Fast Maximum Likelihood estimation via Equilibrium Expectation for Large Network Data

Maksym Byshkin[1], Alex Stivala[4,1], Antonietta Mira[1,3], Garry Robins[2], Alessandro Lomi[1,2]

[1]*Università della Svizzera italiana, Lugano, Switzerland*

[2]*University of Melbourne, Australia*

[3]*Università dell'Insubria, Como, Italy*

[4]Swinburne University of Technology, Melbourne, Australia

Sunbelt, June 26-July1st, 2018, Utrecht, Netherlands

Introduction

We have developed efficient Monte Carlo method to perform Maximum Likelihood parameter estimation for probability distributions with intractable normalizing constants

ERGMs are exponential family of probability distributions for dependent network data

$$\pi(x, \boldsymbol{\theta}) = \frac{\exp(\boldsymbol{\theta} \cdot \mathbf{z}(x))}{k(\boldsymbol{\theta})}$$

Intractable normalizing constant

$$k(\boldsymbol{\theta}) = \sum_x \exp(\boldsymbol{\theta} \cdot \mathbf{z}(x))$$

$$\mathbf{z}(x) = (z_1(x), z_2(x), .., z_S(x))^T$$

$z_i(x)$ is a networks statistic (e.g. number of ties, triangles, stars ..)

Robins G, Snijders T, Wang P, Handcock M, & Pattison P (2007) Recent developments in exponential random graph (p*) models for social networks. *Social Networks* 29(2):192-215.

Maximum Likelihood parameter estimation

Given x_{obs}, find $\boldsymbol{\theta}\ (x_{obs})$ such that

$$E_{\pi(\boldsymbol{\theta})}(z_i(x)) = z_i(x_{obs}) \ \ \forall i$$

Expectations may be computed by **converged** MCMC simulations

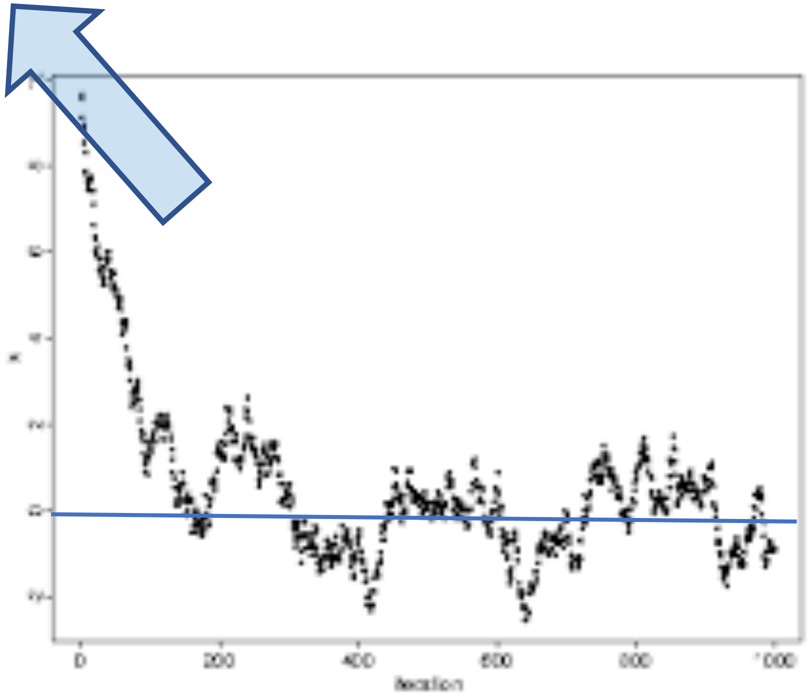
$$E_{\pi(\boldsymbol{\theta})}(z_i(x)) = \sum_x z_i(x) \cdot \pi(x, \boldsymbol{\theta})$$

Stability of statistics at Equilibrium

Theorem

Let a transition probability $P(x \rightarrow x', \boldsymbol{\theta})$ define a Markov chain with a unique stationary distribution. If this distribution is a statistical model from an exponential family $\pi(x,\boldsymbol{\theta})$ and the Markov process has reached its stationary distribution then for any $z_i(x)$

$$\sum_{x,x'} \pi(x,\boldsymbol{\theta}) P(x \rightarrow x', \boldsymbol{\theta}) \left(z_i(x') - z_i(x)\right) = 0$$



$$E_{\pi(\boldsymbol{\theta})}(z_i(x)) = z_i(x_{obs})$$

Byshkin, et al. "Fast Maximum Likelihood estimation via Equilibrium Expectation for Large Network Data" under review in Scientific Reports, *arXiv preprint arXiv:1802.10311* (2018).

Monte Carlo Maximum Likelihood estimation

Theorem

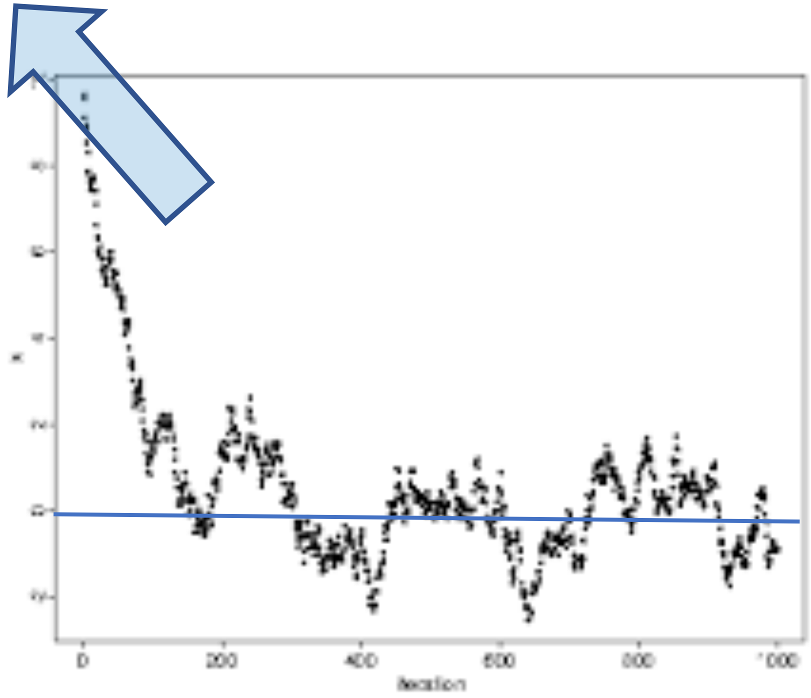
Let a transition probability $P(x \to x', \boldsymbol{\theta})$ define a Markov chain with a unique stationary distribution. If this distribution is a statistical model from an exponential family $\pi(x,\boldsymbol{\theta})$ and the Markov process has reached its stationary distribution then for any $z_i(x)$

$$\sum_{x,x'} \pi(x,\boldsymbol{\theta}) P(x \to x', \boldsymbol{\theta}) \left(z_i(x') - z_i(x)\right) = 0$$

⇧⇩ If MLE exists and is unique

$$E_{\pi(\boldsymbol{\theta})}(z_i(x)) = z_i(x_{obs})$$

s equations with *s* unknowns



MCMC MLE may be computed without MCMC simulation!

Monte Carlo Maximum Likelihood estimation

$$\sum_{x,x'} \pi(x,\mathbf{\theta})P(x \to x',\mathbf{\theta})\,(z_i(x')-z_i(x))=0$$

⇕

$$E_{\pi(\mathbf{\theta})}(dz_i(x,\mathbf{\theta}))=0 \qquad dz_i(x,\mathbf{\theta}) = \sum_{x'} P(x \to x',\mathbf{\theta})\,(z_i(x')-z_i(x))$$

$dz_i(x,\mathbf{\theta})$ may be computed by Monte Carlo integration

If we have data samples i.i.d. from $\pi(x,\mathbf{\theta})$: $x_{S_1}, x_{S_2}, .., x_{S_n}$

$$E_{\pi(\mathbf{\theta})}(dz_i(x,\mathbf{\theta}))=\frac{1}{n}\sum_{j=1}^{n} dz_i(x_{s_j},\mathbf{\theta})$$

may be computed by Monte Carlo integration

Connection to Contrastive Divergence

The theorem

$$E_{\pi(\boldsymbol{\theta})}[\sum_{x'} P(x \to x', \boldsymbol{\theta})\,(z_i(x') - z_i(x))] = 0$$ Unbiased

Contrastive divergence (CD-1):

$$\sum_{x'} P(x \to x', \boldsymbol{\theta})\,(z_i(x') - z_i(x)) = 0$$ Biased for real-world data

$$\sum_{x'} P(x \to x', \boldsymbol{\theta})\,(z_i(x') - z_i(x)) = 0$$ Unbiased for samples of $\pi(\mathrm{x}, \boldsymbol{\theta})$

Hinton, Geoffrey E. "Training products of experts by minimizing contrastive divergence." *Neural computation* 1171 (2002)
Fellows, Ian E. "Why (and When and How) Contrastive Divergence Works." *arXiv preprint arXiv:1405.0602* (2014).

Monotonic dependence of statistics on parameters

How to find MLE for real-world data x_{obs} and data approximating models?
We can simulate data so that $z_i(x) \approx z_i(x_{\text{obs}})$

We can make it because of monotonic dependence of statistics on parameters $\boldsymbol{\theta}$

$$\partial E_{\pi(\boldsymbol{\theta})}(z_i(x))/\partial\theta_i > 0$$

Geyer CJ & Thompson EA, Constrained Monte Carlo Maximum Likelihood for dependent data, Journal of the Royal Statistical Society (1992)

$$\partial dz_i(x, \boldsymbol{\theta})/\partial\theta_i > 0$$

Byshkin, et al. "Fast Maximum Likelihood estimation via Equilibrium Expectation for Large Network Data", under review in Scientific Reports, *arXiv preprint arXiv:1802.10311* (2018).

I) Find a starting point $\boldsymbol{\theta}_0$ so that $dz_i(x_{\text{obs}}, \boldsymbol{\theta}_0)=0\ \ \forall i$

II) Starting from x_{obs} make MCMC by adapting $\boldsymbol{\theta}$ so that $dz_i(x, \boldsymbol{\theta}) \approx 0$ and $z_i(x) \approx z_i(x_{\text{obs}})\ \forall i$

Equilibrium Expectation algorithm for MLE

1. Initialization: t=0; $x = x_{obs}$;
2. Take approximate solution of $dz_i(x_{obs}, \boldsymbol{\theta}_{CD}) = 0$ as a starting point: $\boldsymbol{\theta}(t = 0) = \boldsymbol{\theta}_{CD}$
3. Initialize vector **K:** for i=1 *to s*: $K_i = K0 \cdot \partial(dz_i(x_{obs}, \boldsymbol{\theta}))/\partial\theta_i)^{-2}$ (or any small positive constant)
4. for k=1 to *m2*

   Perform *m* steps of Metropolis-Hastings algorithm on Markov chain *x*

   for i=1 to *s*: that $\theta_i(t+1) = \theta_i(t) - K_i \cdot sgn(z_i(x) - z_i(x_{obs})) \cdot (z_i(x) - z_i(x_{obs}))^2$.

   $t = t + 1$.

   end for
5. Adapt K_i so that $SD(\theta_i(t)) \approx c2 \cdot \max(|\overline{\theta_i(t)}|, c1)$ $\forall i$
6. If t<M then go to step 4.

m=10^3 (from 10^2 to 10^4),

m2=10^2 (from 50 to 10^4),

c2=10^{-3} (from 10^{-5} to 0.1),

c1=10^{-2} (or any small positive constant)

- **Markov chain is not reset between parameters updates**
- **When *c2*=0 the EE algorithm does not differ from the Metropolis-Hastings algorithm**

Byshkin, et al. "Fast Maximum Likelihood estimation via Equilibrium Expectation for Large Network Data", under review in Scientific Reports, *arXiv preprint arXiv:1802.10311* (2018).

Equilibrium Expectation algorithm for MLE

1. Initialization: t=0; $x = x_{obs}$;
2. Take approximate solution of $dz_i(x_{obs}, \boldsymbol{\theta}_{CD}) = 0$ as a starting point: $\boldsymbol{\theta}(t=0) = \boldsymbol{\theta}_{CD}$
3. Initialize vector **K:** for i=1 *to s*: $K_i = K0 \cdot \partial(dz_i(x_{obs}, \boldsymbol{\theta}))/\partial\theta_i)^{-2}$ (or any small positive constant)
4. for k=1 to *m2*

   Perform *m* steps of Metropolis-Hastings algorithm on Markov chain *x*

   for i=1 to *s*: that $\theta_i(t+1) = \theta_i(t) - K_i \cdot sgn(z_i(x) - z_i(x_{obs})) \cdot (z_i(x) - z_i(x_{obs}))^2$.

   $t = t + 1$.

   end for
5. Adapt K_i so that $SD(\theta_i(t)) \approx c2 \cdot \max(|\overline{\theta_i(t)}|, c1) \ \forall i$
6. If t<M then go to step 4.

https://github.com/Byshkin/EquilibriumExpectation

A convergence test :

I) $\left|\frac{\overline{z_i(x) - z_i(x_{obs})}}{SD(z_i(x))}\right| < 0.1. \ \forall i$

II) θ_i converge $\forall i$

If θ_i converge to constant values then this convergence test coincides with that of Tom Snijders

Byshkin, et al. "Fast Maximum Likelihood estimation via Equilibrium Expectation for Large Network Data", under review in Scientific Reports, *arXiv preprint arXiv:1802.10311* (2018).

Comparison with other methods

Most algorithms for precise parameter estimation require many converged MCMC simulations

MCMC-MLE

Geyer CJ & Thompson EA, Constrained Monte Carlo Maximum Likelihood for dependent data, Journal of the Royal Statistical Society (1992)

Handcock, Mark S., et al. "statnet: Software tools for the representation, visualization, analysis and simulation of network data." *Journal of statistical software* 24.1 (2008): 1548

Stochastic approximation

Snijders, Tom AB. Markov chain Monte Carlo estimation of exponential random graph models, Journal of Social Structure, 1-40 (2002)

Bayesian estimation

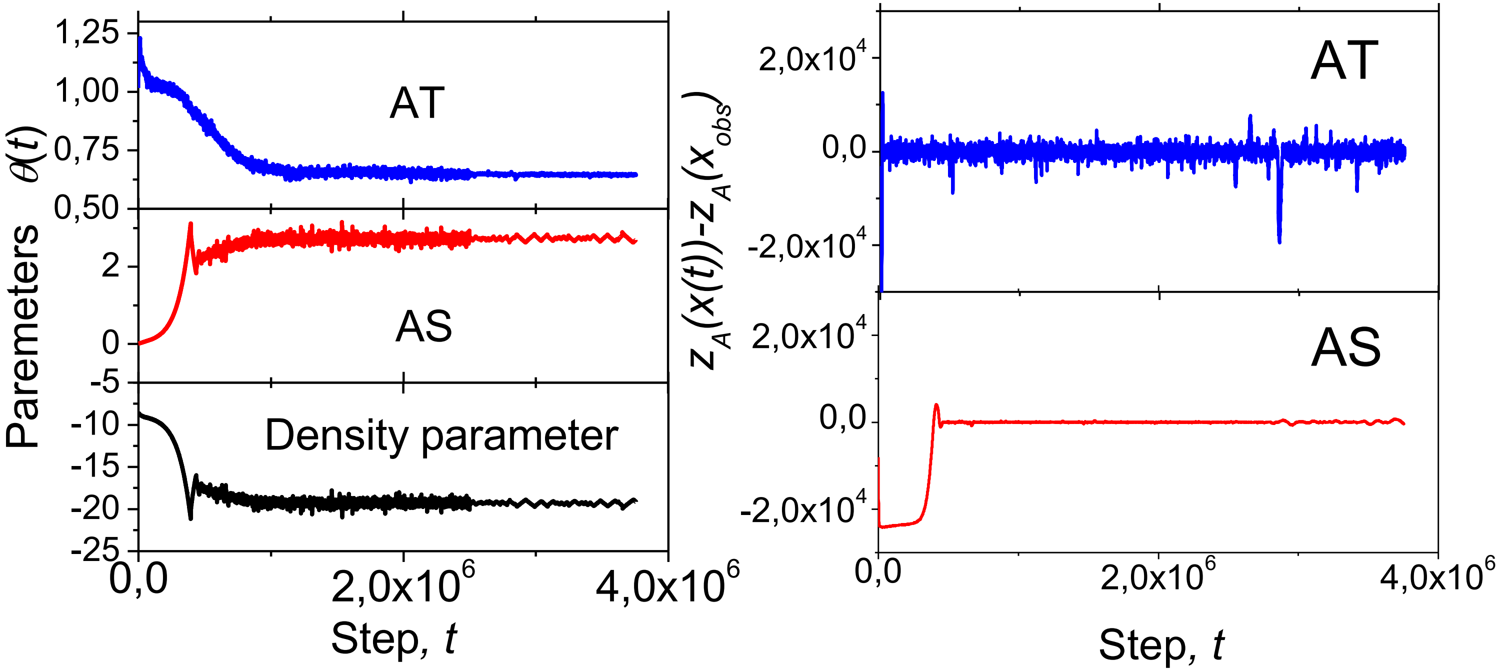
Caimo Alberto, and Nial Friel. "Bayesian inference for exponential random graph models." *Social Networks* 41-55 (2011)

EE algorithm does not require many converged simulations
EE algorithm can be used for models with number of parameters=number of statistics

Illustration. Large undirected network

Livemocha was a social network for language learning
104103 nodes and 2193083 ties

Output of the EE algorithm with IFD sampler

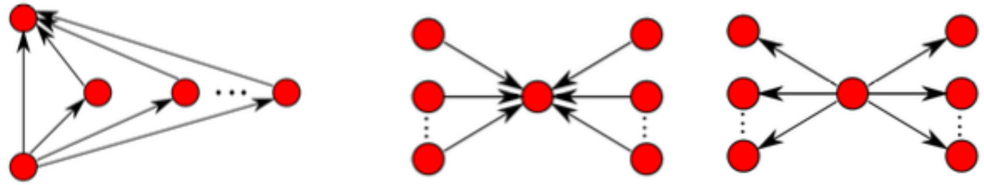


Using existing methods it is not possible to estimate empirical network with more then 5000 nodes even in several months

With EE algorithm the estimation of 104013 node network took several hours

Byshkin, M., Stivala, A., Mira, A., Krause, R., Robins, G., & Lomi, A. Auxiliary parameter MCMC for exponential random graph models. *Journal of Statistical Physics, 165*(4), 740-754 (2016)

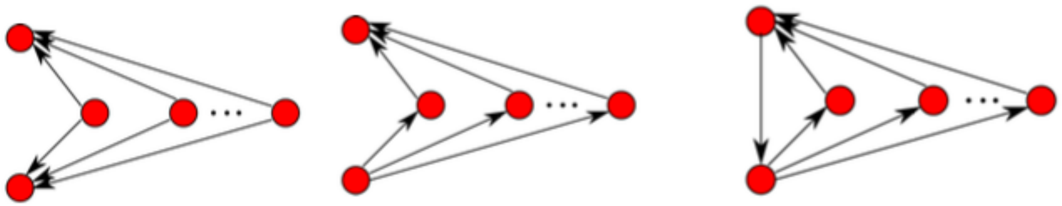
Illustration. Directed network



AltKTrianglesT (path closure)

AltInStars

AltOutStars



AltTwoPathsTD counts both A2P-D (left) and A2P-T (right), adjusting for double-counting.

AltKTrianglesC (cyclic closure)

Robins, G., Pattison P., and Wang P. Closure, connectivity and degree distributions: Exponential random graph (p*) models for directed social networks. Social Networks 105-117 (2009)

Illustration. Directed network

Parameter estimates for the political blogs network (1490 nodes), with the political value (liberal or conservative) coded as a categorical attribute.

Effect	Model 1	Model 2	Model 3	Model 4
Arc	−9.815 (−9.932,−9.699)	−28.631 (−28.655,−28.607)	−9.530 (−9.647,−9.412)	−27.183 (−27.200,−27.167)
AltInStars	1.468 (1.415,1.521)	7.823 (7.792,7.854)	1.473 (1.419,1.526)	8.039 (8.009,8.070)
AltOutStars	0.967 (0.913,1.021)	4.146 (4.109,4.184)	0.968 (0.912,1.024)	4.108 (4.072,4.143)
Reciprocity	2.411 (2.249,2.573)	6.384 (6.281,6.488)	1.227 (1.191,1.264)	3.332 (3.318,3.347)
AltTwoPathsTD	−0.080 (−0.083,−0.077)	−0.074 (−0.075,−0.074)	−0.080 (−0.083,−0.078)	−0.075 (−0.076,−0.074)
AltKTrianglesT	1.141 (1.123,1.158)	—	1.139 (1.122,1.157)	—
Matching	0.294 (0.250,0.337)	1.802 (1.776,1.828)	—	—
MatchingReciprocity	−1.185 (−1.345,−1.025)	−3.050 (−3.153,−2.947)	—	—
Mismatching	—	—	−0.293 (−0.325,−0.261)	−1.814 (−1.838,−1.790)
MismatchingReciprocity	—	—	1.195 (1.031,1.359)	3.116 (3.043,3.188)

Robins, G., Pattison P., and Wang P. Closure, connectivity and degree distributions: Exponential random graph (p*) models for directed social networks. Social Networks 105-117 (2009)

Illustration. Directed network

DBLP, a database of N=12591 scientific publications

Each node in the network is a publication, and each edge represents a citation of a publication by another publication

http://konect.uni-koblenz.de/networks/dblp-cite

Arc	Reciprocity	AltInStars	AltOutStars	AltTwoPathsTD	AltKTrianglesT
-14.91	-5.623	0.7412	3.4385	-0.01237	1.8251
(-15.164; -14.666)	(-5.916; -5.331)	(0.732; 0.749)	(3.313; 3.563)	(-0.0129; -0.0117)	(1.815; 1.835)

Estimation time: 2 hours on laptop.
Basic sampler (random dyad flipping) was used.

IFD sampler would decrease the estimation time by at least one order of magnitude

Byshkin, M., Stivala, A., Mira, A., Krause, R., Robins, G., & Lomi, A. Auxiliary parameter MCMC for exponential random graph models. *Journal of Statistical Physics*, *165*(4), 740-754 (2016)

Illustration. Directed network

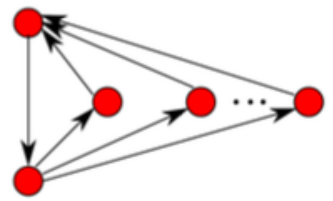
DBLP, a database of N=12591 scientific publications

Each node in the network is a publication, and each edge represents a citation of a publication by another publication

http://konect.uni-koblenz.de/networks/dblp-cite

Arc	Reciprocity	AltInStars	AltOutStars	AltTwoPathsTD	AltKTrianglesT
-14.91	-5.623	0.7412	3.4385	-0.01237	1.8251
(-15.164; -14.666)	(-5.916; -5.331)	(0.732; 0.749)	(3.313; 3.563)	(-0.0129; -0.0117)	(1.815; 1.835)

There are no directed cycles in citation networks



AltKTrianglesC
(cyclic closure)

Sunbelt 2018. pap09.04.01 June 29 Alexander Graham
Modelling Directed Acyclic Graphs in Exponential Random Graph Models

Acknowledgements

We thank Swiss Platform for Advanced Scientific Computing and
Swiss National Science Foundation for financial support,
Melbourne Bioinformatics for high performance computing resources,

We thank Dr. Peng Wang for the source code of the PNet program
We thank Mark Girolami, Illia Horenko, Ritabrata Dutta and Pavel Krivitsky for helpful discussions

The R library for estimation of undirected ERGMs is available from www.estimnet.org

Estimnet for directed ERGMs is available from
https://sites.google.com/site/alexdstivala/home/estimnetdirected