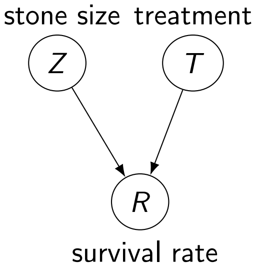
On the Relationship Between Equivariant Predictive Models and Structural Causal Model Identification

Grace Yin
Department of Statistics
University of British Columbia
MSC Presentation

April 19, 2022

Introduction

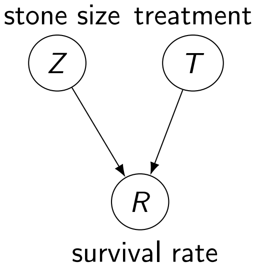
Structural Causal Model (SCM)



$$\begin{cases} Z := f_1(N_Z) \\ T := f_2(N_T) \\ R := f_3(Z, T, N_R) \\ N_Z, N_T, N_R \text{ are jointly independent} \\ \text{noise terms} \end{cases}$$

Introduction

Structural Causal Model (SCM)



$$\begin{cases} Z := f_1(N_Z) \\ T := f_2(N_T) \\ R := f_3(Z, T, N_R) \\ N_Z, N_T, N_R \text{ are jointly independent} \\ \text{ noise terms} \end{cases}$$

Structural Causal Model (SCM) [2]

Definition

A structural causal model (SCM) $\mathfrak{C} := (G, S, P_N)$ consists of a collection S of d (structural) assignments

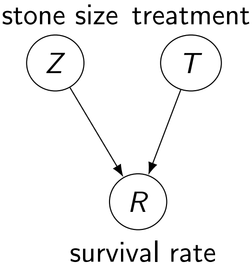
$$X_j := f_j(PA_j, N_j),\ j = 1, \cdots, d$$

where $PA_j \subset \{X_1, \cdots, X_d\} \setminus \{X_j\}$ are parents of X_j and $P_N = P_{N_1, \cdots, N_d}$ is the joint (product) distribution over the noise variables which are assumed to be jointly independent. $G = (V, E)$ is the graph contains vertices and edges.

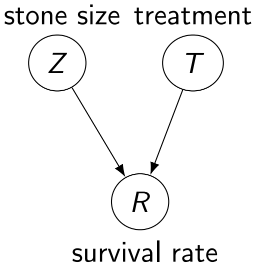
Structural Causal Models (SCM)

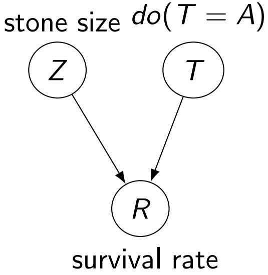
intervention: $do(\cdots)$: sets the variable value, without changing other nodes

treatment: $T = A$ or $T = B$



Structural Causal Models (SCM)
intervention: $do(\cdots)$: sets the variable value, without changing other nodes
treatment: $T = A$ or $T = B$





Introduction

Question: Suppose there are two candidate SCMs, how can we identify which one is correct?

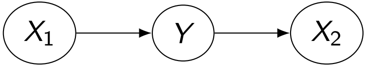


Figure: SCM 1

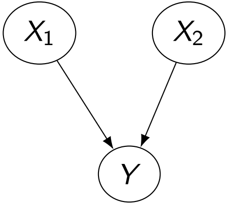


Figure: SCM 2

Introduction

- Potential solution: conditional independence test
  - High computational costs [1]
- $X \rightarrow Y$
  - the functional assignment: $Y := f(X, E) = \beta_0 + \beta_1 X + E$
  - has constant risk for $X' \leftarrow X + a, a \in \mathbb{R}(+)$
- Our approach: apply constant risk theorem
  - across interventions described by the action of a group

Introduction

- Potential solution: conditional independence test
  - High computational costs [1]
- $X \longrightarrow Y$
  - the functional assignment: $Y := f(X, E) = \beta_0 + \beta_1 X + E$
  - has constant risk for $X' \leftarrow X + a, a \in \mathbb{R}(+)$
- Our approach: apply constant risk theorem
  - across interventions described by the action of a group

Introduction

- Potential solution: conditional independence test
  - High computational costs [1]
- $X \longrightarrow Y$
  - the functional assignment: $Y := f(X, E) = \beta_0 + \beta_1 X + E$
  - has constant risk for $X' \leftarrow X + a, a \in \mathbb{R}(+)$
- Our approach: apply constant risk theorem
  - across interventions described by the action of a group

Background and Notation

joint distribution of (X, Y):

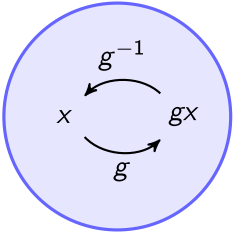
$$\tilde{P} = P \otimes Q_x,$$

P: marginal distribution on $(\mathbf{X}, \mathcal{X})$
Q_x: conditional distribution (Markov kernel) from $(\mathbf{X}, \mathcal{X})$ into $(\mathbf{Y}, \mathcal{Y})$
risk function is defined as

$$R(\tilde{P}, \rho) = \int_{\mathcal{X} \times \mathcal{Y} \times Z} P(dx) Q_x(dy) \rho_x(dz) L(y, z).$$

ρ : decision procedure $\mathbf{X} \times \mathcal{Z} \to [0, 1]$ where $(\mathbf{Z}, \mathcal{Z})$ is the decision space
L: loss function $\mathcal{Y} \times \mathcal{Z} \to [-\infty, \infty)$.

Background and Notation



- $\mathcal{G}$: a group acting measurably on **X** and **Y**
- $g \in \mathcal{G}$: a group action, $\Phi(g, x) = gx$
- **conditional shift**: $g_x \tilde{P} = (P \circ g_x^{-1}) \otimes Q_x$ for $g_x \in \mathcal{G}$

Background and Notation

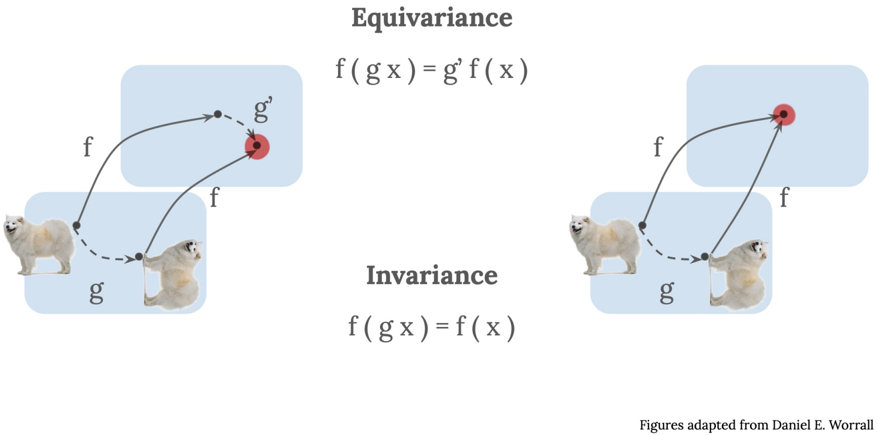


Figure: Equivariance and Invariance map [3]

Constant Risk Theorem

- equivariant markov kernel:
$$Q_{gx} = Q_x \circ g^{-1}$$
- invariant loss function:
$$L(gy, gz) = L(y, z)$$
- equivariant decision procedure ρ:
$$\rho_{gx} = \rho \circ g^{-1}$$

Theorem

For an invariant loss function L, the risk of a decision procedure ρ is constant under the conditional shift $g_x \tilde{P}$ for any group action $g_x \in \mathcal{G}$, if ρ is equivariant and the kernel Q_x is equivariant. i.e.,
$$\forall g_x \in \mathcal{G},\ R(\rho, g_x \tilde{P}) = R(\rho, \tilde{P})$$

Identify the Structure of SCM

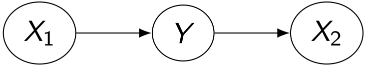


Figure: SCM 1

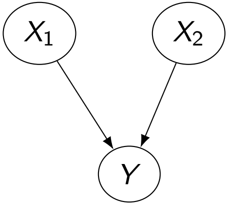


Figure: SCM 2

$$\begin{cases} X_1 \sim \mathcal{N}(0, \sigma^2) \\ Y := \beta_1 X_1 + \varepsilon_y \\ X_2 := \beta_2 Y + \varepsilon_2 \end{cases}$$

with $\varepsilon_2 \sim \mathcal{N}(0,1)$, $\varepsilon_y \sim \mathcal{N}(0, \sigma^2)$

$$\begin{cases} X_1 \sim \mathcal{N}(0, \sigma^2) \\ X_2 := \varepsilon_2 \\ Y := \beta_1 X_1 + \beta_2 X_2 + \varepsilon_y \end{cases}$$

with $\varepsilon_2 \sim \mathcal{N}(0,1)$, $\varepsilon_y \sim \mathcal{N}(0, \sigma^2)$

Simulation Experiments and Discussion

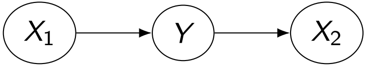


Figure: SCM 1

$$\begin{cases} X_1 \sim \mathcal{N}(0, \sigma^2) \\ Y := \boxed{\beta_1 X_1} + \varepsilon_y \\ X_2 := \boxed{\beta_2 Y} + \varepsilon_2 \end{cases}$$

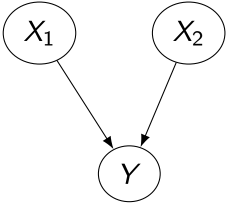


Figure: SCM 2

$$\begin{cases} X_1 \sim \mathcal{N}(0, \sigma^2) \\ X_2 := \varepsilon_2 \\ Y := \boxed{\beta_1 X_1 + \beta_2 X_2} + \varepsilon_y \end{cases}$$

$$\text{intervention}: \begin{cases} X_1' \leftarrow X_1 + g_1 \\ X_2' \leftarrow X_2 + g_2 \end{cases}$$

Simulation Experiments and Discussion

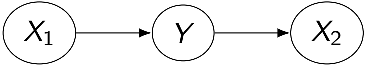


Figure: SCM 1

$$\begin{cases} X_1 \sim \mathcal{N}(0, \sigma^2) \\ Y := \beta_1 X_1 + \varepsilon_y \\ X_2 := \beta_2 Y + \varepsilon_2 \end{cases}$$

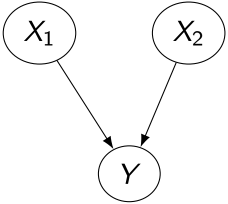


Figure: SCM 2

$$\begin{cases} X_1 \sim \mathcal{N}(0, \sigma^2) \\ X_2 := \varepsilon_2 \\ Y := \beta_1 X_1 + \beta_2 X_2 + \varepsilon_y \end{cases}$$

$$\textbf{intervention} : \begin{cases} X_1' \leftarrow X_1 + g_1 \\ X_2' \leftarrow X_2 + g_2 \end{cases}$$

Simulation Experiments and Discussion

Simulation experiments:

- simulate data from SCM 1
- simulate estimated coefficients in shifting environment 0
- compute risk for three models in different shifting environments

$$\begin{cases} \ell_1 : \hat{y} = \underbrace{\hat{\beta}_0 + \hat{\beta}_1}_{\text{depend on } X_1} x_1 \\ \ell_2 : \hat{y} = \underbrace{\hat{\alpha}_0 + \hat{\alpha}_1}_{\text{depend on } X_2} x_2 \\ \ell_3 : \hat{y} = \underbrace{\hat{\gamma}_0 + \hat{\gamma}_1 x_1 + \hat{\gamma}_2}_{\text{depend on } X_1 \& X_2} x_2 \end{cases}$$

loss function: least square loss function

risk function:

$$R(\rho, P) = \int_{\mathcal{X}\times\mathcal{Y}} L(y, f(x))dP(x,y), \text{ where } \rho = \delta_{f(x)}$$

Simulation Experiments and Discussion

Simulation experiments:

- simulate data from SCM 1
- simulate estimated coefficients in shifting environment 0
- compute risk for three models in different shifting environments

$$\begin{cases} \ell_1 : \hat{y} = \underbrace{\hat{\beta}_0 + \hat{\beta}_1}_{\text{depend on } X_1} x_1 \\ \ell_2 : \hat{y} = \underbrace{\hat{\alpha}_0 + \hat{\alpha}_1}_{\text{depend on } X_2} x_2 \\ \ell_3 : \hat{y} = \underbrace{\hat{\gamma}_0 + \hat{\gamma}_1 x_1 + \hat{\gamma}_2}_{\text{depend on } X_1 \& X_2} x_2 \end{cases}$$

loss function: least square loss function
risk function:

$$R(\rho, P) = \int_{\mathcal{X}\times\mathcal{Y}} L(y, f(x)) dP(x, y), \text{ where } \rho = \delta_{f(x)}$$

Simulation Experiments and Discussion

Simulation experiments:

- simulate data from SCM 1
- simulate estimated coefficients in shifting environment 0
- compute risk for three models in different shifting environments

$$\begin{cases} \ell_1 : \hat{y} = \underbrace{\hat{\beta}_0 + \hat{\beta}_1}_{\text{depend on } X_1} x_1 \\ \ell_2 : \hat{y} = \underbrace{\hat{\alpha}_0 + \hat{\alpha}_1}_{\text{depend on } X_2} x_2 \\ \ell_3 : \hat{y} = \underbrace{\hat{\gamma}_0 + \hat{\gamma}_1 x_1 + \hat{\gamma}_2}_{\text{depend on } X_1 \& X_2} x_2 \end{cases}$$

loss function: least square loss function

risk function:

$$R(\rho, P) = \int_{\mathcal{X}\times\mathcal{Y}} L(y, f(x)) dP(x, y), \text{ where } \rho = \delta_{f(x)}$$

Simulation Experiments and Discussion

Simulation experiments:

- simulate data from SCM 1
- simulate estimated coefficients in shifting environment 0
- compute risk for three models in different shifting environments

$$\begin{cases} \ell_1 : \hat{y} = \underbrace{\hat{\beta}_0 + \hat{\beta}_1}_{\text{depend on } X_1} x_1 \\ \ell_2 : \hat{y} = \underbrace{\hat{\alpha}_0 + \hat{\alpha}_1}_{\text{depend on } X_2} x_2 \\ \ell_3 : \hat{y} = \underbrace{\hat{\gamma}_0 + \hat{\gamma}_1 x_1 + \hat{\gamma}_2}_{\text{depend on } X_1 \& X_2} x_2 \end{cases}$$

loss function: least square loss function

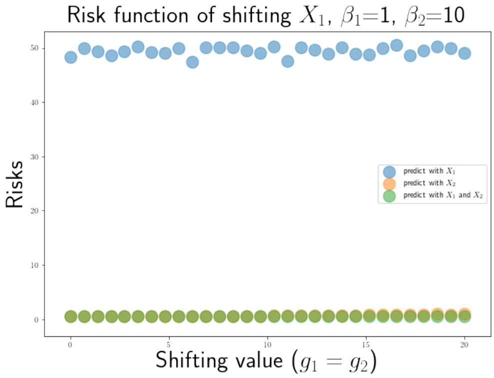
risk function:

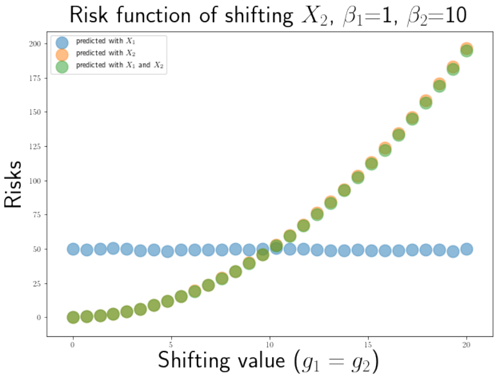
$$R(\rho, P) = \int_{\mathcal{X} \times \mathcal{Y}} L(y, f(x)) dP(x, y), \text{ where } \rho = \delta_{f(x)}$$

Simulation Experiments and Discussion

The simulation results verified our constant risk theorem.

- shifting X_1: constant risk for all three predictive models
- shifting X_2: only ℓ_1 has constant risk

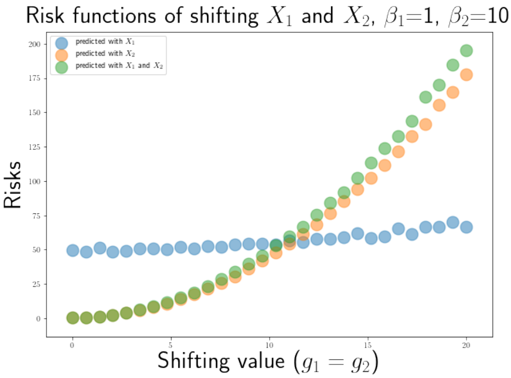




Simulation Experiments and Discussion

Differences between constant risk approach and risk minimization approach

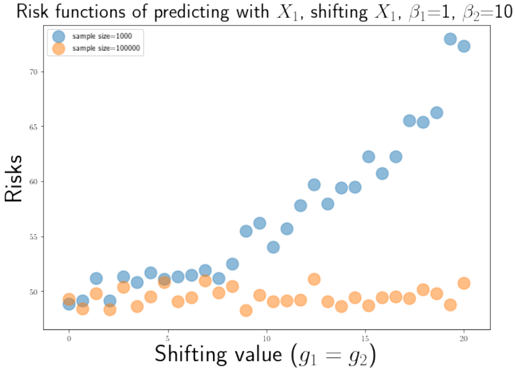
- a crossover among three predictive models when shifting X_1 and X_2



Simulation Experiments and Discussion

Sample size can influence the constant risk results:

- simulated $\hat{\beta}_0$ and $\hat{\beta}_1$ for ℓ_1
- sample size $n_1 = 1000$ vs sample size $n_2 = 100000$
- empirical risks: $\hat{R}(g) \propto (\beta_0 - \hat{\beta}_0)^2 + (x_{1,i} + g)^2(\beta_1 - \hat{\beta}_1)^2$



Potential Extension

Potential directions:

- Apply on other examples of SCMs
- Identify the theoretical interconnection between the risk function among linear regression models
- Implement algorithms
- Explore non-linear predictive models

Reference

[1] J. Peters, P. Bühlmann, and N. Meinshausen.
Causal inference using invariant prediction: identification and confidence intervals.
2015.

[2] J. Peters, D. Janzing, and B. Schlkopf.
Elements of Causal Inference: Foundations and Learning Algorithms.
The MIT Press, 2017.

[3] E. van der Pol.
Geometric deep learning and reinforcement learning, February 2021.

Acknowledgement

Special thanks to :

- my supervisor: Prof. Ben Bloem-Reddy
- the various members of the UBC Department of Statistics

Github QR code:

Thank You for Listening!