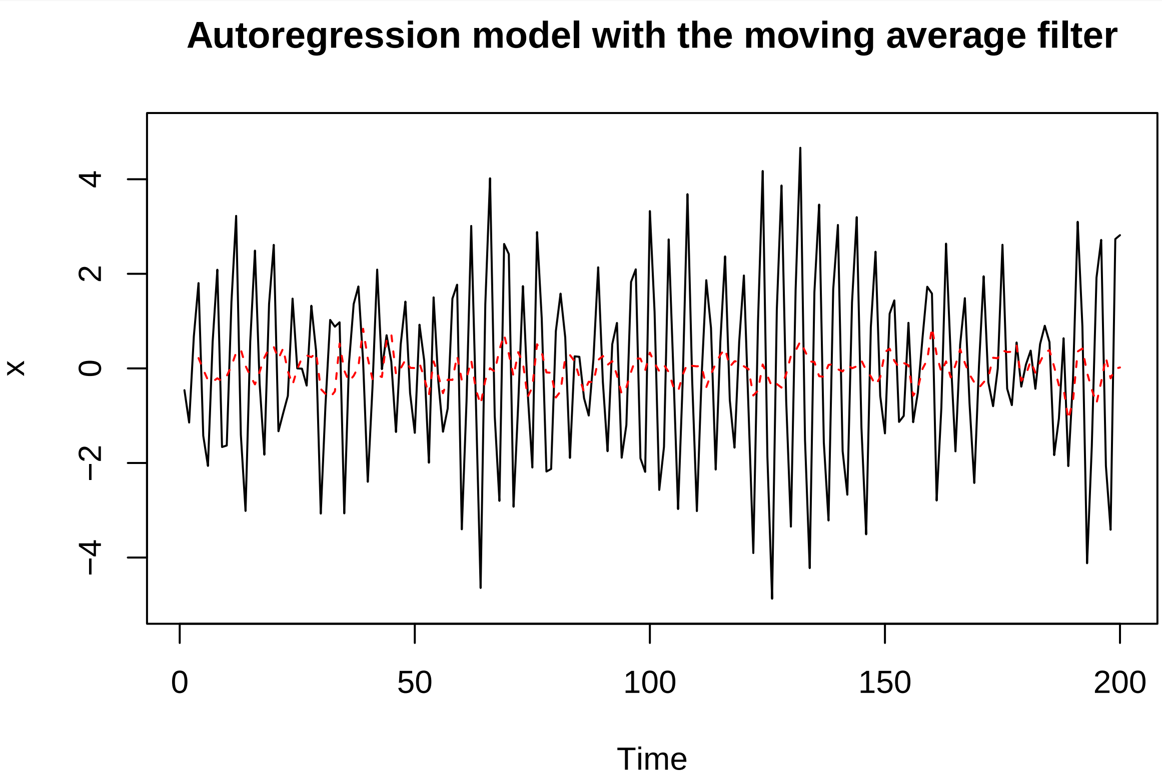
STA2202H1: Time Series Analysis Assignment 1

Student Name: Yulin Wang ID Number: 1003942326

2022-09-23

Q1

```
set.seed(100)
# Simulate 200 observations from N(0,1); Gaussian white noise
w <- rnorm(220, 0, 1) # 20 extra to avoid startup problems
# Generate time series x_t using the second-order autoregression
x = filter(w, filter=c(0, -0.9), method="recursive")[-(1:20)] # remove the first 20
# Apply the moving average filter
v = filter(x, sides=1, filter=rep(1/4, 4)) # sides=1 for past values
# Plot x_t as a line
plot.ts(x, ylim=c(-5, 5), main="Autoregression model with the moving average filter")
# Superimpose the moving average filter v as a red dashed line
lines(v, col='red', lty=2)
```



Comment: After applying the moving average filter v_t, the generated time series data tends to be smooth with a relative constant variance by getting rid of the noise in the series.

Q2

(1) Show that the model can be written as $x_t = \delta t + \sum_{j=1}^{t} w_j$.

$$
\begin{aligned}
x_t &= \delta + x_{t-1} + w_t \\
&= \delta + (\delta + x_{t-2} + w_{t-1}) + w_t \\
&= 2\delta + x_{t-2} + w_{t-1} + w_t \\
&= 2\delta + (\delta + x_{t-3} + w_{t-2}) + w_{t-1} + w_t \\
&= 3\delta + x_{t-3} + w_{t-2} + w_{t-1} + w_t \\
&... \\
&= \delta t + x_0 + (w_1 + w_2 + ... + w_{t-1} + w_t) \quad \text{where } x_0 = 0 \\
&= \delta t + \sum_{j=1}^{t} w_j
\end{aligned}
$$

(2) Find the mean function and the autocovariance function of x_t.

The mean function of x_t

$$
\begin{aligned}
\mu_{x_t} = E(x_t) &= E(\delta t + \sum_{j=1}^{t} w_j) \\
&= E(\delta t) + \sum_{j=1}^{t} E(w_j) \\
&= \delta t + 0 \\
&= \delta t
\end{aligned}
$$

The autocovariance function of x_t

$$
\begin{aligned}
\gamma_x(s,t) &= Cov(x_s, x_t) \\
&= Cov(\delta s + \sum_{j=1}^{s} w_j, \delta t + \sum_{k=1}^{t} w_k) \\
&= min\{s,t\}\sigma_w^2 \qquad \text{because } w_t \text{ are uncorrelated random variables}
\end{aligned}
$$

(3) Show that x_t is not stationary.

Check condition 1: The expected value of x_t is $E(x_t) = \delta t$ from part (1), which is independent of time t only if there is no drift.

Check condition 2: WLOG, we can consider drift $\delta = 0$, then $x_t = \sum_{j=1}^{t} w_j$. The autocovaraince function is

$$
\gamma(h) = Cov(x_t, x_{t+h}) = Cov(\sum_{j=1}^{t} w_j, \sum_{k=1}^{t+h} w_k) = min\{t, t+h\} \cdot \sigma_w^2 = t\sigma_w^2
$$

, which is also dependent of time t.

Hence, x_t is not stationary.

(4) Show that $\rho_x(t-1,t) = \sqrt{\frac{t-1}{t}} \to 1$ as $t \to \infty$.

$$\begin{aligned}
\rho_x(t-1,t) &= \frac{\gamma(t-1,t)}{\sqrt{\gamma(t-1,t-1)\gamma(t,t)}} \\
&= \frac{(t-1)\sigma_w^2}{\sqrt{(t-1)\sigma_w^2 \cdot t\sigma_w^2}} \\
&= \frac{(t-1)}{\sqrt{(t-1)\cdot t}} \\
&= \sqrt{\frac{(t-1)^2}{(t-1)\cdot t}} \\
&= \sqrt{\frac{t-1}{t}} = \sqrt{1-\frac{1}{t}}
\end{aligned}$$

Then, $lim_{t\to\infty}\{\rho_x(t-1,t)\} = lim_{t\to\infty}\{\sqrt{1-\frac{1}{t}}\} = \sqrt{1-0} = 1$.

That is, for large time t with lag h considerably less than t, $\rho_x(t-1,t)$ is nearly 1. This implies that the correlogram for this random walk with drift is characterised by positive autocorrelations with very slow decay down from unity.

(5) Suggest a transformation to make the series stationary, and prove that the transformed series is stationary.

We can take the first difference of x_t. Then,

$$\begin{aligned}
z_t &= x_t - x_{t-1} \\
&= \delta + x_{t-1} + w_t - x_{t-1} \\
&= \delta + w_t
\end{aligned}$$

Thus, the mean of z_t

$$\begin{aligned}
E(z_t) &= E\{\delta + w_t\} \\
&= E(\delta) + E(w_t) \\
&= \delta + 0 \\
&= \delta
\end{aligned}$$

, which is independent of t.

And the autocovariance function of z_t

$$\begin{aligned}
\gamma(h) &= Cov(z_t, z_{t+h}) \\
&= Cov(\delta + w_t, \delta + w_{t+h}) \\
&= \begin{cases} \sigma_w^2 & h = 0 \\ 0 & h \neq 0 \end{cases} \quad \text{because } w_t \text{ are uncorrelated random variables}
\end{aligned}$$

, which is also independent of t.

Hence, the transformed series z_t is stationary.

Q3

(1) Compute the sample autocorrelation function, $\hat{\rho}(h)$, at lags $h = 0$, 1, 2, and 3.

Here is a table of t, x_t, x_{t+1}, x_{t+2}, x_{t+3}.

t	x_t	x_t_1	x_t_2	x_t_3
1	24	20	25	31
2	20	25	31	30
3	25	31	30	32
4	31	30	32	37
5	30	32	37	33
6	32	37	33	40
7	37	33	40	38
8	33	40	38	NA
9	40	38	NA	NA
10	38	NA	NA	NA

The sample mean of these ten x_t values is $\bar{x} = 31$. Thus,

$$
\begin{aligned}
\hat{\rho}(0) &= \frac{\hat{\gamma}(0)}{\hat{\gamma}(0)} = 1 \\
\hat{\rho}(1) &= \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} = \frac{(24-31)(20-31)+(20-31)(25-31)+...+(33-31)(40-31)+(40-31)(38-31)}{(24-31)^2+(20-31)^2+...+(40-31)^2+(38-31)^2} \\
&= \frac{241}{378} \approx 0.638 \\
\hat{\rho}(2) &= \frac{\hat{\gamma}(2)}{\hat{\gamma}(0)} = \frac{(24-31)(25-31)+(20-31)(31-31)+...+(37-31)(40-31)+(33-31)(38-31)}{378} \\
&= \frac{112}{378} \approx 0.296 \\
\hat{\rho}(3) &= \frac{\hat{\gamma}(3)}{\hat{\gamma}(0)} = \frac{(24-31)(31-31)+(20-31)(30-31)+...+(32-31)(40-31)+(37-31)(38-31)}{378} \\
&= \frac{54}{378} \approx 0.143
\end{aligned}
$$

(2) Test the null hypothesis.

Test the null hypothesis $H_0 : \rho(1) = 0$ versus the alternative hypothesis $H_a : \rho(1) \neq 0$.

The t-test statistic is

$$t_{\rho(1)} = \frac{\hat{\rho}(1)}{\sigma_{\hat{\rho}(1)}} = \frac{0.638}{1/\sqrt{10}} \approx 2.018$$

, which is greater than $z_{0.025} = 1.96 \approx 2$. Thus, we reject $H_0 : \rho(1) = 0$ at a 5% level of significance.

(3) Use R and redo parts (1) and (2).

```r
x <- c(24, 20, 25, 31, 30, 32, 37, 33, 40, 38)
# Compute the sample autocorrelation function
acf(x, lag.max=3, plot=FALSE)
```

```
##
```

```
## Autocorrelations of series 'x', by lag
##
##     0     1     2     3
## 1.000 0.638 0.296 0.143
```

```r
# Test the null hypothesis
n = 10 # sample size
t_statistic = 0.638/(1/sqrt(n)) # use 1/sqrt(n) to estimate the standard deviation
if (t_statistic > 1.96) {
  print("Reject the null hypothesis at a significance level of 5%")
} else {
  print("Do NOT reject the null hypothesis at a significance level of 5%")
}
```

```
## [1] "Reject the null hypothesis at a significance level of 5%"
```

Q4

The autocovariance function of x_t is

$$
\begin{aligned}
\gamma(h) &= Cov(x_t, x_{t+h}) \\
&= Cov\{\frac{1}{3}(w_{t-1} + w_t + w_{t+1}), \frac{1}{3}(w_{t+h-1} + w_{t+h} + w_{t+h+1})\} \\
&= \frac{1}{9}Cov\{(w_{t-1} + w_t + w_{t+1}), (w_{t+h-1} + w_{t+h} + w_{t+h+1})\} \\
&= \begin{cases} \frac{3}{9}\sigma_w^2 & h = 0 \\ \frac{2}{9}\sigma_w^2 & h = \pm 1 \\ \frac{1}{9}\sigma_w^2 & h = \pm 2 \\ 0 & |h| > 2 \end{cases}
\end{aligned}
$$

Then the actual autocorrelation function ACF is

$$
\rho(h) = \begin{cases} 1 & h = 0 \\ \frac{2}{3} & h = \pm 1 \\ \frac{1}{3} & h = \pm 2 \\ 0 & |h| > 2 \end{cases}
$$

(1) Simulation with $n = 500$.

```
set.seed(100)
# Simulate 500 observations from N(0,1); Gaussian white noise
w <- rnorm(502, 0, 1) # extra 2 to avoid startup problems
# Apply the moving average filter
x = filter(w, sides=2, filter=rep(1/3, 3))
x = head(x[-1], -1) # remove the first and last NA element
# Compute the sample ACF
acf(x, lag.max=20, plot=FALSE)
```

```
##
## Autocorrelations of series 'x', by lag
##
##      0      1      2      3      4      5      6      7      8      9     10
##  1.000  0.614  0.273 -0.103 -0.108 -0.125 -0.041  0.036  0.089  0.005 -0.025
##     11     12     13     14     15     16     17     18     19     20
## -0.079 -0.088 -0.110 -0.072 -0.030 -0.022 -0.063 -0.054 -0.024  0.053
```

```
#acf(x, lag.max=20, plot=TRUE)
```

Comparing the sample ACF to the actual ACF, we can find that the $\hat{\rho}(1) = 0.614$ and $\hat{\rho}(2) = 0.273$ are close to the actual ACF values $\frac{2}{3}$ and $\frac{1}{3}$, and the sample ACF values for lag h greater than 2 are around the actual ACF value of 0.

(2) Simulation with $n = 50$.

```
set.seed(100)
# Simulate 50 observations from N(0,1); Gaussian white noise
w <- rnorm(52, 0, 1) # extra 2 to avoid startup problems
# Apply the moving average filter
x = filter(w, sides=2, filter=rep(1/3, 3))
x = head(x[-1], -1) # remove the first and last NA element
# Compute the sample ACF
acf(x, lag.max=20, plot=FALSE)
```

```
##
## Autocorrelations of series 'x', by lag
##
##      0      1      2      3      4      5      6      7      8      9     10
##  1.000  0.416  0.237 -0.359 -0.290 -0.340 -0.031  0.123  0.258  0.048 -0.078
##     11     12     13     14     15     16     17     18     19     20
## -0.385 -0.281 -0.366 -0.010  0.016  0.278  0.137  0.140 -0.033  0.032
```

```r
#acf(x, lag.max=20, plot=TRUE)
```

As the sample ACF results from the above two simulations show, when the number of observations n increases, the sample ACF values tend to be closer to the actual ACF values, that is, the empirical values tend to be closer to the theoretical values.

Q5

(1) Show that this process is weak stationary.

$$
\begin{aligned}
x_t &= cos[2\pi(\frac{t}{12} + \phi)] \\
&= cos(\frac{\pi t}{6} + 2\pi\phi) \\
&= cos(\frac{\pi t}{6})cos(2\pi\phi) - sin(\frac{\pi t}{6})sin(2\pi\phi)
\end{aligned}
$$

Then, the mean

$$
\begin{aligned}
E(x_t) &= cos(\frac{\pi t}{6}) \cdot E[cos(2\pi\phi)] - sin(\frac{\pi t}{6}) \cdot E[sin(2\pi\phi)] \\
&= cos(\frac{\pi t}{6}) \cdot \int_0^1 [cos(2\pi\phi)]d\phi - sin(\frac{\pi t}{6}) \cdot \int_0^1 [sin(2\pi\phi)]d\phi \\
&= cos(\frac{\pi t}{6}) \cdot \frac{1}{2\pi}[sin(u)|_0^{2\pi}] - sin(\frac{\pi t}{6}) \cdot \frac{1}{2\pi}[-cos(u)|_0^{2\pi}] \\
&= cos(\frac{\pi t}{6}) \cdot 0 - sin(\frac{\pi t}{6}) \cdot 0 \\
&= 0
\end{aligned}
$$

, which is independent of t.

And the autocovariance function

$$
\begin{aligned}
\gamma(h) = Cov(x_t, x_{t+h}) &= E(x_t x_{t+h}) \\
&= E\{cos(\frac{\pi t}{6} + 2\pi\phi) \cdot cos(\frac{\pi(t+h)}{6} + 2\pi\phi)\} \\
&= E\{\frac{1}{2} \cdot cos[\frac{\pi t}{6} + 2\pi\phi - \frac{\pi(t+h)}{6} - 2\pi\phi] + \frac{1}{2} \cdot cos[\frac{\pi t}{6} + 2\pi\phi + \frac{\pi(t+h)}{6} + 2\pi\phi]\} \\
&= \frac{1}{2}E\{cos(-\frac{\pi h}{6}) + cos[\frac{2\pi t + \pi h}{6} + 4\pi\phi]\} \\
&= \frac{1}{2}cos(\frac{\pi h}{6}) + \frac{1}{2}E\{cos[\frac{2\pi t + \pi h}{6} + 4\pi\phi]\} \\
&= \frac{1}{2}cos(\frac{\pi h}{6}) + \frac{1}{2}E\{cos(\frac{2\pi t + \pi h}{6})cos(4\pi\phi) - sin(\frac{2\pi t + \pi h}{6})sin(4\pi\phi)\} \\
&= \frac{1}{2}cos(\frac{\pi h}{6}) + \frac{1}{2}\{cos(\frac{2\pi t + \pi h}{6}) \cdot \int_0^1 [cos(4\pi\phi)]d\phi - sin(\frac{2\pi t + \pi h}{6}) \cdot \int_0^1 [sin(4\pi\phi)]d\phi\} \\
&= \frac{1}{2}cos(\frac{\pi h}{6}) + \frac{1}{2}\{cos(\frac{2\pi t + \pi h}{6}) \cdot 0 - sin(\frac{2\pi t + \pi h}{6}) \cdot 0\} \\
&= \frac{1}{2}cos(\frac{\pi h}{6}) + 0 \\
&= \frac{1}{2}cos(\frac{\pi h}{6})
\end{aligned}
$$

, which is independent of t.

Hence, the process x_t is weak stationary.

(2) Find its autocorrelation function.

From part (1), we can have $\gamma(0) = \frac{1}{2}cos(0) = \frac{1}{2}$.

Then, the autocorrelation function is

$$
\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\frac{1}{2}cos(\frac{\pi h}{6})}{\frac{1}{2}} = cos(\frac{\pi h}{6})
$$