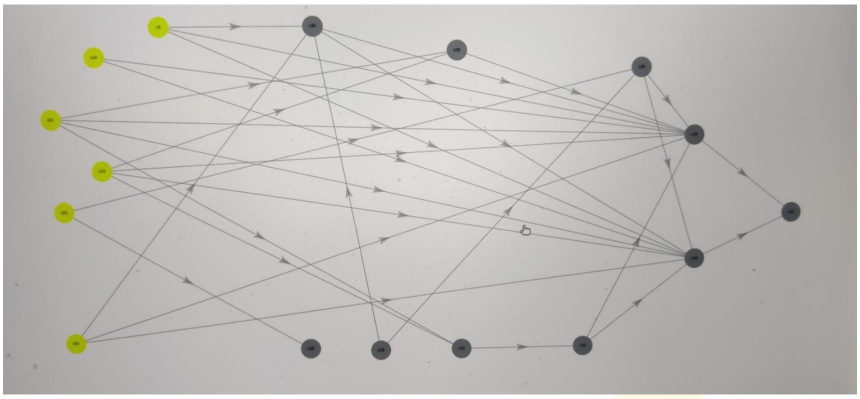
Exercises causality researcher position - Allos AI

November 14, 2025

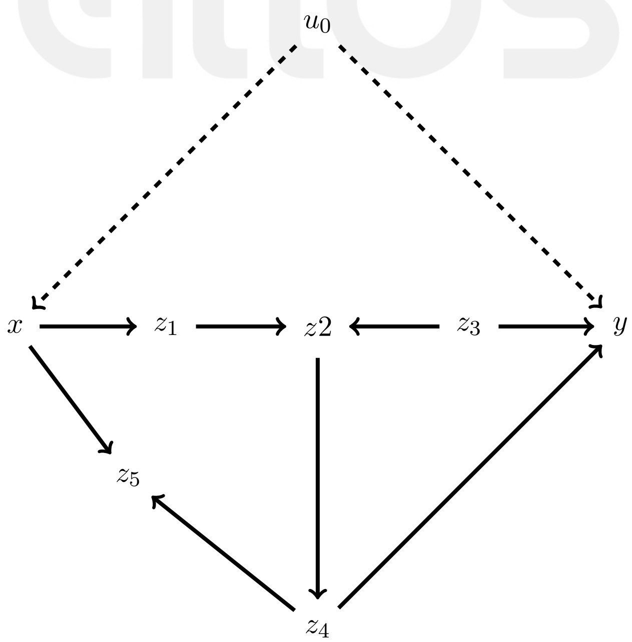
Exercise 0

Given the graph below, with set of nodes V, explain with a simple argument (without computing anything) why we can always say $P(y|\hat{x}, w) = P(y|x, w)$ for all $y, w \in V \setminus \{x\}$ and $x \in \{\text{yellow nodes}\}$



Exercise 1

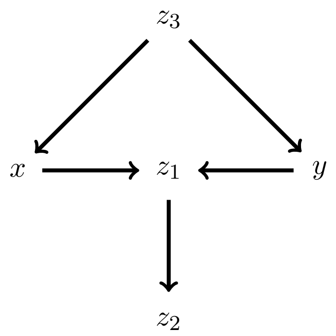
Consider the following graph:



Compute *all* answers to the interventional query $P(y|\hat{x}, z_2)$ (using the rules of do-calculus)

Exercise 2

Consider the following simple graph:



- Try to use the rules of do-calculus to answer the interventional query $P(y|\hat{x}, z_1)$ (or equivalently $P(y|\hat{x}, z_2)$). What is your result?
- how can we re-write the query $P(y|\hat{x}, z_1)$ to compute it?
- What is the result?

Exercise 3

Consider the graph of exercise 2. Re-write the product of probability distributions:

$$P_0 = \sum_{z3} P(y, z_1|x, z_3)P(z_3) \tag{1}$$

in terms of single-outcome probability distributions, using the product rule.

- Find *all* the equivalent ways to re-write the above product (1) s and simplify the final results using D-separation. Let's call these new products P_i
- Build datasets with $10^3, 10^4, 10^5, 10^6$ elements, containing all the variables in the graph as columns. The values of the columns can be generated from a discrete or continuous range of values, by following a specific model (of your choice) for the variables relationships encoded in the graph through the parent(s)/child relationships. (Example: From the graph, we know that $x = f(z_3)$. If z_3 is a random variable with discrete values $(0, 1)$, f could be an exponential function $x = \exp^{z_3}$ so $x \in (0, \exp^1)$; it could be a discrete function between two binary variables x and z_3 with pre-assigned values for the conditional probability: $P(x = 0|z_3 = 0) = x$; $P(x = 1|z_3 = 0) = 1 - x$; $P(x = 0|z_3 = 1) = y$; $P(x = 1|z_3 = 1) = 1 - y$ (with $0 < x, y < 1$)
- Plot the value of P_0 in 1 and all P_i's you found in the previous question, for all datasets
- Check that the results are, within statistical margin, overlapping.
- Try to estimate an "error" between the values plotted, and plot how the error changes with varying dataset size
- Finally, re-do the same exercise, but this time generate a random dataset, without using the causal graph as a guide for the parent/child relationships. Plot the values of P_0 and all P_i's now. What are thee results? Is the error changing in the same fashion as the dataset size changes? Why?

Exercise 4

Write a simple python script that computes, for probability objects to be defined, the (inverse) product rule:

$$P(x_0|x_1, x_2, z_3, \dots, z_n)P(x_1|x_2, x_3, \dots, z_n) \cdots P(z_n) = P(x_0, x_1, \dots, x_n) \tag{2}$$

Also, try to implement simplifications of the type:

$$\frac{\sum_{z_3} P(y, x|z_1, z_2, z_3, z_4)P(z_1, z_2, z_3)}{\sum_{z_3,x,y} P(y, x|z_1, z_2, z_3, z_4)P(z_1, z_2, z_3)} = \sum_{z_3} P(y, x|z_1, z_2, z_3, z_4)P(z_3|z_1, z_2) \tag{3}$$

Hint: whatever way you want to define the probability object, it needs to be a recursive object, since a single probability can contain a multitude of probability distributions summed over variables.