Lecture 10: Markov chains

Admin:
Reading: Chapter 4 (Gallager Ch. 6)
Homework 3: Due Thursday, Oct. 2 in class
4.2, 4.4, 4.5, 4.9, 4.15, 4.18

Outline:
4. Renewal theory: Limit theorems, Wald’s identity, key renewal theorem, branching processes, regenerative processes, stationary point processes
7. Martingales: Definition, martingale differences, level crossings, stopping times, Azuma’s maximal inequality, sub-martingales, super-martingales, and the martingale convergence theorem

Markov Chains

Important concepts:
- Markov property
- Stationary distribution and equilibration
- Period
- Recurrent & transient states
  - (positive recurrent, null recurrent)
- Hitting times
- Reversible chain

Important problems:
- When is there a stationary distribution?
- When is that distribution unique?
- When does the system converge to a stationary distribution?
- How fast does it converge?

Key concept: Equilibration
- “Markov chains forget their past.”
Regardless of the initial state (or initial distribution), distribution at time \( n \) converges to the same stationary distribution, as \( n \to \infty \).

Example:

**Google**

**Problem:** How to rank webpages?

**Model:** User browses the web,
- clicking links (according to some distribution, e.g., uniform)
- to go from page to page.
- With probability \( \alpha = 0.15 \), start again from a uniformly random webpage.

\[
\text{PageRank(webpage)} = \text{stationary probability of the webpage} = \lim_{t \to \infty} [P(\text{user is at webpage at time } t)]
\]

**Problems:**
- There might not be any stationary distribution.
- There might be multiple stationary distributions.

**Observe:** If \( \pi_1 P = \pi_1 \) and \( \pi_2 P = \pi_2 \), then \( (p \pi_1, r(1-p) \pi_2) P = p \pi_1 + (1-p) \pi_2 \).

\( \Rightarrow \) If there are \( \geq 2 \) stationary distributions, then there are infinitely many.

- System might not converge to a stationary distribution.

**Bad examples:**

1) \[\begin{array}{c}
\circ \xrightarrow{\text{click}} \circ \xrightarrow{\text{click}} \circ \xrightarrow{\text{click}} \circ \ldots
\end{array}\]

No stationary distribution: state diverges to \( \infty \)
No stationary distribution: state diverges to $\infty$

2) \[ P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \]

$(p, 1-p)$ is stationary for any $p$.
The graph is disconnected.

3) \[ P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \]

$(\frac{1}{2}, \frac{1}{2})$ is the unique stationary distribution, but the system does not converge.

It just oscillates with period 2:
\[ X_n = \begin{cases} X_0 & \text{if } n \text{ is even} \\ 1-X_0 & \text{if } n \text{ is odd} \end{cases} \]

4) \[ \begin{array}{c}
\begin{array}{c}
\text{closed chain}
\\
\text{w/ stat dist } \pi_1
\end{array}
\\
\frac{1}{2}
\\
\frac{1}{2}
\\
\frac{1}{2}
\\
\frac{1}{2}
\\
\frac{1}{2}
\\
\frac{1}{2}
\\
\frac{1}{2}
\\
\frac{1}{2}
\\
\frac{1}{2}
\\
\frac{1}{2}
\\
\end{array}
\begin{array}{c}
\text{closed chain}
\\
\text{w/ stat dist } \pi_2
\end{array}
\end{array} \]

- Multiple stationary distributions
- State converges to stationarity, but doesn't forget where it started.

\[ X_0 \in \text{chain 1} \Rightarrow (\pi_1, 0) \]
\[ X_0 \in \text{chain 2} \Rightarrow (0, \pi_2) \]
\[ X_0 = 1 \Rightarrow (\frac{1}{2} \pi_1, \frac{1}{2} \pi_2) \]
\[ X_0 = 2 \Rightarrow (\frac{1}{3} \pi_1, \frac{2}{3} \pi_2) \]

We will argue that these are essentially the only bad examples.

- Any chain can be decomposed.
• An “irreducible”, “aperiodic”, “positive recurrent” Markov chain has a unique stationary distribution to which the system converges.

---

**Concept:** Irreducible Markov chain

**Definition:** Communication classes

Write “$i \rightarrow j$” if $P^n_{i,j} > 0$ for some $n$

(i.e., you can get from $i$ to $j$)

“$i \leftarrow j$” means $i \rightarrow j$ and $j \rightarrow i$

Observe: $i \rightarrow j \rightarrow k \Rightarrow i \rightarrow k$

$i \leftarrow j \leftarrow k \Rightarrow i \leftarrow k$

A chain is irreducible if $i \leftrightarrow j$ for all states $i,j$.

(Basically, this just means the web graph is connected)

(in both directions.

---

**Important concept:** Periodicity

**Definition:** The period of a state $i$ is

$$\text{period}(i) = \gcd\{n \mid P^n_{i,i} > 0\}.$$

“Aperiodic” means the period is 1.

(If it is impossible to get back to $i$, say the period is $\infty$.)

**Proposition:** If $i \leftrightarrow j$, then $\text{period}(i) = \text{period}(j)$.

Hence all states in an irreducible component have the same period.

**Examples:**

- Period 1 (aperiodic)
- Period 2
- Period 3
What does period mean?

1) If \( n \) is not a multiple of \( \text{period}(i) \), \( P_{ii}^n = \emptyset \).
   (By definition, \( \gcd \) (a set) divides every element of the set.)

2) \( \text{period}(i) = d \iff P_{ii}^n > 0 \Rightarrow n \) is a multiple of \( d \)
   and \( d \) is the largest integer with this property
   (that's what \( \gcd \) means: greatest common divisor)

3) Proposition:
   For any state \( i \), \( \text{period}(i) \) is the smallest integer \( d \) satisfying
   \[ \exists \text{ no s.t. } \forall n \geq \text{ no multiples of } d, \]
   \[ P_{ii}^n > 0 \]

Proof by example:

```plaintext
HighlightSubset[set_, subset_, style_] :=
    ReplaceValue[set, _/; MemberQ[subset, #] -> Style[#, style], (1)];

DrawSet[set_, highlight_, subhighlight_] := Module[{s},
    s = HighlightSubset[set, highlight, {Bold, Red, 20}];
    HighlightSubset[s, subhighlight, {Bold, Blue, 15}] ];

ReturnTimes[loops_] := Module[{set, i, m},
    set = (0);
    For[i = 1, i \leq 20, i++,
        m = set; 
        set = Union[set, m + loops];
    Print[DrawSet[set, (m + loops)]];
    ];
]

HighlightSubset[set, subset, style] :=
    ReplaceValue[set, _/; MemberQ[subset, #] -> Style[#, style], (1)];

DrawSet[set, highlight, subhighlight] := Module[{s},
    s = HighlightSubset[set, highlight, {Bold, Red, 20}];
    HighlightSubset[s, subhighlight, {Bold, Blue, 15}] ];

ReturnTimes[loops] := Module[{set, i, m},
    set = (0);
    For[i = 1, i \leq 20, i++,
        m = set; 
        set = Union[set, m + loops];
    Print[DrawSet[set, (m + loops)]];
    ];
]

ReturnTimes[] := Module[{set, i, m},
    set = (0);
    For[i = 1, i \leq 20, i++,
        m = set; 
        set = Union[set, m + loops];
    Print[DrawSet[set, (m + loops)]];
    ];
```
times = \{2, 5\};
GCD @ times
RetrunTimes(times)
1
\{0, 2, 5\}
\{0, 2, 4, 5, 7\}
\{0, 2, 4, 5, 6, 7, 9\}
\{0, 2, 4, 5, 6, 7, 9, 10\}
\{0, 2, 4, 5, 6, 7, 8, 9, 10, 11\}
\{0, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12\}

\{0, 4, 14\}
\{0, 4, 8, 14, 18\}
\{0, 4, 8, 12, 14, 18, 22\}
\{0, 4, 8, 12, 14, 16, 18, 22, 26\}
\{0, 4, 8, 12, 14, 16, 18, 22, 28\}
\{0, 4, 8, 12, 14, 16, 19, 22, 26, 28, 30\}
\{0, 4, 8, 12, 14, 16, 19, 20, 22, 26, 28, 30, 32\}
\{0, 4, 8, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34\}
\{0, 4, 8, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36\}

\{0, 6, 10, 15\}
\{0, 6, 10, 12, 15, 16, 21\}
\{0, 6, 10, 12, 15, 16, 20, 21, 25\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 25, 27\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 25, 27, 30\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 25, 27, 30, 31\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 33\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 33\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 33\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 33\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 33\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 33\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 33\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 33\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 33\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 33\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 33\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 33\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 33\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 33\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 33\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 33\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 33\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 33\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 33\}
\{0, 6, 10, 12, 15, 16, 18, 20, 21, 22, 24, 25, 26, 27, 28, 30, 31, 33\}

Proposition: If $d = \text{period}(i) = \text{gcd}(2 \mid \exists n \mid P_{ii}^n > 0)$, then for all sufficiently large $n$ that are multiples of $d$, $P_{ii}^n > 0$. 

EE 512 Page 6
Proof:
List the elements \( n_1, n_2, n_3, \ldots \) of \( \{n \mid p_i^n > 0\} \).
Note that \( \gcd(n, 1), \gcd(n, n_2), \gcd(n, n_1, n_3), \ldots \)
decreases to reach \( d \). So for some \( k \),
\[ \gcd(n_1, \ldots, n_k) = d. \]

Claim: If \( n > (2k-1) \cdot \text{lcm}(n_1, \ldots, n_k) \) is a multiple of \( d \), then \( p_i^n > 0 \).

Proof: Let \( p = \text{lcm}(n_1, \ldots, n_k) \).

By the Chinese Remainder Theorem, any multiple of \( d \)
can be written
\[ n = a_1 n_1 + \cdots + a_k n_k \]
for integers \( a_1, \ldots, a_k \) (which can be negative).

Now write
\[ n = (\sum_{\delta} a_{\delta} n_{\delta}) + a \cdot p \]
where each \( a_{\delta} \) is \( |a_{\delta}| < p/n_{\delta} \). This is possible because if
\[ |a_{\delta}| > p/n_{\delta}, \] then we can reduce \( a_{\delta} \) by \( p/n_{\delta} \) by taking
\[ a \rightarrow a \pm 1. \] Thus
\[ |\sum_{\delta} a_{\delta} n_{\delta}| < k \cdot p. \]

So far, the argument holds for any \( n \).
If \( n > (k-1)p \), then \( a > 0 \) (because if \( a \leq -1 \), \( ap + \sum_{\delta} a_{\delta} n_{\delta} < (k-1)p \leq n \)).
If \( n > (2k-1)p \), then apply the above decomposition to \( n - kp > (k-1)p \),
yielding
\[ n - kp = \sum_{\delta} (a_{\delta} + b_{\delta}) n_{\delta} + a \cdot p \]
with \( a > 0 \) and \( |a_{\delta}| < p/n_{\delta} \).

\[ \Rightarrow n = \sum_{\delta} (a_{\delta} + b_{\delta}) n_{\delta} + a \cdot p, \]
and now all the coefficients \( a_{\delta} + b_{\delta} \) and \( a \) are \( \geq 0 \)
\[ \Rightarrow p_i^n > \left( \prod_{\delta} p_i^{n_{\delta}} (a_{\delta} + b_{\delta}) \right) p_i^{n \cdot p} > 0 \]

Remark: There are probably better proofs!

Proposition: If \( i \leftrightarrow j \), then \( \text{period}(i) = \text{period}(j) \).

Hence all states in an irreducible component have the same period.
Proof: Let $d_i = \text{period}(i)$, $d_j = \text{period}(j)$.

Choose $n$ large enough so $p_0^n > 0$ and $p_{d_j}^n > 0$.

$\Rightarrow n_1 + n_2 + n$ and $n_1 + n_2 + (n + d_j)$ are return times for $i$.

Since $\gcd(a, b) \leq |b - a|$, $d_i \leq d_j$.

By symmetry, $d_j \leq d_i$, so they are equal. \hfill \Box

**Characterizing the States of a Markov Chain**

**Goal:** Understand long-term behavior of the chain.

**Rough initial idea:** Study "flow" of probability through system.

**Recall:** We argued that if you divide the states in two,

then in a stationary distribution, the amount of probability going left to right must balance the probability going right to left.

Consider the following chain:

Since there are no edges going from right to left, and everything on the left is connected, in the long run all probability will leak into the right half $\Rightarrow \Pr(\text{left half}) = 0$.

**Big picture:**

transient states (probability will go to 0 eventually)
transient states (probability will go to 0 eventually)

absorbing state

absorbing class of states (no outgoing edges => probability can't leak)

We will argue that any Markov chain can be divided this way.

**Key concept:** Transient and recurrent states

starting at y, do you ever return?

- Yes: y is Recurrent
  - Expected time between returns?
    - Finite: Positive recurrent
    - Infinite: Null recurrent
  - maybe/no: y is Transient

**Notation:**

\[ X_0, X_1, X_2, \ldots, X_n = \text{state at time } n \]

\[ T_y = \min\{n > 1 : X_n = y\} \quad \text{hitting time} \]
\( f_{xy} = P_x[T_y < \infty] = P[T_y < \infty \mid X_0 = x] \)

- probability that you ever reach \( y \), starting at \( x \)

\( m_{xy} = E_x[T_y] = E[T_y \mid X_0 = x] \)

- expected # of transitions to reach \( y \)

\[ x \rightarrow y \iff \text{for some } n, \ P^n_{xy} > 0 \]

\[ x \rightarrow y \iff f_{xy} > 0 \]

\( N_y(n) = \# \text{ of visits to } y \text{ by time } n \)

\( N_y = \# \text{ of visits to } y \text{ ever} = N_y(\infty) \)

**Remark:** The sequence \( N_y(0), N_y(1), N_y(2), \ldots \) is a (delayed) renewal process, with interarrival times distributed according to \( T_y \mid X_0 = y \).

**Definition:**

\( f_{yy} \)

- \( y \) is transient if \( P_y[\text{ever return to } y] < 1 \).

- \( y \) is recurrent if \( P_y[\text{ever return to } y] = 1 \), i.e., \( T_y < \infty \)

**Properties of transient and recurrent states:**

a) If \( y \) is transient, then for all \( x \),

\[ P_x[N_y < \infty] = 1, \ E_x[N_y] = \frac{f_{xy}}{1 - f_{yy}}. \]

b) If \( y \) is recurrent, then

\[ P_y[N_y = \infty] = 1 \]

you return to \( y \) infinitely many times!

Also, \( P_x[N_y = \infty] = P_x[T_y < \infty] = f_{xy} \).

**Why?** We just have a geometric distribution.

\[ P_x[N_y = m] = f_{xy} \cdot f_{yy}^{m-1} \cdot (1 - f_{yy}) \]
Definition: A recurrent state \( y \) (meaning \( T_y < \infty \) a.s.) is

- **positive recurrent** if \( \mathbb{E}_y[T_y] < \infty \)
- **null recurrent** if \( \mu_y = \mathbb{E}_y[T_y] = \infty \).

Example: Birth-death chain

![Birth-death chain diagram]

Claim: If \( p > \frac{1}{2} \), then all states are transient.
If \( p = \frac{1}{2} \), all states are null recurrent
\( \mathbb{P}_y[N_y = \infty] = 1 \), \( \mathbb{P}_y[T_y < \infty] = 1 \)
but \( \mathbb{E}_y[T_y] = \infty \)
If \( p < \frac{1}{2} \), all states are positive recurrent.
\( \mathbb{E}_y[T_y] < \infty \).

We will prove this later. The \( p > \frac{1}{2} \) and \( p < \frac{1}{2} \) cases
are very intuitive, since in those cases the state will tend
to drift to the right or left, respectively. The $p = \frac{1}{2}$ case
is trickier.

**Example:** Random walk on a lattice

![Random walk on a lattice diagram]

- equal probabilities to go
to any neighbor

**Fact:** All states are
- null recurrent in dimension $\leq 2$
- transient in dimension $> 3$

Proof for $d = 1$:

$$E[\# \text{ returns}] = \sum_{n=0}^{\infty} \left( \frac{1}{2} \right)^{2n} \cdot (2n)!^2 \cdot \frac{\# \text{ of paths of length } 2n}{\sum_{n=0}^{\infty} \frac{1}{(2n)!^2}}$$

$$\approx \sum_{n=0}^{\infty} \frac{1}{\sqrt{\pi n}}$$

using $(2n)! \approx \sqrt{4\pi n} \left( \frac{2n}{e} \right)^{2n}$

$$= \infty \quad \checkmark$$

Recurrence follows. It is null recurrent because there is
no stationary distribution (as we’ll see in a moment).

The proofs in dimensions 2 and 3 are good exercises.

**Theorem:** If $x$ is recurrent and $x \to y$,
then $y$ is recurrent.

(Furthermore, $f_{xy} = P_x[T_y < \infty] = f_{yx} = 1$.)

**Intuition:** $x$ recurrent means it is revisited infinitely
many times. On each visit to $x$, there is a positive
probability of getting to $y \Rightarrow$ you visit $y$ infinitely
many times. Thus $y$ is recurrent (starting at $x$ or $y$,
you’ll visit $y$ infinitely many times).

If $f_{yx} < 1$, then after reaching $y$ the first time.
there would be probability \( 1 - f_{yx} > 0 \) of never getting back to \( x \). That would contradict \( x \) being recurrent.

A symmetrical argument implies \( f_{xy} = 1 \).

**Formal proof.**

\( x \) is recurrent \( \Rightarrow f_{xx} = 1 \).

\( x \rightarrow y \Rightarrow f_{xy} > 0 \)

\( f_{xy} > 0 \) means that there is a positive probability path from \( x \) to \( y \) that does not go back through \( x \); hence \( \mathbb{P}_x[T_y < T_x] > 0 \).

\[ 1 = f_{xx} = \mathbb{P}_x[T_x < \infty] = \text{prob of returning to } x \]

\[ = \mathbb{P}_x[T_x < \infty | T_y > T_x] \cdot \mathbb{P}_x[T_y > T_x] \]

\[ + \mathbb{P}_x[T_x < \infty | T_y < T_x] \cdot \mathbb{P}_x[T_y < T_x] \]

\[ \leq 1 \cdot (1 - \mathbb{P}_x[T_y < T_x]) + f_{yx} \cdot \mathbb{P}_x[T_y < T_x] \]

\[ = 1 - (1 - f_{yx}) \mathbb{P}_y[T_y < T_x] \]

Since \( \mathbb{P}_x[T_y < T_x] > 0 \), this implies \( f_{yx} = 1 \).

The expected number of visits to \( y \), starting at \( y \), satisfies

\[ E_y[N_y] = \sum_{n=1}^{\infty} \mathbb{P}[X_n = x] \cdot \mathbb{P}[T_y < T_x] \]

The right-hand side counts only those visits to \( y \) that occur directly after a visit to \( x \), without an intervening \( y \). For example, here \( x \rightarrow y \rightarrow y \rightarrow y \) it only counts the first visit to \( y \).)

Thus \( E_y[N_y] = \infty \), so \( y \) cannot be transient and must be recurrent. (If \( y \) were transient, \( E_y[N_y] = 1 - f_{yy} = \infty \).)

Now that we know \( y \) is transient and \( y \rightarrow x \), a symmetrical argument implies \( f_{xy} = 1 \).

**Classes of states**

We've seen: \( x \) recurrent, \( x \rightarrow y \Rightarrow f_{xy} = f_{yx} = 1 \), \( y \) recurrent \( (f_{yy} = 1) \)

Thus, once you find one recurrent state, everything it leads to is also recurrent, in the same class. The \( f \) matrix must look like

\[
\begin{pmatrix}
\cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots \\
\cdots & \cdots & 0 \\
\end{pmatrix}
\]

\( \Rightarrow \) classes of
Recall that in each class of recurrent states, all states have the same period.

**LIMIT THEOREMS**

Recall:

**Elementary renewal theorem:**
For a renewal process, with mean interarrival time $\mu$,

$$\lim_{t \to \infty} \frac{N(t)}{t} = \frac{1}{\mu} \quad \text{with probability 1, and}$$

$$\lim_{t \to \infty} \frac{E[N(t)]}{t} = \frac{1}{\mu}.$$  

**Blackwell's Theorem (lattice case):**
For a renewal process with lattice interarrival times, with period $d$,

$$\lim_{n \to \infty} E[\text{# of renewals at time } nd] = \frac{d}{\mu}.$$  

Applying these theorems to the (delayed) renewal process whose events are entries into $y$, we obtain: