ESTIMATION OF MEDIATED EFFECTS

The most common case is the linear structural equation model with continuous outcomes M and Y. We include covariates **X** and an interaction between the mediator and treatment.

It is useful to begin with the model for the potential outcomes and compare it to the analogous model for the observed outcomes.

$$M_i(z) = \alpha_M^{(c)} + \beta_M^{(c)} z + \lambda_M^{\prime(c)} \mathbf{X}_i + \varepsilon_i(z),$$
$$Y_i(z, m) = \alpha_Y^{(c)} + \gamma_Y^{(c)} z + \beta_Y^{(c)} m + \delta_Y^{(c)} (zm) + \lambda_Y^{\prime(c)} \mathbf{X}_i + \varepsilon_i(z, m),$$

where $E(\varepsilon_i(z) \mid \mathbf{X}_i = \mathbf{x}) = E(\varepsilon_i(z, m) \mid \mathbf{X}_i = \mathbf{x}) = 0$. The superscript "c" indicates the parameters of the model are for the potential outcomes, not the parameters for the model for the observed data.

ESTIMATION OF MEDIATED EFFECTS

From the second equation:

$$Y_i(z, m) = \alpha_Y^{(c)} + \gamma_Y^{(c)} z + \beta_Y^{(c)} m + \delta_Y^{(c)} (zm) + \lambda_Y'^{(c)} \mathbf{X}_i + \varepsilon_i(z, m),$$

the controlled direct effects are immediately obtained:

$$E(Y_i(z, m) - Y_i(z^*, m^*) \mid \mathbf{X}_i = \mathbf{x}) = \gamma_Y^{(c)}(z - z^*) + \beta_Y^{(c)}(m - m^*) + \delta_Y^{(c)}(zm - z^* m^*)$$

ESTIMATION OF MEDIATED EFFECTS

Because Z and M interact, the average direct and indirect effects depend on the decompositions of the total effect.

$$\begin{aligned} E(Y_i(z, M_i(z^*)) \mid \mathbf{X}_i = \mathbf{x}) = \alpha_Y^{(c)} + \gamma_Y^{(c)} z + (\beta_Y^{(c)} + \delta_Y^{(c)} z) E(M_i(z^*) \mid \mathbf{X}_i = \mathbf{x}) \\ + \lambda_Y'^{(c)} \mathbf{x} + \mathbf{E}(\varepsilon_{\mathbf{i}}(\mathbf{z}, \mathbf{M_i}(\mathbf{z}^*)) \mid \mathbf{X_i} = \mathbf{x}), \end{aligned}$$

where, under the identification conditions,

$$E(M_i(z^*) \mid \mathbf{X}_i = \mathbf{x}) = \alpha_M^{(c)} + \beta_M^{(c)} z + \lambda_M'^{(c)} \mathbf{x} \text{ and } E(\varepsilon_i(z, M_i(z^*)) \mid \mathbf{X}_i = \mathbf{x}) = 0.$$

ESTIMATION OF MEDIATED EFFECTS

For the analogous observed data model,

$$\begin{aligned} M_i &= \alpha_M + \beta_M Z_i + \lambda_M' \mathbf{X}_i + \varepsilon_{iM}, \\ Y_i &= \alpha_Y + \gamma_Y Z_i + \beta_Y M_i + \delta_Y (Z_i M_i) + \lambda_Y' \mathbf{X}_i + \varepsilon_{iY}, \end{aligned}$$

where $\mathrm{E}(\varepsilon_{iM} \mid Z_i = z, \mathbf{X}_i = \mathbf{x}) = \mathrm{E}(\varepsilon_{iY} \mid Z_i = z, M_i = m, \mathbf{X}_i = \mathbf{x}) = 0$.

Causal effects estimated from the observed data model agree with the potential outcomes model. i.e.

$$E(Y_i(z, m) - Y_i(z^* m^*) \mid \mathbf{X}_i = \mathbf{x}) = \gamma_Y^{(c)}(z - z^*) + \beta_Y^{(c)}(m - m^*) + \delta_Y^{(c)}(zm - z^* m^*)$$

.

ESTIMATION OF MEDIATED EFFECTS

For the remainder of the lesson, we assume the observations $i = 1, ..., n$ are independent and identically distributed (Y, Z, M). The linear structural equation model:

$$M_i = \alpha_M + \beta_M Z_i + \lambda_M' \mathbf{X}_i + \varepsilon_{iM},$$
$$Y_i = \alpha_Y + \gamma_Y Z_i + \beta_Y M_i + \delta_Y (Z_i M_i) + \lambda_Y' \mathbf{X}_i + \varepsilon_{iY},$$

can be estimated from the data by applying ordinary least squares regression to each equation, and the delta method can be used to obtain standard errors for the estimated effects.

ESTIMATION OF MEDIATED EFFECTS

The logistic regression model or probit model is often used when the mediator is continuous and the outcome is binary.

For a logistic model with mediator conditionally normally distributed, VanderWeele and Vansteelandt (2010) give formulae of the model parameters for the averaged direct and indirect effects. The effects are defined in terms of odds ratios.

For an accelerated failure time model with mediator conditionally normally distributed with constant variance, VanderWeele (2015) gives formulae of the model parameters for differences in logged outcome means. e.g.

$$\log(E(Y(z, M(z)) \mid \mathbf{X} = \mathbf{x}) - \log(E(Y(z, M(z^*)) \mid \mathbf{X} = \mathbf{x}).$$

ESTIMATION OF MEDIATED EFFECTS

Hong (2010) proposes a general method for estimating average direct and indirect effects based on an extension of the weighting approach.

The method does not require modeling the outcome Y. It requires modeling the mediator M and the treatment assignment process.

Under the identification assumptions,

$$E(Y(z, M(z^*))) = E(Y \times 1(Z = z)) \frac{1}{\Pr(Z = z \mid \mathbf{X} = \mathbf{x})} \frac{f(m \mid Z = z^*, \mathbf{X} = \mathbf{x})}{f(m \mid Z = z, \mathbf{X} = \mathbf{x})}.$$

ESTIMATION OF MEDIATED EFFECTS

Imai et al. (2010) propose two approaches using simulation. Both require modeling the outcome and the mediator. One uses the sampling distribution of model parameters.

The other draws bootstrapped samples $\ell = 1, ..., L$ from the original data:

1. For each sample, ℓ, the models for the outcome and mediator are fit.
2. For each unit, i, and each possible value z of the treatment, $k = 1, ..., K$, values of the mediator $M_i(z)$ are simulated under the model.
3. From these simulated values, $M_i^{(sk\ell)}(z)$, the corresponding outcome values, $Y_i^{(sk\ell)}(z, M_i^{(sk\ell)}(z^*))$, are simulated.
4. The outcome values are averaged for each unit and then averaged over units. This yields the estimated average direct and indirect effects for bootstrap sample ℓ.
5. The estimates from the L bootstrap samples may be used to compute means, confidence intervals, etc.