Inferring Dynamic Signalling Networks from Steady State Measurements

Harri Lähdesmäki[1,2], Antti Larjo[1,2], Olli Yli-Harja[2] & Ilya Shmulevich[1]

[1]Institute for Systems Biology, Seattle, WA, USA
[2]Institute of Signal Processing, Tampere University of Technology, Tampere, FINLAND

Abstract

Dynamic models of regulatory networks and signalling pathways 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. The proposed approach can also be applied to a combination of time series and steady state data. As an example study, we apply a modified version of our recently proposed Bayesian structure learning algorithm [2] to a subnetwork of a signalling pathway studied in [3].

Methods

Modeling Framework

- Probabilistic graphical models are commonly used to model stochastic biological systems [1]
- Directed edges between nodes represent causal relationships
- We focus on commonly used discrete and first-order dynamic Bayesian networks (DBN)
- A DBN is defined by a directed graphical model structure $\mathcal{M}$ and parameters θ

Steady State Analysis

- The r-step transition probability from state $\mathbf{u}$ to state $\mathbf{v}$ is denoted by $A^{(r)}_{\mathbf{uv}} = P(\mathbf{X}[t+r] = \mathbf{v}|\mathbf{X}[t] = \mathbf{u})$
- DBN (or Markov chain) possesses a stationary distribution if there exists a probability distribution π such that $\lim_{r\to\infty} A^{(r)}_{\mathbf{uv}} = \pi_{\mathbf{v}}$ for all states $\mathbf{u}$ and $\mathbf{v}$
- **A common assumption:** underlying biological model possesses a unique steady state distribution π from which the steady state measurements are also sampled

Learning Model Structure

- DBNs are traditionally learned from time series data $\mathcal{D}_A = (\mathbf{x}[1], \mathbf{x}[2], \ldots, \mathbf{x}[T])$
- Often only stationary data (e.g. stationary microarray data) is available $D_\pi = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M)$
- Bayesian inference is often preferred
- Posterior probability of network $\mathcal{M}$ given data $\mathcal{D} = (\mathcal{D}_A, D_\pi)$

$$P(\mathcal{M}|\mathcal{D}) = \frac{P(\mathcal{D}|\mathcal{M})P(\mathcal{M})}{P(\mathcal{D})},$$

where the marginal likelihood is $P(\mathcal{D}|\mathcal{M}) = \int_\theta P(\mathcal{D}|\mathcal{M},\theta)P(\theta|\mathcal{M})d\theta$

- If $\mathcal{D} = \mathcal{D}_A$, then the marginal likelihood has the well-known analytically tractable solution

⇒ The posterior (over networks) can be explored e.g. with Markov chain Monte Carlo (MCMC)

- No analytical solution if measurements include steady state data $\mathcal{D}_\pi$ (note: steady state data is "scored" relative to π)

⇒ Full Bayesian analysis requires trans-dimensional MCMC (we use reversible jump MCMC)

- RJMCMC samples over "product space" of network structures $\mathcal{M}$ and parameters θ

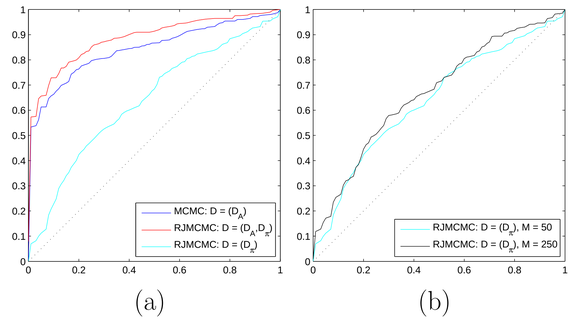
1. Initialize $\mathcal{M}$ and θ
2. propose a new structure $\mathcal{M}'$ (or keep the same $\mathcal{M}$) and propose new parameters θ'
3. accept/reject the proposed move with a probability that the satisfies detailed balance condition [2]
4. repeat steps 2–3

FIGURE 1: Outline of the RJMCMC for DBNs.

- after a burn-in, collect a sample from the chain to estimate the joint probability $P(\mathcal{M}, \theta|\mathcal{D})$ and marginal $P(\mathcal{M}|\mathcal{D})$ (for "all" $\mathcal{M}$ and θ)

Simulation Studies

- Simulation study using known network models
- Figures below show learning performance for three different scenarios



(a) (b)

FIGURE 2: ROC curves. (a) standard time series inference (blue), inference from steady state data only (cyan), combination of time series and steady state data (red). (b) Steady state based inference for two sample sizes, $M = 50$ (cyan) and $M = 250$ (black).

Protein Signalling Pathway Example

- An illustrative example study using a subnetwork from [4] (steady state phosphorylated protein level measurements, including perturbations)
- The above methods are modified to cope with steady state interventions
- Data points are discretized as being up- or down-regulated and subsampled to 20 measurements per experiment

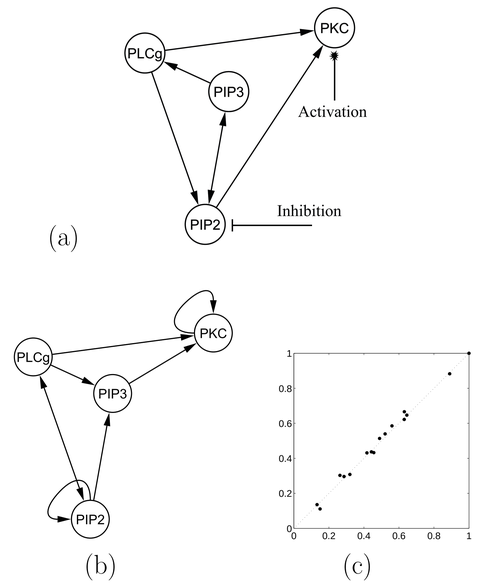


FIGURE 3: (a) A "gold standard". (b) Learned network connections (posterior prob. over 0.5). (c) Convergence diagnostic.

Conclusions & Future Directions

- Method is able to learn dynamic network models even if only non-time series measurements are available
- Inference from a combination of time series and steady state data improves performance relative to the standard state-of-the-art Bayesian inference using time series data alone (no previous method is able to use steady state data)

References

[1] Friedman, N. 2004. Inferring cellular networks using probabilistic graphical models. *Science*, 303:799-805.

[2] Lähdesmäki, H. and Shmulevich, I. Technical report/Submitted. Learning the structure of dynamic Bayesian networks from time series and steady state measurements.

[3] Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D. A. and Nolan, G. P. 2005. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308:523-529.