Causal Mediation Analysis

Luke Keele

Penn State University

September 3, 2014
This talk is based on a series of papers with Kosuke Imai, Dustin Tingley, and Teppei Yamamoto.

Papers are at http://www.personal.psu.edu/ljk20
Experiments and Causal Inference

Treatment: All Girls Classroom

Outcome: Math Scores

**WHY** did the treatment work?

Deaton (2010): it’s a “black-box.”
Causal mediation analysis


![Diagram of causal mediation analysis]

- Treatment: All Girls Classroom
- Mediator: Math Anxiety
- Outcome: Math Test Score

Total Effect = Indirect + Direct
Standard Approach to Mediation Analysis

Linear structural equation models:

\[ Scores_i = \alpha_1 + \beta_1 \text{Treat}_i + \epsilon_{1i}, \]
\[ Anxiety_i = \alpha_2 + \beta_2 \text{Treat}_i + \epsilon_{2i}, \]
\[ Scores_i = \alpha_3 + \beta_3 \text{Treat}_i + \gamma \text{Anxiety}_i + \epsilon_{3i} \]

- Total effect (Average Treatment Effect): \( \beta_1 \).
- Direct effect: \( \beta_3 \).
- Indirect (mediated) effect: \( \beta_2 \gamma \).
- Total effect: \( \beta_1 = \beta_3 + (\beta_2 \gamma) \).
The Complication: Conditioning on a Post-Treatment Quantity

\[ \text{Scores}_i = \alpha_1 + \beta_1 \text{Treat}_i + \epsilon_1_i \]
\[ \text{Scores}_i = \alpha_3 + \beta_3 \text{Treat}_i + \gamma \text{Anxiety}_i + \epsilon_3_i \]
Conditioning on a Post-Treatment Quantity
Observed quantities:
- Binary treatment: $T_i \in \{0, 1\}$, all-girls classroom
- Mediator: $M_i$, math anxiety
- Outcome: $Y_i$, math test scores
- Observed pre-treatment covariates: $X_i$.

The potential outcomes:
- Potential mediators: $M_i(t)$
- Potential outcomes: $Y_i(t, m)$
Example with this notation

\( M_i(1) \) is the \textbf{observed} level of math anxiety reported by individual \( i \), who was assigned to an all girls classroom.

\( Y_i(1, M_i(1)) \) is the \textbf{observed} math performance reported by individual \( i \), who was assigned to an all girls classroom, and had a anxiety level \( M_i(1) \) which is observed under an all girls classroom.

\( M_i(0) \) and \( Y_i = Y_i(0, M_i(0)) \) are the converse.
Total Effect

- Total effect = ATE:
  \[ \tau_i \equiv Y_i(1, M_i(1)) - Y_i(0, M_i(0)) \]

- Randomization of the treatment makes the following true:
  \[ Y_i(1, M_i(1)), Y_i(0, M_i(0)) \perp \perp T_i \]
Indirect or mediation effects

- Causal mediation (indirect) effects:
  \[ \delta_i(t) \equiv Y_i(t, M_i(1)) - Y_i(t, M_i(0)) \]

- Change the mediator from \( M_i(0) \) to \( M_i(1) \) while holding the treatment constant at \( t \).
• $t_i = 1$:

$Y_i(1, M_i(1))$ is the observed math score if student is assigned to an all girls classroom, with an anxiety level produced by the all girls class.

• $t_i = 1$:

$Y_i(1, M_i(0))$ is her math score in the counterfactual world where subject $i$ was assigned to an all girls classroom but her math anxiety level is at the same level as if she were in a co-ed classroom.

• “Identification problem:” we don’t observe $Y_i(1, M_i(0))$
Identification under Sequential Ignorability

- Necessary identification assumption: *Sequential Ignorability* (Imai, Keele, Yamamoto, 2010)

\[
\{ Y_i(t', m), M_i(t) \} \perp\!\!\!\!\!\!\perp T_i \mid X_i = x,
\]

\[
Y_i(t', m) \perp\!\!\!\!\!\!\perp M_i(t) \mid T_i = t, X_i = x
\]
Unpacking Sequential Ignorability in the Example

Figure: We can rule out some confounders but not others when treatment is randomized.
Sequential ignorability in this example

Sequential Ignorability is satisfied if:

- *There are no unobserved pre-treatment covariates that influence both treatment and math anxiety or math scores.*
- *There are no unobserved pre-treatment covariates that influence both math anxiety and math scores.*
Sequential ignorability

- Untestable
- Must defend your specification
Randomizing The Mediator

$T$, All Girls Class

$M$, Math Anxiety

$Y$, Test Scores
IV and mediation analysis

Mediator: Math Anxiety

Treatment: All Girls Classroom

Outcome: Math Test Score

IV: Direct Effect is zero.
Other Issues in Mediation Analysis

• Estimation
• Sensitivity analysis
• Multiple mediators
• Noncompliance
Guidelines

- Articulate and evaluate assumptions: Present total and indirect effects separately
- Measure possible confounders: baseline mediators and outcomes especially.
- Perform a sensitivity analysis
Standard Estimation Methods and Sequential Ignorability

Standard Equations for Mediator and Outcome:

\[ M_i = \alpha_2 + \beta_2 T_i + \epsilon_{2i}, \]
\[ Y_i = \alpha_3 + \beta_3 T_i + \gamma M_i + \epsilon_{3i} \]

Under sequential ignorability, the average causal mediation effect (ACME) is identified as \( \bar{\delta}(0) = \bar{\delta}(1) = \beta_2 \gamma \).

**Product of coefficients is only a valid measure of the causal mediation effect when:**

The M and Y models are linear regressions.

Problem: Linear models are often inappropriate. \( \Rightarrow \) When models aren’t linear regressions \( \beta_2 \gamma \) represents nothing!
Example: Continuous mediator and binary outcome

Suppose

\[ M_i = \alpha_2 + \beta_2 T_i + \epsilon_{2i}, \]
\[ \Pr(Y_i = 1) = \Phi(\alpha_3 + \beta_3 T_i + \gamma M_i + \epsilon_{3i}) \]

Then the ACME is given by

\[ \bar{\delta}(t) = \Phi \left( \frac{\alpha_3 + \beta_3 t + \gamma (\alpha_2 + \beta_2)}{\sqrt{\sigma^2 \gamma^2 + 1}} \right) - \Phi \left( \frac{\alpha_3 + \beta_3 t + \gamma \alpha_2}{\sqrt{\sigma^2 \gamma^2 + 1}} \right) \]

Not \( \beta_2 \gamma \).
Estimation algorithm

- Sketch of the algorithm
  - Step 1: Fit some model for the mediator $g(M_i|T_i, X_i)$ and outcome variable $f(Y_i|T_i, M_i, X_i)$
  - Step 2: Predict values of the mediator under both values of the treatment.
  - Step 3: Predict values of the outcome variable under the simulated values of the mediator but using both $T_i = 1$ and $T_i = 0$
  - Step 4: Take the average difference in these two predictions
  - Confidence intervals using bootstrapping.
Example: Continuous mediator and binary outcome

Estimate the two following models:

\[ M_i = \alpha_2 + \beta_2 T_i + X_i + \epsilon_{2i}, \]
\[ \Pr(Y_i = 1) = \Phi(\alpha_3 + \beta_3 T_i + \gamma M_i + X_i + \epsilon_{3i}) \]

- Predict \( M_i \) for \( T_i = 1 \) and \( T_i = 0 \). This gives you \( \hat{M}_i(1) \) and \( \hat{M}_i(0) \).
- Predict \( Y_i \) with \( T_i = 1 \) and \( \hat{M}_i(0) \) and vice versa.
- Take average of these two predictions.
Sensitivity Analysis

• Sequential ignorability is a unrefutable strong assumption: can’t be tested with any configuration of the data.
• Need to assess the robustness of findings via a sensitivity analysis
• A sensitivity analysis asks: how large a departure from the key assumption must occur for the conclusions to no longer hold?
• Statistical interpretation: Parametric sensitivity analysis by assuming

\[ \{Y_i(t', m), M_i(t)\} \perp \perp T_i \mid X_i = x \]

but not

\[ Y_i(t', m) \perp \perp M_i \mid T_i = t, X_i = x \]
Sensitivity Analysis

Sensitivity Analysis:

- Let the resulting correlation in error terms be \( \rho = \text{corr}(\epsilon_{2i}, \epsilon_{3i}) \).
- If \( \rho = 0 \) sequential ignorability holds. Nonzero values of \( \rho \) represent departures from the key assumption.
- We can calculate mediation effect as a function of \( \rho \) and observe for what value of \( \rho \), the mediation effect and its confidence interval includes zero.
The Product of Coefficients Method

- The $\beta_2$ term represents the magnitude of the relationship between the treatment and the mediator.
- The $\gamma$ term represents the magnitude of the relationship between the mediator and dependent variable after controlling for the effect of the independent variable.
- The product of these two quantities is the amount of variance in the dependent variable that is accounted for by the independent variable through the mechanism of the mediator.