Positivity

PH252E - Advanced Topics in Causal Inference

Assigned: November 13, 2019
Presentation: November 20, 2019
Write up Due: November 27, 2019

Required Reading:
Petersen, ML, Porter KE, Gruber S, Wang Y, van der Laan MJ, Diagnosing and responding to violations in the positivity assumption. Statistical Methods in Medical Research, 2012. 21(1):31-54

Supplementary (not required) Readings:
- R. Neugebauer and M. J. van der Laan. Nonparametric causal effects based on marginal structural models. *J Stat Plan Infer*, 137(2):419–434, 2007
- O. Bembom and M.J. van der Laan. Data-adaptive selection of the truncation level for inverse-probability-of-treatment-weighted estimators. Technical Report 230, Division of Biostatistics, University of California, Berkeley, 2008. Available at http://biostats.bepress.com/ucbbiostat/paper230
- O. Bembom, W.J. Fessel, R.W. Shafer, and M.J. van der Laan. Data-adaptive selection of the adjustment set in variable importance estimation. 231, Division of Biostatistics, University of California, Berkeley, 2008
- S.R. Cole and M.A. Hernán. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*, 168(6):656–664, 2008. PMCID: PMC2732954

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

For point treatment questions, use the point treatment example from the paper. For longitudinal extensions (not covered in the paper), we will consider

• Structural causal model $\mathcal{M}^F$:

$$
\begin{aligned}
L(0) &= f_{L(0)}(U_{L(0)}) \\
A(0) &= f_{A(0)}(L(0), U_{A(0)}) \\
L(t) &= f_{L(t)}(\bar{L}(t-1), \bar{A}(t-1), U_{L(t)}), t = 1, ..., K+1 \\
A(t) &= f_{A(t)}(\bar{L}(t), \bar{A}(t-1), U_{A(t)}), t = 1, ..., K
\end{aligned}
$$

• Through interventions on this SCM, we can define the counterfactual outcome $Y_{\bar{a}}$ as the outcome a subject would have had, if possibly contrary to fact, he or she had received exposure level $\bar{A} = \bar{a}$. Given these counterfactuals, we can define causal parameters of interest. Consider, for example, the average treatment effect:

$$\Psi^F(P_{U,X}) = E_{U,X}(Y_{\bar{a}=1} - Y_{\bar{a}=0})$$

for a binary exposure. Alternatively, we could use a marginal structural model (MSM) to summarize how the expected counterfactual outcome changes as a function of the exposure history $\bar{a}$ and possible baseline effect modifiers, denoted V. Consider, for example, the following working MSM to define the target parameter β as the projection of the true causal curve $E_{U,X}(Y_{\bar{a}}|V)$ onto a summary model $m(\bar{a}, V|\beta)$ with projection function (i.e. weights) $h(\bar{a}, V)$:

$$\beta\big(P_{U,X}, m, h\big) = \underset{\beta}{\operatorname{argmin}}\ E_{U,X}\left[\sum_{\bar{a}\in\mathcal{A}} \big(Y_{\bar{a}} - m(\bar{a}, V|\beta)\big)^2 h(\bar{a}, V)\right]$$

$$\text{where, for example, } m(\bar{a}, V|\beta) = \beta_1 + \beta_2 \sum_{t=0}^{K} a(t) + \beta_3 V + \beta_4 \sum_{t=0}^{K} a(t) \times V$$

The causal parameter is then the value of the β coefficients that minimize the (weighted) sum of squared residuals between the true counterfactual outcome $Y_{\bar{a}}$ and the predicted $m(\bar{a}, V|\beta)$. (If the outcome were binary, we would generally use the negative log loss function.) The function $h(\bar{a}, V)$ determines how to weight the projection. For example $h(\bar{a}, V) = 1$ gives equal weight to the entire curve (i.e. all values of $\bar{a}, v$), while $h(\bar{a}, V) = P_0(\bar{A} = \bar{a})$ gives more weight to exposure levels that are better represented.

- The observed data are n i.i.d. observations of

$$O = \big(L(0), A(0), L(1), A(1), ...L(K), A(K), L(K+1)\big) = \big(\bar{L}(K+1), \bar{A}(K)\big) \sim P_0 \in \mathcal{M}$$

, where $L(t)$ denotes the non-intervention nodes at time t, $A(t)$ denotes the intervention nodes (exposure or treatment) at time t, and $Y \subset L(K+1)$ is the outcome measured at final time point.

1. Consider the point treatment example presented in the paper. What assumptions are needed for identifiability? Why do we need the positivity assumption? Focus on the strong positivity assumption.
2. Consider the longitudinal example presented above. What are the analogous assumptions? Interpret the strong positivity assumption in words.
3. How do theoretical and practical violations of the positivity assumption arise? Give a real world example of each. Do you think practical positivity violations are more or less likely to be a problem for longitudinal versus point treatment interventions? Why?
4. Focus on the point treatment example in the paper. How would you estimate the coefficients of a working MSM with a G computation estimator? Describe the impact of positivity violations on the performance of this estimator.
5. Focus on the point treatment example in the paper. How would you estimate the coefficients of a working MSM with IPTW? Describe the impact of positivity violations on the performance of IPTW. What is the impact of weight truncation?
6. Provide an overview of the properties of the AIPTW and TMLE estimators. Describe impact of positivity violations on the performance of AIPTW and TMLE.
7. What are a few quick ways to diagnose positivity violations? What are some of their short-comings?
8. Discuss the authors proposed parametric bootstrap.
   (a) Formally define bias for an estimator.
   (b) What are some of the causes of bias?
   (c) Describe the parametric bootstrap-based biased estimate and its implementation.
   (d) What is the goal of the proposed algorithm? What are sources of bias does it help identify? What are some of its limitations?
9. Describe the following approaches to responding to positivity violations, using an example. Discuss their pros/cons.
   (a) Changing the projection function $h(a, V)$
   (b) Restricting the adjustment set
   (c) Restricting the sample (trimming)
   (d) Changing the intervention of interest
10. For each approach, consider whether it could be applied in the setting of a longitudinal intervention (in which control for time dependent confounding is required for identifiability). If not, why not (or under what conditions would it break down)? If so, how could it be used to mitigate threats due to positivity? Use an example to illustrate.
11. How could you formalize the selection between different target parameters?