Statistical Analysis of Causal Mechanisms

Kosuke Imai

Princeton University

Joint work with Luke Keele, Dutin Tingley, Teppei Yamamoto

October 12, 2009

Japanese Political Science Association Meeting

Kosuke Imai (Princeton)

Identification of Causal Mechanisms

Tokyo 2009 1 / 15

Experiments, Statistics, and Causal Mechanisms

- Causal inference is a central goal of social science
- Experiments as **gold standard** for estimating *causal effects*
- But, we really care about *causal mechanisms*
- A major criticism of experimentation (and statistics): it can only determine whether the treatment causes changes in the outcome, but not how and why
- Experiments are a **black box**
- Qualitative research uses process tracing
- Key Challenge: How can we use statistics to identify causal mechanisms?

Causal Mediation Analysis



• Popular in psychology and behavioral neuroscience

- Binary treatment: $T_i \in \{0, 1\}$
- Mediator: $M_i \in \mathcal{M}$
- Outcome: $Y_i \in \mathcal{Y}$
- Observed covariates: $X_i \in \mathcal{X}$
- Potential mediators: $M_i(t)$ where $M_i = M_i(T_i)$
- Potential outcomes: $Y_i(t, m)$ where $Y_i = Y_i(T_i, M_i(T_i))$
- Fundamental problem of causal inference: Only one potential outcome is observed

Kosuke Imai (Princeton)

Identification of Causal Mechanisms

Tokyo 2009 5 / 15

Defining and Interpreting Causal Mediation Effects

• Total causal effect:

$$\tau_i \equiv Y_i(1, M_i(1)) - Y_i(0, M_i(0))$$

• Indirect (causal mediation) effects:

$$\delta_i(t) \equiv Y_i(t, M_i(1)) - Y_i(t, M_i(0))$$

- Causal effect of the change in *M_i* on *Y_i* that would be induced by treatment
- Change the mediator from M_i(0) to M_i(1) while holding the treatment constant at t
- Fundamental problem: For each unit *i*, $Y_i(t, M_i(t))$ is observable but one can *never* observe $Y_i(t, M_i(1 t))$

Nonparametric Identification

• Quantity of Interest: Average Causal Mediation Effects

$$\bar{\delta}(t) \equiv \mathbb{E}(\delta_i(t)) = \mathbb{E}\{Y_i(t, M_i(1)) - Y_i(t, M_i(0))\}$$

- Problem: Y_i(t, M_i(t)) is observed but Y_i(t, M_i(1 t)) can never be observed
- Proposed identification assumption: Sequential Ignorability

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

$$Y_i(t', m) \perp M_i \mid T_i = t, X_i = x$$



Kosuke Imai (Princeton)

Identification of Causal Mechanisms

Tokyo 2009 7 / 15

Inference Under Sequential Ignorability

- Model outcome and mediator
- Outcome model: $p(Y_i | T_i, M_i, X_i)$
- Mediator model: $p(M_i | T_i, X_i)$
- A simplest setup: Linear Structural Equation Model (LSEM)

$$\begin{aligned} \mathbf{M}_i &= \alpha_2 + \beta_2 \mathbf{T}_i + \epsilon_{i2}, \\ \mathbf{Y}_i &= \alpha_3 + \beta_3 \mathbf{T}_i + \gamma \mathbf{M}_i + \epsilon_{i3}. \end{aligned}$$

Theorem 2 (Identification Under LSEM)

Under the LSEM and sequential ignorability, the average causal mediation effects are identified as $\overline{\delta}(0) = \overline{\delta}(1) = \beta_2 \gamma$.

- Can include the interaction between T_i and M_i
- Can use parametric or nonparametric regressions; probit, logit, ordered mediator, GAM, quantile regression, etc.

Need for Sensitivity Analysis

- The sequential ignorability assumption is often too strong
- Need to assess the robustness of findings via sensitivity analysis
- Question: How large a departure from the key assumption must occur for the conclusions to no longer hold?
- Parametric sensitivity analysis by assuming

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

but not

$$Y_i(t', m) \perp M_i \mid T_i = t, X_i = x$$

• Possible existence of unobserved pre-treatment confounder

Kosuke Imai (Princeton)	Identification of Causal Mechanisms	Tokyo 2009	9 / 15

Parametric Sensitivity Analysis

- Sensitivity parameter: $\rho \equiv Corr(\epsilon_{i2}, \epsilon_{i3})$
- Sequential ignorability implies $\rho = 0$
- Set ρ to different values and see how mediation effects change

Theorem 3

$$\overline{\delta}(\mathbf{0}) = \overline{\delta}(\mathbf{1}) = \frac{\beta_2 \sigma_1}{\sigma_2} \left\{ \widetilde{\rho} - \frac{\rho}{\sqrt{(1 - \widetilde{\rho}^2)/(1 - \rho^2)}} \right\}$$

where $\sigma_j^2 \equiv \operatorname{var}(\epsilon_{ij})$ for j = 1, 2 and $\tilde{\rho} \equiv \operatorname{Corr}(\epsilon_{i1}, \epsilon_{i2})$.

- When do my results go away completely?
- $\overline{\delta}(t) = 0$ if and only if $\rho = \tilde{\rho}$
- Easy to estimate from the regression of Y_i on T_i :

$$Y_i = \alpha_1 + \beta_1 T_i + \epsilon_{i1}$$

An alternative interpretation via R²

Empirical Illustration: Nelson et al. (APSR)

- How does media framing affect citizens' political opinions?
- News stories about the Ku Klux Klan rally in Ohio
- Treatment: Free speech frame ($T_i = 0$) and public order frame ($T_i = 1$)
- Randomized experiment with sample size = 136
- Mediators: general attitudes about the importance of free speech and public order
- Outcome: tolerance for the Klan rally
- Expected findings: negative mediation effects

Kosuke Imai (Princeton)

Identification of Causal Mechanisms

Tokyo 2009 11 / 15

Analysis under Sequential Ignorability

Average Mediation Effects $\hat{\delta}(0) = \hat{\delta}(1)$	-0.44 [-0.87, -0.01]
Average Direct Effects $\hat{\zeta}(0) = \hat{\zeta}(1)$	-0.02 [-0.49, 0.47]
Average Total Effect $\hat{\tau}$	-0.46 [-1.11, 0.23]

Sensitivity Analysis with Respect to ρ



ACME(p)

Concluding Remarks

- Identification of causal mechanisms is difficult
- An additional assumption is required even in experiments
- General identification and estimation strategy
- Sensitivity analysis to assess the robustness
- Related work "Experimental Identification":
 - New experimental designs for identifying causal mechanisms
 - Assumptions about designs vs. statistical assumptions

Papers and Software

- "Experimental Identification of Causal Mechanisms"
- "Identification, Inference, and Sensitivity Analysis for Causal Mediation Effects."
- "A General Approach to Causal Mediation Analysis."
- "Causal Mediation Analysis in R."
- All available at http://imai.princeton.edu/projects/mechanisms.html
- mediation: R package for causal mediation analysis
- Available at

http://cran.r-project.org/web/packages/mediation/

Kosuke Imai (Princeton)

Identification of Causal Mechanisms

Tokyo 2009 15 / 15