Markov Chains
and
Markov Processes

Prakash Panangaden
McGill University

Some informal definitions

- A stochastic process is the most general kind of probabilistic dynamical system.
- A Markov process is a stochastic process where the dynamics is independent of the past history.
- A Markov chain is a discrete space and time Markov process.
- A continuous time Markov chain is a Markov chain with exponentially distributed delays.

Conditional Probability

- If probability theory is the "logic of science" then conditional probability is the counterpart of implication: the "engine" of deduction.
- Intuitively, $p(A|B)$ means the probability of A given that B has occurred.

A simple puzzle

Suppose that you have three cards. One is green on both sides, one is red on both sides and one is red on one side and green on the other. There are no other distinguishing marks.

A card is chosen at random and a side is chosen at random. This is shown to you and it turns out to be green.

What is the probability that the other side is also green?

Imagine that there is a "random" process governed by some probability distribution P. Let the space of outcomes be $\{\ldots, \omega, \ldots\}$.

$P(A)$ describes what the observer thinks is the probability that the outcome belongs to the set A.

If the observer learns that the outcome is in B, he *revises* his estimate of the probability to

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

If the fact that the outcome is in B conveys no information about A, we say that the events A and B are *independent* and we have

$$P(A|B) = P(A).$$

Markov chains

Let S be a finite or countable set (the states).

An S-valued random variable X, is a function from some probability space (Ξ, P) to S.

We write $P(\{X = s\})$ for $P(X^{-1}(\{s\})$.

Let $X_0, X_1, \ldots, X_n, \ldots$ be a sequence of S-valued random variables.

This is a Markov chain if:

$$P(X_{n+1} = s | X_0 = s_0, \ldots, X_n = s_n) = P(X_{n+1} = s | X_n = s_n)$$

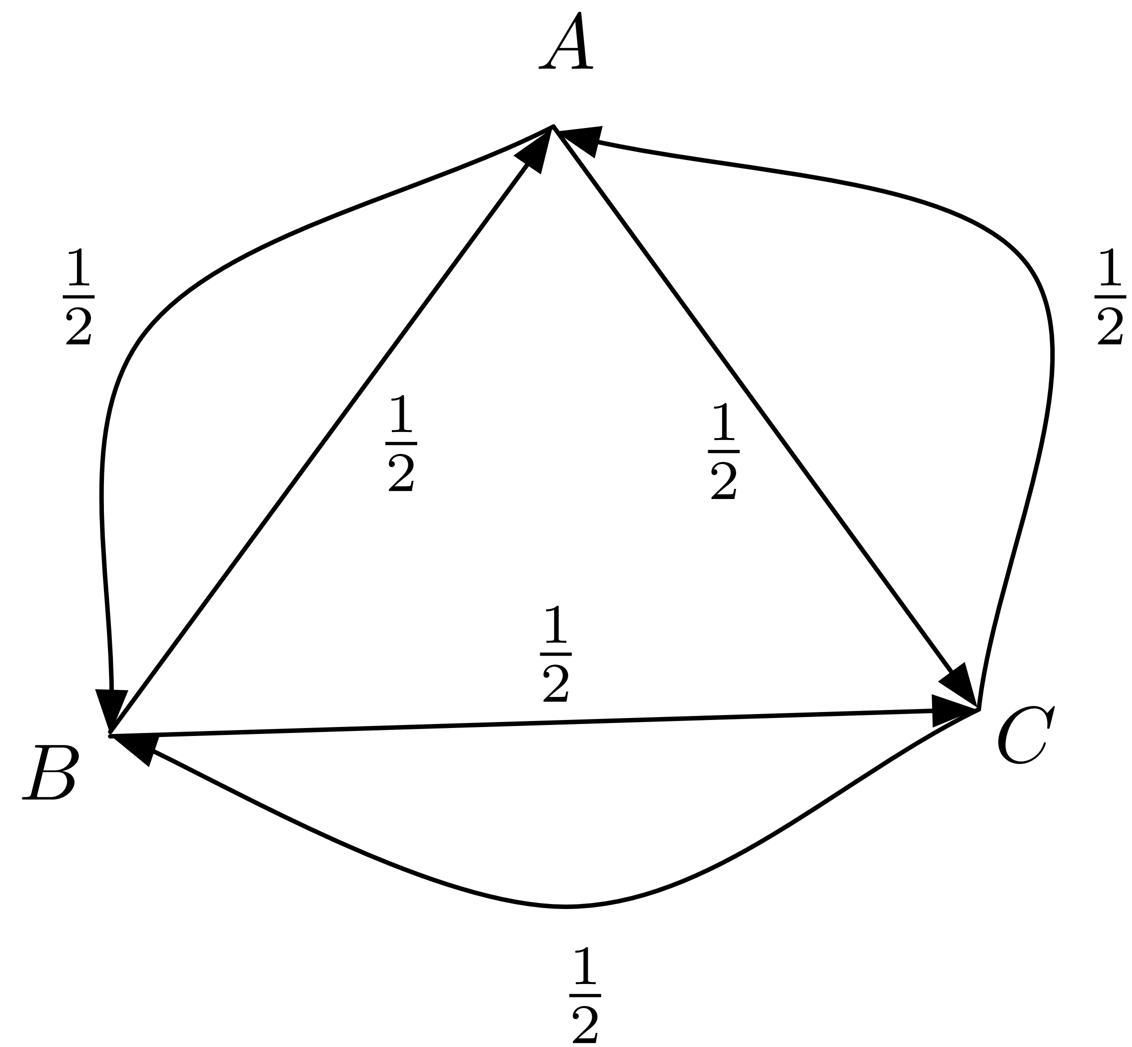
for every n and every sequence $s_0, \ldots, s_n$.

We think of $P(X_{n+1} = s_j | X_n = s_i)$ as *transition probabilities* for the system to move from state s_i to s_j.
We write this as a matrix T_{ij}.

Sometimes we allow the transition probabilities to depend on n. We call these *time-dependent* Markov chains. It it still independent of *which states* the systems has been through.

Markov chains are just the sort of transition systems to which computer scientists are accustomed.

Example



A three-state system. At every step, the system jumps to one of the *other* states with equal probability.

What is the probability that after n steps the system has returned to its starting point?

$$\frac{1}{3}[1 - (-\frac{1}{2})^n]$$

Paths and Classes

There is a path from state s to state t, $s \mapsto t$, if there are states $s = s_0, s_1, \ldots, s_n = t$ with $T(s_i, s_{i+1}) > 0$ for all i; equivalently, when $T^n[s,t] > 0$.

We say s and t are *connected*, $s \leftrightarrow t$, if $s \mapsto t$ and $t \mapsto t$.

The relation $\leftrightarrow$ is an equivalence relation. The equivalence classes are called *communication* classes.

A class C is called *closed* if $s \in C$ and $s \mapsto t$ implies that $t \in C$. If $\{s\}$ is closed, s is an *absorbing* state.

A Markov chain consisting of a single class is said to be *irreducible.*

Recurrence and Transience

Recall X_n is a random variable that tells you the state of the Markov chain at step n.

If $Pr(X_{n+m} = s \text{ for infinitely many } n | X_m = s) = 1$ we say that s is *recurrent*.

Note, this is stronger than saying:
$Pr(X_{n+m} = s | X_m = s) > 0$ for infinitely many n.

If $Pr(X_{n+m} = s \text{ for infinitely many } n | X_m = s) = 0$ we say that s is *transient*.

Note that these two concepts are *not logical opposites*.

However:

Theorem Every state is either recurrent or transient.

Theorem All states in a class are recurrent or all states in the class are transient.

Theorem If all states are recurrent the class is closed.

Theorem Every finite closed class is recurrent.

Stationary distributions

If we have a Markov chain with state space S and transition matrix T, we say μ is a *stationary or invariant* distribution if $T\mu = \mu$.

Theorem Suppose that for some s and all t,

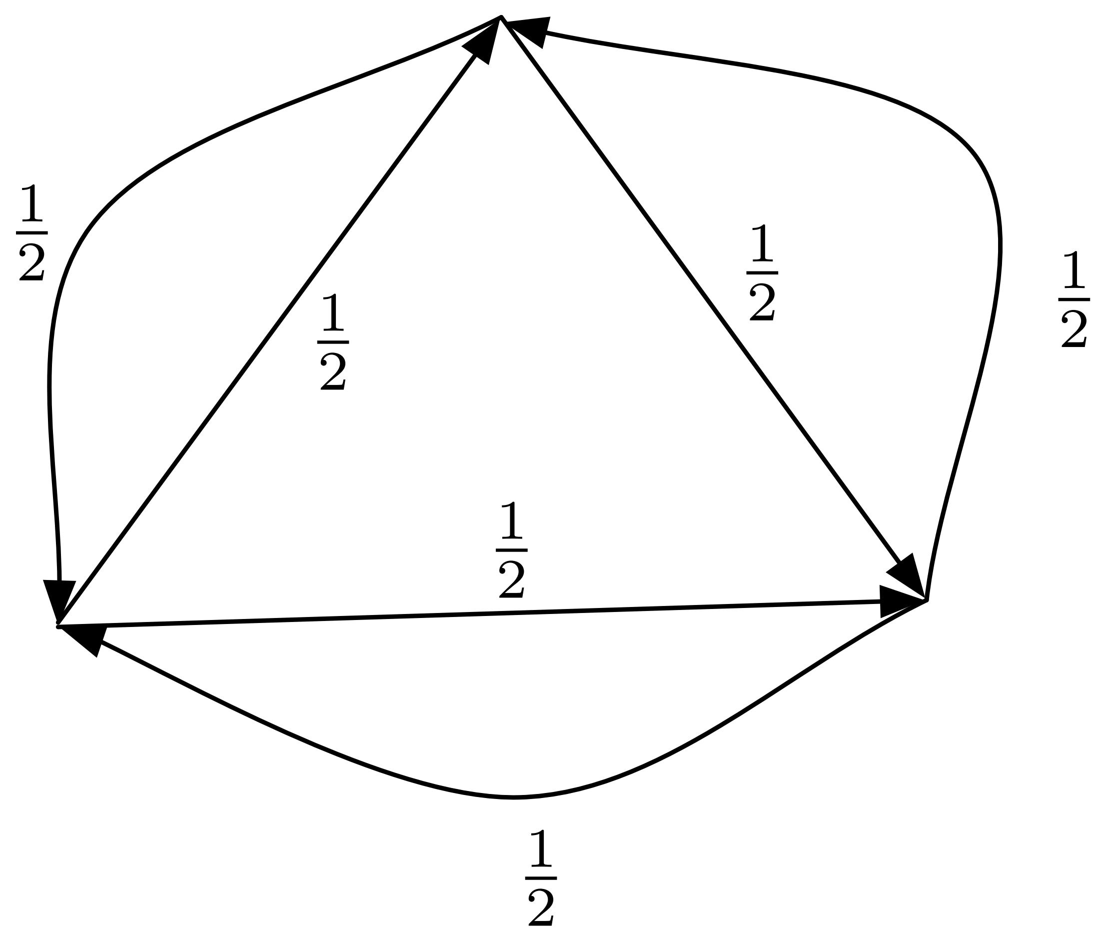
$\lim_{n\to\infty} T^n[s,t] = \pi(t)$

then π is a stationary distribution.

Proof Clearly $\forall n, s \;\; \sum_{t\in S} T^n[s,t] = 1$.
Hence, $\sum_t \pi(t) = \sum_t \lim_{n\to\infty} T^n[s,t] = \lim_{n\to\infty} 1 = 1$.

$$\pi(t) = \lim_{n\to\infty} T^n[s,t] = \lim_{n\to\infty} (T \cdot T^n)[s,t] = (T \cdot \lim_{n\to\infty} T^n)[s,t] =$$

$$T \cdot \pi(t).$$



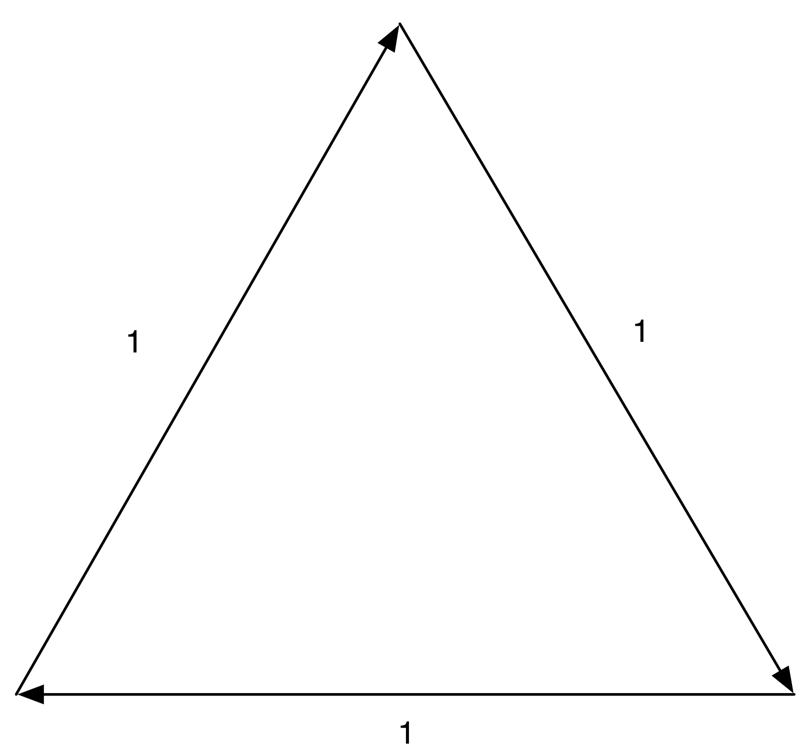
Here the invariant distribution is uniform as expected.

$$\lim_{n\to\infty} \frac{1}{3}[1 - (-\frac{1}{2})^n] = \frac{1}{3}.$$

Aperiodic chains

If d is the largest number such that it divides every n with $T^n[s,s] > 0$ we say that s has period d. Clearly in an irreducible chain every state has the same period.

If the period is 1 we call the chain aperiodic.



This chain has period 3.
It has no stationary distribution.

Theorem In an irreducible aperiodic Markov chain there are three possibilities:

(a) All states are transient: $\forall s, t \ \lim_{n\to\infty} T^n[s,t] = 0.$

(b) The chain is recurrent but there is no stationary distribution: $\forall s, t \ \lim_{n\to\infty} T^n[s,t] = 0.$
The convergence to zero is slow.

(c) There is a stationary distribution: $\forall s, t; \ T^n[s,t] = \pi(t) > 0.$

In a *finite* chain (a) and (b) are impossible so:

A **finite, irreducible, aperiodic** Markov chain has a stationary distribution.

Continuous-time processes

Fix a finite or countable state space S as before.

A continuous-time random process is an I-indexed family of random variables $X_t : \Omega \to S$ where I is an interval in $\mathbb{R}$.

Here (Ω, P) is some probability space.

We want to compute things like $P(\{X_{t_1} = s_1, X_{t_2} = s_2\})$.

Perhaps we want to know $P(\{X_t = s \text{ for some } t\})$.

We need to be careful, we cannot just use additivity because we are now potentially dealing with uncountable sets.

We will restrict to *right-continuous* processes.
In this case, the probability of any event
is determined by the *finite-dimensional* distributions:

$$P(X_{t_1} = s_1, X_{t_2} = s_2, \ldots, X_{t_n} = s_n)$$

$$P(X_t = s \text{ for some } t) =$$

$$1 - \lim_{n\to\infty} \sum_{s_1,\ldots,s_n \neq s} P(X_{q_1} = s_1, \ldots, X_{q_n} = s_n)$$

where q_i is an enumeration of the rationals.

What exactly does “right continuous” mean?

We assume the Markov condition:
$P(X_t = s | X_{t_1} = s_1, \ldots, X_{t_n} = s_n) = P(X_t = s | X_{t_n} = s_n)$,
where $t_1 < t_2 < \ldots < t_n < t$.

For any t, t' and s, s' we have "transition probabilities":
$P_{t,t'}(s, s') \stackrel{def}{=} P(X_{t'} = s' | X_t = s)$.

We assume *temporal homogeneity*: $P_{t,t'}(s, s') = P_{0,t'-t}(s, s')$,

therefore, we can just write P_t.

It follows that $P_{t_1+t_2} = P_{t_1} P_{t_2}$.

Finally, we assume that: $P_t \to I$ as $t \downarrow 0$.

The last property holds if and only if the $P_t(s, s')$ are *continuous* functions of t for fixed s, s'.

The main point: There is a matrix Q given by

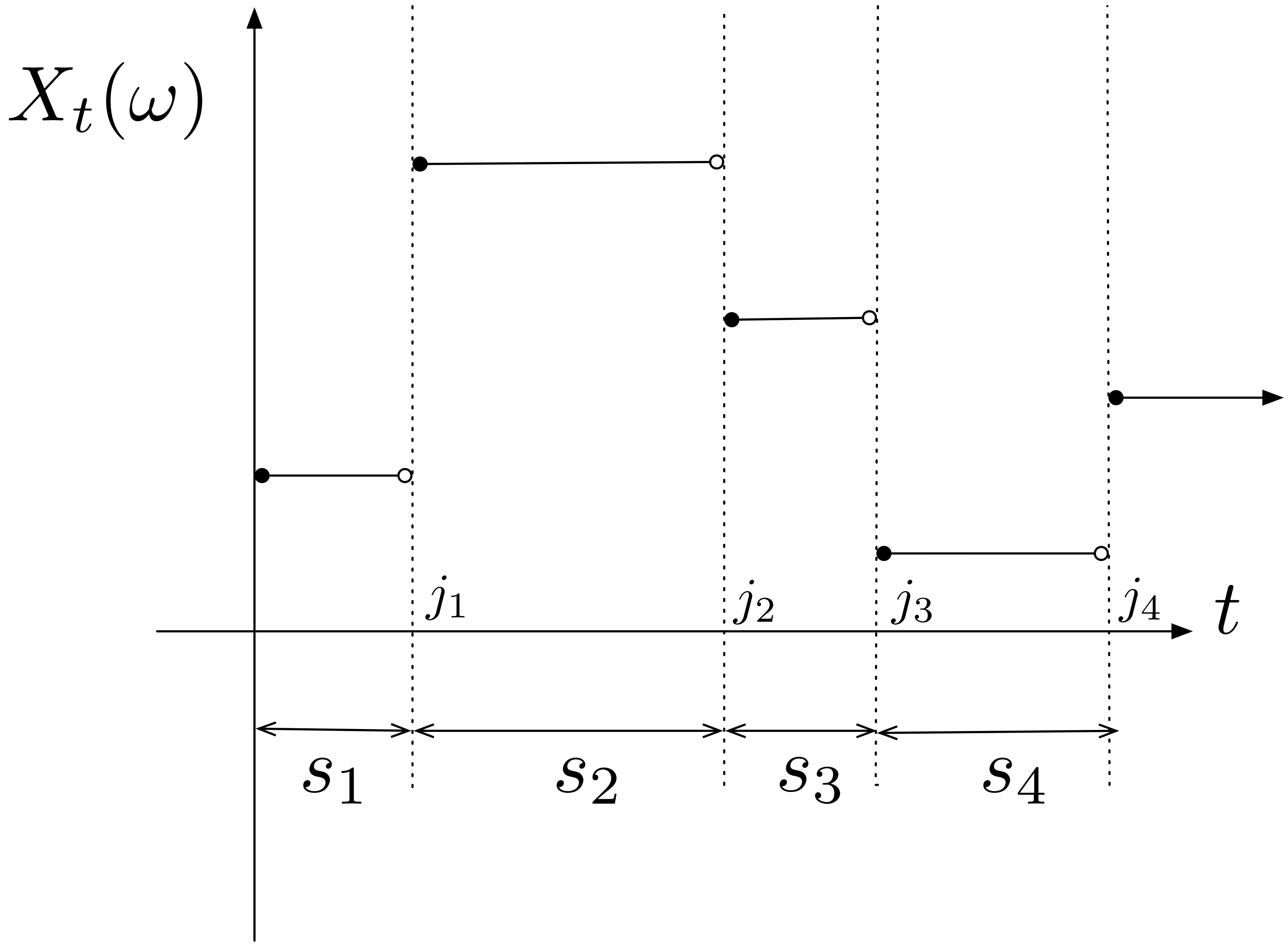
$$\lim_{\Delta t \downarrow 0} \frac{1}{\Delta t}(P_{\Delta t} - I) = Q.$$

Suppose that $X_t = s$, define T by $T = \inf\{t' \geq 0 | X_{t+t'} \neq s\}$.

T is how long the process waits in state s.

Theorem
T is exponentially distributed with parameter $-Q[s, s]$.

A typical history: there are infinitely many jumps but only finitely many in a finite interval.



Another possibility is finitely many jumps. A third is infinitely many jumps in a finite interval; this is ruled out by further technical conditions (cadlag).

Q determines everything

The transition matrices P_t are given by $\exp(Qt)$.

The matrix exponential is given by the usual power series:

$$\exp(Q) = \sum_n \frac{Q^n}{n!}$$

which has an infinite radius of convergence.

Q gives transition rates: class structure, recurrence and transience and invariant distributions can be defined in a way analogous to the discrete case.

Beware the matrix exponential

$$\exp(Q_1 + Q_2) = \exp(Q_1) \cdot \exp(Q_2)$$

but only if $[Q_1, Q_2] \stackrel{def}{=} Q_1 Q_2 - Q_2 Q_1 = 0.$

If $[Q_1, [Q_1, Q_2]] = [Q_2, [Q_1, Q_2]] = 0$ then
$\exp(Q_1 + Q_2) = \exp(Q_1) \exp(Q_2) \exp(-\frac{1}{2}[Q_1, Q_2])$

BREAK!!