Lecture 9:
Introduction to Propensity Score; The Inverse Probability of Treatment Weighted (IPTW) Estimator

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

Overview

1. Introduction to the Inverse Probability of Treatment Weighted (IPTW) Estimator
   - IPTW estimand and G-computation formula
   - Intuition
   - Implementation
2. Limitations of the IPTW Estimator
   1. How to estimate $g(A|W)$?
   2. Strong confounding and positivity violations
3. Alternative IPTW estimator
   - Can improve stability
4. IPTW to estimate the parameters of a marginal structural model
   - Stabilized weights

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

- Petersen, et al. “Diagnosing and responding to violations in the positivity assumption” *Statistical Methods in Medical Research*, 21(1):31-54, 2012

- Petersen, et al "Assessing the Effectiveness of Antiretroviral Adherence Interventions: Using Marginal Structural Models to Replicate the Findings of Randomized Controlled Trials." *JAIDS* 43.Supplement 1 (2006): S96-S103.

- TLB. Chapters 9 and 10

Distribution of the Observed Data Factorizes

- $O=(W,A,Y)$
- $P_0(O)=P_0(W,A,Y)$

  $=P_0(W)\ P_0(A|W)\ P_0(Y|A,W)$

- Different estimators require estimators of distinct factors of the observed data distribution

Different Estimators require estimators of distinct factors of the observed data distribution

- Estimand: $\Psi(P_0)=E_W(E_0(Y|A=1,W)-E_0(Y|A=0,W)$
  - Only a function of $E_0(Y|A,W)$ and $P_0(W)$

- $P_0(O)=P_0(W,A,Y)=P_0(W)P_0(A|W)\ P_0(Y|A,W)$

- Simple substitution estimators
  - Also referred to as "G-computation" estimators
  - Actually, rather than full $P_0(Y|A,W)$, only require estimators of $E_0(Y|A,W)$ (and $P_0(W)$)

Different Estimators require estimators of distinct factors of the observed data distribution

- Estimand: $\Psi(P_0)=E_W(E_0(Y|A=1,W)-E_0(Y|A=0,W)$
  - Only a function of $E_0(Y|A,W)$ and $P_0(W)$

- $P_0(O)=P_0(W,A,Y)=P_0(W)P_0(A|W)\ P_0(Y|A,W)$

- Propensity score-based estimators:
- Inverse Probability (of Treatment) Weighted
- PS matching
- PS as a dimension reduction….

Inverse Probability Weighting: A new estimand

- G-computation estimand:

  $\Psi(P_0)=E(E(Y|A=1,W)-E(Y|A=0,W)$

  - A function of E(Y|A,W) and distribution of W
- However, we can rewrite the same quantity as a function of the distribution of A given W
  - Also referred to as the "treatment mechanism"
  - Notation: g(A=a|W)=P(A=a|W)
- IPTW Estimand:

$$E\left(\frac{I(A=1)}{g(A=1|W)}Y\right)-E\left(\frac{I(A=0)}{g(A=0|W)}Y\right)$$

Equivalence of IPTW and G-computation estimands

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

- For $E_W(E(Y|A=a^*,W))$:

$$E\left[\frac{I(A=a^*)}{g(A|W)}Y\right] = \sum_{w,a,y} \frac{I(a=a*)}{g(a|W=w)} yP(Y=y,A=a,W=w)$$

$$= \sum_{w,y} \frac{1}{g(a*|W=w)} y$$

$$P(Y=y|A=a^*,W=w)P(A=a^*|W=w)P(W=w)$$

$$= \sum_w E(Y|A=a^*,W=w)P(W=w)$$

Standard IPTW Estimator

- IPTW and G-comp estimands are equivalent

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

  - See Extra slides
  - Both are equivalent under Randomization Assumption to $E(Y_a)$ ("treatment specific mean")
- IPTW estimand suggests a new estimator
- IPTW Estimator

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

Example: Effect of Pill Box Use on Adherence to Antiretrovirals

- Intervention: a weekly pill organizer
- Designed to help patients remember to take their prescribed medications

Research Question:
Does use of a pill box improve adherence to antiretroviral drugs?

Example: Effect of Pill Box Use on Adherence to Antiretrovirals

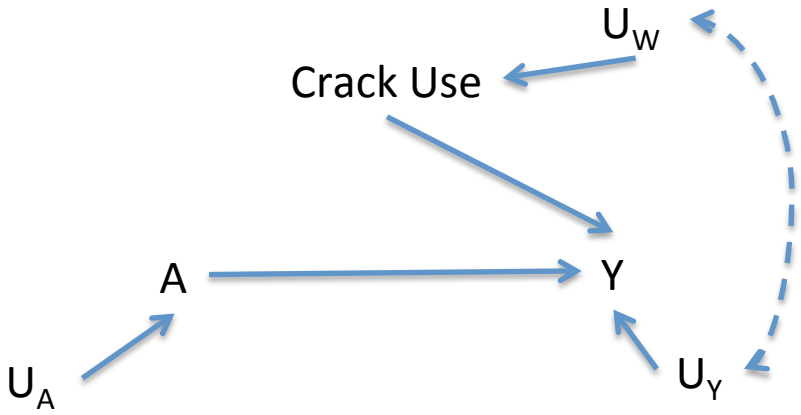
- A= Mediset Use
- Y= Adherence to antiretroviral drugs
  - % of prescribed doses taken
- W= age, sex, recreational drug use, past adherence, type of regimen, CD4 count….

Research Question:
Does use of Pill Box improve adherence to antiretroviral drugs?

Intuition: Inverse Probability of Treatment Weighted Estimator (IPTW)

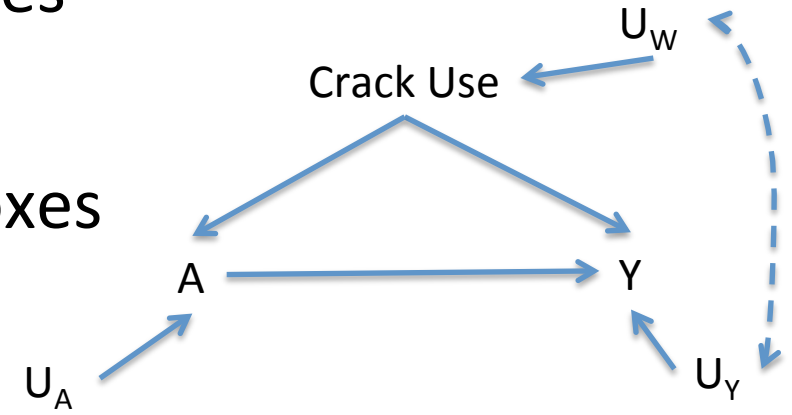
- Confounding can be viewed as a problem of biased sampling
  - In a randomized sample, probability of receiving treatment would be independent of covariates
  - For example, a crack user and non-crack user would have equal probabilities of using a pill box

$P(A=1|Crack=1)=P(A=1|Crack=0)=0.5$
$P(A=0|Crack=1)=P(A=0|Crack=0)=0.5$



Intuition: Inverse Probability of Treatment Weighted Estimator (IPTW)

- In observational data, the probability of receiving treatment is (generally) not independent of covariates
- For example, crack users may be more likely to use pill boxes than non-crack users
  - Crack users that use pill boxes over-represented
  - Crack users don't use pill boxes under represented



Intuition: Inverse Probability of Treatment Weighted Estimator (IPTW)

- Why is estimating g(A|W) helpful?
  - Tells us what types of people (which A,W combinations) are under-represented in our data, and by how much they are under-represented
- We can then up-weight subjects from strata of A,W that are under-represented (compared to the representation they would have had in a randomized trial),

Intuition: Inverse Probability of Treatment Weighted Estimator (IPTW)

- Example: Crack users with pill boxes are over-represented in our sample
  - Relatively more of them than we would have had if pill box use had been randomly assigned
  - Those crack users who don't use pill boxes that we do have in our sample get a bigger weight
  - Those non-crack users who do use pill boxes that we do have in our sample get a bigger weight

Implementation IPTW Estimator (1)

1. Estimate the treatment mechanism

   $g_0(a|W)=P_0(A=a|W)$

   - Predict treatment assignment given confounders

   1. Using an a prior specified parametric model
      - Example: $g(A = 1|W) = expit(\beta_0 + \beta_1 W_1 + \beta_2 W_2)$
   2. Using data-adaptive approaches/ Super learning….

Implementation IPTW Estimator (2)

2. Predict each individual's probability of receiving his/her observed treatment given his/her covariates: $\frac{1}{g_n(A_i|W_i)}$

Denotes Estimator of $g_0(A|W)$

3. Each subject gets a weight inverse to this predicted probability

$$\hat{weight}_i = \frac{1}{g_n(A_i|W_i)}$$

Implementation IPTW Estimator (3)

4. Take empirical mean of the outcome times a weight $$\frac{1}{n}\sum_{i=1}^{n}\frac{I(A_i=a)}{g_n(A_i|W_i)}Y_i$$

- IPTW estimator of $\Psi(P_0)$ (equal under casual assumptions to ATE)

$$\hat{\Psi}(P_n)=\frac{1}{n}\sum_{i=1}^{n}\frac{I(A_i=1)}{g_(A_i|W_i)}Y_i-\frac{1}{n}\sum_{i=1}^{n}\frac{I(A_i=0)}{g_(A_i|W_i)}Y_i$$

Recap: Simple Example of a Substitution Estimator

Original Data

ID	Pill Box (A)	Crack Use (W)	Adherence (Y)
1	1	1	0.7
2	0	0	0.8
3	1	1	0.4
4	1	0	1
5	0	1	0.4
6	0	0	0.7

Simulated G-comp Data

ID	Pill Box (a)	Predicted Adherence ($\hat{Y}$)
1	0	0.4
2	0	0.75
3	0	0.4
4	0	0.75
5	0	0.4
6	0	0.75
1	1	0.55
2	1	1.0
3	1	0.55
4	1	1.0
5	1	0.55
6	1	1.0

$$\hat{E}(Y|A=1,W=1)=0.55$$
$$\hat{E}(Y|A=0,W=1)=0.4$$
$$\hat{E}(Y|A=1,W=0)=1.0$$
$$\hat{E}(Y|A=0,W=0)=0.75$$

Simple Example: Substitution Estimator

Simulated G-comp Data

ID	Pill Box (a)	Predicted Adherence ($\hat{Y}$)
1	0	0.4
2	0	0.75
3	0	0.4
4	0	0.75
5	0	0.4
6	0	0.75
1	1	0.55
2	1	1.0
3	1	0.55
4	1	1.0
5	1	0.55
6	1	1.0

$$\frac{1}{n}\sum_{i=1}^{n}\hat{E}(Y|A=0,W_i)=0.575$$

$$\frac{1}{n}\sum_{i=1}^{n}\hat{E}(Y|A=1,W_i)=0.775$$

Simple Example: IPTW

ID	Pill Box (A)	Adherence (Y)	Crack (W)	g(A|W)	1/g(A|W)
1	1	0.7	1	0.67	1/0.67=1.5
2	0	0.8	0	0.67	1/0.67=1.5
3	1	0.4	1	0.67	1/0.67=1.5
4	1	1.0	0	0.33	1/0.33=3.0
5	0	0.4	1	0.33	1/0.33=3.0
6	0	0.7	0	0.67	1/0.67=1.5

$$\hat{P}(A=1|W=1) = g_n(1|W=1) = 2/3 = 0.67$$

$$\hat{P}(A=0|W=1) = g_n(0|W=1) = 1/3 = 0.33$$

$$\hat{P}(A=1|W=1) = g_n(1|W=0) = 1/3 = 0.33$$

$$\hat{P}(A=0|W=1) = g_n(0|W=1) = 2/3 = 0.67$$

Simple Example: IPTW

ID	Pill Box (A)	Adherence (Y)	Crack (W)	g(A|W)	1/g(A|W)
1	1	0.7	1	0.67	1/0.67=1.5
2	0	0.8	0	0.67	1/0.67=1.5
3	1	0.4	1	0.67	1/0.67=1.5
4	1	1.0	0	0.33	1/0.33=3.0
5	0	0.4	1	0.33	1/0.33=3.0
6	0	0.7	0	0.67	1/0.67=1.5

$$\frac{1}{n}\sum_{i=1}^{n}\frac{I(A_i=0)}{g_n(A_i|W_i)}Y_i = (1.5\times 0.8 + 3\times 0.4 + 1.5\times 0.7)/6 = 0.575$$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{I(A_i=1)}{g_n(A_i|W_i)}Y_i = (1.5\times 0.7 + 1.5\times 0.4 + 3\times 1)/6 = 0.775$$

IPTW vs. Simple Substitution Estimator

- If $E_0(Y|A,W)$ and $g_0(A|W)$ are estimated with NP-MLE the two estimators are identical
- In other settings they may be quite different
  - Based on estimators of different parts of the likelihood
- Simple Substitution Estimator: consistency depends on consistent estimation of $E_0(Y|A,W)$
- IPTW: consistency relies on consistent estimation of $g_0(A|W)$

Simulation results

- "G-comp": Substitution estimator using correctly (Qc) and incorrectly (Qm) specified parametric model for E(Y|A,W)
- IPTW: IPTW estimator using correctly (gc) and incorrectly (gm) specified parametric model for g(a|W)

Table 2: Performance of estimators by specification in Simulation 1: g_0 in [0.48,0.92], shown for unbounded g_n only. Results are based on 250 samples of size 1000.

	Qcgc			Qcgm			Qmgc		
	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
G-COMP	1.5e-03	5.9e-03	5.9e-03	1.5e-03	5.9e-03	5.9e-03	2.6e-01	1.9e-02	8.5e-02
IPTW	6.0e-03	9.2e-03	9.2e-03	2.6e-01	2.1e-02	9.0e-02	6.0e-03	9.2e-03	9.2e-03

Petersen et al, 2012: Diagnosing and responding to violations in the positivity assumption, Tech Report

Estimation of $g_0(A|W)$

- As usual, often our statistical model is non-parametric and W is high dimensional
- Could use an *a priori* specified parametric model, but if it is misspecified, your estimator will be both overly biased and inconsistent
- Alternative: use data-adaptive approaches such as Super Learner…
  - Can help to get a better estimator of $g_0(A|W)$
  - Doesn't solve all of our problems…

Real Example: Effect of Pill Box use on Adherence

- Dimension reduction of W based on univariate associations with Y
- Implementation of Substitution Estimator
  - D/S/A to fit E(Y|A,W)
    - CV to choose # of terms, max degree and order of interactions
- Implementation of IPTW Estimator
  - D/S/A to fit g(A|W)
    - CV to choose # of terms, max degree and order of interactions
- Standard Errors- NP Bootstrap
  - Resample by individual

Results:
Estimated effect of pill box use

Method	Difference in adherence,[a] % (95% CI)	Reduction in viral load, mean $\log_{10}$ copies/mL[b] (95% CI)	Viral load <400 copies/mL, OR [b] (95% CI)
G-computation	4.5 (2.0–7.0)	0.34 (0.08–0.60)	1.81 (1.25–2.62)
IPTW	4.1 (0.0–8.3)	0.37 (0.05–0.69)	1.91 (1.27–2.90)

Petersen ML, et. al.. Pillbox Organizers are Associated with Improved Adherence to HIV Antiretroviral Therapy and Viral Suppression. A Marginal Structural Model Analysis. *Clinical Infectious Diseases*: 2007 45(7):908-15.

Issue #1: How to estimate g(A|W)

- Recall: consistency of IPTW estimator relies on consistent estimation of $g_0(A|W)$
- Use of a misspecified parametric model to estimate $g_0(A|W)$ can result in a biased estimator of $\Psi(P_0)$
- Estimate $g_0(A|W)$ data adaptively? (eg with Super Learner)?

Data adaptive estimation of $g_0(A|W)$

- We can use data adaptive methods to estimate $g_0(A=a|W)$
  - Choose candidate that minimizes the cvRisk estimate, using, eg –log Loss function
- However, this estimator of $g_0(A=a|W)$ is again not the best choice for the target parameter we care about

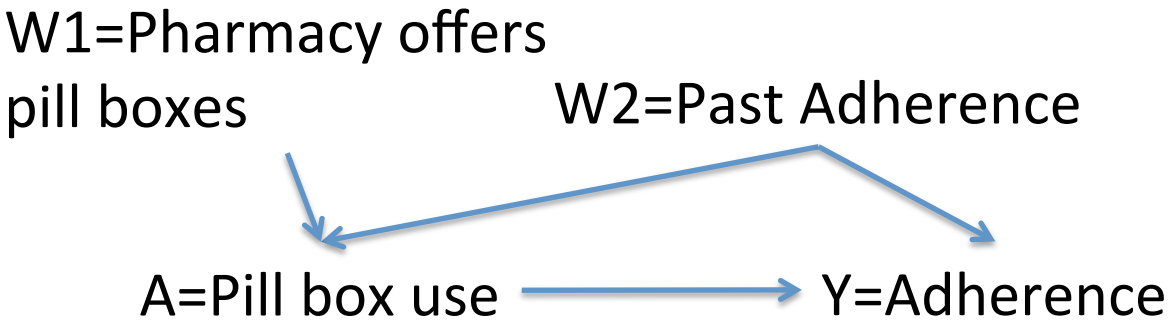
  $E_{W0}(E_0(Y|A=1,W)-E(Y|A=0,W))$

Data adaptive estimation of $g_0(A|W)$

- For example, Including a covariate with a strong effect on A and a weak (or no) effect on Y can hurt the performance of your estimator
  - Potential to increase both bias and variance

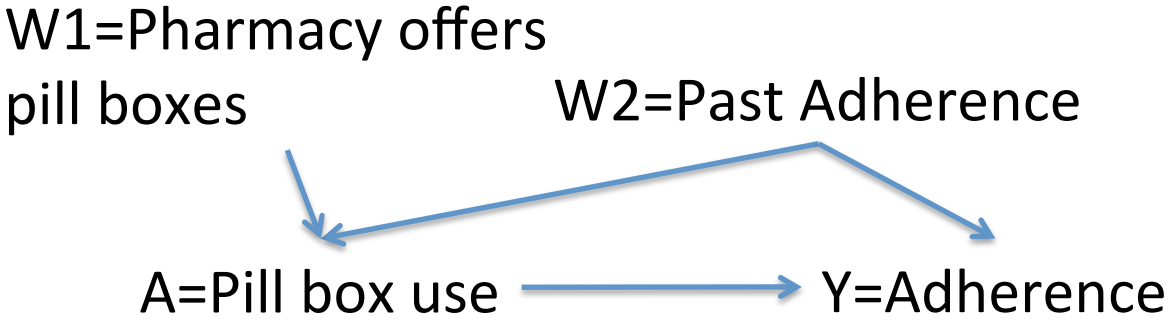
Ex. Limitations of Data-adaptive estimation of $g_0(A|W)$

- Whether your pharmacy offers pill boxes has a strong effect on whether you use a pill box
- Whether your pharmacy offers pill boxes does not affect the outcome



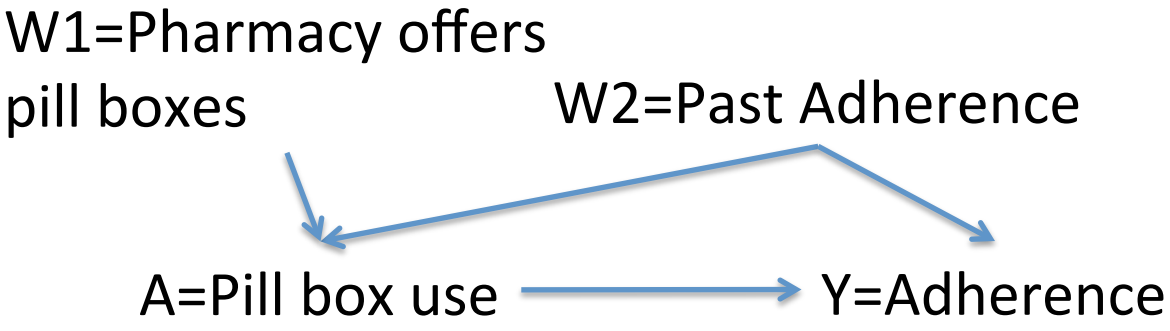
Ex. Limitations of Data-adaptive estimation of $g_0(A|W)$

- SL simply has the task of predicting A given W
  - If W1 is a strong predictor, and W2 is a weak predictor, SL might select an estimator of g(A|W) that includes W1 and not W2
  - You could force it to include W2, but….



Ex. Limitations of Data-adaptive estimation of $g_0(A|W)$

- Including W1 in the fit of $g_0(A|W)$ will still hurt the the performance of the IPTW estimator of $\Psi(P_0)$
  - Increases the variability of the weights (and thus variance)
  - It can also increase bias as a result of positivity violations- more on this shortly



How to proceed? Some simple steps

1. Specify your causal model with care

- If you started by specifying your causal model and evaluating the covariates needed for the backdoor criteria, your estimand would not have involved conditioning on W1

How to proceed? Some simple steps

2. Do an initial dimension reduction of candidate covariates based on univariate associations with Y
   - Part of your pre-specified algorithm- not *ad hoc*
- Neither option is perfect…

Issue #2: Positivity

- To ensure identifiability, we need an additional assumption: positivity or experimental treatment assignment
- Strong version: Assume that every treatment level of interest is represented in every covariate strata

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

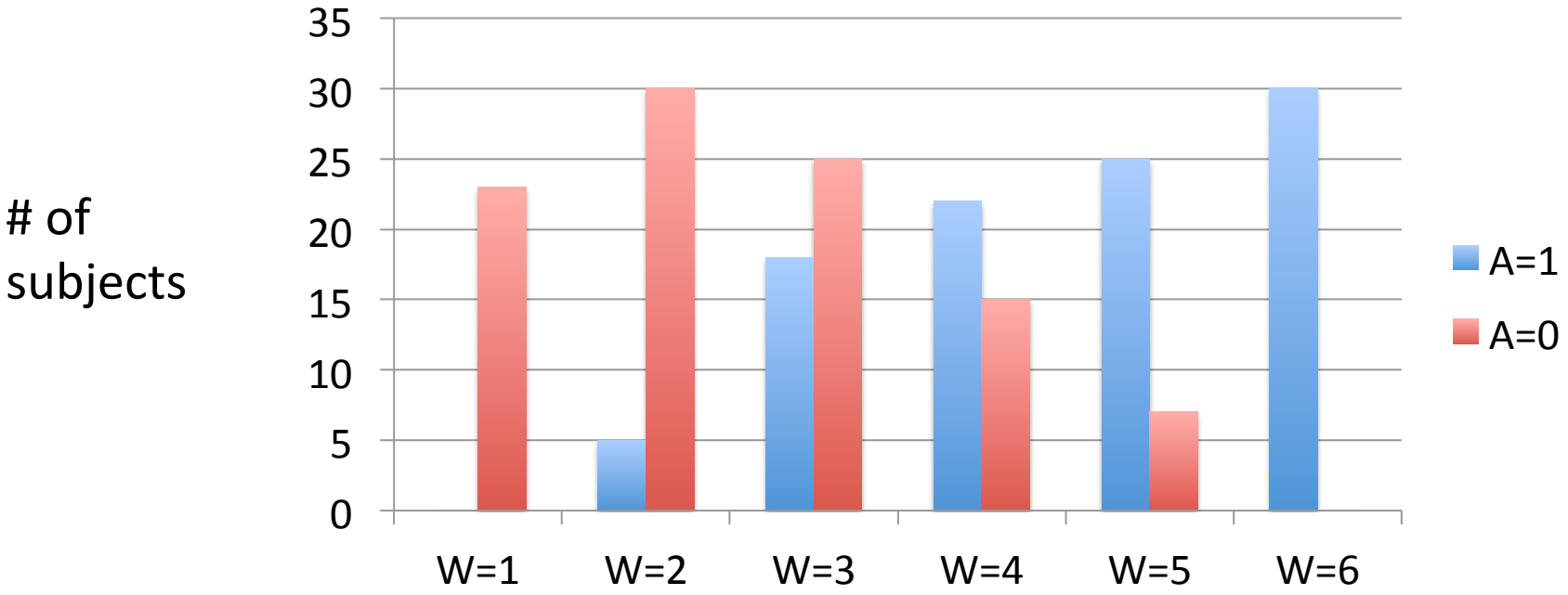
$$\text{for all } w \text{ for which } P_0(W = w) > 0$$

“Practical violations” of the positivity assumption

- Even if the positivity assumption holds for P_0, it may be “practically violated” for a given sample P_n from P_0
- Certain strata of W for which there is no variability in treatment

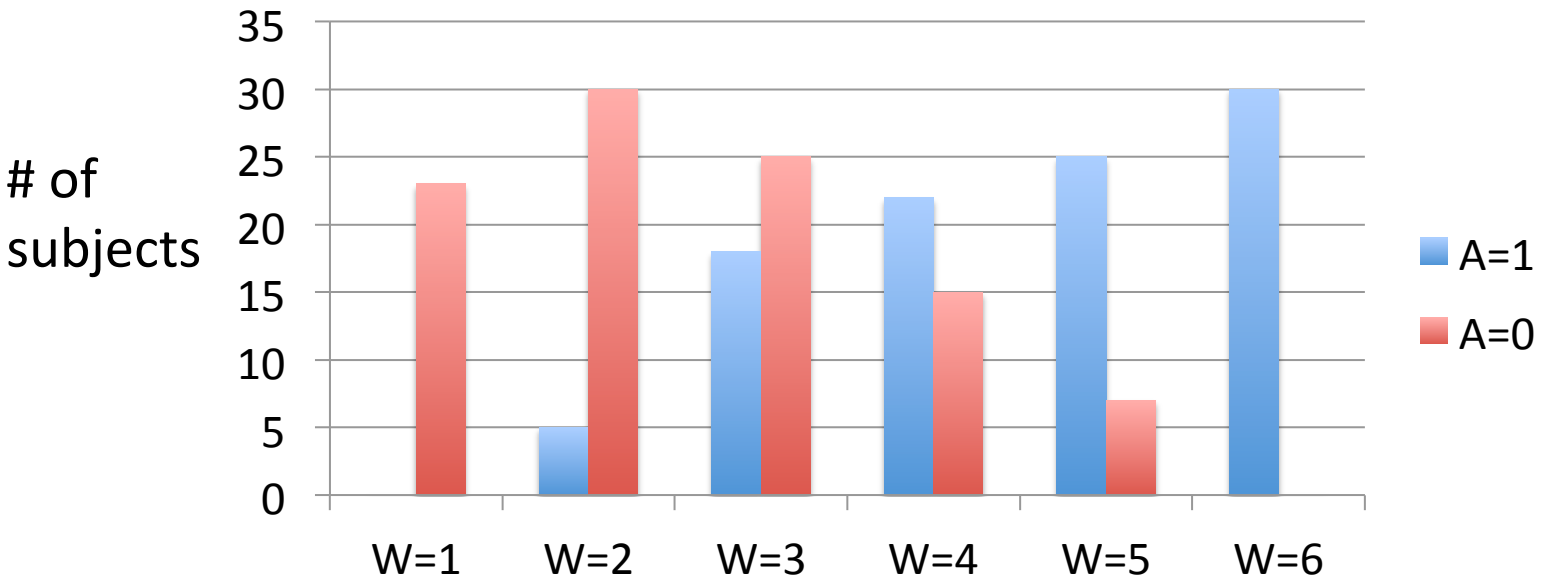
Example

- Target causal parameter: Average effect of drug on incident cardiovascular disease (CVD)
- W=Baseline Risk of CVD (6 level ordinal variable)
  - W=1 low risk; W=6 high risk



Example

- Low risk group very unlikely to get the drug
  - $P_0(A=1|W=1)$ small
- High risk group very unlikely to not get the drug
  - $P_0(A=0|W=6)$ small



NP-MLE?

- Estimate mean of Y in each strata of (A,W)?
- Empty cells

	A=1	A=0
W=1	??? (n=0)	.17 (n=22)
W=2	.20 (n=5)	.18 (n=30)
W=3	.18 (n=17)	.22 (n=25)
W=4	.22 (n=22)	.32 (n=15)
W=5	.25 (n=35)	.14 (n=7)
W=6	.30 (n=30)	?? (n=0)

Basic Problem

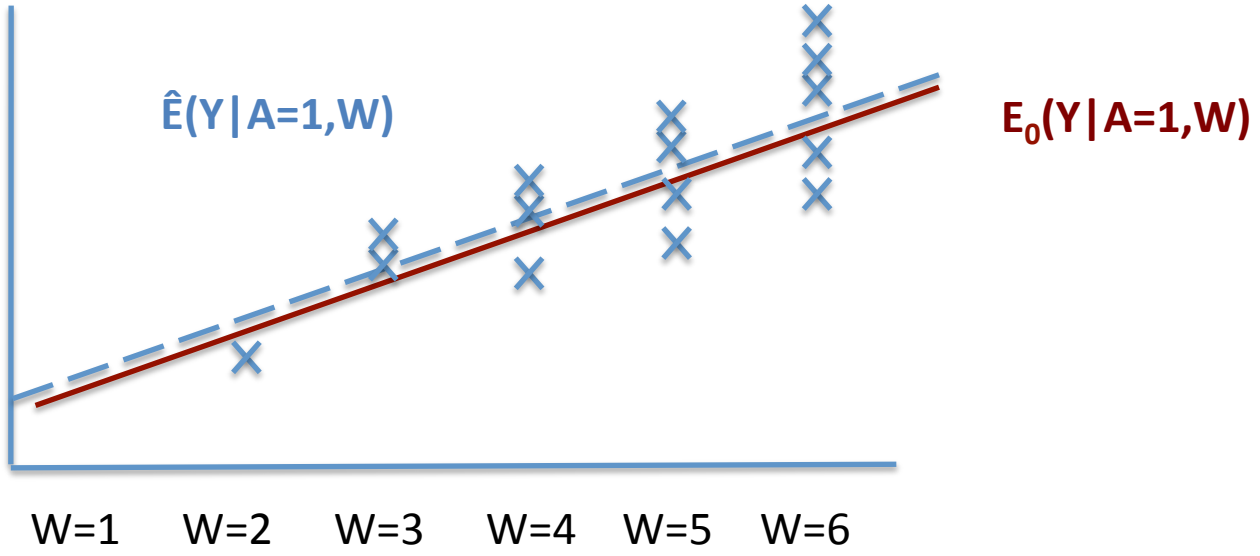
- There is nothing in our observed data to give us information about the effect of the drug in the high and low risk strata
- If we want to say something about the effect of the drug in the whole population (including people in these strata) we are forced to go beyond what the data can tell us…

Substitution estimators and violations of the positivity assumption

- Depending on how $E_0(Y|A,W)$ is estimated, estimator may extrapolate to regions of (A,W) not supported by the data
- If $E_0(Y|A,W)$ is based on a correctly specified parametric model, this extrapolation will be accurate
- However, In general we don't have correctly specified parametric models

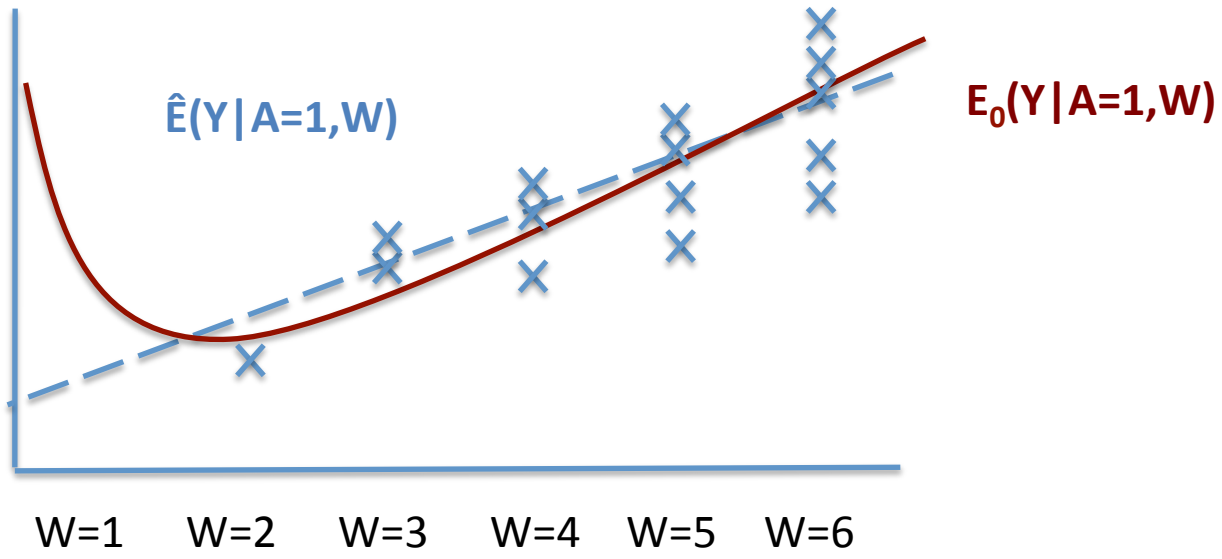
Our estimate of $E_0(Y|A,W)$ may allow us to extrapolate to regions of (A,W) not supported by the data

- Ex. Estimate $E_0(Y|A,W)$ by fitting coefficients β of a linear model: $\beta_0+\beta_1 A+\beta_2 W+\beta_3 AW$
- Extrapolate what the outcome would have been among exposed in low risk strata based on outcomes among exposed in higher risk strata



Bias due to misspecified parametric model in presence of positivity violation

- Ex. Estimate $E_0(Y|A,W)$ by fitting coefficients β of a linear model: $\beta_0+\beta_1 A+\beta_2 W+\beta_3 AW$
  - Extrapolate what the outcome would have been among exposed in low risk strata based on outcomes among exposed in higher risk strata



IPTW and Positivity Violations

- Unlike a substitution estimator, the IPTW estimator cannot extrapolate
- Finite sample performance highly susceptible to positivity violations, including practical violations
- Example:
  - Low risk group underrepresented among treated
  - Treated low risk subjects should be up-weighted
  - However…We don't have any treated subjects in lowest risk group to up-weight…!

“Near” violations of the positivity assumption can also result in poor finite sample performance

- Example: Say there was 1 low risk treated subject with W=1
- A single individual is now providing all of the information about outcomes in the strata (A=1,W=1)
- This individual will get a big weight
  - High variance estimator

How do I know if positivity violations are causing problems?

- First simple thing to do: Look at the distribution of the weights
- Extreme weights-> near positivity violations
- Elevating the variance of your estimate

Truncation of the Weights

- A common approach in the sampling literature: truncate extreme weights
  - At absolute value (eg 10, 20- often set relative to sample size)
  - At a % (eg at 99th percentiles)
- Reduces variability
- However… this comes at a price!

Truncation of the Weights

- Consistency of the IPTW estimator relies on having a consistent estimator of $g_0(A|W)$
- If the true $g_0(A|W)$ is bounded away from 0 and you know this, definitely a good idea to incorporate that model knowledge…
- However, often we either do not know this
  - Or know only enough to bound it as a very small value
  - Still end up with small $g_n(A|W)$ and big weights

Truncation of the Weights

- In this case, by bounding the predicted probabilities, you are ensuring that your estimator of $g_0(A|W)$ is not consistent
- -> Resulting IPTW estimator will be biased
- Intuition: No longer adjusting completely for measured covariates

Simulation results: Positivity violations

Table 4: Performance of estimators by specification and by bound on g_n in Simulation 2: g_0 in [0.001,1]. Results are based on 250 samples of size 1000.

Bound on g_n	Qcgc			Qcgm			Qmgc		
	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
G-COMP									
None	0.007	0.009	0.009	0.007	0.009	0.009	1.145	0.025	1.336
[0.025,0.975]	0.007	0.009	0.009	0.007	0.009	0.009	1.145	0.025	1.336
[0.05,0.95]	0.007	0.009	0.009	0.007	0.009	0.009	1.145	0.025	1.336
[0.1,0.9]	0.007	0.009	0.009	0.007	0.009	0.009	1.145	0.025	1.336
IPTW									
None	0.544	0.693	0.989	1.547	0.267	2.660	0.544	0.693	0.989
[0.025,0.975]	1.080	0.090	1.257	1.807	0.077	3.340	1.080	0.090	1.257
[0.05,0.95]	1.437	0.059	2.123	2.062	0.054	4.306	1.437	0.059	2.123
[0.1,0.9]	1.935	0.043	3.787	2.456	0.043	6.076	1.935	0.043	3.787

Petersen et al, 2012: Diagnosing and responding to violations in the positivity assumption, SMMR

How to choose a truncation level…?

- In summary, truncating your weights can help or hurt your estimator's performance
- Further, choice of truncation level can have a big impact
- Tools for choosing a truncation constant data-adaptively….
  - Bembom & vdL: "Data-Adaptive selection of the truncation level for IPTW estimators"

I don't have any extreme weights- that must mean that all is well…

- No! if there are certain values of W for which no one in your sample gets A=a, won't be any extreme weights
  - If there is one subject in the low risk group that gets treated, that subject will get a large weight
  - If there are no subjects in this group, no large weights
- Looking at the distribution of $g_n(A=a|W)$ for each a of interest can help you realize you have a problem
  - "Propensity Scores"
- However, truncation is not going to solve it

An alternative IPTW estimator…

- 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

An alternative IPTW estimator…

- Standard IPTW estimator of $E(Y_a)$
  - Horvitz-Thompson estimator- from survey sampling $\frac{1}{n}\sum_{i=1}^{n}\frac{I(A_i=a)}{g_n(A_i|W_i)}Y_i$
- Stabilized IPTW estimator
  - Modified Horvitz-Thompson estimator

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\frac{I(A_i=a)g*(A_i)}{g_n(A_i|W_i)}Y_i}{\frac{1}{n}\sum_{i=1}^{n}\frac{I(A_i=a)g*(A_i)}{g_n(A_i|W_i)}}$$

Alternative IPTW Estimator

- A weighted empirical mean
  - Assign people with A=a a weight: $g^*(A)/g_n(A|W)$
  - Everyone else gets weight=0
  - Take the mean of the outcome in the reweighted population

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\frac{I(A_i=a)g*(A_i)}{g_n(A_i|W_i)}Y_i}{\frac{1}{n}\sum_{i=1}^{n}\frac{I(A_i=a)g*(A_i)}{g_n(A_i|W_i)}}$$

Sum of the weighted Outcomes

Sum of the weights

- Ensures the IPTW estimator is bounded
  - Eg if Y is binary, wont get an estimated probability>1

Key Points: IPTW Estimator

- Can rewrite the G-computation estimand as a function of the distribution of the treatment/ exposure given covariates
  - "treatment mechanism": g(A|W)
- Intuition: Use estimate of the treatment mechanism to
  - Up-weight individuals with rare treatment/covariate combinations
  - Down-weight those with common treatment/ covariate combinations
- Relies on doing a good job estimating g(A|W)

Key Points: IPTW

- Often has poor finite sample performance in setting of strong confounding
  - Bias and high variance due to violations/near violations of the positivity assumption
- For IPTW estimator of $E_W(E_0(Y|A=a,W))$
  - Use of the "modified Horvitz Thompson" estimator can help
- Truncation of weights can help or hurt
  - Choosing the right level to truncate is not straightforward