# MSiA 400 Lab Assignment 4

Nov 30, 2015

- Due: 11:59 PM Dec 11, 2015
- This is an open book assignment.
- Using package is allowed only to calculate (i) the power of matrix and (ii) the solution of equation system.
- Please submit one report file (*.pdf, *.doc, *.docx): for each problem, attach (1) short answer if necessary, (2) every single line of codes you ran, (3) output.

## 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 |

### Problem 1(a)

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

Solution

```
> library(expm);
> P10 = P %^% 10;
> P10[1,5]
      V5
0.045091
```

### 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.

Solution

```
> a = rep(0,100);
> a[1:3]=1/3;
> dist10.given.a = a %*% P10;
> dist10.given.a[10];
[1] 0.08268901
```

### Problem 1(c)

Find the steady state probability of being in State 1.

Solution

```
> Q = t(P)-diag(100);
> Q[100,]=1;
> rhs = rep(0,100);
> rhs[100]=1;
> Pi = solve(Q) %*% rhs;
> Pi[1]
[1] 0.01256589
```

**Problem 1(d)**

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

SOLUTION

```
> B = P[1:99,1:99];
> Q = diag(99)-B;
> e = rep(1,99);
> m = solve(Q) %*% e;
> View(m)
[1] 254.9395
```

## Problem 2

You are asked to analyze the data from an website with 8 pages (Page 1 - Page 8). Let us assume that there is a virtual page (Page 9) that a visitor must automatically visit when the visitor leaves the website. The visitors always start their visit from Page 1. Let us formulate a Markov chain for this website. The states are defined as

$$S_i = \text{visitor is at Page } i,\ i = 1, \cdots, 9.$$

For example, suppose that a visitor enters the website (hence visit Page 1), moves to Page 3, Page 5, and then leave the website, sequentially. Then, the user visits States $S_1, S_3, S_5$, and $S_9$, sequentially.

Please find the attached data *webtraffic.txt*. The data includes the record of 1000 visitors (rows). The data has 81 columns labeled as $t11, t12, \cdots, t19, t21, t22, \cdots, t29, \cdots, t91, t92, \cdots, t99$. The label $t_{ij}$ represents the transition from State $i$ to State $j$, for $i = 1, \cdots, 9$ and $j = 1, \cdots, 9$. For example, $t12$ is the transition from State 1 to State 2, and $t84$ is the transition from State 8 to State 4. For each visitor (row), it has 1 for column $t_{ij}$ if the visitor makes transition from State $i$ to State $j$, and it has 0 elsewhere. For example, if a visitor visits States $S_1, S_3, S_5$, and $S_9$, sequentially, then the corresponding row has 1 for columns $t13, t35, t59$ and 0 elsewhere.

The summary of the data set is below.

| | |
|---|---|
| Name of the data set | webtraffic |
| Type of data | binaries (0,1) |
| Number of rows | 1000 |
| Number of columns | 81 |

**Problem 2(a)**

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

SOLUTION

```
> Traffic.vector = colSums(webtraffic);
> Traffic = matrix(Traffic.vector, nrow=9, byrow=T);
> 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.

SOLUTION

```
> Traffic[9,1]=1000;
> P = Traffic;
>
> for(i in 1:9){
+     P[i,]= Traffic[i,]/sum(Traffic[i,]);
+ }
>
> P
      [,1]     [,2]     [,3]     [,4]     [,5]     [,6]     [,7]     [,8]     [,9]
 [1,]    0 0.447000 0.553000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
 [2,]    0 0.036107 0.361068 0.503925 0.000000 0.000000 0.000000 0.000000 0.098901
 [3,]    0 0.202179 0.052058 0.629540 0.000000 0.000000 0.000000 0.000000 0.116223
 [4,]    0 0.000000 0.000000 0.049718 0.178531 0.352542 0.279096 0.000000 0.140113
 [5,]    0 0.000000 0.000000 0.000000 0.043222 0.102161 0.176817 0.249509 0.428291
 [6,]    0 0.000000 0.000000 0.000000 0.139875 0.043841 0.000000 0.613779 0.202505
 [7,]    0 0.000000 0.000000 0.000000 0.000000 0.273256 0.020349 0.537791 0.168605
 [8,]    0 0.000000 0.000000 0.000000 0.411950 0.000000 0.000000 0.047170 0.540881
 [9,]    1 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
```

**Problem 2(c)**

Calculate and display the steady state probability vector Pi.

SOLUTION

```
> Q = t(P)-diag(9);
> Q[9,]=1;
> rhs = rep(0,9);
> rhs[9]=1;
> Pi = solve(Q) %*% rhs;
> Pi
           [,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).

Solution

```
> avgtime.per.page = c(0.1,2,3,5,5,3,3,2,0)
> avgtime = avgtime.per.page %*% Pi
> avgtime
         [,1]
[1,] 2.305731
```

**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 Pi2 to check the effect of the new links. Decide if the link helped balancing the traffic by comparing the variance of Pi and Pi2.
*Hint* Start with matrix Traffic from Problem 2(a).

Solution

```
> Traffic2 = Traffic;
> Traffic2[2,6]=Traffic2[2,3]*0.3;
> Traffic2[2,3]=Traffic2[2,3]*0.7;
> Traffic2[2,7]=Traffic2[2,4]*0.2;
> Traffic2[2,4]=Traffic2[2,4]*0.8;
> P2 = Traffic2;
>
> for(i in 1:9){
+     P2[i,]= Traffic2[i,]/sum(Traffic2[i,]);
+ }
>
> Q2 = t(P2)-diag(9);
> Q2[9,]=1;
> rhs2 = rep(0,9);
> rhs2[9]=1;
> Pi2 = solve(Q2) %*% rhs2;
>
> var(Pi)
            [,1]
[1,] 0.001410675
> var(Pi2)
            [,1]
[1,] 0.001219604
```