STA 137

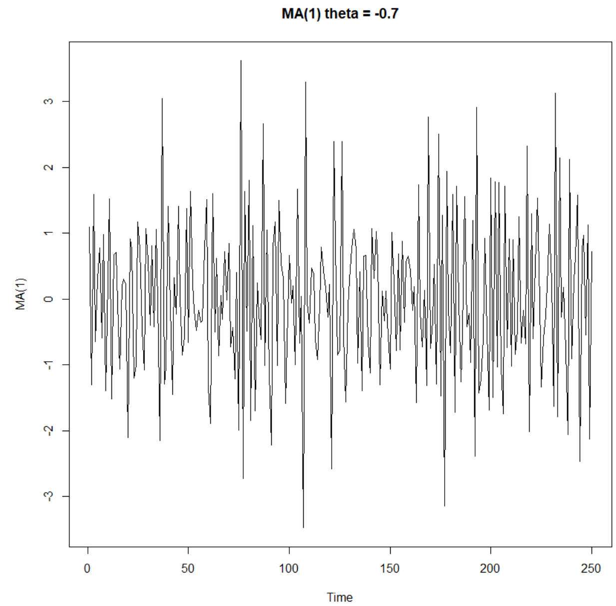
Homework 3

Prof. Burman

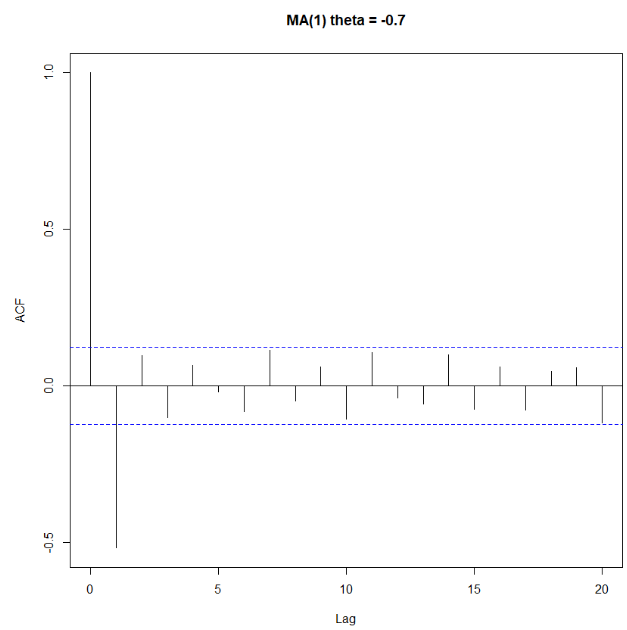
Jared Yu, Danli Zheng

2-5-2019

1. A) Below is a plot of the simulated $MA(1)$ data with $\theta = -0.7$.



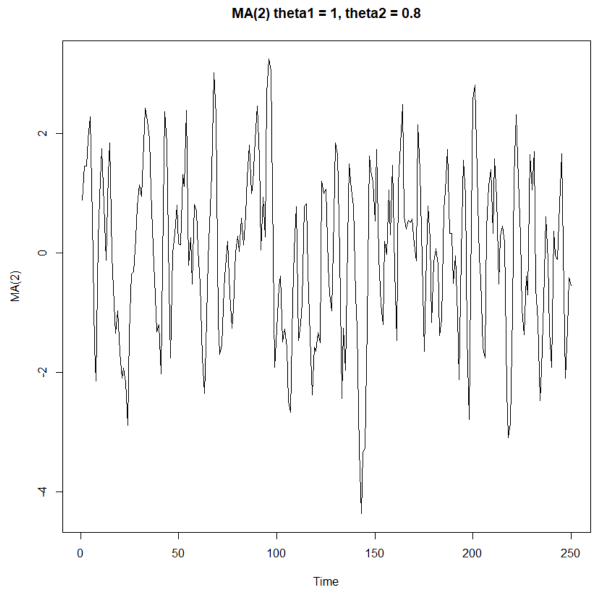
The data appears highly random, with no trend or seasonality, while it remains within a relatively horizontal band. It looks almost like white noise.
Additionally, here is a plot of the ACF,



The ACF shows that there is significant negative correlation up to lag 1. The negative correlation shows that at time $t - 1$, the data moves in an opposite direction for each t. This makes sense, since it is an $MA(1)$ model, and therefore the correlation should drop off after lag 1. The value at $\rho(1) = -0.518$. Using the calculations below in problem 2, the theoretical correlation at lag

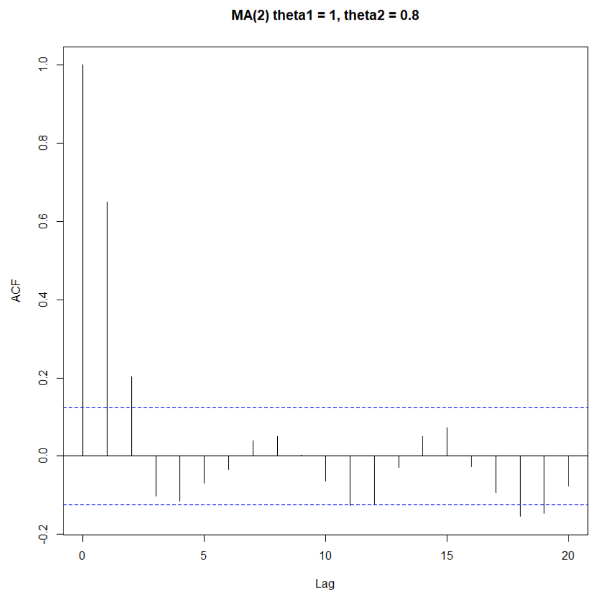
1 for $MA(1)$ is $\frac{\theta\sigma^2}{\sigma^2(1+\theta^2)} = \frac{\theta}{(1+\theta^2)} = \frac{-0.7}{(1+(-0.7)^2)} = -0.4697987$. This value is quite close to the one produced by the simulation.

B) Below is a plot of the simulated $MA(2)$ data with $\theta_1 = 1.0, \theta_2 = 0.8$.



The data looks much less random than the previous model with $q = 1$. Here it seems that the data will have some sort of a lag in the correlation for a short period of time, before the direction seems to change again. If looking closely at the spikes, there are short temporary trends due to neighboring correlation, which can have a few pivots in between before the direction changes more drastically towards the other side.

Below is a plot of the ACF,



There is quite apparently a significant positive correlation up to lag 2. This makes sense considering that the previous plot would show temporary trends which would last a few pivots. The correlation is positive, which makes sense considering that the given coefficients are both positive. Also, it is expected that the correlation would drop off after a lag of 2, because it is a $MA(2)$ model. Using the calculation from the previous homework, the autocorrelations for lag 1 and 2 are as follows:

For $lag = 1, \rho(1) = \frac{\theta_1(1+\theta_2)}{1+\theta_1^2+\theta_2^2} = \frac{1(1+0.8)}{1+1^2+0.8^2} = 0.6818182$, which is similar to the simulated value of 0.649.

For $lag = 2, \rho(2) = \frac{\theta_2}{1+\theta_1^2+\theta_2^2} = \frac{0.8}{1+1^2+0.8^2} = 0.3030303$, which is similar to the simulated value of 0.202. This is somewhat further than the previous simulation was from the theoretical result. If the size of n were increased, then the simulated value would be even closer to the theoretical value. This was tested with $n = 2{,}500$.

2. Given an $MA(1)$ sequence with sample mean $\bar{X}$, $\sigma^2 = 9$, and sample size $n = 90$.
   A) The following are the autocovariance and autocorrelation functions with $\theta = -0.7$.
   The model can be written as,

$$X_t^{(c)} = \varepsilon_t + \theta\varepsilon_{t-1}$$

where $X_t^{(c)} = X_t - \mu$. For convenience, it will simply be written as X_t.

The covariance is as follows:

$$Cov(X_{t+j}, X_t) = Cov(\varepsilon_{t+j} + \theta\varepsilon_{t+j-1}, \varepsilon_t + \theta\varepsilon_{t-1})$$
$$= Cov(\varepsilon_{t+j}, \varepsilon_t) + Cov(\varepsilon_{t+j}, \theta\varepsilon_{t-1}) + Cov(\theta\varepsilon_{t+j-1}, \varepsilon_t) + Cov(\theta\varepsilon_{t+j-1}, \theta\varepsilon_{t-1})$$
$$= Cov(\varepsilon_{t+j}, \varepsilon_t) + \theta Cov(\varepsilon_{t+j}, \varepsilon_{t-1}) + \theta Cov(\varepsilon_{t+j-1}, \varepsilon_t) + \theta^2 Cov(\varepsilon_{t+j-1}, \varepsilon_{t-1})$$

When $j = 0, \Rightarrow Cov(X_t, X_t) = Cov(\varepsilon_t, \varepsilon_t) + \theta Cov(\varepsilon_t, \varepsilon_{t-1}) + \theta Cov(\varepsilon_{t-1}, \varepsilon_t) + \theta^2 Cov(\varepsilon_{t-1}, \varepsilon_{t-1})$

$$= Cov(\varepsilon_t, \varepsilon_t) + 0 + 0 + \theta^2 Cov(\varepsilon_{t-1}, \varepsilon_{t-1})$$
$$= 9 + (-0.7)^2 \times 9 = 13.41$$

When $j = 1, \Rightarrow Cov(X_{t+1}, X_t) = Cov(\varepsilon_{t+1}, \varepsilon_t) + \theta Cov(\varepsilon_{t+1}, \varepsilon_{t-1}) + \theta Cov(\varepsilon_t, \varepsilon_t) + \theta^2 Cov(\varepsilon_t, \varepsilon_{t-1})$

$$= 0 + 0 + \theta Cov(\varepsilon_t, \varepsilon_t) + 0 = -0.7 \times 9 = -6.3$$

When $j = 2, \Rightarrow Cov(X_{t+2}, X_t) = Cov(\varepsilon_{t+2}, \varepsilon_t) + \theta Cov(\varepsilon_{t+2}, \varepsilon_{t-1}) + \theta Cov(\varepsilon_{t+1}, \varepsilon_t) + \theta^2 Cov(\varepsilon_{t+1}, \varepsilon_{t-1})$

$$= 0 + 0 + 0 + 0 = 0.$$

Then, when $|j| \geq 2 \Rightarrow Cov(X_{t+j}, X_t) = 0$, since the left-hand side and right-hand side will no longer have any matching indices.

So, for any t,

$$\gamma(j) = Cov(X_{t+j}, X_t) = \begin{cases} 13.41 & j = 0 \\ -6.3 & j = \pm 1 \\ 0 & j = \pm 2 \end{cases}$$

Hence $\{X_t\}$ is stationary. Note that

$$\rho(j) = \begin{cases} 1 & j = 0 \\ -0.4697987 & j = \pm 1 \\ 0 & j = \pm 2 \end{cases}$$

B) To find the variance of the sample mean $\bar{X}$, we must calculate $Var(\bar{X}) = \frac{\tau_n^2}{n}$, where $\tau_n^2 = \sum_{h=-(n-1)}^{n-1} \left(1 - \frac{|h|}{n}\right) \gamma(h) = \gamma(0) + 2\sum_{h=1}^{n-1} \left(1 - \frac{h}{n}\right) \gamma(h)$

$$= 13.41 + 2\left(\left(1 - \frac{1}{90}\right) \times -6.3\right) = 0.95$$

So $Var(\bar{X}) = \frac{\tau_n^2}{n} = \frac{0.95}{90} = 0.01055556.$

C) We will now repeat the process above for when $\theta = 0.7$. The correlations are now as follows:

When $j = 0, \Rightarrow Cov(X_t, X_t) = Cov(\varepsilon_t, \varepsilon_t) + \theta Cov(\varepsilon_t, \varepsilon_{t-1}) + \theta Cov(\varepsilon_{t-1}, \varepsilon_t) + \theta^2 Cov(\varepsilon_{t-1}, \varepsilon_{t-1})$

$$= Cov(\varepsilon_t, \varepsilon_t) + 0 + 0 + \theta^2 Cov(\varepsilon_{t-1}, \varepsilon_{t-1})$$
$$= 9 + 0.7^2 \times 9 = 13.41$$

When $j = 1, \Rightarrow Cov(X_{t+1}, X_t) = Cov(\varepsilon_{t+1}, \varepsilon_t) + \theta Cov(\varepsilon_{t+1}, \varepsilon_{t-1}) + \theta Cov(\varepsilon_t, \varepsilon_t) + \theta^2 Cov(\varepsilon_t, \varepsilon_{t-1})$

$$= 0 + 0 + \theta Cov(\varepsilon_t, \varepsilon_t) + 0 = 0.7 \times 9 = 6.3$$

When $j = 2, \Rightarrow Cov(X_{t+2}, X_t) = Cov(\varepsilon_{t+2}, \varepsilon_t) + \theta Cov(\varepsilon_{t+2}, \varepsilon_{t-1}) + \theta Cov(\varepsilon_{t+1}, \varepsilon_t) + \theta^2 Cov(\varepsilon_{t+1}, \varepsilon_{t-1})$

$$= 0 + 0 + 0 + 0 = 0.$$

Then, when $|j| \geq 2 \Rightarrow Cov(X_{t+j}, X_t) = 0$, since the left-hand side and right-hand side will no longer have any matching indices.

So, for any t,

$$\gamma(j) = Cov(X_{t+j}, X_t) = \begin{cases} 13.41 & j = 0 \\ 6.3 & j = \pm 1 \\ 0 & j = \pm 2 \end{cases}$$

Hence $\{X_t\}$ is stationary. Note that

$$\rho(j) = \begin{cases} 1 & j = 0 \\ 0.4697987 & j = \pm 1 \\ 0 & j = \pm 2 \end{cases}$$

To find the new $Var(\bar{X})$, we must calculate

$$\tau_n^2 = \sum_{h=-(n-1)}^{n-1} \left(1 - \frac{|h|}{n}\right)\gamma(h) = \gamma(0) + 2\sum_{h=1}^{n-1}\left(1 - \frac{h}{n}\right)\gamma(h)$$

$$= 13.41 + 2\left(\left(1 - \frac{1}{90}\right) \times 6.3\right) = 25.87$$

So $Var(\bar{X}) = \frac{\tau_n^2}{n} = \frac{25.87}{90} = 0.2874444.$

In part B), the $Var(\bar{X}) = 0.01055556$, while in part C), $Var(\bar{X}) = 0.2874444$. In this example, it shows that when there is a negative correlation, the variance of the sample mean is smaller than in comparison to when the correlation is positive. It seems that when the correlation is positive, then the cycles of the rough around the mean is much larger and smoother, while when the correlation is negative there is a quick gyration about a mean. This difference in the appearance is apparent when calculating the variation of the sample mean.

D) If $\bar{X} = 35.2$ and $\theta = 0.7$, then to find an approximate 95% confidence interval for $\mu = E(X_t)$ we must first find an estimate of τ_n, this can be simplified in this case of an $MA(1)$,

$$\tau_n^2 = \gamma(0) + 2\left(1 - \frac{1}{n}\right)\gamma(1) = 13.41 + 2\left(1 - \frac{1}{90}\right)6.3 = 25.87$$

Then $\tau_n = \sqrt{25.87} = 5.086256$. So, a 95% confidence interval for μ is approximately

$$\bar{X} \pm \frac{1.96\tau_n}{\sqrt{n}}, i.e., 35.2 \pm \frac{(1.96)(5.086256)}{\sqrt{90}}, i.e., 35.2 \pm 1.050831, i.e., (34.14917, 36.25083)$$

3. A) Given that $\{X_t\}$ is an $AR(1)$ sequence with $\sigma^2 = 9$, $\bar{X}$ is the sample mean, and the sample size $n = 90$. The model appears as follows:

$$X_t - \mu = \phi(X_{t-1} - \mu) + \varepsilon_t$$

where $\{\varepsilon_t\}$ is white noise with variance σ^2, then we will also understand that

$$\gamma(0) = Var(X_t) = (1 - \phi^2)^{-1}\sigma^2, \rho(h) = \phi^h$$

Then to find $Var(\bar{X}) = \frac{\tau_n^2}{n}$, when $\phi = -0.7$, we must calculate,

$$\tau_n^2 \approx \gamma(0)\left[1+2\sum_{h=1}^{\infty}\rho(h)\right] = (1-\phi^2)^{-1}\sigma^2\left[1+2\sum_{h=1}^{\infty}\phi^h\right] = \frac{\sigma^2}{(1-\phi^2)}\left[1+2\times\frac{\phi}{1-\phi}\right]$$
$$= \frac{\sigma^2}{(1-\phi)(1+\phi)}\left[\frac{1+\phi}{1-\phi}\right] = \frac{\sigma^2}{(1-\phi)^2} = \frac{9}{\left(1-(-0.7)\right)^2} = 3.114187.$$

So $Var(\bar{X}) = \frac{\tau_n^2}{n} = \frac{3.114187}{90} = 0.03460208.$

B) Repeating the above process except when $\phi = 0.7$ yields the following,

$$\tau_n^2 \approx \frac{\sigma^2}{(1-\phi)^2} = \frac{9}{(1-0.7)^2} = 100$$

So $Var(\bar{X}) = \frac{\tau_n^2}{n} = \frac{100}{90} = 1.\bar{1}.$

In part A), the variance of the sample mean was 0.03460208, while in part B) the variance of the sample mean is $1.\bar{1}$. In this case when the sequence is $AR(1)$, the negative coefficient of $\phi = -0.7$ leads to a smaller variance of the sample mean. While in the case of a positive coefficient $\phi = 0.7$, the variance of the sample mean is larger.

4. A) The following series $\{X_t\}$ and $\{Y_t\}$ are independent and stationary, with autocovariances and autocorrelation $\gamma_X(j)$, $\rho_X(j)$ and $\gamma_Y(j)$, $\rho_Y(j)$. We are now to consider the series $W_t = X_t - 2Y_t$. To show that W_t is stationary we must show the following,
$E(W_t) = \mu$ for all t and $Corr\left(X_t, X_{t+j}\right)$ is the same for $j = 0,1,\cdots$ or in other words, the correlation does not depend on time t, but only on lag j.

$$E(W_t) = E(X_t) - 2E(Y_t) = \mu_X - 2\mu_Y = \mu_W \text{ for all } t\text{, since it doesn't depend on time } t.$$

$$\begin{aligned} Cov\left(W_t, W_{t+j}\right) &= Cov\left(X_t - 2Y_t, X_{t+j} - 2Y_{t+j}\right) \\ &= Cov\left(X_t, X_{t+j}\right) + Cov\left(X_t, -2Y_{t+j}\right) + Cov\left(-2Y_t, X_{t+j}\right) + Cov\left(-2Y_t, -2Y_{t+j}\right) \\ &= Cov\left(X_t, X_{t+j}\right) - 2Cov\left(X_t, Y_{t+j}\right) - 2Cov\left(Y_t, X_{t+j}\right) + 4Cov\left(Y_t, Y_{t+j}\right) \\ &= Cov\left(X_t, X_{t+j}\right) + 4Cov\left(Y_t, Y_{t+j}\right) \text{ (because } X \text{ and } Y \text{ are independent)} \\ &= \gamma_X(j) + 4\gamma_Y(j) \text{ for all } t\text{, since it doesn't depend on time } t. \end{aligned}$$

This concludes that W_t is stationary.

B) The autocovariance and autocorrelations are as follows:
The autocovariance were calculated previously in part A), where $\gamma_W(j) = \gamma_X(j) + 4\gamma_Y(j)$. In this case, $\gamma_X(j) = Cov(X_t, X_{t+j})$ and $\gamma_Y(j) = Cov(Y_t, Y_{t+j})$.

The autocorrelation for W_t is $\rho_W(j) = \frac{\gamma_W(j)}{\gamma_W(0)} = \frac{\gamma_X(j)+4\gamma_Y(j)}{\gamma_X(0)+4\gamma_Y(0)} = \frac{\gamma_X(j)+4\gamma_Y(j)}{\frac{\gamma_X(j)}{\rho_X(j)}+\frac{4\gamma_Y(j)}{\rho_Y(j)}} =$
$\frac{(\gamma_X(j)+4\gamma_Y(j))\rho_X(j)\rho_Y(j)}{\gamma_X(j)\rho_Y(j)+4\gamma_Y(j)\rho_X(j)}$