Characteristics of the ARMA models

MS-C2128 Prediction and Time Series Analysis

Fall term 2020

Week 4: Characteristics of the ARMA models

1. Characteristics of the ARMA models
   1. Statistical properties of the stationary ARMA models
   2. ARIMA and SARIMA models
2. Fitting an ARMA model
   1. Estimation
   2. Box-Jenkins method
   3. Decomposition of time series

Content

Characteristics of the MA(q) processes

$$x_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}, \quad (\epsilon_t)_{t \in T} \sim WN(0, \sigma^2)$$

- The expected value

$$\mu_x = \mathrm{E}[x_t] = 0.$$

- The variance

$$\sigma_x^2 = \mathrm{var}(x_t) = \sigma^2 \sum_{i=0}^{q} \theta_i^2, \quad \theta_0 = 1.$$

- The autocovariance

$$\gamma_k = \mathrm{cov}\left(x_t, x_{t-k}\right) = \begin{cases} \sigma^2 \sum_{i=0}^{q-k} \theta_i \theta_{i+k}, & k = 0, 1, 2, \ldots, q \\ 0, \quad k > q. \end{cases}$$

Characteristics of the MA(q) processes

$$x_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}, \quad (\epsilon_t)_{t \in T} \sim WN(0, \sigma^2)$$

- The autocorrelation

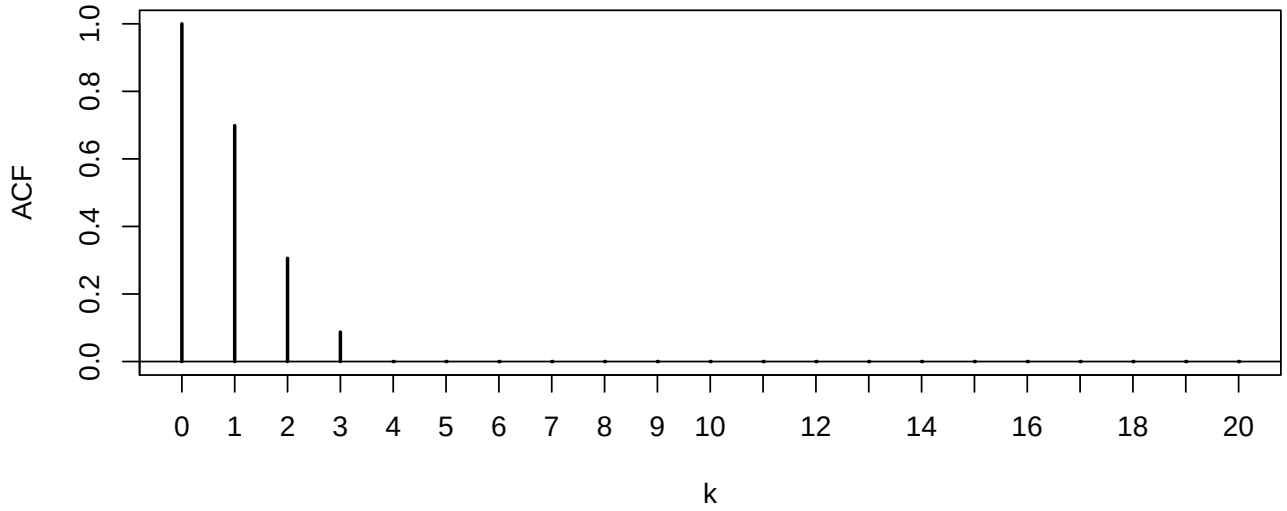
$$\rho_k = \begin{cases} 1, & k = 0 \\ \frac{\sum_{i=0}^{q-k} \theta_i \theta_{i+k}}{\sum_{i=0}^{q} \theta_i^2}, & k = 1, 2, \ldots, q \\ 0, & k > q \end{cases}$$

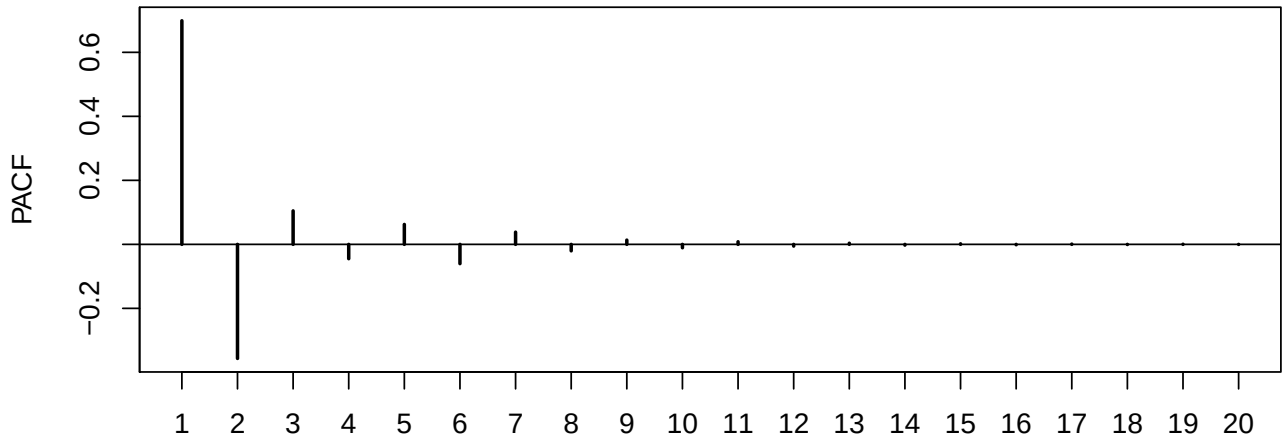
- The AR(∞) representation (if invertible)

$$\sum_{i=0}^{\infty} \pi_i x_{t-i} = \epsilon_t \quad (\pi_0 = 1)$$

- The partial autocorrelation decays exponentially.

MA(3) process, $\theta_1 = 1$, $\theta_2 = 0.5$, $\theta_3 = 0.2$





Characteristics of an invertible MA(1) process

$$x_t = \epsilon_t + \theta_1 \epsilon_{t-1}, \quad (\epsilon_t)_{t \in T} \sim WN(0, \sigma^2)$$

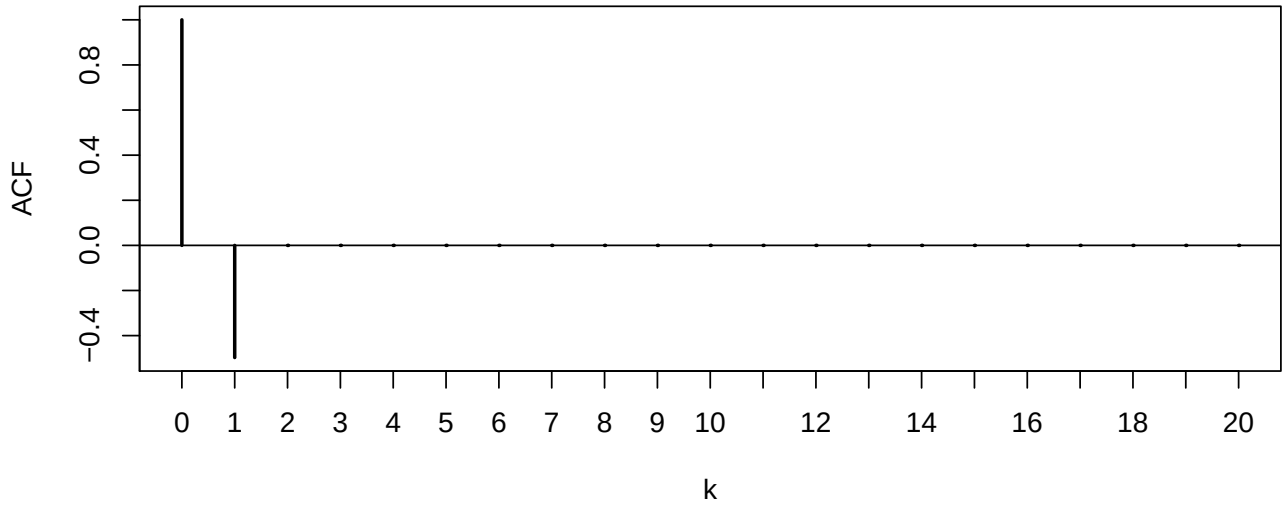
- Since the root of the lag polynomial $\theta(L) = 1 + \theta_1 L$ is outside of the unit circle, we have $|\theta_1| < 1$.
- AR(∞) representation:

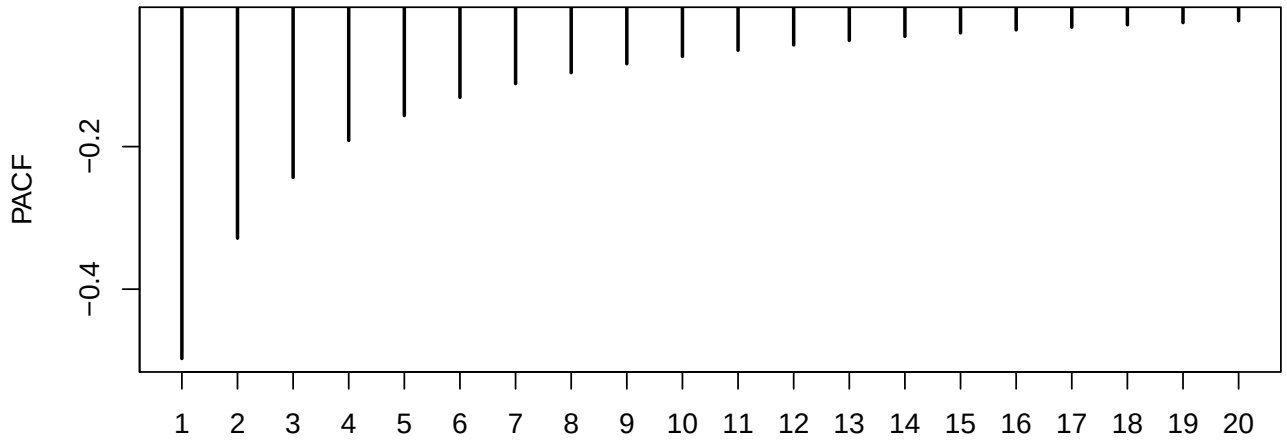
$$\sum_{i=0}^{\infty} (-\theta_1)^i x_{t-i} = \epsilon_t$$

- The autocovariance and the autocorrelation

$$\gamma_k = \begin{cases} \sigma^2(1 + \theta_1^2), & k = 0 \\ \sigma^2 \theta_1, & k = 1 \\ 0, & k > 1 \end{cases}, \qquad \rho_k = \begin{cases} 1, & k = 0 \\ \frac{\theta_1}{1+\theta_1^2}, & k = 1 \\ 0 & k > 1. \end{cases}$$

MA(1) process, $\theta_1 = -0.9$





Characteristics of the stationary AR(p) models

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots \phi_p x_{t-p} + \epsilon_t, \quad (\epsilon_t)_{t\in T} \sim WN(0, \sigma^2).$$

- The MA(∞) representation

$$x_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} \quad (\psi_0 = 1).$$

- The expected value $\mu_x = \mathrm{E}[x_t] = 0.$
- The variance

$$\sigma_x^2 = \mathrm{var}(x_t) = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2.$$

- The autocovariance and the autocorrelation

$$\gamma_k = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}, \qquad \rho_k = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+k}}{\sum_{i=0}^{\infty} \psi_i^2}.$$

Stationary AR(p) models

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots \phi_p x_{t-p} + \epsilon_t, \quad (\epsilon_t)_{t\in T} \sim WN(0, \sigma^2).$$

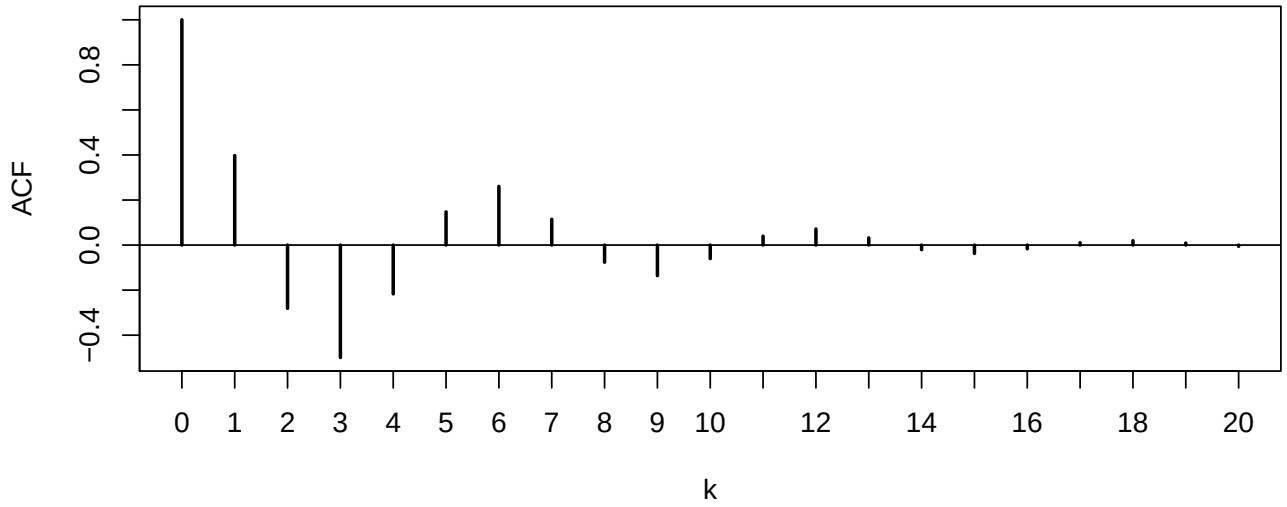
- The autocorrelations satisfy

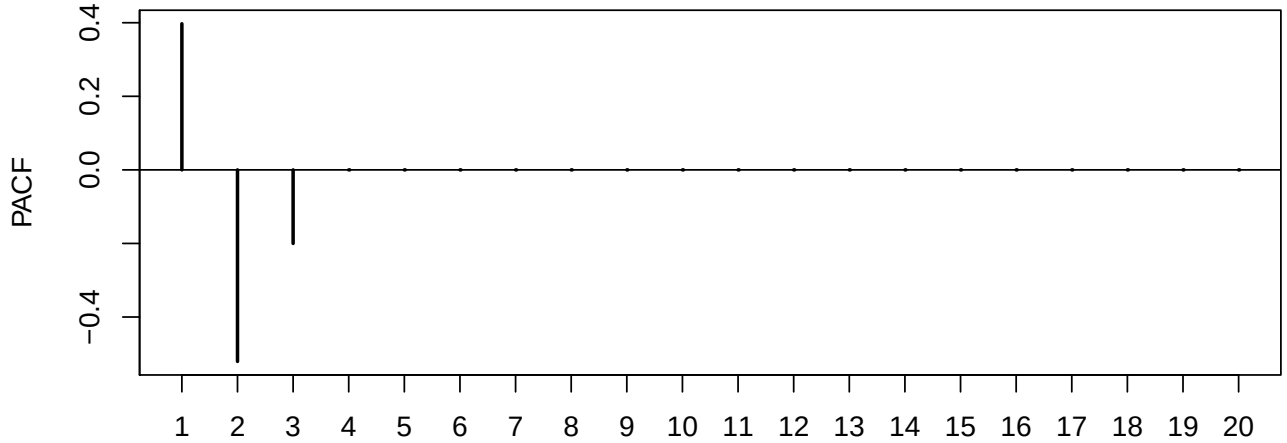
$$\rho_0 = 1$$
$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \ldots + \phi_p \rho_{k-p}, \; k > 0,$$

as

$$\gamma_k = \mathrm{E}[x_t x_{t-k}] = \mathrm{E}\left[x_{t-k}\left(\sum_{i=1}^{p} \phi_i x_{t-i} + \epsilon_t\right)\right]$$
$$= \sum_{i=1}^{p} \phi_i \, \mathrm{E}[x_{t-k} x_{t-i}] + \mathrm{E}[x_{t-k}\epsilon_t] = \sum_{i=1}^{p} \phi_i \gamma_{k-i}.$$

AR(3) process, $\phi_1 = 0.5$, $\phi_2 = -0.4$, $\phi_3 = -0.2$





Stationary AR(1) models: Characteristics

$$x_t = \phi_1 x_{t-1} + \epsilon_t, \quad (\epsilon_t)_{t\in T} \sim WN(0, \sigma^2).$$

- The MA(∞) representation

$$x_t = \sum_{i=0}^{\infty} \phi_1^i \epsilon_{t-i}$$

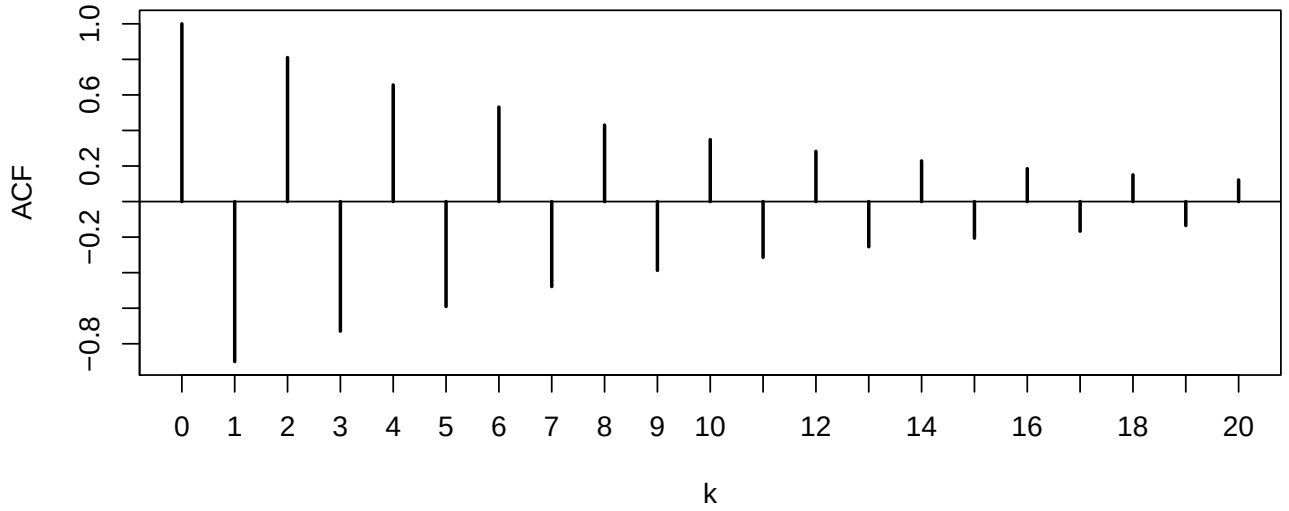
- The expected value $\mu_x = \mathrm{E}[x_t] = 0.$
- The variance

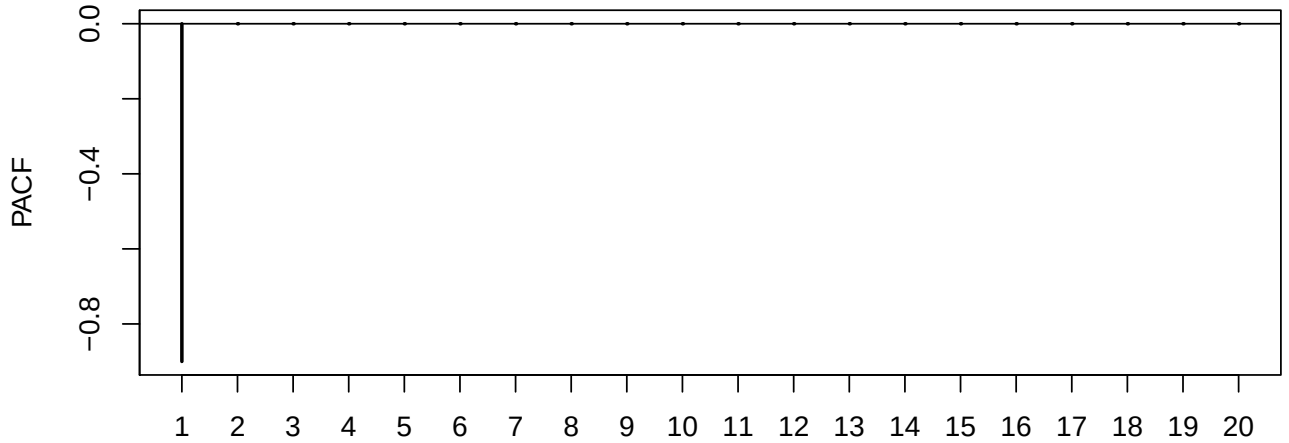
$$\sigma_x^2 = \mathrm{var}(x_t) = \sigma^2 \sum_{i=0}^{\infty} \phi_1^{2i} = \frac{\sigma^2}{1-\phi_1^2}.$$

- The autocovariance and the autocorrelation

$$\gamma_k = \sigma^2 \sum_{i=0}^{\infty} \phi_1^i \phi_1^{i+k} = \phi_1^k \sigma_x^2, \qquad \rho_k = \phi_1^k.$$

AR(1) process, $\phi_1 = -0.9$





Stationary AR(2) models: Characteristics

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \epsilon_t, \quad (\epsilon_t)_{t\in T} \sim WN(0, \sigma^2).$$

- The MA(∞) representation

$$x_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i},$$

$$\psi_0 = 1, \quad \psi_1 - \phi_1 = 0, \quad \psi_i - \phi_1 \psi_{i-1} - \phi_2 \psi_{i-2} = 0, \ i \geq 2.$$

Stationary AR(2) models: Characteristics

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \epsilon_t, \quad (\epsilon_t)_{t \in T} \sim WN(0, \sigma^2).$$

- The expected value $\mu_x = \mathrm{E}[x_t] = 0.$
- The variance

$$\sigma_x^2 = \mathrm{var}(x_t) = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2.$$

- The autocovariance and the autocorrelation

$$\gamma_k = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}, \quad \rho_1 = \frac{\phi_1}{1 - \phi_2}, \quad \rho_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2.$$

Stationary AR(2) models: Characteristics

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \epsilon_t, \quad (\epsilon_t)_{t \in T} \sim WN(0, \sigma^2).$$

- Since the roots of the lag polynomial

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2$$

are outside of the unit circle, we have that

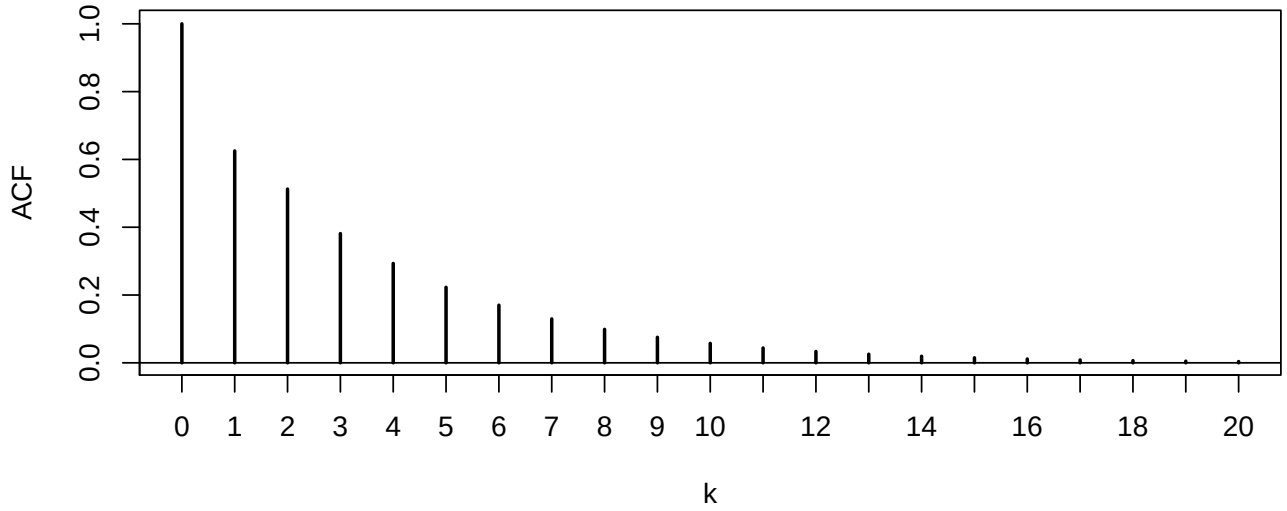
$$\begin{cases} \phi_1 + \phi_2 & < 1, \\ -\phi_1 + \phi_2 & < 1, \\ |\phi_2| & < 1. \end{cases}$$

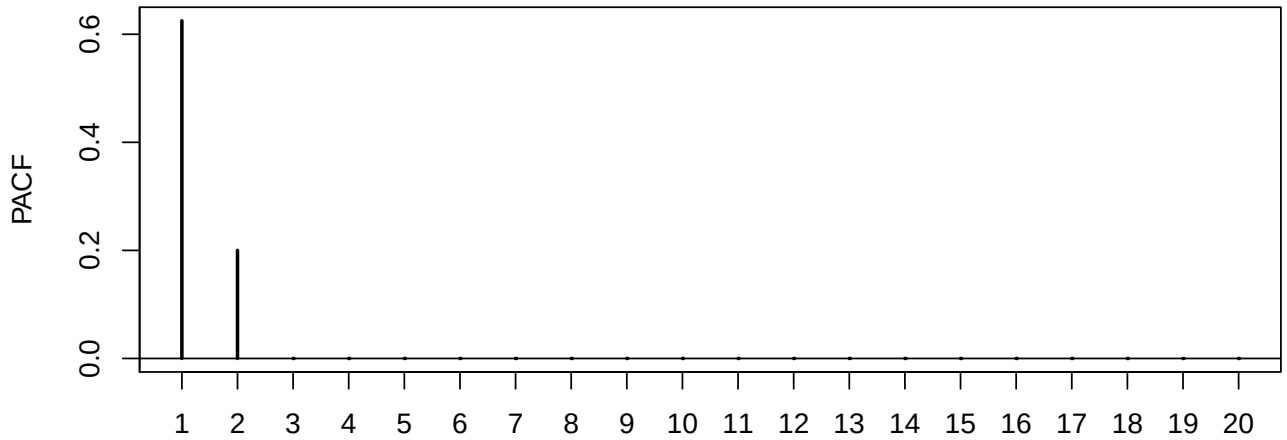
Stationary AR(2) models: Characteristics

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \epsilon_t, \quad (\epsilon_t)_{t \in T} \sim WN(0, \sigma^2).$$

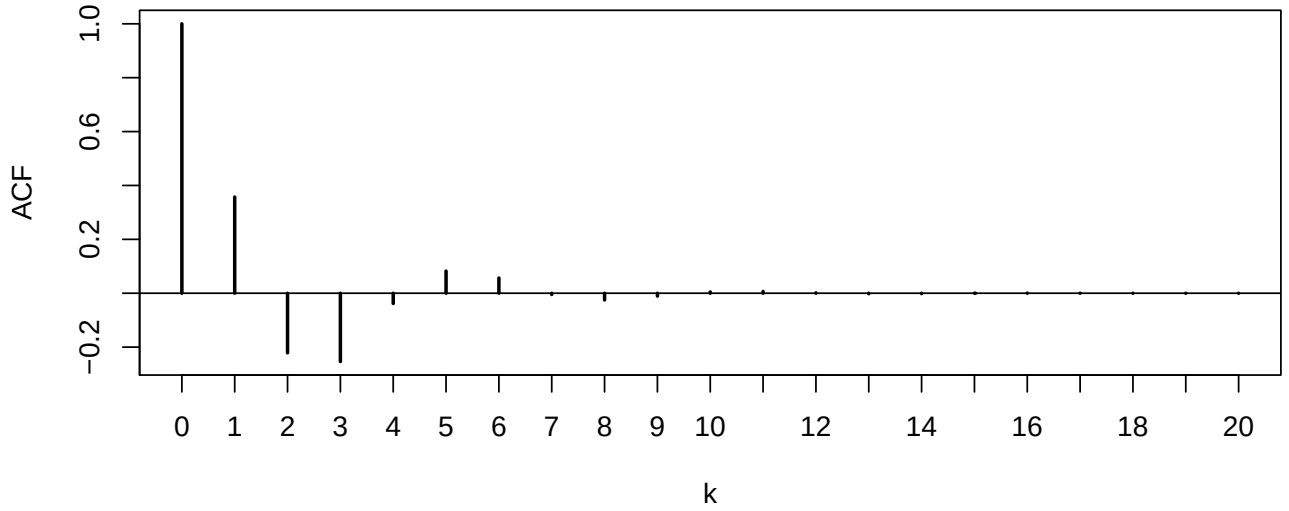
- The roots of the lag polynomial are complex valued if $\phi_1^2 + 4\phi_2 < 0$.
  - If the roots are complex valued, then the autocorrelation function is bounded by an exponentially decaying sine-function.
  - If the roots are real valued, then the autocorrelation function is bounded by an exponential function (or exponential functions).

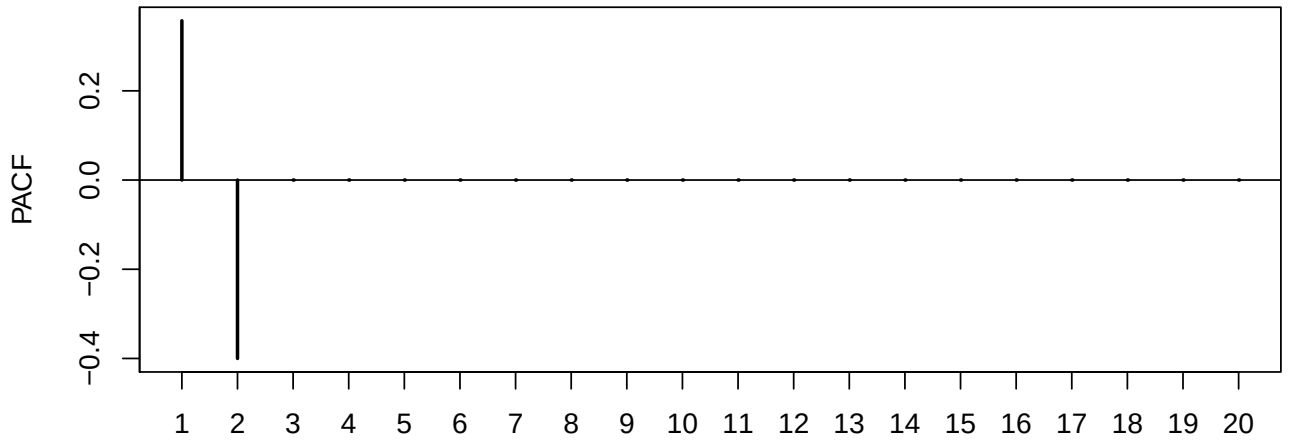
AR(2) process, $\phi_1 = 0.5$, $\phi_2 = 0.2$





AR(2) process, $\phi_1 = 0.5$, $\phi_2 = -0.4$





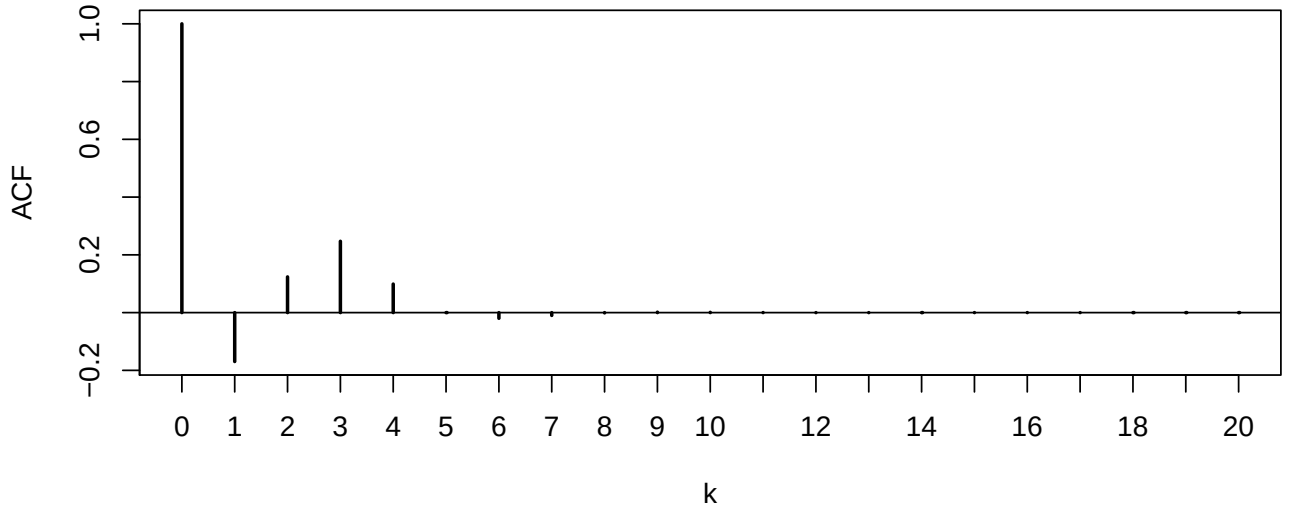
Stationary and invertible ARMA(p, q) processes

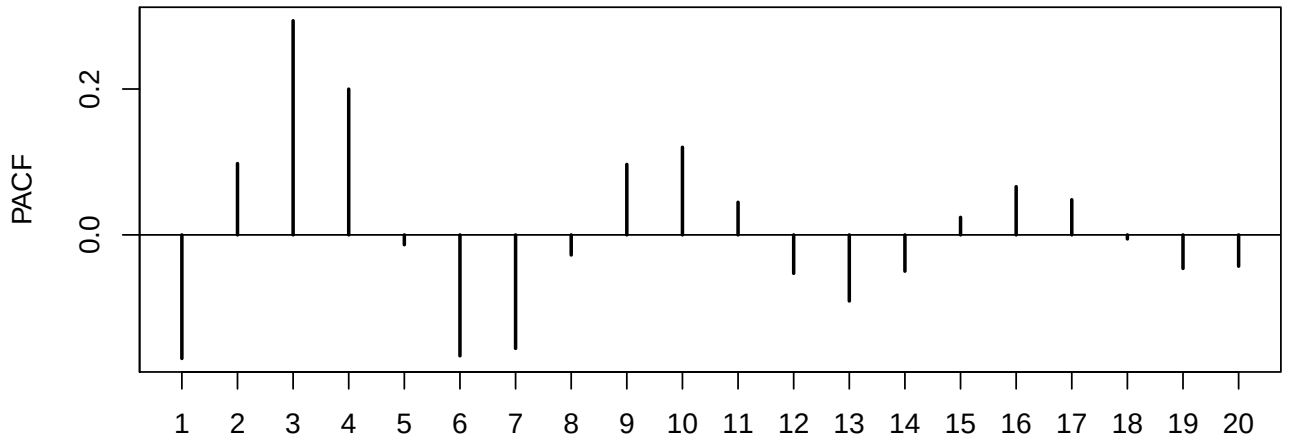
$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \ldots - \phi_p x_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q},$$

where $(\epsilon_t)_{t \in T} \sim WN(0, \sigma^2)$.

- For a stationary AR(p) process
  - The autocorrelation function decays exponentially (geometric series).
  - The partial autocorrelation function is equal to 0 after p.
- For an MA(q) process
  - The autocorrelation function is equal to 0 after q.
  - The partial autocorrelation function decays exponentially.
- The autocorrelation function and the partial autocorrelation of a stationary ARMA(p, q) process decay exponentially.

ARMA(2,3), $\phi = (0.5, -0.2)$, $\theta = (-0.8, 0.6, 0.2)$





Stationary ARMA(p, q) processes

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \ldots - \phi_p x_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q},$$

where $(\epsilon_t)_{t \in T} \sim WN(0, \sigma^2)$.

Model	**Auto-correlation function**	**Partial auto-correlation function**
AR(p)	Decays exponentially	Is equal to 0 after p
MA(q)	Is equal to 0 after q	Decays exponentially
ARMA(p, q)	Decays exponentially	Decays exponentially

Stationary and invertible ARMA(1,1) models

$$x_t - \phi_1 x_{t-1} = \epsilon_t + \theta_1 \epsilon_{t-1}, \quad (\epsilon_t)_{t \in T} \sim WN(0, \sigma^2)$$

- The roots of the lag polynomials

$$\phi(L) = 1 - \phi_1 L, \qquad \theta(L) = 1 + \theta_1 L$$

  are outside of the unit circle, if $|\phi_1| < 1$, $|\theta_1| < 1$.
- The MA(∞) representation

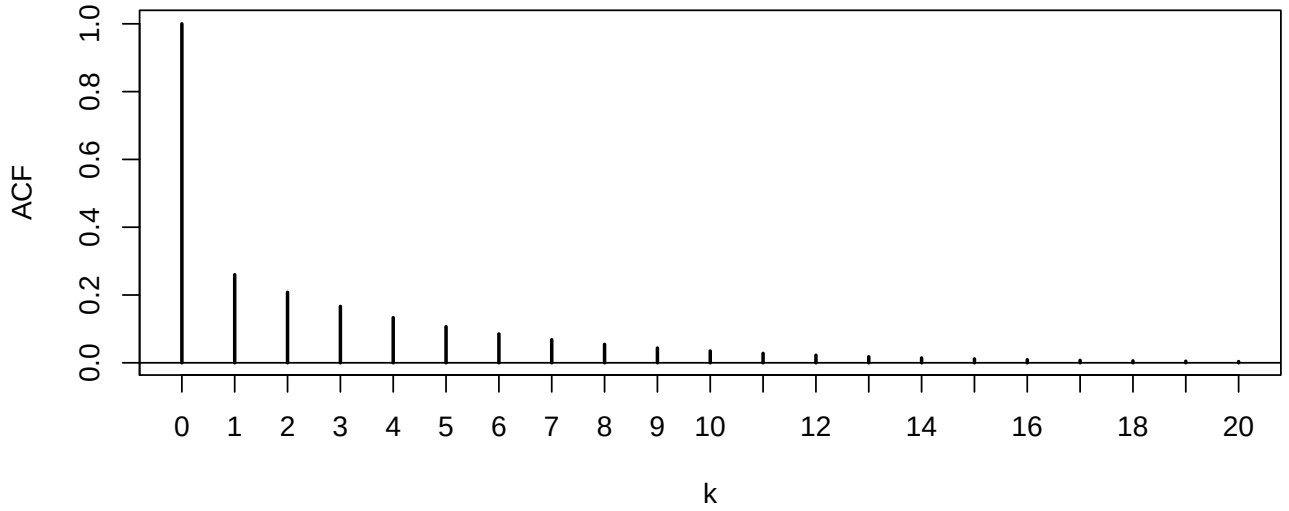
$$x_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i},$$

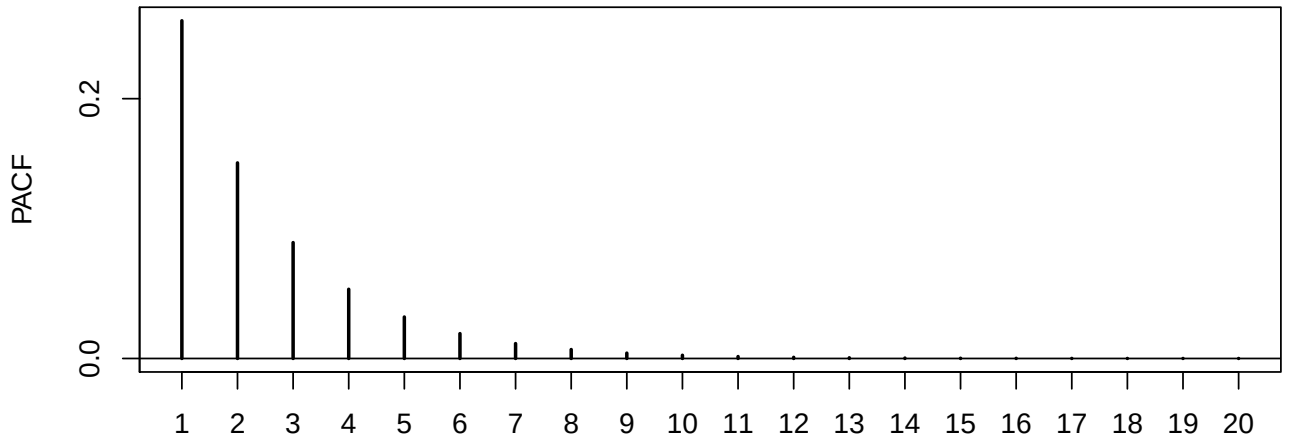
$$\psi_0 = 1, \quad \psi_i = \theta_1 \phi_1^{i-1} + \phi_1^i, \ i > 0.$$

- The expected value, the variance and the autocovariance:

$$\mu_x = \mathrm{E}[x_t] = 0 \quad \sigma_x^2 = \mathrm{var}(x_t) = \sigma^2 \sum_{i=0}^{\infty} \psi_i^2 \quad \gamma_k = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}$$

ARMA(1,1), $\phi = 0.8$, $\theta = -0.6$





Stationary SARMA$(P, Q)_s$ processes

$$x_t - \Phi_1 x_{t-s} - \ldots - \Phi_P x_{t-Ps} = \epsilon_t + \Theta_1 \epsilon_{t-s} + \ldots + \Theta_Q \epsilon_{t-Qs},\ (\epsilon_t)_{t \in T} \sim WN(0, \sigma^2)$$

- At time points $s, 2s, 3s, \ldots$, the autocorrelation function and the partial autocorrelation function of a SARMA$(P, Q)_s$ process behave as the autocorrelation function and the partial autocorrelation function of the corresponding ARMA(p, q)process. At other time other points, the autocorrelation function and the partial autocorrelation function of a SARMA$(P, Q)_s$ process are equal to 0.

Stationary SARMA$(P, Q)_s$ processes

For a stationary SAR$(P)_s$ process

- The values of the autocorrelation function at time points $s, 2s, 3s, ...$ decay exponentially. At other time other points, the autocorrelation function is equal to 0.
- The partial autocorrelation function is non-zero at time points $s, 2s, 3s, ..., Ps$. At other time points, and especially after the time point Ps, it is equal to 0.

Stationary SARMA$(P, Q)_s$ processes

For a stationary SMA$(Q)_s$ process

- The autocorrelation function is non-zero at time points $s, 2s, 3s, ..., Qs$. At other time points, and especially after the time point Qs, it is equal to 0.
- The values of the partial autocorrelation function at time points $s, 2s, 3s, ...$ decay exponentially. At other time other points, the partial autocorrelation function is equal to 0.

Stationary SARMA$(P, Q)_s$ processes

For a stationary SARMA$(P, Q)_s$ process, the values of the autocorrelation function and the values of the partial autocorrelation function at time points $s, 2s, 3s, \ldots$ decay exponentially. At other time points, the functions are equal to 0.

Stationary SARMA$(P, Q)_s$ processes

$$x_t - \Phi_1 x_{t-s} - \ldots - \Phi_P x_{t-Ps} = \epsilon_t + \Theta_1 \epsilon_{t-s} + \ldots + \Theta_Q \epsilon_{t-Qs},\ (\epsilon_t)_{t \in T} \sim WN(0, \sigma^2)$$

Model	**Autocorrelation function at** $s, 2s, 3s, \ldots$	**Partial autocorrelation function at** $s, 2s, 3s, \ldots$
SAR$(P)_s$	Decays exponentially	Is equal to 0 after Ps
SMA$(Q)_s$	Is equal to 0 after Qs	Decays exponentially
SARMA$(P, Q)_s$	Decays exponentially	Decays exponentially

Stationary SARMA$(p, q)(P, Q)_s$ models

$$x_t - \Phi_1 x_{t-s} - \ldots - \Phi_P x_{t-Ps} = \epsilon_t + \Theta_1 \epsilon_{t-s} + \ldots + \Theta_Q \epsilon_{t-Qs},\ (\epsilon_t)_{t \in T} \sim WN(0, \sigma^2)$$

- The behavior of the autocorrelation function and the partial autocorrelation function of a stationary SARMA$(p, q)(P, Q)_s$ process is a (complicated) combination of the behaviours of the correlation functions of the corresponding ARMA(p, q) and SARMA$(P, Q)_s$ processes.

Filtering

- A stochastic process x_t is said to be obtained by filtering from a process y_t using linear time invariant filter, if

$$x_t = \sum_{j=-\infty}^{\infty} w_j y_{t-j}$$

  - The filter is defined by the weights w_j, $\sum_{j=-\infty}^{\infty} |w_j| < \infty$.
- One can show that the spectral density function of the filtered process x_t can be given by

$$f_x(\lambda) = \left|W(\lambda)\right|^2 f_y(\lambda),$$

where $f_y(\lambda)$ is the spectral density function of y_t and where

$$W(\lambda) = \sum_{j=-\infty}^{\infty} w_j e^{-i\lambda j}.$$

  - $\left|W(\lambda)\right|^2$ is the so called transition function of the filter.

ARMA(p, q): Spectrum

A stationary ARMA(p, q) process x_t has an MA(∞) representation

$$x_t = \Psi(L)\epsilon_t, \qquad (\epsilon_t)_{t\in T} \sim WN(0, \sigma^2),$$
$$\Psi(L) = \sum_{j=0}^{\infty} \psi_j L^j, \quad \psi_0 = 1, \qquad \phi(L)\Psi(L) = \theta(L).$$

- The process x_t is hence obtained by filtering from a pure stochastic process ϵ_t using a filter with the transition function

$$\left|\Psi(e^{-i\lambda})\right| = \frac{\left|\theta(e^{-i\lambda})\right|}{\left|\phi(e^{-i\lambda})\right|} = \frac{|1 + \theta_1 e^{-i\lambda} + \ldots + \theta_q e^{-qi\lambda}|}{|1 + \phi_1 e^{-i\lambda} + \ldots + \phi_p e^{-pi\lambda}|}.$$

ARMA(p, q): Spectrum

- Thus, the spectral density function of the process x_t is

$$f_x(\lambda) = \left|W(\lambda)\right|^2 f_\epsilon(\lambda) = \frac{\sigma^2}{2\pi} \frac{\left|1 + \theta_1 e^{-i\lambda} + \ldots + \theta_q e^{-qi\lambda}\right|^2}{\left|1 - \phi_1 e^{-i\lambda} - \ldots - \phi_p e^{-pi\lambda}\right|^2}.$$

  (The spectral density function of the process ϵ_t is the constant function $\sigma^2/2\pi$.)

Spectrums of selected (stationary) processes

$$\text{ARMA(1,1)}: \quad f(\lambda) = \frac{\sigma^2}{2\pi}\frac{|1+\theta_1 e^{-i\lambda}|^2}{|1+\phi_1 e^{-i\lambda}|^2} = \frac{\sigma^2}{2\pi}\frac{1+\theta_1^2+2\theta_1\cos(\lambda)}{1+\phi_1^2-2\phi_1\cos(\lambda)}$$

$$\text{AR}(p): \quad f(\lambda) = \frac{\sigma^2}{2\pi}\frac{1}{|1-\phi_1 e^{-i\lambda}-\ldots-\phi_p e^{-pi\lambda}|^2}$$

$$\text{AR(2)}: \quad f(\lambda) = \frac{\sigma^2}{2\pi}\frac{1}{|1-\phi_1 e^{-i\lambda}-\phi_2 e^{-2i\lambda}|^2}$$

$$= \frac{\sigma^2}{2\pi}\frac{1}{1+\phi_1^2+\phi_2^2-2\phi_1(1-\phi_2)\cos(\lambda)-2\phi_2\cos(2\lambda)}$$

$$\text{MA}(q): \quad f(\lambda) = \frac{\sigma^2}{2\pi}|1+\theta_1 e^{-i\lambda}+\ldots+\theta_q e^{-qi\lambda}|^2$$

$$\text{MA(2)}: \quad f(\lambda) = \frac{\sigma^2}{2\pi}|1+\theta_1 e^{-i\lambda}+\theta_2 e^{-qi\lambda}|^2$$

$$= \frac{\sigma^2}{2\pi}\left(1+\theta_1^2+\theta_2^2+2\theta_1(1-\theta_2)\cos(\lambda)+2\theta_2\cos(2\lambda)\right)$$

SARIMA$(p, h, q)(P, H, Q)_s$

Let x_t be a stochastic process such that

(i) x_t is non-stationary.

(ii) $D_s^G D^g x_t$ is non-stationary, when $g < h$, $G < H$.

(iii) $y_t = D_s^H D^h x_t$ is stationary.

(iv) y_t is a SARMA$(p, q)(P, Q)_s$ process.

- Then we say that the process x_t is integrable of order h and seasonal integrable of order H. The process x_t is called a SARIMA$(p, h, q)(P, H, Q)_s$ process.
- If x_t is a SARIMA$(p, h, q)(P, H, Q)_s$ process, then $y_t = D_s^H D^h x_t$ is a SARMA$(p, q)(P, Q)_s$ process.
- If x_t is an ARIMA(p, h, q) process, then $y_t = D^h x_t$ is an ARMA(p, q) process.

Content

Correlation functions, spectrum and stationarity

- The theoretical autocorrelation and partial autocorrelation functions and the spectral density function are defined only for stationary processes.
- The sample correlation functions and the sample spectral density function can be calculated also when the observed series is not stationary.
  - When the observed series is not stationary, the sample functions should not be interpreted as estimators of the theoretical functions.
  - When the observed series is not stationary, the sample correlation functions and the sample spectral density function may give hints to how one should proceed in order to stationarize the series.

Estimating autocorrelation

- Let x_t, $t = 1, 2, ..., n$ be an observed time series.
- The (arithmetic) sample mean:

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

- The sample variance:

$$c_0 = \frac{1}{n} \sum_{t=1}^{n} (x_t - \bar{x})^2.$$

Estimating autocorrelation

- The k. sample autocovariance:

$$c_k = \frac{1}{n} \sum_{t=k+1}^{n} (x_t - \bar{x})(x_{t-k} - \bar{x}), \quad k = 0, 1, ..., n-1.$$

- The k. sample autocorrelation coefficient:

$$r_k = \frac{c_k}{c_0}, \quad k = 0, 1, 2, ..., n-1.$$

These sample quantities estimate the corresponding parameters.

How many autocorrelations should one estimate?

- Consider an observed series x_t , $t = 1, 2, ..., n$. Now, technically, one can estimate the $n - 1$ first autocovariances c_k and the $n - 1$ first autocorrelations r_k.
- However, one should note that the k. autocovariance

$$c_k = \frac{1}{n} \sum_{t=k+1}^{n} (x_t - \bar{x})(x_{t-k} - \bar{x}), \quad k = 0, ..., n - 1,$$

  is estimated from $n - k$ observations. For large k, the estimates c_k ja r_k are based on few observations only.
- The sample autocovariances and the sample autocorrelations are not reliable if the sample size n is small and k is large. Often, if $n < 50$ and $k > \frac{n}{4}$, the estimates are thought to be unreliable.

Partial autocorrelations coefficient: Estimation

- One can estimate the partial autocorrelations coefficient as follows:
  1. Construct k Yule-Walker equations from the sample

$$\begin{bmatrix} 1 & r_1 & r_2 & \cdots & r_{k-1} \\ r_1 & 1 & r_1 & \cdots & r_{k-2} \\ r_2 & r_1 & 1 & \cdots & r_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{k-1} & r_{k-2} & r_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} a_{k1} \\ a_{k2} \\ a_{k3} \\ \vdots \\ a_{kk} \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \vdots \\ r_k \end{bmatrix},$$

  2. Solve a_{kk} from the equations.
  3. Now $\hat{\phi}_k = a_{kk}$ is an estimate for the k. partial autocorrelations coefficient.
     - Examples: $\hat{\phi}_1 = a_{11}$, $\hat{\phi}_2 = a_{22} = \frac{r_2 - r_1^2}{1 - r_1^2}$.
- The sample partial autocorrelations coefficients $\hat{\phi}_k$ define the sample partial autocorrelation function $\hat{\phi} : \{0, 1, ..., n-1\} \to \mathbb{R}$,

$$\hat{\phi}(k) = \hat{\phi}_k \quad \text{for all } k = 0, 1, ..., n-1.$$

Estimating partial autocorrelations coefficient: AR(p) processes

- Partial autocorrelations coefficient can *alternatively* be estimated from the regression models

$$x_t = \beta_1 x_{t-1} + \beta_2 x_{t-2} + ... \beta_p x_{t-p} + \epsilon_t.$$

  using least squares estimates.
  - Then the estimator for the k. partial autocorrelations coefficient ϕ_k is the least squares estimator b_k for the parameter β_k. That is $\hat{\phi}_k = b_k$, $k = 1, 2, ..., p$.
- This approach is straightforwardly applicable only in the case of AR(p) processes. (If the series contains an MA part, then the noise is not uncorrelated.)

Sample autocovariances

Remark

The k. sample autocovariance c_k is a biased estimator for the autocovariance γ_k . However, c_k is asymptotically unbiased:

$$\lim_{n\to\infty} \mathrm{E}[c_k] = \gamma_k.$$

Autocorrelations: Significance testing

- Consider a stationary stochastic process formed of independent and identically distributed random variables. Then the k. sample autocorrelation r_k of the process is asymptotically normally distributed:

$$r_k \sim_a N\left(0, \frac{1}{n}\right)$$

- This motivates approximative significance testing:
  - Under the null hypothesis ($H_0 : \rho_k = 0$) the estimate r_k lies in the interval

$$\left[-\frac{2}{\sqrt{n}}, \frac{2}{\sqrt{n}}\right] \qquad (2 \approx 1.96).$$

    approximately with probability 95 %.

Remark

If one estimates the 100 first autocorrelations from a time series generated from *IID* random variables, then approximately 5 of the aurocorrelations lie outside of the given interval.

ARMA model: Parameter estimation

- Let x_t, $t = 1, ..., n$ be an observed time series. We now consider fitting an ARMA(p, q) model

$$x_t - \phi_1 x_{t-1} - ... - \phi_p x_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q},$$

  where $(\epsilon_t)_{t \in T} \sim IID(0, \sigma^2)$ and $\epsilon_t \sim N(0, \sigma^2)$ for all $t \in T$.
  - Now, the joint distribution of $x_1, ..., x_n$ is an n-dimensional normal distribution, whose covariance matrix depends, in a non-linear manner, of the parameters of the ARMA(p, q) model.

ARMA model: Parameter estimation

- Using that, one can feed the data $x_1, ..., x_n$ to the corresponding likelihood function and maximize it with respect to the model parameters (R: `arima()`). Maximum likelihood estimators for the parameters of the ARMA(p, q) model are then obtained:

$$\hat{\phi}_1, \hat{\phi}_2, ..., \hat{\phi}_p, \hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_q, \hat{\sigma}^2$$

- It is not possible to obtain closed form solutions for these estimators. For details, see for example Hamilton (1994) or Brockwell & Davis (1991).

ARMA model: Parameter estimation

Assume that we have obtained the following maximum likelihood estimates for the ARMA(p, q) model:

$$\hat{\phi}_1, \hat{\phi}_2, ..., \hat{\phi}_p, \hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_q, \hat{\sigma}^2.$$

- Standard deviations of the maximum likelihood estimators can be obtained using Fisher information.[1]

[1] The second moment of the derivative of the log-likelihood function with respect to each parameter

ARMA model: Parameter estimation

- The maximum likelihood estimators are asymptotically normal, allowing us to construct parametric confidence intervals and conduct significance testing.
  - Normal distribution or t-distribution based confidence intervals can be obtained.
  - Significance can be tested applying the t-test.
- The residuals are given as:

$$e_t = \frac{\hat{\phi}(L)}{\hat{\theta}(L)} x_t,$$

$$\hat{\phi}(L) = 1 - \hat{\phi}_1 L - \ldots - \hat{\phi}_p L^p, \quad \hat{\theta}(L) \quad = 1 + \hat{\theta}_1 L + \ldots + \hat{\theta}_q L^q.$$

ARMA model: Parameter estimation

Maximum likelihood estimators constructed under normality assumption may be applied also when the ϵ_t are not normally distributed. The estimator, however, is not a maximum likelihood estimator anymore. Significance testing and confidence intervals for the model parameters can then be constructed by applying, for example, interval bootstrapping.

ARMA model: Bootstrap confidence intervals

Consider an observed ARMA series x_t , $t = 1, 2, ..., n$. Let w be the minimum time series length required for estimating the model parameters. A $(1 - \alpha)$ bootstrap confidence interval for the ARMA model parameters can be obtained as follows.

1. Select two time points s and u, $0 < s < u \leq n$, $u - s \geq w$, randomly.
2. Calculate a new parameter vector estimate from the series $x_s, x_{s+1}, ..., x_u$.
3. Repeat the previous steps $m - 1$ times.
4. For each ARMA model parameter, order the obtained m estimates from the smallest to the largest.
5. For each ARMA model parameter, set the lower end of the bootstrap confidence interval to be smaller than or equal to the $[\frac{\alpha}{2} \times m]$th ordered estimate and set the upper end of the bootstrap confidence interval to be larger than or equal to the $[(1 - \frac{\alpha}{2}) \times m]$th ordered estimate.

Bootstrap confidence interval

Some remarks:

- In time series context, bootstrapping is based on sampling intervals (that are of random length) from the original series. That ensures that the time series structure is not lost in the resampling procedure.
- Approximate significance tests for the model parameters are obtained by checking whether zero lies on the corresponding bootstrap confidence interval.

Box-Jenkins modeling

The goal is to construct a model that describes the modeled phenomena as well as possible using as few parameters as possible.

- The more parameters to estimate, the more can go wrong.
- Complicated constructions might provide an excellent fit, but forecasting then usually turns out to be difficult.

Box-Jenkins modeling

Box-Jenkins method is a three step strategy for fitting SARIMA models.

1. Model identification:
   (a) In order to stationarize the series (SARIMA $\rightarrow$ SARMA model), the orders h and H (and s) are chosen.
   Remember: h is the order of integrability and H is the order of seasonal integrability.
   (b) The orders (p, q, P, Q) of the lag polynomials of the SARMA model are chosen.
2. Model estimation:
   - Parameters θ_i, Θ_i, ϕ_i, Φ_i (as total $p + q + P + Q$ parameters) are estimated using for example maximum likelihood approach. (Compare to fitting ARMA models.)
3. Diagnostics:
   - The residuals of the estimated SARMA model are analyzed.
     - The residuals are NOT white noise. $\rightarrow$ Return to step 1.
     - The residuals are white noise. $\rightarrow$ The model is fitted.

Box-Jenkins method:

1a) Model identification – Differencing

- Differencing, seasonal differencing (and other transformations) are often applied in order to obtain stationarity.
  - The time series, its correlation function and the spectrum are often plotted and the plots are used as guides in choosing the orders of the differences.
- Differences and seasonal differences are taken until the time series appears to be stationary.
  - If, based on the plots, the time series appears to be stationary, then, naturally, one should not take differences.
  - Usually, taking the necessary differences in order to obtain stationarity, make the variance of the series smaller. Taking too many differences, on the other hand, usually tend to increase the variance of the series.

Box-Jenkins method:

1a) Model identification – Stationarization

- Difference $Dx_t = x_t - x_{t-1}$ removes deterministic linear trend.
  - Similarly, p. difference D^p removes deterministic p. degree polynomial trend.
- Seasonal difference $D_s x_t = x_t - x_{ts}$ removes deterministic seasonality that has period s.
- In addition, sometimes logarithmic transformations, $y_t = \log(x_t)$, are needed.
  - Logarithmic transformations can linearize exponential trends.
- The original time series is obtained by inverse transformation.
  - Example If $y_t = Dx_t$, then $x_1 = y_1$ ja $x_t = y_1 + y_2 + ... + y_t$, $t = 2, 3, ..., n$.
  - Example $x_t = \exp(y_t)$.

Box-Jenkins method:

1b) Model identification – Lag polynomials

- After stationarization of the time series, the order of the lag polynomials of the SARMA model are chosen.
  - The choice of the order of the lag polynomials is based on looking at the plots of the stationarized time series, its correlation functions and its spectrum.
- Choosing the orders of the lag polynomials is a challenging task. Usually one first selects a set of possible orders.
  - Selected degrees are examined by fitting the corresponding models and the final choice is based on comparing the goodness of the fits.
  - Model fits are compared by considering the significance of the estimated parameters and the diagnostics (step 3).

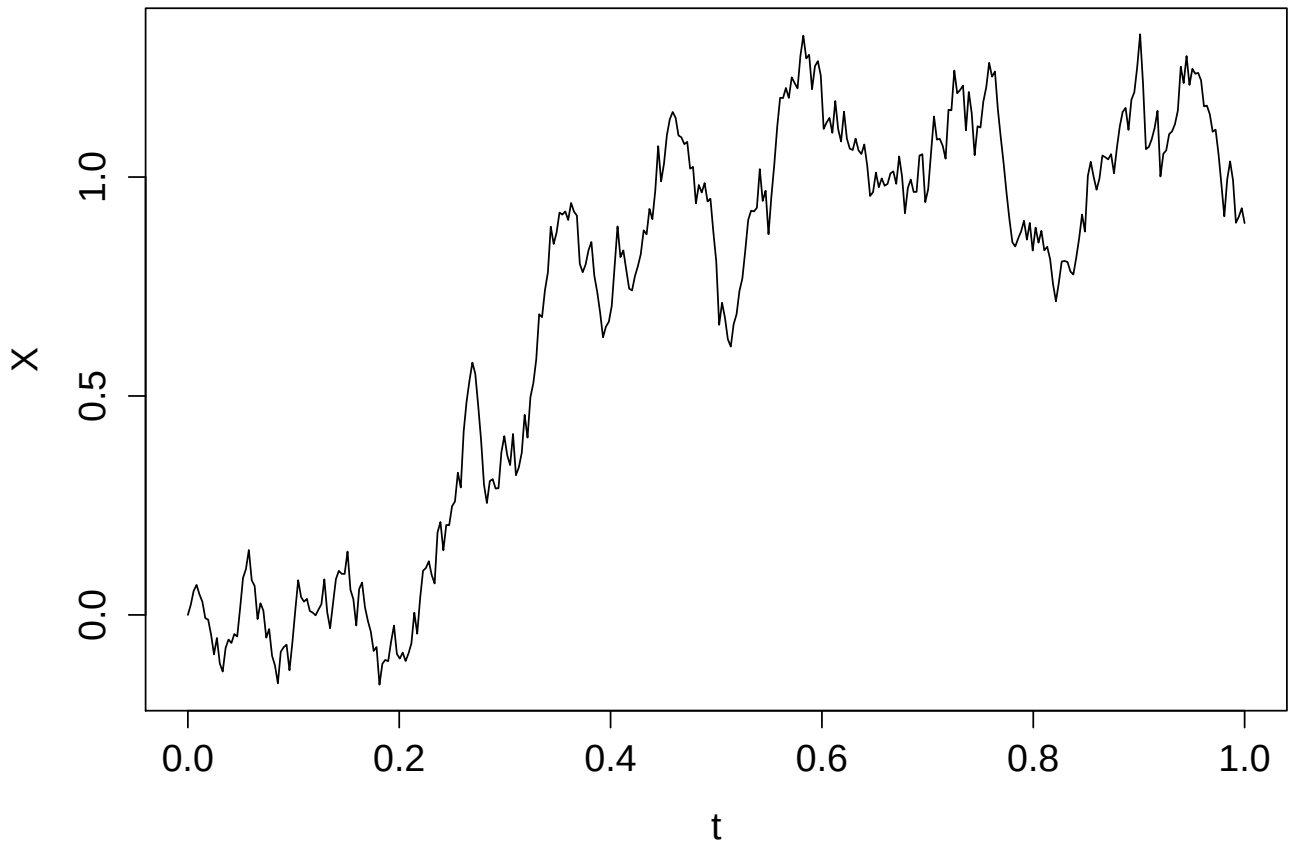
Box-Jenkins method:

1) Model identification – Comments

- When SARIMA$(p, h, q)(P, H, Q)_s$ models are fitted in practical applications, one quite rarely has to consider models where the orders of the differences or the orders of the lag polynomials are other than reasonably small integers.
- Usually (but not always) it is enough to consider the following options:

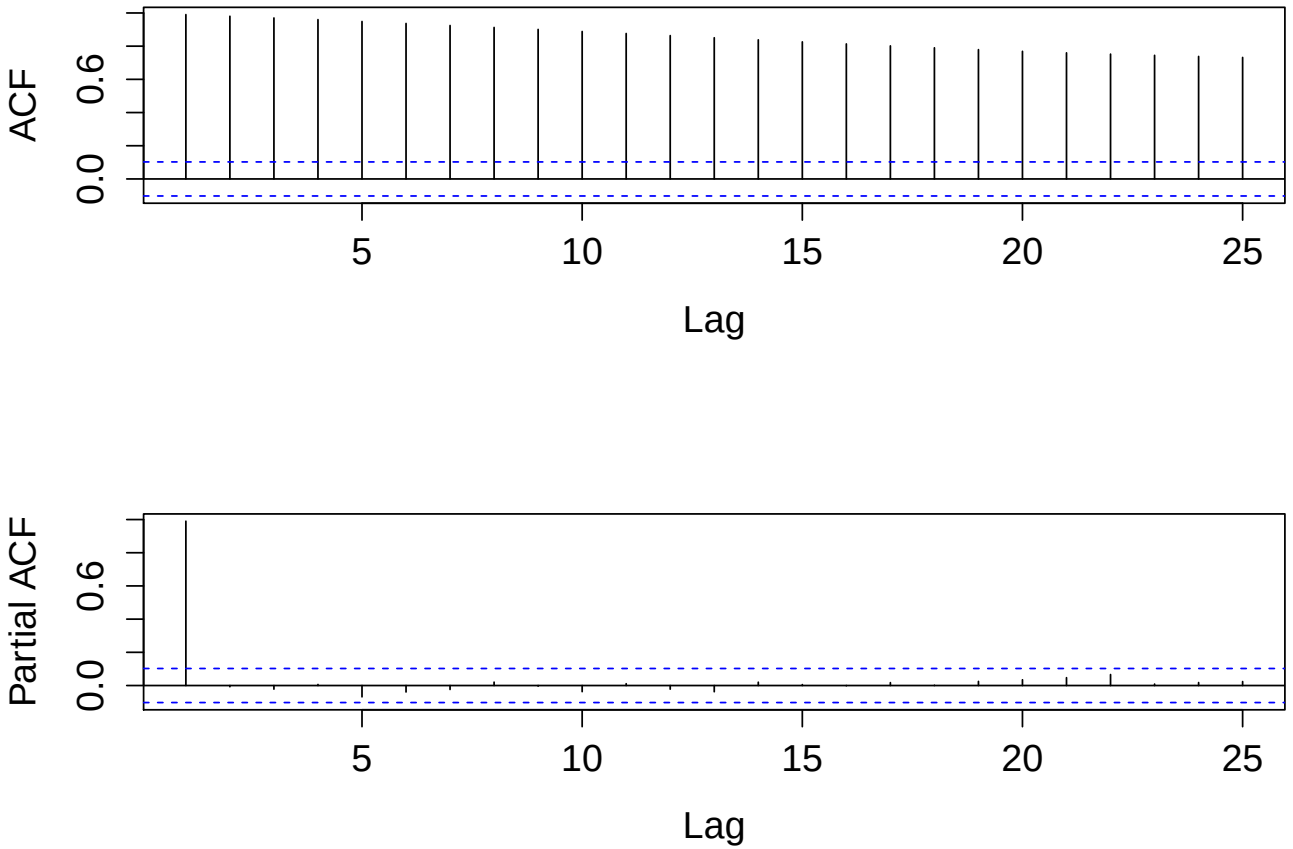
Differencing:	$h = 0, 1$ or 2; $H = 0$ or 1
AR part:	$p = 0, 1$ or 2; $P = 0$ or 1
MA part:	$q = 0, 1$ or 2; $Q = 0$ or 1

Example: Random walk



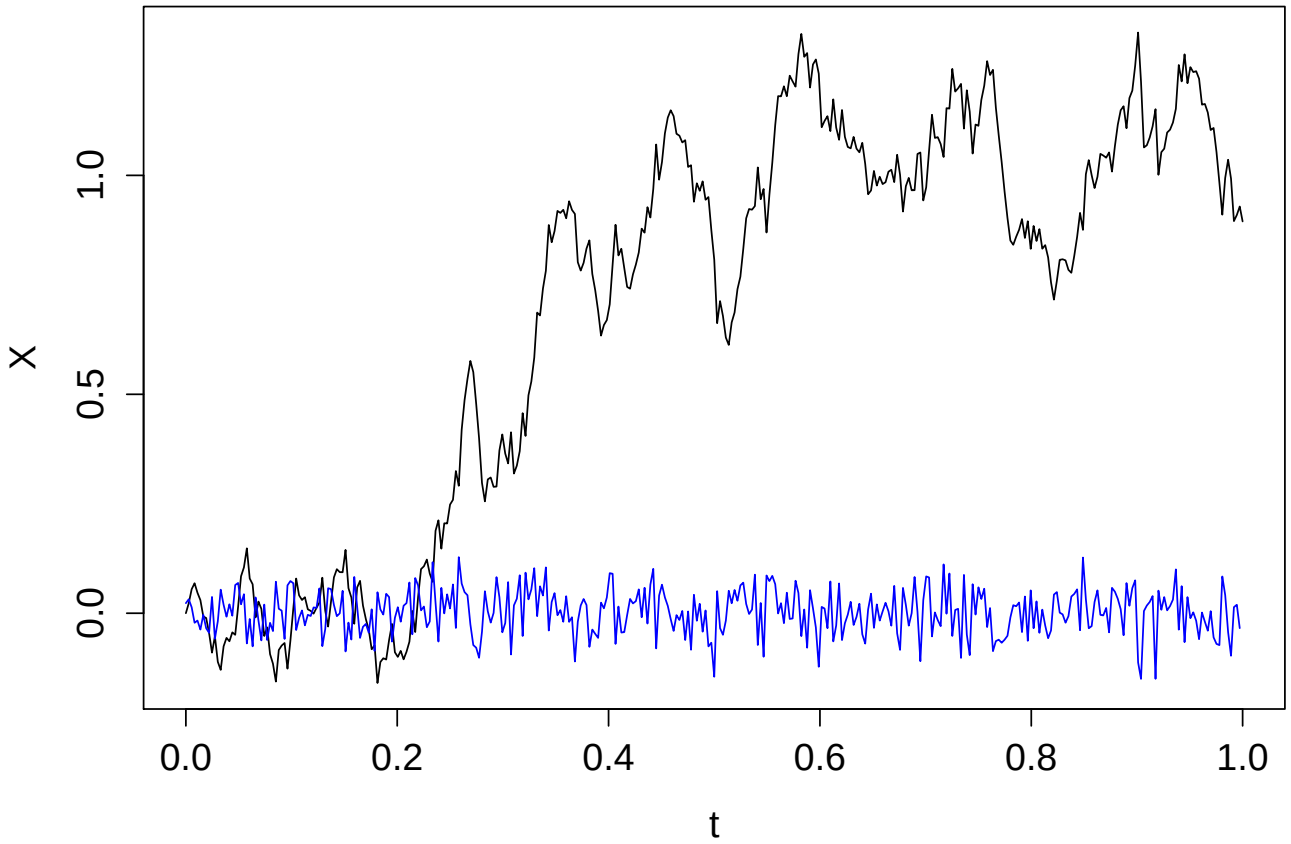
A random walk

Example: Random walk



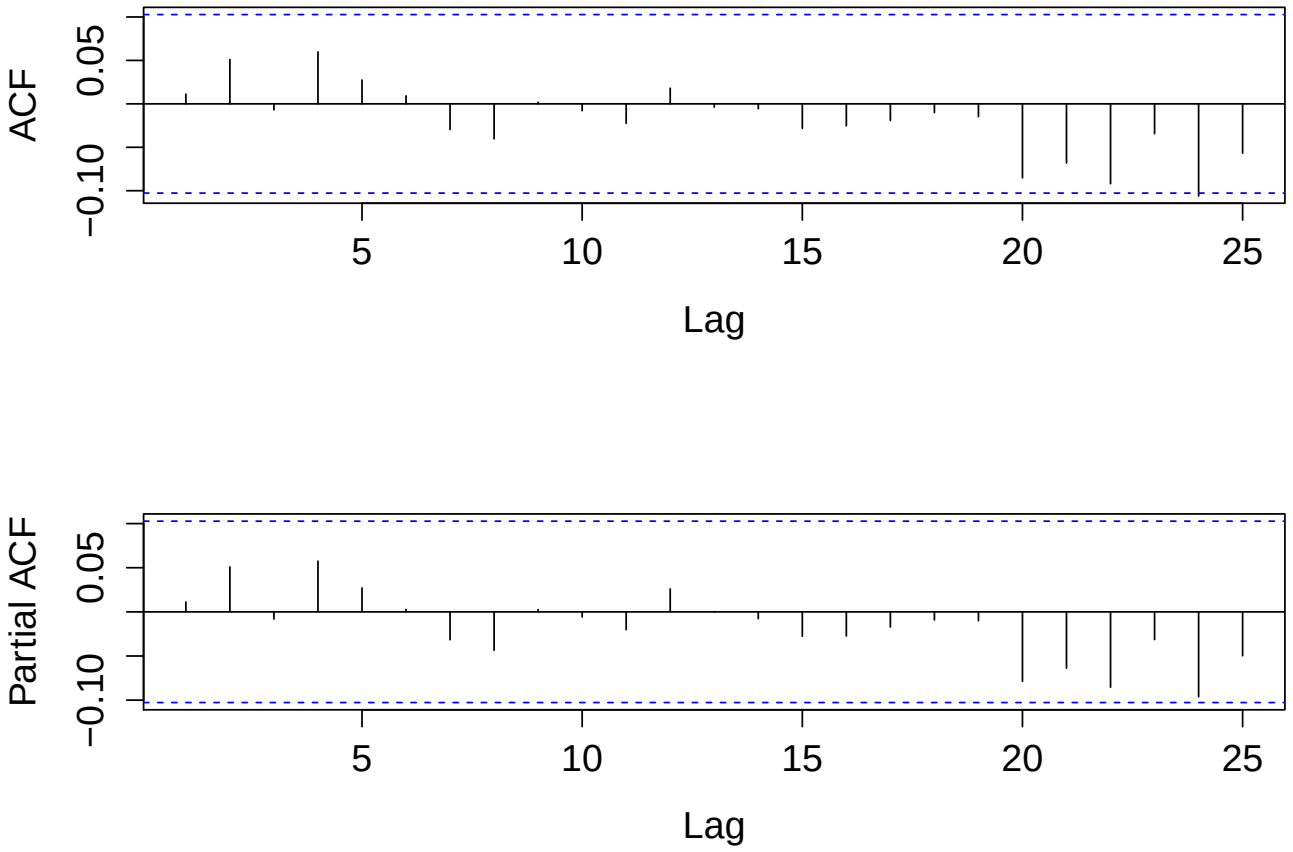
The autocorrelation and the partial autocorrelation of the random walk

Example: Random walk



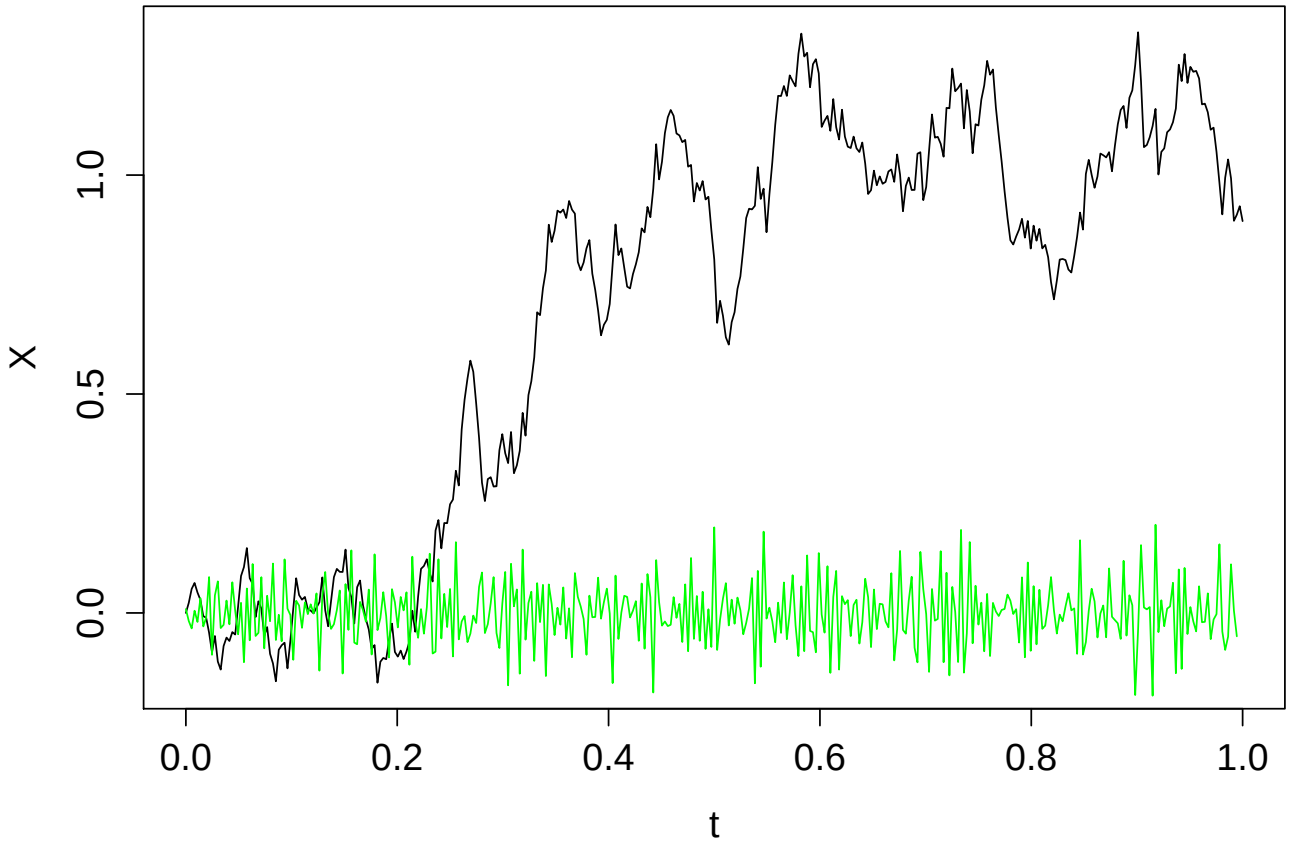
The random walk (in black) and the differenced random walk (in blue)

Example: Random walk



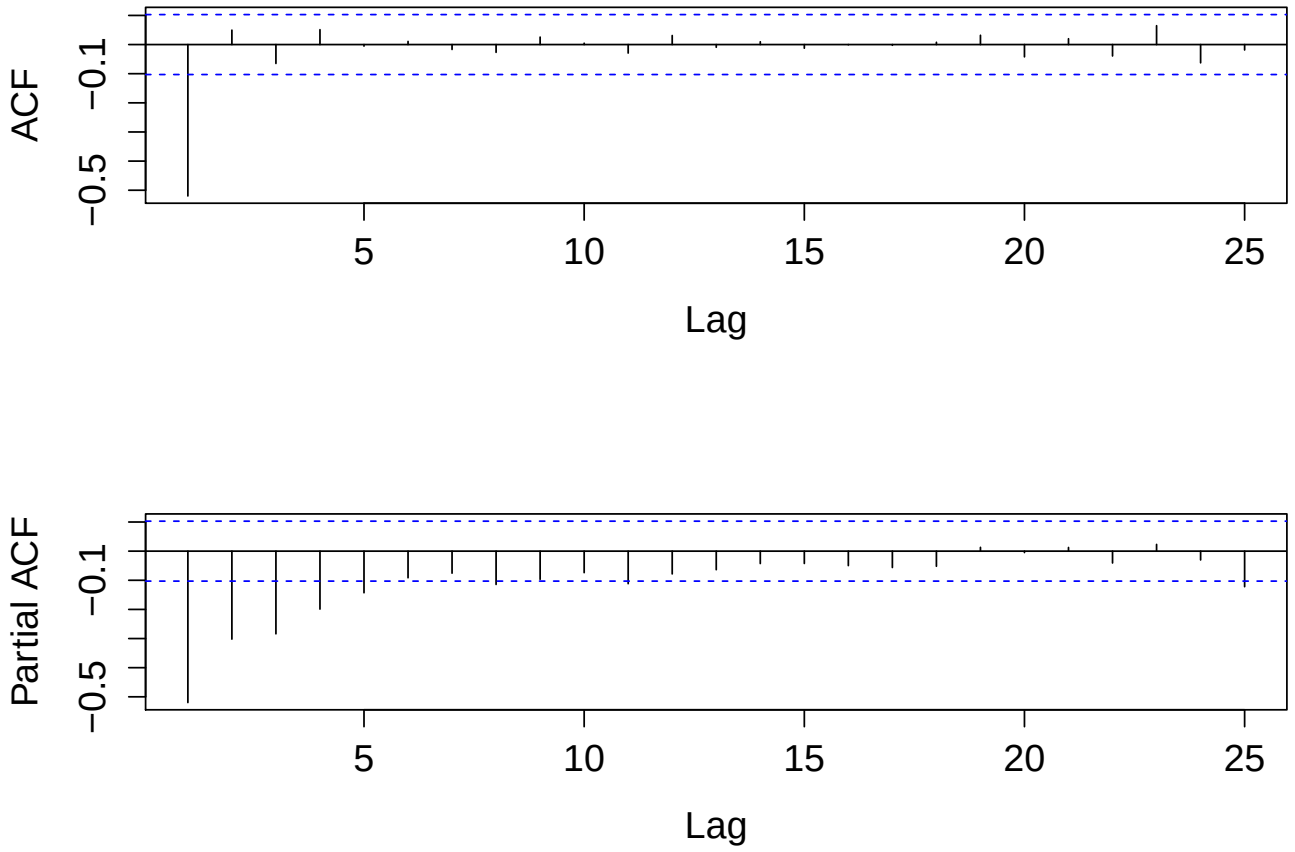
The autocorrelation and the partial autocorrelation of the differenced random walk

Example: Random walk



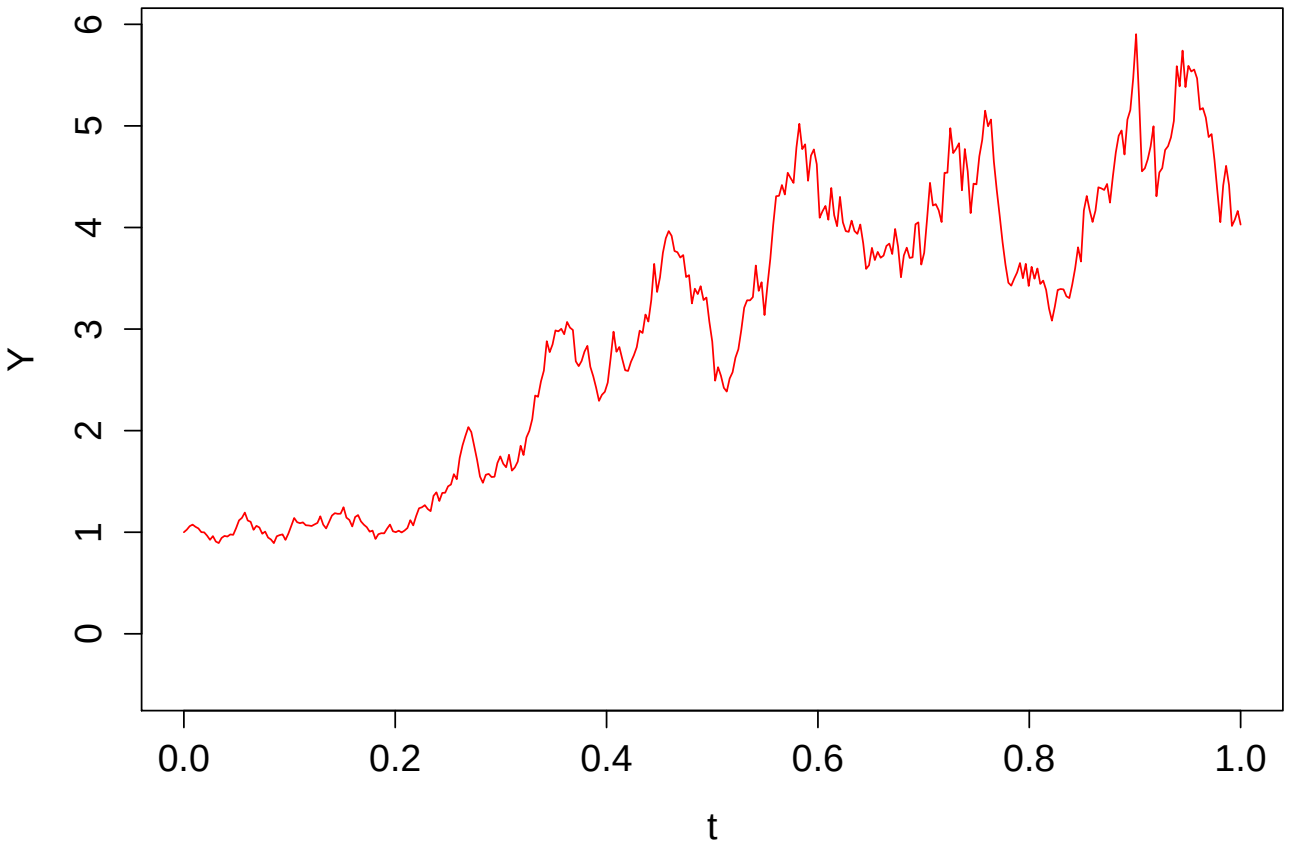
The random walk (in black) and the two times differenced random walk (in green)

Example: Random walk



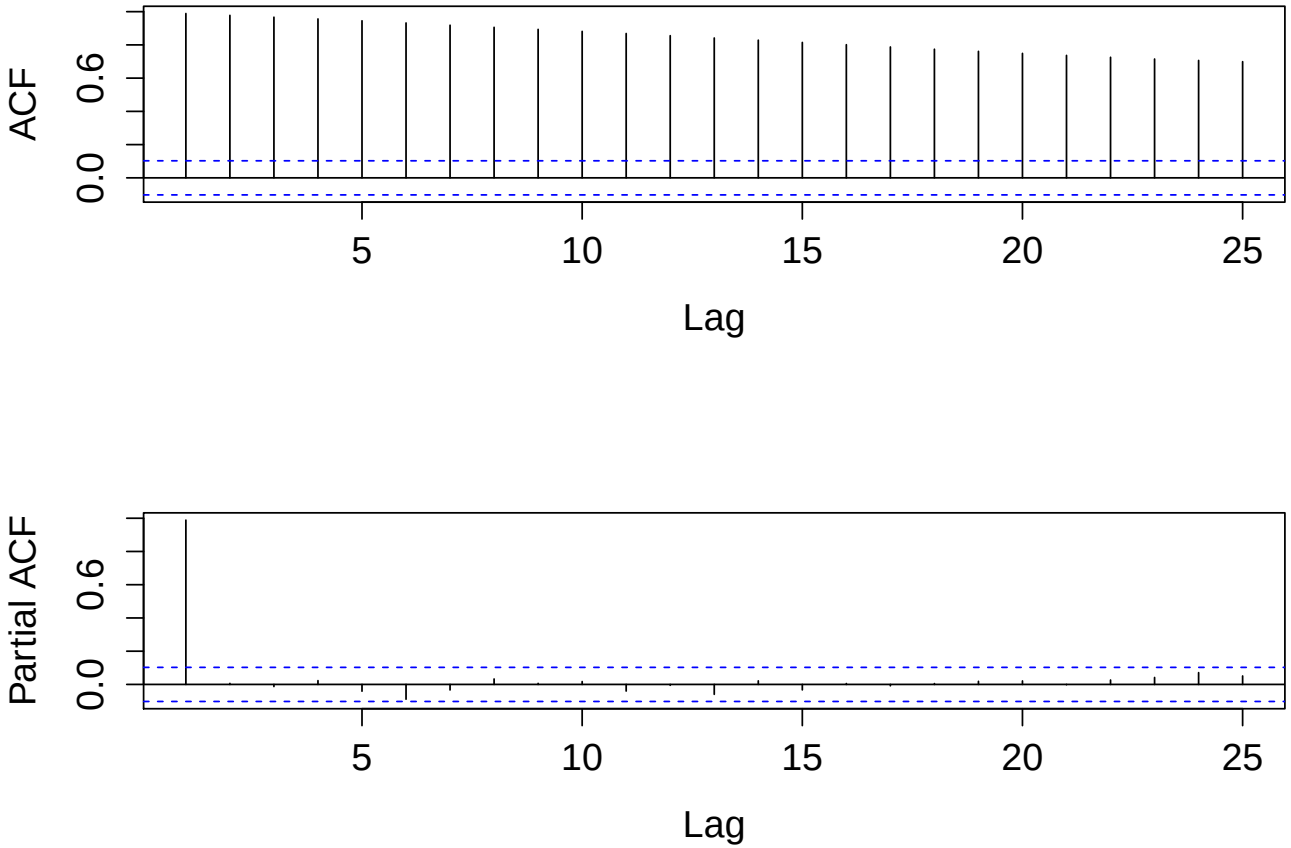
The autocorrelation and the partial autocorrelation of the two times differenced random walk

Example: Geometric random walk (GRW)



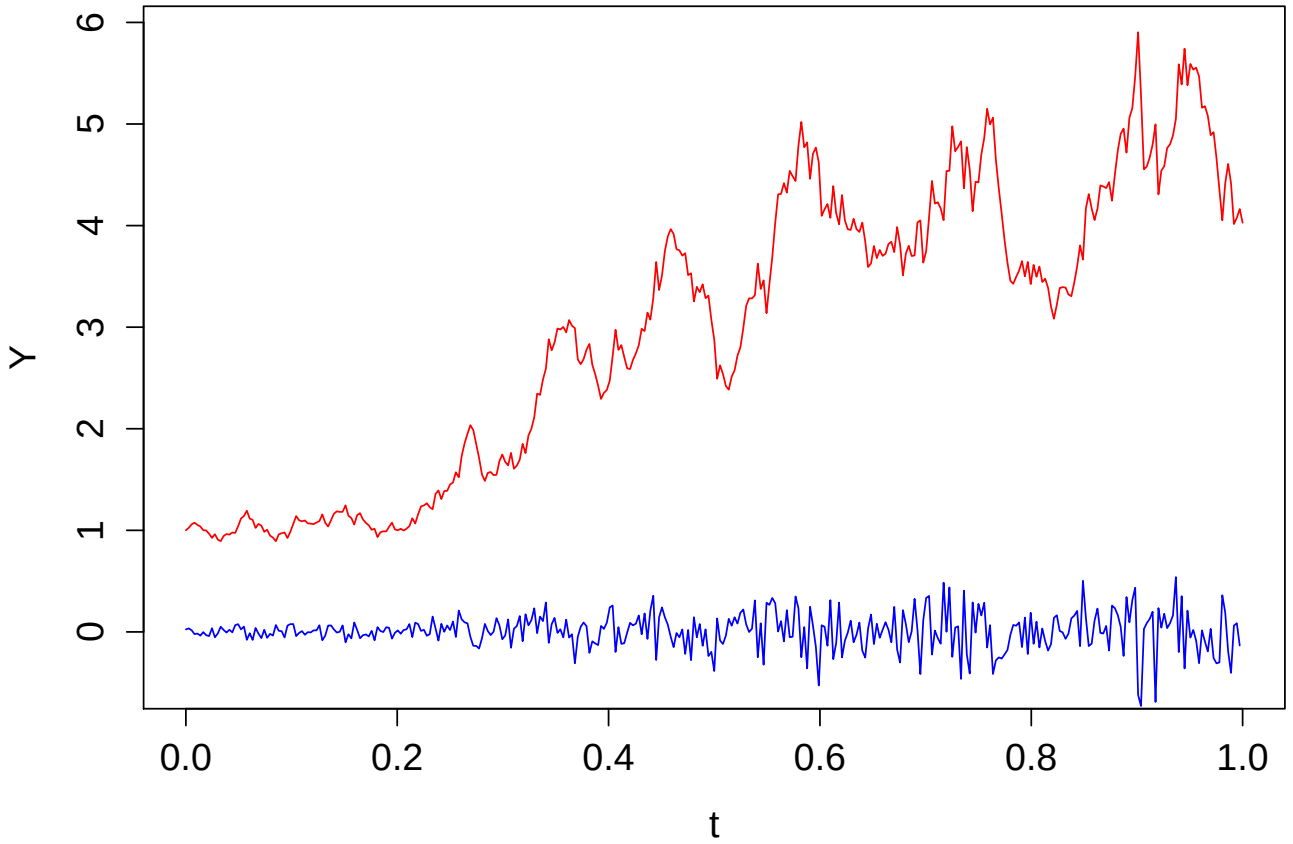
A geometric random walk

Example: Geometric random walk



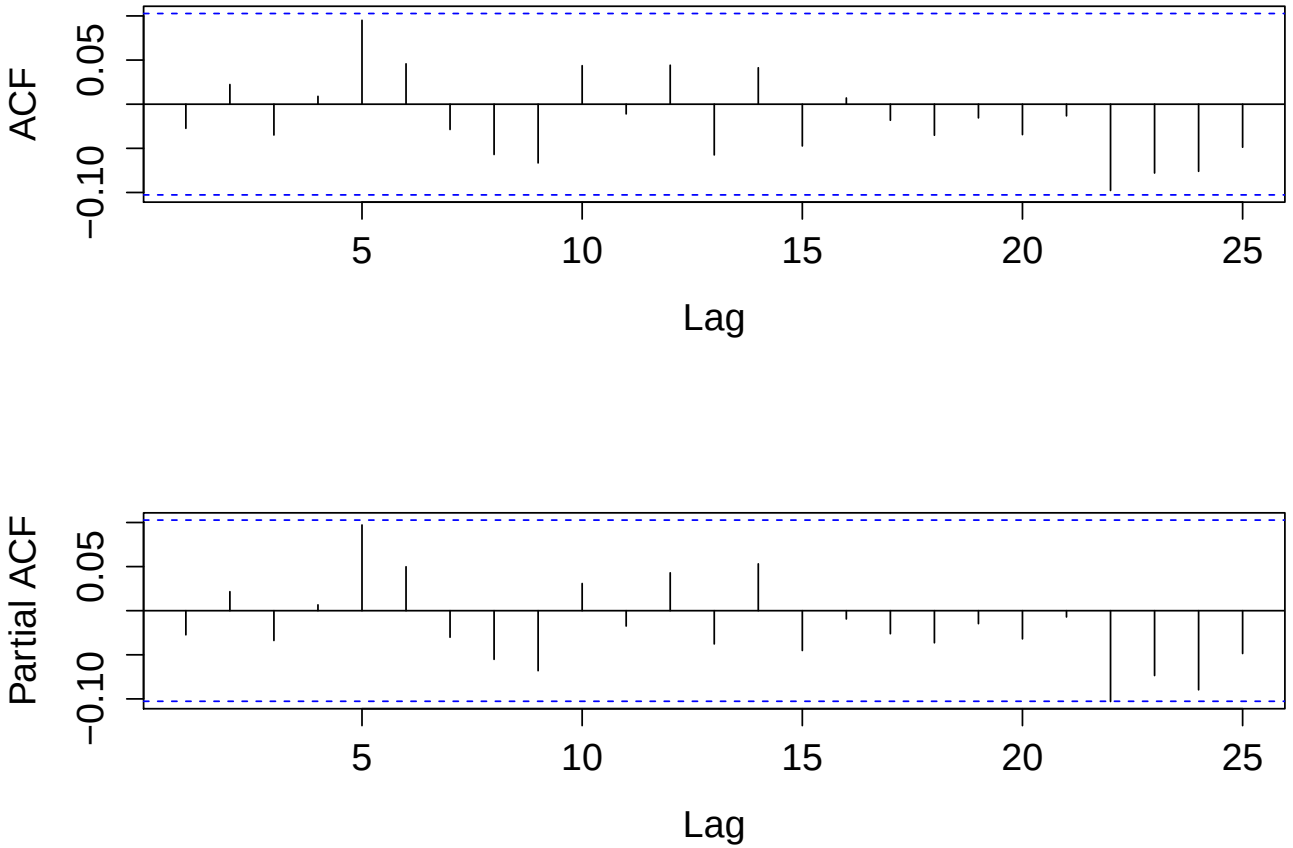
The autocorrelation and the partial autocorrelation of the geometric random walk

Example: Geometric random walk



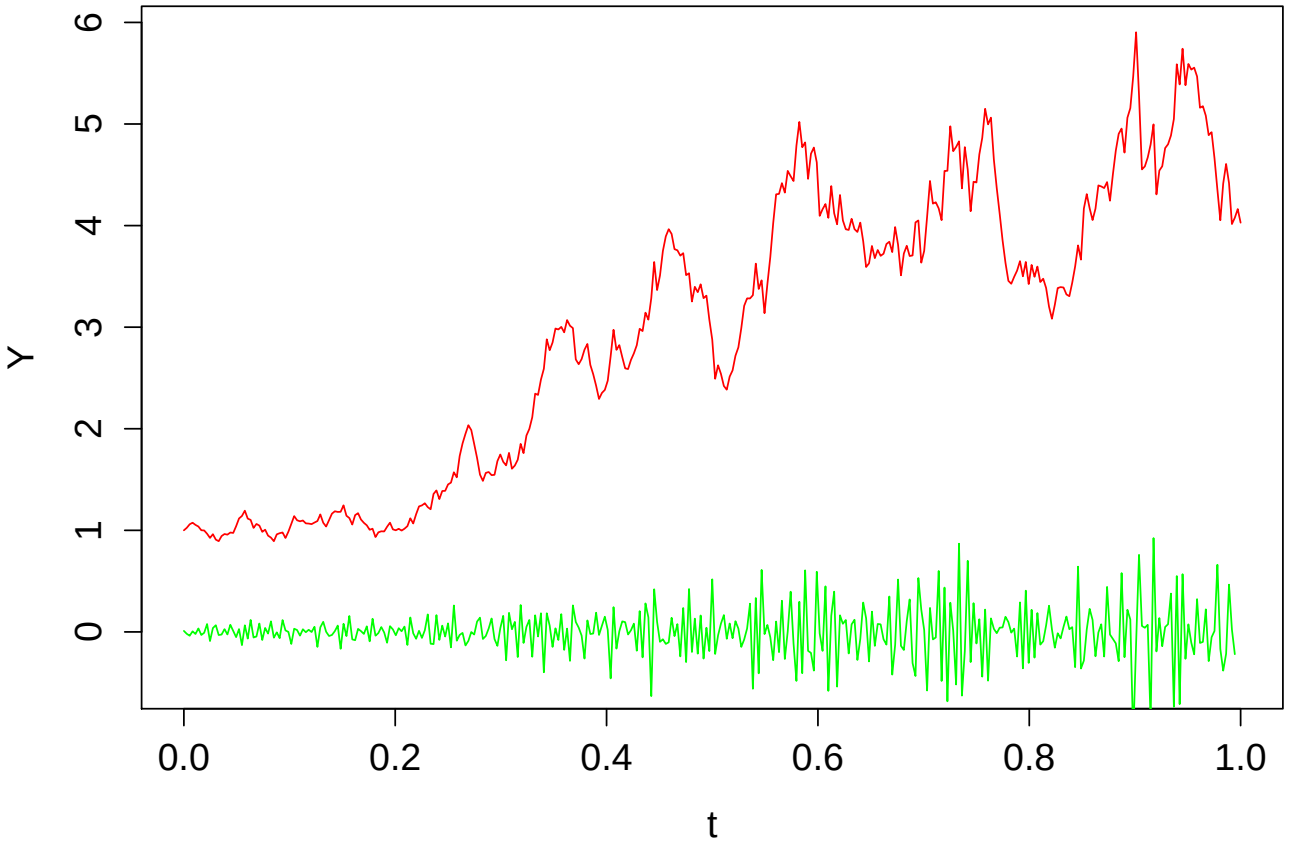
The geometric random walk (in red) and the differenced geometric random walk (in blue)

Example: Geometric random walk



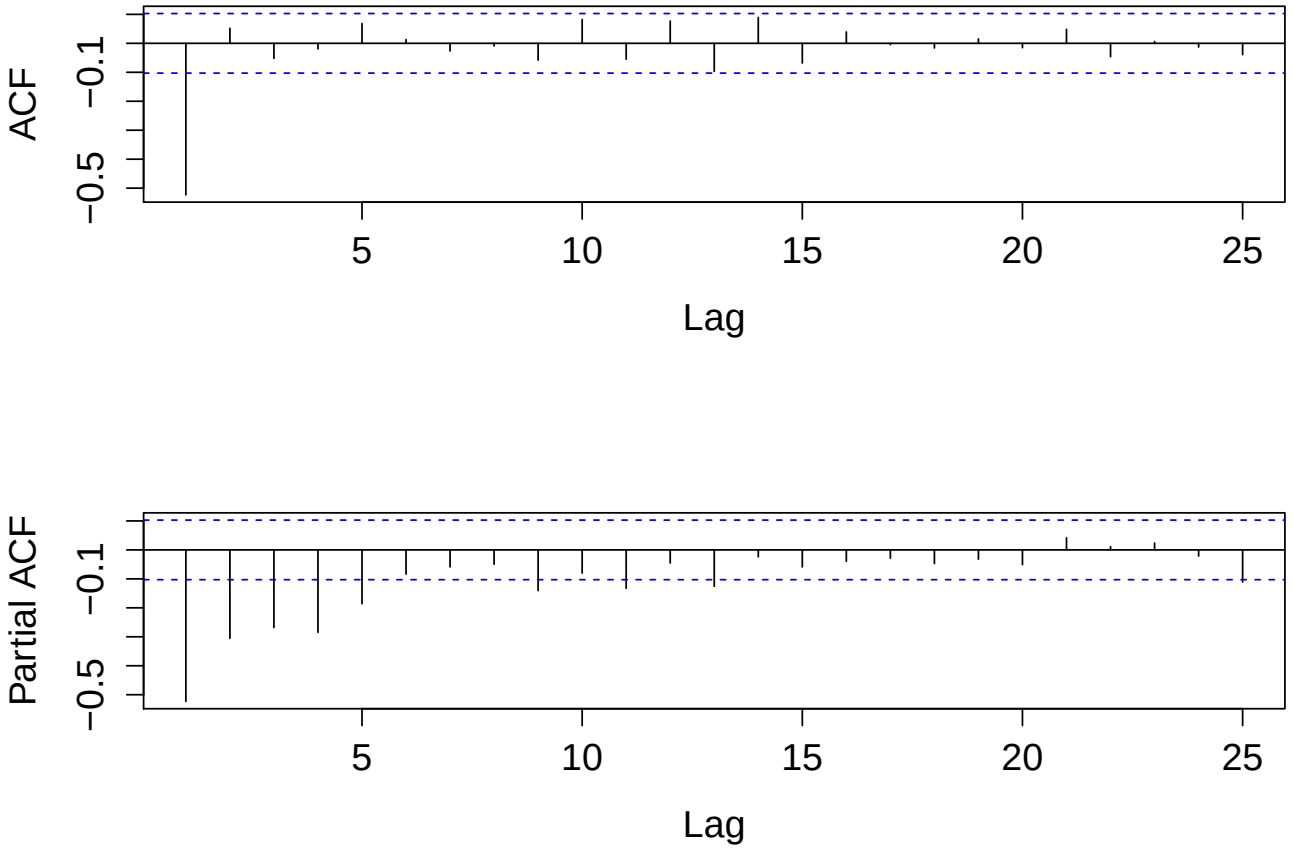
The autocorrelation and the partial autocorrelation of the differenced geometric random walk

Example: Geometric random walk



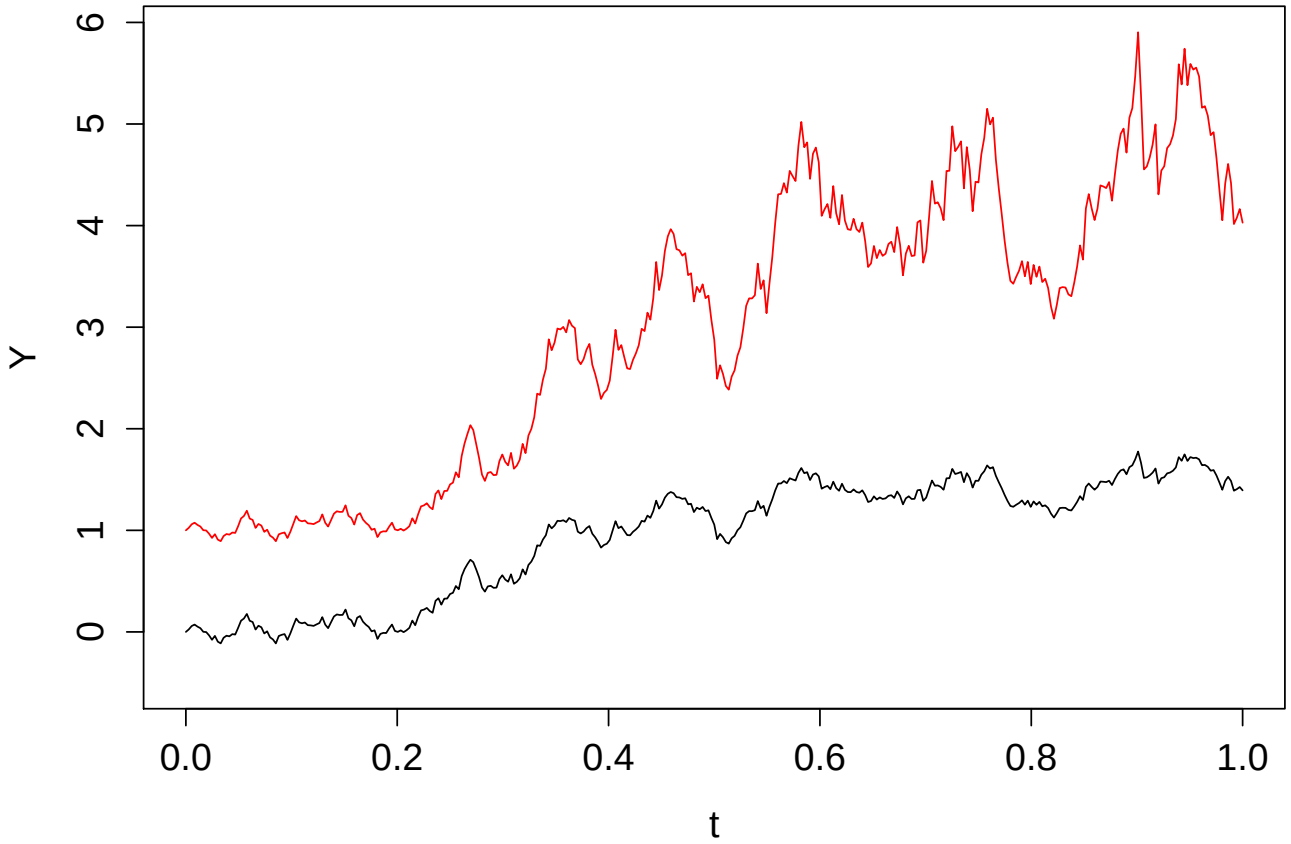
The geometric random walk (in red) and the two times differenced geometric random walk (in green)

Example: Geometric random walk



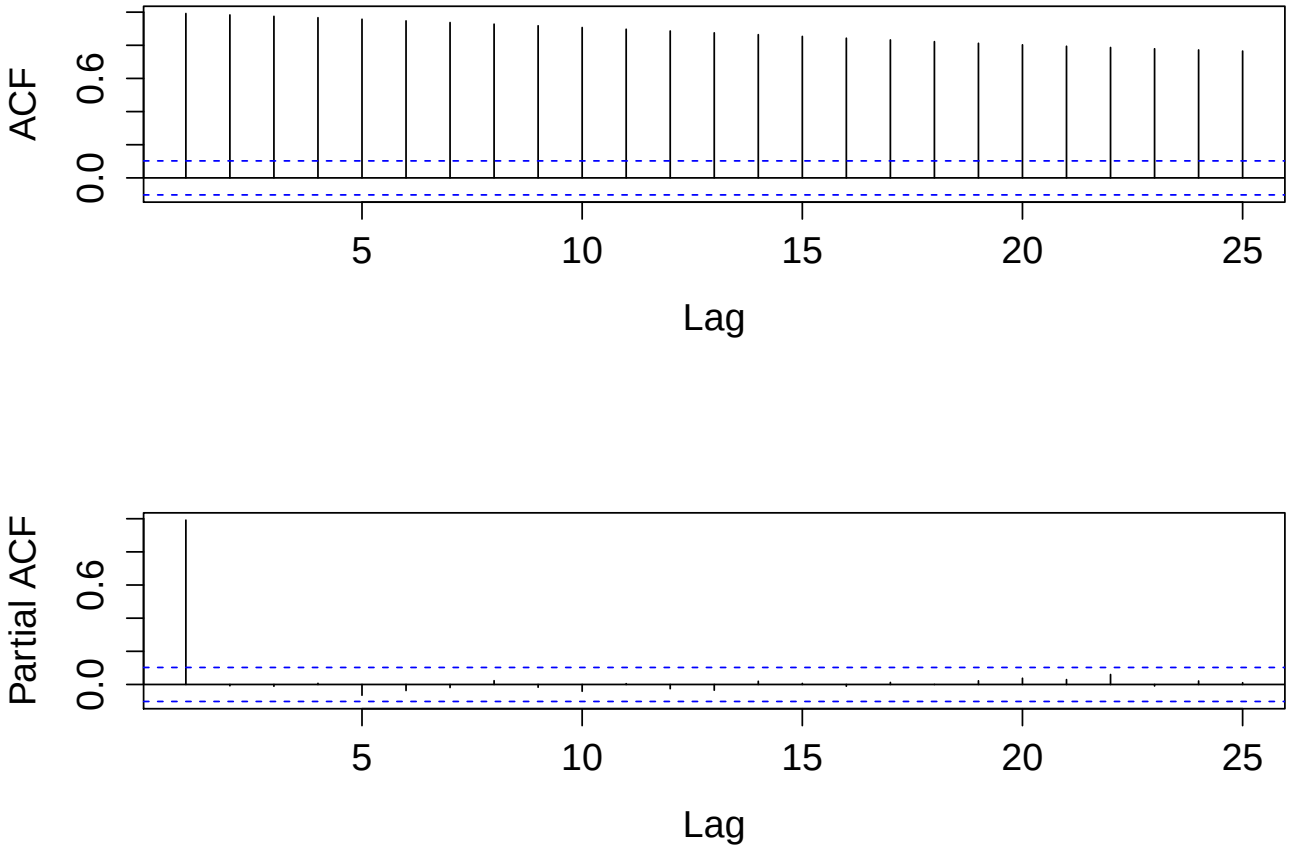
The autocorrelation and the partial autocorrelation of the two times differenced geometric random walk

Example: Geometric random walk



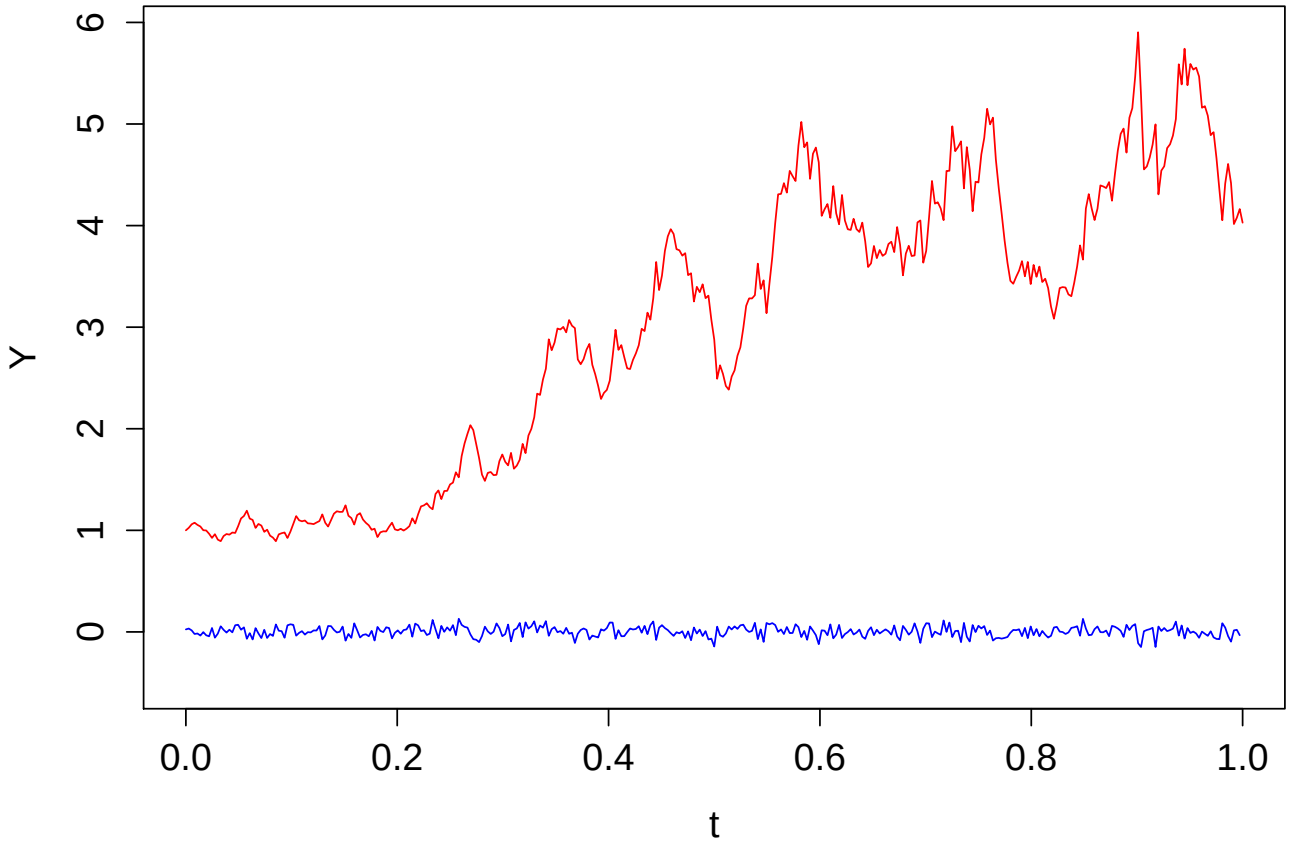
The geometric random walk (in red) and the log-transformed geometric random walk (in black)

Example: Geometric random walk



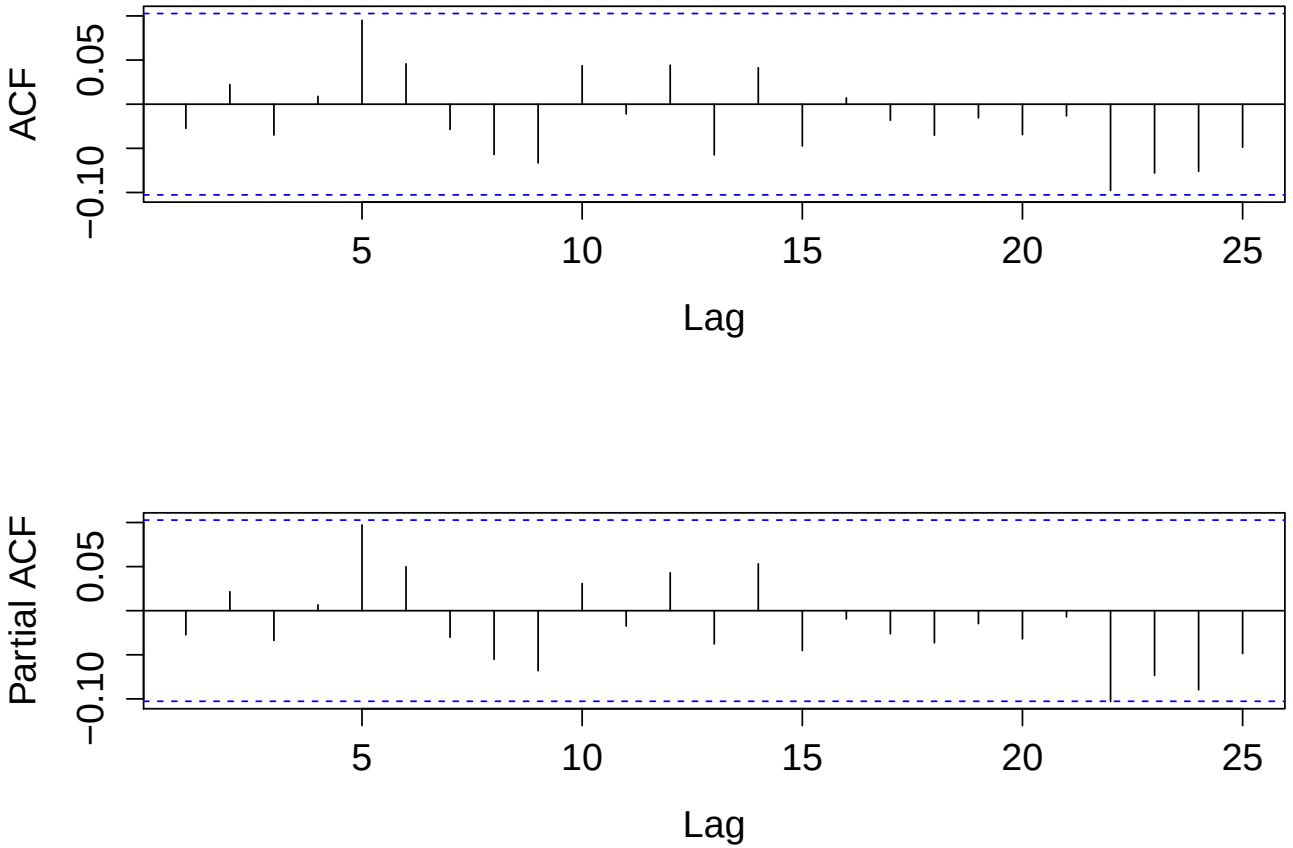
The autocorrelation and the partial autocorrelation of the log-transformed geometric random walk

Example: Geometric random walk



The geometric random walk (in red) and the diffenrenced log-transformed geometric random walk (in blue)

Example: Geometric random walk



The autocorrelation and the partial autocorrelation of the differenced log-transformed geometric random walk

Box-Jenkins method:

2. Model estimation

- SARMA model parameters can be estimated using R. One can use R-functions (for example `arima()`) that estimate the model parameters using some suitable estimation method (for example maximum likelihood estimation).

Box-Jenkins method:

3) Diagnostics

- Diagnostics are based on analyzing the residuals of the estimated SARMA model:
  - The residual time series and the plots of its correlations and spectrum are examined.
  - The uncorrelatedness of the residuals is tested.
- The estimated model is considered good enough if the residuals are white noise.
  - If the model is not good enough, one has to return to step 1 and try again.

- The uncorrelatedness of the residuals can be tested using Ljung-Box Q-statistic

$$Q_K = n(n+2)\sum_{i=1}^{K}\frac{r_i^2}{n-i}.$$

  - The $r_i,\ i \in \{1, 2, ..., K\}$, in the test statistic are the sample autocorrelations of the residuals with lag i.
  - If the residuals are heavily correlated, the value of the test statistic is large.

Box-Jenkins method:

3) Diagnostics

- If the SARMA model null hypothesis H_0: $\epsilon_t \sim WN$ holds, then

$$Q_K \sim_a \chi^2(K - m)$$

  - Above, m is the number of the estimated SARMA model parameter.
  - If the value of the test statistic Q_K is large, the null hypothesis is rejected.
- The value of the Q_K test statistic and its distribution under the null do depend on the number, K, of the autocorrelations considered.
  - It is advisable to calculate Q_K for several different K.

Remark

- Ljung-Box test statistics tests the significance of the K first autocorrelations.
- K has to be larger than the number m of the estimated model parameters.
- In practice, since the degrees of freedom of the asymptotic distribution, $\chi^2(K-m)$, of the test statistic under the null depend on K, the power of the test statistics decreases as K increases.
- If K is small, autocorrelations with larger lag are not tested.
- There are no clear rules for how one should select K.

Decomposition

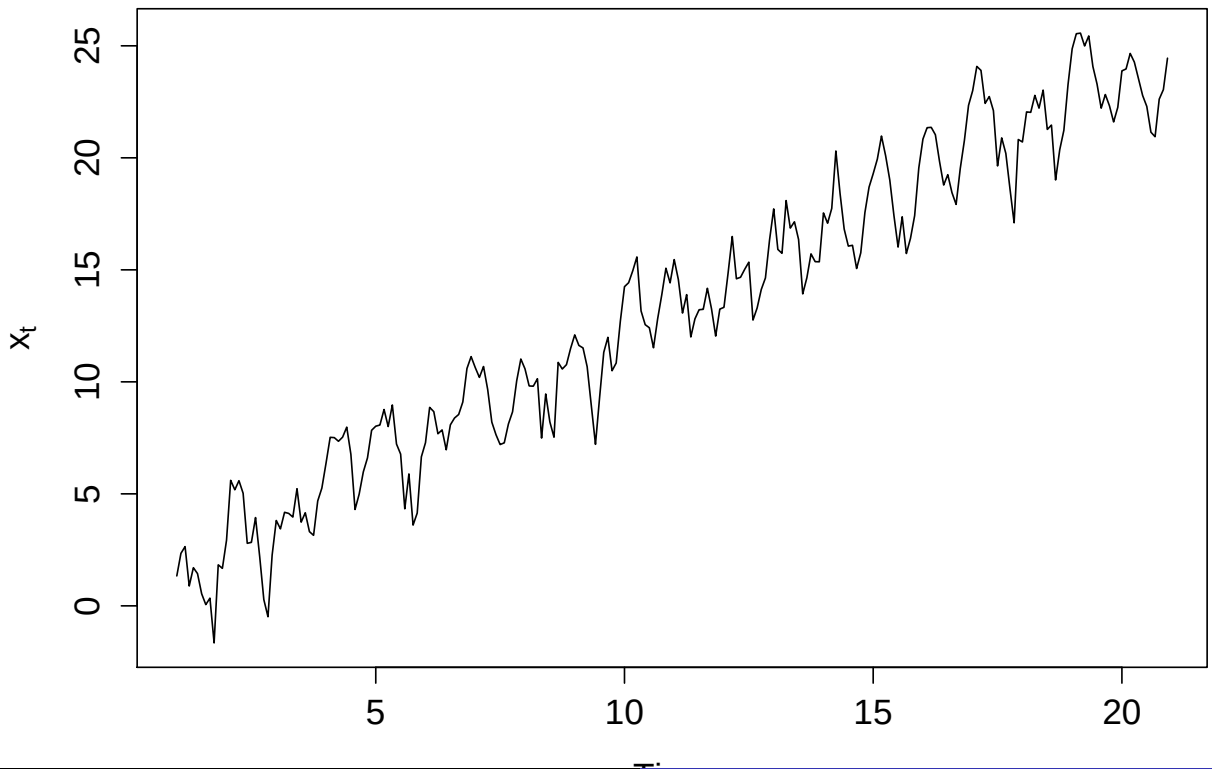
- When one examines time series data, one can often detect:
  - Trends
  - Cyclic changes
  - Seasonal changes
  - Random changes
- This empirical observation has led to the idea that, as part of statistical analysis, one could decompose time series data into corresponding components.

Decomposition: Goals

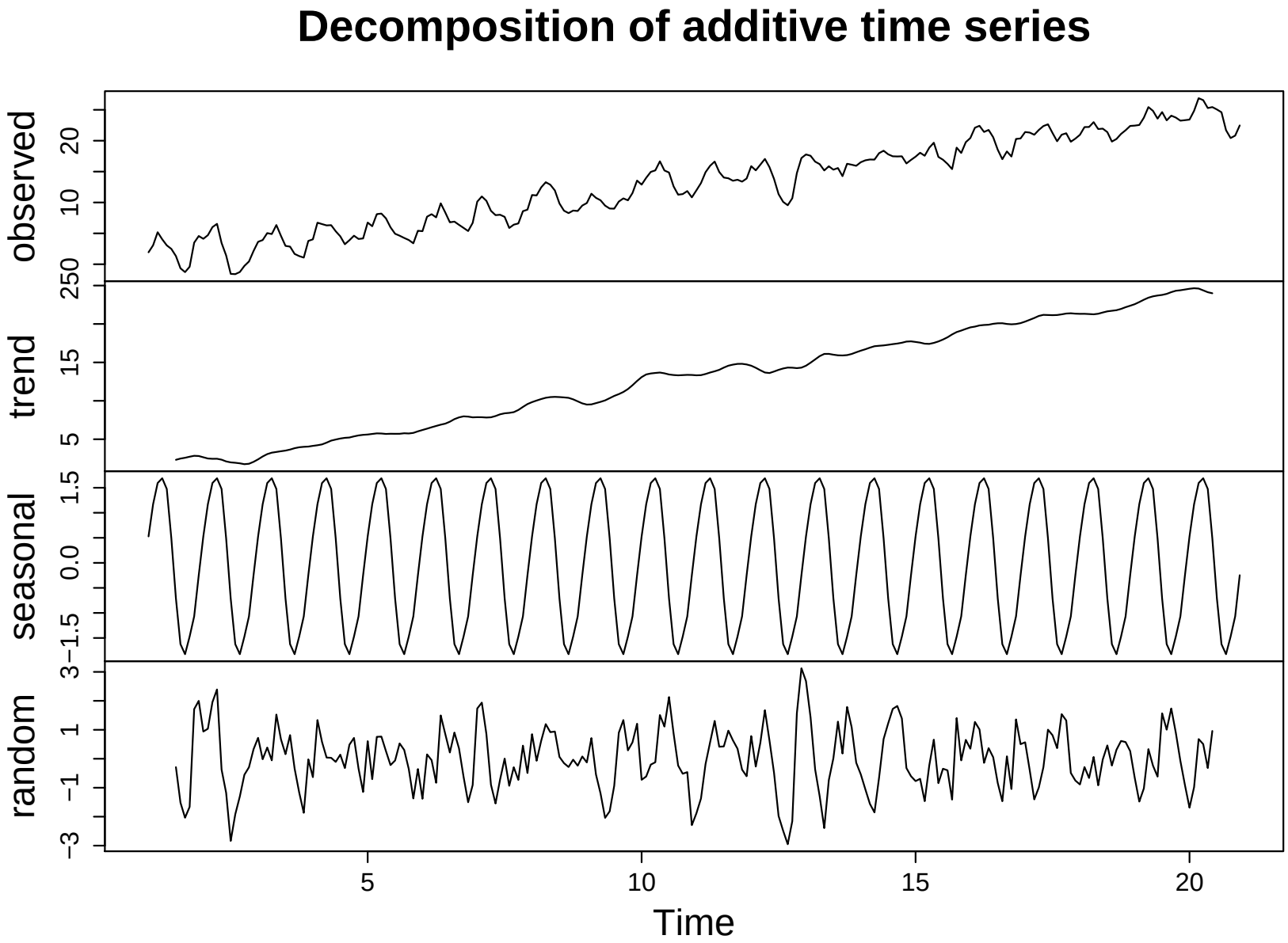
(i) Description of the behavior of time series data using components.

(ii) Analysis of time series data using components.

(iii) Elimination of disturbing seasonal changes in order to conduct better statistical analyses.



Decomposition



Decomposition

- When time series data is decomposed, it is assumed that the series x_t , $t = 1, 2, ..., n$ can be given as a sum or as a product of the following components:

$$\begin{aligned} m_t &= \text{trend} \\ c_t &= \text{cyclic component} \\ s_t &= \text{seasonal component} \\ e_t &= \text{random component.} \end{aligned}$$

- An additive model: $x_t = m_t + c_t + s_t + e_t$.
- A multiplicative model: $x_t = m_t c_t s_t e_t$.
  - A multiplicative model can be transformed into an additive model:

$$\log x_t = \log m_t + \log c_t + \log s_t + \log e_t.$$

Decomposition

Remark

Cyclic changes and seasonal changes are not the same thing:

- Cyclic changes are irregular and cycles may be long.
  - Example: economic cycles (expansion, peak, recession and recovery)
- Seasonal changes are regular and the season length remains the same.
  - Example: Christmas trees sales

Decomposition: Seasonal adjustment

- The goal in decomposing a time series x_t is often seasonal adjustment.
- Seasonal adjustment is based on constructing a new series y_t from which the disturbing seasonal component s_t has been eliminated:
  (i) Seasonal adjustment, additive model:

$$y_t = x_t - s_t = m_t + c_t + e_t$$

  (ii) Seasonal adjustment, multiplicative model:

$$y_t = \frac{x_t}{s_t} = m_t c_t e_t.$$

Decomposition methods

- Commonly applied decomposition methods:
  - X12 (an iterative moving average method)
  - X12-ARIMA (a combination of iterative moving average method and ARIMA modeling)
  - Structural time series models

Decomposition: Comments

- Justification:
  - It is easier to analyze decomposed/seasonally adjusted time series data.
- Critique:
  - The division of a time series into different components is always more or less arbitrary.
  - The components are not real, measurable quantities.
  - Decomposition methods (except structural time series models) are not based on statistical modeling.
  - It is very difficult to statistically measure the goodness of a decomposition.
  - Seasonal adjustment distorts the autocorrelation structure (internal time dependencies) of the series.

Decomposition: Comments

- Conclusion:
  - Decomposition can be used when the behavior of a time series is described, but using the components in formal statistical modeling is questionable.
- Seasonal adjustment can be replaced by methods that are statistically better motivated:
  - Aggregation in time
    - A new time series is obtained by combining (by summing or averaging) consecutive observations.
  - Sampling in time
    - A new time series is obtained by picking observations at regular time interval points (for example every 7th observation).
  - Seasonal differencing

References:

1. Brockwell, P., Davis, R. (2009): Time Series – Theory and Methods, Springer
2. Hamilton, J. (1994): Time Series Analysis, Princeton University Press

Next week

1. Prediction
    1. Predicting using ARMA models
    2. Exponential smoothing
2. Kalman filter
3. Dynamic regression