Problem Set 3

Due on May 19th, 5:00pm

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 programming 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-formatted. 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. Maximum likelihood
Consider the following data generate process. Let $\sigma = 2$, $\beta_1 = 1$, $\beta_2 = 0.5$, $e_i \sim$ i.i.d.$N(0,\sigma^2)$, $x_i \sim$i.i.d.Gamma(2), and

$$y_i = \beta_1 + \beta_2 x_i + e_i.$$

a. Generate a random sample (e_i, x_i, y_i) with $n = 200$.

b. In an econometric exercise, we observe a x_i and y_i but not e_i. Our parameters of interest is $\theta = (\beta_1, \beta_2, \sigma)$. Suppose we know the parametric family of the condition distribution is normal. From this DGP, we know that $y_i \sim$i.i.d.$N(\beta_1 + \beta_2 x_i, \sigma^2)$, that is

$$f(y_i|x_i,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \beta_1 - \beta_2 x_i)^2}{2\sigma^2}\right).$$

Write down the log-likelihood function and use maximum likelihood to estimate θ.

c. Use bootstrap ($B = 200$) to find standard errors of $\hat{\theta}_{ML}$ you obtained in part b.

d. Compute OLS estimate of β_1 and β_2 by OLS. Compare the results from MLE and OLS.

2. GMM. Load the data y.csv.
a. Plot its empirical CDF and PDF using commend from np package.

b. Suppose we know that the distribution of y can be described by two parameters a and b. The following two moment conditions hold:

$$\begin{aligned} E[y_i] &= \frac{a}{b} \\ E[y_i^2] &= \frac{a(a+1)}{b^2} \end{aligned}$$

Find a set of GMM estimate $\hat{a}, \hat{b}$.

c. Use bootstrap to compute standard errors of $\hat{a}$ and $\hat{b}$. Plot the densities of bootstrap estimates of $\hat{a}$ and $\hat{b}$ respectively.

3. Nonparametric regression and classification. Load the data ipod.csv.

b. We want to use number of bidders (BIDRS) to predict PRICE. Use OLS, KNN (with 4 level of κ) and nonparametric regression. Illustrate them in diagrams.

c. Use k-means algorithm to cluster the sample by BIDRS and PRICE. Illustrate the case for $\kappa = 2, 3, 4, 5$.

4. Lasso. Load the data psidps3.csv. Let wagert be the outcome variable.

a. Run a OLS regression with all explanatory variable. Record R-squared.

b. Run a default lasso regression. Plot the coefficient path.

c. Run a lasso with cross-validation. Use MSE+SE rule to find the best model. Compute the R-squared for this model.