

HW1

Kejin Qian

10/28/2018

Problem 1

In Markov100.txt, the one step transition probability matrix for a Markov chain with 100 states (State 1 to State 100) is given. Note that the data has no heading. Name of the data set Markov100 Number of rows 100 Number of columns 100

```
markov <- read.table("markov100.txt", header=FALSE)
P <- as.matrix(markov);
```

problem 1(a)

Suppose we are at State 1 now. Find and display the probability of being in State 5 after 10 transitions.

```
zero <- matrix(0, 1, 99)
one <- matrix(1,1,1)
a <- as.matrix(cbind(one, zero))
library(expm)
```

```
## Loading required package: Matrix
```

```
##
```

```
## Attaching package: 'expm'
```

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

```
##
```

```
##      expm
```

```
prob10 <- a %*% (P %^% 10)
prob10[1,5]
```

```
##      V5
```

```
## 0.045091
```

```
print(paste("the probability of being in State 5 after 10 transitions is ", prob10[1,5]))
```

```
## [1] "the probability of being in State 5 after 10 transitions is  0.0450909981609373"
```

The probability of being in State 5 after 10 transitions is 0.0450909981609373"

problem 1(b)

Suppose we are at one of States 1,2, and 3 with equal probabilities. Find and display the probability of being in State 10 after 10 transitions.

```
first_three <- matrix(1/3, 1, 3)
zeros <- matrix(0, 1, 97)
init <- as.matrix((cbind(first_three, zeros)))
```

```

prob10_b <- init %*% (P %^% 10)
prob10_b[1,10]

##          V10
## 0.08268901

print(paste("the probability of being in State 10 after 10 transitions is ", prob10_b[1,10]))

## [1] "the probability of being in State 10 after 10 transitions is  0.0826890133603156"
The probability of being in State 10 after 10 transitions is 0.0826890133603156

```

problem 1(c)

Find the steady state probability of being in State 1.

```

Q <- t(P) - diag(100)
Q[100,] <- matrix(1,1,100)
zero_c <- matrix(0,1,99)
one_c <- matrix(1,1,1)
rhs <- as.matrix(cbind(zero_c, one_c))
Pi <- solve(Q) %*% t(rhs)
Pi[1,]

## [1] 0.01256589

print(paste("the steady state probability of being in State 1 is ", Pi[1,]))

## [1] "the steady state probability of being in State 1 is  0.0125658937530592"
The steady state probability of being in State 1 is 0.0125658937530593

```

problem 1(d)

Find the mean first passage time from State 1 to State 100.

```

B <- P[1:99,1:99]
Q <- diag(99) - B
e <- matrix(1,1,99)
m <- solve(Q) %*% t(e)
m[1,]

##          V1
## 254.9395

print(paste("the mean first passage time from State 1 to State 100 is ", m[1,]))

## [1] "the mean first passage time from State 1 to State 100 is  254.939463100014"
The mean first passage time from State 1 to State 100 is 254.939463100014

```

Problem 2

```
web <- read.table("webtraffic.txt", header=TRUE)
```

problem 2(a)

Construct 9 by 9 matrix Traffic that counts total traffic between State i to State j for all $i = 1, \dots, 9$ and $j = 1, \dots, 9$. Display Traffic. Hint `colSums()` adds all rows for each column.

```
colsums <- colSums(web)
Traffic <- matrix(colsums, nrow = 9, ncol = 9, byrow = TRUE)
Traffic
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
## [1,]    0  447  553    0    0    0    0    0    0
## [2,]    0   23  230  321    0    0    0    0   63
## [3,]    0  167   43  520    0    0    0    0   96
## [4,]    0    0    0   44  158  312  247    0  124
## [5,]    0    0    0    0   22   52   90  127  218
## [6,]    0    0    0    0   67   21    0  294   97
## [7,]    0    0    0    0    0   94    7  185   58
## [8,]    0    0    0    0  262    0    0   30  344
## [9,]    0    0    0    0    0    0    0    0    0
```

problem 2(b)

Observe that Traffic has 0's in row 9 and 0's in column 1. Set `Traffic[9,1]=1000`. Construct the one step transition probability matrix P and display it.

```
Traffic[9,1] = 1000
P_2 <- Traffic/rowSums(Traffic)
P_2
```

```
##      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
## [1,]    0 0.44700000 0.55300000 0.00000000 0.00000000 0.00000000
## [2,]    0 0.03610675 0.36106750 0.50392465 0.00000000 0.00000000
## [3,]    0 0.20217918 0.05205811 0.62953995 0.00000000 0.00000000
## [4,]    0 0.00000000 0.00000000 0.04971751 0.1785311 0.35254237
## [5,]    0 0.00000000 0.00000000 0.00000000 0.0432220 0.10216110
## [6,]    0 0.00000000 0.00000000 0.00000000 0.1398747 0.04384134
## [7,]    0 0.00000000 0.00000000 0.00000000 0.0000000 0.27325581
## [8,]    0 0.00000000 0.00000000 0.00000000 0.4119497 0.00000000
## [9,]    1 0.00000000 0.00000000 0.00000000 0.0000000 0.00000000
##      [,7]      [,8]      [,9]
## [1,] 0.00000000 0.00000000 0.00000000
## [2,] 0.00000000 0.00000000 0.0989011
## [3,] 0.00000000 0.00000000 0.1162228
## [4,] 0.27909605 0.00000000 0.1401130
## [5,] 0.17681729 0.24950884 0.4282908
## [6,] 0.00000000 0.61377871 0.2025052
## [7,] 0.02034884 0.53779070 0.1686047
## [8,] 0.00000000 0.04716981 0.5408805
## [9,] 0.00000000 0.00000000 0.00000000
```

problem 2(c)

Calculate and display the steady state probability vector Π .

```
Q_2 <- t(P_2) - diag(9)
Q_2[9,] <- matrix(1,1,9)
zero_q2 <- matrix(0,1,8)
one_q2 <- matrix(1,1,1)
rhs_q2 <- as.matrix(cbind(zero_q2, one_q2))
invQ_2 <- solve(Q_2)
Pi_2 <- solve(Q_2) %*% t(rhs_q2)
Pi_2
```

```
##           [,1]
## [1,] 0.15832806
## [2,] 0.10085497
## [3,] 0.13077897
## [4,] 0.14012033
## [5,] 0.08058898
## [6,] 0.07583914
## [7,] 0.05446485
## [8,] 0.10069664
## [9,] 0.15832806
```

problem 2(d)

The following table presents the average time that the visitors spend on each page. Page 1 2 3 4 5 6 7 8
Avg(minute) 0.1 2 3 5 5 3 3 2 Calculate and display the average time a visitor spend on the website (until she leaves).

```
T <- c(0.1, 2, 3, 5, 5, 3, 3, 2)
T <- as.matrix(T)
B2 <- P_2[1:8,1:8]
Q2 <- diag(8) - B2
m2 <- solve(Q2) %*% T
m2[1,]
```

```
## [1] 14.563
```

```
print(paste("the average time(minutes) a visitor spend on the website until she leaves is ", m2[1,]))
```

```
## [1] "the average time(minutes) a visitor spend on the website until she leaves is 14.563"
```

The average time a visitor spend on the website until she leaves is 14.563

problem 2(e)

In the output of Problem 2(c), observe that Pages 3 and 4 are one of the most crowded pages except Pages 1 and 9. To balance the traffic, the owner of the website decided to create links from Page 2 to Pages 6,7 (hence, from State 2 to States 6,7). By adding the links, the owner anticipates that, from Page 2, 30% of the current outgoing traffic to State 3 would move to State 6, and 20% of the current outgoing traffic to State 4 would move to State 7. Calculate new steady state probability vector Π_2 to check the effect of the new links. Decide if the link helped balancing the traffic by comparing the variance of Π and Π_2 . Hint Start with matrix Traffic from Problem 2(a).

$$\begin{aligned}
 t_{ij} &= p_{ij} \cdot t_i + \sum_{k \neq j} p_{ik} (t_i + t_{kj}) \\
 &= p_{ij} \cdot t_i + \sum_{k \neq j} p_{ik} t_i + \sum_{k \neq j} p_{ik} t_{kj} \\
 &= \sum_j p_{ij} t_i + \sum_{k \neq j} p_{ik} t_{kj} \\
 &= t_i \sum_j p_{ij} + \sum_{k \neq j} p_{ik} t_{kj} \\
 &= t_i + \sum_{k \neq j} p_{ik} t_{kj}
 \end{aligned}$$

$$\begin{aligned}
 \Rightarrow t &= T + BT \quad \text{where } T = \text{time spent on each page.} \\
 \Rightarrow (I - B)t &= T \\
 t &= (I - B)^{-1} T
 \end{aligned}$$

Figure 1: 2e)

```

Traffic_e <- as.matrix(Traffic)
t_23 <- Traffic_e[2,3]
t_26 <- t_23 * 0.3
t_23new <- t_23 * (1-0.3)
t_24 <- Traffic_e[2,4]
t_27 <- t_24 * 0.2
t_24new <- t_24 * (1-0.2)
Traffic_e[2,3] <- t_23new
Traffic_e[2,4] <- t_24new
Traffic_e[2,6] <- t_26
Traffic_e[2,7] <- t_27
P2 <- Traffic_e/rowSums(Traffic_e)
Q2 <- t(P2) - diag(9)
Q2[9,] <- matrix(1,1,9)
zeroq2 <- matrix(0,1,8)
oneq2 <- matrix(1,1,1)
rhsq2 <- as.matrix(cbind(zeroq2, oneq2))
invQ2 <- solve(Q2)
Pi2 <- solve(Q2) %*% t(rhsq2)
Pi2

```

```

##           [,1]
## [1,] 0.16162840
## [2,] 0.10034341
## [3,] 0.12104331
## [4,] 0.12275720
## [5,] 0.08164613
## [6,] 0.08250884
## [7,] 0.06003218
## [8,] 0.10841213
## [9,] 0.16162840

```

```
var(Pi_2)
```

```

##           [,1]
## [1,] 0.001410675

```

```
print(paste("the variance of Pi before balancing is ", var(Pi_2)))
```

```
## [1] "the variance of Pi before balancing is 0.00141067501207375"
```

```
var(Pi2)
```

```

##           [,1]
## [1,] 0.001219604

```

```
print(paste("the variance of Pi after balancing is ", var(Pi2)))
```

```
## [1] "the variance of Pi after balancing is 0.00121960422368109"
```

In the output of Problem 2(c), we observed that Pages 3 and 4 are one of the most crowded pages except Pages 1 and 9. The steady state probability of state 3 is 0.13077897 and of state 4 is 0.14012033. After adding the new link, we can see that both of state 3's and state 4's steady state probability dropped (0.12104331 and 0.12275720 respectively). Also, the steady state probabilities of State 6 and State 7 increased after adding the link. Although, the magnitude of the decrease is not very large, but adding this link from State 2 to States 6,7 still have the correct effect that we anticipated.

Also, to get a better understanding on the balancing effect, we calculated the variances of Pi1 and Pi2. Result:

before balancing, $\text{var}(\text{Pi}) = 0.001410675$ and after adding the new link, $\text{var}(\text{Pi2}) = 0.001219604$. Obviously, there is a drop in the variance, which implied that the link helped balancing the traffic.