HW2-Exercise 3

Zining Fan(zf2234), Mutian Wang(mw3386), Siyuan Wang(sw3418)

3.1

a_i is the transition probability.
$a_1 = p(x_1 = 0|x_0 = 0)$, two consecutive rainy days;
$a_2 = p(x_1 = 1|x_0 = 0)$, a rainy day followed by a day with no rain;
$a_3 = p(x_1 = 0|x_0 = 1)$, a day with no rain followed by a rainy day;
$a_4 = p(x_1 = 1|x_0 = 1)$, two consecutive days with no rain.

3.2

From $\pi^T P = \pi^T$, $\pi^T 1_s = 1_s$, we could get that
$\pi = [\frac{a_3}{a_2+a_3}, \frac{a_2}{a_2+a_3}]$.
So, the long term prob of observing a rainy day is $\frac{a_3}{a_2+a_3}$.

3.3

```
n<-nrow(CentralPark_July)-1
p00=0
p01=0
p10=0
p11=0
for (i in 2:n){
  if (CentralPark_July$X[i-1]==0 & CentralPark_July$X[i]==0) {
    p00=p00+1
  }
  if (CentralPark_July$X[i-1]==0 & CentralPark_July$X[i]==1) {
    p01=p01+1
  }
  if (CentralPark_July$X[i-1]==1 & CentralPark_July$X[i]==0) {
    p10=p10+1
  }
  if (CentralPark_July$X[i-1]==1 & CentralPark_July$X[i]==1) {
    p11=p11+1
  }
}
a1=p00/(p00+p01)
a2=p01/(p00+p01)
a3=p10/(p10+p11)
a4=p11/(p10+p11)
a1
```

```
## [1] 0.3089343
```

```
a2
```

```
## [1] 0.6910657
```

```
a3
```

```
## [1] 0.23314
```

```
a4
```

```
## [1] 0.76686
```

From the result, $\hat{a_1} = 0.31$, $\hat{a_2} = 0.69$, $\hat{a_3} = 0.23$, $\hat{a_4} = 0.76$.

3.4

1) $X_{t+1} = 1$

$$p(X_{t+1} = 1|X_t = 1) = a_4,\ p(1 - X_{t+1} = 0|X_t = 0) = a_1.$$

   From the estimated values above, $\hat{a_4} = 0.76$, $\hat{a_1} = 0.30$, we could see that these two probabilities are signiciantly different.

2) $X_{t+1} = 0$

$$p(X_{t+1} = 0|X_t = 1) = a_3,\ p(1 - X_{t+1} = 1|X_t = 0) = a_2.$$

   From the estimated values above, $\hat{a_3} = 0.23$, $\hat{a_2} = 0.69$, we could see that these two probabilities are still signiciantly different.

So we could make a summary that: the probabilities of $X_{t+1}|X_t = 1$ and $1 - X_{t+1}|X_t = 0$ are quite different.

3.5

To compare the first order model and the second order model, we build a hypothsis testing. H_0 :The first markov model fits the data better. H_1 :The second markov model fits the data better. In this question, we still focus on July.

1) we build a second order markov model for July.

```
CentralPark$X<- with(CentralPark, ifelse(CentralPark$PRCP>=1.5, 0, 1))
n<-nrow(CentralPark)-2
p000=0
p001=0
p010=0
p011=0
p100=0
p101=0
p110=0
p111=0
for (i in 3:n){
  if (CentralPark$X[i-2]==0 & CentralPark$X[i-1]==0 & CentralPark$X[i]==0) {
    p000=p000+1
  }
  if (CentralPark$X[i-2]==0 & CentralPark$X[i-1]==0 & CentralPark$X[i]==1) {
    p001=p001+1
  }
  if (CentralPark$X[i-2]==0 & CentralPark$X[i-1]==1 & CentralPark$X[i]==0) {
    p010=p010+1
  }
  if (CentralPark$X[i-2]==0 & CentralPark$X[i-1]==1 & CentralPark$X[i]==1) {
    p011=p011+1
  }
  if (CentralPark$X[i-2]==1 & CentralPark$X[i-1]==0 & CentralPark$X[i]==0) {
    p100=p100+1
  }
  if (CentralPark$X[i-2]==1 & CentralPark$X[i-1]==0 & CentralPark$X[i]==1) {
```

```
    p101=p101+1
  }
  if (CentralPark$X[i-2]==1 & CentralPark$X[i-1]==1 & CentralPark$X[i]==0) {
    p110=p110+1
  }
  if (CentralPark$X[i-2]==1 & CentralPark$X[i-1]==1 & CentralPark$X[i]==1) {
    p111=p111+1
  }
}
```

The transition matrix is

```
m=matrix(c(p000,p001,p010,p011,p100,p101,p110,p111),nrow = 4,dimnames =
           list(c('00','01','10','11'),c('0','1')))
m
```

```
##       0     1
## 00  934  2640
## 01 2640  4329
## 10 1253  5716
## 11 5716 20147
```

Under H_0, the expected matrix is

```
##          0        1
## 00 1107.94  2160.39
## 01  822.02  1602.87
## 10 4808.61 17846.85
## 11 5366.13 19916.05
```

So, the likelihood hood ratio test statistic is 380. The p value is very samll. Thus, we reject H_0, which means the second order morkov chain fits the data better.