STA447/STA2006 Stochastic Processes

Gun Ho Jang

Lecture on January 16, 2014

Note

This lecture note is prepared for the course STA447/STA2006 Stochastic Processes. This lecture note may contain flaws. Please consult text book or reference books for confidence.

- Gun Ho Jang

* indicates graduate level. So you may skip those parts.

2 Markov Chain

2.1 Stochastic Process

Definition 1. A *stochastic process* is a collection of time indexed random variables, that is, $\{X_t : t \in \mathcal{T}\}$.

Definition* 2. A sequence of σ-fields $\mathcal{F}_t$ is called a *filtration* if it is increasing, that is, $\mathcal{F}_s \subset \mathcal{F}_t$ if and only if $s \le t \in \mathcal{T}$. A stochastic process X_t is said to be *adapted to* $\mathcal{F}_t$ if $X_t \in \mathcal{F}_t$ for all t.

Example 1. Brownian motion, Markov chain, renewal process, queuing theory, martingale, Poisson process, jump process, ARMA models, linear processes.

2.2 Markov Chain

Definition 3. A stochastic process X_t is called a *Markov chain* if $P(X_{t+1} \in \cdot \mid X_1, \ldots, X_t) = P(X_{t+1} \in \cdot \mid X_t)$. The initial distribution of a Markov chain X is the distribution of X_0. The *transition probability* is defined by $p_t(i,j) = P(X_t = j \mid X_{t-1} = i)$. A Markov chain is said to be *homogeneous* when the transition probability does not depend on time.

Example 2 (Weather chain)**.** Let X_t be the weather on day t which having values 1 for *rainy* or 2 for *sunny*. Assume that the weather is a (homogeneous) Markov chain having transition probability

	1	**2**
1	0.6	0.4
2	0.2	0.8

Day 0: sunny $\Longrightarrow$ Day 1: sunny(.8), rainy(.2) $\Longrightarrow$ Day 2: sunny(.72), rainy(.28) $\Longrightarrow$ Day 3: sunny(.688), rainy(.312) $\Longrightarrow \cdots \Longrightarrow$ sunny(.667), rainy(.333)

Example 3 (Galton-Watson process)**.** Consider a specie in which each individual lives only one generation and gives birth Y children where Y follows a distribution F having non-negative integers values, $\mathbb{N}_+ = \{0, 1, 2, \ldots\}$. In other words $P(Y = j) = p_j \ge 0$ for $j = 0, 1, \ldots$. Let X_t be the number of individual at time t. Then X_t is a (homogeneous) Markov chain having transition probability

$$p(i,j) = P(X_{t+1} = j \mid X_t = i) = P(Y_1 + \cdots + Y_i = j)$$

where $Y_1, \ldots, Y_i$ are i.i.d copies of Y for $i > 0$ and $p(0,0) = 1$.

Note. If the number of individuals become zero, then it stays forever. Such kind of state is called a *absorbing state*.

Example 4 (Weather example continue)**.** Tomorrow's weather is predictable using the transition probability $p(i,j)$. Using the same transition probability, the weather of two days after is also predictable. When today's weather is sunny, the probability that two days after is sunny is sum of two paths (sunny $->$ sunny $->$ sunny and sunny $->$ rainy $->$ sunny), that is, $p(2,2)p(2,2)+p(2,1)p(1,2) = 0.8^2+0.2\cdot 0.4 = 0.64+0.08 = 0.72$.

Theorem 1 (Chapman-Kolmogorov equation)**.** Let p be the transition matrix of a homogeneous Markov chain X_t. The $(m+n)$-step transition matrix is the multiple of m-step transition matrix and n-step transition matrix, that is, $p^{(m+n)}(i,j) = \sum_k p^{(m)}(i,k)p^{(n)}(k,j)$ where $p^{(k)}(i,j) = P(X_{t+k} = j \mid X_t = i)$.

Proof. Without loss of generality, $t = 0$ can be assumed. All possible value of the Markov chain X_m at time m is a subset of all state space $\mathcal{S}$. Hence, we get

$$\begin{aligned} p^{(m+n)}(i,j) = P(X_{m+n} = j \mid X_0 = i) &= \sum_{k\in\mathcal{S}} P(X_{m+n} = j, X_m = k \mid X_0 = i) \\ &= \sum_{k\in\mathcal{S}} P(X_m = k \mid X_0 = i) P(X_{m+n} = j \mid X_m = k, X_0 = i) \\ &= \sum_{k\in\mathcal{S}} P(X_m = k \mid X_0 = i) P(X_{m+n} = j \mid X_m = k) = \sum_{k\in\mathcal{S}} p^{(m)}(i,k)p^{(n)}(k,j). \end{aligned}$$

Hence $p^{(2)} = p^{(1)} \times p^{(1)} = p \times p = p^2$. In general, $p^{(m)} = p^m$. ☐

Example 5 (Weather example contiued)**.** Note that $p^{(m)} = p^m$.

$$p^{(2)} = \begin{array}{ccc} & \mathbf{1} & \mathbf{2} \\ \mathbf{1} & 0.44 & 0.56 \\ \mathbf{2} & 0.28 & 0.72 \end{array} \quad p^{(3)} = \begin{array}{ccc} & \mathbf{1} & \mathbf{2} \\ \mathbf{1} & 0.376 & 0.624 \\ \mathbf{2} & 0.312 & 0.688 \end{array} \quad p^{(4)} = \begin{array}{ccc} & \mathbf{1} & \mathbf{2} \\ \mathbf{1} & 0.350 & 0.650 \\ \mathbf{2} & 0.325 & 0.675 \end{array} \quad p^{(\infty)} = \begin{array}{ccc} & \mathbf{1} & \mathbf{2} \\ \mathbf{1} & 1/3 & 2/3 \\ \mathbf{2} & 1/3 & 2/3 \end{array}$$

Note. Let $P_x(A) = P(A \mid X_0 = x)$ for convenience and $\mathbb{E}_x$ be the corresponding expectation.

Definition 4. Let $T_y = \min\{t \geq 1 : X_t = y\}$ be the *time of the first return to* y and $\rho_{xy} = P_x(T_y < \infty)$. Let $T_y^1 = T_y$ and $T_y^k = \min\{n > T_y^{k-1} : X_n = y\}$ be the time of the k-th return to y.

By definition, ρ_{yy} is the probability X_t return to y after starting at y.

Definition 5. A time valued random variable T is said to be a *stopping time* if the event $\{T \leq t\}$ can be expressed by $X_0, \ldots, X_t$.

Exercise 1. Show that a time valued random variable T is stopping time if and only if $\{T = t\}$ can be expressed by $X_0, \ldots, X_t$.

Example 6. The first returning time T_y is a stopping time because

$$\{T_y = n\} = \{X_1 \neq y, \ldots, X_{n-1} \neq y, X_n = y\}.$$

Similarly, T_y^k are stopping times.

Theorem 2 (Strong Markov property)**.** Suppose T is a stopping time. Given $T < \infty$ and $X_T = y$, prediction of the future does not rely on $\{X_t : t < T\}$ and the stochastic process $\{X_{T+t}, t \geq 0\}$ behaves like the Markov chain with initial state y.

Proof. Let $k \geq 1$ and $x_1, \ldots, x_k \in \mathcal{S}$.

$$\begin{aligned}
P(X_{T+1} = x_1, \ldots, X_{T+k} = x_k, X_T = y) &= \sum_{n=0}^{\infty} P(X_{T+1} = x_1, \ldots, X_{T+k} = x_k, X_T = y, T = n) \\
&= \sum_{n=0}^{\infty} P(X_{n+1} = x_1, \ldots, X_{n+k} = x_k, X_n = y, T = n) \\
&= \sum_{n=0}^{\infty} P(X_{n+1} = x_1, \ldots, X_{n+k} = x_k \,|\, X_n = y, T = n) P(X_n = y, T = n) \\
&= \sum_{n=0}^{\infty} P(X_{n+1} = x_1, \ldots, X_{n+k} = x_k \,|\, X_n = y) P(X_T = y, T = n) \\
&= \sum_{n=0}^{\infty} P_y(X_1 = x_1, \ldots, X_k = x_k) P(X_T = y, T = n) \\
&= P_y(X_1 = x_1, \ldots, X_k = x_k) P(X_T = y)
\end{aligned}$$

Hence $P(X_{T+1} = x_1, \ldots, X_{T+k} = x_k \,|\, X_T = y) = P_y(X_1 = x_1, \ldots, X_k = x_k)$. □

Example 7. The Markov chain after the first return to y is given by $X_{T_y^1+t}$ which behave a Markov chain starts at y. Both $T_y^3 - T_y^2$ and $T_y^2 - T_y^1$ have the same distribution. Hence $P_y(T_y^k < \infty) = \rho_{yy}^k$. Furthermore $P_x(T_y^k < \infty) = P_x(T_y < \infty) P_y(T_y^{k-1} < \infty) = \rho_{xy}\rho_{yy}^{k-1}$.

Definition 6. A state y is said to be *recurrent* if $\rho_{yy} = 1$ or *transient* otherwise.

Example 8. Consider a Markov chain having transition matrix

$$p = \begin{array}{c|ccc} & \mathbf{1} & \mathbf{2} & \mathbf{3} \\ \hline \mathbf{1} & 1 & 0 & 0 \\ \mathbf{2} & 0.2 & 0.7 & 0.1 \\ \mathbf{3} & 0.3 & 0.1 & 0.6 \end{array}$$

Note that $P_1(X_1 = 1) = 1$ implies $P_(T_1 = 1) = 1$. Hence 1 is recurrent. While $P_2(T_2 = \infty) \geq P_2(X_1 = 1) = p(2,1) = 0.2 > 0$ and $P_3(T_3 = \infty) \geq P_3(X_1 = 1) = p(3,1) = 0.3 > 0$ imply states 2 and 3 are transient.

Let N_y be the number of visits to y, that is, $N_y = \sum_{n=1}^{\infty} 1(X_n = y)$.

Proposition 3. (a) $\mathbb{E}_x N_y = \rho_{xy}/(1 - \rho_{yy})$. (b) $\mathbb{E}_x N_y = \sum_{n=1}^{\infty} p^{(n)}(x,y)$.

Proof. (a) Obvious if $\rho_{xy} = 0$. If $\rho_{xy} > 0$ and $\rho_{yy} = 1$, then the Markov chain visits y infinitely many times with probability at least $\rho_{xy} > 0$. Hence the expectation is infinity. For $\rho_{yy} < 1$, $\mathbb{E}_x N_y = \sum_{k=1}^{\infty} k P_x(T_y^k < \infty, T_y^{k+1} = \infty) = \sum_{k=1}^{\infty} k(P_x(T_y^k < \infty) - P_x(T_y^{k+1} < \infty)) = \sum_{k=1}^{\infty} k(\rho_{xy}\rho_{yy}^{k-1} - \rho_{xy}\rho_{yy}^k) = \sum_{k=1}^{\infty} \rho_{xy}\rho_{yy}^{k-1} = \rho_{xy}/(1 - \rho_{yy})$.
(b) $\mathbb{E}_x N_y = \mathbb{E}_x \sum_{n=1}^{\infty} 1(X_n = y) = \sum_{n=1}^{\infty} \mathbb{E}_x 1(X_n = y) = \sum_{n=1}^{\infty} p^{(n)}(x,y)$. □

Theorem 4. A state x is recurrent if and only if $\mathbb{E}_x N_x = \infty$.

Proof. From Proposition 3 (a), $\mathbb{E}_x N_x = \infty$ if and only if $\rho_{xx} = 1$ if and only if x is recurrent. □

Definition 7. State x *communicates with* state y and write $x \to y$ if y is reachable after starting at x with positive probability, that is, $\rho_{xy} = P_x(T_y < \infty) > 0$.

Example 9. In Example 8, $3 \to 1$, $2 \to 1$ but $1 \not\to 2$, $1 \not\to 3$. Also $2 \to 3$ and $3 \to 2$ (write $2 \leftrightarrow 3$).

Proposition 5. If $x \to y$ and $y \to z$, then $x \to z$.

Proof. Short proof: $\rho_{xz} = P_x(T_z < \infty) \geq P_x(T_y < \infty)P_y(T_z < \infty) = \rho_{xy}\rho_{yz} > 0$.
Long proof: There exists m, n such that $p^{(m)}(x,y) > 0$ and $p^{(n)}(y,z) > 0$. Then $p^{(m+n)}(x,z) \geq p^{(m)}(x,y)p^{(n)}(y,z) > 0$. □

Proposition 6. If $\rho_{xy} > 0$ and $\rho_{yx} < 1$, then x is transient.

Proof. There exists a finite positive integer m such that $p^{(m)}(x,y) > 0$. Then,

$$P_x(T_x = \infty) \geq P_x(T_x = \infty, X_m = y) = P_x(X_m = y)P_y(T_x = \infty) = p^{(m)}(x,y)(1 - \rho_{yx}) > 0.$$

Hence x is transient. □

Proposition 7. If x is recurrent and $\rho_{xy} > 0$, then $\rho_{yx} = 1$ and y is also recurrent.

Proof. If $\rho_{yx} < 1$, then x must be transient by Proposition 6. Since $\rho_{xy}, \rho_{yx} > 0$, there exist $n, m > 0$ such that $p^{(m)}(x,y) > 0$ and $p^{(n)}(y,x) > 0$. Then $\mathbb{E}_y N_y = \sum_{l=1}^{\infty} p^{(l)}(y,y) \geq \sum_{l=m+n+1}^{\infty} p^{(l)}(y,y) \geq \sum_{l=1}^{\infty} p^{(n)}(y,x)p^{(l)}(x,x)p^{(m)}(x,y) = p^{(n)}(y,x)\sum_{l=1}^{\infty} p^{(l)}(x,x)p^{(m)}(x,y) = \infty$. Hence y is recurrent. □

Definition 8. A set C is said to be *closed* if it is impossible to get out, that is, $P_x(X_1 \notin C) = 0$ for all $x \in C$. A set I is said to be *irreducible* if all states in I communicate each other, that is, $x \to y$ for any $x, y \in I$.

Example 10 (Seven-state chain)**.** . Consider a Markov chain having the transition probability of the form

	1	**2**	**3**	**4**	**5**	**6**	**7**
1	0.7	0	0	0	0.3	0	0
2	0.1	0.2	0.3	0.4	0	0	0
3	0	0	0.5	0.3	0.2	0	0
4	0	0	0	0.5	0	0.5	0
5	0.6	0	0	0	0.4	0	0
6	0	0	0	0	0	0.2	0.8
7	0	0	0	1	0	0	0

Then $\{1,5\}$ and $\{4,6,7\}$ are closed and irreducible.

Proposition 8. In a finite closed set C, there exists at least one recurrent state.

Proof. Suppose all states in C are transient. Then $\mathbb{E}_x N_y < \infty$ for all $x, y \in C$. Since $|C| < \infty$, we get

$$\infty > \sum_{y \in C} \mathbb{E}_x N_y = \sum_{y \in C}\sum_{n=1}^{\infty} p^{(n)}(x,y) = \sum_{n=1}^{\infty}\sum_{y \in C} p^{(n)}(x,y) = \sum_{n=1}^{\infty} 1 = \infty.$$

This contradiction leads to the existence of a recurrent state. □

Theorem 9. If a set C is finite, closed, and irreducible, then all states in C are recurrent.

Example 11 (Seven-state chain)**.** States 1,5, 4,6,7 are recurrent.

Theorem 10. If the state space S is finite, then $S = T \cup R_1 \cup \cdots \cup R_k$ for some disjoint sets $T, R_1, \ldots, R_k$ where T is a set of transient states and R_i's are closed and irreducible sets of recurrent states.

Proof. Let T be the set of all transient states, that is, $x \in T$ if and only if there exists y such that $x \to y$ but $y \not\to x$. Let $C_x = \{y : x \to y\}$ for any $x \in S - T$. Since $x \in S - T$, $x \to y$ implies $y \to x$. If $y \in C_x$ and $y \to z$, then $x \to y \to z$ implies $x \to z$ and $z \in C_x$. Hence C_x is closed. If $y, z \in C_x$, then $x \to y$, $x \to z$ and $y \to x$, $z \to x$. Thus $y \to x \to z$ implies $y \to z$. Note that $C_x = C_y$ or $C_x \cap C_y = \emptyset$ for any $x, y \in S - T$. By letting $\{R_1, \ldots, R_k\} = \{C_x : x \in S - T\}$, the theorem holds. □

2.3 Stationary Distribution

Definition 9. A stochastic process X_t is said to be *stationary* if $\{X_t\}$ and $\{X_{t+s}\}$ have the same distribution for any $s \geq 0$.

A (homogeneous) Markov chain X_t can be stationary if X_0 and X_1 have the save distribution. If X_0 and X_1 have the same distribution, then all X_t have the same distribution. For any fixed s. Let $T = s$ be a stopping time. X_0 and X_T have the same distribution and strong Markov property shows $\{X_t\}$ and $\{X_{T+t}\}$ have the same distribution.

Definition 10. A distribution π is called a *stationary distribution* if $\pi p = \pi$ so that $X_0 \equiv^d X_1$.

Example 12 (Two state Markov chain)**.**

$$\begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix} \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix}$$

Solves $\pi_1 = b/(a+b), \pi_2 = a/(a+b)$.

Example 13 (Weather chain)**.** Applying two state Markov chain for

$$\begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix} \begin{pmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{pmatrix} = \begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix}$$

we get $\pi_1 = 0.2/(0.4+0.2) = 1/3$ and $\pi_1 = 0.4/(0.4+0.2) = 2/3$.

Theorem 11. If a $k \times k$ transition matrix p is irreducible, then there exists a unique solution to $\pi p = \pi$ with $\sum_x \pi_x = 1$ and $\pi_x > 0$ for all $x \in S$.

Proof. Since the rank of $p - I$ is at most $k-1$, there exists a solution ν satisfying to $\nu p = \nu$. Let $r = [(I+p)/2]^{k-1}$. Then $\nu(I+p)/2 = \nu$ implies $\nu r = \nu$. For any x, y, there exists $p^{(l)}(x,y) > 0$ with $l \leq k-1$. Thus $r(x,y) > 0$.

Suppose there are two different signs amongν_x. Then $|\nu_y| = |\sum_x \nu_x r(x,y)| < \sum_x |\nu_x| r(x,y)|$ and $\sum_y |\nu_y| < \sum_y \sum_x |\nu_x| r(x,y) = \sum_x |\nu_x|$. It contradicts. Thus $\nu_x \geq 0$ for all x. The fact $\nu_y = \sum_x \nu_x r(x,y)$ implies $\nu_x > 0$. If there exists another solutions w, we can make a new solution $w' = aw + b\nu$ so that $\sum_x w'_x \nu_x = 0$. But both w' and ν are positive. Therefore the solution is unique. □