Contents

1 Foundation

Property 1. If $\hat{\theta}$ is the MLE estimate of θ_0, then it has the following property:

$$\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0, \frac{1}{I(\theta_0)})$$

2 Main content

2.1 G-computation & IPTW

Assumptions(Neal; Benkeser; Nabi)

1. Consistency
If the treatment is T, then the observed outcome Y is the potential outcome under treatment T. Formally,

$$T = t \Longrightarrow Y = Y(t)$$

We could write this equivalently as follow:

$$Y = Y(T)$$

The consistency assumption states that a patients counterfactual outcome under their observed treatment history is the outcome that we actually observe. (It might seem like consistency is obviously true, but that is not always the case. For example, if the treatment specification is simply "get a dog" or "don't get a dog," this can be too coarse to yield consistency. It might be that if I were to get a puppy, I would observe Y = 1 (happiness) because I needed an energetic friend, but if I were to get an old, low-energy dog, I would observe Y = 0 (unhappiness). However, both of these treatments fall under the category of "get a dog," so both correspond to T = 1. This means that Y(1) is not well defined, since it will be 1 or 0, depending on something that is not captured by the treatment specification. In this sense, consistency encompasses the assumption that is sometimes referred to as "no multiple versions of treatment.")

2. Positivity
For all values of covariates x present in the population of interest (i.e. x such that $P(X = x) > .0$),

$$0 < P(T = 1 \mid X = x) < 1$$

3. Conditional Exchangeability (randomization condition)

$$(Y(1), Y(0)) \perp T \mid X$$

4. No Interference

$$Y_i(t_1, \ldots, t_{i-1}, t_i, t_{i+1}, \ldots, t_n) = Y_i(t_i)$$

No interference means that my outcome is unaffected by anyone else's treatment. Rather, my outcome is only a function of my own treatment.

Theorem 2.1. ***G-computation***
Given the assumptions of unconfoundedness, positivity, consistency, and no interference, we can identify the average treatment effect:

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}_X[\mathbb{E}[Y \mid T = 1, X] - \mathbb{E}[Y \mid T = 0, X]]$$

Proof:

$$\begin{aligned}
E[Y(t)] &= \sum_x E[Y(t) \mid X = x] P(X = x) \quad \text{(Double expectation theorem)} \\
&= \sum_x E[Y(t) \mid T = t, X = x] P(X = x) \quad \text{(conditional exchangeability)} \\
&= \sum_x E(Y \mid T = t, X = x) P(X = x) \quad \text{(consistency positivity)}
\end{aligned}$$

Theorem 2.2. ***Inverse Probability Weighting (IPW)***
Given the assumptions of unconfoundedness, positivity, consistency, and no interference, we can identify the average treatment effect:

$$\mathbb{E}[Y(t)] = \mathbb{E}\left[\frac{\mathbb{1}(T = t)Y}{P(t \mid X)}\right]$$

Proof:
The IPTW formula is equivalent to the G-computation formula.

$$\begin{aligned}
E\left[\frac{I(T=1)Y}{P(T=1 \mid X)}\right] &= E\left[E\left[\frac{I(T=1)Y}{P(T=1 \mid X)} \mid T, X\right]\right] \quad \text{(Double expectation theorem)} \\
&= E\left[\frac{I(T=1)}{P(T=1 \mid X)} E(Y \mid T, X)\right] \\
&= E\left[\frac{I(T=1)}{P(T=1 \mid X)} E(Y \mid T=1, X)\right] \\
&= E\left[E\left[\frac{I(T=1)}{P(T=1 \mid X)} E(Y \mid T=1, X) \mid X\right]\right] \quad \text{(Double expectation theorem)} \\
&= E\left[\frac{E(Y \mid T=1, X)}{P(T=1 \mid X)} P(T=1 \mid X)\right] = E[E(Y \mid T=1, X)]
\end{aligned}$$

Theorem 2.3. ***Augmented IPW (Doubly robust)***
Given the assumptions of unconfoundedness, positivity, consistency, and no interference, we can identify:

$$\mathbb{E}\left[Y^a\right] = \mathbb{E}[\mathbb{E}[Y \mid A = a, X]] = \mathbb{E}\left[\frac{\mathbb{I}(A = a)}{p(A \mid X)} \times (Y - \mathbb{E}[Y \mid A, X]) + \mathbb{E}[Y \mid A = a, X]\right]$$

This method is doubly robust that as long as either the propensity score or outcome regression is correctly specified, the estimation is then unbiased.

Proof:

$$\begin{aligned}
g(a; \alpha, \eta) - g(a) &= \mathbb{E}\left[\frac{\mathbb{I}(A = a)}{\pi_a(X; \alpha)} \times (Y - b(a, X; \eta)) + b(a, X; \eta)\right] - \mathbb{E}[b(a, X)] \\
&= \mathbb{E}\left[\frac{\pi_a(X)}{\pi_a(X; \alpha)} \times (b(a, X) - b(a, X; \eta)) + (b(a, X; \eta) - b(a, X))\right] \\
&= \mathbb{E}\left[(b(a, X; \eta) - b(a, X)) \times \left(1 - \frac{\pi_a(X)}{\pi_a(X; \alpha)}\right)\right]
\end{aligned}$$

where $\mathbb{E}[Y \mid A = a, X = x)] = b(a, x)$ and $\pi_a(x) = p(a \mid x)$

Question 2(BIOS760R quiz2): Assume we are interested in the average total effect (ATE) of a point exposure A on an outcome Y, defined as ATE $= \mathbb{E}\left[Y^1 - Y^0\right]$. In order to identify the ATE as a function of observed data, we discussed the conditionally-ignorable model which encodes the following three assumptions:

Go to page 1

- Consistency: the potential outcome is the same of the observed outcome if the (hypothetical) intervention value is the same as the observed value of the treatment, i.e., $Y^a = Y$ if $A = a$.
- Conditional ignorability: $Y^a A \mid X$, for $a = 0, 1$ (where X is a set of observed variables.)
- Positivity: $p(A = a \mid X = x) > 0$ for all x where $p(X = x) > 0$.

Under the above assumptions, show that:

1. $\mathbb{E}\left[Y^a\right]$ can be identified via the adjustment functional, i.e., $\mathbb{E}\left[Y^a\right] = \mathbb{E}[\mathbb{E}[Y \mid A = a, X]]$.

$$
\begin{aligned}
E\left[Y^a\right] &= E[E(y \mid x, A = a)] \quad \text{(Double expectation over } X\text{, conditional ignorability, and consistency} \Longrightarrow \text{G-comp)} \\
&= \sum_{xy} y p(y \mid x, A = a) p(x) \\
&= \sum_{xy, a'} I(A = a) y p\left(y \mid x, A = a'\right) p(x) \\
&= \sum_{xya'} \frac{I\left(A = a'\right)}{p\left(A = a' \mid x\right)} \cdot y p\left(y \mid x, A = a'\right) p\left(A = a' \mid x\right) p(x) \\
&= E\left[\frac{I(A = a)}{P(A \mid x)} y\right] \Longrightarrow \text{IPW}
\end{aligned}
$$

2. The adjustment functional has a dual representation in terms of the IPW functional, i.e., $\mathbb{E}[\mathbb{E}[Y \mid A = a, X]] = \mathbb{E}\left[\frac{\mathbb{I}(A=a)}{p(A \mid X)} \times Y\right]$

$$
\begin{aligned}
E\left[\frac{I(A = a)}{P(A \mid X)} Y\right] &= E\left[\frac{I(A = a)}{P(A = a \mid X)} Y\right] \\
&= E\left[\frac{I(A = a)}{P(A = a \mid X)} E[Y \mid A = a, X]\right] \quad \text{(Double expectation over } A, X\text{)} \\
&= E\left[E[Y \mid A = a, X] \cdot \frac{P(A = a \mid X)}{P(A = a \mid X)}\right] \quad \text{(Double expectation over } X\text{)} \\
&= E[E[Y \mid A = a, X]]
\end{aligned}
$$

3. Bonus point: IPW and adjustment functionals have a third equivalent representation in form of the Augmented IPW functional I^1 That is,

$$
\mathbb{E}[\mathbb{E}[Y \mid A = a, X]] = \mathbb{E}\left[\frac{\mathbb{I}(A = a)}{p(A \mid X)} \times (Y - \mathbb{E}[Y \mid A, X]) + \mathbb{E}[Y \mid A = a, X]\right]
$$

2.2 do-calculus

2.2.1 Intervention

Intervention vs Condition(Glymour et al.): The difference between intervening on a variable and conditioning on that variable should, hopefully, be obvious. When we intervene on a variable in a model, we fix its value. We change the system, and the values of other variables often change as a result. When we condition on a variable, we change nothing; we merely narrow our focus to the subset of cases in which the variable takes the value we are interested in. What changes, then, is our perception about the world, not the world itself.

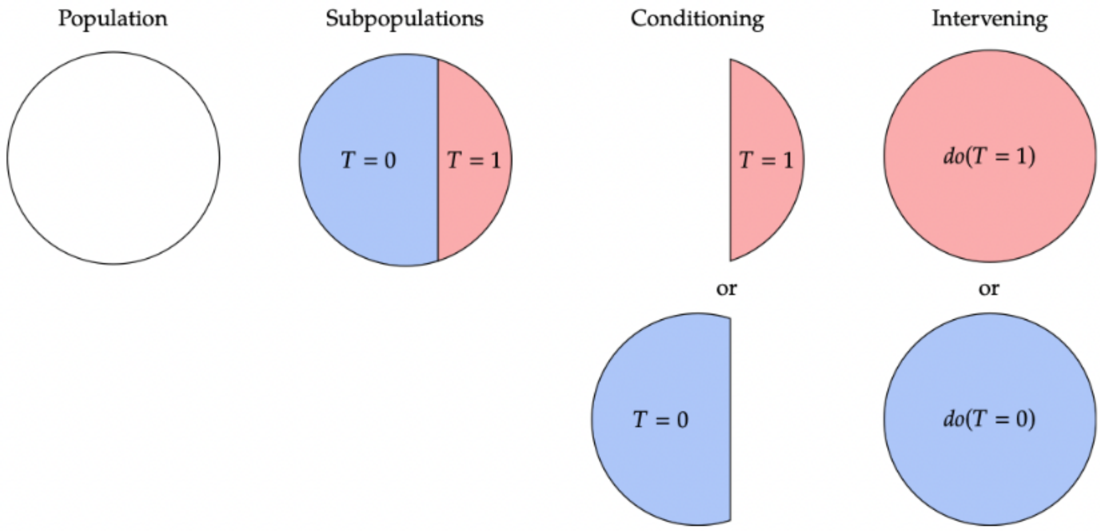


Figure 1: Illustration of the difference between conditioning and intervening (Neal)

Graphically: When we intervene to fix the value of a variable, we curtail the natural tendency of that variable to vary in response to other variables in nature. This amounts to performing a kind of surgery on the graphical model, *removing all edges directed into that variable.*

Notation: In notation, we distinguish between cases where a variable X takes a value x naturally and cases where we fix $X = x$ by denoting the latter $do(X = x)$.

Assumption 2.1 ((Modularity / Independent Mechanisms / Invariance). *If we intervene on a set of nodes $S \subseteq [n]$, then for all i, we have the following:*

1. *if $i \notin S$, then $P(x_i \mid \mathrm{pa}_i)$ remains unchanged*
2. *if $i \in S$, then $P(x_i \mid \mathrm{pa}_i) = 1$ if x_i is the value that X_i was set to by the intervention; otherwise, $P(x_i \mid \mathrm{pa}_i) = 0$*

We assume that intervening on a variable X_i only changes the causal mechanism for X_i; it does not change the causal mechanisms that generate any other variables(Neal). In other words, intervention has no "side effects," that is, that assigning the value x_i for the valuable X_i for an individual does not alter subsequent variables in a direct way. For example, being "assigned" a drug might have a different effect on recovery than being forced to take the drug against one's religious objections. When side effects are present, they need to be specified explicitly in the model(Glymour et al.).

2.2.2 The adjustment formula

Example 1. The causal effect of $P(Y = y \mid do(X = x))$ is equal to the conditional probability $P_m(Y = y \mid X = x)$) that prevails in the manipulated model of 6b. The key to computing the causal effect lies in the observation that P_m, the manipulated probability, shares two essential properties with P.

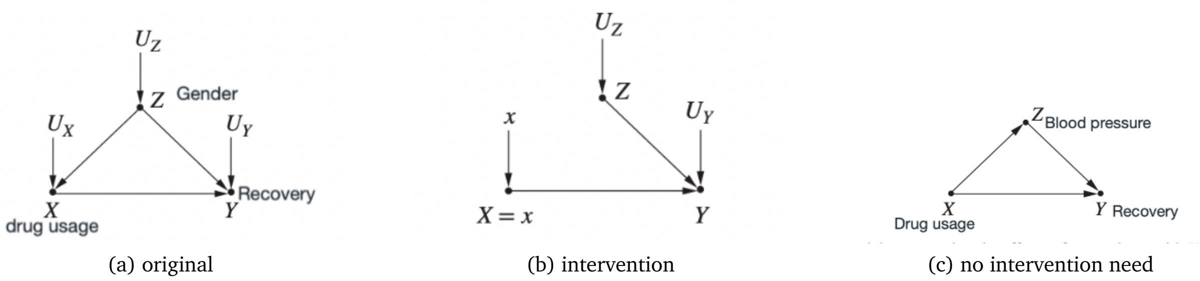


Figure 2: A graphical model representing the effects of a new drug, with Z representing gender, X standing for drug usage, and Y standing for recovery

We have the following observations

1. the marginal probability $P(Z=z)$ is invariant under the intervention, because the process determining Z is not affected by removing the arrow from Z to X
2. the conditional probability $P(Y=y \mid Z=z, X=x)$ is invariant, because the process by which Y responds to X and Z, $Y=f\left(x, z, u_Y\right)$ remains the same, regardless of whether X changes spontaneously or by deliberate manipulation.

Thus,

$$P_m(Y=y \mid Z=z, X=x)=P(Y=y \mid Z=z, X=x) \quad \text{and} \quad P_m(Z=z)=P(Z=z)$$

and

$$\begin{aligned}
P(Y=y \mid do(X=x)) &= P_m(Y=y \mid X=x) \quad \text{(by definition)} \\
&= \sum_z P_m(Y=y \mid X=x, Z=z) P_m(Z=z \mid X=x) \\
&= \sum_z P_m(Y=y \mid X=x, Z=z) P_m(Z=z) \\
&= \sum_z P(Y=y \mid X=x, Z=z) P(Z=z)
\end{aligned}$$

In graphical models, an intervention is simulated by severing all arrows that enter the manipulated variable X. Figure 2c shows no arrow entering X, since X has no parents. This means that no surgery is required, which gives us the license to treat X as a randomized treatment. Thus

$$P(Y=y \mid do(X=x))=P(Y=y \mid X=x)$$

The general rule following this example is

Theorem 2.4 (The Causal Effect Rule). *Given a graph G in which a set of variables PA are designated as the parents of X, the causal effect of X on Y is given by*

$$P\left(Y=y \mid do(X=x)\right)=\sum_z P(Y=y \mid X=x, PA=z) P(PA=z)$$

where z ranges over all the combinations of values that the variables in PA can take.

If we multiply and divide the summand by the probability $P(X=x \mid PA=z)$, we get a form similar to IPW

$$P(y \mid do(x))=\sum_z \frac{P(X=x, Y=y, PA=z)}{P(X=x \mid PA=z)}$$

Multiple intervention
We simply write down the product decomposition of the preintervention distribution, and strike out all factors that correspond to variables in the intervention set

$$P\left(x_1, x_2, \ldots, x_n \mid do(x)\right)=\prod_i P\left(x_i \mid pa_i\right) \quad \text{for all } i \text{ with } X_i \text{ not in } X$$

2.2.3 The backdoor criterion

[Adjusting for variables other than parents]
In the above discussion, we came to the conclusion that we should adjust for a variable's parents, when trying to determine its effect on another variable. But often, we know, or believe, that *the variables have unmeasured parents* that, though represented in the graph, may be inaccessible for measurement. In those cases, we need to find an alternative set of variables to adjust for.

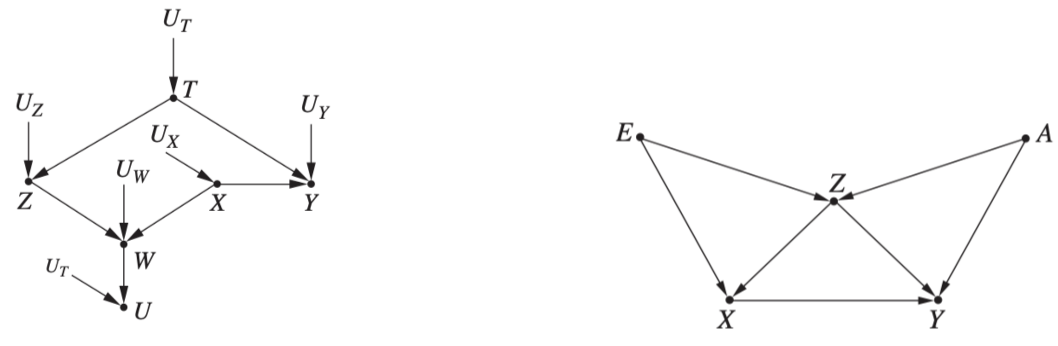
Theorem 2.5 ((The Backdoor Criterion). *Given an ordered pair of variables (X, Y) in a directed acyclic graph G, a set of variables Z satisfies the backdoor criterion relative to (X, Y) if no node in Z is a descendant of X, and Z blocks every path between X and Y that contains an arrow into X.*
If a set of variables Z satisfies the backdoor criterion for X and Y, then the causal effect of X on Y is given by the formula

$$P(Y=y \mid do(X=x))=\sum_z P(Y=y \mid X=x, Z=z) P(Z=z)$$

The logic behind the backdoor criterion is fairly straightforward. In general, we would like to condition on a set of nodes Z such that

1. We block all spurious paths between X and Y
2. We leave all directed paths from X to Y unperturbed
3. We create no new spurious paths

Here are two tricky cases



(a) effect modification

(b) unavoidable collider

Figure 3: Two special cases in applying backdoor criterion

- In 3a, suppose we want to get $P(Y = y \mid do(X = x), W = w)$. However, conditioning on W opens new backdoor path, we need to further condition on T to break such path

$$P(Y = y \mid do(X = x), W = w) = \sum_t P(Y = y \mid X = x, W = w, T = t)P(T = t \mid W = w)$$

- The examples so far give the impression that one should refrain from adjusting for colliders. But in 3b, adjusting for collider Z is unavoidable. Therefore, we need to condition on additional variable(s). So overall, we can condition on $\{E, Z\}, \{A, Z\},$ or $\{E, Z, A\}$

2.2.4 The Front-Door Criterion

[through two consecutive applications of the backdoor criterion]

Theorem 2.6 (The Frontdoor Criterion). *A set of variables Z is said to satisfy the front-door criterion relative to an ordered pair of variables (X, Y) if*

1. *Z intercepts all directed paths from X to Y*
2. *There is no unblocked path from X to Z*
3. *All backdoor paths from Z to Y are blocked by X*

If Z satisfies the front-door criterion relative to (X, Y) and if $P(x, z) > 0$, then the causal effect of X on Y is identifiable and is given by the formula

$$P(y \mid do(x)) = \sum_z P(z \mid x) \sum_{x'} P\left(y \mid x', z\right) P\left(x'\right)$$

One example where the Front-door Criterion is applicable

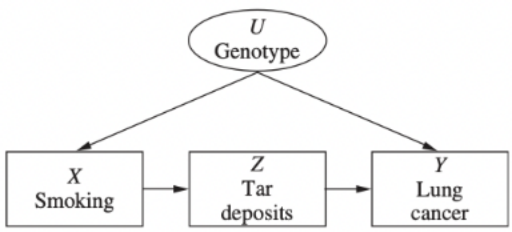


Figure 4: Illustration of the difference between conditioning and intervening (Neal)

Go to page 1

2.3 Semi-parametric odds ratio model (Chen)

2.3.1 Odds ratio related definitions

Disease	Exposure		Total
	Yes	No	
Yes	n_{11}	n_{12}	n_{1+}
No	n_{21}	n_{22}	n_{2+}
Total	n_{+1}	n_{+2}	n_{++}

Table 1: Counts on disease and exposure status

Let's define

$$\begin{aligned} p_{11} &= P(D = \text{ Yes } \mid E = \text{ Yes }) \\ p_{12} &= P(D = \text{ Yes } \mid E = \text{ No }) \\ q_{11} &= P(E = \text{ Yes } \mid D = \text{ Yes }) \\ q_{21} &= P(E = \text{ No } \mid D = \text{ Yes }) \end{aligned}$$

Relative risk

$$RR(D = Yes) = \frac{p_{11}}{p_{12}}$$

Odds

$$\text{Odds}(D = \text{Yes}) = \frac{q_{11}}{q_{21}}$$

$$\text{Odds}(D = \text{No}) = \frac{q_{12}}{q_{22}}$$

Odds ratio

$$OR = \frac{\text{Odds}(D = \text{ Yes })}{\text{Odds}(D = \text{No})}$$

Connection

$$OR = \frac{p_{11}p_{22}}{p_{12}p_{21}} = \frac{RR(D = \text{Yes})}{RR(D = \text{No})}$$

Similarly for the J*K table Define

Y	X				Total
	1	2	$\cdots$	K	
1	n_{11}	n_{12}	$\cdots$	n_{1K}	n_{1+}
2	n_{21}	n_{22}	$\cdots$	n_{2K}	n_{2+}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
J	n_{J1}	n_{J2}	$\vdots$	n_{JK}	n_{J+}
Total	n_{+1}	n_{+2}	$\vdots$	n_{+K}	n_{++}

Table 2: Count in a J*K table

$$\begin{aligned} p_{jk} &= P(Y = j \mid X = k) \\ q_{kj} &= P(X = k \mid Y = j) \\ h_{jk} &= P(Y = j, X = k) \end{aligned}$$

Odds ratio with J, K as reference level

$$\theta_{jk}(J, K) = \frac{p_{kj}p_{KJ}}{p_{kJ}p_{Kj}}$$

$$\theta_{jk}(J, K) = \frac{q_{kj}q_{KJ}}{q_{kJ}q_{Kj}}$$

$$\theta_{jk}(J, K) = \frac{h_{jk}h_{JK}}{h_{jK}h_{Jk}}$$

Connection between OR with different reference

$$\begin{aligned} \theta_{jk}(1, 1) &= \frac{p_{jk}p_{JK}}{p_{jK}p_{Jk}} \times \frac{p_{jK}p_{J1}}{p_{j1}p_{JK}} \times \frac{p_{1K}p_{Jk}}{p_{1k}p_{JK}} \times \frac{p_{JK}p_{11}}{p_{J1}p_{1K}} \\ &= \frac{\theta_{jk}(J, K)\theta_{11}(J, K)}{\theta_{j1}(J, K)\theta_{1k}(J, K)} \end{aligned}$$

Odds ratio representation of conditional and joint distribution

$$\begin{aligned}
p_{jk} &= \frac{\theta_{jk}(J,K)p_{jK}}{\sum_{j=1}^{J}\theta_{jk}(J,K)p_{jK}} \\
q_{kj} &= \frac{\theta_{jk}(J,K)q_{kJ}}{\sum_{k=1}^{K}\theta_{jk}(J,K)q_{kJ}} \\
h_{jk} &= \frac{\theta_{jk}(J,K)h_{jK}h_{Jk}}{\sum_{j=1}^{J}\sum_{k=1}^{K}\theta_{jk}(J,K)h_{jK}h_{Jk}} \\
h_{jk} &= \frac{\theta_{jk}(J,K)p_{jK}q_{kJ}}{\sum_{j=1}^{J}\sum_{k=1}^{K}\theta_{jk}(J,K)p_{jK}q_{kJ}}
\end{aligned}$$

Odds ratio with condition
Define

$$\begin{aligned}
p_{jkm} &= P(Y=j \mid X=k, Z=m) \\
q_{kjm} &= P(X=k \mid Y=j, Z=m) \\
h_{jkm} &= P(Y=j, X=k \mid Z=m)
\end{aligned}$$

then

$$\theta_{jk}(J,K)(m) = \frac{p_{jkm}p_{JKm}}{p_{jKm}p_{Jkm}}$$

Odds ratio representation of conditional and joint distribution

$$\begin{aligned}
p_{jkm} &= \frac{\theta_{jk}(J,K)(m)p_{jKm}}{\sum_{j=1}^{J}\theta_{jk}(J,K)(m)p_{jKm}} \\
q_{kjm} &= \frac{\theta_{jk}(J,K)(m)q_{kJm}}{\sum_{k=1}^{K}\theta_{jk}(J,K)(m)q_{kJm}} \\
h_{jkm} &= \frac{\theta_{jk}(J,K)(m)p_{jKm}q_{kJm}}{\sum_{j=1}^{J}\sum_{k=1}^{K}\theta_{jk}(J,K)(m)p_{jKm}q_{kJm}}
\end{aligned}$$

2.3.2 Odds ratio function

Let $a(y,x)$ be an arbitrary positive domain $\mathcal{Y}\times\mathcal{X}$. Define the odds ratio function corresponding to the positive function relative to a reference point $(y_0,x_0)\in\mathcal{Y}\times\mathcal{X}$ as

$$\eta_a\left\{(y,y_0);(x,x_0)\right\} = \frac{a(y,x)a\left(y_0,x_0\right)}{a\left(y_0,x\right)a\left(y,x_0\right)}$$

it follows

$$\begin{aligned}
\eta_a\left\{(y,y_0);(x,x_0)\right\} &= \left[\eta_a\left\{(y_0,y);(x,x_0)\right\}\right]^{-1} \\
\eta_a\left\{(y,y_0);(x,x_0)\right\} &= \left[\eta_a\left\{(y,y_0);(x_0,x)\right\}\right]^{-1} \\
\eta_a\left\{(y,y_0);(x,x_0)\right\} &= \eta_a\left\{(y_0,y);(x_0,x)\right\}
\end{aligned}$$

Odds ratio function with different reference point can transform as follows

$$\eta\left\{(y,y_1);(x,x_1)\right\} = \frac{\eta\left\{(y,y_0);(x,x_0)\right\}\eta\left\{(y_1,y_0);(x_1,x_0)\right\}}{\eta\left\{(y,y_0);(x_1,x_0)\right\}\eta\left\{(y_1,y_0);(x,x_0)\right\}}$$

The odds ratio function has the following properties

$$\begin{aligned}
\eta\left\{(y,y_0);(x,x_0)\right\} &> 0 \\
\eta\left\{(y_0,y_0);(x,x_0)\right\} &= 1 \\
\eta\left\{(y,y_0);(x_0,x_0)\right\} &= 1
\end{aligned}$$

Any bivariate function has the above three properties is an odds ratio function.
Odds ratio decomposition of density function
Let $p(y \mid x)$ takes the role of $a(y,\, x)$ then

$$\eta_p\left\{(y,y_0);(x,x_0)\right\} = \frac{p(y \mid x)p\left(y_0 \mid x_0\right)}{p\left(y \mid x_0\right)p\left(y_0 \mid x\right)}$$

Example 2. For the logistic regression we have

$$p(y \mid x)=\frac{\exp \left\{y\left(\beta_0+\beta_1 x\right)\right\}}{1+\exp \left(\beta_0+\beta_1 x\right)}$$

then the odds ratio function has the following form

$$\eta\left\{\left(y, y_0\right) ;\left(x, x_0\right)\right\}=\exp \left\{\beta_1\left(y-y_0\right)\left(x-x_0\right)\right\}$$

In particular,

$$\eta\left\{\left(y=1, y_0=0\right) ;\left(x=1, x_0=0\right)\right\}=\exp \left(\beta_1\right)$$

that's why β_1 is odds interpreted as the log odds ratio

Odds ratio representation of density functions

$$\begin{aligned}
p(y \mid x) & =\frac{\eta\left\{\left(y, y_0\right) ;\left(x, x_0\right)\right\} p\left(y \mid x_0\right)}{\int \eta\left\{\left(y, y_0\right) ;\left(x, x_0\right)\right\} p\left(y \mid x_0\right) d y} \\
q(x \mid y) & =\frac{\eta\left\{\left(y, y_0\right) ;\left(x, x_0\right)\right\} q\left(x \mid y_0\right)}{\int \eta\left\{\left(y, y_0\right) ;\left(x, x_0\right)\right\} q\left(x \mid y_0\right) d x} \\
h(y, x) & =\frac{\eta\left\{\left(y, y_0\right) ;\left(x, x_0\right)\right\} h\left(y, x_0\right) h\left(y_0, x\right)}{\iint \eta\left\{\left(y, y_0\right) ;\left(x, x_0\right)\right\} h\left(y, x_0\right) h\left(y_0, x\right) d y d x}
\end{aligned}$$

Odds ratio representation of a joint conditional density From

$$\begin{aligned}
\eta_{c p}\left\{\left(y, y_0\right) ;\left(x, x_0\right) \mid z\right\} & =\frac{p(y \mid x, z) p\left(y_0 \mid x_0, z\right)}{p\left(y_0 \mid x, z\right) p\left(y \mid x_0, z\right)} \\
\eta_{c q}\left\{\left(y, y_0\right) ;\left(x, x_0\right) \mid z\right\} & =\frac{q(x \mid y, z) q\left(x_0 \mid y_0, z\right)}{q\left(x_0 \mid y, z\right) q\left(x \mid y_0, z\right)} \\
\eta_{c h}\left\{\left(y, y_0\right) ;\left(x, x_0\right) \mid z\right\} & =\frac{h(y, x \mid z) h\left(y_0, x_0 \mid z\right)}{h\left(y_0, x \mid z\right) h\left(y, x_0 \mid z\right)}
\end{aligned}$$

we can get

$$\begin{aligned}
p(y \mid x, z) & =\frac{\eta_c\left\{\left(y, y_0\right) ;\left(x, x_0\right) \mid z\right\} p\left(y \mid x_0, z\right)}{\int \eta_c\left\{\left(y, y_0\right) ;\left(x, x_0\right) \mid z\right\} p\left(y \mid x_0, z\right) d y} \\
q(x \mid y, z) & =\frac{\eta_c\left\{\left(y, y_0\right) ;\left(x, x_0\right) \mid z\right\} q\left(x \mid y_0, z\right)}{\int \eta_c\left\{\left(y, y_0\right) ;\left(x, x_0\right) \mid z\right\} q\left(x \mid y_0, z\right) d x} \\
h(y, x \mid z) & =\frac{\eta_c\left\{\left(y, y_0\right) ;\left(x, x_0\right) \mid z\right\} p\left(y \mid, x_0, z\right) q\left(x \mid y_0, z\right)}{\iint \eta_c\left\{\left(y, y_0\right) ;\left(x, x_0\right) \mid z\right\} p\left(y \mid x_0, z\right) q\left(x \mid y_0, z\right) d y d x}
\end{aligned}$$

Relationship between conditional and unconditional odds ratio functions

$$\begin{aligned}
\eta\left[\left(y, y_0\right) ;\left\{(x, z),\left(x_0, z_0\right)\right\}\right] & =\frac{b(y, x, z) b\left(y_0, x_0, z_0\right)}{b\left(y_0, x, z\right) b\left(y, x_0, z_0\right)} \\
& =\frac{b(y, x, z) b\left(y_0, x_0, z\right)}{b\left(y_0, x, z\right) p\left(y, x_0, z\right)} \frac{b\left(y, x_0, z\right) b\left(y_0, x_0, z_0\right)}{b\left(y_0, x_0, z\right) b\left(y, x_0, z_0\right)} \\
& =\eta\left\{\left(y, y_0\right) ;\left(x, x_0\right) \mid z\right\} \eta\left\{\left(y, y_0\right) ;\left(z, z_0\right) \mid x_0\right\}
\end{aligned}$$

which implies that

$$\frac{\eta\left[\left(y, y_0\right) ;\left\{(x, z),\left(x_0, z_0\right)\right\}\right]}{\eta\left\{\left(y, y_0\right) ;\left(x, x_0\right) \mid z\right\}}=\eta\left\{\left(y, y_0\right) ;\left(z, z_0\right) \mid x_0\right\}$$

also note that

$$\eta\left\{\left(y, y_0\right) ;\left(z, z_0\right) \mid x_0\right\}=\eta\left[\left(y, y_0\right) ;\left\{\left(x_0, z\right),\left(x_0, z_0\right)\right\}\right]$$

(cause $\eta\left\{\left(y, y_0\right) ;\left(x_0, x_0\right) \mid z\right\}=1$) thus we have

$$\eta\left\{\left(y, y_0\right) ;\left(x, x_0\right) \mid z\right\}=\frac{\eta\left[\left(y, y_0\right) ;\left\{(x, z),\left(x_0, z_0\right)\right\}\right]}{\eta\left[\left(y, y_0\right) ;\left\{\left(x_0, z\right),\left(x_0, z_0\right)\right\}\right]}$$

Since x and z are in symmetric positions, we have

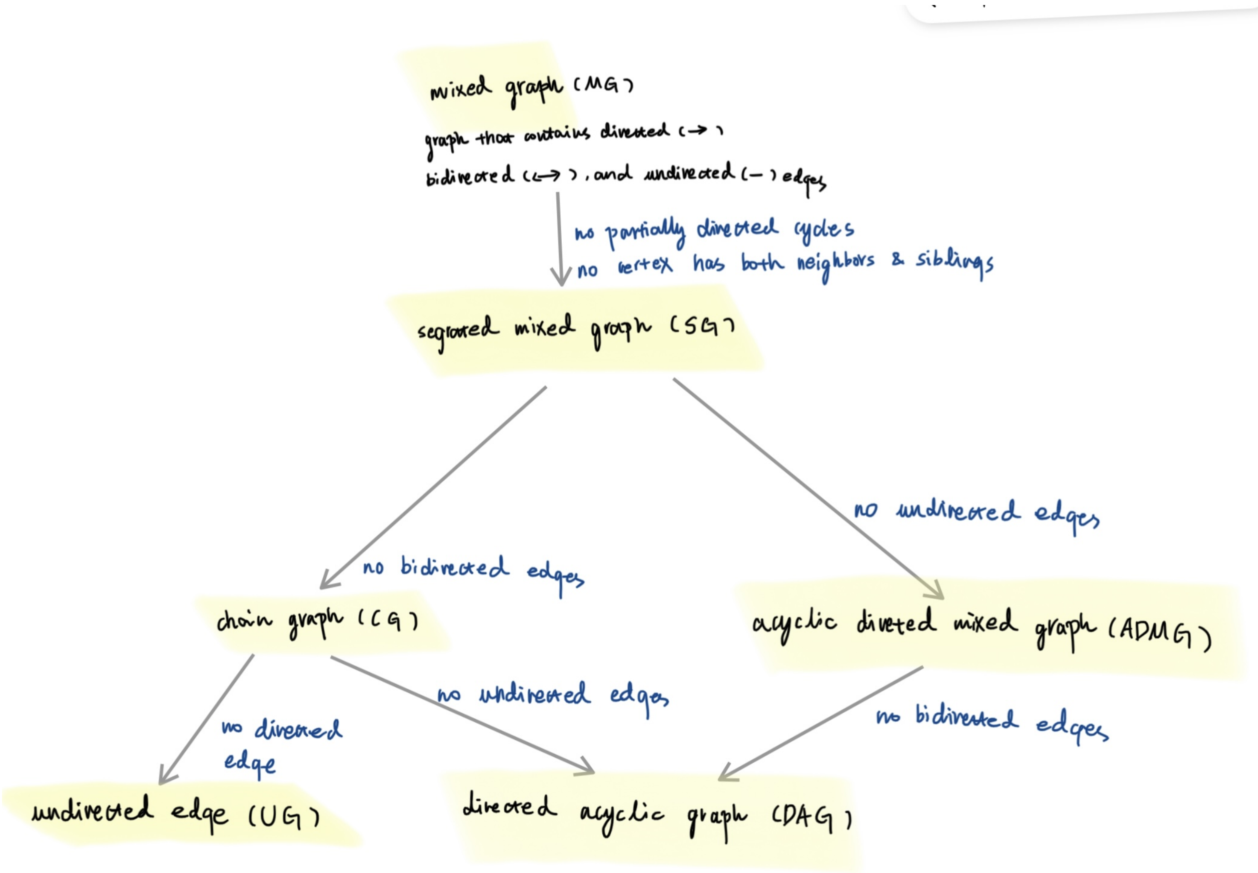
$$\eta\left[\left(y, y_0\right) ;\left\{(x, z),\left(x_0, z_0\right)\right\}\right]=\eta\left\{\left(y, y_0\right) ;\left(z, z_0\right) \mid x\right) \eta\left\{\left(y, y_0\right) ;\left(x, x_0\right) \mid z_0\right\}$$

and

$$\eta\left\{\left(y, y_0\right) ;\left(z, z_0\right) \mid x\right\}=\frac{\eta\left[\left(y, y_0\right) ;\left\{(x, z),\left(x_0, z_0\right)\right\}\right]}{\eta\left[\left(y, y_0\right) ;\left\{\left(x, z_0\right),\left(x_0, z_0\right)\right\}\right]}$$

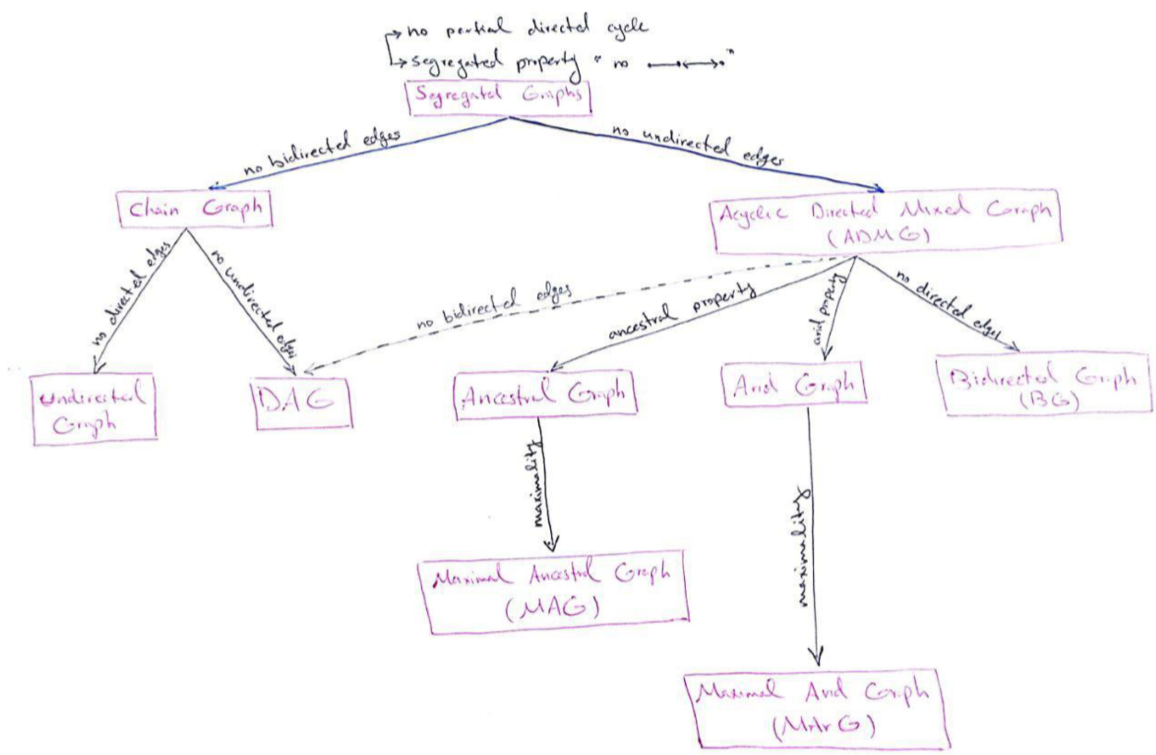
2.4 Types of graphs

From (Sherman and Shpitser)



Razieh's graph

A hierarchy of graphs



3 Mediation analysis

We are interested in

1. Directed effect: $A \longrightarrow Y$
2. Indirect effect: $A \longrightarrow M \longrightarrow Y$

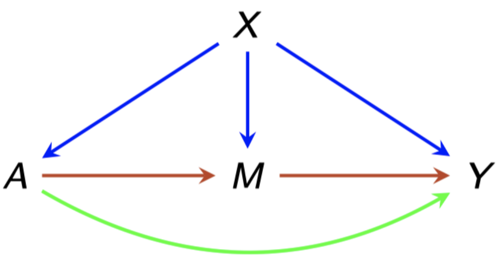


Figure 5: Mediation

Two common violation of mediation analysis is

1. exposure-mediator confounders
2. mediator-outcome confounders

(a) exposure-mediator confounders

(b) mediator-outcome confounders

Figure 6: Issues with the common approach to mediation

3.1 Mediation effects

1. Controlled direct effect (CDE)
$$\text{CDE}(m) = \mathbb{E}[Y(1, m) - Y(0, m)]$$
we need
$$\mathbb{E}[Y(a, m)]$$
to compute the controlled direct effect

2. Natural indirect effect (NID)
$$\text{NIE} = \mathbb{E}[Y(1, M(1)) - Y(1, M(0))]$$

3. Natural direct effect (NDE)
$$\text{NDE} = \mathbb{E}[Y(1, M(0)) - Y(0, M(0))]$$
we need
$$\mathbb{E}\left[Y\left(a, M\left(a'\right)\right)\right]$$
to compute the natural direct and indirect effects

Go to page 1

Relationships

$$\begin{aligned}\text{ATE} &= \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)] \\ &= \mathbb{E}[Y(1, M(1))] - \mathbb{E}[Y(0, M(0))] \\ &= \mathbb{E}[Y(1, M(1))] - \mathbb{E}[Y(1, M(0))] + \mathbb{E}[Y(1, M(0))] - \mathbb{E}[Y(0, M(0))] \\ &= \text{NIE} + \text{NDE}\end{aligned}$$

Identification assumptions for $\mathbb{E}[Y(a,m)]$

1. Conditional randomization/ignorability:
   - $Y(a) \perp A \mid X$
   - $Y(m) \perp M \mid A, X$

   Without additional information, the conditional randomization condition is untestable (or empirically unverifiable) since it does not constrain the observed data distribution. These conditions must be justified by prior knowledge and scrutinized carefully.

2. Positivity:
   - $p(A = a \mid X = x) > 0$ if $p(X = x > 0)$ (treatment positivity)
   - $p(M = m \mid A = a, X = x) > 0$ if $p(X = x > 0)$ (mediator positivity)

   The plausibility of this condition can usually be assessed empirically.

3. Consistency

Under above assumptions

$$\begin{aligned}p(Y(a,m)) &=^{\text{prob}} \sum_x p(Y(a,m) \mid X = x) \times p(X = x) \\ &=^{\text{ig}} \sum_x p(Y(a,m) \mid A = a, M = m, X = x) \times p(X = x) \\ &=^{c} \sum_x p(Y \mid A = a, M = m, X = x) \times p(X = x)\end{aligned}$$

and

$$\begin{aligned}\psi(a,m) &:= \mathbb{E}[Y(a,m)] \\ &= \sum_x \mathbb{E}[Y \mid A = a, M = m, X = x] \times p(X = x) \\ &= \mathbb{E}\left[\mathbb{E}[Y \mid A = a, M = m, X]\right] \\ &= \mathbb{E}\left[\frac{\mathbb{I}(A = a, M = m)}{p(A = a, M = m \mid X)} \times Y\right] \\ &= \mathbb{E}\left[\frac{\mathbb{I}(A = a, M = m)}{p(A = a, M = m \mid X)} \times (Y - \mathbb{E}[Y \mid A, M, X]) + \mathbb{E}[Y \mid A = a, M = m, X]\right]\end{aligned}$$

It's efficient influence function is

$$\begin{aligned}&\frac{\partial}{\partial \epsilon}\psi\left(a, m; P_\epsilon\right)\Big|_{\epsilon=0} \\ &= \int y \frac{\partial}{\partial \epsilon}\left\{dP_\epsilon(y \mid a, m, x) dP_\epsilon(x)\right\}\Big|_{\epsilon=0} \\ &= \int y \frac{\partial}{\partial \epsilon}\left\{dP_\epsilon(y \mid a, m, x)\right\} dP_\epsilon(x)\Big|_{\epsilon=0} + \int y dP_\epsilon(y \mid a, m, x) \frac{\partial}{\partial \epsilon}\left\{dP_\epsilon(x)\right\}\Big|_{\epsilon=0} \\ &= \int y S_\epsilon(y \mid a, m, x) dP(y \mid a, m, x) dP(x) + \int y S_\epsilon(x) dP(y \mid a, m, x) dP(x) \\ &= \int \frac{\mathbb{I}\left(a' = a, m' = m\right)}{p\left(a', m' \mid x\right)} y S_\epsilon\left(y \mid a', m', x\right) dP\left(y, a', m', x\right) + \int \mathbb{E}[Y \mid a, m, x] S_\epsilon(x) dP(x) \\ &=_{*} \int \frac{\mathbb{I}\left(a' = a, m' = m\right)}{p\left(a', m' \mid x\right)}(y - \mathbb{E}[Y \mid a, m, x]) S_\epsilon\left(y \mid a', m', x\right) dP(o) + \int (\mathbb{E}[Y \mid a, m, x] - \psi(a, m)) S_\epsilon(x) dP(x) \\ &=_{**} \int \frac{\mathbb{I}\left(a' = a, m' = m\right)}{p\left(a', m' \mid x\right)}(y - \mathbb{E}[Y \mid a, m, X]) S_\epsilon(o) dP(o) + \int (\mathbb{E}[Y \mid a, m, x] - \psi(a, m)) S_\epsilon(o) dP(o) \\ &= \int \left\{\frac{\mathbb{I}\left(a' = a, m' = m\right)}{p\left(a', m' \mid x\right)}(y - \mathbb{E}[Y \mid a, m, X]) + \mathbb{E}[Y \mid a, m, x] - \psi(a, m)\right\} S_\epsilon(o) dP(o)\end{aligned}$$

Go to page 1

* $\mathbb{E}\left[S_\epsilon\left(y \mid a', m', x\right) \mid a', m', x\right] = 0$ and $\mathbb{E}\left[S_\epsilon(x)\right] = 0$
** $\int \frac{\mathbb{1}(a'=a, m'=m)}{p(a', m' \mid x)}(y - \mathbb{E}[Y \mid a, m, X]) S_\epsilon\left(a', m', x\right) dP(o) = 0$
Therefore

$$
\begin{aligned}
\text{CDE}(m) &= \mathbb{E}\left[\mathbb{E}[Y \mid A=1, M=m, X] - \mathbb{E}[Y \mid A=0, M=m, X]\right] \\
&= \mathbb{E}\left[\frac{\mathbb{1}(A=1, M=m) \times Y}{p(A=1 \mid X) \times p(M=m \mid A=1, X)} - \frac{\mathbb{1}(A=0, M=m) \times Y}{p(A=0 \mid X) \times p(M=m \mid A=0, X)}\right] \\
&= \mathbb{E}\Bigg[\left\{\frac{\mathbb{1}(A=1, M=m)}{p(A=1 \mid X) \times p(M=m \mid A=1, X)} - \frac{\mathbb{1}(A=0, M=m)}{p(A=0 \mid X) \times p(M=m \mid A=0, X)}\right\} Y \\
&\quad + \mathbb{E}[Y \mid A=1, M=m, X] - \mathbb{E}[Y \mid A=0, M=m, X]\Bigg]
\end{aligned}
$$

Identification assumptions for $\mathbb{E}\left[Y\left(a, M\left(a'\right)\right)\right]$

1. Conditional ignorability:
   - $Y(a) \perp A \mid X$ (same as in CDE)
   - $Y(m) \perp M \mid A, X$ (same as in CDE)
   - $M(a) \perp A \mid X$
2. Cross-world assumption:
$$Y(a, m) \perp M\left(a'\right) \mid X$$
   it can be think of as separable direct effect

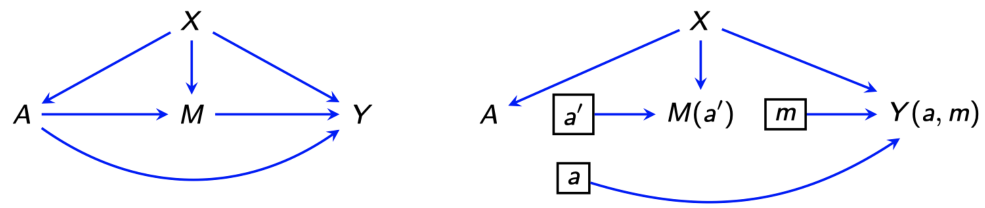


Figure 7: separable direct effect

3. Positivity: (same as in CDE)
   - $p(A=a \mid X=x) > 0$ if $p(X=x>0)$ (treatment positivity)
   - $p(M=m \mid A=a, X=x) > 0$ if $p(X=x>0)$ (mediator positivity)
4. Consistency: (same as in CDE)

With these assumptions

$$
\begin{aligned}
p\left(Y\left(a, M\left(a'\right)\right)\right) &=^{\text{def}} \sum_m p\left(Y(a, m), M\left(a'\right)=m\right) \\
&=^{\text{prob}} \sum_{m,x} p\left(Y(a, m), M\left(a'\right)=m, x\right) \\
&=^{\text{pron}} \sum_{m,x} p\left(Y(a, m) \mid M\left(a'\right)=m, x\right) \times p\left(M\left(a'\right)=m \mid x\right) \times p(x) \\
&=^{(i)} \sum_{m,x} p(Y(a, m) \mid x) \times p\left(M\left(a'\right)=m \mid x\right) \times p(x) \\
&=^{(ii,iii)} \sum_{m,x} p(Y(a, m) \mid A=a, x) \times p\left(M\left(a'\right)=m \mid A=a', x\right) \times p(x) \\
&=^{(c,iv)} \sum_{m,x} p(Y(m) \mid A=a, M=m, x) \times p\left(M=m \mid A=a', x\right) \times p(x) \\
&=^{c} \sum_{m,x} p(Y \mid A=a, m, x) \times p\left(M=m \mid A=a', x\right) \times p(x)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[Y\left(a, M\left(a'\right)\right)\right] &= \sum_{m,x} \mathbb{E}[Y \mid A=a, m, x] \times p\left(M=m \mid A=a', x\right) \times p(x) \\
&= \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}[Y \mid A=a, M, X] \mid A=a', X\right]\right] \\
&= \mathbb{E}\left[\frac{\mathbb{I}(A=a)}{p(A \mid X)} \times \frac{p\left(M \mid A=a', X\right)}{p(M \mid A, X)} \times Y\right] \\
&= E\left[\frac{\mathbb{I}(A=a)}{g_A(a \mid X)} \frac{g_M\left(m \mid a', X\right)}{g_M(m \mid a, X)}(Y - b(m, a, X))\right. \\
&+ \frac{\mathbb{I}\left(A=a'\right)}{g_A\left(a' \mid X\right)}\left(b(m, a, X) - \int b(m, a, X) g_M\left(m \mid a', X\right) d\nu(m)\right) \\
&\left. + \int b(m, a, X) g_M\left(m \mid a', X\right) d\nu(m)\right]
\end{aligned}
$$

where

$$
\begin{aligned}
b(m, a, x) &:= \mathbb{E}[Y \mid M=m, A=a, X=x] \\
g_M(m \mid a, x) &:= p(M=m \mid A=a, X=x) \\
g_A(a \mid x) &:= p(A=a \mid X=x)
\end{aligned}
$$

3.2 Multiple mediators

For 8, there are multiple causal pathways:

$$
\begin{aligned}
&A \to Y \\
&A \to L \to Y \\
&A \to L \to M \to Y \\
&A \to M \to Y
\end{aligned}
$$

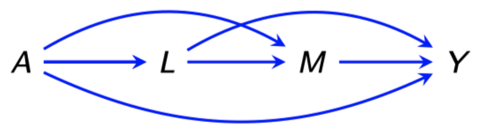


Figure 8: Multiple mediators

we might be interested in

- effect of A on Y through L. That is
  - $A \to L \to Y$
  - $A \to L \to M \to Y$

  which can be expressed as
  $$\mathbb{E}\left[Y\left(a, L\left(a'\right), M\left(a, L\left(a'\right)\right)\right)\right] - E[Y(a)]$$

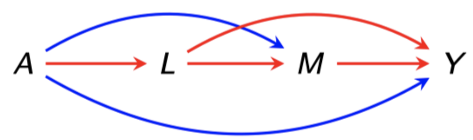


Figure 9: example of an identified effect

- effect of A on Y through L that doesn't go through M. That is
  $$A \to L \to Y$$
  which can be expressed as
  $$\mathbb{E}\left[Y\left(a, L\left(a'\right), M(a, L(a))\right)\right] - E[Y(a)]$$

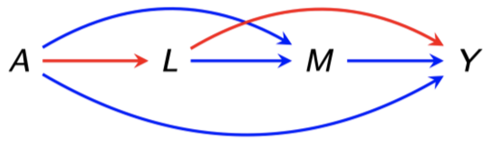


Figure 10: example of an unidentified effect

References

[Benkeser] Benkeser, D. BIOS 761 notes.

[Chen] Chen, H. Y. Semiparametric odds ratio model and its applications.

[Glymour et al.] Glymour, M., Pearl, J., and Jewell, N. P. *Causal inference in statistics: A primer*. John Wiley & Sons.

[Nabi] Nabi, R. BIOS 760r notes.

[Neal] Neal, B. Introduction to causal inference from a machine learning perspective.

[Sherman and Shpitser] Sherman, E. and Shpitser, I. Identification and Estimation of Causal Effects from Dependent Data. page 12.