Simple Time Series Models

We can define time series models using simple constructions. Consider first the sequence of random variables $\{Z_t\}_{t\geq 1}$, where

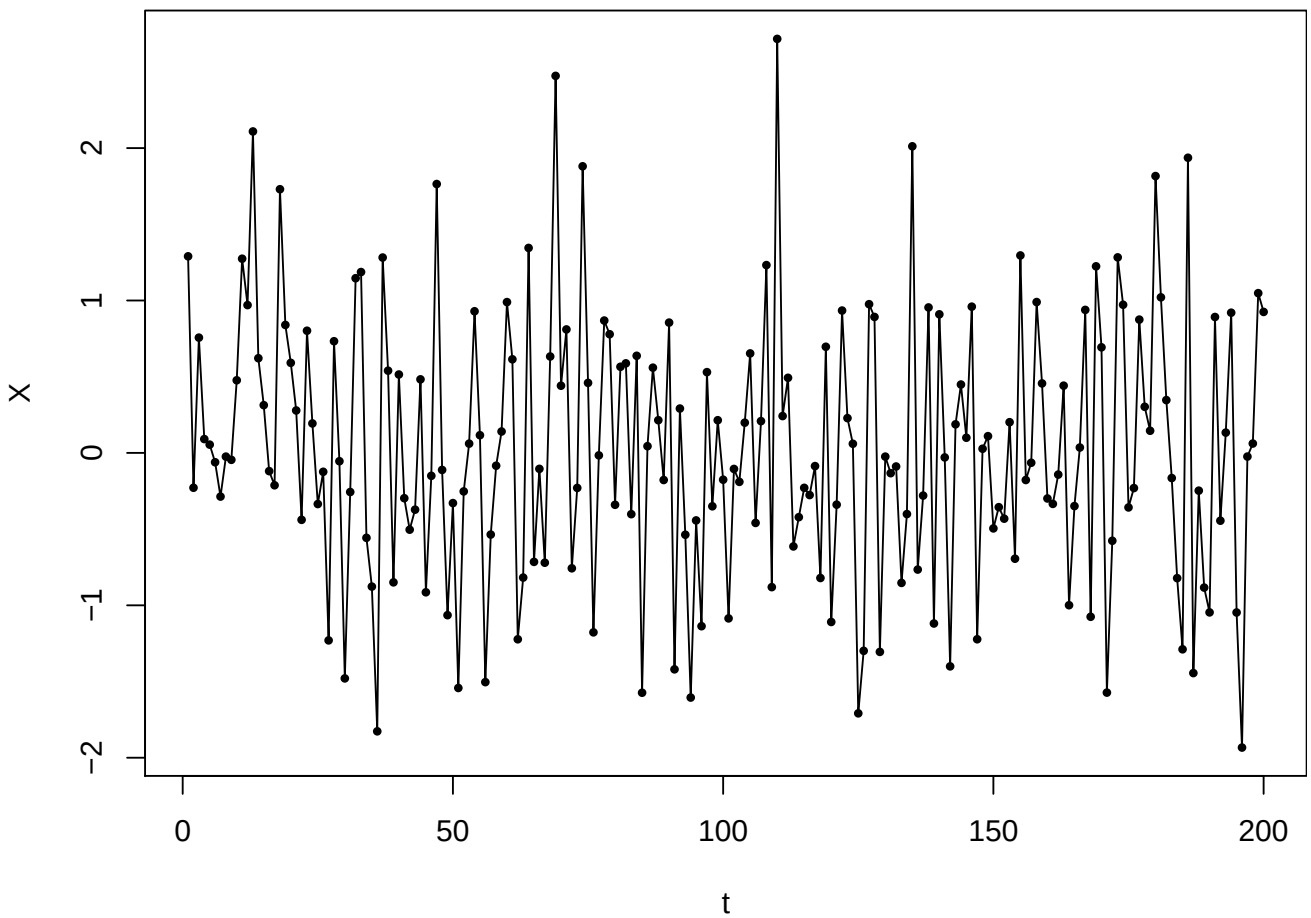
$$Z_t \sim Normal(0,1)$$

with the random variables (mutually) independent. Using these variables, we may define the simple time series $\{X_t\}_{t\geq 1}$ by

$$X_t = Z_t.$$

We can simulate $n = 200$ of these values, and plot them in R using elementary commands.

```
set.seed(2101)                    #Set the random number generator seed value
n<-200                            #Set the sample size
Z<-rnorm(n)                       #Generate the random variables
X<-Z                              #Define the time series X
par(mar=c(4,4,1,0))               #Set up the plotting margins
plot(1:n,X,xlab='t',type='l')   #Constrct a line plot
points(1:n,X,pch=19,cex=0.5)    #Add the points on top of the line
```



For this series, as the variables are independent, there is no correlation between successive values in the series, that is

$$\text{Corr}[X_t, X_{t+1}] = 0 \qquad t = 1, 2, \ldots, n-1$$

We can compute the sample correlation between successive values as follows

```
cor(X[1:(n-1)],X[2:n])

+ [1] -0.01767394
```

The function `cor` applied to the two vectors in `R` computes the correlation between successive values using the formula

$$\frac{\sum_{t=1}^{n-1}(x_t-\overline{x}_1)(x_{t+1}-\overline{x}_2)}{\sqrt{\sum_{t=1}^{n-1}(x_t-\overline{x}_1)^2\sum_{t=2}^{n}(x_t-\overline{x}_2)^2}}$$

where

$$\overline{x}_1=\frac{1}{n-1}\sum_{t=1}^{n-1}x_t \qquad \overline{x}_2=\frac{1}{n-1}\sum_{t=1}^{n-1}x_{t+1}$$

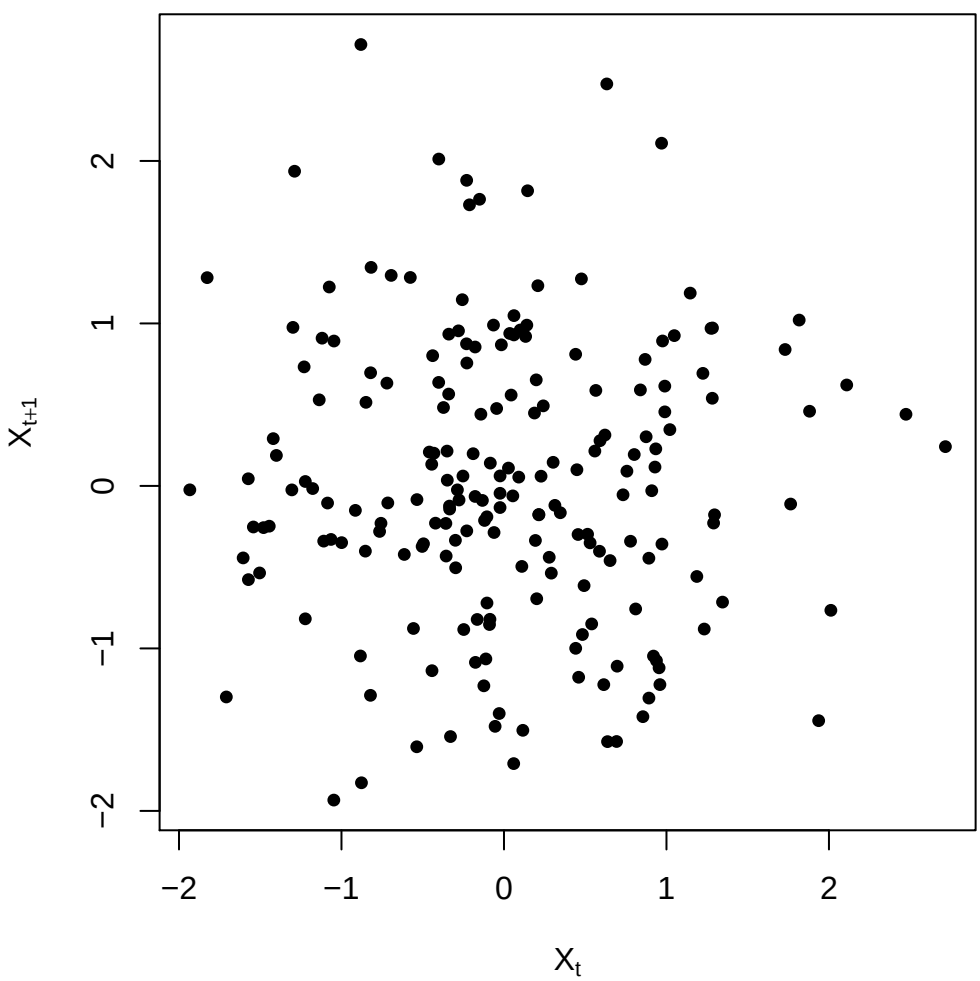
which we can check as follows

```
Xmean1<-mean(X[1:(n-1)])
Xmean2<-mean(X[2:n])
topval<-sum((X[1:(n-1)]-Xmean1)*(X[2:n]-Xmean2))
botval1<-sum((X[1:(n-1)]-Xmean1)^2)
botval2<-sum((X[2:n]-Xmean2)^2)
topval/sqrt(botval1*botval2)

+ [1] -0.01767394
```

A plot of X_{t+1} versus X_t reveals no real pattern:

```
par(mar=c(4,4,1,0),pty='s')
plot(X[1:(n-1)],X[2:n],pch=19,cex=0.75,xlab=expression(X[t]),ylab=expression(X[t+1]))
```

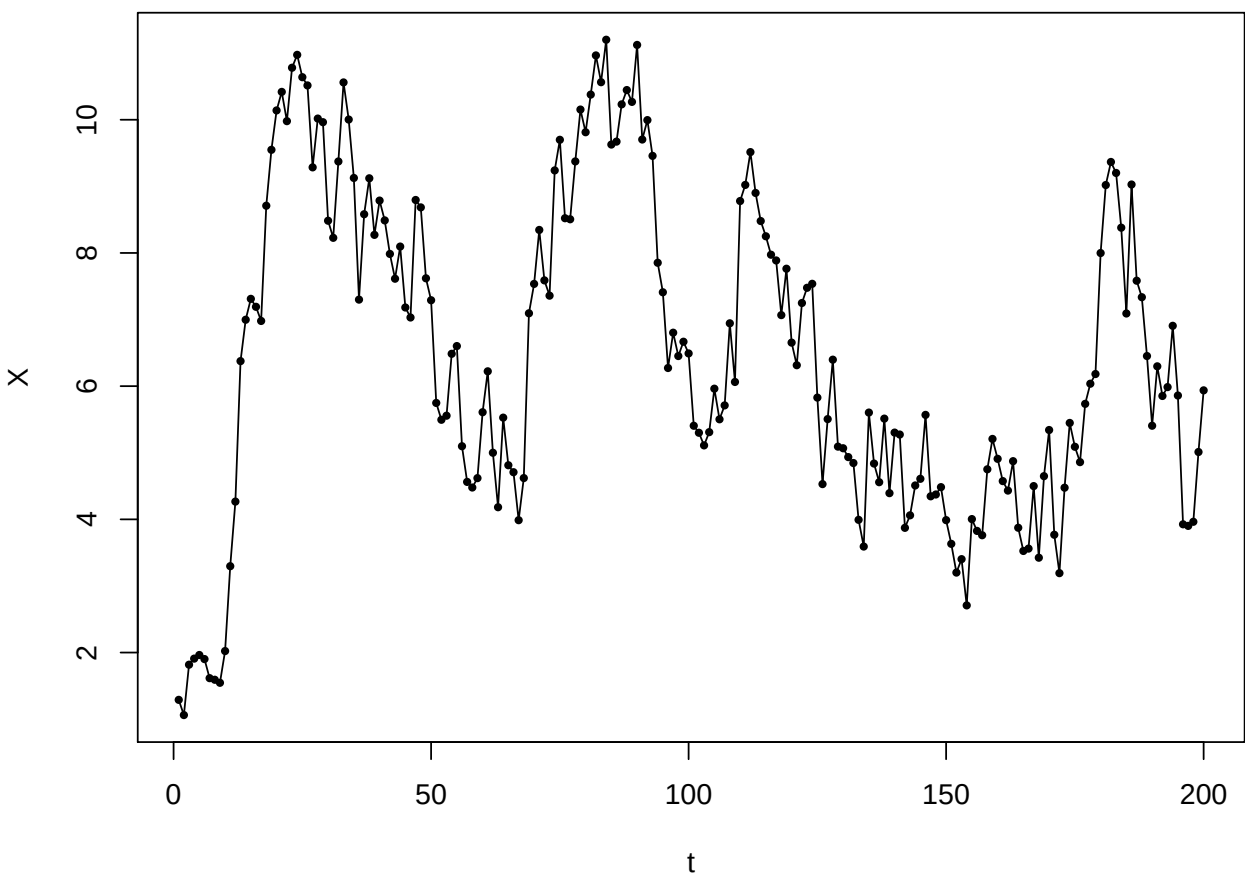


We can use the $\{Z_t\}$ to define different time series in simple ways:

(i) **Random walk:** Suppose we define $Z_1 = X_1$, and then for $t \geq 2$,

$$X_t = X_{t-1} + Z_t$$

```
set.seed(2101)                      #Set the random number generator seed value
n<-200                              #Set the sample size
Z<-rnorm(n)                         #Generate the random variables
X<-rep(0,n)
X[1]<-Z[1]                          #Define the time series X
for(i in 2:n){
    X[i]<-X[i-1]+Z[i]
}
par(mar=c(4,4,1,0))                 #Set up the plotting margins
plot(1:n,X,xlab='t',type='l')       #Constrct a line plot
points(1:n,X,pch=19,cex=0.5)        #Add the points on top of the line
```

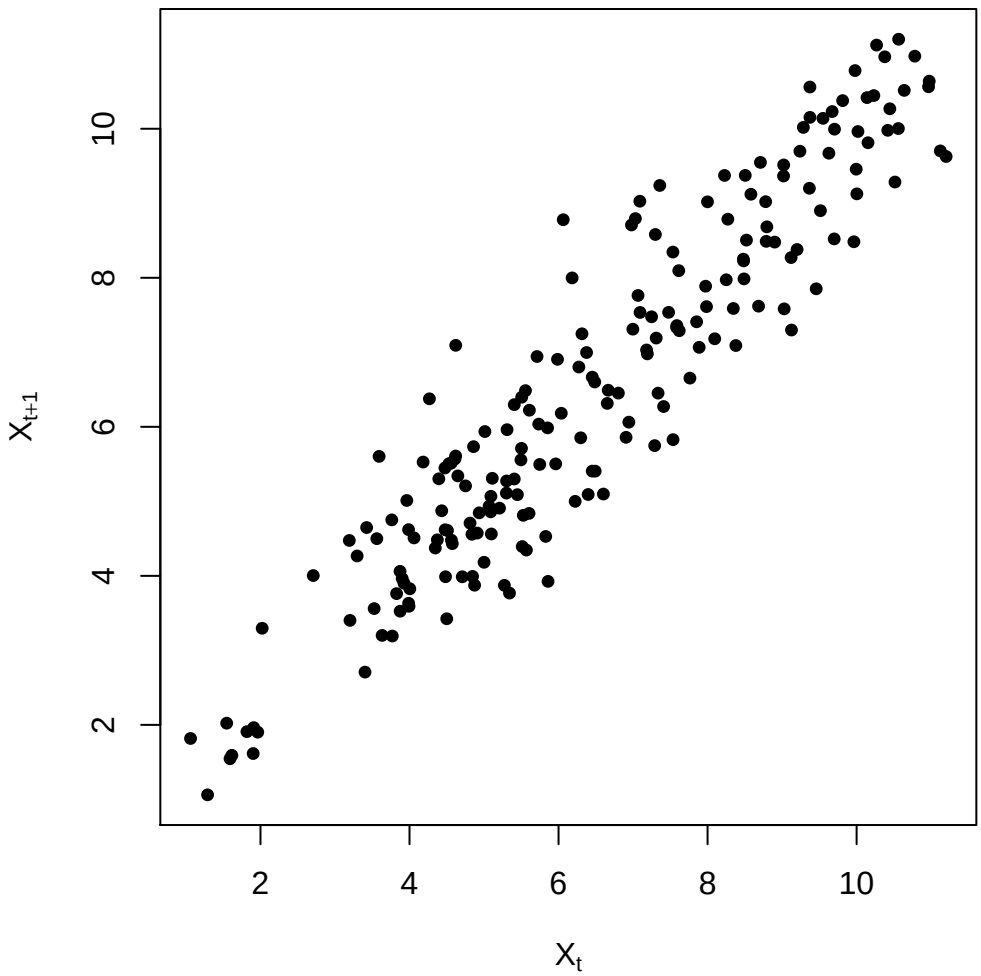



```
cor(X[1:(n-1)],X[2:n])

+ [1] 0.9346292
```

Now we see a very high sample correlation of 0.9346292. A plot of X_{t+1} versus X_t reveals a strong relationship:

```r
par(mar=c(4,4,1,0),pty='s')
plot(X[1:(n-1)],X[2:n],pch=19,cex=0.75,xlab=expression(X[t]),ylab=expression(X[t+1]))
```



The time series process $\{X_t\}$ is termed a *random walk*, here derived from or *driven by* a Normal series (also called a *Gaussian white noise process*) with variance 1. Some of the properties of $\{X_t\}$ are easy to deduce. For example, we have

$$X_t = X_{t-1} + Z_t = (X_{t-2} + Z_{t-1}) + Z_t = ((X_{t-3} + Z_{t-2}) + Z_{t-1}) + Z_t = \cdots = \sum_{i=1}^{t} Z_i$$

with the $\{Z_t\}$ independent and $Normal(0,1)$, we must have that

$$\mathbb{E}[X_t] = 0 \quad \forall t \qquad\qquad \text{Var}[X_t] = t \quad t = 1, 2, 3, \ldots.$$

That is, the variance of X_t grows linearly with t. For the covariance between X_t and X_{t+1}, we have that

$$\text{Cov}[X_t, X_{t+1}] = \text{Cov}[X_t, X_t + Z_{t+1}] = \mathbb{E}[X_t(X_t + Z_{t+1})] = \mathbb{E}[X_t^2 + X_t Z_{t+1}] = \mathbb{E}[X_t^2] + \mathbb{E}[X_t Z_{t+1}]$$

as $\mathbb{E}[X_t] = \mathbb{E}[Z_t] = 0$ for all t. By independence, we have that $\mathbb{E}[X_t Z_{t+1}] = \mathbb{E}[X_t]\mathbb{E}[Z_{t+1}] = 0$ so therefore

$$\text{Cov}[X_t, X_{t+1}] = \mathbb{E}[X_t^2] = \text{Var}[X_t] = t.$$

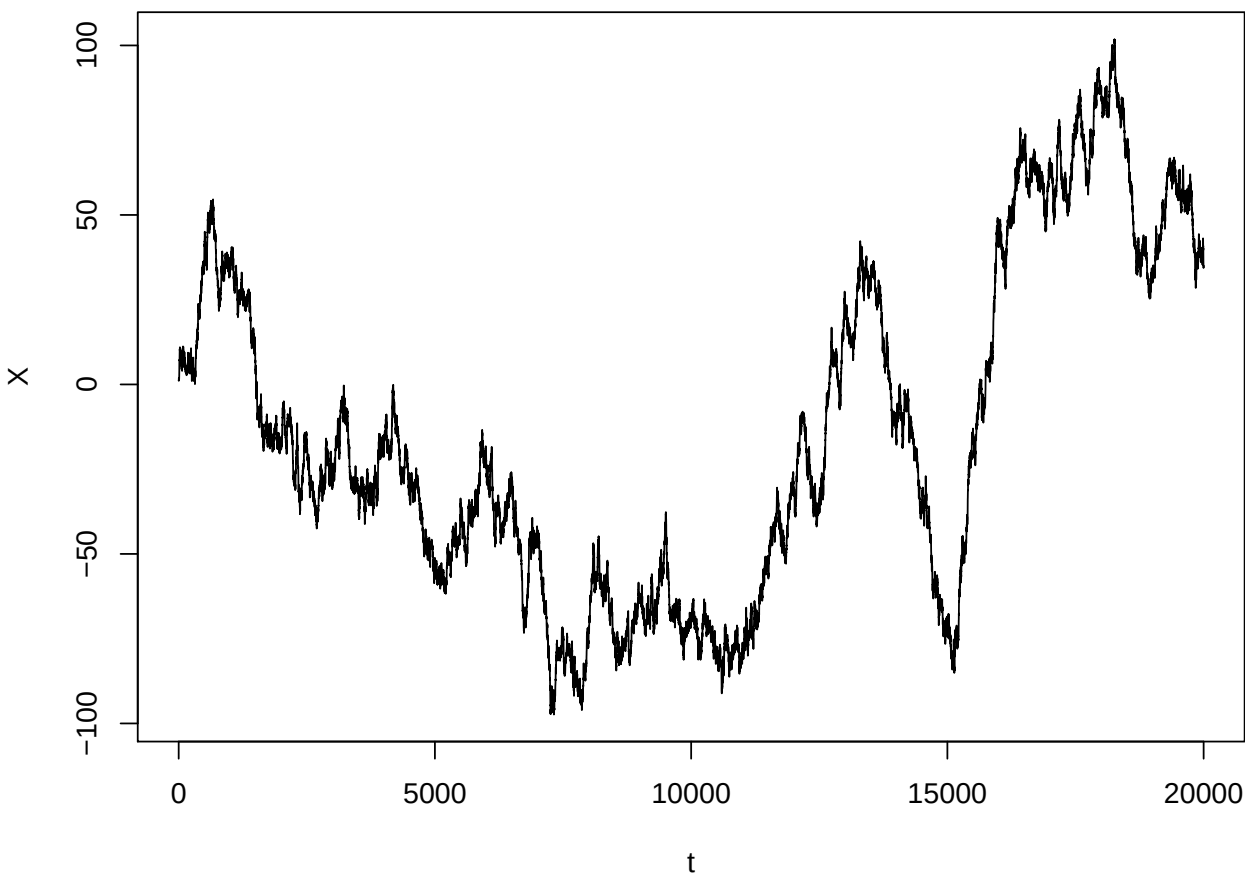
Hence, for the theoretical correlation, we have

$$\text{Corr}[X_t, X_{t+1}] = \frac{\text{Cov}[X_t, X_{t+1}]}{\sqrt{\text{Var}[X_t]\text{Var}[X_{t+1}]}} = \frac{t}{\sqrt{t(t+1)}} = \frac{1}{\sqrt{1+t^{-1}}}.$$

That is, the variance, covariance and correlation are *not constant* in t.

The process looks 'stable' in level in the plot above, but for a much larger sample size, the picture is different

```
set.seed(2101)
n<-20000
Z<-rnorm(n)
X<-rep(0,n)
X[1]<-Z[1]
for(i in 2:n){
    X[i]<-X[i-1]+Z[i]
}
par(mar=c(4,4,1,0))
plot(1:n,X,xlab='t',type='l')
```



Elementary properties of the Normal distribution yield that, for all t, $X_t \sim Normal(0, t)$, so it can be easily computed that for any threshold $x_0 > 0$,

$$P[|X_n| > x_0] = 2(1 - \Phi(x_0/\sqrt{n})) \longrightarrow 1$$

as $n \longrightarrow \infty$. This result implies that the random walk will, eventually, take on unboundedly large values. Thus, the process has an inherent 'instability' in the long run.

Many of the calculations we have carried out above have been facilitated by the use of the properties of the Normal distribution. However, the random walk construction applies even if the driving series $\{Z_t\}$ is non-Normal, and many of the same properties (instability, non-constant variance etc.) still are present.

(ii) **Moving average process:** Recall, for the random walk we defined

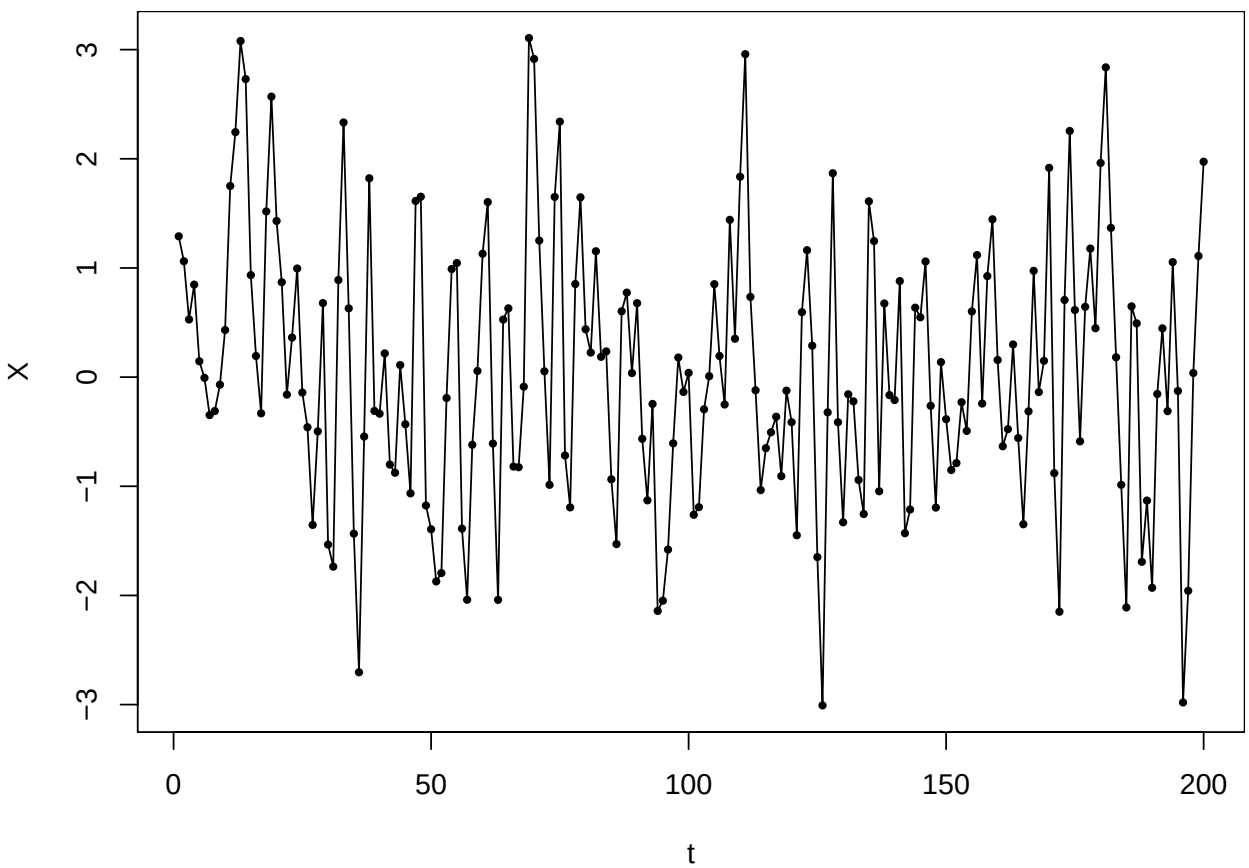
$$X_t = \sum_{i=1}^{t} Z_i = Z_t + Z_{t-1} + \cdots Z_2 + Z_1$$

Suppose instead we define, for $t \geq 2$

$$X_t = Z_t + Z_{t-1}.$$

truncating the previous sum at the second term. Now the simulated series looks somewhat different

```
set.seed(2101)
n<-200;Z<-rnorm(n);X<-rep(0,n)
X[1]<-Z[1]
for(i in 2:n){
    X[i]<-Z[i]+Z[i-1]
}
par(mar=c(4,4,0,0))
plot(1:n,X,xlab='t',type='l')
points(1:n,X,pch=19,cex=0.5)
```




```
cor(X[1:(n-1)],X[2:n])

+ [1] 0.4282162
```

The correlation is reduced compared to the random walk.

It is clear from elementary properties of the Normal distribution that $X_t \sim Normal(0, 2)$ for all $t \geq 2$. Also, for the covariance, we have

$$\text{Cov}[X_t, X_{t+1}] = \mathbb{E}[X_t X_{t+1}] = \mathbb{E}[(Z_t + Z_{t-1})(Z_{t+1} + Z_t)] = \mathbb{E}[Z_t Z_{t+1}] + \mathbb{E}[Z_t^2] + \mathbb{E}[Z_{t-1} Z_{t+1}] + \mathbb{E}[Z_{t-1} Z_t].$$

However, as the Z_ts are mutually independent, $\mathbb{E}[Z_t Z_s] = 0$ if $t \neq s$. Therefore

$$\text{Cov}[X_t, X_{t+1}] = \mathbb{E}[Z_t^2] = 1 \quad \Longrightarrow \quad \text{Corr}[X_t, X_{t+1}] = \frac{1}{\sqrt{2 \times 2}} = \frac{1}{2}.$$

as $Z_t \sim Normal(0, 1)$. Notice that, by a similar calculation

$$\text{Cov}[X_t, X_{t+2}] = \mathbb{E}[X_t X_{t+2}] = \mathbb{E}[(Z_t + Z_{t-1})(Z_{t+2} + Z_{t+1})] = 0$$

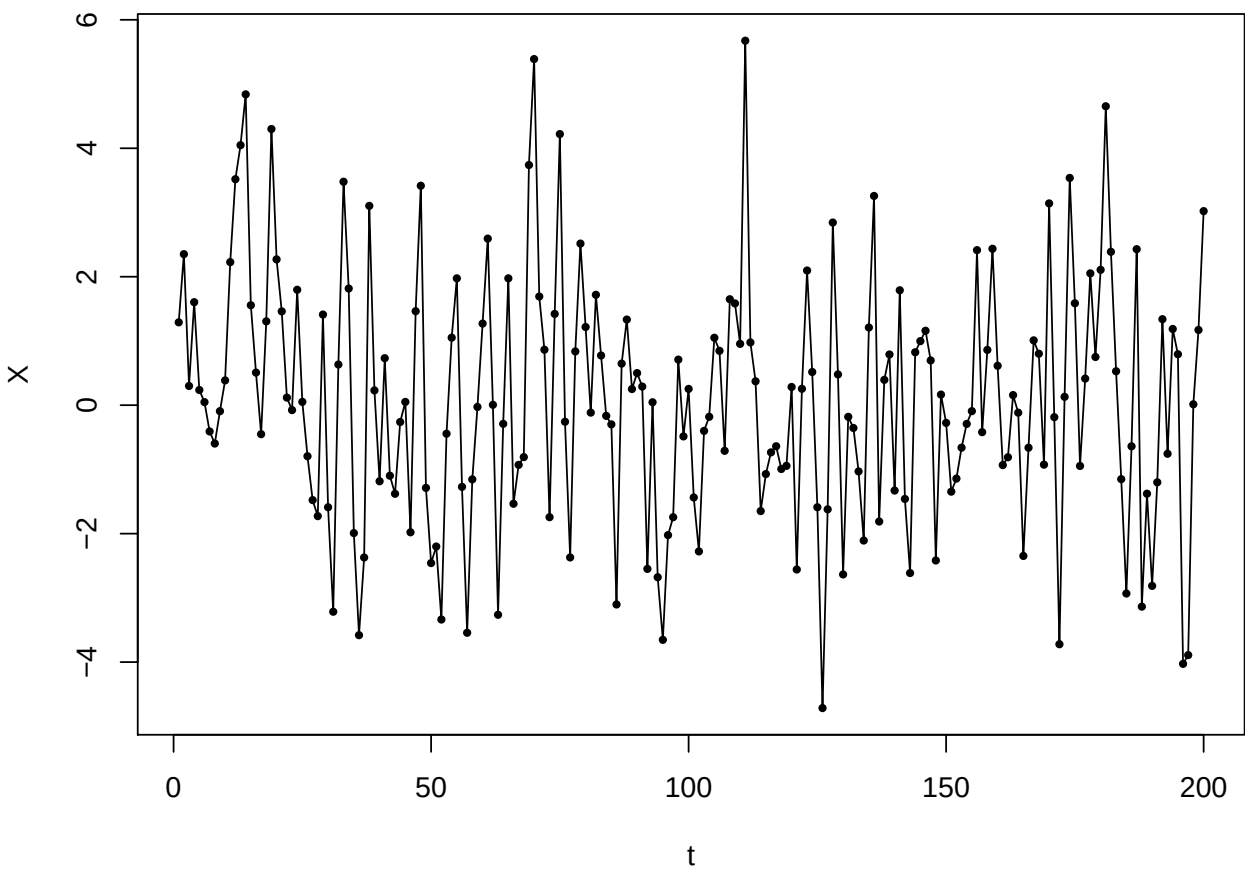
as $(Z_t + Z_{t-1})$ is independent of $(Z_{t+2} + Z_{t+1})$.

We can generalize the construction by allowing an extra parameter into the specification as follows: Suppose instead we define, for $t \geq 2$

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

for some value θ. In the following simulation $\theta = 2$:

```
set.seed(2101)
n<-200;Z<-rnorm(n);X<-rep(0,n)
theta<-2.0
X[1]<-Z[1]
for(i in 2:n){
    X[i]<-Z[i]+theta*Z[i-1]
}
par(mar=c(4,4,0,0))
plot(1:n,X,xlab='t',type='l')
points(1:n,X,pch=19,cex=0.5)
```




```
cor(X[1:(n-1)],X[2:n])

+ [1] 0.3350199
```

From elementary properties of the Normal distribution, we now have that $X_t \sim Normal(0, 1+\theta^2)$ for all $t \geq 2$. Also, for the covariance, we have by a similar calculation to above that

$$\text{Cov}[X_t, X_{t+1}] = \mathbb{E}[X_t X_{t+1}] = \theta \quad \Longrightarrow \quad \text{Corr}[X_t, X_{t+1}] = \frac{\theta}{1+\theta^2}$$

whereas, by a similar calculation

$$\text{Cov}[X_t, X_{t+h}] = 0$$

for $h \geq 2$. If $\theta = 2$, the correlation for $h = 1$ is therefore $\text{Corr}[X_t, X_{t+1}] = \theta/(1+\theta^2) = 0.4$.

This model is termed a *moving average* model, which can in fact be generalized further.

(iii) **Autoregressive model:** Recall that the random walk model can be written for $t \geq 2$ as

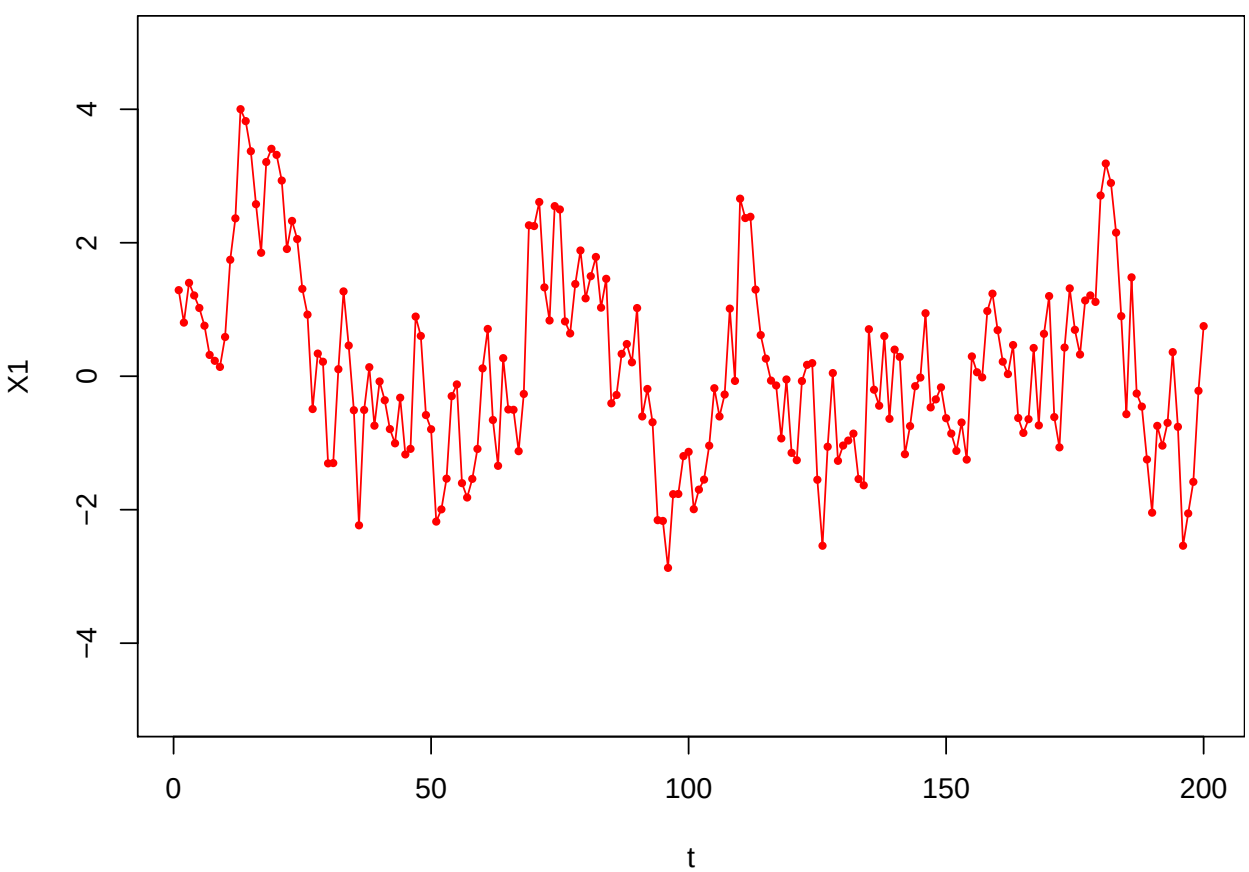
$$X_t = X_{t-1} + Z_t.$$

Suppose that instead we define the process by

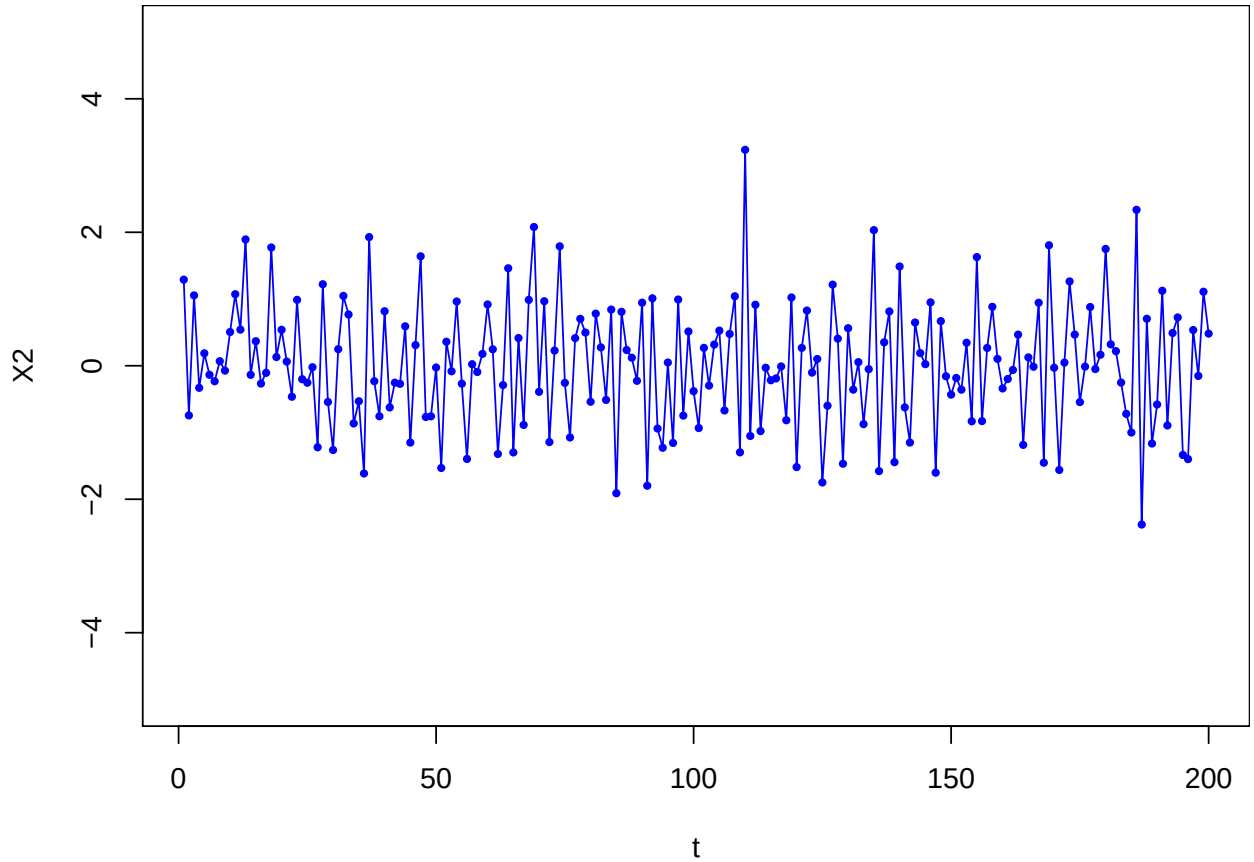
$$X_t = \phi X_{t-1} + Z_t.$$

for some parameter ϕ. In the following simulation, ϕ takes two values, $0.8, -0.4$, with the resulting series plotted in red and blue respectively.

```
set.seed(2101)
n<-200;Z<-rnorm(n);X1<-X2<-rep(0,n)
phi1<-0.8;phi2<--0.4
X1[1]<-X2[1]<-Z[1]
for(i in 2:n){
    X1[i]<-phi1*X1[i-1]+Z[i]
    X2[i]<-phi2*X2[i-1]+Z[i]
}
par(mar=c(4,4,0,0))
plot(1:n,X1,xlab='t',type='l',col='red',ylim=range(-5,5))
points(1:n,X1,pch=19,cex=0.5,col='red')
```



```
plot(1:n,X2,xlab='t',type='l',col='blue',ylim=range(-5,5))
points(1:n,X2,pch=19,cex=0.5,col='blue')
```




```
cor(X1[1:(n-1)],X1[2:n])

+ [1] 0.7793816

cor(X2[1:(n-1)],X2[2:n])

+ [1] -0.3628397
```

Visually, it appears that the red series is more variable than the blue series, and this is in fact the case:

```
var(X1);var(X2)

+ [1] 1.912765
+ [1] 0.8693285
```

In addition the correlation patterns are somewhat different. Formal computation of the covariances and correlations are a little more challenging; heuristically, we have from above that by independence of X_{t-1} and Z_t,

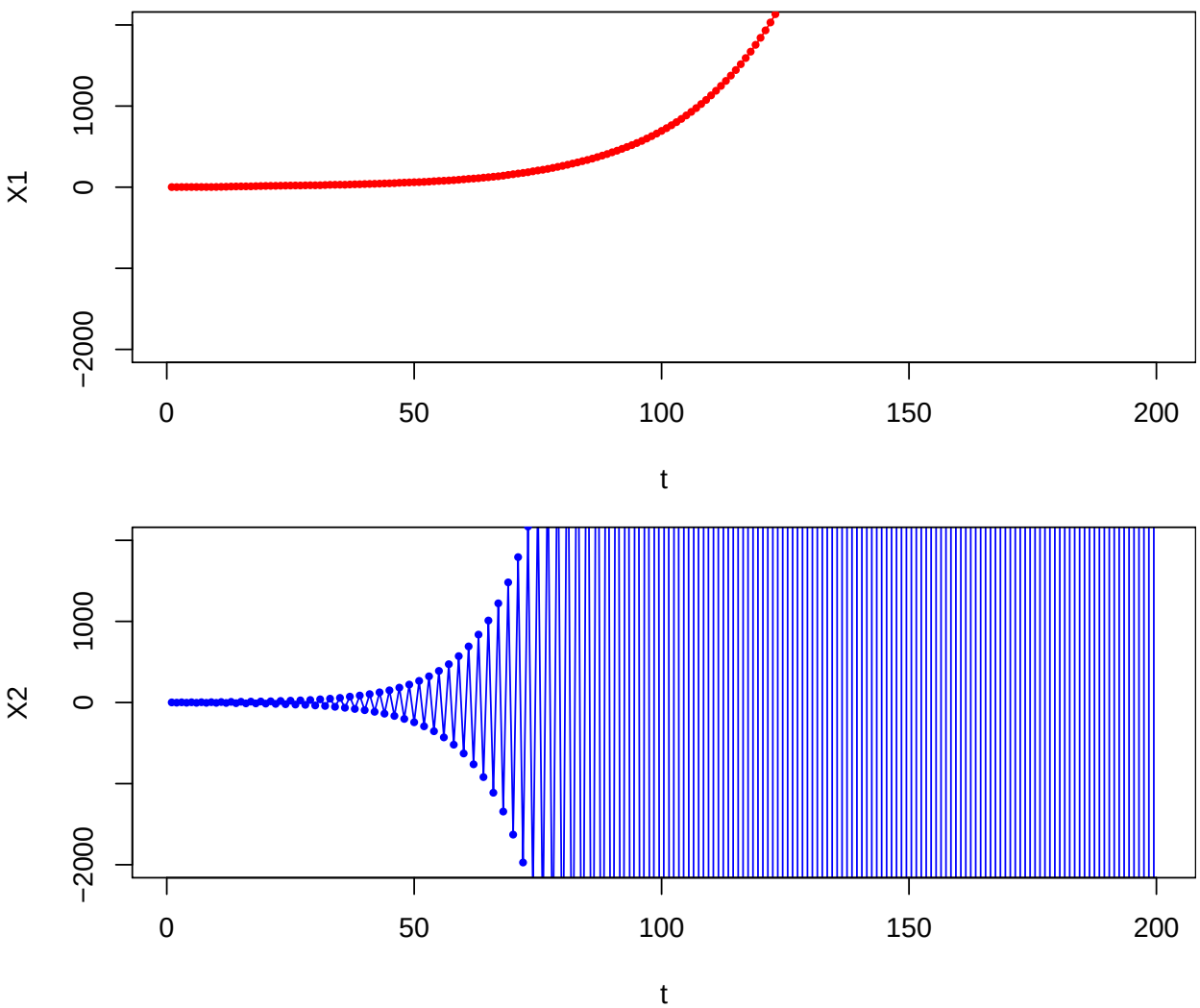
$$\text{Var}[X_t] = \phi^2\text{Var}[X_{t-1}] + \text{Var}[Z_t] = \phi^2\text{Var}[X_{t-1}] + 1$$

from which, under certain conditions, it can be shown that, for all $t \geq 2$,

$$\text{Var}[X_t] = \frac{1}{(1-\phi^2)} \qquad \text{Cov}[X_t, X_{t+1}] = \phi \qquad \text{Corr}[X_t, X_{t+1}] = \frac{\phi}{1-\phi^2}.$$

Finally, in the following simulation, ϕ takes two values, $1.05, -1.1$, with the resulting series plotted in red and blue respectively.

```r
set.seed(2101)
n<-200
Z<-rnorm(n)
X1<-X2<-rep(0,n)
phi1<-1.05
phi2<--1.1
X1[1]<-X2[1]<-Z[1]
for(i in 2:n){
    X1[i]<-phi1*X1[i-1]+Z[i]
    X2[i]<-phi2*X2[i-1]+Z[i]
}
par(mar=c(4,4,1,0),mfrow=c(2,1))
plot(1:n,X1,xlab='t',type='l',col='red',ylim=range(-2000,2000))
points(1:n,X1,pch=19,cex=0.5,col='red')
plot(1:n,X2,xlab='t',type='l',col='blue',ylim=range(-2000,2000))
points(1:n,X2,pch=19,cex=0.5,col='blue')
```



Note the huge vertical scale; clearly, the series both are explosive.

This model is termed a *autoregressive* model of order 1.