Causal Inference Methods and Case Studies

STAT24630

Jingshu Wang

Lecture 15

Topic: IPW, trimming, subclassification

- IPW
  - Connection with weighted least squares
- Trimming
- Subclassification
- IPW V.S. Subclassification
- Textbook chapters: Chapter 16.1, Chapter 17

Connection between IPW estimator and WLS

- Define inverse probability weights

$$\lambda_i = \frac{1}{e(X_i)^{W_i} \cdot (1 - e(X_i))^{1-W_i}} = \begin{cases} 1/(1-e(X_i)) & \text{if } W_i = 0, \\ 1/e(X_i) & \text{if } W_i = 1. \end{cases}$$

- Weighted least square with no covariate adjustments

$$(\hat{\alpha}, \hat{\tau}) = \min_{\alpha,\tau} \sum_{i=1}^{N} \lambda_i \left(Y_i^{\text{obs}} - \alpha - \tau W_i\right)^2$$

  - Solution: $\hat{\alpha} = \frac{\sum_{i=1}^{N}(1-W_i)\lambda_i Y_i^{\text{obs}}}{\sum_{i=1}^{N}(1-W_i)\lambda_i}$ and $\hat{\alpha} + \hat{\tau} = \frac{\sum_{i=1}^{N} W_i \lambda_i Y_i^{\text{obs}}}{\sum_{i=1}^{N} W_i \lambda_i}$

- Solution is the same as IPW with normalizing weights
- If we ignore the uncertainty in estimating the propensity score, we can estimate the variance of $\hat{\tau}$ from Sandwich estimator for WLS

- We can also use WLS to adjust for other pre-treatment covariates

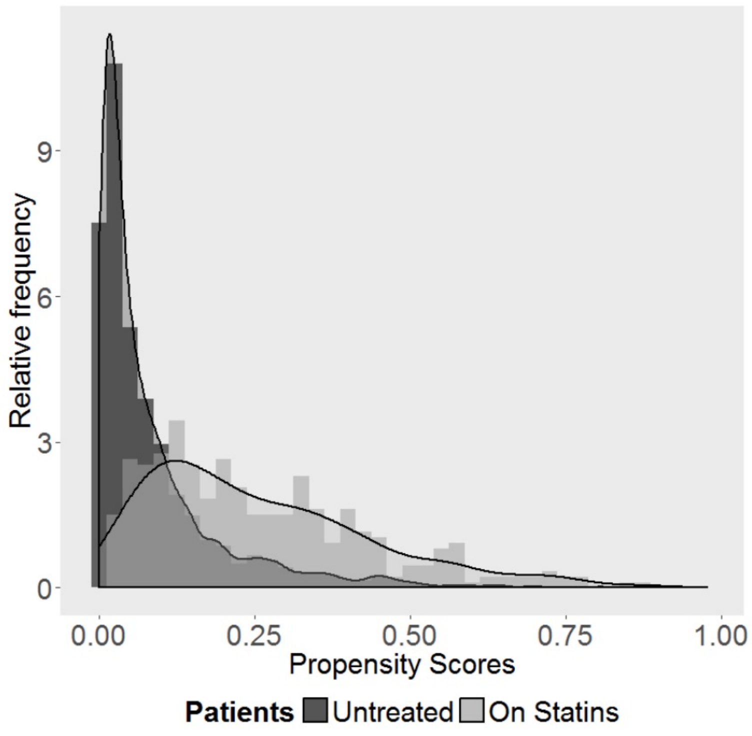
Trimming to improve overlapping

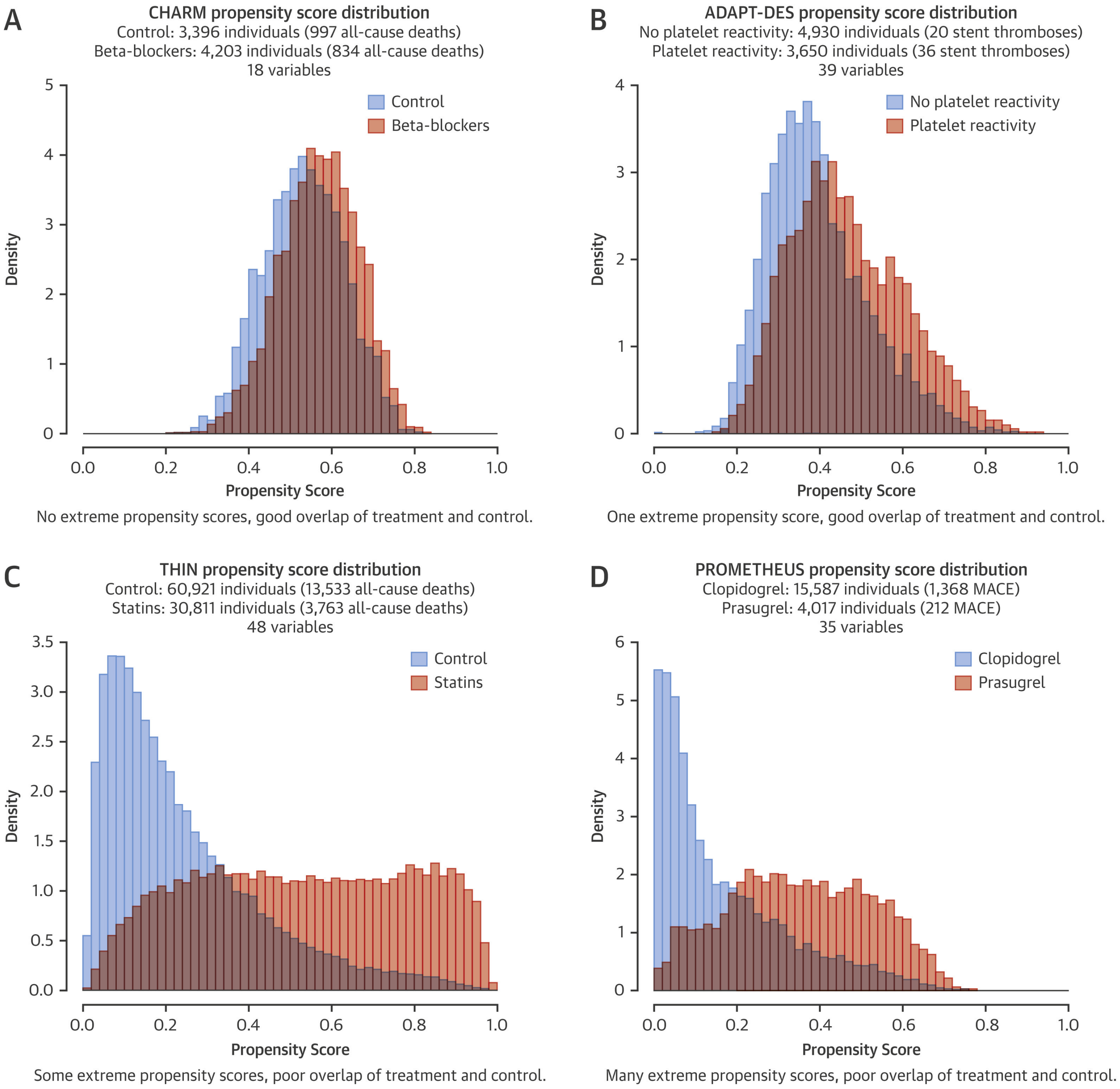
- We implicitly assume the overlap assumption: $e(\boldsymbol{x}) \neq 0 \text{ or } 1$ for any $\boldsymbol{x}$ (otherwise we won't have data to identify $\tau(\boldsymbol{x})$)

- If the estimated propensity scores are close to 0 or 1 for some units, the overlap assumption might be violated at these values' $\boldsymbol{X}_i$

- Trimming: remove units with very small or very large propensity scores
  - Remove all units with estimated propensity scores in the intervals $[0, \alpha_1]$ or $[1-\alpha_2, 1]$
  - $\alpha_1 = \alpha_2 = 0.05$ or 0.1 (ad-hoc)
  - Optimal α_1 and α_2 for trimming (Chapter 16)
  - You may refit the propensity score model after trimming

- Trimming also removes individuals with extremely large weights





Elze, Markus C., et al. "Comparison of propensity score methods and covariate adjustment: evaluation in 4 cardiovascular studies." *Journal of the American College of Cardiology* 69.3 (2017): 345-357.

Subclassification on the estimated PS

- Also called blocking or stratification

$$B_i(j) = \begin{cases} 1 & \text{if } b_{j-1} \leq \hat{e}(X_i) < b_j, \\ 0 & \text{otherwise,} \end{cases}$$

- Stratify individuals into J blocks based on the estimated propensity score

- How to find the boundary points? General guidelines
    - $\max_{j=1,\cdots,J} |b_j - b_{j-1}|$ relatively small
    - There are not too few controls/treated units (say 1 or 2) in each strata/block
    - Covariate balancing within each strata is good

Sequential block splitting

- Introduced in Lecture 12

- Start with a single block $J = 1$ with $b_0 = \underline{e}_t$ and $b_1 = \bar{e}_c$
  - For each of the current blocks, we assess whether we need to further split it into two
    - For block j, calculate the two-sample test statistics (assume equal variance)

$$t_\ell(j) = \frac{\overline{\ell}_t(j) - \overline{\ell}_c(j)}{\sqrt{s_\ell^2(j) \cdot (1/N_c(j) + 1/N_t(j))}}$$

$s_l^2(j)$: pooled sample variance in block j

    - Need to split Block j into two blocks if $|t_j| > t_{\max} = 1.96$
    - Define the two sub-blocks: find the median of $\hat{e}(\boldsymbol{X}_i)$ within block j as b_j'
      - Sub-block 1: all units with $\hat{e}(\boldsymbol{X}_i) < b_j'$; sub-block 2: all units with $\hat{e}(\boldsymbol{X}_i) \geq b_j'$
    - Stop if
      - The block does not need to split $|t_j| \leq t_{\max}$

      or
      - has a small enough size $\min(N_c(j), N_t(j)) < N_{min,1} = 3$ or number of total units of a new stratum $< K + 2$ (K is the number of covariates possibly used in regression adjustment)

The Imbens-Rubin-Sacerdote lottery data

[Estimating the effect of unearned income on labor earnings, savings, and consumption: Evidence from a survey of lottery players. *American economic review*, 2001]

- Goal: Estimate magnitude of lottery prizes (unearned income) on economic behavior, including labor supply, consumption and savings

- Data collection:
  - “Winners”: individuals who had played and won large sums of money in the Massachsetts lottery
  - “Losers”: individuals who played the lottery and had won only small prizes
  - Constructing a comparison group of lottery players who did not win anything was not feasible as the Lottery Commission did not have contact information of such individuals

- Surveys are sent to these individuals with financial incentives

- We analyze a subset of $N_t = 259$ and $N_c = 237$ individuals with complete answers

- We use the model forward selection procedure to estimate the propensity scores

The Imbens-Rubin-Sacerdote lottery data

TABLE 1—RESPONSE RATES BY MAILING

Mailing	Date	Sent		Responses		Response rates		
		Winners	Nonwinners	Winners	Nonwinners	Winners	Nonwinners	Total
Pilot	July '95	50	50	17	25	0.34	0.50	0.42
Main	July '96	752	637	272	262	0.36	0.41	0.38
Follow-up ($50 check)	Sept. '96	248	248	39	40	0.16	0.16	0.16
Follow-up ($10 cash, $40 check)	Sept. '96	49	49	11	12	0.22	0.24	0.23
Total		802	687	339	339	0.42	0.49	0.46

The Imbens-Rubin-Sacerdote lottery data

Table 17.1. ***Normalized Differences in Covariates after Subclassification for the IRS Lottery Data***

Variable	Full Sample		Trimmed Sample			
	One Block	Horvitz-Thompson	One Block	Two Blocks	Five Blocks	Horvitz-Thompson
Year Won	−0.26	0.10	−0.06	−0.03	0.07	0.07
# Tickets	0.91	0.10	0.51	0.17	0.07	−0.04
Age	−0.50	−0.30	−0.09	−0.03	0.05	0.05
Male	−0.19	0.09	−0.11	−0.10	−0.14	−0.13
Education	−0.70	0.48	−0.51	−0.18	−0.10	−0.01
Work Then	0.09	0.05	0.03	0.03	0.01	0.00
Earn Year -6	−0.32	0.01	−0.18	−0.10	−0.03	0.06
Earn Year -5	−0.28	0.01	−0.19	−0.07	−0.00	0.09
Earn Year -4	−0.29	−0.01	−0.23	−0.09	−0.01	0.06
Earn Year -3	−0.26	0.05	−0.18	−0.03	0.03	0.10
Earn Year -2	−0.31	0.06	−0.19	−0.03	0.01	0.09
Earn Year -1	−0.23	0.11	−0.17	−0.01	0.00	0.06
Pos Earn Year -6	0.03	0.16	−0.00	−0.09	−0.09	−0.01
Pos Earn Year -5	0.14	−0.14	0.10	0.01	−0.01	0.06
Pos Earn Year -4	0.10	−0.19	0.06	−0.00	−0.01	0.03
Pos Earn Year -3	0.13	−0.17	0.03	−0.04	−0.05	−0.00
Pos Earn Year -2	0.14	−0.17	0.06	0.00	−0.04	0.01
Pos Earn Year -1	0.10	0.17	−0.01	−0.04	−0.07	−0.01

- Trimming:
  results from optimal trimming
  only keep individuals whose $\hat{e}(\boldsymbol{X}_i) \in [0.0891, 0.9109]$
- Horvitz-Thompson: IPW with normalized weights
- One Block: all individuals
- Two Blocks / Five Blocks: subclassification
  (shown later)

The subclassification estimator

- Treat the data after subclassification as from a stratified randomized experiment
  - Neyman's repeated sampling approach
    1. Apply Neyman's analysis to each stratum / block

$$\hat{\tau}^{\text{dif}}(j) = \overline{Y}_{\text{t}}^{\text{obs}}(j) - \overline{Y}_{\text{c}}^{\text{obs}}(j), \quad \text{and} \quad \hat{\mathbb{V}}^{\text{neyman}}(j) = \frac{s_{\text{c}}(j)^2}{N_{\text{c}}(j)} + \frac{s_{\text{t}}(j)^2}{N_{\text{t}}(j)}$$

    2. Aggregate block-specific estimates and variances

$$\hat{\tau}^{\text{strat}} = \sum_j \frac{N(j)}{N} \hat{\tau}^{\text{dif}}(j), \qquad \hat{\mathbb{V}}(\hat{\tau}^{\text{strat}}) = \sum_j \left(\frac{N(j)}{N}\right)^2 \hat{\mathbb{V}}^{\text{neyman}}(j)$$

  - Regression adjustment
    1. Run separate linear regressions within each stratum
    2. Average regression estimates across strata

$$\hat{\tau}^{\text{reg}} = \sum_j \frac{N(j)}{N} \hat{\tau}^{\text{reg}}(j), \qquad \hat{\mathbb{V}}(\hat{\tau}^{\text{reg}}) = \sum_j \left(\frac{N(j)}{N}\right)^2 \hat{\mathbb{V}}^{\text{reg}}(j)$$

Results on the lottery data

- Set $K = 18$, so each new stratum needs to have at least 20 total units
- Sequential splitting results in 5 blocks (p-scores are after refitting the pscore model)

Table 17.4. ***Final Subclassification for the IRS Lottery Data***

Subclass	Min P-Score	Max P-Score	# Controls	# Treated	t-Stat
1	0.03	0.24	67	13	−0.1
2	0.24	0.32	32	8	0.9
3	0.32	0.44	24	17	1.7
4	0.44	0.69	34	47	2.0
5	0.69	0.99	15	66	1.6

- Comparison with using 2 blocks

Subclass	Min P-Score	Max P-Score	# Controls	# Treated	t-Stat
1	0.03	0.44	123	38	2.8
2	0.44	0.99	49	113	3.8

Results on the lottery data

- Estimates within each block

Covariates	Block 1 ($N = 80$)		Block 2 ($N = 40$)		Block 3 ($N = 41$)		Block 4 ($N = 81$)		Block 5 ($N = 81$)	
	Est	($\widehat{\text{s. e.}}$)	Est	($\widehat{\text{s. e.}}$)	Est	($\widehat{\text{s. e.}}$)	Est	($\widehat{\text{s. e.}}$)	Est	($\widehat{\text{s. e.}}$)
No covariates										
Intercept	20.02	(2.25)	12.70	(2.67)	15.59	(3.07)	19.69	(2.76)	12.75	(3.26)
Treatment	−10.82	(4.70)	2.07	(5.10)	−1.17	(4.97)	−9.43	(3.23)	−2.89	(3.59)
Limited covariates										
Intercept	−20.04	(10.66)	4.47	(9.80)	−9.91	(10.87)	−8.65	(5.58)	−6.70	(5.21)
Treatment	−6.21	(4.01)	−6.51	(3.86)	−4.81	(3.87)	−5.88	(1.82)	−2.56	(2.39)
`# Tickets`	−3.48	(1.39)	1.17	(1.26)	1.85	(1.24)	−0.48	(0.34)	−0.20	(0.37)
`Education`	2.03	(0.87)	−0.37	(0.81)	0.48	(0.93)	1.17	(0.49)	0.59	(0.42)
`Work Then`	−2.66	(2.96)	−0.51	(1.84)	5.98	(4.35)	1.16	(2.18)	5.30	(2.52)
`Earn Year -1`	0.84	(0.06)	0.83	(0.09)	0.60	(0.15)	0.76	(0.07)	0.62	(0.10)

Results on the lottery data

- Estimated ATE

Covariates	Full Sample 1 Block		Trimmed Sample 1 Block		Trimmed Sample 2 Blocks		Trimmed Sample 5 Blocks	
	Est	$(\widehat{\text{s. e.}})$	Est	$(\widehat{\text{s. e.}})$	Est	$(\widehat{\text{s. e.}})$	Est	$(\widehat{\text{s. e.}})$
None	−6.2	(1.4)	−6.6	(1.7)	−6.0	(1.9)	−5.7	(2.0)
`# Tickets, Education, Work Then, Earn Year-1`	−2.8	(0.9)	−4.0	(1.1)	−5.6	(1.2)	−5.1	(1.2)
All	−5.1	(1.0)	−5.3	(1.1)	−6.4	(1.1)	−5.7	(1.1)

Subclassification V.S. IPW estimators

- Subclassification estimator can be treated as a weighting estimator

$$\hat{\tau}^{\text{strat}} = \frac{1}{N}\sum_{i=1}^{N} W_i \cdot Y_i^{\text{obs}} \cdot \lambda_i^{\text{strat}} - \frac{1}{N}\sum_{i=1}^{N}(1 - W_i) \cdot Y_i^{\text{obs}} \cdot \lambda_i^{\text{strat}},$$

where the weights λ_i^{strat} satisfy

$$\lambda_i^{\text{strat}} = \sum_{j=1}^{J} B_i(j) \cdot \left(\frac{1 - W_i}{N_{\text{c}}(j)/N(j)} + \frac{W_i}{N_{\text{t}}(j)/N(j)} \right)$$

$$= \begin{cases} \sum_{j=1}^{J} B_i(j) \cdot \frac{N(j)}{N_{\text{c}}(j)} & \text{if } W_i = 0, \\ \sum_{j=1}^{J} B_i(j) \cdot \frac{N(j)}{N_{\text{t}}(j)} & \text{if } W_i = 1. \end{cases}$$

- Instead of using the eps $\hat{e}(\boldsymbol{X}_i)$ to obtain weights, subclassification estimator estimates the propensity scores as the block proportions (averaging $\hat{e}(\boldsymbol{X}_i)$ within subclasses)

$$\tilde{e}(X_i) = \sum_{j=1}^{J} B_i(j) \cdot \frac{N_{\text{t}}(j)}{N(j)}$$

Subclassification V.S. IPW estimators

- If there are many blocks, then the dispersion within each stratum is limited, two estimators are similar
- The weights will be different only if, in at least some blocks, there is substantial variation in the propensity score, which is most likely to happen in blocks with propensity score values close to zero and one.
- Smoothing the weights by averaging them within blocks, as the subclassification estimator does, may remove some of the biases introduced by the estimation of propensity scores (avoids extreme weights).
- Subclassification is more robust to model mis-specification.
- Subclassification as a coarsening method is more ad-hoc.

Subclassification V.S. IPW estimators

- On the lottery data, summary statistics of the weights

	Full Sample		Trimmed Sample	
	Horvitz-Thompson	Subclass	Horvitz-Thompson	Subclass
Minimum	0.92	1.06	1.00	1.19
Maximum	79.79	17.71	18.18	6.15
Standard deviation	4.20	2.63	1.69	1.35

- Uncertainty and uncertainty on the lottery data

	Full Sample		Trimmed Sample	
	Horvitz-Thompson	Subclass	Horvitz-Thompson	Subclass
Bias	4.34	2.68	1.29	0.30
Variance	2.59^2	0.83^2	1.29^2	1.15^2
Bias^2+Variance	5.06^2	2.81^2	1.83^2	1.19^2