# 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 $\beta_0$ and linear coefficient $\beta_1$.

(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?

**Solution:**

(a) We can assume $Z_i$s are i.i.d. Bernoulli variables and thus $Y$ have Binomial distribution.

(b) We can assume $Z_i$ are independent Bernoulli variables such that

$$\mathbb{E}(Z_i) = g^{-1}(\beta_0 + x_i\beta_1), \quad g(\mu) = \log(\frac{\mu}{1-\mu}).$$

Because $x_i > 0$, and we can expect that increasing the dosage of pesticide, would decrease the probability that fruit flies survive, thus $\beta_1$ should be negative.

(c) The likelihood for the model is

$$l(\beta) = \sum_{i=1}^{n} z_i \log p(x_i, \beta) + (1 - z_i) \log(1 - p(x_i, \beta))$$
$$= \sum_{i=1}^{n} \{z_i(\beta_0 + \beta_1 x_i) - \log(1 + e^{\beta_0 + \beta_1 x_i})\}$$

where $p(x_i, \beta) = \frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}$. Thus we want to solve the equation $\frac{\partial l(\beta)}{\partial \beta} = 0$. We can calculate $\frac{\partial^2 l(\beta)}{\partial \beta \partial \beta^T}$ , for $\beta^T = (\beta_0, \beta_1)$, and starting from some initial $\beta$, use the Newton-Raphson method with the following iterations:

$$\beta^{new} = \beta^{old} - (\frac{\partial^2 l(\beta)}{\partial \beta \partial \beta^T})^{-1} \frac{\partial l(\beta)}{\partial \beta}.$$

(d)

$$\frac{\partial l}{\partial \beta_1} = \sum_{i=1}^{n} x_i \big(z_i - \frac{e^{\beta_0+\beta_1 x_i}}{1 + e^{\beta_0+\beta_1 x_i}}\big).$$

We compute the Fisher information $I(\beta) = -\mathbb{E}[\frac{\partial^2}{\partial \beta^2} l(\beta)]$ by calculating the second derivatives of $l$:

$$\frac{\partial^2}{\partial \beta^2} l(\beta) = -\sum_{i=1}^{n} x_i^2 \frac{e^{\beta_0+\beta_1 x_i}}{(1 + e^{\beta_0+\beta_1 x_i})^2}.$$

Thus for large sample size $n$, one can estimate the variance of $\hat{\beta}$ by the inverse of the calculated Fisher information above.

(e) The logistic regression model is correctly specified that is, when the $Y_i$'s are truly independent random variables with distribution Bernoulli$(p_i)$, where the logit$(p_i)$ is the same linear combination of the covariates $x_i$.

**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.

**Solution:**

(a) The expectation in the denominator is 1, because

$$\begin{aligned} E\left[\frac{I(A=a)}{\pi(A \mid L)}\right] &= E\left[\frac{I(A=a)}{P(A=a \mid L)}\right] \\ &= E\left[\frac{1}{P(A=1 \mid L)} E\left[I(A=a) \,\middle|\, L\right]\right] \\ &= E\left[\frac{P(A=1 \mid L)}{P(A=1 \mid L)}\right] \\ &= 1 \; . \end{aligned}$$

(b) Then,

$$\frac{E\left[\frac{I(A=a)Yf(A)}{\pi(A|L)}\right]}{E\left[\frac{I(A=a)f(A)}{\pi(A|L)}\right]} = \frac{E\left[\frac{I(A=a)Yf(a)}{\pi(A|L)}\right]}{E\left[\frac{I(A=a)f(a)}{\pi(A|L)}\right]}$$

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

$$= E[Y^a] \ .$$

Likewise,

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

$$= \frac{\hat{f}(a)\frac{1}{n}\sum_{i=1}^{n}\frac{I(A_i=a)Y_i}{\pi(A_i|L_i;\gamma)}}{\hat{f}(a)\frac{1}{n}\sum_{i=1}^{n}\frac{I(A_i=a)}{\pi(A_i|L_i;\gamma)}}$$

$$= \hat{\mu}_{STIPW}(a) \ .$$

# References


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