Lecture 15
Discrete-Time Markov Chains (DTMC)

ELE 525: Random Processes in Information Systems

Hisashi Kobayashi

Department of Electrical Engineering
Princeton University
November 11, 2013

15 Discrete-Time Markov Chains

15.1 Markov Processes and Markov Chains

Definition 15.1 (Markov process). A random process $X(t)$ is called a Markov process if for any $t_1 < t_2$

$$P[X(t_2) \leq x | X(t); -\infty < t \leq t_1] = P[X(t_2) \leq x | X(t_1)].$$

(15.1)

Consider a set of $n$ arbitrarily chosen instants $t_1 < t_2 < \ldots < t_{n-1} < t_n$

$$P[X(t_n) \leq x | X(t_i); i = 1, 2, \ldots, n-1] = P[X(t_n) \leq x | X(t_{n-1})].$$

(15.2)

Definition 15.2 ($h$-th order Markov chain). A random sequence $X_n$ is called a Markov chain of order $h$ (or an $h$-th order Markov chain), when $X_n$ takes a finite or countably infinite number of states $S = \{0, 1, 2, \ldots\}$, and if the state of $X_n$ depends only on the last $h$ states, i.e., if

$$P[X_n | X_k; -\infty < k \leq n-1] = P[X_n | X_{n-h}, \ldots, X_{n-1}], \text{ for all } n.$$  

(15.4)

A first-order Markov chain is called a simple Markov chain.
Definition 15.3 (Transition probabilities). The conditional probabilities of a (simple) Markov chain

\[ P[X_{n+1} = j | X_n = i] = P_{ij}(n), \quad i, j \in \mathcal{S}, \quad n = 0, 1, 2, \ldots \]  \hspace{1cm} (15.7)

are called the \textbf{(one-step) transition probabilities}. The matrix

\[ P(n) = [P_{ij}(n)] \]  \hspace{1cm} (15.8)

is called the \textbf{transition probability matrix (TPM)}.

Example 15.1: Random walk.

\[ P[X_n = +1] = p, \quad P[X_n = -1] = 1 - p \triangleq q. \]

\[ S_n = X_1 + X_2 + \ldots + X_n = S_{n-1} + X_n \]

\[ P_{ij}(n) = \begin{cases} p, & j = i + 1, \\ q, & j = i - 1, \\ 0, & \text{otherwise}, \end{cases} \]
\[ p_j(n + 1) = \sum_{i \in S} p_i(n) P_{ij}(n), \quad j \in S, \quad (15.10) \]

If all \( P_{ii}(n) \) are independent of \( n \),

\[ p^T(n + 1) = p^T(n) P \quad \text{for all} \quad n = 0, 1, 2, \ldots, \quad (15.11) \]

called a \textbf{homogeneous} or \textbf{stationary} Markov chain.

\[
\begin{bmatrix}
1 & 1 & 1 \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
0 & 0 & 0 \\
\end{bmatrix}
\]

\[
P = \begin{bmatrix}
0 & 1 & 0 \\
\frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\
0 & 0 & 1 \\
\end{bmatrix}
\]
By applying (15.11) recursively, we have

\[ P_{ij}(n) \geq 0 \quad \text{for all } i, j, \quad \text{and } \quad n = 0, 1, 2, \ldots \quad (15.13) \]

\[ \sum_{j \in S} P_{ij}(n) = 1 \quad \text{for all } i \in S, \quad \text{and } \quad n = 0, 1, 2, \ldots . \quad (15.14) \]

Matrices satisfying (15.13) and (15.14) are called \textbf{stochastic}.

Any stochastic square matrix may serve as a TPM.

By applying (15.11) recursively, we have

\[ p^T(n) = p^T(0) P(0) P(1) \cdots P(n - 1), \quad (15.15) \]

If the Markov chain is time-homogeneous or stationary,

\[ p^T(n) = p^T(0) P^n, \quad (15.16) \]
If we set \( n \leftarrow m + n \)

\[
p^{\top}(m + n) = p^{\top}(0)P^m P^n = p^{\top}(m)P^n. \tag{15.17}
\]

By denoting

\[
P^n = \begin{bmatrix} P_{ij}^{(n)} \end{bmatrix}, \quad i, j \in S, \tag{15.18}
\]

\[
p_j(m + n) = \sum_{i \in S} p_i(m) P_{ij}^{(n)}, \text{ where } i, j \in S, \ m, n = 0, 1, 2, \ldots. \tag{15.19}
\]

**Theorem 15.1 (Chapman-Kolmogorov Equations).** In a homogeneous Markov chain the following equations hold for all states \( i, j, k \in S \) and time (or step) indices \( m, n = 0, 1, 2 \ldots \):

\[
P_{ik}^{(m+n)} = \sum_{j \in S} P_{ij}^{(m)} P_{jk}^{(n)}. \tag{15.20}
\]
Proof. The above result is immediately obtainable from

\[ P^{m+n} = P^m P^n. \]  \hspace{1cm} (15.21)

Since the Markov chain is homogeneous,

\[ P_{i:k}^{(m+n)} = P[X_{m+n} = k | X_0 = i] \]
\[ = \sum_{j \in S} P[X_{m+n} = k, X_m = j | X_0 = i] \]
\[ \overset{(a)}{=} \sum_{j \in S} P[X_{m+n} = k | X_m = j, X_0 = i] P[X_m = j | X_0 = i] \]
\[ \overset{(b)}{=} \sum_{j \in S} P[X_{m+n} = k | X_m = j] P[X_m = j | X_0 = i] \]
\[ = \sum_{j \in S} P_{j:k}^{(n)} P_{i:j}^{(m)}, \] \hspace{1cm} (15.22)

(a) is obtained using the formula \[ P[A \cap B | C] = P[A | B \cap C] P[B | C] \]
(b) makes use of the Markov property
15.2 Computation of State Probabilities

15.2.1 Generating Function Method

\[ g(z) = \sum_{n=0}^{\infty} p(n)z^n, \]  \hfill (15.23)

\[ g^{\top}(z)P = \sum_{n=0}^{\infty} p^{\top}(n+1)z^n = z^{-1} \sum_{n=0}^{\infty} p^{\top}(n+1)z^{n+1} \]

\[ = z^{-1}g^{\top}(z) - z^{-1}p^{\top}(0), \]  \hfill (15.24)

\[ g^{\top}(z) = p^{\top}(0)[I - Pz]^{-1}, \]  \hfill (15.25)
15.2.2 Spectral Expansion Method

\[ P u_i = \lambda_i u_i, \quad i \in S = \{0, 1, 2, \ldots, M - 1\}, \]  
\[ U = [u_0 u_1 \cdots u_{M-1}] \]  
\[ \Lambda = \begin{bmatrix} \lambda_0 & 0 & \cdots & 0 \\ 0 & \lambda_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{M-1} \end{bmatrix}. \]  

\[ P = U \Lambda U^{-1} = U \Lambda V, \]  

\[ V = U^{-1}. \]  

By multiplying \( V \) on the left of (15.36),

\[ VP = \Lambda V. \]
Defining a set of row vectors $\mathbf{v}_i^\top$'s by

$$
V = \begin{bmatrix}
\mathbf{v}_0^\top \\
\mathbf{v}_1^\top \\
\vdots \\
\mathbf{v}_{M-1}^\top 
\end{bmatrix},
$$

we find from (15.38)

$$
\mathbf{v}_i^\top \mathbf{P} = \lambda_i \mathbf{v}_i^\top, \quad i \in \mathcal{S}.
$$

From (15.37)

$$
V U = I,
$$

$\mathbf{v}_i$ and $\mathbf{u}_j$ are bi-orthonormal

$$
\mathbf{v}_i^\top \mathbf{u}_j = \delta_{ij}, \quad i, j \in \mathcal{S}.
$$

From (15.36)

$$
\mathbf{P}^2 = U \Lambda U^{-1} U \Lambda U^{-1} = U \Lambda^2 U^{-1}.
$$
\[ P^n = U \Lambda^n U^{-1} = \sum_{i \in S} \lambda_i^n u_i v_i^\top = \sum_{i \in S} \lambda_i^n E_i, \quad (15.44) \]

\[ E_i = u_i v_i^\top, \quad i \in S, \quad (15.45) \]

are the projection matrices

\[ \sum_{i \in S} E_i = \sum_{i \in S} E_i^\top = I, \quad (15.46) \]

\[ P_{ij}^{(n)} = \sum_{k \in S} \lambda_k^n u_{ki} v_{kj}, \quad i, j \in S, \quad n = 0, 1, 2, \ldots. \quad (15.47) \]

\[ p^\top(n) = p^\top(0) P^n = \sum_{k \in S} \lambda_k^n p^\top(0) u_k v_k^\top. \quad (15.48) \]

\[ p_i(n) = \sum_{k \in S} \lambda_k^n \left( p^\top(0) u_k \right) v_{ki}, \quad i \in S. \quad (15.49) \]
15.3 Classification of States

![Classification Diagram]

**Figure 15.2** Classification of states in a Markov chain.
Figure 15.3 An example of a Markov chain with various states.
15.3.1 Recurrent and Transient States

$T_{ij}$: the **first-passage time** from $i$ to $j$.

Let $f_{ij}^{(n)}$ be

$$f_{ij}^{(n)} \triangleq P[T_{ij} = n], \ i, j \in S, \ n = 1, 2, \ldots. \quad (15.58)$$

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}, \ i, j \in S \quad (15.59)$$

If $f_{ij} = 1$, \( \{f_{ii}^{(n)} : n = 1, 2, \ldots\} \) represents the probability distribution of the first-passage time.

If $f_{ij} < 1$, the process may never reach state $j$.

$$f_{ii} = \sum_{n=1}^{\infty} f_{ii}^{(n)} \quad (15.60)$$

is the probability that the system ever returns to state $i$. 
Definition 15.4 (Recurrent state and transient state). State $i \in S$ is called a recurrent state (or a persistent state) if $f_{ii} = 1$. It is called a transient state if $f_{ii} < 1$.

The mean first-passage time (or expected first-passage time)

$$
\mu_{ij} = E[T_{ij}] = \begin{cases} 
\sum_{n=1}^{\infty} n f_{ij}^{(n)}, & \text{if } f_{ij} = 1 \\
\infty, & \text{if } f_{ij} < 1.
\end{cases}, \quad i, j \in S.
$$

(15.61)

When $f_{ij} = 1$

$$
\mu_{ij} = P_{ij} + \sum_{k \neq j} P_{ik} (\mu_{kj} + 1),
$$

(15.62)

When $j = i$, the mean first-passage time is called the mean recurrence time (or expected recurrence time).
Definition 15.5 (Null-recurrent state and positive-recurrent state). A recurrent state \( i \in S \) is called a null-recurrent state if \( \mu_{ii} = \infty \), and is called a positive-recurrent (or regular-recurrent) state if \( \mu_{ii} < \infty \).

If \( P_{ii}^{(n)} = 0 \) for \( n \neq d_i, 2d_i, \ldots \), then state \( i \) is called periodic with period \( d_i \).

\[
d_i = \gcd \left\{ n : P_{ii}^{(n)} > 0 \right\},
\]

(15.63)

Definition 15.6 (Periodic state and aperiodic state). A state \( i \in S \) is called periodic with period \( d_i \) if \( d_i > 1 \), and is called aperiodic if \( d_i = 1 \).

Definition 15.7 (Ergodic state). State \( i \in S \) is called an ergodic state, if it is positive-recurrent and aperiodic, i.e., \( f_{ii} = 1 \), \( \mu_{ii} < \infty \), and \( d_i = 1 \).

Definition 15.8 (Absorbing state). State \( i \in S \) is called an absorbing state if the transition probability satisfies \( P_{ii} = 1 \).
15.3.2 Criteria for State Classification

\[ G_{ij}(z) \triangleq \sum_{n=0}^{\infty} P_{ij}^{(n)} z^n, \quad i, j \in S, \quad (15.64) \]

\[ F_{ij}(z) \triangleq \sum_{n=0}^{\infty} f_{ij}^{(n)} z^n, \quad i, j \in S. \quad (15.65) \]

**Theorem 15.2** (Generating functions of \( n \)-step transition and first-passage probabilities). The generating functions \( G_{ij}(z) \) and \( F_{ij}(z) \) defined above are related according to

\[ G_{ij}(z) = \delta_{ij} + F_{ij}(z)G_{jj}(z), \quad i, j \in S. \quad (15.66) \]

**Proof.**

Read p. 441

For \( i = j \), we obtain

\[ G_{ii}(z) = \frac{1}{1 - F_{ii}(z)}. \quad (15.70) \]
The first-passage probability

\[
    f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)} = F_{ij}(1).
\]  

(15.71)

The mean first-passage time

\[
    \mu_{ij} = \sum_{n=1}^{\infty} n f_{ij}^{(n)} = F'_{ij}(1).
\]  

(15.72)

**Note:** We shall skip in the class discussion of Theorem 15.3 (Conditions for recurrent and transient states) and Theorem 15.5 (Conditions for null-recurrent, ergodic or periodic states) of pp. 442-445, but be aware that these theorems exist.
15.3.3 Communicating States and an Irreducible Markov Chain

Definition 15.9 (Reachable states and communicating states). We say that state $j \in S$ is reachable (or accessible) from state $i \in S$ if there is an integer $n \geq 1$ such that $P_{ij}^{(n)} > 0$. If state $i$ is reachable from state $j$ and state $j$ is reachable from state $i$, then the states $i$ and $j$ are said to communicate$^4$, written as $i \leftrightarrow j$.

Theorem 15.6 (Reachable states from a recurrent state). Suppose state $j$ is reachable from a recurrent state $i$ (i.e., $j \leftarrow i$). Then state $i$ is also reachable from state $j$ ($i \leftarrow j$); hence states $i$ and $j$ communicate ($i \leftrightarrow j$). Moreover, state $j$ is also recurrent.

Proof. See pp. 445-446.
If two states $i$ and $j$ do not communicate, then either

$$P_{ij}^{(n)} = 0 \text{ for all } n \geq 1,$$

or

$$P_{ji}^{(n)} = 0 \text{ for all } n \geq 1,$$

or both relations are true. The relation of communications “$\leftrightarrow$” is an **equivalence relation**; that is, the following three properties hold:

1. **Reflexive Property**: $i \leftrightarrow i$ for all $i \in S$.
2. **Symmetric Property**: if $i \leftrightarrow j$, then $j \leftrightarrow i$.
3. **Transitive Property**: if $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$. (15.97)

Property 1 is a consequence of $P_{ij}^{(0)} = \delta_{ij}$; Property 2 is apparent, and 3 is easy to verify (Problem 15.4).
Definition 15.10 (Closed or open set and an irreducible set of states). A set \( C \) of states is called closed if

\[
P_{ij} = 0 \text{ for all } i \in C, \ j \notin C,
\]
and is called open, otherwise.

A set \( C \) is called irreducible if

\[
i \leftrightarrow j \text{ for all } i, j \in C.
\]

If all the states \( S \) of a Markov chain are irreducible, the chain is called irreducible.

Theorem 15.7 (Property of an irreducible Markov chain). In an irreducible Markov chain all states belong to the same class: They are all transient, all null-recurrent, or all positive-recurrent. Furthermore, they are either all aperiodic or all periodic with the same period.

Proof: See p. 447.
Theorem 15.8 (Decomposition of a Markov chain). If a Markov chain is not an irreducible chain, its state space $S$ can be uniquely partitioned as

$$S = \mathcal{T} \cup C_1 \cup C_2 \cup \cdots,$$

where $\mathcal{T}$ is the set of transient states, and the $C_r$'s ($r = 1, 2, \ldots$), are irreducible closed sets of recurrent states.

**Proof.** See p. 447
15.3.4 Stationary Distribution of an Aperiodic Irreducible Chain

\[
\lim_{{n \to \infty}} p(n) = \pi. \quad (15.98)
\]

Then by applying this limit to (15.11),

\[
p^\top (n + 1) = p^\top (n) P \quad \text{for all } n = 0, 1, 2, \ldots, \quad (15.11)
\]

\[
\pi^\top = \pi^\top P. \quad (15.99)
\]

**Definition 15.11 (Stationary distribution).** A probability distribution \( \pi \) satisfying (15.99) is called a **stationary distribution** or an **invariant distribution**: If the initial probability \( p(0) \) is set to \( \pi \), then \( p(n) = \pi \) for all \( n \geq 0 \). \( \square \)
Theorem 15.9 (Stationary distribution of an irreducible aperiodic Markov chain). In an irreducible aperiodic Markov chain, one of the following two alternatives holds:

1. The states are all transient or all null-recurrent; in this case there exists no stationary distribution and

\[ \lim_{n \to \infty} P_{ij}^{(n)} = 0, \text{ for all } i, j \in S. \]  \hspace{1cm} (15.100)

2. All states are ergodic; in this case there exists a unique stationary distribution and

\[ \lim_{n \to \infty} P_{ij}^{(n)} = \pi_j, \text{ for all } i, j \in S. \]  \hspace{1cm} (15.101)

Furthermore, \( \pi_i \) is equal to the reciprocal of the mean recurrence time for state \( i \), i.e.,

\[ \pi_i = \frac{1}{\mu_{ii}}, \quad i \in S. \]  \hspace{1cm} (15.102)

\textbf{Proof}. We skip the proof, since it uses Theorems 15.3 and 15.5 which we skipped. Read pp. 448-449.
The theorem assures us that (15.101) holds whenever the states are ergodic.

\[
\lim_{n \to \infty} p_j(n) = \lim_{n \to \infty} \sum_{i \in S} p_i(0) P_{ij}^{(n)} = \sum_{i \in S} p_i(0) \pi_j = \pi_j, \quad j \in S. \tag{15.103}
\]

Thus, the stationary distribution becomes the **steady-state distribution**.

The stationary distribution can be computed by solving linear equations of (15.99) with the linear constraint condition:

\[
\pi^\top 1 = 1, \tag{15.104}
\]

Alternatively, a simple but useful computational formula can be found

\[
\pi^\top = 1^\top (P + E - I)^{-1}. \tag{15.107}
\]