UNIVERSITY OF BRITISH COLUMBIA
Department of Statistics
STAT 443: Time Series and Forecasting
Assignment 2: Analysis in the Frequency Domain

1. A test allows us to check whether our data are from a white noise process, which has a flat spectrum, the periodogram of which should only vary due to random fluctuations. Given a series of length N let

$$M = \max\left\{n \in \mathbb{Z} : n \leq \frac{N}{2}\right\}$$

and define

$$P_k := \sum_{p=1}^{k} I\left(\omega_p\right)$$

for $k \leq M$. Moreover, let

$$C_k := \frac{P_k}{P_M}.$$

A plot of C_k against k/M is called the *cumulative periodogram.* If the data–generating process $X\left(t\right)$ is white noise, the cumulative spectrum should resemble a line of unit slope through the origin. The test statistic suggested is the maximum absolute horizontal distance between the observed cumulative spectrum and the line $y = x$. This defines the test statistic, D. Tables exist of critical values, but for practical purposes if testing at the 5% significance level values of D above

$$D_c := \frac{1.358}{\sqrt{M-1}}$$

would reject the null hypothesis that the series is a realization of white noise. The R commands `spectrum` (or `spec.pgram`) and `cumsum` are helpful in the application of the test.

   (a) Using the `rnorm` command in R generate 120 observations from the standard Normal distribution. Create the raw periodogram for your series, and provide a plot of the cumulative periodogram. Hence perform the above test to determine whether the observed series appears to be consistent with a realization from white noise, stating clearly your test statistic and conclusion. (2 marks)

(b) The dataset “star.dat” gives a record of the luminosity of a star, recorded over 600 consecutive nights. Create the raw periodogram for this series. Create the cumulative periodogram for this series. (2 marks)

(c) Perform the test to determine whether the series appears to be a realization from a white noise process. State clearly your test statistic, and the conclusion. (2 marks)

(d) Plot the acf up to lag 50 of the star luminosity data. From this plot, approximate what appears to be the wavelength of the main periodic component in the series. Use this to estimate what may be the important contributing frequency to the spectrum of the series. (2 marks)

(e) Compare your estimate in (d) with the largest component from the periodogram. (1 mark)

2. A seismograph is a device for measuring the progress of an earthquake. The machine involves a pendulum that will oscillate in response to ground movements due to an earthquake (see for example `www.britannica.com/science/seismograph` for details). A seismograph records a zigzag trace that shows how the ground shakes beneath the instrument as a result of the quake. Modern seismographs can greatly magnify these ground motions and respond to strong earthquakes from anywhere in the world.
   The file “Kobe.csv” contains seismograph readings (recording vertical acceleration in nm/sec^2) of the Kobe earthquake, recorded at Tasmania University, Australia on 16 January 1995. The measurements started at 20:56:51 (GMT) and continued for 51 minutes at 1 second intervals.

   (a) Read the data set into R, and coerce the data into a time series object. Plot the data. Plot the acf of the series up to lag 50, and comment on what you observe. (2 marks)

   (b) Plot the raw (i.e., without taking logarithms) periodogram for the time series. Comment on what you observe. (2 marks)

   (c) Use the raw periodogram to estimate the wavelength of the most important cyclical component in the time series. Comment on your findings in relation to the acf plot of the series. (2 marks)

(d) Recalling that the raw periodogram is not a consistent estimator of the spectral density, smooth the raw periodogram here. Clarifying your working, estimate the amplitude of the most important frequency contributing to the spectrum of the data. (3 marks)

(e) State clearly any reservations you may have regarding your estimate in (d). (2 marks)

BD