Augmented IPW and Doubly Robust Estimators

David M. Vock

PubH 7485/8485

Advantages/Disadvantages of IPW versus regression model

- Both approaches require assuming consistency and no unmeasured confounding
- IPW requires assuming positivity
- IPW requires correctly specifying propensity model; regression approach requires specifying the outcome model
- IPW: can fit a single model (propensity score) and then obtain causal estimates for several outcomes
- Perception that modeling treatment allocation "easier" than outcome
- IPW tends to be more variably (even without extreme weights)

Augmented IPW estimators

- If we take the point of view that the propensity score model is correctly specified, then one can use semiparametric theory to show that all consistent and asymptotically normal estimators for $E(Y^1)$ can be written as

$$\hat{E}(Y^1) = \frac{1}{n}\sum_{i=1}^{n} \frac{A_i Y_i}{\pi(X_i;\hat{\gamma})} - \{A_i - \pi(X_i;\hat{\gamma})\}h(X_i) \qquad (1)$$

- where $h(X_i)$ is any arbitrary function of X
- Show that this is a consistent estimator of $E(Y^1)$ regardless of the choice of $h(X)$
- This class of estimators is known as augmented IPW estimators
- Questions: how to choose $h(X)$

Choosing $h(X)$

- Although all these estimators are consistent for $E(Y^1)$ they will have different variability
- One can show using semiparametric theory that the choice that leads to the estimator with the smallest asymptotic variance is

$$h(X_i) = \frac{E(Y|A=1,X)}{\pi(X)} \tag{2}$$

- Let's think through the implication of this choice. Note that $E(Y|A=1,X) = E(Y^1|X)$ under our standard identifying assumptions

1) When $A_i = 0$ (i.e., Y_i^1 is unobserved) then the summand in equation (1) is $E(Y_i^1|X_i)$ – effectively we are imputing the best guess of the potential outcome given covariates
2) When $A_i = 1$ (i.e., Y_i^1 is observed) then the summand is $E(Y_i^1|X_i) + \frac{1}{\pi(X_i)}(Y_i^1 - E(Y_i^1|X_i))$

Choosing $h(X)$

- Showing that

$$h(X_i) = \frac{E(Y|A=1,X)}{\pi(X)} \tag{3}$$

  leads to the estimator with the smallest asymptotic variance is DIFFICULT
- But showing this choice of $h(X)$ leads to an estimator with smaller variance than the standard IPW estimator (i.e., $h(X) = 0$) is more doable

Choosing $h(X)$

- Of course, $E(Y|A,X)$ is not known to us (if it were, we would have just used regression modeling to estimate $E(Y^1)$)
- Nonetheless, just like regression modeling, we can posit a model for $E(Y|A,X) = \mu(A,X;\eta)$ and then obtain an estimate for η (say $\hat{\eta}$)
- We can then plug this estimate into the formula so that the estimator is

$$\hat{E}(Y^1) = \frac{1}{n}\sum_{i=1}^{n} \frac{A_i Y_i}{\pi(X_i;\hat{\gamma})} - \frac{\{A_i - \pi(X_i;\hat{\gamma})\}}{\pi(X_i;\hat{\gamma})}\mu(A=1,X_i;\hat{\eta}) \qquad (4)$$

Doubly Robust Estimator

$$\hat{E}(Y^1) = \frac{1}{n}\sum_{i=1}^{n} \frac{A_i Y_i}{\pi(X_i;\hat{\gamma})} - \frac{\{A_i - \pi(X_i;\hat{\gamma})\}}{\pi(X_i;\hat{\gamma})}\mu(A=1, X_i;\hat{\eta}) \qquad (5)$$

- Consistent and asymptotically normal if either $\mu(A, X; \eta)$ or $\pi(X; \gamma)$ correctly specified
- Known as doubly robust or doubly protected estimator
- Note that if $\mu(A, X; \eta)$ is misspecified, the augmented estimator is no longer guaranteed to have smaller variance than the "standard" IPW estimator
- Some refer to summand as pseudo-outcome

Estimator for $E(Y^0)$

- If we take the point of view that the propensity score model is correctly specified, then one can use semiparametric theory to show that all consistent and asymptotically normal estimators for $E(Y^0)$ can be written as

$$\hat{E}(Y^0) = \frac{1}{n}\sum_{i=1}^{n} \frac{(1-A_i)Y_i}{1-\pi(X_i;\hat{\gamma})} - \{1 - A_i - (1-\pi(X_i;\hat{\gamma}))\}h(X_i) \quad (6)$$

- The optimal choice of $h(X_i)$ is then $h(X_i) = \frac{E(Y|A=0,X)}{1-\pi(X)}$

Form Pseudo-Outcomes for AIPW

```
PO1 <- imai$VOTED98*w1 - ((imai$PHN.C1-ps)/ps)*pred1
PO0 <- imai$VOTED98*w0 - ((1-imai$PHN.C1-(1-ps))/(1-ps))*pred0
ATE_AIPW <- mean(PO1 - PO0)
```

Voting Example: AIPW Analysss

```
## [1] "Average Treatment Effect"

## [1] 0.102

## [1] "Bootstrap SE"

## [1] 0.033

## [1] "Bootstrap Normal 95% CI"

## [1] 0.0377 0.1672
```

AIPW Results: Key Assumptions

Identifying

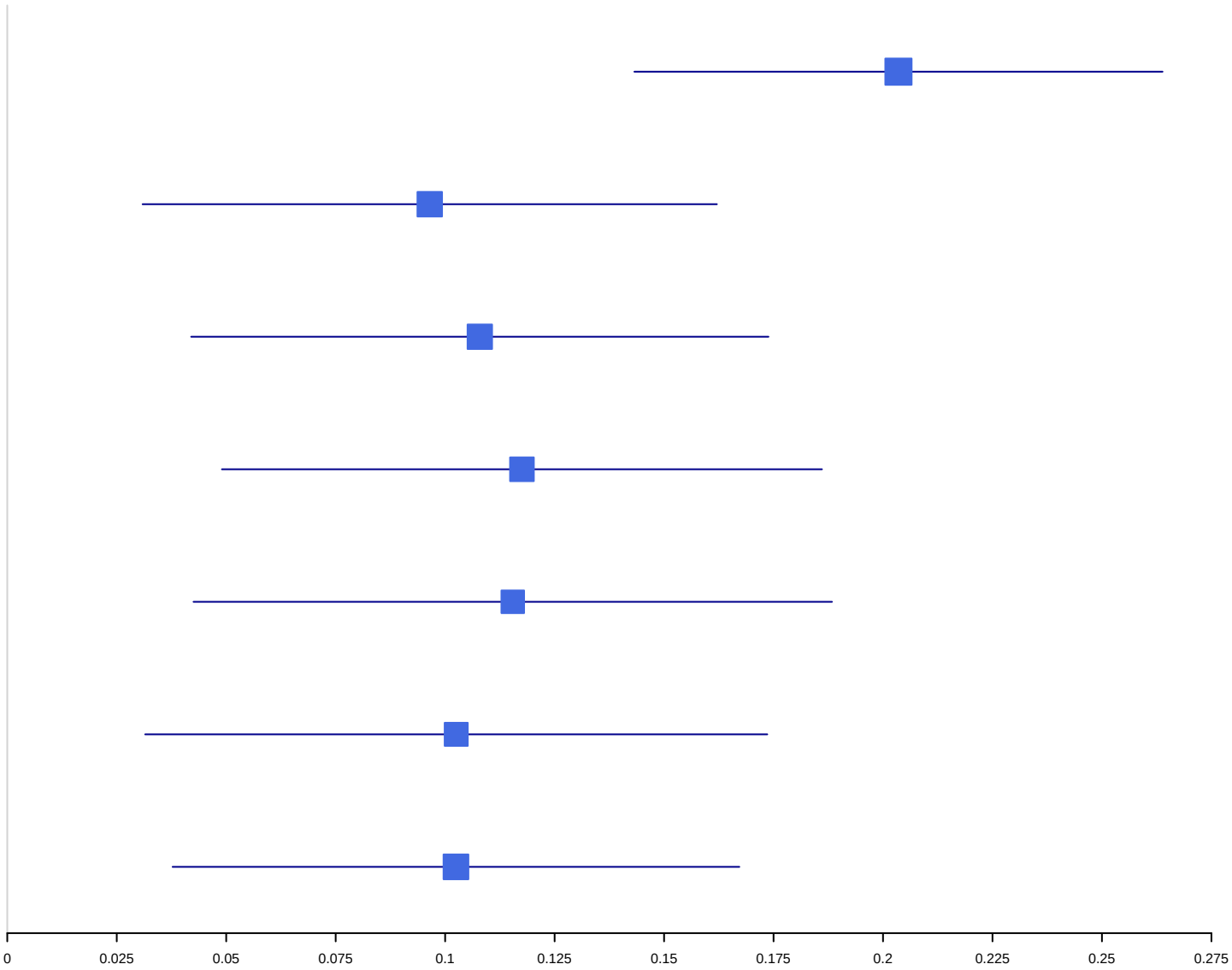
1) Consistency
2) No Unmeasured confounding
3) Positivity

Modeling

1) Propensity score model (given all confounders) OR outcome model (given all confounders) correctly specified.

Putting it All Together

Method	ATE
Per Protocol	0.204
Regression	0.096
PS Reg	0.108
PSS	0.118
IPW1	0.115
IPW2	0.103
AIPW	0.102



Simulation Example

- Let $X_i \sim N(0,1)$, $Y_i^1|X_i \sim N(0.5+\gamma X, 1)$ and $Y_i^0|X_i \sim N(\gamma X, 1)$. ATE $= E(Y^1) - E(Y^0) = 0.5$
- Note that in the "real world" we do not observe $\{Y_i^1, Y_i^0$ but would observe $Y_i = A_i Y_i^1 + (1-A_i)Y_i^0$. This implies that $Y_i|A_i, X_i \sim (0.5A_i + \gamma X_i, 1)$
- Let $A_i|X_i \sim$ Bernoulli(p_i) where $p_i = \exp(0+\alpha X_i)/\{1+\exp(0+\alpha X_i)\}$
- Generate a sample of size 500 consistent with this data generating mechanism with $\gamma = 1$ and $\alpha = 1$.
- With these coefficients the R^2 for regressing the outcome on X in placebo group is 0.5 and C-index for the treatment allocation ≈ 0.75
- 100 Monte Carlo datasets

Simulation Results

```
##            bias    sd
## IPW  -0.00575 0.136
## REG   0.00381 0.109
## AIPW  0.00084 0.116
```