Graphical Model Structure Inference Using Trees

Loïc Schwaller, Stéphane Robin, Michael Stumpf

Introduction

Let V be a finite discrete set and $\mathbf{X} = (X_v)_{v \in V}$ a vector of random variables indexed by V on the space $\mathcal{X} = \bigotimes_{v \in V} \mathcal{X}_v$. For an undirected decomposable graph $G = (V, E_G)$, we denote $M(G)$ the set of distributions for X that are Markov w.r.t. G. An (undirected) graphical model for X is a couple $m_G = (G, \mathcal{F}_G)$ where G is a graph over V and $\mathcal{F}_G \subset M(G)$. Let $m_{G^*} = (G^*, \mathcal{F}_{G^*})$ be a graphical model and $\mathbf{x}$ a sample drawn from a distribution belonging to $\mathcal{F}_{G^*}$. The problem at hand is to retrieve the set of edges E_{G^*} of G^* from a sample. We make the assumption that G^* is in fact a **spanning tree** i.e. a minimal connected subgraph. If we denote $\mathcal{T}$ the set of spanning trees, we have, for an edge $\{k, l\}$,

$$p(\{k,l\} \in E_{G^*} | \mathbf{x}) = \sum_{T \in \mathcal{T} : \{i,j\} \in E_T} p(T|\mathbf{x}). \tag{1}$$

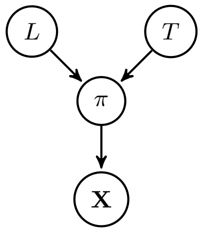
Using Bayes rule, $p(T|\mathbf{x}) \propto p(\mathbf{x}|T)p(T)$ and we have to compute the likelihood of $\mathbf{x}$ under every graphical model with a tree structure. Even under the assumption that G^* is a tree, computing all the likelihoods remain heavy. We describe a family of hyper distributions for $\{\mathcal{F}_T\}_{T \in \mathcal{T}}$ under which the computation can be done efficiently.

Strong Hyper Markov Distributions & Model Description

Let G be a graph and L a distribution over $M(G)$. We will use the term hyper distribution to describe L since it is a distribution over a space of distributions. Let $\pi \in M(G)$ be a random distribution following L. For $A, B \subset V$, we denote π_A the marginal distribution obtained from π on the variables $\mathbf{X}_A$ and $\pi_{B|A}$ the collection of conditional distributions of $\mathbf{X}_B|\mathbf{X}_A$ under π. We also denote L_A the marginal hyper distribution induced by L on π_A and $L_{B|A}$ the collection of hyper distributions induced by L on $\pi_{B|A}$. L is said to be **strong hyper Markov** w.r.t. G if, for any decomposition (A, B) of G, $\pi_{B|A} \perp\!\!\!\perp \pi_A$.

Proposition 1. *If L is strong hyper Markov w.r.t. G, then the marginal likelihood $p(\mathbf{x}) = \int_{M(G)} \pi(\mathbf{x}) L(\pi) d\pi$ is Markov with respect to G.*

$$p(T) = 1/|\mathcal{T}|;$$
$$p(\pi|L,T) = L^T(\pi);$$
$$p(\mathbf{X}|\pi) = \pi(\mathbf{X}).$$



Let L be a hyper distribution over the space of all distributions for $\mathbf{X}$ such that, for any $A \subset V$, $\pi_A \perp\!\!\!\perp \pi_{V \setminus A|A}$ $(\star)$. For any graph G, there exists a unique hyper distribution L^G over $M(G)$ that is strong hyper Markov w.r.t. G and such that, for all clique C of G, $L^G_C = L_C$. A family of hyper distributions over different decomposable graphical models that can be obtained like this is said to be (strongly) **compatible**. We can then easily build our model from a given hyper distribution L verifying $(\star)$ (see opposite).

The marginal likelihood of $\mathbf{X}$ under any tree graphical model can then easily be computed as a product of locally integrated quantities, thanks to compatibility and proposition 1. The computation has only to be done once and for all on any possible edge. In a Bayesian framework, a Dirichlet prior defines an eligible hyper distribution L in the multinomial case. In the multivariate normal case, a normal-Wishart prior over the mean and precision matrix can be used.

Matrix-Tree Theorem

Let $W = (W_{ij})_{i,j \in V}$ be a symmetric weight matrix such that, for all $i \in V$, $W_{ii} = 0$. The Laplacian matrix Q relatively to the weights W is defined by $Q_{ij} = -W_{ij}$ if $i \neq j$ and $\sum_k W_{ik}$ otherwise.

Theorem 1 (Matrix-Tree theorem)**.** *Let Q be the Laplacian matrix associated to weights W. Let $\overline{Q}_{ij}$ denote the $(i,j)^{th}$ minor of Q. Then all $\overline{Q}_{ij}$ are equal and $\overline{Q}_{ij} = \sum_{T \in \mathcal{T}} \prod_{\{i,j\} \in E_T} W_{ij}$.*

We use the weights $W_{ij}(\mathbf{x}) = f_{ij}(x_i, x_j)/f_i(x_i)f_j(x_j)$ where, for all $i, j \in V$, $f_{ij}(x_i, x_j) = \int \pi_{ij}(x_i, x_j) L_{ij}(\pi_{ij}) d\pi_{ij}$ and $f_i(x_i) = \int \pi_i(x_i) L_i(\pi_i) d\pi_i$ to compute the sum over all trees $Z(\mathbf{x}) = \sum_{T \in \mathcal{T}} \prod_{\{i,j\} \in E_T} W_{ij}(x_i, x_j)$. The sum $Z_{kl}(\mathbf{x})$ over the trees that do not contain a particular edge $\{k, l\}$ can similarly be computed by using a weight matrix $W^*(\mathbf{x})$ equal to $W(\mathbf{x})$ except for the terms $W_{kl}(\mathbf{x}) = W_{lk}(\mathbf{x})$ that are set to zero. Then the probability for an edge to belong to a tree T given the observation $\mathbf{x}$ is given by

$$p(\{k, l\} \in E_T | \mathbf{x}) = 1 - Z_{kl}(\mathbf{x})/Z(\mathbf{x}).$$

Simulations

We simulated data according to gaussian graphical models. An array of different topologies were taken (tree, Erdös-Renyi, not connected, etc) to see how robust our method was to the true network not being a tree. We used the multivariate normal framework with a Wishart prior for the inference. We also used the topology of the RAF signalling network presented in [6]. Some results are presented below. The inference using trees performs better when the true graph is not too far from a tree but the performances are still reasonable in the Erdös-Rényi case.

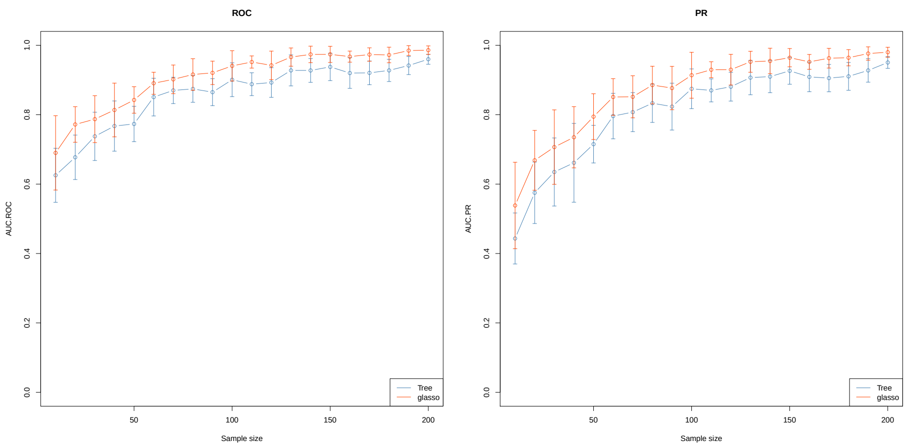


Figure 1: RAF signalling network inference results. Area under the ROC (left) and PR (right) curves for different sample sizes. The gaussian framework of the method presented here (blue) is compared to a regularized shrinkage estimate of the partial correlation matrix [5] (orange).

References

[1] Simon Byrne and A Philip Dawid. "Structural Markov graph laws for Bayesian model uncertainty". 1993.

[2] Seth Chaiken. "A Combinatorial Proof of the All Minors Matrix Tree Theorem". In: *SIAM Journal on Algebraic Discrete Methods* 3.3 (1982), pp. 319–329.

[3] A Philip Dawid and Steffen L. Lauritzen. "Hyper Markov Laws in the Statistical Analysis of Decomposable Graphical Models". In: *The Annals of Statistics* 21.3 (1993), pp. 1272–1317.

[4] Dan Geiger and David Heckerman. "Parameter priors for directed acyclic graphical models and the characterization of several probability distributions". In: *The Annals of Statistics* 30.5 (2002), pp. 1412–1440.

[5] Juliane Schäfer and Korbinian Strimmer. "An empirical Bayes approach to inferring large-scale gene association networks." In: *Bioinformatics (Oxford, England)* 21.6 (Mar. 2005), pp. 754–764. ISSN: 1367-4803.

[6] Adriano V Werhli, Marco Grzegorczyk, and Dirk Husmeier. "Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical gaussian models and bayesian networks." In: *Bioinformatics (Oxford, England)* 22.20 (2006), pp. 2523–2531.