EXERCISES FOR RANDOMIZATION AND CAUSATION (MATH-336)

Exercise Sheet 7

Exercise 1 (IPW and M-estimation)**.** In this exercise we will study the asymptotic properties of the IPW estimator. Consider a sample $\mathcal{S} = \{(A_1, L_1, Y_1), \ldots, (A_n, L_n, Y_n)\}$ of iid replicates of (A, L, Y) such that $Y^a \perp\!\!\!\perp A \mid L$ but $Y^a \not\perp\!\!\!\perp A$, with $A \in \{0, 1\}$ and L and Y discrete (with finite support). Hereafter we will assume the propensity score $\pi(a \mid l) = P(A = a|L = l)$ is known. We also assume consistency and positivity.

(a) Write down the expression for the IPW estimator of the ATE of A on Y,

$$\hat{\text{ATE}}_{\text{IPW}} = \hat{\mu}_{\text{IPW}}(1) - \hat{\mu}_{\text{IPW}}(0).$$

(b) Prove that $\hat{\text{ATE}}_{\text{IPW}}$ is a consistent estimator of $E[Y^1 - Y^0]$, i.e.,

$$\hat{\text{ATE}}_{\text{IPW}} \xrightarrow{p} E[Y^1 - Y^0].$$

(c) Define the $\hat{\text{ATE}}_{\text{IPW}}$ estimator as an M-estimator.

(d) Prove that $\hat{\text{ATE}}_{\text{IPW}}$ is a consistent estimator of $E[Y^1 - Y^0]$ without using the same arguments as in point b).

(e) Suppose now the propensity score is unknown and that

$$\pi(l) := \pi(1 \mid l) = \text{expit}(\gamma_0 + l\gamma_1) \text{ for some } \gamma = (\gamma_0, \gamma_1) \in \Gamma \subseteq \mathbb{R}^2.$$

  (i) write down the expression for the IPW estimator of the ATE of A on Y;

  (ii) prove that $\hat{\text{ATE}}_{\text{IPW}}$ is a consistent estimator of $E[Y^1 - Y^0]$ when we posit a correctly specified model for the propensity score and we estimate β via maximum-likelihood estimation. Can you still use the same arguments as in point b)? Is such an IPW estimator a maximum-likelihood estimator?

Exercise 2 (A comparison of variance)**.** (From [1], Homework 2)

Suppose that the outcome and propensity model are known. Consider two estimators for the average response: $\frac{1}{n}\sum_{i=1}^{n} Y_i^{a=1}$ and $\frac{1}{n}\sum_{i=1}^{n} \frac{A_i Y_i^{a=1}}{\pi(A_i|L_i)}$.[1]

(a) By assuming conditional exchangeability $Y_i^a \perp\!\!\!\perp A_i \mid L_i$, show that the first has lower variance than the second (that is, we pay some penalty for not observing all subjects in the data set being treated).

Hint: Show that the second estimator can be written as the first plus something else, and then demonstrate that the two terms are uncorrelated.

(b) Compute the difference in variance between the estimators in (a) if A is randomized with probability $P(A = 1) = \frac{1}{2}$ (i.e. $\pi = \frac{1}{2}$)

Exercise 3 (Stabilized IPW estimators)**.** (Technical Points 12.1 and 12.2 in [2]) Let A, L, Y denote treatment, baseline covariates and outcome respectively and suppose the usual assumptions of conditional exchangeability, positivity and consistency hold.

[1] The first estimator is an estimator that is typically impossible to compute because all the counterfactuals are not observed. However, in this exercise we have assumed that $Y_i^{a=1}$ is observed.

(a) Show that we can identify $E[Y^a]$ from

$$E[Y^a] = \frac{E\left[\frac{I(A=a)Y}{\pi(A|L)}\right]}{E\left[\frac{I(A=a)}{\pi(A|L)}\right]} .$$

This form of the identification formula motives a modified version of the IPW estimator called the Hajek estimator (or stabilized IPW estimator):

$$\hat{\mu}_{STIPW}(a) = \frac{\frac{1}{n}\sum_{i=1}^{n} \frac{I(A_i=a)Y_i}{\pi(A_i|L_i;\gamma)}}{\frac{1}{n}\sum_{i=1}^{n} \frac{I(A_i=a)}{\pi(A_i|L_i;\gamma)}} . \tag{1}$$

(b) Show that

$$E[Y^a] = \frac{E\left[\frac{I(A=a)Yg(A)}{\pi(A|L)}\right]}{E\left[\frac{I(A=a)g(A)}{\pi(A|L)}\right]}$$

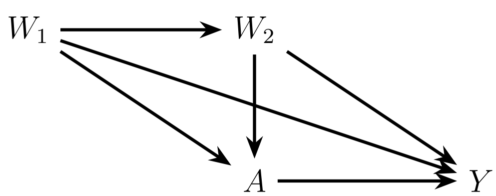
and that

$$\hat{\mu}_{STIPW} = \frac{\frac{1}{n}\sum_{i=1}^{n} \frac{\hat{g}(A_i)}{\pi(A_i|L_i;\gamma)} \cdot I(A_i = a)Y_i}{\frac{1}{n}\sum_{i=1}^{n} \frac{\hat{g}(A_i)}{\pi(A_i|L_i;\gamma)} \cdot I(A_i = a)},$$

where $g(A)$ is a function of A, and is consistently estimated by $\hat{g}(A)$. We refer to $\frac{g(A)}{\pi(A|L)}$ as stabilized weights because they are, in settings where rely on parametric assumptions, often smaller than the regular IPW weights $\frac{1}{\pi}$, and can thus give rise to estimators with a smaller variance.

Exercise 4 (Exploring the IPW estimator)**.** (Based on Lab 4 of [3])

In this exercise we will implement the IPW and Hajek estimators numerically in R in order to explore their efficiency in cases with near violations of positivity. Consider treatment A and outcome Y with baseline covariates W_1, W_2 in the dataset `stabilized_weights.csv`, and suppose these satisfy the causal model below: The data was generated by drawing



$n = 5000$ i.i.d. samples from the distributions

$$\begin{aligned}
W_1 &\sim Ber\left(p = \frac{1}{2}\right) \\
W_2 &\sim Multinom\left(1; (0.125, 0.375, 0.375, 0.125)\right) \\
A &\sim Ber\left(p = \text{logit}^{-1}(-1.3 - 3W_1 + 3W_2)\right) \\
Y &\sim Ber\left(p = \text{logit}^{-1}(-2 - 2W_1 + 3W_2 + 3A + 2AW_2)\right) \\
Y^{a=1} &\sim Ber\left(p = \text{logit}^{-1}(-2 - 2W_1 + 3W_2 + 3 \cdot 1 + 2 \cdot 1 \cdot W_2)\right)
\end{aligned}$$

$$Y^{a=0} \sim Ber\left(p = \text{logit}^{-1}(-2 - 2W_1 + 3W_2 + 3 \cdot 0 + 2 \cdot 0 \cdot W_2)\right),$$

subject to the constraint

$$Y = Y^{a=1} I(A=1) + Y^{a=0} I(A=0) \ .$$

The true effect is given by $E[Y^{a=1} - Y^{a=0}] \approx 0.26$ (computed by evaluating $\frac{1}{n'}\sum_{i=1}^{n'}(Y_i^1 - Y_i^0)$ in a larger realization of the data with $n' = 100000$) .

(a) Import the dataset `stabilized_weights.csv` into R and use the `glm` command to perform the following logistic regression for the treatment mechanism $\pi(A \mid L)$:

$$\text{logit } \pi(A \mid L; \gamma) = \gamma_0 + \gamma_1 W_1 + \gamma_2 W_2 \ .$$

Plot the empirical cumulative distribution function of the IPW weights $\frac{1}{\pi(A_i|W_{1,i},W_{2,i})}$ and use the weights to evaluate the IPW estimator

$$\hat{\mu}_{IPW} = \frac{1}{n}\sum_{i=1}^{n} \frac{I(A_1 = a)Y_i}{\pi(A_i \mid W_{1,i}, W_{2,i}; \gamma)} \ .$$

(b) Compute $\hat{\mu}_{IPW}$ with truncated weights $\frac{I(\pi \leq 10)}{\pi} + 10 \cdot I(\pi > 10)$ instead of the weights $\frac{1}{\pi}$ in part (a).

(c) Evaluate the stabilized IPW estimator given by Eq. 1 using the weights as in part (a).

(d) Estimate the variance of the estimators in parts (a)-(d) by drawing $R = 5000$ different realizations of a population with $n = 5000$ i.i.d. individuals from the data generating mechanism outlined above.

Exercise 5. (Logistic regression model) We would like to estimate the effects of a pesticide on the statue of stink bugs in a farm. We observe the statue of n stink bugs, and let Z_i be the binary outcome of the experiment for the stink bug i. Y is the sum of Z_i and corresponds to the number of stink bugs that are observed to be alive after the termination of experiment.

(a) What distribution is reasonable to assume for Y if each stink bug is given the same dosage of pesticide? What assumption does that require making on the Z_i?

(b) Now assume stink bug i is given a specific dosage of pesticide, namely $x_i > 0$. Using logistic model, state the probability that a bug survives in terms of the constant β_0 and linear coefficient β_1.

(c) Describe how to fit the parameters of the linear model given data Z_i.

(d) Recall from the statistics course that for large sample size n, the variance of the MLE estimator is given by the inverse of the Fisher information (In other words, the variance achieves Cramer-Rao bound asymptotically). Assume $\beta_0 = 0$ and calculate the Fisher information and find an asymptotic estimate for the variance of $\hat{\beta}_1$.

(e) What assumptions were required to write down the likelihood function?

References

[1] David M. Vock. PubH 7485 & 8485: Methods for Causal Inference (University of Minnesota School of Public Health).

[2] Miguel Hernan and James M. Robins. *Causal Inference*. Chapman & Hall, 2018.

[3] Maya L. Petersen and Laura B. Balzer. Labs & Assignments.