PH C240B: Assignment 4

Rachael Phillips

12/1/2017

Load longdata.RData from bCourses. You may use Odat.long or Odat.wide for your analysis. They are identical except for the format. The columns are labeled according to how we normally notate them. A1's are the binary treatments, A2's are the censoring indicators, Y's are the binary survival outcomes and L's are the covariates, which are either continuous or binary as you will see upon looking at the dataframe. Fortunately, the data is already cleaned, as in, NA's are produced for all nodes after outcome is 1 (death) and also after censoring occurs.

1

Assuming sequential randomization and positivity as usual, estimate the difference in counterfactual mean outcomes at time 5 under treatment and not: $\mathbb{E}[Y(5)_{\bar{a}=(0,0,1,1,1)}] - \mathbb{E}[Y(5)_{\bar{a}=(0,0,0,0,0)}]$. Give estimates using IPTW and TMLE (debiasing eachh outcome regression with a targeting step, moving backward in time and otherwise known as L-TMLE) and show all code used to do so. Your superlearner library should include three algorithms, one of each of the following: glm, mean, xgboost. Provide instructions on how I may access your superlearner results in this estimation process.

2

Let us assume you have estimated counterfactual means for all 32 possible regimes, $\bar{a}$, over 5 time points and you have summarized the counterfactual means via the use of a marginal structural model (MSM):

$$\underset{\alpha,\beta}{argmin} \sum_{\bar{a}} (\mathbb{E}Y(5)_{\bar{a}} - (\alpha + \beta sum(\bar{a})))$$

where $sum(\bar{a})$ = sum of treatments over the 5 time points in the study. Notice, this is just a least squares projection of the counterfactual means onto a line. How would you interpret this coefficient to a layperson, i.e. someone not technical?

3

Continuing from the previous question, what if you had some regimes that were much more common than others in the data and you decided to offer estimates of the weighted projection:

$$\underset{\alpha,\beta}{argmin} \sum_{\bar{a}} P_n(\bar{A} = \bar{a})(\mathbb{E}Y(5)_{\bar{a}} - (\alpha + \beta sum(\bar{a})))$$

How would you interpret this coefficient to a layperson, ie someone not technical? $P_n(\bar{A} = \bar{a})$ is just the empirical probability of observing $\bar{a}$. Also, in what situation does this change the parameter defined in the previous sections? Explain briefly.

Collaborators & Resources

Tommy Carpenito, Yue You, Jonathan Levy