

ECEn 370

Homework Problem Set 9

Due on Friday, March 9, 2012.

From Bertsekas and Tsitsiklis, *Introduction to Probability, 2nd Ed.*
A bit more practice is included for you on derived distributions.

1. Chapter 4 Problem 9
2. Chapter 4 Problem 10
3. Chapter 4 problem 14
4. Chapter 4 Problem 17
5. Chapter 4 Problem 18
6. Chapter 4 Problem 19
7. Chapter 4 Problem 23
8. Chapter 4 Problem 24
9. MATLAB Problem

I have taken the following from the CMU DASL statistics website:

“<http://lib.stat.cmu.edu/DASL/Datafiles/carmpgdat.html>”.

Use the following code to import the data using MATLAB:

“`A=importdata('car_data_1991_for_HW.txt')`” to store the data in a variable A.

The headings of the columns are given by “`A.textdata(1,2:6)`” where VOL is the cubic feet of cab space, HP is engine horsepower, MPG is average miles per gallon, SP is top speed, and WT is vehicle weight in 100 lb increments.

The numerical information is found in “A.data”.

Use the “corr” command to find the correlation between two columns of data.

a) Plot weight of the car versus MPG. Are these highly correlated? Are these positively or negatively correlated? Why do you think that is the case?

b) It turns out that the *linear* least-mean-square estimator of \hat{Y} can be found by (see page 438 for more information if you wish):

$$\hat{Y} = E[Y] + \rho \frac{\sigma_Y}{\sigma_X} (X - E[X])$$

What is the line that best fits the data found in part a? Plot this line on your graph in part a.

c) Plot the cab space of the car versus engine horsepower. Plot the best fit line for the data. Are these highly correlated? Are these positively or negatively correlated? Why do you think that is the case?

d) Suppose that $\rho = 0$. What kind of estimator does that give you for Y . Why is that reasonable?