Approximate Variational Estimation for a Model of Network Formation

Lingjiong Zhu

Florida State University
Email: zhu@math.fsu.edu

Mathematical Finance Colloquium
University of Southern California
April 15, 2019
Joint work with Angelo Mele (Carey Business School, Johns Hopkins University)

References

- (S. Chatterjee and P. Diaconis) Estimating and Understanding Exponential Random Graph Models. (2013). *Annals of Statistics*. **41**, 2428-2461.
- (A. Mele) A structural model of dense network formation. (2017). *Econometrica*. **85**, 825-850.
- (A. Mele and L. Zhu) Approximate variational estimation for a model of network formation. Revision Request at *Review of Economics and Statistics*.

Introduction

- Social interactions and social networks
- Strategic vs Random network formation
- Popular models: Erdős-Rényi, ERGM
- Microeconomic foundations
- Estimation is computationally burdensome

Erdős-Rényi Graph Model

- Given n nodes. Two nodes are linked with probability p.
- Edges are independent of each other. That is, if A and B are friends, B and C are friends, it does not provide any information whether A and C are friends.
- Note that there is no spatial dependence in the Erdős-Rényi graph model.

Exponential Random Graph Model (ERGM)

Probability of observing network g is

$$\pi(g,\theta) = \frac{\exp\left[\sum_{k=1}^{K}\theta_k t_k(g)\right]}{\sum_{\omega\in\mathcal{G}}\exp\left[\sum_{k=1}^{K}\theta_k t_k(\omega)\right]}$$

- θ_k are parameters
- $t_k(g)$ are statistics of the network g

Normalizing constant

$$c(\theta) = \sum_{\omega\in\mathcal{G}}\exp\left[\sum_{k=1}^{K}\theta_k t_k(\omega)\right]$$

Examples

1. Erdős-Rényi Model

$$\pi(g,\theta) = \frac{\exp\left[\theta_1 t_1(g)\right]}{c(\theta)}$$

$t_1(g) = \sum_{i,j} g_{ij} = \#$ *links* (total connectivity)

2. Strauss Model

$$\pi(g,\theta) = \frac{\exp\left[\theta_1 t_1(g) + \theta_2 t_2(g)\right]}{c(\theta)}$$

$t_1(g) = \#$ *links*; $t_2(g) = \sum_{i,j,k} g_{ij} g_{jk} g_{ik} = \#$ *triangles* (friends in common)

Network Formation Model

- Population of n players
- Type (observable) of player i is $\tau_i \in \otimes_{i=1}^{m} \mathcal{X}_i$ (gender, education, income, etc). [1]
- Adjacency matrix g, with entry

$$g_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are linked} \\ 0 & \text{ otherwise} \end{cases}$$

- Undirected network: $g_{ij} = g_{ji}$. (by convention $g_{ii} = 0$ for all i)

[1] E.g. $\{\text{male}, \text{female}\} \times \{\text{low income}, \text{medium income}, \text{high income}\}$

Preferences

- Utility depends on direct connections but also link externalities [2]

$$u_i(g,\tau) = \sum_{j=1}^{n} \alpha(\tau_i,\tau_j) g_{ij} + \frac{\beta}{n} \sum_{j=1}^{n} \sum_{k=1}^{n} g_{ij} g_{jk} + \frac{\gamma}{n} \sum_{j=1}^{n} \sum_{k=1}^{n} g_{ij} g_{jk} g_{ki},$$

where $\alpha : \otimes_{i=1}^{m} \mathcal{X}_i \times \otimes_{i=1}^{m} \mathcal{X}_i \to \mathbb{R}$ and $\beta \in \mathbb{R}$.

- Other externalities, e.g. any finite subgraph
- Heterogeneous externalities, e.g. $\beta(\tau_i,\tau_j)$, $\gamma(\tau_i,\tau_j)$ or $\beta(\tau_i,\tau_j,\tau_k)$, $\gamma(\tau_i,\tau_j,\tau_k)$, more technically involved

[2] An externality is the cost or benefit that affects a party who did not choose to incur that cost or benefit.

Preferences: Interpretation

- $\alpha(\tau_i, \tau_j)$ differentiates the likelihood of forming a link between i and j depending on the types of players i and j, e.g. race, gender, age etc.
- Note that $\sum_{j=1}^{n}\sum_{k=1}^{n} g_{ij}g_{jk} = \sum_{j=1}^{n} g_{ij}\sum_{k=1}^{n} g_{jk}$. For individual i, when he forms a link to j, he also considers how many friends j has: $\sum_k g_{jk}$.
- i may be interested in linking popular kids, so the effect of j having many friends will be positive; or i could be afraid that since j has many friends he will not have time to spend with i so that in that case it will be a negative effect.
- Also note that $\sum_{j=1}^{n}\sum_{k=1}^{n} g_{ij}g_{jk}g_{ki} = \sum_{j=1}^{n} g_{ij}\sum_k g_{jk}g_{ki}$, where $\sum_k g_{jk}g_{ki}$ denotes the number of mutual friends between i and j.

Preferences: Example of Homophily

Two finite types (e.g. gender) and homophily [3]

$$\alpha_{ij} := \alpha(\tau_i, \tau_j) = V - c(\tau_i, \tau_j)$$

Cost of direct links is:

$$c(\tau_i, \tau_j) = \begin{cases} c & \text{if } \tau_i = \tau_j \\ C & \text{if } \tau_i \neq \tau_j \end{cases}$$

[3] Homophily is the tendency of individuals to associate and bond with similar others. The opposite of homophly is heterophily.

Network Formation Model

Equilibrium: A network g is **pairwise stable with transfers** if:

1. $g_{ij} = 1 \Rightarrow u_i(g,\tau) + u_j(g,\tau) \geq u_i(g - ij,\tau) + u_j(g - ij,\tau)$;
2. $g_{ij} = 0 \Rightarrow u_i(g,\tau) + u_j(g,\tau) \geq u_i(g + ij,\tau) + u_j(g + ij,\tau)$;

- $g + ij$: network g with the addition of link g_{ij};
- $g - ij$: network g without link g_{ij}.

Assumptions

Sequential network formation

- In each period t a pair of individuals meet with probability $\rho_{ij} > 0$
- Upon meeting, they decide whether to form a link by maximizing the sum of their utility
- Agents are myopic

Assumption 1. The meeting process does not depend on the network, and $\rho_{ij} > 0$ for all *ij* and i.i.d. over time

Assumption 2. Individuals receive a *logistic matching shock* before they decide whether to form a link (i.i.d. over time and players)

Equilibrium Characterization

Proposition

There exists a potential function $Q_n(g;\alpha,\beta)$ that characterizes the incentives of all the players in any state of the network

$$Q_n(g;\alpha,\beta) = \sum_{i,j} \alpha_{ij} g_{ij} + \frac{\beta}{2n} \sum_{i,j,k} g_{ij} g_{jk} + \frac{\gamma}{6n} \sum_{i,j,k} g_{ij} g_{jk} g_{ki}. \qquad (1)$$

Butts (2009), Mele (2017), Badev (2013), Chandrasekhar and Jackson (2014)

Intuition: For any g_{ij}

$$\begin{aligned} & Q_n(g;\tau) - Q_n(g - ij;\tau) \\ & = u_i(g,\tau) + u_j(g,\tau) - [u_i(g-ij,\tau) + u_j(g-ij,\tau)], \end{aligned}$$

and thus Q_n by definition is the potential function. Pairwise stable (with transfers) networks $\Longleftrightarrow$ local maxima of Q_n

Long-run Convergence

Theorem

In the long run, the model converges to the stationary dist. π_n*:*

$$\pi_n(g;\alpha,\beta) = \frac{\exp\left[Q_n(g;\alpha,\beta)\right]}{\sum_{\omega\in\mathcal{G}}\exp\left[Q_n(\omega;\alpha,\beta)\right]} = \exp\left\{n^2\left[T_n(g;\alpha,\beta) - \psi_n(\alpha,\beta)\right]\right\},$$

where

$$T_n(g;\alpha,\beta) = \frac{1}{n^2}Q_n(g;\alpha,\beta)$$

$$\psi_n(\alpha,\beta) = \frac{1}{n^2}\log\sum_{\omega\in\mathcal{G}}\exp\left[n^2 T_n(\omega;\alpha,\beta)\right], \tag{2}$$

Problem: $\mathcal{G}$ contains $2^{\binom{n}{2}}$ networks!

For $n = 20$, there are $2^{190} \approx 1.569275 \times 10^{57}$ networks

To show this, we only need to check the detailed balance condition since the network formation process is a Markov chain. That is, we need to show that

$$P_{gg'}\pi_g = P_{g'g}\pi_{g'}, \tag{3}$$

where

$$P_{gg'} = \mathbb{P}(G_{t+1} = g' | G_t = g), \tag{4}$$

and

$$\pi_g = \pi(G_t = g), \tag{5}$$

where π is the stationary distribution that we will show is given by the ERGM probability distribution.

Let $g = (1, g_{-ij})$ and $g' = (0, g_{-ij})$. Note that

$$\mathbb{P}(G_{ij} = 0 | G_{-ij} = g_{-ij}) = \frac{1}{1 + e^{\Delta Q}}, \tag{6}$$

since the shocks are logistic. We can compute that

$$\begin{aligned}
P_{gg'}\pi_g &= \mathbb{P}(m_t = ij)\mathbb{P}(G_{ij} = 0 | G_{-ij} = g_{-ij}) \frac{e^{Q(1,g_{-ij})}}{\sum_g e^{Q(g)}} \\
&= \mathbb{P}(m_t = ij) \frac{1}{1 + e^{\Delta Q}} \frac{e^{Q(1,g_{-ij})}}{\sum_g e^{Q(g)}} \\
&= \mathbb{P}(m_t = ij) \frac{e^{\Delta Q}}{1 + e^{\Delta Q}} \frac{e^{Q(0,g_{-ij})}}{\sum_g e^{Q(g)}} \\
&= \mathbb{P}(m_t = ij)\mathbb{P}(G_{ij} = 1 | G_{-ij} = g_{-ij}) \frac{e^{Q(0,g_{-ij})}}{\sum_g e^{Q(g)}} = P_{g'g}\pi_{g'}.
\end{aligned}$$

MC-MLE

Lemma

Fix vectors (α_1, β_1) *and* (α_2, β_2). *Then*

$$\frac{e^{n^2\psi_n(\alpha_1,\beta_1)}}{e^{n^2\psi_n(\alpha_2,\beta_2)}} = \mathbb{E}_{\alpha_2,\beta_2} e^{n^2[T_n(\omega;\alpha_1,\beta_1)-T_n(\omega;\alpha_2,\beta_2)]} \tag{7}$$

where $\mathbb{E}_{\alpha_2,\beta_2}$ *is the expectation computed according to* $\pi_n(g, \alpha_2, \beta_2)$

$\Rightarrow$ Estimate the ratio of constants using Monte Carlo

Simulate R networks $g^{(1)}, ..., g^{(R)}$ from $\pi(\cdot, \alpha_2, \beta_2)$

$$R_{\alpha_2,\beta_2}(\alpha_1, \beta_1) = \frac{1}{R}\sum_{r=1}^{R} \exp\left\{n^2\left[T_n\left(g^{(r)}; \alpha_1, \beta_1\right) - T_n\left(g^{(r)}; \alpha_2, \beta_2\right)\right]\right\}$$

$$\to e^{n^2\psi_n(\alpha_1,\beta_1)-n^2\psi_n(\alpha_2,\beta_2)}$$

MC-MLE

Let $\theta \equiv (\alpha, \beta)$.

Find θ_{mle} by maximization of log-likelihood

$$\begin{aligned} \theta_{mle} &= \arg\max_{\theta} \ell(\theta) \\ &= \arg\max_{\theta} \{\ell(\theta) - \mathit{constant}\} \\ &= \arg\max_{\theta} \{\ell(\theta) - \ell(\theta_0)\} \end{aligned}$$

If you subtract a constant, the maximizer does not change

MC-MLE

$$
\begin{aligned}
&\ell(\theta) - \ell(\theta_0) \\
&= n^2 T_n(g, \theta) - n^2 \psi_n(\theta) - n^2 T_n(g, \theta_0) + n^2 \psi_n(\theta_0) \\
&= n^2 \{[T_n(\theta) - T_n(\theta_0)] - [\psi_n(\theta) - \psi_n(\theta_0)]\} \\
&= n^2 [T_n(\theta) - T_n(\theta_0)] - n^2 [\psi_n(\theta) - \psi_n(\theta_0)] \\
&= n^2 [T_n(\theta) - T_n(\theta_0)] - \log \mathbb{E}_{\theta_0} \exp\left\{ n^2 [T_n(\omega; \theta) - T_n(\omega; \theta_0)] \right\}
\end{aligned}
$$

Using Lemma above

$$
\begin{aligned}
\ell(\theta) - \ell(\theta_0) \quad &\approx \quad n^2 [T_n(g; \theta) - T_n(g; \theta_0)] \\
&- \quad \log \frac{1}{R} \sum_{r=1}^{R} \exp\left\{ n^2 \left[T_n(g^{(r)}; \theta) - T_n(g^{(r)}; \theta_0) \right] \right\}
\end{aligned}
$$

MC-MLE

Therefore MC-MLE estimate is

$$\theta_{mcmle} = \arg\max_{\theta} \Bigg\{ n^2 \left[T_n(g;\theta) - T_n(g;\theta_0) \right] - \log \frac{1}{R} \sum_{r=1}^{R} \exp\left\{ n^2 \left[T_n(g^{(r)};\theta) - T_n(g^{(r)};\theta_0) \right] \right\} \Bigg\}$$

Geyer and Thompson (1992) show that as $R \to \infty$

$$\theta_{mcmle} \to \theta_{mle}$$

Variational Inference

Find approximate likelihood $q_n(g)$ to minimize

$$\begin{aligned} KL(q_n|\pi_n) &= \sum_{\omega\in\mathcal{G}} q_n(\omega)\log\left[\frac{q_n(\omega)}{\pi_n(\omega;\alpha,\beta)}\right] \\ &= \mathbb{E}_{q_n}[\log q_n(\omega)] - n^2\mathbb{E}_{q_n}[T_n(\omega;\alpha,\beta)] + n^2\psi_n(\alpha,\beta) \geq 0 \end{aligned}$$

With some algebra we obtain

$$\psi_n(\alpha,\beta) \geq \mathbb{E}_{q_n}[T_n(\omega;\alpha,\beta)] + \frac{1}{n^2}\mathcal{H}(q_n) = \mathcal{L}(q_n),$$

where $\mathcal{H}(q_n)$ = entropy of q_n.

Variational Inference

Therefore the best approximating distribution q_n is the solution of

$$\psi_n(\alpha,\beta) = \sup_{q_n \in \mathcal{Q}_n} \mathcal{L}(q_n) = \sup_{q_n \in \mathcal{Q}_n} \left\{ \mathbb{E}_{q_n}\left[T_n(\omega;\alpha,\beta)\right] + \frac{1}{n^2}\mathcal{H}(q_n) \right\}. \tag{8}$$

- In general no closed-form solution
- In practice we restrict the family $\mathcal{Q}_n$ to tractable distributions

Mean-Field Approximation

Consider only completely factorized q_n

$$q_n(g) = \prod_{i,j} \mu_{ij}^{g_{ij}} (1 - \mu_{ij})^{1-g_{ij}},$$
$$\mu_{ij} = \mathbb{E}_{q_n}(g_{ij}) = \mathbb{P}_{q_n}(g_{ij} = 1)$$

Therefore we get

$$\frac{1}{n^2}\mathcal{H}(q_n) = -\frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\mu_{ij}\log\mu_{ij} + (1-\mu_{ij})\log(1-\mu_{ij})\right],$$

$$\mathbb{E}_{q_n}\left[T_n(\omega;\alpha,\beta)\right] = \frac{\sum_{i,j}\alpha_{ij}\mu_{ij}}{n^2} + \beta\frac{\sum_{i,j,k}\mu_{ij}\mu_{jk}}{2n^3} + \gamma\frac{\sum_{i,j,k}\mu_{ij}\mu_{jk}\mu_{ki}}{6n^3}.$$

Mean-Field Approximation

The maximization problem is now to find a matrix $\mu(\alpha, \beta, \gamma)$

$$\psi_n(\alpha, \beta, \gamma) \geq \psi_n^{MF}(\mu(\alpha, \beta, \gamma))$$

$$:= \sup_{\mu \in [0,1]^{n^2}} \left\{ \frac{\sum_{i,j} \alpha_{ij}\mu_{ij}}{n^2} + \beta \frac{\sum_{i,j,k} \mu_{ij}\mu_{jk}}{2n^3} + \gamma \frac{\sum_{i,j,k} \mu_{ij}\mu_{jk}\mu_{ki}}{6n^3} - \frac{1}{2n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} [\mu_{ij} \log \mu_{ij} + (1 - \mu_{ij}) \log(1 - \mu_{ij})] \right\}.$$

- Take the first order derivatives w.r.t. each μ_{ij} and set it zero.
- We initialize the matrix μ, and iterate, till it converges to a local maximum.

Approximation bounds

Theorem

For fixed n, the approximation error is

$$C_3(\beta,\gamma)n^{-1} \leq \psi_n(\alpha,\beta,\gamma) - \psi_n^{MF}(\mu(\alpha,\beta,\gamma))$$
$$\leq C_1(\alpha,\beta,\gamma)n^{-1/5}(\log n)^{1/5} + C_2(\alpha,\beta,\gamma)n^{-1/2},$$

where $C_1(\alpha,\beta,\gamma)$, $C_2(\alpha,\beta,\gamma)$ are constants depending only on α,β and γ and $C_3(\beta,\gamma)$ are constants depending only on β,γ.

The proof is based on the nonlinear large deviations (Chatterjee-Dembo 2014).

Proposition

Assume (α, β, γ) *lives on a compact set* Θ*. Let* $\hat{\theta}_n := (\hat{\alpha}_n, \hat{\beta}_n, \hat{\gamma}_n)$ *and* $\hat{\theta}_n^{MF} := (\hat{\alpha}_n^{MF}, \hat{\beta}_n^{MF}, \hat{\gamma}_n^{MF})$ *be the maximizers of* ℓ_n *and* ℓ_n^{MF}*, respectively, in the interior of* Θ*. Also assume that* ψ_n *and* ψ_n^{MF} *are differentiable and* μ_n*- and* μ_n^{MF}*-strongly convex in* (α, β, γ)*, respectively, on* Θ*, where* $\mu_n > 0$ *and* $\mu_n^{MF} > 0$*. Then*

$$\|\hat{\theta}_n - \hat{\theta}_n^{MF}\| \leq \frac{2}{(\mu_n + \mu_n^{MF})^{\frac{1}{2}}} \left[\sup_{\alpha,\beta,\gamma\in\Theta} C_1^{\frac{1}{2}}(\alpha, \beta, \gamma) \left(\frac{\log n}{n}\right)^{\frac{1}{10}} + \sup_{\alpha,\beta,\gamma\in\Theta} C_2^{\frac{1}{2}}(\alpha, \beta, \gamma) n^{-\frac{1}{4}} \right],$$

where C_1 *and* C_2 *are defined as before, and* $\|\cdot\|$ *denotes the Euclidean norm.*

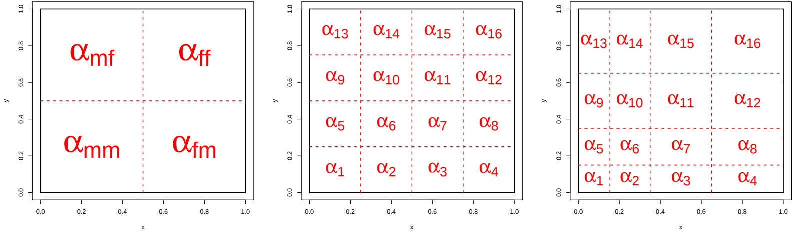
Asymptotics

- Previous theorem gives results for a fixed n
- What happens when $n \to \infty$?
- Graph limits literature and large deviations Lovasz (2012), Borgs et al (2006), (2008), Chatterjee-Diaconis (2011), Chatterjee-Varadhan (2010), Radin-Yin (2011), Aristoff-Zhu (2014)
- When $n \to \infty$ consider a *continuum of nodes* on $[0, 1]$
- Adj. matrix g is replaced by a function, known as *graphon*, $h : [0, 1]^2 \to [0, 1]$

Assumption (Spatial ERGM). Assume that

$$\alpha_{ij} = \alpha\left(i/n, j/n\right), \tag{9}$$

where $\alpha(x, y) : [0, 1]^2 \to \mathbb{R}$, and $\alpha(x, y) = \alpha(y, x)$,



Convergence of Mean-Field Approximation

Proposition

Mean-field converges to exact normalizing constant in large networks, i.e., as $n \to \infty$

$$\psi_n^{MF}(\mu(\alpha,\beta,\gamma)) \to \psi(\alpha,\beta,\gamma).$$

As a corollary, we have $\psi_n(\alpha,\beta,\gamma) \to \psi(\alpha,\beta,\gamma)$, where

$$\psi(\alpha,\beta) := \sup_h \left\{ \int_{[0,1]^2} \alpha(x,y)h(x,y)dxdy + \frac{\beta}{2}\int_{[0,1]^3} h(x,y)h(y,z)dxdydz, \right.$$
$$\left. + \frac{\gamma}{6}\int_{[0,1]^3} h(x,y)h(y,z)h(z,x)dxdydz - \frac{1}{2}\int_{[0,1]^2} I(h(x,y))dxdy \right\},$$

where $I(x) := x\log(x) + (1-x)\log(1-x)$, and the supremum is over symmetric functions $h : [0,1]^2 \to [0,1]$.

Variational problem: homogeneous model

Theorem (Chatterjee-Diaconis 2013)
If $\mathcal{T} : \mathcal{W} \to \mathbb{R}$ is a bounded continuous function, then

$$\psi(\alpha,\beta) \equiv \lim_{n\to\infty} \psi_n(\alpha,\beta) = \sup_{h\in\mathcal{W}} \{\mathcal{T}(h) - \mathcal{I}(h)\}$$

If $\alpha(x,y) = \alpha$ for all x, y

$$\mathcal{T}(h) \equiv \alpha \int_{[0,1]^2} h(x,y)dxdy + \frac{\beta}{2} \int_{[0,1]^3} h(x,y)h(y,z)dxdydz + \frac{\gamma}{6} \int_{[0,1]^3} h(x,y)h(y,z)h(z,x)dxdydz,$$

$$\mathcal{I}(h) \equiv \frac{1}{2} \int_0^1 \int_0^1 I(h(x,y))dxdy,$$

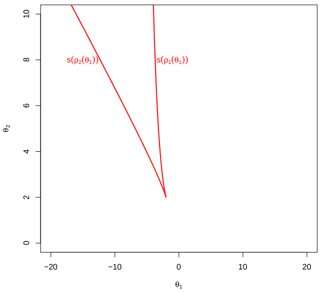
where $I(x) := x\log(x) + (1-x)\log(1-x)$.
The proof is based on the large deviations for Erdős-Rényi graph (Chatterjee-Varadhan 2010).

Variational Problem: Special Cases

Theorem (homogeneous case)Let $\mathcal{T}$ be defined as above and $\gamma = 0$. Then $h(x, y) = \mu$ a.e

$$\lim_{n\to\infty} \psi_n(\alpha, \beta, 0) = \psi(\alpha, \beta, 0) = \sup_{\mu\in[0,1]} \left\{ \alpha\mu + \frac{\beta}{2}\mu^2 - \frac{1}{2}I(\mu) \right\}$$

1. Outside V-shaped region: *unique maximizer* μ^*
2. Inside V-shaped region: *two local maximizers* $\mu_1^* < \frac{1}{2} < \mu_2^*$
3. V-shaped region: there is $\beta = q(\alpha)$, such that $\ell(\mu_1^*) = \ell(\mu_2^*)$



Variational problem: homogeneous model

We recall the variational problem:

$$\psi(\alpha,\beta,\gamma)$$
$$:= \sup_h \left\{ \int_{[0,1]^2} \alpha(x,y)h(x,y)dxdy + \frac{\beta}{2}\int_{[0,1]^3} h(x,y)h(y,z)dxdydz, \right.$$
$$\left. + \frac{\gamma}{6}\int_{[0,1]^3} h(x,y)h(y,z)h(z,x)dxdydz - \frac{1}{2}\int_{[0,1]^2} I(h(x,y))dxdy \right\},$$

where $I(x) := x\log(x) + (1-x)\log(1-x)$, and the supremum is over symmetric functions $h : [0,1]^2 \to [0,1]$.

Proposition

Assume that $\alpha(x,y)$ takes two values:

$$\alpha(x,y)=\begin{cases}\alpha_1, & \text{if } 0<x,y<\frac{1}{2} \text{ or } \frac{1}{2}<x,y<1,\\ \alpha_2, & \text{if } 0<x<\frac{1}{2}<y<1 \text{ or } 0<y<\frac{1}{2}<x<1.\end{cases} \tag{10}$$

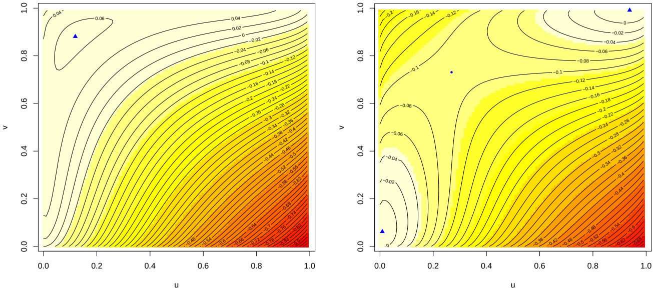
Then, we have

$$\psi(\alpha,\beta,0)=\sup_{0\leq u,v\leq 1}F(u,v), \tag{11}$$

where

$$F(u,v):=\frac{\alpha_1}{2}u-\frac{1}{4}I(u)+\frac{\alpha_2}{2}v-\frac{1}{4}I(v)+\frac{\beta}{8}(u+v)^2. \tag{12}$$

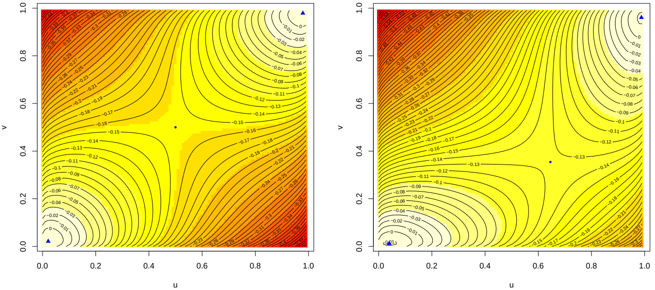
Figure: Examples of maxima with $\beta = 4$



(A) $(\alpha_1, \alpha_2) = (-3, -1)$; (B) $(\alpha_1, \alpha_2) = (-2.5, -1.5)$

The figures show the level curves of $F(u, v)$ for different vectors of parameters. In all the pictures $\beta = 4$. The global maxima are represented as blue triangles.

Figure: Examples of maxima with $\beta = 4$



(C) $(\alpha_1, \alpha_2) = (-2, -2)$; (D)$(\alpha_1, \alpha_2) = (-1.7, -2.3)$

The figures show the level curves of $F(u, v)$ for different vectors of parameters. In all the pictures $\beta = 4$. The global maxima are represented as blue triangles.

Proposition

The specification of the model in (11) *has the following properties*

1. *The solution* (u^*, v^*) *satisfies:*

$$u^* = \frac{e^{2\alpha_1 + \beta(u^* + v^*)}}{1 + e^{2\alpha_1 + \beta(u^* + v^*)}}, \qquad v^* = \frac{e^{2\alpha_2 + \beta(u^* + v^*)}}{1 + e^{2\alpha_2 + \beta(u^* + v^*)}}.$$

2. *If* $\beta \leq 2$, *the maximization problem (11) has a unique solution* (u^*, v^*).

3. *If* $\alpha_1 + \alpha_2 + \beta = 0$ *and* $\beta > \frac{(1+e^{\alpha_1-\alpha_2})^2}{2e^{\alpha_1-\alpha_2}}$ *then the maximization problem (11) has two solutions* (u^*, v^*) *and* $(1 - v^*, 1 - u^*)$ *with* $u^* < \frac{1}{2}$ *and* $v^* < \frac{1}{2}$, *and* $F(u^*, v^*) = F(1 - v^*, 1 - u^*)$. *(Phase transition!)*

Numerical Studies

- We consider two groups of equal size with $\tilde{\alpha}_1 := \alpha_2$ and $\tilde{\alpha}_2 := \alpha_1 - \alpha_2$.
- We have performed simple Monte Carlo experiments to study the performance of our asymptotic approximation in finite networks. We compare the mean-field approximation with the standard simulation-based MCMC-MLE (Geyer and Thompson, 1992).
- We test our approximation technique using artificial network data. Each network is generated using a 10 million run of the Metropolis-Hastings sampler implemented in the `ergm` command in R.
- We report results for networks with 50, 100 and 200 nodes. The results are summarized by the median and several percentiles of the estimated parameters.

Table: Monte Carlo estimates, comparison of three methods. True parameter vector is $(\tilde{\alpha}_1, \tilde{\alpha}_2, \beta, \gamma) = (-2, 1, -1, -1)$

$n = 50$	MCMC-MLE				MEAN-FIELD			
	$\tilde{\alpha}_1$	$\tilde{\alpha}_2$	β	γ	$\tilde{\alpha}_1$	$\tilde{\alpha}_2$	β	γ
median	-1.975	1.054	-1.805	-9.302	-2.016	0.995	-1.000	-1.001
0.05	-2.611	0.598	-9.666	-75.431	-3.879	0.914	-1.269	-1.146
0.95	-1.489	1.397	7.348	58.765	-1.926	4.276	-0.840	-0.901
$n = 100$	MCMC-MLE				MEAN-FIELD			
	$\tilde{\alpha}_1$	$\tilde{\alpha}_2$	β	γ	$\tilde{\alpha}_1$	$\tilde{\alpha}_2$	β	γ
median	-2.035	1.012	-0.765	-2.199	-2.021	0.988	-0.998	-1.001
0.05	-2.312	0.730	-4.937	-52.429	-2.080	0.945	-1.031	-1.031
0.95	-1.662	1.218	3.555	32.974	-1.978	1.139	-0.950	-0.939
$n = 200$	MCMC-MLE				MEAN-FIELD			
	$\tilde{\alpha}_1$	$\tilde{\alpha}_2$	β	γ	$\tilde{\alpha}_1$	$\tilde{\alpha}_2$	β	γ
median	-1.980	1.004	-1.734	-4.100	-2.029	0.988	-0.996	-0.999
0.05	-2.212	0.876	-4.710	-27.070	-2.060	0.968	-1.002	-1.010
0.95	-1.779	1.112	2.792	31.735	-2.005	1.028	-0.969	-0.987

Table: Monte Carlo estimates and comparisons. True parameter vector is $(\tilde{\alpha}_1, \tilde{\alpha}_2, \beta, \gamma) = (-2, 1, -2, 3)$

$n = 50$	MCMC-MLE				MEAN-FIELD			
	$\tilde{\alpha}_1$	$\tilde{\alpha}_2$	β	γ	$\tilde{\alpha}_1$	$\tilde{\alpha}_2$	β	γ
median	-2.033	0.972	-2.239	-6.134	-2.037	0.990	-2.000	3.000
0.05	-2.643	0.614	-10.317	-73.906	-2.212	0.856	-2.652	2.865
0.95	-1.424	1.399	6.763	68.994	-1.887	1.351	-1.875	3.314
$n = 100$	MCMC-MLE				MEAN-FIELD			
	$\tilde{\alpha}_1$	$\tilde{\alpha}_2$	β	γ	$\tilde{\alpha}_1$	$\tilde{\alpha}_2$	β	γ
median	-1.975	0.983	-2.364	3.014	-2.040	0.970	-2.000	3.000
0.05	-2.307	0.779	-7.526	-41.294	-2.108	0.908	-2.044	2.950
0.95	-1.689	1.232	2.959	48.968	-1.995	1.048	-1.939	3.049
$n = 200$	MCMC-MLE				MEAN-FIELD			
	$\tilde{\alpha}_1$	$\tilde{\alpha}_2$	β	γ	$\tilde{\alpha}_1$	$\tilde{\alpha}_2$	β	γ
median	-2.019	1.004	-1.869	7.701	-2.049	0.976	-1.997	2.999
0.05	-2.267	0.890	-6.331	-34.052	-2.113	0.948	-2.020	2.970
0.95	-1.738	1.116	2.277	37.341	-2.017	1.071	-1.953	3.029

Numerical Studies: Conclusions

- While both methods seem to work well, the mean-field approximation gives more robust estimates.
- At the same time, as it is well known, the mean-field can be biased.
- The computational complexity of the mean-field approximation is of order n^2, while it is well known that the simulation methods used in the MCMC-MLE may have complexity of order e^{n^2} for some parameter vector (Bhamidi et al. 2011, Chatterjee and Diaconis 2013, Mele 2017).

Thank You!