Fitting Social Network Models to Data

Mark S. Handcock

handcock@ucla.edu

Statistical Analysis of Networks

October 22, 2024

Fitting a General Model to Data

- Suppose we have network data and want to see how the models fit
- Initially focus on models where the Y_{ij} are independent (given covariates X), so that

$$\text{logit}[P(Y_{ij} = 1 | X = x, \beta)] = \beta_1 x_{1,ij} + \beta_2 x_{2,ij} \ldots + \beta_K x_{K,ij}$$

  for some choice of covariates $x_{1,ij}, x_{2,ij}, \ldots, x_{K,ij}$

Estimation based on the Likelihood Function

- Develop a framework based on the likelihood function
- What is maximum likelihood estimation?
  Find β that makes the observed network most likely under the model with β as the parameter.

The log of the *likelihood* (as a function of β) is

$$\ell(\beta|y,\ x) \equiv \log[P(Y = y|X = x,\ \beta)] \quad \beta \in \mathrm{R}^K, \quad \text{given } y \in \mathcal{Y}$$

Example: Homogeneous Bernoulli model Rényi and Erdős model

- Y_{ij} are independent and equally likely

$$P(Y_{ij} = 1 | X = x, \beta) = \frac{\exp(\beta)}{1 + \exp(\beta)} \qquad \forall \;\; i, j = 1, \ldots, n$$

for some $-\infty < \beta < \infty$

Example: Homogeneous Bernoulli model Rényi and Erdős model

- Y_{ij} are independent and equally likely

$$P(Y_{ij} = 1 | X = x, \beta) = \frac{\exp(\beta)}{1 + \exp(\beta)} \qquad \forall \; i, j = 1, \ldots, n$$

for some $-\infty < \beta < \infty$

- Equivalently

$$\log \text{odds}(Y_{ij} = 1 | X = x, \beta) = \beta \qquad \forall \; i, j = 1, \ldots, n$$

Example: Homogeneous Bernoulli model Rényi and Erdős model

More abstractly:

$$P(Y = y|X = x,\ \beta) = \frac{e^{\beta g(y)}}{\kappa(\beta)} \qquad y \in \mathcal{Y}$$

$\kappa(\beta) = [1 + \exp(\beta)]^N$

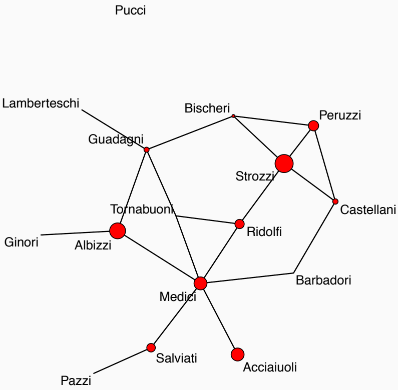
$$g(y) = \sum_{i,j} y_{ij} \quad \text{the number of ties}$$

The log of the *likelihood* (as a function of β) is

$$\ell(\beta|y,\, x) \equiv \log[P(Y = y|X = x,\ \beta)] \ = \ \beta g(y) - N \log[1 + \exp(\beta)] \quad -\infty$$

Example: Effect on the number of Priorates on centrality

Let X be the number of priorates (seats on the civic council) held by the family between 1282- 1344.



Joint specification

Recall

$$P(Y = y|X = x,\ \beta) = \frac{e^{\beta_1 g_1(y)+\beta_2 g_2(y)\ldots+\beta_K g_K(y)}}{c(\beta)} \qquad\qquad y \in \mathcal{Y}$$

where

$$g_k(y) = \sum_{i,j} x_{1,ij} y_{ij}, \qquad k = 1, \ldots, K$$

for the covariates $x_{1,ij}, x_{2,ij}, \ldots, x_{K,ij}$

The *log-likelihood* is

$$\begin{aligned} & \ell(\beta; y,\ x) \equiv \log\left[P(Y = y|X = x,\ \beta)\right] \\ = \ & \beta_1 g_1(y) + \beta_2 g_2(y) \ldots + \beta_K g_K(y) \\ - \ & \sum_{i,j} \log[1 + \exp(\sum_{k=1}^{K} \beta_k x_{k,ij})] \end{aligned}$$

Maximum Likelihood Estimation

Our goal is to find the

Maximum likelihood estimator (MLE):
The maximizer $\hat{\beta}$ of the loglikelihood function, $\ell(\beta|y,\ x)$

Computing standard errors for $\hat{\beta}$

- How accurate an estimate is $\hat{\beta}$ of β?

Computing standard errors for $\hat{\beta}$

- How accurate an estimate is $\hat{\beta}$ of β?
  This depends on how “spiked” or “peaked” the likelihood function is.

Computing standard errors for $\hat{\beta}$

- How accurate an estimate is $\hat{\beta}$ of β?
  This depends on how "spiked" or "peaked" the likelihood function is.
    - if the likelihood function is very concentrated around $\hat{\beta}$

Computing standard errors for $\hat{\beta}$

- How accurate an estimate is $\hat{\beta}$ of β?
  This depends on how "spiked" or "peaked" the likelihood function is.
    - if the likelihood function is very concentrated around $\hat{\beta}$
- Based on a Central Limit Theorem, as $n \to \infty$, the distribution of $\hat{\beta}$ approaches a Gaussian with:

Computing standard errors for $\hat{\beta}$

- How accurate an estimate is $\hat{\beta}$ of β?
  This depends on how “spiked” or “peaked” the likelihood function is.
    - if the likelihood function is very concentrated around $\hat{\beta}$
- Based on a Central Limit Theorem, as $n \to \infty$, the distribution of $\hat{\beta}$ approaches a Gaussian with:

$$\mathbb{E}_Y(\hat{\beta}) \to \beta \qquad \text{as } n \to \infty$$

Computing standard errors for $\hat{\beta}$

- How accurate an estimate is $\hat{\beta}$ of β?
  This depends on how “spiked” or “peaked” the likelihood function is.
  - if the likelihood function is very concentrated around $\hat{\beta}$
- Based on a Central Limit Theorem, as $n \to \infty$, the distribution of $\hat{\beta}$ approaches a Gaussian with:

$$\mathbb{E}_Y(\hat{\beta}) \to \beta \qquad \text{as } n \to \infty$$
$$\mathbb{V}_Y(\hat{\beta}) \to I^{-1}(\beta) \qquad \text{as } n \to \infty$$

where

$$I(\beta) = \mathbb{E}_Y\left[-\frac{\partial^2 \ell(\beta; y,\ x)}{\partial \beta^T \partial \beta}(\beta)\right] \tag{1}$$

is the *information matrix* for β

Computing standard errors for $\hat{\beta}$

- How accurate an estimate is $\hat{\beta}$ of β?
  This depends on how "spiked" or "peaked" the likelihood function is.
    - if the likelihood function is very concentrated around $\hat{\beta}$
- Based on a Central Limit Theorem, as $n \to \infty$, the distribution of $\hat{\beta}$ approaches a Gaussian with:

$$\mathbb{E}_Y(\hat{\beta}) \to \beta \qquad \text{as } n \to \infty$$
$$\mathbb{V}_Y(\hat{\beta}) \to I^{-1}(\beta) \qquad \text{as } n \to \infty$$

where

$$I(\beta) = \mathbb{E}_Y\left[-\frac{\partial^2 \ell(\beta; y,\ x)}{\partial \beta^T \partial \beta}(\beta)\right] \tag{1}$$

is the *information matrix* for β

- We can compute $p-$values and confidence intervals

How well does the model fit the network?

- How can we tell if the estimated model is “likely”?

How well does the model fit the network?

- How can we tell if the estimated model is “likely”?
- Compared to what?

How well does the model fit the network?

- How can we tell if the estimated model is “likely”?
- Compared to what?
- The model were all networks are equally likely?

How well does the model fit the network?

- How can we tell if the estimated model is “likely”?
- Compared to what?
- The model were all networks are equally likely?
    - For this model $\beta = 0$
- Use the likelihood and compare $\ell(\beta = \hat{\beta}|y,\ x)$ to $\ell(\beta = 0|y,\ x)$.

How well does the model fit the network?

- How can we tell if the estimated model is “likely”?
- Compared to what?
- The model were all networks are equally likely?
  - For this model $\beta = 0$
- Use the likelihood and compare $\ell(\beta = \hat{\beta}|y,\ x)$ to $\ell(\beta = 0|y,\ x)$.
- For historical reason, use the *deviance*: $\text{deviance}(\beta) = -2\ell(\beta|y,\ x)$

How well does the model fit the network?

- How can we tell if the estimated model is "likely"?
- Compared to what?
- The model were all networks are equally likely?
    - For this model $\beta = 0$
- Use the likelihood and compare $\ell(\beta = \hat{\beta}|y,\ x)$ to $\ell(\beta = 0|y,\ x)$.
- For historical reason, use the *deviance*: $\text{deviance}(\beta) = -2\ell(\beta|y,\ x)$
- In particular $\text{deviance}(\beta = 0) - \text{deviance}(\beta = \hat{\beta})$

Example: Effect on the number of Priorates on centrality

For $k = 1, ..., n$:

$$x_{ij}^{priorates} = \text{Number of priorates for } i + \text{Number of priorates for } j$$

$$x_{ij}^{density} = 1/N$$

Example: Effect on the number of Priorates on centrality

```
Formula:   flomarriage ~ density + nodecov("priorates")

                    Estimate Std. Error p-value
density             -307.82    64.181     <1e-04 ***
nodecov.priorates    0.01643  0.0073    0.0276 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05

    Null  Deviance: 166.355  on 120  degrees of freedom
 Residual Deviance: 102.921  on 118  degrees of freedom
          Deviance:  63.435  on   2  degrees of freedom
```

Quantifying how well the model fits the network?

Refer to deviance($\beta = 0$) as the *null deviance*

Quantifying how well the model fits the network?

Refer to deviance($\beta = 0$) as the *null deviance*
Refer to deviance($\beta = \hat{\beta}$) as the *residual deviance*

Quantifying how well the model fits the network?

Refer to deviance($\beta = 0$) as the *null deviance*
Refer to deviance($\beta = \hat{\beta}$) as the *residual deviance*
Refer to deviance($\beta = 0$) – deviance($\beta = \hat{\beta}$)
as (just) the *deviance*

Quantifying how well the model fits the network?

Refer to deviance($\beta = 0$) as the *null deviance*
Refer to deviance($\beta = \hat{\beta}$) as the *residual deviance*
Refer to deviance($\beta = 0$) – deviance($\beta = \hat{\beta}$)
as (just) the *deviance*
When there are large numbers of actors, the sampling distribution of the deviance

$$\text{deviance}(\beta = 0) - \text{deviance}(\beta = \hat{\beta})$$

has been worked out.

Quantifying how well the model fits the network?

Refer to deviance($\beta = 0$) as the *null deviance*
Refer to deviance($\beta = \hat{\beta}$) as the *residual deviance*
Refer to deviance($\beta = 0$) – deviance($\beta = \hat{\beta}$)
as (just) the *deviance*
When there are large numbers of actors, the sampling distribution of the deviance

$$\text{deviance}(\beta = 0) - \text{deviance}(\beta = \hat{\beta})$$

has been worked out.

It is approximately χ^2 on K degrees of freedom.

Example: Effect on the number of Priorates on centrality

```
Formula:   flomarriage ~ density + nodecov("priorates")

                    Estimate Std. Error p-value
density            -307.82    64.181     <1e-04 ***
nodecov.priorates    0.01643  0.0073    0.0276 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05

    Null  Deviance: 166.355  on 120  degrees of freedom
 Residual Deviance: 102.921  on 118  degrees of freedom
          Deviance:  63.435  on   2  degrees of freedom
```

Does the model fit perfectly?

Compare the *residual deviance* to a χ^2 on $N - K$ degrees of freedom.

Example: Effect on the number of Priorates on centrality

```
Formula:   flomarriage ~ density + nodecov("priorates")

                    Estimate Std. Error p-value
density            -307.82    64.181     <1e-04 ***
nodecov.priorates    0.01643  0.0073    0.0276 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05

    Null  Deviance: 166.355  on 120  degrees of freedom
 Residual Deviance: 102.921  on 118  degrees of freedom
          Deviance:  63.435  on   2  degrees of freedom
```