Homework 3
STAT 5511
Charles R. Doss
Solution

The usual formatting rules:

- Your homework (HW) should be formatted to be easily readable by the grader.
- You may use knitr or Sweave in general to produce the code portions of the HW. However, the output from knitr/Sweave that you include should *be only what is necessary to answer the question*, rather than just any automatic output that R produces. (You may thus need to avoid using default R functions if they output too much unnecessary material, and/or should make use of `invisible()` or `capture.output()`.)
  - For example: for output from regression, the main things we would want to see are the estimates for each coefficient (with appropriate labels of course) together with the computed OLS/linear regression standard errors and p-values. If other output is not needed to answer the question, it should be suppressed!
- Code snippets that directly answer the questions can be included in your main homework document; ideally these should be preceded by comments or text at least explaining what question they are answering. Extra code can be placed in an appendix.
- All plots produced in R should have appropriate labels on the axes as well as titles. Any plot should have explanation of what is being plotted given clearly in the accompanying text.
- Plots and figures should be *appropriately sized,* meaning they should not be too large, so that the page length is not too long. (The arguments fig.height and fig.width to knitr chunks can achieve this.)
- **Directions for "by-hand" problems:** In general, credit is given for (correct) shown work, not for final answers; so show **all** work for each problem and explain your answer fully.

Questions:

1. ARMA models: Several ARMA models are written below. You can assume in all cases that $W_t \overset{\text{iid}}{\sim} N(0,1)$. For each of the ARMA models, find the roots of the AR and MA polynomials. Identify any parameter redundancy: find the values of p and q for which each model is ARMA(p,q) and write the model in its correct (non-redundant) form. Determine whether each model is causal, and determine whether it is invertible.

   (a) $X_t + 0.81X_{t-2} = W_t + \frac{1}{3}W_{t-1}$
   (b) $X_t - X_{t-1} = W_t - \frac{1}{2}W_{t-1} - \frac{1}{2}W_{t-2}$
   (c) $X_t - 3X_{t-1} = W_t + 2W_{t-1} - 8W_{t-2}$
   (d) $X_t - 2X_{t-1} + 2X_{t-2} = W_t - \frac{8}{9}W_{t-1}$
   (e) $X_t - 4X_{t-2} = W_t - W_{t-1} + \frac{1}{2}W_{t-2}$
   (f) $X_t - \frac{9}{4}X_{t-1} - \frac{9}{4}X_{t-2} = W_t$

   Solution:

   (a) $\phi(z) = 1 + 0.81z^2 = (0.9z - i)(0.9z + i)$, $\theta(z) = 1 + 1/3z$. Let $\phi(z) = 0$, we get $z = i/0.9$, and $z = -i/0.9$. Let $\theta(z) = 0$, we get $z = -3$. Since they are outside the unit circle, the model is invertible and causal. ARMA(2,1).
   (b) $\phi(z) = 1 - z$, $\theta(z) = 1 - 1/2z - 1/2z^2 - 1/2z^3 = 1/2(1 - z)(2 + z)$. Let $\phi(z) = 0$, we have $z = 1$, let $\theta(z) = 0$, we have $z = 1$ or $z = -2$. Write the model as $(1 - B)X_t = (1 - B)(1 + B/2)W_t$ and cancel $(1 - B)$, we have $X_t = W_t + 1/2W_{t-1}$. Since $|-2| > 1$, the model is invertible and causal. MA(1).
   (c) $\phi(z) = 1 - 3z$, $\theta(z) = 1 + 2z - 8z^2 = (1 + 4z)(1 - 2z)$. Let $\phi(z) = 0$, we have $z = 1/3$, since it is inside the unit circle, the model is not causal. Let $\theta(z) = 0$, we have $z = -1/4$ or $z = 1/2$. Since they are inside the unit circle, the model is not invertible. ARMA(1,2).
   (d) $\phi(z) = 1 - 2z + 2z^2$, $\theta(z) = 1 - 8/9z$, let $\phi(z) = 0$, we have $z = (1 + i)/2$ or $z = (1 - i)/2$. Since they are inside the unit circle, the model is not causal. Let $\theta(z) = 0$, we have $z = 9/8$, which is outside the unit circle. So the model is invertibel. ARMA(2,1).

(e) $\phi(z) = 1 - 4z^2$, $\theta(z) = 1 - z + 1/2z^2$, let $\phi(z) = 0$, we have $z = 1/2$ or $z = -1/2$. Since they are inside the unit circle, the model is not causal. Let $\theta(z) = 0$, we have $z = 1 + i$ or $z = 1 - i$, which are outside the unit circle. So the model is invertible. ARMA(2,2).

(f) $\phi(z) = 1 - 9/4z - 9/4z^2$, $\theta(z) = 1$, let$\phi(z) = 0$, we have $z = 1/3$ or $z = -4/3$, since $z = 1/3$ is inside the unit circle, the model is not causal. Let $\theta(z) = 0$, we have no solution. So the model is invertible. AR(2).

2. Linear representation of ARMA: For those models of Question 1 that are causal, compute theoretically the first five coefficients $\psi_0, \ldots, \psi_4$ in the causal linear process representation $X_t = \sum_{j=0}^{\infty} \psi_j W_{t-j}$

Solution:

Based on the results in Q1, we have known that models (a) and (b) are causal.

For the model (a),

$$
\begin{aligned}
\psi(z) = \frac{\theta(z)}{\phi(z)} &= \frac{1 + \frac{1}{3}z}{1 + 0.81z^2} \\
&= \frac{1 + \frac{1}{3}z}{1 - 0.81(zi)^2} \\
&= (1 + \frac{1}{3}z)(1 + (0.9zi)^2 + (0.9zi)^4 + (0.9zi)^6 + \cdots)
\end{aligned}
$$

Hence we have $\psi_0 = 1$, $\psi_1 = \frac{1}{3}$, $\psi_2 = -0.81$, $\psi_3 = -0.27$ and $\psi_4 = 0.6561$.

For the model (b),

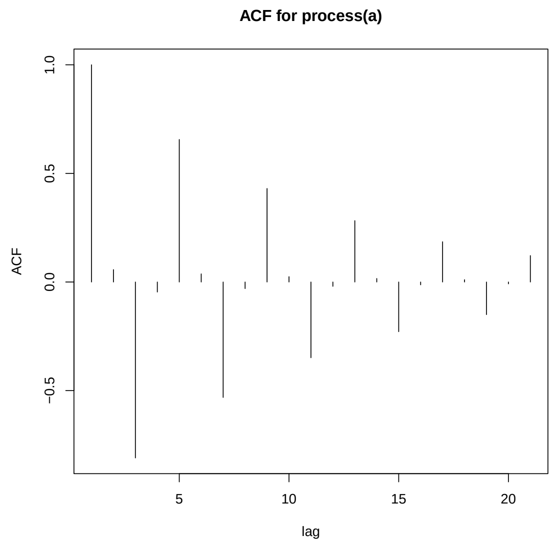
$$
\psi(z) = \frac{\theta(z)}{\phi(z)} = 1 + \frac{1}{2}z
$$

Hence we have $\psi_0 = 1$, $\psi_1 = \frac{1}{2}$, $\psi_2 = 0$, $\psi_3 = 0$ and $\psi_4 = 0$.

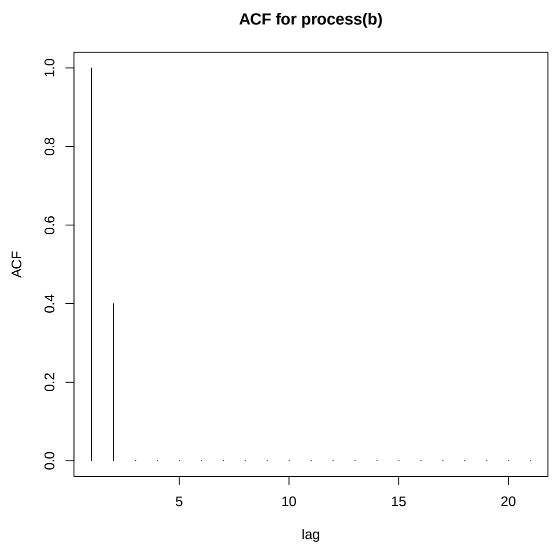
3. Autocorrelation function of ARMA model: For the models in parts (a) and (b) of Question 1: Simulate 100 observations from each model. Compute and plot the sample ACF together with the theoretical ACF. You can have R compute the theoretical ACF using the ARMAacf function.

Solution:

(1)
```r
ACFa<-ARMAacf(ar=c(0,-0.81),ma=c(1/3),20)
plot(ACFa,type="h",xlab = "lag",ylab = "ACF",main="ACF for process(a)")
```

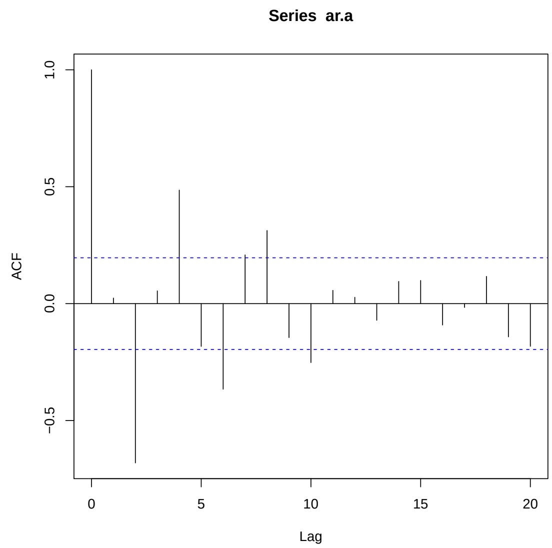



```
ACFb<-ARMAacf(ar=c(0),ma=c(1/2),20)
plot(ACFb,type="h",xlab = "lag",ylab = "ACF",main="ACF for process(b)")
```

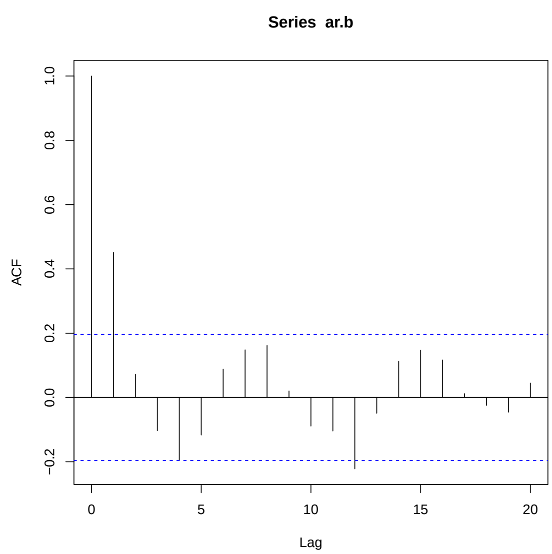


(2)

```
set.seed(20)
ar.a<-arima.sim(list(order=c(2,0,1),ar=c(0,-0.81),ma=c(1/3)),n=100)
t<-acf(ar.a)
```




```
set.seed(20)
ar.b<-arima.sim(list(order=c(0,0,1),ma=c(1/2)),n=100)
t<-acf(ar.b)
```



Solution:

In general $\phi(B)X_t = \theta(B)W_t$ so $\mathrm{Cov}(X_{t-h}, \phi(B)X_t) = \mathrm{Cov}(X_{t-h}, \theta(B)W_t)$.

For the process in 1(a), for $h \geq 2$ we have $\gamma(h) + .81\gamma(h-2) = 0$. Let $f(h) := \gamma(2h) = -.81\gamma(2h-2) = -.81f(h-1)$. This is a first order difference equation. The solution is thus $f(h) = \gamma(0)(-.81)^h$. Thus $\gamma(2h) = \gamma(0)(-.81)^h$.

Next let $g(h) := \gamma(2h+1) = -.81\gamma(2h+1-2) = -.81\gamma(2(h-1)+1) = -.81g(h-1)$. By the same argument as we just made for f, we get that $g(h) = \gamma(2h+1) = \gamma(1)(-.81)^h$.

We need to find $\gamma(0), \gamma(1)$. The initial condition for $\gamma(0)$ is , using the causal representation we found previously,

$$\begin{aligned}\gamma(0) + .81\gamma(2) &= \mathrm{Cov}(X_t, W_t + W_{t-1}/3) \\ &= \mathrm{Cov}(\sum_{j=0}^{\infty} \psi_j W_{t-j}, W_t + W_{t-1}/3) \\ &= \psi_0 + \psi_1/3 = 10/9.\end{aligned}$$

Plugging in $\gamma(2) = -.81\gamma(0)$, we have that $\gamma(0) - .81^2\gamma(0) = 10/9$ so $\gamma(0) = 100000/30951$.

In the same manner one gets that $\gamma(1) + .81\gamma(1) = \psi_0/3 = 1/3$ so that $\gamma(1) = 100/543$.

We are done by setting $\rho(h) = \gamma(h)/\gamma(0)$:

$$\rho(h) = \begin{cases}(-.81)^{|h/2|} & h \text{ even} \\ (57/1000)(-.81)^{(|h|-1)/2} & h \text{ odd.}\end{cases}$$

Solution:

For 1(b): the MA(1) representation gives that $\gamma(h) = 5\mathbb{1}_{h=0}/4 + \mathbb{1}_{h=1}/2$ so that $\rho(h) = \mathbb{1}_{h=0} + 2\mathbb{1}_{h=1}/5$.

4. (Comparing time series with periodic behavior)

   (a) Find an AR(2) process whose periodic ACF $\rho(h)$ has period 9.

   (b) Simulate a time series following the distribution you found in the previous part. Plot both the *true ACF* and the *simulated data series* (not the sample ACF!).

   (c) Simulate a signal-in-noise time series $Y_t = \mu_t + W_t$ where $W_t \overset{\text{iid}}{\sim} N(0, \sigma^2)$ with $\sigma^2 = 0.01$ (or $\sigma = .1$), and where the signal μ_t is given by the ACF ρ you found in the first part. That is: $\mu_t = \rho(t)$. Plot the *simulated data series.* (If you want you may also plot the underlying true signal $\mu_t = \rho(t)$ on the same plot.)

   (d) The point of what we've done so far is to compare different models that lead to periodic behavior. So: discuss/compare both the locations and the frequency/number of the peaks (maxima) and valleys (minima) of the *simulated AR(2) data series* to the locations of the maxima and minima of the *true ACF* and to the locations of the maxima and minima of the *signal in noise data series.*

Solution:

Consider the case where a (stationary, causal) AR(2) polynomial has two different complex roots. Then from class, we saw that $\rho(h) = C_1 Z_1^{-h} + \bar{C}_1 \bar{Z}_1^h = 2|C_1||Z_1|^{-h}\cos(h\theta_z - \theta_c)$ so that the period is $\frac{2\pi}{\theta_z}$. To get period 9, we may take $\theta_z = \frac{2\pi}{9}$. Also, we may take $Z_1 = 1.1\exp(i\frac{2\pi}{9})$, $\bar{Z}_1 = 1.1\exp(\frac{-2\pi}{9}i)$ so we have the autoregressive polynomial is

$$\phi(Z) = (1 - \frac{Z}{Z_1})(1 - \frac{Z}{Z_2}) = 1 - (\frac{1}{Z_1} + \frac{1}{Z_2})Z + \frac{1}{Z_1 Z_2}Z^2 = 1 - (2/1.1)\cos(\frac{2\pi}{9})Z + (1/1.1)^2 Z^2$$

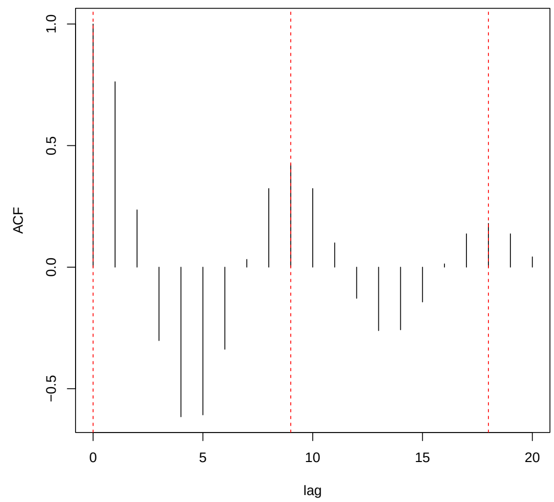
hence we may choose the AR(2)process

$$X_t = (2/1.1)\cos(\frac{2\pi}{9})X_{t-1} - (1/1.1)^2 X_{t-2} + W_t.$$

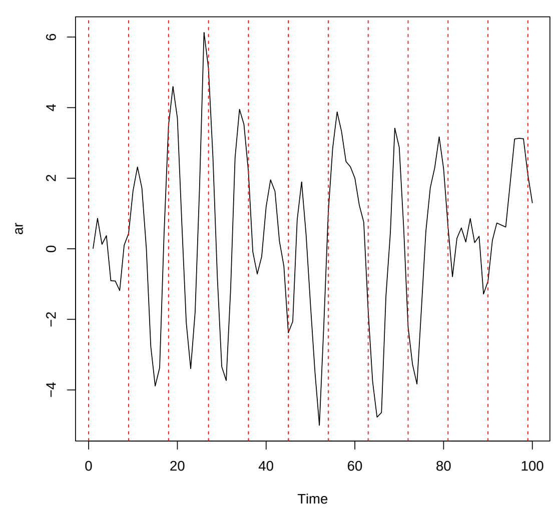
(Note: choosing $|Z_i|$ to be smaller makes the exponential decay slower so the periodicity is more visible.)

Next we plot the true ACF, and we simulate a path and plot its sample path and sample ACF.

```
ACF<-ARMAacf(ar = c((2/1.1)*cos(2*pi/9), -(1/1.1)^2), lag.max = 20)
plot(as.numeric(names(ACF)),ACF,type="h",xlab="lag",ylab = "ACF")
abline(v=(0:25)*9 , lty=2, col="red")
```

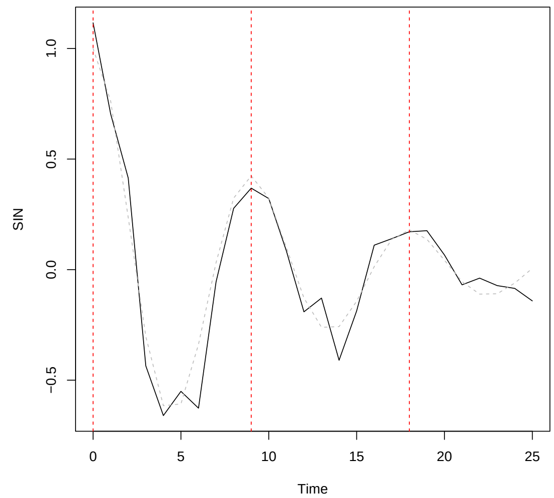



```
set.seed(20)
ar<-arima.sim(list(order=c(2,0,0),ar=c((2/1.1)*cos(2*pi/9),-(1/1.1)^2)),n=100)
plot(ar)
abline(v=(0:25)*9 , lty=2, col="red")
```




```
set.seed(20)
lag.max <- 25
signal<-ARMAacf(ar=c((2/1.1)*cos(2*pi/9),-(1/1.1)^2),lag.max=lag.max)
SIN=signal+rnorm(length(signal),0,0.1)
plot(as.numeric(names(signal)),SIN,type="l",xlab="Time")
xgrid <- seq(0,lag.max, by=1)
```

```r
lines(xgrid, signal, lty=2, col="gray")
abline(v=(0:25)*9,lty=2,col="red")
```



The main idea in comparing the peaks and valleys of the ACF and the actual random time series is that the AR time series has valleys and peaks that are determined by the coefficients and the most recent $p = 2$ values. (So if you have a large peak and a negative ϕ_1 then the next value tends to decrease.) This ends up meaning that the location of peaks and valleys are random rather than predictable. This behavior may be different than other mechanisms that generate periodic observations.

The peaks and valleys of the signal-in-noise (SIN) series are fairly consistent, at least before the mean function decays to 0. So they are (in the first 25 units of time) almost never more than one time step off from the truth. And the number of peaks is (with high probability) equal to the true number of peaks (similar for valleys). Neither of those statements is true for the AR(2) model.

(These statements make less sense once the mean/signal function has decayed to 0.)