Probability and Random Process

COSE382

Wonzoo Chung

Random Process (Stochastic Process)

(Random Process) Random process is an indexed set of random variables.

Examples:

1) Two random variables (X_1, X_2)
2) Random vector $[X_1, \cdots, X_n]^T$
3) Random sequence $\{X_i\}$

The index set can be finite, infinite, or continuous, $\mathbb{R}$, even $\mathbb{R}^m$.

The real line $\mathbb{R}$ is a widely used index set and interpreted as "time"

$$\{X_t | t \in \mathbb{R}\}$$

In order to specify the properties of uncountable random variables $\{X_t\}$ we define several functions of t.

Specifying Random Process

Given a random process X_t,

Mean function

$$m_X(t) = E(X_t) = \int x f_{X_t}(x) dx$$

Autocorrelation function

$$R_X(t,s) = E(X_t X_s) = \int \int x_1 x_2 f_{X_t, X_s}(x_1, x_2) dx_1 dx_2$$

Autocovariance function

$$C_X(t,s) = E((X_t - \mu_t)(X_s - \mu_s)) = R_X(t,s) - m_X(t) m_X(s)$$

Variance function

$$Var_X(t) = C_X(t,t)$$

Gaussian Random Process

- **(Gaussian Process).** A random process $\{X_t\}$ is called Gaussian process if any finite subset of $\{X_t\}$ is jointly Gaussian.

$$(X_{t_1}, \cdots, X_{t_n}) \sim \mathcal{N}(\mathbf{m}, \Sigma)$$

where $\mathbf{m} = [m(t_1), \cdots, m(t_n)]^T$ and $\Sigma_{ij} = C_X(t_i, t_j)$.

Since jointly Gaussian variables are completely determined by mean vector and covariance matrix, a mean function $m(t)$ and a covariance function C_X determines a Gaussian process, denoted by

$$X_t \sim \mathcal{GP}(m(t), C(t, s))$$

- **(Kernel function).** The covariance function of Gaussian process is sometimes called kernel function. A famous example of kernel function is a radial basis function

$$k(t, s) = \sigma^2 \exp\left(-\frac{\|t-s\|^2}{2\ell^2}\right)$$

Example

Wiener Process W_t is a Gaussian Process defined over $\{t \geq 0 | t \in \mathbb{R}\}$ with zero mean function and covariance function $C(t,s) = \min(t,s)$;

$$W_t \sim \mathcal{GP}(m(t) = 0, C(t,s) = \min(t,s))$$

1. For a fixed t, $W_t \sim \mathcal{N}(0,t) = \sqrt{t}\mathcal{N}(0,1)$:
   Since W_t is a GP, W_t for a fixed t should be Gaussian and zero mean $(m(t) = 0)$ with

   $$\mathrm{Var}(W_t) = \min(t,t) = t$$

2. For $t > s$, $W_t - W_s = W_{t-s}$:
   Since W_t is a GP, $W_t - W_s$ is a Gaussian. The mean is zero and the variance is

   $$\mathrm{Var}(W_t - W_s) = \mathrm{Var}(W_t) + \mathrm{Var}(W_s) - 2\mathrm{Cov}(W_t, W_s) = t - s$$

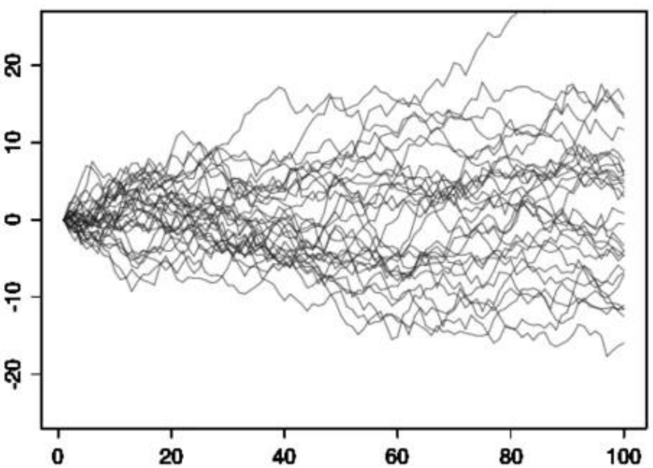
   Since Gaussian is completely determined by mean and variance $W_t - W_s = W_{t-s}$.

3. $W_t - W_s$ and $W_s - W_u$ are independent for $(t > s > u)$:

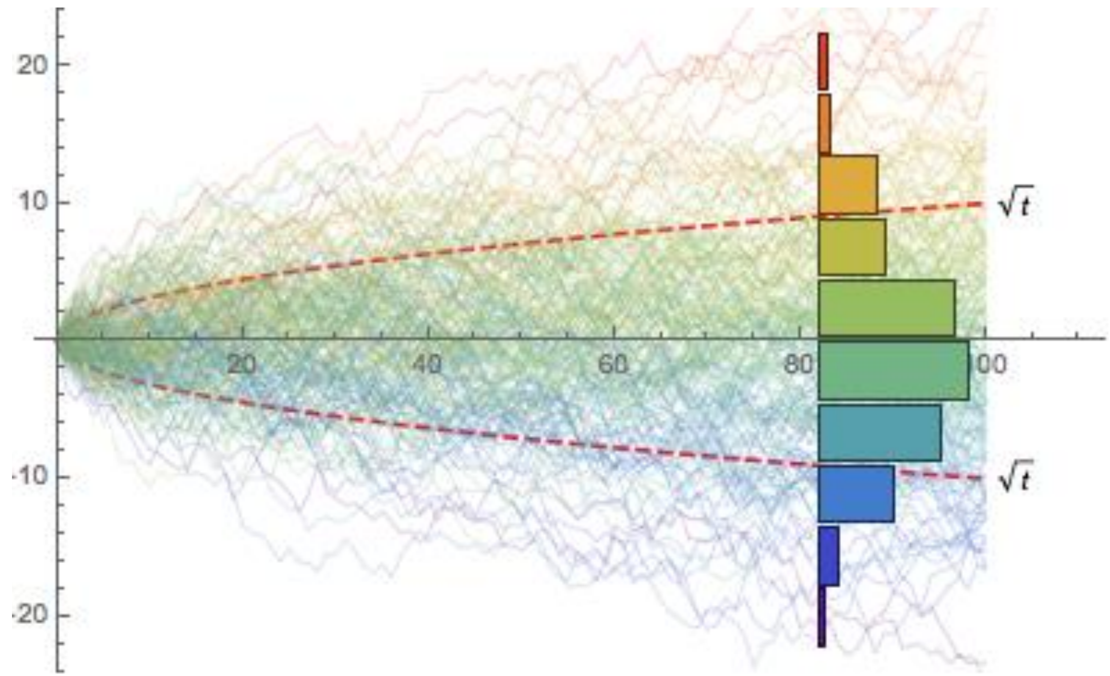
   $$E[(W_t - W_s)(W_s - W_u)] = E(W_t W_s) - E(W_t W_u) - E(W_s W_s) + E(W_s W_u) = s - u - s + u = 0$$

   Since $W_t - W_s$ and $W_s - W_u$ are Gaussian, uncorrelated implies independent.

For a discrete simulation, we can use $W_{t+dt} = W_t + \sqrt{dt}\mathcal{N}(0,1)$ starting from $W_0 = 0$.



20
10
0
-10
-20
0
20
40
60
80
100



20
10
-10
-20
20
40
60
80
100
$\sqrt{t}$
$\sqrt{t}$

Markov Chain

Definition 11.1.1 (Markov chain). Consider a random process $\{X_n\}$ with integer index set and taking values in a fixed integer set $\{1, 2, \cdots, M\}$

$$X_n \in \{1, 2, \cdots, M\} \text{ for all } n.$$

$\{X_n\}$ is called a *Markov chain*, if it has *Markov property*, i.e., for all $n \geq 0$,

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = j | X_n = i).$$

Some jargons and notations:

- State: a specific realization of X_n among $\{1, 2, \cdots, M\}$.
- State space: the integer set of all possible states; $\{1, 2, \ldots, M\}$.
- Transition probability from the state i to state j: $q_{ij} := P(X_{n+1} = j | X_n = i)$
- time-homogeneous: the transition probability $P(X_{n+1} = j | X_n = i)$ is the same for all n.

We assume time-homogeneous

Transition Matrix

Definition 11.1.2 (Transition matrix). Let $X_0, X_1, X_2, \ldots$ be a Markov chain with state space $\{1, 2, \cdots, M\}$, and let $q_{ij} = P(X_{n+1} = j | X_n = i)$ be the transition probability from state i to state j. The $M \times M$ matrix $Q = (q_{ij})$,

$$Q = \begin{pmatrix} q_{11} & q_{12} & \cdots & q_{1M} \\ q_{21} & q_{12} & \cdots & q_{1M} \\ \vdots & \ddots & & \\ q_{M1} & q_{M2} & \cdots & q_{MM} \end{pmatrix}$$

is called the *transition matrix* of the Markov chain.

Note that Q is a nonnegative matrix in which each row sums to 1:

$$\begin{aligned} q_{i1} + q_{i2} + \cdots + q_{iM} &= P(X_{n+1} = 1 | X_n = i) + \cdots + P(X_{n+1} = M | X_n = i) \\ &= P(\{X_{n+1} = 1\} \cup \cdots \cup \{X_{n+1} = M\} | X_n = i) \\ &= 1 \end{aligned}$$

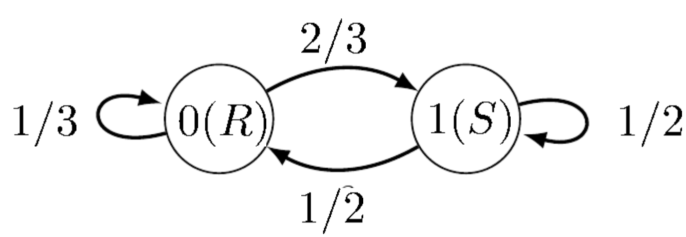
Example 11.1.3 (Rainy-sunny Markov chain). Let X_n be the weather on day n. Suppose the weather can either be rainy (R) or sunny (S).

- If today is R, then tomorrow will be R with $p = 1/3$ and S with $p = 2/3$.
- If today is S, then tomorrow will be R with $p = 1/2$ and S with $p = 1/2$.

X_n is a Markov chain with the state space $\{R(= 0), S(= 1)\}$

The transition matrix of the chain is

$$\begin{array}{cc} & \begin{array}{cc} R(0) & S(1) \end{array} \\ \begin{array}{c} R(0) \\ S(1) \end{array} & \begin{pmatrix} 1/3 & 2/3 \\ 1/2 & 1/2 \end{pmatrix} \end{array}$$



Definition 11.1.4 (n-step transition probability). The n-step transition probability from i to j is the probability of being at j exactly n steps after being at i. We denote this by $q_{ij}^{(n)}$

$$\begin{aligned} q_{ij}^{(n)} &= P(X_n = j | X_0 = i) \\ &= P(X_{n+k} = j | X_k = i) \text{ for any } k \text{ due to homogeneous} \end{aligned}$$

Note that

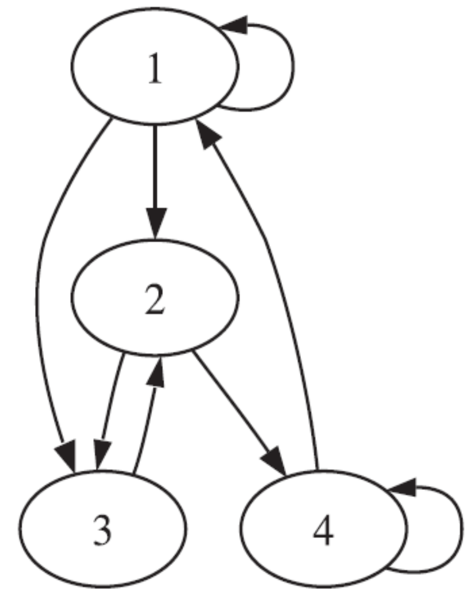
$$\begin{aligned} q_{ij}^{(2)} &= P(X_2 = j | X_0 = i) = \sum_{k=1}^{M} P(X_2 = j | X_1 = k, X_0 = i) P(X_1 = k | X_0 = i) \\ &= \sum_{k=1}^{M} q_{ik} q_{kj} \end{aligned}$$

$q_{ij}^{(2)}$ = the (i, j) entry of Q^2 and by induction we have

$$q_{ij}^{(n)} \text{ is the } (i, j) \text{ entry of } Q^n.$$

Example 11.1.5 (Transition matrix of 4-state Markov chain). Consider the 4-state Markov chain depicted in Figure (no probabilities over the arrows means equally likely).

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



$$Q^5 = \begin{pmatrix} 853/3888 & 509/1944 & 52/243 & 395/1296 \\ 173/864 & 85/432 & 31/208 & 91/288 \\ 37/144 & 29/72 & 1/9 & 11/48 \\ 499/2592 & 395/1296 & 71/324 & 245/864 \end{pmatrix}$$

To compute, $q_{13}^{(5)}$, the probability that the chain is in state 3 after 5 steps, starting at state 1, we would look at the (1,3) entry of Q^5. So

$$q_{13}^{(5)} = 52/243.$$

Proposition 11.1.6 (Marginal distribution of X_n). Define $\mathbf{t} = (t_1, t_2, \ldots, t_M)$ by $t_i = P(X_0 = i)$, and view $\mathbf{t}$ as a row vector. Then the marginal distribution of X_n is given by the vector $\mathbf{t}Q^n$. That is, the jth component of $\mathbf{t}Q^n$ is $P(X_n = j)$.

Proof. By the law of total probability, conditioning on X_0, the probability that the chain is in state j after n steps is

$$\begin{aligned} P(X_n = j) &= \sum_{i=1}^{M} P(X_0 = i) P(X_n = j | X_0 = i) \\ &= \sum_{i=1}^{M} t_i q_{ij}^{(n)}, \end{aligned}$$

which is the jth component of $\mathbf{t}Q^n$ by definition of matrix multiplication.

Example 11.1.7 (Marginal distributions of 4-state Markov chain). Consider the 4-state Markov chain in the previous example.

Given the initial conditions are $\mathbf{t} = (1/4, 1/4, 1/4, 1/4)$, the marginal distribution of X_1 is

$$
\begin{aligned}
\mathbf{t}Q &= \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix} \begin{pmatrix} 1/3 & 1/3 & 1/3 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 1 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 \end{pmatrix} \\
&= \begin{pmatrix} 5/24 & 1/3 & 5/24 & 1/4 \end{pmatrix}
\end{aligned}
$$

The marginal distribution of X_5 is

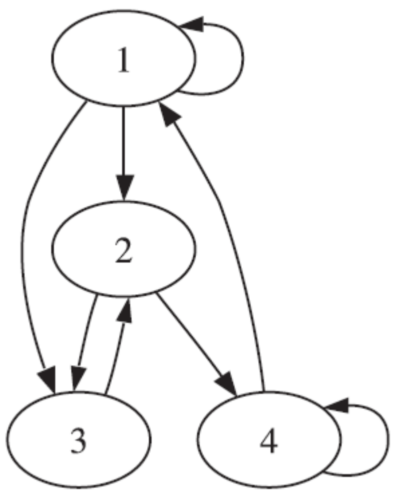
$$
\begin{aligned}
\mathbf{t}Q^5 &= \begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \end{pmatrix} \begin{pmatrix} 853/3888 & 509/1944 & 52/243 & 395/1296 \\ 173/864 & 85/432 & 31/208 & 91/288 \\ 37/144 & 29/72 & 1/9 & 11/48 \\ 499/2592 & 395/1296 & 71/324 & 245/864 \end{pmatrix} \\
&= \begin{pmatrix} 3379/15552 & 2267/7776 & 101/486 & 1469/5184 \end{pmatrix}
\end{aligned}
$$

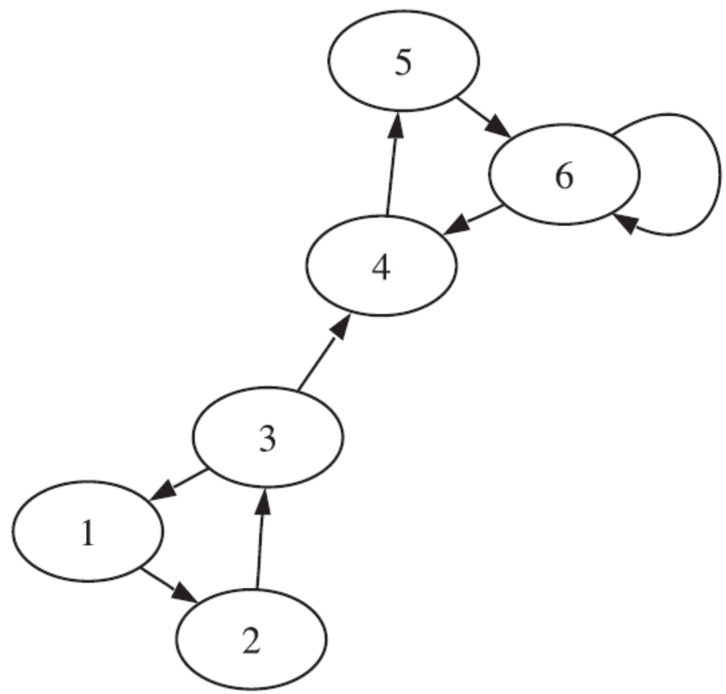
Classification of states

Definition 11.2.1 (Recurrent and transient states).

Recurrent state i: Starting from i, the event of the Markov chain eventually returning to i has probability 1.

Transient state i: Starting from i, the event of the Markov chain never returning to i has positive (non-zero) probability $p > 0$.





Proposition 11.2.2 (Number of returns to transient state is Geometric). Let i be a transient state with the probability p of never returning to i starting from i. Then, the number of events that the chain returns to i starting from i is Geom (p).

Proof. Each time that the chain is at i, we have a Bernoulli trial which results in"failure" if the chain eventually returns to i and "success" if the chain leaves i forever; these trials are independent by the Markov property. The number of returns to state i is simply the number of failures before the first success.

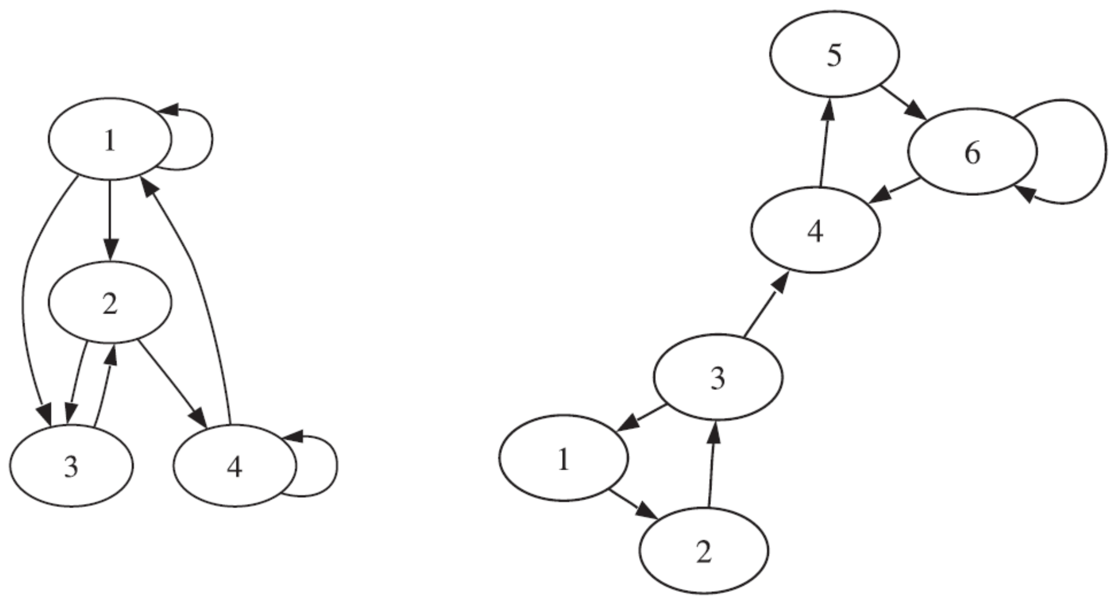
Since a Geometric random variable always takes finite values, this proposition tells us that after a finite number of visits, the chain will leave state i forever.

Definition 11.2.3 (Irreducible and reducible chain). A Markov chain with transition matrix Q is *irreducible* if for any two states i and j, there exists n such that

$$P(X_n = j | X_0 = i) = (Q^n)_{i,j} > 0,$$

i.e., for any states i, j, it is *possible* to go from i to j in a finite number of steps n.

A Markov chain that is not irreducible is called *reducible*.



Proposition 11.2.4 (Irreducible implies all states recurrent). In an irreducible Markov chain with a finite state space, all states are recurrent.

Proof.

At least one state must be recurrent, otherwise the chain leaves eventually all states forever. However, we have only finite states.

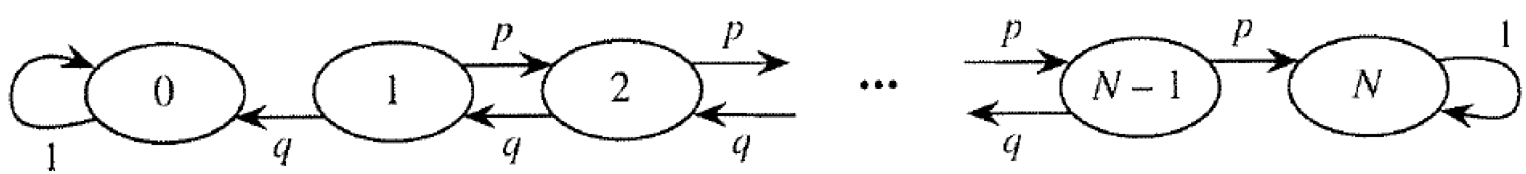
Let state 1 be recurrent. For any i, $q_{1i}^{(n)} > 0$ for some n, by the definition of irreducibility. It implies that there is a transition from 1 to i and should be a transition i to 1 (for 1 is recurrent).

Hence, from the perspective of state i, i is recurrent (for any $i \in \{1, 2, \cdots, M\}$)

Example 11.2.6 (Gambler's ruin as a Markov chain). Two gamblers, A and B, start with i and $N-i$ dollars, respectively, making a sequence of bets for \$1. In each round, A has probability p of winning and $q = 1-p$ of losing.

Let X_n be the wealth of gambler A at time n. Then $X_0, X_1, \ldots$ is a Markov chain on the state space $\{0, 1, \ldots, N\}$ with $X_0 = i$.

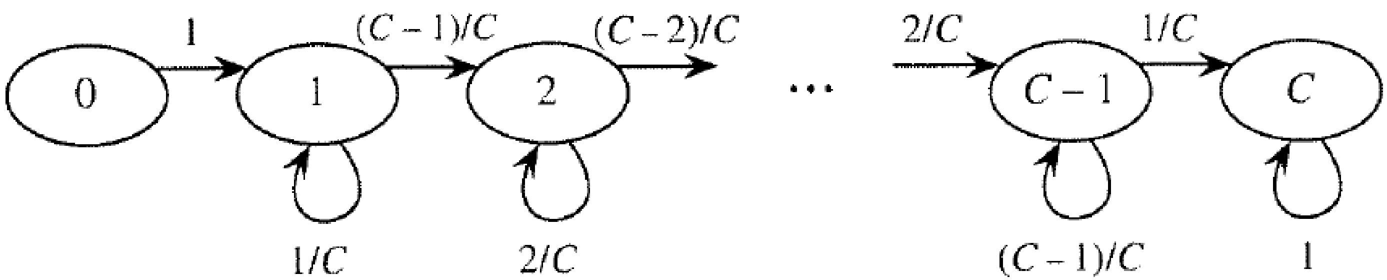
- Once the Markov chain reaches 0 or N, the Markov chain stays in that state forever.
- The probability that either A or B goes bankrupt is 1.
- Hence, states 0 and N are recurrent, and all other states are transient. The chain is reducible.



Example 11.2.7 (Coupon collector as a Markov chain). There are C types of coupons, which we collect one by one, sampling with replacement from the C coupon types each time.

Let X_n be the number of distinct coupon types in our collection after n attempts. Then $X_0, X_1, \ldots$ is a Markov chain on the state space $\{0, 1, \ldots, C\}$ with $X_0 = 0$.

- In this Markov chain, $X_n \le X_{n+1}$ for all n. Thus, all states are transient except for C, which is recurrent.
- Hence, this chain is reducible.



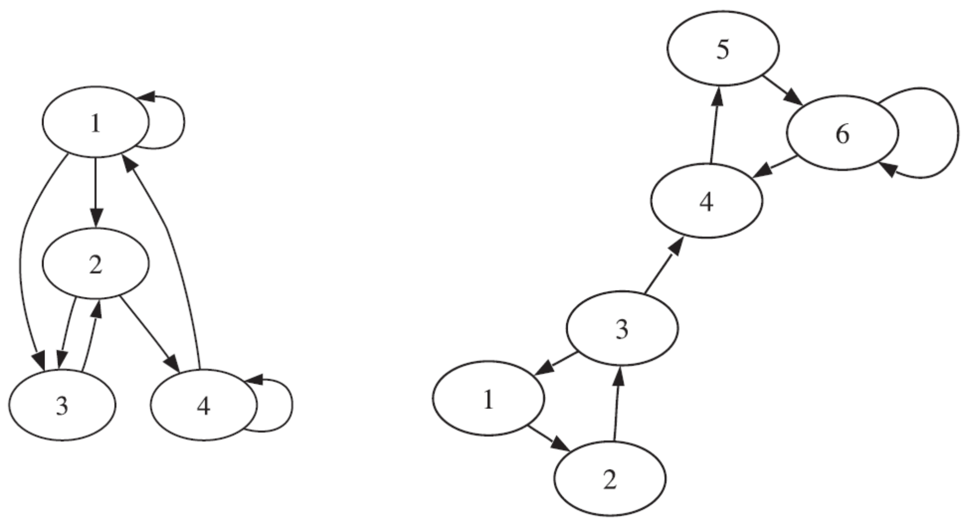
Definition 11.2.8 (Period of a state, periodic and aperiodic chain).

The *period* of a state i is the greatest common divisor (gcd) of the possible numbers of steps it can take to return to i when starting at i.

$$d_i = \gcd\left\{n \mid q_{ii}^{(n)} > 0\right\}$$

A state i is called *aperiodic* if its period equals 1 ($d_i = 1$), and *periodic* otherwise.

The chain is called *aperiodic* if all its states are aperiodic, and *periodic* otherwise.



Facts without proof:

- All states in a irreducible periodic Markov chain have the same period.

$$d_i = d_j \text{ for all } i, j$$

- An irreducible Markov chain is aperiodic if and only if there is a n such that Q^n is a positive matrix

$$Q^n(i, j) > 0 \text{ for all } i, j$$

- An irreducible Markov chain with a self-loop is aperiodic.

Stationary disribution

Definition 11.3.1 (Stationary distribution). A row vector $\mathbf{s} = (s_1, \ldots, s_M)$ such that $s_i \geq 0$ and $\sum_i s_i = 1$ is a *stationary distribution* for a Markov chain with transition matrix Q if

$$\sum_i s_i q_{ij} = s_j$$

for all j,or equivalently,

$$\mathbf{s}Q = \mathbf{s}.$$

Recall that if $\mathbf{s}$ is the distribution of X_0, then $\mathbf{s} = \mathbf{s}Q^k$ is the marginal distribution of X_k. for $k = 1, 2, 3, \cdots$

Example 11.3.4 (Stationary distribution for a two-state chain). Let

$$Q = \begin{pmatrix} 1/3 & 2/3 \\ 1/2 & 1/2 \end{pmatrix}.$$

The stationary distribution is of the form $\mathbf{t} = (s, 1-s)$, and we must solve for s in the system

$$\begin{pmatrix} s & 1-s \end{pmatrix} \begin{pmatrix} 1/3 & 2/3 \\ 1/2 & 1/2 \end{pmatrix} = \begin{pmatrix} s & 1-s \end{pmatrix},$$

which is equivalent to

$$\begin{aligned} \frac{1}{3}s + \frac{1}{2}(1-s) &= s, \\ \frac{2}{3}s + \frac{1}{2}(1-s) &= 1-s. \end{aligned}$$

The only solution is $s = 3/7$, so $(3/7,\, 4/7)$ is the unique stationary distribution of the Markov chain.

Theorem 11.3.5 (Existence and uniqueness of stationary distribution). Any irreducible Markov chain has a unique stationary distribution $\mathbf{s} = (s_1, s_2, \cdots, s_M)$ with all $s_i > 0$.

Proof. The theorem is a consequence of a result from linear algebra called the Perron-Frobenius theorem.

Theorem 11.3.6 without Proof (Convergence to stationary distribution). In any irreducible and aperiodic Markov chain $\{X_i\}$, with transition matrix Q and stationary distribution $\mathbf{s}$, $P(X_n = i)$ converges to the stationary distribution s_i, i.e.,

$$\lim_{n\to\infty} P(X_n = i) = s_i,$$

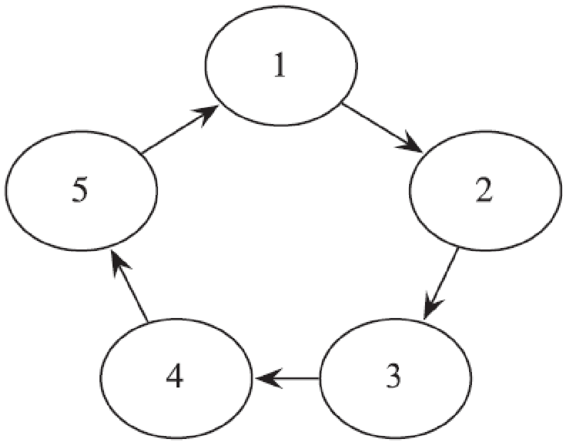
regardless of the initial distribution of X_0. In terms of the transition matrix, Q^n converges to a matrix in which each row is $\mathbf{s}$;

$$\lim_{n\to\infty} Q^n \to \begin{pmatrix} s_1 & s_2 & \cdots & s_M \\ s_1 & s_2 & \cdots & s_M \\ \vdots & \ddots & & \\ s_1 & s_2 & \cdots & s_M \end{pmatrix}$$

Note that for any $(a_1, a_2, \cdots, a_M)$ with $\sum_{i=1}^{M} a_i = 1$, we have

$$(a_1, a_2, \cdots, a_M) \begin{pmatrix} s_1 & s_2 & \cdots & s_M \\ s_1 & s_2 & \cdots & s_M \\ \vdots & \ddots & & \\ s_1 & s_2 & \cdots & s_M \end{pmatrix} = (s_1, s_2, \cdots, s_M)$$

Example 11.3.7 (Periodic chain). Consider the following periodic irreducible chain with the transition matrix Q



$$Q = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Since irreducible it has a unique stationary distribution $\mathbf{s} = (1/5, 1/5, 1/5, 1/5, 1/5)$

However, $P(X_n = i)$ does not converge to s_i as $n \to \infty$: Suppose the chain starts at $X_0 = 1$, then $P(X_n = (n \mod 5) + 1) = 1$.

Simple periodic and aperiodic chain.
Consider the following the transition matrices Q_1 and Q_2

$$Q_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Q_2 = \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Q_1 represents a periodic Markov chain with period 2, while Q_2 does an aperiodic one.

Q_1 has stationary distribution $\mathbf{s}_1 = [1/2, 1/2]$

Q_2 has stationary distribution $\mathbf{s}_2 = [1/3, 2/3]$

$$\begin{aligned} Q_1^2 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad Q_1^{2n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad Q_1^{2n+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \\ Q_2^2 &= \begin{pmatrix} 2/4 & 2/4 \\ 1/4 & 3/4 \end{pmatrix}, \quad Q_2^4 = \begin{pmatrix} 6/16 & 10/16 \\ 5/16 & 11/16 \end{pmatrix}, \quad Q_2^8 = \begin{pmatrix} 86/256 & 170/256 \\ 85/256 & 171/256 \end{pmatrix} \end{aligned}$$

$$Q_2^{16} = \begin{pmatrix} 21846/65536 & 43690/65536 \\ 21845/65536 & 43691/65536 \end{pmatrix} \approx \begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{pmatrix}$$

Long-run behavior of Markov Chain

Theorem 11.3.8 (Expected time to return). Let $X_0, X_1, \ldots$ be an irreducible Markov chain with stationary distribution $\mathbf{s}$. Let r_i be the expected time it takes the chain to return to i, starting at i. Then $s_i = 1/r_i$.

Example 11.3.9 (Long-run behavior of a two-state chain). Consider a Markov chain represented by $Q = \begin{pmatrix} 1/3 & 2/3 \\ 1/2 & 1/2 \end{pmatrix}$ (irreducible and aperiodic). The stationary distribution is $(3/7, 4/7)$ and we have

$$\begin{pmatrix} 1/3 & 2/3 \\ 1/2 & 1/2 \end{pmatrix}^n \to \begin{pmatrix} 3/7 & 4/7 \\ 3/7 & 4/7 \end{pmatrix} \text{as } n \to \infty$$

In the long run, the chain with will spend $3/7$ of its time in state 1 and $4/7$ of its time in state 2. Starting at state 1, it will take an average of $7/3$ steps to return to state 1. Starting at state 2, it will take an average of $7/4$ steps to return to state 2.

Reversibility

Definition 11.4.1 (Reversibility). Let $Q = (q_{ij})$ be the transition matrix of a Markov chain. Suppose there is $\mathbf{s} = (s_1, \ldots, s_M)$ with such that for all states i and j

$$s_i q_{ij} = s_j q_{ji}$$

This equation is called the *reversibility* or *detailed balance* condition, and we say that the chain is *reversible* with respect to $\mathbf{s}$, if it holds.

Proposition 11.4.2 (Reversible implies stationary). Suppose that $Q = (q_{ij})$ is a transition matrix of a Markov chain that is reversible with respect to a nonnegative vector $\mathbf{s} = (s_1, \ldots, s_M)$ whose components sum to 1. Then, $\mathbf{s}$ is a stationary distribution of the chain.

Proof. We have

$$\sum_i s_i q_{ij} = \sum_i s_j q_{ji} = s_j \sum_i q_{ji} = s_j,$$

Proposition 11.4.3. (Doubly stochastic matrix). If each column of the transition matrix Q sums to 1, then the uniform distribution over all states, $(1/M, 1/M, \ldots, 1/M)$, is a stationary distribution. (A nonnegative matrix such that the row sums and the column sums are all equal to 1 is called a *doubly stochastic matrix*.)

Proof. Assuming each column sums to 1, the row vector $\mathbf{v} = (1, 1, \ldots, 1)$ satisfies $\mathbf{v}Q = \mathbf{v}$. It follows that $(1/M, 1/M, \ldots, 1/M)$ is stationary.

- Note that a symmetric transition matrix is doubly stochastic.

Example 11.4.4 (Random walk on an undirected network). A network is a collection of *nodes* joined by *edges*; the network is *undirected* if edges can be traversed in either direction.

The *degree* of a node is the number of edges attached to it, and the *degree sequence* is the vector $\mathbf{d} = (d_1, \ldots, d_n)$ listing the degree d_j of node j. (self-loop is counted as 1)

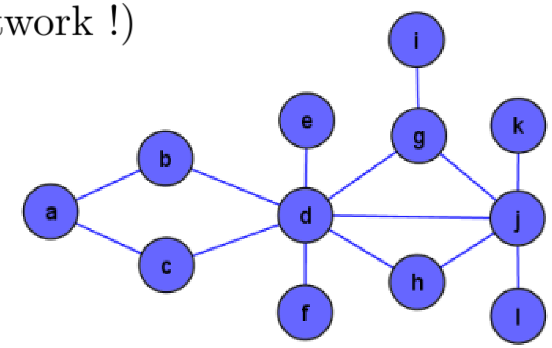
Suppose a wanderer randomly travels nodes through edge, with equal probabilities. Then

$$q_{ij} = \frac{1}{d_i} \quad \text{and} \quad d_i q_{ij} = d_j q_{ji}$$

Hence,

$$\mathbf{s} = \frac{1}{\sum_{i=1}^{n} d_i} \mathbf{d}$$

is a stationary distribution for the random walk. (*Every* reversible Markov chain can be represented as random walk on a weighted undirected network !)



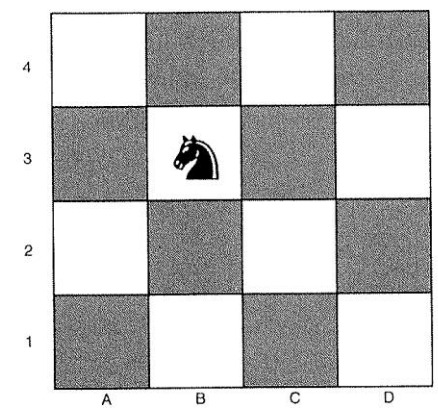
Example 11.4.5 (Knight on a chessboard). Consider a knight randomly moving around on a 4×4 chessboard. Suppose that at each step, the knight moves randomly, with each possibility equally likely. This creates a Markov chain where the states are the 16 squares. Compute the stationary distribution of the chain.

Solution:

There are only three types of squares on the board: 4 center squares that have degree 4, 4 corner squares that have degree 2, and 8 edge squares that have degree 3.

Hence, their stationary probabilities are $4a, 2a, 3a$ respectively for some a. To find a, count the number of squares of each type to get $4a \cdot 4 + 2a \cdot 4 + 3a \cdot 8 = 1$, giving $a = 1/48$.

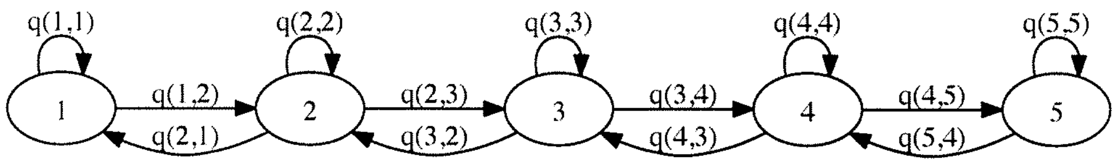
Thus, each center square has stationary probability $= 1/12$, each corner square has stationary probability $= 1/24$, and each edge square has stationary probability $= 1/16$.



Example 11.4.6 (Birth-death chain). A *birth-death* chain on states $1, 2, \dots, M$ is a Markov chain with transition matrix $Q = (q_{ij})$ such that

$$q_{ij} > 0 \text{ if } |i-j| = 1 \text{ and } q_{ij} = 0 \text{ if } |i-j| \geq 2$$

The chain shown below represent a birth-death chain if $q_{ij} > 0$ for $i \neq j$ (q_{ii} can be 0).



We will now show that any birth-death chain is reversible, and construct the stationary distribution. Let s_1 be a positive number. Since we want $s_1 q_{12} = s_2 q_{21}$, let $s_2 = s_1 q_{12}/q_{21}$. Then since we want $s_2 q_{23} = s_3 q_{32}$, let $s_3 = s_2 q_{23}/q_{32} = s_1 q_{12} q_{23}/(q_{32} q_{21})$. Continuing in this way, let

$$s_j = \frac{s_1 q_{12} q_{23} \cdots q_{j-1,j}}{q_{j,j-1} q_{j-1,j-2} \cdots q_{21}},$$

for the states $2 \leq j \leq M$. Choose s_1 so that the $\sum_{j=2}^{M} s_j = 1$. Then the chain is reversible with respect to $\mathbf{s}$, since $q_{ij} = q_{ji} = 0$ if $|i-j| \geq 2$ and by construction $s_i q_{ij} = s_j q_{ji}$ if $|i-j| = 1$. Thus, $\mathbf{s}$ is the stationary distribution.

Example 11.4.7 (Ehrenfest). There are two containers with a total of M distinguishable particles. Transitions are made by choosing a random particle and moving it from its current container into the other container. Initially, all of the particles are in the second container. Let X_n be the number of particles in the first container at time n, so $X_0 = 0$ and the transition from X_n to X_{n+1} is done as described above. This is a Markov chain with state space $\{0, 1, \ldots, M\}$.

We will use the reversibility condition to show that $\mathbf{s} = (s_0, s_1, \ldots, s_M)$ with $s_i = \binom{M}{i}\left(\frac{1}{2}\right)^M$ is the stationary distribution.

Let $s_i = \binom{M}{i}\left(\frac{1}{2}\right)^M$ and check that $s_i q_{ij} = s_j q_{ji}$. If $j = i + 1$ (for $i < M$), then

$$\begin{aligned}
s_i q_{ij} &= \binom{M}{i}\left(\frac{1}{2}\right)^M \frac{M-i}{M} = \frac{M!}{(M-i)!i!}\left(\frac{1}{2}\right)^M \frac{M-i}{M} = \binom{M-1}{i}\left(\frac{1}{2}\right)^M \\
s_j q_{ji} &= \binom{M}{j}\left(\frac{1}{2}\right)^M \frac{j}{M} = \frac{M!}{(M-j)!j!}\left(\frac{1}{2}\right)^M \frac{j}{M} = \binom{M-1}{j-1}\left(\frac{1}{2}\right)^M
\end{aligned}$$

By a similar calculation, if $j = i - 1$ (for $i > 0$), then $s_i q_{ij} = s_j q_{ji}$. For all other values of i and j, $q_{ij} = q_{ji} = 0$. Therefore, $\mathbf{s}$ is stationary.