MARKOV CHAIN

Teruo Nakatsuma

Spring Semester 2021

Faculty of Economics, Keio University

Since

$$\mathbf{P}(A) = \mathbf{P}(A|B)\mathbf{P}(B) + \mathbf{P}(A|B^c)\mathbf{P}(B^c),$$
$$\mathbf{P}(A^c) = \mathbf{P}(A^c|B)\mathbf{P}(B) + \mathbf{P}(A^c|B^c)\mathbf{P}(B^c).$$

we have

$$\begin{bmatrix} \mathbf{P}(A) \\ \mathbf{P}(A^c) \end{bmatrix} = \begin{bmatrix} \mathbf{P}(A|B) & \mathbf{P}(A|B^c) \\ \mathbf{P}(A^c|B) & \mathbf{P}(A^c|B^c) \end{bmatrix} \begin{bmatrix} \mathbf{P}(B) \\ \mathbf{P}(B^c) \end{bmatrix}.$$

Markov Chain ii

Let A_t denote the event that A occur at time t and we replace A and B with A_{t+1} and A_t respectively. Then

$$\begin{bmatrix} \mathbf{P}(A_{t+1}) \\ \mathbf{P}(A_{t+1}^c) \end{bmatrix} = \begin{bmatrix} \mathbf{P}(A_{t+1}|A_t) & \mathbf{P}(A_{t+1}|A_t^c) \\ \mathbf{P}(A_{t+1}^c|A_t) & \mathbf{P}(A_{t+1}^c|A_t^c) \end{bmatrix} \begin{bmatrix} \mathbf{P}(A_t) \\ \mathbf{P}(A_t^c) \end{bmatrix}.$$

This is called a Markov chain. In the context of the Markov chain, each event is often referred to as a state and the conditional probability $\mathbf{P}(A_{t+1}|A_t^c)$ is interpreted as the probability that the chain moves from state A^c at t to state A at $t+1$.

Markov Chain iii

The matrix

$$\begin{bmatrix} \mathbf{P}(A_{t+1}|A_t) & \mathbf{P}(A_{t+1}|A_t^c) \\ \mathbf{P}(A_{t+1}^c|A_t) & \mathbf{P}(A_{t+1}^c|A_t^c) \end{bmatrix}.$$

is called the transition matrix. When the transition matrix does not change over time, the Markov chain is time-homogeneous. The vector of probabilities at time $t = 0$ is called the initial probability vector.

In economics, finance and other fields, the Markov chain is widely used for modeling a time-varying probability.

Markov Chain iv

Let $p_{1t} = \mathbf{P}(A_t)$, $p_{2t} = \mathbf{P}(A_t^c)$, and let π_{ij} denote the (i, j) element of the time-homogeneous transition matrix $(i, j = 1, 2)$. Then the Markov chain is rewritten as

$$\begin{bmatrix} p_{1,t+1} \\ p_{2,t+1} \end{bmatrix} = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix} \begin{bmatrix} p_{1,t} \\ p_{2,t} \end{bmatrix}.$$

In general, the Markov chain with k states is given by

$$\underbrace{\begin{bmatrix} p_{1,t+1} \\ \vdots \\ p_{k,t+1} \end{bmatrix}}_{p_{t+1}} = \underbrace{\begin{bmatrix} \pi_{11} & \cdots & \pi_{1k} \\ \vdots & \ddots & \vdots \\ \pi_{k1} & \cdots & \pi_{kk} \end{bmatrix}}_{\Pi} \underbrace{\begin{bmatrix} p_{1,t} \\ \vdots \\ p_{k,t} \end{bmatrix}}_{p_t}.$$

Properties

1. $\boldsymbol{p}_{t+k} = \boldsymbol{\Pi}^k \boldsymbol{p}_t \ (k = 1, 2, \ldots)$.
2. In particular, $\boldsymbol{p}_t = \boldsymbol{\Pi}^t \boldsymbol{p}_0$. Thus all $\boldsymbol{p}_t$ in a Markov chain is determined by the initial probability vector $\boldsymbol{p}_1$ and the transition matrix $\boldsymbol{\Pi}$.
3. The probability vector $\boldsymbol{p}^*$ satisfies
$$\boldsymbol{p}^* = \boldsymbol{\Pi} \boldsymbol{p}^*,$$
is called the stationary probability vector or stationary distribution of a Markov chain.
4. When the number of states is two, the stationary distribution is given by
$$p_1^* = \frac{1 - \pi_{22}}{2 - \pi_{11} - \pi_{22}}, \quad p_2^* = \frac{1 - \pi_{11}}{2 - \pi_{11} - \pi_{22}}.$$

Example: Business Cycle

Suppose that the business cycle of a country is determined by the following transition matrix:

Table 1: Transition Matrix $\boldsymbol{\Pi}$

	Previous Quarter	
Current Quarter	Expansion (E)	Recession (E^c)
Expansion (E)	0.9	0.25
Recession (E^c)	0.1	0.75

Then the stationary distribution is given by

$$\mathbf{P}(E) = \frac{1 - 0.75}{2 - 0.9 - 0.75} = \frac{5}{7}, \quad \mathbf{P}(E^c) = 1 - \mathbf{P}(E) = \frac{2}{7}.$$

Markov Chain Of Continuous Random Variables

Consider a sequence of continuous random variables $\{X_t\}_{t=0}^{\infty}$. X_t takes a real value in $\mathcal{X} \subseteq \mathbb{R}$.

$\{X_t\}_{t=0}^{\infty}$ is called a Markov chain if, given $\{x_s\}_{s=0}^{t-1}$, the conditional probability that X_t takes a real value in $A \subseteq \mathcal{X}$ is expressed as

$$\begin{aligned} &\mathbf{Pr}\{X_t \in A | X_0 = x_0, \dots, X_{t-1} = x_{t-1}\} \\ &= \mathbf{Pr}\{X_t \in A | X_{t-1} = x_{t-1}\}. \end{aligned} \tag{1}$$

Properties i

- Time-homogeneity
  For any $A \subseteqq \mathcal{X}$, $x \in \mathcal{X}$, $t \geqq 0$,
  $\mathbf{Pr}\{X_{t+1} \in A | X_t = x\} = \mathbf{Pr}\{X_t \in A | X_{t-1} = x\}$.
- Regularity
  Suppose $\int_A f(x)dx > 0$ for any $A \subseteqq \mathcal{X}$. A Markov chain $\{X_t\}_{t=0}^{\infty}$ is regular with respect to f or f-regular if there exists a finite $t \geqq 1$ such that $\mathbf{Pr}\{X_t \in A | X_0 = x\} > 0$ for any $x \in \mathcal{X}$.

Properties ii

- Aperiodicity
  Consider any $t \geqq 1$ satisfies $\mathbf{Pr}\{X_t \in A | X_0 = x\} > 0$ for any $x \in A \subseteqq \mathcal{X}$. The greatest common divisor of such t is called the period. If the period is one for any $A \subseteqq \mathcal{X}$, a Markov chain $\{X_t\}_{t=0}^{\infty}$ is aperiodic.
- Recurrence
  Define $\tau_A = \inf\{t > 0 : X_t \in A\}$. A is recurrent if $\mathbf{Pr}\{\tau_A < \infty | X_0 = x\} = 1$ for any $x \in A$. $\{X_t\}_{t=0}^{\infty}$ is recurrent with respect to f if A is recurrent for a f-regular Markov chain $\{X_t\}_{t=0}^{\infty}$ and $\int_A f(x)dx > 0$.

Transition Kernel i

The conditional p.d.f. of X_t given $X_0, \ldots, X_{t-1}$ is

$$f_t(x_t|x_0, \ldots x_{t-1}) = f(x_t|x_{t-1}), \tag{2}$$

which is called the transition kernel, and the right-hand side of (2) is often expresses as $K(x_{t-1}, x_t)$.

The joint p.d.f. of $\{X_s\}_{s=0}^t$ is

$$\begin{aligned} f(x_0, \ldots, x_t) &= f_0(x_0) f_t(x_1|x_0) f_2(x_2|x_0, x_1) \times \cdots \\ &\quad \times f_t(x_t|x_0, \ldots, x_{t-1}) \\ &= f_0(x_0) \prod_{s=1}^{t} K(x_{s-1}, x_s). \end{aligned} \tag{3}$$

Transition Kernel ii

Note that

$$
\begin{aligned}
f_t(x_t) &= \int_{\mathcal{X}} f(x_{t-1}, x_t) dx_{t-1} \\
&= \int_{\mathcal{X}} f(x_t | x_{t-1}) f_{t-1}(x_{t-1}) dx_{t-1} \\
&= \int_{\mathcal{X}} f_{t-1}(x_{t-1}) K(x_{t-1}, x_t) dx_{t-1}. \qquad (4)
\end{aligned}
$$

Define

$$
f_t = f_{t-1} \circ K = \int_{\mathcal{X}} f_{t-1}(x_{t-1}) K(x_{t-1}, x_t) dx_{t-1}. \qquad (5)
$$

Transition Kernel iii

Then

$$
\begin{aligned}
f_t &= f_{t-1} \circ K = \left\{ \int_{\mathcal{X}} f_{t-2}(x_{t-2}) K(x_{t-2}, x_{t-1}) dx_{t-2} \right\} \circ K \\
&= \int_{\mathcal{X}} \left\{ \int_{\mathcal{X}} f_{t-2}(x_{t-2}) K(x_{t-2}, x_{t-1}) dx_{t-2} \right\} K(x_{t-1}, x_t) dx_{t-1} \\
&= \int_{\mathcal{X}} f_{t-2}(x_{t-2}) \underbrace{\int_{\mathcal{X}} K(x_{t-2}, x_{t-1}) K(x_{t-1}, x_t) dx_{t-1}}_{K \circ K} dx_{t-2} \\
&= f_{t-2} \circ (K \circ K) = f_{t-2} \circ K^2 \quad \cdots \quad = f_0 \circ K^t, \qquad (6) \\
K^t &= \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} K(x_0, x_1) \cdots K(x_{t-1}, x_t) dx_1 \ldots dx_{t-1}.
\end{aligned}
$$

Invariant Distribution

The invariant distribution (density) $\bar{f}$ of a Markov chain with kernel K is

$$\bar{f}(\tilde{x}) = \int_{\mathcal{X}} \bar{f}(x) K(x, \tilde{x}) dx \quad \text{or} \quad \bar{f} = \bar{f} \circ K. \tag{7}$$

If a Markov chain is recurrent and aperiodic,

$$\text{Ergodicity:} \quad \lim_{t \to \infty} \sup_{A \subseteq \mathcal{X}} \left| \int_A (f_t(x) - \bar{f}(x)) dx \right| = 0, \tag{8}$$

$$\text{LLN:} \quad \frac{1}{T} \sum_{t=1}^{T} h(X_t) \xrightarrow{\text{a.s.}} \int_{\mathcal{X}} h(x) \bar{f}(x) dx. \tag{9}$$

Detailed balance condition

$$\bar{f}(x) K(x, \tilde{x}) = \bar{f}(\tilde{x}) K(\tilde{x}, x), \quad {}^{\forall} x, \tilde{x} \in \mathcal{X}. \tag{10}$$

Markov Chain Monte Carlo

Markov chain sampling

Step 1. Set $t = 1$ and $\tilde{x}_0 \leftarrow f_0(x_0)$.

Step 2. $\tilde{x}_t \leftarrow K(\tilde{x}_{t-1}, x_t)$.

Step 3. Increase t by 1 and go to **Step 2**.

- Suppose we can generate $\tilde{x}_t$ from a recurrent and aperiodic Markov chain with K and $\bar{f}$.
- After we repeat Step 1–3 sufficiently many times (say M), the distribution of $\tilde{x}_t$ will be very close to $\bar{f}$.
- $\mathbb{E}[h(X)] = \int_{\mathcal{X}} h(x)\bar{f}(x)dx$ can be approximated by $\frac{1}{N}\sum_{t=M+1}^{M+N} h(\tilde{x}_t)$ due to ergodicity of the Markov chain.
- The first M runs of the Markov chain sampling is called the burnin.