STAD57: Time Series Analysis
Problem Set 1

1. Do the following:

a. Download and install R (https://www.r-project.org/) & RStudio (https://www.rstudio.com/), if you don't have them already.

b. In R, there is a special function for creating time series called `ts()`. This function creates an object of type "ts". Try the following in R:

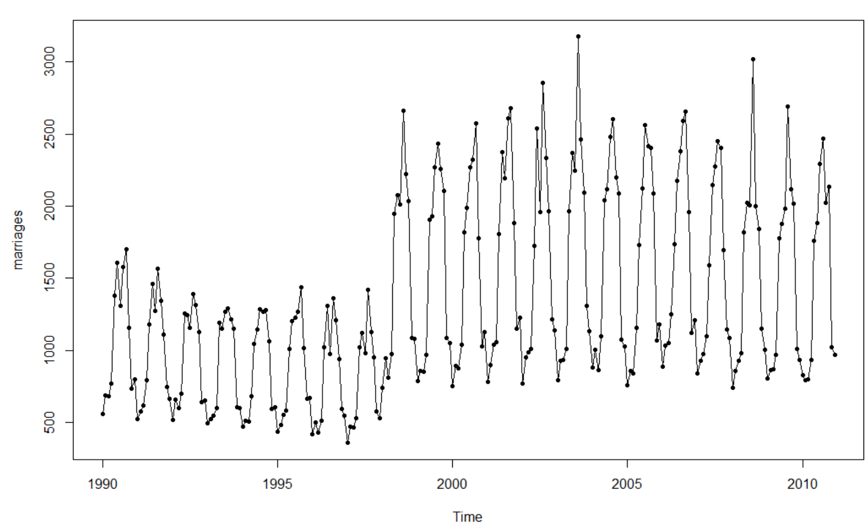
```
X=rnorm(120)
plot(X, pch=20)
X.ts=ts(data=X, start=c(2005,1), frequency=12 )
plot(X.ts)
class(X); class(X.ts)
```

c. The file `ontario_marriages.csv` contains counts of marriages in the province of Ontario by year, month, and city. Load these data in R and create & plot a time series of the total number of marriages in Toronto & Ottawa from Jan 1990 to Dec 2010. You can use the following starter code:

```
my_data=read.csv("ontario_marriages.csv")
X=aggregate(COUNT~MONTH+YEAR, data=my_data, function=sum,
  subset=(CITY %in% c('TORONTO','OTTAWA')) )
```

SOL:

```
idx=(X$YEAR>=1990) # Index for year >= 1990
marriages=ts(X$COUNT[idx],start=c(1990,1), frequency=12)
plot(marriages, type='o', pch=20)
```



2. Consider the series $\{X_1, X_2, X_3\}$, where X_1, X_2 are independent standard Normal and $X_3 = X_1 \times X_2$. Show that this is a White Noise series, i.e. that the three variables are uncorrelated. Is the series strictly stationary?

SOL:

Obviously, X_1, X_2 are uncorrelated since they are independent. We just need to show that X_3 is uncorrelated with X_1, X_2, i.e. show that $\mathrm{Cov}\left(X_1, X_3\right) = \mathrm{Cov}\left(X_2, X_3\right) = 0$.

We have $\mathrm{Cov}(X_1, X_3) = \mathrm{Cov}(X_1, X_1 X_2) = \mathbb{E}\left(X_1(X_1 X_2)\right) - \overbrace{\mathbb{E}(X_1)}^{=0}\mathbb{E}(X_1 X_2) =$

$= \mathbb{E}(X_1^2 X_2) \overset{\text{by indep.}}{=} \mathbb{E}(X_1^2)\underbrace{\mathbb{E}(X_2)}_{=0} = 0$, and similarly for $\mathrm{Cov}(X_2, X_3)$. Since the mean of all variables is (constant) 0: $\mathbb{E}(X_1) = \mathbb{E}(X_2) = 0$ & $\mathbb{E}(X_3) = \mathbb{E}(X_1 X_2) \overset{\text{by indep.}}{=} \mathbb{E}(X_1)\mathbb{E}(X_2) = 0$ and the variance of all variables is (constant) 1: $\mathbb{V}(X_1) = \mathbb{V}(X_2) = 1$ &

$\mathbb{V}(X_3) = \mathbb{V}(X_1 X_2) = \mathbb{E}\left((X_1 X_2)^2\right) - \overbrace{\left(\mathbb{E}(X_1 X_2)\right)^2}^{=0} = \mathbb{E}(X_1^2 X_2^2) \overset{\text{by indep.}}{=} \mathbb{E}(X_1^2)\mathbb{E}(X_2^2) = 1 \cdot 1 = 1$,

the series is a weakly stationary & uncorrelated, i.e. WN(0,1). Nevertheless, the series is *not* strictly stationary, since the marginal distribution of $X_3 = X_1 X_2$ is different from that of X_1, X_2.

3. Exercise 1.6 from textbook.

SOL:

a.

We can simply show that the mean is not constant

$E[X_t] = E[\beta_1 + \beta_2 t + W_t] = \beta_1 + \beta_2 t + \cancel{E[W_t]} = \beta_1 + \beta_2 t$, which is generally a function of *t*.

Moreover, we have: $Var[X_t] = Var[\beta_1 + \beta_2 t + W_t] = Var[W_t] = \sigma_W^2 = \gamma_X(0)$, and

$Cov[X_{t+h}, X_t] = Cov[\beta_1 + \beta_2(t+h) + W_{t+h}, \beta_1 + \beta_2 t + W_t] = Cov[W_{t+h}, W_t] = 0 = \gamma_X(h), \forall h \neq 0$

b.

We have $Y_t = X_t - X_{t-1} = \left(\cancel{\beta_1} + \cancel{\beta_2 t} + W_t\right) - \left(\cancel{\beta_1} + \beta_2(\cancel{t} - 1) + W_{t-1}\right) = \beta_2 + W_t - W_{t-1}$, so that:

$E[Y_t] = E\left[\beta_2 + W_t - W_{t-1}\right] = \beta_2 + \cancel{E[W_t]} - \cancel{E[W_{t-1}]} = \beta_2$, which is constant (indep. of *t*)

$Var[Y_t] = Var\left[\beta_2 + W_t - W_{t-1}\right] = Var[W_t] + Var[W_{t-1}] = 2\sigma_W^2 = \gamma_Y(0)$, which is also constant

$$Cov[Y_{t+h}, Y_t] = Cov\left[\left(\beta_2 + W_{t+h} - W_{t+h-1}\right), \left(\beta_2 + W_t - W_{t-1}\right)\right] =$$

$$= Cov\left[W_{t+h}, W_t\right] - Cov\left[W_{t+h}, W_{t-1}\right] - Cov\left[W_{t+h-1}, W_t\right] + Cov\left[W_{t+h-1}, W_{t-1}\right]$$

For $h=1$, $\gamma_Y(1)=\cancel{Cov[W_{t+1},W_t]}-\cancel{Cov[W_{t+1},W_{t-1}]}-Cov[W_t,W_t]+\cancel{Cov[W_t,W_{t-1}]}=-\sigma_W^2$

For $h\geq 2$, $\gamma_Y(h)=0$ (there is no overlap in the $\{W_t\}$ time subscripts)

So $\{Y_t\}$ is stationary (it behaves like a MA(1) process)

c.

$$E[V_t]=E\left[\frac{1}{2q+1}\sum_{j=-q}^{q}X_{t-j}\right]=\frac{1}{2q+1}\sum_{j=-q}^{q}E\left[X_{t-j}\right]=\frac{1}{2q+1}\sum_{j=-q}^{q}\beta_1+\beta_2(t-j)=$$

$$=\frac{1}{\cancel{(2q+1)}}\left\{\cancel{(2q+1)}(\beta_1+\beta_2 t)-\cancel{\underbrace{\sum_{j=-q}^{q}j}_{=0}}\right\}=\beta_1+\beta_2 t\,.$$

$$Cov[V_{t+h},V_t]=Cov\left[\frac{1}{2q+1}\sum_{i=-q}^{q}X_{t+h-i},\frac{1}{2q+1}\sum_{j=-q}^{q}X_{t-j}\right]=$$

$$=\frac{1}{(2q+1)^2}\sum_{i=-q}^{q}\sum_{j=-q}^{q}Cov\left[X_{t+h-i},X_{t-j}\right]=\frac{1}{(2q+1)^2}\sum_{i=-q}^{q}\sum_{j=-q}^{q}\gamma_X\left((t+h-i)-(t-j)\right)=$$

$$=\frac{1}{(2q+1)^2}\sum_{i=-q}^{q}\sum_{j=-q}^{q}\underbrace{\gamma_X\left(h-i+j\right)}_{=\sigma_W^2\text{ only if }h-i+j=0\Leftrightarrow i=h+j}=\begin{cases}\dfrac{(2q+1)-h}{(2q+1)^2}\sigma_W^2, & \forall 0\leq h\leq 2q\\ 0, & \forall h>2q\end{cases}$$

Note that every non-zero term in the double sum will be equal to $\gamma_X(0)=\sigma_W^2$, which will happen only if $i=h+j$, for $i,j\in\{-q,\ldots,q\}$. The maximum number of non-zero terms occurs for $h=0$, where we have $(2q-1)\times\gamma_X(0)$ terms. If $|h|\geq 2q-1$, then there are no non-zero terms, i.e. we can't have $i=h+j$ for any $i,j\in\{-q,\ldots,q\}$.

4. Exercise 1.7 from textbook.

SOL:

$$E[X_t]=E\left[W_{t-1}+2W_t+W_{t+1}\right]=E\left[W_{t-1}\right]+2E\left[W_t\right]+E\left[W_{t+1}\right]=0$$

$$Cov[X_{t+h}, X_t] = Cov\left[W_{t+h-1} + 2W_{t+h} + W_{t+h+1}, W_{t-1} + 2W_t + W_{t+1}\right] =$$
$$= Cov\left[W_{t+h-1}, W_{t-1}\right] + 2Cov\left[W_{t+h-1}, W_t\right] + Cov\left[W_{t+h-1}, W_{t+1}\right] +$$
$$+2Cov\left[W_{t+h}, W_{t-1}\right] + 4Cov\left[W_{t+h}, W_t\right] + 2Cov\left[W_{t+h}, W_{t+1}\right] +$$
$$+Cov\left[W_{t+h+1}, W_{t-1}\right] + 2Cov\left[W_{t+h+1}, W_t\right] + Cov\left[W_{t+h+1}, W_{t+1}\right] =$$
$$\Rightarrow \gamma(h) = \begin{cases} (1+4+1)\sigma_W^2 = 6\sigma_W^2, & h = 0 \\ (2+2)\sigma_W^2 = 4\sigma_W^2, & h = 1 \\ \sigma_W^2 & h = 2 \\ 0 & h \geq 2 \end{cases}$$
$$\Rightarrow \rho(h) = \frac{\gamma(h)}{\gamma(0)} = \begin{cases} 1, & h = 0 \\ 4/6, & h = 1 \\ 1/6 & h = 2 \\ 0 & h \geq 2 \end{cases}$$

5. Exercise 1.14 from textbook.

SOL:

a.

$$E[Y_t] = E\left[\exp\left\{X_t\right\}\right] = \exp\left\{E[X_t] + Var[X_t]/2\right\} = \exp\left\{\mu_X + \gamma(0)/2\right\} = \mu_Y$$

b.

$$\gamma_Y(h) = Cov\left[Y_{t+h}, Y_t\right] = E[Y_{t+h}Y_t] - E[Y_{t+h}]E[Y_t] = E\left[\exp\left\{X_{t+h}\right\}\exp\left\{X_t\right\}\right] - \mu_Y^2 =$$
$$= E\left[\exp\left\{X_{t+h} + X_t\right\}\right] - \exp\left\{2\mu_X + \gamma(0)\right\}$$

But $\left(X_{t+h} + X_t\right) \sim N\left(E\left[X_{t+h} + X_t\right], Var\left[X_{t+h} + X_t\right]\right)$, where: $E\left[X_{t+h} + X_t\right] = 2\mu_X$ and

$Var\left[X_{t+h} + X_t\right] = Var\left[X_{t+h}\right] + Var\left[X_t\right] + 2Cov\left[X_{t+h}, X_t\right] = 2\gamma(0) + 2\gamma(h), \Rightarrow$

$$\gamma_Y(h) = E\left[\exp\left\{X_{t+h} + X_t\right\}\right] - \exp\left\{2\mu_X + \gamma(0)\right\} =$$
$$= \exp\left\{2\mu_X + \gamma(0) + \gamma(h)\right\} - \exp\left\{2\mu_X + \gamma(0)\right\}$$
$$= \exp\left\{2\mu_X + \gamma(0)\right\}\left[\exp\left\{\gamma(h)\right\} - 1\right]$$

6. Exercise 1.15 from textbook.

SOL:

$$\mu_t = E[X_t] = E[W_t W_{t-1}] = E[W_t]E[W_{t-1}] = 0$$
$$\sigma_t^2 = Var[X_t] = Var[W_t W_{t-1}] = Var[W_t]Var[W_{t-1}] = \sigma_w^4$$
$$\gamma(1) = Cov\left[X_{t+1}, X_t\right] = Cov\left[W_{t+1}W_t, W_t W_{t-1}\right] =$$
$$= E[W_{t+1}W_t^2 W_{t-1}] = E[W_{t+1}]E[W_t^2]E[W_{t-1}] = 0$$
$$\gamma(h) = Cov\left[X_{t+h}, X_t\right] = Cov\left[W_{t+h}W_{t+h-1}, W_t W_{t-1}\right] =$$
$$= E\left[W_{t+h}W_{t+h-1}, W_t W_{t-1}\right] = E[W_{t+h}]E[W_{t+h-1}]E[W_t]E[W_{t-1}] = 0, \ \forall h \ge 1$$

So $\{X_t\}$ is stationary, and is actually a white noise with variance σ_w^4

7. Exercise 1.19 from textbook.

SOL:

a.

$$\mu_t = E[X_t] = E[\mu + W_t - .8W_{t-1}] = \mu + E[W_t] - .8E[W_{t-1}] = \mu$$

b.

For this MA(1) process: $\gamma(0) = \sigma_t^2 = \left[1 + (-.8)^2\right]\sigma_w^2 = 1.64\sigma_w^2$, $\gamma(1) = \gamma(-1) = -.8\sigma_w^2$, a $\gamma(h) = 0, \ \forall |h| \ge 2$. So, the sample mean variance becomes:

$$Var\left[\bar{X}\right] = \frac{1}{n}\sum_{h=-n}^{n}\left(1 - \frac{|h|}{n}\right)\gamma(h) = \frac{1}{n}\sum_{h=-1}^{1}\left(1 - \frac{|h|}{n}\right)\gamma(h) =$$
$$= \frac{1}{n}\left[\left(1 - \frac{1}{n}\right)\gamma(-1) + \gamma(0) + \left(1 - \frac{1}{n}\right)\gamma(1)\right] =$$
$$= \frac{1}{n}\left[\gamma(0) + 2\frac{n-1}{n}\gamma(1)\right] = \frac{1}{n}\left[1.64\sigma_w^2 - 2\frac{n-1}{n}.8\sigma_w^2\right] =$$
$$= \frac{\sigma_w^2}{n}\left[1.64 - 1.6\frac{n-1}{n}\right] \approx .04\frac{\sigma_w^2}{n}, \ \text{for large } n$$

c.

If $\{X_t\}$ was just white noise, then the sample mean variance would just be the usual $\frac{\sigma_w^2}{n}$, which is always smaller than $\frac{\sigma_w^2}{n}\left[1.64 - 1.6\frac{n-1}{n}\right]$. The reason why this MA(1) model's variance is smaller is that the W's will partially cancel out because of the negative -.8 coefficient. To see this, compare what happens when n=2:

[MA(1)] $\bar{X} = \frac{1}{3}(X_1 + X_2 + X_3) = \frac{1}{3}\left[(\mu + W_1 - .8W_0) + (\mu + W_2 - .8W_1) + (\mu + W_3 - .8W_2)\right]$

$$= \frac{1}{3}\left[3\mu - .8W_0 + .2W_1 + .2W_2 + W_3\right] = \mu + \frac{1}{3}(-.8W_0 + .2W_1 + .2W_2 + W_3)$$

[WN] $\bar{X} = \frac{1}{3}(X_1 + X_2 + X_3) = \frac{1}{3}\left[(\mu + W_1) + (\mu + W_2) + (\mu + W_3)\right]$

$$= \frac{1}{3}\left[3\mu + W_1 + W_2 + W_3\right] = \mu + \frac{1}{3}(W_1 + W_2 + W_3)$$

Where $\frac{1}{3}(W_1 + W_2 + W_3)$ has higher variance than $\frac{1}{3}(-.8W_0 + .2W_1 + .2W_2 + W_3)$

8. Assume that the time series $\{Y_t\}$ is weakly stationary with mean μ and ACVF $\gamma_Y(h)$. Show that the differenced series $X_t = \nabla Y_t = Y_t - Y_{t-1}$ is also stationary and find its ACVF $\gamma_X(h)$.

SOL:

The mean of $\{X_t\}$ is constant equal to $\mathbb{E}(X_t) = \mathbb{E}(Y_t - Y_{t-1}) = \mathbb{E}(Y_t) - \mathbb{E}(Y_{t-1}) = \mu - \mu = 0$.

For the ACVF of $\{X_t\}$ let $t = s + h$, so that

$$\gamma_X(s,t) = \gamma_X(s,+h) = \mathrm{Cov}\left(X_s, X_{s+h}\right) = \mathrm{Cov}\left(Y_s - Y_{s-1}, Y_{s+h} - Y_{s+h-1}\right) =$$
$$= \mathrm{Cov}(Y_s, Y_{s+h}) - \mathrm{Cov}(Y_s, Y_{s+h-1}) - \mathrm{Cov}(Y_{s-1}, Y_{s+h}) + \mathrm{Cov}(Y_{s-1}, Y_{s+h-1}) =$$
$$= \gamma_Y(s, s+h) - \gamma_Y(s, s+h-1) - \gamma_Y(s-1, s+h) + -\gamma_Y(s-1, s+h-1) =$$
$$= \gamma_Y(h) - \gamma_Y(h-1) - \gamma_Y(h-1) + \gamma_Y(h) = 2\gamma_Y(h) - 2\gamma_Y(h-1)$$

The result is a function of the time-lag h, i.e. $\gamma_X(s,t) = \gamma_X(h) = 2\gamma_Y(h) - 2\gamma_Y(h-1)$, so $\{X_t\}$ is also weakly stationary.