A **Stochastic process** is a collection of random variables $X = \{X_t : t \in T\}$ (usually $T = \mathbb{N}^0$).

- $X_t$ is the state of the process at time $t \in T$:
  - $X(t)$ is an element of a discrete finite set $\Omega$.

**Examples:**

- A random coin/bit
- Step 1 output random bit, step $t > 1$ or more: if $X_{t-1} = 0$ then $X_t = X_{t-1} = 0$, otherwise toss a coin.
- Step 1 and 2 output random bits, step $t > 2$ or more: if $X_{t-1} = X_{t-2} = 0$ then $X_t = 0$, otherwise toss a coin.

- Process probability $\bar{p}(t) = (p_0(t), p_1(t), ..., p_n(t))$, where $|\Omega| = n$. $\bar{p}$ is a row vector.
Markov chains

Definition (Definition 7.1)
A discrete-time stochastic process is said to be a Markov chain if

\[ \Pr[X_t = a_t \mid X_{t-1} = a_{t-1}, \ldots, X_0 = a_0] = \Pr[X_t = a_t \mid X_{t-1} = a_{t-1}] . \]

Also memoryless property.

Examples:

- A random coin/bit
- Step 1 output random bit, step \( t > 1 \) or more: if \( X_{t-1} = 0 \) then \( X_t = X_{t-1} = 0 \), otherwise toss a coin.
- Step 1 and 2 output random bits, step \( t > 2 \) or more: if \( X_{t-1} = X_{t-2} = 0 \) then \( X_t = 0 \), otherwise toss a coin.
Graph $G = (V, E, w)$ representation of a Markov chain on the state set $\Omega = \{0, 1, 2, 3\}$.

- Vertices $V$ are states of the chain.
- There is an edge $(i, j) \in E$ if $P[j|i] > 0$.
- Edge weight $w(i, j) = P[j|i]$. 

Diagram:

![Diagram of a Markov chain graph with states 0, 1, 2, 3 and edges labeled with transition probabilities.]
The transition matrix $P$, where $P[a_{t-1}, a_t]$ denotes the probability $\Pr[X_t = a_t \mid X_{t-1} = a_{t-1}]$.

- $P$ in terms of a matrix of dimensions $|\Omega| \times |\Omega|$ (if $\Omega$ is finite) or of infinite dimension if $\Omega$ is countably infinite.

$$
\begin{bmatrix}
P[a_1, a_1] & P[a_1, a_2] & \ldots & P[a_1, a_j] & \ldots \\
P[a_2, a_1] & P[a_2, a_2] & \ldots & P[a_2, a_j] & \ldots \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
P[a_j, a_1] & P[a_j, a_2] & \ldots & P[a_j, a_j] & \ldots \\
\vdots & \vdots & \ddots & \vdots & \vdots 
\end{bmatrix}
$$

$P$ is stochastic iff $\forall x : \sum_{y \in \Omega} P(x, y) = 1$. 

$\text{RA (2022/23) – Lecture 13 – slide 5}$
Example Markov chain transition matrix

Previous example corresponds to the following *transition matrix*:

$$M = \begin{bmatrix}
0 & \frac{1}{4} & 0 & \frac{3}{4} \\
\frac{1}{2} & 0 & \frac{1}{3} & \frac{1}{6} \\
0 & 0 & 1 & 0 \\
0 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4}
\end{bmatrix}$$
Iterations of the Markov chain

Suppose we start our Markov process with the initial state $X_0$ being some fixed $a \in \Omega$.

- The “next state” $X_1$ has distribution $\bar{p}_1(y) = P(a, y)$ given by $a$’s row of the transition matrix $P$.

- We define $\bar{p}_0$ to be the row vector with $\bar{p}(a) = 1$ and all other entries 0, then we can define the probability distribution $\bar{p}_1$ by

$$\bar{p}_1 = \bar{p}_0 \cdot P,$$

- Second step of the Markov chain: the random variable $X_2$ will then be distributed according to $\bar{p}_2$:

$$\bar{p}_2 = \bar{p}_1 \cdot M = \bar{p}_0 \cdot M \cdot M = \bar{p}_0 \cdot M^2.$$

- After $t$ steps of the Markov chain $M$, the random variable $X_t$ will then be distributed according to $\bar{p}_t$, where

$$\bar{p}_t = \bar{p}_0 \cdot M^t.$$
Random walk on the $n$-dimensional hypercube

The $n$-dimensional hypercube is a graph whose vertices are the binary $n$-tuples $\{0, 1\}^n$. Two vertices are connected by an edge when they differ in exactly one coordinate.

The simple random walk on the hypercube:

- Choose a coordinate $j \in \{1, 2, \ldots, n\}$ uniformly at random.
- Set $x_j = x_j + 1 \mod 2$ (flip the bit).
Stationary distribution

Many interesting cases of Markov chain converge to their stationary distribution $\pi$, which under mild conditions is unique.

- A stationary distribution satisfies the condition:

\[
\pi = \pi P
\]  

(1)
Random walk on hypercube

The stationary distribution the uniform distribution over the $2^n$ binary $n$-tuples $\{0, 1\}^n$, i.e, $p(x_1, x_2, \ldots, x_n) = 1/2^n$.

Proof.

$\blacktriangleright$ $n = 1$: flipping a bit that is initially 0 or 1 with probability 1/2 does not change the overall distribution.

$\blacktriangleright$ The global uniform distribution is equivalent to the product if its marginals, i.e, $p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2)\ldots p(x_n) = 1/2^n$.

$\blacktriangleright$ The bit flip of $x_j$ does not change the marginal $p(x_j)$, neither $p(x_1, x_2, \ldots, x_n)$.
Convergence: Detail balance (time reversible chains)

Theorem (Th. 7.10 (+ Th. 7.7))

Consider a finite, irreducible, and aperiodic Markov chain with transition matrix $P$. If there is a probability distribution $\pi$ that for each pair of state $i, j$

$$\pi_i P_{i,j} = \pi_j P_{j,i},$$

then $\pi$ is the unique stationary distribution corresponding to $P$. 
Irreducible

Any state must have a non-zero probability to reach any other state.

Lemma (7.4)

A finite Markov chain is irreducible if and only if its graph representation is a strongly connected graph.

Counterexample I: Disconnected graph
Counterexample II: Coupon Collector

A collector desires to complete a collection of $n$ coupons. We suppose each coupon acquired is equally likely. Let $X_t$ denote the number of different types represented among the collector’s $t$ acquired coupons:

- $P(k, k + 1) = (n - k)/n$
- $P(k, k) = k/n$
Periodic

A state $j$ in a Markov chain is periodic if $\exists \Delta > 1$ such that
$\Pr(X_{t+s} = j | X_t = j) = 0$ unless $s$ is divisible by $\Delta$. A Markov chain is periodic if any state in the chain is periodic. A state or chain that is not periodic is periodic.

**Periodic example: Random walk on the $n$-cycle**

Let $\Omega = \mathbb{Z}_n = \{0, 1, ..., n - 1\}$ and consider the transition matrix:

$$P(j, k) = \begin{cases} 
1/2 & \text{if } k \equiv j + 1 \pmod{n} \\
1/2 & \text{if } k \equiv j - 1 \pmod{n} \\
0 & \text{Otherwise.}
\end{cases}$$

(2)
Curing periodicity

On can always turn a periodic Markov chain into an aperiodic one by replacing $P$ by $Q = \frac{P + I}{2}$, where $I$ is the identity matrix. Indeed any convex mixture of $P$ and $I$ that has non-zero probability of $I$ will work.

**Example: Lazy random walk on the $n$-cycle**

Let $\Omega = \mathbb{Z}_n = \{0, 1, \ldots, n-1\}$ and consider the transition matrix:

\[
P(j, k) = \begin{cases} 
1/4 & \text{if } k \equiv j + 1 \pmod{n} \\
1/4 & \text{if } k \equiv j - 1 \pmod{n} \\
1/2 & \text{Otherwise.}
\end{cases}
\]  

(3)
Lazy random walk on the $n$-dimensional hypercube

The random walk on the hypercube is periodic, as it alternates parity at each step of the walk.

Let $\Omega = \{0, 1\}^n$ the $n$-tuple, the following lazy random walk on the hypercube:

- Choose a coordinate $j \in \{1, 2, \ldots, n\}$ uniformly at random.
- Set $x_j = x_j + 1 \pmod{2}$ (flip the bit) with probability $1/2$.
- Set $x_j = x_j \pmod{2}$ with probability $1/2$.
Convergence: Detail balance (time reversible chains)

Theorem (Th. 7.10 (+ Th. 7.7))
Consider a finite, irreducible, and aperiodic Markov chain with transition matrix $P$. If there is a probability distribution $\pi$ that for each pair of state $i, j$

$$\pi_i P_{i,j} = \pi_j P_{j,i},$$

then $\pi$ is the unique stationary distribution corresponding to $P$. 