Base Case TMLE Procedure

November 28, 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 for $W(t)$ and $A(t)$.

Step 3: Compute the estimate

Perform monte-carlo sampling on the initial estimates to estimate EY_{g^*}.

1. sample a W(1) by choosing a random binomial based on the past,
2. Interve on A(1) to set it to 1,
3. Continue with random binomials until the outcome, according to the initial estimates.
4. repeat steps 1 to 3 many times and average to get the estimate

Step 4: Compute the EIC and Clever Covariate

Now we will refer to chapter 19 in the new book on tstmle github.

computing the clever covariate for the $\bar{Q}_y$ update as well as IC

First, this component of the influence curve is given by $\frac{1}{N}\sum_{i=1}^{n} \bar{D}_{\bar{q}_y}(Y(i), C_y(i)) = \frac{1}{N}\sum_{i=1}^{n} H_y(C_y(i))(Y(i) - \bar{Q}_y(1, C_y(i)))$.
Note that:
$C_y(i) = (A(i), W(i))$,
$C_a(i) = (W(i), Y(i-1))$
$C_w(i) = (Y(i-1), A(i-1))$

Notice that the key here is to compute $H_y(C_y(i))$ because this will be our typical clever covariate for the i^{th} time point for dependent observation, $Y(i)$.

In a loop for i in 1:N

for a nested loop from s in 1:N

compute

$H_{y(s)}(C_y(i)) = \mathbb{E}[Y_{g^*} \mid Y(s) = 1, C_y(s) = C_y(i)] - \mathbb{E}[Y_{g^*} \mid Y(s) = 0, C_y(s) = C_y(i)]$

This will take montecarlo sampling using the binaries already estimated in step 2. For each expectation, start with generating $W(s+1)$, then $A(s+1)$ etc until the outcome at time, $\tau = 100$. Do this a lot of times (start with maybe 1000 if it takes a while). Then average the resulting differences. The clever covariates: $\frac{h^*_{cy(s)}(C_y(i))}{\bar{h}_{c_y}(C_y(i))}$ are accomplished via page 327, which entails generating B outcomes under intervention and not under intervention. Again, we just montecarlo sample to generate these observations, given our previous estimate of the distribution. We also (as it says in the book) need not apply logistic regression here as we can do it nonparametrically. Now store $H_{y(s)}(C_y(i))\frac{h^*_{cy(s)}(C_y(i))}{\bar{h}_{c_y}(C_y(i))}$. Concurrently, do the same thing for $H_{g(s)}(C_a(i)) = \mathbb{E}[Y_{g^*} \mid W(s) = 1, C_a(s) = C_a(i)] - \mathbb{E}[Y_{g^*} \mid W(s) = 0, C_a(s) = C_a(i)]$ and $\frac{h^*_{cy(s)}(C_a(i))}{\bar{h}_{c_y}(C_a(i))}$, except only do so on non-intervention nodes as there is no such component in the IC for intervention nodes as usual. On intervention nodes $H_{a(s)}(C_a(i))\frac{h^*_{ca(s)}(C_a(i))}{\bar{h}_{c_a}(C_a(i))} = 0$. Do likewise for w on all nodes.

end the inner loop

store $H_y(C_y(i))$, $H_a(C_a(i))$ and $H_w(C_w(i))$

store $\bar{D}^N(O(i)) = H_y(C_y(i))(Y(i) - \bar{Q}_y(1 \mid C_y(i)) + H_a(C_a(i))(A(i) - \bar{g}(1 \mid C_a(i)) + H_w(C_w(i))(W(i) - \bar{q}_w(1 \mid C_w(i))$ where $\bar{Q}_y, \bar{g}$ and q_w are the initial estimates.

end outer loop

Step 4: check tolerance

compute the sample mean of $\bar{D}^N(O)$ and see if below $\frac{\sigma}{N}$. If so, your last estimate is used as well as the influence curve $\bar{D}^N(O(i))$ to get inference. Otherwise go to Step 5.

Step 5: Run the TMLE and compute estimate

Create a 3N-1 length vector with all the clever covariates and use as outcome, W(t), A(t) and Y(t). The offset is the corresponding initial estimate (each is one of 4 binomial probabilities, considering they are functions of two binaries) . Note that for the intervention nodes there is no outcome and clever covariate. Use logistic regression to find ϵ_n, our fluctuation parameter. Update as usual and return to step 1.