ECE3093 Optimization, Estimation and Numerical Methods. Stochastic Part

Assignment 2 (2013) Part A

This is an individual and not group assignment. This assignment will be released in three separate parts (through Moodle ECE3093 folder Assignments Part2 2013):

Part A Fantasia heart beat data analysis

Part B Gaussian vector simulation

Part C ARMA time and frequency domain simulation

Your work on Part A, Part B and Part C is to be written up and submitted as three separate reports; one for each part. It expected that you will write a MATLAB m-file for each part of this assignment. Advice on how to write these is given to make this as straight forward as possible. If you require further assistance, please seek help as soon as possible.

At Level 3 you are expected to be mature enough to understand that your assignments must be your own work. Accordingly, it is expected that your MATLAB scripts and their output will be your own, original work. All students who submit identical MATLAB scripts or identical output (eg plots of simulations) will all received a mark of zero for these parts of the assessment.

Your reports are to be submitted through the assignment box for Mr Alan Couchman (ground floor, Building 28) at Clayton Campus (Clayton students) or to Dr Nader Kamrani at Sunway Campus (Sunway students).

Deadlines for reports are:

Part A WEEK9, before 6pm Wednesday 8th May 2013

Part B WEEK11, before 6pm Monday 20th May 2013

Part C WEEK12, before 6pm Friday 31st May 2013

NB Include a printout of your m-files for parts A, B and C in the body or as an appendix to your reports.

Part A Fantasia heartbeat interval variation analysis

The RR interbeat period is the period of time between successive depolarisation signals of the heart. It is measured as the period of time between successive R peaks of the QRS complex of the ECG signal. Assignment 2 Part A is about characterising the heart rate variability (HRV); the variation in the interbeat period, uisng the RR time series of the subset of five young Y1,...,Y5, and five elderly O1,...,O3 individuals, is a selected subset of the Fantasia Dataset; http://www.physionet.org/physiobank/database/fantasia/.

(I) Fantasia heartbeat datasets (see Moodle ECE3093 Assignments 2013 Part 2)

(1) Download to a folder, the heartbeat5O.m and heartbeat5Y.m m-files. and the Fantasia heartbeat datasets, Y1.txt,...,Y5.txt and O1.txt,..O5.txt.

(2) Run the m-files in MATLAB to aquaint yourself with these datasets.

The time-series is a plot of interbeat period versus time. The Poincare plot is a plot of the interbeat period versus the next interbeat period, and the histogram plot shows the distribution of interbeat periods for the data set.

Answer the following questions and include your answers in your report.

(a) What are the characteristics of the time-series plot of a typical Y1,..,Y5? How do these differ from the characteristics of the time-series plot of a typical O1,...,O5?

(b) What are the characteristics of the Poincare plot of a typical Y1,..,Y5? How do these differ the characteristics of the Poincare plot of a typical O1,...,O5?

(c) What are the characteristics of the histogram plot of a typical Y1,..,Y5? How do these differ the characteristics of the histogram plot of a typical O1,...,O5?

(d) Do any members of the groups Y1,...,Y5 or O1,...,O5 appear atypical? If yes, how do their time-series, Poincare and histrograms differ from other members of the group?

(e) What features of the time-series plots do you think correspond to the acceleration, deceleration and idling of the heart?

(3) Using suitable MATLAB functions, write an M-file that plots a horizontal stack of boxplots for all Y1.txt,...,Y5.txt,O1.txt,..O5.txt.

Use the axis command to ensure that the 10 boxplots all have the same horizontal (and vertical) scale.

Relevant MATLAB commands are: boxplot, subplot, axis, quantile.

(a) Plot a horizontal stack of 10 boxplots for the raw data of Y1,...,Y5,O1,..O5..

(b) Normalise each data set. Subtract the median (Q2) from each data set and divide it by the interquartile range (IQR = Q3-Q1).

(c) Plot a horizontal stack of 10 boxplots for the normalised data of Y1,...,Y5,O1,..O5.

(d) Include plots of the (a) raw stack and the (c) normalised stack of horizontal boxplots in your report.

(e) interpret the stack boxplots. What do quartiles and the extent of their data say about the variability of the interbeat period among these subjects?

(f) Do the stacked boxplots support your conclusions in (1)(a),(b),(c) and (d)?

(II) Multiscale analysis using detrended fluctuation analysis (DFA).
Look at the diagrams on page R1080 of the research paper by N. Iyengar et al 1996 (see Assignment 2013 Part2 Fantasia heartbeat datasets folder in Moodle). You are to generate a log_{10}F(n) versus log_{10}n plot (see N. Iyengar Fig 2 R1080) for each of the two Fantasia dataset specified by your DFA1 and DFA2; which you will find in the 'ECE3093 Clayton Student ID based numbers Z U V A a b DFA' or 'ECE3093 Sunway Student ID based numbers Z U V A a b DFA' file in the Assignments 2013 Part2 folder in Moodle. You have been allocated two of Y1,Y2,Y3,Y4,Y5,O1,O2,O3,O4, O5, to do DFA on.

Your task is to write a MATLAB m-file that will correctly perform DFA on a heartbeat dataset, with the end result that you obtain the slope or slopes of the log_{10}F(n) versus log_{10}n plot and the corresponding β values, for both of the data sets you have been assigned. Include the plots of the log_{10}F(n) versus log_{10}n in your report. The plots should have the data points and have the lines of best fit plotted or drawn on them.

The basic steps in the DFA procedure are as follows.

(Step 1) Subtract from the N points of your dataset, their average value. eg hmm = heartbeat - mean(heartbeat).

(Step 2) Now integrate the data from Step 1. eg y =cumsum(hmm)

(Step 3) Divide the data up into m boxes (segments). $m - 1$ of equal length n.

(Step 4) For each box. Fit a linear regression line through the data in the box, and substract the trend form the data. ie y = y $-(a_k i + b_k)$ for the kth box.

(Step 5) Calculated F(n). F(n)= $\sqrt{\sum_{i=1}^{N}(y - (a_k i + b_k))^2/N}$ for box $k = 1, ..., m$.

(Step 6) Repeat Step 3 to Step 5 for enough box different sizes in the range 4 to 300, to get a reasonably well defined log_{10}F(n) versus log_{10}n plot.

(Step 7) Determine the piecewise slopes of the log_{10}F(n) versus log_{10}n plot; α_1 and α_2.

(Step8) Determine the value or values of $\beta = 2\alpha - 1$. β = slope of the log-log spectral plot (Bode diagram).

Step 3 and Step 4 are the difficult steps. Step 3: How can the dataset be divided up into boxes? Once boxed, Step 4: How do we find the trend line, y=a+bx for each box?

Hints for dealing with Steps 3-7

There is not skeleton file, but if you need help on how to read the data into MATLAB, see the downloads, heartbeatO5.m etc.

(Step 3) (i) You are going to plot a log-log graph, so choose box lengths, n, in the range 4 to 300, that give approximatley equal spacing of $log(n)$.

(ii) The length of the data is will usually not be exactly divisable by the box length, so there has to be a box on the end with a different length; you will need to decide whether to have a longer box, or to make a extra box. mod() and fix() may be useful.

(iii) To do a linear regression on the data in each box, you need to know the index at the start and end of each data segment. You have to figure out how to do this eg you might set up arrays of 'left' and 'right' indices.

(Step 4) (i) MATLAB provides [r,m,b] = regression(tbox,ybox). You need to construct the input vectors tbox and ybox. Note MATLAB's polyfit can also be used.

(ii) The regession line for the box is, yreg = b + m*tbox;

(iii) The detrended data for the box is, ydet = ybox − yreg.

(iv) You need to work along the dataset, box by box, so that at the end, you have separately detrended every box.

(v) Check the correctness of your detrending by plotting the detrended data and the integrated data from (Step 2) on a single plot.

Include a plot of the detrending for your largest box size in your report.

(Step 5) (i) F(n) is the 'root mean square' of the detrended data; the 'root mean square fluctuation'.

(ii) The sum of the squares of all the detrended data can calculated using a single inner dot product; $\sum_{i=1}^{N} ydet^2 =$ ydet'*ydet.

$$\mathrm{F(n)}= \sqrt{\sum_{i=1}^{N}(y-(a_k i+b_k))^2/N} = \sqrt{\sum_{i=1}^{N}(ydet)^2/N}$$

(Step 6) (i) You need to calculate (n,F(n)) pairs for a range of box sizes; so you can plot, log_{10}F(n) versus log_{10}n.

(ii) To plot log_{10}F(n) versus log_{10}n you only have to plot loglog(n,F(n)).

(Step 7) Step 7 will require you to do a regression in logarithmic coordinates to obtain α_1 and α_2

[r1,alpha1,b1] = regression(log(boxlength(1:n_1)),log(F(1:n_1)));

where n_1 = number of point you use to fit α_1. And, if you are keen, you can plot the regression line on the log-log plot by using data $(n, regF(n))$ where reg$F(n) = 10^{b_1} n_k^{alpha_1}$

(Step 8) As the paper by N. Inyengar explains, $\beta = 2\alpha - 1$ where (see N. Iyengar Fig 1 R1079)

(1) $\beta = 0$ for white noise

(2) $\beta = 1$ for $1/f$ noise

(3) $\beta = 2$ for a random walk (brownian motion)

END OF PART A