1 Introduction

The present work is a simple explanation and implementation of the invariant principle proposed by Jonas Peters et al. in *Causal inference using invariant prediction: identification and confidence intervals* (2015). In it, we will focus on structural linear equation models with additive Gaussian noise.

2 Invariance Assumption

Let ε be a set of experimental settings (henceforth environments), observational or interventional. Where for each environment e, we have a joint distribution over (Y^e, X^e), with $Y^e \in \mathbb{R}$ (the target variable), $X^e \in \mathbb{R}^p$ (the p predictor variables).

We will assume that[1],

$$\exists_{S^* \text{ subset of variables}} \text{ s.t. } \forall_{e \in \varepsilon}\ Y^e | X^e_{S^*} = x \text{ is invariant} \tag{1}$$

Furthermore, if we assume that the underlying distributions are ruled by a set of structural linear equations with additive Gaussian noise (SLG). The assumption translates to

$$\exists_{S^* \text{ subset of variables}, \gamma^* \in \mathbb{R}^{|S^*|}, \mu^* \in \mathbb{R}} \text{ s.t. } \forall_{e \in \varepsilon}\ Y^e = \mu^* + \gamma^* X^e_{S^*} + \epsilon^e \text{ with } \epsilon^e \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \tag{2}$$

where $0 \leq |S^*| \leq p$ denotes the size of the subset S^*, and we assume that there are no hidden confounder to $X^e_{S^*}$ and Y^e for all environments.

Particularly, on a SLG, $S^* = \text{PARENTS}(Y)$ satisfies assumption 2[2].

Please note that by including the intercept in $X^e_{S^*}$ we can rewrite the assumption as

$$\exists_{S^* \text{ subset of variables}, \gamma^* \in \mathbb{R}^{|S^*|+1}} \text{ s.t. } \forall_{e \in \varepsilon}\ Y^e = \gamma^* X^e_{S^*} + \epsilon^e \text{ with } \epsilon^e \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \tag{3}$$

[1] An alternative statement would be $\exists_{S^* \text{ subset of variables}} \text{ s.t. } \forall_{e,f \in \varepsilon} Y^e | X^e_{S^*} = x \overset{d}{=} Y^f | X^f_{S^*} = x$

[2] Relating to the original assumption 1. Note that for a SLG, environment e and $S^* = \text{PARENTS}(Y)$, $P(Y^e | do(X^e_{S^*} := x)) = P(Y^e | X^e_{S^*} = x)$. Thus, it is correct to ask for the invariance of $P(Y^e | X^e_{S^*} = x)$ for both observational and interventional environments.

3 Plausible Causal Predictors

Our goal is to find the parents of the target variable. However, depending on the characteristic of the environments, we might be able to single them out with high confidence or reduce the possibilities to a subset of the parents (including the empty set).

3.1 Algorithm

We call a subset of variables S plausible causal predictors if $H_{0,S,\varepsilon}$ holds true

$$H_{0,S,\varepsilon,\gamma} : \forall_{e \in \varepsilon}\ Y^e = \gamma^* X_S^e + \epsilon^e \text{ with } \epsilon^e \overset{iid}{\sim} \mathcal{N}(0, \sigma^2) \tag{4}$$

$$H_{0,S,\varepsilon} : \exists \gamma \in \mathbb{R}^{|S|+1} \text{ s.t. } H_{0,S,\varepsilon} \text{ holds true .} \tag{5}$$

The identifiable causal predictors under ε are defined as

$$S_\varepsilon = \cap_{S : H_{0,S,\varepsilon} \text{ holds true}}\ S \tag{6}$$

Moreover, since $H_{0,S^*,\varepsilon}$ is true $S_\varepsilon \subseteq S^*$. That is, the identifiable causal predictors are a subset of the parents of Y.

Furthermore, the identifiable causal coefficients can be defined as

$$\Gamma_{S,\varepsilon} = \{\gamma \in \mathbb{R}^{|S|+1} \text{ s.t. } H_{0,S,\varepsilon,\gamma} \text{ holds true } \} \tag{7}$$

$$\Gamma_\varepsilon = \cup_{\text{all possible } S \text{ subsets}} \Gamma_{S,\varepsilon} \tag{8}$$

Having established the above definitions, we can easily create an algorithm for estimating the causal predictors

1. $\forall_{S \text{ subset }, e \in \varepsilon}$ test if $H_{0,S,\varepsilon}$ holds at level α[3].

2. Estimate the plausible causal predictors

$$\hat{S}_\varepsilon = \cap_{S: H_{0,S,\varepsilon} \text{ holds at level } \alpha} S \tag{9}$$

3. Estimate the confidence intervals for the plausible causal predictors' coefficients

$$\hat{\Gamma}_{S,\varepsilon} = \{(1-\alpha) \text{ confidence interval for } \hat{\gamma} \in \mathbb{R}^{|S|+1} \text{ s.t. } H_{0,S,\varepsilon,\hat{\gamma}} \text{ holds true at level } \alpha\} \tag{10}$$

$$\hat{\Gamma}_\varepsilon = \cup_{\text{all possible } S \text{ subsets}} \hat{\Gamma}_{S,\varepsilon} \tag{11}$$

It is very easy to show[4] that $P[\hat{S}_\varepsilon \subseteq S^*] \geq 1-\alpha$ and $P[\gamma^* \in \hat{\Gamma}_\varepsilon] \geq 1-2\alpha$.

[3] In the sense that $\sup_{P: H_{0,S,\varepsilon} \text{ is true}} P[H_{0,S,\varepsilon} \text{is rejected}] \leq \alpha$

[4] The reader is referred to the original paper for the short proof.

3.2 Implementation for SLG

In this section, a particular implementation of the above-mentioned general algorithm is proposed.

In the case of an SLG, given an environment e joint distribution (Y^e, X^e). The causal coefficients γ^* exactly match the regression coefficients[5].

Let β_S^e be the regression coefficient for the environment $e \in \varepsilon$ and subset S, and $\sigma_S^e = \sqrt{E[Y^e - X_S^e \beta_S^e]^2}$ be the standard deviation of the residuals. We can define the following null hypothesis

$$\hat{H}_{0,S,\varepsilon} : \exists_{\beta_S \in \mathbb{R}^{|S|+1}, \sigma_S \in \mathbb{R}_+} \forall_{e \in \varepsilon} \; \beta_S^e = \beta_S \text{ and } \sigma_S^e = \sigma_S \tag{12}$$

Although, $\hat{H}_{0,S,\varepsilon}$ is weaker than $H_{0,S,\varepsilon}$, it is true whenever $H_{0,S,\varepsilon}$ is true.

Furthermore, we redefine

$$\hat{\Gamma}_{S,\varepsilon} = \begin{cases} (1-\alpha) \text{ confidence interval for } \beta_S & \text{if } \hat{H}_{0,S,\varepsilon} \text{ holds true at level } \alpha \\ \emptyset & \text{otherwise} \end{cases} \tag{13}$$

Based on these definitions, we can implement the following algorithm.

[5] Since the linear regression coefficients correspond to the ML coefficients of the Gaussian distribution.

1. $\forall_{S \text{ subset}}$

1.1 $\forall_{e\in\varepsilon}$

1.1.1 Given that we have n_e observations from the environment e, we define $\mathbb{X}_S^e \in \mathbb{R}^{n_e\times(|S|+1)}$ as the design matrix that includes the intercept, and $\mathbb{Y} \in \mathbb{R}^{n_e\times 1}$ as the observed target values.

1.1.2 Moreover, let $n_{-e} = \sum_{\substack{f\in\varepsilon \\ f\neq e}} n_f$ and define the design matrix $\mathbb{X}_S^{-e} \in \mathbb{R}^{n_{-e}\times(|S|+1)}$ and $\mathbb{Y}^{-e} \in \mathbb{R}^{n_{-e}\times 1}$.

1.1.3 Let $\hat{\beta}_S^{-e}$ be the linear regression coefficient using $\mathbb{Y}^{-e}$ and $\mathbb{X}_S^{-e}$, and define $D = Y^e - X_S^e \hat{\beta}_S^{-e}$ to be the residuals when predicting the actual observations at environment e.

1.1.4 Using the Chow test for structural change, we test the residuals under this setting (i.e. holding out the observations from environment e).
Let p_S^e be the p-value for

$$\frac{D^T \Sigma_D^{-1} D}{(\hat{\sigma}_S^{-e})^2 n_e} \sim F(n_e, n_{-e} - (|S|+1)) \tag{14}$$

1.2 Set $p_S = \min_{e\in\varepsilon} p_S^e$ and reject $\hat{H}_{0,S,\varepsilon}$ if p_S is below $\frac{\alpha}{|\varepsilon|}$ (Bonferroni's correction).

2. Estimate the plausible causal predictors

$$\hat{S}_\varepsilon = \cap_{S:\hat{H}_{0,S,\varepsilon} \text{ was not rejected}} S \tag{15}$$

3. Estimate the confidence intervals for the plausible causal predictors' coefficients

3.1 Let $n = \sum_{e\in\varepsilon} n_e$, and define the design matrix of all data $\mathbb{X}_{\hat{S}_\varepsilon} \in \mathbb{R}^{n\times|\hat{S}_\varepsilon|+1}$ and the target variable $\mathbb{Y} \in \mathbb{R}^{n\times 1}$. Based on this, compute the regression coefficients $\hat{\beta}_{\hat{S}_\varepsilon}$ and the standard deviation of the residuals $\hat{\sigma}_{\hat{S}_\varepsilon}$ over all the data.

3.2 Since the regression coefficients follow a t-distribution with $n-(|S|+1)$ degrees of freedom, we define the $(1-\alpha)$ confidence intervals[6] for the plausible causal predictors as

$$\hat{\beta}_{\hat{S}_\varepsilon} \pm \left(t_{1-\frac{\alpha}{2(|S|+1)}, n-(|S|+1)}\right) \hat{\sigma}_{\hat{S}_\varepsilon} \text{diagonal}((\mathbb{X}_{\hat{S}_\varepsilon}^T \mathbb{X}_{\hat{S}_\varepsilon})^{-1}) \tag{16}$$

[6] Using a rectangular confidence region, i.e. assuming the that the coefficients are uncorrelated.

Please note that we must assume that all environments' design matrix $\mathbb{X}^e$ are full rank in order to be able to take the inverse of the corresponding covariance matrices.

It is shown in the original work that if an interventional environment, where only one variable is intervened (i.e. do-intervention, hard intervention)[7], is provided for each one of the variables (except the target Y). Then, the true causal predictors (i.e. parent of Y) are identifiable.

An implementation of the algorithm and some examples can be found in the file `causality.R`.

[7] However if variable X_j is intervened, the value cannot be set to $E[X_j]$