On Markov processes

Yukio NAGAHATA
Niigata univ.
nagahata@eng.niigata-u.ac.jp

2016 January 4-8

Markov processes

Suppose that $(X_n)_{n\geq 0}$ is a Markov process. If $(X_n)_{n\geq 0}$ satisfies

$$P(X_{n+1} = y | X_n = x) = P(X_1 = y | X_0 = x), \quad \forall n$$

then we call time homogeneous Markov process. In this lecture we only consider time homogeneous Markov process.
Let $(X_n)_{n\geq 0}$ be a time homogeneous Markov process. We set a (huge size) square matrix

$$P_{x,y} := P(X_1 = y | X_0 = x)$$

Then we call $P = (P_{x,y})$ transition matrix or transition probability matrix.

matrix formulation?

We only consider $P(X_1 = y | X_0 = x)$. Is it possible to formulate the conditional probability $P(X_{n+m} = y | X_m = x)$ by means of $P_{x,y}$?

Markov processes

We have following identities;

$$
\begin{aligned}
&P(X_{n+m} = y | X_m = x) \\
&= P(X_{n+m} = y, \bigcup_{y_{n+m-1}} X_{n+m-1} = y_{n+m-1} | X_m = x) \\
&= \sum_{y_{n+m-1}} P(X_{n+m} = y, X_{n+m-1} = y_{n+m-1} | X_m = x) \\
&= \sum_{y_{n+m-1}} P(X_{n+m} = y | X_{n+m-1} = y_{n+m-1}) \\
&\qquad \times P(X_{n+m-1} = y_{n+m-1} | X_m = x) \\
&= \sum_{y_{n+m-1}} P(X_{n+m-1} = y_{n+m-1} | X_m = x) P_{y_{n+m-1}, y}
\end{aligned}
$$

Markov processes

Inductively, we have

$$
\begin{aligned}
&P(X_{n+m} = y | X_m = x) \\
&= \sum_{y_{n+m-1}} P(X_{n+m-1} = y_{n+m-1} | X_m = x) P_{y_{n+m-1}, y} \\
&\vdots \\
&= \sum_{y_{m+1}} \sum_{y_{m+2}} \cdots \sum_{y_{n+m-1}} P_{x, y_{m+1}} P_{y_{m+1}, y_{m+2}} \cdots P_{y_{n+m-1}, y} \\
&= (P^n)_{x,y}
\end{aligned}
$$

Namely the conditional probability $P(X_{n+m} = y | X_m = x)$ is given by n-th power of P.

Markov processes

Definition (a random mapping representation)

A random mapping representation of transition matrix P is a function $f : S \times \Lambda \to S$ for some Λ and a random variable Z with state space Λ and probability law Q such that it satisfies

$$Q(\{f(x, Z) = y\}) = P_{x,y}$$

Proposition (a random mapping representation)

Suppose that P be a transition matrix of some time homogeneous Markov process with finite state space S. Then we have a random mapping representation f and Z.

Markov processes

This Proposition is one of the most important Proposition, when we write a program of some computer simulation. In particular, we write a program along the proof of this Proposition.
It is natural and applicable for us to set $\Lambda = [0, 1]$ and Z an uniform random variable on $[0, 1]$.
We can extend this Proposition such that we exchange a finite state space into countable state space (or much wide class). On the other hands, if we consider the application to the computer simulation, then a finite state space condition is reasonable.

Proof

Since S is finite set, without loss of generality, we set $S = \{s_1, s_2, \ldots, s_n\}$. Let $\Lambda = [0,1]$ and Z be an uniform random variable on $\Lambda = [0,1]$.

We set $\{F_{j,k}; 1 \le j \le n, 0 \le k \le n\}$ by $F_{j,0} = 0$, $F_{j,k} = \sum_{i=1}^{k} P_{s_j,s_i}$. Furthermore we set $f : S \times \Lambda \to S$ by $f(s_j, z) := s_k$ if $F_{j,k-1} < z \le F_{j,k}$. Here we have $f(s,0) = s_1$.

Then we can verify

$$Q(\{f(x,Z) = y\}) = P_{x,y}$$

$\square$

Markov processes

Since the matrix P is given by the transition probability, i.e., $P_{x,y} = P(X_1 = y | X_0 = x)$, we have

$$0 \leq P_{x,y} \leq 1, \quad \forall x, y$$
$$\sum_y P_{x,y} = 1, \quad \forall x$$

Conversely, if a matrix P satisfies these two identities, then there is a corresponding Markov process.

Markov processes

We have already seen that the conditional probability $P(X_{n+m} = y | X_m = x)$ is given by the n-th power of a matrix P. Some computation, in particular, eigenvalue, eigenvector and diagonalization is one of the important subject.
Suppose that we can diagonalize the transition matrix P as

$$Q^{-1}PQ = \Lambda, \qquad P = Q\Lambda Q^{-1}$$
$$\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

If the all eigenvalues λ_i are real value, then it is easy to treat. One of the famous sufficient condition is that P is symmetric, i.e., $P_{x,y} = P_{y,x}$ for all x, y.
We have an extension of this result as follows;

Definition (π-symmetric)

An $n \times n$ matrix P is symmetric with respect to $\pi = (\pi_1, \pi_2, \ldots, \pi_n) > 0$, if and only if for all x, y it satisfies

$$\pi_x P_{x,y} = \pi_y P_{y,x}$$

Let Π and $\sqrt{\Pi}$ be diagonal matrices defined via π by

$$\Pi = \begin{pmatrix} \pi_1 & & \\ & \ddots & \\ & & \pi_n \end{pmatrix}, \quad \sqrt{\Pi} = \begin{pmatrix} \sqrt{\pi_1} & & \\ & \ddots & \\ & & \sqrt{\pi_n} \end{pmatrix}$$

and we define a matrix S by

$$S = \sqrt{\Pi} P \sqrt{\Pi}^{-1},$$

Markov processes

then we can have

$$S = \sqrt{\Pi} P \sqrt{\Pi}^{-1} = \sqrt{\Pi}^{-1} (\Pi P) \sqrt{\Pi}^{-1}$$

Problem

Prove that if P is symmetric with respect to π, then the matrix $\sqrt{\Pi}^{-1} (\Pi P) \sqrt{\Pi}^{-1}$ is symmetric (in usual sense).

Since S is symmetric, we can diagonalize it by using real eigenvalues by $S = R \Lambda R^{-1}$. Then we have

$$P = \sqrt{\Pi}^{-1} S \sqrt{\Pi} = (\sqrt{\Pi}^{-1} R) \Lambda (R^{-1} \sqrt{\Pi}) = (\sqrt{\Pi}^{-1} R) \Lambda (\sqrt{\Pi}^{-1} R)^{-1}.$$

Hence the matrix P is diagonalizable and the eigenvalues of P and S coincides.

Conversely, there is a matrix P such that the all eigenvalues are real but for any π, P is not symmetric with respect to π.
Example We set

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}$$

Then the eigenvalues of P is $1, 0, -1/2$.
It is easy to see that if you can find a pair x, y such that $P_{x,y} \neq 0$ and $P_{y,x} = 0$ then for any π, P is not symmetric with respect to π.
We set

$$Q = \frac{1}{12} \begin{pmatrix} 2 & 8 & 2 \\ 5 & 2 & 5 \\ 5 & 2 & 5 \end{pmatrix}$$

then the eigenvalues of Q is $1, 0, -1/4$. But for any π, Q is not symmetric with respect to π. (We shall prove it later.)

Markov processes

equivalence class, $\leftrightarrow$、irreducible

Definition ($\rightarrow$、$\leftrightarrow$)

Let P be a transition matrix, then we define relations $\rightarrow$、$\leftrightarrow$ by ;
(1) $x \rightarrow x$.
(2) if $\exists n$, $x = x_0, x_1, x_2, \ldots, x_n = y$ s.t. $P_{x_{i-1},x_i} > 0, \quad 1 \leq \forall i \leq n$
then $x \rightarrow y$
(3) $x \rightarrow y$ and $y \rightarrow x$ then $x \leftrightarrow y$.

Problem

Show that the condition (2) is equivalent to $\exists n$ s.t.
$P(X_n = y | X_0 = x) > 0$.

Markov processes

Lemma (equivalence relation $\leftrightarrow$)

The relation $\leftrightarrow$ is an equivalence relation, namely it satisfies
(1) $x \leftrightarrow x$
(2) if $x \leftrightarrow y$ then $y \leftrightarrow x$
(3) if $x \leftrightarrow y$ and $y \leftrightarrow z$ then $x \leftrightarrow z$

Problem

It seems obvious, but prove this lemma.

Definition(irreducible)

If it satisfies $\forall x, y (\in S)$ we have $x \leftrightarrow y$then we call that the state space S is irreducible or the (transition) matrix P is irreducible.

If S is not irreducible, then we can decompose S into several classes by means of the equivalence relation $\leftrightarrow$ and we only consider following subject for each classes. Hence we only consider the irreducible case.

Markov processes

period

Definition (period)

For each $x \in S$ we define $N(x) := \{n; (P^n)_{x,x} > 0\}$ and $d_x :=$ the greatest common divisor of $N(x)$. We call d_x period.

Proposition (period)

If $x \leftrightarrow y$, then we have $d_x = d_y$

Proof

Since $x \leftrightarrow y$, we have $x \to y$ and $y \to x$, namely

$$\exists n_1, n_2 \text{ s.t. } P(X_{n_1} = y | X_0 = x) > 0,\ P(X_{n_2} = x | X_0 = y) > 0$$

Suppose that $n \in N(x)$ (note that we allow $n = 0$) then

$$P(X_n = x | X_0 = x) > 0.$$

Hence we have

$$\begin{aligned}
&P(X_{n_1+n+n_2} = y | X_0 = y) \\
&\geq P(X_{n_1+n+n_2} = y, X_{n+n_2} = x, X_{n_2} = x | X_0 = y) \\
&= P(X_{n_1} = x | X_0 = y) P(X_{n_2} = x | X_0 = y) P(X_{n_2} = x | X_0 = y) > 0
\end{aligned}$$

namely, we have $n_1 + n + n_2 \in N(y)$.

Since we allow $n = 0$, we also have $n_1 + n_2 \in N(y)$. By the definition of d_y, we have

$$\exists k \text{ s.t. } n_1 + n_2 = kd_y$$

Similarly for any $n \in N(x)$, we have $n_1 + n + n_2 \in N(y)$ and

$$\exists l \text{ s.t. } n_1 + n + n_2 = ld_y$$

Hence we have

$$n = (l - k)d_y.$$

Namely we have $d_x \geq d_y$.
In this argument, we can exchange the role of x and y. Hence we also have $d_y \geq d_x$. $\square$

Lemma

$\exists n_x$ s.t. $\forall k \geq n_x$, $kd_x \in N(x)$

Problem

Prove this lemma for $d_x = 1$.

Suppose that $d_x = 1$ and $\#S = n$. Then it seems that $n_x \leq n$.
But if we set

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

then $n_3 = 5 > 4 = n$.

Corollary

Suppose that S is irreducible. Then the period d_x does not depend on $x \in S$. Hence we set this period d. We may need to rearrange the order in S and rewrite the transition matrix P, but we have following;
$\exists n$ s.t.

$$P^n = \begin{pmatrix} P_1 & & & \\ & P_2 & & \\ & & \ddots & \\ & & & P_d \end{pmatrix}$$

namely, we can decompose P^n into d block matrix. Furthermore for all $1 \leq i \leq d$, $P_i > 0$ (all entries of P_i is positive). Furthermore for all $1 \leq i \leq d$, P_i are transition matrices.

Markov processes

Note that the matrix sizes of P_i in this corollary may not be the same, for example following is the case such that $d = 2$ and $n = 2$.

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{pmatrix}, \quad P^2 = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

参考文献

少なくとも以下の教科書などは参考にしている。但し、ずいぶん前に読んで勉強した教科書の内容をそれと認識せずに参考にしていることもありえるのでその点に関してはご容赦頂きたい。
Markov 過程の基本的なところでは [K] の教科書の該当部分を多く参考にしている。
ランダムナイトに関しては [AF] の教科書からアイデアを得ているが、証明は [K] の教科書に沿っている。
Perron-Frobenius の定理は [V] の教科書の証明に沿っているが、確率論的には [Sai][Sal][Se] の方がよいかもしれない。なお [Sai] は非負の行列でなく正の行列で証明をしている。
スペクトルギャップに関しては [Sal] の結果を中心に、粒子系（格子気体）への応用を述べた。粒子系への応用に関して明示的に書かれたもの（教科書、論文）はないと思う。

参考文献

[AF] Aldous D.; Fill J.A.
Reversible Markov Chains and Random Walks on Graphs,
(未刊の教科書 現在 (2015/11) も 「Aldous Fill Markov chain」で検索をすれば pdf、 html バージョンを発見できる)
[K] 小谷真一, 測度と確率, 岩波書店
[LPW] Levin D.A.; Peres Y,; Wilmer E.L.
Markov Chains and Mixing Times, AMS
[Sai] 斎藤正彦, 線形代数入門, 東京大学出版会
[Sal] Saloff-Coste L.
Lectures on Finite Markov Chains, Lecture on Probability 1665
[Se] Seneta E.
Non-negative Matrices and Markov Chains, Springer
[V] Varga R.S.
Matrix Iterative Analysis, Springer