Lecture 1

Joshua Goldberg

April, 08 2019

Introduction

What is time series data: data measured at specific time points

$TS\{Y_t\}$

Two types of time series: - uniform: time intervals constant - non-uniform: time intervals varying

Uniform time series focuses on the forecast.

t is now. $+t$ is a forecast and $-t$ is historical.

Random variable values are defined by a probability distribution.

Two interesting time series: - **White noise** is a collection of uncorrelated random variables $\{W_t\}$ with $\mu = 0$ and variance=σw^2 - **Random walk** $y_t = y_{t-1} + w_t$

Orders of moment:

1. mean
2. variance
3. skew
4. kurtosis

Anytime you deal with stochastic data the mean is expected. If it's deterministic, then you just do an average.

Variance: $E[(x-\mu)^2]$

Covariance: $E[(x-\mu_x)(y-\mu_y)]$

Auto covariance: $\gamma_{s,t} = E[(x_s-\mu_x)(x_t-\mu_x)]$

$$\gamma_k = E[(x_{t+k}-\mu_x)(x_t-\mu_x)]$$

$\gamma_0 = E[(x-\mu)^2]$: this is variance at $\gamma = 0$

Auto correlation: $\rho_{s,t} = \frac{\gamma_{s,t}}{\sqrt{cov(x_s,x_s)cov(x_t x_t)}}$

Stationarity

A time series is stationary means that the data generation process is in statistical equilibrium. The statistical properties (orders) of the data do not change. Non-stationary time series has less tools to work with.

Strict: probability behavior of every TS collection $\{x_{t1}, x_{t2}, x_{t3}, ...x_{tk}\}$ is identical to the time-shifted $\{x_{t1+h}, x_{t2+h}, x_{t3+h}, ...x_{tk+h}\}$

Note: the joint probability distribution of the time series is constant over time.

Weak:

1. $\{y_t\}$ TS finite variance process that has a constant mean that does not change with time.

2. Auto convariance that only depends on the lag k.

Weak + Gaussian distribution = strict.

Gaussian distribution is both weak and strict.

If you have a process that is everything **strict**, but is not finite variance, then it is not weak.

Cauchy example, which has no secord order moments.

Everything in class is weak stationary.

I.I.D.

iid => strict distributions that are independent

But strict != iid due to auto correlation.

White noise

Weak since variance is dependent on lag

Random walk

Non-stationary because auto covariance changes?

EDA

If a time series has some auto correlation, you should be able to write:

$y_t = \rho y_{t-1} + \mu_t$, for $\rho < 1$

If $\rho = 1$, this becomes a random walk (non-stationary)

Tests

Augmented Dickey Fuller (ADF) test is a hypothesis testing that says $H_0 = \rho = 0$: non-stationary and $H_a = \rho < 0$: stationary.

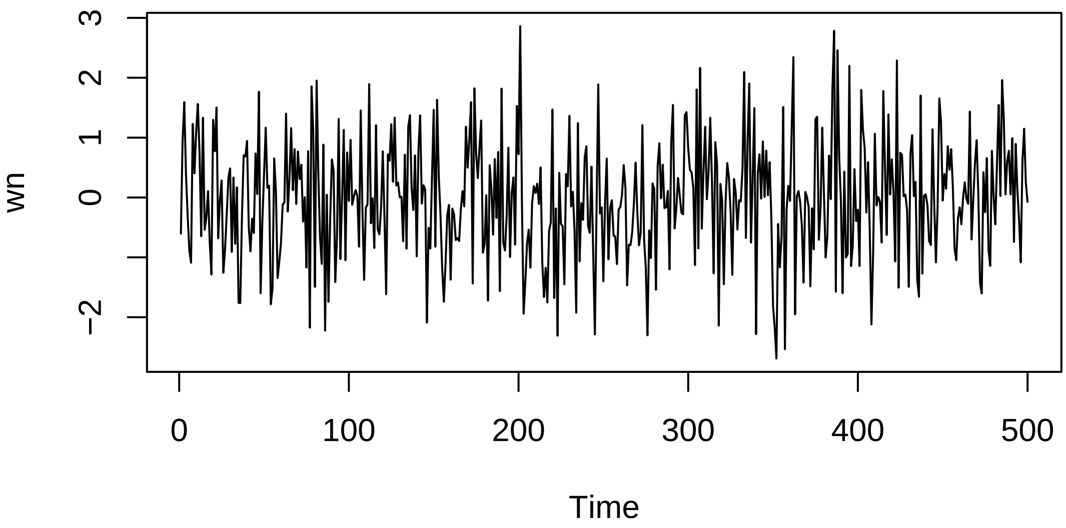
Kwiat Kowski, Phillips Schmidt, Shin (KPSS) test for stationarity. $H_0 =$ stationary $H_a =$ non-stationary

KPSS has been proven to be more stable.

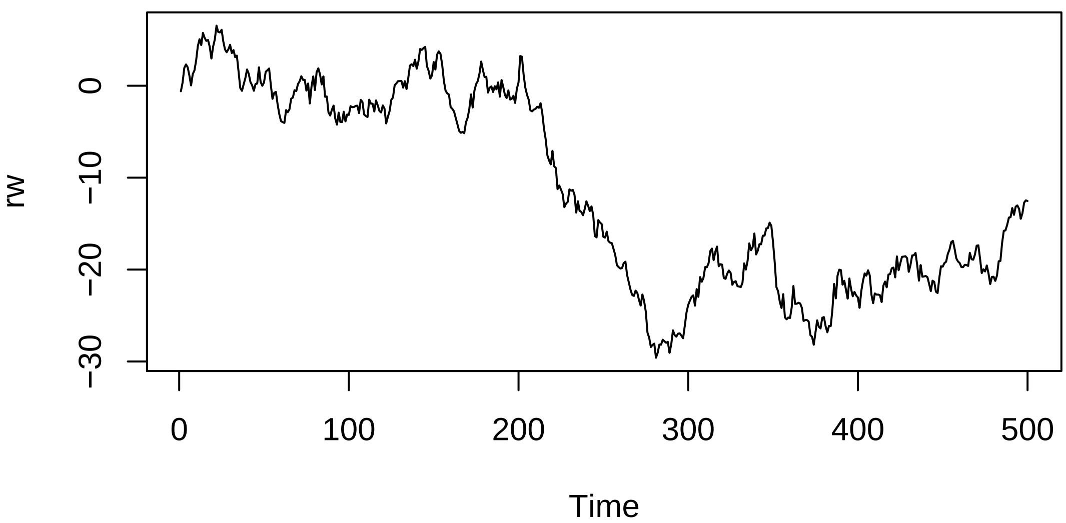
```r
library(tseries)
```

```r
wn <- rnorm(500, 0, 1)
rw <- cumsum(wn)

ts.plot(wn)
```




```r
ts.plot(rw)
```




```r
adf.test(wn)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  wn
## Dickey-Fuller = -7.8347, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

```r
kpss.test(wn)
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  wn
## KPSS Level = 0.17268, Truncation lag parameter = 5, p-value = 0.1
```

```r
adf.test(rw)
```

```
##
##  Augmented Dickey-Fuller Test
```

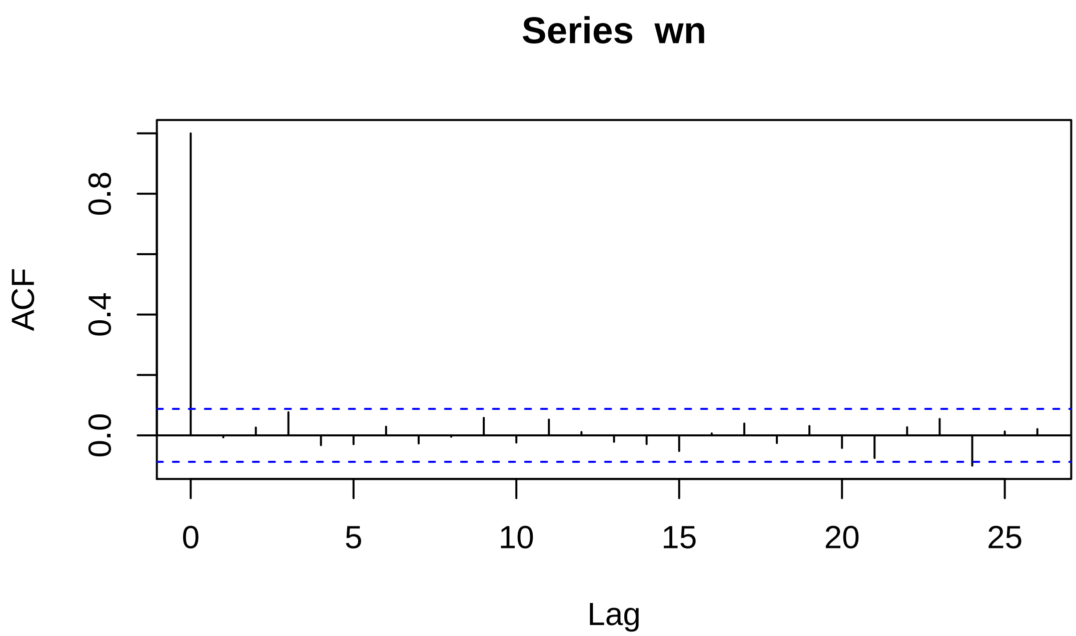
```
##
## data:  rw
## Dickey-Fuller = -1.1848, Lag order = 7, p-value = 0.9089
## alternative hypothesis: stationary
```

```r
kpss.test(rw)
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  rw
## KPSS Level = 6.594, Truncation lag parameter = 5, p-value = 0.01
```

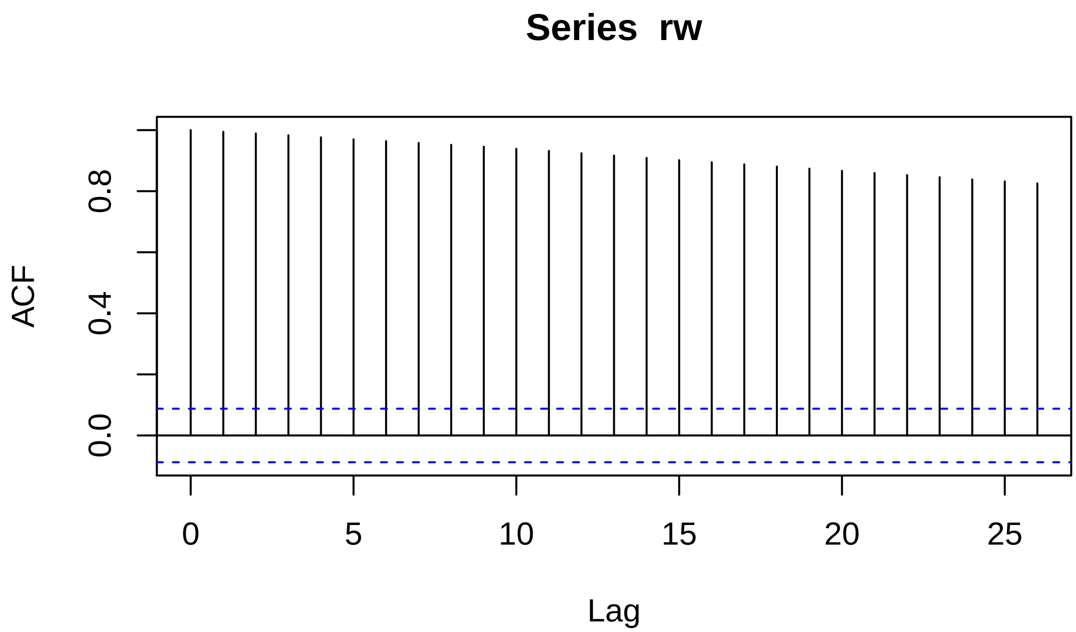
When you see an acf that dies down.

```r
acf(wn)
```



Conversely, if you see an acf that refuses to die down, that is non-stationary.

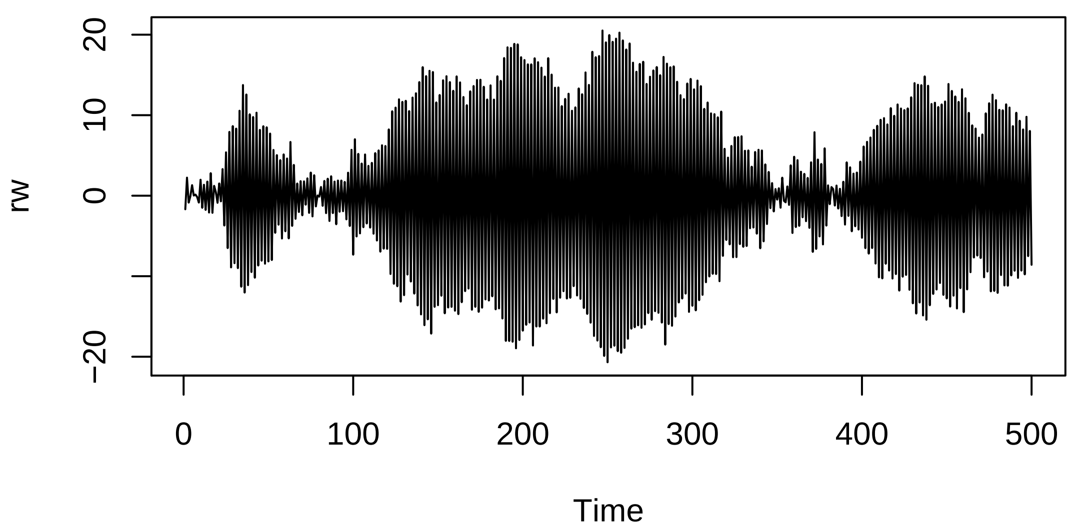
```r
acf(rw)
```




```
rw <- c(0)
wn <- rnorm(500, 0, 1)
rw[1] <- wn[1]

for (i in 2:length(wn)) {
  rw[i] <- -1 * rw[i-1] + wn[i]
}

ts.plot(rw)
```




```
adf.test(rw)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  rw
## Dickey-Fuller = -8.7163, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

```r
kpss.test(rw)
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  rw
## KPSS Level = 0.41378, Truncation lag parameter = 5, p-value =
## 0.07122
```

You cannot randomly sample time series because you end up breaking the auto correlation.