UNIVERSITY OF BRITISH COLUMBIA
Department of Statistics
STAT 443: Time Series and Forecasting
Assignment 2 Solution: 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

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

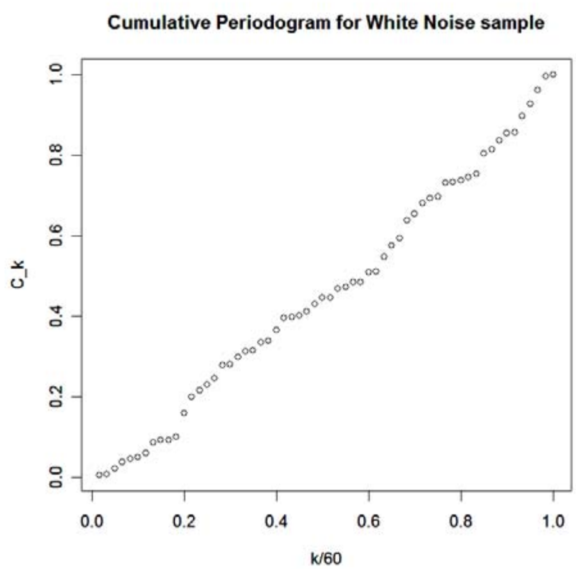
would reject the null hypothesis that the series is a realization of white noise. The R commands `spectrum` (or `spec.pgram`) and `cumsum` are helpful in the application of the test.

   (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)

We use something like

```
> x <- rnorm(120,0,1)
> specx1 <- spec.pgram(x, log="no")
> cumspecx1 <- cumsum(specx1$spec)
>  cumspecx1 <- cumsum(specx1$spec)/cumspecx1[60]
> k1 <- (1:60)/60
> plot(k1, cumspecx1, xlab="k/60", ylab="C_k", main="Cumulative
Periodogram for White Noise sample")
```

A plot looks something like



(1 mark) *To find the test statistic, we can use*

```
> max(cumspecx1-k1)
> min(cumspecx1-k1)
```

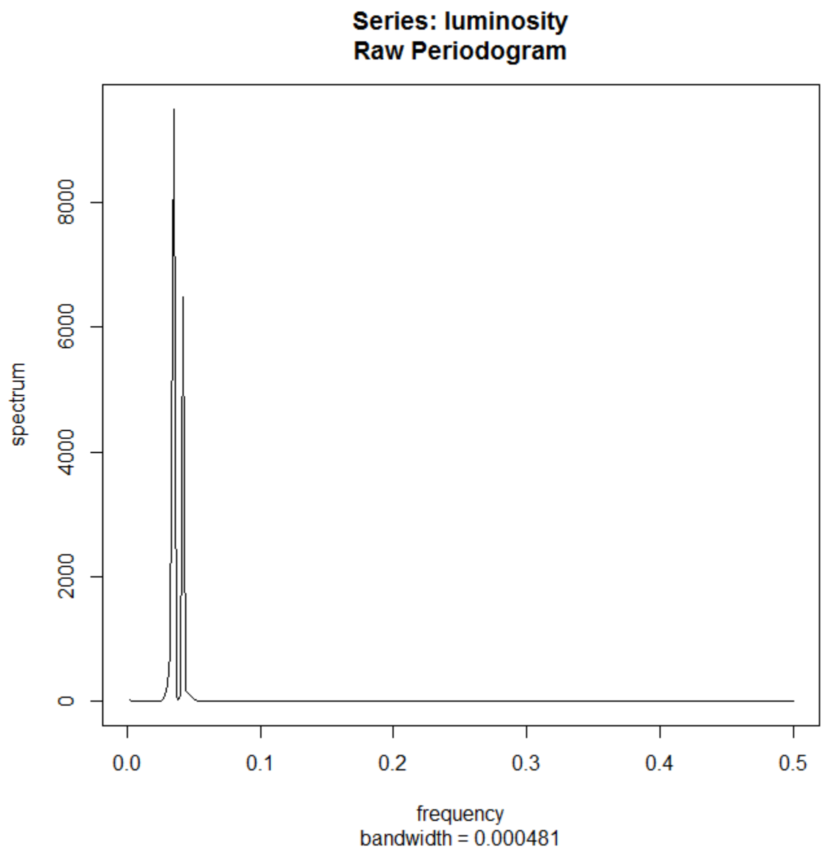
Here $D_c \approx 0.177$. So we compare our observed maximum difference with this, and do not reject (or less likely reject) that the data are from a white noise process. (1 mark)

Comment: Since the line $y = x$ is at an angle $\pi/4$, the vertical distances of the points C_k plotted from the line are the same as the horizontal differences. Rescaling would be required if we changed the scale of the horizontal axis in the cumulutive periodogram.

(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)

We can use

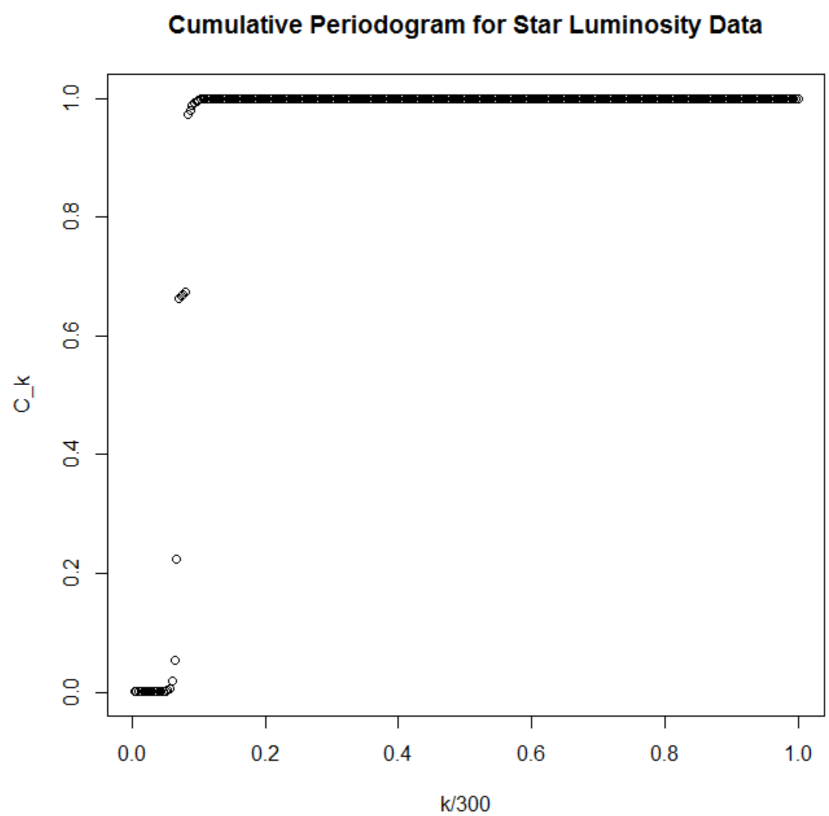
```
> luminosity <- as.ts(stardat$V1)
> spec.pgram(luminosity, log="no")
```



For the cumulative periodogram, we can use

```
> specx2 <- spec.pgram(luminosity, log="no")
> cumspecx2 <- cumsum(specx2$spec)
> cumspecx2 <- cumsum(specx2$spec)/cumspecx2[300]
> k2 <- (1:300)/300
> plot(k2, cumspecx2, xlab="k/300", ylab="C_k", main="Cumulative
```

```
Periodogram for Star Luminosity Data")
```



Evidently nearly all of the power in the spectrum is accounted for by frequencies in a range less than around $\pi/5$. Above values in that range there is little additional power in the spectrum. (2 marks)

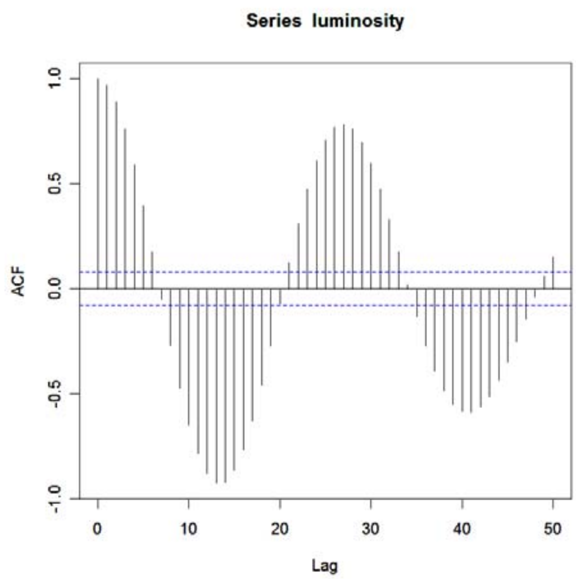
(c) Perform the test to determine whether the series appears to be a realization from a white noise process. State clearly your test statistic, and the conclusion. (2 marks)

The result is somewhat predictable: similar to (a)

```
> max(cumspecx2-k2)
> min(cumspecx2-k2)
```

This gives a test statistic value of 0.8989 (whether the log of the *periodogram was taken or not). Here though* $D_c = 0.078$, *so we reject the null hypothesis that the series is from a white noise process at the 5% level.* (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)
Plotting the acf



the above is consistent with a periodic series having wavelength around 28 days. (1 mark) *This wavelength implies an underlying frequency (in cycles per unit time) of*

$$\frac{2\pi}{28} = \frac{\pi}{14}$$

radians. (1 mark)

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

```
which.max(specx2$spec)
```

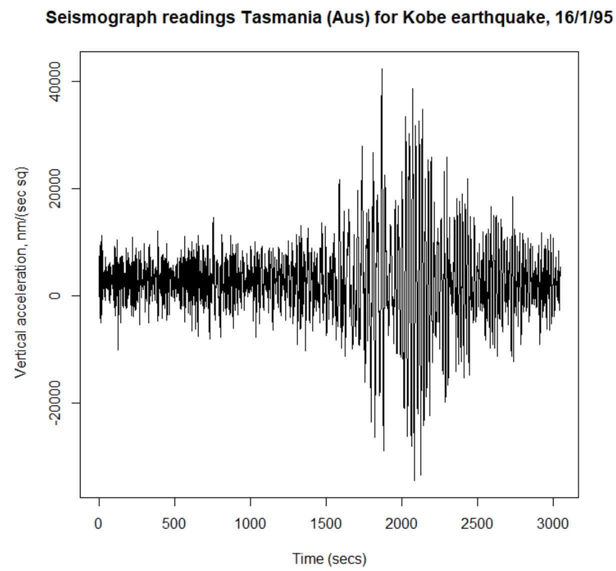
we find the largest value of the spectrum occurs at the 21st frequency, around 0.035. This gives a wavelength of 28.57 days. (1 mark)

Comment: There is a function `bartlettB.test` *in the R package* `hwwntest` *that would appear, based on its description, to apply the test for white noise as described above. However, values of the test statistic the function computes do not concur with the above. For the star luminosity data, the test statistic reported by* `bartlettB.test` *is 15.421, illogical since the maximum possible discrepancy in the test defined in this question is 1. It would appear that* `bartlettB.test` *scales the test statistic proposed here by a factor of approximately* $\sqrt{N/2}$ *and plots the cumuluative periodogram against the frequencies measured in cycles per unit time (which range from 0 to 0.5, hence it is departures from the line* $y = 2x$ *that are computed).*

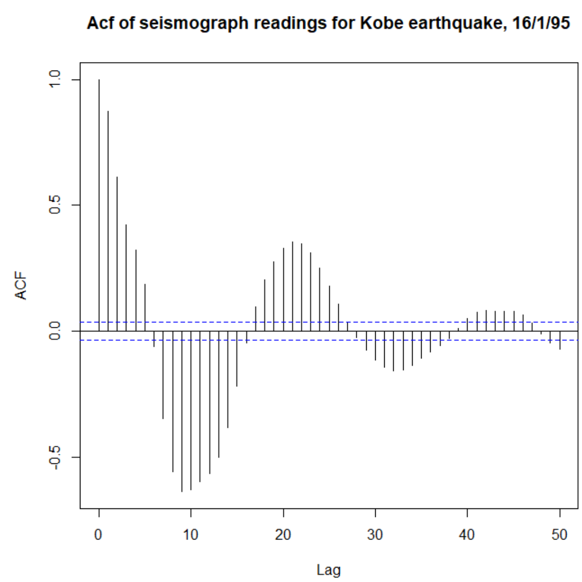
2. A seismograph is a device for measuring the progress of an earthquake. The machine involves a pendulum that will oscillate in response to ground movements due to an earthquake (see for example `www.britannica.com/science/seismograph` for details). A seismograph records a zigzag trace that shows how 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)

The times series plot is as follows:

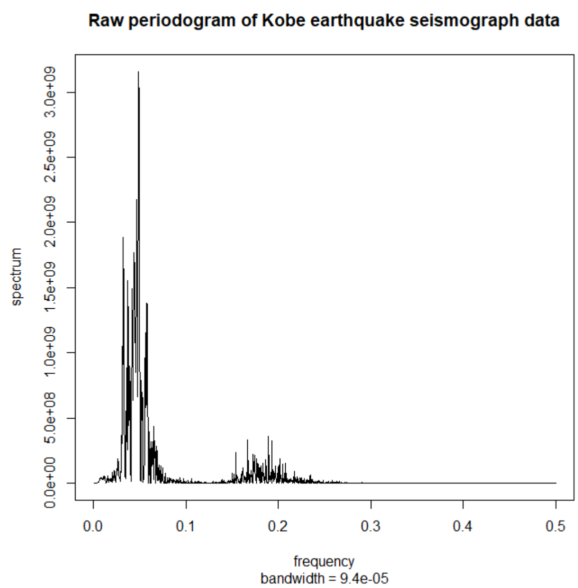


The acf up to lag 50 of the raw data is



The acf shows what appears to be a strong periodic component, likely of period about 21. (2 marks)

(b) Plot the raw (i.e., without taking logarithms) periodogram for the time series. Comment on what you observe. (2 marks)
Note that `spec.pgram` *de-trends by default, though probably not required here and makes little difference. The raw spectrum looks like the following:*



There is a significant frequency component somewhere about

$$\frac{\omega}{2\pi} \approx 0.05.$$

There is also evidence of other lesser contributing frequencies in the range 0.17 to 0.2. (2 marks: 1 for plot, 1 for comment).

(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)
Using
`which.max(prgramx$spec)`
we find the largest value of the spectrum occurs at the 150th frequency. Now R takes the frequencies as

$$\frac{\omega_p}{2\pi} = \frac{p}{N} = \frac{p}{3048}$$

in this case. Hence the 150th frequency is 150/3048=0.049213. This corresponds to a wavelength around 20.3 secs. This concurs with the acf plot, although it provides more accuracy than the acf plot alone. (2 marks: 1 for wavelength, 1 for comment. Note that R pads the data somewhat when computing the periodogram so the value of N adopted may differ slightly from 3048.)

(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)
We create a smoothed version via, for example,
```
smoothpgramx <- spec.pgram(x, spans=7)
```
or, if we wish to smooth twice, something like
```
smoothpgramx <- spec.pgram(x, spans=c(5,3))
```
The frequency at which the maximum value occurs will not change. Recall that the periodogram values from R are for $\pi I(\omega)$, and values can be found via, for example
```
smoothpgramx$spec[150]
```
the value of which will differ depending on the smoothing used. For the second case above the largest value is 2529995096, and so

$$\frac{N\hat{R}_{150}^2}{4} = \mathit{2529995096},$$

giving an estimate of the amplitude at the major frequency as

$$\sqrt{\frac{4 \times \mathit{2529995096}}{\mathit{3048}}} = 1822.1$$

(nm/sec^2). (3 marks: 1 for smoothing, –1 for dropping factor of π. Comment above about "N" also applies here.)

(e) State clearly any reservations you may have regarding your estimate in (d). (2 marks)
Firstly, the estimate depends on the smoothing adopted. More importantly perhaps the raw data show more vigourous seismic activity between around $t = 1500$ and $t = 2500$, suggesting there were two (or more) amplitudes, depending on the time. It may be more sensible to split off the data where the recordings are more

volatile and analyse them separately. (2 marks: 1 for each relevant comment.)

BD