Robustifying doubly-robust estimators

Stephen Cristiano
Jordan Johns
Johns Hopkins University
Department of Biostatistics

August 25, 2017

- In robust regression, we define a weight function such that the estimating equation becomes

$$\sum_{i=1}^{n} w_i(y_i - x^t\beta)x_i^t)$$

- the weight is defined as

$$w(e) = \frac{\psi(e)}{e}$$

  For residual e and some score function ψ.
- Popular choices for weight functions are Huber, Hampel, and bisquare.

- Recall in doubly robust estimation

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

- Estimate the propensity scores $\pi(X_i,\hat{\gamma})$ via logistic regression.
- $\hat{\beta}$ is estimated via complete cases regression (condition on $R = 1$).

- See if we can robustify with

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

- The propensity scores are untouched.
- $\hat{\beta}_{rob}$ is estimated via complete cases regression using a robust method.
- w_i are the weights from that regression.

SIMULATION

- Closely follows the scenario proposed by Tsiatsis and Davidian "More Robust Doubly Robust Estimators"
- $Z_i = (Z_{i1}, \ldots, Z_{i4})^t \sim N(0, 1)$ with $n = 1000$.
- $X_i = (X_{i1}, \ldots, X_{i4})^t$ 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:
  $\pi_0 = expit(-Z_1 + 0.5Z_2 - 0.25Z_3 - 0.1Z_4)$
- Misspecified models use $X's$ instead of $Z's$.
- True $\mu_0 = 210$.

Table : Usual Doubly Robust estimation

	μ	Bias	RMSE
Both Correct	210.01	-0.01	1.37
OR Wrong, PS Correct	210.23	-0.23	1.75
OR Correct, PS Wrong	208.18	1.82	62.49
Both Incorrect	187.89	22.11	418.29

Table : Doubly Robust estimator with Hampel weighting

	μ	Bias	RMSE
Both Correct	210.47	-0.47	1.18
OR Wrong, PS Correct	213.24	-3.24	3.53
OR Correct, PS Wrong	210.47	-0.47	1.27
Both Incorrect	213.03	-3.03	3.36

- While there is a big improvement when the PS is wrong, bias is being introduced when OR is wrong.
- There is sensitivity due to the weighting mechanism chosen.
- Extensions: Estimating regression coefficients, longitudinal data, GLM.