Markov Chains

Let \( \{X_0, X_1, \ldots \} \) be a sequence of random variables that take values in some countable set \( S \), called the state space. Each \( X_n \) is a discrete random variable that takes one of \( N \) possible values where \( N = |S| \) and it is possible that \( N = \infty \).

Definition: The process \( \{X_n\}_{n \geq 1} \) is a Markov Chain if it satisfies

\[
P(X_n = s | X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) = P(X_n = s | X_{n-1} = x_{n-1})
\]

for all \( n \geq 1 \) and all \( s, x_0, x_1, \ldots x_{n-1} \in S \).

Thus a Markov chain is a random process that only has one step memory, i.e., given the present, the past and the future are independent.
The evolution of a chain is described by its transition probabilities. $P(X_{n+1} = j \mid X_n = i)$. In general, this probability depends on $n$, $i$, and $j$.

We shall only be concerned with "homogeneous" chains where the transition probability only depends on $i$ and $j$ and not $n$.

**Definition:** The chain $X$ is called homogeneous if $P(X_{n+1} = j \mid X_n = i) = P(X_1 = j \mid X_0 = i)$ for all $n$, $i$, $j$.

The transition matrix $P = (p_{ij})$ is the $1\times 1$ matrix of transition probabilities.

**Lemma:** For a transition matrix $P$, we have $\sum_j p_{ij} = 1$, $\forall i$.

**Proof:** $p_{ij} = P(X_{n+1} = j \mid X_n = i)$

$$\sum_j p_{ij} = \sum_j P(X_{n+1} = j \mid X_n = i)$$

$$= P(X_{n+1} \in S \mid X_n = i)$$

$$= 1.$$
Example

Random Walk on integers

\[ S = \{0, \pm 1, \pm 2, \ldots \} \] and transition probabilities

\[ P_{ij} = \begin{cases} 
    p & \text{if } j = i+1 \\
    q = 1-p & \text{if } j = i-1 \\
    0 & \text{otherwise}
\end{cases} \]

Population Evolution

Here \( S = \{0, 1, 2, \ldots \} \) and the probabilities could depend upon the birth and death rate of the members of the population.

Definition: The \( n \)-step transition matrix \( \mathbf{P}(m, m+n) = (p_{ij}(m, m+n)) \) is the matrix of \( n \)-step transition probabilities,

\[ p_{ij}(m, m+n) = \mathbb{P}(X_{m+n} = j \mid X_m = i). \]

By homogeneity, assumption \( \mathbf{P}(m, m+1) = \mathbf{P} \).
Theorem (Chapman-Kolmogorov Equations).

\[ p_{ij} (m, m+n) = P(X_{m+n} = j \mid X_m = i) \]

\[ = \sum_k p_{ik} (m, m+n) \cdot p_{kj} (m+n, m+n+\lambda). \]

\[ = \sum_k P(m, m+n+k) = P(m, m+n) \cdot P(m+n, m+n+k). \]

Iterating we obtain \( P(m, m+n) = P^n. \)

Proof:

\[ p_{ij} (m, m+n) = P(X_{m+n} = j \mid X_m = i) \]

\[ = \sum_k p(X_{m+n+k} = j \mid X_m = i) \]

\[ = \sum_k P(X_{m+n+k} = j \mid X_m = i) \cdot P(X_{m+n+k} = j \mid X_{m+n+k} = k, X_m = i) \]

\[ = \sum_k P(X_{m+n+k} = j \mid X_m = i) \cdot p(X_{m+n+k} = j \mid X_{m+n+k} = k) \]

In matrix notation:

\[ P(m, m+n) = P(m, m+n) \times P(m+n, m+n+\lambda). \]

Now \( P(m, m+n) = P(m, m+1) \cdot P(m+1, m+n) \)

\[ = P \cdot P(m+1, m+n) \]

\[ = P^n \]
Let $\mu^{(n)} = (\mu^{(n)}_1, \mu^{(n)}_2, \ldots, \mu^{(n)}_{n+1})$ be a row vector of probabilities where

$\mu^{(n)}_i = P(X_n = i)$

i.e., the probability of being in state $i$ at time $n$.

**Lemma:** $\mu^{(m)} = \mu^{(n)} P^n$ and hence $\mu^{(n)} = \mu^{(0)} P^n$

**Proof:**

$\mu^{(m)}_j = P(X_{m+n} = j) = \sum_{i} P(X_{m+n} = j \mid X_m = i) \cdot P(X_m = i)$

$= \sum_{i} \mu^{(m)}_i \cdot \pi_{ij}(m, m+n)$

$= \mu^{(m)} \cdot \pi_{ij}(m)$

(by homogeneity, $\pi_{ij}(m, m+n) = \pi_{ij}(m)$)

$= (\mu^{(m)} P^n)_j$

i.e., the $j$th component of the row vector $\mu^{(m)} P^n$.

**Example:**

Suppose initially you start in state 0

i.e., $\mu^{(0)} = [1, 0]$.

$\mu^{(0)} \beta = [0, 1] [\begin{array}{c} 1-x \alpha \\ \beta \end{array}] = [1-x \alpha ]$

$\mu^{(0)} P = [1-x \alpha] [\begin{array}{c} 1-x \alpha \\ \beta \end{array}] = [1-x \alpha \beta]$
Thus gradually the distribution starts moving away from the original \[ I \] or \[ J \] distribution, i.e. the chain loses its memory. We shall investigate this phenomenon in more detail later.

Classification of states

Definition: A state \( i \) is called persistent if

\[
P(X_n = i \text{ for some } n \geq 1 \mid X_0 = i) = 1
\]

i.e. having started at state \( i \), the probability of eventual return to state \( i \) is 1.

If

\[
P(X_n = i \text{ for some } n \geq 1 \mid X_0 = i) < 1,
\]

then the state \( i \) is called transient.

Let

\[
j_{ij}(n) = P(X_1 = i, X_2 = j, \ldots, X_n = j, X_{n+1} = i \mid X_0 = i)
\]

i.e. the probability that the first time \( i \) visits state \( j \), having started at \( X_0 = i \) is at time \( n \).

Now, note that the events

\[
\bigcap_{n=1}^{\infty} \{X_1 = i, X_2 = j, \ldots, X_n = j, X_{n+1} = i \mid X_0 = i \}
\]

are disjoint for all \( n \).

Define

\[
s_{ij} = \lim_{n \to \infty} j_{ij}(n) \quad \text{is the same as} \quad P(X_n \to i, \text{ for some } n \geq 1 \mid X_0 = i).
\]
Example of a Markov chain where some states may not be persistent.

Suppose $X_0 = 2$, then with finite probability the chain may never return to 2, since once it goes to state 1, it keeps transitioning between 0 and 1.
This also means that state $j$ is persistent if and only if $f_{i,j} = 1$.

Recall that $f_{i,j}(n) = P(X_n = j | X_0 = i)$

We want to find an analytical characterization of persistence.

Let us define the generating functions

$$P_{i,j}(s) = \sum_{n=0}^{\infty} f_{i,j}(n) s^n, \quad F_{i,j}(s) = \sum_{n=0}^{\infty} f_{i,j}(n) s^n$$

$$P_{i,j}(0) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}, \quad f_{i,j}(0) = 0, \quad \forall i \neq j.$$

Note that $f_{i,j} = F_{i,j}(0)$.

Theorem

(a) $P_{i,i}(s) = 1 + F_{i,i}(s) P_{i,i}(s)$

(b) $P_{i,j}(s) = F_{i,j}(s) P_{i,i}(s)$ if $i \neq j$

Proof:

Let $B_m = \{X_1 = j, X_2 = j, \ldots, X_{m-1} = j, X_m = j\}$

$A_m = \{X_m = j\}$

The $B_m$'s represent the event that the $1^{st}$ visit to $j$ happens at time $m$.

The $B_m$'s are disjoint.
Now,
\[ P(X_m | X_0 = i) = \sum_{n=1}^{m} P(A_m \cap B_n | X_0 = i) \quad \text{(total prob.)} \]

\[ = \sum_{n=1}^{m} P(B_n | X_0 = i) \cdot P(A_m | B_n) \]

\[ = \sum_{n=1}^{m} P(B_n | X_0 = i) \cdot P(A_m | X_n = j) \quad \text{(Markov property)} \]

\[ = \sum_{n=1}^{m} f_{ij}(n) \cdot p_{jj}(m-n). \]

\[ p_{ij}(m) = \sum_{n=1}^{m} f_{ij}(n) \cdot p_{jj}(m-n). \]

Note that this is the discrete convolution of \( f_{ij} \) and \( p_{jj} \).

Now,
\[ \sum_{m=1}^{s} p_{ij}(m) s^m = \sum_{m=1}^{s} s^m \left( \sum_{n=1}^{m} f_{ij}(n) \cdot p_{jj}(m-n) \right) \]
Changing the order of integration

\[ \sum_{n=1}^{\infty} \left( \sum_{m=1}^{n} \int f_{ij}(x) \, p_{ij}(m-x) \, dx \right) \]

\[ = \sum_{n=1}^{\infty} f_{ij}(n) \cdot s^n \cdot \sum_{k=0}^{\infty} p_{ij}(k) s^k \]

\[ = (F_{ij}(s) - f_{ij}(0)) \cdot p_{ij}(s) \]

\[ = F_{ij}(s) \cdot p_{ij}(s) \quad \text{(since } f_{ij}(0) = 0) \]

Now \[ \sum_{m=1}^{\infty} p_{ij}(m) s^m = p_{ij}(s) - F_{ij}(s) \]

- We have shown \[ p_{ij}(s) - F_{ij}(s) = F_{ij}(s) \cdot p_{ij}(s) \]

\[ \Rightarrow p_{ij}(s) = 8_{ij} + F_{ij}(s) \cdot p_{ij}(s) \]

**Corollary:**

(a) State \( j \) is persistent if and only if \[ \lim_{n \to \infty} p_{ij}(n) = \infty. \]

If this holds then \[ \lim_{n \to \infty} p_{ij}(n) = \infty \quad \text{for all } i \text{ such that } f_{ij} > 0. \]

**Proof:** From above, we have that

\[ p_{jj}(s) = 1 + F_{jj}(s) \cdot p_{jj}(s) \]

\[ \Rightarrow p_{jj}(s) = \frac{1}{1 - F_{jj}(s)} \quad \text{if } (s) < 1. \]
As $s \uparrow 1$, $P_{jj}(s) \to \infty$ if and only if $\mathbb{P}_{jj}(1) = 1$.

Now \[ \lim_{s \to 1} P_{jj}(s) = \frac{\sum_n P_{jj}(n)}{n} = \frac{\sum_n \mathbb{P}_{jj}(n)}{n} \to \infty \quad \text{for persistence} \]

Now $P_{jj}(s) = F_{jj}(s) P_{jj}(s)$

If $P_{jj}(s) \to \infty$ as $s \to 1$, noting that $\lim_{s \to 1} F_{jj}(s) \to 0$ (since $F_{jj}(s) > 0$)

we have

\[ \lim_{s \to 1} P_{jj}(s) = \frac{\sum_n P_{jj}(n)}{n} \to \infty \]

(b) State $j$ is transient if $\frac{\sum_n P_{jj}(n)}{n} < \infty$. If this holds, then $\frac{\sum_n \mathbb{P}_{jj}(n)}{n} < \infty$.

Proof: Very similar to (a).

Corollary: If $j$ is transient, then $\mathbb{P}_{jj}(n) \to 0$ as $n \to \infty$.

Proof: From above, we have that

$\frac{\sum_n \mathbb{P}_{jj}(n)}{n} < \infty$

$\mathbb{P}_{jj}(n) \to 0$ as $n \to \infty$
Intuitively, if a state is persistent then it gets visited a number of times while a transient state gets visited only finitely many times. This is because for a persistent state $i$, $\pi_i = 1$, so given that the chain is at state $i$, it is guaranteed that it will be visited again. Arguing recursively we have an intuitive argument for $\infty$ visits. On the other hand, $\pi_i \leq 1$ if $i$ is transient. So there is a possibility that the chain never returns to $i$. Eventually this event will happen and therefore the number of visits to $i$ is finite.

Let $T_j = \min \{ n \geq 1 : X_n = j \}$ be the time of 1st visit to $j$. Note $T_j = \infty$ if this visit never occurs.

$P(T_i = \infty | X_0 = i) > 0$ if and only if $i$ is transient.

And therefore

$E(T_i | X_0 = i) = \infty$ if $i$ is transient.

However, $E(T_i | X_0 = i)$ can be $\infty$ even for persistent $i$, since as we have seen before, there exist distributions with infinite mean even though the right of $\infty$ is undefined.

Definition: The mean recurrence time $\mu_i$ of state $i$ is defined as,

$\mu_i = E(T_i | X_0 = i) = \int_0^\infty f_i(x) \, dx$ if $i$ is persistent

$\infty$ if $i$ is transient.
Definition: \( i \rightarrow j \)

(a) \( i \) communicates with \( j \) if \( \pi_{ij}(m) > 0 \) for some \( m \geq 0 \).

(b) \( i \) inter-communicates with \( j \), denoted \( i \leftrightarrow j \), if \( i \rightarrow j \) and \( j \rightarrow i \).

Theorem: Let \( i \) be a persistent state and let \( i \rightarrow j \), then

1. \( j \) is persistent
2. \( \bar{f}_{ij} = \bar{f}_{ji} = 1 \)

Proof:
If \( i \rightarrow j \), then \( \bar{f}_{ij} = 1 \) by definition.

Let us consider \( i \leftrightarrow j \).

Now, \( i \rightarrow j \) \( \Rightarrow \) \( \bar{f}_{ij} \geq 0 \) \( \Rightarrow \) \( \exists m \geq 0 \) such that \( \pi_{ij}(m) > 0 \).

Let \( n_0 = \min \{ n \geq 1 : \pi_{ij}(n) > 0 \} \).

Then \( \pi_{ij}(n_0) > 0 \) and \( \pi_{ij}(m) = 0 \) for \( 1 \leq m < n_0 \).

Thus, there exists a sequence of states such that

\[ P(X_0 = i, X_1 = y_1, \ldots, X_{n_0-1} = y_{n_0-1}, X_{n_0} = j) > 0 \]

such that \( y_k \neq i, y_k \neq j \) for \( k = 1, \ldots, n_0 - 1 \).

We claim that \( \bar{f}_{ji} = 1 \).
To see this suppose that \( f_{ij} < 1 \).
\[ \Rightarrow \quad 1 - f_{ij} > 0 \]
\[ \text{i.e.} \quad P(\text{never visiting } i | X_0 = j) > 0. \]

Now, this means that
\[ P(X_0 = i, X_1 = y_1, \ldots, X_{n-1} = y_{n-1}, X_n = j, X_{n+1} = i, X_{n+2} = i, \ldots) \]
\[ = P(X_0 = i, X_1 = y_1, \ldots, X_{n-1} = y_{n-1}, X_n = j) \cdot P(X_{n+1} = i, X_{n+2} = i, \ldots | X_n = j) \]
\[ > 0 \quad \text{(by previous argument)} \]
\[ > 0 \quad \text{(as } 1 - f_{ij} > 0) \]
\[ > 0 \]

But we know that \( f_{ii} = 1 \) — the above conclusion is a contradiction.
\[ \therefore \quad f_{ij} = 1. \]
\[ \Rightarrow \quad \exists \, n, > 0 \text{ such that } P^n(j, i) > 0. \]

Now:
\[ P_{ij}(n) + P_{ij}(n+1) \geq P_{ii}(n) \cdot P_{ii}(n) \cdot P_{ij}(n+1) \]
\[ \sum_{n=0}^{\infty} P_{ij}(n) = \prod_{n=0}^{\infty} P_{ii}(n) > 0 \quad \text{(as } i \text{ is recurrent)} \]
\[ \text{permutation} \]
\[ \sum_{n=0}^{\infty} P_{ij}(n) \rightarrow \infty \quad \Rightarrow \quad j \text{ is persistent} \]

Applying a similar argument we can conclude \( f_{ij} = 1. \) \( \square \)
Theorem: Let i be a transient state and let i \(\leftrightarrow\) j. Then j is transient.

Proof:

Note that i \(\leftrightarrow\) j

i.e. \(\exists m \& n \) such that

\[ P_{ij}(m) > 0, \quad P_{ji}(n) > 0 \]

Now,

\[ P_{ii}(m+n) \geq P_{ij}(m) P_{ij}(n) P_{ji}(n) > 0 \]

\[ \Rightarrow \text{LHS term} \leq \frac{P_{ii}(m+n)}{P_{ii}(m)} \geq \frac{P_{ij}(n)}{P_{ii}(m)} \geq 0 \]

Now, LHS, \( \frac{P_{ii}(m+n)}{P_{ii}(m)} < \infty \) since i is transient.

\[ \therefore \text{RHS term} \leq \frac{P_{ji}(n)}{P_{ii}(m)} < \infty \Rightarrow j \text{ is transient.} \]

By a similar argument

Definition: A set of states C is called:

(a) Closed:

If \( P_{ij} = 0 \) for all \( i \in C, j \notin C \)

(b) Irreducible:

If i \(\leftrightarrow\) j for all \( i, j \in C. \)
Observation:
If \( C \) is a closed, irreducible set of states, then either every state in \( C \) is persistent or every state in \( C \) is transient.

- Follows from the previous two theorems.

Lemma: If the state space \( S \) is finite, then at least one state is persistent.

Proof: Suppose that all states in the chain are transient.
Recall that if \( j \) is transient then
\[
P_{ij}(n) \rightarrow 0 \quad \text{as} \ n \rightarrow \infty \quad \text{for all} \ i
\]

Now \( \sum_j P_{ij}(n) = 1 \) for any \( n \) since starting at \( i \) the chain has to transition to some \( j \in S \).

But \( \lim_{n \to \infty} \sum_j P_{ij}(n) \) by the bounded convergence theorem,
\[
= \sum_j \lim_{n \to \infty} P_{ij}(n) = 0
\]

\( \Rightarrow \) We have a contradiction
\( \Rightarrow \) at least one state has to be persistent.
If \( f_n \to f \) almost everywhere, \( |f_n| \leq g \) for all \( n \) and \( g \) is integrable, then \( \int f_n \, dm \to \int f \, dm \).

One version of the bounded convergence theorem

Let \( \lim_{n \to \infty} a_{xy}(n) = a_{xy} \).

Suppose \( |a_{xy}(n)| < b(x) \) for all \( n \).

and \( \int x \leq \lim_{n \to \infty} b(x) < \infty \).

Then,

\[
\lim_{n \to \infty} \int x \leq c(x) a_{xy}(n) = \int x \leq c(x) \lim_{n \to \infty} a_{xy}(n).
\]

\[
= \int x \leq c(x) a_{xy} \, \Box.
\]
Corollary: If \( \lambda \) is irreducible, then all states in \( S \) are persistent.

**Theorem:** For a Markov Chain, the state space \( S \) can be partitioned uniquely as

\[
S = T \cup S_p \cup S_R
\]

where \( S_T \) is the set of transient states, \( S_p \) is the set of persistent states.

Furthermore,

\[
S_R = C_1 \cup C_2 \cup \ldots
\]

where \( C_i \) are disjoint irreducible closed set of persistent states.

**Proof:** As we have seen there are only two types of states, either transient or persistent... \( S = S_T \cup S_R \).

Next we note that the relationships of \( \leftrightarrow \) is an equivalence relation and therefore divides \( S_p \) into equivalence classes that are precisely the \( C_i \)’s.

1. \( i \leftrightarrow i \) when \( i \) is persistent. (Reflexivity)
2. \( i \leftrightarrow j \Rightarrow j \leftrightarrow i \) by definition, and (Symmetry)
3. \( i \leftrightarrow j \), \( j \leftrightarrow k \Rightarrow i \leftrightarrow k \). (Obvious) (Transitivity).

\( \therefore \) The relation \( \leftrightarrow \) induces an equivalence relation.

We only need to show that the sets \( C_i \)’s are closed.
Suppose that there exists \( i \in C_h, j \notin C_h \) such that \( i \to j \), but \( i \not\to j \). (Note that \( i \to j \) is not possible since then \( j \in C_h \).

\[ \therefore \text{there exist } m \text{ such that } p_{ij}(m) > 0. \]

Now
\[ P(X_n = i \text{ for all } n \geq 1 / X_0 = i) = P(X_m = j | X_0 = i) > 0. \]

since once the chain enters \( j \) it can never return to \( C_h \).

This is a contradiction since \( i \) is persistent.
Suppose that a wireless channel has clustered errors. Whenever there is an error, the next packet will have errors with prob. 0.9. Whenever it is error-free, the next packet is error-free with prob. 0.99.

\[ \begin{array}{c}
0.99 \\
0.1 \\
0.01 \\
0.9 \\
\end{array} \]

more generally the state transition matrix can be

\[ P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} \]

Let us try to see intuitively what happens to \( P^n \).

Let \( P^n \) be given by,

\[ P^n = \begin{bmatrix} 1-a_n & a_n \\ b_n & 1-b_n \end{bmatrix} \]

Then \( P^{n+1} = \begin{bmatrix} 1-a_n & a_n \\ b_n & 1-b_n \end{bmatrix} \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} \]

\[ \Rightarrow a_{n+1} = (1-a_n)p + a_n(1-q) = p + a_n(1-p-q) \]

\[ \Rightarrow b_{n+1} = b_n(1-p) + (1-b_n)q = q + b_n(1-p-q) \]

Iterating the recursions we obtain:

\[ a_n = \frac{p}{1-(1-p-q)} = \frac{p}{P_{\text{rev}}} \]

\[ b_n = \frac{q}{1-(1-p-q)} = \frac{q}{P_{\text{rev}}} \]
and \( b_n \rightarrow b = \frac{a}{1 - (1 - p - q)} = \frac{a}{p + q} \).

\[
p^n = \begin{bmatrix}
1 - \frac{t}{p + q} & \frac{t}{p + q} \\
\frac{q}{p + q} & 1 - \frac{q}{p + q}
\end{bmatrix} = \frac{1}{p + q} \begin{bmatrix}
a & p \\
a & p
\end{bmatrix}
\]

i.e. as time progresses the \( n \)-step transition probability settle down.

Now consider any initial distribution \( \pi^{(0)} = [a \quad -a] \).

Then \( \lim_{n \rightarrow \infty} \pi^{(n)} = [a \quad -a] \begin{bmatrix}
a & p \\
a & p
\end{bmatrix} \frac{1}{p + q} = \begin{bmatrix}
a \frac{(a + (a - a)p)}{p + q} & \frac{p}{p + q}
\end{bmatrix}
\]

\[= \begin{bmatrix}
a & p \\
\frac{a}{p + q} & \frac{p}{p + q}
\end{bmatrix} \frac{1}{p + q}
\]

In general, irrespective of the initial distribution after a long time the chain has distribution \( \begin{bmatrix}
\frac{a}{p + q} & \frac{p}{p + q}
\end{bmatrix} \).
Now let us go back to the original example.

\[ p = 0.01 \quad q = 0.1 \]

The limiting distribution is

\[
\begin{bmatrix}
0.1 & 0.01 \\
0.11 & 0.11
\end{bmatrix}
\]

\[ = \begin{bmatrix}
\frac{10}{11} & \frac{1}{11}
\end{bmatrix}
\]

Note that it is more likely for a chain to return to the good state as against returning to a bad state. The difference is not much, but still in the steady state:

\[ P(X_n = 0) = \frac{10}{11} \quad \text{(good)} \]

\[ P(X_n = 1) = \frac{1}{11} \quad \text{(bad)} \]
Transmitter transmits one packet per slot. Prob. of error $\epsilon = 0.1$. $\epsilon = p$.

If there are 5 successive transmission failures, the transmitter concludes that the current link quality is bad & waits for some slots, probabilistically with $\gamma = 0.01$.

\[
\begin{bmatrix}
0 & 1 & 2 & 3 & 4 & 5 \\
0 & 1-p & p & 0 & 0 & 0 & 0 \\
1 & 1-p & 0 & p & 0 & 0 & 0 \\
2 & 1-p & 0 & 0 & p & 0 & 0 \\
3 & 1-p & 0 & 0 & 0 & p & 0 \\
4 & 1-p & 0 & 0 & 0 & 0 & p \\
5 & \gamma & 0 & 0 & 0 & 0 & 1-\gamma \\
\end{bmatrix}
\]

\[
(\pi_0 + \pi_1 \cdots \pi_4)(1-p) + \gamma \pi_5 = \pi_0
\]

\[
\begin{align*}
\pi_0 p &= \pi_1 \\
\pi_1 p &= \pi_2 \\
\pi_2 p &= \pi_3 \\
\pi_3 p &= \pi_4 \\
\pi_4 p + \pi_5(1-\gamma) &= \pi_5 \\
\Rightarrow \pi_5 &= \pi_4 \frac{p}{\gamma}
\end{align*}
\]
Now \( \sum_{i=0}^{\infty} \pi_i = 1 \)

\[ \implies \prod_{0} (1 + p + p^2 + p^3 + p^4 + \frac{p^5}{Q}) = 1 \]

\[ \implies \pi_0 \left( \frac{1 - p^5}{1 - p} + \frac{p^5}{Q} \right) = 1 \]

If \( p = 0.1 \), \( q = 0.01 \), then \( \pi_0 = \frac{1}{1 - 10^{-5} + \frac{10^{-5}}{10^{-2}}} \approx 9 \times 10^{-1} \)

\[ \prod_{i} = 9 \times 10^{-2} \quad \text{and} \quad \prod_{2} = 9 \times 10^{-3} \quad \text{and} \quad \prod_{3} = 9 \times 10^{-4} \quad \text{and} \quad \prod_{4} = 9 \times 10^{-5} \quad \text{and} \quad \prod_{5} = 9 \times 10^{-6} \]

Therefore, under this policy the chain spends about \( 9 \times 10^{-6} \) fraction of its time in state 5.
Stationary Distribution

As we had discussed before, Markov chains tend to lose their memory quite quickly, e.g., if we start in an initial distribution \( \pi^0 \), then the chain starts moving to another distribution quite quickly.

We hope that the distribution \( \pi_n \) converges or settles down after a long time.

Stationary Distribution: A vector \( \pi \) is called a stationary distribution if

(a) \( \pi_j \geq 0 \) for all \( j \in S \) and \( \sum_{j \in S} \pi_j = 1 \).

(b) \( \pi = \pi P \) where \( P \) is the transition matrix of the Markov chain, i.e., \( \pi_j = \sum_{i \in S} \pi_i P_{ij} \) for all \( j \in S \).

Observation:

\[ \pi P = (\pi P) \cdot P = \pi P^2 = \pi P^3 = \ldots = \pi. \]

i.e., if at some time the chain is in the distribution \( \pi \), then for all subsequent times, the distribution remains \( \pi \).

i.e., \( \pi P^n = \pi \) for all \( n \geq 0 \).
We have seen previously that any Markov chain can be decomposed into a set of transient states and closed sets of persistent states.

We now focus our attention on irreducible chains and investigate the existence of stationary distributions in them.

**Theorem:** An irreducible chain has a stationary distribution \( \pi \), if and only if all its states are non-null persistent. In this case, \( \pi \) is the unique stationary distribution and is given by

\[
\pi_i = \frac{1}{\mu_i} \quad \text{for } i \in S
\]

\( \mu_i \) = mean recurrence time of state \( i \).

(Recall \( T_i = \min \{ n \geq 1 : X_n = i \} \) and \( \mu_i = E(T_i | X_0 = i) \).

The above theorem is one of the most important in the theory of Markov chains. We prove it in many steps.

**Explicitly,**

First we demonstrate a solution to the matrix equation

\[
\pi = \pi P
\]

when the chain is irreducible and persistent.
Fix a state $k$ and define
\[ P_i(k) \text{ as the expected number of visits of the chain to state } i \text{ between two successive visits to state } k. \]

\[ P_i(k) = E(N_i \mid X_0 = k), \quad \text{where} \]

\[ N_i = \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n = i \land T_k < n\}} \]

where $T_k \to$ time of first return to state $k$.

\[ \begin{array}{c}
X = \cdots - i - \cdots - i - \cdots \\
X_0 = k & \quad \text{(visit to } k) \quad \text{(visit to } i) \\
\end{array} \]

w.p. 1 (by persistence)

Now $N_k = 1$ from the previous definition as the visit at the time of the $1^{st}$ visit to state $k$ is also counted.

\[ N_k = 1 \quad \Rightarrow \quad P_k(k) = E(N_k \mid X_0 = k) = 1 \]

Next,
\[ P_i(k) = E\left( \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n = i \land T_k > n\}} \mid X_0 = k \right) \]
\[ \frac{\alpha}{\sum_{n=1}^{\infty} P(X_n=i, \tau_k \geq n \mid X_0 = k)}. \]

New: \( \tau_k = \text{time of first return to state } k. \)

\[ \tau_k = \sum_{i \in S} N_i = 1 + \sum_{i \in S} N_i \]

since between two visits to state \( k \), the chain has to be in some state in \( S \).

\[ E(\tau_k \mid X_0 = k) = \sum_{i \in S} E(N_i \mid X_0 = k) \]

\[ = \sum_{i \in S} P_i(k). \]

Let \( p(k) \) denote the vector \([p_1(k), p_2(k), \ldots, p_n(k)]\).

**Lemma:** For any state \( k \) of a irreducible persistent chain, the vector \( p(k) \) is such that \( p_i(k) < \infty \) for all \( i \) and \( p(k) = p(k) P \) i.e. \( p(k) \) satisfies the matrix equation \( X = XP \).

**Proof:**

Let \( E_{ki}(n) = P(X_n=i, \tau_k \geq n \mid X_0 = k). \)

\( E_{ki}(n) \to \) probability that the chain reaches \( i \) in \( n \) steps without ever visiting state \( k \) again.
(Recall \( f_{k,e}(n) = \mathbb{P}(X_1 \neq k, X_2 \neq k, \ldots, X_m = k | X_0 = k) \).

Now,
\[
\frac{f_{k,e}(m+n)}{f_{k,i}(m)} \geq \frac{f_{k,i}(m)}{f_{k,e}(n)}
\]

since the RHS is just the probability of one set of events that comprise the LHS probability.

\( f_{k,i}(m) \) represents the probability of reaching \( i \) in \( m \) steps without visiting \( k \) and \( f_{k,e}(n) \) represents the probability of hitting \( k \) at the first time at the \( n \)th step having started in \( i \).

Next, note that the chain is irreducible (by assumption), i.e. there exists an \( n \) such that \( f_{k,e}(n) > 0 \). (we use irreducibility)

With this choice of \( n \), we have
\[
\frac{f_{k,i}(m)}{f_{k,e}(n)} \leq \frac{f_{k,e}(m+n)}{f_{k,e}(n)}
\]

\[\Rightarrow \sum_{m=1}^{\infty} f_{k,i}(m) = p_i(k) \leq \frac{1}{f_{k,e}(n)} \sum_{m=1}^{\infty} f_{k,e}(m+n) \leq \frac{1}{f_{k,e}(n)} < \infty.
\]

This proves the first assertion.

Now we show the second assertion in the lemma.

Note that
\[
p_i(k) = \sum_{n=1}^{\infty} f_{k,i}(n).
\]

and \( f_{k,i}(i) = p_{ki} \) (one step transition probability from \( k \) to \( i \)).
Now,
\[ l_{ki}(n) = \sum_{j : j \neq k} P(X_n = i, X_{n-1} = j, T_k \geq n | X_0 = k). \]
\[ = \sum_{j : j \neq k} l_{kj}(n-1) P_{ji} \quad \text{for } n \geq 2. \]

Now,
\[ p(n) = \sum_{n=1}^{\infty} l_{ki}(n) = \sum_{n=2}^{\infty} l_{ki}(n) + P_{ki}. \]
\[ \leq \sum_{n=2}^{\infty} l_{ki}(n) = p_i(k) - P_{ki}. \]
\[ = \sum_{j : j \neq k} l_{kj}(n-1) P_{ji} \quad (n \geq 2). \]
\[ = \sum_{j : j \neq k} \left( \sum_{n=2}^{\infty} l_{kj}(n-1) \right) P_{ji} \quad (\text{Interchanging order of summation}). \]
\[ = \sum_{j : j \neq k} p_j(k) P_{ji} \]
\[ \Rightarrow p_i(k) = P_{ki} + \sum_{j : j \neq k} p_j(k) P_{ji}. \]

But recall \( P_i(k) = 1 \) (by the assumption of persistence)
so we can rewrite the above equation as
\[ p_i(k) = \sum_{j} p_j(k) P_{ji}. \]
\[ \Rightarrow p(k) = p(k) P_i. \]
Thus for an irreducible, persistent chain we have a solution to

\[ x = \pi \cdot P \]
given by \( \pi(k) \).

Also note that

\[
\lim_{n \to \infty} \pi(k) = \mu_k \quad (\text{mean recurrence time of } k).
\]

\ implies \( \mu_k < \infty \) then we should be able to normalize

\( \pi(k) \) by \( \mu_k \) and find \( \bar{\pi} = \frac{\pi(k)}{\mu_k} \) as a stationary

distribution.

The condition \( \mu_k < \infty \) implies it is non-null persistent.

We have the following conclusion:

**Theorem:** If a chain is irreducible & persistent, there exists a
positive root \( x \) of the equation \( x = \pi \cdot P \), which is unique
up to a multiplicative constant. The chain is non-null if

\[
\lim_{n \to \infty} \pi(n) < \infty \quad \& \quad \text{null if} \quad \lim_{n \to \infty} \pi(n) = \infty.
\]
We are now in a position to prove the original theorem that we stated.

(i) If the chain has a stationary distribution \( \pi \), all states are non-null persistent.

If all states are transient, then \( \pi_{ij}(n) \to 0 \) as \( n \to \infty \) for all \( i \neq j \).

Now,
\[
\pi_{ij} = \sum_i \pi_i \pi_{ij}(n) \quad \text{for all } i \text{ and } j.
\]

Now, \( \lim_{n \to \infty} \sum_i \pi_i \pi_{ij}(n) = \sum_i \pi_i \lim_{n \to \infty} \pi_{ij}(n) = 0 \)

i.e. if all states are transient, then \( \pi_{ij} \to 0 \) as \( n \to \infty \)

and therefore \( \sum_{j \in S} \pi_{ij} = 1 \) as \( n \to \infty \).

\[\therefore \text{The states have to be persistent.}\]

We now show that all states are non-null and \( \pi_i = \frac{1}{M_i} \) for \( i \in S \).

Let \( X_0 \) have distribution \( \pi \).

Now, \( \pi_j = E(T_j / X_0 = j) \)

Note that \( T_j > 0 \), \( \pi_j = \sum_{n=1}^{\infty} P(T_j \geq n | X_0 = j) \)

\[\therefore \pi_{ij} \pi_j = \sum_{n=1}^{\infty} P(T_j \geq n, X_0 = j) \]
Now \( P(T_i > j, \quad x_0 = j) = P(x_0 = j) \)

and

\[
P(T_i > n, x_0 = j) = P(x_0 = j, x_m = j, 1 \leq m \leq n-1) = P(x_m = j, 1 \leq m \leq n-1) - P(x_m = j, 0 \leq m \leq n-1)
\]

(homogeneity)  

and initial distribution

is stationary.

\[
= P(x_m = j, \quad 1 \leq m \leq n-2) - P(x_m = j, \quad 0 \leq m \leq n-1),
\]

where \( a_n = P(x_m = j, 0 \leq m \leq n) \).

\[
\sum_{n=1}^{\infty} P(T_j > n, x_0 = j) = P(x_0 = j) + \sum_{n=2}^{\infty} (a_{n-2} - a_{n-1})
\]

\[
= P(x_0 = j) + a_0 - \lim_{n \to \infty} a_n
\]

\[
= P(x_0 = j) + P(x_0 \neq j) - \lim_{n \to \infty} a_n
\]

\[
= 1 - \lim_{n \to \infty} a_n
\]

Now \( \lim_{n \to \infty} a_n = P(x_m = j, \quad 0 \leq m \leq n) = 0 \) since \( j \) is persistent.

\[
\Rightarrow M_j = L
\]

\[
\Rightarrow \frac{1}{\Pi_j} M_j = \frac{1}{\Pi_j} \cdot \frac{L}{\Pi_j}
\]

\[
\Rightarrow M_j < \infty \quad \text{if} \quad \Pi_j > 0.
\]
Given that $P(X_0 = j) = \pi_j$

we have that

$$P(X_m + j, 1 \leq m \leq n-1) = P(X_{m+j}, 0 \leq m \leq n-2)$$

since LHS

$$= P(X_1 = j) \cdot P(X_2 = j | X_1 = j) \cdot P(X_3 = j | X_2 = j) \ldots$$

Now $P(X_1 = j) = \sum_{k \in S} P(X_1 = k) = \sum_{k \in \pi_j} P(X_0 = k)$ since $\pi^{(0)} = \pi^{(1)}$

Similarly, $P(X_{i+j} | X_{i+j}) = P(X_{i+j} | X_{i+j})$

since $P(X_{i+j} | X_{i+j})$

$$= \frac{P(X_{i+j}, X_{i+j} = j)}{P(X_{i+j} = j)} = \frac{\sum_{k \in \pi_j} P(X_i = k, X_{i+j} = j)}{P(X_j = j)} = \frac{\sum_{k \in \pi_j} P(X_0 = k, X_1 = j)}{P(X_j = j)}$$

$$= \frac{P(X_1 = j, X_0 = j)}{P(X_0 = j)}$$
Now it has to be the case that $p_{ij} > 0 \quad \text{for all } i \neq j$ if for some $j$ we have $\pi_j = 0$.

$\Rightarrow 0 = \pi_j = \sum_i \pi_i p_{ij}(n)$.

$\Rightarrow \pi_i = 0$ for all $i$ such that $p_{ij}(n) > 0$ for some $n$.

But the chain is irreducible $\Rightarrow \pi_i = 0$ for all $i$ $\Rightarrow \sum_i \pi_i = 1$.

$\Rightarrow \pi_j > 0$ for all $j \in S$.

$\Rightarrow \pi_j < \infty$ for all $j \in S$.

$\Rightarrow$ All states are persistent non-null.

(iii) All states are persistent non-null $\Rightarrow$ chain has a stationary distribution $\pi$.

(Proved earlier).