

# Statistical Considerations in Multilevel Mediation Analysis

William Ruth

Collaborators: Rado Ramasy, Rowin Alfaro, Ariel Mundo, Bruno Remillard, Bouchra Nasri



Université   
de Montréal

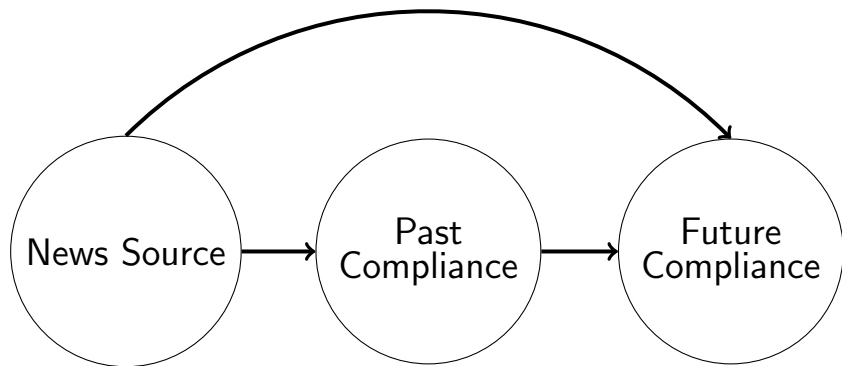
# Outline

- 1) The Problem
- 2) Mediation Analysis
- 3) Causal Inference
- 4) Mixed-Effects Models
- 234) Mixed-Effects Models in Causal Mediation Analysis

# Example

- Goal: Understand adherence to restrictive measures
  - E.g. Lockdowns
  - Both past and future
- Influence of news source
  - How trustworthy?
- Disentangle influence on future from influence on past

# Example

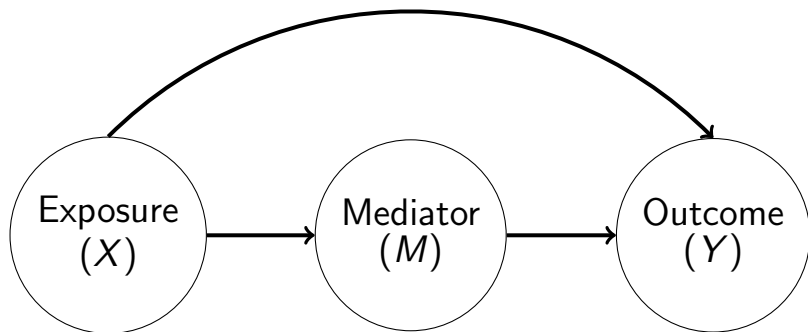


# Example

## Terminology

- Top path: Direct effect
- Center path: Indirect effect
- Combined: Total effect
- Exposure:  $X$
- Outcome:  $Y$
- Mediator:  $M$

# Mediation Analysis



# Mediation Analysis

Separate **Total Effect** of  $X$  on  $Y$  into

- **Direct Effect**
- **Indirect Effect**

Traditionally, use regression

# Mediation Analysis

Continuous outcome and mediator:

- $Y = \alpha_0 + \alpha_1 M + \alpha_2 X + \varepsilon_Y$
- $M = \beta_0 + \beta_1 X + \varepsilon_M$

**Direct Effect:**  $\alpha_2$

- “X in Y”

**Indirect Effect:**  $\alpha_1 \cdot \beta_1$

- “M in Y” · “X in M”

**Total Effect:**  $\alpha_2 + \alpha_1 \cdot \beta_1$



# Mediation Analysis

Popular approach

- A bit outdated...

More popular: Causal mediation analysis

# Causal Inference

Assume that  $X$  *causes*  $Y$

Counterfactuals:

- What value would  $Y$  take if  $X$  were set to a particular level?
- Write  $Y_x$  for the value of  $Y$  when  $X = x$
- If  $X \neq x$  then  $Y_x$  is literally a “counterfactual”

# Causal Inference

Example:

- Alice only reads scientific publications and will follow all lockdown mandates
- What if she instead only read Facebook?
- $Y_{Science}(Alice) = \text{follow}$
- $Y_{Facebook}(Alice) = \text{follow}$

# Causal Inference

Example:

- Bob also only reads scientific publications and will follow all lockdown mandates, but is more susceptible to being influenced
- $Y_{Science}(Bob) = \text{follow}$
- $Y_{Facebook}(Bob) = \text{not follow}$

# Causal Inference

- We only observe one outcome per individual
- Explore population-level effects by averaging
- Define mediation effects in terms of expected counterfactuals

# Causal Inference

**Total Effect:**  $\mathbb{E}(Y_{x'} - Y_x)$

- Effect on outcome when we change exposure from  $X = x$  to  $X = x'$

Other effects involve dependence on a mediator:

- $Y_{xm}$ : Value of outcome when
  - Exposure ( $X$ ) is set to  $x$
  - Mediator ( $M$ ) is set to  $m$
- $M_x$ : Value of mediator when
  - Exposure ( $X$ ) is set to  $x$
- “Nested Counterfactuals”:  $Y_{xM_x}$  or  $Y_{xM_{x'}}$

# Causal Mediation Analysis

**Controlled Direct Effect:**  $\mathbb{E}(Y_{x'm} - Y_{xm})$

- Effect of changing exposure with mediator held fixed

**Natural Direct Effect:**  $\mathbb{E}(Y_{x'M_x} - Y_{xM_x})$

- Effect of changing exposure when we don't interfere with the mediator

**Natural Indirect Effect:**  $\mathbb{E}(Y_{xM_{x'}} - Y_{xM_x})$

- Effect of changing which exposure value is seen by the mediator while holding fixed which exposure value is seen by the outcome

# Causal Mediation Analysis

In our example

- Controlled Direct Effect: Effect of increasing news trustworthiness if the whole population followed guidelines in the past
- Natural Direct Effect: Effect of increasing news trustworthiness independent of any induced change in past compliance
- Natural Indirect Effect: Effect of changing past compliance if everyone only got news from Facebook



# Causal Mediation Analysis

We can't measure all required counterfactuals

- E.g.,  $Y_x$  or  $Y_{x'}$ , not both

Expected counterfactuals related to conditional expectations

- Under strong assumptions,  $\mathbb{E}Y_x = \mathbb{E}(Y|X = x)$

Nested counterfactuals more complicated

- More on this later

# Causal Mediation Analysis

How does causality change our analysis?

Still fit regression models, but include interaction terms between exposure and mediator

- $Y = \alpha_0 + \alpha_1 M + \alpha_2 X + \alpha_3 M \cdot X + \varepsilon_Y$
- $M = \beta_0 + \beta_1 X + \varepsilon_M$

Direct and indirect effects now depend on the levels of the exposure

# Causal Mediation Analysis – Extensions

Discussion so far has involved continuous mediator and outcome

- What about binary or categorical?

Individuals might also be clustered

- E.g. Within countries

# Causal Mediation Analysis – Extensions

Handling binary variables is pretty straightforward

- Instead of linear regression, use logistic regression
- Re-define mediation effects based on expected counterfactuals
  - I.e. Counterfactual probabilities
- New formulas for relating mediation effects to regression coefficients

Extend to more than 2 categories using binary indicators

# Causal Mediation Analysis – Extensions

Clustered data more complicated

Standard approach is multi-level modelling

- I.e. Add random effects which vary across clusters

Combined with categorical variables:

- Generalized linear mixed models (GLMMs)

# Mixed-Effects Models

The core idea is to augment our set of covariates

- Coefficients of these new covariates are random variables which vary across groups/clusters

In the linear setting:

- Old model:  $Y = \alpha_0 + \alpha_1 X_1 + \dots + \alpha_p X_p + \varepsilon$

- New model:

$$Y = \alpha_0 + \alpha_1 X_1 + \dots + \alpha_p X_p + \mathbf{u}_1 \mathbf{Z}_1 + \dots + \mathbf{u}_q \mathbf{Z}_q + \varepsilon$$

# Mixed-Effects Models

The  $Z$ 's are fixed, known covariates The  $u$ 's are random variables

- I.e. Random effects

It's possible for the  $X$ 's and  $Z$ 's to overlap

- The coefficient on such a covariate has the form  
 $\alpha_j + u_k$
- I.e. Mixed effect

# Mixed-Effects Models

Extend to generalized linear models in the usual way

Choose response distribution and link function as for ordinary GLMs

Linear predictor now has a random effects component



# Mixed-Effects Models

Why bother?

- E.g. Measured some but not all levels of a categorical variable
- Estimate covariance matrix of random effects
- Test for non-zero variance of each random effect

“Predict” level of random effects for each group

- Conditional mean or conditional mode of random effects given response

# Mixed-Effects Models

In our example:

- Data collected from 11 different countries
- Explicitly model inter-country variability
- Predict country-specific random effects
- Use country-specific coefficients in formulas for mediation effects
- Test for significant mediation effects within each country

# Multilevel Mediation Analysis

Uncertainty quantification for mixed-effects models can be challenging

Strategies include:

- Bootstrap
- Quasi-Bayesian Monte Carlo
- $\delta$ -method

# Multilevel Mediation Analysis

Uncertainty quantification for mixed-effects models can be challenging

Strategies include:

- **Bootstrap** (Rado Ramasy, yesterday)
- Quasi-Bayesian Monte Carlo
- $\delta$ -method

# Multilevel Mediation Analysis

Uncertainty quantification for mixed-effects models can be challenging

Strategies include:

- Bootstrap
- Quasi-Bayesian Monte Carlo
- **$\delta$ -method**

# Multilevel Mediation Analysis

Recall: Mediation effects defined using nested counterfactuals –  $Y_{xM_{x'}}$

- Value of  $Y$  when  $X$  is set to  $x$ , and
- $M$  is set to whatever value it would have if  $X$  were set to  $x'$

Under strong assumptions, use the “Mediation Formula”

# Multilevel Mediation Analysis

Mediation Formula:

$$\mathbb{E}Y_{xM_{x'}} = \mathbb{E}_{M|X=x'}\mathbb{E}(Y|X = x, M)$$

# Multilevel Mediation Analysis

Mediation Formula:

$$\begin{aligned}\mathbb{E}Y_{xM_{x'}} &= \mathbb{E}_{M|X=x'}\mathbb{E}(Y|X=x, M) \\ &= \sum_{m=0}^1 \mathbb{P}(Y=1|X=x, M=m)\mathbb{P}(M=m|X=x')\end{aligned}$$



# Multilevel Mediation Analysis

Mediation Formula:

$$\begin{aligned}\mathbb{E}Y_{xM_{x'}} &= \mathbb{E}_{M|X=x'}\mathbb{E}(Y|X=x, M) \\ &= \sum_{m=0}^1 \mathbb{P}(Y=1|X=x, M=m)\mathbb{P}(M=m|X=x')\end{aligned}$$

Logistic regression model makes this (relatively) simple:

$$\mathbb{P}(Y=1|X=x, M=m) = \text{logit}^{-1}(\text{linear predictor})$$

$$\mathbb{P}(M=m|X=x') = \text{logit}^{-1}(\text{different linear predictor})$$

# Multilevel Mediation Analysis

Estimating  $\mathbb{E}Y_{xM_{x'}}$ , messy, but not *hard*

Uncertainty quantification based on asymptotic covariance of regression parameters

- Use  $\delta$ -method
- Need derivatives of  $\mathbb{E}Y_{xM_{x'}}$  wrt regression parameters
- Very messy, not particularly hard

# Multilevel Mediation Analysis

Uncertainty quantification for mediation effects similar

- More  $\delta$ -method

Very similar for GLMs and GLMMs

- Latter has more regression parameters
- Use `merDeriv` package to supplement `lme4`

# Putting it All Together

- Define direct, indirect and total effects using counterfactuals
- Estimate these effects across countries using generalized linear mixed models
- Compute standard errors for estimated effects using the  $\delta$ -method

# Acknowledgements

## Collaborators:

- Rado Ramasy
- Rowin Alfaro
- Ariel Mundo
- Bruno Remillard
- Bouchra Nasri

## Funding:

- Canadian Statistical Sciences Institute

# Thank You