# Stationary Distributions for Markov Chains

Sinho Chewi

Spring 2017

University of California, Berkeley

# Introduction

# Introduction

**Markov Chain Convergence Theorem**: A finite, irreducible, aperiodic Markov chain converges to its unique stationary distribution.

**Markov Chain Convergence Theorem**: A finite, irreducible, aperiodic Markov chain converges to its unique stationary distribution.
What about infinite chains?

# Introduction

**Markov Chain Convergence Theorem**: A finite, irreducible, aperiodic Markov chain converges to its unique stationary distribution.

What about infinite chains? Complicated.

**Markov Chain Convergence Theorem**: A finite, irreducible, aperiodic Markov chain converges to its unique stationary distribution.

What about infinite chains? Complicated. We need more subtle notions of recurrence.

**Markov Chain Convergence Theorem**: A finite, irreducible, aperiodic Markov chain converges to its unique stationary distribution.

What about infinite chains? Complicated. We need more subtle notions of recurrence. **Not tested!**

# Introduction

**Markov Chain Convergence Theorem**: A finite, irreducible, aperiodic Markov chain converges to its unique stationary distribution.

What about infinite chains? Complicated. We need more subtle notions of recurrence. **Not tested!**

These slides are for your enrichment.

**Markov Chain Convergence Theorem**: A finite, irreducible, aperiodic Markov chain converges to its unique stationary distribution.

What about infinite chains? Complicated. We need more subtle notions of recurrence. **Not tested!**

These slides are for your enrichment. (Also, to help you with the homework.)

# Recurrent and Transient States

$x$ and $y$ are states.

# Recurrent and Transient States

$x$ and $y$ are states.

Notation:

# Recurrent and Transient States

$x$ and $y$ are states.

**Notation**: $\rho_{x,y}$ is the probability of eventually reaching $y$ from $x$.

# Recurrent and Transient States

$x$ and $y$ are states.

**Notation**: $\rho_{x,y}$ is the probability of eventually reaching $y$ from $x$.

$\mathbb{E}_x$ is the expectation starting from state $x$.

$x$ and $y$ are states.

**Notation**: $\rho_{x,y}$ is the probability of eventually reaching $y$ from $x$. $\mathbb{E}_x$ is the expectation starting from state $x$. $N(y)$ is the number of visits to $y$:

# Recurrent and Transient States

$x$ and $y$ are states.

**Notation**: $\rho_{x,y}$ is the probability of eventually reaching $y$ from $x$. $\mathbb{E}_x$ is the expectation starting from state $x$. $N(y)$ is the number of visits to $y$: $\sum_{n=1}^{\infty} 1\{X_n = y\}$.

# Recurrent and Transient States

$x$ and $y$ are states.

**Notation**: $\rho_{x,y}$ is the probability of eventually reaching $y$ from $x$. $\mathbb{E}_x$ is the expectation starting from state $x$. $N(y)$ is the number of visits to $y$: $\sum_{n=1}^{\infty} 1\{X_n = y\}$.

$$x \text{ is recurrent} \iff \rho_{x,x} = 1$$

# Recurrent and Transient States

$x$ and $y$ are states.

**Notation**: $\rho_{x,y}$ is the probability of eventually reaching $y$ from $x$. $\mathbb{E}_x$ is the expectation starting from state $x$. $N(y)$ is the number of visits to $y$: $\sum_{n=1}^{\infty} 1\{X_n = y\}$.

$$x \text{ is recurrent} \iff \rho_{x,x} = 1 \qquad \text{(guaranteed to return)}$$

$x$ and $y$ are states.

**Notation**: $\rho_{x,y}$ is the probability of eventually reaching $y$ from $x$. $\mathbb{E}_x$ is the expectation starting from state $x$. $N(y)$ is the number of visits to $y$: $\sum_{n=1}^{\infty} 1\{X_n = y\}$.

$$x \text{ is recurrent} \iff \rho_{x,x} = 1 \qquad \text{(guaranteed to return)}$$
$$\iff \mathbb{E}_x[N(x)] = \infty$$

# Recurrent and Transient States

$x$ and $y$ are states.

**Notation**: $\rho_{x,y}$ is the probability of eventually reaching $y$ from $x$. $\mathbb{E}_x$ is the expectation starting from state $x$. $N(y)$ is the number of visits to $y$: $\sum_{n=1}^{\infty} 1\{X_n = y\}$.

$$x \text{ is recurrent} \iff \rho_{x,x} = 1 \qquad \text{(guaranteed to return)}$$
$$\iff \mathbb{E}_x[N(x)] = \infty \qquad \text{(visit infinitely often)}$$

# Recurrent and Transient States

$x$ and $y$ are states.

**Notation**: $\rho_{x,y}$ is the probability of eventually reaching $y$ from $x$. $\mathbb{E}_x$ is the expectation starting from state $x$. $N(y)$ is the number of visits to $y$: $\sum_{n=1}^{\infty} 1\{X_n = y\}$.

$$
\begin{aligned}
x \text{ is recurrent } &\iff \rho_{x,x} = 1 && \text{(guaranteed to return)} \\
&\iff \mathbb{E}_x[N(x)] = \infty && \text{(visit infinitely often)}
\end{aligned}
$$

## Recurrent and Transient States

$x$ and $y$ are states.

**Notation**: $\rho_{x,y}$ is the probability of eventually reaching $y$ from $x$. $\mathbb{E}_x$ is the expectation starting from state $x$. $N(y)$ is the number of visits to $y$: $\sum_{n=1}^{\infty} 1\{X_n = y\}$.

$$
\begin{aligned}
x \text{ is recurrent} &\iff \rho_{x,x} = 1 && \text{(guaranteed to return)} \\
&\iff \mathbb{E}_x[N(x)] = \infty && \text{(visit infinitely often)}
\end{aligned}
$$

$$x \text{ is transient} \iff \rho_{x,x} < 1 \iff \mathbb{E}_x[N(x)] < \infty$$

# Recurrent and Transient States

$x$ and $y$ are states.

**Notation**: $\rho_{x,y}$ is the probability of eventually reaching $y$ from $x$. $\mathbb{E}_x$ is the expectation starting from state $x$. $N(y)$ is the number of visits to $y$: $\sum_{n=1}^{\infty} 1\{X_n = y\}$.

$$\begin{aligned} x \text{ is recurrent} &\iff \rho_{x,x} = 1 && \text{(guaranteed to return)} \\ &\iff \mathbb{E}_x[N(x)] = \infty && \text{(visit infinitely often)} \end{aligned}$$

$$x \text{ is transient} \iff \rho_{x,x} < 1 \iff \mathbb{E}_x[N(x)] < \infty$$

In fact, in the transient case,

$$\mathbb{E}_x[N(x)] = \frac{\rho_{x,x}}{1 - \rho_{x,x}}, \qquad \mathbb{E}_x[N(y)] = \frac{\rho_{x,y}}{1 - \rho_{y,y}}.$$

# Recurrent and Transient States

$x$ and $y$ are states.

**Notation**: $\rho_{x,y}$ is the probability of eventually reaching $y$ from $x$. $\mathbb{E}_x$ is the expectation starting from state $x$. $N(y)$ is the number of visits to $y$: $\sum_{n=1}^{\infty} 1\{X_n = y\}$.

$$\begin{aligned} x \text{ is recurrent } &\iff \rho_{x,x} = 1 && \text{(guaranteed to return)} \\ &\iff \mathbb{E}_x[N(x)] = \infty && \text{(visit infinitely often)} \end{aligned}$$

$$x \text{ is transient } \iff \rho_{x,x} < 1 \iff \mathbb{E}_x[N(x)] < \infty$$

In fact, in the transient case,

$$\mathbb{E}_x[N(x)] = \frac{\rho_{x,x}}{1-\rho_{x,x}}, \qquad \mathbb{E}_x[N(y)] = \frac{\rho_{x,y}}{1-\rho_{y,y}}.$$

Observe that $\mathbb{E}_x[N(x)] = \sum_{n=1}^{\infty} \mathbb{P}_x(X_n = x) = \sum_{n=1}^{\infty} P_n(x,x)$.

# Chain Decomposition

# Chain Decomposition

Any Markov chain can be decomposed into communicating classes.

# Chain Decomposition

Any Markov chain can be decomposed into communicating classes. Graph terminology: strongly connected components.

# Chain Decomposition

Any Markov chain can be decomposed into communicating classes. Graph terminology: strongly connected components. Irreducible means only one communicating class.

# Chain Decomposition

Any Markov chain can be decomposed into communicating classes. Graph terminology: strongly connected components.

Irreducible means only one communicating class.

**Recurrence is a class property.**

# Chain Decomposition

Any Markov chain can be decomposed into communicating classes. Graph terminology: strongly connected components.

Irreducible means only one communicating class.

**Recurrence is a class property.** In a communicating class,

# Chain Decomposition

Any Markov chain can be decomposed into communicating classes. Graph terminology: strongly connected components.

Irreducible means only one communicating class.

**Recurrence is a class property.** In a communicating class, all states are transient,

## Chain Decomposition

Any Markov chain can be decomposed into communicating classes. Graph terminology: strongly connected components.

Irreducible means only one communicating class.

**Recurrence is a class property.** In a communicating class, all states are transient, all states are null-recurrent,

# Chain Decomposition

Any Markov chain can be decomposed into communicating classes. Graph terminology: strongly connected components. Irreducible means only one communicating class.

**Recurrence is a class property.** In a communicating class, all states are transient, all states are null-recurrent, or all states are positive-recurrent.

# Chain Decomposition

Any Markov chain can be decomposed into communicating classes. Graph terminology: strongly connected components.

Irreducible means only one communicating class.

**Recurrence is a class property.** In a communicating class, all states are transient, all states are null-recurrent, or all states are positive-recurrent.

We defined “recurrence” for a particular state,

# Chain Decomposition

Any Markov chain can be decomposed into communicating classes. Graph terminology: strongly connected components.

Irreducible means only one communicating class.

**Recurrence is a class property.** In a communicating class, all states are transient, all states are null-recurrent, or all states are positive-recurrent.

We defined “recurrence” for a particular state, but we will refer to an entire irreducible Markov chain as “recurrent”.

# Chain Decomposition

Any Markov chain can be decomposed into communicating classes. Graph terminology: strongly connected components.

Irreducible means only one communicating class.

**Recurrence is a class property.** In a communicating class, all states are transient, all states are null-recurrent, or all states are positive-recurrent.

We defined "recurrence" for a particular state, but we will refer to an entire irreducible Markov chain as "recurrent".

(Aside: The period of a state is a class property.)

# Null-Recurrence and Positive-Recurrence

## Null-Recurrence and Positive-Recurrence

$\mu$ is an stationary measure.

# Null-Recurrence and Positive-Recurrence

$\mu$ is an stationary measure. (“Stationary” because $\mu P = \mu$. “Measure” because the entries might not add to 1.)

# Null-Recurrence and Positive-Recurrence

$\mu$ is an stationary measure. ("Stationary" because $\mu P = \mu$. "Measure" because the entries might not add to 1.)

irreducible + recurrent $\implies$ existence of $\mu$

# Null-Recurrence and Positive-Recurrence

$\mu$ is an stationary measure. ("Stationary" because $\mu P = \mu$.
"Measure" because the entries might not add to 1.)
irreducible $+$ recurrent $\implies$ existence of $\mu$
(and $\mu$ is unique up to scaling)

# Null-Recurrence and Positive-Recurrence

$\mu$ is an stationary measure. ("Stationary" because $\mu P = \mu$. "Measure" because the entries might not add to 1.)

irreducible $+$ recurrent $\implies$ existence of $\mu$

(and $\mu$ is unique up to scaling)

$S$ is the state space.

# Null-Recurrence and Positive-Recurrence

$\mu$ is an stationary measure. ("Stationary" because $\mu P = \mu$.
"Measure" because the entries might not add to 1.)
irreducible $+$ recurrent $\implies$ existence of $\mu$
(and $\mu$ is unique up to scaling)
$S$ is the state space. $T_x$ is the first passage time to $x$.

## Null-Recurrence and Positive-Recurrence

$\mu$ is an stationary measure. ("Stationary" because $\mu P = \mu$. "Measure" because the entries might not add to 1.)

irreducible + recurrent $\implies$ existence of $\mu$

(and $\mu$ is unique up to scaling)

$S$ is the state space. $T_x$ is the first passage time to $x$.

- $\sum_{x \in S} \mu(x) = \infty$.

# Null-Recurrence and Positive-Recurrence

$\mu$ is an stationary measure. ("Stationary" because $\mu P = \mu$. "Measure" because the entries might not add to 1.)

irreducible + recurrent $\implies$ existence of $\mu$

(and $\mu$ is unique up to scaling)

$S$ is the state space. $T_x$ is the first passage time to $x$.

- $\sum_{x \in S} \mu(x) = \infty$. **Null-recurrent case**.

# Null-Recurrence and Positive-Recurrence

$\mu$ is an stationary measure. ("Stationary" because $\mu P = \mu$. "Measure" because the entries might not add to 1.)

irreducible + recurrent $\implies$ existence of $\mu$

(and $\mu$ is unique up to scaling)

$S$ is the state space. $T_x$ is the first passage time to $x$.

- $\sum_{x \in S} \mu(x) = \infty$. **Null-recurrent case**. $\mathbb{E}_x[T_x] = \infty$ for all $x \in S$.
- $\sum_{x \in S} \mu(x) < \infty$.

# Null-Recurrence and Positive-Recurrence

$\mu$ is an stationary measure. ("Stationary" because $\mu P = \mu$. "Measure" because the entries might not add to 1.)

irreducible $+$ recurrent $\implies$ existence of $\mu$

(and $\mu$ is unique up to scaling)

$S$ is the state space. $T_x$ is the first passage time to $x$.

- $\sum_{x \in S} \mu(x) = \infty$. **Null-recurrent case**. $\mathbb{E}_x[T_x] = \infty$ for all $x \in S$.
- $\sum_{x \in S} \mu(x) < \infty$. **Positive-recurrent case**.

# Null-Recurrence and Positive-Recurrence

$\mu$ is an stationary measure. ("Stationary" because $\mu P = \mu$. "Measure" because the entries might not add to 1.)

irreducible + recurrent $\implies$ existence of $\mu$

(and $\mu$ is unique up to scaling)

$S$ is the state space. $T_x$ is the first passage time to $x$.

- $\sum_{x \in S} \mu(x) = \infty$. **Null-recurrent case**. $\mathbb{E}_x[T_x] = \infty$ for all $x \in S$.
- $\sum_{x \in S} \mu(x) < \infty$. **Positive-recurrent case**. (**Theorem**: Positive-recurrent implies there exists a stationary distribution $\pi$.

# Null-Recurrence and Positive-Recurrence

$\mu$ is an stationary measure. (“Stationary” because $\mu P = \mu$. “Measure” because the entries might not add to 1.)

irreducible $+$ recurrent $\implies$ existence of $\mu$

(and $\mu$ is unique up to scaling)

$S$ is the state space. $T_x$ is the first passage time to $x$.

- $\sum_{x \in S} \mu(x) = \infty$. **Null-recurrent case**. $\mathbb{E}_x[T_x] = \infty$ for all $x \in S$.
- $\sum_{x \in S} \mu(x) < \infty$. **Positive-recurrent case**. (**Theorem**: Positive-recurrent implies there exists a stationary distribution $\pi$. Scale $\mu$.)

# Null-Recurrence and Positive-Recurrence

$\mu$ is an stationary measure. ("Stationary" because $\mu P = \mu$. "Measure" because the entries might not add to 1.)
irreducible + recurrent $\implies$ existence of $\mu$
(and $\mu$ is unique up to scaling)
$S$ is the state space. $T_x$ is the first passage time to $x$.

- $\sum_{x \in S} \mu(x) = \infty$. **Null-recurrent case**. $\mathbb{E}_x[T_x] = \infty$ for all $x \in S$.
- $\sum_{x \in S} \mu(x) < \infty$. **Positive-recurrent case**. (**Theorem**: Positive-recurrent implies there exists a stationary distribution $\pi$. Scale $\mu$.) $\mathbb{E}_x[T_x] < \infty$ for all $x \in S$.

# Convergence Theorem

# Convergence Theorem

*Intuition*: Many interesting Markov chains have a regime in which the chain is transient, and a regime in which the chain is positive-recurrent.

# Convergence Theorem

*Intuition*: Many interesting Markov chains have a regime in which the chain is transient, and a regime in which the chain is positive-recurrent. There is a narrow boundary of null-recurrence.

# Convergence Theorem

*Intuition*: Many interesting Markov chains have a regime in which the chain is transient, and a regime in which the chain is positive-recurrent. There is a narrow boundary of null-recurrence.

**Markov Chain Convergence Theorem**: If the chain is irreducible,

# Convergence Theorem

*Intuition*: Many interesting Markov chains have a regime in which the chain is transient, and a regime in which the chain is positive-recurrent. There is a narrow boundary of null-recurrence.

**Markov Chain Convergence Theorem**: If the chain is irreducible, aperiodic,

# Convergence Theorem

*Intuition*: Many interesting Markov chains have a regime in which the chain is transient, and a regime in which the chain is positive-recurrent. There is a narrow boundary of null-recurrence.

Markov Chain Convergence Theorem: If the chain is irreducible, aperiodic, and positive-recurrent,

# Convergence Theorem

*Intuition*: Many interesting Markov chains have a regime in which the chain is transient, and a regime in which the chain is positive-recurrent. There is a narrow boundary of null-recurrence.

**Markov Chain Convergence Theorem**: If the chain is irreducible, aperiodic, and positive-recurrent, then for any initial distribution $\pi_0$, the distribution of $X_n$ will converge to $\pi$.

# Convergence Theorem

*Intuition*: Many interesting Markov chains have a regime in which the chain is transient, and a regime in which the chain is positive-recurrent. There is a narrow boundary of null-recurrence.

**Markov Chain Convergence Theorem**: If the chain is irreducible, aperiodic, and positive-recurrent, then for any initial distribution $\pi_0$, the distribution of $X_n$ will converge to $\pi$.

Note: Finite irreducible chains are always positive-recurrent.

# Convergence Theorem

*Intuition*: Many interesting Markov chains have a regime in which the chain is transient, and a regime in which the chain is positive-recurrent. There is a narrow boundary of null-recurrence.

**Markov Chain Convergence Theorem**: If the chain is irreducible, aperiodic, and positive-recurrent, then for any initial distribution $\pi_0$, the distribution of $X_n$ will converge to $\pi$.

Note: Finite irreducible chains are always positive-recurrent.

Null-recurrent case?

*Intuition*: Many interesting Markov chains have a regime in which the chain is transient, and a regime in which the chain is positive-recurrent. There is a narrow boundary of null-recurrence.

**Markov Chain Convergence Theorem**: If the chain is irreducible, aperiodic, and positive-recurrent, then for any initial distribution $\pi_0$, the distribution of $X_n$ will converge to $\pi$.

Note: Finite irreducible chains are always positive-recurrent.

Null-recurrent case?

$\pi_n$ is the distribution at time $n$.

# Convergence Theorem

*Intuition*: Many interesting Markov chains have a regime in which the chain is transient, and a regime in which the chain is positive-recurrent. There is a narrow boundary of null-recurrence.

**Markov Chain Convergence Theorem**: If the chain is irreducible, aperiodic, and positive-recurrent, then for any initial distribution $\pi_0$, the distribution of $X_n$ will converge to $\pi$.

Note: Finite irreducible chains are always positive-recurrent.

Null-recurrent case?

$\pi_n$ is the distribution at time $n$. If the chain is irreducible and *not* recurrent, then $\pi_n(x) \to 0$ for every $x \in S$.

# Convergence Theorem

*Intuition*: Many interesting Markov chains have a regime in which the chain is transient, and a regime in which the chain is positive-recurrent. There is a narrow boundary of null-recurrence.

**Markov Chain Convergence Theorem**: If the chain is irreducible, aperiodic, and positive-recurrent, then for any initial distribution $\pi_0$, the distribution of $X_n$ will converge to $\pi$.

Note: Finite irreducible chains are always positive-recurrent.

Null-recurrent case?

$\pi_n$ is the distribution at time $n$. If the chain is irreducible and *not* recurrent, then $\pi_n(x) \to 0$ for every $x \in S$.

**Fact**: $\pi$ exists $\iff$ positive-recurrent.

# Fraction of Time Spent in States

# Fraction of Time Spent in States

$N_t(x)$ is the number of visits to $x$ by time $t$.

# Fraction of Time Spent in States

$N_t(x)$ is the number of visits to $x$ by time $t$.
$N_t(x) = \sum_{n=1}^{t} 1\{X_n = x\}$.

# Fraction of Time Spent in States

$N_t(x)$ is the number of visits to $x$ by time $t$.
$N_t(x) = \sum_{n=1}^{t} 1\{X_n = x\}$.
Irreducible and positive-recurrent implies

$$\frac{1}{t} N_t(x) \to \pi(x) \qquad \text{almost surely, as } t \to \infty.$$

# Fraction of Time Spent in States

$N_t(x)$ is the number of visits to $x$ by time $t$.
$N_t(x) = \sum_{n=1}^{t} 1\{X_n = x\}$.
Irreducible and positive-recurrent implies

$$\frac{1}{t} N_t(x) \to \pi(x) \qquad \text{almost surely, as } t \to \infty.$$

Another way to view the stationary distribution:

# Fraction of Time Spent in States

$N_t(x)$ is the number of visits to $x$ by time $t$.
$N_t(x) = \sum_{n=1}^{t} 1\{X_n = x\}$.
Irreducible and positive-recurrent implies

$$\frac{1}{t} N_t(x) \to \pi(x) \qquad \text{almost surely, as } t \to \infty.$$

Another way to view the stationary distribution:

$$\pi(x) = \frac{1}{\mathbb{E}_x[T_x]} \qquad \forall x \in S.$$

# Fraction of Time Spent in States

$N_t(x)$ is the number of visits to $x$ by time $t$.
$N_t(x) = \sum_{n=1}^{t} 1\{X_n = x\}$.
Irreducible and positive-recurrent implies

$$\frac{1}{t} N_t(x) \to \pi(x) \qquad \text{almost surely, as } t \to \infty.$$

Another way to view the stationary distribution:

$$\pi(x) = \frac{1}{\mathbb{E}_x[T_x]} \qquad \forall x \in S.$$

Not immediately obvious!

# Continuous-Time Markov Chains

…are basically the same as discrete-time Markov chains, with regards to these concepts.

# Continuous-Time Markov Chains

…are basically the same as discrete-time Markov chains, with regards to these concepts.
The CTMC is irreducible if and only if the induced DTMC is irreducible.

…are basically the same as discrete-time Markov chains, with regards to these concepts.
The CTMC is irreducible if and only if the induced DTMC is irreducible. Induced DTMC is the CTMC when you ignore the transition time.

...are basically the same as discrete-time Markov chains, with regards to these concepts.

The CTMC is irreducible if and only if the induced DTMC is irreducible. Induced DTMC is the CTMC when you ignore the transition time.

Notable exception:

…are basically the same as discrete-time Markov chains, with regards to these concepts.

The CTMC is irreducible if and only if the induced DTMC is irreducible. Induced DTMC is the CTMC when you ignore the transition time.

Notable exception:

$$N(x) := \lim_{t\to\infty} \frac{1}{t} \int_0^t 1\{X(s) = x\} \, \mathrm{d}s.$$

Analog of $N(x)$ for DTMCs.

...are basically the same as discrete-time Markov chains, with regards to these concepts.

The CTMC is irreducible if and only if the induced DTMC is irreducible. Induced DTMC is the CTMC when you ignore the transition time.

Notable exception:

$$N(x) := \lim_{t\to\infty} \frac{1}{t} \int_0^t 1\{X(s) = x\} \, \mathrm{d}s.$$

Analog of $N(x)$ for DTMCs. Now, one has

$$N(x) = \frac{1}{Q(x,x)\mathbb{E}_x[T_x]} \qquad \text{almost surely},$$

where $Q(x, x)$ is the transition rate out of state $x$ and $T_x$ is the (continuous) time to return to $x$.

# Continuous-Time Markov Chains

...are basically the same as discrete-time Markov chains, with regards to these concepts.

The CTMC is irreducible if and only if the induced DTMC is irreducible. Induced DTMC is the CTMC when you ignore the transition time.

Notable exception:

$$N(x) := \lim_{t\to\infty} \frac{1}{t} \int_0^t 1\{X(s) = x\} \, \mathrm{d}s.$$

Analog of $N(x)$ for DTMCs. Now, one has

$$N(x) = \frac{1}{Q(x,x)\mathbb{E}_x[T_x]} \qquad \text{almost surely,}$$

where $Q(x,x)$ is the transition rate out of state $x$ and $T_x$ is the (continuous) time to return to $x$. Not trivial to calculate $P_t(x,y)$.