Causal change point detection and localization

Shimeng Huang, Rikke Søndergaard, Jonas Peters, Niklas Pfister

Nordstat 2023, Gothenburg, Sweden

June 19, 2023

UNIVERSITY OF COPENHAGEN

Motivation

- Detecting changes in time series data has long been of interest (e.g., Truong et al., 2020).

Motivation

- Detecting changes in time series data has long been of interest (e.g., Truong et al., 2020). Here: detect causal change points. Example:

$$X_i \longrightarrow Y_i$$

$$i = 1, \cdots, \tau$$

X_i your desire of ice cream, Y_i your actual consumption of ice cream, and at $\tau + 1$ you found out that lactase pills are a thing!

Motivation

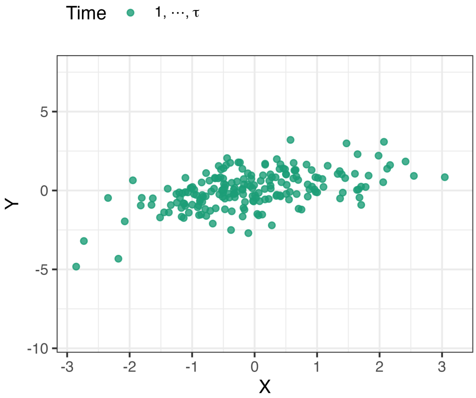
- Detecting changes in time series data has long been of interest (e.g., Truong et al., 2020). Here: detect causal change points. Example:

$$X_i \longrightarrow Y_i \qquad\qquad X_i \longrightarrow Y_i$$

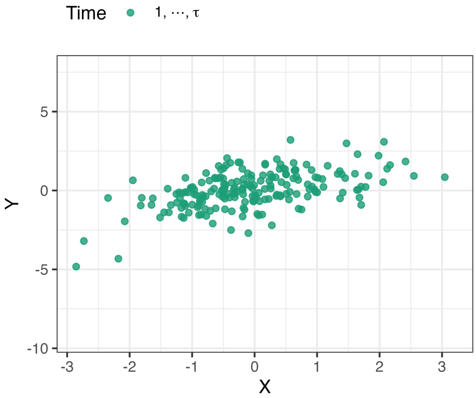
$$i = 1, \cdots, \tau \qquad\qquad i = \tau + 1, \cdots, n$$

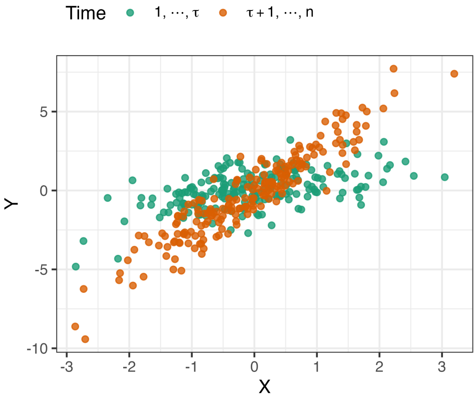
X_i your desire of ice cream, Y_i your actual consumption of ice cream, and at $\tau + 1$ you found out that lactase pills are a thing!

Motivation

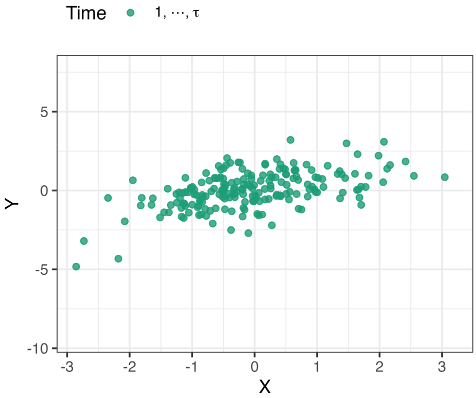


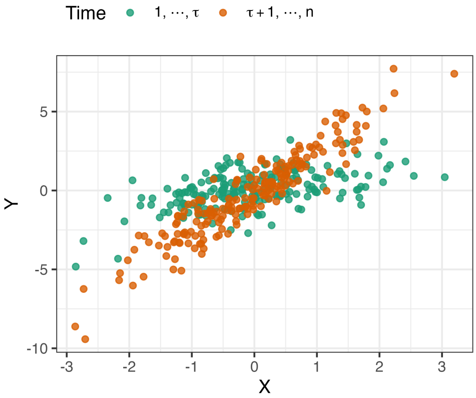
Motivation


Time
$1, \cdots, \tau$
Y
X
5
0
-5
-10
-3
-2
-1
0
1
2
3


Time
$1, \cdots, \tau$
$\tau + 1, \cdots, n$
Y
X
5
0
-5
-10
-3
-2
-1
0
1
2
3

Motivation





Given: data $(X_1, Y_1), \ldots, (X_n, Y_n)$

Ideal output: there is exactly one causal change point and it's located at τ.

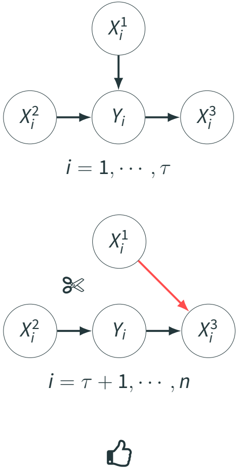
Motivation

- In economics literature, changes in how Y is affected by others are often referred to as structural changes (e.g., Hansen, 2000).

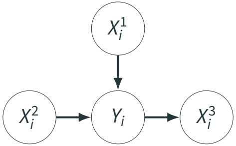
Motivation

- In economics literature, changes in how Y is affected by others are often referred to as structural changes (e.g., Hansen, 2000).
- Here we consider observing **multivariate sequential data** where the causal mechanism affecting a particular variable changes.

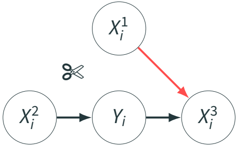
Examples: Are these causal changes?



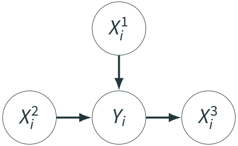
Examples: Are these causal changes?



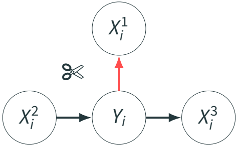
$i = 1, \cdots, \tau$



$i = \tau + 1, \cdots, n$

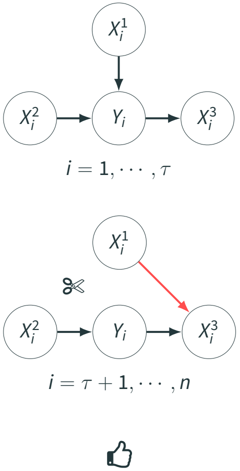


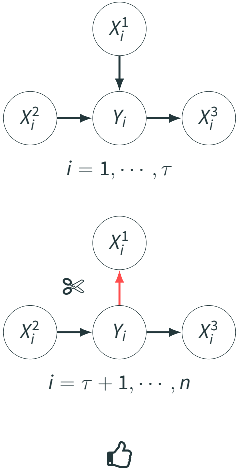
$i = 1, \cdots, \tau$

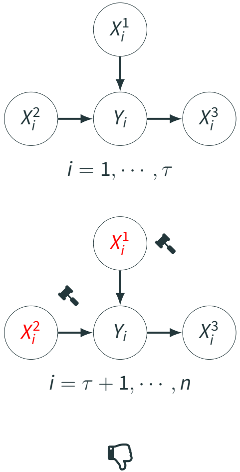


$i = \tau + 1, \cdots, n$

Examples: Are these causal changes?







Setup

Consider the following formal setup:

- Observe $(X_1, Y_1), \dots, (X_n, Y_n)$ independent observations, $X_i \in \mathbb{R}^d$ and $Y_i \in \mathbb{R}$
- The joint distribution of (X_i, Y_i), $\mathbb{P}_i^{X,Y}$, may change over i

Def. Population OLS coefficients and residuals

For all $S \subseteq \{1, \ldots, d\}$ and $i \in \{1, \ldots, n\}$, the population OLS coefficients conditioning on the subset of covariates X_i^S are defined as the vector $\beta_i^{\mathrm{OLS}}(S) \in \mathbb{R}^d$ such that

$$\left(\beta_i^{\mathrm{OLS}}(S)\right)^S = \mathbb{E}[X_i^S (X_i^S)^\top]^{-1} \mathbb{E}[X_i^S Y_i]$$

and $(\beta_i^{\mathrm{OLS}}(S))^j = 0$ for all $j \in \{1, \ldots, d\} \setminus S$. The corresponding OLS residuals is defined as $\epsilon_i(S) := Y_i - X_i^\top \beta_i^{\mathrm{OLS}}(S)$. We denote $\beta_i^{\mathrm{OLS}} = \beta_i^{\mathrm{OLS}}(\{1, \ldots, d\})$ and $\epsilon_i = \epsilon_i(\{1, \ldots, d\})$.

Regression and causal change points

Def. Regression change point (RCP)

A time point $k \in \{2, \dots, n-1\}$ is called a *regression change point* (RCP) if $\beta_k^{\text{OLS}} \neq \beta_{k-1}^{\text{OLS}}$ or $\epsilon_k \overset{d}{\neq} \epsilon_{k-1}$.

Def. Causal change point (CCP)

A time point $k \in \{2, \dots, n-1\}$ is called a *causal change point* (CCP) if for all $S \subseteq \{1, \dots, d\}$, $\beta_k^{\text{OLS}}(S) \neq \beta_{k-1}^{\text{OLS}}(S)$ or $\epsilon_k(S) \overset{d}{\neq} \epsilon_{k-1}(S)$.

We call all RCPs that are not CCPs *non-causal change points* (NCCPs).

Regression and causal change points

Def. Invariant sets

For a time interval $I \in \mathcal{I}$, a set $S \subseteq \{1, \ldots, d\}$ is called an *I-invariant set* if there exists a parameter $\beta \in \mathbb{R}^d$ and a distribution F such that for all $i \in I$,

$$\beta_i^{\mathrm{OLS}}(S) = \beta \quad \text{and} \quad \epsilon_i(S) \sim F.$$

Prop. CCP and RCP

A time point $k \in \{1, \ldots, n\}$ is a CCP if and only if it is an RCP and there does not exist a set $S \subseteq \{1, \ldots, d\}$ that is $\{k-1, k\}$-invariant.

Connection with causal models

Setting. Sequential linear SCM with hidden confounding

Assume we have observed a sequence $(X_1, Y_1), \ldots, (X_n, Y_n) \in \mathbb{R}^d \times \mathbb{R}$ and at each $i \in \{1, \ldots, n\}$, there exists an SCM that generates the observed data,

$$
\begin{aligned}
Y_i &:= \beta_i^\top X_i + g_i(H_i, \varepsilon_i^Y) \\
X_i &:= A_i X_i + \alpha_i Y_i + h_i(H_i, \varepsilon_i^X),
\end{aligned}
$$

where $H_i \in \mathbb{R}^q$ are hidden variables, $\varepsilon_i^X, \varepsilon_i^Y \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$, g_i and h_i are arbitrary measurable functions , and β_i, A_i, and α_i satisfy that the induced graph is directed and acyclic. For all $i \in \{2, \ldots, n-1\}$, the set of parent variables of Y_i is given by $\mathrm{PA}(Y_i) = \{j \in \{1, \ldots, d\} \mid \beta_i^j \neq 0\}$, which includes only observed variables.

Connection with causal models

Prop. CCP in linear SCMs without unobserved confounding

Assume the above setting, let $k \in \{2, \ldots, n-1\}$ be a fixed time point and assume that for all $i \in \{1, \ldots, n\}$ the noise term of Y satisfies that

$$\mathbb{E}[X_i^{\mathrm{PA}(Y_i)} g_i(H_i, \varepsilon_i^Y)] = 0.$$

Then, it holds that

$$k \text{ is a CCP} \quad \Rightarrow \quad \beta_k \neq \beta_{k-1} \text{ or } g_k(H_k, \epsilon_k^Y) \overset{d}{\neq} g_{k-1}(H_{k-1}, \epsilon_{k-1}^Y).$$

Goals

Two goals:

Goals

Two goals:

1. **Detect whether CCPs exist**
   for a time interval $I = \{t, \cdots, t+m\} \subseteq \{1, \cdots, n\}$

$$\mathcal{H}_0(I) : \exists S \subseteq \{1, \ldots, d\} \text{ s.t. } S \text{ is } I\text{-invariant.}$$

Goals

Two goals:

1. **Detect whether CCPs exist**
   for a time interval $I = \{t, \cdots, t+m\} \subseteq \{1, \cdots, n\}$

$$\mathcal{H}_0(I) : \exists S \subseteq \{1, \ldots, d\} \text{ s.t. } S \text{ is } I\text{-invariant.}$$

2. **Estimate where the CCPs are**
   often referred to as “localization” — focus of today.

Causal change point localization

One approach is to localize the CCPs by finding the minima of a loss.

Causal change point localization

One approach is to localize the CCPs by finding the minima of a loss.

For now, consider the case that there is exactly one CCP.

Causal change point localization

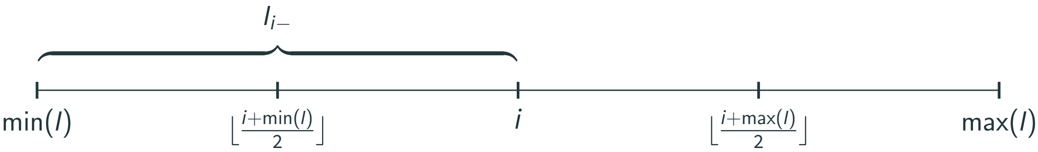
To define the *loss* at any $i \in I$ for any interval $I \in \mathcal{I}$, we introduce some notations.

Causal change point localization

To define the *loss* at any $i \in I$ for any interval $I \in \mathcal{I}$, we introduce some notations.

Causal change point localization

To define the *loss* at any $i \in I$ for any interval $I \in \mathcal{I}$, we introduce some notations.

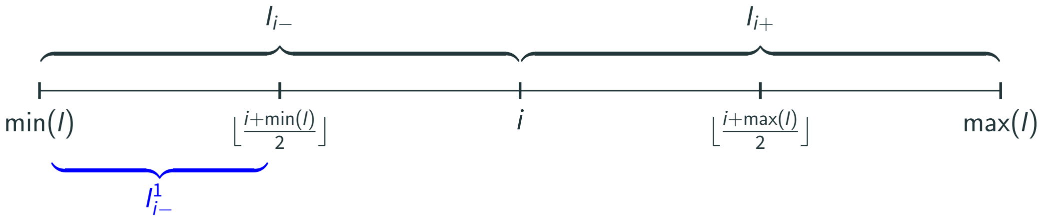


Causal change point localization

To define the *loss* at any $i \in I$ for any interval $I \in \mathcal{I}$, we introduce some notations.

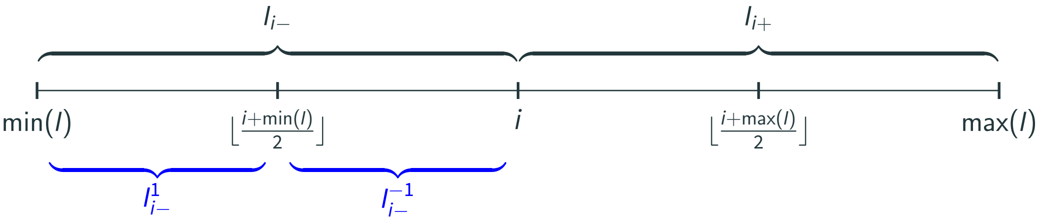
Causal change point localization

To define the *loss* at any $i \in I$ for any interval $I \in \mathcal{I}$, we introduce some notations.



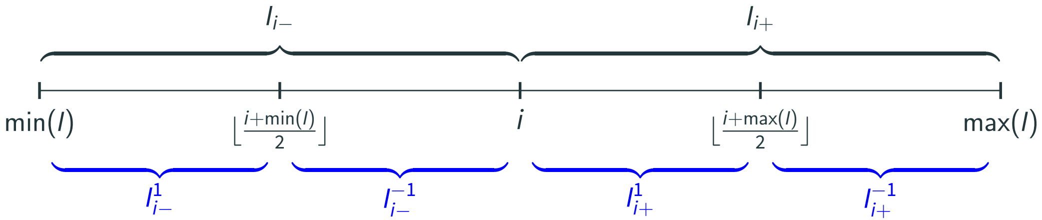
Causal change point localization

To define the *loss* at any $i \in I$ for any interval $I \in \mathcal{I}$, we introduce some notations.



Causal change point localization

To define the *loss* at any $i \in I$ for any interval $I \in \mathcal{I}$, we introduce some notations.



Causal change point localization

We propose the following loss function:

We propose the following loss function:

Def. Invariant loss

Fix $i \in \{2, \ldots, n-1\}$, consider splitting the intervals I_{i-} and I_{i+} into m sub-intervals.

Causal change point localization

We propose the following loss function:

Def. Invariant loss

Fix $i \in \{2, \dots, n-1\}$, consider splitting the intervals I_{i-} and I_{i+} into m sub-intervals. Then for each $r \in \{1, \dots, m\}$ and $\ell \in I_{i-}$, let

$$Z_\ell^{i-}(S, r) = \frac{(Y_\ell - X_\ell^\top \beta^{\text{OLS}}_{I_{i-}^{-r}}(S))^2}{\frac{1}{|I_{i-}^{-r}|} \sum_{\ell \in I_{i-}^{-r}} \mathbb{E}[(Y_\ell - X_\ell^\top \beta^{\text{OLS}}_{I_{i-}^{-r}}(S))^2]},$$

and similarly define $Z_\ell^{i+}(S, r)$ for each $r \in \{1, \dots, m\}$ and $\ell \in I_{i+}$.

Causal change point localization

Def. Invariant loss (Cont.)

Then, the loss at i is defined as

$$\begin{aligned}\mathcal{S}_i^I = \frac{1}{|I|}\Bigg\{ &\min_{S\subseteq\{1,\dots,d\}} \sum_{r=1}^{m} \left| \Big(\sum_{\ell\in I_{i-}^r} \mathbb{E}[Z_\ell^{i-}(S,r)] \Big) - |I_{i-}^r| \right| \\ &+ \min_{S\subseteq\{1,\dots,d\}} \sum_{r=1}^{m} \left| \Big(\sum_{\ell\in I_{i+}^r} \mathbb{E}[Z_\ell^{i+}(S,r)] \Big) - |I_{i+}^r| \right| \Bigg\}.\end{aligned}$$

Property of the loss functions at the population level.

Lem. Invariant loss at population level

Let $I \in \mathcal{I}$. Suppose $\tau \in \{\min(I)+1, \ldots, \max(I)-1\}$ is the only one CCP in I, then $S^I_\tau = 0$.

Property of the estimated fraction.

Conj. Consistency of the estimated fraction

Suppose τ is the only one CCP in $I = \{1, \ldots, n\}$. Let $\hat{\tau}_n = \arg\min_{i \in I} \hat{\mathcal{S}}_i^I$ where $\hat{\mathcal{S}}_i^I$ is the empirical counterpart of $\mathcal{S}_i^I$, and let $\hat{\lambda}_n = \frac{\hat{\tau}_n}{n}$. Denote the true fraction $\lambda = \frac{\tau}{n}$. Then $\hat{\tau}_n \to \tau$ as $n \to \infty$.

Causal change point localization

So far we only considered that there is exactly one CCP. How to localize them when there are multiple CCPs?

Causal change point localization

So far we only considered that there is exactly one CCP. How to localize them when there are multiple CCPs?

It turns out that the loss functions above do not enjoy nice properties when multiple CCPs exist. But the following approaches can be explored:

- Dynamic programming
- Bottom-up approaches such as narrowest-over-threshold (NOT)

Some simulation results

We will compare the invariant loss function with a "naive loss" — at any $i \in I$, the naive loss is the MSE of the concatenated residuals from two OLS to the left and right of i.

Some simulation results

We will compare the invariant loss function with a “naive loss" — at any $i \in I$, the naive loss is the MSE of the concatenated residuals from two OLS to the left and right of i.

We will consider the following settings:

- One CCP only
- One CCP and one NCCP
- No CCP but one NCCP
- One NCCP only

Some simulation results

One CCP only: A simple case is the following two variable SCMs for $i \in \{1, \dots, n\}$

$$\begin{aligned} X_i &:= \varepsilon_i^X \\ Y_i &:= X_i + \varepsilon_i^Y \end{aligned}$$

where $\varepsilon_i^X \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$ for $i \in \{1, \dots, n\}$, $\varepsilon_i^Y \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$ for $i \in \{1, \dots, \frac{n}{2}\}$, and $\varepsilon_i^Y \overset{\text{iid}}{\sim} \mathcal{N}(0, c)$ for $i \in \{\frac{n}{2} + 1, \dots, n\}$ where c controls the amount of change.

Some simulation results

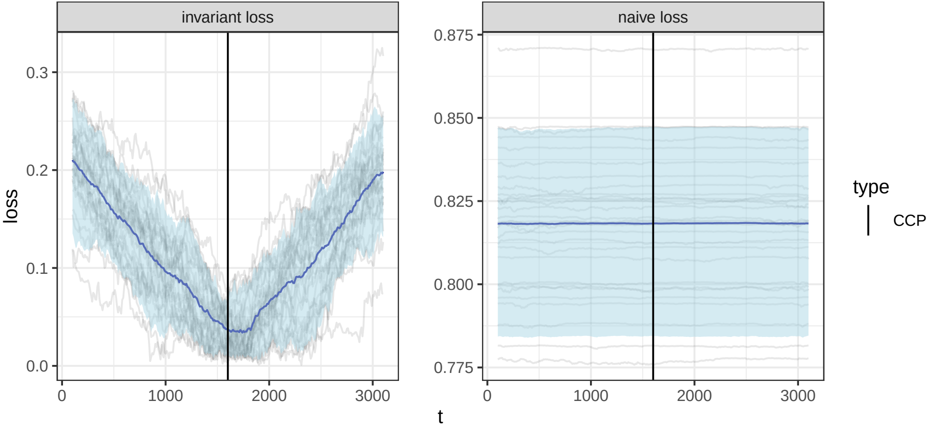
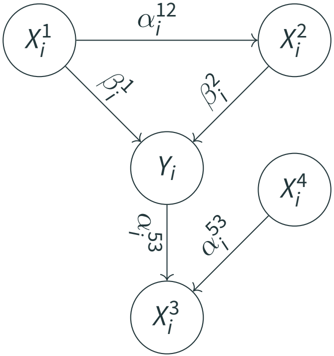


Figure 1: Two variable SCMs where the residual distribution of the response changes ($n = 3200$).

Some simulation results

In the following consider the following Markov Blanket DAG for $i \in \{1, \ldots, n\}$:



The corresponding SCMs are linear Gaussian additive noise.

Some simulation results

One CCP only: one CCP where β_i^1 and β_i^2 changes at $i = \frac{n}{2} + 1$.

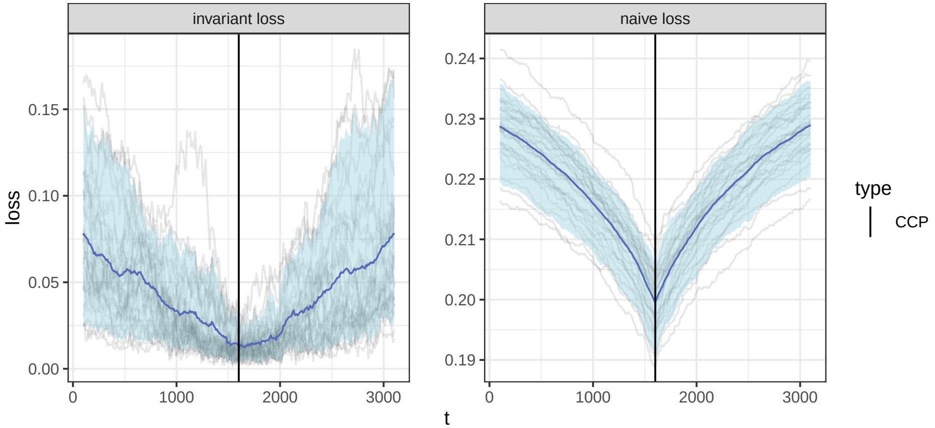


Figure 2: Five variable SCMs with a change in the causal coefficients.

Some simulation results

One CCP and one NCCP: one CCP where β_i^1 and β_i^2 changes at $i = \frac{n}{3} + 1$ and one NCCP where distribution of X_i^3 changes at $i = \frac{2n}{3} + 1$.

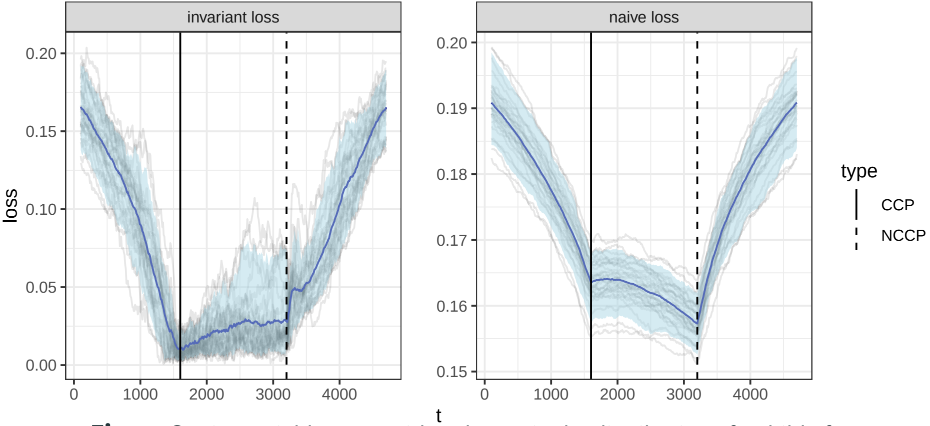


Figure 3: Five variable SCMs with a change in the distribution of a child of Y.

Some simulation results

One NCCP only: one NCCP where distribution of X_i^3 changes at $i = \frac{2n}{3} + 1$.

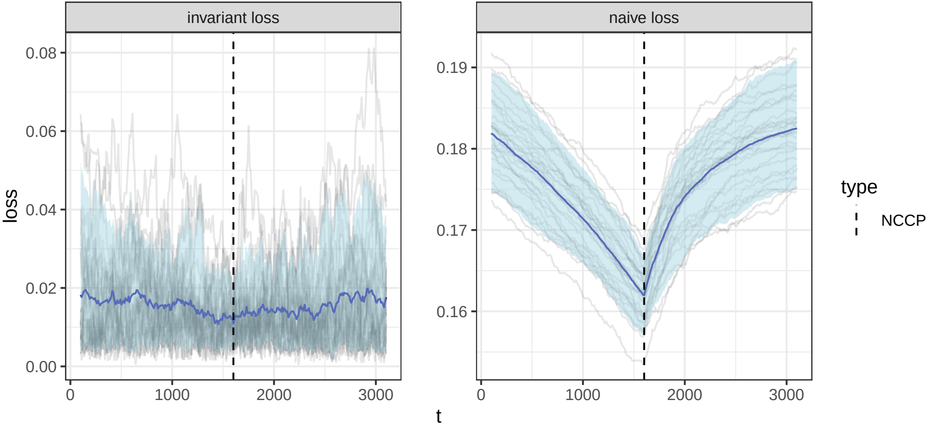


Figure 4: Five variable SCMs with a change in the distribution of a child of Y.

Summary

- Goals:
  1. Test existence of CCP
  2. Estimate locations of CCP
- Approaches:
  a. Testing among known RCPs
  b. Invariant loss

Still investigating

- Properties of the loss functions
- How to localize multiple CCPs

References

J. Bai. Estimating multiple breaks one at a time. *Econometric theory*, 13(3):315–352, 1997.

R. Baranowski, Y. Chen, and P. Fryzlewicz. Narrowest-over-threshold detection of multiple change points and change-point-like features. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 81(3):649–672, 2019.

B. E. Hansen. Testing for structural change in conditional models. *Journal of Econometrics*, 97(1):93–115, 2000.

J. Peters, P. Bühlmann, and N. Meinshausen. Causal inference by using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78 (5):947–1012, 2016.

N. Pfister, P. Bühlmann, and J. Peters. Invariant causal prediction for sequential data. *Journal of the American Statistical Association*, 114(527):1264–1276, 2019.

C. Truong, L. Oudre, and N. Yayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020.

Appendix

Causal change point localization with known RCPs

If the RCPs are known, one can utilize the approach for Goal 1.

Causal change point localization with known RCPs

If the RCPs are known, one can utilize the approach for Goal 1.

Algorithm 1 CCP localization given candidates

Require: data $(\mathbf{X}, \mathbf{Y})$, a set of candidates $\{k_1, \ldots, k_l\}$ with $k_i < k_j$ for $i < j$, and a test φ for $\mathcal{H}_0$.

Let $k_0 := 0$ and $k_{l+1} := n + 1$.
Initiate $\hat{\mathcal{T}} := \emptyset$.
for $i \in \{1, \ldots, l\}$ **do**
    Let $I := \{k_{i-1}, \ldots, k_{i+1} - 1\}$.
    if $\varphi(\mathbf{X}_I, \mathbf{Y}_I) = 1$ **then**
        $\hat{\mathcal{T}} := \hat{\mathcal{T}} \cup \{k_i\}$
    end if
end for
return $\hat{\mathcal{T}}$

Connection with causal models

Prop. (CCP in linear SCMs without unobserved confounding) breaks down if there exists unobserved confounding.

Connection with causal models

Prop. (CCP in linear SCMs without unobserved confounding) breaks down if there exists unobserved confounding.

Example. CCP in linear SCMs without unobserved confounding

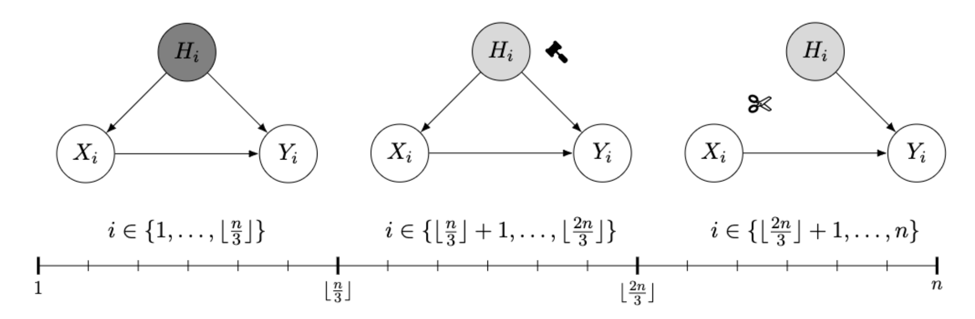
For all $i \in \{1, \dots, n\}$, consider the following linear SCMs where H is unobserved.

$$H_i := \varepsilon_i^H$$
$$X_i := \alpha_i \cdot H_i + \varepsilon_i^X$$
$$Y_i := X_i + H_i + \varepsilon_i^Y,$$

where $\varepsilon_i^X, \varepsilon_i^Y \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ for all $i \in \{1, \dots, n\}$, $\varepsilon_i^H \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ for all $i \in \{1, \dots, \lfloor \frac{n}{3} \rfloor\}$, $\varepsilon_i^H \overset{\text{iid}}{\sim} \mathcal{N}(0,2)$ for all $i \in \{\lfloor \frac{n}{3} \rfloor + 1, \dots, n\}$, $\alpha_i = 1$ for all $i \in \{1, \dots, \lfloor \frac{2n}{3} \rfloor\}$, and $\alpha_i = 0$ for all $i \in \{\lfloor \frac{2n}{3} \rfloor + 1, \dots, n\}$.

Example. CCP in linear SCMs without unobserved confounding (cont.)

The corresponding DAGs are shown below



Example. CCP in linear SCMs without unobserved confounding (cont.)

The OLS coefficient conditioning on the only observed covariate X at each time point is given by

$$\beta_i^{\mathrm{OLS}} = \frac{\mathrm{Cov}(X_i, Y_i)}{\mathbb{V}(X_i)} = \begin{cases} 3/2 & i \in \{1, \ldots, \lfloor \frac{n}{3} \rfloor\} \\ 5/3 & i \in \{\lfloor \frac{n}{3} \rfloor + 1, \ldots, \lfloor \frac{2n}{3} \rfloor\} \\ 1 & i \in \{\lfloor \frac{2n}{3} \rfloor + 1, \ldots, n\}, \end{cases}$$

and the residual distribution conditioning on $\emptyset$ at each time point is given by

$$\varepsilon_i(\emptyset) = \begin{cases} \mathcal{N}(0, 6) & i \in \{1, \ldots, \lfloor \frac{n}{3} \rfloor\} \\ \mathcal{N}(0, 10) & i \in \{\lfloor \frac{n}{3} \rfloor + 1, \ldots, \lfloor \frac{2n}{3} \rfloor\} \\ \mathcal{N}(0, 4) & i \in \{\lfloor \frac{2n}{3} \rfloor + 1, \ldots, n\}. \end{cases}$$