

Laplace approximation for inferring causal directed acyclic structures in gene regulatory networks

G. Nuel

Stochastics and Biology Group, LPSM (CNRS 8001),
Sorbonne University, Paris, France

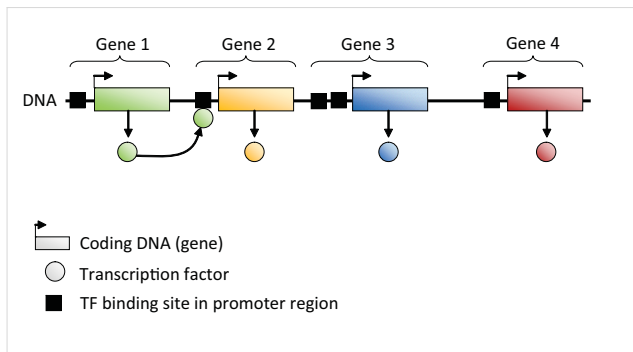
*Journées Francophones sur les Réseaux Bayésiens et les
Modèles Graphiques Probabilistes*

May 31 - June 1, 2018
Toulouse, France



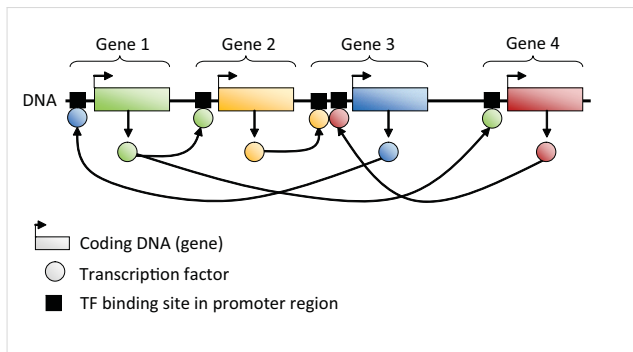
Gene Regulatory Networks (GRN)

- Groups of coordinated genes that interact indirectly with one another through transcription factors



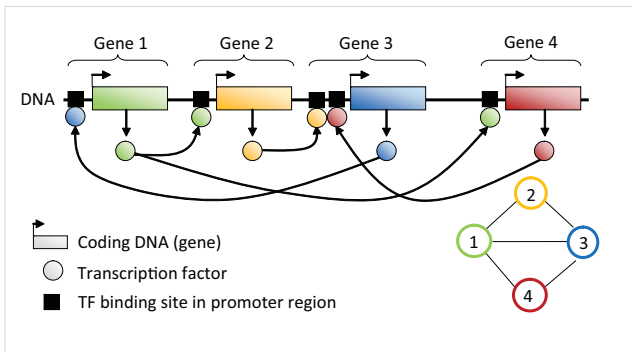
Gene Regulatory Networks (GRN)

- Groups of coordinated genes that interact indirectly with one another through transcription factors



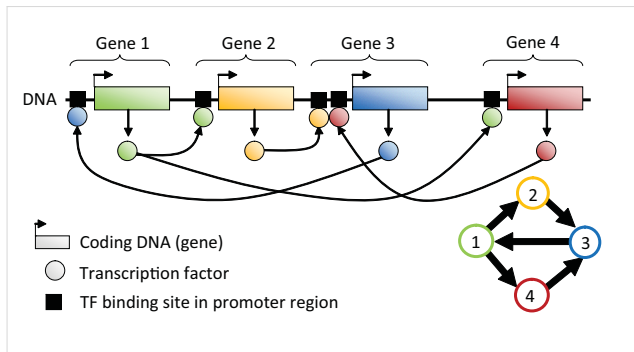
Gene Regulatory Networks (GRN)

- Groups of coordinated genes that interact indirectly with one another through transcription factors

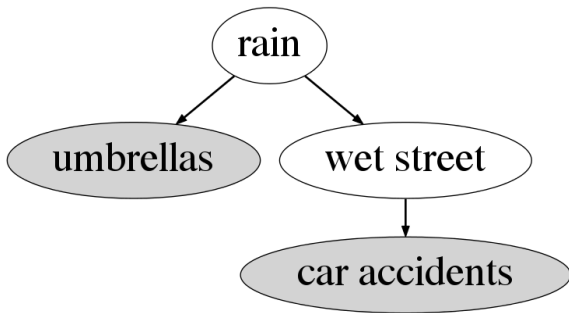


Gene Regulatory Networks (GRN)

- Groups of coordinated genes that interact indirectly with one another through transcription factors



Correlation *versus* Causality



umbrellas and car accidents are correlated

But:

- provoking car accidents does not make appear umbrellas
- distributing umbrellas in the street does not provoke car accidents

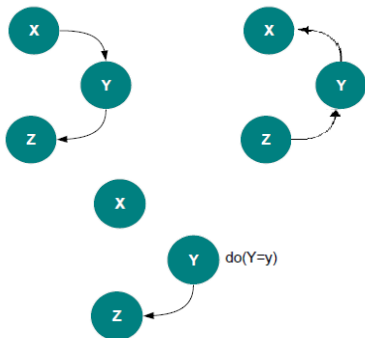
The “do” Operator

- **Markov equivalence**: two different network structures can yield the same joint distribution and **observational data alone generally cannot orient edges**.



The “do” Operator

- **Markov equivalence:** two different network structures can yield the same joint distribution and **observational data alone generally cannot orient edges.**



- **the “do” operator:** a classical way (Pearl, 2009) to model interventions in causal inference.

Previous approaches

Knock-out (KO) experiments (Pinna et al, 2011):

- expensive and time consuming data to obtain;
- rigorous design (systematic KO);
- simple methods (correlation, deviation matrices).

Wild-type (WT) experiments (Maathuis et al 2009):

- cheap and abundant data;
- problem of Markov equivalence (skeleton);
- obtain causal effect bounds (IDA, PCalg).

Mixture of KO and WT experiments (Rau et al, 2013):

- allow complex intervention design (partial, multiple);
- provide a posterior distribution of causal ordering/effects;
- **the present work:** extend to **full posterior DAG distribution**.

Causal Gaussian Bayesian Network

Causal GBN with parameter $\theta = (w, m, \sigma)$: let us denote by X_j the expression of gene $j \in \{1, \dots, p\}$ then we have:

$$X_j = m_j + \sum_{i \in \text{pa}(j)} w_{i,j} X_i + \varepsilon_j \text{ with } \varepsilon_j \sim \mathcal{N}(0, \sigma_j^2)$$

with $w_{i,j} \neq 0$ if and only if $i \in \text{pa}(j)$. NB: with a proper *causal ordering*¹ such that $i \in \text{pa}(j) \Rightarrow i < j$ $\mathbf{W} = (w_{i,j})$ is upper triangular. \mathbf{W} is hence a nilpotent matrix with $\mathbf{W}^p = \mathbf{0}$.

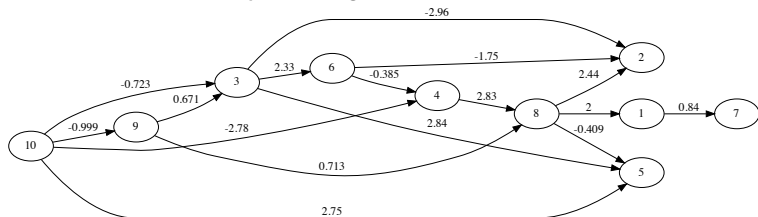
- Direct causal effects $\mathbf{W} = (w_{i,j})$
- Total causal effects $\mathbf{L} = (\ell_{i,j}) = (\mathbf{I} - \mathbf{W})^{-1} = \mathbf{I} + \mathbf{W} + \dots + \mathbf{W}^{p-1}$

$$w_{i,j} = \frac{d}{dx} \mathbb{E}[X_j | X_{-j}, \text{do}(X_i = x)] \quad \ell_{i,j} = \frac{d}{dx} \mathbb{E}[X_j | \text{do}(X_i = x)]$$

¹also called *topological ordering* in a DAG.

A 10 genes example of Causal GBN

A random DAG with $p = 10$ genes



j	1	2	3	4	5	6	7	8	9	10
m	-0.61	-0.41	1.14	-1.84	1.00	0.71	-1.31	-0.96	0.06	0.70
σ	1.90	1.10	0.77	1.30	0.81	0.72	0.98	1.20	0.91	0.41

Some values (a causal ordering 10, 9, 3, 6, 4, 8, 2, 5, 1, 7):

$$\text{pa}(1) = \{8\} \quad \text{pa}(4) = \{6, 10\} \quad \text{pa}(10) = \emptyset$$

$$w_{6,2} = -1.75 \quad l_{6,2} = w_{6,2} + w_{6,4} \times w_{4,8} \times w_{8,2} = -4.41$$

Log-likelihood Results

Exhaustive statistics: For all $i, i', j \in \{1, \dots, p\}$ we define:

$$z_{i,i'}^j = \sum_{k \in \mathcal{K}_j} y_j^{k,i} y_j^{k,i'}$$

Theorem: with the convenient $\sigma_j = \exp(s_j)$ reparametrization we finally get: $\text{loglik}(\mathbf{s}, \mathbf{w}) = \sum_j \text{loglik}_j(\hat{\mathbf{s}}_j, \hat{\mathbf{w}}_{\cdot,j})$

$$\begin{aligned} \text{loglik}_j(\hat{\mathbf{s}}_j, \hat{\mathbf{w}}_{\cdot,j}) &= -\frac{\log(2\pi)}{2} N_j - N_j s_j \\ &- \frac{1}{2} \exp(-2s_j) \left[z_{j,j}^j - 2 \sum_{i \in \text{pa}_j} w_{i,j} z_{i,j}^j + \sum_{i \in \text{pa}_j} \sum_{i' \in \text{pa}_j} w_{i,j} z_{i,j}^j w_{i',j} z_{i',j}^j \right] \end{aligned}$$

with $(\mathbf{s}, \mathbf{w}) \in \mathbb{R}^{p+m}$, m being the number of non-zero terms in W (i.e. the number of edges in the DAG).

Maximum Likelihood Estimator

Each block $(s_j, w_{\cdot,j})$ can be maximized independently by solving:

$$\sum_{i' \in \text{pa}_j} w_{i',j} z_{i,i'}^j = z_{i,j}^j \quad \text{for all } i \in \text{pa}_j.$$

Let us denote by $\mathbf{A}_j \times w_{\cdot,j} = \mathbf{b}_j$ the order $q_j = |\text{pa}_j|$ (symmetric and positive definite) matrix of the corresponding linear system.

Corollary: the MLE of each block is given by

$$\hat{w}_{\cdot,j} = \mathbf{A}_j^{-1} \mathbf{b}_j \quad \text{solved with complexity } \mathcal{O}(q_j^3)$$

$$N_j \exp(2\hat{s}_j) = z_{j,j}^j - 2 \sum_{i \in \text{pa}_j} \hat{w}_{i,j} z_{i,j}^j + \sum_{i \in \text{pa}_j} \sum_{i' \in \text{pa}_j} \hat{w}_{i,j} z_{i,j}^j \hat{w}_{i',j} z_{i',j}^j$$

$$\text{loglik}_j(\hat{s}_j, \hat{w}_{\cdot,j}) = -\frac{\log(2\pi)}{2} N_j - N_j \hat{s}_j - \frac{N_j}{2}$$

A Bayesian Framework: the Laplace Approximation

Posterior DAG distribution:

$$\mathbb{P}(G|\text{data}) \propto \int_{\theta} \underbrace{\mathbb{P}(\text{data}|G, \theta)}_{\text{likelihood}} \times \underbrace{\mathbb{P}(\theta|G)}_{\text{param. prior}} \times \underbrace{\mathbb{P}(G)}_{\text{DAG prior}} d\theta$$

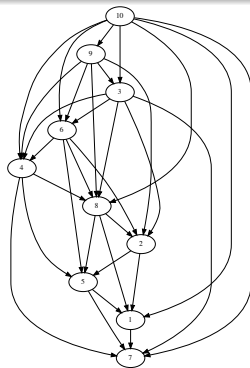
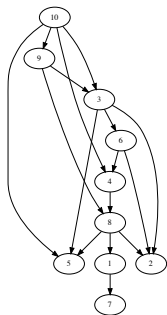
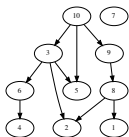
Using **improper uninformative prior** on (s, w) and an order 2 Taylor expansion of the integral above at $\hat{\theta}$ we get:

$$\log \mathbb{P}(G|\text{data}) \simeq \text{Cst.} + \text{loglik}(\hat{s}, \hat{w}|G) + \text{laplace}(\hat{s}, \hat{w}|G) + \log \mathbb{P}(G)$$

where:

$$\text{laplace}(\hat{s}, \hat{w}) = \sum_j \left(\frac{(1 + q_j)}{2} \log(2\pi) - \frac{1}{2} \log \det \left(\frac{\mathbf{A}_j}{\sigma_j} \right) \right)$$

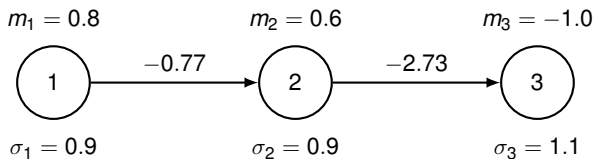
10 genes example: Laplace approximation



	underfitting	reference	overfitting
loglik	-348.76312	-216.82927	-213.6599
laplace	-10.96638	-25.59564	-35.9374
sum	-359.72950	-242.42491	-249.5973

⇒ a classical bias-variance tradeoff

Simple Example



	50 \emptyset		500 \emptyset	500 \emptyset 20 {2}		
DAG	loglik	post	post	loglik	post	
\emptyset {1} {2}	-207.1803	0.322	0.481	-2098.112	0.847	
\emptyset {1} {1, 2}	-206.7987	0.199	0.098	-2098.067	0.142	
{2} \emptyset {2}	-207.1803	0.196	0.288	-2102.267	0.008	
{2} \emptyset {1, 2}	-206.7987	0.121	0.059	-2102.222	0.001	
{2, 3} \emptyset {2}	-206.7987	0.083	0.026	-2102.222	10^{-4}	
{2} {3} \emptyset	-207.1803	0.023	0.033	-2116.247	10^{-10}	

Enumerating DAGs

The problem: compute for all G

$$\mathbb{P}(G|\text{data}) \propto \exp(\text{loglik}(\hat{s}, \hat{w}|G) + \text{laplace}(\hat{s}, \hat{w}|G) + \log \mathbb{P}(G))$$

A simple solution: by enumerating all DAGs.

p	number of DAGs
1	1
2	3
3	25
4	543
5	29,281
6	3,781,503
7	1,138,779,265
8	783,702,329,343

A better idea: through MCMC.

MC3: Metropolis-Hastings in the DAG space

The problem: sample DAG G from

$$\mathbb{P}(G|\text{data}) \propto \exp(\text{loglik}(\hat{S}, \hat{W}|G) + \text{laplace}(\hat{S}, \hat{W}|G) + \log \mathbb{P}(G))$$

Metropolis-Hastings: perform iteratively

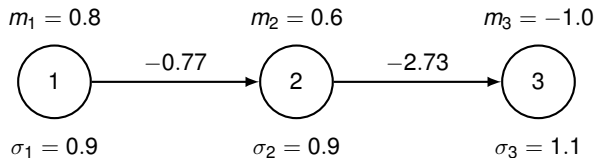
- propose $G' \sim q(\cdot|G)$
- accept G' with rate $\min(1, \alpha)$ with

$$\alpha = \frac{\mathbb{P}(G'|\text{data})}{\mathbb{P}(G|\text{data})} \times \frac{q(G|G')}{q(G'|G)}$$

MC3 (Madigan & Raftery, 1995):

- proposal: add/remove/flip edge *uniformly*
- DAG constraint need smart update of routes tables
- available in `structmcmc` R package (Goudie, 2016)
- more constraints: max number of parents, fixed edges

Back to the Simple Example 3

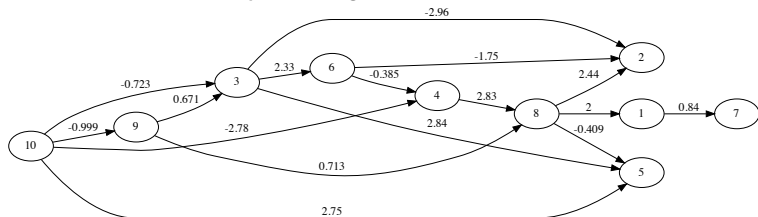


Data: 500 \emptyset and 20 $\{2\}$ – **MC3:** $N = 5000$ with burnin 1000

DAG	loglik	post	rep. #1	rep. #2	rep. #3
$\emptyset \{1\} \{2\}$	-2098.112	0.847	0.857	0.841	0.870
$\emptyset \{1\} \{1, 2\}$	-2098.067	0.142	0.131	0.156	0.119
$\{2\} \emptyset \{2\}$	-2102.267	0.008	0.010	0.003	0.010
$\{2\} \emptyset \{1, 2\}$	-2102.222	0.001	0.001	0.000	0.001
$\{2, 3\} \emptyset \{2\}$	-2102.222	10^{-4}	—	—	—
$\{2\} \{3\} \emptyset$	-2116.247	10^{-10}	—	—	—

Back to our 10 gene example of causal GBN

A random DAG with $p = 10$ genes



j	1	2	3	4	5	6	7	8	9	10
m	-0.61	-0.41	1.14	-1.84	1.00	0.71	-1.31	-0.96	0.06	0.70
σ	1.90	1.10	0.77	1.30	0.81	0.72	0.98	1.20	0.91	0.41

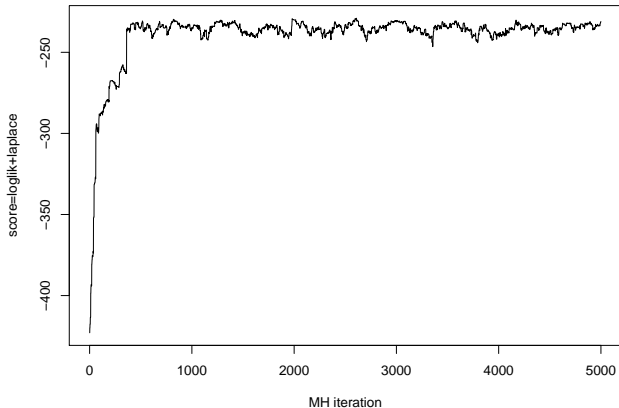
Some values (a causal ordering 10, 9, 3, 6, 4, 8, 2, 5, 1, 7):

$$\text{pa}(1) = \{8\} \quad \text{pa}(4) = \{6, 10\} \quad \text{pa}(10) = \emptyset$$

$$w_{6,2} = -1.75 \quad l_{6,2} = w_{6,2} + w_{6,4} \times w_{4,8} \times w_{8,2} = -4.41$$

10 genes example: MC3

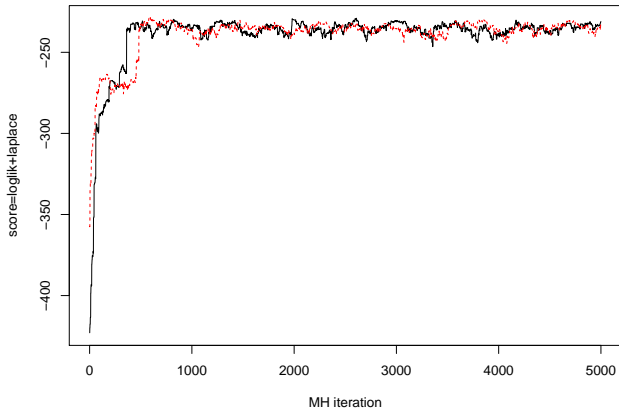
Design: same DAG and data (10 WT, 10 KO), 5,000 MH iterations, unconstrained search, acceptance rate $\sim 40\%$



1 replication

10 genes example: MC3

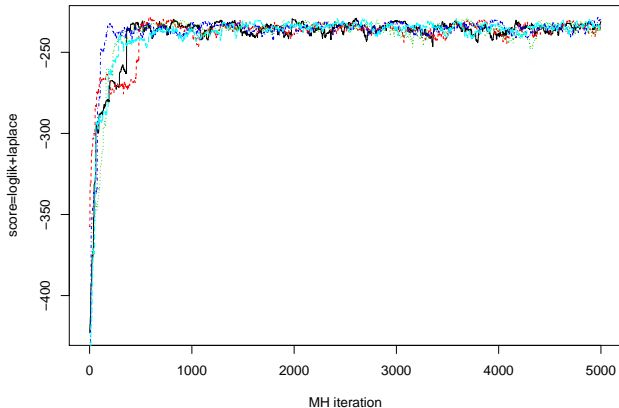
Design: same DAG and data (10 WT, 10 KO), 5,000 MH iterations, unconstrained search, acceptance rate $\sim 40\%$



2 replications

10 genes example: MC3

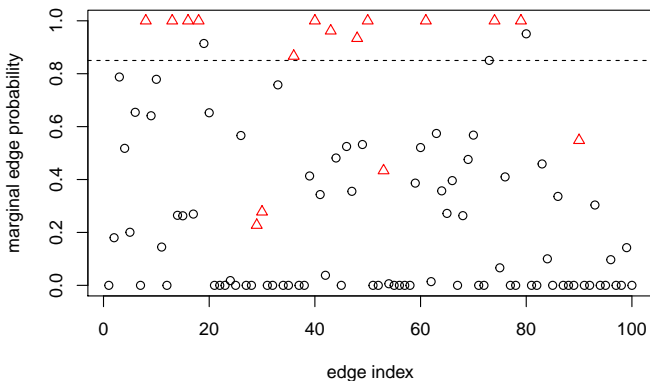
Design: same DAG and data (10 WT, 10 KO), 5,000 MH iterations, unconstrained search, acceptance rate $\sim 40\%$



5 replications

10 genes example: MC3

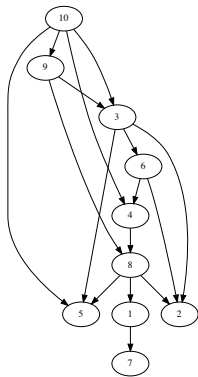
Design: same DAG and data (10 WT, 10 KO), 5,000 MH iterations, unconstrained search, acceptance rate $\sim 40\%$



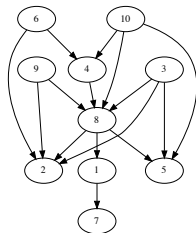
marginal posterior distribution on edges (threshold at 0.85)

10 genes example: MC3

Design: same DAG and data (10 WT, 10 KO), 5,000 MH iterations, unconstrained search, acceptance rate $\sim 40\%$



reference



consensus MC3

Summary:

- new likelihood results with optimization by block in $\mathcal{O}(q_j^3)$;
- Laplace approximation with improper uninformative priors;
- posterior DAG distribution through MC3 (MH-MCMC).

Outlook:

- improvement: parallel tempering with MC4 (Barker, 2010);
- biologically relevant datasets (INRA, Curie, Saint-Antoine);
- extensions: differential expression, time course data, . . .
- original application: mediation analysis mixing observations and randomized trials.

Students:

- Gilles Monneret (PhD, LPSM, Sorbonne Univ.)
- Pascal Fieth (PhD, Univ. Oldenburg, Germany)
- Flaminia Zane (M2, LPSM, Sorbonne Univ.)

Researchers:

- Florence Jaffrézic (DR, GABI, INRA Jouy-en-Josas)
- Andrea Rau (CR, GABI, INRA Jouy-en-Josas)
- Alexander Hartmann (PR, Univ. Oldenburg, Germany)

Funding:

- INRA: Causality and Cosy-net
- French Polynesia: Caristo-pf