Bayesian model selection allows for flexible discovery of bivariate causal relations

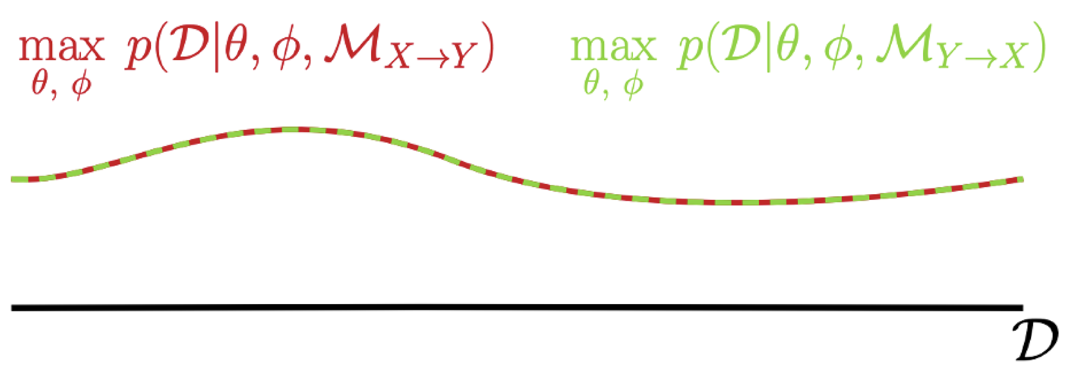
Bivariate Causal Discovery using Bayesian Model Selection

Anish Dhir (Imperial College London), Sam Power (Univ. of Bristol), Mark van der Wilk (Univ. of Oxford)

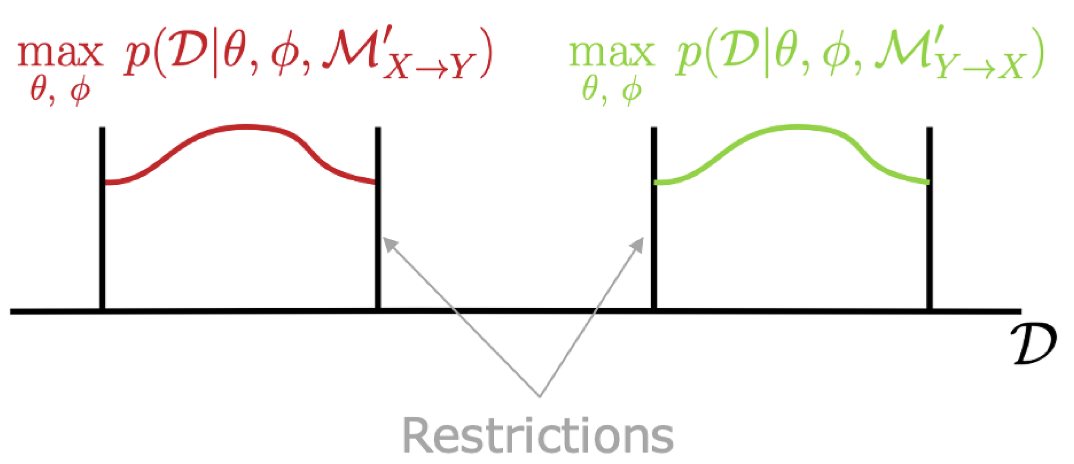
Introduction

Given variables can we find the causal direction?

Problem: $P(Y|X)P(X) = P(X|Y)P(Y)$. Maximising likelihood cannot identify the causal direction.



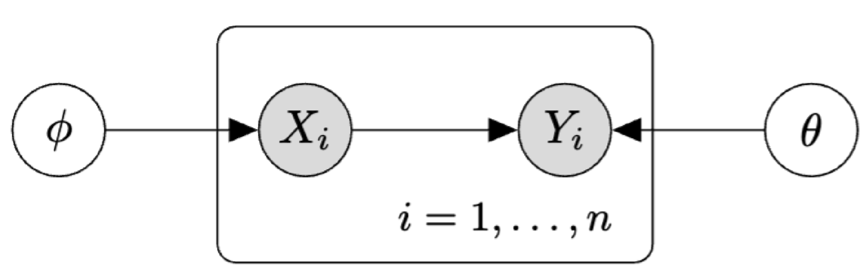
Previous solutions: Restrict model class to allow maximum likelihood to identify, but restricts the datasets you can model!



Idea: Use Bayesian Model selection
Each causal direction is a separate model

$$p(\mathcal{M}_{X\to Y}|\mathcal{D}) = \frac{p(\mathcal{D}|\mathcal{M}_{X\to Y})p(\mathcal{M}_{X\to Y})}{p(\mathcal{D})}$$

$p(\mathcal{D}|\mathcal{M}_{X\to Y})$ is the **Marginal Likelihood**



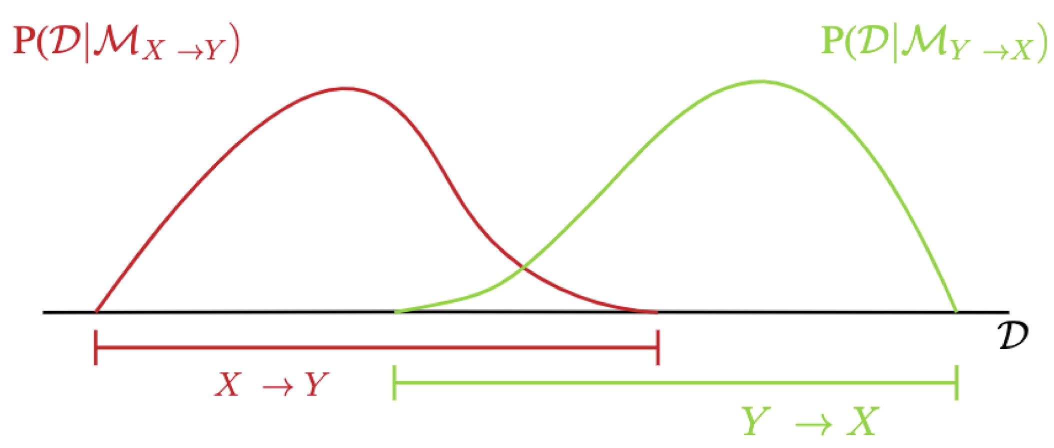
Marginal likelihood $p(x, y|\mathcal{M}_{X\to Y})$**:**

$$\int p(y|x,\theta)p(\theta)d\theta \int p(x|\phi)p(\phi)d\phi$$

- Terms in the causal factorisation are parametrised independently.
- Parameters are independent $\theta \perp\!\!\!\perp \phi$
- Assumptions $p(\theta)p(\phi)$ are required

Assumptions:

- Assumptions can reconstruct previous known identifiability with maximum likelihood.
- Can also identify in the case of **flexible** models, where maximum likelihood cannot.
- Marginal likelihood must sum to 1 over datasets. This **restricts** how well our model can explain all datasets.



Correctness:

- Can quantify probability of finding the correct model

$$P(E) = \frac{1}{2}(1 - \underbrace{\text{TV}[P_{\mathcal{D}}(\cdot|\mathcal{M}_{X\to Y}), P_{\mathcal{D}}(\cdot|\mathcal{M}_{Y\to X})]}_{\text{Total variation between model densities}})$$

- Total variation of 1 corresponds to completely identifiable case
- Under model misspecification our method is not brittle

$$|\underbrace{\Pi(\text{Error})}_{\text{True probability of error}} - \underbrace{P(\text{Error})}_{\text{Model probability of error}}| \leq \text{TV}[\Pi_{\mathcal{D}}(\cdot|X\to Y), P_{\mathcal{D}}(\cdot|\mathcal{M}_{X\to Y})]$$

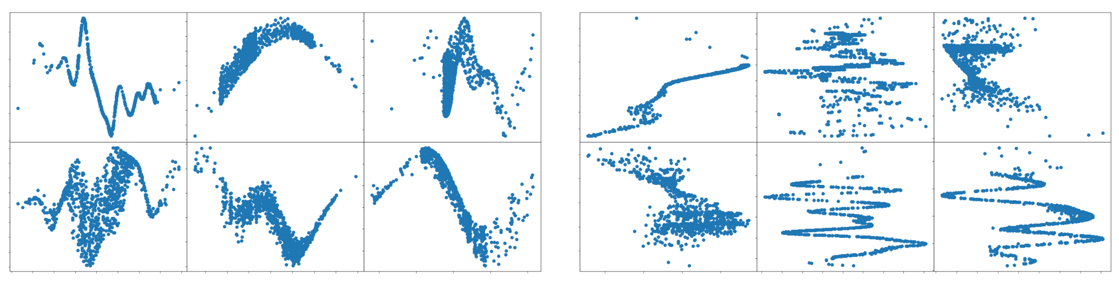
GPLVMs

We want to use a **flexible** Bayesian model with the ability to model:

- Non Gaussian likelihoods
- Heteroscedastic noise

We use **Gaussian Process Latent variable models**

$$f(\cdot,\cdot) \sim \mathcal{N}(0, K_\rho(\mathbf{x},\mathbf{w}))$$
$$p(\mathbf{y}|\mathbf{x},\mathbf{f}) = \int \mathcal{N}(\mathbf{f}(\mathbf{x},\mathbf{w}), \sigma^2)p(\mathbf{w})d\mathbf{w}$$



Results

Flexibility of our method allows for good performance across a wide range of data generating assumptions (Metric: AUPRC, higher is better)

Methods	Cha	Multi	Net	Gauss	Tueb
LiNGAM	57.8	62.3	3.3	72.2	31.1
ANM	43.7	25.5	87.8	90.7	63.9
PNL	78.6	51.7	75.6	84.7	73.8
IGCI	55.6	77.8	57.4	16.0	63.1
RECI	59.0	94.7	66.0	71.0	70.5
SLOPPY	60.1	95.7	79.3	71.4	65.3
CGNN	76.2	94.7	86.3	89.3	76.6
GPI	71.5	73.8	88.1	90.2	70.6
CDCI	72.2	96.0	94.3	**91.8**	58.7
GPLVM	**81.9**	**97.7**	**98.9**	89.3	**78.3**

@dhir_anish @markvanderwilk

Imperial College London