Ma3 - STAT220 - Sigbjørn Fjelland

Sigbjørn Fjelland

2022-10-10

```
library(tinytex)
library(igraph)
```

```
##
## Attaching package: 'igraph'

## The following objects are masked from 'package:stats':
##
##     decompose, spectrum

## The following object is masked from 'package:base':
##
##     union
```

```
library(expm)
```

```
## Loading required package: Matrix

##
## Attaching package: 'expm'

## The following object is masked from 'package:Matrix':
##
##     expm
```

```
library(markovchain)
```

```
## Package:  markovchain
## Version:  0.9.0
## Date:     2022-07-01
## BugReport: https://github.com/spedygiorgio/markovchain/issues
```

```
library(diagram)
```

```
## Loading required package: shape
```

```
library(pracma)
```

```
##
## Attaching package: 'pracma'

## The following objects are masked from 'package:expm':
##
##     expm, logm, sqrtm

## The following objects are masked from 'package:Matrix':
##
##     expm, lu, tril, triu
```

```r
set.seed(123)
```

Problem 3.1

a)

$$
\begin{aligned}
P(X_t = 6) &= \frac{1}{6} \\
P(X_1 = 6 \cap X_2 = 6) &= \left(\frac{1}{6}\right)^2 \\
&= \underline{\frac{1}{36}}
\end{aligned}
$$

b)

Throwing a dice where each throw is indipendent event with a given p, hence it is a bernoulli trail. The sum of indipendent bernoulli trail with failures until success is geometric:

$$X \sim ber(p) \Rightarrow \sum X = T_X \sim geom(p)$$

Since we need two equal in this trail we can construct a new compound sequence where two throws are equal. the process will be the same, hence the distributions will be the same, but the probabilitie p will change:

$$
\begin{aligned}
Y = X^2 \Rightarrow p_y = P(X \cap X) = p^2 \\
Y \sim ber(p^2) \Rightarrow \sum Y = T_Y \sim geom(p^2)
\end{aligned}
$$

c)

By def we have that the expected value and variance of an Geometric distribution is:

$$E[T_Y] = \frac{1}{p}$$
$$V[T_Y] = \frac{1-p}{p^2}$$

$$p_y = p^2 = \frac{1}{6^2}$$
$$= \frac{1}{36}$$

```
P.Y <- 1/6^2

E.T.G <- P.Y^(-1)
V.T.G <- (1-P.Y) / P.Y^2

cat('E(T) = ', E.T.G, ' and Var(T) = ',V.T.G)
```

```
## E(T) =  36  and Var(T) =  1260
```

for the exponential distribution (continious analog of Geometric distribution) variance and expectation is:

$$E[T_Y] = \frac{1}{p}$$
$$V[T_Y] = \frac{1}{p^2}$$

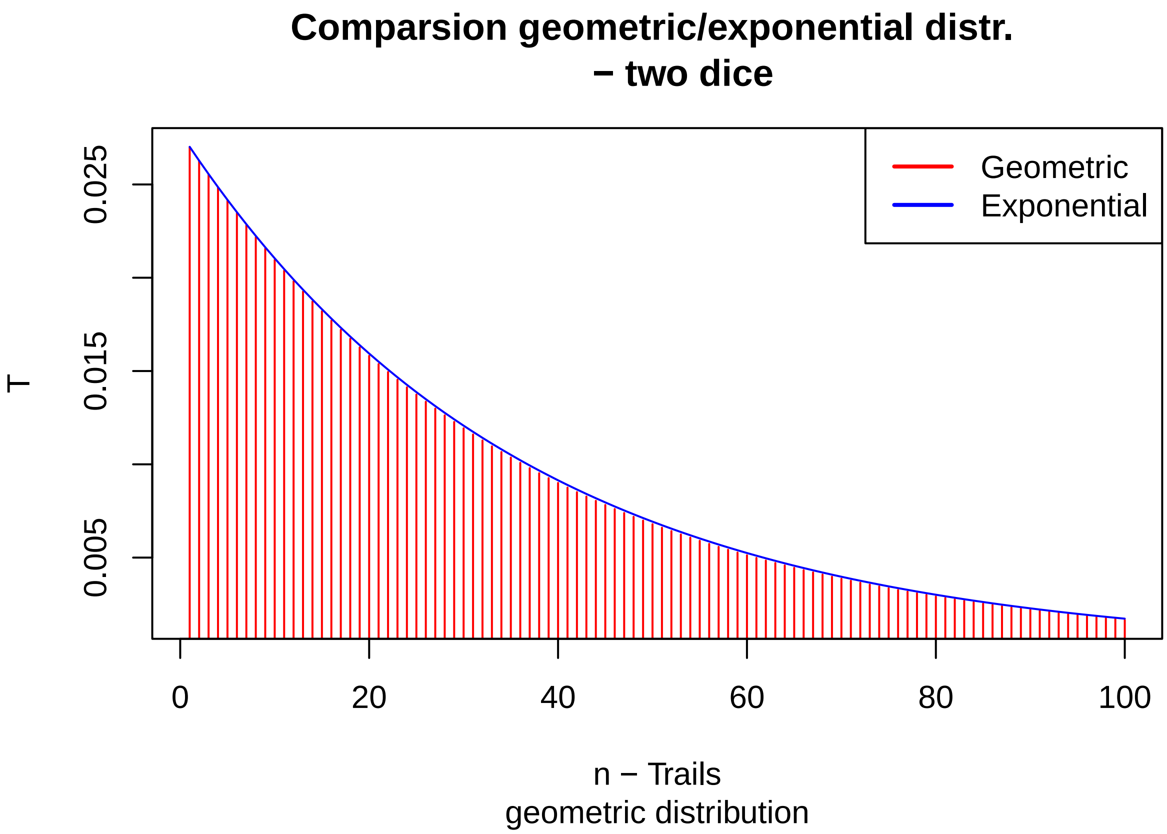
which yields the same expectantion and a larger variance:

```
E.T.E <- P.Y^(-1)
V.T.E <- (1) / P.Y^2

cat('E(T) = ', E.T.E, ' and Var(T) = ',V.T.E)
```

```
## E(T) =  36  and Var(T) =  1296
```

d)

```
n<-1:100

T.geom <-dgeom(x = n, prob = P.Y)
T.exp <- dexp(x = n, rate = P.Y)

plot(T.geom, type = 'h', col='red',xlab = 'n - Trails', ylab = 'T',
     main='Comparsion geometric/exponential distr.
     - two dice', sub='geometric distribution')
lines(T.exp, type = 'l', col='blue')
  legend("topright", col = c("red", "blue"),
  legend = c("Geometric", "Exponential"), lwd = 2)
```

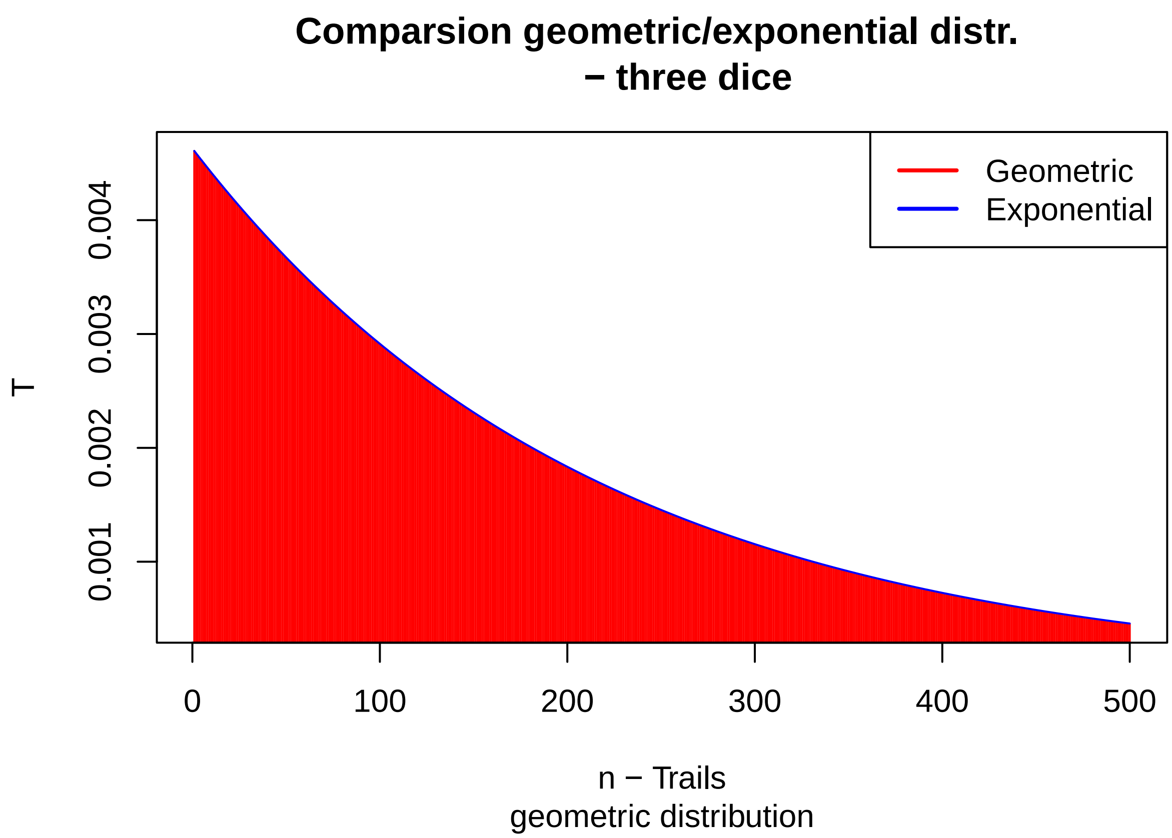


as we see there is a preaty fair chance that we hit withinthe first 10 -20 throws.

e) with three dice the probabilies decrease:

$$P(X_t = 6) = \frac{1}{6}$$
$$P(X_1 = 6 \cap X_2 = 6 \cap X_3 = 6) = \left(\frac{1}{6}\right)^3$$
$$= \underline{\frac{1}{216}}$$

```
P.3 <- (1/6)^3
n<-1:500

T.geom <-dgeom(x = n, prob = P.3)
T.exp <- dexp(x = n, rate = P.3)

plot(T.geom, type = 'h', col='red',xlab = 'n - Trails', ylab = 'T',
     main='Comparsion geometric/exponential distr.
     - three dice', sub='geometric distribution')
lines(T.exp, type = 'l', col='blue')
  legend("topright", col = c("red", "blue"),
  legend = c("Geometric", "Exponential"), lwd = 2)
```



We now need significantly many more throws to achive the goal:

geometric mean and variance:

```
E.T.G <- P.3^(-1)
V.T.G <- (1-P.3) / P.3^2
```

```r
cat('E(T) = ', E.T.G, ' and Var(T) = ',V.T.G)
```

```
## E(T) =  216  and Var(T) =  46440
```

exponential mean and variance:

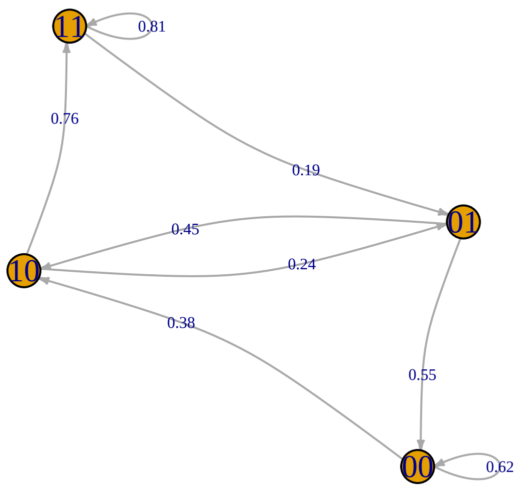
```r
E.T.Ex <- P.3^(-1)
V.T.Ex <- (1-P.3) / P.3^2

cat('E(T) = ', E.T.Ex, ' and Var(T) = ',V.T.Ex)
```

```
## E(T) =  216  and Var(T) =  46440
```

and we can also observe that the variance have converged between the distributions.

Problem 3.2

```
init<-rep(0.25,4)

S <-c(1:4) # state space

mP <- matrix(nrow = 4, ncol = 4, dimnames = list(c("00","01","10","11"),
                                                  c("00","01","10","11")))
mP[1,] <- c(0.62, 0.00, 0.38, 0.00)
mP[2,] <- c(0.55, 0.00, 0.45, 0.00)
mP[3,] <- c(0.00, 0.24, 0.00, 0.76)
mP[4,] <- c(0.00, 0.19, 0.00, 0.81)

mP.plot <- graph_from_adjacency_matrix(mP, weighted = "prob")
E(mP.plot)$prob <- ifelse(is.nan(E(mP.plot)$prob), NA, E(mP.plot)$prob)

plot(mP.plot, edge.label = round(E(mP.plot)$prob, 2), edge.arrow.size =
       .25, edge.curved=-0.2, edge.label.cex = .5)
```



b)

They should all sum up to one:

```
rowSums(mP)
```

```
## 00 01 10 11
##  1  1  1  1
```

and they did…

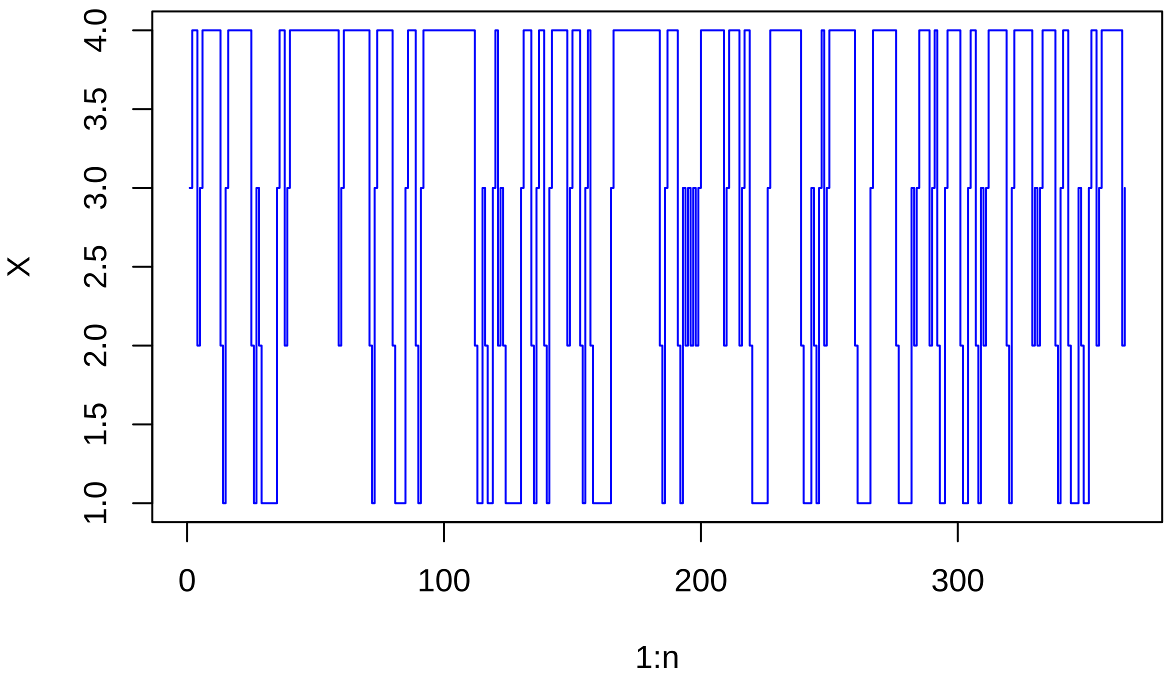
c)

```
set.seed(1234)
n=365
rainsim<-function(n){
    X <- rep(0, n+1)
    X[1] <- sample(S,1, prob=init )

    for(k in 2:(n+1)){
      X[k]<-sample(S,1, prob=mP[X[k-1],] )
    }
  X <- X[-1]
```

```
}

X <-rainsim(n)
```

d)

```
plot(1:n, X, type="s", col="blue")
```



e)

```
pi.hat <- as.numeric(table(X)/n)

cat(' Empirical frequence: ' , pi.hat)
```

```
##  Empirical frequence:  0.1863014 0.1232877 0.1260274 0.5643836
```

f)

```
delta = (sum((pi.hat- (pi.hat %*% mP))^2))^0.5
cat(' delta: ', delta)
```

```
##  delta:  0.01848067
```

g)

The result is a bit weak, probably due to a relativly low "n".

```
mP %^% 2
```

```
##         00     01     10     11
## 00 0.3844 0.0912 0.2356 0.2888
## 01 0.3410 0.1080 0.2090 0.3420
## 10 0.1320 0.1444 0.1080 0.6156
## 11 0.1045 0.1539 0.0855 0.6561
```

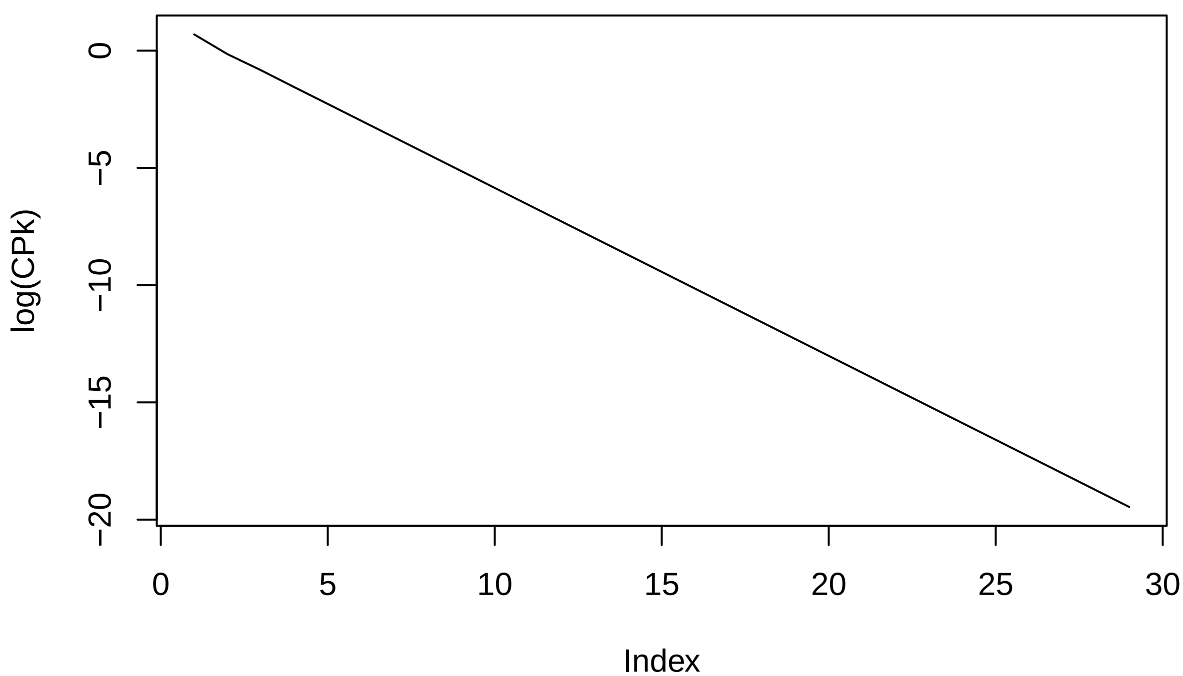
```
m <- 2
while(norm(mP %^% m - mP %^% (m - 1)) > (0.5*10^(-8)) ){
  m <- m+1
```

```r
}

mP_converged <- mP%^%m
mP_converged[1,]
```

```
##        00        01        10        11
## 0.1943463 0.1342756 0.1342756 0.5371025
```

i)

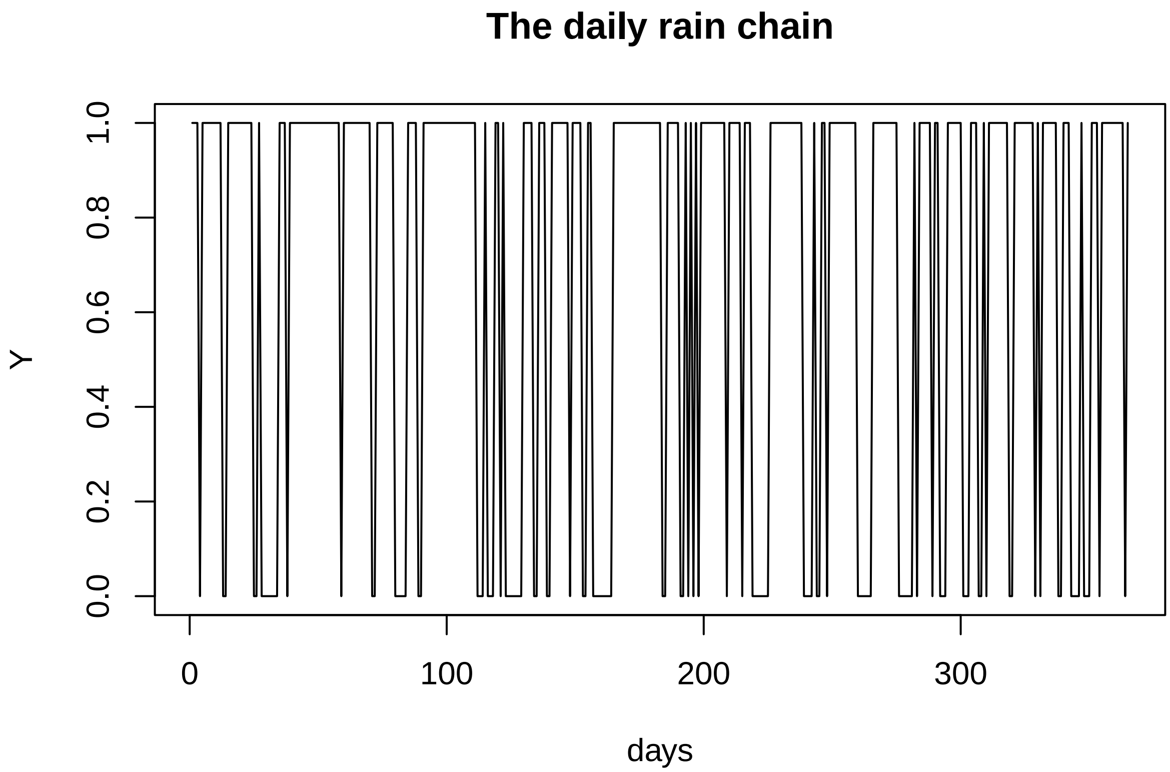
```r
iterations<-c()
CPk <- c()
result <- 10
k<-1
while(result > 0.5*1e-8){
  mP_power_k <- mP%^% k
  sum_l<-c()
  for(i in 1:3){
    for(j in (i+1):4){

      row <- abs(mP_power_k[i, ] - mP_power_k[j, ])
      sum_l <- c(sum_l, sum(row))
    }
  }
  result <- max(sum_l)
  CPk[k] <- result
  k <- k+1
}
plot(log(CPk), type='l')
```



j)

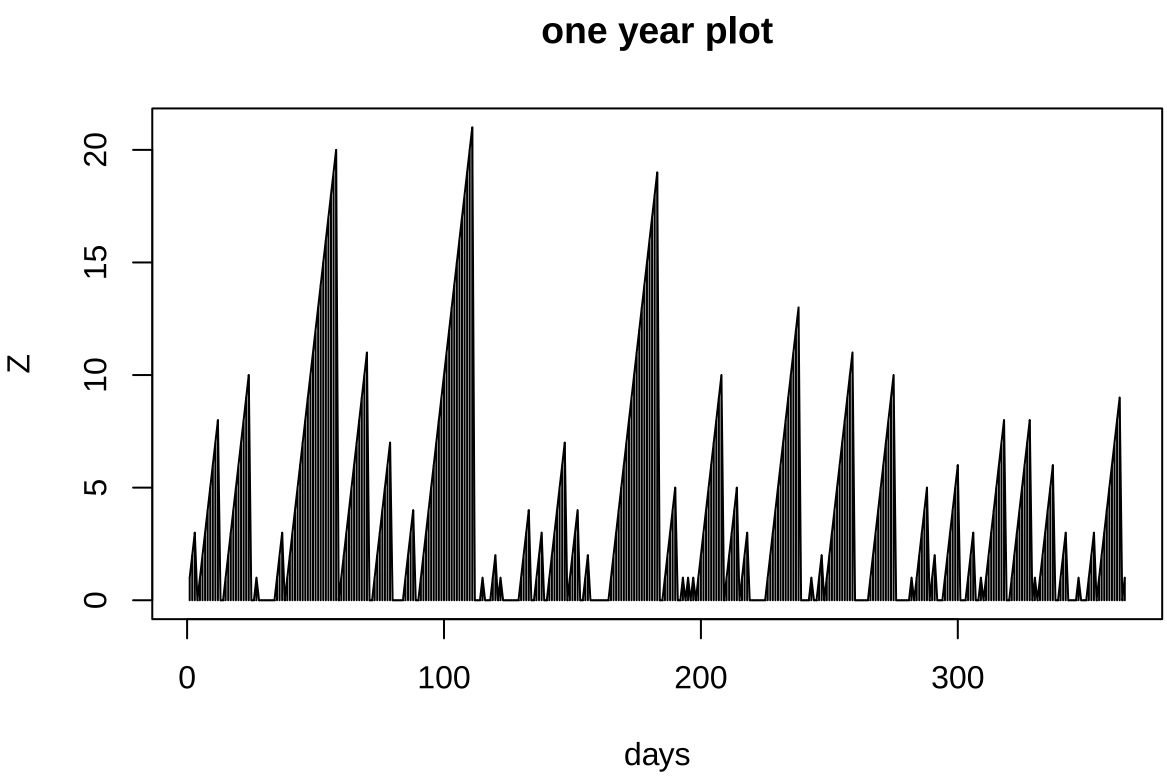
```r
Y<-pmin(pmax(0, X-2),1)
```

```r
plot(Y,main= "The daily rain chain",xlab="days", ylab="Y", type = "l")
```



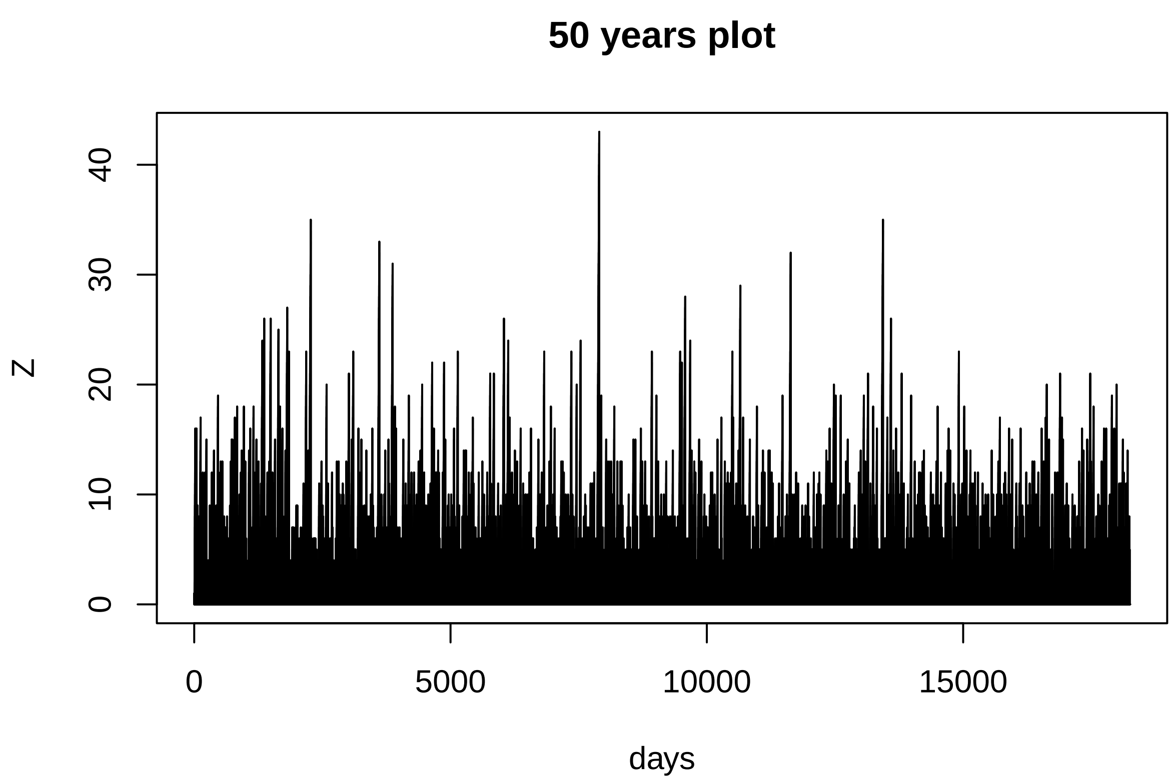
k)

```r
n=365

Z<-c(Y[1])

for(t in 2:n){
  Z_t <- (Z[t-1]+1)*Y[t]
  Z <- c(Z,Z_t)

}
plot(y=Z, x=1:n, type='h',xlab='days', main = "one year plot")
  lines(y=Z, x=1:n, type='l')
```



l)

```
n.50 <- 365*50

X.365.50 <- rainsim(n=n.50)
Y<-pmin(pmax(0, X.365.50-2),1)

Z<-c(Y[1])

for(t in 2:n.50){
  Z_t <- (Z[t-1]+1)*Y[t]
  Z <- c(Z,Z_t)

}
plot(y=Z, x=1:n.50, type='h', xlab="days", main='50 years plot')
  lines(y=Z, x=1:n.50, type='l')
```



m)

```
pi.eigen <- eigen(t(mP))$vectors[, 1] / sum(eigen(t(mP))$vectors[, 1])

pi.hat.50 <- as.numeric(table(X.365.50)/n.50)

cat(' Empirical frequence: ' , pi.hat.50)
```

```
##  Empirical frequence:  0.1896986 0.1349041 0.1348493 0.5405479
```

we have stationary distribution of P from one year and 50 years

```
pi.hat
```

```
## [1] 0.1863014 0.1232877 0.1260274 0.5643836
```

```
pi.hat.50
```

```
## [1] 0.1896986 0.1349041 0.1348493 0.5405479
```

Given we consider the eigen values as closer to true pi we se that pi obtained from the 50 year sample is significantly more accurate than from one year.

```
pi.hat-pi.eigen
```

```
## [1] -0.008044920 -0.010987947 -0.008248221  0.027281088
```

```
pi.hat.50-pi.eigen
```

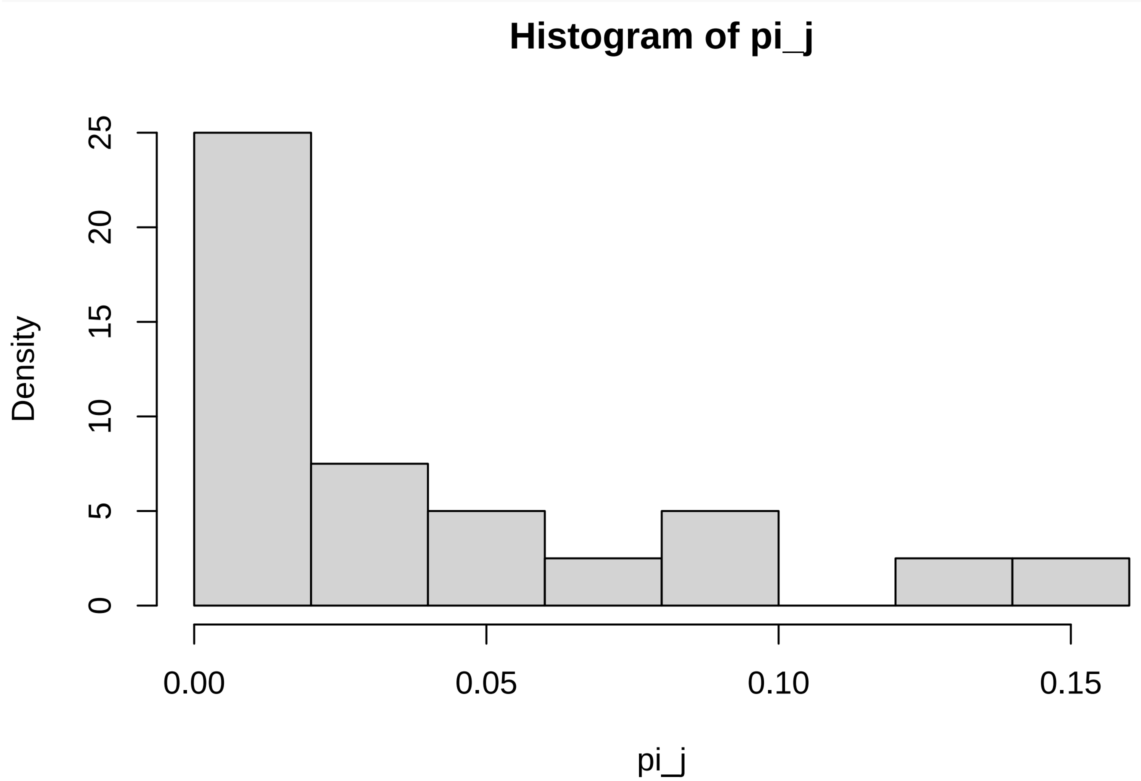
```
## [1] -0.0046476596  0.0006284912  0.0005736967  0.0034454717
```

Problem 3.3

a)

```
p <- 0.45
q <- 1-p

N <-20
n <- 1000

M<-N+1
mP.N<-matrix(0,M,M)
mP.N[1,2]<-1
mP.N[M,M-1]<-1
for(i in 2:(M-1)){
mP.N[i,i-1]<-q
mP.N[i,i+1]<-p
}
```

```
tau <- p/q


pi_0 <- 1-tau
pi_j <- vector(length = N)
pi_j[N] <- (tau)^(N-1)*pi_0
for(j in 1:N-1){
pi_j[j] <- ((tau)^(j)*pi_0)
}


hist(pi_j, probability = TRUE)
```



c)

```r
P_AA <-mP.N[1:20, 1:20]
I <- diag(N)
I_AA<-c(rep(1,N))
mu <- inv(I-P_AA)%*%I_AA
mu
```

```
##            [,1]
##  [1,] 2489.6052
##  [2,] 2488.6052
##  [3,] 2485.1608
##  [4,] 2478.7287
##  [5,] 2468.6450
##  [6,] 2454.0983
##  [7,] 2434.0967
##  [8,] 2407.4282
##  [9,] 2372.6111
## [10,] 2327.8346
## [11,] 2270.8856
## [12,] 2199.0590
## [13,] 2109.0487
## [14,] 1996.8139
## [15,] 1857.4159
## [16,] 1684.8182
## [17,] 1471.6433
## [18,] 1208.8740
## [19,]  885.4893
## [20,]  488.0191
```

Problem 3.4

a)

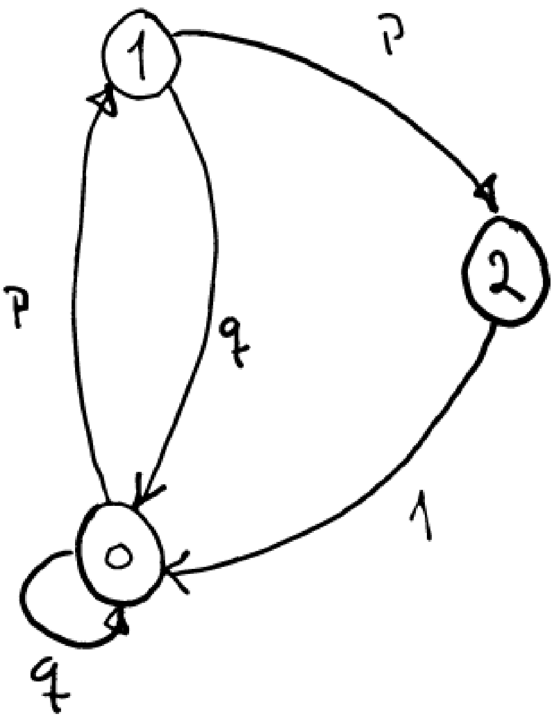


Figure 1: Matrix P illustrated

b)

$$
\begin{aligned}
f_{00} =& P_0(X_1|X_0) \cdot P_1(X_2|X_1) \cdot P_2(X_0|X_2) \\
=& p \cdot p \cdot 1 \\
=& \underline{p^2}
\end{aligned}
$$

since

$$
\begin{aligned}
0 \to 0 \\
0 \to 1 \to 0 \\
0 \to 1 \to 2 \to 0
\end{aligned}
$$

hence $P_0(S_0 < 3) = 1$ and return time is $0 \leq 3$

c)

$$det(P - \lambda I) = 0$$

$$[\pi_0, \pi_1, \pi_2] \cdot \begin{bmatrix} q & p & 0 \\ q & 0 & p \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \pi_0 \\ \pi_1 \\ \pi_2 \end{bmatrix}$$

$$q\pi_0 + \pi_1 = \pi_0 \tag{1}$$
$$q\pi_0 + p\pi_2 = \pi_1 \tag{2}$$
$$\pi_0 = \pi_2 \tag{3}$$
$$\pi_0 + \pi_1 + \pi_2 = 1 \tag{4}$$

by substituting 3 into 2:

$$q\pi_0 + p\pi_0 = \pi_1 \Rightarrow \pi_0(p + q) = \pi_1 \tag{5}$$

and the result of these combination:

$$\pi_0 + \pi_0(p + q) + \pi_0 = 1$$
$$\pi_0(1 + (p + q) + 1) = 1$$

$$\pi_0 = \underline{\frac{1}{p + q + 2}}$$

$$\pi_1 = \underline{\frac{(p + q)}{p + q + 2}}$$

$$\pi_2 = \pi_0 = \underline{\frac{1}{p + q + 2}}$$

d) We know the unique $\vec{\pi}$ and that $p \in (0,1)$ and $q = 1 - p$ hence $p + q = 1$. The chain is also positive recurrent

$$\begin{aligned}
E[S_0] &= \frac{1}{\pi_0} \\
&= \frac{1}{(p+q+2)^{-1}} \\
&= p + q + 2 = 1 + 2 = \underline{3}
\end{aligned}$$

e) No it is not bounded due to:

$$P(x_1 = 0 | X_0 = 0) > 0$$

there is always a chance that it can stop in State "0", however the expectation is equal for S_2 as S_0 since:

$$\begin{aligned}
&\pi_0 = \pi_2 \\
&\Rightarrow \frac{1}{\pi_2} = \frac{1}{\pi_2} \\
&= \frac{1}{(p+q+2)^{-1}} = \underline{3}
\end{aligned}$$