Time Series Analysis

Week 4 - Introduction to Stochastic Processes, Operators and Linear Systems

Peder Bacher
Department of Applied Mathematics and Computer Science
Technical University of Denmark

March 21, 2024

Week 4: Outline of the lecture

- Operators; the backward shift operator; sec. 4.5.
- Stochastic processes in general: Sec 5.1, 5.2, 5.3 [except 5.3.2].

Operators; The backwards shift operator B

- An operator A is (here) a function of a time series $\{x_t\}$ (or a stochastic process $\{X_t\}$).
- Application of an operator on a time series $\{x_t\}$ yields a new time series $\{Ax_t\}$. Likewise of a stochastic process $\{AX_t\}$.

Operators; The backwards shift operator B

- An operator A is (here) a function of a time series $\{x_t\}$ (or a stochastic process $\{X_t\}$).
- Application of an operator on a time series $\{x_t\}$ yields a new time series $\{Ax_t\}$. Likewise of a stochastic process $\{AX_t\}$.
- Most important operator for us: The backwards shift operator $B : Bx_t = x_{t-1}$. Notice $B^j x_t = x_{t-j}$.
- All other operators we shall consider in this lecture may be expressed in terms of B.

Question

	$t-3$	$t-2$	$t-1$	t	$t+1$	$t+2$	$t+3$
x_t	0.2	0.1	1.3	0.8	0.4	0.1	0.5
$B^j x_t$			0.2	0.1	1.3	0.8	0.4

What is j?

Question

	$t-3$	$t-2$	$t-1$	t	$t+1$	$t+2$	$t+3$
x_t	0.2	0.1	1.3	0.8	0.4	0.1	0.5
$B^j x_t$	0.1	1.3	0.8	0.4	0.1	0.5	

What is j?

The forward shift F

The forward shift operator

- $Fx_t = x_{t+1}$; $F^j x_t = x_{t+j}$;

The forward shift F

The forward shift operator

- $Fx_t = x_{t+1}$; $F^j x_t = x_{t+j}$; How can F be expressed in terms of B? ($Bx_t = x_{t-1}$)
- Combining a forward and backward shift yields the identity operator; $BFx_t = Bx_{t+1} = x_t$, ie. F and B are each others inverse: $B^{-1} = F$ and $F^{-1} = B$.

Question

	$t-3$	$t-2$	$t-1$	t	$t+1$	$t+2$	$t+3$
x_t	0.2	0.1	1.3	0.8	0.4	0.1	0.5
$F^j x_t$			0.2	0.1	1.3	0.8	0.4

What is j?

Question

	$t-3$	$t-2$	$t-1$	t	$t+1$	$t+2$	$t+3$
x_t	0.2	0.1	1.3	0.8	0.4	0.1	0.5
$F^j x_t$	0.1	1.3	0.8	0.4	0.1	0.5	

What is j?

The difference operator ∇

The difference operator

- $\nabla x_t = x_t - x_{t-1}$

- How can ∇ be expressed with B? A: $\nabla = 1 - B$ or B: $\nabla = 1 + B^{-1}$

The difference operator ∇

The difference operator

- $\nabla x_t = x_t - x_{t-1}$

> - How can ∇ be expressed with B? A: $\nabla = 1 - B$ or B: $\nabla = 1 + B^{-1}$

$= (1 - B)x_t$.

- Thus: $\nabla = 1 - B$.

The summation S

$$
\begin{aligned}
Sx_t &= x_t + x_{t-1} + x_{t-2} + \ldots \\
&= x_t + Bx_t + B^2 x_t \ldots \\
&= (1 + B + B^2 + \ldots)x_t
\end{aligned}
$$

The summation S

$$\begin{aligned} Sx_t &= x_t + x_{t-1} + x_{t-2} + \dots \\ &= x_t + Bx_t + B^2 x_t \dots \\ &= (1 + B + B^2 + \dots)x_t \end{aligned}$$

- Summation, then difference (using $Sx_t = x_t + Sx_{t-1}$)

$$\nabla Sx_t = Sx_t - Sx_{t-1} = x_t + Sx_{t-1} - Sx_{t-1} = x_t$$

The summation S

$$\begin{aligned} Sx_t &= x_t + x_{t-1} + x_{t-2} + \dots \\ &= x_t + Bx_t + B^2 x_t \dots \\ &= (1 + B + B^2 + \dots)x_t \end{aligned}$$

- Summation, then difference (using $Sx_t = x_t + Sx_{t-1}$)
$$\nabla Sx_t = Sx_t - Sx_{t-1} = x_t + Sx_{t-1} - Sx_{t-1} = x_t$$
- Difference, then summation
$$\begin{aligned} S\nabla x_t &= (1 + B + B^2 \dots)x_t - (1 + B + B^2 \dots)x_{t-1} \\ &= (1 + B + B^2 \dots)x_t - (B + B^2 \dots)x_t = x_t \end{aligned}$$
- So ∇ and S are each others inverse:
$$\nabla^{-1} = \frac{1}{1-B} = 1 + B + B^2 + \dots = S$$

In R:

```
x <- 1:10
x
diff(x)
cumsum(x)
# Not so nice in the ends!
cumsum(diff(x))
diff(cumsum(x))
# Put a zero in beginning to fix it
cumsum(diff(c(0,x)))
diff(cumsum(c(0,x)))
```

Shift operator is simply the most important:

```
lagvec <- function(x, lag){
    if (lag > 0) {
        ## Lag x, i.e. delay x lag steps
        return(c(rep(NA, lag), x[1:(length(x) - lag)]))
    }else if(lag < 0) {
        ## Lag x, i.e. delay x lag steps
        return(c(x[(abs(lag) + 1):length(x)], rep(NA, abs(lag))))
    }else{
        ## lag = 0, return x
        return(x)
    }
}
```

Lag to a data frame:

```
lagdf <- function(x, lag) {
  # Lag x, i.e. delay x lag steps
  if (lag > 0) {
    x[(lag+1):nrow(x), ] <- x[1:(nrow(x)-lag), ]
    x[1:lag, ] <- NA
  }else if(lag < 0) {
    # Lag x "ahead in time"
    x[1:(nrow(x)-abs(lag)), ] <- x[(abs(lag)+1):nrow(x), ]
    x[(nrow(x)-abs(lag)+1):nrow(x), ] <- NA
  }
  return(x)
}
```

Properties of B, F, ∇ and S

- The operators are all linear, ie.

$$H[ax_t + by_t] = aH[x_t] + bH[y_t]$$

Properties of B, F, ∇ and S

- The operators are all linear, ie.

$$H[ax_t + by_t] = aH[x_t] + bH[y_t]$$

- The operators may be combined into new operators:
  The power series

$$a(z) = \sum_{i=0}^{\infty} a_i z^i$$

  defines a new operator from an operator H by linear combinations:

$$a(H) = \sum_{i=0}^{\infty} a_i H^i$$

Examples of combined operators

- ∇^{-1}:

$$\frac{1}{1-z} = \sum_{i=0}^{\infty} z^i \quad \text{so} \quad \nabla^{-1} = \frac{1}{1-B} = \sum_{i=0}^{\infty} B^i = S$$

Examples of combined operators

- ∇^{-1}:

$$\frac{1}{1-z} = \sum_{i=0}^{\infty} z^i \quad \text{so} \quad \nabla^{-1} = \frac{1}{1-B} = \sum_{i=0}^{\infty} B^i = S$$

- Operator polynomial of order q:

$$\theta(z) = \sum_{i=0}^{q} \theta_i z^i$$

ie. $\theta_i = 0$ for $i > q$.

$$\theta(B) = (1 + \theta_1 B + \cdots + \theta_q B^q)$$

where θ_0 is chosen to be 1

Stochastic Processes – in general

- Function: $X(t, \omega)$
- Time: $t \in T$
- Realization: $\omega \in \Omega$

- Index set: T
- Sample Space: Ω

- $X(\cdot, \cdot)$ is a stochastic process
- $X(t, \cdot)$ is a random variable

Stochastic Processes – in general

- Function: $X(t, \omega)$
- Time: $t \in T$
- Realization: $\omega \in \Omega$

- Index set: T
- Sample Space: Ω

- $X(\cdot, \cdot)$ is a stochastic process
- $X(t, \cdot)$ is a random variable
- $X(\cdot, \omega)$ is a time series (i.e. *one* realization). This is what we often denote $\{x_t\}$.
- $X(t, \omega)$ is an observation. This is what we often denote x_t.

Stochastic Processes – in general

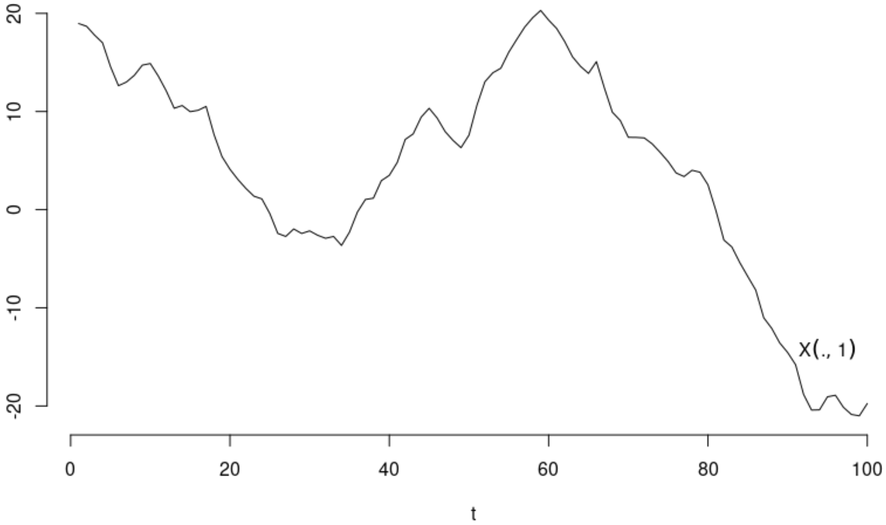
- Function: $X(t, \omega)$
- Time: $t \in T$
- Realization: $\omega \in \Omega$

- Index set: T
- Sample Space: Ω

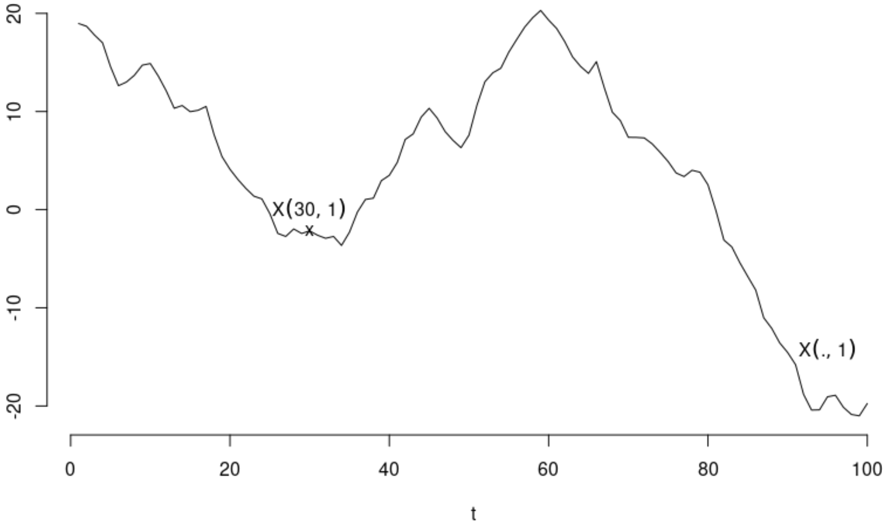
- $X(\cdot, \cdot)$ is a stochastic process
- $X(t, \cdot)$ is a random variable
- $X(\cdot, \omega)$ is a time series (i.e. *one* realization). This is what we often denote $\{x_t\}$.
- $X(t, \omega)$ is an observation. This is what we often denote x_t.

- In this course we restrict ourselves to the case where time is discrete, and the realizations take values on the real numbers (continuous range).

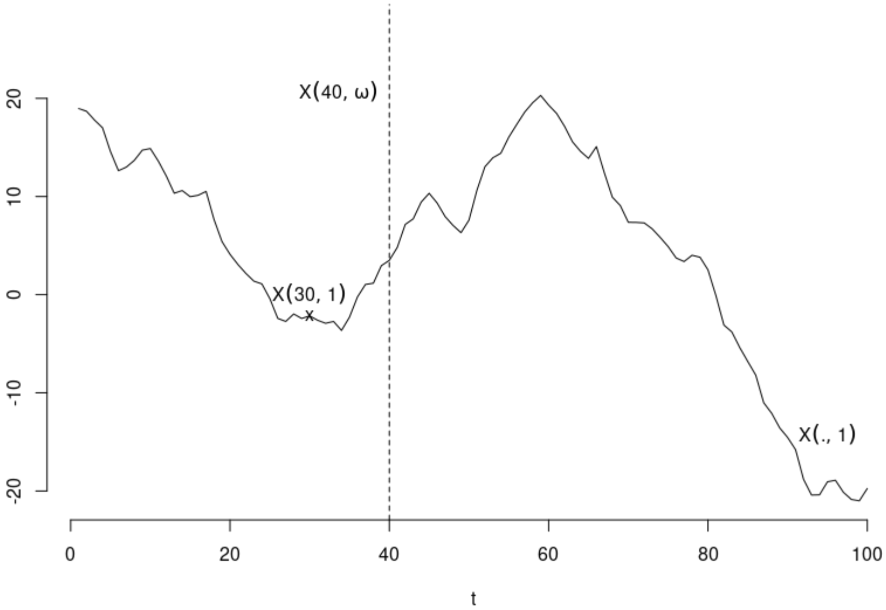
Stochastic Processes – illustration



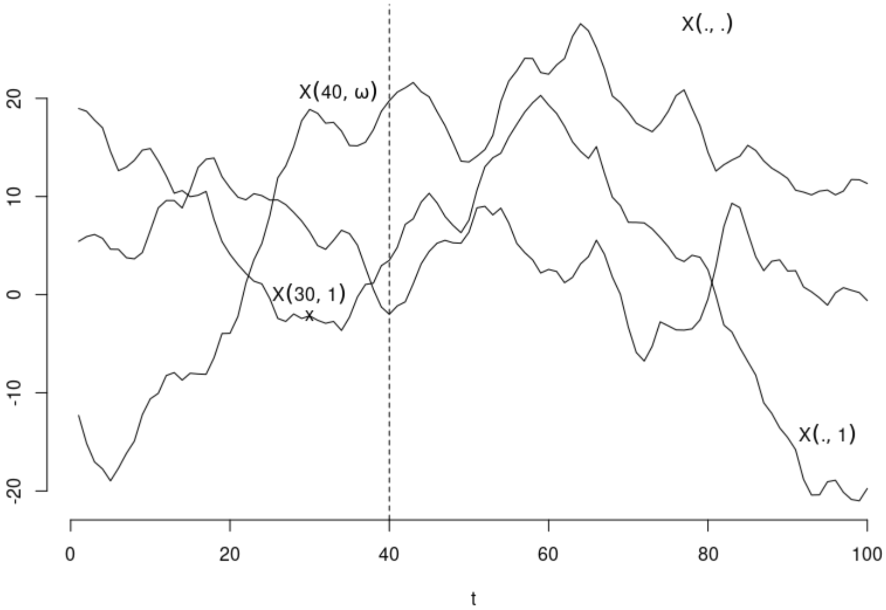
Stochastic Processes – illustration



Stochastic Processes – illustration

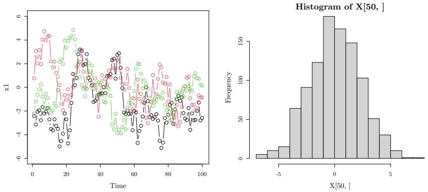


Stochastic Processes – illustration



```
par(mfrow=c(1,2))
# Generate realizations
n <- 100
x1 <- filter(rnorm(n), 0.9, "recursive")
plot(x1, type="b", ylim=c(-6,6))
x2 <- filter(rnorm(n), 0.9, "recursive")
lines(x2, type="b", col=2)
x3 <- filter(rnorm(n), 0.9, "recursive")
lines(x3, type="b", col=3)

# Generate 1000 realizations
X <- matrix(filter(rnorm(1000*n), 0.9, "recursive"), nrow=n)
# One realization (i.e. a time series)
X[ ,1]
# Realization of the stochastic variable at one time point
hist(X[50, ])
```



Complete Characterization

In the end a stochastic process is just a multivariate variable.

Complete Characterization

In the end a stochastic process is just a multivariate variable. Thus, from lecture 1, it is characterised by its n-dimensional probability density:

$$f_{X(t_1),\dots,X(t_n)}(x_1,\dots,x_n)$$

2nd order moment representation

Mean function:

$$\mu(t) = E[X(t)] = \int_{-\infty}^{\infty} x\, f_{X(t)}(x)\, dx,$$

Autocovariance function:

$$\begin{aligned} \gamma_{XX}(t_1, t_2) &= \gamma(t_1, t_2) = \mathrm{Cov}\left[X(t_1), X(t_2)\right] \\ &= E\left[(X(t_1) - \mu(t_1))(X(t_2) - \mu(t_2))\right] \end{aligned}$$

The variance function is obtained from $\gamma(t_1, t_2)$ when $t_1 = t_2 = t$:

$$\sigma^2(t) = V[X(t)] = E\left[(X(t) - \mu(t))^2\right]$$

Stationarity

- A process $\{X(t)\}$ is said to be *strongly stationary* if all finite-dimensional distributions are invariant for changes in time, i.e. for every n, and for any set $(t_1, t_2, \ldots, t_n)$ and for any h it holds

$$f_{X(t_1),\cdots,X(t_n)}(x_1,\cdots,x_n) = f_{X(t_1+h),\cdots,X(t_n+h)}(x_1,\cdots,x_n)$$

Stationarity

- A process $\{X(t)\}$ is said to be *strongly stationary* if all finite-dimensional distributions are invariant for changes in time, i.e. for every n, and for any set $(t_1, t_2, \ldots, t_n)$ and for any h it holds

$$f_{X(t_1),\cdots,X(t_n)}(x_1,\cdots,x_n) = f_{X(t_1+h),\cdots,X(t_n+h)}(x_1,\cdots,x_n)$$

- A process $\{X(t)\}$ is said to be *weakly stationary of order* k if all the first k moments are invariant to changes in time
- A weakly stationary process of order 2 is simply called *weakly stationary* or just *stationary*:

$$\mu(t) = \mu \quad \sigma^2(t) = \sigma^2 \quad \gamma(t_1, t_2) = \gamma(t_1 - t_2)$$

Ergodicity

- In time series analysis we normally assume that we have access to one realization only
- We therefore need to be able to determine characteristics of the process X_t from one realization x_t

Ergodicity

- In time series analysis we normally assume that we have access to one realization only
- We therefore need to be able to determine characteristics of the process X_t from one realization x_t
- It is often enough to require the process to be mean-ergodic:

$$E[X(t)] = \int_\Omega x(t,\omega)f(\omega)\, d\omega = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} x(t,\omega)\, dt$$

i.e. if the *mean of the ensemble* equals the *mean over time*

Some intuitive examples, not directly related to time series:
`http://news.softpedia.com/news/What-is-ergodicity-15686.shtml`

Special processes

- *Normal processes* (also called *Gaussian processes*): All finite-dimensional distribution functions are (multivariate) normal distributions

Special processes

- *Normal processes* (also called *Gaussian processes*): All finite-dimensional distribution functions are (multivariate) normal distributions
- *Markov processes*: The conditional distribution depends only on the latest state of the process:

$$\mathrm{P}\{X(t_n) \leq x | X(t_{n-1}), \cdots, X(t_1)\} = \mathrm{P}\{X(t_n) \leq x | X(t_{n-1})\}$$

Special processes

- *Normal processes* (also called *Gaussian processes*): All finite-dimensional distribution functions are (multivariate) normal distributions
- *Markov processes*: The conditional distribution depends only on the latest state of the process:

$$\mathrm{P}\{X(t_n) \leq x | X(t_{n-1}), \cdots, X(t_1)\} = \mathrm{P}\{X(t_n) \leq x | X(t_{n-1})\}$$

- *Deterministic processes*: Can be predicted without uncertainty from past observations

Special processes

- *Normal processes* (also called *Gaussian processes*): All finite-dimensional distribution functions are (multivariate) normal distributions
- *Markov processes*: The conditional distribution depends only on the latest state of the process:

$$\mathrm{P}\{X(t_n) \leq x | X(t_{n-1}), \cdots, X(t_1)\} = \mathrm{P}\{X(t_n) \leq x | X(t_{n-1})\}$$

- *Deterministic processes*: Can be predicted without uncertainty from past observations
- *Pure stochastic processes*: Can be written as a linear combination of uncorrelated random variables

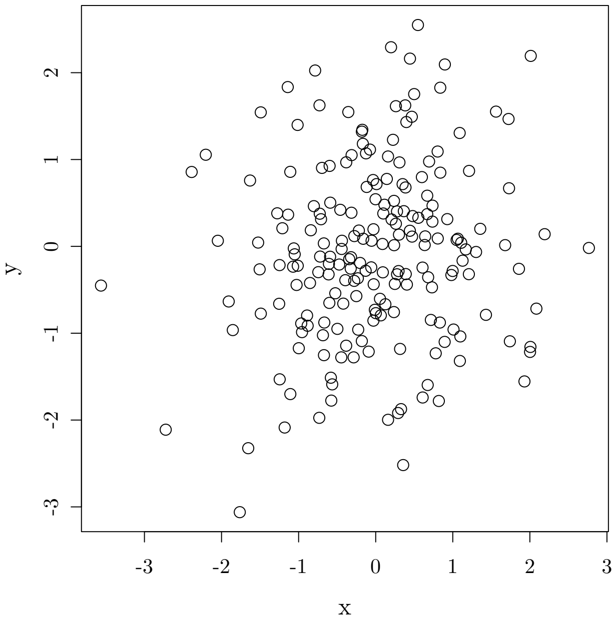
Special processes

- *Normal processes* (also called *Gaussian processes*): All finite-dimensional distribution functions are (multivariate) normal distributions
- *Markov processes*: The conditional distribution depends only on the latest state of the process:

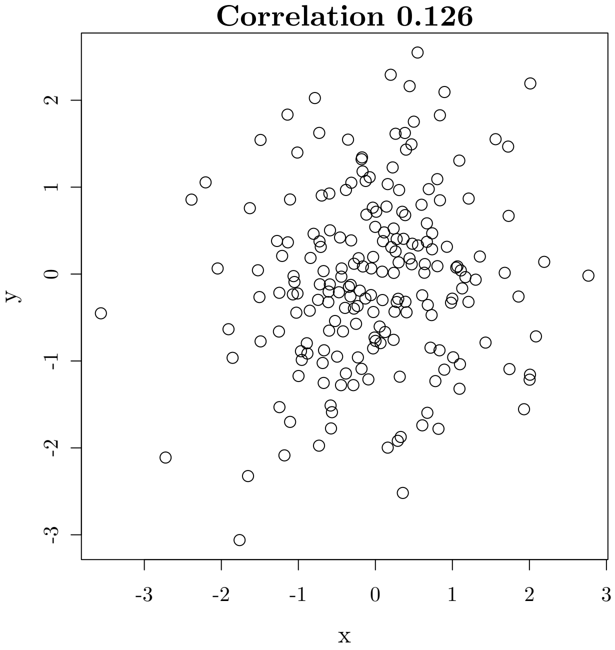
$$\mathrm{P}\{X(t_n) \leq x | X(t_{n-1}), \cdots, X(t_1)\} = \mathrm{P}\{X(t_n) \leq x | X(t_{n-1})\}$$

- *Deterministic processes*: Can be predicted without uncertainty from past observations
- *Pure stochastic processes*: Can be written as a linear combination of uncorrelated random variables
- Decomposition: $X_t = S_t + D_t$, where S_t is a pure stochastic process and D_t is a deterministic process.

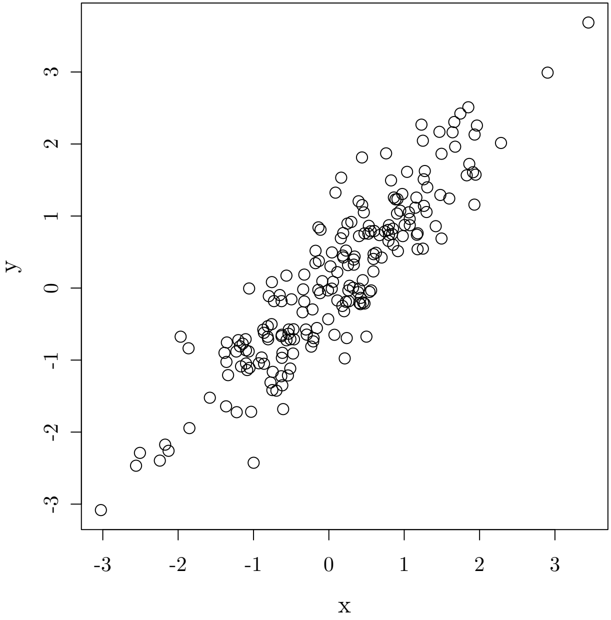
Correlation



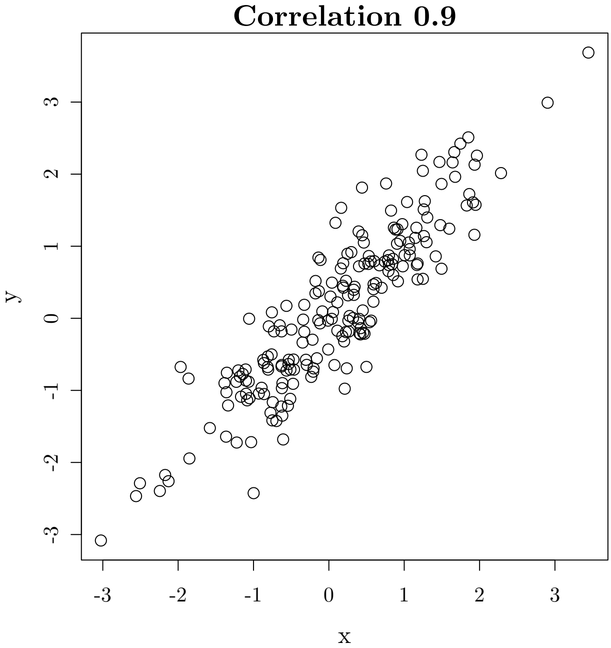
Correlation



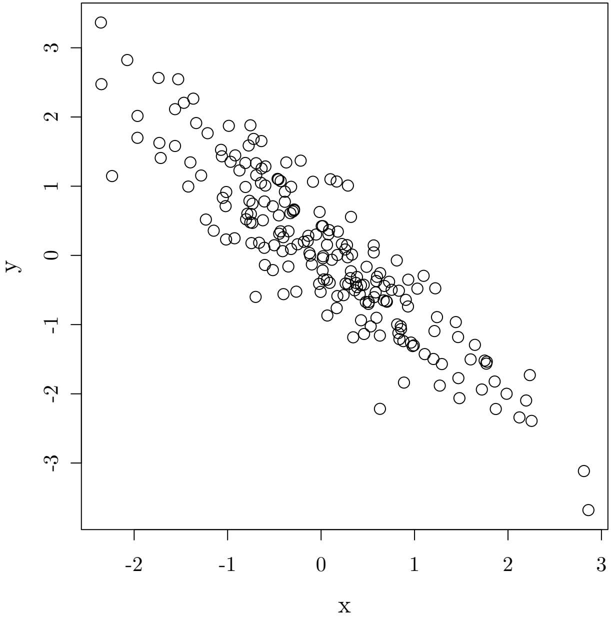
Correlation



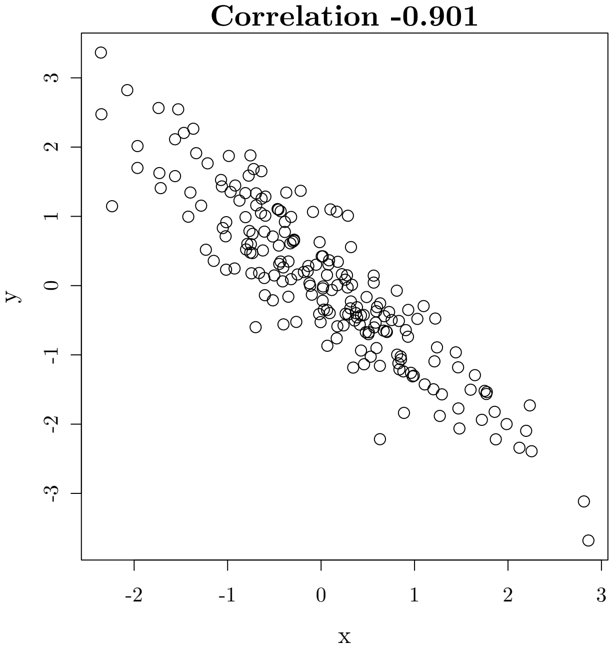
Correlation



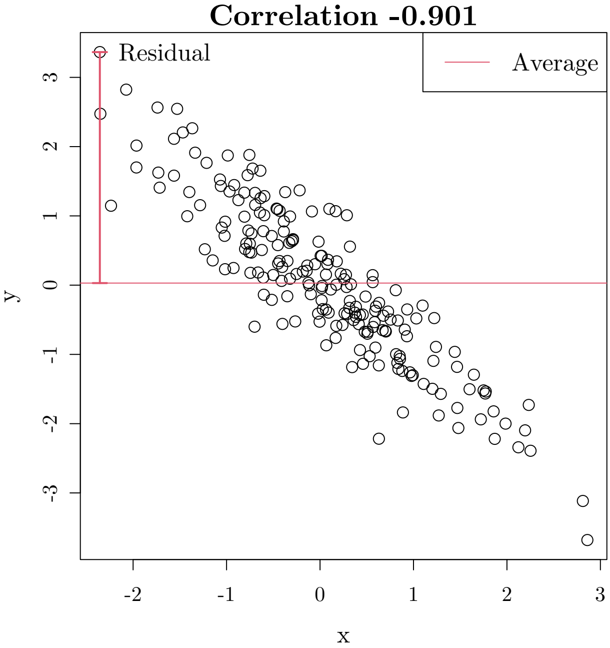
Correlation



Correlation



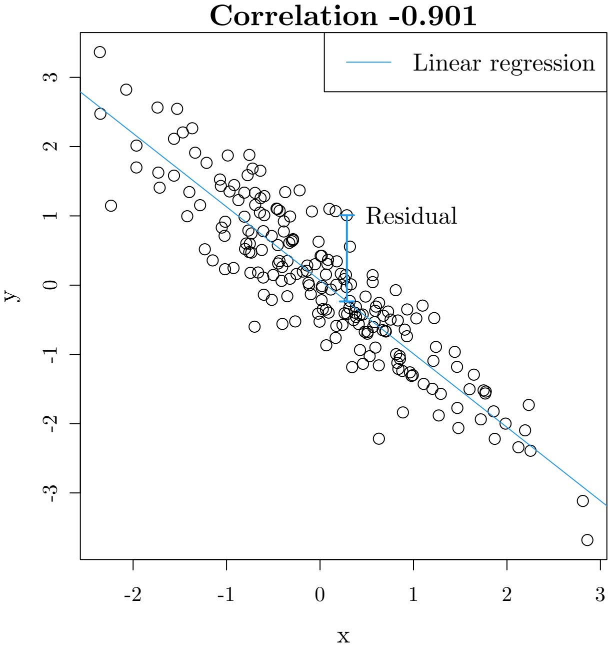
Correlation



Summed Squared Residuals

$$SS_{\text{average}} = \sum_{i}^{n} Residual_i^2$$

Correlation



Summed Squared Residuals

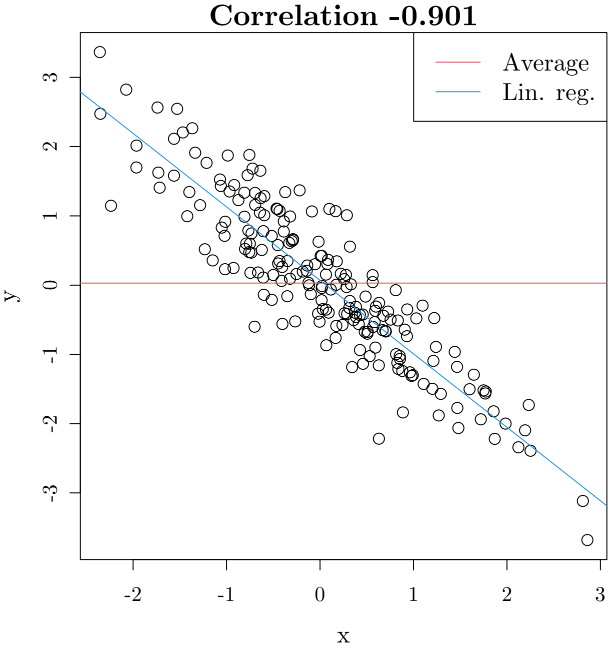
$$SS_{\text{lin}} = \sum_{i}^{n} Residual_i^2$$

How to calculate correlation?

Two vectors with corresponding elements x_i and y_i.

$$R^2 = \frac{SS_{\text{average}} - SS_{\text{lin}}}{SS_{\text{average}}}$$

Correlation

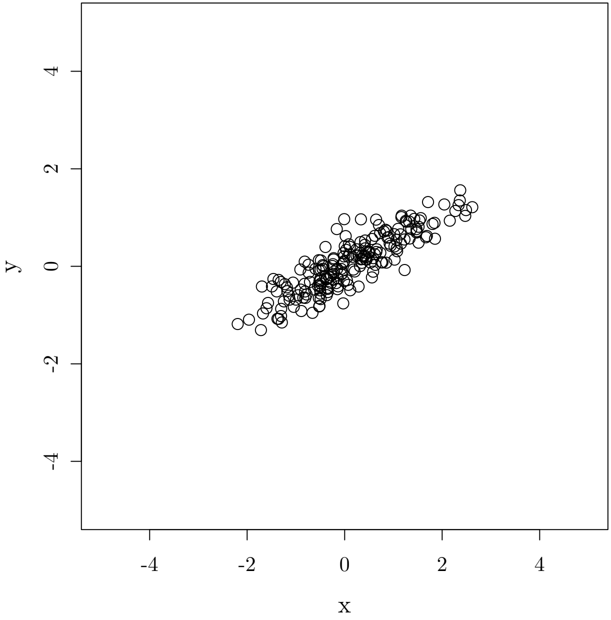


$$R^2 = \frac{SS_{\text{average}} - SS_{\text{lin}}}{SS_{\text{average}}} = \frac{270.548 - 50.81256}{270.548} = 0.8121865$$

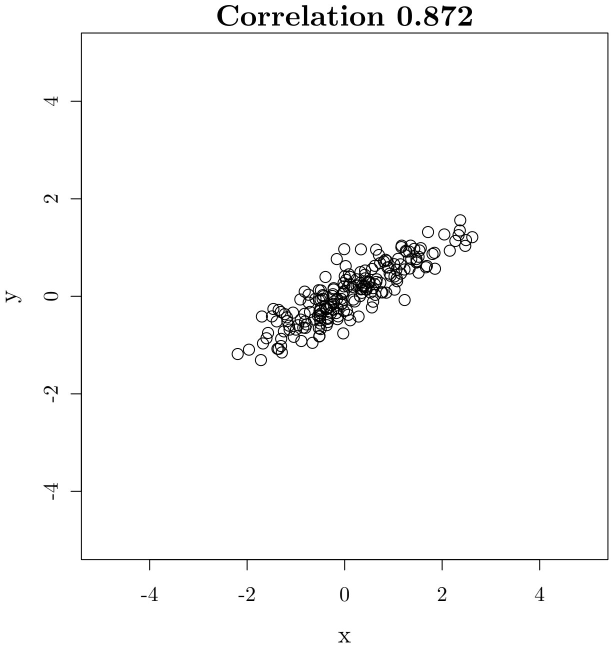
and

$$\text{Cor} = \text{sign}(slope) \cdot R = -0.9012139$$

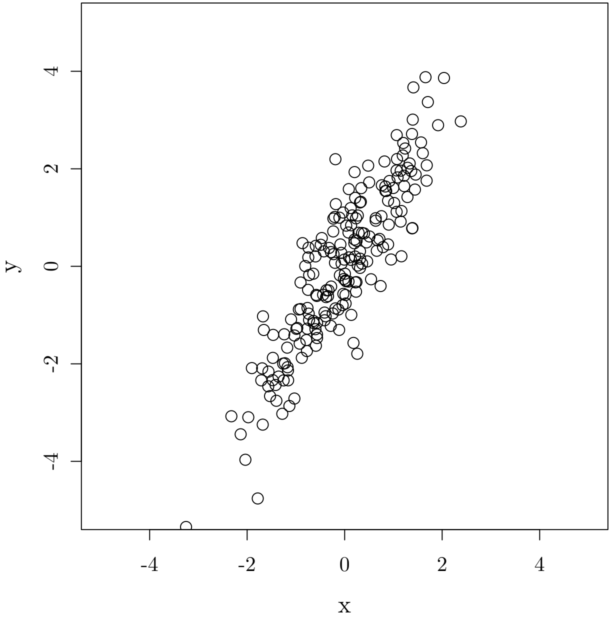
Correlation



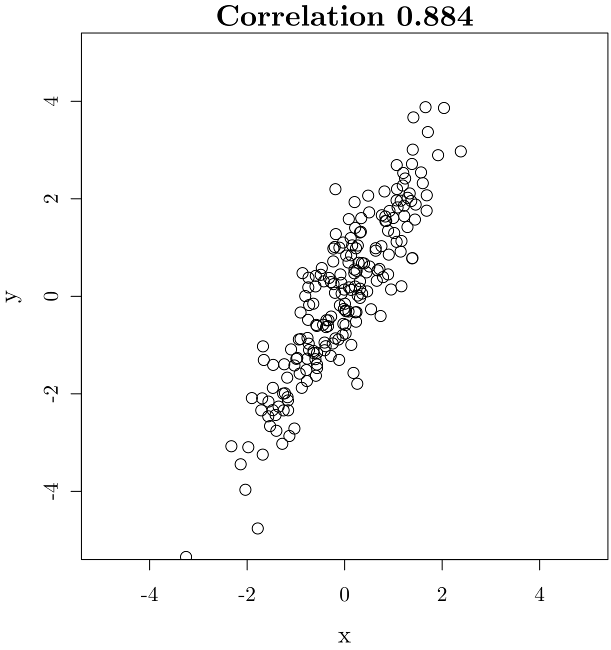
Correlation



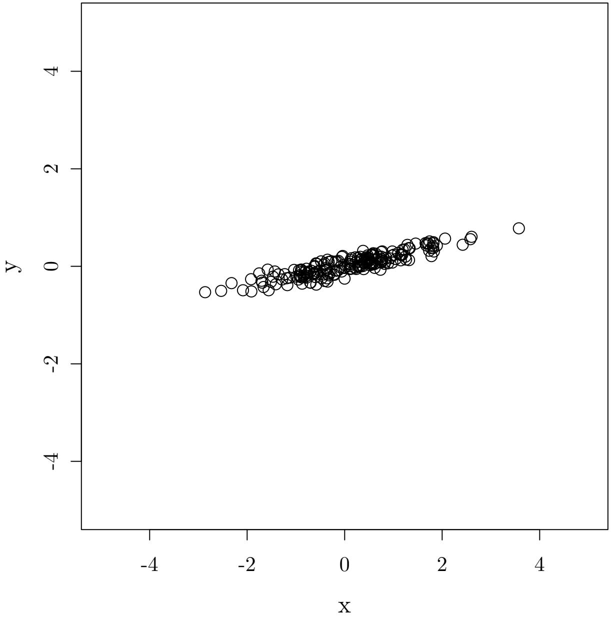
Correlation



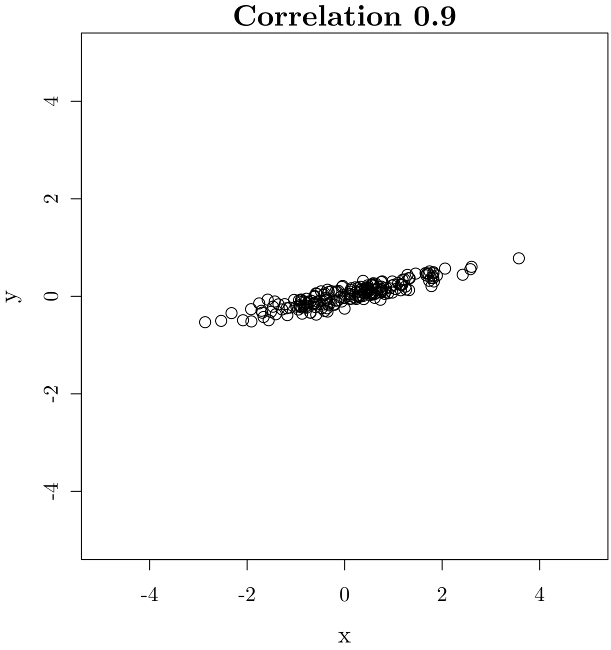
Correlation



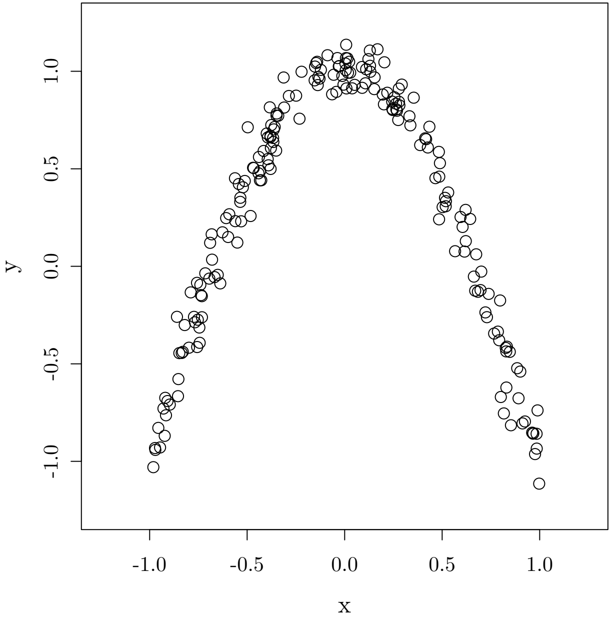
Correlation



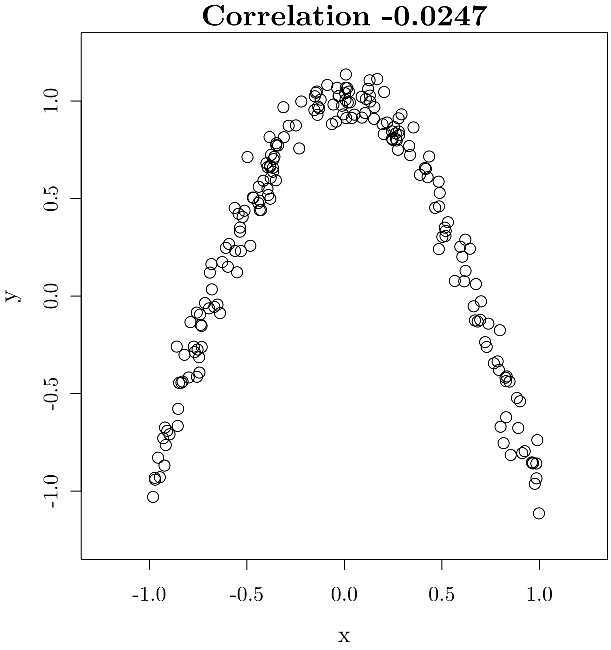
Correlation



Correlation



Correlation



Autocovariance and autocorrelation

- For stationary processes: Only dependent on the time difference $\tau = t_2 - t_1$

Autocovariance and autocorrelation

- For stationary processes: Only dependent on the time difference $\tau = t_2 - t_1$
- Autocovariance:

$$\gamma(\tau) = \gamma_{XX}(\tau) = \mathrm{Cov}[X(t), X(t+\tau)] = E[X(t)X(t+\tau)]$$

Autocovariance and autocorrelation

- For stationary processes: Only dependent on the time difference $\tau = t_2 - t_1$
- Autocovariance:

$$\gamma(\tau) = \gamma_{XX}(\tau) = \text{Cov}[X(t), X(t+\tau)] = E[X(t)X(t+\tau)]$$

- Autocorrelation:

$$\rho(\tau) = \rho_{XX}(\tau) = \gamma_{XX}(\tau)/\gamma_{XX}(0) = \gamma_{XX}(\tau)/\sigma_X^2$$

Autocovariance and autocorrelation

- For stationary processes: Only dependent on the time difference $\tau = t_2 - t_1$
- Autocovariance:

$$\gamma(\tau) = \gamma_{XX}(\tau) = \mathrm{Cov}[X(t), X(t+\tau)] = E[X(t)X(t+\tau)]$$

- Autocorrelation:

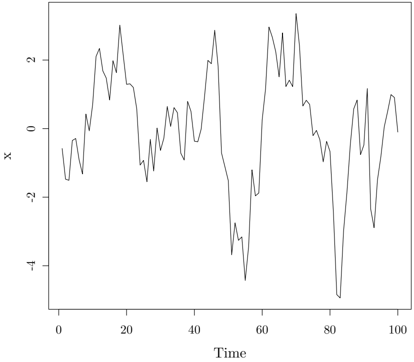
$$\rho(\tau) = \rho_{XX}(\tau) = \gamma_{XX}(\tau)/\gamma_{XX}(0) = \gamma_{XX}(\tau)/\sigma_X^2$$

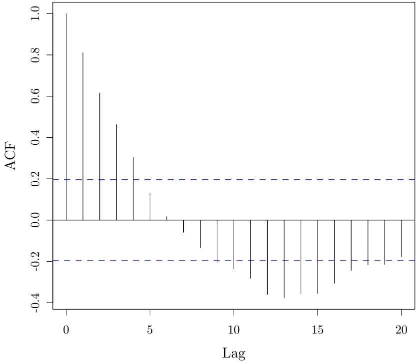
- Some properties of the autocovariance function:
  - $\gamma(\tau) = \gamma(-\tau)$
  - $|\gamma(\tau)| \leq \gamma(0)$

ACF

```
# Simulate a process
n <- 100
x <- filter(rnorm(n), 0.9, "recursive")
plot(x)

# The acf and "height" of "lag 1 bar"
val <- acf(x)
val[1]
# The correlation with lag 1
cor(x, lagvec(x,1), use="complete.obs")
```





White noise

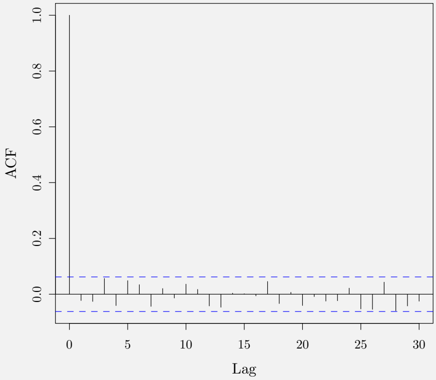
- Def. 5.9: $\{\varepsilon_t\}$ is a sequence of mutually uncorrelated identically distributed random variables, where:
  - $\mu_t = \mathsf{E}[\varepsilon_t] = 0$
  - $\sigma_t = \mathsf{Var}[\varepsilon_t] = \sigma_\varepsilon^2$
  - $\gamma_\varepsilon(k) = \mathsf{Cov}[\varepsilon_t, \varepsilon_{t+k}] = 0$, for $k \neq 0$

How many of the ACF bars stick of the CI if white noise?

White noise

- Def. 5.9: $\{\varepsilon_t\}$ is a sequence of mutually uncorrelated identically distributed random variables, where:
  - $\mu_t = \mathsf{E}[\varepsilon_t] = 0$
  - $\sigma_t = \mathsf{Var}[\varepsilon_t] = \sigma_\varepsilon^2$
  - $\gamma_\varepsilon(k) = \mathsf{Cov}[\varepsilon_t, \varepsilon_{t+k}] = 0$, for $k \neq 0$

How many of the ACF bars stick of the CI if white noise?



A small game! Go try it now: Can you make white noise? Can you make an alternating pattern? Can you make an alternating pattern with "longer period"? Can you make a slow decay to zero?

```
x <- readline("Write a sequence of numbers from 0 to 9\n")
x <- as.numeric(strsplit(x,"")[[1]])
acf(x)
```

More ACF...

```
########
# ACF of a sine
w <- 10
x <- sin((1:1000)*2*pi*(w/1000))
```

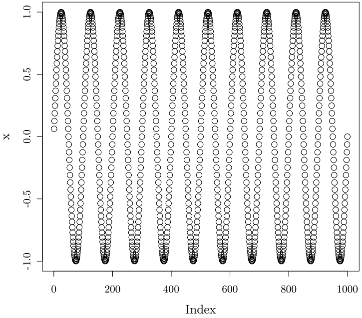
More ACF...

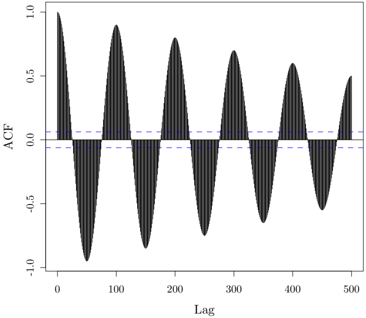
```
########
# ACF of a sine
w <- 10
x <- sin((1:1000)*2*pi*(w/1000))
```

```
# Plot it
par(mfrow=c(1,2))
plot(x)
# The ACF is actually also a sine!
acf(x, lag.max=500)

# Can you "screw it" and make it look like white noise up by altering a single value in x?

#x[100] <- ??
#acf(x, lag.max=500)
```





Linear processes

- A linear process $\{Y_t\}$ is a process that can be written on the form

$$Y_t - \mu = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

  where μ is the mean value of the process and
- $\{\varepsilon_t\}$ is white noise, i.e. a sequence of uncorrelated, identically distributed random variables.
- $\{\varepsilon_t\}$ can be scaled so that $\psi_0 = 1$
- Without loss of generality we assume $\mu = 0$

ψ- and π-weights

- Transfer function and linear process:

$$\psi(B) = 1 + \sum_{i=1}^{\infty} \psi_i B^i \qquad Y_t = \psi(B)\varepsilon_t$$

- Inverse operator (if it exists) and the linear process:

$$\pi(B) = 1 + \sum_{i=1}^{\infty} \pi_i B^i \qquad \pi(B)\,Y_t = \varepsilon_t,$$

- Autocovariance using ψ-weights:

$$\gamma(k) = \mathrm{Cov}\left[Y_t,\, Y_{t+k}\right] = \mathrm{Cov}\left[\sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \sum_{i=0}^{\infty} \psi_i \varepsilon_{t+k-i}\right] = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}$$

Stationarity and invertibility

- The linear process $Y_t = \psi(B)\varepsilon_t$ is *stationary* if

$$\psi(z) = \sum_{i=0}^{\infty} \psi_i z^{-i}$$

  converges for $|z| \geq 1$ (i.e. old values of ε_t are down-weighted)

- The linear process $\pi(B)Y_t = \varepsilon_t$ is said to be *invertible* if

$$\pi(z) = \sum_{i=0}^{\infty} \pi_i z^{-i}$$

  converges for $|z| \geq 1$ (i.e. ε_t can be calculated from recent values of Y_t)

Linear process as a statistical model?

$$Y_t = -\pi_1 Y_{t-1} - \pi_2 Y_{t-2} \ldots + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + +\psi_3 \varepsilon_{t-3} + \ldots$$

- Observations: $Y_1, Y_2, Y_3, \ldots, Y_N$
- Task: Find an infinite number of parameters from N observations!
- Solution: Restrict the sequence $1, \psi_1, \psi_2, \psi_3, \ldots$ - Which gives us the famous ARMA model.