Time Series: A First Course with Bootstrap Starter

In “bis” versions I add - my questions - more code - links to other refs - hints of proof

Lesson 2-1: Random Vectors

- A time series sample is a finite stretch of realizations, i.e., a vector.
- A vector of random variables is called a random vector: $\underline{X} = [X_1, \ldots, X_n]'$.

Mean and Covariance

- The mean of $\underline{X}$ is a vector, each of whose components is the expectation of the corresponding random variable: $\mathbf{E}[X_i]$.
- The covariance matrix of $\underline{X}$ has entries given by the covariance between the corresponding components of the random vector: $\mathrm{Cov}[X_j, X_k]$.
- The covariance matrix $\mathrm{Cov}[\underline{X}]$ of a random vector is non-negative definite and symmetric. Its eigenvalues are real and non-negative.

Affine Transforms

- $\underline{Y} = A\underline{X} + \underline{b}$ is an affine transform of $\underline{X}$.
- If $\mathbf{E}\underline{X} = \underline{\mu}$ and $\mathrm{Cov}[\underline{X}] = \Sigma$, then $\mathbf{E}\underline{Y} = A\underline{\mu} + \underline{b}$ and $\mathrm{Cov}[\underline{Y}] = A\Sigma A'$.

Covariance Decomposition

- We can decompose a symmetric matrix Σ as

$$\Sigma = P\Lambda P'$$

  for an orthogonal matrix P (i.e., $P' = P^{-1}$), and where Λ is a diagonal matrix consisting of the real eigenvalues of Σ.
- A symmetric non-negative definite matrix Σ can be decomposed as $\Sigma = BB'$, and B is called a square root (it is not unique). One such square root is the *Cholesky* factor.
- If $\underline{Z}$ has i.i.d. components with mean zero and variance one, then $\underline{X} = B\underline{Z} + \underline{\mu}$ is a random vector with mean $\underline{\mu}$ and covariance matrix BB'.

Simulation Example

- Simulate a bivariate random vector with mean $[1, 2]$ and covariance matrix

$$\Sigma = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}.$$

```r
Sigma <- rbind(c(2,1),c(1,4))
mu <- c(1,2)
B <- t(chol(Sigma)) # t to get lower triangular
z <- matrix(rnorm(2*100),nrow=2) # 2 vecteurs de 100 obs
x <- B %*% z + mu
print(colMeans(t(x)))
```

```
## [1] 1.026188 1.835320
```

```r
print(var(t(x)))
```

```
##          [,1]     [,2]
## [1,] 2.286739 1.696723
## [2,] 1.696723 5.025462
```

Gaussian Random Vectors

- A random vector $\underline{Y}$ is Gaussian with mean $\underline{\mu}$ and non-singular covariance matrix Σ if its joint pdf is

$$p_{\underline{Y}}(\underline{y}) = (2\pi)^{-n/2} (\det \Sigma)^{-1/2} \exp\{-(\underline{y} - \underline{\mu})' \Sigma^{-1} (\underline{y} - \underline{\mu})/2\}.$$

- Denoted by writing $\underline{Y} \sim \mathcal{N}(\underline{\mu}, \Sigma)$.
- An affine transformation of a Gaussian vector is still Gaussian. In particular, sub-vectors are Gaussian.
- We can decorrelate a Gaussian random vector: $\underline{X} = B^{-1}\underline{Y}$ has $\mathrm{Cov}[\underline{X}] = B^{-1}\Sigma B^{-1\prime} = 1_n$, the identity matrix.
- The quadratic form

$$(\underline{Y} - \underline{\mu})' \Sigma^{-1} (\underline{Y} - \underline{\mu})$$

has a χ^2 distribution on n degrees of freedom.

Lesson 2-2: Stochastic Processes

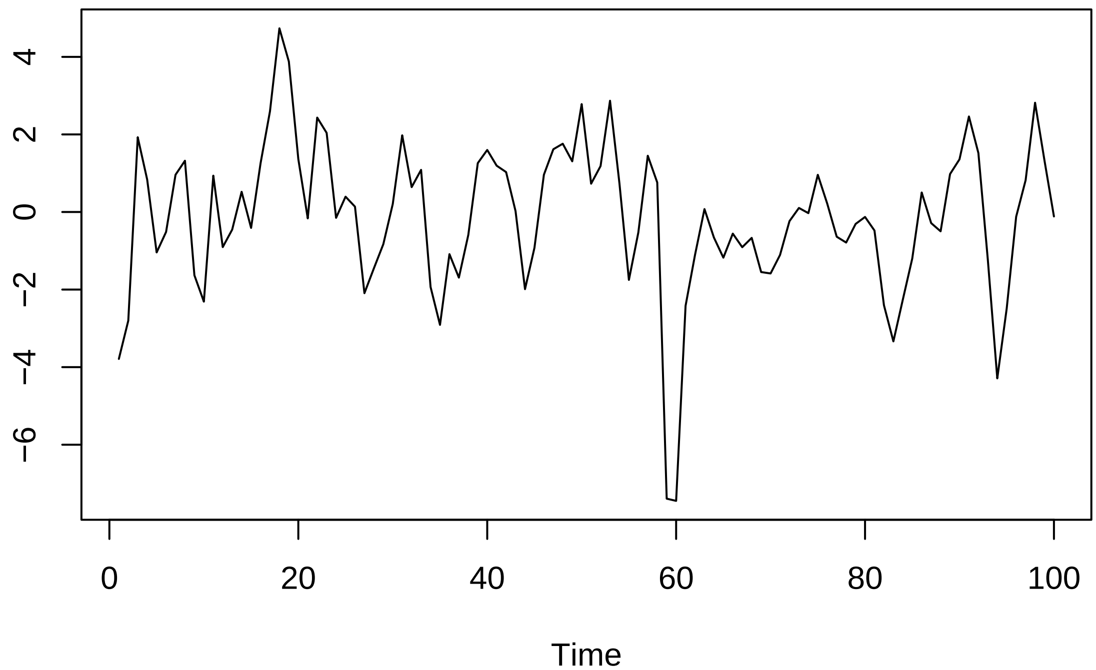
- A collection of random variables indexed by time is called a *stochastic process*, denoted as $\{X_t\}$. The curly brackets let us know $\{X_t\}$ is the process, whereas X_t is a single random variable (at time t).
- Usually time is $t \in \mathbf{Z}$, the integers.
- There are also continuous-time stochastic processes (another subject).

Realization is Sample Path

- A random variable X_t has realization x_t.
- Example: $X_t \sim \mathcal{N}(0, 1)$ has realization -0.5604756.
- Put this together for all $t \in \mathbf{Z}$, and the realization is called the *sample path*.

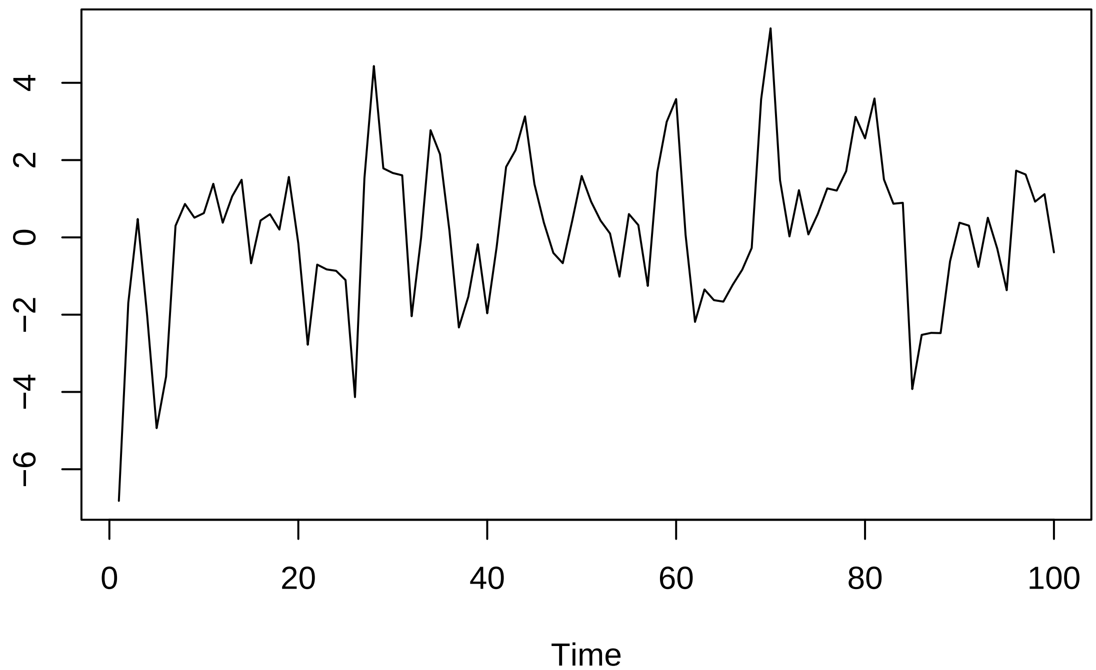
Example: Heavy-tailed Sample Path

```r
set.seed(777)
n <- 100
z <- rt(n+1,df=4)        # heavy-tailed input
theta <- .8
x <- z[-1] + theta*z[-(n+1)]
plot(ts(x),xlab="Time",ylab="")
```



- As usual, we connect the dots when graphing the sample path.
- Here is another realization, or sample path, of the same stochastic process.

```
set.seed(888)
n <- 100
z <- rt(n+1,df=4)        # heavy-tailed input
theta <- .8
x <- z[-1] + theta*z[-(n+1)]
plot(ts(x),xlab="Time",ylab="")
```

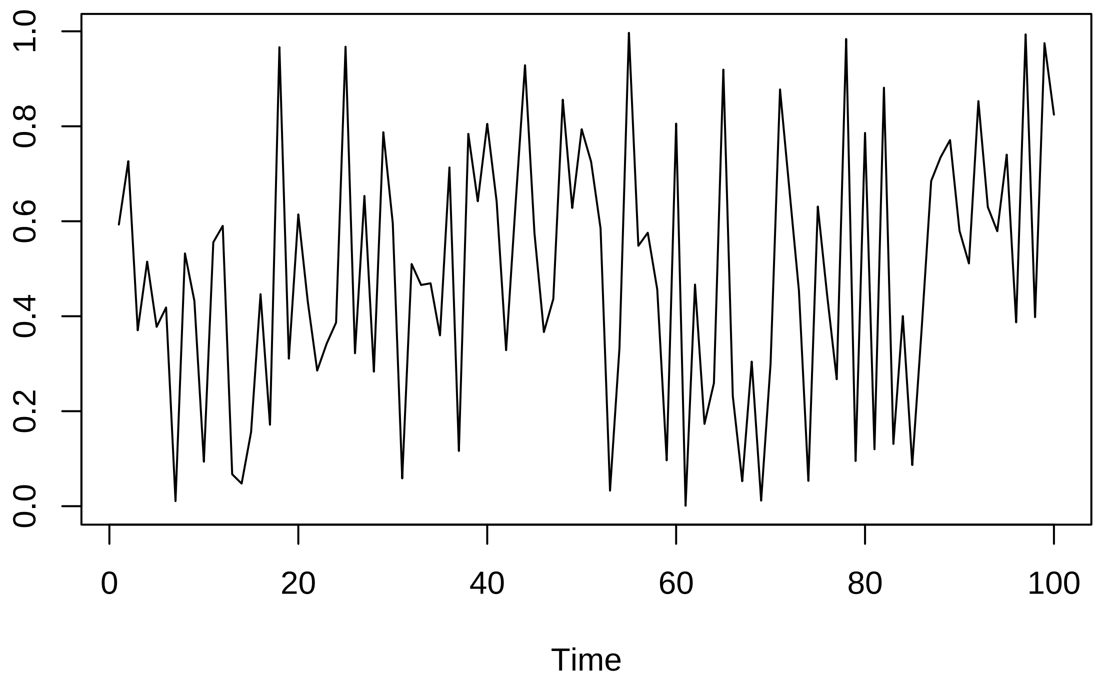


Common Examples of Stochastic Processes

Example 2.2.8. Process A: i.i.d.

- An i.i.d. process, where each X_t has the same distribution and is independent of the rest.

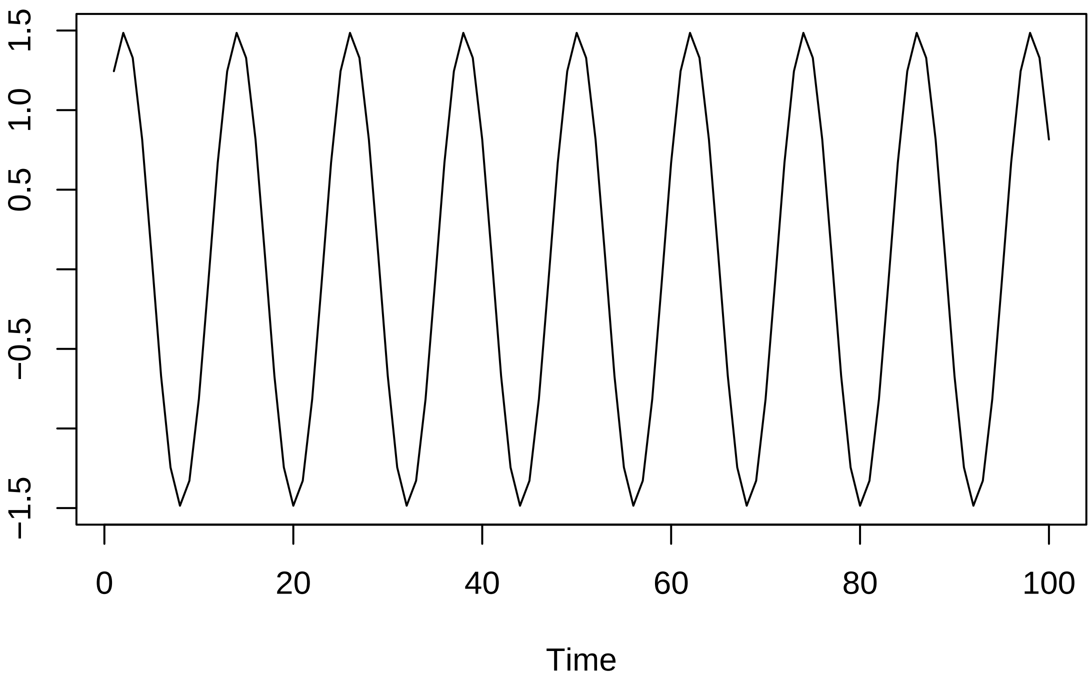
```
set.seed(111)
n <- 100
x <- runif(n)
plot(ts(x),xlab="Time",ylab="")
```



Example 2.2.9. Process B: Cosine

- Suppose $X_t = A\cos(\vartheta t + \Phi)$, where ϑ is given, and A and Φ are independent random variables.

```
n <- 100
set.seed(222)
A <- rnorm(1)
set.seed(223)
phi <- 2*pi*runif(1)
lambda <- pi/6
set.seed(224)
x <- A*cos(seq(1,n)*lambda + phi)
plot(ts(x),xlab="Time",ylab="")
```



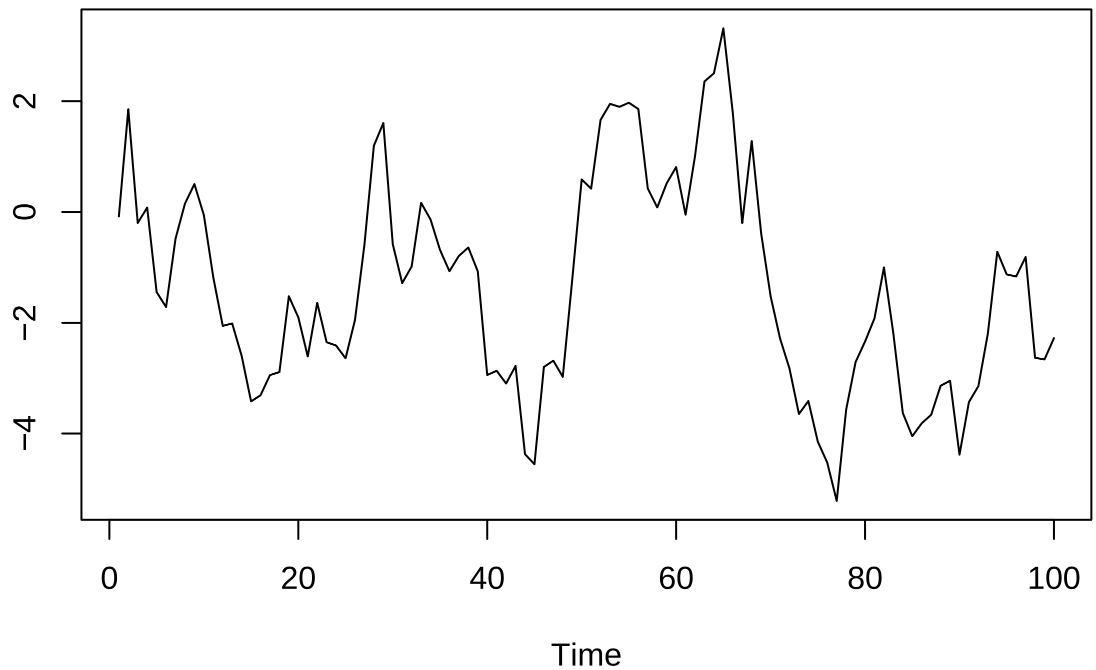
Example 2.2.12. Process E: Random Walk

- Suppose X_t is current location on a straight line, and we step forward or backward at time $t+1$. Let the step size be given by random variable Z_{t+1}, independent of where we are. Then our new location is

$$X_{t+1} = X_t + Z_{t+1}.$$

  This is called a *random walk*.
- We can initialize with $X_0 = 0$, for example.

```
set.seed(333)
n <- 100
z <- rnorm(n)
x <- rep(0,n)
x0 <- 0
x[1] <- x0 + z[1]
for(t in 2:n) { x[t] <- x[t-1] + z[t] }
plot(ts(x),xlab="Time",ylab="")
```



Lesson 2-3: Stationarity

- We want to generalize the concept of *identical distribution* to a stochastic process.

Marginal Distributions

- First marginals are just the X_t random variables' distributions.
- Second marginals are joint distributions for all pairs (X_t, X_s).
- Third marginals are joint distributions for all triplets, etc.

Same First Marginals

- Saying $\{X_t\}$ has same first marginal is same as saying they are identically distributed.
- Sometimes called *First Order Stationary*.
- In particular, all means are the same: $\mathbf{E}[X_t] = \mathbf{E}[X_s]$ for all t, s.

Second Marginals Under Shift

- Suppose all pairs have the same distribution when shifted:

$$(X_1, X_2) \sim (X_2, X_3) \sim (X_3, X_4) \ldots$$

- Then second marginal distribution only depends on lag h, i.e., distribution of (X_t, X_{t-h}) does not depend on t.
- Sometimes called *Second Order Stationary*.

- Then the product mean (the covariance) depends only on lag:

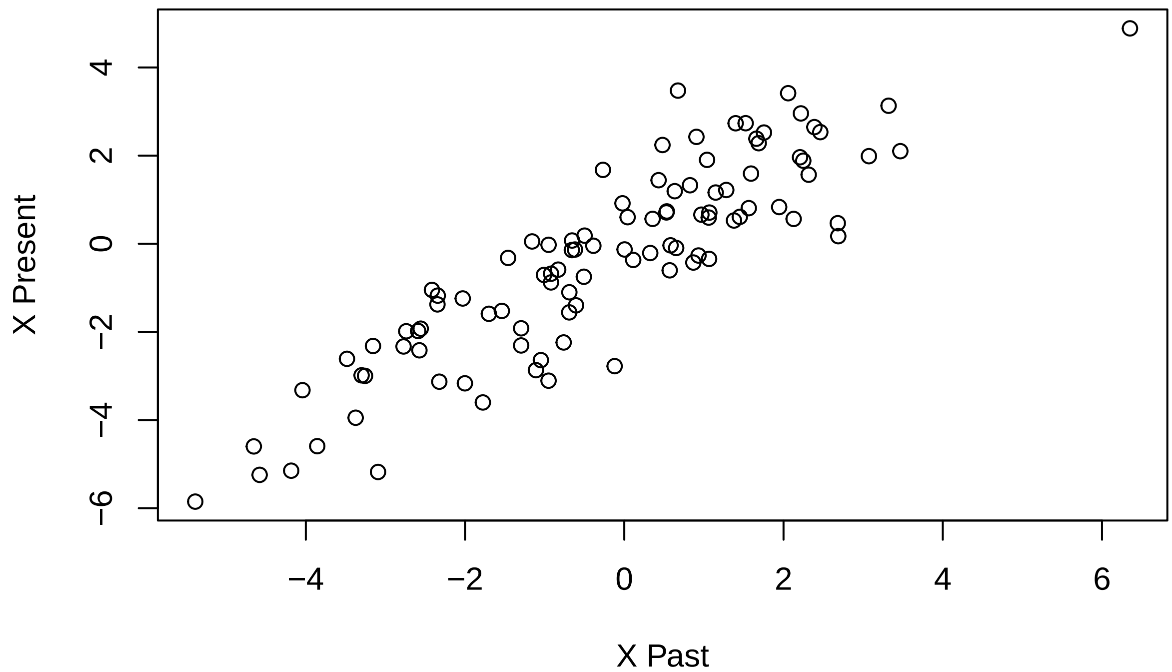
$$\mathbf{E}[X_t \, X_{t-h}].$$

  It does not depend on t.

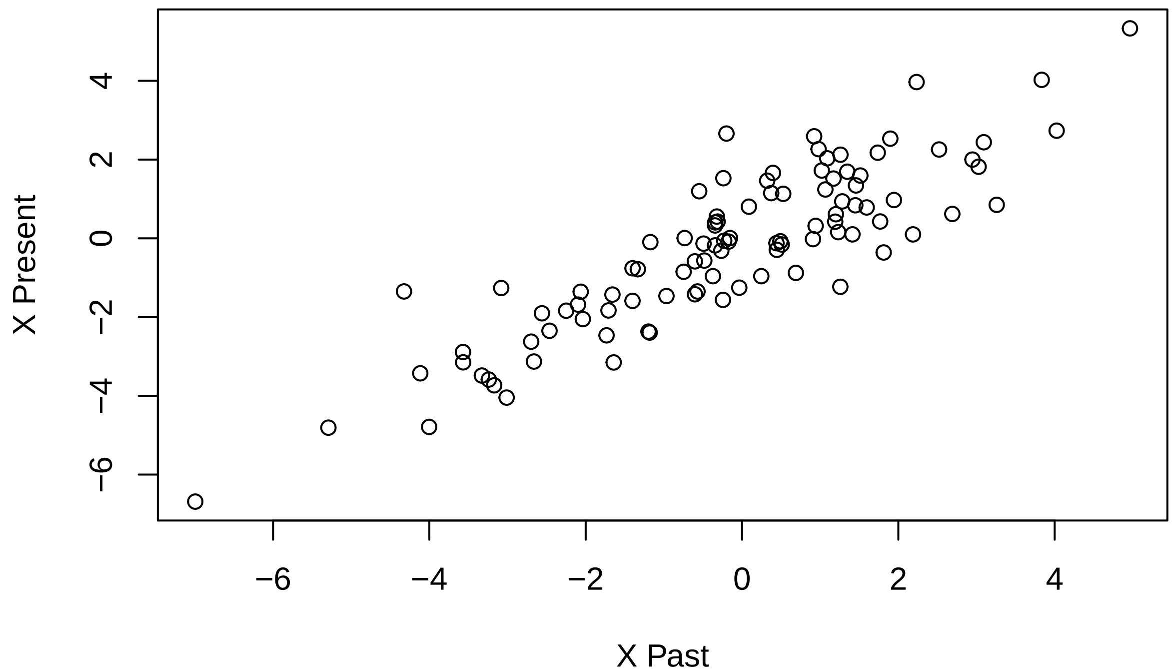
Example: Visualizing Stationarity

- We generate 100 simulations of a Gaussian AR(1), and generate a scatterplot of (X_1, X_2)
- We repeat with (X_3, X_4)

```
x1 <- NULL
x2 <- NULL
x3 <- NULL
x4 <- NULL

for(i in 1:100) {

z <- rnorm(10)
x <- rep(0,10)
phi <- .9
x0 <- rnorm(1)/sqrt(1-phi^2)
x[1] <- phi*x0 + z[1]
for(t in 2:10) { x[t] <- phi*x[t-1] + z[t] }

x1 <- c(x1,x[1])
x2 <- c(x2,x[2])
x3 <- c(x3,x[3])
x4 <- c(x4,x[4])
}

plot(x2,x1,xlab="X Past",ylab="X Present")
```




```r
plot(x4,x3,xlab="X Past",ylab="X Present")
```



Lesson 2-4: Autocovariance

- Now we study the autocovariance function.

Strict and Weak Stationarity

- Strict stationarity: all marginals (of all orders) are time shift invariant.
- Weak stationarity: the time series has finite variance, constant mean μ, and covariance only depends on lag h:

$$\gamma(h) = \mathrm{Cov}[X_t, X_{t-h}] = \mathbf{E}[X_t\, X_{t-h}] - \mu^2.$$

  This function is called the **autocovariance**.
- So the variance is $\gamma(0)$.
- The **autocorrelation** is

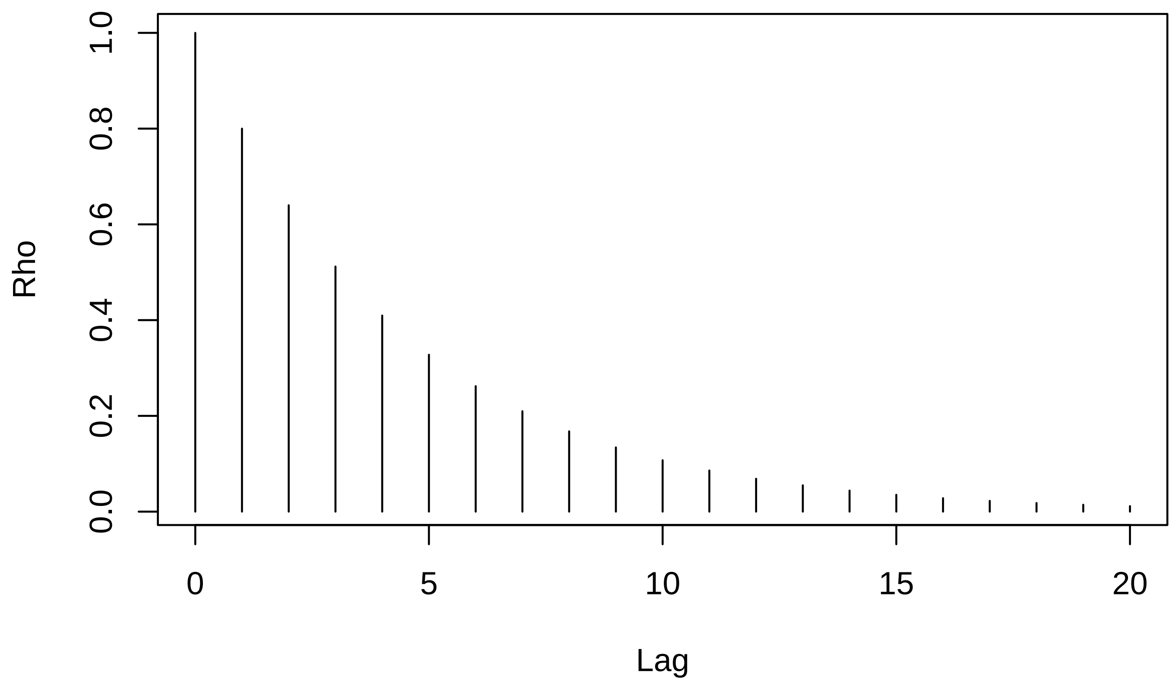
$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}.$$

- Weak stationarity is sometimes called *covariance stationarity*.

Example: Autocorrelation of an AR(1)

- We plot $\rho(h)$ versus h (on x-axis).

```
phi <- .8
rho <- phi^seq(0,20)
plot(ts(rho,start=0),xlab="Lag",ylab="Rho",type ="h")
```



White Noise

- A key example is a *white noise* stochastic process.
- This is any weakly stationary process $\{Z_t\}$ with mean zero such that $\gamma(h) = 0$ for $h \neq 0$.
- Written compactly as $Z_t \sim \mathrm{WN}(0, \sigma^2)$, where $\sigma^2 = \gamma(0)$ is the variance, and the mean is $\mu = 0$.

Covariance Matrix of Sample Vector

- The time series variables corresponding to a sample are $X_1, \ldots, X_n$, which can be put into a random vector $\underline{X}$.
- The covariance matrix of $\underline{X}$ is denoted by Γ_n when the stochastic process is weakly (or strictly) stationary. The entry in row j and column k is

$$\Gamma_n(j,k) = \mathrm{Cov}[X_j, X_k] = \gamma(k-j).$$

  This only depends on the difference between row and column index! Such a matrix is constant along diagonals, and is called *Toeplitz*.

```
rho <- .8
gamma <- rho^seq(0,5)/(1-rho^2)
gamma_mat <- toeplitz(gamma)
gamma_mat
```

```
##           [,1]     [,2]     [,3]     [,4]     [,5]      [,6]
## [1,] 2.7777778 2.222222 1.777778 1.422222 1.137778 0.9102222
## [2,] 2.2222222 2.777778 2.222222 1.777778 1.422222 1.1377778
## [3,] 1.7777778 2.222222 2.777778 2.222222 1.777778 1.4222222
```

```
## [4,] 1.4222222 1.777778 2.222222 2.777778 2.222222 1.7777778
## [5,] 1.1377778 1.422222 1.777778 2.222222 2.777778 2.2222222
## [6,] 0.9102222 1.137778 1.422222 1.777778 2.222222 2.7777778
```

Properties of Autocovariance

1. $\gamma(0) \geq 0$
2. $\gamma(h) = \gamma(-h)$
3. $|\gamma(h)| \leq \gamma(0)$.
4. $\gamma(h)$ is a non-negative definite sequence.

This last property means that Γ_n is a non-negative definite matrix for all n. (Recall from multivariate analysis: covariance matrices are non-negative definite, and are positive definite if all eigenvalues are positive.)

Lesson 2-5: Autoregression and Moving Average

- Examples of weakly stationary stochastic process.

Example 2.5.1. AR(1) Process

- Let $Z_t \sim$ i.i.d.$(0, \sigma^2)$ and $\{X_t\}$ defined via

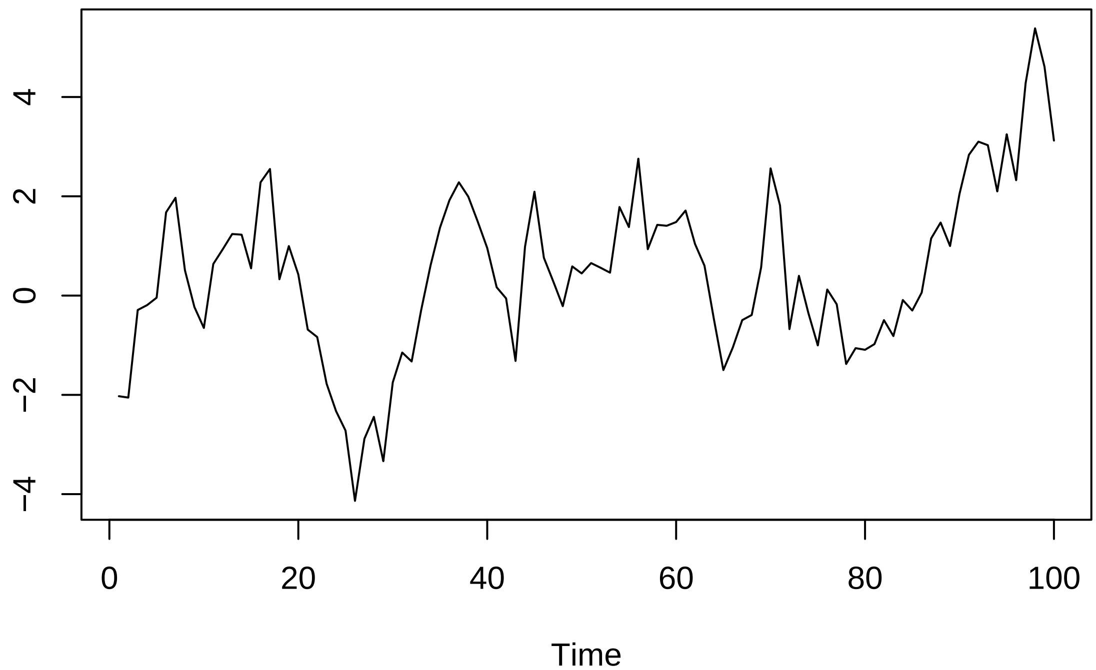
  $$X_t = \phi\, X_{t-1} + Z_t$$

  for $t \geq 1$, where $|\phi| < 1$.
- This is a recursion, called an *order 1 autoregression*, or AR(1).
- How to define X_0? If $X_0 \sim (0, \sigma^2/(1-\phi^2))$, then $\{X_t\}$ is weakly stationary and

  $$\gamma(h) = \sigma^2 \frac{\phi^{|h|}}{1-\phi^2}.$$

  Formula only makes sense when $|\phi| < 1$. This is a *stationarity condition*.

```r
n <- 100
set.seed(123)
z <- rnorm(n)
x <- rep(0,n)
phi <- .9
x0 <- rnorm(1)/sqrt(1-phi^2)
x[1] <- phi*x0 + z[1]
for(t in 2:n) { x[t] <- phi*x[t-1] + z[t] }
plot(ts(x),xlab="Time",ylab="")
```



Example 2.5.5. MA(1) Process

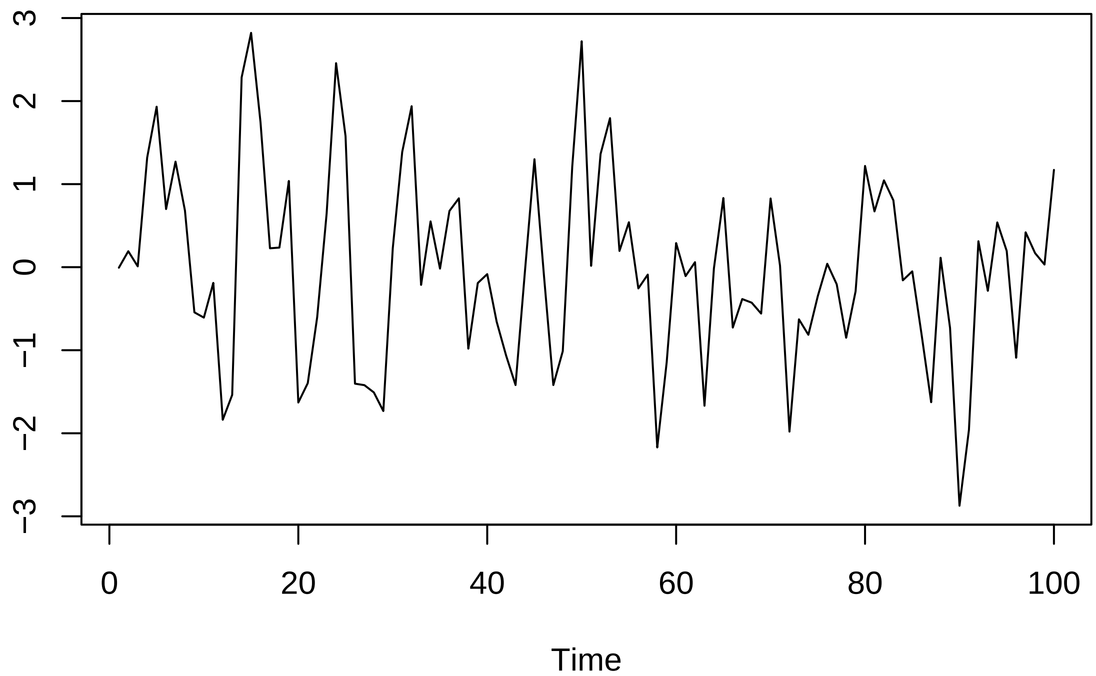
- Let $Z_t \sim \text{i.i.d.}(0, \sigma^2)$ and $\{X_t\}$ defined via

$$X_t = Z_t + \theta\, Z_{t-1}$$

  for $t \geq 1$, where θ is any real number.
- This process is called an *order 1 moving average*, or MA(1).
- It is weakly stationary, with

$$\gamma(h) = \begin{cases} (1+\theta^2)\sigma^2 & \text{if } h = 0 \\ \theta\sigma^2 & \text{if } h = \pm 1 \\ 0 & \text{if } |h| > 1. \end{cases}$$

```
set.seed(777)
n <- 100
z <- rnorm(n+1)      # Gaussian input
theta <- .8
x <- z[-1] + theta*z[-(n+1)]
plot(ts(x),xlab="Time",ylab="")
```



Example 2.5.6. MA(2) Process

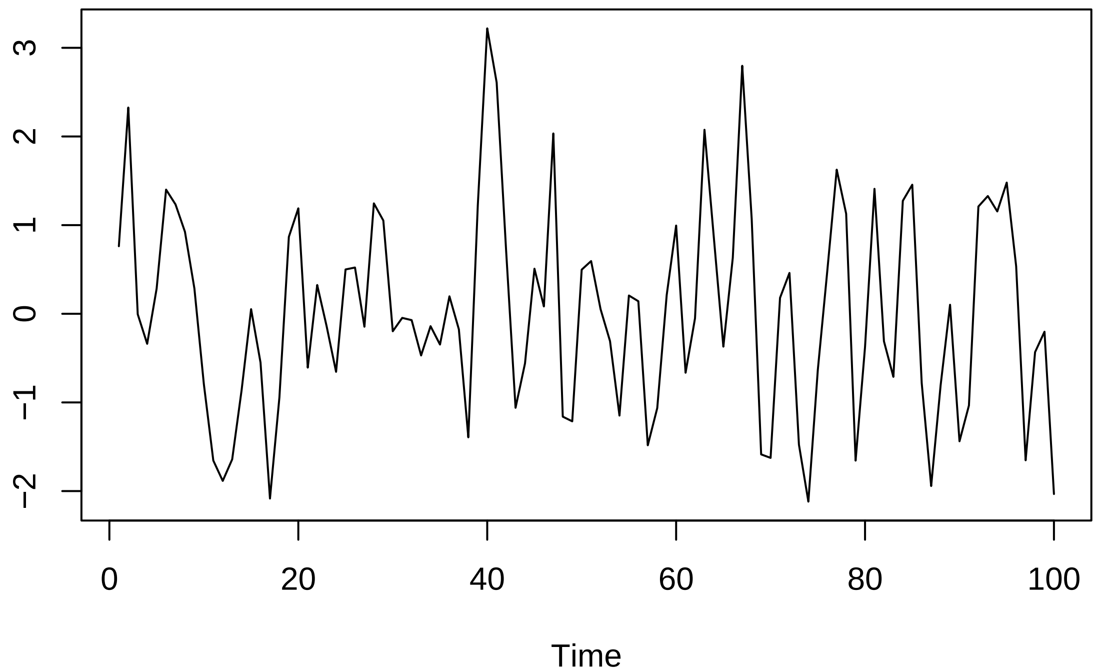
- Let $Z_t \sim$ i.i.d.$(0, \sigma^2)$ and $\{X_t\}$ defined via

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2}$$

  for $t \geq 2$, where θ_1, θ_2 are any real numbers.
- This process is called an *order 2 moving average*, or MA(2).
- It is weakly stationary, with

$$\gamma(h) = \begin{cases} (1+\theta_1^2+\theta_2^2)\sigma^2 \quad \text{if } h = 0 \\ (\theta_1 + \theta_1\theta_2)\sigma^2 \quad \text{if } h = \pm 1 \\ \theta_2\sigma^2 \quad \text{if } h = \pm 2 \\ 0 \quad \text{if } |h| > 2. \end{cases}$$

```
set.seed(555)
n <- 100
z <- rnorm(n+2)      # Gaussian input
theta1 <- .9
theta2 <- .2
x <- z[-c(1,2)] + theta1*z[-c(1,n+2)] + theta2*z[-c(n+1,n+2)]
plot(ts(x),xlab="Time",ylab="")
```



Lesson 2-6: White Noise Processes

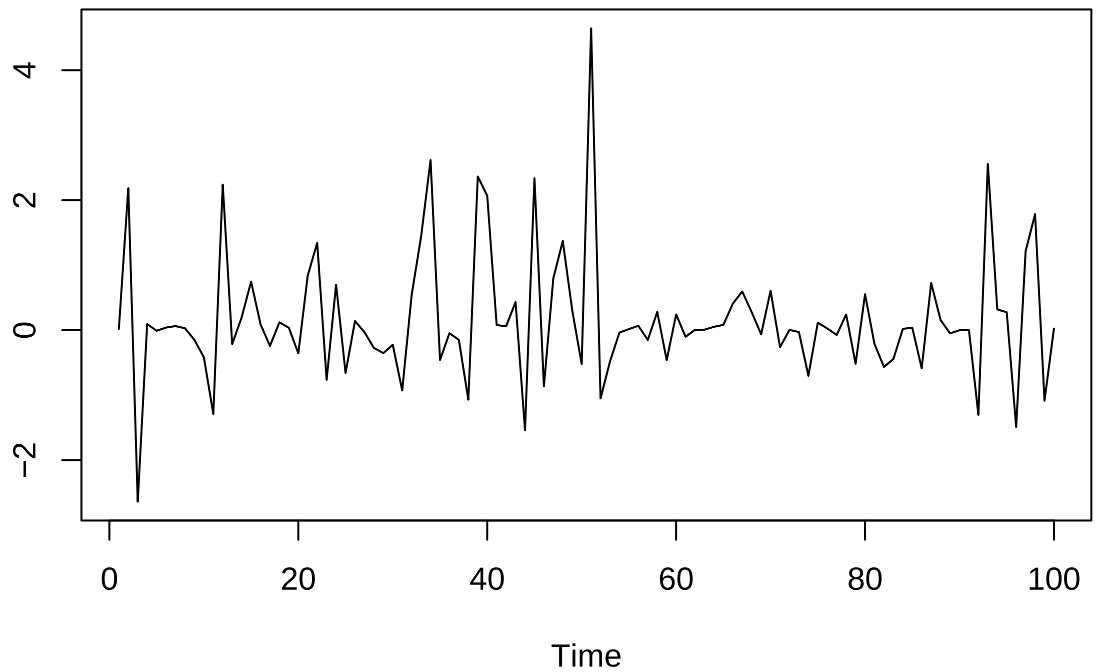
- *White noise* is a fundamental building block for time series models.
- Any $\{X_t\}$ i.i.d. with mean zero and variance σ^2 is a $\text{WN}(0, \sigma^2)$.
- Here we provide three examples of white noise.

Example 2.6.1. Dependent White Noise

- Consider $X_t = Z_t \cdot Z_{t-1}$, where Z_t is i.i.d. $N(0, 1)$.

- Then $X_t \sim \text{WN}(0, 1)$, but $\{X_t\}$ is not i.i.d.

```
n <- 101
z <- rnorm(n)
x <- z[-n]*z[-1]
plot(ts(x),xlab="Time",ylab="")
```




```
#acf(x)
```

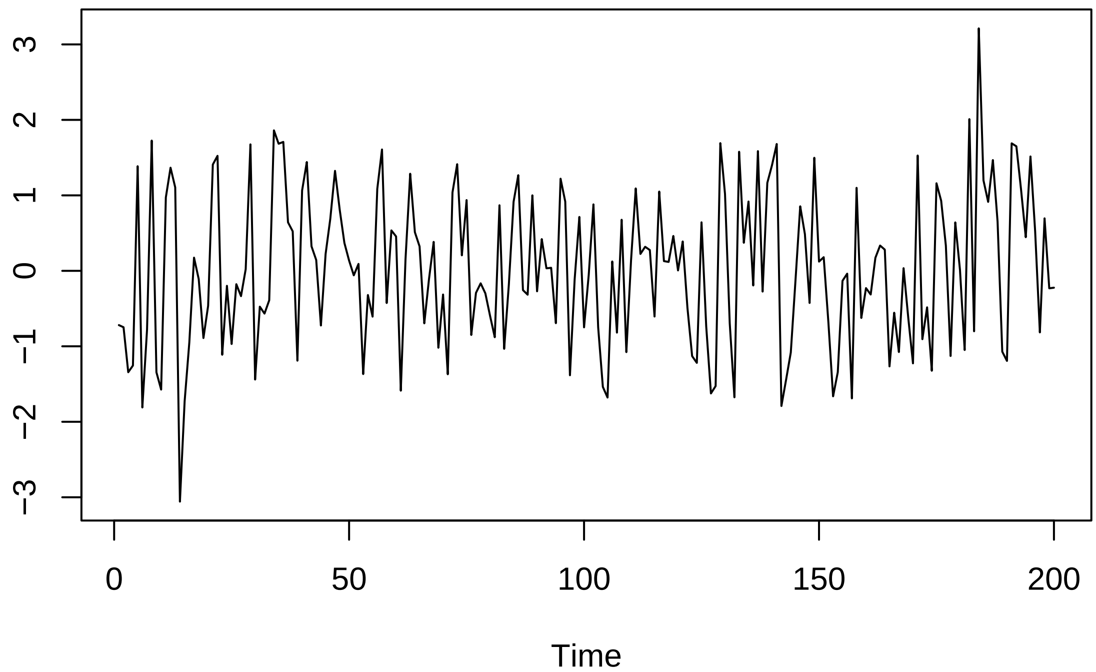
Example 2.6.2. Non-identically Distributed White Noise

- Let $\{Y_t\}$ and $\{Z_t\}$ be independent of each other. Set Z_t i.i.d. $N(0,1)$ and Y_t i.i.d. uniform on $(-\sqrt{3}, \sqrt{3})$. Let

$$X_t = \begin{cases} Z_t & t \text{ even} \\ Y_t & t \text{ odd} \end{cases}$$

- Then $X_t \sim \mathrm{WN}(0,1)$, although the process is not stationary (since the marginal distribution depends on t).

```
n <- 100
z <- rnorm(n)
y <- runif(n,-3^(1/2),3^(1/2))
x <- matrix(t(cbind(y,z)),ncol=1)
plot(ts(x),xlab="Time",ylab="")
```

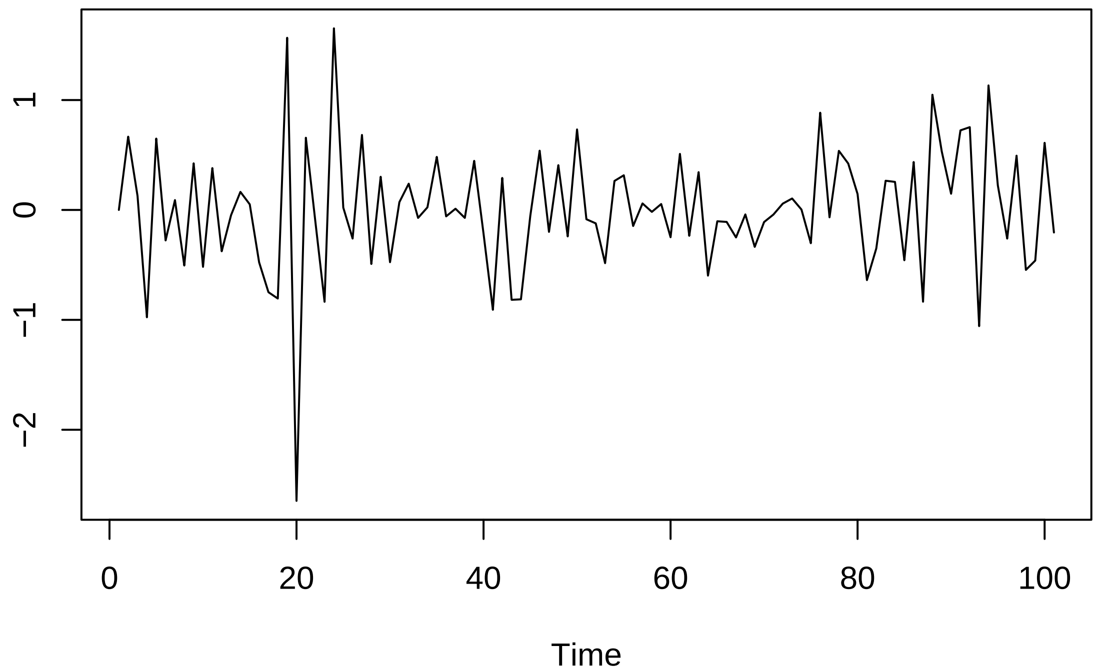



```
#acf(x)
```

Example 2.6.3. ARCH Process

- Model of Engel (1982), a Nobel laureate.
- Set Z_t i.i.d. $N(0,1)$, and $X_t = Z_t\sqrt{\alpha + \beta X_{t-1}^2}$.
- Then $X_t \sim \mathrm{WN}(0, \alpha/(1-\beta))$.

```
n <- 101
z <- rnorm(n)
alpha <- .2
beta <- .3
x <- 0
for(t in 2:n) { x <- c(x,z[t]*sqrt(alpha + beta*x[t-1]^2)) }
plot(ts(x),xlab="Time",ylab="")
```




```
#acf(x)
```