LabAssignment_1

Tian Fu

10/24/2018

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.0.0     v purrr   0.2.5
## v tibble  1.4.2     v dplyr   0.7.6
## v tidyr   0.8.1     v stringr 1.3.1
## v readr   1.1.1     v forcats 0.3.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(expm)
```

```
## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following object is masked from 'package:tidyr':
##
##     expand

##
## Attaching package: 'expm'

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

Problem 1

```r
# load data
markov = read.table("~/Desktop/MSiA400/markov100.txt", head=FALSE)
P = as.matrix(markov) # transition probability matrix
```

```r
names = c(1:100) # assign names to each state, from 1 to 100
colnames(P) = names
rownames(P) = names
```

1(a)

```r
a <- rep(0, 100) # initial distribution
a[1] <- 1 # at State 1 with probability one
```

Since we are at State 1 now, in the initial distribution a, we are at State 1 with probability one.

```r
prob5 <- a %*% (P%^%10)
prob5
```

```
##                1          2          3          4        5        6
## [1,] 0.03210252 0.03315294 0.09941539 0.06462136 0.045091 0.048171
##                7          8         9         10        11         12
## [1,] 0.02981315 0.04957693 0.0678567 0.08126983 0.1194159 0.02849685
##               13         14         15         16         17         18
## [1,] 0.03397078 0.09048926 0.07560999 0.01825296 0.02909586 0.01210839
##               19           20          21          22          23
## [1,] 0.01702716 0.0004803007 0.001104914 0.009981056 0.001970024
##                24          25          26           27          28
## [1,] 0.003182653 0.001596568 0.005224745 0.0009182241 1.76305e-06
##                 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48
## [1,] 1.761169e-06  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##      49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71
## [1,]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##      72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94
## [1,]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##      95 96 97 98 99 100
## [1,]  0  0  0  0  0   0
```

```r
prob5[1,5]
```

```
##        5
## 0.045091
```

Given we are at State 1 now, the probability of being in State 5 after 10 transitions is:

$p_{1,5}(10) = 0.045091$

1(b)

```r
a2 <- rep(0, 100)
a2[1:3] <- 1/3
```

Since we are at one of States 1,2, and 3 with equal probabilities, the initial probability of being in one of these three is $\frac{1}{3}$

```r
prob10 <- a2 %*% (P%^%10)
prob10
```

```
##                1          2        3          4          5          6
## [1,] 0.03365299 0.03473429 0.103148 0.06647602 0.04691916 0.04978609
##                7          8          9         10        11         12
## [1,] 0.03016254 0.04937492 0.06766467 0.08268901 0.1200149 0.02824977
##               13         14         15         16        17         18
## [1,] 0.03319286 0.08936531 0.07294184 0.01694442 0.0272714 0.01097558
##               19          20           21          22          23
## [1,] 0.01564754 0.000398366 0.0009532201 0.008850776 0.001593833
##                24          25          26           27           28
## [1,] 0.002601547 0.001301921 0.004332957 0.0007160195 2.002671e-05
##                 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48
## [1,] 2.000535e-05  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
```

```
##      49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71
## [1,]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##      72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94
## [1,]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##      95 96 97 98 99 100
## [1,]  0  0  0  0  0   0
```

```r
prob10[1,10]
```

```
##         10
## 0.08268901
```

Given we are initially at State 1,2,3 with equal probability, the probability of being in State 10 after 10 transitions is:

$p_{i,10}(10) = 0.08268901$

where i=1,2,3 with equal probability

1(c)

```r
Q = t(P)-diag(100)
Q[100,] = rep(1,100) # replace the last row with all ones
rhs = rep(0,100)
rhs[100] = 1
Pi = solve(Q) %*% rhs
Pi # steady state probabilities
```

```
##             [,1]
## 1   0.0125658938
## 2   0.0091915303
## 3   0.0255231969
## 4   0.0187240817
## 5   0.0115751391
## 6   0.0121239209
## 7   0.0083136486
## 8   0.0152539239
## 9   0.0201849886
## 10  0.0227006031
## 11  0.0338552495
## 12  0.0083157699
## 13  0.0117828234
## 14  0.0289037115
## 15  0.0298638379
## 16  0.0095552449
## 17  0.0139102321
## 18  0.0079487586
## 19  0.0126688098
## 20  0.0047145165
## 21  0.0039498476
## 22  0.0143966186
## 23  0.0120881289
## 24  0.0173623697
## 25  0.0069709745
```

```
## 26  0.0265861556
## 27  0.0130348800
## 28  0.0037966314
## 29  0.0036423185
## 30  0.0010794500
## 31  0.0016783263
## 32  0.0074044721
## 33  0.0068897274
## 34  0.0014189398
## 35  0.0017930324
## 36  0.0109944087
## 37  0.0120631618
## 38  0.0129222379
## 39  0.0178424966
## 40  0.0082080694
## 41  0.0138441177
## 42  0.0046340255
## 43  0.0182176477
## 44  0.0023285120
## 45  0.0031248893
## 46  0.0009744419
## 47  0.0018880951
## 48  0.0037302200
## 49  0.0017600842
## 50  0.0037672559
## 51  0.0033658985
## 52  0.0036948907
## 53  0.0018605719
## 54  0.0047856565
## 55  0.0050767072
## 56  0.0023760149
## 57  0.0028493610
## 58  0.0101897104
## 59  0.0044646936
## 60  0.0108659367
## 61  0.0157483299
## 62  0.0080770433
## 63  0.0324202507
## 64  0.0214827961
## 65  0.0234131029
## 66  0.0624730859
## 67  0.0629301714
## 68  0.0220613102
## 69  0.0072403690
## 70  0.0132948300
## 71  0.0123211134
## 72  0.0163058220
## 73  0.0168143679
## 74  0.0062696308
## 75  0.0119441327
## 76  0.0048514374
## 77  0.0021763481
## 78  0.0014480905
## 79  0.0010413500
```

```
## 80  0.0026629249
## 81  0.0019927676
## 82  0.0007585553
## 83  0.0014826699
## 84  0.0013118807
## 85  0.0013298209
## 86  0.0025981346
## 87  0.0030361321
## 88  0.0062437668
## 89  0.0040033770
## 90  0.0044848282
## 91  0.0025584427
## 92  0.0008277174
## 93  0.0017875498
## 94  0.0015061251
## 95  0.0022677010
## 96  0.0032878670
## 97  0.0015725118
## 98  0.0019396680
## 99  0.0025299448
## 100 0.0039071739
```

As shown above, the steady state probability of being in State 1 is: $\pi_1 = 0.0125658938$

1(d)

```r
B = P[1:99,1:99] # exclude destination State 100
Q2 = diag(99) - B
e = rep(1,99)
m = solve(Q2) %*% e
m # mean first passage time from State i to State 100, i != 100
```

```
##          [,1]
## 1  254.939463
## 2  255.756780
## 3  255.553434
## 4  252.020994
## 5  254.689848
## 6  253.741872
## 7  253.572908
## 8  251.866892
## 9  249.390794
## 10 247.986623
## 11 244.808027
## 12 249.271139
## 13 244.048246
## 14 242.112390
## 15 232.030348
## 16 233.108471
## 17 217.761197
## 18 230.098329
## 19 207.242442
## 20 207.618923
```

```
## 21 202.962858
## 22 195.385117
## 23 201.024569
## 24 197.401354
## 25 197.152996
## 26 190.532650
## 27 172.667559
## 28 158.833969
## 29 153.487433
## 30 152.755309
## 31 156.096022
## 32 150.594867
## 33 149.507037
## 34 147.958820
## 35 148.265543
## 36 149.636556
## 37 146.005518
## 38 147.766862
## 39 143.005036
## 40 143.644636
## 41 148.465041
## 42 146.009273
## 43 138.935358
## 44 121.384733
## 45 120.998033
## 46 121.322578
## 47 129.179597
## 48 118.386090
## 49 120.272528
## 50 115.940519
## 51 115.605339
## 52 120.678477
## 53 112.276507
## 54 112.637460
## 55 111.270399
## 56 109.257070
## 57 113.623850
## 58 109.555265
## 59 109.371776
## 60 106.390883
## 61 107.104019
## 62 106.498192
## 63 103.143540
## 64 103.791399
## 65 104.727505
## 66 102.882204
## 67  99.471192
## 68  99.637840
## 69  72.153204
## 70  94.184715
## 71  33.476861
## 72  54.763689
## 73  50.445604
## 74  47.521308
```

```
## 75  46.317640
## 76  44.836599
## 77  16.035969
## 78  19.133208
## 79  14.031529
## 80  12.607691
## 81  11.518157
## 82  13.244137
## 83  12.369547
## 84  13.130539
## 85  11.202843
## 86  11.427886
## 87  10.839776
## 88   9.471288
## 89   8.418722
## 90   6.158265
## 91   6.321574
## 92   5.554380
## 93   4.674446
## 94   6.190252
## 95   4.431775
## 96   3.795454
## 97   3.521875
## 98   2.931884
## 99   2.389126
```

From above, the mean first passage time from State 1 to State 100 is: $m_{1,100} = 254.939463$

Problem 2

```r
# load data
webtraffic <- read.table("~/Desktop/MSiA400/webtraffic.txt", head=TRUE)
```

2(a)

```r
columnSum <- unname(colSums(webtraffic))
Traffic <- t(matrix(columnSum, nrow = 9, ncol = 9))
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
```

From above, matrix Traffic that counts total traffic between State i to State j for all i = 1,2,3,4,5,6,7,8,9 and j = 1,2,3,4,5,6,7,8,9 is:

$$Traffic = \begin{bmatrix} 0 & 447 & 553 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 23 & 230 & 321 & 0 & 0 & 0 & 0 & 63 \\ 0 & 167 & 43 & 520 & 0 & 0 & 0 & 0 & 96 \\ 0 & 0 & 0 & 44 & 158 & 312 & 247 & 0 & 124 \\ 0 & 0 & 0 & 0 & 22 & 52 & 90 & 127 & 218 \\ 0 & 0 & 0 & 0 & 67 & 21 & 0 & 294 & 97 \\ 0 & 0 & 0 & 0 & 0 & 94 & 7 & 185 & 58 \\ 0 & 0 & 0 & 0 & 262 & 0 & 0 & 30 & 344 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

2(b)

```r
Traffic[9,1] <- 1000

for (i in 1:nrow(Traffic)){
  rowSum = sum(Traffic[i,])
  Traffic[i,] = Traffic[i,] / rowSum
}

Traffic # P
```

```
##       [,1]       [,2]       [,3]       [,4]      [,5]       [,6]
##  [1,]    0 0.44700000 0.55300000 0.00000000 0.0000000 0.00000000
##  [2,]    0 0.03610675 0.36106750 0.50392465 0.0000000 0.00000000
##  [3,]    0 0.20217918 0.05205811 0.62953995 0.0000000 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 0.0000000
##  [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 0.0000000
```

One step transition probability matrix is:

$$P = \begin{bmatrix} 0 & 0.44700000 & 0.55300000 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.03610675 & 0.36106750 & 0.50392465 & 0 & 0 & 0 & 0 & 0.0989011 \\ 0 & 0.20217918 & 0.05205811 & 0.62953995 & 0 & 0 & 0 & 0 & 0.1162228 \\ 0 & 0 & 0 & 0.04971751 & 0.1785311 & 0.35254237 & 0.27909605 & 0 & 0.1401130 \\ 0 & 0 & 0 & 0 & 0.0432220 & 0.10216110 & 0.17681729 & 0.24950884 & 0.4282908 \\ 0 & 0 & 0 & 0 & 0.1398747 & 0.04384134 & 0 & 0.61377871 & 0.2025052 \\ 0 & 0 & 0 & 0 & 0 & 0.27325581 & 0.02034884 & 0.53779070 & 0.1686047 \\ 0 & 0 & 0 & 0 & 0.4119497 & 0 & 0 & 0.04716981 & 0.5408805 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

2(c)

```
Q3 = t(Traffic)-diag(9)
Q3[9,] = rep(1,9)
rhs2 = rep(0,9)
rhs2[9] = 1
Pi2 = solve(Q3) %*% rhs2
Pi2
```

```
##             [,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
```

From above, the steady state probability vector is:

$$\pi = \begin{bmatrix} 0.15832806 \\ 0.10085497 \\ 0.13077897 \\ 0.14012033 \\ 0.08058898 \\ 0.07583914 \\ 0.05446485 \\ 0.10069664 \\ 0.15832806 \end{bmatrix}$$

2(d)

```
avg_time <- c(0.1,2,3,5,5,3,3,2,0)
expected <- t(Pi2) %*% avg_time
expected
```

```
##          [,1]
## [1,] 2.305731
```

The average time a visitor spend on a page is 2.305731 minutes

```r
B2 = Traffic[1:8,1:8] # exclude destination State 9
Q_m = diag(8) - B2
e2 = rep(1,8)
m2 = solve(Q_m) %*% e2
m2 # mean first passage time from State i to State 9, i != 9
```

```
##          [,1]
## [1,] 5.316000
## [2,] 4.401776
## [3,] 4.246666
## [4,] 3.392390
## [5,] 2.429794
## [6,] 2.749343
## [7,] 2.940475
## [8,] 2.100010
```

From above, the mean first passage time from Page 1 to Page 9 is: $m_{1,9} = 5.316$

Therefore, average time a visitor spend on the website (until she leaves) is:

5.316*2.305731 = 12.25727 minutes

2(e)

In the output 2(c), Pages 3 and 4 have higher values than others, excluding Pages 1 and 9.

```r
Traff <- t(matrix(columnSum, nrow = 9, ncol = 9))
Traff # Traffic matrix
```

```
##       [,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
```

```r
two_to_three <- Traff[2,3] #current outgoing traffic to State3 from State2
two_to_four <- Traff[2,4] #current outgoing traffic to State4 from State2
# new assignments after linking Page 2 to 6,7
Traff[2,3] <- two_to_three*0.7
Traff[2,6] <- two_to_three*0.3
Traff[2,4] <- two_to_four*0.8
Traff[2,7] <- two_to_four*0.2
Traff
```

```
##       [,1] [,2] [,3]  [,4] [,5] [,6]  [,7] [,8] [,9]
##  [1,]    0  447  553   0.0    0    0   0.0    0    0
##  [2,]    0   23  161 256.8    0   69  64.2    0   63
##  [3,]    0  167   43 520.0    0    0   0.0    0   96
##  [4,]    0    0    0  44.0  158  312 247.0    0  124
```

```
## [5,]    0    0    0  0.0   22   52  90.0  127  218
## [6,]    0    0    0  0.0   67   21   0.0  294   97
## [7,]    0    0    0  0.0    0   94   7.0  185   58
## [8,]    0    0    0  0.0  262    0   0.0   30  344
## [9,]    0    0    0  0.0    0    0   0.0    0    0
```

calculate new steady state probability

```
Traff[9,1] <- 1000
# calculate one step transtion probability matrix and stored in Traff
for (i in 1:nrow(Traff)){
  rowSum = sum(Traff[i,])
  Traff[i,] = Traff[i,] / rowSum
}

Q4 = t(Traff)-diag(9)
Q4[9,] = rep(1,9)
rhs3 = rep(0,9)
rhs3[9] = 1
Pi3 = solve(Q4) %*% rhs3
Pi3
```

```
##            [,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
```

The new steady state probability vector is:

$$\pi_2 = \begin{bmatrix} 0.16162840 \\ 0.10034341 \\ 0.12104331 \\ 0.12275720 \\ 0.08164613 \\ 0.08250884 \\ 0.06003218 \\ 0.10841213 \\ 0.16162840 \end{bmatrix}$$

Comparing with original π, after creating new links from Page 2 to Page 6,7, steady state probabilities in Page 3,4 decrease and steady state probability in Page 2 slightly decreases. Steady state probabilities in other Pages all increase.

```
# variance of pi before change
var(Pi2)
```

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

```r
# variance of pi after change
var(Pi3)
```

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

From above, $var(\pi) = 0.001410675 > var(\pi_2) = 0.001219604$

As the variance of steady state probability decreases after creating new links, there is less variation after introducing the new links. So, the link helped balancing the traffic.