Active & Passive Causal Inference: Introduction

Anonymous authors
Paper under double-blind review

Abstract

This paper serves as a starting point for machine learning researchers, engineers and students who are interested in but not yet familiar with causal inference. We start by laying out an important set of assumptions that are collectively needed for causal identification, such as exchangeability, positivity, consistency and the absence of interference. From these assumptions, we build out a set of important causal inference techniques, which we do so by categorizing them into two buckets; active and passive approaches. We describe and discuss randomized controlled trials and bandit-based approaches from the active category. We then describe classical approaches, such as matching and inverse probability weighting, in the passive category, followed by more recent deep learning based algorithms. By finishing the paper with some of the missing aspects of causal inference from this paper, such as collider biases, we expect this paper to provide readers with a diverse set of starting points for further reading and research in causal inference and discovery.

1 Introduction

The human curiosity and the desire to understand how things work lead to studying causality (Kidd & Hayden, 2015; Zheng et al., 2020; Ferraro et al., 2019; Bender, 2020). Causality is about discovering the causal relationship between variables. Causality delineates an asymmetric relationship where it can only say *"A causes B"* or *"B causes A"*, while correlation expresses a symmetric relationship that measures the co-occurrence of A and B. They can extend to more than two variables. Being able to depict the causal relationship is an ideal framework for humans to explain how the system works.

An important concept in causality, that we are particularly interested in, is causal effect. It refers to the impact of a choice of action on an outcome. For example, what is the impact of receiving COVID-19 vaccine on catching COVID-19? In order to know the effect of Covid-19 vaccination, we must be able to predict the outcomes of both taking and not taking the vaccine, respectively. That is, we must know the potential outcome of taking an arbitrary action. We call this process of inferring the potential outcome given an action causal inference (CI).

Correlation does not imply causation, and at the same time, it is possible to have causation with no correlation. CI is tricky like that because observation is not the same as intervention. The former is about passively observing what happens, while the latter is about observing what happens when one actively intervenes in the world. Taking the CI approach allows us to distinguish between correlation and causation by clearly defining the two probabilities.

To infer causal effects we must measure intervention probabilities. Intervention probability is the probability of a particular outcome resulting from our intervention in the system by imposing a specific action. This is different from conditioning, as we actively alter the action. We often have access to conditional and joint probabilities, but not intervention probabilities directly. It has thus been a major research topic to infer the intervention probability from conditional and joint probabilities in many disciplines (Rubin, 1974; Berry & Fristedt, 1985; Heckman et al., 1997; Weitzen et al., 2004; Hernán & Robins, 2006; Breslow et al., 2009; Graepel et al., 2010; Abhijit V. & Esther, 2012; Yazdani & Boerwinkle, 2015; Bouneffouf et al., 2020).

In this paper, we assume the availability of an underlying causal graph, determined by domain or prior knowledge, and focus on inferring the causal effect of an action on the outcome under such a causal graph. We examine a variety of CI methods from a specific causal graph shown in Figure 2b condition, so that we can readily determine which CI approaches to take given a new causal graph in the future. More specifically, we consider a problem setup in which we have a covariate X, an action A, and an outcome Y. The action A is a variable of interest and has a causal effect on the outcome. The outcome Y is a variable that is affected by the treatment and is what we often want to maximize. Covariates X are all the other variables that may affect and be affected by the action A and the outcome Y. We are particularly interested in the case where these covariates are confounders, i.e., affect the action and the outcome together. We measure causal effects by computing average and individual treatment effects.

There is enormous work on various assumptions and conditions that allow us to infer causal effects (Rubin, 1974). The most fundamental assumptions are i) exchangeability, ii) positivity, iii) consistency, and iv) no interference. These assumptions must be satisfied at the time of data collection rather than at the time of causal inference. When these assumptions are met, we can then convert statistical quantities, estimated from collected data, into causal quantities, including the causal effect of the action on the outcome (Hernán et al., 2019; Musci & Stuart, 2019; Zheng & Kleinberg, 2019). One way to satisfy all of these assumptions is to collect data actively by observing the outcome after randomly assigning an action independent of the covariate. Such an approach, which is often referred to as a random controlled trial (RCT), is used in clinical trials, where patients are assigned randomly to an actual treatment or placebo (Chalmers et al., 1981; Kendall, 2003; P. M. et al., 2021). RCT is deliberately designed to prevent confounding the treatment and the outcome, so that the conditional probabilities estimated from collected data approximate the intervention probabilities as well.

Randomized data collection is not always feasible and often suffers in efficiency from running large-scale experiments. There has been an enormous amount of work from various disciplines on estimating causal effects without such randomized data collection (Rubin, 1977; 1979; Chalmers et al., 1981; Lu & Rosenbaum, 2004). As an alternative, different approaches have been proposed, including figuring out how to work with the non-randomized dataset and finding a more efficient way to collect data than the randomized approach. In this paper, we organize these CI methods into passive and active learning categories. In the passive CI category exist methods that work *given* a dataset which was *passively* collected by the experts. In contrast, active CI category includes methods that may actively intervene in the data collection process. RCT for instance belongs to the active CI category, as it actively collects data by randomizing the treatment assignment. There are however other methods in the same category that aims also to maximize the outcome by making a trade-off between exploration and exploitation.

The organization of this literature review paper is as follows. In §2, we introduce the definitions and metrics for estimating causal effects and discuss in depth the assumptions necessary for identification of causal effects. We then cover naive conditional mean estimator and ordinary square estimator, both of which are widely used with randomized datasets (Rubin, 1974; Pearl, 2010).

In §3, we describe RCT and move on to bandit approaches in the active CI category. While bandits are used in many practical applications, the research community has been adding more emphasis on theoretical analysis of minimizing the regret bounds of different policy algorithms (Berry & Fristedt, 1985; Langford & Zhang, 2007). We look at bandits through the lens of CI, where many of the bandit algorithms can be seen as learning the classic causal graph in Figure 1 with different exploration and exploitation rates. We examine different constrained contextual bandit problems that correspond to different causal graphs, respectively. We also compare passive CI learning methods to bandits on naive causal graphs. We furthermore review causal bandits which consider graphs with unknown confounding variables (Bareinboim et al., 2015; Lattimore et al., 2016; Sachidananda & Brunskill, 2017). In this survey, we limit our scope to bandits and do not consider causal reinforcement learning which we leave for the future.

In §4, we start with classical approaches in the passive CI category, such as matching (RUBIN & THOMAS, 1992; Gu & Rosenbaum, 1993), inverse probability weighting (ROSENBAUM & RUBIN, 1983; RUBIN & THOMAS, 1992; Hirano et al., 2003) and doubly robustness methods (Heejung & James M., 2005; Shardell et al., 2014; Seaman & Vansteelandt, 2018). We then discuss deep learning based CI methods (Zhihong

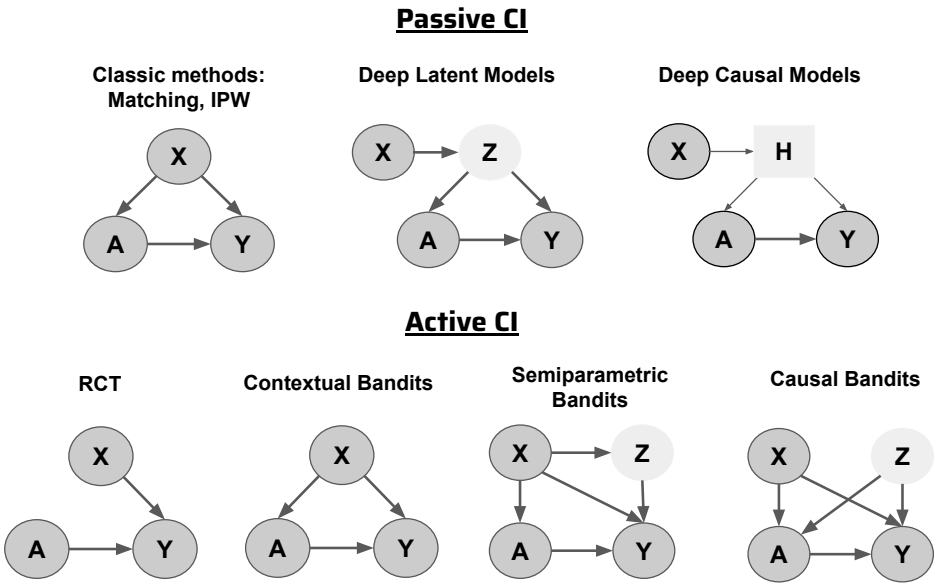
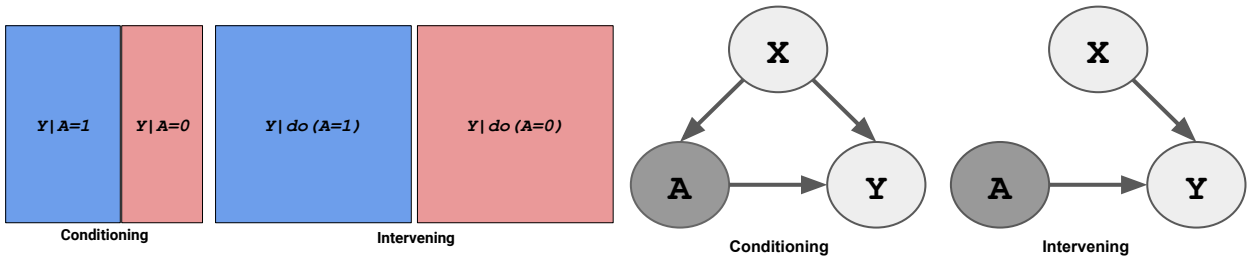


Figure 1: Examples of passive and active causal inference methods

& Manabu, 2012; Pearl, 2015; Johansson et al., 2016; Wang et al., 2016; Louizos et al., 2017a). Deep learning is particularly useful when we need to conduct causal inference on high dimensional data with a very complicated mapping from input to output, as deep neural networks can learn a compact representation of action as well as covariate that captures the intrinsic and semantic similarities underlying the data (Kingma & Welling, 2014; Danilo Jimenez & Shakir, 2014). Deep learning is applied to CI in order to infer causal effects by learning the hidden/unknown confounder representations from complicated data and causal graph relationships. Such a capability of learning a compact representation from a high-dimensional input allows it to work with challenging problems such as those involving raw medical images and complex treatments (Castro et al., 2020; jiwoong Im et al., 2021; Puli et al., 2022; van Amsterdam et al., 2022).

CI is an important topic in various disciplines, including statistics, epidemiology, economics, and social sciences, and is receiving an increasingly higher level of interest from machine learning and natural language processing due to the recent advances and interest in large-scale language models and more generally generative artificial intelligence. In this paper, we cover various CI algorithms and categorize them into active and passive CI families. The goal of this paper is to serve as a concise and readily-available resource for those who are just starting to grow their interest in causal inference.



(a) Population

(b) Causal Graph

Figure 2: Condition versus intervention

2 Background

2.1 Preliminary

Let X, A and Y be *covariate*, *action*, and *outcome* variables, respectively. We define the *joint*, *conditional* and *intervention* probabilities as

$$\text{Joint: } p(Y=y, A=a, X=x) = p(X=x)p(A=a|X=x)p(Y=y|A=a, X=x),$$
$$\text{Conditional: } p(Y=y|A=a) = \frac{\sum_x p(X=x)p(A=a|X=x)p(Y=y|A=a, X=x)}{\sum_{x',y'} p(X=x')p(A=a|X=x')p(Y=y'|A=a, X=x')}, \text{and}$$
$$\text{Intervention: } p(Y=y|do(A=a)) = \sum_x p(X=x)p(Y=y|A=a, X=x),$$

respectively. We marginalize the covariate variable to obtain the conditional and intervention probabilities. We observe that the conditional probability is often different from the intervention probability, $p(Y|A=a) \neq p(Y|do(A=a))$. The conditional probability takes into account the prevalence of a particular action a in the population and checks how often a particular outcome y is associated with it. On the other hand, the intervention probability does not consider the prevalence of the action a and only considers what the outcome would be had the action been forced to be a. See Figure 2a for a graphical illustration. This can be understood as removing all the incoming edges to the action variable, i.e. $X \to A$, when we intervene on A, as in Figure 2b. The intervention probability formula above reflects it by ignoring the probability of a particular action.

Consider the following example that shows how intervention and conditional probabilities are sometimes different and sometimes the same, given two variables.

Example 1. *let C indicate whether coffee is hot or cold and T be the thermometer reading that measures the temperature of the coffee. Based on our everyday observation, T and C are highly correlated. For instance, $p(C=hot|T=70°)$ and $p(T=70°|C=hot)$ are both high. However, it is clear that forcing the thermostat's reading to be high does not cause the temperature of coffee to go up, that is, $p(C=hot|do(T=70°))$ is low despite high $p(C=hot|T=70°)$. On the other hand, boiling coffee would indeed cause the thermostat's reading to go up, that is, both $p(T=70°|C=hot)$ and $p(T=70°|do(C=hot))$ are high.*

Another example, demonstrating the discrepancy between the intervention and conditional probabilities, is *Simpson's paradox* in which different assumptions about outcome, treatment, and confounder variables lead to different conclusions on causation:

Example 2 (Simpson's paradox illustration (Carlson, 2019))**.** *You are studying sex bias in graduate school admission. According to the data, men were more likely to be admitted to graduate school than women were, where 40% of male applicants and 25% of female applicants were admitted. In other words, there was a strong association between being a man and being admitted. You however found that such association varies across different sub-populations. For example, 80% and 46% of men and women were admitted to natural science respectively, and 20% and 4% of women and men were admitted to social science respectively. It turns out*

that the social science department has a much lower acceptance rate than the natural science department, while women were more likely to apply to social science and men were more likely to apply to natural science departments. In summary, we can derive different conclusions about sex and admission rate association by either combining or separating sub-populations.

We can notice how adding the department to the covariate variable or not changes the result. If the department is not part of the covariate, then the conditional and intervention probability coincide with each other, since there is no confounding via the department choice. Otherwise, these two probabilities deviate from each other due to $p(S = s|X = x)$. This demonstrates that CI is inherently dependent upon our modelling assumption.

A *potential outcome* is an outcome given a covariate under a potential action (Rubin, 1974; 2005). Assuming a binary outcome variable $Y \in \{0, 1\}$ and a discrete action variable $A \in \mathcal{A}$, the potential outcome is defined as

$$Y_X(a) = Y|X, do(A = a).$$

For instance, there are two potential outcomes, $Y_X(1)$ and $Y_X(0)$, for a binary action $A \in \{0, 1\}$. It is often impossible to compute these potential outcomes given a fixed covariate directly due to the fundamental problem of causal inference, that is, we cannot perform two distinct actions for a given covariate X simultaneously and observe their outcomes (Saito & Yasui, 2019). We can only observe the outcome of one action that has been performed, to which we refer as the *factual outcome*, but cannot observe that of the other action, to which we refer as the *counterfactual outcome*. We instead consider the potential outcome averaged over the entire population, represented by the covariate distribution $p(X)$. We call it the *expected potential outcome* $\mathbb{E}_{Y,X}[Y_X(A = a)]$, as opposed to the *conditional potential outcome*, where we marginalize out the covariate X.

The goal of causal inference is then to infer these potential outcomes, individual as well as expected, from a set of observed outcomes $(y_i|do(A = a_i), x_i, a_i) \overset{\text{iid}}{\sim} P$, where P is the data distribution and $a_i \in \mathcal{A}$ is the action performed for the i-th covariate x_i. This dataset does not, or perhaps more correctly cannot, contain any counterfactual potential outcome $(y_i|do(A = \neg a_i), x_i)$ for all i.

We often hear about the *treatment effect* of an experimental drug or a new surgical procedure in the medical literature. Treatment effect measures whether and how much the treatment caused the difference in the outcome (or in the potential outcome). Often treatment effect is used for binary actions in medical research. Unless confusing, we will interchangingly use the treatment effect and causal effect throughout this paper. In general, the treatment effect is defined as the difference between two potential outcomes $Y_X(1) - Y_X(0)$. The *average treatment effect* (ATE) is then the difference between the potential outcomes averaged over the covariate distribution (Rubin, 1974; Imbens, 2004):

$$ATE := \mathbb{E}_{X,Y}[Y_X(1) - Y_X(0)] = \mathbb{E}_{X,Y}[Y_X(1)] - \mathbb{E}_{X,Y}[Y_X(0)]. \tag{1}$$

We may be interested in the average treatment effect over a subpopulation, defined by a subset $X' \subseteq X$ of covariates. We then compute the *conditional average treatment effect* (CATE). CATE is defined as averaging the treatment effect for an individual patient characterized by X' (Radcliffe, 2007; Athey et al., 2015):

$$CATE(x') := \mathbb{E}_{Y,X\setminus X'}[Y_X(1) - Y_X(0)|X' = x'] = \mathbb{E}_{Y,X\setminus X'}[Y_X(1)|X' = x'] - \mathbb{E}_{Y,X\setminus X'}[Y_X(0)|X' = x'], \tag{2}$$

where $X\setminus X'$ is the remainder of the covariate over which we compute the expectation.

There are a few alternatives to the ATE. The first one is an *expected precision in the estimation of heterogeneous effect* (PEHE (Imbens, 2004)):

$$PEHE := \mathbb{E}_{X,Y}[(Y_X(1) - Y_X(0))^2].$$

Another alternative is the *population average treatment effect* (PATT) for the treated, i.e. $a = 1$ (Rubin, 1977; Heckman & Robb, 1985):

$$PATT(a) := \mathbb{E}_{X,Y|A=a}[Y_X(1) - Y_X(0)],$$

which is a helpful quantity when a particular, treated sub-population is more relevant in the context of narrowly targeted experiments. All of these metrics have their own places. For instance, it is more common to see PEHE in medical research, while PATT can be used to study the effect on the treated group programs (e.g. individuals disadvantaged in the labour market (Heckman & Robb, 1985)).

2.2 Assumptions for Causal Inference

Unfortunately, we cannot simply average the factual outcomes for each action to estimate the average treatment effect, because this corresponds to measuring the expected conditional outcome which is a biased estimate of the expected potential outcome. This bias is usually due to confounders. For instance, socioeconomic status can be a confounder in the study of the treatment effect of a medication. Socioeconomic status often affects both patients' access to medication and their general health, which makes it a confounder between the treatment and health outcome. In this case, the expected conditional outcome estimate is biased, because those patients who receive the medication are also likely to receive better healthcare, resulting in a better outcome. We must isolate the effect of medications on the health outcome by separating out the effect of having better access to healthcare due to patients' higher socioeconomic status, in order to properly estimate the expected treatment effect. In this section, we review the assumptions required for us to obtain an unbiased estimator for the average potential outcome.

The main strategy for estimating the (average) potential outcome is to compute causal quantities, such as intervention probabilities, from statistical quantities that are readily estimated from a set of samples, i.e., a dataset. In this context, we can say that a causal quantity is *identifiable* if we can compute it from statistical quantities. In doing so, there are a number of assumptions that must be satisfied.

Positivity/Overlap. The first step in estimating the potential outcome is to estimate the conditional probabilities from data. In particular, we need to compute

$$p(Y|X, A) = \frac{p(Y, A|X)}{p(A|X)}$$

for X with $p(X) > 0$. This implies that $p(A = a|X)$ for the action a, of which we are interested in computing the potential outcome, must be positive. We call it the *positivity*. Positivity is a necessary condition for computing ATE as we need the $p(A|X = x) > 0$ in the denominator for data x. The *overlap* assumption is similar to the positivity assumption but applies to the covariate. It requires that the distributions $p(X|A = 0)$ and $p(X|A = 1)$ have common support. Partial overlap occurs when you are missing on particular action for a certain area of covariate space. For example, we can have treated units of certain patient groups but no control units, or vice versa.

Ignorability/Exchangeability. Even if we can estimate the statistical quantities, such as the conditional probability $p(Y|X, A)$, we need an additional set of assumptions in order to turn them into causal quantities. The first such assumption is *exchangeabilty* which states that the potential outcome $\hat{Y}(a)$ must be preserved even if the choice of an action to each covariate configuration $p(A|X)$ changes. That is, the causal effect of A on Y does not depend on how we assign an action A to each sample X. This is also called *ignorability*, as this is equivalent to ignoring the associated covariate when assigning an action to a sample, i.e., $A \perp\!\!\!\perp X$. This enables us to turn the conditional probabilities into intervention probabilities, eventually allowing us to estimate the potential outcome, which we describe in more detail later.

The exchangeability is a strict condition that may not be easily met in practice. This can be due to the existence of confounding variables, selection bias, or due to time-dependency between action selections (violation of Markov Assumption). We can relax this by assuming *conditional exchangeability*. As we did for defining the CATE above, we partition the covariate into X and X' and condition the latter on a particular value, i.e., $X' = x'$. If the exchangeability is satisfied conditioned on $X' = x'$, we say that conditional exchangeability was satisfied. This however implies that we are only able to estimate the potential outcome given a particular configuration of X'. In order to estimate the potential outcome over all possible configurations of $[X, X']$, we must further assume *unconfoundedness*. It states that we must *observe* all configurations of X' for which $p(X, X') > 0$ and that there cannot be an unobserved covariate that affects

the choice of action for each sample. These two assumptions together allow us to turn the statistical quantities into causal quantities, just like when (unconditional) exchangeability alone was satisfied.

Consistency and Markov assumption. Data for causal inference is often collected in a series of batches rather than at once in parallel. Each t-th batch has an associated potential outcome $\hat{Y}(a_t)$ for action a. If the potential outcome changes over these batches, we cannot concatenate all these batches and use them as one dataset to estimate the potential outcome. That is, we must ensure that $\hat{Y}(a_t) = \hat{Y}(a_{t'})$ for all t, t' where $a_t = a_{t'}$. This condition is called *consistency*. Furthermore, we must ensure that the past batches do not affect the future batches in terms of the potential outcome, i.e., $\hat{Y}(a_1, \ldots, a_t) = \hat{Y}(a_t)$, as this would effectively increase the action space dramatically and make it impossible to satisfy positivity. We call this condition a *Markov* assumption.

Combining all the assumptions together, ATE quantity is computed as follows:

$$
\begin{aligned}
ATE &= \mathbb{E}_X[\mathbb{E}_{X'}[Y(1) - Y(0)|X']] \\
&= \mathbb{E}_X[\mathbb{E}_{X'}[Y(1)|X'] - \mathbb{E}_{X'}[Y(0)|X']] && \text{(Conditional exchangeability)} \\
&= \mathbb{E}_X[\mathbb{E}_{X'}[Y(1)|A = 1, X']] - \mathbb{E}_X[\mathbb{E}_{X'}[Y(0)|A = 0, X']] && \text{(Ignorability)} \\
&= \mathbb{E}_X[\mathbb{E}[Y|A = 1, X]] - \mathbb{E}_X[\mathbb{E}[Y|A = 0, X]] && \text{(Consistency)} \\
&= \mathbb{E}[Y|A = 1] - \mathbb{E}[Y|A = 0]].
\end{aligned}
$$

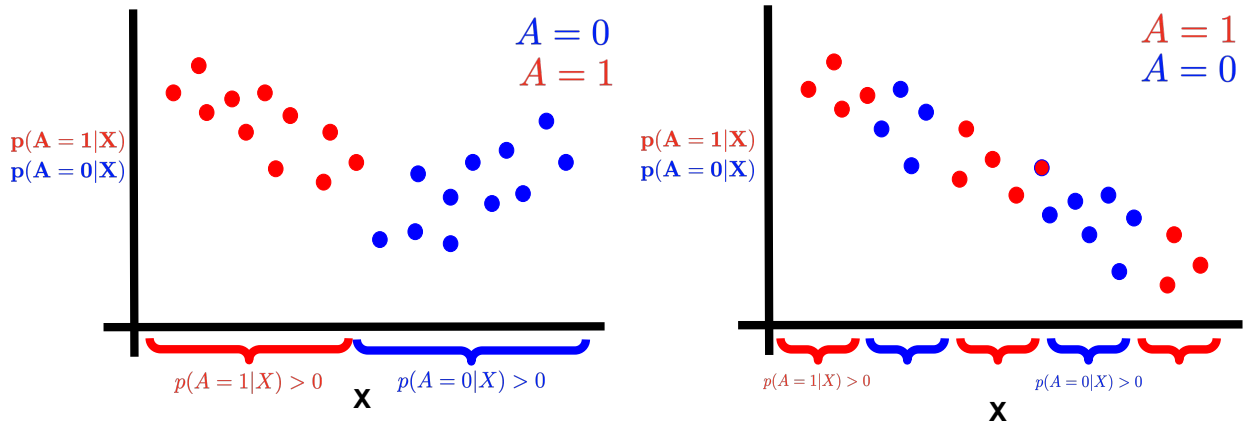
2.3 Discussion on the assumptions in practice

These assumptions above, or some of their combinations, enable us to derive causal quantities from statistical quantities. It is thus important to carefully consider these assumptions when faced with causal inference and how they may be violated, as most of these are often impossible to verify in practice. We discuss a few important points regarding these assumptions in practice.

Unconfoundness & Conditional Exchangeability. We face the challenge of verifying whether the potential outcome remains the same across all possible confounder configurations because we cannot enumerate all possible confounders. In practice, we use conditional exchangeability because we often condition each individual data point (e.g., covariates per patient). However, the conditional exchangeability assumption is impossible to test and verify in practice. Estimating the potential outcome given a particular configuration of X' means removing all the existing confounders for X', however, we cannot know every possible confounder out there. In practice, we condition $X' = x'$ to be an individual data point (e.g., x' being a patient) and often conditional exchangeability is taken for granted for that data point.

Unconfoundedness vs. Overlap. In order to estimate ATE, one must assume both unconfoundedness and positivity. It is however difficult to satisfy both of them in practice, because there is a natural trade-off between them. We are likely to satisfy the unconfoundedness by adding more features to the covariate, which in turn increases the dimensionality of data. This in turn increases the chance of violating the overlap assumption due to the curse of dimensionality. Similarly, we can satisfy the overlap assumption by choosing only the minimum number of features as a covariate, but we may unintentionally create unobserved confounders along the way.

Positivity via generalization. Positivity is hard to satisfy in a strict sense, as we must have as many data points as there are actions for each and every possible x with $p(x)$. This is largely impossible when x is continuous, and even when x is discrete, it is difficult if the support of $p(x)$, i.e., $\{x \in \mathcal{X} | p(x) > 0\}$ is large. Instead, we can fit a parametric or non-parametric model to model $p(A = a|X)$ that can generalize to an unseen combination of (X, A). In such a case, the success of generalization depends on the support of data points we have access to. Figure 3 presents a depiction of two cases where the fitted model must respectively extrapolate and interpolate. In the latter case of interpolation, even without positivity, we may be successful at correctly inferring the causal effect, but in the former case, it would be much more challenging. This suggests we must be careful at relying on generalization to overcome the issue of positivity (or lack thereof.)



(a) Generalization through extrapolation

(b) Generalization through interpolation

Figure 3: Generalization of the propensity score in two scenarios where the positivity assumption is violated. (a) requires extrapolation and (b) requires interpolation for generalizing to unseen counterfactual examples respectively.

Algorithm 1 Active CI protocol

A actions, T rounds (both known); potential outcome $Y(a)$ for each action a (unknown *a priori*).
while each round $t \in [T]$ **do**
  Observe a covariate x_t
  Pick an action according to policy, $a_t \sim \pi(x)$.
  Outcome observed $y_t \in [0, 1]$ is sampled given $X = x_t, A = a_t$.
  Update the potential outcome $\mathbb{E}_{p(X)}[Y_X(a)] \leftarrow t^{-1} \sum_{t'=1}^{t} \mathbb{I}(a_{t'}) y_t$
  [Optional] Update the policy π.
end while

3 Active Causal Inference Learning

Because counterfactual information is rarely available together with factual information in observed data (Graepel et al., 2010; Li et al., 2010; Chapelle et al., 2015), one approach to causal inference (CI) is to design an online algorithm that actively explores and acquires data. We introduce and describe an active CI framework that combines causal inference and data collection, inspired by works of literature on contextual bandit(Slivkins, 2019; Bouneffouf et al., 2020). In this framework, an algorithm estimates the expected potential outcome $Y_X(A)$ of each action and collects further data for underexplored counterfactual actions. A general active CI protocol is presented in Algorithm 1. We denote (x_t, a_t, y_t) for observed covariate, action, and outcome at time t, and denote $Y_X(A)$ for potential outcome. We update the estimated expected potential outcome $\mathbb{E}_X[Y_X(A)]$ based on a new observed data point (x_t, a_t, y_t). Although active CI algorithms are inspired by contextual bandit, there is a major difference between these two; that is, CI focuses much more on estimating the expected potential outcomes, while bandit algorithms on finding an optimal policy function that maximizes the expected outcome values.[1]

Most of the time CI researchers and practitioners face the challenge of being limited to data exploration not by our choice, but by other factors that make it inaccessible to A/B testing from real-world constraints. Active CI learning benefits from the policy function $\pi(x)$ when it can compromise between exploration and exploitation for understanding the causal effect and optimizing the decision in real-world applications respectively. In this section, we review active CI literature by first examining RCT and then expanding it to contextual bandit methods.

[1]We use *outcome* and *reward* interchangeably.

3.1 Randomized Controlled Trial

Randomized controlled trial (RCT) is the most well-known and widely practiced CI method. It is *de facto* standard in e.g. clinical trials, where participants are randomly divided into treatment and control groups, and the outcomes of these two groups are used to compute the average treatment effect (ATE). RCT falls under the active CI method group since it collects data proactively by randomly assigning an action to each data point (Rubin, 1974). Although it is typical to choose a uniform distribution over an action set $\pi = \mathcal{U}[\mathcal{A}]$ as a policy function in RCT, we can in principle pick any random distribution as long as it is independent of the covariate $\pi = \mathcal{P}[\mathcal{A}]$, as shown in Algorithm 2.[2]

Algorithm 2 Randomized Controlled Trial

Observes a covariate x_t.
Pick an action according to a policy, $a_t \sim \pi = \mathcal{P}[\mathcal{A}]$.
Observe the outcome $y_t \in [0, 1] \sim Y_X(A)|X = x_t, A = a_t$.
Estimate $\mathbb{E}_{p(X)}[\hat{Y}_X(a)]$ for all a.

A set of points $\{(x_t, a_t, y_t)\}_{t=0}^{T}$ collected by RCT automatically satisfy the exchangeability assumption. Because actions are selected independently of covariate X as shown in Figure 2b, the potential outcome $Y_X[A]$, estimated from this data, must be preserved even if action $P(A|X)$ changes. With these we can show that the conditional and intervention probabilities coincide with each other, allowing us to perform causal inference from data collected by RCT:

$$
\begin{aligned}
p(Y = y|do(A = a)) &= \sum_x p(Y = y|A = a, X = x)p(x) \\
&= \sum_x \frac{p(Y = y|A = a, X = x)p(A = a|X = x)p(x)}{p(A = a|X = x)} \\
&= \sum_x \frac{p(Y = y, A = a, X = x)}{p(A = a|X = x)} \\
&= \sum_x \frac{p(Y = y, A = a, X = x)}{p(A = a)} \text{ (since } A \perp\!\!\!\perp X) \\
&= \sum_x p(Y = y, X = x|A = a) \\
&= p(Y = y|A = a). \qquad (3)
\end{aligned}
$$

Let us consider Simpson's paradox (Ex. 2 in §2.3). Although there was an overall strong association between being a man and being more likely to be admitted, we found different associations when we considered different sub-populations due to the uneven sex distribution among the applicants and the acceptance rates, across departments. With RCT using the uniform action distribution, we would end up with an even number of each sex independent of the department choice, i.e., $p(A = a|S = \text{man}) = p(A = a|S = \text{woman})$. This would allow us to verify whether men are more likely to get admitted to graduate school without being confounded by the department's choice.

It is often tedious and difficult to plan and design an RCT due to many biases that are difficult to mitigate. One such example is a control bias which arises when the control group behaves differently or is treated differently from the treatment group. For instance, participants in the control group may be more likely to drop out of the study than those in the treatment group, due to the lack of progress they perceive themselves. In order to avoid a control bias, one must carefully consider eligibility criteria, selection of study groups, baseline differences in the available population, variability in indications in covariates, and also the management of intermediate outcomes (Simon, 2001; Jane-wit et al., 2010). There are techniques that help you assess the quality of RCT studies with an emphasis on measuring control bias in the planning

[2]Existing literature often conflates having an equal chance of sampling each action with independence, but this is not true.

and implementation of experiment designs (Chalmers et al., 1981; 1989; Stewart & Parmar, 1996; Moreira & Susser, 2002; Olivo et al., 2008).

RCT is widely used in clinical trials in medicine (Concato et al., 2000; Rothwell, 2005; Green & Glasgow, 2006; Frieden, 2017), policy research in public health (Sibbald & Roland, 1998; García & Wantchekon, 2010; Deaton & Cartwright, 2016; Choudhry, 2017), econometrics (Heckman & Robb, 1985; LaLonde, 1986; Abhijit V. & Esther, 2012) and advertisements in marketing (Graepel et al., 2010; Chapelle et al., 2015; Gordon et al., 2019). In these real-life applications, we often cannot afford to blindly assign actions but to determine the action based on the covariate X, due to ethical, legal and economical reasons. Because it is challenging to apply RCT in practice (Saturni et al., 2014), some studies explore combining RCT with observational study data (Rubin, 1974; Concato et al., 2000; Hannan, 2008). For example, one can use inverse probability weighting or matching techniques to de-bias the potential outcome estimation, as we will discuss in §4. We thus review active CI methods with a covariate-dependent data collection policy in the rest of this section.

3.2 Causal inference with contextual bandits

It is often impractical to use RCT in real-life settings due to (but not limited to) the following three limitations (Rubin, 1974; Olivo et al., 2008; Schafer & Kang, 2009; Saturni et al., 2014). First, a sample size must be large enough to detect a meaningful difference between the outcomes of actions, due to the high variance of RCT. Second, complete randomization or complete independence from the covariate is often neither feasible nor ethical in practice. Lastly, complete randomization often goes against the real-world objective of maximizing the outcome which is different from correctly inferring the causal effect. For example, doctors should not randomly assign different treatments to patients in order to test their causal effects on the patients, because this could end up harming many patients. Instead, a doctor makes the best decision for each patient given their expertise (i.e. their own policy), and on the fly adjusts their policy online in order to maximize the outcome of each patient:

$$\arg\max_{\pi_e} \mathbb{E}_{x\sim p(X)}\mathbb{E}_{a\sim\pi_{e(x)}}\left[Y_X(a)\right].$$

RCT on the other hand does not maximize the outcome at all, which makes it less desirable to use in many real-world scenarios.

In this section, we review different ways of performing both tasks together, that is, finding an optimal policy and estimating causal effects. In particular, we examine various ways to intervene and actively collect data under the framework of contextual bandits.[3] There are two primary approaches to applying contextual bandit algorithms to CI. The first approach actively gathers interventional data and then estimates ATE from this actively collected randomized dataset.[4] The second approach uses a contextual bandit method to learn a causal model utilizing all the collected data points including both randomized and non-randomized actions.

3.2.1 Estimating ATE from interventional data collected with a bandit method

The general idea is to keep track of data points for which intervention, i.e. randomization, happened. It is no different from RCT except that random intervention happens only occasionally based on your choice of bandit algorithm. For the purpose of illustration, we use the ϵ-greedy strategy together with the expert's policy as an example in Algorithm 3. The epsilon greedy algorithm is an easy way to add exploration to the basic greedy algorithm; we greedily choose an action based on the estimated highest outcome values but once in a while randomly select an action independently of the covariate with the probability ϵ. We use a *randomized set* to refer to a collection of these randomized actions together with associated outcomes and covariates. We then estimate the average treatment effect (ATE) directly using the randomized dataset only (see Appendix **??**). The efficiency in estimating the causal effect is determined solely by how often we randomize action, i.e., the probability ϵ. In contrast to RCT, as we select actions based on the covariate-aware

[3]See Appendix **??** for the basic description of contextual bandits.

[4]A randomized dataset refers to a dataset consisting of tuples collected using actions chosen independently of covariates.

Algorithm 3 ϵ-greedy protocol

A actions, T rounds (both known); potential outcome $Y(a)$ for each action a (unknown).
while In each round $t \in [T]$ **do**
  Toss a coin with the exploration probability ϵ_t.
  if explore **then**
    explore: choose an action $a_t \sim \mathcal{U}[a]$
  else
    Observe a covariate x_t
    Pick an action according to the expert $a_t \sim \pi_e(x)$.
  end if
  Observed the outcome $y_t \sim Y|X = x_t, A = a_t \in [0, 1]$.
  Store y_t in set $\mathcal{D}$ if explore
  Update the expected potential outcome $\mathbb{E}_{p(X)}[Y_X(a)] \leftarrow |\mathcal{D}|^{-1} \sum_d^{\mathcal{D}} \mathbb{I}[A = a] y_d$ for all a.
end while

policy occasionally with the probability $1 - \epsilon$, we fulfill both outcome maximization and ATE estimation, although the efficiency in ATE estimation is typically worse than that of RCT.

There are other, more sophisticated methods such as high-confidence elimination and upper-confidence bound (UCB) algorithms Auer et al. ("2002"); Slivkins (2019).[5] Unlike the ϵ-greedy strategy, these approaches choose when to randomize action based on a learned policy so far. A *high-confidence elimination method* alternates between two actions, a and a' until the confidence bounds of the two actions' potential outcomes do not overlap, where the confidence bounds are defined as

$$UBC_t(a) = \mathbb{E}[Y(a)] + r_t(a)$$
$$LBC_t(a) = \mathbb{E}[Y(a)] - r_t(a)$$

with the confidence radius $r_t(a) = \sqrt{2\log(T)/n_t(a)}$[6] and the number $n_t(a)$ of rounds with the action aSlivkins (2019).

Once the lower-confidence bound of the potential outcome of one action is greater than the upper-confidence bound of that of the other action, i.e., $UCB(a) < LCB(a')$, the former action a' is selected indefinitely from there on because the abandoned action cannot be the best action in terms of maximizing the outcome value. The rationale behind this method is to make sure to explore until we are confident about the expected potential outcome of each action. In short, the high-confidence elimination method fully explores until the best action is determined with a high level of confidence, after which it converts to exploiting the discovered best action only. The first phase of exploration is thus similar to performing RCT. This approach is interesting because as soon as the lower bound of a' and the upper bound of a separate, we automatically get some level of assurance about the potential outcome estimates.

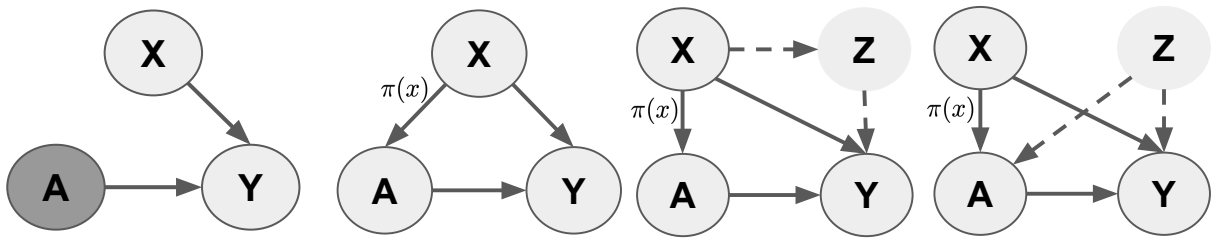
The *UCB algorithm* on the other hand picks an action that maximizes $UCB_t(a)$ at every round t. This choice automatically balances exploration and exploitation. a' is selected over another action a if $UCB_t(a') > UCB_t(a)$, which can happen for one of two reasons; the uncertainty is high (exploration) and the actual reward is high (exploitation). For the purpose of estimating ATE, we only use the randomized (explored) data points where the confidence radius $r_t(a)$ was large. There are other methods such as Thompson-sampling causal forest Dimakopoulou et al. (2017) and random-forest bandit Féraud et al. (2016), which share a similar flavour with the UCB algorithm.

3.2.2 Learning a causal graph using a bandit algorithm

So far in this section, we have discussed how to estimate ATE directly by gathering an interventional dataset using bandit methods. While such an approach allows us to measure the expected potential outcomes and

[5] See Appendix **??** for more details.

[6] There are ways to estimate a tighter radius based on different assumptions according to the previous research Slivkins (2019).



(a) Unconfounded graph (b) Contextual Bandit (c) SemiparametricBandit (d) Causal Bandit

Figure 4: Bandit methods for Active CI

ATE, we cannot infer conditional potential outcomes. One of the CI goals is to measure the causal effect using ATE but a bigger and more ambitious goal is to answer counterfactual questions. The latter is only possible if we can infer individual potential outcomes. Here, we discuss how to learn an underlying graphical model using a contextual bandit method in order to infer individual potential outcomes.

At each trial t, we approximate the potential outcome $\hat{Y}_X(a_t)$ using a parametrized model $g_\theta(x_t, a_t)$ that computes the outcome based on the context x_t and action a_t,

$$\hat{Y}_X(a_t) = g_\theta(x_t, a_t) + \epsilon_t, \tag{4}$$

where $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ is noise which comes from unknown confounders at time t. Although data was collected by the generic probabilistic graph that includes a policy π in Figure 4b, g_θ learns a causal graph in Figure 4a, because the parameters θ are estimated from the intervention dataset alone. This allows us to use g_θ to infer the potential outcome of a counterfactual action and/or of unseen covariates.

Contextual bandits can handle causal graphs that are more complicated. For instance, in Fig. 4c, there is an extra variable $Z = f(X)$ that mediates the effect of the covariate X on the outcome Y but does not affect the action A. In this case, the outcome function is in the form of

$$\hat{Y}_X(a_t) = g_\theta(x_t, a_t) + f_\phi(x_t) + \epsilon_t,$$

where $g_\theta(x_t, a_t)$ takes into account the confounder X as well as policy function $\pi(X)$ and maps them to an output Y. $f_\phi(X)$ learns to isolate the effect of the covariates that affect the outcome independent of the action. ϵ_t is the noise. Having $f_\phi(X)$ to learn covariates that are not confounders can reduce the complexity of learning for $g_\theta(x_t, a_t)$.

When both g_θ and f_ϕ are non-parametric, it is often computationally intractable to tackle this problem. In order to avoid these issues of computational tractability and undesirable regret bound, *semiparametric contextual bandits* consider parametric policies (Krishnamurthy et al., 2018; Greenewald et al., 2017; Peng et al., 2019) and learn a linear bandit model using regression oracles to estimate potential outcome (Swaminathan et al., 2017)

$$\hat{Y}_X(a_t) = w^\top x_t + f_\phi(x_t) + \epsilon_t,$$

where $w^\top$ is a parameter vector and ϵ_t is noise. They have a nice property where the regret bound is the same as the regular contextual bandit's regret. The action-independent features $f_\phi(x_t)$ get cancelled out in the regret formulation together with the noise term ϵ_t, because they are independent of the action choice.

3.2.3 Correcting a bias from unobserved confounders

It is unrealistic to observe all confounders in practice, nor for us to verify whether all confounders have been observed. Any confounding factor not included in the covariate leads to difficulties in accurately estimating the causal effect. Here, we consider a way to detect and correct such a bias that arises from having unobserved confounders (see Figure 4d).

Consider a policy function $\pi^*(x')$ that is optimized for maximizing $\arg\max_a CATE(x', a)$:[7]

$$\pi^*(x') = \arg\max_a CATE(x', a) = \arg\max_a \mathbb{E}_{Y,X\setminus X'}\left[Y_X(a) - \max_{\bar{a}} Y_X(\bar{a}) | X' = x'\right].$$

This $\pi^*(x')$ is estimated without considering unobserved confounders like in Figure 4b, and yet, with the realization of unobserved confounders shown in Figure 4d, we now correct the estimation. Let $\bar{a}$ be the counteraction to $\pi^*(x')$ that returns a larger CATE than the originally selected action $\pi^*(x')$. Suppose we found x' where $CATE(x', \pi^*(x')) < CATE(x', \bar{a})$ where $\bar{a} \neq \pi^*(x')$. Since the policy function π^* is optimal, this can only happen if there were unobserved confounders. This illustrates that our learned policy might not be the best in such a situation. We then need to search for a new policy function π' that can handle this situation better.

Because our counteraction could not be better than the action $\pi^*(x')$ without unobserved confounders, we compare the potential outcomes of the action and the counteraction $\bar{a}$;

$$\mathbb{E}_{Y,X\setminus X'}\left[Y_X(\pi^*(x')) | X' = x'\right] \text{ and} \tag{5}$$

$$\mathbb{E}_{Y,X\setminus X'}\left[Y_X(\bar{a}) | A = \pi^*(x') | X' = x'\right], \tag{6}$$

respectively. In order to compute the outcome of the counteraction, we actively intervene with the counteraction and collect a new sample $(x', y', \bar{a}, a)$. We also count the frequency of the expected potential outcome of counteraction being higher than of action $\pi^*(x')$,

$$f(\pi^*) = \sum_i^N \mathbb{I}\left[\mathbb{E}_{Y,X\setminus X'}\left[Y_X(\pi^*(x'_i)) | X' = x'_i\right] > \mathbb{E}_{Y,X\setminus X'}\left[Y_X(\bar{a}) | A = \pi^*(x'_i), X' = x'_i\right]\right],$$

where $\frac{f(\pi^*)}{N}$ tells us how often π^* is wrong. Our new policy function $\pi'(x')$ samples a new action according to the Bernoulli distribution with probability $\frac{f(\pi^*)}{N}$. This strategy copes with the fact that our policy is biased and incorrect with the probability $\frac{f(\pi^*)}{N}$ due to unobserved confounders.

Bareinboim et al. (2015) show that while this method requires collecting enough data for f to converge, we can optimize a new policy function faster by weighting the samples from the Beta distribution $\mathcal{B}(f, (1-f))$ by the bias as shown in Algorithm 4. The bias from the unobserved confounder is defined as one minus the absolute treatment effect (Bareinboim et al., 2015),

$$\text{bias} = 1 - |\mathbb{E}_{Y,X\setminus X'}[Y_X(\bar{a}) | A = \pi^*(x'), X' = x'] - \mathbb{E}_{Y,X\setminus X'}[Y_X(\pi^*(x')) | X' = x']|. \tag{7}$$

Eq. 7 quantifies how effective applying action $\pi^*(x')$ over the counteraction $\bar{a}$ is. If this quantity is large, the re-weight of the probability of choosing action $\pi^*(x')$ by the bias remains high. Otherwise, the re-weight of the probability of choosing counteraction $\bar{a}$ by bias remains high. Overall this encourages faster convergence of re-estimating the potential outcome (see the last three lines of Algorithm 4). This specific algorithm is called causal Thompson sampling.

[7]Unlike $CATE(x)$ with a binary action in Equation 2, we define $CATE(x, a)$ for a set of more than two actions.

Algorithm 4 Causal Thompson Sampling (TS^{C}) (Bareinboim et al., 2015)

Let $a = \pi^*(x')$ be intuitive action.
Let $\bar{a}$ be counteraction to a.
Let $\mathbb{E}_{Y,X\setminus X'}[Y_X A = a | A = a, X' = x']$ be the expected payout for intuitive action
Let $\mathbb{E}_{Y,X\setminus X'}[Y_X A = \bar{a} | A = a, X' = x']$ be the expected payout for counter-intuitive action

while $t = 1 \cdots T$ **do**
  Let $w = [1, 1]$
  Sample $a \sim \text{intuition}(x_t, t)$

  // Estimate the potential outcomes and bias
  Compute $bias = 1 - |\mathbb{E}[Y_X(\bar{a})|A = a, X' = x'] - \mathbb{E}[Y_X(a)|X' = x']|$ (Equation 7)
  if $\mathbb{E}[Y_X(\bar{a})|A = a, X' = x'] > \mathbb{E}[Y_X(a)|X' = x']$ **then**
    Set $w[a] = bias$
  else
    Set $w[\bar{a}] = bias$
  end if

  // Choose a new policy $\pi'(x')$ based on new weighting
  $\beta_1 = \mathcal{B}(f(a), (1 - f(a)))$
  $\beta_2 = \mathcal{B}(f(\bar{a}), (1 - f)(\bar{a}))$
  Set $\dot{\pi}(x') \leftarrow \max(\beta_1 \cdot w[a], \beta_2 \cdot w[\bar{a}])$
  Set $\dot{y} = \text{simulate}(\dot{\pi}(x'))$
  Update $\mathbb{E}[Y_X(\dot{\pi}(x'))|A = a, X' = x']$ with $(x', \dot{y}, \dot{\pi}(x'), a)$
end while

4 Passive Causal Inference

Unlike in active causal inference (CI), in passive CI, we must calculate ATE given a non-randomized dataset which was gathered in advance with the actions chosen based on covariates (Rubin, 1974; Holland, 1986). As discussed earlier in §2.3, we assume that the dataset satisfies positivity, although we discuss how to relax it with deep learning later in this section. In this section, we present and examine passive approaches to CI, which are gaining more popularity. Most of them are primarily grounded in one of the following three general approaches: matching, inverse probability weighting and doubly-robust methods. We first review these basic approaches and introduce some of the more recent approaches.

4.1 A Naive Estimator

Before we begin, we start with a naive estimator $\mu_A(X)$ of the outcome variable Y. Let $\mu_A : \mathcal{X} \to \mathcal{Y}$ be the generic estimator that predicts the outcome. For example, this estimator can be simple empirical averaging:

$$\mu_a(x) = \frac{\sum_{i=1}^{N} y_i \mathbb{I}[a_i = a, x_i = x]}{\sum_{i'=1}^{N} \mathbb{I}[x_{i'} = x]},$$

where $D = \{(x_i, a_i, y_i)\}_{i=1}^{N}$ is a dataset that consists of N data points, and $\mathbb{I}$ is an indicator function. This estimator looks at the average outcome of the action a given a particular covariate x. Another example would be to have a parametrized estimator, $\mu_A(X; \theta)$ where θ is a parameter.

Such a naive estimator $\mu_A(X)$, which often maximizes the log-likelihood, is a biased estimator of the potential outcome Y, due to the discrepancy between the conditional and invention probabilities. In the subsequent sections, we introduce and discuss modifications either to the estimator or to the dataset, that allows us to obtain an unbiased estimate of the potential outcome.

4.2 Matching

ATE estimation is hard when working with a real-life dataset due to confounding. A common approach is to construct a randomized dataset, that is free of confounding, from a non-randomized one. The matching method achieves this by pairing each treated instance with a control instance and ignoring any unmatched instance. This process balances the treatment and controlled group (Scotina & Gutman, 2019).

Ideally, we should have both factual and counterfactual outcomes for every data point ($\langle(x_i, y_i(1), y_i(0)\rangle$ for all i). This is often impossible due to the fundamental problem of CI, where the problem arises from the very fact that we can only observe the outcome of a single action given a particular covariate. We instead aim for approximate one-to-one matching where we pair treated data with controlled data that are similar enough. That is, we pair two instances if $\{\langle(x_i, 1, y_i(1)), (x_j, 0, y_j(0))\rangle | D(x_i, x_j) < \epsilon \text{ for } i \neq j\}$ and for some small ϵ, where $D(\cdot, \cdot)$ is a problem-specific distance metric. We then compute ATE using our new balanced dataset. The advantages and disadvantages of such a matching method lie in the bias-variance tradeoff. The advantage of the matching method is that it reduces the confounding bias, but it increases the variance because we removed (potentially many) unmatched data points.

There are many standard metrics of *closeness* that are widely used. One such metric is Mahalanobis distance:

$$D_{ij} = (x_i - x_j)^\top \Sigma^{-1} (x_i - x_j),$$

where Σ is the covariance metric. Many alternative metrics have been proposed over the past decades, such as those relying on a coarsened data space or other feature spaces (Cochran & Rubin, 1973; Rubin, 1979; ROSENBAUM & RUBIN, 1983; RUBIN & THOMAS, 1992; Rubin & Stuart, 2006; Stuart, 2010; Iacus et al., 2012; Zubizarreta, 2012; Zhao, 2004; Resa & Zubizarreta, 2016). The choice of a metric must be determined for each problem separately.

Matching methods have evolved from a greedy algorithm based (heuristic-based search) to optimal/full matching (Kim & Steiner, 2016). The greedy method ends up with a sub-optimal solution since the ordering of pairing matters (Weitzen et al., 2004). Moving away from greedy matching, one can use an optimal non-bipartite matching algorithm that runs in polynomial time (Lu & ROSENBAUM, 2004; Dehejia & Wahba, 1999). Such an algorithm generates a series of matched sets that consists of at least one treated individual and at least one control individual. It is optimal in terms of minimizing the average distance between each treated individual and each controlled individual within each matched set (Hansen, 2004). There are other approaches such as weighting methods where one adjusts the importance of distance between data points (Heckman et al., 1997; Hirano et al., 2003; Imbens, 2004). These methods can be helpful when the data samples are unevenly spread out over the domain.

There however remain challenges with matching. First, it is difficult to have exact matches for all data points with a non-randomized or imbalanced dataset (the dataset is unevenly distributed w.r.t actions). We thus end up eliminating a significant number of unpaired data points after matching between treated and controlled groups, resulting in the loss of information and statistical power. There are however some algorithms that enable many-to-one matching, with the simplest being the K-nearest neighbour method (Karp, 1972; Schafer & Kang, 2009; Zubizarreta, 2012). Second, matching works well when the dimensionality of the covariate is low, but it easily fails when the dimensionality is high due to the curse of dimensionality (Gu & Rosenbaum, 1993). It can be helpful to use deep learning for obtaining a more concise and dense representation, similar to what we will discuss in §4.5.

4.3 Inverse Probability Weighting

Inverse probability weighting (IPW) removes the effect of confounders by weighting the potential outcome of each action by its inverse probability weight (ROSENBAUM & RUBIN, 1983; Robins et al., 1994; Hirano et al., 2003):

$$\mathbb{E}\left[\frac{Y\mathbb{I}[A=a]}{p(A|X)}\right] = \mathbb{E}_{p(X)}\left[\mathbb{E}\left[\frac{Y(a)\mathbb{I}[A=a]}{p(A|X)}\middle|X\right]\right] = \mathbb{E}\left[\frac{\mathbb{E}\left[Y(a)\middle|X\right]\mathbb{E}\left[\mathbb{I}[A=a|X]\right]}{p(A|X)}\right] = \mathbb{E}\left[Y(a)\right]. \quad (8)$$

Equation 8 illustrates that we can take the subset of data that corresponds to a particular action a as long as we can divide by the so-called propensity score in order to compute $\mathbb{E}[Y(A)]$. $p(A = a|X = x)$ is known as the *propensity score* $e(x)$.[8]

The propensity score theorem (ROSENBAUM & RUBIN, 1983) tells us that if ignoreability is satisfied given X, then ignoreability conditioned on $e(X)$ is also satisfied (Imbens & Rubin, 2015):

$$(Y(1), Y(0)) \perp\!\!\!\perp A|X \implies (Y(1), Y(0)) \perp\!\!\!\perp A|e(X).$$

This predicate tells us that the potential outcomes are independent of A given confounder X, and this is due to the blocking back-door criterion conditioning on confounder X (Pearl, 2009). Similarly, the potential outcomes are independent of A given the propensity score $e(x)$, and the propensity score has the same effect as removing the blocking back-door path in a causal graph by conditioning on the edge between X and A. This illustrates that the 1-dimensional score function, that is the propensity score, is enough to compress the high-dimensional confounder X.

While simple finite-sample averaging $\widehat{ATE}_{AGG}$ is a biased estimator of ATE,

$$\widehat{ATE}_{AGG} = \frac{1}{N}\sum_{i}^{n}\left[\frac{\mathbb{I}[A_i = 1]y_i(1)}{\hat{e}(x_i)}\right] - \frac{1}{N}\sum_{i}^{n}\left[\frac{\mathbb{I}[A_i = 0]y_i(0)}{1 - \hat{e}(x_i)}\right],$$

$\widehat{ATE}_{IPW}$ is an unbiased estimator of ATE,

$$\widehat{ATE}_{IPW} = \frac{1}{N}\sum_{i}^{n}\left[\frac{\mathbb{I}[A_i = 1]y_i(1)}{e(x_i)}\right] - \frac{1}{N}\sum_{i}^{n}\left[\frac{\mathbb{I}[A_i = 0]y_i(0)}{1 - e(x_i)}\right],$$

where $\hat{e}(x) = \frac{N_{x1}}{N}$ and N_{x1} is the number of treated samples. Note that $\widehat{ATE}_{AGG}$ uses a naive estimator from §4.1 to estimate ATE. $\widehat{ATE}_{IPW}$ is known as *an oracle IPW estimator* since the propensity score $e(\cdot)$ is known. The estimation error, defined as $\sqrt{N}|ATE - \widehat{ATE}_{IPW}|$, follows a Normal distribution with zero mean and the variance of

$$\mathrm{VAR}_{\mathrm{IPW}} = \mathrm{Var}[ATE_{\mathrm{AGG}}(X)] + \mathbb{E}_p(x)\left[\frac{\sigma^2(X)}{e(X)(1 - e(X))} + \frac{c(X)^2}{e(X)(1 - e(X))}\right], \tag{9}$$

where $c(X)$ is a function that satisfies

$$\begin{aligned} Y(0) &= c(X) - (1 - e(X))ATE(X) + \epsilon(X) \\ Y(1) &= c(X) + e(X)ATE(X) + \epsilon(X), \end{aligned}$$

where $\epsilon(X)$ is a Gaussian noise with variance of $\sigma^2(x)$. $\widehat{ATE}_{AGG}$ can be seen as a version of the IPW estimator with an imperfect propensity score being $\hat{e}(x) = \frac{N_{x1}}{N}$. $\mathrm{VAR}_{\mathrm{IPW}}$ demonstrates that even with using the true propensity score $e(x)$, $\widehat{ATE}_{IPW}$ has a worse asymptotic variance than $\widehat{ATE}_{AGG}$ which can be thought of as inverse probability reweighting with an incorrect propensity score $\hat{e}(x)$. This is an example of the bias-variance trade-off.

A plethora of methods have been proposed since then, that are unbiased and exhibit lower variance than the oracle IPW above by replacing the propensity score with other weighting schemes $\hat{e}'$. For example, one can

[8] Additionally, propensity scores can be used for matching methods where one compares two covariates with similar propensity values $D(\hat{e}(x_i), \hat{e}(e_j))$ for $i \neq j$ and some distance metric

$$D_{ij} = |e(x_i) - e(x_j)|,$$

where $e(x_i)$ and $e(x_j)$ are the propensity scores for the data point x_i and x_j, respectively. The propensity score is a popular method as it summarizes the entire covariate into a single scalar and has been shown to be effective in theory (ROSENBAUM & RUBIN, 1983; RUBIN & THOMAS, 1992; Rubin & Stuart, 2006; Zubizarreta, 2012; Diamond & Sekhon, 2013; Resa & Zubizarreta, 2016; Abadie & Imbens, 2016) and in practice (Jalan & Ravallion, 2001; Dehejia & Wahba, 2002; Monahan et al., 2011; Amusa, 2018).

reduce the variance of an IPW estimator by normalizing the weights (Hirano et al., 2003):

$$\widehat{ATE}_{SW} = \frac{\sum_i^n \mathbb{I}[A_i = 1] y_i(1) w(x_i)}{\sum_i^n \mathbb{I}[A_i = 1] w_1(x_i)} - \frac{\sum_i^n \mathbb{I}[A_i = 0] y_i(0) w_0(x_i)}{\sum_i^n \mathbb{I}[A_i = 0] w_0(x_i)},$$

where $w_1 = \frac{1}{e(x_i)}$ and $w_0 = \frac{1}{1-e(x_i)}$. This leads to a lower variance in the estimate, and these weights are thus called *stabilized weights* (Robins et al., 2000). According to Hirano et al. (2003), $\widehat{ATE}_{SW}$ outperforms $\widehat{ATE}_{IPW}$ in terms of asymptotic convergence rate (Hirano et al., 2003). Lunceford & Davidian (2004) review various versions of the IPW estimator and suggest a way to take into account the uncertainty in estimating the propensity score using a closed-form sandwich estimator (M-estimator (Stefanski & Boos, 2002)) (Lunceford & Davidian, 2004).

4.4 Doubly Robust Methods

It is often hard to obtain the propensity score in advance nor guarantee that our propensity estimate $\hat{e}(x)$ is accurate. Furthermore, even with the oracle propensity score, we have just shown that the IPW has a high variance. On the other hand, the naive estimator (without IPW) from §4.1 is unbiased only when the actions in the dataset were sampled independently of the associated covariates. In other words, it is often not enough to rely on either of these approaches on their own to perform causal inference (Belloni et al., 2011). We can do better by combining IPW with the naive estimator $\mu_A(X)$, to which we refer as a doubly robust estimator. In doubly robust estimation, a bias from one method is addressed by the other method, and vice versa (Robins et al., 1994; Lunceford & Davidian, 2004; Kang & Schafer, 2007; Chernozhukov et al., 2017). Such a method is both consistent and unbiased, as long as at least one of the IPW and the naive estimator is consistent and unbiased. The doubly robust estimator for ATE in the case of two actions is then

$$ATE_{DR} = \mathbb{E}_{Y,X}\left[\mu_1(X) - \mu_0(X)\right] + \mathbb{E}_{Y,X,A}\left[A\frac{(Y - \mu_1(X))}{e(X)} - (1-A)\frac{(Y - \mu_0(X))}{1 - e(X)}\right] \tag{10}$$

$$= \mathbb{E}_{Y,X,A}\left[A\frac{(Y - \mu_1(X))}{e(X)} + \mu_1(X)\right] - \mathbb{E}_{Y,X,A}\left[(1-A)\frac{(Y - \mu_0(X))}{1 - e(X)} + \mu_0(X)\right]. \tag{11}$$

If the estimated potential outcomes $\mu_A(X)$ are correct, we do not need to worry about propensity score estimation, since $Y - \hat{\mu}_1(X)$ and $Y - \hat{\mu}_0(X)$ will be zero in Equation 11. $\widehat{ATE}_{DR}$ hence reduces to approximating $\mathbb{E}\left[\hat{\mu}_1(X) - \hat{\mu}_0(X)\right]$. In contrast, it is okay for our estimated potential outcome to be wrong if propensity score estimation is consistent. If the propensity score is correctly estimated, then $\mathbb{E}_{Y,X,A}\left[\frac{AY}{\hat{e}(X)}\right]$ and $\mathbb{E}_{Y,X,A}\left[\frac{(1-A)Y}{(1-\hat{e}(X))}\right]$ will be weighted correctly as well, and we recover the IPW estimator (Robins et al., 1994; Robins & Rotnitzky, 1995; SCHARFSTEIN et al., 1999). Consequently, such a doubly robust method is consistent (Hahn, 1998; Heejung & James M., 2005; Shardell et al., 2014; Farrell, 2015).

Re-arranging terms and expressing in terms of Monte Carlo estimation of ATE_{DR}, we get

$$\widehat{ATE}_{DR} \approx \frac{1}{N}\sum_i^N \left[\frac{a_i y_i}{\hat{e}(x_i)} - \frac{a_i - \hat{e}(x_i)}{\hat{e}(x_i)}\hat{\mu}_1(x_i)\right] - \frac{1}{N}\sum_i^N \left[\frac{(1-a_i)y_i}{1-\hat{e}(x_i)} - \frac{\hat{e}(x_i) - a_i}{1-\hat{e}(x_i)}\hat{\mu}_0(x_i)\right]. \tag{12}$$

The empirical estimate converges the true ATE, $\hat{ATE}_{DR} \to ATE$, with an asymptotic variance of

$$\mathrm{Var}[ATE_{DR}(X)] = \mathrm{Var}[ATE_{AGG}(X)] + \mathbb{E}\left[\frac{\sigma_1^2(X)}{e(X)}\right] + \mathbb{E}\left[\frac{\sigma_0^2(X)}{1 - e(X)}\right],$$

where $\sigma_T(X) = \mathrm{Var}[Y_i(T)|X]$. Despite its greater variance, the doubly robust method often exhibits greater efficiency and robustness to model misspecification.

Multiple studies have shown that doubly robust methods for ATE estimation with missing data perform better than their non-doubly robust baselines and theoretically have shown to converge faster to the true ATE than individual methods (Robins et al., 1994; Mayer et al., 2020). A naive doubly-robust estimator

is asymptotically optimal among non-parametrized estimators, meaning the semiparametric variances are bounded and asymptotically convergent if either the propensity score or the estimator $\mu_A(X)$ are correct (Robins & Rotnitzky, 1995; Kang & Schafer, 2007). Subsequently, other estimators with bounded asymptotic variances have been proposed even when IPW exhibits a high variance (Robins et al., 2007; Tan, 2010; Waernbaum & Pazzagli, 2017).

4.5 Causal inference with representation learning

While matching, IPW, and doubly robust methods have their own merits for estimating potential outcomes, it may be necessary to utilize a powerful model that can express highly complex functions given high dimensional data with a limited amount of data. For instance, working with non-linear high-dimensional data such as X-ray images may require a non-linear parametric model to transform data from its original space to a better representation that facilitates CI. The goal is to extract a causal representation from high dimensional data (covariate and action) and more accurately predict potential outcomes from the extracted causal representation. In this section, we review methods that extend the previous CI approaches in this way by (deep) representation learning (Schölkopf et al., 2021; Wang & Jordan, 2021).

Representation learning involves automatically learning features or representations of raw data. Rather than manual feature engineering, representation learning enables models to learn to extract high-level features from raw, high-dimensional data such as images, audio, and texts, themselves. Here, the potential outcome estimator $u_A(X;\phi)$ is a deep neural network with parameter ϕ with m hidden layers $H = \{\mathbf{h}^{(i)}\}_1^m$. Each hidden representation at i-th layer ($i < k$) is a function of all the previous layers, $\mathbf{h}_i(x) = f_i(\mathbf{h}_{i-1};\phi)$ with $\mathbf{h}_0 = x$. Each hidden representation at $j \geq k$, $\mathbf{h}_j(x) = f_j(\mathbf{h}_{j-1}, a;\phi)$, depends also on the action.

The model is trained to minimize the factual loss l_{Factual} which is typically the negative log-likelihood of the dataset with inverse probability weighting. $u_A(X;\phi)$ learns to predict the potential outcome however only well on observed covariate-action pairs due to the challenges in generalization. We thus need to add a regularization term to the loss function in order to encourage the model to generalize better to (unseen) counterfactual actions:

$$\mathcal{L} = \mathbb{E}_{Y,X,A}[wl_{\text{Factual}}(X, Y, A; u_A(X;\phi)) + \lambda\mathcal{R}(H)]. \tag{13}$$

The loss function is weighted by the inverse probability weight $w = \frac{A}{2e(X)} + \frac{1-A}{2(1-e(X))}$ (see §4.3). Regularization often imposes certain properties on the hidden layers, which is why we refer to it as $\mathcal{R}(H)$, with the regularization coefficient λ (Johansson et al., 2016; Uri et al., 2017; Yao et al., 2020; Wu & Fukumizu, 2021). Regularization reduces the hypothesis space in a way that encourages $\mu_A(X,\phi)$ to capture a causal relationship rather than a spurious relationship between the action and outcome. Deep representation learning based CI is vast and fast-growing (Nabi & Shpitser, 2017; Yoon et al., 2018; Veitch et al., 2020; Zhang et al., 2020; Wang & Jordan, 2021; Zhang et al., 2021). We characterize a majority of these approaches as minimizing a combination of a factual regression loss and a regularizer, as in Equation 13, and we discuss four representative ones in this section.

Among these methods, we separately discuss counterfactual smoothing regularization and deep latent variable models (see Figure 5) with a focus on how both of these approaches lead to better generalization of ATE.

4.5.1 Counterfactual Smoothing Regularization

The objective in counterfactual smoothing is to train a model to generalize to a counterfactual potential outcome even if it only saw factual data during training (see Figure 5a). Although the model $u_A(X,\phi)$ is trained to estimate the potential outcome using the inverse probability weighted factual loss function, it may still underperform for an unseeded data or a data paired with unseen action (see Figure 3). In such a case, the variance of the potential outcome estimate over counterfactual actions is high, implying that an individual model's prediction may be inaccurate.

Counterfactual smoothing reduces the variance of the predicted potential outcome whose variance can be controlled by ensuring that the learned representations of treated and controlled data points to follow (largely) indistinguishable distributions in the feature space, respectively (Johansson et al., 2016; Shalit et al., 2017;

(a) (b)

Figure 5: Causal Graph - (a) Deterministic Representation: H corresponds to a hidden layer representation that is a deterministic variable, X is a covariate, A is an action, and Y is an outcome variable. (b) Stochastic Representation: Z is a latent variable, X is a noise version of confounder Z, A is an action, and Y is an outcome variable.

Johansson et al., 2018; Yao et al., 2018). The divergence between factual and counterfactual hidden representations' distributions influences the maximum variance of the ATE since covariate and an action propagate through deep neural networks, assuming that the deep neural network has a finite Lipschitz constant. We can thus reduce the variance by optimizing the factual loss function while, for instance, minimizing the integral probability metric (IPM) as regularization. The overall objective function is

$$\mathcal{L}_{CFR} = \mathbb{E}_{Y,X,A}[wl_{\text{Factual}}(X, Y, A; u_A(X; \phi))] + \lambda \text{IPM}_{\mathcal{F}}(\hat{p}(\mathbf{h}_i(X)|A = 1), \hat{p}(\mathbf{h}_i(X)|A = 0)), \tag{14}$$

where $\mathbf{h}_i(X)$ is a hidden representation from the i-th layer of u_A with $i < k$. $\text{IPM}_{\mathcal{F}}$ is the integral probability metric with a class $\mathcal{F}$ of real-valued bounded measurable functions. The hidden representation $\mathbf{h}_i(X)$ does not depend on action A, although the data point x_i was collected together with some action a. $\hat{p}(\mathbf{h}_i(X)|A = 1)$ and $\hat{p}(\mathbf{h}_i(X)|A = 0)$ are thus the empirical probability distributions over the representations of treated and controlled groups, respectively. The 1-Lipschitz function and universal reproducing Hilbert kernel space, such as 1-Wasserstein distance (Cuturi & Doucet, 2013) and MMD (Gretton et al., 2012), is the most commonly used class of functions for IPMs (Shalit et al., 2017; Johansson et al., 2018).

Suppose $\text{IPM}_{\mathcal{F}}(\hat{p}(\mathbf{h}_i(X)|A = 1), \hat{p}(\mathbf{h}_i(X))|A = 0) = 0$. Then, we would not be able to tell whether the hidden representation $\mathbf{h}_i$ is likely conditioned on action or counterfactual action. The generalization error of ITE is bounded by the CFR objective and it has been empirically shown to have lower ATE for the out of sample data. However, this does not theoretically guarantee that generalization error will be always lower. We explain how the approach upper bounds the PEHE estimation in Appendix **??**.

Zeng et al. (2020) extend this approach to use a doubly robust estimator instead of IPW (see §4.4) and simultaneously minimize the Jensen-Shannon divergence between the treated and controlled group instead of IPM. Hassanpour & Greiner (2019) view ITE estimation problem from a domain adaption perspective where factual data is assumed to come from the source and the counteraction data from a different target distribution. They use importance sampling to re-weight the loss function for each factual data to make it look like it was sampled from the target distribution instead of the source distribution (Hassanpour & Greiner, 2019). Instead of globally balancing the treatment and controlled posterior distributions, Yao et al. (2018) propose a local similarity preserved individual treatment effect (SITE) estimation method based on deep representation learning (Yao et al., 2018). Their method preserves the local similarity and balances between the factual and counterfactual distributions over the set of actions, simultaneously. Furthermore, there is a line of work where they separately extract the representations of confounders and non-confounders and re-weight the confounder representation only (Kuang et al., 2017; Wu et al., 2020).

Domain invariance, or equivalently the full overlap between the factual and counterfactual distributions, can be an overly restrictive criterion, as it may remove information from input variables (covariates and sometimes action) that may be necessary for accurately estimating the treatment effect. (Stojanov et al., 2021). Yao et al. (2020) demonstrate why it is not ideal to use a distributional divergence to balance the treated and controlled representations (Yao et al., 2020). Instead, they propose to minimize the counterfactual variance and make the hidden representation invertible by adding a reconstruction loss function, which is, they claim, enough to have sufficient overlap in the factual and counterfactual supports.

Deep kernel learning for individual treatment effect (DKLITE) is a representative passive CI method that uses variance reduction (Yao et al., 2020). Unlike the methods above, this algorithm manipulates the action-dependent hidden representation. More formally, it uses kernel regression to estimate the potential outcome $\hat{y}_i = W_a \mathbf{h}_{m-1}(x_i) + \epsilon_{i,a}$ on top of the final layer of deep neural network $\mathbf{h}_{m-1}(x)$, where W_a is the parameter for action a and $\epsilon_{i,a}$ is an action-dependent noise variable. The posterior distribution is $\mathcal{N}(m_a \mathbf{h}_{m-1}(x), \sigma^2(x; \mathcal{X}, \Theta_a))$, where $\sigma^2(x; \mathcal{X}, \Theta_a) = \mathbf{h}_{m-1}^T K_a^{-1} \mathbf{h}_{m-1}$ is the variance. DKLITE objective function minimizes both the negative log-likelihood and posterior variance given the counterfactual actions. Although DKLITE uses kernel regression to derive the posterior distribution, it is not necessary to use kernel regression.

The objective function for training a causal inference model with posterior variance reduction can be written as

$$\mathcal{L}_{VR} = w\mathcal{L}_{\text{Factual}}(X, Y, A; u_A(X; \phi)) + \lambda \mathbb{E}_{p(X,A)}[g(\text{VAR}[h_\theta(X, 1 - A))]], \tag{15}$$

where $\text{VAR}[h(X, 1 - A)]$ is the posterior variance given a data point $X = x$ and a counterfactual action. The function $g : \mathbb{R}^D \to \mathbb{R}$ aggregates the covariance of high dimensional representations into a single scalar. For example, $g(\cdot)$ can be a sum of the element-wise variances of the hidden representation. By reducing the variance in the representations $h(X, 1 - A)$, we also reduce the variance in the ATE.

4.5.2 Deep Latent Variables CI Models

In deep latent variable models for causal inference, we assume a particular data-generating process described with a probabilistic graphical model that contains latent (or hidden) stochastic variables (Louizos et al., 2017a; Rakesh et al., 2018; Vowels et al., 2020; Wu & Fukumizu, 2021; jiwoong Im et al., 2021; Kumor et al., 2021; Lu et al., 2022). The inclusion of such latent variables enables us to model the potential outcome with a much richer distribution. Without latent variables, it is challenging to build a function approximator, such as a deep neural network, that captures a multimodal distribution of the potential outcome, regardless of how complex a form such a function takes.

Causal effects are however not identifiable in general when there are latent variables. Identifiability requires that the model parameter can be uniquely estimated from the data. However, since the latent variables do not directly measure the unobserved variable but infer from the observed variables, it introduces ambiguity that leads to violates the assumption of unconfoundedness. In order to overcome this issue, one has to make two assumptions; (1) X is a proxy variable that is a noisy version of a hidden confounder, and (2) this unknown confounder can be modelled by these latent variables (see Figure 5b). While these extra assumptions are a major drawback of deep latent variable-based CI (Louizos et al., 2017b; Kocaoglu et al., 2017), we nevertheless review this literature as they are increasingly more widely used in practice (Pearl et al., 2016; Wu & Fukumizu, 2021; Trifunov et al., 2020; Kumor et al., 2021; Rissanen & Marttinen, 2021).

With these assumptions, we can consider the latent variable Z a hidden, i.e. unobserved, confounder, to which we have access via its noisy realization X. We can recover the joint distribution $p(Z, X, Y, A)$ from observational data (X, Y, A). We can then compute $p(Y|X = x, do(A = 1))$ and $p(Y|X = x, do(A = 0))$, which allows us to compute ITE, by

$$\begin{aligned} p(Y|X = x, do(A = a)) &= \int_z p(Y|X, do(A = a), Z) p(Z|X = x, do(A = a)) dZ \\ &= \int_z p(Y|X = x, do(A = a), Z) p(Z|X = x, do(A = a)) dZ \\ &= \int_z p(Y|X = x, A = a, Z) p(Z|X = x) dZ. \end{aligned}$$

The third equality follows from the rule of *do*-calculus. Therefore, we can estimate $p(Y|X = x, do(A = a))$ as long as we can approximate $p(Y|A = a, Z)$ and $p(X|Z)$.

The causal effect variational autoencoder (CEVAE) is a particular type of variational inference framework which allows us to estimate $p(Y|A, Z)$ and $p(Z|X)$ using deep neural networks (Kingma & Welling, 2014;

Louizos et al., 2017a). With this VAE, the true posterior distribution is defined as $p_\theta(Z|X,Y,A) \propto p_\theta(Y,A|Z)p_\theta(X|Z)p(Z)$, where both$p_\theta(Y,A|Z)$ and $p_\theta(X|Z)$ are modelled using a deep neural network parametrized by θ and $p(Z)$ is a prior distribution and is not parameterized by θ. We approximate the posterior distribution with a variational posterior $q_\phi(Z|X,Y,A)$ which is modelled by a deep neural network with a variational parameter ϕ. We infer the hidden confounder Z from the observation (X,Y,A) using this variational posterior neural network.

We estimate $p(Y|A,Z)$ and $p(Z|X)$ directly by training both generative and inference networks on observational data. Training is done to maximize the following variational lower bound with respect to the parameters θ and ϕ:

$$\mathcal{L}_{\text{CEVAE}} = \mathbb{E}_{p_{data}(X,Y,A)}\left[\mathbb{E}_{q_\phi(Z|X,Y,A)}\left[\log p_\theta(X,Y,A|Z)\right] - \mathbb{KL}\left[q_\phi(Z|X,Y,A)\|p(Z)\right]\right].$$

The first term is the reconstruction of observable variables from the inferred confounder Z, and the the second term is a regularizer which enforces the approximate posterior to be close to the prior and maximizes the entropy of the posterior distribution over the confounder Z. We jointly update both generative and inference network parameters by using backpropagation and the re-parameterization trick (Danilo Jimenez & Shakir, 2014; Kingma & Welling, 2014).

Similar to CEVAE, linked causal variational autoencoder (LCVA) treats the latent attributes directly as confounders with the assumption that these confounders affect both the treatment and the outcome of units (Rakesh et al., 2018). The main difference is that the authors want to measure the causal effect when there exists *a spillover effect*[9] between pairs of two covariates through the confounders. Another variant is the Causal Effect by using Variational Information Bottleneck (CEVIB) (Lu et al., 2022). Just like any other variational latent model, it learns to fit the model to observation data and learns the confounders that affect treatments and outcomes using variational information bottleneck (Alemi et al., 2016). CEVIB does this in a way that allows the model to forget some latent variables that are not confounders and learn to extract only the confounding information from covariate. Deep entire space cross networks for individual treatment effect estimation (DESCN) attempt to learn the latent confounders ("the hidden treatment effect") through a cross-network in a multi-task learning manner (Zhong et al., 2022). It reduces treatment biases, that favour one treatment over another, by learning from multiple tasks and overcoming sample imbalance.

In CEVAE, the conditional distribution $p_\theta(X,Y,A|Z)$ is learned from data sampled from $p(A|Z)$, $p(X|Z)$ and $p(Y|A,Z)$, where $Z \sim p(Z)$. This conditional distribution, which is used for computing treatment effect, is however used with an action A sampled from $p(A)$ rather than $p(A|Z)$ in the inference time. This discrepancy is known as covariate shift or more generally distribution shift and is detrimental to generalization in deep learning (Shimodaira, 2000; Sergey & Christian, 2015; Jeong & Namkoong, 2020; Louizos et al., 2017b), which in turn results in a degradation in the quality of ATE estimation.

Replacing the observational distribution with a uniform treatment distribution, which is independent of the covariate, provides randomized treatment samples for training a CEVAE. A uniform treatment selection process decouples Z and A, thereby making A independent of the covariate X, i.e. $p(A|X) = p(A)$. This is similar to a randomized clinical trial over treatment A in Section 3.1. For this reason, it may be beneficial to train a CEVAE using a uniform treatment distribution. Here, the observational data-based distribution is $p(X,Y,A) = p(A|X)p(X)p(Y|X,A)$ and the corresponding uniform treatment distribution is $r(X,Y,A) = r(A)p(X)p(Y|X,A)$.

jiwoong Im et al. (2021) use importance weighting to write the variational lower bound objective under the uniform treatment distribution (jiwoong Im et al., 2021):

$$\mathcal{L}_{\text{UTVAE}} = \mathbb{E}_{p(X,Y,A)}\left[w(X,A)\mathbb{E}_{q_\phi(Z|X,Y,A)}\left[\log \frac{p_\theta(X,Y,A|Z)p(Z)}{q_\phi(Z|X,Y,A)}\right]\right],$$

where $w(X,A) = \frac{r(A|X)}{p(A|X)} = \frac{1}{2p(A|X)}$ is the importance weight. $\frac{r(A|X)}{p(A|X)} = \frac{r(X,Y,A)}{p(X,Y,A)}$ and $r(A|X) = r(A) = \frac{1}{2}$, because of the independence between X and A in the causal graph and the uniformly distributed treatment

[9] A spillover happens when something in one situation affects something else in a different situation, even though they may not seem related.

selection procedure. UTVAE generalizes better than CEVAE especially when there is a distribution shift between training and inference. See the details of the training procedure of CEVAE and UTVAE for generative and inference networks respectively at (jiwoong Im et al., 2021).

4.5.3 Combining active and passive methods

So far, we have discussed active and passive CI learning separately. It is however often necessary to combine active and passive approaches in order to mitigate the bias arising from non-randomized data. Here, we review some of the methods that combine the two.

Sawant et al. (2018) use a bandit method to collect data online and estimate the ATE offline using the potential outcome model (Sawant et al., 2018). At each time step, they sample data using a bandit algorithm like Thompson sampling and aggregate a dataset (see Algorithm 5). Using this dataset, they update the model and re-estimate the potential outcome (see Algorithm 6). This approach combines active and passive learning since the model is updated in a batch training setting while the dataset can be collected asynchronously. Similarly, Ye et al. (2020) proposes a framework that combines the two methods. Here they use inverse probability weighting and matching algorithms for passive learning and UCB and LinUCB (Slivkins, 2019)) for active CI algorithms (Ye et al., 2020). In the inference time, Ye et al. (2020) estimate the potential outcomes using a passive method given a new unseen example $X = x_t$ for every action. If the new data point $(X = x_t, A = a_t)$ with an action a_t is not in a non-randomized dataset, then the algorithm resorts to exploring the action for $X = x_t$ and adds the new data point $(X = x_t, A = a_t, Y = y_t)$ to the non-randomized dataset.

Algorithm 5 Online Scoring and Batch training

Iteration $t = 0$; Log $L = \{\}$; contextual distribution parameter θ_0 and θ_1.
for $i = 1, 2, \cdots, T$ **do**
  for $t = 1, 2, \cdots, T$ **do**
    Sample data x_t
    Predict $\hat{Y}(1) = f(x_t, \theta_1)$
    Predict $\hat{Y}(0) = f(x_t, \theta_0)$
    Choose action $a_t = \text{argmax}_{a \in \{0,1\}} \hat{Y}(a)$
    Compute $p_t = p(a_t | x_t)$
    $L = L \cup (x_t, a_t, p_t)$
  end for
  Update θ_0 and θ_1 using Algo 6
end for

Algorithm 6 Offline Batch Training

Dataset $D = \{\}$.
for $i = 1, 2, \cdots, T$ **do**
  Sample $(x_i, a_i, p_i, y_i(a_i))$ from L
  $\hat{y}_i(a_i) = y_i(a_i)/p_i$ - bias correction
  $\hat{y}_i(\neg a_i) = 0$
  for $m = 1, 2, \cdots, M$ **do**
    Sample $(x_m, \neg a_i, p_m, y_m(\neg a_m))$ from L
    $\hat{y}_i(\neg a_i) \leftarrow \hat{y}_i(\neg a_i) + \frac{1}{M} y_m(\neg a_i)/p_m$
  end for
  $\text{CATE}_i = \hat{y}_i(a_i) - \hat{y}_i(\neg a_i)$
  $D = D \cup (x_i, a_i, \text{CATE}_i)$
end for
update θ_0 and θ_1 by maximizing the likelihood on D.

5 Conclusion

The objective of this paper is to introduce various algorithms and frameworks in causal inference by categorizing them into passive and active algorithms. We have particularly focused on estimating average treatment effect (ATE), after outlining the standard assumptions necessary for the identification of causal effects: positivity, ignoreability, conditional exchangeability, consistency and Markov assumptions.

We first present the randomized controlled trial (RCT) as a representative example of an active causal inference algorithm. We then delve into bandit approaches that aim to balance the outcome itself and the quality of estimating the treatment effect. We explore different contextual bandit algorithms by considering various causal graph scenarios such as taking account of non-confounding variables or dealing with unknown confounding variables.

We then move on to discussing passive CI methods, including matching, inverse probability weighting and doubly robust methods. We touched upon deep learning-based approaches as well. A majority of these studies focus on converting a causal estimand into a statistical estimand, and subsequently, estimating the statistical estimand in order to obtain the causal estimate. These classical methods unfortunately do not perform well when they are put to work with high dimensional data. In order to overcome this challenge, deep learning has been proposed as a way to learn a compact representation suitable for estimating ATE. We thus discussed several deep learning-based approaches that learn to infer causal effects by automatically extracting hidden or unknown confounders' representations in problems with high dimensional data.

After reading the main part of this paper, readers should notice a clear resemblance between offline policy evaluation and passive causal inference methods. This resemblance is due to the similarity in estimating the reward in policy evaluation and estimating the potential outcomes in causal inference (Swaminathan & Joachims, 2015; Li, 2015). For example, some of the methods discussed in Section 4.5.3 have been used for offline policy evaluation in contextual bandit algorithms (Li et al., 2012; Sawant et al., 2018). Although we do not explore this connection further here, this is an important avenue to pursue both causal inference and reinforcement learning.

This review of causal inference is limited in two ways. First, we do not consider collider bias, and second, we assume a stationary conditional probability distribution. We discuss these two limitations briefly before ending the paper.

Collider bias happens when there is an extra variable, caused by both action and outcome variables, that is observed to be (conditioned on) a particular value. For example, suppose we investigate the effect of a new versus old medication on the patient outcome, and we gather data from a hospital. We divide the patients into two groups: those who received both the new and old medications. The study finds that the patients who received the old medication had better outcomes than the new one and the hospital conclude that the old medication is more effective on patient outcomes. It turns out this conclusion has a collider bias. This is because the decision to give the medication is based on the patient's recovery rate, which encourages doctors to prescribe a more potent drug. Thus, patients are more likely to receive the old medication which is stronger. In this case, the patient's recovery rate becomes a collider variable because it is caused by both the decision to give which medication and the patient's outcome. this is very separate from confounder bias and it doesn't address the collider bias issue

Besides collider bias, we have assumed that all conditional distributions are stationary (i.e. do not change over the course of causal inference and data collection). The question is what happens if these conditional distributions shift over time? One can introduce temporal dependencies to a causal graph to describe such shifts over time. Still, it is challenging to work with such a graph because collected data points over time are correlated with each other. This violates the no interference/Markov assumption we discussed earlier in Section 2.3. To address this problem, various time series methods have been developed that take into account the temporal dependence of the data. For instance, deep sequential weighting (DSW) and sequential causal effect variational autoencoder (SEVAE) estimate ITE with time-varying confounders (Trifunov et al., 2022; Liu et al., 2020; Kumor et al., 2021).

References

Alberto Abadie and Guido W. Imbens. Matching on the estimated propensity score. *Econometrica*, 84(2): 781–807, 2016.

Banerjee Abhijit V. and Duflo Esther. Poor economics: A radical rethinking of the way to fight global poverty. In *Public Affairs*, 2012.

Alexander A. Alemi, Ian Fischer, Joshua V. Dillon, and Kevin Murphy. Deep variational information bottleneck. 2016.

Lateef Amusa. Reducing bias in observational studies: An empirical comparison of propensity score matching methods. *Turkiye Klinikleri Journal of Biostatistics*, 10:13–26, 01 2018. doi: 10.5336/biostatic.2017-58633.

Susan Athey, Guido Imbens, and Vikas Ramachandra. Machine learning methods for estimating heterogeneous causal effects. 04 2015.

"Peter Auer, Paul Fischer, and Nicolo Cesa-Bianchi". "finite-time analysis of the multiarmed bandit problem". *"Machine Learning"*, "47"("3"):"235–256", "2002".

Elias Bareinboim, Andrew Forney, and Judea Pearl. Bandits with unobserved confounders: A causal approach. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL `https://proceedings.neurips.cc/paper/2015/file/795c7a7a5ec6b460ec00c5841019b9e9-Paper.pdf`.

Alexandre Belloni, Victor Chernozhukov, and Christian Hansen. Inference on treatment effects after selection amongst high-dimensional controls. *Operations Research eJournal*, 2011.

Andrea Bender. What is causal cognition? *Frontiers in Psychology*, 11, 01 2020. doi: 10.3389/fpsyg.2020.00003.

Donald A Berry and Bert Fristedt. Bandit problems: sequential allocation of experiments (monographs on statistics and applied probability). *London: Chapman and Hall*, 5(71-87):7–7, 1985.

Djallel Bouneffouf, Irina Rish, and Charu Aggarwal. Survey on applications of multi-armed and contextual bandits. In *2020 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1–8, 2020. doi: 10.1109/CEC48606.2020.9185782.

Norman E. Breslow, Thomas Lumley, Christie M. Ballantyne, Lloyd E. Chambless, and Michal Kulich. Using the whole cohort in the analysis of case-cohort data. *American journal of epidemiology*, 169:1398–1405, 2009.

Bruce W. Carlson. Simpson's paradox. *Encyclopedia Britannica*, 2019. doi: https://www.britannica.com/topic/Simpsons-paradox.

Daniel C. Castro, Ian Walker, and Ben Glocker. Causality matters in medical imaging. *Nature Communication*, 3673, 2020. doi: https://doi.org/10.1038/s41467-020-17478-w.

T. C. Chalmers, Jr Smith, H., B. Blackburn, B. Silverman, B. Schroeder, D. Reitman, and A. Ambroz. A method for assessing the quality of a randomized control trial. *Controlled clinical trials*, 2:31–49, 1981.

Thomas C. Chalmers, Paul Hewett, Dinah Reitman, and Henry S Sacks. Selection and evaluation of empirical research in technology assessment. *International Journal of Technology Assessment in Health Care*, 5:521 – 536, 1989.

Olivier Chapelle, Eren Manavoglu, and Romer Rosales. Simple and scalable response prediction for display advertising. *ACM Transaction Intelligence System Technology*, 5(4), 2015.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James M. Robins. Double/debiased machine learning for treatment and structural parameters. *Econometrics: Econometric & Statistical Methods - Special Topics eJournal*, 2017.

Niteesh K. Choudhry. Randomized, controlled trials in health insurance systems. *The New England Journal of Medicine*, 377:957–964, 2017.

William G. Cochran and Donald B. Rubin. Controlling bias in observational studies: A review. *The Indian Journal of Statistics, Series A*, 35:417–446, 1973. doi: JSTOR, http://www.jstor.org/stable/25049893.

John Concato, Nirav Shah, and Ralph I. Horwitz. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *New England Journal of Medicine*, 342(25):1887–1892, 2000.

Marco Cuturi and A. Doucet. Fast computation of wasserstein barycenters. In *International Conference on Machine Learning*, 2013.

Rezende Danilo Jimenez and Mohamed Shakir. Stochastic backpropagation and approximate inference in deep generative models. In *arXiv preprint arXiv:1401.4082*, 2014.

Angus Deaton and Nancy Cartwright. Understanding and misunderstanding randomized controlled trials. *Behavioral & Experimental Economics eJournal*, 2016.

Rajeev Dehejia and Sadek Wahba. Propensity score matching methods for non-experimental causal studies. *The Review of Economics and Statistics*, 84:151–161, 02 2002. doi: 10.1162/003465302317331982.

Rajeev H. Dehejia and Sadek Wahba. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94(448):1053–1062, 1999. doi: 10.1080/01621459.1999.10473858. URL `https://www.tandfonline.com/doi/abs/10.1080/01621459.1999.10473858`.

Alexis Diamond and Jasjeet S. Sekhon. Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. *The Review of Economics and Statistics*, 95(3):932–945, 2013.

Maria Dimakopoulou, Susan Athey, and Guido Imbens. Estimation considerations in contextual bandits. 11 2017.

Max H. Farrell. Robust inference on average treatment effects with possibly more covariates than observations. *Journal of Econometrics*, 189(1):1–23, 2015. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2015.06.017. URL `https://www.sciencedirect.com/science/article/pii/S0304407615001864`.

Paul J. Ferraro, James N. Sanchirico, and Martin D. Smith. Causal inference in coupled human and natural systems. *Proceedings of the National Academy of Sciences*, 116(12):5311–5318, 2019. doi: 10.1073/pnas.1805563115. URL `https://www.pnas.org/doi/abs/10.1073/pnas.1805563115`.

Thomas R. Frieden. Evidence for health decision making — beyond randomized, controlled trials: The changing face of clinical trials. *The New England Journal of Medicine*, 377:465–475, 2017.

Raphaël Féraud, Robin Allesiardo, Tanguy Urvoy, and Fabrice Clérot. Random forest for the contextual bandit problem. In Arthur Gretton and Christian C. Robert (eds.), *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51 of *Proceedings of Machine Learning Research*, pp. 93–101, Cadiz, Spain, 09–11 May 2016. PMLR.

Fernando Martel García and Léonard Wantchekon. Theory, external validity, and experimental inference: Some conjectures. *The ANNALS of the American Academy of Political and Social Science*, 628:132 – 147, 2010.

Brett R. Gordon, Florian Zettelmeyer, Neha Bhargava, and Dan Chapsky. A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook. *Marketing Science*, 38(2): 193–225, 2019.

Thore Graepel, Joaquin Quiñonero Candela, Thomas Borchert, and Ralf Herbrich. Web-scale bayesian click-through rate prediction for sponsored search advertising in microsoft's bing search engine. In *Proceedings of the 27th International Conference on Machine Learning ICML 2010, Invited Applications Track (unreviewed, to appear)*, June 2010. Invited Applications Track.

L. W. Green and Russell E. Glasgow. Evaluating the relevance, generalization, and applicability of research. *Evaluation & the Health Professions*, 29:126 – 153, 2006.

Kristjan Greenewald, Ambuj Tewari, Susan Murphy, and Predag Klasnja. Action centered contextual bandits. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL `https://proceedings.neurips.cc/paper/2017/file/4fa177df22864518b2d7818d4db5db2d-Paper.pdf`.

Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13(25):723–773, 2012. URL http://jmlr.org/papers/v13/gretton12a.html.

Xing Gu and Paul R. Rosenbaum. Comparison of multivariate matching methods: Structures, distances, and algorithms. *Journal of Computational and Graphical Statistics*, 2:405–420, 1993.

Jinyong Hahn. On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects. *Econometrica*, 66(2):315–332, March 1998.

Edward L Hannan. Randomized clinical trials and observational studies: guidelines for assessing respective strengths and limitations. *JACC. Cardiovascular interventions*, 1:211–217, 2008.

Ben Hansen. Full matching in an observational study of coaching for the sat. *Journal of the American Statistical Association*, 99:609–618, 02 2004. doi: 10.1198/016214504000000647.

Negar Hassanpour and Russell Greiner. Counterfactual regression with importance sampling weights. In *International Joint Conference on Artificial Intelligence*, 2019.

James Heckman, Hidehiko Ichimura, and Petra Todd. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies*, 64:605–54, 02 1997. doi: 10.2307/2971733.

James J. Heckman and Richard Robb. Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics*, 30(1):239–267, 1985. ISSN 0304-4076. doi: https://doi.org/10.1016/0304-4076(85)90139-3. URL https://www.sciencedirect.com/science/article/pii/0304407685901393.

Bang Heejung and Robins James M. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61(4):962–973, 07 2005. doi: 10.1111/j.1541-0420.2005.00377.x.

Miguel A Hernán and James M Robins. Estimating causal effects from epidemiological data. *Journal of epidemiology and community health*, 60:578–586, 2006.

Miguel A. Hernán, John Hsu, and Brian Healy. A second chance to get causal inference right: A classification of data science tasks. *CHANCE*, 32(1):42–49, 2019. doi: 10.1080/09332480.2019.1579578. URL https://doi.org/10.1080/09332480.2019.1579578.

Keisuke Hirano, Guido Imbens, and Geert Ridder. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, 71:1161–1189, 02 2003. doi: 10.1111/1468-0262.00442.

Paul W. Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986. doi: 10.1080/01621459.1986.10478354. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1986.10478354.

Stefano M. Iacus, Gary King, and Giuseppe Porro. Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20(1):1–24, 2012. doi: 10.1093/pan/mpr013.

Guido W Imbens. Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and statistics*, 86(1):4–29, 2004.

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

Jyotsna Jalan and Martin Ravallion. Estimating the benefit incidence of an antipoverty program by propensity-score matching. *Journal of Business and Economic Statistics*, 21, 12 2001. doi: 10.1198/073500102288618720.

Dan Jane-wit, Ralph I Horwitz, and John Concato. Variation in results from randomized, controlled trials: stochastic or systematic? *Journal of clinical epidemiology*, 63 1:56–63, 2010.

Sookyo Jeong and Hongseok Namkoong. Robust causal inference under covariate shift via worst-case subpopulation treatment effects. *ArXiv*, abs/2007.02411, 2020.

Daniel jiwoong Im, Kyunghyun Cho, and Narges Razavian. Causal effect variational autoencoder with uniform treatment for overcoming covariate shifts. In *arXiv preprint arXiv:2111.08656*, 2021.

Fredrik D. Johansson, Uri Shalit, and David Sontag. Learning representations for counterfactual inference. In *arXiv preprint arXiv:1605.03661*, 2016.

Fredrik D. Johansson, Nathan Kallus, Uri Shalit, and David A. Sontag. Learning weighted representations for generalization across designs. *arXiv: Machine Learning*, 2018.

Joseph Kang and Joseph L. Schafer. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22:523–539, 2007.

Richard Karp. Reducibility among combinatorial problems. volume 40, pp. 85–103, 01 1972. ISBN 978-3-540-68274-5. doi: 10.1007/978-3-540-68279-0-8.

J M Kendall. Designing a research project: randomised controlled trials and their principles. *Emergency Medicine Journal*, 20(2):164–168, 2003. ISSN 1472-0205. doi: 10.1136/emj.20.2.164. URL `https://emj.bmj.com/content/20/2/164`.

Celeste Kidd and Benjamin Y. Hayden. The psychology and neuroscience of curiosity. *Neuron*, 88(3):449–460, November 2015. ISSN 0896-6273. doi: 10.1016/j.neuron.2015.09.010. Funding Information: This research was supported by a grant from the NIH, R01 (DA038615) (to B.Y.H.). We thank Sarah Heilbronner, Steve Piantadosi, Shraddha Shah, Maya Wang, Habiba Azab, and Maddie Pelz for helpful comments. Publisher Copyright: © 2015 Elsevier Inc.

Yongnam Kim and Peter Steiner. Quasi-experimental designs for causal inference. *Educ Psychol*, 51:395–405, 2016. doi: 10.1080/00461520.2016.1207177.

Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *arXiv preprint arXiv:1312.6114*, 2014.

Murat Kocaoglu, Karthikeyan Shanmugam, and Elias Bareinboim. Experimental design for learning causal graphs with latent variables. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL `https://proceedings.neurips.cc/paper/2017/file/291d43c696d8c3704cdbe0a72ade5f6c-Paper.pdf`.

Akshay Krishnamurthy, Zhiwei Steven Wu, and Vasilis Syrgkanis. Semiparametric contextual bandits. arXiv, 2018. doi: 10.48550/ARXIV.1803.04204. URL `https://arxiv.org/abs/1803.04204`.

Kun Kuang, Peng Cui, B. Li, Meng Jiang, Shiqiang Yang, and Fei Wang. Treatment effect estimation with data-driven variable decomposition. In *AAAI Conference on Artificial Intelligence*, 2017.

Daniel Kumor, Junzhe Zhang, and Elias Bareinboim. Sequential causal imitation learning with unobserved confounders. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 14669–14680. Curran Associates, Inc., 2021. URL `https://proceedings.neurips.cc/paper/2021/file/7b670d553471ad0fd7491c75bad587ff-Paper.pdf`.

Robert J. LaLonde. Evaluating the econometric evaluations of training programs with experimental data. *The American economic review*, pp. 604–620, 1986.

John Langford and Tong Zhang. The epoch-greedy algorithm for multi-armed bandits with side information. In J. Platt, D. Koller, Y. Singer, and S. Roweis (eds.), *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2007. URL `https://proceedings.neurips.cc/paper/2007/file/4b04a686b0ad13dce35fa99fa4161c65-Paper.pdf`.

Finnian Lattimore, Tor Lattimore, and Mark D. Reid. Causal bandits: Learning good interventions via causal inference, 2016. URL https://arxiv.org/abs/1606.03203.

Lihong Li. Offline evaluation and optimization for interactive systems. In *Proceedings of the 8th ACM International Conference on Web Search and Data Mining*. ACM - Association for Computing Machinery, February 2015. URL https://www.microsoft.com/en-us/research/publication/offline-evaluation-and-optimization-for-interacti

Lihong Li, Wei Chu, John Langford, and Robert E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pp. 661–670, New York, NY, USA, 2010. Association for Computing Machinery. ISBN 9781605587998. doi: 10.1145/1772690.1772758. URL https://doi.org/10.1145/1772690.1772758.

Lihong Li, Wei Chu, John Langford, Taesup Moon, and Xuanhui Wang. An unbiased offline evaluation of contextual bandit algorithms with generalized linear models. In Dorota Glowacka, Louis Dorard, and John Shawe-Taylor (eds.), *Proceedings of the Workshop on On-line Trading of Exploration and Exploitation 2*, volume 26 of *Proceedings of Machine Learning Research*, pp. 19–36, Bellevue, Washington, USA, 02 Jul 2012. PMLR.

Ruoqi Liu, Changchang Yin, and Ping Zhang. Estimating individual treatment effects with time-varying confounders, 2020.

Christos Louizos, Uri Shalit, Joris Mooij, David Sontag, Richard Zemel, and Max Welling. Causal effect inference with deep latent-variable models. In *arXiv preprint arXiv:1705.08821*, 2017a.

Christos Louizos, Uri Shalit, Joris M. Mooij, David A. Sontag, Richard S. Zemel, and Max Welling. Causal effect inference with deep latent-variable models. In *NIPS*, 2017b.

Bo Lu and Paul ROSENBAUM. Optimal pair matching with two control groups. *Journal of Computational and Graphical Statistics - J COMPUT GRAPH STAT*, 13:422–434, 06 2004. doi: 10.1198/1061860043470.

Bo Lu and Paul R. Rosenbaum. Optimal pair matching with two control groups. *Journal of Computational and Graphical Statistics*, 13:422 – 434, 2004.

Zhenyu Lu, Yurong Cheng, Mingjun Zhong, George Stoian, Ye Yuan, and Guoren Wang. *Causal Effect Estimation Using Variational Information Bottleneck*, pp. 288–296. 12 2022. ISBN 978-3-031-20308-4. doi: 10.1007/978-3-031-20309-1-25.

Jared K Lunceford and Marie Davidian. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in Medicine*, 23, 2004.

Imke Mayer, Erik Sverdrup, Tobias Gauss, Jean-Denis Moyer, Stefan Wager, and Julie Josse. Doubly robust treatment effect estimation with missing attributes. *The Annals of Applied Statistics*, 14(3):1409 – 1431, 2020. doi: 10.1214/20-AOAS1356. URL https://doi.org/10.1214/20-AOAS1356.

Kathryn Monahan, Joanna Williams, and Laurence Steinberg. Revisiting the impact of part-time work on adolescent adjustment: Distinguishing between selection and socialization using propensity score matching. *Child development*, 82:96–112, 01 2011. doi: 10.1111/j.1467-8624.2010.01543.x.

Edson Duarte Moreira and Ezra S Susser. Guidelines on how to assess the validity of results presented in subgroup analysis of clinical trials. *Revista do Hospital das Clinicas*, 57 2:83–8, 2002.

Rashelle J. Musci and Elizabeth A. Stuart. Ensuring causal, not casual, inference. *Prevention Science*, 20: 452–456, 2019.

Razieh Nabi and Ilya Shpitser. Semi-parametric causal sufficient dimension reduction of high dimensional treatments. *arXiv: Methodology*, 2017.

Susan Armijo Olivo, Luciana Gazzi Macedo, Inae Caroline Gadotti, Jorge Fuentes, Tasha R. Stanton, and D J Magee. Scales to assess the quality of randomized controlled trials: A systematic review. *Physical Therapy*, 88:156 – 175, 2008.

Aronow P. M., Robins James M., Saarinen Theo, Sävje Fredrik, and Sekhon Jasjeet. Nonparametric identification is not enough, but randomized controlled trials are. In *arXiv preprint arXiv:2108.11342*, 2021.

Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2nd edition, 2009.

Judea Pearl. An introduction to causal inference. *The international journal of biostatistics*, 6:1557–4679, 2010.

Judea Pearl. Detecting latent heterogeneity. In *Sociological Methods & Research*, 2015.

Judea Pearl, M Maria Glymour, and Nicholas P Jewell. Causal inference in statistics: A primer. 2016.

Yi Peng, Miao Xie, Jiahao Liu, Xuying Meng, Nan Li, Cheng Yang, Tao Yao, and Rong Jin. A practical semi-parametric contextual bandit. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, IJCAI'19, pp. 3246–3252. AAAI Press, 2019. ISBN 9780999241141.

Aahlad Puli, Nitish Joshi, He He, and Rajesh Ranganath. Nuisances via negativa: Adjusting for spurious correlations via data augmentation, 2022. URL `https://arxiv.org/abs/2210.01302`.

Nicholas Radcliffe. Using control groups to target on predicted lift: Building and assessing uplift model. 2007.

Vineeth Rakesh, Ruocheng Guo, Raha Moraffah, Nitin Agarwal, and Huan Liu. Linked causal variational autoencoder for inferring paired spillover effects. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, pp. 1679–1682, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450360142.

María Resa and José Zubizarreta. Evaluation of subset matching methods and forms of covariate balance. *Statistics in medicine*, 35, 07 2016. doi: 10.1002/sim.7036.

Severi Rissanen and Pekka Marttinen. A critical look at the consistency of causal estimation with deep latent variable models. 2021.

James Robins and A G Rotnitzky. Semiparametric efficiency in multivariate regression models with missing data. *Journal of The American Statistical Association - J AMER STATIST ASSN*, 90:122–129, 03 1995. doi: 10.1080/01621459.1995.10476494.

James M. Robins, Andrea Rotnitzky, and Lue Ping Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866, 1994. doi: 10.1080/01621459.1994.10476818. URL `https://doi.org/10.1080/01621459.1994.10476818`.

James M. Robins, Miguel A. Hernán, and Babette A. Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11:550–560, 2000.

James M. Robins, Mariela Sued, Quanhong Lei-Gomez, and Andrea Rotnitzky. Comment: Performance of double-robust estimators when "inverse probability" weights are highly variable. *Statistical Science*, 22: 544–559, 2007.

PAUL R. ROSENBAUM and DONALD B. RUBIN. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 04 1983. ISSN 0006-3444. doi: 10.1093/biomet/70.1.41.

Peter M. Rothwell. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *The Lancet*, 365:82–93, 2005.

Donald Rubin and Elizabeth Stuart. Affinely invariant matching methods with discriminant mixtures of proportional ellipsoidally symmetric distributions. *The Annals of Statistics*, 34, 12 2006. doi: 10.1214/009053606000000407.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

Donald B Rubin. Assignment to treatment group on the basis of a covariate. *Journal of educational Statistics*, 2(1):1–26, 1977.

Donald B. Rubin. Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, 74:318–328, 1979.

Donald B Rubin. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

DONALD B. RUBIN and NEAL THOMAS. Characterizing the effect of matching using linear propensity score methods with normal distributions. *Biometrika*, 79(4):797–809, 12 1992. ISSN 0006-3444. doi: 10.1093/biomet/79.4.797. URL `https://doi.org/10.1093/biomet/79.4.797`.

Vin Sachidananda and Emma Brunskill. Online learning for causal bandits. In *Advances in Neural Information Processing Systems*, 2017.

Yuta Saito and Shota Yasui. Counterfactual cross-validation: Effective causal model selection from observational data. In *arXiv preprint arXiv:1909.05299*, 2019.

Sara Saturni, Federico Bellini, Fulvio Braido, Pierluigi Paggiaro, Alessandro Sanduzzi, Nicola Scichilone, Pierachille Santus, Luca Morandi, and Alberto Papi. Randomized controlled trials and real life studies. approaches and methodologies: a clinical point of view. *Pulmonary pharmacology & therapeutics*, 27 2: 129–38, 2014.

Neela Sawant, Chitti Babu Namballa, Narayanan Sadagopan, and Houssam Nassif. Contextual multi-armed bandits for causal marketing. *CoRR*, abs/1810.01859, 2018. URL `http://arxiv.org/abs/1810.01859`.

Joseph Schafer and Joseph Kang. Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological methods*, 13:279–313, 01 2009. doi: 10.1037/a0014268.

Daniel SCHARFSTEIN, A G Rotnitzky, and James Robins. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *JASA. Journal of the American Statistical Association*, 94, 12 1999. doi: 10.2307/2669923.

Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Towards causal representation learning, 2021. URL `https://arxiv.org/abs/2102.11107`.

Anthony D. Scotina and Roee Gutman. Matching algorithms for causal inference with multiple treatments. *arXiv preprint arXiv:1809.00269*, 2019.

Shaun Seaman and Stijn Vansteelandt. Introduction to double robust methods for incomplete data. *Statistical Science*, 33:184–197, 05 2018. doi: 10.17863/CAM.23913.

Ioffe Sergey and Szegedy Christian. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *arXiv preprint arXiv:1502.03167*, 2015.

Uri Shalit, Fredrik D. Johansson, and David Sontag. Estimating individual treatment effect: Generalization bounds and algorithms. ICML'17, pp. 3076–3085. JMLR.org, 2017.

Michelle Shardell, Gregory E. Hicks, and Luigi Ferrucci. Doubly robust estimation and causal inference in longitudinal studies with dropout and truncation by death. *Biostatistics*, 16(1):155–168, 07 2014. doi: 10.1093/biostatistics/kxu032. URL `https://doi.org/10.1093/biostatistics/kxu032`.

Hidetoshi Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, 90:227–244, 2000.

Bonnie Sibbald and Martin Roland. Understanding controlled trials: Why are randomised controlled trials important? *BMJ*, 316:201, 1998.

Stephen D. Simon. Is the randomized clinical trial the gold standard of research? *Journal of andrology*, 22 6:938–43, 2001.

Aleksandrs Slivkins. Introduction to multi-armed bandits. *CoRR*, abs/1904.07272, 2019. URL `http://arxiv.org/abs/1904.07272`.

Leonard A. Stefanski and Dennis D. Boos. The calculus of m-estimation. *The American Statistician*, 56:29 – 38, 2002.

Lesley A Stewart and Mahesh K. B. Parmar. Bias in the analysis and reporting of randomized controlled trials. *International Journal of Technology Assessment in Health Care*, 12:264 – 275, 1996.

Petar Stojanov, Zijian Li, Mingming Gong, Ruichu Cai, Jaime Carbonell, and Kun Zhang. Domain adaptation with invariant representation learning: What transformations to learn? In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 24791–24803. Curran Associates, Inc., 2021. URL `https://proceedings.neurips.cc/paper/2021/file/cfc5d9422f0c8f8ad796711102dbe32b-Paper.pdf`.

Elizabeth A. Stuart. Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science*, 25(1):1 – 21, 2010. doi: 10.1214/09-STS313. URL `https://doi.org/10.1214/09-STS313`.

Adith Swaminathan and Thorsten Joachims. Counterfactual risk minimization: Learning from logged bandit feedback. *CoRR*, abs/1502.02362, 2015. URL `http://arxiv.org/abs/1502.02362`.

Adith Swaminathan, Akshay Krishnamurthy, Alekh Agarwal, Miroslav Dudík, John Langford, Damien Jose, and Imed Zitouni. Off-policy evaluation for slate recommendation. NIPS'17, pp. 3635–3645, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Zhiqiang Tan. Bounded, efficient and doubly robust estimation with inverse weighting. *Biometrika*, 97: 661–682, 2010.

Violeta Teodora Trifunov, Maha Shadaydeh, Jakob Runge, Veronika Eyring, Markus Reichstein, and Joachim Denzler. Causal link estimation under hidden confounding in ecological time series. 01 2020.

Violeta Teodora Trifunov, Maha Shadaydeh, and Joachim Denzler. Time series causal link estimation under hidden confounding using knockoff interventions, 2022.

Shalit Uri, Johansson Fredrik D., and David Songtag. Estimating individual treatment effect: generalization bounds and algorithms. In *arXiv preprint arXiv:1606.03976*, 2017.

Wouter A. C. van Amsterdam, Pim A. de Jong, Joost J. C. Verhoeff, Tim Leiner, and Rajesh Ranganath. Decision making in cancer: Causal questions require causal answers, 2022. URL `https://arxiv.org/abs/2209.07397`.

Victor Veitch, Dhanya Sridhar, and David Blei. Adapting text embeddings for causal inference. In Jonas Peters and David Sontag (eds.), *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 124 of *Proceedings of Machine Learning Research*, pp. 919–928. PMLR, 03–06 Aug 2020.

Matthew James Vowels, Necati Cihan Camgöz, and R. Bowden. Targeted vae: Structured inference and targeted learning for causal parameter estimation. *ArXiv*, abs/2009.13472, 2020.

Ingeborg Waernbaum and Laura Pazzagli. Model misspecification and bias for inverse probability weighting and doubly robust estimators. *arXiv: Statistics Theory*, 2017.

Miao Wang, Geng Zhi, and Tchetgen Eric Tchetgen. Identifying causal effects with proxy variables of an unmeasured confounder. In *arXiv preprint arXiv:1705.08821*, 2016.

Yixin Wang and Michael I. Jordan. Desiderata for representation learning: A causal perspective, 2021.

Sherry Weitzen, Kate Lapane, Alicia Toledano, Anne Hume, and Vincent Mor. Principles for modeling propensity scores in medical research: A systematic literature review. *Pharmacoepidemiology and drug safety*, 13:841–53, 12 2004. doi: 10.1002/pds.969.

Anpeng Wu, Kun Kuang, Junkun Yuan, Bo Li, Pan Zhou, Jianrong Tao, Qiang Zhu, Yueting Zhuang, and Fei Wu. Learning decomposed representation for counterfactual inference. *ArXiv*, abs/2006.07040, 2020.

Pengzhou (Abel) Wu and Kenji Fukumizu. Intact-vae: Estimating treatment effects under unobserved confounding. *ArXiv*, abs/2101.06662, 2021.

Liuyi Yao, Sheng Li, Yaliang Li, Mengdi Huai, Jing Gao, and Aidong Zhang. Representation learning for treatment effect estimation from observational data. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL `https://proceedings.neurips.cc/paper/2018/file/a50abba8132a77191791390c3eb19fe7-Paper.pdf`.

Zhang Yao, Bellot Alexis, and Schaar Mihaela van der. Learning overlapping representations for the estimation of individualized treatment effects. In *International Conference of Machine Learning*, 2020.

A Yazdani and E Boerwinkle. Causal inference in the age of decision medicine. *Journal of data mining in genomics & proteomics*, 6, 2015.

Li Ye, Yishi Lin, Hong Xie, and John C.s Lui. Combining offline causal inference and online bandit learning for data driven decisions. 01 2020. URL `http://arxiv.org/abs/2001.05699`.

Jinsung Yoon, James Jordon, and Mihaela van der Schaar. Ganite: Estimation of individualized treatment effects using generative adversarial nets. In *International Conference on Learning Representations*, 2018.

Shuxi Zeng, Serge Assaad, Chenyang Tao, Shounak Datta, Lawrence Carin, and Fan Li. Double robust representation learning for counterfactual prediction. *ArXiv*, abs/2010.07866, 2020.

Junzhe Zhang, Daniel Kumor, and Elias Bareinboim. Causal imitation learning with unobserved confounders. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020.

Yao Zhang, Jeroen Berrevoets, and Mihaela van der Schaar. Identifiable energy-based representations: An application to estimating heterogeneous causal effects. *ArXiv*, abs/2108.03039, 2021.

Zhong Zhao. Using matching to estimate treatment effects: Data requirements, matching metrics, and monte carlo evidence. *The Review of Economics and Statistics*, 86:91–107, 02 2004. doi: 10.1162/003465304323023705.

Min Zheng and Samantha Kleinberg. Using domain knowledge to overcome latent variables in causal inference from time series. In Finale Doshi-Velez, Jim Fackler, Ken Jung, David Kale, Rajesh Ranganath, Byron Wallace, and Jenna Wiens (eds.), *Proceedings of the 4th Machine Learning for Healthcare Conference*, volume 106 of *Proceedings of Machine Learning Research*, pp. 474–489. PMLR, 09–10 Aug 2019. URL `https://proceedings.mlr.press/v106/zheng19a.html`.

Min Zheng, Jessecae K. Marsh, Jeffrey V. Nickerson, and Samantha Kleinberg. How causal information affects decisions. *Cognitive Research: Principles and Implications*, 5, 2020.

Cai Zhihong and Kuroki Manabu. On identifying total effects in the presence of latent variables and selection bias. In *arXiv preprint arXiv:1206.3239*, 2012.

Kailiang Zhong, Fengtong Xiao, Yan Ren, Yaorong Liang, Wenqing Yao, Xiaofeng Yang, and Ling Cen. Descn: Deep entire space cross networks for individual treatment effect estimation. *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022.

José Zubizarreta. Using mixed integer programming for matching in an observational study of kidney failure after surgery. *JASA. Journal of the American Statistical Association*, 107, 12 2012. doi: 10.1080/01621459.2012.703874.