Lecture C2. Discrete Time Markov Chain 2

Sim, Min Kyu, Ph.D., mksim@seoultech.ac.kr

서울과학기술대학교 데이터사이언스학과

1. I. Stationary distribution
2. II. Numerical approach to find a stationary distribution
3. III. Limiting probailities
4. IV. When [limit=stationary] falls apart.

I. Stationary distribution

Warm-up

Exercise 1

Suppose we are considering the soda problem with $\mathrm{P} = \begin{array}{c} coke \\ pepsi \end{array} \begin{pmatrix} 0.7 & 0.3 \\ 0.5 & 0.5 \end{pmatrix}$, *and there are 20 people who drink coke today and 10 people who drink pepsi today. What will happen tomorrow?*

Coke: 14 + 5 = 19 Pepsi: 6 + 5 = 11

$a_0 = (2/3 \quad 1/3)$

$a_0 P = (2/3 \quad 1/3) \begin{pmatrix} .7 & .3 \\ .5 & .5 \end{pmatrix} = \left(\frac{14}{30} + \frac{5}{30} \quad \frac{6}{30} + \frac{5}{30} \right)$

$= \left(\frac{19}{30} \quad \frac{11}{30} \right)$

Exercise 2

Again with the soda problem with

$$\mathrm{P} = \begin{array}{c} \textit{coke} \\ \textit{pepsi} \end{array} \begin{pmatrix} 0.7 & 0.3 \\ 0.5 & 0.5 \end{pmatrix}$$

Suppose that we happen to have a distribution of S_k on day k such as $a_k = (5/8, 3/8)$.

- *What is a_{k+1}?*

  $(5/8 \quad 3/8)\begin{pmatrix} .7 & .3 \\ .5 & .5 \end{pmatrix} = \left(\frac{35}{80} + \frac{15}{80} \quad \frac{15}{80} + \frac{15}{80}\right) = \left(\frac{50}{80} \quad \frac{30}{80}\right)$

- *What is a_{k+2}?* $(5/8 \quad 3/8)$

- *What is a_∞?* $(5/8 \quad 3/8)$

  $\lim_{n\to\infty} a_n$

Starting from "this" distribution, the distribution will not ever change no matter how many transitions occur! Does it worth having a special name? How would you name it?

Definition

Definition 1 (stationary distribution)

For a DTMC with state space S and transition probability matrix P, a vector v whose length is $|S|$ is said to be a *stationary distribution* if

- $\mathrm{v}_i \geq 0$ for all $i \in S$ and $\sum_{i \in S} \mathrm{v}_i = 1$
- $\mathrm{v} = \mathrm{v}P$

Remark 1

The two bullets in the above definition can be understood as:

- v is a legit distribution.
- Going through a transition does not change the distribution.

Discussion

- "The distribution does not change" does not mean that there is no movement between states.
- It is rather that movement flows coincide for each state.
- In other words, it is not a *static* equilibrium but a *dynamic* equilibrium.
- It is called *steady state*, because it looks steady from outside look.
- Under the *steady state*, $(\text{inflow})_i = (\text{outflow})_i$ for $\forall i \in S$

Flow balance equation

$$(\text{inflow})_{\text{coke}} = (\text{outflow})_{\text{coke}}$$

$$P = \begin{pmatrix} P_{cc} & P_{cp} \\ P_{pc} & P_{pp} \end{pmatrix}$$

$$v = (v_c \quad v_p)$$

$$v_p \underline{P_{pc}} = v_c P_{cp}$$

$$0.5 v_p = 0.3 v_c \quad \text{— (1)}$$

$$v_p + v_c = 1 \quad \text{— (2)}$$

Computation of stationary distribution

- Using definition

$v = vP$

$(v_1 \quad v_2) = (v_1 \quad v_2)\begin{pmatrix} .7 & .3 \\ .5 & .5 \end{pmatrix}$

$= (.7v_1 + .5v_2 \quad .3v_1 + .5v_2)$

$.3v_1 = .5v_2$ — ①

$.5v_2 = .3v_1$

$v_1 + v_2 = 1$ — ②

$v_1 = \frac{5}{8}, \quad v_2 = \frac{3}{8}$

- Using flow balance equation

(inflow)coke = (outflow)coke

$v_p P_{pc} = v_c P_{cp}$

$0.5 v_p = 0.3 v_c$ — ①

$v_c + v_p = 1$ — ②

II. Numerical approach to find a stationary distribution

Mathematical aspects

- Remind that for a DTMC with S and P, a vector v of length $|S|$ is a stationary distribution if 1) $\mathrm{v}_i \geq 0$ for all $i \in S$ and $\sum_{i\in S} \mathrm{v}_i = 1$ and 2) $\mathrm{v} = \mathrm{v}P$.

Remark 2

?? With a DTMC's transition matrix P, the number of solution to $\mathrm{x} = \mathrm{xP}$ is either one or infinite. In other words, the stationary distribution always exists, and it may be unique or infinite.

Method 1 - eigen-decomposition

$Ax = \lambda x$

$(AB)^t = B^t A^t$

Remark 3

$xP = x \Rightarrow P^t x^t = x^t \Rightarrow P^t x^t = 1 \cdot x^t$. That is, a stationary distribution v is nothing but an eigenvector of P^t, which corresponds to its eigenvalue 1.

```
P <- array(c(0.7, 0.5, 0.3, 0.5), dim = c(2,2))
eigen(t(P)) # eigen-decomposition for P^t
```

```
## eigen() decomposition
## $values
## [1] 1.0 0.2
##
## $vectors
##           [,1]       [,2]
## [1,] 0.8574929 -0.7071068
## [2,] 0.5144958  0.7071068
```

- It can be seen that the matrix P^t has eigenvalue 1.

- The eigenvector corresponding to eigenvalue 1 needs to be normalized so that it becomes a legit distribution.

```
x_1 <- eigen(t(P))$vectors[,1]
x_1
```

```
## [1] 0.8574929 0.5144958
```

```
v <- x_1/sum(x_1)
v
```

```
## [1] 0.625 0.375
```

- The stationary distribution is found!

Method 2 - system of linear equation

$(v_1 \quad v_2)\begin{pmatrix}1\\1\end{pmatrix} = 1$

Remark 4

The two conditions for stationary distribution can be written in vector notation as follows. column vector

1. $\text{v}1 = 1$, where 1 is a column vector whose length is $|S|$.
2. $\text{vP} = \text{v}$ $\Rightarrow \text{vP} = \text{vI} \Rightarrow \text{v}\cdot(\text{P}-\text{I}) = 0$

- The first condition can be described as

$$\begin{pmatrix}-\text{v}-\end{pmatrix}\begin{pmatrix}1\\1\\1\end{pmatrix} = 1$$

- The second condition can be developed as $\text{vP} = \text{v} \Rightarrow \text{vP} = \text{vI} \Rightarrow \text{v}\big(\text{P} - \text{I}\big) = 0$

$$\begin{pmatrix}- & \text{v} & -\end{pmatrix}\begin{pmatrix} & | & \\ - & P - I & - \\ & | & \end{pmatrix} = \begin{pmatrix}0 & 0 & 0\end{pmatrix}$$

- These can be concatenated to a single system of linear equations.

$$\begin{pmatrix} - & \mathrm{v} & - \end{pmatrix} \begin{pmatrix} & | & & 1 \\ - & P-I & - & 1 \\ & | & & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix}$$

Now, we are ready to make the following remark.

Remark 5

Letting $\mathrm{A} = [\mathrm{P} - \mathrm{I}|1]$ and $\mathrm{b} = [0^t \ \ 1]^t$ gives $\mathrm{vA} = \mathrm{b}$. Since A is not a square matrix but a dimension of $|S| \times (|S| + 1)$ the stationary distribution can be found by:

$$\begin{aligned} & \mathrm{vA} = \mathrm{b} \\ \Rightarrow \quad & \mathrm{A}^t \mathrm{v}^t = \mathrm{b}^t \\ \Rightarrow \quad & \mathrm{AA}^t \mathrm{v}^t = \mathrm{Ab}^t \\ \Rightarrow \quad & \mathrm{v}^t = (\mathrm{AA}^t)^{-1} \mathrm{Ab}^t \end{aligned}$$

```r
P <- array(c(0.7, 0.5, 0.3, 0.5), dim = c(2,2))
n <- nrow(P) # n=|S|
I <- diag(n) # identity matrix
A <- cbind(P-I, rep(1,n))
b <- array(c(rep(0,n),1), dim = c(1, n+1))
A
```

```
##      [,1] [,2] [,3]
## [1,] -0.3  0.3    1
## [2,]  0.5 -0.5    1
```

```r
b
```

```
##      [,1] [,2] [,3]
## [1,]    0    0    1
```

- Using $AA^t v^t = Ab^t$,

```r
v <- solve(A %*% t(A), A %*% t(b))
v
```

```
##       [,1]
## [1,] 0.625
## [2,] 0.375
```

III. Limiting probailities

Motivation

- n-step transition probability: $\mathbb{P}(S_{t+n} = j | S_t = i) = \mathrm{P}_{ij}^n$
- Letting $n \to \infty$ to see what happens!

```
library(expm) # provides matrix power
P <- array(c(0.7, 0.5, 0.3, 0.5), dim = c(2,2))
P
```

```
##      [,1] [,2]
## [1,]  0.7  0.3
## [2,]  0.5  0.5
```

```
P %*% P # matrix multiplication
```

```
##      [,1] [,2]
## [1,] 0.64 0.36
## [2,] 0.60 0.40
```

```
P %^% 3
```

```
##       [,1]  [,2]
## [1,] 0.628 0.372
## [2,] 0.620 0.380
```

```
P %^% 4
```

```
##        [,1]   [,2]
## [1,] 0.6256 0.3744
## [2,] 0.6240 0.3760
```

```
P %^% 20
```

```
           C     P
##       [,1]  [,2]
C ## [1,] 0.625 0.375
P ## [2,] 0.625 0.375
```

- The limiting distribution exists in this case.
- Each row of matrix converges to stationary distribution.

$$\mathrm{P}^{\infty} = \begin{pmatrix} 5/8 & 3/8 \\ 5/8 & 3/8 \end{pmatrix} = \begin{pmatrix} - & \mathrm{v} & - \\ - & \mathrm{v} & - \end{pmatrix}$$

- In the soda example, what happens today has little effect in the long run. That is, limiting probability is independent of initial state. Initial distribution does not matter in the long run.

- The limiting distribution may or may not exist. For example,

```
P <- array(c(0, 1, 1, 0), dim = c(2,2))
P
```

```
##      [,1] [,2]
## [1,]    0    1
## [2,]    1    0
```

```
P %^% 2
```

```
##      [,1] [,2]
## [1,]    1    0
## [2,]    0    1
```

```
P %^% 3
```

```
##      [,1] [,2]
## [1,]    0    1
## [2,]    1    0
```

$vP = v$

$(v_1 \quad v_2)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = (v_1 \quad v_2)$

$(v_2 \quad v_1) = (v_1 \quad v_2)$

$\Rightarrow \therefore v = (1/2 \quad 1/2)$

Using [limiting probabilities = stationary distribution]

- If I do this for 10 years (3650 days) from now, then how many days I will drink Pepsi?

  $3650 \times \frac{3}{8}$

- Suppose Coke is \$1.5 and Pepsi is \$1. How much on average I spend on soda in a month?

  $1.5 \times 30 \times \frac{5}{8} + 1.0 \times 30 \times \frac{3}{8}$

- In the above question of '*staying at a certain state costs*' is a motivation for upcoming *Markov reward process (MRP)*.

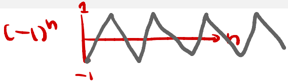
- Suppose there are a billion customers (who has same type of consuming pattern) like me in the world. You are working for Pepsi and like to boost Pepsi $\rightarrow$ Pepsi probability from 0.5 to 0.6 by marketing. On average, how much additional revenue will be generated by this change for a day?

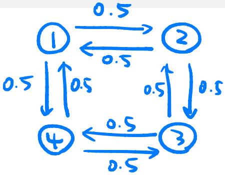
$$\begin{bmatrix} 0.7 & 0.3 \\ 0.5 & 0.5 \end{bmatrix} \rightarrow v = (5/8 \;\; 3/8) \qquad \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix} \rightarrow v = ?(\quad \bigcirc)$$

$$\underbrace{1003 \times 3/8} \qquad \underbrace{1003 \times \bigcirc}$$

IV. When [limit=stationary] falls apart.

When things not going well 1: Periodic MC





- Transition diagram
- Demonstration

$$P = \begin{pmatrix} 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix}$$

$$P^2 = \begin{bmatrix} 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \end{bmatrix}$$

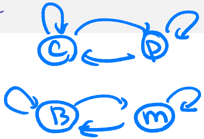
cf) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = P$

$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $= P^2$

- Observations
  1. Limiting probability NOT exists
  2. Stationary distribution is unique
- Remedy
  - $\lim_{n\to\infty} \frac{P^{n+1}+P^{n+2}+\ldots+P^{n+d}}{d}$ exists and same as stationary distribution.

$$\lim_{n\to\infty} \frac{P^n + P^{n+1}}{2} = \begin{bmatrix} - v - \\ - v - \end{bmatrix}$$

When things not going well 2: Reducible MC



- Transition diagram
- Demonstration

$$P^{\infty} = \begin{bmatrix} 5/8 & 3/8 & 0 & 0 \\ 5/8 & 3/8 & 0 & 0 \\ 0 & 0 & 3/7 & 4/7 \\ 0 & 0 & 3/7 & 4/7 \end{bmatrix}$$

$$\mathrm{P} = \begin{array}{c} \textit{Coke} \\ \textit{Pepsi} \\ \textit{Bud} \\ \textit{Miller} \end{array} \begin{pmatrix} 0.7 & 0.3 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.6 & 0.4 \\ 0 & 0 & 0.3 & 0.7 \end{pmatrix}$$

3 4

① Stationary ?

② P^{∞} = ? (limit) !

- Observations
  1. Limiting probability exists.
  2. But, Limiting probability depends on the initial state.
  3. Stationary distribution is not unique (∞)

$v = \alpha(5/8 \;\; 3/8 \;\; 0 \;\; 0)$
$\quad + (1-\alpha)(0 \;\; 0 \;\; 3/7 \;\; 4/7)$

for $\alpha \in [0,1]$

$v = (5/8 \;\; 3/8 \;\; 0 \;\; 0)$

$v = (0 \;\; 0 \;\; 3/7 \;\; 4/7)$

$v = (\frac{1}{2} \times 5/8 \;\; \frac{1}{2} \times 3/8 \;\; \frac{1}{2} \times 3/7 \;\; \frac{1}{2} \times 4/7)$

$v = (\frac{2}{3} \times 5/8 \;\; \frac{2}{3} \times 3/8 \;\; \frac{1}{3} \times 3/7 \;\; \frac{1}{3} \times 4/7)$

Summary of observations

For a *finite* state space MC,

MC	Limiting	Stationary	Remark
Irreducible Aperiodic	Exists, indep. of initial state	Unique	NICE!
Irreducible Periodic	Not Exists	Unique	Remedy by average of d
Reducible Aperiodic	Exists, dependent on initial state	maybe ∞	Deeper look

A few definitions (1)

- Accessibility
  - Def. A state i can *reach* state j and write $i \to j$ if $\exists n$ s.t. $P_{ij}^{n} > 0$.
    -
    -
    -
  - Def. State i and j are said to *communicate* and write $i \leftrightarrow j$ if $i \to j$ and $j \to i$.
  - Def. A group of states that communicate is said to be a *class*.

A few definitions (2)

- Reducibility
  - Def. MC S_n is said to be *irreducible* if all states communicate.
  - Def. MC S_n is said to be *irreducible* if $\exists$ only one class.
  - Def. MC S_n is said to be *reducible* unless *irreducible*.

A few definitions (3)

- Periodicity
  - Def. For a state $i \in S$, *period* $d(i) := gcd\{n, P_{ii}^{n} > 0\}$.
  - Def. MC S_n is said to be *periodic* if $\exists\, i$ with $d(i) > 1$.
  - Def. MC S_n is said to be *aperiodic* if not *periodic*.
  - Remark: Periodicity is class property.
    (Class shares period; $i \leftrightarrow j \Rightarrow d(i) = d(j)$)

So, when does it go well?

Theorem 1

If a finite DTMC S_n is aperiodic and irreducible, then all of the followings hold:

- *Limiting probabilities exists*
- *Stationary distribution is unique.*
- *Stationary distribution = Limiting probabilities.*

Remark 6

Above theorem implies the following:

- Finite, Aperiodic, Irreducible $\Rightarrow \lim_{n\to\infty} \mathrm{P}_{ij}^{n} = \mathrm{v}_j, \forall i, j \in S$

- In these "nice" cases, we can talk about things like "The long-run fraction of time that the MC spends in each state".
- In these "nice" cases, we can calculate limiting probability by solving stationary distribution.

```
"Faber est suae quisque fortunae - 운명을 만드는 사람은 그 자신이다."
```