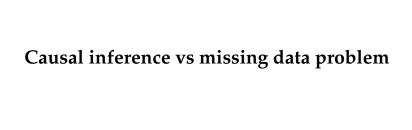
### Causal Inference: A View from Missing Data Problem



#### Counterfactual outcome

- ▶ Let A denote exposure, treatment or action of interest and take values from a set A.
- ► Counterfactual outcomes are the outcome values associated with each  $a \in \mathcal{A}$  if the treatment a is given, denoted by  $\mathcal{Y} \equiv \{Y(a) : a \in \mathcal{A}\}$ .
- ▶ Causal inference is essentially a framework to make inferences about the random variables in  $\mathcal{Y}$ , including evaluation of E[Y(a)], comparison between E[Y(a)] and E[Y(a')], or comparison of these means given covariates.
- ► However, observed data consist of the actual treatment/exposure *A* and the corresponding outcome *Y* for any individual, instead of the set of counterfactual outcomes.
- ▶ Therefore, causal inference based on the observed data is essentially a missing data problem where  $\mathcal{Y}$  is missing!

## Casted into missing data notations under SUTV condition

- ► Consistency/Stable Unit Treatment Value (SUTV) condition: Y = Y(a) when A = a.
- ► Full data: A and  $\mathcal{Y} = \{Y(a) : a \in A\}$
- ightharpoonup Observed data: A and Y(A)
- ightharpoonup Missing data: Y(a) for all a's different from A's value
- ▶ Missing indicator:  $R = \{R(a) : a \in \}$  where R(a) = 1 for a = A and 0, otherwise.
- ► Auxiliary information: (*X*, *Z*) denotes the covariate information (*X* denotes the covariates of interest for treatment effect modifiers)

## Missingness mechanism for causal inference

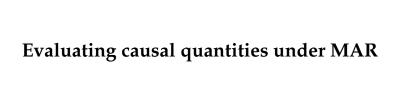
- ightharpoonup MCAR: A is completely independent of  $\mathcal{Y}$  and (X, Z)
- ▶ MAR: *A* is independent of  $\mathcal{Y}$  conditional on (X, Z)
- ▶ MNAR: *A* is not independent of  $\mathcal{Y}$  conditional on (X, Z)
- ► MAR is equivalent to no unobserved confounder (NUC) assumption in causal inference

### Causal assumptions

- ▶ Ignorability/no unobserved confounder condition (NUC): A is independent of  $\{Y(a) : a = 1, -1\}$  conditional on X and Z.
- ► Consistency/Stable Unit Treatment Value (SUTV) condition: Y = Y(a) when A = a.
- ► The first condition says that the treatment assignment is independent of all potential outcomes given *X* and *Z*.
- ► The second condition says that the observed outcome is the same as the potential outcome for the given treatment.

### Discussion of causal assumptions

- ► The first condition holds in randomization trials unless randomization is imperfect.
- ► The second condition is natural but may not hold when there is treatment interference or non-compliance.



## Why MAR or NUC is important?

- For simplicity, we only consider 2 treatment options  $A = \{-1, 1\}$ .
- ► It removes all potential confounders so the observed difference between two arms is purely due to their treatment assignments.
- ► It can provide an unbiased estimator for the average treatment effect

$$E[Y(1)] - E[Y(-1)].$$

► It also gives an unbiased estimator for the feature-specific treatment effect

$$E[Y(1)|X = x] - E[Y(-1)|X = x].$$

### **Justification**

► The first key relationship:

$$E[Y(a)|X = x, Z = z] = E[Y(a)|A = a, X = x, Z = z].$$

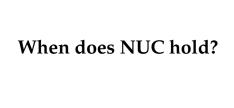
► The second key relationship:

$$Y(a) = Y$$
 when  $A = a$ .

► These two equations yield

$$E[Y(a)|X = x, Z = z] = E[Y|A = a, X = x, Z = z]$$

which can be estimated with bias from data.



#### Randomized Trials

- ► In randomized trials, patients are randomized into one of each treatment arm.
- ► The randomization probability can be different for each patient, i.e.,

$$\pi(a, x, z) \equiv P(A = a | X = x, Z = z)$$
 can be a function of  $x$  and  $z$ ,

where *X* denotes covariates of interest, *Z* contains all other auxiliary covariates, and *A* is treatment.

▶ Note that  $\pi(a, x, z)$  is known by the design.

### Observational Studies

- ► Most data are collected from observational studies (cross-sectional or longitudinal).
- Randomization of treatment options is no longer controlled so individuals choose or are given treatments in a naturalistic and unknown way.
- ▶ NUC assumption may not hold; however, if we can collect additional features/covaraites as much as possible, say *Z*, it may be more plausible to assume the following NUC condition:

#### *General NUC (GNUC) condition:*

A is independent of  $\{Y(a)\}$  conditional on both X and Z.

Note that P(A = a|X, Z) is unknown so needs to be estimated.

## Applying likelihood method for missing data under GNUC

► Likelihood approach under MAR:

$$f(X,Z) \prod_{a \in \mathcal{A}} f(Y(a)|X,Z)^{I(A=a)} f(A|X,Z)$$
$$= (X,Z) \prod_{a \in \mathcal{A}} f(Y|X,Z)^{I(A=a)} f(A|X,Z)$$

- ► Estimation procedure:
  - ▶ We estimate E[Y|X, Z, A = a] using either semiparametric regression models or nonparametrically.
  - $\blacktriangleright$  We estimated the conditional distribution of Z|X semiparametrically and nonparametrically.
  - ► We then calculate  $E[Y|X, A = a] = \int E[Y|X, Z, A = a]f(Z|X)dZ$ .
- ▶ This is the essential algorithm in G-computation.

## Applying semiparametric method (IPW) for missing data under GNUC

► This is based on the fact

$$E\left[\frac{YI(A=a)}{P(A=a|X,Z)}\Big|X\right] = E\left[\frac{Y(a)I(A=a)}{P(A=a|X,Z)}\Big|X\right]$$
$$= E\left[E[Y(a)\Big|X,Z]\frac{E[I(A=a)|X,Z]}{P(A=a|X,Z)}\Big|X\right]$$
$$= E\left[Y(a)\Big|X\right].$$

▶ We estimate E[Y(a)|X] using the inver probability weighted expectation of R among subjects with A = a.

### Interpretation of semiparametric method

► The observed data for A = a are sampled from the distribution

$$f_{obs} \equiv f(X, Z) \prod_{a} \{ P(A = a | X, Z) f(Y(a) | X, Z) \}^{I(A=a)}$$
.

▶ To estimate the distribution of Y(a), the data we wish to obtain should be sampled from the distribution

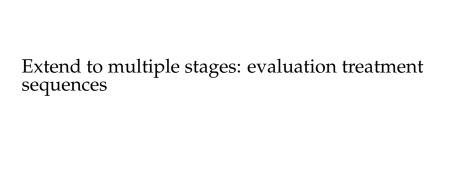
$$f_{wish} \equiv f(X,Z)f(Y(a)|X,Z)$$
 A is always held at a.

- ► We can estimate  $E_{wish}[Y(a)|X]$  from  $f_{wish}$  but data are from  $f_{obs}$ .
- ► By important sampling,

$$E_{wish}[Y(a)|X] = E_{obs} \left[ Y(a) \frac{f_{wish}}{f_{obs}} |X \right] = E \left[ \frac{Y(a)I(A=a)}{P(A=a|X,Z)} |X \right]$$
$$= E \left[ \frac{YI(A=a)}{P(A=a|X,Z)} |X \right].$$

## Positivity assumption implied in these approaches

- ▶ One implicit assumption is that the probability measure  $f_{wish}$  is dominated by the probability measure  $f_{obs}$ , i.e., P(A = a|X,Z) > 0 almost surely.
- ▶ Note that this assumption holds in randomization trials if randomization probabilities are positive.
- ► AIPW can be further constructed to achieve double robustness and local efficiency using semiparametric efficiency theory.



### Constant Treatment Sequences

- Now we consider treatment sequences, say  $(a_1, a_2, ...)$ , and wish to estimate  $E[Y(a_1, a_2, ...)]$ .
- ▶ We focus on two stages, so the sequence is  $(a_1, a_2)$ .
- ► Individual data obtained from either trials or observational studies are

$$X_1, A_1, X_2, A_2, Y$$
.

▶ Sequentially,  $A_1$  may depend on  $X_1$  while  $A_2$  may depend on anything before that.

### View from Missing Data Problem

- ► Full data consist of  $\{Y(a_1, a_2) : a_1 \in A_1, a_2 \in A_2\}$ .
- ▶ Observed data:  $X_1, A_1, X_2, A_2, Y$  and under SUTV assumption, only  $Y(a_1, a_2) = Y$  is observable when  $A_1 = a_1, A_2 = a_2$ .
- ▶ Missing data include all  $Y(a'_1, a'_2)$  for  $A_1 \neq a'_1$  or  $A_2 \neq a'_2$ .

### MAR vs Sequential Ignorability Conditions

► Similar to NUC/MAR condition, we need conditions

 $A_1$  is independent of the potential outcomes given  $X_1$ 

 $A_2$  is independent of the potential outcomes given  $(X_1, A_1, X_2)$ .

- ▶ These conditions hold if  $A_1$  is randomized with randomization probability dependent on  $X_1$  and  $A_2$  is randomized with randomization probability dependent on  $(X_1, A_1, X_2)$ —sequential randomization.
- ▶ When the study is observational,  $X_1$  and  $X_2$  should contain all possible confounders to make these conditions hold.

### Unbiased Estimator for $E[Y(a_1, a_2)]$ under MAR

▶ Using IPWE or important sampling, we consider

$$E\left[\frac{YI(A_1=a_1,A_2=a_2)}{P(A_2=a_2|X_1,A_1,X_2)P(A_1=a_1|X_1)}\right].$$

▶ By SUTV condition, it is equivalent to

$$E\left[\frac{Y(a_1,a_2)I(A_1=a_1)}{P(A_1=a_1|X_1)}\frac{I(A_2=a_2)}{P(A_2=a_2|X_1,A_1,X_2)}\right].$$

► From the second NCU condition, it becomes

$$E\left[\frac{Y(a_1,a_2)I(A_1=a_1)}{P(A_1=a_1|X_1)}\right].$$

- ▶ From the first NCU condition, it reduces to  $E[Y(a_1, a_2)]$ .
- ► Clearly, we require  $P(A_1 = a_1|X_1)$  and  $P(A_2 = a_2|A_1 = a_1, X_1, X_2)$  to be positive.

### Other missing data methods

- ► G-computation (likelihood method)
- Marginal structural equation model based on inverse probability weighted estimating equations (semiparametric method)
- Structural nested models based on sequentially modelling blip functions (comparing potential outcomes at each stage)
- ► Double robust estimators (AIPW)



# What Can We Learn from Constant Treatment Strategies?

- ► Sequential ignorability conditions (sequential MAR) are the key conditions; thus, sequential randomization design is important. These conditions guarantees the removal of bias due to unobserved confounders when comparing different strategies.
- ► Positivity assumption ensures that data contain the observations for treatment strategies of interest.
- ► Important sampling is useful to infer the desired situation under treatment strategies of interest from data under a different probability sampling distribution.