Discrete-State Markov Chains

(A) Basic Concepts. Transition Probabilities
(B) Classification of States and Chains
(C) Invariant Distributions and Large-Time Behavior
(D) Finite-State Markov Chains

References
[GS] Grimmett and Stirzaker, 2001.
[HPS] Hoel, Port, and Stone, 1972.
[L] Lawler, 2006.

(A) Basic Concepts. Transition Probabilities

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let S be a finite or countably infinite set. We will specify it to be

$$S = \{1, 2, \cdots, N\} \quad \text{for some } N \in \mathbb{N}$$

or

$$S = \mathbb{N} = \{1, 2, \cdots\}.$$

We consider a discrete-time stochastic process

$$X_0, X_1, \cdots, X_n, \cdots,$$

where each $X_n : \Omega \to S$ is a random variable, describing the state of an underlying stochastic system at time n.

We call S the <u>state space</u>, and each $i \in S$ a <u>state</u>. It is a discrete space, meaning that it is finite or countably infinite.

Two general questions:

(1) Determine <u>the joint distributions</u>

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n)$$

for any $n \geq 0$ and $i_0, \dots, i_n \in S$. Note that

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n)$$
$$= \mathbb{P}(X_0 = i_0)\,\mathbb{P}(X_1 = i_1 \mid X_0 = i_0)\,\mathbb{P}(X_2 = i_2 \mid X_0 = i_0, X_1 = i_1)$$
$$\cdots \mathbb{P}(X_n = i_n \mid X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}).$$

(2) Determine the large-time behavior of the distribution for X_n:

$$\mathbb{P}(X_n = j), \quad j \in S, \qquad n \gg 1.$$

<u>Definition</u> (1) A stochastic process $\{X_n\}_{n=0}^{\infty}$ is a <u>Markov process</u>, if

$$\mathbb{P}(X_{n+1} = i_{n+1} \mid X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n)$$

for any $n \geq 0$ and any $i_0, i_1, \dots, i_{n+1} \in S$.

(2) A stochastic process $\{X_n\}_{n=0}^{\infty}$ is <u>time-homogeneous</u>, if

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \mathbb{P}(X_1 = j \mid X_0 = i)$$

for any $n \geq 0$ and any $i, j \in S$.

<u>Assumption</u>

Unless otherwise stated, we assume always that the discrete-time stochastic process $\{X_n\}_{n=0}^{\infty}$ is a time-homogeneous Markov chain with a discrete state space $S=\{1,\cdots,N\}$ or $S=\mathbb{N}$.

The following two identities are very useful:

(1) $$\mathbb{P}(X_{n_m}=i_m,\cdots,X_{n_{k+1}}=i_{k+1}\mid X_{n_k}=i_k,\cdots,X_{n_1}=i_1)$$
$$=\mathbb{P}(X_{n_m}=i_m,\cdots,X_{n_{k+1}}=i_{k+1}\mid X_{n_k}=i_k)$$
where $0\le n_1<\cdots<n_k<n_{k+1}<\cdots<n_m$, and $i_1,\cdots,i_m\in S$.

(2) $$\mathbb{P}(X_{n+1}\in B_1,\cdots,X_{n+k}\in B_k\mid X_0\in A_0,\cdots,X_{n-1}\in A_{n-1},X_n=i_0)$$
$$=\mathbb{P}(X_{n+1}\in B_1,\cdots,X_{n+k}\in B_k\mid X_n=i_0)$$
$$=\mathbb{P}(X_1\in B_1,\cdots,X_k\in B_k\mid X_0=i_0),$$
where n,k are positive integers, $A_1,\cdots,A_{n-1}$ and $B_1,\cdots,B_k$ are all <u>subsets</u> of S, and $i_0\in S$.
(They can be S or ϕ or other subsets.)

We prove these at the end of this set of notes.

With the Markovian property, we can rewrite the joint distribution
$$\mathbb{P}(X_0=i_0,X_1=i_1,\cdots,X_n=i_n)$$
$$=\mathbb{P}(X_0=i_0)\,\mathbb{P}(X_1=i_1\mid X_0=i_0)\cdots\mathbb{P}(X_n=i_n\mid X_0=i_0,\cdots,X_{n-1}=i_{n-1})$$
$$=\mathbb{P}(X_0=i_0)\,\mathbb{P}(X_1=i_1\mid X_0=i_0)\cdots\mathbb{P}(X_n=i_n\mid X_{n-1}=i_{n-1})$$

<u>Definition</u> Let $i, j \in S$. The <u>transition probability</u> is

$$p(i,j) := \mathbb{P}(X_{n+1} = j \mid X_n = i) = \mathbb{P}(X_1 = j \mid X_0 = i)$$

for any $n \geq 0$.

The joint distribution can finally be written as

$$\begin{aligned}
&\mathbb{P}(X_0 = i_0, \dots, X_n = i_n) \\
&= \mathbb{P}(X_0 = i_0)\, \mathbb{P}(X_1 = i_1 \mid X_0 = i_0) \cdots \mathbb{P}(X_n = i_n \mid X_{n-1} = i_{n-1}) \\
&= \mathbb{P}(X_0 = i_0)\, p(i_0, i_1) \cdots p(i_{n-1}, i_n).
\end{aligned}$$

Thus, the transition probabilities $p(i,j)$ $(i, j \in S)$, together with the distribution of the initial random variable X_0, $\mathbb{P}(X_0 = i_0)$ $(i_0 \in S)$, determine completely the joint distributions.

<u>Definition</u> The <u>transition matrix</u> is $P = (p(i,j))_{i,j \in S}$

Here, $i, j \in S$ means that $i, j = 1, \dots, N$ for $S = \{1, \dots, N\}$ and $i, j = 1, 2, \dots$ for $S = \mathbb{N}$.

The transion matrix P is a stochastic matrix, i.e., it satisfies the following

(1) all entries $p(i,j) \geq 0$ $\forall i, j \in S$;

(2) $\sum_{j \in S} p(i,j) = 1$ (each row sums to 1) $\forall i \in S$.

Note: If $|S| < \infty$, i.e., S is finite, the transition matrix is the usual, square, stochastic matrix of finitely many rows and columns. Here, we allow infinite matrices.

- For a (finite) stochastic matrix $P = (p(i,j))_{i,j=1}^{n}$, we observe that 1 is an eigenvalue of P, and $\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$ is a corresponding eigenvector. This is true for an infinite stochastic matrix.

- It is easy to check that the product of two stochastic matrices is still a stochastic matrix.

<u>Definition</u> Let $n \geq 0$ be an integer and $i, j \in S$. The <u>n-step transition probability</u> is

$$p_n(i,j) = \mathbb{P}(X_n = j \mid X_0 = i).$$

For $n=0$, $p_0(i,j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} = \delta_{ij}$.

For $n=1$, $p_1(i,j) = p(i,j)$.

The <u>n-step transition matrix</u> is $P_n = (p_n(i,j))_{i,j \in S}$.

A probability vector $a = (a(i))_{i \in S}$ is a vector of nonnegative components and all summed up to 1.

For each $n \geq 0$, we define <u>the distribution (vector)</u> π_n for X_n by

$$\pi_n(i) = \mathbb{P}(X_n = i), \quad i \in S.$$

Call π_0 the <u>initial distribution</u> of the chain $\{X_n\}_{n=0}^{\infty}$.

<u>Proposition</u>

(1) For any integers $n, k \geq 0$ and any $i, j \in S$,

$$p_n(i,j) = \mathbb{P}(X_{n+k} = j \mid X_k = i)$$

(2) <u>The Chapman-Kolmogorov equation</u>

$$p_{n+m}(i,j) = \sum_{k\in S} p_n(i,k)\, p_m(k,j)$$

$$\forall i,j \in S,\ \forall n,m \geq 0 : \text{integers}.$$

In particular, this means

$$P_n P_m = P_{n+m}. \quad (\cancel{\text{Hence}})$$

Also, $\quad P_n = P^n$

i.e., ~~the~~ $p_n(i,j)$ is the (i,j)-entry of P^n. This also implies that P_n is a stochastic matrix.

(3) For any integer $n \geq 0$,

$$\pi_n = \pi_0 P_n = \pi_0 P^n,$$

which is

$$\mathbb{P}(X_n = j) = \sum_{i\in S} \pi_0(i)\, p_n(i,j) \qquad \forall j \in S.$$

Also, $\quad \pi_{n+1} = \pi_n P.$

We end this part by providing a brief proof of the identities in page $\boxed{3}$

To prove (1), we first show

(*) $$\mathbb{P}(X_{n_k+m} = j \mid X_{n_k} = i_k, \dots, X_{n_1} = i_1) = \mathbb{P}(X_{n_k+m} = j \mid X_{n_k} = i_k),$$

where $m \geq 1$, $n_1 < n_2 < \dots < n_k$, and all $i_1, \dots, i_k, j \in S$.

By induction on $m \geq 1$.

$m=1$. Let I denote the set of "missing indices": $l \in \{0, 1, \dots, n_k\}$ ~~such that $l \notin$~~ $\{n_1, n_2, \dots, n_k\}$.

If $I = \phi$, then (*) is true for $m=1$ by the definition of a Markov chain. Assume $I \neq \phi$. Then, using the elementary but useful identity

$$\mathbb{P}(A\cap B|C) = \mathbb{P}(A|B\cap C)\,\mathbb{P}(B|C),$$

and denote $I=\{l_1,\dots,l_s\}$, we have

$$\mathbb{P}(X_{n_k+1}=j \mid X_{n_k}=i_k,\dots,X_{n_1}=i_1)$$

$$=\sum_{j_1,\dots,j_s\in S}\mathbb{P}(X_{n_k+1}=j, X_{l_1}=j_1,\dots,X_{l_s}=j_s \mid X_{n_k}=i_k,\dots,X_{n_1}=i_1)$$

$$=\sum_{j_1,\dots,j_s\in S}\mathbb{P}(X_{n_k+1}=j \mid X_{n_k}=i_k,\dots,X_{n_1}=i_1, X_{l_1}=j_1,\dots,X_{l_s}=j_s)$$

$$\cdot\,\mathbb{P}(X_{l_1}=j_1,\dots,X_{l_s}=j_s \mid X_{n_k}=i_k,\dots,X_{n_1}=i_1)$$

$$=\sum_{j_1,\dots,j_s\in S}\mathbb{P}(X_{n_k+1}=j \mid X_{n_k}=i_k)\cdot$$

$$\cdot\,\mathbb{P}(X_{l_1}=j_1,\dots,X_{l_s}=j_s \mid X_{n_k}=i_k,\dots,X_{n_1}=i_1)$$

$$=\mathbb{P}(X_{n_k+1}=j \mid X_{n_k}=i_k)\sum_{j_1,\dots,j_s\in S}\mathbb{P}(X_{l_1}=j_1,\dots,X_{l_s}=j_s \mid$$

$$X_{n_k}=i_k,\dots,X_{n_1}=i_1)$$

$$=\mathbb{P}(X_{n_k+1}=j \mid X_{n_k}=i_k)$$

So, $(*)$ is true for $m=1$.

Suppose $(*)$ is true for ~~any~~ $m\geq 1$. Then,

$$\mathbb{P}(X_{n_k+m+1}=j \mid X_{n_k}=i_k,\dots,X_{n_1}=i_1)$$

$$=\sum_{t\in S}\mathbb{P}(X_{n_k+m+1}=j, X_{n_k+m}=t \mid X_{n_k}=i_k,\dots,X_{n_1}=i_1)$$

$$=\sum_{t\in S}\mathbb{P}(X_{n_k+m+1}=j \mid X_{n_k+m}=t, X_{n_k}=i_k,\dots,X_{n_1}=i_1)$$

$$\cdot\,\mathbb{P}(X_{n_k+m}=t \mid X_{n_k}=i_k,\dots,X_{n_1}=i_1)$$

$$=\sum_{t\in S}\mathbb{P}(X_{n_k+m+1}=j \mid X_{n_k+m}=t)\,\mathbb{P}(X_{n_k+m}=t \mid X_{n_k}=i_k)$$

$$=\sum_{t\in S}\mathbb{P}(X_{n_k+m+1}=j, X_{n_k+m}=t \mid X_{n_k}=i_k)$$

$$=\mathbb{P}(X_{n_k+m+1}=j \mid X_{n_k}=i_k).$$

Thus (*) is true for $m+1$. Hence it is true for all $m \geq 1$.

Now, we prove the first identity on page $\boxed{3}$

$$\mathbb{P}(X_{n_m}=i_m, \cdots, X_{n_{k+1}}=i_{k+1} \mid X_{n_k}=i_k, \cdots, X_{n_1}=i_1)$$

$$= \frac{\mathbb{P}(X_{n_m}=i_m, \cdots, X_{n_{k+1}}=i_{k+1}, X_{n_k}=i_k, \cdots, X_{n_1}=i_1)}{\mathbb{P}(X_{n_k}=i_k, \cdots, X_{n_1}=i_1)}$$

$$= \frac{\mathbb{P}(X_{n_1}=i_1)\,\mathbb{P}(X_{n_2}=i_2 \mid X_{n_1}=i_1)\cdots \mathbb{P}(X_{n_m}=i_m \mid X_{n_1}=i_1, \cdots X_{n_{m-1}}=i_{m-1})}{\mathbb{P}(X_{n_1}=i_1)\,\mathbb{P}(X_{n_2}=i_2 \mid X_{n_1}=i_1)\cdots \mathbb{P}(X_{n_k}=i_k \mid X_{n_1}=i_1, \cdots, X_{n_{k-1}}=i_{k-1})}$$

$$= \mathbb{P}(X_{n_{k+1}}=i_{k+1} \mid X_{n_k}=i_k) \cancel{\cdots \mathbb{P}(X_{n_m}=i_m \mid X_{n_{m-1}}=i_{m-1})}$$

$$\cdot\, \mathbb{P}(X_{n_{k+2}}=i_{k+2} \mid X_{n_{k+1}}=i_{k+1}, X_{n_k}=i_k) \cdots$$

$$\cdot\, \mathbb{P}(X_{n_m}=i_m \mid X_{n_{m-1}}=i_{m-1}, \cdots, X_{n_k}=i_k)$$

$$= \frac{\mathbb{P}(X_{n_m}=i_m, X_{n_{m-1}}=i_{m-1}, \cdots, X_{n_k}=i_k)}{\mathbb{P}(X_{n_k}=i_k)}$$

$$= \mathbb{P}(X_{n_m}=i_m, \cdots, X_{n_{k+1}}=i_{k+1} \mid X_{n_k}=i_k).$$

This completes the proof of the 1st identity on page $\boxed{3}$

For the second identy on page $\boxed{3}$ we first show that

$$\mathbb{P}(X_{n_1}=i_1, \cdots, X_{n_k}=i_k \mid X_0 \in A_1, X_1 \in A_2, \cdots, X_{n-1} \in A_{n-1}, X_n = i_0\}$$

$$= \mathbb{P}(X_{n_1}=i_1, \cdots, X_{n_k}=i_k \mid X_n = i_0),$$

where $n \leq n_1 < \cdots < n_k$, $i_0, i_1, \cdots, i_k \in S$, and all $A_1, \cdots, A_{n-1}$ are subsets of S. This is proved by using the following property: If D_i are disjoint, and $\mathbb{P}(C \mid D_i) = p$ is independent of i, then $\mathbb{P}(C \mid \bigcup_i D_i) = p$.

We now have

$$\mathbb{P}(X_{n+1}\in B_1,\cdots,X_{n+k}\in B_k \mid X_0\in A_0,\cdots,X_{n-1}\in A_{n-1},X_n=i_0)$$

$$=\sum_{i_1\in B_1,\cdots,i_k\in B_k}\mathbb{P}(X_{n+1}=i_1,\cdots,X_{n+k}=i_k \mid X_0\in A_0,\cdots,X_{n-1}\in A_{n-1},X_n=i_0)$$

$$=\sum_{i_1\in B_1,\cdots,i_k\in B_k}\mathbb{P}(X_{n+1}=i_1,\cdots,X_{n+k}=i_k \mid X_n=i_0)$$

$$=\mathbb{P}(X_{n+1}\in B_1,\cdots,X_{n+k}\in B_k \mid X_n=i_0).$$

The proof is complete.

(B) Classification of States and Chains

Given a state $i\in S$. The chain may visit this state infinitely many times or otherwise only finitely many times, probabilistically. These two distinguished situations can be describe using different notions.

<u>Definition</u> A state $i\in S$ is <u>recurrent</u> if

$$\mathbb{P}(X_n=i \text{ for some } n\geq 1 \mid X_0=i)=1.$$

A state $i\in S$ is <u>transient</u> if

$$\mathbb{P}(X_n=i \text{ for some } n\geq 1 \mid X_0=i)<1.$$

So, any state is either recurrent or transient, but not both.

A useful concept is the hitting time.

<u>Definition</u> The <u>hitting time</u> for a state $i \in S$ is
$$T_i = \min\{n \geq 1 : X_n = i\}$$
if the minimum exist. Otherwise, we define $T_i = \infty$.

Clearly, the event $\{X_n = i \text{ for some } n \geq 1\} = \{T_i < \infty\}$.
Hence, we conclude:
$$i \in S \text{ is recurrent} \iff \mathbb{P}(T_i < \infty \mid X_0 = i) = 1,$$
$$i \in S \text{ is transient} \iff \mathbb{P}(T_i < \infty \mid X_0 = i) < 1.$$

The following result connect the hitting times with transition probabilities:

<u>Proposition</u> We have for any $n \geq 1$ and $i, j \in S$ that
$$p_n(i,j) = \sum_{k=1}^{n} \mathbb{P}(T_j = k \mid X_0 = i)\, p_{n-k}(j,j).$$

See [HPS] (Section 1.4.1).

The first passage times are also useful quantities.

<u>Definition</u> Let $i, j \in S$. Denote for any $n \geq 1$
$$f_n(i,j) = \mathbb{P}(X_1 \neq j, \dots, X_{n-1} \neq j, X_n = j \mid X_0 = i)$$
= probability that the first visit to state j, starting from state i, takes place at the nth step

Denote

$$f(i,j) = \sum_{n=1}^{\infty} f_n(i,j)$$

= probability that the chain ever visits state j, starting from state i.

We have

$$f_n(i,j) = \mathbb{P}(T_j = n \mid X_0 = i),$$

$$f(i,j) = \mathbb{P}(T_j < \infty \mid X_0 = i),$$

= probability of ever returning to j, starting from i,

$$1 - f(i,j) = \mathbb{P}(T_j = \infty \mid X_0 = i)$$

= probability of never returning to j, starting from i

Therefore

$i \in S$ is recurrent $\iff f(i,i) = 1$,

$i \in S$ is transient $\iff f(i,i) < 1$.

A more intuitive quantity is the number of visits to a state i.

<u>Definition</u> Let $i \in S$. Define

$$N(i) = \sum_{n=1}^{\infty} \mathbb{1}_i(X_n)$$

= # times the chain visits state i (after the initial time)

where $\mathbb{1}_i = \mathbb{1}_{\{i\}}$ is the indicator function of $\{i\}$. $\mathbb{1}_i(X_n) = \begin{cases} 1 & \text{if } X_n = i, \\ 0 & \text{if } X_n \neq i. \end{cases}$

We have clearly that

$$\{N(j) \geq 1\} = \{T_j < \infty\} \quad \forall j \in S.$$

In particular,

$$\mathbb{P}(N(j) \geq 1 \mid X_0 = i) = \mathbb{P}(T_j < \infty \mid X_0 = i) \quad \forall i, j \in S.$$

Hence,

$i \in S$ is recurrent $\iff \mathbb{P}(N(j) \geq 1 \mid X_0 = i) = 1$,

$i \in S$ is transient $\iff \mathbb{P}(N(i) \geq 1 \mid X_0 = i) < 1$.

Note also that for any $i \in S$,

$$\{N(i) = \infty\} = \{X_n = i \text{ for infinitely many } n\}$$

The expected number of visits to a state $j \in S$, starting from state i, is $\mathbb{E}(\mathbb{1}_j(X_n) \mid X_0 = i) = \mathbb{E}(X_n = j \mid X_0 = i) = p_n(i,j)$ for any $n \geq 0$.

Thus

$$\boxed{\mathbb{E}(N(j) \mid X_0 = i) = \sum_{n=1}^{\infty} \mathbb{E}(\mathbb{1}_j(X_n) \mid X_0 = i) = \sum_{n=1}^{\infty} p_n(i,j)}$$

The following result characterizes the fundamental difference between a recurrent state and a transient state.

<u>Theorem</u>

(1) A state $i \in S$ is recurrent

$$\iff \sum_{n=1}^{\infty} p_n(i,i) = \infty$$

$$\iff \mathbb{P}(N(i) = \infty \mid X_0 = i) = 1.$$

In this case, for any $k \in S$,

$$\mathbb{P}(N(i)=\infty \mid X_0=k) = \mathbb{P}(T_i<\infty \mid X_0=k) = f(k,i)$$

If $f(k,i)=0$ then all $p_n(k,i)=0$ $(n=1,2,\cdots)$

If $f(k,i)>0$ then $\sum_{n=1}^{\infty} p_n(k,i)=\infty$.

(2) A state $i\in S$ is transient

$$\iff \sum_{n=1}^{\infty} p_n(i,i) < \infty$$

$$\iff \mathbb{P}(N(i)=\infty \mid X_0=i)=0$$

In this case, for any $k\in S$,

$$\mathbb{P}(N(i)=\infty \mid X_0=k)=0,$$

$$\sum_{n=1}^{\infty} p_n(k,i) = \frac{f(k,i)}{1-\cancel{f(k,i)}\, f(i,i)}$$

which is finite.

Note that the theorem particularly indicates that

$$\mathbb{P}(N(i)=\infty \mid \cancel{X_0=i}) = \begin{cases} 1 & \text{if } i \text{ is recurrent,} \\ 0 & \text{if } i \text{ is transient.} \end{cases}$$

There are at least two different proofs of the theorem. One of them is based on the generating functions

$$P_{ij}(s) = \sum_{n=0}^{\infty} s^n p_n(i,j)$$

$$F_{ij}(s) = \sum_{n=0}^{\infty} s^n f_n(i,j),$$

where $f_0(i,j)=0$, $i,j\in S$, and $s\in(-1,1)$. They satisfy

$$P_{ij}(s) = \delta_{ij} + F_{ij}(s)\, P_{jj}(s), \quad \forall i,j\in S,\ \forall s\in(-1,1).$$

This, together with Abel's Theorem on the convergence of power series at the end points of the interval of convergence interval, will leads the desired result. See [GS].

A different proof presented in [HPS] is more of a probabilistic proof, using some relations between the hitting times and total number of visiting times, e.g. the formula in Proposition on page 10, and also the expected number of visits.

<u>Corollary</u> If $i \in S$ is a transient state, then

$$\lim_{n\to\infty} p_n(k,i) = 0 \quad \forall k \in S.$$

<u>Corollary</u> If S is finite, there there exists at least one recurrent state.

<u>Definition</u> The mean recurrence time μ_i of a state $i \in S$ is

$$\mu_i = \mathbb{E}(T_i \mid X_0 = i) = \begin{cases} \sum_{n=1}^{\infty} n f_n(i,i) & \text{if } i \text{ is recurrent,} \\ \infty & \text{if } i \text{ is transient.} \end{cases}$$

Note that μ_i can be infinite even it is a recurrent state. Note also that, if i is transient, then

$$\mathbb{P}(T_i = \infty \mid X_0 = i) = 1 - \mathbb{P}(T_i < \infty \mid X_0 = i) > 0.$$

Hence, $\mathbb{E}(T_i \mid X_0 = i) = \infty$.

<u>Definition</u> A recurrent state $i \in S$ is a null recurrent state if $\mu_i = \infty$, and a positive recurrent state if $\mu_i < \infty$.

<u>Theorem</u> Let $i \in S$ be a recurrent state.

(1) i is null $\iff \lim_{n\to\infty} p_n(i,i) = 0$

In this case, $\lim_{n\to\infty} p_n(j,i) = 0 \quad \forall j \in S$.

(2) If i is positive recurrent, then

$$\lim_{n\to\infty} \frac{N_n(i)}{n} = \frac{\mathbb{P}\,\mathbb{1}_{\{T_i<\infty\}}}{\mu_i} \quad \text{with probability } 1.$$

where $N_n(i) = \sum_{k=1}^{n} \mathbb{1}_i(X_n)$, and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_k(j,i) = \frac{\mathbb{P}(T_i<\infty \mid X_0=j)}{\mu_i} \quad \forall j \in S.$$

(When $j=i$, the right-hand side is just $\frac{1}{\mu_i}$.)

Note that Part (2) above improves the result in a theorem on page $\boxed{12}$.

<u>Corollary</u> If S is finite, then any recurrent state must be positive recurrent.

We now consider states that communicate with each other.

<u>Definition</u> A state $i \in S$ <u>leads to</u> a state $j \in S$ if $\mathbb{P}(T_j<\infty \mid X_0=i) > 0$. We write $i \to j$.

<u>Proposition</u> $i \to j \iff \exists\, n \geq 1$ such that $p_n(i,j) > 0$.

<u>Theorem</u> (1) If $i \in S$ is recurrent and $i \to j \in S$, then j is also recurrent. Moreover, $\mathbb{P}(T_j<\infty \mid X_0=i) = \mathbb{P}(T_i<\infty \mid X_0=j) = 1$, and $j \to i$.

(2) If $i \in S$ is null recurrent (resp. positive recurrent) and $i \to j$, then j is also null recurrent (resp. positive recurrent), and $j \to i$.

See [HPS] for a proof of this theorem.

By the previous theorem, we immediately have: $\forall i, j \in S$
If $i \leftrightarrow j$, then: (1) i is recurrence $\Leftrightarrow$ j is. ~~Mor~~ Moreover, i is null rec. (positive) $\Leftrightarrow$ j is null rec. (positive)
(2) i is transient $\Leftrightarrow$ j is.

<u>Definition</u> Two states $i, j \in S$ <u>communicate</u> with each other, if $i \to j$ and $j \to i$, written $i \leftrightarrow j$.

<u>Proposition</u> The relation $\leftrightarrow$ is an equivalent relation, i.e., it satisfies the following:

(1) $i \leftrightarrow i \quad \forall i \in S$;

(2) $i \leftrightarrow j \Longrightarrow j \leftrightarrow i \quad \forall i, j \in S$;

(3) $i \leftrightarrow j,\ j \leftrightarrow k \Longrightarrow i \leftrightarrow k \quad \forall i, j, k \in S$.

<u>Definition</u> The equivalent relation $\leftrightarrow$ classifies S into disjoint equivalent classes, called <u>communication classes</u>.

In each communication class, any two states communicate with each other. States in different communication classes will not communicate with each other. But, one way "communication" is still possible. For instance, if $i \in S$ is transient and $j \in S$ is recurrent, then it is possible that $i \to j$. But: $j \to i$ is impossible.

Move this to somewhere after defining period(i)

<u>Definition</u> A nonempty subset of states $C \subseteq S$ is <u>closed</u>, if

$$p(i,j) = 0 \quad \forall i \in C \quad \forall j \in S \setminus C.$$

Move this part to page 15

Proposition If S is finite, then any recurrent state is positive recurrent.

Proof. Use Part (1) of Theorem 1 on page 15. Consider a recurrent communication class.

We now describe another characterist property of a state, other than recurrence and transience.

Definition The *period* $d(i)$ of a state $i \in S$ is

$$d(i) = \text{g.c.d.}\{n \geq 0 : p_n(i,i) > 0\},$$

where g.c.d. = greatest common division.

If $d(i) > 1$, call i a *periodic* state.

If $d(i) = 1$, call i an *aperiodic* state.

Clearly $1 \leq d(i) \leq \min\{n \geq 1 : p_n(i,i) > 0\}$.

Hence, $\boxed{p(i,i) > 0 \implies d(i) = 1}$

Definition A state $i \in S$ is *ergodic* if it is positively recurrent and aperiodic.

Theorem If $i \in S$ is aperiodic, then $\lim_{n\to\infty} p_n(i,i) = \frac{1}{\mu_i}$, and $\lim_{n\to\infty} p_n(j,i) = \frac{f(j,i)}{\mu_i}$, $\forall j \in S$. In particular, $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} p_k(j,i) = \frac{f(j,i)}{\mu_i}$ $\forall j \in S$.

Proposition Let $i, j \in S$. If $i \leftrightarrow j$ then $d(i) = d(j)$.

Therefore, we can define the period of a communication class of states as that of any state in the class.

A summary

Theorem Let $i, j \in S$ be such that $i \leftrightarrow j$. Then

(1) $d(i) = d(j)$

(2) i is transient $\iff$ j is.

(3) i is rec. $\iff$ j is

(4) i is null rec. $\iff$ j is.

(5) i is positive rec. $\iff$ j is.

Definition A nonempty subset of states $C \subseteq S$ is *closed*, if

$$p(i,j) = 0 \quad \forall i \in C \quad \forall j \in S \setminus C.$$

Proposition Let $\phi \neq C \subseteq S$. If C is closed ~~if $\exists n \geq 1$~~ then ~~such that~~

$$p_n(i,j) = 0 \quad \forall i \in C, \forall j \in S \setminus C, \forall n \geq 1.$$

<u>Definition</u> A nonempty subset of states, $C \subseteq S$, is <u>irreducible</u>, if $i \leftrightarrow j$ for any $i, j \in C$.

If S is itself irreducible, then the chain $\{X_n\}_{n=0}^{\infty}$ is <u>irreducible</u>.

Clearly, any communication class of states is irreducible. If there exists only one communication class, then the chain is irreducible.

<u>Decomposition Theorem</u> The state space S is the union of disjoint subsets of states

$$S = T \cup \left(\bigcup_{i \geq 1} R_i \right),$$

where T is the set of all transient states, and each R_i is a recurrent communication class. There are finitely many or countably infinitely many R_i's.

Note. Among those recurrent classes, some of them can be null, i.e., states in such a class are null recurrent. Some of them can be positive.

For a finite S, there exists at least an R_i.

In general, $T = \phi$ is possible. If $T \neq \phi$, T itself may not be a communication class. But, we have

$$T = \bigcup_{j \geq 1} T_j$$

with each T_j a communication class of transient states, and $T_i \cap T_j = \phi$ if $i \neq j$.

If the chain starts with a state in some recurrent class R_j, then the chain will stay in R_j. This means each recurrent class defines itself a (sub) state space.

If the chain starts at a transient state then the ~~chain~~ chain may never leave the set of all transient states T, or it may eventually get into a recurrent class and ~~stay~~ stays there forever. If S is finite, then the first situation will not occur.

(C) Invariant Distributions and Large-Time Behavior

<u>Definition</u> A probability vector $\pi = (\pi(i))_{i \in S}$ is an <u>invariant</u> (or <u>stationary</u>) distribution (for the transition matrix P) if $\pi = \pi P$, i.e., $\forall j \in S$

$$\pi(j) = \sum_{i \in S} \pi(i)\, p(i,j).$$

<u>Remarks</u>

(1) Motivation. If for some initial distribution π_0, we have

$$\lim_{n \to \infty} \pi_n = \lim_{n \to \infty} \pi_0 P^n = \pi$$

for some probability vector π, where $\pi_n = \pi_0 P^n$ is the distribution of X_n, then

$$\pi = \lim_{n \to \infty} \pi_{n+1} = \lim_{n \to \infty} \pi_0 P^{n+1} = \lim_{n \to \infty} (\pi_0 P^n) P = \pi P.$$

Hence the limit π is invariant.

(2) $\pi = \pi P$ means that π is a left eigenvector of P corresponding to the eigenvalue 1 of P. Recall that $\begin{bmatrix} 1 \\ 1 \\ \vdots \end{bmatrix}$ is a (right) eigenvector of P corresponding to the eigenvalue 1.

<u>Proposition</u> Let π be an invariant distribution. Then

$$\lim_{n\to\infty} \pi_0 P^n = \pi$$

for any probability vector π_0, if and only if

$$\lim_{n\to\infty} P^n = \begin{bmatrix} \pi \\ \pi \\ \vdots \end{bmatrix}$$

i.e., $\lim_{n\to\infty} p_n(i,j) = \pi(j) \quad \forall i,j \in S$.

In this case, the invariant distribution is unique. Moreover

$$\lim_{n\to\infty} \mathbb{P}(X_n = i) = \pi(i) \quad \forall i \in S.$$

<u>Proof</u> Suppose $\pi_0 P^n \to \pi$ as $n\to\infty$ for any probability vector π_0. Let $i,j \in S$. Choose $\pi_0 = (\pi_0(k))_{k\in S}$ by $\pi(k) = \delta_{ki}$. Then, $\pi_0 P^n \to \pi$ implies

$$\sum_{k\in S} \pi_0(k)\, p_n(k,j) \longrightarrow \pi(j) \quad \forall j \in S.$$

But $\sum_{k\in S} \pi_0(k)\, p_n(k,j) = p_n(i,j)$. Hence, $p_n(i,j) \to \pi(j)$ as $n\to\infty$ for any j.

Conversely, assume that $p_n(i,j) \to \pi(j)$ $\forall i,j \in S$. If $\pi_0 = (\pi_0(k))_{k\in S}$ is any probability vector then

$$(\pi_0 P^n)(i) = \sum_{k\in S} \pi_0(k)\, p_n(k,i) \longrightarrow \sum_{k\in S} \pi_0(k)\, \pi(i)$$
$$= \pi(i) \sum_{k\in S} \pi_0(k) = \pi(i), \quad \forall i \in S.$$

[Note that the convergence here can be justified.]

If $\tilde{\pi}$ is also an invariant distribution, then

$$\tilde{\pi} = \tilde{\pi} P = \tilde{\pi} P^2 = \cdots = \tilde{\pi} P^n \longrightarrow \pi .$$

Hence, $\tilde{\pi} = \pi$. Moreover,

$$\mathbb{P}(X_n = i) = \pi_n(i) = (\pi_0 P^n)(i) \longrightarrow \pi(i) \quad \forall i \in S. \quad \square$$

<u>Theorem</u> Let π be an invariant distribution. If $i \in S$ is a transition state or a null recurrent state, then $\pi(i) = 0$.

This follows from the fact $\lim_{n\to\infty} p_n(j,i) = 0$ and $\pi(i) = (\pi P^n)(i)$
$= \sum_{j\in S} \pi(j)\, p_n(j,i) \to 0$.

<u>Definition</u> A stochastic matrix P (or order $|S| \times |S|$) and a probability vector of order $|S|$ (i.e., in $\mathbb{R}^{|S|}$) satisfy the <u>detailed balance</u> if

$$\pi(i)\, p(i,j) = \pi(j)\, p(j,i) \quad \forall i \in S.$$

<u>Theorem</u> If $\pi = (\pi(i))_{i\in S}$ is a probability vector and the transition matrix P and π satisfy the detailed balance, then π is an invariant distribution.

<u>Proof</u> We have for any $j \in S$ that

$$(\pi P)(j) = \sum_{i\in S} \pi(i)\, p(i,j) = \sum_{i\in S} \pi(j)\, p(j,i) = \pi(j) \sum_{i\in S} p(j,i) = \pi(j) \quad \square$$

<u>Theorem</u> An irreducible chain has an invariant distribution $\pi \iff$ all the states are positively recurrent. In this case, π is the unique invariant distribution given by

$$\pi(i) = \frac{1}{\mu_i} > 0 \quad \forall i \in S.$$

[In particular, $\sum_{i\in S} \frac{1}{\mu_i} = 1$.]

<u>Theorem</u> For an irreducible aperiodic chain, we have that

$$\lim_{n\to\infty} p_n(i,j) = \frac{1}{\mu_j} \qquad \forall i,j \in S.$$

If, in addition, the chain is positive recurrent, then the vector $\pi \in \mathbb{R}^{|S|}$, defined by $\pi(i) = \frac{1}{\mu_i}$ $\forall i \in S$, is the unique invariant distribution, and

$$\lim_{n\to\infty} \pi_n(i) = \pi(i) \qquad \forall i \in S$$

where $\pi_n(i) = \mathbb{P}(X_n = i)$.

Note that the convergence $\pi_n(i) \to \pi(i) = \frac{1}{\mu_i}$ is independent of an initial distribution π_0:

$$\begin{aligned}
\pi_n(i) = \mathbb{P}(X_n = i) &= \sum_{j\in S} \mathbb{P}(X_0 = j)\,\mathbb{P}(X_n = i \mid X_0 = j) \\
&= \sum_{j\in S} \mathbb{P}(X_0 = j)\, p_n(j,i) \\
&\longrightarrow \sum_{j\in S} \mathbb{P}(X_0 = j)\, \frac{1}{\mu_i} \\
&= \pi(i) \sum_{j\in S} \mathbb{P}(X_0 = j) \\
&= \pi(i).
\end{aligned}$$

<u>Theorem</u> Suppose the chain is irreducible, recurrent, and periodic with period $d > 1$. Let $i, j \in S$. Then $\exists$ integer r: $0 \le r < d$, such that

$$p_n(i,j) = 0 \quad \text{if } n \geq 1 \text{ and } n \not\equiv r \pmod{d}$$

and

$$\lim_{m\to\infty} p_{md+r}(i,j) = d\,\pi(j).$$

(D) Finite-State Markov Chains

Assume in this part $|S| < \infty$. In particular, $S = \{1, 2, \cdots, N\}$ for some integer $N \geq 1$.

<u>Theorem</u> Let $P = (p(i,j))_{N \times N}$ be a stochatic matrix. Suppose:

(1) 1 is a simple eigenvalue of P and all other eigenvalues of P have absolute values less than 1;

(2) There exists a left eigenvector of P corresponding to the eigenvalue 1 that has nonnegative components.

Then P has a unique invariant distribution $\pi \geq 0$, and for any initial distribution π_0,

$$\lim_{n \to \infty} \pi_0 P^n = \pi, \quad \text{i.e.,} \quad \lim_{n \to \infty} \pi_n = \pi,$$

where π_n is the distribution of X_n $(n \geq 0)$.

<u>Proof</u> We have the Jordan decomposition $P = QDQ^{-1}$, where $D = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & M & \\ 0 & & & \end{bmatrix}$, $Q^{-1} = \begin{bmatrix} a_1 \\ \vdots \\ a_N \end{bmatrix}$ (a_j: row vectors),

$Q = [\vec{1}, b_2, \cdots, b_N]$ $\left(\vec{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^N, \ b_j: \text{column vectors}\right)$.

From $DQ^{-1} = Q^{-1}P$, we can get $a_1 = a_1 P$.

$D^n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & M^n & \\ 0 & & & \end{bmatrix}$. M consists of Jordan blocks. But, all eigenvalues of P have absolute values < 1 except 1, which is simple. So, $M^n \to 0$. Thus $P^n = QD^nQ^{-1} \to \begin{bmatrix} a_1 \\ a_1 \\ \vdots \\ a_1 \end{bmatrix}$

If necessary, normalize a_1 so that $a_1 \geq 0$, $\sum_{i=1}^{N} a_1(i) = 1$.

Thus $\pi := a_1$ is an invariant distribution. It is unique and $\pi_n \to \pi$ by a previous ~~Thm~~ proposition. Cf. page 20. $\square$

<u>Theorem</u> If S is finite, and there exists $n \geq 1$ such that $p_n(i,j) > 0$ for all $i,j \in S$. Then (1) and (2) in the above theorem holds true, and hence the conclusions of the above theorem follow.

<u>Proof</u> This follows from the Perron-Frobenius Theorem. $\square$

Note that, for a finite-state chain, there exists at least one recurrent state, and any recurrent state is positively recurrent. Therefore, if a finite-state chain is irreducible, all states are positively recurrent.

<u>Theorem</u> If a finite-state chain is irreducible and aperiodic, then $\exists\, n \geq 1$ such that $p_n(i,j) > 0$ $\forall i,j \in S$. Hence, $\exists!$ invariant distribution $\pi > 0$ determined by $\pi(i) = \frac{1}{\mu_i}$ $\forall i \in S$ and satisfies $\lim_{n \to \infty} \pi_0 P^n = \pi$ $\forall$ initial distribution π_0.

<u>Proof</u> Let $i,j \in S$. Since P is irreducible, $\exists\, m = m(i,j)$ s.t. $p_m(i,j) > 0$. Since P is aperiodic, $\exists\, M(i)$ for $i \in S$ such that $p_n(i,i) > 0$ if $n \geq M(i)$. Hence, if $n \geq M(i)$ then $p_{n+m}(i,j) \geq p_n(i,i)\, p_m(i,j) > 0$. Let M be the maximum value of $M(i) + m(i,j)$ over all $i,j \in S$. The maximum exists, since S is finite.

Then $p_n(i,j) > 0$ if $n \geq M$, $\forall i,j \in S$. Then, apply the previous theorem. $\square$

<u>Theorem</u> If ~~the tran~~ S is finite, and P is symmetric ($p(i,j) = p(j,i)$ $\forall i,j \in S$) and irreducible, then the invariant distribution is uniform: $\pi(i) = \frac{1}{|S|}$ $\forall i \in S$.

Brief discussion on the case of a reducible and/or periodic finite-state chain.

Suppose the state S is finite. If the chain is reducible, then there exist disjoint recurrent communication classes $R_1, \cdots, R_r$. Each R_k is irreducible and (positively) recurrent. Thus, for each R_k, there exists an invariant distribution $\pi^{(k)}$ which is concentrated on R_k, i.e., $\pi^{(k)}(i) = 0$ if $i \notin R_k$. The eigenvalue 1 of the transition matrix P has r linearly independent eigenvectors, & no more than r eigenvectors of the eigenvalue 1 are linearly independent &. i.e., the multiplicity of 1 is r. One eigenvector for each R_k. Suppose each R_k is also aperiodic. Then

$$\lim_{n\to\infty} p_n(i,j) = \pi^{(k)}(j) \quad \forall j \in R_k,$$

$$p_n(i,j) = 0 \quad \forall j \notin R_k.$$

If $i \in S$ is a transient state, then the chain starting at i eventually ends up in a recurrent class, i.e.,

$$\lim_{n\to\infty} p_n(i,j) = 0 \quad \text{if } j \text{ is transient.}$$

Let ~~α_k~~ $\alpha_k(i)$ ($1 \leq k \leq r$) be the probability that the chain starting at i eventually ends up in R_k. Then

$$\lim_{n\to\infty} p_n(i,j) = \alpha_k(i)\pi^{(k)}(j) \quad \forall j \in R_k.$$

The limit $\pi_0 P^n$ exists but may depend on π_0.

Suppose now the transition matrix P for a finite-state chain ~~&~~ is irreducible, but periodic with period $d>1$. Then P has exactly d eigenvalues with absolute values equal to 1. These are d roots of the equation $z^d=1$. They ~~and except 1, the others~~ are complex ~~but not real~~ numbers. Each of such eigenvalues is simple.

In this case, there exists a unique invariant distribution π. For any initial distribution π_0,

$$\lim_{n\to\infty} \frac{1}{d}\pi_0\left(P^{n+1}+\cdots+P^{n+d}\right)=\pi.$$