UNIVERSITY OF BRITISH COLUMBIA
Department of Statistics
Stat 443: Time Series and Forecasting
Assignment 4: Analysis in the Frequency Domain

The assignment is due on **Thursday, April 7** at **9:00pm**.

- Submit your assignment online in the **pdf format** under module "Assignments".
- Include all steps of your derivations as partial marks will be given.
- Please make sure your submission is clear and neat. It is the student's responsibility that the submitted file is in good order (e.g., not corrupted and is what you intend to submit).
- **Late submission penalty**: 1% per hour or fraction of an hour. (In the event of technical issues with submission, you can email your assignment to the instructor to get a time stamp but submit on canvas as soon as it becomes possible to make it available for grading.)

1. Consider the following stochastic process:

$$X_t = Z_t + 0.7Z_{t-1} + 0.2Z_{t-2} - 0.1Z_{t-3}, \qquad t \in \mathbb{Z},$$

where $\{Z_t\}_{t\in\mathbb{Z}} \sim WN(0,1)$.

(a) Derive the power spectral density function of $\{X_t\}_{t\in\mathbb{Z}}$.

(b) Write down the normalized spectral density function of $\{X_t\}_{t\in\mathbb{Z}}$.

(c) Plot the normalized spectral density and comment on its behaviour.

2. (a) Show that if $\{X_t\}_{t\in\mathbb{Z}}$ and $\{Y_t\}_{t\in\mathbb{Z}}$ are independent, stationary processes with power spectral density functions $f_X(\omega)$ and $f_Y(\omega)$, respectively, then the process $\{W_t\}_{t\in\mathbb{Z}}$ with $W_t = X_t + Y_t$ is also stationary with power spectral density function

$$f_W(\omega) = f_X(\omega) + f_Y(\omega), \qquad \omega \in (0, \pi).$$

(b) Suppose $\{X_t\}_{t\in\mathbb{Z}}$ is an AR(1) process with $X_t = -0.5X_{t-1} + Z_t$, and $\{Y_t\}_{t\in\mathbb{Z}}$ and $\{Z_t\}_{t\in\mathbb{Z}}$ are two independent white noise processes with mean zero and common variance σ^2. Derive the power spectral density function of $\{W_t\}_{t\in\mathbb{Z}}$, where $W_t = X_t + Y_t$.

3. (This question must be completed in R Markdown; display all the R code used to perform your data analysis.)

The Southern Oscillation Index (SOI) measures the normalised pressure difference between Tahiti and Darwin. The data file `soi.txt` gives the annual SOI between 1866 and 2010.

(a) Read the data into R and coerce the data into a time series object. Plot the resulting time series and its sample acf. Comment on what you observe. (Make sure to properly label the axes and provide titles for the plots.)

(b) Plot the periodogram with a logarithmic vertical scale for the series, and smooth the periodogram by setting the argument `spans = sqrt(2 * length(x))`, where `x` is your time series object. Comment on what you observe and estimate the wavelength and angular frequency for the dominating frequency.

(c) Build a function in `R` that generates the Fourier frequency ω_p for a given time series and given constant $p \in \{0, 1, ..., N/2\}$. Document the inputs and outputs of this function, so that another person would be able to understand how to use your function. What is the output of your function for $p = 10$?

(d) To determine which Fourier frequencies are "significant", suppose we were to fit the linear model

$$Y_t = a_0 + a_p \cos(\omega_p t) + b_p \sin(\omega_p t) + \epsilon_t, \qquad t = 1, \ldots, N,$$

where we assume $\epsilon_t \sim N(0, \sigma^2)$ for all t and are independent. Let $\{Y_t\}_{t=1,...,N}$ be the time series and $\omega_p = \frac{2\pi p}{N}$. Note that the model will be fitted $N/2$ times for each $p = 1, 2, ..., N/2$. On fitting the above model for a given p by least squares, a test of the "significance" of the contribution of frequency ω_p is a test with null hypothesis

$$H_0 : a_p = b_p = 0,$$

that uses the F-test statistic

$$F_p = \frac{\frac{1}{k-1} \sum_{t=1}^{N} (\hat{y}_{t,p} - \overline{y})^2}{\frac{1}{N-k} \sum_{t=1}^{N} (y_t - \hat{y}_{t,p})^2},$$

where k is the number of estimated coefficients in the linear model, N is the number of observations, $\hat{y}_{t,p} = \hat{a}_0 + \hat{a}_p \cos(\omega_p t) + \hat{b}_p \sin(\omega_p t)$ and $\overline{y} = \frac{1}{N} \sum_{t=1}^{N} y_t$. Asymptotically, $F_p \sim F_{2,N-3}$.

Find all values of p that give significant Fourier frequencies at the 5% confidence level.

Hint: Use function `lm()` to fit the linear model. The output of this function can also be used to extract the value of the F-statistic, or compute it directly.

(e) Give the estimated coefficients for the linear model that results from using **all significant frequencies** found in part (d).

(f) Plot the data and the estimated model's fitted values on the same plot in `R`. Remember to properly label the axes, specify a legend and title for the plot.