Multi-Domain Causal Discovery in Bijective Causal Models

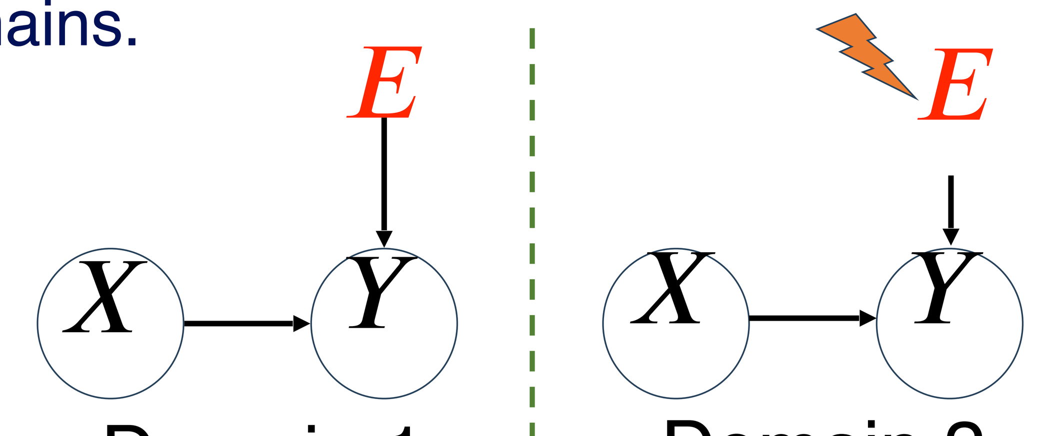
Kasra Jalaldoust[1], Saber Salehkaleybar[2], Negar Kiyavash[3]

1 Columbia University, 2 Leiden University, 3 EPFL

• Problem Statement

- A **multi-domain** setting where the underlying causal mechanisms are invariant.

- The **distribution of exogenous noise varies** between domains.



Bivariate case:
$Y := f(X, E)$

- $\mathbb{P}^i$: Probability measure over the observed variables in i-th domain.

- Having access to m domains: $\mathcal{M} = \{\mathbb{P}^1, \ldots, \mathbb{P}^m\}$.

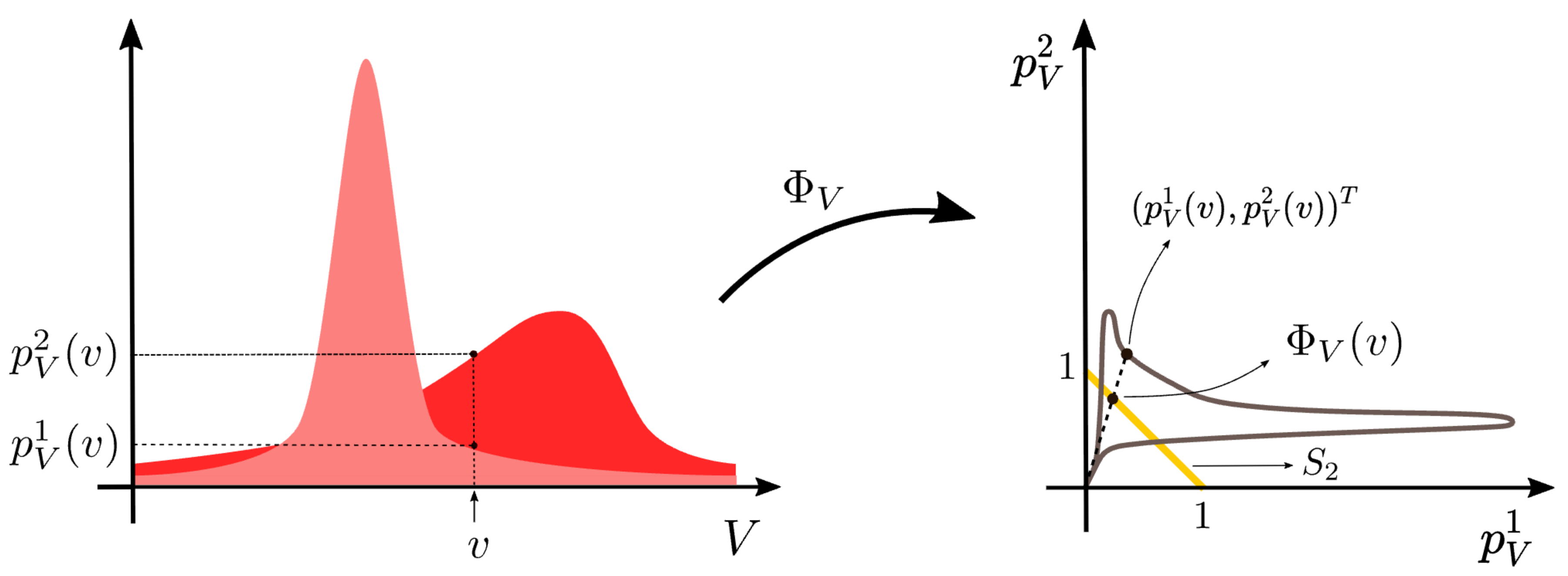
• Assumptions

Definition: A bijective function $g: \mathcal{A} \to \mathcal{B}$ is a **diffeomorphism** if it is differentiable and also has a differentiable inverse.

Assumption 1 (Bijective Generation Mechanisms (BGM)): For every $x \in \mathcal{X}$, we assume that the fixed-cause functional $f(x, \cdot)$ is a diffeomorphism.

Assumption 2 (Causal Sufficiency): E and X are independent under every $\mathbb{P}^i$.

• Main Contribution



Definition: For a random variable (RV) V, the density-vectorization associated with V is defined as, $\Phi_V(v) = h/\|h\|_1$ where $h = (p_V^1(v), \ldots, p_V^m(v))$.

Let $\tilde{Y}_x$ be $Y \mid X = x$. We define $\Gamma_{X\to Y} := \Phi_{\tilde{Y}_X}(\tilde{Y}_X)$.

Theorem: Under Assumptions 1 and 2, for continuous RVs X and Y, the causal diagram $X \not\to Y$ is rejected if $\Gamma_{X\to Y} \perp\!\!\!\perp X$ for some $\mathbb{P}^i \in \mathcal{M}$.

• More on the Identifiability Result

BGM extends the classical noise models:

Model	Functional Form
Additive noise model	$f(X, E) = g(X) + E$
LiNGAM	$f(X, E) = \beta.X + E$
Post-nonlinear model	$f(X, E) = h(g(X) + E)$
location-scale noise model	$f(X, E) = g(X) + h(X)E$

- The identifiability result can be extended to: 1- the discrete-valued E, 2- the multivariate case.

- The main assumption in the multivariate case: For each value of X_i's parents (such as pa_i), $f_i(pa_i, .)$ is a diffeomorphism.

• Estimation Method

Sampling from $\Gamma_{X\to Y}$:

- Estimate $p^i_{Y|X}$ in each domain i.

- Sample x from $p^i(x)$ and form the vector $w := (p^1_{Y|X}(x), p^2_{Y|X}(x), \ldots, p^m_{Y|X}(x))$ and then obtain $\gamma := \frac{w}{\|w\|_1}$ as a sample of $\Gamma_{X\to Y}$.

H1: $\hat{PA} = \cup_{S\subseteq A: L(S)>c} S$ ($L(S)$ is the minimum p-value of testing $\Gamma_{S\to V} \perp\!\!\!\perp S$, and A is the neighboring nodes of V).

H2: Similar to **H1** with some prior knowledge on the size of parent set.

• Experiments

Bivariate case:

The noise model: $Y = f(X) + g(X)E$

#samples=1000, f(X) is linear, g(X) is selected randomly from $\sqrt{|\beta^T X| + 1}$ or $\log(|\beta^T X| + 2)$.

Alg.	H1	ICP	NLICP	LiNGAM	MC	IB	LRE	CD-NOD	CdF	MGL
Acc.	82%	43%	51%	66%	62%	59%	43%	10%	49%	63%

Accuracy of detecting the causal direction