INVERSE PROBABILITY WEIGHTING
Estimation

Miguel Hernán, Barbra Dickerman, Joy Shi
DEPARTMENTS OF EPIDEMIOLOGY AND BIOSTATISTICS
HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Learning objectives
At the end of this lecture you will be able to

- Estimate IP weights nonparametrically and parametrically
- Discuss violations of positivity and its consequences
- Understand the relation between IP weighting and standardization

- Key concepts
  - IP weights
  - Structural non-positivity
  - Random non-positivity

Study population

- 1629 cigarette smokers
- Aged 25-74 years when interviewed in 1971-75 (baseline)
- Interviewed again in 1982
- Known sex, age, race, weight, height, education, alcohol use, and smoking intensity at both baseline and follow-up visits, and who answered the general medical history questionnaire at baseline

Key variables

Treatment A	Quit smoking between baseline and 1982 1: yes, 0: no
Continuous outcome Y	Weight gain, kg Weight in 1982 minus baseline weight Available for 1566 individuals
Dichotomous outcome D	Death by 1992 1: yes, 0: no
Baseline (pre-treatment) covariates	Age, sex, race, alcohol use, intensity of smoking, weight…

Causal questions of interest

1. What is the effect of smoking cessation on weight gain?
2. What is the effect of smoking cessation on risk of death?

- This is an informal statement of the questions

A more formal version of causal question #1 First define the counterfactual means

if everybody had quit smoking

- $\mathrm{E}[Y^{a=1}]$
- $Y^{a=1}$ is an individual's outcome under $a=1$

if nobody had quit smoking

- $\mathrm{E}[Y^{a=0}]$
- $Y^{a=0}$ is an individual's outcome under $a=0$

Then the formal question is:

- What is the average causal effect $\mathrm{E}[Y^{a=1}] - \mathrm{E}[Y^{a=0}]$?

Plan for today

A. Estimation of the causal effect by IP weighting

- Nonparametric estimation of IP weights
- Parametric estimation of IP weights
- Positivity

B. IP weighting vs. standardization

What if this were an ideal randomized experiment…

- … in which individuals had been randomly assigned to "smoking cessation" with a probability that depends on their age group?
  - Randomization is **conditional** on age group, rather than unconditional (or marginal)
- Probability of being assigned to smoking cessation is
  - 33.3% if age>50 years ($L=1$)
  - 22.5% if age≤50 years ($L=0$)

We need to estimate $E[Y^{a=1}]$ and $E[Y^{a=0}]$

$E[Y^{a=1}]$: the mean outcome had everyone been treated

- consistently estimated by the IP weighted mean in the treated

$E[Y^{a=0}]$: the mean outcome had everyone been untreated

- consistently estimated by the IP weighted mean in the untreated

Under

✓ Exchangeability: expected
✓ Positivity: guaranteed
✓ Consistency: guaranteed

Reminder: IP weights

$$W^A = \frac{1}{f(A|L)}$$

For the treated

- $1 / \Pr[A=1|L]$
- One over the probability of smoking cessation

For the untreated

- $1 / \Pr[A=0|L] = 1 / (1 - \Pr[A=1|L])$
- One over the probability of no smoking cessation

☐ Each individual in the population is weighted to create W^A individuals in the pseudo-population

- The denominator of your W^A is (informally) the probability of having your observed treatment value given your L value

Is the denominator of the weight W^A the same as the propensity score?

Recall: The PS is the probability of being treated conditional on the confounders or $Pr[A = 1|L]$.

0

Yes 0%

No 0%

Don't know 0%

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

IP weights for smoking cessation

- If the study were a randomized experiment conditional on age group
  - These 2 probabilities (and thus the weights) are known because they were chosen by the investigators
- If the study were observational
  - These 2 probabilities (and therefore the weights) need to be estimated from the data
  - Let's estimate these 2 probabilities both without and with models

Nonparametric estimation
of IP weights using sample proportions

- Older age
  - the estimate of $\Pr[A=1|L=1]$ is 156/468 = 0.3333
  - Treated W^A : 1 / 0.3333 = 3.000
  - Untreated W^A: 1 / (1 − 0.3333) = 1.500
- Younger age
  - the estimate of $\Pr[A=1|L=0]$ is 247/1098 = 0.2250
  - Treated W^A : 1 / 0.2250 = 4.444
  - Untreated W^A : 1 / (1 − 0.2250) = 1.290

See ipw_weightestimation.R, lines 12-18

The pseudo-population

- We assign a weight W^A to each individual in the population
- Mathematically, the weighting simulates a (pseudo-)population in which the relation between variables change

Is treatment independent of the measured confounders *L* in the pseudo-population?

0

(A) Yes

0%

(B) No

0%

(C) Don't know

0%

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Is treatment independent of all confounders in the pseudo-population?

0

Yes

0%

No

0%

Don't know

0%

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

IP weighted averages

- Compute the IP weighted average (that is, the average in the pseudo-population) of the outcome in the
  - treated: 4.87 kg
  - untreated: 1.87 kg
- Difference: 3.0 kg *See ipw_weightestimation.R, lines 20-21*
  - causal interpretation if treatment had been conditionally randomized within levels of age group
    - if conditional exchangeability given age group

Nonparametric estimation
of IP weights using a saturated model

- Logistic model $\text{logit}\Pr[A=1|L]=\theta_0+\theta_1 L$
  - Older age: the estimate of $\Pr[A=1|L=1]$ is 0.3333
  - Younger age: the estimate of $\Pr[A=1|L=0]$ is 0.2250
- A saturated model because
  - 2 parameters, 2 quantities to be estimated
  - No restrictions
- The estimates from the model are equal to the nonparametric estimates we obtained before
  - Same IP weighted means

See ipw_weightestimation.R, lines 24-28

Plan for today

A. Estimation of the causal effect by inverse probability (IP) weighting

- Nonparametric estimation of IP weights
- Parametric estimation of IP weights
- Positivity

B. IP weighting vs. standardization

This is an observational study

- Smoking cessation A was *not* conditionally randomized by the investigators
- However, we may believe that A occurred at random within levels of
  - L: sex, M: race, N: age group
  - and use IP weighting to adjust for them
- If our assumption of conditional exchangeability given those 3 covariates is correct, then
  - the IP weighted means equal the counterfactual means

Estimating IP weighted averages with several confounders

- Estimate the probability of treatment A for every combination of values of the variables in L,M,N
- Then calculate each individual's corresponding IP weight, and the IP weighted averages of Y in the treated and the untreated
  - i.e., the averages in the pseudo-population
- The difference of IP weighted averages consistently estimates the average causal effect $E[Y^{a=1}] - E[Y^{a=0}]$

Nonparametric estimation of IP weights with 3 dichotomous confounders

- We need to estimate 2×2×2=8 quantities
  - $\Pr[A=1| L=1, M=1, N=0]$, $\Pr[A=1| L=1, M=0, N=0]$
  - $\Pr[A=1| L=0, M=1, N=0]$, $\Pr[A=1| L=0, M=0, N=0]$
  - $\Pr[A=1| L=1, M=1, N=1]$, $\Pr[A=1| L=1, M=0, N=1]$
  - $\Pr[A=1| L=0, M=1, N=1]$, $\Pr[A=1| L=0, M=0, N=1]$
- Can calculate the sample proportions or fit a saturated logistic model with 8 parameters

$$\text{logit} \Pr[A=1|L,M,N]=$$
$$\theta_0 + \theta_1 L + \theta_2 M + \theta_3 N + \theta_4 LM + \theta_5 LN + \theta_6 MN + \theta_7 LMN$$

Parametric estimation
of IP weights

- Now suppose the variable 'age' takes 50 values from 25 to 74 years
  - need to estimate 2×2×50=200 probabilities $\Pr[A=1|L,M,N]$
- Either by calculating sample proportions or fitting a logistic model with 200 parameters?
  - I don't think so. We have only ~1500 individuals
- Typical logistic model in this setting

$$\text{logit} \Pr[A=1|L,M,N] = \theta_0 + \theta_1 L + \theta_2 M + \theta_3 N$$

Is this model a saturated model?

$$\text{logit} \Pr[A=1|L,M,N] = \theta_0 + \theta_1 L + \theta_2 M + \theta_3 N$$

0

Yes 0%

No 0%

Don't know 0%

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Estimating IP weighted averages with many confounders (I)

- In our study, L is a vector of 9 confounders (4 continuous)
- Impossible to nonparametrically estimate the probability of treatment A for each combination of values of the variables in L

- We will need to use a nonsaturated model
  - e.g., a logistic regression model with a few parameters

Estimating IP weighted averages with many confounders (II)

- For example, a logistic model with
  - indicators for discrete variables
  - linear+quadratic terms for continuous variables
  - no product terms
- The predicted values from this model are estimates of the probability of treatment in each combination of values of the confounders
  - Then calculate each individual's corresponding IP weight, and the IP weighted averages of Y in the treated and the untreated

Parametric estimation of IP weights
Nonsaturated logistic model

Estimates of IP weighted average of the outcome
- 5.2 kg in the treated
- 1.8 kg in the untreated

95% confidence interval can be obtained via bootstrapping or other methods

Difference: 3.4 kg *See ipw_weightestimation.R, lines 44-57*
- This difference would have a causal interpretation as $E[Y^{a=1}] - E[Y^{a=0}]$ if
  - all confounders had been included in the IP weighting procedure
  - the model for treatment is correctly specified

Plan for today

A. Estimation of the causal effect by inverse probability (IP) weighting
- Nonparametric estimation of IP weights
- Parametric estimation of IP weights
- Positivity

B. IP weighting vs. standardization

What about positivity?

- There are 4 white women aged 66 yrs
  - None of them quit smoking
- The nonparametric estimate of the probability of quitting is 0
  - The weight $W^A = 1/0$ is undefined
- Positivity is violated:
  - Nonparametric IP weighting cannot be used to estimate average causal effect
  - Stratification and standardization cannot be used either

See ipw_weightestimation.R, lines 60-61

Two types of violations of positivity

- Structural
  - Individuals with certain confounder values cannot possibly be treated (untreated)
    - e.g., those off work cannot receive occupational exposures
    - always zero cells at those values when conditioning on confounders
- Random
  - Our sample is not infinite so, if we stratify on many confounders, we'll start finding zero cells
    - At different places in different samples

What can be done about positivity violations?

- Structural
  - No causal inferences for subsets with structural zeroes
  - Need to restrict the study population to the others
- Random
  - Use parametric models to smooth over the zeros
  - That is, to borrow information from other individuals

In our example

- Our nonsaturated logistic model will estimate the probability of quitting in white women aged 66 by using data from white women age 65 and 67, etc.

- If we go ahead with parametric estimation of IP weights in the presence of zero cells, we are effectively assuming random non-positivity

If we find no zeroes in our data, structural positivity holds

0

True

0%

False

0%

If we find zeroes in our data, structural positivity does not hold

0

True

0%

False

0%

Plan for today

A. Estimation of the causal effect by inverse probability (IP) weighting

- Nonparametric estimation of IP weights
- Parametric estimation of IP weights
- Positivity

B. IP weighting vs. standardization

IP weighting vs. standardization

- Both methods provide consistent estimators of the difference $E[Y^{a=1}] - E[Y^{a=0}]$
  - under exchangeability, positivity, consistency
- Both methods are algebraically equivalent
  - when using nonparametric estimates
- The standardized averages are equal to the IP weighted averages
  - when using nonparametric estimates

IP weighting = Standardization

- But only in the nonparametric case
  - That is, only when saturated models are used

- Each method computes a different component of the joint distribution
  - IP weighting: $f[A|L]$
  - Standardization: $f[L], f[Y|A,L]$
- Even slight model misspecification will result in different estimates
  - because misspecification of different models will generally have different effects on the final estimate

IP weighting ≠ Standardization when modeling high-dimensional data

- Data may be high-dimensional because
  - many categorical variables
  - continuous variables
  - (time-varying variables)
  - all of the above
- We need to use parametric models to obtain effect estimates
  - IP weighting: estimate $f[A|L]$
  - Standardization: estimate $f[Y|A,L], f[L]$

So what's better? IP weighting or standardization?

- Not a simple answer
  - We'll come back to this question after describing marginal structural models in the next lecture
- But anyway, why choose?
  - Use both methods to triangulate estimates

Readings

- *Causal Inference: What If*. Chapter 12

In observational studies, (conditional) exchangeability is 0

Expected 0%

Guaranteed 0%

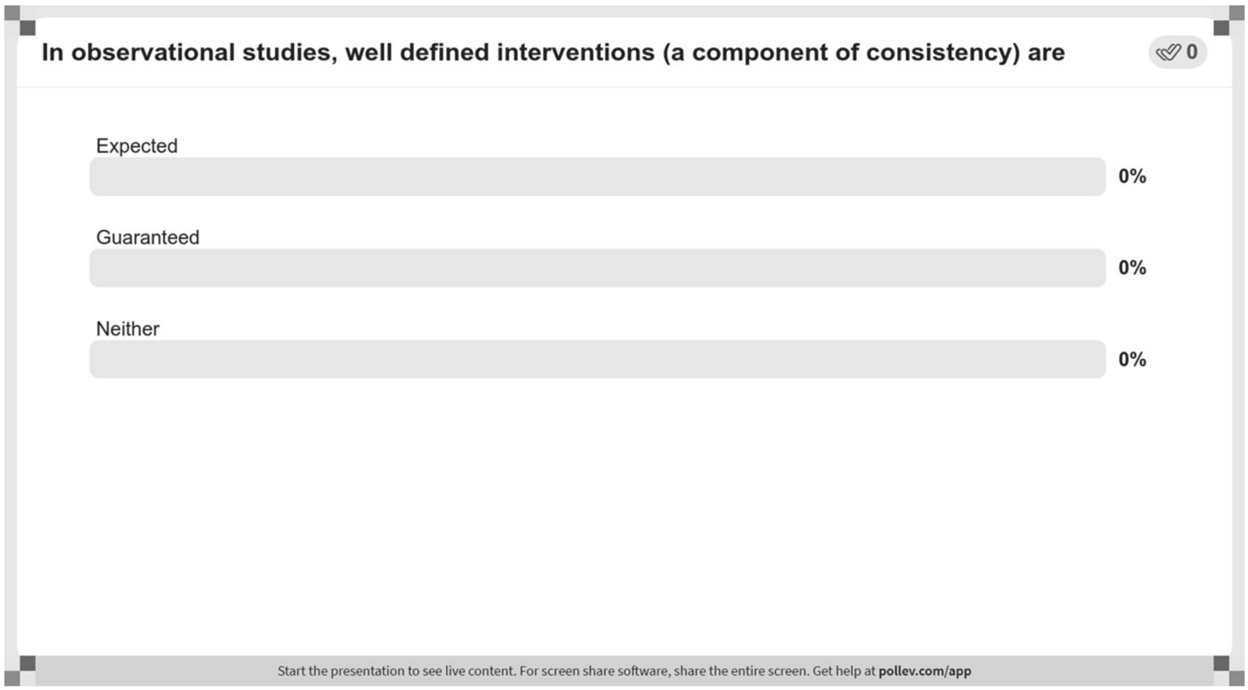
Neither 0%

In observational studies, (conditional) positivity is 0

Expected 0%

Guaranteed 0%

Neither 0%



Progress report

1. Introduction to modeling
2. Stratified analysis:
   - outcome regression
   - propensity scores
3. Standardization
4. IP weighting