Discussion Assignment #1

Sabrina Boyce, Shelley Facente, and Steph Holm

10/14/2019

Question 1: Specify the question(s) of interest

How does detectability of HIV RNA in patients' blood vary as a function of the cumulative AZT dose history?

Question 2: For the AZT example, specify the longitudinal causal model $\mathcal{M}^F$. (See Lecture 1 for help.) What is the corresponding DAG?

Structural Causal Model

$$
\begin{aligned}
O &= \{\bar{L}(k), \bar{A}(k), Y(k+1)\} \\
\bar{L}(k) &= \text{all measured risk factors for Y (CD4 count, age, white blood count, hematocrit, AIDS diagnosis, symptoms)} \\
\bar{A}(k) &= \text{AZT dosage (100mg)} \\
Y(k+1) &= \text{HIV RNA detected in blood}
\end{aligned}
$$

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

The DAG is represented as Figure 1a in the article.

Question 3: Specify the counterfactuals of interest. How are these counterfactuals generated using an NPSEM?

Our counterfactuals of interest are the values of $Y_{\bar{a_k}}$ we would have observed if A_0 had been set to values 0, 1, 2, 3,... 15 at each time point, all the 16 possible AZT doses at each time point. We are able to generate these counterfactuals by plugging in every combination of treatment level across all time points and observing the outcome for each.

Question 4: Specify the target causal parameters. Discuss their interpretation

The target causal parameter is given by the following marginal structural model:

$$E(Y_{\bar{a_k}}) = m(\bar{a}|\beta = \beta_0 + \beta_1 \sum_{t=1}^{k} a(t)$$

Our target causal parameter is β_1 so that we can calculate e_1^{β} which would give us the causal OR for HIV RNA detection associated with an increasing AZT dose of 100mg. β_1 represents the causal parameter if the relationship between AZT dose and HIV RNA is unconfounded by measured or unmeasured factors.

Question 5: What are the observed data? What is the link between the observed data and the structural causal model $\mathcal{M}^F$? Do we place any restrictions on the observed data model $\mathcal{M}$?

N i.i.d. copies of $O = (A(\bar{t}), L(\bar{t}))$ P_0

$\mathcal{M}^F$ contains the distribution of $\mathcal{M}$ and the true distribution (P_0) of O is an element of $\mathcal{M}$. There are no restrictions on the observed data model M.

Question 6: Identifiability and the statistical estimand

(a) Consider a single time point intervention. Discuss the assumptions needed to identify the average treatment effect $E_{U,X}(Y_1 - Y_0)$ from the observed data distribution. Discuss the backdoor criteria. (See Lecture Notes from 252D for a refresher.)

In order to identify the average treatment effect from the observed data distribution in a single point treatment context, we would need to verify the randomization assumption (that $a(\bar{t})|L$ is independent of Y), the independence assumption (met by conditioning on L so that U_A is independent of U_Y and U_L is independent of U_Y), and the positivity assumption ($\min_{aEA} P(A = a|L) > 0$).

The backdoor criteria are that 1) All spurious sources of association between A and Y are blocked, 2) No new spurious associations are created (no colliders are conditioned upon), and 3) No part of the causal pathway between A and Y are conditioned upon (we are not conditioning on a factor on the causal pathway between A and Y).

In this example of a point treatment setting, if we condition on L we would close all backdoor paths between A and Y without conditioning on a collider or mediator which would satisfy our backdoor criteria.

(b) Consider a longitudinal setting and the following DAG. Can you specify a set of covariates that satisfy the backdoor criteria to identify the expected counterfactual outcome $E_{U,X}(Y_{\bar{a}})$ under a joint intervention on A_0 and A_1?

No, we do not have a set of covariates to satisfy the backdoor criteria. If we want to have a joint intervention on A(0) and A(1) and we try to condition on L(0) and L(1), we would be conditioning on a factor (L(1) in this case) that is on the causal pathway from A(0) to Y. If, instead, you condition on just L(0), you leave open the backdoor pathway between A(1) and Y. These show us that both conditioning or not conditioning on L(1) is problematic in the longitudinal setting if you use the strategies we employ in the point-treatment setting.

(c) What alternative identifiability assumptions would be sufficient in this case? What statistical estimand is equal to $E_{U,X}(Y_{\bar{a}})$ under this assumption?

To satisfy the backdoor pathway in the longitudinal setting in which we have a joint intervention on A(0) and A(1), we could employ the sequential backdoor criteria to take care of the problem described in 6b, as well as our randomization and positivity assumptions, in order to identify our target parameter.

The statistical estimand is

$$\Psi(P_0) = \sum_{\bar{l}}(E(Y|\bar{A}=1,\bar{L}=\bar{l})xP(L(1)=l(1)|A(0)=1,L(0)=l(0))xP(L(0)=l(0))-$$
$$\sum_{\bar{l}}(E(Y|\bar{A}=0,\bar{L}=\bar{l})xP(L(1)=l(1)|A(0)=0,L(0)=l(0))xP(L(0)=l(0))$$

Question 7: What is the intuition behind longitudinal IPTW?

In general, with IPTW, the idea is that some pattern of covariates can predict the probability of having received the exposure. But because the covariates are related to exposure, some covariate-exposure combinations will be over-represented in the data and some will be under-represented. By upweighting the contributions of the under-represented exposure combinations (and vice-versa) we can approximate the data that would result from a randomized controlled trial. In the longitudinal setting specifically, the individuals who are upweighted are those whose *entire* exposure history is rare. An individual's weight is the inverse of the probability that the subject received the treatment that they did.

Question 8: Give a detailed implementation of longitudinal IPTW to estimate parameters of an MSM without effect modifiers.

Implementing the Estimator

The stabilized weights for a marginal structural model are:

$$sw_i = \frac{\prod_{k=0}^{K} g_n(A_k = a_{ki})}{\prod_{k=0}^{K} g_n(A_k = a_{ki}|\bar{A}_{k-1} = \bar{a}_{(k-1)i}, \bar{L}_k = \bar{l}_{ki})}$$

Step 1: Calculate appropriate stabilized weights

Part A:

Estimate the probability of receiving treatment, based on prior treatment and covariates using correctly specified parametric regression models (logistic regression) to get the denominator of the weights.

- **In this example** we are estimating the probability of being treated with AZT at each time point, given the covariate pattern and the prior history of AZT treatment, and the authors use this model:

$$\begin{aligned} logit(p[A_k = 1|\bar{A}_{k-1} = \bar{a}_{k-1}, \bar{L}_k = \bar{l}_k)]) &= \alpha_0 + \alpha_1 k + \alpha_2 a_{k-1} + \alpha_3 a_{k-2} \\ &+ \alpha_4 l_k + \alpha_4 l_{k-1} + \alpha_5 l_{k-2} + \alpha_6 a_{k-1} l_k \\ &+ \alpha_7 l_0 \end{aligned}$$

Where l_k is the covariate vector.

Step 1: Calculate appropriate stabilized weights

Part B:

Estimate the probability of receiving treatment, based on prior treatment (but NOT covariates) to get the numerator of the weights.

- **In this example** We simply remove those terms from the model that depended on l, our covariates, resulting in this model:

$$logitp[A_k = 1|\bar{A}_{k-1} = \bar{a}_{k-1}] = \alpha_0^* + \alpha_1^* k + \alpha_2^* a_{k-1} + \alpha_3^* a_{k-2}$$

Step 1: Calculate appropriate stabilized weights

Part C:

Predict each subject's probability of the entire exposure history, from each of the models (based on either ONLY exposure history or exposure history PLUS covariates). Those predicted values are then plugged into this equation to get the stabilized weight for each subject, i.

- **In this example** we are estimating the probability of their entire AZT treatment history pattern.
- Say that ρ_{ki} is the probability of treatment for the ith subject at time k given exposure and covariates, but ρ_{ki}^* is the probability of treatment for the ith subject at time k given exposure history only. Then each subject's weight is:

$$sw_i = \frac{\prod_{k=1}^{K} (\rho_{ki}^*)^{a_{ki}} (1 - \rho_{ki}^*)^{1-a_{ki}}}{\prod_{k=1}^{K} (\rho_{ki})^{a_{ki}} (1 - \rho_{ki})^{1-a_{ki}}}$$

Step 2: Take the weighted average of observed outcomes across the population

Here we are trying to get our estimate of the β using the stabilized weights we calculated before. We fit a parsimonious MSM (with our stablized weights) such as:

$$logit(p[Y = 1|\bar{A} = \bar{a}]) = \beta_0 + \beta_1 \sum_{t=0}^{K} a(t)$$

β_1 from this model is then our MSM IPTW estimate.

Question 9: How would you modify the above procedure when the target causal parameter is a MSM with effect modification?

Including Baseline Covariates in the Model

If effect modification by baseline covariates V (a subset of $L(t)$) is of interest to the target causal parameter, inclusion of those baseline characteristics in the MSM allows for their incorporation into the counterfactual pseudo-populations.

Therefore, we want to condition on baseline covariates V in our model, where the true model is now:

$$logit(Pr[Y_{\bar{a}_k}|V]) = m(\bar{a}, V|\beta) = \beta_0 + \beta_1 \sum_{t=0}^{K} a(t) + \beta_2 V + \beta_3 \sum_{t=0}^{K} a(t) \times V$$

and the working model is now:

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

Including Baseline Covariates in the Model

The choice of numerator in the MSM changes the target parameter being measured, and the stabilized weights can be improved:

$$\hat{sw}_i = \frac{g_n(\bar{A}_i(K)|V_i)}{\prod_{t=1}^{K} g_n(A_i(t)|\bar{A}(t-1), \bar{L}_i(t))}$$

While we can use a variety of predicted values in the numerator, ultimately we want the one that is going to make the weights closest to one (i.e. the numerator has as many overlapping factors with the denominator as possible, so that most terms in the fraction cancel out).

Question 10: How does censoring change (a) the scientific question, (b) the causal model $\mathcal{M}^F$, (c) the counterfactuals and target causal parameter, (d) the observed data, (e) identifiability and (f) estimation

The article suggests that censoring be considered as an additional time-varying treatment. This causes the following changes: * the scientific question: does not change * the causal model: add an indicator C_k at each time point, which is an indicator of whether the outcome was observed at time k, and is equal to 1 if the subject was censored * the counterfactuals: the counterfactual outcomes must now include both that the subject followed a particular treatment regimen ($\bar{a}$) and that they were never censored ($\bar{C} = 0$),so the counterfactuals are $Y(\bar{A}(t) = \bar{a}(t), \bar{C}(K) = 0$ * the causal parameter: the same as previous, but with the addition to the counterfactuals noted above. So the parameter is $\psi^F(\mathcal{P}_{U,X}) = E_{U,X}(Y_{\bar{a},\bar{C}=0})$ * the observed data: presumably if there is censoring some subjects will be missing data at later time points * identifiability: we must assume that there are no unmeasured factors in common for A_k and C_k for all values of k, and that censoring is independent of the outcome Y at all times t. * estimation: it is now an inverse-probability-of-treatment-and-censoring estimator.

Question 11: In Section 11, the authors note "our IPTW estimators will be biased and thus MSMs should not be used in studies in which at each time k there is a covariate level l_k such that all subjects with that level of the covariate are certain to receive the identical treatment a_k". What assumption are the authors referring to?

They are referring to the positivity assumption. In order for IPTW estimators to be unbiased, the probability of exposure for any given set of covariates should be neither close to zero nor to 1. If a covariate l determines treatment, then this will be a practical positivity violation as all individuals with a specific value of l_k will get the treatment (ie.probability will be one) and all individuals with other value(s) of l_k will not get the treatment (ie. probability will be zero).

Question 12: What are some potential advantages or disadvantages to longitudinal IPTW?

Advantages:

- Allows us to use confounded observational data to approximate data wherein the exposure was randomized.

Disadvantages:

- It has high variance compared to alternative methods (it is inefficient).
- It is easy to run into ositivity violations or near violations, which leadsto a biased estimator. Because the weights are the inverse of the product of the probabilities, if there are many time points even moderately small probabilities can end up producing quite large weights once everything is multiplied through. (Stablizing the weights does help.)
- Relies on consistent estimation of the probabilities, often using parametric models. If model is misspecified, the parameter can be biased.