Inferring Dynamic Signalling Networks from Steady State Measurements

Harri Lähdesmäki123, Antti Larjo123, Olli Yli-Harja23, Ilya Shmulevich13

Keywords: network inference, steady state data, dynamic Bayesian network, RJMCMC

1 Introduction.

A central goal in systems biology is to reveal the underlying molecular mechanisms that control various biological processes. This problem can be studied in silico with the help of high-throughput measurement techniques. Dynamic regulatory models of such control mechanisms are typically learned from time series measurements. However, it is not always possible to monitor changes over time, and only steady state behavior can be measured. Inference is then commonly restricted to learning static models. Here, we propose to infer dynamic network models directly from steady state measurements. As an example study, we apply a modified version of our recently proposed Bayesian structure learning algorithm [3] to a subnetwork of a signalling pathway studied in [4].

2 Methods.

Probabilistic graphical models have been found to be a useful model class for representing and inferring stochastic biological control mechanisms [2]. Here, we focus on the class of dynamic Bayesian networks (DBN), which is one of the most often used classes of dynamic probabilistic graphical models. We follow the standard first-order, discrete DBN model, which is defined by a directed graphical model structure \( M \) and a family of conditional distributions \( F \) and their parameters \( \theta \). Directed edges between nodes (variables representing e.g. molecular concentrations) at consecutive time points represent causal relationships. The conditional distribution family \( F \) is assumed to consist of multinomial distributions. A DBN \((M, \theta)\) defines a stochastic (Markov) process \( P(X[t]|X[t-1], \ldots, X[1]) = P(X[t]|X[t-1]) \) where \( X[t] \) represents the variables at time \( t \) and \( P(X[t]|X[t-1]) \) factorizes according to \( M \).

In a Bayesian setting, the most natural score function for network structure learning is the posterior probability of \( M \) given the data \( D \), i.e., \( P(M|D) \). For time series data, \( P(M|D) \) has an analytically tractable solution and the posterior can be explored with various sampling strategies, such as Markov chain Monte Carlo or bootstrap. Learning a dynamic network model is more complicated in the case of steady state measurements. Let \( A^{(r)}_{uv} = P(X[t+r] = v|X[t] = u) \) denote the \( r \)-step transition probability of a homogeneous Markov chain from state \( u \) to state \( v \). A Markov chain is said to possess a steady state distribution if there exists a probability distribution \( \pi \) such that \( \lim_{r \to \infty} A^{(r)}_{uv} = \pi_v \) for all states \( u \) and \( v \). We assume that the underlying biological model possesses a unique steady state distribution \( \pi \) from which the steady state measurements are also sampled. In most applications, this is a reasonable assumption. Since there is no analytical solution for the Bayesian structure score, we have developed a reversible jump Markov chain Monte Carlo (RJMCMC) strategy for sampling from the full posterior [3].

---

1Institute for Systems Biology, Seattle, WA, USA.
2Institute of Signal Processing, Tampere University of Technology, Tampere, FINLAND.
3Emails: hlahdesmaki@systemsbiology.org, antti.larjo@tut.fi, yliharja@cs.tut.fi, ishmulevich@systemsbiology.org
Here, we modify our RJMCMC estimation method to cope with steady state interventions and apply it to a subnetwork of a signalling pathway studied in [4] where some of the phosphorylated protein level measurements are perturbed. This modification can be achieved similarly to the case of time series inference except that in our case interventions must be incorporated via the steady state analysis of the system. For illustration purposes, we focus on a subsystem consisting of only four proteins; see Figure 1 (a) for a “gold standard” network [4]. Data points are discretized as being up- or down-regulated using the same method as in [4]. We ignore one of the experiments (PKC inhibited) as data characteristics are remarkably different and that can also affect discretization. We also subsample data to 20 measurements per experiment.

3 Results and conclusions.

Inferred posterior probabilities of network edges, averaged from two independent RJMCMC chains and thresholded with $P = 0.5$, are shown in Figure 1 (b). We assess convergence by plotting estimated probabilities from the two runs, see Figure 1 (c). An important feature of our dynamic model inference is the possibility of inferring feedback loops, which is not possible with static BNs. Our preliminary results indeed suggest various feedback mechanisms, including self-loops. PLCg–PIP2–PIP3 loop is not correctly inferred but we find PLCg–PIP2 loop. In particular, as also reported in [4], PIP3→PLCg connection is inferred to the opposite direction.

Future work includes improving the computational complexity/convergence properties of the RJMCMC procedure in the case of interventional data and incorporating a more realistic intervention model similar to the one, e.g., in [1].

References


