Predictive Modeling

Forecasting

Mirko Birbaumer

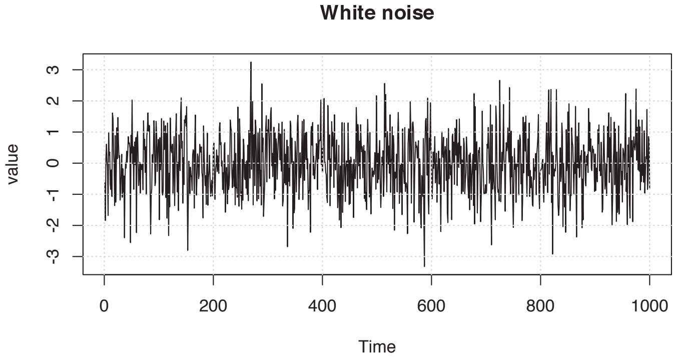
HSLU T&A

1. Repetition : Models of Stochastic Processes

2. Forecasting

Repetition : White Noise

- A white noise process consists of independent and identically distributed random variables $\{W_1, W_2, \dots\}$ where each W_i has mean 0 and variance σ^2

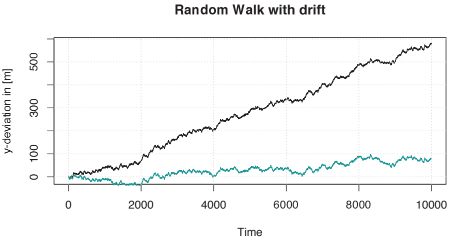


Repetition : Random Walk

- Choose n independent Bernoulli random variables $D_1, \ldots, D_n$ that take on the values 1 and -1 with equal probability, i.e. $p = 0.5$

- Define the random variables $X_i = D_1 + \cdots + D_i$ for each $1 \leq i \leq n$. Then $\{X_1, X_2, \ldots\}$ is a discrete stochastic process modeling the random walk

- Random walk with a drift : $X_i = X_{i-1} + \delta + W_i$



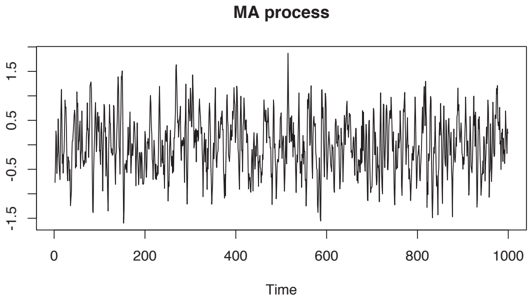
Repetition : Moving Average Process

- We apply a sliding window filter to the white noise process

$$\{W_1, W_2, \dots\}$$

- We obtain a **moving average** process
- If we choose the window length to be 3, we obtain

$$V_i = \frac{1}{3}(W_{i-1} + W_i + W_{i+1}).$$



Repetition : Autoregressive Model

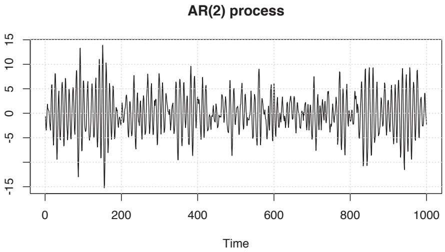
- We consider again the white noise process

$$\{W_1, W_2, \dots, W_n\}$$

- We recursively define the following sequence:

$$X_i = 1.5X_{i-1} - 0.9X_{i-2} + W_i$$

- In other words, the value of the process at time instance i is modeled as a linear combination of the past two values plus some random component : **autoregressive** process



Measures of Dependence

- Aside to the individual (also called **marginal**) distributions F_i of the random variables in the process

Measures of Dependence

- Aside to the individual (also called **marginal**) distributions F_i of the random variables in the process we define **first** and **second order** moments to analyse the whole process.

Measures of Dependence

- Aside to the individual (also called **marginal**) distributions F_i of the random variables in the process we define **first** and **second order** moments to analyse the whole process.

- We start with the first order moments of the process, the **mean sequence**:

Mean sequence

The mean sequence $\{\mu(1), \mu(2), \dots\}$ (or mean function) of a discrete stochastic process $\{X_1, X_2, \dots\}$ is defined as the sequence of the means:

$$\mu(i) = \mathrm{E}[X_i].$$

Covariance

- As second order moments, we consider the covariance within a *single* process

Covariance

- As second order moments, we consider the covariance within a *single* process

Autocovariance and autocorrelation

Let $\{X_1, X_2, \dots\}$ be a discrete stochastic process.

1. The *autocovariance* γ_X is defined as

$$\gamma_X(i,j) = \mathrm{Cov}(X_i, X_j) = \mathrm{E}[(X_i - \mu(i))(X_j - \mu(j))].$$

Covariance

- As second order moments, we consider the covariance within a *single* process

Autocovariance and autocorrelation

Let $\{X_1, X_2, \dots\}$ be a discrete stochastic process.

1. The *autocovariance* γ_X is defined as

$$\gamma_X(i,j) = \mathrm{Cov}(X_i, X_j) = \mathrm{E}[(X_i - \mu(i))(X_j - \mu(j))].$$

2. The *autocorrelation* ρ_X is defined as

$$\rho_X(i,j) = \frac{\gamma_X(i,j)}{\sqrt{\gamma_X(i,i)\gamma_X(j,j)}}.$$

Repetition : Weak Stationarity

Repetition : Weak stationarity

A discrete stochastic process X_i is called *weakly stationary* if

1. the mean sequence $\mu_X(i)$ is constant and does not depend on the time index i and

Repetition : Weak Stationarity

Repetition : Weak stationarity

A discrete stochastic process X_i is called *weakly stationary* if

1. the mean sequence $\mu_X(i)$ is constant and does not depend on the time index i and
2. the autocovariance sequence $\gamma_X(i,j)$ depends on i and j only through their difference $|i-j|$

Repetition : Weak Stationarity

Repetition : Weak stationarity

A discrete stochastic process X_i is called *weakly stationary* if

1. the mean sequence $\mu_X(i)$ is constant and does not depend on the time index i and
2. the autocovariance sequence $\gamma_X(i,j)$ depends on i and j only through their difference $|i-j|$

- The first and simplest way to **test** whether a time series is weakly stationary consists of looking for evidence of trend in mean sequence or in the autocorrelation function

Repetition : Weak Stationarity

Repetition : Weak stationarity

A discrete stochastic process X_i is called *weakly stationary* if

1. the mean sequence $\mu_X(i)$ is constant and does not depend on the time index i and
2. the autocovariance sequence $\gamma_X(i, j)$ depends on i and j only through their difference $|i - j|$

- The first and simplest way to **test** whether a time series is weakly stationary consists of looking for evidence of trend in mean sequence or in the autocorrelation function
- If any such patterns are present then these are **signs of non-stationarity**

Repetition : Estimation of the autocovariance

Due to the stationarity of a process we know that the mean sequence $\mu(k) = \mu$ is constant. A canonical estimator : $\widehat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

Repetition : Estimation of the autocovariance

Due to the stationarity of a process we know that the mean sequence $\mu(k) = \mu$ is constant. A canonical estimator : $\widehat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Sample Autocovariance

1. The *sample autocovariance* is defined by

$$\widehat{\gamma}(h) = \frac{1}{n}\sum_{i=1}^{n-h}(x_{i+h} - \bar{x})(x_i - \bar{x})$$

with $\widehat{\gamma}(-h) = \widehat{\gamma}(h)$ for $h = 0, 1, \ldots, n-1$

Repetition : Estimation of the autocovariance

Due to the stationarity of a process we know that the mean sequence $\mu(k) = \mu$ is constant. A canonical estimator : $\widehat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Sample Autocovariance

1. The *sample autocovariance* is defined by

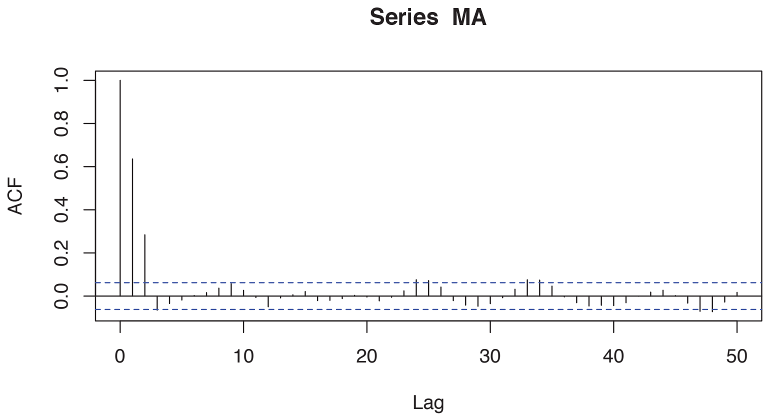
$$\widehat{\gamma}(h) = \frac{1}{n}\sum_{i=1}^{n-h}(x_{i+h} - \bar{x})(x_i - \bar{x})$$

   with $\widehat{\gamma}(-h) = \widehat{\gamma}(h)$ for $h = 0, 1, \ldots, n-1$

2. The *sample autocorrelation* is defined by

$$\widehat{\rho}(h) = \frac{\widehat{\gamma}(h)}{\widehat{\gamma}(0)}.$$

Repetition : Sample ACF of Simulated MA(5) Process



Repetition : ACF of White Noise

- For a **white noise process** for large n :

Repetition : ACF of White Noise

- For a **white noise process** for large n : distribution of the sample autocorrelation $\widehat{\rho}(h)$ for a fixed h is approximately normal

Repetition : ACF of White Noise

- For a **white noise process** for large n : distribution of the sample autocorrelation $\widehat{\rho}(h)$ for a fixed h is approximately normal with mean $-1/n$

Repetition : ACF of White Noise

- For a **white noise process** for large n : distribution of the sample autocorrelation $\widehat{\rho}(h)$ for a fixed h is approximately normal with mean $-1/n$ and standard deviation

$$\sigma_{\widehat{\rho}} = \frac{1}{\sqrt{n}}.$$

Repetition : ACF of White Noise

- For a **white noise process** for large n : distribution of the sample autocorrelation $\widehat{\rho}(h)$ for a fixed h is approximately normal with mean $-1/n$ and standard deviation

$$\sigma_{\widehat{\rho}} = \frac{1}{\sqrt{n}}.$$

- Used to check whether a time series stems from a white noise process or not:

Repetition : ACF of White Noise

- For a **white noise process** for large n : distribution of the sample autocorrelation $\widehat{\rho}(h)$ for a fixed h is approximately normal with mean $-1/n$ and standard deviation

$$\sigma_{\widehat{\rho}} = \frac{1}{\sqrt{n}}.$$

- Used to check whether a time series stems from a white noise process or not: If the series is white, it should stay 95% of the time within the $-1/n \pm 2/\sqrt{n}$ limits.

Repetition : ACF of White Noise

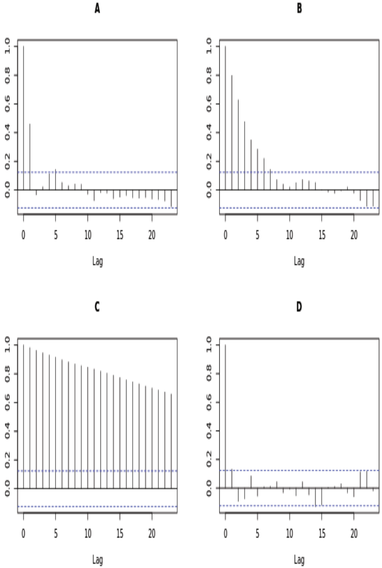
- For a **white noise process** for large n : distribution of the sample autocorrelation $\widehat{\rho}(h)$ for a fixed h is approximately normal with mean $-1/n$ and standard deviation

$$\sigma_{\widehat{\rho}} = \frac{1}{\sqrt{n}}.$$

- Used to check whether a time series stems from a white noise process or not: If the series is white, it should stay 95% of the time within the $-1/n \pm 2/\sqrt{n}$ limits.

- These limits are automatically drawn by `R` or `Pyton` when calling `acf` (**blue lines**)

Clicker Question



A time series from each of the four models you have considered in this course was simulated and their sample autocorrelation functions (ACF) are shown in one of the four panels in the adjoining figure. They include the white noise (WN), random walk (RW), autoregressive (AR), and simple moving average (MA) models. Match each sample ACF plot with one of our models WN, RW, AR, MA.

- (A) MA, (B) RW, (C) AR, (D) WN
- (A) RW, (B) MA, (C) WN, (D) AR
- (A) WN, (B) RW, (C) AR, (D) MA
- (A) MA, (B) AR, (C) RW, (D) WN

Forecasting

- Now we will model given time series data with a proper stochastic process to **forecast** future and unobserved values of the process

Forecasting

- Now we will model given time series data with a proper stochastic process to **forecast** future and unobserved values of the process
- Goal : **predict** future values x_{n+k} with $k = 1, 2, \ldots$ given a time series up to the present time $\{x_1, \ldots, x_n\}$

Forecasting : Procedure

There are three steps that have to be carried out subsequently, in order to achieve this goal:

1. We need to be certain that the underlying process is **predictible**, i.e. in the future the process will not change dramatically but continues as it has up to the present (in a probabilistic sense)

Forecasting : Procedure

There are three steps that have to be carried out subsequently, in order to achieve this goal:

1. We need to be certain that the underlying process is **predictible**, i.e. in the future the process will not change dramatically but continues as it has up to the present (in a probabilistic sense)

2. We choose a **model class** by explorative data analysis of the given time series. After that we **fit** the model to the training data and receive model parameters that fully describe the fitted model

Forecasting : Procedure

There are three steps that have to be carried out subsequently, in order to achieve this goal:

1. We need to be certain that the underlying process is **predictible**, i.e. in the future the process will not change dramatically but continues as it has up to the present (in a probabilistic sense)

2. We choose a **model class** by explorative data analysis of the given time series. After that we **fit** the model to the training data and receive model parameters that fully describe the fitted model

3. With the fitted model we **predict** future values of the process

Autoregressive Models

Autoregressive models are based on the idea that the current value of the series can be **explained** by a **linear combination** of the p previous values:

Autoregressive Models

Autoregressive models are based on the idea that the current value of the series can be **explained** by a **linear combination** of the p previous values:

Autoregressive model AR(p)

The autoregressive model of order p is a discrete stochastic process

Autoregressive Models

Autoregressive models are based on the idea that the current value of the series can be **explained** by a **linear combination** of the p previous values:

Autoregressive model AR(p)

The autoregressive model of order p is a discrete stochastic process that satisfies

$$X_n = a_1 X_{n-1} + a_2 X_{n-2} + \cdots + a_p X_{n-p} + W_n$$

where $a_1, a_2, \ldots, a_n$ are the model parameters and $W_1, W_2, \ldots$ is a white noise process with variance σ^2.

Autoregressive Models

Autoregressive models are based on the idea that the current value of the series can be **explained** by a **linear combination** of the p previous values:

Autoregressive model AR(p)

The autoregressive model of order p is a discrete stochastic process that satisfies

$$X_n = a_1 X_{n-1} + a_2 X_{n-2} + \cdots + a_p X_{n-p} + W_n$$

where $a_1, a_2, \ldots, a_n$ are the model parameters and $W_1, W_2, \ldots$ is a white noise process with variance σ^2. For a given autoregressive process, the **characteristic polynomial** is defined as

$$\Phi(x) = 1 - a_1 x - a_2 x^2 - \cdots - a_p x^p.$$

Example : AR(1)

- Autoregressive models are **exclusively** used for the modeling of **stationary** processes

Example : AR(1)

- Autoregressive models are **exclusively** used for the modeling of **stationary** processes

- We will give a sufficient condition on the parameters of an autoregressive process that render the process stationary

Example : AR(1)

- Autoregressive models are **exclusively** used for the modeling of **stationary** processes
- We will give a sufficient condition on the parameters of an autoregressive process that render the process stationary
- AR(1) process is defined by

$$X_n = a_1 X_{n-1} + W_n$$

- Remark : random walk a special case of an AR(1) with $a_1 = 1$, but non-stationary

Example : AR(1)

- Expected value of the process by taking expectations on both sides of the equation: $X_n = a_1 X_{n-1} + W_n$

$$\mu = \mathrm{E}(X_n) = a_1 \mathrm{E}(X_{n-1}) + \mathrm{E}(W_n) = a_1 \mu + 0$$

Example : AR(1)

- Expected value of the process by taking expectations on both sides of the equation: $X_n = a_1 X_{n-1} + W_n$

$$\mu = \mathrm{E}(X_n) = a_1 \mathrm{E}(X_{n-1}) + \mathrm{E}(W_n) = a_1 \mu + 0$$

- Thus, $(a_1 - 1)\mu = 0$ which implies that $\mu = 0$, if $a_1 \neq 1$

Example : AR(1)

- Expected value of the process by taking expectations on both sides of the equation: $X_n = a_1 X_{n-1} + W_n$

$$\mu = \mathrm{E}(X_n) = a_1 \mathrm{E}(X_{n-1}) + \mathrm{E}(W_n) = a_1 \mu + 0$$

- Thus, $(a_1 - 1)\mu = 0$ which implies that $\mu = 0$, if $a_1 \neq 1$

- Conclusion :
  - I.e. if the process is **stationary**, then the mean function is $\mu(i) = 0$

Example : AR(1)

- Expected value of the process by taking expectations on both sides of the equation: $X_n = a_1 X_{n-1} + W_n$

$$\mu = \mathrm{E}(X_n) = a_1 \mathrm{E}(X_{n-1}) + \mathrm{E}(W_n) = a_1 \mu + 0$$

- Thus, $(a_1 - 1)\mu = 0$ which implies that $\mu = 0$, if $a_1 \neq 1$

- Conclusion :
  - I.e. if the process is **stationary**, then the mean function is $\mu(i) = 0$
  - If the process is a **random walk**, i.e. non-stationary, with $a_1 = 1$, then μ may take on any value

Example : AR(1)

- Next we compute the **variance** of the process:

$$\sigma_X^2 = \mathrm{Var}(X_n) = a_1^2 \mathrm{Var}(X_{n-1}) + \mathrm{Var}(W_n) = a_1^2 \sigma_X^2 + \sigma^2.$$

Example : AR(1)

- Next we compute the **variance** of the process:

$$\sigma_X^2 = \mathrm{Var}(X_n) = a_1^2 \mathrm{Var}(X_{n-1}) + \mathrm{Var}(W_n) = a_1^2 \sigma_X^2 + \sigma^2.$$

- So we find that $\sigma_X^2 = a_1^2 \sigma_X^2 + \sigma^2$ or, after some rearrangements:

$$\sigma_X^2 = \frac{\sigma^2}{1 - a_1^2}$$

Example : AR(1)

- Next we compute the **variance** of the process:

$$\sigma_X^2 = \mathrm{Var}(X_n) = a_1^2 \mathrm{Var}(X_{n-1}) + \mathrm{Var}(W_n) = a_1^2 \sigma_X^2 + \sigma^2.$$

- So we find that $\sigma_X^2 = a_1^2 \sigma_X^2 + \sigma^2$ or, after some rearrangements:

$$\sigma_X^2 = \frac{\sigma^2}{1 - a_1^2}$$

- To have a **non-negative** and constant value for the variance σ_X^2, the absolute value $|a_1|$ must be less than 1

Example : AR(1)

- Next we compute the **variance** of the process:

$$\sigma_X^2 = \mathrm{Var}(X_n) = a_1^2\mathrm{Var}(X_{n-1}) + \mathrm{Var}(W_n) = a_1^2\sigma_X^2 + \sigma^2.$$

- So we find that $\sigma_X^2 = a_1^2\sigma_X^2 + \sigma^2$ or, after some rearrangements:

$$\sigma_X^2 = \frac{\sigma^2}{1 - a_1^2}$$

- To have a **non-negative** and constant value for the variance σ_X^2, the absolute value $|a_1|$ must be less than 1

- Interpretation : in order to be **stationary**, the dependence of the process on past values should **not** be too strong.

General Condition for an AR(p) to be Weakly Stationary

Stationarity of AR(p)

An AR(p) stochastic process is weakly stationary, if all (complex) roots of the characteristic polynomial

$$\Phi(x) = 1 - a_1 x - a_2 x^2 - \cdots - a_p x^p.$$

exceed 1 in absolute value.

- Example AR(1) : characteristic polynomial is given by

$$\Phi(x) = 1 - a_1 x$$

General Condition for an AR(p) to be Weakly Stationary

Stationarity of AR(p)

An AR(p) stochastic process is weakly stationary, if all (complex) roots of the characteristic polynomial

$$\Phi(x) = 1 - a_1 x - a_2 x^2 - \cdots - a_p x^p.$$

exceed 1 in absolute value.

- Example AR(1) : characteristic polynomial is given by

$$\Phi(x) = 1 - a_1 x$$

- Single root of the polynomial Φ is

$$x = 1/a_1$$

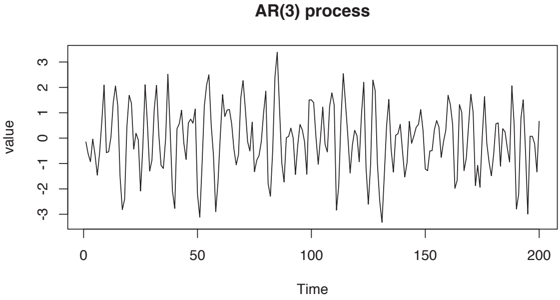
which exceeds 1 in absolute value if and only if $|a_1| < 1$

Example : AR(3) Process

- AR(3) process

$$X_n = 0.5X_{n-1} - 0.5X_{n-2} - 0.1X_{n-3} + W_n$$

- Roots of characteristic polynomial : $1.28, 1.28, 6.09 \rightarrow$ **stationary**
- Simulation of AR(3) process:



- See example 1.3 of Time Series Forecasting chapter

Autocorrelation of AR(p) processes

- If we intend to **fit an autoregressive model** to a given time series data set, two things have to be clarified in advance
  1. Is the **autoregressive model** the right choice for the given data?

Autocorrelation of AR(p) processes

- If we intend to **fit an autoregressive model** to a given time series data set, two things have to be clarified in advance
  1. Is the **autoregressive model** the right choice for the given data?
  2. What is the proper **model order** p for the given data?

Autocorrelation of AR(p) processes

- If we intend to **fit an autoregressive model** to a given time series data set, two things have to be clarified in advance
  1. Is the **autoregressive model** the right choice for the given data?
  2. What is the proper **model order** p for the given data?
- In practice : **sample autocorrelation** of a given series is computed and compared with the **theoretical autocorrelation** of the AR(p) model. In this way we can judge if the time series is autoregressive

Autocorrelation of AR(p) processes

- If we intend to **fit an autoregressive model** to a given time series data set, two things have to be clarified in advance
  1. Is the **autoregressive model** the right choice for the given data?
  2. What is the proper **model order** p for the given data?

- In practice : **sample autocorrelation** of a given series is computed and compared with the **theoretical autocorrelation** of the AR(p) model. In this way we can judge if the time series is autoregressive

- **Partial autocorrelation** is introduced as a further measure well suited for determining the **order** of an autoregressive model

Example : Theoretical Autocorrelation of AR(p) Process

- AR(1) process : theoretical autocorrelation at lag h is a_1^h (see lecture notes)

Example : Theoretical Autocorrelation of AR(p) Process

- AR(1) process : theoretical autocorrelation at lag h is a_1^h (see lecture notes)

- Autocorrelation of the AR(1) process decays exponentially fast

- Theoretical autocorrelation of a general AR(p) process with $p > 1$ can be computed in `R` and `Python`

Example : Theoretical Autocorrelation of AR(p) Process

- AR(1) process : theoretical autocorrelation at lag h is a_1^h (see lecture notes)

- Autocorrelation of the AR(1) process decays exponentially fast

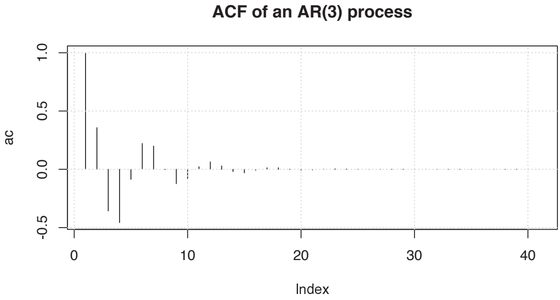
- Theoretical autocorrelation of a general AR(p) process with $p > 1$ can be computed in `R` and `Python`

- Example : AR(3) process defined by coefficients $a_1 = 0.5$, $a_2 = -0.5$ and $a_3 = -0.1$

Example : Theoretical Autocorrelation of AR(p) Process

- AR(1) process : theoretical autocorrelation at lag h is a_1^h (see lecture notes)
- Autocorrelation of the AR(1) process decays exponentially fast
- Theoretical autocorrelation of a general AR(p) process with $p > 1$ can be computed in `R` and `Python`
- Example : AR(3) process defined by coefficients $a_1 = 0.5$, $a_2 = -0.5$ and $a_3 = -0.1$
- Please check example `1.5` of `Forecasting` chapter

Example : Theoretical ACF of AR(3) Process



- As it can be seen in above figure, the theoretical autocorrelation of the given AR(3) is **oscillating** and **decreasing** - essentially following an exponential function
- This is the typical autocorrelation behaviour of an autoregressive process

Partial autocorrelation

- Previous examples indicate that the autocorrelation of an AR(p) process is nonzero for a wide span of lags

Partial autocorrelation

- Previous examples indicate that the autocorrelation of an AR(p) process is nonzero for a wide span of lags

- This is due to propagation of correlation through the model: If X_k is strongly correlated with X_{k+1} and in turn X_{k+1} is strongly correlated with X_{k+2}, then X_k will likely be correlated with X_{k+2} as well

Partial autocorrelation

- Previous examples indicate that the autocorrelation of an AR(p) process is nonzero for a wide span of lags
- This is due to propagation of correlation through the model: If X_k is strongly correlated with X_{k+1} and in turn X_{k+1} is strongly correlated with X_{k+2}, then X_k will likely be correlated with X_{k+2} as well
- If we want to study the **direct** correlation between X_k and X_{k+2}, i.e. the proportion of correlation that is **not** due to X_{k+1}, we have to compute the **partial autocorrelation**

Partial Autocorrelation

Partial autocorrelation

For a weakly stationary stochastic process $\{X_1, X_2, \dots\}$ the partial autocorrelation is defined as

$$\pi(h) = \mathrm{Cor}(X_k, X_{k+h} | X_{k+1}, \dots, X_{k+h-1})^a.$$

[a] The quantity $\mathrm{Cor}(X, Y|Z)$ denotes the conditional correlation of X and Y given the value of Z

Two important properties:

1. The p-th coefficient α_p of an AR(p) process equals $\pi(p)$, i.e. the partial autocorrelation value at lag p of the process

Partial Autocorrelation

Partial autocorrelation

For a weakly stationary stochastic process $\{X_1, X_2, \dots\}$ the partial autocorrelation is defined as

$$\pi(h) = \mathrm{Cor}(X_k, X_{k+h} | X_{k+1}, \dots, X_{k+h-1})^a.$$

[a] The quantity $\mathrm{Cor}(X, Y|Z)$ denotes the conditional correlation of X and Y given the value of Z

Two important properties:

1. The p-th coefficient α_p of an AR(p) process equals $\pi(p)$, i.e. the partial autocorrelation value at lag p of the process
2. For an autoregressive process AR(p) the partial autocorrelation at lags greater than p is zero, i.e. $\pi(k) = 0$ for $k > p$.

Order of AR(p) Process

- How do we investigate a given autoregressive time series and determine the order of the model?
  1. Compute the partial autocorrelation

Order of AR(p) Process

- How do we investigate a given autoregressive time series and determine the order of the model?
  1. Compute the partial autocorrelation
  2. Choose the largest lag k for which the value $\pi(k)$ is not zero

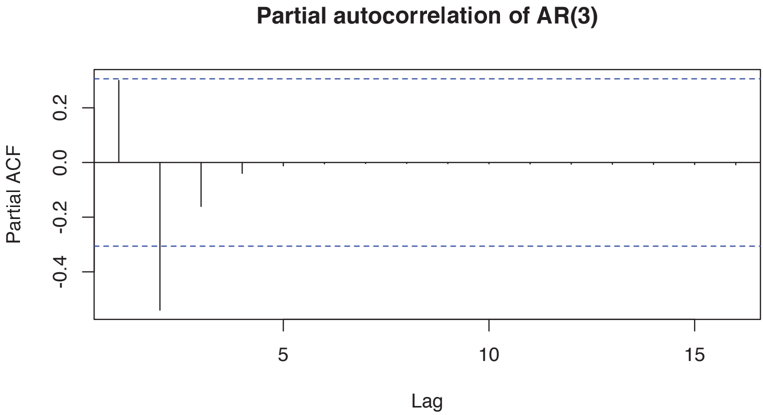
Order of AR(p) Process

- How do we investigate a given autoregressive time series and determine the order of the model?
  1. Compute the partial autocorrelation
  2. Choose the largest lag k for which the value $\pi(k)$ is not zero $\rightarrow$ order p of AR(p) model

Order of AR(p) Process

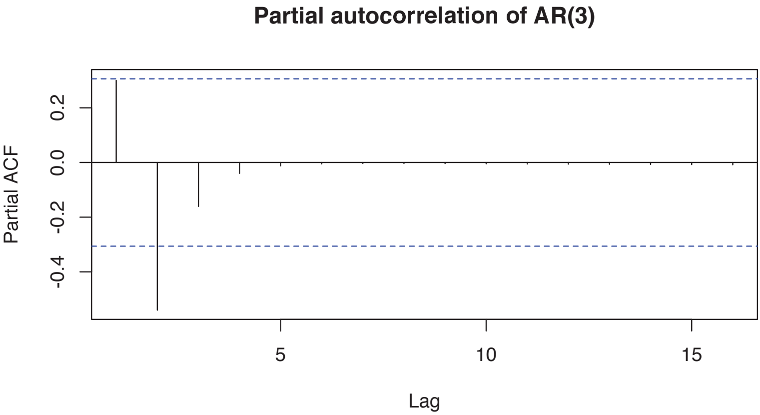
- How do we investigate a given autoregressive time series and determine the order of the model?
  1. Compute the partial autocorrelation
  2. Choose the largest lag k for which the value $\pi(k)$ is not zero $\rightarrow$ order p of AR(p) model
- Please check example `1.6` of the **`Forecasting`** chapter

Example : AR(3) Process



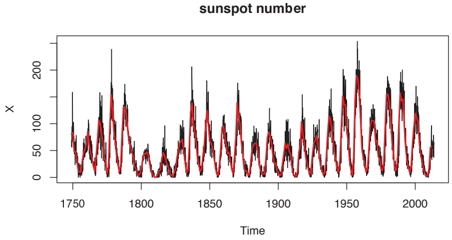
Partial autocorrelation coefficients **larger** than 3 are almost zero :

Example : AR(3) Process



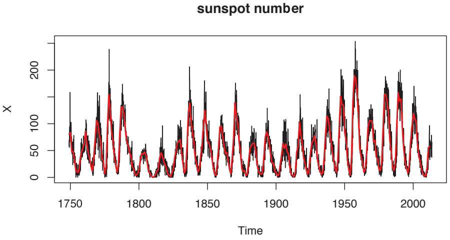
Partial autocorrelation coefficients **larger** than 3 are almost zero : choose an autoregressive model of **order** 3 for the present sequence

Example : Sunspot number



- Forcasting solar activity is important for satellite drag, telecommunication outages and solar winds in connection with blackout of powerplants

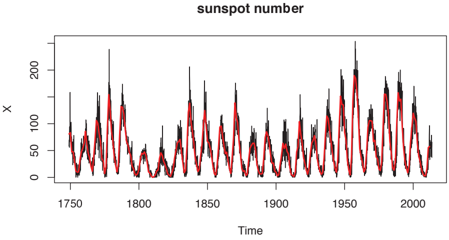
Example : Sunspot number



- Forcasting solar activity is important for satellite drag, telecommunication outages and solar winds in connection with blackout of powerplants

- Indicator of solar activity : **sunspot number**

Example : Sunspot number

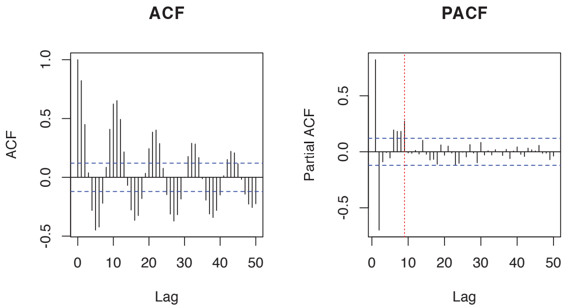


- Forcasting solar activity is important for satellite drag, telecommunication outages and solar winds in connection with blackout of powerplants

- Indicator of solar activity : **sunspot number**

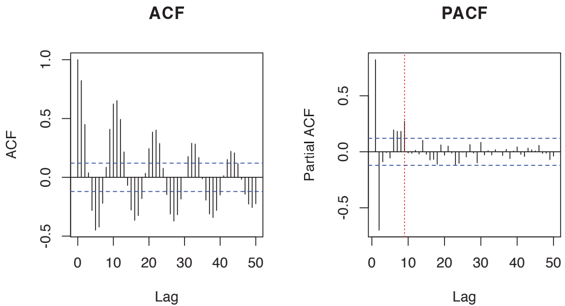
- Swiss astronomer Johann Rudolph Wolf introduced the sunspot number in 1848 and the number of sunspots has been recorded on a monthly basis back to the year 1749

Example : Sunspot Number



- Autocorrelation shows typical behaviour of an autoregressive process: an **oscillating pattern with an exponential decay**

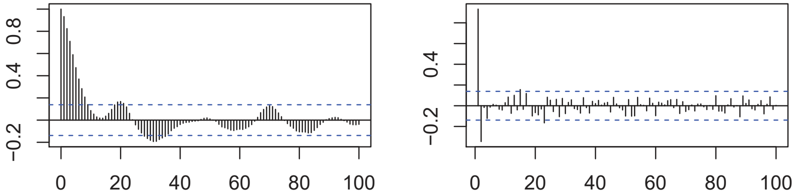
Example : Sunspot Number



- Autocorrelation shows typical behaviour of an autoregressive process: an **oscillating pattern with an exponential decay**
- Partial autocorrelation : **maximum lag of** 9 which we will use as our model parameter p
- Please check example 1.7 of the Forecasting chapter and solve exercise 1

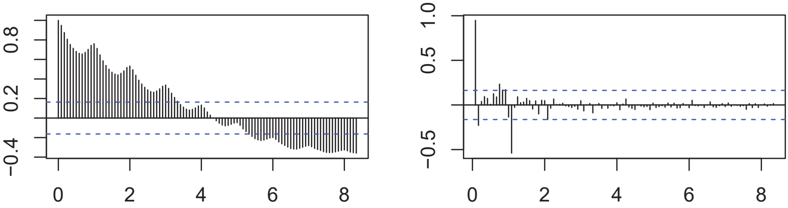
Clicker Question 1

In the following there is a pair of sample autocorrelation/partial autocorrelation plots belonging to a time series that is not shown. Decide whether the process is autoregressive or not. If yes, estimate the model order p.

a) 

Clicker Question 2

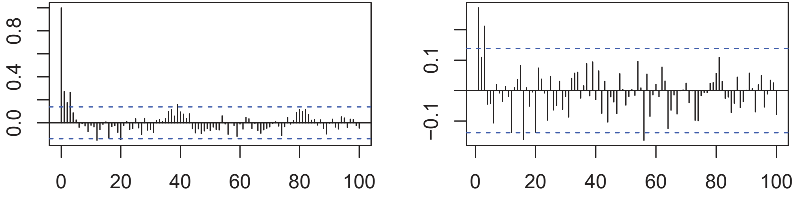
In the following there is a pair of sample autocorrelation/partial autocorrelation plots belonging to a time series that is not shown. Decide whether the process is autoregressive or not. If yes, estimate the model order p.



b)

Clicker Question 3

In the following there is a pair of sample autocorrelation/partial autocorrelation plots belonging to a time series that is not shown. Decide whether the process is autoregressive or not. If yes, estimate the model order p.



c)

Model Fitting Procedure

- *Task*: estimation of an AR(p) model from given time series data

Model Fitting Procedure

- *Task*: estimation of an AR(p) model from given time series data

- *Assumption*: Time series $\{x_1, x_2, \ldots, x_n\}$ is a realization of a **stationary** stochastic process

Model Fitting Procedure

- *Task*: estimation of an AR(p) model from given time series data

- *Assumption*: Time series $\{x_1, x_2, \ldots, x_n\}$ is a realization of a **stationary** stochastic process

- Fitting procedure:
  1. Decide whether a given time series stems from an autoregressive process, by investigating the autocorrelation and partial autocorrelation of the sequence.

Model Fitting Procedure

- *Task*: estimation of an AR(p) model from given time series data

- *Assumption*: Time series $\{x_1, x_2, \ldots, x_n\}$ is a realization of a **stationary** stochastic process

- Fitting procedure:
  1. Decide whether a given time series stems from an autoregressive process, by investigating the autocorrelation and partial autocorrelation of the sequence. If the autocorrelation is **decaying exponentially** to zero (possibly with oscillations) and the partial autocorrelation is zero for larger lags, then the autoregressive model is appropriate

Model Fitting Procedure

- *Task*: estimation of an AR(p) model from given time series data
- *Assumption*: Time series $\{x_1, x_2, \ldots, x_n\}$ is a realization of a **stationary** stochastic process
- Fitting procedure:
  1. Decide whether a given time series stems from an autoregressive process, by investigating the autocorrelation and partial autocorrelation of the sequence. If the autocorrelation is **decaying exponentially** to zero (possibly with oscillations) and the partial autocorrelation is zero for larger lags, then the autoregressive model is appropriate
  2. Choose the **largest lag** p such that the partial autocorrelation is **not** zero

Model Fitting Procedure

- *Task*: estimation of an AR(p) model from given time series data

- *Assumption*: Time series $\{x_1, x_2, \ldots, x_n\}$ is a realization of a **stationary** stochastic process

- Fitting procedure:
  1. Decide whether a given time series stems from an autoregressive process, by investigating the autocorrelation and partial autocorrelation of the sequence. If the autocorrelation is **decaying exponentially** to zero (possibly with oscillations) and the partial autocorrelation is zero for larger lags, then the autoregressive model is appropriate

  2. Choose the **largest lag** p such that the partial autocorrelation is **not** zero

  3. Choose **parameters** $a_1, \ldots, a_p$ such that given data are likely to be realizations of corresponding autoregressive process

Model Fitting

- Given the data $\{x_1, x_2, \ldots, x_n\}$ and a model order p, we fit the AR(p) process to the data by solving the following linear equation system

$$
\begin{aligned}
x_{p+1} &= a_1 x_p + a_2 x_{p-1} + \cdots + a_p x_1 + W_{p+1} \\
x_{p+2} &= a_1 x_{p+1} + a_2 x_p + \cdots + a_p x_2 + W_{p+2} \\
&\vdots \\
x_n &= a_1 x_{n-1} + a_2 x_{n-2} + \cdots + a_p x_{n-p} + W_n
\end{aligned}
$$

- System is solved in the least squares sense with the solution : estimates $\widehat{a}_1, \ldots, \widehat{a}_p$

Model Fitting

- Given the data $\{x_1, x_2, \ldots, x_n\}$ and a model order p, we fit the AR(p) process to the data by solving the following linear equation system

$$
\begin{aligned}
x_{p+1} &= a_1 x_p + a_2 x_{p-1} + \cdots + a_p x_1 + W_{p+1} \\
x_{p+2} &= a_1 x_{p+1} + a_2 x_p + \cdots + a_p x_2 + W_{p+2} \\
&\vdots \\
x_n &= a_1 x_{n-1} + a_2 x_{n-2} + \cdots + a_p x_{n-p} + W_n
\end{aligned}
$$

- System is solved in the least squares sense with the solution : estimates $\widehat{a}_1, \ldots, \widehat{a}_p$
- 3 further standard methods for estimating the AR(p) coefficients:
  1. Burg's algorithm

Model Fitting

- Given the data $\{x_1, x_2, \ldots, x_n\}$ and a model order p, we fit the AR(p) process to the data by solving the following linear equation system

$$
\begin{aligned}
x_{p+1} &= a_1 x_p + a_2 x_{p-1} + \cdots + a_p x_1 + W_{p+1} \\
x_{p+2} &= a_1 x_{p+1} + a_2 x_p + \cdots + a_p x_2 + W_{p+2} \\
&\vdots \\
x_n &= a_1 x_{n-1} + a_2 x_{n-2} + \cdots + a_p x_{n-p} + W_n
\end{aligned}
$$

- System is solved in the least squares sense with the solution : estimates $\widehat{a}_1, \ldots, \widehat{a}_p$
- 3 further standard methods for estimating the AR(p) coefficients:
  1. Burg's algorithm
  2. Yule-Walker

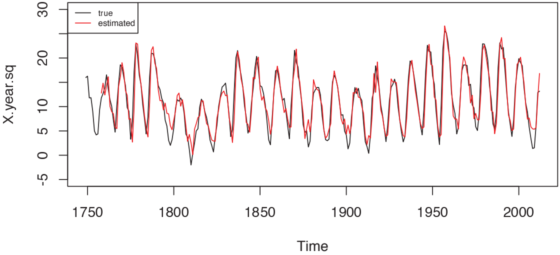
Model Fitting

- Given the data $\{x_1, x_2, \ldots, x_n\}$ and a model order p, we fit the AR(p) process to the data by solving the following linear equation system

$$\begin{aligned}
x_{p+1} &= a_1 x_p + a_2 x_{p-1} + \cdots + a_p x_1 + W_{p+1} \\
x_{p+2} &= a_1 x_{p+1} + a_2 x_p + \cdots + a_p x_2 + W_{p+2} \\
&\vdots \\
x_n &= a_1 x_{n-1} + a_2 x_{n-2} + \cdots + a_p x_{n-p} + W_n
\end{aligned}$$

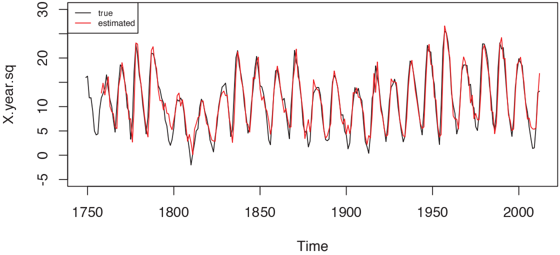
- System is solved in the least squares sense with the solution : estimates $\widehat{a}_1, \ldots, \widehat{a}_p$
- 3 further standard methods for estimating the AR(p) coefficients:
  1. Burg's algorithm
  2. Yule-Walker
  3. Maximum Likelihood Method

Example : Sunspot Number



- Plot shows the annually averaged time series in black and the output of the model in red

Example : Sunspot Number



- Plot shows the annually averaged time series in black and the output of the model in red
- Please check example 1.8 of `Forecasting` chapter

Forecasting AR(p) processes

- The general methodology of forecasting stationary time series can be summarized as follows:

k-step ahead forecast

Assume that $\{X_1, X_2, \ldots,\}$ is a stationary process and that we have observed a times series $\{x_1, x_2, \ldots, x_n\}$. The k-step ahead forecast is an estimate of the random variable X_{n+k} given by

$$\widehat{X}_{n+k} = \mathrm{E}[X_{n+k} | X_1 = x_1, \ldots, X_n = x_n].$$

Here $\mathrm{E}[X|Y = y]$ denotes the conditional expectation of X given that $Y = y$

Example : AR(1)

- Consider AR(1) with $|a_1| < 1$, i.e. the process is **stationary**:

$$X_k = a_1 X_{k-1} + W_k.$$

Example : AR(1)

- Consider AR(1) with $|a_1| < 1$, i.e. the process is **stationary**:

$$X_k = a_1 X_{k-1} + W_k.$$

- If data $\{x_1, x_2, \ldots, x_n\}$ is given, then we estimate the value of X_{n+1}:

$$\widehat{X}_{n+1} = \mathrm{E}[X_{n+1} | X_1 = x_1, \ldots, X_n = x_n] = a_1 x_n.$$

Example : AR(1)

- Consider AR(1) with $|a_1| < 1$, i.e. the process is **stationary**:

$$X_k = a_1 X_{k-1} + W_k.$$

- If data $\{x_1, x_2, \ldots, x_n\}$ is given, then we estimate the value of X_{n+1}:

$$\widehat{X}_{n+1} = \mathrm{E}[X_{n+1} | X_1 = x_1, \ldots, X_n = x_n] = a_1 x_n.$$

- For $k > 1$ we plug-in the model equation several times and obtain

$$\begin{aligned}
\widehat{X}_{n+k} &= \mathrm{E}[X_{n+k} | X_1 = x_1, \ldots, X_n = x_n] \\
&= \mathrm{E}[a_1 X_{n+k-1} + W_{n+k} | X_1 = x_1, \ldots, X_n = x_n] \\
&= a_1 \mathrm{E}[X_{n+k-1} | X_1 = x_1, \ldots, X_n = x_n] \\
&= \cdots = a_1^k x_n.
\end{aligned}$$

- k-step ahead forecast of an AR(1) process is **exponentially decaying to zero** from last observation x_n and not dependent on earlier observations

Confidence Intervals

- The standard error σ_k is the square root of the conditional variance

$$\sigma_k^2 = \operatorname{Var}(X_{n+k} | X_1 = x_1, \ldots, X_n = x_n).$$

Confidence Intervals

- The standard error σ_k is the square root of the conditional variance

$$\sigma_k^2 = \mathrm{Var}(X_{n+k}|X_1 = x_1, \ldots, X_n = x_n).$$

- This quantity is increasing with k and converges to the process variance σ_X^2

Confidence Intervals

- The standard error σ_k is the square root of the conditional variance

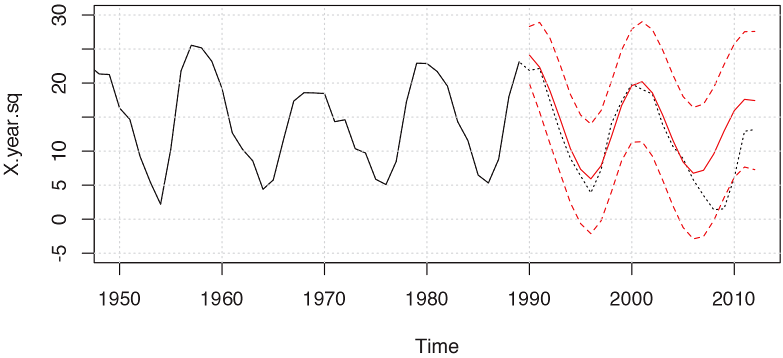
$$\sigma_k^2 = \mathrm{Var}(X_{n+k}|X_1 = x_1, \ldots, X_n = x_n).$$

- This quantity is increasing with k and converges to the process variance σ_X^2

- With this standard error a 95% **confidence interval** for the conditional expectation $\mathrm{E}[X_{n+k}|X_1 = x_1, \ldots, X_n = x_n]$ can be computed

$$\widehat{X}_{n+k} \pm 1.96\sigma_k.$$

Example : Sunspot Number Prediction with Confidence Intervals



Please check example 1.11 of the Forecasting chapter and solve exercises 3