Introduction to causal inference with machine learning

Chapter 4:
Efficient, doubly-robust estimation of an average treatment effect

David Benkeser
Emory University

Iván Díaz
Weill Cornell Medical College

6th Seattle Symposium in Biostatistics

November 14, 2020

Contents of this chapter

1. Limitations of naive G-computation and IPTW estimation
2. Augmented IPTW estimation
3. Targeted minimum loss-based estimation
4. Doubly-robust inference
5. Use in randomized trials

Limitations of naive G-computation or IPTW estimation

The G-computation and IPTW formulas are useful because they allow us to connect the counterfactual and observable worlds.

We require consistent estimation of the outcome regression or propensity score as an intermediate step. For this reason, we need a flexible strategy (e.g., Super Learner).

Even then, does naive use of these formulas lead to good estimators?

These naive estimators have two important shortcomings.

1. **Lack of robustness:**
   - G-computation estimators rely heavily on correct estimation of the outcome regression.
   - IPTW estimators rely heavily on correct estimation of the propensity score.
2. **Lack of basis for statistical inference:**
   - What is the distribution of naive estimators when $\bar{Q}$ and g are estimated flexibly?
   - The bootstrap is then also generally invalid.

Augmented IPTW estimation

To address lack of robustness, can we combine G-computation and IPTW estimators?

The **augmented IPTW (AIPTW) estimator** of ψ_1 is defined as

$$\psi_{n,AIPTW,1} \;:=\; \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{I(A_i=1)}{g_n(W_i)}\right]Y_i}_{\text{IPTW estimator}} \;+\; \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left[1-\frac{I(A_i=1)}{g_n(W_i)}\right]\bar{Q}_n(1,W_i)}_{\text{augmentation term}} \;.$$

Augmentation seeks to rectify any incorrect estimation of g in the IPTW estimator.

Suppose Y is non-negative and g_n underestimates g throughout.

On average, the IPTW estimator overshoots the target but the augmentation term is negative and brings it back down on target.

Augmented IPTW estimation

To address lack of robustness, can we combine G-computation and IPTW estimators?

The AIPTW estimator can also be seen as an augmented G-computation estimator as

$$\psi_{n,AIPTW,1} = \underbrace{\frac{1}{n}\sum_{i=1}^{n} \bar{Q}_n(1, W_i)}_{\text{G estimator}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n} \frac{I(A_i = 1)}{g_n(W_i)} \left[Y_i - \bar{Q}_n(1, W_i)\right]}_{\text{augmentation term}} .$$

Augmentation seeks to rectify any incorrect estimation of $\bar{Q}$ in the G-computation estimator.

Suppose $\bar{Q}_n$ overestimates $\bar{Q}$ throughout.

On average, the G-computation estimator overshoots the target but the augmentation term is negative and brings it back down on target.

Augmented IPTW estimation

This idea can be made precise through the concept of double robustness, a property enjoyed by the AIPTW estimator.

Suppose that $\bar{Q}_n \xrightarrow{P} \bar{Q}_*$ and $g_n \xrightarrow{P} g_*$, where $\bar{Q}_*$ and g_* are not necessarily equal to the true values $\bar{Q}$ and g. In large samples, we expect that

$$\begin{aligned}
\psi_{n,AIPTW,1} &= \frac{1}{n}\sum_{i=1}^{n} \bar{Q}_n(1, W_i) + \frac{1}{n}\sum_{i=1}^{n} \frac{I(A_i = 1)}{g_n(W_i)}\left[Y_i - \bar{Q}_n(1, W_i)\right] \\
&\approx \frac{1}{n}\sum_{i=1}^{n} \bar{Q}_*(1, W_i) + \frac{1}{n}\sum_{i=1}^{n} \frac{I(A_i = 1)}{g_*(W_i)}\left[Y_i - \bar{Q}_*(1, W_i)\right] \\
&\approx E\left[\bar{Q}_*(1, W)\right] + E\left\{\frac{I(A = 1)}{g_*(W)}\left[Y - \bar{Q}_*(1, W)\right]\right\}
\end{aligned}$$

In fact, under mild conditions, it will be true that

$$\psi_{n,AIPTW,1} \xrightarrow{P} H(\bar{Q}_*, g_*) := E\left\{\bar{Q}_*(1, W) + \frac{I(A = 1)}{g_*(W)}\left[Y - \bar{Q}_*(1, W)\right]\right\}.$$

Augmented IPTW estimation

We must take a closer look at this limit.

Through repeated uses of the law of total expectation, we observe that

$$\begin{aligned}
H(\bar{Q}_*, g_*) &= E\left[E\left\{\bar{Q}_*(1,W) + \frac{I(A=1)}{g_*(W)}\left[Y - \bar{Q}_*(1,W)\right] \,\middle|\, A, W\right\}\right] \\
&= E\left\{\bar{Q}_*(1,W) + \frac{I(A=1)}{g_*(W)}\left[\bar{Q}(1,W) - \bar{Q}_*(1,W)\right]\right\} \\
&= E\left[E\left\{\bar{Q}_*(1,W) + \frac{I(A=1)}{g_*(W)}\left[\bar{Q}(1,W) - \bar{Q}_*(1,W)\right] \,\middle|\, A\right\}\right] \\
&= E\left\{\bar{Q}_*(1,W) + \frac{g(W)}{g_*(W)}\left[\bar{Q}(1,W) - \bar{Q}_*(1,W)\right]\right\}.
\end{aligned}$$

It follows directly that $H(\bar{Q}_*, g_*) = \psi_1$ if either $\bar{Q}_* = \bar{Q}$ or $g_* = g$.

Augmented IPTW estimation

This observation indicates that

the AIPTW estimator is consistent for ψ_1
if at least one of $\bar{Q}_n$ and g_n used in its construction is itself consistent.

The same is true for the AIPTW estimator of ψ_0 and of the ATE $\gamma = \psi_1 - \psi_0$.

This property is **double robustness**, and the AIPTW estimator is called doubly-robust.

Colloquially, we say that we have two chances to get it right!

A few comments on double robustness:

- The extra robustness is not an excuse to use simplistic estimation strategies (e.g., parametric models) since both $\bar{Q}_n$ and g_n then generally fail to be consistent.
- While this is really an asymptotic property, it has direct relevance in practice.
- As we shall see soon, it is intimately tied to efficiency in this problem.

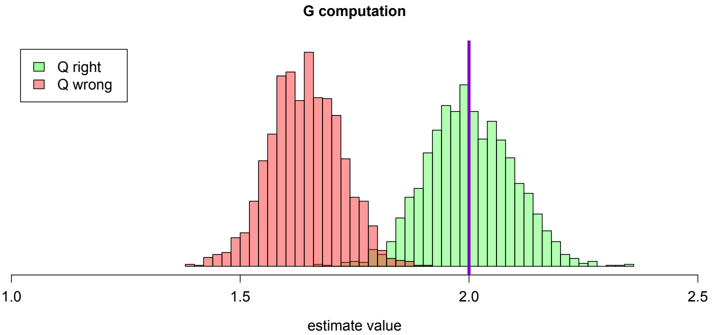
Augmented IPTW estimation

```r
# fit super learner to outcome
fit_or <- SuperLearner(Y = Y, X = data.frame(A,W),
                       SL.library = SL.lib,
                       method="method.CC_LS")

# fit super learner to propensity
fit_ps <- SuperLearner(Y = A, X = data.frame(W),
                       SL.library = SL.lib,
                       method = "method.CC_LS")

# get super learner fits
Qbar1 <- predict(fit_or, newdata = data.frame(A=1, W))
Qbar0 <- predict(fit_or, newdata = data.frame(A=0, W))
g1W <- fit_ps$SL.predict

# compute gcomp + augmentation
psi_nAIPTW1 <- mean(Qbar1) +
                  mean(as.numeric(A==1)/g1W * (Y - Qbar1))
psi_nAIPTW0 <- mean(Qbar0) +
                  mean(as.numeric(A==0)/(1-g1W) * (Y - Qbar0))

# compute ate
gamma_nAIPTW <- psi_nAIPTW1 - psi_nAIPTW
```

Augmented IPTW estimation

$$W \sim U(-2,+2),\ A \mid W = w \sim \text{Bernoulli}(g(w)),\ Y \mid A = a, W = w \sim N(\bar{Q}(a,w), \sigma^2)$$
$$\text{with } g(w) := \text{expit}\left[\tfrac{3}{2}(w+1)^2 - 3\right] \text{ and } \bar{Q}(a,w) := 1 + a - w - aw$$

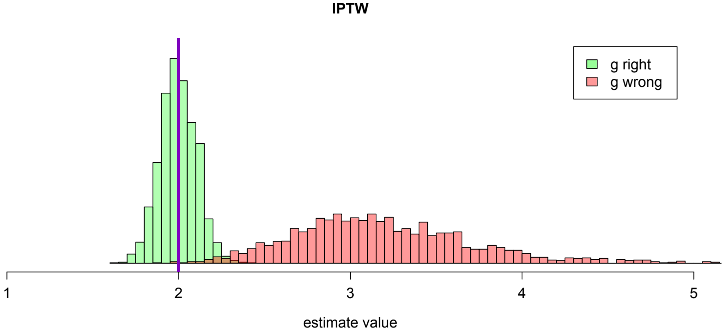
In this simulation study, we contrast estimating $\bar{Q}$ using linear regression with main terms only ($\bar{Q}$ wrong) vs also including an interaction ($\bar{Q}$ right).



Augmented IPTW estimation

$$W \sim U(-2,+2),\ A \mid W = w \sim \text{Bernoulli}(g(w)),\ Y \mid A = a, W = w \sim N(\bar{Q}(a,w), \sigma^2)$$
$$\text{with } g(w) := \text{expit}\left[\tfrac{3}{2}(w+1)^2 - 3\right] \text{ and } \bar{Q}(a,w) := 1 + a - w - aw$$

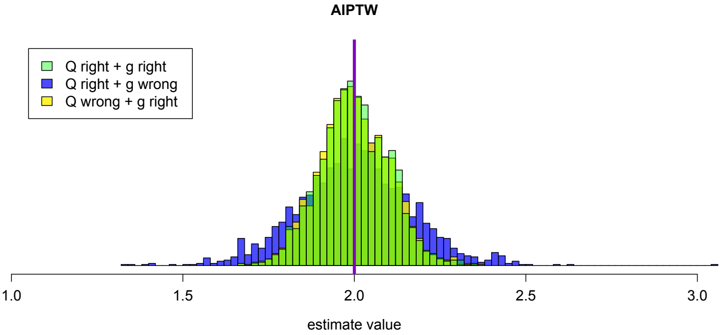
In this simulation study, we contrast estimating $\bar{g}$
using logistic regression with main terms only (g wrong) vs also quadratic terms (g right).



Augmented IPTW estimation

$$W \sim U(-2,+2),\ \ A \mid W = w \sim \text{Bernoulli}(g(w)),\ \ Y \mid A = a, W = w \sim N(\bar{Q}(a,w), \sigma^2)$$
$$\text{with } g(w) := \text{expit}\left[\tfrac{3}{2}(w+1)^2 - 3\right] \text{ and } \bar{Q}(a,w) := 1 + a - w - aw$$

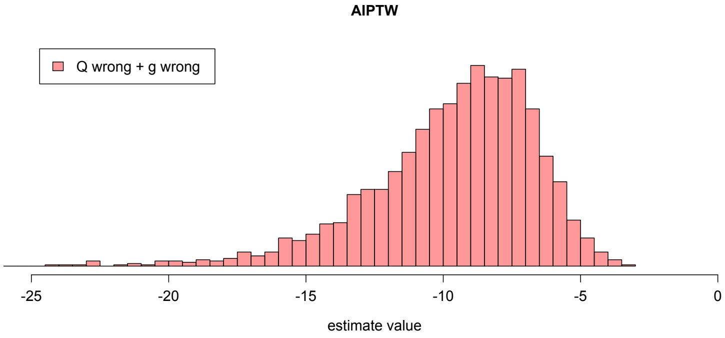
In this simulation study, we contrast different patterns of inconsistent estimation of either $\bar{Q}$ and g.



Augmented IPTW estimation

$$W \sim U(-2,+2),\ A \mid W = w \sim \text{Bernoulli}(g(w)),\ Y \mid A = a, W = w \sim N(\bar{Q}(a,w), \sigma^2)$$
$$\text{with } g(w) := \text{expit}\left[\tfrac{3}{2}(w+1)^2 - 3\right] \text{ and } \bar{Q}(a,w) := 1 + a - w - aw$$

In this simulation study, we contrast different patterns of inconsistent estimation of either $\bar{Q}$ and g.



Augmented IPTW estimation

If $\bar{Q}_n$ and g_n are both consistent, then $\psi_{n,AIPTW,1} - \psi_1$ can be approximated by

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{I(A_i=1)}{g(W_i)}\left[Y-\bar{Q}(1,W_i)\right]+\bar{Q}(1,W_i)-\psi_1\right\}$$

under certain regularity conditions. In particular, this implies that

$$n^{1/2}\left(\psi_{n,AIPTW,1}-\psi_1\right)\xrightarrow{d}N(0,\tau_1^2)\ ,$$

where the asymptotic variance τ_1^2 can be estimated consistently using

$$\tau_{n,1}^2 := \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{I(A_i=1)}{g_n(W_i)}\left[Y_i-\bar{Q}_n(1,W_i)\right]+\bar{Q}_n(1,W_i)-\psi_{n,AIPTW,1}\right\}^2\ .$$

As a consequence, for example, an approximate 95% CI for ψ_1 is given by

$$\left(\psi_{n,AIPTW,1}-1.96n^{-1/2}\tau_{n,1},\ \psi_{n,AIPTW,1}+1.96n^{-1/2}\tau_{n,1}\right).$$

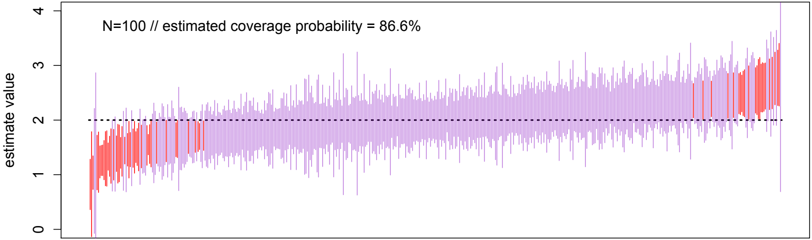
Augmented IPTW estimation

```
# compute asymptotic variance estimate of psi_nAIPTW1
tau2_n1 <-  mean((as.numeric(A==1)/g1W * (Y - Qbar1) +
                                Qbar1 - psi_nAIPTW1)^2)

# confidence interval for psi_n1APTW
ci1 <- c(psi_nAIPTW1 - 1.96*sqrt(tau2_n1/n),
         psi_nAIPTW1 + 1.96*sqrt(tau2_n1/n))

# compute asymptotic variance estimate of psi_nAIPTW0
tau2_n0 <-  mean((as.numeric(A==0)/(1-g1W) * (Y - Qbar0) +
                                        Qbar0 - psi_nAIPTW0)^2)

# confidence interval for psi_n1APTW
ci0 <- c(psi_nAIPTW0 - 1.96*sqrt(tau2_n0/n),
         psi_nAIPTW0 + 1.96*sqrt(tau2_n0/n))
```

Augmented IPTW estimation

$$W \sim U(-2,+2),\ A \mid W = w \sim \text{Bernoulli}(g(w)),\ Y \mid A = a, W = w \sim N(\bar{Q}(a,w),\sigma^2)$$
$$\text{with } g(w) := \text{expit}\left[\tfrac{3}{2}(w+1)^2 - 3\right] \text{ and } \bar{Q}(a,w) := 1 + a - w - aw$$

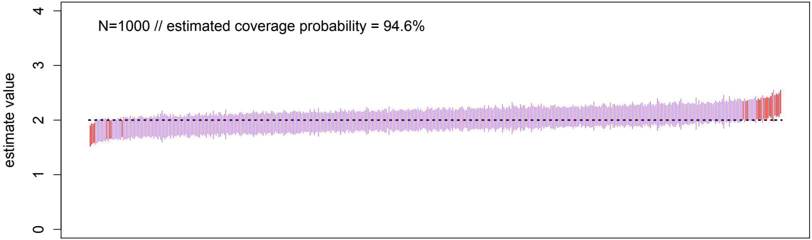
In this simulation study, correct regression models were used for both $\bar{Q}$ and g.



Augmented IPTW estimation

$$W \sim U(-2,+2),\ A \mid W = w \sim \text{Bernoulli}(g(w)),\ Y \mid A = a, W = w \sim N(\bar{Q}(a,w), \sigma^2)$$
$$\text{with } g(w) := \text{expit}\left[\tfrac{3}{2}(w+1)^2 - 3\right] \text{ and } \bar{Q}(a,w) := 1 + a - w - aw$$

In this simulation study, correct regression models were used for both $\bar{Q}$ and g.



Augmented IPTW estimation

What about improved estimation of the ATE and testing of $H_0 : ATE = 0$?

For ease of notation, we define, for $i = 1, 2, \ldots, n$,

$$D_{i,n} := \bar{Q}_n(1, W_i) - \bar{Q}_n(0, W_i) + \frac{A_i}{g_n(W_i)} \left[Y - \bar{Q}_n(1, W_i) \right] - \frac{1 - A_i}{1 - g_n(W_i)} \left[Y - \bar{Q}_n(0, W_i) \right].$$

The AIPTW estimator of the ATE is

$$\gamma_{n,AIPTW} := \overline{D}_n = \frac{1}{n} \sum_{i=1}^{n} D_{i,n}$$

and its variance can be approximated by $\sigma_n^2 := \tau_n^2 / n$, where $\tau_n^2 := \frac{1}{n} \sum_{i=1}^{n} (D_{i,n} - \overline{D}_n)^2$ is the empirical variance of $D_{1,n}, D_{2,n}, \ldots, D_{n,n}$.

As before, Wald CIs can be easily constructed, and an approximate p-value of the test of $H_0 : ATE = 0$ versus $H_1 : ATE \neq 0$ can be obtained as

$$p = 2 \left[1 - \Phi \left(\frac{|\gamma_{n,AIPTW}|}{\sigma_n} \right) \right],$$

where Φ is the distribution function of the standard normal distribution.

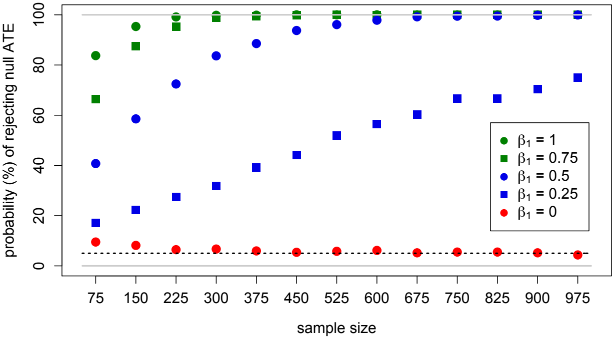
Augmented IPTW estimation

```r
# compute asymptotic variance estimate of gamma_nAIPTW
D <- Qbar1 - Qbar0 + A/g1W * (Y - Qbar1) -
      (1-A)/(1-g1W) * (Y - Qbar0)

tau2_n <- var(D)

sigma_n <- sqrt(tau2_n/n)

# confidence interval for psi_n0APTW
ci <- c(gamma_nAIPTW - 1.96*sigma_n,
        gamma_nAIPTW + 1.96*sigma_n)

# upper tail wald test p-value
pval <- 2 * (1 - pnorm(abs(gamma_nAIPTW/sigma_n)))
```

Augmented IPTW estimation

$$W \sim U(-1,+1),\, A \mid W = w \sim \text{Bernoulli}(g(w)),\, Y \mid A = a, W = w \sim N(\bar{Q}(a,w), \sigma^2)$$
$$\text{with } g(w) := \text{expit}(3w) \text{ and } \bar{Q}(a,w) := \beta_0 + \beta_1 a + \beta_2 w$$

(In the simulations below, we set $\beta_0 = 1$ and $\beta_2 = -1$.)



We have discussed several properties we may wish an estimator to have. One property we have not discussed yet is whether it is a compatible plug-in estimator or not.

What is a compatible plug-in estimator?

An estimator ψ_n is said to be a compatible plug-in estimator if there exists a fixed parameter mapping Ψ and a distribution $\widehat{P}_n$ for the data unit O such that $\psi_n = \Psi(\widehat{P}_n)$.

The sample mean is the most popular example of a compatible plug-in estimator.

To check this property, we can verify whether or not ψ_n uses two different estimators of the same (or related) component of the joint distribution P_0 of the data unit O.

Which of the following estimators are compatible plug-in estimators?

- the G-computation estimator $\psi_{n,G,1} = \frac{1}{n}\sum_{i=1}^{n} \bar{Q}_n(1, W_i)$?
- the IPTW estimator $\psi_{n,IPTW,1} = \frac{1}{n}\sum_{i=1}^{n} \frac{I(A_i=1)}{g_n(W_i)} Y_i$?
- the AIPTW estimator $\psi_{n,IPTW,1} = \psi_{n,G,1} + \frac{1}{n}\sum_{i=1}^{n} \frac{I(A_i=1)}{g_n(W_i)}[Y_i - \bar{Q}_n(1, W_i)]$?

Why are compatible plug-in estimators preferred?

- they satisfy natural parameter constraints (e.g., probability between 0 and 1);
- they often are more difficult to derail (e.g., in near-violations of positivity);
- they often have reduced mean-squared error in small samples.

To find a compatible plug-in counterpart of the AIPTW estimator, we could try to find a smarter estimator $\bar{Q}_n^*$ of $\bar{Q}$ such that the revised G-computation estimator

$$\frac{1}{n}\sum_{i=1}^{n}\bar{Q}_n^*(1,W_i)$$

is a (nonparametric) efficient, doubly-robust estimator of ψ_1.

This can be shown to happen if $\bar{Q}_n^*$ is both a good estimator of $\bar{Q}$ and satisfies

$$0 \;=\; B_n(Q_n^*,g_n) := \frac{1}{n}\sum_{i=1}^{n}\frac{I(A_i=1)}{g_n(W_i)}\left[Y_i-\bar{Q}_n^*(1,W_i)\right].$$

Targeted minimum loss-based estimation

Given an estimator $\bar{Q}_n$ of $\bar{Q}$, we can construct such a $\bar{Q}_n^*$ using the framework of targeted minimum loss-based estimation (TMLE).

For simplicity, suppose that the outcome Y is bounded between 0 and 1.

Implementation of TMLE algorithm for ψ_1:

1. Get good estimates $\bar{Q}_n$ and g_n (e.g., using Super Learner with a rich library).
2. Run logistic regression with outcome Y, single covariate $Z := \frac{A}{g_n(W)}$ and offset $K := \text{logit}\, \bar{Q}_n(1, W)$ using the subset of the data with $A = 1$.
3. Save the fitted coefficient α_n of Z.
4. Set $\bar{Q}_n^*(a, w) = \text{expit}\left[\text{logit}\, \bar{Q}_n(a, w) + \alpha_n \frac{a}{g_n(w)}\right]$, the TMLE of $\bar{Q}$.

We note that this only changes $\bar{Q}_n(1, w)$ but not $\bar{Q}_n(0, w)$.

This algorithm can be implemented using standard software for logistic regression, and has been shown to have better performance in some contexts (Porter et al., 2011).

Targeted minimum loss-based estimation

In general, we would care about estimating each of ψ_0, ψ_1 and γ.

Can we find a revised (i.e., targeted) estimator $\bar{Q}_n^*$ that simultaneously leads to optimal estimators of each of these targets?

Implementation of TMLE algorithm for ψ_0, ψ_1 and γ:

1. Get good estimates $\bar{Q}_n$ and g_n (e.g., using Super Learner with a rich library).
2. Run logistic regression with outcome Y, covariates $Z^0 := \frac{1-A}{1-g_n(W)}$ and $Z^1 := \frac{A}{g_n(W)}$, and offset $K :=$ logit $\bar{Q}_n(A, W)$ using the entire dataset.
3. Save the fitted coefficients α_n^0 of Z^0 and α_n^1 of Z^1.
4. Set $\bar{Q}_n^*(a, w) = \text{expit}\left\{\text{logit}\, \bar{Q}_n(a, w) + \alpha_n^0 \left[\frac{1-a}{1-g_n(w)}\right] + \alpha_n^1 \left[\frac{a}{g_n(w)}\right]\right\}$, the TMLE of $\bar{Q}$.

Confidence intervals and p-values can be obtained as for the AIPTW estimator.

Targeted minimum loss-based estimation

```r
# get Qbar(A,W) from super learner fit
QbarA <- fit_or$SL.predict

# create covariates
Z1 <- A/g1W; Z0 <- (1-A)/(1-g1W)

# create scaled outcome
l <- min(Y); u <- max(Y)
Ystar <- (Y - l)/(u-l)

# fit logistic regression, ignore warnings about non 0/1 outcome
logistic_fit <- glm(Ystar ~ -1 + offset(qlogis(QbarA)) + Z0 + Z1,
                    family = binomial())

# save fitted coefficients
alpha <- coef(logistic_fit)

# compute Qbarstar1 Qbarstar0 by rescaling
Qbarstar0 <- (u-l)*plogis(qlogis(Qbar0) + alpha[1]/(1-g1W)) + l
Qbarstar1 <- (u-l)*plogis(qlogis(Qbar1) + alpha[2]/g1W) + l

# see AIPTW slides for computing ci and pvalues by
# swapping in Qbarstar0 and Qbarstar1 for Qbar0 and Qbar1.
```

Targeted minimum loss-based estimation

We analyzed data from the BOLD study using TMLE and estimation of the propensity score and outcome regression using the super learner (as described in Chapter 3).

Average counterfactual score corresponding to early imaging intervention:

estimate = 8.13, 95% CI: (7.90, 8.35)

Average counterfactual score corresponding to control (no early imaging):

estimate = 8.56, 95% CI: (8.34, 8.77)

Average treatment effect comparing early imaging to control:

estimate = -0.43, 95% CI: (-0.66, -0.19), $p < 0.001$

Based on these results, we would conclude that obtaining early imaging appears to lower disability scores on average at the 12-month mark.

Doubly-robust inference

The AIPTW and TMLE estimators discussed so far are said to be doubly-robust. To be precise though, we should say that they enjoy doubly-robust consistency.

When both $\bar{Q}_n$ and g_n are consistent, we can easily construct valid CIs and p-values.

What about when only one of $\bar{Q}_n$ and g_n is consistent?

- If parametric models are used, the bootstrap can be safely used.
- If flexible estimation techniques are used, we are out of luck – there is generally no nice limit distribution we can lean on.

Doubly-robust inference (i.e., CIs and p-values) therefore appears difficult to perform.

This likely requires a tractable distribution even if one of $\bar{Q}_n$ or g_n is inconsistent.

Doubly-robust inference

If flexible estimation techniques are employed, the usual AIPTW and TMLE estimators do not easily allow inference when only of $\bar{Q}_n$ and g_n is inconsistent.

It is possible to use the TMLE framework to construct an estimator that

1. is efficient when both $\bar{Q}_n$ and g_n are consistent;
2. is consistent when at least one of $\bar{Q}_n$ and g_n is consistent;
3. when suitably normalized, tends to a mean-zero normal distribution with variance we can consistently estimate, even when only one of $\bar{Q}_n$ or g_n is consistent.

It does not appear possible to adapt the AIPTW estimator for this purpose.

Details are provided in Benkeser, Carone, van der Laan & Gilbert (2017).

Doubly-robust inference

From initial estimates $\bar{Q}_n$ and g_n, we must find revised estimates $\bar{Q}_n^*$ and g_n^* that not only satisfy $B_n(Q_n^*, g_n^*) = 0$ but also

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\bar{Q}_{n,r}^*(W_i)}{g_n^*(W_i)}\left[A_i - g_n^*(W_i)\right] = 0 = \frac{1}{n}\sum_{i=1}^{n} A_i \frac{g_{2,n,r}^*(W_i)}{g_{1,n,r}^*(W_i)}\left[Y_i - \bar{Q}_n^*(W_i)\right],$$

where $\bar{Q}_{n,r}$, $g_{1,n,r}$ and $g_{2,n,r}$ are consistent estimators of

$$\begin{aligned}
\bar{Q}_{0,r}(w) &:= E[Y - \bar{Q}_n^*(W) \mid g_n^*(W) = g_n^*(w)] \\
g_{1,0,r}(w) &:= E[A \mid \bar{Q}_n^*(W) = \bar{Q}_n^*(w)] \\
g_{2,0,r}(w) &:= E[\{A - g_n^*(W)\}/g_n^*(W) \mid \bar{Q}_n^*(W) = \bar{Q}_n^*(w)] \ ,
\end{aligned}$$

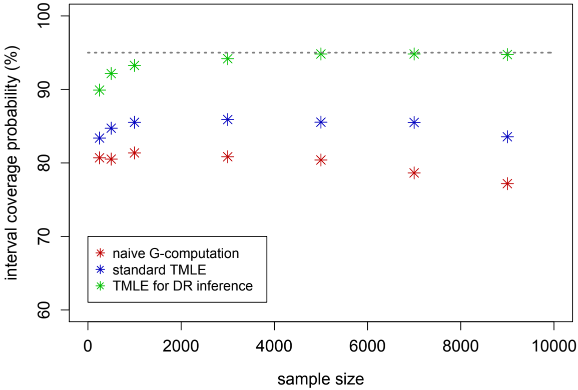
and $\bar{Q}_n^*$ and g_n^* are considered fixed in the definition of these conditional expectations.

In contrast to the standard TMLE, in this case, both $\bar{Q}_n$ and g_n need to be updated, and the updating process is iterative.

Doubly-robust inference

$$W = (W_1, W_2) \text{ is generated as } W_1 \sim U(-2, +2),\ \ W_2 \sim \text{Bernoulli}(0.5) \text{ and } W_1 \perp W_2,$$
$$A \mid W = w \sim \text{Bernoulli}(g(w)),\quad Y \mid A = a, W = w \sim \text{Bernoulli}(\bar{Q}(a, w))$$
$$\text{with } g(w) := \text{expit}(-w_1 + 2w_1 w_2) \text{ and } \bar{Q}(a, w) := \text{expit}(0.2a - w_1 + 2w_1 w_2)$$

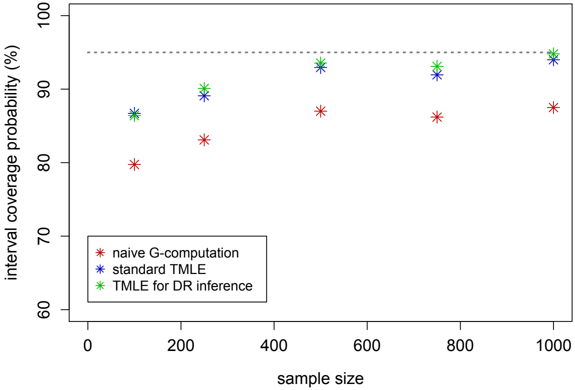
In this simulation study, a logistic regression model with main terms only (i.e., incorrect) is used for g, while a nonparametric kernel smoother is used to estimate $\bar{Q}$.



Doubly-robust inference

$$W = (W_1, W_2) \text{ is generated as } W_1 \sim U(-2,+2),\ W_2 \sim \text{Bernoulli}(0.5) \text{ and } W_1 \perp W_2,$$
$$A \mid W = w \sim \text{Bernoulli}(g(w)),\quad Y \mid A = a, W = w \sim \text{Bernoulli}(\bar{Q}(a,w))$$
$$\text{with } g(w) := \text{expit}(-w_1 + 2w_1w_2) \text{ and } \bar{Q}(a,w) := \text{expit}(0.2a - w_1 + 2w_1w_2)$$

In this simulation study, a nonparametric kernel smoother is used for g and $\bar{Q}$.



Use in randomized trials

We have been focusing on the analysis of data from observation studies. In such settings, adjustment for confounding is mandatory. However, even in randomized trials, the methods discussed can be useful.

In a randomized trial, use of the G-computation formula is not necessary since then

$$\gamma = ATE = E(Y \mid A = 1) - E(Y \mid A = 0) \ ,$$

and the unadjusted estimator $\gamma_{n,unadjusted} := \frac{\sum_i Y_i I(A_i=1)}{\sum_i I(A_i=1)} - \frac{\sum_i Y_i I(A_i=0)}{\sum_i I(A_i=0)}$ is consistent.

However, we could still use the AIPTW or TMLE estimators with g_n equal to the true propensity g, known by design (usually $g(w) = 0.5$ for each w).

What are the pros and cons of doing so?

Use in randomized trials

A few notes on using AIPTW or TMLE estimators in randomized trials:

- ⊕ Because g is exactly know, consistency is guaranteed by double robustness.
- ⊕ Valid confidence intervals and p-values can be obtained even if $\bar{Q}_n$ is inconsistent.
- ⊕ If baseline covariates are predictive of the outcome, their inclusion generally yields tighter confidence intervals / more powerful tests.

  The relative decrease in variance is given by

$$\frac{\text{var}(\gamma_{n,unadjusted}) - \text{var}(\gamma_{n,TMLE})}{\text{var}(\gamma_{n,unadjusted})} = \frac{\text{var}\left[\frac{\bar{Q}(0,W)+\bar{Q}(1,W)}{2}\right]}{\left[\frac{\text{var}(Y|A=0)+\text{var}(Y|A=1)}{2}\right]} .$$

- ⊖ Estimating $\bar{Q}$ requires effort not needed in the unadjusted approach.
- ⊖ If the baseline covariates are not predictive at all, or a very poor estimator $\bar{Q}_n$ is used, it is possible to have slightly decreased efficiency.

See Moore & van der Laan (2009) for more details.

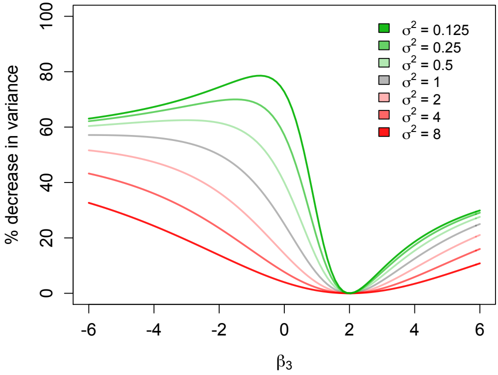
Use in randomized trials

An illustration of potential gains:

$$W \sim U(-1,+1), \quad A \mid W = w \sim \text{Bernoulli}(0.5), \quad Y \mid A = a, W = w \sim N(\bar{Q}(a,w), \sigma^2)$$

$$\text{with} \quad \bar{Q}(a,w) := \beta_0 + \beta_1 a + \beta_2 w + \beta_3 aw$$

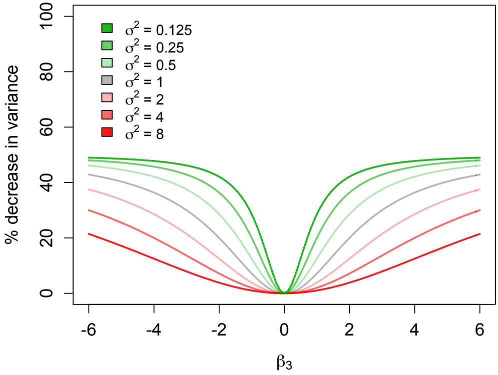
(Throughout, we set $\beta_2 = -1$ as a reference. Both β_0 and β_1 are irrelevant here.)



An illustration of potential gains:

$$W \sim U(-1,+1), \quad A \mid W = w \sim \text{Bernoulli}(0.5), \quad Y \mid A = a, W = w \sim N(\bar{Q}(a,w), \sigma^2)$$
$$\text{with} \quad \bar{Q}(a,w) := \beta_0 + \beta_1 a + \beta_2 w + \beta_3 aw$$

(Throughout, we set $\beta_2 = 0$ as a reference. Both β_0 and β_1 are irrelevant here.)

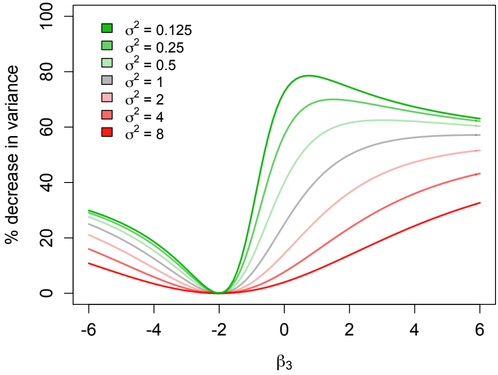


Use in randomized trials

An illustration of potential gains:

$$W \sim U(-1,+1), \quad A \mid W = w \sim \text{Bernoulli}(0.5), \quad Y \mid A = a, W = w \sim N(\bar{Q}(a,w), \sigma^2)$$
$$\text{with} \quad \bar{Q}(a,w) := \beta_0 + \beta_1 a + \beta_2 w + \beta_3 aw$$

(Throughout, we set $\beta_2 = 1$ as a reference. Both β_0 and β_1 are irrelevant here.)



Key points of Chapter 4

- If flexible estimators of $\bar{Q}$ and/or g are used, naive estimators based on G-computation or IPTW do not generally allow for valid inference!
- G-computation and IPTW estimators do not offer any robustness.
- The AIPTW estimator is a doubly-robust hybrid of the two approaches.
- Double robustness generally refers to consistency even when one of $\bar{Q}$ and g is incorrectly estimated.
- Inference can be carried out using sandwich variance estimator.
- The AIPTW estimator may suffer from not being a compatible plug-in estimator.
- The TMLE algorithm can be used to produce a clever $\bar{Q}$ estimator such that the resulting G-computation estimator is doubly-robust.
- It can also be used to produce an estimator that allows doubly-robust inference.
- Though the methods we have discussed are necessary for analyzing observational studies, they can also be used to gain efficiency in randomized trials.

References and additional reading

References:

Benkeser D, Carone M, van der Laan MJ, Gilbert P (2017). Doubly-robust nonparametric inference on the average treatment effect. *Biometrika*; 104(4)863-880. doi: 10.1093/biomet/asx053.

Moore K & van der Laan MJ (2009). Covariate adjustment in randomized trials with binary outcomes: targeted maximum likelihood estimation. *Statistics in Medicine*, 28(1)39-64. doi: 10.1002/sim.3445.

Porter K, Gruber S, van der Laan MJ, Sekhon JS (2011). The relative performance of targeted maximum likelihood estimators. *International Journal of Biostatistics*, 7(1). doi: 10.2202/1557-4679.1308.

Robins JM (1986). A new approach to causal inference in mortality studies with a sustained exposure period – application to control of the healthy worker survivor effect. *Mathematical Modelling*, 7(9–12)1393-1512. doi: 10.1016/0270-0255(86)90088-6.

Robins JM, Rotnitzky A, Zhao LP (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427)846-866. doi: 10.1080/01621459.1994.10476818.

van der Laan MJ, Rubin D (2006). Targeted maximum likelihood learning. *The International Journal of Biostatistics*, 2(1). doi: 10.2202/1557-4679.1043.

Additional reading:

Kang, JDY, Schafer, JL (2007). Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22(4)523-539. doi: 10.1214/07-STS227

van der Laan, MJ & Rose, S (2011). Targeted learning: causal inference for observational and experimental data; Chapters 4–6, 11. Springer New York. doi: 10.1007/978-1-4419-9782-1.