random Graphs

Frank & Strauss (1986, JASA)
Def: An array $Y = (Y_{i,j})$ is a Markov Graph when for every set of distinct nodes $\{i, j, h, k\}$,

$$Y_{i,j} \perp Y_{h,k} \mid Y \setminus \{Y_{i,j}, Y_{h,k}\}$$

Thm: If Y is a Markov Graph
and $(Y_{i,j}) \stackrel{d}{=} (Y_{\sigma(i)\sigma(j)})$ for all permutations σ. Then,

$$\mathcal{P}\{Y = y\} \propto \exp\{\theta L(Y) + \sum_{k=2}^{n-1} \sigma_k \mathcal{S}_k(y) + \tau T(y)\}$$

- $L(Y) \equiv \#$ edges
- $\mathcal{S}_k(y) \equiv \#$ k-stars
- $T(y) \equiv \#$ triangles

4.7 Fitting ERGMs

$$\mathcal{P}_\theta\{Y = y\} \propto \exp\{\theta^T g(y)\}$$
$$\text{with } \theta \in \mathbb{R}^P, g : \Phi_n \to \mathbb{R}^p$$
$$\text{MLE } \hat{\theta} \equiv \arg\max \; l(\theta)$$
$$\text{where } l(\theta) = \theta^T g(y) - \Psi(\theta)$$
$$\text{with } \Psi(\theta) = \sum_{y'} e^{\theta^T g(y')}$$

Because $\mathcal{P}_\theta$ is an exponential family, hence,

$$\mathbb{E}_\theta(g(Y)) = \frac{\partial \Psi(\theta)}{\partial \theta}$$

4.8 Stochastic approximation of log-likelihood $l(\theta)$

$$\begin{aligned}
\arg\max_\theta \; l(\theta) &= \arg\max_\theta [l(\theta) - l(\theta_0)] \\
&= \arg\max_\theta \; [(\theta - \theta_0)^T g(y) - (\Psi_\theta - \Psi_{\theta_0})]
\end{aligned}$$

$$\begin{aligned}
&\exp\{\Psi(\theta) - \Psi(\theta_0)\} \\
&\quad = \sum_y \exp\{(\theta - \theta_0)^T g(y)\} \times \left(\frac{\exp\{\theta_0^T g(y)\}}{\exp\{\Psi(\theta_0)\}} \right) \\
&\quad = \mathbb{E}_{\theta_0}[\exp\{(\theta - \theta_0)^T g(Y)\}]
\end{aligned}$$

We can approximate this expectation by drawing samples $\tilde{Y_1}, \ldots, \tilde{Y_n}$ from $\mathbb{P}_{\theta_0}$, then the above equation can be approximate as,

$$\frac{1}{m} \sum_{i=1}^{m} \exp\{(\theta - \theta_0)^T g(\tilde{Y_i}\}$$

4.9 How to optimize $l(\theta)$

R-package: ergm

- Generate $\tilde{Y_1}, \ldots, \tilde{Y_n}$
- Use approximation $\theta \to \exp\{\Psi(\theta) - \Psi(\theta_0)\} \approx \frac{1}{m} \sum_{i=1}^{m} \exp\{(\theta - \theta_0)^T g(\tilde{Y_i})\}$
- Move along derivative to update θ_0
- Repeat until convergence