Oct 25 Notes

Hongdou Li

10/25/2018

Recap:

- strict vs. weak stationary.
- Mean function $\mu(t) = E[Y_t]$
- Covariance function: $\gamma(t, t+h) = Cov(Y_t, Y_{t+h})$

In the case that Y_t is a stationary time series, $\gamma(t, t+h) = \gamma(h)$ is called the **autocovariance function**.

In this context, we define the **autocorrelation function** of lag h to be

$\rho(h) = Corr(Y_t, Y_{t+h})$

$= \frac{Cov(Y_t, Y_{t+h})}{\sqrt{Var(Y_t)Var(Y_{t+h})}}.$

$= \frac{Cov(Y_t, Y_{t+h})}{\sqrt{Cov(Y_t, Y_t)Cov(Y_{t+h}, Y_{t+h})}}$

$\frac{\gamma(h)}{\gamma(0)}$

- Properties:
- $\gamma(0) \geq 0 <=> Var[Y_t] \geq 0$
- $|\gamma(h)| \leq \gamma(0) <=> \frac{|\gamma(h)|}{\gamma(0)} = |\rho(h)| \leq 1$
- $\gamma(h) = \gamma(-h)$, $\rho(h) = \rho(-h)$ (just need to care about the positive side)

Example1:

Recall that if $Y_t \sim MA(1)$,

$$\gamma(h) = Cov(Y_t, Y_{t+h}) = \begin{cases} \sigma^2(1+\theta^2) & h = 0 \\ \sigma^2\theta & h = \pm 1 \\ 0 & otherwise \end{cases}$$

Notice $\gamma(0) = \sigma^2(1+\theta^2)$ and so

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \begin{cases} 1 & h = 0 \\ \theta/(1+\theta^2) & h = \pm 1 \\ 0 & otherwise \end{cases}$$

we can make the ACF plot based on it.

Example2:

First order Autoregressive Model[AR(1)]. Assume Y_t is a stationary times series satisfying the following coditions

$$Y_t = \phi Y_{t-1} + \epsilon_t$$

where $|\phi| < 1$ (we require this for it to be stationary, this is called the stationary condition) and $\epsilon_t \sim WN(0, \sigma^2)$ and ϵ_t and Y_s are uncorrelated for s<t. Calculate $\gamma(h)$ and $\rho(h)$.

- $E[Y_t] = E[\phi Y_{t-1} + \epsilon_t] = \phi E[Y_{t-1}] + E[\epsilon_t]$

Thus, $E[Y_t] = \phi E[Y_{t-1}]$, $\mu = \phi\mu$

Therefore $\mu = 0$. This is the only posibility for the condition to hold.

$$\begin{aligned}
\gamma(h) &= Cov(Y_t, Y_{t+h}) \\
&= E(Y_t Y_{t+h}) - E(Y_t)E(Y_{t+h}) \\
&= E(Y_t Y_{t+h}) \\
&= E[Y_t(\phi Y_{t-1+h} + \epsilon_{t+h})] \\
&= E[\phi Y_t Y_{t-1+h} + Y_t \epsilon_{t+h})] \\
&= \phi E[Y_t Y_{t-1+h}] + E[Y_t \epsilon_{t+h}]
\end{aligned}$$

also known that $E[Y_t \epsilon_{t+h}] = Cov(Y_t \epsilon_{t+h}) = 0$ because current value should not depend on future error. there is no independence between those two thing, so the cov is 0.

$$\begin{aligned}
\gamma(h) = E[Y_t, Y_{t+h}] &= \phi E[Y_t Y_{t-1+h}] \\
&= \phi^2 \gamma(h-2) \\
&= \phi^3 \gamma(h-3) \\
&\dots \\
&= \phi^h \gamma(0)
\end{aligned}$$

$$\begin{aligned}
\gamma(0) = Var[Y_t] = E[Y_t^2] - E[Y_t] &\text{ Known } E[Y_t] = 0 \\
&= E[(\phi Y_{t-1} + \epsilon_t)^2] \\
&= E[\phi^2 Y_{t-1}^2 + 2\phi Y_{t-1}\epsilon_t + \epsilon_t^2] \\
&= \phi^2 E[Y_{t-1}^2] + 2\phi E[Y_{t-1}\epsilon_t] + E[\epsilon_t^2] \\
&= \phi^2 \gamma(0) + 0 + \sigma^2
\end{aligned}$$

$=> \gamma(0) = \frac{\sigma^2}{1-\phi^2}$

Therefore $\gamma(h) = \frac{\phi^{|h|}\sigma^2}{1-\phi^2}$ for $h \in Z$ and $\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \phi^{|h|}, h \in Z$.

we will have a decreasing ACF. we would expect to see a quick exponential decay.

Whereas we've calculated ACF's from specified models, in practice we observe data and calculate sample estimates of the quantities.

Given an observed time series $\{Y_t\} = \{Y_1, Y_2, Y_3 ..., Y_n\}$

- Sample Mean Function: $\hat{\mu(t)} = \bar{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t$
- Sample Autocovariance Function: $\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(Y_t - \bar{Y})(Y_{t+|h|} - \bar{Y})$

The bias is not significant when in a large enough time series, thus we may use n instead of n-1.

- Sample Autocorrelation Function: $\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$

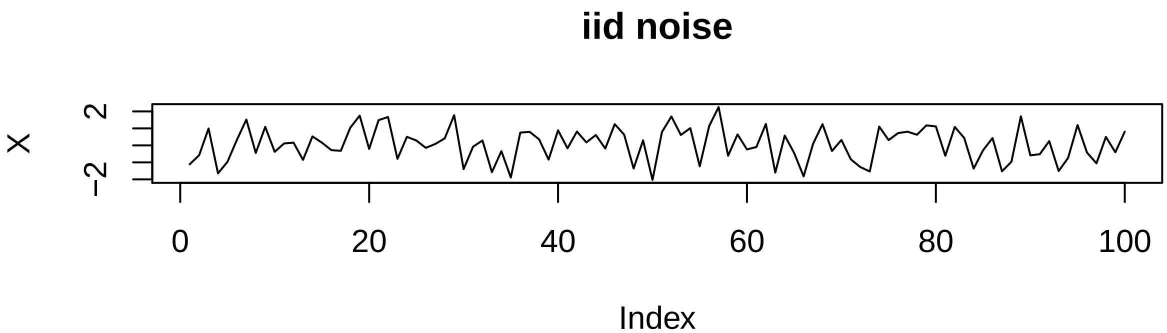
In practice, we use the $\hat{\rho}(h)$ to determine whether an observed time series is correlated. We can determine a threshold where if $\hat{\rho}(h)$ lies beyond this threshold, The correlation at lag h is deemed to be significant.

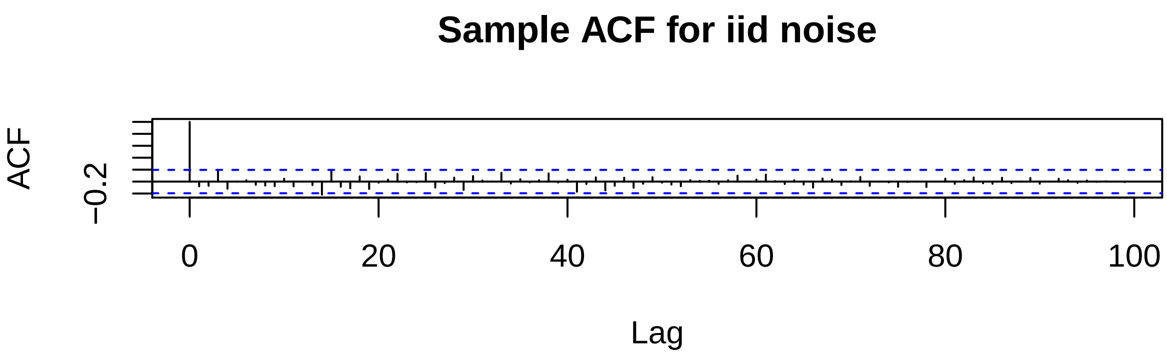
These threshold are calculated in the context of the fact that asymptotically $\tilde{\rho}(h) \sim N(0, \frac{1}{n})$

if the time series is uncorrelated. The threshold is calculated as $\pm 1.96/\sqrt{n}$ and $\hat{\rho}(h) \notin [-1.96/\sqrt{n}, 1.96/\sqrt{n}]$ indicates significant lag h correlation.

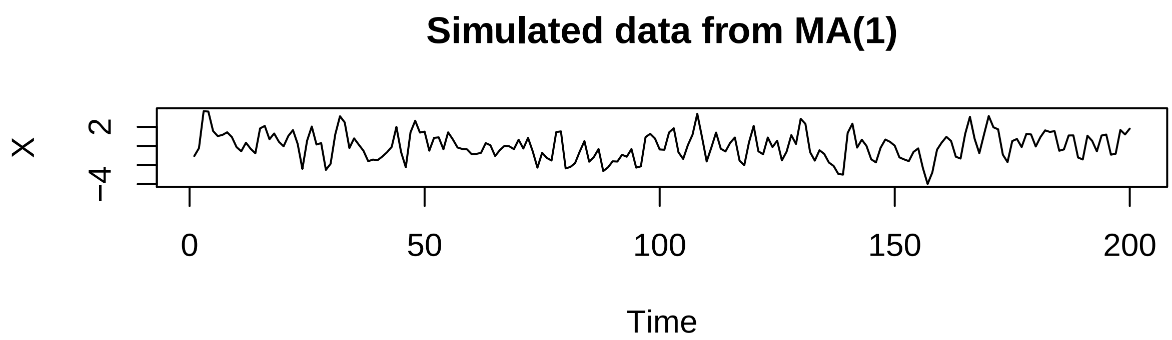
Sample ACF example

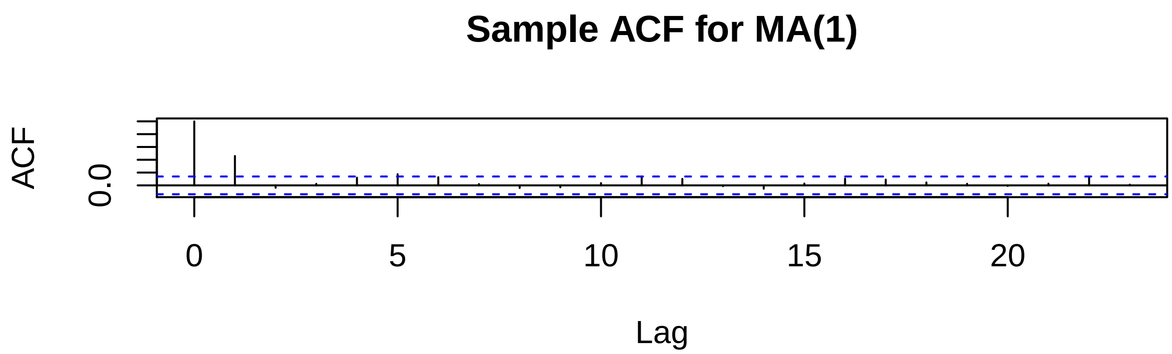
```r
# sample ACF for iid noise N(0,1)
X <- rnorm(100) # generating (independently) 100 realizations of N(0,1)
par(mfrow=c(2,1)) #dividing the page into 2 rows and one column
plot(X,type='l',main='iid noise') #plotting the data
acf(X,main='Sample ACF for iid noise', lag.max = 100) # plotting the acf
```



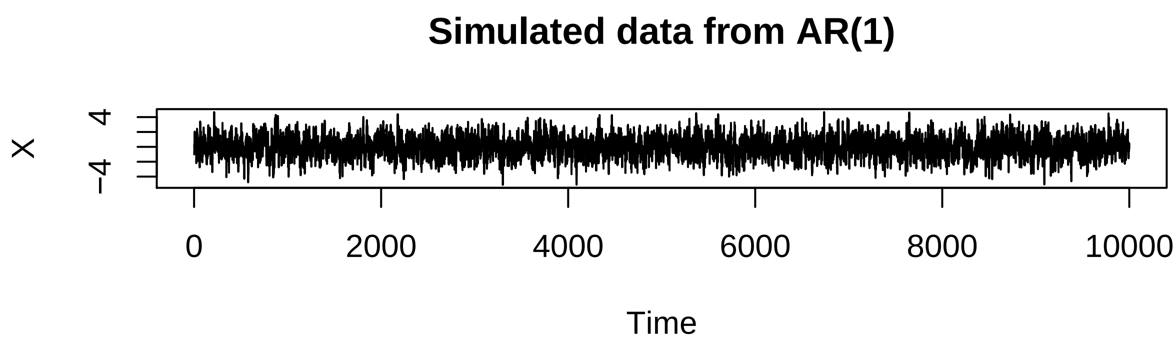


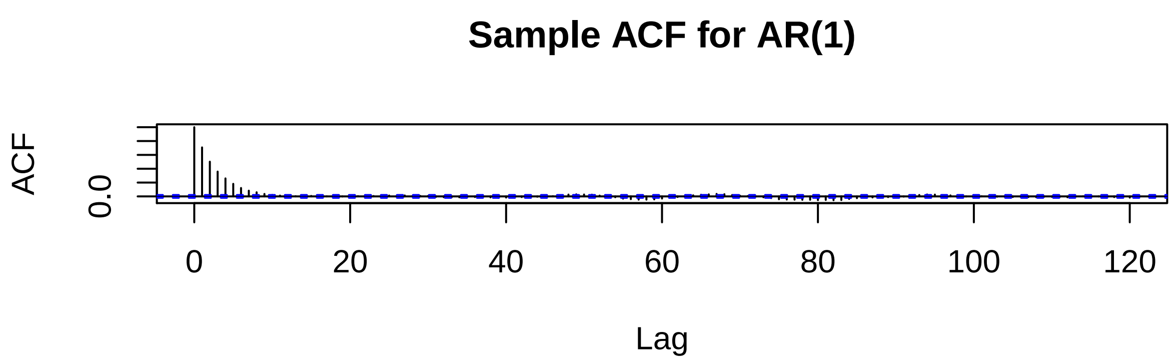

```r
# sample ACF for MA(1) process
X <- arima.sim(list(order = c(0,0,1), ma = 0.85), n = 200) # simulating data from an MA(1) process
par(mfrow=c(2,1)) #dividing the page into 2 rows and one column
plot(X,type='l',main='Simulated data from MA(1)') #plotting the data
acf(X,main='Sample ACF for MA(1)') # plotting the acf
```






```r
# sample ACF for AR(1) process
X <- arima.sim(list(order = c(1,0,0), ar = .7), n = 10000) # simulating data from an AR(1) process
par(mfrow=c(2,1)) #dividing the page into 2 rows and one column
plot(X,type='l',main='Simulated data from AR(1)') #plotting the data
acf(X,main='Sample ACF for AR(1)', lag.max = 120) # plotting the acf
```



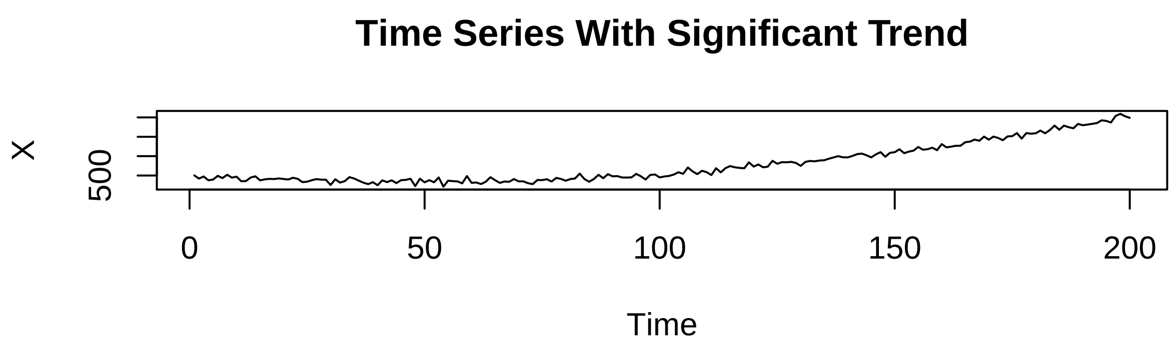


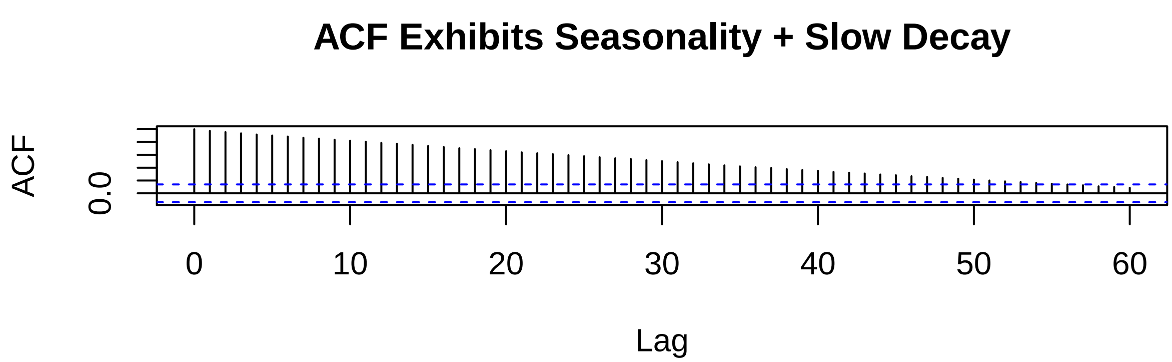
```r
arima(X,order=c(1,0,0))
```

```
##
## Call:
## arima(x = X, order = c(1, 0, 0))
##
## Coefficients:
##          ar1  intercept
##       0.7082    -0.0103
## s.e.  0.0071     0.0338
##
## sigma^2 estimated as 0.9753:  log likelihood = -14064.57,  aic = 28135.14
```

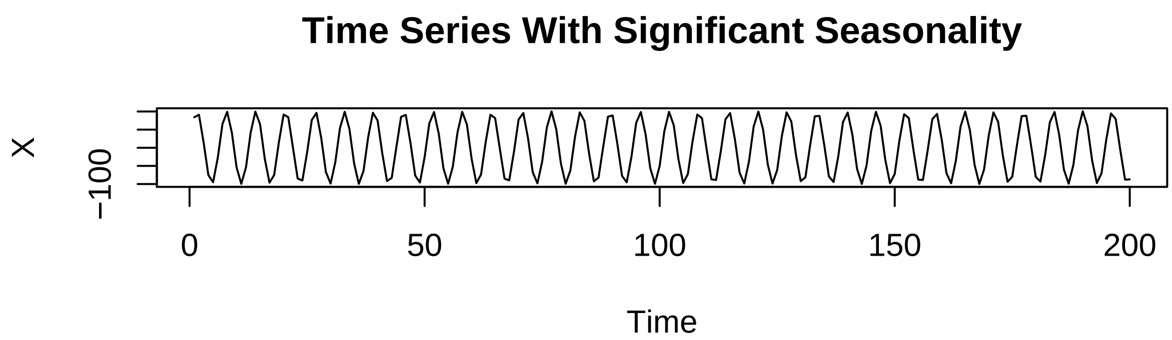
```r
#arima(X,order=c(2,0,0))
```

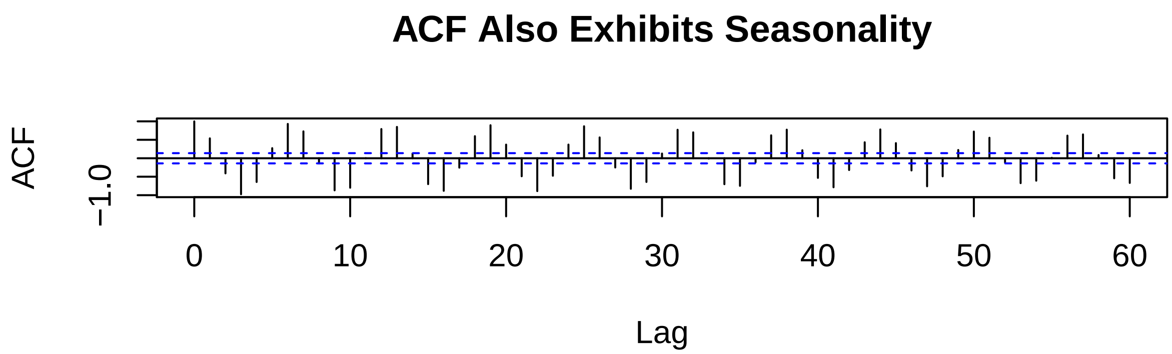
```r
# sample ACF for data with trend
a <- seq(1,100,length=200)
X <- 22-15*a+0.3*a^2+rnorm(200,500,50)
par(mfrow=c(2,1)) #dividing the page into 2 rows and one column
ts.plot(X, main = "Time Series With Significant Trend")
acf(X, main = "ACF Exhibits Seasonality + Slow Decay", lag.max = 60)
```



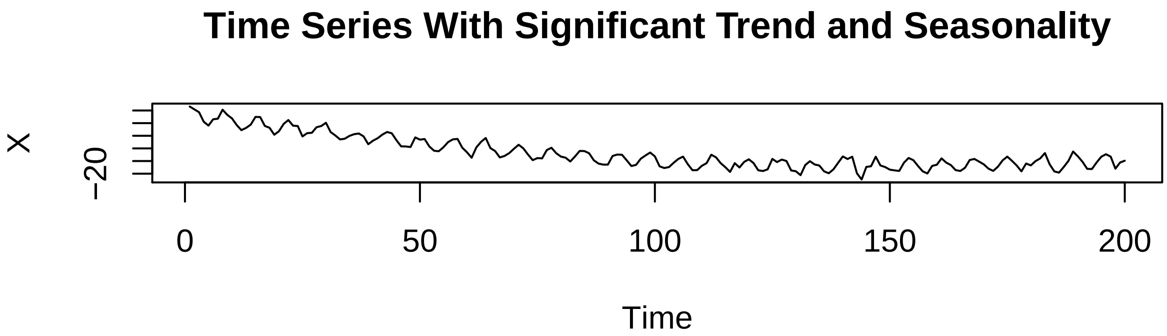


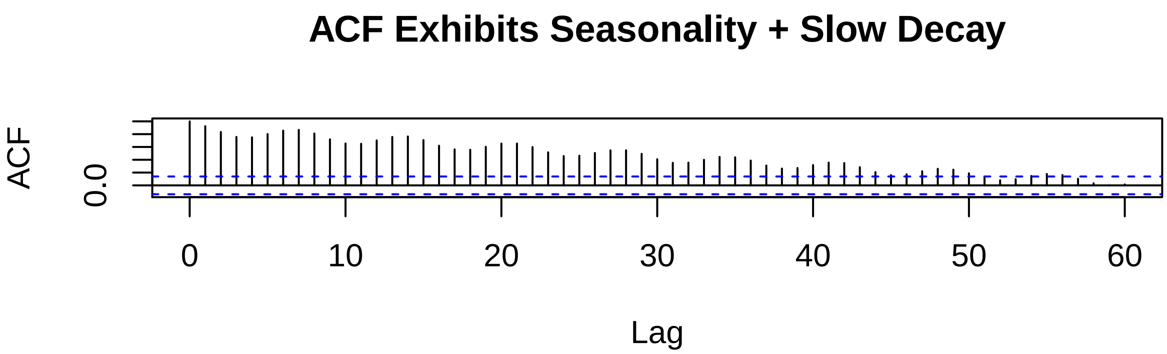

```r
# sanple ACF for data with seasonality
X <- 100*sin(1:200)+rnorm(200, 0, 0.5)
par(mfrow=c(2,1))
ts.plot(X, main = "Time Series With Significant Seasonality")
acf(X, main = "ACF Also Exhibits Seasonality", lag.max=60)
```






```r
# sample ACF for data with trend and seasonal component
a <- seq(1,10,length=200)
X <- 22-15*a+a^2+5*sin(20*a)+rnorm(200,20,2)
par(mfrow=c(2,1)) #dividing the page into 2 rows and one column
ts.plot(X, main = "Time Series With Significant Trend and Seasonality")
acf(X, main = "ACF Exhibits Seasonality + Slow Decay", lag.max = 60)
```






```r
# sample ACF for US Accidental Deaths data (data with seasonality)
par(mfrow=c(2,1)) #dividing the page into 2 rows and one column
plot(USAccDeaths,type='l',main='# of Accidental deaths in US') #plotting the data
acf(USAccDeaths,main='Sample ACF for US accidental deaths data',lag.max=48) # plotting the acf
```