Introduction to Causal Calculus

Sanna Tyrväinen

University of British Columbia

August 1, 2017

Bayesian network

Bayesian networks are *Directed Acyclic Graphs* (DAGs) whose nodes represent random variables:

$$p(x_1, \ldots, x_d) = \prod_{j=1}^{d} p(x_j | x_{pa(j)})$$

Assumes that a variable is independent of previous non-parents given the parents, that is, $p(x_j | x_1, \ldots, x_d) = p(x_j | x_{pa(j)})$

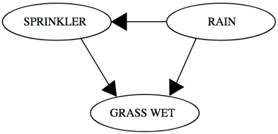
Captures how variables are conditionally dependent: If there are no any arrows between the nodes then they are independent:

$$p(A, B) = p(A)p(B)$$

The joint probability distribution:

$$p(G, S, R) = p(G|S, R)p(S|R)p(R)$$

	SPRINKLER	
RAIN	T	F
F	0.4	0.6
T	0.01	0.99



RAIN	
T	F
0.2	0.8

		GRASS WET	
SPRINKLER	RAIN	T	F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

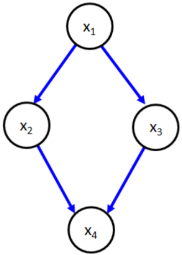
Sampling from a DAG

Ancestral sampling:

- Sample x_1 from $p(x_1)$
- If x_1 is a parent of x_2, sample x_2 from $p(x_2|x_1)$ Otherwise, sample x_2 from $p(x_2)$
- Go through the subsequent j in order sampling x_j from $p(x_j|x_{pa(j)})$

Conditional Sampling:

- easy if condition on the first variables: fix these and run ancestral sampling
- Hard if condition on the last variables: Conditioning on descendent makes ancestors dependent



It's hard to separate out causality from correlation

DAGs can be viewed as a causal process: the parents "cause" the children to take different values

The below equations are equivalent and the graphs have same conditional independences, but the causalities are not the same. Graphs tells us something useful that equations miss.

$$Y = X + 1$$
$$Z = 2Z$$

$$X = Y - 1$$
$$Y = Z/2$$

There is *observational data* ("seeing") and *interventional data* ("doing")

Usually the DAG is designed for observational data, but that ignores the possibility of unobserved variables, also without interventional data you can't distinguish the direction of causality.

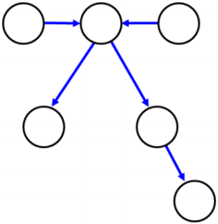
Simplest external intervention: a single variable is forced to take some fixed value (in a graph remove arrows entering that variable)

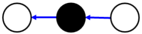
D-separation

d-separation is a criterion for deciding, from a causal graph, whether a set A of variables is independent of another set B (given a third set C)

$$A \perp\!\!\!\perp B|C$$

A and B are d-separated if for all paths P from A to B, at least one of the following holds:



- P includes a "chain" with an observed middle node
  
- P includes a "fork" with an observed parent node
  
- P includes a "v-structure" or "collider"

A and B are *d-separated*, give C, iff corresponding random variables are conditionally independent:

$$p(A, B|C) = p(A|C)p(B|C)$$

If A and B are not d-separated they are *d-connected*

The Causal Calculus (do-calculus, Pearl's Causal Calculus, Calculus of Actions)

Shortly: Calculus to discuss causality in a formal language by Judea Pearl

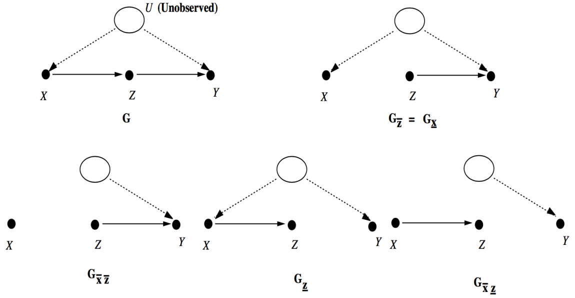
A new operator, *do()*, marks an action or an intervention in the model. In an algebraic model we replace certain functions with a constant $X = x$, and in a graph we remove edges going into the target of intervention, but preserve edges going out of the target.

The causal calculus uses Bayesian conditioning, $p(y|x)$, where x is observed variable, and causal conditioning, $p(y|do(x))$, where an action is taken to force a specific value x.

Goal is to generate probabilistic formulas for the effect of interventions in terms of the observed probabilities.

Judea Pearl, *Causality: Models, reasoning, and inference.* Cambridge University Press, 2000.

Notations



Notation: a graph G, W, X, Y, Z are disjoint subsets of the variables. $G_{\overline{X}}$ denotes the perturbed graph in which all edges pointing to X have been deleted, and $G_{\underline{X}}$ denotes the perturbed graph in which all edges pointing from X have been deleted. $Z(W)$ denote the set of nodes in Z which are not ancestors of W

Image: Judea Pearl

Pearl's 3 rules

Pearl's 3 rules

- Ignoring observations

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

- Action/Observation exchange (the back-door criterion)

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

- Ignoring actions/interventions

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

Notation: a graph G, W, X, Y, Z are disjoint subsets of the variables. $G_{\overline{X}}$ denotes the perturbed graph in which all edges pointing to X have been deleted, and $G_{\underline{X}}$ denotes the perturbed graph in which all edges pointing from X have been deleted. $Z(W)$ denote the set of nodes in Z which are not ancestors of W

Intuition behind the Pearl's first rule

With condition $(Y \perp\!\!\!\perp Z|X, W)_{G_{\overline{X}}}$ we have

$$p(y|do(x), z, w) = p(y|do(x), w)$$

- Let's start with a simple case where we assume that there are no W or X. We get a condition $(Y \perp\!\!\!\perp Z)_G$, so Y is independent of Z, that is, $p(y|z) = p(y)$
- In the second case assume we have passively observed W, but no variable X: $(Y \perp\!\!\!\perp Z|W)_G$. Earlier we mentioned connection of d-separation and conditionally independent, that is, $p(y|z, w) = p(y|w)$
- The third case assume we don't know W, but we have X that's value is set by intervention: $(Y \perp\!\!\!\perp Z|X)_{G_{\overline{X}}}$. By the same theorem, that is, $p(y|z, do(x)) = p(y|do(x))$

Combining these gives the full rule.

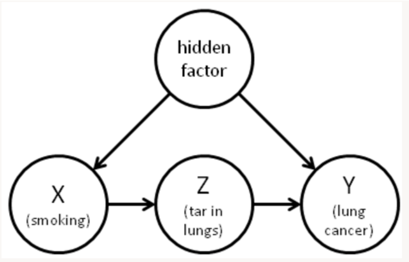
Example: Smoking and lung cancer

Randomized Controlled Trials (RCT)

- AKA Randomized Control Trial, Randomized clinical trial
- The participants in the trial are randomly allocated to either the group receiving the treatment under investigation or to the control group
- The control group removes the confounding factor of the placebo effect
- Double-blind studies remove further confounding factors
- Sometimes impractical or impossible

We can try to use causal calculus to analyze the probability that someone would get cancer given that they are smoking, without doing an actual RCT:

$$p(y|do(x))$$

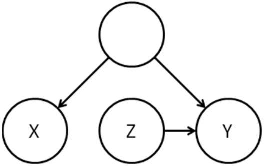


Note: We have no information about the hidden variable that could cause both smoking and cancer

Example

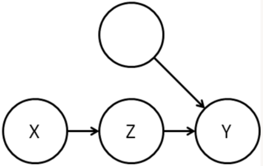
We can't try to apply rule 1 because there is no observations to ignore, we would just have $p(y|do(x)) = p(y|do(x))$.

Try apply rule 2: We would have $p(y|do(x)) = p(y|x)$, that is, the intervention doesn't matter. It's condition is $(Y \perp\!\!\!\perp X)_{G_{\underline{X}}}$:



Y and X are not d-separated, because they have a common ancestor.
$\Longrightarrow$ Rule 2 can't be applied

Try apply rule 3: We would have $p(y|do(x)) = p(y)$, that is, an intervention to force someone to smoke has no impact on whether they get cancer. It's condition is $(Y \perp\!\!\!\perp X)_{G_{\overline{X}}}$:

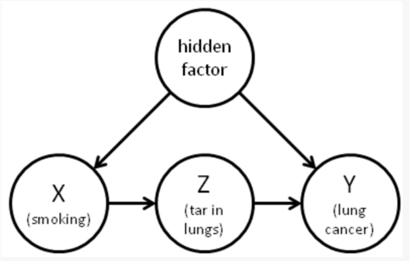


Y and X are not d-separated, because we have unblocked path between them.
$\Longrightarrow$ Rule 3 can't be applied

Example

New attempt:

$$
\begin{aligned}
p(y|do(x)) &= \sum_z p(y|z, do(x))p(z|do(x)) \\
&= \sum_z p(y|z, do(x))p(z|x) && (\text{rule 2: } (Z \perp\!\!\!\perp X)_{G_{\underline{X}}}) \\
&= \sum_z p(y|do(z), do(x))p(z|x) && (\text{rule 2: } (Y \perp\!\!\!\perp Z|X)_{G_{\overline{X},\underline{Z}}}) \\
&= \sum_z p(y|do(z))p(z|x) && (\text{rule 3: } (Y \perp\!\!\!\perp X|Z)_{G_{\overline{Z},\overline{X}}})
\end{aligned}
$$



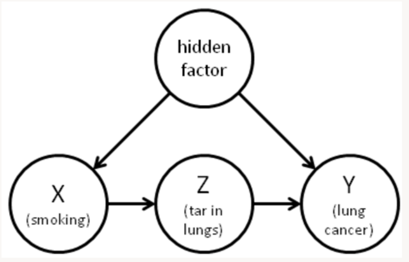
Example

We can use the same approach to the first term on the right hand side:

$$p(y|do(z)) = \sum_x p(y|x, do(z))p(x|do(z))$$
$$= \sum_x p(y|x, z)p(x) \qquad \text{(rule 2 + rule 3)}$$

Finally we can combine these results:

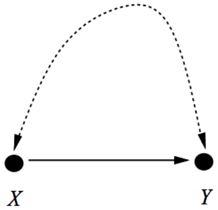
$$p(y|do(x)) = \sum_{z,x'} p(y|x', z)p(z|x)p(x')$$



We can now compare $p(y)$ and $p(y|x)$. The needed probabilities can be observed directly from experimental data: What part of smokers have lung cancer, how many of them have tar in their lungs etc.

Example: end

- The analysis would have not worked if the graph had missed the tar variable, Z, because there is no general way to compute $p(y|do(x))$ from any observed distributions whenever the causal model includes subgraph shown the figure below
- Causal Calculus can be used to analyze causality in more complicated (and more unethical) situations than RCT
- Causal Calculus can also be used to test whether unobserved variables are missed by removing all do terms from the relation
- Not all models are acyclic. See for example Modeling Discrete Interventional Data Using Directed Cyclic Graphical Models (UAI 2009) by Mark Schmidt and Kevin Murphy



Literature

- A Probabilistic Calculus of Actions (UAI1994) by Judea Pearl
- Tutorial by Michael Nielsen: http://www.michaelnielsen.org/ddi/if-correlation-doesnt-imply-causation-then-what-does/
- Tutorial in Formalised Thinking: https://formalisedthinking.wordpress.com/2010/08/20/pearls-formalisation-of-causality-sequence-index/
- Judea Pearl, Causality: Models, reasoning, and inference, Cambridge University Press, 2000
- Introduction to Judea Pearl's Do-Calculus, Robert Tucci, 2013

Thank you for listening. Questions?