Exponential Random Graph Models and Their Polytopes

Johannes Rauh

York University
(the one in Canada)

AMS Sectional Meeting
Chicago 2015

Outline

1. Discrete Exponential Families
2. Graphical Models
3. Exponential Random Graph Models

Discrete Exponential Families

Given a finite sample space $\mathcal{X}$ and a vector of statistics $f : \mathcal{X} \to \mathbb{R}^h$, we can associate a discrete exponential family

$$\mathcal{E} = \left\{ P_\theta \;:\; P_\theta(x) = \frac{1}{Z_\theta} \exp(\langle \theta, f_x \rangle) \right\}$$

Theorem

Let $m = \mathbb{E}_{P_\theta}[f]$. Then $P_\theta = \underset{P:\mathbb{E}_P[f]=m}{\operatorname{argmax}} \{H(P)\}$.

Jaynes' principle of maximum entropy: If you know nothing about a distribution but its expectation value $\mathbb{E}[f]$, you should use P_θ.

Theorem (MLE)

Let P be the empirical distribution of some data set, and suppose that $\mathbb{E}_P[f] = \mathbb{E}_{P_\theta}[f]$. Then P_θ is the unique maximum likelihood estimate (MLE).

The Moment Map

$P_\theta(x) > 0$ for all x, θ. However, $P(x) = 0$ is possible in the closure/boundary of $\mathcal{E}$. The closure is denoted by $\overline{\mathcal{E}}$.

Definition

The map $\mu : P \mapsto \mathbb{E}_P[f]$ is the moment map. The image

$$\mathbf{M} = \operatorname{conv}\{f_x : x \in \mathcal{X}\}.$$

is the convex support polytope

Theorem

μ restricts to a bijection $\overline{\mathcal{E}} \cong \mathbf{M}$.

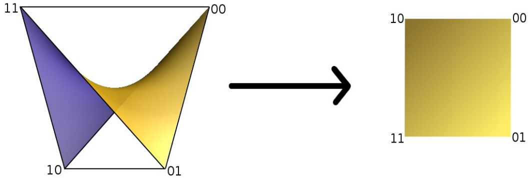
The inverse will be denoted by μ^{-1}.

Theorem

Suppose m belongs to the face $\mathbf{F} \subseteq \mathbf{M}$, and let $P = \mu^{-1}(m)$. Then $P(x) > 0$ if and only if $f_x \in \mathbf{F}$.

Example

The independence model of two binary variables:



- The idea is that $\mathcal{E}$ “looks like” $\mathbf{M}$.
- The combinatorics of $\mathbf{M}$ reflect properties of $\mathcal{E}$.

Generalized MLEs

Theorem (Generalized MLE)

Let P be the empirical distribution, and let $m = \mathbb{E}_P[f]$. Then $\mu^{-1}(m)$ is the unique maximum likelihood estimate (MLE) within $\overline{\mathcal{E}}$.

- The MLE within $\overline{\mathcal{E}}$ is also called the generalized MLE (GMLE) for $\mathcal{E}$.
- If $\mu^{-1}(m)$ belongs to the boundary $\partial\mathcal{E} := \overline{\mathcal{E}} \setminus \mathcal{E}$, one says that "the MLE does not exist."
- If the MLE does not exist:
  1. not enough data?
  2. structural zero?

  The GMLE has no parameters, so how do you estimate?

Graphical Models

Let $G = (V, E)$ be a graph with $n = |V|$ nodes. To each node $v \in V$ associate a finite random variable X_v, taking values in $\mathcal{X}_v$. Thus, $\mathcal{X} = \times_{v \in V} \mathcal{X}_v$. A joint distribution of $(X_v)_{v \in V}$ is a $|V|$-dimensional tensor

$$(p_{x_1,x_2,\dots,x_n})_{x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, \dots, x_n \in \mathcal{X}_n}.$$

Denote by $C(G)$ the set of cliques of G (i.e. the complete subgraphs).

Definition

The graphical model is the set of all probability distributions P on $\mathcal{X}$ of the form

$$P(x_1, \dots, x_n) = \prod_{C = \{i_1,\dots,i_k\} \in C(G)} \phi_C(x_{i_1}, \dots, i_k),$$

where ϕ_C is a positive function.

Lemma

The graphical model is a discrete exponential family. The vector of statistics contains the C-marginals for all $C \in C(G)$.

Marginal Polytopes

The convex support $\mathbf{M}$ of a graphical model is a marginal polytope.

- Marginal polytopes are 0/1-polytopes (each f_x is a 0/1-vector)
- Therefore, each f_x is a vertex of $\mathbf{M}$.
- Moreover, $\mathbf{M}$ is a subpolytope of a hypercube.
- Every f_x has the same number of ones (in some parametrization. . .)
- Therefore, all f_x lie on a sphere.
- Marginal polytopes are symmetric
  (the symmetry group acts transitively on the vertices)

Graphical models have been proven to be a versatile modelling platform in many applications.

Exponential Random Graph Models

Exponential Random Graph Models (ERGMs) are discrete exponential families where the sample space $\mathcal{G}$ is a set of graphs, e.g.

$$\mathcal{G} = \mathcal{G}_n := \{\text{graphs on } n \text{ nodes}\}.$$

Any choice of graph statistics defines an ERGM, for example:

- Subgraph counts (triangles, cycles, k-stars, ...)
- Degree statistics (average degree, degree distribution, degree sequence, ...)

Changing the Number of Nodes

Often, the number of nodes n is fixed in theory; i.e. the sample space is

$$\mathcal{G} = \mathcal{G}_n := \{\text{graphs on } n \text{ nodes}\}.$$

In applications it is often a variable. What happens when changing n?

- Many graph statistics behave "continuously" when changing n. However: Need to take into account scaling of the parameters!
- For subgraph densities (normalized counts), the convex support polytopes converge (Engström, Norén 2011):

$$\mathbf{M}_n \supseteq \mathbf{M}_{n+1} \supseteq \cdots \supseteq \mathbf{M}_\infty := \bigcap_{n'} \mathbf{M}_{n'}.$$

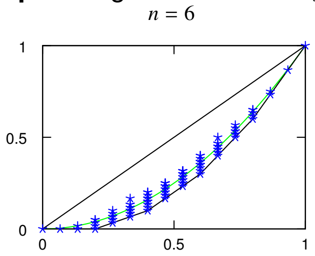
- For other statistics (e.g. degree sequences), the number of parameters depends on n (Chatterjee, Diaconis 2011)
- n plays a role similar to the sample size.
- sparse vs. dense graphs: Restrict $\mathcal{G}$ or penalize large edge density?

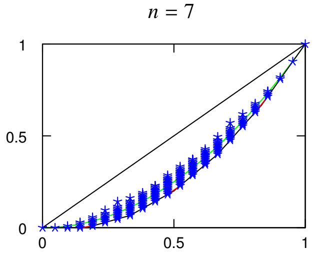
ERGMs and Small Samples

When analyzing networks, often there is just a single observation.

- The MLE might still exist—provided most f_G are not vertices.
- For many ERGMs, this is indeed the case.
  The "geometry" of such an ERGMs is determined not only by their convex support $\mathbf{M}$, but also by the location of the points f_G within $\mathbf{M}$.

Example: Edges and 2-stars (Rauh 2012)





An LDP for ERGMs

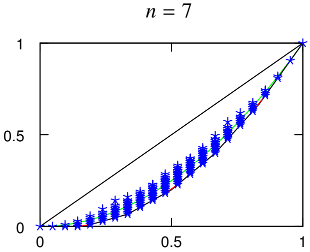
Chatterjee and Varadhan (2010) proved a Large Deviation Principle that describes the distribution of f_G for many statistics for large n.

- **Intuition:** The uniform distribution on (labelled!) graphs on n nodes is equivalent to $\binom{n}{2}$ independent random variables.
- As a consequence, most f_G lie close to the center of mass of $\mathbf{M}$ (corresponding to the uniform distribution).
- Of those f_G that don't lie close to the center of mass, most lies close to some Erdős-Rényi graph.

Asymptotics of ERGMs and Degeneracy

Diaconis and Chatterjee (2011) showed for certain examples of subgraph counts that for large n and "reasonably scaled parameters" θ, the random graph P_θ lies close to some Erdős-Rényi graph.

In principle, you can always escape the Erdős-Rényi by choosing large parameters—but maybe not in all directions!



A similar degeneracy has been observed in applications: The MLEs of real networks are often close to Erdős-Rényi.

$\Longrightarrow$ If you are not happy with Erdős-Rényi, you need to change f.

Degree Sequences and Partitions

Some ERGMs don't have interior points:

- Let $f_{\text{ds}}(G)$ = the degree sequence. $\mathcal{M}_{\text{ds}}$ many interior points.
- Let $f_{\text{dp}}(G)$ = the degree sequence ordered by magnitude, larger degrees first (degree partition). $\mathcal{M}_{\text{dp}}$ has no interior point.

  (reason: every degree partition has a repetition $d_i = d_{i+1}$, and $d_i \geq d_{i+1}$ defines a facet)
- Let $f_{\text{dd}}(G)$ = the degree distribution ($\cong$ all k-stars). $\mathcal{M}_{\text{dd}}$ has no interior point. (reason: a graph cannot have both an isolated node and a fully connected node)

$\mathcal{M}_{\text{dp}}$ and $\mathcal{M}_{\text{ds}}$ are related (Bhattacharya, Sivasubramanian, Srinivasan 2006): $\mathcal{M}_{\text{ds}}$ consists of $n!$ copies of $\mathcal{M}_{\text{dp}}$, corresponding to all ways of ordering the nodes.

Further example (joint work with K. Sadeghi, T. Short, É. Czabarka, L. Szekely):

- Bi-degree statistics: Find the smallest face containing some f_G in its interior.

So. . . what's the difference between graphical models and ERGMs?

So. . . what's the difference between graphical models and ERGMs?

- In graphical models, the nodes are random, but in ERGMS, the edges are random.

So…what's the difference between graphical models and ERGMs?

- ~~In graphical models, the nodes are random, but in ERGMS, the edges are random.~~

Technically true, but…

Comparing graphical models and ERGMs

- Both are very different examples of discrete exponential families.
- Graphical models are a fixed (but very flexible) class of models.
- Properties of ERGMs depend very much on the chosen statistics.

Typically:

graphical models		ERGMs
every f_x is a vertex	$\longleftrightarrow$	most f_G are not vertices
f_x distributed on a sphere	$\longleftrightarrow$	f_G cluster at the center of mass
high symmetry	$\longleftrightarrow$	almost no symmetry

Comparing graphical models and ERGMs

- Both are very different examples of discrete exponential families.
- Graphical models are a fixed (but very flexible) class of models.
- Properties of ERGMs depend very much on the chosen statistics.

Typically:

graphical models		ERGMs
every f_x is a vertex	$\longleftrightarrow$	most f_G are not vertices
f_x distributed on a sphere	$\longleftrightarrow$	f_G cluster at the center of mass
high symmetry	$\longleftrightarrow$	almost no symmetry

Actually, there is a way to combine ERGMs and graphical models:
Study conditional independences between edges (Lauritzen, Rinaldo, Sadeghi)