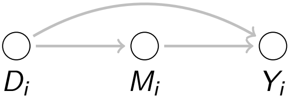
Mediation Analysis

Ye Wang
University of North Carolina at Chapel Hill

Advanced Topics in Causal Inference
POLI891

Mediation

- Researchers are often interested in mechanisms underlying a causal relationship:



- Variables that stand for mechanisms are known as "mediators".
- Sailors know for a long time that eating fruits prevents you from getting scurvy.
- But only fresh fruits are effective as they contain vitamin C.
- Isolating such mechanisms is thus crucial for policy interventions.
- They may also deepen our understanding of the world.
- E.g., how does a message shown to the respondents change their opinion?
- Does it increase their knowledge, change their belief, or evoke certain emotions?

Define the mediation effect

- Consider a sample with N units, for each we observe Y_i, $D_i \in \{0, 1\}$, and a mediator M_i.
- The outcome Y_i is decided by both D_i and M_i: $Y_i = Y_i(D_i, M_i)$.
- Therefore multiple potential outcomes for each unit i:

$$Y_i = \begin{cases} Y_i(1, m), D_i = 1, M_i = m \\ Y_i(0, m), D_i = 0, M_i = m \\ Y_i(0, m'), D_i = 0, M_i = m' \\ Y_i(1, m'), D_i = 1, M_i = m'. \end{cases}$$

- The mediator's value is decided by D_i hence post-treatment:

$$M_i = \begin{cases} M_i(1), D_i = 1 \\ M_i(0), D_i = 0. \end{cases}$$

Define the mediation effect

- We can define the total effect for unit i as

$$\tau_{i,total} = Y_i(1, M_i(1)) - Y_i(0, M_i(0)).$$

- The natural mediation effect is

$$\tau_{i,nm}(d) = Y_i(d, M_i(1)) - Y_i(d, M_i(0)).$$

- The natural direct effect is

$$\tau_{i,nd}(d) = Y_i(1, M_i(d)) - Y_i(0, M_i(d)).$$

- We can see that $\tau_{i,total} = \tau_{i,nd}(d) + \tau_{i,nm}(1-d)$.

Define the mediation effect

- Recall our previous example where i stands for a country.
- D_i means whether country i has a high ethnic diversity; M_i indicates whether the country is developed; Y_i is the frequency of civil conflicts.
- The total effect captures the effect on civil conflicts generated by ethnic diversity through all possible channels.
- The mediation effect: the effect on civil conflicts when economic development changes from the level under control to the level under treatment, while ethnic diversity is fixed at d.
- Note that it differs from $Y_i(d, 1) - Y_i(d, 0)$.
- The direct effect: the effect of ethnic diversity on civil conflicts when development is fixed at the level under d.

Define the mediation effect

- The average total effect is $\tau = E[\tau_{i,total}]$, which equals the ATE.
- Similarly, the average natural direct effect is $\tau_{ANDE}(d) = E[\tau_{i,nd}(d)]$, and the average natural mediation effect is $\tau_{ANME}(d) = E[\tau_{i,nm}(d)]$.
- Imai, Keele, and Tingley (2010) call them "average direct effect" (ADE) and "average causal mediation effect" (ACME), respectively.
- The same decomposition holds

$$\tau = \tau_{ADE}(d) + \tau_{ACME}(1-d).$$

Identify the mediation effect

- For simplicity, let's assume that D_i is randomly assigned:

$$D_i \perp \{Y_i(1,m), Y_i(0,m), M_i(0), M_i(1)\},$$
$$\varepsilon < P(D_i = 1) < 1 - \varepsilon.$$

- It is sufficient to identify the average total effect and the ATE on the mediator.
- To identify the ADE or ACME, we need to further assume that

$$M_i(d) \perp \{Y_i(1,m), Y_i(0,m)\} | D_i = d,$$
$$\varepsilon < P(M_i(d) = m) < 1 - \varepsilon.$$

- Imai, Keele, and Tingley (2010) call the two assumptions "sequential ignorability".
- It is different from what we saw in panel data analysis.

Identify the mediation effect

- We can easily estimate $E[Y_i(1, M_i(1))]$ and $E[Y_i(0, M_i(0))]$.
- Sequential ignorability allows us to estimate $E[Y_i(1, M_i(0))]$ and $E[Y_i(0, M_i(1))]$.
- Note that the assumption requires the manipulation of $M_i(d)$ rather than M_i.
- It cannot be simply guaranteed "by design" as the value of potential outcomes cannot be altered.
- Ideally, we need at least three parallel universes.
- In universe one, everyone is assigned with fresh oranges, and we observe $M_i(1)$ and $Y_i(1, M_i(1))$.
- In universe two, no one is assigned with fresh oranges, and we observe $M_i(0)$ and $Y_i(0, M_i(0))$.
- In universe three, everyone is assigned with fresh oranges, and we fix their level of vitamin C at $M_i(0)$ and observe $Y_i(1, M_i(0))$.
- The difference between universes one and three captures the ACME.

Identify the mediation effect

- Therefore, the assumption of sequential ignorability is also called "cross-world independence."
- Think of the following experiment: we randomize two treatments, fresh orange and vitamin C, simultaneously.
- Can we identify the ADE or ACME from the experiment?
- No, as what we have is

$$(D_i, M_i) \perp \{Y_i(d, m)\}, \text{ and } D_i \perp M_i.$$

- We are randomizing M_i rather than its potential level, $M_i(d)$.
- We can identify $E[Y_i(d, m)]$ for each combination (d, m).
- Additional assumptions are needed for identifying the ACME.

Estimate the mediation effect

- In reality, randomizing $M_i(0)$ or $M_i(1)$ is impossible as we do not know their values for everyone.
- We can only design experiments to identify the ACME indirectly under structural restrictions.
- Consider the parallel designs proposed by Imai, Tingley, and Yamamoto (2013).
- The idea is to estimate first τ and $\tau_{ADE}(d)$, and use their difference as an estimate of $\tau_{ACME}(1-d)$.
- We randomly divide the sample into two groups, G_1 and G_2.
- D_i is randomized in G_1, while both D_i and $M_i \in \{0, 1\}$ are randomized in G_2.
- We do not assume sequential ignorability.
- From G_1, we can estimate τ as before.

Estimate the mediation effect

- To estimate $\tau_{ADE}(0)$ or $\tau_{ADE}(1)$, we need a restriction that $Y_i(1,1) - Y_i(1,0) = Y_i(0,1) - Y_i(0,0)$ (no interaction).
- It implies that $\tau_{ADE}(0) = \tau_{ADE}(1) = \tau_{ADE}$.
- Define $p = P(D_i = 1)$ and $q = P(M_i = 1)$ in G_2, then

$$\begin{aligned}\hat{\tau}_{ADE} =& \frac{1}{N}\sum_{i=1}^{N}\frac{D_i M_i Y_i}{p} - \frac{1}{N}\sum_{i=1}^{N}\frac{(1-D_i)M_i Y_i}{1-p} \\ &+ \frac{1}{N}\sum_{i=1}^{N}\frac{D_i(1-M_i)Y_i}{p} - \frac{1}{N}\sum_{i=1}^{N}\frac{(1-D_i)(1-M_i)Y_i}{1-p}.\end{aligned}$$

- This estimator identifies

$$qE\left[Y_i(1,1) - Y_i(0,1)\right] + (1-q)E\left[Y_i(1,0) - Y_i(0,0)\right] = \tau_{ADE}.$$

Estimate the mediation effect

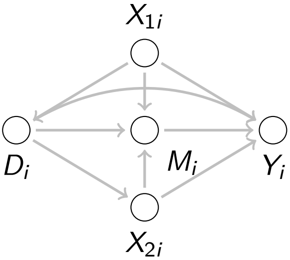
- Sequential ignorability is necessary in observational studies.
- The classical approach (Baron and Kenny 1986) is built upon the following linear models:

$$Y_i = \tau D_i + \beta M_i + \varepsilon_i,$$
$$M_i = \delta D_i + \nu_i.$$

- It assumes linearity, no interaction, and homogeneous treatment effects.
- Imai, Keele, and Yamamoto (2010) show that if all these restrictions are satisfied, then $\tau_{ACME} = \delta * \beta$ and $\tau_{ADE} = \tau$.
- It is straightforward to extend the first equation and assume $Y_i = \tau D_i + \beta M_i + \eta D_i * M_i + \varepsilon_i$.
- Then, $\tau_{ACME}(0) = \delta * \beta$, $\tau_{ACME}(1) = \delta * (\beta + \eta)$, $\tau_{ADE}(0) = \tau$, and $\tau_{ADE}(1) = \tau + \eta * \delta$.

Identify the mediation effect under strong ignorability

- Previous discussions have not accounted for the existence of confounders.
- We need to distinguish different types of confounders in mediation analysis.
- Consider the following graph:



- We can assume sequential ignorability conditional on X_{1i} but not X_{2i}.

Estimate the mediation effect under sequential ignorability

- We can control for X_{1i} by adding them into the equations.
- The modern approach is built upon nonparametric regression estimators.
- We need to estimate conditional expectations such as $\delta(D_i, M_i, X_{1i}) = E[Y_i | D_i, M_i, X_{1i}]$.
- Imai, Keele, and Tingley (2010) show that

$$
\begin{aligned}
&\tau_{ACME}(d) \\
=&E\left[E_{M|D=1,X_1}[\delta(D_i, M_i, X_{1i})] - E_{M|D=0,X_1}[\delta(D_i, M_i, X_{1i})]\right]
\end{aligned}
$$

- They suggest an estimation algorithm based on Bayesian methods.
- This approach can be applied to continuous treatments or mediators.
- Sensitivity analysis is necessary to ensure that sequential ignorability holds.

The binary case

- Let's consider a simple case where there is no confounder and M_i is dichotomous.
- In this case, the previous result indicates that

$$
\begin{aligned}
\tau_{ACME}(1) =& (E[Y_i(1,1) - Y_i(1,0)]) \\
& * (P(M_i(1) = 1) - P(M_i(0) = 1)) \\
=& (E[Y_i \mid D_i = 1, M_i = 1] - E[Y_i \mid D_i = 1, M_i = 0]) \\
& * (P(M_i = 1 \mid D_i = 1) - P(M_i = 1 \mid D_i = 0))
\end{aligned}
$$

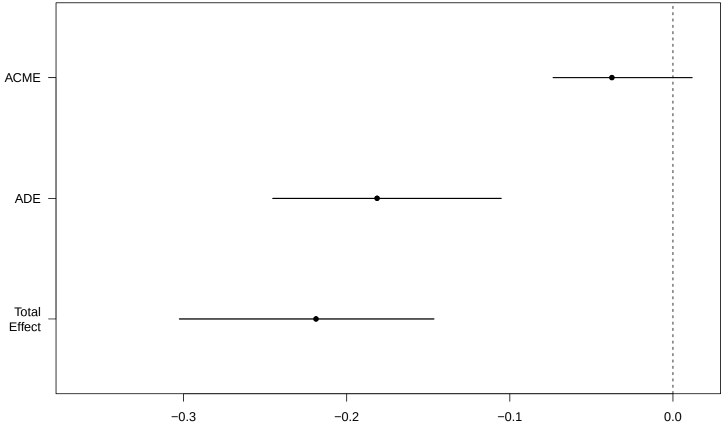
- The first equality unveils the relationship between $\tau_{ACME}(1)$ and $E[Y_i(1,1) - Y_i(1,0)]$.
- It implies that $|\tau_{ACME}(1)| \leq E[Y_i(1,1) - Y_i(1,0)]$.
- $\tau_{ACME}(0)$ can be similarly identified.

Mediation analysis: application

- Consider the study in Lupu and Peisakhin (2017), which investigated the political legacy of Stalin's deportation of the Crimean Tatars.
- The authors conducted a survey on households with senior members who were born before the deportation.
- The treatment is whether any of the family members were victims of the deportation.
- The outcome is their support for Russia's annexation of Crimea.
- Mediators include multiple indicators about their identity and feelings over generations.
- We focus on the sub-sample of the third generation in these households.

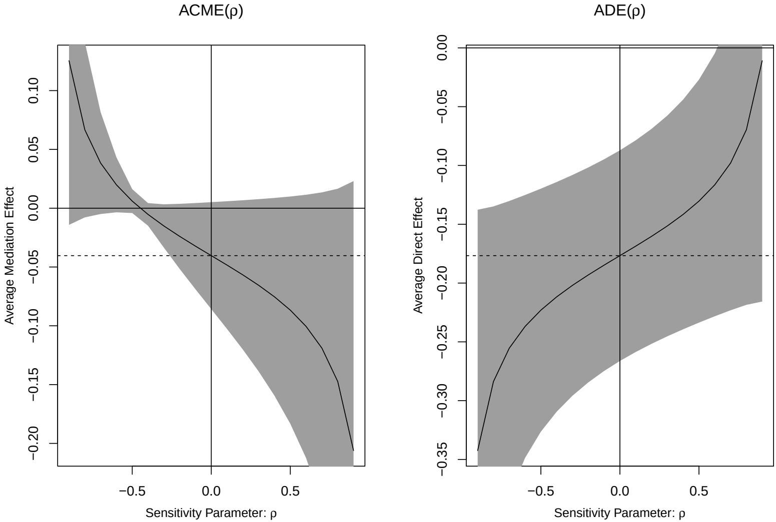
Mediation analysis: application

- We consider a single mediator, fear among the second generation in these households.



Mediation analysis: application

- The sensitivity analysis is built upon the linear model, and ρ is the correlation between ε_i and ν_i.



The average controlled direct effect

- What if we have confounders like X_{2i}?
- Acharya, Blackwell, and Sen (2016) show that we can identify a quantity known as the average controlled direct effect (ACDE):

$$\tau_{ACDE}(m) = E[Y_i(1, m) - Y_i(0, m)].$$

- We can replace the second part of sequential ignorability with

$$M_i \perp \{Y_i(1, m), Y_i(0, m)\} | D_i, X_{1i}, X_{2i}$$

- It is a familiar problem from panel data analysis.
- We need to account for the influence of the post-treatment variable X_{2i}.
- This can be done via IPW estimators.
- Or we can rely on regression models under structural restrictions.

The average controlled direct effect

- How do we interpret the ACDE?
- Acharya, Blackwell, and Sen (2016) provide the following decomposition:

$$\tau = \tau_{ACDE} + \tau_{ACME} + E\left[\tau_{i,int}\mathbf{1}\{M_i(0) = 1\}\right],$$

  where $\tau_{i,int} = Y_i(1,1) - Y_i(0,1) - (Y_i(1,0) - Y_i(0,0))$ is the interaction effect of D_i and M_i.
- In other words, τ_{ADE} equals τ_{ACDE} plus the interaction effect.
- The difference between τ and τ_{ACDE} captures two channels through which the mediator affects the outcome.
- If $\tau_{ACDE} \neq 0$, there must exist causal mechanisms (mediators) other than M_i.

Bound the mediation effect

- Recall the hypothetical experiment where we randomize two treatments simultaneously.
- We can identify each $\mu_{dm} = E[Y_i(d, m)]$ and $p_d = P(M_i(d) = 1)$.
- Then, the following bounds exist for the ACME:

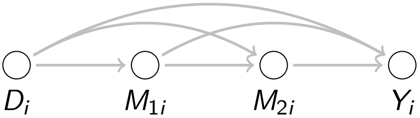
$$\tau_{LB}(1) \leq \tau_{ACME}(1) \leq \tau_{UB}(1),$$

where

$$\begin{aligned}
\tau_{LB}(1) =& \big(\max\{0,\ \mu_{11} + p_1 - 1\} + \max\{0,\ \mu_{10} - p_1\}\big) \\
&- \big(\min\{\mu_{11}, p_0\} + \min\{\mu_{10}, 1 - p_0\}\big), \\
\tau_{UB}(1) =& \big(\min\{\mu_{11}, p_1\} + \min\{\mu_{10}, 1 - p_1\}\big) \\
&- \big(\max\{0,\ \mu_{11} + p_0 - 1\} + \max\{0,\ \mu_{10} - p_0\}\big).
\end{aligned}$$

Ordered mediators

- Sometimes, there can be multiple mediators on the path from the treatment to the outcome.



- Mediators for the first generation affect mediators for the second generation.
- Zhou and Yamamoto (2020) show that we can similarly define the average causal mediation effect for each mediator.
- In the previous example, we can isolate the direct effect (D - Y), the effect through M_2 (D - M_2 - Y), and the effect through M_1 (D - M_1 - Y and D - M_1 - M_2 - Y).

Ordered mediators

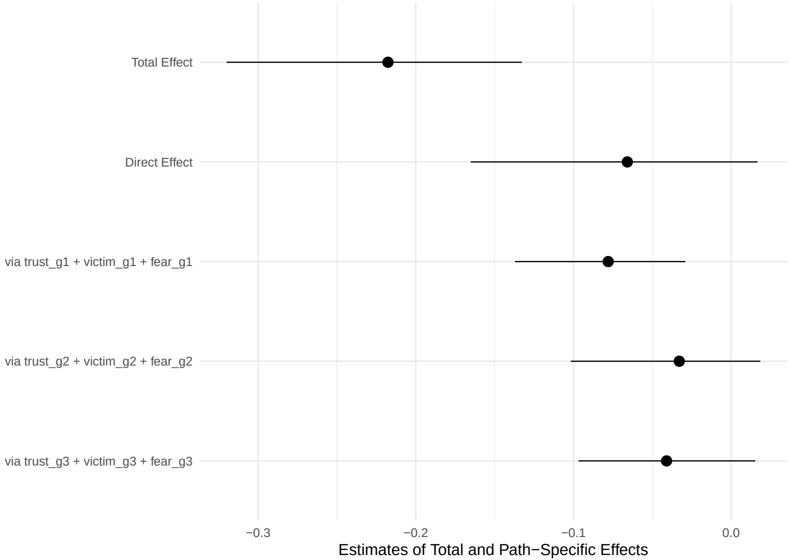
- Zhou and Yamamoto (2020) demonstrate the following decomposition of τ:

$$
\begin{aligned}
& E\left[Y_i\left(1, M_{1i}(0), M_{2i}(0, M_{1i}(0))\right) - Y_i\left(0, M_{1i}(0), M_{2i}(0, M_{1i}(0))\right)\right] \\
& + E\left[Y_i\left(1, M_{1i}(1), M_{2i}(1, M_{1i}(1))\right) - Y_i\left(1, M_{1i}(0), M_{2i}(1, M_{1i}(0))\right)\right] \\
& + E\left[Y_i\left(1, M_{1i}(0), M_{2i}(1, M_{1i}(0))\right) - Y_i\left(1, M_{1i}(0), M_{2i}(0, M_{1i}(0))\right)\right].
\end{aligned}
$$

- Identification requires that sequential ignorability holds for each mediator on the path.
- With binary mediators, each natural mediation effect equals the product of two components, a difference in expectations and a difference in probabilities.

Ordered mediators: application

....



References I

Acharya, Avidit, Matthew Blackwell, and Maya Sen. 2016. "Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects." *American Political Science Review* 110 (3): 512–29.

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.

Imai, Kosuke, Luke Keele, and Dustin Tingley. 2010. "A General Approach to Causal Mediation Analysis." *Psychological Methods* 15 (4): 309.

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

References II

Imai, Kosuke, Dustin Tingley, and Teppei Yamamoto. 2013. "Experimental Designs for Identifying Causal Mechanisms." *Journal of the Royal Statistical Society Series A: Statistics in Society* 176 (1): 5–51.

Lupu, Noam, and Leonid Peisakhin. 2017. "The Legacy of Political Violence Across Generations." *American Journal of Political Science* 61 (4): 836–51.

Zhou, Xiang, and Teppei Yamamoto. 2020. "Tracing Causal Paths from Experimental and Observational Data."