ICML
International Conference
On Machine Learning

Estimating Identifiable Causal Effects on Markov Equivalence Class through Double Machine Learning

Yonghan Jung
Purdue University

Jin Tian
Iowa State University

Elias Bareinboim
Columbia University in the City of New York

The task of Identification (ID)

Given a causal graph (2) &
the observational distribution (3),
is an interventional distribution (1)
computable?

J. Pearl, Causal Diagrams for empirical research. Biometrika, Vol. 82, p. 669-709, 1995.

The task of Identification (ID)

1 Query

$$Q = P_{\mathbf{x}}(\mathbf{y}) \equiv P(\mathbf{y} \,|\, do(\mathbf{x}))$$

Given a causal graph (2) &
the observational distribution (3),
is an interventional distribution (1)
computable?

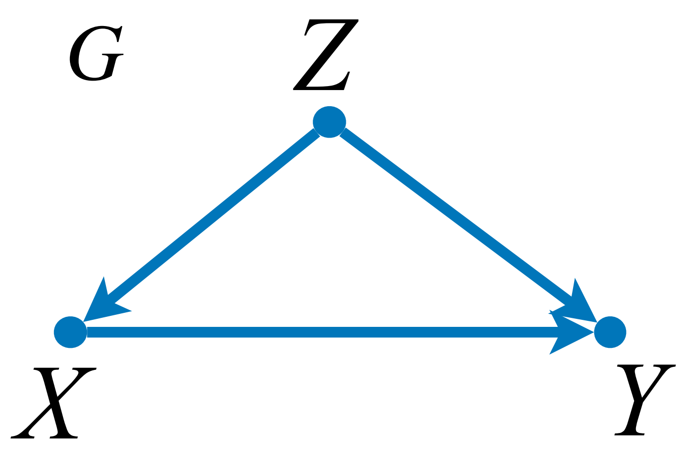
J. Pearl, Causal Diagrams for empirical research. Biometrika, Vol. 82, p. 669-709, 1995.

The task of Identification (ID)

1 Query

$$Q = P_{\mathbf{x}}(\mathbf{y}) \equiv P(\mathbf{y} \mid do(\mathbf{x}))$$

2 Graph



Given a causal graph (2) &
the observational distribution (3),
is an interventional distribution (1)
computable?

J. Pearl, Causal Diagrams for empirical research. Biometrika, Vol. 82, p. 669-709, 1995.

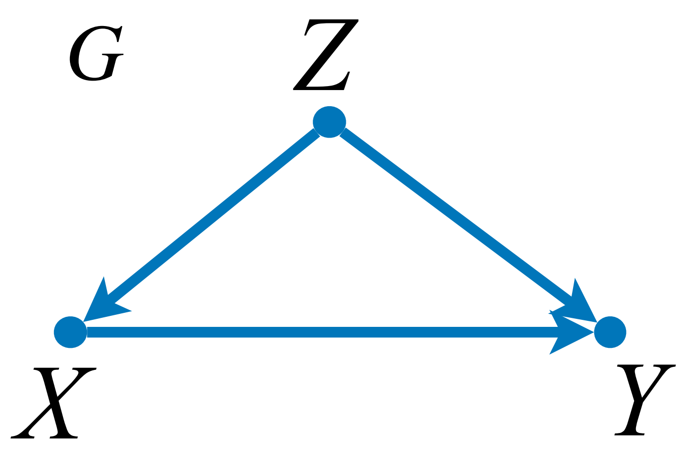
The task of Identification (ID)

Given a causal graph (2) &
the observational distribution (3),
is an interventional distribution (1)
computable?

1 Query

$Q = P_{\mathbf{x}}(\mathbf{y}) \equiv P(\mathbf{y} \mid do(\mathbf{x}))$

2 Graph



3 Distribution

$P(\mathbf{V})$

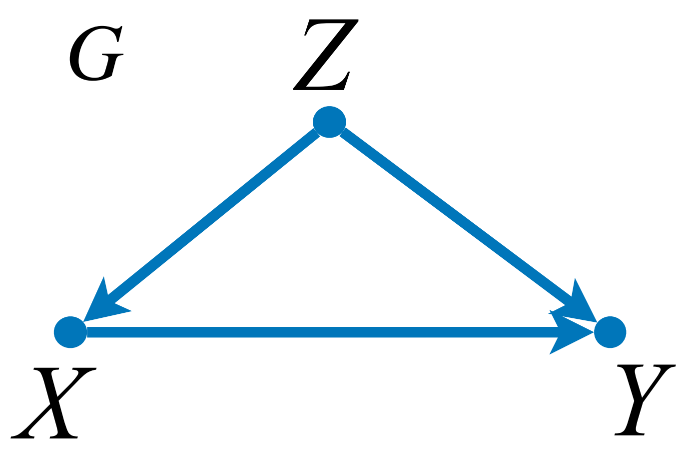
J. Pearl, Causal Diagrams for empirical research. Biometrika, Vol. 82, p. 669-709, 1995.

The task of Identification (ID)

1 Query

$Q = P_{\mathbf{x}}(\mathbf{y}) \equiv P(\mathbf{y} \mid do(\mathbf{x}))$

2 Graph



3 Distribution

$P(\mathbf{V})$

Given a causal graph (2) &
the observational distribution (3),
is an interventional distribution (1)
computable?

ID (G, P, Q)

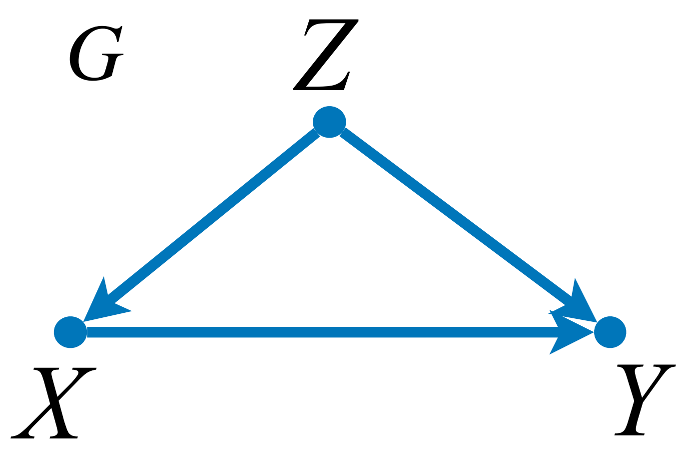
J. Pearl, Causal Diagrams for empirical research. Biometrika, Vol. 82, p. 669-709, 1995.

The task of Identification (ID)

1 Query

$Q = P_{\mathbf{x}}(\mathbf{y}) \equiv P(\mathbf{y} \mid do(\mathbf{x}))$

2 Graph



3 Distribution

$P(\mathbf{V})$

Given a causal graph (2) &
the observational distribution (3),
is an interventional distribution (1)
computable?

ID (G, P, Q) → solution

yes / no

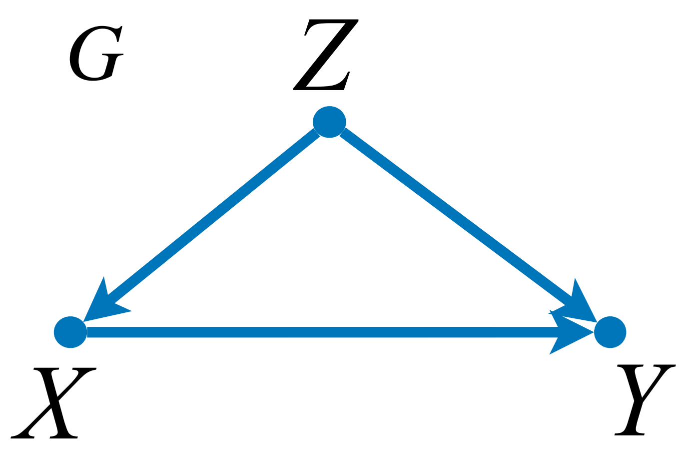
J. Pearl, Causal Diagrams for empirical research. Biometrika, Vol. 82, p. 669-709, 1995.

The task of Identification (ID)

1 Query

$Q = P_{\mathbf{x}}(\mathbf{y}) \equiv P(\mathbf{y} \mid do(\mathbf{x}))$

2 Graph



3 Distribution

$P(\mathbf{V})$

Given a causal graph (2) &
the observational distribution (3),
is an interventional distribution (1)
computable?

ID (G, P, Q) → solution

yes / no

Causal Functional

$P_{\mathbf{x}}(\mathbf{y}) = f(P)$

J. Pearl, Causal Diagrams for empirical research. Biometrika, Vol. 82, p. 669-709, 1995.

The task of Identification (ID)

1 Query

$Q = P_{\mathbf{x}}(\mathbf{y}) \equiv P(\mathbf{y} \mid do(\mathbf{x}))$

Given a causal graph (2) &
the observational distribution (3),
is an interventional distribution (1)
computable?

What if the graph is unknown?

3 Distribution

$P(\mathbf{V})$

ID (G, P, Q) → solution

yes / no

Causal Functional
$P_{\mathbf{x}}(\mathbf{y}) = f(P)$

J. Pearl, Causal Diagrams for empirical research. Biometrika, Vol. 82, p. 669-709, 1995.

Data-Driven Causal Identification

With the causal graph (2) discovered from data (0) and the available distribution (3), we can answer the research question (1).

1 Jaber, Amin, Jiji Zhang, and Elias Bareinboim. "Causal identification under Markov equivalence: Completeness results." *ICML*. PMLR, 2019.

Data-Driven Causal Identification

With the causal graph (2) discovered from data (0) and the available distribution (3), we can answer the research question (1).

0 Data

$\mathscr{D}$

1 Jaber, Amin, Jiji Zhang, and Elias Bareinboim. "Causal identification under Markov equivalence: Completeness results." *ICML*. PMLR, 2019.

Data-Driven Causal Identification

1 Query

$$Q = P_{\mathbf{x}}(\mathbf{y}) \equiv P(\mathbf{y} \mid do(\mathbf{x}))$$

With the causal graph (2) discovered from data (0) and the available distribution (3), we can answer the research question (1).

0 Data

$\mathscr{D}$

1 Jaber, Amin, Jiji Zhang, and Elias Bareinboim. "Causal identification under Markov equivalence: Completeness results." *ICML*. PMLR, 2019.

Data-Driven Causal Identification

1 Query

$Q = P_{\mathbf{x}}(\mathbf{y}) \equiv P(\mathbf{y} \mid do(\mathbf{x}))$

With the causal graph (2) discovered from data (0) and the available distribution (3), we can answer the research question (1).

0 Data $\mathscr{D}$

Causal discovery algorithm (e.g., FCI algorithm)

2 Markov Equiv. Class (MEC) $\mathscr{P}$

1 Jaber, Amin, Jiji Zhang, and Elias Bareinboim. "Causal identification under Markov equivalence: Completeness results." *ICML*. PMLR, 2019.

Data-Driven Causal Identification

1 Query

$Q = P_{\mathbf{x}}(\mathbf{y}) \equiv P(\mathbf{y} \mid do(\mathbf{x}))$

With the causal graph (2) discovered from data (0) and the available distribution (3), we can answer the research question (1).

0 Data $\mathscr{D}$

→ Causal discovery algorithm (e.g., FCI algorithm)

2 Markov Equiv. Class (MEC) $\mathscr{P}$

3 Distribution

$P(\mathbf{V})$

1 Jaber, Amin, Jiji Zhang, and Elias Bareinboim. "Causal identification under Markov equivalence: Completeness results." *ICML*. PMLR, 2019.

Data-Driven Causal Identification

1 Query

$Q = P_{\mathbf{x}}(\mathbf{y}) \equiv P(\mathbf{y} \mid do(\mathbf{x}))$

With the causal graph (2) discovered from data (0) and the available distribution (3), we can answer the research question (1).

0 Data $\mathscr{D}$

Causal discovery algorithm (e.g., FCI algorithm)

2 Markov Equiv. Class (MEC) $\mathscr{P}$

Relaxed identification (IDP)[1]

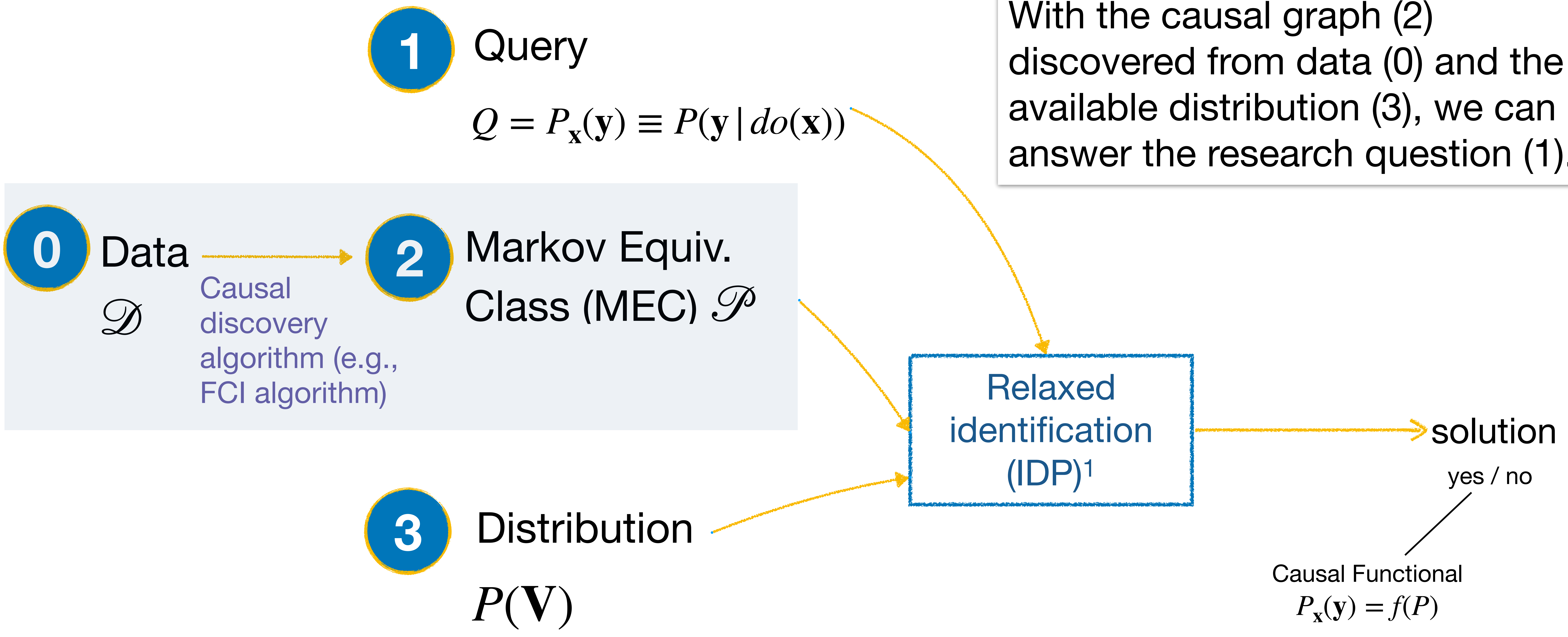
3 Distribution

$P(\mathbf{V})$

1 Jaber, Amin, Jiji Zhang, and Elias Bareinboim. "Causal identification under Markov equivalence: Completeness results." *ICML*. PMLR, 2019.

Data-Driven Causal Identification

With the causal graph (2) discovered from data (0) and the available distribution (3), we can answer the research question (1).

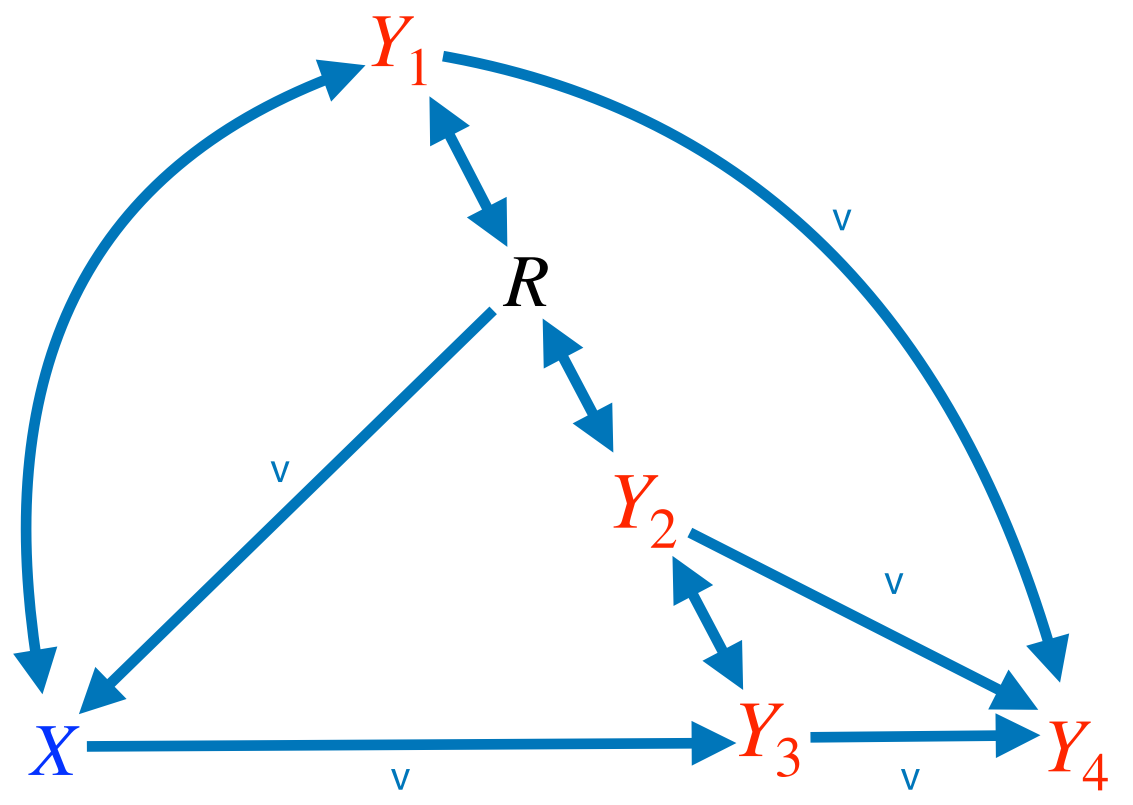


1 Jaber, Amin, Jiji Zhang, and Elias Bareinboim. "Causal identification under Markov equivalence: Completeness results." *ICML*. PMLR, 2019.

Research question

How do we estimate an identifiable functional from the Markov Equiv. Class?

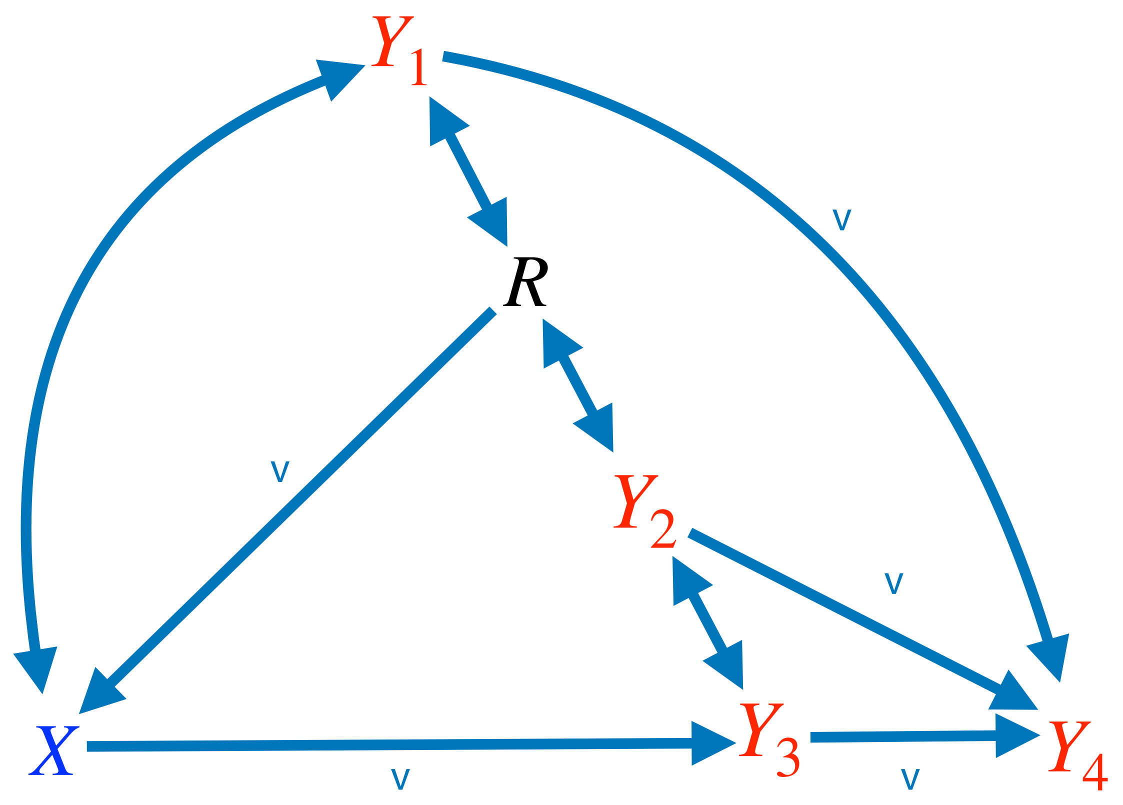
Plug-in (PI) estimator: Example



$$P_x(y_1, y_2, y_3, y_4)$$

$$= P(y_4 \mid y_3, y_2, y_1, x, r) P(y_1) \sum_r P(y_2, y_3 \mid x, r) P(r)$$

Plug-in (PI) estimator: Example



$$P_x(y_1, y_2, y_3, y_4)$$

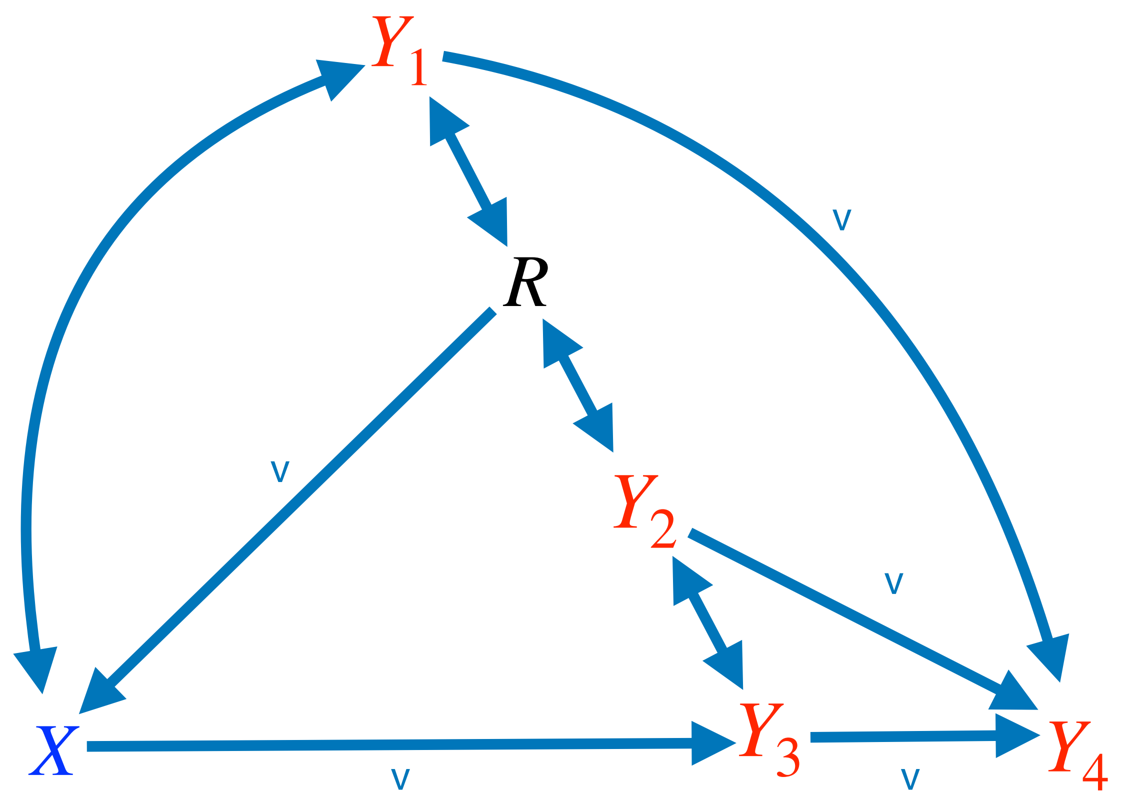
$$= P(y_4 \mid y_3, y_2, y_1, x, r)P(y_1) \sum_r P(y_2, y_3 \mid x, r)P(r)$$

"Nuisances"

Let $\mathscr{P} \equiv \left\{ P(y_4 \mid y_3, y_2, y_1, x, r), P(y_1), P(y_2, y_3 \mid x, r), P(r) \right\}$

$$\hat{\mathscr{P}} \equiv \left\{ \hat{P}(y_4 \mid y_3, y_2, y_1, x, r), \hat{P}(y_1), \hat{P}(y_2, y_3 \mid x, r), \hat{P}(r) \right\}$$

Plug-in (PI) estimator: Example



$$P_x(y_1, y_2, y_3, y_4)$$
$$= P(y_4 \mid y_3, y_2, y_1, x, r) P(y_1) \sum_r P(y_2, y_3 \mid x, r) P(r)$$

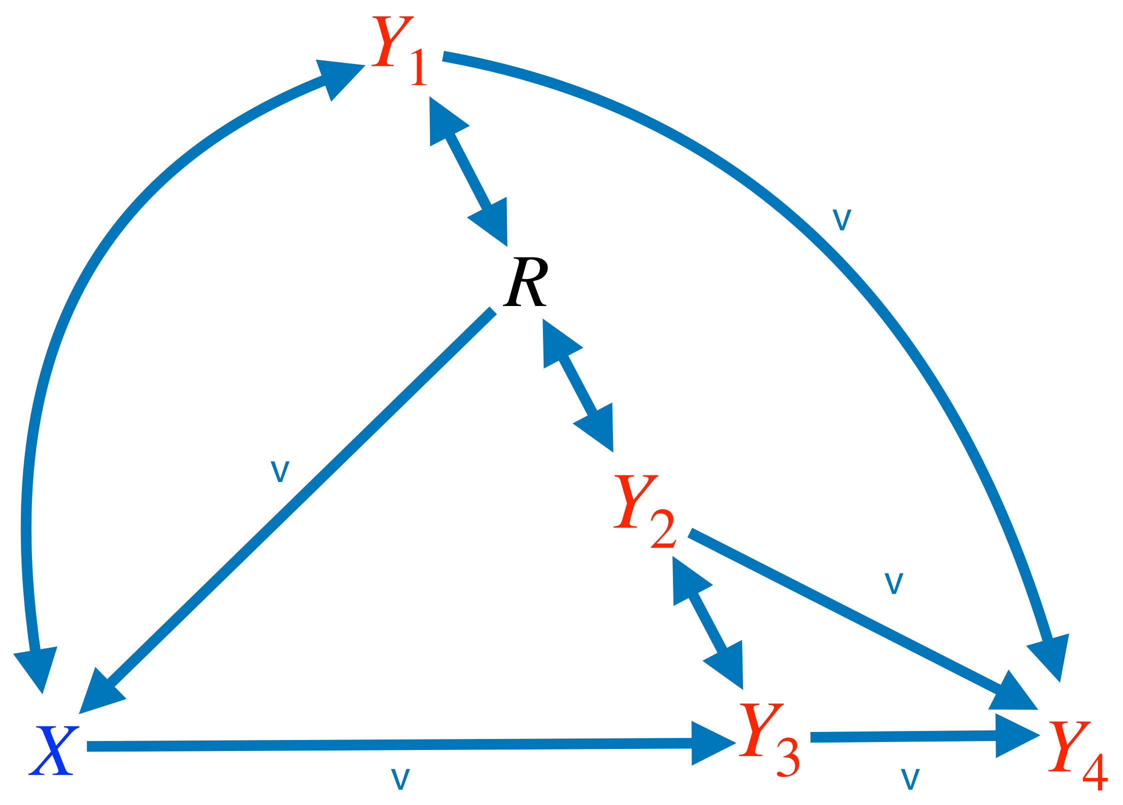
"Nuisances"

Let $\mathscr{P} \equiv \{P(y_4 \mid y_3, y_2, y_1, x, r), P(y_1), P(y_2, y_3 \mid x, r), P(r)\}$

$$\hat{\mathscr{P}} \equiv \left\{\hat{P}(y_4 \mid y_3, y_2, y_1, x, r), \hat{P}(y_1), \hat{P}(y_2, y_3 \mid x, r), \hat{P}(r)\right\}$$

Plug-in (PI) estimator	$\hat{P}(y_4 \mid y_3, y_2, y_1, x, r) \hat{P}(y_1) \sum_r \hat{P}(y_2, y_3 \mid x, r) \hat{P}(r)$

Plug-in (PI) estimator: Example



$$P_x(y_1, y_2, y_3, y_4)$$
$$= P(y_4 \mid y_3, y_2, y_1, x, r)P(y_1) \sum_r P(y_2, y_3 \mid x, r)P(r)$$

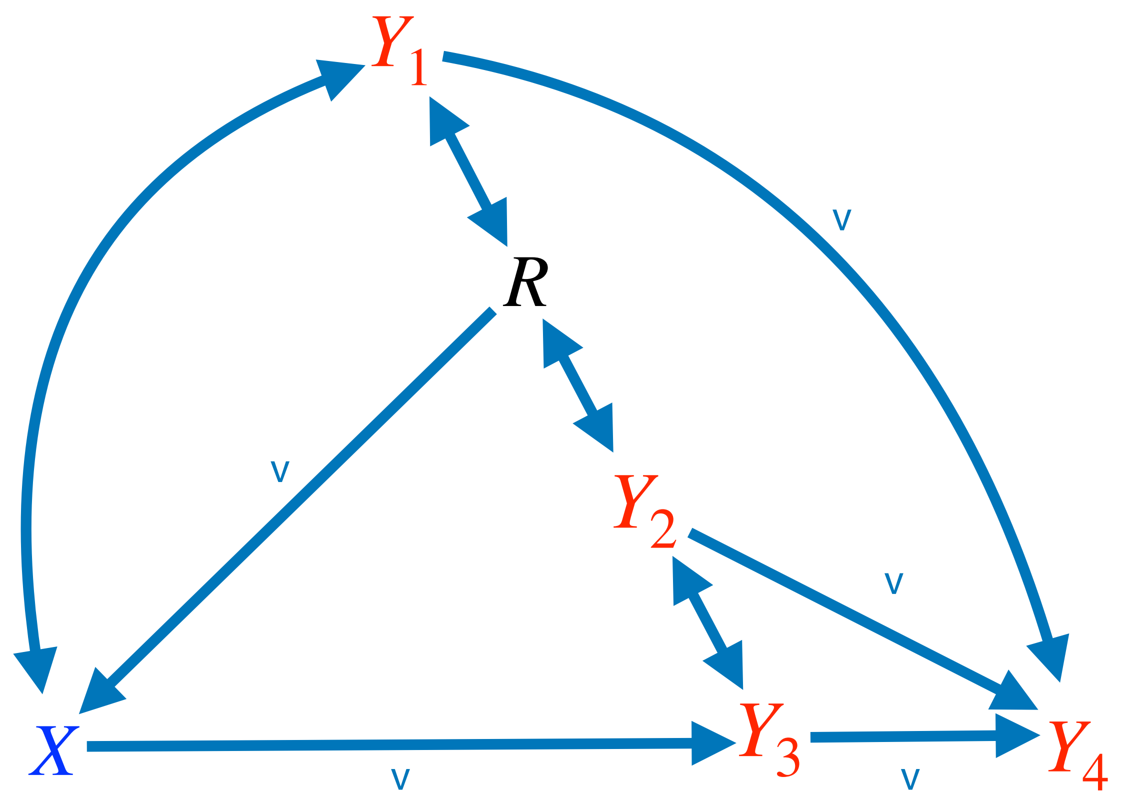
"Nuisances"

Let $\mathscr{P} \equiv \{P(y_4 \mid y_3, y_2, y_1, x, r), P(y_1), P(y_2, y_3 \mid x, r), P(r)\}$

$$\hat{\mathscr{P}} \equiv \left\{\hat{P}(y_4 \mid y_3, y_2, y_1, x, r), \hat{P}(y_1), \hat{P}(y_2, y_3 \mid x, r), \hat{P}(r)\right\}$$

Plug-in (PI) estimator	$\hat{P}(y_4 \mid y_3, y_2, y_1, x, r)\hat{P}(y_1) \sum_r \hat{P}(y_2, y_3 \mid x, r)\hat{P}(r)$
For consistent estimation	$\hat{\mathscr{P}}$ converges to the true $\mathscr{P}$.

Plug-in (PI) estimator: Example



$$P_x(y_1, y_2, y_3, y_4)$$
$$= P(y_4 \mid y_3, y_2, y_1, x, r)P(y_1) \sum_r P(y_2, y_3 \mid x, r)P(r)$$

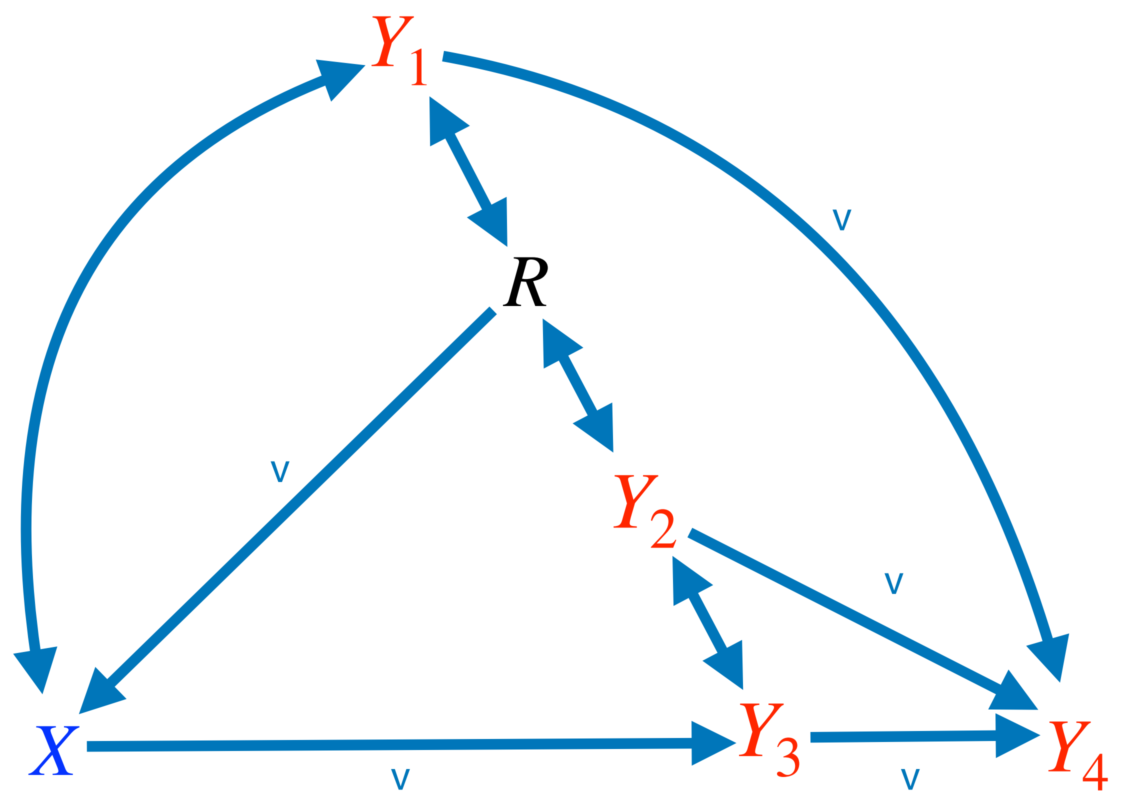
"Nuisances"

Let $\mathscr{P} \equiv \left\{P(y_4 \mid y_3, y_2, y_1, x, r), P(y_1), P(y_2, y_3 \mid x, r), P(r)\right\}$

$$\hat{\mathscr{P}} \equiv \left\{\hat{P}(y_4 \mid y_3, y_2, y_1, x, r), \hat{P}(y_1), \hat{P}(y_2, y_3 \mid x, r), \hat{P}(r)\right\}$$

Plug-in (PI) estimator	$\hat{P}(y_4 \mid y_3, y_2, y_1, x, r)\hat{P}(y_1) \sum_r \hat{P}(y_2, y_3 \mid x, r)\hat{P}(r)$
For consistent estimation	$\hat{\mathscr{P}}$ converges to the true $\mathscr{P}$.
For fast ($N^{-1/2}$ rate) convergence	$\hat{\mathscr{P}}$ converges fast ($N^{-1/2}$-rate).

Plug-in (PI) estimator: Example



$$P_x(y_1, y_2, y_3, y_4)$$

$$= P(y_4 \mid y_3, y_2, y_1, x, r) P(y_1) \sum_r P(y_2, y_3 \mid x, r) P(r)$$

"Nuisances"

Let $\mathscr{P} \equiv \left\{ P(y_4 \mid y_3, y_2, y_1, x, r), P(y_1), P(y_2, y_3 \mid x, r), P(r) \right\}$

$$\hat{\mathscr{P}} \equiv \left\{ \hat{P}(y_4 \mid y_3, y_2, y_1, x, r), \hat{P}(y_1), \hat{P}(y_2, y_3 \mid x, r), \hat{P}(r) \right\}$$

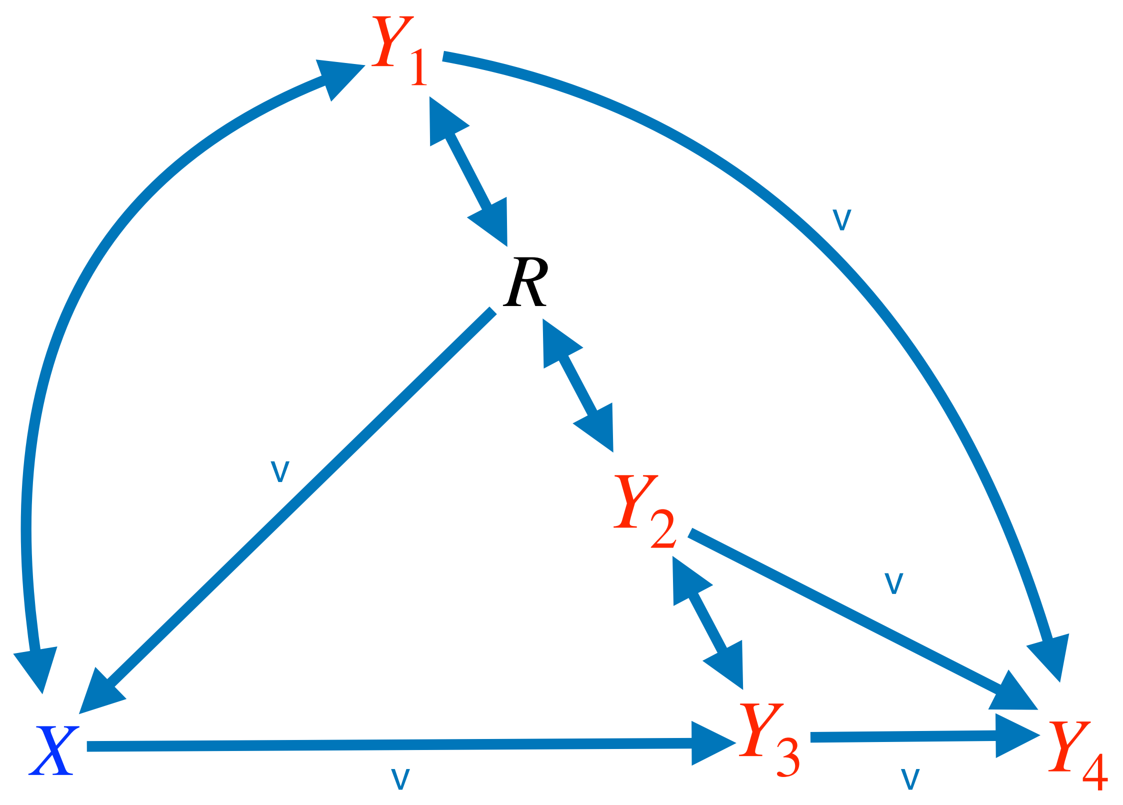
Plug-in (PI) estimator

For consistent estimation

For fast ($N^{-1/2}$ rate) convergence

These requirements are strong!

Plug-in (PI) estimator: Example



$$P_x(y_1, y_2, y_3, y_4)$$
$$= P(y_4 \mid y_3, y_2, y_1, x, r)P(y_1) \sum_r P(y_2, y_3 \mid x, r)P(r)$$

"Nuisances"

Let $\mathscr{P} \equiv \{P(y_4 \mid y_3, y_2, y_1, x, r), P(y_1), P(y_2, y_3 \mid x, r), P(r)\}$

$$\hat{\mathscr{P}} \equiv \left\{\hat{P}(y_4 \mid y_3, y_2, y_1, x, r), \hat{P}(y_1), \hat{P}(y_2, y_3 \mid x, r), \hat{P}(r)\right\}$$

Plug-in (PI) estimator

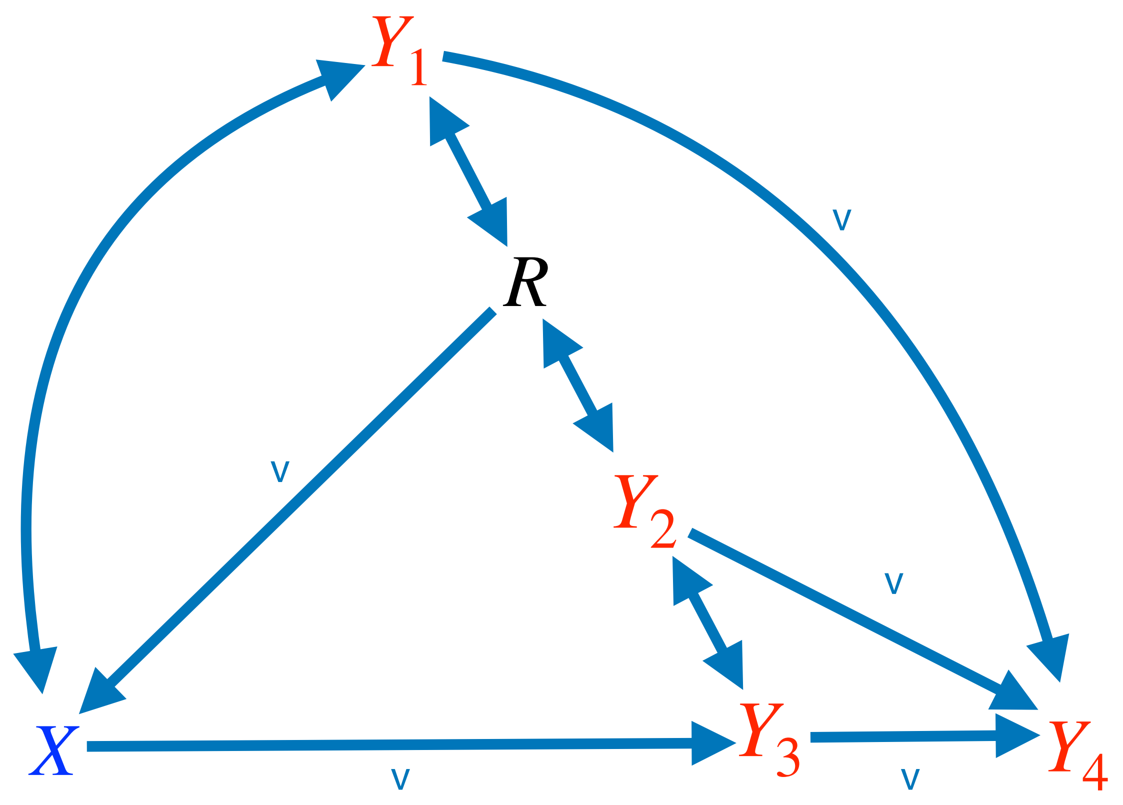
For consistent estimation

For fast ($N^{-1/2}$ rate) convergence

These requirements are strong!

- **Model misspecification** — If any of nuisances is misspecified, then PI is not consistent.

Plug-in (PI) estimator: Example



$$P_x(y_1, y_2, y_3, y_4)$$
$$= P(y_4 \mid y_3, y_2, y_1, x, r)P(y_1) \sum_r P(y_2, y_3 \mid x, r)P(r)$$

"Nuisances"

Let $\mathscr{P} \equiv \{P(y_4 \mid y_3, y_2, y_1, x, r), P(y_1), P(y_2, y_3 \mid x, r), P(r)\}$

$$\hat{\mathscr{P}} \equiv \left\{\hat{P}(y_4 \mid y_3, y_2, y_1, x, r), \hat{P}(y_1), \hat{P}(y_2, y_3 \mid x, r), \hat{P}(r)\right\}$$

Plug-in (PI) estimator

For consistent estimation

For fast ($N^{-1/2}$ rate) convergence

These requirements are strong!

- **Model misspecification** — If any of nuisances is misspecified, then PI is not consistent.
- **Bias** — If any of nuisances converges slowly (e.g., $N^{-1/4}$ -rate), then PI fails to converge fast (e.g., in $N^{-1/2}$ rate).

Our approach: Estimation Engine using Double/Debiased Machine Learning (DML)

Chernozhukov et al., 2018

Our approach: Estimation Engine using Double/Debiased Machine Learning (DML)

Data $\mathscr{D}$

↓

Markov Equiv. Class $\mathscr{P}$

↓

IDP($\mathscr{P}, P_{\mathbf{x}}(\mathbf{y}), P$)

↓

Causal Functional

$$P_{\mathbf{x}}(\mathbf{y}) = f(P)$$

Our approach: Estimation Engine using Double/Debiased Machine Learning (DML)

Data $\mathscr{D}$

Markov Equiv. Class $\mathscr{P}$

1 Derivation of Influence function(IF) (Algorithm 1)

IDP($\mathscr{P}, P_{\mathbf{x}}(\mathbf{y}), P$)

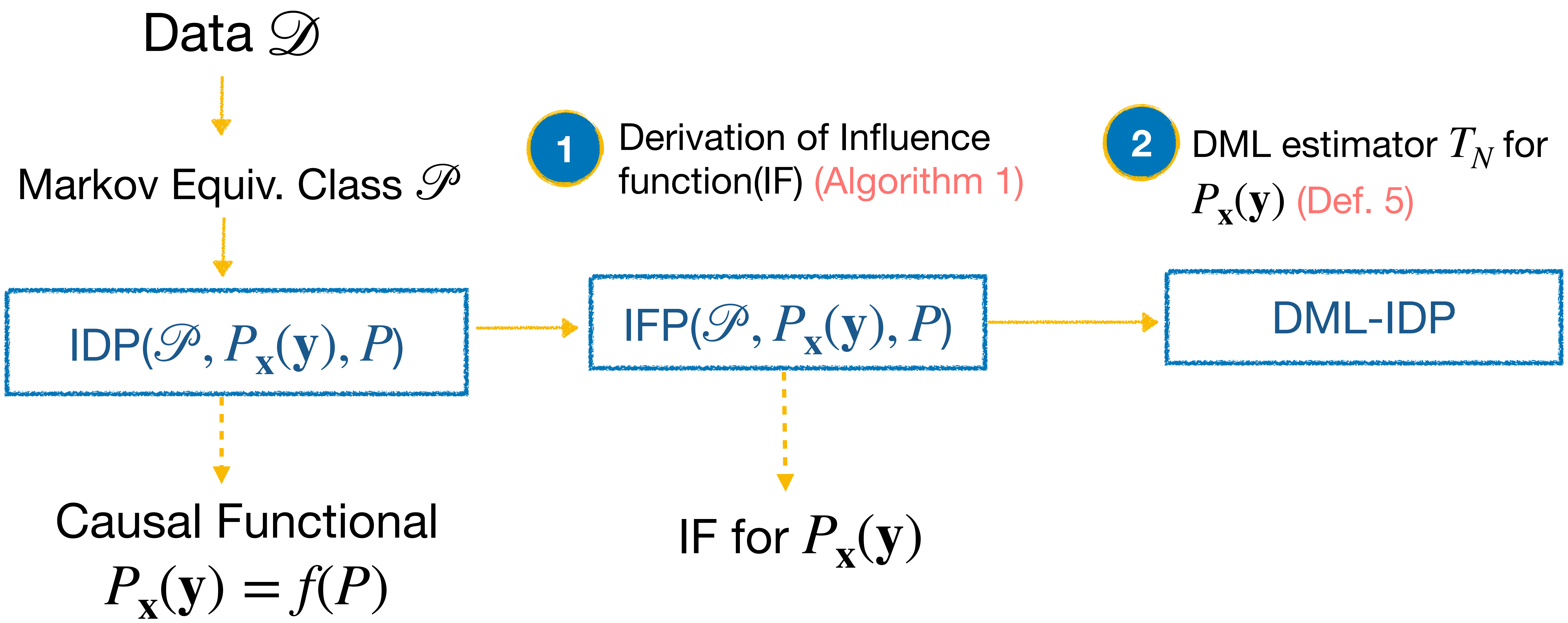
IFP($\mathscr{P}, P_{\mathbf{x}}(\mathbf{y}), P$)

Causal Functional

$P_{\mathbf{x}}(\mathbf{y}) = f(P)$

IF for $P_{\mathbf{x}}(\mathbf{y})$

Our approach: Estimation Engine using Double/Debiased Machine Learning (DML)



Our approach: Estimation Engine using Double/Debiased Machine Learning (DML)

Data $\mathscr{D}$

Markov Equiv. Class $\mathscr{P}$

IDP($\mathscr{P}, P_{\mathbf{x}}(\mathbf{y}), P$)

Causal Functional

$P_{\mathbf{x}}(\mathbf{y}) = f(P)$

1 Derivation of Influence function(IF) (Algorithm 1)

2 DML estimator T_N for $P_{\mathbf{x}}(\mathbf{y})$ (Def. 5)

Robustness of DML estimator (Thm. 2)

Our approach: Estimation Engine using Double/Debiased Machine Learning (DML)

Data $\mathscr{D}$

Markov Equiv. Class $\mathscr{P}$

IDP($\mathscr{P}, P_{\mathbf{x}}(\mathbf{y}), P$)

Causal Functional
$P_{\mathbf{x}}(\mathbf{y}) = f(P)$

1 Derivation of Influence function(IF) (Algorithm 1)

2 DML estimator T_N for $P_{\mathbf{x}}(\mathbf{y})$ (Def. 5)

Robustness of DML estimator (Thm. 2)

(1) **Debiasedness** — T_N converges fast ($\sqrt{N}$ rate) even when nuisances (e.g., $\mathscr{P}, \hat{\mathscr{P}}$) converges slow ($N^{-1/4}$ rate)

Our approach: Estimation Engine using Double/Debiased Machine Learning (DML)

Data $\mathscr{D}$

Markov Equiv. Class $\mathscr{P}$

IDP($\mathscr{P}, P_{\mathbf{x}}(\mathbf{y}), P$)

Causal Functional

$P_{\mathbf{x}}(\mathbf{y}) = f(P)$

1 Derivation of Influence function(IF) (Algorithm 1)

2 DML estimator T_N for $P_{\mathbf{x}}(\mathbf{y})$ (Def. 5)

Robustness of DML estimator (Thm. 2)

(1) **Debiasedness** — T_N converges fast ($\sqrt{N}$ rate) even when nuisances (e.g., $\mathscr{P}$, $\hat{\mathscr{P}}$) converges slow ($N^{-1/4}$ rate)

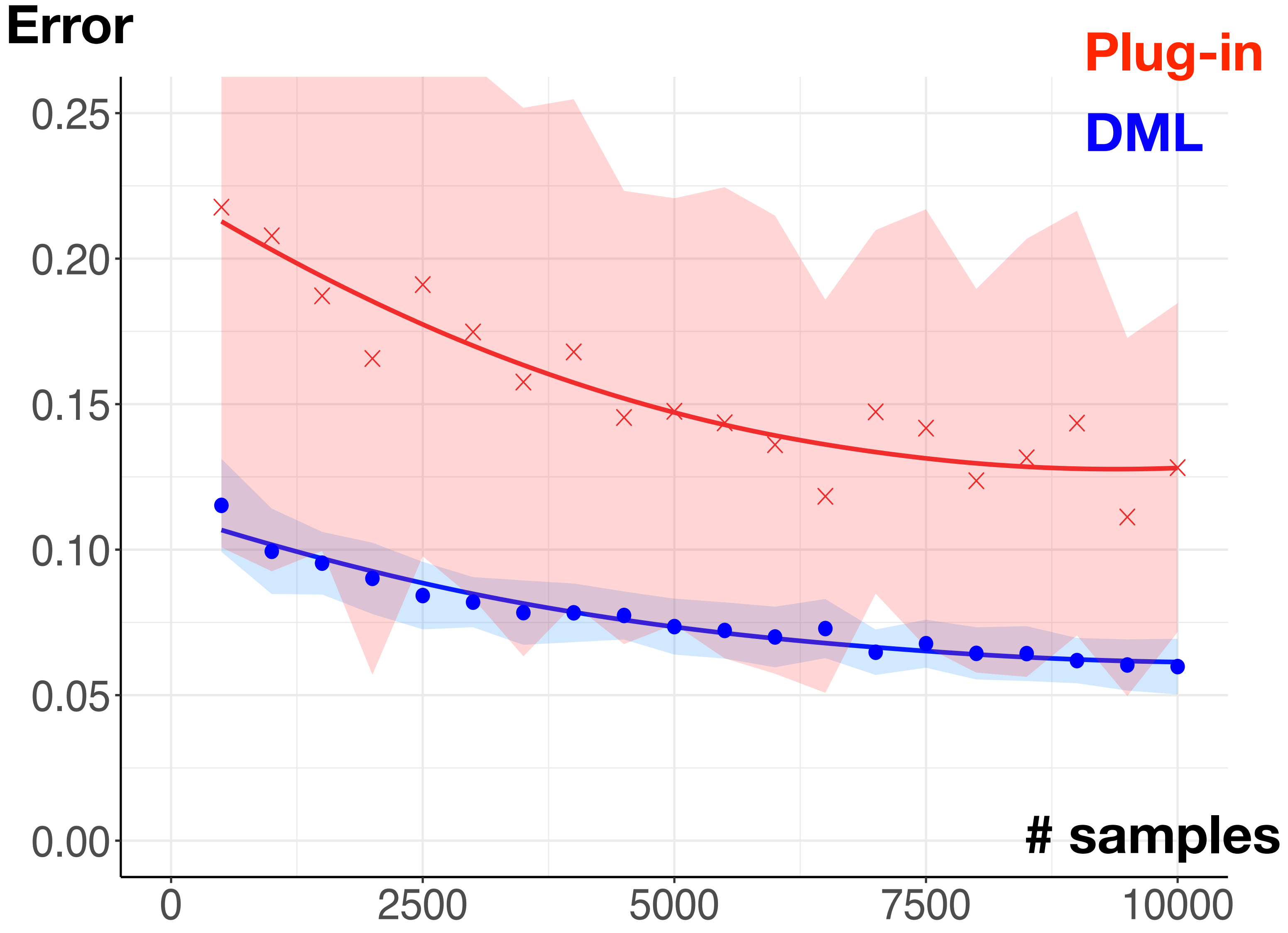
(2) **Doubly robustness** — T_N is consistent even when a portion of nuisance is misspecified.

Simulation results

DB	DR-1	DR-2

Simulation results

Debiasedness (DB)	All nuisances converges slow (at $N^{-1/4}$ rate).

DB	DR-1	DR-2
		

Simulation results

Debiasedness (DB)	All nuisances converges slow (at $N^{-1/4}$ rate).
Doubly Robust: Case 1 (DR-1)	A portion of nuisance ($P(y_4 \mid y_3, y_2, y_1, r, x)$) is misspecified.
Doubly Robust: Case 2 (DR-2)	Another portion of nuisance $\{P(y_2 \mid x, r), P(y_3 \mid y_2, x, r)\}$ are misspecified.

Simulation results

Debiasedness (DB)	All nuisances converges slow (at $N^{-1/4}$ rate).
Doubly Robust: Case 1 (DR-1)	A portion of nuisance ($P(y_4 \mid y_3, y_2, y_1, r, x)$) is misspecified.
Doubly Robust: Case 2 (DR-2)	Another portion of nuisance $\{P(y_2 \mid x, r), P(y_3 \mid y_2, x, r)\}$ are misspecified.

DML estimator exhibits Debiasedness and Doubly Robustness properties.

Conclusion

Conclusion

- We derived an influence function (Alg. 1, Thm. 1) for ***any*** causal effects identifiable given a Markov equivalence class (MEC), learnable from data.

Conclusion

- We derived an influence function (Alg. 1, Thm. 1) for ***any*** causal effects identifiable given a Markov equivalence class (MEC), learnable from data.

- We developed DML estimators (Def. 5), which enjoy ***debiasedness*** and ***doubly robustness***, for any identifiable functional.

Conclusion

- We derived an influence function (Alg. 1, Thm. 1) for ***any*** causal effects identifiable given a Markov equivalence class (MEC), learnable from data.

- We developed DML estimators (Def. 5), which enjoy ***debiasedness*** and ***doubly robustness***, for any identifiable functional.

- ***Summary*** — We introduced the **first general algorithm** for estimating *any* identifiable functional which enjoy ***debiasedness*** and ***doubly robustness.***

Data $\mathscr{D}$ → MEC $\mathscr{P}$ → Causal identification $P_{\mathbf{x}}(\mathbf{y}) = f(P)$ → Causal estimation