# Math 456: Mathematical Modeling

Tuesday, April 3rd, 2018

# Markov Chains:
# More on limit theorems
# (asymptotic frequency of visits)

Tuesday, April 3rd, 2018

# Last time

1. Computing the limit $\lim_{n\to\infty} \mathrm{p}^n(x,y)$ for recurrent $y$ (The "Convergence Theorem") or for transient $y$ (the two formulas for $\mathbb{E}_x[N(y)]$)
2. Decomposition of a Markov chain in closed and irreducible sets plus a transient part.

# Today

1. Review / Practice midterm discussion.
2. Maybe? Proof of the convergence theorem.
3. Asymptotic frequency (or: How to use the stationary distribution to estimate the average amount of time a chain lies in a given state.)

# Proof of the Convergence Theorem

## Proof

An **sketch** of the proof: consider two realizations of the chain, $X_n$ and $Y_n$, and which are independent from one another.

# Proof of the Convergence Theorem

## Proof

An **sketch** of the proof: consider two realizations of the chain, $X_n$ and $Y_n$, and which are independent from one another.

# Proof of the Convergence Theorem

## Proof

An **sketch** of the proof: consider two realizations of the chain, $X_n$ and $Y_n$, and which are independent from one another.

First, note that if $X_0$ and $Y_0$ have the same initial distribution, then the same would be true of $X_n$ and $Y_n$ for all $n \geq 0$.

What if $X_0$ and $Y_0$ do not have the same initial distribution?

# Proof of the Convergence Theorem

## Proof

An **sketch** of the proof: consider two realizations of the chain, $X_n$ and $Y_n$, and which are independent from one another.

First, note that if $X_0$ and $Y_0$ have the same initial distribution, then the same would be true of $X_n$ and $Y_n$ for all $n \geq 0$.

What if $X_0$ and $Y_0$ do not have the same initial distribution?

If at some finite $n_0$ we have $X_{n_0} = Y_{n_0}$, then the subsequent evolution of the chain must have the same probabilities. A moment of reflection would suggest that for the probabilities $\mathrm{P}(X_n = a)$ and $\mathrm{P}(Y_n = a)$ to differ for large $n$, we need to have $\{X_k \neq Y_k \text{ for } k = 1, 2, \ldots, n\}$ has **high probability** – but the aperiodicity and irreducibility will prevent this!

# Proof of the Convergence Theorem

## Proof

Let $X_n, Y_n$ be two independent realizations of the same chain, their initial distributions are TBD.
This chain living on the state space $S \times S$ will be referred to as the **auxiliary chain**.

Then, each of them has the same transition probability matrix $\mathrm{p}(x, y)$. Furthermore,

$$\begin{aligned}
&\mathrm{P}(X_{n+1} = y, Y_{n+1} = w \mid X_n = x, Y_n = z) \\
&= \mathrm{P}(X_{n+1} = y \mid X_n = x)\mathrm{P}(Y_{n+1} = w \mid Y_n = z) \\
&= \mathrm{p}(x, y)\mathrm{p}(z, w)
\end{aligned}$$

# Proof of the Convergence Theorem

## Proof

The way to think about this is that on the state space $S \times S$ we have a chain with transition matrix

$$\mathrm{p}\left((x_1, y_1), (x_2, y_2)\right) = \mathrm{p}(x_1, x_2)\mathrm{p}(y_1, y_2).$$

# Proof of the Convergence Theorem

## Proof

The way to think about this is that on the state space $S \times S$ we have a chain with transition matrix

$$\mathrm{p}\left((x_1, y_1), (x_2, y_2)\right) = \mathrm{p}(x_1, x_2)\mathrm{p}(y_1, y_2).$$

It is not hard to see that

$$\mathrm{p}^n\left((x_1, y_1), (x_2, y_2)\right) = \mathrm{p}^n(x_1, x_2)\mathrm{p}^n(y_1, y_2).$$

Since the original chain is **irreducible and aperiodic**,

$$\mathrm{p}^N(x_1, x_2) > 0, \mathrm{p}^N(y_1, y_2) > 0 \;\; \forall \, x_1, x_2, y_1, y_2$$

# Proof of the Convergence Theorem

## Proof

This means that

$$\mathrm{p}^N\left((x_1, y_1), (x_2, y_2)\right)) > 0$$

so the auxuliary chain in $S \times S$ is irreducible!. In particular, every state in $S \times S$ is recurrent.

Let us show this means that eventually $X_n$ and $Y_n$ must land simultaneously on the same state. That is,

$$\mathrm{P}(X_n = Y_n \text{ for some finite } n) = 1$$

# Proof of the Convergence Theorem

**Proof**

Simply define the set $\Delta \subset S \times S$ by

$$\Delta := \{(x, x) \mid x \in S\}$$

and consider the first hitting time for $\Delta$,

$$T_\Delta = \min\{n \mid (X_n, Y_n) \in \Delta\}$$

Note that, $T_\Delta \leq T_{(x,x)}$ for every $x \in S$.

# Proof of the Convergence Theorem

## Proof

As the chain $(X_n, Y_n)$ is irreducible, $\mathrm{P}(T_{(x,x)} < \infty) = 1$.

Then, since $T_\Delta \leq T_{(x,x)}$ for any $x$, we must have

$$\mathrm{P}(T_\Delta < \infty) = 1$$

Which means that

$$\lim_{n\to\infty} \mathrm{P}(T_\Delta \geq n) = 0.$$

Now, **here is why this matters**: we have that for every $n$, and every state $y$,

$$\mathrm{P}(X_n = y, T_\Delta \leq n) = \mathrm{P}(Y_n = y, T_\Delta \leq n)$$

# Proof of the Convergence Theorem

## Proof

Ok, so, to repeat, for every $n \geq 1$

$$\mathrm{P}(X_n = y, T_\Delta \leq n) = \mathrm{P}(Y_n = y, T_\Delta \leq n)$$

Taking the difference between the distributions for $X_n$ and $Y_n$,

$$\begin{aligned}
&|\mathrm{P}(X_n = y) - \mathrm{P}(Y_n = y)| \\
&= |\mathrm{P}(X_n = y, T_\Delta > n) - \mathrm{P}(Y_n = y, T_\Delta > n)|
\end{aligned}$$

# Proof of the Convergence Theorem

**Proof.**

Adding over all states $y$,

$$\begin{aligned}
&\sum_y |\mathrm{P}(X_n = y) - \mathrm{P}(Y_n = y)| \\
&= \sum_y \mathrm{P}(X_n = y, T_\Delta > n) + \sum_y \mathrm{P}(Y_n = y, T_\Delta > n) \\
&\leq 2\mathrm{P}(T_\Delta > n)
\end{aligned}$$

Then,

$$\lim_{n\to\infty} \sum_y |\mathrm{P}(X_n = y) - \mathrm{P}(Y_n = y)| \leq 2 \lim_{n\to\infty} \mathrm{P}(T_\Delta > n) = 0$$

the last equality being thanks to the irreducibility of the auxiliary chain.

□

# Asymptotic frequency of visits

Fix a chain $X_n$. Recall that $T_y$ denotes the **first arrival time at** $y$ and that

$$N_n(y) = \#\{k \mid 1 \leq k \leq n \text{ and } X_k = y\}$$

which is simply **number of visits to a state** $y$ **up to time** $n$.

**Theorem**

*For an irreducible chain we have the limit*

$$\mathrm{P}\left(\lim_{n\to\infty} \frac{N_n(y)}{n} = \frac{1}{\mathbb{E}_y[T_y]}\right) = 1$$

# Asymptotic frequency of visits

The proof of this theorem will follow from the **law of large numbers**, which we now recall.

Given variables $Y_1, Y_2, \ldots$ which are independent, and identically distributed, we have

$$\mathrm{P}\left(\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} Y_k = \mathbb{E}[Y_1]\right) = 1$$

# Asymptotic frequency of visits

## Proof

Fix $y \in S$, let $T_y^k$ denote the **time of the $k$-th visit to $y$**, and consider the sequence of random variables

$$Y_k := T_y^k - T_y^{k-1}, \;\; k \geq 1, \; Y_1 := T_y^1.$$

# Asymptotic frequency of visits

## Proof

Fix $y \in S$, let $T_y^k$ denote the **time of the $k$-th visit to $y$**, and consider the sequence of random variables

$$Y_k := T_y^k - T_y^{k-1}, \;\; k \geq 1, \; Y_1 := T_y^1.$$

By the Strong Markov Property, the sequence $Y_1, Y_2, \ldots$ is made out of independent, identically distributed random variables.

# Asymptotic frequency of visits

## Proof

Fix $y \in S$, let $T_y^k$ denote the **time of the $k$-th visit to $y$**, and consider the sequence of random variables

$$Y_k := T_y^k - T_y^{k-1}, \;\; k \geq 1, \; Y_1 := T_y^1.$$

By the Strong Markov Property, the sequence $Y_1, Y_2, \dots$ is made out of independent, identically distributed random variables.

Therefore,

$$\mathrm{P}_y \left( \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} Y_k = \mathbb{E}_y[T_y] \right) = 1$$

# Asymptotic frequency of visits

## Proof

Written in terms of $T_y^k$

$$\mathrm{P}_y \left( \lim_{n \to \infty} \frac{T_y^n}{n} = \mathbb{E}_y[T_y] \right) = 1$$

Now, a moment of reflection (and a drawing) shows that

$$T_y^{N_n} \leq n \leq T_y^{N_n+1}$$

for every $n$.

# Asymptotic frequency of visits

## Proof

Written in terms of $T_y^k$

$$\mathrm{P}_y\left(\lim_{n\to\infty}\frac{T_y^n}{n}=\mathbb{E}_y[T_y]\right)=1$$

Now, a moment of reflection (and a drawing) shows that

$$T_y^{N_n}\le n\le T_y^{N_n+1}$$

for every $n$. Dividing all sides by $N_n$, we have

$$\frac{T_y^{N_n}}{N_n}\le\frac{n}{N_n(y)}\le\frac{T_y^{N_n+1}}{N_n+1}\frac{N_n+1}{N_n}$$

# Asymptotic frequency of visits

**Proof.**

Considering that

- $N_n \to \infty$ as $n \to \infty$,
- $\frac{n}{N_n(y}$ lies in between two sequences having the same limit,

we conclude that

$$\mathrm{P}_y \left( \lim_{n\to\infty} \frac{N_n(y)}{n} = \frac{1}{\mathbb{E}_y[T_y]} \right) = 1$$

and the theorem is proved.

$\square$

# Asymptotic frequency of visits

## Putting it all together

If one's goal is to estimate $N_n(y)/n$, then the previous theorem is of no use if we cannot compute $\mathbb{E}_y[T_y]$ for every state $y$.

### Theorem (Durrett, p. 50, Theorem 1.22)

*For an irreducible chain, we have*

$$\mathbb{E}_y[T_y] = \frac{1}{\pi(y)} \ \forall\, y \in S.$$

In **particular**, the stationary distribution encodes what percentage of the time is the chain in each of the states, so that $N_n(y)/n \approx \pi(y)$ for large enough $n$.

# Asymptotic frequency of visits

## Putting it all together

**Proof.**

Take the chain with initial distribution given by $\pi$ itself. Then,

$$\mathrm{P}(X_n = y) = \pi(y) \ \forall\, n, \ \forall\, y \in S.$$

On the other hand $N_n(y)$ is equal to $\sum_{k=1}^{n} \chi_{\{X_k=y\}}$, so

$$\mathbb{E}[N_n(y)] = \sum_{k=1}^{n} \mathrm{P}(X_k = y) \Rightarrow \mathbb{E}[N_n(y)] = \sum_{k=1}^{n} \pi(y) = n\pi(y)$$

Then, previous theorem yields

$$\pi(y) = 1/\mathbb{E}_y[T_y]$$

□

# Examples

For any problem involving **computing the time spent in a given state**, we proceed as follows:

- Verify the chain is irreducible.

# Examples

For any problem involving **computing the time spent in a given state**, we proceed as follows:

- Verify the chain is irreducible.
- Find its stationary distribution.

# Examples

For any problem involving **computing the time spent in a given state**, we proceed as follows:

- Verify the chain is irreducible.
- Find its stationary distribution.
- Use the asymptotic frequency theorem.

**Problem:** Take the chain with transition probability matrix

$$\mathrm{p} = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 1/3 & 0 & 2/3 \\ 0 & 0 & 1/3 & 2/3 \end{pmatrix}$$

Find $\lim \mathrm{p}^n(x, y)$, and estimate how often the chain occupies each state after a large number of steps.

# Examples

## Reflective Random walk

**Solution:**

$$p = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 1/3 & 0 & 2/3 \\ 0 & 0 & 1/3 & 2/3 \end{pmatrix}$$

- Is the chain irreducible?

# Examples

## Reflective Random walk

**Solution:**

$$p = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 1/3 & 0 & 2/3 \\ 0 & 0 & 1/3 & 2/3 \end{pmatrix}$$

- Is the chain irreducible? **Answer:** yes.
- Is the chain aperiodic?

**Solution:**

$$p = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 1/3 & 0 & 2/3 \\ 0 & 0 & 1/3 & 2/3 \end{pmatrix}$$

• Is the chain irreducible? **Answer:** yes.

• Is the chain aperiodic? **Answer:** yes, note that $p(1,1) > 0$, so this state has period 1, which by irreducibility means all states have period 1.

# Examples

## Reflective Random walk

**Solution:**

$$p = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 1/3 & 0 & 2/3 \\ 0 & 0 & 1/3 & 2/3 \end{pmatrix}$$

• By the convergence theorem, and the "ergodic theorem", all we need to do is solve the eigenfunction system to determine $\pi(y)$. Doing so yields the vector

$$\pi^t = (\tfrac{1}{15}, \tfrac{2}{15}, \tfrac{4}{15}, \tfrac{8}{15})$$

**Solution:**

$$
\mathrm{p} = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 1/3 & 0 & 2/3 \\ 0 & 0 & 1/3 & 2/3 \end{pmatrix}
$$

- How often does the chain occupy each state? **Answer:** Since,

$$
\pi^t = (\tfrac{1}{15}, \tfrac{2}{15}, \tfrac{4}{15}, \tfrac{8}{15}),
$$

the system spends about $1/15$ of the time in state $x = 1$, about $2/15$ of the time in state $x = 2$, about $4/15$ of the time in state $x = 4$, and finally, about $8/15$ of the time (which is more than half) in state $x = 4$.

# Next week

## The Metropolis Algorithm

Next class, we will talk about a reverse procedure: we will want to compute a certain distribution $\pi$, and we are going to use a Markov chain **to approximate it by sampling paths from the chain**.

# Next week

## Ergodic Dynamical Systems

*Erdogic*

*ergon* (work) *odos* (path)

The term *ergodic* was introduced by Ludwig Boltzmann, in his attempts at understanding the behavior of molecules in a gas, ultimately founding the field of **statistical mechanics**.

Today, the adjective **ergodic** is used in a dynamical system whenever it has the following property:

*The average of any quantity over a long period of time equals the average of the quantity over the state space*

# Ergodic Dynamical Systems

*The average of any quantity over a long period of time equals the average of the quantity over the state space*

Note, however, the above statement is a big vague: there are many ways of "averaging over the state space".

Heuristically, for this to happen, every trajectory of the system must cover the entire state space, and must do so according to a some distribution –this distribution is the invariant measure.