Exercises for Causal Thinking (Math-352)

November 17, 2025

1 Exercise Sheet 9

Exercise 1. (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) *(Challenging)* 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?

Exercise 2 (Stabilized IPW estimators)**.** (Technical Points 12.1 and 12.2 in Hernan and Robins [2018]) Let A, L, Y denote treatment, baseline covariates and outcome respectively and suppose the usual assumptions of conditional exchangeability, positivity and consistency hold.

(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.

References

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