Time Series Analysis
Coursework

Handout: Monday 23 November 2020.

Deadline: **Submit electronic copy to the turnitin assignment on Blackboard before Friday 18 December 2020, 4pm.**

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

There are 4 questions. **Together they are worth 20 marks.**

Plots and tables should be 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, e.g. overly unnecessary uses of `for` loops.

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 or LaTeXis fine), including all your code within the main text (not as appendices or screen shots), and submit a pdf.

Your report must be no more than 12 pages (not including a cover page).

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.

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

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 (3 marks)

This question is related to material learnt in weeks 3, 7 and 8.

(a) Write a function `S_AR(f,phis,sigma2)` that evaluates the parametric form of the spectral density function for an $\mathrm{AR}(p)$ process on a designated set of frequencies.

The inputs should be:

`f`: a vector of frequencies at which to evaluate the spectral density function.

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

`sigma2`: the variance of the white noise.

Your function should not take p as an input, but instead determine it from the stated input parameters.

It should return a single output:

`S`: a vector of values of the spectral density function evaluated at the elements of `f`.

(b) Write a function `AR2_sim(phis,sigma2,N)` that simulates a *Gaussian* $\mathrm{AR}(2)$ process of length `N`. The function will use a burn in method, i.e. set $X_1 = 0$, $X_2 = 0$, and $X_t = \phi_{1,2}X_{t-1}+\phi_{2,2}X_{t-2}+\epsilon_t$ for $t > 2$. The function should then discard the first 100 values and return $X_{101}, ..., X_{100+N}$ as the time series of length N.

The inputs should be:

`phis`: the vector $[\phi_{1,2}, \phi_{2,2}]$.

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

`N`: the length of the process to be simulated.

It should return a single output:

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

(c) Write a function `acvs_hat(X,tau)` that computes the estimate of autocovariance $\hat{s}_\tau^{(p)}$ for an input time series `X`, at designated values of `tau`.

The inputs should be:

`X`: the vector $[X_1, X_2, ..., X_N]$ for the time series.

`tau`: a vector of the values of τ at which to evaluate $\hat{s}_\tau^{(p)}$. E.g. `tau = [0,1,...,N-1]`.

It should return a single output:

`s_hat`: a vector of values of the autocovariance sequence estimate evaluated at the elements of `tau`.

Question 2 (7 marks)

This question is related to material learnt in weeks 7 and 8. Specifically, you will explore the bias properties of the periodogram and direct spectral estimator for AR processes.

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

- `periodogram(X)` that computes the periodogram at the Fourier frequencies of a time series stored as a vector `X`.

- `direct(X)` that computes the direct spectral estimate at the Fourier frequencies using the Hanning taper, defined here as

$$h_t = \frac{1}{2}\left[\frac{8}{3(N+1)}\right]^{1/2}\left[1 - \cos\left(\frac{2\pi t}{N+1}\right)\right], \quad t = 1, \ldots, N.$$

(b) A. Using `AR2_sim(phis,sigma2,N)`, simulate 10000 realizations, each of length $N = 16$, of an AR(2) process with complex conjugate roots

$$\frac{1}{r}e^{i2\pi f'} \quad \text{and} \quad \frac{1}{r}e^{-i2\pi f'},$$

where $f' = 1/8$ and $r = 0.95$. Set $\sigma_\epsilon^2 = 1$. For each realization, compute the periodogram and direct spectral estimates and store the values for each at frequencies $1/8$, $2/8$ and $3/8$.

B. Compute the *empirical* bias (using `S_AR(f,phis,sigma2)`) of the periodogram and direct spectral estimators at frequencies $1/8$, $2/8$ and $3/8$ from the 10000 realizations.

C. Repeat steps A and B for $N = 32, 64, 128, 256, 512, 1024, 2048$ and 4096.

D. Present plots that compare the two spectral estimators for different values of N. You should consider using a log scale for the x-axis.

(c) Briefly comment on the results. You may want to consider plotting the true spectral density function for the process to support your explanation.

---

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

Question 3 (7 marks)

This question relates to material learnt in week 9.

You each have your own individual time series of length 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 this out. 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 Hanning taper (see Question 2) at the Fourier frequencies. Plot each of these on separate axes over the frequency range $[-1/2, 1/2]$. You may want to read up about `fftshift`.

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

   - Yule-Walker (untapered);
   - (forward) Least Squares;
   - approximate maximum likelihood.

(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.

(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.

(e) For the three selected models, plot the associated spectral density functions on a single axis.

Question 4 (3 marks)

This question relates to material learnt in week 11. You will use the same time series as you used in Question 3.

Assume that in fact you only observed values $X_1, ..., X_{118}$. For each of the three methods (YW, LS and ML), forecast $X_{119}, ..., X_{128}$ using the selected models and parameter estimates from Question 3(d). Compare them to the actual values using a table and plot the 4 time series (true, YW forecast, LS forecast, ML forecast) from time point 110 to 128.

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.