Augmented IPW and Doubly Robust Estimators

David S. Rosenberg

NYU: CDS

February 19, 2021

Contents

Applying control variates to IPW estimators

Recap: MAR and the IPW mean

- Observed data: $(X, R, RY), (X_1, R_1, R_1 Y_1), \ldots, (X_n, R_n, R_n Y_n)$ i.i.d.
- $R_1, \ldots, R_n \in \{0, 1\}$ is the response indicator.
- In missing at random (MAR) setting, $R_i \perp\!\!\!\perp Y_i \mid X_i$
- Probability of response is given by the **propensity score function**:

$$\pi(x) = \mathbb{P}(R_i = 1 \mid X_i = x) \quad \forall i.$$

- The inverse propensity weighted mean estimate of $\mathbb{E}Y$ is

$$\hat{\mu}_{\text{ipw}} = \frac{1}{n} \sum_{i=1}^{n} \frac{R_i Y_i}{\pi(X_i)}$$

- We found $\hat{\mu}_{\text{ipw}}$ had high variance. Can a control variate help?

Prep for IPW with control variate

- Suppose we have $f(x) \approx \mathbb{E}[Y \mid X = x]$.
- Let's try to use $f(X)$ as a control variate in

$$\hat{\mu}_{\text{ipw}} = \frac{1}{n}\sum_{i=1}^{n}\frac{R_i Y_i}{\pi(X_i)}$$

- A natural estimate for $R_i Y_i/\pi(X_i)$ is $R_i f(X_i)/\pi(X_i)$.
- Following our pattern for control-variate based estimators, we get the following:

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left(\frac{R_i Y_i}{\pi(X_i)} - \frac{R_i f(X_i)}{\pi(X_i)} + \mathbb{E}\left[\frac{R_i f(X_i)}{\pi(X_i)}\right]\right) \\
= \quad & \frac{1}{n}\sum_{i=1}^{n}\left(\frac{R_i Y_i}{\pi(X_i)} - \frac{R_i f(X_i)}{\pi(X_i)}\right) + \mathbb{E}\left[\frac{R f(X)}{\pi(X)}\right]
\end{aligned}$$

- How to compute $\mathbb{E}\left[\frac{Rf(X)}{\pi(X)}\right]$?

Computing the expectation

- If we try to estimate $\mathbb{E}\left[\frac{Rf(X)}{\pi(X)}\right]$ from our sample with

$$\frac{1}{n}\sum_{i=1}^{n}\frac{R_i f(X_i)}{\pi(X_i)},$$

we'll end up back with $\hat{\mu}_{\text{ipw}}$:

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n}\left(\frac{R_i Y_i}{\pi(X_i)} - \frac{R_i f(X_i)}{\pi(X_i)}\right) + \frac{1}{n}\sum_{i=1}^{n}\frac{R_i f(X_i)}{\pi(X_i)} \\ = \quad & \frac{1}{n}\sum_{i=1}^{n}\frac{R_i Y_i}{\pi(X_i)} = \hat{\mu}_{\text{ipw}}. \end{aligned}$$

- Luckily, in this setting there's a workaround...

Computing the expectation

- Recall that

$$
\begin{aligned}
\mathbb{E}\left[\frac{Rf(X)}{\pi(X)}\right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{Rf(X)}{\pi(X)} \mid X\right]\right] && \text{Adam's Law} \\
&= \mathbb{E}\left[\frac{f(X)}{\pi(X)}\mathbb{E}[R \mid X]\right] && \text{Taking out what is known} \\
&= \mathbb{E}[f(X)]
\end{aligned}
$$

- Can we use the sample we already have? In this situation, yes!
- We'll plug in

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i)$$

for $\mathbb{E}[f(X)]$.

"Deriving" the Augmented IPW estimator

- Consider

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left(\frac{R_iY_i}{\pi(X_i)}-\frac{R_if(X_i)}{\pi(X_i)}+\mathbb{E}\left[\frac{R_if(X_i)}{\pi(X_i)}\right]\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{R_iY_i}{\pi(X_i)}-\frac{R_if(X_i)}{\pi(X_i)}\right)+\mathbb{E}\left[f(X)\right]\\
&\approx \frac{1}{n}\sum_{i=1}^{n}\left(\frac{R_iY_i}{\pi(X_i)}-\frac{R_if(X_i)}{\pi(X_i)}\right)+\frac{1}{n}\sum_{i=1}^{n}f(X_i)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{R_iY_i}{\pi(X_i)}-\frac{R_if(X_i)}{\pi(X_i)}+f(X_i)\right)
\end{aligned}$$

- Note that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{R_iY_i}{\pi(X_i)}-\frac{R_if(X_i)}{\pi(X_i)}+f(X_i)\right)\right]=\mathbb{E}Y.$$

The augmented IPW (AIPW) estimator

- Let $f(x;\theta) : \mathcal{X} \to \mathbb{R}$ for $\theta \in \mathbb{R}^d$.
- We can fit $f(x;\theta)$ by least squares on the complete cases:

$$\hat{\theta} = \operatorname*{arg\,min}_{\theta \in \mathbb{R}^d} \sum_{i=1}^{n} R_i \left(f(X_i;\theta) - Y_i\right)^2.$$

- Then the **augmented IPW (AIPW)** estimator is defined as

$$\hat{\mu}_{\text{aipw}} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{R_i Y_i}{\pi(X_i)} - \frac{R_i f(X_i;\hat{\theta})}{\pi(X_i)} + f(X_i;\hat{\theta}) \right).$$

The augmented IPW (AIPW) estimator

- Let $f(x;\theta): \mathcal{X} \to \mathbb{R}$ for $\theta \in \mathbb{R}^d$.
- We can fit $f(x;\theta)$ by least squares on the complete cases:

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^d}{\arg\min} \sum_{i=1}^{n} R_i \left(f(X_i;\theta) - Y_i\right)^2.$$

- Then the **augmented IPW (AIPW)** estimator is defined as

$$\hat{\mu}_{\text{aipw}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{R_i Y_i}{\pi(X_i)} - \frac{R_i f(X_i;\hat{\theta})}{\pi(X_i)} + f(X_i;\hat{\theta})\right).$$

- Another way to look at this is that our control variate is $\frac{R_i f(X_i)}{\pi(X_i)} - f(X_i)$, which has expectation 0. From this perspective, this exactly follows our recipe for a control-variate adjusted estimator.

Experimental results

Performance on SeaVan1

- Response bias, linear model (well-specified)

estimator	mean	SD	SE	bias	RMSE
mean	0.3564	0.0515	0.0016	-0.6431	0.6452
ipw_mean	1.0127	0.2968	0.0094	0.0132	0.2971
sn_ipw_mean	0.9906	0.1890	0.0060	-0.0089	0.1892
impute_linear	1.0022	0.0781	0.0025	0.0027	0.0782
aipw_mean	1.0014	0.1405	0.0044	0.0019	0.1406

2021-02-21

- Response bias, linear model (well-specified)

estimator	mean	SD	SE	bias	RMSE
mean	0.3564	0.0515	0.0016	-0.6431	0.6452
ipw_mean	1.0127	0.2968	0.0094	0.0132	0.2971
sn_ipw_mean	0.9906	0.1890	0.0060	-0.0089	0.1892
impute_linear	1.0022	0.0781	0.0025	0.0027	0.0782
aipw_mean	1.0014	0.1405	0.0044	0.0019	0.1406

- The AIPW estimator is significantly better in RMSE than the IPW mean estimator. The improvement is coming primarily from the reduction in variance / SD. It's even better than the self-normalized IPW estimator.
- Still not as good as the regression imputation estimator, though that's a tough one to beat when the model is well-specified.

Performance on SeaVan2

- Response bias, linear model (misspecified)

estimator	mean	SD	SE	bias	RMSE
mean	0.3442	0.0508	0.0016	-0.3218	0.3258
ipw_mean	0.6740	0.1898	0.0060	0.0080	0.1900
sn_ipw_mean	0.6650	0.1412	0.0045	-0.0010	0.1412
impute_linear	0.9403	0.0794	0.0025	0.2743	0.2855
aipw_mean	0.6696	0.1814	0.0057	0.0036	0.1814

Performance on SeaVan2

- Response bias, linear model (misspecified)

estimator	mean	SD	SE	bias	RMSE
mean	0.3442	0.0508	0.0016	-0.3218	0.3258
ipw_mean	0.6740	0.1898	0.0060	0.0080	0.1900
sn_ipw_mean	0.6650	0.1412	0.0045	-0.0010	0.1412
impute_linear	0.9403	0.0794	0.0025	0.2743	0.2855
aipw_mean	0.6696	0.1814	0.0057	0.0036	0.1814

- For this distribution, the linear model is misspecified. Together with response bias, this explains the poor performance of the linear regression imputation estimator.
- The AIPW estimator still manages to reduce the SD over the IPW mean a bit, and does show a slight improvement. (Though not as good as the self-normalized IPW mean.)

Performance on MAR_normal_nonlinear

- Response bias, linear model (very misspecified)

estimator	mean	SD	SE	bias	RMSE
mean	2.4075	0.0476	0.0015	0.9063	0.9075
ipw_mean	1.4985	0.0851	0.0027	-0.0027	0.0852
sn_ipw_mean	1.5070	0.1224	0.0039	0.0057	0.1225
impute_linear	2.4060	0.0583	0.0018	0.9048	0.9066
aipw_mean	1.4989	0.3061	0.0097	-0.0023	0.3061

Performance on MAR_normal_nonlinear

- Response bias, linear model (very misspecified)

estimator	mean	SD	SE	bias	RMSE
mean	2.4075	0.0476	0.0015	0.9063	0.9075
ipw_mean	1.4985	0.0851	0.0027	-0.0027	0.0852
sn_ipw_mean	1.5070	0.1224	0.0039	0.0057	0.1225
impute_linear	2.4060	0.0583	0.0018	0.9048	0.9066
aipw_mean	1.4989	0.3061	0.0097	-0.0023	0.3061

- For this distribution, the linear model is a very poor fit. Together with response bias, leads to very bad performance the linear regression imputation estimator. Almost all of the RMSE for the regression imputation estimator comes from the bias.
- The IPW estimators are actually comparatively quite good.
- The AIPW estimator is significantly worse than the IPW estimators, but significantly better than the regression imputation estimator. As we know from the theory, AIPW is unbiased when using the true propensity score. Thus the bias being indistinguishable from zero is expected.

IPW, regression imputation, and model misspecification

Unknown propensity function and IPW

- In practice, for response bias situations,
  - we usually do not know $\pi(x) = \mathbb{P}(R = 1 \mid X = x)$.
- But we can learn it from our data.
- Let $\pi(x;\gamma) : \mathcal{X} \to (0,1)$ for $\gamma \in \mathbb{R}^d$ be a parametrized space of functions.
  - Typically $\pi(x;\gamma)$ is a logistic regression model.
- Fit to $(X_1, R_1), \ldots, (X_n, R_n)$ with maximum likelihood:

$$\hat{\gamma} = \operatorname*{arg\,max}_{\gamma \in \mathbb{R}^d} \prod_{i=1}^{n} [\pi(X_i;\gamma)]^{R_i} [1 - \pi(X_i;\gamma)]^{1-R_i}.$$

- Then, for example, our IPW mean estimator becomes

$$\hat{\mu}_{\text{ipw}} = \frac{1}{n} \sum_{i=1}^{n} \frac{R_i Y_i}{\pi(X_i, \hat{\gamma})}.$$

IPW mean with estimated propensity

- Propensity model: $\pi(x;\gamma) : \mathcal{X} \to (0,1)$ for $\gamma \in \mathbb{R}^d$.
- Suppose $\hat{\gamma} \xrightarrow{P} \gamma^*$ as $n \to \infty$.
- Then

$$\hat{\mu}_{\text{ipw}} = \frac{1}{n}\sum_{i=1}^{n} \frac{R_i Y_i}{\pi(X_i, \hat{\gamma})} \xrightarrow{P} \mathbb{E}\left[\frac{RY}{\pi(X,\gamma^*)}\right]$$

and

$$\begin{aligned}
\mathbb{E}\left[\frac{RY}{\pi(X,\gamma^*)}\right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{RY}{\pi(X,\gamma^*)} \mid X\right]\right] \\
&= \mathbb{E}\left[\frac{1}{\pi(X,\gamma^*)}\mathbb{E}[R \mid X]\,\mathbb{E}[Y \mid X]\right] \\
&= \mathbb{E}\left[\frac{\pi(X)}{\pi(X,\gamma^*)}\mathbb{E}[Y \mid X]\right]
\end{aligned}$$

IPW mean under misspecification

- Propensity score function: $\pi(x) = \mathbb{P}(R = 1 \mid X = x)$.
- If $\pi(x;\gamma)$ is **well-specified** and $\pi(x;\gamma^*) = \pi(x)$, then

$$\hat{\mu}_{\text{ipw}} \xrightarrow{P} \mathbb{E}\left[\frac{\pi(X)}{\pi(X,\gamma^*)}\mathbb{E}[Y \mid X]\right] = \mathbb{E}Y.$$

- But if $\pi(x;\gamma)$ is **misspecified,** then $\pi(x;\gamma^*) \neq \pi(x)$ and $\hat{\mu}_{\text{ipw}} \overset{P}{\not\to} \mathbb{E}Y$ (in general).
- So a misspecified propensity score model can be an issue.

Asymptotics of regression imputation

- Let $f(x;\theta) : \mathcal{X} \to \mathbb{R}$ for $\theta \in \mathbb{R}^d$.
- We can fit $f(x;\theta)$ by least squares on complete cases:

$$\hat{\theta} = \operatorname*{arg\,min}_{\theta \in \mathbb{R}^d} \sum_{i=1}^{n} R_i \left(f(X_i;\theta) - Y_i\right)^2.$$

- If $\hat{\theta} \xrightarrow{P} \theta^*$ as $n \to \infty$, then under reasonable assumptions,
  - the regression imputation estimator converges as

$$\hat{\mu}_{f(x;\hat{\theta})} = \frac{1}{n} \sum_{i=1}^{n} \left[R_i Y_i + (1 - R_i) f(X_i;\hat{\theta})\right] \xrightarrow{P} \mathbb{E}\left[R_i Y_i + (1 - R_i) f(X_i;\theta^*)\right].$$

Regression imputation under misspecification

- If $f(x;\theta)$ is **well-specified** and $f(x;\theta^*) = \mathbb{E}[Y \mid X = x]$, then

$$\hat{\mu}_{f(x;\hat{\theta})} \xrightarrow{P} \mathbb{E}\left[R_i Y_i + (1 - R_i)\,\mathbb{E}\left[Y \mid X = x\right]\right] = \mathbb{E} Y.$$

  - (We'll show the last equality in the homework.)
- If $f(x;\theta)$ is **misspecified**, then generally regression imputation is not consistent:

$$\hat{\mu}_{f(x;\hat{\theta})} \xrightarrow{P} \mathbb{E}\left[R_i Y_i + (1 - R_i)\, f(X_i;\theta^*)\right] \neq \mathbb{E} Y.$$

Double robustness

AIPW in practice

- Propensity model: $\pi(x;\gamma) : \mathcal{X} \to (0,1)$ for $\gamma \in \mathbb{R}^d$.
- Regression model: $f(x;\theta) : \mathcal{X} \to \mathbb{R}$ for $\theta \in \mathbb{R}^d$.
- Fit as described above. Then the AIPW estimator is

$$\hat{\mu}_{\text{aipw}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{R_i Y_i}{\pi(X_i;\hat{\gamma})} - \frac{R_i f(X_i;\hat{\theta})}{\pi(X_i;\hat{\gamma})} + f(X_i,\hat{\theta})\right).$$

- We can also use cross-fitting in this situation.
- Note that if $\pi(x;\hat{\gamma}) \equiv 1$, then $\hat{\mu}_{\text{aipw}} = \frac{1}{n}\sum_{i=1}^{n}\left(R_i Y_i - R_i f(X_i;\hat{\theta}) + f(X_i,\hat{\theta})\right)$,
  - which is **exactly** the regression imputation estimator.

AIPW in practice

- Propensity model: $\pi(x;\gamma):\mathcal{X}\to(0,1)$ for $\gamma\in\mathbb{R}^d$.
- Regression model: $f(x;\theta):\mathcal{X}\to\mathbb{R}$ for $\theta\in\mathbb{R}^d$.
- Fit as described above. Then the AIPW estimator is

$$\hat{\mu}_{\text{aipw}} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{R_iY_i}{\pi(X_i;\hat{\gamma})} - \frac{R_if(X_i;\hat{\theta})}{\pi(X_i;\hat{\gamma})} + f(X_i,\hat{\theta})\right).$$

- We can also use cross-fitting in this situation.
- Note that if $\pi(x;\hat{\gamma})\equiv 1$, then $\hat{\mu}_{\text{aipw}} = \frac{1}{n}\sum_{i=1}^{n}\left(R_iY_i - R_if(X_i;\hat{\theta}) + f(X_i,\hat{\theta})\right)$,
  - which is **exactly** the regression imputation estimator.

- In the last bullet, note that we're examining the case that $\pi(X_i;\hat{\gamma})\equiv 1$, not that $\pi(x)\equiv 1$. If we had $\pi(x)\equiv 1$, then then we would only ever observe $R_i = 1$, and $\hat{\mu}_{\text{aipw}}$ would reduce to the complete case mean.

Asymptotics of AIPW

- Propensity model: $\pi(x;\gamma) : \mathcal{X} \to (0,1)$ for $\gamma \in \mathbb{R}^d$.
- Regression model: $f(x;\theta) : \mathcal{X} \to \mathbb{R}$ for $\theta \in \mathbb{R}^d$.
- Suppose we have $\hat{\theta} \xrightarrow{P} \theta^*$ and $\hat{\gamma} \xrightarrow{P} \gamma^*$.
- Then one can show (e.g. [SV18]) that

$$
\begin{aligned}
\hat{\mu}_{\text{aipw}} &\xrightarrow{P} \mathbb{E}\left[\frac{RY}{\pi(X;\gamma^*)} - \frac{Rf(X;\theta^*)}{\pi(X;\gamma^*)} + f(X;\theta^*)\right] \\
&= \mathbb{E}\left[\frac{1}{\pi(X;\gamma^*)}\left[RY - Rf(X;\theta^*) + \pi(X;\gamma^*)f(X;\theta^*)\right]\right] \\
&= \mathbb{E}\left[Y + \frac{1}{\pi(X;\gamma^*)}\left[RY - Rf(X;\theta^*) - \pi(X;\gamma^*)Y + \pi(X;\gamma^*)f(X;\theta^*)\right]\right] \\
&= \mathbb{E}\left[Y + \frac{1}{\pi(X;\gamma^*)}\left(R - \pi(X;\gamma^*)\right)\left(Y - f(X;\theta^*)\right)\right]
\end{aligned}
$$

Asymptotics of AIPW (continued)

$$
\begin{aligned}
\hat{\mu}_{\text{aipw}} &\xrightarrow{P} \mathbb{E}\left[Y + \frac{1}{\pi(X;\gamma^*)}\left(R - \pi(X;\gamma^*)\right)\left(Y - f(X;\theta^*)\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[Y + \frac{1}{\pi(X;\gamma^*)}\left(R - \pi(X;\gamma^*)\right)\left(Y - f(X;\theta^*)\right) \mid X, Y\right]\right] \\
&= \mathbb{E}\left[Y + \frac{1}{\pi(X;\gamma^*)}\left(\mathbb{E}\left[R \mid X, Y\right] - \pi(X;\gamma^*)\right)\left(Y - f(X;\theta^*)\right)\right] \\
&= \mathbb{E}\left[Y + \frac{1}{\pi(X;\gamma^*)}\left(\pi(X) - \pi(X;\gamma^*)\right)\left(Y - f(X;\theta^*)\right)\right]
\end{aligned}
$$

Asymptotics of AIPW, $f(x,\theta)$ well-specified

- We have

$$\hat{\mu}_{\text{aipw}} \xrightarrow{P} \mathbb{E}\left[Y + \underbrace{\frac{1}{\pi(X;\gamma^*)}\left(\pi(X) - \pi(X;\gamma^*)\right)}_{h(X)}\left(Y - f(X;\theta^*)\right)\right].$$

- Suppose $f(x;\theta)$ is well-specified and $f(x;\theta^*) = \mathbb{E}[Y \mid X = x]$.
- Then by the **projection interpretation** of conditional expectation,
- $(Y - f(X;\theta^*) = Y - \mathbb{E}[Y \mid X = x]$ is orthogonal to $h(X)$.
    - i.e. $\mathbb{E}[h(X)(Y - \mathbb{E}[Y \mid X = x])] = 0$.
- Therefore,

$$\hat{\mu}_{\text{aipw}} \xrightarrow{P} \mathbb{E}Y.$$

Asymptotics of AIPW, $\pi(x,\gamma)$ well-specified

- We have

$$\hat{\mu}_{\text{aipw}} \xrightarrow{P} \mathbb{E}\left[Y + \underbrace{\frac{1}{\pi(X;\gamma^*)}\left(\pi(X) - \pi(X;\gamma^*)\right)}_{h(X)}\left(Y - f(X;\theta^*)\right)\right].$$

- Suppose $\pi(x;\gamma)$ is well-specified and $\pi(x;\gamma^*) = \pi(x) = \mathbb{P}(R = 1 \mid X = x)$.
- Then $h(X) \equiv 0$ and

$$\hat{\mu}_{\text{aipw}} \xrightarrow{P} \mathbb{E}Y.$$

AIPW is doubly robust

- To summarize, if **either**
  - $\pi(x;\gamma^*) = \pi(x)$ (i.e. $\pi(x;\gamma)$ is well-specified) **OR**
  - $f(x,\theta^*) = \mathbb{E}[Y \mid X = x]$ (i.e. $f(x,\theta)$ is well-specified)
- Then $\hat{\mu}_{\text{aipw}} \xrightarrow{P} \mathbb{E}Y$.
- An estimator with this property is called **doubly robust**.
- In words, no matter how bad a propensity or regression model is,
  - if at least one of them is well-specified, then
  - $\hat{\mu}_{\text{aipw}}$ is consistent.
- Although many estimators can have this property, $\hat{\mu}_{\text{aipw}}$ is often referred to as
  - **the doubly robust estimator**.

AIPW is doubly robust

- To summarize, if **either**
  - $\pi(x;\gamma^*) = \pi(x)$ (i.e. $\pi(x;\gamma)$ is well-specified) **OR**
  - $f(x,\theta^*) = \mathbb{E}[Y \mid X = x]$ (i.e. $f(x,\theta)$ is well-specified)
- Then $\hat{\mu}_{\text{aipw}} \xrightarrow{P} \mathbb{E}Y$.
- An estimator with this property is called **doubly robust**.
- In words, no matter how bad a propensity or regression model is,
  - if at least one of them is well-specified, then
  - $\hat{\mu}_{\text{aipw}}$ is consistent.
- Although many estimators can have this property, $\hat{\mu}_{\text{aipw}}$ is often referred to as
  - **<u>the</u> doubly robust estimator**.

- For example, suppose we know the true propensity function $\pi(x)$, but we our regression model is misspecified. Then regression imputation may not be asymptotically consistent. But the augmented IPW estimator will be.
- Conversely, if our regression model satisfies $f(x,\theta^*) = \mathbb{E}[Y \mid X = x]$, s well-specified,

Other directions to investigate for AIPW

- Should we use IPW or IW weighting for fitting $f(x)$?
- Based on asymptotics, [CTD09] suggests weighting by $\frac{1-\pi(X_i)}{\pi(X_i)^2}$!
- Would results improve with cross-fitting?
- Should we be using a regression approach with $\hat{\beta}_{\text{opt}}$, as in the control variate module?

All these would be interesting directions for projects.

References

Resources

- A quick introduction to AIPW, similar to our treatment here, can be found in [Tsi06, Ch 6].
- The early sections of [KS07] and [SV18] also give introductions to AIPW estimators that may give additional flavor.
- These slides give the most accessible summary I've found on the asymptotics of the estimators we've discussed.

References I

[CTD09] Weihua Cao, Anastasios A. Tsiatis, and Marie Davidian, *Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data*, Biometrika **96** (2009), no. 3, 723–734.

[KS07] Joseph D. Y. Kang and Joseph L. Schafer, *Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data*, Statistical Science **22** (2007), no. 4, 523–539.

[SV18] Shaun R. Seaman and Stijn Vansteelandt, *Introduction to double robust methods for incomplete data*, Statistical Science **33** (2018), no. 2, 184–197.

[Tsi06] Anastasios A. Tsiatis, *Semiparametric theory and missing data*, Springer, 2006.