Causality Primer: Justin Chumbley Ph.D

Information box: Defining and representing causation

Modern definitions of total, and path-specific effects - direct and indirect - are general, and not tied to any specific statistical model.

A causal variable is defined as any variable which changes the potential outcome of another variable. This idea can be interpreted as follows. First suppose we know the equations which dictate the natural directions of causation between variables in some system. Next override the equation governing one focal variable, and instead switch this variable between two different values. By definition, this focal variable is a cause of any variable which responds to this intervention (through the remaining equations). The difference between these definitions is purely notational; potential outcome definitions can easily be converted to structural definitions. Potential outcomes can be viewed as a short hand notation for general structural equations (not necessarily linear or parametric). For example, take the following trivial, linear parametric structural equation model: we can abbreviate the structural causal equations $CTRA_i(X_i = 1) = d + c + e_i$, and $CTRA_i(X_i = 0) = d + e_i$ as $CTRA_i(1)$ and $CTRA_i(0)$ respectively. Note that only one potential outcome can be observed, the other is counterfactual. Causal inference, i.e. on $CTRA_i(1) - CTRA_i(0)$, thus requires identifying conditions which justify imputing the missing counterfactual. See (Pearl 2014) to explicitly compare the structural formulation of mediation side by side with the potential outcome formulation.

Causation is defined *ceteris parabis*, i.e. at the level of each individual "unit" subjected to intervention. Various statistical methods aim to infer population parameters of these unit-level causal effects, such as propensity score matching and nearest-neighbor matching (which often uses the Mahalanobis metric, also called Mahalanobis matching), attempt to correct for the assignment mechanism by finding control units similar to treatment units on variables which confound causal effects (implied by ceteris parabis).

Information box: what is identification?

A parameter is said to be identified if different parameter settings of the underlying data generating process imply different distributions over observed variables. This identifiability - or lack thereof - is not a statistical problem related to the challenges of statistical inference with small samples. Pearl (2009) provides one way to think about identification. Dependence between observed variables reflects some unknown mix of causal and noncausal ("backdoor") effects. A causal effect is identified when the observed association can been adjusted somehow to remove these noncausal components. For nonparametric identification, the analyst would describe the set of assumptions that will allow us to identify a causal effect without any distributional or functional form assumptions.

To take a famous example, randomized treatment and the SUTVA identification (Rubin 1974) together nonparametrically identify the average total effect. To identify the indirect and direct effects, additional assumptions are necessary, e.g. "sequential ignorability".

Causal identification assumes the investigator has domain knowledge to judge the plausibility of no confounding type of assumptions which underly all mediation methods, whether under the rubric of sequential ignorability (e.g., Imai et al., 2010b), uncorrelated error terms, or graphical criterea. The assumptions identifying mediation can be stated most succinctly in the latter.

Identification conditions can be expressed in diverse ways, e.g. judging conditional independencies among counterfactual variables, often called strong ignorability, conditional ignorability , or sequential ignorability, presents a formidable task without structural models. Efforts to replace ignorability vocabulary - with notions such as no unmeasured confounders, no unmeasured confounding, as if randomized, effectively randomly assigned, or essentially random - create ambiguity. First, the notion of a confounder varies significantly from author to author. Some define a confounder (say of the NP-CTRA relationship) as a variable that affects both NP and CTRA. Some define confounder as a variable that is associated with both NP and CTRA. Others allow for a confounder to affect NP and be associated with CTRA. Worse yet, the expression no unmeasured confounders is sometimes used to exclude the very existence of such confounders and sometimes to affirm our ability to neutralize them by controlling other variables, not necessarily confounders. Second, the interpretations have taken sequential ignorability as a starting point and consequently are overly stringent – sequential ignorability is a sufficient but not necessary condition for identifying natural effects. Weaker conditions can be articulated in a transparent and unambiguous language which provide a greater identification power and a greater conceptual clarity.

Information box: Alternatives to sequential ignorability conditions for

identification

Instrumental variables offer a very different answer from a causal mediation analysis (Keele 2015). Mechanisms based on IV have the advantage that one can allow for the possibility of unobserved confounding between the mediator and the outcome. However, to identify the indirect effect, one must assume that the direct effect is zero. The assumption that the direct effect is zero is widely referred to as the exclusion restriction (Angrist, Imbens, & Rubin, 1996). Thus, one must assume that there is only an indirect effect, which implies that the effect of the treatment is entirely mediated. Under this form of mechanism, we must assume that the effect of a NP only works through Chen: There cannot be any other mechanisms for the intervention.

Statistically "controlling" for M in the analysis (by including M in the regression equation) does not physically disable the paths going through M ; it merely matches samples with equal M values, and thus induces spurious correlations among other factors in the analysis, see (Pearl 2014). This can be readily shown using classical path-tracing rules. Such dependence cannot be detected by statistical means, so theoretical knowledge must be invoked to identify the sources of these correlations and control for common causes (so called "confounders") of M and CTRA

whenever they are observable. This approach to mediation has two major drawbacks. One (mentioned above) is its reliance on the untested assumption of uncorrelated errors, and the second is its reliance on linearity and, in particular, on a property of linear systems called effect constancy (or no interaction): The effect of one variable on another is independent of the level at which we hold a third. This property does not extend to nonlinear systems; in such systems, the level at which we control M would in general modify the effect of T on CTRA. For example, if the output CTRA requires both T and M to be present, then holding M at zero would disable the effect of T on CTRA , while holding M at a high value would enable the latter.

Information Box: Modern mediation

Although one could define mediation statistically, we follow the causal definition.

The conventional mediation analysis entails fitting a set of linear regression models: ``mediation effects'' are defined in terms of these estimated model parameters. One problem with *defining* mediation in terms of statistical changes induced by adding a third mediator variable into a regression equation, is that mediation is inherently a causal notion hence should not be defined in statistical terms. Modern approaches therefore define mediation in terms of potential outcomes, or equivalently causal graphs. In the language of the latter, a mediator is then an intermediate variable that lies on the causal path from the treatment to the outcome. This definition is grounded in the notion of a causal path and emphasizes the difference between "fixing a variable" and "statistically adjusting for" (conditioning on) a variable as in regression.

To illustrate our measure of ACME more formally, consider a binary measure of negative parenting, a variable we call t which takes 0 or 1. We will now define indirect effect of NP - via mediator Chen M - within the modern framework. $M_i(t)$ is the effect of NP on Chen for subject i under treatment (NP) status t. Let $CTRA_i(t, m)$ denote the potential outcome if NP and Chen took values t, m respectively. We only observe one of these potential outcomes $CTRA_i(t_i, M_i(t_i))$, where $M_i(t_i)$ is the observed value of Chen at the observed NP level t_i. $CTRA_i(t, M_i(t))$ is the effect of t on CTRA, which in general and be transmitted both indirectly, through $M_i(t)$, and "directly" (i.e. not through M but possibly through some independent mediators). Let the total causal effect for unit (subject) i be

$$\tau_i = CTRA_i(1, M_i(1)) - CTRA_i(0, M_i(0))$$

and the unit-level indirect effect be

$$\delta_i = CTRA_i(t, M_i(1)) - CTRA_i(t, M_i(0)).$$

This latter relates to the following counterfactual question: how would CTRA change in this indidividual if we were to physically (counterfactually) change Chen's value under $t = 0$ (no negative parenting) to that under $t = 1$ (negative parenting), while keeping NP at its observed value t? Because these two values of Chen would naturally occur as responses to changes in NP, this quantity formalizes the notion of a causal mechanism that the causal effect of the treatment is transmitted through changes in the mediator of interest. Similarly, we define the

unit direct effect, corresponding to all other possible causal mechanisms (sometimes refered to en masse as the "direct effect"), as:

$$\gamma_i = CTRA_i(1, M_i(t)) - CTRA_i(0, M_i(t)).$$

The counterfactual question here is: how would CTRA respond to NP change $T_i = 0$ to $T_i = 1$, if (counterfactually) Chen was held constant?

Mediation analysis creates an identification problem. The quantity $CTRAi(1, M_i(0))$, for example, is unobservable, but to estimate the mediation effect we need assumptions which link this unobserved counterfactual to observed quantities. We examine these assumptions.

Such definitions can easily be extended to continuous treatments (NP not binary) (Imai, Keele, and Yamamoto 2010).

References

Baron, Reuben M, and David A Kenny. 1986. "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51 (6): 1173–82. *http://webcom.upmf-grenoble.fr/LIP/Perso/DMuller/GSERM/Articles/Journal of Personality and Social Psychology 1986 Baron.pdf*.

Bolstad, B M, R A Irizarry, M Astrand, and T P Speed. 2003. "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias." *Bioinformatics* 19 (2): 185–93. *http://www.stat.berkeley.edu/bolstad/normalize/*.

Bullock, John G., Donald P. Green, and Shang E. Ha. 2010. "Yes, but what's the mechanism? (don't expect an easy answer)." *Journal of Personality and Social Psychology* 98 (4). American Psychological Association: 550–58. doi:*10.1037/a0018933*.

Goeman, Jelle J, and Peter Bühlmann. 2007. "Analyzing gene expression data in terms of gene sets: methodological issues." *Bioinformatics (Oxford, England)* 23 (8). Oxford University Press: 980–7. doi:*10.1093/bioinformatics/btm051*.

Goeman, Jelle J., Sara A. van de Geer, and Hans C. van Houwelingen. 2006. "Testing against a high dimensional alternative." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68 (3). Blackwell Publishing Ltd: 477–93. doi:*10.1111/j.1467-9868.2006.00551.x*.

Greenland, Sander, and James M Robins. 1986. "Identifiability, Exchangeability, and Epidemiological Confounding." *International Journal of Epidemiology* 15 (3): 413–19. doi:*10.1093/ije/15.3.413*.

Imai, Kosuke, Luke Keele, and Teppei Yamamoto. 2010. "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects." *Statistical Science* 25 (1): 51–71. doi:*10.1214/10-STS321*.

Irizarry, Rafael A, Benjamin M Bolstad, Francois Collin, Leslie M Cope, Bridget Hobbs, and Terence P Speed. 2003. “Summaries of Affymetrix GeneChip probe level data.” *Nucleic Acids Research* 31 (4). Oxford University Press: e15. *http://www.ncbi.nlm.nih.gov/pubmed/12582260*.

Kaufman, Jay S, Richard F Maclehose, and Sol Kaufman. 2004. “A further critique of the analytic strategy of adjusting for covariates to identify biologic mediation.” *Epidemiologic Perspectives & Innovations* 1 (4). doi:*10.1186/1742-5573-1-4*.

Keele, Luke. 2015. “Causal Mediation Analysis.” *American Journal of Evaluation* 36 (4). SAGE PublicationsSage CA: Los Angeles, CA: 500–513. doi:*10.1177/1098214015594689*.

Kraemer, Helena, Eric Stice, Alan Kazdin, David Offord, and David Kupfer. 2001. “How Do Risk Factors Work Together? Mediators, Moderators, and Independent, Overlapping, and Proxy Risk Factors.” *Am J Psychiatry* 1586 (158): 848–56. *http://ajp.psychiatryonline.org/doi/pdf/10.1176/appi.ajp.158.6.848*.

MacKinnon, David P., Chondra M. Lockwood, Jeanne M. Hoffman, Stephen G. West, and Virgil Sheets. 2002. “A comparison of methods to test mediation and other intervening variable effects.” *Psychological Methods* 7 (1). American Psychological Association: 83–104. doi:*10.1037/1082-989X.7.1.83*.

Pearl, Judea. 2009. “Causal inference in statistics: An overview.” *Statistics Surveys* 3: 96–146. doi:*10.1214/09-SS057*.

———. 2014. “Interpretation and identification of causal mediation.” *Psychological Methods* 19 (4): 459–81. doi:*10.1037/a0036434*.

Preacher, Kristopher J., and Andrew F. Hayes. 2004. “SPSS and SAS procedures for estimating indirect effects in simple mediation models.” *Behavior Research Methods, Instruments, & Computers* 36 (4). Springer-Verlag: 717–31. doi:*10.3758/BF03206553*.

Rubin, Donald B. 1974. “Estimating causal effects of treatments in randomized and nonrandomized studies.” *Journal of Educational Psychology* 66 (5): 688–701. doi:*10.1037/h0037350*.

Smyth, Gordon, Natalie Thorne, and James Wettenhall. 2005. “Limma: linear models for microarray data.” In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor.*, 397–420. Springer New York.

Sobel, Michael E. 1982. “Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models.” *Sociological Methodology* 13: 290. doi:*10.2307/270723*.

Subramanian, Aravind, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, Amanda Paulovich, et al. 2005. “Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.” *Proceedings of the National Academy of Sciences of the United States of America* 102 (43). National Academy of Sciences: 15545–50. doi:*10.1073/pnas.0506580102*.