

Fast maximum likelihood estimation via equilibrium expectation for large network data

Author and Presenter:

Maksym Byshkin (Università della Svizzera italiana, Switzerland)

Co-authors:

Alex Stivala (Università della Svizzera italiana, Switzerland)

Antonietta Mira (Università della Svizzera italiana, Switzerland)

Alessandro Lomi (Università della Svizzera italiana, Switzerland)

Abstract: Network data may be analyzed by constructing statistical models that accurately reproduce the structural properties that may be of theoretical relevance or empirical interest. Examples of such models are Exponential Random Graph Models for social networks and Markov Random Field for image processing. Typically, Markov chain Monte Carlo (MCMC) methods are used when normalizing constants of statistical models cannot be computed. We improve on a recently developed Auxiliary Parameter Markov Chain Monte Carlo method [1] and propose a new MCMC approach for the Maximum Likelihood Estimation (MLE) of parameters of statistical models from exponential family. The existing MCMC approaches for parameter estimation (Bayesian, MCM-CMLE of Geyer and Thompson [2], and stochastic approximation for Method of Moments [3]) require many MCMC simulations until convergence. The approach we propose does not require many converged simulations and, in a result, is much less computationally expensive. It relies on properties of Markov chains at equilibrium and, for this reason, we call it Equilibrium Expectation (EE). Using this approach we design a simple and efficient algorithm to find the MLE when it exists and is unique. The EE algorithm is similar to the Metropolis-Hastings algorithm, but allows MCMC simulation to be performed while constraining the desired networks properties.

We demonstrate the performance of the EE algorithm in the context of Exponential Random Graph Models (ERGMs) - a family of statistical models for network data. The EE algorithm is first tested on simulated networks. We compute bias and variance of the estimates and show that the estimates obtained with the proposed method are not less accurate than those obtained with stochastic approximation. Thus far, the lack of efficient computational methods has limited the empirical scope of ERGMs to relatively small networks with a few thousand nodes. Good scaling properties of the EE algorithm allow a dramatic increase in the size of networks that may be analyzed with ERGMs. This is illustrated in an analysis of several biological and one social network with 104,103 nodes.

References

- [1] M. Byshkin, A. Stivala, A. Mira, R. Krause, G. Robins and A. Lomi. *Auxiliary Parameter MCMC for Exponential Random Graph Models*, Journal of Statistical Physics, **165**(4):740-754, 2016.
- [2] C.J. Geyer and E.A. Thompson. *Constrained Monte Carlo maximum likelihood for dependent data*, Journal of the Royal Statistical Society, Series B (Methodological), 657-699, 1992.
- [3] T.A. Snijders. *Markov chain Monte Carlo estimation of exponential random graph models*, Journal of Social Structure, **3**(2):1-40, 2002.