Chapter 2: Discrete-Time Markov Models (Part ii , iii)

Andrey Markov (1856-1922, Russian mathematician)

Part ii: Limiting Behavior of DTMC

— What happens if a DTMC runs for a very long time ?

Start from these (simple two-state) examples

What happens to these Markov Chains as $n \to \infty$?

draw a picture, study the dynamics, see what you get ...

1. $\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
2. $\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
3. $\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}$
4. $\mathbf{P} = \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix}$

Start from these (simple two-state) examples

What happens to these Markov Chains as $n \to \infty$?
draw a picture, study the dynamics, see what you get ...

1. $\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
2. $\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
3. $\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}$
4. $\mathbf{P} = \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix}$

Observation : let $a^{(n)} = \Pr(X_n = \cdot)$ be transient distribution, then $a^{(n)} = a^{(0)}\mathbf{P}^n$. Let $n \to \infty$. If $a^{(n)} \to \pi$, then $\pi = a^{(0)} \lim_{n\to\infty} \mathbf{P}^n$
What is $\lim_{n\to\infty} \mathbf{P}^n$ for these examples ?

Start from these (simple two-state) examples

What happens to these Markov Chains as $n \to \infty$?
draw a picture, study the dynamics, see what you get ...

1. $\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
2. $\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$
3. $\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 1/2 & 1/2 \end{bmatrix}$
4. $\mathbf{P} = \begin{bmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{bmatrix}$

Observation : let $a^{(n)} = \Pr(X_n = \cdot)$ be transient distribution, then $a^{(n)} = a^{(0)}\mathbf{P}^n$. Let $n \to \infty$. If $a^{(n)} \to \pi$, then $\pi = a^{(0)} \lim_{n\to\infty} \mathbf{P}^n$

What is $\lim_{n\to\infty} \mathbf{P}^n$ for these examples ?

See also Textbook: Example 2.15, 2.16 (page 25)

Limiting distribution

Definition (limiting or *steady-state* distribution)

The (row) probability vector $\pi = [\pi_1, \pi_2, \cdots, \pi_N]$ is the limiting distribution where

$$\pi_j = \lim_{n \to \infty} \Pr(X_n = j).$$

Limiting distribution

Definition (limiting or *steady-state* distribution)

The (row) probability vector $\pi = [\pi_1, \pi_2, \cdots, \pi_N]$ is the limiting distribution where

$$\pi_j = \lim_{n \to \infty} \Pr(X_n = j).$$

Questions to answer:

- Does the limiting distribution π exists ?

Limiting distribution

Definition (limiting or *steady-state* distribution)

The (row) probability vector $\pi = [\pi_1, \pi_2, \cdots, \pi_N]$ is the limiting distribution where

$$\pi_j = \lim_{n \to \infty} \Pr(X_n = j).$$

Questions to answer:

- Does the limiting distribution π exists ?
- If it exists, is it unique (for *any* initial distribution)?

Limiting distribution

Definition (limiting or *steady-state* distribution)

The (row) probability vector $\pi = [\pi_1, \pi_2, \cdots, \pi_N]$ is the limiting distribution where

$$\pi_j = \lim_{n \to \infty} \Pr(X_n = j).$$

Questions to answer:

- Does the limiting distribution π exists ?
- If it exists, is it unique (for *any* initial distribution)?
- How can we compute it?

Limiting distribution

Definition (limiting or *steady-state* distribution)

The (row) probability vector $\pi = [\pi_1, \pi_2, \cdots, \pi_N]$ is the limiting distribution where

$$\pi_j = \lim_{n\to\infty} \Pr(X_n = j).$$

Questions to answer:

- Does the limiting distribution π exists ?
- If it exists, is it unique (for *any* initial distribution)?
- How can we compute it?

Limiting distribution

Definition (limiting or *steady-state* distribution)

The (row) probability vector $\pi = [\pi_1, \pi_2, \cdots, \pi_N]$ is the limiting distribution where

$$\pi_j = \lim_{n\to\infty} \Pr(X_n = j).$$

Questions to answer:

- Does the limiting distribution π exists ?
- If it exists, is it unique (for *any* initial distribution)?
- How can we compute it?

We first derive a *steady-state or balance equation* for π: a necessary condition that π must satisfy.

Theorem

If a limiting distribution π exists, it satisfies the following balancing equation

$$\pi_j = \sum_{i=1}^{N} \pi_i p_{ij}$$

or in matrix form

$$\pi = \pi \mathbf{P}$$

The limiting distribution π is the left-eigenvector of $\mathbf{P}$ corresponding to eigenvalue 1 satisfying the normalizing equation $\sum_{j=1}^{N} \pi_j = 1$.

Theorem

If a limiting distribution π exists, it satisfies the following balancing equation

$$\pi_j = \sum_{i=1}^{N} \pi_i p_{ij}$$

or in matrix form

$$\pi = \pi \mathbf{P}$$

The limiting distribution π is the left-eigenvector of $\mathbf{P}$ corresponding to eigenvalue 1 satisfying the normalizing equation $\sum_{j=1}^{N} \pi_j = 1$.

Proof.

The n-step transient distribution (the law of $\{X_n\}$), $a^{(n+1)}$, satisfies $a^{(n+1)} = a^{(n)}\mathbf{P}$. Let $n \to \infty$. Since $a^{(n)}, a^{(n+1)}$ both go to π, we have $\pi = \pi \mathbf{P}$. □

Theorem

If a limiting distribution π exists, it satisfies the following balancing equation

$$\pi_j = \sum_{i=1}^{N} \pi_i p_{ij}$$

or in matrix form

$$\pi = \pi \mathbf{P}$$

The limiting distribution π is the left-eigenvector of $\mathbf{P}$ corresponding to eigenvalue 1 satisfying the normalizing equation $\sum_{j=1}^{N} \pi_j = 1$.

Proof.

The n-step transient distribution (the law of $\{X_n\}$), $a^{(n+1)}$, satisfies $a^{(n+1)} = a^{(n)}\mathbf{P}$. Let $n \to \infty$. Since $a^{(n)}, a^{(n+1)}$ both go to π, we have $\pi = \pi \mathbf{P}$. $\square$

Textbook: Example 2.15, 2.16 (page 25)

Stationary or invariant distribution

Definition

$\mu = [\mu_1, \mu_2, \cdots, \mu_N]$ is a *stationary* (or *invariant*) distribution if

$$\Pr(X_0 = i) = \mu_i, \forall i \in S \implies \Pr(X_n = i) = \mu_i, \forall n > 0, \forall i \in S.$$

It is clear that the condition holds for $n = 1$ is sufficient for μ being invariant distribution.

Textbook: Example 2.17, 2.18 (page 28)

Definition

$\mu = [\mu_1, \mu_2, \cdots, \mu_N]$ is a *stationary* (or *invariant*) distribution if

$$\Pr(X_0 = i) = \mu_i, \forall i \in S \implies \Pr(X_n = i) = \mu_i, \forall n > 0, \forall i \in S.$$

It is clear that the condition holds for $n = 1$ is sufficient for μ being invariant distribution.

Theorem

A probability vector μ is a stationary distribution of the Markov chain with transition matrix $\mathbf{P}$ if and only if

$$\mu = \mu \mathbf{P}.$$

Textbook: Example 2.17, 2.18 (page 28)

If the limiting distribution π uniquely exists, then

- Since $a^{(n)} = a^{(0)}\mathbf{P}^n$, then $\pi = a^{(0)}\lim_n \mathbf{P}^n$ and π is independent of $a^{(0)}$. Particularly,

If the limiting distribution π uniquely exists, then

- Since $a^{(n)} = a^{(0)}\mathbf{P}^n$, then $\pi = a^{(0)}\lim_n \mathbf{P}^n$ and π is independent of $a^{(0)}$. Particularly,
  - choose $a^{(0)} = (1, 0, 0, \ldots, 0)$, then $\pi =$ the first row of $\lim_n \mathbf{P}^n$

If the limiting distribution π uniquely exists, then

- Since $a^{(n)} = a^{(0)}\mathbf{P}^n$, then $\pi = a^{(0)} \lim_n \mathbf{P}^n$ and π is independent of $a^{(0)}$. Particularly,
    - choose $a^{(0)} = (1, 0, 0, \ldots, 0)$, then $\pi =$ the first row of $\lim_n \mathbf{P}^n$
    - choose $a^{(0)} = (0, 1, 0, \ldots, 0)$, then $\pi =$ the second row of $\lim_n \mathbf{P}^n$

If the limiting distribution π uniquely exists, then

- Since $a^{(n)} = a^{(0)}\mathbf{P}^n$, then $\pi = a^{(0)}\lim_n \mathbf{P}^n$ and π is independent of $a^{(0)}$. Particularly,
  - choose $a^{(0)} = (1, 0, 0, \ldots, 0)$, then $\pi =$ the first row of $\lim_n \mathbf{P}^n$
  - choose $a^{(0)} = (0, 1, 0, \ldots, 0)$, then $\pi =$ the second row of $\lim_n \mathbf{P}^n$
  - ......

If the limiting distribution π uniquely exists, then

- Since $a^{(n)} = a^{(0)}\mathbf{P}^n$, then $\pi = a^{(0)}\lim_n \mathbf{P}^n$ and π is independent of $a^{(0)}$. Particularly,
  - choose $a^{(0)} = (1, 0, 0, \ldots, 0)$, then $\pi =$ the first row of $\lim_n \mathbf{P}^n$
  - choose $a^{(0)} = (0, 1, 0, \ldots, 0)$, then $\pi =$ the second row of $\lim_n \mathbf{P}^n$
  - ......
  - Each row of $\lim_n \mathbf{P}^n$ is exactly π.

If the limiting distribution π uniquely exists, then

- Since $a^{(n)} = a^{(0)}\mathbf{P}^n$, then $\pi = a^{(0)}\lim_n \mathbf{P}^n$ and π is independent of $a^{(0)}$. Particularly,
  - choose $a^{(0)} = (1,0,0,\dots,0)$, then $\pi =$ the first row of $\lim_n \mathbf{P}^n$
  - choose $a^{(0)} = (0,1,0,\dots,0)$, then $\pi =$ the second row of $\lim_n \mathbf{P}^n$
  - ......
  - Each row of $\lim_n \mathbf{P}^n$ is exactly π.
- So, the limit $\lim\limits_{n\to\infty} \mathbf{P}^n$ must be the following very special rank-1 matrix whose each row is the liming distribution $\pi = [\pi_1, \pi_2, \dots, \pi_N]$.

$$\lim_n \mathbf{P}^n = \begin{bmatrix} \pi_1 & \pi_2 & \dots & \pi_N \\ \pi_1 & \pi_2 & \dots & \pi_N \\ \vdots & & & \\ \pi_1 & \pi_2 & \dots & \pi_N \end{bmatrix} = \begin{bmatrix} \pi \\ \pi \\ \vdots \\ \pi \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \pi = \mathbb{1}\pi$$

(note $\mathbb{1}$ is column vector, π is row vector.)

If the limiting distribution π uniquely exists, then

- Since $a^{(n)} = a^{(0)}\mathbf{P}^n$, then $\pi = a^{(0)}\lim_n \mathbf{P}^n$ and π is independent of $a^{(0)}$. Particularly,
  - choose $a^{(0)} = (1,0,0,\dots,0)$, then $\pi =$ the first row of $\lim_n \mathbf{P}^n$
  - choose $a^{(0)} = (0,1,0,\dots,0)$, then $\pi =$ the second row of $\lim_n \mathbf{P}^n$
  - ......
  - Each row of $\lim_n \mathbf{P}^n$ is exactly π.
- So, the limit $\lim\limits_{n\to\infty} \mathbf{P}^n$ must be the following very special rank-1 matrix whose each row is the liming distribution $\pi = [\pi_1, \pi_2, \dots, \pi_N]$.

$$\lim_n \mathbf{P}^n = \begin{bmatrix} \pi_1 & \pi_2 & \dots & \pi_N \\ \pi_1 & \pi_2 & \dots & \pi_N \\ \vdots & & & \\ \pi_1 & \pi_2 & \dots & \pi_N \end{bmatrix} = \begin{bmatrix} \pi \\ \pi \\ \vdots \\ \pi \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \pi = \mathbb{1}\pi$$

(note $\mathbb{1}$ is column vector, π is row vector.)

- The balancing equation $\pi = \pi\mathbf{P}$ has a unique non-negative solution.

How to use MATLAB to solve the balancing equation?

Example 2.15 (page 25).

```
>> P=[.2 .3 .5; .1 0 .9; .55 0 .45];
>> [V,D]=eig(P') %  solve the eigenvectors of transpose of P
V = %eigenvectors
    -0.5726      -0.62554                          -0.62554
    -0.17178     0.24327 -  0.4491i     0.24327 +   0.4491i
    -0.80164     0.38228 +  0.4491i     0.38228 -   0.4491i
D = %eigenvalues
          1              0                    0
          0         -0.175 - 0.32307i    -0.175 + 0.32307i
          0              0                    0
>> mu=V(:,1)./sum(V(:,1))  % normalise the first column
mu =
     0.37037
     0.11111
     0.51852
```

How to use MATLAB to solve the balancing equation?

Example 2.15 (page 25).

```
>> P=[.2 .3 .5; .1 0 .9; .55 0 .45];
>> [V,D]=eig(P') %  solve the eigenvectors of transpose of P
V = %eigenvectors
    -0.5726      -0.62554                          -0.62554
    -0.17178      0.24327 -  0.4491i      0.24327 +   0.4491i
    -0.80164      0.38228 +  0.4491i      0.38228 -   0.4491i
D = %eigenvalues
          1              0                      0
          0          -0.175 - 0.32307i      -0.175 + 0.32307i
          0              0                      0
>> mu=V(:,1)./sum(V(:,1))  % normalise the first column
mu =
     0.37037
     0.11111
     0.51852
```

try Example 2.16 by yourself in which no limiting distribution exists, but the balancing equation still has one unique solution.

application: Markov chain with discounted reward

Assume that (X_n) is a DTMC with the transition matrix $\mathbf{P}$. $0 < \alpha < 1$ is a constant. c is a scalar-valued function defined on the state space $S = \{1, \ldots, N\}$. Let

$$u(x) = \mathrm{E}\left[\sum_{n=0}^{\infty} \alpha^n c(X_n) | X_0 = x\right], \quad x \in S.$$

Then show that the vector $\boldsymbol{u}$ satisfies the linear equation

$$\boxed{\boldsymbol{u} = \boldsymbol{c} + \alpha \mathbf{P} \boldsymbol{u}}$$

where $\boldsymbol{u} = [u(1), \ldots, u(N)]^{\mathsf{T}}$ and $\boldsymbol{c} = [c(1), \ldots, c(N)]^{\mathsf{T}}$.

Proof: $\mathrm{E}\, c(X_n | X_0 = i) = \sum_{j \in S} c(j) \Pr(X_n = j | X_0 = i) = \sum_j (\mathbf{P}^n)_{ij} c(j) = \mathbf{P}^n \boldsymbol{c}$. Then $\boldsymbol{u} = \sum_{n=0}^{\infty} \alpha^n \mathbf{P}^n \boldsymbol{c} = (\mathbf{I} - \alpha \mathbf{P})^{-1} \boldsymbol{c}$, which is $\boldsymbol{u} + \alpha \mathbf{P} \boldsymbol{u} = \boldsymbol{c}$.

Occupancy Distribution: frequency of time when DTMC visited a given state

Definition

The occupancy distribution $v = [v_1, v_2, \cdots, v_N]$ is

$$v_j = \lim_{n\to\infty} \frac{\mathrm{E}\left(N_j^{(n)} \mid X_0 = i\right)}{n+1}.$$

(formally v_j depends on the initial i from this def.)

- Recall that $N_j^{(n)} = \sum_{t=0}^{n} 1_{\{X_t=j\}}$ and $m_{i,j}^{(n)} = \mathrm{E}(N_j^{(n)} | X_0 = i)$ is the occupancy time up to n of state j starting from state i.

Definition

The occupancy distribution $\nu = [\nu_1, \nu_2, \cdots, \nu_N]$ is

$$\nu_j = \lim_{n\to\infty} \frac{\mathrm{E}\left(N_j^{(n)} \mid X_0 = i\right)}{n+1}.$$

(formally ν_j depends on the initial i from this def.)

- Recall that $N_j^{(n)} = \sum_{t=0}^{n} 1_{\{X_t=j\}}$ and $m_{i,j}^{(n)} = \mathrm{E}(N_j^{(n)} | X_0 = i)$ is the occupancy time up to n of state j starting from state i.
- It is clear that $\sum_{j\in S} \nu_j = 1$ since $\sum_{j\in S} 1_{\{X_t=j\}} = 1$.

Definition

The occupancy distribution $\nu = [\nu_1, \nu_2, \cdots, \nu_N]$ is

$$\nu_j = \lim_{n\to\infty} \frac{\mathrm{E}\left(N_j^{(n)} \mid X_0 = i\right)}{n+1}.$$

(formally ν_j depends on the initial i from this def.)

- Recall that $N_j^{(n)} = \sum_{t=0}^{n} 1_{\{X_t=j\}}$ and $m_{i,j}^{(n)} = \mathrm{E}(N_j^{(n)}|X_0 = i)$ is the occupancy time up to n of state j starting from state i.
- It is clear that $\sum_{j\in S} \nu_j = 1$ since $\sum_{j\in S} 1_{\{X_t=j\}} = 1$.
- By [Thm 2.4] ($\mathbf{M}^{(n)} = \sum_{t=0}^{n} \mathbf{P}^t$), the connection to the transition matrix is

$$\nu_j = \lim_{n\to\infty} \frac{m_{i,j}^{(n)}}{n+1} = \lim_{n\to\infty} \frac{1}{n+1} \sum_{t=0}^{n} p_{ij}^{(t)}$$

where $p_{ij}^{(t)} = (\mathbf{P}^t)_{i,j}$ is the (i,j) of the t-step transition matrix.

Theorem (Thm 2.7)

If the occupancy distribution ν exists, then it satisfies the balance and normalizing equations:

$$\nu = \nu \mathbf{P}, \qquad \sum_{j \in S} \nu_j = 1.$$

*Proof:**

$$\begin{aligned}
\nu_j &= \lim_{n\to\infty} \frac{m_{i,j}^{(n)}}{n+1} = \lim_{n\to\infty} \frac{1}{n+1} \sum_{t=0}^{n} p_{i,j}^{(t)} \\
&= \lim_{n\to\infty} \frac{1}{n+1} \left(p_{ij}^{(0)} + \sum_{t=1}^{n} p_{i,j}^{(t)} \right) \\
&= \lim_{n\to\infty} \frac{1}{n+1} \left(p_{ij}^{(0)} + \sum_{t=1}^{n} \sum_{k\in S} p_{i,k}^{(t-1)} p_{k,j} \right) (\because \text{C-K eqn.}) \\
&= \lim_{n\to\infty} \frac{n}{n+1} \sum_{k\in S} \frac{1}{n} \sum_{t=0}^{n-1} p_{i,k}^{(t)} p_{k,j} \\
&= \sum_{k\in S} \lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} p_{i,k}^{(t)} p_{k,j} \\
&= \sum_{k\in S} \nu_k p_{k,j}
\end{aligned}$$

*textbook page 29.

Application: Section 2.6.2 Long-Run Expected Cost Rate

Let $c(x)$ be a cost function. The expected long-run cost rate is defined as

$$\mathscr{C}_i \triangleq \lim_{n\to\infty} \frac{1}{n+1} \mathrm{E}\left[\sum_{t=0}^{n} c(X_t)|X_0 = i\right].$$

Its calculation derived below leads to Theorem 2.12 (page 38). By Thm 2.11 and the definition of ν,

$$\begin{aligned}\mathscr{C}_i &= \lim_{n\to\infty} \sum_{j\in S} \frac{1}{n+1} c(j) m_{ij}^{(n)} \\ &= \sum_{j\in S} c(j)\nu_j \qquad \text{(indepenent of } i\text{)}\end{aligned}$$

which is the expectation of the function c under the occupancy distribution:

$$\boxed{\mathscr{C} = \mathrm{E}_\nu(c)}$$

Remark:* A stronger property ("strong ergodicity") is that: the empirical measure $\frac{1}{n+1}\sum_{t=0}^{n}\delta_{X_t}(\cdot)$ (which is *random*) weakly converges to a unique probability measure ν: as $n\to\infty$, $\frac{1}{n+1}\sum_{t=0}^{n} c(X_t) \to \mathrm{E}_\mu[c] := \sum_{x\in S} c(x)\nu(x)$.

---

* *optional*

Theories

Existence/Uniqueness of π, μ, ν ?

Theories

Existence/Uniqueness of π, μ, ν ?
depends on the structure of the transition diagram

- Why the very simple DTMC $\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ does not have a unique limiting distribution? The transient distribution $a^{(n)} = a^{(0)}\mathbf{I}^n = a^{(0)}$ is unchanged at all! state 1 does not talk to state 2; state 2 does not talk to state 1, either. — No communication, no mixing of the initial distribution.

- Why the very simple DTMC $\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ does not have a unique limiting distribution? The transient distribution $a^{(n)} = a^{(0)}\mathbf{I}^n = a^{(0)}$ is unchanged at all! state 1 does not talk to state 2; state 2 does not talk to state 1, either. — No communication, no mixing of the initial distribution.
- (unique) limiting distribution should *forget* any information of the initial state and uniquely exists: wherever the DTMC starts, the process should converge to the same destination.

- Why the very simple DTMC $\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ does not have a unique limiting distribution? The transient distribution $a^{(n)} = a^{(0)}\mathbf{I}^n = a^{(0)}$ is unchanged at all! state 1 does not talk to state 2; state 2 does not talk to state 1, either. — No communication, no mixing of the initial distribution.
- (unique) limiting distribution should *forget* any information of the initial state and uniquely exists: wherever the DTMC starts, the process should converge to the same destination.

- Why the very simple DTMC $\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ does not have a unique limiting distribution? The transient distribution $a^{(n)} = a^{(0)}\mathbf{I}^n = a^{(0)}$ is unchanged at all! state 1 does not talk to state 2; state 2 does not talk to state 1, either. — No communication, no mixing of the initial distribution.
- (unique) limiting distribution should *forget* any information of the initial state and uniquely exists: wherever the DTMC starts, the process should converge to the same destination.

Definition

A state j is said to be **accessible** from a state i (written $i \to j$) if there is an $n \geq 0$ such that

$$\Pr(X_n = j \,|\, X_0 = i) = (\mathbf{P}^n)_{ij} > 0.$$

A state i is said to **communicate** with state j (written $i \leftrightarrow j$) if both $i \to j$ and $j \to i$.

$i \to j$ means that it is possible (i.e. with nonzero probability) to go from i to j in one or more steps or, alternatively, there is a directed path from node i to node j in the transition diagram of the DTMC.

Properties of "$\leftrightarrow$" relationship

- $n \geq 0$ implies that n could be 0, thus $i \leftrightarrow i$ for any state i. (Reflexivity)

Properties of "$\leftrightarrow$" relationship

- $n \geq 0$ implies that n could be 0, thus $i \leftrightarrow i$ for any state i. (Reflexivity)
- if $i \leftrightarrow j$, then $j \leftrightarrow i$. (Symmetry)

Properties of "$\leftrightarrow$" relationship

- $n \geq 0$ implies that n could be 0, thus $i \leftrightarrow i$ for any state i. (Reflexivity)
- if $i \leftrightarrow j$, then $j \leftrightarrow i$. (Symmetry)
- if $i \rightarrow j$ and $j \rightarrow k$, then $i \rightarrow k$ (Transitivity), and if $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$. (exercise: prove this claim)

Properties of "$\leftrightarrow$" relationship

- $n \geq 0$ implies that n could be 0, thus $i \leftrightarrow i$ for any state i. (Reflexivity)
- if $i \leftrightarrow j$, then $j \leftrightarrow i$. (Symmetry)
- if $i \rightarrow j$ and $j \rightarrow k$, then $i \rightarrow k$ (Transitivity), and if $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$. (exercise: prove this claim)

Properties of "$\leftrightarrow$" relationship

- $n \geq 0$ implies that n could be 0, thus $i \leftrightarrow i$ for any state i. (Reflexivity)
- if $i \leftrightarrow j$, then $j \leftrightarrow i$. (Symmetry)
- if $i \to j$ and $j \to k$, then $i \to k$ (Transitivity), and if $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$. (exercise: prove this claim)

Remark: for each state i, we can define its **communicating class**:

$$[i] := \{j \in S : i \leftrightarrow j\}$$

Note $[i] = [j]$ if and only if $i \leftrightarrow j$.

Properties of "$\leftrightarrow$" relationship

- $n \geq 0$ implies that n could be 0, thus $i \leftrightarrow i$ for any state i. (Reflexivity)
- if $i \leftrightarrow j$, then $j \leftrightarrow i$. (Symmetry)
- if $i \to j$ and $j \to k$, then $i \to k$ (Transitivity), and if $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$. (exercise: prove this claim)

Remark: for each state i, we can define its **communicating class**:

$$[i] := \{j \in S : i \leftrightarrow j\}$$

Note $[i] = [j]$ if and only if $i \leftrightarrow j$. Therefore the relation "$\leftrightarrow$" is a called an **equivalence relation** and it induces a partition of the state space S into disjoint subsets $S_1, \ldots, S_K$ such that

$$S = S_1 \cup \ldots \cup S_K$$

and any two states communicate if and only if they are in the same subset. The sets $S_1, \ldots, S_K$ are called the communicating classes of the chain.
If there is only one communicating class, i.e., S itself, then we call this chain is **irreducible**; Otherwise, we call it **reducible** since it can be "reduced" to multiple communicating classes.

Definition

A DTMC $X_n, n \geq 0$ on state space $S = \{1, 2, \cdots, N\}$ is said to be **irreducible** if its *any* two states communicate, *i.e.*, $i \leftrightarrow j, \forall i, j \in S$

A DTMC that is not irreducible is called reducible.

A chain is irreducible if for any two states i, j, there exists an integer n(possibly depending on i and j) such that $(\mathbf{P}^n)_{ij} > 0$.

Definition

A DTMC $X_n, n \geq 0$ on state space $S = \{1, 2, \cdots, N\}$ is said to be **irreducible** if its *any* two states communicate, *i.e.*, $i \leftrightarrow j, \forall i, j \in S$
A DTMC that is not irreducible is called reducible.

A chain is irreducible if for any two states i, j, there exists an integer n(possibly depending on i and j) such that $(\mathbf{P}^n)_{ij} > 0$.

Theorem (Thm2.8, 2.9 (Ergodicity Theorems))

If a finite-state DTMC is irreducible, then

1. *its stationary distribution μ uniquely exists.*
2. *its occupancy distribution ν uniquely exists.*
3. *$\mu = \nu$, which is the unique solution of the balance equation (together with normalizing equation and non-negativity for probability vector).*
4. $\mu_i > 0, \forall i \in S$

Definition

A DTMC $X_n, n \geq 0$ on state space $S = \{1, 2, \cdots, N\}$ is said to be **irreducible** if its *any* two states communicate, *i.e.*, $i \leftrightarrow j, \forall i, j \in S$
A DTMC that is not irreducible is called reducible.

A chain is irreducible if for any two states i, j, there exists an integer n(possibly depending on i and j) such that $(\mathbf{P}^n)_{ij} > 0$.

Theorem (Thm2.8, 2.9 (Ergodicity Theorems))

If a finite-state DTMC is irreducible, then

1. *its stationary distribution μ uniquely exists.*
2. *its occupancy distribution ν uniquely exists.*
3. $\mu = \nu$*, which is the unique solution of the balance equation (together with normalizing equation and non-negativity for probability vector).*
4. $\mu_i > 0, \forall i \in S$

Remark: The last property is called positive recurrent, which may not hold true if the state space S is infinite.

Definition

A DTMC $X_n, n \geq 0$ on state space $S = \{1, 2, \cdots, N\}$ is said to be **irreducible** if its *any* two states communicate, *i.e.*, $i \leftrightarrow j, \forall i, j \in S$
A DTMC that is not irreducible is called reducible.

A chain is irreducible if for any two states i, j, there exists an integer n(possibly depending on i and j) such that $(\mathbf{P}^n)_{ij} > 0$.

Theorem (Thm2.8, 2.9 (Ergodicity Theorems))

If a <u>*finite-state*</u> *DTMC is irreducible, then*

1. *its stationary distribution* μ *uniquely exists.*
2. *its occupancy distribution* ν *uniquely exists.*
3. $\mu = \nu$, *which is the unique solution of the balance equation (together with normalizing equation and non-negativity for probability vector).*
4. $\mu_i > 0, \forall i \in S$

Remark: The last property is called positive recurrent, which may not hold true if the state space S is infinite.
Ergodicity : time average (occupancy distribution ν) = ensemble average (stationary distribution μ)

What about limiting distribution? : aperiodicity

- Why the *irreducible* DTMC $\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ still does not have a limiting distribution ? $(\mathbf{P}^{2n} = \mathbf{I}, \mathbf{P}^{2n+1} = \mathbf{P})$. state 1 and 2 do communicate. But state 1(or 2) returns to itself at every two steps.

What about limiting distribution? : aperiodicity

- Why the *irreducible* DTMC $\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ still does not have a limiting distribution ? $(\mathbf{P}^{2n} = \mathbf{I}, \mathbf{P}^{2n+1} = \mathbf{P})$. state 1 and 2 do communicate. But state 1(or 2) returns to itself at every two steps.
- This DTMC is periodic with period 2.

What about limiting distribution? : aperiodicity

- Why the *irreducible* DTMC $\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ still does not have a limiting distribution ? $(\mathbf{P}^{2n} = \mathbf{I}, \mathbf{P}^{2n+1} = \mathbf{P})$. state 1 and 2 do communicate. But state 1(or 2) returns to itself at every two steps.
- This DTMC is periodic with period 2.

What about limiting distribution? : aperiodicity

- Why the *irreducible* DTMC $\mathbf{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ still does not have a limiting distribution ? $(\mathbf{P}^{2n} = \mathbf{I}, \mathbf{P}^{2n+1} = \mathbf{P})$. state 1 and 2 do communicate. But state 1(or 2) returns to itself at every two steps.
- This DTMC is periodic with period 2.

Definition

A state i has period d if any return to state i must occur in multiples of d steps. Formally, the period of a state is defined as

$$d = \gcd\left\{n \geq 1 : \Pr(X_n = i | X_0 = i) > 0\right\}$$

(where "gcd" is the greatest common divisor). If $d = 1$, then the state i is said to be **aperiodic**.

Refer to Example 2.21, 2.22, page 31.

It is possible that

$$\{n \geq 1 : \Pr(X_n = i | X_0 = i) > 0\}$$

is an empty set! Consider the example $\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$. Then $\Pr(X_n = 1 | X_0 = 1) = 0$ for any $n \geq 1$ since $\mathbf{P}^n = \mathbf{P}$ for any $n \geq 1$. In this (degenerate) case of empty set, we say the period of the state 1 is 0.

*optional: Refer to details at `http://www.mathematik.uni-ulm.de/stochastik/lehre/ss06/markov/skript_engl/node12.html` (Thm 2.8,2.9 therein)

It is possible that

$$\{n \geq 1 : \Pr(X_n = i | X_0 = i) > 0\}$$

is an empty set! Consider the example $\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$. Then $\Pr(X_n = 1 | X_0 = 1) = 0$ for any $n \geq 1$ since $\mathbf{P}^n = \mathbf{P}$ for any $n \geq 1$. In this (degenerate) case of empty set, we say the period of the state 1 is 0.

It can be proved (by elementary number theory*) that

1. a state i is aperiodic if and only if there exists $0 \leq n < +\infty$ such that for **all** $k \geq n$, $\Pr(X_k = i | X_0 = i) > 0$, *i.e.*, $(\mathbf{P}^k)_{ii} > 0$.
2. If $i \leftrightarrow j$, then i and j have the same period.

*optional: Refer to details at `http://www.mathematik.uni-ulm.de/stochastik/lehre/ss06/markov/skript_engl/node12.html` (Thm 2.8,2.9 therein)

It is possible that

$$\{n \geq 1 : \Pr(X_n = i | X_0 = i) > 0\}$$

is an empty set! Consider the example $\mathbf{P} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$. Then $\Pr(X_n = 1 | X_0 = 1) = 0$ for any $n \geq 1$ since $\mathbf{P}^n = \mathbf{P}$ for any $n \geq 1$. In this (degenerate) case of empty set, we say the period of the state 1 is 0.

It can be proved (by elementary number theory*) that

1. a state i is aperiodic if and only if there exists $0 \leq n < +\infty$ such that for **all** $k \geq n$, $\Pr(X_k = i | X_0 = i) > 0$, *i.e.*, $(\mathbf{P}^k)_{ii} > 0$.
2. If $i \leftrightarrow j$, then i and j have the same period.

Definition

All states in one communicating class share the same period; This period is then called as the period of the communicating class. All states in an *irreducible* DTMC share the same period, which is also called as the period of this irreducible DTMC.

---

*optional: Refer to details at `http://www.mathematik.uni-ulm.de/stochastik/lehre/ss06/markov/skript_engl/node12.html` (Thm 2.8,2.9 therein)

proof:* (2).

- If $i = j$, the conclusion is trivial. Now we assume $i \neq j$.
  Denote $A_j := \{n \geq 1 : X_n = j | X_0 = j\}$. Let d_i, d_j be the period of state i, j, respectively. Since $i \leftrightarrow j$, then there exit two integers k_1, k_2 such that $(\mathbf{P}^{k_1})_{ij} > 0$ and $(\mathbf{P}^{k_2})_{ji} > 0$, and $k_1 \geq 1, k_2 \geq 1$ since $i \neq j$. This fact implies that $k_1 + k_2$ (or $k_2 + k_1$) belongs to both A_i and A_j. (why? C-K eqn.)
- We claim that if A_i is empty, then A_j is empty too. If not so, then there is $m \in A_j$, i.e., $(\mathbf{P}^m)_{jj} > 0$. So $(\mathbf{P}^{k_1+m})_{ii} \geq (\mathbf{P})^{k_1}_{ij}(\mathbf{P}^m)_{jj} > 0$, implying $m + k_1 \in A_i$ which is a contradiction.
- Now we assume neither of A_i nor A_i is empty. We shall show that any number in A_j is divisible by d_i. Then by definition of "gcd" for A_j, we know $d_i \leq d_j$. By symmetry, it is also true that $d_j \leq d_i$. So, $d_i = d_j$. Here is the detail.
  Pick up any number n from A_j, then $k_1 + n + k_2$ belongs to A_i (consider $i \to j \to j \to i$). So n must be divisible by d_i since both $k_1 + n + k_2$ and $k_1 + k_2$ are divisible by d_i (because they are in A_i).

* *optional*

Theorem (Theorem 2.10 (Unique Limiting Distribution))

A finite-state irreducible aperiodic DTMC has a unique limiting distribution.

Exercise

Consider the following DTMC: $S = \{1,2,3,4\}$ *and the transition matrix is*

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1/2 & 0 & 1/3 & 1/6 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}.$$

There are three communication classes $\{1,2\}$ $\{3\}$ and $\{4\}$. The solution of the balance equation is unique: $\mu = (0,0,1,0)$. The limit of $\mathbf{P}^n$ is $= \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$. So, the limiting distribution and the occupancy distribution are both unique and equal to μ. But this is a reducible chain. So, the above theorems we learnt are "sufficient" but not "necessary" conditions.

The trouble comes from the absorbing state 3, which is like a black hole.

Part iii: Reversible Markov chain and Introduction to Markov Chain Monte Carlo

MCMC* is one of the most important sampling methods with wide applications in many areas. It can be used to sample any distribution up to an unknown constant where the pdf is in the form of $\pi(x) = Z^{-1} f(x)$ where $Z = \int f(x) dx$ may be unknown.

The concept of reversibility is important in statistical physics and network sciences, and most physical models satisfy this condition.

The mechanism of why MCMC works is related to the limiting behavior of the reversible DTMC.

*The interested reader for this part can refer to the reference: *Understanding Markov Chains: Examples and Applications* by Nicolas Privault. Springer. 2013. §8.3

Definition (detailed balance)

DTMC $\{X_n\}$ with transition matrix $\mathbf{P} = [p_{ij}]$ is said to satisfy the *detailed balance condition* with respect to the probability distribution $\pi = (\pi_i)_{i \in S}$ satisfying $\pi_i > 0, \forall i$, if

$$\pi_i p_{ij} = \pi_j p_{ji}, \quad \forall i \in S, \; j \in S. \tag{1}$$

Take sum over j for the detail balance condition, then we have

Theorem

If the detailed balance condition holds for $\mathbf{P}$ *and* π, *then* π *is a stationary distribution for* $\mathbf{P}$, *i.e., the balancing equation* $\pi = \pi\mathbf{P}$ *holds.*

Understanding detailed balance condition – reversibility

We assume the chain initially is already in equilibrium, i.e., the distribution of X_0 is its stationary distribution π. Then we know X_n follows π, too. Note that

$$\pi_i p_{ij} = \Pr(X_n = i)\Pr\left(X_{n+1} = j \mid X_n = i\right) = \Pr(X_n = i, X_{n+1} = j)$$

So, the detailed balance (1) actually says that

$$\boxed{\Pr(X_n = i, X_{n+1} = j) \quad = \quad \Pr(X_n = j, X_{n+1} = i), \quad \forall i, j}$$

The joint distribution of (X_n, X_{n+1}) is symmetric:
the probability of the move from i to j is equal to the probability of the move from j to i.

Exercise

Show that the symmetric random walk on $S = \{1, 2, \ldots, N\}$ with periodic boundary condition is reversible.

Understanding detailed balance condition – reversibility

- Introduce a new matrix $\mathbf{P}^* = [p^*_{ij}]$ whose (i,j) entry is

$$\boxed{p^*_{ij} := p_{ji}\pi_j/\pi_i, \text{ or }, \quad \mathbf{P}^* := \mathbf{D}^{-1}\mathbf{P}^\mathsf{T}\mathbf{D}},$$

where $\mathbf{D} \triangleq \text{diag}\{\pi_1, \pi_2, \ldots, \pi_N\}$ is the diagonal matrix with diagonal entry $\{\pi_i\}$.

Definition

A **reversible** DTMC means a DTMC satisfying the detailed balance condition.

*Actually, this DTMC is the *time-reversed* chain $\left\{Y_n \triangleq X_{N-n}\right\}$ with initial distribution just being π for a fixed N. See Exercise 8.15 in reference of "Understanding...".

Understanding detailed balance condition – reversibility

- Introduce a new matrix $\mathbf{P}^* = [p^*_{ij}]$ whose (i,j) entry is

$$\boxed{p^*_{ij} := p_{ji}\pi_j/\pi_i, \text{ or },\quad \mathbf{P}^* := \mathbf{D}^{-1}\mathbf{P}^\mathsf{T}\mathbf{D}},$$

  where $\mathbf{D} \triangleq \operatorname{diag}\{\pi_1,\pi_2,\ldots,\pi_N\}$ is the diagonal matrix with diagonal entry $\{\pi_i\}$.
- Note that $\sum_j p^*_{ij} = (\pi\mathbf{P})_i/\pi_i = 1$ since π is the stationary measure of $\mathbf{P}$. So, for the transition matrix $\mathbf{P}$ and stationary measure π, which are associated with the chain $\{X_n : n = 0,1,\ldots\}$, the stochastic matrix $\mathbf{P}^*$ defines a new DTMC *. Show that π is also the stationary distribution for this new chain (exercise).

Definition

A **reversible** DTMC means a DTMC satisfying the detailed balance condition.

*Actually, this DTMC is the *time-reversed* chain $\left\{Y_n \triangleq X_{N-n}\right\}$ with initial distribution just being π for a fixed N. See Exercise 8.15 in reference of "Understanding...".

Understanding detailed balance condition – reversibility

- Introduce a new matrix $\mathbf{P}^* = [p^*_{ij}]$ whose (i,j) entry is

$$\boxed{p^*_{ij} := p_{ji}\pi_j/\pi_i, \text{ or }, \quad \mathbf{P}^* := \mathbf{D}^{-1}\mathbf{P}^{\mathsf{T}}\mathbf{D}},$$

  where $\mathbf{D} \triangleq \text{diag}\{\pi_1, \pi_2, \ldots, \pi_N\}$ is the diagonal matrix with diagonal entry $\{\pi_i\}$.
- Note that $\sum_j p^*_{ij} = (\pi\mathbf{P})_i/\pi_i = 1$ since π is the stationary measure of $\mathbf{P}$. So, for the transition matrix $\mathbf{P}$ and stationary measure π, which are associated with the chain $\{X_n : n = 0, 1, \ldots\}$, the stochastic matrix $\mathbf{P}^*$ defines a new DTMC *. Show that π is also the stationary distribution for this new chain (exercise).
- Clearly, the detailed balance condition is equivalent to $\boxed{\mathbf{P} = \mathbf{P}^*}$. So, the detail balance condition is also called "reversibility", which means the time-reversed chain has the same transition matrix as the original chain.

Definition

A **reversible** DTMC means a DTMC satisfying the detailed balance condition.

*Actually, this DTMC is the *time-reversed* chain $\left\{Y_n \triangleq X_{N-n}\right\}$ with initial distribution just being π for a fixed N. See Exercise 8.15 in reference of "Understanding...".

Understanding detailed balance condition – reversibility

- Introduce a new matrix $\mathbf{P}^* = [p^*_{ij}]$ whose (i,j) entry is

$$\boxed{p^*_{ij} := p_{ji}\pi_j/\pi_i, \text{ or }, \quad \mathbf{P}^* := \mathbf{D}^{-1}\mathbf{P}^\mathsf{T}\mathbf{D}},$$

  where $\mathbf{D} \triangleq \operatorname{diag}\{\pi_1, \pi_2, \ldots, \pi_N\}$ is the diagonal matrix with diagonal entry $\{\pi_i\}$.
- Note that $\sum_j p^*_{ij} = (\pi\mathbf{P})_i/\pi_i = 1$ since π is the stationary measure of $\mathbf{P}$. So, for the transition matrix $\mathbf{P}$ and stationary measure π, which are associated with the chain $\{X_n : n = 0, 1, \ldots\}$, the stochastic matrix $\mathbf{P}^*$ defines a new DTMC *. Show that π is also the stationary distribution for this new chain (exercise).
- Clearly, the detailed balance condition is equivalent to $\boxed{\mathbf{P} = \mathbf{P}^*}$. So, the detail balance condition is also called "reversibility", which means the time-reversed chain has the same transition matrix as the original chain.
- $f(x) := \sum_{i,j=1}^N x_i(\pi_i p_{ij} - \pi_i\delta_{ij})x_j$ is a quadratic form, which is called the Dirichlet form of the Markov chain.

Definition

A **reversible** DTMC means a DTMC satisfying the detailed balance condition.

*Actually, this DTMC is the *time-reversed* chain $\left\{Y_n \triangleq X_{N-n}\right\}$ with initial distribution just being π for a fixed N. See Exercise 8.15 in reference of "Understanding..".

MCMC is to generate random samples distributed on S according to an given distribution $\pi_i, i \in S$.

- When DTMC with transition matrix $\mathbf{P}$ is positive recurrent, aperiodic, and irreducible, then the limiting distribution uniquely exists and equals the stationary distribution π if the detailed balance condition holds for $\mathbf{P}$ and π. When DTMC satisfies these coditions, then it is called "**reversible**" Markvo chain.
- In general, it is easy to propose an irreducible Markov chain but its transition matrix $\hat{\mathbf{P}}$ may *not* satisfy the detailed balance condition for the given π.
- Metropolis (1953) modified the proposed $\hat{\mathbf{P}}$ to build a new DTMC whose transition matrix $\mathbf{P}$ as follows

$$p_{ij} := \hat{p}_{ij} \times \left(1 \wedge \frac{\pi_j \hat{p}_{ji}}{\pi_i \hat{p}_{ij}}\right) = \begin{cases} \hat{p}_{ji} \frac{\pi_j}{\pi_i}, & \text{if } \pi_j \hat{p}_{ji} < \pi_i \hat{p}_{ij}, \\ \hat{p}_{ij}, & \text{if } \pi_j \hat{p}_{ji} \geq \pi_i \hat{p}_{ij}. \end{cases}$$

  for $i \neq j$, and $\hat{p}_{ii} := 1 - \sum_{k \neq i} \hat{p}_{ik}$. "$\wedge$" means the minimum of two numbers.
- Homework Verify that $\mathbf{P} = [p_{ij}]$ given above is a stochastic matrix and satisfies the detailed balance condition with π.

Metropolis-Hastings algorithm*

The Markov chain generated from $\mathbf{P}$ can be implemented in two steps: Given current state $X_n = i$

- Draw j from the proposal distribution $\hat{p}_{ij}$.
- Accept $X_{n+1} = j$ with probability r_{ij} where

  $$r_{ij} = 1 \wedge \frac{\pi_j \hat{p}_{ji}}{\pi_i \hat{p}_{ij}}$$

  Otherwise, set $X_{n+1} = i$.
- Then the probability $\Pr\left(X_{n+1} = j \mid X_n = i\right) = \hat{p}_{ij} \times r_{ij} = p_{ij}$

*Ranked No. 1 in Top 10 algorithms in 20th century.
https://www.siam.org/pdf/news/637.pdf

Metropolis-Hastings algorithm*

The Markov chain generated from $\mathbf{P}$ can be implemented in two steps: Given current state $X_n = i$

- Draw j from the proposal distribution $\hat{p}_{ij}$.
- Accept $X_{n+1} = j$ with probability r_{ij} where

$$r_{ij} = 1 \wedge \frac{\pi_j \hat{p}_{ji}}{\pi_i \hat{p}_{ij}}$$

  Otherwise, set $X_{n+1} = i$.
- Then the probability $\Pr\left(X_{n+1} = j \mid X_n = i\right) = \hat{p}_{ij} \times r_{ij} = p_{ij}$

The convergence is guaranteed for any proposal density $\hat{\mathbf{P}} = [\hat{p}_{ij}]$. The convergence speed depends on π and $\hat{\mathbf{P}}$. A practical guidance for tune $\hat{\mathbf{P}}$ is to make sure the acceptance ratio r should not be too small (huge waste, low efficiency) or too large (too local, hard to sample other importance regions). One trade-off value is around 23.4%.

---

*Ranked No. 1 in Top 10 algorithms in 20th century.
`https://www.siam.org/pdf/news/637.pdf`

Homework

- Consider the two-state DTMC with the transition matrix $\mathbf{P} = \begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}$ $a, b \in [0,1]$. Discuss the reducibility, periodicity and limiting/stationary/occupancy distributions for different a, b.
- Prove the transition matrix used by Metropolis $[p_{ij}]$ satisfies the detailed balance condition with π.
- Define the matrix $\widetilde{\mathbf{P}}$ whose (i,j) entry is $\pi_i^{1/2} p_{ij} \pi_j^{-1/2}$, i.e., $\widetilde{\mathbf{P}} := \mathbf{D}^{1/2} \mathbf{P} \mathbf{D}^{-1/2}$, where $\mathbf{D} \triangleq \operatorname{diag}\{\pi_1, \pi_2, \ldots, \pi_N\}$. Prove that the detailed balance condition (for the chain $\mathbf{P}$) is equivalent to the symmetry of $\widetilde{\mathbf{P}}$: $\widetilde{\mathbf{P}}^\mathsf{T} = \widetilde{\mathbf{P}}$. So,if $\mathbf{P}$ is reversible, one can diagonalized $\widetilde{\mathbf{P}} = U^\mathsf{T} \Lambda U$ where U and Λ are real, then can you find the diagonalization of $\mathbf{P}$? All eigenvalues of the reversible chain are real.
- Let $\mathbf{P} = [p_{ij}]$ be the transition matrix of a reversible DTMC on $S = \{1, 2, \ldots, N\}$ with the invariant measure $\pi = (\pi_1, \ldots, \pi_N)$. Define a restricted DTMC on a subset of $S' \subset S$, whose transition matrix is denoted by $\mathbf{P}' = [p'_{ij}]$ for $i, j \in S'$, where $p'_{ij} := p_{ij}$ if $i \neq j$ and $p'_{ii} = p_{ii} + \sum_{k \notin S'} p_{ik}$. Is $\mathbf{P}'$ reversible ? What is its invariant measure if yes.
- [Textbook] page 55-57: 2.19, 2.20, 2.21, 2.28.

MATLAB command for eigenvectors of a matrix is : eigs(A)