

# PLSC 30600

Week 5: More estimation. Overlap and positivity.

Molly Offer-Westort

Department of Political Science,  
University of Chicago

Winter 2026

## Science table: strong ignorability and the propensity score

$i$	$X_{[1]i}$	$X_{[2]i}$	$p_D(\mathbf{X}_i)$	$Y_i(0)$	$Y_i(1)$	$D_i$	$Y_i$
1	A	0	?	0	?	0	0
2	A	0	?	?	1	1	1
3	B	0	?	1	?	0	1
4	B	0	?	?	1	1	1
5	A	1	?	0	?	0	0
6	A	1	?	?	1	1	1
7	B	1	?	1	?	0	1
8	B	1	?	?	0	1	0
9	A	0	?	1	?	0	0
10	B	1	?	?	1	1	1

## Code: science table data

```
> df <- data.frame(  
+   X_1 = c("A", "A", "B", "B", "A", "A", "B", "B", "A", "B"),  
+   X_2 = c(0, 0, 0, 0, 1, 1, 1, 1, 0, 1),  
+   D    = c(0, 1, 0, 1, 0, 1, 0, 1, 0, 1),  
+   Y    = c(0, 1, 1, 1, 0, 1, 1, 0, 0, 1)  
+ )  
> df$X_1 <- factor(df$X_1)  
> # propensity model (same as week 4)  
> X <- model.matrix(~ X_1 + X_2, data = df)  
> D <- df$D  
> neg_loglik <- function(beta, X, D) {  
+   eta <- as.vector(X %*% beta)  
+   p <- pnorm(eta)  
+   -sum(D * log(p) + (1 - D) * log(1 - p))  
+ }  
> fit <- optim(par = rep(0, ncol(X)),  
+             fn = neg_loglik,  
+             X = X,  
+             D = D)  
> df$p_hat <- pnorm(as.vector(X %*% fit$par))
```

# IPW vs. stabilized IPW

- IPW ATE estimator:

$$\hat{E}_{IPW}[\tau_i] = \frac{1}{n} \sum_{i=1}^n \left( \frac{Y_i D_i}{\hat{p}_D(\mathbf{X}_i)} - \frac{Y_i(1 - D_i)}{1 - \hat{p}_D(\mathbf{X}_i)} \right).$$

- Stabilized IPW (Hájek / ratio estimator):

$$\hat{E}_{SIPW}[\tau_i] = \frac{\frac{1}{n} \sum_{i=1}^n \frac{Y_i D_i}{\hat{p}_D(\mathbf{X}_i)}}{\frac{1}{n} \sum_{i=1}^n \frac{D_i}{\hat{p}_D(\mathbf{X}_i)}} - \frac{\frac{1}{n} \sum_{i=1}^n \frac{Y_i(1 - D_i)}{1 - \hat{p}_D(\mathbf{X}_i)}}{\frac{1}{n} \sum_{i=1}^n \frac{1 - D_i}{1 - \hat{p}_D(\mathbf{X}_i)}}.$$

# Why stabilize?

- The denominators re-normalize when the sample has unusually many large or small weights.
- With the true propensity score,

$$E \left[ \frac{D_i}{p_D(\mathbf{X}_i)} \right] = 1, \quad E \left[ \frac{1 - D_i}{1 - p_D(\mathbf{X}_i)} \right] = 1,$$

so stabilization targets the same estimand.

- In practice, stabilized IPW reduces variance relative to raw IPW.

## Code: stabilized IPW

```
> # IPW and stabilized IPW using the science table data frame df
> w_t <- df$D / df$p_hat
> w_c <- (1 - df$D) / (1 - df$p_hat)
> ipw_ate <- mean(w_t * df$Y - w_c * df$Y)
> sipw_ate <- (sum(w_t * df$Y) / sum(w_t)) -
+   (sum(w_c * df$Y) / sum(w_c))
> c(ipw_ate = ipw_ate, sipw_ate = sipw_ate)

      ipw_ate  sipw_ate
0.3499516 0.3499739
```

## Doubly robust theorem (Thm. 7.2.8)

- Let  $\tilde{m}_d(x)$  approximate  $E[Y_i \mid D_i = d, \mathbf{X}_i = x]$  and let  $\tilde{p}_D(x)$  approximate  $p_D(x) = \Pr[D_i = 1 \mid \mathbf{X}_i = x]$ .
- Under strong ignorability, if either:
  - $\tilde{m}_d(x)$  is correct for  $d \in \{0, 1\}$  and  $0 < \tilde{p}_D(x) < 1$ , or
  - $\tilde{p}_D(x) = p_D(x)$  for all  $x$ ,

then

$$E[\tau_i] = E \left[ \tilde{m}_1(\mathbf{X}_i) - \tilde{m}_0(\mathbf{X}_i) + \frac{D_i \{Y_i - \tilde{m}_1(\mathbf{X}_i)\}}{\tilde{p}_D(\mathbf{X}_i)} - \frac{(1 - D_i) \{Y_i - \tilde{m}_0(\mathbf{X}_i)\}}{1 - \tilde{p}_D(\mathbf{X}_i)} \right].$$

# Why it works

$$E[\tau_i] = E \left[ \tilde{m}_1(\mathbf{X}_i) - \tilde{m}_0(\mathbf{X}_i) + \frac{D_i \{Y_i - \tilde{m}_1(\mathbf{X}_i)\}}{\tilde{p}_D(\mathbf{X}_i)} - \frac{(1 - D_i) \{Y_i - \tilde{m}_0(\mathbf{X}_i)\}}{1 - \tilde{p}_D(\mathbf{X}_i)} \right]$$

- If the outcome model is correct, then

$$E[Y_i - \tilde{m}_d(\mathbf{X}_i) \mid D_i = d, \mathbf{X}_i] = 0,$$

so the weighted residual terms vanish.

- If the propensity model is correct, the IPW terms reweight the residuals so their expectation is zero even when  $\tilde{m}_d(x)$  is misspecified.
- Intuition: with correct  $p_D(\mathbf{X}_i)$ , weighting makes the residuals average over the *target*  $X$  distribution, so

$$E \left[ \frac{D_i}{p_D(\mathbf{X}_i)} (Y_i - \tilde{m}_1(\mathbf{X}_i)) \right] = E[E[Y_i(1) - \tilde{m}_1(\mathbf{X}_i) \mid \mathbf{X}_i]],$$

and the analogous control term offsets bias.

- Thus the expression for  $E[\tau_i]$  holds if *either* model is correct.



# Doubly robust estimator

$$\hat{E}_{DR}[\tau_i] = \frac{1}{n} \sum_{i=1}^n \underbrace{\hat{m}_1(\mathbf{X}_i) - \hat{m}_0(\mathbf{X}_i)}_{\text{outcome model}} + \frac{1}{n} \sum_{i=1}^n \underbrace{\frac{D_i (Y_i - \hat{m}_1(\mathbf{X}_i))}{\hat{p}_D(\mathbf{X}_i)} - \frac{(1 - D_i) (Y_i - \hat{m}_0(\mathbf{X}_i))}{1 - \hat{p}_D(\mathbf{X}_i)}}_{\text{residual correction}} + \underbrace{\frac{D_i (Y_i - \hat{m}_1(\mathbf{X}_i))}{\hat{p}_D(\mathbf{X}_i)}}_{\text{residual}}$$

- The **outcome model** can be estimated by regression or ML.
- The **propensity model** can be estimated by logit, probit, or ML.

## Code: doubly robust estimator

```
> # Outcome models
> fit_dr <- lm(Y ~ D + X_1 + X_2, data = df)
> df$m1_hat <- predict(fit_dr, newdata = transform(df, D = 1))
> df$m0_hat <- predict(fit_dr, newdata = transform(df, D = 0))
> # DR estimator (uses p_hat from propensity model)
> dr_ate <- mean(
+   df$m1_hat - df$m0_hat +
+     df$D * (df$Y - df$m1_hat) / df$p_hat -
+     (1 - df$D) * (df$Y - df$m0_hat) / (1 - df$p_hat)
+ )
> dr_ate

[1] 0.3499874
```

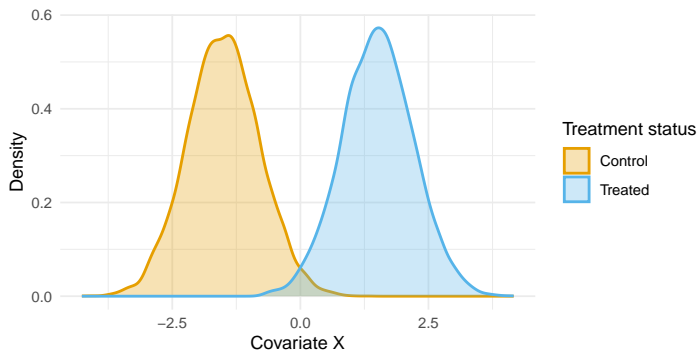
# Overlap and positivity: definitions

- **Positivity (overlap in treatment probabilities):** There exists  $\varepsilon > 0$  such that for all  $x \in \text{Supp}[\mathbf{X}_i]$ ,

$$\varepsilon < p_D(x) < 1 - \varepsilon.$$

- Interpretation: every covariate profile has nonzero probability of both treatment and control.
- **Complete population overlap:** the supports of  $\mathbf{X}_i$  in treated and control groups overlap.
- Failure of positivity  $\Rightarrow$  incomplete overlap (regions of  $x$  with only treated or only control).

## Example: failure of overlap



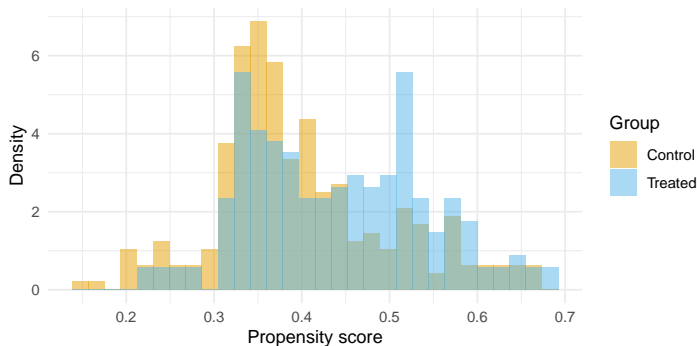
Little overlap in  $X$  implies  $p_D(x)$  near 0 or 1 for some  $x$ .

## Code: Lalonde propensity score overlap

```
> library(ggplot2)
> lalonde_url <-
+   "https://raw.githubusercontent.com/xuyiqing/lalonde/master/data/lalonde.RData"
> load(url(lalonde_url))
> covar <- c("age", "education", "black", "hispanic", "married", "nodegree",
+           "re74", "re75", "u74", "u75")
> ps_fit <- glm(treat ~ age + education + black + hispanic + married + nodegree +
+           re74 + re75 + u74 + u75,
+           data = ldw, family = binomial())
> ldw$ps <- predict(ps_fit, type = "response")
> ldw$group <- factor(ifelse(ldw$treat == 1, "Treated", "Control"),
+           levels = c("Control", "Treated"))
> ggplot(ldw, aes(x = ps, fill = group)) +
+   geom_histogram(aes(y = after_stat(density)),
+           position = "identity", alpha = 0.5, bins = 30) +
+   scale_fill_manual(values = c("#E69F00", "#56B4E9")) +
+   labs(x = "Propensity score", y = "Density", fill = "Group") +
+   theme_minimal()
```

# Figure: Lalonde propensity score overlap

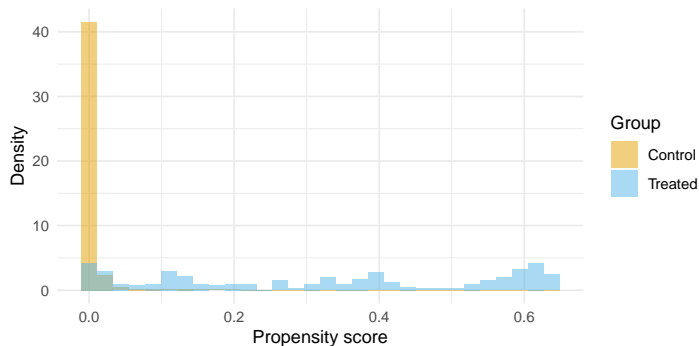
- Treated vs. control propensity score overlap in the LDW data. (Xu et al. tutorial; see References)



# Figure: CPS1 overlap (observational controls)

- Treated LDW units + CPS controls.

(Xu et al. tutorial; see References)



# Why overlap matters

- If  $p_D(x) = 0$  or  $1$  for some  $x$ , then  $Y_i(1)$  or  $Y_i(0)$  is not identified at that  $x$ .
- If  $p_D(x)$  is near  $0$  or  $1$ , weights explode and estimates become unstable.
- IPW: extreme weights  $\Rightarrow$  high variance.
- Matching: few/no close matches  $\Rightarrow$  bias or large variance.
- Regression: extrapolation where data are missing.



# Diagnostics and responses

- Diagnose overlap with propensity score plots (treated vs. control) and extreme  $\hat{p}_D(x)$ .
- Report share of units outside common support (e.g.,  $\hat{p}_D(x) < 0.05$  or  $> 0.95$ ).
- **Trimming / common support:** restrict to overlap region.
- Trimming changes the estimand (e.g., ATT or overlap population).
- Stabilized IPW reduces variance but does not fix identification when overlap fails.

# ATT as a pseudo-population

- For ATT, we reweight controls to match the treated covariate distribution.
- Define weights:

$$w_i^{ATT} = \begin{cases} 1, & D_i = 1 \\ \frac{\hat{p}_D(\mathbf{X}_i)}{1 - \hat{p}_D(\mathbf{X}_i)}, & D_i = 0 \end{cases}$$

- In the weighted pseudo-population, controls look like treated units in  $X$ .
- Then ATT is a difference in weighted means:

$$\hat{E}_{ATT}[\tau_i] = \bar{Y}_{D=1} - \frac{\sum_{i:D_i=0} w_i^{ATT} Y_i}{\sum_{i:D_i=0} w_i^{ATT}}.$$

# Science table: ATT weights

$i$	$X_{[1]i}$	$X_{[2]i}$	$\hat{p}_D(\mathbf{X}_i)$	$D_i$	$Y_i$	$w_i^{ATT}$
1	A	0	$\hat{p}_D(\mathbf{X}_i)$	0	0	$\frac{\hat{p}_D(\mathbf{X}_i)}{1-\hat{p}_D(\mathbf{X}_i)}$
2	A	0	$\hat{p}_D(\mathbf{X}_i)$	1	1	1
3	B	0	$\hat{p}_D(\mathbf{X}_i)$	0	1	$\frac{\hat{p}_D(\mathbf{X}_i)}{1-\hat{p}_D(\mathbf{X}_i)}$
4	B	0	$\hat{p}_D(\mathbf{X}_i)$	1	1	1
5	A	1	$\hat{p}_D(\mathbf{X}_i)$	0	0	$\frac{\hat{p}_D(\mathbf{X}_i)}{1-\hat{p}_D(\mathbf{X}_i)}$
6	A	1	$\hat{p}_D(\mathbf{X}_i)$	1	1	1
7	B	1	$\hat{p}_D(\mathbf{X}_i)$	0	1	$\frac{\hat{p}_D(\mathbf{X}_i)}{1-\hat{p}_D(\mathbf{X}_i)}$
8	B	1	$\hat{p}_D(\mathbf{X}_i)$	1	0	1
9	A	0	$\hat{p}_D(\mathbf{X}_i)$	0	0	$\frac{\hat{p}_D(\mathbf{X}_i)}{1-\hat{p}_D(\mathbf{X}_i)}$
10	B	1	$\hat{p}_D(\mathbf{X}_i)$	1	1	1

## Code: ATT weights

```
> # ATT weights: treated = 1, control = p_hat / (1 - p_hat)
> w_att <- ifelse(df$D == 1, 1, df$p_hat / (1 - df$p_hat))
> att_ipw <- mean(df$Y[df$D == 1]) -
+   weighted.mean(df$Y[df$D == 0], w_att[df$D == 0])
> c(att_ipw = att_ipw)

att_ipw
0.199926
```

# ATT-specific IPW formulas

- Target estimand:

$$E[\tau_i \mid D_i = 1] = E[Y_i(1) - Y_i(0) \mid D_i = 1].$$

- IPW ATT estimator (controls reweighted by  $p_D$ ):

$$= \frac{1}{n_1} \sum_{i=1}^n Y_i D_i - \frac{1}{n_1} \sum_{i=1}^n Y_i (1 - D_i) \frac{\hat{p}_D(\mathbf{X}_i)}{1 - \hat{p}_D(\mathbf{X}_i)}, \quad n_1 = \sum_{i=1}^n D_i.$$

- Stabilized (Hájek) ATT:

$$= \frac{\sum_{i=1}^n Y_i D_i}{\sum_{i=1}^n D_i} - \frac{\sum_{i=1}^n Y_i (1 - D_i) \frac{\hat{p}_D(\mathbf{X}_i)}{1 - \hat{p}_D(\mathbf{X}_i)}}{\sum_{i=1}^n (1 - D_i) \frac{\hat{p}_D(\mathbf{X}_i)}{1 - \hat{p}_D(\mathbf{X}_i)}}.$$

- Doubly robust ATT (with outcome model  $\hat{m}_d$  and propensity  $\hat{p}_D$ ):

$$\begin{aligned} & \frac{1}{n_1} \sum_{i=1}^n D_i [\hat{m}_1(\mathbf{X}_i) - \hat{m}_0(\mathbf{X}_i)] \\ = & \frac{1}{n_1} \sum_{i=1}^n D_i \{Y_i - \hat{m}_1(\mathbf{X}_i)\} - \frac{1}{n_1} \sum_{i=1}^n (1 - D_i) \frac{\hat{p}_D(\mathbf{X}_i)}{1 - \hat{p}_D(\mathbf{X}_i)} \{Y_i - \hat{m}_0(\mathbf{X}_i)\}. \end{aligned}$$

# Placebo testing

- Placebo tests probe the plausibility of ignorability by checking balance on *outcomes that should not be affected by treatment*.
- If treated and control units differ on placebo outcomes after adjustment, that is evidence against the identifying assumptions.
- Placebo tests are *not* the same as balance tests: balance checks whether the reweighting/matching worked, placebo tests probe the causal identification.

# Placebo outcomes (implementation)

- Choose an outcome measured *before* treatment (or otherwise unaffected by treatment).
- Re-estimate the propensity score without using that placebo outcome as a covariate.
- Apply the same estimator(s) (IPW, stabilized IPW, DR, etc.) to the placebo outcome.
- Large or significant placebo effects suggest lack of overlap or residual confounding.

# References I

Dehejia, R. H. and Wahba, S. (1998). Propensity score matching methods for non-experimental causal studies. Working Paper 6829, National Bureau of Economic Research.

Imbens, G. W. and Xu, Y. (2025). Comparing experimental and nonexperimental methods: What lessons have we learned four decades after lalonde (1986)? *Journal of Economic Perspectives*, 39(4):173–201.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *The American Economic Review*, pages 604–620.

Tutorial: <https://yiqingxu.org/tutorials/lalonde/>