3 Markov Chains

When we deal with a stochastic process, we are often interested in *prediction*. That is, we observe the process until time t, and use this knowledge to try and predict what the process will do at time $t+1$. In a white process, of course, we may just as well save the time and effort to observe: the past will give no indication about the future, none of our observations will help, and a prediction based on $P(x;t)$ is the best we can do. But, in general, we can rely on the kindness of processes: the past will indeed tell us something about the future, so we are often interested in calculating

$$P\big(m_{t+1}; t+1 \big| m_t, \ldots, m_0; t, \ldots, 0\big) \tag{143}$$

(we use the symbols m instead of x because of tradition: in this section, the range M will be finite or countable, and x is usually the name of a real variable). If the process is not white, then

$$P\big(m_{t+1}; t+1\big) \neq P\big(m_{t+1}; t+1 \big| m_t, \ldots, m_0; t, \ldots, 0\big) \tag{144}$$

and, in general, the latter is a distribution with a smaller variance, allowing us a better prediction. Finding this conditional distribution is, in the general case, very complex, since it may depend on all the values $m_0, \ldots, m_t$, so that the amount of information that we have to keep track of grows with time without bound. Fortunately for us, many interesting processes have a nice feature: knowing just the immediate past is enough for the prediction to be as good as it gets, and it is not necessary to keep track of remote values. We refer to this as the **Markov Property**.

Markov property

Definition 3.1. *Let $\Lambda = [\lambda_1, \ldots, \lambda_n]'$ be a distribution that is, a vector $\Lambda \in \mathbb{R}^n$ such that $\lambda_m \geq 0$, $\sum_m \lambda_m = 1$.*

A stochastic process X_t, $t \geq 0$, with values in M (M finite or countable) is Markov(λ, P) if

i) $P(m; 0) = \lambda_m$;

ii) $P\big(m_t; t \big| m_0, \ldots, m_{t-1}; 0, \ldots, t-1\big) = P\big(m_t; t \big| m_{t-1}; t-1\big)$

That is, a process is Markov if knowing the value of the process at time $t-1$ gives us the same information about the value at t as it would knowing the whole history of the process. We define the **transition matrix** at time t $\mathbf{P}(t)$ as

transition matrix

$$\mathbf{P}(t)\big|_{n,m} = P(m; t | n; t-1) \tag{145}$$

The values $m \in M$ are referred to as the **states** of the process, and $P(m; t|n; t-1)$ are the **transition probabilities** at t. A process is **stationary** if the transition probabilities are independent

states

transition probabilities

stationary process

of time, that is, for all m, n, t, τ,

$$P(m;t|n;t-1) = P(m;t+\tau|n;t+\tau-1) \tag{146}$$

In this case, $\mathbf{P}$ is a constant matrix

$$\mathbf{P}_{n,m} \stackrel{\triangle}{=} P(m|n) = P(m;t|n;t-1) \tag{147}$$

Note that $\sum_m P(m|n) = 1$ (from state n it is certain that we move to some state m), therefore

$$\sum_{m \in M} \mathbf{P}_{n,m} = 1 \tag{148}$$

for all n.

Example IX:
The process of die throw of Example 3 is white and, therefore, Markov, since

$$P\big(m_t;t\big|m_0,\ldots,m_{t-1};0,\ldots,t-1\big) = P\big(m_t;t\big|m_{t-1};t-1\big) = P_X(m_t) \tag{149}$$

Note that the definition of Markov process doesn't imply that knowing that $X_{t-1} = m_{t-1}$ will tell us something about X_t. The only requirement is that knowing that $X_{t-1} = m_{t-1}$ will give us the *same* information as knowing that $X_{t-1} = m_{t-1}, X_{t-1} = m_{t-2}, \ldots, X_0 = m_0$. In this process, the two will give us indeed the same information: none.

The process takes values in $\{-2, \ldots, 2\}$ and the probability $\mathbf{P}_{m,n}$ is independent of m (since the process is white) and constant (since the die is fair). It is

$$\mathbf{P} = \frac{1}{5}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \tag{150}$$

(note that the indices of the matrix span $[-2, \ldots, 2]$) (end of example)

Example X:
Consider the process Y_t of example 4. The process is Markov with

$$P\big(m_t;t\big|m_0,\ldots,m_{t-1};0,\ldots,t-1\big) = P\big(m_t;t\big|m_{t-1};t-1\big) = \begin{cases} \frac{1}{5} & m_{t-1} - 2 \le m_t \le m_{t-1} + 2 \\ 0 & \text{otherwise} \end{cases} \tag{151}$$

The transition matrix is, in this case

$$\mathbf{P}_{n,m} = \begin{cases} \frac{1}{5} & n-2 \le m \le n+2 \\ 0 & \text{otherwise} \end{cases} \tag{152}$$

The structure is that of a block-diagonal matrix

$$\mathbf{P} = \frac{1}{5}\begin{bmatrix} \ddots & & & 0 & & & & & & \\ & 1 & 1 & 1 & 1 & 1 & & & & \\ & & 1 & 1 & 1 & 1 & 1 & & & \\ & 0 & & 1 & 1 & 1 & 1 & 1 & & 0 \\ & & & & 1 & 1 & 1 & 1 & 1 & \\ & & & & & 1 & 1 & 1 & 1 & 1 \\ & & & & & & & 0 & & \ddots \end{bmatrix} \tag{153}$$

(end of example)

Example XI:
The process Z_t of Example 5 is **not** Marvkov. As we have seen in Example 5,

$$P(m_t; t|m_{t-1}; t-1) = \sum_k P_X(m_t - k) P_X(m_{t-1} - k) \tag{154}$$

On the other hand, assume that we have complete information, that is, we know that $Z_\tau = m_\tau$, $\tau = 0, \ldots, t-1$. In this case we can reconstruct X_{t-1} as

$$\begin{aligned} X_{t-1} &= X_{t-1} + X_{t-2} - X_{t-2} - X_{t-3} + X_{t-3} + \cdots \pm X_0 \\ &= Z_{t-1} - Zt - 2 + \cdots \pm Z_0 \end{aligned} \tag{155}$$

That is:

$$X_{t-1} = \sum_{\tau=1}^{t} (-1)^{\tau+1} Z_{t-\tau} \tag{156}$$

So that

$$P\big(m_t; t\big|m_0, \ldots, m_{t-1}; 0, \ldots, t-1\big) = \begin{cases} \frac{1}{5} & \big|\sum_{\tau=1}^{t} (-1)^{\tau+1} m_{t-\tau} - m_t\big| \le 2 \\ 0 & \text{otherwise} \end{cases} \tag{157}$$

which is different from the distribution (154). (end of example)

Theorem 3.1. *The process X_t, $t \geq 0$ is Markov(λ, P) iff*

$$P\big(m_0, \dots, m_t; 0, \dots, t\big) = \lambda_{m_0} \mathbf{P}_{m_0,m_1} \mathbf{P}_{m_1,m_2} \cdot \mathbf{P}_{m_{t-1},m_t} \tag{158}$$

Proof. Suppose X_t is Markov(λ, P), then

$$\begin{aligned} P\big(m_0, \dots, m_t; 0, \dots, t\big) &= P\big(m_t; t \big| m_0, \dots, m_{t-1}; 0, \dots, t-1\big) P\big(m_0, \dots, m_{t-1}; 0, \dots, t-1\big) \\ &\overset{(M)}{=} P\big(m_t; t \big| m_{t-1}; t-1\big) P\big(m_0, \dots, m_{t-1}; 0, \dots, t-1\big) \\ &\ \vdots \\ &= P\big(m_0; 0\big) P\big(m_1; 1 \big| m_0; 0\big) \cdots P\big(m_t; t \big| m_{t-1}; t-1\big) \\ &= \lambda_{m_0} \mathbf{P}_{m_0,m_1} \mathbf{P}_{m_1,m_2} \cdots \mathbf{P}_{m_{t-1},m_t} \end{aligned} \tag{159}$$

where the equaliy (M) is the Markov property.

Suppose not that (158) holds. We sum over $m_1, \dots, m_t$ obtaining

$$\sum_{m_1,\dots,m_t} P\big(m_0, \dots, m_t; 0, \dots, t\big) = P(m; 0) = \sum_{m_1,\dots,m_t} \lambda_{m_0} \mathbf{P}_{m_0,m_1} \mathbf{P}_{m_1,m_2} \cdots \mathbf{P}_{m_{t-1},m_t} = \lambda_{m_0} \tag{160}$$

(the last equality is a consequence of (148)). This proves property i) of definition 3.1. Summing (158) on m_t yields

$$P\big(m_0, \dots, m_{t-1}; 0, \dots, t-1\big) = \lambda_{m_0} \mathbf{P}_{m_0,m_1} \mathbf{P}_{m_1,m_2} \cdot \mathbf{P}_{m_{t-2},m_{t-1}} \tag{161}$$

Therefore

$$P\big(m_t; t \big| m_0, \dots, m_{t-1}; 0, \dots, t-1\big) = \frac{P\big(m_0, \dots, m_t; 0, \dots, t\big)}{P\big(m_0, \dots, m_{t-1}; 0, \dots, t-1\big)} = \mathbf{P}_{m_{t-1},m_t} \tag{162}$$

which proves ii). □

The one step transition matrix $\mathbf{P}$ can easily be extended to t steps. As a convenient notation, if A is an event, let $P_m(A) \triangleq P(A|X_0 = m)$ so that, for example, $P_m(X_1 = n) = \mathbf{P}_{m,n}$. Given an initial distribution Λ, we have

$$P(X_1 = n) = \sum_{m \in M} \lambda_m P_m(X_1 = n) = \sum_{m \in M} \lambda_m \mathbf{P}_{m,n} \tag{163}$$

Similarly,

$$
\begin{aligned}
P_m(X_2 = n) &= \sum_{k \in M} P_m(X_1 = k, X_2 = n) \\
&= \sum_{k \in M} P_m(X_2 = n | X_1 = k) P_m(X_1 = k) \\
&\overset{(M)}{=} \sum_{k \in M} P(X_2 = n | X_1 = k) P_m(X_1 = k) \\
&= \sum_{k \in M} \mathbf{P}_{m,k} \mathbf{P}_{k,n} \\
&= \mathbf{P}^2\big|_{m,n}
\end{aligned}
\tag{164}
$$

Note that in (M) we have removed the subscript m (that is, the conditioning on $X_0 = m$) because of the Markov property: if we know that $X_1 = k$, the knowledge that $X_0 = m$ does not change the conditional distribution of X_2. We also have

$$
\begin{aligned}
P(X_2 = n) &= \sum_{m,k \in M} \lambda_m P_m(X_1 = k, X_2 = n) \\
&= \sum_{m,k \in M} \lambda_m \mathbf{P}_{m,k} \mathbf{P}_{k,n} \\
&= \sum_{m \in M} \lambda_m \; \mathbf{P}^2\big|_{m,k} \\
&= \Lambda' \mathbf{P}^2\big|_{m,n}
\end{aligned}
\tag{165}
$$

Continuing we get

$$
\begin{aligned}
P_m(X_t = n) &= \mathbf{P}^t\big|_{m,n} \stackrel{\triangle}{=} \mathbf{P}^{(t)}_{m,n} \\
P(X_t = n) &= \sum_{m \in M} \lambda_m P_m(X_t = n) = \Lambda' \mathbf{P}^t\big|_n
\end{aligned}
\tag{166}
$$

Let $p(t) = [p_1, \ldots, p_n]'$, with $p_i = P(X_t = i)$ the probability distribution of the system at time t. Then:

$$
p'(t) = \Lambda' \mathbf{P}^t \tag{167}
$$

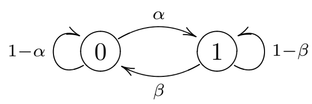
It is easy to show that the following **Chapman-Kolmogorov** equation holds:

$$
\mathbf{P}^{(t+s)}_{n,m} = \sum_{k \in M} \mathbf{P}^{(t)}_{n,k} \mathbf{P}^{(s)}_{k,m} \tag{168}
$$

Therefore, if the chain is stationary, the t step transition matrix is simply $\mathbf{P}^t$, if the chain is not stationary, then

$$
\mathbf{P}^{(t,t+\tau)} = \mathbf{P}_t \cdot \mathbf{P}_{t+1} \cdots \mathbf{P}_{t+\tau} \tag{169}
$$

Example XII:
The following diagram represents a simple two-state Markov chain



$$\mathbf{P} = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} \tag{170}$$

Assume, for the time being, that initially the system is in state 0, that is, $\Lambda = [1, 0]'$.

To compute $\mathbf{P}^{(t)}_{m,n}$ we use a trick that will be useful in many occasions, based on the eigenvalues of $\mathbf{P}$. These can easily be calculated, and they are

$$\lambda_1 = 1 \qquad \lambda_2 = 1 - \alpha - \beta \tag{171}$$

The matrix $\mathbf{P}$ can then be written as

$$\mathbf{P} = \mathbf{U} \begin{bmatrix} 1 & 0 \\ 0 & 1-\alpha-\beta \end{bmatrix} \mathbf{U}^{-1} \tag{172}$$

and

$$\mathbf{P}^t = \mathbf{U} \begin{bmatrix} 1 & 0 \\ 0 & (1-\alpha-\beta)^t \end{bmatrix} \mathbf{U}^{-1} \tag{173}$$

The elements of $\mathbf{P}^t$ are linear combinations of the elements of the vectors $[1, 0]'$ and $[0, (1-\alpha-\beta)^t]'$ so we can write, for example,

$$\mathbf{P}^{(t)}_{1,1} = A + B(1-\alpha-\beta)^t \tag{174}$$

for some A and B. The initial conditions give us

$$\begin{aligned} \mathbf{P}^{(0)}_{1,1} &= 1 = A + B \\ \mathbf{P}^{(1)}_{1,1} &= 1 - \alpha = A + B(1-\alpha-\beta) \end{aligned} \tag{175}$$

from which we derive

$$\begin{aligned} A &= \frac{\beta}{\alpha+\beta} \\ B &= \frac{\alpha}{\alpha+\beta} \end{aligned} \tag{176}$$

Repeating the argument for all the $\mathbf{P}^{(t)}_{m,n}$ we obtain

$$\mathbf{P}^t = \frac{1}{\alpha+\beta} \begin{bmatrix} \beta + \alpha(1-\alpha-\beta)^t & \alpha - \alpha(1-\alpha-\beta)^t \\ \beta - \beta(1-\alpha-\beta)^t & \alpha + \beta(1-\alpha-\beta)^t \end{bmatrix} \tag{177}$$

Many times, we are interested in what the system does in the log run, that is, in finding out in which state will it be and with which probability as $t \to \infty$. Let us assume that $0 < \alpha, \beta < 1$ (we shall consider the limit case shortly) so that $|1 - \alpha - \beta| < 1$. In this case

$$\lim_{t\to\infty} (1-\alpha-\beta)^t = 0 \tag{178}$$

So that

$$\mathbf{P}^{\infty} = \frac{1}{\alpha+\beta}\left[\begin{array}{cc} \beta & \alpha \\ \beta & \alpha \end{array}\right] \tag{179}$$

and

$$\Lambda'\mathbf{P}^{\infty} = \frac{1}{\alpha+\beta}[\beta,\alpha] \tag{180}$$

That is, as $t \to \infty$, the probability of finding the system in state 0 will be $\beta/(\alpha+\beta)$, an that of finding it in state 1 will be $\alpha/(\alpha+\beta)$. This chain is ergodic, therefore the result can be interpreted temporally, and we can say that, as $t \to \infty$, the system spends a fraction $\beta/(\alpha+\beta)$ of its time in state 0 and a fraction $\alpha/(\alpha+\beta)$ in state 1.

We have assumed that at $t = 0$ the system is in state 0. What if this is not the case? The most general condition that we can assume is that at $t = 0$ the system is in state 0 with probability u and in state 1 with probability $1-u$. In this case, the probabilities as $t \to \infty$ are

$$\frac{1}{\alpha+\beta}[u,1-u]\left[\begin{array}{cc} \beta & \alpha \\ \beta & \alpha \end{array}\right] = \frac{1}{\alpha+\beta}[u\beta+(1-u)\beta, u\alpha+(1-u)\alpha] = \frac{1}{\alpha+\beta}[\beta,\alpha], \tag{181}$$

the same as we have found before. The steady state of the system does not depend on its initial state. As time progresses, the system, "firgets" its initial state and settles to a steady state in which the probability of occupying each state depends only on the structure of the Markov model.

This is not always the case. If $\alpha = \beta = 0$, then

$$\mathbf{P} = \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right] \tag{182}$$

and $\mathbf{P}^{\infty} = \mathbf{P}$ so that $p(\infty) = \Lambda$. For example, if $\Lambda = [1,0]'$, the system starts in state 0 and will always remain there. If $\Lambda = [0,1]'$ the system will always be in state 1. The limit may also fail to exist. If $\alpha = \beta = 1$, then

$$\mathbf{P} = \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}\right] \tag{183}$$

and

$$\mathbf{P}^{2n} = \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right] \qquad \mathbf{P}^{2n+1} = \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}\right] \tag{184}$$

The limit doesn't exist: the system will bounce back and forth from a state to another and, at each time, we will know for sure (with probability 1) in which state it will be: there is no stable probability distribution. (end of example)

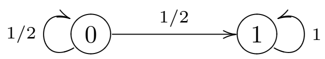
One useful property of Markov chains is that at any time we can "forget the past and sart anew" taking as initial state the one in which we are placed.

Theorem 3.2. *Let X_t, $t \geq 0$ be Markov(λ, P). Then, conditioned on $X_t = m$, the sequence $X_{t+s}|_{s\geq 0}$ is Markov(δ_m, P), and is independent on the variables $X_0, \ldots, X_t$.*

Here, $\delta_m = [0, \ldots, 0, 1, 0, \ldots, 0]'$, with the 1 in the mth position.

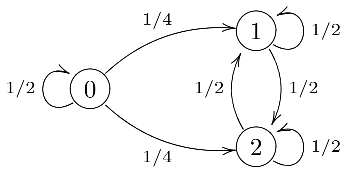
3.1 Class Structure

Consider the following chain



(185)

It is clear that, once the chain enters the state 2, it will never leave it. State 2 acts like a "sink" or (if you feel musically inclined), like Hotel California: "you can check out any time you like, but you can never leave". In the following chain, the idea is the same but the situation is a bit more complex



(186)

In this case there is no single state in which the chain remains forever, but once it has entered the *set* of states $\{1, 2\}$ it will never leave it, although it will keep jumping from one state to another. There is obviously something special about the set $\{1, 2\}$. In this section we shall study this and other characteristics of sets of states.

* * *

We say that n **leads to** m ($n \to m$) if $P_n(X_t = m) > 0$ for some $t > 0$. We say that n leads to

communicates with m ($n \leftrightarrow m$) if $n \to m$ and $m \to n$.

communicates with

Theorem 3.3. *Given two distinct states n and m, the following are equivalent:*

i) $n \to m$

ii) *There are $n_0, \ldots, n_t$ with $n_0 = n$ and $n_t = m$ such that*

$$\mathbf{P}_{n_0,n_1} \cdot \mathbf{P}_{n_1,n_2} \cdot \cdots \cdot \mathbf{P}_{n_{t-1},n_t} > 0 \tag{187}$$

iii) $\mathbf{P}^{(t)}_{n,m} > 0$ *for some $t \geq 1$.*

Proof. Note that

$$\mathbf{P}^{(t)}_{n,m} \leq P\Big[\exists t.X_t = m\Big] \leq \sum_{s=0}^{\infty} \mathbf{P}^{(s)}_{n,m} \tag{188}$$

proving the equivalence of i) and ii). Also

$$\mathbf{P}^{(t)}_{n,m} = \sum_{n_0,\ldots,n_t, n_0=n, n_t=m} \mathbf{P}_{n_0,n_1} \cdot \mathbf{P}_{n_1,n_2} \cdot \cdots \cdot \mathbf{P}_{n_{t-1},n_t} > 0 \tag{189}$$

proving the equivalence of ii) and iii). □

It is quite easy to see that $\leftrightarrow$ is an equivalence relation and, as such, it partitions M into equivalence classes that we call **communicating classes**. Moreover, we say that $C \subseteq M$ is a **closed class** if

communicating classes

closed class

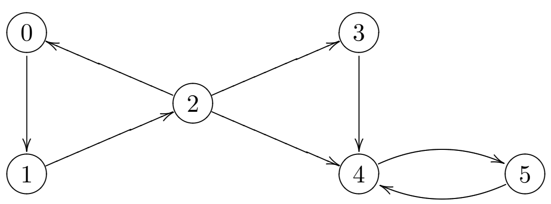
$$c \in C, c \to m \implies m \in C \tag{190}$$

A closed class is a set of states that constitutes a trap from which there is no escape: once the chain has reached a state in C, it will forever stay in C. A state $m \in M$ is **absorbing** is $\{m\}$ is a closed class. If C is not closed, then it is **open**.

absorbing state

open class

Example XIII:
Given the diagram



(191)

The sets $\{0,1,2\}$ $\{3\}$ and $\{4,5\}$ are communicating classes. $\{4,5\}$ is closed, while $\{0,1,2\}$ and $\{3\}$ are open. (end of example)

A chain for which M is a single class is called **irreducible**.

irreducible chain

3.2 Hitting probabilities and hitting mean time

Let $X_t|_{t\geq 0}$ be a Markov chain taking values in M, with transition matrix $\mathbf{P}$. Given a subset $A \subseteq M$, the **first hitting time** of A is the random variable $H^A : \Omega \to \mathbb{N} \cup \{\infty\}$ given by

first hitting time

$$H^A(\omega) = \inf\{t \geq 0 \big| X_t(\omega) \in A\} \tag{192}$$

where, by convention, the infimum of the empty set is ∞. The probability that starting from state m the chain ever hits A is

$$h_m^A = P(H^A < \infty) \tag{193}$$

The most interesting case, and the one that we shall more often consider in the following, is that in which A is closed. In this case h_m^A is called the **absorption probability** of A. The **mean hitting time** for reaching A is

absorption probability
mean hitting time

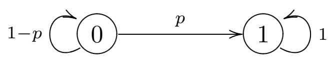
$$k_m^A = \mathbf{E}[H^A] = \sum_{t<\infty} t P_m(H^A = t) + \infty P_m(H^A - \infty) \tag{194}$$

Informally, we shall use the notation

$$\begin{aligned} h_m^A &\triangleq P_m(\text{hit } A) \\ k_m^A &\triangleq \mathbb{E}[t \text{ to } A] \end{aligned} \tag{195}$$

and we shall omit the superscript A when the set we are considering is obvious from the context.

Example XIV:
Consider the chain



$$\mathbf{P} = \begin{bmatrix} 1-p & p \\ 0 & 1 \end{bmatrix} \tag{196}$$

where 1 is an absorption state. The probability of absorption starting from 1 is clearly $h_1 = 1$. The absorption probability starting from 0 is

$$h_0 = \sum_t P(H^2 = t | X_0 = 0) = \sum_t (1-p)^{t-1} p = 1 \tag{197}$$

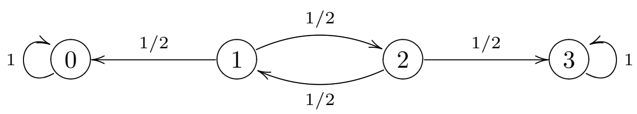
For the mean hitting time, clearly $k_1 = 0$, and

$$k_0 = \sum_t t P(H^2 = t | X_0 = 0) = \sum_t t(1-p)^{t-1} p = -p \frac{d}{dp} \sum_t (1-p)^t = -p \frac{d}{dp} \frac{1}{1-(1-p)} = \frac{1}{p} \tag{198}$$

(end of example)

Calculating the hitting probability and the mean hitting time can get quite complex, as it can well be understood by looking at all the work we had to do and the tricks we had to put in place to solve this trivial example. Fortunately for us, these quantities can be calculated as a solutions of a simple system of linear equations

Example XV:
Consider the chain



(199)

The chain has two absorbing states, 0 and 3. Lat us say that we start from 2; what is the probability of being absorbed in 3? Let $h_m = P_m(\text{hit } 3)$. Consider now the state 1. From 1 one can get to 3 in two ways: either moving to 0 (with probability 1/2) and then from 0 to 3 (with probability h_0), or moving to 2 (with probability 1/2) and then from 2 to 3 (with probability h_2). So h_1 is equal to the probability of moving from 1 to 0 multiplied by the probability of hitting 3 starting from 0, plus the probability of moving from 1 to 2 multiplied by the probability of hitting 3 starting from 2. Reasoning in this way for all states, we have

$$\begin{aligned} h_0 &= 0 \qquad\qquad \text{(This is so because 0 is absorbing)} \\ h_1 &= \frac{1}{2} h_0 + \frac{1}{2} h_2 \\ h_2 &= \frac{1}{2} h_1 + \frac{1}{2} h_3 \\ h_3 &= 1 \end{aligned} \tag{200}$$

which gives solutions $h_1 = 1/3$ and $h_2 = 2/3$. For the average time, the reasoning is similar, but this time each transition "costs" one time step, and the time from an absorbing state other than 3 is ∞:

$$\begin{aligned} k_0 &= \infty \\ k_1 &= 1 + \frac{1}{2}k_0 + \frac{1}{2}k_2 \\ k_2 &= 1 + \frac{1}{2}k_1 + \frac{1}{2}k_3 \\ k_3 &= 0 \end{aligned} \tag{201}$$

which gives us $k_2 = \infty$ (if we start from 2, one third of the times we will end up in 0, from which the time to reach 3 is infinity, therefore $k_2 \le \frac{1}{3}\infty + \frac{2}{3}A$, where A is a finite quantity.) On the other hand, if we want to know the average time to be absorbed eiter in 0 or 3, setting

$$k_m = \mathbb{E}[t \text{ to } \{0,3\} \big| X_0 = m] \tag{202}$$

we have

$$\begin{aligned} k_0 &= 0 \\ k_1 &= 1 + \frac{1}{2}k_0 + \frac{1}{2}k_2 \\ k_2 &= 1 + \frac{1}{2}k_1 + \frac{1}{2}k_3 \\ k_3 &= 0 \end{aligned} \tag{203}$$

which gives $k_1 = k_2 = 2$. (end of example)

The equations for the absorption probabiliy of this example don't come out of nowhere, they are the example of a general property.

Theorem 3.4. *The vector of hitting probabilities of $A \subseteq M$, $h^A = [h_m^A \big| m \in M]'$ is the minimal non-negative solution of*

$$\begin{array}{lll} h_m^A & = 1 & m \in A \\ h_m^A & = \displaystyle\sum_{n \in M} \mathbf{P}_{m,n} h_n^A & m \notin A \end{array} \tag{204}$$

Proof. We first show that h_m^A is a solution of (204). If $X_0 = m$, $m \in A$, then $h_m^A = 1$, which satisfies the first of (204). If $X_0 \notin A$, then $H^A \geq 1$. By the Markov property,

$$\begin{aligned} h_m^A = P_m(H^A < \infty) &= \sum_{n \in M} P_m(H^A < \infty, X_1 = n) \\ &= \sum_{n \in M} P_m(H^A < \infty \big| X_1 = n) P_m(X_1 = n) \\ &\overset{(*)}{=} \sum_{n \in M} \mathbf{P}_{m,n} h_n^A \end{aligned} \tag{205}$$

where equality (*) derives from Teorem 3.2

$$P_m(H^A < \infty \big| X_1 = n) = P_n(H^A < \infty) = h_n^A \tag{206}$$

Assume now that $x = [x_m \big| m \in M]'$ is a solution of (204). If $m \in A$, then $x_m = 1 = h_m^A$. Suppose $m \notin A$, then

$$x_m = \sum_{n \in M} \mathbf{P}_{m,n} x_n = \sum_{n \in A} \mathbf{P}_{m,n} x_n + \sum_{n \notin A} \mathbf{P}_{m,n} x_n \tag{207}$$

Expanding the x_n, we have

$$\begin{aligned} x_m &= \sum_{n \in A} \mathbf{P}_{m,n} + \sum_{n \notin A} \mathbf{P}_{m,n} \Big[\sum_{k \in A} \mathbf{P}_{n,k} x_k + \sum_{k \notin A} \mathbf{P}_{n,k} x_k \Big] \\ &= \sum_{n \in A} \mathbf{P}_{m,n} + \sum_{n \notin A, k \in A} \mathbf{P}_{m,n} \mathbf{P}_{n,k} + \sum_{n \notin A, k \notin A} \mathbf{P}_{m,n} \mathbf{P}_{n,k} x_k \\ &= P_m(X_1 \in A) + P_m(X_1 \notin A, X_2 \in A) + \sum_{n \notin A, k \notin A} \mathbf{P}_{m,n} \mathbf{P}_{n,k} x_k \end{aligned} \tag{208}$$

Iterating we have

$$\begin{aligned} x_m &= P_m(X_1 \in A) + P_m(X_1 \notin A, X_2 \in A) + \cdots + P_m(X_1 \notin A, X_2 \in A, \ldots, X_{t-1} \notin A, X_t \in A) \\ &\quad + \sum_{n_i \notin A} \mathbf{P}_{m,n_1} \mathbf{P}_{n_1,n_2} \cdots \mathbf{P}_{n_{t-1},n_t} x_{n_t} \\ &= P_m(H^A \leq t) + \sum_{n_i \notin A} \mathbf{P}_{m,n_1} \mathbf{P}_{n_1,n_2} \cdots \mathbf{P}_{n_{t-1},n_t} x_{n_t} \end{aligned} \tag{209}$$

Since $x_m \geq 0$, we have $x_m \geq P_m(H^A \leq t)$, therefore

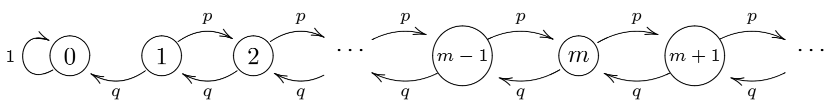
$$x_m \geq \lim_{t \to \infty} P_m(H^A \leq t) = P_m(H^A < \infty) = h_m^A \tag{210}$$

□

Note that, in (200), the equality $h_0 = 0$ cannot be derived from equations (204), which only yields $h_0 = h_0$. The condition $h_0 = 0$ derives from the minimality of the solution.

Example XVI:
(Gambler's ruin.) You are a gambler playing against the house repeated rounds of a zero-sum game in which you have in each round a probability p of winning and a probability $q = 1 - p$ of losing. You gamble \$1 each round, and have an initial capital of m dollars. The house has an infinite supply of money, while you must stop playing when you go bust, that is, when you are left without money (in this casino, the house doesn't give credit, fortunately for you). If you play forever, what is the probability that you will end up being ruined?

The game is represented by a Markov chain, in which the state is the amount of money that you have.



(211)

Set $h_n = P_n(\text{hit } 0)$. The transition probabilities are

$$\begin{aligned} \mathbf{P}_{0,0} &= 1 \\ \mathbf{P}_{m,m-1} &= q \\ \mathbf{P}_{m,m+1} &= p \end{aligned} \tag{212}$$

therefore the probabilities that we seek are the solution of

$$\begin{aligned} h_0 &= 1 \\ h_m &= p h_{m-1} + q h_{m+1} \quad m = 1, 2, \ldots \end{aligned} \tag{213}$$

If $p \neq q$, these equations have general solution

$$h_n = 1 - A + A \left(\frac{q}{p} \right)^n \tag{214}$$

and the minimal solution must be a distribution, that is, $0 \leq h_m \leq 1$. If $p < q$, the condition $h_m \leq 1$ forces $A = 0$, therefore $h_0 = 1$. If $p > q$, the minimal solution has A as large as possible consistent with $h_m \geq 0$, that is, $A = 1$, and $h_n = (q/p)^n$.

If $p = q = 1/2$, the recurrence has solution $h_n = 1 + Bn$ and the condition $h_n \leq 1$ forces $B = 0$, so $h_n = 1$: even if you find a fair casino[4], the fact that you have a limited amount of money will, in

[4]Which, of course, you won't: Las Vegas casinos are required by law to publish the odds that you have in all the games, and in general $p/q \approx 0.98$.

the long run, bankrupt you. (end of example)

A theorem analogous to 3.4 holds for mean hitting times (we omit the proof).

Theorem 3.5. *The vector of mean hitting times of $A \subseteq M$, $k^A = [k^A_m | m \in M]'$ is the minimal non-negative solution of*

$$\begin{aligned} k^A_m &= 0 & m \in A \\ k^A_m &= 1 + \sum_{n \in M} \mathbf{P}_{m,n} k^A_n & m \notin A \end{aligned} \tag{215}$$

3.3 Recurrence and transience

Let us begin with a bunch of definitions, some of which we shall need only in a while (but, since they are all related, we define them here and get it over with).

Let $X_t|_{t \geq 0}$ a Markov chain. Define

i) The **hitting time** of m is $H_m = \inf\{n \geq 0 | X_t = m\}$ | hitting time |

ii) The **first passage time** to m is $T_m = \inf\{n \geq 1 | X_t = m\}$ | first passage time |

Note that H_m and T_m differ only if $X_0 = m$.

iii) The **number of visits** to m is $V_m = \sum_{t=0}^{\infty} \chi_{\{X_t = m\}}$, where χ_A is the indicator function of a set A. | number of visits |

iv) The **return probability** to m is $f_m = P_m(T_m < \infty)$ | return probability |

v) The **mean return time** to m is $\mu_m = \mathbb{E}_m[T_m]$ | mean return time |

vi) The **number of visits** to m **before τ** is | number of visits before τ |

$$V_m(\tau) = \sum_{t=0}^{\tau-1} \chi_{\{X_t = m\}} \tag{216}$$

vii) The **number of visits** to m **before the first return** to n is $V_m^n = V_m(T_n)$

number of visits before first return

viii) The The **mean number of visits** to m **between successive visits** to n is $\gamma_m^n = \mathbb{E}_m[V_m^n]$

mean number of visits

Notice that if $X_0 = m$, then $V_m^m = 1$ and therefore $\gamma_m^m = 1$.

recurrent and transient states

We say that m is **recurrent** if $P_m(V_m = \infty) = 1$, otherwise we say that m is **transient**. A state is recurrent is the chain keeps returning to it, while in the case of a transient state, the chain will at some point leave it, never to return. Note that being transient doesn't mean that the chain will pass only once through the state. However, it does mean that, with probability 1, the chain will visit it only a finite number of times[5]. This doesn't quite seem to be the case if one gives a look at the definition: *prima facie* all the definition says is that if teh state is transient, the probability of visiting it infinite times is less than one. That it is really zero is a consequence of the following lemma:

Lemma 3.1. *For all $\tau \geq 0$, $P_m(V_m \geq \tau + 1) = (f_m)^\tau$.*

Proof. The proof is by induction over τ. The lemma is true for $\tau = 0$ by definition. Assume that it is true for $\tau - 1$. Then

$$\begin{aligned} P_m(V_m \geq \tau + 1) &= P_m(V_m \geq \tau + 1 \big| V_m \geq \tau) P_m(V_m \geq \tau) \\ &= P_m(T_m < \infty)(f_m)^{\tau - 1} \\ &= (f_m)^\tau \end{aligned} \tag{217}$$

□

From this, it is easy to show that the following is true

Theorem 3.6.

$$m \textit{ recurrent} \quad \Leftrightarrow \quad f_m = 1 \quad \Leftrightarrow \quad \sum_{t=0}^{\infty} \mathbf{P}_{m,m}^{(t)} = \infty \tag{218}$$

$$m \textit{ transient} \quad \Leftrightarrow \quad f_m < 1 \quad \Leftrightarrow \quad \sum_{t=0}^{\infty} \mathbf{P}_{m,m}^{(t)} < \infty \tag{219}$$

[5] The notion of the event E happening "with probability 1" is important here. It is not the same as saying that E will *always* happen or, equivalently, that $\neg E$ will *never* happen. It does mean that the set of outcomes in which $\neg E$ happens has measure zero:

$$\mu(\{\omega \in \Omega : (\neg E)(\omega)\}) = 0$$

Recurrence and transience have been defined for a state, but they are really class prperties

Theorem 3.7. *Let C be a communicating class, then either all the states in C are transient or they are all recurrent.*

Proof. Take $n, m \in C$ and suppose that m is transient. Then, by definition of communicating class, there are $t, s > 0$ with $\mathbf{P}^{(t)}_{m,n} > 0$ and $\mathbf{P}^{(s)}_{n,m} > 0$. By Chapman-Kolmorogov, for all $\tau > 0$,

$$\mathbf{P}^{(t+s+\tau)}_{m,m} \geq \mathbf{P}^{(t)}_{m,n}\mathbf{P}^{(\tau)}_{n,n}\mathbf{P}^{(s)}_{n,m} \tag{220}$$

so

$$\sum_{\tau=0}^{\infty} \mathbf{P}^{(\tau)}_{n,n} \leq \frac{1}{\mathbf{P}^{(t)}_{m,n}\mathbf{P}^{(s)}_{n,m}} \sum_{\tau=0}^{\infty} \mathbf{P}^{(t+s+\tau)}_{m,m} < \infty \tag{221}$$

□

Because of this theorem, we can speak of a transient or a recurrent class.

Theorem 3.8. *Every recurrent class is closed.*

Proof. If C is not closed, then there are $m \in C$, $n \notin C$ with $m \to n$ and $\tau \geq 1$ such that

$$P_m(X_\tau = n) > 1 \tag{222}$$

However, $n \not\to m$, so

$$P_m\big(V_m = \infty \big| X_\tau = n\big) = 0 \tag{223}$$

Therefore

$$P_m\big(V_m = \infty\big) = \sum_k P_m\big(V_m = \infty \big| X_\tau = k\big) < 1 \tag{224}$$

So m is not recurrent, and neither is C □

Theorem 3.9. *Every finite closed class is recurrent.*

Proof. Let C be such a class. Pick any initial distribution on C. Then $\sum_{m \in C} V_m = \infty$ (since C is closed). Since C is finite, some state must be visited infinitely often, so

$$1 = P\Big[\bigcup_{m\in C}\Big\{V_m = \infty\Big\}\Big] \leq \sum_{m \in C} P\big(V_m = \infty\big) \tag{225}$$

So, for some m,

$$0 < P\big(V_m = \infty\big) = P\big(H_m < \infty\big) P_m\big(V_m = \infty\big) \tag{226}$$

But $P_m(V_m = \infty)$ can only be 0 or 1, so it must be 1. Thus m is recurrent, and so is C. $\square$

Example XVII:
We shall talk about random walks later on, but this is a good example to warm up. A random walk on $\mathbb{Z}$ is a process in which a walker jumps from spot m to $m+1$ with probability p, and from m to $m-1$ with probability q:

$$\begin{tikzcd} \cdots \arrow[r, bend left, "p"] & -2 \arrow[r, bend left, "p"] \arrow[l, bend left, "q"] & -1 \arrow[r, bend left, "p"] \arrow[l, bend left, "q"] & 0 \arrow[r, bend left, "p"] \arrow[l, bend left, "q"] & 1 \arrow[r, bend left, "p"] \arrow[l, bend left, "q"] & 2 \arrow[r, bend left, "p"] \arrow[l, bend left, "q"] & \cdots \arrow[l, bend left, "q"] \end{tikzcd} \tag{227}$$

The walk starts from 0, and the question that we pose is: is 0 recurrent? We shall apply the first condition of theorem 3.6 to show that if the walk is unbiased ($p = q = 1/2$) it is, while if the walk is biased ($p \neq 1/2$, $q = 1-p$) it is not. If we start at 0, we can never be back at 0 after an odd number of steps, so $\mathbf{P}_{0,0}^{(2n+1)} = 0$ for all n. Consider now a sequence of $2n$ steps from 0 to 0. Independently of how many twists and turns we make, this will consist of n steps forward, and n steps backward and will have probability $p^n q^n$. We identify a trajectory of length $2n$ by deciding where we do the n steps forward, so there are $\left(\begin{array}{c} 2n \\ n \end{array}\right)$ different trajectories, that is

$$\mathbf{P}_{0,0}^{(2n)} = \left(\begin{array}{c} 2n \\ n \end{array}\right) p^n q^n \tag{228}$$

Using stirling approximation

$$n! \sim \sqrt{2\pi}\sqrt{n}\left(\frac{n}{e}\right)^n \qquad n \to \infty \tag{229}$$

we have

$$\mathbf{P}_{0,0}^{(2n)} = \frac{(2n)!}{(n!)^2}(pq)^n \sim \frac{(4pq)^n}{\sqrt{2\pi}\sqrt{\dfrac{n}{2}}} \tag{230}$$

If $p = q = 1/2$, we have $4pq = 1$ so, for N large enough and $n > N$,

$$\mathbf{P}_{0,0}^{(2n)} \geq \frac{1}{2\sqrt{2\pi n}} \tag{231}$$

and

$$\sum_{n=0}^{\infty} \mathbf{P}_{0,0}^{(2n)} \geq \frac{1}{2\sqrt{2\pi}} \sum_{n=N}^{\infty} \frac{1}{\sqrt{n}} = \infty \tag{232}$$

which shows that the walk is recurrent. If $p \neq q$ then $4pq = r < 1$ so, reasoning in a similar way,

$$\sum_n \mathbf{P}_{0,0}^{(2n)} \leq k \sum_n r^n < \infty \tag{233}$$

which shows that the walk is transient. (end of example)

3.4 Invariant distribution

Over long periods of time, the occupancy probability of the states of a Markov chain may settle to a stable distribution. A **measure** $\Lambda = [\lambda_m | m \in M]'$ is a vector with $\lambda_m \geq 0$ for all m. If $\sum \lambda_m = 1$, the vector is a **distribution**. An **invariant measure** is a measure such that

measure

invariant measure

$$\Lambda'\mathbf{P} = \Lambda' \qquad \lambda_m \geq 0 \tag{234}$$

An **invariant distribution** is defined analogously:

invariant distribution

$$\Pi'\mathbf{P} = \Pi' \qquad \pi_m \geq 0 \qquad \sum_{m \in M} \pi_m = 1 \tag{235}$$

If Λ is an invariant measure and $\sum_m \lambda_m < \infty$, then

$$\pi_m = \frac{\lambda_m}{\sum_{n \in M} \lambda_n} \tag{236}$$

is an invariant distribution. This is always possible if M is finite.

If M is finite, an invariant distribution always exists. We know that $\sum_n \mathbf{P}_{m,n} = 1$, which implies that for the vector $\mathbf{1} = [1, 1, \ldots, 1]'$ it holds

$$\mathbf{P}\mathbf{1} = \mathbf{1} \tag{237}$$

Therefore $\mathbf{P}$ has an eigenvalue equal to 1, and $\mathbf{1}$ is a right eigenvector of this eigenvalue. Standard linear algrbra arguments show that a left eigenvalue also exists and, by the Frobenius-Perron theorem, all its components are positive. Normalizing this left eigenvector we obtain (235).

The importance of invariant distributions is related to the long-term behavior of a chain. In particular, if the Markov chain settles down to a distribution, this is an invariant one.

Theorem 3.10. *Let M be finite and, for some $m \in M$ and all $n \in M$,*

$$\lim_{t \to \infty} \mathbf{P}^{(t)}_{m,n} = \pi_n \tag{238}$$

Then $\pi = [\pi_m | m \in M]'$ is an invariant distribution

Proof. We have

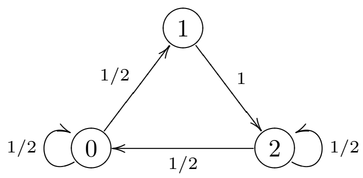
$$\sum_{m\in M}\pi_m = \sum_{m\in M}\lim_{t\to\infty}\mathbf{P}^{(t)}_{m,n} = \lim_{t\to\infty}\sum_{m\in M}\mathbf{P}^{(t)}_{m,n} = 1 \tag{239}$$

and

$$\pi_n = \lim_{t\to\infty}\mathbf{P}^{(t)}_{m,n} = \lim_{t\to\infty}\sum_{k\in M}\mathbf{P}^{(t-1)}_{m,k}\mathbf{P}_{k,n} = \sum_{k\in M}\lim_{t\to\infty}\mathbf{P}^{(t-1)}_{m,k}\mathbf{P}_{k,n} = \sum_{k\in M}\pi_k\mathbf{P}_{k,n} \tag{240}$$

that is $\pi' = \pi'\mathbf{P}$. □

Example XVIII:
Consider the chain



$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \tag{241}$$

We determine the stationary distribution by solving the equilibrium equation $\pi' = \pi'\mathbf{P}$:

$$[\pi_1, \pi_2, \pi_3] = [\pi_1, \pi_2, \pi_3]\begin{bmatrix} 0 & 1 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \tag{242}$$

that is:

$$\begin{aligned} \pi_1 &= \frac{1}{2}\pi_3 \\ \pi_2 &= \pi_1 + \frac{1}{2}\pi_2 \\ \pi_3 &= \frac{1}{2}\pi_2 + \frac{1}{2}\pi_3 \end{aligned} \tag{243}$$

These equations not independent and therefore under-constrained, but we can determine a unique solution by adding the normalization condition

$$\pi_1 + \pi_2 + \pi_3 = 1 \tag{244}$$

obtaining $\pi = \frac{1}{5}[1, 2, 2]$ (end of example)

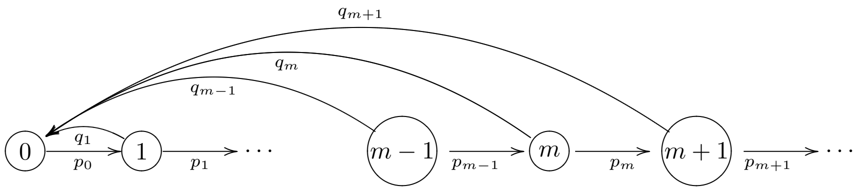
Sometimes it is more convenient to use the **balance equations**

Balance equations

$$\pi_m \mathbf{P}_{m,n} = \pi_n \mathbf{P}_{n,m} \qquad \text{for all } n, m \in M \tag{245}$$

If M is infinite, a stationary distribution doesn't necessarily exist.

Example XIX:
Consider the chain



(246)

with

$$\mathbf{P}_{m,m+1} = p_m \qquad \mathbf{P}_{m,0} = q_m = 1 - p_m \tag{247}$$

Then

$$\pi_0 = \sum_{m=0}^{\infty} \pi_m q_m \implies p_o \pi_0 = \sum_{m=1}^{\infty} \pi_m q_m \tag{248}$$

$$\pi_m = \pi_{m-1} p_{m-1} \quad m \geq 1 \tag{249}$$

Define the sequence r_m, $m \geq 0$ as

$$\begin{aligned} r_0 &= 1 \\ r_m = p_0 \cdots p_{m-1} &= r_{m-1} p_{m-1} \end{aligned} \tag{250}$$

and choose the p_m in such a way that

$$r \triangleq \prod_{m=0}^{\infty} p_m > 0 \tag{251}$$

The definition of r_m implies that $\pi_m = r_m \pi_0$ and

$$\begin{aligned}
\pi_0 = \sum_m \pi_m q_m &= \lim_{n\to\infty} \sum_{m=0}^{n} \pi_m q_m \\
&= \lim_{n\to\infty} \sum_{m=0}^{n} (r_m - r_{m-1})\pi_0 \\
&= \lim_{n\to\infty} (r_0 - r_{n+1})\pi_0 \\
&= (1-r)\pi_0
\end{aligned} \tag{252}$$

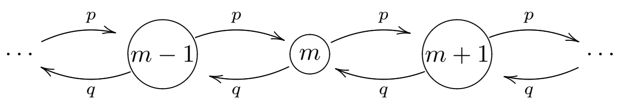
and, since $r > 0$, the only solution is $\pi_0 = 0$, from which, via (249), we have $\pi = 0$, therefore π is not a distribution. (end of example)

Even if a stationary distribution exists, it needs not be unique.

Example XX:
Let $\mathbf{P} = \mathbf{I}$ (the identity matrix), then any distribution is stationary. (end of example)

In this example, the chain is not irreducible, but the same is true for irreducible chains.

Example XXI:
Consider the chain



(253)

Then $\lambda_m = 1$ and $\lambda_m = (p/q)^m$ are both invariant measures. The measures are different (viz., the solution is not unique) if $p \neq q$, that is, if the chain is transient. (end of example)

On the other hand, the following lemma guarantees unicity under certain conditions.

Lemma 3.2. *If* $\mathbf{P}$ *is irreducible and recurrent, then* $\gamma_k = [\gamma_m^k | m \in M]'$ *satisfies* $\gamma_k' = \gamma_k' \mathbf{P}$*, and the solution is unique*

If the limit exists, then it must be a stationary distribution. But the limit may fail to exist.

Example XXII:
Consider the chain

$$\begin{tikzcd} 0 \arrow[r, bend left, "1"] & 1 \arrow[l, bend left, "1"] \end{tikzcd} \qquad \mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{254}$$

Then $\mathbf{P}^{2n} = \mathbf{I}$ and $\mathbf{P}^{2n+1} = \mathbf{P}$ so the limit of $\mathbf{P}^n$ does not exist. Intuitively, the chain jumps frpm one state to another without ever setting down to a stable solution. (end of example)

A state m is **aperiodic** if there is a t_0 such that $\mathbf{P}^{(t)}_{m,m} > 0$ for all $t > t_0$. The states in the previous example are not aperiodic since, for example, if the chain has $X_0 = 0$, then $\mathbf{P}^{(2t+1)}_{0,0} = 0$ for all t. A state that is not aperiodic is called **periodic** and it can be shown that there is a $\tau > 0$ such that $\mathbf{P}^{(t)}_{m,m} > 0$ is t is a multiple of τ and $\mathbf{P}^{(t)}_{m,m} = 0$ otherwis; the value τ is called the period of the state.

aperiodic state

periodic state

Lemma 3.3. *Let* $\mathbf{P}$ *be irreducible, and have an aperiodic state* m*. Then, for all* $n, k \in M$*,* $\mathbf{P}^{(t)}_{n,k} > 0$ *for* t *sufficiently large. In particular, all states are aperiodic.*

We define a state m **positive recurrent** if $mu_m < \infty$. It can be shown that if $\mathbf{P}$ is irreducible, then if m is recurrent, all the states are.

positive recurrent state

A chain is **regular** is for some t, all the elements of $\mathbf{P}^{(t)}$ are positive.

regular chain

A chain is **ergodic** if it is aperiodic, irreducible, and positive recurrent.

ergodic chain

Theorem 3.11. *Let* $\mathbf{P}$ *be the transition matrix of an ergodic Markov chain with invariant distribution* π*. Then, for any initial distribution,*

$$\lim_{t\to\infty} P(X_t = m) = \pi_m \tag{255}$$

In particular, for any initial distribution and any $m, n \in M$*,*

$$\lim_{t\to\infty} \mathbf{P}^{(t)}_{n,m} = \pi_m \tag{256}$$

In this case, we have:

$$\begin{aligned} \mu_m &= \frac{1}{\pi_m} \\ \gamma_m^n &= \frac{\pi_m}{\pi_n} \\ P\Big[\lim_{t\to\infty} \frac{V_t}{t} = \pi_m\Big] &= 1 \end{aligned} \tag{257}$$