MCI revision

Basics

conditional gaussian: $p(A=a|B=b)=N(a;\tilde{\mu},\tilde{\sigma}^2)$, $\tilde{\mu}=\mu_A+\rho\frac{\sigma_A}{\sigma_B}(b-\mu_B)$, $\tilde{\sigma}^2=\sigma_A^2(1-\rho^2)$, $\rho=\frac{\mathrm{Cov}(A,B)}{\sigma_A\sigma_B}$.

graphs: vertices, edges, acyclic, tree, chain, DAG

- **D-separation**: fork & chain nodes in Z; collider (and its descendants) not in Z
- directed / undirected graph encode strictly different (conditional) independence information
- Markov blanket $X_i \perp\!\!\!\perp S \backslash S_1 | S_1$. Markov boundary: parents, children, coparents

examples: Monte Hall problem, Simpson's paradox (multiple regression when covariates are not independent)

ladder of causation: association, intervention, counterfactual.

structural Causal Models (SCMs): exogenous variables (U, jointly independent), endogenous variables (V), a set of functions (f). G acyclic. $X_j := f_j(PA_j, N_j), \quad j=1,\ldots,d.$

treatment (T), **outcome (Y)**, **confounders (U)**

Randomized Control Trials (RCT): Subjects are assigned at random to various groups (treatment / control)

stratification: dividing study subjects into different groups or strata to better understand causal effects within a group & differences between groups.

identifiability: hypothetical variables identifiable from observational data.

Causal Effect Estimation

Rubin's

1. observed confounders

   potential outcome ($y_0^{(i)}$, $y_1^{(i)}$): value y would have taken if individual i had been under treatment t (not observed).

   > observed outcomes / counterfactual: ($y_0^{(i)}$ or $y_1^{(i)}$, depend on the treatment)
   >
   > **counterfactual**: the outcome would have been observed if taken the other treatment
   >
   > $y_{obs}^{(i)}=t^{(i)}y_1^{(i)}+(1-t^{(i)})y_0^{(i)}$
   > $y_{CF}^{(i)}=t^{(i)}y_0^{(i)}+(1-t^{(i)})y_1^{(i)}$

   assumptions:

   - **Stable Unit Treatment Value Assumption (SUTVA)**
     - Consistency: if $T=t$ then $Y_t=Y$. Well-defined treatment, potential outcome independent of how treatment is assigned
     - no interference: individuals in a population do not influence each other
   - **Positivity**: $P(T=1|X=x)\in(0,1)$ if $P(X=x)>0$, non-zero chance of individual receiving treatment.
   - **Unconfoundedness (ignorability)**: $y_1^{(i)}, y_0^{(i)} \perp\!\!\!\perp t^{(i)}|x$ ($P(Y_t|T,X)=P(Y_t|X)$), Given confounding features X, treatment assignment is random; no unobserved confounders.

   adjustment formula: (can be estimated from observational data)

   - **average treatment effect (ATE)**: $\tau=\hat{\mathbb{E}}[\tau^{(i)}]=\hat{\mathbb{E}}[y_1^{(i)}-y_0^{(i)}]$.

     $$\begin{aligned} ATE=\mathbb{E}[Y_1-Y_0] &= \mathbb{E}_X[\mathbb{E}[Y_1-Y_0|X]] \\ &= \mathbb{E}_X[\mathbb{E}[Y_1|X]-\mathbb{E}[Y_0|X]] \\ &= \mathbb{E}[\mathbb{E}[Y_1|T=1,X]-\mathbb{E}[Y_0|T=0,X]] \quad \text{Unconfoundedness} \\ &= \mathbb{E}[\mathbb{E}[Y|T=1,X]-\mathbb{E}[Y|T=0,X]] \quad \text{consistency} \end{aligned}$$

     under linearity assumption: $\mathbb{E}[Y|T,X]=\alpha_0+\beta_x X+\beta_t T+\epsilon$, $ATE=\beta_t$.
   - **average treatment effect of the treated (ATT)**: $ATT=\hat{\mathbb{E}}[y_1^{(i)}-y_0^{(i)}|t^{(i)}=1]$ (same as ATE if no confounders!).

     $$\begin{aligned} ATT &= \mathbb{E}[Y_1-Y_0|T=1] \\ &= \mathbb{E}[Y_1|T=1]-\mathbb{E}[Y_0|T=1] \quad \text{(counterfactual)} \\ &= \mathbb{E}[Y|T=1]-\Sigma_y yp(Y_0=y|T=1) \\ &= \mathbb{E}[Y|T=1]-\Sigma_{x,y} yp(Y_0=y|T=1,X=x)p(X=x|T=1) \quad \text{marginalize} \\ &= \mathbb{E}[Y|T=1]-\Sigma_{x,y} yp(Y_0=y|T=0,X=x)p(X=x|T=1) \quad \text{Unconfoudedness} \\ &= \mathbb{E}[Y|T=1]-\Sigma_{x,y} yp(Y=y|T=0,X=x)p(X=x|T=1) \quad \text{consistency} \\ &= \mathbb{E}[Y|T=1]-\Sigma_x \mathbb{E}[Y|T=0,X=x]p(X=x|T=1) \end{aligned}$$
   - **conditional average treatment effect (CATE)**: $CATE=\mathbb{E}[Y_1-Y_0|X=x]$.
   - **causal interactions of two treatments on outcome** (with confounder X):

$I^a_{i,j} = [\mathbb{E}(Y|(T_1,T_2)=(1,1),X) - \mathbb{E}(Y|(T_1,T_2)=(0,1),X)] - [\mathbb{E}(Y|(T_1,T_2)=(1,0),X) - \mathbb{E}(Y|(T_1,T_2)=(0,0),X)]$
.

under linearity assumption: $\mathbb{E}[Y|T,X] = \alpha_0 + \beta_x X + \alpha_1 T_1 + \alpha_2 T_2 + \gamma T_1 T_2$, $I^a_{1,2} = \gamma = I^a_{2,1}$.

balancing score: function $b(x)$ making $x \perp\!\!\!\perp t|b(x)$.

- *finest*: $b(x) = x$. OK for binary confounders, but only give point estimates for continuous ones.
- *coarsest*: **propensity score** $e(x) = P(T=1|X=x)$. better estimates.
  - it is a function of every balancing score: $e(x) = f(b(x))$.
- **Unconfoundedness given a balancing score**: $y_1^{(i)}, y_0^{(i)} \perp\!\!\!\perp t^{(i)}|x^{(i)} \Rightarrow y_1^{(i)}, y_0^{(i)} \perp\!\!\!\perp t^{(i)}|b(x^{(i)})$.

(Propensity Score) Matching: match control & treatment individuals based on their propensity score.

- *greedy matching*: always find the one with smallest distance for current unit.
- *optimal matching*: minimize the global distance, computationally demanding.

Inverse Probability of Treatment Weighting (IPTW): inflate the weight for under represented-subjects due to missing data (considering covariate imbalances in receiving treatments).

weight:

$$w_i = \begin{cases} \frac{1}{e(x_i)} & \text{if } t_i = 1 \\ \frac{1}{1-e(x_i)}, & \text{if } t_i = 0 \end{cases}$$

$$\hat{\tau} = \frac{1}{N}\Sigma_{\text{treated}} y_1^{(i)} \frac{1}{e(x_i)} - \frac{1}{N}\Sigma_{\text{not treated}} y_0^{(i)} \frac{1}{1-e(x_i)}$$

may have inaccurate weights with very low propensity scores.

Ideal scenario: randomized control trial (RCT), no imbalances, t independent to x, without weight differences.

Sensitivity check: unobserved confounders may exist (confounders fundamentally unverifiable), hidden bias (not *uncertainty*) & its severity.

- Quick and simple **sanity checks** (do the *drawings*)
  - *Random 'unobserved' common cause*: add an independently and randomly simulated confounder affecting treatment & outcome (noise, shifts), re-run analysis. result do not change much, then original CE significant.
  - *Placebo treatment effect*: replace treatment with randomly generated placebo. new estimate should be statistically 0.
  - *Subset / validate the data*: cross-validation / bootstrap. results statistically the same.
- **Super Learning** other potential confounders: treat considering different confounder sets as different models for the SL to choose from with cross-validation (feature selection), test if the estimates change / stabilize at some order.
- **Deriving bounds** on the estimates
  - degree of bias, Γ: take individuals i and j such that $X^{(i)} = X^{(j)}$, then $e^{(i)} = e^{(j)}$, but $\frac{1}{\Gamma} \leq \frac{\frac{e^{(i)}_{\text{true}}}{1-e^{(i)}_{\text{true}}}}{\frac{e^{(j)}_{\text{true}}}{1-e^{(j)}_{\text{true}}}} \leq \Gamma$ (only in hypothesis).
  - p-value, t-test: $\frac{\text{ATE}}{\sigma_{\text{ATE}}} \sim t/z\text{-distributed}, p\text{-value} = P(|\frac{signal}{noise}| > t_0|H_0)$.

    Too small p-value, reject H_0 (True / False Positive, Type I error); Otherwise, not reject H_0 (True / False Negative, Type II error).

2. unobserved confounders

   violate unconfoundedness, introducing bias. e.g. $Y = \tau T + \delta_U U, T = \gamma_U U$, then naive regression of Y on T will yield $\frac{\text{Cov}[T,Y]}{\text{Var}[T]} = \tau + \frac{\delta_U}{\gamma_U}$, instead of τ.

   instrumental variable (IV): intention-to-treat variable (randomized)

   assumptions:

   - **SUTVA**
   - **Z relevant**: $\mathbb{E}[(T^{(i)}|z=1) - (T^{(i)}|z=0)] \neq 0$, treatment assignment Z associated with the treatment is not zero (Z actually doing something).
   - **Z random**: $(Y^{(i)}|z=1,t) = (Y^{(i)}|z=0,t)$, Z & Y do not share a cause.
   - **Exclusion Restriction**: any effect of Z on Y is via an effect of Z on T (Z should not affect Y when T is held constant). (can be ensured by **double-blind studies** if possible)
   - **Monotonicity**: $(T^{(i)}|z=1) \geq (T^{(i)}|z=0)$, increasing encouragement "dose" increases probability of treatment, no *defiers*, only *compliers*.

   General case:

$$\begin{aligned} \tau &= \mathbb{E}[Y_1 - Y_0] \quad \text{(potential outcomes)} \\ &= \frac{\mathbb{E}[(Y|z=1) - (Y|z=0)]}{\mathbb{E}[(T|z=1) - (T|z=0)]} \\ \hat{\tau} &= \frac{\frac{1}{n_{z=1}}\Sigma_{i\in z=1} Y^{(i)} - \frac{1}{n_{z=0}}\Sigma_{i\in z=0} Y^{(i)}}{\frac{1}{n_{z=1}}\Sigma_{i\in z=1} T^{(i)} - \frac{1}{n_{z=0}}\Sigma_{i\in z=0} T^{(i)}} \quad \text{no bias given randomized z} \end{aligned}$$

   Linear case:

- *estimand*: $\tau = \frac{\text{Cov}(Y,Z)}{\text{Cov}(T,Z)}$, $\hat{\tau} = \frac{\hat{\text{Cov}}(Y,Z)}{\hat{\text{Cov}}(T,Z)} = \frac{\beta_{\text{Y on Z}}}{\beta_{\text{T on Z}}}$.
- *two-stage OLS*: estimate $\mathbb{E}[T|Z]$, obtain $\hat{T}$; estimate $\mathbb{E}[Y|\hat{T}]$, obtain $\hat{\tau}$.

3. considering change over time**!!!!!!!**

   Difference in Difference: treatment effect on outcome is estimated as the **difference in changes** over time between the two groups (difference in trends), since treatment is **not random** and have **different starting points**.

   components: treatment & control *group*, data before & after the treatment is applied.

   assumptions:

   - **parallel trends assumption**: $(Y_0(s_1) - Y_0(s_0)) \perp\!\!\!\perp \text{Group}$, $\mathbb{E}[Y_0(s_1)|\text{Group} = T] = \mathbb{E}[Y_0(s_0)|\text{Group=T}] + \mathbb{E}[Y_0(s_1) - Y_0(s_0)|\text{Group} = C]$.
   - **no pre-treatment effect**: participants knowing being assigned to the treatment group do not change their behaviors.

   estimand:

   $$\begin{aligned} ATT &= \mathbb{E}[Y_1 - Y_0|\text{Group} = T] \\ &= \mathbb{E}[Y_1(s_1) - Y_0(s_1)|\text{Group} = T] \\ &= \mathbb{E}[Y(s_1)|\text{Group=T})] - \mathbb{E}[Y_0(s_1)|\text{Group=T})] \quad \text{consistency} \\ &= \mathbb{E}[Y(s_1)|\text{Group=T})] - (\mathbb{E}[Y_0(s_0)|\text{Group=T})] + \mathbb{E}[Y_0(s_1) - Y_0(s_0)|\text{Group=C})]) \quad \text{parallel trends} \\ &= (\mathbb{E}[Y(s_1)|\text{Group} = T] - \mathbb{E}[Y(s_0)|\text{Group} = T]) - (\mathbb{E}[Y(s_1)|\text{Group} = C] - \mathbb{E}[Y(s_0)|\text{Group} = C]) \quad \text{consistency} \end{aligned}$$

4. violation of positivity

   Sharp Regression Discontinuity (SRD): looks at discontinuity in outcome at the *cut-off*, any discontinuity at cut-off is *only* due to the treatment.

   - design: $T(W) = \mathbb{I}\{W \geq c\} = \begin{cases} 1, & \text{if} \quad W \geq c \\ 0, & \text{if} \quad W < c \end{cases}$(cut-off $W = c$).
   - assumptions: function $\mu_t(W) = \mathbb{E}[Y_t|W = w]$ are continuous (at least at $w = c$).
   - estimand:

   $$\begin{aligned} \hat{\tau}_{\text{SRD}} = \mathbb{E}[Y_1 - Y_0|W = c] &= \lim_{w \downarrow c} \mathbb{E}[Y_1|W = w] - \lim_{w \uparrow c} \mathbb{E}[Y_0|W = w] \\ &= \lim_{w \downarrow c} \mathbb{E}[Y|W = w] - \lim_{w \uparrow c} \mathbb{E}[Y|W = w] \quad \text{consistency} \end{aligned}$$

5. **Counterfactual**: $\mathbb{E}[Y_{T=1}|T = 0, Y = Y_0 = 40\text{mins}]$, 'if' statement in which the condition is unrealized (hypothetical world vs. actual world), *defining* it should not require approximation. Used in scenarios when we don't want to specify T to some amount by disabling all pre-existing causes of T (applying *do*), we want to keep incoming arrows of T. *Comparison*:

	Counterfactual $\mathbb{E}[Y_t|T = t', Y_{t'}]$	**do operator** $\mathbb{E}[Y|do(T = t)]$
condition	condition on the actual world	do not reference the other world whatsoever
captures what	describes behaviors of a specific individual $U = u$ under such intervention	captures behaviors of population under intervention
estimate	ATT, $P(X = x|T = t')$	ATE, $P(X = x)$
policy / scientific	**policy question**, population dependent (**free choice**, people with different X may tend to choose different T, we then average over to get the result. Not complete randomization), reveal population-based effects	**scientific intervention**, population independent (**experimental design**, work with random sample), reveal micro-level meaningful effect

   applications:

   - **ATT**, $\mathbb{E}[Y_x|X = x']$: recruitment of a program, additive interventions;
   - **probability of necessity**, $\text{PN} = P(Y_x = y|X = x', Y = y')$: cancer treatment, legal liability (attribution, "but for");
     - probability of sufficiency: $\text{PC} = P(Y_{x'} = y'|T = x, Y = y)$;
     - probability of necessity & sufficiency: $\text{PNS} = P(Y_x = y, Y_{x'} = y')$, estimated by $P(Y = 1|do(T = 1)) - P(Y = 1|do(T = 0))$ under monotonicity.

     theorem: under monotonicity assumption + do identifiable:

     $$\begin{aligned} \text{PN} &= \frac{p(y) - p(y|do(x'))}{p(x, y)} \\ &= \frac{p(y|x) - p(y|x')}{p(y|x)} + \frac{p(y|x') - p(y|do(x'))}{p(x, y)} \end{aligned}$$

     - **Excess Risk Ratio (ERR)** or **Attributable Risk Fraction among the exposed**: how much likely is y when $X = x$ than not, $X = x'$.
     - **Confounding Factor (CF)**: corrects for confounding bias due to confounding of the causal effect of X on Y (do $\neq$ conditional)

     In experimental settings, $\text{PN} = \text{ERR}$, gives a false impression (need the confounding bias for real PN!)
   - **nested counterfactual expression**, $\mathbb{E}[Y_{x, M_{x'}}]$: key quantity in mediation (see mediation).

Pearl's: causal graphical models + structual equations

do vs. condition: intervention vs. observation (graph surgeries)

1. observed confounders

   adjustment formula: $p(Y=y|do(T=t))=\Sigma_x p(Y=y|T=t,X=x)p(X=x)$.

   $ATE=p(Y=1|do(T=1))-p(Y=1|do(T=0))$. same as Rubin's.

   (* set of parents of T is always an adjustment set for (T,Y))

   backdoor criterion: variable set X satisfies the backdoor criterion relative to (T,Y) if:

   - no node in X is a descendent of T;
   - X block every path between T and Y that contains an arrow into T ("spurious path").

   then the causal effect of T on Y is based on the adjustment formula adjusted on X.

   optimal adjustment set (for smaller error, minimize $\frac{\text{variance in }Y}{\text{variance in }X}$): $\mathrm{pa}_G(\mathrm{cn}_G(X\to Y))\backslash(\mathrm{cn}_G(X\to Y)\cup\{X\})$ ($\mathrm{cn}_G(X\to Y)$: all nodes on a *directed* path from X to Y, excluding X)

2. unobserved confounders

   front-door formula:

$$\begin{aligned}p(Y=y|do(X=x))&=\Sigma_z p(Y=y|do(Z=z))p(Z=z|do(X=x))\\&=\Sigma_z p(Z=z|X=x)\Sigma_{x'}p(Y=y|Z=z,X=x')p(X=x')\end{aligned}$$

   front-door criterion: variable set Z satisfies the front-door criterion relative to (X,Y) if:

   - Z intercepts all *directed* paths from X to Y;
   - all paths from X to Z are blocked;
   - All backdoor paths from Z to Y are blocked by X.

   if also $p(x,z)>0$, then the causal effect of X on Y is based on the front-door formula on Z.

3. generalization: **do-calculus**: (*derivation of front-door criterion)

   ($G_{\overline{X}}$: graph with all arrows pointing to nodes in X deleted; $G_{\underline{X}}$: graph with all arrows emerging from nodes in X deleted)

   - **Rule1** (insertion / deletion of observations): $p(Y|do(X=x),Z,W)=p(Y|do(X=x),W)$ if $(Y\perp\!\!\!\perp Z)|X,W$ in $G_{\overline{X}}$.

     (generalization of d-separation with intervention $do(X=x)$, special case: $X=\emptyset$)

   - **Rule2** (Action / observation exchange): $p(Y|do(X=x),do(Z=z),W)=p(Y|do(X=x),z,W)$ if $(Y\perp\!\!\!\perp Z)|X,W$ in $G_{\overline{X}\underline{Z}}$.

     (generalization of backdoor criterion, special case: $X=\emptyset$)

   - **Rule3** (Insertion / deletion of actions): $p(Y|do(X=x),do(Z=z),W)=p(Y|do(X=x),W)$ if $(Y\perp\!\!\!\perp Z)|X,W$ in $G_{\overline{X}\overline{Z(W)}}$ ($Z(W)$: the set of Z-nodes not ancestors of any W-node in $G(\overline{X})$).

4. **IPW (From Pearl's)**: computational savings when Z contains too many values (but few actually appears) / Number of $Z=z$ samples too small. *when Z satisfies backdoor*:

$$\begin{aligned}p(Y=y|do(X=x))&=\Sigma_z p(Y=y|X=x,Z=z)p(Z=z)\\&=\Sigma_z\frac{p(Y=y,X=x,Z=z)}{p(X=x|Z=z)}\quad\text{redistribution of population with a factor (propensity)}\end{aligned}$$

5. z**-specific effect** (similar to CATE): $p(Y=y|do(T=t),Z=z)=\Sigma_s P(Y=y|T=t,S=s,Z=z)P(S=s|Z=z)$ ($S\cup Z$ satisfies backdoor criterion)

6. **conditional interventions**: involving z-dependent policies.

$$\begin{aligned}p(Y=y|do(T=g(Z)))&=\Sigma_z p(Y=y|do(T=g(Z)),Z=z)p(Z=z|do(T=g(z)))\\&=\Sigma_z p(Y=y|do(T=g(Z)),Z=z)p(Z=z))\quad\text{Z occurs before T}\\&=\Sigma_z p(Y=y|do(T=t),Z=z)|_{t=g(z)}p(Z=z)\quad\text{can continue with z-specific effect}\end{aligned}$$

7. **Mediation**: discriminate between direct & indirect effects.

   (a vivid demonstration of difference between Counterfactual & do)

   - do-expressions: can be estimated from experimental / observational data with back / front -door criteria. (**intervene the indirect effect $T\to M\to Y$ from M**)

     1. **Total Effect (TE)**: $\mathrm{TE}=\mathbb{E}[Y_1-Y_0]=\mathbb{E}[Y|do(T=1)]-\mathbb{E}[Y|do(T=0)]$, measures expect increase in Y as treatment changes from $T=0$ to $T=1$ while mediator M changes freely (as per the structural function f_M).
     2. **Controlled Direct Effect (CDE(m))**: $\mathrm{CDE}=\mathbb{E}[Y_{1,m}-Y_{0,m}]=\mathbb{E}[Y|do(T=1,M=m))-p(Y|do(T=0,M=m))$, measures expect increase in Y as treatment changes from $T=0$ to $T=1$ while mediator is set to $M=m$ uniformly (focusing on the direct one; condition on M may open backdoor paths, thus need two dos here).

        criterion: CDE related to (T,Y) meditated by X identifiable if:

        - exists S_1 blocking all backdoor paths from X to Y (for $do(X=x)$);
        - exists S_2 blocking all backdoor paths from T to Y after $do(X=x)$ (for $do(T=t)$).

   - Counterfactuals: not do expressions (**intervene the indirect effect $T\to M\to Y$ from T**)

1. **Natural Direct Effect (NDE)**: $\text{NDE} = \mathbb{E}[Y_{1,M_0} - Y_{0,M_0}]$, measures expected increase in Y as treatment changes from $T = 0$ to $T = 1$ while mediator is set to whatever value it would have attained (for each individual) prior to change, that is, under $T = 0$.
2. **Natural Indirect Effect (NIE)**: $\text{NIE} = \mathbb{E}[Y_{0,M_1} - Y_{0,M_0}]$, measures the expected increase in Y as treatment is held constant at $T = 0$, and the mediator M changes to whatever value it would have attained (for each individual, in a *natural* unfrozen way) as treatment changes from $T = 1$ to $T = 0$. captures the portion of the effect that can be explained by mediation alone, while disabling (or "freezing") the capacity of Y respond to T (direct effect).

cannot just remove the edge, since the observational data is still based on the original graph with (in)direct paths

allow X vary naturally between applicants, as oppose to CDE (do).

criterion: There exists a set of measured covariates W, s.t.

1. no member of W is descendant of T;
2. W blocks all backdoor paths from M to Y (after removing the arrows $T \to M$ and $T \to Y$);
3. W-specific effect of T on M is identifiable, possibly using experiments;
4. W-specific joint effect of {T, M} on Y is identifiable, possibly using experiments.
5. The exogenous variables $U = (U_T, U_M, U_Y)$ are mutually independent (no confounder W)

When I and II hold, NDE experimentally identifiable:
$\text{NDE} = \Sigma_m \Sigma_w [\mathbb{E}[Y|do(T = 1, M = m), W = w] - \mathbb{E}[Y|do(T = 0, M = m), W = w]] \times p(M = m|do(T = 0), W = w)p(W =$
. with III and IV, do expressions are further guaranteed identifiable with back / front door. If W deconfound the relationships in III and IV,
$\text{NDE} = \Sigma_m \Sigma_w [\mathbb{E}[Y|T = 1, M = m, W = w] - \mathbb{E}[Y|T = 0, M = m, W = w]] \times p(M = m|T = 0, W = w)p(W = w)$
. with V (all satisfies), then $\text{NDE} = \Sigma_m [\mathbb{E}[Y|T = 1, M = m] - \mathbb{E}[Y|T = 0, M = m]]p(M = m|T = 0)$. **NDE is a weighted average of CDE**.

same with $\text{NIE} = \Sigma_m \mathbb{E}[Y|T = 0, M = m](p(M = m|T = 1) - p(M = m|T = 0))$.

- response factors
  1. NDE/TE: measures fraction of response that is transmitted directly, with M 'frozen' *naturally*.
  2. NIE/TE: measures fraction of response that may be transmitted through M, with Y blinded to T.
  3. $(TE - NDE)/TE$: measures fraction of response that is necessary due to M.

under linearity assumption1: $\begin{cases} y = \beta_1 m + \beta_2 t + u_y \\ m = \gamma_1 t + u_m \end{cases}, \begin{cases} TE = \beta_2 + \gamma_1 \beta_1 \\ NDE = \beta_2 \\ NIE = \gamma_1 \beta_1 \end{cases}.$

under linearity assumption2: $\begin{cases} y = \beta_1 m + \beta_2 t + \beta_3 tm + \beta_4 w + u_y \\ m = \gamma_1 t + \gamma_2 w + u_m \\ w = \alpha t + u_w \end{cases}, \begin{cases} TE = (\beta_1 + \beta_3)(\gamma_1 + \gamma_2 \alpha) + \beta_2 + \beta_4 \alpha \\ NDE = \beta_2 + \beta_4 \alpha \\ NIE = \beta_1(\gamma_1 + \gamma_2 \alpha) \end{cases}.$

Causal discovery: learning set of edges from data (causal structure constraints)

1. constraint-based

   assumptions: **Markov condition**; **causal sufficiency**; **faithfulness** (probability distribution P presents no CI relations other than the ones entailed by DAG G, possibly fails when paths exactly cancels or regulatory systems)

   - by Markov Equivalence Class (MEC) & d-separations, super inefficient, search space grows exponentially in the number of nodes.
   - **Peter-Clark (PC) algorithm**: start with complete graph; based on n-th order CI (conditioning sets only need to contain neighbors of the 2 nodes), remove edges by faithfulness, until no higher order CI observed; add directions, take triplets where 2 nodes are connected to the 3rd (based on colliders, $A \not\perp\!\!\!\perp B|C$)