Introduction to Stochastic Processes
(Based on the book by G. F. Lawler)

CHAPTER 1: Finite Markov Chains.

§1.1. Definitions and Examples.

Consider a discrete-time stochastic process, X_n, $n=0,1,2,\ldots$, where X_n is a random variable with values in the finite set $S=\{1,2,\ldots,N\}$ or $\{0,1,2,\ldots,N-1\}$. The values of X_n are called the states of the system. The probabilities of this process are defined by $\mathbb{P}(X_0=i_0,\ldots,X_n=i_n)$, for every n and $i_0,\ldots,i_n\in S$. Alternatively, we could define the initial probabilities $\phi(i)=\mathbb{P}(X_0=i)$, $i=1,\ldots,N$, and the transition probabilities

$$q_n(i_n|i_0,\ldots,i_{n-1})\doteq\mathbb{P}(X_n=i_n|X_0=i_0,\ldots,X_{n-1}=i_{n-1}),$$ so that

$$\mathbb{P}(X_0=i_0,\ldots X_n=i_n)=\phi(i_0)\,q_1(i_1|i_0)\,q_2(i_2|i_0,i_1)\cdots q_n(i_n|i_0,\ldots,i_{n-1}).$$

$$=\mathbb{P}(X_0=i_0)\cdot\mathbb{P}(X_0=i_0,X_1=i_1)/\mathbb{P}(X_0=i_0)\cdot\frac{\mathbb{P}(X_0=i_0,X_1=i_1,X_2=i_2)}{\mathbb{P}(X_0=i_0,X_1=i_1)}\cdots$$

A stochastic process X_n is said to satisfy the Markov property (and is called a Markov Chain) if

$$\mathbb{P}(X_n=i_n|X_0=i_0,\ldots,X_{n-1}=i_{n-1})=\mathbb{P}(X_n=i_n|X_{n-1}=i_{n-1})$$

for all n and $i_0,\ldots,i_n\in S$.

A Markov chain X_n is called homogeneous or time-homogeneous if

$$\mathbb{P}(X_n=i_n|X_0=i_0,\ldots X_{n-1}=i_{n-1})=\mathbb{P}(X_n=i_n|X_{n-1}=i_{n-1})$$
$$=p(i_{n-1},i_n)$$ for all n and $i_0,\ldots,i_n$, for some function $p:S\times S\to[0,1]$.

Note that in this case, we have

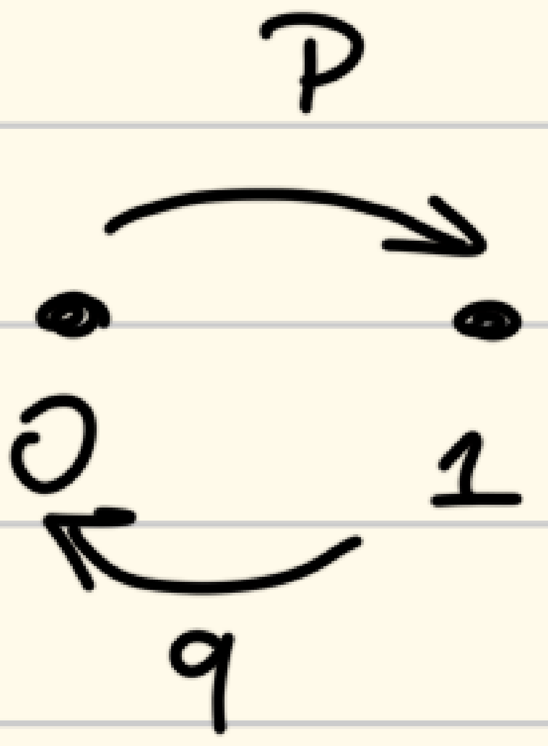
$$P(X_0=i_0,\dots,X_n=i_n)=\phi(i_0)\,q_1(i_1|i_0)\,q_2(i_2|i_0,i_1)\cdots q_n(i_n|i_0,\dots,i_{n-1})$$
$$=\phi(i_0)\,q_1(i_1|i_0)\,q_2(i_2|i_1)\cdots q_n(i_n|i_{n-1})$$
$$=\phi(i_0)\,p(i_0,i_1)\,p(i_1,i_2)\cdots p(i_{n-1},i_n)$$

The matrix $P=(p(i,j))_{n\times n}$ is called the transition matrix of the process. Note that the entries $P_{ij}=p(i,j)$ of P satisfy

$$0\leq P_{ij}\leq 1, \text{ and } \sum_{j=1}^{N} P_{ij}=1, \ \forall i,j=1,\dots,n.$$

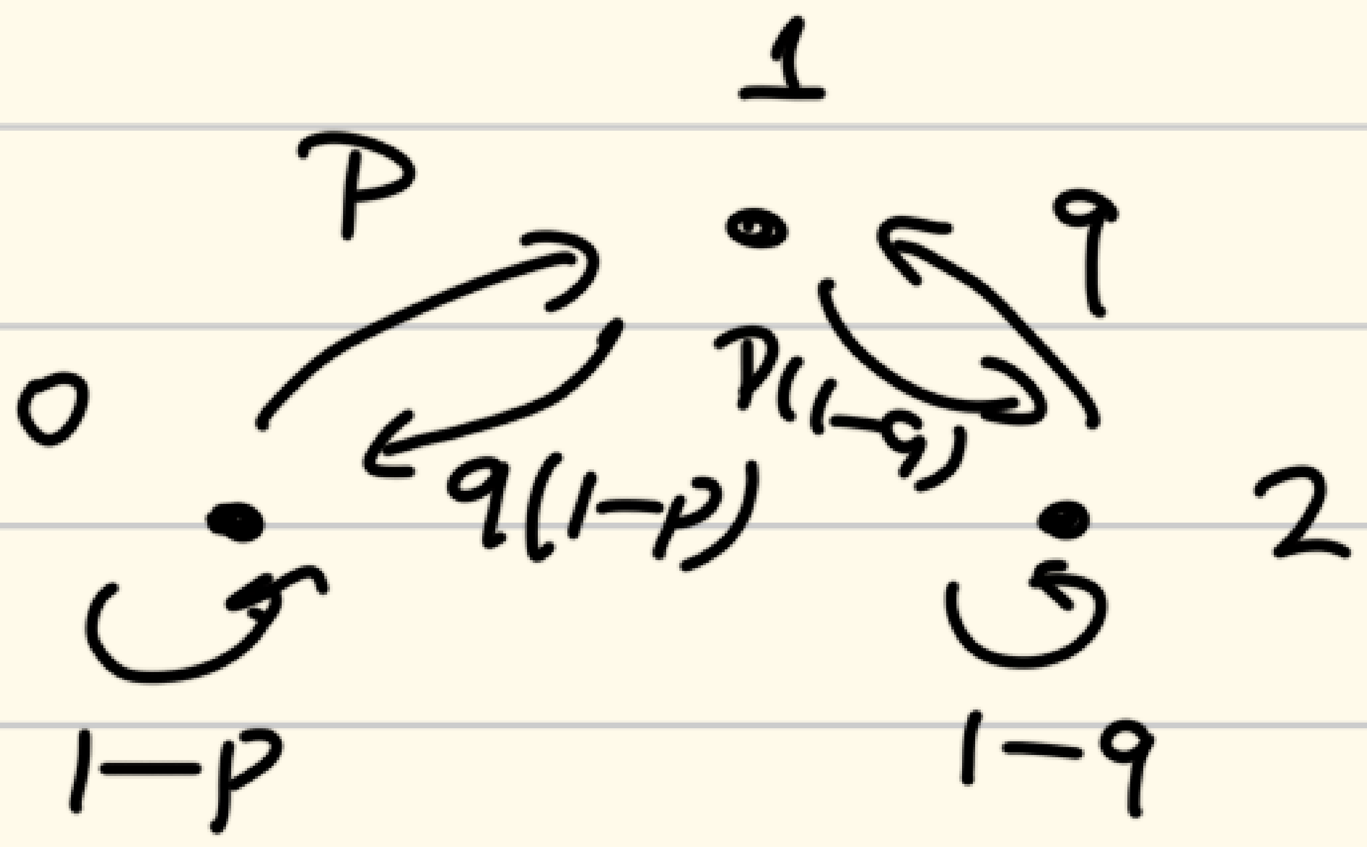
Example 1 Two-state Markov chain.

$X_n\in S=\{0,1\}$. Let $p=P_{01}$ and $q=P_{10}$



$$P=\begin{bmatrix}1-p & p\\ q & 1-q\end{bmatrix}.$$

Example 2 $S=\{0,1,2\}$ $$P=\begin{bmatrix}1-p & p & 0\\ q(1-p) & 1-p(1-q)-q(1-p) & p(1-q)\\ 0 & q & 1-q\end{bmatrix}$$



Example 3. Random walk with reflecting Boundary

$p(i,i+1)=p$, $p(i,i-1)=1-p$, $0<i<N$, and
$P(0,1)=1$, $P(N,N-1)=1$.

If $p=1/2$ it is called symmetric or unbiased random walk with reflecting boundary. If $p\neq 1/2$

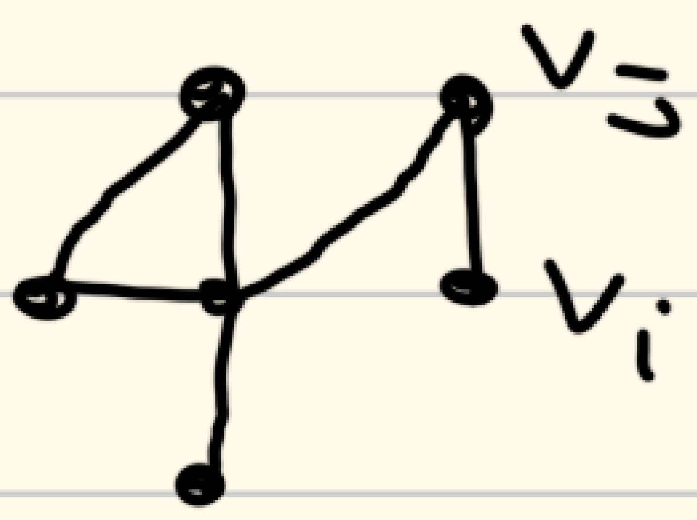
It is called biased random walk.

Example 4. Random walk with Absorbing Boundaries.

$p(i,i+1)=p$, $p(i,i-1)=1-p$, $0<i<N$ and $P(0,0)=P(N,N)=1$.

(We adopt the convention that if $p(i,j)$ is not specified then it should be taken as 0.)

Example 5. Simple Random Walk on a Graph.



$d(i)$ = # of edges at the vertex i.

$v_i \sim v_j$ if there is an edge between v_i and v_j.

(Simple, finite undirected graph)

If $v_i \sim v_j$ then $P(v_i, v_j) = 1/d(v_i)$.

Given a transition matrix P and initial probability distribution ϕ we define

$$P_n(i,j) = \mathbb{P}\{X_n=j \mid X_0=i\} = \mathbb{P}\{X_{n+k}=j \mid X_k=i\}$$

Then $\mathbb{P}\{X_n=j\} = \sum_{i\in S} \phi(i)\,\mathbb{P}\{X_n=j \mid X_0=i\}$ Moreover,

$$\mathbb{P}\{X_{n+1}=j \mid X_0=i\} = \sum_{k\in S} \mathbb{P}\{X_n=k \mid X_0=i\}\,\mathbb{P}\{X_{n+1}=j \mid X_n=k\}$$

$$= \sum_{k\in S} P_n(i,k)\,P(k,j)$$

So $P_{n+1}(i,j) = (P_n \cdot P)(i,j)$. Hence, $P_n = P^n$.

Let $\tilde{\phi}_0 \doteq (\phi_0(1), \ldots, \phi_0(N))$, the initial probability distribution vector. Also define $\tilde{\phi}_n = \tilde{\phi}_0 \mathbb{P}^n$, the state vector at step n (time n).

Example 6. Consider the first example with $p = 1/4$ and $q = 1/6$. Then $\mathbb{P}^6 = \begin{bmatrix} 3/4 & 1/4 \\ 1/6 & 5/6 \end{bmatrix}^6 = \begin{bmatrix} 0.424 & 0.576 \\ 0.384 & 0.616 \end{bmatrix}$.

Suppose $\tilde{\phi}_0 = (1, 0)$. $\tilde{\phi}_0 \mathbb{P}^6 = (0.424 \;\; 0.576)$. So,
$\mathbb{P}\{X_6 = 0 \mid X_0 = 0\} = 0.424$ and $\mathbb{P}\{X_6 = 1 \mid X_0 = 0\} = 0.576$.
Similarly, $\mathbb{P}\{X_6 = 0 \mid X_0 = 1\} = 0.384$ and $\mathbb{P}\{X_6 = 1 \mid X_0 = 1\} = 0.616$.

This model suits any Pass-Fail trial experiment.

§1.2. Large Time Behavior and Invariant Probability.

Clearly the matrix $\mathbb{P}^n$, for large n values, gives an idea about the large-time behavior of a Markov chain with transition matrix $\mathbb{P}$.

For example, if $\mathbb{P} = \begin{bmatrix} 3/4 & 1/4 \\ 1/6 & 5/6 \end{bmatrix}$ and for n large

$\mathbb{P}^n \approx \begin{bmatrix} 0.4 & 0.6 \\ 0.4 & 0.6 \end{bmatrix}$. In other words, $\lim_{n\to\infty} \mathbb{P}^n = \Pi$ exists

and its rows are identical. Moreover, if $v = (v(1), \ldots, v(N))$ is any probability vector, i.e. $v(i) \geq 0$ and $\sum v(i) = 1$, then $\lim_{n\to\infty} v\mathbb{P}^n = \overline{\pi}$, where $\overline{\pi}$ is one of the rows of Π.
In other words, any initial vector of probability distributions yields the same limiting vector of probability distributions.

Also $\bar\pi = \lim_{n\to\infty} v P^n = (\lim_{n\to\infty} v P^{n-1}) P = \bar\pi P$ so that $\bar\pi$ is a left eigenvector for P, for the eigenvalue 1. Such a vector is called a stationary, equilibrium or steady-state probability distribution (vector).

Immediately, the following questions come to mind:

1) Does every stochastic matrix P have an invariant probability distributions for stochastic matrices?
2) Is the invariant probability distribution unique?
3) When can we conclude that
$\lim_{n\to\infty} P^n = [\bar\pi \ \bar\pi \cdots \bar\pi]^T$ and hence, for all initial probability vector $\bar v$ we have
$\lim_{n\to\infty} \bar v P^n = \bar\pi$?

Example Let's investigate these questions on the stochastic matrix $P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$.

This matrix P is diagonalizable with eigenvalues 1 and $1-p-q$ so that $Q^{-1}PQ = D = \begin{bmatrix} 1 & 0 \\ 0 & 1-p-q \end{bmatrix}$, where $Q = \begin{bmatrix} 1 & -p \\ 1 & q \end{bmatrix}$. The columns of Q are right eigenvalues of P, and the rows of Q^{-1} are the left eigenvectors of P. The eigenvectors are unique if we require them to be probability vectors.

$$P^n = (QDQ^{-1})^n = QD^nQ^{-1} = Q\begin{bmatrix} 1 & 0 \\ 0 & (1-p-q)^n \end{bmatrix}Q^{-1} \to Q\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}Q^{-1}$$

which is

$$\lim_{n\to\infty} P^n = \begin{bmatrix} q/(p+q) & p/(p+q) \\ q/(p+q) & p/(p+q) \end{bmatrix} = \begin{bmatrix} \bar\pi \\ \bar\pi \end{bmatrix},$$ because $|1-p-q|<1$.

As in this specific example, the vector $[1 \cdots 1]^T$ is a right eigenvector for any stochastic matrix because the sum of each row is equal 1. Hence, if P is any stochastic matrix then it has at least one left eigenvector with eigenvalue 1.

Suppose further that we've shown:

A) The left eigenvector can be chosen to have all non-negative entries.

B) The eigenvalue 1 is simple and all other eigenvalues have absolute values less than 1.

Then the Jordan form of P can be chosen so that $D = Q^{-1}PQ$, where $D = \begin{bmatrix} 1 & 0 \cdots 0 \\ 0 & \\ \vdots & M \\ 0 & \end{bmatrix}$ so that $\lim_{n\to\infty} M^n = 0$.

So, $\lim_{n\to\infty} P^n = \lim_{n\to\infty} Q^{-1}PQ = Q \begin{bmatrix} 1 & 0 \cdots 0 \\ 0 & \\ \vdots & 0 \\ 0 & \end{bmatrix} Q^{-1} = \begin{bmatrix} \overline{\pi} \\ \vdots \\ \overline{\pi} \end{bmatrix}$

The Perron-Frobenius Theorem tells us that all stochastic matrices with positive entries satisfy (A) and (B).

Some stochastic matrices may not have all positive entries, some entries might be zero. However, we have

Fact. If P is a stochastic matrix such that for some n, P^n has all positive entries then P satisfy (A) and (B).

In this next section we'll classify all stochastic matrices P so that some P^n has all positive entries.

§1.3. Classification of States.

First let's see some examples of stochastic matrices P so that no P^n is (strictly) positive.

Example 1 Simple random walk on $\{0,1,2,3,4\}$ with reflecting boundaries. In this case

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \text{ Then } P^n \approx \begin{bmatrix} 0.25 & 0 & 0.5 & 0 & 0.25 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0.25 & 0 & 0.5 & 0 & 0.25 \\ & & \vdots & & \end{bmatrix}$$

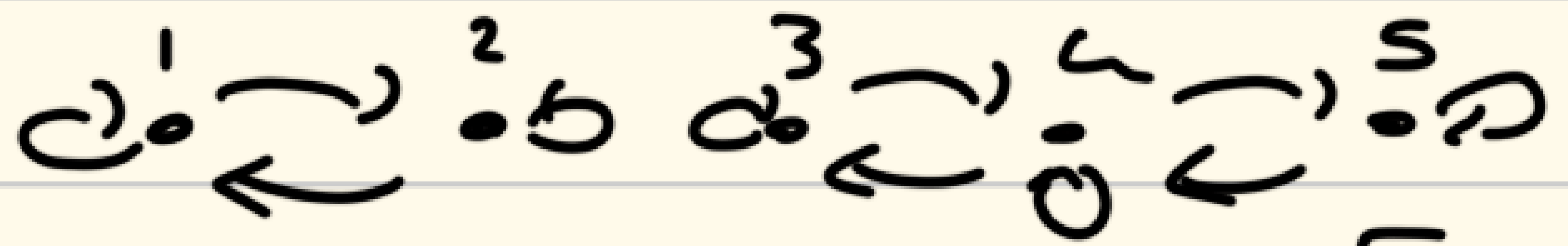
if n is even and $P^n \approx \begin{bmatrix} 0 & 0.5 & 0 & 0.5 & 0 \\ 0.25 & 0 & 0.5 & 0 & 0.25 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ & & \vdots & & \end{bmatrix}$ if n is odd.

Note that $P_n(i,j) = 0$ if i is even, j is odd, n is even and $P_n(i,j) = 0$ if i, j are even and n is odd.

Example 2. Simple random walk with absorbing boundary on $\{0,..,4\}$. This time $P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ & \vdots & & & \\ 0 & 0 & - & - & 1 \end{bmatrix}$.

If n is large $P^n \approx \begin{bmatrix} 1 & 0 & 0 \\ 3/4 & \vdots & 1/4 \\ 1/2 & \vdots & 1/2 \\ 1/4 & \vdots & 3/4 \\ 0 & 0 & 1 \end{bmatrix}$ so that some entries are always zero, e.g. $P_n(0,i) = 0$, $\forall n$, $\forall i = 1,2,3,4$. The states $1,2,3$ are called transient, the states 0 and 4 are called absorbing.

Example 3. $S = \{1,2,3,4,5\}$ and $P = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/6 & 5/6 & 0 & 0 & 0 \\ 0 & 0 & 3/4 & 1/4 & 0 \\ 0 & 0 & 1/8 & 2/3 & 5/24 \\ 0 & 0 & 0 & 1/6 & 5/6 \end{bmatrix}$



In this case, $P^n \approx \begin{bmatrix} 0.25 & 0.75 & 0 & 0 & 0 \\ 0.25 & 0.75 & 0 & 0 & 0 \\ 0 & 0 & 0.182 & 0.364 & 0.455 \\ 0 & 0 & 0.182 & 0.364 & 0.455 \\ 0 & 0 & 0.182 & 0.364 & 0.455 \end{bmatrix}$, we have Indeed, two disjoint chains.

Reducibility. We say that two states i and j of a Markov chain communicate if there are $m, n \geq 0$ such that $P_m(i,j) > 0$ and $P_n(j,i) > 0$. In this case we write $i \leftrightarrow j$. This is clearly an equivalence relation: $P_0(i,i) = 1 > 0$ (reflexive), symmetric by definition. For transitivity assume that $P_{m_1}(i,j) > 0$ and $P_{m_2}(j,k) > 0$. Then

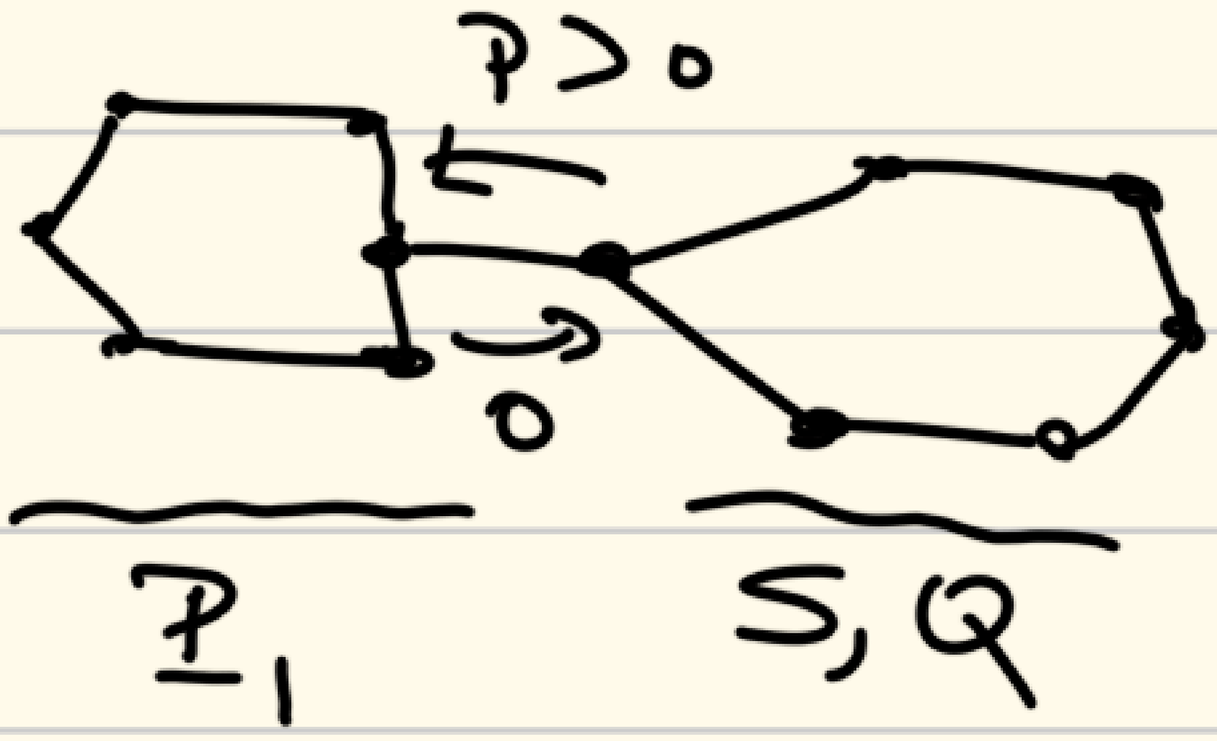
$$\begin{aligned}
\mathbb{P}\{X_{m_1+m_2}=k \mid X_0=i\} &= \sum_{l} \mathbb{P}\{X_{m_1+m_2}=k, X_{m_1}=l \mid X_0=i\} \\
&= \sum_{l} \mathbb{P}\{X_{m_1+m_2}=k, X_{m_1}=l, X_0=i\} / \mathbb{P}\{X_0=i\} \\
&= \sum_{l} \frac{\mathbb{P}\{X_{m_1+m_2}=k, X_{m_1}=l, X_0=i\}}{\mathbb{P}\{X_{m_1}=l, X_0=i\}} \, \frac{\mathbb{P}\{X_{m_1}=l, X_0=i\}}{\mathbb{P}\{X_0=i\}} \\
&= \sum_{l} \mathbb{P}\{X_{m_1+m_2}=k \mid X_{m_1}=l, X_0=i\} \, \mathbb{P}\{X_{m_1}=l \mid X_0=i\} \\
\text{(Since } X_n \text{ is a Markov chain)} \quad &= \sum_{l} \mathbb{P}\{X_{m_1+m_2}=k \mid X_{m_1}=l\} \, \mathbb{P}\{X_{m_1}=l \mid X_0=i\}, \\
&\geq \mathbb{P}\{X_{m_1+m_2}=k \mid X_{m_1}=j\} \, \mathbb{P}\{X_{m_1}=j \mid X_0=i\} \\
&= P_{m_1}(i,j) \, P_{m_2}(j,k) > 0.
\end{aligned}$$

Hence, $\leftrightarrow$ is transitive and thus is an equivalence relation. The equivalence classes of this chain is a disjoint partition of the states. Each class is called a communication class. If there is only one class then we say that the chain is irreducible.

If a chain starts in a state of a class and leaves this class so that it never comes back that class is called transient with transient states. Other classes and states are called recurrent.

Suppose $\mathbb{P}$ is the matrix of a reducible chain with recurrent communication classes $R_1, \ldots, R_s$ and transient classes $T_1, \ldots, T_s$. Since there are finitely many classes there must be at least one recurrent class. For each recurrent class R, the submatrix of P obtained from considering only the rows and columns for states in R is itself a stochastic matrix. Now, we see that by reordering the states if necessary the matrix P can be written as

$$P = \left[\begin{array}{ccc|c} P_1 & & 0 & \\ & P_2 & & 0 \\ 0 & & \ddots P_r & \\ \hline & S & & Q \end{array}\right]$$



$$\text{Then } P^n = \left[\begin{array}{ccc|c} P_1^n & \ddots & & 0 \\ & & P_r^n & 0 \\ \hline & S_n & & Q^n \end{array}\right], \text{ for some matrix } S_n.$$

Now, we will study P_i^n and Q^n seperately.

§1.3.2. Periodicity Suppose P is irreducible (so it is recurrent!). For a state i of P its period is defined as $d = d(i)$ to be the greatest common divisor of

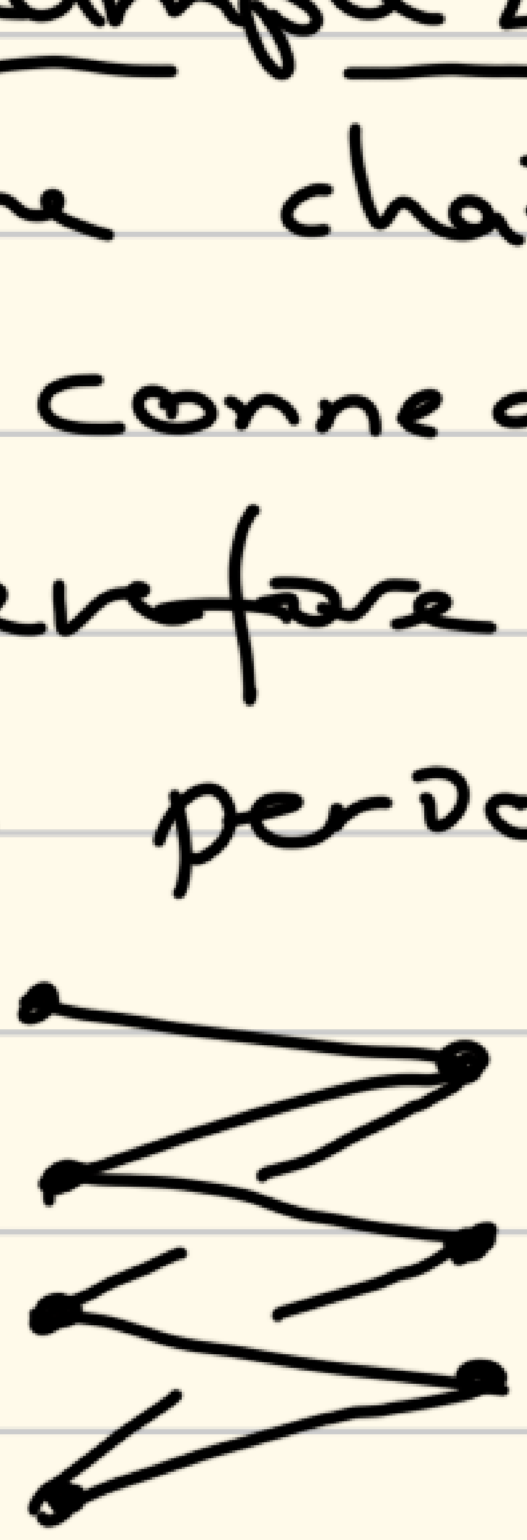
$$J_i \doteq \{n \geq 0 \mid P_n(i,i) > 0\}.$$

In Example 1 we have $P_{2n}(i,i) > 0$ and $P_{2n+1}(i,i) = 0, \forall n, i$, and thus the period of each class is 2. Note that the set J_i is closed under addition, because if $m, n \in J_i$, then $P_{m+n}(i,i) \geq P_m(i,i) P_{n,n}(i,i)$.

Clearly, $J_i \subseteq \{0, d, 2d, 3d, \ldots\}$. Exercise 1.21 also shows that J_i contains all but finitely many element of $\{0, d, 2d, 3d, \ldots\}$. Suppose that j is another state and m, n are such that $P_m(i,j) > 0$ and $P_n(j,i) > 0$. Then $m+n \in J_i \cap J_j$ and thus $m+n = kd$ for some integer k, $(P_{m+n}(i) \geq P_m(i,j) P_n(j,i) > 0)$. Also, if $l \in J_j$, then $P_{m+n+l}(i,j) \geq P_m(i,j) P_l(j,j) P_n(j,i) > 0$ and hence $d \mid l$.

So, we have shown that if d divides every element of J_i then it divides every element of J. In other words, all states have the same period.

Example 4. Consider simple random walk on a graph. The chain is irreducible if and only if the graph is connected. If $v \sim w$, then $P_2(v,v) \geq P_1(v,w)\, P_1(w,v) > 0$. Therefore the period is either 1 or 2. Note that the period is 2 if and only if the graph is bipartite:



§1.3.3. Irreducible, aperiodic Chains

An irreducible chain is called aperiodic if $d=1$.

Now assume P is irreducible and aperiodic. Take any states i and j. Since P is irreducible there is some $m(i,j)$ so that $P_{m(i,j)}(i,j) > 0$. Moreover, since P is aperiodic there is some $M(i)$ so that for all $n \geq M(i)$ $P_n(i,i) > 0$. Hence if $n \geq M(i)$, $P_{n+m(i,j)}(i,j) \geq P_n(i,i)\, P_{m(i,j)}(i,j) > 0$.

Now let $M = \max\{M(i), m(i,j) \mid \forall i,j\}$. Then $P_n(i,j) > 0$ for all $n \geq M$ and all states i,j. So, with the results of §1.2 we have proved:

Theorem. If P is the transition matrix for an irreducible, aperiodic Markov chain, then there is a unique invariant probability vector π satisfying $\pi P = \pi$. Moreover, if ϕ is any initial probability vector $\lim_n \phi P^n = \pi$. Moreover, $\pi(i) > 0$, for each i.

§1.3.4. Reducible or Periodic Chains

Now we'll first consider the case where P is reducible with recurrent classes $R_1,\dots,R_r$ and transient classes $T_1,\dots,T_s$. Each R_i is an irreducible class and thus there exist r different invariant probability vector $\bar{\pi}^1,\dots,\bar{\pi}^r$ with π^k concentrated on R_k ($\bar{\pi}^k(i)=0$ if $i\notin R_k$). For simplicity let's assume first that each recurrent class R_k is aperiodic. Then for each $i\in R_k$, $\lim_{n\to\infty} P_n(i,j)=\pi^k(j)$, $\forall j\in R_k$, and $P_n(i,j)=0$, $\forall j\notin R_k$.

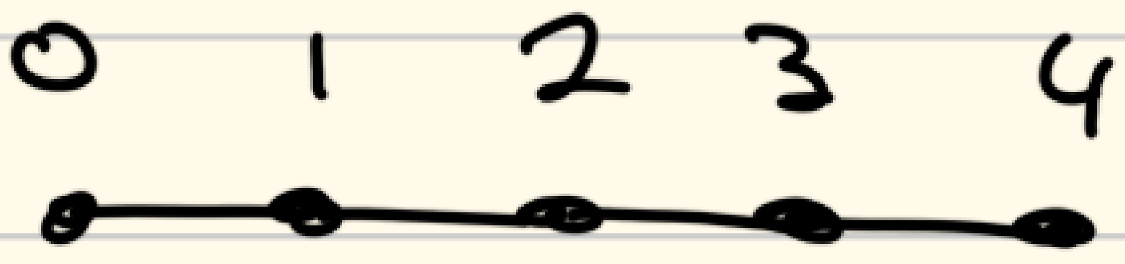
If i is any transient state, then any chain starting in state i eventually ends up in a recurrent state. Hence for any transient state j we have

$$\lim_{n\to\infty} P_n(i,j)=0.$$

For any recurrent class R_k, let $\alpha_k(i)$, $k=1,\dots,r$, denote the probability that the transient class eventually ends up in the recurrent class k. Once the chain reaches a state in R_k it will settle down to the equilibrium distribution on R_k. From this we see that if $j\in R_k$ then $\lim_{n\to\infty} P_n(i,j)=\alpha_k(i)\pi^k(j)$

If $\bar{\phi}$ is an initial vector, $\lim_{n\to\infty}\bar{\phi}P^n$ exists but depends on $\bar{\phi}$.

Example:



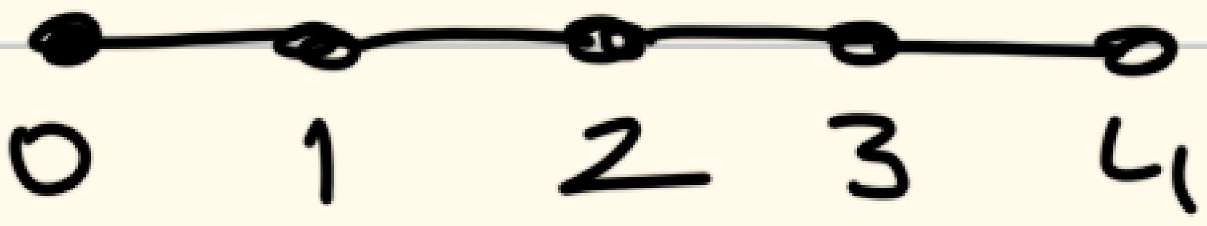
0, 4 are absorbing states. All other $p(i,j)=1/2$, $i,j\in\{1,2,3\}$, $i\neq j$.

$$P=\begin{bmatrix}1&0&0&0&0\\ 1/2&0&1/2&0&0\\ 0&1/2&0&1/2&0\\ 0&0&1/2&0&1/2\\ 0&0&0&0&1\end{bmatrix},\quad \lim_n P^n=\begin{bmatrix}1&0&0&0&0\\ 3/4&0&0&0&1/4\\ 1/2&0&0&0&1/2\\ 1/4&0&0&0&3/4\\ 0&0&0&0&1\end{bmatrix}$$

exists, but $\lim_n \bar{\phi}P^n$ depends on $\bar{\phi}$.

Now suppose that $\mathbb{P}$ is irreducible but has period $d>1$. In this case the state space splits nicely into d sets, $A_1, \dots, A_d$ such that the chain always move from A_i to A_{i+1} (A_d to A_1).

To illustrate this consider Example 1, the linear chain with reflecting boundaries:



$$\mathbb{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

The eigenvalues of P are $1, -1, 0, 1/\sqrt{2}, -1/\sqrt{2}$. $\mathbb{P}$ is diagonalizable (it is symmetric)

$$D = Q^{-1}\mathbb{P}Q, \quad P^n = QD^nQ^{-1} \approx Q \begin{bmatrix} 1 & & & & \\ & (-1)^n & & & \\ & & 0 & & \\ & & & 0 & \\ & & & & 0 \end{bmatrix} Q^{-1}$$

$$= \begin{bmatrix} 1/8 & 1/4 & 1/4 & 1/4 & 1/8 \\ 1/8 & 1/4 & 1/4 & 1/4 & 1/8 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1/8 & 1/4 & 1/4 & 1/4 & 1/8 \end{bmatrix} + (-1)^n \begin{bmatrix} 1/8 & -1/4 & 1/4 & -1/4 & 1/8 \\ -1/8 & + & - & \vdots & \vdots \\ 1/8 & - & + & \vdots & \vdots \\ -1/8 & + & - & & \\ 1/8 & - & + & & \end{bmatrix}$$

The asymptotic value for $\mathbb{P}^n$ depends on whether n is even or odd. In this case, the invariant probability at a state i, $\pi(i)$, does not represent the limit of $P_n(i,j)$. Instead it represents the average amount of time that is spent in state i. In fact, the average of $P_n(j,i)$ and $P_{n+1}(j,i)$ approaches to $\pi(i)$ for any initial state j:

$$\pi(i) = \lim_{n\to\infty} \frac{1}{2}\left(P_n(j,i) + P_{n+1}(j,i)\right).$$

In general, if $\mathbb{P}$ is irreducible with period d, $\mathbb{P}$ will have d eigenvalues with absolute value 1. Each is simple and there is a unique invariant probability $\bar{\pi}$.

Moreover, if $\bar{\phi}$ is any initial probability vector, then
$$\lim_{n\to\infty} \frac{1}{d}\, \bar{\phi}\left[\mathbb{P}^{n+1}+\cdots+\mathbb{P}^{n+d}\right]=\bar{\pi}.$$

§1.4 Return Times.

Let X_n be an irreducible possibly periodic Markov chain with transition matrix $\mathbb{P}$. For any state j we let $Y(j,n)$ denote the amount of time spent in state j up to and including time n, $Y(j,n)=\sum_{m=0}^{n} I\{X_m=j\}$, when I denotes the indicator function of an event. If $\bar{\pi}$ denotes the invariant probability distribution for $\mathbb{P}$, then it follows from the results of the previous sections that
$$\lim_{n\to\infty} \frac{1}{n+1}\, \mathbb{E}\left(Y(j,n)\,|\,X_0=i\right)=\lim_{n\to\infty} \frac{1}{n+1}\sum_{m=0}^{n} \mathbb{P}\{X_m=j\,|\,X_0=i\}$$
$$=\pi(j).$$

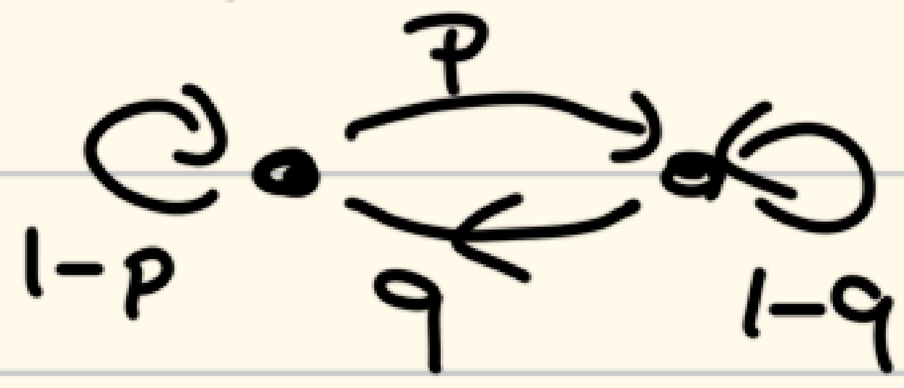
In other words, $\pi(j)$ represents the fraction of time that the chain spends in state j.

Now fix a state i and assume $X_0=i$. Define T to be the first time after 0 that the Markov chain visits the state i,
$$T=\min\{n\geq 1 : X_n=i\}.$$
Since the chain is irreducible $T<+\infty$ with probability 1 (Exercise 1.7 even shows $\mathbb{E}(T)<+\infty$). If T_i is identical to T, then $T_1+\cdots+T_k$ represent the time the chain visits the state i kth time. Here T_m represents the time between the $(m-1)$st and mth return. For large k, the Law of Large Numbers tells us that

$1/k\,(T_1+\dots+T_k) \simeq \mathbb{E}[T]$. On the other hand, we have seen that in n steps we expect $n\pi(i)$ visits to the state i. So putting $n = k\mathbb{E}[T]$ we see that $\mathbb{E}[T] = 1/\pi(i)$.

This says that the expected number of steps to return to i, assuming we start at i, is given by the reciprocal of $\pi(i)$. This argument can be made rigorous (for example see Exercise 1.11).

__Example__ Consider the 2-state Markov chain, $S=\{0,1\}$ and $\mathbb{P} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$,



In Section 1.2 we computed $\pi = (q/_{p+q},\ p/_{p+q})$ and hence $\mathbb{E}[T] = \dfrac{1}{\pi(0)} = \dfrac{p+q}{q}$, where we assume that the chain starts at 0 and T is its return time to 0. Note we can compute directly that

$$\mathbb{P}(T \geq n) = \mathbb{P}(X_1=1,\dots,X_{n-1}=1 \mid X_0=0\} = p\,(1-q)^{n-2}.$$

If Y is any random variable taking values in the nonnegative integers,

$$\mathbb{E}[Y] = \sum_{n=1}^{\infty} n\,\mathbb{P}\{Y=n\} = \sum_{n=1}^{\infty}\sum_{k=1}^{n} \mathbb{P}\{Y=n\}$$
$$= \sum_{k=1}^{\infty}\sum_{n=k}^{\infty} \mathbb{P}\{Y=n\} = \sum_{k=1}^{\infty} \mathbb{P}\{Y \geq k\}.$$

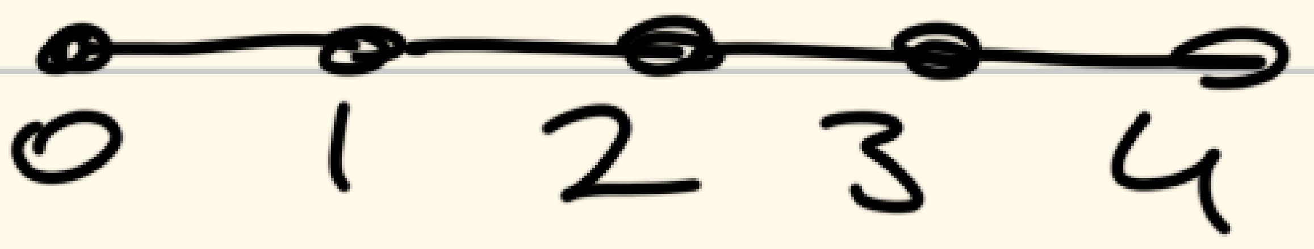
So,
$$\mathbb{E}[T] = \sum_{n=1}^{\infty} n\,\mathbb{P}\{T=n\} = \sum_{n=1}^{\infty} \mathbb{P}\{T \geq n\}$$
$$= 1 + \sum_{n=2}^{\infty} p\,(1-q)^{n-2} = \frac{p+q}{q}.$$

§1.5. Transient States.

Recall that a state i is transient if the chain visits i only finitely many times. We've seen before that we may arrange states so that the transition matrix has the form

$$\mathbb{P} = \left[\begin{array}{c|c} \bar{P} & 0 \\ \hline S & Q \end{array}\right] \text{ and } \mathbb{P}^n = \left[\begin{array}{c|c} \bar{P}^n & 0 \\ \hline S^n & Q^n \end{array}\right],$$

where the square Q matrix corresponds to the transient states. For example recall the example of random walk with absorbing boundaries:



$$\mathbb{P} = \begin{array}{c|c} \left[\begin{array}{cc|ccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hline 1/2 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 & 1/2 & 0 \end{array}\right] & \begin{array}{c} 0 \\ 4 \\ 1 \\ 2 \\ 3 \end{array} \\ \begin{array}{ccccc} 0 & 4 & 1 & 2 & 3 \end{array} & \end{array}, \text{ where}$$

$$Q = \begin{array}{cc} \begin{array}{ccc} 1 & 2 & 3 \end{array} & \\ \left[\begin{array}{ccc} 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 1/2 & 0 \end{array}\right] & \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \end{array}.$$

The matrix Q is a substochastic matrix, whose entries are all non negative with row sums atmost 1. Since all the states in Q are transient $Q^n \to 0$ so that all eigenvalues of Q have norm less than 1. So, $I-Q$ is invertible. Say $M = (I-Q)^{-1}$.

Let i be a transient state and Y_i the total number of visits to i,

$$Y_i = \sum_{n=0}^{\infty} I\{X_n = i\}.$$

Since i is transient $Y_i < +\infty$ with probability 1. Suppose $X_0 = j$, where j is another transient state. Then

$$\mathbb{E}[Y_i \mid X_0 = j] = \mathbb{E}\left[\sum_{n=0}^{\infty} I\{X_n = i\} \mid X_0 = j\right]$$
$$= \sum_{n=0}^{\infty} \mathbb{E}\left[X_n = i \mid X_0 = j\right]$$
$$= \sum_{n=0}^{\infty} \mathbb{P}\{X_n = i \mid X_0 = j\}$$
$$= \sum_{n=0}^{\infty} P_n(j,i).$$

= the (j,i)th entry of $I+P+P^2+\cdots$

= " " $I+Q+Q^2+\cdots$

= " " of $M=(I-Q)^{-1}$!

So, for the above example, $M = (I-Q)^{-1} = \begin{array}{cc} \begin{bmatrix} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} \\ \begin{matrix} 1 & 2 & 3 \end{matrix} & \end{array}$
so that for a chain starting at state 1 it will be around before absorbed only 3 steps.

One may use this technique to determine the expected number of steps that an irreducible Markov chain takes to go from one state j to another state i. First write i as the first stage of the chain.

$$P = \left[\begin{array}{c|c} P(i,i) & R \\ \hline S & Q \end{array}\right].$$ Then change i to an absorbing state. Then P becomes
$$\tilde{P} = \left[\begin{array}{c|c} 1 & 0 \\ \hline S & Q \end{array}\right].$$

Let T_i be the number of steps to reach step i (starting from j). In other words, T_i is the smallest time n s.t. $X_n = i$, while $X_0 = j$. For any other k, let $T_{i,k}$ be the number of visits to k before reaching k (if we start at k, then we include this as one visit to k). Then,

$$\mathbb{E}\left[T_i \mid X_0 = j\right] = \mathbb{E}\left[\sum_{k \neq i} T_{i,k} \mid X_0 = j\right]$$
$$= \sum_{k \neq i} M_{jk}$$

In other words, $M\bar{1}$ is the vector whose j^{th} component is the number of steps starting at j until reaching i.

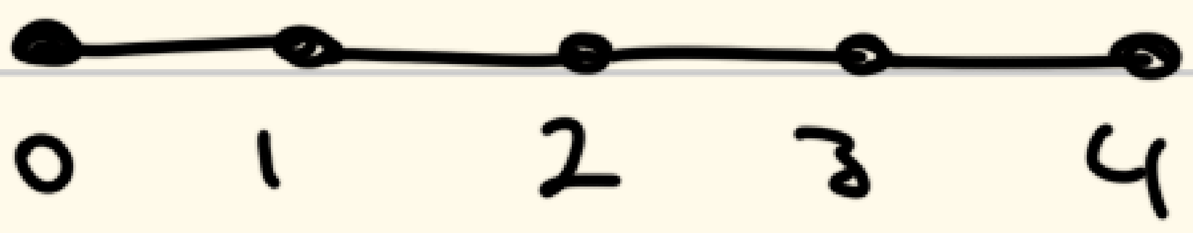
Example 1. Again let P be the transition matrix for random walk with reflecting boundaries:

$$P = \begin{array}{c} \left[\begin{array}{ccccc} 0 & 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right] \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \\ \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \end{array}$$

Let $i=0$, then $Q = \begin{array}{c} \left[\begin{array}{cccc} 0 & 1/2 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \end{array}\right] \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \\ \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \end{array}$

because if we let $i=0$ an absorbing state then P becomes:

$$\tilde{P} = \overset{O}{\underset{S \qquad Q}{\left[\begin{array}{c|cccc} 1 & 0 & 0 & 0 & 0 \\ \hline 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right]}}$$



Then $M = (I-Q)^{-1} = \begin{array}{c} \left[\begin{array}{cccc} 2 & 2 & 2 & 1 \\ 2 & 4 & 4 & 2 \\ 2 & 4 & 6 & 3 \\ 2 & 4 & 6 & 4 \end{array}\right] \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \\ \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \end{array}$

$M\bar{1} = (7, 12, 15, 16)$. In particular, the expected number of steps from 4 to 0 is 16.

Suppose now we have at least two different recurrent classes and ask the question: Starting at a given transient state j, what is the probability that the Markov chain eventually ends up in a particular recurrent class? In order to answer this question, we can assume that the recurrent classes consist of single points $r_1, \ldots, r_k$ with $p(r_i, r_i) = 1$. Ordering the states as $r_1, \ldots, r_k, t_1, \ldots, t_s$ (t_i's are transient states), we get

$\mathbb{P} = \left[\begin{array}{c|c} I & 0 \\ \hline S & Q \end{array}\right]$. For $i=1,\dots,s$, $j=1,\dots,k$, let $\alpha(t_i,r_j)$ be the probablity that the chain starting at t_i eventually ends up in recurrent state r_j. We set $\alpha(r_i,r_i)=1$ and $\alpha(r_i,r_j)=0$ if $i\neq j$. So, for any transient state t_i,

$$\alpha(t_i,r_j) = \mathbb{P}\{X_n = r_j \text{ eventually} \mid X_0=t_i\}$$

(S is the set of transient states)

$$= \sum_{x\in S} \mathbb{P}\{X_1=x \mid X_0=t_i\}\,\mathbb{P}\{X_n=r_j \text{ eventually} \mid X_1=x\}$$

$$+ \sum_{x\in R} \mathbb{P}\{X_1=x \mid X_0=t_i\}\,\mathbb{P}\{X_n=r_j \text{ eventually} \mid X_1=x\}$$

$$\|$$

$$\sum_{x\in R} \mathbb{P}\{X_1=x \mid X_0=t_i\}\,\delta_{x,r_j}$$

$$\|$$

$$\sum_{j} \mathbb{P}\{X_1=r_j \mid X_0=t_i\}$$

$$\|$$

$$S_{ij}$$

$$= S_{ij} + (QA)_{ij}, \text{ where } A = (\alpha(t_i,r_j))$$

So, $A = S + QA$ and thus, $A = Q^{-1}S = MS$.

<u>Example 2.</u> Again consider the example of random walk on $\{0,1,2,3,4\}$ with absorbing boundaries.

$$\mathbb{P} = \begin{array}{c} \left[\begin{array}{cc|ccc} 1&0&0&0&0\\ 0&1&0&0&0\\ \hline 1/2&0&0&1/2&0\\ 0&0&1/2&0&1/2\\ 0&1/2&0&1/2&0 \end{array}\right]\begin{array}{c}0\\4\\1\\2\\3\end{array} \\ \begin{array}{ccccc}0&4&1&2&3\end{array} \end{array}$$

$$S = \begin{array}{c} \left[\begin{array}{cc} 1/2 & 0 \\ 0 & 0 \\ 0 & 1/2 \end{array}\right]\begin{array}{c}1\\2\\3\end{array} \\ \begin{array}{cc}0 & 4\end{array} \end{array} \qquad M = \begin{array}{c} \begin{array}{ccc}1&2&3\end{array} \\ \left[\begin{array}{ccc} 3/2 & 1 & 1/2 \\ 1 & 2 & 1 \\ 1/2 & 1 & 3/2 \end{array}\right]\begin{array}{c}1\\2\\3\end{array} \end{array}$$

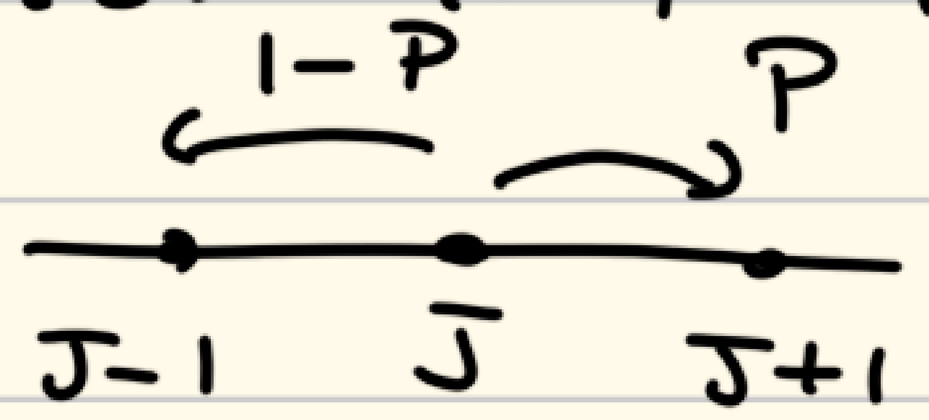
$$MS = \begin{array}{c} \begin{array}{cc}0 & 4\end{array} \\ \left[\begin{array}{cc} 3/4 & 1/4 \\ 1/2 & 1/2 \\ 1/4 & 3/4 \end{array}\right]\begin{array}{c}1\\2\\3\end{array} \end{array}$$

So, the probablity of a chain starting at state 1 ending at state 0 is 3/4.

<u>Example 3.</u> <u>Gambler's Ruin</u>. Consider the random walk on $\{0,1,\dots,N\}$ with absorbing boundaries.

Let $\alpha(J)=\alpha(J,N)$ be the probability that the walker starting at state J eventually ending at the absorbing end N. Clearly, $\alpha(0)=0$ and $\alpha(N)=1$. For $0<J<N$, we have then

$$\alpha(J)=(1-p)\alpha(J-1)+p\,\alpha(J+1).$$



So, we have $N-1$ linear equations in $N-1$ unknowns, $\alpha(1),\dots,\alpha(N-1)$. Writing this as $\alpha(J+1)=\frac{1}{p}\alpha(J)+\frac{p-1}{p}\alpha(J-1)$ we obtain a linear difference equation. We can solve this as follows:

$$r^2-\frac{1}{p}r-\frac{p-1}{p}=0 \Rightarrow pr^2-r-p+1=0$$

$$\Rightarrow (pr+p-1)(r-1)=0$$

$$\Rightarrow r=1 \text{ and } r=\frac{1-p}{p}$$

So, $\alpha(J)=c_1\cdot 1^J+c_2\left(\frac{1-p}{p}\right)^J$, for $J\neq 1/2$, and
$\alpha(J)=c_1+c_2J$ if $J=1/2$.

Using the boundary conditions we get:

$$p=1/2 \Rightarrow \alpha(J)=J/N.$$

$$p\neq 1/2 \Rightarrow \alpha(J)=\frac{1-\left(\frac{1-p}{p}\right)^J}{1-\left(\frac{1-p}{p}\right)^N}.$$

Note that if $p\leq 1/2$, $\lim_{N\to\infty}\alpha(J)=0$.

So, if $p\leq 1/2$ and if the gambler has fixed amount of money he/she will eventually loose with a high probability since the house has large, N, amount of money.

However, if $p>1/2$, then $\lim_{N\to\infty}\alpha(J)=1-\left(\frac{1-p}{p}\right)^J>0$.

Suppose now $p=1/2$ and T be the time it takes for the random walk to reach 0 or N, and let $G(J)=G(J,N)=E[T \mid X_0=J]$, the amount of time for the game to end.

Clearly, $G(0)=G(N)=0$ and for $1 \leq J \leq N-1$,

$$G(J)=1+\frac{1}{2}G(J-1)+\frac{1}{2}G(J+1).$$

This is an inhomogeneous linear difference equation, which can be written as

$G(J+1)-2G(J)+G(J-1)=-2$, which is equivalent to the ODE

$y''-2y'+y=-2e^{t}$, under the correspondence, $y(t)=\sum_{J=0}^{\infty} G(J)\frac{t^J}{J!}$.

Ch. Eqn. $r^2-2r+1=0 \Rightarrow r_{1,2}=1$ (double root)

So, $y_c=c_1e^{t}+c_2te^{t}$. For y_p, we should try $y_p=At^2e^t$.

$y_p'=2Ate^t+At^2e^t$
$y_p''=2Ae^t+4Ate^t+At^2e^t$

$y_p''-2y_p'+y_p=-2e^t \Rightarrow 2A=-2 \Rightarrow A=-1.$

So, $y_p=-t^2e^t \Rightarrow y_g=c_1e^t+c_2te^t-t^2e^t.$

$$\sum_{J=0}^{\infty} G(J) \frac{t^J}{J!} = c_1 e^t + c_2 t e^t - t^2 e^t$$

$$= c_1 \sum_{J=0}^{\infty} \frac{t^J}{J!} + c_2 \sum_{J=1}^{\infty} J \frac{t^J}{J!}$$

$$- \sum_{J=2}^{\infty} J(J-1) \frac{t^J}{J!},$$

so that $G(J) = c_1 + c_2 J - J(J-1)$.

$$G(0) = 0 \Rightarrow c_1 = 0 \Rightarrow G(J) = c_2 J - J(J-1).$$

$$G(N) = 0 \Rightarrow c_2 N - N(N-1) = 0$$
$$\Rightarrow c_2 = N-1$$

$$G(J) = (N-1)J - J(J-1)$$
$$= J(N-J)$$

$$\Rightarrow \mathbb{E}[T \mid X_0 = J] = J(N-J).$$

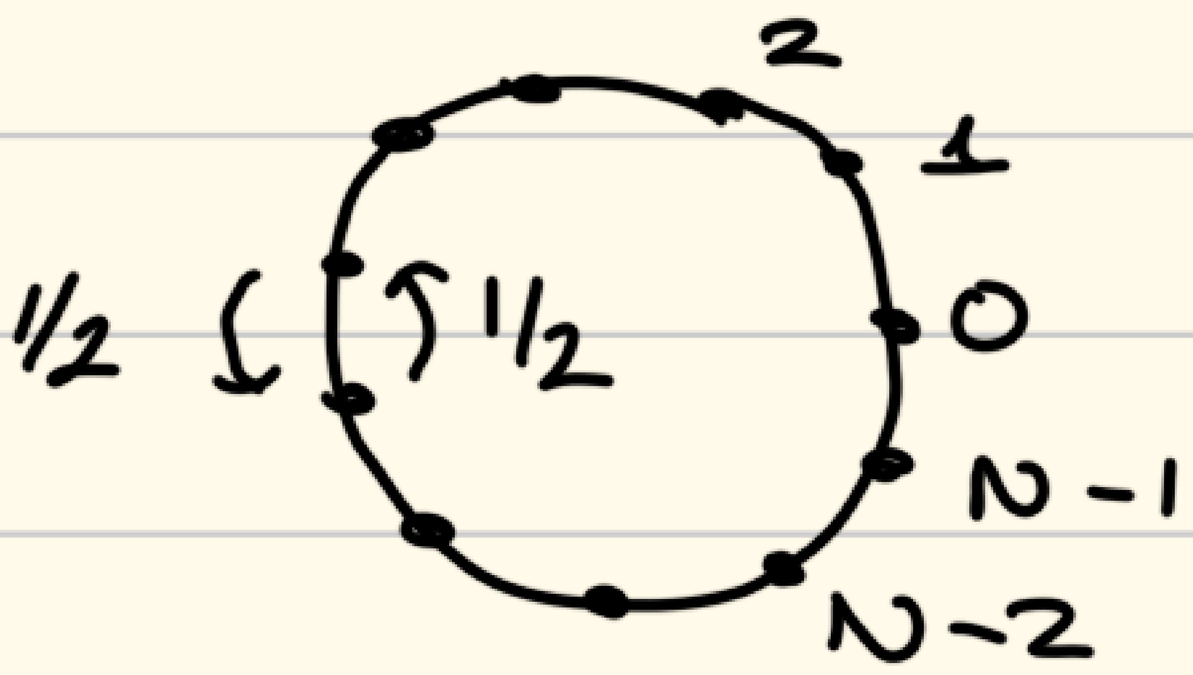
§1.6. Examples:

Simple Random Walk on a Graph (Ex. 5, §1.1)

Assume that the graph is connected so that it is irreducible. Let e denote the total number of edges in the graph and $d(v)$ the number of edges that have v as one of their edges. Since each edge has two ends $\sum_v d(v) = 2e$. $\pi(v) = d(v)/2e$ is the invariant probability measure for this chain.

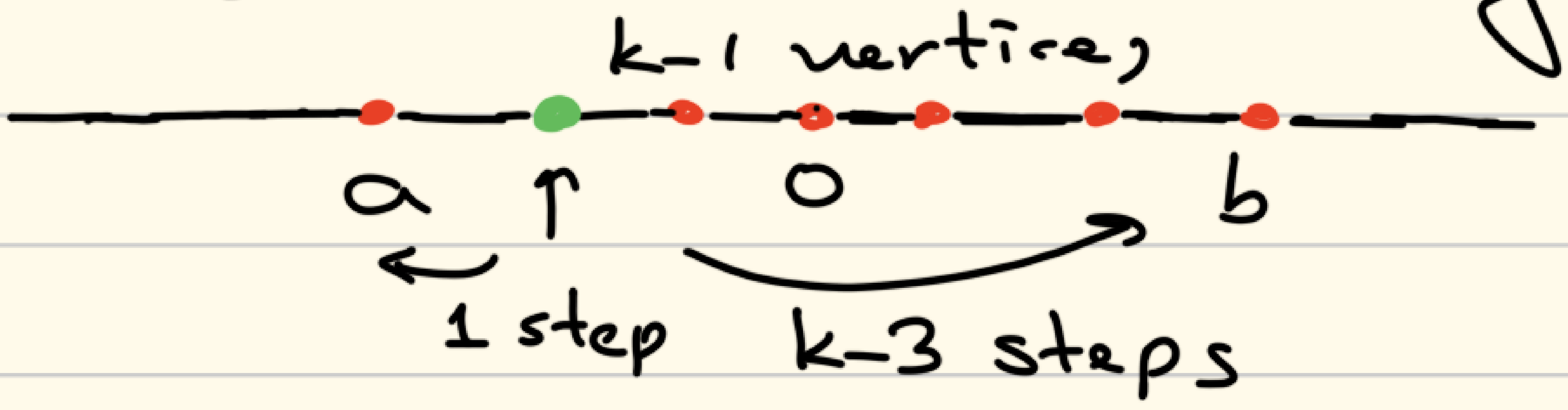
In other words, it is the unique distribution that is invariant under the automorphisms of the graps and satisfies $\sum_{v} \pi(v) = \sum_{v} d(v)/2e = 2e/2e = 1$.

Simple Random Walk on a Circle.



For any vertices i and j of an edge $p(i,j) = 1/2$.

Assume that $X_0 = 0$ and let T_k denote the first time at which the number of distinct points visited equals k. So T_N is the first time every vertex has been visited. By definition $T_1 = 0$ and $T_2 = 1$. Let's compute $r(k) = \mathbb{E}[T_k - T_{k-1}]$, for $k = 3, \dots, N$. Note that as the notation suggests $r(k)$ depends only on k, not N. Note that at time T_{k-1} the chain is at a boundary is a boundary point so that one of the neighbors of $X_{T_{k-1}}$ has been visited and the other has not. In the next step one will either visit the new point or will go to an interior point. So, the expected time that it takes the random walk from the interior point (next to the boundary point) to reach a boundary point is k-3:



Thus, $r(k) = 1 + \frac{1}{2}\left[(k-3) + r(k)\right]$ or $r(k) = k-1$.

It follows that

$$\mathbb{E}[T_N] = \underset{\overset{\shortparallel}{1}}{\mathbb{E}[T_2]} - \underset{\overset{\shortparallel}{0}}{\mathbb{E}[T_1]} + \mathbb{E}[T_3 - T_2] + \cdots + \mathbb{E}[T_k - T_{k-1}] + \cdots + \mathbb{E}[T_N - T_{N-1}].$$

$$= 1 + \sum_{k=3}^{N} \mathbb{E}[T_k - T_{k-1}]$$

$$= 1 + \sum_{k=3}^{N} (k-1) = 1 - \left(\frac{1}{2} N(N-1) - 1\right) = \frac{N(N-1)}{2}.$$

Urn Model. An urn contains N balls each colored either red or green. In each time period a ball is chosen at random from the urn and with probability $1/2$ is replaced with a ball of the other color; otherwise, the ball is returned to the urn. Let X_n be the number of red balls after n picks. X_n is an irreducible Markov chain with state space $\{0,1,\ldots,N\}$. The transition matrix is given by

$$p(j,j+1) = \frac{N-j}{2N}, \quad p(j,j-1) = \frac{j}{2N}, \quad p(j,j) = \frac{1}{2}, \quad j = 0,1,\ldots,N.$$

It turns out that the invariant probability is given by the binomial distribution

$$\pi(j) = \binom{N}{j} 2^{-N}:$$

$$(\pi P)(j) = \sum_{k=0}^{N} \pi(k) P(k,j)$$

$$= \pi(j-1)P(j-1,j) + \pi(j)P(j,j) + \pi(j+1)P(j+1,j)$$

$$= 2^{-N}\left[\binom{N}{j-1}\frac{N-(j-1)}{2N} + \binom{N}{j}\frac{1}{2} + \binom{N}{j+1}\frac{j+1}{2N}\right]$$

$$= 2^{-N}\left[\binom{N}{N-(j-1)}\frac{N-(j-1)}{2N} + \binom{N}{j}\frac{1}{2} + \binom{N}{j+1}\frac{j+1}{2N}\right]$$

Now use $\binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1}$ to get

$$= 2^{-N}\left[\binom{N-1}{N-1-(J-1)}\frac{1}{2} + \binom{N}{J}\frac{1}{2} + \binom{N-1}{J}\frac{1}{2}\right]$$

$$= 2^{-N}\left[\binom{N-1}{J-1}\frac{1}{2} + \binom{N}{J}\frac{1}{2} + \binom{N-1}{J}\frac{1}{2}\right]$$

$$= 2^{-N}\left[\binom{N}{J}\frac{1}{2} + \binom{N}{J}\frac{1}{2}\right]$$

$$= 2^{-N}\binom{N}{J} = \pi(J).$$

Note that this distribution is the same as the distributions for the number of heads in N flips of a fair coin.

<u>Cell Genetics.</u> We'll consider a Markov chain that models reproduction of cells. Suppose each cell contains N particles each of either one of two types, I or II. Let J be the number particles of type I. In reproduction, we assume that the cell duplicates itself and then splits, randomly distributing the particles. After duplication the cell has 2J particles of type I and 2(N−J) particles of type II. Then it selects N of these 2N particles for the next cell. So, by hypergeometric distribution we see that the transition probabilities are given as

$$p(J,k) = \frac{\binom{2J}{k}\binom{2(N-J)}{N-k}}{\binom{2N}{N}}.$$

The states 0 and N are absorbing. Eventually, all cells will have only particles of type I or II.

Assuming we start with a large number of cells each with J particles of type I. After a long time, the population will be full of cells of all with one type of particle. What fraction of these will be all type I. Since the fraction of type I particles does not change in this procedure we would expect that the fraction would be J/N. In other words, if we let $\alpha(J)$ be the probability that the Markov chain starting in state J eventually absorbed in state N, then we expect that $\alpha(J) = J/N$.

One can see that for $J=1,\dots,N-1$,

$$\alpha(J) = \sum_{k=0}^{N} p(k,J)\alpha(k),$$ and hence gives the absorbtion probabilities.

<u>Card Shuffling</u>. Consider a deck of cards numbered 1,2,...,n. At each time we will shuffle the cards by drawing a card at random and placing it at the top of the deck. This can be thought of as a Markov chain on S_n, the set of permutations of n elements. If λ denotes any permutation, and ν_J denotes the permutation corresponding to moving the $J^{\underline{th}}$ card to the top of the deck, then the transition probabilities for this chain are given by

$$p(\lambda, \nu_J\lambda) = \frac{1}{n}, \quad J=1,\dots,n.$$

This chain is irreducible and aperiodic. It is easy to verify that the unique invariant probability is the uniform measure on S_n, the measure that assigns the probability $1/n!$ to each permutation. Thus, if we start with any ordering of the cards, after enough moves of this kind the deck will be well shuffled. A much harder question is how many such moves are enough so that the deck of cards is well shuffled.

CHAPTER 2. Countable Markov Chains.

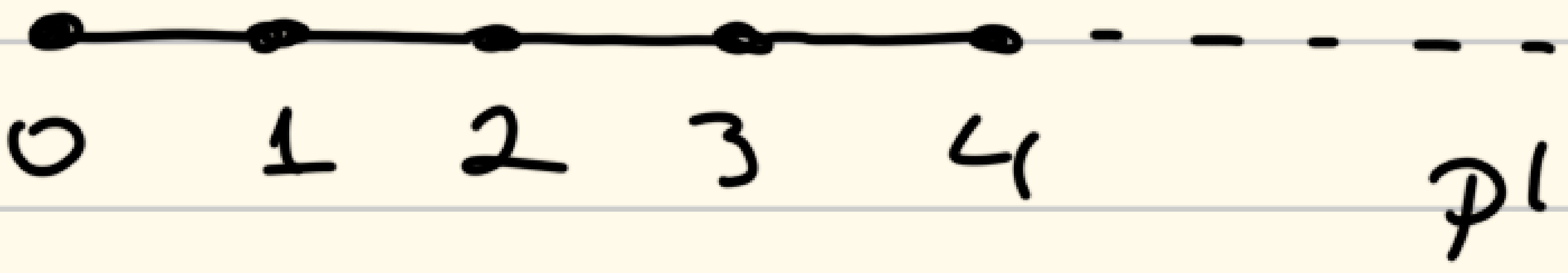
§2.1. Introduction. Let S be a countably infinite set and X_n a Markov chain on S. For $x, y \in S$ denote $p(x,y) = \mathbb{P}\{X_1 = y \mid X_0 = x\}$ transition probability from state x to state y. Then clearly "row sums" are equal to 1: $\sum_y p(x,y) = 1$.

Similarly, denote the n-step transition probabilities $p_n(x,y) = \mathbb{P}\{X_n = y \mid X_0 = x\}$. Again clearly we have, for $0 < m, n < +\infty$,

$$\begin{aligned} p_{m+n}(x,y) &= \mathbb{P}\{X_{m+n} = y \mid X_0 = x\} \\ &= \sum_{z \in S} \mathbb{P}\{X_{m+n} = y, X_m = z \mid X_0 = x\} \\ &= \sum_{z \in S} p_m(x,z)\, p_n(z,y). \end{aligned}$$

This equation, also known as Chapman-Kolmogorov equation, can be considered as the definition of matrix multiplication for infinite matrices.

Example 1. Random walk with partially reflecting Boundary at 0. Let $0 < p < 1$ and $S = \{0, 1, 2, \ldots\}$.



$p(x, x-1) = 1-p$, $p(x, x+1) = p$, $\forall x \in S$.

Also, $p(0,0) = 1-p$, $p(0,1) = p$.

Example 2. Let $S = \mathbb{Z}^d$ d-dimensional lattice.

$$p(x,y) = \begin{cases} 1/2d, & \text{if } |x-y| = 1 \\ 0 & \text{; otherwise.} \end{cases}$$

Example 3. Queueing Model.

Let X_n be the number of customers waiting in line for some service. We think of the first person in line as being served while all others are waiting their turn. During each time interval there is a probability p that a new customer arrives. With probability q, the service for the first customer is completed and that customer leaves the queue. There is no limit on the number of customers waiting in line. This is a Markov chain on $S = \{0, 1, 2, \dots\}$ and transition probabilities are given by

$$p(x, x-1) = q(1-p), \quad p(x,x) = pq + (1-p)(1-q), \text{ and}$$
$$p(x, x+1) = p(1-q), \quad x > 0; \text{ and } p(0,0) = 1-p, \ p(0,1) = p.$$

Clearly, the notion of communication of states and irreducibility of chains are defined the same way. We can also talk about periodicity of irreducible chains. Example 1 and 3 are aperiodic, whereas Example 2 has period 2. Unfortunately, an irreducible, aperiodic Markov chain with infinite state space converges to an equilibrium probability distribution.

§ 2.2. Recurrence and Transience.

Suppose X_n is an irreducible Markov chain with infinite countable state space S and transition probabilities $p(x,y)$. The chain X_n is called recurrent if for each state x, one has
$\mathbb{P}\{X_n = x, \text{ for infinitely many } n\} = 1$, i.e., the chain returns to x infinitely many times.

In an irreducible chain if one state is infinitely many times visited then all states are infinitely many times visited: Suppose $x \in S$ is visited infinitely many times $X_{k_i} = x$, for $k_1 < k_2 < k_3 < \cdots < k_n < \cdots \forall n$. If $y \in S$ is another state $p(x,y) > 0$ since the chain is irreducible. So $\mathbb{P}\{X_{k_{n+1}} \neq y\} = 1 - p(x,y)$. Hence $\mathbb{P}\{X_{k_{n+1}} \neq y, \forall n = 1, 2, \ldots\} = \lim_{n \to \infty} (1 - p(x,y))^n = 0$, since $0 \leq 1 - p(x,y) < 1$.

Hence, it follows that if the chain is not recurrent then every state is visited finitely many times. In this case, we call the chain transient. In this section, we'll study two criteria for determining this.

Fix a state x and assume that $X_0 = x$. Define the random variable

$$R = \sum_{n=0}^{\infty} I\{X_n = x\},$$ where I is the indicator function so that R counts the visits of X_n to x. If the chain is recurrent then clearly $R = +\infty$. Otherwise, $R < +\infty$ with probability 1.

$$\text{Clearly, } \mathbb{E}[R] = \sum_{n=0}^{\infty} E[I\{I_n = x\}] = \sum_{n=0}^{\infty} \mathbb{P}\{X_n = x\}$$
$$= \sum_{n=0}^{\infty} P_n(x,x).$$

Claim: $\sum_{n=0}^{\infty} P_n(x,x) < +\infty$ if and only if the chain is transient.

Now, define $T = \min\{n > 0, X_n = 0\}$. $T = \infty$ if the chain never returns to x. Suppose $\mathbb{P}\{T < \infty\} = 1$. Then with probability the chain returns to x and by continuing it returns to x infinitely often with probability 1. Hence, the chain is recurrent. Now suppose $\mathbb{P}\{T<\infty\}=q<1$. Let's compute the distribution of the random variable R.

Clearly, $R=1$ if and only if the chain never returns. Hence, $\mathbb{P}\{R=1\}=1-q$. If $m>1$, then $R=m$ if and only if the chain returns $m-1$ times and then never returns. Hence
$\mathbb{P}\{R=m\}=q^{m-1}(1-q)$. Therefore, in the transient case, $q<1$,

$$E[R]=\sum_{m=1}^{\infty} m\,\mathbb{P}\{R=m\}=\sum_{m=1}^{\infty} m q^{m-1}(1-q)=\frac{1}{1-q}<\infty.$$

So, we've proved the above claim.

Example. Simple Random Walk in $\mathbb{Z}^d$.

First let $d=1$ and consider the Markov chain with $p(x,x+1)=p(x,x-1)=1/2$. Let $X_0=0$. Since the chain has period 2, $p_n(0,0)=0$ for odd n. Let's compute $p_{2n}(0,0)$. If the walker will be back at state 0 after 2n steps then the walker must take n steps the left and n steps to the right. Each such path has probability $(1/2)^{2n}$ and clearly there are $\binom{2n}{n}$ such paths. So $p_{2n}(0,0)=\binom{2n}{n} 1/2^{2n}=\frac{(2n)!}{(n!)^2}\frac{1}{2^{2n}}$.

To compute this we use so called Stirling's formula $n!\sim\sqrt{2\pi n}\, n^n e^{-n}$, which means that $\lim_{n\to\infty} n!/(\sqrt{2\pi n}\, n^n e^{-n})=1$. Hence,

$$p_{2n}(0,0)\sim\frac{(2n)!}{(n!)^2}\frac{1}{2^{2n}}\sim\frac{\sqrt{4\pi n}\,(2n)^{2n}e^{-2n}}{(\sqrt{2\pi n})^2 n^{2n}e^{-2n}}\frac{1}{2^{2n}},$$

so that $P_{2n}(0,0) \sim \frac{1}{\sqrt{\pi n}}$. In particular, $\sum P_{2n}(0,0) = +\infty$, so that the chain is recurrent.

Now, let's take $d>1$. For $|x-y|=1$, $p(x,y) = 1/2d$ and for $|x-y| \neq 1$, let $p(x,y)=0$. Again assume $X_0=(0,\dots,0)$. As before $P_n(0,0)=0$ for odd n. What about $P_{2n}(0,0)$? Then by Law of Large Numbers, we expect, for large n, $2n/d$ of these $2n$ steps in each component. We'll need the number of steps in each component to be even if we have any chance of being at 0 in n steps. For large n, the probability of this occuring is about $(1/2)^{d-1}$ (since the total number of steps is $2n$, an even number, if $d-1$ directions have all even steps, then the last one must be even too). Now using the computation of the case $d=1$, we see that

$$P_{2n}(0,0) \sim (1/2)^{d-1} \left(\frac{d}{n\pi}\right)^{d/2}.$$

$$\sum_{n=0}^{\infty} P_{2n}(0,0) = \infty$$ if and only if $n=1$ or 2 and thus we have proved the following.

<u>Fact:</u> Simple random walk on $\mathbb{Z}^d$ is recurrent if $d=1$ or 2, and is transient if $d \geq 3$.

Next we'll see another method for determining recurrence or transience. Again assume X_n is an irreducible chain and $z \in S$ be any fixed state. For any other $x \in S$, define

$$\alpha(x) = \mathbb{P}\{X_n = z \text{ for some } n \geq 0 \mid X_0 = x\}.$$

Clearly, $\alpha(z)=1$. If the chain is recurrent then $\alpha(x)=1$, for all x. So, if the chain is transient there is at least one state x with $\alpha(x)<1$. In fact, there are points with $\alpha(x)$ is arbitrarily small.

Let $x \neq z$, then

$$\alpha(z) = \mathbb{P}\{X_n = z \text{ for some } n\geq 0 \mid X_0 = x\}$$
$$= \mathbb{P}\{X_n = z \text{ for some } n\geq 1 \mid X_0 = x\}$$
$$= \sum_{y\in S} \mathbb{P}\{X_1 = y \mid X_0 = x\}\, \mathbb{P}\{X_n = z, \text{for some } n \mid X_1 = y\}$$
$$= \sum_{y\in S} p(x,y)\,\alpha(y).$$

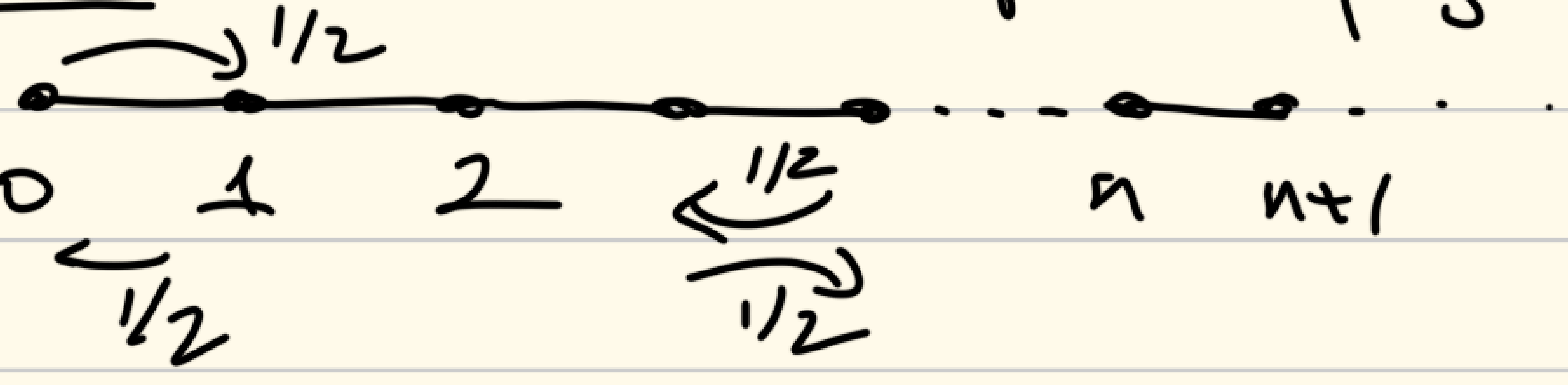
$$(*)\left\{\begin{array}{l}\text{Hence, } \alpha(x) \text{ satisfies, } 0\leq\alpha(x)\leq 1,\ \alpha(z)=1, \\ \inf\{\alpha(x) \mid x\in S\}=0 \text{ and } \alpha(x) = \sum_{y\in S} p(x,y)\alpha(y), \\ x\neq z.\end{array}\right.$$

We'll see in Chapter 5, if X_n is transient, then there is a unique solution to $(*)$. Also, if X_n is recurrent then there is no solution $\alpha(x)$ to $(*)$.

Thus, this gives the following:

<u>Fact</u>: An irreducible Markov chain is transient if and only if for any z we can find a function $\alpha(x)$ satisfying $(*)$.

<u>Example</u>. Consider Example 1 of §2.1.



Let $z=0$ and let's try to find $\alpha(x)$ satisfying $(*)$.

The third equation in (*) yields
$\alpha(x) = (1-p)\alpha(x-1) + p\alpha(x+1)$, $x > 0$. We can solve $\alpha(x)$ as $\alpha(x) = c_1 + c_2\left(\frac{1-p}{p}\right)^x$, $p \neq 1/2$, and
$\alpha(x) = c_1 + c_2 x$, if $p = 1/2$.

The first condition of (*) gives $\alpha(0) = 1$ and thus we get
$$\alpha(x) = (1-c_2) + c_2\left(\frac{1-p}{p}\right)^x, \; p \neq 1/2 \text{ and}$$

$$\alpha(x) = 1 + c_2 x, \; p = 1/2.$$

If $p = 1/2$ and $c_2 \neq 0$, then the solution is not bunded hence does not satisfy (*). Similarly, if $p < 1/2$ then again the solution is unbounded. So, we conclude that the chain is recurrent for $p \leq 1/2$. For $p > 1/2$, there is a solution to (*). The second condition in (*) boils down to $\alpha(x) \to 0$ as $x \to \infty$. Thus, we get $\alpha(x) = \left(\frac{1-p}{p}\right)^x$. Thus, for $p > 1/2$, the chain is transient.

§2.3. Positive Recurrence and Null Recurrence

Suppose X_n is an irreducible, aperiodic Markov chain on the infinite state space S. We'll now investigate when a limiting probability distribution exists: In other a function $\pi : S \to [0,1]$ s.t. for all $x, y \in S$, $\lim_{n\to\infty} p_n(y,x) = \pi(x)$.
If X_n is transient, then $\lim_{n\to\infty} p_n(x,y) = 0$ so that in this case, no such limiting distribution exist.
On the other hand, we may have $\lim_{n\to\infty} p_n(x,y) = 0$

for a recurrent chain also. Such recurrent chains are called null recurrent. Otherwise, it is called positive recurrent. Positive recurrent chains are very much like finite Markov chains. For example, an irreducible, aperiodic, positive recurrent chain then for every x, y, the limit $\lim_{n\to\infty} p_n(y,x) = \pi(x) > 0$, exists and is independent of the initial state y. It follows that $\pi(x)$ is invariant: $\sum_{y\in S} \pi(y) p(y,x) = \pi(x)$.

Moreover, if we consider the return time to a state x, $T = \min\{n>0 \mid X_n = x\}$, then for a positive recurrent chain $E[T \mid X_n = x] = 1/\pi(x)$.
If X_n is null recurrent, then $T<\infty$ with probability 1, but $E[T] = \infty$. If X_n is transient, then $T=\infty$ with positive probability.

Example. Consider again the random walk example with partially reflecting boundary on $S=\{0,1,2,\dots\}$. Let us try to find a probability distribution that satisfies $\sum_{y\in S} \pi(y) p(y,x) = \pi(x)$ and $\sum_{x\in S} \pi(x) = 1$.

The first sum gives $\pi(x+1)(1-p) + \pi(x-1)p = \pi(x)$, $x>0$, and $\pi(1)(1-p) + \pi(0)(1-p) = \pi(0)$. Solving these difference equations we obtain the solutions

$$\pi(x) = c_1 + c_2\left(\frac{p}{1-p}\right)^x,\ p \neq 1/2,\ \pi(x) = c_1 + c_2 x,\ p = 1/2.$$

Now the initial condition gives $\pi(0) = [(1-p)/p]\pi(1)$. Plugging this into the above gives $\pi(x) = c_2 (p/1-p)^x$, $p \neq 1/2$ and $\pi(x) = c_1$, $p = 1/2$.

Finally, the condition $\sum \pi(x) = 1$. For $p = 1/2$ clearly there is <u>no</u> solution. for $p > 1/2$, $\sum [p/(1-p)]^x = \infty$ and we cannot find such a c_2, hence <u>no</u> solution. Finally, if $p < 1/2$, the sum is finite and we can choose

$$\pi(x) = \left(\frac{p}{1-p}\right)^x \left[\sum_{y=0}^{\infty} \left(\frac{p}{1-p}\right)^y\right]^{-1} = \frac{1-2p}{1-p} \left(\frac{p}{1-p}\right)^x.$$

<u>Summary</u>: positive recurrent if $p < 1/2$
null recurrent if $p = 1/2$
transient if $p > 1/2$.

§2.4. Branching Process. (A stochastic process for Population Growth)

Consider a population of individuals and let X_n denote the number of individuals at time n. At each time interval each individual will produce a random number of offsprings, after that the individual dies and leaves the system. The system obeys the following rules:

1) Each individual produces offsprings with the same probability distribution: There are given nonnegative numbers $p_0, p_1, \ldots$ summing to 1 such that the probability that an individual produces exactly k offsprings is p_k.

2) The individuals produce offsprings independently.

X_n, the number of individuals at stage n is then a Markov chain on the state space $\{0, 1, 2, \ldots\}$ so that 0 is an absorbing state. Let's write the transition probabilities for this chain.

Suppose $X_n = k$. Then k individuals produce offspring for the $(n+1)$-st generation. If $Y_1, \dots, Y_k$ are independent random variables each with distribution $\mathbb{P}\{Y_i = j\} = p_j$, then

$$p(k, j) = \mathbb{P}\{X_{n+1} = j \mid X_n = k\} = \mathbb{P}\{Y_1 + \cdots + Y_k = j\}.$$

The actual distribution of $Y_1 + \cdots + Y_k$ can be computed by convolutions but instead we'll use a shortcut. Let $\mu = \sum_{i=0}^{\infty} i p_i$, the expectation of number of each Y_i. Then

$$\mathbb{E}[X_{n+1} \mid X_n = k] = \mathbb{E}[Y_1 + \cdots + Y_k] = k\mu.$$

Now,
$$\begin{aligned}\mathbb{E}[X_n] &= \sum_{k=0}^{\infty} \mathbb{P}\{X_{n-1} = k\} \mathbb{E}[X_n \mid X_{n-1} = k] \\ &= \sum_{k=0}^{\infty} k\mu \, \mathbb{P}\{X_{n-1} = k\} \\ &= \mu \, \mathbb{E}[X_{n-1}].\end{aligned}$$

Hence, by induction $\mathbb{E}[X_n] = \mu^n \mathbb{E}[X_0]$.

Some conclusions:

1) If $\mu < 1$, then
$$\begin{aligned}\mu^n \mathbb{E}[X_0] = \mathbb{E}[X_n] &= \sum_{k=0}^{\infty} k \, \mathbb{P}\{X_n = k\} \\ &\geq \sum_{k=1}^{\infty} \mathbb{P}\{X_n = k\} \\ &= \mathbb{P}\{X_n \geq 1\}.\end{aligned}$$

So,
$$\lim_{n\to\infty} \mathbb{P}\{X_n = 0\} = \lim_{n\to\infty} 1 - \mathbb{P}\{X_n \geq 1\} = 1 - \lim_{n\to\infty} \mu^n \mathbb{E}(X_0)$$
$$= 1$$

so that the population extincts.

For $\mu=1$ the population size remains constant while for $\mu>1$, the expected population grow. Below we investigate the probability that the population dies out. For meaningful results assume that $p_0>0$ and $p_0+p_1<1$.

Let $a_n(k)=\mathbb{P}\{X_n=0 / X_0=k\}$ and $a(k)=\lim_{n\to\infty} a_n(k)$, the probability that the population eventually dies out assuming that initially there are k individuals.

If the population has k individuals at a certain time, then the only way for the population to die out is for all k individuals to die out. Since the branches act independently we deduce that $a(k)=(a(1))^k$.

The number $a(1)$, which we'll denote simply as a, is called the extinction probability.

Assume now that $X_0=1$. Then

$$\begin{aligned} a &= \mathbb{P}\{\text{population dies out} \mid X_0=1\} \\ &= \sum_{k=0}^{\infty} \mathbb{P}\{X_1=k \mid X_0=1\}\, \mathbb{P}\{\text{population dies out} \mid X_1=k\} \\ &= \sum_{k=0}^{\infty} p_k\, a(k) = \sum_{k=0}^{\infty} p_k\, a^k. \end{aligned}$$

Definition: If X is a random variable taking values in $\{0,1,2,\ldots\}$, the generating function of X is the function

$$\phi(s)=\phi_X(s)=\mathbb{E}[s^X]=\sum_{k=0}^{\infty} s^k\, \mathbb{P}\{X=k\}.$$

Hence, we have $a = \phi_X(a)$.

Clearly, $\phi(s)$ is an increasing function on $[0,\infty)$ and $\phi(S) = \mathbb{P}\{X = 0\}$ and $\phi(1) = \sum_{k=0}^{\infty} \mathbb{P}\{X=k\} = 1$.

Also, $\phi'(s) = \sum_{k=1}^{\infty} k\, s^{k-1}\, \mathbb{P}\{X=k\}$, $\phi''(s) = \sum_{k=2}^{\infty} k(k-1) s^{k-2} \mathbb{P}\{X=k\}$.

Hence, $\phi'(1) = \sum_{k=1}^{\infty} k\, \mathbb{P}\{X=k\} = \mathbb{E}[X]$. $(*)$

Also, for $s > 0$, if $\mathbb{P}\{X \geq 2\} > 0$, then $\phi''(s) > 0$. $(**)$

If $X_1, \dots, X_m$ are independent random variables taking values on $\{0,1,2,\dots\}$ then

$$\phi_{X_1 + \dots + X_m}(s) = \mathbb{E}\left[s^{X_1 + \dots + X_m}\right] = \mathbb{E}\left[s^{X_1} \cdots s^{X_m}\right]$$

$$= \phi_{X_1}(s) \cdots \phi_{X_m}(s).$$

Now, let's return back to the equation $a = \phi(a)$. As we've seen above $a=1$ is a solution, since $\phi(1) = 1$. We want to find all other solutions of the equation $a = \phi(a)$.

Assume that $X_0 = 1$. The generating function of X_0 is $\phi(s) = \phi_{X_0}(s) = \mathbb{E}[s^{X_0}] = \sum_{k=0}^{\infty} s^k \mathbb{P}\{X_0 = k\}$

Let $\phi^n(a)$ be the generating function of X_n.

Claim: $\phi^n(a) = \phi(\phi^{n-1}(a))$.

Proof:
$$\phi^n(a) = \sum_{k=0}^{\infty} \mathbb{P}\{X_n = k\} a^k$$
$$= \sum_{k=0}^{\infty} \left[\sum_{J=0}^{\infty} \mathbb{P}\{X_1 = J\} \mathbb{P}\{X_n = k | X_1 = J\} \right] a^k$$
$$= \sum_{J=0}^{\infty} P_J \sum_{k=0}^{\infty} \mathbb{P}\{X_{n-1} = k | X_0 = J\} a^k$$

Now, if $X_0 = J$, then X_{n-1} is the sum of J independent random variables each with the distribution of X_{J-1} given $X_0 = 1$. Hence, the sum over k is just the generating function of the sum of J independent random variables each with generating function $\phi^{n-1}(a)$ and hence $\sum_{k=0}^{\infty} \mathbb{P}\{X_{n-1} = k | X_0 = J\} a^k = [\phi^{n-1}(a)]^J$ and
$$\phi^n(a) = \sum_{J=0}^{\infty} P_J [\phi^{n-1}(a)]^J = \phi(\phi^{n-1}(a)).$$

So, we have a recursive way to find $\phi^n(a)$ and hence to find $a_n(1) = \mathbb{P}\{X_n = 0 | X_0 = 1\} = \phi^n(0)$.

Claim: The extinction probability a is the smallest positive root of the equation $a = \phi^n(a)$.

Proof: We know that a is a solution of $a = \phi^n(a)$. Let $\hat{a}$ be the smallest root of the equation. We'll show that $a_n = \mathbb{P}\{X_n = 0\} < \hat{a}$, $\forall n$, which implies that $a = \lim a_n \leq \hat{a}$. This is clearly true for $n = 0$ since $a_0 = 0$. Now assume that $a_{n-1} \leq \hat{a}$.

Then $\mathbb{P}\{X_n = 0\} = \phi^n(0) = \phi(\phi^{n-1}(0)) = \phi(a_{n-1}) \leq \phi(\hat{a}) = \hat{a}$, which finishes the proof, since ϕ is an increasing function. This finishes the proof. ■

Example 1. Suppose $P_0 = 1/4$, $P_1 = 1/4$, $P_2 = 1/2$. Then $\mu = 5/4$ and $\phi(a) = 1/4 + 1/4\, a + 1/2\, a^2$.
$a = \phi(a)$ gives $a = 1$ or $1/2$. Hence, the extinction probability is $1/2$.

Example 2. Suppose $P_0 = 1/2$, $P_1 = 1/4$, $P_2 = 1/4$. Then $\mu = 3/4$ and $\phi(a) = 1/2 + 1/4 + 1/4\, a^2$.
$a = \phi(a) \Rightarrow a = 1, 2$ so that the extinction probability is 1.

Example 3. Suppose $P_0 = 1/4$, $P_1 = 1/2$, $P_2 = 1/4$. Then $\mu = 1$, $\phi(a) = 1/4 + 1/2\, a + 1/4\, a^2$. This time $a = \phi(a)$ gives $a = 1, 1$ so that $a = 1$.

We've already seen that if $\mu < 1$ then $a = 1$. Suppose now $\mu = 1$. By ($*$, p. 38) we have $\phi'(1) = 1$ and thus by ($**$, p. 38) $\phi'(s) < 1$, for $s > 1$. Hence, for any $s < 1$,
$1 - \phi(s) = \int_s^1 \phi'(s)\, ds < 1 - s \Rightarrow \phi(s) > s$. Thus, if $\mu = 1$ then the extinction probability is 1.

Now assume $\mu > 1$. Then $\phi'(s) > 1$ and hence there must be some $s < 1$ with $\phi(s) < s$. However, $\phi(0) > 0$. So by continuity of ϕ there must be some $a \in (0, s)$ with $a = \phi(a)$. Since, $\phi''(s) > 0$ for $s \in (0, 1)$, the curve is convex and there can be at most one $s \in (0,1)$ with $\phi(s) = s$. In this case, with positive probability the population lives forever.

Summary:

<u>Theorem</u>: If $\mu \leq 1$ and $p_0 > 0$, the extinction probability is $a = 1$, so that the population eventually dies out. If $\mu > 1$, then the extinction probability $a < 1$ and equals to the root of the equation $t = \phi(t)$ with $0 < t < 1$.

CHAPTER 3. Continuous-time Markov Chains

§3.1. Poisson Process. Let X_t denote the number of customers arriving at a store by time t, where $t \in [0, \infty)$ is considered as a continuous parameter. Suppose we make three assumption given below:

1) The number of customers arriving during a certain time interval does not affect the number of arriving customer during another time interval.

2) The "average" rate at which customers arrive remains constant.

3) Customers arrive one at a time.

A mathematical formulation of the first assumption is as follows: For $s_1 \leq t_1 \leq s_2 \leq t_2 \leq \cdots \leq s_n \leq t_n$, the variables $X_{t_1} - X_{s_1}, \ldots, X_{t_n} - X_{s_n}$ are all independent. For the second assumption let λ be the rate at which customers arrive, i.e., on the average we expect λt customers at time t. In a small interval of time $[t, t+\Delta t]$ we expect that $\lambda \Delta t$ new customers to arrive.

The third assumption can be formulated in several ways:

$$\mathbb{P}\{X_{t+\Delta t} = X_t\} = 1 - \lambda \Delta_t + o(\Delta t)$$
$$\mathbb{P}\{X_{t+\Delta t} = X_t + 1\} = \lambda \Delta_t + o(\Delta t)$$
$$\mathbb{P}\{X_{t+\Delta t} \geq X_t + 2\} = o(\Delta t),$$

where $o(\Delta t)$ represents a function that is much smaller that Δt so that $\lim_{\Delta t\to 0}\frac{o(\Delta t)}{\Delta t}=0$.

A stochastic process X_t with $X_0=0$ satisfying these assumptions is called a Poisson process with rate parameter λ.

Let's now determine the distribution of X. We'll see two methods for this. For the first one, choose a big integer n and write X as a sum of n independent identically distributed variables:

$$X_t=\sum_{j=1}^{n}\left(X_{jt/n}-X_{(j-1)t/n}\right) \qquad (*)$$

Note that $\mathbb{P}\{X_{jt/n}-X_{(j-1)t/2}\geq 2 \text{ for some } j\leq n\}$

$$\leq\sum_{j=1}^{n}\mathbb{P}\{X_{jt/n}-X_{(j-1)t/n}\geq 2\}$$

$$=n\,\mathbb{P}\{X_{t/n}-\underset{0}{\underset{\shortparallel}{X_0}}\geq 2\},$$ where the last term goes to zero as $n\to\infty$ by the third assumption ($\mathbb{P}\{X_{t+\Delta t}\geq X_t+2\}=o(\Delta t)$).

This implies that we can approximate the sum $(*)$ by a sum of independent random variables which equals 1 with probability $\lambda(t/n)$ and equals 0 with probability $1-\lambda(t/n)$. By the binomial distribution formula,

$$\mathbb{P}\{X_t=k\}\approx\binom{n}{k}\left(\frac{\lambda t}{n}\right)^k\left(1-\frac{\lambda t}{n}\right)^{n-k}.$$ Indeed, one

one can show that $\mathbb{P}\{X_t=k\}=\lim_{n\to\infty}\binom{n}{k}\left(\frac{\lambda t}{n}\right)^k\left(1-\frac{\lambda t}{n}\right)^{n-k}$.

Now, let's compute the limit:

Note that $\lim_{n\to\infty} \binom{n}{k} n^{-k} = \lim_{n\to\infty} \frac{n(n-1)\cdots(n-k+1)}{k!\, n^k} = \frac{1}{k!}$

and $\lim_{n\to\infty} \left(1-\frac{\lambda t}{n}\right)^{n-k} = \lim_{n\to\infty} \left(1-\frac{\lambda t}{n}\right)^{n} \lim_{n\to\infty} \left(1-\frac{\lambda t}{n}\right)^{-k}$

$$= e^{-\lambda t} \cdot 1 = e^{-\lambda t}.$$

Hence, $\mathbb{P}\{X_t = k\} = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$, i.e., X_t has a Poisson distribution with parameter λt.

The second method is as follows: Let $P_k(t) = \mathbb{P}\{X_t = k\}$. Note that $P_0(0) = 1$ and $P_k(0) = 0$, for $k > 0$. Using again the third assumption we can derive a system of differential equations. Namely,

$$P_k'(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \left(\mathbb{P}\{X_{t+\Delta t} = k\} - \mathbb{P}\{X_t = k\}\right) \text{ and}$$

$$\begin{aligned}\mathbb{P}\{X_{t+\Delta t} = k\} = {} & \mathbb{P}\{X_t = k\}\mathbb{P}\{X_{t+\Delta t} = k \mid X_t = k\} \\ & + \mathbb{P}\{X_t = k-1\}\mathbb{P}\{X_{t+\Delta t} = k \mid X_t = k-1\} \\ & + \mathbb{P}\{X_t \leq k-2\}\mathbb{P}\{X_{t+\Delta t} = k \mid X_t \leq k-2\}\end{aligned}$$

$$= P_k(t)(1-\lambda \Delta t) + P_{k-1}(t)\lambda \Delta t + o(\Delta t)$$

$$\Rightarrow \mathbb{P}\{X_{t+\Delta t} = k\} - \mathbb{P}\{X_t = k\} = \lambda \Delta t \left(P_{k-1}(t) - P_k(t)\right) + o(\Delta t)$$

$$\Rightarrow P_k'(t) = \lambda\left(P_{k-1}(t) - P_k(t)\right)$$

One can solve this system recursively:

$P_0'(t) = -\lambda P_0(t)$, $P_0(0)=1$ (taking $P_{-1}(t)=0$.)

$\Rightarrow P_0(t) = e^{-\lambda t}$. Now, let $f_k(t) = e^{\lambda t} P_k(t)$. Then

$$\begin{aligned} f_k'(t) &= \lambda e^{\lambda t} P_k(t) + e^{\lambda t} P_k'(t) \\ &= \cancel{\lambda e^{\lambda t} P_k(t)} + e^{\lambda t} (\lambda P_{k-1}(t) - \cancel{\lambda P_k(t)}) \\ &= \lambda e^{\lambda t} P_{k-1}(t) \\ &= \lambda f_{k-1}(t) \text{ and } f_k(0)=0, \text{ for } k \geq 1. \end{aligned}$$

So, $f_1'(t) = \lambda f_0(t) = \lambda e^{\lambda t} P_0(t) = \lambda e^{\lambda t} e^{-\lambda t} = \lambda$.

So, $f_1(t) = \lambda t + C$. $f_1(0) = 0 \Rightarrow C=0 \Rightarrow f_1(t) = \lambda t$.

Similarly, $f_2'(t) = \lambda f_1(t) = \lambda^2 t \Rightarrow f_2(t) = \frac{\lambda^2 t^2}{2} + C$.

$f_2(0) = 0 \Rightarrow C=0 \Rightarrow f_2(t) = \frac{1}{2}\lambda^2 t^2$. Inductively

$f_k(t) = \lambda^k t^k / k!$ and hence, $P_k(t) = e^{-\lambda t} (\lambda t)^k / k!$.

One can view the Poisson process considering the waiting times between cutomers. Let T_n, $n=1,2,\dots$ be the time between the arrivals of the $(n-1)$st and the nth cutomers. Also let $Y_n = T_1 + \dots + T_n$ be the total amount of time until the nth customers arrives. Clearly, we can write

$$Y_n = \inf\{t \mid X_t = n\}, \quad T_n = Y_n - Y_{n-1}.$$

We'll assume that T_i's are independent an identically distributed. This can be written in mathematical terms as

$$\mathbb{P}\{T_i \geqslant s+t \mid T_i \geqslant s\} = \mathbb{P}\{T_i \geqslant t\}.$$

$$\Rightarrow \mathbb{P}\{T_i \geqslant s+t, T_i \geqslant s\} = \mathbb{P}\{T_i \geqslant t\}\mathbb{P}\{T_i \geqslant s\}$$

$$\Rightarrow \mathbb{P}\{T_i \geqslant s+t\} = \mathbb{P}\{T_i \geqslant t\}\mathbb{P}\{T_i \geqslant s\}$$

If $g(t) = \mathbb{P}\{T_i \geqslant t\}$, then the above equation becomes $g(t+s) = g(t)g(s)$.

[$g(t+s) = g(t)g(s) \Rightarrow g(0) = 1$ and $g(n) = g(1)^n$. Say $\lambda = g(1)$. Then $g(n) = \lambda^n$ and

$$\lambda = g(1) = g(1/n + \cdots + 1/n) = g(1/n)^n \text{ so that}$$

$$g(1/n) = \lambda^{1/n}. \text{ Hence, } g(t) = \lambda^t \text{ for } t \in \mathbb{Q}.$$

If we further assume that g is continuous then $g(t) = \lambda^t = e^{-bt}$, $b = -\ln \lambda$.]

The only solution to this functional equation is that $g(t) = e^{-bt}$.

So, we may assume that $\mathbb{P}\{T_i \geqslant t\} = e^{-bt}$. Hence, $\mathbb{P}\{T_i \leqslant t\} = 1 - e^{-bt}$. $(t \geqslant 0, = 0 \text{ if } t < 0)$

If we let, $F(z) = \mathbb{P}\{T_i \leqslant z\} = 1 - e^{-bt}$, we can compute that

$$\mathbb{E}[T_i] = \int_0^\infty z F'(z)\,dz = \int_0^\infty z\, b e^{-bt}\,dz = b\frac{z e}{-b}\Big|_0^\infty - \int_0^\infty b\frac{e^{-bz}}{-b}\,dz = 0 - \frac{e^{-bz}}{b}\Big|_0^\infty = \frac{1}{b}.$$

So, $\mathbb{E}[Y_n] = \mathbb{E}[T_1 + \cdots + T_n] = n/b$.

Moreover, $X_t = n$ if $Y_n \leq t < Y_{n+1}$.

§3.2. Finite State Space.

We'll start by establishing some facts about random variables. Suppose $T_1, \ldots, T_n$ are independent random variables, each of which is an exponential with rates $b_1, \ldots, b_n$, respectively. Set $T = \min\{T_1, \ldots, T_n\}$. We may consider T_i's as the times of n alarm clocks goes off.

Note that
$$\begin{aligned}\mathbb{P}\{T \geq t\} &= \mathbb{P}\{T_1 \geq t, \ldots, T_n \geq t\} \\ &= \mathbb{P}\{T_1 \geq t\} \cdots \mathbb{P}\{T_n \geq t\} \\ &= e^{-b_1 t} \cdots e^{-b_n t} \\ &= e^{-(b_1 + \cdots + b_n)t},\end{aligned}$$
so that T is also exponential with parameter $b_1 + \cdots + b_n$. Now,

$$\begin{aligned}\mathbb{P}\{T_1 = T\} &= \int_0^\infty \mathbb{P}\{T_2 \geq t, \ldots, T_n \geq t\}\, d\mathbb{P}\{T_1 \leq t\} \\ &= \int_0^\infty e^{-(b_2 + \cdots + b_n)t}\, b_1 e^{-b_1 t}\, dt, \text{ since } \mathbb{P}\{T_1 \leq t\} = 1 - e^{-b_1 t} \\ &= \frac{b_1}{b_1 + \cdots + b_n}.\end{aligned}$$

Now suppose we have a finite state space S. A continuous-time process X_t on S is said to have the Markov property if
$$\mathbb{P}\{X_t = y \mid X_r, 0 \leq r \leq s\} = \mathbb{P}\{X_t = y \mid X_s\},$$
and is time-homogeneous
$$\mathbb{P}\{X_t = y \mid X_s = x\} = \mathbb{P}\{X_{t-s} = y \mid X_0 = x\}.$$

For each $x, y \in S$ with $x \neq y$, we assign a non-negative real number $\alpha(x,y)$ that we think of as the rate the chain moves from x to y. Also let, $\alpha(x) = \sum_{y \neq x} \alpha(x,y)$.

A (time-homogeneous) continuous-time Markov chain with rates α is a stochastic process X_t taking values in S satisfying

$$\mathbb{P}\{X_{t+\Delta t} = x \mid X_t = x\} = 1 - \alpha(x)\Delta t + o(\Delta t),$$
$$\mathbb{P}\{X_{t+\Delta t} = x \mid X_t = y\} = \alpha(y,x)\Delta t + o(\Delta t), \; y \neq x.$$

In other words, the probability that the chain in state y jumps to a different state x in a small time interval of length Δt is about $\alpha(y,x)\Delta t$. Now, we'll obtain a system of differential equations that describes the process: Let $p_x(t) = \mathbb{P}\{X_t = x\}$. Now,

$$\mathbb{P}\{X_{t+\Delta t} = x, \; X_t = x\} = \mathbb{P}\{X_t = x\}\left(1 - \alpha(x)\Delta t + o(t)\right), \text{and}$$
$$\mathbb{P}\{X_{t+\Delta t} = x, \; X_t = y\} = \mathbb{P}\{X_t = y\}\left(\alpha(y,x)\Delta t + o(t)\right), \; y \neq x.$$

$$\Rightarrow \mathbb{P}\{X_{t+\Delta x} = x\} - \mathbb{P}\{X_t = x\} = \left[-\mathbb{P}\{X_t = x\}\alpha(x) + \sum_{y \neq x} \mathbb{P}\{X_t = y\}\alpha(y,x)\right]\Delta t + o(t)$$

$$\Rightarrow p_x'(t) = -\alpha(x)\, p_x(t) + \sum_{y \neq x} \alpha(y,x) p_y(t)$$

This can be written in matrix form

$$\bar{p}'(t) = \bar{p}(t) A, \text{ where } A_{xy} = \begin{cases} -\alpha(x), & x = y \\ \alpha(x,y)), & x \neq y \end{cases}.$$

The matrix A is called the infitesimal generator of the system. Note that the row sums are all zero,

the diagonal entries are non-positive and nondiagonal entries are all non-negative. This system has the solution
$$\bar{p}(t) = \bar{p}(0)e^{tA}.$$

Now let $p_t(x,y) = \mathbb{P}\{X_t = y \mid X_0 = x\}$ and $P_t = (p_t(x,y))$. Then the above system can be written as a single matrix equation
$$\frac{d}{dt}P_t = P_t A, \quad P_0 = I,$$ and its solution is given by $P_t = e^{tA}$.

Example 1. Consider a two state chain with states 0 and 1 and $\alpha(0,1) = 1$, $\alpha(1,0) = 2$. Then the infinitesimal generator is
$$A = \begin{array}{cc} \begin{array}{cc} 0 & 1 \end{array} & \\ \begin{bmatrix} -1 & 1 \\ 2 & -2 \end{bmatrix} & \begin{array}{c} 0 \\ 1 \end{array} \end{array}.$$
Now $A = QDQ^{-1}$, where $D = \begin{bmatrix} 0 & 0 \\ 0 & -3 \end{bmatrix}$, $Q = \begin{bmatrix} 1 & 1 \\ 1 & -2 \end{bmatrix}$ and $Q^{-1} = \frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}$.

$$P_t = e^{tA} = Q\begin{bmatrix} 1 & 0 \\ 0 & e^{-3t} \end{bmatrix}Q^{-1} = \begin{bmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{bmatrix} + e^{-3t}\begin{bmatrix} 1/3 & -1/3 \\ -2/3 & 2/3 \end{bmatrix}$$

so that $\lim_{t\to\infty} P_t = \begin{bmatrix} 2/3 & 1/3 \\ 2/3 & 1/3 \end{bmatrix} = \begin{bmatrix} \pi \\ \pi \end{bmatrix}$, $\pi = [2/3 \;\; 1/3]$.

Example 2. Consider a chain with ~~states~~ states 0,1,2,3 and infinitesimal generator $A = \begin{array}{cc} \begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & -3 & 1 & 1 \\ 0 & 1 & -2 & 1 \\ 0 & 1 & 1 & -2 \end{bmatrix} & \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} \\ \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} & \end{array}$. Then

$$D = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & -4 \end{bmatrix}, \quad Q = \begin{bmatrix} 1 & 1 & 0 & -1/3 \\ 1 & 0 & 0 & 1 \\ 1 & -1/2 & -1/2 & -1/3 \\ 1 & -1/2 & 1/2 & -1/3 \end{bmatrix}, \text{ and}$$

$$\lim_{t\to\infty} P_t = \begin{bmatrix} 1/4 & \cdots & 1/4 \\ \vdots & & \vdots \\ 1/4 & \cdots & 1/4 \end{bmatrix}.$$

As in the case of discrete time we are interested in the large-time behavior. The above examples suggests that $\lim_{t\to\infty} P_t = \Pi_t = \begin{bmatrix} \overline{\pi} \\ \vdots \\ \overline{\pi} \end{bmatrix}$, where $\overline{\pi}$ represents the limiting probability.

Since the limiting probability should not change in t, considering the equation $\overline{p}'(t) = \overline{p}(t)A$, we see that $0 = \Pi_t A$.

Exercise 3.4. show that, as in the discrete case, for an irreducible chain we have:

1) There is a unique probability vector $\overline{\pi}$ satisfying $0 = \overline{\pi} A$ (so 0 is a simple eigenvalue)

2) All other eigenvalues of A have negative real parts.

Moreover, in this case one can prove that $\lim_{t\to\infty} P_t = [\overline{\pi} \ldots \overline{\pi}]^T$.

§3.3. Birth-and-Death Process.