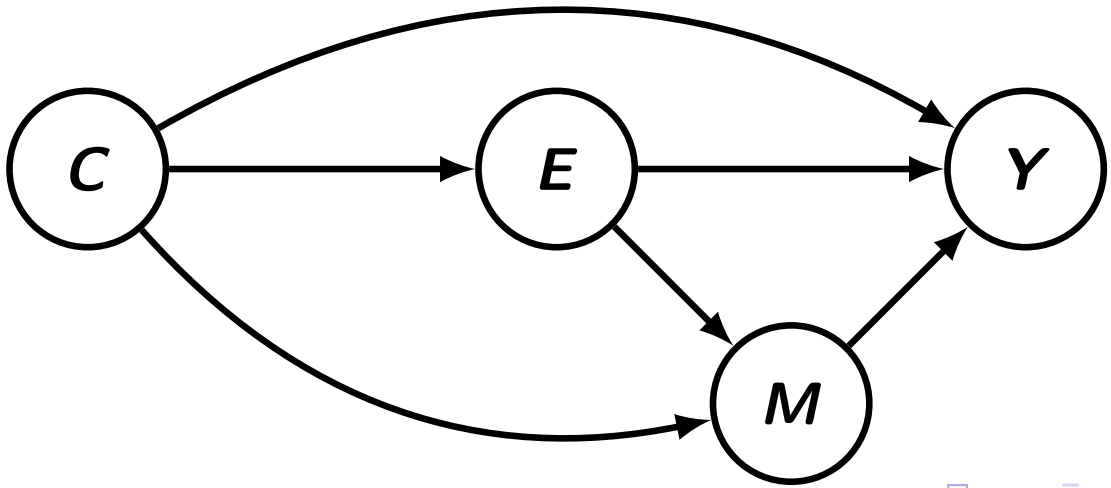
Statistical Learning
Mediation Analysis

Introduction

- An important scientific goal of studies in the health and social sciences is increasingly to determine to what extent the total effect of a point exposure is mediated by an intermediate variable on the causal pathway between the exposure and the outcome.
- Notions of direct and indirect effects are important as they provide a framework to quantify and characterize the mechanism by which an intervention affects a given outcome.
- Understanding such mechanistic pathways is of scientific interest.

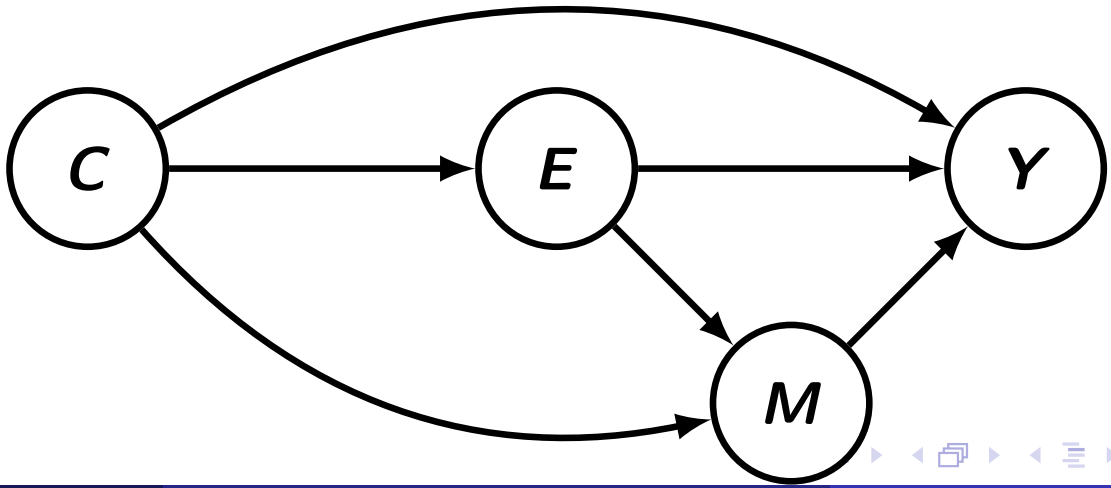
Introduction

- HIV example: consider an observational HIV study conducted in Nigeria concerned with a comparison of two highly active anti-retroviral combination therapies, to figure out which one should be preferred as a first line therapy.
- Let E indicate HAART regimen ($E = 0$: regimen 1 vs $E = 1$: regimen 2), M denotes adherence to a given prescribed regimen, Y is viral load measured after 1 year of follow-up and C is a set of pre-treatment covariates that confound the joint effects of (E, M) on Y.



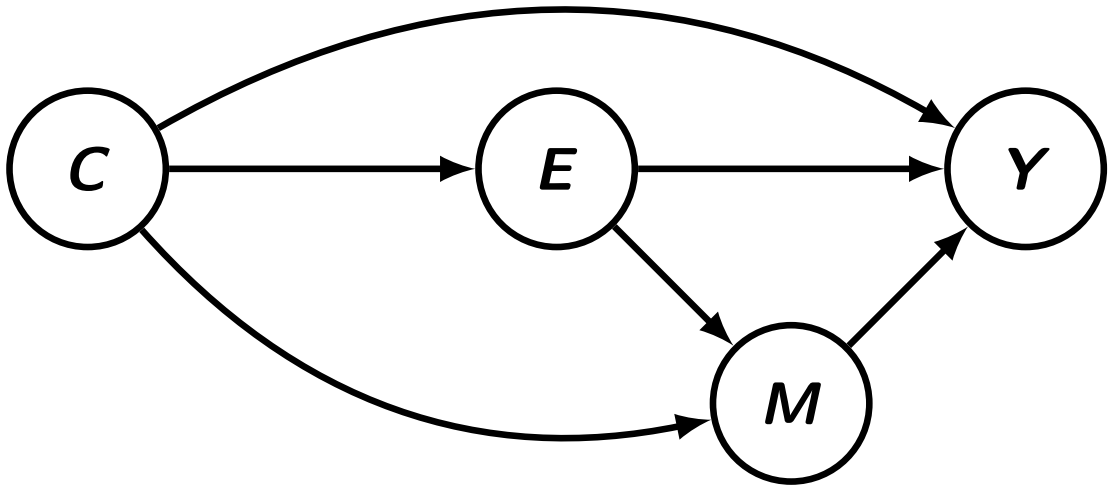
Introduction

- Suppose a non-null effect of E on Y is found, so that the two regimen are known not to be equivalent.
- It would then be of interest to understand the extent to which this observed difference may be due to a difference in the adherence rate of patients on one regimen vs another, or to a differential effect of HAART regimen not mediated by adherence.
- Such understanding would tell us whether eliminating differences in adherence rate between the two regimen could in fact eliminate differences in the effects of the two regimen, or whether such differences would prevail upon eliminating differential adherence.



Direct and indirect effects

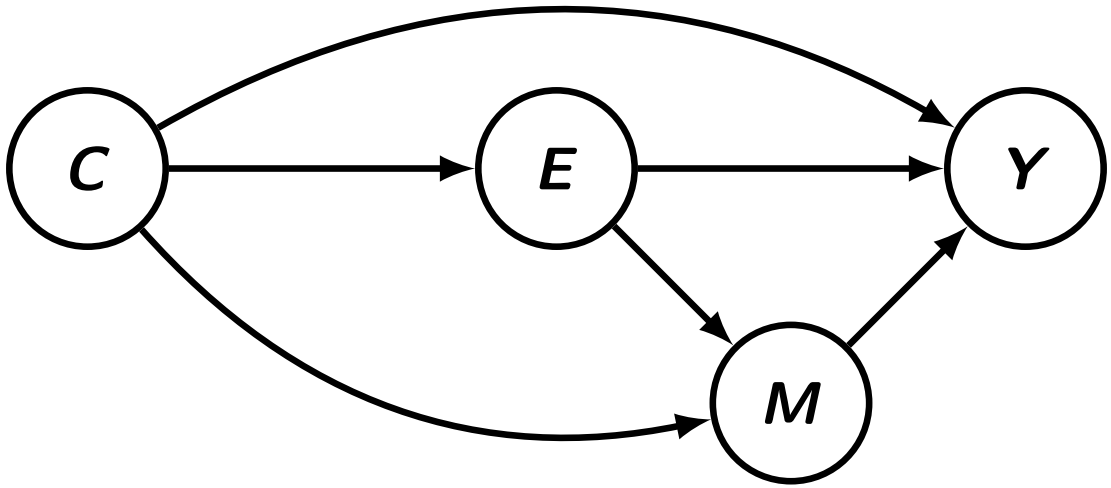
- The differential effect of HAART regimen on viral load is known in mediation analysis as the average total effect of E on Y (adjusting for C)
- The average total effect of E on Y includes two possible causal paths from E to Y,
  - the path $E \rightarrow Y$ is the direct effect of E on Y relative to M.
  - the path $E \rightarrow M \rightarrow Y$ is the indirect effect of E on Y through M.



Classical approach

- Traditionally, direct and indirect effects have been evaluated in health and social sciences using linear models specifications.
- Mainly, they considered three models for M and Y respectively:

$$Y = \theta_0 + \theta_e E + \theta_c C + \varepsilon_Y$$
$$M = \alpha_0 + \alpha_e E + \alpha_c C + \varepsilon_M$$
$$Y = \beta_0 + \beta_e E + \beta_m M + \beta_c C + \varepsilon_Y^*$$



Classical approach

- Under linearity, the classical approach interprets θ_e in the first regression which includes E and C but not M, as the total effect, i.e., the sum of paths $E \rightarrow Y$ and $E \rightarrow M \rightarrow Y$

$$Y = \theta_0 + \theta_e E + \theta_c C + \varepsilon_Y$$

- The approach interprets β_e in the third linear model which includes E, C and M as a direct effect of E on Y not mediated by M, i.e., the path $E \rightarrow Y$.
- Under the classical approach, two different expressions have been given for the indirect effect:
  - Difference formula: $\theta_e - \beta_e$
  - Product formula: $\beta_m \alpha_e$, where

$$M = \alpha_0 + \alpha_e E + \alpha_c C + \varepsilon_M$$
$$Y = \beta_0 + \beta_e E + \beta_m M + \beta_c C + \varepsilon_Y^*$$

Classical approach

- The two formulae for the indirect effect are in fact equal $\theta_e - \beta_e = \beta_m \alpha_e$
- These "parametric" notions of direct and indirect effects follow from the observation that:
  - $\theta_e \neq 0$ if and only if E and Y are associated given C, irrespective of whether this association is due to a direct or indirect effect.
  - In the event of no direct effect, the total effect should equate the indirect effect, i.e., $\theta_e - \beta_e = \theta_e$ since $\beta_e = 0$.
  - If either E does not affect M $(\alpha_e = 0)$ or M does not affect Y $(\beta_m = 0)$, it must be that the indirect effect is also null $\beta_m \alpha_e = 0$ and the total effect (θ_e) and direct (β_e) are equal $\theta_e - \beta_e = 0$

$$
\begin{aligned}
Y &= \theta_0 + \theta_e E + \theta_c C + \varepsilon_Y \\
M &= \alpha_0 + \alpha_e E + \alpha_c C + \varepsilon_M \\
Y &= \beta_0 + \beta_e E + \beta_m M + \beta_c C + \varepsilon_Y^*
\end{aligned}
$$

Classical approach

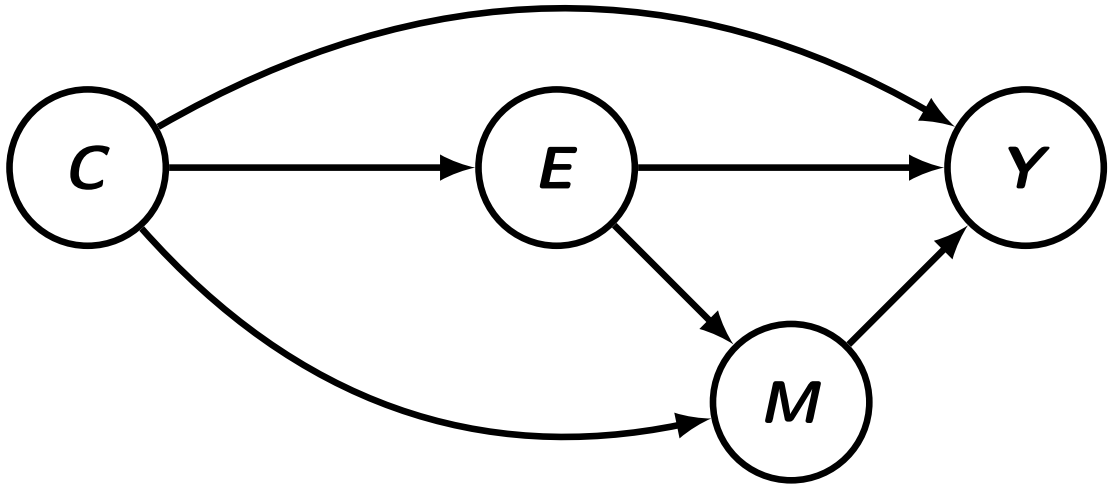
- The classical approach is widely used, for example, one of the original papers on the approach by Baron and Kenny (1984) has over 30000 citations.
- However, the approach has some serious limitations:
  - It relies on a linear specification of models for the observed data, and assumes no interaction or nonlinearities involving E and M.
  - Thus restricted to simple problems with continuous outcome and mediator.
  - Because these parametric definitions of direct and indirect effects rely on the specific linear model, it is unclear what assumptions about confounding are needed for these effects to be considered causal.
  - It is not clear whether the parametric assumptions can be relaxed and whether nonparametric notions of direct and indirect effects are possible, which can be used for binary, count, or time to event outcomes.

Causal mediation

- There has recently developed a fast growing literature in causal inference concerned with the definition, identification and estimation of direct and indirect effects in fully nonparametric models.
- The recent literature uses the language of potential outcomes/counterfactuals to give a nonparametric (i.e., that do not rely on a linear model specification) definition of effects involved in mediation analysis known as controlled direct effects, natural direct and indirect effects, and path-specific effects.
- These effects, despite being defined in a fully nonparametric way, can nevertheless be sometimes identified and estimated from observational data.
- Conditions for identification of these mediation effects have been made explicit using the counterfactual language of causal inference.

Causal mediation

- Consider the previous DAG



- The counterfactual definitions of direct and indirect effects requires defining the counterfactuals:
  - Y_e :?
  - Y_{em} :?
  - M_e :?
- Note that under a slight strengthening of consistency assumption, $Y_e = Y_{eM_e}$, which explains why generally, $Y_e \neq Y_{em}$

Controlled direct effect

- The controlled direct effect refers to the exposure effect that arises upon intervening to set the mediator to a fixed level that may differ from its actual observed value. For instance for E and M binary, the individual controlled direct effect for a fixed value m is given by:

$$Y_{1m} - Y_{0m}$$

- We cannot hope except possibly under strong assumptions to be able to identify individual CDE, but we can also define the population average CDE:

$$\text{CDE}(m) = E\left(Y_{1m} - Y_{0m}\right)$$

- This effect has a policy implication, e.g., $\text{CDE}(0) = E\left(Y_{10} - Y_{00}\right)$ is the direct effect one would observe if we were to completely remove the mediator from the underlying population.
- In HIV example, what would be the differential effect of two HAART regimen in the population if one could force all individuals to fully adhere (i.e., intervene to remove non-adherence).

Controlled direct effect

- While controlled direct effect is informative about possible intervention on the mediator, it is only potentially relevant for settings where one can design an intervention to force the mediator to take on a certain value for all individuals.
- This may not be reasonable for certain biological mediators such as say CD4 count, cholesterol, etc... Also, even if one can imagine intervening on the mediator, such interventions for behavioral mediators such as adherence, would need to be carefully delineated since these can possibly be manipulated in a number of ways.
- Not immediately clear whether there is a mechanistic notion of indirect effect as a counterpart of the controlled direct effect.

Natural effects

- Consider the following decomposition of the individual total effect:

$$
\begin{aligned}
TE &= \overbrace{Y_{e=1} - Y_{e=0}}^{total\ \ effect} = Y_{e=1,M_{e=1}} - Y_{e=0,M_{e=0}} \\
&= \overbrace{Y_{e=1,M_{e=0}} - Y_{e=0,M_{e=0}}}^{natural\ \ direct\ \ effect} + \overbrace{Y_{e=1,M_{e=1}} - Y_{e=1,M_{e=0}}}^{natural\ \ indirect\ \ effect} \\
&= \text{NDE} + \text{NIE}
\end{aligned}
$$

- This decomposition is obtained upon adding and subtracting the term $Y_{e=1,M_{e=0}}$, the counterfactual outcome we would observe in a world where we intervene to set E to 1 and M is made to take the value it would naturally have under no exposure $M_{e=0}$.

NDE

$$\overbrace{Y_{e=1} - Y_{e=0}}^{total\ \ effect} = \overbrace{Y_{e=1,M_{e=0}} - Y_{e=0,M_{e=0}}}^{natural\ \ direct\ \ effect} + \overbrace{Y_{e=1,M_{e=1}} - Y_{e=1,M_{e=0}}}^{natural\ \ indirect\ \ effect}$$

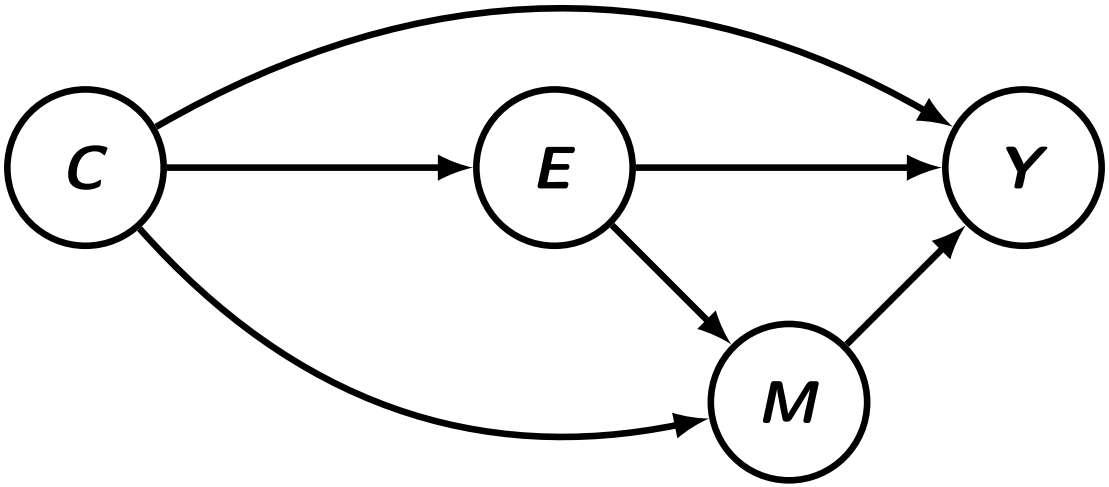
- The term $Y_{e=1,M_{e=0}} - Y_{e=0,M_{e=0}}$ compares, upon holding the mediator to its natural value under no treatment, the value of the outcome under treatment and control conditions. Because the mediator value is held fixed for the individual in the contrast, this is the individual natural direct effect, since it will only be non-null if the exposure continues to have an effect on the outcome without allowing the mediator to vary. The population version of this effect is $NDE = E\left(Y_{e=1,M_{e=0}} - Y_{e=0,M_{e=0}}\right)$.

NIE

$$\overbrace{Y_{e=1} - Y_{e=0}}^{total\ \ effect} = \overbrace{Y_{e=1,M_{e=0}} - Y_{e=0,M_{e=0}}}^{natural\ \ direct\ \ effect} + \overbrace{Y_{e=1,M_{e=1}} - Y_{e=1,M_{e=0}}}^{natural\ \ indirect\ \ effect}$$

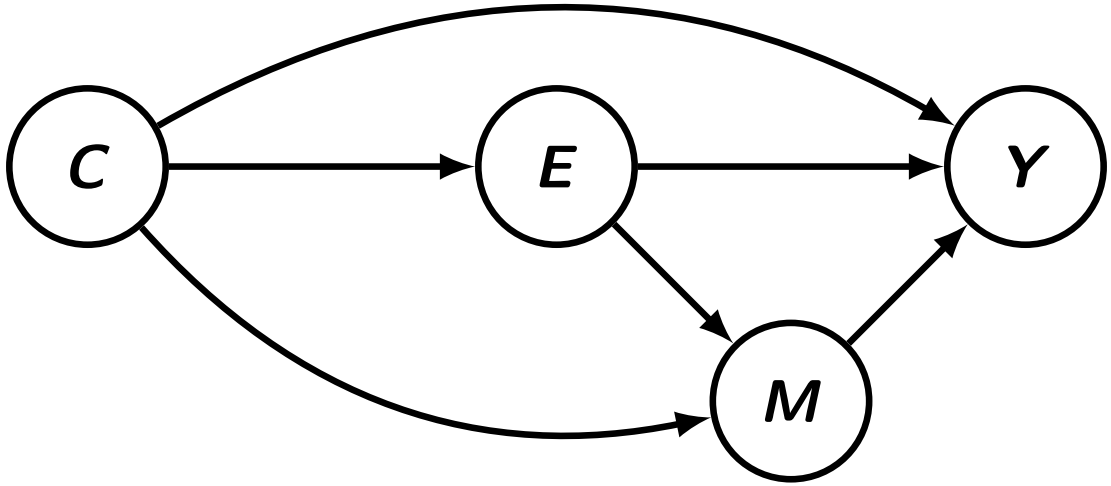
- The second term $Y_{e=1,M_{e=1}} - Y_{e=1,M_{e=0}}$ compares, upon setting the exposure to be one, the outcome under the natural value the mediator would take with and without the active exposure. Because this contrast will be non-null only if the exposure has an effect on the mediator, which in turn has an effect on the outcome, it is known as the individual natural indirect effect. The population version of this effect is $NIE = E\left(Y_{e=1,M_{e=1}} - Y_{e=1,M_{e=0}}\right)$
- An advantage of NDE is that it adds up with NIE to produce the total effect TE

Controlled vs natural direct effects



- Although the existence of an individual with non-null controlled direct effect $Y_{1m} - Y_{0m}$ and the existence of an individual with non-null natural direct effect $Y_{e=1,M_{e=0}} - Y_{e=0,M_{e=0}}$ are both depicted graphical by the presence of the path $E \rightarrow Y$, these are generally two different notions of direct effect.

Controlled vs natural direct effects



- To see why, note that the intervention that produces the CDE removes any variation in M, so that say $M = 0$ for all individuals, while the NDE allows for variation in values of M across individuals which is a reflection of variation in M in the absence of the exposure, thus some individuals will have $M_{e=0} = 0$ and some $M_{e=0} = 1$
- The first intervention removes the mediator, while the second type of intervention merely deactivates the mediated pathway.

HIV example

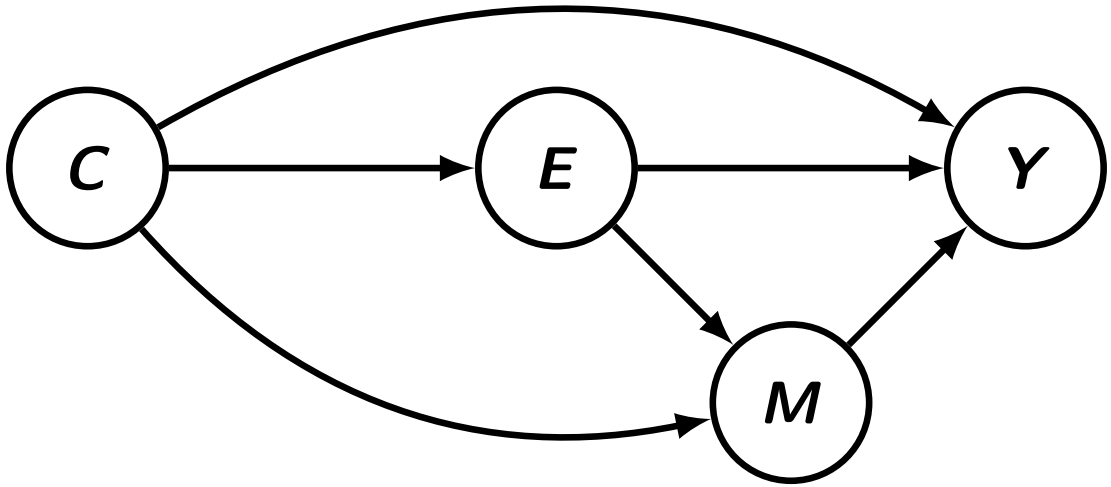
- In the HIV example, NDE is the differential effects of HAART regimens if one were to intervene and force the adherence of all individuals to be what it would be under the reference HAART regime.
- To think of such a physical intervention may be challenging in practice, and would possibly require isolating and removing the specific component of the second HAART regimen that is the source of a difference in adherence when compared to the reference regimen.
- May not be an easy task, thus NDE is not particularly useful to consider from an interventionist or policy point of view and can simply be understood from a mechanistic perspective as effect explanation. e.g., proportion mediated = NIE / TE.

NIE

- Consider the NIE, it can be shown with some algebra that

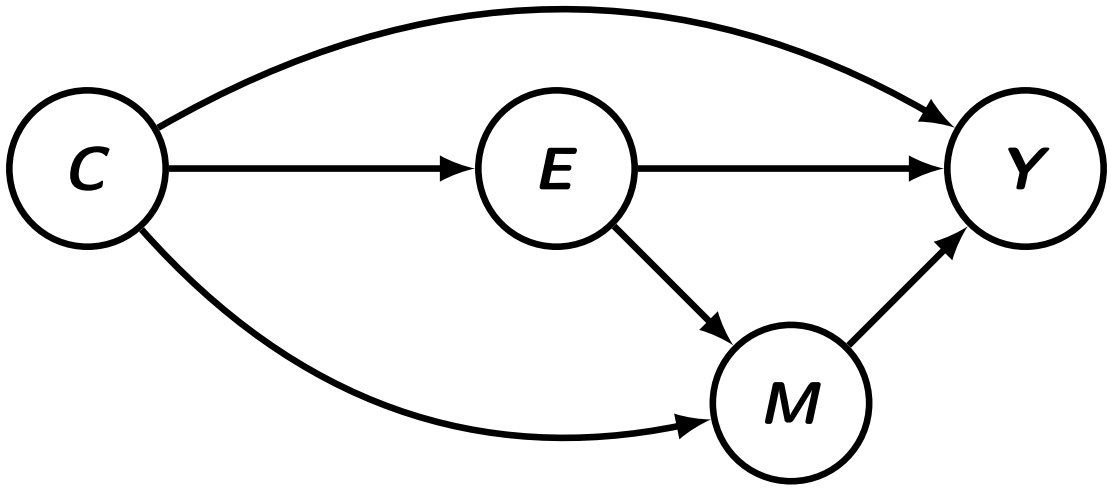
$$\begin{aligned}\text{NIE} =& E\left(Y_{e=1,M_{e=1}} - Y_{e=1,M_{e=0}}\right) \\ =& E\left(Y_{e=1} - Y_{e=0}\right) - E\left(Y_{e=1,m=0} - Y_{e=0,m=0}\right) \\ & - E\left\{\left(Y_{e=1,m=1} - Y_{e=1,m=0} - Y_{e=0,m=1} + Y_{e=0,m=0}\right) M_{e=0}\right\}\end{aligned}$$

Therefore identification of NIE requires identification of $E\left(Y_{e=1} - Y_{e=0}\right)$, is this identified in the DAG below?



NIE

- Next identification of NIE requires identification of $E\left(Y_{e=1,m=0} - Y_{e=0,m=0}\right)$, is this identified in the DAG below?

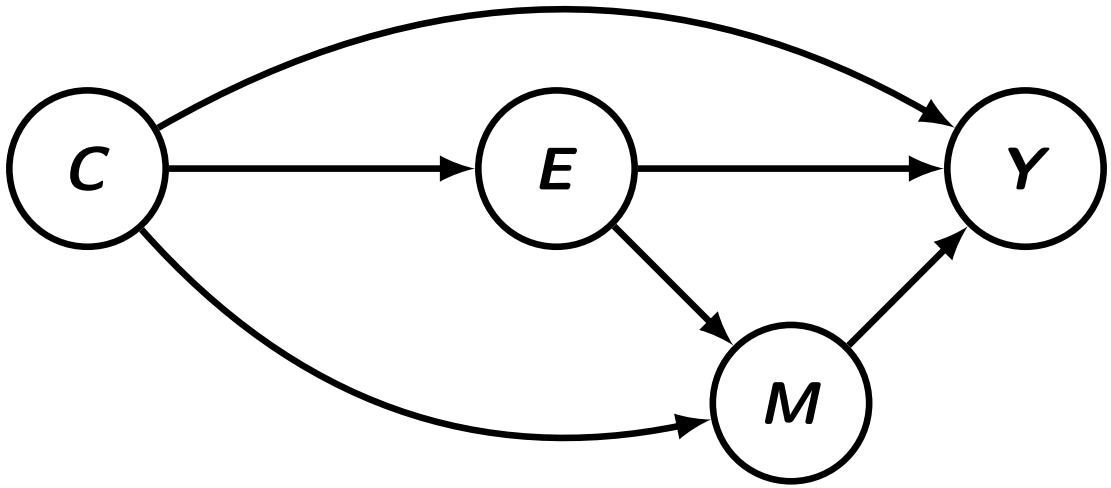


NIE

- Finally, note that identification of NIE requires identification of the last term

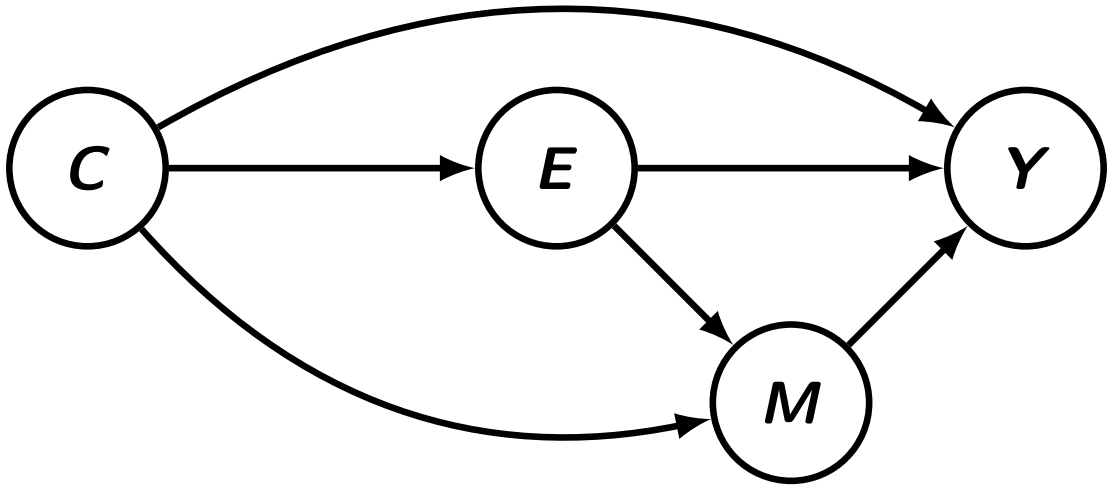
$$E\left\{\left(Y_{e=1,m=1}-Y_{e=1,m=0}-Y_{e=0,m=1}+Y_{e=0,m=0}\right)M_{e=0}\right\}$$

- This term has an interesting interpretation, note that the random variable is only non-zero for individuals with non-null causal additive interaction $\left(Y_{e=1,m=1}-Y_{e=1,m=0}-Y_{e=0,m=1}+Y_{e=0,m=0}\right)$ and with $M_{e=0}=1$.



NIE

- It can be shown that this last term is identified under one of the following assumptions,
  (1) $M_{e=0} = 1$ for all individuals in the population
  (2) $M_{e=0} = 0$ for all individuals in the population
  (3) $Y_{e=1,m=1} - Y_{e=1,m=0} - Y_{e=0,m=1} + Y_{e=0,m=0} = 0$ for all individuals with $M_{e=0} = 1$
  (4) $Y_{e=1,m=1} - Y_{e=1,m=0} - Y_{e=0,m=1} + Y_{e=0,m=0} \perp\!\!\!\perp M_e \mid C$



NIE

- Assumptions (1) or (2) might sometimes apply but not in general.
- For example suppose E: Maternal HIV status; M: maternal use of anti-retroviral therapy during pregnancy; Y: Birth outcome, then $M_{e=0} = 0$ as mothers who are not HIV positive are not offered ART. Note: This may no longer hold in era of Pre-exposure prophylaxis (or PrEP), i.e., when people at very high risk for HIV take HIV medicines daily to lower their chances of getting infected.
- Assumption (3) is an assumption of no individual causal interaction between E and M for the subset of the population with $M_{e=0} = 1$, it is a strong assumption which is unlikely to hold exactly in most applications.
- The fourth assumption would hold if

$$Y_{e^*,m} \perp\!\!\!\perp M_e \mid C$$

for $e, e^* = 0, 1$ and $m = 0, 1$.

Cross-world independence

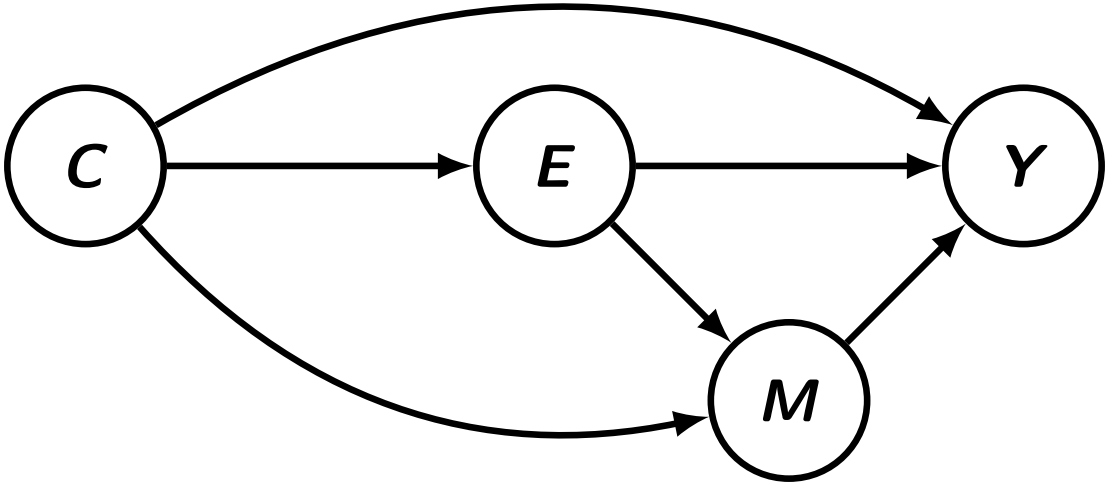
- This assumption is a cross-world independence assumption that is now routinely made in modern causal mediation analysis, and it would hold under a so-called nonparametric structural equation interpretation of the causal DAG.
- The "cross-world" label comes from the fact that one is relating counterfactual variables under conflicting potential treatment values $Y_{e^*,m}$ and M_e, and therefore the assumed independence cannot be imposed experimentally, say by randomization.
- This in contrast to standard no unmeasured confounding assumption such as say $Y_e \perp\!\!\!\perp E \mid C$ which in principle can be enforced by randomization.

Identification

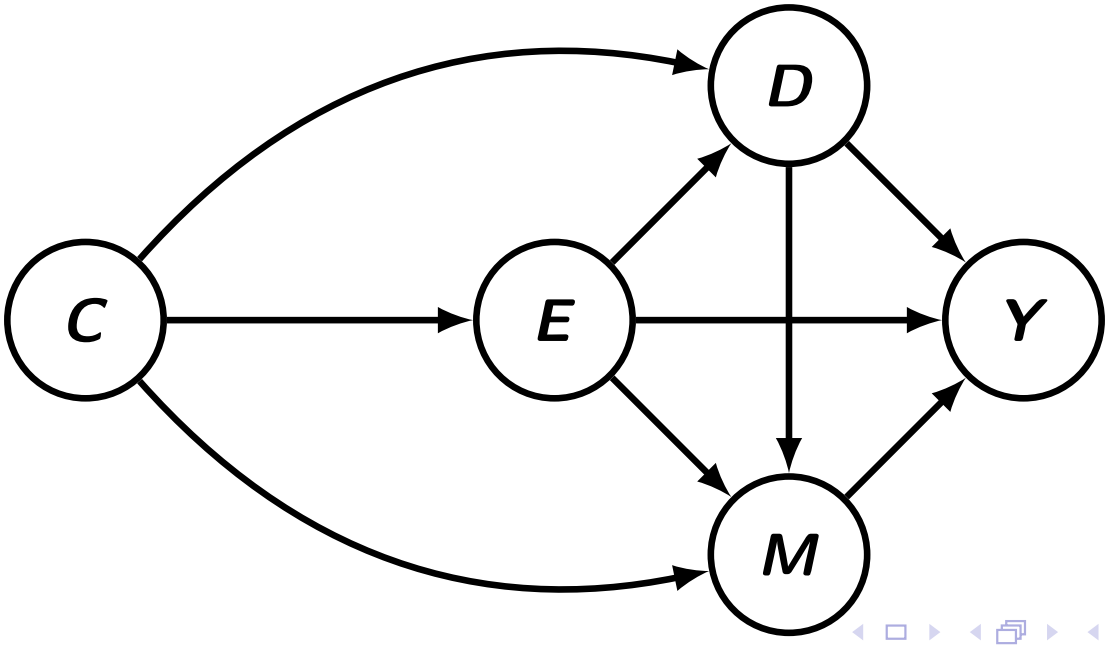
The nonparametric assumptions needed for identification of NDE and NIE under the causal DAG are that upon conditioning on pre-exposure variables C,

1. There is no unobserved confounding of the effects of E on M
2. There is no unobserved confounding of the effects of E on Y
3. There is no unobserved confounding of the effects of M on Y
4. There is no confounder of the effect of M on Y that is affected by E

Violation of Assumption 4 so that NDE is not identified



vs



Mediation formula

- Under Assumptions 1-4, Pearl (2001) established that for $e, e^* = 0, 1$

$$E\left(Y_{e,M_{e^*}}\right) = \sum_{m,c} E(Y \mid m, e, c) f\left(m \mid e^*, c\right) f(c)$$

and therefore

$$\begin{aligned} & NDE \\ =& E\left(Y_{e=1,M_{e=0}}\right) - E\left(Y_{e=0,M_{e=0}}\right) \\ =& \sum_{m,c} \{E(Y \mid m, e=1, c) - E(Y \mid m, e=0, c)\} f(m \mid e=0, c) f(c) \end{aligned}$$

and

$$\begin{aligned} & NIE \\ =& E\left(Y_{e=1,M_{e=1}}\right) - E\left(Y_{e=1,M_{e=0}}\right) \\ =& \sum_{m,c} E(Y \mid m, e=1, c)\{f(m \mid e=1, c) - f(m \mid e=0, c)\} f(c) \end{aligned}$$

Baron and Kenny

- Under linear models

$$M = \alpha_0 + \alpha_e E + \alpha_c C + \varepsilon_M$$
$$Y = \beta_0 + \beta_e E + \beta_m M + \beta_c C + \varepsilon_Y^*$$

and Assumptions 1-4, the mediation formula gives

$$NDE = \beta_e$$
$$NIE = \alpha_e \beta_m$$

reducing to the product rule and therefore the approach of Baron and Kenny is thus formally justified.

Mediation formula

- The advantage of the mediation formula is that unlike B&K it can be used with any choice of parametric model. For example, suppose that the linear model specification now includes an interaction between E and M,

$$M = \alpha_0 + \alpha_e E + \alpha_c C + \varepsilon_M$$
$$Y = \beta_0 + \beta_e E + \beta_m M + \beta_{em} EM + \beta_c C + \varepsilon_Y^*$$

- Under this model B&K no longer holds, yet the mediation formula gives

$$NDE = \beta_e + \beta_{em} (\alpha_0 + \alpha_c E(C))$$
$$NIE = \alpha_e (\beta_m + \beta_{em})$$

which appropriately accounts for the interaction