Introduction to Time Series Analysis

PRACTICAL TIME SERIES ANALYSIS

THISTLETON AND SADIGOV

Objectives

Define a time series

Get familiar with ‘astsa’ package

Definition

Time series is a data set collected through time.

Correlation

Sampling adjacent points in time introduce a correlation.

Areas

Economics and financial time series

Physical time series

Marketing time series

Demographic time series

Population time series

Etc.

“astsa” package

Package by Robert H. Shumway and David S. Stoffer

Contains data sets and scripts to accompany “Time Series Analysis and Its Applications: With R Examples”

https://cran.r-project.org/web/packages/astsa/astsa.pdf

What We’ve Learned

Definition of a time series (we will re-define it in a slightly different way)

The package titled ‘astsa’

Some Time Plots

PRACTICAL TIME SERIES ANALYSIS

THISTLETON AND SADIGOV

Objectives

See some examples of time series.

Produce meaningful time plots.

Some time series from ‘astsa’

jj

flu

globtemp

globtempl

star

Johnson and Johnson Quarterly Earnings (jj)

US company Johnson and Johnson

Quarterly earnings

84 quarters

1st quarter of 1960 to 4th quarter of 1980

Pneumonia and influenza deaths in the U.S. (flu)

Monthly pneumonia and influenza deaths per 10,000 people

11 years

From 1968 to 1978

Land-ocean temperature deviations (globtemp)

Global mean land-ocean temperature deviations

Deviations from base period 1951-1980 average

Measured in degrees centigrade

For the years 1880-2015.

http://data.giss.nasa.gov/gistemp/graphs/

Land (only) temperature deviations (globtempl)

Global mean [land only] temperature deviations

Deviations from base period 1951-1980 average

Measured in degrees centigrade

For the years 1880-2015.

http://data.giss.nasa.gov/gistemp/graphs/

Variable Star (star)

The magnitude of a star taken at midnight

For 600 consecutive days

The data are from “The Calculus of Observations, a Treatise on Numerical Mathematics”, by E.T. Whittaker and G. Robinson

What We’ve Learned

Time series exist in variety of areas

How to produce meaningful time plots

Stationarity

PRACTICAL TIME SERIES ANALYSIS

THISTLETON AND SADIGOV

Objectives

Get some intuition for (weak) stationary time series

No systematic change in mean

i.e., No trend

No systematic change in variation

No periodic fluctuations

The properties of one section of a data are much like the properties of the other sections of the data

For an non-stationary time series, we will do some transformations to get stationary time series

What We've Learned

In a (weak) stationary time series, there is no

systematic change in mean (no trend)

systematic change in variance

periodic variations

Autocovariance function

PRACTICAL TIME SERIES ANALYSIS

THISTLETON AND SADIGOV

Objectives

recall random variables and covariance of two random variables

characterize time series as a realization of a stochastic process

define autocovariance function

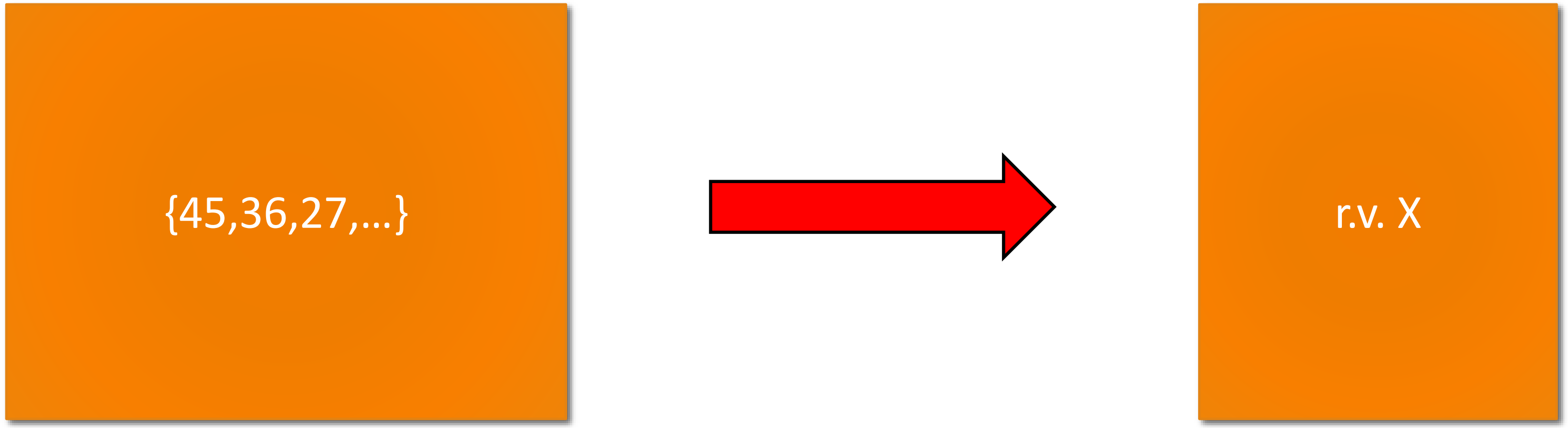
Random variables

Random variable is defined

$$X: S \to \mathbb{R}$$

where S is the sample space of the experiment.

From data to a model



Discrete vs. Continuous r.v.

X= {20, 37, 57, …}

VS.

Y in (10, 60)

20 is a realization of r.v. X

30.29 is a realization of a r.v. Y

Covariance

X, Y are two random variables.

Measures the linear dependence between two random variables

$$CoV(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = Cov(Y,X)$$

Stochastic Processes

Collection of a random variables

$$X_1, X_2, X_3, \ldots$$

$$X_t \sim distribution\ (\mu, \sigma^2)$$

Time series as a realization of a stochastic process

$$X_1, X_2, X_3, \dots$$

$$30, 29, 57, \dots$$

Autocovariance function

$$\gamma(s,t) = Cov(X_s, X_t) = E[(X\_s - \mu_s)(X_t - \mu_t)]$$

$$\gamma(t,t) = E[(X_t - \mu_t)^2] = Var(X_t) = \sigma_t^2$$

Autocovariance function cont.

$$\gamma_k = \gamma(t, t+k) \approx c_k$$

What We've Learned

the definition of a stochastic processes

how to characterize time series as realization of a stochastic process

how to define autocovariance function of a time series

Autocovariance coefficients

PRACTICAL TIME SERIES ANALYSIS

THISTLETON AND SADIGOV

Objectives

Recall the covariance coefficient for a bivariate data set

Define autocovariance coefficients for a time series

Estimate autocovariance coefficients of a time series at different lags

Covariance

X, Y are two random variables.

Measures the <u>linear</u> dependence between two random variables

$$Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

Estimation of the covariance

We have a paired dataset

$$(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$$

Estimation of covariance (cov() in R)

$$s_{xy} = \frac{\sum_{t=1}^{N}(x_t - \bar{x})(y_t - \bar{y})}{N-1}$$

Autocovariance coefficients

Autocovariance coefficients at different lags $\gamma_k = Cov(X_t, X_{t+k})$

c_k is an estimation of γ_k.

We assume (weak) stationarity

Estimation

$$c_k = \frac{\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{N}$$

where

$$\bar{x} = \frac{\sum_{t=1}^{N} x_t}{N}$$

Routine in R

acf() routine (next video lecture)

acf(time_series, type='covariance')

Purely random process

Time series with no special pattern

We use rnorm() routine

What We've Learned

Definition of autocovariance coefficients at different lags

Estimate autocovariance coefficients of a time series using acf() routine

The autocorrelation function

PRACTICAL TIME SERIES ANALYSIS

THISTLETON AND SADIGOV

Objectives

Define the autocorrelation function

Obtain corellograms using acf() routine

Estimate autocorrelation coefficients at different lags using acf() routine

The autocorrelation function (ACF)

We assume weak stationarity

The autocorrelation coefficient between X_t and X_{t+k} is defined to be

$$-1 \leq \rho_k = \frac{\gamma_k}{\gamma_0} \leq 1$$

Estimation of autocorrelation coefficient at lag k

$$r_k = \frac{c_k}{c_0}$$

Another way to write r_k

$$r_k = \frac{\sum_{t=1}^{N-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{N}(x_t - \bar{x})^2}$$

acf() routine

We have already used it for autocovariance coefficients

It plots autocorrelation coefficients at different lags: Correlogram

It always starts at 1 since $r_0 = \frac{c_0}{c_0} = 1$

What We've Learned

Definition of the autocorrelation function (ACF)

How to produce correlograms using acf() routine

How to estimate the autocorrelation coefficients at different lags using acf() routine.

Random Walk

PRACTICAL TIME SERIES ANALYSIS

THISTLETON AND SADIGOV

Objectives

Get familiar with the random walk model

Simulate a random walk in R

Obtain the correlogram of a random walk

See the difference operator in action

Model

Location at previous step
(or price of the stock yesterday)

Location at time t
(or a price of a stock today)

$$X_t = X_{t-1} + Z_t$$

$$Z_t \sim \; Normal\ (\mu, \sigma^2)$$

White noise
(residual)

$X_0 = 0$ → $X_1 = Z_1$ → $X_2 = Z_1 + Z_2$ → … → $$X_t = \sum_{i=1}^{t} Z_i$$

$$E[X_t] = E\left[\sum_{i=1}^{t} Z_i\right] = \sum_{i=1}^{t} E[Z_i] = \mu t$$

$$Var[X_t] = Var\left[\sum_{i=1}^{t} Z_i\right] = \sum_{i=1}^{t} Var[Z_i] = \sigma^2 t$$

Simulation

$X_1 = 0$

$Z_t \sim Normal\ (0, 1)$

$X_t = X_{t-1} + Z_t$ for $t = 2,3, \ldots, 1000$

Plot and ACF

Removing the tend

$X_t = X_{t-1} + Z_t$

$X_t - X_{t-1} = Z_t$

$\nabla X_t = X_t - X_{t-1}$

$\nabla X_t = Z_t$

Purely Random Process

Difference operator

diff() to remove the trend

Plot and ACF differenced time series

What We’ve Learned

Random Walk model

How to simulate a random walk in R

How to get stationary time series from a random walk using diff() operator

Introduction to Moving Average processes

PRACTICAL TIME SERIES ANALYSIS

THISTLETON AND SADIGOV

Objectives

Identify Moving average processes

Intuition

X_t is a stock price of a company

Each daily announcement of the company is modeled as a noise

Effect of the daily announcements (noises Z_t) on the stock price (X_t) might last few days (say 2 days)

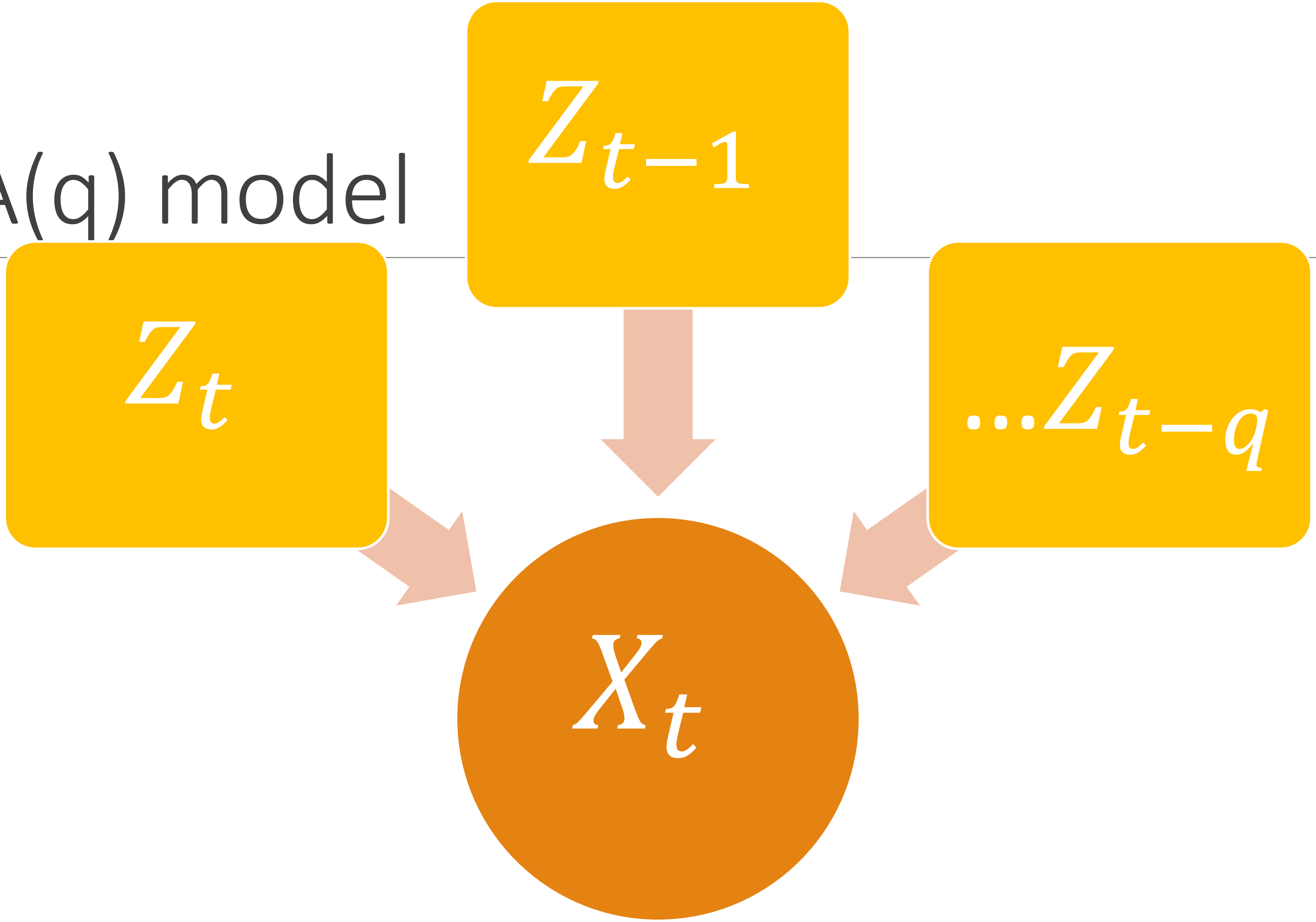
Stock price is linear combination of the noises that affects it

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$$

Moving average model of order 2

MA(2)

MA(q) model



$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \ldots + \theta_q Z_{t-q}$$

Z_i are i.id. & $Z_i \sim Normal\ (\mu, \sigma^2)$

What We’ve Learned

How to identify Moving average processes MA(q)

Simulating MA(2) process

PRACTICAL TIME SERIES ANALYSIS

THISTLETON AND SADIGOV

Objectives

Simulate a moving average process

Interpret correlogram of a Moving average process

MA(2) process

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$$

Simulation - MA(2) model

$$X_t = Z_t + 0.7\, Z_{t-1} + 0.2\, Z_{t-2}$$

$$Z_t \sim Normal\ (0, 1)$$

What We've Learned

How to simulate MA processes in R

That ACF of MA(q) cuts off at lag q