

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

Problem Set 2

Professor: Christopher Adolph, Political Science and CSSS

Fall Quarter 2018

Due in class on Tuesday 23 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 1: Working with profile likelihoods

[16 points.] Consider the following dataset: $\mathbf{y} = (1, 0, 0, 1, 0, 0, 0, 0)$, which happens to be a series of independent realizations of a Bernoulli distributed random variable. Plot the likelihood function for the Bernoulli parameter π given \mathbf{y} .¹ If the experiment were repeated, roughly what fraction of the observations would you expect to be successes ($y = 1$)? Why?

¹ Use the Bernoulli likelihood to solve this problem, rather than the Binomial likelihood. It is easy to check your result using the Binomial applied to a single observation, but the goal is to learn to apply a likelihood to a vector of data.

Problem 2: Writing and testing a useful new maximum likelihood estimator

To solve this problem, it will help to consult King, *Unifying Political Methodology*, on the Poisson likelihood.

- a. [10 points.] Derive the log-likelihood function corresponding to this pdf:

$$f_{\text{Poisson}}(\lambda) = \exp(-\lambda)\lambda^y/y!$$

- b. [10 points.] Suppose that for periods $i \in \{1, \dots, i, \dots, n\}$, each of (potentially varying) length t_i , we observe counts y_i . We would like to use the Poisson distribution to model these data, but an assumption of the Poisson distribution (and hence the Poisson regression model) is that the periods observed are of equal length. Relax this assumption, and revise your derivation in part a accordingly.
- c. [10 points.] Write an R function, usable by `optim()`, implementing this variable-period Poisson estimator.
- d. [10 points.] Generate an artificial dataset with 1000 observations consisting of three variables:

$$x_{1i} \sim \text{Uniform}(0, 1) \tag{1}$$

$$x_{2i} \sim \text{Uniform}(0, 1) \tag{2}$$

$$y_i \sim \text{Poisson}(t_i\lambda_i) \tag{3}$$

where

$$\lambda_i = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}) \tag{4}$$

and t_i is a random draw from the integers $\{1, 2, 3, 4, 5\}$; this is the length of the period. Assume that the true values of the model parameters are $\beta_0 = 0$, $\beta_1 = 1$, $\beta_2 = 2$. Confirm that the mean of y_i is approximately 16.5, and the standard deviation of y_i is approximately 14.7 (your results may differ by as much as one unit in each case due to Monte Carlo error).

- e. [10 points.] Using the artificial data you generated in part d and the R function for the unequal periods Poisson likelihood you wrote in part c, estimate the parameters β_0 , β_1 , and β_2 and report your point estimates and their standard

errors. Does this exercise give you confidence that you have correctly derived and implemented the Poisson MLE for unequal periods? What could you do to increase your confidence further?

Problem 3: Solving the birthday problem using simulation

[30 points.] A famous probability problem asks “What is the probability that at least two people in a classroom of n people share the same birthday?” Write an R program to solve this problem using simulation. Produce as output a plot of the probabilities of at least one shared birthday for $n = \{2, \dots, 50\}$.