Exploiting Independent Instruments: Identification and Distribution Generalization

Sorawit Saengkyongam

Copenhagen Causality Lab (CoCaLa), University of Copenhagen

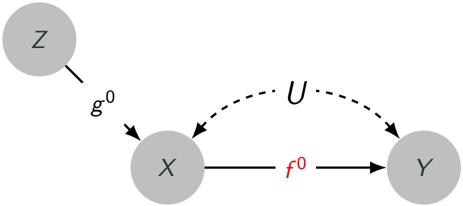
Joint work with Leonard Henckel, Niklas Pfister and Jonas Peters.

UNIVERSITY OF COPENHAGEN

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:

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

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:

$$Y - X^\top \theta \perp\!\!\!\perp Z. \qquad (2)$$

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).

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:

$$\hat{f} := \arg\min_{f \in \mathcal{F}} \widehat{\mathrm{HSIC}}(\boldsymbol{Y} - f(\boldsymbol{X}), \boldsymbol{Z}), \tag{3}$$

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:

$$\hat{f} := \arg\min_{f \in \mathcal{F}} \widehat{\mathrm{HSIC}}(\boldsymbol{Y} - f(\boldsymbol{X}), \boldsymbol{Z}), \tag{3}$$

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

$$\inf_{f \in \mathcal{F}_{\text{inv}}} \mathbb{E}_{M^0}\big[\ell(Y - f(X))\big] = \inf_{f \in \mathcal{F}} \sup_{i \in \mathcal{I}} \mathbb{E}_{M^0(i)}\big[\ell(Y - f(X))\big], \tag{4}$$

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 (4), we propose the HSIC-X-pen ('pen' for 'penalization') estimator:

$$\hat{f}^{\lambda} = \arg\min_{f \in \mathcal{F}} \widehat{\text{HSIC}}(\boldsymbol{Y} - f(\boldsymbol{X}), \boldsymbol{Z})) + \lambda \sum_{i=1}^{n} \ell(Y_i - f(X_i)), \quad (5)$$

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.

Have some questions? See you all at the poster session: Tue 19 Jul 6:30 p.m.