CONCISE NOTES FOR MTH6141 RANDOM PROCESSES

MARK JERRUM

- What is in these notes: definitions, theorems, proofs, methods.
- What is not in these notes: examples, motivation, chit-chat.

1. Discrete-time Markov chains

Definition 1.1. A *stochastic process* is a collection of random variables (r.v's) $\{X_t : t \in T\}$ indexed by a set T (usually thought of as "time"). The X_t take values in some *state space* S.

In the first half of the course we take $T = \mathbb{N}$, so the stochastic process is $X_0, X_1, X_2, \ldots$. This is a *discrete-time* process. Later we consider continuous time $T = \mathbb{R}_{\geq 0} = \{x \in \mathbb{R} : x \geq 0\}$.

1.1. The Markov property.

Definition 1.2. A stochastic process $X_0, X_1, X_2, \ldots$ on a (finite or countably infinite) state space S is a *Markov chain* if it satisfies, for all $t \in \mathbb{N}$,

$$\mathbb{P}(X_{t+1} = s \mid X_t = s_t, X_{t-1} = s_{t-1}, \ldots, X_0 = s_0) = \mathbb{P}(X_{t+1} = s \mid X_t = s_t),$$

for all $s, s_t, s_{t-1}, \ldots, s_0 \in S$ for which the conditional probabilities are defined.

This is called the *Markov property.* Informally: "The future conditioned on the present does not depend on the past."

If $\mathbb{P}(X_{t+1} = j \mid X_t = i)$ is independent of t then we say that the Markov chain is *homogeneous.* (We only consider homogeneous Markov chains in this course.) Then we write $\mathbb{P}(X_{t+1} = j \mid X_t = i) = p_{ij}$ and call p_{ij} the *transition probabilities.*

Lemma 1.1. *If $X_0, X_1, X_2, \ldots$ is a Markov chain with $\mathbb{P}(X_0 = i) = \mu_i$ for $i \in S$, then*

$$\mathbb{P}(X_0 = s_0, X_1 = s_1, X_2 = s_2, \ldots, X_t = s_t) = \mu_{s_0} p_{s_0 s_1} p_{s_1 s_2} \ldots p_{s_{t-1} s_t},$$

for all $s_0, s_1, \ldots, s_t \in S$.

These notes are condensed and adapted from lecture notes of Dr Robert Johnson.

Proof.

$$\begin{aligned}
&\mathbb{P}(X_0 = s_0, X_1 = s_1, \dots, X_t = s_t) \\
&\quad = \mathbb{P}(X_0 = s_0) \times \mathbb{P}(X_1 = s_1 \mid X_0 = s_0) \times \mathbb{P}(X_2 = s_2 \mid X_1 = s_1, X_0 = s_0) \\
&\qquad \times \cdots \times \mathbb{P}(X_t = s_t \mid X_{t-1} = s_{t-1}, \dots, X_0 = s_0) \\
&\quad = \mu_{s_0}\, \mathbb{P}(X_1 = s_1 \mid X_0 = s_0)\, \mathbb{P}(X_2 = s_2 \mid X_1 = s_1) \cdots \mathbb{P}(X_t = s_t \mid X_{t-1} = s_{t-1}) \\
&\qquad \text{(by the Markov property)} \\
&\quad = \mu_{s_0} p_{s_0 s_1} p_{s_1 s_2} \cdots p_{s_{t-1} s_t} \\
&\qquad \text{(by definition of } p_{ij}\text{)}.
\end{aligned}$$

□

A convenient means to visualise Markov chains is by *transition graphs.* The vertex set of a transition graph of a Markov chain is the set of states S. For each pair or states i, j for which the transition probability p_{ij} is non-zero, there is a directed edge (arc) from vertex i to vertex j labelled p_{ij}.

1.2. **Transition matrices.** The events $X_t = s$, as s ranges over S, partition the sample space. Thus

$$\sum_{j \in S} \mathbb{P}(X_t = j) = 1$$

and

$$\sum_{j \in S} \mathbb{P}(X_t = j \mid X_{t-1} = i) = 1,$$

for all $i \in S$, i.e.,

$$\sum_{j \in S} p_{ij} = 1.$$

The *transition matrix* is the matrix P with rows and columns indexed by S and with ij th entry p_{ij}. We have just seen that it is *stochastic*, i.e., that all row sums are 1. For convenience in numbering the rows and columns of the matrix P, we usually take $S = \{1, 2, \dots, n\}$.

The *r-step transition probabilities* are

$$p_{ij}^{(r)} = \mathbb{P}(X_{t+r} = j \mid X_t = i).$$

(These are independent of t as the chain is homogeneous.) Note that

$$p_{ij}^{(0)} = \begin{cases} 1, & \text{if } i = j; \\ 0 & \text{otherwise} \end{cases} \qquad \text{and} \qquad p_{ij}^{(1)} = p_{ij}.$$

Theorem 1.2 (Chapman-Kolmogorov relations)**.**

$$p_{ij}^{(r+s)} = \sum_{k \in S} p_{ik}^{(r)} p_{kj}^{(s)}, \quad \textit{for all } i, j \in S \textit{ and } r, s \geq 0.$$

Proof.

$$\begin{aligned}
p_{ij}^{(r+s)} &= \mathbb{P}(X_{r+s} = j \mid X_0 = i) \\
&= \sum_{k\in S} \mathbb{P}(X_{r+s} = j, X_r = k \mid X_0 = i) \\
&\qquad \text{(since the events } X_r = k \text{ partition the sample space)} \\
&= \sum_{k\in S} \mathbb{P}(X_r = k \mid X_0 = i)\ \mathbb{P}(X_{r+s} = j \mid X_r = k, X_0 = i) \\
&= \sum_{k\in S} \mathbb{P}(X_r = k \mid X_0 = i)\ \mathbb{P}(X_{r+s} = j \mid X_r = k) \\
&\qquad \text{(by the Markov property)} \\
&= \sum_{k\in S} p_{ik}^{(r)} p_{ik}^{(s)}.
\end{aligned}$$

□

Corollary 1.3. *If P is the transition matrix of a Markov chain, then P^r is the matrix of r-step transition probabilities, i.e.,*

$$P^r = \begin{pmatrix} p_{11}^{(r)} & p_{12}^{(r)} & \cdots & p_{1n}^{(r)} \\ p_{21}^{(r)} & p_{22}^{(r)} & \cdots & p_{2n}^{(r)} \\ \vdots & \vdots & & \vdots \\ p_{n1}^{(r)} & p_{n2}^{(r)} & \cdots & p_{nn}^{(r)} \end{pmatrix}.$$

Proof. The proof is by induction on r, with the base case $r = 1$ being immediate. If $r > 1$ then

$$\begin{aligned}
(P^r)_{ij} &= (P^{r-1}P)_{ij} \\
&= \sum_{k\in S} (P^{r-1})_{ik} P_{kj} \\
&= \sum_{k\in S} p_{ik}^{(r-1)} p_{kj}^{(1)} && \text{(by inductive hypothesis)} \\
&= p_{ij}^{(r)} && \text{(by Theorem 1.2).}
\end{aligned}$$

□

If the matrix P is diagonalisable, i.e., there is an invertible matrix M and a diagonal matrix D such that $P = MDM^{-1}$, then

$$P^r = \underbrace{(MDM^{-1})(MDM^{-1})\cdots(MDM^{-1})}_{r \text{ copies}} = MD^rM^{-1},$$

which provides an easy way to compute P^r.

1.3. **First-step analysis.** A state is *absorbing* if $p_{ii} = 1$; when we reach an absorbing state we never leave it.

Several quantities — what is the probability of reaching a certain absorbing state? what is the expected time to absorption? — can be computed by conditioning on the first step. The computation reduces to solving simultaneous linear equations.

Theorem 1.4. *Suppose (X_t) is a Markov chain with state space S and absorbing states $A \subseteq S$, and suppose $w : S \setminus A \to \mathbb{R}$ is a weight function on non-absorbing states. Let $T = \min\{t : X_t \in A\}$ and $W = \sum_{t=0}^{T-1} w(X_t)$.*

(1) *Fix $k \in A$. Let $a_i = \mathbb{P}(X_T = k \mid X_0 = i)$. Then the a_i satisfy*

$$a_i = \begin{cases} 1, & \text{if } i = k; \\ 0, & \text{if } i \in A \setminus \{k\}; \\ p_{ik} + \sum_{j \in S \setminus A} p_{ij} a_j & \text{if } i \in S \setminus A. \end{cases}$$

(2) *Let $w_i = \mathbb{E}(W \mid X_0 = i)$. Then the w_i satisfy*

$$w_i = \begin{cases} 0, & \text{if } i \in A; \\ w(i) + \sum_{j \in S \setminus A} p_{ij} w_j & \text{if } i \notin A. \end{cases}$$

(3) *If S is finite and from every state it is possible to reach an absorbing state then the above systems of equations have unique solutions.*

Proof. (1) If $i \in A$ (i.e., we start in an absorbing state) then $T = 0$; thus $a_k = \mathbb{P}(X_T = k \mid X_0 = k) = 1$, and if $i \in A \setminus \{k\}$ then $a_i = \mathbb{P}(X_T = k \mid X_0 = i) = 0$. Otherwise (i.e., if $i \notin A$), conditioning on the first step:

$$\begin{aligned} a_i &= \sum_{j \in S} \mathbb{P}(X_1 = j \mid X_0 = i)\, \mathbb{P}(X_T = k \mid X_1 = j) \\ &\qquad \text{(Law of Total Probability)} \\ &= \sum_{j \in S} p_{ij} a_j \\ &= p_{ik} + \sum_{j \in S \setminus A} p_{ij} a_j \end{aligned}$$

(2) If $i \in A$ then $T = 0$ and $w_i = 0$. Otherwise, conditioning on the first step:

$$\begin{aligned} w_i &= w(i) + \sum_{j \in S} \mathbb{P}(X_1 = j \mid X_0 = i)\, \mathbb{E}\Big(\sum_{t=1}^{T-1} w(X_t) \mid X_1 = j \Big) \\ &\qquad \text{(Law of Total Expectation)} \\ &= w(i) + \sum_{j \in S} p_{ij} w_j \\ &= w(i) + \sum_{j \in S \setminus A} p_{ij} w_j. \end{aligned}$$

(3) Outside the scope of the module.

□

1.4. **Long-term behaviour.** A *probability vector* is a row vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)$ with $\mu_i \geq 0$ and $\sum_{i=1}^n \mu_i = 1$. Let $X_0, X_1, \ldots$ be a Markov chain with $S = \{1, 2, \ldots, n\}$. Let the distribution of X_0 be given by a probability vector $\boldsymbol{\mu}^{(0)} = (\mu_1^{(0)}, \mu_2^{(0)}, \ldots, \mu_n^{(0)})$ (i.e., $\mathbb{P}(X_0 = k) = \mu_k^{(0)}$), the *initial distribution.* Let $\boldsymbol{\mu}^{(t)}$ be the distribution of X_t (i.e., $\mathbb{P}(X_t = k) = \mu_k^{(t)}$).

Lemma 1.5. *With the above definitions,* $\boldsymbol{\mu}^{(t)} = \boldsymbol{\mu}^{(0)} P^t$.

Proof.

$$\mu_k^{(t)} = \mathbb{P}(X_t = k) = \sum_{i=1}^n \mathbb{P}(X_t = k \mid X_0 = i)\, \mathbb{P}(X_0 = i) = \sum_{i=1}^n \mu_i^{(0)} p_{ik}^{(t)}.$$

Now appeal to Corollary 1.3. □

1.4.1. *The general two-state chain.* Suppose $S = \{1, 2\}$ and

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$

Boring cases are

- $\alpha = \beta = 0$, when $P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $P^t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, and $\lim_{t\to\infty} P^t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$; and
- $\alpha = \beta = 1$, when $P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $P^t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ if t is even, and $P^t = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ if t is odd. In this case $\lim_{n\to\infty} P^t$ does not exist.

The interesting case is when $0 < \alpha + \beta < 2$. Then P has eigenvalue 1 with eigenvector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, and $\lambda = 1 - \alpha - \beta$ with eigenvector $\begin{pmatrix} \alpha \\ -\beta \end{pmatrix}$. Thus

$$P = \begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{1}{\alpha+\beta} & \frac{-1}{\alpha+\beta} \end{pmatrix},$$

and so

$$\begin{aligned} P^t &= \frac{1}{\alpha+\beta} \begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \lambda^t \end{pmatrix} \begin{pmatrix} \beta & \alpha \\ 1 & -1 \end{pmatrix} \\ &= \frac{1}{\alpha+\beta} \begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix} \begin{pmatrix} \beta & \alpha \\ \lambda^t & -\lambda^t \end{pmatrix} \\ &= \frac{1}{\alpha+\beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} + \frac{\lambda^t}{\alpha+\beta} \begin{pmatrix} \alpha & -\alpha \\ -\beta & \beta \end{pmatrix}. \end{aligned}$$

Since $-1 < \lambda < 1$,

$$\lim_{t\to\infty} P^t = \frac{1}{\alpha+\beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} = W,$$

say, and

$$\lim_{t\to\infty} \boldsymbol{\mu}^{(t)} = \lim_{t\to\infty} \boldsymbol{\mu}^{(0)} P^t = \boldsymbol{\mu}^{(0)} W = \left(\tfrac{\beta}{\alpha+\beta}, \tfrac{\alpha}{\alpha+\beta}\right),$$

for any initial distribution $\boldsymbol{\mu}^{(0)}$. Note that, in the limit as $t \to \infty$, the distribution of X_t does not depend on the initial distribution of X_0: this is because $\lim_{t\to\infty} P^t$ exists (which is not so when $\alpha = \beta = 1$) and has equal rows (which is not so when $\alpha = \beta = 0$).

Definition 1.3. A probability vector $\boldsymbol{w} = (w_1, \dots, w_n)$ is a *limiting distribution* for P if

$$P^t \to \begin{pmatrix} w_1 & w_2 & \dots & w_n \\ w_1 & w_2 & \dots & w_n \\ \vdots & & \vdots & \\ w_1 & w_2 & \dots & w_n \end{pmatrix}, \quad \text{as } t \to \infty.$$

It is an *equilibrium distribution* if $\boldsymbol{w}P = \boldsymbol{w}$.

For the 2-state example,

- If $\alpha = \beta = 0$ then any vector $(w, 1-w)$ is an equilibrium distribution, and there is no limiting distribution.
- If $\alpha = \beta = 1$ them $(\frac{1}{2}, \frac{1}{2})$ is the unique equilibrium distribution, but there is no limiting distribution.
- If $0 < \alpha + \beta < 2$ then $(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta})$ is the unique equilibrium distribution and the limiting distribution.

Definition 1.4. A Markov chain is *irreducible* if, for all $i, j \in S$, there is a $k \geq 0$ with $p_{ij}^{(k)} > 0$; it is *regular* if there is a $k \geq 0$ with $p_{ij}^{(k)} > 0$ for all $i, j \in S$.

Theorem 1.6. *Let P be the transition matrix of a regular Markov chain on a finite state space. Then P has a limiting distribution $\boldsymbol{w}$. Moreover, $\boldsymbol{w}$ is the unique equilibrium distribution for P.*

We will prove a slightly weaker statement.

Lemma 1.7. *Let $P = (p_{ij} : 1 \leq i, j \leq n)$ be the transition matrix of a Markov chain on a finite state space. Suppose that $p_{ij} > 0$ for all i, j. Then P has a limiting distribution $\boldsymbol{w}$. Moreover, $\boldsymbol{w}$ is the unique equilibrium distribution for P.*

Proof. Suppose $p_{ij} \geq \varepsilon > 0$ for all $1 \leq i, j \leq n$. Recall that $p_{ij}^{(t)}$ are the t-step transition probabilities of the Markov chain. Denote by $m_j^{(t)}$ and $M_j^{(t)}$ the maximum and minimum entries in the jth column of P^t. That is,

$$m_j^{(t)} = \min\{p_{ij}^{(t)} : 1 \leq i \leq n\}$$

and

$$M_j^{(t)} = \max\{p_{ij}^{(t)} : 1 \leq i \leq n\}$$

Fix j, and let k' and k'' be such that $p_{k'j}^{(t)} = m_j^{(t)}$ and $p_{k''j}^{(t)} = M_j^{(t)}$; that is, k' and k'' are positions in column j of P^t at which the minimum and maximum are attained.

$$\begin{aligned} p_{ij}^{(t+1)} &= \sum_{k=1}^{n} p_{ik} p_{kj}^{(t)} \\ &= \sum_{k \neq k'} p_{ik} p_{kj}^{(t)} + p_{ik'} p_{k'j}^{(t)} \\ &\leq \sum_{k \neq k'} p_{ik} M_j^{(t)} + p_{ik'} m_j^{(t)} \\ &= \sum_{k} p_{ik} M_j^{(t)} - p_{ik'} (M_j^{(t)} - m_j^{(t)}) \\ &\leq M_j^{(t)} - \varepsilon (M_j^{(t)} - m_j^{(t)}). \end{aligned}$$

Since this ineq holds for all i, he have

$$M_j^{(t+1)} \leq M_j^{(t)} - \varepsilon (M_j^{(t)} - m_j^{(t)}).$$

A similar calculation yields

$$m_j^{(t+1)} \geq m_j^{(t)} + \varepsilon (M_j^{(t)} - m_j^{(t)}).$$

Notice that the sequence $M_j^{(1)}, M_j^{(2)}, M_j^{(3)}, \ldots$ is monotonically decreasing and $m_j^{(1)}, m_j^{(2)}, m_j^{(3)}, \ldots$ monotonically increasing. So (Convergence and Continuity) both sequences converge to a limit.

Subtracting,

$$M_j^{(t+1)} - m_j^{(t+1)} \leq (1 - 2\varepsilon)(M_j^{(t)} - m_j^{(t)}),$$

So $(M_j^{(t)} - m_j^{(t)}) \to 0$ $(t \to \infty)$, and $m_j^{(t)}$ and $M_j^{(t)}$ tend to the same limit: call this limit w_j. Since the t-step transition probability $p_{ij}^{(t)}$ is bounded by $m_j^{(t)}$ and $M_j^{(t)}$ for all t, i.e.,

$$m_j^{(t)} \leq p_{ij}^{(t)} \leq M_j^{(t)},$$

it must also be the case that $p_{ij}^{(t)} \to w_j$ $(t \to \infty)$. Thus $\boldsymbol{w} = (w_1, \ldots, w_n)$ is a limiting distribution for P.

The distribution $\boldsymbol{w}$ is also the unique equilibrium distribution. On the one hand, $\boldsymbol{w}P^{t+1} \to \boldsymbol{w}$, as $t \to \infty$; on the other, $\boldsymbol{w}P^{t+1} = (\boldsymbol{w}P^t)P \to \boldsymbol{w}P$ $(t \to \infty)$. So $\boldsymbol{w}P = \boldsymbol{w}$, i.e., $\boldsymbol{w}$ is an equilibrium distribution for P. To see uniqueness, let $\boldsymbol{w}'$ be any equilibrium distribution for P. Then $\boldsymbol{w}'P^t = \boldsymbol{w}'$ for all $t \in \mathbb{N}$ and so $\boldsymbol{w}'P^t \to \boldsymbol{w}'$ as $t \to \infty$. However, $\boldsymbol{w}'P^t \to \boldsymbol{w}$ as $t \to \infty$ since $\boldsymbol{w}$ is the limiting distribution. Thus $\boldsymbol{w}' = \boldsymbol{w}$. □

Fix $k \in S$ and define the indicator r.v.

$$A_s = \begin{cases} 1, & \text{if } X_s = k; \\ 0, & \text{otherwise.} \end{cases}$$

Then $\mathbb{E}(A_s) = \mathbb{P}(X_s = k)$. The proportion of time spent in state k up to but not including time t is $\frac{1}{t}\sum_{s=0}^{t-1} A_s$. We are interested in the limit $\lim_{t\to\infty} \mathbb{E}\left(\frac{1}{t}\sum_{s=0}^{t-1} A_s\right)$, assuming it exists: this is the expected proportion of time spent in state k in the long term.

Lemma 1.8. *If $\boldsymbol{w}$ is the limiting distribution of P (see Definition 1.3) then*

$$\mathbb{E}\left(\frac{1}{t}\sum_{s=0}^{t-1} A_s\right) \to w_k, \quad \textit{as } t \to \infty.$$

Proof.

$$\mathbb{E}\left(\frac{1}{t}\sum_{s=0}^{t-1} A_s\right) = \frac{1}{t}\,\mathbb{E}\left(\sum_{s=0}^{t-1} A_s\right) = \frac{1}{t}\sum_{s=0}^{t-1}\mathbb{E}(A_s) = \frac{1}{t}\sum_{s=0}^{t-1} P(X_s = k).$$

But $\mathbb{P}(X_s = k) \to w_k$ as $s \to \infty$, so

$$\frac{1}{t}\sum_{s=0}^{t-1}\mathbb{P}(X_s = k) \to w_k, \quad \text{as } t \to \infty.$$

(If you took MTH5124 *Convergence and Continuity*, you could pause to verify this claim.) □

Theorem 1.9. *Any irreducible Markov chain on a finite state space S has a unique equilibrium distribution $\boldsymbol{w}$. Moreover,*

$$\mathbb{E}(\textit{proportion of time spent in state } k \textit{ up to time } t) \to w_k, \quad t \to \infty.$$

Proof. Let P be the transition matrix. Consider the transition matrix $Q = \frac{1}{2}(I + P)$. Since P is irreducible, there exists $t \in \mathbb{N}$ such that for all $i, j \in S$, $p_{ij}^s > 0$ for some $s \le t$. Now,

$$Q^t = \tfrac{1}{2}(I+P)^t = 2^{-t}\left[I + \binom{t}{1}P + \binom{t}{2}P^2 + \cdots + \binom{t}{t-1}P^{t-1} + P^t\right]$$

For each pair $i, j \in S$, there is some term (matrix) in the sum which has a non-zero entry in row i and column j. So Q^t has no non-zero entries, and Q is regular. By Theorem 1.6 there is a unique distribution $\boldsymbol{w}$ such that $\boldsymbol{w}Q = \boldsymbol{w}$, i.e., such that $\boldsymbol{w}(\frac{1}{2}P + \frac{1}{2}I) = \boldsymbol{w}$. It follows that $\frac{1}{2}\boldsymbol{w}P + \frac{1}{2}\boldsymbol{w} = \boldsymbol{w}$, and hence $\boldsymbol{w}P = \boldsymbol{w}$. In other words, $\boldsymbol{w}$ is an equilibrium distribution for P also. Conversely, any equilibrium distribution for P is an equilibrium distribution for Q also. So $\boldsymbol{w}$ is the unique equilibrium distribution for P.

The claim about the proportion of time spent in state k in the long term is beyond the scope of the module. □

Regular	⇒	Irreducible
⇓ (Thm 1.6)		⇓ (Thm 1.9)
Limiting distribution	⇒	! Stationary distribution
		⇓ (Thm 1.9)
		Long-run averages

FIGURE 1. Summary of the situation for finite-state Markov chains.

1.5. **Recurrence and Transience.** The definitions and theorems in the previous section were phrased for finite state spaces. We often want to consider (countably) infinite state spaces. If $S = \{0, 1, 2, \ldots\} = \mathbb{N}$ we say that $(w_0, w_1, w_2, \ldots)$ is an *equilibrium distribution* if $w_k = \sum_{i=0}^{\infty} w_i p_{ik}$ for all k, and $\sum_{i=0}^{\infty} w_i = 1$. We say that $(w_0, w_1, w_2, \ldots)$ is a *limiting distribution* if $\mathbb{P}(X_t = k \mid X_0 = i) \to w_k$ $(t \to \infty)$, for all $i, k \in \mathbb{N}$.

Definition 1.5. Suppose $a, b \in S$ are states. We say that *a communicates with b* (written $a \to b$) if $p_{ab}^{(t)}$ for some $t \geq 0$. We say that *a and b intercommunicate* (written $a \leftrightarrow b$) if $a \to b$ and $b \to a$. The *communicating classes* of the chain are the equivalence classes of S under $\leftrightarrow$.

Note (i) $a \leftrightarrow a$ for all $a \in S$, (ii) a Markov chain is irreducible if $a \leftrightarrow b$ for all $a, b \in S$.
Define

$$f_{ii}^{(t)} = \mathbb{P}(X_t = i, X_{t-1} \neq i, X_{t-2} \neq i, \ldots, X_1 \neq i \mid X_0 = i)$$

(informally, $f_{ii}^{(t)}$ is the probability that the first return to i occurs at time t). Then define $f_{ii} = \sum_{t=1}^{\infty} f_{ii}^{(t)}$ (informally, the probability of ever returning to i).

Definition 1.6. We say that a state i is *recurrent* (or persistent) if $f_{ii} = 1$ and *transient* otherwise (i.e., if $f_{ii} < 1$).

Lemma 1.10. *For all* $t \geq 1$, $p_{ii}^{(t)} = \sum_{s=1}^{t} f_{ii}^{(s)} p_{ii}^{(t-s)}$.

Proof. Condition on the time s of first return to state i:

$$\begin{aligned} p_{ii}^{(t)} &= \sum_{s=1}^{t} \mathbb{P}(X_1 \neq i, \ldots, X_{s-1} \neq i, X_s = i \mid X_0 = i)\, \mathbb{P}(X_t = i \mid X_s = i) \\ &= \sum_{s=1}^{t} f_{ii}^{(s)} p_{ii}^{(t-s)}. \end{aligned}$$

□

Thus, if we know the values $p_{ii}^{(t)}$, for $t = 1, 2, \ldots$, we can calculate the first return probabilities $p_{ii}^{(t)}$ as follows:

$$\begin{aligned} f_{ii}^{(1)} &= p_{ii}^{(1)} \\ f_{ii}^{(2)} + p_{ii}^{(1)} f_{ii}^{(1)} &= p_{ii}^{(2)}, \quad \text{i.e.,} \quad f_{ii}^{(2)} = p_{ii}^{(2)} - p_{ii}^{(1)} f_{ii}^{(1)} \\ f_{ii}^{(3)} + p_{ii}^{(1)} f_{ii}^{(2)} + p_{ii}^{(2)} f_{ii}^{(1)} &= p_{ii}^{(3)}, \quad \text{i.e.,} \quad f_{ii}^{(3)} = p_{ii}^{(3)} - p_{ii}^{(2)} f_{ii}^{(1)} - p_{ii}^{(1)} f_{ii}^{(2)}, \end{aligned}$$

etc.

Let M_i be the number of returns to i, given that $X_0 = i$, and write $M_i = \infty$ if the Markov chain returns infinitely often.

Lemma 1.11. (1) *If i is recurrent then $\mathbb{P}(M_i < \infty) = 0$ and $\mathbb{E}(M_i)$ is undefined.*
(2) *If i is transient then $\mathbb{P}(M_i < \infty) = 1$ and $\mathbb{E}(M_i) = f_{ii}/(1 - f_{ii})$.*

Proof. Observe that $\mathbb{P}(M_i = k) = f_{ii}^k (1 - f_{ii})$, since f_{ii}^k expresses the probability that we return k times to state i, and $(1 - f_{ii})$ that we never return again. So if i is transient then $f_{ii} < 1$ and $M_i + 1 \sim \text{Geom}(1 - f_{ii})$; in this case $\mathbb{P}(M_i < \infty) = 1$ and $\mathbb{E}(M_i) + 1 =$

$\mathbb{E}(M_i + 1) = 1/(1 - f_{ii})$, i.e., $\mathbb{E}(M_i) = f_{ii}/(1 - f_{ii})$. On the other hand, if state i is recurrent then $f_{ii} = 1$ and $\mathbb{P}(M_i = k) = f_{ii}^k(1 - f_{ii}) = 0$ for all $k \in \mathbb{N}$; in this case $\mathbb{P}(M_i < \infty) = 0$. ☐

Lemma 1.12. (1) *State i is recurrent iff $\sum_{t=1}^\infty p_{ii}^{(t)} = \infty$.*
(2) *State i is transient iff $\sum_{t=1}^\infty p_{ii}^{(t)}$ is finite.*

Proof. Let

$$A_t = \begin{cases} 1, & \text{if } X_t = i; \\ 0, & \text{otherwise.} \end{cases}$$

(So A_t is the indicator r.v. of the event $X_t = i$.) Then

$$\mathbb{E}(M_i) = \mathbb{E}\Big(\sum_{t=1}^\infty A_t\Big) = \sum_{t=1}^\infty \mathbb{E}(A_t) = \sum_{t=1}^\infty p_{ii}^{(t)}. \tag{1}$$

If i is transient, then $\mathbb{E}(M_i)$ exists, by Lemma 1.11, and hence the sum $\sum_{t=1}^\infty p_{ii}^{(t)}$ is finite. Conversely, if the sum is finite then $\mathbb{E}(M_i)$ exists and i is transient. These deals with (2). But parts (1) and (2) of the lemma are actually logically equivalent! ☐

Theorem 1.13. *Suppose $a, b \in S$ are states.*
(1) *If $a \leftrightarrow b$ and a is recurrent then b is recurrent.*
(2) *If $a \leftrightarrow b$ and a is transient then b is transient.*

Proof. Since (a) and (b) are equivalent, we just prove (a).

Since $a \to b$ there exists r with $p_{ab}^{(r)} > 0$; since $b \to a$ there exists s with $p_{ba}^{(s)} > 0$. Suppose that a is recurrent. Then

$$\sum_{t=1}^\infty p_{bb}^{(t)} \ge \sum_{k=1}^\infty p_{bb}^{(r+s+k)} \ge \sum_{k=1}^\infty p_{ba}^{(s)} p_{aa}^{(k)} p_{ab}^{(r)} = \alpha \sum_{k=1}^\infty p_{aa}^{(k)},$$

where $\alpha = p_{ba}^{(s)} p_{ab}^{(r)} > 0$. But $\sum_{k=1}^\infty p_{aa}^{(k)} = \infty$ by Lemma 1.12 and the fact that a is recurrent. So $\sum_{t=1}^\infty p_{bb}^{(t)} = \infty$ and b is recurrent. ☐

So we can speak of a communicating class as being recurrent or transient, knowing that all states are the same. We say that recurrence and transience are *class properties.*

1.6. **Random walk on $\mathbb{Z}$.** As an example application of Lemma 1.12, consider the Markov chain (X_t) with state space $S = \mathbb{Z}$ and transition probabilities given by

$$p_{ij} = \begin{cases} p, & \text{if } j = i + 1; \\ 1 - p, & \text{if } j = i - 1; \\ 0, & \text{otherwise.} \end{cases}$$

First consider the symmetric random walk, which is the case $p = \frac{1}{2}$. Since $\sum_{k=0}^{2n} \binom{2n}{k} = 2^{2n}$, the average of the coefficients $\binom{2n}{k}$ taken over $0 \le k \le 2n$ is $2^{2n}/(2n + 1)$. Now the largest of these binomial coefficients is $\binom{2n}{n}$ (you can easily check that the coefficients increase for $k < n$ and decrease for $k > n$) and the largest coefficient is certainly at least as large as the average. It follows that $\binom{2n}{n} \ge 2^{2n}/(2n + 1)$.

We now estimate the $2n$-step transition probability $p_{00}^{(2n)}$. In order to be back at state 0 after $2n$ steps, the Markov chain must make n right (i.e., increasing) transitions, and n left (i.e., decreasing). There are $\binom{2n}{n}$ possible ways to choose the sequence of n left and n right transitions. Each such sequence has a probability 2^{-2n} of occurring. Thus

$$p_{00}^{(2n)} = \binom{2n}{n} 2^{-2n} \geq \frac{2^{2n}}{2n+1} \times 2^{-2n} = \frac{1}{2n+1}.$$

Now

$$\sum_{t=1}^{\infty} p_{00}^{(t)} \geq \sum_{n=1}^{\infty} p_{00}^{(2n)} \geq \sum_{n=1}^{\infty} \frac{1}{2n+1} \geq \frac{1}{2} \sum_{n=1}^{\infty} \frac{1}{n+1} = \infty.$$

Hence, by Lemma 1.12, state 0 is recurrent. By symmetry (or by Theorem 1.13) all states are recurrent.

Now consider an asymmetric walk with $p \neq \frac{1}{2}$. For concreteness, suppose $p = \frac{2}{3}$. By counting sequences of transitions as before, we have

$$p_{00}^{(2n)} = \binom{2n}{n} \left(\frac{1}{3}\right)^n \left(\frac{2}{3}\right)^n \leq 2^{2n} \left(\frac{2}{9}\right)^n = \left(\frac{8}{9}\right)^n,$$

where we have used the fact that the binomial coefficient $\binom{2n}{n}$ is certainly less than the sum $\sum_{k=0}^{2n} \binom{2n}{k} = 2^{2n}$ of all coefficients. Then

$$\sum_{t=1}^{\infty} p_{00}^{(t)} = \sum_{n=1}^{\infty} p_{00}^{(2n)} \leq \sum_{n=1}^{\infty} \left(\frac{8}{9}\right)^n = 9 - 1 = 8,$$

where we have used the fact that $p_{00}^{(t)} = 0$ when t is odd. Hence, by Lemma 1.12, state 0 is transient. By symmetry (or by Theorem 1.13) all states are transient.

1.7. **First return time.** If i is recurrent, let $R_i = \min\{t : t \geq 1, X_t = i\}$, where we assume $X_0 = i$. This is the *first return time.* Then

$$\mathbb{P}(R_i = k) = f_{ii}^{(k)}, \text{ and}$$
$$\mathbb{E}(R_i) = \sum_{k=1}^{\infty} k f_{ii}^{(k)}.$$

$\mathbb{E}(R_i)$ is the expected first return time or the expected time between visits to i.

Definition 1.7. Suppose i is a recurrent state. We say that state i is *positive recurrent* if $\mathbb{E}(R_i)$ is finite, and *null recurrent* if $\mathbb{E}(R_i)$ is infinite.

Theorem 1.14. *Suppose i is recurrent.*

- *If $p_{ii}^{(t)} \to 0$ as $t \to \infty$ then i is null recurrent.*
- *If $p_{ii}^{(t)} \not\to 0$ as $t \to \infty$ then i is positive recurrent.*

Proof. Omitted. □

The above theorem can be used to show that positive/null recurrence are class properties.

Theorem 1.15. *Consider an irreducible Markov chain.*

- *If the Markov chain is positive recurrent, then it has a unique equilibrium distribution, which is given by* $w_i = 1/\mathbb{E}(R_i)$ *for all* $i \in S$.
- *Otherwise (i.e., it is null recurrent or transient) the Markov chain has no equilibrium distribution.*

Proof. Omitted. □

1.7.1. *Summary of recurrence/transience.* For a Markov chain with state space S (possibly infinite) and $i \in S$,

$$i \text{ recurrent} \iff f_{ii} = 1 \iff \sum_{t=1}^{\infty} p_{ii}^{(t)} = \infty \iff \mathbb{E}(M_i) = \infty;$$

$$i \text{ transient} \iff f_{ii} < 1 \iff \sum_{t=1}^{\infty} p_{ii}^{(t)} < \infty \iff \mathbb{E}(M_i) = f_{ii}/(1 - f_{ii}),$$

where M_i is the number of returns to state i.

Now assume the Markov chain is irreducible and that state i is recurrent. Then

$$i \text{ positive recurrent} \iff \mathbb{E}(R_i) < \infty \iff p_{ii}^{(t)} \not\to 0 \iff \exists\,! \text{ equilibrium distribution};$$

$$i \text{ null recurrent} \iff \mathbb{E}(R_i) = \infty \iff p_{ii}^{(t)} \to 0 \iff \neg\exists\,! \text{ equilibrium distribution},$$

where R_i is the time of first return to i. When a unique equilibrium distribution exists, it is given by $w_i = 1/\mathbb{E}(R_i)$.