Lecture 1: Causal Inference

Tianpei Xie

Sep. 5th., 2022

Contents

1 Motivation: association does not imply causation **2**

1.1 Simpson's Paradox 2

1.2 Main Themes 3

2 Potential Outcomes **4**

2.1 Potential Outcomes and Individual Treatment Effects 4

2.2 The Fundamental Problem of Causal Inference 5

2.3 Average Treatment Effects and Causal Assumptions 5

2.3.1 Ignorability and Exchangeability 5

2.3.2 Conditional Exchangeability and Unconfoundedness 6

2.3.3 Positivity/Overlap and Extrapolation 7

2.3.4 No interference, Consistency, and SUTVA 8

3 Structural Causal Models **10**

3.1 Structural Equations 11

4 Basic process in Causal Inference **12**

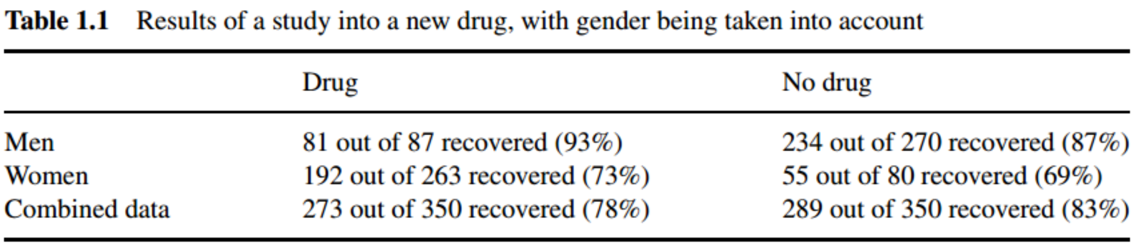

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug	No drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

Figure 1: The example for Simpson paradox. The drug ha positive impact on each gender but not in total population. [Glymour et al., 2016]

1 Motivation: association does not imply causation

- The aim of ***standard statistical analysis***, typified by regression, estimation, and hypothesis testing techniques, is to assess parameters of a **distribution** from samples drawn of that distribution.

- ***Causal analysis*** aim at inferring not only beliefs or probabilities under static conditions, but also the ***dynamics*** of beliefs under ***changing conditions***, for example, changes induced by treatments or external interventions.

- there is nothing in a distribution function to tell us how that distribution would differ ***if external conditions were to change*** – say from ***observational*** to ***experimental*** setup – because the laws of probability theory do not dictate how one property of a distribution ought to change when another property is modified. This information must be provided by ***causal assumptions*** which identify relationships that remain invariant when external conditions change. [Pearl, 2009]

- ***Association does not imply causation***: one cannot *substantiate* **causal** claims from ***associations*** alone, even at the population level – behind every causal conclusion there must lie some *causal assumption* that is **not testable in observational studies**. [Pearl, 2009]

- An ***associational concept*** is any relationship that can be defined in terms of a **joint distribution** of observed variables, and a ***causal concept*** is any relationship that ***cannot*** *be defined* from the distribution alone.

  - Examples of *associational concepts* are: *correlation*, *regression*, *dependence*, *conditional independence*, *likelihood*, *collapsibility*, propensity score, *risk ratio*, *odds ratio*, *marginalization*, *conditionalization*, controlling for.

  - Examples of *causal concepts* are: ***randomization***, influence, effect, ***confounding***, holding constant, disturbance, ***spurious correlation***, faithfulness/stability, instrumental variables, ***intervention***, explanation, ***attribution***, and so on.

1.1 Simpson's Paradox

- **Simpson's Paradox** refers to the existence of data in which a *statistical association* that holds for an ***entire population*** is **reversed** in ***every subpopulation***.

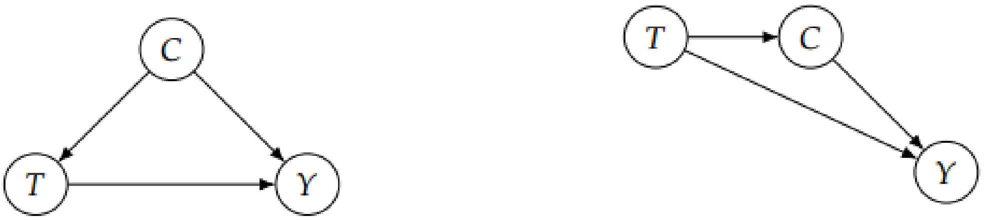

Figure 1.1: Causal structure of scenario 1, where condition C is a common cause of treatment T and mortality Y. Given this causal structure, treatment B is preferable.

Figure 1.2: Causal structure of scenario 2, where treatment T is a cause of condition C. Given this causal structure, treatment A is preferable.

Figure 2: The causal dependency of external condition C to precscription of treatment T and its outcome Y. [Neal, 2020]

Figure 1 shows an example for a drug test. The first row shows the outcome for male patients; the second row shows the outcome for female patients; and the third row shows the outcome for all patients, regardless of gender. In male patients, drug takers had a better recovery rate than those who went without the drug (93% vs 87%). In female patients, again, those who took the drug had a better recovery rate than nontakers (73% vs 69%). However, in the combined population, those who did not take the drug had a better recovery rate than those who did (83% vs 78%).

The data seem to say that if we know the patients gender – male or female – we can prescribe the drug, but if the gender is unknown we should not!

- ***Causality*** is essential to solve Simpsons paradox. The solution depends on the causal structure.

Scenario 1: If there is an external condition C that **both causes the treatment** T of drug, **and affects** the recovery Y. Like Figure 2 left. For instance, in the table in 1, a portion of element in the drug will negatively affect women's recovery more than men. Also we see that women are significantly more likely to take the drug than men are. So, the reason the drug appears to be harmful overall is that, if we select a drug user at random, that person is more likely to be a woman and hence less likely to recover than a random person who does not take the drug. Condition C is "being women".

Scenario 2: If the ***prescription*** of treatment T is a ***cause*** of condition C, and both T and C affect the outcome Y. Like Figure 2 right. For instance, presciption of the drug requires a long delay before the treatment and the long delay negatively affects the recovery.

- The key assumption missing in Simpson's paradox is that the *treatment cannot not affect sex*. In other words, it assumes that *the **partition variable** itself cannot be a factor determing the outcome*. If it could, there would be no paradox since the causal strcuture will explain the result.

1.2 Main Themes

- **Statistical vs. Causal**: Even with an infinite amount of data, **we sometimes cannot compute some causal quantities**. In contrast, much of statistics is about *addressing* *uncertainty* in finite samples. Association, a statistical concept, is not causation. There is more work to be done in causal inference, even after starting with infinite data. This is the **main distinction** motivating ***causal inference***.

- **Identification vs. Estimation**: *Identification* of causal effects is **unique** and challenging in causal inference. However, causal inference also shares *estimation* with traditional statistics and machine learning.
- **Interventional vs. Observational**: If we can **intervene/experiment**, identification of causal effects is relatively easy. We can simply measure the effect after we take that action. *Observational* data is where it gets more complicated because **confounding** is almost always introduced into the data.
- **Assumptions**: There will be a large focus on what assumptions we are using to get the results that we get. Each assumption will have its own box to help make it difficult to not notice. **Clear assumptions** should make it easy to see where critiques of a given causal analysis or causal model will be. The hope is that presenting assumptions clearly will lead to more lucid discussions about causality.

2 Potential Outcomes

2.1 Potential Outcomes and Individual Treatment Effects

We will use $T \in \{0, 1\}$ to denote the random variable for treatment, Y to denote the random variable for the outcome of interest and X to denote covariates.

- The ***potential outcome*** $Y(t)$ denotes what your outcome **would be**, if you **were** to take treatment $T = t$. A potential outcome $Y(t)$ is distinct from the ***observed*** outcome Y in that *not all potential outcomes are observed*. Rather all potential outcomes can potentially be observed. The one that is *actually observed* depends on the value that the treatment T takes on.
- **Before** treatment, any outcome is *potential outcome.* **After** treatment, one outcome is *observed outcome* (i.e. $(Y(1) \,|\, T = 1)$ and $(Y(0) \,|\, T = 0)$) and the rest are ***counterfactual outcomes*** (i.e. $(Y(1) \,|\, T = 0)$ and $(Y(0) \,|\, T = 1)$).
- Although only one of the outcome can be observed for one individual, there are many people in the population. We will denote the treatment, covariates, and outcome of the i-th individual using T_i, X_i, and Y_i. Then, we can define the ***individual treatment effect (ITE)*** [Neal, 2020] for individual i:

$$\tau_i = Y_i(1) - Y_i(0) \tag{1}$$

  Whenever there is more than one individual in a population, $Y(t)$ is a random variable because different individuals will have different potential outcomes. In contrast, $Y_i(t)$ is usually treated as **non-random** because the subscript i means that we are conditioning on so much individualized (and context-specific) information, that we restrict our focus to a single individual (in a specific context) whose potential outcomes are deterministic.

- ITEs are some of the main quantities that we care about in causal inference.

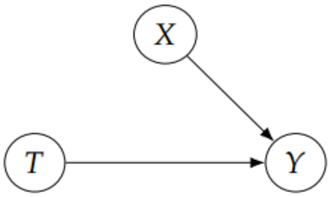

Figure 2.2: Causal structure when the treatment assignment mechanism is ignorable. Notably, this means there's no arrow from X to T, which means there is no confounding.

Figure 3: The Causal structure when the treatment assignment mechanism is ignorable. [Neal, 2020]

2.2 The Fundamental Problem of Causal Inference

It is impossible to observe **all potential outcomes** for a given individual in (1). This is known as ***the fundamental problem of causal inference***. It is fundamental because if we cannot observe both $Y_i(1)$ and $Y_i(0)$, then we cannot observe the causal effect $Y_i(1) - Y_i(0)$. This problem is **unique** to causal inference because, in causal inference, we care about making causal claims, which are defined in terms of potential outcomes.

The potential outcomes that you *do not (and cannot) observe* are known as ***counterfactuals*** because they are counter to fact (reality). In contrast, the potential outcome that is **observed** is sometimes referred to as a *factual*.

2.3 Average Treatment Effects and Causal Assumptions

- Given a popuIation, we can compute the ***average treatment effect (ATE)***, or, *average causal effect (ACE)* [Neal, 2020] by taking an average over the ITEs:

$$\tau := \mathbb{E}\left[Y_i(1) - Y_i(0)\right] = \mathbb{E}\left[Y(1)\right] - \mathbb{E}\left[Y(0)\right] \tag{2}$$

  where the average is over the individuals i if $Y_i(t)$ is deterministic. If $Y_i(t)$ is random, the average is also over any other randomness.

- A natural quantity that comes to mind is the ***associational difference***:

$$\delta := \mathbb{E}\left[Y|T=1\right] - \mathbb{E}\left[Y|T=0\right] \tag{3}$$

  Note that in general, (3) is not equal to (2) since $\mathbb{E}\left[Y|T=1\right]$ is the expectation of outcome on a ***sub-population*** with treatment $T = 1$ while $\mathbb{E}\left[Y(1)\right]$ is the expectation of outcome if the treatment ***on the whole population was*** $T = 1$. The former is an associational quality and association does not equal to causation. δ is comparing ***two populations*** of people while τ is comparing what would happen if the **same people** were given different treatments.

2.3.1 Ignorability and Exchangeability

- Although one of potential outcomes are not observed, we can choose to ignore them and compute $\mathbb{E}\left[Y(1)\right]$ by taking average over observed outcomes. Assuming **ignorability** is like

ignoring how people ended up **selecting** the treatment they selected and just **assuming they were randomly assigned** their treatment. Formally, we have

Assumption 2.1 *(**Ignorability / Exchangeability**) [Neal, 2020]*

$$(Y(1), Y(0)) \perp\!\!\!\perp T \tag{4}$$

The ignorability is a key assumption to causal inference because it allows us to reduce the ATE to the associational difference, i.e. $\tau = \mathbb{E}\left[Y(1)\right] - \mathbb{E}\left[Y(0)\right] = \mathbb{E}\left[Y|T=1\right] - \mathbb{E}\left[Y|T=0\right] = \delta$. Under this assumption, there is no confounding factor that allows the treatment selection have effects on the outcome after it is selected (i.e. ***no unmeasured confounders***). Figure 3 shows the causal structure behind exhangeablility.

- Remember $Y(0)$ and $Y(1)$ are both random variables themselves even if the treatment $T = 0, 1$ is fixed. $P(Y(0)\,|\,T=1)$ answers the ***counterfactual question*** ”what potential outcome *would be* if I *did not* select treatment ?”
- Another perspective on this assumption is that of ***exchangeability***. This assumption means $\mathbb{E}\left[Y(1)\,|T=0\right] = \mathbb{E}\left[Y(1)\,|T=1\right]$ and $\mathbb{E}\left[Y(0)\,|T=0\right] = \mathbb{E}\left[Y(0)\,|T=1\right]$, respectively. Then, this implies $\mathbb{E}\left[Y(1)\,|T=t\right] = \mathbb{E}\left[Y(1)\right]$ and $\mathbb{E}\left[Y(0)\,|T=t\right] = \mathbb{E}\left[Y(0)\right]$ for all t. Exchangeability means that the ***treatment groups are exchangeable*** in the sense that if they were **swapped**, the new treatment group would observe the **same outcomes** as the old treatment group, and the new control group would observe the same outcomes as the old control group. In this sense, our choice of test subjects are nothing special for treatement vs. control.

  An important **intuition** to have about exchangeability is that it guarantees that the ***treatment groups are comparable***. In other words, the control and treatment groups are the same in all relevant aspects other than the treatment. This intuition is what underlies the concept of ”**controlling** for” or ”adjusting for” variables.
- **Definition** (***Identifiability***) [Neal, 2020]
  A *causal* quantity (e.g. $\mathbb{E}\left[Y(t)\right]$) is ***identifiable*** if we can compute it from a purely statistical quantity (e.g. $\mathbb{E}\left[Y\,|T=t\right]$).

  To identify a causal effect is to reduce a causal expression to a purely statistical expression. It means to reduce the potential outcomes to conditioning and expectation.
- ***Randomized experiments*** are the key for ignorability assumption to hold. With *randomization*, we can safely assume that both observed and unobserved covariates are ***balanced***.
- Under randomized experiments, the ***expectation*** of ITE can be estimated by the ***average outcomes*** from both control and treatment groups. Moreover, the expected outcome given only one treatment $T = 1$ for entire population $\mathbb{E}\left[Y(1)\right]$ can be estimated by the average outcome from the treatment group. Same for $T = 0$.

2.3.2 Conditional Exchangeability and Unconfoundedness

- There is no reason to expect that the groups are the same in all relevant variables other than the treatment. However, if we **control** for relevant variables by conditioning, then maybe the subgroups will be exchangeable.

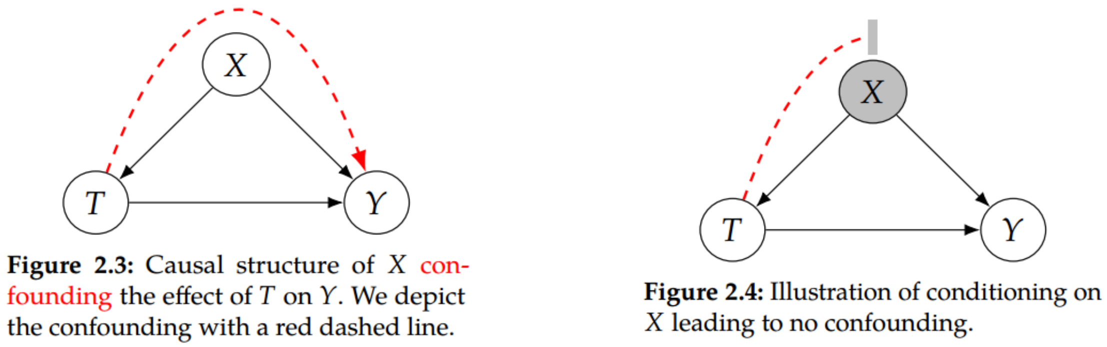

Figure 2.3: Causal structure of X confounding the effect of T on Y. We depict the confounding with a red dashed line.

Figure 2.4: Illustration of conditioning on X leading to no confounding.

Figure 4: Illustration of conditioning on X leading to no confounding. [Neal, 2020]

Assumption 2.2 ***(Conditional Exchangeability / Unconfoundedness)*** *[Neal, 2020]*

$$(Y(1), Y(0)) \perp\!\!\!\perp T \mid X \tag{5}$$

The idea is that although the treatment and potential outcomes may be unconditionally associated (due to confounding), ***within levels of*** X, they are not associated. In other words, there is no confounding within levels of X because controlling for X has made the treatment groups comparable (i.e. randomly assigned). Figure 4 illustrate conditioning on X the treatement and potential outcomes are independent.

- **Conditional exchangeability is the main assumption necessary for causal inference**. Armed with this assumption, we can identify the causal effect within levels of X, just like we did with (unconditional) exchangeability

$$\begin{aligned} \mathbb{E}\left[Y_i(1) - Y_i(0)|X\right] &= \mathbb{E}\left[Y_i(1) \mid X\right] - \mathbb{E}\left[Y_i(0) \mid X\right] \\ &= \mathbb{E}\left[Y \,|T = 1,\, X\right] - \mathbb{E}\left[Y \,|T = 0,\, X\right] \qquad (6) \\ \tau = \mathbb{E}\left[Y_i(1) - Y_i(0)\right] &= \mathbb{E}_X\left[\mathbb{E}\left[Y_i(1) - Y_i(0)\right] |X\right] \\ &= \mathbb{E}_X\left[\mathbb{E}\left[Y \,|T = 1,\, X\right]\right] - \mathbb{E}_X\left[\mathbb{E}\left[Y \,|T = 0,\, X\right]\right] \qquad (7) \end{aligned}$$

- **Theorem 2.3** ***(Adjustment Formula)*** *[Neal, 2020]*
  Given the assumptions of ***unconfoundedness***, ***positivity***, ***consistency***, *and* ***no interference***, *we can identify the average treatment effect:*

$$\tau = \mathbb{E}\left[Y(1) - Y(0)\right] = \mathbb{E}_X\left[\mathbb{E}\left[Y \,|T = 1,\, X\right] - \mathbb{E}\left[Y \,|T = 0,\, X\right]\right]$$

- The conditioning process $(\mathbb{E}\left[Y \,|T = 1,\, X\right] - \mathbb{E}\left[Y \,|T = 0,\, X\right])$ in (7) is called ***stratification***, i.e. we focus our estimation of treatment effects within each stratum $X = x$. The process of stratification and averaging is called ***standarization***.

2.3.3 Positivity/Overlap and Extrapolation

- ***Positivity*** is the condition that all subgroups of the data with different covariates have some probability of receiving ***any value of treatment***. Formally, we define positivity for binary treatment as follows

Assumption 2.4 ***(Positivity / Overlap / Common Support)*** *[Neal, 2020]*
For all values of covariates x present in the population of interest (i.e. x such that $P(X = x) > 0$),

$$0 < P(T = 1|X = x) < 1 \tag{8}$$

This assumption make sure for all choice of $X = x$, the condition on both $T = 1$ and $X = x$ is well defined.

- If we have a positivity violation, that means that within some subgroup of the data, everyone always receives treatment or everyone always receives the control. It wouldnt make sense to be able to estimate a causal effect of treatment vs. control in that subgroup since we see only treatment or only control. We never see the alternative in that subgroup.
- Another name for positivity is ***overlap***. The intuition for this name is that we want the covariate distribution of the treatment group to overlap with the covariate distribution of the control group.
- There is a ***Positivity-Unconfoundedness Tradeoff***: Conditioning on more covariates could lead to a higher chance of satisfying unconfoundedness. But it can lead to a higher chance of violating positivity.
- Violations of the positivity assumption can actually lead to demanding too much from models and getting very bad behavior in return. Note that the estimate $\mathbb{E}[Y|T = t,\ X = x]$ requires all possible input pairs $(1, x)$ and $(0, x)$. These models will be forced to extrapolate in regions (using their parametric assumptions) where $P(T = 1, X = x) = 0$ and regions where $P(T = 0, X = x) = 0$.

2.3.4 No interference, Consistency, and SUTVA

- ***No interference*** means that my outcome is unaffected by anyone elses treatment. Rather, my outcome is only a function of my own treatment.

  Assumption 2.5 ***(No Interference)*** *[Neal, 2020]*

$$Y_i(t_1, \ldots, t_i, \ldots, t_n) = Y_i(t_i) \tag{9}$$

- ***Consistency*** is the assumption that the outcome we **observe** Y is actually the **potential outcome** under the **observed treatment** T.

  Assumption 2.6 ***(Consistency)*** *[Neal, 2020]*
  If the treatment is T, then the observed outcome Y is the potential outcome under treatment T. Formally,

$$T = t \Rightarrow Y = Y(t) \tag{10}$$

We could write this equivalently as follow:

$$Y = Y(T) \tag{11}$$

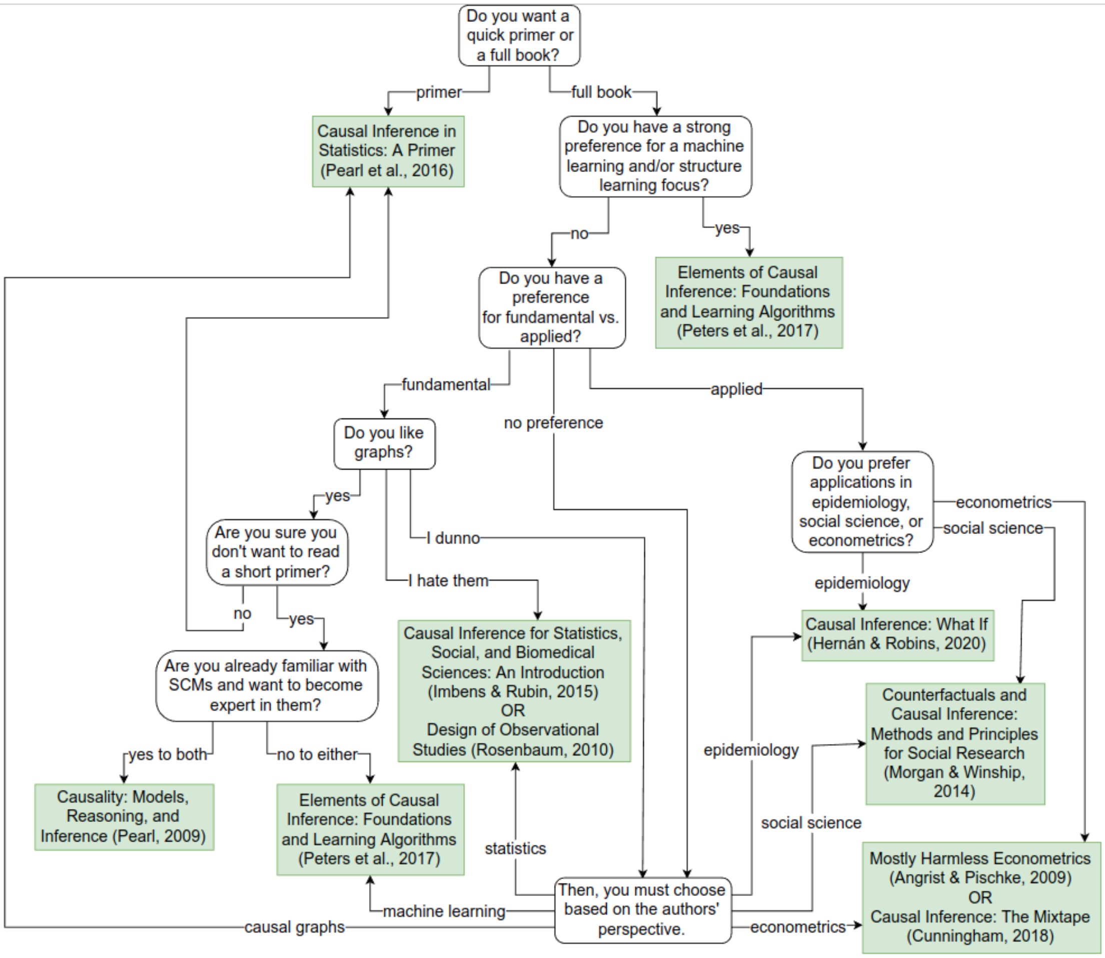
Do you want a quick primer or a full book?
primer
full book
Causal Inference in Statistics: A Primer (Pearl et al., 2016)
Do you have a strong preference for a machine learning and/or structure learning focus?
no
yes
Elements of Causal Inference: Foundations and Learning Algorithms (Peters et al., 2017)
Do you have a preference for fundamental vs. applied?
fundamental
no preference
applied
Do you like graphs?
yes
I dunno
I hate them
Are you sure you don't want to read a short primer?
no
yes
Are you already familiar with SCMs and want to become expert in them?
yes to both
no to either
Causality: Models, Reasoning, and Inference (Pearl, 2009)
Elements of Causal Inference: Foundations and Learning Algorithms (Peters et al., 2017)
Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction (Imbens & Rubin, 2015) OR Design of Observational Studies (Rosenbaum, 2010)
Do you prefer applications in epidemiology, social science, or econometrics?
econometrics
social science
epidemiology
Causal Inference: What If (Hernán & Robins, 2020)
Counterfactuals and Causal Inference: Methods and Principles for Social Research (Morgan & Winship, 2014)
Mostly Harmless Econometrics (Angrist & Pischke, 2009) OR Causal Inference: The Mixtape (Cunningham, 2018)
epidemiology
social science
statistics
machine learning
causal graphs
econometrics
Then, you must choose based on the authors' perspective.

Figure 5: Books for different areas in causal inference and analysis [Neal, 2020]

Note that T is different from t, and $Y(T)$ is different from $Y(t)$. T is a random variable that corresponds to the **observed treatment**, whereas t is a specific value of treatment. Similarly, $Y(t)$ is the potential outcome for some specific value of treatment, whereas $Y(T)$ is the potential outcome for the *actual* value of treatment that we **observe**.

- It may be confusing but the **potential outcome** $Y(t)$ **is a random variable itself due to intervention** $T = t$, and it is conceptually different from the random variable Y. Only via the consistency assumption (10) these two variables are connected.
- By consistency assumption, we have

$$\mathbb{E}\left[Y(1)\,|T=1\right] - \mathbb{E}\left[Y(0)\,|T=0\right] = \mathbb{E}\left[Y|T=1\right] - \mathbb{E}\left[Y|T=0\right]$$

- ***Stable unit-treatment value assumption (SUTVA)*** is satisfied if unit (individual) i's outcome is simply a **function** of unit i's treatment. Therefore, SUTVA is a combination of consistency and no interference (and also ***deterministic*** potential outcomes).

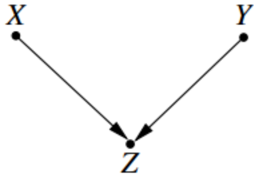

Figure 1.9 The graphical model of SCM 1.5.1, with X indicating years of schooling, Y indicating years of employment, and Z indicating salary

Figure 6: The graphical model representation of a structural causal model [Pearl, 2009]

3 Structural Causal Models

An alternative viewpoint to the ***Potential Outcome (PO) model*** [Imbens and Rubin, 2015, Rosenbaum, 2017] is the ***Structural Causal Model*** **(SCM)** [Pearl, 2000, 2009, Glymour et al., 2016]. This model as well as **Causal Bayesian Networks** [Pearl, 2000, 2009] are ways of describing the relevant features of the world and how they interact with each other based on **Directed Acyclic Graphs (DAG)**.

- **Definition** (***Structural Causal Model (SCM)***)
  A *structural causal model* is a tuple of the following sets:

  1. a set of *internal variables* $\boldsymbol{V}$ called ***endogenous*** variables;
  2. a set of *external variables* $\boldsymbol{U}$ called ***exogenous*** variables;
  3. and a set of ***functions*** $\boldsymbol{F}$ that determine the values of the the variables in $\boldsymbol{V}$ based on the values of the variables in $\boldsymbol{U}$.

- The exogenous variables in $\boldsymbol{U}$ are **external** to the model; we choose not to explain how they are caused. *Exogenous variables* ***cannot be descendants*** *of any other variables*, and in particular, cannot be a descendant of an endogenous variable; they have **no ancestors** and are represented as **root nodes** in graphs.

  Every endogenous variable in a model is a **descendant** of at least one ***exogenous*** variable.

- If we know the value of every exogenous variable, then using the functions in $\boldsymbol{F}$, we can determine with *perfect* certainty the value of every endogenous variable.

- A variable X is a *direct cause* of a variable Y if X appears in the function that assigns Y's value.

- Every SCM is associated with a ***graphical causal model***. Graphical models consist of a set of nodes representing the variables in $\boldsymbol{U}$ and $\boldsymbol{V}$, and a set of ***edges*** between the nodes representing the functions in $\boldsymbol{F}$.

- We can define **causality** in terms of directed graphical models: If, in a graphical model, a variable X is the **child** of another variable Y, then Y is a **direct cause** of X; if X is a ***descendant*** of Y, then Y is a ***potential cause*** of X. In this way, **the causal graphical model encodes the *causal relationship assumptions***. This is in contrast to the traditional directed graphical model, which encodes the *associational relationship assumptions* such as conditional independence.

Figure 6 shows an example for SCM. This model represents the salary Z that an employer pays an individual with X years of schooling and Y years in the profession. X and Y both appear in f_Z, so X and Y are both direct causes of Z. If X and Y had any ancestors, those ancestors would be potential causes of Z.

$$\boldsymbol{U} = \{X, Y\}, \boldsymbol{V} = \{Z\}, \boldsymbol{F} = \{f_Z\}$$
$$f_Z : Z = 2X + 3Y$$

3.1 Structural Equations

- In its general form, a ***functional causal model*** consists of a set of equations of the form

$$x_i = f_i(pa_i, u_i), \quad i = 1, \dots, n, \tag{12}$$

  where *parents* pa_i, stands for the set of variables that directly determine the value of X_i and where the U_i represent ***errors*** (or ***"disturbances"***) due to omitted factors. Equation (12) is a *nonlinear*, *nonparametric generalization* of the ***linear structural equation models*** *(SEMs)*

$$x_i = \sum_{k \neq 1} \alpha_{k,i} x_k + u_i, \quad i = 1, \dots, n, \tag{13}$$

  which have become a standard tool in economics and social science. In linear models, pa_i corresponds to those variables on the r.h.s. of (13) that have nonzero coefficients.

- The function forms in (12) can be *laws* in physics and the natural sciences. They describe a set of ***derterministic*** causal dependencies of X_i to its parents and external factors (PA_i, U_i). A set of equations in the form of (12) and in which each equation represents an *autonomous mechanism* is called a ***structural model***; if each variable has a distinct equation in which it appears on the left-hand side (called the *dependent variable*), then the model is called a ***structural causal model*** (or, a causal model for short).

- Note that compared to **potential outcome theory**,
  - the SCM views the intervention $do(x)$ as an ***operation*** that **changes the model** (and the distribution) but keeps all variables the same;
  - the potential-outcome approach views the **outcome variable** Y under $do(x)$ to be a **different variable**, $Y(x)$, loosely connected to Y through relations such as (10).

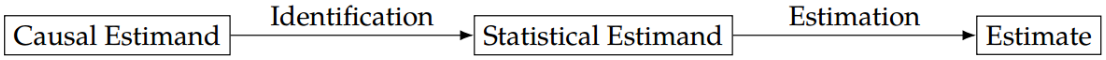

Figure 2.5: The Identification-Estimation Flowchart – a flowchart that illustrates the process of moving from a target causal estimand to a corresponding estimate, through identification and estimation.

Figure 7: The basic process of identification and estimation in causal inference. [Neal, 2020]

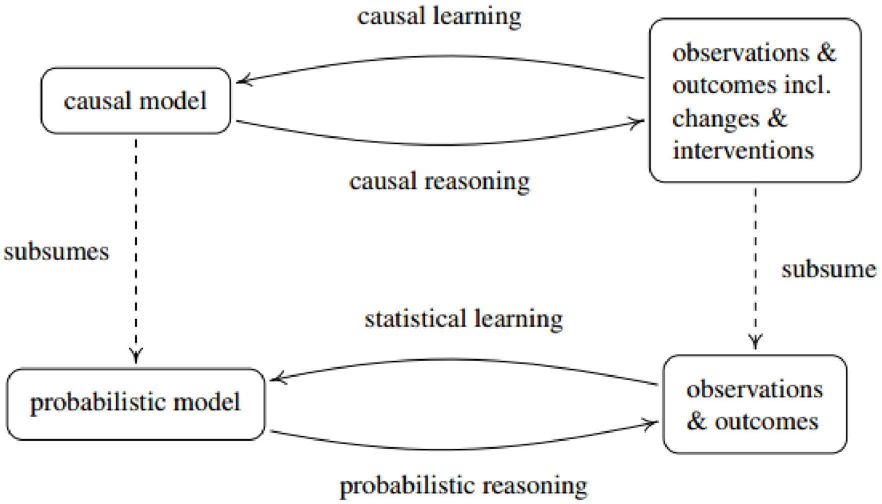

Figure 1.1: Terminology used by the present book for various **probabilistic inference** problems (bottom) and **causal inference** problems (top); see Section 1.3. Note that we use the term "inference" to include both learning and reasoning.

Figure 8: The probabilistic inference problems (bottom) and causal inference problems. [Peters et al., 2017]

4 Basic process in Causal Inference

From the Adjustment Formula (7), we see that there is a process for causal inference:

1. ***Causal Identification***. In this process, we move from a ***causal estimand*** to an equivalent ***statistical estimand*** under various causal assumptions (*unconfoundedness, positivity, and consistency*). That is from $\mathbb{E}\left[Y(1) - Y(0)|X\right] \rightarrow \mathbb{E}_X\left[\mathbb{E}\left[Y\,|T = 1,\, X\right]\right] - \mathbb{E}_X\left[\mathbb{E}\left[Y\,|T = 0,\, X\right]\right]$

2. ***Statistical Estimation***. In this process, we move from a *statistical estimand* to an **estimate**. That is to find an estimate of $\mathbb{E}\left[Y\,|T = t,\, X = x\right]$ using models or statistical methods.

Figure 8 shows the relationships between statistical modeling and causal modeling [Peters et al., 2017]. Causal modeling starts from another, arguably more fundamental, structure. A causal structure ***entails*** a probability model, but it contains additional information not contained in the latter.

- ***Causal reasoning***, according to the terminology used in [Peters et al., 2017], denotes the process of drawing conclusions from a causal model, similar to the way probability theory allows us to reason about the outcomes of random experiments.

- Just like statistical learning denotes the inverse problem to probability theory, ***causal learning*** studies how to infer causal structures from its empirical implications. The empirical implications can be purely observational, but they can also include data under interventions (e.g., randomized trials) or distribution changes. Researchers use various terms to refer to these problems, including ***structure learning*** and ***causal discovery***.
- ***Structure identifiability*** is mainly concerned about the question of which parts of the ***causal structure*** can in principle be inferred from the *joint distribution*.
- ***Causal learning*** is the *inverse problem* of causal reasoning. Unlike the standard problems of statistical learning, *even full knowledge of P does not make the solution trivial*, and we need **additional assumptions** (see above). The diffculties in causal learning not only include the ill-posed-ness of the usual statistical problems, but also arise from the fact that we are trying to estimate a richer structure than just a probabilistic one.
- It is less well known that one may postulate that while we cannot infer a concrete causal structure, we may at least **infer the existence** of causal links from statistical dependences.

Proposition 4.1 *(**Reichenbach's common cause principle**) [Peters et al., 2017]*
*If two random variables X and Y are **statistically dependent** ($X \not\perp\!\!\!\perp Y$), then there exists a third variable Z that **causally influences both**. (As a special case, Z may coincide with either X or Y.) Furthermore, this variable Z screens X and Y from each other in the sense that given Z, they become independent, $X \perp\!\!\!\perp Y|Z$.*

References

Madelyn Glymour, Judea Pearl, and Nicholas P Jewell. *Causal inference in statistics: A primer.* John Wiley & Sons, 2016.

Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press, 2015.

Brady Neal. Introduction to causal inference from a machine learning perspective. *Course Lecture Notes (draft)*, 2020.

Judea Pearl. *Causality: Models, Reasoning, and Inference.* Cambridge University Press, 2000.

Judea Pearl. Causal inference in statistics: An overview. *Statistics surveys*, 3:96–146, 2009.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms.* The MIT Press, 2017.

Paul Rosenbaum. *Observation and Experiment: An Introduction to Causal Inference.* Harvard University Press, 2017.