Introduction to Time Series Analysis. Lecture 3.

Peter Bartlett

Last lecture:

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

Introduction to Time Series Analysis. Lecture 3.

Peter Bartlett

1. Review: Autocovariance, linear processes
2. Sample autocorrelation function
3. ACF and prediction
4. Properties of the ACF

Mean, Autocovariance, Stationarity

A time series $\{X_t\}$ has **mean function** $\mu_t = \mathrm{E}[X_t]$ and **autocovariance function**

$$\begin{aligned}\gamma_X(t+h,t) &= \mathrm{Cov}(X_{t+h}, X_t) \\ &= \mathrm{E}[(X_{t+h} - \mu_{t+h})(X_t - \mu_t)].\end{aligned}$$

It is **stationary** if both are independent of t.
Then we write $\gamma_X(h) = \gamma_X(h, 0)$.
The **autocorrelation function (ACF)** is

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

Linear Processes

An important class of stationary time series:

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

where $\{W_t\} \sim WN(0, \sigma_w^2)$

and μ, ψ_j are parameters satisfying

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

Linear Processes

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

Examples:

- White noise: $\psi_0 = 1$.
- MA(1): $\psi_0 = 1$, $\psi_1 = \theta$.
- AR(1): $\psi_0 = 1$, $\psi_1 = \phi$, $\psi_2 = \phi^2$, ...

Estimating the ACF: Sample ACF

Recall:

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

Its **mean** is

$$\mu = \mathrm{E}[X_t].$$

Its **autocovariance function** is

$$\begin{aligned}\gamma(h) &= \mathrm{Cov}(X_{t+h}, X_t) \\ &= \mathrm{E}[(X_{t+h} - \mu)(X_t - \mu)].\end{aligned}$$

Its **autocorrelation function** is

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

Estimating the ACF: Sample ACF

For observations $x_1, \ldots, x_n$ of a time series,

the **sample mean** is $$\bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t.$$

The **sample autocovariance function** is

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

The **sample autocorrelation function** is

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

Estimating the ACF: Sample ACF

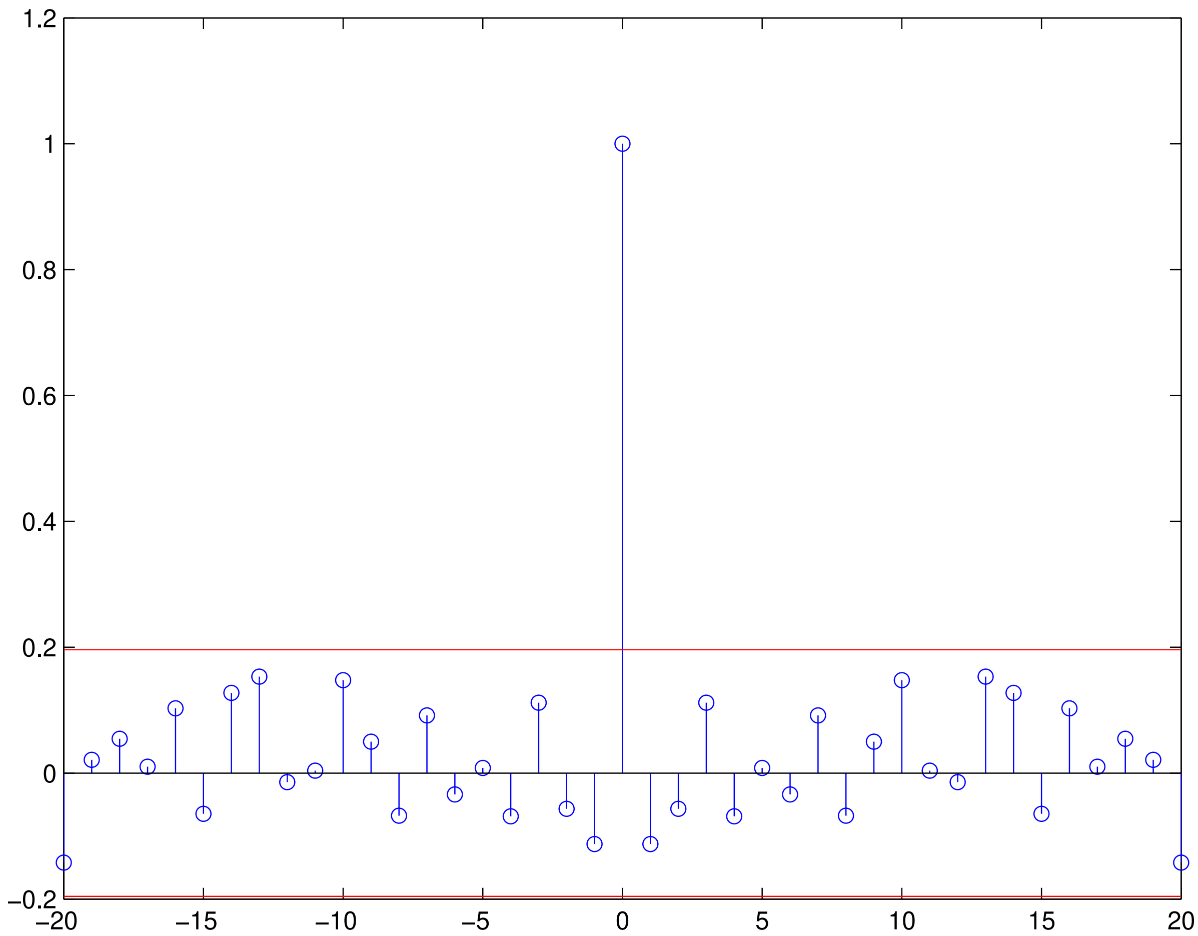
Sample autocovariance function:

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}).$$

$\approx$ the sample covariance of $(x_1, x_{h+1}), \ldots, (x_{n-h}, x_n)$, except that

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

Sample ACF for white Gaussian (hence i.i.d.) noise



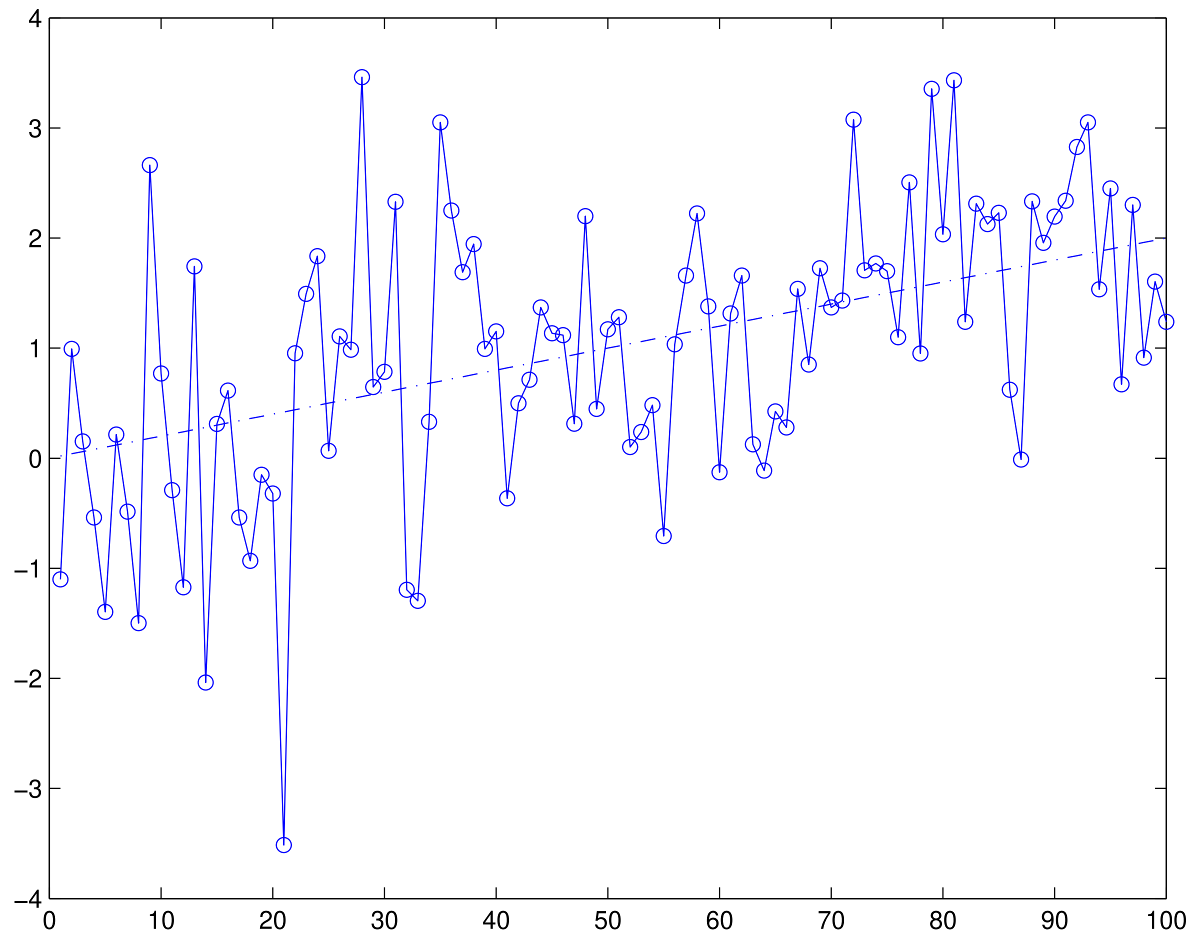
Red lines=c.i.

Sample ACF

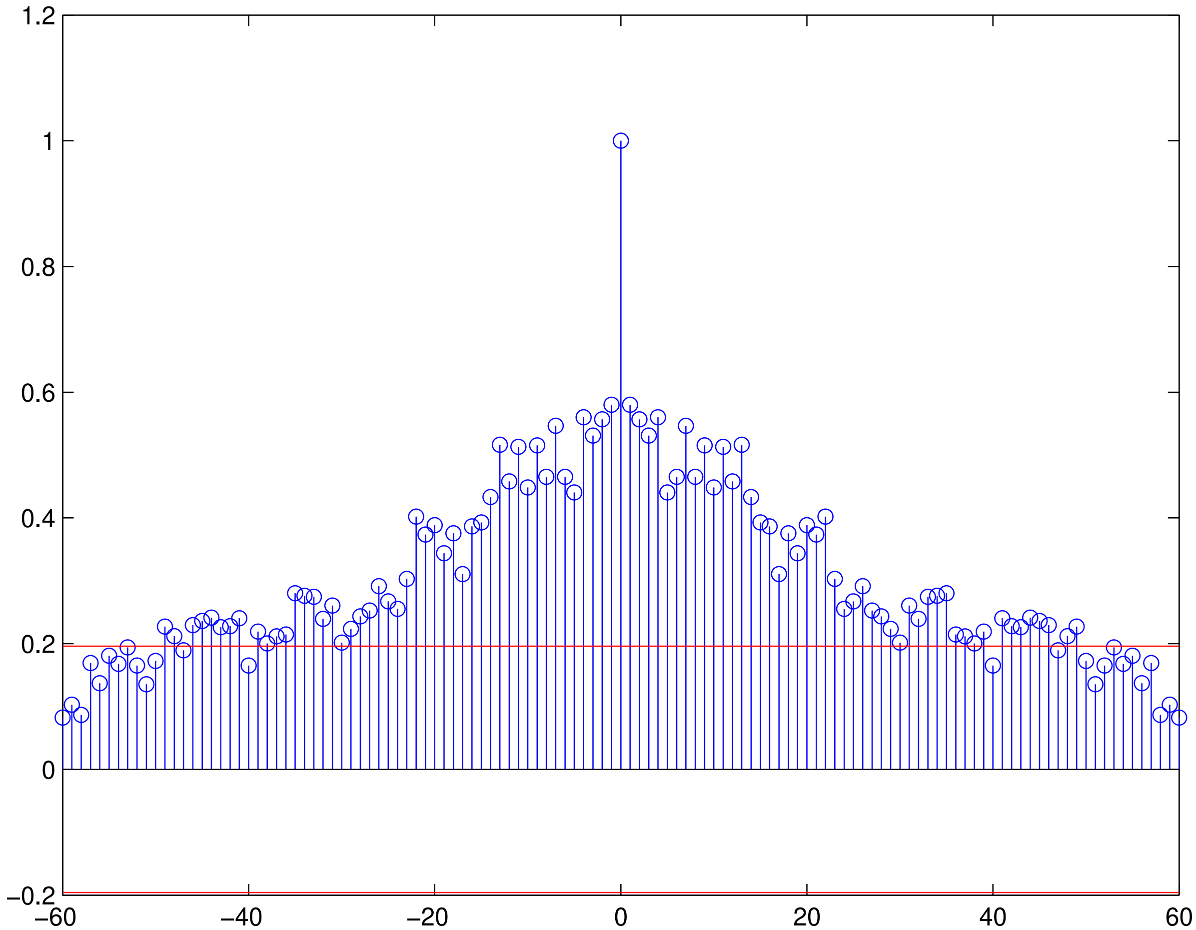
We can recognize the sample autocorrelation functions of many non-white (even non-stationary) time series.

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

Sample ACF: Trend



Sample ACF: Trend

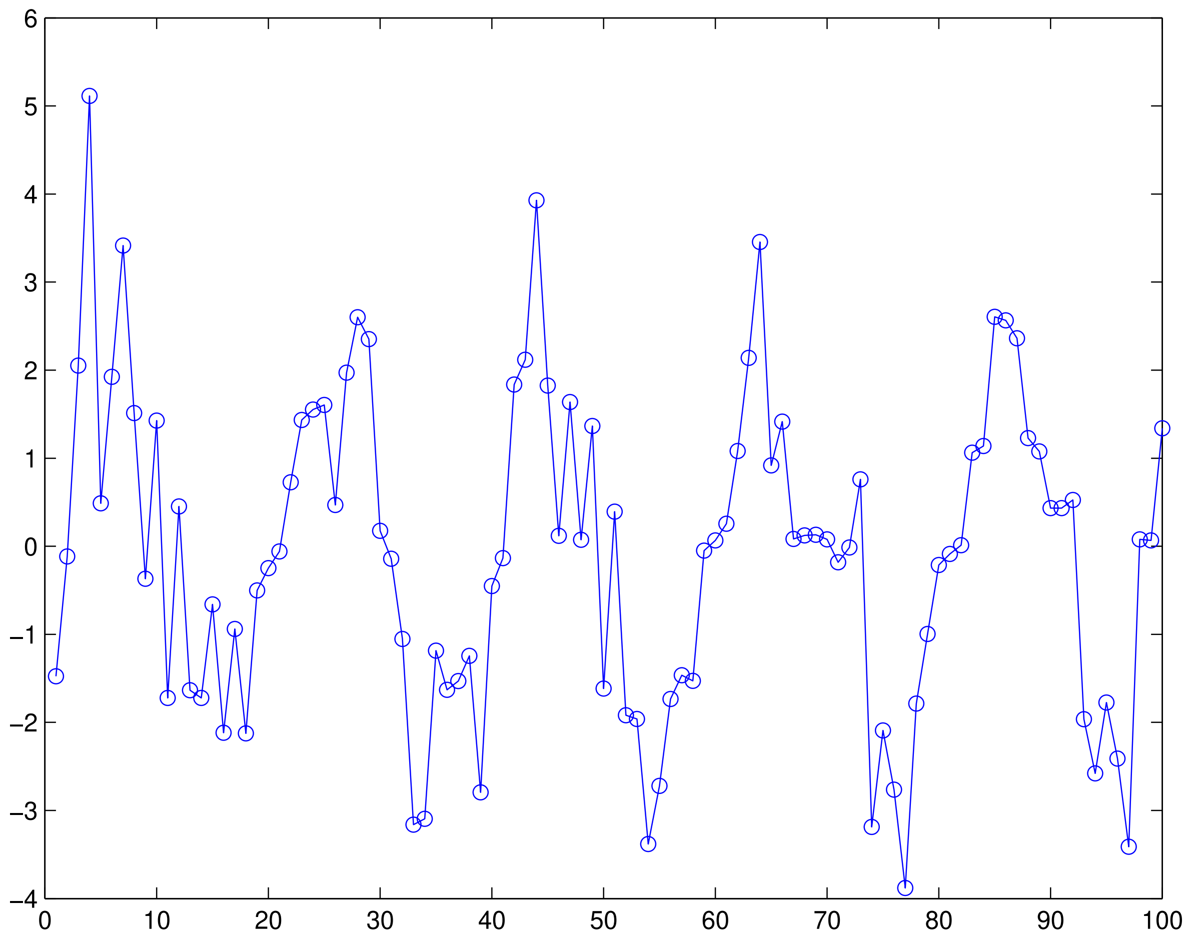


(why?)

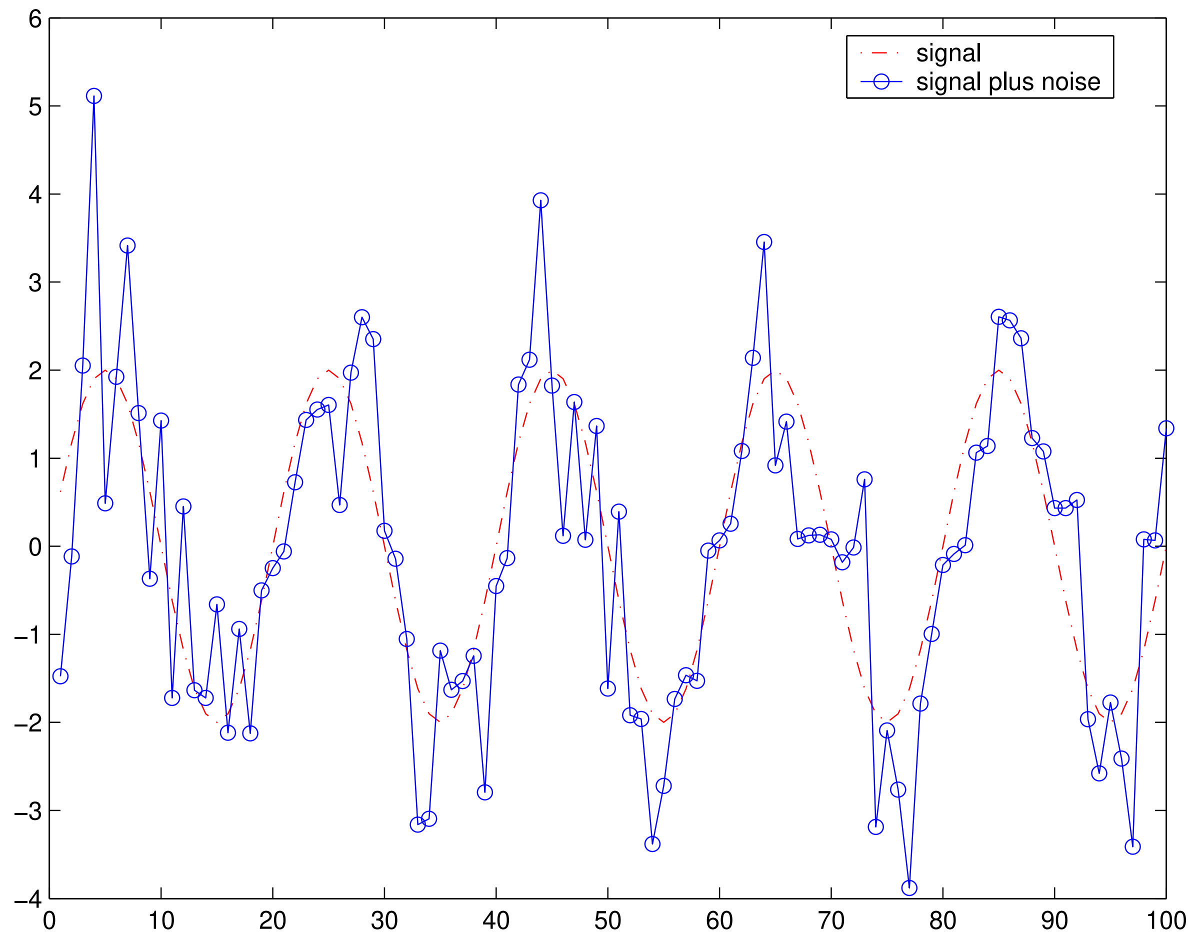
Sample ACF

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

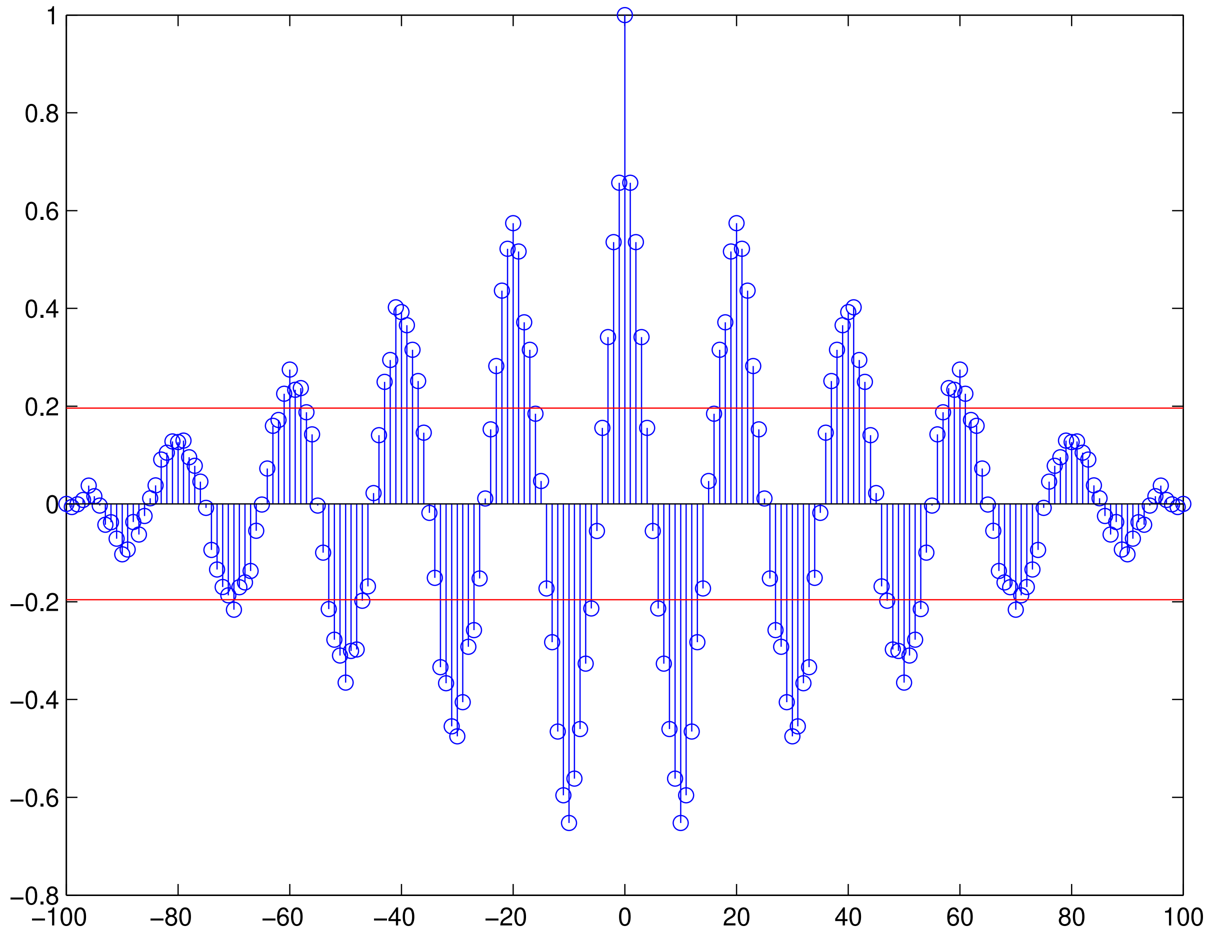
Sample ACF: Periodic



Sample ACF: Periodic



Sample ACF: Periodic

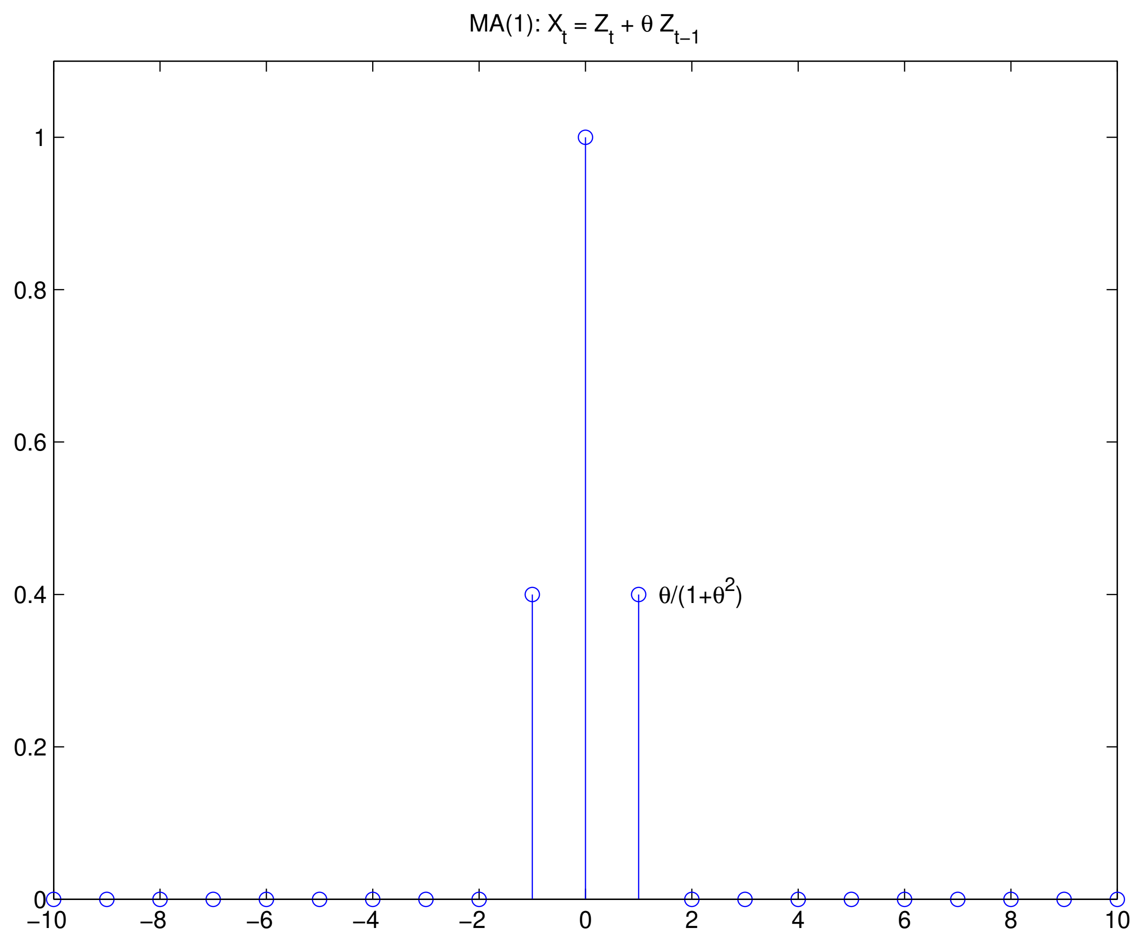


(why?)

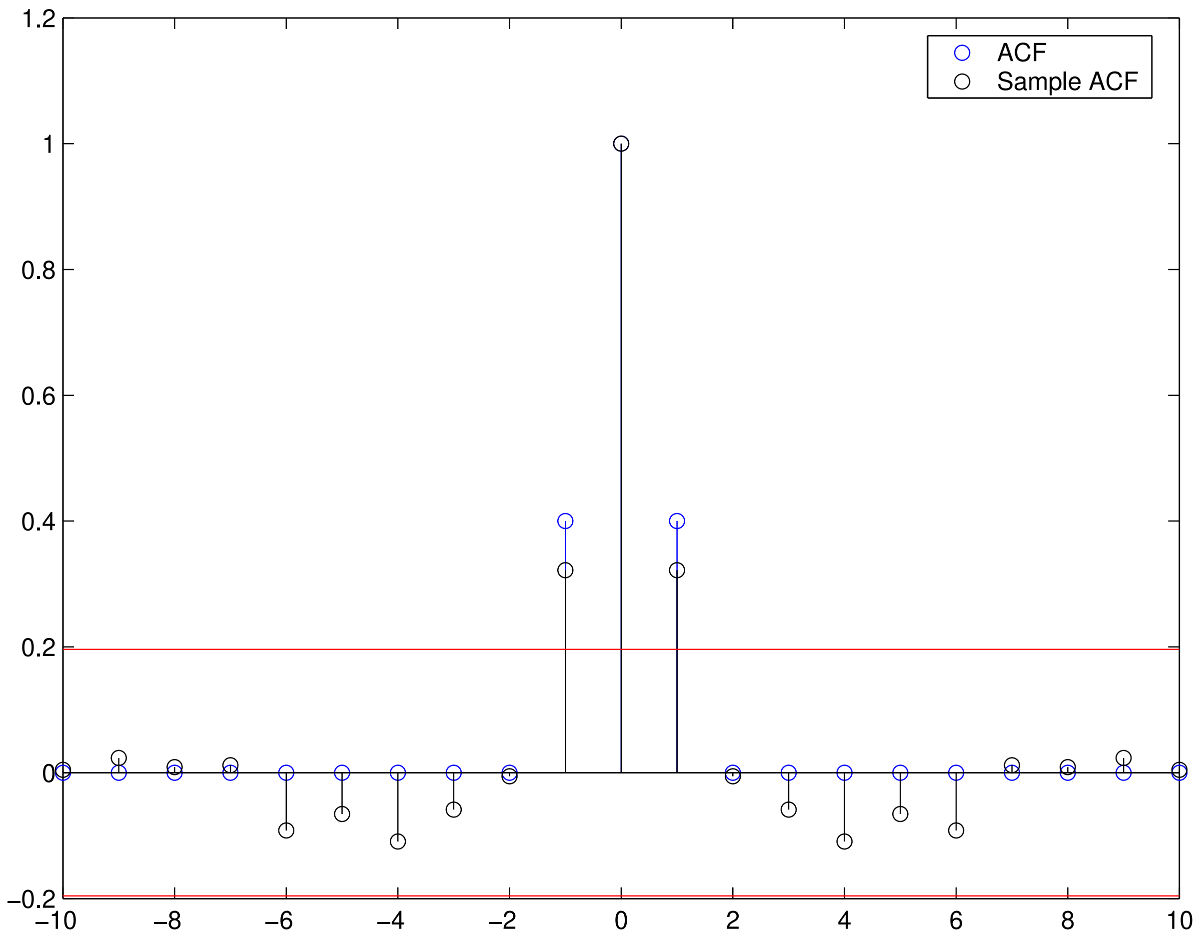
Sample ACF

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

ACF: MA(1)



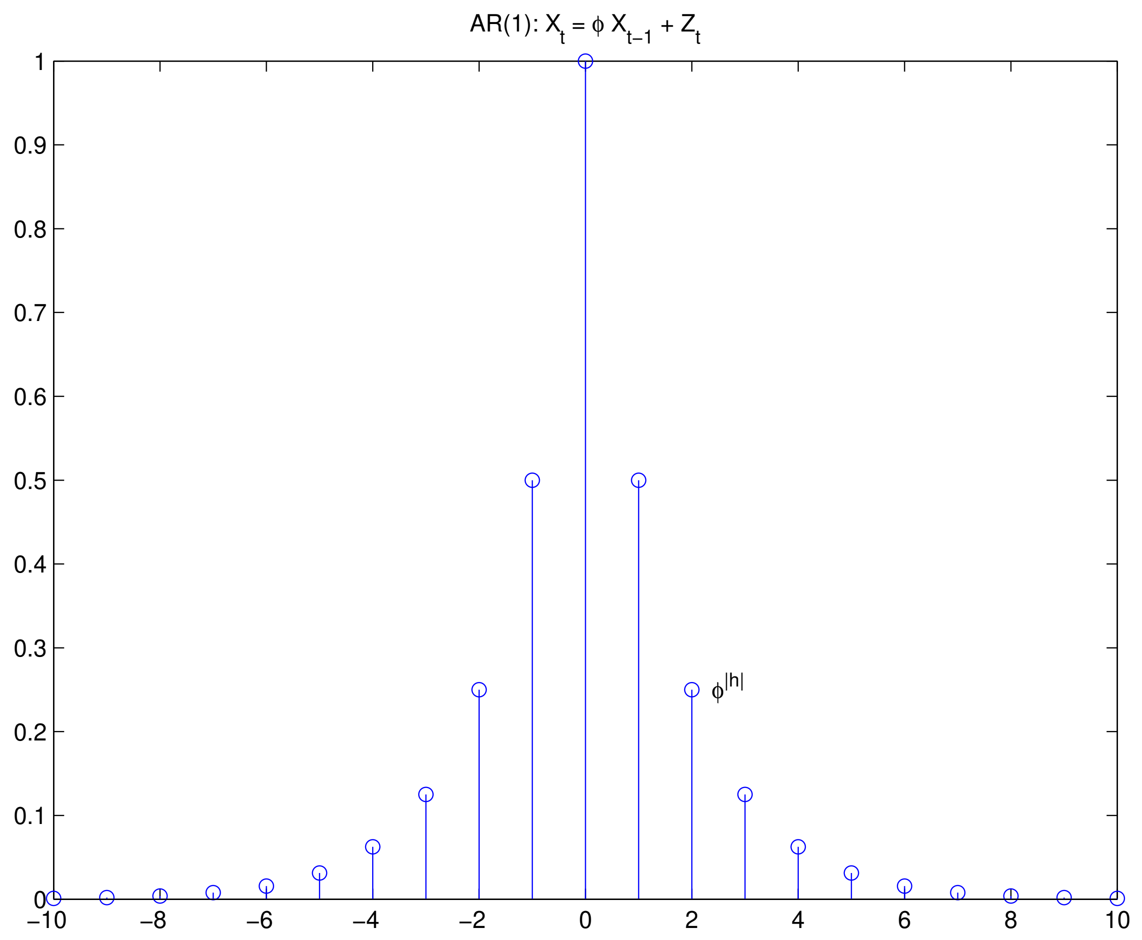
Sample ACF: MA(1)



Sample ACF

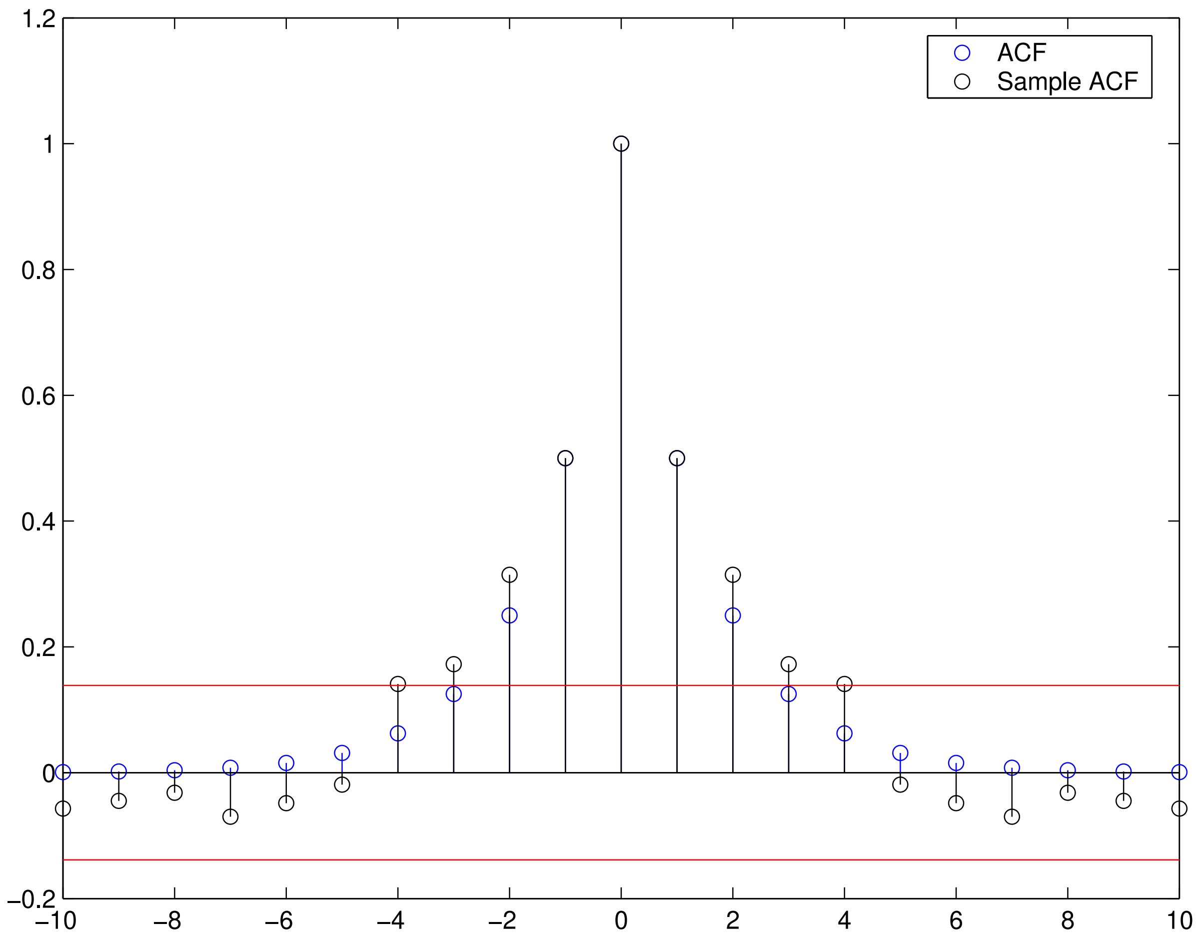
Time series:	**Sample ACF:**
White	zero
Trend	Slow decay
Periodic	Periodic
MA(q)	Zero for $\|h\| > q$
AR(p)	Decays to zero exponentially

ACF: AR(1)



AR(1): $X_t = \phi X_{t-1} + Z_t$
$\phi^{|h|}$
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
−10
−8
−6
−4
−2
0
2
4
6
8
10

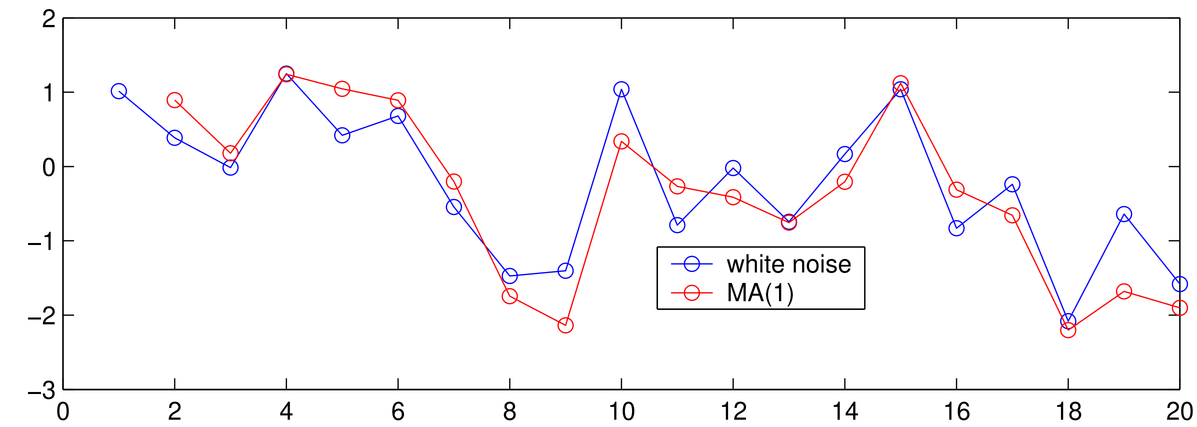
Sample ACF: AR(1)

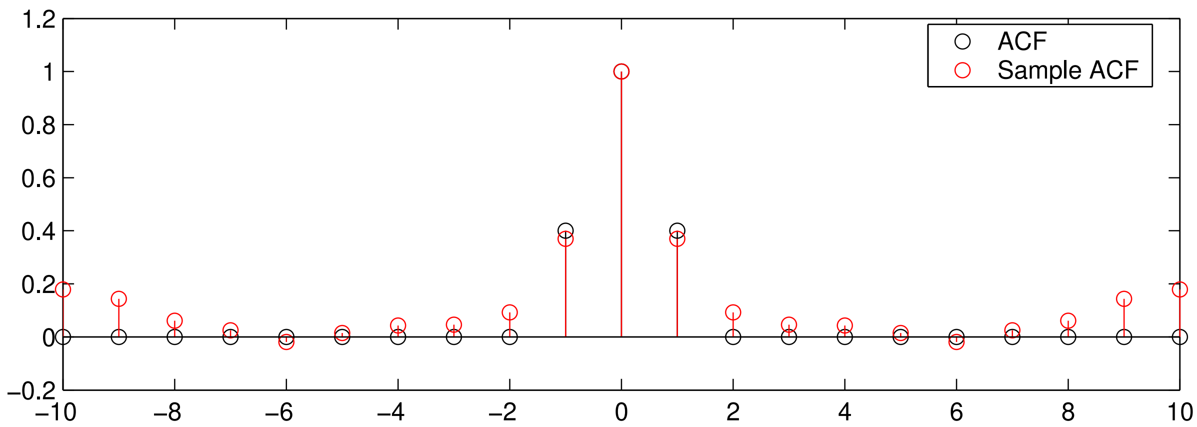


Introduction to Time Series Analysis. Lecture 3.

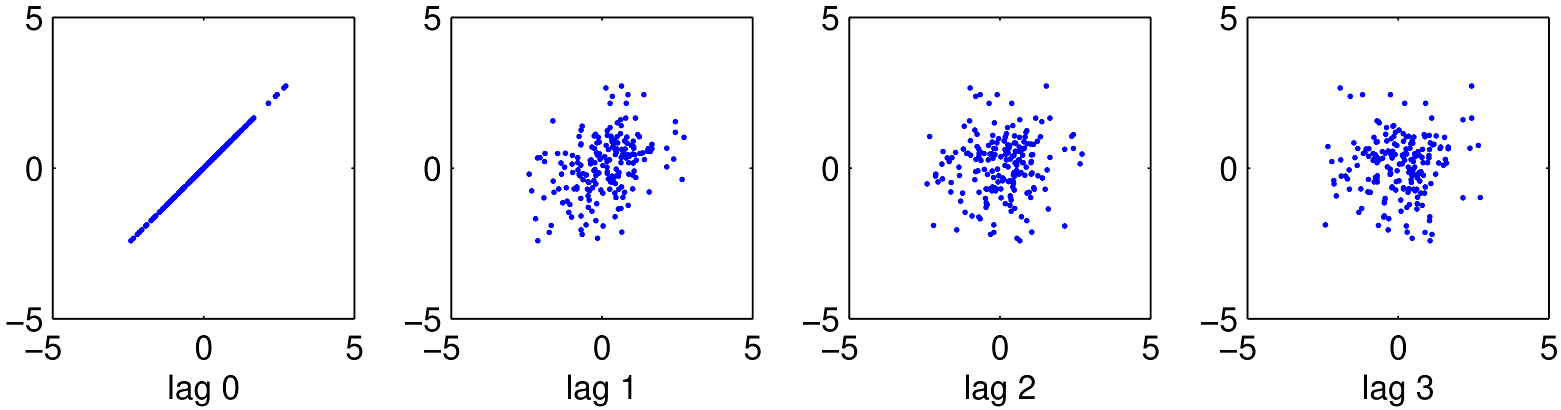
1. Sample autocorrelation function
2. ACF and prediction
3. Properties of the ACF

ACF and prediction





ACF of a MA(1) process



ACF and least squares prediction

Best least squares estimate of Y is $\mathrm{E}Y$:

$$\min_c \mathrm{E}(Y-c)^2 = \mathrm{E}(Y-\mathrm{E}Y)^2.$$

Best least squares estimate of Y given X is $\mathrm{E}[Y|X]$:

$$\begin{aligned}
\min_f \mathrm{E}(Y-f(X))^2 &= \min_f \mathrm{E}\left[\mathrm{E}[(Y-f(X))^2|X]\right] \\
&= \mathrm{E}\left[\mathrm{E}[(Y-\mathrm{E}[Y|X])^2|X]\right] \\
&= \mathrm{var}[Y|X].
\end{aligned}$$

Similarly, the best least squares estimate of X_{n+h} given X_n is $f(X_n) = \mathrm{E}[X_{n+h}|X_n]$.

ACF and least squares prediction

Suppose that $X = (X_1, \ldots, X_{n+h})$ is jointly Gaussian:

$$f_X(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right).$$

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

$$N\left(\left(\begin{array}{c} \mu_n \\ \mu_{n+h} \end{array}\right), \left(\begin{array}{cc} \sigma_n^2 & \rho\sigma_n\sigma_{n+h} \\ \rho\sigma_n\sigma_{n+h} & \sigma_{n+h}^2 \end{array}\right)\right),$$

and the conditional distribution of X_{n+h} given X_n is

$$N\left(\mu_{n+h} + \rho\frac{\sigma_{n+h}}{\sigma_n}(x_n - \mu_n), \sigma_{n+h}^2(1-\rho^2)\right).$$

ACF and least squares prediction

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

$$f(x_n) = \mu + \rho(h)(x_n - \mu),$$

and the mean squared error is

$$\mathrm{E}(X_{n+h} - f(X_n))^2 = \sigma^2(1 - \rho(h)^2).$$

Notice:

- Prediction accuracy improves as $|\rho(h)| \to 1$.
- Predictor is linear: $f(x) = \mu(1 - \rho(h)) + \rho(h)x$.

ACF and least squares linear prediction

Consider a **linear predictor** of X_{n+h} given $X_n = x_n$. Assume first that $\{X_t\}$ is stationary with $\mathrm{E}X_n = 0$, and predict X_{n+h} with $f(x_n) = ax_n$. The best linear predictor minimizes

$$\begin{aligned}\mathrm{E}\left(X_{n+h} - aX_n\right)^2 &= \mathrm{E}\left(X_{n+h}^2\right) - \mathrm{E}\left(2aX_{n+h}X_n\right) + \mathrm{E}\left(a^2X_n^2\right) \\ &= \sigma^2 - 2a\gamma(h) + a^2\sigma^2,\end{aligned}$$

and this is minimized when $a = \rho(h)$, that is,

$$f(x_n) = \rho(h)X_n.$$

For this optimal linear predictor, the mean squared error is

$$\begin{aligned}\mathrm{E}(X_{n+h} - f(X_n))^2 &= \sigma^2 - 2\rho(h)\gamma(h) + \rho(h)^2\sigma^2 \\ &= \sigma^2(1 - \rho(h)^2).\end{aligned}$$

ACF and least squares linear prediction

Consider the following **linear predictor** of X_{n+h} given $X_n = x_n$, when $\{X_n\}$ is stationary and $\mathrm{E}X_n = \mu$:

$$f(x_n) = a(x_n - \mu) + b.$$

The linear predictor that minimizes

$$\mathrm{E}\left(X_{n+h} - (a(X_n - \mu) + b)\right)^2$$

has $a = \rho(h)$, $b = \mu$, that is,

$$f(x_n) = \rho(h)(X_n - \mu) + \mu.$$

For this optimal linear predictor, the mean squared error is again

$$\mathrm{E}(X_{n+h} - f(X_n))^2 = \sigma^2(1 - \rho(h)^2).$$

Least squares prediction of X_{n+h} given X_n

$$f(X_n) = \mu + \rho(h)(X_n - \mu).$$

$$\mathrm{E}(f(X_n) - X_{n+h})^2 = \sigma^2(1 - \rho(h)^2).$$

- 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.
- Extends to longer histories, $(X_n, X_n - 1, \ldots)$.

Introduction to Time Series Analysis. Lecture 3.

1. Sample autocorrelation function
2. ACF and prediction
3. Properties of the ACF

Properties of the autocovariance function

For the autocovariance function γ of a stationary time series $\{X_t\}$,

1. $\gamma(0) \geq 0$, (variance is non-negative)
2. $|\gamma(h)| \leq \gamma(0)$, (from Cauchy-Schwarz)
3. $\gamma(h) = \gamma(-h)$, (from stationarity)
4. γ is positive semidefinite.

Furthermore, any function $\gamma : \mathbb{Z} \to \mathbb{R}$ that satisfies (3) and (4) is the autocovariance of some stationary time series.

Properties of the autocovariance function

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

A matrix $F_n \in \mathbb{R}^{n \times n}$ is positive semidefinite if, for all vectors $a \in \mathbb{R}^n$,

$$a'Fa \geq 0.$$

To see that γ is psd, consider the variance of $(X_1, \ldots, X_n)a$.

Properties of the autocovariance function

For the autocovariance function γ of a stationary time series $\{X_t\}$,

1. $\gamma(0) \geq 0$,
2. $|\gamma(h)| \leq \gamma(0)$,
3. $\gamma(h) = \gamma(-h)$,
4. γ is positive semidefinite.

Furthermore, any function $\gamma : \mathbb{Z} \to \mathbb{R}$ that satisfies (3) and (4) is the autocovariance of some stationary time series (in particular, a Gaussian process).

e.g.: (1) and (2) follow from (4).

Introduction to Time Series Analysis. Lecture 3.

1. Sample autocorrelation function
2. ACF and prediction
3. Properties of the ACF