Homework 7

Ron Cordell, Lei Yang, Subhashini Raghunathan

March 23, 2016

Question 1:

1.1 Load hw07__series1.csv 1.2 Describe the basic structure of the data and provide summary statistics of the series

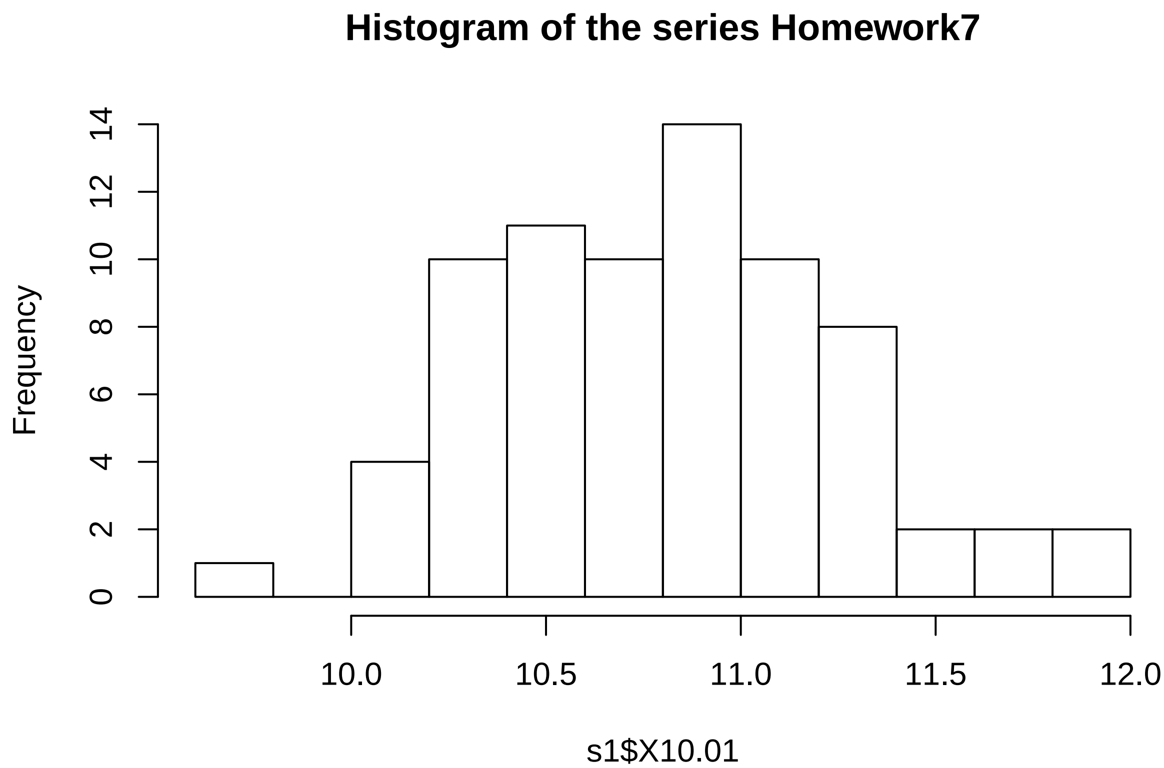
```
setwd('C:/Subha/WS271-Regression/Labs/data')

#read the series from file
s1 = read.csv("hw07_series1.csv")

#describe the basic structure and summary stats
str(s1)
```

```
## 'data.frame':    74 obs. of  1 variable:
##  $ X10.01: num  10.07 10.32 9.75 10.33 10.13 ...
```

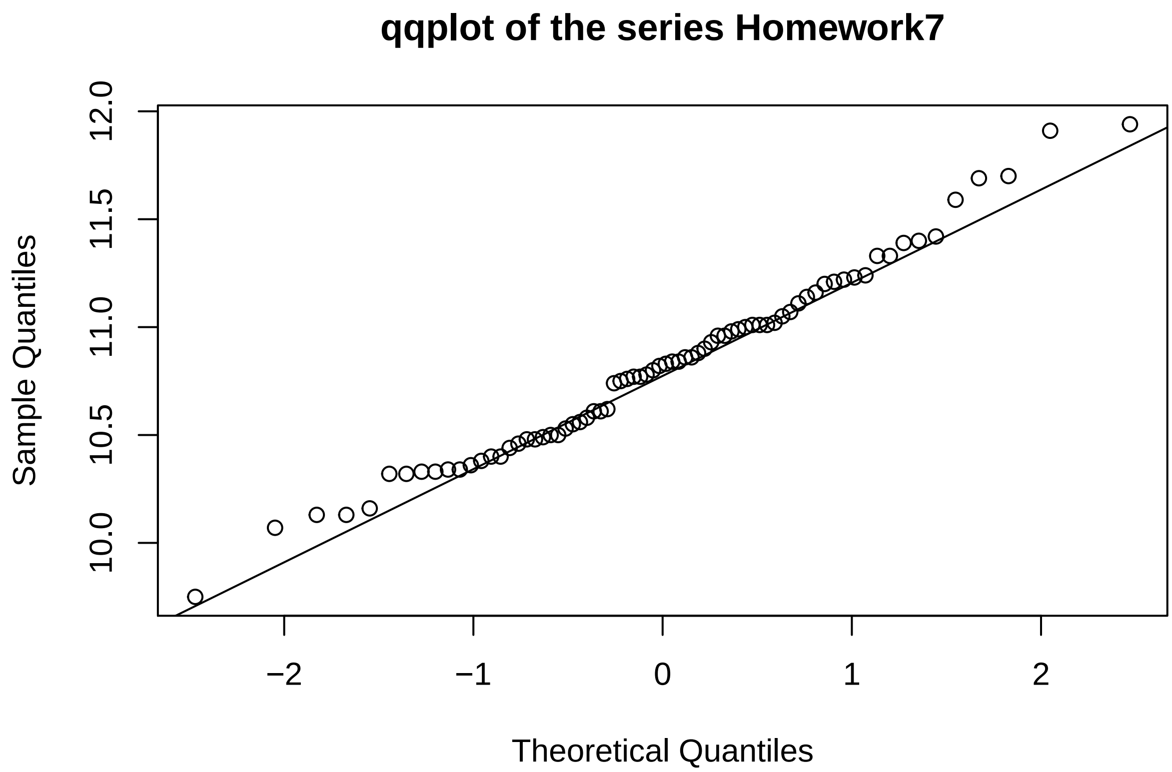
```
summary(s1)
```

```
##      X10.01
##  Min.   : 9.75
##  1st Qu.:10.48
##  Median :10.82
##  Mean   :10.82
##  3rd Qu.:11.06
##  Max.   :11.94
```

```
hist(s1$X10.01, breaks="FD", main="Histogram of the series Homework7")
```



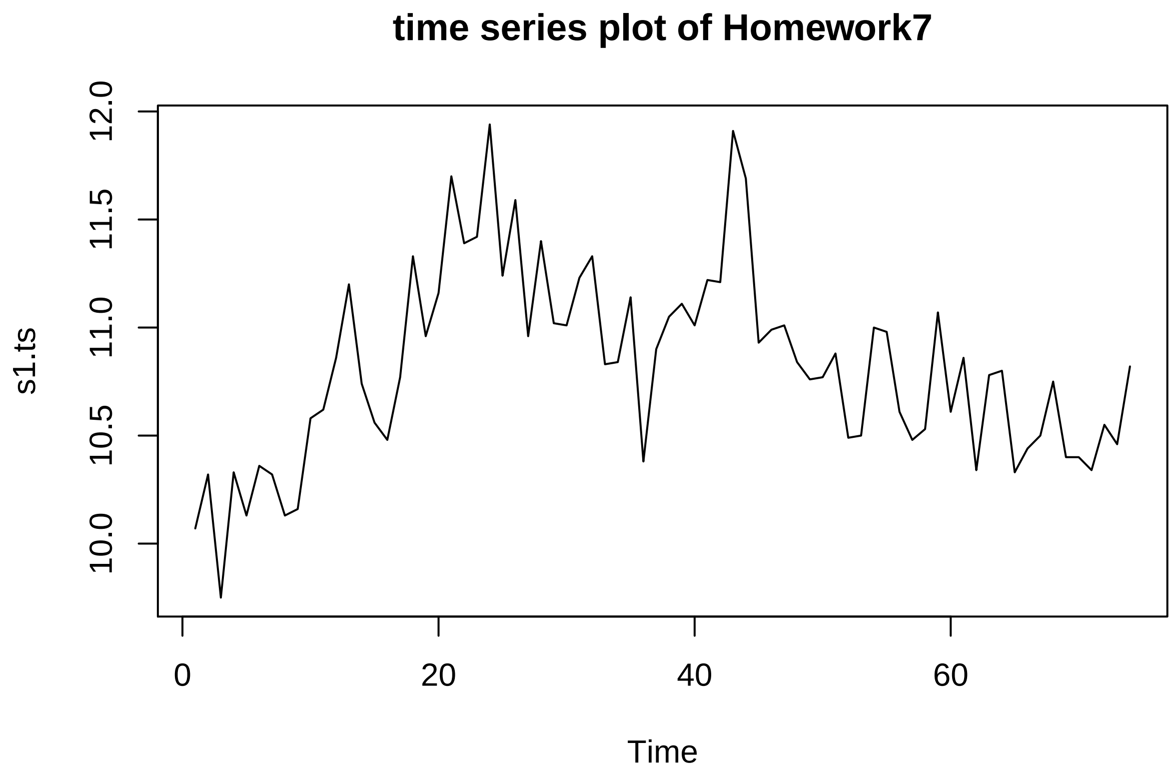

```r
qqnorm(s1$X10.01,main="qqplot of the series Homework7", type ="p")
qqline(s1$X10.01)
```



- Basic structure - the series has 74 observations with no frequncy. The data are somewhat normally distributed with a positive skew.

1.3 Plot histogram and time-series plot of the series. Describe the patterns exhibited in histogrram and time-series plot. For time series analysis, is it sufficient to use only histogram to describe a series?

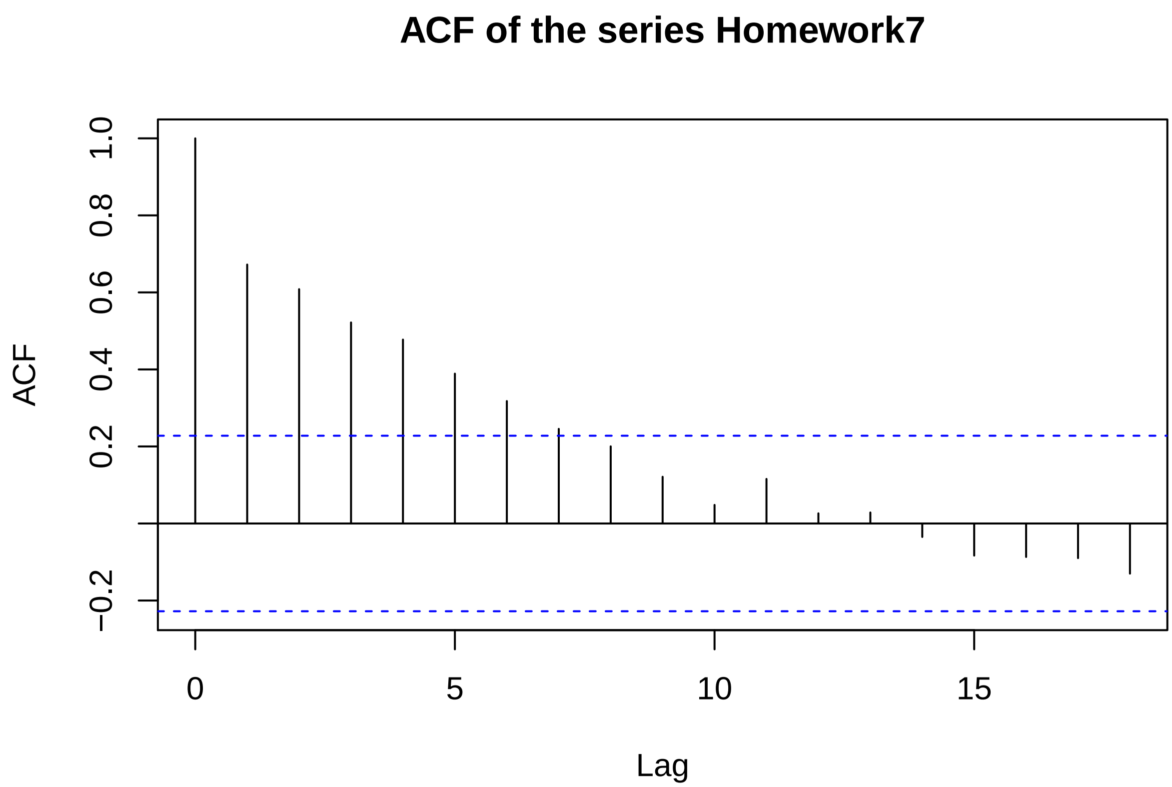
```
#plot time series
s1.ts= ts(data=s1$X10.01)
plot(s1.ts, main="time series plot of Homework7")
```



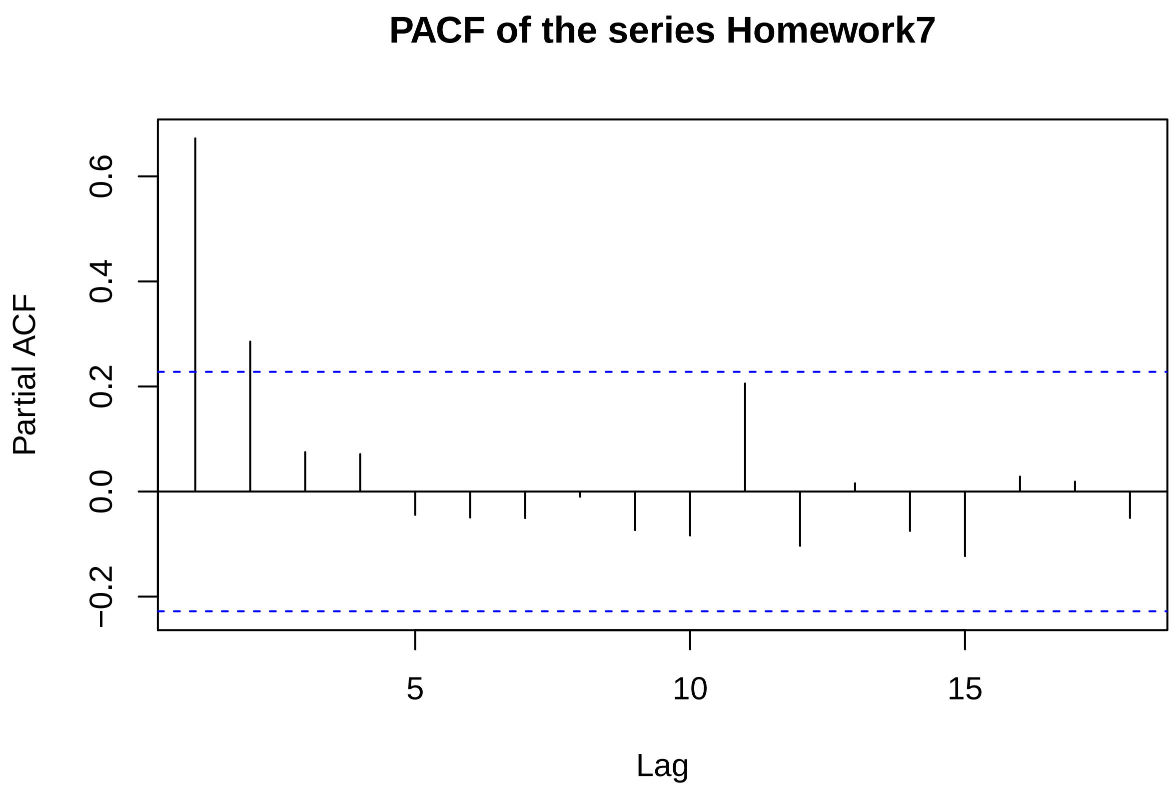
- Patterns - the time series plot has upwards and downwards trends that are somewhat persistent. It resembles a random walk model.
- There is no frequency attached to the time series, hence it is not possible to run decompose() and examine trends and seasonal variations.
- For time series analysis, histogram alone is not enough. ACF and PACF graphs much more important, as they show the dependency structures in the data.

1.4 Plot the ACF and PACF of the series. Describe the patterns exhibited in the ACF and PACF.

```
acf(s1.ts, main="ACF of the series Homework7")
```




```
pacf(s1.ts, main="PACF of the series Homework7")
```



- The ACF is significant until lag 7. It gradually drops to 0.
- The PACF abruptly drops off to 0 after the second lag. This is indicative of an AR(2) model.

1.5 Estimate the series using the ar() function. 1.6 Report the estimated AR parameters, the order of the model, and standard errors.

```
s1.arfit <- ar(s1.ts, method = "mle")
str(s1.arfit)
```

```
## List of 14
##  $ order       : int 2
##  $ ar          : num [1:2] 0.482 0.305
##  $ var.pred    : num 0.0919
##  $ x.mean      : num 10.8
##  $ aic         : Named num [1:13] 49.45 5.01 0 1.33 1.8 ...
##   ..- attr(*, "names")= chr [1:13] "0" "1" "2" "3" ...
##  $ n.used      : int 74
##  $ order.max   : num 12
##  $ partialacf  : NULL
##  $ resid       : Time-Series [1:74] from 1 to 74: NA NA -0.59 0.189 -0.117 ...
##  $ method      : chr "MLE"
##  $ series      : chr "s1.ts"
##  $ frequency   : num 1
##  $ call        : language ar(x = s1.ts, method = "mle")
##  $ asy.var.coef: num [1:2, 1:2] 0.01184 -0.00796 -0.00796 0.01184
##  - attr(*, "class")= chr "ar"
```

```
s1.arfit$order # order of the AR model with lowest AIC
```

```
## [1] 2
```

```
s1.arfit$ar    # parameter estimate
```

```
## [1] 0.4821482 0.3049637
```

```
s1.arfit$aic   # AICs of the fit models
```

```
##         0         1         2         3         4         5         6
## 49.447162  5.009512  0.000000  1.333477  1.801587  3.756255  5.687737
##         7         8         9        10        11        12
##  7.438148  9.436637 11.026353 12.580168 11.691192 12.148725
```

```
sqrt(s1.arfit$asy.var) # asy. standard error
```

```
## Warning in sqrt(s1.arfit$asy.var): NaNs produced
```

```
##           [,1]      [,2]
## [1,] 0.1088067       NaN
## [2,]       NaN 0.1088067
```

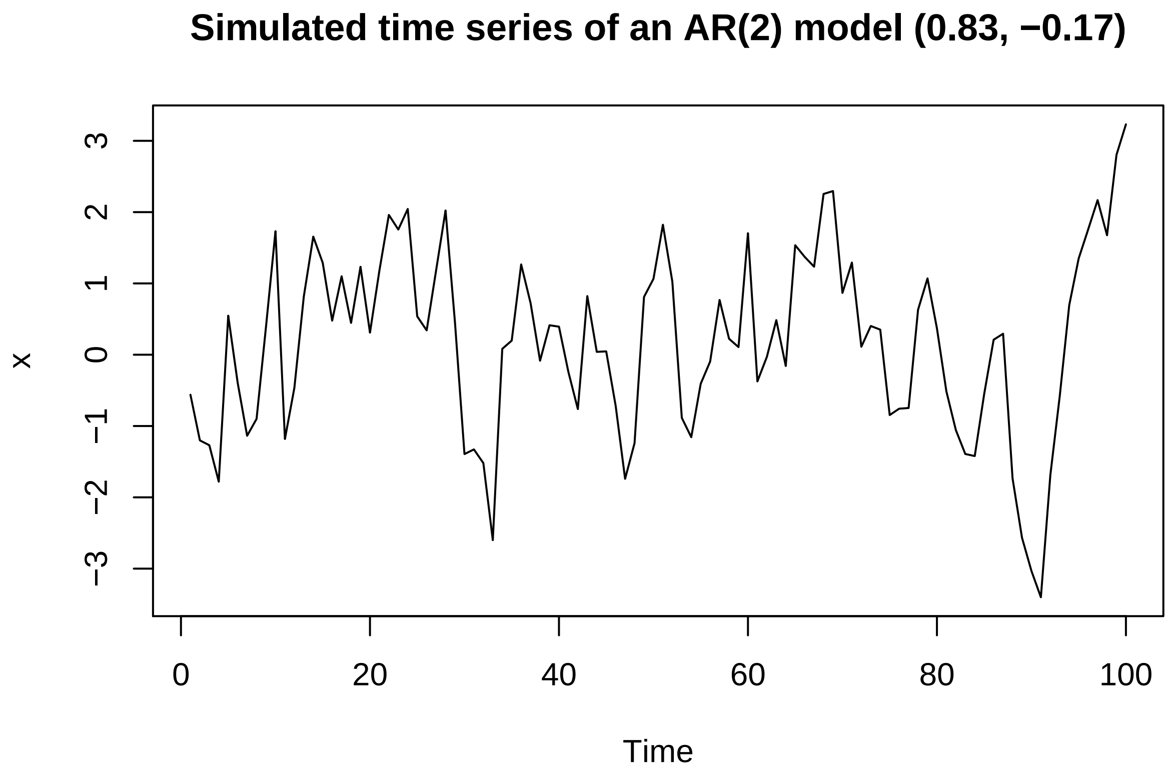
- The estimated model is AR(2): **x[t] = 0.48x[t-1] + 0.3x[t-2] + w[t]**
- The model is order: 2 with a standard error of 0.11 for each parameter.

Question 2:

2.1 Simulate a time series of lenght 100 for the following model. Name the series x. x[t] = 5/6 x[t???1] ??? 1/6 x[t???2] + w

2.2 Plot the correlogram and partial correlogram for the simulated series. Comments on the plots.

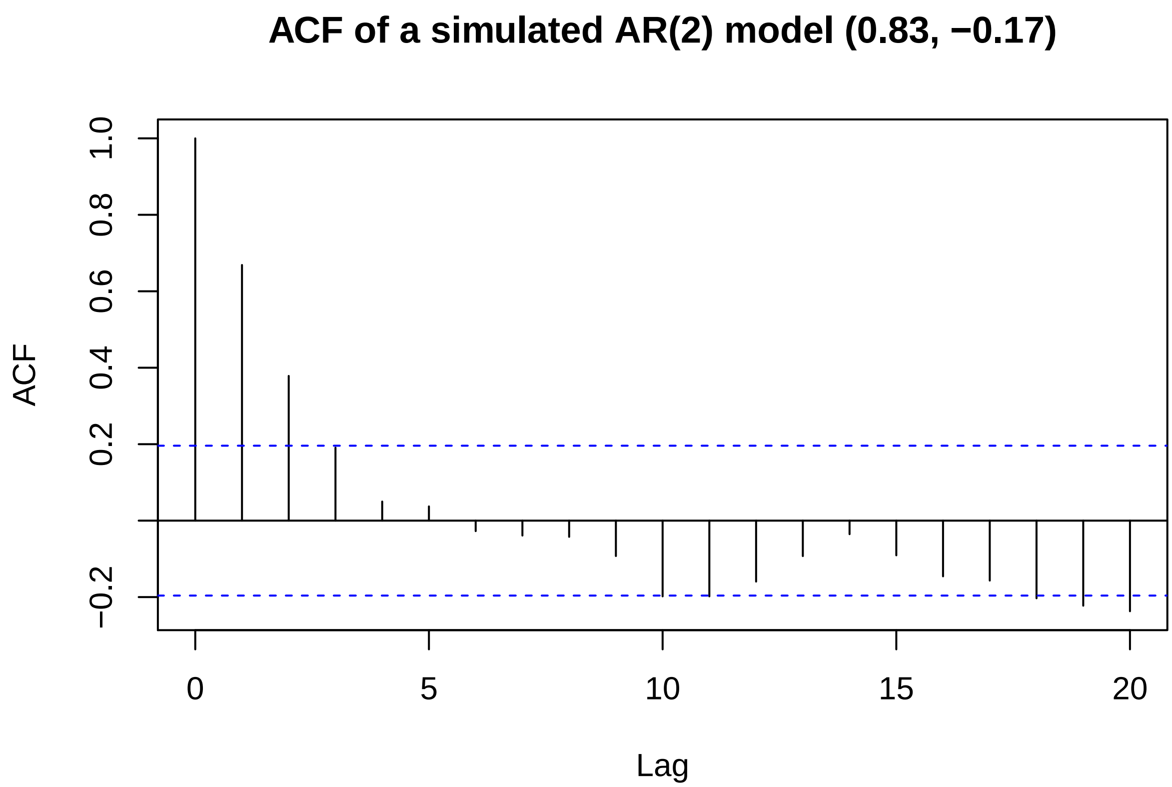
```
set.seed(898)
x = arima.sim(n = 100, list(ar = c(0.83, -0.17)))

plot(x, main="Simulated time series of an AR(2) model (0.83, -0.17)")
```



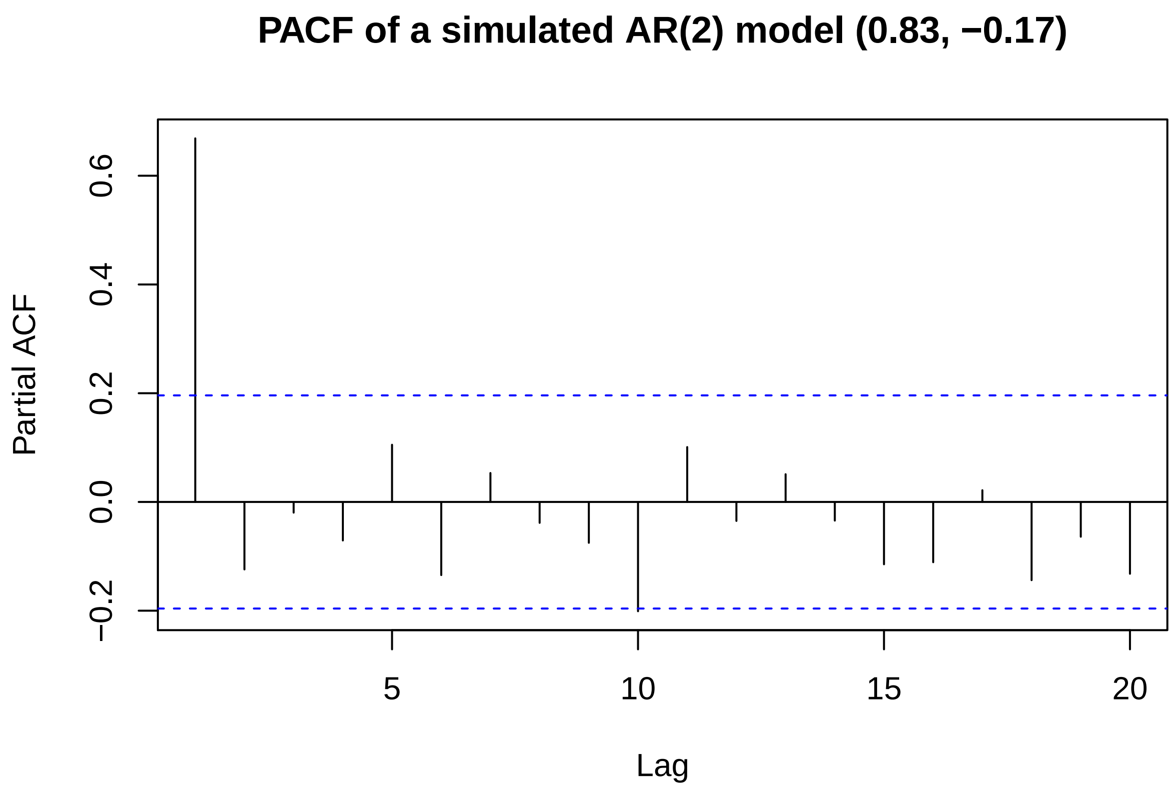

```r
str(x)
```

```
## Time-Series [1:100] from 1 to 100: -0.561 -1.202 -1.269 -1.781 0.547 ...
```

```r
acf(x, main="ACF of a simulated AR(2) model (0.83, -0.17)")
```




```r
pacf(x, main = "PACF of a simulated AR(2) model (0.83, -0.17)")
```



- The ACF graph has 3 significant lags (the 4th one is marginally significant) and gradually drops off to zero.
- The ACF has some periodicity but not the damped cosine function typical of an AR model.
- The PACF graph has 1 significant lags and drops off significantly after that. This is suggestive of an AR(1) model.But we know our true modell is AR(2). The difference could be due to sampling variations.

2.3 Estimate an AR model for this simulated series. Report the estimated AR parameters, standard erros, and the order of the AR model.

```
x.arfit <- ar(x, method = "mle")
str(x.arfit)
```

```
## List of 14
##  $ order       : int 1
##  $ ar          : num 0.704
##  $ var.pred    : num 0.863
##  $ x.mean      : num 0.229
##  $ aic         : Named num [1:13] 61.912 0 0.688 2.62 4.405 ...
##   ..- attr(*, "names")= chr [1:13] "0" "1" "2" "3" ...
##  $ n.used      : int 100
##  $ order.max   : num 12
##  $ partialacf  : NULL
##  $ resid       : Time-Series [1:100] from 1 to 100: NA -0.875 -0.491 -0.955 1.733 ...
##  $ method      : chr "MLE"
```

```
##  $ series      : chr "x"
##  $ frequency   : num 1
##  $ call        : language ar(x = x, method = "mle")
##  $ asy.var.coef: num [1, 1] 0.00524
##  - attr(*, "class")= chr "ar"
```

```r
x.arfit$order # order of the AR model with lowest AIC
```

```
## [1] 1
```

```r
x.arfit$ar    # parameter estimate
```

```
## [1] 0.7038557
```

```r
x.arfit$aic   # AICs of the fit models
```

```
##          0          1          2          3          4          5
## 61.9115542  0.0000000  0.6876973  2.6201961  4.4048578  5.2159341
##          6          7          8          9         10         11
##  5.2210686  6.9174319  8.5366452  9.1276604  1.7961945  2.3821580
##         12
##  4.1830824
```

```r
sqrt(x.arfit$asy.var) # asy. standard error
```

```
##            [,1]
## [1,] 0.07239899
```

```r
x.arfit$aic
```

```
##          0          1          2          3          4          5
## 61.9115542  0.0000000  0.6876973  2.6201961  4.4048578  5.2159341
##          6          7          8          9         10         11
##  5.2210686  6.9174319  8.5366452  9.1276604  1.7961945  2.3821580
##         12
##  4.1830824
```

- The ar() function estimated an AR(1) model but we can see that the AIC difference for order 2 is very small, suggesting that either AR(1) or AR(2) model would be a good fit.

2.4 Construct a 95% confidence intervals for the parameter estimates of the estimated model. Do the "true" mode parameters fall within the confidence intervals? Explain the 95% confidence intervals in this context.

```r
x.arfit$ar + c(2,-2)*sqrt(x.arfit$asy.var)
```

```
## [1] 0.8486537 0.5590578
```

- The confidence interval in this case includes the true population parameter alpha1 = 0.83. The model was estimated as AR(1), hence the second true population parameter was not estimated.

- The 95% confidence interval is the interval within which, for 95% of the samples, the true population parameters lie. The interpretation is the same in the context of this problem.

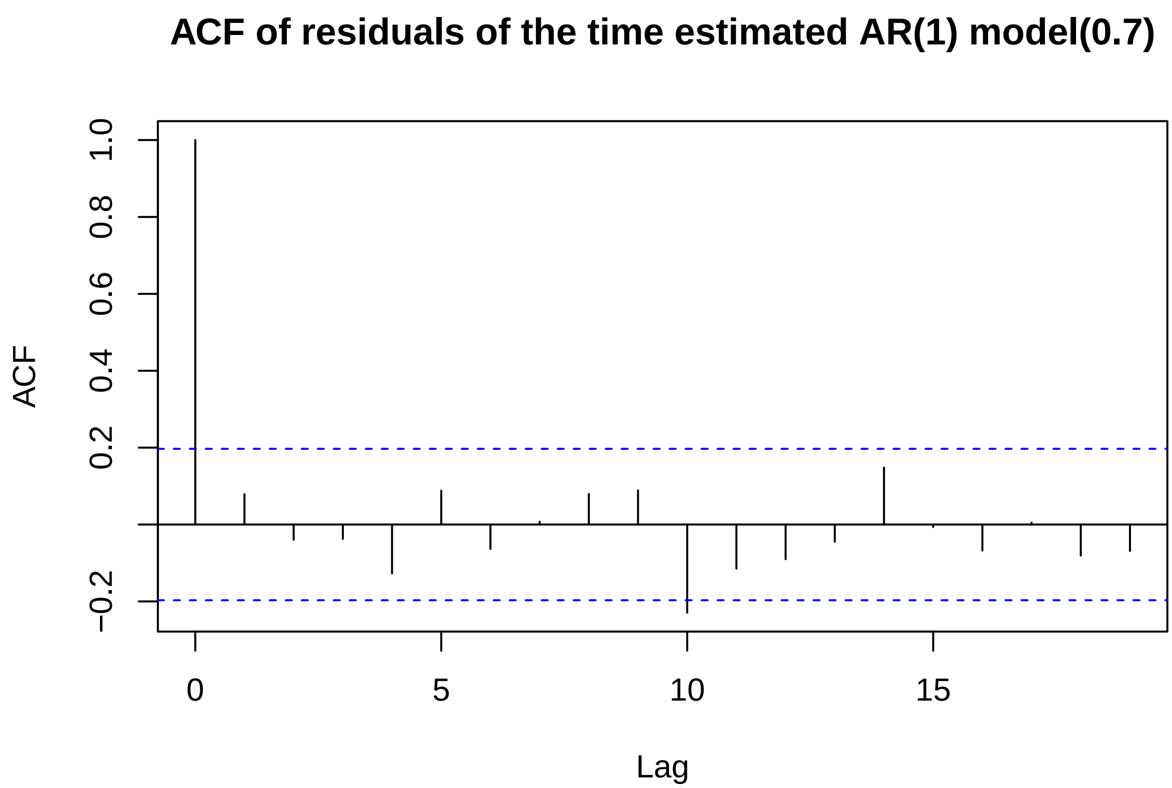
2.5 Is the estimated model stationary or non-stationary?

The estimated model is AR(1): x[t] = 0.7x[t-1] + w

The characteristic equation is (1- 0.7*B) = 0, and the root of the is B = 1/0.7 = 1.4 As the root is > 1, the estimated model is stationary.

2.6 Plot the correlogram of the residuals of the estimated model. Comment on the plot.

```
acf(x.arfit$resid[-(1:x.arfit$order)], main="ACF of residuals of the time estimated AR(1) model(0.7)")
```



- The ACF of the residuals resembles that of white noise. This indicates that the estimated AR(1) model is a good fit to the data.