Problem Set 1

Due on Mar 5th 17:00

Please note:

- All question shall be relevant from class materials but not every part is taught in class. In other word, some questions are a little "innovative" so to train your programing and problem solving skill.
- It is normal to get stuck. Keep working. Go back to class code and exercise in the textbook for practice. Discuss, try, and retry.
- Submit your answers (in Microsoft Word or PDF format) and your code. Your answer shall be well written. Graph and Table shall be well-formated. Your code shall be easy for TA to run and check. Your grade will be affected if your code does not provide proper output, or it is confusing so that TA cannot run it.

1. Consider the data set "SAT.csv"

a. Load this data set into R. Compile a summary statistics table for all numerical and dummy variables for this data set.

b Make a plot with 2×2 diagrams. Each diagram shows the histogram of these variables: GPA, SAT, APMath, APEng.

c. Make a scatter plot of SAT and GPA. Add a linear regression line and a non-linear (lowess) regression line on the plot. You may refer to this link <http://www.statmethods.net/graphs/scatterplot.html>.

d. Use matrix algebra to compute a least square estimator for regression model:

$$\text{GPA}_i = \beta_0 + \beta_1 \text{SAT}_i + \beta_1 \text{APMath} + \beta_2 \text{APEng} + \beta_3 \text{ESL} + \beta_4 \text{gender} + \beta_5 \text{race}$$

You may follow the following insturctions:

(i) Generate a vector with 65 one's as the constant regressor.
(ii) Make a X matrix that each regressor constitute a column. The matrix shall be 65×7.
(iii) Make a column vector y using GPA.
(iv) Compute $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$.
(v) Compare with the result from build-in least-square function "lm".

2. In this exercise, we will simulate rock-paper-scissor game and show how memorizing and infering from past "data" can help a player increase winning probability.
Two player submit their action simultaneously and the payoff is shown as follow

		2		
		Rock $[r_2]$	Paper $[p_2]$	Scissor
	Rock $[r_1]$	0, 0	-1, 1	1, -1
1	Paper $[p_1]$	1, -1	0, 0	-1, 1
	Scissor	-1, 1	1, -1	0, 0

For programming convenience, denote Rock=1, Paper=2, Scissor=3.

a. Write a function that compute **player 1**'s payoff given player 1 and 2's actions as input. (e.g. $\pi_1(1,2) = -1$)

b. Think of two players that take each action with certain probability: e.g. $\Pr(i \text{ plays Rock}) = r_i$. Write a function that compute expected payoff of **player 1**, given r_1, p_1, r_2, p_2, that is

$$E\pi_1(r_1, p_1, r_2, p_2) = r_1 [p_2 - (1 - r_2 - p_2)] + p_1 [r_2 - (1 - r_2 - p_2)] + (1 - r_1 - p_1) [-r_2 + p_2].$$

Show that a Scissor lover ($r_1 = 0.1,\ p_1 = 0.1$) has disadvantage facing Rock lover ($r_2 = 0.8$, $p_2 = 0.1$) .

c. Write a function that can simulate a naïve player that play with fixed probabilities r_i and p_i (a typical mixed strategy). Suppose player 2 that plays $r_2 = 0.4$, $p_2 = 0.3$ (a slight Rock lover), simulate his action for $S = 5000$ games.

d. Create an Artificial Intelligent (AI) player BetaGo that can beat a naïve player. Specifically, follow these instructions:

(i) At the beginning, without any information, BetaGo shall play the best mixed strategy with $p_1 = 1/3$, $r_1 = 1/3$.

(ii) BetaGo can memorize all actions that it seen from its opponent. After 10 games, BetaGo starts to compute empirical probabilities $\hat{r}_2$,$\hat{p}_2$ of his opponents.

(iii) If BetaGo find out his opponent is incline to play certain action with more than $1/3$ probability, then it adjusts his strategy to take advantage.

(iv) Show that BetaGo can "beat" a naïve player 2 with $r_2 = 0.4$, $p_2 = 0.3$.

(v) Show that BetaGo has more advantage as number of games (S) becomes larger.

[Note: This last exercise is difficult and take time. There is lots of way to make even BetaGo stronger. For example, using Bayesian learning, instead of frequentiest learning. It is beyond the scope of this course.]

3. Simulation of two-stage least square (2SLS) estimator with instrumental variables.
We want to use Monte-Carlo simulation to study the property of 2SLS estimator with IV. We will first generate a data set with some x variable being endogenous and some z variable can serve as IV. Specifically, consider the following regression model of interest:

$$y_i = \beta_0 x_i + e_i, \quad i = 1, 2, ..., n.$$

However, the data here suffer from endogeneity issue, i.e. $E[x_i e_i] \neq 0$. There is an instrumental variable z_i that is both valid and relevant. Consider a first-stage regression

$$z_i = \gamma_0 x_i + u_i, \quad i = 1, 2, ..., n,$$

where z_i satisfied $E[z_i e_i] = 0$.

a. Set the sample size as $n = 100$. We first generate a random sample of error terms with the following bivariate normal distribution

$$\begin{pmatrix} e_i \\ u_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}\right).$$

Plot a scatter plot of e_i's and u_i's. [Hint: directly programing a bivariate normal distribution function is very difficult. You shall seek a package that can generate data from such a distribution.]

b. Let $\beta_0 = 1$, $\gamma_0 = 1$, $z_i \sim \text{Gamma}(5)$. generate a data set $\{x_i, z_i, y_i\}$ that satisfiies the setting above.

c. Set the number of simulation $S = 1000$, graphically show that OLS estimator of β_0 is biased.

d. The two-stage least square (2SLS) estimator using instrumental variable is

$$\hat{\beta}_{2SLS} = (\sum_{i=1}^{n} z_i x_i)^{-1} (\sum_{i=1}^{n} z_i y_i).$$

Graphically show that 2SLS estimator of β_0 is unbiased. (This formula is for one variable case. A more general formula shall be found in Hansen book, page 253).

e. Graphically show that 2SLS estimator of β_0 is consistent.

4. In this exercise, we will write a function that can numerically solve utility maximization problem. This function feeds in a utility function u, prices p_x, p_y, and income w. It then solves the utility maximization problem numerically:

$$\max_{x,y} u(x,y), \quad \text{s.t. } p_x x + p_y y \leq w.$$

This utility maximization problem gives output of indirect utility function $v(p_x, p_y, w)$ and demand functions $x^*(p_x, p_y, w)$, $y^*(p_x, p_y, w)$.

You may follow the following instruction:

a. Start with a particular example: $u(x,y) = x^2 y^3$, $p_x = 1$, $p_y = 2$, $w = 10$. Define this utility function in R.

[Hint: for later convenience of picking a bundle, it's better to define the utility function as function of bundle, i.e. $\mathbf{x} = (x,y)$, $u(\mathbf{x})$.]

b. Define a grid of points representing bundles. For example $(0.01, 0.01), (0.01, 0.02), (0.01, 0.03)$, $..., (0.02, 0.01), (0.02, 0.02), (0.02, 0.03), ...$

c. Find a set of bundles that satisfies the budget constraint $p_x x + p_y y \leq w$. You may use a scatter plot to check whether it is indeed a triangle.

d. Compute utility for each bundle in the budget set, find the maximum. The bundle is (x^*, y^*) and the maximum is the value of indirect utility $v = u(x^*, y^*)$. Report the result.

e. Generalize the above step into a big function. The function shall take u, p_x, p_y, w as input, and report x^*, y^*, v as output.

f. Check whether the function works by these inputs: $u(x,y) = x^2\sqrt{xy}$, $p_x = 5$, $p_y = 4$, $w = 100$. Report the result.

g. Following the setting in step f, compute $x^*$for each p_x in a sequence $(1, 1.5, 2, 2.5, ..., 5)$. Show a demand curve of x.