STAT33600 Homework 6

Sarah Adilijiang

Section 4.1 Cyclical Behavior and Periodicity

4.1

(a)

First, we prove that for an positive integer n and any integer $0 \leq m \leq n$, since:

$$e^{\pm i2\pi m} = \cos(2\pi m) \pm i\sin(2\pi m) = 1$$

So, we have:

$$\sum_{t=1}^{n} e^{\pm i2\pi m/n\ t} = \frac{e^{\pm i2\pi m/n}(1 - e^{\pm i2\pi m/n\ n})}{1 - e^{\pm i2\pi m/n}} = \frac{e^{\pm i2\pi m/n}(1 - e^{\pm i2\pi m})}{1 - e^{\pm i2\pi m/n}} = 0$$

where $e^{\pm i2\pi m/n} \neq 1$, i.e. $m \neq 0, m \neq n$.

Consequently,

$$\cos(\pm 2\pi m/n\ t) = \cos(2\pi m/n\ t) = \frac{1}{2}(e^{i2\pi m/n\ t} + e^{-i2\pi m/n\ t}) = 0$$

$$\sin(\pm 2\pi m/n\ t) = \pm\sin(2\pi m/n\ t) = \pm\frac{1}{2i}(e^{i2\pi m/n\ t} - e^{-i2\pi m/n\ t}) = 0$$

Therefore, take $m = 2j$, so for $j \neq 0$ and $j \neq n/2$, we have:

Use $\cos(2\theta) = 2\cos^2(\theta) - 1 = 1 - 2\sin^2(\theta)$:

$$\sum_{t=1}^{n} \cos^2(2\pi tj/n) = \sum_{t=1}^{n} \frac{1}{2}(1 + \cos(4\pi tj/n)) = \frac{n}{2} + \frac{1}{2}\sum_{t=1}^{n} \cos(2\pi 2j/n\ t) = \frac{n}{2}$$

$$\sum_{t=1}^{n} \sin^2(2\pi tj/n) = \sum_{t=1}^{n} \frac{1}{2}(1 - \cos(4\pi tj/n)) = \frac{n}{2} - \frac{1}{2}\sum_{t=1}^{n} \cos(2\pi 2j/n\ t) = \frac{n}{2}$$

So:

$$\sum_{t=1}^{n} \cos^2(2\pi tj/n) = \sum_{t=1}^{n} \sin^2(2\pi tj/n) = \frac{n}{2}$$

(b)

When $j = 0$, we have:

$$\sum_{t=1}^{n} \cos^2(2\pi tj/n) = n, \quad \sum_{t=1}^{n} \sin^2(2\pi tj/n) = 0$$

When $j = n/2$, we have:

$$\sum_{t=1}^{n} \cos^2(2\pi tj/n) = \frac{n}{2} + \frac{1}{2}\sum_{t=1}^{n}\cos(4\pi tj/n) = \frac{n}{2} + \frac{1}{2}\sum_{t=1}^{n}\cos(2\pi t) = \frac{n}{2} + \frac{n}{2} = n$$

$$\sum_{t=1}^{n} \sin^2(2\pi tj/n) = \frac{n}{2} - \frac{1}{2}\sum_{t=1}^{n}\cos(4\pi tj/n) = \frac{n}{2} - \frac{1}{2}\sum_{t=1}^{n}\cos(2\pi t) = \frac{n}{2} - \frac{n}{2} = 0$$

In conclusion, when $j = 0$ or $j = n/2$:

$$\sum_{t=1}^{n} \cos^2(2\pi tj/n) = n, \qquad \sum_{t=1}^{n} \sin^2(2\pi tj/n) = 0$$

(c)

Since $j, k = 0, 1, ..., [[n/2]]$, where $[[n/2]]$ is $n/2$ when n is even and $(n-1)/2$ when n is odd, so $|j-k| = 0, ..., [[n/2]]$ and $j + k = 0, .., 2 \times [[n/2]]$

(1) For $j \neq k$

We have $|j-k| \neq 0$ or n, $j + k \neq 0$ or n, so we can set $m = |j-k|$ or $j + k$, thus:

Use $\cos\alpha\cos\beta = \frac{1}{2}[\cos(\alpha-\beta) + \cos(\alpha+\beta)]$:

$$\sum_{t=1}^{n} \cos(2\pi tj/n)\cos(2\pi tk/n) = \frac{1}{2}\sum_{t=1}^{n}\cos(2\pi(j-k)/n\ t) + \frac{1}{2}\sum_{t=1}^{n}\cos(2\pi(j+k)/n\ t) = 0 + 0 = 0$$

Use $\sin\alpha\sin\beta = \frac{1}{2}[\cos(\alpha-\beta) - \cos(\alpha+\beta)]$:

$$\sum_{t=1}^{n} \sin(2\pi tj/n)\sin(2\pi tk/n) = \frac{1}{2}\sum_{t=1}^{n}\cos(2\pi(j-k)/n\ t) - \frac{1}{2}\sum_{t=1}^{n}\cos(2\pi(j+k)/n\ t) = 0 - 0 = 0$$

(2) For any $j, k = 0, 1, ..., [[n/2]]$

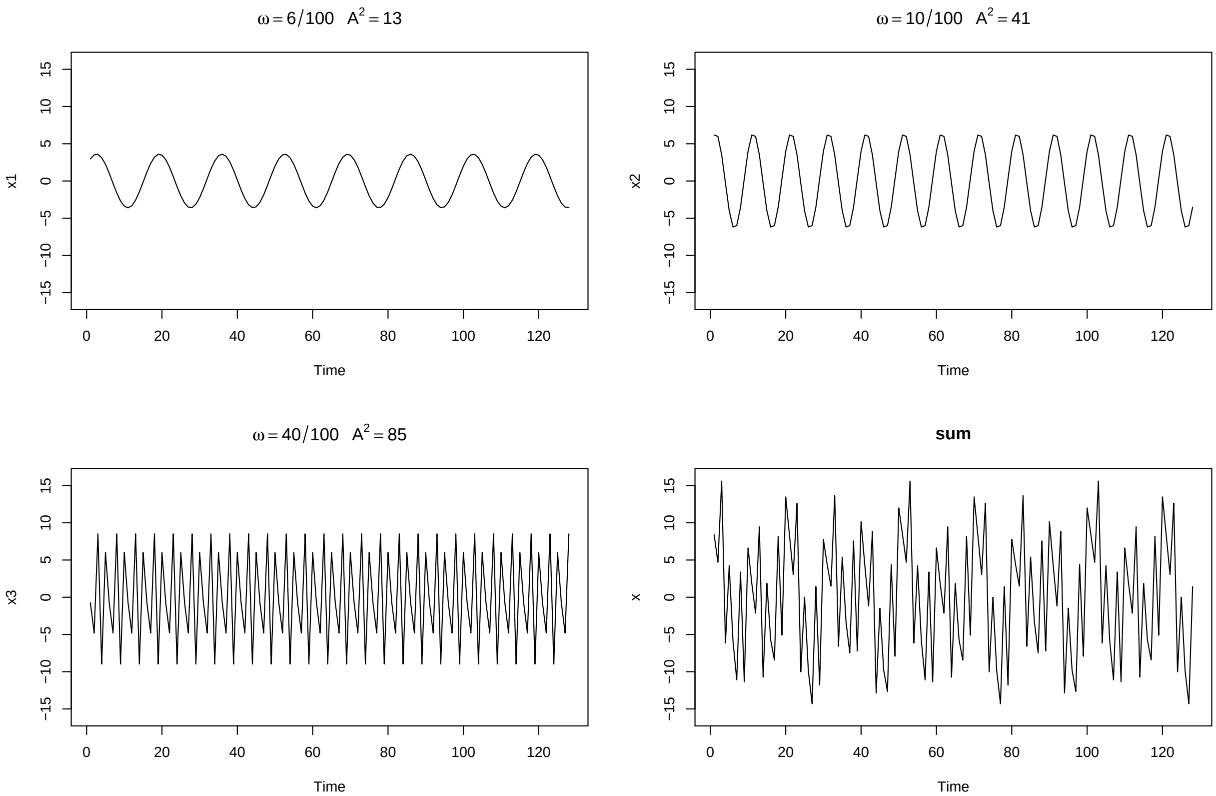
Use $\cos\alpha\sin\beta = \frac{1}{2}[\sin(\alpha+\beta) - \sin(\alpha-\beta)]$:

$$\sum_{t=1}^{n} \cos(2\pi tj/n)\sin(2\pi tk/n) = \frac{1}{2}\sum_{t=1}^{n}\sin(2\pi(j+k)/n\ t) - \frac{1}{2}\sum_{t=1}^{n}\sin(2\pi(j-k)/n\ t) = 0 - 0 = 0$$

4.2

(a)

```r
n = 128
x1 = 2*cos(2*pi*1:n*6/100) + 3*sin(2*pi*1:n*6/100)
x2 = 4*cos(2*pi*1:n*10/100) + 5*sin(2*pi*1:n*10/100)
x3 = 6*cos(2*pi*1:n*40/100) + 7*sin(2*pi*1:n*40/100)
x = x1 + x2 + x3
par(mfrow=c(2,2))
plot.ts(x1, ylim=c(-16,16), main=expression(omega==6/100~~~A^2==13))
plot.ts(x2, ylim=c(-16,16), main=expression(omega==10/100~~~A^2==41))
plot.ts(x3, ylim=c(-16,16), main=expression(omega==40/100~~~A^2==85))
plot.ts(x, ylim=c(-16,16), main="sum")
```



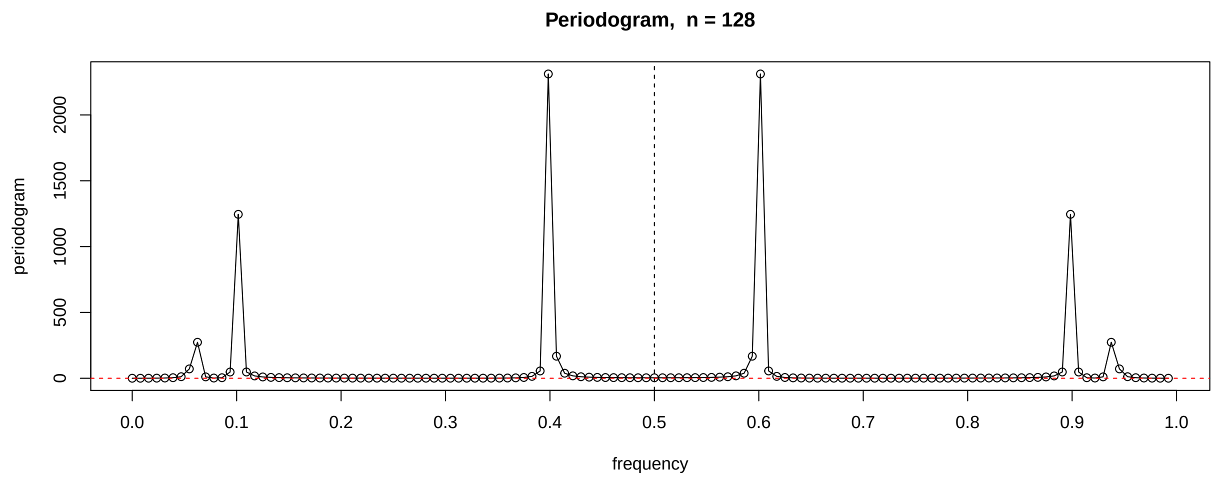
Comment:

The major difference between these seires where $n = 128$ and the series generated in Example 4.1 where $n = 100$ is that the Fourier or fundamental frequencies $w_j = \frac{j}{n}$ for $j = 0, 1, ..., n-1$ are different for different sample size n.

The three frequency components of the true data are $\frac{6}{100}, \frac{10}{100}, \frac{40}{100}$, which are among the fundamental frequencies $w_j = \frac{j}{100}$ for $j = 0, 1, ..., n-1$ in Example 4.1. However, the three frequency components are not among the fundamental frequencies $w_j = \frac{j}{128}$ for $j = 0, 1, ..., n-1$ any more here in this question.

(b)

```
per = 1/n * Mod(fft(x))^2
freq = 0:(n-1) / n
plot(freq, per, type="o", xlab="frequency", ylab="periodogram",
     main="Periodogram,  n = 128", xaxt="n")
axis(at=(0:10)/10, side=1)
abline(v = 0.5, lty=2)
abline(h = 0, lty=2, col=2)
```




```r
# periodogram
m = length(0:ceiling(n/2))
y = cbind(0:ceiling(n/2), per[1:m], freq[1:m], 1/freq[1:m])
colnames(y) = c('j','periodogram','frequency','time units/cycle')
y[order(y[,'periodogram'], decreasing=TRUE)[1:10], ]
```

```
##        j periodogram frequency time units/cycle
##  [1,] 51  2311.31728 0.3984375         2.509804
##  [2,] 13  1245.09478 0.1015625         9.846154
##  [3,]  8   273.22182 0.0625000        16.000000
##  [4,] 52   167.38303 0.4062500         2.461538
##  [5,]  7    71.06318 0.0546875        18.285714
##  [6,] 50    55.46593 0.3906250         2.560000
##  [7,] 14    47.62709 0.1093750         9.142857
##  [8,] 12    47.40848 0.0937500        10.666667
##  [9,] 53    38.28875 0.4140625         2.415094
## [10,] 54    18.29889 0.4218750         2.370370
```

```r
# non-zero values
sum(per>10^(-5)) / 2
```

```
## [1] 64
```

Comment:

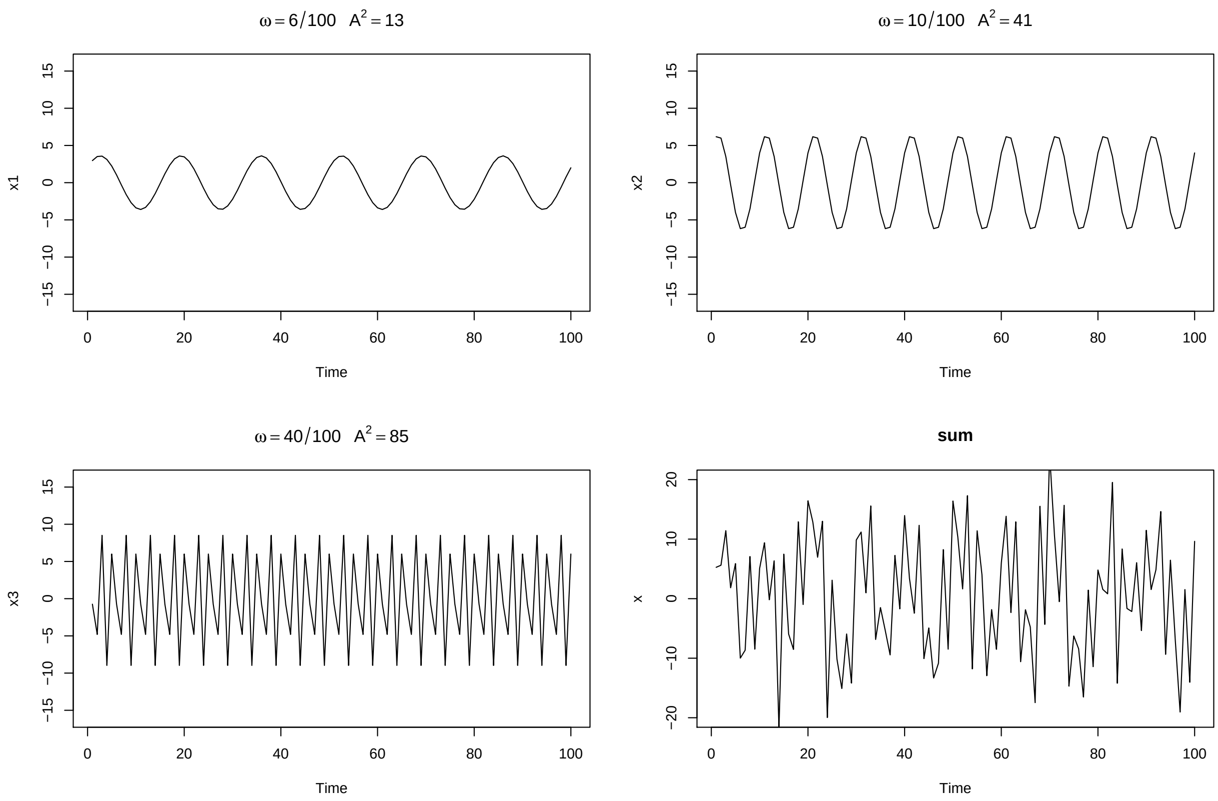
When $n = 100$ as in Example 4.1, since the three frequency components are among the fundamental frequencies, so the periodogram gets high peaks exactly at $\frac{6}{100}, \frac{10}{100}, \frac{40}{100}$, and equal to zeros at all the other frequencies.

However, in this question, the three frequency components are not among the fundamental frequencies, so the periodogram gets high peaks near but not exactly at $\frac{6}{100}, \frac{10}{100}, \frac{40}{100}$, and there are also many nonzero points at other frequencies.

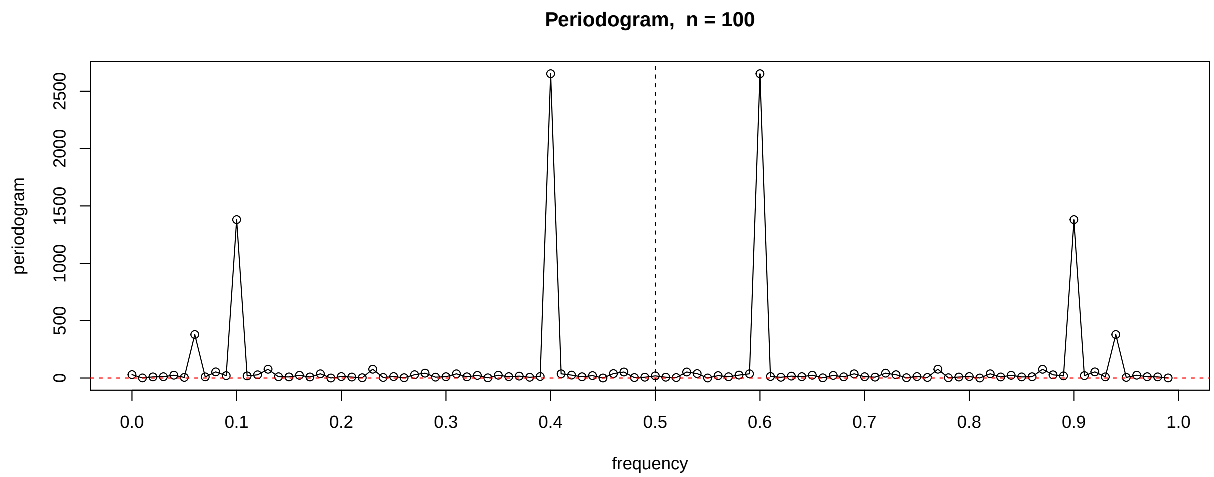
(c)

```r
set.seed(1)
n = 100
x1 = 2*cos(2*pi*1:n*6/100) + 3*sin(2*pi*1:n*6/100)
x2 = 4*cos(2*pi*1:n*10/100) + 5*sin(2*pi*1:n*10/100)
```

```
x3 = 6*cos(2*pi*1:n*40/100) + 7*sin(2*pi*1:n*40/100)
x = x1 + x2 + x3 + rnorm(n,0,5)
par(mfrow=c(2,2))
plot.ts(x1, ylim=c(-16,16), main=expression(omega==6/100~~~A^2==13))
plot.ts(x2, ylim=c(-16,16), main=expression(omega==10/100~~~A^2==41))
plot.ts(x3, ylim=c(-16,16), main=expression(omega==40/100~~~A^2==85))
plot.ts(x, ylim=c(-20,20), main="sum")
```




```
per = 1/n * Mod(fft(x))^2
freq = 0:(n-1) / n
plot(freq, per, type="o", xlab="frequency", ylab="periodogram",
     main="Periodogram,  n = 100", xaxt="n")
axis(at=(0:10)/10, side=1)
abline(v = 0.5, lty=2)
abline(h = 0, lty=2, col=2)
```




```r
# periodogram
m = length(0:ceiling(n/2))
y = cbind(0:ceiling(n/2), per[1:m], freq[1:m], 1/freq[1:m])
colnames(y) = c('j','periodogram','frequency','time units/cycle')
y[order(y[,'periodogram'], decreasing=TRUE)[1:10], ]
```

```
##        j periodogram frequency time units/cycle
##  [1,] 40  2652.49021      0.40         2.500000
##  [2,] 10  1380.57143      0.10        10.000000
##  [3,]  6   379.12403      0.06        16.666667
##  [4,] 23    76.84801      0.23         4.347826
##  [5,] 13    76.52126      0.13         7.692308
##  [6,]  8    53.57770      0.08        12.500000
##  [7,] 47    52.52757      0.47         2.127660
##  [8,] 28    42.64237      0.28         3.571429
##  [9,] 46    38.59361      0.46         2.173913
## [10,] 41    36.79365      0.41         2.439024
```

```r
# non-zero values
sum(per>10^(-5)) / 2
```

```
## [1] 50
```

Comment:

In Example 4.1, the three frequency components are among the fundamental frequencies and there are no noise term, so the periodogram gets high peaks exactly at $\frac{6}{100}, \frac{10}{100}, \frac{40}{100}$, and equal to zeros at all the other frequencies.

However, in this question, although the three frequency components are still among the fundamental frequencies, the added noise term make the periodogram get high peaks still exactly at $\frac{6}{100}, \frac{10}{100}, \frac{40}{100}$ but with many nonzero points at other frequencies due to the noise term.

4.3

(a)

Set $X = A^2 = Z_1^2 + Z_2^2,\ Y = \phi = \tan^{-1}(Z_2/Z_1)$, since Z_1, Z_2 are independent standard normal variables, so:

$$J = \left|\frac{\partial(z_1,z_2)}{\partial(x,y)}\right| = \frac{1}{\left|\frac{\partial(x,y)}{\partial(z_1,z_2)}\right|} = \frac{1}{\begin{vmatrix}\frac{\partial x}{\partial z_1} & \frac{\partial x}{\partial z_2} \\ \frac{\partial y}{\partial z_1} & \frac{\partial y}{\partial z_2}\end{vmatrix}} = \frac{1}{\begin{vmatrix} 2z_1 & 2z_2 \\ \frac{-z_2}{z_1^2+z_2^2} & \frac{z_1}{z_1^2+z_2^2}\end{vmatrix}} = \frac{1}{\frac{2z_1^2}{z_1^2+z_2^2}+\frac{2z_2^2}{z_1^2+z_2^2}} = \frac{1}{2}$$

$$\Rightarrow \qquad f_{X,Y}(x,y) = f_{Z_1,Z_2}(z_1,z_2)\ |J| = \frac{1}{2}\ f_{Z_1}(z_1)f_{Z_2}(z_2)$$

$$= \frac{1}{2}\ \frac{1}{\sqrt{2\pi}}\exp\{-\frac{z_1^2}{2}\}\frac{1}{\sqrt{2\pi}}\exp\{-\frac{z_2^2}{2}\} = \frac{1}{4\pi}\exp\{-\frac{z_1^2+z_2^2}{2}\} = \frac{1}{4\pi}\exp\{-\frac{x}{2}\}$$

where $0 \le x < \infty$ and $-\pi < y < \pi$.

So the marginal densities of x and y are:

$$f_X(x) = \int_{-\pi}^{\pi} f_{X,Y}(x,y)dy = \int_{-\pi}^{\pi}\frac{1}{4\pi}\exp\{-\frac{x}{2}\}\ dy = \frac{1}{2}\exp\{-\frac{x}{2}\} \qquad (0 \le x < \infty)$$

which is the density function of a chi-squared distribution: χ_2^2 (df =2).

$$f_Y(y) = \int_{0}^{\infty} f_{X,Y}(x,y)dx = \int_{0}^{\infty}\frac{1}{4\pi}\exp\{-\frac{x}{2}\}\ dx = \frac{1}{2\pi} \qquad (-\pi < y < \pi)$$

which is the density function of a uniform distribution: $Unif(-\pi,\pi)$.

And X, Y are independent since:

$$f_{X,Y}(x,y) = f_X(x)\times f_Y(y)$$

(b)

Set $X = A^2$ $(A = \sqrt{X})$, $Y = \phi$, since X, Y are independent, X is chi-squared with 2 df, and Y is uniformly districution on $(-\pi,\pi)$, so:

$$J = \left|\frac{\partial(x,y)}{\partial(z_1,z_2)}\right| = \frac{1}{1/2} = 2$$

$$\Rightarrow \qquad f_{Z_1,Z_2}(z_1,z_2) = f_{X,Y}(x,y)\ |J| = 2\ f_X(x)f_Y(y)$$

$$= 2\times\frac{1}{2}\exp\{-\frac{x}{2}\}\times\frac{1}{2\pi} = \frac{1}{2\pi}\exp\{-\frac{x}{2}\} = \frac{1}{2\pi}\exp\{-\frac{z_1^2+z_2^2}{2}\}$$

where $-\infty < z_1 < \infty$ and $-\infty < z_2 < \infty$.

So the marginal densities of z_1 and z_2 are:

$$f_{Z_1}(z_1) = \int_{-\infty}^{\infty} f_{Z_1,Z_2}(z_1,z_2)dz_2 = \int_{-\infty}^{\infty}\frac{1}{2\pi}\exp\{-\frac{z_1^2+z_2^2}{2}\}\ dz_2 = \frac{1}{\sqrt{2\pi}}\exp\{-\frac{z_1^2}{2}\} \qquad (-\infty < z_1 < \infty)$$

which is the density function of a standard normal distribution.

$$f_{Z_2}(z_2) = \int_{-\infty}^{\infty} f_{Z_1,Z_2}(z_1,z_2)dz_1 = \int_{-\infty}^{\infty}\frac{1}{2\pi}\exp\{-\frac{z_1^2+z_2^2}{2}\}\ dz_1 = \frac{1}{\sqrt{2\pi}}\exp\{-\frac{z_2^2}{2}\} \qquad (-\infty < z_2 < \infty)$$

which is the density function of a standard normal distribution.

And Z_1, Z_2 are independent since:

$$f_{Z_1,Z_2}(z_1,z_2) = f_{Z_1}(z_1)\times f_{Z_2}(z_2)$$

Section 4.2 The Spectral Density

4.5

(a)

Since $w_t \sim \text{iid} N(0,1)$, we have $E(w_t) = 0$, and:

$$\gamma_w(h) = Cov(w_{t+h}, w_t) = \begin{cases} 1 & h = 0 \\ 0 & h \neq 0 \end{cases}$$

Therefore, the mean function of w_t is a constant that does not depend on time t, and the autocovariance function only depends on lag h, so w_t is stationary.

For $x_t = w_t - \theta w_{t-1}$, we have $E(x_t) = 0$, and:

$$\gamma_x(h) = Cov(x_{t+h}, x_t) = Cov(w_{t+h} - \theta w_{t+h-1}, w_t - \theta w_{t-1})$$

$$= (\theta^2 + 1)\gamma_w(h) - \theta\ (\gamma_w(h-1) + \gamma_w(h+1)) = \begin{cases} \theta^2 + 1 & h = 0 \\ -\theta & |h| = 1 \\ 0 & |h| \geq 2 \end{cases}$$

Therefore, the mean function of x_t is a constant that does not depend on time t, and the autocovariance function only depends on lag h, so x_t is also stationary.

(b)

For series w_t, the spectral density is:

$$f_w(\omega) = \sum_{h=-\infty}^{\infty} \gamma_w(h) e^{-2\pi i \omega h} = \sigma_w^2 = 1$$

Hence for series $x_t = w_t - \theta w_{t-1}$, the spectral density (power spectrum) is:

$$f_x(\omega) = \sum_{h=-\infty}^{\infty} \gamma_x(h) e^{-2\pi i \omega h} = (\theta^2 + 1) - \theta(e^{-2\pi i \omega} + e^{2\pi i \omega}) = \theta^2 + 1 - 2\theta \cos(2\pi\omega)$$

4.6

(a)

For causal stationary series $x_t = \phi x_{t-1} + w_t,\ |\phi| < 1$, i.e. $(1 - \phi B)x_t = w_t$, where $\phi(z) = 1 - \phi z,\ \theta(z) = 1$. Then, the spectral density (power spectrum) is:

$$f_x(\omega) = \sigma_w^2\ \left|\frac{\theta(e^{-2\pi i \omega})}{\phi(e^{-2\pi i \omega})}\right|^2 = \sigma_w^2\ \frac{1}{|1 - \phi e^{-2\pi i \omega}|^2} = \frac{\sigma_w^2}{|1 - \phi(\cos(2\pi\omega) - i \sin(2\pi\omega))|^2}$$

$$= \frac{\sigma_w^2}{|1 - \phi \cos(2\pi\omega) + i\phi \sin(2\pi\omega))|^2} = \frac{\sigma_w^2}{(1 - \phi \cos(2\pi\omega))^2 + \phi^2 \sin^2(2\pi\omega)} = \frac{\sigma_w^2}{1 + \phi^2 - 2\phi \cos(2\pi\omega)}$$

(b)

If the autocovariance function is:

$$\gamma_x(h) = \frac{\sigma_w^2 \ \phi^{|h|}}{1-\phi^2}$$

Note that for any θ, $|e^{i\theta}| = 1$. Since $|\phi| < 1$, so $|\phi e^{i\theta}| < 1$.

Then the inverse transform of $\gamma_x(h)$ is:

$$g_x(\omega) = \sum_{h=-\infty}^{\infty} \gamma_x(h) e^{-2\pi i\omega h} = \sum_{h=-\infty}^{\infty} \frac{\sigma_w^2 \ \phi^{|h|}}{1-\phi^2} e^{-2\pi i\omega h} = \frac{\sigma_w^2}{1-\phi^2} \left(\sum_{h=-\infty}^{0} \phi^{-h} e^{-2\pi i\omega h} + \sum_{h=1}^{\infty} \phi^{h} e^{-2\pi i\omega h} \right)$$

$$= \frac{\sigma_w^2}{1-\phi^2} \left(\sum_{h=0}^{\infty} (\phi e^{2\pi i\omega})^h + \sum_{h=1}^{\infty} (\phi e^{-2\pi i\omega})^h \right) = \frac{\sigma_w^2}{1-\phi^2} \left(\frac{1}{1-\phi e^{2\pi i\omega}} + \frac{\phi e^{-2\pi i\omega}}{1-\phi e^{-2\pi i\omega}} \right)$$

$$= \frac{\sigma_w^2}{1-\phi^2} \times \frac{1-\phi^2}{1+\phi^2-2\phi\cos(2\pi\omega)} = \frac{\sigma_w^2}{1+\phi^2-2\phi\cos(2\pi\omega)} = f_x(\omega)$$

This spectral density is the same as in question (a). Due to the uniqueness of the Fourier transform pairs between the autocovariance function and the spectral density, we can conclude that the autocovariance function is indeed:

$$\gamma_x(h) = \frac{\sigma_w^2 \ \phi^{|h|}}{1-\phi^2}$$

4.7

The series $x_t = s_t + As_{t-D} + n_t$ is stationary, because s_t, n_t are independent and stationary with mean zero. So the autocovariance function of x_t is:

$$\begin{aligned}
\gamma_x(h) &= Cov(x_{t+h}, x_t) = Cov(s_{t+h} + As_{t+h-D} + n_{t+h},\ s_t + As_{t-D} + n_t) \\
&= (A^2+1)\gamma_s(h) + A\gamma_s(h-D) + A\gamma_s(h+D) + \gamma_n(h) \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left\{ \left[A^2 + 1 + Ae^{-2\pi i\omega D} + Ae^{2\pi i\omega D}\right] f_s(\omega) + f_n(\omega) \right\} e^{2\pi i\omega h} d\omega \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left\{ \left[A^2 + 1 + 2A\cos(2\pi\omega D)\right] f_s(\omega) + f_n(\omega) \right\} e^{2\pi i\omega h} d\omega = \int_{-\frac{1}{2}}^{\frac{1}{2}} f_x(\omega) e^{2\pi i\omega h} d\omega
\end{aligned}$$

Due to the uniqueness of the Fourier transform, we have that:

$$f_x(\omega) = \left[A^2 + 1 + 2A\cos(2\pi\omega D)\right] f_s(\omega) + f_n(\omega)$$

4.8

Because x_t, y_t are independent and stationary with zero-mean. So the mean of z_t is: $E(z_t) = E(x_t y_t) = E(x_t)E(y_t) = 0$.

So the autocovariance function of z_t is:

$$\begin{aligned}\gamma_z(h) &= Cov(z_{t+h}, z_t) = E(z_{t+h}z_t) - E(z_{t+h})E(z_t) = E(z_{t+h}z_t)\\ &= E(x_{t+h}y_{t+h}x_t y_t) = E(x_{t+h}x_t)E(y_{t+h}y_t) = \gamma_x(h)\gamma_y(h)\end{aligned}$$

Therefore, the spectral density function is:

$$f_z(\omega) = \sum_{h=-\infty}^{\infty} \gamma_z(h)e^{-2\pi i\omega h} = \sum_{h=-\infty}^{\infty} \gamma_x(h)\gamma_y(h)e^{-2\pi i\omega h} = \sum_{h=-\infty}^{\infty} \left(\gamma_x(h)e^{-2\pi i\omega h}\int_{-\frac{1}{2}}^{\frac{1}{2}} f_y(\nu)e^{2\pi i\nu h}d\nu\right)$$

$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\sum_{h=-\infty}^{\infty} \gamma_x(h)e^{-2\pi i\omega h}e^{2\pi i\nu h}\right) f_y(\nu)d\nu = \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\sum_{h=-\infty}^{\infty} \gamma_x(h)e^{-2\pi i(\omega-\nu)h}\right) f_y(\nu)d\nu$$

$$= \int_{-\frac{1}{2}}^{\frac{1}{2}} f_x(\omega-\nu)f_y(\nu)d\nu$$

Section 4.3 Periodogram and Discrete Fourier Transform

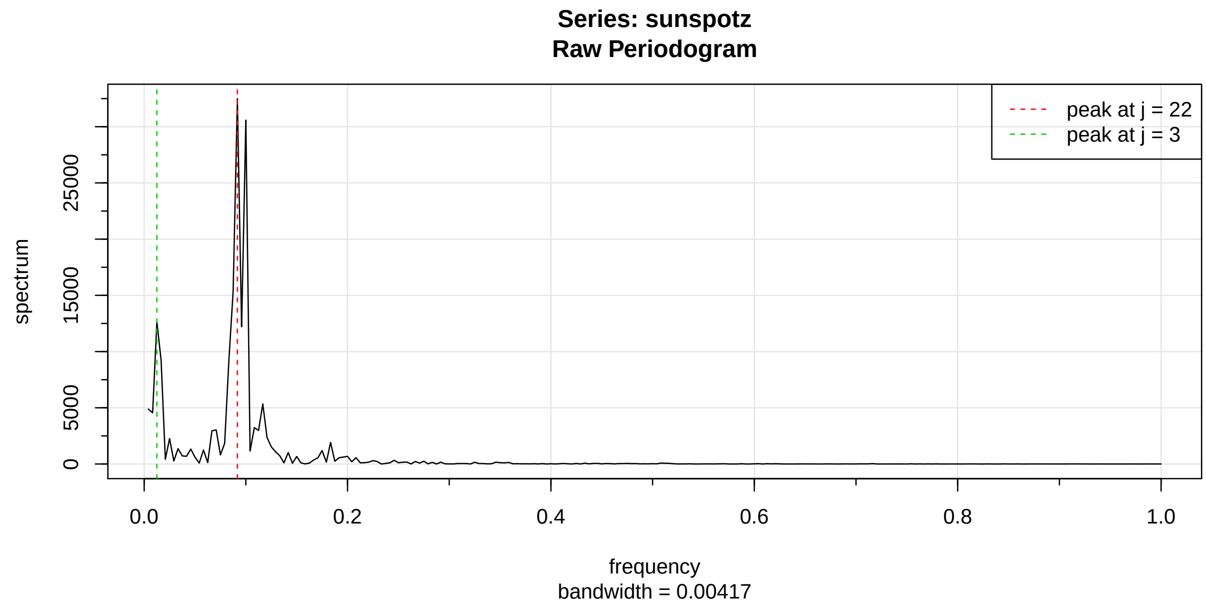
4.9

```r
library(astsa)
per = mvspec(sunspotz, log="no")

# periodogram
n = per$n.used    # or:  n = nextn(length(sunspotz))
m = length(per$freq)
y = cbind(1:m, per$spec, per$freq, 1/per$freq)
colnames(y) = c('j','periodogram','frequency(cycle/year)','year/cycle')
y[order(y[,'periodogram'], decreasing=TRUE)[1:10], ]
```

```
##        j periodogram frequency(cycle/year) year/cycle
##  [1,] 22   32477.873           0.091666667  10.909091
##  [2,] 24   30583.289           0.100000000  10.000000
##  [3,] 21   15346.750           0.087500000  11.428571
##  [4,]  3   12645.799           0.012500000  80.000000
##  [5,] 23   12205.532           0.095833333  10.434783
##  [6,] 20    9265.959           0.083333333  12.000000
##  [7,]  4    9226.914           0.016666667  60.000000
##  [8,] 28    5340.559           0.116666667   8.571429
##  [9,]  1    4883.863           0.004166667 240.000000
## [10,]  2    4551.155           0.008333333 120.000000
```

```r
# add line to plot
time_unit = 0.5
abline(v=22/n/time_unit, lty=2, col=2)
abline(v=3/n/time_unit, lty=2, col=3)
```

```r
legend("topright", legend=c("peak at j = 22","peak at j = 3"),
       lty=2, col=c(2,3))
```



The above figure shows the periodogram of the series, where the frequency axis is labeled in multiples of time unit $\triangle = 0.5$ year. Note that the centered data have been padded to a series of lenghth 480 here while the original data is of lenghth 459.

The two main peaks are at:

(1) $w_j = j/n \div \triangle = 22/480 \div 0.5 = 22/240$ cycles/year, which means about one cycle every $10.909091 \approx 11$ years.

(2) $w_j = j/n \div \triangle = 3/480 \div 0.5 = 3/240$ cycles/year, which means one cycle every 80 years.

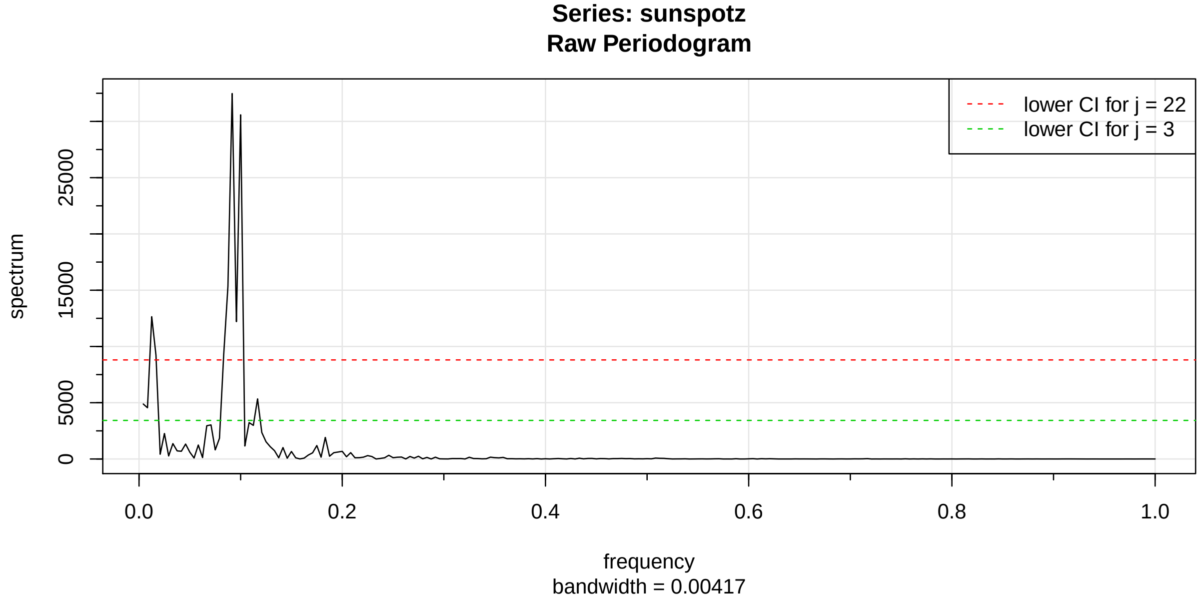
```r
# 95% confidence intervals
chi_right = qchisq(0.975,df=2,lower.tail=TRUE)
chi_left = qchisq(0.025,df=2,lower.tail=TRUE)

# peak j=22
CI_22 = c(2*per$spec[22]/chi_right, 2*per$spec[22]/chi_left); CI_22
```

```
## [1]    8804.265 1282807.445
```

```r
# peak j=3
CI_3 = c(2*per$spec[3]/chi_right, 2*per$spec[3]/chi_left); CI_3
```

```
## [1]   3428.087 499482.385
```

```r
# plot CI in periodogram figure
per = mvspec(sunspotz, log="no")
abline(h=CI_22[1], lty=2, col=2)
abline(h=CI_3[1], lty=2, col=3)
legend("topright",legend=c("lower CI for j = 22","lower CI for j = 3"),
       lty=2, col=c(2,3))
```



Comment:

The 95% confidence interval for peak at $j = 22$ (11-year cycle) is [8804.265, 1282807.445]. And the 95% confidence interval for peak at $j = 3$ (80-year cycle) is [3428.087, 499482.385].

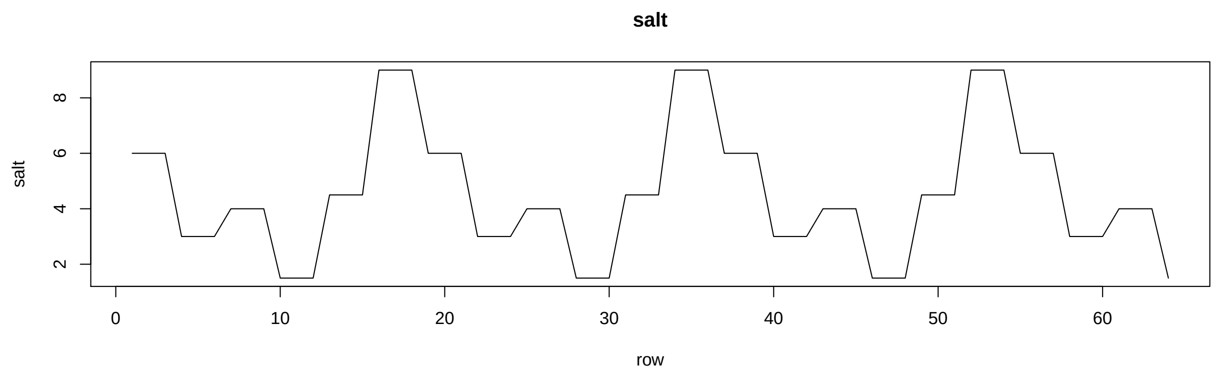
We can see that both the confidence intervals are wide.

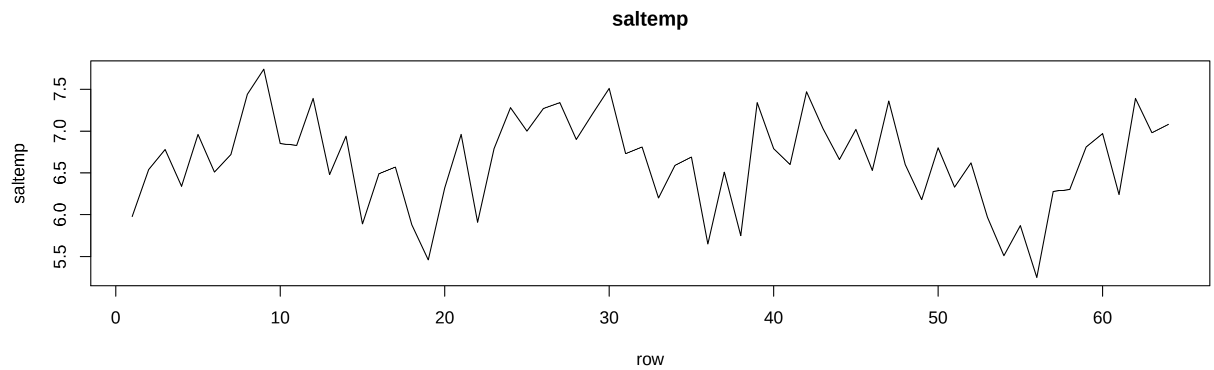
For peak at $j = 22$ (11-year cycle), the lower value of confidence interval is higher than any other periodogram ordinate except for the second peak at $j = 3$, so it is safe to say that this value is significant.

For peak at $j = 3$ (80-year cycle), the lower value of confidence interval is higher than most of the other periodogram ordinate, so it is also fine to say that this value is roughly significant.

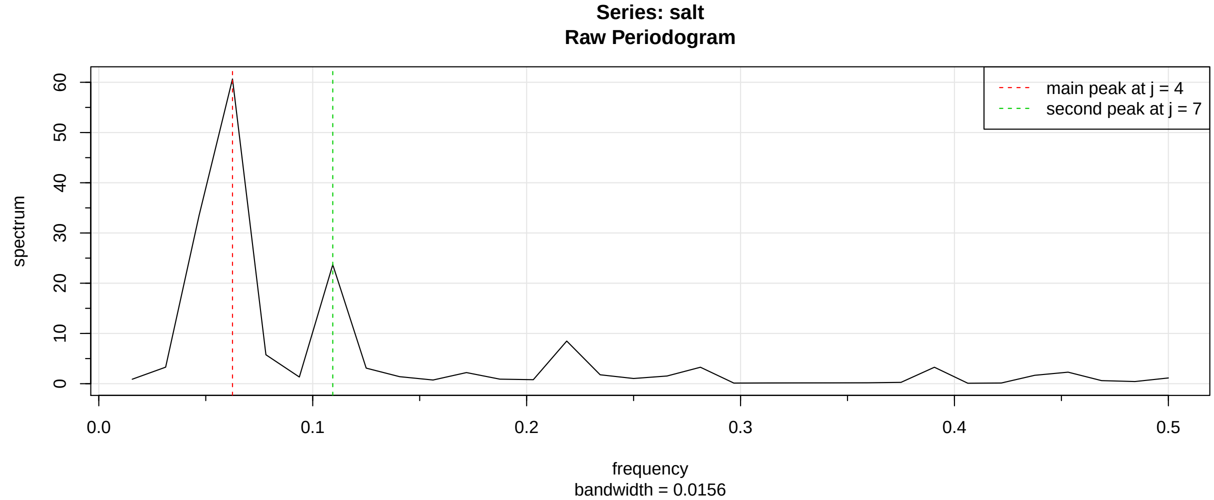
4.10

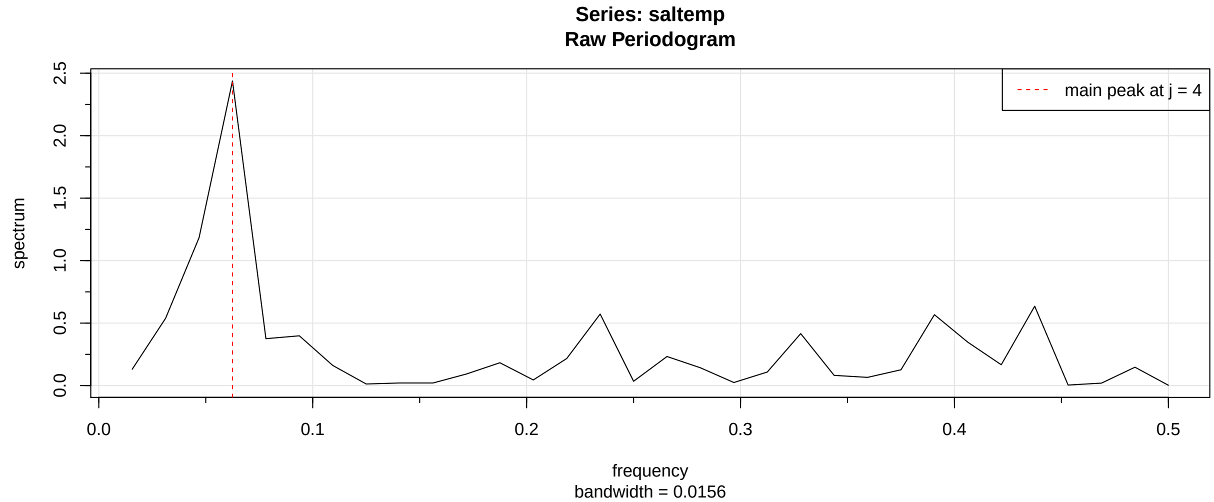
```
par(mfrow=c(2,1))
plot(salt, xlab="row", main="salt")
plot(saltemp, xlab="row", main="saltemp")
```






```r
par(mfrow=c(2,1))

salt.per = mvspec(salt, log="no")
n = salt.per$n.used    # or:  n = nextn(length(salt))
time_unit = 1
abline(v=4/n/time_unit, lty=2, col=2)
abline(v=7/n/time_unit, lty=2, col=3)
legend("topright", legend= c("main peak at j = 4","second peak at j = 7"), lty=2, col=c(2,3))

temp.per = mvspec(saltemp, log="no")
abline(v=4/n/time_unit, lty=2, col=2)  # same n & time_unit as in salt
legend("topright", legend= "main peak at j = 4", lty=2, col=2)
```






```r
# periodogram: salt
m = length(salt.per$freq)
y = cbind(1:m, salt.per$spec, salt.per$freq, 1/salt.per$freq)
colnames(y) = c('j','periodogram','frequency(cycle/row)','row/cycle')
y[order(y[,'periodogram'], decreasing=TRUE)[1:10], ]
```

```
##        j periodogram frequency(cycle/row) row/cycle
##  [1,]  4   60.666480             0.062500 16.000000
##  [2,]  3   33.485893             0.046875 21.333333
##  [3,]  7   23.690288             0.109375  9.142857
##  [4,] 14    8.486221             0.218750  4.571429
##  [5,]  5    5.759764             0.078125 12.800000
##  [6,]  2    3.288731             0.031250 32.000000
##  [7,] 25    3.274281             0.390625  2.560000
##  [8,] 18    3.273662             0.281250  3.555556
##  [9,]  8    3.098553             0.125000  8.000000
## [10,] 29    2.292958             0.453125  2.206897
```

```r
# periodogram: saltemp
m = length(temp.per$freq)
y = cbind(1:m, temp.per$spec, temp.per$freq, 1/temp.per$freq)
```

```
colnames(y) = c('j','periodogram','frequency(cycle/row)','row/cycle')
y[order(y[,'periodogram'], decreasing=TRUE)[1:10], ]
```

```
##        j periodogram frequency(cycle/row) row/cycle
##  [1,]  4   2.4378701             0.062500 16.000000
##  [2,]  3   1.1835676             0.046875 21.333333
##  [3,] 28   0.6353287             0.437500  2.285714
##  [4,] 15   0.5724505             0.234375  4.266667
##  [5,] 25   0.5671983             0.390625  2.560000
##  [6,]  2   0.5390863             0.031250 32.000000
##  [7,] 21   0.4158753             0.328125  3.047619
##  [8,]  6   0.3983407             0.093750 10.666667
##  [9,]  5   0.3753379             0.078125 12.800000
## [10,] 26   0.3470017             0.406250  2.461538
```

The above figure shows the periodogram of the series, where the frequency axis is labeled in multiples of space unit $\triangle = 1$ row. Note that the length of both series is 64, which is already a factor of 2.

(1) For series **salt**, there seems to be two main peaks: one is at $w_j = j/n \div \triangle = 4/64 \div 1 = 1/16$ cycles/row, which means one cycle every 16 rows; the other is at $w_j = j/n \div \triangle = 7/64 \div 1 = 7/64$ cycles/row, which means about one cycle every $9.142857 \approx 9$ rows

(2) For series **saltemp**, there seems to be only one main peak, which is also at $w_j = j/n \div \triangle = 4/64 \div 1 = 1/16$ cycles/row, which means one cycle every 16 rows.

```
# 95% confidence intervals
chi_right = qchisq(0.975,df=2,lower.tail=TRUE)
chi_left = qchisq(0.025,df=2,lower.tail=TRUE)

### (1) salt
# peak j=4
salt.CI_4 = c(2*salt.per$spec[4]/chi_right, 2*salt.per$spec[4]/chi_left); salt.CI_4
```

```
## [1]   16.44577 2396.19797
```

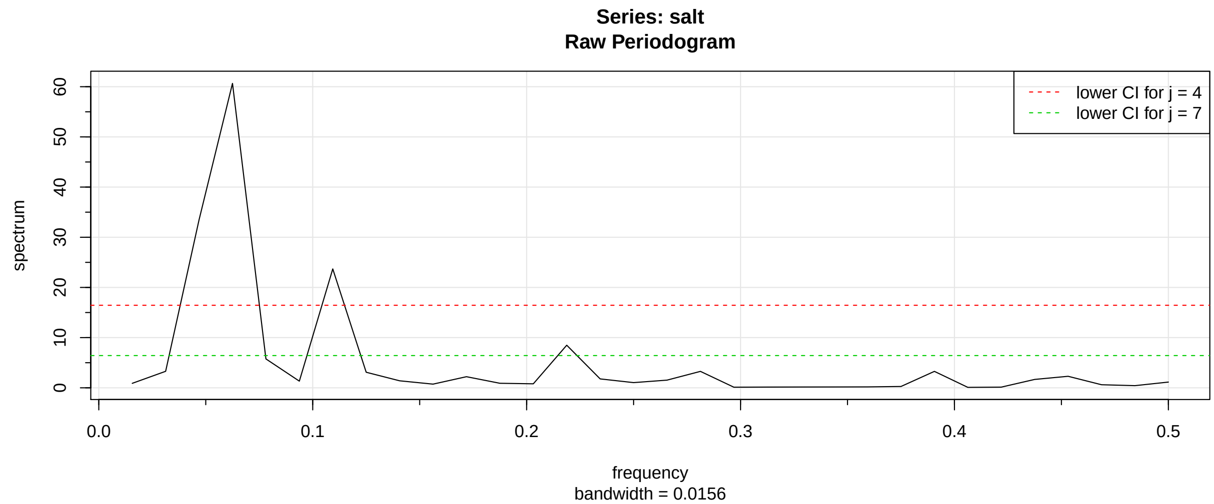
```
# peak j=7
salt.CI_7 = c(2*salt.per$spec[7]/chi_right, 2*salt.per$spec[7]/chi_left); salt.CI_7
```

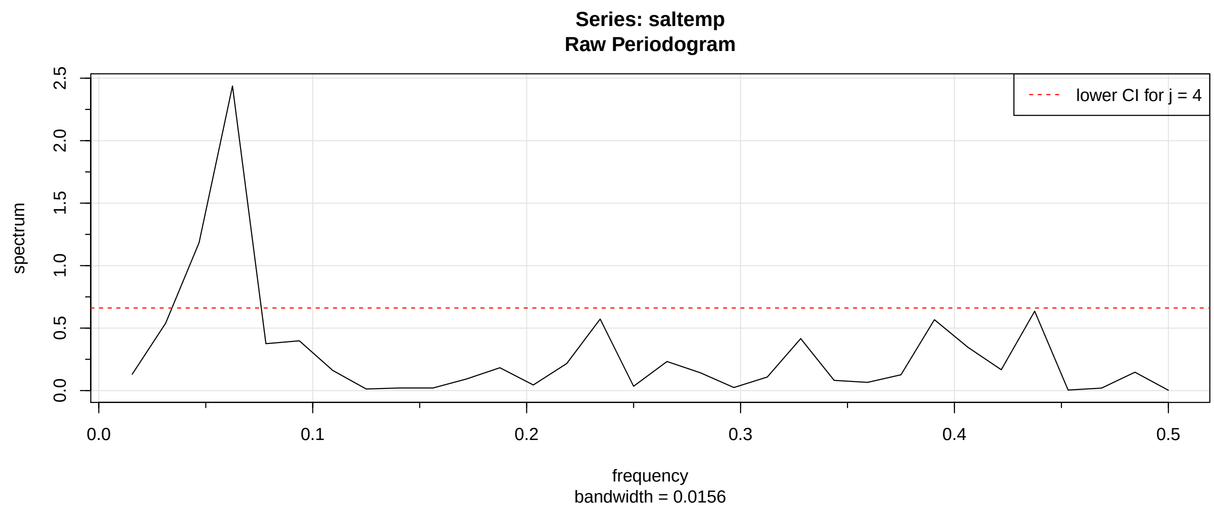
```
## [1]   6.422082 935.716379
```

```
### (2) saltemp
# peak j=4
temp.CI_4 = c(2*temp.per$spec[4]/chi_right, 2*temp.per$spec[4]/chi_left); temp.CI_4
```

```
## [1]  0.6608701 96.2907243
```

```
# plot CI in periodogram figure
par(mfrow=c(2,1))

salt.per = mvspec(salt, log="no")
abline(h=salt.CI_4[1], lty=2, col=2)
abline(h=salt.CI_7[1], lty=2, col=3)
legend("topright", legend= c("lower CI for j = 4","lower CI for j = 7"), lty=2, col=c(2,3))

temp.per = mvspec(saltemp, log="no")
abline(h=temp.CI_4[1], lty=2, col=2)
legend("topright", legend= "lower CI for j = 4", lty=2, col=2)
```





Comment:

(1) For series **salt**:

The 95% confidence interval for peak at $j = 4$ (16-row cycle) is [16.44577, 2396.19797]. And the 95% confidence interval for peak at $j = 3$ (9-row cycle) is [6.422082, 935.716379].

We can see that both the confidence intervals are wide.

For peak at $j = 4$ (16-row cycle), the lower value of confidence interval is higher than any other periodogram ordinate except for the second peak at $j = 7$, so it is safe to say that this value is significant.

For peak at $j = 7$ (9-row cycle), the lower value of confidence interval is higher than most of the other periodogram ordinate, so it is okay to say that this value is roughly significant.

However, comparing with the series saltemp, since the series salt has a correlation with the series saltemp, we should consider their common dominant frequency, which is the peak at $j = 4$ (16-row cycle). Therefore, we can also only consider this significant major peak for series salt.

(2) For series **saltemp**:

The 95% confidence interval for peak at $j = 4$ (16-row cycle) is [0.6608701, 96.2907243]. This confidence intervals is also wide.

The lower value of confidence interval is higher than any other periodogram ordinates, so it is safe to say that this value is significant.

4.12

$$d_x(\omega_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t e^{-2\pi i \omega_k t} \qquad \Longleftrightarrow \qquad x_t = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} d_x(\omega_k) e^{2\pi i \omega_k t}$$

$$d_A(\omega_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} a_t e^{-2\pi i \omega_k t} \qquad \Longleftrightarrow \qquad a_t = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} d_A(\omega_k) e^{2\pi i \omega_k t}$$

Since $x_t = x_{t+n}$ is periodic, so we have:

$$\sum_{s=1}^{n} a_s x_{t-s} = \sum_{s=1}^{n} a_s \left(\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} d_x(\omega_k) e^{2\pi i \omega_k (t-s)} \right) = \sum_{k=0}^{n-1} d_x(\omega_k) e^{2\pi i \omega_k t} \left(\frac{1}{\sqrt{n}} \sum_{s=1}^{n} a_s e^{-2\pi i \omega_k s} \right)$$

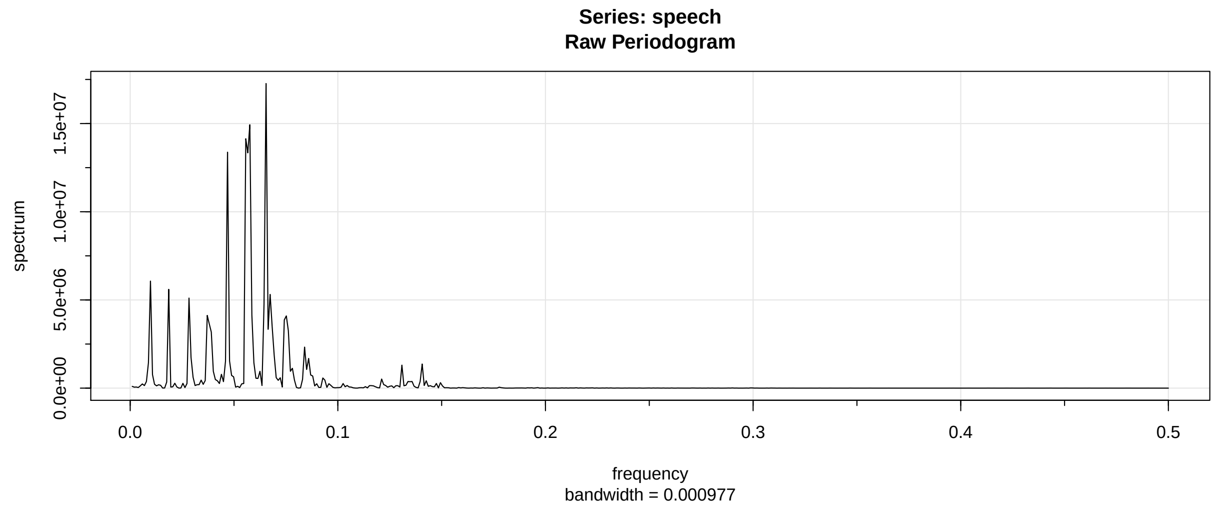
$$= \sum_{k=0}^{n-1} d_x(\omega_k)\ d_A(\omega_k)\ e^{2\pi i \omega_k t}$$

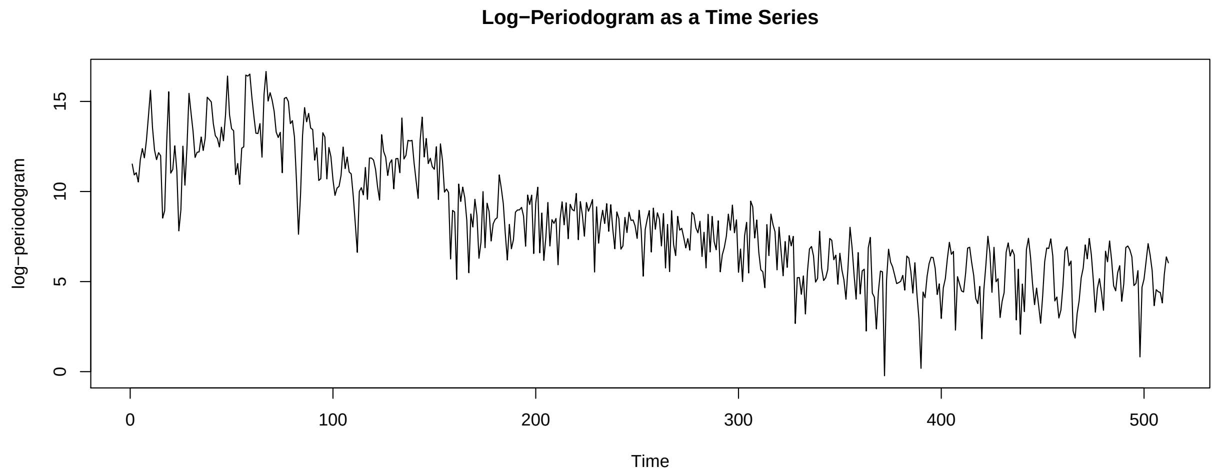
Section 4.4 Nonparametric Spectral Estimation

4.16

(a)

```
par(mfrow=c(2,1))

# raw periodogram
per = mvspec(speech, log="no")

# calculate and plot log-periodogram
x = log(per$spec)
plot.ts(x, ylab="log-periodogram", main="Log-Periodogram as a Time Series")
```





Comment:

From the plot we can see that the log-periodogram of the data does look like a periodic time series as predicted.

Note that the speech data have been padded to a series of lenghth 1024 here while the original data is of lenghth 1020.

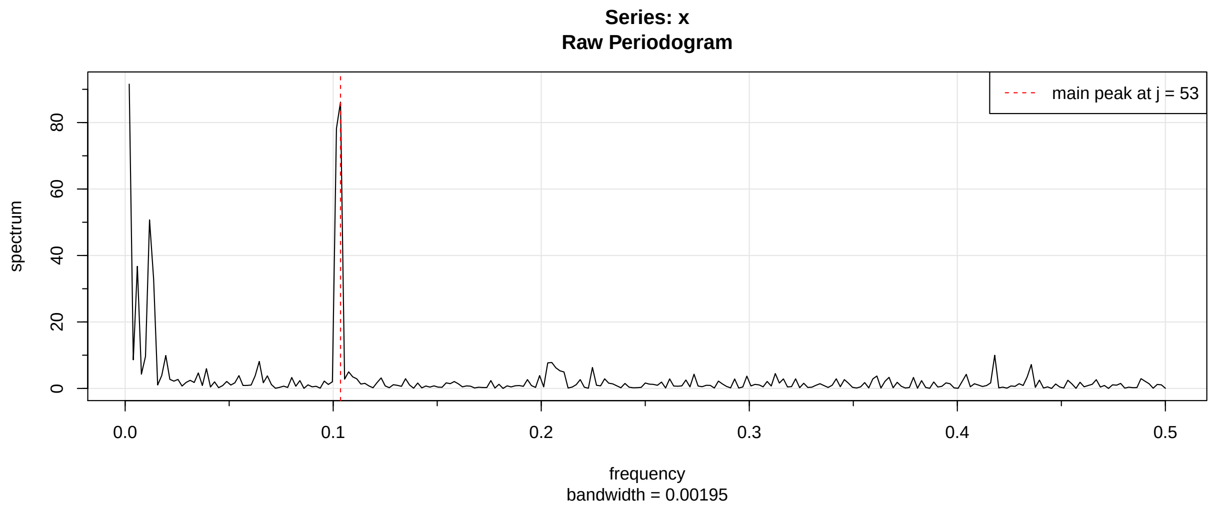
(b)

```r
x.per = mvspec(x, log="no")

# periodogram
n = x.per$n.used  # 512
m = length(x.per$freq)
y = cbind(1:m, x.per$spec, x.per$freq, 1/x.per$freq)
colnames(y) = c('j','periodogram','frequency','1/frequency')
y[order(y[,'periodogram'], decreasing=TRUE)[1:10], ]
```

```
##          j periodogram   frequency 1/frequency
##  [1,]    1   91.564025 0.001953125  512.000000
##  [2,]   53   86.163030 0.103515625    9.660377
```

```
##  [3,]  52   78.232924 0.101562500    9.846154
##  [4,]   6   50.708294 0.011718750   85.333333
##  [5,]   3   36.739175 0.005859375  170.666667
##  [6,]   7   33.026117 0.013671875   73.142857
##  [7,] 214    9.983166 0.417968750    2.392523
##  [8,]  10    9.898654 0.019531250   51.200000
##  [9,]   5    9.609871 0.009765625  102.400000
## [10,]   2    8.621042 0.003906250  256.000000
```

```r
# add line to plot
time_unit = 1
abline(v=53/n/time_unit, lty=2, col=2)
legend("topright", legend="main peak at j = 53", lty=2, col=2)
```



In this question the time series is the log-periodogram calculated in question (a). Note that the log-periodogram data is detrened by the *mvspec*() function be default, and the detrended log-periodogram data is of lenghth 512.

The main peak is at $j/n = 53/512 \approx 0.103515625$.

Since the log-periodogram is derived from:

$$f_x(\omega_j) = [1 + A^2 + 2A\cos(2\pi\omega_j D)]f_s(\omega_j)$$
$$x_j = \log f_x(\omega_j) = \log[1 + A^2 + 2A\cos(2\pi\omega_j D)] + \log f_s(\omega_j)$$
$$x_j = \log f_x(\omega_j) = \log[1 + A^2 + 2A\cos(2\pi j\frac{D}{n})] + \log f_s(\omega_j)$$

So x_j has a periodic part $\cos(2\pi j\frac{D}{n})$ with a period of $\frac{D}{n}$, where $n = 1024$ is the length of the padded origianl speech data.

Therefore, we have:

$$\frac{D}{n} = \frac{53}{512} \approx 0.103515625 \qquad \Rightarrow \qquad \hat{D} = \frac{53}{512} \times n = \frac{53}{512} \times 1024 = 106$$

As a result, the estimated delay $\hat{D} = 106$ points, i.e. 0.0106 seconds (the speech data is sampled at 10,000 points per second).

In Example 1.27 the ACF of the speech signal shows repeating peaks spaced at about 106-109 points, i.e. the pitch period (the distance between the repeating signals) is between 0.0106 and 0.0109 seconds. This agrees with the result estimated in this question.