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 3** at **9:00pm**.

- Submit your assignment online on `canvas.ubc.ca` in the **pdf format** through Gradescope (as a single file).
- Solutions to Questions 1 and 2 can be hand-written and scanned or typeset (e.g., using either LaTeX or R Markdown).
- Question 3 should be completed in **RStudio** and written up using **R Markdown**. Display the R code used to perform your data analysis.
- Please make sure your submission is clear and neat. The student is responsible for the submitted file being in good order (i.e., not corrupted).
- **Late submission penalty**: 1% per hour or fraction of an hour.

1. Consider the AR(2) process from Assignment 2 (Q1):

$$X_t = 0.1X_{t-1} + 0.2X_{t-2} + Z_t, \quad \{Z_t\}_{t\in\mathbb{Z}} \sim WN(0, \sigma^2),$$

and recall that its autocorrelation function has the form

$$\rho(h) = \frac{15}{36}(-0.4)^{|h|} + \frac{21}{36}0.5^{|h|}, \quad h \in \mathbb{Z}.$$

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

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

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

2. Given the power spectral density function

$$f(\omega) = \frac{1}{\pi}\Big(1.1 + 0.54\cos(\omega) - 0.2\cos(2\omega)\Big), \qquad \omega \in (0, \pi),$$

compute the autocovariance function $\gamma(k)$ of the underlying stochastic process for $k = 0, 1, 2$ and $k > 2$.

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

The data file `predator_prey.csv` contains population density of a type of unicellular algae and its predator, an animal called rotifer, from a chemostat trial. The daily rotifer data was recorded in animals per millilitre.

(a) Read the data into R and coerce the `rotifer` column into a time series object. Plot the resulting time series. Comment on what you observe. (Make sure to properly label the axes and provide the title for the plot.)

(b) Plot the raw periodogram for the series. Estimate the wavelength (in days) and angular frequency for the dominating frequency.

(c) Approximately how many long-term predator-prey cycles does this rotifer population experience over the course of the trial?

(d) Why would we want to modify the raw periodogram? Smooth the raw periodogram for the rotifer time series and comment on the behaviour of the resulting spectral density estimate.

(e) Build a function in `R` that generates the Fourier frequency ω_p for a given time series and given constant $p \in \{0, 1, \ldots, N/2\}$. 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 $p = 59$?

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

$$X_t = a_0 + a_p \cos(\omega_p t) + b_p \sin(\omega_p t) + \epsilon_t, \qquad t = 1, 2, \ldots, N,$$

where X_t the number of animals per milliliter at time index t and $\omega_p = \frac{2\pi p}{N}$ for each $p = 1, 2, \ldots, N/2$. Assume that $\epsilon_t \sim N(0, \sigma^2)$ for all t and are independent. 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

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

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 $\bar{y} = \frac{1}{N}\sum_{t=1}^{N} y_t$. Asymptotically, as $N \to \infty$,

$$F_p \sim F_{2,N-k},$$

where $F_{2,N-k}$ denotes the F-distribution with 2 and $N - k$ degrees of freedom.

Find all Fourier frequencies which are significant at the 99% confidence level.

Hint: use the 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.

(g) Give the estimated coefficients and write down the linear model that results from using the three frequencies with the smallest p-values found in part (f).

(h) Plot the rotifer data and the estimated model's fitted values on the same plot. Remember to properly label the axes, specify a legend and title for the plot. Comment on the model fit.