Stat 443: Time Series and Forecasting

Assignment 4: Analysis in the Frequency Domain

Caden Hewlett

April 04, 2024

Question 1

Consider the following second-order AR process AR(2) process for $\{X_t\}_{t\in\mathbb{Z}}$, where $\{Z_t\}_{t\in\mathbb{Z}} \overset{\text{iid}}{\sim} \text{WN}(0,\sigma^2)$.

$$X_t = \frac{7}{10}X_{t-1} - \frac{1}{10}X_{t-2} + Z_t$$

We have previously shown that the autocorrelation function $\gamma(h)$ for $h \in \mathbb{Z}$ is given by:

$$\rho(h) = \frac{16}{11}\Big(\frac{1}{2}\Big)^{|h|} - \frac{5}{11}\Big(\frac{1}{5}\Big)^{|h|}, \qquad h \in \mathbb{Z}$$

Part A: Normalized Spectral Density

Derive the normalized spectral density function $f^\star(\omega)$ for $\{X_t\}_{t\in\mathbb{Z}}$.

Solution

We begin by verifying that the Fourier Transform is well defined.

$$\begin{aligned}
\sum_{h=-\infty}^{\infty}|\rho(h)| &= \sum_{h=-\infty}^{\infty}\left|\frac{16}{11}\Big(\frac{1}{2}\Big)^{|h|} - \frac{5}{11}\Big(\frac{1}{5}\Big)^{|h|}\right| \overset{?}{<} \infty \\
\sum_{t=-\infty}^{\infty}|\rho(h)| &= \Big(\frac{16}{11} - \frac{5}{11}\Big) + 2\bigg(\frac{16}{11}\sum_{h=1}^{\infty}\Big(\frac{1}{2}\Big)^{h} - \frac{5}{11}\sum_{h=1}^{\infty}\Big(\frac{1}{5}\Big)^{h}\bigg) \\
\sum_{t=-\infty}^{\infty}|\rho(h)| &= 1 + 2\bigg(\frac{16}{11}\Big(\frac{1/2}{1-1/2}\Big) - \frac{5}{11}\Big(\frac{1/5}{1-1/5}\Big)\bigg) \\
\sum_{t=-\infty}^{\infty}|\rho(h)| &= 1 + 2\bigg(\frac{16}{11} - \frac{5}{11}\Big(\frac{1}{4}\Big)\bigg) = \boxed{\frac{81}{22} < \infty, \ \therefore \text{well-defined.}}
\end{aligned}$$

Now, we evaluate given ρ, recalling that for $\omega \in (0,1)$ and even functions, the normalized spectral density is given by:

$$f^\star(\omega) = \frac{1}{\pi}\bigg(\rho(0) + 2\sum_{h=1}^{\infty}\rho(h)\cos(\omega h)\bigg), \qquad \omega \in (0,1)$$

Where $\rho(0) = 1$.

We will evaluate the infinite sum and substitute the result into the equation above. We will re-instate coefficients A_1 and A_2 from the previous assignment during intermediate steps for simplicity. In addition, we will let $d_1 = 1/2$ and $d_2 = 1/5$, noting that the geometric series equation is usable here as $|d_1|$ and $|d_2|$ are both less than 1.

$$\sum_{h=1}^{\infty} \rho(h)\cos(\omega h) = \sum_{h=1}^{\infty} \left(\frac{16}{11}\left(\frac{1}{2}\right)^{|h|} - \frac{5}{11}\left(\frac{1}{5}\right)^{|h|} \right) \cos(\omega h)$$

$$\sum_{h=1}^{\infty} \rho(h)\cos(\omega h) = \sum_{h=1}^{\infty} \left(A_1(d_1)^{|h|} - A_2(d_2)^{|h|} \right) \cos(\omega h), \quad \text{using variable form.}$$

$$\sum_{h=1}^{\infty} \rho(h)\cos(\omega h) = \underbrace{\sum_{h=1}^{\infty} \left(A_1(d_1)^{|h|} \cos(\omega h) \right)}_{\text{Term 1}} - \underbrace{\sum_{h=1}^{\infty} \left(A_2(d_2)^{|h|} \cos(\omega h) \right)}_{\text{Term 2}}$$

We will evaluate Term 1 and Term 2 separately. We will use the following identities without proof:

$$\cos(\omega h) = \frac{1}{2}\left(e^{\mathrm{i}h\omega} + e^{-\mathrm{i}h\omega}\right), \qquad \mathrm{i} = \sqrt{-1} \tag{1}$$

$$\sum_{n=1}^{\infty} a \cdot r^n = \frac{ar}{(1-r)}, \qquad |r| < 1, \ a \in \mathbb{R} \tag{2}$$

Evaluating Term 1, noting that $|h| = h$ since the summation spans $h \in \mathbb{Z}^+$.

$$\sum_{h=1}^{\infty} \left(A_1(d_1)^{|h|} \cos(\omega h) \right) = A_1 \sum_{h=1}^{\infty} (d_1)^{|h|} \left(\frac{1}{2}\left(e^{\mathrm{i}h\omega} + e^{-\mathrm{i}h\omega}\right) \right), \qquad \text{by (1)}$$

$$\sum_{h=1}^{\infty} \left(A_1(d_1)^{|h|} \cos(\omega h) \right) = \frac{A_1}{2} \sum_{h=1}^{\infty} (d_1)^{h} \left(e^{\mathrm{i}h\omega} + e^{-\mathrm{i}h\omega}\right)$$

$$\sum_{h=1}^{\infty} \left(A_1(d_1)^{|h|} \cos(\omega h) \right) = \frac{A_1}{2} \left(\sum_{h=1}^{\infty} (d_1)^{h} e^{\mathrm{i}h\omega} + \sum_{h=1}^{\infty} (d_1)^{h} e^{-\mathrm{i}h\omega} \right)$$

$$\sum_{h=1}^{\infty} \left(A_1(d_1)^{|h|} \cos(\omega h) \right) = \frac{A_1}{2} \left(\sum_{h=1}^{\infty} \left(d_1\, e^{\mathrm{i}\omega}\right)^{h} + \sum_{h=1}^{\infty} \left(d_1 e^{-\mathrm{i}\omega}\right)^{h} \right)$$

$$\sum_{h=1}^{\infty} \left(A_1(d_1)^{|h|} \cos(\omega h) \right) = \frac{A_1}{2} \left(\frac{d_1 e^{\mathrm{i}\omega}}{1 - d_1 e^{\mathrm{i}\omega}} + \frac{d_1 e^{-\mathrm{i}\omega}}{1 - d_1 e^{-\mathrm{i}\omega}} \right), \qquad \text{by (2)}$$

$$\sum_{h=1}^{\infty} \left(A_1(d_1)^{|h|} \cos(\omega h) \right) = \frac{A_1}{2} \left(\frac{d_1 e^{\mathrm{i}\omega}(1 - d_1 e^{-\mathrm{i}\omega}) + d_1 e^{-\mathrm{i}\omega}(1 - d_1 e^{\mathrm{i}\omega})}{(1 - d_1 e^{\mathrm{i}\omega})(1 - d_1 e^{-\mathrm{i}\omega})} \right)$$

$$\sum_{h=1}^{\infty} \left(A_1(d_1)^{|h|} \cos(\omega h) \right) = \frac{A_1}{2} \left(\frac{d_1(e^{\mathrm{i}\omega} + e^{-\mathrm{i}\omega}) - 2d_1^2}{1 - d_1(e^{\mathrm{i}\omega} + e^{-\mathrm{i}\omega}) + d_1^2} \right)$$

$$\sum_{h=1}^{\infty} \left(A_1(d_1)^{|h|} \cos(\omega h) \right) = \frac{A_1}{2} \left(\frac{2d_1\cos(\omega) - 2d_1^2}{1 - 2d_1\cos(\omega) + d_1^2} \right)$$

$$\sum_{h=1}^{\infty} \left(A_1(d_1)^{|h|} \cos(\omega h) \right) = \frac{A_1(d_1\cos(\omega) - d_1^2)}{1 - 2d_1\cos(\omega) + d_1^2}$$

Similarly, if we repeat this exact same process with A_2 and d_2, noting that $|d_2| < 1$ and $A_2 \in \mathbb{R}$ also satisfy the requirements of (1) and (2), we arrive at Term 2:

$$\sum_{h=1}^{\infty} \left(A_2(d_2)^{|h|} \cos(\omega h) \right) = \frac{A_2(d_2\cos(\omega) - d_2^2)}{1 - 2d_2\cos(\omega) + d_2^2}$$

Then, we can recombine these into our original expression for the infinite sum:

$$
\begin{aligned}
\sum_{h=1}^{\infty} \rho(h)\cos(\omega h) &= \sum_{h=1}^{\infty}\left(A_1(d_1)^{|h|}\cos(\omega h)\right) - \sum_{h=1}^{\infty}\left(A_2(d_2)^{|h|}\cos(\omega h)\right) \\
\sum_{h=1}^{\infty} \rho(h)\cos(\omega h) &= \frac{A_1(d_1\cos(\omega) - d_1^2)}{1 - 2d_1\cos(\omega) + d_1^2} - \frac{A_2(d_2\cos(\omega) - d_2^2)}{1 - 2d_2\cos(\omega) + d_2^2} \\
\sum_{h=1}^{\infty} \rho(h)\cos(\omega h) &= \left(\frac{16}{11}\right)\frac{(\frac{1}{2}\cos(\omega) - \left(\frac{1}{2}\right)^2)}{1 - 2\left(\frac{1}{2}\right)\cos(\omega) + \left(\frac{1}{2}\right)^2} - \left(\frac{5}{11}\right)\frac{(\frac{1}{5}\cos(\omega) - (\frac{1}{5})^2)}{1 - 2(\frac{1}{5})\cos(\omega) + (\frac{1}{5})^2} \\
\sum_{h=1}^{\infty} \rho(h)\cos(\omega h) &= \left(\frac{16}{11}\right)\frac{2\cos(\omega) - 1}{5 - 4\cos(\omega)} - \left(\frac{5}{11}\right)\frac{5\cos(\omega) - 1}{2(13 - 5\cos(\omega))}
\end{aligned}
$$

Then, combining the expressions we can get the final equation for the normalized spectral density:

$$
\begin{aligned}
f^{\star}(\omega) &= \frac{1}{\pi}\left(1 + 2\sum_{h=1}^{\infty}\rho(h)\cos(\omega h)\right) \\
f^{\star}(\omega) &= \frac{1}{\pi}\left(1 + 2\left(\left(\frac{16}{11}\right)\frac{2\cos(\omega) - 1}{5 - 4\cos(\omega)} - \left(\frac{5}{11}\right)\frac{5\cos(\omega) - 1}{2(13 - 5\cos(\omega))}\right)\right) \\
f^{\star}(\omega) &= \boxed{\frac{1}{\pi} + \frac{32}{11\pi}\left(\frac{2\cos(\omega) - 1}{5 - 4\cos(\omega)}\right) - \frac{5}{11\pi}\left(\frac{5\cos(\omega) - 1}{13 - 5\cos(\omega)}\right)}
\end{aligned}
$$

We can verify these results below for a fixed $\omega = \pi/4$.

```r
w = pi/4
# define acf
rho <- function(h){  (16/11)*((1/2)^h) - (5/11)*((1/5)^h) }
# define the "infinite sum"
sum_vals = sum(sapply(1:10000, function(h){
  rho(h)*cos(w*h)
}))
# from the equation...
eqnval = (1/pi)*(rho(0) + 2*sum_vals)
# from our simplification
fw1 = 1/pi
fw2 = (32/(11*pi)) * ( (2 * cos(w) - 1) / (5 - 4*cos(w)) )
fw3 = (5/(11*pi)) * ( (5 * cos(w) - 1) / (13 - 5*cos(w)) )
# comparison
c(eqnval, fw1 + fw2 - fw3)
```

```
## [1] 0.4561755 0.4561755
```

We see that the values are identical for at least the first 10,000 lags at fixed $\omega = \pi/4$.

Part B: Power Spectral Density

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

Solution

We recall from the definition of normalized spectral density that

$$f^\star(\omega) = \frac{f(\omega)}{\sigma_X^2}$$

Where σ_X^2 is the variance of $\{X_t\}_{t\in\mathbb{Z}}$.

Directly, then, we can write $f(\omega)$ as:

$$f(\omega) = \sigma_X^2 f^\star(\omega) = \gamma(0) f^\star(\omega)$$

To find $\gamma(0)$ we re-establish the Yule-Walker equations from Assignment 2, letting $\alpha_1 = 7/10$ and $\alpha_2 = -1/10$

$$\begin{aligned}
\mathbb{E}(X_tX_t) &= \alpha_1\mathbb{E}(X_tX_{t-1}) - \alpha_2\mathbb{E}(X_tX_{t-2}) + \mathbb{E}(X_tZ_t)\\
\mathbb{E}(X_tX_{t-1}) &= \alpha_1\mathbb{E}(X_{t-1}X_{t-1}) - \alpha_2\mathbb{E}(X_{t-1}X_{t-2}) + \mathbb{E}(Z_tX_{t-1})\\
\mathbb{E}(X_tX_{t-2}) &= \alpha_1\mathbb{E}(X_{t-1}X_{t-2}) - \alpha_2\mathbb{E}(X_{t-2}X_{t-2}) + \mathbb{E}(Z_tX_{t-2})
\end{aligned}$$

Which becomes the following system of three equations:

$$\begin{aligned}
\gamma(0) &= \alpha_1\gamma(1) + \alpha_2\gamma(2) + \sigma^2\\
\gamma(1) &= \alpha_1\gamma(0) + \alpha_2\gamma(1)\\
\gamma(2) &= \alpha_1\gamma(1) + \alpha_2\gamma(0)
\end{aligned}$$

We begin by isolating for $\gamma(1)$ as follows:

$$\begin{aligned}
\gamma(1) &= \alpha_1\gamma(0) + \alpha_2\gamma(1)\\
\gamma(1) &= \frac{\alpha_1}{1-\alpha_2}\gamma(0)
\end{aligned}$$

Then, we express $\gamma(2)$ in terms of $\gamma(0)$ by substitution:

$$\begin{aligned}
\gamma(2) &= \alpha_1\gamma(1) + \alpha_2\gamma(0)\\
\gamma(2) &= \alpha_1\Big(\frac{\alpha_1}{1-\alpha_2}\gamma(0)\Big) + \alpha_2\gamma(0)\\
\gamma(2) &= \Big(\frac{\alpha_1^2+\alpha_2-\alpha_2^2}{1-\alpha_2}\Big)\gamma(0)
\end{aligned}$$

Finally, we can simplify to acquire an expression for $\gamma(0)$ in terms of σ^2, α_1 and α_2.

$$\begin{aligned}
\gamma(0) &= \alpha_1\gamma(1) + \alpha_2\gamma(2) + \sigma^2\\
\gamma(0) &= \alpha_1\Big(\frac{\alpha_1}{1-\alpha_2}\Big)\gamma(0) + \alpha_2\Big(\frac{\alpha_1^2+\alpha_2-\alpha_2^2}{1-\alpha_2}\Big)\gamma(0) + \sigma^2\\
\gamma(0) &= \Big(\frac{\alpha_1^2+\alpha_1^2\alpha_2+\alpha_2^2-\alpha_2^3}{1-\alpha_2}\Big)\gamma(0) + \sigma^2\\
\gamma(0)\Big(1 - \frac{\alpha_1^2+\alpha_1^2\alpha_2+\alpha_2^2-\alpha_2^3}{1-\alpha_2}\Big) &= \sigma^2\\
\gamma(0)\Big(\frac{(1-\alpha_2)-\alpha_1^2-\alpha_1^2\alpha_2-\alpha_2^2+\alpha_2^3}{1-\alpha_2}\Big) &= \sigma^2\\
\gamma(0) &= \frac{\sigma^2(1-\alpha_2)}{1-\alpha_2-\alpha_1^2-\alpha_1^2\alpha_2-\alpha_2^2+\alpha_2^3} \qquad \square
\end{aligned}$$

Then, substituting $\alpha_1 = 7/10$ and $\alpha_2 = -1/10$ and simplifying, we arrive at $\gamma(0)$ in terms of σ^2.

$$\gamma(0) = \frac{275\sigma^2}{162}$$

Then, we can verify our results and simplification as follows, assuming $\sigma^2 = 1$.

```r
set.seed(3)
# declare constants
alpha_1 <- 7/10
alpha_2 <- -1/10
sigma_sq <- 1
# compute gamma(0) according to equation
gamma_0 <- (sigma_sq * (1 - alpha_2) /
        (1 - alpha_2 - alpha_1^2 - alpha_1^2 * alpha_2 - alpha_2^2 + alpha_2^3))
# check gamma_0 computation with simplified fraction
if(!all.equal(gamma_0, 275/162)){print("Simplifcation Doesn't Match.")}
# simulate many draws from the AR(2) model
ar_params <- c(alpha_1, alpha_2)
simulated_data <- arima.sim(n = 5000000,
                            model = list(ar = c(7/10, -1/10)))
# compute the variance
simulated_gamma_0 <- var(simulated_data)
# use both regular and simplified fraction version
round(c((275/162), simulated_gamma_0, abs(275/162 - simulated_gamma_0)), 4)
```

```
## [1] 1.6975 1.6979 0.0004
```

It seems the computation very closely approximates the truth. Hence, we conclude that:

$$f(\omega) = \gamma(0)f^{\star}(\omega) = \frac{275\sigma^2}{162}\left(\frac{1}{\pi} + \frac{32}{11\pi}\left(\frac{2\cos(\omega) - 1}{5 - 4\cos(\omega)}\right) - \frac{5}{11\pi}\left(\frac{5\cos(\omega) - 1}{13 - 5\cos(\omega)}\right)\right)$$

Part C: Plot and Comments

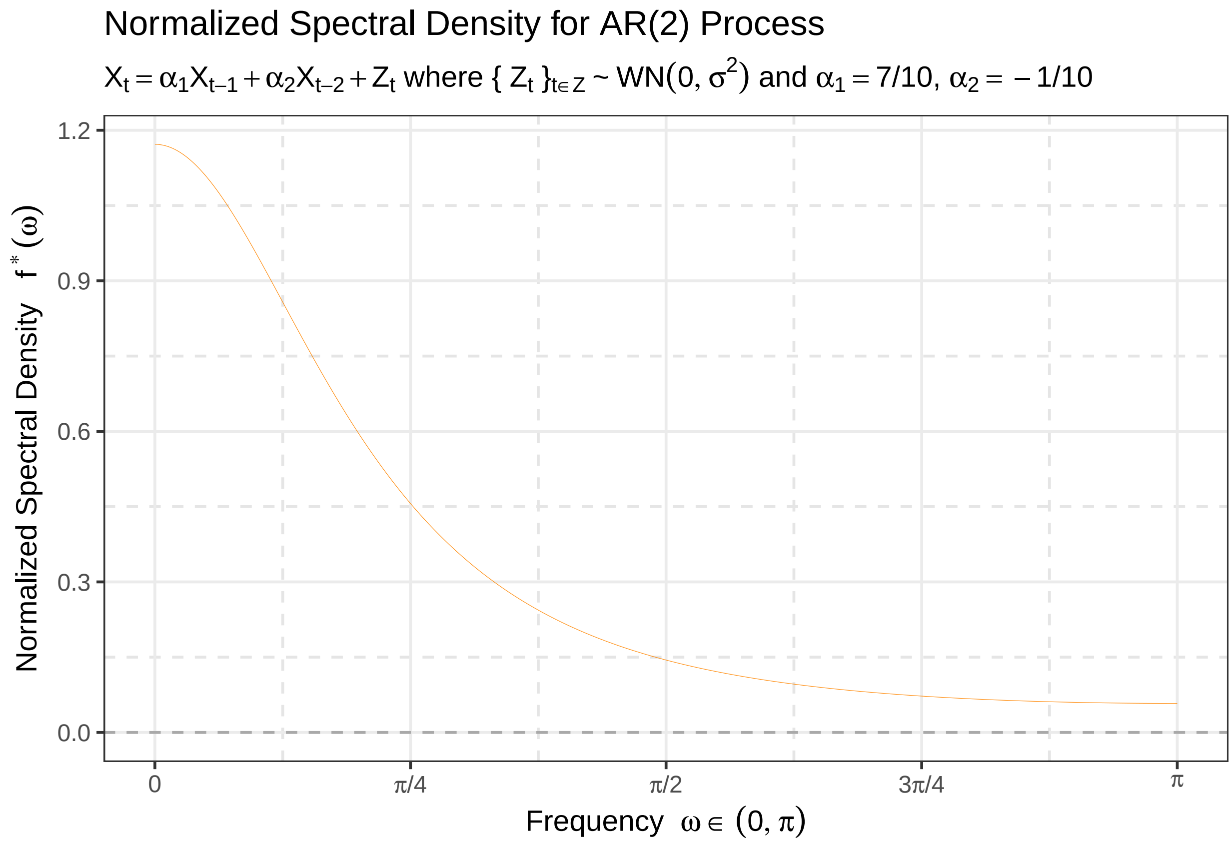
Plot the normalized spectral density and comment on its behaviour.

The normalized spectral density equation ended up being quite long, so we'll define each term one-by-one in the function below:

```
f = function(w){
  term_1 = 1
  term_2 = (32/11)*((2*cos(w) - 1) / (5 - 4*cos(w)))
  term_3 = (5 /11)*((5*cos(w) - 1) / (13 - 5*cos(w)))
  return ( (1/pi)*(term_1 + term_2 - term_3) )
}
```

Then, we plot the function for a long sequence of ω values in $(0, \pi)$:

```
omega = seq(from = 0, to = pi, length.out = 1e4)
# define data frame for values
p1df = data.frame(omega = omega,
           f_star_omega = f(omega))
# build plot
p1 <- ggplot(p1df, aes(x = omega, y = f_star_omega)) +
  geom_line(color = "#ff8600", linewidth = 0.1) +
  labs(
    title = "Normalized Spectral Density for AR(2) Process",
    subtitle = TeX(paste(
      "$X_t = \\alpha_1 X_{t-1} + \\alpha_2 X_{t-2} + Z_t$",
      "where",
      "$\\{\\,Z_t\\,\\}_{t \\in {Z}} \\sim WN(0, \\sigma^2)$",
      "and $\\alpha_1 =$7/10, $\\alpha_2 =-$1/10")),
    y = TeX("Normalized Spectral Density   $f^{*}(\\omega)$"),
    x = TeX("Frequency  $\\omega \\in (0, \\pi)$")
  ) + theme_bw() +
  geom_hline(yintercept = 0, lty = 'dashed', col = "darkgrey")+
  scale_x_continuous(
    breaks = c(0, pi/4, pi/2, 3*pi/4, pi),
    labels = c(TeX("0"), TeX("$\\pi$/4"), TeX("$\\pi$/2"),
               TeX("$3\\pi$/4"), TeX("$\\pi$")))+
  theme(panel.grid.minor = element_line(
    color = "grey90",
    linetype = "dashed",
    linewidth = 0.5
  ))
print(p1)
```



Comments: It appears that the normalized spectral density plot is largely dominated by low frequencies. We can tell this is the case due to the fact that the largest values of $f^\star(\omega)$ are for small $\omega \in (0, \pi)$. Specifically, for $\omega \in (0, \pi/4)$ we find the majority of frequencies with observed normalized density greater than 0.5. This tells us that a greater proportion of the variance inherent in the process X_t can be attributed to the lower frequencies. As we would expect of a normalized spectral density dominated by low ω values, we see that $f^\star(\omega)$ is strictly decreasing as $\omega \to \pi$.

Question 2

Given the spectral density function

$$f(\omega) = \frac{1}{\pi}\Big(62 - 70\cos(\omega) + 12\cos(2\omega)\Big), \qquad \omega \in (0, 1)$$

compute the autocovariance function $\gamma(h)$ and autocorrelation function $\rho(h)$ of the underlying stochastic process, where $h \in \mathbb{Z}$.

We will begin with computing the autocovariance function $\gamma(h)$ at lag $h = 0$, then at $h = 1$, $h = 2$ and finally $h > 2$.

Further, for notational simplicity, we let $a = 62$, $b = 70$ and $c = 12$, meaning that:

$$f(\omega) = \frac{1}{\pi}\Big(a - b\cos(\omega) + c\cos(2\omega)\Big), \qquad \omega \in (0, 1)$$

Part A: Lag 0

For $h = 0$, we have the following:

$$\gamma(0) = \int_0^{\pi} f(\omega)\cos(h\omega)\mathrm{d}\omega = \int_0^{\pi}\left[\frac{1}{\pi}\Big(a - b\cos(\omega) + c\cos(2\omega)\Big)\right]\cos(0 \times \omega)\mathrm{d}\omega$$

Let's evaluate this integral to find the autocovariance at lag zero.

$$\begin{aligned}
\gamma(0) &= \int_0^{\pi}\left[\frac{1}{\pi}\Big(a - b\cos(\omega) + c\cos(2\omega)\Big)\right]\cos(0)\mathrm{d}\omega \\
\gamma(0) &= \frac{a}{\pi}\int_0^{\pi}\mathrm{d}\omega - \frac{b}{\pi}\int_0^{\pi}\cos(\omega)\mathrm{d}\omega + \frac{c}{\pi}\int_0^{\pi}\cos(2\omega)\mathrm{d}\omega \\
\gamma(0) &= \frac{a}{\pi}\Big(\omega\,\Big|_0^{\pi}\Big) - \frac{b}{\pi}\Big(\sin(\omega)\,\Big|_0^{\pi}\Big) + \frac{c}{\pi}\Big(\frac{1}{2}\sin(2\omega)\,\Big|_0^{\pi}\Big) \\
\gamma(0) &= \frac{a}{\pi}(\pi - 0) - \frac{b}{\pi}\Big(\sin(\pi) - \sin(0)\Big) + \frac{c}{2\pi}\Big(\sin(2\pi) - \sin(0)\Big) \\
\gamma(0) &= a - \frac{b}{\pi}\Big(0 - 0\Big) + \frac{c}{2\pi}\Big(0 - 0\Big) \\
\gamma(0) &= a = \boxed{62}
\end{aligned}$$

Part B: Lag 1

We repeat these calculations, but now for $h = 1$. We retain $a = 62$, $b = 70$ and $c = 12$.

For $h = 1$, we have the following:

$$\gamma(1) = \int_0^{\pi} f(\omega)\cos(h\omega)\mathrm{d}\omega = \int_0^{\pi}\left[\frac{1}{\pi}\Big(a - b\cos(\omega) + c\cos(2\omega)\Big)\right]\cos(1 \times \omega)\mathrm{d}\omega$$

As before, we expand and evaluate this integral.

We will use the following trigonometric identities without proof:

$$\cos^2(\theta) = 1 - \sin^2(\theta) \tag{3}$$

$$\sin^2(\theta) = \frac{1}{2}\big(1 - \cos(2\theta)\big) \tag{4}$$

$$\gamma(1) = \int_0^\pi \left[\frac{1}{\pi}\Big(a - b\cos(\omega) + c\cos(2\omega)\Big)\right]\cos(\omega)\mathrm{d}\omega$$
$$\gamma(1) = \frac{a}{\pi}\int_0^\pi \cos(\omega)\mathrm{d}\omega - \frac{b}{\pi}\int_0^\pi \cos(\omega)\cos(\omega)\mathrm{d}\omega + \frac{c}{\pi}\int_0^\pi \cos(2\omega)\cos(\omega)\mathrm{d}\omega$$
$$\gamma(1) = \frac{a}{\pi}\Big(\sin(\omega)\Big|_0^\pi\Big) - \frac{b}{\pi}\int_0^\pi \cos^2(\omega)\mathrm{d}\omega + \frac{c}{\pi}\int_0^\pi \cos(\omega)\cos(2\omega)\mathrm{d}\omega$$
$$\gamma(1) = 0 - \frac{b}{\pi}\int_0^\pi \big(1 - \sin^2(\omega)\big)\mathrm{d}\omega + \frac{c}{\pi}\int_0^\pi \cos(\omega)\cos(2\omega)\mathrm{d}\omega \qquad \text{applying (3)}$$
$$\gamma(1) = -\frac{b}{\pi}\int_0^\pi \mathrm{d}\omega + \frac{b}{\pi}\int_0^\pi \sin^2(\omega)\mathrm{d}\omega + \frac{c}{\pi}\int_0^\pi \cos(\omega)\cos(2\omega)\mathrm{d}\omega$$
$$\gamma(1) = -b + \frac{b}{\pi}\int_0^\pi \frac{1}{2}\big(1 - \cos(2\theta)\big)\mathrm{d}\omega + \frac{c}{\pi}\int_0^\pi \cos(\omega)\cos(2\omega)\mathrm{d}\omega \qquad \text{applying (4)}$$
$$\gamma(1) = -b + \frac{b}{\pi}\int_0^\pi \frac{1}{2}\mathrm{d}\omega - \frac{b}{2\pi}\underbrace{\int_0^\pi \cos(2\theta)\mathrm{d}\omega}_{\text{shown to be 0 in Part A}} + \frac{c}{\pi}\int_0^\pi \cos(\omega)\cos(2\omega)\mathrm{d}\omega$$
$$\gamma(1) = -b + \frac{b}{\pi}\int_0^\pi \frac{1}{2}\mathrm{d}\omega + \frac{c}{\pi}\int_0^\pi \cos(\omega)\cos(2\omega)\mathrm{d}\omega$$
$$\gamma(1) = -b + \frac{b}{2} + \frac{c}{\pi}\int_0^\pi \cos(\omega)\cos(2\omega)\mathrm{d}\omega$$

We apply a manipulation of equation (4) to evaluate the final integral:

$$\gamma(1) = -\frac{b}{2} + \frac{c}{\pi}\int_0^\pi \cos(\omega)\cos(2\omega)\mathrm{d}\omega$$
$$\gamma(1) = -\frac{b}{2} + \frac{c}{\pi}\int_0^\pi \cos(\omega)\Big(1 - 2\sin^2(\omega)\Big)\mathrm{d}\omega$$
$$\gamma(1) = -\frac{b}{2} + \frac{c}{\pi}\int_0^\pi \cos(\omega)\mathrm{d}\omega - \frac{2c}{\pi}\int_0^\pi \cos(\omega)\sin^2(\omega)\mathrm{d}\omega$$
$$\gamma(1) = -\frac{b}{2} + \frac{c}{\pi}\Big(\sin(\omega)\Big|_0^\pi\Big) - \underbrace{\frac{2c}{\pi}\int_0^\pi \cos(\omega)\sin^2(\omega)\mathrm{d}\omega}_{\text{let } q=\sin(\omega)}$$
$$\gamma(1) = -\frac{b}{2} + 0 - \frac{2c}{\pi}\int_0^\pi \cos(\omega)q^2\Big(\frac{1}{\cos(\omega)}\Big)\mathrm{d}q$$
$$\gamma(1) = -\frac{b}{2} - \frac{2c}{3\pi}\left(q^3\Big|_{\omega=0}^{\pi}\right)$$
$$\gamma(1) = -\frac{b}{2} - \frac{2c}{3\pi}\left(\sin^3(\pi) - \sin^3(0)\right)$$
$$\gamma(1) = -\frac{b}{2} = -\frac{70}{2} = \boxed{-35}$$

Part C: Lag 2

We repeat these calculations, but now for $h = 2$. We retain $a = 62$, $b = 70$ and $c = 12$.

For $h = 2$, we have the following:

$$\gamma(2) = \int_0^\pi f(\omega)\cos(h\omega)\mathrm{d}\omega = \int_0^\pi \left[\frac{1}{\pi}\Big(a - b\cos(\omega) + c\cos(2\omega)\Big)\right]\cos(2\times\omega)\mathrm{d}\omega$$

As before, we expand $f(\omega)$ and evaluate this integral.

$$\begin{aligned}
\gamma(2) &= \int_0^\pi \left[\frac{1}{\pi}\Big(a - b\cos(\omega) + c\cos(2\omega)\Big)\right]\cos(2\omega)\mathrm{d}\omega \\
\gamma(2) &= \frac{a}{\pi}\int_0^\pi \cos(2\omega)\mathrm{d}\omega - \frac{b}{\pi}\underbrace{\int_0^\pi \cos(\omega)\cos(2\omega)\mathrm{d}\omega}_{\text{Shown in Part B to be zero.}} + \frac{c}{\pi}\int_0^\pi \cos(2\omega)\cos(2\omega)\mathrm{d}\omega \\
\gamma(2) &= \frac{a}{2\pi}\Big(\sin(2\omega)\Big|_0^\pi\Big) - \frac{b}{\pi}(0) + \frac{c}{\pi}\int_0^\pi \cos^2(2\omega)\mathrm{d}\omega, \qquad \text{applying (3)} \\
\gamma(2) &= \frac{a}{2\pi}\Big(\sin(2\pi) - \sin(0)\Big) + \frac{c}{\pi}\int_0^\pi \Big(1 - \sin^2(2\omega)\Big)\mathrm{d}\omega \\
\gamma(2) &= \frac{a}{2\pi}(0) + \frac{c}{\pi}\int_0^\pi 1\mathrm{d}\omega - \frac{c}{\pi}\int_0^\pi \sin^2(2\omega)\mathrm{d}\omega \\
\gamma(2) &= \frac{c}{\pi}(\pi) - \frac{c}{\pi}\int_0^\pi \frac{1}{2}\Big(1 - \cos(2\cdot 2\omega)\Big)\mathrm{d}\omega, \qquad \text{applying (4)} \\
\gamma(2) &= c - \frac{c}{2\pi}\int_0^\pi 1\mathrm{d}\omega + \frac{c}{2\pi}\int_0^\pi \cos(4\omega)\mathrm{d}\omega \\
\gamma(2) &= c - \frac{c}{2\pi}(\pi) + \frac{c}{2\pi}\Big(\frac{1}{4}\sin(4\omega)\Big|_0^\pi\Big) \\
\gamma(2) &= c - \frac{c}{2} + \frac{c}{8\pi}\Big(\sin(4\pi) - \sin(0)\Big) \\
\gamma(2) &= \frac{c}{2} = \frac{12}{2} = \boxed{6}
\end{aligned}$$

Part D: Lags Greater Than 2

Before we evaluate lags $h > 2 \subset \mathbb{Z}$, we will establish the following lemma:

Lemma 1: Let $m, n \in \mathbb{Z}^+$. We will show that:

$$\forall (m,n) \in \mathbb{Z}^+ \text{ s.t. } m \neq n, \int_0^\pi \cos(m\omega)\cos(n\omega)\,\mathrm{d}\omega = 0$$

Proof: In order to show this equality holds, we will use the following product-to-sum identity of cosines without proof. Let $\theta, \vartheta \in \mathbb{R}$.

$$\cos(\theta)\cos(\vartheta) = \frac{1}{2}\cos(\theta + \vartheta) + \frac{1}{2}\cos(\theta - \vartheta) \tag{2.1}$$

We will use (2.1) to evaluate the integral.

$$\int_0^\pi \cos(m\omega)\cos(n\omega)\,\mathrm{d}\omega = \int_0^\pi \left(\frac{1}{2}\cos(m\omega+n\omega)+\frac{1}{2}\cos(m\omega-n\omega)\right)\mathrm{d}\omega$$
$$\int_0^\pi \cos(m\omega)\cos(n\omega)\,\mathrm{d}\omega = \frac{1}{2}\int_0^\pi \underbrace{\cos\big((m+n)\omega\big)}_{\text{Let } q=(m+n)\omega}\,\mathrm{d}\omega + \frac{1}{2}\int_0^\pi \underbrace{\cos\big((m-n)\omega\big)}_{\text{Let } \nu=(m-n)\omega}\,\mathrm{d}\omega$$
$$\int_0^\pi \cos(m\omega)\cos(n\omega)\,\mathrm{d}\omega = \frac{1}{2(m+n)}\int_0^\pi \cos(q)\,\mathrm{d}q + \frac{1}{2(m-n)}\int_0^\pi \cos(\nu)\,\mathrm{d}\nu, \qquad \text{requires } m \neq n$$
$$\int_0^\pi \cos(m\omega)\cos(n\omega)\,\mathrm{d}\omega = \frac{1}{2(m+n)}\left(\sin(q)\Big|_{\omega=0}^{\pi}\right) + \frac{1}{2(m-n)}\left(\sin(\nu)\Big|_{\omega=0}^{\pi}\right)$$
$$\int_0^\pi \cos(m\omega)\cos(n\omega)\,\mathrm{d}\omega = \frac{1}{2(m+n)}\left(\sin(m\omega+n\omega)\Big|_0^{\pi}\right) + \frac{1}{2(m-n)}\left(\sin(m\omega-n\omega)\Big|_0^{\pi}\right)$$
$$\int_0^\pi \cos(m\omega)\cos(n\omega)\,\mathrm{d}\omega = \frac{1}{2(m+n)}\Big(\sin\big((m+n)\pi\big)-\sin(0)\Big) + \frac{1}{2(m-n)}\Big(\sin\big((m-n)\pi\big)-\sin(0)\Big)\Big)$$
$$\int_0^\pi \cos(m\omega)\cos(n\omega)\,\mathrm{d}\omega = \frac{1}{2(m+n)}\sin\big((m+n)\pi\big) + \frac{1}{2(m-n)}\sin\big((m-n)\pi\big)$$

We note that $m \in \mathbb{Z}$ and $n \in \mathbb{Z}$. Further, by properties of integers, we know that $m+n \in \mathbb{Z}$ and $m-n \in \mathbb{Z}$.

We can hence note that:

$$(m \in \mathbb{Z} \wedge n \in \mathbb{Z}) \implies \exists \ell \in \mathbb{Z} \text{ s.t. } \ell = m+n$$

Since $(m,n) \in \mathbb{Z}^+$ and $m \neq n$, we know that $m+n>0$, $\therefore \ell > 0$.

Similarly, we have that:

$$(m \in \mathbb{Z} \wedge n \in \mathbb{Z}) \implies \exists \varrho \in \mathbb{Z} \text{ s.t. } \varrho = m-n$$

Further, since $(m,n) \in \mathbb{Z}^+$ and $m \neq n$, we know that $\varrho \neq 0$. We also note by properties of sine that $\forall k \in \mathbb{Z}, \sin(k \cdot \pi) = 0$. This will be used below without proof.

Finally, we can rewrite our simplified integral in terms of ℓ and ϱ and solve.

$$\int_0^\pi \cos(m\omega)\cos(n\omega)\,\mathrm{d}\omega = \frac{1}{2\ell}\sin(\ell\pi) + \frac{1}{2\varrho}\sin(\varrho\pi)$$
$$\int_0^\pi \cos(m\omega)\cos(n\omega)\,\mathrm{d}\omega = \frac{1}{2\ell}(0) + \frac{1}{2\varrho}(0), \quad \text{since } \forall k \in \mathbb{Z}, \sin(k\cdot\pi)=0$$
$$\int_0^\pi \cos(m\omega)\cos(n\omega)\,\mathrm{d}\omega = (0)+(0), \qquad \text{since } \ell \neq 0 \text{ and } \varrho \neq 0$$
$$\int_0^\pi \cos(m\omega)\cos(n\omega)\,\mathrm{d}\omega = 0$$

Hence, we conclude that

$$\forall (m,n) \in \mathbb{Z}^+ \text{ s.t. } m \neq n, \int_0^\pi \cos(m\omega)\cos(n\omega)\,\mathrm{d}\omega = 0, \quad \text{as required.} \ \square$$

Application of the Lemma

Now, we can show that $\forall h > 2 \subseteq \mathbb{Z}^+$ that:

$$\gamma(h) = \int_0^\pi f(\omega)\cos(h\omega)\mathrm{d}\omega = 0$$

Proof: We begin by expanding $f(\omega)$ with our previous variable assignments:

$$\int_0^\pi f(\omega)\cos(h\omega)\mathrm{d}\omega = \int_0^\pi \left[\frac{1}{\pi}\Big(a - b\cos(\omega) + c\cos(2\omega)\Big)\right]\cos(h\omega)\mathrm{d}\omega$$
$$\int_0^\pi f(\omega)\cos(h\omega)\mathrm{d}\omega = \frac{a}{\pi}\int_0^\pi \cos(h\omega)\mathrm{d}\omega - \frac{b}{\pi}\int_0^\pi \cos(\omega)\cos(h\omega)\mathrm{d}\omega + \frac{c}{\pi}\int_0^\pi \cos(2\omega)\cos(h\omega)\mathrm{d}\omega$$
$$\int_0^\pi f(\omega)\cos(h\omega)\mathrm{d}\omega = \frac{a}{\pi}\int_0^\pi \cos(0\cdot\omega)\cos(h\cdot\omega)\mathrm{d}\omega - \frac{b}{\pi}\int_0^\pi \cos(1\cdot\omega)\cos(h\cdot\omega)\mathrm{d}\omega + \frac{c}{\pi}\int_0^\pi \cos(2\cdot\omega)\cos(h\cdot\omega)\mathrm{d}\omega$$

Now, from the expansion on the third line, we can draw an interesting parallel to the lemma established previously. Let $M = \{0, 1, 2\}$ and $N = \{n \in \mathbb{Z} : n > 2\}$.

We see by construction of these sets $M \cap N = \emptyset$. Importantly,

$$M \cap N = \emptyset \implies \forall n \in N, \forall m \in M,\ m \neq n$$

Now, if we let $M = \{m_1, m_2, m_3\} = \{0, 1, 2\}$ and $n \in N := \{h \in \mathbb{Z} : h > 2\}$, we can rewrite our integral in terms of m and n and simplify using our lemma.

$$\int_0^\pi f(\omega)\cos(h\omega)\mathrm{d}\omega = \frac{a}{\pi}\int_0^\pi \underbrace{\cos(m_1\cdot\omega)\cos(n\cdot\omega)}_{\forall n\in N,\, m_1\neq n}\mathrm{d}\omega - \frac{b}{\pi}\int_0^\pi \underbrace{\cos(m_2\cdot\omega)\cos(n\cdot\omega)}_{\forall n\in N,\, m_2\neq n}\mathrm{d}\omega + \frac{c}{\pi}\int_0^\pi \underbrace{\cos(m_2\cdot\omega)\cos(n\cdot\omega)}_{\forall n\in N,\, m_3\neq n}\mathrm{d}\omega$$
$$\int_0^\pi f(\omega)\cos(h\omega)\mathrm{d}\omega = \frac{a}{\pi}(0) - \frac{b}{\pi}(0) + \frac{c}{\pi}(0),\ \text{by set construction and Lemma 1.}$$
$$\int_0^\pi f(\omega)\cos(h\omega)\mathrm{d}\omega = \boxed{0}$$

So, we can conclude that:

$$\forall h > 2, \gamma(h) = \int_0^\pi f(\omega)\cos(h\omega)\mathrm{d}\omega = 0,\ \text{as required.}\ \square$$

Conclusions

Hence, bringing all the previous parts together, we can conclude that for $h \in \mathbb{Z}$:

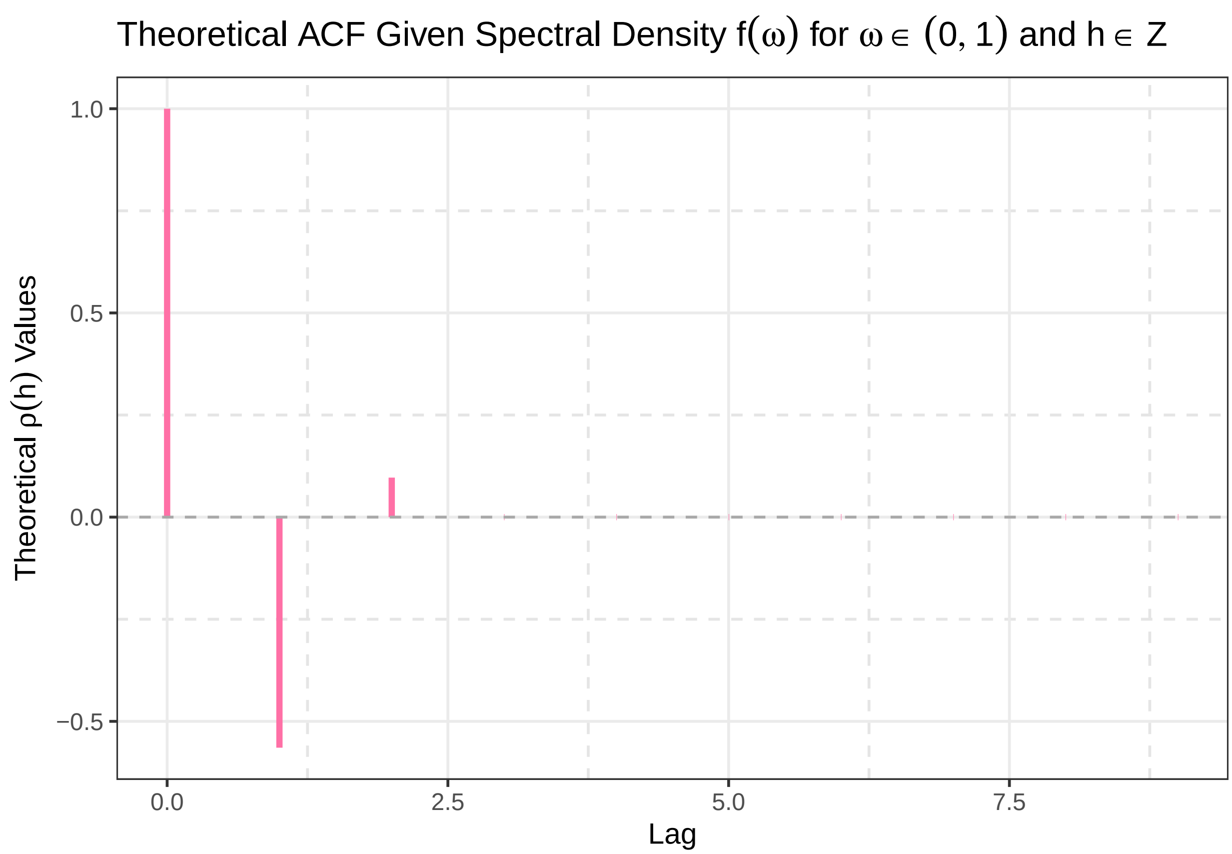
$$\gamma(h) = \begin{cases} \gamma(-h), & h < 0 \\ 62, & h = 0 \\ -35, & h = 1 \\ 6, & h = 2 \\ 0, & \text{otherwise} \end{cases}$$

Similarly, we find $\rho(h)$ for $h \in \mathbb{Z}$ by dividing $\gamma(h)$ by $\sigma_X^2 = \gamma(0) = 62$.

$$\rho(h) = \begin{cases} \rho(-h), & h < 0 \\ 1, & h = 0 \\ -35/62, & h = 1 \\ 3/31, & h = 2 \\ 0, & \text{otherwise} \end{cases}$$

We can also provide a plot of the theoretical ACF to get a more easily interpretable idea of how the autocorrelation of X_t changes for lags $h \in \mathbb{Z}$.

```
p2DF = data.frame(
  h = 0:9,
  rh = c(1, -35/62, 3/31, rep(0, times = 7))
)
p2 <- ggplot(p2DF, aes(x = h, y = rh)) +
  geom_segment(aes(xend = h, yend = 0),
               color = "#FF70A6",
               linewidth = 1.1) +
  geom_hline(yintercept = 0,
             linetype = "dashed",
             color = "darkgray") +
  labs(x = "Lag",  y = TeX("Theoretical $\\rho(h)$ Values"),
       title =  TeX(paste("Theoretical ACF Given Spectral Density $f(\\omega)$",
       "for $\\omega \\in (0, 1)$ and $h \\in Z$")) )+
  theme_bw() +
  theme(panel.grid.minor = element_line(color = "grey90",
                                        linetype = "dashed", linewidth = 0.5))
print(p2)
```



Question 3

The data file `accel_watch.csv` contains three axes of accelerometer data for a test subject that was walking at a steady pace while wearing a biologging watch. The accelerometer data was recorded in meters per second squared and measured every 0.05 seconds (i.e., at a rate of 20 observations per second)

Part A: Data Preprocessing and Investigation

Part A.1.

Read the data into R.

```
df = read.csv("accel_watch.csv")
```

Part A.2.

Create a vector that contains the magnitude of acceleration at each time index.

We will use the following to define the column:

$$\|\vec{a}\| = \sqrt{a_x^2 + a_y^2 + a_z^2}$$

```
df$Mag = sqrt(df$Ax^2 + df$Ay^2 + df$Az^2)
kable( t( head(df$Mag) ), caption = "Preview of $\\| \\vec{a} \\|$ Vector")
```

Table 1: Preview of $\|\vec{a}\|$ Vector

17.88641	15.61458	12.2083	12.02874	11.79417	12.63943

Part A.3.

Coerce the vector into a time series object.

We are told that the accelerometer data was recorded in meters per second squared and measured every 0.05 seconds (i.e., at a rate of 20 observations per second.) Hence, we set `frequency = 20`, as `?ts` informs us that the `frequency` parameter is "the number of observations per unit of time."

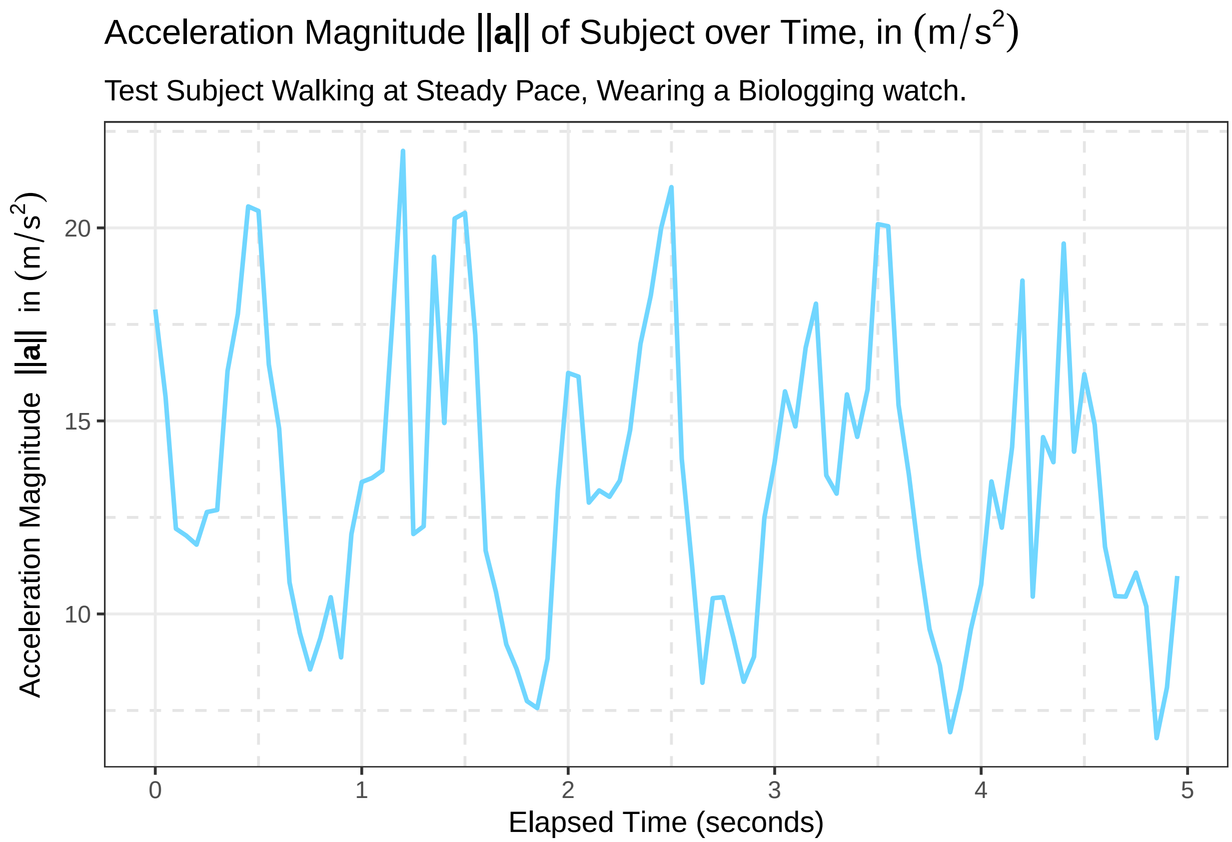
```
accel_ts = ts(df$Mag, frequency = 20)
```

Part A.4.

Plot the resulting time series and its sample acf. (Make sure to properly label the axes and provide titles for the plots.)

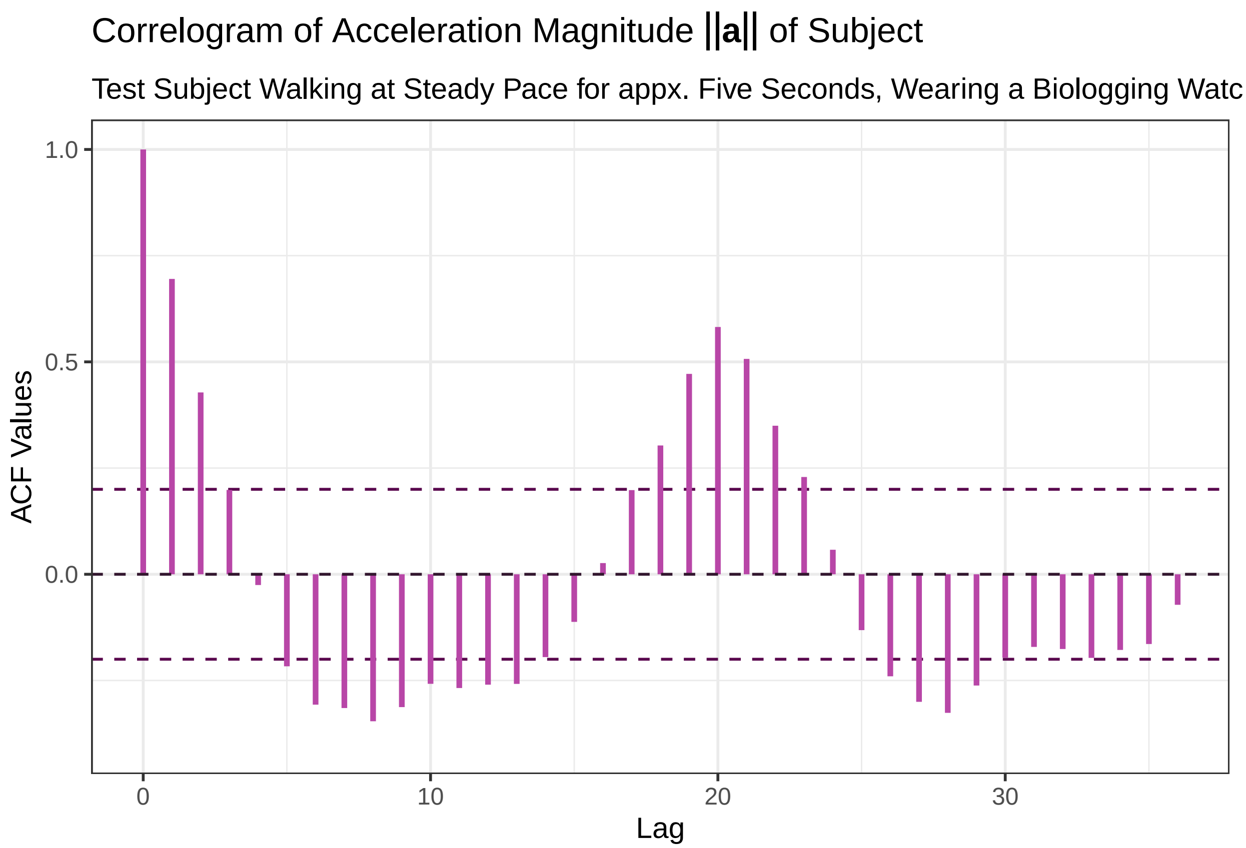
Time Series Plot

```r
# coerce time series into a plottable object
p3DF = fortify.zoo(accel_ts)
# build plot
p3 <- ggplot(p3DF, aes(x = Index, y = accel_ts)) +
  geom_line(color = "#70D6FF", linewidth = 0.8) +
  labs(
    title = TeX(paste("Acceleration Magnitude",
    "$\\| \\textbf{a} \\|$ of Subject over Time, in $(m / s^2)$")),
    subtitle = "Test Subject Walking at Steady Pace, Wearing a Biologging watch.",
    y = TeX("Acceleration Magnitude  $\\| \\textbf{a} \\|$  in $(m / s^2)$"),
    x = TeX("Elapsed Time (seconds)")
  ) + theme_bw() +
  scale_x_continuous(labels = 0:5, breaks = 1:6) +
  theme(panel.grid.minor = element_line(
    color = "grey90",
    linetype = "dashed",
    linewidth = 0.5
  ))
# show results
print(p3)
```



Sample ACF

```r
# get sample acf values
p5DF = data.frame(
  h = 0:36,
  rh = acf(accel_ts, plot = FALSE, lag.max = 36)$acf
)
# find n for WN bound calculation
n = length(accel_ts)
# create plot
p5 <- ggplot(p5DF, aes(x = h, y = rh)) +
  geom_hline(yintercept = 2/sqrt(n),
             linetype = "dashed",
             col = "#59054d") +
  geom_hline(yintercept = -2/sqrt(n),
             linetype = "dashed",
             col = "#59054d") +
  ylim(-0.4, 1) +
  geom_segment(aes(xend = h, yend = 0),
               color = "#b846a7",
               linewidth = 1) +
  geom_hline(yintercept = 0,
             linetype = "dashed",
             color = "#33172f") +
  labs(x = "Lag",  y = "ACF Values",
   title = TeX(paste("Correlogram of Acceleration Magnitude",
    "$\\| \\textbf{a} \\|$ of Subject")),
  subtitle = paste("Test Subject Walking at Steady Pace for appx.",
  "Five Seconds, Wearing a Biologging Watch")) +
  theme_bw()
print(p5)
```



Part A.5.

Comment on what you observe.

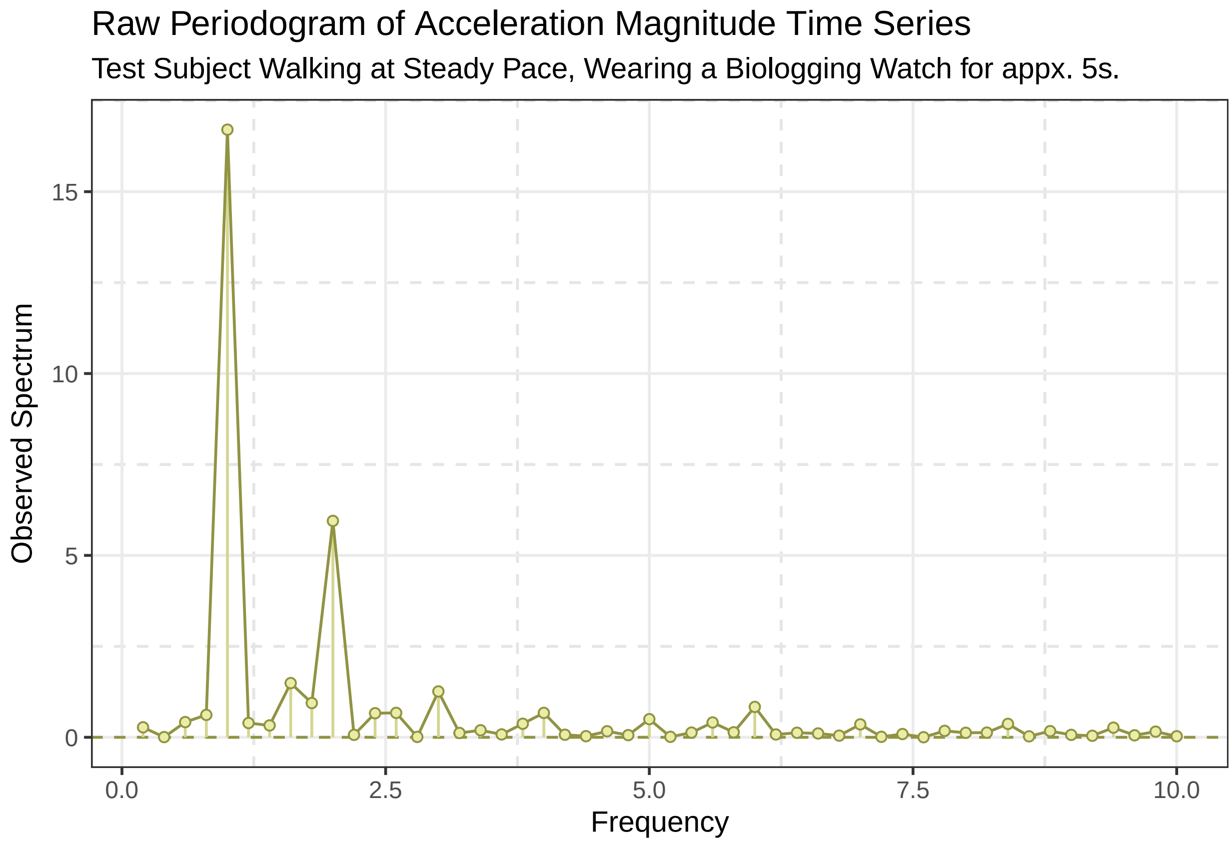
In both the Sample ACF and the plot of the time series, there is evidence of a seasonal component s_t. By this fact, we know that the series is not stationary (by the first property of weakly stochastic processes.) The ACF shows a distinct periodic component which is not increasing or decreasing exponentially in magnitude over time; this is evidence of perhaps an additive seasonal component. This would match our intuition with an individual walking while wearing a biologging watch: a potential cause of the periodicity lies with the natural swinging of the subject's arm as they walk. This back-and-forth motion would cause a periodic z-axis acceleration a_z which we would then see exemplified in a varying magnitude of acceleration. Further, in the plot of the time series itself, there may be a slight negative trend in the data, though this is difficult to discern purely from a visual standpoint. Depending on how the experiment was conducted, a negative trend would not be surprising as it is possible that the subject would begin to slow down their walking speed as the experiment drew to a close. This would then be reflected in a decrease of a_x and/or a_y, which would yield an overall decline in $\|\vec{a}\|$.

Part B: Raw Periodogram

Part B.1.

Plot the raw periodogram for the series.

```r
N = length(accel_ts)
# compute raw periodogram
raw_pdg = spec.pgram(accel_ts, log = "no", plot = F)
# coerce to data frame
p6DF = data.frame(
  Frequency = raw_pdg$freq,
  Spectrum =  raw_pdg$spec
)
# plot
p6 <- ggplot(p6DF, aes(x = Frequency, y = Spectrum)) +
  geom_hline(yintercept = 0,  linetype = "dashed", color = "#909342") +
  geom_segment(aes(
    x = Frequency,   xend = Frequency,
    y = 0,  yend = Spectrum,
  ), color = "#d4d692")+
  geom_line(color = "#909342", linewidth = 0.5) +
  geom_point(color = "#909342", fill = "#ebeda6", pch = 21) +
  labs(
    title = "Raw Periodogram of Acceleration Magnitude Time Series",
    subtitle = paste(
      "Test Subject Walking at Steady Pace,",
      "Wearing a Biologging Watch for appx. 5s."
    ),
    y = TeX("Observed Spectrum"),
    x = TeX("Frequency")
  ) +
  theme_bw() +
  theme(panel.grid.minor = element_line(
    color = "grey90",
    linetype = "dashed",
    linewidth = 0.5
  ))
print(p6)
```



Comments: The raw periodogram appears to be dominated by low frequencies; we know this is likely the case as we see two peaks in spectrum magnitude at indexes ω_5 and ω_{10}, while $I(\omega)$ tails off towards zero for higher frequencies. Recalling that the periodogram is an unbiased estimator for spectral density, from this plot we can conclude it is likely that a larger proportion of the total variation in the series is attributed to the lower frequencies in the series. Specifically, the two noted earlier (ω_5 and ω_{10}) that have large observed spectrum. It should be noted that while there may also be signs of potential harmonics, as we can see roughly equidistant declining peaks in $I(\omega)$ for $\omega_5, \omega_{10}, \omega_{15}$ and ω_{20}.

Part B.2.

To estimate the angular frequency and wavelength of the dominating frequency, we must first determine which frequency dominates the raw periodogram. Let $f^{\mathrm{p}}(\omega)$ be the periodogram spectrum. Let Ω be the set of cardinality $N/2$ containing the frequency values. We hence denote the dominating frequency as ϖ, where:

$$\varpi = \underset{\omega \in \Omega}{\operatorname{argmax}} \Big(f^{\mathrm{p}}(\omega) \Big)$$

In this case, the dominating frequency is found via the following:

```
LOC = which.max(raw_pdg$spec)
dfreq = raw_pdg$freq[LOC]
dfreq
```

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

The Angular Frequency from can be found by converting back to its corresponding Fourier Frequency (in radians.)

$$\varpi_{\otimes} = \frac{2\pi\varpi}{N} = \frac{2\pi(1)}{100} = \frac{\pi}{50}$$

Then, we recall that the wavelength is found with $1/\phi$, where ϕ is the location of the peak in spectrum. In our case, this is $\varpi = 1$. Hence, the wavelength is approximately $1/\phi = 1/\varpi = 1$ second(s).

Part C

Build a function in `R` that generates the Fourier frequency ω_p for a given time series and given constant $p \in \{0, 1, \dots N/2\}$.

We will use the equation:

$$\omega_p := \frac{2\pi p}{N}, \quad p \in \{0, 1, \dots N/2\}$$

We are tasked to "Document the inputs and outputs of this function so that another person would be able to understand how to use your function." Hence, we define the function below in Roxygen-style `R` function documentation with input validation.

```r
#' @param ts A `ts` object of length `N`, where `N` is even.
#' @param p A numeric $p \in \{0, 1, \dots N/2\}$
#' @return the Fourier Frequency `omega_p` for the given p and time series
gen_freq <- function(ts, p){
  # check that the function is being given a time series
  if (!is.ts(ts)) {
    stop("Input `ts` must be a time series object.")
  }
  # get the time series length if the above check passes
  N <- length(ts)
  # then validate that this is even
  if (N %% 2 != 0){
    warning("Detected Odd-Length ts. Using Floor...")
    N <- N - 1 # this will cause N/2 to be floor(N/2)
  }
  # check that the input p is a numeric
  if (!is.numeric(p)) {
    stop("Input `p` must be numeric.")
  }
  # check that the input p is within [0, N/2]
  if (p < 0 || p > (N / 2)) {
    stop("Input `p` must be in the range 0 to N/2.")
  }
  # check that the input p is an integer
  if (round(p) != p){
    p = round(p)
    warning("Input `p` is not an integer. Using Rounded Value")
  }
  # now that all tests pass...
  omega_p = (2*pi*p)/N
  # return p-th Fourier Frequency
  return(omega_p)
}
```

What is the output of your function for $p = 10$?

When $p = 10$, the function returns $\pi/5$.

```
gen_freq(accel_ts, 10) # == (pi / 5)
```

```
## [1] 0.6283185
```

Part D

In this question, we will fit all the fourier frequencies and determine which are significant, do some testing, and fit a finalized model.

Part D.1.

We will first compute the set of frequencies for $p \in \{0, 1, \ldots N/2\}$ using the function from the previous question.

We will denote this set Ω, where $|\Omega| = \frac{N}{2} + 1$ is defined as follows, where ω_p is the p-th set element:

$$\Omega = \left\{ \frac{2\pi p}{N} : p \in \mathbb{Z},\, 0 \leq p \leq \frac{N}{2} \right\}, \text{ hence } \forall \omega_p \in \Omega,\ \omega_p \in [0, \pi] \subseteq \mathbb{R}$$

```
obsv_w = sapply(0:(N/2), gen_freq, ts = accel_ts)
```

Part D.2.

Then, we will fit a model of the following one-at-a-time $\forall \omega_p \in \Omega$ where X_t is the magnitude of acceleration at index t:

$$X_t = a_0 + a_p \cos(\omega_p \cdot t) + b_p \sin(\omega_p \cdot t) + \varepsilon,\ \text{where } \varepsilon \overset{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2) \text{ and } t \in \{1, 2, \ldots, N\}$$

We can equivalently formulate this model as:

$$X_t^p = a_0 + \sum_{t=1}^{N} \Big(a_p \cos(\omega_p \cdot t) + b_p \sin(\omega_p \cdot t) \Big) + \varepsilon,\ \text{ where } \varepsilon \overset{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2)$$

Where the superscript p indicates the Fourier Frequency $\omega_p \in \Omega$.

Hence, the p-th model fitted will have the form:

$$\widehat{X}_t^p = \hat{a}_0 + \sum_{t=1}^{N} \Big(\hat{a}_p \cos(\omega_p \cdot t) + \widehat{b}_p \sin(\omega_p \cdot t) \Big),\ \text{ where } \omega_p \in \Omega$$

We hence let $\mathfrak{X}$ be the set of all models fitted such that the p-th entry of $\mathfrak{X}$ is the model fitted $\widehat{X}_t^p$, i.e. $\mathfrak{X} = \left\{ \widehat{X}_t^{\,0}, \widehat{X}_t^{\,1}, \ldots, \widehat{X}_t^{\,N/2} \right\}$. Then, finally, we compute the F-statistic for each model in $\mathfrak{X}$, creating the following set of statistics, letting $F(\widehat{X})$ be the F-statistic of model $\widehat{X}$:

$$\mathcal{F} : \left\{ F\big(\, \widehat{X}_t^{\,p} \,\big) : \widehat{X}_t^{\,p} \in \mathfrak{X} \right\},\ \text{ where } \forall F \in \mathcal{F},\ F \sim \mathrm{F}_{k-1, N-k}$$

Where $k = 3$ and $N = 100$. In each instance, the F-statistic is computed by the standard ratio of sum of squares, i.e. for arbitrary $\widehat{X}_t^{\,p}$, we have that

$$F : \mathbb{R}^N \mapsto \mathbb{R}, \text{ s.t. } F\big(\,\widehat{X}_t^{\,p}\,\big) = \frac{\sum_{t=1}^{N}\big(\,\widehat{X}_t^{\,p} - \overline{X}\,\big)^2/(k-1)}{\sum_{t=1}^{N}\big(\,X_t - \widehat{X}_t^{\,p}\,\big)^2/(N-k)}$$

Where $\widehat{X}_t^{\,p}$ are the fitted values at time t, $\overline{X}$ is the sample mean of the series and X_t is the observed value at time t.

Then, after all of this discussion, the whole procedure can be done directly as follows:

```
bigF = unlist(sapply(obsv_w, function(w){
  summary(lm(df$Mag ~ cos(w*(1:N)) + sin(w*(1:N))))$fstatistic[[1]]
}))
```

It should be clarified that each test statistic was against the model fit hypothesis, in other words:

$$H_0 : a_p = b_p = 0 \quad \text{against} \quad H_A : \text{ at least one of } \{a_p, b_p\} \neq 0$$

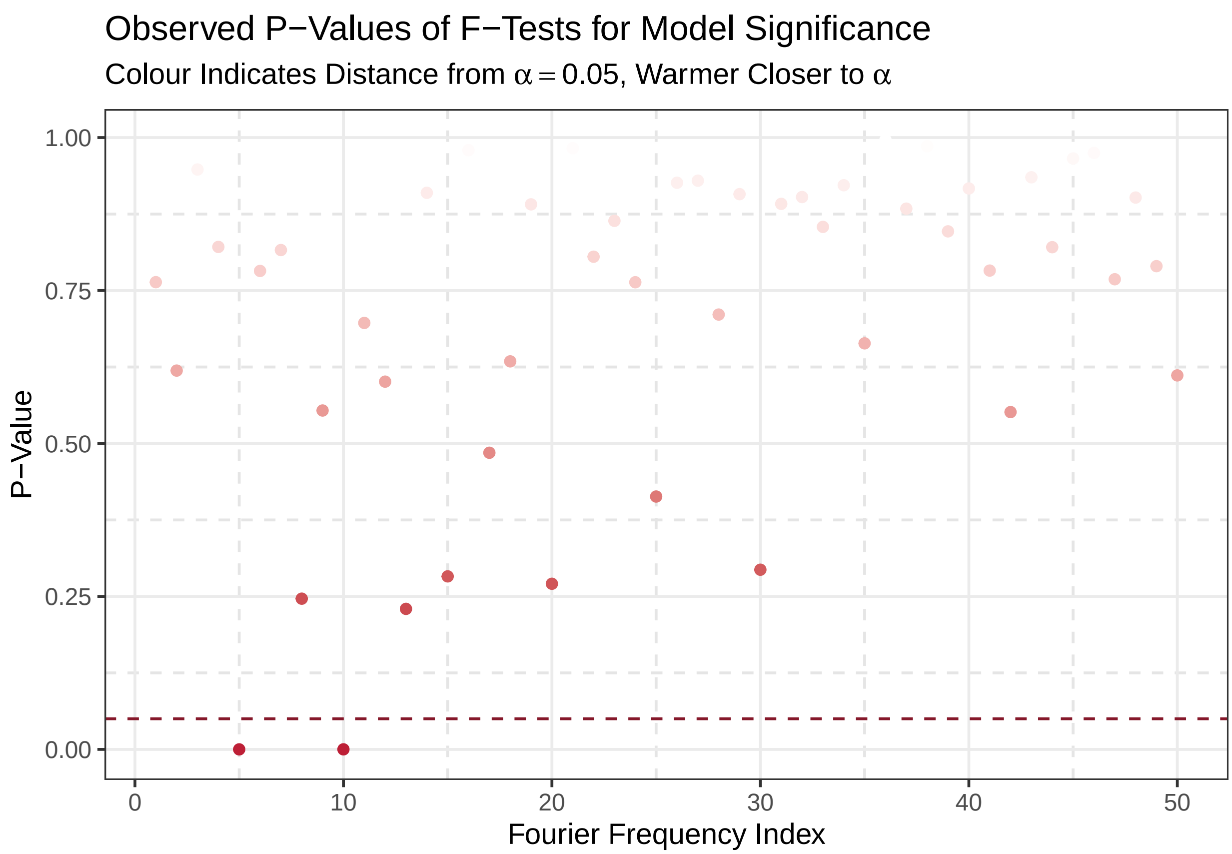
Thus, the set of p-values can be defined as follows:

$$\mathcal{P} : \Big\{\mathbb{P}\big(\,\mathrm{F}_{k-1,N-k} > F_p\big) : F_p \in \mathcal{F}\Big\}, \text{ hence } \forall p \in \mathcal{P},\ p \in (0,1) \subseteq \mathbb{R}$$

We compute and plot the p-values in $\mathcal{P}$ below. We omit the code for creating the `ggplot` object.

```
# declare K as noted above
k = 3
# compute P set as defined above, in a data frame for plotting
p6df = data.frame(
  pval = pf(bigF, k - 1, N - k, lower.tail = FALSE),
  index = 1:length(bigF))
```

```
print(p6)
```



We then extract the index of the significant coefficients. These correspond to ω_5 and ω_{10} (as shown below) which makes sense given the peaks in the original raw periodogram. We add `1` to the indexing of `signif_ws` to adjust for the list starting from ω_0 (which has an `NA` p-value in the regression model.)

```
signif_ws = obsv_w[ p6df$index[p6df$pval < 0.05] + 1 ]
# here, we verify that the significant ones correspond to w_5 and w_10
signif_ws == sapply(c(5, 10), gen_freq, ts = accel_ts)
```

```
## [1] TRUE TRUE
```

Part E

Give the estimated coefficients for the linear model that results from using all significant frequencies found in part **(d)**

We define the unified model using both ω_5 and ω_{10}. We will hence fit a model of the following form:

$$\widehat{X}_{\text{unified}} = \hat{a}_0 + \sum_{t=1}^{N} \Big(\hat{a}_5 \cos(\omega_5 \cdot t) + \widehat{b}_5 \sin(\omega_5 \cdot t) \Big) + \sum_{t=1}^{N} \Big(\hat{a}_{10} \cos(\omega_{10} \cdot t) + \widehat{b}_{10} \sin(\omega_{10} \cdot t) \Big)$$

Which can be simplified to:

$$\widehat{X}_{\text{unified}} = \hat{a}_0 + \sum_{t=1}^{N} \Big(\hat{a}_5 \cos(\omega_5 \cdot t) + \widehat{b}_5 \sin(\omega_5 \cdot t) + \hat{a}_{10} \cos(\omega_{10} \cdot t) + \widehat{b}_{10} \sin(\omega_{10} \cdot t) \Big)$$

We compute the unified model using both ω_5 and ω_{10} below.

```
unified_model = lm(df$Mag ~ (
    cos(signif_ws[1] * (1:N)) + sin(signif_ws[1] * (1:N)) +
    cos(signif_ws[2] * (1:N)) + sin(signif_ws[2] * (1:N))
))
```

We then report coefficients $\{\hat{a}_0, \hat{a}_5, \hat{b}_5, \hat{a}_{10}, \widehat{b}_{10}\}$ in the table below, to three decimal places.

```
# round and transform
un_coefs = t(data.frame(as.numeric(round(coefficients(unified_model), 3))))
rownames(un_coefs) = "Estimate"
# format in a nice table
kable(un_coefs, "latex", escape = FALSE,
      col.names = c("$\\hat{a}_0$", "$\\hat{a}_5$", "$\\hat{b}_5$",
                    "$\\hat{a}_{10}$", "$\\hat{b}_{10}$"),
  caption = "Fitted Coefficients for Significant Frequencies")  %>%
  kable_styling(latex_options = "hold_position") %>%
  kable_styling(position = "center")
```

Table 2: Fitted Coefficients for Significant Frequencies

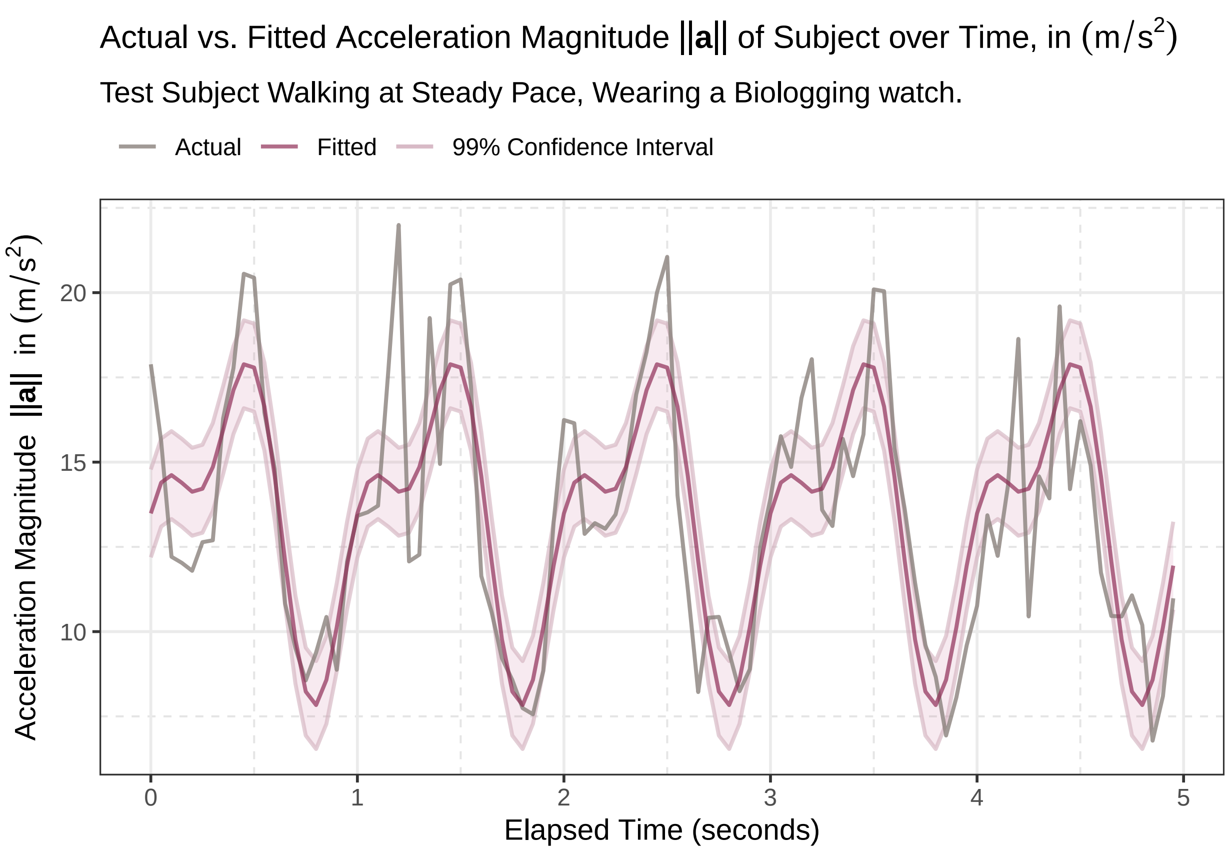
	$\hat{a}_0$	$\hat{a}_5$	$\hat{b}_5$	$\hat{a}_{10}$	$\hat{b}_{10}$
Estimate	13.431	-2.97	2.181	1.485	1.711

Part F

Below, we plot X_t alongside $\widehat{X}_{\text{unified}}$ for the training data. We also add a 99% confidence interval for the in-sample fit, to have a rough quantification of the uncertainty inherent in the fitted unified model.

```
model_fit = as.data.frame(predict(unified_model, interval = "confidence",
                                  level = 0.99))
# format as data frame
p7DF = data.frame(
  Time = time(accel_ts),
  Actual = as.numeric(accel_ts),
  Predicted = as.numeric(model_fit$fit),
  Lower = as.numeric(model_fit$lwr),
  Upper = as.numeric(model_fit$upr)
)
# create plot
p7 = ggplot(data = p7DF, aes(x = Time)) +
  geom_ribbon(aes(ymin = Lower, ymax = Upper), fill = "#c97598", alpha = 0.15) +
  # lines for obsv, pred and interval bounds
  geom_line(aes(y = Lower, color = "99% Confidence Interval"),  lwd = 0.7, alpha = 0.6) +
  geom_line(aes(y = Upper, color = "99% Confidence Interval"), lwd = 0.7, alpha = 0.6) +
  geom_line(aes(y = Actual, color = "Actual"),  lwd = 0.7, alpha = 0.8) +
  geom_line(aes(y = Predicted, color = "Fitted"), lwd = 0.7, alpha = 0.7) +
  # legend and colour assignment
  scale_color_manual(name = "",
    values = c(
    "Actual" = "#8a817c",
```

```
    "Fitted" = "#8f2d56",
    "99% Confidence Interval" = "#d1aebc"
  ),limits = c("Actual", "Fitted", "99% Confidence Interval")) +
  # add titles with units
  labs(
    title = TeX(paste("Actual vs. Fitted Acceleration Magnitude",
    "$\\| \\textbf{a} \\|$ of Subject over Time, in $(m / s^2)$")),
    subtitle = "Test Subject Walking at Steady Pace, Wearing a Biologging watch.",
    y = TeX("Acceleration Magnitude  $\\| \\textbf{a} \\|$  in $(m / s^2)$"),
    x = TeX("Elapsed Time (seconds)")
  ) +
  scale_x_continuous(labels = 0:5, breaks = 1:6) +
  theme_bw() +
  theme(panel.grid.minor =  element_line(
    color = "grey90",
    linewidth  = 0.35,
    linetype = "dashed"
  ), legend.position = "top", legend.justification = "left",
  legend.margin = margin(0,0,0,0),
  plot.title = element_text(size = 12 ))
print(p7)
```



Comments: We see that the fitted model using the significant frequencies does a decent job at capturing the pattern of the original series X_t; however, it is far from the performance we would expect from a more formal fit/forecast method such as those discussed in the previous chapters. Considering that the 99% confidence interval does not capture a significant portion of the true series, it is very likely that our fitted $\widehat{X}_t$

is not entirely specified for these data. This makes sense considering that we are only using two frequencies, so it is very unlikely we would fully capture the complexity of the data from this small subset of ω values. Also, we are fitting the model solely using the deterministic sequence of time indices. This being said, the fact that the general shape and pattern of the data are well-encapsulated by the model fit by ω_5 and ω_{10} help to support our previous findings (i.e. from the periodogram) of the significance of these frequencies to the variation of the overall time series.

Part G

Approximately how many steps per minute was the subject taking?

We can derive a very informal approximation for the steps per minute by observing the number of complete cycles ϕ the fitted $\widehat{X}_t$ makes in a 1-second time frame. As commented earlier, the general shape of $\widehat{X}_t$ closely follows X_t, so it can serve nicely as an approximation for the number of cycles. We assume that the periodicity in the time series are caused by the steps of the individual changing the magnitude of acceleration in a cyclical pattern. This isn't an egregious assumption, as we know the individual was wearing a biologging watch. Since most people swing their arms while they walk (which forms a sinusoidal pattern as seen in the series) it's not unlikely that the individual's arm completes one "swing" every step they take. In the given time frame of 5 seconds, we observe approximately 5 complete cycles. Hence, we can estimate that in $6 \times 12 = 60$ seconds the subject would take roughly $6 \times 12 = 60$ steps, which is in-line with a slow walking pace.