MLE Exercise

Grid Search - Poisson Distribution

The following data show the number of appointments to the United States Supreme Court from 1961 through 1992 by presidential term. Given that these are events which occur over a fixed interval of time, it is not unreasonable to suppose that they follow a poisson distribution.

The Poisson distribution is:

$$y_i \sim f(\lambda, y_i) = \frac{e^{-\lambda}\lambda^{y_i}}{y_i!} \tag{1}$$

The likelihood is:

$$L(y_1, y_2, \ldots, y_8|\lambda) = \frac{e^{-8\lambda}\lambda^{\sum y_i}}{\prod y_i!} \tag{2}$$

The data are:

Years	Number Appointed (y_i)
1961-1964	2
1965-1968	2
1969-1972	4
1973-1976	1
1977-1980	0
1981-1984	1
1985-1988	3
1989-1992	2

1. Use maximum likelihood estimation via a grid search to estimate the parameter (λ) of the poisson distribution (to 3 decimal places). Hint: λ lies between 0 and 6. Do this in STATA.
   - First, generate a range for lambda from 0 to 6 with increments of 0.001.

     ```
     set obs 6000
     egen l=seq(), from (0) to (6001)
     gen lambda=l/1000
     ```

   - Calculate the likelihood function for each value of λ.

     ```
     generate likelihood= (you fill in)
     ```

   - Graph likelihood function against λ and see where the maximum is. Show graph.
   - Get STATA to tell you what λ maximizes the likelihood function.
2. We usually maximize the log of the likelihood function rather than the likelihood function. Write down the log of the likelihood function for the sample. Now repeat the steps from above

ie. graph the log-likelihood function against λ to find the maximum and then get STATA to tell you the λ that maximizes the log-likelihood function. Show graph. (Do this directly i.e. write the log-likelihood equation in STATA - do not simply take the log of the likelihood column obtained in question 1).

3. Now you can take the log of the likelihood column from question 1 and make sure that you get the same result from question 2.

4. You will observe that only part of the log-likelihood is a function of λ. Thus, the maximum does not depend on the stuff that does not vary with λ. Redo question 2 but dropping the parts of the log-likelihood function that do not vary with λ.

5. With the information that you have gathered copy the following table and fill it in.

Table 2: Results

	Value	λ
Likelihood (L)		
Log Likelihood (lnL)		
lnL'		

where lnL' is just the part of the log-likelihood function that varies with λ.

6. Using the answer to question 4, do we need to know the individual values of y_i to find λ? If not, what information about y_i is sufficient to find λ?

7. Use the estimated parameter $\hat{\lambda}$ to estimate the probability that Bill Clinton would have 0, 1, 2, 3, 4, or 5 appointments to the Supreme Court during his first term of office. Show this information in a graph or a table.

Exercise 2: Grid Search - Binomial Distribution

Until recently, Japan's electoral system used multi-member districts with single non-transferable votes. The following table shows the number of seats won by the LDP (the ruling party during most of the period in which this electoral system was employed) in 6 elections in a single district. The district has 5 seats. We observe only the number of seats won. From this, we would like to estimate the probability that the LDP wins each seat. We must assume that the winner of each seat is independent of the winner of all other seats (probably a dubious assumption, but . . .). Since the district magnitude is fixed at 5 seats, the binomial distribution seems appropriate for this problem.

The binomial distribution is:

$$y_i \sim f(\pi, y_i) = \frac{N!}{y_i!(N-y_i)!}\pi^{y_i}(1-\pi)^{N-y_i} \tag{3}$$

where N is the number of possible events (winning 5 seats in this case).

The likelihood is:

$$L(y_1, y_2, y_3, \ldots, y_6) = \frac{N!^6}{\prod y_i!(N-y_i)!}\pi^{\sum y_i}(1-\pi)^{\sum(N-y_i)} \tag{4}$$

The data are:

Election	LDP Seats (y_i)
1	2
2	4
3	2
4	4
5	5
6	3

1. Calculate the maximum likelihood estimate of π using a grid search (to 3 decimal places). Obviously, π lies between 0 and 1. Do this in STATA the same way you did for the previous exercise. Graph the likelihood function against π and see where the maximum is. Show graph. Get STATA to tell you what π maximizes the likelihood function.

2. Write down the log of the likelihood function for the sample. Now repeat the steps from above ie. graph the log-likelihood function against π to find the maximum and then get STATA to tell you the π that maximizes the log-likelihood function. Show graph. (Do this directly i.e. write the log-likelihood equation in STATA - do not simply take the log of the likelihood column obtained in question 1).

3. Now you can take the log of the likelihood column from question 1 and make sure that you get the same result from question 2.

4. You will again observe that only part of the log-likelihood is a function of π. Thus, the

maximum does not depend on the stuff that does not vary with π. Redo question 2 but dropping the parts of the log-likelihood function that do not vary with π.

5. With the information that you have gathered copy the following table and fill it in.

Table 4: Results

	Value	π
Likelihood (L)		
Log Likelihood (lnL)		
lnL'		

where lnL' is just the part of the log-likelihood function that varies with π.

6. Using the answer to question 4, do we need to know the individual values of y_i to find π? If not, what are the sufficient statistics to find π?

7. Use the estimate of $\hat{\pi}$ to calculate the probability distribution for seat outcomes from 0 to 5 LDP seats won.

Exercise 3: OLS via MLE

tab9s.dta is a STATA data set based on congressional elections in 1992; it has all elections where the incumbent ran for election. The dependent variable (cv92) is the challenger's vote in the 1992 election. The independent variables are all labelled in obvious ways. To see what the variables are, type DESC in STATA.

1. Run a regression of CV92 on prior office (PO), LNCE92 (log of challenger expenditure) and a constant. Put the output of interest in a table of results.

2. Now using the log likelihood for a regression (assuming independent observations), use STATA to estimate the same model via MLE. The code will look something like that shown below - note that this is a do-file which sets up the ML program, then calls it twice - once to set it up and once to run it.

```
clear;

capture program drop mlols;

program mlols;

version 10.0;

args lnf mu sigma;

quietly replace ‘lnf’=ln(normalden($ML_y1,‘mu’,‘sigma’));

end;

use "E:\essex2009\mle\tab9s.dta", clear;

ml model lf mlols (mu: cv92 = po lnce92) (sigma:);

ml maximize;
```

3. This code uses numerical derivatives and Hessians. Run it and put the results in another column of your table of results. Compare the OLS regression results with the MLE version. What is identical? What is close but not identical? Why?

4. STATA offers an option to check that you have written the ML code correctly. Become familiar with the output by typing:

```
ml model lf mlols (mu: cv92 = po lnce92) (sigma:), tech(nr);

ml check
```

5. You can also look at the individual steps of the maximization process by typing:

   ```
   ml model lf mlols (mu: cv92 = po lnce92) (sigma:), tech(nr);
   ```

   ```
   ml maximize, showstep
   ```

   Take a look at this.

6. You should now try to use different maximization algorithms. You should try the Newton-Raphson and the BHHH. To do this, you will have to type:

   ```
   ml model lf mlols (mu: cv92 = po lnce92) (sigma:), tech(nr);
   ```

   ```
   ml maximize;
   ```

   or

   ```
   ml model lf mlols (mu: cv92 = po lnce92) (sigma:), tech(bhhh);
   ```

   ```
   ml maximize;
   ```

   or

   ```
   ml model lf mlols (mu: cv92 = po lnce92) (sigma:), tech(bfgs);
   ```

   ```
   ml maximize;
   ```

7. Now graph the process of convergence by typing:

   ```
   ml graph
   ```

   after each maximization.

8. Run the ML program with the Newton-Raphson technique. However, before the program has finished iterating, hit the 'break' key. Now type ML QUERY and/or ML REPORT. What is the information that STATA is now giving you. You may have to use STATA's help command - you should have some intuition though. Now type ML MAXIMIZE again. What happens?

9. This should not be an exercise in whether you can copy from your notes. The only way you are going to learn this is if you can do this yourself - thus, don't do this question in a group, though you may ask for help from others and check the notes if necessary.

   - Write down the likelihood for a regression
   - Write down the log-likelihood for a regression
   - What is the gradient for the log-likelihood function?
   - What is the Hessian?

- Take the expectation of the Hessian and then the inverse of this. What does this tell you?

10. In less than one paragraph, what does your regression tell you about what affects challenger vote? What is the effect of a one-unit change in the log of challenger expenditure? Don't forget that your estimate is uncertain, so report the uncertainty.