Time Series is Great!

James Balamuta

2016-05-23

Contents

Chapter 1

Preface

1.1 A foreword

The book presented below is a compilation of material and extended explanations that I have sought during my studies of time series. The material is not necessarily traditional, however, the hope is that it is at least helpful and/or provides an alternative explanation for the concepts under observation.

Any typos or other issues should be reported to James Balamuta forthwith.

1.2 Notation

The following notation will be adopted throughout the book.

- X denotes a (continuous) RV.
- X_t is X at time $t \in N$.
- $E\left(X_t\right)$ is the Mean of X at time t.
- $Var\left(X_t\right)$ is the Variance of X at time t.
- $X_1, X_2, \ldots, X_k$ are sequence of random variables.
- $f\left(x\right)$ denotes the density function of X and $f\left(x, y\right)$ denotes the joint density function of x and Y.
- $\left(X_t\right)_{t=1,\ldots,T} := \left(X_t\right) := \left(X_1, \ldots, X_T\right)$.

1.3 R Code

The code used throughout the book is R code. The R code should be able to used as-is. Alongside the PDF download of the book, you should find the R code used within each chapter.

Chapter 2

The World of Time Series

2.1 Objective of Time Series

1. View a set of observations made sequentially in "time."
   - "One damned thing after another." ~ R. A. Fisher
2. Find a suitable model to describe an observed process
   - "All models are wrong, but some are useful" ~ George Box
3. Forecast future observations
   - "Prediction is very difficult, especially if it's about the future." ~ Niels Bohr

In essence, we seek to be able to predict, classify, and associate observed data with a theoretical backend. To do so, one must first create a model that provides an explanation of the data as a mixture of a pattern and noise (error). That is, we must be able to formulate the data in terms of:

Model = Pattern + Noise

In Time Series, the pattern represents the association between values observed over time (e.g. autocorrelation).

Since these patterns are correlated in time, methods that assume independence are unable to be used.

2.2 What is a Time Series and can I eat it?

The definition of ***Time Series (TS)*** is as follows:

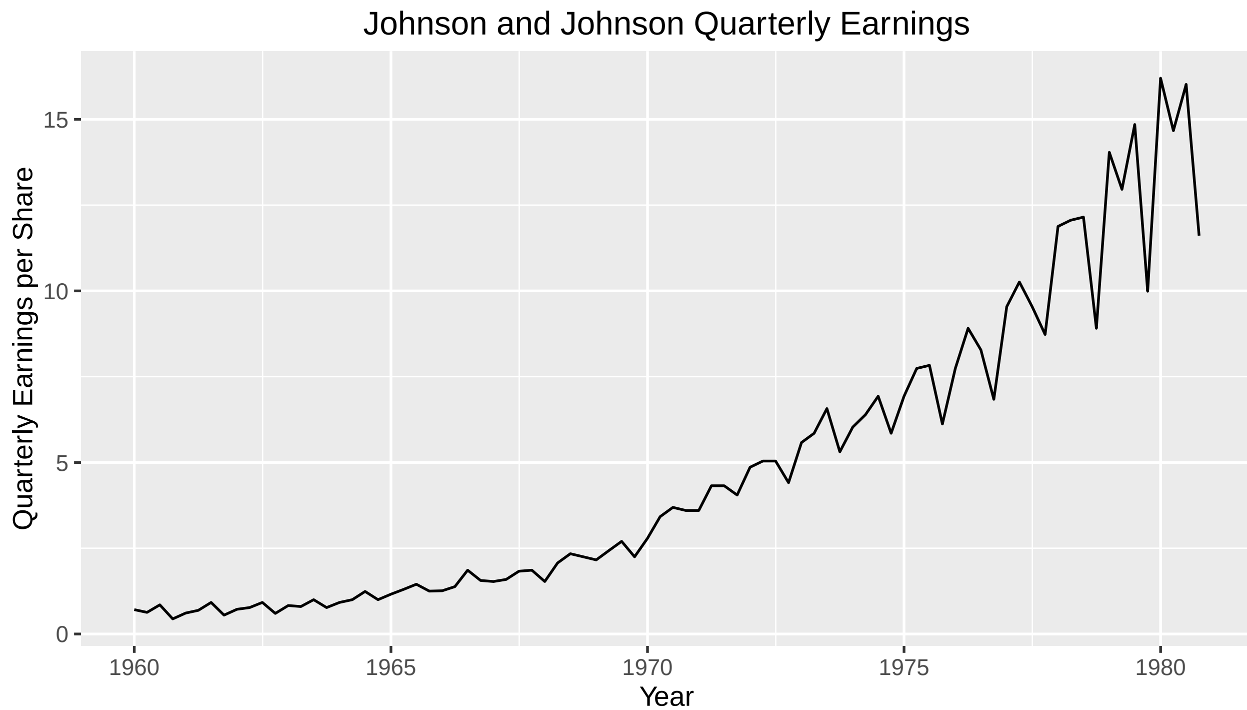
A ***Time Series*** is a stochastic process, a sequence of random variables (RV) defined on a common probability space denoted as $(X_t)_{t=1,\ldots,T}$ (i.e. $X_1, X_2, \ldots, X_T$).

Note: The time t belongs to discrete index sets ($\in \mathbb{Z}$) not continuous ($\notin \mathbb{R}$). After all, TS data is always collected at discrete time points. Furthermore, by time belonging to $\mathbb{Z}$ it can take upon itself negative and positive integer values (e.g. $-2, -1, 0, 1, 2$).

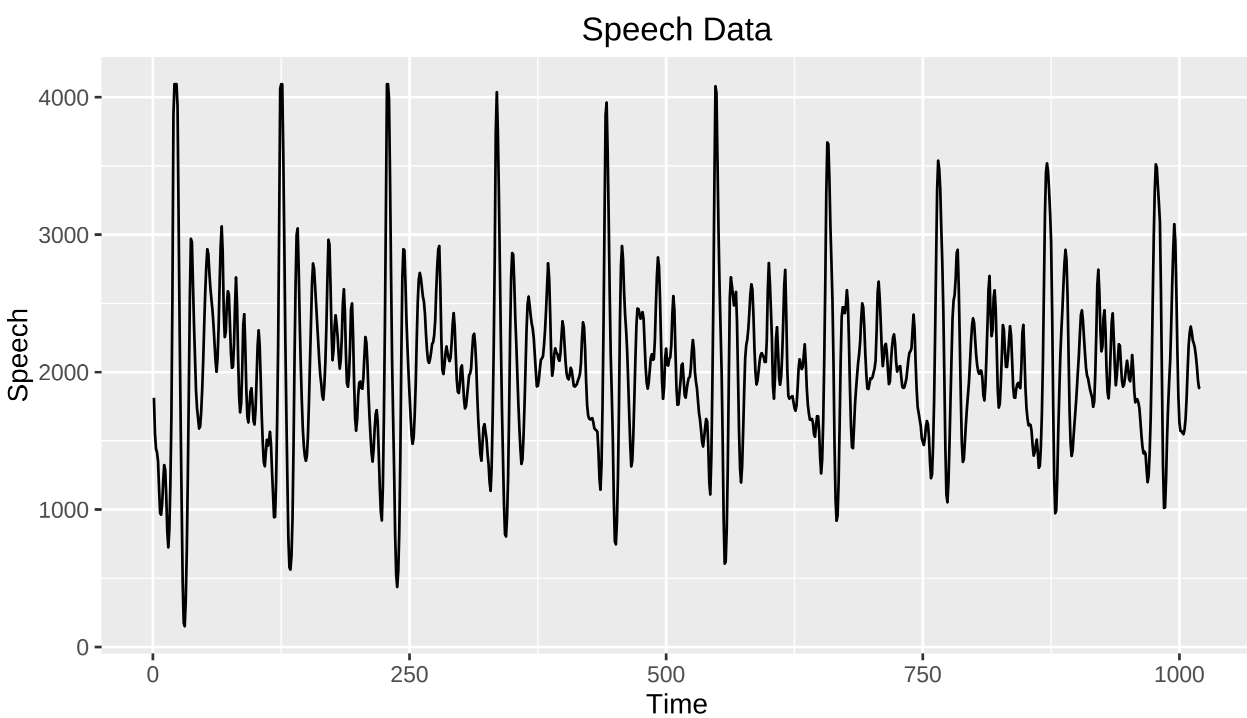
In laypersons terms, a time series is a variable that gets measured sequentially at fixed intervals of time, which are oftenly spaced apart at equal distances (e.g. equispaced).

Examples of Time Series:

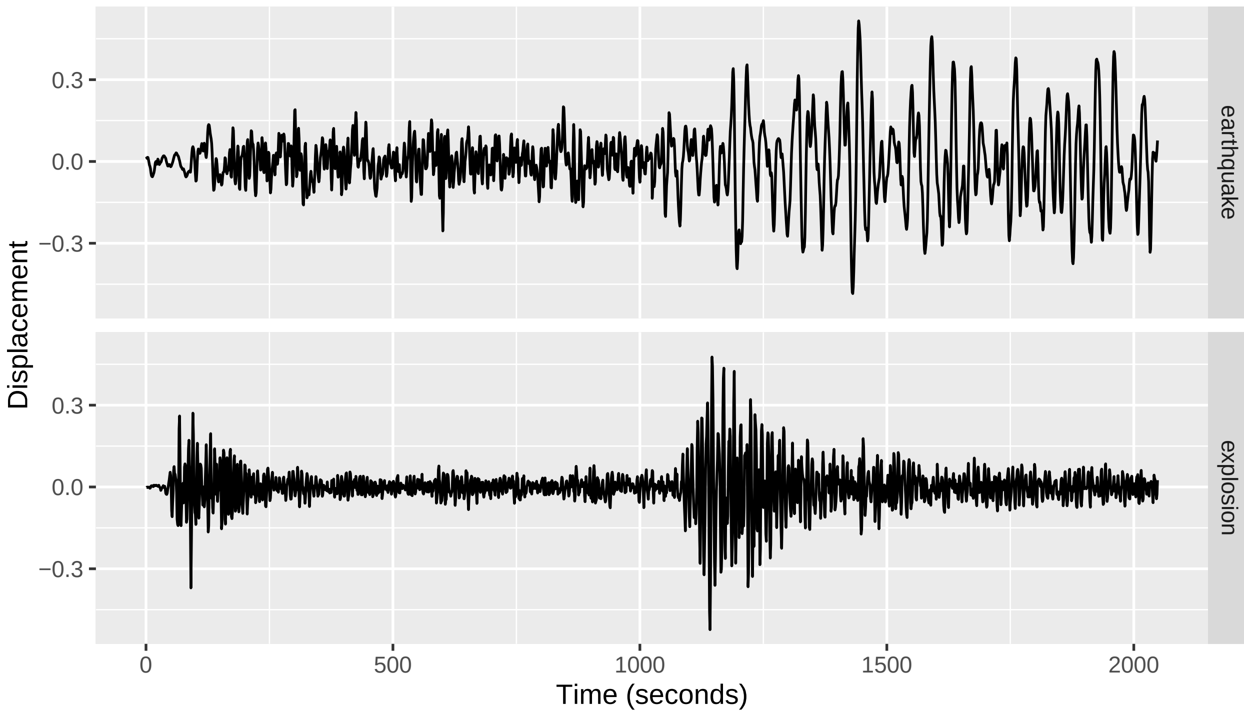
1. Stock Data from Johnson and Johson's Quarterly earnings...



2. Speech data from someone talking



3. Earthquake and explosion data



2.3 Exploratory Data Analysis (EDA) for Time Series

A large part of time series involves looking at graphs of time series. The graphs provide us information as to what kind of trends and outliers the data maybe hiding.

2.3.1 Identifying Trends

A trend exists when there is a long term increase or decrease or combination of increases or decreases (polynomial) in the data. It could be linear or non-linear.

Note: Long-term trend might change direction, indicating non-linear trends!

Examples of non-linear trends:

1. Seasonal trends (periodic): These are the cyclical patterns which repeat after a fixed/regular time period.
    - Business cycles (bust/recession, recovery, boom)
    - Seasons (summer, fall, winter, spring)
2. Non-seasonal trends (periodic): These patterns cannot be associated to seasonal variation.
    - Impact of economic indicators on stock returns.
3. “Other” trends: These trends have no regular patterns. They could be just local, short-term. They change statistical properties of a TS over a segment of time (“window”).
    - Earthquakes!
    - El Nino

2.3.2 Noticing changes in time and outliers

Change in time and outliers yields interesting results. These results can be seen as:

1. Change in Means
    - Change in means of a TS can be related to long-term, cyclical, and short-term trends.
2. Change in Variance
    - Change in variance can be related to change in the amplitude of the fluctuations of a TS.
3. Change in State
    - An event which causes change in statistical properties of TS for short term and long term! Some events cause abrupt changes in statistical properties of TS. They are often associated with "explosive" nature of TS.
4. Outliers
    - These are the "extreme" observations in the time series. May be related to data collection or change in state.

Chapter 3

Basic Models

3.1 White Noise

The process name of white noise has meaning in the notion of colors of noise. Specifically, the white noise is a process that mirrors white light's flat frequency spectrum. So, the process has equal frequencies in any interval of time.

Definition: **White Noise**

w_t or ε_t is a **white noise process** if w_t are uncorrelated identically distributed random variables with $E\left[w_t\right]=0$ and $Var\left[w_t\right]=\sigma^2$, for all t. We can represent this algebraically as:

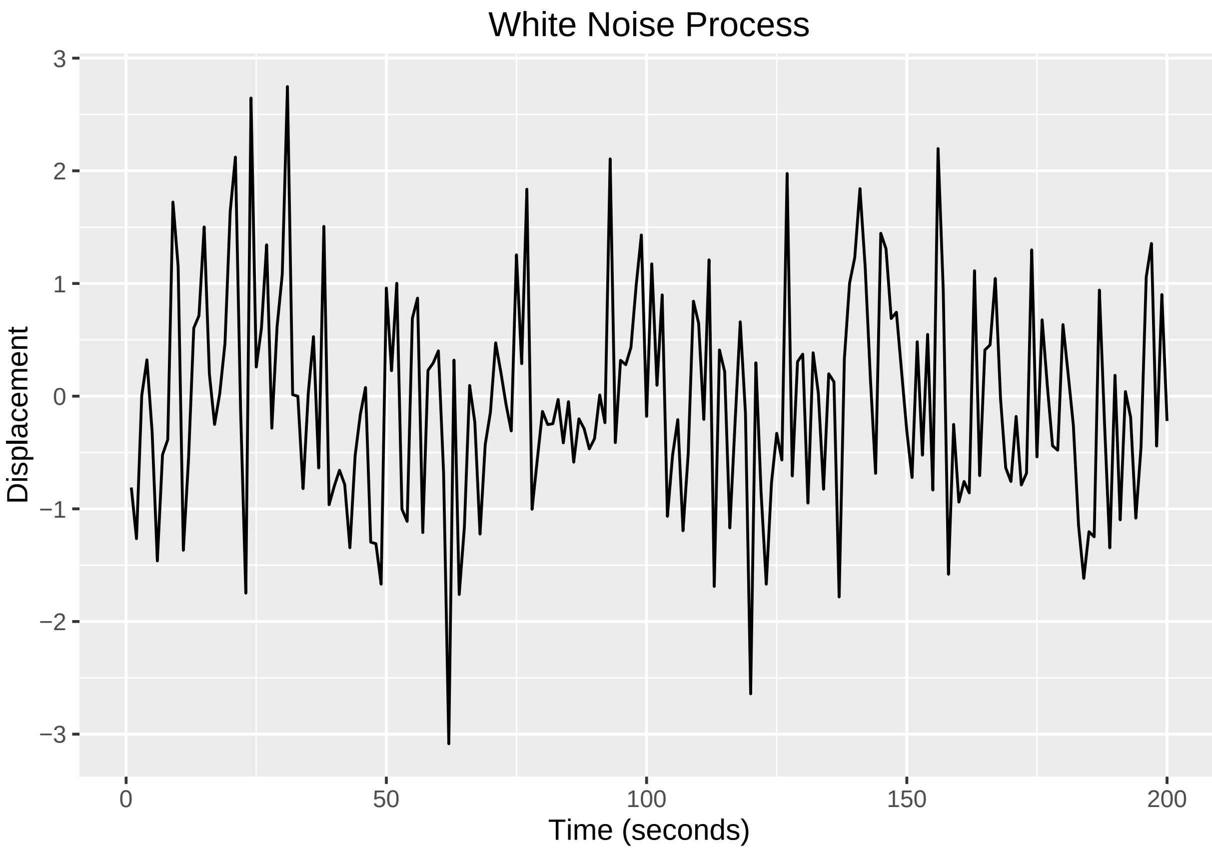
$$y_t = w_t,$$

where $w_t \overset{id}{\sim} WN\left(0, \sigma_w^2\right)$

Now, if the w_t are **Normally (Gaussian) distributed**, then the process is known as a **Gaussian White Noise** e.g. $w_t \overset{iid}{\sim} N\left(0, \sigma^2\right)$

To generate gaussian white noise use:

```
set.seed(1336)            # Set seed to reproduce the results
n  = 200                  # Number of observations to generate
wn = ts(rnorm(n,0,1))     # Generate Guassian white noise.

autoplot(wn) +
  ggtitle("White Noise Process") +
  ylab("Displacement") + xlab("Time (seconds)")
```



3.2 Moving Average Process of Order q = 1 a.k.a MA(1)

Definition: **Moving Average Process of Order (q = 1)**

The concept of a **Moving Average Process of Order q** is a way to remove "noise" and emphasize the signal. The moving average achieves this by taking the local averages of the data to produce a new smoother time series series. The newly created time series is more descriptive, but it does influence the dependence within the time series.

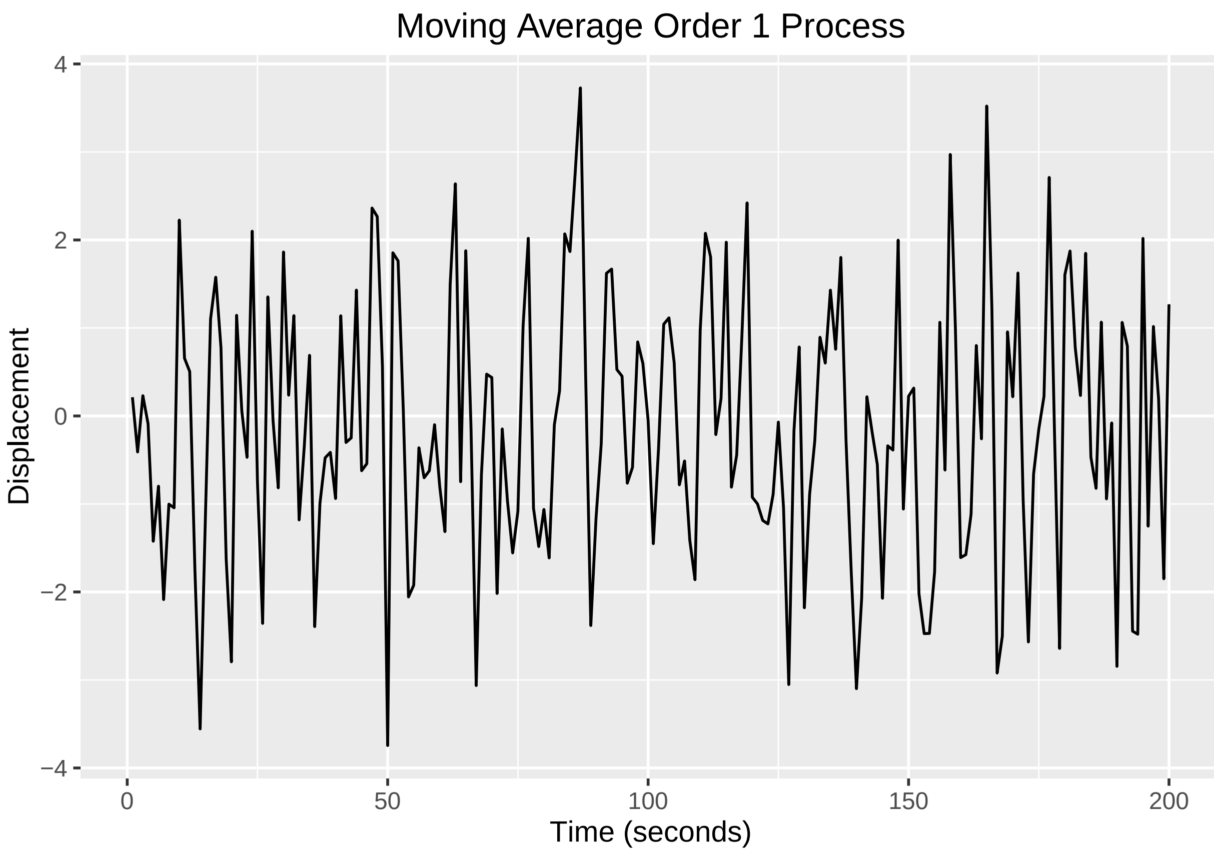
This process is generally denoted as **MA(1)** and is defined as:

$$y_t = \theta_1 w_{t-1} + w_t,$$

where $w_t \overset{iid}{\sim} WN\left(0, \sigma_w^2\right)$

```
set.seed(1345)  # Set seed to reproduce the results
n      = 200    # Number of observations to generate
sigma2 = 2      # Controls variance of Guassian white noise.
theta  = 0.3    # Handles the theta component of MA(1)


# Generate a white noise
wn = rnorm(n+1, sd = sqrt(sigma2))

# Simulate the MA(1) process
```

```r
ma = rep(0, n+1)
for(i in 2:(n+1)) {
  ma[i] = theta*wn[i-1] + wn[i]
}

ma = ts(ma[2:(n+1)])      # Remove first item

autoplot(ma) +
  ggtitle("Moving Average Order 1 Process") +
  ylab("Displacement") + xlab("Time (seconds)")
```



3.3 Drift

Definition: **Drift**

A **drift process** has two components: time and a slope. As more points are accumlated over time, the drift will match the common slope form.

Specifically, the drift process has the following form:

$$y_t = y_{t-1} + \delta$$

with the initial condition $y_0 = c$.

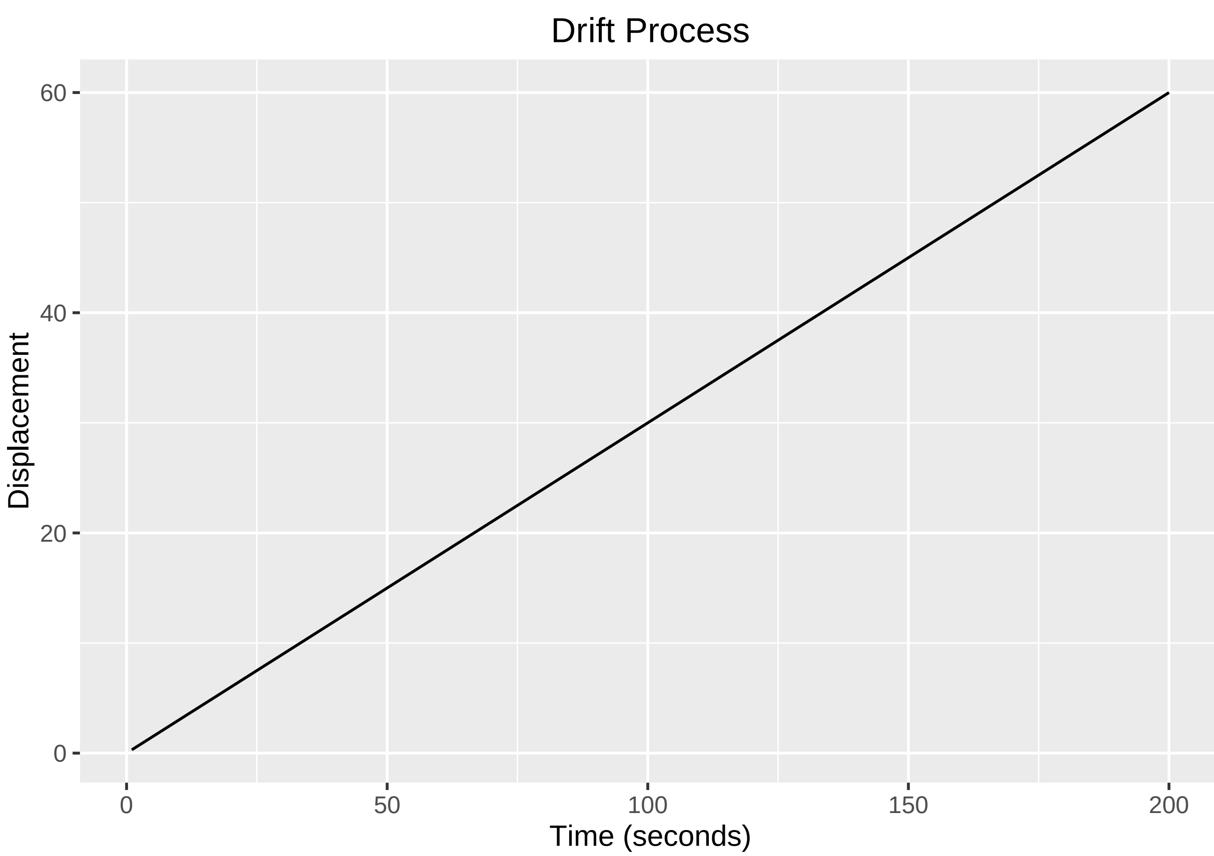
The process can be simplified using **backsubstitution** to being:

$$
\begin{aligned}
y_t &= y_{t-1} + \delta \\
&= (y_{t-2} + \delta) + \delta \\
&\vdots \\
&= \sum_{i=1}^{t} \delta + y_0 \\
y_t &= t\delta + c
\end{aligned}
$$

Again, note that a drift is similar to the slope-intercept form a linear line. e.g. $y = mx + b$.

To generate a drift use:

```
n     = 200              # Number of observations to generate
drift = .3               # Drift Control
dr    = ts(drift*(1:n)) # Generate drift sequence (e.g. y = drift*x + 0)

autoplot(dr) +
  ggtitle("Drift Process") +
  ylab("Displacement") + xlab("Time (seconds)")
```



3.4 Random Walk

In 1906, Karl Pearson coined the term 'random walk' and demonstrated that "the most likely place to find a drunken walker is somewhere near his starting point." Empirical evidence of this phenomenon is not too hard to find on a Friday night in Champaign.

Definition: **Random Walk**

A **random walk** is defined as a process where the current value of a variable is composed of the past value plus an error term that is a white noise. In algebraic form,

$$y_t = y_{t-1} + w_t$$

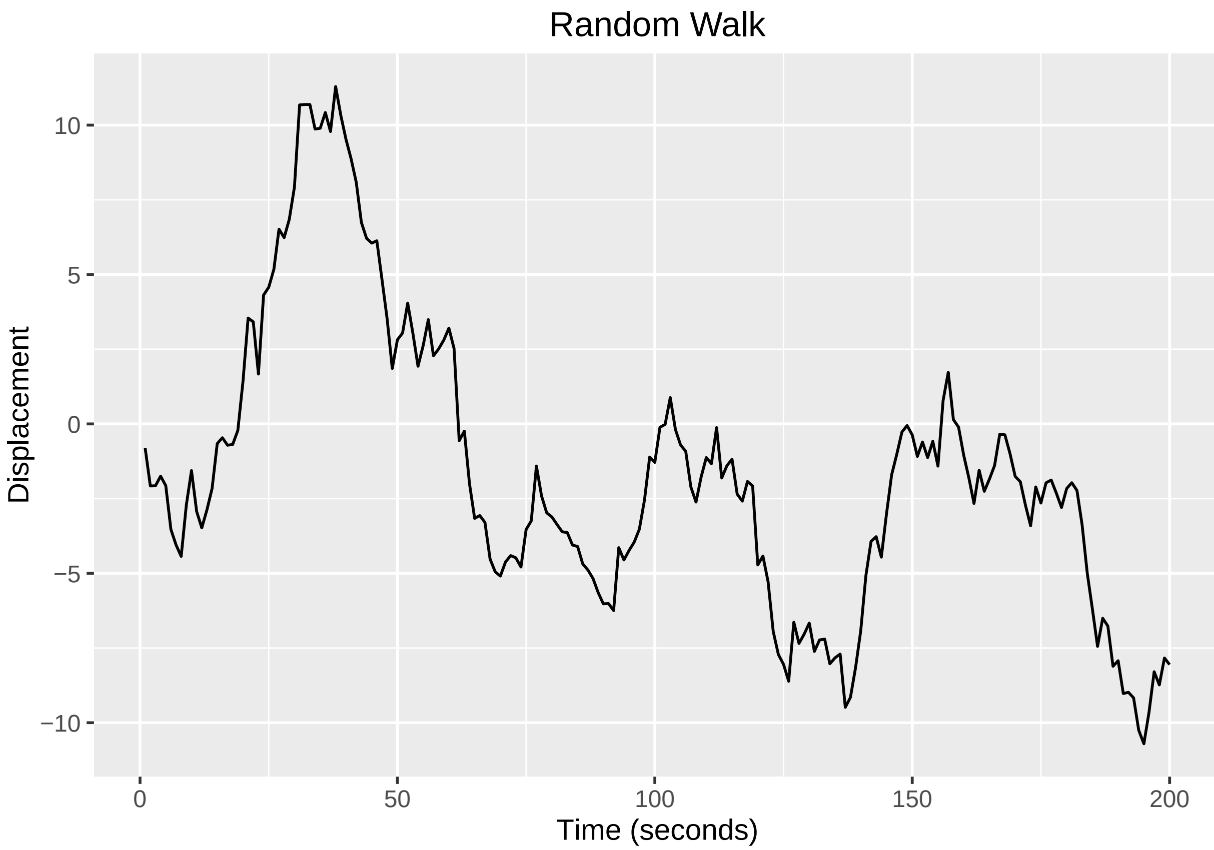
with the initial condition $y_0 = c$.

The process can be simplified using **backsubstitution** to being:

$$\begin{aligned} y_t &= y_{t-1} + w_t \\ &= (y_{t-2} + w_{t-1}) + w_t \\ &\vdots \\ y_t &= \sum_{i=1}^{t} w_i + y_0 = \sum_{i=1}^{t} w_i + c \end{aligned}$$

To generate a random walk, we use:

```r
set.seed(1336)        # Set seed to reproduce the results
n  = 200              # Number of observations to generate
w  = rnorm(n,0,1)     # Generate Guassian white noise.
rw = ts(cumsum(w))    # Cumulative sum

# Create a data.frame to graph in ggplot2
autoplot(rw) +
  ggtitle("Random Walk") +
  ylab("Displacement") + xlab("Time (seconds)")
```



3.5 Random Walk with Drift

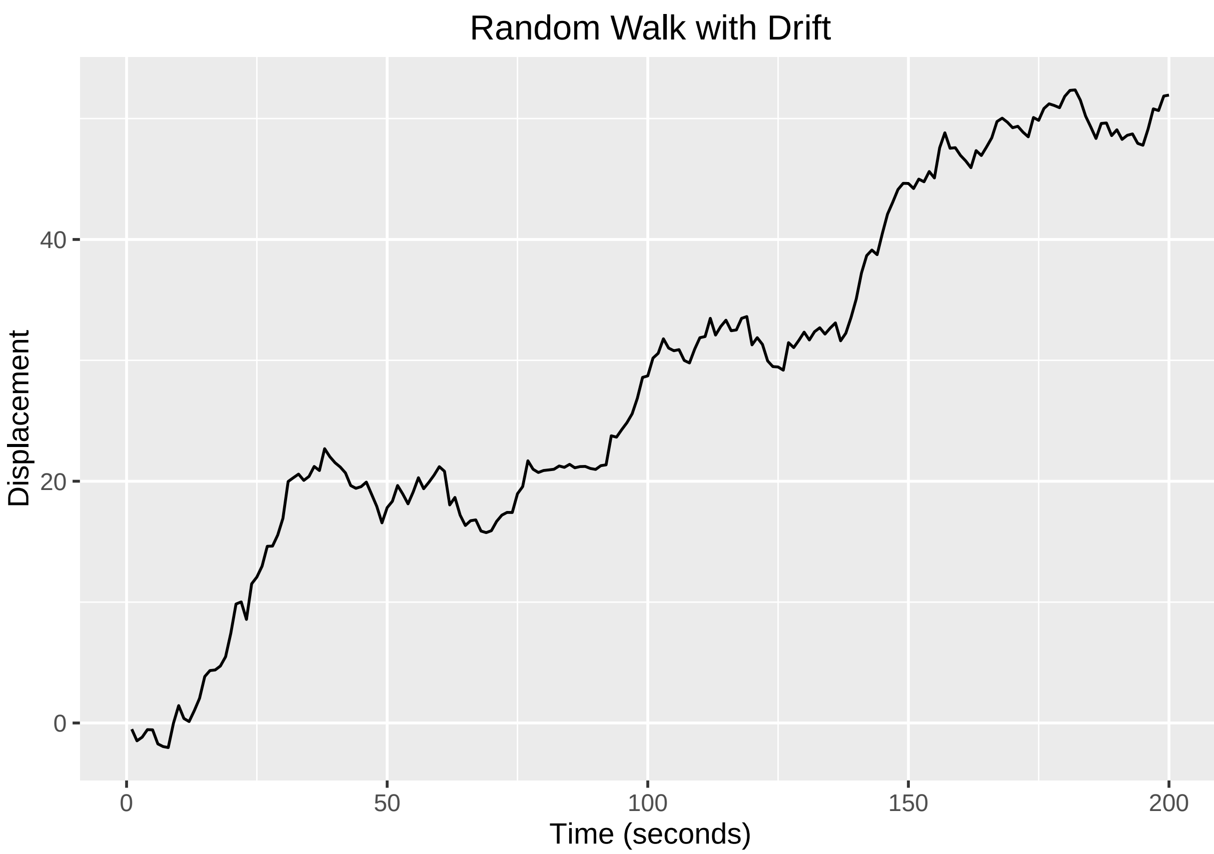
In the previous case of a random walk, we assumed that drift, δ, was equal to 0. What happens to the random walk if the drift is not equal to zero? That is, what happens with the initial condition $y_0 = c$?

$$\begin{aligned} y_t &= y_{t-1} + w_t + \delta \\ &= (y_{t-2} + w_{t-1} + \delta) + w_t + \delta \\ &\vdots \\ y_t &= \sum_{i=1}^{t} (w_i + \delta) + y_0 = \sum_{i=1}^{t} w_i + t\delta + c \end{aligned}$$

To generate a random walk with drift we use:

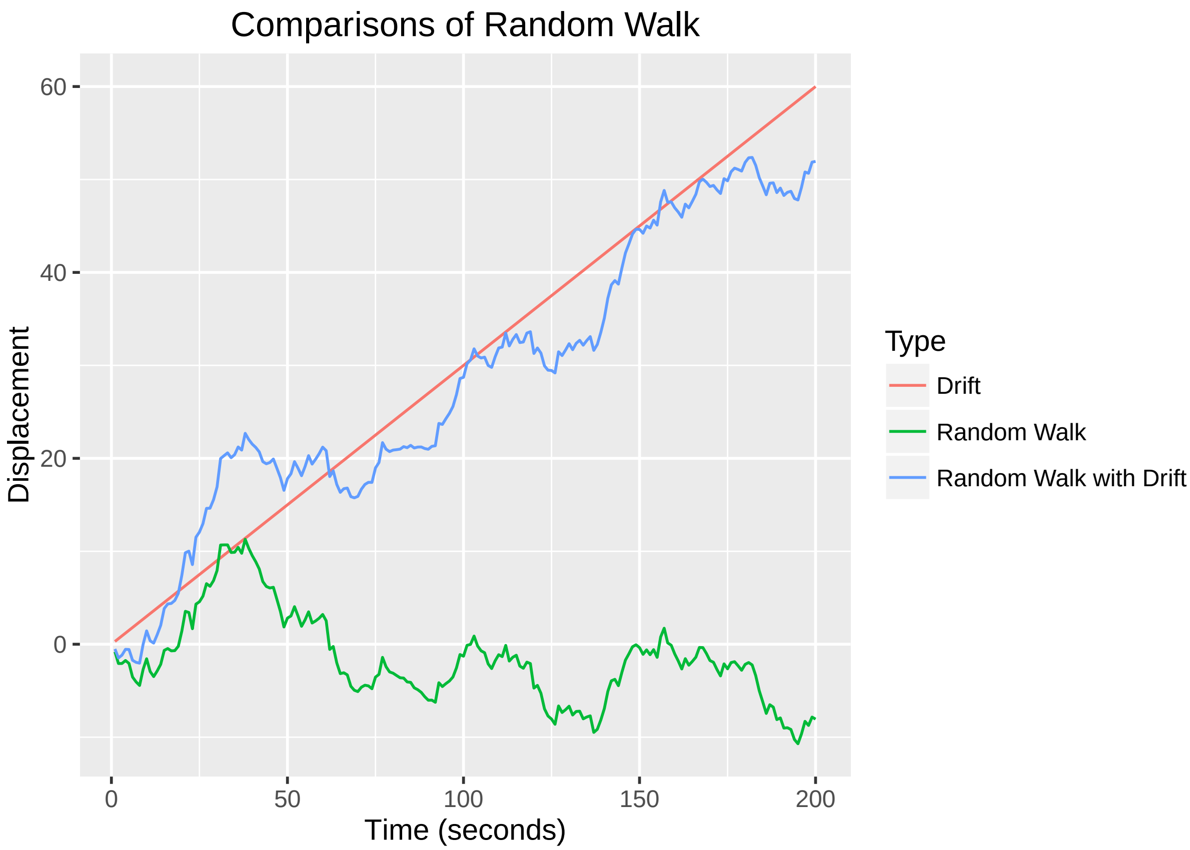
```
set.seed(1336)         # Set seed to reproduce the results
n      = 200           # Number of observations to generate
drift = .3             # Drift Control

w = rnorm(n,0,1)       # Generate Guassian white noise.
wd = w + drift         # Add a drift
rwd = ts(cumsum(wd)) # Cumulative sum

# Create a data.frame to graph in ggplot2
autoplot(rwd) +
```

```
  ggtitle("Random Walk with Drift") +
  ylab("Displacement") + xlab("Time (seconds)")
```



Notice the difference the drift makes upon the random walk:

```
# Add identifiers
drift.df = data.frame(Index = 1:n, Data = drift*(1:n), Type = "Drift")

rw.df = data.frame(Index = 1:n, Data = rw, Type = "Random Walk")

rwd.df = data.frame(Index = 1:n, Data = rwd, Type = "Random Walk with Drift")

combined.df = rbind(drift.df, rw.df, rwd.df)

ggplot(data = combined.df, aes(x = Index, y = Data, colour = Type)) +
  geom_line() +
  ggtitle("Comparisons of Random Walk") +
  ylab("Displacement") + xlab("Time (seconds)")
```



3.6 Autoregressive Process of Order p = 1 a.k.a AR(1)

Definition: **Autoregressive Process of Order p = 1**

This process is generally denoted as **AR(1)** and is defined as: $y_t = \phi_1 y_{t-1} + w_t$,

where $w_t \overset{iid}{\sim} WN\left(0, \sigma_w^2\right)$

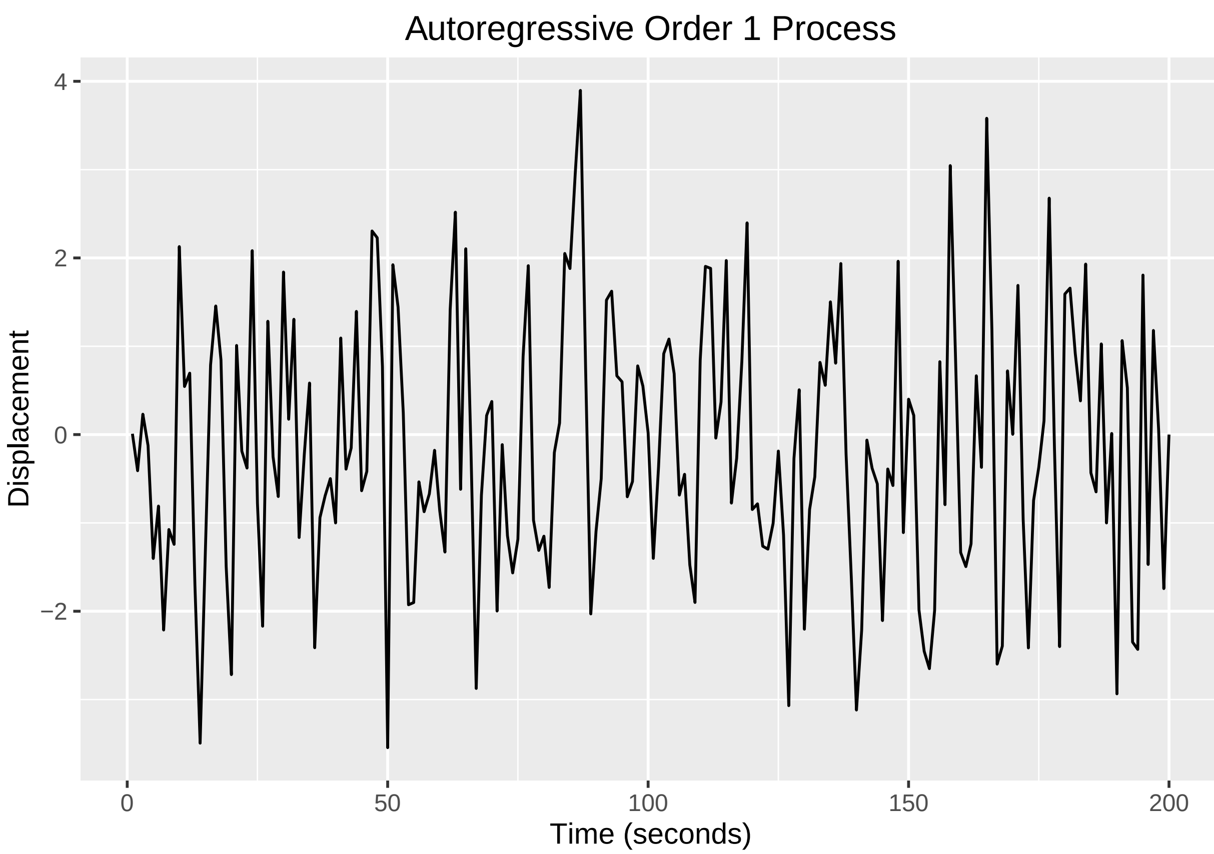
If $\phi_1 = 1$, then the process is equivalent to a random walk.

The process can be simplified using **backsubstitution** to being:

$$\begin{aligned} y_t &= \phi_t y_{t-1} + w_t \\ &= \phi_1 \left(\phi_1 y_{t-2} + w_{t-1}\right) + w_t \\ &= \phi_1^2 y_{t-2} + \phi_1 w_{t-1} + w_t \\ &\vdots \\ &= \phi^t y_0 + \sum_{i=0}^{t-1} \phi_1^i w_{t-i} \end{aligned}$$

```
set.seed(1345)   # Set seed to reproduce the results
n      = 200     # Number of observations to generate
sigma2 = 2       # Controls variance of Guassian white noise.
phi    = 0.3     # Handles the phi component of AR(1)
```

```
wn = rnorm(n+1, sd = sqrt(sigma2))

# Simulate the MA(1) process
ar = rep(0,n+1)
for(i in 2:n) {
  ar[i] = phi*ar[i-1] + wn[i]
}

ar = ts(ar[2:(n+1)])

autoplot(ar) +
  ggtitle("Autoregressive Order 1 Process") +
  ylab("Displacement") + xlab("Time (seconds)")
```



Chapter 4

Dependency

Generally speaking, there is a dependence that within the sequence of random variables.

Recall the difference between independent and dependent data:

Definition: **Independence**

$X_1, X_2, \ldots, X_T$ are independent and identically distributed if and only if

$$P\left(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_T \leq x_T\right) = P\left(X_1 \leq x_1\right) P\left(X_2 \leq x_2\right) \cdots P\left(X_T \leq x_T\right)$$

for any $T \geq 2$ and $x_1, \ldots, x_T \in \mathbb{R}$.

Definition: **Dependence**

$X_1, X_2, \ldots, X_T$ are identically distributed but dependent, then

$$\left|P\left(X_1 < x_1, X_2 < x_2, \ldots, X_T < x_T\right) - P\left(X_1 < x_1\right) P\left(X_2 < x_2\right) \cdots P\left(X_T < x_T\right)\right| \neq 0$$

for some $x_1, \ldots, x_T \in \mathbb{R}$.

4.1 Measuring (Linear) Dependence

There are many forms of dependency…

However, the methods, covariance and correlation, that we will be using are specific to measuring linear dependence. As a result, these tools are less helpful to measure monotonic dependence and they are much less helpful to measure nonlinearly dependence.

4.1.1 Autocovariance Function

Dependence between T different RV is difficult to measure in one shot! So we consider just two random variables, X_t and X_{t+h}. Then one (linear) measure of dependence is the covariance between (X_t, X_{t+h}). Since X is the same RV observed at two different time points, the covariance between X_t and X_{t+h} is defined as the Autocovariance.

Definition: **Autocovariance Function**

The **Autocovariance Function** is defined as the second moment product

$$\gamma_x\left(t, t+h\right) = \operatorname{cov}\left(x_t, x_{t+h}\right) = E\left[\left(x_t - \mu_t\right)\left(x_{t+h} - \mu_{t+h}\right)\right]$$

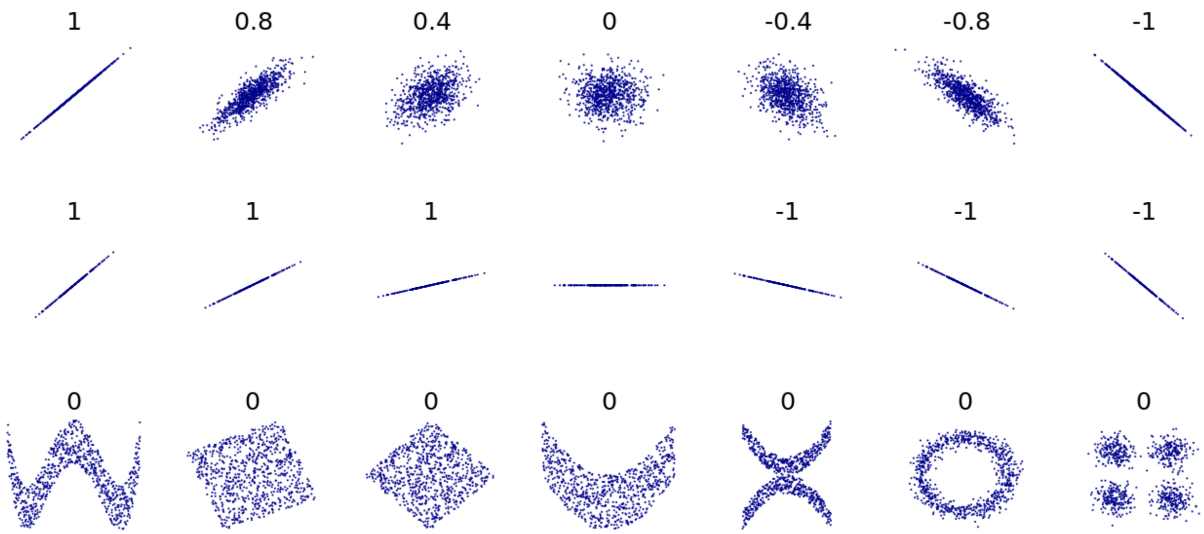


Figure 4.1: dependency

for all t and $t+h$.

The notation used above corresponds to:

$$\text{cov}\left(X_t, X_{t+h}\right) = E\left[X_t X_{t+h}\right] - E\left[X_t\right] E\left[X_{t+h}\right]$$

$$E\left[X_t\right] = \int\limits_{-\infty}^{\infty} x \cdot f_x\left(x\right) dx$$

$$E\left[X_t X_{t+h}\right] = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} x_1 x_2 \cdot f\left(x_1, x_2\right) dx_1 dx_2$$

We normally drop the subscript referring to the time series if it is clear to the time series the autocovariance function is referencing. e.g. $\gamma_x\left(t, t+h\right) = \gamma\left(t, t+h\right)$.

The more commonly used formulation for weakly stationary processes (more next section) is:

$$\gamma\left(X_t, X_{t+h}\right) = \text{cov}\left(X_t, X_{t+h}\right) = \gamma\left(h\right)$$

A few other notes:

1. The covariance function is symmetric. That is, $\gamma\left(t, t+h\right) = \gamma\left(t+h, t\right)$
2. Just as any covariance, the $\gamma\left(t, t+h\right)$ is "scale dependent", $\gamma\left(t, t+h\right) \in \mathbb{R}$, or $-\infty \leq \gamma\left(t, t+h\right) \leq +\infty$
    1. If $\left|\gamma\left(t, t+h\right)\right|$ is "close" to 0, then they are "less dependent"
    2. If $\left|\gamma\left(t, t+h\right)\right|$ is "far" from 0, X_t and X_{t+h} are "more dependent".
3. $\gamma\left(t, t+h\right) = 0$ does not imply X_t and X_{t+h} are independent.
4. If X_t and X_{t+h} are joint normally distributed then X_t and X_{t+h} are independent.

4.1.2 Autocorrelation Function (ACF)

A "simplified" $\gamma\left(t, t+h\right)$ is the Autocorrelation (AC) between X_t and X_{t+h}, which is scale free! It is simply defined as

$$\rho\left(X_t, X_{t+h}\right) = Corr\left(X_t, X_{t+h}\right) = \frac{Cov\left(X_t, X_{t+h}\right)}{\sigma_{X_t} \sigma_{X_{t+h}}}$$

The more commonly used formulation for weakly stationary processes (more next section) is:

$$\rho\left(X_t, X_{t+h}\right) = \frac{Cov\left(X_t, X_{t+h}\right)}{\sigma_{X_t}\sigma_{X_{t+h}}} = \frac{\gamma\left(h\right)}{\gamma\left(0\right)} = \rho\left(h\right)$$

Therefore, the autocorrelation function is only a function of the lag h between observations.

Just as any correlation:

1. $\rho\left(X_t, X_{t+h}\right)$ is scale free
2. $\rho\left(X_t, X_{t+h}\right)$ is closer to $\pm 1 \Rightarrow \left(X_t, X_{t+h}\right)$ "more dependent."

Remember… When using correlation….

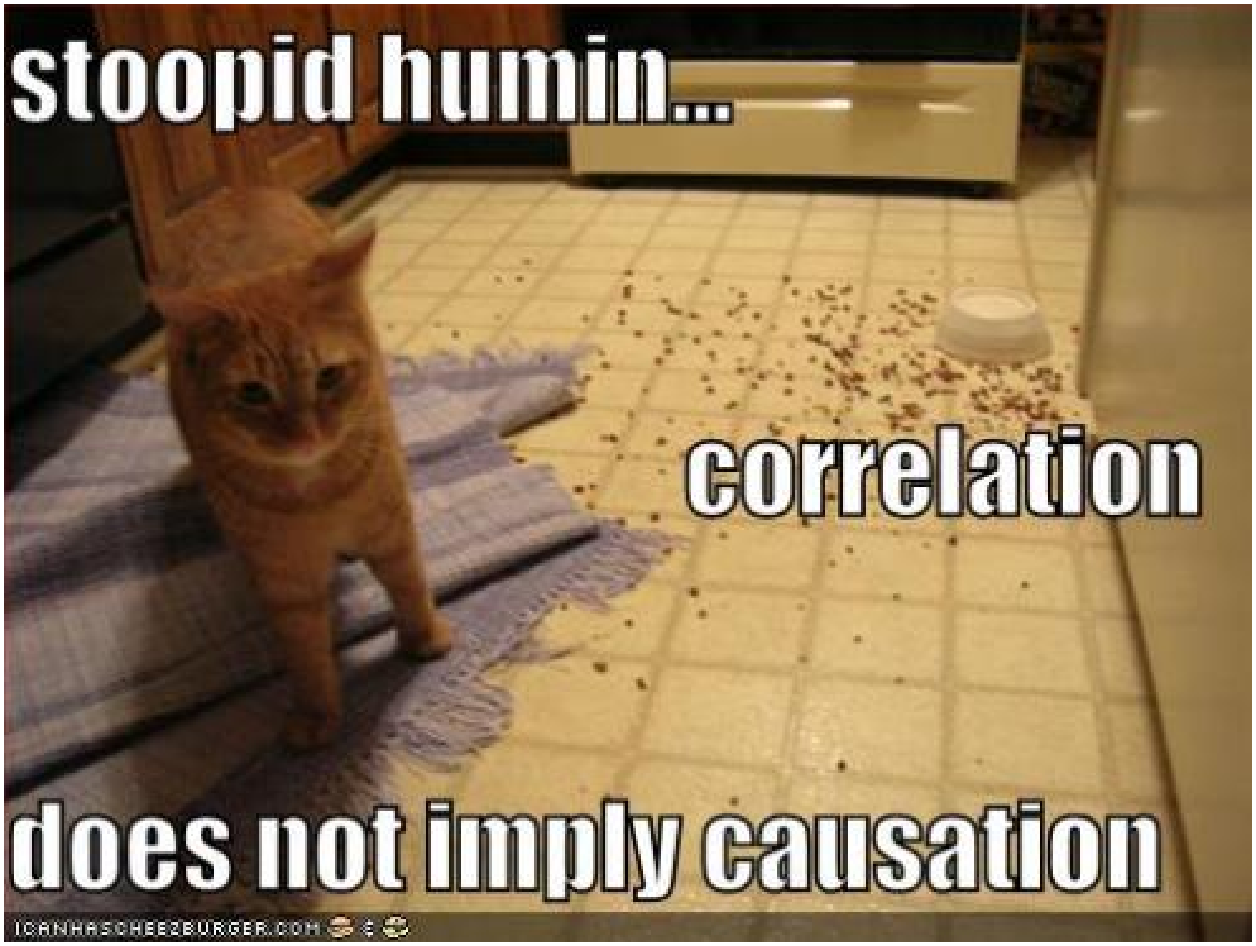


Figure 4.2: correlation_sillies

4.1.3 Cross dependency functions

Consider two time series: $\left(X_t\right)$ and $\left(Y_t\right)$.

Then the cross-covariance function between two series $\left(X_t\right)$ and $\left(Y_t\right)$ is:

$$\gamma_{XY}\left(t, t+h\right) = \operatorname{cov}\left(X_t, Y_{t+h}\right) = E\left[\left(X_t - E\left[X_t\right]\right)\left(Y_{t+h} - E\left[Y_{t+h}\right]\right)\right]$$

The cross-correlation function is given by

$$\rho_{XY}\left(t, t+h\right) = Corr\left(X_t, Y_{t+h}\right) = \frac{\gamma_{XY}\left(t, t+h\right)}{\sigma_{X_t}\sigma_{Y_{t+h}}}$$

These ideas can extended beyond the bivariate case to a general multivariate setting.

4.1.4 Sample Autocovariance and Autocorrelation Functions

Definition: **Sample Autocovariance Function**

The **Sample Autocovariance Function** is defined as:

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

Definition: **Sample Autocorrelation function**

The **Sample Autocorrelation function** is defined as:

$$\hat{\rho}\left(h\right) = \frac{\hat{\gamma}\left(h\right)}{\hat{\gamma}\left(0\right)}$$

Chapter 5

Stationarity

There are two kinds of stationarity: Strong and Weak Stationarity. These types of stationarity are **not equivalent** and the presence of **one kind of stationarity does not imply the other**. That is, a time series can be strongly stationary but not weakly stationary and vice versa. In very rare cases, a time series can be both strong and weakly stationary. The most common form of stationarity is that of weakly stationarity.

5.1 Weak Stationarity

Definition: **Weak Stationarity** or **Second-order Stationarity**

The mean and autocovariance of the stochastic process are finite and invariant under a shift in time, i.e.

$$E\left[X_t\right] = \mu_t = \mu < \infty$$
$$cov\left(X_t, X_{t+h}\right) = cov\left(X_{t+k}, X_{t+h+k}\right) = \gamma\left(h\right)$$

5.2 Strong Stationarity

Definition: **Strong Stationarity** or **Strict Stationarity**

The joint probability distribution of (X_t), $t \in N$ is invariant under a shift in time, i.e.

$$P\left(X_t \leq x_1, \ldots, X_{t+k} \leq x_k\right) = P\left(X_{t+h} \leq x_1, \ldots, X_{t+h+k} \leq x_k\right)$$

for any time shift h and any $x_1, x_2, \ldots, x_k$ belong to the domain of $X_{t_1}, X_{t_2}, \ldots, X_{t_k}$ respectively.

So to summarize, we have **weak stationarity** that relies on how *separated each observation* is rather than their location in time. On the flip side, we have **strict stationarity** that relies on the *location in time* instead of how separated observations are.

Stationarity of X_t matters, because it provides the framework in which averaging makes sense. Unless properties like mean and covariance are either fixed or "evolve" in a known manner, we cannot average the observed data.

With this being said, here are a few examples of stationarity:

1. $X_t \sim Cauchy$ is strictly stationary but **NOT** weakly stationary.
    - The strong stationarity exists due to the symmetric properties of the distribution.
    - It cannot be weakly stationary because it has an infinite variance!

2. $X_{2t} = U_{2t}, X_{2t+1} = V_{2t+1} \forall t$ where $U_t \overset{iid}{\sim} N(1,1)$ and $V_t \overset{iid}{\sim} Exponential(1)$ is weakly stationary but **NOT** strictly stationary.
   - The weak stationary exists since the mean is constant ($\mu = 1$) and the variance does not depend on time ($\sigma^2 = 1$).
   - It cannot be strongly stationary due to values not aligning in time.

Regarding white noises, we can obtain different levels of stationarity depending on the assumption:

1. If $X_t \sim WN$, e.g. **uncorrelated observations** with a finite variance, then it is weakly stationary but **NOT** strictly stationary.
2. If $X_t \overset{iid}{\sim} NWN$, e.g. **normally distributed independent observations** with a finite variance, then it is weakly stationary *AND* strictly stationary.

The autocovariance of weakly stationary processes has the following properties:

1. $\gamma(0) = var[X_t] \geq 0$ (variance)
2. $\gamma(h) = \gamma(-h)$ (function is even / symmetric)
3. $|\gamma(h)| \leq \gamma(0) \forall h$.

We obtain these properties through:

1. $$\gamma(0) = Var(x_t) = E\left[(x_t - \mu)^2\right] = \sum_{t=1}^{T} p_t (x_t - \mu)^2 = p_1(x_1 - \mu)^2 + \cdots + p_T(x_T - \mu)^2 \geq 0$$
2. $$\begin{aligned} \gamma(h) &= \gamma(t + h - t) \\ &= E[(x_{t+h} - \mu)(x_t - \mu)] \\ &= E[(x_t - \mu)(x_{t+h} - \mu)] \\ &= \gamma(t - (t+h)) \\ &= \gamma(-h) \end{aligned}$$
3. Using the Cauchy-Schwarz Inequality, $(E[XY])^2 \leq E\left[X^2\right] E\left[Y^2\right]$, we have:
$$\begin{aligned} (|\gamma(h)|)^2 &= (\gamma(h))^2 \\ &= (E[(x_t - \mu)(x_{t+h} - \mu)])^2 \\ &\leq E\left[(x_t - \mu)^2\right] E\left[(x_{t+h} - \mu)^2\right] \\ &= (\gamma(0))^2 \\ (\gamma(h))^2 &\leq (\gamma(0))^2 \\ |\gamma(h)| &\leq \gamma(0) \end{aligned}$$

5.3 Is it weakly stationary?

In order to verify if it is weakly stationary, we must make sure the time series satisfies:

1. $E[y_t] = \mu_t = \mu < \infty$
2. $cov(y_t, Y_{t+h}) = \gamma(h)$

5.3.1 Is a random walk weakly stationary?

First up, is a RW stationary? By intuition, the answer should be "no" since there is a randomness component that cannot be accounted for when looking for a pattern. But, we need to prove that.

1.
$$\begin{aligned} E\left[y_t\right] &= E\left[y_{t-1}+w_t\right] \\ &= E\left[\sum_{i=1}^{t} w_t + Y_0\right] \\ &= E\left[\sum_{i=1}^{t} w_t\right] + Y_0 \\ &= 0 + c \\ &= c \end{aligned}$$

Note, the mean here is constant since it depends only on the value of the first term in the sequence.

2.
$$\begin{aligned} Var\left(y_t\right) &= Var\left(\sum_{i=1}^{t} w_t + Y_0\right) \\ &= Var\left(\sum_{i=1}^{t} w_t\right) + \underbrace{Var\left(Y_0\right)}_{=0 \text{ constant}} \\ &= \sum_{i=1}^{t} Var\left(w_t\right) \\ &= t\sigma_w^2 \\ &\Rightarrow Cov\left(y_t, y_{t+h}\right) \neq \gamma\left(h\right) \end{aligned}$$

Alas, the variance has a dependence on time. This causes the $Var\left(y_t\right) \geq \infty$ as $t \to \infty$. As a result, the process is not weakly stationary.

Continuing on just to obtain the covariance, we have:

$$\begin{aligned} \gamma\left(h\right) &= Cov\left(y_t, y_{t+h}\right) \\ &= Cov\left(\sum_{i=1}^{t} w_i, \sum_{j=1}^{t+h} w_j\right) \\ &= Cov\left(\sum_{i=1}^{t} w_i, \sum_{j=1}^{t} w_j\right) \\ &= \min\left(t, t+h\right)\sigma_w^2 \\ &= \left(t + \min\left(0, h\right)\right)\sigma_w^2 \end{aligned}$$

5.3.2 Is an MA(1) Stationary?

1.
$$\begin{aligned} E\left[y_t\right] &= E\left[\theta_1 w_{t-1} + w_t\right] \\ &= \theta_1 E\left[w_{t-1}\right] + E\left[w_t\right] \\ &= 0 \end{aligned}$$

The mean is constant over time. So the first criterion is okay.

2.

$$\begin{aligned} Cov\left(y_t, y_{t+h}\right) &= E\left[\left(y_t - E\left[y_t\right]\right)\left(y_{t+h} - E\left[y_{t+h}\right]\right)\right] \\ &= E\left[y_t y_{t+h}\right] - \underbrace{E\left[y_t\right]}_{=0}\underbrace{E\left[y_{t+h}\right]}_{=0} \\ &= E\left[\left(\theta_1 w_{t-1} + w_t\right)\left(\theta_1 w_{t+h-1} + w_{t+h}\right)\right] \\ &= E\left[\theta_1^2 w_{t-1} w_{t+h-1} + \theta w_t w_{t+h} + \theta_1 w_{t-1} w_{t+h} + w_t w_{t+h}\right] \end{aligned}$$

$$E\left[w_t w_{t+h}\right] = \operatorname{cov}\left(w_t, w_{t+h}\right) + E\left[w_t\right]E\left[w_{t+h}\right] = 1_{\{h=0\}}\sigma_w^2$$

$$\Rightarrow Cov\left(y_t, y_{t+h}\right) = \left(\theta_1^2 1_{\{h=0\}} + \theta_1 1_{\{h=1\}} + \theta_1 1_{\{h=-1\}} + 1_{\{h=0\}}\right)\sigma_w^2$$

$$\gamma(h) = \begin{cases} \left(\theta_1^2 + 1\right)\sigma_w^2 & h = 0 \\ \theta_1 \sigma_w^2 & |h| = 1 \\ 0 & |h| > 1 \end{cases}$$

Note, the autocovariance function does not depend on time. Thus, the second weakly stationary criterion is satisfied.

The MA(1) process is weakly stationary since both the mean and variance are constant over time.

As a bonus, note that we also can easily obtain the autocorrelation function (ACF)

$$\Rightarrow \rho(h) = \begin{cases} 1 & h = 0 \\ \frac{\theta_1 \sigma_w^2}{\left(\theta_1^2+1\right)\sigma_w^2} = \frac{\theta_1}{\theta_1^2+1} & |h| = 1 \\ 0 & |h| > 1 \end{cases}$$

5.3.3 Is an AR(1) Stationary?

Consider the AR(1) process given as:

$$y_t = \phi_1 y_{t-1} + w_t, \text{ where } w_t \overset{iid}{\sim} WN\left(0, \sigma_w^2\right)$$

This process was shown to simplify to:

$$y_t = \phi^t y_0 + \sum_{i=0}^{t-1} \phi_1^i w_{t-i}$$

In addition, we add the requirement that $|\phi_1| < 1$. This requirement allows for the process to be stationary. If $\phi_1 \geq 1$, the process would not converge. This way the process will be able to be written as a geometric series that converges:

$$\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}, \quad |r| < 1$$

Next, we demonstrate how crucial this property is:

$$
\begin{aligned}
\lim_{t\to\infty} E\left[y_t\right] &= \lim_{t\to\infty} E\left[\phi^t y_0 + \sum_{i=0}^{t-1} \phi_1^i w_{t-i}\right] \\
&= \lim_{t\to\infty} \underbrace{\phi^t y_0}_{|\phi|<1 \Rightarrow t\to\infty \;=\; 0} + \sum_{i=0}^{t-1} \phi_1^i \underbrace{E\left[w_{t-i}\right]}_{=0} \\
&= 0 \\
\lim_{t\to\infty} Var\left(y_t\right) &= \lim_{t\to\infty} Var\left(\phi^t y_0 + \sum_{i=0}^{t-1} \phi_1^i w_{t-i}\right) \\
&= \lim_{t\to\infty} \underbrace{Var\left(\phi^t y_0\right)}_{=0 \text{ since constant}} + Var\left(\sum_{i=0}^{t-1} \phi_1^i w_{t-i}\right) \\
&= \lim_{t\to\infty} \sum_{i=0}^{t-1} \phi_1^{2i} Var\left(w_{t-i}\right) \\
&= \lim_{t\to\infty} \sigma_w^2 \sum_{i=0}^{t-1} \phi_1^{2i} \\
&= \sigma_w^2 \cdot \underbrace{\frac{1}{1-\phi^2}}_{\text{Geometric Series}}
\end{aligned}
$$

This leads us to being able to conclude the autocovariance function is:

$$
\begin{aligned}
Cov\left(y_t, y_{t+h}\right) &= Cov\left(y_t, \phi y_{t+h-1} + w_{t+h}\right) \\
&= Cov\left(y_t, \phi y_{t+h-1}\right) \\
&= Cov\left(y_t, \phi^{|h|} y_t\right) \\
&= \phi^{|h|} Cov\left(y_t, y_t\right) \\
&= \phi^{|h|} Var\left(y_t\right) \\
&= \phi^{|h|} \frac{\sigma_w^2}{1-\phi_1^2}
\end{aligned}
$$

Both the mean and autocovariance function do not depend on time and, thus, the AR(1) process is stationary if $|\phi_1| < 1$.

If we assume that the AR(1) process is stationary, we can derive the mean and variance in another way. Without a loss of generality, we'll assume $y_0 = 0$.

Therefore:

$$
\begin{aligned}
y_t &= \phi_t y_{t-1} + w_t \\
&= \phi_1 \left(\phi_1 y_{t-2} + w_{t-1}\right) + w_t \\
&= \phi_1^2 y_{t-2} + \phi_1 w_{t-1} + w_t \\
&\vdots \\
&= \sum_{i=0}^{t-1} \phi_1^i w_{t-i}
\end{aligned}
$$

$$
\begin{aligned}
E\left[y_t\right] &= E\left[\sum_{i=0}^{t-1} \phi_1^i w_{t-i}\right] \\
&= \sum_{i=0}^{t-1} \phi_1^i \underbrace{E\left[w_{t-i}\right]}_{=0} \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
Var\left(y_t\right) &= E\left[\left(y_t - E\left[y_t\right]\right)^2\right] \\
&= E\left[y_t^2\right] - \left(E\left[y_t\right]\right)^2 \\
&= E\left[y_t^2\right] \\
&= E\left[\left(\phi_1 y_{t-1} + w_t\right)^2\right] \\
&= E\left[\phi_1^2 y_{t-1}^2 + w_t^2 + 2\phi_1 y_t w_t\right] \\
&= \phi_1^2 E\left[y_{t-1}^2\right] + \underbrace{E\left[w_t^2\right]}_{=\sigma_w^2} + 2\phi_1 \underbrace{E\left[y_t\right]}_{=0} \underbrace{E\left[w_t\right]}_{=0} \\
&= \underbrace{\phi_1^2 Var\left(y_{t-1}\right) + \sigma_w^2 = \phi_1^2 Var\left(y_t\right) + \sigma_w^2}_{\text{Assume stationarity}} \\
Var\left(y_t\right) &= \phi_1^2 Var\left(y_t\right) + \sigma_w^2 \\
Var\left(y_t\right) - \phi_1^2 Var\left(y_t\right) &= \sigma_w^2 \\
Var\left(y_t\right)\left(1 - \phi_1^2\right) &= \sigma_w^2 \\
Var\left(y_t\right) &= \frac{\sigma_w^2}{1 - \phi_1^2}
\end{aligned}
$$

Chapter 6

Joint Stationarity

Two time series, say (X_t) and (Y_t), are said to be jointly stationary if they are each stationary, and the cross-covariance function

$$\gamma_{XY}\left(t, t+h\right) = Cov\left(X_t, Y_{t+h}\right) = \gamma_{XY}\left(h\right)$$

is a function only of lag h.

The cross-correlation function for jointly stationary times can be expressed as:

$$\rho_{XY}\left(t, t+h\right) = \frac{\gamma_{XY}\left(t, t+h\right)}{\sigma_{X_t}\sigma_{Y_{t+h}}} = \frac{\gamma_{XY}\left(h\right)}{\sigma_{X_t}\sigma_{Y_{t+h}}} = \rho_{XY}\left(h\right)$$

6.1 The Backshift Operator and Differencing Operations

Definition: **Backshift Operator**

The **Backshift Operator** is helpful when manipulating time series. When we backshift, we are changing the indices of the time series. e.g. $t \to t-1$. The operator is defined as:

$$Bx_t = x_{t-1}$$

If we were to repeatedly apply the backshift operator, we would receive:

$$\begin{aligned} B^2 x_t &= B\left(Bx_t\right) \\ &= B\left(x_{t-1}\right) \\ &= x_{t-2} \end{aligned}$$

We can generalize this behavior as:

$$B^k x_t = x_{t-k}$$

The backshift operator is helpful for later decompositions in addition to making differencing operations more straightforward.

Definition: **Differencing Operator**

The **Differencing Operator** is defined as the gradient symbol applied to a time series:

$$\nabla x_t = x_t - x_{t-1}$$

The differencing operator is helpful when trying to remove trend from the data.

We can take higher moments of differences by:

$$\begin{aligned}
\nabla^2 x_t &= \nabla\left(\nabla x_t\right) \\
&= \nabla\left(x_t - x_{t-1}\right) \\
&= \left(x_t - x_{t-1}\right) - \left(x_{t-1} - x_{t-2}\right) \\
&= x_t - 2x_{t-1} + x_{t-2}
\end{aligned}$$

So, the difference operator has the following properties:

$$\begin{aligned}
\nabla^k x_t &= \nabla^{k-1}\left(\nabla x_t\right) \\
\nabla^1 x_t &= \nabla x_t
\end{aligned}$$

Notice, within the difference operation, we are backshifting the timeseries.

If we rewrite the difference operator to use the backshift operator, we receive:

$$\nabla x_t = x_t - x_{t-1} = \left(1 - B\right) x_t$$

This holds for later incarnations as well:

$$\begin{aligned}
\nabla^2 x_t &= x_t - 2x_{t-1} + x_{t-2} \\
&= \left(1 - B\right)\left(1 - B\right) x_t \\
&= \left(1 - B\right)^2 x_t
\end{aligned}$$

Thus, we can generalize this to:

$$\nabla^k x_t = \left(1 - B\right)^k x_t$$

Bibliography