Algorithmic approaches to fitting ERG models

Ruth Hummel, Penn State University
Mark Handcock, University of Washington
David Hunter, Penn State University

Research funded by Office of Naval Research Award No. N00014-08-1-1015

MURI meeting, April 24, 2009
Outline

1. Introduction
2. Partial stepping
3. Biological network example
4. References
The class of Exponential-family Random Graph Models (ERGMs):

**Definition**

\[ P_\eta(Y = y) = \frac{\exp\{\eta^t g(y)\}}{\kappa(\eta)} \]

where

- \( Y \) is a random network written as an adjacency matrix so that \( Y_{ij} \) is the indicator of an edge from \( i \) to \( j \);
- \( g(y) \) is a vector of the network statistics of interest;
- \( \eta \) is a vector of parameters corresponding to the vector \( g(y) \);
- \( \kappa(\eta) \) is the constant of proportionality which makes the probabilities sum to one; intractable.
The loglikelihood for this class of models is

\[ l(\eta) = \eta^t g(y_{\text{obs}}) - \log \sum_{z \in \mathcal{Y}} \exp(\eta^t g(z)). \]  

(1)

which can be written:

\[ l(\eta) - l(\eta_0) = \left( \eta - \eta_0 \right)^t g(y_{\text{obs}}) - \log E_{\eta_0} \left[ \exp \left\{ \left( \eta - \eta_0 \right)^t g(Y) \right\} \right]. \]  

(2)
Maximum Pseudolikelihood Estimation

Notation: For a network $y$ and a pair $(i, j)$ of nodes,

- $y_{ij} = 0$ or $1$, depending on whether there is an edge
- $y_{ij}^c$ denotes the status of all pairs in $y$ other than $(i, j)$
- $y_{ij}^+$ denotes the same network as $y$ but with $y_{ij} = 1$
- $y_{ij}^-$ denotes the same network as $y$ but with $y_{ij} = 0$

- Assume no dependence among the $Y_{ij}$.
- In other words, assume $P(Y_{ij} = 1) = P(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c)$.
- Then some algebra gives

$$
\log \frac{P(Y_{ij} = 1)}{P(Y_{ij} = 0)} = \theta^t \left[ g(y_{ij}^+) - g(y_{ij}^-) \right],
$$

so $\theta$ is estimated by straightforward logistic regression.

- Result: The maximum pseudolikelihood estimate.
Maximum Pseudolikelihood Estimation

Notation: For a network $y$ and a pair $(i, j)$ of nodes,

- $y_{ij} = 0$ or $1$, depending on whether there is an edge
- $y_{ij}^c$ denotes the status of all pairs in $y$ other than $(i, j)$
- $y_{ij}^+$ denotes the same network as $y$ but with $y_{ij} = 1$
- $y_{ij}^-$ denotes the same network as $y$ but with $y_{ij} = 0$

Assume no dependence among the $Y_{ij}$.
In other words, assume $P(Y_{ij} = 1) = P(Y_{ij} = 1 | Y_{ij}^c = y_{ij}^c)$.
Then some algebra gives

$$\log \frac{P(Y_{ij} = 1)}{P(Y_{ij} = 0)} = \theta^t \left[ g(y_{ij}^+) - g(y_{ij}^-) \right],$$

so $\theta$ is estimated by straightforward logistic regression.

Result: The maximum pseudolikelihood estimate.
**MLE’s behavior**

- **MLE (maximum likelihood estimation):** Well-established method but very hard because the normalizing constant $\kappa(\alpha)$ is difficult (usually impossible) to evaluate, so we approximate it instead.

- **MPLE (maximum pseudo-likelihood estimation):** Easy to do using logistic regression, but based on an independence assumption that is often not justified.

Several authors, notably van Duijn et al. (2009), argue forcefully against the use of MPLE.
Remember that the loglikelihood

\[ l(\eta) = \eta^t g(y_{obs}) - \log \sum_{z \in Y} \exp(\eta^t g(z)). \]  

(3)

can be written:

\[ l(\eta) - l(\eta_0) = (\eta - \eta_0)^t g(y_{obs}) - \log E_{\eta_0} \left[ \exp \left\{ (\eta - \eta_0)^t g(Y) \right\} \right], \]

This leads to our first approximation for the MLE:

- MCMC MLE idea: Pick \( \theta_0 \), draw \( Y_1, \ldots, Y_m \) from this model using MCMC, then approximate the population mean above by a sample mean.

- We can take \( \theta_0 \) to be, for example, the MPLE.
Remember that the loglikelihood
\[ l(\eta) = \eta^t g(y_{obs}) - \log \sum_{z \in \mathcal{Y}} \exp(\eta^t g(z)). \] (3)

This can be written:
\[ l(\eta) - l(\eta_0) = (\eta - \eta_0)^t g(y_{obs}) - \log E_{\eta_0} \left[ \exp \left\{ (\eta - \eta_0)^t g(Y) \right\} \right], \]

This leads to our first approximation for the MLE:

- MCMC MLE idea: Pick \( \theta_0 \), draw \( Y_1, \ldots, Y_m \) from this model using MCMC, then approximate the population mean above by a sample mean.
- We can take \( \theta_0 \) to be, for example, the MPLE.
Back to the loglikelihood

- Remember that the loglikelihood

\[ l(\eta) = \eta^t g(y_{\text{obs}}) - \log \sum_{z \in Y} \exp(\eta^t g(z)). \]  

(3)

can be written:

\[ l(\eta) - l(\eta_0) = (\eta - \eta_0)^t g(y_{\text{obs}}) - \log E_{\eta_0} \left[ \exp \left\{ (\eta - \eta_0)^t g(Y) \right\} \right], \]

This leads to our first approximation for the MLE:
- MCMC MLE idea: Pick \( \theta_0 \), draw \( Y_1, \ldots, Y_m \) from this model using MCMC, then approximate the population mean above by a sample mean.
- We can take \( \theta_0 \) to be, for example, the MPLE.
Existence of MLE?

There are two issues we need to be concerned with when using MCMC MLE:

- First, if the observed $g(y_{obs})$ is not in the interior of the convex hull of the sampled (MCMC generated) $g(y_i)$, then no maximizer of the approximate loglikelihood ratio exists.

- Also, the approximation of the likelihood surface,

$$l(\eta) - l(\eta_0) \approx (\eta - \eta_0)^t g(y_{obs}) - \log \frac{1}{m} \sum_{i=1}^{m} \exp((\eta - \eta_0)^t g(Y_i)),$$

is not very good when we get far from $\eta_0$. 
The Erdős-Rényi model can be written as an ERGM, if $g(y)$ is the number of edges of $y$.

Since each edge exists independently with probability $b$, 

$$ P(Y = y) = (b)^{g(y)} (1 - b)^{n^* - g(y)} = \left( \frac{b}{1 - b} \right)^{g(y)} (1 - b)^{n^*}. $$
Comparison of the true loglikelihood ratio and its estimation for an Erdos-Renyi graph.

True and estimated loglikelihood ratios plotted for $\eta = -0.69$ (binomial probability = 1/3).

- True LLR
- Estimated LLR
- True parameter
- MLE
- Starting value

$\eta_0 = -1$ (binomial probability = 0.269, MLE is $-0.55$ (binomial probability = 0.365).

April 2009

Algorithmic approaches to fitting ERG models
We would like to move our arbitrary initial value $\eta_0$ close enough to the MLE that samples (of the statistics) generated from $\eta_0$ cover the observed statistics, $g(y_{obs})$.

New Idea: Intuitively, then, we might try to find a way to make the journey from $\eta_0$ to the MLE in pieces.

In order to take steps toward the MLE, we need to have some idea where we are going. We obviously don’t know where the MLE is, but we do know that the MLE has the following property:

**Definition**

The MLE, if it exists, is the unique parameter vector $\eta$ satisfying $E_{\eta}g(Y) = g(Y_{obs}) = \hat{\xi}$.

This is true for any exponential family (Barndorff-Nielson, 1978, and Brown, 1986).
Mean value parametrization

1. We would like to move our arbitrary initial value $\eta_0$ close enough to the MLE that samples (of the statistics) generated from $\eta_0$ cover the observed statistics, $g(y_{obs})$.

2. New Idea: Intuitively, then, we might try to find a way to make the journey from $\eta_0$ to the MLE in pieces.

In order to take steps toward the MLE, we need to have some idea where we are going. We obviously don’t know where the MLE is, but we do know that the MLE has the following property:

**Definition**

The MLE, if it exists, is the unique parameter vector $\eta$ satisfying $E_{\eta} g(Y) = g(y_{obs}) = \hat{\xi}$.

This is true for any exponential family (Barndorff-Nielson, 1978, and Brown, 1986).
We would like to move our arbitrary initial value $\eta_0$ close enough to the MLE that samples (of the statistics) generated from $\eta_0$ cover the observed statistics, $g(y_{obs})$.

New Idea: Intuitively, then, we might try to find a way to make the journey from $\eta_0$ to the MLE in pieces.

In order to take steps toward the MLE, we need to have some idea where we are going. We obviously don’t know where the MLE is, but we do know that the MLE has the following property:

**Definition**

The MLE, if it exists, is the unique parameter vector $\eta$ satisfying $E_{\eta}g(Y) = g(y_{obs}) = \hat{\xi}$.

This is true for any exponential family (Barndorff-Nielson, 1978, and Brown, 1986).
\[ \eta_0 \text{ is the (given) initial value of } \eta \text{ in the Markov Chain} \]

\[ \xi_0 \text{ is the vector of mean statistics corresponding to } \eta_0 \]

\[ \hat{\eta} \text{ is the (unknown) MLE} \]

\[ \hat{\xi} \text{ is the observed statistics, } g(y_{\text{obs}}), \text{ which is the corresponding mean statistics vector for } \hat{\eta}. \]
We now want to move toward $\hat{\xi} = g(y_{\text{obs}})$ in mean value parameter space. Here we specify a step length, $0 \leq \gamma \leq 1$, as a fraction of the distance toward $\hat{\xi}$ that we want to traverse. We move the fraction $\gamma$ toward $\hat{\xi}$ and call this point $\xi_1 = \gamma \hat{\xi} + (1 - \gamma)\hat{\xi}_0$.

At each step, we choose this fraction to be the biggest move toward $g(y_{\text{obs}})$ that does not leave the convex hull of the current sample.
We now want to move toward $\hat{\xi} = g(y_{\text{obs}})$ in mean value parameter space. Here we specify a step length, $0 \leq \gamma \leq 1$, as a fraction of the distance toward $\hat{\xi}$ that we want to traverse. We move the fraction $\gamma$ toward $\hat{\xi}$ and call this point $\xi_1 = \gamma \hat{\xi} + (1 - \gamma)\hat{\xi}_0$.

At each step, we choose this fraction to be the biggest move toward $g(y_{\text{obs}})$ that does not leave the convex hull of the current sample.
Next, using the sample from the model defined by $\eta_0$, we maximize the approximate loglikelihood

$$l(\eta) - l(\eta_0) \approx (\eta - \eta_0)^t g(y_{obs}) - \log \frac{1}{m} \sum_{i=1}^{m} \exp((\eta - \eta_0)^t g(Y_i)), \quad (6)$$

but with $\xi_1$ substituted in place of $g(y_{obs})$. The resulting maximizer will be called $\eta_1$. The process then repeats, with $\eta_1$ taking the place of $\eta_0$. 
0. Set \( \eta_0 \).

1. Take an MCMC sample from the model defined by \( \eta = \eta_0 \) to get \( \xi_0 \).
1. Let \( \eta_0 = \text{given} \) and \( \hat{\eta} = ??? \).

2. Go \( \gamma \% \) toward \( \hat{\xi} = g(y_{\text{obs}}) \) in mean value parameter space. Call this \( \xi_1 \).
3. Use $\eta_0$ in MCMC MLE to find the $\eta$ that corresponds to $\xi_1$. Call this $\eta_1$. 
4. Re-estimate $\xi_1$ from an MCMC sample from the model defined by $\eta = \eta_1$. Call this $\hat{\xi}_1$. 
5. Repeat step (2) by going $\gamma\%$ toward $\hat{\xi} = g(y_{obs})$ from $\hat{\xi}_1$. Call this $\xi_2$. Also repeat steps (3) and (4) to obtain $\hat{\xi}_2$. Keep going until $g(y_{obs})$ is in the convex hull of the new sample.
6. Use MCMC MLE with $\eta_i$ as the initial value to find $\hat{\eta}$. If we have made it into the appropriate neighborhood around $\hat{\eta}$, this will now be possible.
In review:

In other words, for each $t \geq 0$, we first use MCMC to draw a random sample $Y_1, \ldots, Y_m$ from the model determined by $\eta_t$, then we set

$$\hat{\xi}_t = \frac{1}{m} \sum_{i=1}^{m} g(Y_i);$$

$$\xi_{t+1} = \gamma_t \hat{\xi} + (1 - \gamma_t) \hat{\xi}_t;$$

$$\eta_{t+1} = \arg \max_{\eta} \left\{ (\eta - \eta_0)^t \xi_{t+1} - \log \left[ \frac{1}{m} \sum_{i=1}^{m} \exp \left\{ (\eta - \eta_0)^t g(y_i) \right\} \right] \right\};$$

We iterate until $\hat{\xi} = g(y_{\text{obs}})$ is in the convex hull of the statistics generated from $\hat{\xi}_t$. 
A second approximation: the lognormal

- Here, the ERGM is very simple, $g(y) = \text{edges}$.
- In other words, the model is binomial.

Nevertheless, naive MCMC approximation of the log likelihood ratio is not good far from $\theta_0$, even for gigantic samples.

One possible remedy: Assume $(\theta - \theta_0)^t g(Y)$ is normally distributed in

$$
\ell(\theta) - \ell(\theta_0) = (\theta - \theta_0)^t g(y^{\text{obs}}) - \log \mathbb{E}_{\theta_0} \left[ \exp \left\{ (\theta - \theta_0)^t g(Y) \right\} \right].
$$
Outline

1. Introduction
2. Partial stepping
3. Biological network example
4. References
Example Bionet: E. Coli (Salgado et al 2001)

- A node is an operon
- Edge $A \rightarrow B$ means $A$ encodes a transcription factor that regulates $B$.
- Green indicates self-regulation
Another ERGM for the E. Coli network

We fit a model similar to that of Saul and Filkov (2007)

<table>
<thead>
<tr>
<th>Term(s)</th>
<th>Description:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edges</td>
<td>Number of Edges</td>
</tr>
<tr>
<td>2-Deg, . . . , 5-Deg</td>
<td>Nodes with degree 2, . . . , 5</td>
</tr>
<tr>
<td>GWDeg</td>
<td>Single statistic: Weighted sum of 1-Deg, . . . , (n – 1)-Deg with weights tending to 1 at a geometric rate</td>
</tr>
</tbody>
</table>

```r
model <- ergm(ecoli2 ~ edges + degree(2:5) + gwdegree(0.25, fixed=TRUE), MPLEonly=TRUE)

MPLE
   edges degree2 degree3 degree4 degree5 gwdegree
     -5.35   -2.58   -3.06   -2.39   -1.85     8.13
```

Algorithmic approaches to fitting ERG models
Another ERGM for the E. Coli network

We fit a model similar to that of Saul and Filkov (2007)

<table>
<thead>
<tr>
<th>Term(s)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edges</td>
<td>Number of Edges</td>
</tr>
<tr>
<td>2-Deg, . . . , 5-Deg</td>
<td>Nodes with degree 2, . . . , 5</td>
</tr>
<tr>
<td>GWDeg</td>
<td>Single statistic: Weighted sum of 1-Deg, . . . , (n – 1)-Deg with weights tending to 1 at a geometric rate</td>
</tr>
</tbody>
</table>

model <- ergm(ecoli2 ~ edges + degree(2:5) + gwdegree(0.25, fixed=TRUE), MPLEonly=TRUE)

MPLE

<table>
<thead>
<tr>
<th></th>
<th>edges</th>
<th>degree2</th>
<th>degree3</th>
<th>degree4</th>
<th>degree5</th>
<th>gwdegree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-5.35</td>
<td>-2.58</td>
<td>-3.06</td>
<td>-2.39</td>
<td>-1.85</td>
<td>8.13</td>
</tr>
</tbody>
</table>
MPL fit is degenerate

With the MPL we encounter problems:

Here is a time-series plot of the edge-count of the 25 networks generated from the MPL:
A sample from the MPLE-fitted model

<table>
<thead>
<tr>
<th>Iteration 1: Yellow = 0.06(Red) + 0.94(Green)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image.png" alt="Graph with edge and degree scatter plots" /></td>
</tr>
</tbody>
</table>

- **Red**: \( g(y^{\text{obs}}) \)
- **Green**: Sample mean
- **Theory**: No maximizer of the approximated likelihood exists because \( g(y^{\text{obs}}) \) is not in the interior of the convex hull of the sampled points.
- **However**, the likelihood depends on the data only through \( g(y^{\text{obs}}) \)
- **Idea**: What if we pretend \( g(y^{\text{obs}}) \) is the yellow point?
A sample from the MPLE-fitted model

Iteration 1: Yellow = 0.06(Red) + 0.94(Green)

- Red: $g(y^{\text{obs}})$
- Green: Sample mean
- Theory: No maximizer of the approximated likelihood exists because $g(y^{\text{obs}})$ is not in the interior of the convex hull of the sampled points.
- However, the likelihood depends on the data only through $g(y^{\text{obs}})$
- Idea: What if we pretend $g(y^{\text{obs}})$ is the yellow point?
A sample from a model with a better $\theta_0$

For $\theta_0$ we have replaced the MPLE by the MCMC “MLE” obtained by pretending that $g(y^{obs})$ was the yellow point on the previous slide.
A sample from a model with a better $\theta_0$

Iteration 3: Yellow = 0.18(Red) + 0.82(Green)

Continue to iterate...
A sample from a model with a better $\theta_0$

Iteration 4: Yellow = 0.36(Red) + 0.64(Green)
A sample from a model with a better $\theta_0$

**Iteration 5:** Yellow = 0.53(Red) + 0.47(Green)
A sample from a model with a better $\theta_0$

- Finally, we don't need to pretend; the true $g(y^{\text{obs}})$ is actually interior to the convex hull of sampled points...
A sample from a model with a better $\theta_0$

Final Iteration (#7): Green = mean, Yellow=observed

\[ \text{edges} \]
\[ \text{degree2} \]
\[ \text{degree3} \]
\[ \text{degree4} \]
\[ \text{degree5} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]

\[ \text{gwdegree} \]
Finally, an MLE

Original MPLE:

<table>
<thead>
<tr>
<th>edges</th>
<th>degree2</th>
<th>degree3</th>
<th>degree4</th>
<th>degree5</th>
<th>gwdegree</th>
</tr>
</thead>
<tbody>
<tr>
<td>-5.35</td>
<td>-2.58</td>
<td>-3.06</td>
<td>-2.39</td>
<td>-1.85</td>
<td>8.13</td>
</tr>
</tbody>
</table>

Final (approximated) MLE:

<table>
<thead>
<tr>
<th>edges</th>
<th>degree2</th>
<th>degree3</th>
<th>degree4</th>
<th>degree5</th>
<th>gwdegree</th>
</tr>
</thead>
<tbody>
<tr>
<td>-5.06</td>
<td>-1.45</td>
<td>-2.35</td>
<td>-2.28</td>
<td>-2.91</td>
<td>1.77</td>
</tr>
</tbody>
</table>

We know this is close to the true MLE $\hat{\theta}$ because the true MLE uniquely gives $E_{\hat{\theta}}g(Y) = g(y^{obs})$. 
Failure of first approximation in this example

Iteration 5: Yellow = 0.13(Red) + 0.87(Green)

A steplength of 0.01 is too small in this case.
Conclusions

- MPLE looks increasingly dangerous; it can mask problems when they exist and miss badly when they don’t.
- Naive MCMC MLE may not perform well even in very simple problems, but it may be modified. Here, we had success in a hard problem using two ideas:
  - (a) Partial stepping toward the MLE in mean-value parameter space;
  - (b) A log-normal approximation to the normalizing constant.
- By making MLE more and more automatic, we hope that scientists will be able to focus on modeling, not programming.
Cited References

Thank You!