Probabilities of Causation

Three counterfactual interpretations and their identification

Judea Pearl

March 4, 2022

Presentation by Edgar Robles @ UCI

Table of contents

Introduction

Motivation

There is great value in knowing how to quantify causal relationships between variables, this can be used for applications in

- epidemiology
- legal reasoning
- psychology
- natural sciences.

Quantifying causality

This paper tries to look at quantifying this through three statistics

- Probability of necessity (PN): $X \rightarrow Y$
- Probability of sufficiency (PS): $X \leftarrow Y$
- Probability of necessity and sufficiency (PNS): $X \leftrightarrow Y$

Definitions

Notation

For simplicity, we will only work with binary random variables. We will denote the true value of X as x and the false value of X as x'.

For example, Y_x is the variable Y had X been true, and $Y_{x'}$ is the variable Y had X been false.

Consistency condition (Robins 1987)

Definition
For two directly observed variables $X(u)$ and $Y(u)$,

$$(X = x) \implies (Y_x = Y) \tag{1}$$

this essentially just means that, in a world where $X = x$, the Y has to be the same variable as Y_x.

Probability of Necessity

The probability of necessity is the probability that y would not occur in the absence of x, in the case where x and y did occur.

Definition

$$PN \triangleq P(Y_{x'} = \text{false} | X = \text{true}, Y = \text{true}) \triangleq P(y'_{x'} | x, y) \tag{2}$$

This is **not** identifiable in the general case.

Probability of Sufficiency

The probability of sufficiency is the probability the capacity of x to produce y, in the case where both x and y are absent.

Definition

$$PS \triangleq P(Y_x = \text{true} | X = \text{false}, Y = \text{false}) \triangleq P(y_x | x', y') \tag{3}$$

This is also not identifiable in the general case.

Probability of Necessity and Sufficiency

As the "if and only if" of the relationship, this measures the world in which y is true when x is true, and y is false when x is false.

Definition

$$PNS \triangleq P(Y_x = \text{true}, Y_{x'} = \text{false}) \triangleq P(y_x, y'_{x'}) \tag{4}$$

also not identifiable in the general case.

Identification of variables

A simple relationship

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

Proof.
Using the consistency conditions,

$$x \implies (y_x = y), x' \implies (y_{x'} = y),$$

we can write

$$P(y_x, y'_{x'}) = P((y_x \cap y'_{x'}) \cap (x \cup x')) = P(y'_{x'}, x, y) + P(y_x, x', y')$$

$$P(y_x, y'_{x'}) = P(y'_{x'}|x, y)P(x, y) + P(y_x|x', y')P(x', y')$$

$\square$

Insensitivity to potential inhibitors of PN

Lemma
Let $PN(x, y)$ stand for the probability that x is a necessary cause of y and $z = y \wedge q$ a consequence of y, potentially inhibited by q'. If $q \perp\!\!\!\perp X, Y_x, Y_{x'}$, then

$$PN(x, z) \triangleq P(z'_{x'}|x, z) = P(y'_{x'}|x, y) \triangleq PN(x, y)$$

Proof

Proof.
We have

$$P(z'_{x'}|x,z) = \frac{P(z'_{x'},x,z)}{P(x,z)} = \frac{P(z'_{x'},x,z|q)P(q) + P(z'_{x'},x,z|q')P(q')}{P(x,z,q) + P(x,z,q')}$$

since $z = y \wedge q$: $q \Rightarrow (z = y)$, $q \Rightarrow (z'_{x'} = y'_{x'})$, and $q' \Rightarrow z'$

$$P(z'_{x'}|x,z) = \frac{P(y'_{x'},x,y|q)P(q) + 0}{P(x,y,q) + 0} = P(y'_{x'}|x,y)$$

$\square$

Insensitivity to alternative causes of PS

Lemma
Let $PS(x, y)$ stand for the probability that x is a sufficient cause of y and $z = y \vee r$ a consequence of y, potentially triggered by r. If $r \perp\!\!\!\perp X, Y_x, Y_{x'}$, then

$$PS(x, z) \triangleq P(z_x | x', z') = P(y_x | x', y') \triangleq PS(x, y)$$

Proof.
HOMEWORK (Hint: it's analogous to the last proof) $\square$

Exogeneity

Definition of exogeneity

Definition
A variable X is said to be exogenous to Y in model M iff

$$P(y_x, y_{x'}|x) = P(y_x, y_{x'}),$$

namely, the way Y would potentially respond to conditions x or x' is independent of the actual value of X.

Consequences of exogeneity

Under condition of exogeneity, PNS is bounded as follows:

$$\max(0, P(y|x) + P(y'|x') - 1) \leq PNS \leq \min(P(y|x), P(y'|x')),$$

while PN and PS are related to PNS as following:

$$PN = \frac{PNS}{P(y|x)}$$

$$PS = \frac{PNS}{1 - P(y|x')}$$

Monotonicity

Definition of monotonicity

Definition
A variable Y is said to be monotonic relative to variable X in a causal model M iff the junction $Y_x(u)$ is monotonic in x for all u. Equivalently, Y is monotonic relative to X if

$$y'_x \wedge y_{x'} = \mathit{false}$$

This assumption is used a lot in fields like epidemiology, it is essentially equivalent to assuming that no individual can be helped by exposure to the risk factor.

Consequences of monotonicity and exogeneity

Under both these conditions, we have that

$$PNS = P(y|x) - P(y|x'),$$

which is the risk difference often used in many fields. Additionally,

$$PN = (P(y|x) - P(y|x'))/P(y|x) = 1 - P(y|x')/P(y|x),$$

which is the excess risk ratio, and

$$PS = (P(y|x) - P(y|x'))/(1 - P(y|x')),$$

which is called the "relative difference", and is interpreted as the suceptibility of a population to risk factor x.

Non-exogenous monotonicity

We can still identify PN, PS and PNS with just monotonicity, namely

$$PNS = P(y_x, y'_{x'}) = P(y_x) - P(y_{x'})$$

$$PN = P(y'_{x'}|x, y) = \frac{P(y) - P(y_{x'})}{P(x, y)}$$

$$PS = P(y_x|x', y') = \frac{P(y_x) - P(y)}{P(x', y')},$$

we can then identify y_x if M is markovian, with the formula

$$P(y_x) = \sum_{pa_X} P(y|pa_X, x)P(pa_X)$$

Conclusion

Conclusion

Using this framework, we were able to properly define statistics commonly used in fields like epidemiology,

- Risk difference
- Excess risk ratio
- Relative difference

this lets us think about what the general case is for these kinds of statistics, along with the assumptions made to get to them.