Causal Analysis in Theory and Practice

On Mediation, counterfactuals and manipulations

Filed under: Discussion, Opinion — moderator @ 9:00 pm , 05/03/2010

Opening remarks

A few days ago, Dan Sharfstein posed a question regarding the "well-defineness" of "direct effects" in situations where the mediating variables cannot be manipulated. Dan's question triggered a private email discussion that has culminated in a posting by Thomas Richardson and Jamie Robins (below) followed by Judea Pearl's reply.

We urge more people to join this important discussion.

Thomas Richardson and James Robins' discussion:

Hello,

There has recently been some discussion of mediation and direct effects.

There are at least two issues here:

(1) Which counterfactuals are well defined.

(2) Even when counterfactuals are well defined, should we include assumptions that identify effects (ie the natural direct effect) that could never be confirmed even in principle by a Randomized Controlled Trial (RCT).

As to (1) it is clear to most that all counterfactuals are vague to a certain extent and can be made more precise by carefully describing the (quite possibly only hypothetical) intervention you want the counterfactual to represent. For this reason, whether you take manipulation or causality as ontologically primary, we need to relate causation to manipulation to clarify and make more precise which counterfactual world we are considering.

On (2) we have just finished a long paper on the issue, fleshing out considerably an argument I (Jamie) made at the American Statistical Association (in 2005) discussing a talk by Judea on natural (pure and total) direct effects.

"Alternative Graphical Causal Models and the Identification of Direct Effects"

It is available at
http://www.csss.washington.edu/Papers/wp100.pdf.

Here is a brief summary:

(continue)

The paper is long and gives various arguments pro and con for different epistemological positions. Ultimately we argue all view points are of interest as long as one is aware of the implications and justifications for each.

Here, in response to Judea, we describe one of our central arguments in the paper wherein we argue against Non-Parametric Structural Equation Models (NPSEMs) as a logical basis for counterfactual causal models, because they assume conditional independence relations that involve counterfactual variables in different possible worlds.

Specifically we argue this may lead to erroneous conclusions being drawn from causal models that have passed every direct empirical test corresponding to intervening on the original variables in the model.

In particular, in the context of the simple Directed Acyclic Graph (DAG):

X --> Z --> Y with X --> Y

the NPSEM assumes:

$Y(x=1,z) \perp\!\!\!\perp Z(x=0) \mid X=0$

and

$Y(x=1,z) \perp\!\!\!\perp Z(x=0) \mid X=1.$

This assumes independence across possible worlds, since Y(x=1,z) and Z(x=0) are never observed together. Thus the NPSEM contains assumptions that are intrinsically untestable without additional assumptions.

More significantly, these untestable assumptions lead to the identification of natural direct effects (called 'pure and total direct effect' by Jamie & Sander).

This has the following practical consequence: If someone posits the DAG:

X --> Z --> Y with X --> Y

when in fact there is confounding between Z and Y, then, if they are `unlucky', it may be the case that this confounding is intrinsically undetectable in the sense that

$p(y \mid do(x), do(z)) = p(y \mid x,z)$

$p(z \mid do(x)) = p(z \mid x)$

so that experiments intervening on x and z produce empirical results as predicted by the factorization of the causal DAG, but yet, as a consequence of the confounding between Y and Z the natural direct effect is not equal to the formula given by Judea, specifically:

E[Y(x=1,z=0)] does not equal sum_z E[Y | X=1,Z=z]f(z | X=0)

hence the assumption that the DAG is an NPSEM will have led us to an inconsistent estimate of a counterfactual contrast. To make matters worse, this error is undetectable even in principle, because, as has been pointed out by Jamie & Sander and Judea there is no intervention on the variables X, Y and Z that provides an independent means of directly estimating the natural direct effect. (Hence Judea's slogan "causation before manipulation".)

In the paper we present an alternative framework, called the Minimal Conterfactual Model (MCM), for associating counterfactual independence relations with a DAG that does not imply any untestable independence relations.

(The MCM is a minor extension of the Finest Fully Randomized Causally Interpretable Structured Tree Graph (FFRCISTG) model described in Robins (1986) and is equivalent to the FFRCISTG model for binary outcomes).

[The distinction may be made concrete as follows. For the DAG

X --> Z --> Y with X --> Y

with X,Y,Z binary both the NPSEM and the MCM contain the following 7 binary variables:

X, Z(x=0), Z(x=1), Y(0,0), Y(0,1), Y(1,0), Y(0,0).

Thus without any independence assumptions the set of distributions is of dimension (2^7) - 1 = 127.

The independence assumptions imposed by the NPSEM associated with the DAG, lead to a model of dimension 19, in contrast, the MCM independence assumptions lead to a model of dimension 113.

Hence the NPSEM is imposing 94 separate (untestable) independence constraints on the distribution of counterfactuals:

p(X, Z(x=0), Z(x=1), Y(0,0), Y(0,1), Y(1,0), Y(0,0)).]

Furthermore we obtain bounds on the natural direct effect that can be guaranteed to hold under the MCM counterfactual independence relations that are all subject to empirical test via interventions on X and Z. Thus providing 'safer' inferences for the natural direct effect than the identification formula given by Judea. (Judea's identified estimate of the natural direct effect is always contained within our bounds.)

We also show that by considering interventions on variables (not occurring in the original graph) but present on an extended graph on which certain deterministic relations hold, it is possible to obtain Judea's identification formula for the natural direct effect via the usual g-computation intervention formula i.e. the truncated factorization expression, see e.g. Pearl (2000), (1.37), p. 24.

Thus for those who find the counterfactuals involved in the natural direct effect appealing, but worry about the (untestable) epistemic commitments implicit within the NPSEM framework, we describe a means of "reducing" the pure direct effect to the classical causal DAG framework (involving intervention distributions). Our explication of the natural direct effect also provides a means of saving this effect while also maintaining the dictum "no causation without manipulation" By doing so it has the added benefit that because we need to explicitly specify a particular intervention (on the new substantive variables on the expanded graph), we can better evaluate whether the causal assumptions (embedded in the new expanded causal graph) required to identify the natural direct effect of the initial graph as an identified interventional effect on the expanded graph, are substantively plausible.

Finally, we show that if someone (not us !) has no qualms about making untestable (cross-world) counterfactual independence assumptions then by making even more assumptions than those embedded in NPSEMs it is possible to identify path specific effects that are not identified within the usual NPSEM framework, because the expressions involve what Judea has termed "recanting witnesses"; see Avin, Shpitser, Pearl, 2005 http://ftp.cs.ucla.edu/pub/stat_ser/r321-ijcai05.pdf

Thus for someone who is very courageous (heroic?) in their epistemic commitments and is not alarmed by the lack of empirical testability, there is no need to be restricted to identifying only those path specific effects given by an NPSEM. Such a person may "help themselves" to many more identification claims than offered by the standard NPSEM framework.

A key point in our development is the distinction between "actual randomization", such as occurs in a randomized trial, or sequential randomized trial, and mere "hypothesized randomization".

If randomization was actually performed then we believe there is no reason not to assume a "cross-world" independence such as:

$X \perp\!\!\!\perp Y(x=0), Y(x=1)$.

However, assumptions such as:

$Y(x=1,z) \perp\!\!\!\perp Z(x=0) \mid X=0$

and

Y(x=1,z) _||_ Z(x=0) | X=1

implied by the NPSEM:

X --> Z --> Y with X --> Y

crucially posit that it is "as if" nature has randomized Z (conditional on X), but this assumption is not subject to empirical test, because in any intervention we might perform we would not be sure that we were assigning treatment using all of the information to which nature was privy.

Though we do not mention it in the paper we conjecture that an alternative route to trying to eliminate cases in which an NPSEM that has passed all possible empirical tests conducted (through performing randomized experiments, so p(y | do(x), do(z)) = p(y | x,z) and p(z | do(x)) = p(z | x)) yet the formula for identifying the natural direct effect fails to hold, would be to make some kind of faithfulness or stability assumption. However, we do not explore this approach in our paper.

In almost all situations (possibly excepting quantum mechanics) we agree with Judea and others that it is reasonable to posit that the underlying generating mechanism is an NPSEM with unknown structure, possibly including unobserved variables. Where the methodological and epistemological differences arise is over whether it is reasonable to posit NPSEMs with only the observed variables (or the observed variables and a small number of unobserved variables), if such a hypothesis leads to untestable inference.

We would welcome comments on this paper - which is to appear in an edited volume.

Best wishes,

Jamie Robins and Thomas Richardson

Judea Pearl's reply:

1.
As to the which counterfactuals are "well defined", my position is that counterfactuals attain their "definition" from the laws of physics and, therefore, they are "well defined" before one even contemplates any specific intervention. Newton concluded that tides are DUE to lunar attraction without thinking about manipulating the moon's position; he merely envisioned how water would react to gravitaional force in general.

In fact, counterfactuals (e.g., f=ma) earn their usefulness precisely because they are not tide to specific manipulation, but can serve a vast variety of future inteventions, whose details we do not

know in advance; it is the duty of the intervenor to make precise how each anticipated manipulation fits into our store of counterfactual knowledge, also known as "scientific theories".

2.
Regarding identifiability of mediation, I have two comments to make; ' one related to your Minimal Causal Models (MCM) and one related to the role of structural equations models (SEM) as the logical basis of counterfactual analysis.

In my previous posting in this discussion I have falsely assumed that MCM is identical to the one I called "Causal Bayesian Network" (CBM) in Causality Chapter 1, which is the midlevel in the three-tier hierarchy: probability-intervention-counterfactuals?

A Causal Bayesian Network is defined by three conditions (page 24), each invoking probabilities under interventions, and none relies on counterfactuals. Given that the three conditions imply the truncated product decomposition (G-formula), I have concluded that it is the same mathematical object as MCM, and I was glad to see a counterfactual characterization of this object,

It turns out, however, (and this was brought to my attention by Thomas, yesterday) that there is a 4th layer in the hierarchy, containing those couterfactuals that can be tested empirically (i.e., reducible to do(x) expressions) yet not preresented directly in CBN. A well known example is the binary ETT (Effect of Treatment' on the Treated = P(Y(0)|X=1), see Causality p. 396-7) The idea here is that assumptions from the 1st layer of the hierarchy (i.e., binary variables is a probabilistic restriction) can combine with assumptions from the 2nd lever (i.e., interventional) to lift a third-level quantity like ETT to the second level. Such "liftable" counterfactuals may deserve to be recognized as a 4th, intermediate level, between CBN and SEM.

Note-1, Shpitzer and Pearl (2007) (in "what counterfactuals can be tested" also characterized a set of "liftable" counterfactuals, but they allowed assumptions from the 3rd level, and prohibited distribution-specific assumptions (e.g., binary, normal))

Note-2, Another "liftable counterfactual" is the Probability of Necessity (PN= P(Y(0)=0|X=1, Y=1)) which can be identified given a specific combination of experimental and observational data, see Causality p. 302-3.

3.
Regarding the role of SEM in causal modeling, I do not agree with your critique; let me explain why SEM or NPSEM remains the logical basis for counterfactual analysis. SEM represents the scientific view of nature: a collection of invariant functions (or laws) that connect variables together. To deny SEM is to deny physics and physical thinking, according to which there ARE invariant laws in nature, regardless of whether or not we can test them. In many cases the assumption that such laws exist is all that is needed, not their details.

Indeed, the laws of physics, because they are counterfactual, cannot be given empirical tests. For example, we cannot test the statement: "If the weight on this spring was different, its length would be different as well." We can never be sure that the spring is the same spring, or that the time of day did not change its elasticity. Yet, thus far, this untested Laplacian assumption of

invariance has not caused us a major disappointment. Only subatomic processes, governed by quantum mechanics managed to defy the counterfactual interpretation of physical laws (Causality, page 26).

The real question is whether reliance on SEM is more dangerous in epidemiology than our reliance on the untestable invariance of physical laws in every day life, a reliance that gave rise to the scientific revolution and, in my opinion, is not more dangerous than saying that a patient will remain the same patient even if we close our eyes for a split second.

Consider two variables, X and Y. If we assume that the laws of nature are deterministic and invariant we can write:

Y=f(x,u)

where u stands for all factors not affected by X, and then we can write:

Y(0) = f(0,u)
Y(1) = f(1,u).

Assuming that U is independent of X gives X _||_ {Y(0),Y(1)}, which asserts independence accross worlds. We see that, the untestable assumption of independence across worlds carries the same information as the standard unconfoundedness assumption that X is independent of the factor U that nature consults before determining the value of Y.

Judgmentally, this sort of assumption is routinely made whenever we deal with observational studies. True, to merely predict causal effects we can get by with weaker assumptions, expressible in do(x) notation, and verifiable (in principle) from manipulative experiments. But, realistically, human store scientific experience in the form of invariant laws of nature, not in the form of experimental probabilities, and it is this store of knowledge that we tap when we judge the plausibility of causal assumptions.

This sort of assumption is the basis for postulating unit level counterfactuals, sometimes written Y_x(u) . Indeed, what gives us the audacity to assume that Y(0) has a unique value for every unit regardless of the treatment actually received by that unit? I contend that he who resists writing the SEM equation Y = f(x,u) should also refrain from using unit-based counterfactuals as a basis for analysis. Purging NPSEM from epidemiology is tantamount to purging most counterfactuals from causal analysis, with the exception of the thin layer of "liftable" counterfactuals.

I take it, therefore, that you do not object to SEM as an analytical and conceptual tool, only to assumptions of form X_||_U, which cannot be tested by RCT.

Next let us attend to practical issues. You say "if someone posit the DAG X--->Z--->Y when in fact there is confounding between Z and Y, then...the assumption that the DAG is an NPSEM will have led us to an inconsistent estimate of a counterfactual contrast." This is a contradictory

premise. "if someone posit the DAG X--->Z--->Y when in fact there is confounding between Z and Y," then this someone cannot simultaneously assume that that DAG is an NPSEM, because if it were, the confounding between Z and Y would be representable as a latent variables affecting both. This is what NPSEM is entrusted to represent: functional, not probabilistic dependencies.

Perhaps you are concerned with the possibility that someone will posit the DAG X--->Z--->Y being unaware of of confounding that exists between Z and Y. This is a mispecification omission, not a deficiency of NPSEM as a representation scheme. It is a failure of the modeler to take advantage of the powerful machinery of NPSEM and express the existence of confounding between Z and Y.

.

In summary, I submit that SEM retains its status as the logical basis for counterfactuals. I would like, however, to go further and argue that even untestable assumptions such as X_||_U are useful. They are useful, because scientists are useful. In other words, whenever our analysis depends on scientific knowledge obtained from human beings, beware, that knowledge is already molded, filtered, contaminated, perhaps distorted by assumptions of the type X_||_U, exactly the type that you warn us to avoid.

I believe the ability to defend those assumptions on the basis of scientific theories more than makes up for our inability to verify them experimentally in each specific case. I therefore] stand behind the Mediation Formula:

Indirect Effect = sum_z E(Y(x=0,Z=z)[P(z|X=1)-P(z|x=0)]

(see: http://ftp.cs.ucla.edu/pub/stat_ser/r363.pdf), as long as I am willing to defend its underlying assumptions, here cast in a solid and transparent scientific language (i.e., NPSEM):

X=h(U_x)
Z=f(X, U_z)
Y=g(X, Z, U_y)
and {U_x,U_z, U_y} mutually independent.

Best regards,

Judea Pearl