MTH 665

Lectures 8 - 16

Yevgeniy Kovchegov
Oregon State University
Topics:

- Markov chains
- Stationary distribution
- Reversible Markov chains
- Coupling method
- Mixing times
Markov chains.

Consider a sequence of random variables $X_0, X_1, X_2, \ldots$ with values in a discrete state space $S$.

The sequence $\{X_t\}_{t=0,1,\ldots}$ is said to be a discrete time Markov chain if it satisfies the following property, known as Markov property:

$$P(X_{t+1} = j \mid X_t = i, X_{t-1} = i_{t-1}, \ldots, X_1 = i_1, X_0 = i_0) = P(X_{t+1} = j \mid X_t = i)$$

In other words,

$$P(X_{t+1} \in A \mid \mathcal{F}_t) = E[1_{X_{t+1} \in A} \mid \mathcal{F}_t] = E[1_{X_{t+1} \in A} \mid X_t] = P(X_{t+1} \in A \mid X_t)$$

A Markov chain $\{X_t\}_{t=0,1,\ldots}$ is said to be time homogeneous if

$$P(X_{t+1} = j \mid X_t = i) = p(i,j) \quad \text{for all } t = 0,1,2,\ldots$$
Markov chains.

Consider a time homogeneous Markov chain $X_0, X_1, X_2, \ldots$ with a discrete state space $S$ and transition probabilities

$$P(X_{t+1} = j \mid X_t = i) = p(i,j) \quad \text{for all } t = 0, 1, 2, \ldots$$

Matrix (operator) $P = (p(i,j))_{i,j \in S}$ is called the transition probability matrix (operator). Then

$$\sum_{j \in S} p(i,j) = 1 \quad \forall i \in S$$

Example. Consider $S = \{0, 1\}$ (two states) and

$$P = \begin{pmatrix} 1 - p & p \\ q & 1 - q \end{pmatrix}$$
**Birth-and-death chain.** Consider state space

\[ S = \{0, 1, 2, \ldots\} \]

and a Markov chain \( \{X_t\}_{t=0,1,\ldots} \) on \( S \) with transition probabilities

\[ p(i, i + 1) = p_i, \quad p(i, i - 1) = q_i, \quad \text{and} \quad p(i, i) = r_i \]

satisfying \( q_0 = 0 \) and \( q_i + r_i + p_i = 1 \) \( \forall i \)

\[
P = \begin{pmatrix}
  r_0 & p_0 & 0 & 0 & \cdots \\
  q_1 & r_1 & p_1 & 0 & \cdots \\
  0 & q_2 & r_2 & p_2 & \cdots \\
  0 & 0 & q_3 & r_3 & \cdots \\
  \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}
\]

This is a Markov chain with only nearest neighbor transitions.
**Distribution of $X_t$.**

Consider a time homogeneous Markov chain $\{X_t\}_{t=0,1,...}$ on a discrete state space $S$.

Define the distribution of $X_t$ as a row vector of probability mass functions

$$\mu_t = \left( P(X_t = i) \right)_{i \in S}$$

ordered in the same way as enumeration of $S$ used in

$$P = \left( p(i,j) \right)_{i,j \in S}.$$ 

**Proposition.**

$$\mu_{t+1} = \mu_t P$$
**Distribution of** \( X_t \).

Define the distribution of \( X_t \) as a row vector of probability mass functions

\[
\mu_t = \left( P(X_t = i) \right)_{i \in S}
\]

ordered in the same way as enumeration of \( S \) used in

\[
P = \left( p(i, j) \right)_{i, j \in S}.
\]

**Proposition.**

\[
\mu_{t+1} = \mu_t P
\]

**Proof.** Denote by \( \mu_t(j) \) the \( j \)-th coordinate of \( \mu_t \).

Then

\[
\mu_{t+1}(j) = P(X_{t+1} = j) = \sum_{i \in S} P(X_t = i) P(X_{t+1} = j | X_t = i) = \sum_{i \in S} \mu_t(i) p(i, j)
\]
Chapman-Kolmogorov equation.

Consider a time homogeneous Markov chain $\{X_t\}_{t=0,1,...}$ on a discrete state space $S$. For an integer $s \geq 0$, let

$$p_s(i, j) = P(X_{t+s} = j \mid X_t = i) \quad \forall t \geq 0$$

be the $s$ step transition probabilities, and let

$$P_s = \left( p_s(i, j) \right)_{i,j \in S}$$

Observe that $P_0 = I$ and $P_1 = P$.

Chapman-Kolmogorov equation.

$$P_{s+t} = P_s P_t \quad \forall s, t \geq 0, \quad \text{and therefore} \quad P_t = P^t$$

Proof. $p_{s+t}(i, j) = \sum_{k \in S} p_s(i, k)p_t(k, j)$ follows from

$$P(X_{s+t} = j \mid X_0 = i) = \sum_{k \in S} P(X_s = k \mid X_0 = i)P(X_{s+t} = j \mid X_s = k) \quad \square$$
**Stationary distribution.**

For a homogeneous Markov chain with the transition probability matrix \( P = \left( p(i, j) \right)_{i,j \in S} \), the stationary distribution (aka ‘equilibrium distribution’) \( \pi \) is defined as follows:

\[
\pi P = \pi \iff \sum_{i \in S} \pi(i) p(i, j) = \pi(j).
\]

Thus \( \sum_{i} \pi(i) p(i, j) = \pi(j) \sum_{i} p(j, i) \), and for any state \( j \in S \),

\[
\sum_{i: i \neq j} \pi(i) p(i, j) = \sum_{i: i \neq j} \pi(j) p(j, i).
\]

Thus when restated in terms of traffic flow, the influx to the state \( j \) is equal to outflow from \( j \), for each \( j \). Thus the distribution stays unchanged.
Stationary distribution.

Consider a homogeneous Markov chain over a discrete state space $S$ with transition matrix $P$.

**Definition.** A homogeneous Markov chain is said to be **irreducible** if for any pair $x, y \in S$, there exists an integer $k \geq 1$ such that $p_k(x, y) > 0$.

**Definition.** A homogeneous Markov chain is said to be **aperiodic** if

$$\gcd\{k \geq 1 : p_k(x, x) > 0\} = 1 \quad \forall x \in S$$

Notice that if the Markov chain is **irreducible**, then

$$\gcd\{k \geq 1 : p_k(x, x) > 0\} = \gcd\{k \geq 1 : p_k(y, y) > 0\}$$

for any pair $x, y \in S$. 
**Irreducible Markov chains.**

**Definition.** A homogeneous Markov chain is said to be **irreducible** if for any pair \( x, y \in S \), there exists an integer \( k \geq 1 \) such that \( p_k(x, y) > 0 \).

**Example (reducible).** Consider \( S = \{1, 2, 3, 4, 5\} \) and

\[
P = \begin{pmatrix}
0.5 & 0.5 & 0 & 0 & 0 \\
0.5 & 0.5 & 0 & 0 & 0 \\
0 & 0 & 0.3 & 0.2 & 0.5 \\
0 & 0 & 0.5 & 0.3 & 0.2 \\
0 & 0 & 0.2 & 0.5 & 0.3 \\
\end{pmatrix}
\]

\( \pi' = (1/2, 1/2, 0, 0, 0) \) and \( \pi'' = (0, 0, 1/3, 1/3, 1/3) \) are both stationary distributions, as well as all

\[
\pi = \lambda \pi' + (1 - \lambda) \pi'' \quad \lambda \in [0, 1].
\]
Aperiodic Markov chains.

**Definition.** A homogeneous Markov chain is said to be aperiodic if
\[
\gcd\{k \geq 1 : p_k(x, x) > 0\} = 1 \quad \forall x \in S
\]

**Example (periodic).** Consider \( S = \{0, 1\} \) (two states) and
\[
P = \begin{pmatrix}
0 & 1 \\
1 & 0
\end{pmatrix}
\]
Here
\[
\gcd\{k \geq 1 : p_k(x, x) > 0\} = 2 \quad \forall x \in S
\]
The stationary distribution \( \pi = (0.5, 0.5) \) is unique.
However, \( \lim_{t \to \infty} p_t(x, y) \) does not exist.
Stationary distribution.

The following is version of ergodicity result for Markov chains over a finite state space $S$.

**Theorem (Ergodicity).** Consider an irreducible homogeneous Markov chain over a finite state space $S$. Then there exists a unique stationary distribution $\pi$. Furthermore, if the Markov chain is aperiodic, then

$$\lim_{t \to \infty} p_t(x, y) = \pi(y) \quad \forall x, y \in S.$$ 

**Example.** Consider $S = \{0, 1\}$ (two states) and

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \quad p, q \in (0, 1)$$
**Stationary distribution.**

**Example.** Consider $S = \{0, 1\}$ (two states) and

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \quad p, q \in (0, 1)$$

**Eigenvalues:** $\lambda_1 = 1$ and $\lambda_2 = 1 - p - q$.

**Matrix theory:**
For an eigenvalue $\lambda_i$ of $P$, the **left eigenvector** $u_i$ is a nonzero **row vector** satisfying $u_i P = \lambda_i u_i$.

The **right eigenvector** $v_i$ is a nonzero **column vector** satisfying $P v_i = \lambda_i v_i$.

**Spectral Theorem:** one can construct biorthogonal ($u_i v_j = 0$ for $i \neq j$) pair of left and right eigenbases $\{u_i\}_{i \in S}$ and $\{v_i\}_{i \in S}$, and

$$P = \sum_i \lambda_i \frac{v_i u_i}{u_i v_i}, \quad \text{and therefore,} \quad P^t = \sum_i \lambda_i^t \frac{v_i u_i}{u_i v_i}$$
Stationary distribution.

Example. Consider $S = \{0, 1\}$ (two states) and

$$P = \begin{pmatrix} 1 - p & p \\ q & 1 - q \end{pmatrix} \quad p, q \in (0, 1)$$

Eigenvalues: $\lambda_1 = 1$ and $\lambda_2 = 1 - p - q$.

Left eigenvectors: $u_1 = \pi = \left( \frac{q}{p+q}, \frac{p}{p+q} \right)$ and $u_2 = (1, -1)$

Right eigenvectors: $v_1 = 1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} p \\ -q \end{pmatrix}$

Spectral Theorem:

$$P^t = \sum_i \lambda_i^t \frac{v_i u_i}{u_i v_i} = 1\pi + (1 - p - q)^t \frac{v_2 u_2}{p + q}$$

$$= \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{q}{p+q} & \frac{p}{p+q} \end{pmatrix} + \frac{(1 - p - q)^t}{p + q} \begin{pmatrix} p & -p \\ -q & q \end{pmatrix}$$
Stationary distribution.

Example. Consider $S = \{0, 1\}$ (two states) and

$$P = \begin{pmatrix} 1 - p & p \\ q & 1 - q \end{pmatrix} \quad p, q \in (0, 1)$$

Eigenvalues: $\lambda_1 = 1$ and $\lambda_2 = 1 - p - q$.

Left eigenvectors: $u_1 = \pi = \left( \frac{q}{p+q}, \frac{p}{p+q} \right)$ and $u_2 = (1, -1)$

Right eigenvectors: $v_1 = 1 = \left( \begin{smallmatrix} 1 \\ 1 \end{smallmatrix} \right)$ and $v_2 = \left( \begin{smallmatrix} p \\ -q \end{smallmatrix} \right)$

Spectral Theorem:

$$P^t = \sum_i \lambda_i^t \frac{v_i u_i}{u_i v_i} = 1\pi + (1-p-q)^t \frac{v_2 u_2}{p+q} \rightarrow 1\pi = \left( \begin{smallmatrix} \frac{q}{p+q} \\ \frac{p}{p+q} \end{smallmatrix} \right)$$

as $t \rightarrow \infty$. Hence, $\lim_{t \rightarrow \infty} p_t(x, y) = \pi(y) \quad \forall x, y \in S$. 
Stationary distribution.

**Theorem (Ergodicity).** Consider an irreducible homogeneous Markov chain over a finite state space $S$. Then there exists a unique stationary distribution $\pi$. Furthermore, if the Markov chain is aperiodic, then

$$\lim_{t \to \infty} p_t(x, y) = \pi(y) \quad \forall x, y \in S.$$

The linear algebraic proof uses the Perron-Frobenius Theorem to prove the above version of ergodicity for Markov chains with finite state space $S$. 
Stationary distribution.

Perron-Frobenius Theorem. If $A = (a_{i,j}) \in \mathbb{R}^{d \times d}$ is an irreducible nonnegative matrix, then there exists a real eigenvalue $\rho \geq 0$, called Perron-Frobenius eigenvalue, such that

- $|\lambda| \leq \rho$ for any other eigenvalue $\lambda$.
- $\rho$ is simple (i.e., its left and right eigenspaces are one-dimensional), and it has left and right eigenvectors whose coordinates are all positive.

- $\min_{i} \sum_{j=1}^{d} a_{i,j} \leq \rho \leq \max_{i} \sum_{j=1}^{d} a_{i,j}$.

- $A$ has exactly $h$ eigenvalues in $\mathbb{C}$ with absolute value $\rho$:
  $$\rho e^{2\pi i k/h} \quad k = 0, \ldots, h-1,$$
  where $h$ is the period of $A$. 
Stationary distribution.

Matrix $A$ is irreducible if $\forall i, j \exists m \in \mathbb{N}$ such that $(A^m)_{i,j} > 0$.

Consider an irreducible homogeneous Markov chain over a finite state space $S$. Then $P$ is an irreducible nonnegative matrix. Hence, Perron-Frobenius Theorem applies.

The Perron-Frobenius eigenvalue $\rho$ of $P$ satisfies

$$1 = \min_i \sum_{j=1}^{d} p(i,j) \leq \rho \leq \max_i \sum_{j=1}^{d} p(i,j) = 1.$$ 

Hence, $\rho = 1$.

Finally, $\rho = 1$ is a simple eigenvalue, and it has left and right eigenvectors with all positive coordinates:

There exists a unique distribution $\pi$ such that $\pi P = \pi$.

We know that $P1 = 1$ as the rows of $P$ add up to 1.
Stationary distribution.

Consider an irreducible homogeneous Markov chain over a finite state space $S$. Then $P$ is an irreducible nonnegative matrix. Hence, Perron-Frobenius Theorem applies with $\rho = 1$.

There exists a unique distribution $\pi$ such that $\pi P = \pi$.

We know that $P1 = 1$ as the rows of $P$ add up to 1.

For an irreducible homogeneous Markov chain, matrix $P$ has period

$$h = \gcd\{k \geq 1 : p_k(x,x) > 0\} \text{ same for all } x \in S.$$

Spectral Theorem:

$$P^t = \sum_{i=1}^{\lfloor S \rfloor} \lambda_i^t \frac{v_i u_i}{u_i v_i} = 1\pi + \sum_{i=2}^{\lfloor S \rfloor} \lambda_i^t \frac{v_i u_i}{u_i v_i}.$$
**Stationary distribution.**

For an irreducible homogeneous Markov chain, matrix $P$ has period

$$h = \gcd\{k \geq 1 : p_k(x,x) > 0\} \text{ same for all } x \in S.$$ 

By Perron-Frobenius Theorem, $P$ has exactly $h$ eigenvalues with absolute value $\rho = 1$:

$$e^{2\pi ik/h} \quad k = 0, \ldots, h-1.$$ 

**Spectral Theorem:**

$$P^t = \sum_{i=1}^{\left|S\right|} \lambda_i^t \frac{v_i u_i}{u_i v_i} = 1\pi + \sum_{i=2}^{\left|S\right|} \lambda_i^t \frac{v_i u_i}{u_i v_i}.$$ 

Hence, if $P$ is aperiodic ($h = 1$), $P^t \to 1\pi$ as $t \to \infty$. 

If $h > 1$, $P^t$ does not have a limit as $t \to \infty$. 
Recurrent and transient states.

We will use the following notations:

\[ P_x(A) = P(A \mid X_0 = x) \quad \text{and} \quad E_x[Y] = E[Y \mid X_0 = x]. \]

For \( x \in S \), consider the first hitting time

\[ T_x = \min\{n \geq 1 : X_n = x\}. \]

**Definition.** A state \( x \in S \) is said to be **recurrent** if

\[ P_x(T_x < \infty) = 1. \]

A recurrent state \( x \in S \) is **positive recurrent** if

\[ E_x[T_x] < \infty. \]

Otherwise it is **null recurrent**.

**Definition.** A state \( x \in S \) is said to be **transient** if

\[ P_x(T_x < \infty) < 1. \]
**Stationary distribution.**

**Definition.** A state $x \in S$ is said to be positive recurrent if

$$E_x[T_x] < \infty$$

The following is a version of ergodicity theorem for a general discrete state space $S$.

**Theorem (Ergodicity).** Consider an irreducible homogeneous Markov chain over a discrete state space $S$. If all of its states are positive recurrent, then there exists a unique stationary distribution $\pi$ such that

$$\pi(x) = \frac{1}{E_x[T_x]}.$$  

Furthermore, if the Markov chain is aperiodic,

$$\lim_{t \to \infty} p_t(x, y) = \pi(y) \quad \forall x, y \in S.$$
Stationary distribution.

Theorem (Ergodicity). Consider an irreducible homogeneous Markov chain over a discrete state space $S$. If all of its states are positive recurrent, then there exists a unique stationary distribution $\pi$ such that

$$\pi(x) = \frac{1}{E_x[T_x]}.$$ 

Furthermore, if the Markov chain is aperiodic,

$$\lim_{t \to \infty} p_t(x, y) = \pi(y) \quad \forall x, y \in S.$$

The probabilistic proof can be done in two steps. First, proving existence/uniqueness, and then establishing convergence.
Lemma (Existence/Uniqueness). Consider an irreducible homogeneous Markov chain over a discrete state space $S$, all of whose states are positive recurrent. Then

$$
\pi(x) = \frac{1}{E_x[T_x]}
$$

is the unique stationary distribution.

Lemma (Convergence). Consider an irreducible aperiodic homogeneous Markov chain over a discrete state space $S$. If there is a stationary distribution $\pi$, then

$$
\lim_{t \to \infty} p_t(x, y) = \pi(y) \quad \forall x, y \in S.
$$
**Stationary distribution.**

**Lemma (Existence/Uniqueness).** Consider an irreducible homogeneous Markov chain over a discrete state space $S$, all of whose states are positive recurrent. Then

$$
\pi(x) = \frac{1}{E_x[T_x]}
$$

is the unique stationary distribution.

**Proof.** For a given $x \in S$, let

$$
\nu_x(y) = \sum_{n=0}^{\infty} P_x(X_n = y, T_x > n) = E_x \left[ \sum_{n=0}^{\infty} 1\{X_n = y, T_x > n\} \right]
$$

be the mean number of visits to state $y \in S$ between the times 0 and $T_x$. Then,

$$
\sum_{z \in S} \nu_x(z)p(z, y) = \nu_x(y), \quad \text{where } \nu_x(x) = 1.
$$
Proof. For a given $x \in S$, let $\nu_x(y) = \sum_{n=0}^{\infty} P_x(X_n = y, T_x > n)$.
Then,
$$\sum_{z \in S} \nu_x(z)p(z, y) = \nu_x(y), \text{ where } \nu_x(x) = 1.$$
Indeed, for all $y \neq x$, we have

$$\sum_{z \in S} \nu_x(z)p(z, y) = \sum_{z \in S} \sum_{n=0}^{\infty} P_x(X_n = z, T_x > n)P_x(X_{n+1} = y \mid X_n = z)$$

$$= \sum_{n=0}^{\infty} \sum_{z \neq x} P_x(X_n = z, T_x > n)P_x(X_{n+1} = y \mid X_n = z, T_x > n)$$

$$= \sum_{n=0}^{\infty} P_x(X_{n+1} = y \mid T_x > n) = \sum_{n=0}^{\infty} P_x(X_{n+1} = y \mid T_x > n+1) = \nu_x(y)$$

by Markov property, as $\{T_x > n-1\} \in \mathcal{F}_{n-1}$ and for $z \neq x$,
$$\{X_n = z, T_x > n\} = \{X_n = z, T_x > n-1\}.$$
Proof. For a given \( x \in S \), let \( \nu_x(y) = \sum_{n=0}^{\infty} P_x(X_n = y, T_x > n) \).

Then,

\[
\sum_{z \in S} \nu_x(z)p(z, y) = \nu_x(y), \quad \text{where} \quad \nu_x(x) = 1.
\]

In case of \( y = x \), we have

\[
\sum_{z \in S} \nu_x(z)p(z, x) = \sum_{z \in S} \sum_{n=0}^{\infty} P_x(X_n = z, T_x > n)P_x(X_{n+1} = x \mid X_n = z)
\]

\[
= P_x(X_1 = x) + \sum_{n=1}^{\infty} \sum_{z \neq x} P_x(X_n = z, T_x > n)P_x(X_{n+1} = x \mid X_n = z, T_x > n)
\]

\[
= P_x(X_1 = x) + \sum_{n=1}^{\infty} P_x(X_{n+1} = x, T_x > n) = P_x(T_x < \infty) = 1 = \nu_x(x)
\]

as \( x \) is a recurrent state.
Stationary distribution.

Proof (continued). Notice that

$$\sum_{y \in S} \nu_x(y) = \sum_{y \in S} \sum_{n=0}^{\infty} P_x(X_n = y, T_x > n) = \sum_{n=0}^{\infty} P_x(T_x > n) = E_x[T_x].$$

Thus

$$\sum_{z \in S} \nu_x(z)p(z, y) = \nu_x(y)$$

yields

$$\pi(y) = \frac{\nu_x(y)}{E_x[T_x]}$$

is a stationary measure \(\forall x \in S\). The existence of stationary distribution follows.

Since we have shown the existence of a stationary distribution, our next goal is to show it is unique. Consider a stationary distribution \(\pi\).
Stationary distribution.

Proof (continued). For any pair $a, b \in S$, a stationary distribution $\pi$ should satisfy

$$\pi(b) = \sum_{x_1 \in S} \pi(x_1)p(x_1, b) = \pi(a)p(a, b) + \sum_{x_1 \neq a} \pi(x_1)p(x_1, b)$$

Similarly, $$\pi(x_1) = \pi(a)p(a, x_1) + \sum_{x_2 \neq a} \pi(x_2)p(x_2, x_1)$$

Substituting, we have

$$\pi(b) = \pi(a)p(a, b) + \pi(a)\sum_{x_1 \neq a} p(a, x_1)p(x_1, b) + \sum_{x_1 \neq a} \pi(x_2)p(x_2, x_1)p(x_1, b),$$

where $$\pi(x_2) = \pi(a)p(a, x_2) + \sum_{x_3 \neq a} \pi(x_3)p(x_3, x_2)$$
Stationary distribution.

Proof (continued). We have

\[ \pi(b) = \pi(a) p(a, b) + \pi(a) \sum_{x_1 \neq a} p(a, x_1) p(x_1, b) + \sum_{\substack{x_1 \neq a \ \ x_2 \neq a}} \pi(x_2) p(x_2, x_1) p(x_1, b), \]

where \[ \pi(x_2) = \pi(a) p(a, x_2) + \sum_{x_3 \neq a} \pi(x_3) p(x_3, x_2) \]

Substituting, we have

\[ \pi(b) = \pi(a) p(a, b) + \pi(a) \sum_{x_1 \neq a} p(a, x_1) p(x_1, b) + \pi(a) \sum_{\substack{x_1 \neq a \ \ x_2 \neq a}} p(a, x_2) p(x_2, x_1) p(x_1, b) \]

\[ + \sum_{\substack{x_1 \neq a \ \ x_2 \neq a \ \ x_3 \neq a}} \pi(x_3) p(x_3, x_2) p(x_2, x_1) p(x_1, b) \]

and so on...
Proof (continued). After \( n \) iterations we have

\[
\pi(b) = \pi(a)p(a, b) + \pi(a) \sum_{x_1 \neq a} p(a, x_1)p(x_1, b)
\]

\[+ \ldots + \pi(a) \sum_{x_1 \neq a} p(a, x_{n-1})p(x_{n-1}, x_{n-2})\ldots p(x_1, b)\]

\[+ \sum_{x_1 \neq a} \ldots \sum_{x_{n-1} \neq a} \pi(x_n)p(x_n, x_{n-1})p(x_{n-1}, x_{n-2})\ldots p(x_1, b)\]

which rewrites as

\[
\pi(b) = \pi(a)P_a(T_a \geq 1, X_1 = b) + \pi(a)P_a(T_a \geq 2, X_2 = b) + \ldots
\]

\[+ \ldots + \pi(a)P_a(T_a \geq n, X_n = b) + \sum_{y \neq a} \pi(y)P_y(T_a \geq n, X_n = b)\]
Proof (continued). After \( n \) iterations we have

\[
\pi(b) = \pi(a) \sum_{k=1}^{n} P_a(T_a \geq k, X_k = b) + \sum_{y \neq a} \pi(y) P_y(T_a \geq n, X_n = b)
\]

Summing over all \( b \in S \), we obtain

\[
1 = \sum_{b \in S} \pi(b) = \pi(a) \sum_{k=1}^{n} P_a(T_a \geq k) + \sum_{y \neq a} \pi(y) P_y(T_a \geq n)
\]

Since for all \( y \in S \),

\[
P_y(T_a \geq n) \leq \frac{E_y[T_a]}{n} \rightarrow 0
\]

as \( n \to \infty \), we have

\[
1 = \pi(a) \sum_{k=1}^{\infty} P_a(T_a \geq k) = \pi(a) E_a[T_a]. \quad \text{Hence, } \pi(a) = \frac{1}{E_a[T_a]}.
\]
Stationary distribution.

**Lemma (Convergence).** Consider an irreducible aperiodic homogeneous Markov chain over a discrete state space $S$. If there is a stationary distribution $\pi$, then

$$
\lim_{t \to \infty} p_t(x, y) = \pi(y) \quad \forall x, y \in S.
$$

**Proof.** We use the coupling method. We let $(X_n, Y_n)$ be a homogeneous Markov chain on $S \times S$ with transition probabilities

$$
p((i_1, i_2), (j_1, j_2)) = \begin{cases} 
p(i_1, j_1)p(i_2, j_2) & \text{if } i_1 \neq i_2, \\
p(i_1, j_1) & \text{if } i_1 = i_2 \text{ and } j_1 = j_2, \\
0 & \text{if } i_1 = i_2 \text{ but } j_1 \neq j_2.
\end{cases}
$$

The process evolves according to

$$
P\left((X_{n+1}, Y_{n+1}) = (j_1, j_2) \mid (X_n, Y_n) = (i_1, i_2)\right) = p((i_1, i_2), (j_1, j_2)).$$
Proof (continued). We use the coupling method.
Proof (continued). We let \((X_n, Y_n)\) be a homogeneous Markov chain on \(S \times S\) with transition probabilities

\[
p((i_1, i_2), (j_1, j_2)) = \begin{cases} 
p(i_1, j_1)p(i_2, j_2) & \text{if } i_1 \neq i_2, \\
p(i_1, j_1) & \text{if } i_1 = i_2 \text{ and } j_1 = j_2, \\
0 & \text{if } i_1 = i_2 \text{ but } j_1 \neq j_2.
\end{cases}
\]

Notice that each margin, \(X_n\) and \(Y_n\), is a Markov chain with transition probabilities \(P = \left(p(i, j)\right)_{i, j \in S}\).

Matrix \(P\) is associated with an irreducible aperiodic homogeneous Markov chain. Thus,

\[
\exists m \in \mathbb{N} \text{ s.t. } P^m > 0, \text{ i.e., } P^m \text{ has all positive coordinates.}
\]

Consequently, there exists a finite coupling time

\[
\tau = \min\{n \geq 0 \mid X_n = Y_n\}.
\]
Proof (continued). Matrix $P$ is associated with an irreducible aperiodic homogeneous Markov chain. Thus, \[ \exists m \in \mathbb{N} \text{ s.t. } P^m > 0, \] i.e., $P^m$ has all positive coordinates.

Consequently, there exists a finite coupling time
\[ \tau = \min\{n \geq 0 \mid X_n = Y_n\}. \]
For $n \geq \tau$, they $X_n$ and $Y_n$ evolve as a single Markov chain with transition probabilities $p(i,j)$:
\[ X_{\tau} = Y_{\tau}, \quad X_{\tau+1} = Y_{\tau+1}, \quad X_{\tau+2} = Y_{\tau+2}, \quad X_{\tau+3} = Y_{\tau+3}, \ldots \]

Indeed, there exists $\varepsilon > 0$ such that for all $x, y \in S$,
\[ P(\tau \leq m \mid (X_0, Y_0) = (x, y)) = \sum_{z \in S} P(X_m = Y_m = z \mid (X_0, Y_0) = (x, y)) \]
\[ \geq \sum_{z \in S} p_m(x, z)p_m(y, z) \geq \varepsilon. \]
Hence, $P(\tau < \infty) = 1$. 

Proof (continued). Each margin, $X_n$ and $Y_n$, is a Markov chain with transition probabilities $P = \left( p(i, j) \right)_{i, j \in S}$.

There exists a finite coupling time

$$\tau = \min \{ n \geq 0 \mid X_n = Y_n \}.$$ 

For any given $x \in S$, let $X_0 = x$, and $Y_0$ be distributed with probabilities

$$P(Y_0 = y) = \pi(y) \quad \forall y \in S.$$ 

Then, for all $n \in \mathbb{N}$, $P(Y_n = y) = \pi(y) \ \forall y \in S$. Hence,

$$\sum_{y \in S} |p_t(x, y) - \pi(y)| = \sum_{y \in S} |P(X_t = y) - P(Y_t = y)|$$

$$= \sum_{y \in S} \left| P(X_t = y, X_t \neq Y_t) - P(Y_t = y X_t \neq Y_t) \right|$$

$$\leq 2P(X_t \neq Y_t) = 2P(\tau \geq t) \rightarrow 0 \text{ as } t \rightarrow \infty. \quad \square$$
Stationary distribution.

For a homogeneous Markov chain with the transition probability matrix $P = \left( p(i, j) \right)_{i, j \in S}$, the stationary distribution (aka ‘equilibrium distribution’) $\pi$ is defined as follows:

$$\pi P = \pi \iff \sum_{i \in S} \pi(i)p(i, j) = \pi(j).$$

Thus $\sum_i \pi(i)p(i, j) = \pi(j)\sum_i p(j, i)$, and for any state $j \in S$,

$$\sum_{i: i \neq j} \pi(i)p(i, j) = \sum_{i: i \neq j} \pi(j)p(j, i).$$

Thus when restated in terms of traffic flow, the influx to the state $j$ is equal to outflow from $j$, for each $j$. Thus the distribution stays unchanged.
Stationary distribution and reversibility.

The following are the detailed balance conditions (d.b.c.) also called time reversibility:

\[ \pi(i)p(i, j) = \pi(j)p(j, i) \quad \forall i, j \in S. \]

Restated in terms of traffic flow: for every pair of states \( i \) and \( j \) the traffic in between them is balanced (equalized), i.e. the traffic flow from \( i \) to \( j \) equals to the traffic flow from \( j \) to \( i \).

Observe that if d.b.c. are satisfied, the distribution will not change with time, i.e. \( \pi \) is stationary;

\[ \sum_{i: \, i \neq j} \pi(i)p(i, j) = \sum_{i: \, i \neq j} \pi(j)p(j, i) \quad \forall j \in S. \]
Stationary distribution and reversibility.

Observe that in the case of a birth-and-death chain, the definition of a stationary distribution

\[ \pi P = \pi \iff \sum_{i \in S} \pi(i)p(i, j) = \pi(j). \]

can be rewritten as

\[ \pi_k = p_{k-1}\pi_{k-1} + r_k\pi_k + q_{k+1}\pi_{k+1} \quad \text{for } k = 1, 2, \ldots \]

The above equations can be shown to be equivalent to the detailed balance conditions (d.b.c.)

\[ \pi_{k-1}p_{k-1} = q_k\pi_k. \]

Hence, \( \pi_k = \frac{p_{k-1}}{q_k}\pi_{k-1} \) for \( k = 1, 2, \ldots \).
Stationary distribution and reversibility.

In the case of a birth-and-death chain, \( \pi_k = \frac{p_{k-1}}{q_k} \pi_{k-1} \) and

\[
\pi_k = \frac{p_0 \cdots p_{k-1}}{q_1 \cdots q_k} \pi_0 \quad \text{for } k = 1, 2, \ldots.
\]

Next, \( \sum_{k=0}^{\infty} \pi_k = 1 \) implies

\[
\pi_0 + \sum_{k=0}^{\infty} \frac{p_0 \cdots p_{k-1}}{q_1 \cdots q_k} \pi_0 = 1.
\]

Hence,

\[
\pi_0 = \left(1 + \sum_{j=1}^{\infty} \frac{p_0 \cdots p_{j-1}}{q_1 \cdots q_j} \right)^{-1}
\] and

\[
\pi_k = \frac{p_0 \cdots p_{k-1}}{q_1 \cdots q_k} \quad \text{for } k = 1, 2, \ldots.
\]
Stationary distribution.

Example (Random walk on weighted graph). Consider a finite simply connected graph $G = (V, E)$ with the weights assigned to all of its edges:

$$W_{x,y} = W_{y,x} > 0$$

for all $x, y \in V$ connected by an edge in $E$.

Denote by $W_x = \sum_{y \in V} W_{x,y}$ the total weight of the edges adjacent to $x \in V$.

Next, consider a random walk $X_n$ on state space $S = V$ evolving according to the following transition probabilities

$$p(x, y) = \frac{W_{x,y}}{W_x} \quad \forall x, y \in V.$$ 

Then, $\pi(x) = W_x/Z_G$ with $Z_G = 2 \sum_{x \in V} W_x$ satisfies the d.b.c.
Stationary distribution.

Example (Knight walk). Here is an example from an unpublished book by Aldous and Fill.

Consider the following random walk: Start with a knight at one of the corner squares of otherwise-empty chessboard. Each step, we move the knight by choosing uniformly from all the possible knight moves. What is the mean number of moves until the knight returns to the starting square?
Harmonic functions.

Suppose \( \{X_n\} \) is a time-homogeneous Markov chain. Then \( h(\cdot) \) is a probability harmonic function if \( h \) satisfies the averaging property

\[
\sum_{y \in S} p(x, y) h(y) = h(x).
\]

Then \( E[h(X_{n+1}) \mid X_n = x] = \sum_y p(x, y) h(y) = h(x) \) and

\[
E[h(X_{n+1}) \mid \mathcal{F}_n] = h(X_n) - h(X_n)
\]

is a martingale.

Equation \( \sum_{y \in S} p(x, y) h(y) = h(x) \) is often written in the operator form as

\[
Ph = h
\]