SPECTRAL ANALYSIS

The periodogram

The **periodogram** $I(\omega)$ of a time series $y_1,\dots,y_n$ is for a Fourier frequency ω_k with $k\in\{1,2,\dots,[\frac{n-1}{2}]\}$ defined by

$$\begin{aligned}
I(\omega_k) &= \tfrac{n}{8\pi}\hat{R}_k^2 \\
&= \tfrac{n}{8\pi}(\hat{A}_k^2+\hat{B}_k^2) \\
&= \tfrac{n}{8\pi}\Big(\big(\tfrac{2}{n}\sum_{t=1}^{n} y_t\cos(\omega_k t)\big)^2+\big(\tfrac{2}{n}\sum_{t=1}^{n} y_t\sin(\omega_k t)\big)^2\Big) \\
&= \tfrac{1}{2\pi n}\Big(\big(\sum_{t=1}^{n} y_t\cos(\omega_k t)\big)^2+\big(\sum_{t=1}^{n} y_t\sin(\omega_k t)\big)^2\Big).
\end{aligned}$$

Apart from an unimportant scaling factor, $I(\omega_k)$ is just the squared estimate of the amplitude of a sinusoid with frequency ω_k, hence its size indicates how important that particular frequency is.

Defining the periodogram for an arbitrary frequency ω by

$$I(\omega)=\tfrac{1}{2\pi n}\Big(\big(\sum_{t=1}^{n} y_t\cos(\omega t)\big)^2+\big(\sum_{t=1}^{n} y_t\sin(\omega t)\big)^2\Big),$$

it can be written in complex form as

$$\begin{aligned}
I(\omega) &= \frac{1}{2\pi n}\left|\sum_{t=1}^{n} y_t\cos(\omega t)+i\sum_{t=1}^{n} y_t\sin(\omega t)\right|^2 \\
&= \frac{1}{2\pi n}\left|\sum_{t=1}^{n} y_t(\cos(\omega t)+i\sin(\omega t))\right|^2 \\
&= \frac{1}{2\pi n}\left|\sum_{t=1}^{n} y_t e^{i\omega t}\right|^2.
\end{aligned}$$

1P

The sequence

$$\sum_{t=1}^{n} y_t \, e^{-i\omega_k t}, \; k=0,\ldots,n-1,$$

is called the **discrete Fourier transform (DFT)** of the sequence $y_1,\ldots,y_n$.

The R function **fft** uses a fast algorithm (the **fast Fourier transform**) to calculate the DFT.

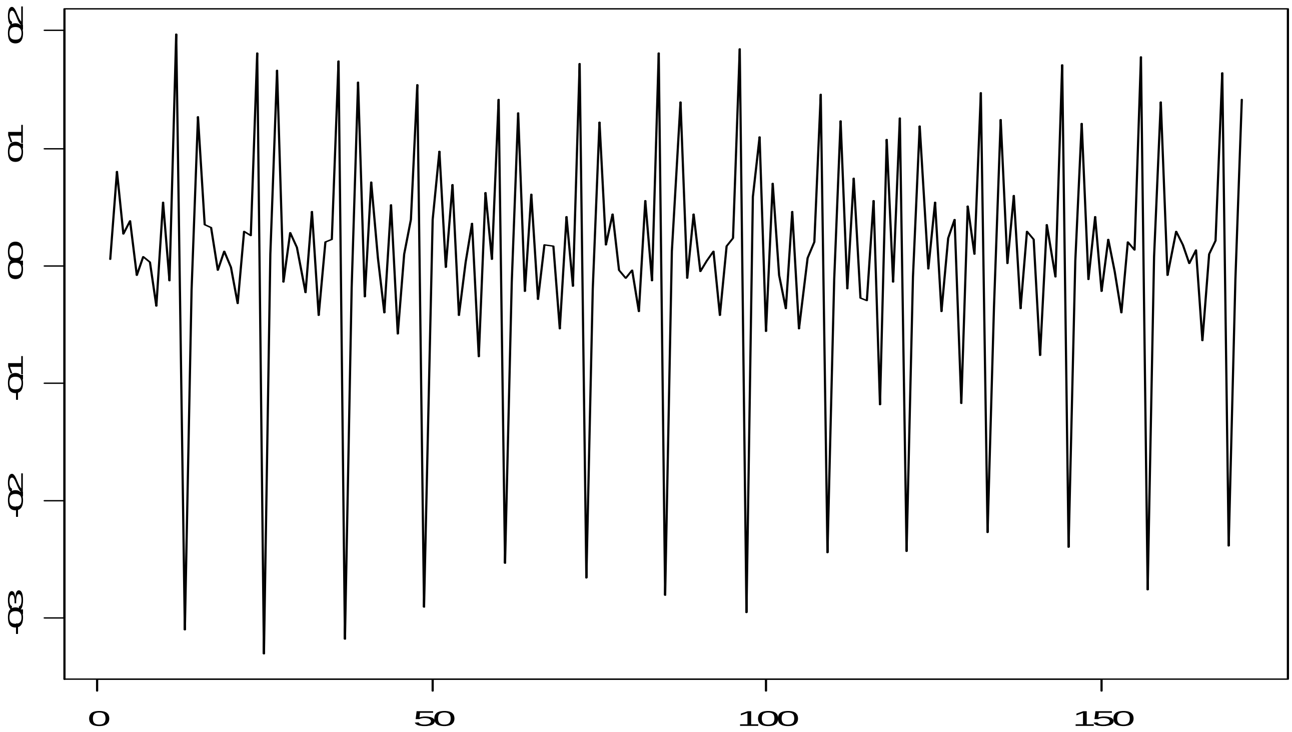
Exercise: Show that $I(\omega_k) \neq \frac{n}{8\pi} \hat{R}_k^2$ if $k=\frac{n}{2}$. 1N

Exercise: Show that it is sufficient to consider $I(\omega)$ on the interval $0 \leq \omega \leq \pi$ by proving that $I(\omega)$ is periodic with period 2π and symmetric about the y-axis. 1E

Using the periodogram to detect seasonal patterns

Exercise: Revisit the not seasonally adjusted **Retail and Food Services Sales**. Plot the differenced log series.

```
y.diff <- diff(y.log); plot(y.diff,xlab="",ylab="")
```

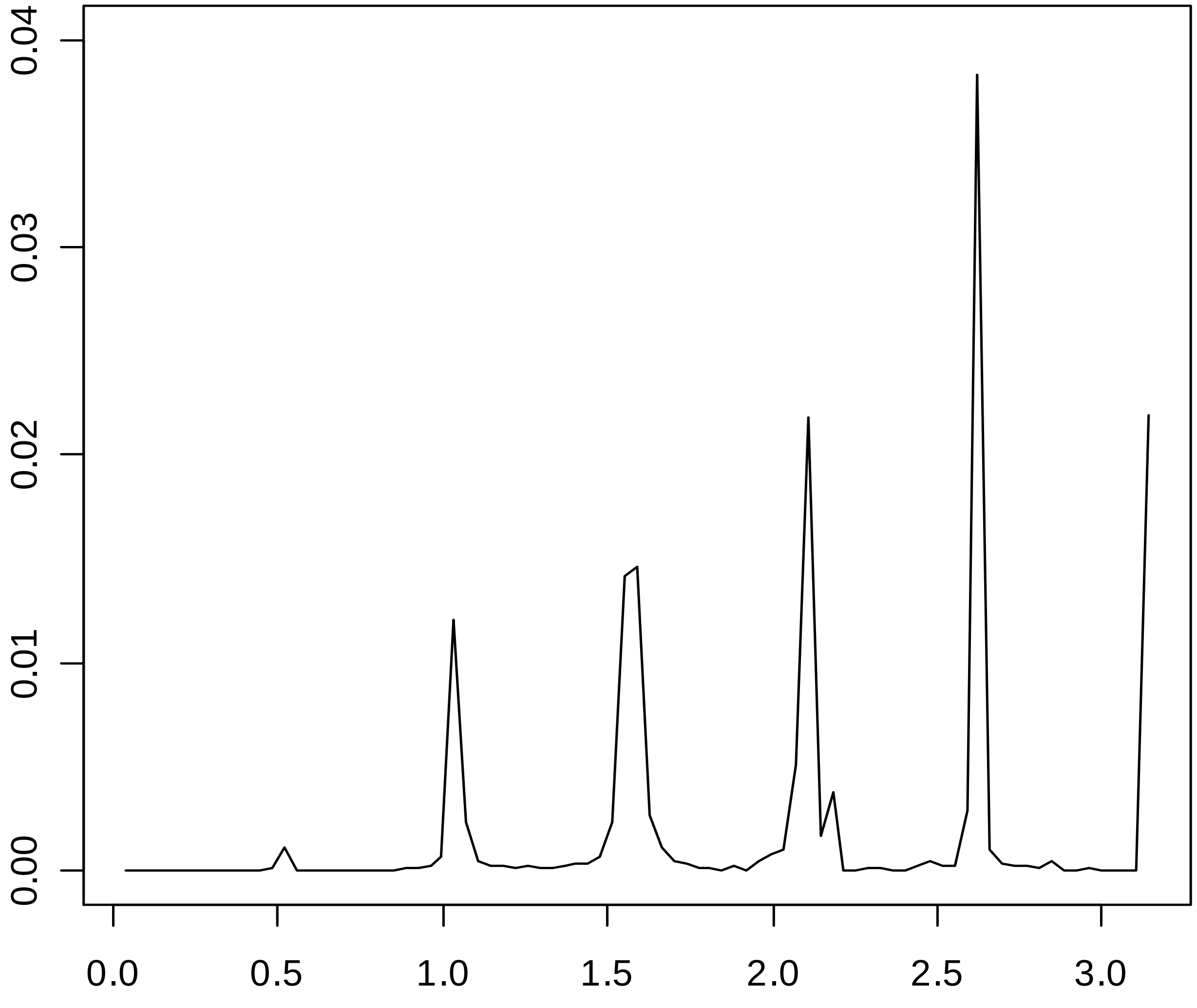


Exercise: Calculate the discrete Fourier transform of **y.diff** at the Fourier frequencies ω_k, k=1,…,[n/2].

```
m <- floor(n/2)  # number of frequencies
        # floor = the largest integer not greater than
y.dft <- fft(y.diff)  # use the fast Fourier transform
y.dft <- y.dft[2:(m+1)]  # excl. k=0, k=m+1, m+2, …,n-1
```

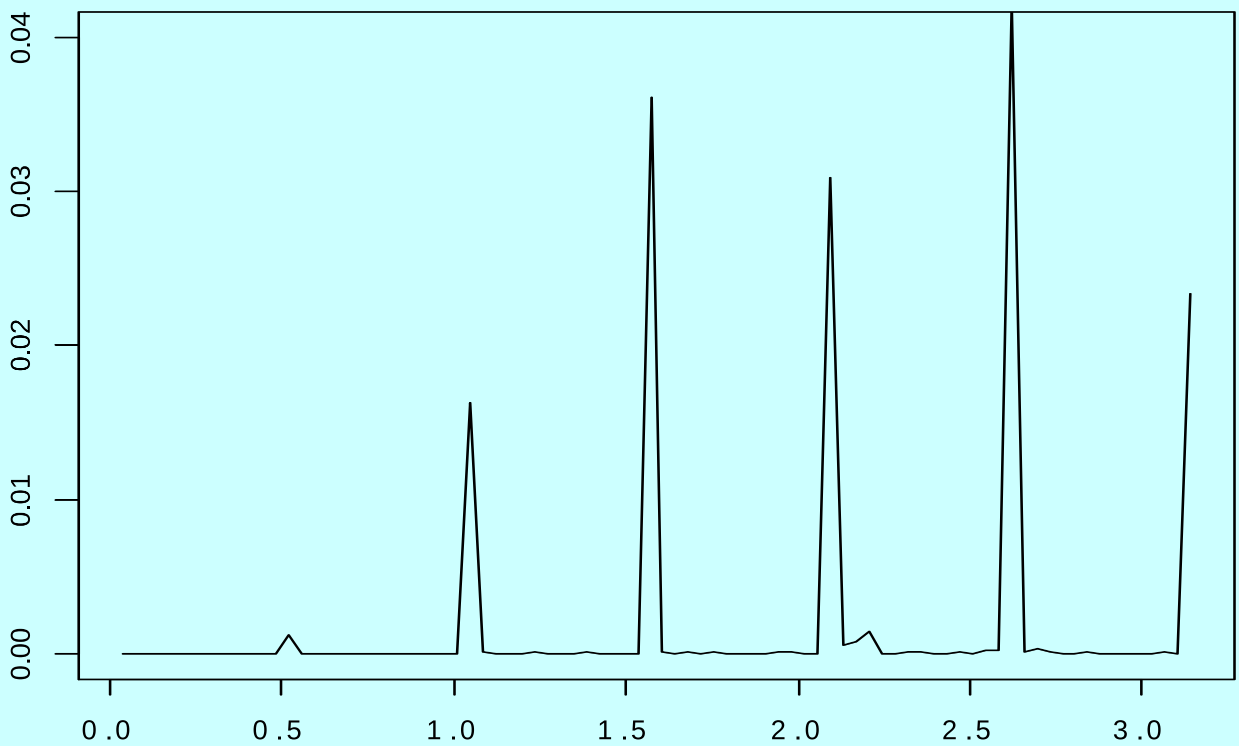
Exercise: Calculate the periodogram of **y.diff** at the Fourier frequencies ω_k, k=1,…,[n/2], and plot it.

```
y.p <- (1/(2*pi*n))*(Mod(y.dft))^2   # Mod = modulus
f <- (2*pi/n)*(1:m)  # vector of m Fourier frequencies
plot(f,y.p,type="l",xlab="",ylab="",ylim=c(0,0.04))
```



Exercise: Redo the last exercise, but this time omit some observations to guarantee that the seasonal frequencies $2\pi j/12$, j=1,…,6, are Fourier frequencies.

```
n.12 <- n - n%%12  # n%%12 = n modulo 12
        #  the remainder we get when we divide n by 12
y.diff.12 <- y.diff[1:n.12]
        #  length of new series y.diff.12 is divisible by 12
m.12 <- floor(n.12/2); f.12 <- (2*pi/n.12)*(1:m.12)
y.dft.12 <- fft(y.diff.12); y.dft.12 <- y.dft.12[2:(m.12+1)]
y.p.12 <- (1/(2*pi*n))*(Mod(y.dft.12))^2
plot(f.12,y.p.12,"l",xlab="",ylab="",ylim=c(0,0.04))
```

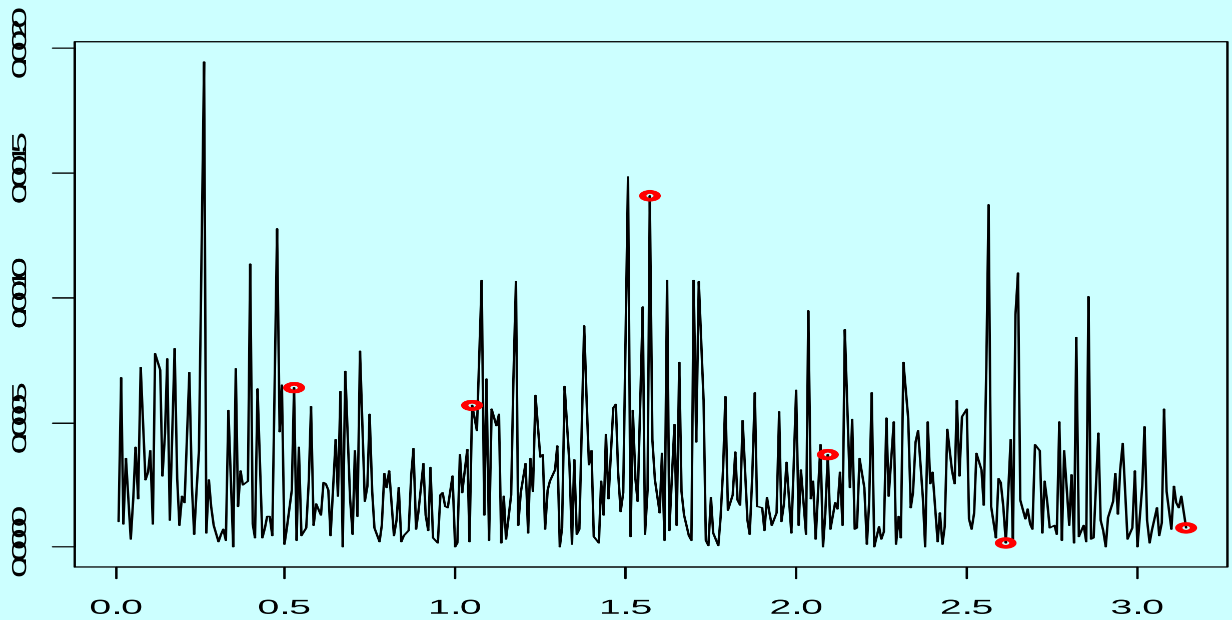


The peaks at the seasonal frequencies are much more pronounced than the slightly displaced peaks in the last exercise. 1F

The periodogram of a financial time series

Exercise: Get the monthly S&P 500 index (**^GSPC**) as **SP500.csv** and plot the periodogram of the returns. Make sure that the number of returns is divisible by 12.

```
SP500 <- read.table("SP500.csv",sep=",",header=T)
N <- length(SP500$Date); SP500[1:N,] <- SP500[N:1,]
y <- ts(data=log(SP500[,7])); r <- diff(y,k=-1)
n <- length(r); n.12 <- n - n%%12; r.12 <- r[1:n.12]
m.12 <- floor(n.12/2); f.12 <- (2*pi/n.12)*(1:m.12)
r.dft.12 <- fft(r.12); r.dft.12 <- r.dft.12[2:(m.12+1)]
r.p.12 <- (1/(2*pi*n))*(Mod(r.dft.12))^2
plot(f.12,r.p.12,"l",xlab="",ylab="")
n.years <- n.12/12  # number of full years
s <- n.years*(1:6)  # indices of seasonal frequencies
lines(f.12[s],r.p.12[s],"p",lwd=3)
```



Since $I(2\pi j/12)$, j=1,2,3 are relatively large, the returns could possibly contain a small seasonal component. 1Y

Statistical properties of the periodogram

Exercise: Show that

$\mathrm{Cov}(x+y,u)=\mathrm{Cov}(x,u)+\mathrm{Cov}(y,u).$ 1C

Suppose that x_t, $t\in\mathbb{Z}$, is white noise with mean μ and variance σ^2 and $\omega_k=\frac{2\pi\cdot k}{n}$, $0<k<\frac{n}{2}$. Then

$$E\hat{A}_k=\frac{2}{n}\sum_{t=1}^{n}E(x_t)\cos(\omega_k t)=\frac{2\mu}{n}\sum_{t=1}^{n}\cos(\omega_k t)=0,$$

$$E\hat{B}_k=\frac{2}{n}\sum_{t=1}^{n}E(x_t)\sin(\omega_k t)=\frac{2\mu}{n}\sum_{t=1}^{n}\sin(\omega_k t)=0,$$

$$\mathrm{Var}(\hat{A}_k)=\frac{4}{n^2}\sum_{t=1}^{n}\mathrm{Var}(x_t)\cos^2(\omega_k t)=\frac{4\sigma^2}{n^2}\sum_{t=1}^{n}\cos^2(\omega_k t)=\frac{4\sigma^2}{n^2}\frac{n}{2}=\frac{2\sigma^2}{n},$$

$$\mathrm{Var}(\hat{B}_k)=\frac{4}{n^2}\sum_{t=1}^{n}\mathrm{Var}(x_t)\sin^2(\omega_k t)=\frac{4\sigma^2}{n^2}\sum_{t=1}^{n}\sin^2(\omega_k t)=\frac{4\sigma^2}{n^2}\frac{n}{2}=\frac{2\sigma^2}{n},$$

$$\begin{aligned}\mathrm{Cov}(\hat{A}_j,\hat{B}_k)&=\frac{4}{n^2}\sum_{s=1}^{n}\sum_{t=1}^{n}\mathrm{Cov}(x_s,x_t)\cos(\omega_j s)\sin(\omega_k t)\\&=\frac{4}{n^2}\sum_{t=1}^{n}\mathrm{Var}(x_t)\cos(\omega_j t)\sin(\omega_k t)\\&=\frac{4\sigma^2}{n^2}\sum_{t=1}^{n}\cos(\omega_j t)\sin(\omega_k t)=0.\end{aligned}$$

Similarly, $\mathrm{Cov}(\hat{A}_j,\hat{A}_k)=\mathrm{Cov}(\hat{B}_j,\hat{B}_k)=0$ if $j\neq k$. 1M

Suppose that $x_1,\ldots,x_n$ are i.i.d. $N(\mu,\sigma^2)$. Then

$$\hat{A}_k=\frac{2}{n}\sum_{t=1}^{n}x_t\cos\left(\frac{2\pi k}{n}t\right),\ \hat{B}_k=\frac{2}{n}\sum_{t=1}^{n}x_t\sin\left(\frac{2\pi k}{n}t\right),$$

are for $1\le k\le m=[\frac{n-1}{2}]$ normally distributed with mean zero and
variance $\frac{2\sigma^2}{n}$. This implies that

$$\sqrt{\frac{n}{2\sigma^2}}\,\hat{A}_k,\ \sqrt{\frac{n}{2\sigma^2}}\,\hat{B}_k$$

have a standard normal distribution. Because of the joint normality of $\hat{A}_k$, $\hat{B}_k$ it follows already from $\mathrm{Cov}(\hat{A}_k,\hat{B}_k)=0$ that $\hat{A}_k$ and $\hat{B}_k$ are independent. For each $1\le k\le m$, the random variable

$$\frac{4\pi}{\sigma^2}I(\omega_k)=\frac{4\pi}{\sigma^2}\frac{n}{8\pi}\hat{R}_k^2=\frac{n}{2\sigma^2}(\hat{A}_k^2+\hat{B}_k^2)$$

is therefore the sum of the squares of two independent standard normal random variables, i.e., it has a chi-squared distribution with 2 degrees of freedom, denoted

$$\frac{4\pi}{\sigma^2}I(\omega_k)\sim\chi^2(2).$$

The random variables $I(\omega_k)$ are not only identically distributed, they are also independent. This follows from the independence of the pairs $(\hat{A}_k,\hat{B}_k)$, k=1,…,m, and the fact that each $I(\omega_k)$ is a function of $\hat{A}_k$ and $\hat{B}_k$. 2D

<u>Exercise:</u> Show that if $k=\frac{n}{2}$, then $\frac{2\pi}{\sigma^2}I(\omega_k)=\frac{2\pi}{\sigma^2}I(\pi)\sim\chi^2(1)$. 21

Simple tests of white noise

If $x_1,\ldots,x_n$ are i.i.d. $N(\mu,\sigma^2)$, then $(4\pi/\sigma^2)\, I(\omega_k)$, $1\le k<\frac{n}{2}$, are i.i.d. $\chi^2(2)$. Since the $\chi^2(2)$ distribution is identical to the exponential distribution with mean 2, $(2\pi/\sigma^2)I(\omega_k)$, $1\le k<\frac{n}{2}$, are i.i.d. Exp(1). Hence,

$$P\left(\tfrac{2\pi}{\sigma^2}I(\omega_k)\le c\right)=\int_0^c e^{-\lambda}d\lambda=-e^{-\lambda}\Big|_0^c=1-e^{-c}=1-\alpha,$$

if $c=-\log(\alpha)$. When n is large we can approximate the unknown parameter σ^2 by the sample variance s^2.

The null hypothesis of Gaussian white noise can be rejected at the approximate 5% level, if for a specified k or for specified $k_1,\ldots,k_j$,

(i) $\frac{2\pi}{s^2}I(\omega_k)>-\log(0.05)$, 2T

or (ii) $\max\left(\frac{2\pi}{s^2}I(\omega_{k_1}),\ldots,\frac{2\pi}{s^2}I(\omega_{k_j})\right)>-\log(1-\sqrt[j]{0.95})$,

or (iii) $\frac{4\pi}{s^2}I(\omega_{k_1})+\ldots+\frac{4\pi}{s^2}I(\omega_{k_j})>\chi^2_{0.95}(2j)$,

where $\chi^2_{0.95}(2j)=5.991, 9.488,\ldots$ for $j=1,2,\ldots$.

Exercise: Show that $Ex=2$ if $x\sim\chi^2(2)$. 22

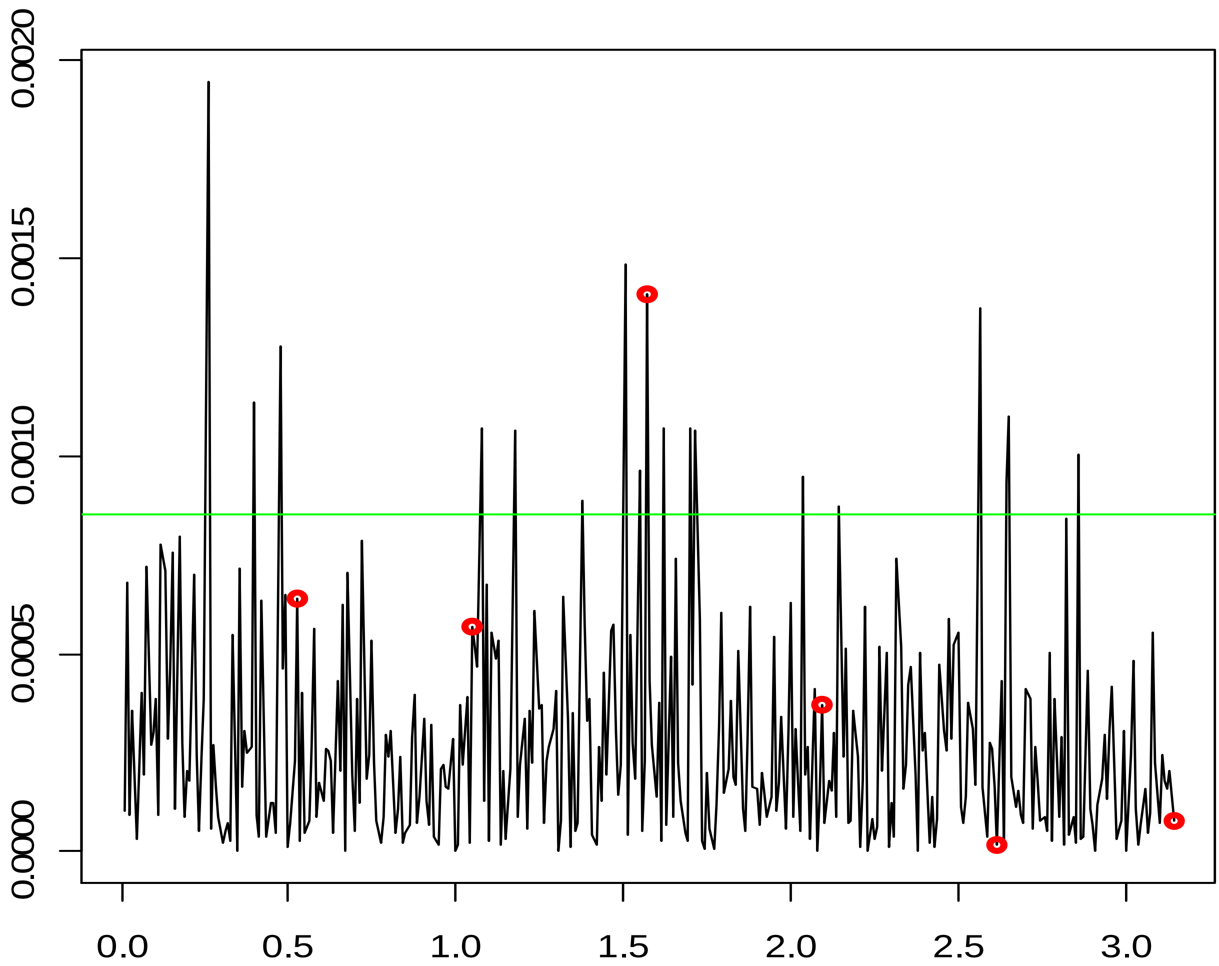
Exercise: Suppose that $x_1,\ldots,x_n$ are i.i.d. $N(\mu,\sigma^2)$. Show that $E\,I(\omega_k)=\frac{\sigma^2}{2\pi}$ for all $1\le k<\frac{n}{2}$. 2W

Exercise: Suppose that J_1, J_2 are i.i.d. Exp(1). Show that $P(\max(J_1,J_2) \le -\log(1-\sqrt{0.95})) = 0.95$.

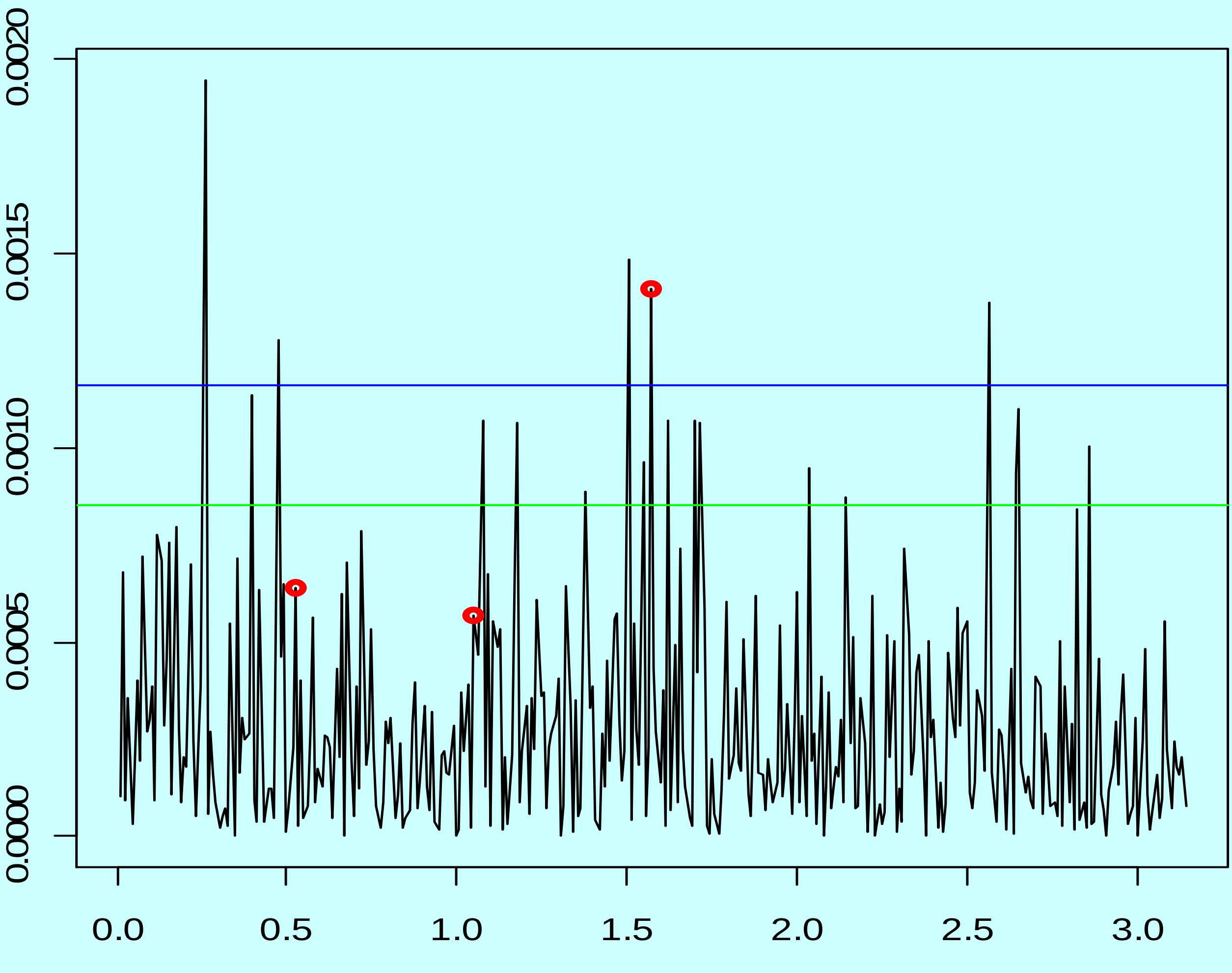
Hint: $P(\max(J_1,J_2) \le c) = P(J_1 \le c \wedge J_2 \le c) = P(J_1 \le c)P(J_2 \le c)$ 2M

Exercise: Add a horizontal line to the S&P 500 periodogram, which represents the critical value (α=0.05) for a test based on the periodogram value at a fixed frequency.

s2 <- var(r.12)
abline(h=-log(0.05)*s2/(2*pi),col="green")



Exercise: Add another horizontal line to the S&P 500 periodogram, which represents the critical value (α=0.05) for a test based on the maximum of three periodogram values at three fixed frequencies.



2C

Exercise: Use the sum of j periodogram values at fixed frequencies as test statistic.

(i) j=3: First three seasonal frequencies

```
sum(r.p.12[s[1:3]])*4*pi/s2 # test statistic
18.34568
qchisq(0.95,6) # 0.95-quantile of a chi2(6) distr.
12.59159
```

$18.34568 > 12.59159 \Rightarrow H_0$ is rejected.

(ii) j=6: All seasonal frequencies

```
sum(r.p.12[s[1:5]])*4*pi/s2+r.p.12[s[6]]*2*pi/s2
21.31183
qchisq(0.95,11) # 0.95-quantile of a chi2(11) distribution
19.67514
```

$21.31183 > 19.67514 \Rightarrow H_0$ is rejected.

Note: The periodogram at frequency π, which is the 6^{th} seasonal frequency has a different distribution under H_0, i.e., $\frac{2\pi}{\sigma^2} I(\omega_k) \sim \chi^2(1)$ if $\omega_k = \pi$ and $\frac{4\pi}{\sigma^2} I(\omega_k) \sim \chi^2(2)$ if $\omega_k \neq \pi$.

2S