Robustifying doubly-robust estimators

Stephen Cristiano

02 May, 2019

Introduction

Consider the following setting:

- We wish to estimate the mean of Y, an outcome of interest, given observed covariates X.
- Incomplete data: $R \in 0, 1$ is an indicator, where $R = 0$ when Y is missing
- We assume $P(R = 1|Y, X) = P(R = 1|X) = P(X)$, i.e. missing at random.

Recall a doubly robust estimator for the mean, μ, has the form:

$$\hat{\mu}_{dr} = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{R_i Y_i}{\pi(X_i, \hat{\gamma})} - \frac{R_i - \pi(X_i, \hat{\gamma})}{\pi(X_i, \hat{\gamma})} m(X, \hat{\beta}) \right\}$$

Where $\hat{\beta}$ is estimated via complete cases regression, i.e. condition on $R = 1$, $\pi(X_i, \hat{\gamma})$ are the propensity scores estimated via logistic regression, and $m(X, \hat{\beta})$ is the estimated mean of the regression. Doubly robust estimators have the property where if either the outcome regression model or the propensity score model is misspecified (but not both), $\hat{\mu}_{dr}$ is still a consistent estimator for μ.

When outliers are present and also dependent on the distribution of X, doubly robust estimators become very unstable with regards to MSE. In particular, inverse probability weighting by the propensity score poses a problem when outliers exist, as it can lead to values close to zero in the denominator and hence unreliable estimation. Our goal is to modify the doubly robust estimator by weighting by some function of the influence each observation has on the overall estimates to mitigate the the effect outlying samples may have.

A common approach in robust regression settings for parameter estimation while mitigating the effect of outliers is to define a score function in which influential observations are down-weighted with regards to some weighting function. An iterated re-weighted least squares algorithm is used to iteratively update the weights and estimates for parameters of interest until some convergence rule.

Model

To robustly solve for μ, we set up a vector of weighted of estimating equations:

$$U(Y, X; \gamma, \beta, \mu, w) = \begin{bmatrix} u_1(R, Y, X; \gamma, \beta, \mu, w) \\ u_2(R, X; \gamma, w) \\ u_3(R, Y, X; \beta, w) \end{bmatrix}$$

Where u_1, u_2, u_3 are the estimating equations for μ, γ, and β, respectively.

Our goal is to iteratively solve

$$\begin{aligned} u_1 &= \sum_i w_i \left(\frac{RY}{\pi(x_i, \gamma)} - \frac{R - \pi(x_i, \gamma)}{R - \pi(x_i, \gamma)} m(X\beta) - \mu \right) = 0 \\ u_2 &= \sum_i w_i \left(R_i - \pi(x_i, \gamma) \right) x_i = 0 \\ u_3 &= \sum_i w_i \left(Y_i - m(x_i, \beta) \right) x_i = 0 \end{aligned}$$

Define

$$\hat{A} = \frac{1}{n} \sum_{i=1}^{n} U_i(y_i, x_i; \theta) U_i^t(y_i, x_i; \theta).$$

For notational convenience, let $\theta = (\gamma, \beta, \mu)$. Define the norm $\|\cdot\|_M$ of a $p \times p$ positive definite matrix M on $\mathbb{R}^p$ by

$$\|x\|_M = [x^t M^{-1} x]^{1/2}$$

Let ψ be Hampel's weight function, defined as

$$\psi(u) = \begin{cases} 1 & \text{if } |u| < a \\ \frac{a}{|u|} & \text{if } a \le |u| < b \\ a \frac{c/|u| - 1}{c - b} & \text{if } b \le |u| < c \\ 0 & \text{otherwise} \end{cases}$$

We apply ψ with $a = 2, b = 4, c = 8$ to weight each observation according to its influence with the rule

$$w_i = \psi \left(\frac{\|U_i\|_{\hat{A}}}{1.48 \times \text{median}(\|U_i\|_{\hat{A}})} \right)$$

Using an iterated reweighted least squares approach for estimation, our algorithm for solving $\hat{\theta}$ is as follows:

---

1. Fix $\epsilon = 0.0001$. Let $s = 1$ be the current iteration. Set our starting values:

$$\begin{aligned}
w_i^{(1)} &= 1 \text{ for all } i \\
\tilde{\beta}^{(1)} &= (X_{R=1}^t X_{R=1})^{-1} X_{R=1}^t Y_{R=1} \\
\tilde{\gamma}^{(1)} &= \text{coefficents of logistic regression} \\
\tilde{\mu}^{(1)} &= \frac{1}{n} \sum_{i=1}^{n} \left(\frac{RY}{\pi(x_i, \tilde{\gamma}^{(1)})} - \frac{R - \pi(x_i, \tilde{\gamma}^{(1)})}{R - \pi(x_i, \tilde{\gamma}^{(1)})} m(X\tilde{\beta}^{(1)}) \right)
\end{aligned}$$

2. Solve

$$\hat{A}^{(s+1)} = \frac{1}{n} \sum_{i=1}^{n} U_i(y_i, x_i; \tilde{\theta}^{(s)}) U_i^t(y_i, x_i; \tilde{\theta}^{(s)}).$$

   k

3. Update the weights:

$$w_i^{(s+1)} = \psi \left(\frac{\|U_i^{(s)}\|_{\hat{A}^{(s)}}}{a \times \text{median}(\|U_i^{(s)}\|_{\hat{A}^{(s)}})} \right)$$

   Set $W^{(s+1)}$ to be the vector of $w_i^{(s+1)}$'s. If $w_i^{(s)} = 0$, then $w_i^{(s+1)} = 0$.

4. Solve the estimating equations. Our parameters become

$$\begin{aligned}
\tilde{\beta}^{(s+1)} &= \text{coefficients of weighted least squares using} W^{(s+1)} \\
\tilde{\gamma}^{(s+1)} &= \text{coefficents of weighted logistic regression using} W^{(s+1)} \\
\tilde{\mu}^{(s+1)} &= \sum_{i=1}^{n} w_i^{(s+1)} \left(\frac{RY}{\pi(x_i, \tilde{\gamma}^{(s+1)})} - \frac{R - \pi(x_i, \tilde{\gamma}^{(s+1)})}{R - \pi(x_i, \tilde{\gamma}^{(s+1)})} m(X\tilde{\beta}^{(s+1)}) \right)
\end{aligned}$$

5. If $|\theta^{(s+1)} - \theta^{(s)}| < \epsilon$, end. Else, return to step 2.

To estimate the coefficiences from weighted regressions, we use the `glm` function in `R` with the weight option set to W.

Simulation

Closely follows the scenario proposed by Tsiatsis and Davidian 'More Robust Doubly Robust Estimators', but with an additional step to randomly "corrupt" outcomes, creating outliers.

- $Z_i = (Z_{i1}, \ldots, Z_{i4})^t \sim N(0, 1)$ with $n = 500$.
- $X_i = (X_{i1}, \ldots, X_{i4})^t$ where
  - $X_{i1} = \exp(Z_{i1}/2)$
  - $X_{i2} = Z_{i2}/\{1 + \exp(Z_{i1})\} + 10$
  - $X_{i3} = (Z_{i1} Z_{i3}/25 + 0.6)^3$ and

  - $X_{i4} = (Z_{i3} + Z + i4 + 20)^2$.
- Let the true outcome model be $Y|X \sim N(m_0(X), 1)$.
  - $m_0(X) = 210 + 24.7Z_1 + 13.7Z_2 + 13.7Z_3 + 13.7Z_4$
  - "Corrupt" 10% of the y_i's by simulating $y_i|x_i \sim N(m_0(x_i), 7)$ to create outliers.
- True propensity score model:
  - $\pi_0 = \text{expit}(-Z_1 + 0.5Z_2 - 0.25Z_3 - 0.1Z_4)$
  - Misspecified models use $X's$ instead of $Z's$.
- True $\mu_0 = 210$.

Implementation

From our simulations, we apply both the standard doubly robust estimator (DR) and our re-weighted doubly robust estimator (RWDR). Evaluating the bias and root mean squared error for the two approaches, we see:

	Bias_{DR}	RMSE_{DR}	$\text{Bias}_{\text{RWDR}}$	$\text{RMSE}_{\text{RWDR}}$
Both Correct	-0.04	1.43	-0.24	1.26
OR Wrong	-0.25	1.86	-0.57	1.82
PS Wrong	3.74	88.39	-0.26	1.21
Both Wrong	42.01	654.75	0.52	2.28

Hence our reweighted approach greatly reduces the variance of the estimator in the cases where the propensity score model is wrong, but seems to introduce small amount of bias when both models are correct.