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	$X_{[1]i}$	$X_{[2]i}$	$p_D(\boldsymbol{X}_i)$	$Y_i(0)$	$Y_i(1)$	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:

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

- Stabilized IPW (Hájek / ratio estimator):

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

Why stabilize?

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

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

  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 $\mathrm{E}\left[Y_i \mid D_i = d, \boldsymbol{X}_i = x\right]$ and let $\tilde{p}_D(x)$ approximate $p_D(x) = \Pr[D_i = 1 \mid \boldsymbol{X}_i = x]$.

- 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

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

Why it works

- If the outcome model is correct, then

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

  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.

- Thus the expression for $\mathrm{E}\left[\tau_i\right]$ holds if *either* model is correct.

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

Doubly robust estimator

- Replace expectations with sample averages and plug in $\tilde{m}_d(\cdot)$ and $\tilde{p}_D(\cdot)$:

$$\widehat{\mathrm{E}}_{DR}[\tau_i] = \frac{1}{n}\sum_{i=1}^{n}\Big[\tilde{m}_1(\boldsymbol{X}_i) - \tilde{m}_0(\boldsymbol{X}_i) + \frac{D_i\{Y_i - \tilde{m}_1(\boldsymbol{X}_i)\}}{\tilde{p}_D(\boldsymbol{X}_i)} - \frac{(1-D_i)\{Y_i - \tilde{m}_0(\boldsymbol{X}_i)\}}{1-\tilde{p}_D(\boldsymbol{X}_i)}\Big].$$

Code: doubly robust estimator (science table)

```
> # 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 $x \in \text{Supp}\,[\boldsymbol{X}_i]$,

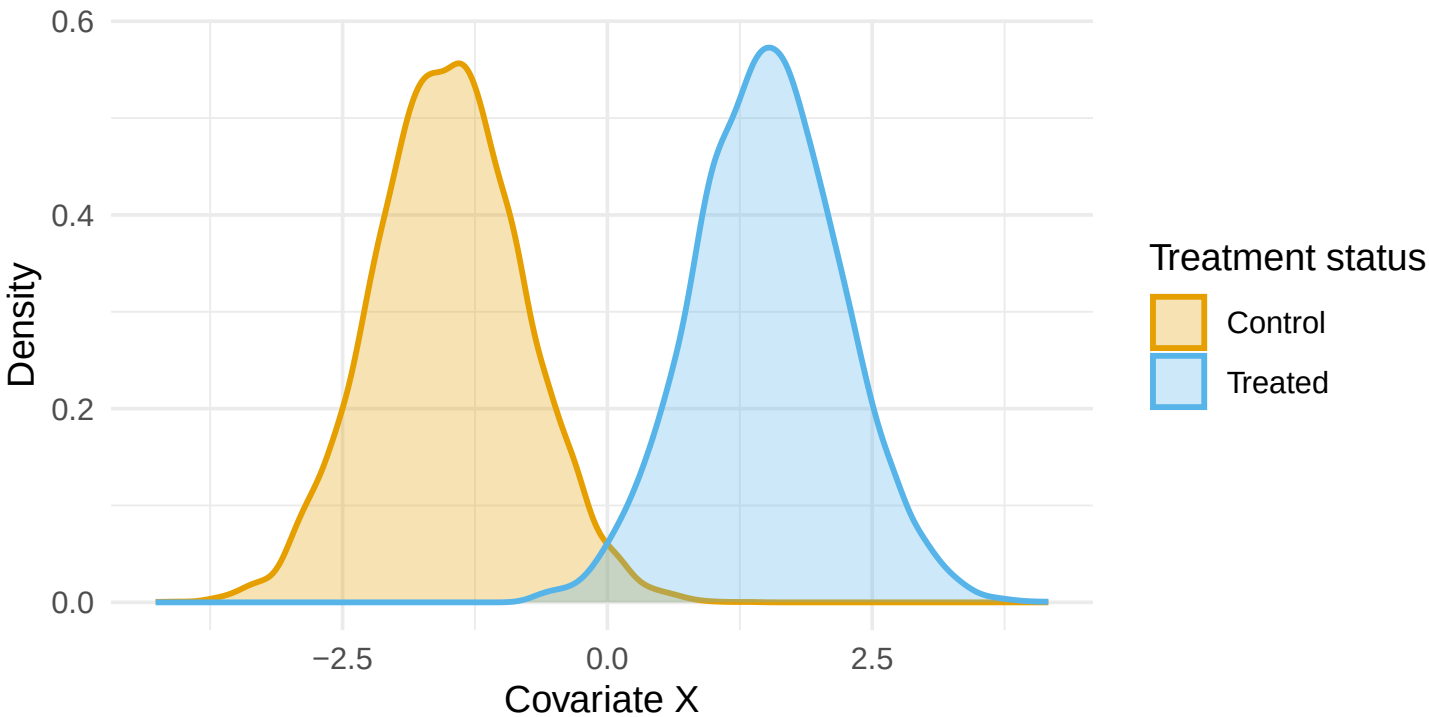
$$\varepsilon < p_D(x) < 1 - \varepsilon.$$

- Interpretation: every covariate profile has nonzero probability of both treatment and control.

- **Complete population overlap:** the supports of $\boldsymbol{X}_i$ 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)

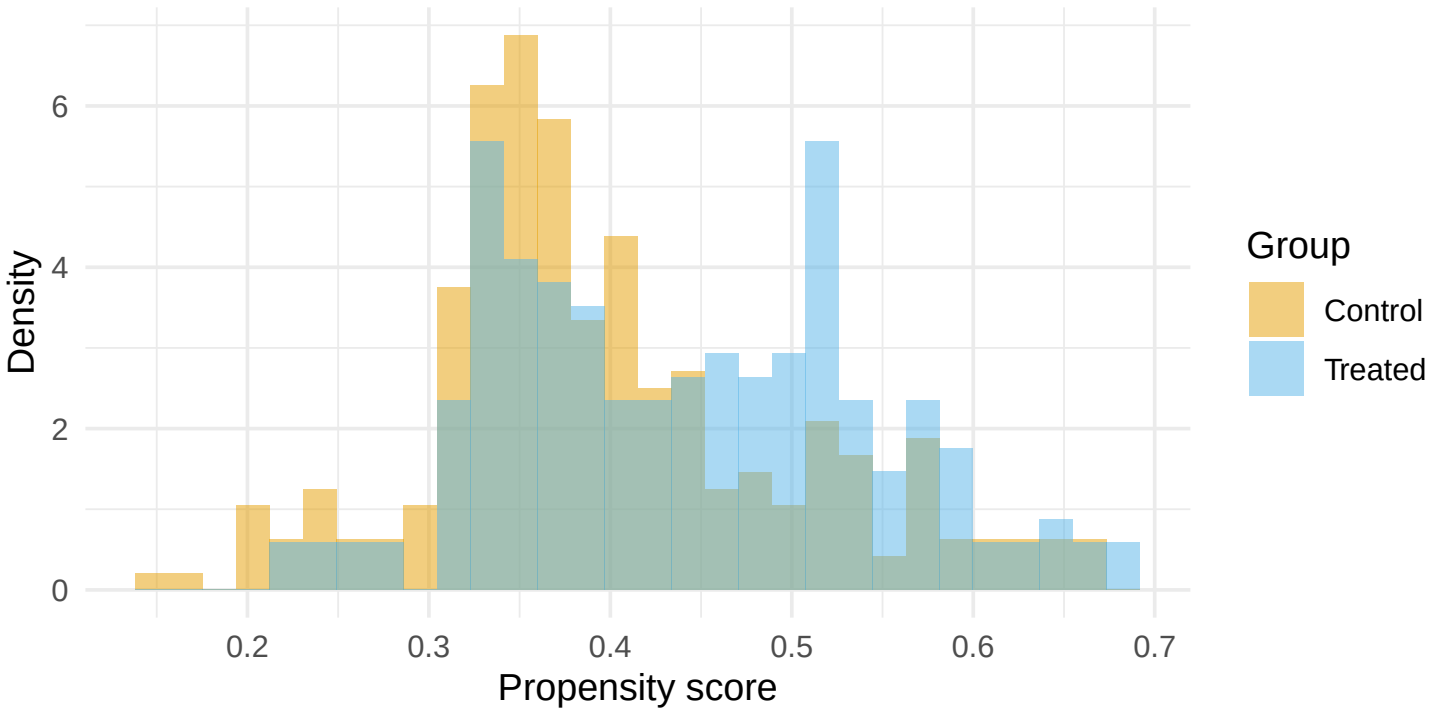
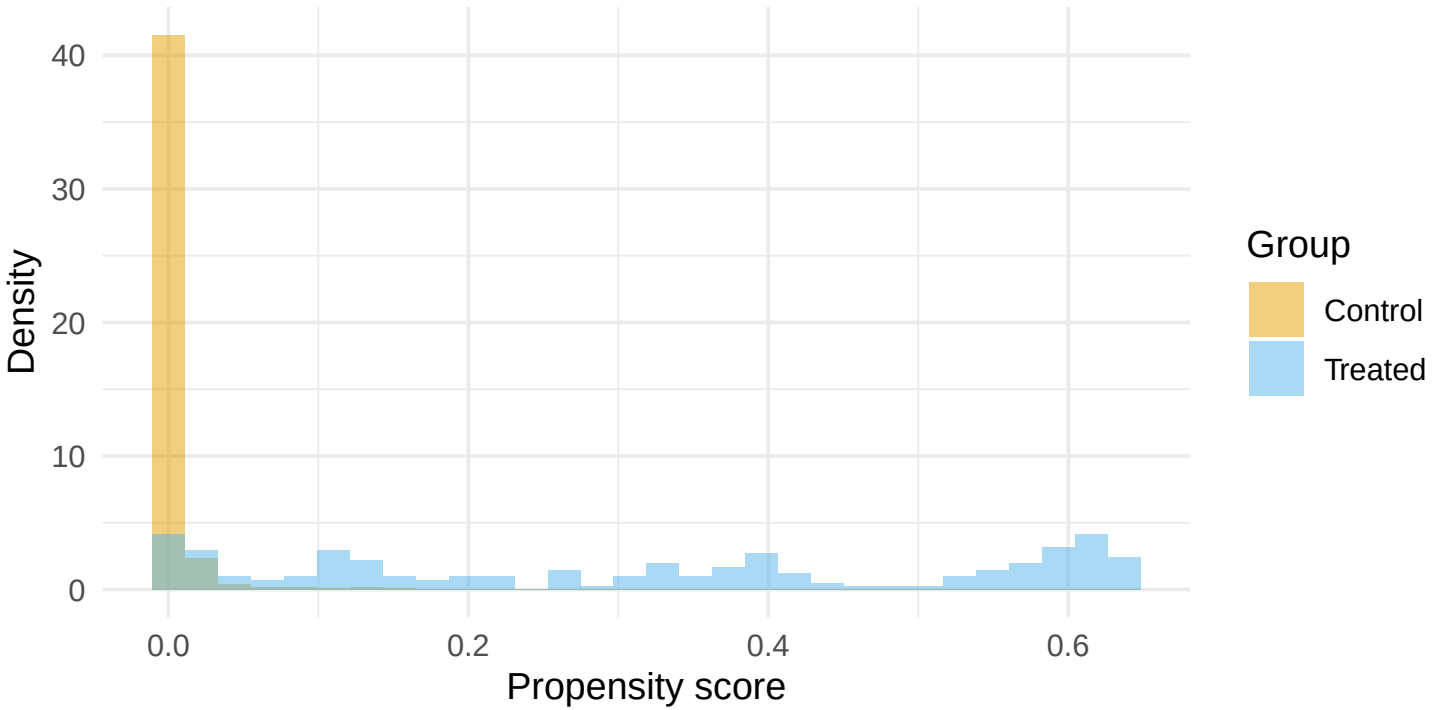


Figure: CPS1 overlap (observational controls)

- 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-specific IPW formulas

- Target estimand:

$$\mathrm{E}\left[\tau_i \mid D_i = 1\right] = \mathrm{E}\left[Y_i(1) - Y_i(0) \mid D_i = 1\right].$$

- IPW ATT estimator (controls reweighted by p_D):

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

- Stabilized (Hájek) ATT:

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

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

$$= \frac{1}{n_1}\sum_{i=1}^{n} D_i\left[\tilde{m}_1(\boldsymbol{X}_i) - \tilde{m}_0(\boldsymbol{X}_i)\right] + \frac{1}{n_1}\sum_{i=1}^{n} D_i\{Y_i - \tilde{m}_1(\boldsymbol{X}_i)\} - \frac{1}{n_1}\sum_{i=1}^{n}(1-D_i)\frac{\tilde{p}_D(\boldsymbol{X}_i)}{1-\tilde{p}_D(\boldsymbol{X}_i)}\{Y_i - \tilde{m}_0(\boldsymbol{X}_i)\}, \quad n_1 = \sum_{i=1}^{n} D_i.$$

Placebo testing (idea)

- 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

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