Exponential Random graph models

Sumit Mukherjee

sumitm@stanford.edu

January 25, 2011

Introduction

- Exponential random graph models are a family of probability distributions on graphs. Depending on the application, we may consider simple, loopy, multiple-edged, weighted or directed graphs.

- **Definition**: Let $\mathcal{G}_n$ be the set of all graphs on n vertices. Consider the following model

$$P_\theta(G = g) = exp\{\sum_{i=1}^{k} \theta_i T_i(g) - c(\theta)\}$$

where $\theta_i, i = 1, 2 \cdots, k$ are real valued parameters, and $T_i, i = 1, 2 \cdots, k$ are real valued statistic defined on $\mathcal{G}_n$.

- ERGM are mostly used for modeling "snap-shots" of relational networks in sociology. A few examples of possible sufficient statistics for such models are
  - the degrees of the vertices (representing how popular that vertex is in the network),

  - the number of edges (representing the total number of friendships in the network),

  - the number of triangles, or other sub-graph counts (representing the number of groups of 3 friends, and so on),

  - the number of connected components in the graph (representing the number of sub-communities in the network),

    and so on. In general, any interpretable statistic can be used to model ERGM.

An example[Holland, Leinhardt]

- Consider a network on n vertices, with directed edges but no multiple edges or loops. An example can be the case where we allow for unidirectional friendship, i.e. it is possible that A is a friend of B, even though B does not reciprocate. :(

- We model the scenario by $\{0, 1\}$ random variables $X_{ij}, 1 \leq i \neq j \leq n$, where $X_{ij} = 1$ if there is an edge from $i \rightarrow j$, or equivalently, if i considers j to be a friend. Let $X_{ii} = 0$ for convenience.

- Let $D_{ij} = (X_{ij}, X_{ji}), i < j$. Assume D_{ij} are mutually independent, i.e. the relation between (i, j) is not affected by the relation between any other pair.
- For a given pair (i, j), let

$$
\begin{align}
P(D_{ij} = (1, 1)) &= m_{ij} \tag{1} \\
P(D_{ij} = (1, 0)) &= a_{ij} \tag{2} \\
P(D_{ij} = (0, 0)) &= n_{ij} \tag{3}
\end{align}
$$

with $m_{ij} + a_{ij} + a_{ji} + n_{ij} = 1$.

- This gives the model

$$
P(X = x) \propto \exp\{\sum_{i<j} \rho_{ij} x_{ij} x_{ji} + \sum_{i<j} \theta_{ij} x_{ij}\}
$$

with $\rho_{ij} = \log \frac{m_{ij} n_{ij}}{a_{ij} a_{ji}}, \theta ij = \log \frac{a_{ij}}{n_{ij}}$.

- ρ_{ij} is a log-odds ratio which is the relative increase of odds of $X_{ij} = 1$ given $X_{ji} = 1$. Thus ρ_{ij} models reciprocity.

- θ_{ij} is a log odds which is $X_{ij} = 1$ given $X_{ji} = 0$. θ_{ij} thus models the probability for asymmetric dyads.

- This model has $O(n^2)$ parameters and so may over-fit. To simplify the model, let $\rho_{ij} = \rho, \theta_{ij} = \theta + \alpha_i + \beta_j$, with the constraints $\sum \alpha_i = \sum \beta_j = 0$. Then the model becomes

$$P(X = x) \propto \exp\{\rho m + \theta x_{++} + \sum_i \alpha_i x_{i+} + \sum_j \beta_j x_{.+j}\}$$

- The sufficient statistics for modeling the social behavior are
  - $M = \sum\limits_{i<j} X_{ij} X_{ji}$, which is the number of reciprocated relations.
  - $X_{i+} = \sum\limits_{j} X_{ij}$, which is the out-degree (or number of persons whom i considers a friend.
  - $X_{+j} = \sum\limits_{i} X_{ij}$, which is the in-degree (or number of persons who consider j to be a friend)
  - $X_{++} = \sum\limits_{i<j} X_{ij}$, which is the total number of friendships in the model (and is a function of X_{i+}/X_{+j}).

- Suppose we want to test $\rho = 0$, or there is no reciprocity in the model. A way to do that is to use the LRT test.
- In [Holland, Leinhardt], data from the null and alternate model are drawn, and is compared to the χ_1^2 cut-off to analyze the performance of the LRT test.
- In all the simulations, the observed test statistic was higher than the corresponding χ^2 value.
- This does not violate the classical theory, as the number of nuisance parameters are growing at $O(n)$, but since there are $O(\frac{n^2}{2})$ independent observations, this does come as a surprise, as the p/n ratio is of order $\frac{2}{n}$.

Advantages of ERGM models

- ERGM models are very easy to interpret.
- In cases where we don't know normalizing constant, we cannot carry out estimation by ML. When the normalizing constant is unknown, the estimation is carried out by first estimating the normalizing constant via MCMLE.
- Even if the asymptotic χ_1^2 approximation is not appropriate, we can always get the cut-offs by sampling from the null models via MCMC.

Caveat

- In some ERGM models, most MCMC samples are degenerate in the sense that they are either very dense or very sparse [Handcock].
- One reason for this can be because the Markov chain is slow mixing.
- Another reason can be that the parameters (in mean parameterization) are very close to the boundary of the parameter space.

Scope for work

- One direction can be to develop the testing of parameters in graphical models by deriving the limiting distribution of the involved test statistic.

- Another direction can be to analyze the rate of convergence of Markov chains to assess the degeneracy issue. Alternatively, one can look for other efficient ways to estimate normalizing constants.

- In case the model itself is degenerate, one can look for alternative models to explain the data. One interesting thing to try is to model the power law of the degree distribution within the ERGM framework. A somewhat related work is by [Dyer, Owen]

References

📄 Holland, P. W., and Leinhardt, S. (1981), *An Exponential Family of Probability Distributions for Directed Graphs,* J. Am. Stat. Assoc., 76, 33-50

📄 Handcock, M. S. (2003), *Assessing Degeneracy in Statistical Models of Social Networks,*Working paper no. 39, Center for Statistics and the Social Sciences, University of Washington-Seattle

📄 Dyer, J. and Owen, A.B.,(2010), *Correct ordering in the Zipf-Poisson ensemble*