Algorithms and solutions to exercises

Ivano Lodato

February 15, 2024

The rules of do-calculus are equalities valid (and applicable in both directions) whenever certain conditional independecies exist in a mutilated version of the causal graph. The whole point of conditioning is to force independencies between variables. In short, d-separation on a graph, or conditional independence in probability distributions, is the name of the game.

Say we have an interventional query, $P(Y|\hat{X}, Z)$, where Y, X, Z are sets of nodes/variables. The set Z may be empty, but the set Y and X must contain some elements, otherwise we would have no outcome Y (the probability distribution is defined over nothing) or intervention(s) X (the query would be non-interventional, hence trivial). The procedure to solve a query of this kind on the graph G is explained below, with images and counterexamples.

Solve interventional queries - single outcome

Start by fixing one of the interventions $x \in X$. It can be proven easily that the final result will not depend on the order of interventions.

1. Find *all* the active paths between an intervention $x \in X$ and the outcome $y \in Y$. Call the set of these paths $\mathcal{P}$
   If a path from $x \in X$ to $y \in Y$ is inactive, it either means the flow of information is blocked by an observation of a fork or chain node $z \in Z$, or an unobserved collider c ($c \notin Z$)

2. Divide $\mathcal{P}$ into backdoor ($\mathcal{P}_B$) and frontdoor ($\mathcal{P}_F$) paths

   (a) If $\mathcal{P}_B = \emptyset$ and $\mathcal{P}_F = \emptyset$, try to use rule 3 of do-calculus (next steps to be added)

   (b) If the set $\mathcal{P}_B = \emptyset$ is empty (there is no active backdoor path), and $\mathcal{P}_F \neq \emptyset$, use rule 2 of do-calculus

   (c) If the set $\mathcal{P}_B \neq \emptyset$ is empty (there is no active backdoor path), and $\mathcal{P}_F = \emptyset$, try to use rule 3 of do-calculus (next steps to be added)

   (d) $\mathcal{P}_B \neq \emptyset$ and $\mathcal{P}_F \neq \emptyset$, conditioning (observing) some variable is necessary.
   If we have both backdoor and frontdoor paths active, then we cannot use either rule 2 or rule 3 (or no sequence of both) of do-calculus to identify the interventional query.

   Check among the active paths which are *unblockable*: i.e. (1) conditioning (i.e. observing a fork or a chain) on that active path is not possible, because either there are no fork/chain nodes on that path (paths of length 1) or they are unobserved; or (2) a collider on a path is observed, and there are no fork/chain nodes on that path (or they are unobserved) that we can condition the probability distribution on).
   Call the sets of unblockable paths $\mathcal{P}_B^U$ (backdoor) and $\mathcal{P}_F^U$ (frontdoor)

      i. if there are no unblockable backdoor paths $\mathcal{P}_B^U = \emptyset$, then we condition on fork/chain nodes on those paths and block them (see below *). We are left with some active frontdoor paths (unblockable and blockable): use rule 2 of do-calculus

ii. if there are no unblockable frontdoor paths, $\mathcal{P}_F^U = \emptyset$ then we condition on fork/chain nodes on those paths, and block them (see below *). We are left with some active backdoor paths (unblockable and blockable): try to use rule 3 of do-calculus (next steps to be added)

iii. if there are no unblockable (backdoor or frontdoor) paths, then decide which paths you want to block (see below *), and use the corresponding rule of do-calculus to (try to) simplify the intervention x

iv. if there is at least one frontdoor and one backdoor unblockable paths, then the query is **non-identifiable**

Repeat this procedure for all interventions in all probability distributions of the query (remember, after conditioning the joint probability distributions is expanded and new probability distributions appear with interventions as conditions) until they are either transformed into observations (via rule 2) or erased from the query (via rule 3). Finally, check if rule 1 can be applied on any observations - as the straightforward application of rule 2 (whenever only front-door paths are active) may lead to an superfluous observation, i.e. we could have deleted the intervention via rule 3 instead of using rule 2 and then rule 1 on it.
If even one intervention in one probability distribution cannot be transformed or deleted after conditioning, then the graph is non-identifiable.
[**Exercise 3: Check and identify, in the solutions of Exercise 1 below, which intervention was deleted via rule 3 and which observation was deleted via rule 1. Pay attention to how and when rule 3 is used.**].

[*** finding the correct variables to condition on is not trivial. Conditioning on a variable z on a path between x and y to block it, may activate some other path on which that variable was a collider. So now we may have to block that path (See worked out example below.)**]

1 Solve interventional queries - many outcomes

The above procedure can be generalized, by still considering one intervention $x \in X$ at a time, but there are now many outcomes $y_1, \ldots, y_n \in Y$. Step 1 of the algorithm above gets modified into:

① Find *all* active paths between x and all the outcomes in Y. Call the paths from x to each outcome y_i $\mathcal{P}^i$.

② Divide each of these path sets into backdoor $\mathcal{P}_B^i$ and frontdoor $\mathcal{P}_F^i$ active paths.

- ⓐ If we have $\mathcal{P}_B^i = \emptyset$ and $\mathcal{P}_F^i = \emptyset$ for all outcomes y_i, then we try to use rule 3 directly on x
- ⓑ If we have $\mathcal{P}_B^i = \emptyset$ but $\mathcal{P}_F^i \neq \emptyset$, then we can use rule 2 on x
- ⓒ If we have $\mathcal{P}_F^i = \emptyset$ but $\mathcal{P}_B^i \neq \emptyset$, then try to use rule 3 on x
- ⓓ If $\mathcal{P}_F^i \neq \emptyset$ and $\mathcal{P}_B^i \neq \emptyset$, conditioning on some variable is necessary.
  Check, for each $\mathcal{P}^i$ which are the unblockable paths. Call them $P_B^{i,U}$ and $P_F^{i,U}$. For each outcome y_i, any of these two sets may be empty, so there are many possible such combinations. Let us go in order:
  - ⓘ Suppose there is a specific outcome y_i for which exist at least one unblockable active backdoor and one unblockable active backdoor. In this case, we can already say that the query is non-identifiable
  - ⓘⓘ Suppose for no outcome $y_i \in Y$, there exist both an unblockable backdoor and frontdoor paths from x. This implies that, for each outcome y_i there may be at most one type of paths which are unblockable, either $\mathcal{P}_B^{i,U} = \emptyset$ or $\mathcal{P}_F^{i,U} = \emptyset$. If for all outcomes, only

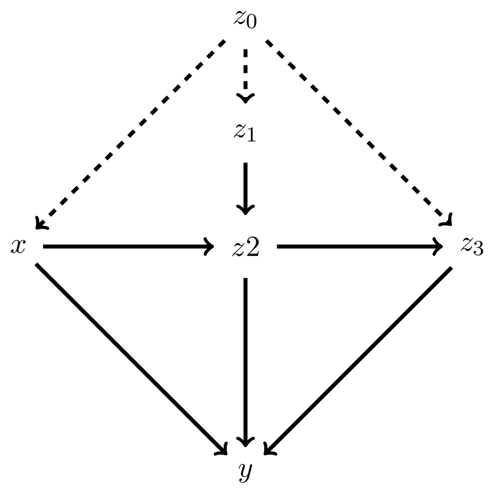


Figure 1: (simplified) Causal graph presented in Figure 3.1 of the book "Causality: models,reasoning, and inference"

unblockable frontdoor paths exist, we condition on forks/chain on the backdoor paths from x to all outcomes y_i, then apply rule 2 on the intervention

(iii) If for all outcomes, only unblockable backdoor paths exist, we condition on forks/chain on the frontdoor paths from x to all outcomes y_i, then try to apply rule 3 on the intervention

(iv) If for k outcomes the unblockable paths are frontdoor, and for $k - n$ outcomes the unblockable paths are backdoor, we find all variables which block as many of the active backdoor paths connecting to the k outcomes and, at the same time, as many of the active frontdoor paths connecting to the $n - k$ outcomes (pay attention to the observation * above). Eventually, we will need to decompose the full joint probability distribution $P(y_1, \ldots, y_n|\hat{x})$ into the product of many smaller joint probability distributions, as dictated by the blocking variables in the blocked paths. Then start from step 1 again.

Let us work out an exercise together. We try to answer the query $P(y|\hat{x})$ int he graph of Figure 1. We use the steps for one outcome y. First we identify (in)active paths between x and y.

Active frontdoor paths

$$x \to y \quad ; \tag{1}$$
$$x \to z_2 \to y \quad ; \tag{2}$$
$$x \to z_2 \to z_3 \to y \quad ; \tag{3}$$

Inactive frontdoor paths

$$x \to z_2 \leftarrow z_1 \leftarrow z_0 \to z_3 \to y \quad ; \tag{4}$$

Active backdoor paths

$$x \leftarrow z_0 \to z_3 \to y \quad ; \tag{5}$$
$$x \leftarrow z_0 \to z_1 \to z_2 \to y \quad ; \tag{6}$$
$$x \leftarrow z_0 \to z_1 \to z_2 \to z_3 \to y \quad ; \tag{7}$$

Inactive backdoor paths

$$x \leftarrow z_0 \to z_3 \leftarrow z_2 \to y \quad ; \tag{8}$$

Among the active frontdoor paths, one is unblockable $x \to y$. The other two paths can be blocked by (z_2) and (z_2 and/or z_3) respectively. Note also that observing z_2 activates the inactive frontdoor path (z_2 is a collider there), but observing also z_3 blocks (keep inactive) this path.
Among the backdoor paths, none is unblockable: the active backdoor paths can be blocked respectively by (z_3), (z_1 and/or z_2), (z_1 and/or z_2 and/or z_3). Note that, again, observing z_3 activates an inactive frontdoor path, but observing also z_2 keeps the path inactive. Hence, by looking at the intersection of the sets of variables blocking the backdoor paths, we immediately obtain (z_2,z_3) (z_3 is the only variable that can block the path $x \leftarrow z_0 \to z_3 \to y$, once z_3 is conditioned on, then we

need to condition on z_2 to keep the backdoor path $x \leftarrow z_0 \rightarrow z_3 \leftarrow z_3 \rightarrow y$ inactive. Finally, we note that all other backdoor paths are blocked by (z_2, z_3), and no other spurious path is activated by these variables.
The conditioning goes as follows:

$$P(y|\hat{x}) = \sum_{z_2,z_3} P(y|\hat{x}, z_2, z_3)P(z_2, z_3|\hat{x}) = \sum_{z_2,z_3} P(y|x, z_2, z_3)P(z_2, z_3|\hat{x}) \tag{9}$$

where we applied rule 2 on the first probability distribution in the second equality. We have another probability distribution to analyse, $P(z_2, z_3|\hat{x})$ (see the steps to follow for interventional queries with many outcomes). Again, we find all active paths from x to both z_2 and z_3.

Active frontdoor paths $x \rightarrow \dots z_2$

$$x \rightarrow z_2 \quad ; \tag{10}$$

Inactive frontdoor paths $x \rightarrow \dots z_2$

$$x \rightarrow y \leftarrow z_2 \quad ; \tag{11}$$
$$x \rightarrow y \leftarrow z_3 \leftarrow z_2 \quad ; \tag{12}$$
$$x \rightarrow y \leftarrow z_3 \leftarrow z_0 \rightarrow z_1 \rightarrow z_2 \quad ; \tag{13}$$

Active backdoor paths $x \leftarrow \dots z_2$

$$x \leftarrow z_0 \rightarrow z_1 \rightarrow z_2 \quad ; \tag{14}$$

Inactive backdoor paths $x \leftarrow \dots z_2$

$$x \leftarrow z_0 \rightarrow z_3 \leftarrow z_2 \quad ; \tag{15}$$
$$x \leftarrow z_0 \rightarrow z_3 \rightarrow y \leftarrow z_2 \quad ; \tag{16}$$
$$x \rightarrow y \leftarrow z_2 \leftarrow z_1 \leftarrow z_0 \rightarrow z_3 \tag{22}$$

Active frontdoor paths $x \rightarrow \dots z_3$

$$x \rightarrow z_2 \rightarrow z_3 \quad ; \tag{17}$$

Inactive frontdoor paths $x \rightarrow \dots z_3$

$$x \rightarrow z_2 \leftarrow z_1 \leftarrow z_0 \rightarrow z_3 \quad ; \tag{18}$$
$$x \rightarrow y \leftarrow z_3 \quad ; \tag{19}$$
$$x \rightarrow z_2 \rightarrow y \leftarrow z_3 \quad ; \tag{20}$$
$$x \rightarrow y \leftarrow z_2 \rightarrow z_3 \quad ; \tag{21}$$

Active backdoor paths $x \leftarrow \dots z_3$;

$$x \leftarrow z_0 \rightarrow z_3 \quad ; \tag{23}$$
$$x \leftarrow z_0 \rightarrow z_1 \rightarrow z_2 \rightarrow z_3 \quad ; \tag{24}$$

Inactive backdoor paths $x \leftarrow \dots z_3$

$$x \leftarrow z_0 \rightarrow z_1 \rightarrow z_2 \rightarrow y \leftarrow z_3 \quad ; \tag{25}$$

Now there is an unblockable frontdoor path from $x \rightarrow z_2$. The active backdoor path to z_2 can be blocked by conditioning on z_1 (and this will not activate any other paths). On the other hand, there is an unblockable backdoor path to z_3, $x \leftarrow z_0 \rightarrow z_3$. The directed frontdoor path $x \rightarrow z_2 \rightarrow z_3$ can be blocked only by conditioning on z_2. However, this activates the path $x \rightarrow z_2 \leftarrow z_1 \leftarrow z_0 \rightarrow z_3$ which we can block only by observing z_1. So conditioning the joint probability distribution on z_1 and z_2, we block at the same time the active backdoor path to z_2 $x \leftarrow z_0 \rightarrow z_1 \rightarrow z_2$, the frontdoor path $x \rightarrow z_2 \rightarrow z_3$ (and frontdoor path, activated by observing z_2 $x \rightarrow z_2 \leftarrow z_1 \leftarrow z_0 \rightarrow z_3$). It is also to be noticed that one of the blocking variable z_2, in this case, is also an outcome. So *we first condition on all blocking variables available which are not outcomes*, i.e. z_1:

$$P(z_2, z_3|\hat{x}) = \sum_{z_1} P(z_2, z_3|\hat{x}, z_1)P(z_1|\hat{x}) \tag{26}$$

Now, since we have an unblockable backdoor path from x to z_3 and an unblockable frontdoor path from x to z_2, we cannot solve the intervention on the joint probability distributions by using any rules of do-calculus (or sequence thereof). Instead, we factorize the join probability distribution. Now, there is $n!$ way to factorize a joint probability distribution with n outcomes. In this case, we see immediately that z_2 is needed as an observation to block the front-door path from x to z_3, while z_3 is not needed to block any active paths from x to z_2. So we can write:

$$\sum_{z_1} P(z_2, z_3|\hat{x}, z_1)P(z_1|\hat{x}) = \sum_{z_1} P(z_3|\hat{x}, z_2, z_1)P(z_2|\hat{x}, z_1)P(z_1|\hat{x}) \tag{27}$$

Clearly, the ordering of probability distributions is related to the ancestral relation between z_2 and z_3.

The above procedure can be inverted, i.e. instead of conditioning on z_1 and observing (redundantly) z_2 and then factorizing the joint probability distribution, we can do the opposite: first we factorize the joint probability distribution, then condition on the variables needed. Now, following the ancestry relation of the variables z_2, z_3, it is easy to write:

$$P(z_2, z_3|\hat{x}) = P(z_3|\hat{x}, z_2)P(z_2|\hat{x}) \tag{28}$$

To identify the first factor, we need to note that the observation of z_2 activates a frontdoor path to z_3, which can be blocked by conditioning on z_1. There is again an unblockable backdoor path. We can then write two possible conditioning:

$$P(z_3|\hat{x}, z_2)P(z_2|\hat{x}) = \sum_{z_1} P(z_3|\hat{x}, z_2, z_1)P(z_1|\hat{x}, z_2)P(z_2|\hat{x}) \tag{29}$$

or alternatively:

$$P(z_3|\hat{x}, z_2)P(z_2|\hat{x}) = \sum_{z_1} P(z_3|\hat{x}, z_2, z_1)P(z_2|\hat{x}, z_1)P(z_1|\hat{x}) \tag{30}$$

The choice is not arbitrary: the first identity is wrong, because we are not accounting for the fact that z_1 is also needed to block paths from x to z_2, not just paths from x to z_3. Hence, we need to condition both $P(z_3|..)$ and $P(z_2|...)$ with z_1. Another way to see this is to consider the ancestry relations of all the outcomes that follow z_3 (including z_3 itself): in this case, we are manipulating the probability distribution with outcome z_3, so we need to order z_3,z_2 and z_1 (the last conditioning variable, for now) based on their ancestry.
It is clear that this alternative procedure gives equivalent results: however, the interpretation can become ambiguous and it complicates much in the way of generalizations, so we will stick with the first resolution: when considering a query with more than 1 outcome, we condition on the variables which are needed to block, at the same time, paths from x to every outcome, paying attention to conditioning variables which are outcomes in the joint probability distribution. Then we decompose the joint probability distribution in factors guided by the conditioning variables needed (i.e. ancestry relations), as we have done in the example above.
[**Exercise 4: Complete the identification of the query $P(y|\hat{x})$ from the above results. Obtain the formula presented in eq. 3.1 of the book "Causality"**]

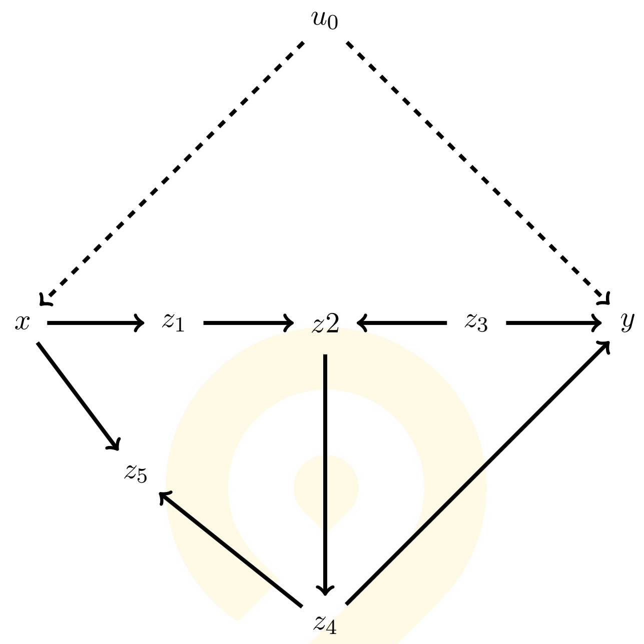
Solution to Exercise 0

Every intervention (yellow node) $\hat{x}$ in the query is connected to other nodes in the graphs *only* through front door paths, so we can always use rule 2 of do-calculus. Note however, that in some cases, we can simplify the query further, by using rule 1 of do-calculus on an observation x, after using rule 2 on the intervention $\hat{x}$. Take for instance the case (it is in fact a subgraph in the picture) in which $x \rightarrow z \leftarrow y$. If z is not observed, this path is inactive, so x is d-separated from y. Furthermore, if the query is $P(y|\hat{x})$ (hence z, descendant of x is not observed), I can directly use rule 3 of do-calculus and get rid of the intervention. An even simpler example is the case where no (active or inactive) path exist between x and an outcome y: in this case, trivially, we can use rule 3.
Note 1: if the query on the above graph would have been $P(y|\hat{x}, z)$ then the path is active and unblockable! We can still use rule 2 of do-calculus, and transform the intervention $\hat{x} \rightarrow x$, but we cannot use rule 3 anymore!!!!
Note 2: if a variable x satisfies rule 1 of do-calculus, $\hat{x}$ automatically satisfies rule 2 and rule 3. The viceversa is not always true.
In the above example, for instance, with the graph $x \rightarrow z \leftarrow y$ and z is not observed, we can apply rule 3 or apply rule 2 and then rule 1, but this is not always the case.
[**Exercise 5: Draw a counterexample causal graph, for which I can apply rule 3 on the intervention $\hat{x}$ in the query $P(y|\hat{x}, w)$, but I cannot use rule 1 on the observation x in the**

query $P(y|x,w)$. Do the same for a graph where rule 2 is applicable on $\hat{x}$ but rule 1 on x is not. Why is that so? Hint: think about the mutilations of graphs appearing in the rules of do-calculus]

Solution to Exercise 1

In the graph: the interventional query $P(y|\hat{x}, z_2)$ has 3 answers.



Blocking variables: z_3, x

$$\begin{aligned} P(y|\hat{x}, z_2) &= \sum_{z_3} P(y|\hat{x}, z_2, z_3)P(z_3|\hat{x}, z_2) \\ &= \sum_{z_3} P(y|\hat{x}, \hat{z_2}, z_3)P(z_3|\hat{x}, z_2) \\ &= \sum_{z_3} P(y|\hat{z_2}, z_3)P(z_3|\hat{x}, z_2) \\ &= \sum_{z_3,x'} \Big(P(y|\hat{z_2}, z_3, x')P(x'|\hat{z_2}, z_3)\Big)P(z_3|\hat{x}, z_2) \\ &= \sum_{z_3} P(z_3|x, z_2) \sum_{x'} P(y|z_2, z_3, x')P(x'|z_3) \\ &= \sum_{z_3} P(z_3|x, z_2) \sum_{x'} P(y|z_2, z_3, x')P(x') \end{aligned} \tag{31}$$

Blocking variables: z_3, z_1

$$\begin{aligned} P(y|\hat{x}, z_2) &= \sum_{z_3} P(y|\hat{x}, z_2, z_3)P(z_3|\hat{x}, z_2) \\ &= \sum_{z_3} P(y|\hat{x}, \hat{z_2}, z_3)P(z_3|\hat{x}, z_2) \\ &= \sum_{z_3} P(y|\hat{z_2}, z_3)P(z_3|\hat{x}, z_2) \\ &= \sum_{z_3,z_1} \Big(P(y|\hat{z_2}, z_3, z_1)P(z_1|\hat{z_2}, z_3)\Big)P(z_3|\hat{x}, z_2) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{z_3} P(z_3|x, z_2) \sum_{z_1} P(y|z_2, z_3, z_1) P(z_1|z_3) \\
&= \sum_{z_3} P(z_3|x, z_2) \sum_{z_1} P(y|z_2, z_3, z_1) P(z_1) \qquad (32)
\end{aligned}
$$

Note that the above results are basically identical, where instead of observing x, we observe its descendant z_1 (the paths in the graph are such that this is the only distinction between the two results, but this is not necessarily a general feature).

Blocking variables: z_1, x

$$
\begin{aligned}
P(y|\hat{x}, z_2) &= \sum_{z_1} P(y|\hat{x}, z_2, z_1) P(z_1|\hat{x}, z_2) \\
&= \sum_{z_1} P(y|\hat{x}, \hat{z_1}, z_2) P(z_1|\hat{x}, z_2) \\
&= \sum_{z_1} P(y|\hat{z_1}, z_2) P(z_1|\hat{x}, z_2) \\
&= \sum_{z_1, x'} \Big(P(y|\hat{z_1}, z_2, x') P(x'|\hat{z_1}, z_2) \Big) P(z_1|\hat{x}, z_2) \\
&= \sum_{z_1} P(z_1|x, z_2) \sum_{x'} P(y|z_1, z_2, x') P(x'|\hat{z_1}, \hat{z_2}) \\
&= \sum_{z_1} P(z_1|x, z_2) \sum_{x'} P(y|z_1, z_2, x') P(x'|\hat{z_2}) \\
&= \sum_{z_1} P(z_1|x, z_2) \sum_{x'} P(y|z_1, z_2, x') P(x') \qquad (33)
\end{aligned}
$$

Solution to exercise 2

In the following simple graph:

$$
\begin{tikzcd}
& z_3 \arrow[dl] \arrow[dr] & \\
x \arrow[r] & z_1 \arrow[d] & y \arrow[l] \\
& z_2 &
\end{tikzcd}
$$

Let us consider the query: $P(y|\hat{x}, z_1)$ (the same arguments can be applied to the query $P(y|\hat{x}, z_2)$). Now:

$$
\begin{aligned}
P(y|\hat{x}, z_1) &= \sum_{z_3} P(y|\hat{x}, z_3, z_1) P(z_3|\hat{x}, z_1) \\
&= \sum_{z_3} P(y|\hat{x}, z_3, z_1) \sum_{y'} P(z_3|\hat{x}, z_1, y') P(y'|\hat{x}, z_1) \qquad (34)
\end{aligned}
$$

Note now that 1) we are conditioning on y, but y is the original outcome of our query (fishy); the last probability distribution on the right is *identical* to the original query on the left-hand side of the equation (very fishy).

[Exercise 6: show that the above equality is not actually an infinite recursion, but rather a trivial identity $P(y|\hat{x}, z_1) = P(y|\hat{x}, z_1)$!!!]

[Exercise 7: now that we have seen our usual methods do not work, how can we re-write the query $P(y|\hat{x}, z_1)$ so that it is identifiable? what is the result? Hint: remember the decomposition of the joint probability of A and B, given the condition C, into a product of two (conditional) probabilities. Invert the identity to re-write the query above. Now solve it using the steps explained above for multiple outcomes queries.]

[Exercise 8: Try to answer the query $P(y|\hat{x})$ for the graph below
Prove that there are 3 solutions to this query, all requiring to condition on z_2 and z_3

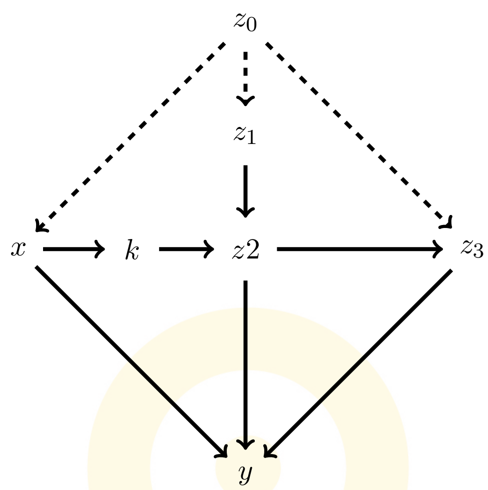


Figure 2: Slightly modified version of Figure 1

first, then (z_1, k), (z_1, x) and finally (k, x). Note in the last, case a subgraph similar to the one in Exercise 2 is analyzed, and as before, we cannot seem to be able to identify with usual methods one of the probabilities in the product. Use what you learned in exercise 6, then identify the above query when the conditioning variables are (z_2, z_3, k, x).]

2 Clarification of some algorithmic steps

Up until now we have left unspecified few delicate points of the algorithmic steps required to implement do-calculus:

a - the use of rule 3

b - selection of blocking variable for conditioning

c - types of conditioning

d - solve an interventional query with multiple outcomes

In this section we will highlight some examples and leave you to find the general conditions that a query and a graph must satisfy for the above steps to be implemented in an algorithm.

Solution to exercise 3

Going line by line in the first solution (blocking variables z_3, x):

1. conditioning on z_3 to block the front door path from x to y (preparing for the use of rule 3 on x)
2. Rule 3 cannot be applied on $\hat{x}$ in the first probability distribution (p.b.), z_2 observed descendant of $\hat{x}$. Rule 2 applied on z_2.

3. Rule 3 applied to x i the first p.b.

4. Condition on x to block the backdoor path from z_2 to y.
Note: In eq. (29)-(30) we showed a similar situation: a product of two probability distributions, and conditioning on the variable z_1 in the first p.d. has an *effect* on the second p.d., as z_1 is used has a blocking variable for the paths from x to z_3 and the paths from x to z_2.
In this case, instead, when we condition for x, we do not modify the second p.d. $P(z_3|\hat{x}, z_2)$.
[**Exercise 9: Can you try to find a way to condition differently, in a fashion similar to eq. (29)-(30) ? If you cannot, can you understand what are the obstacles? Hint: Remember that when we condition on a variable, we are just re-writing the *same* probability distribution, in a different way, a way that helps us solve interventions (by blocking paths). The appearance of other probability distributions when conditioning is *necessary* for the identity to be true. Conditioning comes from the product rule applied to a joint probability distribution, which is also a mathematical identity. Once we have used any rules of do-calculus after conditioning, the identity is not immediately evident (although one could recover it by going backwards in the application of the rules). What happens when we apply rule 3? what changes between the various probability distributions in the product?**

5. Rule 3 applied on z_2 in the p.d. $P(x|\hat{z_2}, z_3)$. There is only an open backdoor path from z_2 to x, and the observed z_3 is a parent (not a descendant) of z_2.
Rule 2 applied on the other p.d.'s on z_2 and .

6. Rule 1 applied on z_3 because it has no open paths to x (all paths are blocked by a collider - y, z_2 and z_5

Note that rule 1 is much stronger than any other rule, as no paths at all must exist for it to be applied. For the second solution, we have, line by line (blocking variables z_3, z_1):

1. conditioning on z_3 to block the front door path from x to y (preparing for the use of rule 3 on x)

2. Rule 3 cannot be applied on $\hat{x}$ in the first probability distribution (p.b.), z_2 observed descendant of $\hat{x}$. Rule 2 applied on z_2.

3. Rule 3 applied to x in the first p.b.

4. Condition on z_1 to block the backdoor path from z_2 to y.
Note: same as before, conditioning on z_1 does not affect the p.d. $P(z_3|\hat{x}, z_2)$

5. Rule 3 applied on z_2 in the p.d. $P(z_1|\hat{z_2}, z_3)$. There is only an open,unblockable backdoor path from z_2 to z_1, and the observed z_3 is a parent (not a descendant) of z_2.
Rule 2 applied on the other p.d.'s on z_2 and x .

6. Rule 1 applied on z_3 because it has no open paths to x (all paths are blocked by a collider - y, z_2 and z_5

Finally, for the last solution we have:

1. conditioning on z_1 to block the front door path from x to y (preparing for the use of rule 3 on x)

2. Rule 3 cannot be applied on $\hat{x}$ in the first probability distribution (p.b.), z_1 and z_2 observed descendants of $\hat{x}$. Rule 2 applied on z_1.
Note: compare to the second line of the first (or second) solution: there you use rule 2 on z_2 because z_3 is blocking one backdoor path to y and $\hat{x}$ (graph mutilated $G_{\overline{X}}$) eliminates the other backdoor path to y.
Exercise 10: show that rule 2 cannot be applied on z_2 in this case, but MUST be

applied on z_1. Explain also why, once z_1 has been transformed into an intervention, we can still apply rule 3 on x even though z_2 - a descendant of x in the graph G, is observed.

3. Rule 3 applied to x in the first p.b.

4. Condition on x to block the backdoor path from z_1 to y. There is no choice of blocking variables this time.
Note: same as before, conditioning on x does not affect the p.d. $P(z_1|\hat{x}, z_2)$. Same exact reasons lead to this conditioning being unique.

5. Rule 2 applied on z_2 (descendant of z_1) in the p.d. $P(x'|\hat{z_1}, z_2)$. Also Rule 2 applied on the other p.d.'s on z_1 and x .

6. Rule 3 applied on z_1 (z_2 is an intervention now).

7. Rule 3 applied on z_2 because it has no open paths to x in $G_{\overline{Z_2}}$ (frontdoor paths are blocked by the collider z_5)

[Exercise 11: in line 5. of the last solution above we used rule 2 of z_2 to apply rule 3 on z_1 (since z_2 is a descendant of z_1). Can you find a different set of rules to apply to $P(x'|\hat{z_1}, z_2)$ to simplify the query? Hint: if you have really checked the paths from z_2 to x before applying rule 2 on z_2 you should have noticed something, already]
The above solution shows already many details about conditioning and the application of rule 3 (points (c) and (a) above). We have seen for instance that whenever a descendant z of an intervention $\hat{x}$ is observed, we need to apply a rule on it to allow the use of rule 3. From what we have seen, there are only two choices: either one tries to apply rule 1 on the observation, or rule 2 is applied to make that observation into an intervention. Also, we have seen that whenever multiple descendants z_1, z_2 of $\hat{x}$ are observed, and they are all in chain $x \to z_1 \to z_2$, then it is most convenient to try to apply rule 2 on z_1, if rule 1 fails on both observations. The reasons should be clear: once z_1 is an intervention, z_2 is not a descendant of x anymore, as all incoming edges in z_1 have been erased. If instead we transform z_2 into an intervention, z_1 is still a descendant of x we need to try an apply rule 2 on.
Exercise 12: Are there any other choices for the application of rule 3, i.e. can you figure out a counter-example graph where all frontdoor paths are blocked, rule 3 cannot be applied directly because of observed descendants Z, Z cannot be deleted via rule 1 OR transformed via rule 2, but the graph is still identifiable? Hint: try first to find all types of graph which are non-identifiable by rule 3, because they have an observed descendant which cannot be transformed via rule 2 or deleted via rule 1. Then it will be easier to answer the question.
Exercise 13: what happens if the observed descendants of x are not connected by directed paths(e.g. $x \to z_1$ and $x \to z_2$ but no arrow between z_1 or z_2 - or $z_1 \leftarrow u \to z_2$ with u confounder)?
Draw a causal graph for each of the two cases and create for both an *identifiable* query where there are more than 1 observed descendants of x.
Another thing we have discovered is that, when the interventional query becomes - through conditioning on n variables- a product of $n+1$ probability distributions, when any of the rules of do-calculus are applied, the identity imposed by the product rule (i.e. conditioning) is not immediately evident anymore. In fact, in some cases, there could be two consecutive probability distributions which have different interventions or the outcome of the right one may not appear as a condition on the left one. These differences force us to define two "types" of conditioning:

- non-isolated conditioning: the condition on one p.d. modifies also (some of) the probability distributions following it (see eq. (30))
- isolated conditioning: the condition on one p.d. does not (cannot) modify the probability distributions following it (see 3 solutions to exercise 3 above - lines 4.)

Exercise 14: can you figure out the exact constraints on the probability distribution and on the graph that need to be satisfied for the isolated conditioning to be applicable? what about the non-isolated one?

Solution to exercise 4

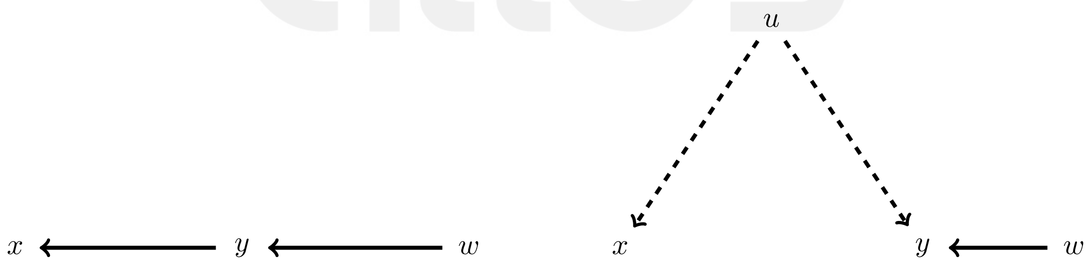
Starting from the given results, we have:

$$
\begin{aligned}
P(y|\hat{x}) &= \sum_{z_3,z_2} P(y|x,z_2,z_3) \sum_{z_1} P(z_3|\hat{x},z_2,z_1)P(z_2|\hat{x},z_1)P(z_1|\hat{x}) \\
&= \sum_{z_3,z_2} P(y|x,z_2,z_3) \sum_{z_1} P(z_3|\hat{x},\hat{z}_2,z_1)P(z_2|\hat{x},z_1)P(z_1|\hat{x}) \\
&= \sum_{z_3,z_2} P(y|x,z_2,z_3) \sum_{z_1} \Big(\sum_{x'} P(z_3|\hat{z}_2,z_1,x')P(x'|\hat{z}_2,z_1) \Big) P(z_2|\hat{x},z_1)P(z_1|\hat{x}) \\
&= \sum_{z_3,z_2} P(y|x,z_2,z_3) \sum_{z_1} \Big(\sum_{x'} P(z_3|z_2,z_1,x')P(x'|\hat{z}_2,z_1) \Big) P(z_2|\hat{x},z_1)P(z_1|\hat{x}) \\
&= \sum_{z_3,z_2} P(y|x,z_2,z_3) \sum_{z_1} \Big(\sum_{x'} P(z_3|z_2,z_1,x')P(x'|z_1) \Big) P(z_2|\hat{x},z_1)P(z_1|\hat{x}) \\
&= \sum_{z_3,z_2} P(y|x,z_2,z_3) \sum_{z_1} \Big(\sum_{x'} P(z_3|z_2,z_1,x')P(x'|z_1) \Big) P(z_2|x,z_1)P(z_1|\hat{x}) \\
&= \sum_{z_3,z_2,z_1} P(y|x,z_2,z_3)P(z_2|x,z_1) \sum_{x'} P(z_3|z_2,z_1,x')P(x'|z_1)P(z_1) \\
&= \sum_{z_3,z_2,z_1} P(y|x,z_2,z_3)P(z_2|x,z_1) \sum_{x'} P(z_3|z_2,z_1,x')P(x',z_1)
\end{aligned}
$$

Note that, in line 1 we have already used a non-isolated conditioning on z_1 but the conditioning on x in line 3 is isolated.

Solution to exercise 5

There are infinitely many graphs that satisfy the conditions required by the exercise. Let us consider the simplest examples, and pinpoint other examples in previous exercises solutions. However, in the



solutions of exercise 3, there was already one such example. You are invited to find it in exercise 11! In there we give an example of a variable for which rule 1 can be applied (with certain interventions present). If so, then the same variable definitely satisfies the conditions for both rule 2 AND rule 3, once again showing that rule 1 is the most restrictive of all rules. The reasons are easy to figure out: the d-separation of two nodes, given the observations on other nodes, is much faster to check and more likely to happen in a graph which less edges. If an observed variable z can be deleted through rule 1, it means it is completely d-separated from the outcome y in the graph G (or $G_{\overline{X}}$ with X set of interventions), given the observations W. If that is the case, the same variable will obviously keep being d-separated in the mutilated graphs $G_{\underline{Z}}$ or $G_{\overline{X(W)}}$, which have even less edges that G.

Solution to exercise 6

The above equality arises from use of conditioning, so it must be an identity coming from the product rule. To see this, let us simplify the expression the above query by calling $Z = \{\hat{x}, z_1\}$, $U = z_3$ and $Y = y$ (just grouping and renaming some variables)

$$\begin{aligned} P(Y|Z) &= \sum_U P(Y|U,Z) \sum_{Y'} P(U|Z,Y')P(Y'|Z) = \sum_U P(Y|U,Z) \sum_{Y'} P(U,Y'|Z) \\ &= \sum_U P(Y|U,Z)P(U|Z) = \sum_U P(Y,U|Z) = P(Y|Z) \end{aligned} \tag{35}$$

which is an identity, as expected.
I suggested in the text of the exercise that one could rewrite the query (in this simplified form) as:

$$P(Y) = \sum_{Y'} \Big(\sum_U P(Y|U)P(U|Y') \Big) P(Y') = \sum_{Y'} \mathbb{A}(Y,Y')P(Y') \tag{36}$$

and that the matrix $\mathbb{A}(y, y') = \mathbb{I}(Y, Y')$, namely that the matrix could only take value 1 whenever $Y = Y'$ but 0 whenever $Y \neq Y'$. This suggestion was wrong, my mistake!
[**Exercise 15: check that the above statement is never true by direct computation, namely consider a (random) dataset with two binary random variables:**

x	**y**
0	**1**
0	**0**
0	**0**
0	**0**
1	**1**
1	**1**
1	**0**
1	**0**
1	**0**
1	**0**

and compute the matrix $\mathbb{A}(Y, Y')$ (after computing all conditional probabilities). Is any of its (4, because we have 2 binary variables so 2^2 possibilities) components zero? If the off-diagonal elements $\mathbb{A}(Y = 0, Y' = 1)$ and $\mathbb{A}(Y = 1, Y' = 0)$ would be zero, what value would the diagonal elements take? Hint: Bayes' rule (do you remember it?) connects the conditional probability $P(Y|U)$ in terms of $P(U|Y)$, and the proportionality constants (the two marginal probabilities) cannot be zero. Hence, if $P(Y|U) = 0$ then $P(U|Y) = 0$ straightforwardly.

Solution to exercise 7

Let us re-write the original query using the division rule:

$$P(y|\hat{x}, z_1) = \frac{P(y, z_1|\hat{x})}{P(z_1|\hat{x})} \tag{37}$$

Now the paths from x to both y and z_1 are the following:

Frontdoor paths

$$x \to z_1 \quad ; \tag{38}$$

$$x \to z_1 \to y \quad ; \tag{39}$$

Backdoor paths

$$x \leftarrow z_3 \to y \to z_1 \quad ; \tag{40}$$

$$x \leftarrow z_3 \to y \quad ; \tag{41}$$

Now while there could be a bit of confusion about the outcome z_1 (which is a collider on the frontdoor path between x and y) it is very easy to see that, by observing z_3 we would block all backdoor paths from x to both z_1 and y. So we condition on z_3 and use rule 2 on $\hat{x}$ in the first query, since all backdoor paths are blocked. We then get:

$$\begin{aligned} P(y|\hat{x}, z_1) &= \frac{\sum_{z_3} P(y, z_1|\hat{x}, z_3)P(z_3|\hat{x})}{P(z_1|\hat{x})} \\ &= \frac{\sum_{z_3} P(y, z_1|x, z_3)P(z_3)}{P(z_1|\hat{x}} \\ &= \frac{\sum_{z_3} P(y, z_1|x, z_3)P(z_3)}{\sum_{z_3} P(z_1|\hat{x}, z_3)P(z_3|\hat{x})} \\ &= \frac{\sum_{z_3} P(y, z_1|x, z_3)P(z_3)}{\sum_{z_3} P(z_1|x, z_3)P(z_3)} \end{aligned}$$

Note: There are two different results to this query, depending on how we use the product rule for the joint probability distribution in the numerator:

$$\begin{aligned} \sum_{z_3} P(y, z_1|x, z_3)P(z_3) &= \sum_{z_3} P(y|x, z_3, z_1)P(z_1|x, z_3)P(z_3) \\ \sum_{z_3} P(y, z_1|x, z_3)P(z_3) &= \sum_{z_3} P(z_1|x, z_3, y)P(y|x, z_3)P(z_3) \\ &= \sum_{z_3} P(z_1|x, y)P(y|z_3)P(z_3) \sum_{z_3} P(z_1|x, y)P(y, z_3) = P(z_1|x, y)P(y) \end{aligned}$$

So the two equivalent results to the query read:

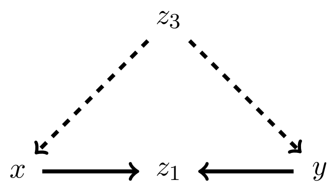
$$P(y|\hat{x}, z_1) = \frac{\sum_{z_3} P(y|x, z_3, z_1)P(z_1|x, z_3)P(z_3)}{\sum_{z_3} P(z_1|x, z_3)P(z_3)} \tag{42}$$

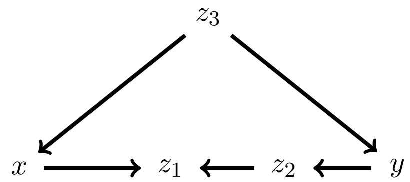
$$P(y|\hat{x}, z_1) = \frac{P(z_1|x, y)P(y)}{\sum_{z_3} P(z_1|x, z_3)P(z_3)} \tag{43}$$

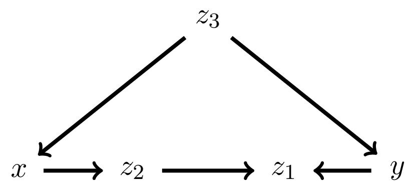
[Exercise 16: Can you see what changed between the joint-probability distribution approach and the previous approach, that lead to a circular recursion? What have we prevented, by joining the outcomes y and z_1, that has stopped the recursion and allowed us to identify the query? Can you see why this graph (and similar) require the use of the division rule?]

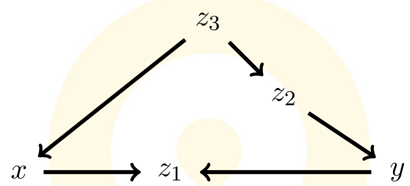
[Exercise 17: Create a dataset following the causal graph arrows for the graph of Exercise 2 (you can neglect z_2 there). Once that is done, compute the numerical values of the above formulas eq. (42) **and** (43) **and show they give the same results within statistical errors (try to increase the size of your dataset, see if the error/differences between the two solutions above decrease)]**

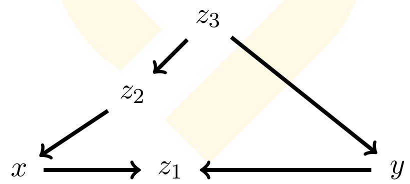
[Exercise 18: To double check your answers to the above exercise, consider the same query $P(y|\hat{x}, z_1)$ for all the graphs below. Find the non-identifiable queries and give the result for the identifiable ones.]

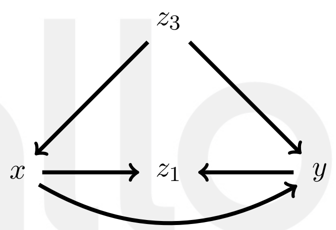












Solution to exercise 8

Below all the possible solutions for the query $P(y|\hat{x})$ with causal graph as in Figure 2.
Blocking variables (z_1, k)

$$
\begin{aligned}
P(y|\hat{x}) &= \sum_{z_2,z_3} P(y|\hat{x}, z_2, z_3) P(z_3|\hat{x}, z_2) P(z_2|\hat{x}) \\
&= \sum_{z_2,z_3} P(y|x, z_2, z_3) P(z_3|\hat{x}, z_2) P(z_2|\hat{x}) \\
&= \sum_{z_2,z_3} P(y|x, z_2, z_3) \sum_{z_1} P(z_3|\hat{x}, z_2, z_1) P(z_2|\hat{x}, z_1) P(z_1|\hat{x}) \\
&= \sum_{z_2,z_3} P(y|x, z_2, z_3) \sum_{z_1} P(z_3|\hat{z_2}, z_1) P(z_2|\hat{x}, z_1) P(z_1|\hat{x}) \\
&= \sum_{z_2,z_3} P(y|x, z_2, z_3) \sum_{z_1} \Big(\sum_{k} P(z_3|\hat{z_2}, z_1, k) P(k|\hat{z_2}, z_1) \Big) P(z_2|\hat{x}, z_1) P(z_1|\hat{x})
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{z_2,z_3} P(y|x,z_2,z_3) \sum_{z_1} \Big(\sum_{k} P(z_3|z_2,z_1,k)P(k|z_1) \Big) P(z_2|\hat{x},z_1)P(z_1|\hat{x}) \\
&= \sum_{z_2,z_3} P(y|x,z_2,z_3) \sum_{z_1} P(z_2|x,z_1)P(z_1) \sum_{k} P(z_3|z_2,z_1,k)P(k|z_1) \qquad (44)
\end{aligned}
$$

Blocking variables (z_1, x)

$$
\begin{aligned}
P(y|\hat{x}) &= \sum_{z_2,z_3} P(y|\hat{x},z_2,z_3)P(z_3|\hat{x},z_2)P(z_2|\hat{x}) \\
&= \sum_{z_2,z_3} P(y|x,z_2,z_3)P(z_3|\hat{x},z_2)P(z_2|\hat{x}) \\
&= \sum_{z_2,z_3} P(y|x,z_2,z_3) \sum_{z_1} P(z_3|\hat{x},z_2,z_1)P(z_2|\hat{x},z_1)P(z_1|\hat{x}) \\
&= \sum_{z_2,z_3} P(y|x,z_2,z_3) \sum_{z_1} P(z_3|\hat{z}_2,z_1)P(z_2|\hat{x},z_1)P(z_1|\hat{x}) \\
&= \sum_{z_2,z_3} P(y|x,z_2,z_3) \sum_{z_1} \Big(\sum_{x'} P(z_3|\hat{z}_2,z_1,x')P(x'|\hat{z}_2,z_1) \Big) P(z_2|\hat{x},z_1)P(z_1|\hat{x}) \\
&= \sum_{z_2,z_3} P(y|x,z_2,z_3) \sum_{z_1} \Big(\sum_{x'} P(z_3|z_2,z_1,x')P(x'|z_1) \Big) P(z_2|\hat{x},z_1)P(z_1|\hat{x}) \\
&= \sum_{z_2,z_3} P(y|x,z_2,z_3) \sum_{z_1} P(z_2|x,z_1)P(z_1) \sum_{x'} P(z_3|z_2,z_1,x')P(x'|z_1) \qquad (45)
\end{aligned}
$$

These 2 results are quite similar, in fact they are they can be transformed into one another, by exchanging $k \leftrightarrow x'$.

Blocking variables (k, x')

$$
\begin{aligned}
P(y|\hat{x}) &= \sum_{z_2,z_3} P(y|\hat{x},z_2,z_3)P(z_3|\hat{x},z_2)P(z_2|\hat{x}) \\
&= \sum_{z_2,z_3} P(y|x,z_2,z_3)P(z_3|\hat{x},z_2)P(z_2|\hat{x}) \\
&= \sum_{z_2,z_3} P(y|x,z_2,z_3) \sum_{k} P(z_3|\hat{x},z_2,k)P(z_2|\hat{x},k)P(k|\hat{x}) \\
&= \sum_{z_2,z_3} P(y|x,z_2,z_3) \sum_{k} P(z_3|\hat{x},\hat{k},z_2)P(z_2|\hat{x},k)P(k|\hat{x}) \\
&= \sum_{z_2,z_3} P(y|x,z_2,z_3) \sum_{k} P(z_3|\hat{k},z_2)P(z_2|\hat{x},k)P(k|\hat{x}) \\
&= \sum_{z_2,z_3} P(y|x,z_2,z_3) \sum_{k} \Big(\sum_{x'} P(z_3|\hat{k},z_2,x')P(x'|\hat{k},z_2) \Big) P(z_2|\hat{x},k)P(k|\hat{x}) \qquad (46) \\
&= \sum_{z_2,z_3} P(y|x,z_2,z_3) \sum_{k} \Big(\sum_{x'} P(z_3|k,z_2,x'){\color{red}P(x'|\hat{k},z_2)} \Big) P(z_2|\hat{x},k)P(k|\hat{x}) \\
&= \sum_{z_2,z_3} P(y|x,z_2,z_3) \sum_{k} \Big[\sum_{x'} P(z_3|k,z_2,x') \Big(\sum_{z_1} P(x'|\hat{k},z_2,z_1)P(z_1|\hat{k},z_2) \Big) \Big] P(z_2|\hat{x},k)P(k|\hat{x}) \\
&= \sum_{z_2,z_3} P(y|x,z_2,z_3) \sum_{k} \Big[\sum_{x'} P(z_3|k,z_2,x') \Big(\sum_{z_1} P(x'|z_1)(\sum_{x''} P(z_1|\hat{k},z_2,x''){\color{red}P(x''|\hat{k},z_2)} \Big) \Big] P(z_2|\hat{x},k)P(k|\hat{x}) \\
&\qquad (47)
\end{aligned}
$$

In this last solution, we have highlighted in red the repeated query, that cannot be solved by considering single outcome queries.

[**Exercise 19: Following the procedure explained for Exercise 7, rewrite the query** $P(x|\hat{k}, z_2)$ **in the above graph, and find all its solutions, if any**].

Note these last exercises, especially 6-7 and 8 have forced us to think about point (d), i.e. the resolution of interventional queries with multiple outcomes. Next week more exercises to come on that!

3 Paths finding and Conditioning unveiled - Project 1

4 Do-calculus Rule 3 unveiled - Project 2

5 Multiple-outcomes queries unveiled - Project 2

6 Craete a real-world dataset