Lecture 10: IPTW for Marginal Structural Models, Inference for IPTW

A roadmap for causal inference

1. Specify **Causal Model** representing real background knowledge
2. Specify **Causal Question**
3. Specify **Observed Data** and link to causal model
4. **Identify** : Knowledge + data sufficient?
5. Commit to an **estimand** as close to question as possible, and a **statistical model** representing real knowledge.
6. **Estimate**
7. **Interpret** Results

Outline

1. IPTW for Working MSM
2. IPTW for MSM conditional on baseline covariates
3. Inference for IPTW

References

- Hernan and Robins. “Estimating causal effects from epidemiological data” *J. Epidem and Community Health*, 60(7): 578-586, 2006
- Robins and Hernan. “Estimation of the causal effects of time-varying exposures” In Fitzmaurice, Davidian, Verbeke, Molenberghs, editors, Longitudinal Data Analysis, chapter 23. Chapman & Hall/CRC Press, Boca Raton, FL, 2009
- Robins, Hernan, Brumback. “Marginal Structural Models and Causal Inference in Epidemiology”. *Epidemiology,* 11(5):550-560, 2000
  - Longitudinal MSM
- Petersen, et al. “Diagnosing and responding to violations in the positivity assumption” *Statistical Methods in Medical Research*, 21(1):31-54, 2012
- Neugebauer and van der Laan. “Nonparametric causal effects based on marginal structural models” *Journal of Statistical Planning and Inference*, 137(2): 419-434, 2007.

Recap: Marginal Structural Models

- Another way to define your target parameter…
- Provides a summary measure of how the counterfactual outcome changes as a function of treatment
  - and possibly pre-treatment covariates that are effect modifiers of interest
- Useful when A (or (A,V)) has many possible levels…

Defining our target parameter using a marginal structural model

- Example: we are interested in how counterfactual mean blood pressure (Y) varies as a function of drug dose (A)
  - Drug can be given at one of five doses
  - $\mathcal{A}$={0,1,2,3,4}
- Defining our target causal parameter
- Option 1: Estimate all the pairwise comparisons
  - (or those of interest)
  - Ex. $E(Y_2-Y_1)$; $E(Y_4-Y_0)$; etc…

Defining our target parameter using a marginal structural model

- Option 2: Summarize how counterfactual outcome (blood pressure) varies as a function of dose using a marginal structural model
- Ex. $E(Y_a)=m(a|\beta)= \beta_0 +\beta_1 a$
  - Assumes a linear change in expected counterfactual blood pressure with increasing dose
    - Can modify to be a "working" MSM- projection of true causal curve onto line…

"Marginal Structural" Model

- "Marginal": model on the marginal distribution (typically the expectation) of the counterfactual outcome Y_a
  - Vs. a model on the joint distribution of the counterfactual outcomes Y_a, a in $\mathcal{A}$
  - Note: a <u>marginal</u> structural model can still be <u>conditional</u> on baseline covariates
  - Ex: $E(Y_a|V) = \beta_0 + \beta_1 a + \beta_2 V + \beta_3 aV$
- "Structural": parameters of the model (eg. coefficients β) have causal as opposed to statistical interpretation

IPTW estimator for a linear marginal structural model

- IPTW estimator of β is solution in β to the following estimating equation:

$$0 = \frac{1}{n}\sum_{i=1}^{n} \frac{g^*(A_i)}{g_n(A_i|W_i)} \frac{d}{d\beta} m(A_i|\beta)(Y_i - m(A_i|\beta)$$

  - g*(A) is any non-null function of A
    - More on this to come..

Some intuition about the IPTW estimating equation

- Ex: Say have a linear regression model:

$E(Y|A)=m(A|\beta)=\beta_0+\beta_1 A$

- One option to estimate β: minimize the sum of squared errors

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} (Y_i - m(A_i|\beta))^2$$

- Take the derivative with respect to β, set equal to zero, and solve for β

$$0 = \frac{1}{n} \sum_{i=1}^{n} \frac{d}{d\beta} m(A_i|\beta)(Y_i - m(A_i|\beta))$$

Some intuition about the IPTW estimating equation

- Standard least squares estimator for β corresponds to solving

$$0 = \frac{1}{n}\sum_{i=1}^{n} \frac{d}{d\beta} m(A_i|\beta)(Y_i - m(A_i|\beta))$$

- IPTW estimator (for particular g*(A)) solves weighted version of this estimating equation
  - Intuition: Weights used to correct for confounding

$$0 = \frac{1}{n}\sum_{i=1}^{n} \frac{g^*(A_i)}{g_n(A_i|W_i)} \frac{d}{d\beta} m(A_i|\beta)(Y_i - m(A_i|\beta)$$

IPTW Estimator of a Marginal Structural Model Parameter

1. Estimate probability of treatment given covariates: $g_0(A|W)$
   - One option when A has multiple levels- multinomial logistic regression
   - Other (data-adaptive) approaches also possible…
2. Each subject i gets a weight: $g^*(A_i)/g_n(A_i|W_i)$
   - We can choose g*(A)- more to come

IPTW Estimator of a Marginal Structural Model Parameter

3. In the reweighted population, treatment is no longer associated with covariates W
   - If W is sufficient to control for confounding (ie satisfies back door criterion), in re-weighted population can estimate effect simply by fitting a regression of the observed exposure on the observed outcome
4. Regress Y on A according to MSM, using estimated weights
   - Can use standard software, and provide the weights

Choice of numerator for the weights

- When we use IPTW to estimate the parameters of a **correctly specified** marginal structural model

– Under the (strong) positivity assumption, choice of numerator for the weights (i.e. g*(A)) does not affect consistency of the estimator

– However, clever choice of numerator

1. Can allow us to weaken the positivity assumption
2. Can improve efficiency
   - Particular choices of g*(A) can result in lower variance estimators (for MSM not saturated in A)

Choosing a numerator for the weights in an MSM

- Let's start by assuming that our MSM is correctly specified

  Ex: $E(Y_a) = \beta_0 + \beta_1 a$ for some β
- "Stabilized weights:" Choose g*(A) equal to the marginal probability of A
  - Estimate g*(a) as proportion of people with A=a

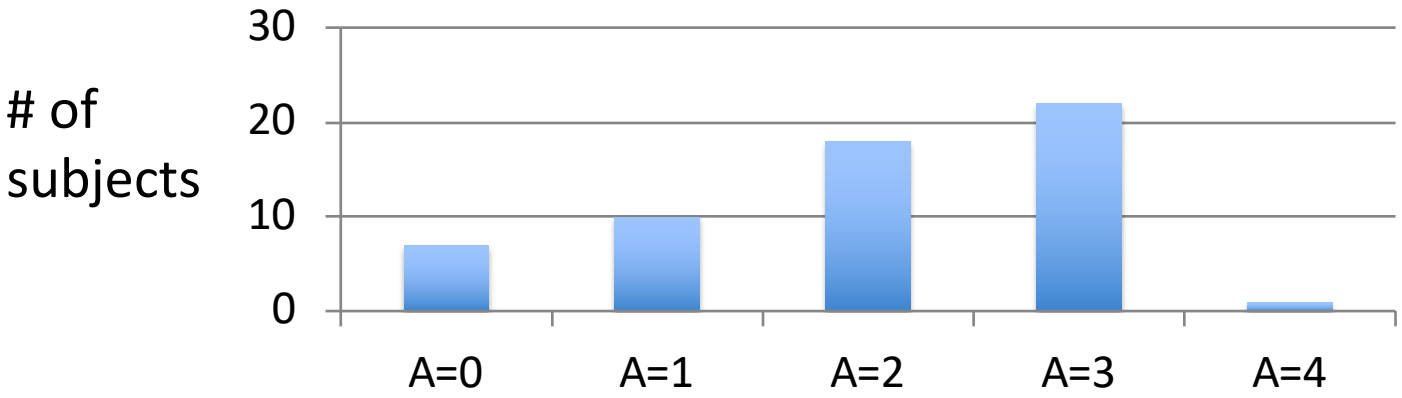
$$st.wt = \frac{g_n^*(A)}{g_n(A|W)}$$

Why is this choice of g*(A) commonly recommended?

1. Reduce variability in the weights
   - If some levels of A are rare, the few individuals who have those values will have small predicted probability of having their observed treatment
     - $g_n(A|W)$ will be small
   - Unstabilized weights for these individuals will be large
     - Drive up variability of estimator
   - Stabilization can help avoid these extreme weights

Running example

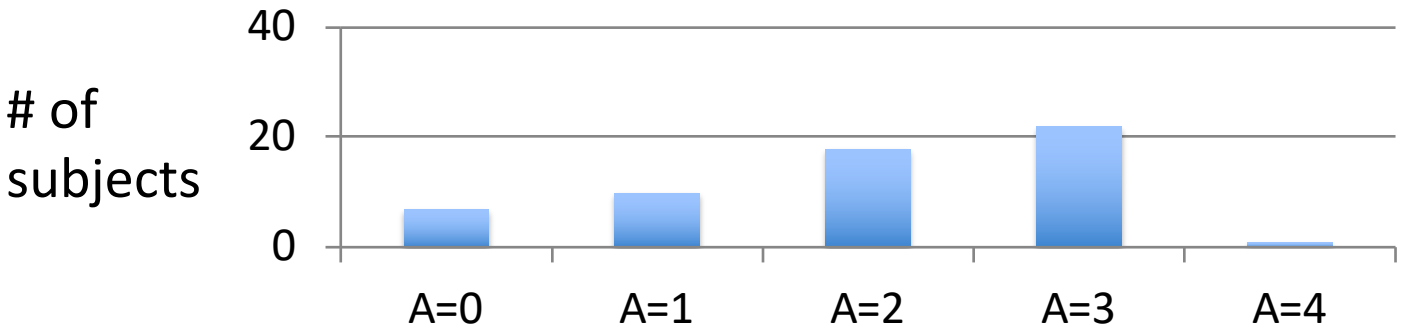
- Say one person in our sample gets a high dose



- MSM: $E(Y_a) = \beta_0 + \beta_1 a$
- Unstabilized weights: $1/g_n(A_i|W_i)$
  - Denominator small (and thus weight large) for subject who gets highest dose

Running Example

- Say one person in our sample gets a high dose



- MSM: $E(Y_a) = \beta_0 + \beta_1 a$
  - Stabilized weights: $g_n^*(A_i)/g_n(A_i|W_i)$
  - Denominator for subject who gets highest dose will be small
  - Numerator also small for this subject
  - Less extreme weight

Why is this choice of g*(A) commonly recommended?

2. Weaker positivity assumption
   - Recall: Unlike the substitution estimators, the IPTW estimator of $E(Y_a)$ cannot extrapolate to areas of the data with no support
   - However, the IPTW estimator of an MSM parameter can extrapolate
     - It extrapolates using the MSM itself

Weakening the positivity assumption

- The positivity assumption we need for identifiability depends on our choice of g*(A)
- Positivity assumption for parameter defined using MSM: m(a|β)
  - **For any a in $\mathcal{A}$ for which g*(a)≠0**, we need:

    $P_0(A=a|W=w)> 0$ for all w for which $P_0(W=w)>0$
  - If we choose $g^*(a)=P_0(A=a)$, only values of exposure that occur with non zero probability must occur with non-zero probability in all strata of W

Positivity assumption for MSM (A bit more formally):

- The positivity assumption we need for identifiability depends on our choice of g*(A)
- Positivity assumption for parameter defined using MSM: m(a|β) (define 0/0=0)

$$\sup_{a \in \mathcal{A}} \frac{g^*(a)}{g_0(a|w)} < \infty \quad \text{for all } w \text{ for which } P_0(W = w) > 0$$

If this is 0

Then this can be 0

- If we choose g*(a)=P(A=a), only values of exposure that occur with non zero probability must occur with non-zero probability in all strata of W

Why is this helpful?

- It means that our target causal quantity is still defined if some levels of the exposure of interest occur with zero probability
- Ex: $E(Y_a)= \beta_0 + \beta_1 a$
  - MSM here is assuming a linear relationship between dose and expected blood pressure
  - If we believe this model, it doesn't matter if there are some values of the exposure with zero probability of occurring (or that don't occur in our sample)
- The MSM itself allows us to extrapolate

Running Example

- Now say that <u>no one</u> in our target population gets the highest dose of the drug
  - $P_0(A=4)=0$
- The strong positivity assumption will fail

$$\min_{a \in \mathcal{A}} P_0(A = a|W = w) > 0,$$
$$\text{for all } w \text{ for which } P_0(W = w) > 0$$

- The Standard IPTW estimator will fail for a=4

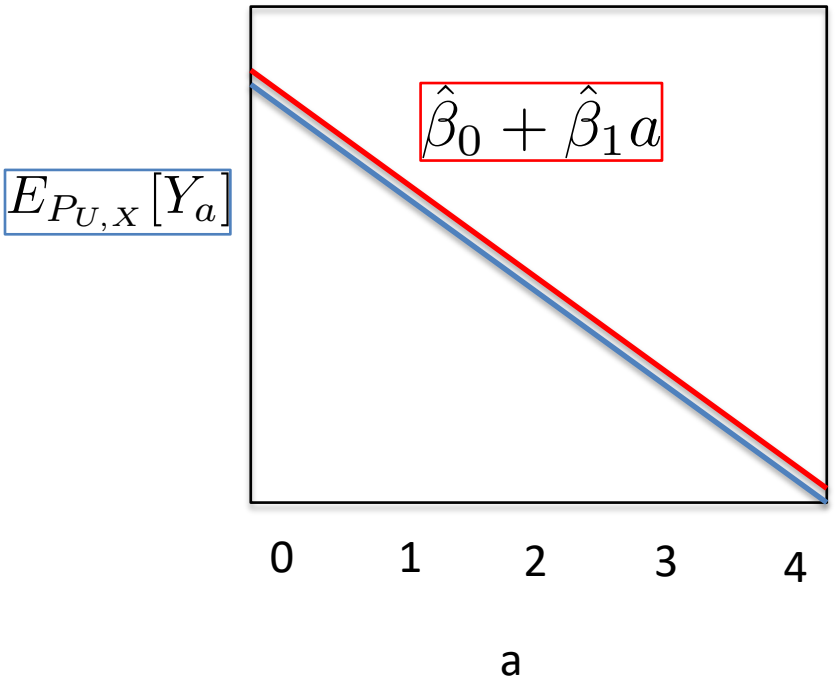
$$\hat{\Psi}_{IPTW}(P_n) = \frac{1}{n} \sum_{i=1}^{n} \frac{I(A_i = a)}{g_n(A_i|W_i)} Y_i$$

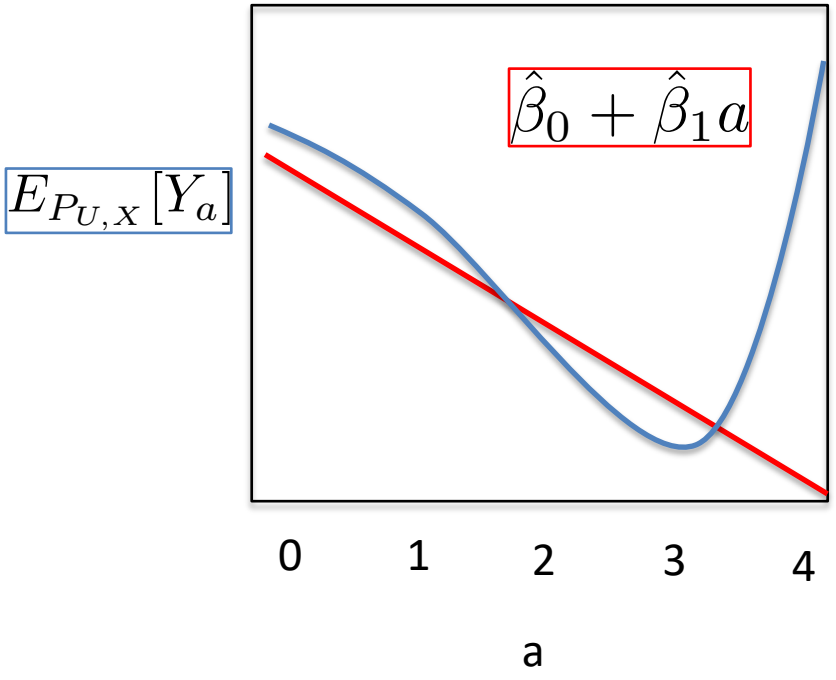
Running example

- We assume MSM: $E(Y_a) = \beta_0 + \beta_1 a$
- Now we need a weaker positivity assumption
  - Positive probability of getting doses 0-3 regardless of covariate values
- If this holds, IPTW estimator of β still defined
- Resulting estimate of β allows us to extrapolate in order to generate estimates of the expected counterfactual outcome at the highest dose

MSM allows us to extrapolate

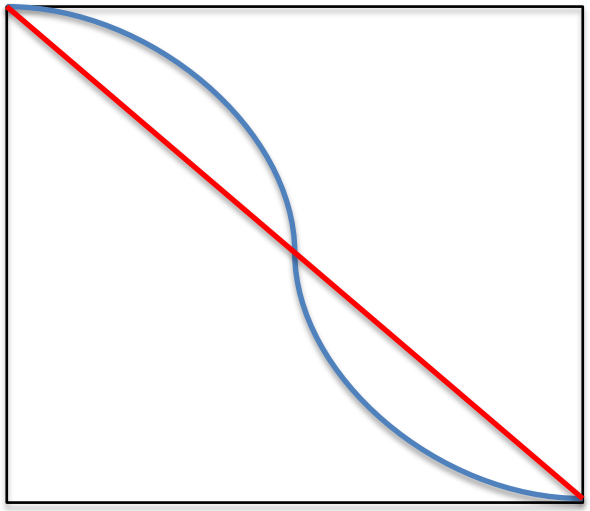
- Fine if the MSM is correctly specified
- Dangerous if it isn't





Recall: Marginal Structural Working Models

- We usually don't know enough to confidently specify a parametric model for our dose response curve
- Nonetheless, we may be willing to settle for some summary measure of the true dose-response curve….



Define target parameter as a projection of the true causal curve onto a working model

$$m(a|\beta) = \beta_0 + \beta_1 a$$

$$\beta(P_{U,X}|m) \equiv \arg\min_{\beta} E_{P_{U,X}} \left[\sum_{a \in \mathcal{A}} (Y_a - m(a|\beta))^2 \right]$$

Recap: IPTW to estimate the parameters of a *correctly specified* MSM

1. Choice of numerator g*(A) for the weights does not affect consistency of the estimator (under strong positivity assumption)
2. Choice of numerator g*(A) may affect efficiency

- "Stabilized weights:" Choose g*(A) equal to the marginal probability of A
  - Estimate g*(a) as proportion of people with A=a

$$st.wt = \frac{g_n^*(A)}{g_n(A|W)}$$

IPTW estimator for a working MSM

- If we assume our MSM is correctly specified, choice of numerator only affects efficiency
- If we define our target parameter using a working MSM, choice of numerator changes the target parameter…

$$\beta(P_{U,X}, m, g^*) \equiv \arg\min_{\beta} E_{P_{U,X}} \left[\sum_{a \in \mathcal{A}} (Y_a - m(a|\beta))^2 g^*(a) \right]$$

  - g*(a) how much weight we put on specific values of the exposure

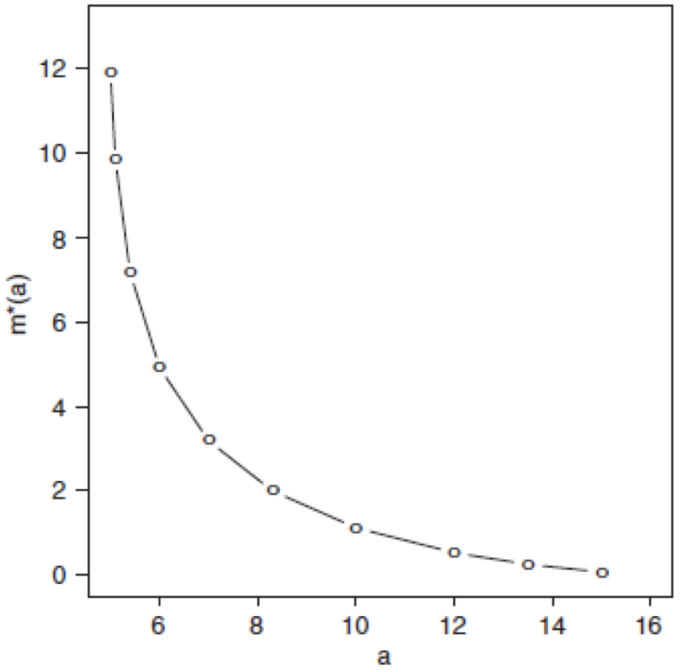
IPTW estimator for a working MSM

- Choice of numerator changes the target parameter
  - Option1: Unstabilized weights
    - Numerator=1 : puts equal weight on all possible values a in $\mathcal{A}$
  - Option 2: Stabilized weights
    - Numerator=empirical proportion with a subject's observed A: puts less weight on values of A with few (or no) observations
    - Puts zero weight on values (a) such that g(A=a)=0
    - Relies entirely on extrapolation for these values

Simulation Example

- Neugebauer R, et al. Nonparametric causal effects based on marginal structural models *JSPI* (2007)

- True Causal Curve



$$E(Y_a) = 4 - 3\log(a - 4.9) + 0.2a$$

Simulation Example

- True causal curve

$$E(Y_a) = 4 - 3\log(a - 4.9) + 0.2a$$

- Working MSM $m(a|\beta) = \beta_0 + \beta_1 a$
- Target Parameter

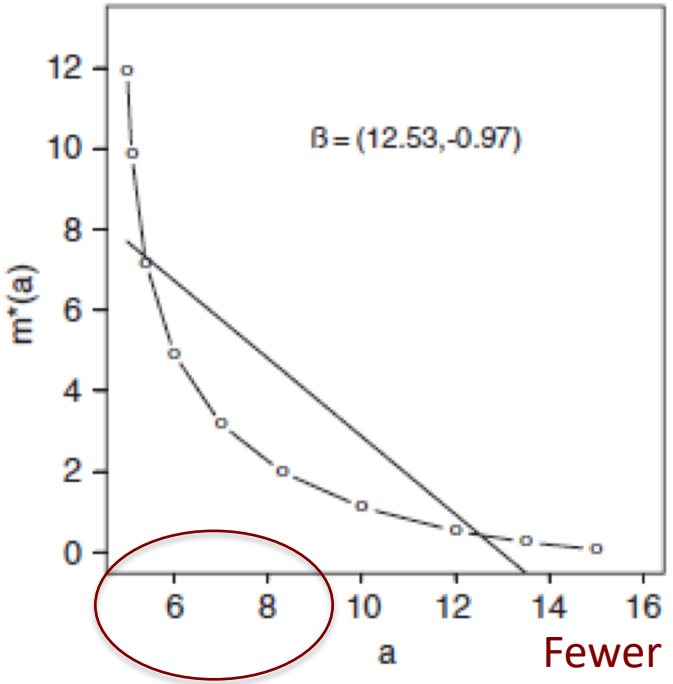
$$\beta(P_{U,X}, m, g^*) \equiv \arg\min_{\beta} E_{P_{U,X}} \left[\sum_{a \in \mathcal{A}} (Y_a - m(a|\beta))^2 g^*(a) \right]$$

- Option 1: g*(a)=1
- Option 2: g*(a)=P(A=a)

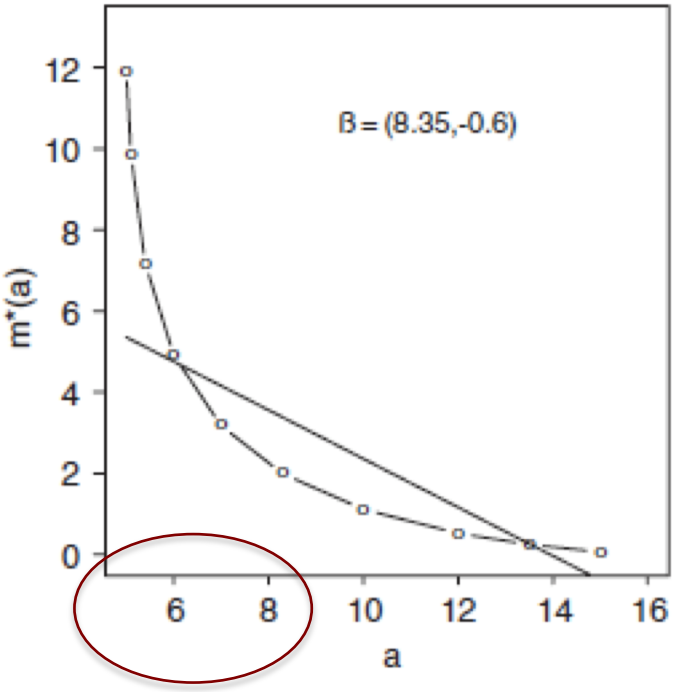
Neugebauer and van der Laan 2007

Simulation Example

- Option 1: g*(a)=1
- Option 2: g*(a)=P(A=a)



Fewer Observations



Marginal treatment probabilities in the simulation with point treatment data

a	5	5.1	5.4	6	7	8.3	10	12	13.5	15
$g(a)$	0.01	0.01	0.035	0.035	0.04	0.065	0.085	0.2	0.26	0.26

Neugebauer and van der Laan 2007

Marginal structural models used to define modification of causal effects by baseline covariates

- Example: Interested in a summary of how counterfactual blood pressure varies as a function of dose and baseline risk (BR)
  - Hypothesis: the dose response curve will differ by V=BR
- A possible MSM: (BR=V)

$$E(Y_a|V)= \beta_0 +\beta_1 a +\beta_2 V +\beta_3 a\times V$$

How would we implement the IPTW estimator of this parameter?

- As before,

1. Estimate $g_0(A|W)$
2. Generate predicted probability that each subject received his/her observed treatment given covariates
3. Use these predicted probabilities to estimate the weights
4. Regress Y on A and BR according to MSM, using weights

How does estimation of causal effects conditional on baseline effect modifiers change the IPTW estimator?

- We have more choices for the numerator of the weights
- For non-conditional MSMs such as

  $E(Y_a)= \beta_0 +\beta_1 a$,

  the numerator can be any non-null function of A
- Example: P(A=a) for subjects for whom (A=a)
  - Ie numerator is each subject's estimated probability of getting her observed treatment
    - Ex: the proportion of the sample that got the same treatment as that subject

Choosing a numerator for the weights in a conditional MSM

- For conditional MSMs such as

  $E(Y_a|V)= \beta_0 +\beta_1 a +\beta_2 V +\beta_3 a\times V$

  the <u>numerator can be any non-null function of (A,V)</u>
- IPTW estimator solves the following estimating equation

$$0 = \frac{1}{n}\sum_{i=1}^{n} \frac{g^*(A_i, V_i)}{g_n(A_i|W_i)} \frac{d}{d\beta} m(A_i, V_i|\beta)(Y_i - m(A_i, V_i|\beta)$$

Stabilized weights in a conditional MSM

- Choose $g^*(A|V) = P_0(A|V)$
  - Often can estimate as the empirical proportion
- Numerator is each subject's probability of getting his/her observed treatment, given his/her effect modifiers of interest

$$\hat{st.wt}_i = \frac{g^*(A_i|V_i)}{g_n(A_i|W_i)}$$

- Why? Same arguments apply
  - Weaker positivity assumption
  - Reduce variability in the weights

Weakening the positivity assumption

- Stabilized weights: Choose $g^*(a,V)=P_0(A=a|V)$
- Weaker positivity assumption:
  - For any a in $\mathcal{A}$ for which $P_0(A=a|V)>0$, we need:

    $P_0(A=a|W=w)> 0$ (for all w for which $P_0(W=w)>0$)
- Now, it is OK if certain exposure levels are not represented in certain strata of V
  - For these (a,v) combinations, we rely on the MSM to extrapolate
  - $E(Y_a|V)= \beta_0 +\beta_1 a +\beta_2 V +\beta_3 a\times V$

Positivity assumption for MSM conditional on V

- The positivity assumption we need for identifiability depends on our choice of g*(A,V)
- Positivity assumption for parameter defined using MSM: m(a,V|β)

$$\sup_{a \in \mathcal{A}} \frac{g^*(a,V)}{g_0(a|W)} < \infty,$$

$$\text{for all } w \text{ for which } P_0(W = w) > 0$$

If this is 0

Then this can be 0

- In other words, for any (a,V) combinations for which g*(a,V)>0, we need $g_0(a|W)>0$

Stabilized Weights

- Why can stabilized weights improve efficiency?
- Intuition: Resulting estimator only uses weights to control for confounding beyond baseline covariates V

$$\hat{stwt} = \frac{g^*(A_i|V_i)}{g_n(A_i|W_i)}$$

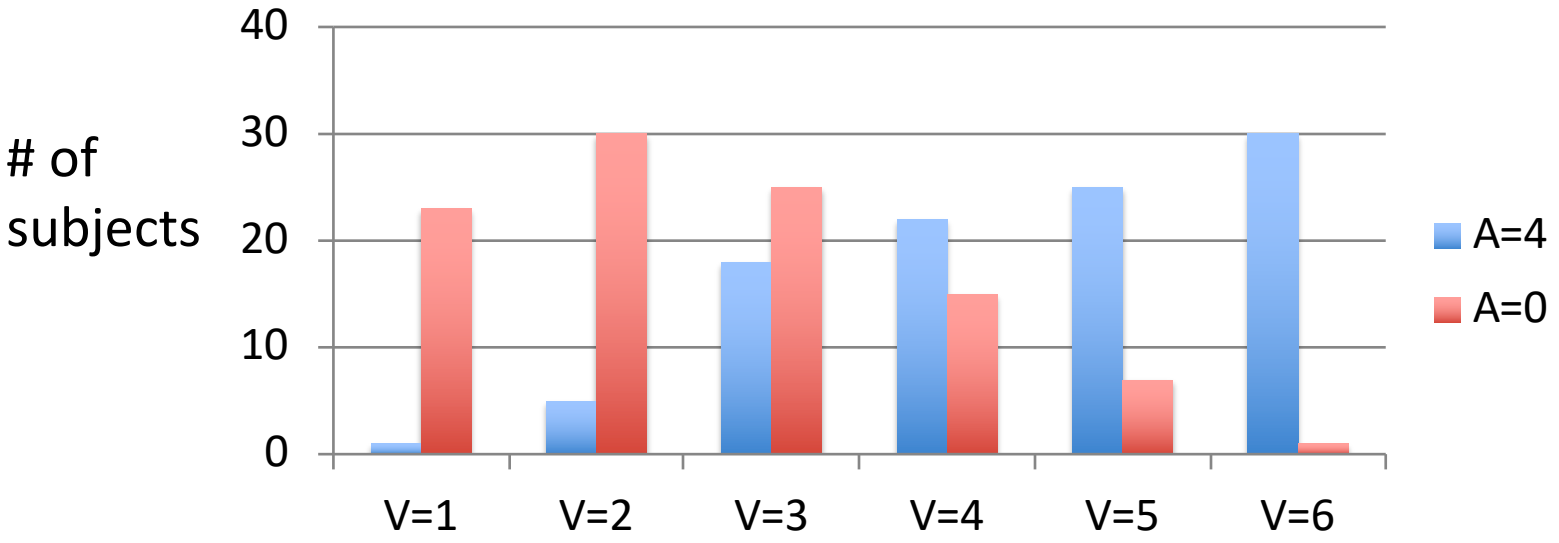
- If covariates in W no longer predict A after controlling for V, then the weights will be 1
  - Avoid adding additional variability with the weights

Stabilized weights

- Stabilized weights can still add efficiency in less extreme cases
  - i.e. even when the estimated weights are not one, stabilizing them can make them (and thus the IPTW estimator) less variable
- Subjects get up-weighted based on their probability of having their observed treatment given observed covariates relative to their probability of observed treatment given V

Ex: Baseline risk and drug dose

- Again: 5 drug doses (0,1,2,3,4)
- Baseline risk is effect modifier of interest (V)
  - Strongly associated with dose
- In lowest risk group only one person gets the highest dose
  - P(A=4|V=1)= small
- In highest risk group only one person gets the lowest dose
  - P(A=0|V=6)=small



Example: Advantage of Stabilized Weights

- Unstabilized weights:
  - $1/g_n(A_i|W_i)$
  - Denominator small for
    - Subject in the lowest risk group who gets highest dose
    - Subject in the highest risk group who gets lowest dose
  - These subjects thus get big weights

Example: Advantage of Stabilized Weights

- Stabilized weights
  - $g_n^*(A_i|V_i)/g_n(A_i|W_i)$
  - Denominator small for
    - Subject in the lowest risk group who gets highest dose
    - Subject in the highest risk group who gets lowest dose
  - However… numerator is also small for these subjects
  - Their weights will only differ from 1 to the extent that additional covariates affect the probability of treatment
    - Less extreme weights

Estimating the Standard Error of the IPTW estimator for an MSM

- Can use standard software
  - Regress Y on A according to MSM, specifying weights, specify robust standard error estimate
  - Ex. geeglm in R
- This provides standard error estimate for the estimates of β (the coefficients)
  - This is an influence curve-based approach
  - You could also code it yourself! (will do for TMLE)

Estimating the Standard Error of the IPTW estimator for an MSM

- This approach treats the weights as fixed when in fact they were estimated
- If the weights were estimated using MLE and a correctly specified parametric model- resulting standard error estimates conservative
    - Too big
- If the weights were estimated using machine learning (eg Super-Learner)- no such guarantee

Alternative: Bootstrap

- If g was estimated using MLE of a correctly specified parametric model, NP-Boot can provide smaller CIs
  - More precise estimates
- If g was estimated with machine learning…
  - Make sure to rerun entire ML algorithm in each bootstrap sample
  - Still no theory guaranteeing that estimator meets conditions needed for non-parametric bootstrap to work

Key points: IPTW

- Relies on consistent estimation of $g_0(A|W)$
- Data-adaptive algorithms often called for
  - Challenge: Not targeting the estimand

1. Be smart in your covariate selection
   - Want to avoid covariates that have strong effect on exposure and no or minimal effect on outcome
     - Dimension reduction based on marginal association with Y
2. NP-boot probably better for inference
   - Look at the bootstrap distribution of your estimator
   - Use quantiles to get 95% CI

Brief intro to estimating equations

- An Estimating Function $D(O|\psi)$ is a function of the observed data and the (unknown) parameter of interest
  - Observe n i.i.d. copies of O_i, i=1,…n; $O \sim P_0$
  - Parameter of interest $\Psi(P_0)=\psi$
  - Define $D(O|\psi)$ such that $E_0[D(O|\psi)]=0$
- Estimating Equation: $$0 = \frac{1}{n}\sum_{i=1}^{n} D(O_i|\psi)$$
- Estimator: ψ_n defined as the solution satisfying $$\frac{1}{n}\sum_{i=1}^{n} D(O_i|\psi_n) = 0$$

Simple example: Population mean

- Observe n i.i.d. copies of $O_i=Y_i$; $O \sim P_0$
- Parameter of interest $\Psi(P_0)= \psi = E_0(Y)$
- Let $D(O| \psi)=Y- \psi$
  - Note: $E[D(O| \psi)]=E(Y)- \psi=0$
- Estimating Equation: $0 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \psi)$
- Estimator $\psi_n = \frac{1}{n}\sum_{i=1}^{n}(Y_i)$
  - Sample mean as estimator of population mean can be understood as root of an estimating equation

IPTW estimator defined as solution to an estimating equation

- Observe n i.i.d. copies of $O_i=(W_i\, A_i\, Y_i)$; $O \sim P_0$
- Parameter of interest: $\Psi(P_0) = E_0\left(\frac{I(A=a)}{g_0(A|W)}Y\right)$
- Estimating function: $D_{IPTW}(O|g,\psi) = \frac{I(A=a)}{g(A|W)}Y - \psi$
  - Note: if treatment mechanism *g* is not known, then it is a "nuisance parameter" which must be estimated
- Estimating Equation: $0 = \frac{1}{n}\sum_{i=1}^{n}\frac{I(A_i=a)}{g_n(A_i|W_i)}Y_i - \psi$

(Non-stabilized) IPTW estimator: $\psi_n = \frac{1}{n}\sum_{I=1}^{n}\frac{I(A_i=a)}{g_n(A_i|W_i)}Y_i$

Influence Curves and Estimating Functions

- Recall: An estimator is asymptotically linear with influence curve $IC(O_i)$ if it satisfies

$$\psi_n - \psi = \frac{1}{n}\sum_{i=1}^{n} IC(O_i) + o_{P_0}\left(\frac{1}{\sqrt{n}}\right)$$

$E_0(IC(O))=0$
$Var(IC(O))$ Finite

Converges to 0 in probability as n->∞, even when multiplied by √n

- Because $E_0(IC(O))=0$, **if we know the IC of an estimator, then we can use it as an estimating function**

Example: IC of IPTW

- Assume g_0 is known, and strong positivity
- IC of the IPTW estimator = $D_{IPTW}(O|g_0, \psi)$
  - $E_0 D_{IPTW}(O|g_0, \psi)=0$
  - $D_{IPTW}(O|g_0, \psi)$ has finite variance

$$\psi_n - \psi = \frac{1}{n}\sum_{i=1}^{n} D(O_i|g_0, \psi)$$

Exact equality- no remainder term

$$= \frac{1}{n}\sum_{i=1}^{n} \frac{I(A_i = a)}{g_0(A_i|W_i)} Y_i - \psi$$

  - For known g_0, variance of the IPTW estimator well-approximated by $var[D(O_i|g_0, \psi_n)]/n$

Example: IC of IPTW

- For known g_0, variance of the IPTW estimator well-approximated by $\text{var}[D(O_i|g_0, \psi_n)]/n$
- If g_0 is estimated (using a correctly specified parametric model), using an IC that treats g_0 as known gives conservative variance estimates (too big)
  - IC of IPTW estimator with g_0 estimated=

    IC of IPTW estimator with g_0 known – *projection*

  -> Estimating g_0 reduces the variance of the IC and thus of the IPTW estimator

Influence Curves vs. The Efficient Influence Curve

- For a given a statistical estimation problem:
  - n i.i.d. copies of O_i, $O \sim P_0 \in \mathcal{M}$
  - Target parameter $\Psi(P_0)=\psi$

1. Influence curves (or influence functions) are *estimator- specific*
   - Each asymptotically linear estimator has an influence curve
   - Teaches us about the asymptotic variance of that estimator
2. The efficient influence curve are *parameter-specific*
   - Teaches us about the asymptotic variance of the **most efficient** estimator
   - Lowest asymptotic variance (among "reasonable" competitors)- "generalized Cramer Rao lower bound"

The Efficient Influence Curve

- Can be derived for a given statistical estimation problem
  - A tiny bit here- beyond score of class
- An estimator is efficient if and only if it has IC equal to the eff IC
- For our running example:

- Suggests a new (optimal) estimating estimator defined as solution to corresponding estimating equation (A-IPW)
  - Two nuisance params, double robustness property
  - Other nice thing- an incorporate data adaptive estimation/ML for nuisance parameters…(see kennedy)

TMLE..

- Like AIPW:
  - DR
  - Solves estimating equation corresponding to the eff IC
    - Efficient if g and Q consistent
    - Can incorporate ML for estimation
- But also… a substitution estimator…
  - Review what this means
  - Why is this nice…

Recall can rewrite IPTW estimand

- Can write

$$E\left[\frac{I(A=a)}{g(A|W)}Y\right] = E\left[\frac{I(A=a)}{g(A|W)}Y\right] / E\left[\frac{I(A=a)}{g(A|W)}\right]$$

  - Note: $E\left[\frac{I(A=a)}{g(A|W)}\right] = E\left[E\left[\frac{I(A=a)}{g(A|W)}|W\right]\right] = 1$

- And

$$E\left[\frac{I(A=a)g^*(A)}{g(A|W)}Y\right] / E\left[\frac{I(A=a)g^*(A)}{g(A|W)}\right]$$

  - g*(A) any non-null function of A

Defining the target causal parameter using a linear working MSM

- Ex. $m(a|\beta) = \beta_0 + \beta_1 a$

$$\beta(P_{U,X}, m, g^*) \equiv \arg\min_{\beta} E_{P_{U,X}} \left[\sum_{a \in \mathcal{A}} (Y_a - m(a|\beta))^2 g^*(a) \right]$$

- Target parameter defined as solution (in β) to:

$$0 = E_{U,X} \left[\sum_{a \in \mathcal{A}} \frac{d}{d\beta} m(a|\beta)(Y_a - m(a|\beta)) g^*(A) \right]$$

  - This defines the target casual parameter (function on $P_{U,X}$)- not a statistical estimand or estimator (yet)

IPTW Estimator of Working MSM parameter

- Target parameter defined as solution (in β) to:

$$0 = E_{U,X}\left[\sum_{a\in\mathcal{A}} \frac{d}{d\beta} m(a|\beta)(Y_a - m(a|\beta))g^*(A)\right]$$

- IPTW estimator defined as solution (in β) to corresponding weighted (observed data) estimating equation:

$$0 = \frac{1}{n}\sum_{i=1}^{n} \frac{d}{d\beta} m(A_i|\beta)\frac{g*(A_i)}{g_n(A_i|W_i)}(Y_i - m(A_i|\beta))$$

  - Consistent and asymptotically linear if g_n is consistent for g_0 and positivity holds

Logistic working MSM with V

- Define target causal parameter with logistic working MSM:

$$\Psi^F(P_{U,X}) =$$

$$\arg\min_{\beta} -E_{U,X}\left[\sum_a g^*(a,V)\{Y_a \log m_\beta(a,V) + (1-Y_a)\log(1-m_\beta(a,V))\}\right]$$

- Target parameter defined as solution (in β) to:

$$0 = E_{U,X}\left[\sum_{a\in\mathcal{A}} g^*(a,V)\frac{\frac{d}{d\beta}m_\beta(a,V)}{m_\beta(1-m_\beta)}(E_{U,X}(Y_a|V) - m_\beta(a,V))\right]$$

Logistic working MSM with V

- Target parameter defined as solution (in β) to:

$$0 = E_{U,X}\left[\sum_{a\in\mathcal{A}} g^*(a,V)\frac{\frac{d}{d\beta}m_\beta(a,V)}{m_\beta(1-m_\beta)}(E_{U,X}(Y_a|V) - m_\beta(a,V))\right]$$

- IPTW estimator defined as solution (in β) to corresponding weighted (observed data) estimating equation:

$$0 = \frac{1}{n}\sum_{i=1}^{n}\frac{\frac{d}{d\beta}m_\beta(A_i,V_i)}{m_\beta(1-m_\beta)}\frac{g^*(A_i,V_i)}{g_n(A_i|W_i)}(Y_i - m_\beta(A_i,V_i))$$

  - Consistent and asymptotically linear if g_n is consistent for g_0 and positivity holds