Randomised Algorithms
Lecture 4: Markov Chains and Mixing Times

Thomas Sauerwald (tms41@cam.ac.uk)
Outline

Recap of Markov Chain Basics

Irreducibility, Periodicity and Convergence

Total Variation Distance and Mixing Times

Application 1: Card Shuffling

Application 2: Markov Chain Monte Carlo (non-examin.)

Appendix: Remarks on Mixing Time (non-examin.)
Applications of Markov Chains in Computer Science

Broadcasting

Clustering

Ranking Websites

\[
A = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0
\end{pmatrix}
\]
Applications of Markov Chains in Computer Science

Broadcasting

\[ A = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{bmatrix} \]
Applications of Markov Chains in Computer Science

Broadcasting

Clustering

$A = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{pmatrix}$
Applications of Markov Chains in Computer Science

Broadcasting

Clustering

\[ A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \]
Applications of Markov Chains in Computer Science

- Broadcasting
- Ranking Websites
- Clustering

\[ A = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
\end{bmatrix} \]
Applications of Markov Chains in Computer Science

Broadcasting

Ranking Websites

Clustering

Sampling and Optimisation

A =

\[
\begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 \\
\end{pmatrix}
\]
Applications of Markov Chains in Computer Science

- Broadcasting
- Ranking Websites
- Clustering
- Sampling and Optimisation

\[ A = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} \]
Applications of Markov Chains in Computer Science

- Broadcasting
- Ranking Websites
- Clustering
- Sampling and Optimisation

\[ A = \begin{bmatrix}
0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 & 0
\end{bmatrix} \]
Applications of Markov Chains in Computer Science

- Broadcasting
- Ranking Websites
- Load Balancing
- Clustering
- Sampling and Optimisation

$$A = \begin{pmatrix}
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0
\end{pmatrix}$$
Applications of Markov Chains in Computer Science

- Broadcasting
- Ranking Websites
- Load Balancing
- Clustering
- Sampling and Optimisation
- Particle Processes

A = \begin{pmatrix}
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\
\end{pmatrix}
We say that \((X_t)_{t=0}^{\infty}\) is a Markov Chain on State Space \(\Omega\) with Initial Distribution \(\mu\) and Transition Matrix \(P\) if:

1. For any \(x \in \Omega\), \(P[X_0 = x] = \mu(x)\).
2. The Markov Property holds: for all \(t \geq 0\) and any \(x_0, \ldots, x_{t+1} \in \Omega\),
   \[P[X_{t+1} = x_{t+1} | X_0 = x_0, \ldots, X_t = x_t] = P(x_t, x_{t+1}).\]
We say that \((X_t)_{t=0}^{\infty}\) is a Markov Chain on State Space \(\Omega\) with Initial Distribution \(\mu\) and Transition Matrix \(P\) if:

1. For any \(x \in \Omega\), \(P [ X_0 = x ] = \mu(x)\).
Markov Chains

We say that \((X_t)_{t=0}^\infty\) is a Markov Chain on State Space \(\Omega\) with Initial Distribution \(\mu\) and Transition Matrix \(P\) if:

1. For any \(x \in \Omega\), \(P [ X_0 = x ] = \mu(x)\).
2. The Markov Property holds: for all \(t \geq 0\) and any \(x_0, \ldots, x_{t+1} \in \Omega\),
   \[
P \left[ X_{t+1} = x_{t+1} \mid X_t = x_t, \ldots, X_0 = x_0 \right] = P \left[ X_{t+1} = x_{t+1} \mid X_t = x_t \right] = P(x_t, x_{t+1}).
   \]
Markov Chains (Discrete Time and State, Time Homogeneous)

We say that \((X_t)_{t=0}^\infty\) is a Markov Chain on State Space \(\Omega\) with Initial Distribution \(\mu\) and Transition Matrix \(P\) if:

1. For any \(x \in \Omega\), \(P[X_0 = x] = \mu(x)\).
2. The Markov Property holds: for all \(t \geq 0\) and any \(x_0, \ldots, x_{t+1} \in \Omega\),

\[
P \left[ X_{t+1} = x_{t+1} \mid X_t = x_t, \ldots, X_0 = x_0 \right] = P \left[ X_{t+1} = x_{t+1} \mid X_t = x_t \right] = P(x_t, x_{t+1}).
\]

From the definition one can deduce that (check!)
Markov Chains

We say that \((X_t)_{t=0}^\infty\) is a Markov Chain on State Space \(\Omega\) with Initial Distribution \(\mu\) and Transition Matrix \(P\) if:

1. For any \(x \in \Omega\), \(P[ X_0 = x ] = \mu(x) \).
2. The Markov Property holds: for all \(t \geq 0\) and any \(x_0, \ldots, x_{t+1} \in \Omega\),

\[
P \left[ X_{t+1} = x_{t+1} \mid X_t = x_t, \ldots, X_0 = x_0 \right] = P \left[ X_{t+1} = x_{t+1} \mid X_t = x_t \right] \]
\[
:= P(x_t, x_{t+1}).
\]

From the definition one can deduce that (check!)

- For all \(t, x_0, x_1, \ldots, x_t \in \Omega\),

\[
P \left[ X_t = x_t, X_{t-1} = x_{t-1}, \ldots, X_0 = x_0 \right] \\
= \mu(x_0) \cdot P(x_0, x_1) \cdot \ldots \cdot P(x_{t-2}, x_{t-1}) \cdot P(x_{t-1}, x_t).
\]
We say that $(X_t)_{t=0}^\infty$ is a Markov Chain on State Space $\Omega$ with Initial Distribution $\mu$ and Transition Matrix $P$ if:

1. For any $x \in \Omega$, $P[ X_0 = x ] = \mu(x)$.
2. The Markov Property holds: for all $t \geq 0$ and any $x_0, \ldots, x_{t+1} \in \Omega$,

$$
P \left[ X_{t+1} = x_{t+1} \mid X_t = x_t, \ldots, X_0 = x_0 \right] = P \left[ X_{t+1} = x_{t+1} \mid X_t = x_t \right]$$

$$
:= P(x_t, x_{t+1}).
$$

From the definition one can deduce that (check!)

- For all $t, x_0, x_1, \ldots, x_t \in \Omega$,

$$
P \left[ X_t = x_t, X_{t-1} = x_{t-1}, \ldots, X_0 = x_0 \right]
= \mu(x_0) \cdot P(x_0, x_1) \cdot \ldots \cdot P(x_{t-2}, x_{t-1}) \cdot P(x_{t-1}, x_t).
$$

- For all $0 \leq t_1 < t_2, x \in \Omega$,

$$
P \left[ X_{t_2} = x \right] = \sum_{y \in \Omega} P \left[ X_{t_2} = x \mid X_{t_1} = y \right] \cdot P \left[ X_{t_1} = y \right].
What does a Markov Chain Look Like?

**Example:** the carbohydrate served with lunch in the college cafeteria.

This has transition matrix:

\[
P = \begin{bmatrix}
0 & 1/2 & 1/2 \\
1/4 & 0 & 3/4 \\
3/5 & 2/5 & 0
\end{bmatrix}
\]
The Transition Matrix $P$ of a Markov chain $(\mu, P)$ on $\Omega = \{1, \ldots, n\}$ is given by
The Transition Matrix $P$ of a Markov chain $(\mu, P)$ on $\Omega = \{1, \ldots, n\}$ is given by

$$P = \begin{pmatrix}
P(1, 1) & \ldots & P(1, n) \\
\vdots & \ddots & \vdots \\
P(n, 1) & \ldots & P(n, n)
\end{pmatrix}.$$
The Transition Matrix $P$ of a Markov chain $(\mu, P)$ on $\Omega = \{1, \ldots, n\}$ is given by

$$P = \begin{pmatrix}
P(1, 1) & \ldots & P(1, n) \\
\vdots & \ddots & \vdots \\
P(n, 1) & \ldots & P(n, n)
\end{pmatrix}.$$ 

- $\rho^t = (\rho^t(1), \rho^t(2), \ldots, \rho^t(n))$: state vector at time $t$ (row vector).
The Transition Matrix $P$ of a Markov chain $(\mu, P)$ on $\Omega = \{1, \ldots, n\}$ is given by

$$P = \begin{pmatrix} P(1, 1) & \ldots & P(1, n) \\ \vdots & \ddots & \vdots \\ P(n, 1) & \ldots & P(n, n) \end{pmatrix}.$$ 

- $\rho^t = (\rho^t(1), \rho^t(2), \ldots, \rho^t(n))$: state vector at time $t$ (row vector).
- Multiplying $\rho^t$ by $P$ corresponds to advancing the chain one step:

$$\rho^t(y) = \sum_{x \in \Omega} \rho^{t-1}(x) \cdot P(x, y) \quad \text{and thus} \quad \rho^t = \rho^{t-1} \cdot P.$$
The Transition Matrix $P$ of a Markov chain $(\mu, P)$ on $\Omega = \{1, \ldots, n\}$ is given by

$$P = \begin{pmatrix} P(1,1) & \ldots & P(1,n) \\ \vdots & \ddots & \vdots \\ P(n,1) & \ldots & P(n,n) \end{pmatrix}.$$ 

- $\rho^t = (\rho^t(1), \rho^t(2), \ldots, \rho^t(n))$: state vector at time $t$ (row vector).
- Multiplying $\rho^t$ by $P$ corresponds to advancing the chain one step:

$$\rho^t(y) = \sum_{x \in \Omega} \rho^{t-1}(x) \cdot P(x, y) \quad \text{and thus} \quad \rho^t = \rho^{t-1} \cdot P.$$

- The Markov Property and line above imply that for any $t \geq 0$

$$\rho^t = \rho \cdot P^{t-1} \quad \text{and thus} \quad P^t(x,y) = \mathbf{P} [ X_t = y \mid X_0 = x ].$$
The Transition Matrix $P$ of a Markov chain $(\mu, P)$ on $\Omega = \{1, \ldots n\}$ is given by

$$
P = \begin{pmatrix}
P(1, 1) & \ldots & P(1, n) \\
\vdots & \ddots & \vdots \\
P(n, 1) & \ldots & P(n, n)
\end{pmatrix}.
$$

- $\rho^t = (\rho^t(1), \rho^t(2), \ldots, \rho^t(n))$: state vector at time $t$ (row vector).
- Multiplying $\rho^t$ by $P$ corresponds to advancing the chain one step:

$$
\rho^t(y) = \sum_{x \in \Omega} \rho^{t-1}(x) \cdot P(x, y) \quad \text{and thus} \quad \rho^t = \rho^{t-1} \cdot P.
$$

- The Markov Property and line above imply that for any $t \geq 0$

$$
\rho^t = \rho \cdot P^{t-1} \quad \text{and thus} \quad P^t(x, y) = \mathbb{P}[X_t = y \mid X_0 = x].
$$

Thus $\rho^t(x) = (\mu P^t)(x)$ and so $\rho^t = \mu P^t = (\mu P^t(1), \mu P^t(2), \ldots, \mu P^t(n))$. 
Transition Matrices and Distributions

The Transition Matrix $P$ of a Markov chain $(\mu, P)$ on $\Omega = \{1, \ldots, n\}$ is given by

$$P = \begin{pmatrix}
P(1, 1) & \ldots & P(1, n) \\
\vdots & \ddots & \vdots \\
P(n, 1) & \ldots & P(n, n)
\end{pmatrix}.$$

- $\rho^t = (\rho^t(1), \rho^t(2), \ldots, \rho^t(n))$: state vector at time $t$ (row vector).
- Multiplying $\rho^t$ by $P$ corresponds to advancing the chain one step:
  $$\rho^t(y) = \sum_{x \in \Omega} \rho^{t-1}(x) \cdot P(x, y) \quad \text{and thus} \quad \rho^t = \rho^{t-1} \cdot P.$$
- The Markov Property and line above imply that for any $t \geq 0$
  $$\rho^t = \rho \cdot P^{t-1} \quad \text{and thus} \quad P^t(x, y) = P [ X_t = y \mid X_0 = x ].$$

Thus $\rho^t(x) = (\mu P^t)(x)$ and so $\rho^t = \mu P^t = (\mu P^t(1), \mu P^t(2), \ldots, \mu P^t(n))$.

- Everything boils down to deterministic vector/matrix computations
The Transition Matrix $P$ of a Markov chain $(\mu, P)$ on $\Omega = \{1, \ldots n\}$ is given by

$$P = \begin{pmatrix}
P(1, 1) & \ldots & P(1, n) \\
\vdots & \ddots & \vdots \\
P(n, 1) & \ldots & P(n, n)
\end{pmatrix}.$$ 

- $\rho^t = (\rho^t(1), \rho^t(2), \ldots, \rho^t(n))$: state vector at time $t$ (row vector).
- Multiplying $\rho^t$ by $P$ corresponds to advancing the chain one step:
  $$\rho^t(y) = \sum_{x \in \Omega} \rho^{t-1}(x) \cdot P(x, y) \quad \text{and thus} \quad \rho^t = \rho^{t-1} \cdot P.$$ 

- The Markov Property and line above imply that for any $t \geq 0$
  $$\rho^t = \rho \cdot P^{t-1} \quad \text{and thus} \quad P^t(x, y) = P \left[ X_t = y \mid X_0 = x \right].$$ 
  Thus $\rho^t(x) = (\mu P^t)(x)$ and so $\rho^t = \mu P^t = (\mu P^t(1), \mu P^t(2), \ldots, \mu P^t(n)).$

- Everything boils down to deterministic vector/matrix computations
  \Rightarrow can replace $\rho$ by any (load) vector and view $P$ as a balancing matrix!
Stopping and Hitting Times

A non-negative integer random variable $\tau$ is a stopping time for $(X_t)_{t \geq 0}$ if for every $s \geq 0$ the event $\{\tau = s\}$ depends only on $X_0, \ldots, X_s$. 

Example - College Carbs Stopping times:

✓ “We had rice yesterday”; $\tau := \min\{t \geq 1 : X_t = \text{"rice"}\}$ × “We are having pasta next Thursday”

For two states $x, y \in \Omega$ we call $\text{h}(x, y)$ the hitting time of $y$ from $x$:

$h(x, y) := \mathbb{E}_x[\tau_y] = \mathbb{E}_x[\tau_y | X_0 = x]$ where $\tau_y = \min\{t \geq 1 : X_t = y\}$.

Some distinguish between $\tau_y + y = \min\{t \geq 1 : X_t = y\}$ and $\tau_y = \min\{t \geq 0 : X_t = y\}$.

Hitting times are the solution to a set of linear equations:

$h(x, y) = \text{Markov Prop.} = 1 + \sum_{z \in \Omega \{y\}} P(x, z) \cdot h(z, y) \quad \forall x \neq y \in \Omega$.

A Useful Identity
A non-negative integer random variable $\tau$ is a stopping time for $(X_t)_{t \geq 0}$ if for every $s \geq 0$ the event $\{\tau = s\}$ depends only on $X_0, \ldots, X_s$.

**Example** - College Carbs Stopping times:
- “We had rice yesterday”
Stopping and Hitting Times

A non-negative integer random variable $\tau$ is a stopping time for $(X_t)_{t \geq 0}$ if for every $s \geq 0$ the event $\{\tau = s\}$ depends only on $X_0, \ldots, X_s$.

Example - College Carbs Stopping times:

✓ “We had rice yesterday” $\sim \tau := \min \{t \geq 1 : X_{t-1} = \text{“rice”}\}$
Stopping and Hitting Times

A non-negative integer random variable $\tau$ is a stopping time for $(X_t)_{t\geq 0}$ if for every $s \geq 0$ the event \{\(\tau = s\)\} depends only on $X_0, \ldots, X_s$.

**Example** - College Carbs Stopping times:

✓ “We had rice yesterday” $\leadsto \tau := \min \{t \geq 1: X_{t-1} = \text{“rice”}\}$

✗ “We are having pasta next Thursday”
A non-negative integer random variable $\tau$ is a stopping time for $(X_t)_{t \geq 0}$ if for every $s \geq 0$ the event $\{\tau = s\}$ depends only on $X_0, \ldots, X_s$.

**Example** - College Carbs Stopping times:

- ✓ “We had rice yesterday” $\sim \tau := \min \{t \geq 1 : X_{t-1} = \text{“rice”}\}$
- × “We are having pasta next Thursday”

For two states $x, y \in \Omega$ we call $h(x, y)$ the hitting time of $y$ from $x$:

$$h(x, y) := E_x[\tau_y] = E[\tau_y | X_0 = x] \quad \text{where } \tau_y = \min\{t \geq 1 : X_t = y\}.$$
A non-negative integer random variable $\tau$ is a stopping time for $(X_t)_{t \geq 0}$ if for every $s \geq 0$ the event $\{\tau = s\}$ depends only on $X_0, \ldots, X_s$.

**Example** - College Carbs Stopping times:

✓ “We had rice yesterday” $\leadsto \tau := \min \{t \geq 1 : X_{t-1} = \text{“rice”} \}$

✗ “We are having pasta next Thursday”

For two states $x, y \in \Omega$ we call $h(x, y)$ the hitting time of $y$ from $x$:

$$h(x, y) := E_x[\tau_y] = E[\tau_y \mid X_0 = x] \quad \text{where} \quad \tau_y = \min\{t \geq 1 : X_t = y\}.$$ 

Some distinguish between $\tau_y^+ = \min\{t \geq 1 : X_t = y\}$ and $\tau_y = \min\{t \geq 0 : X_t = y\}$. 

---

**Stopping and Hitting Times**

4. Markov Chains and Mixing Times © T. Sauerwald

Recap of Markov Chain Basics
Stopping and Hitting Times

A non-negative integer random variable $\tau$ is a stopping time for $(X_t)_{t \geq 0}$ if for every $s \geq 0$ the event $\{\tau = s\}$ depends only on $X_0, \ldots, X_s$.

Example - College Carbs Stopping times:

✓ “We had rice yesterday” $\sim \tau := \min \{t \geq 1 : X_{t-1} = \text{“rice”}\}$

✗ “We are having pasta next Thursday”

For two states $x, y \in \Omega$ we call $h(x, y)$ the hitting time of $y$ from $x$:

$$h(x, y) := E_x[\tau_y] = E[\tau_y \mid X_0 = x]$$

where $\tau_y = \min\{t \geq 1 : X_t = y\}$.

A Useful Identity

Hitting times are the solution to a set of linear equations:

$$h(x, y) = 1 + \sum_{z \in \Omega \setminus \{y\}} P(x, z) \cdot h(z, y) \quad \forall x \neq y \in \Omega.$$
Outline

Recap of Markov Chain Basics

Irreducibility, Periodicity and Convergence

Total Variation Distance and Mixing Times

Application 1: Card Shuffling

Application 2: Markov Chain Monte Carlo (non-examin.)

Appendix: Remarks on Mixing Time (non-examin.)
A Markov Chain is irreducible if for every pair of states $x, y \in \Omega$ there is an integer $k \geq 0$ such that $P^k(x, y) > 0$. 
A Markov Chain is **irreducible** if for every pair of states \( x, y \in \Omega \) there is an integer \( k \geq 0 \) such that \( P^k(x, y) > 0 \).
A Markov Chain is **irreducible** if for every pair of states $x, y \in \Omega$ there is an integer $k \geq 0$ such that $P^k(x, y) > 0$.

**Exercise:** Which of the two chains (if any) are irreducible?
Irreducible Markov Chains

A Markov Chain is **irreducible** if for every pair of states \( x, y \in \Omega \) there is an integer \( k \geq 0 \) such that \( P^k(x, y) > 0 \).

![Diagram of two Markov chains]

✓ irreducible

× not irreducible (thus reducible)

**Exercise:** Which of the two chains (if any) are irreducible?
A Markov Chain is **irreducible** if for every pair of states \( x, y \in \Omega \) there is an integer \( k \geq 0 \) such that \( P^k(x, y) > 0 \).

\[ \begin{array}{c}
\text{✓ irreducible} \\
\end{array} \quad \begin{array}{c}
\times \text{ not irreducible (thus reducible)} \\
\end{array} \]

**Finite Hitting Time Theorem**

For any states \( x \) and \( y \) of a **finite irreducible** Markov Chain \( h(x, y) < \infty \).
Stationary Distribution

A probability distribution \( \pi = (\pi(1), \ldots, \pi(n)) \) is the stationary distribution of a Markov Chain if \( \pi P = \pi \) (\( \pi \) is a left eigenvector with eigenvalue 1).
Stationary Distribution

A probability distribution \( \pi = (\pi(1), \ldots, \pi(n)) \) is the stationary distribution of a Markov Chain if \( \pi P = \pi \) (\( \pi \) is a left eigenvector with eigenvalue 1)

College carbs example:

\[
\begin{pmatrix}
\frac{4}{13}, & \frac{4}{13}, & \frac{5}{13}
\end{pmatrix}
\cdot
\begin{pmatrix}
0 & \frac{1}{2} & \frac{1}{2} \\
\frac{1}{4} & 0 & \frac{3}{4} \\
\frac{3}{5} & \frac{2}{5} & 0
\end{pmatrix}
= 
\begin{pmatrix}
\frac{4}{13}, & \frac{4}{13}, & \frac{5}{13}
\end{pmatrix}
\]
Stationary Distribution

A probability distribution \( \pi = (\pi(1), \ldots, \pi(n)) \) is the **stationary distribution** of a Markov Chain if \( \pi P = \pi \) (**\( \pi \) is a left eigenvector with eigenvalue 1)**

College carbs example:

\[
\begin{pmatrix}
\frac{4}{13}, \frac{4}{13}, \frac{5}{13} \\
\pi
\end{pmatrix}
\cdot
\begin{pmatrix}
0 & 1/2 & 1/2 \\
1/4 & 0 & 3/4 \\
3/5 & 2/5 & 0
\end{pmatrix}
=
\begin{pmatrix}
\frac{4}{13}, \frac{4}{13}, \frac{5}{13} \\
\pi
\end{pmatrix}
\]

- A Markov Chain reaches stationary distribution if \( \rho^t = \pi \) for some \( t \).
Stationary Distribution

A probability distribution \( \pi = (\pi(1), \ldots, \pi(n)) \) is the stationary distribution of a Markov Chain if \( \pi P = \pi \) (\( \pi \) is a left eigenvector with eigenvalue 1)

College carbs example:

\[
\left( \frac{4}{13}, \frac{4}{13}, \frac{5}{13} \right) \cdot \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 0 & 3/4 \\ 3/5 & 2/5 & 0 \end{pmatrix} = \left( \frac{4}{13}, \frac{4}{13}, \frac{5}{13} \right)
\]

- A Markov Chain reaches stationary distribution if \( \rho^t = \pi \) for some \( t \).
- If reached, then it persists: If \( \rho^t = \pi \) then \( \rho^{t+k} = \pi \) for all \( k \geq 0 \).
Stationary Distribution

A probability distribution $\pi = (\pi(1), \ldots, \pi(n))$ is the stationary distribution of a Markov Chain if $\pi P = \pi$ ($\pi$ is a left eigenvector with eigenvalue 1).

College carbs example:

\[
\begin{pmatrix}
\frac{4}{13} & \frac{4}{13} & \frac{5}{13} \\
\pi & & \\
\end{pmatrix}
\cdot
\begin{pmatrix}
0 & 1/2 & 1/2 \\
1/4 & 0 & 3/4 \\
3/5 & 2/5 & 0 \\
\end{pmatrix}
= 
\begin{pmatrix}
\frac{4}{13} & \frac{4}{13} & \frac{5}{13} \\
\pi & & \\
\end{pmatrix}
\]

- A Markov Chain reaches stationary distribution if $\rho^t = \pi$ for some $t$.
- If reached, then it persists: If $\rho^t = \pi$ then $\rho^{t+k} = \pi$ for all $k \geq 0$.

Existence and Uniqueness of a Positive Stationary Distribution

Let $P$ be finite, irreducible M.C., then there exists a unique probability distribution $\pi$ on $\Omega$ such that $\pi = \pi P$ and $\pi(x) = 1/h(x, x) > 0$, $\forall x \in \Omega$. 

Periodicity

- A Markov Chain is **aperiodic** if for all \( x \in \Omega \), \( \gcd\{ t \geq 1 : P^t(x, x) > 0 \} = 1 \).
Periodicity

- A Markov Chain is aperiodic if for all $x \in \Omega$, $\gcd\{t \geq 1 : P^t(x, x) > 0\} = 1$.
- Otherwise we say it is periodic.
A Markov Chain is **aperiodic** if for all $x \in \Omega$, $\gcd\{t \geq 1 : P^t(x, x) > 0\} = 1$. Otherwise we say it is **periodic**.

**Question:** Which of the two chains (if any) are aperiodic?
A Markov Chain is **aperiodic** if for all \( x \in \Omega \), \( \gcd\{t \geq 1 : P^t(x, x) > 0\} = 1 \).
Otherwise we say it is **periodic**.

---

**Question:** Which of the two chains (if any) are aperiodic?
Periodicity

- A Markov Chain is aperiodic if for all $x \in \Omega$, $\gcd\{t \geq 1 : P^t(x, x) > 0\} = 1$.
- Otherwise we say it is periodic.

**Question:** Which of the two chains (if any) are aperiodic?
Convergence Theorem

Let $P$ be any finite, irreducible, aperiodic Markov Chain with stationary distribution $\pi$. Then for any $x, y \in \Omega$,

$$\lim_{t \to \infty} P^t(x, y) = \pi(y).$$
Convergence Theorem

Let \( P \) be any finite, irreducible, aperiodic Markov Chain with stationary distribution \( \pi \). Then for any \( x, y \in \Omega \),

\[
\lim_{t \to \infty} P^t(x, y) = \pi(y).
\]
Convergence Theorem

Let $P$ be any finite, irreducible, aperiodic Markov Chain with stationary distribution $\pi$. Then for any $x, y \in \Omega$,

$$\lim_{t \to \infty} P^t(x, y) = \pi(y).$$

- mentioned before: For finite irreducible M.C.'s $\pi$ exists, is unique and

$$\pi(y) = \frac{1}{h(y, y)} > 0.$$
Convergence Theorem

Let $P$ be any finite, irreducible, aperiodic Markov Chain with stationary distribution $\pi$. Then for any $x, y \in \Omega$,

$$\lim_{t \to \infty} P^t(x, y) = \pi(y).$$

- mentioned before: For finite irreducible M.C.'s $\pi$ exists, is unique and

$$\pi(y) = \frac{1}{h(y, y)} > 0.$$  

- We will prove a simpler version of the Convergence Theorem after introducing Spectral Graph Theory.
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P_t(1, x)$.

Step: 0
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$. 

Step: 1
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.
Convergence to Stationarity (Example)

- **Markov Chain:** stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$. 

Step: 3
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$. 

Step: 4
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

\[
\begin{array}{ccccccccccc}
0.246 & 0.205 & 0.117 & 0.044 & 0.010 & 0.001 & 0.000 & 0.001 & 0.010 & 0.044 & 0.117 & 0.205 & 0.246
\end{array}
\]
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

\[
\begin{array}{cccccccc}
0.226 & 0.054 & 0.016 & 0.003 & 0.000 & 0.003 & 0.016 & 0.054 & 0.121 & 0.193 & 0.226 \\
\end{array}
\]

Step: 6
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

Step: 7
**Convergence to Stationarity (Example)**

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$. 

![Diagram showing Markov chain transitions and values at step 8]
Convergence to Stationarity (Example)

- Markov Chain: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$. 

![Diagram](image-url)
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with 1/2 and moves left (or right) w.p. 1/4
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$. 

```
| 0.009 | 0.016 |
| 0.016 | 0.037 |
| 0.037 | 0.074 |
| 0.074 | 0.120 |
| 0.120 | 0.160 |
| 0.160 | 0.176 |
| 0.176 | 0.016 |
```

Step: 10
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

![Diagram showing a Markov chain with values at vertices and transition probabilities.](image-url)
**Convergence to Stationarity (Example)**

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

```
0.161
0.149
0.117
0.078
0.044
0.023
0.016
0.023
0.044
0.078
0.117
0.149
0.161
```

Step: 12
- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

[Diagram showing a Markov chain with probabilities at each vertex at step 13.]
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$. 

Step: 14
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

Step: 15
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

Step: 16
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

Step: 18
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with 1/2 and moves left (or right) w.p. 1/4
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$. 

Step: 19
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

![Diagram showing a Markov chain with vertex values at step 20]
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

Step: 21
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

Step: 22
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

Step: 23
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with 1/2 and moves left (or right) w.p. 1/4
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

Step: 25
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.
Convergence to Stationarity (Example)

- **Markov Chain:** stays put with 1/2 and moves left (or right) w.p. 1/4
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

Step: 28
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

![Diagram of Markov Chain](image-url)
**Convergence to Stationarity (Example)**

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P_t(1, x)$.

![Diagram showing a cycle of states with probabilities]
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with 1/2 and moves left (or right) w.p. 1/4
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

Step: 33
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

Step: 36
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

Step: 37
Convergence to Stationarity (Example)

- Markov Chain: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

Step: 38
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

Step: 40
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.
- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

Step: 43
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$. 

Step: 44
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

Step: 45
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

Step: 46
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

Step: 47
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with $1/2$ and moves left (or right) w.p. $1/4$
- At step $t$ the value at vertex $x \in \{1, 2, \ldots, 12\}$ is $P^t(1, x)$.

Step: 49
Convergence to Stationarity (Example)

- **Markov Chain**: stays put with \(1/2\) and moves left (or right) w.p. \(1/4\)
- At step \(t\) the value at vertex \(x \in \{1, 2, \ldots, 12\}\) is \(P^t(1, x)\).
Outline

Recap of Markov Chain Basics

Irreducibility, Periodicity and Convergence

Total Variation Distance and Mixing Times

Application 1: Card Shuffling

Application 2: Markov Chain Monte Carlo (non-examin.)

Appendix: Remarks on Mixing Time (non-examin.)
How Similar are Two Probability Measures?

Loaded Dice

- You are presented three loaded (unfair) dice $A$, $B$, $C$:

<table>
<thead>
<tr>
<th>$x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P[A = x]$</td>
<td>1/3</td>
<td>1/12</td>
<td>1/12</td>
<td>1/12</td>
<td>1/12</td>
<td>1/3</td>
</tr>
<tr>
<td>$P[B = x]$</td>
<td>1/4</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/4</td>
</tr>
<tr>
<td>$P[C = x]$</td>
<td>1/6</td>
<td>1/6</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>9/24</td>
</tr>
</tbody>
</table>

Question 1: Which dice is the least fair? Most choose $A$. Why?

Question 2: Which dice is the most fair? Dice $B$ and $C$ seem "fairer" than $A" but which is fairest?

Loaded Dice
We need a formal "fairness measure" to compare probability distributions!
How Similar are Two Probability Measures?

**Loaded Dice**

- You are presented three loaded (unfair) dice $A$, $B$, $C$:

<table>
<thead>
<tr>
<th>$x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P[A = x]$</td>
<td>1/3</td>
<td>1/12</td>
<td>1/12</td>
<td>1/12</td>
<td>1/12</td>
<td>1/3</td>
</tr>
<tr>
<td>$P[B = x]$</td>
<td>1/4</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/4</td>
</tr>
<tr>
<td>$P[C = x]$</td>
<td>1/6</td>
<td>1/6</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>9/24</td>
</tr>
</tbody>
</table>

**Question 1:** Which dice is the least *fair*?
How Similar are Two Probability Measures?

**Loaded Dice**

- You are presented three *loaded* (unfair) dice $A$, $B$, $C$:

<table>
<thead>
<tr>
<th>$x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P[A=x]$</td>
<td>$1/3$</td>
<td>$1/12$</td>
<td>$1/12$</td>
<td>$1/12$</td>
<td>$1/12$</td>
<td>$1/3$</td>
</tr>
<tr>
<td>$P[B=x]$</td>
<td>$1/4$</td>
<td>$1/8$</td>
<td>$1/8$</td>
<td>$1/8$</td>
<td>$1/8$</td>
<td>$1/4$</td>
</tr>
<tr>
<td>$P[C=x]$</td>
<td>$1/6$</td>
<td>$1/6$</td>
<td>$1/8$</td>
<td>$1/8$</td>
<td>$1/8$</td>
<td>$9/24$</td>
</tr>
</tbody>
</table>

**Question 1:** Which dice is the least *fair*?

**Question 2:** Which dice is the most *fair*?
How Similar are Two Probability Measures?

**Loaded Dice**

- You are presented three **loaded** (unfair) dice $A$, $B$, $C$:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P[A = x]$</td>
<td>1/3</td>
<td>1/12</td>
<td>1/12</td>
<td>1/12</td>
<td>1/12</td>
<td>1/3</td>
</tr>
<tr>
<td>$P[B = x]$</td>
<td>1/4</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/4</td>
</tr>
<tr>
<td>$P[C = x]$</td>
<td>1/6</td>
<td>1/6</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>9/24</td>
</tr>
</tbody>
</table>

**Question 1:** Which dice is the least **fair**?

**Question 2:** Which dice is the most **fair**?

---

4. Markov Chains and Mixing Times © T. Sauerwald  
Total Variation Distance and Mixing Times 15
How Similar are Two Probability Measures?

**Loaded Dice**

- You are presented three loaded (unfair) dice $A$, $B$, $C$:

<table>
<thead>
<tr>
<th>$x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbb{P}[A = x]$</td>
<td>1/3</td>
<td>1/12</td>
<td>1/12</td>
<td>1/12</td>
<td>1/12</td>
<td>1/3</td>
</tr>
<tr>
<td>$\mathbb{P}[B = x]$</td>
<td>1/4</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/4</td>
</tr>
<tr>
<td>$\mathbb{P}[C = x]$</td>
<td>1/6</td>
<td>1/6</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>9/24</td>
</tr>
</tbody>
</table>

**Question 1:** Which dice is the least fair?

**Question 2:** Which dice is the most fair?
How Similar are Two Probability Measures?

**Loaded Dice**

- You are presented three loaded (unfair) dice $A, B, C$:

<table>
<thead>
<tr>
<th>$x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P[A = x]$</td>
<td>$1/3$</td>
<td>$1/12$</td>
<td>$1/12$</td>
<td>$1/12$</td>
<td>$1/12$</td>
<td>$1/3$</td>
</tr>
<tr>
<td>$P[B = x]$</td>
<td>$1/4$</td>
<td>$1/8$</td>
<td>$1/8$</td>
<td>$1/8$</td>
<td>$1/8$</td>
<td>$1/4$</td>
</tr>
<tr>
<td>$P[C = x]$</td>
<td>$1/6$</td>
<td>$1/6$</td>
<td>$1/8$</td>
<td>$1/8$</td>
<td>$1/8$</td>
<td>$9/24$</td>
</tr>
</tbody>
</table>

**Question 1:** Which dice is the least fair?

**Question 2:** Which dice is the most fair?
How Similar are Two Probability Measures?

Loaded Dice

- You are presented three loaded (unfair) dice $A$, $B$, $C$:

<table>
<thead>
<tr>
<th>$x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P[A=x]$</td>
<td>1/3</td>
<td>1/12</td>
<td>1/12</td>
<td>1/12</td>
<td>1/12</td>
<td>1/3</td>
</tr>
<tr>
<td>$P[B=x]$</td>
<td>1/4</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/4</td>
</tr>
<tr>
<td>$P[C=x]$</td>
<td>1/6</td>
<td>1/6</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>9/24</td>
</tr>
</tbody>
</table>

**Question 1:** Which dice is the least fair?

**Question 2:** Which dice is the most fair?
How Similar are Two Probability Measures?

**Loaded Dice**

- You are presented three loaded (unfair) dice $A$, $B$, $C$:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P[A = x]$</td>
<td>1/3</td>
<td>1/12</td>
<td>1/12</td>
<td>1/12</td>
<td>1/12</td>
<td>1/3</td>
</tr>
<tr>
<td>$P[B = x]$</td>
<td>1/4</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/4</td>
</tr>
<tr>
<td>$P[C = x]$</td>
<td>1/6</td>
<td>1/6</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>9/24</td>
</tr>
</tbody>
</table>

**Question 1:** Which dice is the least fair? Most choose $A$.

**Question 2:** Which dice is the most fair?
How Similar are Two Probability Measures?

### Loaded Dice

- You are presented three loaded (unfair) dice $A$, $B$, $C$:

<table>
<thead>
<tr>
<th>$x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P[A = x]$</td>
<td>1/3</td>
<td>1/12</td>
<td>1/12</td>
<td>1/12</td>
<td>1/12</td>
<td>1/3</td>
</tr>
<tr>
<td>$P[B = x]$</td>
<td>1/4</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/4</td>
</tr>
<tr>
<td>$P[C = x]$</td>
<td>1/6</td>
<td>1/6</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>9/24</td>
</tr>
</tbody>
</table>

#### Question 1: Which dice is the least fair? Most choose $A$. Why?

#### Question 2: Which dice is the most fair?
How Similar are Two Probability Measures?

You are presented three loaded (unfair) dice $A$, $B$, $C$:

<table>
<thead>
<tr>
<th>$x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P[A=x]$</td>
<td>1/3</td>
<td>1/12</td>
<td>1/12</td>
<td>1/12</td>
<td>1/12</td>
<td>1/3</td>
</tr>
<tr>
<td>$P[B=x]$</td>
<td>1/4</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/4</td>
</tr>
<tr>
<td>$P[C=x]$</td>
<td>1/6</td>
<td>1/6</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>9/24</td>
</tr>
</tbody>
</table>

**Question 1:** Which dice is the least fair? Most choose $A$. Why?

**Question 2:** Which dice is the most fair? Dice $B$ and $C$ seem “fairer” than $A$ but which is fairest?

---

4. Markov Chains and Mixing Times © T. Sauerwald

Total Variation Distance and Mixing Times
How Similar are Two Probability Measures?

Loaded Dice

- You are presented three loaded (unfair) dice $A$, $B$, $C$:

<table>
<thead>
<tr>
<th>$x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P[A = x]$</td>
<td>1/3</td>
<td>1/12</td>
<td>1/12</td>
<td>1/12</td>
<td>1/12</td>
<td>1/3</td>
</tr>
<tr>
<td>$P[B = x]$</td>
<td>1/4</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>1/4</td>
</tr>
<tr>
<td>$P[C = x]$</td>
<td>1/6</td>
<td>1/6</td>
<td>1/8</td>
<td>1/8</td>
<td>1/8</td>
<td>9/24</td>
</tr>
</tbody>
</table>

**Question 1:** Which dice is the least fair? Most choose $A$. Why?

**Question 2:** Which dice is the most fair? Dice $B$ and $C$ seem “fairer” than $A$ but which is fairest?

We need a formal “fairness measure” to compare probability distributions!
The Total Variation Distance between two probability distributions $\mu$ and $\eta$ on a countable state space $\Omega$ is given by

$$\|\mu - \eta\|_{tv} = \frac{1}{2} \sum_{\omega \in \Omega} |\mu(\omega) - \eta(\omega)|.$$
Total Variation Distance

The Total Variation Distance between two probability distributions $\mu$ and $\eta$ on a countable state space $\Omega$ is given by

$$\|\mu - \eta\|_{tv} = \frac{1}{2} \sum_{\omega \in \Omega} |\mu(\omega) - \eta(\omega)|.$$ 

Loaded Dice: let $D = Unif\{1, 2, 3, 4, 5, 6\}$ be the law of a fair dice:

$$\|D - A\|_{tv} = \frac{1}{2} \left( 2 \left| \frac{1}{6} - \frac{1}{3} \right| + 4 \left| \frac{1}{6} - \frac{1}{12} \right| \right) = \frac{1}{3}$$
The **Total Variation Distance** between two probability distributions $\mu$ and $\eta$ on a countable state space $\Omega$ is given by

$$\|\mu - \eta\|_{tv} = \frac{1}{2} \sum_{\omega \in \Omega} |\mu(\omega) - \eta(\omega)|.$$ 

**Loaded Dice:** let $D = \text{Unif}\{1, 2, 3, 4, 5, 6\}$ be the law of a fair dice:

$$\|D - A\|_{tv} = \frac{1}{2} \left( 2 \left| \frac{1}{6} - \frac{1}{3} \right| + 4 \left| \frac{1}{6} - \frac{1}{12} \right| \right) = \frac{1}{3}$$

$$\|D - B\|_{tv} = \frac{1}{2} \left( 2 \left| \frac{1}{6} - \frac{1}{4} \right| + 4 \left| \frac{1}{6} - \frac{1}{8} \right| \right) = \frac{1}{6}$$

$$\|D - C\|_{tv} = \frac{1}{2} \left( 3 \left| \frac{1}{6} - \frac{1}{8} \right| + 4 \left| \frac{1}{6} - \frac{9}{24} \right| \right) = \frac{1}{6}.$$ 

Thus $\|D - B\|_{tv} = \|D - C\|_{tv}$ and $\|D - B\|_{tv}, \|D - C\|_{tv} < \|D - A\|_{tv}$. So $A$ is the least “fair”, however $B$ and $C$ are equally “fair” (in TV distance).
The Total Variation Distance between two probability distributions $\mu$ and $\eta$ on a countable state space $\Omega$ is given by

$$\|\mu - \eta\|_{tv} = \frac{1}{2} \sum_{\omega \in \Omega} |\mu(\omega) - \eta(\omega)|.$$ 

Loaded Dice: let $D = Unif\{1, 2, 3, 4, 5, 6\}$ be the law of a fair dice:

$$\|D - A\|_{tv} = \frac{1}{2} \left( 2 \left| \frac{1}{6} - \frac{1}{3} \right| + 4 \left| \frac{1}{6} - \frac{1}{12} \right| \right) = \frac{1}{3}$$

$$\|D - B\|_{tv} = \frac{1}{2} \left( 2 \left| \frac{1}{6} - \frac{1}{4} \right| + 4 \left| \frac{1}{6} - \frac{1}{8} \right| \right) = \frac{1}{6}$$

$$\|D - C\|_{tv} = \frac{1}{2} \left( 3 \left| \frac{1}{6} - \frac{1}{8} \right| + \left| \frac{1}{6} - \frac{9}{24} \right| \right) = \frac{1}{6}.$$ 

Thus

$$\|D - B\|_{tv} = \|D - C\|_{tv} \quad \text{and} \quad \|D - B\|_{tv}, \|D - C\|_{tv} < \|D - A\|_{tv}.$$ 

So $A$ is the least “fair”, however $B$ and $C$ are equally “fair” (in TV distance).
Let $P$ be a finite Markov Chain with stationary distribution $\pi$. 

Exercise 4/5.5

For any $\mu$, $\|P^t \mu - \pi\|_{TV} \leq \max_{x \in \Omega} \|P^t x - \pi\|_{TV}$.

For any finite, irreducible, aperiodic Markov Chain

$$\lim_{t \to \infty} \max_{x \in \Omega} \|P^t x - \pi\|_{TV} = 0.$$ 

Convergence Theorem (Implication for TV Distance)

We will see a similar result later after introducing spectral techniques (Lecture 12)!
Let $P$ be a finite Markov Chain with stationary distribution $\pi$.

- Let $\mu$ be a prob. vector on $\Omega$ (might be just one vertex) and $t \geq 0$. Then

$$P_\mu^t := P \left[ X_t = \cdot \mid X_0 \sim \mu \right],$$

is a probability measure on $\Omega$. 

[Exercise 4/5.5] For any $\mu$,

$$\|P_\mu^t \mu - \pi\|_{tv} \leq \max_{x \in \Omega} \|P_\mu^t x - \pi\|_{tv}.$$
Let $P$ be a finite Markov Chain with stationary distribution $\pi$.

- Let $\mu$ be a prob. vector on $\Omega$ (might be just one vertex) and $t \geq 0$. Then

$$P^t_{\mu} := P \left[ X_t = \cdot \mid X_0 \sim \mu \right],$$

is a probability measure on $\Omega$.

- [Exercise 4/5.5] For any $\mu$,

$$\left\| P^t_{\mu} - \pi \right\|_{tv} \leq \max_{x \in \Omega} \left\| P^t_x - \pi \right\|_{tv}.$$
Let $P$ be a finite Markov Chain with stationary distribution $\pi$.

- Let $\mu$ be a prob. vector on $\Omega$ (might be just one vertex) and $t \geq 0$. Then

$$P^t_\mu := P \left[ X_t = \cdot \mid X_0 \sim \mu \right],$$

is a probability measure on $\Omega$.

- [Exercise 4/5.5] For any $\mu$,

$$\left\| P^t_\mu - \pi \right\|_{tv} \leq \max_{x \in \Omega} \left\| P^t_x - \pi \right\|_{tv}.$$

---

**Convergence Theorem (Implication for TV Distance)**

For any finite, irreducible, aperiodic Markov Chain

$$\lim_{t \to \infty} \max_{x \in \Omega} \left\| P^t_x - \pi \right\|_{tv} = 0.$$
Let $P$ be a finite Markov Chain with stationary distribution $\pi$.

- Let $\mu$ be a prob. vector on $\Omega$ (might be just one vertex) and $t \geq 0$. Then

$$P^t_{\mu} := P \left[ X_t = \cdot \mid X_0 \sim \mu \right],$$

is a probability measure on $\Omega$.

- [Exercise 4/5.5] For any $\mu$,

$$\left\| P^t_{\mu} - \pi \right\|_{tv} \leq \max_{x \in \Omega} \left\| P^t_x - \pi \right\|_{tv}.$$

**Convergence Theorem (Implication for TV Distance)**

For any finite, irreducible, aperiodic Markov Chain

$$\lim_{t \to \infty} \max_{x \in \Omega} \left\| P^t_x - \pi \right\|_{tv} = 0.$$

We will see a similar result later after introducing spectral techniques (Lecture 12)!
Mixing Time of a Markov Chain

Convergence Theorem: “Nice” Markov Chains converge to stationarity.
Mixing Time of a Markov Chain

Convergence Theorem: “Nice” Markov Chains converge to stationarity.

Question: How fast do they converge?
Mixing Time of a Markov Chain

Convergence Theorem: “Nice” Markov Chains converge to stationarity.

Question: How fast do they converge?

Mixing Time

The mixing time $\tau_x(\epsilon)$ of a finite Markov Chain $P$ with stationary distribution $\pi$ is defined as

$$\tau_x(\epsilon) = \min \left\{ t \geq 0 : \left\| P_x^t - \pi \right\|_{tv} \leq \epsilon \right\},$$
Mixing Time of a Markov Chain

Convergence Theorem: “Nice” Markov Chains converge to stationarity.

Question: How fast do they converge?

The mixing time \( \tau_x(\epsilon) \) of a finite Markov Chain \( P \) with stationary distribution \( \pi \) is defined as

\[
\tau_x(\epsilon) = \min \left\{ t \geq 0 : \left\| P_x^t - \pi \right\|_{tv} \leq \epsilon \right\},
\]

and,

\[
\tau(\epsilon) = \max_x \tau_x(\epsilon).
\]
Convergence Theorem: “Nice” Markov Chains converge to stationarity.

Question: How fast do they converge?

The mixing time $\tau_{x}(\epsilon)$ of a finite Markov Chain $P$ with stationary distribution $\pi$ is defined as

$$\tau_{x}(\epsilon) = \min \left\{ t \geq 0 : \left\| P_{x}^{t} \pi - \pi \right\|_{tv} \leq \epsilon \right\},$$

and,

$$\tau(\epsilon) = \max_{x} \tau_{x}(\epsilon).$$

- This is how long we need to wait until we are “$\epsilon$-close” to stationarity
Convergence Theorem: “Nice” Markov Chains converge to stationarity.

Question: How fast do they converge?

Mixing Time

The mixing time $\tau_x(\epsilon)$ of a finite Markov Chain $P$ with stationary distribution $\pi$ is defined as

$$
\tau_x(\epsilon) = \min \left\{ t \geq 0 : \left\| P^t_x - \pi \right\|_{tv} \leq \epsilon \right\},
$$

and,

$$
\tau(\epsilon) = \max_x \tau_x(\epsilon).
$$

- This is how long we need to wait until we are “$\epsilon$-close” to stationarity
- We often take $\epsilon = 1/4$, indeed let $t_{mix} := \tau(1/4)$
Mixing Time of a Markov Chain

Convergence Theorem: “Nice” Markov Chains converge to stationarity.

Question: How fast do they converge?

Mixing Time

The mixing time \( \tau_x(\epsilon) \) of a finite Markov Chain \( P \) with stationary distribution \( \pi \) is defined as

\[
\tau_x(\epsilon) = \min \left\{ t \geq 0 : \| P^t_x - \pi \|_{tv} \leq \epsilon \right\},
\]

and,

\[
\tau(\epsilon) = \max_x \tau_x(\epsilon).
\]

- This is how long we need to wait until we are \( \epsilon \)-close to stationarity
- We often take \( \epsilon = 1/4 \), indeed let \( t_{mix} := \tau(1/4) \)

See final slides for some comments on why we choose 1/4.
Outline

Recap of Markov Chain Basics

Irreducibility, Periodicity and Convergence

Total Variation Distance and Mixing Times

Application 1: Card Shuffling

Application 2: Markov Chain Monte Carlo (non-examin.)

Appendix: Remarks on Mixing Time (non-examin.)
Experiment Gone Wrong...

Thanks to Krzysztof Onak (pointer) and Eric Price (graph)

Source: Slides by Ronitt Rubinfeld
What is Card Shuffling?

How long does it take to shuffle a deck of 52 cards?

Source: wikipedia
What is Card Shuffling?

How long does it take to shuffle a deck of 52 cards?

Persi Diaconis (Professor of Statistics and former Magician)

Source: www.soundcloud.com
What is Card Shuffling?

How long does it take to shuffle a deck of 52 cards?

One of the leading experts in the field who has related card shuffling to many other mathematical problems.

Persi Diaconis (Professor of Statistics and former Magician)

Source: www.soundcloud.com
What is Card Shuffling?

Source: wikipedia

Here we will focus on one shuffling scheme which is easy to analyse.

How long does it take to shuffle a deck of 52 cards?

One of the leading experts in the field who has related card shuffling to many other mathematical problems.

Persi Diaconis (Professor of Statistics and former Magician)

Source: www.soundcloud.com

4. Markov Chains and Mixing Times © T. Sauerwald
What is Card Shuffling?

Here we will focus on one **shuffling scheme** which is easy to analyse.

How long does it take to **shuffle a deck of 52 cards**?

How quickly do we converge to the **uniform distribution** over all $n!$ permutations?

One of the leading experts in the field who has related card shuffling to many other mathematical problems.

Persi Diaconis (Professor of Statistics and former Magician)
The Card Shuffling Markov Chain

**TOPTORANDOMSHUFFLE** (Input: A pile of $n$ cards)

1: **For** $t = 1, 2, \ldots$
2: Pick $i \in \{1, 2, \ldots, n\}$ uniformly at random
3: Take the top card and insert it behind the $i$-th card
The Card Shuffling Markov Chain

**The Algorithm:**

**TOPTORANDOMSHUFFLE** (Input: A pile of \( n \) cards)

1. For \( t = 1, 2, \ldots \)
2. Pick \( i \in \{1, 2, \ldots, n\} \) uniformly at random
3. Take the top card and insert it behind the \( i \)-th card

This is a slightly informal definition, so let us look at a small example...
The Card Shuffling Markov Chain

**TOPTORANDOMSHUFFLE** (Input: A pile of $n$ cards)

1. **For** $t = 1, 2, \ldots$
2. Pick $i \in \{1, 2, \ldots, n\}$ uniformly at random
3. Take the top card and insert it behind the $i$-th card

This is a slightly informal definition, so let us look at a small example...

We will focus on this “small” set of cards ($n = 8$)
Even if we know which set of cards come after 8, every permutation is equally likely! The deck of cards is perfectly mixed after the last card "8" reaches the top and is inserted to a random position!
Even if we know which set of cards come after 8, every permutation is equally likely!

The deck of cards is perfectly mixed after the last card "8" reaches the top and is inserted to a random position!
Even if we know which set of cards come after 8, every permutation is equally likely!

The deck of cards is perfectly mixed after the last card "8" reaches the top and is inserted to a random position!
Even if we know which set of cards come after 8, every permutation is equally likely!

the deck of cards is perfectly mixed after the last card "8" reaches the top and is inserted to a random position!
Even if we know which set of cards come after 8, every permutation is equally likely!
Even if we know which set of cards come after 8, every permutation is equally likely!
Even if we know which set of cards come after 8, every permutation is equally likely!
Even if we know which set of cards come after 8, every permutation is equally likely!

→ the deck of cards is perfectly mixed after the last card “8” reaches the top and is inserted to a random position!
Analysing the Mixing Time (Intuition)

Deck of cards is perfectly mixed after the last card “8” reaches the top and is inserted to a random position!
How long does it take for the last card “n” to become top card?

A deck of cards is perfectly mixed after the last card “8” reaches the top and is inserted to a random position!
Analysing the Mixing Time (Intuition)

A deck of cards is perfectly mixed after the last card “8” reaches the top and is inserted to a random position!

- How long does it take for the last card “n” to become top card?
- At the last position, card “n” moves up with probability $\frac{1}{n}$ at each step.
Analysing the Mixing Time (Intuition)

A deck of cards is perfectly mixed after the last card “8” reaches the top and is inserted to a random position!

- How long does it take for the last card “n” to become top card?
- At the last position, card “n” moves up with probability $\frac{1}{n}$ at each step
- At the second last position, card “n” moves up with probability $\frac{2}{n}$
How long does it take for the last card “n” to become top card?

At the last position, card “n” moves up with probability $\frac{1}{n}$ at each step.

At the second last position, card “n” moves up with probability $\frac{2}{n}$.

...
How long does it take for the last card “n” to become top card?
- At the last position, card “n” moves up with probability $\frac{1}{n}$ at each step
- At the second last position, card “n” moves up with probability $\frac{2}{n}$
  
  
  
  :  
- At the second position, card “n” moves up with probability $\frac{n-1}{n}$
How long does it take for the last card “n” to become top card?

- At the last position, card “n” moves up with probability \( \frac{1}{n} \) at each step.
- At the second last position, card “n” moves up with probability \( \frac{2}{n} \).
- \[ \vdots \]
- At the second position, card “n” moves up with probability \( \frac{n-1}{n} \).
- One final step to randomise card “n”.

\( \sim \) deck of cards is perfectly mixed after the last card “8” reaches the top and is inserted to a random position!
How long does it take for the last card “n” to become top card?

- At the last position, card “n” moves up with probability $\frac{1}{n}$ at each step.
- At the second last position, card “n” moves up with probability $\frac{2}{n}$
- ...
- At the second position, card “n” moves up with probability $\frac{n-1}{n}$
- One final step to randomise card “n” (with probability 1)
Analysing the Mixing Time (Intuition)

- How long does it take for the last card “n” to become top card?
- At the last position, card “n” moves up with probability \(\frac{1}{n}\) at each step
- At the second last position, card “n” moves up with probability \(\frac{2}{n}\)
  ...
- At the second position, card “n” moves up with probability \(\frac{n-1}{n}\)
- One final step to randomise card “n” (with probability 1)

This is a “reversed” coupon collector process with \(n\) cards, which takes \(n \log n\) in expectation.
Analysing the Mixing Time (Intuition)

How long does it take for the last card “n” to become top card?

- At the last position, card “n” moves up with probability $\frac{1}{n}$ at each step.
- At the second last position, card “n” moves up with probability $\frac{2}{n}$
  
  
  
  
  
  
  
  
  
  
  
  
  

- At the second position, card “n” moves up with probability $\frac{n-1}{n}$.
- One final step to randomise card “n” (with probability 1)

This is a “reversed” coupon collector process with n cards, which takes $n \log n$ in expectation.

Using the so-called coupling method, one could prove $t_{mix} \leq n \log n$. 

---

This deck of cards is perfectly mixed after the last card “8” reaches the top and is inserted to a random position!
Riffle Shuffle

1. Split a deck of \( n \) cards into two piles (thus the size of each portion will be Binomial)

2. Riffle the cards together so that the card drops from the left (or right) pile with probability proportional to the number of remaining cards

\[
\begin{array}{cccccccccc}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
1.000 & 1.000 & 1.000 & 1.000 & 0.924 & 0.614 & 0.334 & 0.167 & 0.085 & 0.043 \\
\end{array}
\]

Figure: Total Variation Distance for \( t \) riffle shuffles of 52 cards.
1. **Split** a deck of $n$ cards into two piles (thus the size of each portion will be Binomial)
Riffle Shuffle

1. **Split** a deck of \( n \) cards into two piles (thus the size of each portion will be *Binomial*)

2. **Riffle** the cards together so that the card drops from the left (or right) pile with probability proportional to the number of remaining cards
Riffle Shuffle

1. Split a deck of $n$ cards into two piles (thus the size of each portion will be Binomial)

2. Riffle the cards together so that the card drops from the left (or right) pile with probability proportional to the number of remaining cards

\[
\begin{array}{cccccccccc}
\text{a} & A & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & J & Q & K \\
\text{b} & A & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & J & Q & K \\
\text{c} & A & 7 & 2 & 8 & 9 & 3 & 10 & 4 & 5 & 6 & J & Q & K \\
\text{d} & A & 7 & 2 & 8 & 9 & 3 & 10 & 4 & 5 & J & 6 & Q & K \\
\end{array}
\]
Riffle Shuffle

1. Split a deck of $n$ cards into two piles (thus the size of each portion will be Binomial)

2. Riffle the cards together so that the card drops from the left (or right) pile with probability proportional to the number of remaining cards

<table>
<thead>
<tr>
<th>$t$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td></td>
<td>P^t - \pi</td>
<td></td>
<td>_{tv}$</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.924</td>
<td>0.614</td>
</tr>
</tbody>
</table>

Figure: Total Variation Distance for $t$ riffle shuffles of 52 cards.
Riffle Shuffle

1. Split a deck of \( n \) cards into two piles (thus the size of each portion will be Binomial)

2. Riffle the cards together so that the card drops from the left (or right) pile with probability proportional to the number of remaining cards

\[
\begin{array}{cccccccccccc}
\hline
\text{t} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
\|P^t - \pi\|_{tv} & 1.000 & 1.000 & 1.000 & 1.000 & 0.924 & 0.614 & 0.334 & 0.167 & 0.085 & 0.043 \\
\hline
\end{array}
\]

**Figure:** Total Variation Distance for \( t \) riffle shuffles of 52 cards.
Outline

Recap of Markov Chain Basics

Irreducibility, Periodicity and Convergence

Total Variation Distance and Mixing Times

Application 1: Card Shuffling

Application 2: Markov Chain Monte Carlo (non-examin.)

Appendix: Remarks on Mixing Time (non-examin.)
Given an undirected graph $G = (V, E)$, an independent set is a subset $S \subseteq V$ such that there are no two vertices $u, v \in S$ with $\{u, v\} \in E(G)$. 
Given an undirected graph $G = (V, E)$, an independent set is a subset $S \subseteq V$ such that there are no two vertices $u, v \in S$ with $\{u, v\} \in E(G)$.
Given an undirected graph $G = (V, E)$, an independent set is a subset $S \subseteq V$ such that there are no two vertices $u, v \in S$ with $\{u, v\} \in E(G)$. 

$S = \{2, 6, 8\}$ is an independent set.

[Diagram showing a graph with labeled vertices 1, 2, 3, 4, 5, 6, 7, 8, with edges connecting some of these vertices, and the independent set $S$ highlighted.]
Given an undirected graph $G = (V, E)$, an independent set is a subset $S \subseteq V$ such that there are no two vertices $u, v \in S$ with $\{u, v\} \in E(G)$.

$S = \{1, 7, 8\}$ is not an independent set ×
Given an undirected graph $G = (V, E)$, an independent set is a subset $S \subseteq V$ such that there are no two vertices $u, v \in S$ with $\{u, v\} \in E(G)$. 
Given an undirected graph $G = (V, E)$, an independent set is a subset $S \subseteq V$ such that there are no two vertices $u, v \in S$ with $\{u, v\} \in E(G)$.

How can we take a sample from the space of all independent sets?
Given an undirected graph $G = (V, E)$, an independent set is a subset $S \subseteq V$ such that there are no two vertices $u, v \in S$ with $\{u, v\} \in E(G)$.

How can we take a sample from the space of all independent sets?

Naive brute-force would take an insane amount of time (and space)!
Given an undirected graph $G = (V, E)$, an independent set is a subset $S \subseteq V$ such that there are no two vertices $u, v \in S$ with $\{u, v\} \in E(G)$.

How can we take a sample from the space of all independent sets?

Naive brute-force would take an insane amount of time (and space)!

We can use a generic Markov Chain Monte Carlo approach to tackle this problem!
INDEPENDENT SET SAMPLER

1: Let $X_0$ be an arbitrary independent set in $G$
2: For $t = 0, 1, 2, \ldots$
3: Pick a vertex $v \in V(G)$ uniformly at random
4: If $v \in X_t$ then $X_{t+1} \leftarrow X_t \setminus \{v\}$
5:elif $v \notin X_t$ and $X_t \cup \{v\}$ is an independent set then $X_{t+1} \leftarrow X_t \cup \{v\}$
6: else $X_{t+1} \leftarrow X_t$

Key Question: What is the mixing time of this Markov Chain? (not covered here, see the textbook by Mitzenmacher and Upfal)
INDEPENDENT SET SAMPLER

1: Let $X_0$ be an arbitrary independent set in $G$
2: For $t = 0, 1, 2, \ldots$:
3: Pick a vertex $v \in V(G)$ uniformly at random
4: If $v \in X_t$ then $X_{t+1} \leftarrow X_t \setminus \{v\}$
5: elif $v \notin X_t$ and $X_t \cup \{v\}$ is an independent set then $X_{t+1} \leftarrow X_t \cup \{v\}$
6: else $X_{t+1} \leftarrow X_t$

$X_0 = \{1, 4\}$

Diagram:

```
      5
     / \  \
    8   7
   /   /  \
  2   4   6
 /   /   /
1   3
```
INDEPENDENT_SET_SAMPLER

1: Let \( X_0 \) be an arbitrary independent set in \( G \)
2: \textbf{For} \( t = 0, 1, 2, \ldots \):
3: \hspace{1em} Pick a vertex \( v \in V(G) \) uniformly at random
4: \hspace{1em} \textbf{If} \( v \in X_t \) \textbf{then} \( X_{t+1} \leftarrow X_t \setminus \{v\} \)
5: \hspace{1em} \textbf{elif} \( v \notin X_t \) \textbf{and} \( X_t \cup \{v\} \) is an independent set \textbf{then} \( X_{t+1} \leftarrow X_t \cup \{v\} \)
6: \hspace{1em} \textbf{else} \( X_{t+1} \leftarrow X_t \)

\[
X_0 = \{1, 4\}
\]

\[
v = 1
\]
**INDEPENDENT SET SAMPLER**

1. Let $X_0$ be an arbitrary independent set in $G$
2. **For** $t = 0, 1, 2, \ldots$
3. Pick a vertex $v \in V(G)$ uniformly at random
4. **If** $v \in X_t$ **then** $X_{t+1} \leftarrow X_t \setminus \{v\}$
5. **elif** $v \notin X_t$ **and** $X_t \cup \{v\}$ is an independent set **then** $X_{t+1} \leftarrow X_t \cup \{v\}$
6. **else** $X_{t+1} \leftarrow X_t$

\begin{itemize}
    \item $X_0 = \{1, 4\}$
    \item $X_1 = \{4\}$
\end{itemize}

$v = 1$
Markov Chain for Sampling Independent Sets (2/2) (non-examin.)

**INDEPENDENT_SET_SAMPLER**

1: Let $X_0$ be an arbitrary independent set in $G$
2: For $t = 0, 1, 2, \ldots$
3: Pick a vertex $v \in V(G)$ uniformly at random
4: If $v \in X_t$ then $X_{t+1} \leftarrow X_t \setminus \{v\}$
5: elif $v \notin X_t$ and $X_t \cup \{v\}$ is an independent set then $X_{t+1} \leftarrow X_t \cup \{v\}$
6: else $X_{t+1} \leftarrow X_t$

Key Question: What is the mixing time of this Markov Chain?

not covered here, see the textbook by Mitzenmacher and Upfal

---

$X_0 = \{1, 4\}$

$v = 1$

$v = 8$

$X_1 = \{4\}$
**INDEPENDENT_SET_SAMPLER**

1: Let $X_0$ be an arbitrary independent set in $G$
2: For $t = 0, 1, 2, \ldots$
3: Pick a vertex $v \in V(G)$ uniformly at random
4: If $v \in X_t$ then $X_{t+1} \leftarrow X_t \setminus \{v\}$
5: elif $v \notin X_t$ and $X_t \cup \{v\}$ is an independent set then $X_{t+1} \leftarrow X_t \cup \{v\}$
6: else $X_{t+1} \leftarrow X_t$

---

$X_0 = \{1, 4\}$

$v = 1$

$X_1 = \{4\}$

$v = 8$

$X_1 = \{1, 4, 8\}$
INDEPENDENT_SET_SAMPLER

1: Let $X_0$ be an arbitrary independent set in $G$
2: For $t = 0, 1, 2, \ldots$
3: Pick a vertex $v \in V(G)$ uniformly at random
4: If $v \in X_t$ then $X_{t+1} \leftarrow X_t \setminus \{v\}$
5: elif $v \not\in X_t$ and $X_t \cup \{v\}$ is an independent set then $X_{t+1} \leftarrow X_t \cup \{v\}$
6: else $X_{t+1} \leftarrow X_t$

Key Question:
What is the mixing time of this Markov Chain?

not covered here, see the textbook by Mitzenmacher and Upfal
Markov Chain for Sampling Independent Sets (2/2) (non-examin.)

**INDEPENDENTSET_SAMPLER**

1. Let $X_0$ be an arbitrary independent set in $G$
2. **For** $t = 0, 1, 2, \ldots$
3. Pick a vertex $v \in V(G)$ uniformly at random
4. **If** $v \in X_t$ **then** $X_{t+1} \leftarrow X_t \setminus \{v\}$
5. **elif** $v \notin X_t$ **and** $X_t \cup \{v\}$ is an independent set **then** $X_{t+1} \leftarrow X_t \cup \{v\}$
6. **else** $X_{t+1} \leftarrow X_t$

Key Question: What is the mixing time of this Markov Chain?

not covered here, see the textbook by Mitzenmacher and Upfal

---

4. Markov Chains and Mixing Times © T. Sauerwald Application 2: Markov Chain Monte Carlo (non-examin.) 27
Markov Chain for Sampling Independent Sets (2/2) (non-examin.)

INDEPENDENT SET SAMPLER

1: Let $X_0$ be an arbitrary independent set in $G$
2: For $t = 0, 1, 2, \ldots$
3: Pick a vertex $v \in V(G)$ uniformly at random
4: If $v \in X_t$ then $X_{t+1} \leftarrow X_t \setminus \{v\}$
5: elif $v \notin X_t$ and $X_t \cup \{v\}$ is an independent set then $X_{t+1} \leftarrow X_t \cup \{v\}$
6: else $X_{t+1} \leftarrow X_t$

Remark
**INDEPENDENTSETSampler**

1: Let $X_0$ be an arbitrary independent set in $G$
2: For $t = 0, 1, 2, \ldots$
3: Pick a vertex $v \in V(G)$ uniformly at random
4: If $v \in X_t$ then $X_{t+1} \leftarrow X_t \setminus \{v\}$
5: elif $v \notin X_t$ and $X_t \cup \{v\}$ is an independent set then $X_{t+1} \leftarrow X_t \cup \{v\}$
6: else $X_{t+1} \leftarrow X_t$

**Remark**

- This is a **local** definition (no explicit definition of $P$!)
Markov Chain for Sampling Independent Sets (2/2) (non-examin.)

INDEPENDENT SET Sampler

1: Let $X_0$ be an arbitrary independent set in $G$
2: For $t = 0, 1, 2, \ldots$
3: Pick a vertex $v \in V(G)$ uniformly at random
4: If $v \in X_t$ then $X_{t+1} \leftarrow X_t \setminus \{v\}$
5: elif $v \notin X_t$ and $X_t \cup \{v\}$ is an independent set then $X_{t+1} \leftarrow X_t \cup \{v\}$
6: else $X_{t+1} \leftarrow X_t$

Remark

- This is a local definition (no explicit definition of $P$!)
- This chain is irreducible (every independent set is reachable)
Markov Chain for Sampling Independent Sets (2/2) (non-examin.)

**INDEPENDENT SET Sampler**

1: Let $X_0$ be an arbitrary independent set in $G$
2: For $t = 0, 1, 2, \ldots$
3: Pick a vertex $v \in V(G)$ uniformly at random
4: If $v \in X_t$ then $X_{t+1} \leftarrow X_t \setminus \{v\}$
5: elif $v \not\in X_t$ and $X_t \cup \{v\}$ is an independent set then $X_{t+1} \leftarrow X_t \cup \{v\}$
6: else $X_{t+1} \leftarrow X_t$

**Remark**

- This is a **local** definition (no explicit definition of $P$!)
- This chain is **irreducible** (every independent set is reachable)
- This chain is **aperiodic** (Check!)

This is a local definition (no explicit definition of $P$!)
**Markov Chain for Sampling Independent Sets (2/2) (non-examin.)**

**INDEPENDENT_SET_SAMPLER**

1: Let $X_0$ be an arbitrary independent set in $G$
2: For $t = 0, 1, 2, \ldots$
3: Pick a vertex $v \in V(G)$ uniformly at random
4: If $v \in X_t$ then $X_{t+1} \leftarrow X_t \setminus \{v\}$
5: elif $v \not\in X_t$ and $X_t \cup \{v\}$ is an independent set then $X_{t+1} \leftarrow X_t \cup \{v\}$
6: else $X_{t+1} \leftarrow X_t$

Remark

- This is a local definition (no explicit definition of $P$!)
- This chain is irreducible (every independent set is reachable)
- This chain is aperiodic (Check!)
- The stationary distribution is uniform, since $P_{u,v} = P_{v,u}$ (Check!)
INDEPENDENTSETSampler

1: Let $X_0$ be an arbitrary independent set in $G$
2: For $t = 0, 1, 2, \ldots$
3: Pick a vertex $v \in V(G)$ uniformly at random
4: If $v \in X_t$ then $X_{t+1} \leftarrow X_t \setminus \{v\}$
5: elif $v \notin X_t$ and $X_t \cup \{v\}$ is an independent set then $X_{t+1} \leftarrow X_t \cup \{v\}$
6: else $X_{t+1} \leftarrow X_t$

Remark

- This is a local definition (no explicit definition of $P$!)
- This chain is irreducible (every independent set is reachable)
- This chain is aperiodic (Check!)
- The stationary distribution is uniform, since $P_{u,v} = P_{v,u}$ (Check!)

Key Question: What is the mixing time of this Markov Chain?
**Markov Chain for Sampling Independent Sets (2/2) (non-examin.)**

**INDEPENDENTSET_SAMPLER**

1. Let $X_0$ be an arbitrary independent set in $G$
2. For $t = 0, 1, 2, \ldots$:
3. Pick a vertex $v \in V(G)$ uniformly at random
4. If $v \in X_t$ then $X_{t+1} \leftarrow X_t \setminus \{v\}$
5. elif $v \notin X_t$ and $X_t \cup \{v\}$ is an independent set then $X_{t+1} \leftarrow X_t \cup \{v\}$
6. else $X_{t+1} \leftarrow X_t$

**Remark**

- This is a local definition (no explicit definition of $P$!)
- This chain is irreducible (every independent set is reachable)
- This chain is aperiodic (Check!)
- The stationary distribution is uniform, since $P_{u,v} = P_{v,u}$ (Check!)

**Key Question:** What is the mixing time of this Markov Chain?

not covered here, see the textbook by Mitzenmacher and Upfal
Recap of Markov Chain Basics

Irreducibility, Periodicity and Convergence

Total Variation Distance and Mixing Times

Application 1: Card Shuffling

Application 2: Markov Chain Monte Carlo (non-examin.)

Appendix: Remarks on Mixing Time (non-examin.)
Further Remarks on the Mixing Time (non-examin.)

- One can prove $\max_x \| P^t_x - \pi \|_{tv}$ is non-increasing in $t$ (this means if the chain is “$\epsilon$-mixed” at step $t$, then this also holds in future steps) \[ \text{[Mitzenmacher, Upfal, 12.3]} \]
Further Remarks on the Mixing Time (non-examin.)

- One can prove $\max_x \| P^t_x - \pi \|_{tv}$ is non-increasing in $t$ (this means if the chain is “$\varepsilon$-mixed” at step $t$, then this also holds in future steps) \[\text{[Mitzenmacher, Upfal, 12.3]}\]

- We chose $t_{mix} := \tau(1/4)$, but other choices of $\varepsilon$ are perfectly fine too (e.g, $t_{mix} := \tau(1/e)$ is often used); in fact, any constant $\varepsilon \in (0, 1/2)$ is possible.
Further Remarks on the Mixing Time (non-examin.)

- One can prove \( \max_x \| P^t_x - \pi \|_{tv} \) is non-increasing in \( t \) (this means if the chain is "\( \epsilon \)-mixed" at step \( t \), then this also holds in future steps) \[ \text{[Mitzenmacher, Upfal, 12.3]} \]

- We chose \( t_{mix} := \tau(1/4) \), but other choices of \( \epsilon \) are perfectly fine too (e.g, \( t_{mix} := \tau(1/e) \) is often used); in fact, any constant \( \epsilon \in (0, 1/2) \) is possible.

Remark: This freedom on how to pick \( \epsilon \) relies on the sub-multiplicative property of a (version) of the variation distance. First, let

\[
d(t) := \max_x \| P^t_x - \pi \|_{tv}
\]

be the variation distance after \( t \) steps when starting from the worst state. Further, define

\[
\overline{d}(t) := \max_{\mu, \nu} \| P^t_\mu - P^t_\nu \|_{tv}.
\]

These quantities are related by the following double inequality

\[
d(t) \leq \overline{d}(t) \leq 2d(t).
\]

Further, \( \overline{d}(t) \) is sub-multiplicative, that is for any \( s, t \geq 1 \),

\[
\overline{d}(s + t) \leq \overline{d}(s) \cdot \overline{d}(t).
\]

Hence for any fixed \( 0 < \epsilon < \delta < 1/2 \) it follows from the above that

\[
\tau(\epsilon) \leq \left[ \frac{\ln \epsilon}{\ln(2\delta)} \right] \tau(\delta).
\]

In particular, for any \( \epsilon < 1/4 \)

\[
\tau(\epsilon) \leq \left[ \log_2 \epsilon^{-1} \right] \tau(1/4).
\]
Further Remarks on the Mixing Time (non-examin.)

- One can prove $\max_x \| P^t_x - \pi \|_{tv}$ is non-increasing in $t$ (this means if the chain is “$\epsilon$-mixed” at step $t$, then this also holds in future steps) \[\text{[Mitzenmacher, Upfal, 12.3]}\]

- We chose $t_{mix} := \tau(1/4)$, but other choices of $\epsilon$ are perfectly fine too (e.g, $t_{mix} := \tau(1/e)$ is often used); in fact, any constant $\epsilon \in (0, 1/2)$ is possible.

Remark: This freedom on how to pick $\epsilon$ relies on the sub-multiplicative property of a (version) of the variation distance. First, let

$$d(t) := \max_x \| P^t_x - \pi \|_{tv}$$

be the variation distance after $t$ steps when starting from the worst state. Further, define

$$\overline{d}(t) := \max_{\mu, \nu} \| P^t_{\mu} - P^t_{\nu} \|_{tv}.$$ 

These quantities are related by the following double inequality

$$d(t) \leq \overline{d}(t) \leq 2d(t).$$

Further, $\overline{d}(t)$ is sub-multiplicative, that is for any $s, t \geq 1$,

$$\overline{d}(s + t) \leq \overline{d}(s) \cdot \overline{d}(t).$$

Hence for any fixed $0 < \epsilon < \delta < 1/2$ it follows from the above that

$$\tau(\epsilon) \leq \left\lceil \frac{\ln \epsilon}{\ln(2\delta)} \right\rceil \tau(\delta).$$

In particular, for any $\epsilon < 1/4$

$$\tau(\epsilon) \leq \left\lceil \log_2 \epsilon^{-1} \right\rceil \tau(1/4).$$

This 2 is the reason why we ultimately need $\epsilon < 1/2$ in this derivation. On the other hand, see [Exercise (4/5).8] why $\epsilon < 1/2$ is also necessary.
Further Remarks on the Mixing Time (non-examin.)

- One can prove $\max_x \| P^t_x - \pi \|_{tv}$ is non-increasing in $t$ (this means if the chain is $\epsilon$-mixed” at step $t$, then this also holds in future steps). \[\text{[Mitzenmacher, Upfal, 12.3]}\]

- We chose $t_{mix} := \tau(1/4)$, but other choices of $\epsilon$ are perfectly fine too (e.g, $t_{mix} := \tau(1/e)$ is often used); in fact, any constant $\epsilon \in (0, 1/2)$ is possible.

Remark: This freedom on how to pick $\epsilon$ relies on the sub-multiplicative property of a (version) of the variation distance. First, let

$$d(t) := \max_x \| P^t_x - \pi \|_{tv}$$

be the variation distance after $t$ steps when starting from the worst state. Further, define

$$\bar{d}(t) := \max_{\mu, \nu} \| P^t_\mu - P^t_\nu \|_{tv}.$$  

These quantities are related by the following double inequality

$$d(t) \leq \bar{d}(t) \leq 2d(t).$$

Further, $\bar{d}(t)$ is sub-multiplicative, that is for any $s, t \geq 1$,

$$\bar{d}(s + t) \leq \bar{d}(s) \cdot \bar{d}(t).$$

Hence for any fixed $0 < \epsilon < \delta < 1/2$ it follows from the above that

$$\tau(\epsilon) \leq \left\lceil \frac{\ln \epsilon}{\ln(2\delta)} \right\rceil \tau(\delta).$$

In particular, for any $\epsilon < 1/4$

$$\tau(\epsilon) \leq \left\lceil \log_2 \epsilon^{-1} \right\rceil \tau(1/4).$$

Hence smaller constants $\epsilon < 1/4$ only increase the mixing time by some constant factor.

This 2 is the reason why we ultimately need $\epsilon < 1/2$ in this derivation. On the other hand, see \[\text{[Exercise (4/5).8]}\] why $\epsilon < 1/2$ is also necessary.