Universidad Carlos III de Madrid

MSc in Statistics for Data Science

First Assignment

Stochastic Processes

Authors:

David de la Fuente López

Diego Perán Vacas

Katherine Fazioli

Adrian White

24^{th} ***November 2020***

Exercise 1

a)

Markov chain.

Let the event $X_n = j$ as "the chain hits state j at time n" or "the chain visits state j at time n". This pollution scenario sequence meets the Markov property

$$P(X_n = j \,|X_0 = x_0; X_1 = x_1; ...; X_{n-1} = i) = P(X_n = j \,|X_{n-1} = i)$$

This means that the probability of going from one state to another in the chain does not depend on all the states it has previously passed through, but it just depends on the immediately previous one. This fact occurs in our example, where the probability of arriving the "non pollution" state, scenarios 1-4 or Alert state only depends on the last state it has visit.

Let us define the following two random variables, indicating both the status of the city and the level of pollution for the n-th unit time. The time unit for this Markov chain is the "day", since we consider a change in the pollution status of Madrid only for different days.

$$X_n = \text{status of Madrid} = (0, 1, 2, 3, 4, 5)$$

We label the state NR (no pollution state), with 0; then the different scenarios 1-4 with the corresponding numbers; and finally, we label the Alert status as 5.

$$Y_n = \text{level of pollution} \begin{cases} \alpha = \text{threshold for activating } 1, 2, 3, 4 \\ \beta = \text{threshold for activating } 5 \end{cases}$$

$$X_n = \begin{cases} X_{n-1} + 1 & \text{if } X_{n-1} = 0, 1, 2, 3 \text{ and } \alpha \leq Y_n < \beta \\ X_{n-1} & \text{if } X_{n-1} = 4 \text{ and } \alpha \leq Y_n < \beta \\ 5 & \text{if } X_{n-1} = 0, 1, 2, 3, 4 \text{ and } Y_n \geq \beta \\ X_{n-1} & \text{if } X_{n-1} = 5 \text{ and } Y_n \geq \alpha \\ 0 & \text{if } X_{n-1} = 0, 1, 2, 3, 4, 5 \text{ and } Y_n < \alpha \end{cases}$$

States and transitions.

AS we have mentioned above, we are defining the states as in the given data set used in following parts of the exercise. Thus, we represent the non-pollution state with "NR"; the scenarios 1-4 with "Sc1-4"; and the Alert status as "Alert".

The possible transitions are defined with the *figure 1.* For instance, if the city is in the state "NR", in the next day it is able to be in states "Sc1" (if α is reached), in "Alert" (if β is reached) or remain in "NR" if none of them are reached. Thus, the arrows in the *figure 1* represent the possible transitions.

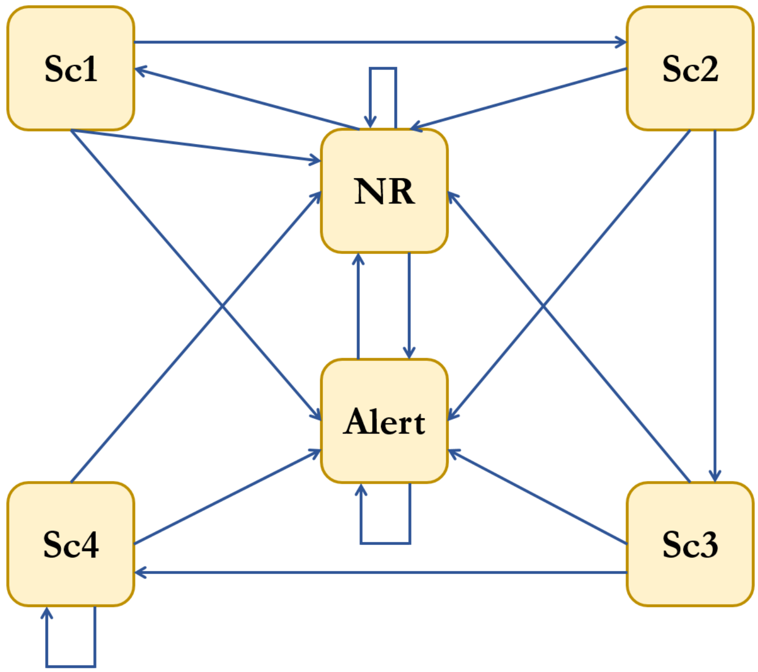


Figure 1: transition graph of the Markov chain representing the pollution scenario of Madrid.

Stationary distribution.

In order to give an answer to this question, we have to consider the **limit theorems** of the discrete-time Markov chains. First, we have to take into account that our stochastic process is a finite state Markov chain (in fact the number of states is 6). Second, we have to consider whether the Markov chain is **irreducible** or not. In the first case, the stationary distribution of the Markov chain is unique; in the second one, it is not. A Markov chain is irreducible if it has exactly one communication class (i.e. all states communicate with each other). We can observe in *figure 1* that this is our case, so we can affirm that our Markov chain is irreducible; thus it has a **unique stationary distribution**.

We had doubts about how to proof it is a Markov chain, whether it was necessary to proof it with our own words or mathematically. We asked the prof and she told us we can do it in both ways, so we have selected the previous mathematical way. THIS SHOULD BE REVIEWED TO CHECK EVERYTHING IS CORRECT IN THE OPINION OF THE FOUR MEMBERS

b)

In order to estimate the transition probabilities of the chain, we have to consider ML estimators of the proportion of transitions from one state to another (for all the possible transitions). The ML estimators of the real proportions are the sample proportions. We have to consider a Bernoulli variable, with a certain probability of success θ. Then, the estimated value of θ in a sample will be the proportion of successes in that sample. We are including here a brief mathematical proof of that.

X = "Bernoulli trial" [for instance having a transition NR-NR or any other allowed transition]

$$P(X=x;\theta)=\theta^x(1-\theta)^{1-x}$$
$$L(\theta;X_1,...,X_n)=\prod_{i=1}^{n}\theta^{x_i}(1-\theta)^{1-x_i}$$
$$l(\theta)=\log(L)=\sum_i x_i\log\theta+(n-\sum_i x_i)\log(1-\theta)$$
$$\frac{\partial l}{\partial\theta}=0 \quad\rightarrow\quad \frac{\bar{X}_n}{\theta}(\bar{X}_n-1)\frac{1}{1-\theta} \quad\rightarrow\quad (1-\theta)\bar{X}_n=-(\bar{X}_n-1)\theta$$
$$\boxed{\hat{\theta_{ML}}=\bar{X}_n}$$

It is easy to verify that $\hat{\theta}_{ML}$ maximizes the likelihood, and so is the MLE for the real proportions.

```r
# In the following line, one must include the directory in which this file is saved.
# setwd("C:/Users/Diego/OneDrive/Documents/Master Data Science/Asignaturas2º/Procesos estocásticos")
setwd("C:/Users/TEMP/Documents/MU Statistics for Data Science/Stochastic Processes/First assignment")
load("PollutionMadrid.RData")
X=X[6,] # we are group 6.

count = rep(0,17) # Initiate the vector to count the transitions.

for(i in 2:1460){
  count[1] = count[1] + (X[i-1] =="NR" & X[i] == "NR")*1
  count[2] = count[2] + (X[i-1] =="NR" & X[i] == "Sc1")*1
  count[3] = count[3] + (X[i-1] =="NR" & X[i] == "Alert")*1
  count[4] = count[4] + (X[i-1] =="Sc1" & X[i] == "NR")*1
  count[5] = count[5] + (X[i-1] =="Sc1" & X[i] == "Sc2")*1
  count[6] = count[6] + (X[i-1] =="Sc1" & X[i] == "Alert")*1
  count[7] = count[7] + (X[i-1] =="Sc2" & X[i] == "NR")*1
  count[8] = count[8] + (X[i-1] =="Sc2" & X[i] == "Sc3")*1
  count[9] = count[9] + (X[i-1] =="Sc2" & X[i] == "Alert")*1
  count[10] = count[10] + (X[i-1] =="Sc3" & X[i] == "NR")*1
  count[11] = count[11] + (X[i-1] =="Sc3" & X[i] == "Sc4")*1
  count[12] = count[12] + (X[i-1] =="Sc3" & X[i] == "Alert")*1
  count[13] = count[13] + (X[i-1] =="Sc4" & X[i] == "NR")*1
  count[14] = count[14] + (X[i-1] =="Sc4" & X[i] == "Sc4")*1
  count[15] = count[15] + (X[i-1] =="Sc4" & X[i] == "Alert")*1
  count[16] = count[16] + (X[i-1] =="Alert" & X[i] == "NR")*1
  count[17] = count[17] + (X[i-1] =="Alert" & X[i] == "Alert")*1
}
```

```r
sum(count)
```

```
## [1] 1459
```

As the sum of all the components of the “count” vector is equal to 1459, we know that we are including all the transitions on our sample. Now, we create the vector of proportions (simply dividing the counts by the total amount of transitions of our sample), and we verify that its sum is equal to 1.

```r
prop = count/sum(count) # Vector of proportions

print(round(prop,4))
```

```
##  [1] 0.8944 0.0336 0.0000 0.0137 0.0185 0.0014 0.0082 0.0089 0.0014 0.0048
## [11] 0.0041 0.0000 0.0041 0.0027 0.0000 0.0027 0.0014
```

```r
sum(prop)
```

```
## [1] 1
```

Finally, we use the previous calculated proportions to calculate the transition probabilities and form the transition matrix for our Markov chain.

```r
M=matrix(rep(0,36),6,6)
M[1,1]=prop[1]/(prop[1]+prop[2]+prop[3])
M[1,2]=prop[2]/(prop[1]+prop[2]+prop[3])
M[1,6]=prop[3]/(prop[1]+prop[2]+prop[3]) # This estimated transition is 0.
M[2,1]=prop[4]/(prop[4]+prop[5]+prop[6])
M[2,3]=prop[5]/(prop[4]+prop[5]+prop[6])
M[2,6]=prop[6]/(prop[4]+prop[5]+prop[6])
M[3,1]=prop[7]/(prop[7]+prop[8]+prop[9])
M[3,4]=prop[8]/(prop[7]+prop[8]+prop[9])
M[3,6]=prop[9]/(prop[7]+prop[8]+prop[9])
M[4,1]=prop[10]/(prop[10]+prop[11]+prop[12])
M[4,5]=prop[11]/(prop[10]+prop[11]+prop[12])
M[4,6]=prop[12]/(prop[10]+prop[11]+prop[12]) # This estimated transition is 0.
M[5,1]=prop[13]/(prop[13]+prop[14]+prop[15])
M[5,5]=prop[14]/(prop[13]+prop[14]+prop[15])
M[5,6]=prop[15]/(prop[13]+prop[14]+prop[15]) # This estimated transition is 0.
M[6,1]=prop[16]/(prop[16]+prop[17])
M[6,6]=prop[17]/(prop[16]+prop[17])

rownames(M) = c("NR","Sc1","Sc2","Sc3","Sc4","Alert")
colnames(M) = c("NR","Sc1","Sc2","Sc3","Sc4","Alert")
print(round(M,4))
```

```
##           NR    Sc1   Sc2    Sc3    Sc4  Alert
## NR    0.9638 0.0362 0.000 0.0000 0.0000 0.0000
## Sc1   0.4082 0.0000 0.551 0.0000 0.0000 0.0408
## Sc2   0.4444 0.0000 0.000 0.4815 0.0000 0.0741
## Sc3   0.5385 0.0000 0.000 0.0000 0.4615 0.0000
## Sc4   0.6000 0.0000 0.000 0.0000 0.4000 0.0000
## Alert 0.6667 0.0000 0.000 0.0000 0.0000 0.3333
```

```r
rowSums(M) # Verify the property of a transition matrix.
```

```
##    NR   Sc1   Sc2   Sc3   Sc4 Alert
##     1     1     1     1     1     1
```

Our estimated transition matrix does not show any of the theoretically impossible transitions that we have not mentioned or included in the *figure 1*. These transitions are not possible due to the problem conditions (for example, going from Sc1 to Sc3, since it is necessary to pass through Sc2 in an intermediate step).

However, the estimated probabilities obtained with our sample for the transitions NR to Alert, Sc3 to Alert, Sc4 to Alert are equal to 0. These transitions are theoretically possible, although they are not present in the sample.

c)

```r
X[1:6]
```

```
## [1] "NR" "NR" "NR" "NR" "NR" "NR"
```

We can observe that the first 6 observations are all in the state "NR" (we consider the initial state $X_0 =$ "NR" on Monday, so our first six observations correspond to Tuesday to Sunday). Let's calculate the joint probability of the observed sequence of scenarios for the first week of data. The initial distribution (it is given that the first step is at "NR") is given by the vector:

$$\alpha = (1, 0, 0, 0, 0, 0)$$

The joint distribution is computed from the estimated transition matrix as follows:

$$P(X_1 = NR, X_2 = NR, ..., X_7 = NR) = (\alpha M)_1 \cdot M_{1,1} \cdot M_{1,1} \cdot M_{1,1} \cdot M_{1,1} \cdot M_{1,1} = M_{1,1}^6 = 0.8016$$

Note that $M_{1,1}$ represents the transition between the state 0 and the state 0 (or equivalently from state NR to state NR).

```r
M[1,1]^6 # Compute the joint probability from the estimated transition matrix.
```

```
## [1] 0.8015878
```

This can be interpreted in two ways: the 0-th state is Monday, or the 0-th state is the previous Sunday. THIS SHOULD BE REVIEWED TO CHECK EVERYTHING IS CORRECT IN THE OPINION OF THE FOUR MEMBERS

d)

In order to compute the stationary distribution of the chain, we can implement a function to the estimated transition matrix. This function solves the equation:

```r
stat_dist = function(P){
  # ARGUMENT:
  # P is the transition matrix of a finite state Markov chain.

  # The equation that we are solving is $\hat{(P-I)}*pi = (1,0,...,0)

  # We create that matrix:
```

```r
  m = nrow(P) #dimension of the matrix.
  P_tild = matrix(0,nrow=m,ncol=m)
  P_tild = P - diag(m)
  P_tild[,1] = rep(1,m)

  b = c(1,rep(0,m-1))
  solve(t(P_tild),b) # Here, we are using row vectors, not column vectors, because of this
  # we use the transpose of the matrix.
}
```

Stationary distribution **from the estimated transition matrix**:

```r
round(stat_dist(M),4)
```

```
##     NR    Sc1    Sc2    Sc3    Sc4  Alert
## 0.9280 0.0336 0.0185 0.0089 0.0069 0.0041
```

Estimation of the stationary distribution **from the data**. In order to make this estimation, we should consider the definition of the π_i element of the stationary distribution. It represents the proportion of time that the Markov chain will spend in the state i (in the long-term). Following this definition and assuming 1460 time units are enough to consider the long-term situation, we can estimate from the data by counting the proportion of times in each state.

```r
est = c(sum(X == "NR"),sum(X == "Sc1"),sum(X == "Sc2"),
        sum(X == "Sc3"),sum(X == "Sc4"),sum(X == "Alert"))
est = est/length(X)
names(est) = c("NR","Sc1","Sc2","Sc3","Sc4","Alert")
round(est,4)
```

```
##     NR    Sc1    Sc2    Sc3    Sc4  Alert
## 0.9281 0.0336 0.0185 0.0089 0.0068 0.0041
```

Both methods are two different procedures to obtain the same quantity. The first one would be more convenient if we have the transition matrix (the real transition matrix) of our Markov chain, since it computes the stationary distribution directly from it by solving the equation:

$$\pi \cdot (\widetilde{P - \mathbb{I}}) = (1, 0, ..., 0)$$

Nevertheless, if we do not have the transition matrix, but instead we have a very long sample chain, we can estimate the stationary distribution using the second procedure. This method is much more simple, and requires less calculations. Nevertheless, in our case, we have commented previously that our particular sample does not even include all theoretically possible transitions between states. This fact is an indicator that our sample chain may not be long enough to estimate this stationary distribution, and so we consider that the first method is more convenient for this specific case.

However, as we can observe above, the results are very similar with both methods. In fact, they only differ in the fourth decimal for the elements of the stationary distribution corresponding to the states "NR" and "Alert".

Discussion of the differences between the two methods. THIS SHOULD BE REVIEWED TO CHECK EVERYTHING IS CORRECT IN THE OPINION OF THE FOUR MEMBERS

e)

From what we have seen in the lessons, an irreducible an aperiodic Markov Chain having a stationary distribution π, have a limiting distribution equal to π. We only need to show that it is an aperiodic Markov chain, that is, the greatest common divisor of the set of possible return times to i, is equal to 1.

$$d(i) = gcd\{n > 0 \,|\, (P^n)_{ii} > 0\} = 1$$

Let us see that the period of each state is equal to 1 and thus we will show that it is an aperiodic Markov chain.

Possibles transitions $(P^n)_{NR,NR}$

- $NR - Sc1 - NR \rightarrow n = 2$
- $NR - Sc1 - Sc2 - NR \rightarrow n = 3$
- $gcd(2, 3) = 1 \Rightarrow$ aperiodic

Possibles transitions $(P^n)_{Sc1,Sc1}$

- $Sc1 - NR - Sc1 \rightarrow n = 2$
- $Sc1 - Sc2 - NR - Sc1 \rightarrow n = 3$
- $gcd(2, 3) = 1 \Rightarrow$ aperiodic

Possibles transitions $(P^n)_{Sc2,Sc2}$

- $Sc2 - NR - Sc1 - Sc2 \rightarrow n = 3$
- $Sc2 - Sc3 - NR - Sc1 - Sc2 \rightarrow n = 4$
- $gcd(3, 4) = 1 \Rightarrow$ aperiodic

Possibles transitions $(P^n)_{Sc3,Sc3}$

- $Sc3 - NR - Sc1 - Sc2 - Sc3 \rightarrow n = 4$
- $Sc3 - Sc4 - NR - Sc1 - Sc2 - Sc3 \rightarrow n = 5$
- $gcd(4, 5) = 1 \Rightarrow$ aperiodic

Possibles transitions $(P^n)_{Sc4,Sc4}$

- $Sc4 - NR - Sc1 - Sc2 - Sc3 - Sc4 \rightarrow n = 5$
- $Sc4 - Sc4 - NR - Sc1 - Sc2 - Sc3 - Sc4 \rightarrow n = 6$
- $gcd(5, 6) = 1 \Rightarrow$ aperiodic

Possibles transitions $(P^n)_{Alert,Alert}$

- $Alert - NR - Alert \rightarrow n = 2$

- $Alert - NR - Sc1 - Alert \to n = 3$
- $gcd(2, 3) = 1 \Rightarrow$ aperiodic

So the Markov chain is aperiodic and irreducible, and thus, it has a unique stationary distribution which is also the limiting distribution.

The limiting distribution means that, when n tends to ∞, the probability of being in a pollution episode is constant and equal to the sum of probabilities of being in all of the states implying a pollution episode. I.e. being in states "Sc1", "Sc2", "Sc3", "Sc4" or "Alert". In our case, we have proved that the limiting distribution is equal to the stationary one. This means that, with the sample we got, the estimated proportion of time that we expect to find a pollution episode in the city is: $\hat{p} = 0.0336+0.0185+0.0089+0.0069+0.0041 = 0.072$

Therefore, for example, during a whole year, we expect to find about 26 days in which the pollution episode protocol is activated.

Discussion of the meaning in terms of pollution episodes. THIS SHOULD BE REVIEWED TO CHECK EVERYTHING IS CORRECT IN THE OPINION OF THE FOUR MEMBERS

f)

Let $\pi = (0.9280, 0.0336, 0.0185, 0.0089, 0.0069, 0.0041)$ be the stationary distribution of our Markov chain. Then, the expected number of days over a whole year in which you won't be allowed to use the car is calculated as follows:

$$(0.0089 + 0.0069 + 0.0041) \cdot 365 = 7.2270$$

Thus, we expected to have about 7 days a year without being able to drive.

```
round(365*(round(sum(X == "Sc3")/length(X),4)+round(sum(X == "Sc4")/length(X),4)+
             round(sum(X == "Alert")/length(X),4)),4)
```

```
## [1] 7.227
```

Exercise 2

a)

We are given the following Markov Chain defined on the set of Natural Numbers and Zero, where p takes values in (0, 1).

$$P_{i,0} = p, \ P_{i,i+1} = 1 - p, \ p_{i,j} = 0 \ if \ j \neq 0, i + 1, \ i \geq 0$$

The purpose of this exercise is to analyze this process, determine the classification of its states, and eventually understand its limiting behavior. We will first determine its stationary distribution analytically.

In order to determine the stationary distribution of our Markov Chain, we must find an expression for its transition matrix, P. Firstly, we note that our process has an infinite state space, S = N U {0}, where N is the set of natural numbers.

We may immediately set the elements of the first column of P to our probability p, since the probability of moving to zero at any state i is p. We then consider the accessibility of the remaining states, where for any state i > 0, state i may only be accessed by state (i-1) with probability (1-p). Thus, we may represent the second block of P as a diagonal matrix with elements (1-p), where state i may move to state i+1 with probability (1-p).

$$P = \begin{bmatrix} p & (1-p) & 0 & 0 & 0 & \dots \\ p & 0 & (1-p) & 0 & 0 & \dots \\ p & 0 & 0 & (1-p) & 0 & \dots \\ p & 0 & 0 & 0 & (1-p) & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

We now have our transition matrix P for the Markov Chain in question. We may use this to determine our stationary distribution Π. The stationary distribution of a Markov Chain has the following property:

$$\Pi P = \Pi$$

In other words, once the process enters its stationary distribution, it will remain at this state of equilibrium for the following transitions. To determine the stationary distribution, we use the following approach.

We have,

$$\Pi P = \Pi$$

Where,

$$\Pi = [\Pi_0, \Pi_1, \Pi_2, \Pi_3 ...]^T$$

We solve the system of equations

$$p(\Pi_0 + \Pi_1 + \Pi_2 + \Pi_3 + ...) = \Pi_0 (1-p)\Pi_0 = \Pi_1 (1-p)\Pi_1 = \Pi_2 (1-p)\Pi_2 = \Pi_3 ...$$

Which gives us,

$$\Pi_i = (1-p)^i \Pi_0$$

We must verify that this is a valid distribution by taking the sum across all possible states in S.

$$\sum_{n \geq 0} \Pi_i = \sum_{n \geq 0} (1-p)^i \Pi_0 = 1$$

We have an infinite geometric series where (1-p) < 1.

$$\frac{1}{1-(1-p)} \Pi_0 = 1$$

So,

$$\Pi_0 = p$$

Thus, we have the stationary distribution

$$\Pi_i = (1-p)^i p$$

We will now use our knowledge of the Markov Chain to determine the expected return time to zero. We may understand this as a random variable, k, where k equals the number of steps before the process returns to zero with an initial position at zero. We can model k using a geometric distribution, k ~ Geom(p), where the process returns to zero after k = r transitions, with (r-1) transitions from i to (i+1) with probability (1-p), then one transition from i to zero with probability p.

We have k ~ Geom(p), with probability density function

$$P(k = r) = p(1-p)^{r-1}$$

We may now express the expected return time to zero as the expectation of the random variable k.

$$E[k] = \sum_{k\geq 1} kP(k) = \sum_{k\geq 1} kp(1-p)^{k-1} = p\sum_{k\geq 1} k(1-p)^{k-1} = p\frac{d}{dp}(-\sum_{k\geq 1}(1-p)^k) = p\frac{d}{dp}(-\frac{1}{p}) = p(\frac{1}{p^2})$$

This gives us

$$E[k] = \frac{1}{p}$$

We may also note that the distribution of k, our return time to zero from an initial state zero, and our stationary distribution are the same.

b)

We will now analyze the communication classes of the Markov Chain. First, may discuss the accessibility of the states. As noted earlier, zero is accessible from all states, including itself. For states i greater than zero, state i is only accessible from state (i-1). However, by induction, this tells us that any state i greater or equal to zero is accessible from zero.

$$\sum_{n\geq 1} P^n_{i,0} > 0$$

Since the Markov Chain has a non zero probability of reaching any state i $>$ 0 in some n number of steps.

As such, we have any state i $>$ 0 accessible from zero and state zero accessible from all states. Thus, all states of this Markov Chain communicate, so we have one communication class. The Markov Chain is irreducible.

Moreover, we know that state zero has Period 1, since the process may move to state zero in a single step with probability p from any state i in S. Since there are no obstructions for this state and we have an irreducible chain, the process is aperiodic.

We now consider whether the process is recurrent or transient. Since our chain is irreducible, we may assess the possibility of recurrence for a single state in order to classify the full chain. We will focus on state zero.

We have recurrence at state zero if the probability of returning to zero eventually, from any point in the process, is one. Previously, we modeled the return time to zero as a geometric distribution, k ~ Geom(p), where the probability density function, P(k = n), is the probability of returning to zero from zero in n steps.

We have

$$P(k = n) = f^n_{0,0} = p(1-p)^{n-1}$$

Where, by the property of the probability measure,

$$\sum_{n\geq 1} f^n_{0,0} = \sum_{n\geq 1} p(1-p)^{n-1} = 1$$

So, we have a 100% probability of returning to zero from state zero. Thus, state zero is recurrent, as is the Markov Chain. We now consider the type of recurrence, whether positive or null, by considering the expected value of the number of steps to return to zero.

We determined previously that our expected return time to zero is expressed as the expectation of the geometric random variable k.

$$E[k] = \frac{1}{p} p \in (0, 1)$$

This expectation is always a finite value for our range of possible values of p. The expected return time does diverge to infinity as p goes to zero, but for any p such that 0 < p < 1, we we have a finite expected return time. Thus, we have positive recurrence.

Lastly, since the Markov Chain is irreducible and aperiodic, we know that the chain has a limiting distribution. Moreover, aperiodicity ensures that the limiting distribution is equal to the stationary distribution.

c)

The following function "sim" can simulate sequences of a Markov chain, with arguments probability p, initial value of the chain a, and number of steps to generate n. We create a vector x to write the values of our chain. The sequence begins with the specified initial value. Then, the next step at any given state is to either return to 0 with probability p or go to $i + 1$ with probability $1 - p$.

```
sim<- function(a,p,n){
  x=numeric(n+1)
  x[1]=a
  for (i in 1:n){
    x[i+1]=sample(c(0,x[i]+1),1,prob=c(p,1-p))
  }
  return(x)
}
```

Now, we can simulate sequences of our chain with our function. First, we generate one trajectory. We sample an initial value on [100,1000]. The probability is defined as $\frac{1}{k+1} = \frac{1}{7}$, where k=6. Lastly, we generate 1000 steps of this sequence.

In this simulation, the initial value for the chain was 524. Overall, 49 states were visited during this simulation, including state 524 and states 0 to 47. We can observe the frequencies of visiting each state during the simulation. We see that 0 was the state visited with the highest frequency followed by states closer to 0.

```
set.seed(333)
a=sample(100:1000,1)
p=1/7
n=1000
x=sim(a,p,n)
t=as.data.frame(table(x))
colnames(t)<-c("State", "Frequency")
knitr::kable(list(t[1:17,],t[18:33,],t[34:49,]), row.names=F, align='l')
```

The proportion of visits to individual states in this one trajectory begin to reflect the stationary distribution found in part a. For example, the proportion of visits to states 0, 1, 2, and 3 observed in this one trajectory were 0.1538, 0.1289, 0.1119, and 0.0909. These values correspond to the probabilities defining the stationary distribution $\pi_i = p(1-p)^i$, which are 0.1429, 0.1225, 0.1050, and 0.0900 for states 0, 1, 2, and 3, respectively.

```
options(digits=4)
prob=as.numeric(t$Frequency)/(n+1)
prob_t=as.data.frame(prob)
prob_t$x=c(0:47,524)
```

State	Frequency
0	154
1	129
2	112
3	91
4	80
5	70
6	52
7	46
8	41
9	33
10	30
11	25
12	21
13	18
14	17
15	14
16	13

State	Frequency
17	10
18	10
19	5
20	2
21	1
22	1
23	1
24	1
25	1
26	1
27	1
28	1
29	1
30	1
31	1
32	1

State	Frequency
33	1
34	1
35	1
36	1
37	1
38	1
39	1
40	1
41	1
42	1
43	1
44	1
45	1
46	1
47	1
524	1

```
colnames(prob_t)<-c('Probability', 'State')
prob_t=prob_t[c('State', 'Probability')]
knitr::kable(list(prob_t[1:17,],prob_t[18:33,],prob_t[34:49,]),row.names=F, align='l')
```

We can also see that the average return time to 0 in simulating this one trajectory reflects the expected return time to 0 found in part a, $\frac{1}{p} = 7$, where $p = \frac{1}{7}$. The average return time to 0 can be estimated by the average distance between 0's in the sequence. The average return time to 0 was 6.4575 in this simulation.

```
options(digits=5)
rt=mean(diff(which(x==0)))
```

d)

Now, we generate 8 trajectories with different initial values on [100,1000]. We create a matrix to store each of the eight sequences. We again see the states visited with the highest frequency across the eight trajectories is 0 followed by states closest to 0. However, we note that these are the frequencies across the 8 trajectories, and some states were only visited in one trajectory given the different initial values.

```
set.seed(333)
k=8
X=matrix(nrow=k,ncol=n+1)
for (i in 1:k){
  a=sample(100:1000,1)
  x=sim(a,p,n)
  X[i,]=x
}
t2=as.data.frame(table(X))
colnames(t2)<-c("State", "Frequency")
knitr::kable(list(t2[1:33,],t2[34:66,],t2[66:98,],
                  t2[99:128,]),row.names=F, align='l')
```

The proportion of visits to individual states across the eight trajectories closely reflects the stationary distribution found in part a; however, note again that some states were only visited in one trajectory due

State	Probability	State	Probability	State	Probability
0	0.1538	17	0.010	33	0.001
1	0.1289	18	0.010	34	0.001
2	0.1119	19	0.005	35	0.001
3	0.0909	20	0.002	36	0.001
4	0.0799	21	0.001	37	0.001
5	0.0699	22	0.001	38	0.001
6	0.0519	23	0.001	39	0.001
7	0.0460	24	0.001	40	0.001
8	0.0410	25	0.001	41	0.001
9	0.0330	26	0.001	42	0.001
10	0.0300	27	0.001	43	0.001
11	0.0250	28	0.001	44	0.001
12	0.0210	29	0.001	45	0.001
13	0.0180	30	0.001	46	0.001
14	0.0170	31	0.001	47	0.001
15	0.0140	32	0.001	524	0.001
16	0.0130				

to different initial values. Across the trajectories, the proportion of visits to states 0, 1, 2, and 3 observed were 0.1465, 0.1256, 0.1061, and 0.0918. These values correspond to the probabilities defining the stationary distribution $\pi_i = p(1-p)^i$, which are 0.1429, 0.1225, 0.1050, and 0.0900 for states 0, 1, 2, and 3, respectively.

```
options(digits=4)
prob2=as.numeric(t2$Freq)/((n+1)*k)
prob_t2=as.data.frame(prob2)
val=unique(as.vector(X))
prob_t2$x=c(sort(val))
colnames(prob_t2)<-c('Probability', 'State')
prob_t2=prob_t2[c('State', 'Probability')]
knitr::kable(list(prob_t2[1:33,],prob_t2[34:66,],prob_t2[66:98,],
                  prob_t2[99:128,]),row.names=F, align='l')
```

We can also see that the average return time to 0 across the 8 trajectories. The average return time to 0 was 6.8148 in this simulation, which corresponds to the expected return time to 0 determined in part a $\frac{1}{p} = 7$, where $p = \frac{1}{7}$.

```
options(digits=5)
rt2=mean(diff(which(X==0)))
```

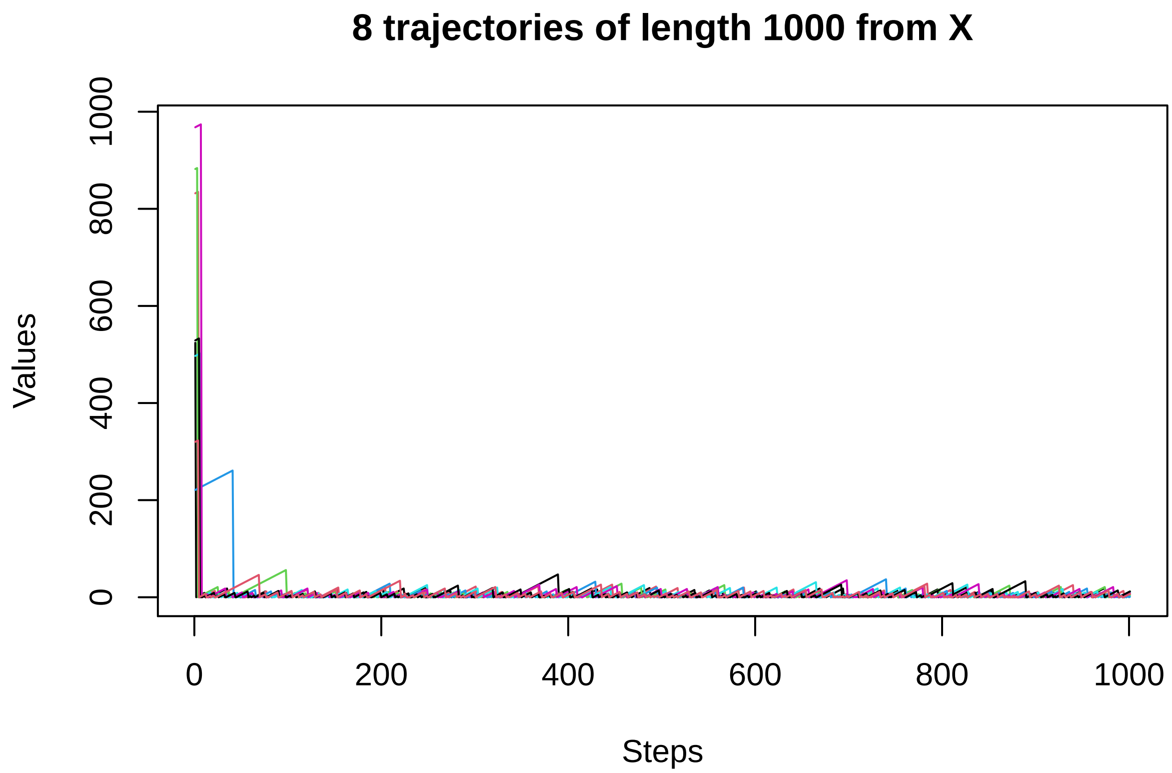
We can also represent the trajectories as a function of time in a graph. In this graph, the values of the chains at each step, or states (y-axis) are plotted versus the successive steps of the chain (x-axis).

The graph reflects the expected return time to 0. The increases and decreases in the graph correspond to the chain moving away from 0 (increase) and back to 0 (decrease). The average distance covered by the peaks appears to be approximately 7. In addition, the graph also reflects the stationary distribution in that it shows the most frequently visited state is 0 followed by states closest to 0.

```
matplot(1:(n+1),t(X),type="l",lty=1,
        main=paste(k,"trajectories of length",n,"from X"), xlab="Steps",
        ylab="Values")
```

State	Frequency	State	Frequency	State	Frequency	State	Frequency
0	1173	33	7	229	1	320	1
1	1006	34	6	230	1	321	1
2	850	35	5	231	1	322	1
3	735	36	4	232	1	323	1
4	613	37	4	233	1	497	1
5	531	38	3	234	1	498	1
6	442	39	3	235	1	499	1
7	384	40	3	236	1	500	1
8	339	41	3	237	1	501	1
9	285	42	3	238	1	502	1
10	237	43	3	239	1	524	1
11	190	44	3	240	1	529	1
12	155	45	3	241	1	530	1
13	130	46	3	242	1	531	1
14	117	47	2	243	1	532	1
15	104	48	1	244	1	533	1
16	93	49	1	245	1	832	1
17	79	50	1	246	1	833	1
18	70	51	1	247	1	834	1
19	56	52	1	248	1	835	1
20	50	53	1	249	1	882	1
21	40	54	1	250	1	883	1
22	32	55	1	251	1	884	1
23	30	56	1	252	1	968	1
24	27	221	1	253	1	969	1
25	23	222	1	254	1	970	1
26	19	223	1	255	1	971	1
27	14	224	1	256	1	972	1
28	13	225	1	257	1	973	1
29	10	226	1	258	1	974	1
30	9	227	1	259	1		
31	9	228	1	260	1		
32	8	229	1	261	1		

State	Probability	State	Probability	State	Probability	State	Probability
0	0.1465	33	9e-04	229	1e-04	320	1e-04
1	0.1256	34	7e-04	230	1e-04	321	1e-04
2	0.1061	35	6e-04	231	1e-04	322	1e-04
3	0.0918	36	5e-04	232	1e-04	323	1e-04
4	0.0765	37	5e-04	233	1e-04	497	1e-04
5	0.0663	38	4e-04	234	1e-04	498	1e-04
6	0.0552	39	4e-04	235	1e-04	499	1e-04
7	0.0480	40	4e-04	236	1e-04	500	1e-04
8	0.0423	41	4e-04	237	1e-04	501	1e-04
9	0.0356	42	4e-04	238	1e-04	502	1e-04
10	0.0296	43	4e-04	239	1e-04	524	1e-04
11	0.0237	44	4e-04	240	1e-04	529	1e-04
12	0.0194	45	4e-04	241	1e-04	530	1e-04
13	0.0162	46	4e-04	242	1e-04	531	1e-04
14	0.0146	47	2e-04	243	1e-04	532	1e-04
15	0.0130	48	1e-04	244	1e-04	533	1e-04
16	0.0116	49	1e-04	245	1e-04	832	1e-04
17	0.0099	50	1e-04	246	1e-04	833	1e-04
18	0.0087	51	1e-04	247	1e-04	834	1e-04
19	0.0070	52	1e-04	248	1e-04	835	1e-04
20	0.0062	53	1e-04	249	1e-04	882	1e-04
21	0.0050	54	1e-04	250	1e-04	883	1e-04
22	0.0040	55	1e-04	251	1e-04	884	1e-04
23	0.0037	56	1e-04	252	1e-04	968	1e-04
24	0.0034	221	1e-04	253	1e-04	969	1e-04
25	0.0029	222	1e-04	254	1e-04	970	1e-04
26	0.0024	223	1e-04	255	1e-04	971	1e-04
27	0.0017	224	1e-04	256	1e-04	972	1e-04
28	0.0016	225	1e-04	257	1e-04	973	1e-04
29	0.0012	226	1e-04	258	1e-04	974	1e-04
30	0.0011	227	1e-04	259	1e-04		
31	0.0011	228	1e-04	260	1e-04		
32	0.0010	229	1e-04	261	1e-04		


8 trajectories of length 1000 from X
Values
0
200
400
600
800
1000
0
200
400
600
800
1000
Steps