Time Series Analysis
Coursework

Handout: Monday 22 November 2021.
Deadline: **Submit electronic copy to the turnitin assignment on Blackboard before Friday 17 December 2021, 4pm.**

This coursework is worth 10% of your mark for Time Series Analysis.

There are 3 questions. **Together they are worth 40 marks**

Plots and tables should be clear, well labelled and captioned. **Marks will be deducted for a poorly presented report.**

Comment your code. Marks will be deducted for uncommented code and *very* inefficient coding.

For the computational elements, you must use `R`, `MATLAB` or `Python`. It is up to you which you choose. All three are equally valid.

You may use any results from the notes you wish, but you must properly cite them.

You must type up your report (MS Word, LaTeXor a notebook (e.g. Jupyter) is fine), including all your code within the main text (not as appendices or screen shots). **You must submit a pdf**.

Your report must be no more than 15 pages (not including a cover page). This does not mean you have to use all 15 pages - you should be able to do this in fewer than 15 pages.

NB: Stationarity by itself always means 'second-order' stationarity. $\{\epsilon_t\}$ denotes white noise with mean zero and variance σ_ϵ^2. Assume $\{\epsilon_t\}$ is Gaussian/normal here. You can assume a sampling interval of $\Delta t = 1$ throughout.

PLAGIARISM IS A VERY SERIOUS OFFENCE.
YOU MUST SUBMIT YOUR OWN PIECE OF WORK.
CASES OF PLAGIARISM WILL BE REPORTED TO COLLEGE REGISTRY.

Question 1

This question is related to material learnt in weeks 3, 6 and 7.

(a) Write a function `S_ARMA(f,phis,thetas,sigma2)` that computes the (theoretical) spectral density function for an ARMA(p,q) process. If $p = 0, q > 0$, i.e. you pass in an empty array for `phis`, then it will compute the spectrum of an MA(q) process. If $p > 0, q = 0$, i.e. you pass in an empty array for `thetas`, it will compute the spectrum of an AR(p) process.

The inputs should be:

`f`: the vector of frequencies at which it should be evaluated.

`phis`: the vector $[\phi_{1,p}, ..., \phi_{p,p}]$.

`thetas`: the vector $[\theta_{1,q}, ..., \theta_{q,q}]$

`sigma2`: a scalar for the variance σ_ϵ^2 of the white noise $\{\epsilon_t\}$.

Orders p and q should not be inputs to the function, instead they need to be computed from the lengths of `phis` and `theta`. **3 marks**

(b) Write *your own* function `ARMA22_sim(phis,thetas,sigma2,N)` that simulates a *Gaussian* ARMA(2,2) process of length `N`. The function will use a burn in method, i.e. set $X_1 = X_2 = 0$, and $X_t = \phi_{1,2}X_{t-1} + \phi_{2,2}X_{t-2} + \epsilon_t - \theta_{1,2}\epsilon_{t-1} - \theta_{2,2}\epsilon_{t-2}$ for $t > 2$. The function should then discard the first 100 values so that it reaches an (approx.) stationary state. Return $X_{101}, ..., X_{101+N}$ as the time series of length N.

It should return a single output:

`X`: a vector of length `N` for values of the time series. **3 marks**

(c) Using `fft`[1], write two functions.

- `periodogram(X)` that computes the periodogram at the Fourier frequencies for a time series `X`.
- `direct(X,p)` that computes the direct spectral estimate at the Fourier frequencies using the $p \times 100\%$ cosine taper for a time series `X`.

3 marks

(d) An ARMA process exhibits pseudo-cyclical behaviour when its autoregressive part exhibits pseudo-cyclical behaviour, as described in the notes. In this question, we will explore how dynamic range effects bias. Recall, the roots of the AR characteristic function which exhibits pseudo-cyclical behaviour can be written as

$$z_1 = \frac{1}{r}e^{i2\pi f'} \qquad z_2 = \frac{1}{r}e^{-i2\pi f'}.$$

You need to write a script that calls the functions you made in (a), (b) and (c). It should perform the following tasks:

A. Using $\theta_{1,2} = -0.5$ and $\theta_{2,2} = -0.2$, simulate 10,000 realizations, each of length $N = 128$, of an ARMA$(2,2)$ process that shows pseudo-cyclical behaviour at $f' = 12/128$ where $r = 0.8$. For each realization, compute the periodogram and four direct spectral estimates using a cosine taper with $p = 0.05$, $p = 0.1$, $p = 0.25$ and $p = 0.5$. Store the values for each at frequencies 12/128, 32/128 and 60/128 (i.e. on the oscillating frequency and two frequencies away from the oscillating frequency).

[1] the fft algorithm computes the Fourier transform at the Fourier frequencies $f_k = k/N$, $k = 0, ..., N-1$.

B. Compute the *sample* bias (using `S_ARMA`) of the periodogram and four direct spectral estimators at frequencies 12/128, 32/128 and 60/128 from the 10,000 realizations.

C. Repeat steps A and B for $r = 81, 82, 83, ..., 99$.

D. Present three plots, one for each frequency, that compare the spectral estimators for different values of r.

6 marks

(e) Briefly comment on the results. You may want to consider plotting the true spectral density function and computing the dynamic range of the process at two or three values of r to support your explanation. **2 mark**

Question 2

This question relates to material learnt in week 8.

You each have your own individual time series $x_1, ..., x_{128}$ which can be retrieved by downloading

www2.imperial.ac.uk/∼eakc07/time_series/number.csv

where instead of "number" you insert the number next to your name in `time_series_number.pdf` on blackboard, e.g. www2.imperial.ac.uk/∼eakc07/time_series/72.csv.

You will need to type out this URL. Copying from the pdf will not work.

You can assume your time series has a mean of zero.

(a) Compute the periodogram and direct spectral estimate using the 50% cosine taper at the Fourier frequencies. Plot each of these on separate axes over the range $[-1/2, 1/2)$. You may want to read up about `fftshift`.

2 marks

(b) Write **your own code** to fit an AR(p) model using each of the following *methods*:

- Yule-Walker (untapered);
- Yule-Walker (50% cosine tapered);
- approximate maximum likelihood.

4 marks

(c) It can be shown that for stationary Gaussian AR processes, the AIC reduces down to

$$\text{AIC} = 2p + N \ln(\hat{\sigma}_\epsilon^2).$$

For $p = 1, 2, ..., 20$, fit an AR(p) model. Create a 20×3 table to neatly display the AIC for each method and each value of p. **3 marks**

(d) For each method, which order model do you select as being the best fit for your data? Report the $p + 1$ estimated parameter values for each method. **2 marks**

(e) For the three selected models, plot the associated spectral density functions on a single axis. HINT: use your function from Question 1(a).

2 marks

Question 3

This question relates to material learnt in week 10. You will use the same time series as you used in Question 2.

(a) Assume that in fact you only observed values $x_1, ..., x_{118}$. Fit and select the model again using approximate maximum likelihood on just the observed values. Forecast $X_{119}, ..., X_{128}$ using the selected model and parameter estimates. Compare them to the actual values using a table.

HINT: An argument similar to that used for AR(1) processes follows for forecasting an AR(p) process, i.e. set future innovation terms to zero.

5 marks

(b) Point forecasts, like you have just given, on their own only have limited use. What is much more useful is supplying an accompanying prediction interval. This is an interval which has some designated probability of containing the realised trajectory.

We will look at a Monte Carlo simulation method of providing these.

If I simulate the innovation terms $\{\epsilon_{119}, ..., \epsilon_{128}\}$ and iterate the AR model, I can obtain one possible future trajectory of $X_{119}, ..., X_{128}$. If I repeat this procedure 999 times, I can get 999 possible trajectories for $X_{119}, ..., X_{128}$. Then, if at each time I take the 50th and 950th largest values of those 999, I can build 90% prediction intervals at times $t = 119, ..., 128$. That is, a set of intervals which have a 90% chance of containing the true, realised trajectory $x_{119}, ..., x_{128}$.

On a single axes, from $t = 100$ to $t = 128$ plot the true trajectory, the predicted trajectory and the 90% prediction interval (the last two will only be from $t = 119$ to $t = 128$.)

5 marks