Instability, Sensitivity, and Degeneracy of Discrete Exponential Families

Michael Schweinberger

Pennsylvania State University

ONR grant N00014-08-1-1015

Scalable Methods for the Analysis of Network-Based Data

Structure

The problem: simulating large networks and learning the structure of large networks is based on models. Some models of large networks are viable, others are not—impeding MCMC simulation and learning.

The question: which models are non-viable?

The key to answers: notion of sufficient statistics (Fisher 1922): key to MCMC simulation and learning.

Here:

- Introduce notion of unstable sufficient statistics.
- Discuss implications of unstable sufficient statistics: excessive sensitivity and degeneracy.
- Discuss impact of unstable sufficient statistics on MCMC simulation.
- Discuss impact of unstable sufficient statistics on learning.

Models

Frank and Strauss (1986), Wasserman and Pattison (1996): the probability mass function of graph y can be parameterized in exponential family form:

$$P_\theta(Y = y) = \exp(\theta^T g(y) - \psi(\theta))$$

$\Rightarrow$ the mass of graph y is an exponential function of

- $g(y)$: vector of sufficient statistics (Fisher 1922): e.g. number of edges and triangles.
- θ: vector of natural parameters.
- $\mu(\theta) = E_\theta(g(Y))$: vector of mean-value parameters.

Notes:

- $\psi(\theta) = \log \sum_x \exp(\theta^T g(x))$.
- *Natural parameter space:* $\{\theta \in \mathbb{R}^K : \psi(\theta) < \infty\}$.
- *Here: focus on linear exponential families* $\eta(\theta) = A\theta$*; both linear and non-linear exponential families* $\eta(\theta)$ *in Schweinberger (2011).*

Non-viable models

Model may be non-viable, because

- $P_\theta(Y = y)$ is near-degenerate (negative impact on MCMC simulation and learning).
- $P_\theta(Y = y)$ is excessively sensitive to small changes of y (negative impact on MCMC simulation).
- $P_\theta(Y = y)$ is excessively sensitive to small changes of θ (negative impact on learning).

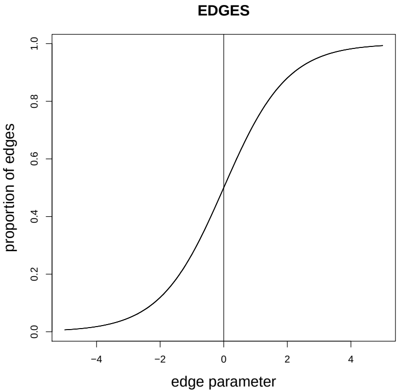
Which models are non-viable?

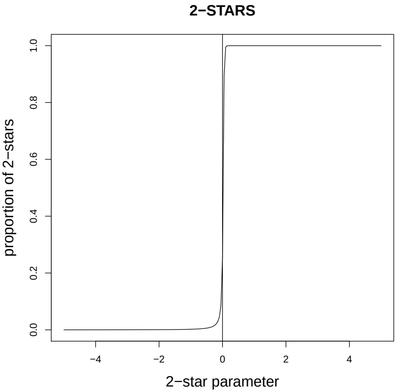
Models with number of 2-stars and triangles (Strauss 1986, Jonasson 1999, Häggström and Jonasson 1999, Snijders 2002, Handcock 2003, Park and Newman 2004a,b, 2005, Rinaldo et al. 2009, Koskinen et al. 2010).

Which models, in general, are non-viable? Which sufficient statistics tend to problematic?

Simple examples

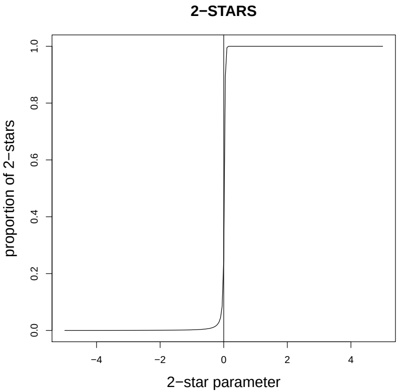
One-parameter exponential families ($n = 32$):

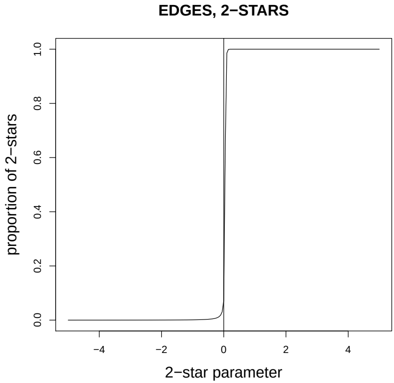




michael.schweinberger@stat.psu.edu

One- and two-parameter exponential families ($n = 32$):





Unstable sufficient statistics

Definition

- **Stable sufficient statistic (SSS):** bounded by number of possible edges N.
- **Unstable sufficient statistics (USS):** not bounded by number of possible edges N.

Examples

- **SSS:** number of edges $\sum_{i<j}^{n} y_{ij}$ is $O(N)$.
- **USS:** number of 2-stars $\sum_{i<j<k}^{n} y_{ij} y_{ik}$ is $O(N^{3/2})$ and number of triangles $\sum_{i<j<k}^{n} y_{ij} y_{jk} y_{ik}$ is $O(N^{3/2})$.

K-parameter exponential families with one USS

Excessive sensitivity

If n is large, $P_\theta(Y = y)$ tends to be extremely sensitive to small, local changes of y: some, but not necessarily all, single-site log odds $\log \frac{P_\theta(Y = x)}{P_\theta(Y = y)}$ tend to be extremely large. A walk through the set of y resembles a walk through rugged, mountaineous landscape: small increases in y can lead to dramatic increases and descreases in probability mass. Example: models with number of 2-stars and triangles.

Degeneracy

If n is large, model tends to be degenerate wrt USS $g_1(y)$:

- all $\theta_1 < 0$: probability mass tends to be concentrated on graphs close to the (greatest) lower bound of USS; so is mean-value parameter $\mu_1(\theta) = E_\theta(g_1(Y))$.
- all $\theta_1 > 0$: probability mass tends to be concentrated on graphs close to the (lowest) upper bound of USS; so is mean-value parameter $\mu_1(\theta) = E_\theta(g_1(Y))$.

K-parameter exponential families with multiple USS

Excessive sensitivity and degeneracy

- **One dominating USS:** same excessive sensitivity and degeneracy issues as above.
- **No dominating USS:** unless clever parametrization is chosen, counterbalancing unstable statistics may not work.

Impact on MCMC simulation and learning

MCMC simulation:

- By excessive sensitivity: small, local changes in y can result in extremely large changes in probability mass.
- By degeneracy: simulated networks tend to be degenerate wrt USS.

Learning:

- If y is "extreme" in terms of $g(y)$, maximum likelihood estimator of θ does not exist (Handcock 2003).
- Even if y is not "extreme" in terms of $g(y)$, learning is problematic:
  (a) the effective parameter space tends to be small.
  (b) the estimating function of the method of maximum likelihood estimation

$$\nabla_\theta \log P_\theta(Y = y) = g(y) - E_\theta(g(Y)) = g(y) - \mu(\theta)$$

  tends to be extremely sensitive to changes in θ.
  (c) MCMC simulation-based maximum likelihood estimation algorithms—exploiting simulated network to estimate θ—do not simulate networks which cluster around observed networks in terms of sufficient statistics and therefore tend to suffer from computational failure (Handcock 2003).

Discussion

- **Question:** can unstable sufficient be stabilized?
- **Tentative answer:** simple stabilization strategies fail to address the problem of lack of fit (Hunter et al. 2008) due to the non-uniqueness of the canonical form of exponential families and the paramerization-invariance of maximum likelihood estimators.
- **Question:** instability implies non-viability, does stability imply viability?
- **Tentative answer:** stability may be too weak; maybe semi-group structure (Lauritzen 1988); semi-group structure implies stability.
- **Technical details:** see Schweinberger (2011).

Conclusion: the notion of instability is useful for characterizing, detecting, and penalizing non-viable models which are useless for simulating large networks and learning parameters from large networks.

References

Fisher, R. A. (1922), "On the mathematical foundations of theoretical statistics," *Philosophical Transactions of the Royal Society of London, Series A*, 222, 309–368.

Frank, O., and Strauss, D. (1986), "Markov graphs," *Journal of the American Statistical Association*, 81, 832–842.

Häggström, O., and Jonasson, J. (1999), "Phase transition in the the random triangle model," *Journal of Applied Probability*, 36, 1101–1115.

Handcock, M. (2003), "Assessing degeneracy in statistical models of social networks," Tech. rep., Center for Statistics and the Social Sciences, University of Washington, http://www.csss.washington.edu/Papers.

Hunter, D. R., Goodreau, S. M., and Handcock, M. S. (2008), "Goodness of fit of social network models," *Journal of the American Statistical Association*, 103, 248–258.

Jonasson, J. (1999), "The random triangle model," *Journal of Applied Probability*, 36, 852–876.

Koskinen, J. H., Robins, G. L., and Pattison, P. E. (2010), "Analysing exponential random graph (p-star) models with missing data using Bayesian data augmentation," *Statistical Methodology*, 7, 366–384.

Lauritzen, S. L. (1988), *Extremal Families and Systems of Sufficient Statistics*, Heidelberg: Springer.

Park, J., and Newman, M. E. J. (2004a), "Solution of the two-star model of a network," *Physical Review E*, 70, 66–146.

— (2004b), "The statistical mechanics of networks," *Physical Review E*, 70, 66–117.

— (2005), "Solution for the properties of a clustered network," *Physical Review E*, 72, 026136.

Rinaldo, A., Fienberg, S. E., and Zhou, Y. (2009), "On the geometry of discrete exponential families with application to exponential random graph models," *Electronic Journal of Statistics*, 3, 446–484.

Schweinberger, M. (2011), "Instability, sensitivity, and degeneracy of discrete exponential families," *Journal of the American Statistical Association*, in press.

Snijders, T. A. B. (2002), "Markov chain Monte Carlo Estimation of exponential random graph models," *Journal of Social Structure*, 3, 1–40.

Strauss, D. (1986), "On a general class of models for interaction," *SIAM Review*, 28, 513–527.

Wasserman, S., and Pattison, P. (1996), "Logit Models and Logistic Regression for Social Networks: I. An Introduction to Markov Graphs and p^*," *Psychometrika*, 61, 401–425.