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

Estimand-agnostic Estimation

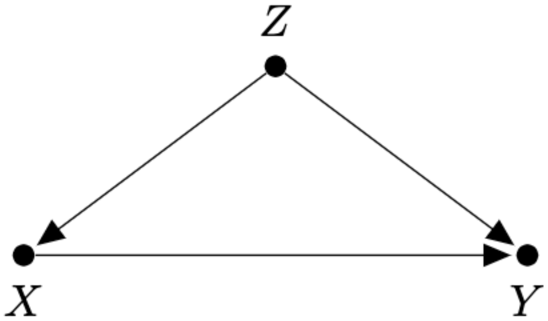
Counterfactuals

Jordi Vitrià
jordi.vitria@ub.edu

Universitat de Barcelona

SCMs

One key piece of information that is not included in the representation of the graph is the **functional relationship between nodes**.



$$P(X,Y,Z) = \overset{f_1}{P(Z)}\,\overset{f_2}{P(X\,|\,Z)}\,\overset{f_3}{P(Y\,|\,X,Z)}$$

SCMs

The causal diagram can be seen as a representation of an underlying **structural causal model** (generative model).

A structural causal model (SCM) is comprised of three components:

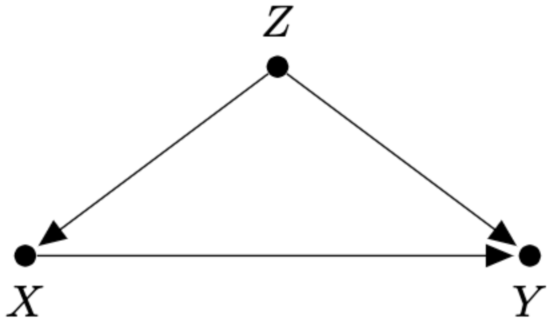
1. A set of **variables** describing the state of the universe and how it relates to a particular data set we are provided.

2. **Causal model** (DAG), which describe the causal effect variables have on one another.

3. A **probability distribution** defined over observed variables in the model, describing the likelihood that each variable takes a particular value.

SCMs

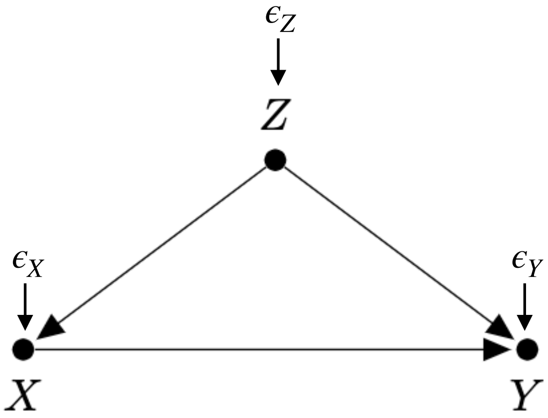
	Sex	Race	Height	Income	Marital Status	Years of Educ.	Liberal-ness
R1001	M	1	70	50	1	12	1.73
R1002	M	2	72	100	2	20	4.53
R1003	F	1	55	250	1	16	2.99
R1004	M	2	65	20	2	16	1.13
R1005	F	1	60	10	3	12	3.81
R1006	M	1	68	30	1	9	4.76
R1007	F	5	66	25	2	21	2.01
R1008	F	4	61	43	1	18	1.27
R1009	M	1	69	67	1	12	3.25



$$P(X,Y,Z) = \overset{f_1}{P(Z)}\,\overset{f_2}{P(X \mid Z)}\,\overset{f_3}{P(Y \mid X,Z)}$$

SCMs

We can estimate f_i from data by using statistical/ML methods.



$$Z \leftarrow f_1(\epsilon_Z)$$
$$X \leftarrow f_2(Z, \epsilon_X)$$
$$Y \leftarrow f_3(X, Z, \epsilon_Y)$$

Structural Causal Model

ϵ_i are independent **exogenous background factors** represented by an arbitrary noise distribution.

SCMs

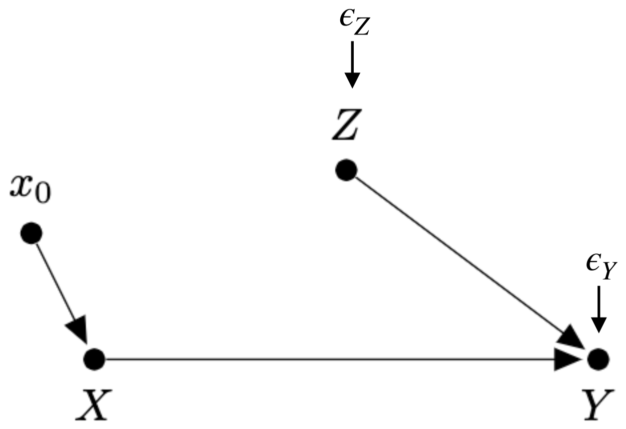
Actions can now be defined as interventions on variables in the model. For example, intervening on X amounts to deleting f_2 and setting X to a constant value x_0.

$$Z \leftarrow f_1(\epsilon_Z)$$

$$X \leftarrow x_0$$

$$Y \leftarrow f_3(X, Z, \epsilon_Y)$$

Modified
Structural Causal Model



SCMs

Given a certain observational sample $e = (x_e, y_e, z_e)$ and an intervention $do(X = x_q)$, a **counterfactual** is the result of an hypothetical experiment in the past, what would have happened to the value of variable Y had we intervened on X by assigning value x_q.

Identifiable counterfactuals can be computed as a three-step process by using a SCM:

1. **Abduction**: compute the posterior distribution of $(\epsilon_X, \epsilon_Y, \epsilon_Z)$ conditioned on e.
2. **Intervention**: apply the desired intervention $do(X = x)$
3. **Prediction**: compute the required prediction in the intervened distribution.

Counterfactuals (example)

Let's support that we want to rent an apartment and we train a model with real data to predict a price.

After entering all the details about size, location, whether pets are allowed and so on, the model tells us that we can charge 900€.

How could we get (by doing an intervention) 1000€? We can play with the feature values of the apartment to see how we can improve the value of the apartment!

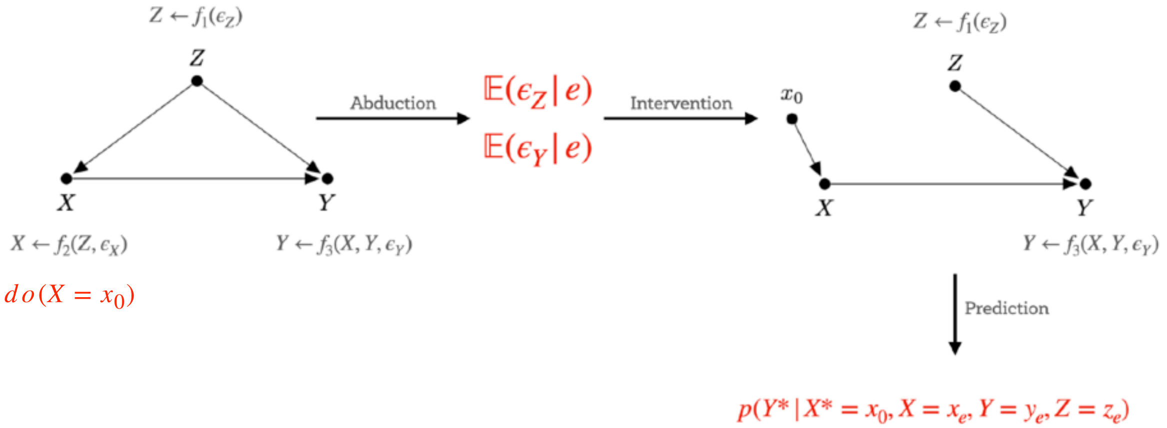
We find out that the apartment could be rented out for over 1000 Euro, if it were 15 m2 larger. Interesting, but non-actionable knowledge, because we cannot enlarge the apartment.

Finally, by tweaking only the feature values under our control (built-in kitchen yes/no, pets allowed yes/no, type of floor, etc.), we find out that if we allow pets and install windows with better insulation, we can charge 1000€.

SCMs

e

R1001	M	1	70	50	1	12	1.73
R1002	M	2	72	100	2	20	4.53
R1003	F	1	55	250	1	16	2.99
R1004	M	2	65	20	2	16	1.13
R1005	F	1	60	10	3	12	3.81
R1006	M	1	68	30	1	9	4.76
R1007	F	5	66	25	2	21	2.01



SCMs

An SCM encodes the intervened distribution, from which we can **sample** and compute causal queries and counterfactual queries (if they are identifiable).

For example, we can compute this causal query

$$\mathbb{E}(Y \mid do(X = x_1)) - \mathbb{E}(Y \mid do(X = x_0))$$

as

$$\sum_e P(Y* \mid X* = x_1, X = x_e, Y = y_e, Z = z_e) - \sum_e P(Y* \mid X* = x_0, X = x_e, Y = y_e, Z = z_e)$$

SCMs

We can also compute counterfactual queries, which are very interesting for explainability and fairness analysis.

Would my salary be higher if I were non-black?

For every individual e we only see $Pr(Y = y_e | A = \texttt{black})$ or $Pr(Y = y_e | A = \texttt{non\_black})$ (not both!), but we can consider its counterfactual.

SCMs

Individual Counterfactual Fairness (ICF), for individual

$$Pr(Y* = y_e | A* = \texttt{non\_black}) = Pr(Y = y_e | A = \texttt{black})$$

Would the salary be different if I were $A =$ *non_black* instead of $A =$ *black*?

Counterfactual Parity (CP),

$$\mathbb{E}[Pr(Y* | A* = \texttt{non\_black})] = \mathbb{E}[Pr(Y* | A* = \texttt{black})]$$

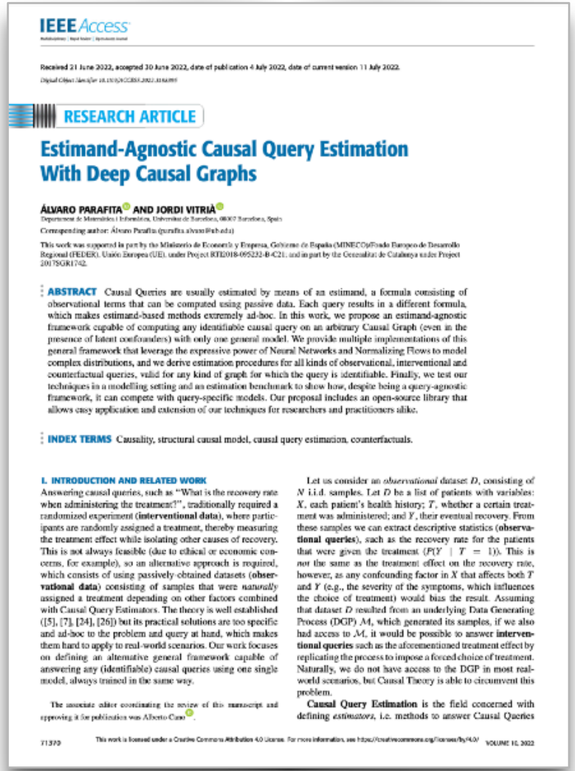
Would the mean salary be different if everyone were *black*?

Conditional Counterfactual Parity (CCP),

$$\mathbb{E}[Pr(Y* | A* = \texttt{non\_black}, X)] = \mathbb{E}[Pr(Y* | A* = \texttt{black}, X)]$$

Would the mean salary be different if everyone were black, **conditioned on education**?

SCMs



IEEE Access

Received 21 June 2022, accepted 30 June 2022, date of publication 4 July 2022, date of current version 11 July 2022.

RESEARCH ARTICLE

Estimand-Agnostic Causal Query Estimation With Deep Causal Graphs

ÁLVARO PARAFITA AND JORDI VITRIÀ
Departament de Matemàtiques i Informàtica, Universitat de Barcelona, 08007 Barcelona, Spain
Corresponding author: Álvaro Parafita (parafita.alvaro@ub.edu)

This work was supported in part by the Ministerio de Economía y Empresa, Gobierno de España (MINECO)/Fondo Europeo de Desarrollo Regional (FEDER), Unión Europea (UE), under Project RTI2018-095232-B-C21, and in part by the Generalitat de Catalunya under Project 2017SGR1742.

ABSTRACT Causal Queries are usually estimated by means of an estimand, a formula consisting of observational terms that can be computed using passive data. Each query results in a different formula, which makes estimand-based methods extremely ad-hoc. In this work, we propose an estimand-agnostic framework capable of computing any identifiable causal query on an arbitrary Causal Graph (even in the presence of latent confounders) with only one general model. We provide multiple implementations of this general framework that leverage the expressive power of Neural Networks and Normalizing Flows to model complex distributions, and we derive estimation procedures for all kinds of observational, interventional and counterfactual queries, valid for any kind of graph for which the query is identifiable. Finally, we test our techniques in a modelling setting and an estimation benchmark to show how, despite being a query-agnostic framework, it can compete with query-specific models. Our proposal includes an open-source library that allows easy application and extension of our techniques for researchers and practitioners alike.

INDEX TERMS Causality, structural causal model, causal query estimation, counterfactuals.

I. INTRODUCTION AND RELATED WORK

Answering causal queries, such as "What is the recovery rate when administering the treatment?", traditionally required a randomized experiment (**interventional data**), where participants are randomly assigned a treatment, thereby measuring the treatment effect while isolating other causes of recovery. This is not always feasible (due to ethical or economic concerns, for example), so an alternative approach is required, which consists of using passively-obtained datasets (**observational data**) consisting of samples that were *naturally* assigned a treatment depending on other factors combined with Causal Query Estimators. The theory is well established ([5], [7], [24], [26]) but its practical solutions are too specific and ad-hoc to the problem and query at hand, which makes them hard to apply to real-world scenarios. Our work focuses on defining an alternative general framework capable of answering any (identifiable) causal queries using one single model, always trained in the same way.

The associate editor coordinating the review of this manuscript and approving it for publication was Alberto Cano.

Let us consider an *observational* dataset D, consisting of N i.i.d. samples. Let D be a list of patients with variables: X, each patient's health history; T, whether a certain treatment was administered; and Y, their eventual recovery. From these samples we can extract descriptive statistics (**observational queries**), such as the recovery rate for the patients that were given the treatment ($P(Y \mid T = 1)$). This is *not* the same as the treatment effect on the recovery rate, however, as any confounding factor in X that affects both T and Y (e.g., the severity of the symptoms, which influences the choice of treatment) would bias the result. Assuming that dataset D resulted from an underlying Data Generating Process (DGP) $\mathcal{M}$, which generated its samples, if we also had access to $\mathcal{M}$, it would be possible to answer **interventional queries** such as the aforementioned treatment effect by replicating the process to impose a forced choice of treatment. Naturally, we do not have access to the DGP in most real-world scenarios, but Causal Theory is able to circumvent this problem.

Causal Query Estimation is the field concerned with defining *estimators*, i.e. methods to answer Causal Queries

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9815055