Probability of Causation

Amine Hadji

University of Leiden

May 27, 2019

Overview

- Model and notations
- Definitions
- Identifiability (under some conditions)
- Interesting examples
- Identifiability (in general)

Causal model

Let V a set of random variables.

Definition

A functional causal model consists of a set of equations of the form

$$X_i = f_i(PA_i, U_i)$$

where $PA_i \subseteq V$, and U_i represent errors.

From a certain causal model, we can draw a causal diagram $G = (V, E)$, such that there is an arrow from each of the PA_i to X_i.

- If G is acyclic, then it is said to be *semi-Markovian*
- If moreover the U_i's are jointly independent, G is said to be *Markovian*

Example

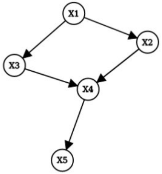

Figure: Causal diagram G of the functional model M

The Markovian model M:

- $X_1 = f_1(U_1)$
- $X_2 = f_2(X_1, U_2)$
- $X_3 = f_3(X_1, U_3)$
- $X_4 = f_4(X_2, X_3, U_4)$
- $X_5 = f_5(X_1, U_5)$

with the U_i's jointly independent.

In Pearl's book, the X_i's are almost always considered binary variables; therefore U_i's can also be summed up to binary variables.

Notations

According to Pearl's notation, for all $X, Y \in V$

- $x : X = 1$
- $x' : X = 0$
- $y_x : Y^{X=1} = 1$
- $y_{x'}: Y^{X=0} = 1$
- $y'_x : Y^{X=1} = 0$
- $y'_{x'}: Y^{X=0} = 0$

where $Y^{X=x}$ denotes the counterfactual of Y according to X.
An important assumption for the following results is consistency, i.e.

$$P(Y = y|X = x) = P(Y^{X=x} = y), \text{ or}$$
$$x \Rightarrow (y = y_x), \text{ and } x' \Rightarrow (y = y_{x'})$$

Main definitions

Definition

According to the previous notations, the probability of necessity is defined as the expression

$$PN = P(y'_{x'}|y, x) = P(Y^{X=0} = 0|Y = 1, X = 1).$$

Definition

According to the previous notations, the probability of sufficiency is defined as the expression

$$PS = P(y_x|y', x') = P(Y^{X=1} = 1|Y = 0, X = 0).$$

Main definitions

Definition

The probability of necessity and sufficience is defined as the expression

$$PNS = P(y_x, y'_{x'}) = P(Y^{X=1} = 1, Y^{X=0} = 0).$$

Main definitions

Definition

The probability of necessity and sufficience is defined as the expression

$$PNS = P(y_x, y'_{x'}) = P(Y^{X=1} = 1, Y^{X=0} = 0).$$

Lemma

Under the consistency condition, we have

$$PNS = P(x, y)PN + P(x', y')PS.$$

Lemmas

Lemma

Let $X, Y, Z, Q \in V$ such that X may be an (indirect) cause of Y, and $Z = Y \wedge Q$. If $Q \perp\!\!\!\perp \{X, Y, Y^{X=0}\}$

$$PN(x, z) = PN(x, y).$$

Lemma

Let $X, Y, Z, R \in V$ such that X may be an (indirect) cause of Y, and $Z = Y \vee R$. If $R \perp\!\!\!\perp \{X, Y, Y^{X=1}\}$

$$PS(x, z) = PS(x, y).$$

Sketch of the proof

- $$\begin{aligned} PN(x,z) &= PN(z'_{x'}|x,z) = \frac{P(z'_{x'},x,z)}{P(x,z)} \\ &= \frac{P(z'_{x'},x,z|q)P(q)+P(z'_{x'},x,z|q')P(q')}{P(x,z,q)+P(x,z,q')} \end{aligned}$$
- Since $Z = Y \wedge Q$, $Q = 1$ implies $Z = Y$ and $Q = 0$ implies $Z = 0$.
- $PN(x,z) = \frac{P(y'_{x'},x,y|q)P(q)}{P(x,y,q)} = PN(x,y)$ because $Q \perp\!\!\!\perp \{X, Y, Y^{X=0}\}$

Exogeneity

Definition

A variable X is said to be *exogenous* to Y in model M iff $\{Y^{X=1}, Y^{X=0}\} \perp\!\!\!\perp X$.

Exogeneity is quite a strong assumption, but it also allows the identification of the laws of $Y^{X=1}$ and $Y^{X=0}$. Indeed

$$P(Y^{X=x} = y) = P(Y^{X=x} = y|X = x) = P(Y = y|X = x)$$

which can be estimated.

Bounds for PNS

Theorem

Under condition of exogeneity, PNS is bounded as follows:

$$\max\{0, P(Y=1|X=1) - P(Y=1|X=0)\} \leq PNS$$
$$PNS \leq \min\{P(Y=1|X=1), P(Y=0|X=0)\}.$$

Moreover, the bounds are sharp in the sense that there exists a functional causal model $Y = f(X, U)$ that realizes any value permitted by the bounds.

The theorem is directly derived by the sharp bounds

$$\max\{0, P(A) + P(B) - 1\} \leq P(A, B) \leq \min\{P(A), P(B)\},$$

and the identification of the laws of $Y^{X=1}$ and $Y^{X=0}$.

Identifiability of PN and PS

Theorem

Under condition of exogeneity:

$$PN = \frac{PNS}{P(Y=1|X=1)}$$

$$PS = \frac{PNS}{P(Y=0|X=0)}.$$

Therefore, the bounds on PNS provide corresponding bounds for PN and PS.

Identifiability of PN and PS

Corollary

If $P(Y^{X=1}=1)$ and $P(Y^{X=0}=1)$ are measured in an experimental study, then for any point p in the range

$$\frac{\max\{0, P(Y^{X=1}=1) - P(Y^{X=0}=1)\}}{P(Y^{X=1}=1)} \leq p$$

$$p \leq \frac{\min\{P(Y^{X=1}=1), P(Y^{X=0}=1)\}}{P(Y^{X=1}=1)}$$

there exists a causal model M that agrees with $P(Y^{X=1}=1)$ and $P(Y^{X=0}=1)$, and for which $PN = p$.

Proof of the Theorem

- $PN = P(y'_{x'}|x, y) = \frac{P(y'_{x'},x,y)}{P(x,y)}$
- Using consistency, we can see that $P(y'_{x'}, x, y) = P(y'_{x'}, x, y_x)$
- Since X is exogenous to Y, the last quantity is also equal to $P(y'_{x'}, y_x)P(x)$
- Replacing this quantity in the first equation leads directly to $PN = P(y'_{x'}|x, y) = \frac{PNS\ P(x)}{P(x,y)} = \frac{PNS}{P(Y=1|X=1)}$.

Monotonicity

Definition

A variable Y is said to be *monotonic* relative to X in a causal model M iff $Y^{X=1} \geq Y^{X=0}$ (modulo complemeting).

Monotonicity can be equivalently described as $Y = f(X, U) \neq \neg X$ for all possible values taken by U.

Identifiability of PNS

Theorem

If X is exogenous and Y is monotonic relative to X, then PNS (and therefore PN and PS) is identifiable, with

$$PNS = P(Y = 1|X = 1) - P(Y = 1|X = 0).$$

- The derived PN is known as the excess risk ration.
- The derived PS is known as the relative difference.

Proof of the Theorem

- $y_{x'} \vee y'_{x'}$ =true
- $y_x = y_x \wedge (y_{x'} \vee y'_{x'}) = (y_x \wedge y_{x'}) \vee (y_x \wedge y'_{x'})$
- $y_{x'} = y_{x'} \wedge (y_x \vee y'_x) = y_{x'} \wedge y_x$ because of monotonicity.
- Substituting this quantity into the previous equations yields
  $y_x = y_{x'} \vee (y_x \wedge y'_{x'})$.
- $P(y_x) = P(y_{x'}) + P(y_x, y'_{x'})$
- $P(y_x, y'_{x'}) = P(y_x) - P(y_{x'})$.

Using the argument of exogeneity, we derive directly the result of the theorem.

Identifiability of PNS

The previous proof showed that exogeneity is not necessary to identify PNS. Indeed

Identifiability of PNS

The previous proof showed that exogeneity is not necessary to identify PNS. Indeed

Theorem

If Y is monotonic relative to X, and the laws of $Y^{X=1}$ and $Y^{X=0}$ are identifiable, then PN, PS and PNS are identifiable, with

$$PNS = P(Y^{X=1} = 1) - P(Y^{X=0} = 1),$$

$$PN = \frac{P(Y = 1) - P(Y^{X=0} = 1)}{P(X = 1, Y = 1)},$$

$$PS = \frac{P(Y^{X=1} = 1) - P(Y = 1)}{P(X = 0, Y = 0)}.$$

Identifiability of PNS

The proof resembles the previous proof, without the condition of exogeneity

Identifiability of PNS

The proof resembles the previous proof, without the condition of exogeneity

Besides, since PN and PS must be nonnegative, the theorem provides a simple test for the assumption of monotonicity: $P(Y^{X=1} = 1) \geq P(Y = 1) \geq P(Y^{X=0} = 1)$

Identifiability of PNS

Asking for identifiability of the laws of $Y^{X=1}$ and $Y^{X=0}$ might seem unreasonable. However, for every X and Y variables in a positive Markovian model M

$$P(Y^{X=1}=1)=\sum_{pa_X} P(Y=1|pa_X, X=1)P(pa_X)$$

Identifiability of PNS

Asking for identifiability of the laws of $Y^{X=1}$ and $Y^{X=0}$ might seem unreasonable. However, for every X and Y variables in a positive Markovian model M

$$P(Y^{X=1} = 1) = \sum_{pa_X} P(Y = 1|pa_X, X = 1)P(pa_X)$$

Corollary

For any positive Markovian model M, if Y is monotonic relative to X, then PN, PS and PNS are identifiable.

Example 1 - Betting against a fair coin

Assume we must bet heads or tails on the outcome of a fair coin toss, and we win a dollar if we guess correctly.

Example 1 - Betting against a fair coin

Assume we must bet heads or tails on the outcome of a fair coin toss, and we win a dollar if we guess correctly.

- x stands for "we bet on heads"
- y stands for "we win a dollar"
- u stands for "the coin turned up heads"

$$y = (x \wedge u) \vee (x' \wedge u')$$

Example 1 - Betting against a fair coin

Assume we must bet heads or tails on the outcome of a fair coin toss, and we win a dollar if we guess correctly.

- x stands for "we bet on heads"
- y stands for "we win a dollar"
- u stands for "the coin turned up heads"

$$y = (x \wedge u) \vee (x' \wedge u')$$

$$PN = P(y'_{x'}|x, y) = P(y'_{x'}|u) = 1$$
$$PS = P(y_x|x', y') = P(y_x|u) = 1$$

Example 1 - Betting against a fair coin

Assume we must bet heads or tails on the outcome of a fair coin toss, and we win a dollar if we guess correctly.

- x stands for "we bet on heads"
- y stands for "we win a dollar"
- u stands for "the coin turned up heads"

$y = (x \wedge u) \vee (x' \wedge u')$

$$PN = P(y'_{x'}|x, y) = P(y'_{x'}|u) = 1$$
$$PS = P(y_x|x', y') = P(y_x|u) = 1$$
$$PNS = P(y_x, y'_{x'}) = P(y_x, y'_{x'}|u)P(u) = 1/2$$

Example 2 - Legal Responsibility

A lawsuit is filed against a manufacturer, chargin that a certain drug is likely to have caused the death of Mr. A.

Both experimental and nonexperimental data are provided in this table

	Experimental		*Nonexperimental*	
	$X = 1$	$X = 0$	$X = 1$	$X = 0$
Deaths $Y = 1$	16	14	2	28
Survival $Y = 0$	984	986	998	972

Our main interest is to compute PN.

Example 2 - Legal Responsibility

The experimental data provide the estimates $\hat{P}(Y^{X=1} = 1) = 0.016$ and $\hat{P}(Y^{X=0} = 1) = 0.014$, whereas the nonexperimental data provide the estimates $\hat{P}(Y = 1) = 0.015$ and $\hat{P}(Y = 1, X = 1) = 0.001$.

$$\hat{PN} \geq \frac{\hat{P}(Y=1) - \hat{P}(Y^{X=0}=1)}{\hat{P}(X=1, Y=1)} = 1$$

Therefore, the plaintiff was correct. The data show that the drug was responsible for the death of Mr. A.

Example 2 - Legal Responsibility

The experimental data provide the estimates $\hat{P}(Y^{X=1} = 1) = 0.016$ and $\hat{P}(Y^{X=0} = 1) = 0.014$, whereas the nonexperimental data provide the estimates $\hat{P}(Y = 1) = 0.015$ and $\hat{P}(Y = 1, X = 1) = 0.001$.

$$\hat{PN} \geq \frac{\hat{P}(Y=1) - \hat{P}(Y^{X=0}=1)}{\hat{P}(X=1, Y=1)} = 1$$

Therefore, the plaintiff was correct. The data show that the drug was responsible for the death of Mr. A.Compare with the experimental excess risk ratio $\frac{\hat{P}(Y^{X=1}=1) - \hat{P}(Y^{X=0}=1)}{\hat{P}(Y^{X=1}=1)} = 0.125$.

Nonmonotonicity

Monotonocity played a big role in the identification of PN, PS and PNS, but how essential is it?

Consider any arbitrary equation in the model M:
$Y = f(PA_Y, U_Y)$
Since the observable variables are binary, there is only a finite number of functions from PA_Y to Y, and for any point $U_Y = u$, only one of those functions is realized.

Nonmonotonicity

Monotonocity played a big role in the identification of PN, PS and PNS, but how essential is it?

Consider any arbitrary equation in the model M:
$Y = f(PA_Y, U_Y)$
Since the observable variables are binary, there is only a finite number of functions from PA_Y to Y, and for any point $U_Y = u$, only one of those functions is realized.

We can then partition the domain of U_Y into a set S of equivalence classes. $s \in S$ induces the same function $f^{(s)}$ from PA_Y to Y.

Nonmonotonicity

Monotonocity played a big role in the identification of PN, PS and PNS, but how essential is it?

Consider any arbitrary equation in the model M:
$Y = f(PA_Y, U_Y)$
Since the observable variables are binary, there is only a finite number of functions from PA_Y to Y, and for any point $U_Y = u$, only one of those functions is realized.

We can then partition the domain of U_Y into a set S of equivalence classes. $s \in S$ induces the same function $f^{(s)}$ from PA_Y to Y.

We can now regard S as a new background variable whose values correspond to $\{f^{(s)} : s \in S\}$

Nonmonotonicity

Some direct computations show that
$P(y|pa_Y) = \sum_{s:f^{(s)}(pa_Y)=true} P(s)$.

Nonmonotonicity

Some direct computations show that
$P(y|pa_Y) = \sum_{s:f^{(s)}(pa_Y)=true} P(s)$.

If we can invert the process, and determine $P(s)$ from $P(y|pa_Y)$, then the model is directly identifiable. This boils down to invert the matrix $\mathbf{R}$ in
$\vec{p} = \mathbf{R}\vec{q}$.

Nonmonotonicity

Some direct computations show that $P(y|pa_Y) = \sum_{s:f^{(s)}(pa_Y)=true} P(s)$.

If we can invert the process, and determine $P(s)$ from $P(y|pa_Y)$, then the model is directly identifiable. This boils down to invert the matrix $\mathbf{R}$ in $\vec{p} = \mathbf{R}\vec{q}$.

However, in general $\mathbf{R}$ is not invertible because the dimensionality of $\vec{q}$ is much larger than that of $\vec{p}$.

Local invertibility

We note though that the invertibility of $\mathbf{R}$ is not a necessary condition for identifiability

Local invertibility

We note though that the invertibility of **R** is not a necessary condition for identifiability

Definition

A model M is said to be *locally invertible* if, for every $Y_i \in V$, the set of equations

$$P(y_i|pa_i) = \sum_{s:f^{(s)}(pa_i)=true} q_i(s)$$

$$\sum_s q_i(s) = 1$$

has a unique solution for $q_i(s)$, where each $f_i^{(s)}(pa_i)$ corresponds to the function $f_i(pa_i, u_i)$ induced by u_i in equivalence class s.

Local invertibility

Theorem

Given a Markovian model $M = (U, V, \{f_i\})$ in which the functions $\{f_i\}$ are known and the variables U are unobserved, if M is locally invertible then the probability of every counterfactual sentence is identifiable from the joint probability $P(v)$.