Chapter 17. Subclassification on the Propensity Score

Yeonho Jung

August 16, 2022

Seoul National University

Contents

17.1 Introduction

Subclassification : Blocking or Stratification

1. **Propensity Score**(PS) : A class of functions of the covariates
2. Estimating casual effects that relies on **Subclassification**
    - The sample is partitioned into subclasses based on PS
    - Within the Subclasses, the estimated PS are almost constant
    - Estimate causal effect, as if assignment is completely at random within subclass
3. **Implementation of Subclassification**
    - The choice of the number of blocks and boundary values
4. **The bias reduction**
    - Because blocking cannot eliminate all biases
    - Regression or model-based adjustments can improve the precision of estimators for causal effects

17.2 Lottery Data(1/2)

Data : Full-Sample and Trimmed Sample

i . (Table 17.1) Normalized Differences

- Designed to improve the overlap by dropping units

ii . (Table 17.2) Dropping units with PS outside the interval

- Small/Large estimated PS-values are dropped
- Trimming improves the covariate balance
- Trimmed sample with 323 units out of 496 full samples

iii . (Table 17.3) Estimates of PS in trimmed sample

- Selecting linear and second-order terms(In Ch.13)
- The estimated PS : for the purpose of subclassification
- Fewer terms for the specification of the PS **after trimming**

Data : Full-Sample and Trimmed Sample

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

Table 17.2. ***Number of Units within Selected Subsamples Defined by the Estimated Propensity Score for the IRS Lottery Data***

	Low	Middle	High	All
	$\hat{e}(X_i) < 0.0891$	$0.0891 \leq \hat{e}(X_i) \leq 0.9109$	$0.9109 < \hat{e}(X_i)$	
Losers	82	172	5	259
Winners	4	151	82	237
All	86	323	87	496

Table 17.3. ***Estimates of Propensity Score in Trimmed Sample for the IRS Lottery Data***

Covariate	Est	$(\widehat{\text{s.e.}})$	t-Stat
Intercept	21.77	(0.13)	164.8
Linear terms			
# Tickets	−0.08	(0.46)	−0.2
Education	−0.45	(0.08)	−5.7
Working Then	3.32	(1.95)	1.7
Earnings Year -1	−0.02	(0.01)	−1.4
Age	−0.05	(0.01)	−3.7
Pos Earnings Year -5	1.27	(0.42)	3.0
Year Won	−4.84	(1.53)	−3.2
Earnings Year -5	−0.04	(0.02)	−2.1
Quadratic terms			
Year Won × Year Won	0.37	(0.12)	3.2
Tickets Bought × Year Won	0.14	(0.06)	2.2
Tickets Bought × Tickets Bought	−0.04	(0.02)	−1.8
Working Then × Year Won	−0.49	(0.30)	−1.6

17.3. Subclassification on the PS and Bias Reduction(1/2)

<Subclassification>

- Partition the range of the PS into J blocks
  - $B_i(j) = \begin{cases} 1 & : \quad if \quad b_{j-1} \leqslant \hat{e}(X_i) < b_j \\ 0 & : \qquad\qquad\qquad otherwise \end{cases}$
  - Within an interval $(b_{j-1}, b_j]$, they have identical PS.
  - $N_c(j) = \sum_{i=1}^{N} (1 - W_i) * B_i(j), N_t(j) = \sum_{i=1}^{N} (W_i) * B_i(j)$
  - $q(j) = \frac{N(j)}{N} (j = 1, 2, , , , J)$
- Aseess the adequacy of the current number of blocks
  - For each stratum, a t-statistic for the null hypothesis :
    the average value of the estimated **linearized PS is the same in that stratum**
  - Check if the strata is adequately balanced by 3 parameters

$\rightarrow$ $t_{max} = 1.96$, $N_{min,1} = 3$, $N_{min,2} = K + 2$

17.3. Subclassification on the PS and Bias Reduction(2/2)

<Subclassification Estimator for the Average Treatment Effect>

- **(Blocking estimator)** $\hat{\tau}^{dif}(j) = \bar{Y_t}^{obs}(j) - \bar{Y_c}^{obs}(j)$, where
  - $\bar{Y_t}^{obs}(j) = \frac{1}{N_t(j)} \sum_{i=1}^{N} W_i * B_i(j) * Y_i^{obs}$
  - $\bar{Y_c}^{obs}(j) = \frac{1}{N_c(j)} \sum_{i=1}^{N} (1 - W_i) * B_i(j) * Y_i^{obs}$
  - ☞ $\hat{\tau}^{strat} = \sum_{j=1}^{J} q(j) * \hat{\tau}^{dif}(j)$, weighted by relative block sizes
- Assess the **adequacy of the current number of blocks**
  - For each stratum, a t-statistic for the null hypothesis : the average value of the estimated linearized PS is the same in that stratum
  - Check if the strata is adequately balanced by 3 parameters
  - $\rightarrow$ $t_{max} = 1.96$, $N_{min,1} = 3$, $N_{min,2} = K + 2$

17.4 Subclassification and the Lottery data

<Determine the number of subclasses by cutoff values>

- **Cutoff Values :** $t_{max} = 1.96$, $N_{min,1} = 3$, $N_{min,2} = K + 2$
  - $\gamma_k = \frac{\sum_{j=1}^{J} q(j) * (\bar{X}_{t,k}(j) - \bar{X}_{c,k}(j))}{\bar{X}_{t,k} - \bar{X}_{c,k}}$
  - $\bar{X}_{t,k}(j)$: (k)th elements of $\bar{X}_t(j)$
  - γ_k : ratio of **the bias reduction** from the subclassification

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

Table 17.5. ***Subclassification with Two Subclasses, Split at Median Propensity Score for the IRS Lottery Data***

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

17.7. Average Treatment Effects for the Lottery data

<The estimates of the overall treatment effect on earning>

- The algorithm for choosing the number of blocks led to **5 blocks**
- Both trimming and subclassification **reduce the sensitivity** to the inclusion of covariates in the regression specification
- Lead to **more credible estimates of casual effects**

Table 17.7. ***Estimated Average Treatment Effects with Final Subclassification for the IRS Lottery Data (regression estimates as in Table 17.6)***

Covariates	Full Sample 1 Block		Selected Sample 1 Block		Selected Sample 2 Blocks		Selected 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)

17.9. Conclusion

- **Subclassification estimator** for ATE under unconfoundedness
  - Use Propensity-Score to construct strata within which **the covariates are well balanced**
  - In settings, trim the units as PS values close to zero or one

The simple subclassification estimator for ATE

1. $\hat{\tau}^{strat} = \sum_{j=1}^{J}[q(j) * \hat{\tau}^{dif}(j)]$, where $\hat{\tau}^{dif}(j) = \overline{Y_t}^{obs}(j) - \overline{Y_c}^{obs}(j)$

   - $Y_i^{obs} = \alpha(j) + \tau(j) * W_i + \epsilon_i$ (17.2)

   ☞ (17.2) does not eliminate all correlations within blocks

   - $Y_i^{obs} = \alpha(j) + \tau(j) * W_i + X_i * \beta(j) + \epsilon_i$ (17.3)

   ☞ (17.3) : The inclusion of the covariates aims to improve precision

2. Analysis estimates the ATE within the blocks **by linear regression**

   - $\left(\hat{\alpha}(j), \hat{\tau}^{adj}(j), \hat{\beta}(j)\right) = \underset{\alpha,\tau,\beta}{argmin} \sum_{i=1}^{N} B_i(j)(Y_i^{obs} - \alpha - \tau * W_i - X_i\beta)^2$ (17.4)

3. Within block least squares estimates, $\hat{\tau}^{\text{adj}}(j)$ are averaged

   - $\hat{\tau}^{\text{strat,adj}} = \sum_{j=1}^{J} q(j) \cdot \hat{\tau}^{\text{adj}}(j)$

17.6. Neymanian Inference(1/3)

The sampling variance of $\hat{\tau}^{dif(j)}$ is

1. $V(\hat{\tau}^{dif}(j)) = \frac{S_c(j)^2}{N_c(j)} + \frac{S_t(j)^2}{N_t(j)} - \frac{S_{ct}(j)^2}{N(j)}$
   - $S_c(j)^2 = \frac{1}{N(j)-1}\sum_{i=1}^{N} B_i(j)(Y_i(0) - \overline{Y}(0,j))^2$
   - $S_t(j)^2 = \frac{1}{N(j)-1}\sum_{i=1}^{N} B_i(j)(Y_i(1) - \overline{Y}(1,j))^2$
   - $S_{ct}(j)^2 = \frac{1}{N-1}\sum_{i=1}^{N} B_i(j)(Y_i(1) - Y_i(0) - \tau(j))^2$

2. $\hat{V}(\hat{\tau}^{dif}(j))$

$$= \frac{1}{N_c(j)(N_c(j)-1)}\sum_{i:W_i=0} B_i(j)(Y_i^{obs} - \overline{Y_c^{obs}}(j))^2 + \frac{1}{N_t(j)(N_t(j)-1)}\sum_{i:W_i=1} B_i(j)(Y_i^{obs} - \overline{Y_t^{obs}}(j))^2$$

3. $\hat{V}(\hat{\tau}^{strat}) = \sum_{j=1}^{J}\hat{V}(\hat{\tau}^{dif}(j))q(j)^2 = \hat{V}(\hat{\tau}^{dif}(j))(\frac{N(j)}{N})^2$

Further covariance adjustment to reduce the remaining bias

- $(\hat{\alpha}(j), \hat{\tau}^{adj}(j), \hat{\beta}(j))$ be the ordinary least squares estimates in (17.4)

- $\hat{\Delta}(j) = \frac{1}{N(j)} \sum_{i=1}^{N} B_i(j) \begin{pmatrix} 1 & W_i & \acute{X}_i \\ & W_i & W_i\acute{X}_i \\ & & X_i\acute{X}_i \end{pmatrix}$

- $\Gamma\hat{(j)} = \frac{1}{N(j)} \sum_{i=1}^{N} B_i(j)(Y_i - \hat{\alpha}(j) - \hat{\tau}^{adj}(j)W_i - \hat{\beta}(j)\grave{X}_i)^2 \begin{pmatrix} 1 & W_i & \acute{X}_i \\ & W_i & W_i\acute{X}_i \\ & & X_i\acute{X}_i \end{pmatrix}$

- $\hat{V}(\hat{\tau}^{adj})(j) = \frac{1}{N(j)} \left(\hat{\Gamma}(j)\hat{\Delta}(j)^{-1}\hat{\Gamma}(j)\right)^{-1}_{(2,2)}$
- $\hat{V}(\hat{\tau}^{strat,adj}) = \sum_{i=1}^{J} \hat{V}(\hat{\tau}^{adj}(j))q(j)^2$ (17.5)

17.6. Neymanian Inference(3/3)

Table 17.7. ***Estimated Average Treatment Effects with Final Subclassification for the IRS Lottery Data (regression estimates as in Table 17.6)***

Covariates	Full Sample 1 Block		Selected Sample 1 Block		Selected Sample 2 Blocks		Selected 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)

Weighting Estimators

- Subclassification estimator & Horbitz-Thomson estimator

$$\mathbb{E}\left[\frac{W_i \cdot Y_i^{\text{obs}}}{e(X_i)}\right] = \mathbb{E}_{\text{sp}}[Y_i(1)] \quad \text{and} \quad \mathbb{E}\left[\frac{(1-W_i)\cdot Y_i^{\text{obs}}}{1-e(X_i)}\right] = \mathbb{E}_{\text{sp}}[Y_i(0)]$$

$$\mathbb{E}\left[\frac{W_i \cdot Y_i^{\text{obs}}}{e(X_i)}\right] = \mathbb{E}\left[\frac{W_i \cdot Y_i(1)}{e(X_i)}\right]$$

By iterated expectations, we can write this as

$$\mathbb{E}\left[\frac{W_i \cdot Y_i(1)}{e(X_i)}\right] = \mathbb{E}\left[\mathbb{E}\left[\frac{W_i \cdot Y_i(1)}{e(X_i)} \mid X_i\right]\right]$$

$$\mathbb{E}\left[\frac{W_i \cdot Y_i(1)}{e(X_i)} \mid X_i\right] = \frac{\mathbb{E}_W[W_i \mid X_i] \cdot \mathbb{E}_{\text{sp}}[Y_i(1) \mid X_i]}{e(X_i)} = \frac{e(X_i) \cdot \mathbb{E}_{\text{sp}}[Y_i(1) \mid X_i]}{e(X_i)}$$
$$= \mathbb{E}_{\text{sp}}[Y_i(1) \mid X_i]$$

and thus

$$\mathbb{E}\left[\frac{W_i \cdot Y_i(1)}{e(X_i)}\right] = \mathbb{E}_{\text{sp}}[\mathbb{E}_{\text{sp}}[Y_i(1) \mid X_i]] = \mathbb{E}_{\text{sp}}[Y_i(1)]$$

Weighting Estimators

- $\mathbb{E}\left[\frac{W_i \cdot Y_i(1)}{e(X_i)}\right] = \mathbb{E}_{\text{sp}}\left[\mathbb{E}_{\text{sp}}\left[Y_i(1) \mid X_i\right]\right] = \mathbb{E}_{\text{sp}}\left[Y_i(1)\right]$ (17.6)
- Equation (17.6) suggest estimating $\mathbb{E}\left[Y_i(1)\right]$ and $\mathbb{E}\left[Y_i(0)\right]$ as

$$\widehat{\mathbb{E}_{\text{sp}}\left[Y_i(1)\right]} = \frac{1}{N}\sum_{i=1}^{N}\frac{W_i \cdot Y_i^{\text{obs}}}{e\left(X_i\right)}$$

and estimating **the average treatment effect** $\tau_{\text{sp}} = \mathbb{E}_{\text{sp}}\left[Y_i(1) - Y_i(0)\right]$ as a HorvitzThompson estimator,

$$\tilde{\tau}^{\text{ht}} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{W_i \cdot Y_i^{\text{obs}}}{e\left(X_i\right)} - \frac{\left(1 - W_i\right) \cdot Y_i^{\text{obs}}}{1 - e\left(X_i\right)}\right)$$

Weighting Estimators

We rarely know the PS, so using the estimated propensity score $\hat{e}(X_i)$,

$$\hat{\tau}^{\text{ht}} = \sum_{i=1}^{N} \frac{W_i \cdot Y_i^{\text{obs}}}{\hat{e}(X_i)} \Big/ \sum_{i=1}^{N} \frac{W_i}{\hat{e}(X_i)} - \sum_{i=1}^{N} \frac{(1-W_i) \cdot Y_i^{\text{obs}}}{1-\hat{e}(X_i)} \Big/ \sum_{i=1}^{N} \frac{1-W_i}{1-\hat{e}(X_i)}.$$

- Normalizing the weights to one imporves the MSE of the estimator

17.8. Weighting Estimators and Subclassification(4/4)

Subclassification estimator VS the Horvitz-Thompson estimator

- As the propensity score is in the denominator of the weights : Horbitz-Thompson Estimator **(Bias)**
- **Smaller estimated Sampling Variance** in Subclassification estimator Relative merits of the Subclassification Estimator because of $MSE = bias^2 + var$

Table 17.10. ***Estimated Bias and Estimated Sampling Variance for Horvitz-Thompson and Subclassification Estimators under Linear Model for the IRS 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