Causal Inference in Layman's Terms

Maarten Vonk

August 12, 2021

Contents

1 Fundamental Problem of Causal Inference

On January 20th 2021 the Netherlands initiated a curfew to counter the Covid-19 contagious cases. The idea is that when people did not go out after 9pm, there would be less interaction and consequentially less spreading of the disease. Indeed the curfew coincided with a reduction of positive corona cases, but the question remains: to what degree was this due to the curfew? At the same time people got vaccinated and the weather got better, all variables that play a role in reducing spreading of the virus. A legitimate question then arises: what would the number of contagious have been if we did not have a curfew?

We can compare the effect of our curfew compared to similar countries without a curfew, but these countries have different cultures, temperatures or vaccination degrees, so even if we did that, this would not be a fair comparison. This brings us to the fundamental problem of causal inference: *individual causal effects can never be known.* In causal inference, this unknown object is called counterfactual. So the outcome of Covid-19 cases in the case of the lockdown that we know is denoted by Y_{1i} where 1 denote we are considering lockdown and i is this particular case (e.g. Netherlands). Similarly we Y_{0i} to be the Covid-19 outcome cases if we did not implement a lockdown. Because we did have a lockdown, the latter is unknown and hence the individual causal effect of implementing a lockdown $Y_{1i} - Y_{0i}$ is not know.

Even though we cannot say anything about individual causal effects, we can gather data about all lockdowns of the entire world together with their

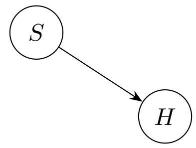

Figure 1: Illustration direct causal effect

	Sleeping with shoes	**Not Sleeping with Shoes**
Headache	80	300
No Headache	20	1200

temperatures, vaccination degrees and of course impact on positive Covid-19 cases to say something about average treatment effects. The average treatment effect is of course the mean of the individual treatment effects.

$$\frac{1}{n}\sum_i Y_{1i} - Y_{0i} = \mathbb{E}[Y_1 - Y_0]$$

We now consider various cases where we can identify and estimate causal effects. For the estimation in Python using DoWhy, look at this notebook and find out that the results in this presentation correspond to the causal estimators in the notebook.

2 Average Causal Effect in Naive Way.

On this internet blog there are discussions debating whether shoes can play a part in causes of headaches or migraines. The idea behind this is that headaches are caused by muscular tension in the neck, which are triggered by muscular tension in the feet. This muscular tension can arise when wearing certain shoes. We want to tackle this problem causally, but we simplify it a little bit: let for the sake of simplicity say that there are only two variables. We have collected data about people sleeping with shoes on and people waking up with a headache and we assume that there is a direct causal effect of sleeping with your shoes on S waking up with a headache H.

The data can be summarized in the following table. In this case, extracting the causal effect is easy, because the way we set up the graph implies that we are not suffering from the fundamental problem of causal inference: unlike the lockdown example every observation is identical hence the potential outcomes are interchangeable. That means that if in one observation someone slept with shoes on and we want to know how they wake up when they did not sleep with shoes on, we can just fill it with another observation who did sleep with shoes on: In order to calculate the average causal effect we just look at the percentage of who got a headache given that they are sleeping with shoes on and subtract the percentage of people who got a headache given they did not sleep with shoes on:

$$\mathbb{E}[H_1 - H_0] = P(H=1 \mid S=1) - P(H=1 \mid S=1) = \frac{80}{100} - \frac{300}{1500} = 0.8 - 0.2 = 0.6.$$

We call this the average causal effect in this case.There is however one underlying assumption when calculating this causal effect (except for the fact that we are only dealing with two variables).

Assumption 1 ***Stable Unit Treatment Value Assumption.*** *The observation on one unit should be unaffected by the particular assignment of treatments to the other units.*

So in our example this means that a person waking up with a headache does not depend on whether other people slept with shoes on. This seems a fair assumption in this case, but in general this is not true. Think about people getting vaccinated and about your probabilities to get infected with Covid-19. If I get vaccinated, this affects your chance of being infected with Covid-19. When deviating from this assumption, causal inference becomes really complex really fast.

2.1 Causal Inference with Mediator

We can expand the above example with another variable. On the website I learned that people with shoes on can suffer from cramps during the night and these cramps cause the headache consequentially. In this example we restrict ourselves by thinking that these cramps can only be caused by sleeping with shoes on, but sleeping with shoes on also causes other ways of waking up with headaches. In this example we then consider the causal relations as in the Figure 2 and the table containing the same data but know split up in people also having cramps:

We can calculate the causal effect via the cramps X by looking only at the percentage of people having a headache, given they had cramps and slept with shoes times probability they have cramps given they have shoes on and subtract the people with headaches given they had cramps and did not sleep with shoes

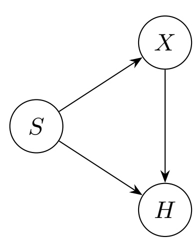

Figure 2: Illustration causal effect with Mediator

	Shoes	No Shoes
Headache and Cramp	70	260
Headache, but no Cramp	10	40
No Headache and Cramp	5	50
No Headache and no Cramp	15	1150

times probability they have cramps given they did not have shoes on:

$$\begin{aligned}\text{Causal effect via cramp: } & P(H=1 \mid S=1, X=1)P(X=1 \mid S=1) \\ & - P(H=1 \mid S=0, X=1)P(X=1 \mid S=0) \\ & = \frac{70}{75}\frac{75}{100} - \frac{260}{310}\frac{310}{1500} = \frac{79}{150}\end{aligned}$$

We can then calculate the direct effect of shoes on waking up with headache by considering the cases where people did not have cramp:

$$\begin{aligned}\text{Causal effect not via cramp: } & P(H=1 \mid S=1, X=0)P(X=0 \mid S=1) \\ & - P(H=1 \mid S=0, X=0)P(X=0 \mid S=0) \\ & = \frac{10}{25}\frac{25}{100} - \frac{40}{1190}\frac{1190}{1500} = \frac{11}{150}\end{aligned}$$

Adding the two together then yields:

$$\begin{aligned}\text{Total Causal Effect: } & = \sum_X P(H=1 \mid S=1, X)P(X \mid S=1) \\ - \sum_X P(H=1 \mid S=0, X)P(X \mid S=0) & = \frac{79}{150} + \frac{11}{150} = \frac{3}{5}.\end{aligned}$$

We can derive two conclusions from this result. First is that the causal effect via cramps is far larger than the direct causal effect of sleeping with shoes on. This is very logical if we look at the data. The second conclusion is that the solution is the exact same as the previous chapter. This is also logical once we observe that the causal graph does not allow other variables affect either cramp or waking up with a headache.

2.2 Causal effect with Treatment Assignment Mechanism

We can now modify this problem a little bit to arrive at another causal graph that contains *treatment assignment*: suppose we are not dealing with a mediating value like cramps but we are dealing with a pre-treatment variable: cold feet. We now consider the case that people with cold feet can cause people to sleep with their shoes on, but we also assume that having cold feet does not directly affect waking up with headaches apart from maybe via sleeping with shoes on. We consider the same data as before, but now instead of cramps we deal with cold feet X and we have causal graph of Figure 5.

	Shoes	No Shoes
Headache and Cold Feet	70	260
Headache, but no Cold Feet	10	40
No Headache and Cold Feet	5	50
No Headache and no Cold Feet	15	1150

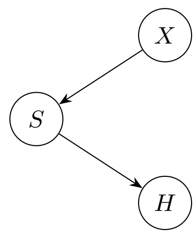

Figure 3: Illustration causal effect with Treatment Assignment Mechanism

Because cold feet does not affect waking up with a headache directly, we can easily calculate the causal effect of sleeping with shoes on waking up with headaches by breaking the relation between cold feet and sleeping with shoes on or, in other words, just ignoring the cold feet:

$$\text{Average Causal Effect:} \; = P(H = 1 \mid S = 1) - P(H \mid S = 0) = \frac{79}{150} + \frac{11}{150} = \frac{3}{5}.$$

The result of the above causal model is relatively trivial, but it is considered because it encompasses one of the most central assumption within causal inference:

Assumption 2 ***Ignorabilty Assumption.*** *The potential outcomes are independent of the treatment assignment.*

In the example, the treatment is considered sleeping with shoes on S, the treatment assignment is cold feet X and the potential outcome is headache X and we specifically assumed that headache is independent of cold feet, hence we can *ignore* cold feet. Like the SUTVA assumption, when we let this assumption go, causal inference (calculating causal effects) becomes more complex, but unlike the previous example, this still becomes manageable, thanks to the Do-calculus of Judea Pearl. But first we look at another case where this assumption *is* satisfied.

2.3 Conditional Causal Effects

We now consider a case that also satisfies the ignorability assumption, but with a different kind of causal graph structure. Imagine that we now not consider people with a cold feet, but people with migraine history X. For the sake of this example we assume that migraine history has a causal effect on waking up with

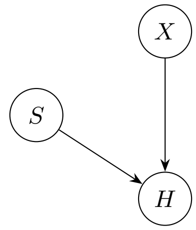

Figure 4: Illustration conditional causal effect.

headaches, but has no relation with sleeping with shoes on. The new causal graph structure that then arises is as depicted in Figure 4.

In this case, it does not make sense to talk about average causal effect of S on H, since this depends on the value of X. This is why we better talk about causal effects conditional on X. Suppose we want to know what the average causal effects are of S on H conditional on $X = 1$ Then we look at the percentage of people having a headache that slept with shoes on and had migraines compared to the percentage of people not having a headache that slept with shoes on and had migraines. Mathematically:

$$\begin{aligned} P(H = 1 \mid S = 1, X = 1) - P(H = 1 \mid S = 0, X = 1) \\ = \frac{70}{75} - \frac{260}{310} = \frac{44}{465}. \end{aligned}$$

We can do the exact same thing for people without migraines:

$$\begin{aligned} P(H = 1 \mid S = 1, X = 0) - P(H = 1 \mid S = 0, X = 0) \\ = \frac{10}{25} - \frac{40}{1150} = \frac{42}{115}. \end{aligned}$$

2.4 Causal Inference with Confounding

Suppose now that we have a similar causal problem, but instead of taking into account cold feet, we replace that variable X with heavily drinking the night before. Of course, heavy drinking not only affects the probability of sleeping with shoes on, but also the probability of waking up with headaches, see Figure 5 for the causal effects. In this case we cannot just calculate the causal effect by ignoring the variable X like we did in the previous example, cause if we did, we would also ignore the fact that X affects H and we would completely ascribe that causal affect to S.

We do not satisfy the ignorability assumption anymore, but applying Pearls do-operator for intervention, we can still calculate causal effects. In order to break the relation between X and H we intervene on sleeping with shoes on (not on drinking). This intervention is to be interpreted as breaking the relation between X and H. You can imagine this as sneaking in the middle of the night

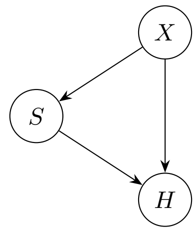

Figure 5: Illustration causal effect with Confounding

to randomly putting shoes back on or off people regardless of their drinking. In this way, to calculate the causal effect of S on H we make sure there are as many drunk as sober people with their shoes on, so any causal effect we find, must be due to sleeping with shoes on S and cannot be due to drinking the night before. The mathematics required to prove that in this case causal effects can be calculated is tedious and therefore omitted here.

In order to calculate the causal effect of S on H we need to know the probability of drinking the night before.

$$\begin{aligned} P(X=1) &= \frac{70+260+5+50}{1600} = \frac{39}{160} \\ P(X=0) &= \frac{121}{160} \end{aligned}$$

Then we can start calculating causal effects in a similar way as before. Note that in contrast to the conditional average treatment effect, we now multiply by the probability that someone has been drinking the night before.

$$\begin{aligned} &\mathbb{E}(H \mid do(S=1)) - \mathbb{E}(H \mid do(S=0)) \\ &\qquad = \sum_x P(H=1 \mid S=1, x)P(x) - \sum_x P(H=1 \mid S=0, x)P(x) \\ &\qquad = (\frac{70}{75}\frac{39}{160} + \frac{10}{25}\frac{121}{160}) - (\frac{260}{310}\frac{39}{160} + \frac{40}{1190}\frac{121}{160}) \\ &\qquad \approx 0.3. \end{aligned}$$

Of course the solution is not surprising: it is mostly due to heavy drinking that one can wakes up with headaches the day after. This is the famous *back-door theorem.* In order to fully grasp the causal effect the relation of S on H we require a lot of observations. For if there were only a couple observations, this relation could have been coincidental. Therefore we require enough *statistical power.*

2.5 Policy Intervention

The example of the previous chapter is the most trivial application of the do-calculus of interventions to calculate causal effects. Pearl now has generalized this idea, so that in the context of many variables, we can identify whether we can calculate causal effects.

We can also modify our interests a little bit. Suppose we are not actually interested in causal effect, but we are interested in minimizing the probability that we wake-up with a headache. What policy do we need to adopt? Should we be cutting down on the drinking, or maybe stop sleeping with shoes on, or a combination of both? This is the policy intervention model.

The above policy is actually trivial, since we have calculated the causal effect of sleeping with shoes on and we can do the same for drinking. However, in this case we are only considering binary variables, you either drink or you don't. But we can also consider continuous cases taking into account how much you actually drink or cases where there are multiple sorts of shoes or many other variables included. It often happens that the quantity where we are after, the intervention distribution is intractable. We cannot calculate this exactly, but we need to approximate this. This is where the applied mathematics kicks in.