Final Report COMS-W4995
Causal Inference and Deterministic Relations

Nilaksh Agarwal
Department of Computer Science
Columbia University
New York, USA
na2886@columbia.edu

Abstract

Deterministic relationships form an integral part of many mechanisms from physics, chemistry, etc. and current Causal Models are unequipped to handle these relations. Through this paper, I aim to provide a framework to handle determinism within the Causal Framework, building on the D-separation framework from (Geiger et al., 1990).

The main rules of Causal Calculus (called Do-calculus) provide a complete method of identifying causal interventional effects through expressions. In this paper, I provide a similar set of rules for the identification of causal interventional effects in the presence of determinism, and provide some examples where these rules can be utilized.

1 Background

We first start with a recap of relevant topics from literature, which we will utilize to build our Deterministic Causal Calculus.

1.1 Causal Diagrams and SCMs

To completely understand the causal relationship of a system, we aim to create a Structured Causal Model (SCM) (Galles and Pearl, 1998). It consists of a 4-tuple $< \mathbf{V}, \mathbf{U}, \mathscr{F}, P(\mathbf{u}) >$

- $\mathbf{V} = \{V_1, \ldots, V_n\}$ are endogenous variables
- $\mathbf{U} = \{U_1, \ldots, U_m\}$ are exogenous variables
- $\mathscr{F} = \{f_1, \ldots, f_n\}$ functions determining V
  $v_i \rightarrow f_i(pa_i, u_i), Pa_i \subset \mathbf{V}, U_i \subset \mathbf{U}$
- $P(\mathbf{u})$ is a distribution over $\mathbf{U}$

Moreover, every SCM induces a graphical model called a Causal Diagram (DAG) where:

- Each $V_i \in \mathbf{V}$ is a node
- There is a $W \rightarrow V_i$ for $W \in Pa_i$
- There is a $V_i \dashleftarrow\dashrightarrow V_j$ whenever $U_i \cap U_j \neq \emptyset$

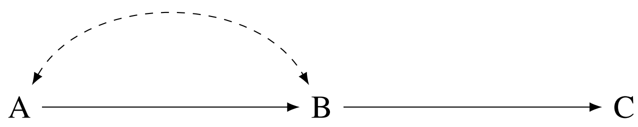


Figure 1: A sample Causal Graph.

If the graph contains no bi-directed edges (i.e., no nodes share any exogenous variables) the graph is said to be Markovian. Otherwise, it is said to be semi-Markovian.

We can employ the following method to convert a semi-markovian graph to a markovian one:

The only caveat here is we are not allowed to condition on the newly introduced variables (e.g. X).

1.2 Deterministic Relationships

In the real world, we do not just deal with probabilistic relationships between variables. Sometimes, we have variables that are deterministically dependant on their parents (i.e.) they do not contain any exogenous variables in their function.

$$v_i \rightarrow f_i(pa_i), Pa_i \subset \mathbf{V}$$

These variables, conditioned on their parents, are independent of all other variables, not merely it's ancestors.

1.3 d-Separation

In a standard Graphical Model (without determinism) we can identify independence between variables using a concept called d-Separation.
Consider the question of whether sets of nodes X and Y are independent given Z.

1. Look at every path from X to Y in the graph
2. A path is active if every triplet in it is active (given Z).
3. If any path is active, X and Y are not independent

For any combination of triplets, the following possibilities can occur

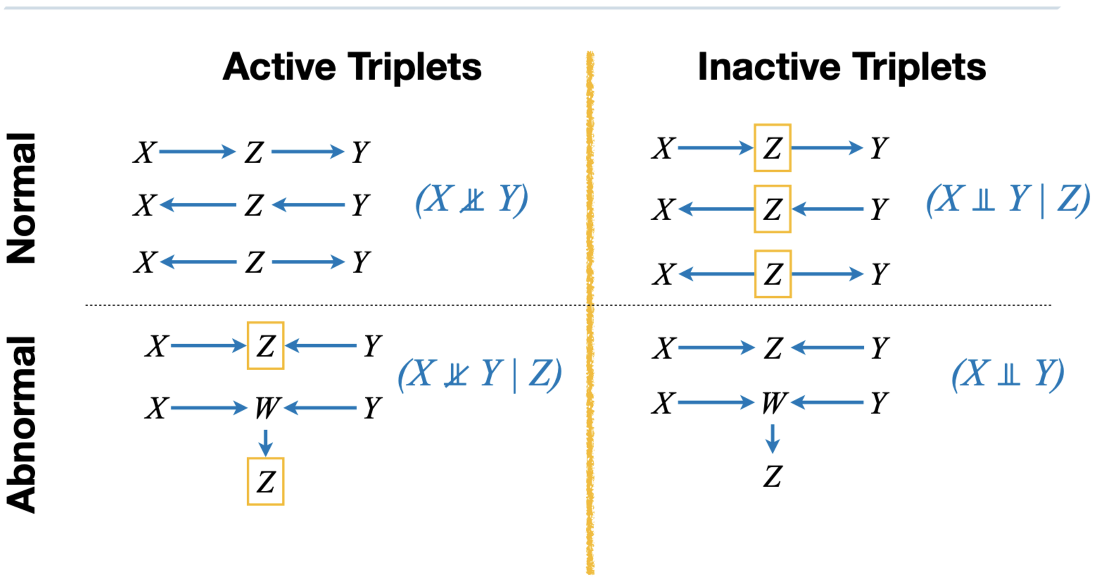


Figure 2: Active/Inactive Triplets in Graph

So, if X and Y are independent (given Z) in the graph, they are said to be d-separated by Z.

A node of the form $X \rightarrow Z \leftarrow Y$ is said to be a collider. Another equivalent definition of a path being active is:
A path p is active by Z if every collider node (wrt p) either is or has a descendant in Z and every other node along p is outside Z. Otherwise, the path is said to be blocked by Z.

1.4 D-Separation

In addition to the conditions present in d-separation, consider the following graph:

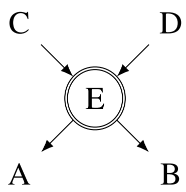
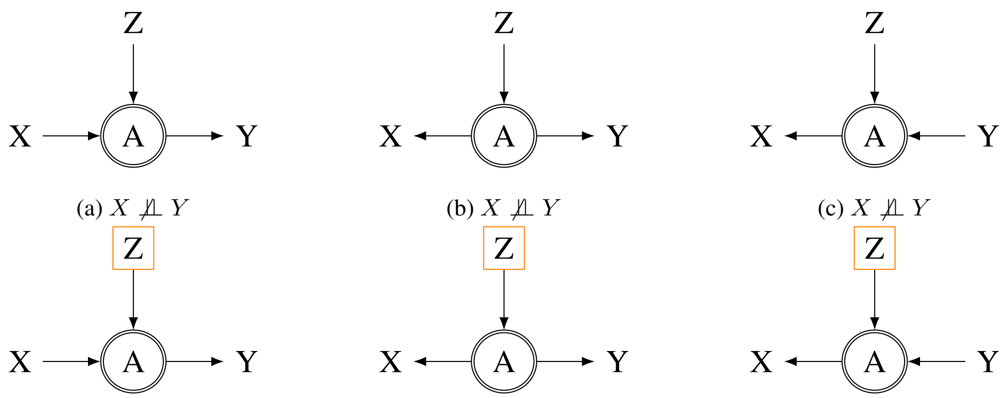


Figure 3: Causal Graph with Determinism

Here a double circle is used for a deterministic (functionally determinable) node. So, Node E is determined entirely by the value of node C and D. Hence, A and B are not D Separated conditioned on C and D (even though they are d-separated given the same conditions).

This leads to additional considerations in our algorithm for D-Separation. For a path to now be declared active, we also must look at the impact the conditioning set has on the deterministic nodes. Hence, if any node is functionally determined given the conditioning set, WLOG it can be added to the conditioning set. Moreover, any deterministic node without a parent is always conditioned on. (Geiger et al., 1990)

This leads us to a definition of Independence given determinism, i.e., D-Separation.



(a) $X \not\perp\!\!\!\perp Y$ (b) $X \not\perp\!\!\!\perp Y$ (c) $X \not\perp\!\!\!\perp Y$

(d) $X \not\perp\!\!\!\perp Y|Z$ (e) $X \perp\!\!\!\perp Y|Z$ (f) $X \not\perp\!\!\!\perp Y|Z$

Figure 4: Some example of active and inactive triplets

Here in part (e) we see that by d-separation, node A should have been inactive, but since it's parents Z are conditioned on, and it is a deterministic node, it is active as well.

1.5 Do-Calculus

In Causal Analysis, a primary tool is using interventions. To understand the effects of these interventions from the data available in the observable world, we employ the following transformations:

- Rule 1: Adding/removing Observations

$$P(y|do(x), z, w) = P(y|do(x), w) \quad if(Y \perp\!\!\!\perp Z|W)_{G_{\overline{X}}}$$

- Rule 2: Action/observation exchange

$$P(y|do(x), do(z), w) = P(y|do(x), z, w) \quad if(Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\underline{Z}}}$$

- Rule 3: Adding/removing Actions

$$P(y|do(x), do(z), w) = P(y|do(x), w) \quad if(Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}\overline{Z(W)}}}$$

  Where $Z(W)$ are the set of Z nodes which are not ancestors of any W-node in $G_{\overline{X}}$

These rules, along with the probability axioms are used to identify interventional experiments in Causal Graphs.

Soundness and Completeness

For the task of causal identifiablity of interventional distribution from $P(v)$, the causal quantity $Q = P(y|do(x))$ is identifiable from $P(v)$ and G if and only if there exists a sequence of application of the rules of do-calculus and the probability axioms that reduces Q into a do-free expression.

2 My Contributions

- Deterministic Do-Calculus (D-Do Calculus)
- Soundness of D-Do Calculus
- Examples where Do Calculus fails and we utilize D-Do Calculus

In the following sections, we will look at these contributions in detail

2.1 D-Do-Calculus

Now shifting to the deterministic relaxation, we notice that if we were to rewrite the Independence relations in terms of D-Separation (in place of d-separation) we would get the corresponding D-Do-Calculus equations. (Here $\perp\!\!\!\perp_D$ corresponds to the D-Separation Independence)

- Rule 1: Adding/removing Observations

$$P(y|do(x), z, w) = P(y|do(x), w) \quad if(Y \perp\!\!\!\perp_D Z|W)_{G_{\overline{X}}}$$

- Rule 2: Action/observation exchange

$$P(y|do(x), do(z), w) = P(y|do(x), z, w) \quad if(Y \perp\!\!\!\perp_D Z|X, W)_{G_{\overline{X}\underline{Z}}}$$

- Rule 3: Adding/removing Actions

$$P(y|do(x), do(z), w) = P(y|do(x), w) \quad if(Y \perp\!\!\!\perp_D Z|X, W)_{G_{\overline{X}\overline{Z(W)}}}$$

  Where $Z(W)$ are the set of Z nodes which are not ancestors of any W-node in $G_{\overline{X}}$

2.1.1 Soundness

In this section, we prove soundness for these rules of Deterministic Do-Calculus in a manner similar to (Pearl, 1995)

Rule 1

In any given causal model, we can write the values of each endogenous variable as:

$$X_i = f_i(pa_i, u_i)$$

Where pa_i signify the parents of the variable X_i in G and u_i are the exogenous variables which influence X_i For deterministic variables, this equation becomes:

$$X_i^{(D)} = f_i(pa_i)$$

Also from (Pearl, 1995) we have the following definition:
Given two disjoint sets of variables, X and Y, the causal effect of X on Y, denoted $pr(y|do(x))$, is a function from X to the space of probability distributions on Y For each realisation x of X, $pr(y|do(x))$ gives the probability of $Y = y$ induced on deleting from the model all equations corresponding to variables in X and substituting x for X in the remainder.

Moreover, by the Markovian property, we have the following:

$$pr(x_1, \ldots, x_n) = \prod_i pr(x_i|pa_i)$$

Hence, we can get the following expression:

$$pr(x_1, \ldots, x_n|do(x_i')) = \begin{cases} pr(x_1, \ldots, x_n)/pr(x_i|pa_i) & \text{if } x_i = x_i' \\ 0 & \text{if } x_i \neq x_i' \end{cases}$$

Hence this expression just showcases the removal of the $pr(x_i|pa_i)$ term in the product.

Since our equations for deterministic nodes are just special cases of independence from exogenous variables, we can see these equations apply to our deterministic case as well. In the last case however, we would be dividing by 0 (in certain assignments due to the deterministic nature), however if we look at it from removing certain terms from the product itself, we will not run into that problem (can be fixed with a conditional check)

Hence, the probability equations left after a do-intervention are the same as the ones which would infer the removal of all links between X_i and Pa_i, namely the graph $G \rightarrow G_{\overline{X}}$

Now we note that if $(Y \perp\!\!\!\perp Z|W)$ then $pr(y|z, w) = pr(y|w)$, this follows from basic probability axioms and the definition of independence.

So, for our deterministic scenario, we have already proved that intervening on X (do(x)) is equivalent to working with an alternative graph $G_{\overline{X}}$. Moreover, this intervention fixes the value of $X = x$.

$$P(y|do(x), z, w) = P(y|do(x), w) \quad if (Y \perp\!\!\!\perp_D Z|W)_{G_{\overline{X}}}$$

Note that we use the D-separation independence in contrast to the d-separation one since we are dealing with deterministic nodes.

This proves soundness for Rule 1 of the Deterministic Causal Calculus.

Rule 2

The graph $G_{\overline{X}\underline{Z}}$ only differs from $G_{\overline{X}}$ in terms of edges from Z, hence all paths with an incoming edge to Z (backdoor paths) are still present. The condition $(Y \perp\!\!\!\perp_D Z|X, W)_{G_{\overline{X}\underline{Z}}}$ ensures that all backdoor paths from Z to Y are blocked by $\{X, W\}$ in $G_{\overline{X}}$. In this scenario, intervening $Z = z$ or conditioning on $Z = z$ will have the same effect on Y. This is because all backdoor paths are blocked, so even on conditioning $Z = z$, no active paths will exist, hence causal information flow would be the same as if intervening $Z = z$ which would remove all incoming edges (and hence backdoor-paths) from Z to Y. Moreover, we use $\perp\!\!\!\perp_D$ here since for all causal information flow from Z to Y to be blocked, we need to account for deterministic nodes in our model. This proves soundness for Rule 2 of the Deterministic Causal Calculus.

Rule 3

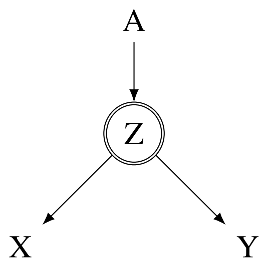
Now consider the graph $G_{\overline{X}}$ to which the intervention $F_z \to Z$ is added. If $(F_z \perp\!\!\!\perp Y|W, Z)_{G_{\overline{X}}}$ then $pr(y|do(x), do(z), w) = pr(y|do(x), w)$. Moreover, if $(Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{XZ(W)}}}$ and $(Y \not\perp\!\!\!\perp F_z|X, W)_{G_{\overline{XZ(W)}}}$, then there must be an unblocked path from $A \in F_z$ to Y that either is a collider ($\to A \leftarrow$) or is a directed path ($\to A \to$). If there is such a path, let P be the shortest such path.

If the path has a collider then we know $(Y \perp\!\!\!\perp A|X, W)_{G_{\overline{XZ(W)}}}$ but $(Y \not\perp\!\!\!\perp A|X, W)_{G_{\overline{X}}}$. So there must be a unblocked path from Y to A that passes through some $B \in Z(W)$ which is either a collider or a directed path. If B is a collider, then some descendant of $B \in W$ for the path to be unblocked, but then $B \not\in Z(W)$. Similarly, if it is a directed path $\to B \to$ then either the path from A to B ends in $\to B$, or $B \to$. If it ends in an arrow pointing away from B, then there must be a collider junction along the path from A to B (since arrow is outgoing from A). In that case, for the path to be unblocked, W must be a descendant of B, but then B would not be in Z(W). If it ends in an arrow pointing to B, then there must be an unblocked path from B to Y in $G_{\overline{X}}$ that is blocked in $G_{\overline{XZ(W)}}$. If this is true, then there is an unblocked path from B to Y that is shorter than P, the shortest path. Due to these contradictions, this means that $(Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{XZ(W)}}} \implies pr(y|do(x), do(z), w) = pr(y|do(x), w)$.

Since in our deterministic case, we have to additionally check for independencies arising from the functional determinism, we replace the equation with $(Y \perp\!\!\!\perp_D Z|X, W)_{G_{\overline{XZ(W)}}} \implies pr(y|do(x), do(z), w) = pr(y|do(x), w)$. This proves soundness for Rule 3 of our Deterministic Do-calculus.

2.2 Examples

Here we will showcase cases where the introduction of determinism causes regular Causal Calculus to fail.



Consider the above figure, where Z is a deterministic node (Z = f(A)). Since Z is functionally determined by A, if we condition on A, w.l.o.g we can condition on A,Z (by the deterministic case).

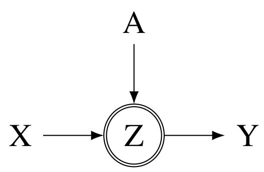
Rule 1

If we wish to simplify $Pr(y|do(a), x)$, we can use D-Do calculus Rule 1, to write $Pr(y|do(a), x) = Pr(y|do(a))$ since $(Y \perp\!\!\!\perp_D X|A)_{G_{\overline{A}}}$. However, $(Y \not\perp\!\!\!\perp X|A)_{G_{\overline{A}}}$ in the non-deterministic case. So in this situation, regular Do-Calculus falls through.

Rule 2

If we wish to simplify $Pr(y|do(a), do(x))$, we can use D-Do calculus Rule 2, to write $Pr(y|do(a), do(x)) = Pr(y|do(a), x)$ since $(Y \perp\!\!\!\perp_D X|A)_{G_{\overline{A}\underline{X}}}$. However, $(Y \not\perp\!\!\!\perp X|A)_{G_{\overline{A}\underline{X}}}$ in the non-deterministic case. So in this situation, regular Do-Calculus falls through.

Rule 3



In the figure above, if we wish to simplify $Pr(y|do(x), do(a))$, we can use D-Do calculus Rule 3, to write $Pr(y|do(x), do(a)) = Pr(y|do(a))$ since $(Y \perp\!\!\!\perp_D X|A)_{G_{\overline{AX}}}$, as intervening on X and A implies intervening on Z as well. However, $(Y \not\perp\!\!\!\perp X|A)_{G_{\overline{AX}}}$ in the non-deterministic case. So in this situation, regular Do-Calculus falls through.

3 Conclusion/Future Scope

This paper serves to provide a framework to unify causality with determinism, and gives certain rules/examples for its use. Furthermore, subsequent work in this field can be to extend determinism to identification algorithms, counterfactuals, and other causal frameworks.

References

David Galles and Judea Pearl. 1998. An axiomatic characterization of causal counterfactuals. *Foundations of Science*, 3(1):151–182.

Dan Geiger, Thomas Verma, and Judea Pearl. 1990. Identifying independence in bayesian networks. *Networks*, 20(5):507–534.

Yimin Huang and Marco Valtorta. 2012. Pearl's calculus of intervention is complete. *arXiv preprint arXiv:1206.6831*.

Judea Pearl. 1995. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688.

Judea Pearl and Rina Dechter. 2013. Identifying independencies in causal graphs with feedback. *arXiv preprint arXiv:1302.3595*.