POLS/CSSS 510:
Maximum Likelihood Methods for the Social Sciences

Problem Set 1

Professor: Christopher Adolph, Political Science and CSSS

Fall Quarter 2018

Due in class on Thursday 11 October 2018

General instructions for homeworks: Homework can be handwritten or typed. For any exercises done with R or other statistical packages, you should attach all code you have written and all (interesting) output. Materials should be stapled together in order by problem. The most readable and elegant format for homework answers incorporates student comments, code, output, and graphics into a seamless narrative, as one would see in a textbook.

Problem 0: R refresher (optional)

[0 points.] If you are new to R, I recommend you work through the R practice exercises contained in Problem Set 1, Problem 1, parts **a** to **r** from POLS/CSSS 503, the preceding course in this sequence. This problem set can be found at `http://faculty.washington.edu/cadolph/503/503hw1.pdf`, and the data used can be found at `http://faculty.washington.edu/cadolph/503/democracy.csv`. (This optional problem carries no credit but will be corrected if submitted.)

Problem 1: Three ways to solve probability problems

[20 points.] The fraction of males in a given population is 0.49. What is the probability of sampling 22 males in a random sample of size 30? How about sampling 16 males in a random sample of size 30? Calculate your answers by hand. Next, verify your results using R's built-in functions for calculating probabilities. Finally, write R code to simulate each of these scenarios many times, and report the results of your simulations.

Problem 2: Working with marginal, joint, and conditional probabilities

[24 points.] Consider two random variables X and Y. X can take values equal to 1, 2, or 3, and Y can take values equal to 1, 2, 3, 4. We know the following joint probabilities:

- $f_{X,Y}(1,1) = 0.05$
- $f_{X,Y}(1,2) = 0.08$
- $f_{X,Y}(1,3) = 0.13$
- $f_{X,Y}(1,4) = 0.03$
- $f_{X,Y}(2,1) = 0.07$
- $f_{X,Y}(2,3) = 0.16$
- $f_{X,Y}(2,4) = 0.04$
- $f_{X,Y}(3,1) = 0.06$
- $f_{X,Y}(3,2) = 0.07$
- $f_{X,Y}(3,3) = 0.15$
- $f_{X,Y}(3,4) = 0.06$

Using what you know about the relationships between marginal, conditional, and joint probability functions and the information above, compute the following quantities by hand (you should check your results in R):

a. $f_X(1)$

b. $f_X(2)$

c. $f_X(3)$

d. $f_Y(1)$

e. $f_Y(2)$

f. $f_Y(3)$

g. $f_Y(4)$

h. $f_{X,Y}(2,2)$

i. $f_{X|Y}(1|2)$

j. $f_{X|Y}(3|4)$

k. $f_{Y|X}(4|3)$

l. $f_{Y|X}(2|2)$

Hint: It may help to arrange the available joint probabilities in a table.

Problem 3: Student project checkpoint

This problem must be submitted separately, one copy per paper-writing group.

[40 points.] Identify and describe the dependent variable to be analyzed in your research paper, and provide an (approximately) 2 paragraph sketch of your research design. Among other things, this sketch should answer the following questions:

a. What is the unit of analysis and source of data?

b. What probability distribution best describes the generation of your data? (If you have the data in hand, you should provide histograms and/or density plots of the dependent variable, and use them to support your answers.)

c. Are there any assumptions underlying this distribution that you suspect may be violated by the data?

d. Based on what you have learned so far, what methods might be appropriate for analyzing your data?

e. What are the covariates of interest (or treatment variables), and what other covariates do you plan to include in your models?

NB: The best answers to this question take the form of a brief essay presenting your planned research, rather than a series of bullet points. Your instructor will provide feedback on your paper proposals, so see this as a starting point subject to revision, rather than a commitment to a certain research plan or a polished prospectus proposal.