Base Case TMLE Procedure

November 21, 2017

Step 1: Simulate from above, 10 time points

Assume the end time, $\tau = 10$ and we start with binaries $W(0), A(0), Y(0)$. From there,
$W(t)$ depends on $Y(t-1), A(t-1)$
$A(t)$ depends on $W(t), Y(t-1)$
$Y(t)$ depends on $(A(t), W(t))$

Step 2: Estimate the W,A,Y mechanisms which determine the likelihood

$g^*(A(1) = 1 \mid W) = 1$ is the intervention, leaving everything else in tact.
First we use the data to estimate coefficients of previous variables so we have $Pr(Y(t) = 1) = expit(\alpha_0 + \alpha_1 A(t) + \alpha_2 W(t))$ estimates and likewise so for $W(t)$ and $A(t)$.

Step 3 Compute the EIC

With all of those binaries in hand, you first can compute the estimate by drawing a $W(1)$ from the binomial, intervening to set treatment to 1, then choosing binaries from there until you draw the outcome. Repeat this 1 million times and average the results for your initial estimate. You can do this exact procedure for all of the rest of the conditional densities, which are clarified for our simple scenario by noticing the following: We can see on the bottom of page 270 and top of page 271, the three components of the efficient influence curve in their relative tangents spaces related to outcome, covariates and treatment respectively. Here, $C_y(s) = (A(s), W(s))$ and $C_w(s) = (Y(s-1), A(s-1))$ and $C_a(s) = (W(s), Y(s-1))$ and $J = 1$ with $\tau_J = 10$. Therefore we have no sum from 1 to J. W e also only have g^* defined above and $\mathcal{I}_j = \mathcal{I} = \{1\}$ because we only intervene at that time point. The clever covariates: $\frac{h^*_{c_y(s)}(C_y(i))}{h_{c_y}(C_y(i))}$ are accomplished via page 273, which entails generating B outcomes under intervention and not under intervention. For our case, we will have 4 empirical probability ratios possible for $C_y(i)$ being one of (0,0), (0,1), (1,0) and (1,1) and likewise for $C_w(i)$ and $C_a(i)$. Once these are done we can evaluate our efficient influence curve on bottom of page 270. If you look at the IC on bottom of page 270, you can see we have only 8 (2^3) possibilities for $A(i), C_a(i)$ and same for $Y(i), C_y(i)$ and $W(i), C_w(i)$. This will be a nested loop as on the bottom of page 270 but could be made more efficient I'm certain.

Step 4: Run the TMLE

We can then run the TMLE on all the factors of the likelihood using logistic regression and the h ratios as the clever covariates, each with its own ϵ, to update the conditional expectations and then re-evaluate the parameter and empirical mean of IC to check if it is below a tolerance of $\frac{\hat{\sigma}(IC_n)}{n}$. Just how to do this, is a

bit much right now to explain because I am leaving and not working on this until Monday but we are very close at this point.