Section 25

Lecture 10

Lemma of IPW theorem

Lemma

Consider weights on the form

$$\frac{g(\overline{A})}{\prod_{k=0}^{K} p(A_k \mid \overline{L}_k, \overline{A}_{k-1})},$$

then

$$b_{\overline{a}}(y) = \frac{1}{g(\overline{a})}\mathbb{E}\left\{ \frac{g(\overline{A})I(\overline{A} = \overline{a})}{\prod_{k=0}^{K} p(A_k \mid \overline{L}_k, \overline{A}_{k-1})} p(y \mid \overline{L}_K, \overline{A}_K) \right\}.$$

Proof.

$$
\begin{aligned}
& b_{\overline{a}}(y) \\
&= \sum_{\overline{l}_K} p(y \mid \overline{l}_K, \overline{a}_K) \prod_{j=0}^{K} p(l_j \mid \overline{l}_{j-1}, \overline{a}_{j-1}) \\
&= \sum_{\overline{l}_K} p(y \mid \overline{l}_K, \overline{a}_K) \frac{\prod_{k=0}^{K} p(a_k \mid \overline{l}_k, \overline{a}_{k-1})}{\prod_{k=0}^{K} p(a_k \mid \overline{l}_k, \overline{a}_{k-1})} \prod_{j=0}^{K} p(l_j \mid \overline{l}_{j-1}, \overline{a}_{j-1}) \\
&= \sum_{\overline{l}_K} \frac{1}{\prod_{k=0}^{K} p(a_k \mid \overline{l}_k, \overline{a}_{k-1})} p(y \mid \overline{l}_K, \overline{a}_K) \prod_{k=0}^{K} p(a_k \mid \overline{l}_k, \overline{a}_{k-1}) \prod_{j=0}^{K} p(l_j \mid \overline{l}_{j-1}, \overline{a}_{j-1}) \\
&= \sum_{\overline{l}_K} \frac{1}{\prod_{k=0}^{K} p(a_k \mid \overline{l}_k, \overline{a}_{k-1})} p(y, \overline{l}_K, \overline{a}_K).
\end{aligned}
$$

□

Proof.

$$
\begin{aligned}
&= \sum_{\bar{l}_K} \frac{1}{\prod_{k=0}^{K} p(a_k \mid \bar{l}_k, \bar{a}_{k-1})} p(y, \bar{l}_K, \bar{a}_K) \\
&= \sum_{\bar{l}_K} \sum_{\bar{a}^*} \frac{I(\bar{a}^* = \bar{a})}{\prod_{k=0}^{K} p(a_k^* \mid \bar{l}_k, \bar{a}_{k-1}^*)} p(y, \bar{l}_K, \bar{a}_K^*) \\
&= \frac{1}{g(\bar{a})} \sum_{\bar{l}_K} \sum_{\bar{a}^*} \frac{g(\bar{a}^*) I(\bar{a}^* = \bar{a})}{\prod_{k=0}^{K} p(a_k^* \mid \bar{l}_k, \bar{a}_{k-1}^*)} p(y \mid \bar{l}_K, \bar{a}_K^*) p(\bar{l}_K, \bar{a}_K^*) \\
&= \frac{1}{g(\bar{a})} \mathbb{E}\left\{ \frac{g(\overline{A}) I(\overline{A} = \bar{a})}{\prod_{k=0}^{K} p(A_k \mid \overline{L}_k, \overline{A}_{k-1})} p(y \mid \overline{L}_K, \overline{A}_K) \right\}.
\end{aligned}
$$

where the expectation is taken over $\overline{A}_K, \overline{L}_K$ under the distribution that generated the observed data, and positivity is used in the last line. □

So the lemma from Slide 270 follows.

Another (simple) lemma

Lemma (individuals with $\overline{A} = \overline{a}$ in the psedopopulation)

$$\mathbb{E}\left\{ \frac{g(\overline{A})I(\overline{A} = \overline{a})}{\prod_{k=0}^{K} p(A_k \mid \overline{L}_k, \overline{A}_{k-1})} \right\} = g(\overline{a}).$$

Proof.

We use that the g-formula is a density, i.e. that $\int b_{\overline{a}}(y)dy = 1$,

$$1 = \int b_{\overline{a}}(y)dy = \int \frac{1}{g(\overline{a})} \mathbb{E}\left\{ \frac{g(\overline{A})I(\overline{A} = \overline{a})}{\prod_{k=0}^{K} p(A_k \mid \overline{L}_k, \overline{A}_{k-1})} p(y \mid \overline{L}_K, \overline{A}_K) \right\} dy$$

$$g(\overline{a}) = \mathbb{E}\left\{ \frac{g(\overline{A})I(\overline{A} = \overline{a})}{\prod_{k=0}^{K} p(A_k \mid \overline{L}_k, \overline{A}_{k-1})} \right\},$$

where we used that integrals of sums are sums of integrals. □

A Theorem

Now we will show that $\mathbb{E}(b_{\overline{a}}(Y)) = \mathbb{E}_{ps}(Y \mid \overline{A} = \overline{a})$.

Proof.

$$
\begin{aligned}
&\int y b_{\overline{a}}(y) dy \\
&= \int y \frac{1}{g(\overline{a})} \mathbb{E}\left\{ \frac{g(\overline{A}) I(\overline{A} = \overline{a})}{\prod_{k=0}^{K} p(A_k \mid \overline{L}_k, \overline{A}_{k-1})} p(y \mid \overline{L}_K, \overline{A}_K) \right\} dy \\
&= \frac{1}{g(\overline{a})} \int \mathbb{E}\left\{ \frac{g(\overline{A}) I(\overline{A} = \overline{a})}{\prod_{k=0}^{K} p(A_k \mid \overline{L}_k, \overline{A}_{k-1})} y p(y \mid \overline{L}_K, \overline{A}_K) \right\} dy \\
&= \frac{1}{g(\overline{a})} \mathbb{E}\left\{ \frac{g(\overline{A}) I(\overline{A} = \overline{a})}{\prod_{k=0}^{K} p(A_k \mid \overline{L}_k, \overline{A}_{k-1})} Y \right\} \text{ (by def of expectation)}
\end{aligned}
$$

□

Finally: A proof of the Theorem

Proof.

plugging in for $g(\overline{a})$ in the Expression from the Corollary on slide 274,

$$= \frac{\mathbb{E}\left\{ \frac{g(\overline{A}) I(\overline{A}=\overline{a})}{\prod_{k=0}^{K} p(A_k|\overline{L}_k,\overline{A}_{k-1})} Y \right\}}{\mathbb{E}\left\{ \frac{g(\overline{A}) I(\overline{A}=\overline{a})}{\prod_{k=0}^{K} p(A_k|\overline{L}_k,\overline{A}_{k-1})} \right\}} \quad \text{(i.e. an IPW formula)}$$

$$= \frac{\mathbb{E}_{ps}(I(\overline{A}=\overline{a})Y)}{P_{ps}(\overline{A}=\overline{a})}$$

$$= \mathbb{E}_{ps}(Y \mid \overline{A}=\overline{a}).$$

□

This allows us to say "association is causation" in the pseudopopulation.

PS: Pseudopopulation vs observed population

The lemma allows us to characterize the number of treated in the pseudopopulation vs the original population. Recall that $\mathbb{E}(I(\overline{A} = \overline{a}))$ is the fraction of individuals with $\overline{A} = \overline{a}$ in the observed population. Let n be the total size of the observed population. Then

$$n \times \mathbb{E}(I(\overline{A} = \overline{a}))$$

is the expected number of individuals with $\overline{A} = \overline{a}$ in the observed population and

$$n \times \mathbb{E}\left\{ \frac{g(\overline{A}) I(\overline{A} = \overline{a})}{\prod_{k=0}^{K} p(A_k \mid \overline{L}_k, \overline{A}_{k-1})} \right\} = n \times g(\overline{a})$$

is the expected number of individuals with $\overline{A} = \overline{a}$ in the pseudopopulation.

We can encode various assumptions in MSMs

- Suppose we hypothesize that the causal effect of treatment history $\overline{a}$ on the mean of Y is a linear function of the cumulative exposures, i.e.

$$\text{cum}(\overline{a}) = \sum_{k=0}^{K} a_k.$$

- This hypothesis is included in the MSM

$$\mathbb{E}(Y^{\overline{a}}) = \mathbb{E}_{ps}(Y \mid \overline{A} = \overline{a}) = \eta_0 + \eta_1 \text{cum}(\overline{a}).$$

  That is, we model the marginal mean of the counterfactuals $Y^{\overline{a}}$. Whereas there are 2^K treatment combinations (unknowns on the left-hand side of the equation), we have now reduced the model such that there are only two unknowns on the right-hand side of the equation.

- Obviously, this model could also be misspecified, e.g. if the counterfactual outcome depends on some other function of the regime or if the outcome depends nonlinearly on the cumulative exposure.

Motivating the weighted regressions

Lemma (Result for weighted least squares)

Suppose excheangeability, consistency and positivity hold. Then $\mathbb{E}_{ps}(Y \mid \overline{A} = \overline{a}) = \int y b(\overline{a}) dy = \mathbb{E}(Y^{\overline{a}})$. *Then,*

$$
\begin{aligned}
&\mathbb{E}\left\{ \frac{g(\overline{A})}{\prod_{k=0}^{K} p(A_k \mid \overline{L}_k, \overline{A}_{k-1})} [Y - \mathbb{E}(Y^{\overline{A}})] \right\} \\
&\mathbb{E}_{ps}\left\{ [Y - \mathbb{E}(Y^{\overline{A}})] \right\} \\
=&\mathbb{E}_{ps}\left\{ \mathbb{E}_{ps}\left\{ [Y - \mathbb{E}(Y^{\overline{A}})] \mid \overline{A} \right\} \right\} \\
=&0, \ \textit{because the inner expectation above is 0.}
\end{aligned}
$$

Consider now the estimating equations

We use the results from the previous slide and the parameterisation

$$\mathbb{E}(Y^{\overline{a}}) = \eta_0 + \eta_1 \text{cum}(\overline{a}).$$

Now, consider the (two-dimensional) estimating equation

$$\sum_{i=1}^{n} M(\overline{L}_{k,i}, \overline{A}_i; \eta_0, \eta_1, \gamma) = 0,$$

where

$$M(\overline{L}_k, \overline{A}; \eta_0, \eta_1, \gamma) = \frac{g(\overline{A})}{\prod_{k=0}^{K} p(A_k | \overline{L}_k, \overline{A}_{k-1}; \gamma)} \begin{pmatrix} 1 \\ \text{cum}(\overline{A}) \end{pmatrix} [Y - \eta_0 - \eta_1 \text{cum}(\overline{A})].$$

This is an estimating equation for the weighted least squares estimator, where we first must have solved the estimating equations for the propensities. Together, we denote the estimating equations for the counterfactual model and the propensity scores a "stacked estimating equation".

Side note: parametric g-formula estimator

Another estimation strategy is to do so-called g-computation, using the time-varying g-formula,

$$
\begin{aligned}
&\int b_{\overline{a}}(y)dy \\
&= \int \sum_{\bar{l}_K} p(y \mid \bar{l}_K, \overline{a}_K) \prod_{j=0}^{K} p(l_j \mid \bar{l}_{j-1}, \overline{a}_{j-1}) dy,
\end{aligned}
$$

to motivate the parametric g-formula estimator

$$
\sum_i \mathbb{E}(Y \mid \overline{L}_{K,i}, \overline{a}_K; \beta_y) \prod_{j=0}^{K} p(L_{j,i} \mid \overline{L}_{j-1,i}, \overline{a}_{j-1}; \beta_{l_j}).
$$

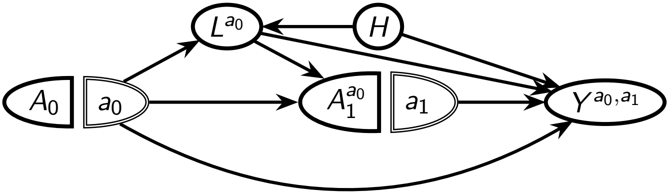
Null hypotheses in MSMs

Note that under the null hypothesis of no effect of any a_k, $k = 0, \dots, K$ the MSM is correctly specified with

$$\mathbb{E}(Y^{\bar{a}}) = \eta_0.$$

However, the standardisation estimator (parametric g-formula estimator) suffers from the so-called "g-null-paradox". That is, it is possible to show that it will always reject the null hypothesis – even if the null hypothesis is true – when the sample size grows.

Treatment-confounder feedback



- we cannot adjust for L using traditional methods, like stratification, outcome regression, and matching.
- But we read off that $Y^{a_0,a_1} \perp\!\!\!\perp A_0$ and $Y^{a_0,a_1} \perp\!\!\!\perp A_1^{a_0} \mid L_0^{a_0}, A_0 = a_0$, and we can fit MSMs.

MSMs and effect modification

Suppose that an investigator believes that for a particular component V of the vector of baseline covariates L_0, there might exist qualitative effect modification with respect to V. For example, suppose that $A = 1$ is harmful to subjects with $V = 0$ and beneficial to those with $V = 1$.
To examine this hypothesis, we would elaborate the MSM,

$$\mathbb{E}(Y^{\overline{a}} \mid V) = \eta_0 + \eta_1 \text{cum}(\overline{a}) + \eta_2 V + \eta_3 \text{cum}(\overline{a}) V.$$

Then we have qualitative effect modification if $\text{sign}(\eta_1) \neq \text{sign}(\eta_1 + \eta_3)$.
We can e.g. use the weights,

$$\frac{\prod_{k=0}^{K} p(A_k \mid V, \overline{A}_{k-1})}{\prod_{k=0}^{K} p(A_k \mid \overline{L}_k, \overline{A}_{k-1})}$$

in a weighted least squares regression model.
One thing to remember: Here, IPW is used to adjust for confounding and regression modelling is used to study effect modification.

MSMs and direct effects

To illustrate a point, consider the saturated MSM for two binary treatments A_0, A_1,

$$\mathbb{E}(Y^{\bar{a}}) = \mathbb{E}(Y^{a_0,a_1}) = \eta_0 + \eta_1 a_0 + \eta_2 a_1 + \eta_3 a_0 a_1.$$

Now, the direct effect of A_0 when A_1 is set to 1 is $\mathbb{E}(Y^{1,1}) - \mathbb{E}(Y^{0,1})$.
How do we articulate the hypothesis that $\mathbb{E}(Y^{1,1}) = \mathbb{E}(Y^{0,1})$?

$$\begin{aligned} \mathbb{E}(Y^{1,1}) &= \mathbb{E}(Y^{0,1}) \\ \eta_0 + \eta_1 + \eta_2 + \eta_3 &= \eta_0 + \eta_2 \\ 0 &= \eta_1 + \eta_3 \end{aligned}$$

Optimal regimes and dynamic MSMs

Suppose that we aim to find the optimal treatment regime g^* in a given class of regimes $\{g = x : x \in \mathcal{X}\}$, where $|\mathcal{X}| = m$. Suppose that $x \in \{0, 1, \dots, 999\}$. Let $n = 2000$ individuals.

- Suppose I come up with the following strategy: Run an experiment and randomly assign the regime g.
- Maximize $\hat{\mathbb{E}}(Y \mid X = x)$.
- Problem: We have m regimes, but only 2000 people so $\hat{\mathbb{E}}(Y \mid X = x)$ will be too variable...we will expect to have two people receiving the regime.
- Running example: Once we have started treatment (say, antiretroviral therapy in patients with HIV), then we never stop treatment. The question is: what is the best X to start treatment?

Dynamic MSMs

- Constructing an MSM allows us to impose assumptions, and then borrow strength across the regimes g, for example by assuming that $\mathbb{E}(Y \mid X = x) = \mathbb{E}(Y^x)$ is smooth in x.
- Note that we have to do this even if the data are from an experiment.
- Idea: for example, suppose we fit the model

$$\mathbb{E}(Y^x) = \eta_0 + \eta_1 x + \eta_2 x^2 + \eta_3 x^3.$$

- Then, we find the optimal regime g^* by maximising $\eta_1 x + \eta_2 x^2 + \eta_3 x^3$ over x.
- However, because there may be qualitative effect modification, we can expand the model to

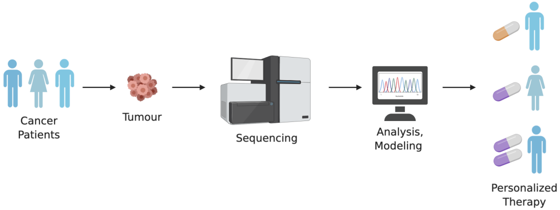
$$\mathbb{E}(Y^x \mid V) = \eta_0 + \eta_1 x + \eta_2 x^2 + \eta_3 x^3 + \eta_4 x V,$$

and for each value of V maximize $\eta_1 x + \eta_2 x^2 + \eta_3 x^3 + \eta_4 x V$ over x, $g(v) = \arg\max_{x \in \mathcal{X}} \eta_1 x + \eta_2 x^2 + \eta_3 x^3 + \eta_4 x v$

Advantages of MSMs

- Easy to understand.
- Can be fitted with standard statistical software.

Precision medicine is a buzz word



My claim:

Modelling the disease process is of secondary importance in precision medicine, except when it helps support the identification (and estimation) of optimal regimes.

Precision medicine is a buzz word, and the idea is simple

- The idea is to tailor treatment decisions to patient characteristics.
- The premise: *individual heterogeneity* can be leveraged to *individualize therapy*.
- Work on causal inference gives us theory for *optimizing* individual decisions.
  - What if patient i receives treatment A vs. treatment B?
    That is, what is the causal effect of taking A vs. B...

Algorithmic vs. human decisions

- Decision rules might be algorithmically individualized.
- Yet these rules will be implemented under supervision of humans (e.g., doctors).
- Are optimal algorithmic regimes better than human-decision rules?
  - Care providers may have information that is not recorded in the observed data.

    $\Longrightarrow$ unmeasured confounding in the data.
  - So, when should we let humans override algorithmic treatment recommendations?

...but causal inference requires strong assumptions, no?

- We need to take the causal question seriously.

  Scientists who choose not to give up causal inference must understand that, without selecting a definition of a causal effect, it is impossible to evaluate whether we have reasonably estimated one.

- Can we deal with unmeasured confounding?
  - Sometimes we can point identify effects in the presence of unmeasured confounding.

    Instrumental variables, front-door variables, negative controls (proximal inference) ...
  - Other times we can bound the causal effects.

  Transparency about study goals and the assumptions we make to justify an analysis are required to discuss bias, refine our questions and improve our answers.