STAT 436 / 536 - Lecture 8: Key

September 28, 2018

Autoregressive Models

- The random walk model can be written more generally as

$$x_t = \alpha x_{t-1} + w_t,$$

where $\alpha = 1$. In the general case, this is known as an autoregressive model.

- If a time series can be written as

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p}$$

then it is known as an autoregressive process of order p, also denoted by AR(p)

- The AR model can also be written in terms of the backward shift operator **B**.

$$\theta_p(\boldsymbol{B}) x_t = (1 - \alpha_1 \boldsymbol{B} - \alpha_2 \boldsymbol{B}^2 - \cdots - \alpha_p \boldsymbol{B}^p) x_t = w_t$$

- We have seen that the random walk is a special case of an AR(1) model. *The exponential smoothing model is also a special case where $\alpha_i = \alpha(1-\alpha)^i$ for $i = 1, 2, \ldots$ and $p \to \infty$.*

- The name autoregressive comes from the fact that the model is a regression of x_t on past terms.

- The prediction (of a point estimate) at time t is given by plugging in point estimates for the α values.

$$\hat{x}_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p}$$

- Stationarity of the AR process can be determined using the $\theta_p(\boldsymbol{B})x_t$ representation of the series, where $\boldsymbol{B}$ is treated as a number. This equation is known as the characteristic equation.
  - The roots of the characteristic equation determine the stationarity of the series. The absolute value of all of the roots must be greater than one for stationarity.

  - Consider the AR(1) model, $x_t = \frac{1}{2}x_{t-1} + w_t$

$$(1 - \frac{1}{2}\boldsymbol{B})x_t = 0$$

$$1 - \frac{1}{2}\boldsymbol{B} = 0$$

    thus $\boldsymbol{B} = 2$ and we have stationarity

  - Consider the AR(2) model, $x_t = x_{t-1} + \frac{1}{4}x_{t-2} + w_t$

$$(1 - \boldsymbol{B} - \frac{1}{4}\boldsymbol{B}^2)x_t = 0$$

$$\frac{1}{4}(\boldsymbol{B}^2 - 4\boldsymbol{B} + 4) = 0$$

$$\frac{1}{4}(\boldsymbol{B} - 2)^2 = 0$$

    so both roots are equal to 2 and we have stationarity.
  - Consider the random walk model $x_t = x_{t-1} + w_t$

$$(1 - \boldsymbol{B})x_t = 0$$

$$(1 - \boldsymbol{B}) = 0$$

    so $\boldsymbol{B} = 1$ and this is a non-stationary model.

- For an AR(1) process, $x_t = \alpha x_{t-1} + w_t$, the second order properties are: *mean = 0 and* $\gamma_k = \frac{\alpha_k \sigma^2}{1-\alpha^2}$. Note these are for $|\alpha| < 1$.

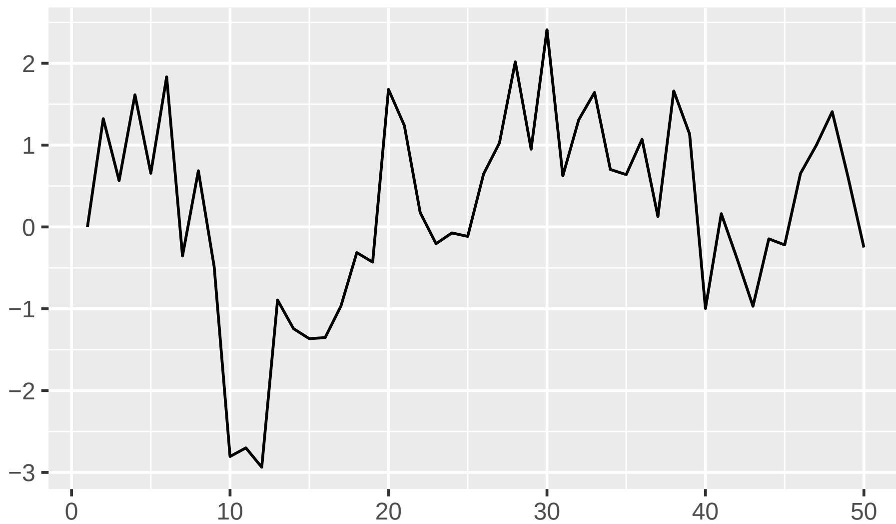
- The autocorrelation function for an AR(1) process is

$$\rho_k = \alpha^k$$

  Thus the autocorrelation decays more quickly with small α.

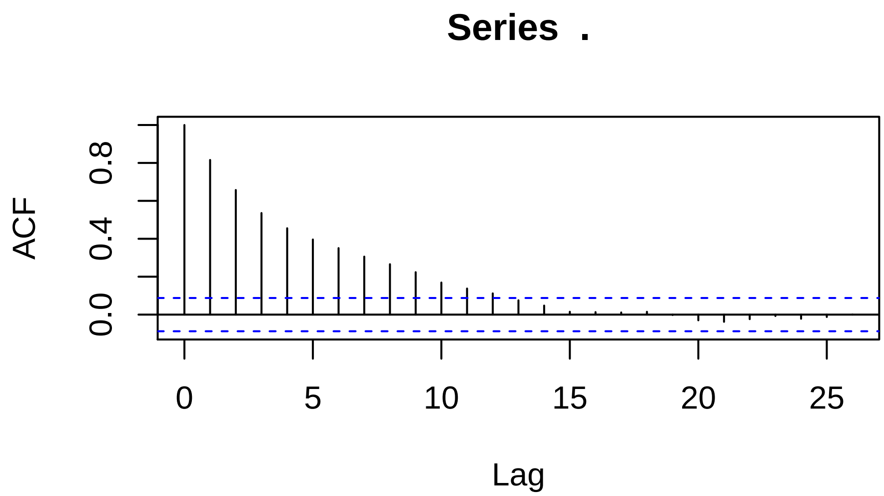
- Write a function to simulate an AR(1) process

```
simAR <- function(alpha, sigma, time.pts){
  # function to simulate and AR process
  # inputs: alpha - the alpha coefficient
  #         : sigma - standard deviation of noise
  #         : time.pts - number of time points
  # outputs: the time series vector as a ts object
  x <- rep(0, time.pts)
  for (t in 2:time.pts){
    x[t] <- alpha * x[t-1] + rnorm(1,0,sigma)
  }
  return(ts(x))
}
ar <- simAR(alpha=.8, sigma=1, time.pts = 50)
library(ggfortify)
library(dplyr)
ar %>% autoplot
```



- Now let's examine the correlogram

```
set.seed(09192018)
ar.series <- simAR(alpha=.8, sigma=1, time.pts = 500)
acf.ar <- ar.series %>% acf
```




```
acf.ar
```

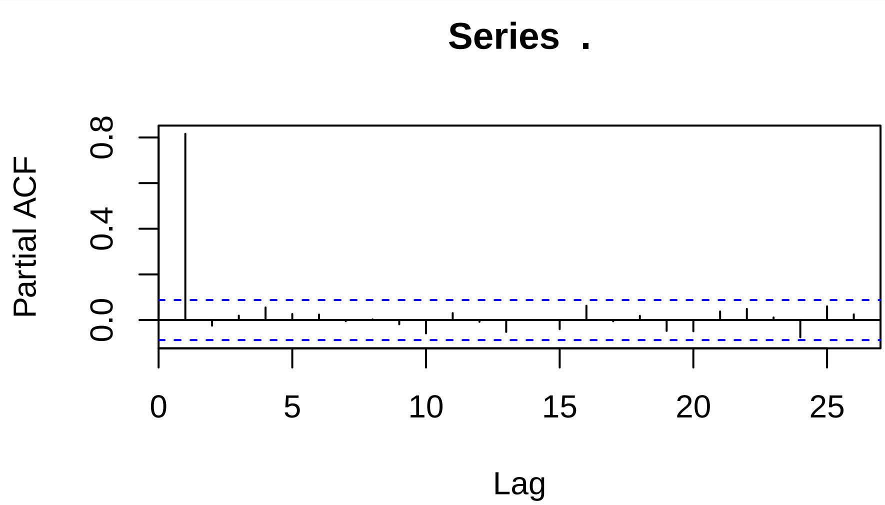
```
##
## Autocorrelations of series '.', by lag
##
##      0      1      2      3      4      5      6      7      8      9
##  1.000  0.816  0.657  0.536  0.456  0.396  0.351  0.306  0.266  0.224
##     10     11     12     13     14     15     16     17     18     19
##  0.169  0.138  0.112  0.076  0.048  0.015  0.013  0.011  0.015 -0.002
##     20     21     22     23     24     25     26
## -0.030 -0.038 -0.024 -0.007 -0.022 -0.012  0.001
```

this is fairly close to the empirical correlation term.

- The autocorrelation will be non-zero for all lags, even though the model for time t only depends on the value from time $t-1$. *So instead of looking at the autocorrelation, we are interested in the partial autocorrelation that results after removing the effect of correlations at the shorter levels.*

- *The partial autocorrelation of an AR(p) process will be the p^{th} coefficient of the fitted model. Hence, it will be zero for all k greater than p.*

```
set.seed(09192018)
ar.series <- simAR(alpha=.8, sigma=1, time.pts = 500)
pacf.ar <- ar.series %>% pacf
```



```
pacf.ar
```

```
##
## Partial autocorrelations of series '.', by lag
##
##      1      2      3      4      5      6      7      8      9     10
##  0.816 -0.025  0.019  0.055  0.027  0.024 -0.005  0.003 -0.019 -0.058
##     11     12     13     14     15     16     17     18     19     20
##  0.031 -0.008 -0.052  0.000 -0.040  0.063 -0.005  0.019 -0.048 -0.049
##     21     22     23     24     25     26
##  0.038  0.049  0.012 -0.076  0.060  0.025
```

- *The PACF is useful for determining the order of an AR process*

- The `ar()` function in R can be used to fit AR models and has several useful properties - *the order of the AR model can be fit using AIC*

- *the AR coefficients can be estimated through several methods*

- *the AR function can be used for forecasting*

```
ar.vals <- ar(ar.series, order.max = 2)
predict(ar.vals, n.ahead = 5)
```

```
## $pred
## Time Series:
## Start = 501
## End = 505
## Frequency = 1
## [1] -0.28474478 -0.22254696 -0.17180572 -0.13041082 -0.09664069
##
## $se
## Time Series:
## Start = 501
## End = 505
## Frequency = 1
## [1] 0.964170 1.244316 1.400032 1.494702 1.554516
```