# Homework 1
# UMN STAT 5511
# Charles R. Doss

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.

Further instructions:

- You may see `acf` and/or `astsa::acf1`.

Questions:

1. (MA process and ACF) Shumway and Stoffer, Question 1.7. (Page 39.) In addition to plotting the true ACF, you will plot a sample ACF: Generate $n = 100$ observations for $\{W_t\} \sim WN(0, 1)$ (white noise with mean 0 and variance 1). Compute and plot the sample ACF for $X_t = W_{t+1} + 2W_t + W_{t-1}$, and (as the book asks) also plot the true ACF. Compute the sample ACF "by hand".[1] Either make two plots, one above the other (you can use `par(mfrow=)` for instance), or put the two functions on the same plot (visually distinguished in some way).

   [Note: you can use the "type" argument with the plot function to change the way the plot is drawn.]

   **Solution:**

   Assume without loss of generality that $\mathrm{Var}(W_t) = 1$. The linear process $X_t$ is stationary. We have $\mathrm{Cov}(X_t, X_{t-h}) = \mathrm{Cov}(X_t, X_{t+h})$ so we can assume $h > 0$ here.

$$\begin{aligned}\mathrm{Cov}(X_t, X_{t-h}) &= \mathrm{Cov}(2W_t + W_{t-1}, W_{t+1-h} + 2W_{t-h} + W_{t-1-h}) \\ &= 2\,\mathrm{Cov}(W_t, W_{t+1-h}) + \mathrm{Cov}(W_{t-1}, W_{t+1-h}) + 2\,\mathrm{Cov}(W_{t-1}, W_{t-h}) \\ &= 4\mathbb{1}_{(h=1)} + \mathbb{1}_{(h=2)},\end{aligned}$$

   and $\mathrm{Var}(X_t) = 1 + 4 + 1 = 6$. For autocorrelation: $\rho(h) = I_{(h=0)} + \frac{2}{3}I_{(h=1)} + \frac{1}{6}I_{(h=2)}$

2. Shumway & Stoffer, Question 2.1 (page 70, 4th ed.) You can skip the part of the question in (c) that asks "And, by what percentage does it increase or decrease."

   **Solution:**

   **(a)**

[1] Here "by hand" means: do not use acf(), but rather write your own code. (It does not mean that you should actually use a calculator.) You could use acf() to check your results for your own edification, but this should not be included in the output.

```r
set.seed(1)
w<-rnorm(100)
x<-filter(w, sides=2, filter=c(1,2,1))
x<-na.omit(x)
par(mfrow=c(1,1))
##write a function to compute sample acf
sample_cov=function(x,h){
  n=length(x)
  xbar=mean(x)
  sum_term=sum((x[1:(n-h)]-xbar)*(x[(1+h):n]-xbar))
  return(sum_term/n)
}
y<-seq(0,19)
SampleAcf=unlist(lapply(y,sample_cov,x=x))/sample_cov(x,0)
TrueAcf<-c(1,2/3,1/6,rep(0,17))
plot(SampleAcf~y,type="h",ylab="ACF value",xlab="lag",las=1)
lines(TrueAcf~y,type="p")
abline(h=0)
legend("topright", legend = c("True ACF","Sample ACF"),
              pch = c(1, NA),lty=c(NA,1))
```

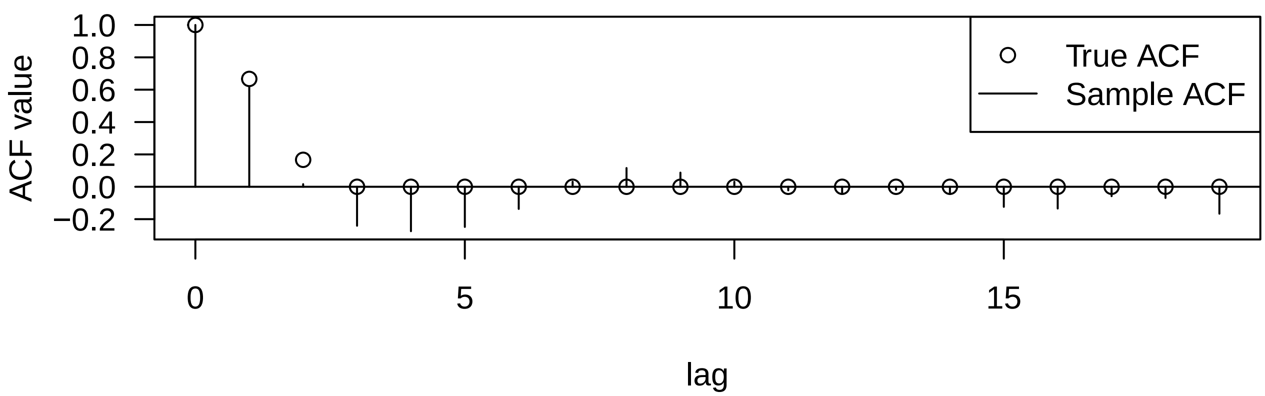
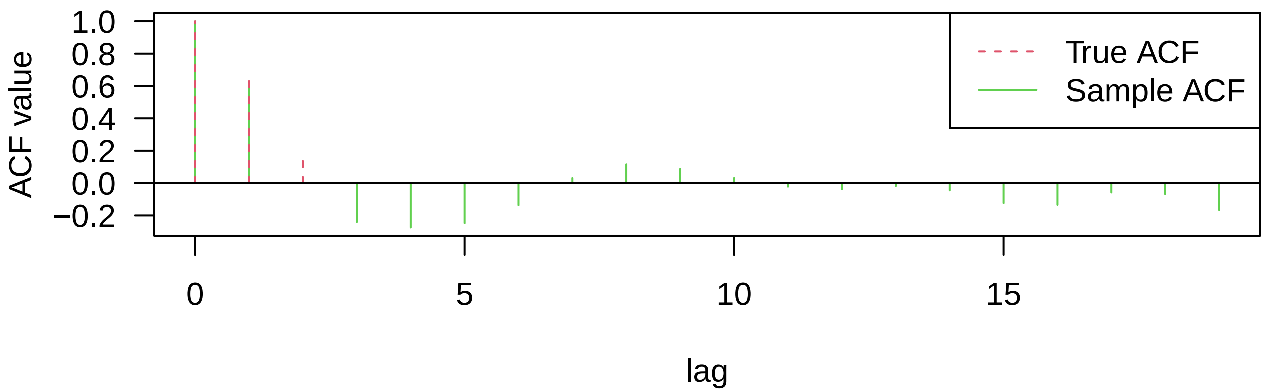



Figure 1: Sample ACF and true ACF

```r
##or distinguish two lines by color
plot(SampleAcf~y,type="h",col=3,ylab="ACF value",xlab="lag",las=1)
lines(TrueAcf~y,type="h",col=2,lty=2)
abline(h=0)
legend("topright", legend = c("True ACF","Sample ACF"),col=c(2,3)
       ,lty=c(2,1))
```




```r
library(astsa)
length(jj)

## [1] 84

trend = time(jj) - 1970 # helps to 'center' time
Q =factor(cycle(jj))# make (Q)uarter factors
reg1 =lm(log(jj)~0 + trend + Q, na.action = NULL)# no intercept
summary(reg1)$coef

##        Estimate  Std. Error  t value     Pr(>|t|)
## trend 0.1671722 0.002259113 73.99902 9.919298e-75
## Q1    1.0527931 0.027359218 38.48038 6.215327e-53
## Q2    1.0809159 0.027365047 39.49987 8.711575e-54
## Q3    1.1510242 0.027382526 42.03499 7.974132e-56
## Q4    0.8822665 0.027411632 32.18584 3.608412e-47
```

### (b)

The estimated average annual increase in the logged earnings per share is given by the coefficient $\beta$. (The variable $y_t$ is the earnings per share, and $x_t = \log(y_t)$, on which we ran the regression, is the logged earnings per share. Note that for any time $t$, according to the model, $E(x_{t+1} - x_t) = \beta$; time is measured at .25 intervals (quarters) but one unit of time is a year.) Our estimate of $\beta$ is 0.167.

### (c)

The estimated change from third quarter to fourth quarter is

$$0.25 * \hat{\beta} + \hat{\alpha}_4 - \hat{\alpha}_3 = -0.227 < 0$$

The percentage of decrease varies for different $t$; the percentage is given by $\frac{0.227}{0.167(t-1970)+\hat{\alpha}_3}$.

Alternative method: We can use $\frac{\hat{\alpha}_4-\hat{\alpha}_3}{\hat{\alpha}_3} = -0.233$ as the overall approximation.

### (d)

```
reg2 =lm(log(jj)~ trend + Q, na.action = NULL) # with intercept
summary(reg2)$coef

##                Estimate  Std. Error    t value     Pr(>|t|)
## (Intercept)  1.05279315 0.027359218 38.4803813 6.215327e-53
## trend        0.16717216 0.002259113 73.9990241 9.919298e-75
## Q2           0.02812273 0.038695899  0.7267626 4.695195e-01
## Q3           0.09823103 0.038708262  2.5377277 1.312480e-02
## Q4          -0.17052665 0.038728858 -4.4030902 3.311318e-05
```

The coefficient for $\alpha_1$ is no longer avaiable. The reason is that if we add intercept to the model, the design matrix will not be full rank. In order to get the unique solution, R drop the first quarter automatically.

### (e)

It is hard to tell from the residual plot directly, but the ACF plot suggests that there the residuals are not white (with seemingly significant correlations at 1 and 2 year lags).

3. (Stationarity) For each of the following, state if it is a stationary process. If so, give the mean and autocovariance functions. Here, $\{W_t\}$ is i.i.d. $N(0, 1)$.

   (a) $X_t = W_t - W_{t-3}$.
   (b) $X_t = W_3$.
   (c) $X_t = t + W_3$.
   (d) $X_t = W_t^2$.
   (e) $X_t = W_t W_{t-2}$.

   **Solution:**

   a Stationary. $EX_t = 0$, and

$$\begin{aligned}
\mathrm{Cov}(X_t, X_{t-h}) &= \mathrm{Cov}(W_t - W_{t-3}, W_{t-h} - W_{t-h-3}) \\
&= \mathrm{Cov}(W_t, W_{t-h}) - \mathrm{Cov}(W_t, W_{t-h-3}) - \mathrm{Cov}(W_{t-3}, W_{t-h}) + \mathrm{Cov}(W_{t-3}, W_{t-h-3}) \\
&= \mathrm{Cov}(W_t, W_{t-h}) - \mathrm{Cov}(W_{t-3}, W_{t-h}) + \mathrm{Cov}(W_{t-3}, W_{t-h-3}) \\
&= 2\mathbb{1}_{(h=0)} - \mathbb{1}_{(h=3)}
\end{aligned}$$

   b Stationary. $EX_t = 0$ and $\mathrm{Cov}(X_t, X_{t-h}) = \mathrm{Cov}(W_3, W_3) = 1$

```r
par(mfrow=c(1,2))
plot(log(jj), ylab="logged earning",
     las=1,xlab=NULL) # data
lines(fitted(reg1), lty=2) # fitted
legend("topleft", legend = c("Data","Fitted value"),lty=c(1,2))
plot(log(jj)-fitted(reg1),las=1,
     ylab="residual",xlab=NULL)
```

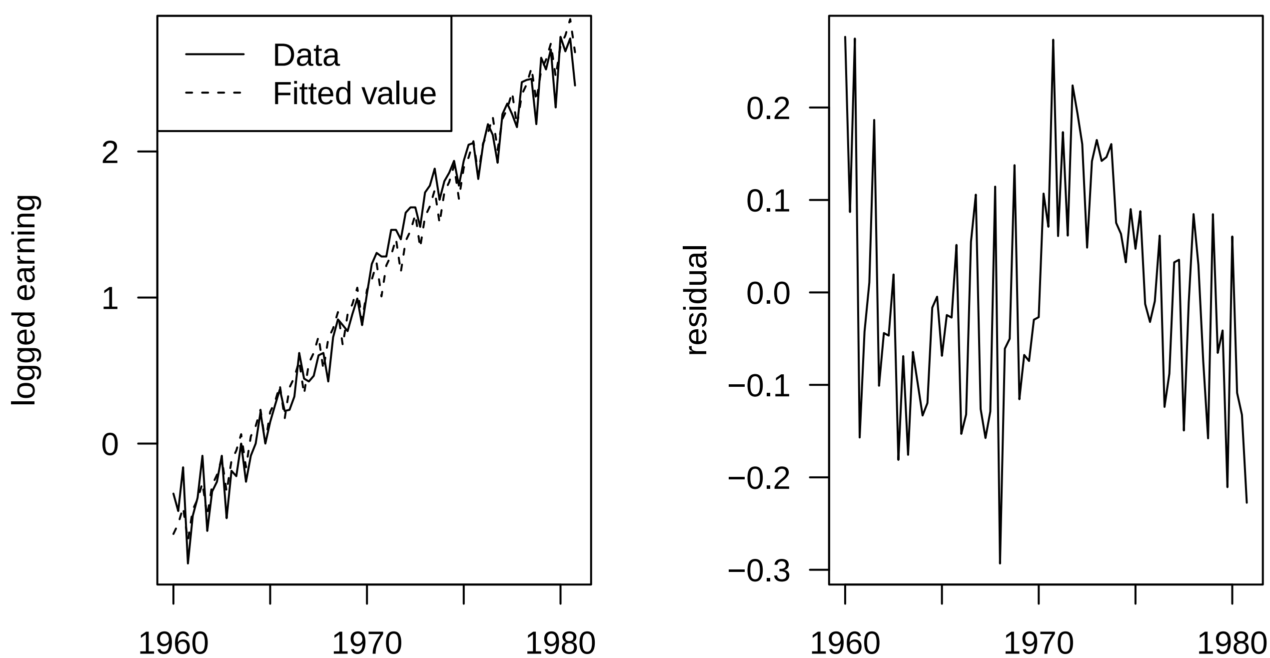



Figure 2: Left: data and fitted value; Right: residuals

```r
par(mfrow=c(1,1))
acf(resid(reg1),las=1,ylab="residual ACF",main="")
```

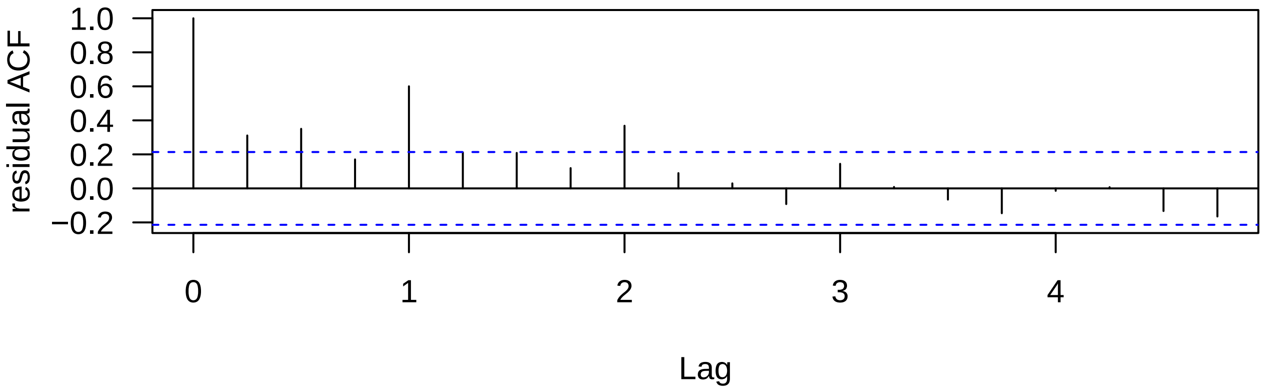



Figure 3: Residual ACF

c Non-stationary, for, e.g., $EX_1 = 1 \neq 2 = EX_2$.

d Stationary. $EX_t = 1$ and $\text{Cov}(X_t, X_{t-h}) = 2\mathbb{1}_{(h=0)}$

e Stationary. $EX_t = 0$ and

$$\begin{aligned}\text{Cov}(X_t, X_{t-h}) &= EX_tX_{t-h} \\ &= EW_tW_{t-2}W_{t-h}W_{t-h-2} \\ &= \mathbb{1}_{(h=0)}\end{aligned}$$

4. Pick *one* of the following two questions to do:

- (White noise is not necessarily i.i.d.). Suppose that $\{W_t\}$ and $\{Z_t\}$ are independent and identically distributed (i.i.d.) sequences, also independent of each other, with $P(W_t = 0) = P(W_t = 1) = 1/2$ and $P(Z_t = -1) = P(Z_t = 1) = 1/2$. Define the time series $X_t$ by $X_t = W_t(1 - W_{t-1})Z_t$. Show that $\{X_t\}$ is white but not i.i.d.

  **Solution:**
  For all sample paths such that $X_{t-1} \neq 0$ it must be that $W_{t-1} = 1$ and, consequently, $X_t = 0$. Thus, $P(X_t = 0, X_{t-1} \neq 0) = P(X_{t-1} \neq 0) = P(W_{t-1} = 1, W_{t-2} = 0) = 1/4 \neq P(X_t = 0)P(X_{t-1} \neq 0) = P(X_t = 0)/4 = 3/16$, where we have used that $W_t$ is i.i.d. and that the distribution of $X_t$ and $X_s$ is the same for every $t$ and $s$ (why?).
  By the assumed independence and the fact that $EZ_t = 0$, we have $EX_t = (EW_t)(1-EW_{t-1})EZ_t = 0$ and, for $h > 0$ (why does it suffice to check only $h > 0$?),

$$\begin{aligned}\text{Cov}(X_t, X_{t-h}) &= EX_tX_{t-h} \\ &= EW_t(1 - W_{t-1})Z_tW_{t-h}(1 - W_{t-h-1})Z_{t-h} \\ &= E\left(W_t(1 - W_{t-1})Z_tW_{t-h}(1 - W_{t-h-1})\right)EZ_{t-h} \\ &= E\left(W_t(1 - W_{t-1})Z_tW_{t-h}(1 - W_{t-h-1})\right) \times 0.\end{aligned}$$

Lastly, $\mathrm{Var}(X_t) = EX_t^2 = EW_t^2 E(1-W_{t-1})^2 EZ_t^2 = EW_t^2 E(1-W_t)^2 EZ_t^2$ does not depend on $t$ since the moments of $\{W_t\}$ and $\{Z_t\}$ do not. So, $X_t$ is white noise but not i.i.d.

- Shumway and Stoffer, Question 1.25 (page 42).

  **Solution:**

  (a) Prove that $\gamma(h)$ for a stationary process is a nonnegative definite function.
  Note that $G := (\gamma(t_j - t_i))_{i,j}$, $i, j = 1, \dots, n$, is the covariance matrix of the vector $(X_{t_1}, \dots, X_{t_n})$ and so is positive semi definite. (One can check that a covariance matrix $\Sigma$ of any random vector $Y \in \mathbb{R}^p$ is positive semidefinite by seeing that $0 \le \mathrm{Var}(a'Y) = a'\Sigma a$.)

  **b**

  The sample autocovariance is

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}(X_{t+h} - \bar{X})(X_t - \bar{X})$$

$$\hat{\Gamma} = \begin{pmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(n-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(n-2) \\ \vdots & \vdots & \cdots & \vdots \\ \hat{\gamma}(n-1) & \hat{\gamma}(n-2) & \cdots & \hat{\gamma}(0) \end{pmatrix}$$

  Let

$$M = \begin{pmatrix} 0, 0, \cdots, 0, 0, X_1 - \bar{X}, X_2 - \bar{X}, \cdots, X_n - \bar{X} \\ 0, 0, \cdots, 0, X_1 - \bar{X}, X_2 - \bar{X}, \cdots, X_n - \bar{X}, 0 \\ \vdots \\ X_1 - \bar{X}, X_2 - \bar{X}, \cdots, X_n - \bar{X}, 0, 0, \cdots, 0, 0 \end{pmatrix}$$

  The $(i, j)$th element of $\hat{\Gamma}$ is

$$\hat{\gamma}(i, j) = \frac{1}{n} M_{i\cdot} M_{j\cdot},$$

  where $M_{i\cdot}$ is the $i$th row of $M$.
  So we have

$$\hat{\Gamma} = \frac{1}{n} M M'$$

  Hence, we have

$$a'\hat{\Gamma}a = \frac{1}{n} a' M M' a = \frac{1}{n}\|a'M\|_2^2 \ge 0$$

  Therefore, the sample autocovariance matrix is nonnegative definite.

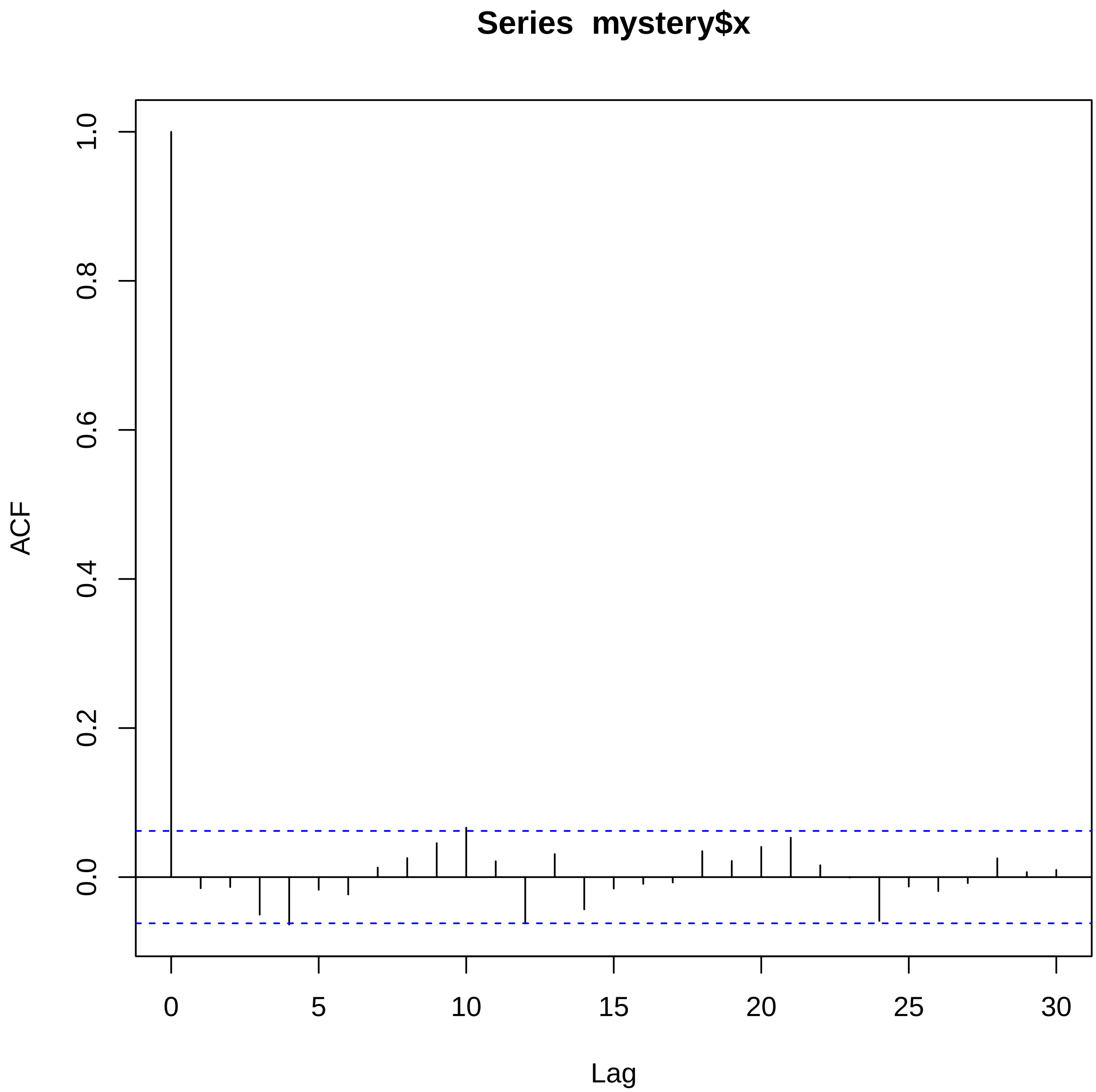
5. ("Plot your data!" [This question is due to Andy Poppick.]) Find the file "mystery.csv" which contains a data series of mysterious origin.

   (a) Plot the sample ACF of the data series. Based on the sample ACF, suggest a model for these data.

   **Solution:**

   The ACF suggests white noise; the correlations are small.

```
mystery  = read.csv('mystery.csv', header=TRUE);
acf(mystery$x)
```

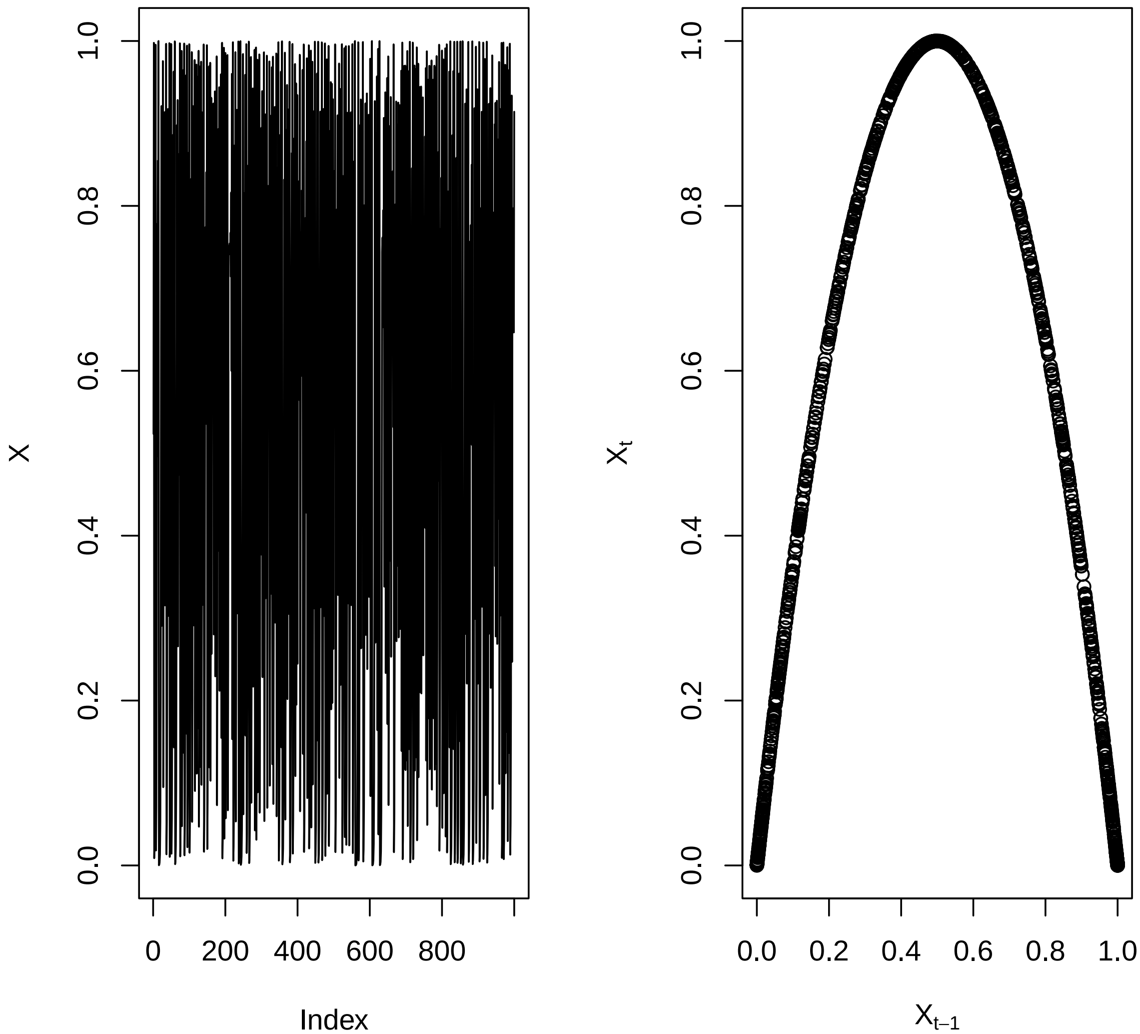



(b) Plot your data: make a time series plot (a plot of the series over time) and a "lag-1" plot (meaning plot $X_t$ against $X_{t-1}$). Do you still believe your answer to the previous part?

**Solution:**

The series is in fact deterministic, there is no noise at all, but the relationship is nonlinear.

```
par(mfrow=c(1,2))
plot(mystery$x, type="l", main="", ylab="X")
plot(mystery$x[1:999], mystery$x[2:1000], ylab=expression(X[t]), xlab=expression(X[t-1]
```


X
Index
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
X_t
X_{t−1}