STA447/STA2006 Stochastic Processes

Gun Ho Jang

Lecture on January 30, 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.

2.6 Doubly Stochastic

Definition 30. A transition matrix p is *doubly stochastic* if its columns sum to 1.

Theorem 47. If p is a doubly stochastic transition probability matrix with rank N, then the uniform distribution is a stationary distribution.

Proof. Let $\pi(x) = 1/N$ for all x. Then,

$$\sum_y \pi(y)p(y,x) = (1/N)\sum_y p(y,x) = 1/N = \pi(x).$$

Hence π is a stationary distribution. □

Example 34 (Symmetric reflecting random walk on the line)**.** The states are $\mathcal{S} = \{0, 1, \ldots, L\}$. The chain goes to the left or right with probability 1/2 under the restriction that the left of 0 is treated as 0 and the right of L is treated as L. For example, the transition probability for $L = 3$ is

	0	**1**	**2**	**3**
0	1/2	1/2	0	0
1	1/2	0	1/2	0
2	0	1/2	0	1/2
3	0	0	1/2	1/2

Since it is doubly stochastic, the uniform distribution $\pi(x) = 1/4$ is a stationary distribution. It is irreducible because $0 \to 1 \to 2 \to 3 \to 2 \to 1 \to 0$. Then π is the unique stationary distribution.

Example 35. Consider the following transition matrix

	1	**2**	**3**	**4**
1	0.2	0	0.8	0
2	0	0.7	0	0.3
3	0.4	0	0.6	0
4	0	0.3	0	0.7

Since is doubly stochastic, the uniform distribution $\pi(x) = 1/4$ is a stationary distribution. Define $\mu(1) = 1/9, \mu(2) = 1/3, \mu(3) = 2/9, \mu(4) = 1/3$, then μ is also a stationary distribution. There are multiple stationary distributions because transition matrix p is not irreducible, that is, $1 \to 3 \to 1$ and $2 \to 4 \to 2$ imply that $\{1, 3\}$ and $\{2, 4\}$ two distinct irreducible sets.

2.7 Detailed Balance Condition

Definition 31. A distribution π is said to satisfy *detailed balance condition* if

$$\pi(x)p(x,y) = \pi(y)p(y,x)$$

for all x, y.

By summing over x, we get

$$\sum_x \pi(x)p(x,y) = \sum_x \pi(y)p(y,x) = \pi(y)$$

which is the equation for the stationary distribution. Hence detailed balance condition is stronger than stationary distribution.

Example 36. Consider a doubly stochastic transition matrix

	1	**2**	**3**	**4**
1	0.5	0.5	0	0
2	0.3	0.2	0.5	0
3	0.1	0.1	0.1	0.7
4	0.1	0.2	0.4	0.3

It is irreducible because $1 \to 2 \to 3 \to 4 \to 1$. But $\pi(1)p(1,3) = (1/4)0 = 0 \neq 0.025 = (1/4)0.1 = \pi(3)p(3,1)$. Hence it does not satisfy detailed balance condition.

Example 37 (Birth and death chain)**.** Let $\mathcal{S} = \{l, l+1, \ldots, r\}$ be the state set. It is impossible to jump more than one, that is, $p(x,y) = 0$ if $|x-y| > 1$. The transition matrix p satisfies

$$p(x,x+1) = p_x \text{ for } x < r, p(x,x-1) = q_x \text{ for } x > l, p(x,x) = 1 - p_x - q_x \text{ for } l \leq x \leq r.$$

Note that $p_r = q_l = 0$. For $x < r$, the detailed balance condition for x and $x+1$ implies $\pi(x)p_x = \pi(x+1)q_{x+1}$. Hence,

$$\pi(x+1) = \frac{p_x}{q_{x+1}}\pi(x) = \frac{p_x p_{x-1}}{q_{x+1}q_x}\pi(x-1) = \cdots = \pi(l)\frac{p_l p_{l+1}\cdots p_x}{q_{l+1}q_{l+2}\cdots q_{x+1}}.$$

If $p_x = p_0 > 0$ for $x = l, \ldots, r-1$ and $q_x = q_0 > 0$ for $x = l+1, \ldots, r$, then $\pi(x) = \pi(l)(p_0/q_0)^{x-l}$ and $\pi(l) = (1-p_0/q_0)/(1-(p_0/q_0)^{r-l})$. Hence $\pi(x) = (p_0/q_0)^{x-l}(1-p_0/q_0)/(1-(p_0/q_0)^{r-l})$ satisfies the detailed balance condition.

2.8 Reversibility

Let X_n be a HMC with transition probability p having a stationary distribution π.

Theorem 48. Let $Y_m = X_{n-m}$ for $0 \leq m \leq n$. Then Y_m is a HMC with transition probability

$$\hat{p}(x,y) = P(Y_{m+1} = y \,|\, Y_m = x) = \frac{\pi(y)p(y,x)}{\pi(x)}.$$

Proof. For any m and states $x_0, \ldots, x_{m+1}$,

$$P(Y_{m+1} = x_{m+1} \,|\, Y_0 = x_0, \ldots, Y_m = x_m) = \frac{P(X_{n-m-1} = x_{m+1}, \ldots, X_n = x_0)}{P(X_{n-m} = x_m, \ldots, X_n = x_0)}$$

$$= \frac{P(X_{n-m-1} = x_{m+1}, X_{n-m} = x_m)P(X_{n-m+1} = x_{m-1}, \ldots, X_n = x_0 \,|\, X_{n-m} = x_m)}{P(X_{n-m} = x_m)P(X_{n-m+1} = x_{m-1}, \ldots, X_n = x_0 \,|\, X_{n-m} = x_m)}$$

$$= \frac{P(X_{n-m-1} = x_{m+1}, X_{n-m} = x_m)}{P(X_{n-m} = x_m)} = \frac{\pi(x_{m+1})p(x_{m+1}, x_m)}{\pi(x_m)}.$$

Hence Y_m is a HMC with transition probability $\hat{p}(x,y) = \pi(y)p(y,x)/\pi(x)$. ☐

If π satisfies the detailed balance condition, then

$$\hat{p}(x, y) = \pi(y)p(y, x)/\pi(x) = \pi(x)p(x, y)/\pi(x) = p(x, y).$$

Hence the transition probability of the reversed HMC is the same to the original HMC.

2.9 Metropolis-Hastings Algorithm

Numerical integration computes $\mathbb{E}_\pi f(X) = \int f(x)\pi(x)\, dx$ where π is a distribution. If a sequence of random numbers X_n can be generated from π, the problem can be solved numerically using strong law of large numbers. Assume random number generation using π is computationally very hard. Even in this case it is possible to generate random variable from a homogeneous Markov chain having π as the stationary distribution.

Let q be a *proposal distribution* for random number generation and r be an acceptance distribution having density $r(x, y) = \min(1, \pi(y)q(y, x)/(\pi(x)q(x, y)))$. Then, $p(x, y) = q(x, y)r(x, y)$ is a transition probability satisfying detailed balance condition, that is, when $\pi(y)q(y, x) > \pi(x)q(x, y)$,

$$\pi(x)p(x, y) = \pi(x)q(x, y) \times 1, \quad \pi(y)p(y, x) = \pi(y)q(y, x) \times \frac{\pi(x)q(x, y)}{\pi(y)q(y, x)} = \pi(x)p(x, y).$$

Besides π is the stationary distribution of p. Finally using Ergodic theorem,

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \to \mathbb{E}_\pi f(X).$$

2.10 Exit Distribution and Time

It is well known that $p^{(n)}(x, y) \to 0$ as $n \to \infty$ for any transient states x and y. If a HMC started from a transient state, it eventually absorbed in a state or a set of irreducible state.

Recall a state x is an *absorbing state* if $p(x, x) = 1$. If a state y communicate with x, then a HMC started from y visits x with positive probability then it stays forever in x. Obviously y is a transient state.

Example 38 (Two year college)**.** At a local 2 year college, 60% of freshmen become sophomores, 25% remain freshmen, and 15% drop out. Seventy percent of sophomores graduate and transfer to a 4 year college, twenty percent remain sophomores and ten percent drop out. What fraction of new students eventually graduate? What is an expected year for a new student to graduate if it happens?

The transition matrix is

	1	**2**	**G**	**D**
1	0.25	0.6	0	0.15
2	0	0.2	0.7	0.1
G	0	0	1	0
D	0	0	0	1

It is easy to see that $1 \to 2 \to G$, that is, $\rho_{1G}, \rho_{2G} > 0$. But $\rho_{G1} = \rho_{G2} = 0$. Hence $1, 2$ are transient while $p(G, G) = p(D, D) = 1$ show G, D are absorbing states.

Let $h_y(x)$ be the probability of state x being absorbed into y eventually. Then,

$$h_G(1) = 0.25h_G(1) + 0.6h_G(2), \quad h_G(2) = 0.2h_G(2) + 0.7.$$

Hence $h_G(2) = 0.7/0.8 = 7/8$ and $h_G(1) = h_G(2)0.6/0.75 = 0.7$.

Let $t_y(x)$ be the expected time of state x being absorbed into y.

$$t_G(1) = 0.25(1 + t_G(1)) + 0.6(1 + t_G(2)), \quad t_G(2) = 0.2(1 + t_G(2)) + 0.7 \cdot 1.$$

It solves $t_G(2) = 0.9/0.8 = 9/8$ and $t_G(1) = (0.85 + 0.6t_G(2))/0.75 = 61/30 = 2.0333$.

Similarly $h_D(1) = 0.3, h_D(2) = 0.125$ and $t_D(1) = 49/30, t_D(2) = 3/8$.

Proposition 49. Let $\mathcal{T}$ be the set of all transient states and z be an absorbing state. Define $h_y(x)$ is the probability of a transient state x absorbed into y. Define $t_y(x)$ is the expected time of a transient state x absorbed into y if it happens. Then, h_z solves $(I_{\mathcal{T}} - p_{\mathcal{T},\mathcal{T}})h_z = p_{\mathcal{T},z}$ and t_z solves $(I_{\mathcal{T}} - p_{\mathcal{T},\mathcal{T}})t_z = p_{\mathcal{T},\mathcal{T}}\mathbf{1}_{\mathcal{T}} + p_{\mathcal{T},z}$.

Proof. The equations to be solved for h_z are

$$h_z(x) = \sum_{y\in\mathcal{T}} p(x,y)h_z(y) + p(x,z)$$

for all $x \in \mathcal{T}$. Similarly t_z solves

$$t_z(x) = \sum_{y\in\mathcal{T}} p(x,y)(1 + t_z(y)) + p(x,z)$$

for all $x \in \mathcal{T}$. Hence the proposition holds. □

2.11 Hitting Times

Definition 32. For a subset $A \subset \mathcal{S}$ of state space, the hitting time H_A to A is defined by $H_A = \inf\{n \geq 0 : X_n \in A\}$.

Hitting times are similar to the first returning time. The difference is that hitting times take account the initial distribution X_0 while returning times do not.

Theorem 50. Suppose A, B are disjoint subset of the state space. If $P_x(\min(H_A, H_B) < \infty) > 0$ for all $x \notin A \cup B$, then $h(x) = P_x(H_A < H_B)$ satisfies $h(x) = 1$ for all $x \in A$, $h(x) = 0$ for all $x \in B$ and $h(x) = \sum_y p(x,y)h(y)$ for $x \notin A \cup B$. The expected hitting time $g(x) = \mathbb{E}_x(H_A \,|\, H_A < \infty)$ satisfies $g(x) = 0$ for all A and $g(x) = 1 + \sum_{y\in C_A} p(x,y)g(y)$ for $x \in C_A$ where $C_A = \{y \in \mathcal{S} : y \notin A, P_y(H_A < \infty) > 0\}$.

Proof. By the definition $h(x) = 1$ for all $x \in A$ and $h(x) = 0$ for all $x \in B$. For any $x \notin A \cup B$, $h(x) = P_x(H_A < H_B) = \sum_{y\in\mathcal{S}} P_x(X_1 = y, H_A < H_B) = \sum_{y\in A} P_x(X_1 = y) + \sum_{y\notin A\cup B} P_x(X_1 = y)P_y(H_A < H_B) = \sum_{y\in A} p(x,y)h(y) + \sum_{y\in B} p(x,y)h(y) + \sum_{y\notin A\cup B} p(x,y)h(y) = \sum_y p(x,y)h(y)$.

By the definition $g(x) = 0$ for all $a \in A$. If $x \in C_A$, then $g(x) = \mathbb{E}_x(H_A \,|\, H_A < \infty) = \sum_{n=1}^{\infty} P_x(H_A \geq n \,|\, H_A < \infty) = P_x(H_A \geq 1 \,|\, H_A < \infty) + \sum_{n=2}^{\infty}\sum_{y\in C_A} P_x(X_1 = y, H_A \geq n \,|\, H_A < \infty) = 1 + \sum_{n=2}^{\infty}\sum_{y\in C_A} P_x(X_1 = y)P_y(H_A \geq n-1 \,|\, H_A < \infty) = 1 + \sum_{y\in C_A} p(x,y)g(y)$. □

Example 39. Consider a HMC with transition probability

$$p = \begin{array}{c|ccccc} & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} \\ \mathbf{1} & 1 & 0 & 0 & 0 & 0 \\ \mathbf{2} & 0 & 2/3 & 0 & 1/3 & 0 \\ \mathbf{3} & 1/8 & 1/4 & 5/8 & 0 & 0 \\ \mathbf{4} & 0 & 1/6 & 0 & 5/6 & 0 \\ \mathbf{5} & 1/3 & 0 & 1/3 & 0 & 1/3 \end{array}$$

State 1 is absorbing because $p(1,1) = 1$. $I = \{2,4\}$ is irreducible by considering $2 \to 4 \to 2$. States 3 and 5 are transient because $\rho_{31} \geq p(3,1) = 1/8 > 0$, $\rho_{51} \geq p(5,1) = 1/3 > 0$ while $\rho_{13} = \rho_{15} = 0$. As $n \to \infty$, $p^{(n)}(1,1) = 1$, $p^{(n)}(2,2) \to (1/6)/(1/6 + 1/3) = 1/3$, $p^{(n)}(4,4) \to 2/3$.

$$\begin{aligned}\lim_{n\to\infty} \begin{pmatrix} p^{(n)}(3,1) \\ p^{(n)}(5,1) \end{pmatrix} &\to \left(I_2 - \begin{pmatrix} 5/8 & 0 \\ 1/3 & 1/3 \end{pmatrix}\right)^{-1} \begin{pmatrix} 1/8 \\ 1/3 \end{pmatrix} = \begin{pmatrix} 3/8 & 0 \\ -1/3 & 2/3 \end{pmatrix}^{-1} \begin{pmatrix} 1/8 \\ 1/3 \end{pmatrix} \\ &= \frac{1}{(3/8)(2/3) - (0)(-1/3)} \begin{pmatrix} 2/3 & 0 \\ 1/3 & 3/8 \end{pmatrix} \begin{pmatrix} 1/8 \\ 1/3 \end{pmatrix} = 4 \begin{pmatrix} 1/12 \\ 1/6 \end{pmatrix} = \begin{pmatrix} 1/3 \\ 2/3 \end{pmatrix}\end{aligned}$$

In sum, the limit becomes

$$\lim_{n\to\infty} p^{(n)}(x,y) = \begin{array}{c|ccccc} & \mathbf{1} & \mathbf{2} & \mathbf{3} & \mathbf{4} & \mathbf{5} \\ \hline \mathbf{1} & 1 & 0 & 0 & 0 & 0 \\ \mathbf{2} & 0 & 1/3 & 0 & 2/3 & 0 \\ \mathbf{3} & 1/3 & 2/9 & 0 & 4/9 & 0 \\ \mathbf{4} & 0 & 1/3 & 0 & 2/3 & 0 \\ \mathbf{5} & 2/3 & 1/9 & 0 & 2/9 & 0 \end{array}$$

Note that both $\pi_1 = (1/2, 1/6, 0, 1/3, 0)$ and $\pi_2 = (1/3, 2/9, 0, 4/9, 0)$ are stationary distributions.

The expected exit time are

$$\begin{pmatrix} \mathbb{E}_3 T_1 \\ \mathbb{E}_5 T_1 \end{pmatrix} = \left(I_2 - \begin{pmatrix} 5/8 & 0 \\ 1/3 & 1/3 \end{pmatrix} \right)^{-1} \begin{pmatrix} 5/8 + 0 + 1/8 \\ 1/3 + 1/3 + 1/3 \end{pmatrix} = 4 \begin{pmatrix} 2/3 & 0 \\ 1/3 & 3/8 \end{pmatrix} \begin{pmatrix} 3/4 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 5/2 \end{pmatrix}$$

Let $A = \{2, 4\}$ and $B = \{1\}$. Then $P_x(\min(H_A, H_B) < \infty) = 1$ for all x. Let $h(x)$ be the probability of $H_A < H_B$ with initial distribution $X_0 \equiv x$. Then $h(2) = h(4) = 1$, $h(1) = 0$ and

$$\begin{pmatrix} h(3) \\ h(5) \end{pmatrix} = \begin{pmatrix} 1/8 & 1/4 & 5/8 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 \end{pmatrix} \begin{pmatrix} \mathrm{h}(1) & \mathrm{h}(2) & \mathrm{h}(3) & \mathrm{h}(4) & \mathrm{h}(5) \end{pmatrix}^T = \begin{pmatrix} 1/4 + 5/8h(3) \\ 1/3h(3) + 1/3h(5) \end{pmatrix}$$

solves $h(3) = 2/3$ and $h(5) = h(3)/2 = 1/3$. Expected exit time to A from initial state x denoted by $h(x)$ satisfies $h(2) = h(4) = 0$, $h(1) = \infty$, and

$$\begin{pmatrix} g(3) \\ g(5) \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 5/8 & 0 \\ 1/3 & 1/3 \end{pmatrix} \begin{pmatrix} \mathrm{g}(3) & \mathrm{g}(5) \end{pmatrix}^T = \begin{pmatrix} 1 + (5/8)h(3) \\ 1 + (1/3)h(3) + (1/3)h(5) \end{pmatrix}$$

Which solves $g(3) = 8/3$ and $g(5) = (3/2)(1 + h(3)/3) = 3/2 + 4/3 = 17/6$.

2.12 Proof of Theorem 42 (b)

Define $q(x,y) = (\mu_z(y)/\mu_z(x))p(y,x)$.
Claim: q is a transition probability.

$$\sum_y q(x,y) = \frac{1}{\mu_z(x)} \sum_y \mu_z(y) p(y,x) = \frac{1}{\mu_z(x)} \mu_z(x) = 1.$$

Claim: $q^{(n)}(x,y) = (\mu_z(y)/\mu_z(x))p^{(n)}(y,x)$.
It is true for $n = 1$ by definition. Assume it is true for n. Then,

$$\begin{aligned} q^{(n+1)}(x,y) &= \sum_w q(x,w) q^{(n)}(w,y) = \sum_w \frac{\mu_z(w)}{\mu_z(x)} p(w,x) \frac{\mu_z(y)}{\mu_z(w)} p^{(n)}(y,w) \\ &= \frac{\mu_z(y)}{\mu_z(x)} \sum_w p^{(n)}(y,w) p(w,x) = \frac{\mu_z(y)}{\mu_z(x)} p^{(n+1)}(y,x). \end{aligned}$$

Claim: q is irreducible.
For any x and y, there exists $l > 0$ so that $p^{(l)}(y,x) > 0$. Then $q^{(l)}(x,y) = \frac{\mu_z(y)}{\mu_z(x)} p^{(l)}(y,x) > 0$.
Let Y_n be a HMC having q as the transition probability and $U_x = \inf\{n \geq 1 : Y_n = x\}$. Define $g(x,y,n) = \mu_z(y) Q_y(U_x = n)$. Then,

$$\begin{aligned} g(x,y,n+1) = \mu_z(y) Q_y(U_x = n+1) &= \mu_z(y) \sum_{w \neq x} Q_y(U_w = 1, U_x = n+1) \\ &= \mu_z(y) \sum_{w \neq x} Q_y(U_w = 1) Q(U_x = n+1 \,|\, Y_1 = w) = \mu_z(y) \sum_{w \neq x} q(y,w) Q_w(U_x = n) \\ &= \mu_z(y) \sum_{w \neq x} \frac{\mu_z(w)}{\mu_z(y)} p(w,y) \frac{g(x,w,n)}{\mu_z(w)} = \sum_{w \neq y} g(x,w,n) p(w,y). \end{aligned}$$

Define $f(x,y,n) = \mu_z(x)P_x(X_n = y, T_x > n)$. Then, for any $x \neq y$,

$$f(x,y,n+1) = \mu_z(x)P_x(X_{n+1} = y, T_x > n+1) = \mu_z(x)\sum_{w\neq x} P_x(X_n = w, T_x > n, X_{n+1} = y)$$
$$= \mu_z(x)\sum_{w\neq x} P_x(X_n = w, T_x > n)P(X_{n+1} = y \mid X_n = w) = \sum_{w\neq x} f(x,w,n)p(w,y).$$

Hence f and g have the same generating function with initial values $f(x,y,1) = \mu_z(x)P_x(X_1 = y, T_x > 1) = \mu_z(x)p(x,y) = \mu_z(y)q(y,x) = \mu_z(y)Q_y(U_x = 1) = g(x,y,1)$. Thus f and g are identical functions. By definition,

$$\mu_x(y) = \sum_{n=0}^{\infty} P_x(X_n = y, T_x > n) = \sum_{n=1}^{\infty} \frac{f(x,y,n)}{\mu_z(x)} = \frac{1}{\mu_z(x)}\sum_{n=1}^{\infty} g(x,y,n)$$
$$= \frac{1}{\mu_z(x)}\sum_{n=1}^{\infty} \mu_z(y)Q_y(U_x = n) = \frac{\mu_z(y)}{\mu_z(x)}Q_y(U_x < \infty) = \frac{\mu_z(y)}{\mu_z(x)}.$$