POL SCI 231B (Spring 2022): Problem Set 3

Prof. Thad Dunning / GSI Clara Bicalho
Dept. of Political Science
University of California, Berkeley

Distributed Friday, March 4; due Wednesday, March 16 at 3:59 PM.

Please email only one copy of your solution set per group to clara.bicalho@berkeley.edu. Make sure to *include "Problem Set [GROUP NAME]" in the email subject or body field.*

Remember to work out the problems on your own, before you meet with your group to agree on solutions.

1. After reading Gerber and Green Chapter 8, complete

- Exercise 1, (a)-(d) on p. 283. Note that the reference to section 8.4 in (b) should refer to section 8.2 or 8.3.
- Exercise 9, (a)-(c) on p. 286 (note that while you apprear to need an ISPS login for the .csv file containing data for Figure 4 at htpp://isps.research.yale.edu/FEDAI, you can access data for exercise 9(c) at https://isps.yale.edu/research/data/d081
- Exercise 11, (a)-(b)

2. Consider an experimental sample of groups with three units each. Half of all units are assigned to treatment at random. Suppose a model where there is spillover within groups but not across groups. Specifically, if any individual in a group receives treatment, the spillover effect is spread equally among members of the same group.

(a) Define the set of all potential outcomes involved in this set up.

(b) Define the direct and spillover effects under the potential outcomes framework.

(c) Using the data set `data_spillover.RDS` provided on bCourses, estimate the direct and spillover effects.

(d) Should we use inverse propensity weights in this setting? Why / why not?

3. **Maximum Likelihood Estimation**

Likelihood Function. Suppose that we observe the following independent data:

```
y <- c(10,4,5,3,9,2,7,3,6,4)
```

(a) Plot the log-likelihood function using the Poisson distribution (hint: how do we get the likelihood function of independent data?). Recall that a Poisson distribution describes a discrete variable X using a parameter λ such that for $k = 0, 1, 2, ...$

$$Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

The positive number λ is equal to the expected value of X and also its variance.

$$\lambda = E(X) = \text{Var}(X)$$

Looking at the plot, what does the maximum appear to be? Here is some code that might help you.

```
# 1. write a function for the log-likelihood
# with lambda as the input
log.lik <- function(lambda) { LOG LIKELIHOOD FUNCTION }

# 2. generate a vector with the possible values of lambda
p.l <- seq(0,30,0.1)

# 3. plot the log likelihood function
plot(p.l, lapply(p.l, log.lik))
```

(b) Now use R to solve for the MLE. There are a variety of commands you can use (you can check optim or optimize, for example). Below is some example code using optim.

```
mle <- optim(
  c(1), # starting value to look for max

  log.lik, # the function we are optimizing

  control=list(fnscale=-1), # this tells optim to maximize (vs minimize)

  lower=0, # lower bound for optimization
  upper=100, # upper bound

  method="Brent" # one dimensional optimization method
)

mle
```

(c) Now compare the MLE estimate of λ to the mean of y. Interpret your results.

4. **Logistic regression.** A scholar is interested in evaluating the effect of $X \in \mathbb{R}$ (X is continuous from $-\infty$ to ∞) on the probability that $Y = 1$. (Y is dichotomous). The

scholar estimates two different logistic regression models M_1 and M_2.

M_1:

$$\text{Prob}(Y_i = 1) = \Lambda(\alpha_1 + \beta_{X1} X_i + \gamma_1 Z_{i,M_1}) \tag{1}$$

and

M_2:

$$\text{Prob}(Y_i = 1) = \Lambda(\alpha_2 + \beta_{X2} X_i + \gamma_2 Z_{i,M_2}). \tag{2}$$

Here, Λ is the logistic distribution function. There are possibly different covariates in each model: Z_{i,M_1} in the first model and Z_{i,M_2} in the second model. Also, γ_j and Z_{i,M_j} may be vectors. X is the same variable in models $M1$ and $M2$.

Now, suppose the scholar first estimates logistic regression model M1; this yields an estimated coefficient for X of $\hat{\beta}_{X1} = 0.3$. The scholar then estimates the second model M2; this yields an estimated coefficient for X of $\hat{\beta}_{X2} = 0.8$.

For the following questions, we are setting aside the more vexing problems for causal inference—like, both models can't be simultaneously true. Just assume the models in each case and answer the questions, taking each model as given:

(a) For each model $M1$ and $M2$, express the marginal effect of a one-unit increase in X on the probability that $Y_i = 1$ in terms of the models in equations (1) and (2), respectively.

(b) The M2 estimate implies that the estimated effect on $Pr(Y = 1)$ of increasing X by one unit is greater than is implied by the M1 estimate. True or false? Explain your answer.

(c) Holding the baseline probability of Y constant, the M2 estimate implies that the estimated effect on $Pr(Y = 1)$ of increasing X by one unit is greater than is implied by the M1 estimate. True or false? Explain your answer.

(d) For any particular unit (say India-Pakistan-1986), the M2 estimate implies that the estimated effect on $Pr(Y = 1)$ of increasing X by one unit is greater than is implied by the M1 estimate. True or false? Explain your answer.

5. **Probit Model and Interpretation.** For this exercise you will need the `camp1.dta` dataset available on bcourses.[1] Consider the probit model with democratic win (`DWIN`) as the dependent variable, and the following independent variables:
   - `JULYECQ2`: 2nd Quarter GNP Growth
   - `PRESINC`: Elected Incumbent Seeking Reelection (1=Democrat, 0=no incumbent, 1-=Republican)
   - `ADAACA`: State Liberalism Index (ADA & ACA)
   - An interaction term between `PRESINC` and `JULYECQ2`.

We have the following hypotheses:

H1: *Economic growth in the months prior to the election increases the chances that the Democrat will win.*

[1] You will need to use the `foreign` library to open the dataset.

H2: *The Democrat has a better chance of winning if he/she is the incumbent President.*

H3: *The more liberal the state, the more likely the Democrat will win.*

H4: *Growth prior to the election only helps the Democrat if he or she is the incumbent.*

(a) Run the analysis in R including an intercept in the model. You can write your own function or use *glm* in R.

```
glm(y ~ x, family = binomial(link = "probit"))
```

(b) Suppose you could manipulate *ADAACA*, the state liberalism index. (Just suppose). What is the estimated effect of that variable on a Democratic win, holding all other variables at their median values in the dataset?

(c) Arbitrarily define two levels (low and high) for the variable measuring growth, `JULYECQ2`, so that 1.5 is `high" growth and $-1.5$ is`low” growth.

Consider the following four scenarios:

1) High growth, incumbent Democratic candidate running.

2) High growth, incumbent candidate not running.

3) Low growth, incumbent Democratic candidate running.

4) Low growth, incumbent candidate not running.

Across the range of values of `ADAACA` in the dataset, plot the predicted value of Prob(`DWIN` $= 1$) for each of the four scenarios. (You can use the `predict()` function in `R` to generate their respective predicted values). Generate two figures: one with the plots for scenarios 1 and 3, and the other with those for scenarios 2 and 4. Please provide a substantive discussion of your results in light of the above hypotheses.