

Chapter 12: IP Weighting and Marginal Structural Models

Contents

| | | |
|----------|--|----------|
| 1 | 12.1 The Causal Question (pp. 157-159) | 1 |
| 1.1 | Research Question | 2 |
| 1.2 | Measured Confounders | 2 |
| 2 | 12.2 Estimating IP Weights (pp. 159-163) | 2 |
| 2.1 | IP Weights Definition | 2 |
| 2.2 | Estimating Propensity Scores | 3 |
| 2.3 | Example: NHEFS Data | 3 |
| 3 | 12.3 Stabilized IP Weights (pp. 163-165) | 3 |
| 3.1 | Properties of Stabilized Weights | 4 |
| 3.2 | Example: NHEFS Data | 4 |
| 4 | 12.4 Marginal Structural Models (pp. 165-169) | 4 |
| 4.1 | Fitting Marginal Structural Models | 4 |
| 4.2 | Example: NHEFS Study | 4 |
| 5 | 12.5 Effect Modification and Marginal Structural Models (pp. 169-171) | 5 |
| 5.1 | MSM with Effect Modification | 5 |
| 5.2 | Fitting MSMs with Effect Modification | 5 |
| 5.3 | Example: Effect Modification by Sex | 5 |
| 6 | 12.6 Censoring and Missing Data (pp. 171-173) | 5 |
| 6.1 | IP Weights for Censoring | 6 |
| 6.2 | Joint Weights for Treatment and Censoring | 6 |
| 6.3 | Example: NHEFS with Loss to Follow-up | 6 |
| 7 | 12.7 A Likelihood Approach (pp. 173-174) | 6 |
| 7.1 | Weighted Likelihood | 6 |
| 7.2 | Connection to Maximum Likelihood | 7 |
| 8 | Summary | 7 |

This chapter introduces **inverse probability (IP) weighting**, a method for estimating causal effects that creates a pseudo-population in which treatment is independent of measured confounders. IP weighting is used to fit **marginal structural models**, which provide a natural framework for estimating marginal causal effects when treatment and confounding vary over time.

This chapter is based on Hernán and Robins (2020, chap. 12, pp. 157-174).

Key concepts: IP weighting creates a pseudo-population where confounding is eliminated by construction. This approach is particularly powerful for time-varying treatments and confounders, which will be explored further in Part III.

1 12.1 The Causal Question (pp. 157-159)

We return to the NHEFS study to estimate the average causal effect of quitting smoking on weight gain.

1.1 Research Question

Population: 1,566 cigarette smokers from NHEFS who had a baseline visit and were seen again approximately 10 years later.

Treatment: $A = 1$ if quit smoking between visits, $A = 0$ if continued smoking

Outcome: Y = weight change in kg between visits

Causal estimand:

$$E[Y^{a=1}] - E[Y^{a=0}]$$

The average treatment effect of smoking cessation on weight gain.

1.2 Measured Confounders

We have measured baseline covariates L that may confound the relationship:

- Sex
- Age

- Race
- Education
- Intensity and duration of smoking
- Physical activity
- Weight and weight change in past year
- Other lifestyle and health factors

Assumption: Conditional exchangeability given L :

$$Y^a \perp\!\!\!\perp A \mid L$$

Why we need adjustment: People who quit smoking differ systematically from those who continue. For example, those who quit may be more health-conscious, may have experienced health scares, or may differ in baseline weight. These factors could also affect weight gain independently of smoking cessation.

The assumption of conditional exchangeability says that within levels of the measured covariates L , treatment assignment is as good as random with respect to the potential outcomes.

2 12.2 Estimating IP Weights (pp. 159-163)

The core idea of IP weighting is to create a pseudo-population by weighting each individual by the inverse of their probability of receiving the treatment they actually received.

2.1 IP Weights Definition

Definition 2.1 (Inverse Probability Weights). For individual i , the **IP weight** is:

$$W_i^A = \frac{1}{f(A_i \mid L_i)}$$

where $f(A_i \mid L_i) = \Pr[A = A_i \mid L = L_i]$ is the **propensity score** - the probability of receiving the treatment actually received, given confounders.

For a dichotomous treatment:

- If $A_i = 1$: $W_i^A = \frac{1}{\Pr[A=1|L_i]}$
- If $A_i = 0$: $W_i^A = \frac{1}{\Pr[A=0|L_i]} = \frac{1}{1-\Pr[A=1|L_i]}$

2.2 Estimating Propensity Scores

Step 1: Fit a model for $\Pr[A = 1 | L]$

For dichotomous treatment, use logistic regression:

$$\text{logit } \Pr[A = 1 | L] = \beta_0 + \beta_1 L_1 + \beta_2 L_2 + \dots + \beta_p L_p$$

Step 2: Predict $\hat{f}(A_i | L_i)$ for each individual

- For $A_i = 1$: $\hat{f}(1 | L_i) = \widehat{\Pr}[A = 1 | L_i]$
- For $A_i = 0$: $\hat{f}(0 | L_i) = 1 - \widehat{\Pr}[A = 1 | L_i]$

Step 3: Calculate IP weights

$$\widehat{W}_i^A = \frac{1}{\hat{f}(A_i | L_i)}$$

Why this works: In the pseudo-population created by IP weighting, each individual is weighted by how “surprising” their treatment assignment was given their covariates.

- Individuals with $\Pr[A | L]$ close to 1 receive small weights (their treatment was expected)
- Individuals with $\Pr[A | L]$ close to 0 receive large weights (their treatment was unexpected)

This reweighting creates a pseudo-population in which treatment is independent of L .

2.3 Example: NHEFS Data

In the NHEFS study:

Propensity score model: Logistic regression including sex, age, race, education, smoking intensity, smoking duration, exercise, weight, etc.

Typical weights: - Median weight: approximately 1.0 - Range: 0.3 to 16.7 - Mean: approximately 1.0 (by construction in simple settings)

Some individuals have very large weights, indicating their treatment was unusual given their covariates.

3 12.3 Stabilized IP Weights (pp. 163-165)

Standard IP weights can have extreme values, leading to unstable estimates. **Stabilized weights** reduce variability.

Definition 3.1 (Stabilized IP Weights).

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

where $f(A) = \Pr[A]$ is the marginal probability of treatment.

For dichotomous A :

- If $A = 1$: $SW^A = \frac{\Pr[A=1]}{\Pr[A=1|L]}$
- If $A = 0$: $SW^A = \frac{\Pr[A=0]}{\Pr[A=0|L]} = \frac{1-\Pr[A=1]}{1-\Pr[A=1|L]}$

3.1 Properties of Stabilized Weights

Advantages: 1. Mean is exactly 1.0 2. Smaller range than unstandardized weights 3. More stable variance estimates 4. Still create pseudo-population with $A \perp\!\!\!\perp L$

Estimation: - Numerator: Fit model for $\Pr[A = 1]$ (intercept-only logistic regression) - Denominator: Same as unstabilized weights

3.2 Example: NHEFS Data

Stabilized weights: - Median: approximately 1.0 - Range: 0.3 to 13.3 (compared to 0.3 to 16.7 for unstabilized) - Mean: exactly 1.0

Intuition: Stabilized weights still create a pseudo-population where treatment is independent of confounders, but they “shrink” extreme weights toward 1.0. The numerator ensures that the marginal distribution of treatment is preserved, while the denominator still removes the association between treatment and confounders.

4 12.4 Marginal Structural Models (pp. 165-169)

IP weighting is used to fit **marginal structural models** - models for the marginal distribution of the potential outcomes.

Definition 4.1 (Marginal Structural Model). A **marginal structural model (MSM)** is a model for the marginal mean of the potential outcome Y^a as a function of treatment a (and possibly other variables):

$$E[Y^a] = \beta_0 + \beta_1 a$$

For dichotomous A , parameter β_1 equals the average causal effect:

$$\beta_1 = E[Y^{a=1}] - E[Y^{a=0}]$$

4.1 Fitting Marginal Structural Models

Procedure:

1. Estimate IP weights $\hat{S}W^A$ for all individuals
2. Fit a weighted regression model:
 - Outcome: Y
 - Predictor: A
 - Weights: $\hat{S}W^A$
3. The coefficient of A estimates the marginal causal effect

Important: The model is fit using the observed data (A, Y) , but weighted by IP weights. This approximates what we would see if we fit an unweighted model in the pseudo-population.

4.2 Example: NHEFS Study

MSM:

$$E[Y^a] = \beta_0 + \beta_1 a$$

Weighted linear regression:

```
# Pseudo-code
fit <- lm(weight_change ~ quit_smoking,
          weights = stabilized_weights,
          data = nhefs)
```

Results: - $\hat{\beta}_1 \approx 3.4$ kg (95% CI: 2.4, 4.5) - Interpretation: Quitting smoking causes an average weight gain of 3.4 kg

Comparison to stratification:

- IP weighting estimates the marginal effect: $E[Y^{a=1}] - E[Y^{a=0}]$
- Stratification + standardization can estimate the same quantity
- IP weighting and standardization are mathematically equivalent under certain conditions
- IP weighting is often more convenient, especially for continuous confounders or time-varying treatments

5 12.5 Effect Modification and Marginal Structural Models (pp. 169-171)

MSMs can model effect modification by including interactions with baseline covariates.

5.1 MSM with Effect Modification

Definition 5.1 (MSM with Effect Modifier). To assess effect modification by variable V :

$$E[Y^a | V] = \beta_0 + \beta_1 a + \beta_2 V + \beta_3 a \times V$$

where β_3 quantifies effect modification:

- If $\beta_3 \neq 0$, the causal effect differs across levels of V
- The causal effect at $V = v$ is $\beta_1 + \beta_3 v$

5.2 Fitting MSMs with Effect Modification

Procedure:

1. Estimate IP weights as before (adjustment set includes V and other confounders)
2. Fit weighted regression including A , V , and $A \times V$
3. Interpret β_3 as the change in causal effect per unit increase in V

5.3 Example: Effect Modification by Sex

MSM:

$$E[Y^a | \text{Sex}] = \beta_0 + \beta_1 a + \beta_2 \text{Sex} + \beta_3 a \times \text{Sex}$$

Results (hypothetical): - $\hat{\beta}_1 = 2.5$ kg (effect in men) - $\hat{\beta}_3 = 1.8$ kg (additional effect in women) - Effect in women: $2.5 + 1.8 = 4.3$ kg

Important distinction:

- The MSM $E[Y^a | V]$ models the mean of the potential outcome within levels of V
- This is a **conditional causal effect** (conditional on V)
- V appears in the model, but we still adjust for confounders L through IP weighting
- V must not be affected by treatment (must be a pre-treatment variable)

6 12.6 Censoring and Missing Data (pp. 171-173)

IP weighting can also handle censoring and missing data under appropriate assumptions.

6.1 IP Weights for Censoring

Let $C = 1$ if censored (data missing), $C = 0$ if uncensored (data observed).

IP weight for censoring:

$$W^C = \frac{1}{\Pr[C = 0 \mid A, L]}$$

These weights create a pseudo-population of only uncensored individuals.

6.2 Joint Weights for Treatment and Censoring

When we have both confounding and censoring:

$$W^{A,C} = W^A \times W^C = \frac{1}{\Pr[A \mid L]} \times \frac{1}{\Pr[C = 0 \mid A, L]}$$

Stabilized version:

$$SW^{A,C} = \frac{\Pr[A]}{\Pr[A \mid L]} \times \frac{\Pr[C = 0 \mid A]}{\Pr[C = 0 \mid A, L]}$$

6.3 Example: NHEFS with Loss to Follow-up

Setting: Some individuals lost to follow-up by the second visit

Assumption: Censoring is independent of potential outcomes given (A, L) :

$$C \perp\!\!\!\perp Y^a \mid A, L$$

Procedure:

1. Fit model for $\Pr[C = 0 \mid A, L]$ among all individuals
2. Calculate censoring weights for uncensored individuals
3. Multiply by treatment weights: $SW^{A,C}$
4. Fit weighted MSM using uncensored data only

Assumptions required:

1. **Treatment:** $Y^a \perp\!\!\!\perp A \mid L$ (conditional exchangeability)
2. **Censoring:** $C \perp\!\!\!\perp Y^a \mid A, L$ (missing at random)
3. **Positivity:** $\Pr[A = a \mid L] > 0$ and $\Pr[C = 0 \mid A, L] > 0$ for all a, L

Missing at random (MAR) means censoring may depend on observed variables but not on the (unobserved) outcome. This is weaker than missing completely at random (MCAR).

7 12.7 A Likelihood Approach (pp. 173-174)

IP weighting can be viewed through the lens of likelihood theory.

7.1 Weighted Likelihood

The IP weighted estimator solves the **weighted estimating equations**:

$$\sum_{i=1}^n W_i^A \times \frac{\partial \log f(Y_i \mid A_i; \beta)}{\partial \beta} = 0$$

This is equivalent to maximizing a **weighted likelihood**:

$$L_W(\beta) = \prod_{i=1}^n [f(Y_i | A_i; \beta)]^{W_i^A}$$

7.2 Connection to Maximum Likelihood

Without confounding: Standard MLE of β in model $E[Y | A] = g(A; \beta)$

With confounding: IP weighted MLE of β in MSM $E[Y^a] = g(a; \beta)$

The IP weights “adjust” the likelihood to account for confounding.

Theoretical advantages:

1. IP weighted estimators have well-studied asymptotic properties
2. Can derive standard errors from weighted likelihood theory
3. Provides a unified framework for various causal inference methods

Practical note: Most software provides sandwich (robust) standard errors for IP weighted analyses, which account for: - Uncertainty in IP weight estimation - Model misspecification - Clustering

8 Summary

Key concepts introduced:

1. **Inverse probability weighting:** Create a pseudo-population where treatment is independent of confounders
2. **Stabilized weights:** Reduce variability while maintaining confounding control
3. **Marginal structural models:** Models for the mean potential outcome as a function of treatment
4. **Effect modification in MSMs:** Include interactions to assess heterogeneity
5. **Censoring weights:** Handle missing data under MAR assumption
6. **Weighted likelihood:** Theoretical foundation for IP weighting

Advantages of IP weighting: - Natural for marginal effects - Handles continuous confounders easily - Extends naturally to time-varying treatments (Part III) - Can combine treatment and censoring weights

Limitations: - Requires correct specification of treatment model - Can be unstable with extreme weights - Positivity violations lead to extreme weights - Efficiency loss compared to outcome modeling (when that model is correct)

Looking ahead: Part III will extend these ideas to time-varying treatments and confounders, where IP weighting and MSMs are particularly powerful. We’ll see that handling time-varying confounding affected by prior treatment is impossible with standard regression, but natural with IP weighting.

Hernán, Miguel A, and James M Robins. 2020. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC. <https://miguelhernan.org/whatifbook>.