ICE representation of the Longitudinal G-Comp Estimand & Longitudinal TMLE

PH252E - Advanced Topics in Causal Inference

Assigned: October 30, 2019
Present in class: November 6, 2019
Write up Due: November 13, 2019

Please read the above article and come prepared to present the following questions with your group

Required Reading:
Schnitzer ME, van der Laan MJ, Moodie EE, Platt RW. EFFECT OF BREASTFEEDING ON GASTROINTESTINAL INFECTION IN INFANTS: A TARGETED MAXIMUM LIKELIHOOD APPROACH FOR CLUSTERED LONGITUDINAL DATA. Ann Appl Stat. 2014 Jun;8(2):703-725.

1. Specify the question of interest. How does this question differ from that addressed in the original analysis of the randomized trial? What was the authors' motivation for doing a secondary analysis?

2. Specify the longitudinal causal model $\mathcal{M}^F$ for individual level data (with endogenous nodes and time ordering corresponding to the observed data as specified in Section 3). What do you think about the temporal ordering assumptions? Any concerns? *Bonus:* How might you modify this causal model to more accurately describe knowledge about the true data generating process?

3. Specify the counterfactual outcomes of interest. Hint: First define the interventions of interest. Then define the counterfactual outcome under these interventions (the counterfactual outcome is a random variable). How are these counterfactual outcomes generated using an intervention on the causal model?

4. Specify the target causal parameter.

5. What are the observed data? What is the assumed link between the observed data and the structural causal model $\mathcal{M}^F$? Factorize the observed data distribution $P_0(O)$ according to the time-ordering.

6. What are the needed identifiability assumptions? Do they seem reasonable here? Any particular concerns?

7. Under the assumption of sequential randomization and positivity, write out the (counterfactual or "post intervention") distribution $Q^{\bar{a}}$ of the counterfactual non-intervention variables $(W, \bar{L}_5^{\bar{a}}, Y^{\bar{a}})$ as a function of the observed data distribution P_0.

8. Specify the statistical estimand using the traditional G-computation formula. (Fine to assume W is discrete to keep the discrete notation we have been using in class, and write this in the form we have been using in class.)

9. Briefly review implementation of "traditional" longitudinal parametric G-Computation for this estimand. What are some possible pros/cons to this approach?

10. What is the sequential representation of the statistical estimand? (Note- try to write it out for yourself rather than copying as there appears to be a typo in eq. 2)

11. Give an overview of implementation of the sequential G computation estimator. What are some advantages over the traditional parametric G comp estimator?

12. Give an overview of implementation of the sequential TMLE.