STAT 3006: Statistical Computing
Lecture 7*

26 February

When we want to sample from a distribution $f(x)$, if $f(x)$ is univariate, we can make use of inverse method or accept-reject method; if $f(x)$ is multivariate and it is easy to sample from $f(x_i|x_1, \ldots, x_{i-1})$, we select sequential sampling; if $f(x)$ is multivariate and hard to be directly sampled, our idea is constructing a sequence of$\{X_n\}_{n=1}^{\infty}$ such that X_n's limiting distribution is $f(x)$. Give an initial value X_0, we iteratively obtain X_{n+1} $(n = 0, 1, \ldots,)$. When n is large enough (X_n's distribution is very close to $f(x)$), we treat X_n as a sample from $f(x)$. This kind of approximation sampling method is called Markov chain Monte Carlo method. In this lecture, we will first review the Markov chain.

6 Review of Markov Chain

Definition 6.1. A Markov chain is a sequence of random variables $X_n(n \geq 0)$, where each X_n takes value at a discrete (finite or countable) set S, s.t. $P(X_{n+1} = j|X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = j|X_n = i)$ $\forall n \geq 0$. S is state space. If S is finite, the Markov chain is finite state Markov chain. If $P(X_{n+1} = j|X_n = i)$ is nothing to do with n, the Markov chain is time-homogeneous. For time-homogeneous Markov chain, $P = (p_{ij})_{i,j \in S}$ is transition matrix satisfying $p_{ij} \geq 0$ and $\sum_{j \in S} p_{ij} = 1$.

Throughout this course, for simplicity, we use Markov chain to represent time-homogeneous Markov chain.

(Example) In the simple symmetric random walk, $P(Y_n = 1) = P(Y_n = -1) = \frac{1}{2}$. Let $X_n = \sum_{i=1}^{n} Y_i$, then $\{X_n\}_{n=1}^{\infty}$ construct a Markov chain. Given the state i of the current step, the state of the next step is either $i+1$ or $i-1$. It is independent of the states of the past.

6.1 Chapman-Kolmogorov Equation

Now we investigate the transition matrix P. Noteworthy, (i, j) entry of P denotes the probability that the state of the chain changes from i to j by one step. What is the probability that the

*If you have any question about the note, please send an email to xyluo@link.cuhk.edu.hk state of the chain changes from i to j by two steps?

$$
\begin{aligned}
p_{ij}^{(2)} := P(X_{n+2} = j | X_n = i) &= \sum_{k \in S} P(X_{n+2} = j, X_{n+1} = k | X_n = i) \\
&= \sum_{k \in S} P(X_{n+2} = j | X_{n+1} = k, X_n = i) P(X_{n+1} = k | X_n = i) \\
&= \sum_{k \in S} P(X_{n+2} = j | X_{n+1} = k) P(X_{n+1} = k | X_n = i) \\
&= \sum_{k \in S} p_{ik} p_{kj}. \qquad (6.1)
\end{aligned}
$$

If we put all $p_{ij}^{(2)}$ $(i, j \in S)$ into a matrix denoted by $P^{(2)}$, then $P^{(2)}$ is $P \cdot P = P^2$ based on Equation 6.1. Generally, for a n-step transition matrix $P^{(n)} = \{p_{ij}^{(n)}\}_{i,j \in S}$, $P^{(n)} = P^n$, the matrix product of n the same transition matrices. The Chapman-Kolmogorov equation provides a more general representation of $P^{(n)}$

Theorem 6.1. *For a $n+m$-step transition probability $p_{ij}^{(n+m)}$, we have $p_{ij}^{(n+m)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}^{(m)}$.*

Proof. $P^{(n+m)} = P^{n+m} = P^n \cdot P^m = P^{(n)} \cdot P^{(m)}$.

The theorem tells us the probability from i to j by $n + m$ steps is the sum of probabilities that the chain jumps from i to k by n steps and then from k to j by m steps across all $k \in S$.

6.2 Classification of States

Definition 6.2. A state j is accessible from state i if $\exists n \geq 0$ s.t. $p_{ij}^{(n)} > 0$. In this case, we write it as $i \to j$. If $i \to j$ and $j \to i$, then states i and j communicate with each other, denoted by $i \leftrightarrow j$.

Proposition 6.2. *The relation "$\leftrightarrow$" is equivalent relation which satisfies the three following conditions: 1)(reflexivity) $i \leftrightarrow i$; 2) (symmetric) $i \leftrightarrow j \Rightarrow j \leftrightarrow i$; 3) transitivity: $i \leftrightarrow j, j \leftrightarrow i \Rightarrow i \leftrightarrow k$.*

If two states communicate, then we say they belong to the same equivalent class. Therefore, we can divide state set S into several equivalent classes.

Definition 6.3. Markov Chain is said to be irreducible if there is only one equivalent class.

Definition 6.4. A state i is absorbing if $p_{ii} = 1$.

Definition 6.5. A state i is periodic with period d if d is the greatest common divisor of $T(i) = \{t \geq 1 : p_{ii}^{(t)} > 0\}$. If $d = 1$, the state is aperiodic.

Definition 6.6. Let $f_i := P(\exists t > 0 s.t. X_t = i | X_0 = i)$ denote the probability to return to state i when starting from state i. If $f_i < 1$, state i is transient. If $f_i = 1$, state i is recurrent.

Definition 6.7. When $f_i = 1$, denote the first return time to state i by T_i $(T_i > 0)$. If $E(T_i | X_0 = i) < \infty$, state i is positive recurrent. If $E(T_i | X_0 = i) = \infty$, state i is null recurrent.

Theorem 6.3. *All states in an equivalent class have the same period and are either all recurrent or all transient.*

Theorem 6.4. *In finite state Markov chain, all recurrent states are positive recurrent.*

6.3 Stationary and Limiting Distribution

Definition 6.8. $\pi^* = (\pi_1^*, \pi_2^*, \ldots)$ is a stationary distribution for Markov chain $\{X_n\}_{n\geq 0}$ if $\pi^* = \pi^* P$.

Remark 1. π^* does not necessarily exist, nor is it necessarily unique.

Definition 6.9. π^* is the limiting distribution for Markov chain $\{X_n\}$ if for any initial distribution, $\lim_{n\to\infty} P(X_n = i) = \pi_i^*$.

Remark 2. If π^* is a limiting distribution, it is also stationary. Limiting distribution does not necessarily exist, but if it exists then it is unique.

Theorem 6.5. *Let $(X_n; n \geq 0)$ be an irreducible, aperiodic Markov chain. Assuming it admits a stationary distribution π^*, then π^* is also the limiting distribution.*

Definition 6.10. If π^* satisfies $\pi_i^* p_{ij} = \pi_j^* p_{ji}\ \forall i, j$, then we say π^* satisfies the detailed balance condition.

Proposition 6.6. *If π^* satisfies detailed balance condition, then π^* is a stationary distribution.*

Proof. $\sum_{i\in S} \pi_i^* p_{ij} = \sum_{i\in S} \pi_j^* p_{ji} = \pi_j^*$.

Example, consider a two-state Markov chain $\{X_n\}$, $P(X_{n+1} = 1|X_n = 0) = p$, $P(X_{n+1} = 0|X_n = 0) = 1 - p$, $P(X_{n+1} = 0|X_n = 1) = q$ and $P(X_{n+1} = 1|X_n = 1) = 1 - q$ $(p > 0, q > 0)$. We can prove that the Markov chain is irreducible. $P($*the chain does not return to* $0|X_0 = 0) = \lim_{n\to\infty} p(1-q)^n = 0$, so the Markov chain is also recurrent. By solving $\pi^* = \pi^* P$, we have $\pi_1^* = \frac{q}{p+q}$ and $\pi_2^* = \frac{p}{p+q}$. When $p + q < 2$, the Markov chain is aperiodic. In this case, π^* is also the limiting distribution by theorem. When $p = 1, q = 1$, $\pi^* = (1/2, 1/2)$, the Markov chain has period 2, so π^* is not the limiting distribution any more. There is no limiting distribution, as $P(X_n = 0|X_0 = 0) = 1$ if n is even; $P(X_n = 0|X_0 = 0) = 0$ if n is odd. The limit of $P(X_n = 0)$ does not exist.