HW 1 Solutions

1 Data analysis problems

1. Read sections 1, 2, 3, 6, 7, and 8 of An Introduction of R (do this in front of your computer), and answer the following questions.

   (a) What’s the difference in output between the commands `2*1:5` and `(2*1):5`? Why is there a difference?

   [Sol] The first will result in a vector 2, 4, 6, 8, 10, and the later will yield 2, 3, 4, 5. The difference is simply due to the order of operations.

   (b) If you wanted to enter the odd numbers from 1 to 19 in the variable `x`, what command would you use?

   [Sol] One way to do this is `2*(1:10)-1`. There may be some other legitimate answers here.

   (c) If you create a variable using the following command `y=c(-1,2,-3,4,-5)`, what command would put the positive values of `y` into the variable `z`?

   [Sol] The answer here could be either `z<-abs(y)` or `z<-y` combined with `z[z<0]<- -y[y<0]`. Note: give full credit if they do `z<-y[y>0]`, that is, they put the positive numbers in `z`.

   (d) What R command would give you the 95th percentile for a chi-squared distribution with 10 degrees of freedom?

   [Sol] `qchisq(0.95,10)` yielding [1] 18.30704

   (e) Generate a vector of 1000 standard normal random variables using the command `x=rnorm(1000)`, use R to give a five number summary of your simulated data; what is the mean and variance of your `x` variable? Make and print a histogram for this data.

   [Sol] Use the command `summary(x)` to get the summary. If the work looks reasonable give the students full credit.
2. From our course website, download “gas.dat”.

(a) Import the data from “gas.dat”. This data file contains monthly heating oil prices from July 1973 to June 1988.

(b) Make a plot of the data using the commands for the time-series object and/or time series plots. Make sure to label the plot and the x and y axis appropriately. Print out the plot to turn in.

[Sol]

![Average wholesale US gas and oil prices](image)

gas <- scan("gas.dat")
gas <- ts(gas, start=c(1973,7), frequency=12)
plot(gas, main="Average wholesale US gas and oil prices", xlab="Time", ylab="Price")

3. Write a function called \texttt{masim} to generate a $MA(q)$ series of length $T$.

[Sol] The function should look something like this (there may be other versions):
masim=function(thetas, sigsq, T){
    q=length(thetas)
    noise=rnorm(T+q, sd=sqrt(sigsq))
    #put the initial noise terms in and set the rest to zero
    x=c(noise[1:q],rep(0,T))
    for (i in (q+1):(T+q)){ #this loop generates the MA series
        x[i]=thetas %*% noise[i-(1:q)] +noise[i]
    }
    x=x[(q+1):(T+q)] #throw away those initial starting points
    x
}

There could be some slight variations.

4. Make sure that your function works. Simulate a model with $\sigma^2 = 1$, $\theta_1 = 0.5$, and $\theta_2 = 2$ with $T = 10,000$. Make an ACF plot. Is this consistent with the model that you generated? Try generating a model with the same parameters but only 200 observations. Make an ACF. Repeat this a few times, what do you notice about the autocorrelations and the dotted blue lines?

[Sol]

Figure 1:

The thing to notice with the smaller sample size is that many of the correlations that should be zero are outside the blue
With bigger sample size, the blue lines are much narrower.

5. Use the \texttt{arsim()} function from class to find one \(AR(2)\) that seems to be stationary and another that seems not to be stationary. Do this with simulations and plots of the simulations. Report the parameters, \(\phi_1\) and \(\phi_2\), from your stationary and non-stationary models.

[\textbf{Sol}] For this problem, basically the students should give two plots of data where one appears stationary and another appears non-stationary and report their parameter values. They will vary wildly just make sure that they’ve done the problem.

2 \textbf{Theoretical problems}

1. Find the autocovariance function for the random walk model \(x_t = x_{t-1} + w_t\), where \(w_t\) are white noise with variance \(\sigma_w^2\).

[\textbf{Sol}] For \(k \geq 0\) and \(x_0 = 1\), the autocovariance function is

\[
\gamma(t, t + k) = \text{Cov}(x_t, x_{t+k}) = \text{Cov}(\sum_{i=1}^{t} w_t, \sum_{i=1}^{t+k} w_t) = t\sigma_w^2.
\]

Note different variations. For example, some students may write \(\text{Cov}(x_{t-k}, x_t) = (t - k)\sigma_w^2\).

2. For each of the following, state if it is a stationary process. If it is a stationary process, compute the mean and autocovariance functions. Note that \(\{w_t\}\) is i.i.d. \(N(0, 1)\).

(a) \(x_t = w_t - w_{t-3}\)

[\textbf{Sol}] \(x_t = w_t - w_{t-3}\) is a stationary process: \(E(x_t) = E(w_t) - E(w_{t-3}) = 0\) and

\[
\gamma(s, t) = E(x_s x_t) = E(w_s w_t) - E(w_s w_{t-3}) - E(w_{s-3} w_t) + E(w_{s-3} w_{t-3})
\]

\[
= I\{s=t\} - I\{s=t-3\} - I\{s-3=t\} + I\{s-3=t-3\} = 2 \cdot I\{|s-t|=0\} - I\{|s-t|=3\}
\]

which is a function of \(|s - t|\).

(b) \(x_t = t + w_t\)
[Sol] $x_t = t + w_t$ is not a stationary process because its mean is not constant over time $t$: $E(x_t) = t$.

(c) $x_t = w_tw_{t-2}$

[Sol] $x_t = w_tw_{t-2}$ is a stationary process because $E(x_t) = E(w_tw_{t-2}) = 0$ and $\gamma(s, t) = E(w_sw_{s-2}w_tw_{t-2}) = I_{\{s=t\}}$.

3. Problem 1.7 in the textbook.

[Sol] If $x_t = w_{t-1} + 2w_t + w_{t+1}$, then $\gamma(t, t) = E(w_{t-1} + 2w_t + w_{t+1})^2 = E(w_{t-1}^2) + 4E(w_t^2) + E(w_{t+1}^2) = 6\sigma_w^2$,

$\gamma(t, t + 1) = E(w_{t-1} + 2w_t + w_{t+1})(w_t + 2w_{t+1} + w_{t+2}) = 2E(w_t^2) + 2E(w_{t+1}^2) = 4\sigma_w^2$,

$\gamma(t, t + 2) = E(w_{t-1} + 2w_t + w_{t+1})(w_{t+1} + 2w_{t+2} + w_{t+3}) = E(w_{t+1}^2) = \sigma_w^2$, and $\gamma(t, t + h) = 0$ for $h \geq 3$. By symmetry, $\gamma(t, t - h) = \gamma(t, t + h)$.

For the autocorrelation function, $\rho(h) = \frac{\gamma(t, t + h)}{\gamma(t, t)}$ and so $\rho(0) = 1$, $\rho(1) = 2/3$, $\rho(2) = 1/6$ and $\rho(h) = 0$ for $h \geq 3$.

4. Problem 1.10 (a) and (b) in the textbook.

[Sol] (a) If we differentiate with respect to $A$, we obtain

$$\frac{d}{dA}MSE(A) = \frac{d}{dA}(E(x_{t+\ell}^2) + A^2E(x_t^2) - 2AE(x_t x_{t+\ell})) = 2AE(x_t^2) - 2E(x_t x_{t+\ell}) = 2A\gamma(0) - 2\gamma(\ell).$$

Setting this zero, we see that the minimum is obtained at $A = \gamma(\ell)/\gamma(0) = \rho(\ell)$.

(b) Replacing $A$ with $\rho(\ell)$ in $MSE$, we have

$$MSE(A) = \gamma(0) + \rho^2(\ell)\gamma(0) - 2\rho(\ell)\gamma(\ell) = \gamma(0)(1 - \rho^2(\ell)).$$

Note that $\gamma(\ell) = \rho(\ell)\gamma(0)$. 

5