MATH 7339 - Machine Learning and Statistical Learning Theory 2

Section Stationary Process

1. Stationarity
2. Autocovariance, autocorrelation
3. MA, AR, linear processes
4. Sample autocorrelation function

➢ Strictly Stationarity

The Times Series $\{X_t\}$ is **strictly stationary** if

$$\{X_{t_1}, \dots, X_{t_k}\} \text{ and } \{X_{t_1+h}, \dots, X_{t_k+h}\}$$

have the **same joint distribution** for every $k, t_1, \dots, t_k,$ and h

That is, for all $k, t_1, \dots, t_k, x_1, \dots, x_k$ and h,

$$P\left(X_{t_1} \le x_1, \dots, X_{t_k} \le x_k\right) = P\left(X_{t_1+h} \le x_1, \dots, X_{t_k+h} \le x_k\right)$$

i.e., shifting the time axis does not affect the distribution.

Location does not matter-ONLY the window size.

We shall consider **second-order moments** properties only.

❑ Mean, Autocovariance and Autocorrelation

Suppose that $\{X_t\}$ is a time series with $E[X_t^2] < \infty$

- The **mean function** of $\{X_t\}$ is $\mu_t := E[X_t]$

- Its **autocovariance function** of $\{X_t\}$ is

$$\gamma_X(s,t) := \mathrm{Cov}(X_s, X_t) = E[(X_s - \mu_s)(X_t - \mu_t)]$$

- The **autocorrelation function** (**ACF**) of $\{X_t\}$ is defined as

$$\rho_X(s,t) := \mathrm{Corr}(X_s, X_t) = \frac{\gamma_X(s,t)}{\sqrt{\gamma_X(s,s)\gamma_X(t,t)}}$$

A common feature of time series is that the observations are **dependent**.

The autocovariance function is simply the **covariance** between x_s and x_t evaluated at all combinations.

Covariance measures the strength of the **linear dependence** between random variables.

A covariance that is small for s, t close together generally implies random variables that are closer to white noise. Smoother series tend to have a large autocovariance even for s and t which are far apart.

Weak Stationarity

We say that $\{X_t\}$ is (**weakly**) **stationary** if

1. The mean function μ_t is independent of t, (constant) and
2. For each h, $\gamma_X(t + h, t)$ is independent of t. (only depending on h)

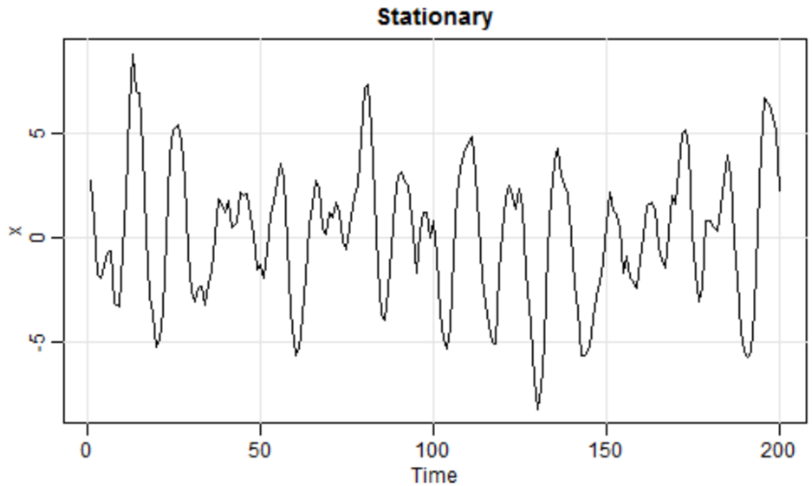
In that case, we write $\gamma_X(h) = \gamma_X(h, 0)$

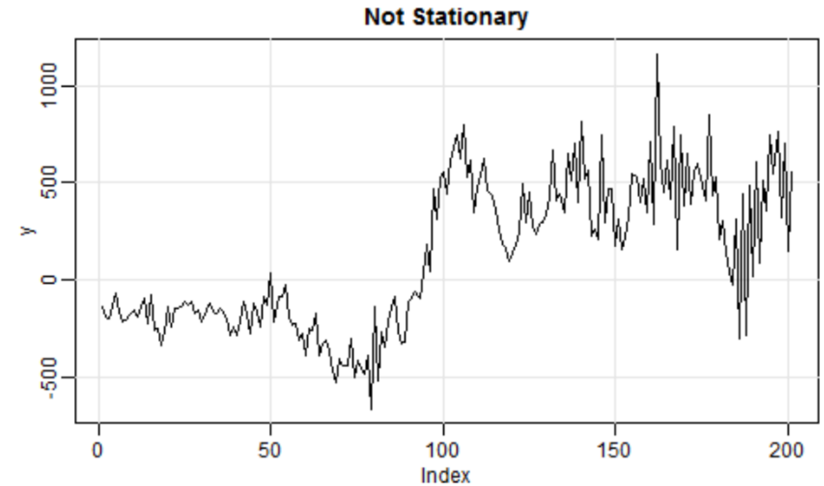
When we say stationary, we'll mean weakly stationary.

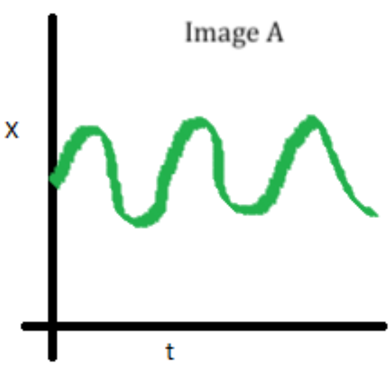
All strongly stationary time series are also weakly stationary, but the reverse may not be the case.

Most of the time we are going to be working with Gaussian time series, and in this case the two concepts coincide.

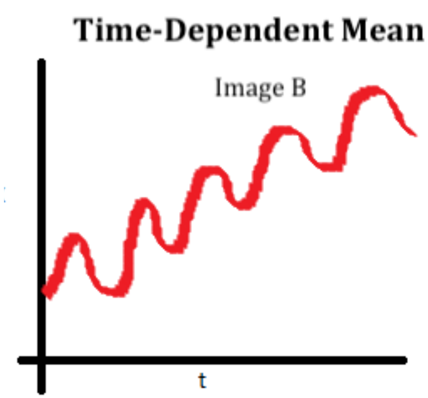
A time series is stationary when it is “stable”



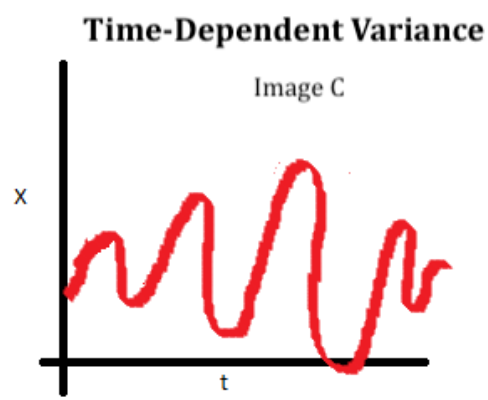




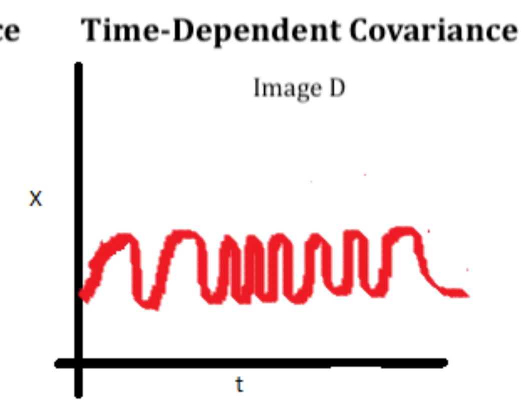
Stationary series



Non-Stationary series



Non-Stationary series



Non-Stationary series

Definition: The **autocorrelation function** (**ACF**) of stationary $\{X_t\}$ is defined as

$$\rho_X(h) := \frac{\gamma_X(h)}{\gamma_X(0)} = \frac{\mathrm{Cov}(X_{t+h}, X_t)}{\mathrm{Cov}(X_t, X_t)} = \mathrm{Corr}(X_{t+h}, X_t)$$

Example: (IID noise**)**

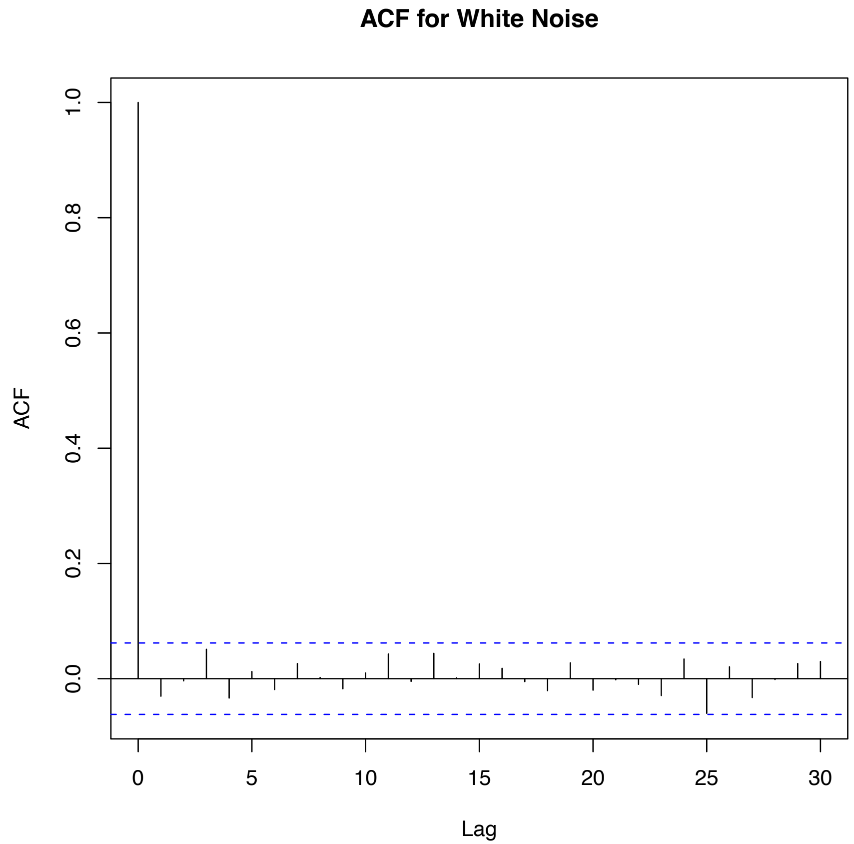
$$E[X_t] = 0, \ \mathrm{Var}(X_t) = \sigma^2,$$

$$\gamma_X(t+h, t) := \mathrm{Cov}(X_{t+h}, X_t) = \begin{cases} \sigma^2 & \text{if } h = 0 \\ 0 & \text{otherwise} \end{cases}$$

1. $\mu_t = 0$ is independent of t.
2. $\gamma_X(t+h, t) = \gamma_X(h, 0)$ for all t.

So, $\{X_t\}$ is stationary.

Similarly for any white noise (uncorrelated, zero mean).



A technical results states that when the true model is white noise, $\hat{\rho}$(h) is approximately normally distributed with zero mean and standard deviation of $1/\sqrt{n}$.

This is very useful for conducting tests concerning hypothesis about the true autocorrelation function.

Example: (Random Walk)

$$S_t = \sum_{i=1}^{t} X_i \qquad \text{where } X_i \text{ is the } \textbf{iid} \text{ noise.}$$

So, $E[S_t] = 0, \ \text{Var}(S_t) = t\sigma^2,$

$$\gamma_X(t+h, t) = \text{Cov}(S_{t+h}, S_t) = \text{Cov}\left(S_t + \sum_{s=1}^{h} X_{t+s}\ ,\ S_t\right) = \text{Cov}(S_t, S_t) = t\sigma^2$$

1. $\mu_t = 0$ is independent of t.
2. $\gamma_X(t+h, t)$ is <u>depending</u> on t.

So, $\{S_t\}$ is NOT stationary.

Example: MA(1) process (Moving Average):

$$X_t = W_t + \theta W_{t-1} \qquad \text{where } W_t \sim WN(0, \sigma^2)$$

So,

$$E(X_t) = 0$$

$$\gamma_X(t+h, t) = \text{Cov}(X_{t+h}, X_t)$$

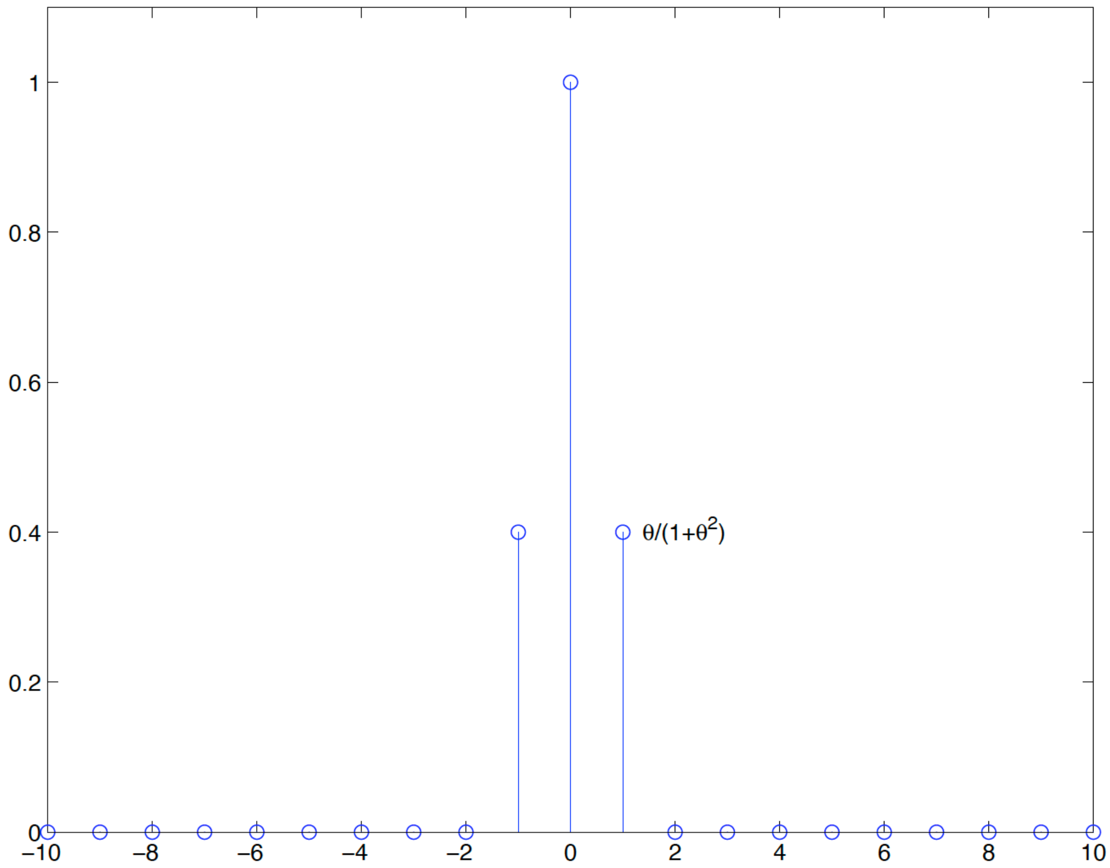
$$= \text{E}(X_{t+h} X_t)$$

$$= \text{E}((W_{t+h} + \theta W_{t+h-1})(W_t + \theta W_{t-1}))$$

$$= \begin{cases} \sigma^2(1+\theta^2) & \text{if } h = 0 \\ \sigma^2\theta & \text{if } h = 1, or\ -1 \\ 0 & \text{others} \end{cases}$$

So, $\{X_t\}$ is stationary.

ACF of the MA(1) process



Example: AR(1) process (AutoRegressive):

$$X_t = \phi X_{t-1} + W_t \qquad \text{where } W_t \sim WN(0, \sigma^2)$$

Assume that X_t is **stationary** and $|\phi| < 1$, then $X_t = \sum_{j=0}^{\infty} \phi^j W_{t-j}$

$$E[X_t] = \phi E[X_{t-1}] = 0$$

$$E[X_t^2] = \phi^2 E[X_{t-1}^2] + \sigma^2 = \frac{\sigma^2}{1-\phi^2}$$

$$\gamma_X(h) = \mathrm{Cov}(X_{t+h}, X_t) = \mathrm{Cov}(\phi X_{t+h-1} + W_{t+h}, X_t)$$

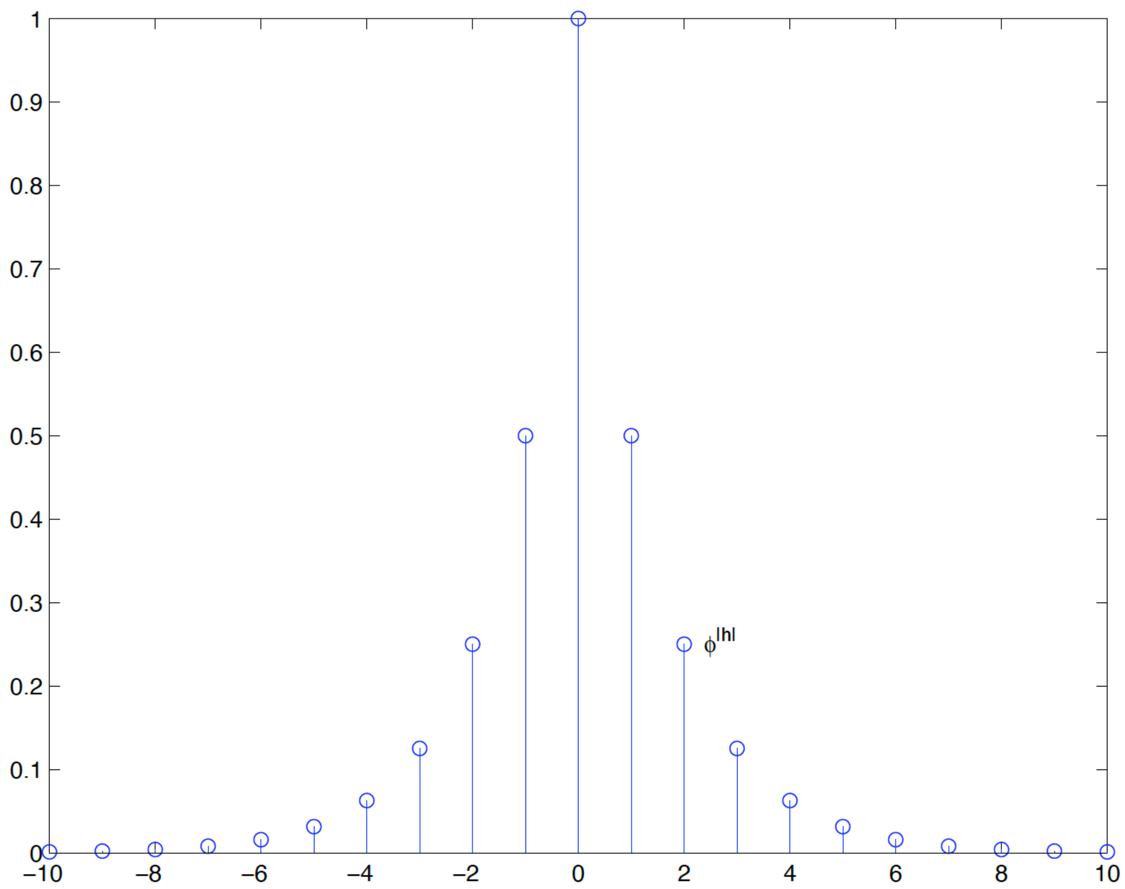
$$= \phi \mathrm{Cov}(X_{t+h-1} + W_{t+h}, X_t) = \phi \mathrm{Cov}(X_{t+h-1}, X_t)$$

$$= \phi \gamma_X(h-1) = \phi^{|h|} \gamma_X(0)$$

Check for $h > 0$ and $h < 0$

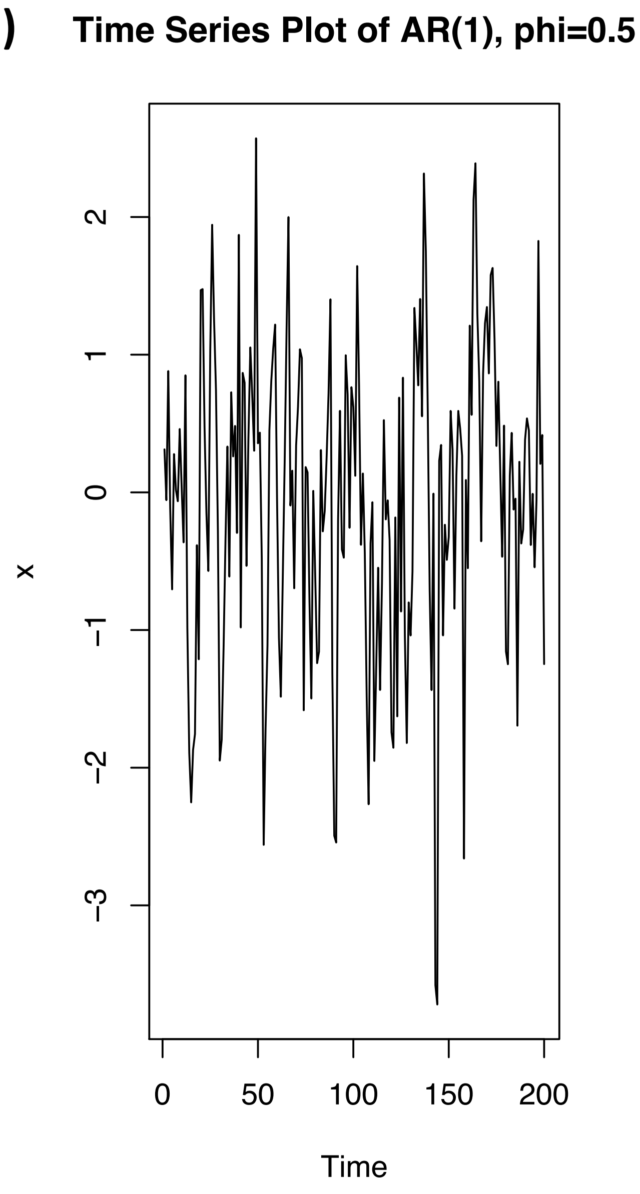
$$= \frac{\phi^{|h|}\sigma^2}{1-\phi^2}$$

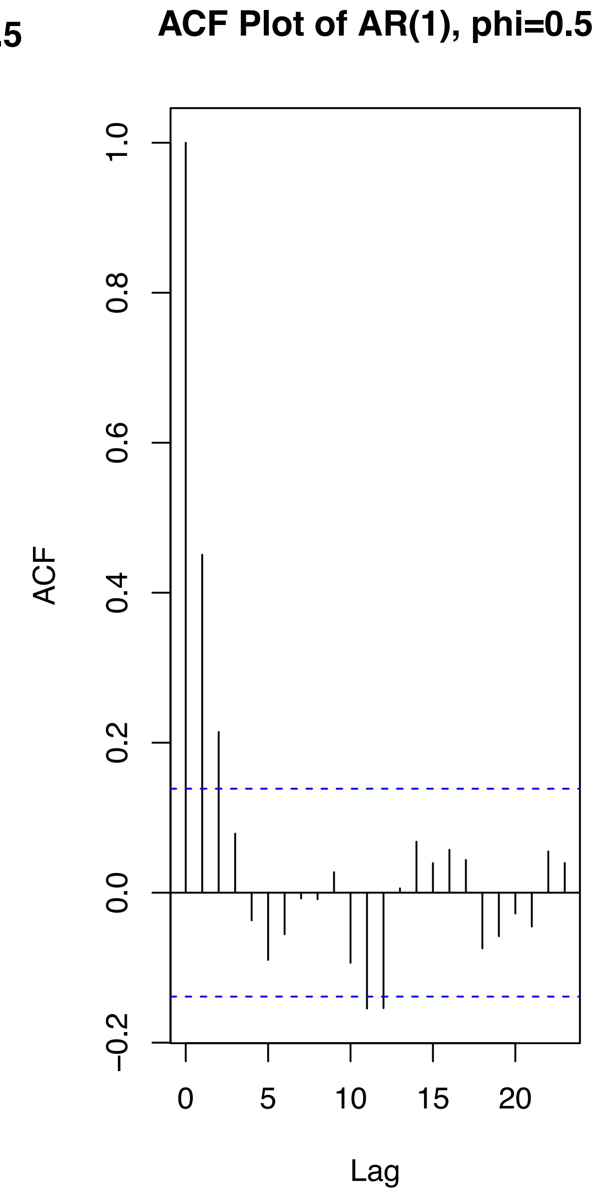
ACF of the AR(1) process



The ACF should show a geometric decline{but never truly go to zero. However, after a number of lags the series will be essentially zero. An ACF that does not fall off quickly but where the series does not appear stationary -may not be indicative of an AR model.

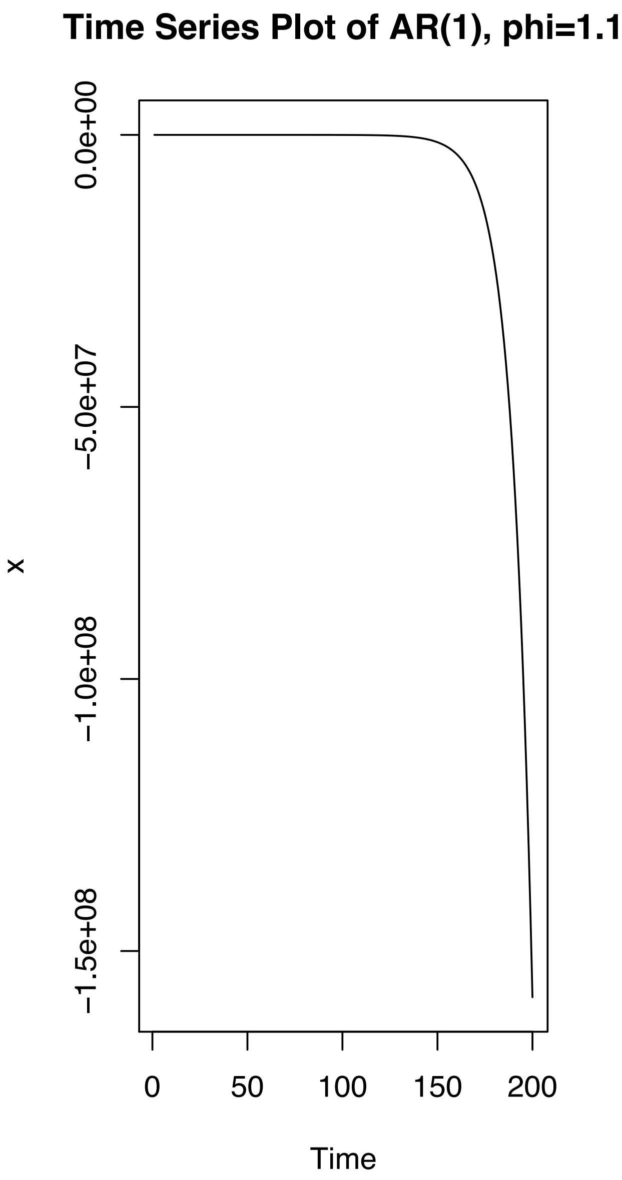
Stationary AR(1)

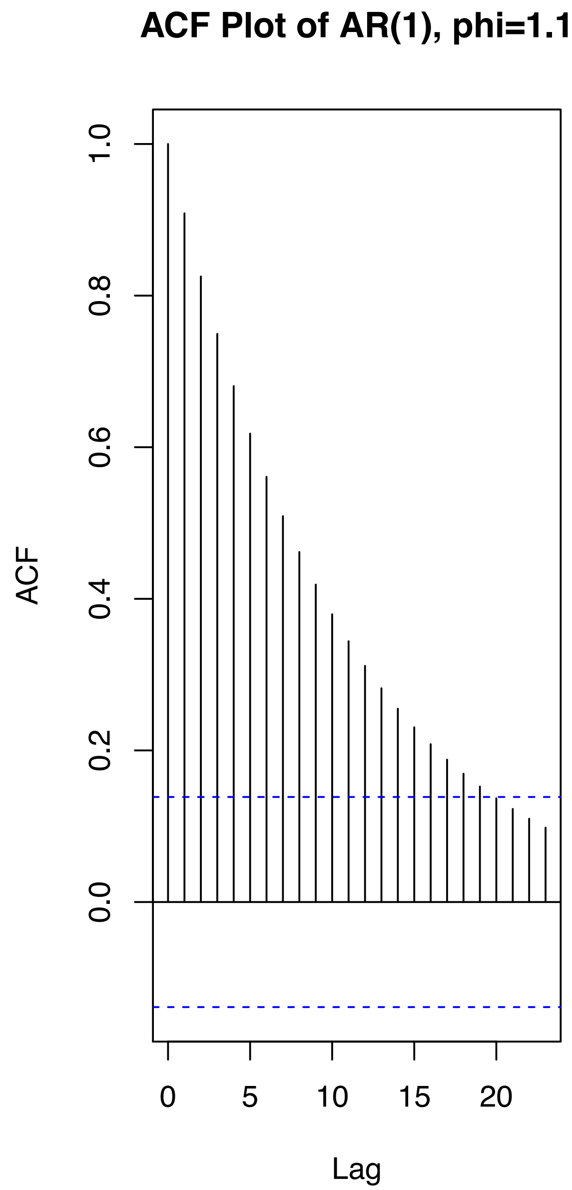




https://demonstrations.wolfram.com/AutocorrelationAndPartialAutocorrelationFunctionsOfAR1Proces/

Explosive AR(1)





Python Code:

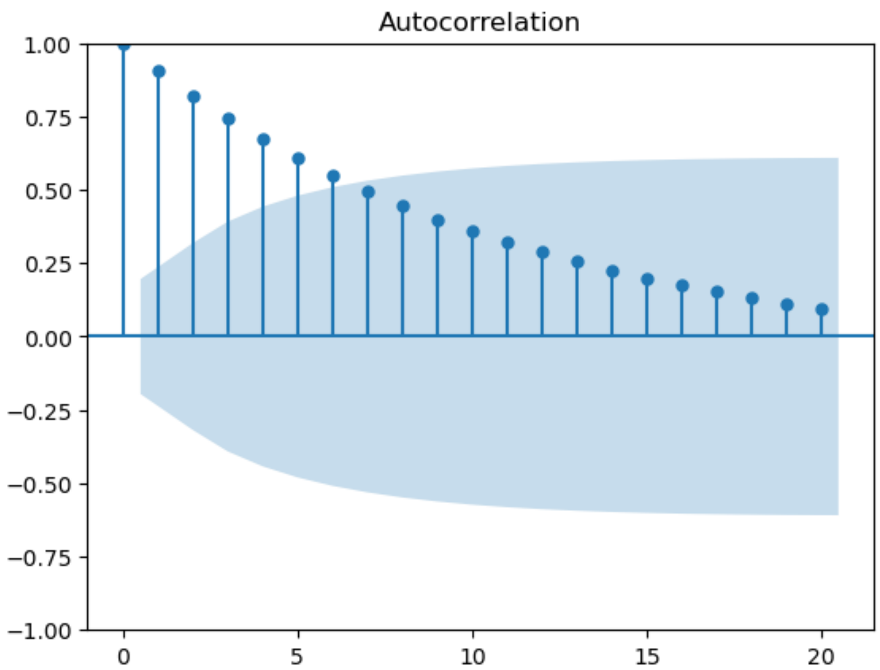
```
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf
from statsmodels.tsa.arima_process import ArmaProcess

# Set phi value
phi = 1.1

# Define AR(1) process
ar1 = ArmaProcess(ar=[1, -phi])

# Generate time series
np.random.seed(123)
x = ar1.generate_sample(nsample=100)

# Plot ACF
plot_acf(x)
plt.show()
```



R code:

```
# Generate AR(1) time series with phi=1.1
set.seed(123)
x <- arima.sim(n = 100, list(ar = 1.1), sd = 1)

# Create ACF plot
acf(x)
```

MATLAB code:

```
% Set phi value
phi = 1.1;

% Define AR(1) process
ar1 = arima('AR', [1, -phi], 'Variance', 1);

% Generate time series
rng(123);
x = simulate(ar1, 100);

% Plot ACF
autocorr(x)
```

Linear Processes

An important class of stationary time series:

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}$$

where $W_t \sim WN(0, \sigma^2), \mu, and\ \psi_j$ are parameters satisfying $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$

So, $E(X_t) = \mu$

$$\gamma_X(h) = \mathrm{Cov}(X_{t+h}, X_t) = \sigma^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{h+j}$$

- **Examples of Linear Processes: White noise**

Choose μ and $\psi_0 = 1\ and\ \psi_j = 0$ for $j \neq 0$.

Then $X_t = \mu + W_t \sim WN(\mu, \sigma^2)$

- **Examples of Linear Processes: MA(1)**

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j W_{t-j}$$

Choose $\mu = 0 \; and \;\; \psi_j = \begin{cases} 1 & if \; j = 0 \\ \theta & if \; j = 1 \\ 0 & others \end{cases}$

Then $X_t = W_t + \theta W_{t-1}$

- **Examples of Linear Processes: AR(1)**

We have $X_t = \phi X_{t-1} + W_t$. For $|\phi| < 1, \;\; X_t = \sum_{j=0}^{\infty} \phi^j W_{t-j}$

Choose $\mu = 0 \; and \;\; \psi_j = \begin{cases} \phi^j & if \; j \geq 0 \\ 0 & others \end{cases}$

➢ Estimating the ACF: Sample ACF

Suppose that $\{X_t\}$ is a **stationary** time series. Recall

- The **mean function** of $\{X_t\}$ is $\mu_t := E[X_t]$

- The **autocovariance function** of of $\{X_t\}$ is

$$\gamma_X(h) := \text{Cov}(X_{t+h}, X_t) = E[(X_{t+h} - \mu)(X_t - \mu)]$$

- The **autocorrelation function** (**ACF**) of $\{X_t\}$ is

$$\rho_X(h) := \frac{\gamma_X(h)}{\gamma_X(0)}$$

For observations $x_1, x_2, \dots, x_n$ of a time series, the **sample mean**

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

- The **sample autocovariance** function is

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}\left(x_{t+|h|} - \bar{x}\right)(x_t - \bar{x}) \qquad \text{for } -n < h < n$$

- The **sample autocorrelation (ACF)** function is

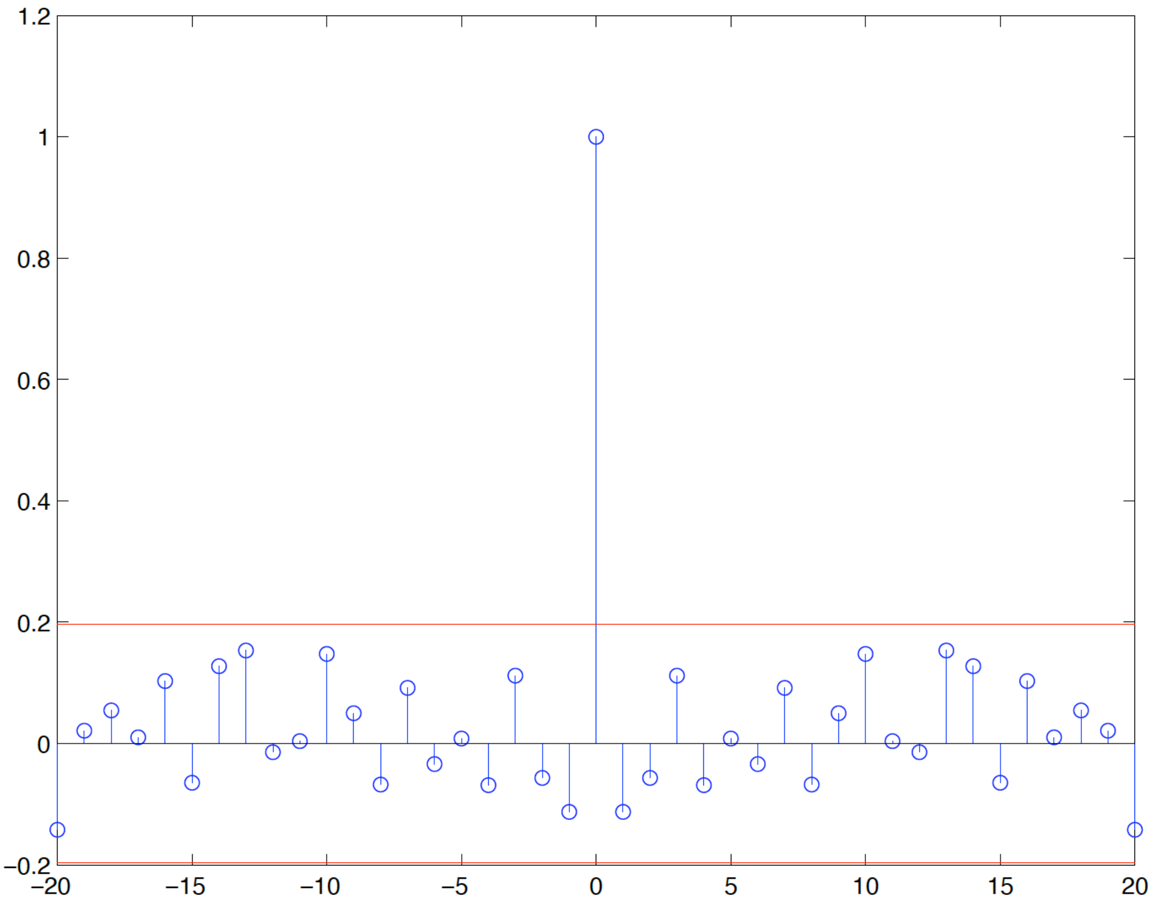
$$\hat{\rho}(h) := \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

Sample autocovariance function ≈ the sample covariance of $(x_1, x_{h+1}), \ldots, (x_{n-h}, x_n)$

Except:

- we normalize by n instead of $n - h$, and
- we subtract the full sample mean.

Sample ACF $\widehat{\rho}(h)$ for white Gaussian (hence i.i.d.) noise



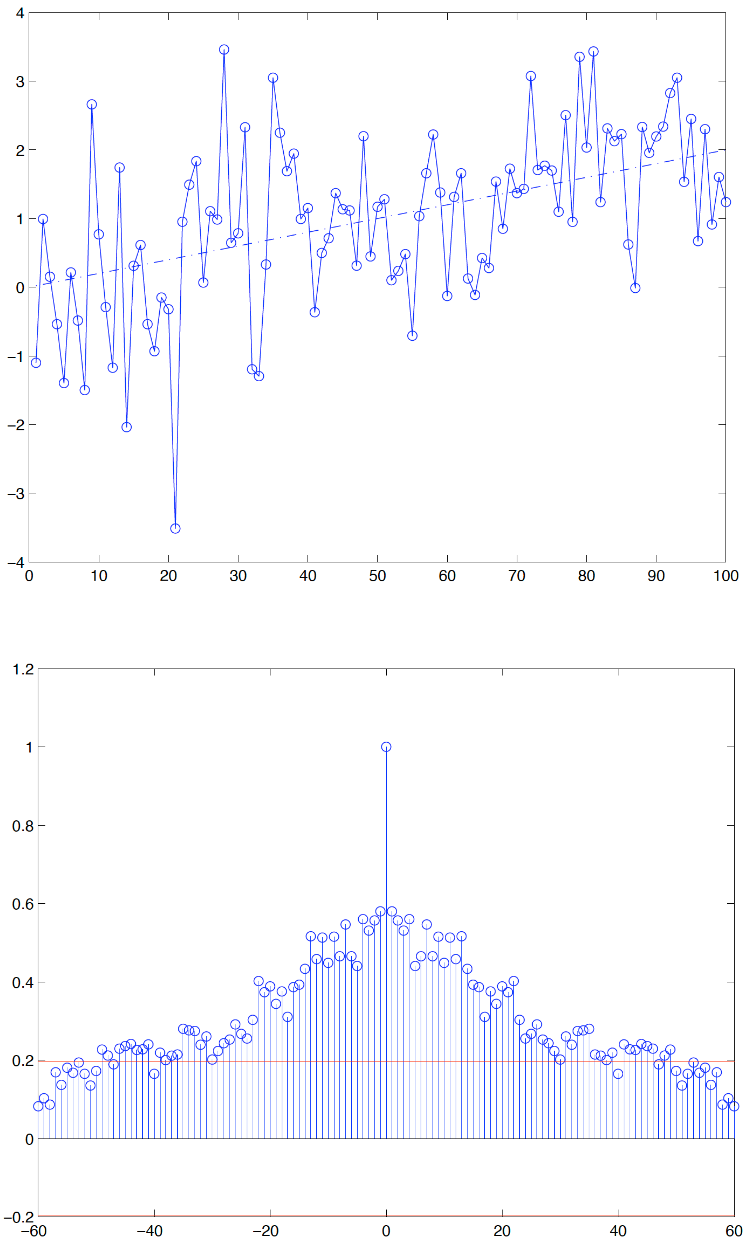
Red line: confidence interval

Sample ACF

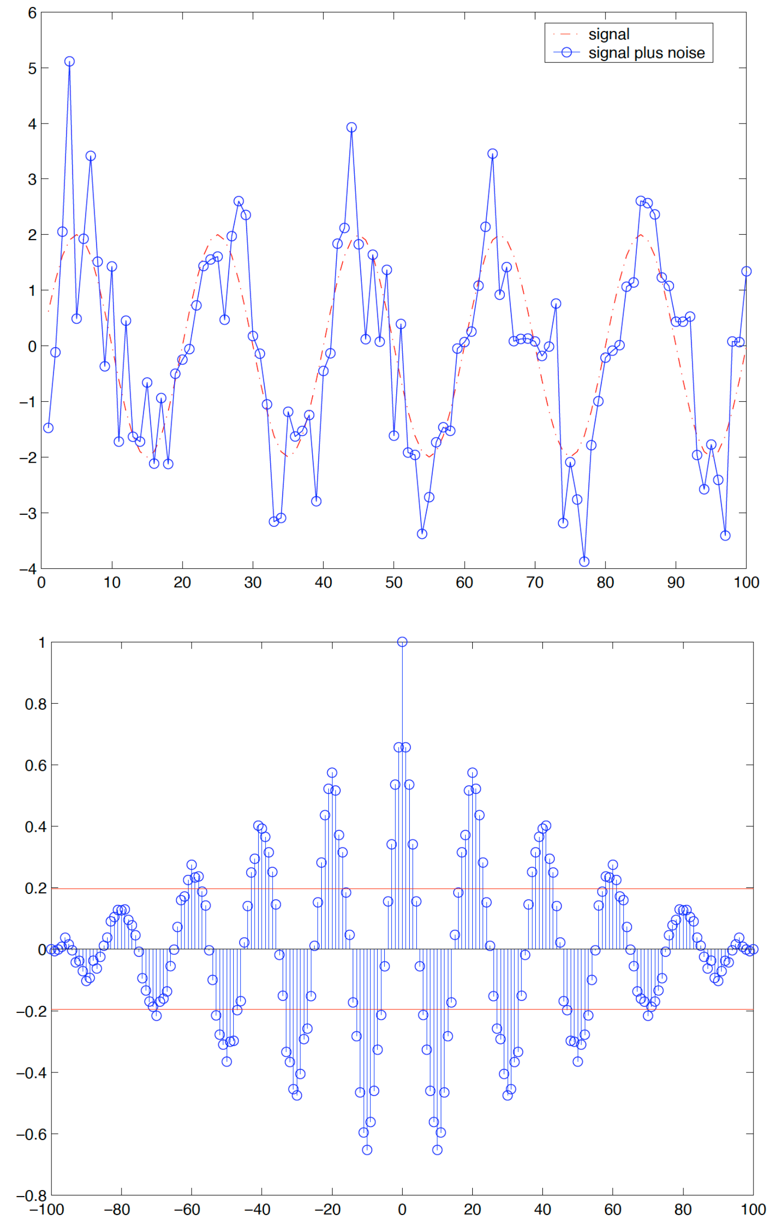
We can recognize the sample autocorrelation functions of many non-white (even non-stationary) time series.

Time Series:	**Sample ACF:**
White	Zero
Trend	Slow decay
Periodic	Periodic
MA(q)	Zero for $\lvert h \rvert > q$
AR(p)	Decays to zero exponentially

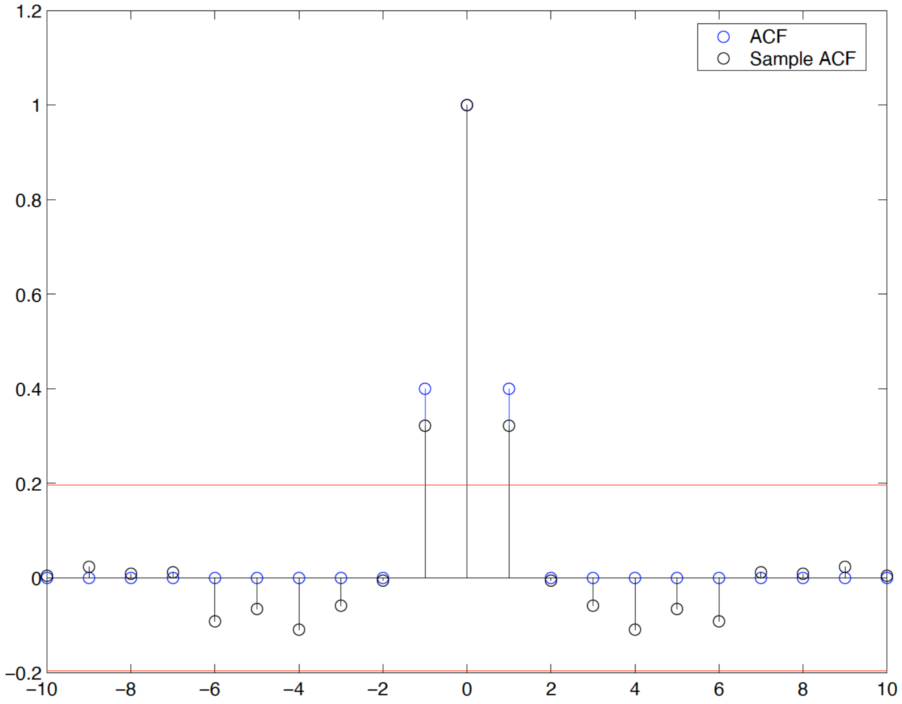
Sample ACF: Trend



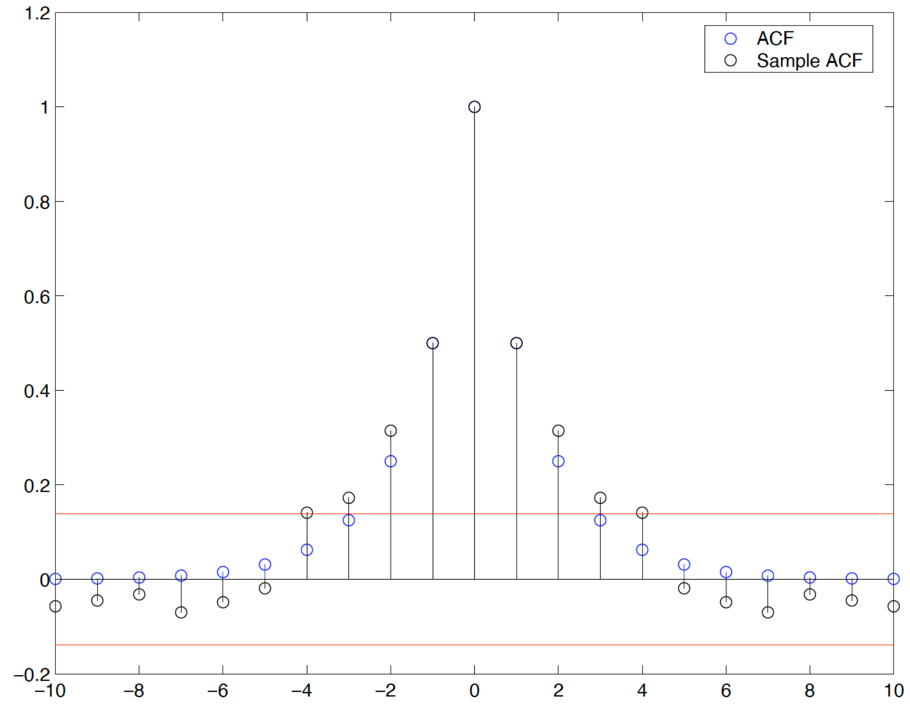
Sample ACF: Periodic



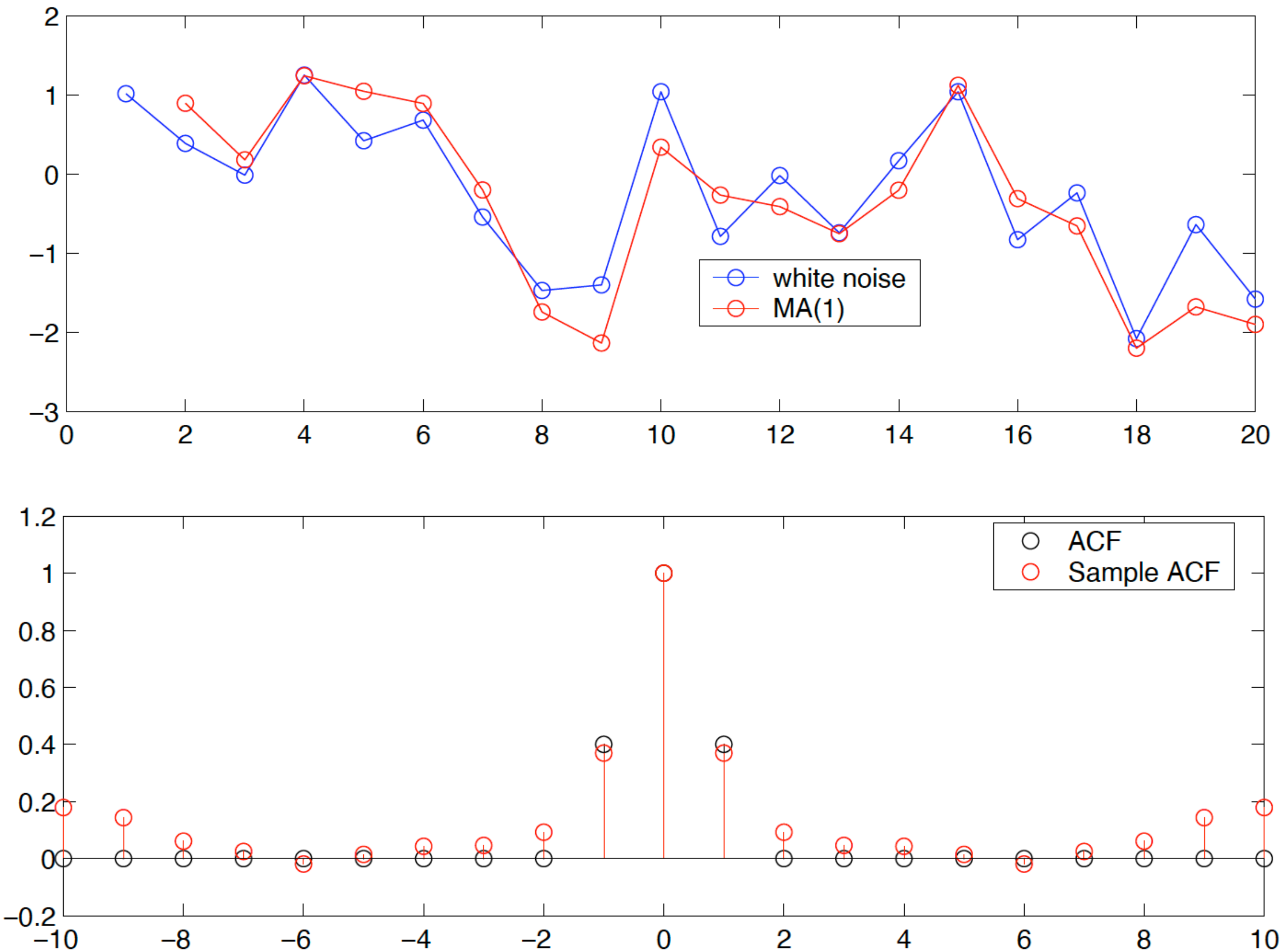
Sample ACF: MA(1)



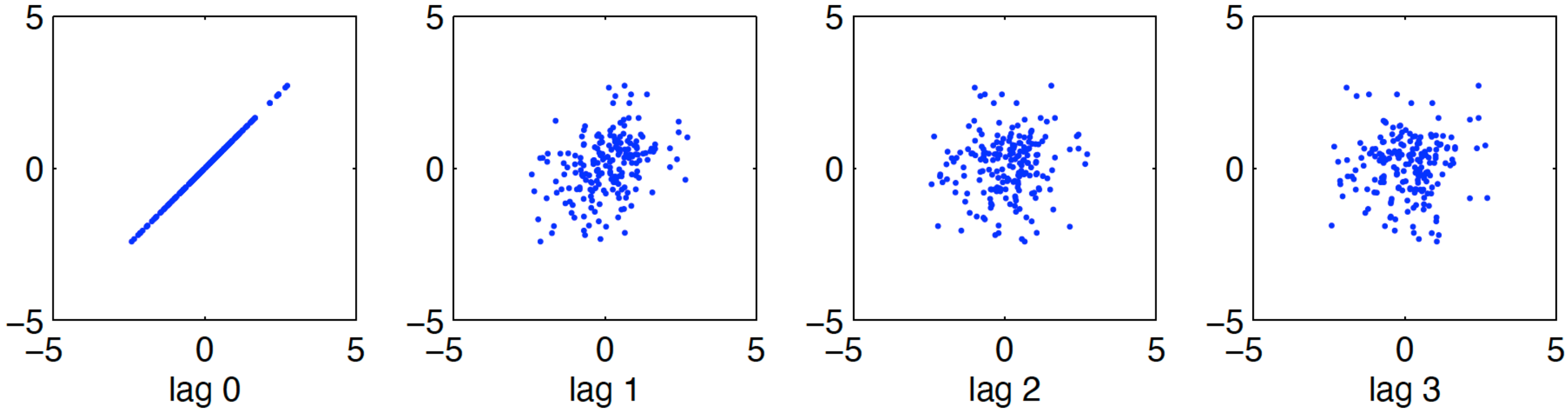
Sample ACF: AR(1)



ACF and prediction



ACF of a MA(1) process



Properties of the **autocovariance function** of a **stationary** time series:

1. $\gamma(h) = \gamma(-h)$
2. γ is positive semidefinite.

A function $f\colon \mathbb{Z} \to \mathbb{R}$ is **positive semidefinite** if for all n, the matrix F_n, with entries $(F_n)_{ij} := f(i-j)$, is positive semidefinite.

In particular, from positive semidefinite, we have $\gamma(0) \geq 0$ and $|\gamma(h)| \leq \gamma(0)$

Furthermore, any function $f\colon \mathbb{Z} \to \mathbb{R}$ that satisfies above two properties is the autocovariance of some stationary time series (in particular, a Gaussian process).

❑ ACF and least squares prediction

Best least squares estimate of *Y* is $E[Y]$ with minimum value $Var(Y)$.

$$E[Y] = \underset{c}{\operatorname{argmin}}\, E(Y - c)^2$$

Proved in section 1.

Best least squares estimate of $Y|X$ is $E[Y|X]$

$$E[Y|X] = \underset{f}{\operatorname{argmin}}\, E\left[E\big(Y - f(X)\big)^2 |X\right]$$

The minimum value is $Var(Y|X)$

Similarly, the **best least squares estimate of** $\boldsymbol{X_{n+h}}$ **given** $\boldsymbol{X_n}$ is

$$f(X_n) = E[X_{n+h}\,|X_n]$$

Example: (stationary Gaussian)

Suppose $\vec{X} = (X_1, \ldots, X_{n+h})$ is the Gaussian distribution:

$$\vec{X} \sim \text{Normal}(\vec{\mu}, \Sigma)$$

Then the joint distribution of (X_n, X_{n+h}) is normal

$$(X_n, X_{n+h}) \sim \text{Normal}\left(\begin{pmatrix} \mu_n \\ \mu_{n+h} \end{pmatrix}, \begin{pmatrix} \sigma_n^2 & \rho\sigma_n\sigma_{n+h} \\ \rho\sigma_n\sigma_{n+h} & \sigma_{n+h}^2 \end{pmatrix} \right)$$

Here, $\rho = Cor(X_n, X_{n+h})$ is the autocorrelation.

The conditional distribution of X_{n+h} given X_n is normal with

Mean: $\mu_{n+h} + \rho \frac{\sigma_{n+h}}{\sigma_n}(x_n - \mu_n)$

Variance: $\sigma_{n+h}^2(1 - \rho^2)$

So, for **stationary** Gaussian, the best **estimate** of X_{n+h} given $X_n = x_n$ is

$$f(x_n) = E[X_{n+h} \,|X_n = x_n] = \mu + \rho(h)(x_n - \mu)$$

The mean square error is

$$Var[X_{n+h} \,|X_n = x_n] = \sigma^2(1 - \rho(h)^2)$$

Note: 1. Prediction accuracy improves as $|\rho(h)| \to 1$

2. Predictor is linear: $f(x) = \mu\big(1 - \rho(h)\big) + \rho(h)x$

➢ ACF and least squares linear prediction

Assume that $\{X_t\}$ is **stationary** with $E(X_t) = 0$.

Consider a linear predictor of X_{n+h} given $X_n = x_n$, with $f(x_n) = ax_n$

The best linear predictor minimizes

$$E\left(X_{n+h} - f(X_n)\right)^2 = E(X_{n+h} - aX_n)^2$$

$$= E\left(X_{n+h}^2\right) - E\left(2aX_nX_{n+h}\right) + E(a^2X_n^2)$$

$$= \sigma^2 - 2a\rho(h) + a^2\sigma^2$$

So, $\rho(h) = \underset{a}{\operatorname{argmin}}\, E\left(X_{n+h} - f(X_n)\right)^2$

That is, the optimal linear predictor is $f(x_n) = \rho(h)\, x_n$

The mean square error is $E\left(X_{n+h} - f(X_n)\right)^2 = \sigma^2\left(1 - \rho(h)\right)^2$

Assume that $\{X_t\}$ is **stationary** with $E(X_t) = \mu$.

Consider a linear predictor of X_{n+h} given $X_n = x_n$, with $f(x_n) = a(x_n - \mu) + b$

The best linear predictor minimizes

$$E\big(X_{n+h} - f(X_n)\big)^2 = E(X_{n+h} - a(X_n - \mu) - b)^2$$

is the optimal linear predictor is $f(x_n) = \rho(h)(x_n - \mu) + \mu$

The mean square error is $E\big(X_{n+h} - f(X_n)\big)^2 = \sigma^2\big(1 - \rho(h)\big)^2$

Note:

- If $\{X_t\}$ is stationary, f is the optimal linear predictor.
- If $\{X_t\}$ is also Gaussian, f is the optimal predictor. Linear prediction is optimal for Gaussian time series.
- Over all stationary processes with that value of $\rho(h)$ and σ^2, the optimal
- mean squared error is maximized by the Gaussian process.
- Linear prediction needs only second order statistics.