≡ ⚙️ DoWhy   🔍 ⊟

🏠 > **User Guide** > ⋯ > **Estimating Causal Effects** > **Estimating...**

Estimating average causal effect using GCM

One of the most common causal questions is how much does a certain target quantity differ under two different interventions/treatments. This is also known as average treatment effect (ATE) or, more generally, average causal effect (ACE). The simplest form is the comparison of two treatments, i.e. what is the difference of my target quantity on average given treatment A vs treatment B. For instance, do patients treated with a certain medicine ($T := 1$) recover faster than patients who were not treated at all ($T := 0$). The ACE API allows to estimate such differences in a target node, i.e. it estimates the quantity $\mathbb{E}[Y|\text{do}(T := A)] - \mathbb{E}[Y|\text{do}(T := B)]$

How to use it

Lets generate some data with an obvious impact of a treatment.

```
>>> import networkx as nx, numpy as np, pandas as pd
>>> import dowhy.gcm as gcm
>>> X0 = np.random.normal(0, 0.2, 1000)
>>> T = (X0 > 0).astype(float)
>>> X1 = np.random.normal(0, 0.2, 1000) + 1.5 * T
>>> Y = X1 + np.random.normal(0, 0.1, 1000)
>>> data = pd.DataFrame(dict(T=T, X0=X0, X1=X1, Y=Y))
```

Here, we see that T is binary and adds 1.5 to Y if it is 1 and 0 otherwise. As usual, lets model the cause-effect relationships and fit it on the data:

```
>>> causal_model = gcm.ProbabilisticCausalModel(nx.DiGraph([('X0', 'T'), ('T', 'X1
>>> gcm.auto.assign_causal_mechanisms(causal_model, data)
>>> gcm.fit(causal_model, data)
```

Now we are ready to answer the question: "What is the causal effect of setting $T := 1$ vs $T := 0$

Skip to main content

```
>>> gcm.average_causal_effect(causal_model,
>>>                                   'Y',
>>>                                   interventions_alternative={'T': lambda x: 1},
>>>                                   interventions_reference={'T': lambda x: 0},
>>>                                   num_samples_to_draw=1000)
1.5025054682995396
```

The average effect is ~1.5, which coincides with our data generation process. Since the method expects an dictionary with interventions, we can also intervene on multiple nodes and/or specify more complex interventions.

Note that although it seems difficult to correctly specify the causal graph in practice, it often suffices to specify a graph with the correct causal order. This is, as long as there are no anticausal relationships, adding too many edges from upstream nodes to a downstream node would still provide reasonable results when estimating causal effects. In the example above, we get the same result if we add the edge $X0 \rightarrow Y$ and $T \rightarrow Y$:

```
>>> causal_model.graph.add_edge('X0', 'Y')
>>> causal_model.graph.add_edge('T', 'Y')
>>> gcm.auto.assign_causal_mechanisms(causal_model, data, override_models=True)
>>> gcm.fit(causal_model, data)
>>> gcm.average_causal_effect(causal_model,
>>>                                   'Y',
>>>                                   interventions_alternative={'T': lambda x: 1},
>>>                                   interventions_reference={'T': lambda x: 0},
>>>                                   num_samples_to_draw=1000)
1.509062353057525
```

To further account for potential interactions between root nodes that were not modeled, we can also pass in observational data instead of generating new ones:

```
>>> gcm.average_causal_effect(causal_model,
>>>                                   'Y',
>>>                                   interventions_alternative={'T': lambda x: 1},
>>>                                   interventions_reference={'T': lambda x: 0},
>>>                                   observed_data=data)
1.4990885925844586
```

Related example notebooks

[Skip to main content](#)

- [Impact of 401(k) eligibility on net financial assets](#)

Understanding the method

Estimating the average causal effect is straightforward seeing that this only requires to compare the two expectations of a target node based on samples from their respective interventional distribution. This is, we can boil down the ACE estimation to the following steps:

1. Draw samples from the interventional distribution of Y under treatment A.
2. Draw samples from the interventional distribution of Y under treatment B.
3. Compute their respective means.
4. Take the differences of the means. This is, $\mathbb{E}[Y|\mathrm{do}(T := A)] - \mathbb{E}[Y|\mathrm{do}(T := B)]$, where we do not need to restrict the type of interventions or variables we want to intervene on.

Previous
‹ **Estimating conditional average causal effect**

Next
Quantify Causal Influence ›

© Copyright 2022, PyWhy contributors.

Created using [Sphinx](#) 7.1.2.

Built with the [PyData Sphinx Theme](#) 0.14.4.