Exponential Family and Random Graph Models (ERGMS)

Shengming Luo

3 October 2016

1 Overview

- Definitions
  - Graph modeling
  - Examples: Erdös-Renyi, p_1, 2-star, triangle
- Properties
  - Edge prediction
  - Moments
- Estimation
  - MLE equation
  - Stochastic approximation
  - MCMCMLE

1.1 Definitions

Definition 1.1. *Recall: For a graph with n nodes, we denote its adjacency matrix as* $A \in \mathbb{R}^{n,n}$. $A_{ij} = 1$ *if* $(i,j) \in E$, $A_{ij} = 0$ *if* $(i,j) \notin E$.

Definition 1.2. *A* graph model *is a probability distribution over the space of graphs.*

2 Model a random graph

We can start by selecting a subset of features to model. Let's start with the important features of data, hoping only some aspects of the data matter for the probability distribution.

2.1 Sufficient statistics

Postulate statistics: $T_1, T_2, \cdots, T_d$ are sufficient with respect to a statistical model and its associated unknown parameter if "no other statistic that can be calculated from the same sample provides any additional information as to the value of the parameter".

Lemma 2.1. *(Neyman's factorization criterion)T is sufficient for* Θ *if and only if nonnegative functions g and h can be found such that:*

$$P_\Theta(x) = h(x) g(\Theta, T(x)).$$

In other words, the data only interacts with parameter Θ via T.

As a consequence, the likelihood ratios:

$$P_\Theta(x)/P_{\Theta_0}(x) = \frac{h(x)g(\Theta, T(x))}{h(x)g(\Theta_0, T(x))} = \frac{g(\Theta, T(x))}{g(\Theta_0, T(x))}$$

Hence, the likelihood ratios only depends on sufficient statistics.

2.2 Exponential family

One way to get $T(\cdot)$ to be sufficient is an exponential family, i.e. $P_\Theta(x) \propto exp\{\sum_{i=1}^{d} T_i(x)\theta_i\} = e^{T(x)\Theta}$.

In fact, $P_\Theta(x) = \frac{e^{T(x)\Theta}}{\sum_x e^{T(x)\Theta}} = P_\Theta(x) = \frac{e^{T(x)\Theta}}{Z(\Theta)} = e^{T(x)\cdot\Theta - \psi(\Theta)}$, where Z and ψ are notations for convenience.

T is sufficient here, by construction. On the other hand, if T is to be sufficient and support does not change with Θ, $P_\Theta(x)$ must be exponential family.

Remark. Counter − example: *$X \sim Unif(0, \Theta)$, $X_{(n)}$ is sufficient but uniform distribution does not belong to exponential family, because the support change with Θ.*

Theorem 1 (Fisher-Pitman-Koopman-Darmois)**.** *Let $T = (T_1, T_2, \ldots, T_d)$ be a finite set of sufficient statistics for a model $p_\Theta(x)$ with support that does not depend on Θ. Then, $p_\Theta(x)$ must either be an exponential family distribution, or a uniform distribution.*

2.3 Exponential Random Graph Model(ERGM)

Definition 2.1. Exponential-family Random Graph Models *(ERGMs) are exponential families over graphs. In other words, the sufficient statistics are functions of the graph/adjacency matrix.*

In this setting, sufficient statistics(for graph) are functions of the graph, hence also functions of the adjacency matrix A(or a).

A recipe for creating an ERGM is therefore:

1. Pick d (distinct) functions of the graph; they might be chosen through appeals to theory, experience, guesswork, tradition, referee pressure, trial and error, etc.
2. Then, calculate these statistics, and forget the original graph for all within-model work: simulating, testing, estimation, etc.

Examples :

1. Random graph/ Erdos-Renyi: we only have one parameter. $\Theta = \{\text{Prob of an edge}\}$. $T(a) = \sum_{i,j} a_{ij}.\}$ To see that, notice

$$\begin{aligned} P_\Theta(a) &= \Pi_{i=1}^n \Pi_{j=1}^n \Theta^{a_{ij}} (1-\Theta)^{1-a_{ij}} \\ &= exp\{\sum_{i,j} a_{ij} log(\Theta) + (1-a_{ij}) log(1-\Theta)\} \\ &= exp\{\sum_{i,j} a_{ij} log(\frac{\Theta}{1-\Theta}) + n^2 log(1-\Theta)\} \end{aligned}$$

2. Block Model: parameters $\Theta = P$(or B), the affinity matrix. Sufficient statistics: n_{rs}.
3. p_1 model: directed graphs only. Sufficient statistics: out degrees of each node(denoted as a_i), in degrees of each node(denoted as b_i), number of reciprocal edges, denoted as r. If we do not include r, this is instead called a configuration model.

4. **Not ERGM**: Stochastic block model with latent blocks. It's a mixture of expo-families but not an expo-family.

5. Graph motif counts: edges + 2-stars or edges+triangles. We'd have

$$P_\Theta(a) \propto exp\{\Theta_1 e(a) + \Theta_2 t(a)\},$$

where e(a)= number of edges in graph, t(a)= number of triangles in graph.

Remark : No reason (except practical) not to include motif.

1. can include node attributes. e.g. # edges between nodes of same type("homophily").

2. can combine with node degrees.

3. can include global functions, like diameter (very rare, however).

3 Properties

3.1 Edge or link prediction

Having seen most of the graph, we'd like to guess whether $(i, j) \in E$ or $(i, j) \notin E$.

Notations: a_{+ij}= adjacency matrix with $(i, j) \in E$; a_{-ij}= adjacency matrix with $(i, j) \notin E$.

Therefore, the likelihood could be written as

$$P_\Theta(a_{+ij}) = e^{T(a_{+ij})\Theta}/Z(\Theta),$$

$$P_\Theta(a_{-ij}) = e^{T(a_{-ij})\Theta)}/Z(\Theta).$$

Hence, the likelihood ratio is:

$$P_\Theta(a_{+ij})/P_\Theta(a_{-ij}) = e^{T(a_{+ij})\Theta - T(a_{-ij})\Theta} = e^{\Delta_{ij}\Theta}.$$

So, $log(P_\Theta(a_{+ij})/P_\Theta(a_{-ij})) = \Delta_{ij}\Theta$. (logistic regression.)

This lends an easy interpretation to the parameters: "For any given configuration of the graph, if I toggle an edge and it leads to an increase in the statistics, it is more likely to see that configuration."

However, we cannot apply any causal interpretation, since there is no reason to think that any given edge (i, j) was generated after all of the other edges in the graph.

3.2 Moments of sufficient statistics

Recall,

$$Z(\Theta) = \sum_x exp\{\sum_{i=1}^{d} T_i(x)\theta_i\}.$$

We have

$$\begin{aligned}\frac{\partial Z(\Theta)}{\partial\theta_i} &= \sum_x \frac{\partial exp\{\sum_{i=1}^d T_i(x)\theta_i\}}{\partial\theta_i}\\ &= \sum_x exp\{\sum_{j\neq i}^d T_j(x)\Theta_j\}\frac{\partial exp\{T_i(x)\theta_i\}}{\partial\theta_i}\\ &= \sum_x exp\{\sum_{j\neq i}^d T_j(x)\Theta_j\}\cdot T_i(x)exp(T_i(x)\theta_i)\\ &= \sum_x T_i(x)exp\{\sum_{i=1}^d T_i(x)\theta_i\}\\ &= \sum_x T_i(x)P_\Theta(x)Z(\Theta) = Z(\Theta)\mathbb{E}_\Theta(T).\end{aligned}$$

Hence, $\mathbb{E}_\Theta(T_i) = \frac{1}{Z(\Theta)}\frac{\partial Z(\Theta)}{\partial\theta_i}$. We can rewrite it as

$$\begin{aligned}\mathbb{E}_\Theta[T_i] &= \frac{1}{Z(\Theta)}\frac{\partial}{\partial\theta_i}Z(\Theta)\\ &= \frac{\partial}{\partial\theta_i}\log Z(\Theta).\end{aligned}$$

Remark. *We can get higher moments by taking higher-order derivatives.*

4 Estimation

4.1 MLE

The likelihood of an exponential family is

$$L(\Theta) = p_\Theta(x) = e^{T(x)\cdot\Theta}/Z(\Theta).$$

We find the MLE by maximizing $L(\Theta)$.

By takin derivative, we have

$$\begin{aligned}&\frac{\partial}{\partial\theta_i}\frac{e^{T(x)\cdot\Theta}}{Z(\Theta)}\bigg|_{\Theta=\hat\Theta} = 0\\ \iff\ &\frac{Z(\Theta)\frac{\partial}{\partial\theta_i}[e^{T(x)\cdot\Theta}] - \frac{\partial}{\partial\theta_i}[Z(\Theta)]e^{T(x)\cdot\Theta}}{(Z(\Theta))^2}\bigg|_{\Theta=\hat\Theta} = 0\\ \iff\ &Z(\hat\Theta)T_i(x)e^{T(x)\cdot\hat\Theta} - e^{T(x)\cdot\hat\Theta}\mathbb{E}_{\hat\Theta}[T_i]Z(\hat\Theta) = 0\\ \iff\ &T_i(x) = \mathbb{E}_{\hat\Theta}[T_i].\end{aligned}$$

Note that all of this applies to general exponential families, not just ERGMs.

So to find the MLE of Θ, we "just" need to solve for $\hat\Theta$ in the equation $T_i(x) = \mathbb{E}_{\hat\Theta}[T_i]$, where

$$\mathbb{E}_\Theta[T_i] = \frac{\partial}{\partial\theta_i}\log Z(\Theta) = \frac{\partial}{\partial\theta_i}\log\{\sum_x e^{\Theta\cdot T}\}.$$

For undirected graphs, "x" is a full graph, so there are $2^{\binom{n}{2}}$ terms in this sum. This sum is too large to simply brute-force it.

What can we do to get around this? In some special cases, including the edge-triangle ERGM, the block model, and the Erdös-Renyi random graph model, we can get a closed-form solution for the MLE.

4.2 Stochastic approximation

When this fails, we can try simulating:

- Start with an initial graph configuration $\mathbf{a}^{(0)}$.
- Pick an edge (i, j) at random.
- Flip that edge with probability
$$\frac{p_\Theta(\mathbf{a}^{(0)}_{+ij})}{p_\Theta(\mathbf{a}^{(0)}_{-ij})},$$
which does not involve $Z(\Theta)$.

This is a Gibbs sampling procedure for finding the correct equilibrium distribution. In particular, since $Z(\Theta)$ is not involved, we circumvent the issue of computing its value.

Now, to solve $T(x) = \mathbb{E}_{\hat{\Theta}}[T]$ for $\hat{\Theta}$,

1. Start with a guestimate $\hat{\Theta}^{(0)}$.
2. Use simulation to get many graphs from $\hat{\Theta}^{(0)}$.
3. Approximate $\mathbb{E}_{\hat{\Theta}}[T]$ by sample averages.
4. Adjust $\hat{\Theta}^{(0)}$ to $\hat{\Theta}^{(1)}$ to bring $\mathbb{E}_{\hat{\Theta}}[T]$ closer to $T(x)$.

4.3 MCMCMLE

By the definition of the MLE,

$$\hat{\Theta} = argmax_\Theta L(\Theta) = argmax_\Theta \frac{L(\Theta)}{L(\Theta_0)}$$

where Θ_0 is some fixed initial guess for Θ. We can rewrite the likelihood ratio as

$$\begin{aligned}\frac{L(\Theta)}{L(\Theta_0)} &= \frac{e^{T(x)\cdot\Theta}/Z(\Theta)}{e^{T(x)\cdot\Theta_0}/Z(\Theta_0)} \\ &= \exp\left\{T(x)\cdot(\Theta - \Theta_0)\right\} / \frac{Z(\Theta)}{Z(\Theta_0)},\end{aligned}$$

and we can rewrite the denominator as

$$\begin{aligned}\frac{Z(\Theta)}{Z(\Theta_0)} &= \sum_{x'} \frac{e^{T(x')\cdot\Theta}}{Z(\Theta_0)} \\ &= \sum_{x'} \frac{e^{T(x')\cdot(\Theta-\Theta_0+\Theta_0)}}{Z(\Theta_0)} \\ &= \sum_{x'} e^{T(x')\cdot(\Theta-\Theta_0)} \frac{e^{T(x')\cdot\Theta_0}}{Z(\Theta_0)} \\ &= \mathbb{E}_{\Theta_0}\left[e^{T(x')\cdot(\Theta-\Theta_0)}\right].\end{aligned}$$

We can estimate this by simulating with Θ_0 *only*. That is, we never need to use any of the updates of Θ.

However, if $\Theta - \Theta_0$ is large, this estimate will oscillate. We would either need a ton of samples, or to update Θ_0 to a better value at some point. But in principle, we do not need anything but our initial guess Θ_0 to perform this step.