Lecture C2. Discrete Time Markov Chain 2

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

서울과학기술대학교 데이터사이언스학과

I. Stationary distribution

Warm-up

Exercise 1

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

Exercise 2

Again with the soda problem with

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

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}?*
- *What is a_{k+2}?*
- *What is a_∞?*

Once you get into "this" distribution, the distribution will not ever change over time! Does it worth having a special name?

Definition

Definition 1 (stationary distribution)

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

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

Remark 1

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

- $\mathbf{v}$ 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* in a way, it looks steady from outside look.
- Under the *steady state*, $(\text{inflow})_i = (\text{outflow})_i$ for $\forall i \in S$

Flow balance equation

Computation of stationary distribution

- Using definition
- Using flow balance equation

II. Numerical approach for stationary distribution

Mathematical aspects

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

Remark 2

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

Exercise 3

General linear system $A\mathbf{x} = \mathbf{b}$ has the number of solutions: 0, or 1, or ∞. However, the Remark implies that $\mathbf{x} = \mathbf{x}\mathbf{P}$ is always consistent. Prove that it always has a solution.

Method 1 - eigen-decomposition

Remark 3

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

```r
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.

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

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

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

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

- The stationary distribution is found!

Method 2 - system of linear equation

Remark 4

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

1. $\mathbf{v1} = 1$, where $\mathbf{1}$ is a column vector whose length is $|S|$.
2. $\mathbf{vP} = \mathbf{v}$

- The first condition can be described as

$$(-\mathbf{v}-)\begin{pmatrix}1\\1\\1\end{pmatrix} = 1$$

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

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

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

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

Now, we are ready to make the following remark.

Remark 5

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

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

```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 $\mathbf{A}\mathbf{A}^t\mathbf{v}^t = \mathbf{A}\mathbf{b}^t$,

```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: $\mathbb{P}(S_{t+n} = j | S_t = i) = \mathbf{P}_{ij}^n$
- 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
```

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

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

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

- In the long run, what happens today has little effect. 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
```

Using [limiting probabilities = stationary distribution]

- If I do this for 10 years (3650 days) from now, then how many days I will drink Pepsi?
- Suppose Pepsi is \$1 and Coke is \$1.5. How much on average I spend on soda in a month?
- In the above question of '*staying at a certain state costs*' is an example of *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?

IV. When [limit=stationary] falls apart.

When things not going well 1: Periodic MC

- Transition diagram
- Demonstration

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

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

When things not going well 2: Reducible MC

- Transition diagram
- Demonstration

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

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

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: 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.
- Above theorem can be also written as:
  - Finite, Aperiodic, Irreducible $\Rightarrow \lim_{n\to\infty} \mathbf{P}_{ij}^n = \mathbf{v}_j, \forall i, j \in S$
- 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.

```r
cat(str)
```

```
## If I only had an hour to chop down a tree, I would spend the first 45 minutes sharpening my axe. -
A. Lincoln
```