UNIVERSITAT DE BARCELONA

CAUSAL INFERENCE AND MACHINE LEARNING

About the course

The relationship between causality and artificial intelligence can be seen from two points of view: how causality can help solve some of the current problems of AI and how causal inference can leverage machine learning techniques. In this course we will review the two points of view with special emphasis on examples and practical cases.

01 Jordi
Introduction
Observational and Interventional Distributions. Causal Thinking.

02 Roger
Potential Outcomes
Fundamental Problem of Causal Inference

03 Jordi
Causal Graphs
Do Calculus

11:45-12:10 Coffee Break

04 Roger
Estimand-based Estimation
Metalearners

05 Jordi
Estimand-agnostic Estimation
Counterfactuals

14:00-15:30 Lunch

06 Jordi & Enrique
Causal Machine Learning
Supervised and Reinforcement Learning

07 Enrique
Practical Causal Inference
Exercises

17:30 End

UNIVERSITAT DE BARCELONA

Causal Graphs

DoCalculus

Jordi Vitrià
jordi.vitria@ub.edu

Universitat de Barcelona

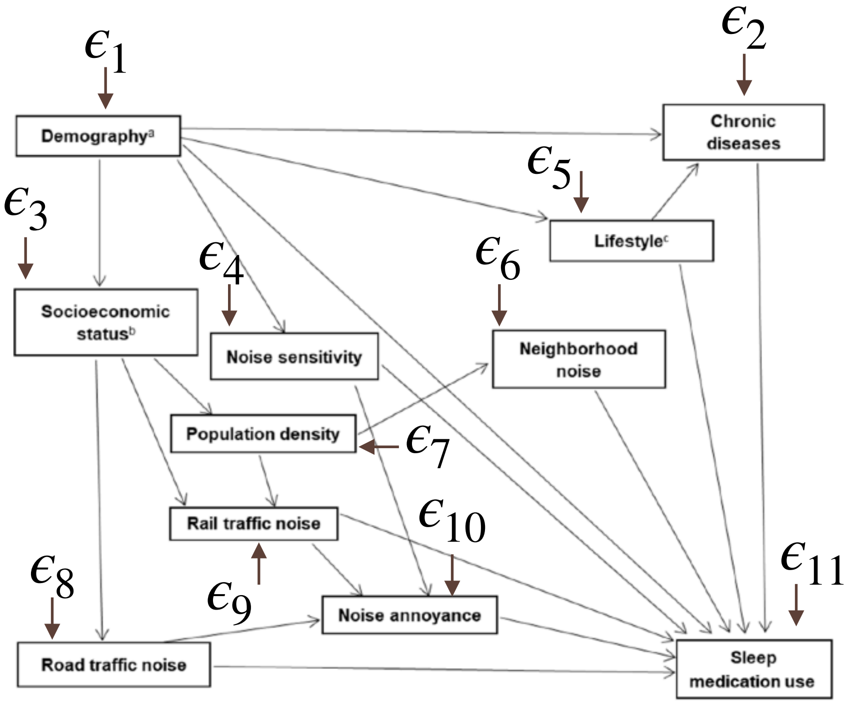
DAGs

DAGs encode the **analyst's qualitative causal assumptions** about the data-generating process in the population.

We assume that (i) the DAGs <u>display all observed and unobserved causes</u> in the process under investigation, and (ii) <u>all variables have independent error terms</u>.

DAGs

In this graph there are no hidden causes.



Error terms are not displayed in the DAG because they do not play any role in **non-parameteric identification**.

They are important to estimate SCMs.

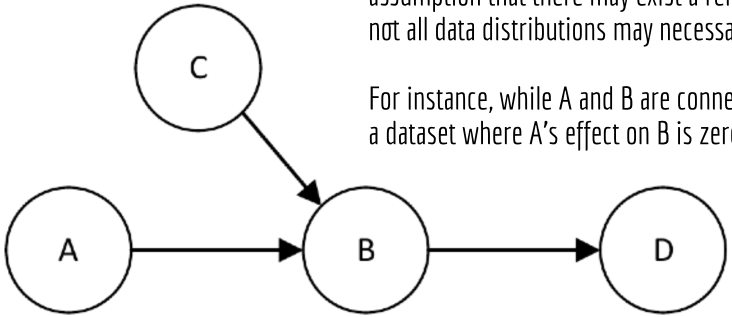
DAGs

Formally, a causal graph specifies a factorization of the joint probability distribution of data. Any probability distribution consistent with the graph needs to follow the specific factorization.

An edge between two nodes in a causal graph conveys the assumption that there may exist a relationship between them, but not all data distributions may necessarily follow it.

For instance, while A and B are connected via an edge, there can be a dataset where A's effect on B is zero.

The assumptions asserted by a causal graph are encoded by the missing edges in a graph, and the direction of edges



$$P(A, B, C, D) = P(D \mid A, B, C)P(B \mid C, A)P(C \mid A)P(A)$$ Chain Rule of Probability

$$= P(D \mid B)P(B \mid C, A)P(C)P(A)$$ Structure of the causal graph

DAGs

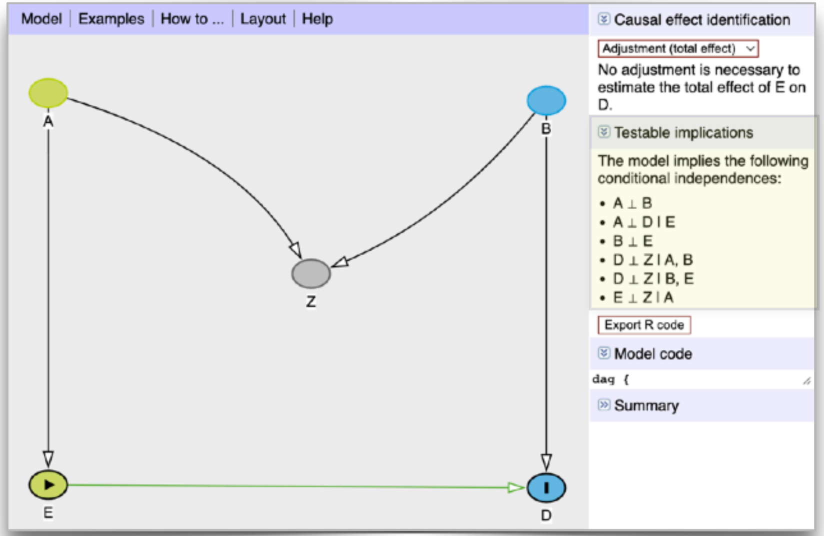
More generally, for any causal graph $\mathcal{G}$ over variables $V_1, V_2, \ldots, V_m$, the probability distribution of data is given by,

$$P(V_1, V_2, \ldots, V_m) = \Pi_{i=1}^{m} P(V_i \,|\, Pa(V_i))$$

where $Pa(V_i)$ refers to parents of V_i in the causal graph $\mathcal{G}$.

DAGs

DAGs are **falsifiable** through testable implications over the observed distributions, including conditional independence relationships between variables in the model.

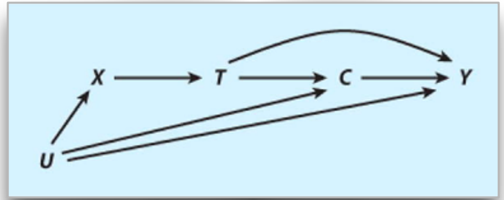


http://dagitty.net/

However, causal graphs cannot be learned from data alone: **every unique causal graph does not imply a unique set of independence tests**.

Every causal graph has an *equivalence class* of graphs that generate the same independence tests.

DAGs



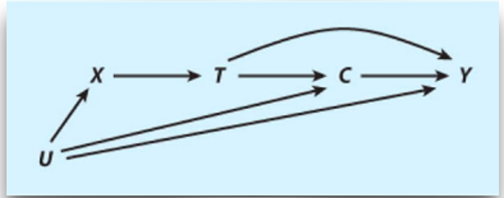
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6089543/

A **path** is sequence of arrows connecting two variables regardless of the direction of the arrowheads. A path can traverse one variable only once.

A **causal path** between two variables is a path in which all arrows point in the same direction.

$T \to C \to Y$ and $U \to X \to T \to C \to Y$ are causal paths

DAGs



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6089543/

The **set of all causal paths** between two variables comprise the **total causal effect**. All other paths are called non-causal or **spurious**.

The total causal effect of T on Y comprises $T \rightarrow Y$ and $T \rightarrow C \rightarrow Y$

$T \rightarrow C \leftarrow U \rightarrow Y$ and $T \leftarrow X \leftarrow U \rightarrow Y$ are noncausal paths

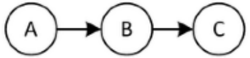
If two arrows along a path both point directly into the same variable, the variable is called a **collider variable**.

C is a collider on the path $T \rightarrow C \leftarrow U$

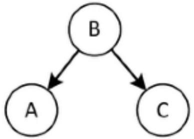
DAGs

The power of DAGs lies in their ability to reveal all marginal and conditional associations and independences implied by a qualitative causal model.

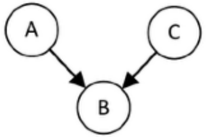
Absent sampling bias, **all observable associations** originate from just three elementary configurations: **chains**, **forks** and **inverted forks**.



A indirectly causes C via B
A and C are independent conditional on B



A and C are not independent.
A and C are independent conditional on B



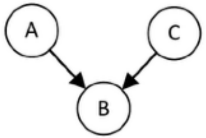
A via C are independent,
A and C are not independent conditional on B

These configurations correspond to the following sources of association between A and C: **causality**, **confounding bias** and **endogenous selection bias**.

If conditioning on B

$$P(A|B) \neq P(A|B,C)$$

DAGs



A via C are independent
A and C are not independent conditional on B

Consider the relationship between talent, A, beauty, C, and Hollywood success, B.

> Beauty and talent are not associated (<u>beauty does not cause talent, talent does not cause beauty, and beauty and talent do not share a common cause</u>), but beauty and talent are sufficient for becoming a successful Hollywood actor.

Now, condition on the collider by looking at the relationship between beauty and talent only among successful Hollywood actors: knowing that a talentless person is a successful actor implies that the person must be beautiful.

DAGs

All DAGs are build from chains, forks and inverted forks. Therefore, understanding the associational implications of these structures is sufficient for conducting **nonparametric identification analysis in arbitrarily complicated causal models**.

A path between two variables A and B does not transmit association and is said to be blocked, closed, or **d-separated** if:

1. the path contains a non-collider, C, that has been conditioned on (f.e. $A \rightarrow \boxed{C} \rightarrow B$ or $A \leftarrow \boxed{C} \rightarrow B$); or if

2. the path contains a collider, C, and neither the collider nor any of its descendants have been conditioned on, (f.e. $A \rightarrow C \leftarrow B$)

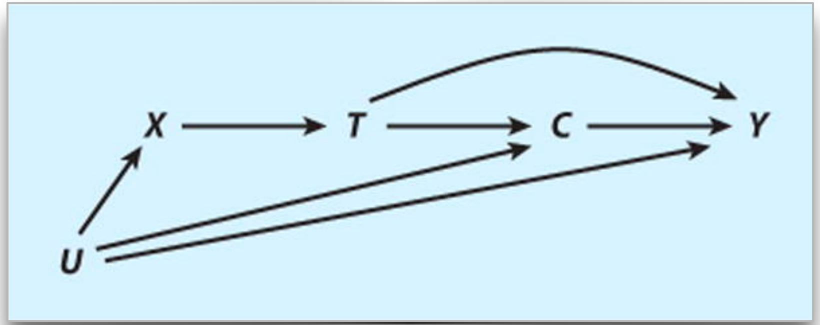
DAGs

1. Two variables that are d-separated along all paths are statistically independent.

2. Two variables that are d-connected along at least one path are associated.

One can determine the identifiability of a causal effect between A and B by checking whether one can block all non-causal paths between A and B by conditioning on a suitable set of observed features.

DAGs



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6089543/

There are 5 paths between T and Y.

The total causal effect of T and Y can be **identified** by conditioning on $\boxed{X}$ because:

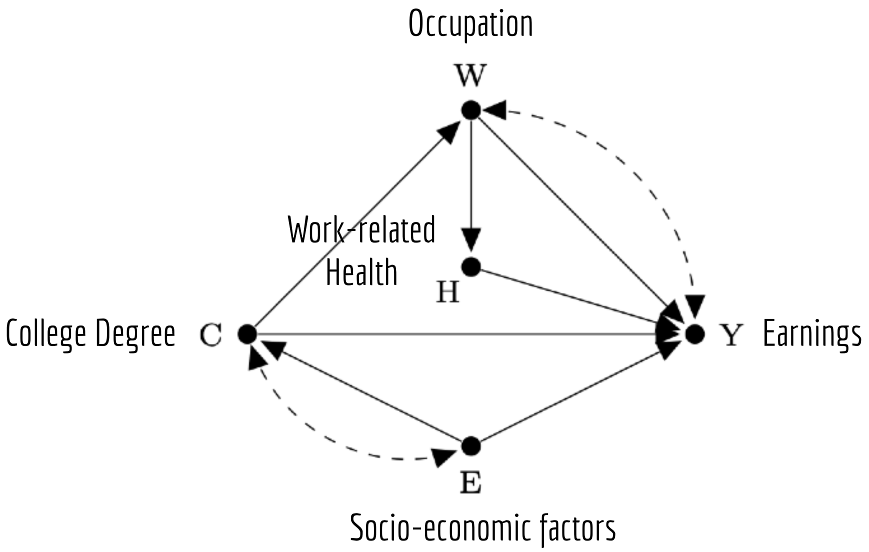
1. X does not sit on a causal path from T to Y.

2. Conditioning on X block the two non-causal paths between T and Y ($T \leftarrow \boxed{X} \leftarrow U \rightarrow Y$ and $T \leftarrow \boxed{X} \leftarrow U \rightarrow C \rightarrow Y$)

A third non-causal path $T \rightarrow C \leftarrow U \rightarrow Y$ is unconditionally blocked because it contains C.
Be aware that conditioning on C would ruin identification.

DAGs

One of the biggest threats to causal inference is **confounding bias** (correlation driven by a set of common causes).



We want to estimate the effect of a college degree C in earnings Y.

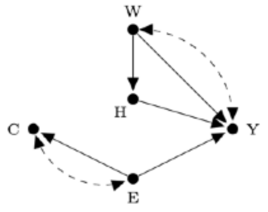
DAGs

The following graphical criterion can be used to find admissible adjustment sets that eliminate any confounding influences between C and Y.

> ***Backdoor criterion***: *Given and ordered pair of treatment and outcome variables (X, Y) in G, a set of variables Z is backdoor admissible if it blocks every path between X and Y in G^X.*
>
> *If a set of variables satisfies the backdoor criterion relative to (X, Y), the causal effect of X on Y can be identified by:*

$$P(Y \mid do(X)) = \sum_z P(Y \mid X, Z)P(Z)$$

In our example, $Z = \{E\}$ satisfies the backdoor criterion.



G^X

DAGs

Identification via backdoor adjustment requieres that all backdoor paths can be bloacked by a set of observed nodes, which is not always feasible. But there are other strategies!

> ***Conditional frontdoor criterion****: A set of variables Z is said to satisfy the conditional frontdoor criterion relative to a triplet (X, Y, W) if:*
>
> 1. *Z intercepts all causal paths from X to Y.*
> 2. *There is no unblocked backdoor path from X to Z given W.*
> 3. *All backdoor paths from Z to Y are blocked by $\{X, W\}$.*
>
> *In this case, the causal effect of X on Y can be identified by:*

$$P(Y = y | do(X = x)) = \sum_{m,w} P(m | X = x, w)P(W) \sum_{x'} P(Y = y | w, m, X = x')P(X = x' | w)$$

DAGs

The backdoor and frontdoor offer simple graphical rules that are easy to check, but they only represent a subset of the overall identification results that are derivable in DAGs.

In more generality, identifiability of any query of the form $P(Y \mid do(X))$ can be decided systematically by using a symbolic causal inference engine called **do-calculus**.

DAGs

Let X, Y, Z, W be arbitrary disjoints sets of nodes in G.
Let G_X be the mutilated graph that is obtained by removing all arrows pointing to nodes of X.
Let G^X be the graph that results from deleting all arrows that are emitted by X.

Do-calculus is based on 3 rules.

Rule 1: Insertion/deletion of observations:

$$p(Y | do(X), Z, W) = p(Y | do(X), W) \text{ if } (Y \perp\!\!\!\perp Z | X, W)_{G_X}$$

Rule 2: Observation exchange:

$$p(Y | do(X), do(Z), W) = p(Y | do(X), Z, W) \text{ if } (Y \perp\!\!\!\perp Z | X, W)_{G_X^Z}$$

Rule 3: Insertion/deletion of actions:

$$p(Y | do(X), do(Z), W) = p(Y | do(X), W) \text{ if } (Y \perp\!\!\!\perp Z | X, W)_{G_{X,Z(W)}}$$

where $Z(W)$ is the set of nodes of Z that are not ancestors of any node of W in G_X.

DAGs

It is guaranteed to return a solution,
whenever this solution exists.

Do-calculus was proved **sound and complete** for general queries of the form $P(Y \mid do(X), Z)$ by Pearl et al. (for graphs including unobserved confounders)

This result can also be seen algorithmically and the **identification of causal effects becomes a straightforward exercise.**