# Time Series Forecasting Using Python

*Novri Suhermi, Suhartono, Dedy Dwi Prastyo

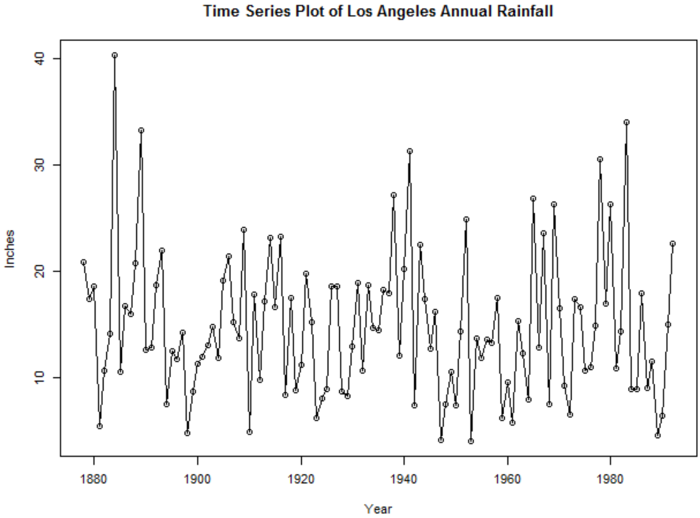
Department of Statistics - Institut Teknologi Sepuluh Nopember

Presented in PyCon Indonesia 2017
December 9, 2017
PENS, Surabaya

# Outline

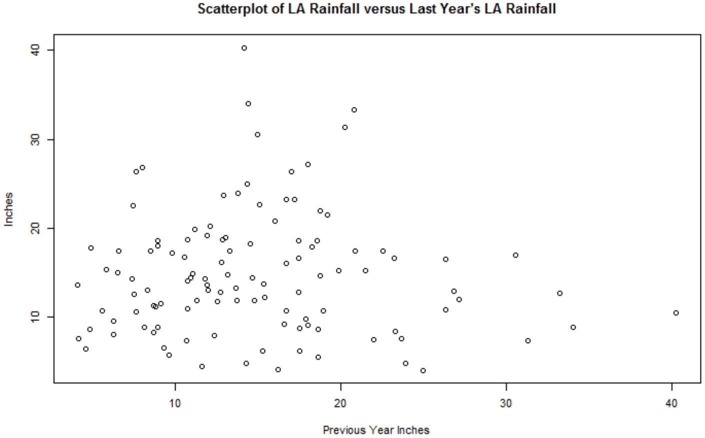
# Outline

# Motivation



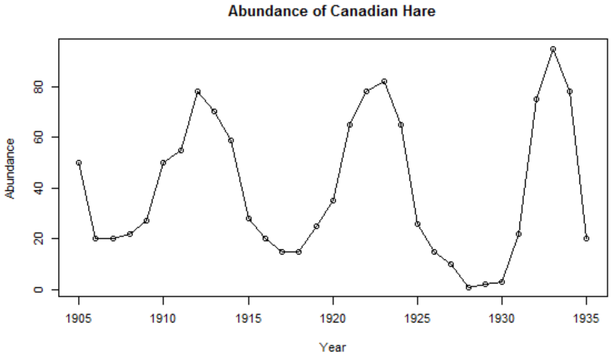
Time Series Plot of Los Angeles Annual Rainfall
Inches
10
20
30
40
1880
1900
1920
1940
1960
1980
Year

# Motivation



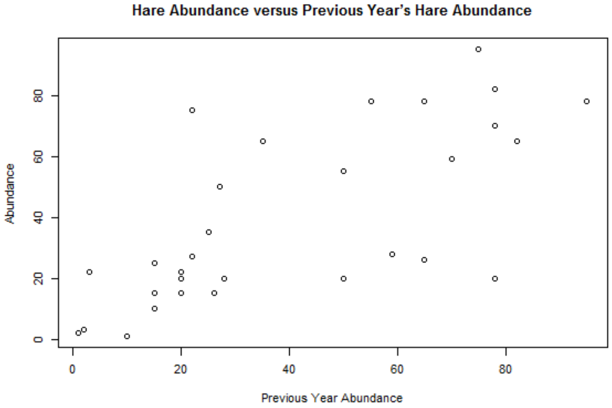
Scatterplot of LA Rainfall versus Last Year's LA Rainfall
Inches
Previous Year Inches
10
20
30
40

# Motivation



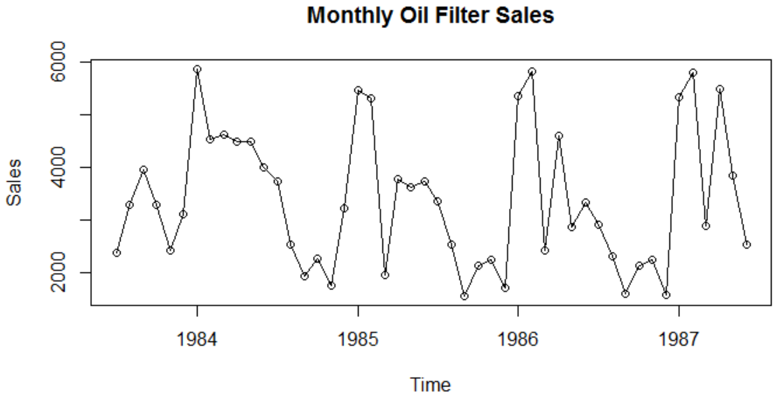
Abundance of Canadian Hare
Abundance
0
20
40
60
80
1905
1910
1915
1920
1925
1930
1935
Year

# Motivation



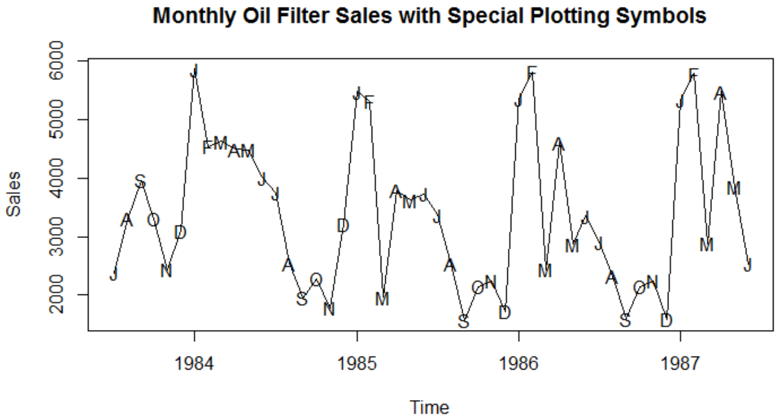
Hare Abundance versus Previous Year's Hare Abundance
Abundance
0
20
40
60
80
0
20
40
60
80
Previous Year Abundance

# Motivation



Monthly Oil Filter Sales
Sales
6000
4000
2000
1984
1985
1986
1987
Time

# Motivation



Monthly Oil Filter Sales with Special Plotting Symbols
Sales
2000
3000
4000
5000
6000
1984
1985
1986
1987
Time

# Outline

# Formal Definitions

A stochastic process is a collection of time indexed random variables

$$(Z_t, t \in T) = (Z_t(\omega), t \in T, \omega \in \Omega)$$

defined on some space $\Omega$.
Suppose that

- for a fixed $t \rightarrow Z_t(\omega)$, $Z_t : \Omega \rightarrow \mathbb{R}$ This is just a random variable.
- for fixed $\omega \rightarrow Z_\omega : T \rightarrow \mathbb{R}$ This is a realization or sample function.

Changing the time index, we can generate several random variables:

$$Z_{t_1}(\omega), Z_{t_2}(\omega), ..., Z_{t_n}(\omega)$$

From which a realization is:

$$Z_{t_1}, Z_{t_2}, ..., Z_{t_n}$$

This collection of random variables is called a **stochastic process**
A realization of the stochastic process is called a **time series**

# Example of stochastic processes

## Example 1

Let the index set be $T = \{1, 2, 3\}$ and let the space of outcomes $(\Omega)$ be the possible outcomes associated with tossing one dice:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Define

$$Z(t, \omega) = t + [\text{value on dice}]^2 t$$

Therefore for a particular $\omega$, say $\omega_3 = \{3\}$, the realization or path would be $(10, 20, 30)$.

# Example of stochastic processes

## Example 2

A Brownian Motion $B = (B_t, t \in [0, \infty])$:

- it starts at zero, $B_0 = 0$
- It has stationary, independent increments
- For every $t > 0$, $B_t$ has a normal $N(0, t)$ distribution
- It has continuous sample paths: no jumps.

# Means, Autocovariances, and autocorrelations

For a stochastic process $\{Y_t : t = 0, \pm 1, \pm 2, ...\}$:

- **Mean function**:

$$\mu_t = \mathbb{E}[Y_t]$$

  for $t = 0, \pm 1, \pm 2, ...$

- **Autocovariance function**:

$$\gamma_{t,s} = Cov(Y_t, Y_s)$$

  for $t, s = 0, \pm 1, \pm 2, ...$

- **Autocorellation function**:

$$\rho_{t,s} = Corr(Y_t, Y_s)$$

  for $t, s = 0, \pm 1, \pm 2, ...$

# White Noise Processes

A process $\{a_t\}$ is called a **white noise process** if it is a sequence of uncorrelated random variables from a fixed distribution with constant mean $\mathbb{E}[a_t] = \mu_a$, usually assumed to be 0, constant variance $Var(a_t) = \sigma_a^2$, and $\rho_k = Cov(a_t, a_{t+k}) = 0$ for all $k \neq 0$. By definition, white noise process $\{a_t\}$ is **stationary**.

# The Random Walk

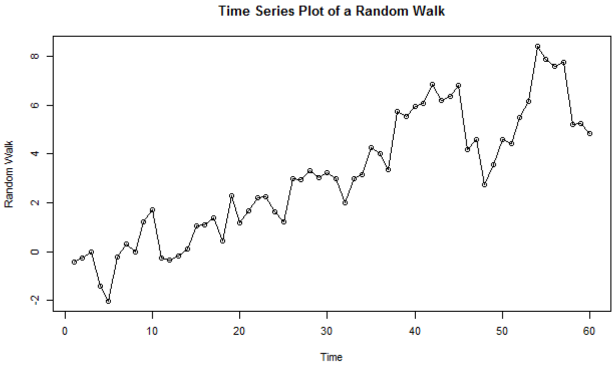
Let $a_1, a_2, ...$ be a sequence of independent, identically distributed random variables each with zero mean and variance $\sigma_a^2$. The observed time series, $\{Y_t : t = 1, 2, ...\}$, is constructed as follows:

$$
\begin{aligned}
Y_1 &= a_1 \\
Y_2 &= a_1 + a_2 \\
&... \\
Y_t &= a_1 + a_2 + ... + a_t
\end{aligned}
$$

Alternatively, we can write:

$$Y_t = Y_{t-1} + a_t$$

# The Random Walk



# Stationarity

A process $\{Y_t\}$ is said to be **strictly stationary** if the joint distribution of $Y_{t_1}, Y_{t_2}, ..., Y_{t_n}$ is the same as the joint distribution of $Y_{t_1-k}, Y_{t_2-k}, ..., Y_{t_n-k}$.

A process $\{Y_t\}$ is said to be **weakly stationary** if

1. The mean function $(\mu_t)$ is constant for all $t$, $\mu_t = \mu$
2. The variance function $(\sigma_t^2)$ is constant for all $t$, $\sigma_t^2 = \sigma^2$
3. The autocovariance function between $Y_{t_1}$ and $Y_{t_2}$ only depends on the interval $t_1$ and $t_2$

# Outline

# Moving Average Processes

A general linear process, $\{Y_t\}$, is one that can be represented as a weighted linear combination of present and past white noise terms as:

$$Y_t = a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + \dots$$

where $\sum_{i=0}^{\infty} \Psi_i^2 < \infty$
In the case where only a finite number of the $\Psi$-weights are nonzero, we have what is called a **moving average process**:

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots$$

# Autoregressive Processes

Autoregressive processes are as their name suggests —regressions on themselves. Specifically, a $p$th-order **autoregressive process** $\{Y_t\}$ satisfies the equation

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + e_t$$

## ARIMA Model

General form of ARIMA model:

$$\phi_p(B)(1-B)^d Y_t = \theta_0 + \theta_q(B)a_t$$

where,

$$\begin{aligned}
\theta_0 &= \mu(1-\phi_1-\phi_2-...-\phi_p) \\
\phi_p(B) &= 1-\phi_1 B-\phi_2 B^2-...-\phi_p B^p \\
\theta_q(B) &= 1-\theta_1 B-\theta_2 B^2-...-\theta_q B^q \\
BY_t &= Y_{t-1}
\end{aligned}$$

$Y_t$: actual value, $B$: backshift operator, $a_t$: white noise, $a_t \sim WN(0, \sigma^2)$, $\phi_i(i=1,2,...,p)$,$\theta_j(j=1,2,...,q)$, $\mu$: model parameters, $d$: differencing order.

# Autocorrelations and Partial Autocorrelations

ACF:

$$\rho_k = \frac{Cov(Y_t, Y_{t-k})}{Var(Y_t)} = \frac{\gamma_k}{\gamma_0}$$

Sample ACF:

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{T}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{T}(Y_t - \bar{Y})^2}$$

**Partial autocorrelation (PACF)** between $Y_t$ and $Y_{t-k}$ is defined as the correlation between $Y_t$ and $Y_{t-k}$ after the intervening variables $Y_{t-1}, Y_{t-2}, ..., Y_{t-(k-1)}$ have been removed. The conditional correlation

$$Corr(Y_t, Y_t - k \mid Y_{t-1}, ..., Y_{t-(k-1)})$$

is called partial autocorrelation.

# Autocorrelations and Partial Autocorrelations

Sample PACF:

$$\hat{\phi}_{11} = \hat{\rho}_1$$

$$\hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^{k} \hat{\phi}_{kj} \hat{\rho_{k+1-j}}}{1 - \sum_{j=1}^{k} \hat{\phi}_{kj} \hat{\rho}_j}$$

# Build ARIMA Model: Box-Jenkins Procedure

The Box-Jenkins method refers to the iterative application of the following three steps:

1. **Identification**. Using plots of the data, autocorrelations, partial autocorrelations, and other information, a class of simple ARIMA models is selected. This amounts to estimating appropriate values for $p$, $d$, and $q$.
2. **Estimation**. The $\phi$ and $\theta$ of the selected model are estimated using conditional least square, maximum likelihood techniques, backcasting, etc.
3. **Diagnostic Checking**. The fitted model is checked for inadequacies by considering the autocorrelations of the residual series (the series of residual, or error, values).

# ACF and PACF Patterns

| **Model** | **ACF** | **PACF** |
|---|---|---|
| ARIMA$(p, d, 0)$ | Infinite. Tails off. | Finite. |
| | | Cuts off after $p$ lags. |
| ARIMA$(0, d, q)$ | Finite. Cuts off after q lags. | Infinite. Tails off. |
| ARIMA$(p, d, q)$ | Infinite. Tails off. | Infinite. Tails off. |

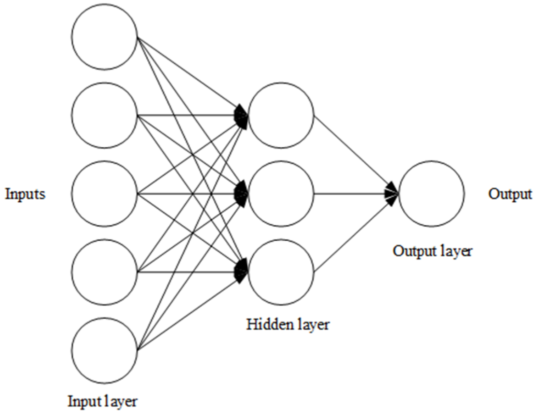
## ANN Model

General form of single hidden layer feedforward network:

$$Y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j \, g\left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} Y_{t-i}\right) + \varepsilon_t$$

where $\alpha_j (j = 0, 1, 2, ..., q)$,$\beta_{ij} (i = 0, 1, 2, ..., p; j = 1, 2, ..., q)$ are model parameters (connection weights), $p$ is the number of nodes in input layer, $q$ is the number of nodes in hidden layer, and $g(.)$ is the hidden layer activation function.

## Neural Architecture



# Outline

# Prediction of Roll Motion

Data Set: series of rolling motion (degree) of a Floating Production Unit
Number of observation : 3250 time points

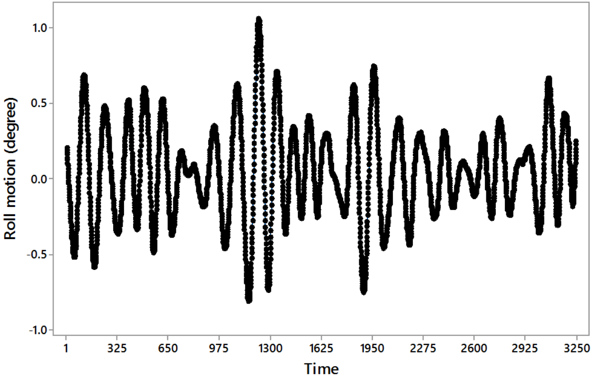



Figure: Time series plot of roll motion

# Ship motions

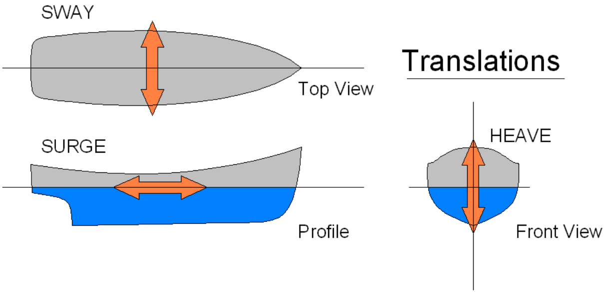



Figure: Translation motions

# Ship motions

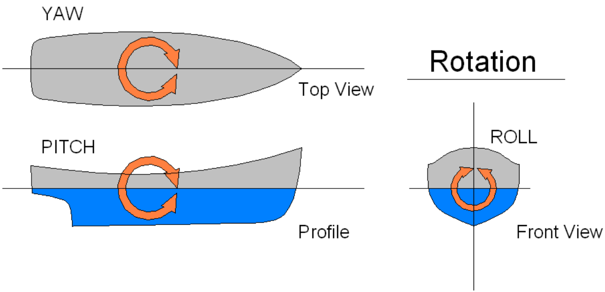



Figure: Rotation motions

# Methodology

- Partitioning the data: 3000 in-sample, 250 out-of-sample
- in-sample: modeling the data
- out-of-sample: forecast evaluation
- Measuring forecast accuracy: Root Mean Squared Error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{L}\sum_{l=1}^{L}\left(Y_{n+l} - \hat{Y}_n(l)\right)^2}$$

$L$: size of out-of-sample
$Y_{n+l}$: $l$-th actual value of out-of-sample data
$\hat{Y}_n(l)$: $l$-th forecast

DEMO

http://bit.ly/TSPyConID2017