Do-calculus and identifiability

Charles K. Assaad, Emilie Devijver, Eric Gaussier

emilie.devijver@univ-grenoble-alpes.fr

Table of contents

Reminder: back-door and front-door criterion

Notations

Rules of do-calculus

Examples

Completeness of do-calculus

Reminder: back-door criterion

The back-door criterion: Consider a causal graph $\mathcal{G}$ and a causal effect $P(y|do(x))$. A set of variables $\mathcal{Z}$ satisfies the back-door criterion iff:

- no node in $\mathcal{Z}$ is a descendant of X;
- $\mathcal{Z}$ blocks every path between X and Y that contains an arrow into X.

Theorem: If $\mathcal{Z}$ satisfies the back-door criterion relative to (X, Y) and if $P(x, z) > 0$, then the causal effect of X on Y is identifiable and is given by

$$P(y|do(x)) = \sum_{z} P(y|x, z)P(z).$$

The causal effect in Markovian models is always identifiable using the back-door criterion and is given by the back-door adjustment.

Reminder: front-door criterion

The front-door criterion: Consider a causal graph $\mathcal{G}$ and a causal effect $P(y|do(x))$. A set of variables $\mathcal{Z}$ satisfies the front-door criterion iff:

- $\mathcal{Z}$ intercepts all directed paths from X to Y;
- there is no back-door path from X to $\mathcal{Z}$;
- All back-door paths from $\mathcal{Z}$ to Y are blocked by X.

Theorem: If $\mathcal{Z}$ satisfies the front-door criterion relative to (X, Y) and if $P(x, z) > 0$, then the causal effect of X on Y is identifiable and is given by

$$P(y|do(x)) = \sum_z P(z|x) \sum_{x'} P(y|x', z) P(x').$$

Reminder: back-door and front-door criterion

If there exists a set that satisfy the back-door criterion for $P(y|do(x))$, then $P(y|do(x))$ is identifiable,

If there exists a set that satisfy the front-door criterion for $P(y|do(x))$, then $P(y|do(x))$ is identifiable,

If there exists no set that satisfy the back-door or the front-door criterion for $P(y|do(x))$, then $P(y|do(x))$ is not necessarily not identifiable.

Rules of do-calculus!

Goal: identify any causal quantity that is identifiable!

Reminder: back-door and front-door criterion

If there exists a set that satisfy the back-door criterion for $P(y|do(x))$, then $P(y|do(x))$ is identifiable,

If there exists a set that satisfy the front-door criterion for $P(y|do(x))$, then $P(y|do(x))$ is identifiable,

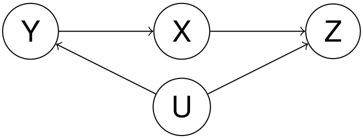
If there exists no set that satisfy the back-door or the front-door criterion for $P(y|do(x))$, then $P(y|do(x))$ is not necessarily not identifiable.

Rules of do-calculus!

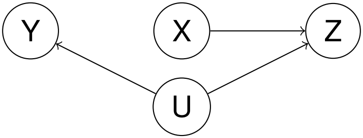
Goal: identify any causal quantity that is identifiable!

Preliminary notations

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and let $X \subseteq \mathcal{V}$.

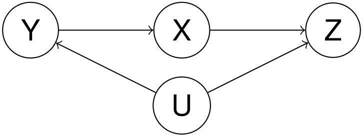


We define $G_{\overline{X}}$ to be the graph obtained by removing from G all edges from $Pa(X)$ to X.

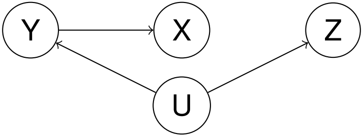


Preliminary notations

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and let $X \subseteq \mathcal{V}$.

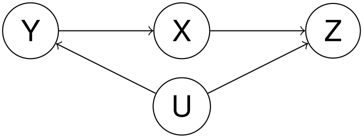


Analogously, we define $G_{\underline{X}}$ to be the graph obtained by removing from G all edge from X to $Ch(X)$.



Preliminary notations

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and let $X \subseteq \mathcal{V}$.



Exercise: Draw $G_{\overline{Z}}$, $G_{\underline{Z}}$, $G_{\overline{XZ}}$, $G_{\overline{X}\underline{Z}}$.

Rule 1: Insertion / deletion of observations

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a causal graph. Let $X, Y, Z, W \subseteq V$ be disjoint. We have:

$$P(y|do(x), z, w) = P(y|do(x), w) \quad \text{if} \quad (X \perp\!\!\!\perp Z|Z, W)_{G_{\underline{X}}}$$

(proof on board)

Remark: removing $do(x)$,we recognize the following fact: d-separation implies conditional independence

Rule 1: Insertion / deletion of observations

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a causal graph. Let $X, Y, Z, W \subseteq V$ be disjoint. We have:

$$P(y|do(x), z, w) = P(y|do(x), w) \quad \text{if} \quad (X \perp\!\!\!\perp Z|Z, W)_{G_{\underline{X}}}$$

(proof on board)

Remark: removing $do(x)$,we recognize the following fact: d-separation implies conditional independence

Rule 2: Action/observation exchange

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a causal graph. Let $X, Y, Z, W \subseteq V$ be disjoint. We have:

$$P(y|do(x), do(z), w) = P(y|do(x), z, w) \quad \text{if} \quad (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\underline{Z}}}$$

Remark: again, removing $do(x)$,we recognize the following fact: back-door criterion in term of d-separation.

Rule 2: Action/observation exchange

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a causal graph. Let $X, Y, Z, W \subseteq V$ be disjoint. We have:

$$P(y|do(x), do(z), w) = P(y|do(x), z, w) \quad \text{if} \quad (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\underline{Z}}}$$

Remark: again, removing $do(x)$,we recognize the following fact: back-door criterion in term of d-separation.

Rule 3: insertion / deletion of actions

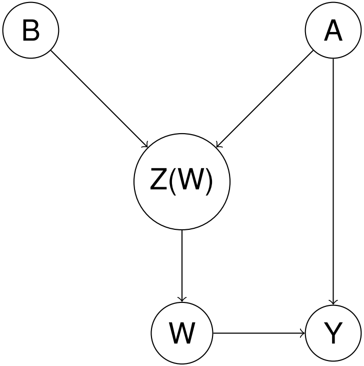
Let $Z(W) = Z \setminus An_{G_{\underline{X}}}(W)$. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a causal graph. Let $X, Y, Z, W \subseteq V$ be disjoint. We have:

$$P(y|do(x), do(z), w) = P(y|do(x), w) \quad \text{if} \quad (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{XZ(W)}}}$$

Rule 3: insertion / deletion of actions

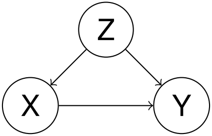
Let $Z(W) = Z \setminus An_{G_{\overline{X}}}(W)$. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a causal graph. Let $Y, Z, W \subseteq V$ be disjoint. We have:

$$P(y|do(z), w) = P(y|w) \quad \text{if} \quad (Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{Z(W)}}}$$



Remove potential colliders → needs to define $Z(W)$!

Example: Back-door criteria



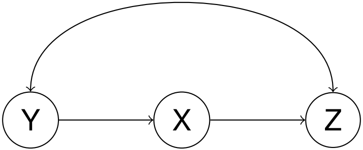
$$P(y|do(x)) = \sum_z P(y|do(x), z)P(z|do(x)) \quad \text{Marginalization over } Z$$

$$P(y|do(x), z) = P(y|x, z) \quad \text{by Rule 2 over } G_{\underline{X}}$$

$$P(z|do(x)) = P(z) \quad \text{by Rule 3 over } G_{\overline{X}}$$

$$P(y|do(x)) = \sum_z P(y|x, z)P(z)$$

Example: Front-door criteria



$$
\begin{aligned}
P(y|do(x)) &= \sum_z P(y|do(x), z)P(z|do(x)) \quad \text{Margin. over } Z \\
P(z|do(x)) &= P(z) \quad \text{by Rule 2} \\
P(y|do(x), z) &= P(y|do(x), do(z)) \quad \text{by Rule 2 in } G_{\overline{X}\underline{Z}} \\
P(y|do(x), do(z)) &= P(y|do(x), do(z)) \quad \text{by Rule 3 in} G_{\overline{Z}} \\
P(y|do(x)) &= \sum_z P(y|do(z))P(z|x)
\end{aligned}
$$

Example: Front-door criteria

$$\begin{tikzcd} Y \arrow[r] \arrow[rr, bend left, leftrightarrow] & X \arrow[r] & Z \end{tikzcd}$$

$$P(y|do(x)) = \sum_z P(y|do(x), z)P(z|do(x)) \quad \text{Margin. over } Z$$

$$P(y|do(x)) = \sum_z P(y|do(z))P(z|x)$$

$$P(y|do(x)) = \sum_z P(z|x) \sum_{x'} P(y|do(z), x')P(x'|do(z))$$

$$\text{Margin. over } X$$

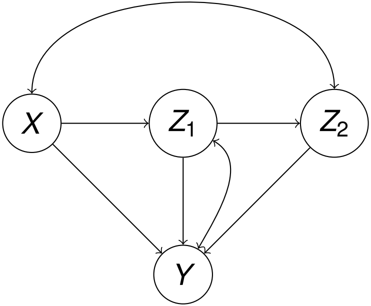
$$P(y|do(z), x') = P(y|z, x') \quad \text{by Rule 2}$$

$$P(x'|do(z)) = P(x') \quad \text{by Rule 3}$$

$$P(y|do(x)) = \sum_z P(z|x) \sum_{x'} P(y|z, x')P(x')$$

Examples

An example not satisfying back-door and front-door criteria but being identifiable



Completeness of do-calculus

The do-calculus is complete: if a causal estimand is identifiable, we can identify it by a sequence of rules of do-calculus[1]

Nonparametric identification: do-calculus tells us if we can identify a given causal estimand using only the causal assumptions encoded in the causal graph.
If we introduce more assumptions about the distribution (e.g. linearity), we can identify more causal estimands.

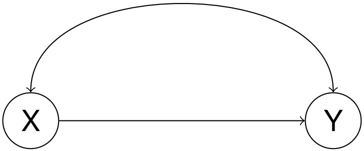
Proofs are constructive: there exist polynomial time algorithms for identification.

ID algorithm!

[1] Shipster and Pearl, 2006, Huand and Valtorta, 2006

Bow graph

The bow graph is not identifiable.



Proof:

- As there are only 2 variables, you can check what happens with every rule: it appears that nothing is doable.
- We can construct two causal models with the same joint distribution but different marginal distributions with $do(x)$. See Shpitser and Pearl, JMLR 2008, Thm 10

Details on board!

Some definitions

A graph $\mathcal{G}$ such that each vertex has at most one child, and only one vertex (called the root) has no children is called a tree.

A graph $\mathcal{G}$ such that each vertex has at most one child is called a forest.

A path where all directed arrowheads point at observable nodes, and never away from observable nodes is called a confounded path.

A graph $\mathcal{G}$ where any pair of observable nodes is connected by a confounded path is called a c-component (confounded component).

A graph $\mathcal{G}$ which is both a C-component and a tree is called a C-tree. We call a C-tree with a root node Y Y rooted.

First results

Theorem Let $\mathcal{G}$ be a Y-rooted C-tree. Let X be any subset of observable nodes in $\mathcal{G}$ which does not contain Y. Then $P(y|do(x))$ is not identifiable.

Theorem $P(y|do(pa(y)))$ is not identifiable if and only if there exists a subgraph of $\mathcal{G}$ which is a Y-rooted C-tree.

Remark For X a direct cause of Y, the arrow between X and Y is not enough, we should fix all other parents of Y.

First problem

Downward extension lemma Assume that $P(y|do(x))$ is not identifiable in a graph $\mathcal{G}$. Consider $\mathcal{G}'$ that contains all the nodes and edges of $\mathcal{G}$, and an additional node Z which is a child of all nodes in Y. Then, $P(z|do(x))$ is not identifibale in $\mathcal{G}'$.

Remark: identification of effects on a singleton is not any simpler than the general problem of identification of effect on a set.

Hedge

c-forest A graph $\mathcal{G}$ which is both a C-component and a forest is called a C-forest.

Hedge Let X, Y be sets of variables in $\mathcal{G}$. Let F, F' be R-rooted C-forest in $\mathcal{G}$ such that F' is a subgraph of F, X only occur in F, and $R \in An(Y)_{G_{\underline{X}}}$. Then F and F' form a hedge for $P(y|do(x))$.

Theorem Let F, F' be subgraphs of $\mathcal{G}$ which form a hedge for $P(y|do(x))$. Then $P(y|do(x))$ is not identifiable.

Theorem If $P(y|do(x))$ is not identifiable, then there is a hedge structure involved.

ID algorithm

ID(y,x,P,$\mathcal{G}$)
input: x, y, value assignments, P a probability distribution, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ a causal diagram.
output: expression for $P(y|do(x))$ in terms of P or FAIL(F, F')

If $x = \varnothing$ return $\sum_{v \in \mathcal{V} \setminus y} P(y, v)$

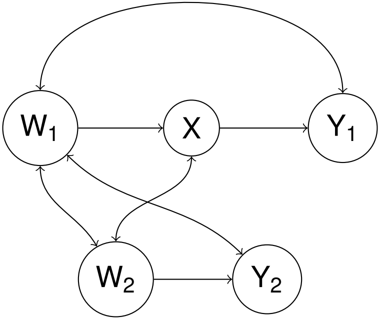
If $\mathcal{V} \setminus An(Y)_{\mathcal{G}} \neq \varnothing$ return $ID(y, x \cap An(Y)_{\mathcal{G}}, \sum_{\mathcal{V} \setminus An(Y)_{\mathcal{G}}} P, \mathcal{G}_{An(Y)})$

Let $W = (V \setminus X) \setminus An(Y)_{\mathcal{G}_{\overline{X}}}$. If $W \neq \varnothing$, return $ID(y, x \cup w, P, \mathcal{G})$

If $C(\mathcal{G} \setminus X) = \{S_1, \ldots, S_k\}$ return $\sum_{\mathcal{V} \setminus (y \cup x)} \prod_i ID(s_i, v \setminus s_i, P, \mathcal{G})$

If $C(\mathcal{G} \setminus X) = \{S\}$
If $C(\mathcal{G}) = \{\mathcal{G}\}$ throw FAIL$(\mathcal{G}, \mathcal{G} \cap S)$
If $S \in C(\mathcal{G})$ return $\sum_{s \setminus y} \prod i | V_i \in S P(v_i | v_{\pi}^{(-i1)})$
If $\exists S'$ such that $S \subset S' \in C(\mathcal{G})$ return
$ID(y, x \cap S', \prod_{i | V_i \in S'} P(V_i V_{\pi}^{(i-1)} \cap S', v_{\pi}^{(i-1)} \setminus S'), \mathcal{G}_{S'})$

Illustration of ID algorithm



Completeness of ID algorithm

Theorem ID is complete.

Whenever the algorithm fails, it is possible to cover a hedge from the C-components S and $\mathcal{G}$ considered for the subproblem where the failure occurs.
It can be shown that this hedge implies the non-identifiability of the original query with which the algorithm was invoked.
Then, hedges can be used to characterize all cases where effects of the form $P(y|do(x))$ cannot be identified from the observational distribution.

Theorem $P(y|do(x))$ is not identifiable if and only if $\mathcal{G}$ contains a hedge for some $P(y'|do(x'))$, where $y' \subseteq y, x' \subseteq x$.

Estimation

As soon as we deal with an identifiable request, we get a formula without 'do', so we can estimate every term using classical statistics (conditional probabilties).

This is particularly easy when dealing with linear regression or discrete variables.

Keep in mind that the theoretical properties of the estimators are not classical (sum of products of estimators: the convergence will be slower!)

Estimation

As soon as we deal with an identifiable request, we get a formula without 'do', so we can estimate every term using classical statistics (conditional probabilties).

This is particularly easy when dealing with linear regression or discrete variables.

Keep in mind that the theoretical properties of the estimators are not classical (sum of products of estimators: the convergence will be slower!)

Estimation

As soon as we deal with an identifiable request, we get a formula without 'do', so we can estimate every term using classical statistics (conditional probabilties).

This is particularly easy when dealing with linear regression or discrete variables.

Keep in mind that the theoretical properties of the estimators are not classical (sum of products of estimators: the convergence will be slower!)

References

- *Complete identification methods for the causal hierarchy*, Shpitser and Pearl, JMLR 2008
- *Causal inference in statistics: a primer*, Pearl, Glymour and Jewell, 2016
- *Causation, Prediction, and Search*, Spirtes, Glymour, Scheines, 1993