Markov Chains

Statistics 110

Summer 2006

Markov Chains

For independent RVs (finitely valued) $X_1, X_2, \ldots$ from $F_X(x)$, $x \in \{1, 2, \ldots, K\}$ we can regard them as draws with replacement from an urn containing K types of balls, with proportions $p_j, j = 1, \ldots, K$.

A Markov Chain $X_1, X_2, X_3, \ldots$ is a simple generalization: there are K urns, each containing K types of balls.

The composition of urn$_i$ is described by proportions $p_{ij}, j = 1, \ldots, K$.

- Start from an initial state $X_0 = j_0$;
- Draw X_1 with replacement from urn$_{j_0} \to j_1$;
- Draw X_2 with replacement from urn$_{j_1} \to j_2$;
- $P[X_{n+1} = j | X_n = i] = p_{ij}$ are the **transition probabilities**.

The transition probabilities are often given by the **transition matrix**

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1K} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2K} \\ p_{31} & p_{32} & p_{33} & \cdots & p_{3K} \\ \vdots & & \ddots & & \vdots \\ p_{K1} & p_{K2} & p_{K3} & \cdots & p_{KK} \end{bmatrix}$$

Rows in the matrix correspond to which state you are in at time n and columns corresponds to which state you move to at time $n+1$. The matrix is square with K rows and K columns.

Note the for each $\{i, j\}, p_{ij} \geq 0$ and for each i, $\sum_{j=1}^{K} p_{ij} = 1$. So the K^2 entries in the matrix are non-negative who's rows sum to 1.

Any event relating to the first n steps is a subset of the sample space with sample points of the form

$$j_0 \; j_1 \; j_2 \; \cdots \; j_n$$

$$
\begin{aligned}
&P[j_0\ j_1\ j_2\ \ \ldots\ \ j_n] = \\
&\qquad P[X_0 = j_0]P[X_1 = j_1|X_0 = j_0]P[X_2 = j_2|X_0 = j_0X_1 = j_1] \\
&\qquad \times \ldots \times P[X_n = j_n|X_0 = j_0, X_1 = j_1, \ldots X_{n-1} = j_{n-1}] \\
&= \quad P[X_0 = j_0]P[X_1 = j_1|X_0 = j_0]P[X_2 = j_2|X_1 = j_1] \\
&\qquad \ldots P[X_n = j|X_{n-1} = j_{n-1}] \\
&= \quad P[X_0 = j_0]p_{j_0j_1}p_{j_1j_2}\ldots p_{j_{n-1}j_n}
\end{aligned}
$$

Markov chains are an example of conditional independence. In all cases

$$
\begin{aligned}
&P[X_{n+1} = j|X_0 = j_0, X_1 = j_1, \ldots, X_{n-1} = j_{n-1}, X_n = j_n] \\
&\qquad = P[X_{n+1} = j|X_n = j_n]
\end{aligned}
$$

In this case, once you know what X_n is, the distribution of X_{n+1} doesn't depend on times $1, 2, \ldots, n-1$.

We will also be interested in events that are subsets of the complete sample space, which has sample points that are infinite sequences

$$j_0 \; j_1 \; j_2 \; j_3 \; \cdots$$

Probabilities of these events are obtained as limits of probabilities of events involving finite sequences.

In this course, we will limit our treatment to only Markov chains with a finite state space (i.e. $x \in \{1, 2, \ldots, K\}$) or a countable (but infinite) state space (i.e. $x \in \{1, 2, 3, \ldots\}$).

It is also possible to have Markov chain with continuous state spaces. However we will not discuss this situation.

Markov chains are an example of a stochastic processes, which are used to model many phenomena. In many of these, the dependence of the present state on the past decreases as the past becomes more distant.

n-step transitions: The transition probabilities describe how the states change 1 step at a time. However is can be interest to know how the state change over 2, 3, or more steps.

The n-step transition probabilities are

$$p_{ij}^{(n)} = P[X_{m+n} = j | X_m = i] = P[X_n = j | X_0 = i]$$

How can we get these. Well the following idea (Chapman-Kolmogov equation) suggests an approach.

$$P[X_{m+n} = j | X_0 = i] = \sum_{k=1}^{K} P[X_n = k | X_0 = i] P[X_{n+m} = j | X_n = k]$$

Which can be rewritten as

$$p_{ij}^{(n+m)} = \sum_{k=1}^{K} p_{ik}^{(n)} p_{kj}^{(m)}$$

If I want to go from i to j in $n+m$ steps, I could go from i to k in n steps and then from k to j in m steps.

Note that these probabilities don't make any restriction on how you get from i to k or k to j. For example for a three step transition from state 1 to state 4 it could be $1 \to 2 \to 1 \to 4$ or $1 \to 2 \to 3 \to 4$ (or many other possibilities).

The matrix notation gives an easy approach to determining the n-step transition probabilities ($P^{(n)} = \left[p_{ij}^{(n)}\right]$ is the n-step transition matrix).

In matrix notation

$$P^{(2)} = P^2 = PP$$

$$P^{(3)} = P^3 = P^2 P$$

$$P^{(4)} = P^4 = P^3 P = P^2 P^2$$

$$P^{(n)} = P^n$$

Review: Matrix Multiplication of square matrices.

Let A and B both be K by K matrices

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1K} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2K} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3K} \\ \vdots & & \ddots & & \vdots \\ a_{K1} & a_{K2} & a_{K3} & \cdots & a_{KK} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1K} \\ b_{21} & b_{22} & b_{23} & \cdots & b_{2K} \\ b_{31} & b_{32} & b_{33} & \cdots & b_{3K} \\ \vdots & & \ddots & & \vdots \\ b_{K1} & b_{K2} & b_{K3} & \cdots & b_{KK} \end{bmatrix}$$

Then $C = AB = [c_{ij}]$ is a K by K matrix with entries

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{iK}b_{Kj}$$

The entry in the ith row and jth column of C depends on the ith row of A and the jth column of B.

Note that usually $AB \neq BA$ (order matters with matrix multiplication).

So lets look at the $\{i, j\}$ entry of $P^{(2)} = P^2$ (call it c_{ij} for now)

$$c_{ij} = p_{i1}p_{1j} + p_{i2}p_{2j} + \ldots + p_{iK}p_{Kj} = p_{ij}^{(2)}$$

so by induction, we can show that $P^{(n)} = P^n$ for any n.

Example Markov chains

1. Random walk with absorbing boundaries ($x \in \{0, \ldots, K\}$)

$$p_{i,i+1} = p; \quad p_{i,i-1} = q; \quad p + q = 1$$

so you go to the right 1 step with probability p or left 1 step with probability $q = 1 - p$, except for states 0 and K which are absorbing with

$$p_{00} = p_{KK} = 1$$

(once you get to state 0 or state K you never leave.)

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ q & 0 & p & 0 & \cdots & 0 \\ 0 & q & 0 & p & \cdots & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & \cdots & q & 0 & p \\ 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix}$$

2. Random walk with reflecting boundaries ($x \in \{0, \ldots, K\}$)

$$p_{i,i+1} = p; \quad p_{i,i-1} = q; \quad p + q = 1$$

so you go to the right 1 step with probability p or left 1 step with probability $q = 1 - p$, except for states 0 and K with

$$p_{00} = q; p_{01} = p; p_{K-1,K} = q; p_{KK} = p$$

(once you get to state 0 or state K you can leave.)

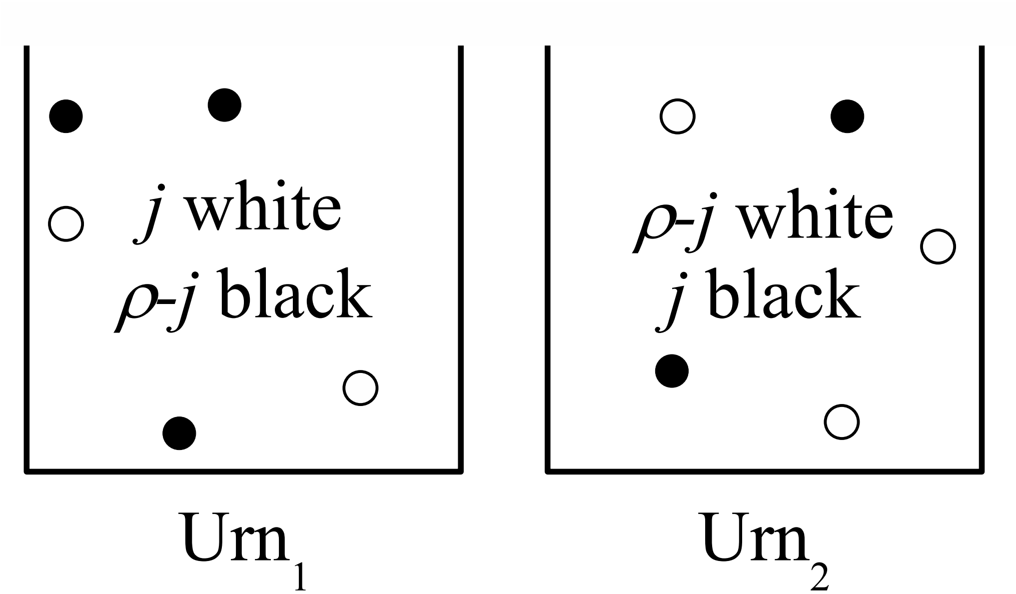
$$P = \begin{bmatrix} q & p & 0 & 0 & \cdots & 0 \\ q & 0 & p & 0 & \cdots & 0 \\ 0 & q & 0 & p & \cdots & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & \cdots & q & 0 & p \\ 0 & 0 & \cdots & 0 & q & p \end{bmatrix}$$

3. Bernoulli-Laplace diffusion

   Urn_1 and Urn_2 both contains ρ particles and half the particles in the two urns white and the other half are black.

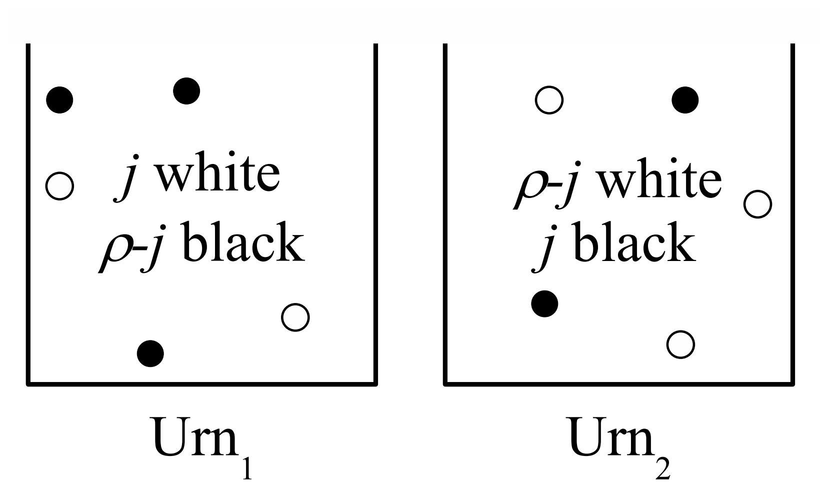
   This is model used to describe the flow of 2 incompressible liquids.

   The system is said to be in state j if Urn_1 contains j white particles $(j = 0, 1, \ldots, \rho)$.



If $X_n = j$, then X_{n+1} is generated by choosing 1 particle from each urn and exchanging them.

$$p_{jk} = \begin{cases} \left(\frac{j}{\rho}\right)^2 & \text{if } k = j-1 \\ 2\left(\frac{j}{\rho}\right)\left(\frac{\rho-j}{\rho}\right) & \text{if } k = j \\ \left(\frac{\rho-j}{\rho}\right)^2 & \text{if } k = j+1 \\ 0 & \text{otherwise} \end{cases}$$



4. Rain model

Suppose

$$P[\text{Rain Tomorrow}|\text{Today, Yesterday}] = \begin{cases} & & TD & YD \\ 0.7 & \text{if} & 1 & 1 \\ 0.5 & \text{if} & 1 & 0 \\ 0.4 & \text{if} & 0 & 1 \\ 0.2 & \text{if} & 0 & 0 \end{cases}$$

If

$$X_n = \begin{cases} 1 & \text{if rains on day } n \\ 0 & \text{otherwise} \end{cases}$$

This model, as described, is not a Markov chain, since it depends on more than just the current state.

However we can construct a Markov chain with an extended state space to represent this process

State	Present Day	Previous Day
1	1	1
2	1	0
3	0	1
4	0	0

Then

	Next Day	Today		Today	Yesterday	
$p_{11} = 0.7$	1	1	given	1	1	$\surd$
$p_{12} = 0$	1	0	given	1	1	$\times$
$p_{13} = 0.3$	0	1	given	1	1	$\surd$
$p_{14} = 0$	0	0	given	1	1	$\times$

Similarly for the other 3 states, which gives a transition matrix of

$$P = \begin{bmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{bmatrix}$$

Now lets assume that $X_0 = x$, i.e we start in state x.

Definition. [Entrance Times] *The entrance times to $y \neq x$ are defined by*

$$\begin{aligned}
T_y = T_y^{(1)} &= \inf\{n > 0 : X_n = y\} \\
&= \textit{first entrance time to } y \\
T_y^{(2)} &= \inf\{n > T_y^{(1)} : X_n = y\} \\
&= \textit{second entrance time to } y \\
T_y^{(k)} &= \inf\{n > T_y^{(k-1)} : X_n = y\} \\
&= k\textit{th entrance time to } y
\end{aligned}$$

Definition. [Recurrence Times] *If $y = x$, then the above definitions give the recurrence times to x.*

Let

$$\begin{aligned}\rho_{xx} &= P_x[T_x < \infty] \quad \text{(i.e. } P[T_x < \infty | X_0 = x]) \\ &= \text{probability of ever returning to } x \\ \rho_{xy} &= P_x[T_y < \infty] \\ &= \text{probability of ever visiting } y\end{aligned}$$

Definition. *A state y is said to be* **recurrent** *if $\rho_{yy} = 1$ and is said to be* **transient** *if $\rho_{yy} < 1$.*

Lemma. $P_x[T_y^{(k)} < \infty] = \rho_{xy}\rho_{yy}^{k-1}$

Proof. To make k visits to y, the chain must make a first entrance to y, and then must return to y $k-1$ times, with all of these steps taking finite times. $\square$

Let $N(y) = \sum_{i=1}^{\infty} I\{X_n = y\}$ = number of visits to state y.

Theorem. *y is a recurrent state if and only if $E_y[N(y)] = \infty$. If y is transient, then*

$$E_x[N(y)] = \frac{\rho_{xy}}{1 - \rho_{yy}}$$

Another way of thinking of the first part of this result, if y is recurrent, then $P_y[y \text{ recurs infinitely often}] = 1$.

Before proving this result, we need one result

Lemma. *Let N be a discrete random variable with possible outcomes $1, 2, 3, \ldots$. Then $E[N] = \sum_{k=1}^{\infty} P[N \geq k]$.*

To prove this result, notice that $P[N = k]$ occurs exactly k times in the sum.

Back to the theorem

Proof.

$$
\begin{aligned}
E_x[N(y)] &= \sum_{k=1}^{\infty} P_x[N(y) \geq k] = \sum_{k=1}^{\infty} P_x[T_y(k) < \infty] \\
&= \sum_{k=1}^{\infty} \rho_{xy} \rho_{yy}^{k-1} \\
&= \begin{cases} \infty & \text{if } y \text{ is recurrent and } x = y \\ \frac{\rho_{xy}}{1-\rho_{yy}} & \text{if } y \text{ is transient} \end{cases}
\end{aligned}
$$

□

Lemma 1. *If x is recurrent and $\rho_{xy} > 0$, then y is recurrent and $\rho_{yx} = 1$*

Proof. If $\rho_{xy} > 0$ and $\rho_{yx} < 1$, then starting from x, there is a positive probability of visiting y and then not returning to x in finite time. In that case x cannot be recurrent. Thus if x is recurrent and $\rho_{xy} > 0$, then ρ_{yx} must be 1. □

Note that this lemma implies that if x is a recurrent state and $\rho_{xy} > 0$, then $\rho_{xy} = 1$ as well.

Definition. *Denote the state space of the chain by S. A subset $C \subset S$ is* **Closed** *if $x \in C$ and $\rho_{xy} > 0$ implies $y \in C$.*

Clearly, if C is closed, then a chain starting from $x \in C$ can never leave C.

Definition. *A subset $D \subset S$ is* **Irreducible** *if any two states x and y in D must be mutually reachable from each other, i.e. $\rho_{xy} > 0$ and $\rho_{yx} > 0$.*

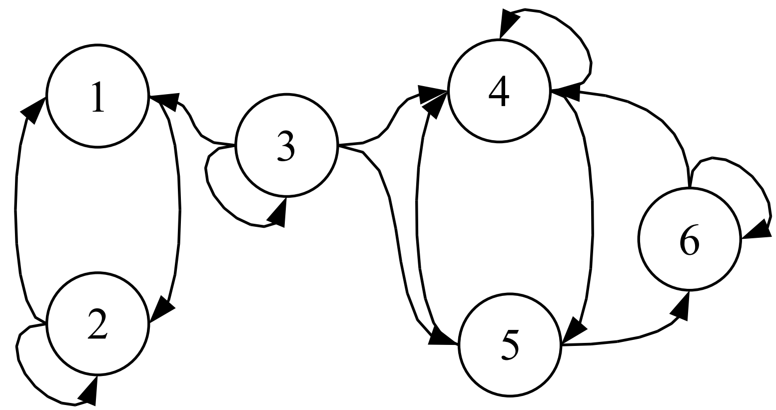
Lemma 2. *Let C be a finite closed set, then C contains a recurrent state. If further, C is irreducible, then all states in C are recurrent.*

Proof. Clearly C cannot contain only transient states. If C is irreducible, by lemma 1, all states must be recurrent if one is. $\square$

Lemmas 1 and 2 are useful for deciding which state are recurrent and which are transient.

Example

$$P: \begin{array}{c|cccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 2 & 0.4 & 0.6 & 0 & 0 & 0 & 0 \\ 3 & 0.3 & 0 & 0.4 & 0.2 & 0.1 & 0 \\ 4 & 0 & 0 & 0 & 0.3 & 0.7 & 0 \\ 5 & 0 & 0 & 0 & 0.5 & 0 & 0.5 \\ 6 & 0 & 0 & 0 & 0.8 & 0 & 0.2 \end{array}$$



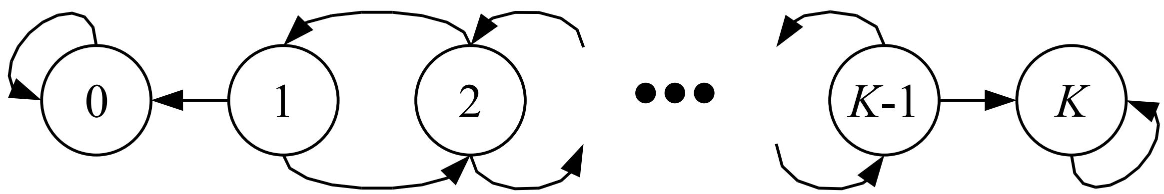
- By lemma 1, state 3 can't be recurrent.
- $\{1, 2\}$ is a finite closed, irreducible set. So by lemma 2, both are recurrent.
- Similarly, $\{4, 5, 6\}$ is a finite, closed, irreducible set, hence all are recurrent.

Given that the only way you can get to state 3 is if you start there, how long will you stay on average.

$$E_3[N(3)] = \frac{0.4}{1-0.4} = \frac{2}{3}$$

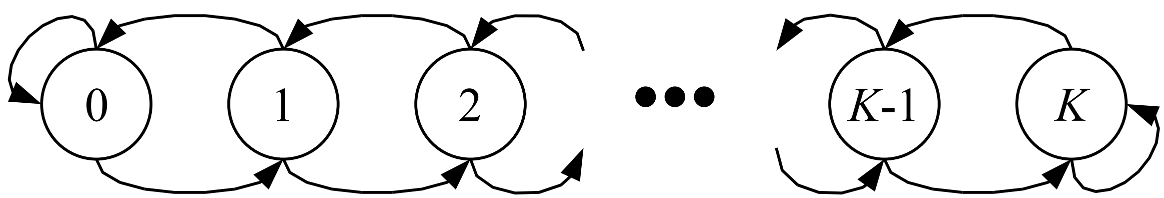
Lets examine the other 4 examples

1. Random walk with absorbing boundaries ($x \in \{0, \ldots, K\}$)



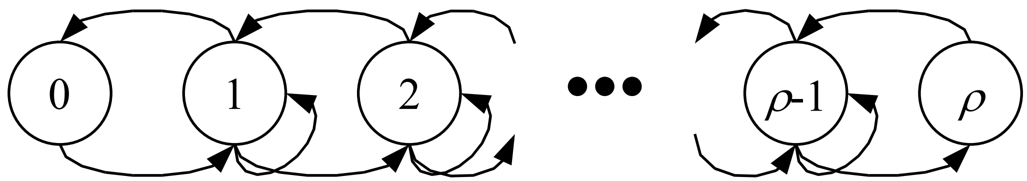
   States 0 and K are obviously recurrent, since once you enter each of those states you never leave them. All other states are transient.

2. Random walk with reflecting boundaries ($x \in \{0, \ldots, K\}$)



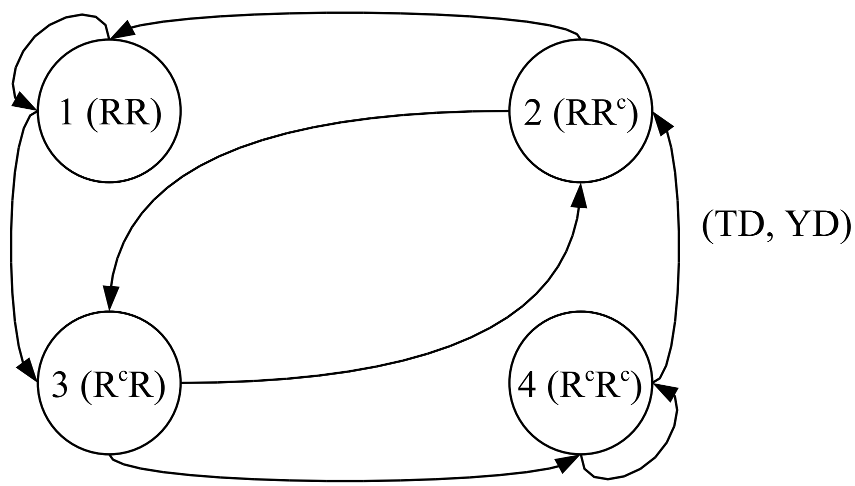
   With the reflecting boundaries modification, all states are now recurrent.

3. Bernoulli-Laplace diffusion



As there are no absorbing states here, all states must be recurrent.

4. Rain model



Similarly for this example, all states must be recurrent.

Lets rearrange the ordering status in the example: $\{1,2\},\{4,5,6\},\{3\}$

P :

	1	2	4	5	6	3
1	0	1	0	0	0	0
2	0.4	0.6	0	0	0	0
4	0	0	0.3	0.7	0	0
5	0	0	0.5	0	0.5	0
6	0	0	0.8	0	0.2	0
3	0.3	0	0.2	0.1	0	0.4

So the transition matrix P can be written in the following block form

$$P = \begin{bmatrix} P_1 & 0 & 0 \\ 0 & P_2 & 0 \\ Q_1 & Q_2 & P_3 \end{bmatrix}$$

The higher order transitions have matrices of the form

$$P^2 = \begin{bmatrix} P_1 & 0 & 0 \\ 0 & P_2 & 0 \\ Q_1 & Q_2 & P_3 \end{bmatrix} \begin{bmatrix} P_1 & 0 & 0 \\ 0 & P_2 & 0 \\ Q_1 & Q_2 & P_3 \end{bmatrix} = \begin{bmatrix} P_1^2 & 0 & 0 \\ 0 & P_2^2 & 0 \\ Q_1^{(2)} & Q_2^{(2)} & P_3^2 \end{bmatrix}$$

$$P^n = \begin{bmatrix} P_1^n & 0 & 0 \\ 0 & P_2^n & 0 \\ Q_1^{(n)} & Q_2^{(n)} & P_3^n \end{bmatrix}$$

Each of the recurrent sets $\{1,2\}$ and $\{4,5,6\}$ evolves independently as Markov chains with transitions matrices P_1 and P_2 respectively.

P_3 is a "sub-stochastic" matrix (i.e. row sum < 1) and $P_3^n \to 0$

Theorem. [Decomposition theorem for Markov chains]
Let $R = \{x : \rho_{xx} = 1\}$ be the recurrent states. Then $R = R_1 \cup R_2 \cup R_3 \cup \ldots$, where R_i's are mutually exclusive, irreducible, closed sets.

Proof. For each $x \in R$, let $C_x = \{y : \rho_{xy} > 0\}$. C_x is the set of all states reachable from x.

1. By lemma 1, all states in C_x are recurrent.

2. C_x is obviously closed by definition.

3. For any $x, y \in R$, if $C_x \cap C_y \neq \phi$, then we can find a $z \in C_x \cap C_y$ so that y is reachable from x via z and vice versa. Hence $C_x \cap C_y = \phi$ or $C_x = C_y$.

$\square$

Thus the state space can always be decomposed as $S = T \cup R_1 \cup R_2 \cup R_3 \cup \ldots$, where T contains the transient states that eventually disappear. Once the chain leaves the transient states, the chain evolves as an independent Markov chain on R_i once it enters R_i.

So to understand the long term behaviour of the chain, it is sufficient to understand the irreducible, recurrent chains.

Stationary Distributions

Theorem. *For an irreducible, recurrent chain*

$$P_x\left[\frac{N_n(y)}{n} \to \frac{1}{E_y[T_y]}\right] = 1$$

where

$$\begin{aligned} N_n(y) &= \sum_{m=1}^{n} I\{x_m = y\} \\ &= \textit{Number of visits to } y \textit{ by time } n \end{aligned}$$

(The proportion of times you enter state y converges almost surely to the reciprocal of the expected time between visits.)

Remarks:

- This theorem holds whether $E_y[T_y] < \infty$ or not.

- If $g(x)$ a function on S, then $N_n(y) = \#$ of times $g(X_m) = g(y)$ by time n

$$\frac{1}{n}\sum_{m=1}^{n} g(X_m) = \sum_{y\in S} \frac{\#(g(X_m) = g(y) \text{ by time } n)}{n} g(y)$$

$$\rightarrow \sum_{y\in S} \frac{1}{E_y[T_y]} g(y)$$

almost surely.

An example of where this is useful would be with the rainfall example. It would allow us to estimate the proportion of days that it actually rains. Let

$$g(y) = \begin{cases} 1 & \text{if } y = 1, 2 \\ 0 & \text{if } y = 3, 4 \end{cases}$$

Thus we can just look at the proportion of times we fall in either state 1 (RR) or 2 (RR^c) to estimate the long-run proportion of days with rain.

This theorem implies that for an irreducible recurrent chain, all "time averages" converge almost surely.

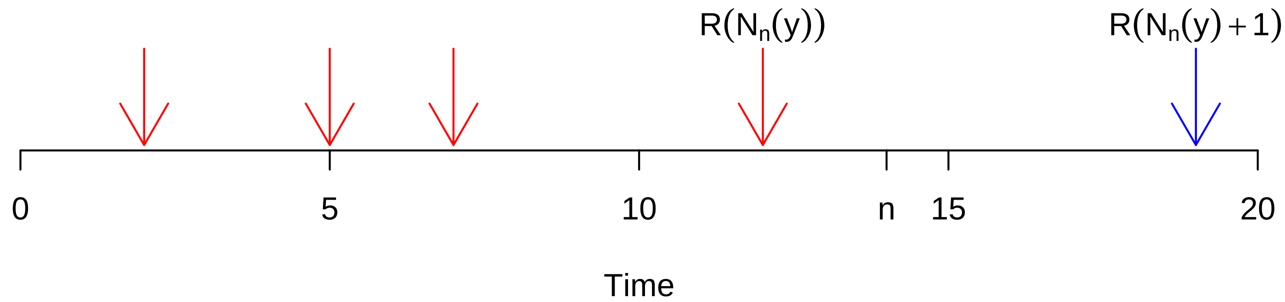
Proof. First suppose that $x = y$ (i.e. $X_0 = y$), and let

$$
\begin{aligned}
R(k) = T_y^{(k)} &= \text{time to } k\text{th visit to } y \\
&= \text{time to } k\text{th return to } y \\
&= \text{sum of } k \text{ independent draws from the distribution of } T_y
\end{aligned}
$$

So by the Strong Law of Large Numbers,

$$\frac{R(k)}{k} \to E_y[T_y]$$

almost surely if $X_0 = y$.



In general we have $R(N_n(y)) \leq n \leq R(N_n(y)+1)$ which implies

$$\frac{R(N_n(y))}{N_n(y)} \leq \frac{n}{N_n(y)} \leq \frac{R(N_n(y)+1)}{N_n(y)+1}\frac{N_n(y)+1}{N_n(y)}$$

Note that $N_n(y) \to \infty$ with probability 1 (since y is recurrent).

Since $\frac{R(N_n(y)+1)}{N_n(y)+1}$ is a subsequence of $\frac{R(k)}{k}$, it converges to $E_y[T_y]$ with probability 1.

Thus

$$P\left[\frac{n}{N_n(y)} \to E_y[T_y]\right] = 1$$

when $X_0 = y$. With a minor modification, it also holds when $X_0 = x \neq y$. $\square$

Corollary. *If the chain is irreducible and recurrent, then*

$$\frac{1}{n}\sum_{m=1}^{n} p_{xy}^{(m)} \to \frac{1}{E_y[T_y]}$$

for all $x, y \in S$. That is the m-step transition probabilities converge to some average value.

Proof. The almost sure convergence says that

$$\frac{N_n(y)}{n} \to \frac{1}{E_y[T_y]}$$

this then implies (with a couple of technical points omitted)

$$E\left[\frac{N_n(y)}{n}\right] \to \frac{1}{E_y[T_y]}$$

Finally

$$\begin{aligned}
E_x[N_n(y)] &= E_x\left[\sum_{m=1}^{n} I\{X_m = y\}\right] \\
&= \sum_{m=1}^{n} P_x[X_m = y] \\
&= \sum_{m=1}^{n} p_{xy}^{(m)}
\end{aligned}$$

□

Question: How to compute the limit $\frac{1}{E_y[T_y]}$?

The existence of

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} p_{xy}^{(m)} \stackrel{Def}{=} \pi(y)$$

suggests that under some conditions,

$$p_{xy}^{(m)} \to \pi(y)$$

If so, by the Chapman-Kolmogorov

$$\lim_{n\to\infty} p_{xy}^{(n+1)} = \sum_z \left(\lim_{n\to\infty} p_{xz}^{(n)} \right) p_{zy}$$

or

$$\pi(y) = \sum_z \pi(z) p_{zy} \qquad (*)$$

($\boldsymbol{\pi} = \boldsymbol{\pi} P$ in matrix formulation ($\boldsymbol{\pi}$ is a row vector of length K)).

We want to find a solution to (*) that represents a PMF.

Definition. *A solution of (*) satisfying*

$$\pi(y) \geq 0, \qquad \sum_{y \in S} \pi(y) = 1$$

is known as the **stationary distribution** *of the Markov chain.*

If $\pi(\cdot)$ is a stationary distribution and $X_0 \sim \pi(\cdot)$, then $p_{xy}^{(n)} = \pi(y)$ for all n, i.e. $X_n \sim \pi(\cdot)$ for all n.

Comment: If $S = \{1, 2, \ldots, K\}$ is finite, then $\boldsymbol{\pi} = (\pi_1 \ \pi_2 \ \ldots \ \pi_K)$ and P is a $K \times K$ matrix. Then $\boldsymbol{\pi}$ is stationary if

- $\pi_i \geq 0, \quad \sum_{i=1}^{K} \pi_i = 1$
- $\boldsymbol{\pi}$ is a (left) eigenvector of P with a non-zero eigenvalue

Examples:

- Rainfall model:

  Let us assume that $p_{xy}^{(m)} \to \pi(y)$ (which it does), then

$$(\pi_1 \ \pi_2 \ \pi_3 \ \pi_4) \begin{bmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{bmatrix} = (\pi_1 \ \pi_2 \ \pi_3 \ \pi_4)$$

  or

$$0.7\pi_1 + 0.5\pi_2 = \pi_1 \tag{1}$$

$$0.4\pi_1 + 0.2\pi_2 = \pi_2 \tag{2}$$

$$0.3\pi_1 + 0.5\pi_2 = \pi_3 \tag{3}$$

$$0.6\pi_1 + 0.8\pi_2 = \pi_4 \tag{4}$$

$$
\begin{aligned}
(1) + (3) &\Rightarrow \pi_2 = \pi_3 \\
(1) &\Rightarrow \pi_1 = \frac{5}{3}\pi_2 \\
+ \, (2) &\Rightarrow \pi_4 = 3\pi_2
\end{aligned}
$$

If we set $\pi_2 = 1$, then $\boldsymbol{\pi} \propto (\frac{5}{3} \;\, 1 \;\, 1 \;\, 3)$. Normalize this so it sums to 1 (making it a proper distribution) gives

$$
\boldsymbol{\pi} = \begin{pmatrix} \frac{5}{20} & \frac{3}{20} & \frac{3}{20} & \frac{9}{20} \end{pmatrix}
$$

State	Current Day	Previous Day	Probability
1	1	1	$\frac{5}{20}$
2	1	0	$\frac{3}{20}$
3	0	1	$\frac{3}{20}$
4	0	0	$\frac{9}{20}$

$$P[\text{Rain on current day}] = \frac{8}{20} = P[\text{Rain on previous day}]$$

$$P[\text{No rain on current day}] = \frac{12}{20} = \frac{3}{5} = P[\text{No rain on previous day}]$$

These imply that we should expect to go 2.5 days with between rainy days and 1.67 days between sunny days since the expected return time to a state is the reciprocal of its stationary probability.

- Random walk with reflecting boundaries ($x \in \{0, \ldots, K\}$)

  To find the stationary probabilities for this chain, we need to solve the system

$$(\pi_0 \; \pi_2 \; \ldots \; \pi_K) \begin{bmatrix} q & p & 0 & 0 & \cdots & 0 \\ q & 0 & p & 0 & \cdots & 0 \\ 0 & q & 0 & p & \cdots & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & \cdots & q & 0 & p \\ 0 & 0 & \cdots & 0 & q & p \end{bmatrix} = (\pi_0 \; \pi_2 \; \ldots \; \pi_K)$$

  subject to $\sum_{i=0}^{K} \pi_i = 1$.

This leads to

$$q\pi_0 + q\pi_1 = \pi_0 \Rightarrow \pi_1 = \frac{p}{q}\pi_0$$

$$p\pi_0 + q\pi_2 = \pi_1 \Rightarrow \pi_2 = \frac{p}{q}\pi_1 = \left(\frac{p}{q}\right)^2 \pi_0$$

$$p\pi_1 + q\pi_3 = \pi_2 \Rightarrow \pi_3 = \frac{p}{q}\pi_2 = \left(\frac{p}{q}\right)^3 \pi_0$$

$$\cdots$$

$$p\pi_{K-2} + q\pi_K = \pi_{K-1} \Rightarrow \pi_K = \frac{p}{q}\pi_{K-1} = \left(\frac{p}{q}\right)^K \pi_0$$

$$p\pi_{K-1} + p\pi_K = \pi_K \Rightarrow \pi_{K-1} = \frac{q}{p}\pi_K$$

So

$$\pi_j = \left(\frac{p}{q}\right)^j \pi_0 \quad \text{and} \quad \pi_{K-j} = \left(\frac{q}{p}\right)^j \pi_K$$

– If $p = 0.5 = q$, then $\pi_0 = \pi_1 = \ldots \pi_K = \frac{1}{K+1}$. So you expect to spend the same amount of time in each state in the long-run.

– If $p < 0.5$ $(q > 0.5)$ (tend to move to lower state), then

$$\pi_0 = \frac{1 - p/q}{1 - (p/q)^{K+1}}; \qquad \pi_j = \left(\frac{p}{q}\right)^j \frac{1 - p/q}{1 - (p/q)^{K+1}}$$

So the lower the state number, the more time you tend to spend in that state.

– If $p > 0.5$ $(q < 0.5)$ (tend to move to higher state), then

$$\pi_K = \frac{1 - q/p}{1 - (q/p)^{K+1}}; \qquad \pi_{K-j} = \left(\frac{q}{p}\right)^j \frac{1 - q/p}{1 - (q/p)^{K+1}}$$

So the higher the state number, the more time you tend to spend in that state.

The expected recurrence times are

- $p = q = 0.5$: $E_j[T_j] = K + 1$
- $p < 0.5$:

$$E_j[T_j] = \left(\frac{q}{p}\right)^j \frac{1 - (p/q)^{K+1}}{1 - p/q}$$

  So the higher the state number, the longer it tends to be before you return.
- $p > 0.5$:

$$E_{K-j}[T_{K-j}] = \left(\frac{p}{q}\right)^j \frac{1 - (q/p)^{K+1}}{1 - q/p}$$

  So the lower the state number (bigger j's), the longer it tends to be before you return.

Theorem. *If a Markov chain has a stationary distribution, then any state with a positive probability under the stationary distribution is recurrent*

Proof. Suppose $\pi(\cdot)$ is the stationary distribution and let $N(y) = \sum_{n=1}^{\infty} I\{X_n = y\}$. Then by lemma 1,

$$E_x[N(y)] = \frac{\rho_{xy}}{1 - \rho_{yy}}$$

thus we have

$$\sum_{n=1}^{\infty} p_{xy}^{(n)} = \frac{\rho_{xy}}{1 - \rho_{yy}}$$

Then $E[N(y)]$ satisfies

$$\sum_x \pi(x) \sum_{n=1}^{\infty} p_{xy}^{(n)} = \sum_x \pi(x) \frac{\rho_{xy}}{1-\rho_{yy}}$$

$$\sum_{n=1}^{\infty} \underbrace{\left(\sum_x \pi(x) p_{xy}^{(n)} \right)}_{y\text{th component of } \boldsymbol{\pi} P^n (= \boldsymbol{\pi})} = \sum_x \pi(x) \frac{\rho_{xy}}{1-\rho_{yy}}$$

So

$$\sum_{n=1}^{\infty} \pi(y) = \frac{\left(\sum_x \pi(x) \rho_{xy}\right)}{1-\rho_{yy}}$$

If $\pi(y) > 0$, then the left hand side must be ∞, which can only occur if $\rho_{yy} = 1$ (i.e. y is a recurrent state). $\square$

Theorem. *If a Markov chain is irreducible and has a stationary distribution* $\pi(\cdot)$, *then*

$$\pi(y) = \frac{1}{E_y[T_y]}$$

Proof. Since S is countable (or finite), there is a $x \in S$ such that $\pi(x) > 0$. By the previous theorem, this x is recurrent, and hence by the irreducibility, all state are recurrent. Since the chain is irreducible and recurrent, the corollary (page 33) applies, giving

$$\frac{1}{n}\sum_{m=1}^{n} p_{xy}^{(m)} \to \frac{1}{E_y[T_y]}$$

$$\frac{1}{n}\sum_{m=1}^{n}\sum_{x} \pi(x) p_{xy}^{(m)} \to \sum_{x} \pi(x)\frac{1}{E_y[T_y]}$$

$$\frac{1}{n}\sum_{m=1}^{n} \pi(y) \to \sum_{x} \pi(x)\frac{1}{E_y[T_y]}$$

□

Definition. *If $E_y[T_y] < \infty$, then $\pi(y) > 0$ and y is said to be a* **positive recurrent** *state. If y is recurrent with $E_y[T_y] = \infty$, then $\pi(y) = 0$ and y is said to be a* **null recurrent** *state.*

Theorem. *For a irreducible chain, the following are equivalent:*

1. *There is a positive recurrent state*

2. *There is a unique stationary distribution*

3. *All states are positive recurrent*

Proof. $(1 \Rightarrow 2)$

If x is recurrent, let $T = T_x =$ first return time. Then

$$\begin{aligned}
\mu_x(y) &= E_x[\#\text{ of visits to } y \text{ by } X_0, X_1, \ldots X_{T-1}] \\
&= E_x\left[\sum_{n=0}^{T-1} I\{X_n = y\}\right] \\
&= \sum_{n=0}^{\infty} P_x[X_n = y, T > n]
\end{aligned}$$

Since $X_0 = X_T$ (by construction), we also have

$$\begin{aligned}
\mu_x(y) &= E_x[\#\text{ of visits to } y \text{ by } X_1, \ldots X_{T-1}] \\
&= \sum_{n=1}^{\infty} P_x[X_n = y, T \geq n] \\
&= \sum_{n=1}^{\infty} \left[\sum_z P_x[X_{n-1} = z, T > n-1] p_{zy} \right] \\
&= \sum_z \left(\sum_{m=0}^{\infty} P_x[X_m = z, T > m-1] p_{zy} \right) \\
&= \sum_z \mu_x(z) p_{zy}
\end{aligned}$$

Thus

$$\frac{\mu_x(y)}{\sum_z \mu_x(z)}$$

is a stationary distribution since

$$\sum_z \mu_x(z) = \sum_{n=0}^{\infty} P_x[T > n] = E_x[T] < \infty$$

$\square$

Proof. $(2 \Rightarrow 3)$

Pick x with $\pi(x) > 0$, then for any y,

$$\pi(y) = \sum_z \pi(z) p_{zy}^{(n)} \geq \pi(x) p_{xy}^{(n)}$$

By irreducibility, $p_{xz}^{(n)} > 0$ for some n, hence $\pi(y) > 0$ and y is positive recurrent. $\square$

Proof. $(3 \Rightarrow 1)$

Obvious. :)

$\square$

The assumption about irreducibility is important. Otherwise there can be a multitude of stationary distributions.

Example: 6 state example with

$$P = \begin{bmatrix} P_1 & 0 & 0 \\ 0 & P_2 & 0 \\ Q_1 & Q_2 & P_3 \end{bmatrix}$$

Let $\mathbf{a}$ be a stationary distribution of P_1 (i.e. $a_i \geq 0, \sum a_i = 1, \mathbf{a}P_1 = \mathbf{a}$) and let $\boldsymbol{\pi}_1 = (\mathbf{a}\ \mathbf{0}\ \mathbf{0})$. Then

$$\boldsymbol{\pi}_1 P^n = (\mathbf{a}P_1^n\ \mathbf{0}\ \mathbf{0}) = \boldsymbol{\pi}_1$$

Similarly, if $\mathbf{b}$ is a stationary distribution for P_2, then $\boldsymbol{\pi}_2 = (\mathbf{0}\ \mathbf{b}\ \mathbf{0})$ is also a stationary distribution for P.

In addition, any PMF of the form

$$\alpha\boldsymbol{\pi}_1 + (1-\alpha)\boldsymbol{\pi}_2$$

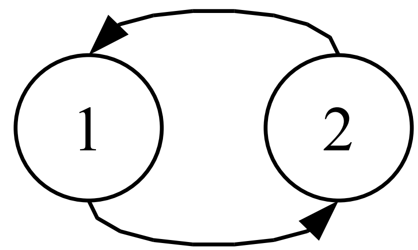
is also a stationary distribution of P.

Periodic Chains

Example

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

$$P^{(2)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



Then for $n = 1, 2, 3, \ldots, P^{(2n-1)} = P$ and $P^{(2n)} = P^{(2)}$

For each $x, E_x[T_x] = 2$ and

$$\frac{1}{n} \sum_{m=1}^{\infty} p_{xy}^{(m)} \rightarrow (0.5\ 0.5)$$

This chain is irreducible, recurrent, and has a stationary distribution, and generally satisfies all of the previous results. However $p_{xy}^{(m)}$ does not converge. The problem is due to periodicity.

Definition. *Let x be a recurrent state, and let*

$$I_x = \{n : n \geq 1,\ p_{xx}^{(n)} > 0\}$$

$$d_x = \textit{Greatest common divisor of } I_x$$

d_x is called the **period** *of x.*

In the above example, $I_x = \{2, 4, 6, 8, \ldots\}$ for $x = 1$ or 2. Thus the chain has period 2.

Lemma. *For an irreducible, recurrent chain, all states have the same period.*

Proof. By irreducibility, for any $x, y \in S$, there exist $k, l > 0$ such that

$$p_{xy}^{(k)} > 0 \quad p_{yx}^{(l)} > 0$$

Thus

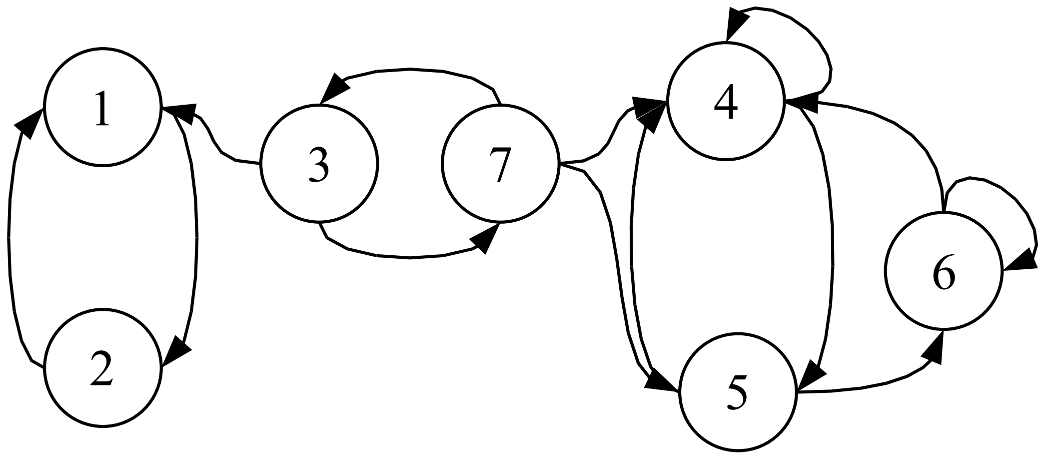
$$p_{yy}^{(k+l)} \geq p_{yx}^{(l)} p_{xy}^{(k)} > 0$$

Thus d_y divides $k + l$. Then for $n \in I_x$, we have

$$p_{yy}^{(k+l+n)} \geq p_{yx}^{(l)} p_{xx}^{(n)} p_{xy}^{(k)} > 0$$

so d_y also divides $k + l + n$. For this to hold d_y must divide n for any $n \in I_x$. Since d_x is the greatest common divisor of $I_x, d_x \geq d_y$. By an equivalent argument $d_y \geq d_x$. Thus $d_x = d_y$. $\square$

Note that the irreducible assumption is important as can be seen in the following example



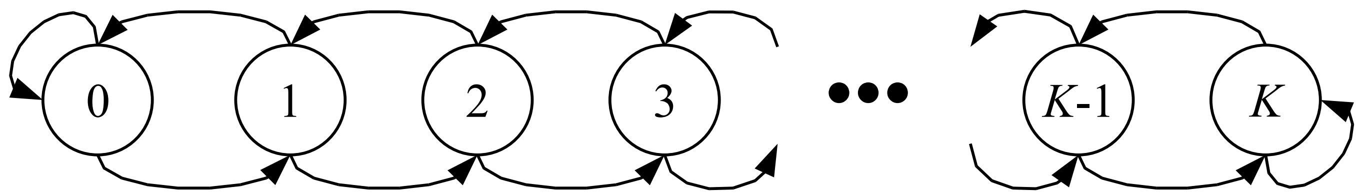
States 1,2,3, and 7 all have period 2, where states 4,5,6 have period 1.

Definition. *An irreducible, recurrent chain is said to be* **aperiodic** *is the period of the states in the chain is 1. If the chain has a period* $>$ *1, the chain is said to be* **periodic***.*

Lemma. *If* $d_x = 1$*, then there is a* $m_0 > 0$ *such that*

$$p_{xx}^{(m)} > 0 \quad \textit{for all } m \geq m_0$$

Example: Random walk with reflecting boundaries ($x \in \{0, \ldots, K\}$)



For state 2, $I_2 = \{2, 4, 5, 6, 7, 8, \ldots\}$ (assuming $K \geq 4$), so $d_2 = 1$ and $m_0 = 4$.

For state 3, $I_3 = \{2, 4, 6, 7, 8, \ldots\}$ (assuming $K \geq 6$), so $d_3 = 1$ and $m_0 = 6$.

Proof.

1. Suppose I_x contains 2 consecutive integers, then the result must hold. To see this, suppose $N, N+1 \in I_x$. Then for any $m \geq N^2$, we can write

$$m - N^2 = kN + r; \quad k \geq 0, 0 \leq r < N$$

or

$$m = r + N^2 + kN = r(N+1) + (N+k-r)N$$

Now if n_1 and n_2 are in I_x, so must be $n_1 + n_2$ and jn_1 for $j \geq 1$. Therefore $m \in I_x$.

2. Show that I_x must contain 2 consecutive integers. Choose $n_0, n_0 + k \in I_x$. If $k = 1$, we're done. If $k > 1$, then k does not divide some $n_1 \in I_x$ (since $d_x = 1$ is the only common divisor). Write $n_1 = mk + r$ with $0 < r < k, m \geq 0$. Then

$$N_1 \stackrel{Def}{=} (m+1)n_0 + n_1 \in I_x$$

$$N_2 \stackrel{Def}{=} (m+1)(n_0 + k) \in I_x$$

Then

$$N_2 - N_1 = (m+1)k - (mk + r) = k - r$$

which satisfies $0 < k - r < k$. So we have found a pair in I_x with a smaller difference. Repeat this argument until the difference is 1.

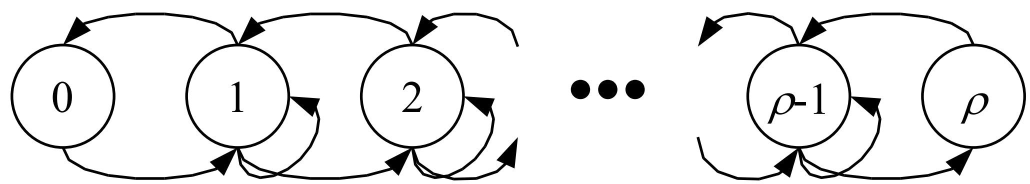
□

Theorem. *If a Markov chain, is irreducible, recurrent, aperiodic, and has a stationary distribution* $\pi(\cdot)$, *then*

$$p_{xy}^{(n)} \to \pi(y)$$

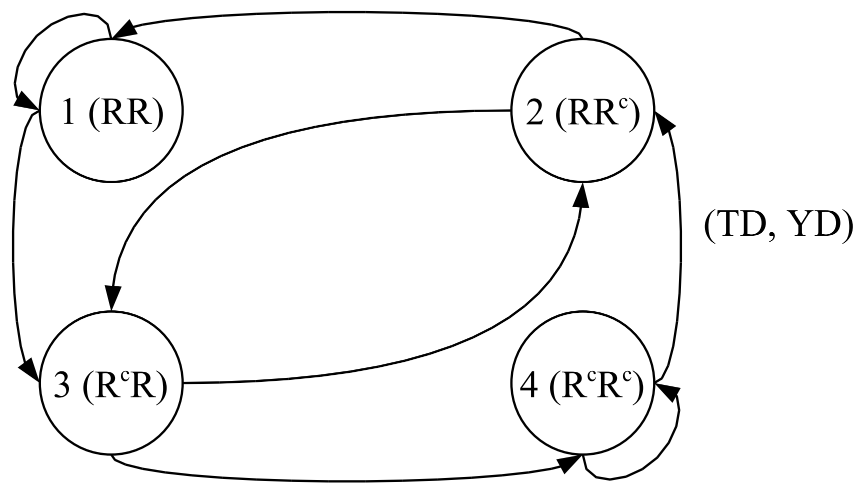
Examples:

- Bernoulli-Laplace diffusion



Since this chain is irreducible, recurrent and aperiodic, the n-step transition probabilities will converge to $\pi(\cdot)$. We can see that this chain is aperiodic since for some states it is possible to return in 1 step.

- Rain model



Similarly, this chain is also irreducible, recurrent and aperiodic, so n-step transition probabilities will converge to the stationary distribution $\pi(\cdot)$ calculated last class.

Proof. This proof relies on the trick known as "coupling of two independent copies". Let $\{X_n\}$ be a Markov chain evolving according to P, and $\{Y_n\}$ be another chain evolving according to P independently of $\{X_n\}$. Let $Z_n = (X_n, Y_n)$. Then $\{Z_n\}$ is a Markov chain on the sample space $S \times S$

with transition probabilities

$$\tilde{p}(z_1, z_2) = p_{x_1 x_2} p_{y_1 y_2}$$

if $z_1 = (x_1, y_1)$ and $z_2 = (x_2, y_2)$.

There are 4 steps we need to justify to prove the result.

1. $\{Z_n\}$ is a irreducible chain.

   Since $\{X_n\}$ and $\{Y_n\}$ are irreducible chains, $p^{(k)}_{x_1 x_2} > 0, p^{(l)}_{y_1 y_2} > 0$ for some $k, l > 0$. By aperiodicity, we can find m large enough so that $p^{(l+m)}_{x_2 x_2} > 0, p^{(k+m)}_{y_2 y_2} > 0$. Then

   $$\tilde{p}^{(k+l+m)}((x_1, y_1), (x_2, y_2)) = p^{(k+l+m)}_{x_1 x_2} p^{(k+l+m)}_{y_1 y_2} > 0$$

   since $p^{(k+l+m)}_{x_1 x_2} \geq p^{(k)}_{x_1 x_2} p^{(l+m)}_{x_2 x_2} > 0$ (similarly for $p^{(k+l+m)}_{y_1 y_2}$). Thus $\{Z_n\}$ is irreducible.

2. Let T = first time $Z_n = (X_n, Y_n)$ hits the diagonal $\{(y, y) : y \in S\}$. Then $T < \infty$ almost surely and $P[T > n] \to 0$.

   $\tilde{\pi}(x, y) = \pi(x)\pi(y)$ is the stationary distribution for $\{Z_n\}$. Thus by an earlier result, all states are positive recurrent. In particular, $T_{(y,y)} < \infty$ almost surely, which implies that $\min_{y \in S} T_{(y,y)} < \infty$ almost surely. Also its easy to see that $P[T < \infty] = 1 \Rightarrow P[T > n] \to 0$.

3. $P[X_n = y, T \le n] = P[Y_n = y, T \le n]$, i.e. on the set $\{T \le n\}, X_n$ and Y_n have the same distribution.

$$
\begin{aligned}
P[X_n = y, T \le n] &= \sum_{m=1}^{n} P[T = m, X_n = y] \\
&= \sum_{m=1}^{n} \sum_{x \in S} P[T = m, X_m = x, X_n = y] \\
&= \sum_{m=1}^{n} \sum_{x \in S} P[T = m, X_m = x] P[X_n = y | X_m = x] \\
&= \sum_{m=1}^{n} \sum_{x \in S} P[T = m, Y_m = x] P[Y_n = y | Y_m = x] \\
&= P[Y_n = y, T \le n]
\end{aligned}
$$

4. $\sum_{y \in S} |P[X_n = y] - P[Y_n = y]| \leq 2P[T > n]$

$$\begin{aligned} P[X_n = y] &= P[X_n = y, T \leq n] + P[X_n = y, T > n] \\ &= P[Y_n = y, T \leq n] + P[Y_n = y, T > n] \end{aligned}$$

which implies

$$P[X_n = y] \leq P[Y_n = y] + P[X_n = y, T > n]$$
$$P[Y_n = y] \leq P[X_n = y] + P[Y_n = y, T > n]$$

$$|P[X_n = y] - P[Y_n = y]| \leq P[X_n = y, T > n] + P[Y_n = y, T > n]$$

$$\begin{aligned} \sum_{y \in S} |P[X_n = y] - P[Y_n = y]| &\leq \sum_{y \in S} P[X_n = y, T > n] + P[Y_n = y, T > n] \\ &= P[T > n] + P[T > n] \end{aligned}$$

5. Now take $X_0 = x$, then $P[X_n = y] = p_{xy}^{(n)}$. Also take $Y_0 \sim \pi(\cdot)$, then $P[Y_n = y] = \pi(y)$. Then by 4., we have

$$\sum_{y \in S} |p_{xy}^{(n)} - \pi(y)| \le 2P[T > n] \to 0 \text{ by 2.}$$

□