Predictive Modeling

Series 12

Exercise 12.1

Let $\{X_1, X_2, \dots\}$ be a discrete stochastic process. Show that

$$\gamma(i,j) = \mathrm{E}(X_i X_j) - \mu(i)\mu(j).$$

Exercise 12.2

We consider a moving average process of the form

$$X_i = W_{i-1} + 2W_i + W_{i+1},$$

where W_i are independet random variables with zero mean and variance σ^2.

a) Compute the mean sequence of the process.

b) Compute the autocovariance and -correlation sequence.

c) Plot the autocorrelation $\rho(i,j)$ as a function of the lag $h = i - j$.

Exercise 12.3

In this exercise we study a simulated discrete process $\{X_1, X_2, \dots\}$ given by the following construction rule

- Set $X_1 = -1$
- For each $k \geq 1$ toss a fair coin and set $D_k = 1$ for head and $D_k = -1$ for tail. Define

  $$X_k = a + D_k + bD_{k-1}$$

  for some numbers $a, b \in \mathbb{R}$.

a) Generate a time series $\{x_1, x_2, \dots, x_{200}\}$ from the process and create a time series plot for $a = 2$ and $b = -0.7$.

Hint: The coin tossing can be modelled by a binomial variable with $n = 1$ and $p = 0.5$. In **R** binomial random variables can be created with the command **rbinom**. The following snipped creates the sequence x

```
# Create the process Dk
D = 2 * (rbinom(200, size = 1, p = 0.5) - 0.5)

# Create the process Xk
X = rep(-1, 1, 200)
for (k in 2:200) {
    X[k] = 2 + D[k] - 0.7 * D[k - 1]
}
```

b) From your data compute the sample autocorrelation $\hat{\rho}(k)$ using the **acf()** function. Create a correlogram up to lag 50.

c) Compute the theoretical mean $\mu(k)$ and the autocorrelation $\rho(k)$ of the process from the definition (i.e. for general a and b). Compare $\hat{\rho}(1)$ and $\rho(1)$ for $a = 2$ and $b = -0.7$.

d) Is the process X_k weakly stationary?

Exercise 12.4

This example deals with real world data: the monthly land air temperature anomalities on the northern hemisphere[1].

a) Load the data file **GlobalTemperature.Rdata** file simply by typing

```
load("Daten/GlobalTemperature.Rdata")
```

You should then find an object **Global.ts** of class **ts** which contains the monthly temperature starting from January 1850 until March 2017. Make a time series plot of the data.

b) Perform an STL decomposition of the data. Vary the parameter **s.window** and find a proper value by visual inspection of the trend and seasonal components. Comment on the result.

c) Is the remainder sequence generated by a weakly stationary process? Compute the correlogram and interpret it. Are there statistically significant correlations?

[1] The data are updated regularly and can be downloaded free of charge at: http://www.cru.uea.ac.uk/data/

Hint: The `stl` function returns a data.frame with a component `time.series`. The third column is the remainder sequence.

Result Checker

Predictive Modeling

Solutions to Series 12

Solution 12.1

$$\begin{aligned}\gamma(i,j) &= \mathrm{E}\left((X_i - \mu(i))(X_j - \mu(j))\right)\\ &= \mathrm{E}\left(X_i X_j) - \mu(i)X_j - \mu(j)X_i + \mu(i)\mu(j)\right)\\ &= \mathrm{E}(X_i X_j) - \mu(i)\mu(j) - \mu(j)\mu(i) + \mu(i)\mu(j) = \mathrm{E}(X_i X_j) - \mu(i)\mu(j).\end{aligned}$$

Solution 12.2

a) Since the W_i's are random variabels with zero mean, it follows that any linear combination of these random variables has zero expected value as well. Hence, $\mu(i) = 0$.

b) We compute the autocovariance for lag $h = 0$, i.e. $i = j$. Using the fact that $\mathrm{E}(X_i X_j) = 0$ if $i \neq j$, we find

$$\begin{aligned}\gamma(i,i) &= \mathrm{E}\left((W_{i-1} + 2W_i + W_{i+1})(W_{i-1} + 2W_i + W_{i+1})\right)\\ &= \mathrm{E}(W_{i-1}^2) + 4\mathrm{E}(W_i^2) + \mathrm{E}(W_{i+1}^2) = \sigma^2 + 4\sigma^2 + \sigma^2 = 6\sigma^2.\end{aligned}$$

Analogously, we find for $j = i + 1$

$$\begin{aligned}\gamma(i,i+1) &= \mathrm{E}\left((W_{i-1} + 2W_i + W_{i+1})(W_i + 2W_{i+1} + W_{i+2})\right)\\ &= 2\mathrm{E}(W_i^2) + 2\mathrm{E}(W_{i+1}^2) = 2\sigma^2 + 2\sigma^2 = 4\sigma^2.\end{aligned}$$

Finally we find for $j = i + 2$

$$\begin{aligned}\gamma(i,i+2) &= \mathrm{E}\left((W_{i-1} + 2W_i + W_{i+1})(W_{i+1} + 2W_{i+2} + W_{i+3})\right)\\ &= \mathrm{E}(W_{i+1}^2) = \sigma^2.\end{aligned}$$

Thus we summarize that

$$\gamma(i,j) = \begin{cases} 6\sigma^2 & \text{if } j = i \\ 4\sigma^2 & \text{if } j = i+1 \\ \sigma^2 & \text{if } j = i+2 \\ 0 & \text{else.} \end{cases}$$

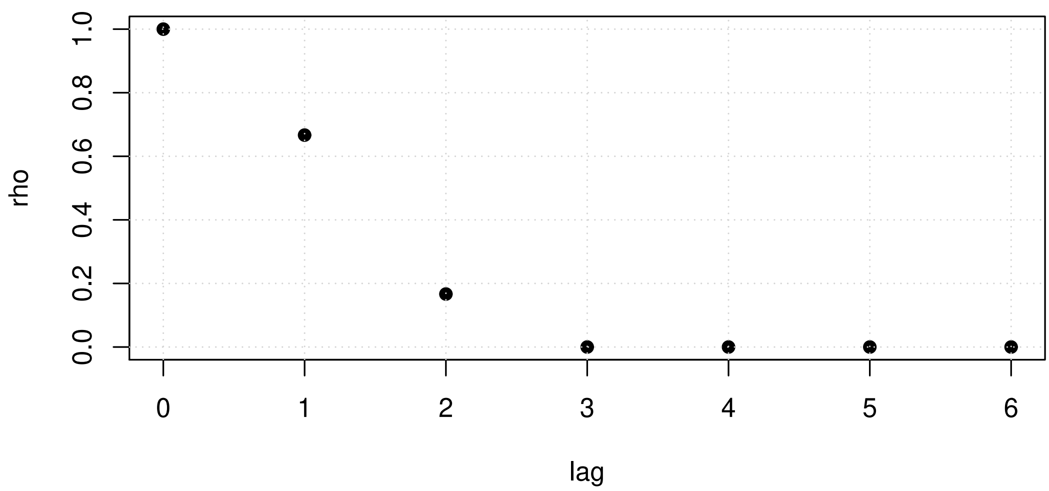
The autocorrelation is obtained by $\rho(i,j) = \gamma(i,j)/\gamma(i,i)$, since the autocovariance only depends on the lag of the times $h = i - j$. This results in

$$\rho(i,j) = \begin{cases} 1 & \text{if } j = i \\ \frac{2}{3} & \text{if } j = i+1 \\ \frac{1}{6} & \text{if } j = i+2 \\ 0 & \text{else.} \end{cases}$$

c) The autocorrelation function as a function of the lag h reads

$$\rho(h) = \begin{cases} 1 & \text{if } h = 0 \\ \frac{2}{3} & \text{if } h = 1 \\ \frac{1}{6} & \text{if } h = 2 \\ 0 & \text{else.} \end{cases}$$

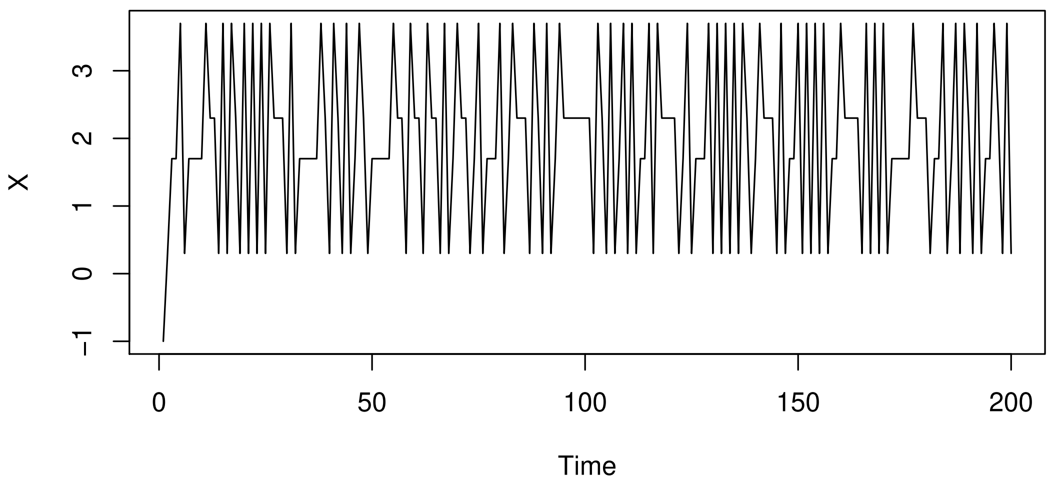
Thus we obtain a plot like this



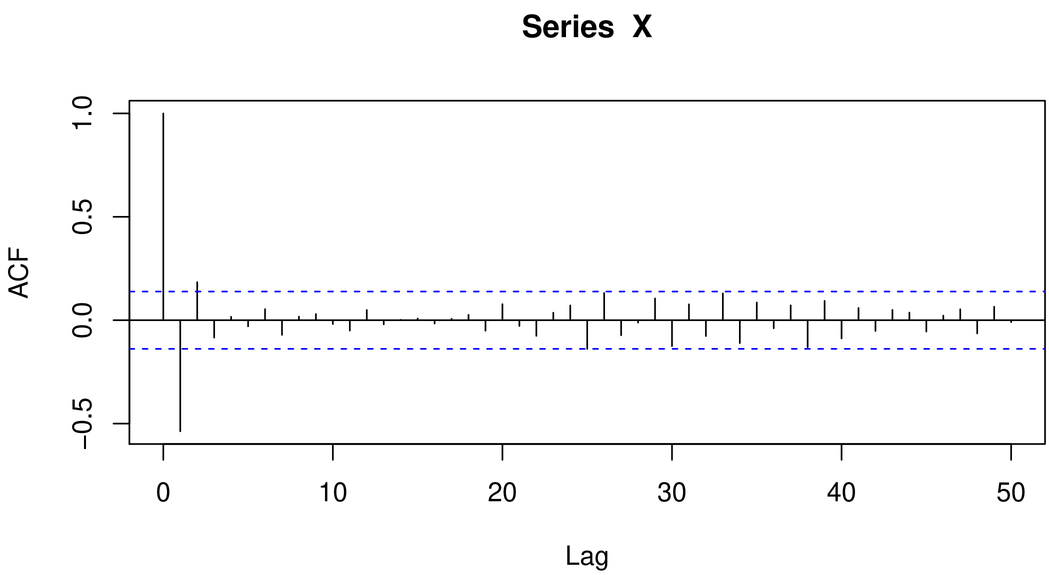
Solution 12.3

a)
```
# Create the process Dk
D = 2 * (rbinom(200, size = 1, p = 0.5) - 0.5)

# Create the process Xk
X = rep(-1, 1, 200)
```

```
for (k in 2:200) {
    X[k] = 2 + D[k] - 0.7 * D[k - 1]
}
X = ts(X)

plot(X)
```



b)
```
acf(X, lag.max = 50)
```



c) For the mean, we compute

$$\mu(i) = \mathrm{E}(X_i) = a + \mathrm{E}(D_k) + b\mathrm{E}(D_{k-1}) = a.$$

When computing the autocovariance, only those pairs of X_k and X_j have to considered that contain D_ks with common indices, since the coin tossing is independent. So we have to consider $k = j$ and $k = j + 1$- For all other cases, the autocovariance is zero.

$$\begin{aligned}\gamma(k,k) &= \mathrm{E}\left((D_k + bD_{k-1})(D_k + bD_{k-1})\right) \\ &= \mathrm{E}\left(D_k^2 + 2bD_{k-1}D_k + b^2 D_{k-1}^2\right) \\ &= 1 + b^2\end{aligned}$$

The last equality follows from the fact that $\mathrm{E}(D_k D_{k-1}) = 0$ and $\mathrm{E}(D_k^2) = 1$. Now we compute

$$\begin{aligned}\gamma(k,k+1) &= \mathrm{E}\left((D_k + bD_{k-1})(D_{k+1} + bD_k)\right) \\ &= \mathrm{E}\left(D_k D_{k-1} + bD_k D_{k+1} + bD_{k-1}D_k + bD_k^2\right) \\ &= b\end{aligned}$$

Here all terms in the product that have zero expectation have not been computed. Eventually we find that

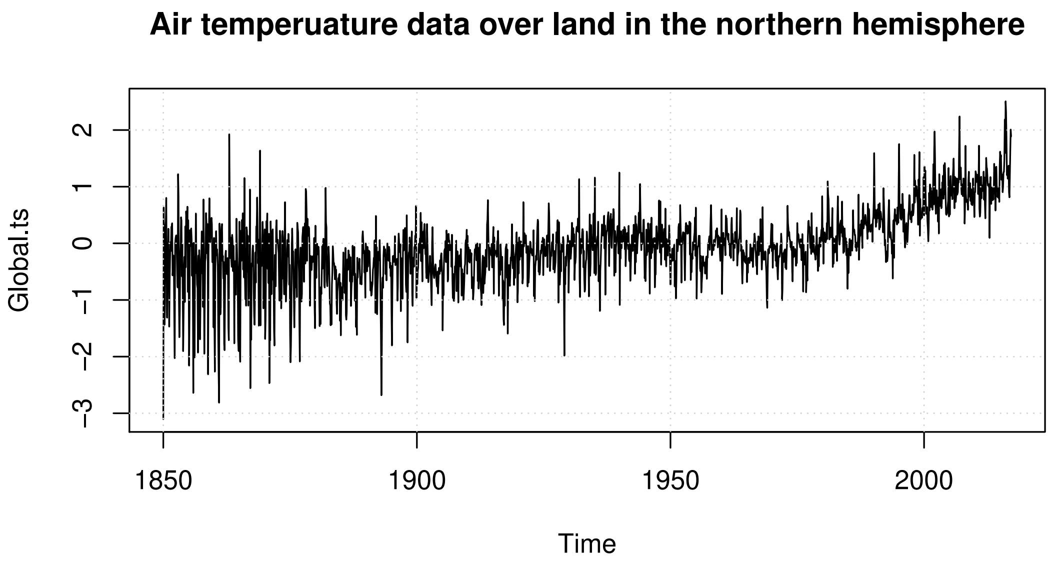
$$\rho(k,j) = \begin{cases} 1 & \text{if } j = k \\ \frac{b}{1+b^2} & \text{if } j = k+1 \\ 0 & \text{else.} \end{cases}$$

We have $\hat{\rho}(1) = -0.54$ and $\rho(1) = -0.47$.

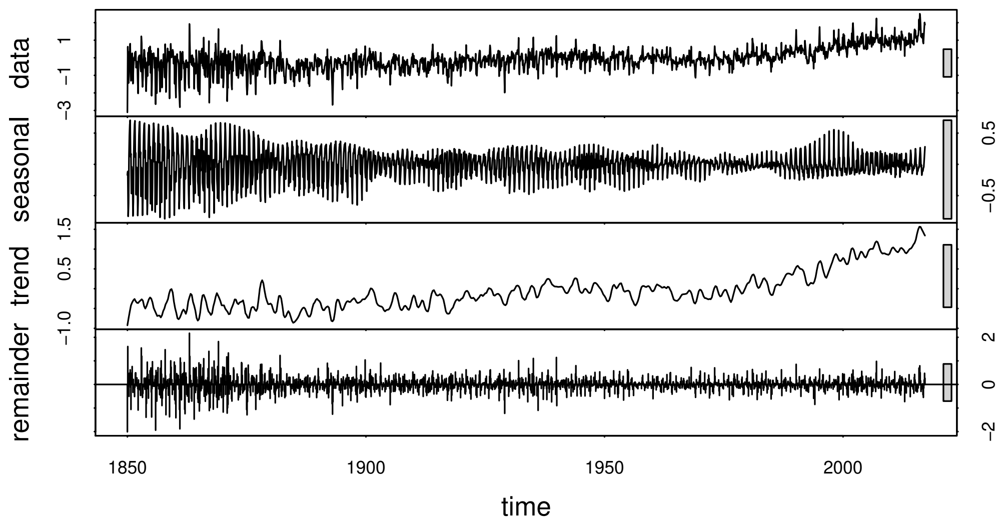
d) Yes. The mean is constant and the autocorrelation sequence only depends on the lag of the arguments.

Solution 12.4

a)
```
load("Daten/GlobalTemperature.Rdata")
plot(Global.ts, main = "Air temperuature data over land in the northe
grid()
```

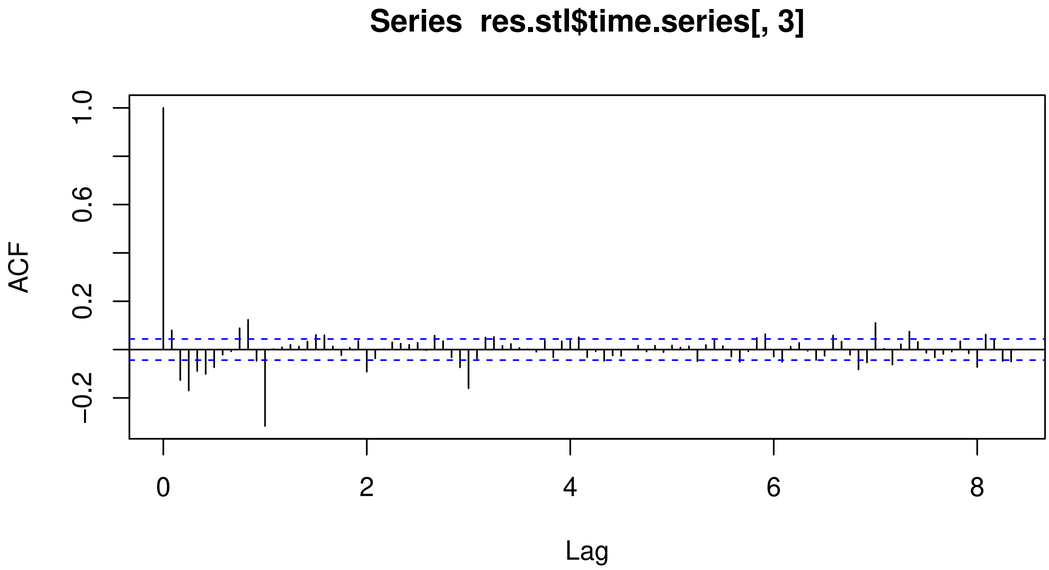


b)
```r
res.stl = stl(Global.ts, s.window = 10)
plot(res.stl)
```



The decomposition yields a trend component with a significant increase between the 70s and now. In the 18th century the data seems to have higher variance. This could be due to inferior measurment equipment etc. during that time.

c)
```
acf(res.stl$time.series[, 3], lag.max = 100)
```



The correlogram shows surprisingly low (linear) dependencies. A significant dependency can be seen at lag $h = 12$ which corresponds to remaining seasonal components in the series.