The I-MAP Parameterization of Gaussian DAG Models

Alessandro Mascaro[1] David Rossell[1]

[1]Departament d'Economia i Empresa
Universitat Pompeu Fabra, Barcelona

September 5, 2025

SISBayes 2025, session on Bayesian Causal Inference

Gaussian DAG Models and Estimation of Causal Effects

Gaussian SEMs and DAG Models

Suppose we have observations from a q-variate Gaussian random variable X, generated by a linear Gaussian Structural Equation Model (SEM) of the form

$$X = \boldsymbol{B}^T X + \epsilon, \qquad \epsilon \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D}) \tag{1}$$

where $\boldsymbol{B}$ is a (q,q) matrix s.t. $\boldsymbol{B}_{ij} \neq 0$ iff $i \to j \in \mathcal{D}$, with $\mathcal{D}$ a Directed Acyclic Graph (DAG) and $\boldsymbol{D}$ a (q,q) diagonal matrix of node variances

Gaussian SEMs and DAG Models

Suppose we have observations from a q-variate Gaussian random variable X, generated by a linear Gaussian Structural Equation Model (SEM) of the form

$$X = \boldsymbol{B}^T X + \epsilon, \qquad \epsilon \sim \mathcal{N}(\mathbf{0}, \boldsymbol{D}) \tag{1}$$

where $\boldsymbol{B}$ is a (q, q) matrix s.t. $\boldsymbol{B}_{ij} \neq 0$ iff $i \to j \in \mathcal{D}$, with $\mathcal{D}$ a Directed Acyclic Graph (DAG) and $\boldsymbol{D}$ a (q, q) diagonal matrix of node variances

The linear SEM induces a Gaussian DAG model on X, i.e.

$$X \sim \mathcal{N}(\mathbf{0}, \boldsymbol{L}^{-T} \boldsymbol{D} \boldsymbol{L}^{-1})$$

where $\boldsymbol{L} = \boldsymbol{I_q} - \boldsymbol{B}$ and the joint pdf factorises according to the Markov property of $\mathcal{D}$:

$$f(x) = \prod_{j=1}^{q} d\mathcal{N}\left(x_j; -\boldsymbol{L}^T_{pa_j(\mathcal{D}),j} x_{pa_j(\mathcal{D})}, \boldsymbol{D}_{jj}\right), \quad \text{where } pa_j(\mathcal{D}) = \{i : i \to j \in \mathcal{D}\}$$

Causal effects: definition and identification

We are interested in estimating the **total causal effect** τ_{to} of a *treatment* variable X_t on an outcome variable X_o. Using the language of the *do-calculus*

$$\tau_{to} := \frac{\partial \mathbb{E}\left[X_o \mid \mathrm{do}(X_t = \tilde{x}_t)\right]}{\partial \tilde{x}_t} \tag{2}$$

Causal effects: definition and identification

We are interested in estimating the **total causal effect** τ_{to} of a *treatment* variable X_t on an outcome variable X_o. Using the language of the *do-calculus*

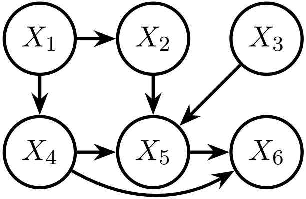
$$\tau_{to} := \frac{\partial \mathbb{E}\left[X_o \mid \mathrm{do}(X_t = \tilde{x}_t)\right]}{\partial \tilde{x}_t} \tag{2}$$

! The numerator Equation 2 requires a marginalization over a **post-intervention** distribution that we do not observe;
If $\mathcal{D}$ is known, it is possible to identify a **valid adjustment set** $z \subset [q]$ s.t. $\{t, o\} \notin z$ and

$$\mathbb{E}(X_o \mid \mathrm{do}(X_t = \tilde{x}_t) = \beta_0 + \tau_{to}\tilde{x}_t + \beta_{zo}^T \mathbb{E}(X_z) \tag{3}$$

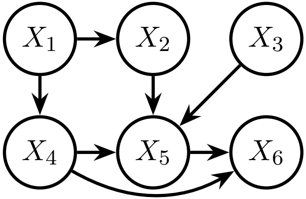
i.e. τ_{to} corresponds to a regression coefficient in a specific regression model!

Causal effects: Example



If we are interested in τ_{46} possible valid adjustment sets are
$z_1 = \{1, 2, 3\}, z_2 = \{1, 3\}, z_3 = \{2, 3\}, z_4 = \{1\}, z_5 = \{2\}$

Causal effects: Example



If we are interested in τ_{46} possible valid adjustment sets are
$z_1 = \{1, 2, 3\}, z_2 = \{1, 3\}, z_3 = \{2, 3\}, z_4 = \{1\}, z_5 = \{2\}$
All these valid adjustment sets would provide an unbiased estimate of the total causal effect of interest. **However**, the correspoding estimator may have different properties;

Causal effects: estimation

The regression coefficient in (3) can be very easily estimated via OLS.

Causal effects: estimation

The regression coefficient in (3) can be very easily estimated via OLS. Alternatively, in the Gaussian case, one may also

i) Obtain an estimate $(\hat{\boldsymbol{L}}, \hat{\boldsymbol{D}})$ via MLE

ii) Derive the MLE of the covariance function of the Gaussian DAG as

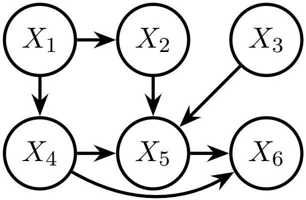
$$\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{L}}^{-T} \hat{\boldsymbol{D}} \hat{\boldsymbol{L}}^{-1}$$

iii) Compute the MLE of τ_{to} as

$$\hat{\tau}_{to} = (\hat{\boldsymbol{\Sigma}}_{z^*,z^*})^{-1} \hat{\boldsymbol{\Sigma}}_{z^*,o}$$

where $z^* = t \cup z$ and z

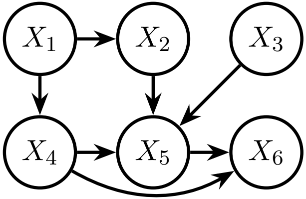
Causal effects: Best Valid Adjustment set



Henckel et al. (2022): graphical characterization of **best valid adjustment set** $\mathcal{O}_{to}(\mathcal{D})$ as the one minimising asymptotic variance of the corresponding OLS estimator:

$$\mathcal{O}_{to}(\mathcal{D}) = \mathsf{pa}_{\mathcal{D}}\left(\mathsf{med}_{\mathcal{D}}(X_t, X_o) \cup X_o\right) \setminus \{X_t \cup \mathsf{med}_{\mathcal{D}}(X_t, X_o)\}$$

Causal effects: Best Valid Adjustment set



Henckel et al. (2022): graphical characterization of **best valid adjustment set** $\mathcal{O}_{to}(\mathcal{D})$ as the one minimising asymptotic variance of the corresponding OLS estimator:

$$\mathcal{O}_{to}(\mathcal{D}) = \mathsf{pa}_{\mathcal{D}}\left(\mathsf{med}_{\mathcal{D}}(X_t, X_o) \cup X_o\right) \setminus \{X_t \cup \mathsf{med}_{\mathcal{D}}(X_t, X_o)\}$$

In our example, $\mathcal{O}_{46}(\mathcal{D}) = \{2, 3\}$

Causal effects: unknown causal structure

If $\mathcal{D}$ is **not** known, one may:

i) Learn $\mathcal{D}$ from data using a **causal discovery** algorithm such as the PC-algorithm;

ii) Identify a valid adjustment set from the learnt DAG, and use it to estimate τ_{to};

Problem: Causal discovery algorithm return Markov equivalence classes of DAGs
Solution: Enumerate all the possible causal effects compatible with the Markov equivalence class (IDA, Maathuis et al. 2009)

Causal effects: unknown causal structure

If $\mathcal{D}$ is **not** known, one may:

i) Learn $\mathcal{D}$ from data using a **causal discovery** algorithm such as the PC-algorithm;

ii) Identify a valid adjustment set from the learnt DAG, and use it to estimate τ_{to};

Problem: Causal discovery algorithm return Markov equivalence classes of DAGs
Solution: Enumerate all the possible causal effects compatible with the Markov equivalence class (IDA, Maathuis et al. 2009)

! This approach would still not account for the uncertainty in the estimation of the Markov equivalence class $\implies$ **Bayesian approaches**

Bayesian Causal Discovery and Causal Effect Estimation

In the Bayesian setting, causal discovery can be tackled as a Bayesian Model Selection task, and causal effect estimation under DAG uncertainty as a Bayesian Model Averaging (BMA) estimation problem;

Bayesian Causal Discovery and Causal Effect Estimation

In the Bayesian setting, causal discovery can be tackled as a Bayesian Model Selection task, and causal effect estimation under DAG uncertainty as a Bayesian Model Averaging (BMA) estimation problem;

Let $\boldsymbol{X}$ be a (n, q) matrix of observations from a causal Gaussian DAG Model. We are interested in the posterior distribution

$$p((\boldsymbol{L}, \boldsymbol{D}), \mathcal{D}|\boldsymbol{X}) = p(\boldsymbol{L}, \boldsymbol{D}|\mathcal{D}, \boldsymbol{X})p(\mathcal{D}|\boldsymbol{X}) \tag{4}$$

Bayesian Causal Discovery and Causal Effect Estimation

In the Bayesian setting, causal discovery can be tackled as a Bayesian Model Selection task, and causal effect estimation under DAG uncertainty as a Bayesian Model Averaging (BMA) estimation problem;
Let $\boldsymbol{X}$ be a (n, q) matrix of observations from a causal Gaussian DAG Model. We are interested in the posterior distribution

$$p((\boldsymbol{L}, \boldsymbol{D}), \mathcal{D} | \boldsymbol{X}) = p(\boldsymbol{L}, \boldsymbol{D} | \mathcal{D}, \boldsymbol{X}) p(\mathcal{D} | \boldsymbol{X}) \tag{4}$$

As the causal effect is a function of the parameters, by sampling from the posterior distribution (4), we can also approximate the posterior distribution over τ_{to} and hence obtain a BMA estimate of τ.

Bayesian Causal Discovery: A simple model

A typical model for Bayesian causal effect estimation under DAG uncertainty in the Gaussian setting is the following:

$$\boldsymbol{X}_{i\cdot}|(\boldsymbol{L}, \boldsymbol{D}), \mathcal{D} \sim \mathcal{N}_q\left(\boldsymbol{0}, \boldsymbol{L}^{-T}\boldsymbol{D}\boldsymbol{L}^{-1}\right) \qquad i \in [n]$$
$$(\boldsymbol{L}, \boldsymbol{D})|\mathcal{D} \sim \text{DAG-Wishart}(\boldsymbol{a}(\mathcal{D}), \boldsymbol{U})$$
$$p(\mathcal{D}) \propto \omega^{|\mathcal{S}_\mathcal{D}|}(1-\omega)^{\frac{q(q-1)}{2}-|\mathcal{S}_\mathcal{D}|}$$

where the DAG-Wishart prior (Ben-David, 2015) is a prior on the Cholesky space of a DAG $\mathcal{D}$, with hyperparameters $\boldsymbol{a}(\mathcal{D}) = (\boldsymbol{a}_1(\mathcal{D}), \dots, \boldsymbol{a}_q(\mathcal{D}))$ and a (q, q) s.p.d. matrix $\boldsymbol{U}$, while $\omega \in (0, 1)$ is a prior probability of edge inclusion.

DAG-Wishart prior: properties

The DAG-Wishart prior has many useful properties:

- It's conjugate, so that

$$p(\boldsymbol{L}, \boldsymbol{D} | \boldsymbol{X}, \boldsymbol{\mathcal{D}}) \sim \text{DAG-Wishart}\left(\boldsymbol{a}(\mathcal{D}) + n, \boldsymbol{U} + \boldsymbol{X}^T \boldsymbol{X}\right)$$

- Its **marginal likelihood** $p(\boldsymbol{X}|\mathcal{D})$ is **available in closed form** and **decomposable**
- Sampling from $p(\boldsymbol{L}, \boldsymbol{D}|\boldsymbol{X}, \boldsymbol{\mathcal{D}}$ occurs by direct sampling.
- Nice selection consistency properties in high-dimensions (Cao et al. 2019);
- The hyperparameters can be chosen in a way that guarantees **score equivalence**, i.e. that Markov equivalent DAGs are assigned the same marginal likelihood (Peluso & Consonni, 2020);

DAG-Wishart prior: properties

The DAG-Wishart prior has many useful properties:

- It's conjugate, so that

$$p(\boldsymbol{L}, \boldsymbol{D} | \boldsymbol{X}, \boldsymbol{\mathcal{D}}) \sim \text{DAG-Wishart}\left(\boldsymbol{a}(\mathcal{D}) + n, \boldsymbol{U} + \boldsymbol{X}^T \boldsymbol{X}\right)$$

- Its **marginal likelihood** $p(\boldsymbol{X} | \mathcal{D})$ is **available in closed form** and **decomposable**
- Samples from $p(\boldsymbol{L}, \boldsymbol{D} | \boldsymbol{X}, \boldsymbol{\mathcal{D}})$ can be obtained by direct sampling.
- Nice selection consistency properties in high-dimensions (Cao et al. 2019);
- The hyperparameters can be chosen in a way that guarantees **score equivalence**, i.e. that Markov equivalent DAGs are assigned the same marginal likelihood (Peluso & Consonni, 2020);

DAG-Wishart prior on causal effects

When doing causal inference, our ultimate goal is testing or estimating causal effects
The DAG-Wishart is a distribution over the non-null elements of the matrices $(\boldsymbol{L}, \boldsymbol{D})$, but causal effects are identified as **DAG-specific** functions of these parameters;

DAG-Wishart prior on causal effects

When doing causal inference, our ultimate goal is testing or estimating causal effects
The DAG-Wishart is a distribution over the non-null elements of the matrices $(\boldsymbol{L}, \boldsymbol{D})$, but causal effects are identified as **DAG-specific** functions of these parameters;
$\implies$ When evaluating different DAGs, we have **very little control** over the prior that we specify on the causal effect of interest!

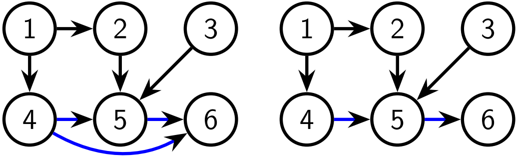


Figure: In the first DAG, $\tau_{46} = \boldsymbol{L}_{45}\boldsymbol{L}_{56} - \boldsymbol{L}_{46}$, in the second $\tau_{46} = \boldsymbol{L}_{45}\boldsymbol{L}_{56}$

DAG-Wishart prior on causal effects

When doing causal inference, our ultimate goal is testing or estimating causal effects
The DAG-Wishart is a distribution over the non-null elements of the matrices $(\boldsymbol{L}, \boldsymbol{D})$, but causal effects are identified as **DAG-specific** functions of these parameters;
$\implies$ When evaluating different DAGs, we have **very little control** over the prior that we specify on the causal effect of interest!

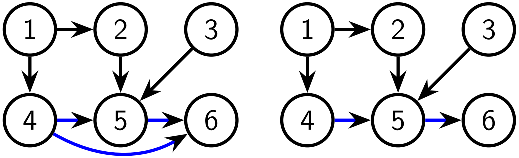


Figure: In the first DAG, $\tau_{46} = \boldsymbol{L}_{45}\boldsymbol{L}_{56} - \boldsymbol{L}_{46}$, in the second $\tau_{46} = \boldsymbol{L}_{45}\boldsymbol{L}_{56}$

Solution: Novel parameterization, the **I-MAP parameterization**, which always contains a **single parameter** corresponding to the total causal effect of interest;

The I-MAP parameterization

Minimal Independence MAPs: Constructive procedure

Suppose X is a Gaussian q-variate random vector and let π be an ordering of the variables. For any π, the joint pdf can be written as

$$f(x) = \prod_{j=1}^{q} f(x_j | x_{a_j(\pi)})$$

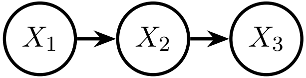
where $a_j(\pi)$ denotes all the variables that precede j in π;
If X is Markov w.r.t. to a DAG $\mathcal{D}$, then a set of conditional independencies $\mathcal{I}(\mathcal{D})$ holds, and for each j we can cancel out conditionally indep. variables from $a_j(\pi)$, so that

$$f(x) = \prod_{j=1}^{q} f(x_j | x_{\mathsf{S}_j(\mathcal{D},\pi)}) \qquad S_j(\mathcal{D},\pi) \subseteq a_j(\pi)$$

Def: A **Minimal Independence Map** (I-MAP) is the DAG $\mathcal{D}^*_\pi$ obtained by drawing, for each $j \in [p]$ an edge $k \to j$ if $k \in S_j(\mathcal{D},\pi)$

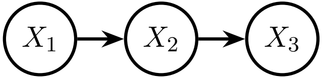
Minimal Independence MAPs: Example

Suppose that our random vector X is Markov w.r.t to the following DAG, implying only that $X_1 \perp X_3 \mid X_2$:

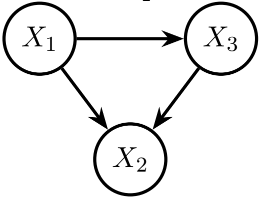


If we consider the two orderings $\pi_1 = X_1 \prec X_2 \prec X_3$ and $\pi_2 = X_1 \prec X_3 \prec X_2$, we obtain the following minimal I-MAPs:

$\mathcal{D}^*_{\pi_1}$



$\mathcal{D}^*_{\pi_2}$



Minimal I-MAPs: Comments

Different orderings imply different Minimal I-MAPs;

If π is topological of $\mathcal{D}$ (i.e., if $i \to j \in \mathcal{D}$ implies that $i \prec j$ (i precedes j) in π), then the Minimal I-MAP corresponds to the data-generating DAG;

Otherwise it is a strictly denser DAG whose Markov property implies a subset of the conditional independencies originally implied by the DAG;

Minimal I-MAPs: Comments

Different orderings imply different Minimal I-MAPs;
If π is topological of $\mathcal{D}$ (i.e., if $i \to j \in \mathcal{D}$ implies that $i \prec j$ (i precedes j) in π), then the Minimal I-MAP corresponds to the data-generating DAG;
Otherwise it is a strictly denser DAG whose Markov property implies a subset of the conditional independencies originally implied by the DAG;

In the **Gaussian case**, one can obtain the parameters of the minimal I-MAP $(\boldsymbol{L}_\pi, \boldsymbol{D}_\pi)$ by permuting the rows and column of $\boldsymbol{\Sigma}_\mathcal{D}$ according to the ordering π, and then computing the LDL decomposition of the permuted covariance matrix
If π is not topological of $\mathcal{D}$, the I-MAP parameters $(\boldsymbol{L}_\pi, \boldsymbol{D}_\pi)$ will have a strictly larger number of non-null elements, **but** the **free parameters** will be the same;

Minimal I-MAPs: Comments

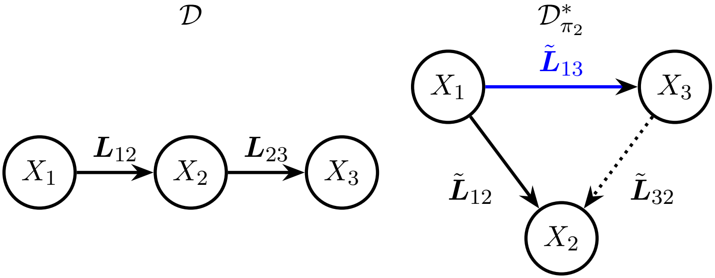
Different orderings imply different Minimal I-MAPs;
If π is topological of $\mathcal{D}$ (i.e., if $i \to j \in \mathcal{D}$ implies that $i \prec j$ (i precedes j) in π), then the Minimal I-MAP corresponds to the data-generating DAG;
Otherwise it is a strictly denser DAG whose Markov property implies a subset of the conditional independencies originally implied by the DAG;

In the **Gaussian case**, one can obtain the parameters of the minimal I-MAP $(\boldsymbol{L}_\pi, \boldsymbol{D}_\pi)$ by permuting the rows and column of $\boldsymbol{\Sigma}_{\mathcal{D}}$ according to the ordering π, and then computing the LDL decomposition of the permuted covariance matrix
If π is not topological of $\mathcal{D}$, the I-MAP parameters $(\boldsymbol{L}_\pi, \boldsymbol{D}_\pi)$ will have a strictly larger number of non-null elements, **but** the **free parameters** will be the same;

Our **I-MAP Parameterization** involves the free parameters of a **specific** minimal I-MAP

I-MAP constraints: Example

Let's consider again the 3-nodes DAG and its minimal I-MAP for π_2:



For the minimal I-MAP to represent the same model as the $\mathcal{D}$, it must hold that

$$\tilde{\boldsymbol{L}}_{32} = -\frac{\tilde{\boldsymbol{L}}_{13}\tilde{\boldsymbol{D}}_{22}}{\tilde{\boldsymbol{L}}_{12}\tilde{\boldsymbol{D}}_{33}}$$

The Optimal I-MAP: Definition

Let

- $\mathcal{D}$ be such that the treatment node X_t is an ancestor of the outcome node X_o;
- π^* be an ordering of the variables such that (i) π^* is topological of $\mathcal{D}$ and (ii) X_t and X_o are separated only by their mediators;
- $\tilde{\pi}$ be the ordering obtained from a topological ordering π and moving all the mediators of X_t and X_o after X_o, preserving their relative ordering relation

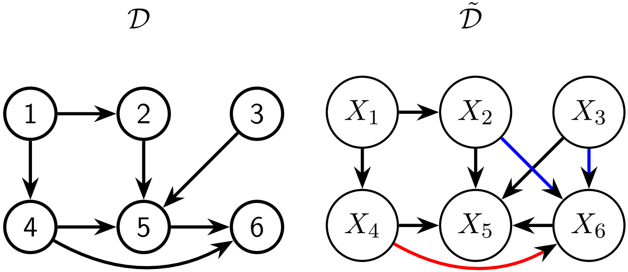
We call $\tilde{\mathcal{D}} := \mathcal{D}^*_{\tilde{\pi}}$ the **optimal I-MAP** of $\mathcal{D}$. It holds that

$$\mathsf{pa}_{\tilde{\mathcal{D}}}(X_o) = \tilde{z} = \mathcal{O}(X_t, X_o, \mathcal{D}) \quad \text{and} \quad \tilde{\boldsymbol{L}}_{\mathsf{pa}_o(\tilde{\mathcal{D}}),o} = -(\hat{\boldsymbol{\Sigma}}_{\tilde{z},\tilde{z}})^{-1}\hat{\boldsymbol{\Sigma}}_{\tilde{z},o}$$

The parameter of the edge $t \to o \in \tilde{\mathcal{D}}$ corresponds to the **total causal effect** of interest;

The Optimal I-MAP: Example

.



When interested in τ_{46} the Optimal I-MAP includes the **best valid adjustment set** $z^* = \{4, 2, 3\}$ as the parent set of the outcome node, and $\tilde{\boldsymbol{L}}_{46}$ is the corresponding parameter in the I-MAP parameterization;

I-MAP Parameterization: Further technical results

We fully characterise the connection between the canonical Gaussian DAG Model parameterization and our I-MAP parameterization.
We developed a **fast** procedure that transform the parameters $(\boldsymbol{L}, \boldsymbol{D})$ of a Gaussian DAG Model into the corresponding parameters of the I-MAP parameterization, using only $m := |\text{med}_{\mathcal{D}}(t, o)|$ steps of **Gaussian elimination**.
The same procedure can be also used in computer algebra systems to derive the **constraints** that hold among the parameters of the I-MAP parameterization.

I-MAP Parameterization: Further technical results

We fully characterise the connection between the canonical Gaussian DAG Model parameterization and our I-MAP parameterization.
We developed a **fast** procedure that transform the parameters $(\boldsymbol{L}, \boldsymbol{D})$ of a Gaussian DAG Model into the corresponding parameters of the I-MAP parameterization, using only $m := |\text{med}_{\mathcal{D}}(t, o)|$ steps of **Gaussian elimination**.
The same procedure can be also used in computer algebra systems to derive the **constraints** that hold among the parameters of the I-MAP parameterization.

In practice, identifying the constraints of the I-MAP parameterization remains quite inconvenient;

Bayesian Model Specification on the I-MAP

The statistical model can be written as

$$\boldsymbol{X} \mid \boldsymbol{\Sigma}_{\mathcal{D}}, \mathcal{D} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathcal{D}})$$
$$\boldsymbol{\Sigma}_{\mathcal{D}} = \tilde{\boldsymbol{L}}^{-T} \tilde{\boldsymbol{D}} \tilde{\boldsymbol{L}}^{-1}$$

The parameter prior must then be defined only on the free-parameters of the I-MAP parameterization. For instance, in the simplest case, we may specify for any $j \in [q]$ and any pair i, j such that $\boldsymbol{L}_{ij}$ is a free parameter:

$$\tilde{\boldsymbol{L}}_{ij} \mid \mathcal{D} \sim \mathcal{N}(0, \sigma_0^2)$$
$$\tilde{\boldsymbol{D}}_{jj} \mid \mathcal{D} \sim \text{Inv-Gamma}(\nu_0/2, \nu_0\sigma_0/2)$$

The DAG prior can be specified as before

$$p(\mathcal{D}) =\propto \omega^{|\mathcal{S}_{\mathcal{D}}|}(1-\omega)^{\frac{q(q-1)}{2} - |\mathcal{S}_{\mathcal{D}}|}$$

Bayesian Model Specification: Comments

Because of the constraints holding among the parameters of the I-MAP parameterization, computing the marginal likelihood becomes a difficult task;

Solution: We propose to use a Laplace approximation centered on the MLE of the I-MAP $(\tilde{\boldsymbol{L}}^{\mathsf{MLE}}, \tilde{\boldsymbol{D}}^{\mathsf{MLE}})$, leading to a **BIC**-style approximation of the marginal likelihood;

As deriving the said constraints is computationally costly, even evaluating the likelihood and deriving the MLE can be difficult

We thus first estimate the MLE of the DAG parameters $(\boldsymbol{L}, \boldsymbol{D})$ and then use our Gaussian elimination procedure to derive $(\tilde{\boldsymbol{L}}^{\mathsf{MLE}}, \tilde{\boldsymbol{D}}^{\mathsf{MLE}})$ by the **invariance property** of MLE;

Similarly, sampling from the posterior $p(\tilde{\boldsymbol{L}}, \tilde{\boldsymbol{D}} | \boldsymbol{X}, \mathcal{D})$ occurs via Importance Sampling, by first sampling from a DAG-Wishart distribution and then resampling the transformed samples with weights given by the two priors ratio;

Conclusions

- We propose an alternative parameterization of Gaussian DAG Models, the I-MAP parameterization, that always contains a causal effect of interest as a single parameter;
- Our parameterization allows for full control over the specification over the prior distribution on the causal effect of interest in a Bayesian setting;
- As the I-MAP parameterization involves constraints among the parameters, we developed both a procedure to (more easily) derive those constraints, and a procedure to quickly transform the parameters of a Gaussian DAG model in the typical parameterization to the ones of the I-MAP parameterization;
- The second procedure can be used to both derive an approximation to the marginal likelihood and to sample from the posterior distribution over the parameters.
- As numerical experiments are still unsatisfactory, we are currently exploring new possibilities for the approximation of the marginal likelihood

Thank you!

alessandro.mascaro@upf.edu

https://alesmascaro.github.io

References

- Ben-David, E., Li, T., Massam, H., & Rajaratnam, B. (2011). High dimensional Bayesian inference for Gaussian directed acyclic graph models. arXiv preprint arXiv:1109.4371.
- Cao, X., Khare, K., & Ghosh, M. (2019). Posterior graph selection and estimation consistency for high-dimensional Bayesian DAG models. Annals of Statistics;
- Henckel, L., Perković, E., & Maathuis, M. H. (2022). Graphical criteria for efficient total effect estimation via adjustment in causal linear models. Journal of the Royal Statistical Society Series B: Statistical Methodology, 84(2), 579-599.
- Maathuis, M. H., Kalisch, M., & Bühlmann, P. (2009). Estimating high-dimensional intervention effects from observational data. The Annals of Statistics
- Peluso, S., & Consonni, G. (2020). Compatible priors for model selection of high-dimensional Gaussian DAGs. Electronic Journal of Statistics