Invariant Policy Learning: A Causal Perspective

Sorawit Saengkyongam

Copenhagen Causality Lab (CoCaLa), University of Copenhagen

Joint work with Nikolaj Thams, Jonas Peters and Niklas Pfister.

UNIVERSITY OF COPENHAGEN

Contextual bandits: Covariates, Action, Reward

Goal: We consider the problem of learning policies that are robust with respect to shifts in the environments.

Setting: (Offline) Contextual Bandits

X: context (observed); **A**: action;

R: reward; **U**: context (unobserved)

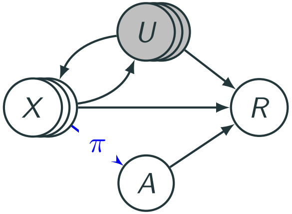


Figure 1: Graphical model of the setting

Contextual bandits: Covariates, Action, Reward

Goal: We consider the problem of learning policies that are robust with respect to shifts in the environments.

Setting: (Offline) Contextual Bandits

X: context (observed); **A**: action;

R: reward; **U**: context (unobserved)

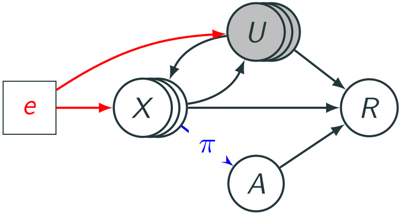


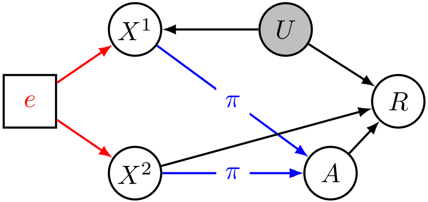
Figure 1: Graphical model of the setting

We assume additionally that data is collected from different environments, $e \in \mathcal{E}$, changing the covariate distributions. Future changes in distribution is represented as new environments.

- E.g. records from different hospitals, countries or experimental setups.

Example: Association flips between environments

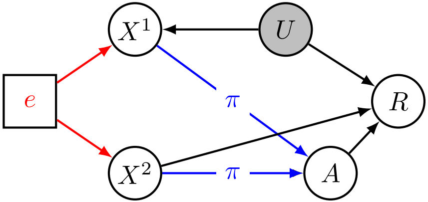
$$\mathcal{S}(\pi, e): \quad \begin{cases} U := \epsilon_U \\ X^1 := \gamma_e U + \epsilon_{X^1} \\ X^2 := \alpha_e + \epsilon_{X^2} \\ A := g_\pi(X^1, X^2, \epsilon_A) \\ R := \begin{cases} \beta_1 X^2 + U + \epsilon_R, & \text{if } A = 0 \\ \beta_2 X^2 - U + \epsilon_R, & \text{if } A = 1 \end{cases} \end{cases}$$



In some environments, X^1 is positively correlated with R under $A = 0$, and in others it is negatively correlated.

Example: Association flips between environments

$$\mathcal{S}(\pi, e): \quad \begin{cases} U := \epsilon_U \\ X^1 := \gamma_e U + \epsilon_{X^1} \\ X^2 := \alpha_e + \epsilon_{X^2} \\ A := g_\pi(X^1, X^2, \epsilon_A) \\ R := \begin{cases} \beta_1 X^2 + U + \epsilon_R, & \text{if } A = 0 \\ \beta_2 X^2 - U + \epsilon_R, & \text{if } A = 1 \end{cases} \end{cases}$$

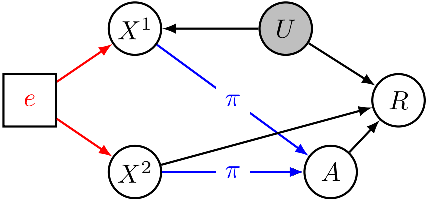


In some environments, X^1 is positively correlated with R under $A = 0$, and in others it is negatively correlated.

$\implies$ We say that X^1 is non-invariant. This poses a threat to generalization.

Example: Association flips between environments

$$\mathcal{S}(\pi, e): \quad \begin{cases} U := \epsilon_U \\ X^1 := \gamma_e U + \epsilon_{X^1} \\ X^2 := \alpha_e + \epsilon_{X^2} \\ A := g_\pi(X^1, X^2, \epsilon_A) \\ R := \begin{cases} \beta_1 X^2 + U + \epsilon_R, & \text{if } A = 0 \\ \beta_2 X^2 - U + \epsilon_R, & \text{if } A = 1 \end{cases} \end{cases}$$



In some environments, X^1 is positively correlated with R under $A = 0$, and in others it is negatively correlated.

$\implies$ We say that X^1 is non-invariant. This poses a threat to generalization.

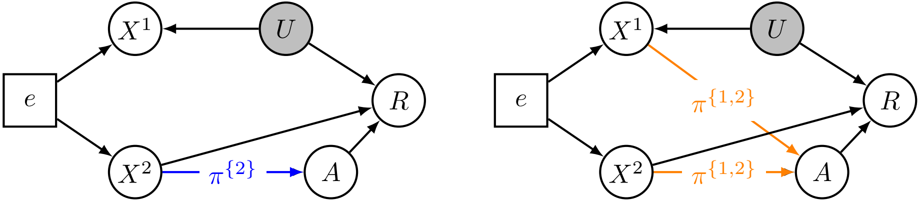
We do not assume that the graph is known! Instead, we seek for non-invariant features and exclude those from policy learning.

Invariance

A set of covariates X^S is invariant if it holds that

$$e \perp\!\!\!\perp_{\mathcal{G}^S} R \mid X^S.$$

A policy π is invariant w.r.t. a set S if π depends only on X^S.



$e \perp\!\!\!\perp_{\mathcal{G}^{\{2\}}} R \mid X^{\{2\}}$ but $e \not\perp\!\!\!\perp_{\mathcal{G}^{\{1,2\}}} R \mid X^{\{1,2\}}$.

Maximizing the worst-case reward

Objective: Distributional Robustness

$$\text{argmax}_{\pi \in \Pi} V^{\mathcal{E}}(\pi), \quad \text{where } V^{\mathcal{E}}(\pi) := \inf_{e \in \mathcal{E}} \mathbb{E}^{\pi,e}\left[R\right].$$

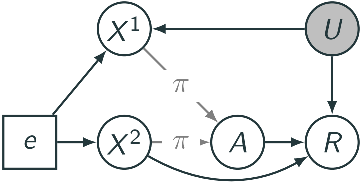
Under certain assumptions solving the distributionally robust objective amounts to finding an optimal invariant policy.

Theorem

Consider an invariant policy $\pi^* \in \text{argmax}_{\pi \in \Pi_{\text{inv}}} \sum_{e \in \mathcal{E}^{\text{obs}}} \mathbb{E}^{\pi,e}[R]$. Under "strong environments" assumption, it holds that

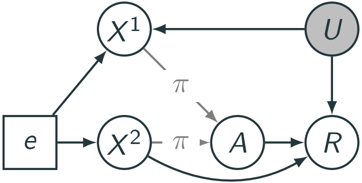
$$\forall \pi \in \Pi : \qquad V^{\mathcal{E}}(\pi) \leq V^{\mathcal{E}}(\pi^*).$$

"Strong environments" Assumption

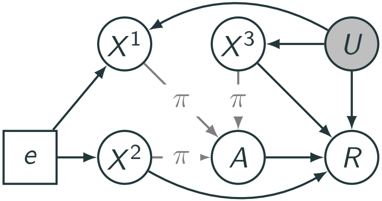


Strong environments: There exists $e \in \mathcal{E}$ such that $X^1 \perp\!\!\!\perp U$ in e.

"Strong environments" Assumption

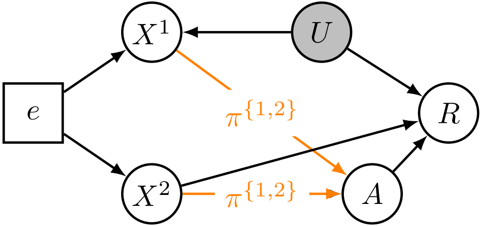


Strong environments: There exists $e \in \mathcal{E}$ such that $X^1 \perp\!\!\!\perp U$ in e.



Strong environments: There exists $e \in \mathcal{E}$ such that $X^1 \perp\!\!\!\perp U$ in e. (Note that X^3 can still be confounded in e).

Testing invariance



We have,

$e \not\perp\!\!\!\perp_{\pi^{\{1,2\}}} R \mid X^{\{1,2\}}$

and $e \not\perp\!\!\!\perp_{\pi^{\{1,2\}}} R \mid X^{\{2\}}$

(but $e \perp\!\!\!\perp_{\pi^{\{2\}}} R \mid X^{\{2\}}$).

Given $S \subseteq \{1, \ldots, d\}$, we resample the data to mimick[1] the policy π^S .

To test invariance: 1) bundle all environments, 2) fit regression, 3) test whether prediction residuals are equally distributed across environments[2].

[1]Nikolaj Thams et al. (2021). "Statistical Testing under Distributional Shifts". In: *arXiv preprint arXiv:2105.10821*

[2]Christina Heinze-Deml et al. (2018). "Invariant Causal Prediction for Nonlinear Models". In: *Journal of Causal Inference* 6.2

Limitations of Subset Search

(i) Computational efficiency
   - Variable screening
   - Greedy search

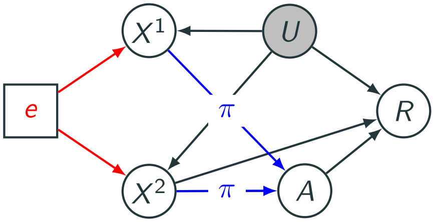
Limitations of Subset Search

(i) Computational efficiency
   - Variable screening
   - Greedy search

(ii) No invariant sets when U acts on the parents

Limitations of Subset Search

$$\mathcal{S}(\pi, e): \quad \begin{cases} U := \epsilon_U \\ X^1 := \gamma_e U + \epsilon_{X^1} \\ X^2 := \alpha_e U + \epsilon_{X^2} \\ A := g_\pi(X^1, X^2, \epsilon_A) \\ R := \begin{cases} \beta_1 X^2 + U + \epsilon_R, & \text{if } A = 0 \\ \beta_2 X^2 - U + \epsilon_R, & \text{if } A = 1 \end{cases} \end{cases}$$



There is no invariant set!!

HSIC-X: Exploiting Independent Instruments Identification and Distribution Generalization

Sorawit Saengkyongam

Copenhagen Causality Lab (CoCaLa), University of Copenhagen

Joint work with Leonard Henckel, Niklas Pfister and Jonas Peters.

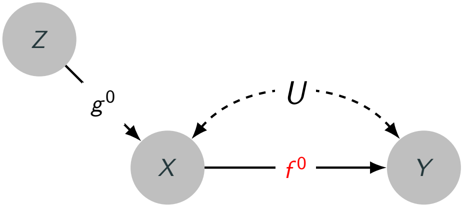
UNIVERSITY OF COPENHAGEN

CoCaLab
Copenhagen Causality Lab

Instrumental Variable (IV) Setting

We consider the following structural causal model M^0

$$
\begin{aligned}
Z &:= \epsilon_Z \\
U &:= \epsilon_U \\
X &:= g^0(Z, U, \epsilon_X) \\
Y &:= f^0(X) + h^0(U, \epsilon_Y)
\end{aligned}
$$



where $Z \in \mathbb{R}^r$ are instruments, $U \in \mathbb{R}^q$ are unobserved variables, $X \in \mathbb{R}^d$ are predictors, $Y \in \mathbb{R}$ is a response, and $(\epsilon_Z, \epsilon_U, \epsilon_X, \epsilon_Y)$ are jointly independent noise variables. The causal function f^0 satisfies independence restriction $Y - f^0(X) \perp\!\!\!\perp Z$.

Identification of f^0: Moment restriction vs Independence restriction

E.g., consider a linear causal function $f^0(x) = x^\top \theta^0$ for some $\theta^0 \in \mathbb{R}^d$.

Classical IV approach

Identification of f^0 is based on the (conditional) moment restriction:

$$\mathbb{E}[Y - X^\top \theta \mid Z] = 0. \tag{1}$$

f^0 is not identifiable when $\mathbb{E}[X \mid Z] = 0$.

Independence-based IV

Identification of f^0 is based on the independence restriction:

$$Y - X^\top \theta \perp\!\!\!\perp Z. \tag{2}$$

We can identify f^0 even if $\mathbb{E}[X \mid Z] = 0$.

Identification of f^0: Moment restriction vs Independence restriction

E.g., consider a linear causal function $f^0(x) = x^\top \theta^0$ for some $\theta^0 \in \mathbb{R}^d$.

Classical IV approach

Identification of f^0 is based on the (conditional) moment restriction:

$$\mathbb{E}[Y - X^\top \theta \mid Z] = 0. \tag{1}$$

f^0 is not identifiable when $\mathbb{E}[X \mid Z] = 0$.

Independence-based IV

Identification of f^0 is based on the independence restriction:

$$Y - X^\top \theta \perp\!\!\!\perp Z. \tag{2}$$

We can identify f^0 even if $\mathbb{E}[X \mid Z] = 0$.

The independence restriction (2) yields

(i) Strictly stronger identifiability results.

(ii) (in some settings) More efficient estimators (e.g., under weak instruments).

Example: Non-additive Instruments

Consider the following SCM

$$\begin{aligned} Z &:= \epsilon_Z \\ U &:= \epsilon_U \\ X &:= ZU + \epsilon_X \\ Y &:= X + U + \epsilon_Y, \end{aligned} \tag{3}$$

with $\mathcal{F} = \{f \mid f(x) = \theta x\}$, where $(\epsilon_Z, \epsilon_U, \epsilon_X, \epsilon_Y)$ are jointly independent standard Gaussian variables.

Example: Non-additive Instruments

Consider the following SCM

$$
\begin{aligned}
Z &:= \epsilon_Z \\
U &:= \epsilon_U \\
X &:= ZU + \epsilon_X \\
Y &:= X + U + \epsilon_Y,
\end{aligned} \tag{3}
$$

with $\mathcal{F} = \{f \mid f(x) = \theta x\}$, where $(\epsilon_Z, \epsilon_U, \epsilon_X, \epsilon_Y)$ are jointly independent standard Gaussian variables.

Here, we have $\mathbb{E}[X|Z] = Z\,\mathbb{E}[U] + \mathbb{E}[\epsilon_X] = 0$ and therefore one cannot identify the causal function based only on the moment restriction. Nonetheless, the causal function can be identified with the independence restriction.

Independence-based IV with HSIC

Given $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$, our method aims to find a function $\hat{f}$ that minimizes the dependency between the residuals $\boldsymbol{Y} - \hat{f}(\boldsymbol{X})$ and the instruments $\boldsymbol{Z}$.

We propose the HSIC-X ('X' for 'exogenous') estimator:

$$\hat{f} := \arg\min_{f \in \mathcal{F}} \widehat{\mathrm{HSIC}}(\boldsymbol{Y} - f(\boldsymbol{X}), \boldsymbol{Z}), \tag{4}$$

Independence-based IV with HSIC

Given $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$, our method aims to find a function $\hat{f}$ that minimizes the dependency between the residuals $\boldsymbol{Y} - \hat{f}(\boldsymbol{X})$ and the instruments $\boldsymbol{Z}$.

We propose the HSIC-X ('X' for 'exogenous') estimator:

$$\hat{f} := \arg\min_{f \in \mathcal{F}} \widehat{\mathrm{HSIC}}(\boldsymbol{Y} - f(\boldsymbol{X}), \boldsymbol{Z}), \tag{4}$$

Two heuristics to alleviate the non-convexity issue:

(i) Initialize the parameters in the first trial at the OLS/2SLS solutions.

(ii) Restarting heuristic: Test for the independence restriction at the solution. If the test is rejected, randomly re-initialize the parameters and restart the optimization.

Under-identified IV and Distribution Generalization

In the under-identified case when Z is not rich enough to identify f^0, we can still get a meaningful estimator where we find the most predictive invariant function.

Under-identified IV and Distribution Generalization

In the under-identified case when Z is not rich enough to identify f^0, we can still get a meaningful estimator where we find the most predictive invariant function.

Theorem [Generalization to interventions on Z]

Let $\ell : \mathbb{R} \to \mathbb{R}$ be a convex loss function and $\mathcal{I}$ be a set of interventions on Z. If the interventions $\mathcal{I}$ is 'strong enough', then

$$\inf_{f \in \mathcal{F}_{\text{inv}}} \mathbb{E}_{M^0}\big[\ell(Y - f(X))\big] = \inf_{f \in \mathcal{F}} \sup_{i \in \mathcal{I}} \mathbb{E}_{M^0(i)}\big[\ell(Y - f(X))\big], \tag{5}$$

where $\mathcal{F}_{\text{inv}} := \{f_\diamond \in \mathcal{F} \mid Z \perp\!\!\!\perp Y - f_\diamond(X) \text{ under } \mathbb{P}_{M^0}\}$ is the space of invariant functions.

Under-identified IV and Distribution Generalization

Motivated by (5), we propose the HSIC-X-pen ('pen' for 'penalization') estimator:

$$\hat{f}^{\lambda} = \arg\min_{f \in \mathcal{F}} \widehat{\mathrm{HSIC}}(\boldsymbol{Y} - f(\boldsymbol{X}), \boldsymbol{Z})) + \lambda \sum_{i=1}^{n} \ell(Y_i - f(X_i)), \tag{6}$$

where the tuning parameter $\lambda \in [0, \infty)$ is selected as the largest possible value for which an HSIC-based independence test between the residuals and the instruments is not rejected.

Contributions

Three contributions:

(i) We discuss the use of the independence restriction $Y - f(X) \perp\!\!\!\perp Z$ in IV estimation and its implication on the identifiability of f^0.

(ii) We propose HSIC-X, a gradient-based learning method that exploits the independence restriction to estimate f^0 and prove its consistency.

(iii) We propose to use the independence restriction for distribution generalization and prove theoretical guarantees.

Future Work

(i) How to estimate the prediction intervals?

(ii) How to handle non-additive confounding effect?