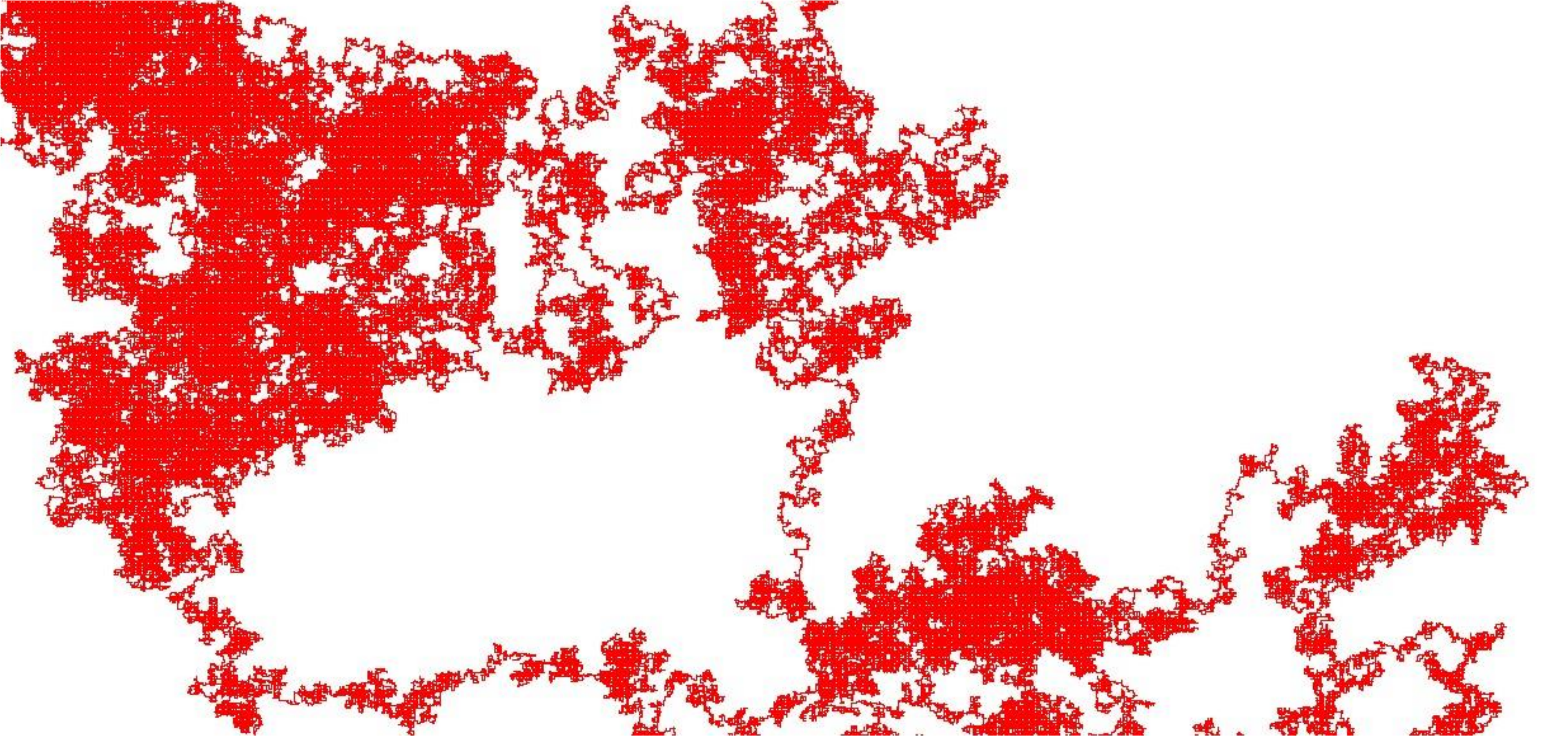
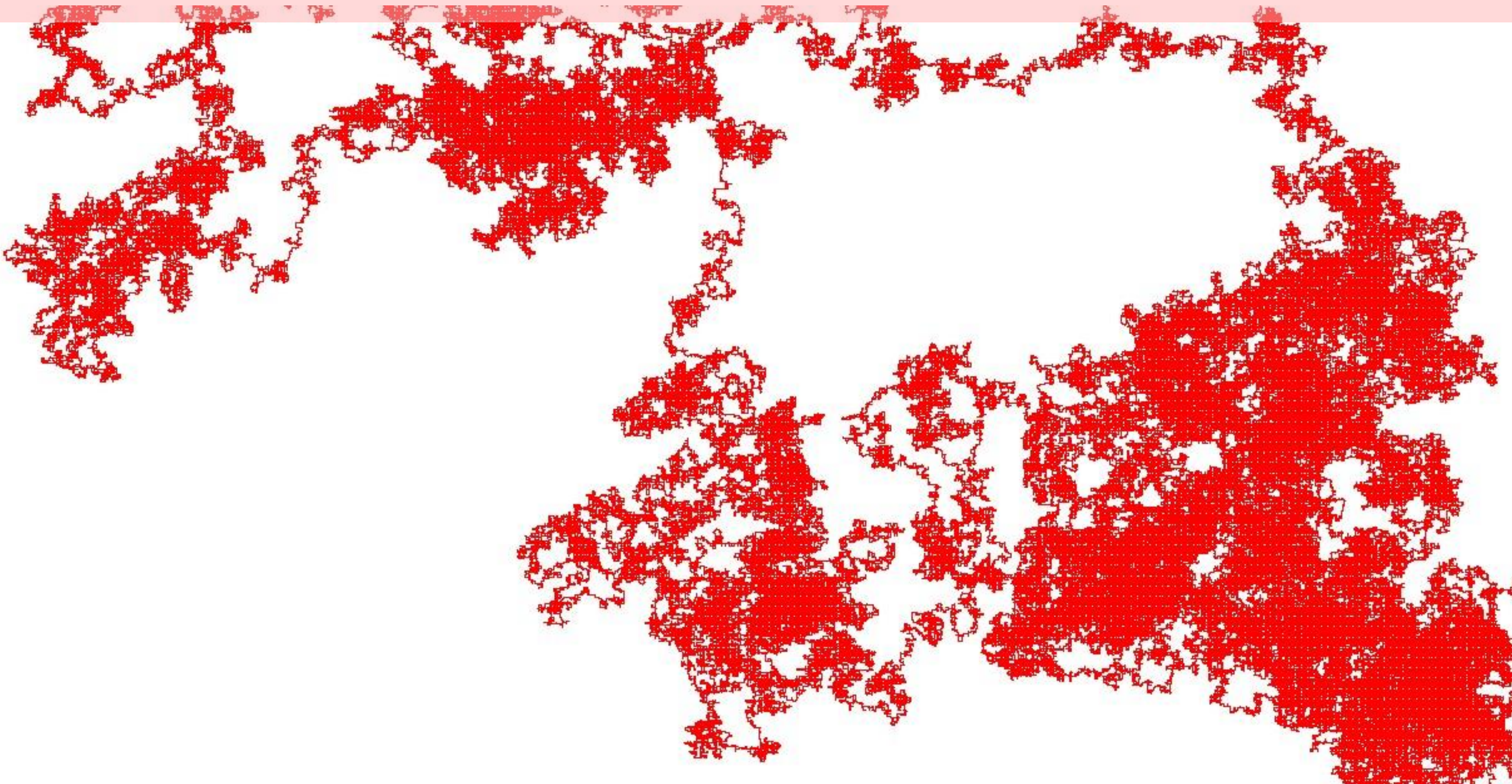


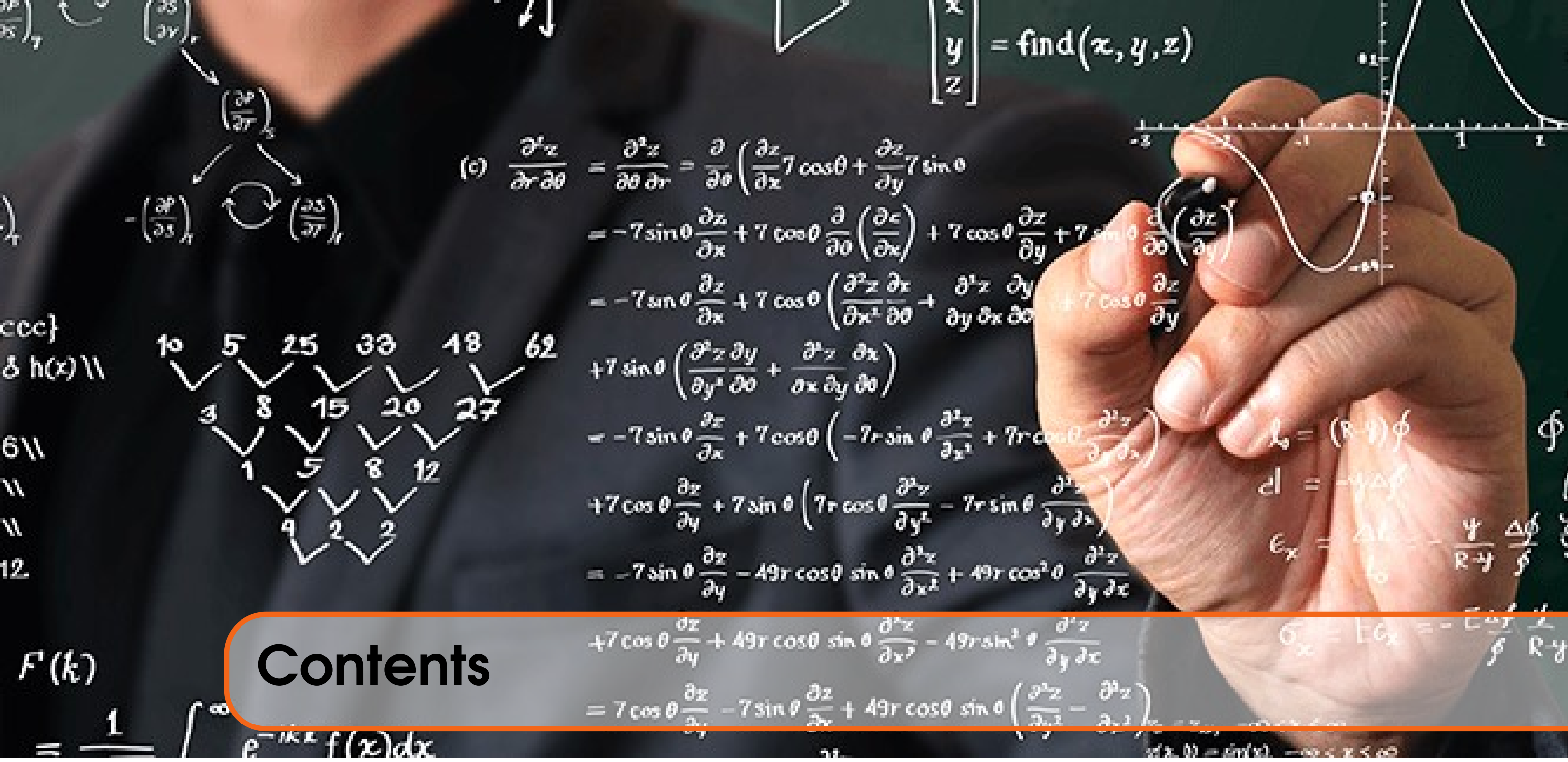
STAT 333 Course Notes

University of Waterloo

The One And Only
Waterloo 76er

Bill Zhuo





Contents

Basic Settings

Review & Introduction

0.1 General Outline

1. Bi-weekly homework (10%), Monday before tutorial starts
2. Assessments:
   (a) 6 Homework (drop lowest one) (10%)
   (b) 2 Midterms ($50\% = 2 \times 25\%$)
   (c) Final (40%)

0.2 Basic Settings

Definition 0.2.1 — Probability Space. A probability space consists of a triplet $(\Omega, \mathscr{E}, \mathbf{P})$ where

1. Ω is the **sample space**, the collection of all the possible outcomes of a random experiment
2. $\mathscr{E}$ is the **σ-algebra**, the collection of all the "events". An **event** E is a subset of Ω, for which we can talk about probability.
3. $\mathbf{P}$ is a **probability measure**, that is a set function (a mapping from event to a real number) $\mathbf{P} : \mathscr{E} \to \mathbb{R}$ with $\mathbf{P}(E)$ is the result. This mapping needs to satisfy the **probability axioms**:
   (a) $\forall E \in \mathscr{E}, 0 \leq \mathbf{P}(E) \leq 1$
   (b) $\mathbf{P}(\Omega) = 1$
   (c) For countably many disjoint events $E_1, E_2, \ldots$, we have

$$\mathbf{P}(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mathbf{P}(E_i)$$

Definition 0.2.2 — Random Variable. A random variable (r.v.) X is a mapping from Ω to $\mathbb{R}$.

$$\begin{aligned} X :&\Omega \to \mathbb{R} \\ &\omega \mapsto X(\omega) \end{aligned}$$

Definition 0.2.3 — Stochastic Processes.
1. **Process** is idea of change overtime
2. **Stochastic** is just the idea of randomness. The **Etymology** is from the ancient greek, stands for "aim at", "guess". This particular vocabulary had not been used until the beginning of the 20-th century by Khinchin. There is a political story behind this, the note will not go into details of this.
   This was later translated by Feller and Doob into English. And the rest is history.

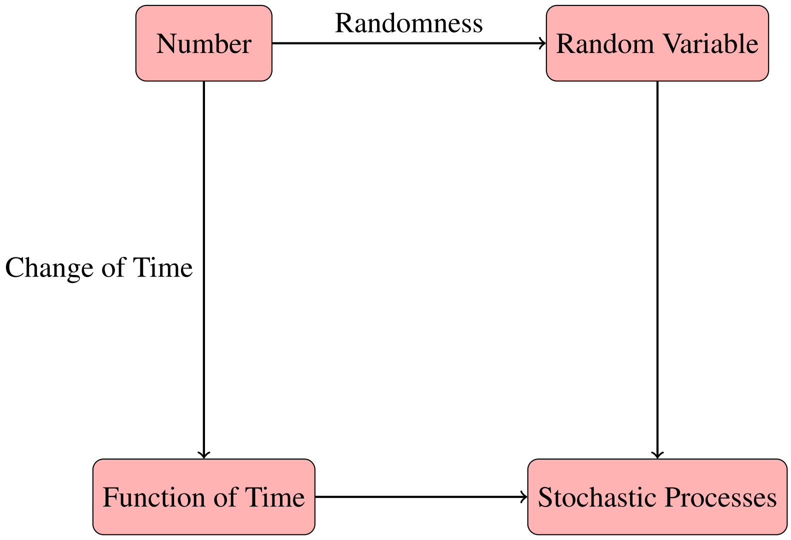


Figure 0.2.1: *Formulation of Stochastic Process Definition*

   (a) A sequence/family of random variables (simpler, so we take this as the definition)
   (b) A random function (hard to formulate)
3. **Formal Definition**: A stochastic process $\{X_t\}_{t\in T}$ is a collection of random variables defined on the common probability space, where T is an index set. In most cases, T corresponds to "time", which can be either discrete or continuous.
   (a) In discrete case, we rather write $\{X_n\}_{n=0,1,...}$

Definition 0.2.4 — States. The possible values of $X_t, t \in T$ are called the states of the process. There collection is called the **state space**, often denoted by S. The state space can either be discrete or continuous. The famous example of continuous state space is the **Brownian motion**. But in this course, we will be more focusing on the discrete state space so that we can relabel the states in S into the so-called **standardized state space**

$$S^* = \{0,1,2,\dots\} \longleftarrow \textbf{Countable State Space}$$

or

$$S^* = \{0,1,2,\dots,n\} \longleftarrow \textbf{Finite State Space}$$

■ **Example 0.1 — White Noise.** Let $X_0, X_1, \dots$ be i.i.d r.v.s following certain distribution. Then, $\{X_n\}_{n=0,1,2,...}$ is a stochastic process. This is also known as (**"White Noise"**). ■

■ **Example 0.2 — Simple Random Walk.** Let $X_1, X_2, \dots$ be i.i.d r.v.s. For each one of them

$$\begin{cases} \mathbf{P}(X_i = 1) = p \\ \mathbf{P}(X_i = -1) = 1-p \end{cases}$$

define $S_0 = 0$, for the other S_i, we have

$$S_n = \sum_{i=1}^{n} X_i$$

Then, $\{S_n\}_{n=0,1,2,...}$ is a stochastic process. With the state space,

$$\mathbf{S} = \mathbb{Z}$$

This $\{S_n\}_{n=0,1,2,...}$ is called a "simple random walk". Note that

$$S_n = \begin{cases} S_{n-1}+1 & \text{with probability } p \\ S_{n+1}-1 & \text{with probability } 1-p \end{cases}$$

■

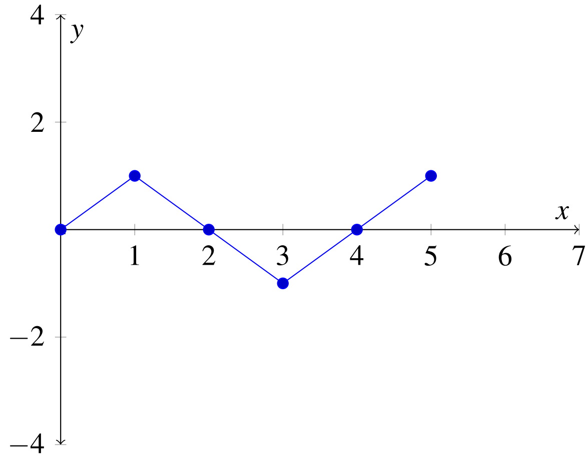


Figure 0.2.2: *Simple Random Walk*

1. **Q:** Now the question is why do we need the concept of stochastic process? Why don't we just look at the joint distribution of $(X_0, X_1, \dots, X_n)$?
2. **A:** The joint distribution of a large number of r.v.s is very complicated, because it does not take advantage of the special structure of T, which is time. For example, for simple random walk, the full distribution of $(S_0, S_1, \dots, S_n)$ is complicated for n large. However, the structure is actually simple if we focus on adjacent terms.

$$S_{n+1} = S_n + X_{n+1}$$

where S_n is the last value and X_{n+1} is a dirichlet type of output. Also, note that S_n and X_{n+1} are actually independent by definition. By introducing time into the framework, things can be often greatly simplified. More precisely, from this **simple random walk** example, we found that for $\{S_n\}_{n=0,1,2,...}$, if we know S_n, then the distribution of S_{n+1} will be depent on the history $S_i, \forall 0 \le i \le n-1$. This is a **very useful** property and it motivates the notion of **Markov Chain**.

II STAT 333 Main Part

1. (Discrete-Time) Markov Chains

1.1 Review on Conditional Probability

Definition 1.1.1 — Conditional Probability. The conditional probability of an event given an event A with $\mathbf{P}(A) > 0$ is given by

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

Theorem 1.1.1 Let $A_1, A_2, \ldots$ be disjoint events such that

$$\bigcup_i A_i = \Omega$$

then we have

1. **Law of Total Probability**:

$$\mathbf{P}(B) = \sum_{i=1}^{\infty} \mathbf{P}(B|A_i) \cdot \mathbf{P}(A_i)$$

2. **Bayes' Rule**:

$$\mathbf{P}(A_i|B) = \frac{\mathbf{P}(B|A_i) \cdot \mathbf{P}(A_i)}{\sum_{j=1}^{\infty} \mathbf{P}(B|A_j) \cdot \mathbf{P}(A_j)}$$

1.2 Markov Chains (DTMC)

Definition 1.2.1 — Discrete-Time Markov Chain (DTMC). A discrete-time stochastic process $\{X_n\}_{n=0,1,2,\ldots}$ is called a discrete-time Markov Chain (DTMC) with transition matrix

$$P = [P_{ij}]_{i,j \in S}$$

if for any $j, i, i_{n-1}, \ldots, i_0 \in S$,

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

R In a more general setting, the Markov property can be defined as

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

The previous definition only depends on i, j, but this one depends on i, j and time n.

What does the Markov property tells us?

Note that X_n is the present states and $X_{n-1}, \ldots, X_0$ are the historical states, and X_{n+1} is about the future. The Markov property means the history has no impact on the future, all we care is the present. In other words, the past will only influence the future through current state.

> [Yi Shen] "Given the present state, the history and the future are independent."

In our definition, we focus on the **time-homogeneous** case, i.e

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

this one does not depend on time n (probability structure is time-invariant).

1.3 Transitional Matrix

Proposition 1.3.1 — Transition Matrix Properties P. We have a corresponding transition matrix $P = \{P_{ij}\}_{i,j \in S}$

1. $$P_{ij} \geq 0, \forall i, j \in S$$

2. Row sums of P are always 1:

$$\sum_{j \in S} P_{ij} = 1, \forall i \in S$$

   The reason is

$$\sum_{j \in S} P_{ij} = \sum_{j \in S} \mathbf{P}(X_{n+1} = j | X_n = i) = \mathbf{P}(X_{n+1} \in S | X_n = i) = 1$$

3. Matrix form:

$$\begin{pmatrix} & & \\ & P_{ij} & \\ & & \end{pmatrix}$$

   the row index i gives the **current state** and the column index j gives the **next state**.

■ **Example 1.1 — Simple Random Walk Revisit.** We have not yet shown **simple random walk** is DTMC. **Left as an exercise**.

Exercise 1.1 Show that the simple random walk is a DTMC. ■

$$
\begin{aligned}
P_{i,i+1} &= \mathbf{P}(S_1 = i+1 | S_0 = i) \\
&= \mathbf{P}(X_1 = 1 | S_0 = i) && \textbf{equivalent to first move is } 1 \\
&= \mathbf{P}(X_1 = 1) && X_1 \perp\!\!\!\perp S_0 \\
&= p
\end{aligned}
$$

$$
\begin{aligned}
P_{i,i-1} &= \mathbf{P}(S_1 = i-1 | S_0 = i) \\
&= \mathbf{P}(X_1 = -1 | S_0 = i) && \textbf{equivalent to first move is } 1 \\
&= \mathbf{P}(X_1 = -1) && X_1 \perp\!\!\!\perp S_0 \\
&= 1-p
\end{aligned}
$$

Thus, $P_{ij} = 0$ for all $j \neq i \pm 1$. The transition matrix P is infinite-dimensional

$$
P = \begin{pmatrix}
\ddots & \ddots & & & & \\
& 0 & p & & & \\
& q & 0 & p & & \\
& & q & 0 & p & \\
& & & \ddots & \ddots & \ddots
\end{pmatrix}
$$

■

■ **Example 1.2 — Ehrenfest's Urn.** Two urns A, B, total M balls. Each time, pick one ball at random uniformly and leave it to the opposite urn.

1. Let X_n be the number of balls in A after step n.
2. The state space $S = \{0, \dots, M\}$.
3. $X_0 = i$ means i balls in A and $M - i$ balls in B

$$
P_{ij} = \mathbf{P}(X_1 = j | X_0 = i) = \begin{cases} \frac{i}{M} & j = i-1 \textbf{ (The ball is from } A\textbf{)} \\ \frac{M-i}{M} & j = i+1 \textbf{ (The ball is from } B\textbf{)} \\ 0 & j \neq i \pm 1 \end{cases}
$$

then, the transition matrix is

$$
P = \begin{pmatrix}
0 & 1 & & & & \\
\frac{1}{M} & 0 & \frac{M-1}{M} & & & \\
& \frac{2}{M} & 0 & & & \\
& & & 0 & \ddots & \\
& & & \ddots & \ddots & \frac{1}{M} \\
& & & & 1 & 0
\end{pmatrix}
$$

■

1.4 C-K Equation

Theorem 1.4.1 — Chapman-Kolmogorov Equation (C-K Equation).
The question is that we are given the 1-step transition matrix $P = \{P_{ij}\}_{i,j \in S}$. How can we decide the n-step transition probabilities? (with time-homogeneity)

$$
P_{ij}^{(n)} := \mathbf{P}(X_n = j | X_0 = i) = \mathbf{P}(X_{m+n} = j | X_m = i), m = 0, 1, 2, \dots
$$

Let us start with 2 step transition matrix

$$P^{(2)} = \left\{P_{ij}^{(2)}\right\}_{i,j\in S}$$

Condition on what happens at time 1.

$$P_{ij}^{(2)} = \mathbf{P}(X_2 = j|X_0 = i) = \sum_{k\in S} \mathbf{P}(X_2 = j|X_0 = i, X_1 = k)\mathbf{P}(X_1 = k|X_0 = i) \qquad (*)$$

this is nothing else but the **conditional version** of the law of total probability. Details are as follow

$$\begin{aligned}
\mathbf{P}(X_2 = j|X_0 = i) &= \sum_{k\in S} \mathbf{P}(X_2 = j, X_1 = k|X_0 = i) \\
&= \sum_{k\in S} \frac{\mathbf{P}(X_2 = j, X_1 = k, X_0 = i)}{\mathbf{P}(X_0 = i)} \\
&= \sum_{k\in S} \frac{\mathbf{P}(X_2 = j, X_1 = k, X_0 = i)}{\mathbf{P}(X_1 = k, X_0 = i)} \cdot \frac{\mathbf{P}(X_1 = k, X_0 = i)}{\mathbf{P}(X_0 = i)} \\
&= \sum_{k\in S} \mathbf{P}(X_2 = j|X_0 = i, X_1 = k)\mathbf{P}(X_1 = k|X_0 = i)
\end{aligned}$$

Now, we have $(*)$ proved. We have not yet used Markov property! By Markov property, $X_0 \perp\!\!\!\perp X_2$

$$(*) \to P_{ij}^{(2)} = \sum_{k\in S} \mathbf{P}(X_2 = j|X_1 = k)\mathbf{P}(X_1 = k|X_0 = i) = \sum_{k\in S} P_{kj}P_{ik} = \sum_{k\in S} P_{ik}P_{kj} = (P^2)_{ij}$$

Or we can interpret

$$P_{ij}^{(2)} = [P]_i^T [P]_j$$

Thus, $P^{(2)} = P^2$. Using the same idea, for $n, m = 0, 1, 2, \dots$, apply $X_0 \perp\!\!\!\perp X_m$

$$\begin{aligned}
P_{ij}^{(n+m)} &= \mathbf{P}(X_{n+m} = j|X_0 = i) \\
&= \sum_{k\in S} \mathbf{P}(X_{n+m} = j|X_m = k, X_0 = i) \cdot \mathbf{P}(X_m = k|X_0 = i) \\
&= \sum_{k\in S} \mathbf{P}(X_{n+m} = j|X_m = k) \cdot \mathbf{P}(X_m = k|X_0 = i) \\
&= \sum_{k\in S} P_{ik}^{(m)} P_{kj}^{(n)} \\
&= (P^{(m)} \cdot P^{(n)})_{ij} \to P^{(n+m)} = P^{(n)}P^{(m)} \qquad \Longleftarrow \textbf{C-K Equation}
\end{aligned}$$

We can also write the entry form $P_{ij}^{(m+n)} = \sum_{k\in S} P_{ik}^{(m)} P_{kj}^{(n)}$.
By definition,

$$\begin{aligned}
&P^(1) = P \\
&P^{(2)} = P^{(1)}P^{(1)} = P^2 \\
&P^{(3)} = P^{(1)}P^{(2)} = P^3 \\
&\vdots \\
&P^{(n)} = P^n
\end{aligned}$$

The left hand side is the n-step transition matrix

$$P^{(n)} = \left\{P_{ij}^{k}\right\}_{i,j\in S} \to P_{ij}^{n} = \mathbf{P}(X_n = j|X_0 = i)$$

while the right hand side is the n-th power of the one-step transition matrix

$$P^n = P \cdot P \ldots P$$

R

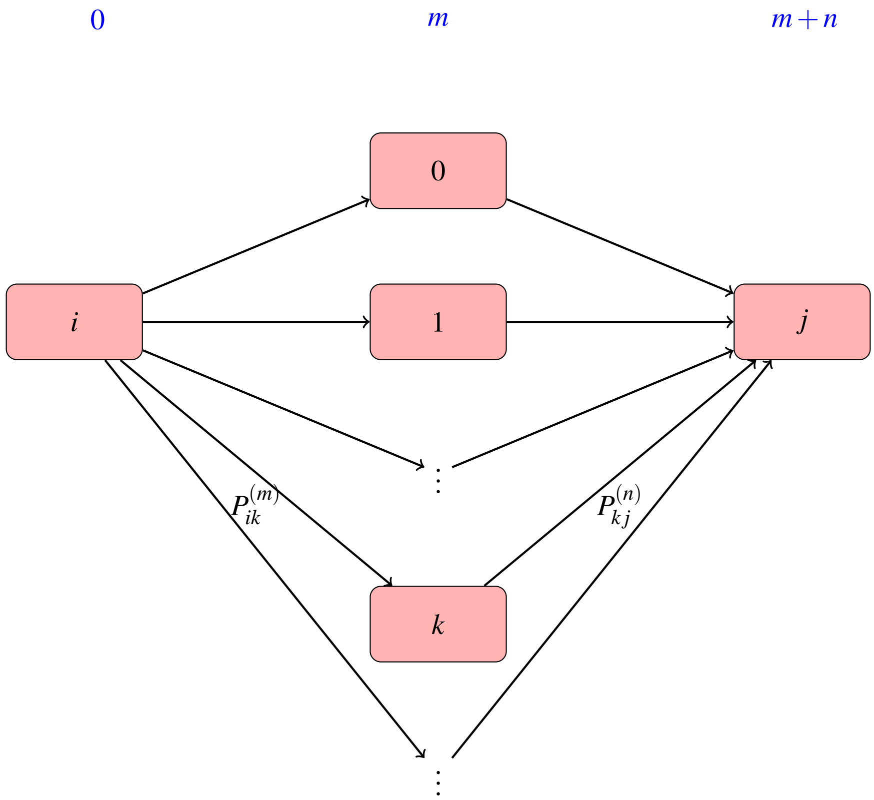


Figure 1.4.1: *Intuition of C-K Equation*

[Yi Shen]“Condition at timee m (on X_m) and sum up all the possibilities"

1.4.1 Visualization of Markov Chain

Markov chains can be represented by weighted directed graphs.

1. We define states to be nodes
2. Possible (one-step) transitions to be directed edges
3. Draw an edge from i to j if and only if $P_{ij} > 0$
4. The transition probabilities to be the weights of the edges

■ **Example 1.3 — Markov Chain Visualization.** With $S = \{0,1,2\}$

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \\ & & 1 \\ \frac{2}{5} & \frac{3}{5} & \end{pmatrix}$$

■

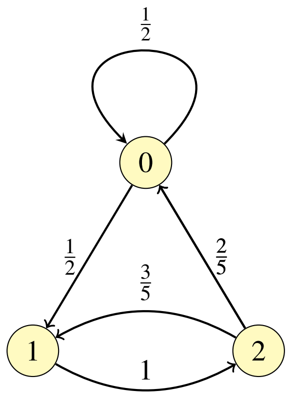


Figure 1.4.2: *Visualization*

■ **Example 1.4 — Simple Random Walk Revisit Again.** Another example for Markov Chain visualization with a graph representation

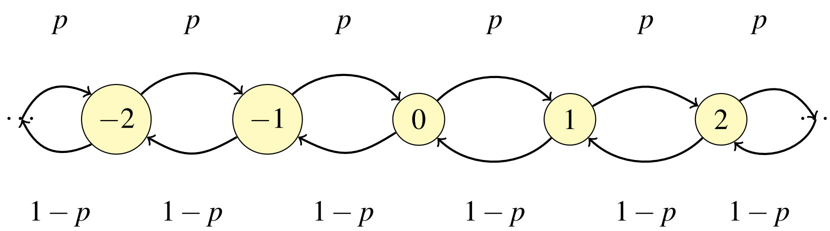


Figure 1.4.3: *Random Walk Visualization*

1.4.2 Vector-Matrix Form

What is the distribution of X_n?
So far, we have seen transition probabilities

$$P_{ij}^{(n)} = \mathbf{P}(X_n = j | X_0 = i)$$

if this DTMC starts from state i for sure (i.e. $\mathbf{P}(X_0 = i) = 1$), then $\left(P_{ij}^{(n)}\right)_j$ is the also the distribution of X_n at time n. That is the row i of $P^{(n)}$ is the distribution of X_n if the chain starts from state i.

$$P^{(n)} = \begin{array}{c} \\ i \end{array} \begin{pmatrix} - & - & - \end{pmatrix}$$

what if the chain has a random starting state, i.e, X_0 is random? ■

Definition 1.4.1 — Initial Distribution. Let $\mu = (\mu(0), \mu(1), \dots)$, where $\mu(i) = \mathbf{P}(X_0 = i)$. The row vector μ is called the initial distribution of this Markov Chain.

R This is the initial distribution of the initial state X_0.

Definition 1.4.2 — μ_n. Similarly, we can define $\mu_n = (\mu_n(0), \mu_n(1), \dots)$ to be the distribution of X_n, where $\mu_n(i) = \mathbf{P}(X_n = i)$. In this notation, we think of this as a **row vector** representing a distribution. Sometimes, written as $\mu_n(X_n = i)$. While in this notation, we think of it as a probability.

R We note that $\mu = \mu_0$.

Proposition 1.4.2 — Property of μ_n. The row vector μ_n represents a distribution, hence we have

1. $\mu_n(i) \geq 0, \forall i$
2. $$\sum_{i \in S} \mu_n(i) = 1 = \mu_n \cdot \mathbb{I}'$$

Q: $\mu_n =$? given μ and P

Theorem 1.4.3 — A Piece of Useful Fact.

$$\mu_n = \mu P^n$$

Proof. For any $j \in S$,

$$\begin{aligned}
\mu_n(j) &= \mathbf{P}(X_n = j) \\
&= \sum_{i \in S} \mathbf{P}(X_n = j | X_0 = i)\mathbf{P}(X_0 = i) && \textbf{Law of Total Probability} \\
&= \sum_{i \in S} \mu(i) P_{ij}^{(n)} && \textbf{A Bunch of Definitions} \\
&= \left[\mu P^{(n)}\right]_j \\
&= \left[\mu P^{(n)}\right]_j && \textbf{C-K}
\end{aligned}$$

Thus,

$$\mu_n = \mu P^n$$

■

Note that μ_n is the distribution of X_n, a row vector.
μ is the initial distribution, a row vector.
P^n is the transition matrix.

R The distribution of a DTMC is completely determined by two things
1. Initial distribution μ
2. Transition matrix P

1.5 Expected Value of A Function of X_n

1.5.1 Review of Conditional Expectation

Definition 1.5.1 — Conditional Distribution. Let X, Y be discrete random variables. Then, the conditional distribution of X given $Y = y$ is given by

$$\mathbf{P}(X = x|Y = y) = \frac{\mathbf{P}(X = x, Y = y)}{\mathbf{P}(Y = y)}, \forall x \in \mathscr{A}_{\S} \text{ and } \mathbf{P}(Y = y) \neq 0$$

where $\mathscr{A}_{\S}$ is the support of X and denote

$$\mathbf{f}_{X|Y=y}(x) = \mathbf{P}(X = x|Y = y) = \mathbf{f}_{X|Y}(x|y)$$

the **conditional probability mass function**.

Proposition 1.5.1 — Conditional pmf is a legitimate pmf. Given any y such that $\mathbf{P}(Y = y) \neq 0$,

1. $$\mathbf{f}_{X|Y=y}(x) \geq 0, \forall x \in \mathscr{A}_{\S}$$
2. $$\sum_{x \in \mathscr{A}_{\S}} \mathbf{f}_{X|Y=y}(x) = 1$$

Continuous case is similar.

R Since the conditional pmf is a legitimate pmf, the conditional distribution is also a legitimate probability distribution. (Potentially different from the unconditional distribution) As a result, we can define expectation on the conditional distribution.

Definition 1.5.2 — Conditional Expectation. Let X, Y be discrete random variables and g is a continuous function on X. Then, the conditional expectation of $g(X)$ given $Y = y$ is defined as

$$\mathbb{E}(g(X)|Y = y) = \sum_{x \in \mathscr{A}_{\S}} g(x)\mathbf{P}(X = x|Y = y)$$

R Fixed y, the conditional expectation is nothing else but the expectation under the conditional distribution.

Different Ways to Understand Conditional Expectations (Important!)

1. **Value:** Fix a value y, $\mathbb{E}(g(X)|Y = y)$ is a number
2. **Function:** As y changes,
$$h(y) = \mathbb{E}(g(X)|Y = y)$$
is a function of y.
3. **Random Variable:** since Y is a random variable, we can define
$$\mathbb{E}(g(X)|Y) = h(Y)$$
thus, $\mathbb{E}(g(X)|Y)$ is a function of Y. Hence, also a random variable. Then,
$$\mathbb{E}(g(X)|Y)(\omega) = \mathbb{E}(g(X)|Y = Y(\omega)), \omega \in \Omega$$
this random variable takes value $\mathbb{E}(g(X)|Y = y)$ when $Y = y$.

Proposition 1.5.2 — Properties of Conditional Expectation.

1. **Linearity**: inherited from expectation
$$\mathbb{E}(aX+b|Y=y) = a\mathbb{E}(X|Y=y)+b$$
and
$$\mathbb{E}(X+Z|Y=y) = \mathbb{E}(X|Y=y)+\mathbb{E}(Z|Y=y)$$
2. **Plug-in Property**
$$\mathbb{E}(g(X,Y)|Y=y) = \mathbb{E}(g(X,y)|Y=y) \neq \mathbb{E}(g(X,Y))$$
Proof. **Discrete case**
$$\begin{aligned}
&\mathbb{E}(g(X,Y)|Y=y) = \sum_{x_i}\sum_{y_j} g(x_i,y_j)\mathbf{P}(X=x_i,Y=y_j|Y=y) \\
&\to \sum_{x_i}\sum_{y_i} g(x_i,y_j)\mathbf{P}(X=x_i,Y=y_j|Y=y_j) = \begin{cases} 0 & y_i = y \\ \frac{\mathbf{P}(X=x_i,Y=y_j)}{\mathbf{P}(Y=y)} = \mathbf{P}(X=x_i|Y=y) & y_j = y \end{cases} \\
&\to \mathbb{E}(g(X,Y)|Y=y) = \sum_{x_i} g(x_i,y)\mathbf{P}(X=x_i|Y=y) \\
&= \mathbb{E}(g(X,y)|Y=y) \qquad\qquad g(X,y) \text{ is a function of } X
\end{aligned}$$
■

In particular, for $y \in \mathscr{A}_\dagger$
$$\mathbb{E}(g(X)h(Y)|Y=y) = h(y)\mathbb{E}(g(X)|Y=y)$$
in random variable form,
$$\mathbb{E}(g(X)h(Y)|Y) = h(Y)\mathbb{E}(g(X)|Y)$$
3. **Independence:** If $X \perp\!\!\!\perp Y$, then $\mathbb{E}(g(X)|Y) = \mathbb{E}(g(X))$

Proof. Due to independence, the conditional distribution is the same as the unconditional distribution ■
4. **Law of Iterated Expectation**
$$\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X)$$
Alert: $\mathbb{E}(X|Y)$ is a random variable, as a function of Y

Proof. **Discrete case:**
when $Y=y_j$, then random variable
$$\mathbb{E}(X|Y) = \mathbb{E}(X|Y=y_j) = \sum_{x_i} x_i\mathbf{P}(X=x_i|Y=y_j)$$

This happens with probability $\mathbf{P}Y=y_j$. Thus, we have the following
$$\begin{aligned}
\mathbb{E}(\mathbb{E}(X|Y)) &= \sum_{y_j}\mathbb{E}(X|Y=y_j)\mathbf{P}(Y=y_j) \\
&= \sum_{y_j}\left(\sum_{x_i} x_i\mathbf{P}(X=x_i|Y=y_j)\right)\mathbf{P}(Y=y_j) \\
&= \sum_{x_i} x_i \sum_{y_j}\mathbf{P}(X=x_i|Y=y_j)\mathbf{P}(Y=y_j) \\
&= \sum_{x_i} x_i\mathbf{P}(X=x_i) \qquad\qquad \textbf{Law of Total Probability} \\
&= \mathbb{E}(X)
\end{aligned}$$

■

■ **Example 1.5** Let Y be the number of claims received by an insurance company, X is some random parameter, such that

$$Y|X \sim \textbf{Poi}(X), X \sim \textbf{Exp}(\lambda)$$

what is $\mathbb{E}(Y)$?
Solution: $\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X))$, since $Y|X \sim \textbf{Poi}(X)$, then $\mathbb{E}(Y|X = x) = x \to \mathbb{E}(Y|X) = X$. Thus, $\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(X) = \frac{1}{\lambda}$ ■

1.5.2 Expectation of $f(X_n)$

Let $f : S \to \mathbb{R}$ given by $i \mapsto f(i)$. We want to know the expectation of $f(X_n)$.

Proposition 1.5.3 — Two Approaches.

1. **Method 1**:

$$\begin{aligned}\mathbb{E}(f(X_n)) &= \sum_{i\in S} f(i)\mathbf{P}(X_n = i) \\ &= \sum_{i\in S} f(i)\mu_n(i) \\ &= \mu_n f'\end{aligned}$$

where

$$f' = \begin{pmatrix} f(0) \\ f(1) \\ \vdots \\ f(i) \\ \vdots \end{pmatrix}$$

is the column vector giving all the values of f on different states. Then,

$$\begin{aligned}\mathbb{E}(f(X_n)) &= \mu_n f' \\ &= \mu P^n f'\end{aligned}$$

where
(a) μ is the row vector, initial distribution
(b) P^n is the transition matrix
(c) f' is the column vector, function of states

R [what happens if we calculate $P^n f'$ first?] This corresponds to finding $\mathbb{E}(f(X_n)|X_0 = i), i \in S$ first

2. **Method 2**:

$$\begin{aligned}
\mathbb{E}\left(f(X_n)\right) &= \mathbb{E}\left(\mathbb{E}\left(f(X_n)|X_0\right)\right) \\
&= \sum_{i\in S} \mathbb{E}\left(f(X_n)|X_0=i\right)\mathbf{P}(X_0=i) \\
&= \sum_{i\in S} \mathbb{E}\left(f(X_n)|X_0=i\right)\mu(i) \\
&= \sum_{i\in S} f^{(n)}(i)\mu(i) \\
&= \mu f^{(n)'} = \mu \begin{pmatrix} f^{(n)}(0) \\ f^{(n)}(1) \\ \vdots \\ f^{(n)}(i) \\ \vdots \end{pmatrix} \\
&= \mu P^n f'
\end{aligned}$$

Note that $\mathbb{E}\left(f(X_n)|X_0=i\right)$ is a function X_0 denoted by $f^{(n)}(X_0)$.

R [How to find $f^{(n)'}$?]

$$\begin{aligned}
f^{(n)'}(i) &= \mathbb{E}\left(f(X_n)|X_0=i\right) \\
&= \sum_j f(j)\mathbf{P}(X_n=j|X_o=i) \\
&= \sum_j P^n_{ij} f(j) \\
&= \left(P^n f'\right)_i
\end{aligned}$$

Thus, $f^{(n)'} = P^n f'$

This agrees with what we get in **Method 1**.

R Hence, if we calculate $P^n f'$ first, what we get is

$$f^{(n)'} = \begin{pmatrix} \mathbb{E}\left(f(X_n)|X_0=0\right) \\ \mathbb{E}\left(f(X_n)|X_0=1\right) \\ \vdots \\ \mathbb{E}\left(f(X_n)|X_0=i\right) \\ \vdots \end{pmatrix}$$

which is a column vector represeneting $f^{(n)}(i) = \mathbb{E}\left(f(X_n)|X_0=i\right)$

R Let's think about the interpretation

$$\mathbb{E}\left(f(X_n)\right) = \mu P^n f'$$

1. **Method 1**: $\mathbb{E}\left(f(X_n)\right) = \mu_n f'$ go directly to the states of time n, get a row vector first.
2. **Method 2**: $\mathbb{E}\left(f(X_n)\right) = \mu\mathbb{E}\left(f(X_n)\right)$, standing still at time 0, but use $f^{(n)'}$ to project the future expected states, get the column vector first.

In any case, row vectors are **distributions** $(\mu, \mu_1, \ldots, \mu_n, \ldots)$ while column vectors are **fucntions** $\left(f', f^{(n)'}, \ldots\right)$

1.6 Stationary Distribution

Definition 1.6.1 — Stationary Distribution. A probability distribution $\pi = (\pi_0, \pi_1, \dots)$ is called a stationary distribution (invariant distribution) of the DTMC $\{X_n\}_{n=0,1,\dots}$ with transition matrix P if

1. $\pi = \pi P$ as a system of equations
2. $\sum_{i \in S} \pi_i = 1$ by the definition of probability distribution

1.6.1 Why such π is called stationary?

Assume thee DTMC starts with initial distribution $\mu = \pi$. Then, the distribution of X_1 is

$$\mu_1 = \mu P = \pi P = \pi = \mu$$

the distribution of X_2:

$$\mu_2 = \mu P^2 = \pi P P = \pi = \mu$$

Thus, $\mu_n = \mu$. If the markov chain starts from a stationary distribution, its distribution will never change (stationary/invariant).

■ **Example 1.6** An electron has two states: ground state $|0\rangle$ and excited state$|1\rangle$. Let $X_n = \{0, 1\}$ be the state at time n. At each step, the electron changes state with probability α if it is in state $|0\rangle$, with probability β if it is in state $|1\rangle$.
Then, $\{X_n\}$ is a DTMC. Its transitional matrix is

$$\begin{array}{cc} & \begin{array}{cc} 0 & \quad 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \end{array}$$

solve for stationary distribution (s)

Solution: $\pi = \pi P$

$$\begin{pmatrix} \pi_0 & \pi_1 \end{pmatrix} \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} = \begin{pmatrix} \pi_0 & \pi_1 \end{pmatrix}$$

$$\rightarrow \begin{cases} \pi_0(1-\alpha) + \pi_1 \beta = \pi_0 & (1) \\ \pi_0 \alpha + \pi_1(1-\beta) = \pi_1 & (2) \end{cases}$$

Even though we have 2 equations with 2 unknowns. However, note that they are not linearly independent. Subtracting (1) from indentity $\pi_0 + \pi_1 = \pi_1 + \pi_0$, gives (2). Hence, (2) is redundant. From (1), we have $\alpha \pi_0 = \beta \pi_1 \rightarrow \frac{\pi_0}{\pi_1} = \frac{\beta}{\alpha}$. Now, we need $\pi_0 + \pi_1 = 1$. Thus,

$$\pi_0 = \frac{\beta}{\alpha + \beta} \qquad \pi_1 = \frac{\alpha}{\alpha + \beta}$$

Thus, there exists a unique stationary distribution $\pi = \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix}$ ■

R The stationary distribution example from before is typical.

1. Use $\pi = \pi P$ to get proportions between different components of π
2. Use $\pi \cdot \mathbf{1}' = 1$ to normalize and get exact values

1.6.2 Existence & Uniqueness of π? Convergence to π?

We can note that $\pi = \pi P$ implies that π is the transpose of an eigenvector of P with eigenvalue of 1.

1.7 Recurrence & Transience

Definition 1.7.1 — First Revisit Time. Let $y \in S$ be a state. Define

$$T_y = \min\{n \geq 1 : X_n = y\}$$

as the time of the first (re)visit to y. And we define

$$\rho_{yy} = \mathbf{P}_y(T_y < \infty) = \mathbf{P}(T_y < \infty | X_0 = y)$$

Definition 1.7.2 — Recurrent & Transient State. A state $y \in S$ is called recurrent, if $\rho_{yy} = 1$, which means always return to y.
It is called transient if $\rho_{yy} < 1$, which means there is a pass probability that the chain never visits y again.

R Note that

$$1 - \rho_{yy} = \mathbf{P}_y(P_y = \infty) > 0$$

■ **Example 1.7** Consider the following transition matrix

$$P = \begin{array}{c c} & \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \end{array} & \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \\ & \frac{1}{2} & \frac{1}{2} \\ & & 1 \end{pmatrix} \end{array}$$

its related graph representation is given below Given $X_0 = 0$, we have

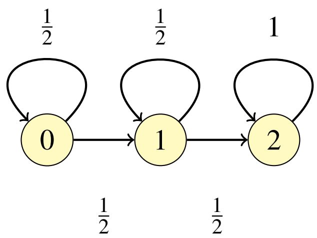


Figure 1.7.1: *Markov Chain Graph Representation*

$$\mathbf{P}(X_1 = 0 | X_0 = 0) = \mathbf{P}(X_1 = 1 | X_0 = 0) = \frac{1}{2}$$

Note that $X_1 = 0$ means $T_0 = 1$ and $X_1 = 1$ means $T_0 = \infty$ since 1 and 2 will never go to 0. This analysis implies that

$$\rho_{00} = \mathbf{P}_0(T_0 < \infty) = \frac{1}{2} < 1$$

Thus, by definition, state 0 is transient. By similar deduction, we have state 1 is also transient. Given $X_0 = 2$, we have

$$\mathbf{P}(X_1 = 2 | X_0 = 2) = 1, X_1 = 2 \equiv T_2 = 1 \rightarrow \rho_{22} = \mathbf{P}_2(T_2 < \infty) = 1$$

thus, state 2 is recurrent. ■

R This is an example for which recurrence/transience can be directly checked by definitions. However, this is very rare, as the distribution of T_i is very hard to derive in general. Thus, we need better criteria for recurrence/transience.

Definition 1.7.3 — x communicates to y. Let $x, y \in S$ (x, y can be the same state). x is said to communicate to y, denoted by $x \to y$, means it starting from x, the probability that the chain eventually (re)visits state y is positive. i.e,

$$\rho_{xy} = \mathbf{P}_x(T_y < \infty) > 0$$

Note that this is equivalent to say

$$\exists n \geq 1, \ni P^n_{xy} > 0$$

or say "x can go to y".

Lemma 1.8 — Transitivity of Communication. If $x \to y, y \to z$, then $x \to z$.

Proof. Note that

$$x \to y \Longrightarrow \exists M \geq 1 \ni P^M_{xy} > 0 \qquad y \to z \Longrightarrow \exists N \geq 1 \ni P^N_{yz} > 0$$

then, by C-K Equation, we have

$$P^{M+N}_{xz} = \sum_{k \in S} P^M_{xk} P^N_{kz} \geq P^M_{xy} P^N_{yz} > 0 \Longrightarrow x \to z$$

this is true since $P^M_{xy} P^N_{yz}$ just specifies one $M+N$-long path from x to z via y. While the quantity P^{M+N}_{xz} captures all possible paths. ■

Theorem 1.8.1 If $\rho_{xy} > 0$ but $\rho_{yx} < 1$, then x is transient.

Proof. Define $\kappa = \min\left\{k : P^k_{xy} > 0\right\}$ being the smallest length of a path from x to y. Since $P^k_{xy} > 0$, there are states

$$y_1, \ldots, y_{\kappa-1} \longrightarrow P_{xy_1} \cdot P_{y_1 y_2} \cdot \cdots \cdot P_{y_{\kappa-1} y} > 0$$

Note that none of $y_1, \ldots, y_{\kappa-1}$ is x, since otherwise, this is not the shortest path. Now, we have

$$\mathbf{P}_x(T_x = \infty) \geq P_{xy_1} \cdot P_{y_1 y_2} \cdot \cdots \cdot P_{y_{\kappa-1} y}(1 - P_{yx})$$

Where $P_{xy_1} \cdot P_{y_1 y_2} \cdot \cdots \cdot P_{y_{\kappa-1} y}$ is the path going from x to y without returning to x and $(1 - P_{yx})$ corresponds to the idea of once in y, never goes back to x. This quantity is larger than 0 since $\rho_{yx} < 1$. These two together describes one way not to visit x ever again via y. Thus,

$$\rho_{xx} = P_x(T_x < \infty) < 1$$

This means x is transient. ■

Corollary 1.8.2 If x is recurrent, and $\rho_{xy} > 0$, then $\rho_{yx} = 1$. (Result of contrapositive)

1.8.1 Communicating Class

Definition 1.8.1 — Communicating Class. A set of states $C \subseteq S$ is called a communicating class, if it satisfies the following properties

1. $\forall i, j \in C, i \to j, j \to i$
2. $\forall i \in C, j \notin C, i \not\to j$ or $j \not\to i$.

The idea is

> "States in the same class communicate with each other, states in different classes do not communicates in both ways"

R Communication and communicating classes are graphs. $i \to j$ means can go from i to j by following the arrows (directed edges).

How to Find classes: "find the loops"

■ **Example 1.8**

$$\begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & & \\ \frac{1}{2} & \frac{1}{2} & & \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} \\ & & & 1 \end{pmatrix} \end{array}$$

Note that

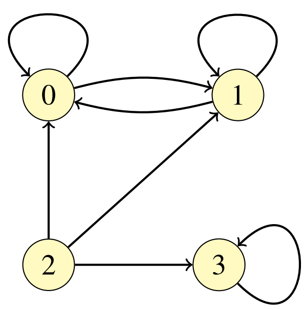


Figure 1.8.1: *Markov Chain Graph Representation*

$$\rho_{01} > 0, \rho_{10} > 0 \Longrightarrow 0, 1 \in \text{ same communication class}$$

State 2 is not in any class since $\rho_{i2} = 0, \forall i \in S$.

State 3 is a class on its own as $3 \to 3$ but $\rho_{3i} = 0, \forall i \in S$.

Then, there are two recurrent classes, $\{0,1\}, \{3\}$, with one transient state 2. And 2 does not below to any class. ■

■ **Example 1.9** We start the graphical representation:

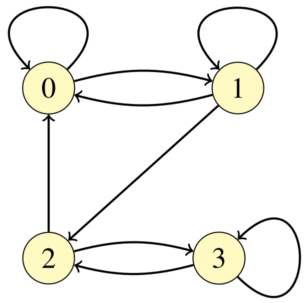
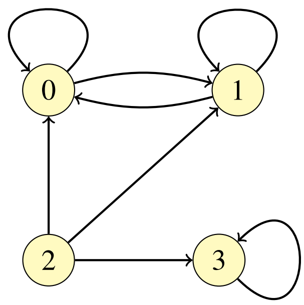


Figure 1.8.2: *Markov Chain Graph Representation*

Note that $\rho_{01}, \rho_{12}, \rho_{20} > 0 \implies$ 0,1,2 are in the same class.
$\rho_{23}, \rho_{32} > 0 \implies$ 2,3 are in the same class. Then, by transivity, we have $0, 1, 2, 3$ are all in the same class. ■

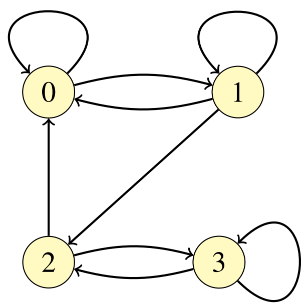
Definition 1.8.2 — Irreducile Markov Chain. A Markov Chain is called irreducible if all the states are in the same class. In other words, $i \leftrightarrow j$, for all $i, j \in S$.
A set B is called irreducible if $i \leftrightarrow j, \forall i, j \in B$.

■ **Example 1.10** From a previous example:



This is not a irreducible Markov Chain. ■

■ **Example 1.11** For the second example we have just seen,



This is indeed an irreducible Markov Chain. ■

Theorem 1.8.3 Let $i, j \in C$ be in the same communicating class. Then j is recurrent/transient if and only if i is recurrent/transient.

"Recurrence/Transience are class properties."

Proof. The proof will be included later. ■

R As a result of the theorem. We can call a class recurrent/transient, if all its states are recurrent/transient.

(This is equivalent to one state in the class if recurrent/transient.)

In order to decide if a class is recurrent/transient, we just need to check an element is recurrent/transient is enough.

Definition 1.8.3 — Closed Set. A set A is called closed if $i \in A, j \notin A$ implies $P_{ij} = 0$ (not from i to j in one step). Equivalently, this is to say, if $i \in A$ and $j \notin A$, then $i \not\rightarrow j$ (not from i to j in no-matter-how-many steps).

"Cannot get out once the chain goes into A."

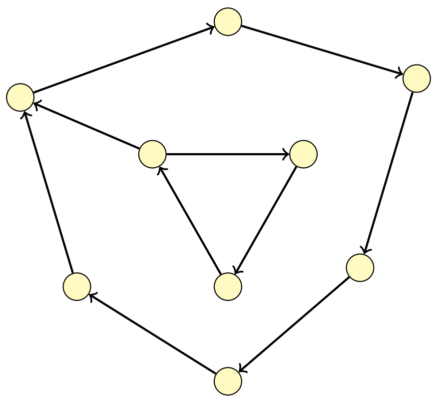


Figure 1.8.3: *Once get into the outer class, cannot get out*

Theorem 1.8.4 — Decomposition of the State Space. The state space S can be written as a disjoint union

$$S = T \cup R_1 \cup R_2 \cup \ldots$$

where T is the set of all transient states (T is not necessarily one class), and $R_i, i = 1, 2, \ldots$ are closed recurrent classes. Equivalently, in other textbook, we can say irreducible sets of recurrent states.

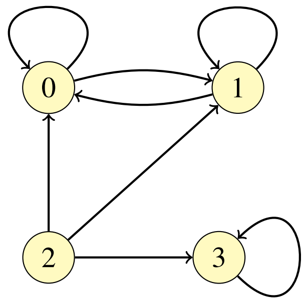
(This proof is trivial if we can recognize communicating classes are equivalence classes with the equivalence relation $\leftrightarrow$, two-way communication)

Proof. First wee collect all the transient states and put them into T. For each recurrent states, note that it must belong to at least one class, since it communicates with itself. We collect one class for each state, remove the identical classes, and get $\{R_k\}_{k=1,2,\ldots}$. Also, since recurrence is a class property, $R_k, k = 1, 2, \ldots$ are all recurrent classes. By construction, we have

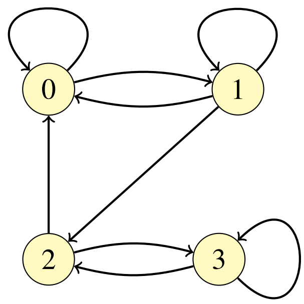
$$S = T \cup R_1 \cup R_2 \cup \ldots$$

1. **Disjoint:** We still need to show $R_i \cap R_j = \emptyset, \forall i \neq j \in \{1,2,\dots\}$.
   For the sake of contradiction, suppose $\exists i \in S$ and $i \in R_k, i \in R_{k'}$. But then, this means $\forall j \in R_k, j' \in R_{k'}$, we have
   $$i \leftrightarrow j, i \leftrightarrow j' \implies j' \leftrightarrow j'$$
   Thus, j, j' are in the same class. Since j and j' are arbitrary states in R_k and $R_{k'}$. We must have $R_k = R_{k'}$ contradicting the construction that $R_k \neq R_{k'}$ for $k \neq k'$.
2. **Closeness:** for the sake of contradiction, say R_k is not closed, then there exists $i \in R_k$ and $j \notin R_k$ such that $P_{ij} > 0 \to i \to j$. Or equivalently, we can say $\rho_{ij} > 0$. As $i \notin R_k, i \not\leftrightarrow j$, which implies that $j \not\to i$, then $\rho_{ji} = 0 < 1$. Thus, i is a transient state, which yields a contradiction.

■



■ **Example 1.12** We have $T = \{2\}$, $R_1 = \{0,1\}, R_2 = \{3\}$, so $S = T \cup R_1 \cup R_2$ ■



■ **Example 1.13** Only one recurrence class $S = R = \{0,1,2,3\}$ ■

1.8.2 Strong Markov Property

Recall that $T_y = \min\{n \geq 1 : X_n = y\}$.

Theorem 1.8.5 — Strong Markov Property for (time-homogeneous) MC. The process $\{X_{T_y+k}\}_{k=0,1,2,\dots}$ behaves like the MC with initial state y. (forget about the history and restart at state y)

Proof. It suffices to show that for $T = T_y$

$$\mathbf{P}(X_{T+1} = z | X_T = y, T = n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = P_{yz}$$

for all $n, x_{n-1}, \dots, x_1 \neq y$.

This is obvious since

$$
\begin{aligned}
&\mathbf{P}(X_{T+1} = z | X_T = y, T = n, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) \\
&= \mathbf{P}(X_{n+1} = z | X_n = y, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) && \textbf{Conditioning on } T = n \textbf{ and } T = T_y \\
&= \mathbf{P}(X_{n+1} = z | X_n = y) = P_{yz} && \textbf{Markov Property}
\end{aligned}
$$

■

1.8.3 Alternative Characterizations for Recurrence/Transience

Definition 1.8.4 Define

$$T_y^k = \min\left\{ n > T_y^{k-1}, X_n = y \right\}$$

time of the $k-$th visit to state y.

By **Strong Markov Property**,

$$\mathbf{P}_y(T_y^k < \infty) = \rho_{yy}^k$$

the LHS means revisiting y for at least k times, while the ρ_{yy} means revisit y for the first time.

Two Possibilities:

1. $\rho_{yy} < 1$ and y is transient, then $\rho_{yy}^k \to 0$ as $k \to \infty$, where

$$\mathbf{P}_y(\text{visits } y \text{ for infinite number of times}) = 0$$

$$\iff \mathbf{P}_y(\text{there is a last visit tp } y) = 1$$

2. $\rho_{yy} = 1$ and y is recurrent, then $\rho_{yy}^k = 1$ for all k, then

$$\mathbf{P}_y(\text{visits } y \text{ for infinite number of times}) = 1$$

Indeed, we know more. Let $N(y)$ be the total number of visits to state y. Then,

$$
\begin{aligned}
\mathbf{P}_y(N(y) \geq k) = \mathbf{P}_y(T_y^k < \infty) &= \rho_{yy}^k \\
\implies \mathbf{P}_y(N(y) \geq k+1) &= \rho_{yy}^{k+1} \\
\implies \mathbf{P}_y(N(y) \leq k) &= 1 - \rho_{yy}^{k+1}
\end{aligned}
$$

This is the cdf of $\mathbf{Geo}(1-\rho_{yy})$ and

$$N(y)|_{X_0=y} \sim \mathbf{Geo}(1-\rho_{yy})$$

1. Expectation can be used to characterize transience/recurrence:

$$\mathbb{E}_y N(y) = \frac{1}{1-\rho_{yy}} - 1 = \frac{\rho_{yy}}{1-\rho_{yy}}$$

then

$$
\begin{aligned}
\rho_{yy} = 1 &\implies \mathbb{E}_y N(y) = \infty \\
\rho_{yy} < 1 &\implies \mathbb{E}_y N(y) < \infty
\end{aligned}
$$

More generally, we have the following lemma

Lemma 1.9

$$\mathbb{E}_x N(y) = \frac{\rho_{xy}}{1-\rho_{yy}}$$

Proof.

$$\begin{aligned}
\mathbb{E}_x N(y) &= \mathbf{P}_x(\text{the chain ever visits } y)\mathbb{E}_x(N(y)|T_y<\infty) \\
&+ \mathbf{P}_x(\text{the chain never visits } y)\cdot 0 \\
&= \mathbf{P}_x(T_y<\infty)\mathbb{E}_x(N(y)|T_y<\infty) \\
&= \rho_{xy}\mathbb{E}_y(N(y)) && \textbf{Strong Markov Property} \\
&= \frac{\rho_{xy}}{1-\rho_{yy}}
\end{aligned}$$

■

Lemma 1.10

$$\mathbb{E}_x N(y) = \sum_{n=1}^{\infty} P_{xy}^n$$

Proof.

$$N(y) = \sum_{n=1}^{\infty} \mathbf{1}_{\{X_n=y\}} \Longrightarrow \mathbb{E}_x N(y) = \sum_{n=1}^{\infty} \mathbf{E}_x(\mathbf{1}_{\{X_n=y\}})$$
$$= \sum_{n=1}^{\infty} \mathbf{P}_x(X_n=y) = \sum_{n=1}^{\infty} P_{xy}^n$$

■

R Combine this lemma with the previous result that

$$\rho_{yy}=1 \Longrightarrow \mathbb{E}_y N(y)=\infty$$
$$\rho_{yy}<1 \Longrightarrow \mathbb{E}_y N(y)<\infty$$

we have the following theorem

Theorem 1.10.1 y is recurrent if and only if $\mathbb{E}_y(N(y)) = \sum_{n=1}^{\infty} P_{yy}^n = \infty$ diverges to infinity

1.10.1 Short Review on Indicator

Let A be an event, then

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

and we have

$$\mathbb{E}(\mathbf{1}_A) = \mathbf{P}(A)$$

Theorem 1.10.2 Let $i, j \in C$ be in the same communicating class. Then j is recurrent/transient if and only if i is recurrent/transient.

"Recurrence/Transience are class properties."

Proof. Now, we can do this proof. Suppose x, y are in the same class, which means $x \leftrightarrow y$ and x is recurrent. Since $x \to y$ and $y \to x$, there exists $M, N\mathbb{N}$ such that

$$P_{xy}^{(M)} > 0 \qquad\qquad P_{yx}^{(N)} > 0$$

Note that

$$\begin{aligned}
&\mathbf{P}(X_{M+N+k} = y | X_0 = y) = P_{yy}^{M+N+k} \geq P_{yx}^{(N)} P_{xx}^{(k)} P_{xy}^{(M)} \\
&= \mathbf{P}(X_{M+N+k} = y, X_{N+k} = x, X_n = x | X_0 = y) \\
&\implies \sum_{l=1}^{\infty} P_{yy}^{(l)} \geq \sum_{l=M+N+1}^{\infty} P_{yy}^{(l)} = \sum_{k=1}^{\infty} P_{yy}^{M+N+k} && \text{Let } k = l - M - N \\
&\geq \sum_{k=1}^{\infty} P_{yx}^{(N)} P_{xx}^{(k)} P_{xy}^{(M)} = P_{yx}^{(N)} P_{xy}^{(M)} \sum_{k=1}^{\infty} P_{xx}^{k} = \infty && x \text{ is recurrent}
\end{aligned}$$

Thus, y is recurrent. Therefore, x recurrent implies y recurrent and recurrence is a class property. This also implies that transience is a class property as well. ■

Lemma 1.11 In a finite closed set, there has to be at least one recurrent state.

Proof. For the sake of contradiction, suppose all states in C are transient. Then, for any two $x, y \in C$

$$\begin{aligned}
&\mathbb{E}_x(N(y)) = \frac{\rho_{xy}}{1 - \rho_{yy}} < \infty \\
&\implies \sum_{y \in C} \mathbb{E}_x(N(y)) < \infty
\end{aligned}$$

However,

$$\begin{aligned}
\sum_{y \in C} \mathbb{E}_x(N(y)) &= \mathbb{E}_x\left(\sum_{y \in C} N(y)\right) \\
&= \mathbb{E}_x\left(\sum_{y \in C} \sum_{n=1}^{\infty} \mathbf{1}_{\{X_n = y\}}\right) \\
&= \mathbb{E}_x\left(\sum_{n=1}^{\infty} \sum_{y \in C} \mathbf{1}_{\{X_n = y\}}\right)
\end{aligned}$$

For each n, exactly one indicator takes value 1, the rest ones are 0, this implies

$$\sum_{y \in C} \mathbf{1}_{\{X_n = y\}}) = 1$$

then, we have

$$\mathbb{E}_x\left(\sum_{n=1}^{\infty} \sum_{y \in C} \mathbf{1}_{\{X_n = y\}}\right) = \mathbb{E}_x\left(\sum_{n=1}^{\infty} 1\right) = \infty$$

This yields a contradiction. Hence, there must exist at least one recurrent state ■

Combine this result with the fact that recurrence/transience are class properties, we have

Theorem 1.11.1 A **finite** closed class must be recurrent. In particular, an irreducible Markov Chain with finite state space must be recurrent.

1.12 Existence of a Stationary Distribution

In this part, we show that an irreducible and recurrent DTMC "almost" has a stationary distribution. If the state space is finite, then it has a stationary distribution.

Definition 1.12.1 — Stationary Measure (Invariant Measure). Let a row vector

$$\mu^* = (\mu^*(0), \mu^*(1), \ldots, \mu^*(i), \ldots)$$

is called a stationary measure (invariant measure), if $\mu^*(i)^* \geq 0, \forall i \in S$ and $\mu^* P = \mu^*$.

R A stationary measure is a stationary distribution without normalization. If $\sum_i \mu^*(i) < \infty$, then it can be normalized to get a stationary distribution.

Theorem 1.12.1 Let $\{X_n\}_{n=0,1,2,\ldots}$ be an irreducible and recurrent DTMC with transition matrix P. Let $x \in S$ and $Tx := \min\{n \geq 1 : X_n = x\}$, then

$$\mu_x(y) = \sum_{n=0}^{\infty} \mathbf{P}_x(X_n = y, T_x > n), y \in S$$

defines a stationary measure with $0 < \mu_x(y) < \infty, \forall y \in S$.

Proof. Define $\bar{P}_{xy}^n = \mathbf{P}_x(X_n = y, T_x > n)$, then

$$\mu_x(y) = \sum_{n=0}^{\infty} \bar{P}_{xy}^n$$

We have the following two cases:

1. **For $z \neq x$:**

$$\begin{aligned}(\mu_x P)(z) = \sum_y \mu_x(y) P_{yz} \\ \sum_y \left(\sum_{n=0}^{\infty} \bar{P}_{xy}^n \right) P_{yz} \\ \sum_{n=0}^{\infty} \sum_y \bar{P}_{xy}^n P_{yz}\end{aligned}$$

then,

$$\begin{aligned}\sum_y \bar{P}_{xy}^n P_{yz} &= \sum_y \mathbf{P}_x(X_n = y, T_x > n, X_{n+1} = z) \\ &= \mathbf{P}_x(T_x > n+1, X_{n+1} = z) = \bar{P}_{xz}^{n+1}\end{aligned}$$

Thus,

$$\begin{aligned}\sum_{n=0}^{\infty} \sum_y \bar{P}_{xy}^n P_{yz} &= \sum_{n=0}^{\infty} \bar{P}_{xz}^{n+1} \\ &= \sum_{n=0}^{\infty} \bar{P}_{xz}^n = \mu_x(z)\end{aligned}$$

Since $\bar{P}^0_{xz} = \mathbf{P}_x(X_0 = z, T_x > 0) = 0$.

2. **For $z = x$:** similarly, we have

$$\begin{aligned}\sum_y \bar{P}^n_{xy} P_{yx} &= \sum_y \mathbf{P}_x(X_n = y, T_x > n, X_{n+1} = x) \\ &= \mathbf{P}_x(T_x = n+1)\end{aligned}$$

Thus,

$$\begin{aligned}(\mu_x P)(x) &= \sum_{n=0}^{\infty} \sum_y \bar{P}^n_{xy} P_{yx} = \sum_{n=0}^{\infty} \mathbf{P}_x(T_x = n+1) = 1 \quad \textbf{By total probability and recurrence} \\ &= \mu_x(x) = \sum_{n=0}^{\infty} \mathbf{P}_x(X_n = x, T_x > n)\end{aligned}$$

The last line is true since the probability will be 1 only when $n = 0$, 0 for all other n.
Combine part 1 and 2, we have

$$(\mu_x P)(z) = \mu_x(z), \forall z \in S \implies \mu_x P = \mu_x$$

Next, we show $0 < \mu_x(y) < \infty$ for all $y \in S$. First,

$$\begin{aligned}1 = \mu_x(x) &= (\mu_x P^n)_x \\ &= \sum_z \mu_x(z) P^n_{zx} \\ &\geq \mu_x(y) P^n_{yx}\end{aligned}$$

Since we have irreducibility which implies $y \to x$. Then, there exists n such that $P^n_{yx} > 0$. This implies that $\mu_x(y) < \infty$.
Second, recall that the chain can visit y before returning to x is

$$\begin{aligned}&\mathbf{P}_x(\text{number of visits to } y \text{ before returning to } x \geq 1) > 0 \\ &\implies \mathbb{E}_x(\text{number of visits to } y \text{ before returning to } x \geq 1) = \mu_x(y) > 0\end{aligned}$$

■

R Note that

$$\begin{aligned}\mu_x(y) &= \sum_{n=0}^{\infty} \mathbf{P}_x(X_n = y, Tx > n) \\ &= \sum_{n=0}^{\infty} \mathbb{E}_x\left(\mathbf{1}_{\{X_n = y\}} \mathbf{1}_{\{T_x > n\}}\right) = \mathbb{E}_x \sum_{n=0}^{T_x - 1} \mathbf{1}_{\{X_n = y\}} \\ &= \mathbb{E}_x(\text{number of visits to } y \text{ before returning to } x)\end{aligned}$$

1.13 The Period of A State

Definition 1.13.1 — Period of a State. The period of a state x is defined as

$$d(x) := \gcd\{n \geq 1 : P^n_{xx} > 0\}$$

R Note that we are taking the gcd of the steps when the probability of state x going back to x is not 0. There is no guarantee for going-back.

Definition 1.13.2 — Aperiodic. If x has period 1, then we say x is **aperiodic**. If all states in a MC is **aperiodic**, then we call this MC is **aperiodic**.

If $P_{xx} > 0$, then x is obviously aperiodic. But the converse is not true, x is aperiodic does not imply $P_{xx} > 0$.

If x is **aperiodic**, we do not necessarily have $P_{xx} > 0$.

■ **Example 1.14** Note that $P_{00}^3 > 0$ and $P_{00}^2 > 0$ means $d(0) = 1$. However, $P_{00} = 0$.

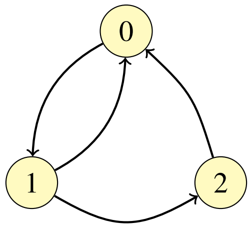


Figure 1.13.1: *Aperiodic Graph Example*

■

■ **Example 1.15 — Simple Random Walk Revisit.** Note that

$$\begin{cases} P_{00}^n = 0 & n = 1 \mod 2 \\ P_{00}^n = \binom{n}{n/2}\left(\frac{1}{2}\right)^n \end{cases}$$

where $\binom{n}{n/2}$ is the number of ways to get $\frac{n}{2}$ steps to the right, and $\left(\frac{1}{2}\right)^n$ is the probability of a particular ordering with $\frac{n}{2}$ steps to the right. By this analysis, we have that

$$d(0) = 2$$

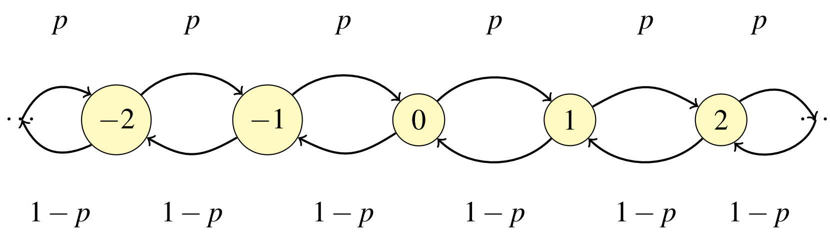


Figure 1.13.2: *Random Walk Visualization with* $p = \frac{1}{2}$

■

Lemma 1.14 — "Period is a class property".

$$x \to y, y \to x \Longrightarrow d(x) = d(y)$$

Proof. Since $x \to y$ and $y \to x$, there exists m, n such that

$$P_{xy}^m > 0 \qquad\qquad P_{yx}^n > 0$$

then

$$P_{xx}^{m+n} \geq P_{xy}^m P_{yx}^n > 0$$

Moreover, for any l such that $P^l_{yy} > 0$, we have

$$P^{m+n+l}_{xx} \geq P^m_{xy} P^l_{yy} P^n_{yx} > 0$$

As a result,

$$d(x)|m+n \qquad\qquad d(x)|m+n+l$$

thus,

$$d(x)|l$$

this holds for all l such that $P^l_{yy} > 0$, thus $d(x)$ is the common divisor for all elements in

$$\left\{l \geq 1 : P^l_{yy} > 0\right\}$$

Thus, $d(x)|d(y)$. Symmetrically, $d(y)|d(x)$. Thus,

$$d(x) = d(y)$$

as desired. ■

1.15 Main Theorems

Overall Conditions:

1. I: The MC is irreducible (one and only one class, everything communicates with everything)
2. A: The MC is aperiodic (i.e, all the states have period 1)
3. R: All the states are recurrent
4. S: There exists a stationary distribution π.

Theorem 1.15.1 — Convergence Theorem. Suppose I,A,S. Then,

$$P^n_{xy} \longrightarrow_{n\to\infty} \pi(y), \forall x, y \in S$$

no-matter where you starts, only depends on the target states.

> "The limiting transition probability, hence also the limiting distribution, does not depend on where we start. (Under the conditions of I,A,S)"

Or we can write

$$\lim_{n\to\infty} P^n_{xy} = \pi(y), \forall x, y \in S \Longrightarrow \lim_{n\to\infty} \mathbf{P}(X_n = y) = \pi(y)$$

Proof. To show this convergence theorem is true, we need to prove a lemma first.

Lemma 1.16 If there exists a stationary distribution π such that $\pi(y) > 0$, then y is recurrent.

Proof. Assume the DTMC $\{X_n\}_{n=0,1,\ldots}$ starts from the stationary distribution π. This means that

$$\begin{aligned}
\mathbf{P}(X_n = y) = \pi(y), \forall n = 0, 1, \ldots \\
\infty &= \sum_{n=1}^{\infty} \mathbf{P}(X_n = y) \\
&= \sum_{n=1}^{\infty} \mathbb{E}(\mathbf{1}_{\{X_n=y\}}) \\
&= \mathbb{E}\left(\sum_{n=1}^{\infty} \mathbf{1}_{\{X_n=y\}}\right) \\
&= \mathbb{E}(N(y)) \\
&= \sum_{x\in S} \mathbb{E}_x N(y)\pi(x) \\
&= \sum_{x\in S} \pi(x)\frac{\rho_{xy}}{1-\rho_{yy}} \\
&\leq \sum_{x\in S} \pi(x)\frac{1}{1-\rho_{yy}} \\
&= \frac{1}{1-\rho_{yy}} =
\end{aligned}$$

Thus,

$$\rho_{yy} = 1$$

This means the state y is recurrent. ■

By taking the contrapositive the lemma above, we have the following corollaries

Corollary 1.16.1 If y is transient, then $\pi(y) = 0$ for any stationary distribution π.

Corollary 1.16.2

$$\text{I and S} \implies \text{R}$$

Alright, back to the Convergence Theorem Proof. The proof is freaking long, but the main idea is

"Coupling"

Consider two independent DTMCs $\{X_n\}_{n=0,1,\ldots}$ and $\{Y_n\}_{n=0,1,\ldots}$, both having the same transition matrix P with arbitrary initial distributions.
It is easy to show that the pairs $Z_n = (X_n, Y_n)$ can also result in a DTMC with transition matrix

$$\bar{P}_{(x_1,y_1),(x_2,y_2)} = P_{x_1,x_2}P_{y_1,y_2}$$

Next, we show that under I and A, $\{Z_n\}_{n=1,\ldots}$ is also inreducible. Consider the following lemma

Lemma 1.17 If y is aperiodic, then there exists n_0 such that $P^n_{yy} > 0$ for all $n \geq n_0$.

Proof. We will use a **fact** in number theory, as a corollary of Bezout's Lemma. It states that

> If we have a set of coprime numbers I, then there are integers $i_1, \ldots, i_m$ from this I and a n_0, such that for any $n \geq n_0$, n can be written as
>
> $$n = a_1 i_1 + a_2 i_2 + \cdots + a_n i_n$$
>
> where a_i are positive integers.

Here, $I = \{n \geq 1 : P_{yy}^n > 0\}$, by aperiodicity of y, we know that all elements in I are coprime. By the Bezout's Corollary, there exists $n_0 \in \mathbb{N}$ such that $n \geq n_0$,

$$P_{yy}^n \geq \underbrace{P_{yy}^{i_1} P_{yy}^{i_1} \dots P_{yy}^{i_1}}_{a_1 \text{ terms}} \cdot \underbrace{P_{yy}^{i_2} P_{yy}^{i_2} \dots P_{yy}^{i_2}}_{a_2 \text{ terms}} \dots \underbrace{P_{yy}^{i_m} P_{yy}^{i_m} \dots P_{yy}^{i_m}}_{a_m \text{ terms}} > 0$$

■

Since we know that $\{X_n\}$ is irreducible, we have for any x_1, x_2 there exists K such that

$$P_{x_1,x_2}^K > 0$$

similarly, for any y_1, y_2, there exists L such that

$$P_{y_1,y_2}^L > 0$$

Since the DTMCs are aperiodic, we can apply the above lemma, and take M large enough such that $P_{x_2,x_2}^m > 0$ and $P_{y_2,y_2}^m > 0$ for any $m \leq M$. In particular, for any $m \geq M + \max K, L$, then

$$\begin{matrix} P_{x_1,x_2}^m \geq P_{x_1,x_2}^K P_{x_2,x_2}^{m-K} > 0 \\ P_{y_1,y_2}^m \geq P_{y_1,y_2}^L P_{y_2,y_2}^{m-L} > 0 \end{matrix} \implies \bar{P}_{(x_1,x_2),(y_1,y_2)}^m = P_{x_1,x_2}^m P_{y_1,y_2}^m > 0$$

Since this holds for any $(x_1, x_2), (y_1, y_2)$. We have $\{Z_n\}_{n=0,1,\dots}$ is irreducible. Moreover, we want to show $\{Z_n\}_{n=0,1,\dots}$ is recurrent. To see this, note that $\bar{\pi}$ given by $\bar{\pi}(x,y) = \pi(x)\pi(y)$ is a stationary distribution. Take x such that $\pi(x) > 0$, then $\bar{\pi}(x,x) > 0$. By the first lemma within this proof, we have (x,x) as a recurrent state. And since $\{Z_n\}$ is irreducible, then it must be recurrent as well. Now, define $T = \min\{n \geq 0 : X_n = Y_n\}$, this is the **first time that the two chains meet** and

$$V(x,x) = \min\{n \geq 0 : X_n = Y_n = x\} = \min\{n \geq 0 : Z_n = (x,x)\}$$

then, from what have deduced so far,

$$T \leq V(x,x) < \infty$$

with certainty.

Recall that if i is recurrent and $\rho_{ij} > 0$, then $\rho_{ji} = 1$.

Finally, we have proved that

"The two independent Markov Chains will eventually meet"

For any state $y \in S$, discuss the values of T and X_T, we have

$$\begin{aligned} \mathbf{P}(X_n = y, T \leq n) &= \sum_{m=0}^{n} \sum_{x \in S} \mathbf{P}(T = m, X_m = x, X_n = y) \\ &= \sum_{m=0}^{n} \sum_{x} \mathbf{P}(T = m, X_m = x)\mathbf{P}(X_n = y | X_m = x) \quad \textbf{Markov Property} \\ &= \sum_{m=0}^{n} \sum_{y} \mathbf{P}(T = m, Y_m = x)\mathbf{P}(Y_n = y | Y_m = x) \\ &= \mathbf{P}(Y_n = y, T \leq n) \end{aligned}$$

"After meeting, they have the same distribution."

Then,

$$
\begin{aligned}
&|\mathbf{P}(X_n = y) - P(Y_n = y)| \\
&= |(\mathbf{P}(X_n = y, T \leq n) + \mathbf{P}(X_n = y, T > n)) - (\mathbf{P}(Y_n = y, T \leq n) + \mathbf{P}(Y_n = y, T > n))| \\
&\leq \mathbf{P}(X_n = y, T > n) + \mathbf{P}(Y_n = y, T > n) \\
&\leq 2\mathbf{P}(T > n) \xrightarrow{n \to \infty} 0
\end{aligned}
$$

Recall that we have the freedom to choose the initial distributions of the $\{X_n\}$ and $\{Y_n\}$. Take $X_0 = x$ and $Y_0 \sim \pi$. Then,

$$\left|P_{xy}^n - \pi(y)\right| = |\mathbf{P}(X_n = y) - \pi(y)| = |\mathbf{P}(X_n = y) - \mathbf{P}(Y_n = y)| \xrightarrow{n \to \infty} 0$$

Thus, we have

$$P_{xy}^n \xrightarrow{n \to \infty} \pi(y)$$

■

Theorem 1.17.1 — Existence of Stationary Measure. Suppose I, R, then there exists a stationary measure μ^* with $0 < \mu^* < \infty, \forall x \in S$.

Proof. Proved by **Theorem 1.12.1** ■

R Remember that stationary measure is a weaker definition of stationary distribution.

Theorem 1.17.2 — Asymptotic Frequency. Suppose I, R, if $N_n(y)$ is the **number of visits to** y **up to time** n, then

$$\frac{N_n(y)}{n} \underset{n \to \infty}{\longrightarrow} \frac{1}{\mathbb{E}_y(T_y)}$$

where

$$T_y = \min\{n \geq 1 : X_n = y\}$$

we consider $\frac{N_n(y)}{n}$ as the fraction of time spent in y (up to time n).

"Long run fraction of time spent in y is $\frac{1}{\mathbb{E}_y(T_y)}$"

where $\mathbb{E}_y(T_y)$ is the expected revisit time to y given that we start with y, which is also the "expected cycle length".

Proof. We can chop the time line into difference cycles. Let $T_y^{(1)}, T_y^{(2)}, \ldots$ bee the time that the chain (re)visits y after time 0. By **Strong Markov Property**, $T_y^{(2)} - T_y^{(1)}, T_y^{(3)} - T_y^{(2)}, \ldots$ are i.i.d

Figure 1.17.1: *Chop the Time Line*

random variables. By the Strong Law of Large Number,

Theorem 1.17.3 — Strong Law of Large Number. For $X_1, X_2, \dots$ i.i.d r.v.s, then

$$\frac{\sum_{i=1}^{n} X_i}{n} \underset{a.s}{\to} \mathbb{E}(X_i)$$

we have

$$\Longrightarrow \frac{\sum_{i=1}^{k-1} T_y^{(i+1)} - T_y^{(i)}}{k-1}$$
$$\Longrightarrow \mathbb{E}\left(T_y^{(i+1)} - T_y^{(i)}\right) = \mathbb{E}_y(T_y)$$

With negligible changes (THIS IS VERY ANNOYING)

$$\Longrightarrow \frac{T_y^{(k)}}{k} \xrightarrow{n\to\infty} \mathbb{E}_y(T_y)$$

Observe that $T_y^{(N_n(y))} \le n \le T_y^{(N_n(y)+1)}$, then

$$\frac{T_y^{(N_n(y))}}{N_n(y)} \le \frac{n}{N_n(y)} \le \frac{T_y^{(N_n(y))+1}}{N_n(y)+1}\frac{N_n(y)+1}{N_n(y)} \quad \xrightarrow{n\to\infty} \quad \begin{array}{l} \dfrac{T_y^{(N_n(y))}}{N_n(y)} \Longrightarrow \mathbb{E}_y(T_y) \\ \dfrac{T_y^{(N_n(y))+1}}{N_n(y)+1}\dfrac{N_n(y)+1}{N_n(y)} \longrightarrow \mathbb{E}_y(T_y) \end{array}$$

By Squeeze Theorem, we have

$$\lim_{n\to\infty} \frac{n}{N_n(y)} = \mathbb{E}_y(T_y) \Longrightarrow \frac{N_n(y)}{n} \xrightarrow{n\to\infty} \frac{1}{\mathbb{E}_y(T_y)}$$

■

Theorem 1.17.4 — How to find a stationary distribution?. Suppose I and S, then,

$$\pi(y) = \frac{1}{\mathbb{E}_y(T_y)}$$

Proof. As I, S implies R, we can apply the result above, with the initial distribution being π. Taking expectation on both sides,

$$\mathbb{E}\left(\frac{N_n(y)}{n}\right) \longrightarrow \frac{1}{\mathbb{E}_y(T_y)}$$

This is a result from **Dominant Convergence Theorem (DCT)** from real analysis. Then, note that

$$\begin{aligned} \mathbb{E}(N_n(y)) &= \mathbb{E}\left(\sum_{m=1}^{n} \mathbf{1}_{\{X_m=y\}}\right) \\ &= \sum_{m=1}^{n} \mathbb{E}(\mathbf{1}_{\{X_m=y\}}) \qquad \textbf{By DCT} \\ &= \sum_{m=1}^{n} \mathbf{P}(X_m = y) \end{aligned}$$

Since the Chain is stationary, we have

$$\mathbf{P}(X_m = y) = \mathbf{P}(X_0 = y) = \pi(y)$$
$$\Longrightarrow \mathbb{E}(N_n(y)) = n\pi(y)$$

Thus,

$$\pi(y) = \frac{1}{\mathbb{E}(T_y)}$$

■

Corollary 1.17.5 — Nicest Case. Suppose I, A, S, (R), then

$$\pi(y) = \lim_{n\to\infty} P^n_{xy} = \lim_{n\to\infty} \frac{N_n(y)}{n} = \frac{1}{\mathbb{E}_y(T_y)}$$

$$\begin{aligned}\textbf{Stationary Distribution} &= \textbf{Limiting transition probability}\\ &= \textbf{Long-run fraction of time}\\ &= \frac{1}{\textbf{Expected revisit time}}\end{aligned}$$

"Everything exists and things are all equal!"

This is very philosophical LMAO!

R We don't need R actually, but we have not yet shown that relationship.

Theorem 1.17.6 — Long-run Average. Suppose I, S, and $\sum_x |f(x)|\pi(x) < \infty$. Then,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} f(X_m) = \sum_x f(x)\pi(x) = \pi f'$$

Proof. Recall that I, R, from Theorem 1.12.1, we have

$$\begin{aligned}\mu_x(y) &= \sum_{n=0}^{\infty} \mathbf{P}_x(X_n = y, T_x > n), y \in S\\ &= \mathbb{E}(\textbf{Number of visits to } y \textbf{ before returning to } x)\\ &= \mathbb{E}_x N_{T_x}(y)\end{aligned}$$

forms a stationary measure. Then, note that $\sum_y \mathbb{E}_x N_{T_x}(y) = \mathbb{E}_x(T_x) = \frac{1}{\pi(x)}$. We need a lemma to proceed.

Lemma 1.18 If I,S then $\pi(x) > 0$ for all $x \in S$.

"This is pretty much telling you that the stationary distribution exists and it is unique."

Proof. Say π is a distribution, then at $\pi(y) > 0$ for some $y \in S$. Since the Markov Chain is irreducible, we know that $y \to x$ for all $x \in S$. It means that there exists $n\mathbb{N}$ such that $P^n_{yx} > 0$, then

$$\pi = \pi P^n n$$

hence,

$$\begin{aligned}\pi(x) &= \sum_{z\in S} \pi(z) P^n_{zx}\\ &\geq \pi(y) P^n_{yx} > 0\end{aligned}$$

Then, we have $\mathbb{E}_x(T_x) < \infty$. Thus,

$$\left(\frac{E_x(N_{T_x}(y))}{\mathbb{E}_x(T_x)}\right)_{y\in S}$$

gives a stationary distribution. ∎

> R This is an amazing result to perform stationary measure **normalization** to get a stationary distribution.

Thus,

$$\frac{E_x(N_{T_x}(y))}{\mathbb{E}_x(T_x)} = \pi(y)$$

while $\mathbb{E}_x(T_x) = \frac{1}{\pi(x)}$, we have

$$E_x(N_{T_x}(y)) = \frac{\pi(y)}{\pi(x)}$$

Proposition 1.18.1 If I, then the stationary distribution is unique if it exists.

Proof. The proof is postponed to "harmonic function" part. ∎

Oh boy, we still need to get back to our reward function proof... The reward collected in k-th cycle (defined by returns to x) is

$$Y_k := \sum_{m=T_{k-1}+1}^{T_k} f(X_m)$$

we take the expectation of Y_k, then

$$\mathbb{E}(T_y) = \sum_{y\in S} \mathbb{E}_x(N_{T_x}(y))f(y) = \frac{\sum_{y\in S}\pi(y)f(y)}{\pi(x)}$$

The average reward over time is

$$\frac{\sum_y Y_k + \textbf{negligible terms}}{\sum_k (T_k - T_{k-1}) + \textbf{negligible terms}}$$

where $T_k = T_x^{(k)}$ We have each cycle to be i.i.d shown in the graph above. The negligible terms are

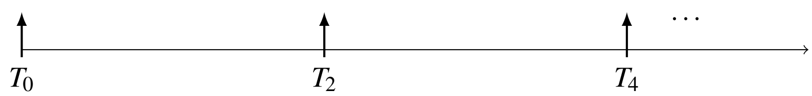


Figure 1.18.1: *Chop the Time Line*

from the first and last cycles. Then,

$$\begin{aligned}
&\frac{\sum_y Y_k + \textbf{negligible terms}}{\sum_k (T_k - T_{k-1}) + \textbf{negligible terms}} \\
&= \frac{\frac{1}{k}\sum_k Y_k + \cdots}{\frac{1}{k}\sum_k (T_k - T_{k-1}) + \cdots} \\
&\longrightarrow \frac{\mathbb{E}(Y_k)}{\mathbb{E}_x(T_x)} = \frac{\frac{\sum_{y\in S}\pi(y)f(y)}{\pi(x)}}{\frac{1}{\pi(x)}} \\
&= \sum_y \pi(y)f(y) = \pi f^{'}
\end{aligned}$$

∎

R

1. **Reminder:** f' is the transpose of f and f' is a column vector. Thus, $\pi f'$ is the dot product.
2. **Interpretation:** where $\sum_{m=1}^{n} f(X_m)$ is the total of rewards/costs up to time n based on f. Then, we can interpret $\frac{1}{n}\sum_{m=1}^{n} f(X_m)$ as the average rewards/costs up to time n per step. Then, the whole left-hand side $\lim_{n\to\infty}\frac{1}{n}\sum_{m=1}^{n} f(X_m)$ can be interpreted as the **long-run average rewards/costs per step**

■ **Example 1.16**

$$P = \begin{array}{c c} & \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 0.1 & 0.2 & 0.4 & 0.3 \\ 0.1 & 0.2 & 0.4 & 0.3 \\ 0.3 & 0.4 & 0.3 & 0 \\ 0.1 & 0.2 & 0.4 & 0.3 \end{pmatrix} \end{array}$$

1. This is **irreducible** since $P_{03}, P_{32}, P_{21}, P_{10} > 0$
2. This is **aperiodic** since $P_{00} > 0$ and aperiodicity is a class property. (check for the diagonal entries)
3. This is **recurrent** since there is only one irreducible class with finite state space and, by theorem, we have the whole class as a recurrent class.
4. This has a **stationary** distribution (measure), by solve using brute force.

$$\begin{cases} \pi P = \pi \\ \sum_x \pi_x = 1 \end{cases} \implies \pi = \left(\frac{19}{110}, \frac{30}{110}, \frac{40}{110}, \frac{21}{110}\right)$$

By using previous results, we have

1. **Limiting Transition Probability**

$$\lim_{n\to\infty} P^n_{xy} = \pi(y)$$

for example, $\lim_{n\to\infty} P^n_{12} = \pi(2) = \frac{4}{11}$.
Again, this limits does not depend on x.

2. **Long-run fraction/frequency of visiting state** y**:**

$$\lim_{n\to\infty} \frac{N_n(y)}{n} = \pi(y)$$

for example, the long-run fraction of time that the chain visits/ spends in state 0 is given by $\pi(0) = \frac{19}{110}$.

3. **Expected cycle length given by visits to** y or **the expected time that the chain visits state** y **again given the chain starts with state** y**:**

$$\mathbb{E}_y(T_y) = \frac{1}{\pi(y)}$$

for example, given the chain starts from state 3, the expected time that it returns to state 3 (by default, first time) is

$$\frac{1}{\pi(3)} = \frac{110}{21}$$

4. **Long-run average:** for example, according to the state, have a **holding cost** of $2x$, for $x \in \{0,1,2,3\}$, then the average/mean holding cost per term in the long-run is

$$\pi f' = \left(\frac{19}{110}, \frac{30}{110}, \frac{40}{110}, \frac{21}{110}\right)\begin{pmatrix} 0 \\ 2 \\ 4 \\ 6 \end{pmatrix} = \frac{19}{110}\times 0 + \frac{30}{110}\times 2 + \frac{40}{110}\times 4 + \frac{21}{110}\times 6 = \frac{173}{55}$$

■

R Typically, we want to solve for the stationary distribution first for questions like this.

1.19 Roles of Different Conditions

1.19.1 Irreducibility (I)

I (irreducibility) is related to the uniqueness of the stationary distribution

Irreducible $\implies$ stationary distribution is unique (if exists)

we have seen an example in Assignment 2. Or simpler

■ Example 1.17

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

both $(1,0)$ and $(0,1)$ are stationary distributions. Thus, any **convex** combination of them:

$$\alpha(1,0) + (1-\alpha)(0,1) = (\alpha, 1-\alpha), \forall 0 \leq \alpha \leq 1$$

is a stationary distribution. Therefore, π is not unique. As a result, the limiting transition probability $\lim_{n\to\infty} P^n_{xy}$ and the limiting distribution $\lim_{n\to\infty} \mathbf{P}(X_n = y)$ will depend on the initial state/distribution. ■

1.19.2 Aperiodicity (A)

A (Aperiodicity) is related to the existence of $\lim_{n\to\infty} P^n_{xy}$ (limiting transitional probability). In particular,

Aperiodic $\implies$ limit exists

■ Example 1.18

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

we have $d(0) = d(1) = 2$. Note that $P^2 = I$ and $P^{2n} = I$ and

$$P^{2n+1} = P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Thus, by the uniqueness of limits, we have

$$\lim_{n\to\infty} P^n_{xy}$$

does not exist for all $x, y \in \{0,1\}$. ■

1.19.3 Recurrent (R)

R (recurrence) is related to the existence of a stationary measure.

The MC is recurrent $\implies$ a stationary measure exists

1.20 Special Examples

1.20.1 Detailed Balance Condition

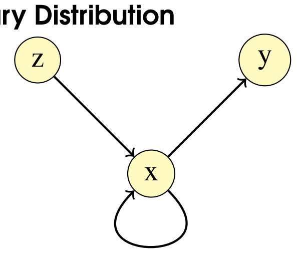
Definition 1.20.1 — Detailed Balance Condition. A distribution $\pi = \{\pi(x)\}_{x \in S}$ is said to satisfy the detailed balance condition if

$$\pi(x)P_{xy} = \pi(y)P_{yx}, \forall x, y \in S$$

Proposition 1.20.1 — Detailed Balance Condition $\implies$ Stationary Distribution. If a distribution π satisfies the detailed balance condition, then π is a stationary distribution.

Proof. See Assignment 2 Q2 (a) ■

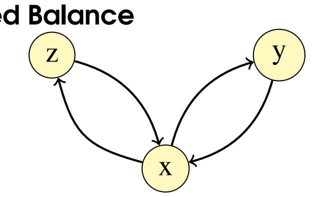
Stationary Distribution



Total probability "flow" entering x should be equal to the total probability flow leaving x

$$\sum_z \pi(z)P_{zx} = (\pi P)_x = \pi(x)$$

Detailed Balance



Flow $x \to y$ should be the flow $y \to x$

$$\pi(x)P_{xy} = \pi(y)P_{yx}$$

R Overall balance is not pair-wise balance.

1.20.2 Time Reversibility

Start with a DTMC $\{X_m\}_{m=0,1,\ldots}$. Fix n, then $\{Y_m\}_{m=0,1,\ldots,n}$ given by $Y_m = X_{n-m}$ is called the reversed process of $\{X_m\}$.

Theorem 1.20.2 If $\{X_m\}$ starts from a stationary distribution π satisfying $\pi(i) > 0$ for any $i \in S$, then its reversed process $\{Y_m\}$ is a DTMC with transition matrix given by

$$\begin{aligned}\hat{P}_{ij} &= \mathbf{P}(Y_{m+1} = j | Y_m = i) \\ &= \frac{\pi(j)P_{ji}}{\pi(i)}\end{aligned}$$

Proof. Check the Markov Property:

$$\begin{aligned}
&\mathbf{P}(Y_{m+1} = i_{m+1} | Y_m = i_m, \ldots, Y_0 = i_0) \\
&= \frac{\mathbf{P}(Y_{m+1} = i_{m+1}, Y_m = i_m, \ldots, Y_0 = i_0)}{\mathbf{P}(Y_m = i_m, \ldots, Y_0 = i_0)} \\
&= \frac{\mathbf{P}(X_{n-(m+1)} = i_{m+1}, X_{n-m} = i_m, \ldots, X_n = i_0)}{\mathbf{P}(X_{n-m} = i_m, \ldots, X_n = i_0)} \\
&= \frac{\mathbf{P}(X_{n-(m+1)} = i_{m+1})P_{i_{m+1},i_m}P_{i_m,i_{m-1}} \ldots P_{i_1,i_0}}{\mathbf{P}(X_{n-m} = i_m)P_{i_m,i_{m-1}} \ldots P_{i_1,i_0}} \\
&= \frac{\mathbf{P}(X_{n-(m+1)} = i_{m+1})P_{i_{m+1},i_m}}{\mathbf{P}(X_{n-m} = i_m)P_{i_m,i_{m-1}}}
\end{aligned}$$

Since $\{X_m\}$ starts from a stationary distribution $\mathbf{P}(X_{n-(m+1)} = i_{m+1}) = \pi(i_{m+1})$ and $\mathbf{P}(X_{n-m} = i_m) = \pi(i_m)$. Now, we have

$$\begin{aligned}&\mathbf{P}(Y_{m+1} = i_{m+1}|Y_m = i_m, \ldots, Y_0 = i_0)\\ &= \frac{\pi(i_{m+1})P_{i_{m+1},i_m}}{\pi(i_m)}\end{aligned}$$

This shows:

1. This transitional probability does not depend on the history $i_{m-1}, \ldots, i_0$. Hence, $\{Y_m\}_{m=0}^n$ is a DTMC
2. The transition probability is given by

$$\hat{P}_{ij} = \mathbf{P}(Y_{m+1} = j|Y_m = i) = \frac{\pi(j)P_{ji}}{\pi(i)}$$

■

R

1. We note this theorem means that the reversability requires the stationary distribution to exist
2. We can check that $\hat{P} = \{\hat{P}_{ij}\}_{i,j\in S}$ is actually a valid transition matrix

$$\hat{P}_{ij} = \frac{\pi(j)P_{ji}}{\pi(i)} \geq 0$$

and

$$\sum_{j\in S} \hat{P}_{ij} = \frac{\sum_{j\in S} \pi(j)P_{ji}}{\pi(i)} = \frac{(\pi P)_i}{\pi(i)} = \frac{\pi(i)}{\pi(i)} = 1$$

Definition 1.20.2 — Time-Reversable DTMC. A DTMC $\{X_m\}_{m=0,1,\ldots}$ is called time-reversable, of its reversed chain $\{Y_m\}_{m=0}^n$ has the same distribution as $\{X_m\}_{m=0}^n$ for all n.

R This is much stronger than reversability, not all reversable DTMC is time-reversable. But the other way around is clearly true. This is somehow related to the detailed balance condition in an intuitive way. Since the detail balance condition provides a two-way transition for each state, which provides the "time-reversability".

Proposition 1.20.3 A DTMC $\{X_m\}_{m=0,1,\ldots}$ is time-reversable **if and only if** it satisfies the detailed balance condition.

Proof.

1. **Assume Detailed Balance Condition:** we have stationarity for free. Say $\{X_m\}$ starts from π and $\pi(i)P_{ij} = \pi(j)P_{ji}$. Then, $Y_0 = X_n$ starts from π and the transition matrix

$$\hat{P}_{ij} = \frac{\pi(j)P_{ji}}{\pi(i)} = P_{ij}, \forall i, j \in S$$

Note that the starting stationary distribution and their transition matrices are identical, then we must have them as identical DTMC with same distributions.

2. **Assume Time-Reversability:** by definition, we have X_0 adn $X_n = Y_0$ have the same distribution for all n. Then, X_0 follows a stationary distribution.

$$P_{ij} = \hat{P}_{ij} = \frac{\pi(j)P_{ji}}{\pi(i)} \Longrightarrow \pi(i)P_{ij} = \pi(j)P_{ji}, \forall i, j \in S$$

Thus, we have detailed balance condition.

■

1.20.3 The Metropolis-Hastings Algorithm

Goal:

the sample from a distribution $\pi = \{\pi(x)\}_{x\in S}$ when a direct sampling is hard to implement. This is an "MCMC" algorithm, which stand for Monte-Carlo Markov Chain

Algorithm 1.1 — Metropolis-Hastings.

- Start an irreducible DTMC with transition matrix $Q = \{Q_{xy}\}_{x,y\in S}$ and certain initial distribution (typically, an initial state)
- In each time,
    1. Propose a move from the current state x to state $y \in S$ according to probability Q_{xy}
    2. Accept this move with probability

$$r_{xy} = \min\left\{\frac{\pi(y)Q_{yx}}{\pi(x)Q_{xy}}, 1\right\}$$

if the move is rejected, stay in x

Wait for a long time, then sample from this MC.

R The transition matrix of this MC is given by

$$P_{xy} = Q_{xy}r_{xy}, x \neq y$$
$$P_{xx} = 1 - \sum_{y\neq x} P_{xy}$$

we show π is the stationary distribution of this MC.

Indeed, π satisfies the detailed balance condition:

For any two states $x,y \in S$, by symmetry, assume $\pi(y)Q_{yx} \geq \pi(x)Q_{xy}$ and $r_{xy} = 1$ and $r_{yx} = \frac{\pi(x)Q_{xy}}{\pi(y)Q_{yx}}$, then

$$P_{xy} = Q_{xy} \qquad\qquad P_{yx} = \frac{\pi(x)Q_{xy}}{\pi(y)}$$

$$\implies \pi(x)P_{xy} = \pi(x)Q_{xy} = \pi(y)P_{yx}$$

Since typical, the rejection rate is positive, the Markov Chain is automatically aperiodic. Thus, the convergence of the transition matrix is guaranteed by the **Convergence Theorem**

$$\lim_{n\to\infty} \mathbf{P}(X_n = x) = \pi(x)$$

1.21 Exit Distribution

1.21.1 Basic Setting

The subsets A,B of the state space. $C = S - (A\cup B)$ is finite. The question is

> "Starting in a state in C, what is the probability that the chain exits C by entering A or B."

Mathematical Formulation

We define $V_A = \min\{n \geq 0 : X_n \in A\}$ and $V_B = \min\{n \geq 0, X_n \in B\}$. Then, what is $\mathbf{P}_x(V_A < V_B)$?

■ **Example 1.19**

$$P = \begin{array}{c c} & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 0.25 & 0.6 & 0 & 0.15 \\ 0 & 0.2 & 0.7 & 0.1 \\ & & 1 & \\ & & & 1 \end{pmatrix} \end{array}$$

then, let's say

$$C = \{1,2\} \qquad A = \{3\} \qquad B = \{4\}$$

We think the entry $P_{33} = 1$ and $P_{44} = 1$ are not that important since we only care about the chain before going to 3 or 4. ■

Let $h(1) = \mathbf{P}_1(V_3 < V_4), h(2) = \mathbf{P}_2(V_3 < V_4)$. Discuss the first-step

$$\begin{aligned} h(1) &= \mathbf{P}_1(V_3 < V_4) \\ &= \sum_{x=1}^{4} \mathbf{P}(V_3 < V_4 | X_1 = x, X_0 = 1)\mathbf{P}(X_1 = x | X_0 = 1) \end{aligned}$$

$$\mathbf{P}(V_3 < V_4 | X_1 = x, X_0 = 1) = \begin{cases} \mathbf{P}_1(V_3 < V_4) = h(1) & x = 1 \\ \mathbf{P}_2(V_3 < V_4) = h(2) & x = 2 \\ 1 & x = 3 \\ 0 & x = 4 \end{cases}$$

$$\implies h(1) = 0.25h(1) + 0.6h(2)$$

similarly,

$$h(2) = 0.2h(2) + 0.7$$

Solve this system of equations, we have

$$\begin{cases} h(1) = 0.7 \\ h(2) = \frac{7}{8} \end{cases}$$

R To solve this question, we need to introduce the idea of First-Step Analysis.

1.21.2 First-Step Analysis

Theorem 1.21.1 Let $S = A \cup B \cup C$, where A, B, C are disjoint sets, and C is finite. If $\mathbf{P}_x(V_A \wedge V_B < \infty) > 0$, for all $x \in C$. Then,

$$h(x) := \mathbf{P}_x(V_A < V_B)$$

is the unique solution of the system of equations

$$h(x) = \sum_y P_{xy} h(y), x \in C$$

with boundary conditions

$$h(a) = 1, a \in A \qquad\qquad h(b) = 0, b \in B$$

Proof. By first-step analysis,

$$\begin{aligned}
h(x) &= \mathbf{P}(V_A < V_B | X_0 = X) \\
&= \sum_{y \in S} \mathbf{P}(V_A < V_B | X_0 = x, X_1 = y) \cdot \mathbf{P}(X_0 = x | X_1 = y) \\
&= \sum_{y \in S} P_{xy} h(y)
\end{aligned}$$

boundary conditions hold trivially.
Hence, we only need to look at the uniqueness. Note that the system of equations can be written as

$$h' = Qh' + R'_A$$

where $h = (h(x_1), h(x_2), \dots)$ for $x_1, x_2, \dots \in C$. Note that

$$\begin{array}{cc}
 & \begin{array}{ccc} C & \qquad & A \quad B \end{array} \\
\begin{array}{c} C \\ \\ A \\ B \end{array} & \begin{pmatrix} Q & | & R \\ -- & -- & -- \\ & & \\ & & \end{pmatrix}
\end{array}
\qquad\qquad
R'_A = \begin{pmatrix} \sum_{y \in A} P_{x_1, y} \\ \sum_{y \in A} P_{x_2, y} \\ \vdots \end{pmatrix}$$

The reason is that

$$\begin{aligned}
h(x) &= \sum_{y \in S} P_{xy} h(y) \\
&= \underbrace{\sum_{y \in C} P_{xy} h(y)}_{=(Qh)(x)} + \underbrace{\sum_{y \in A} P_{xy}}_{=(R'_A)(x)}
\end{aligned}$$

then,

$$\begin{aligned}
I \cdot h' &= Qh' + R'_A \\
(I - Q)h' &= R'_A
\end{aligned} \implies h' = (I - Q)^{-1} R'_A$$

is unique as long as $I - Q$ is invertible. Note that

$$\begin{array}{cc}
 & \begin{array}{ccc} C & \qquad & A \quad B \end{array} \\
\begin{array}{c} C \\ \\ A \\ B \end{array} & \begin{pmatrix} Q & | & R & \\ -- & -- & -- & \\ 0 & | & I & \\ & | & & I \end{pmatrix}
\end{array}$$

Since for $\mathbf{P}_X(V_A < V_B)$, we only need to observe the chain before it hits A or B, the change of the rows in P corresponding to A and B will not change the result of this problem.
By doing this change, A and B are now **absorbing**, and all the states in C becomes transient! (since $\mathbf{P}_x(V_A \wedge V_B < \infty) > 0$). Now, we are working with a modified transition matrix P' and, therefore, a modified DTMC $\{X'_n\}_n$ To show $I - Q$ is invertible, note that since the states in C are transient (in P') and C is finite

$$\begin{aligned}
0 &= \lim_{n \to \infty} \mathbf{P}_x(X'_n \in C) \\
&= \lim_{n \to \infty} \sum_{y \in C} \left(P'^n\right)_{xy} \\
&= \lim_{n \to \infty} \sum_{y \in C} Q^n_{xy}
\end{aligned}$$

The last equality holds because of the block structure of P', recall that

$$P' = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix} \Longrightarrow P'^n = \begin{pmatrix} Q^n & \cdots \\ 0 & I \end{pmatrix}$$

This corresponds to the fact that in order to have $X'_n \in C$, we must have $X'_0, \ldots, X'_{n-1} \in C$ which implies that

$$\lim_{n\to\infty} Q^n = 0$$

as the zero matrix. Then, all the eigenvalues of Q have norm smaller than 1. Thus, there does not exist a non-zero f' such that

$$I \cdot f' = f' = Qf' \iff (I-Q)f' = 0$$

Thus, $I-Q$ is invertible. We are done! ■

R

1. We see that the function h in the above theorem satisifies

$$h(x) = \sum_y P_{xy} h(y) = \mathbb{E}_x(h(X_1)) =: h^{(1)}(x), \forall x \in C$$

Definition 1.21.1 — Harmonic Function. In general, a function h is called harmonic at state x, if

$$h(x) = \sum_y P_{xy} h(y) = \mathbb{E}_x(h(X_1)) = h^{(1)}(x)$$

h is called harmonic in $A \subseteq S$, if

$$h(x) = \sum_y P_{xy} h(y) = \mathbb{E}_x(h(X_1)) = h^{(1)}(x), \forall x \in A$$

h is called harmonic if

$$h(x) = \sum_y P_{xy} h(y) = \mathbb{E}_x(h(X_1)) = h^{(1)}(x), \forall x \in S \iff \underbrace{\begin{pmatrix} h(0) \\ h(1) \\ \vdots \end{pmatrix}}^{h'} = Ph' = h^{(1)'}$$

2. **Matrix Formula**: In the proof we have seen $h' = (I-Q)^{-1} R'_A$, where

$$\begin{array}{c|c} & \begin{array}{ccc} C & A & B \end{array} \\ \begin{array}{c} C \\ \\ A \\ B \end{array} & \begin{pmatrix} Q & | & R \\ -- & -- & -- \\ & | & \\ & | & \end{pmatrix} \end{array} \qquad R'_A = \begin{pmatrix} \sum_{y\in A} P_{x_1,y} \\ \sum_{y\in A} P_{x_2,y} \\ \vdots \end{pmatrix} = \begin{pmatrix} \mathbf{P}(X \in A | X_0 = x_1) \\ \mathbf{P}(X \in A | X_0 = x_2) \\ \vdots \end{pmatrix}$$

This is the matrix formula to calculate

$$\mathbf{P}_x(V_A < V_B) = h(x)$$

1.22 Exit Time

1.22.1 Basic Setting

Similar to the exit distribution part, but now we are interested in the **expected time** that the chain exits a part of the state space.
More precisely, let $S = A \cup C$, A, C disjoint and C is finite. Define

$$V_A := \min\{n \geq 0 : X_n \in A\}$$

which is the **first time the chain exits** C, which is also the **first time the chain hits/visits** A.
We want to know

$$\mathbb{E}_x(V_A) = \mathbb{E}(V_A|X_0 = x), x \in C$$

■ **Example 1.20 — Sample Example as for the Exit Distribution.**

$$P = \begin{array}{c c} & \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 0.25 & 0.6 & 0 & 0.15 \\ 0 & 0.2 & 0.7 & 0.1 \\ & & 1 & \\ & & & 1 \end{pmatrix} \end{array}$$

then, let's say

$$C = \{1,2\} \qquad A = \{3,4\}$$

Now, we want to know

$$g(1) = \mathbb{E}(V_A|X_0 = 1)$$
$$g(2) = \mathbb{E}(V_A|X_0 = 2)$$

Note that $g(3) = g(4) = 0$.

First-step Analysis

1.

$$\begin{aligned} g(1) &= \mathbb{E}(V_A|X_0 = 1) \\ &= \sum_{x=1}^{4} \mathbb{E}(V_A|X_1 = x, X_0 = 1)\mathbf{P}(X_1 = x|X_0 = 1) \end{aligned}$$

$$\mathbb{E}(V_A|X_1 = x, X_0 = 1) = \begin{cases} g(1)+1 & x = 1 \\ g(2)+1 & x = 2 \\ 1 & x = 3 \\ 1 & x = 4 \end{cases}$$

Note that this "1" comes from the time passed.

$$\Longrightarrow \begin{aligned} g(1) &= 0.25(g(1)+1) + 0.6(g(2)+1) + 0.15 \times 1 \\ &= 1 + 0.25g(1) + 0.6g(2) \qquad [1] \end{aligned}$$

2. Similarly, we have

$$g(2) = 1 + 0.2g(2) \qquad [2]$$

We can solve for $g(1), g(2)$ with $[1], [2]$. Then,

$$(g(1), g(2)) = \left(\frac{7}{3}, \frac{5}{4}\right)$$

Thus, starting from the state 1, the expected time until the chain reaches 3 or $3 = 4$ is $\frac{7}{3}$. If starting from state 2, the expected time until the chain reaches 3 or $3 = 4$ is $\frac{5}{4}$. ■

1.22.2 General Results

Theorem 1.22.1 — Unique Solution of $g(x)$. Let $S = A \cup C$ where A, C are disjoint and C is finite. If $\mathbf{P}_x(V_A < \infty) > 0$ for any $x \in C$, then $g(x) := \mathbb{E}_x(V_A), x \in C$ is the unique solution of the system of equations

$$g(x) = 1 + \sum_{y \in S} P_{xy} g(y), x \in C$$

with the boundary conditions $g(a) = 0, \forall a \in A$.

Proof.

1. **Existence:** by first-step analysis, we have

$$\begin{aligned} g(x) &= \sum_{y \in C} P_{xy}(g(y)+1) + \sum_{y \in A} P_{xy} \times 1 \\ &= 1 + \sum_{y \in C} P_{xy} g(y) \\ &= 1 + \sum_{y \in S} P_{xy} g(y) && S = A \cup C \text{ given the boundary condition} \end{aligned}$$

2. **Uniqueness:** observe that

$$g(x) = 1 + \sum_{y \in S} P_{xy} g(y), x \in S$$

means

$$g' = \mathbf{1}' + Qg'$$

$$\begin{pmatrix} g(x_1) \\ g(x_2) \\ \vdots \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \end{pmatrix} + Q \begin{pmatrix} g(x_1) \\ g(x_2) \\ \vdots \end{pmatrix}$$

where

$$\begin{array}{cc} & \begin{array}{ccc} C & & A \end{array} \\ \begin{array}{c} C \\ \\ A \end{array} & \begin{pmatrix} Q & | & R \\ -- & | & -- \\ & | & \end{pmatrix} \end{array}$$

Then,

$$\begin{aligned} Ig' &= \mathbf{1}' + Qg' \\ (I-Q)g' &= \mathbf{1}' \\ g' &= (I-Q)^{-1}\mathbf{1}' \end{aligned}$$

We are looking at exactly the same matrix $I - Q$ as in the exit distribution part. By Theorem 1.21.1, we have $I - Q$ is invertible, g' is the unique solution.

■

R

1. $g' = |C| \times$ column vector
2. $Q = |C| \times |c|$ matrix
3. $\mathbf{1}' = |C| \times 1$ column vector

1.23 Infinite State Space

1.23.1 Basic Setting

All the results covered in the previous parts hold for both finite and infinite state spaces (unless otherwise specified).
There is one distribution (one pair of two notions) which only makes sense in finite state space.

Definition 1.23.1 — Positive Recurrent and Null Recurrent. A state x is called positive recurrent if $\mathbb{E}_x(T_x) < \infty$ (recall $T_x = \min\{n \geq 1 : X_n = X\}$).
A recurrent state x is called null recurrent, if $\mathbb{E}_x(T_x) = \infty$.

R Recall that recurrence means $\mathbf{P}(T_x < \infty) = 1$ and transient means $\mathbf{P}(T_x = \infty) > 0$. The classification should be as follow:

Category	Subcategory
Recurrent	Positive Recurrent
Recurrent	Null Recurrent
Transient	Transient

■ **Example 1.21 — St. Petersburg paradox.** A random variable which is finite with probability 1, but has infinite mean. Let $X = 2^n$ with probability 2^{-n} for $n = 1, 2, \ldots$. Then,

$$\sum_{n=1}^{\infty} 2^{-n} = 1 \Longrightarrow \mathbf{P}(X = \infty) = 1$$

But

$$\mathbb{E}(X) = 2 \times \frac{1}{2} + 4 \times \frac{1}{4} + 8 \times \frac{1}{8} + \cdots = 1 + 1 + \cdots + \infty$$

■

1.23.2 Main Results

Theorem 1.23.1 For irreducible MC, the followings are equivalent

1. Some state is positive recurrent
2. There exists a stationary distribution π
3. All the states are positive recurrent

Proof. 1. $3 \to 1$: **TRIVIAL!!!!!!!!**

2. $1 \to 2$: Let x be a positive recurrent state. By Theorem 1.12.1, we know that a recurrent state can give us a stationary measure

$$\mu_x(y) = \sum_{n=0}^{\infty} \mathbf{P}_x(X_n = y | T_x > n), y \in S$$

starting with x, the expected number of visits to y before returning to x.
It can be normalized to become a stationary distribution if and only if

$$\mu_y \mu_x(y) < \infty$$

Note that

$$
\begin{aligned}
\sum_{y\in S}\mu(y) &= \sum_{y\in S}\sum_{n=0}^{\infty}\mathbf{P}_x(X_n=y|T_x>n)\\
&= \sum_{y\in S}\sum_{n=0}^{\infty}\mathbb{E}_x(\mathbf{1}_{X_n=y}\cdots\mathbf{1}_{T_x>n})\\
&= \mathbb{E}_x\left(\sum_{n=0}^{\infty}\mathbf{1}_{T_x>n}\sum_{y\in S}\mathbf{1}_{X_n=y}\right)\\
&= \mathbb{E}_x\left(\sum_{n=0}^{\infty}\mathbf{1}_{T_x>n}\right)\\
&= \mathbb{E}_x(T_x)<\infty \qquad\qquad \mathbf{1}_{T_x>n}=1, n=0,1,\dots,n-1\\
&\Longrightarrow \pi(y)=\frac{\mu(y)}{\mathbb{E}_x(T_x)}
\end{aligned}
$$

since x is positive recurrent.

3. $2\to 3$: Recall that we have

$$
\begin{aligned}
&\text{I,S} \Longrightarrow \pi(x)>0,\forall x\in S\\
&\text{I,R} \Longrightarrow \pi(x)=\tfrac{1}{\mathbb{E}_x(T_x)}
\end{aligned}
$$

Hence, $\mathbb{E}_x(T_x)<\infty$ for any $x\in S$. Thus, all $x\in S$ are positive recurrent.

■

Corollary 1.23.2 Positive recurrence and null recurrence are class properties.

$$x\leftrightarrow y\Longrightarrow(x\textbf{ is positive recurrent }\iff y\textbf{ is positive recurrent})$$

and

$$x\leftrightarrow y\Longrightarrow(x\textbf{ is null recurrent }\iff y\textbf{ is null recurrent})$$

Proof. This is just a sketch of the proof. Let x be a positive recurrent state and C be the communicating class containing x. Since C is recurrent, it is closed. Hence, $\{X_n\}\big|_C$ is a DTMC and its transition matrix is given by $\{P_{ij}\}_{i,j\in C}$.

$$
\begin{array}{cc}
 & \begin{array}{cc} C & C^c \end{array}\\
\begin{array}{c} C \\ C^c \end{array} & \begin{pmatrix} P\big|_C & 0 \\ & \end{pmatrix}
\end{array}
$$

The top right block is 0 since C is closed. Thus, $P\big|_C$ is a transition matrix. This restricted Markov Chain is irreducible. By the Theorem above, x is positive recurrent, will imply all the states in C are positive recurrent. (Note that for every $y\in C$,

$$\mathbb{E}_y(T_y)\big|_P=\mathbb{E}_y(T_y)\big|_{P|_C}$$

Since starting from y, the chain will only move in C ■

since both positive recurrence and recurrence are closed property, so is null recurrence.

Corollary 1.23.3 A state x is positive recurrent if and only if there exists a stationary distribution π such that $\pi(x) > 0$.

Proof. Note that for both directions, we have x is recurrent. Hence, it suffices to prove the result for the case when the chain is irreducible (otherwise, we can consider the restricted chain as the closed class containing x)

1. $\Longrightarrow$: From a previous result, we know that

$$\mu(y) = \sum_{n=0}^{\infty} \mathbf{P}_x(X_n = y, T_x > n), y \in S$$

   gives a stationary measure. Then,

$$\mu(x) = \sum_{n=0}^{n} \mathbf{P}_x(X_n = x, T_x > n) = 1$$
$$\sum_{y \in S} \mu(y) = \mathbb{E}_x(T_x) < \infty$$
$$\pi(x) = \frac{\mu(x)}{\mathbb{E}_x(T_x)} > 0$$

2. $\Longleftarrow$: Given by Theorem above

■

Corollary 1.23.4 A DTMC with finite state space must have at least one positive recurrent state.

Proof. Again, we can assume the Markov Chain is irreducible. Fix $x \in S$,

$$\mu(y) = \sum_{n=0}^{\infty} \mathbf{P}_x(X_n = y, T_x > n)$$

gives a stationary distribution. Moreover,

$$\sum_{y \in S} \mu(y) < \infty \Longleftarrow |S| < \infty$$
$$\Longrightarrow \pi(y) = \frac{\mu(y)}{\sum_{y \in S} \mu(y)}$$

is a stationary distribution. Thus, by the theorem above, one state is positive recurrent. ■

Corollary 1.23.5 A DTMC with finite state space does not have null recurrent state.

R **Why is this?** consider a nulll recurrent class. Since it is recurrent, it is closed. This implies that the class forms an irreducible Markov Chain. Assume the state space (hence the class only have finite states, then by the previous corollary, we have a positive recurrent state, which contradicts the class being null recurrent. The corollay holds automatically.

Moreover, this tells us that

> "A null recurrent class must have infinite states."

The intution is

$$\frac{1}{\mathbb{E}_x(T_x)} = \lim_{n\to\infty} \frac{N_n(x)}{n}$$

if we have $\mathbb{E}_x(T_x) = \infty$, this is saying that the long run fraction to each state is 0. This will not be possible with only finitely many states.

■ **Example 1.22 — Simple Random Walk (Revisit for $n+1$ times).**

$$\begin{aligned} &S = \mathbb{Z} \\ &P_{i,i+1} = p \qquad\qquad p \in (0,1) \\ &P_{i,i-1} = 1-p = q \end{aligned}$$

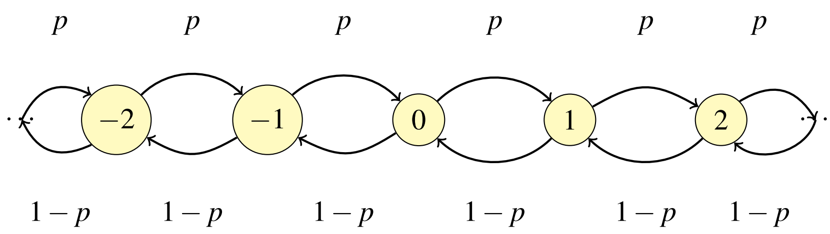


Figure 1.23.1: *Random Walk Visualization*

1. Irreducible
2. The period is 2
3. **Fact:** The simple random walk transient for $p \neq \frac{1}{2}$, null recurrent for $p = \frac{1}{2}$.

Proof. 1. **Case 1**: $p \neq \frac{1}{2}$. By symmetry, assume $p > \frac{1}{2}$, then

$$X_n = Y_1 + \cdots + Y_n$$

where $\{Y_n\}$ are iid with distribution

$$Y_n = \begin{cases} 1 & p \\ -1 & 1-p \end{cases}$$

Then,

$$\mathbb{E}(Y_n) = 1 \cdot p + (-1)(1-p) = 2p - 1 > 0$$

By **Strong Law of Large Number**,

$$\frac{X_n}{n} = \frac{1}{n}\sum_{m=1}^{n} Y_m \xrightarrow{a.s.} \mathbb{E}(Y_1) = 2p - 1$$

as $n \to \infty$. Thus,

$$X_n \xrightarrow{n\to\infty} \infty$$

Thus, for any state $i \geq 0$ (in particular, state 0), there is a last visit time to i. This implies that 0 is transient and $\{X_n\}$ is transient.

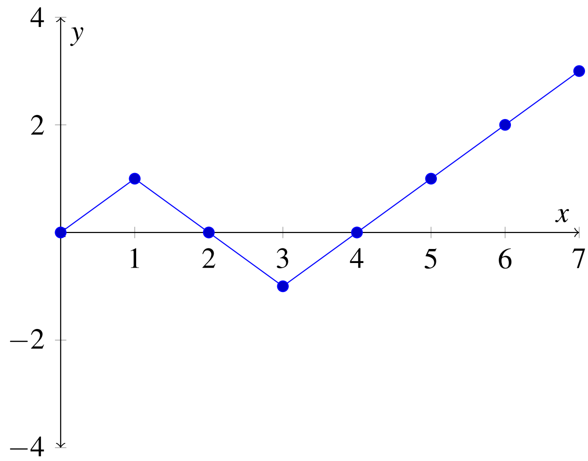


Figure 1.23.2: *Simple Random Walk*

2. **Case 2:** $p = \frac{1}{2}$, recall that a state i is recurrent, if and only if $\sum_{n=0}^{\infty} P_{ii}^{(n)} = \infty$. For state 0, we have

$$P_{00}^{2n+1} = 0 \qquad \text{period is 2}$$

$$\begin{aligned} P_{00}^{2n} &= \mathbf{P}(n \text{ stes to the left}, n \text{ steps to the right}) \\ &= \binom{2n}{n} \underbrace{\left(\frac{1}{2}\right)^n}_{n \text{ to the left}} \underbrace{\left(\frac{1}{2}\right)^n}_{n \text{ to the right}} \\ &= \binom{2n}{n} \left(\frac{1}{4}\right)^n \end{aligned}$$

This is hard to compute! But we have a good way to approximate!

Theorem 1.23.6 — Stirling's Formula.

$$n! \sim \sqrt{2\pi} e^{-n} n^{n+\frac{1}{2}}, \text{ as } n \to \infty$$

This is based on the fact that

$$\frac{a(n)}{b(n)} \xrightarrow{n\to\infty} 1 \Longrightarrow a(n) \sim b(n)$$

Now, we have

$$\begin{aligned} &\Longrightarrow \binom{2n}{n} = \frac{(2n)!}{n!n!} \sim \frac{\sqrt{2\pi} e^{-2n} (2n)^{2n+\frac{1}{2}}}{\left(\sqrt{2\pi} e^{-n} n^{n+\frac{1}{2}}\right)^2} \\ &= \frac{1}{\sqrt{2\pi}} 2^{2n+\frac{1}{2}} \frac{1}{\sqrt{n}} \\ &\Longrightarrow \binom{2n}{n} \left(\frac{1}{4}\right)^n \sim \frac{1}{\sqrt{\pi n}} \end{aligned}$$

Thus,
When $p = \frac{1}{2}$, there does not exist a stationary distribution.

Try to solve $\pi P = \pi$

Component corresponding to state i.

$$\pi(i) = \frac{1}{2}\pi(i-1) + \frac{1}{2}\pi(i+1)$$
$$\pi(i+1) - \pi(i) = \pi(i) - \pi(i-1)$$

Since this holds for all $i \in \mathbb{Z}$, $\{\pi(i)\}$ is an arithmetic series. The general form is

$$\pi(i) = \pi(0) + ia, \forall i \qquad a = \pi(1) - \pi(0)$$

also, $\pi(i) \in [0,1], \forall i$. This will force $a = 0$. However, this implies $\pi(i) = \pi(0), \forall i$.

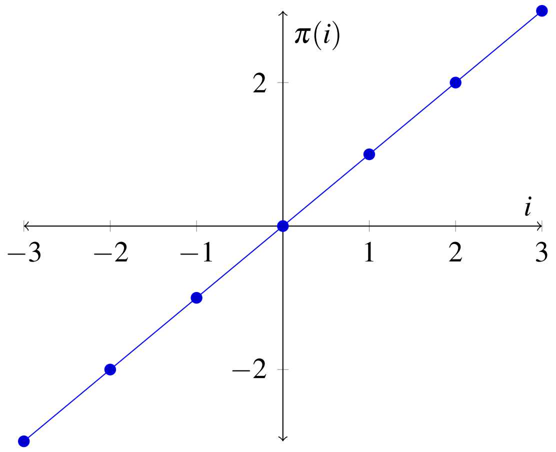


Figure 1.23.3: *Simple Random Walk $\pi(i)$ with $p = \frac{1}{2}$*

(a) If $\pi(0) = 0$, $\pi(i) = 0, \forall i \Longrightarrow \sum_i \pi(i) = 0 \neq 1$
(b) If $\pi(0) > 0$, $\sum_i \pi(i) = \infty \neq 1$

Thus, the normalization condition $\sum_i \pi(i) = 1$ cannot hold, a stationary distribution does not exist. In particular, 0 is not a positive recurrent state and the whole chain is not positive recurrent. In other words, the whole Chain is null recurrent.

■

■

■ **Example 1.23 — Simple Random Walk with A Reflecting Barrier.** A random walk with a reflecting barrier.

$$S = \overline{\mathbb{Z}^+} = \{0, 1, 2, \dots\}$$

and $P_{i,i+1} = p, P_{i,i-1} = 1 - p, i \geq 1, P_{01} = 1$. When $p < \frac{1}{2}$, such a chain is positive recurrent.

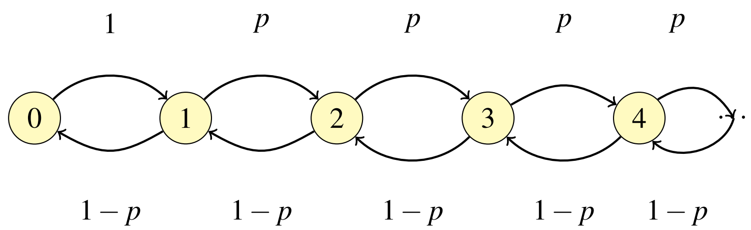


Figure 1.23.4: *Random Walk Visualization*

To see this, we solve for the stationary distribution. Only $P_{i,i+1}$ and $P_{i,i-1}$ are non-zero, we can use the detailed balanced condition (tridiagonal transition matrix). Then,

$$\begin{aligned}
\pi(0)\cdot 1 = \pi(1)(1-P) &\Longrightarrow \pi(1) = \frac{1}{1-p}\pi(0) \\
\pi(i)\cdot p = \pi(i+1)(1-p), i = 1,2,\dots &\Longrightarrow \pi(i+1) = \frac{p}{1-p}\pi(i) \\
\pi(i) = \frac{p}{1-p}\pi(i-1) &= \left(\frac{p}{1-p}\right)^2 \pi(i-2) = \cdots \\
&= \left(\frac{p}{1-p}\right)^{i-1} \pi(i) \\
&= \left(\frac{p}{1-p}\right)^{i-1} \frac{1}{1-p}\pi(0)
\end{aligned}$$

A stationary distribution exists if and only if $\sum_{i=0}^{\infty} \pi(i) < \infty$.
Here $\pi(1), \pi(2), \dots$ forms a geometric series with ratio a $\frac{p}{1-p}$, which is smaller than 1 if and only if $p < \frac{1}{2}$. ■

R The reflected simple random walk is

1. Positive recurrent $\iff p < \frac{1}{2}$
2. Null recurrent $\iff p = \frac{1}{2}$
3. Transient $\iff p > \frac{1}{2}$

1.24 Branching Process (Galton-Watson Process)

1.24.1 Basic Setup

Consider a population. Each organism, at the end of its life, produces a random number Y of offsprings. The distribution of Y is denoted as

$$\mathbf{P}(Y = k) = P_k, P_k \geq 0, k = 0, 1, \dots$$

and $\sum_{k=0}^{\infty} P_k = 1$.
Start from one common ancestor, $X_0 = 1$. **The number of offsprings of different individuals are independent**.
Let

$X_n :=$ the number fo individual (in the population) in the n-th generation.

Then,

$$X_{n+1} = Y_1^n + Y_2^{(n)} + \cdots + Y_{X_n}^{(n)}$$

where $Y_1^n, Y_2^{(n)}, \dots, Y_{X_n}^{(n)}$ are **independent** copies of Y and $Y_i^{(n)}$ is the ubmer of offsprings of the i-th individual in the n-th generation.
One thing we care about is the expectation of the number of offsprings of the n-th generation:

$$\mathbb{E}(X_n) = ?$$

Assuming $\mathbb{E}(Y) = \mu$.

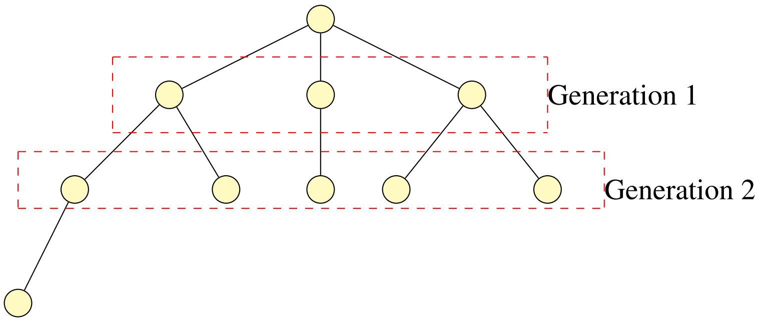


Figure 1.24.1: *Branching Process*

Solution:

$$\begin{aligned}\mathbb{E}(X_{n+1}) &= \mathbb{E}\left(Y_1^n + Y_2^{(n)} + \cdots + Y_{X_n}^{(n)}\right) \\ &= \mathbb{E}\left(\mathbb{E}\left(Y_1^n + Y_2^{(n)} + \cdots + Y_{X_n}^{(n)} | X_n\right)\right) \\ &= \mathbb{E}(\mu X_n) = \mu \mathbb{E}(X_n)\end{aligned}$$

R This result is known as **Wald's Identity** in statistics

$$\mathbb{E}(X_{n+1}) = \mu \mathbb{E}(X_n)$$

We can continue this inductively to see that

$$\mathbb{E}(X_n) = \mu^n \mathbb{E}(X_0) = \mu^n, n = 0, 1, \ldots$$

1.24.2 Extinction Probability

As long as $P_0 > 0$, state 0 is absorbing and all the other states are transient.
But it does not mean that the population will extinct for sure.
If the population on average keeps growing and tends to infinity with positive probability, then we probability of extension is smaller than 1.
To find the extinction probability, we introduce the mathematical tool, **Generating Functions**.

Definition 1.24.1 — Generating Function. Let $P = \{P_0, P_1, \ldots\}$ be a distribution on on $\{0, 1, \ldots\}$. Let η be a random variable following distribution P. That is

$$\mathbf{P}(\eta = i) = P_i$$

The **generating function** of η or of P is defined by

$$\begin{aligned}\varphi(s) &= \mathbb{E}(s^\eta) \\ &= \sum_{k=0}^{\infty} P_k s^k \quad 0 \le s \le 1\end{aligned}$$

Proposition 1.24.1 — Properties of Generating Functions. Let $\varphi(s)$ be a generating function

1. $\varphi(0) = P_0$, $\varphi(1) = \sum_{k=0}^{\infty} P_k = 1$
2. Generating function determines the distribution.

$$P_k = \frac{1}{k!} \frac{d^k \varphi(s)}{ds^k}\bigg|_{s=0}$$

Reason (This is immediate by Taylor Expansion, a short illustration is included here):

$$\varphi(s) = P_0 + P_1 s^1 + \cdots + P_{k-1}s^{k-1} + P_k s^k + P_{k+1}s^{k+1} + \ldots$$

then,

$$\frac{d^k\varphi(s)}{ds^k} = k!P_k + (\cdots)s + (\cdots)s^2 + \ldots$$

Then,

$$\left.\frac{d^k\varphi(s)}{ds^k}\right|_{s=0} = k!P_k \Longrightarrow P_k = \frac{1}{k!}\left.\frac{d^k\varphi(s)}{ds^k}\right|_{s=0}$$

In particular, given $P_1, P_2, \cdots \geq 0$, this implies $\varphi(s)$ is **increasing** and **convex** (all of its derivatives will be positive).

3. Let $\eta_1, \ldots, \eta_n$ be independent random variables with generating functions $\varphi_1, \ldots, \varphi_n$, then

$$X = \eta_1 + \cdots + \eta_n$$

have generating function

$$\varphi_X(s) = \varphi_1(s) \ldots \varphi_n(s)$$

Proof.

$$\begin{aligned}\varphi_x(s) &= \mathbb{E}\big(s^X\big) \\ &= \mathbb{E}(s^{\eta_1} \ldots s^{\eta_n}) \\ &= \mathbb{E}(s^{\eta_1}) \ldots \mathbb{E}(s^{\eta_n}) \quad \textbf{independence} \\ &= \varphi_1(s) \ldots \varphi_n(s)\end{aligned}$$

■

4. **Pseudo Moments:** very useful!

$$\left.\frac{d^k\varphi(s)}{ds^k}\right|_{s=1} = \left.\frac{d^k\mathbb{E}(s^\eta)}{ds^k}\right|_{s=1} = \mathbb{E}\left(\frac{d^k s^\eta}{ds^k}\right)\Bigg|_{s=1} = \mathbb{E}\Big(\eta(\eta-1)\ldots(\eta-k+1)s^{\eta-k}\Big)\Bigg|_{s=1}$$

$$= \mathbb{E}(\eta(\eta-1)\ldots(\eta-k+1))$$

In particular,

$$\mathbb{E}(\eta) = \varphi'(1)$$

and

$$\mathbf{Var}(\eta) = \varphi''(1) + \varphi'(1) - \big(\varphi'(1)\big)^2$$

The Graph of a Generating Function

Back to extinction probability. Define

$$N = \min\{n \geq 0 : X_n = 0\}$$

to be the **extinction time** and

$$u_n = \mathbf{P}(N \leq n) = \mathbf{P}(X_n = 0)$$

where $\mathbf{P}(N \leq n)$ is the extinction happens before or at time n. Note that $\{u_n\}$ is an increasing sequence and bounded above, by **Monotone Convergence Theorem**, it is well-defined to define

$$u := \lim_{n\to\infty} u_n = \mathbf{P}(N < \infty) = \mathbf{P}(\text{the population eventually die out}) = \text{extinction probability}$$

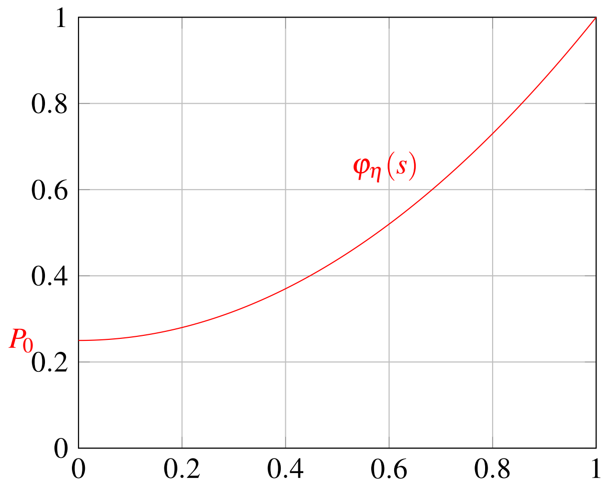


Figure 1.24.2: *The Graph of a Generating Function*

Our goal is to find u
(meiyouzhuya)
Note that we have the following relation between u_n and u_{n-1}. Then,

$$u_n = \sum_{k=0}^{\infty} P_k (u_{n-1})^k = \varphi(u_{n-1})$$

where φ is the generating fucntion of Y.

Reason:
Note that each sub-population has the same distribution as the whole population. The whole population dies out in n steps if and only if each sub-population initiated by an individual in generating 1 dies out in $n-1$ steps.
Then,

$$\begin{aligned} u_n &= \mathbf{P}(N \le n) \\ &= \sum_k \mathbf{P}(N \le n | X_1 = k)\mathbf{P}(X_1 = k) \\ &= \sum_k \mathbf{P}(N_1 \le n-1, \ldots, N_k \le n-1 | X_1 = k) P_k \\ &= \sum_k P_k u_{n-1}^k = \varphi(u_{n-1}) \end{aligned}$$

where N_m is the number of steps for the sub-population to die out. And we can also write $u_{n-1}^k = \mathbb{E}\left(u_{n-1}^Y\right)$.
Thus, the problem becomes:

> With an initial value $u_0 = 0$ since $(X_0 = 1)$ and the relation
>
> $$u_n = \varphi(u_{n-1})$$
>
> what is $\lim_{n\to\infty} u_n = u$?

Recall that

1. $\varphi(0) = P_0 > 0$
2. $\varphi(1) = 1$
3. $\varphi(s)$ is increasing

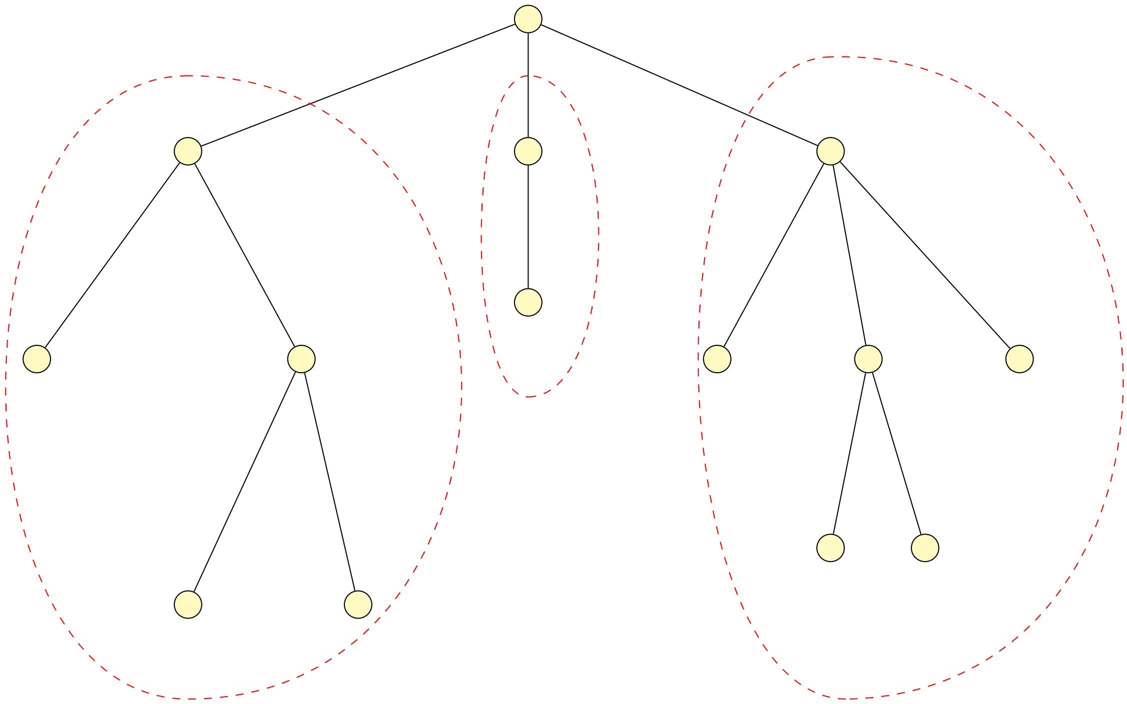
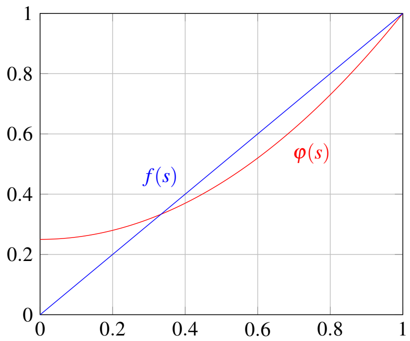


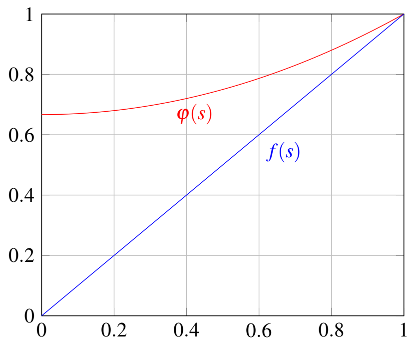
Figure 1.24.3: *We can divide the whole population into sub-populations*

4. $\varphi(s)$ is convex

Draw $\varphi(s)$ and the function $f(s) =$ between 0 and 1. We have two possibilities.



Case 1: $u < 1$



Case 2: $u = 1$ **extinction happens for sure**

Theorem 1.24.2 The extinction probability u will be the smallest intersection of $\varphi(s)$ and $f(s)$. Equivalently, it is the smallest solution of $\varphi(s) = s$ between 0 and 1.

Reason:

see the dynamics of the graph

This dynamic process verifies the results for **case 1** and **case 2**.

Q: How to tell whether we are in Case 1 or Case 2?

A: We can check the derivative at $s = 1$!

Note that $\varphi'(1) = \mathbb{E}(Y)$ and

$$\varphi'(1) > 1 \Longrightarrow \textbf{Case 1}$$
$$\varphi'(1) \leq 1 \Longrightarrow \textbf{Case 2}$$

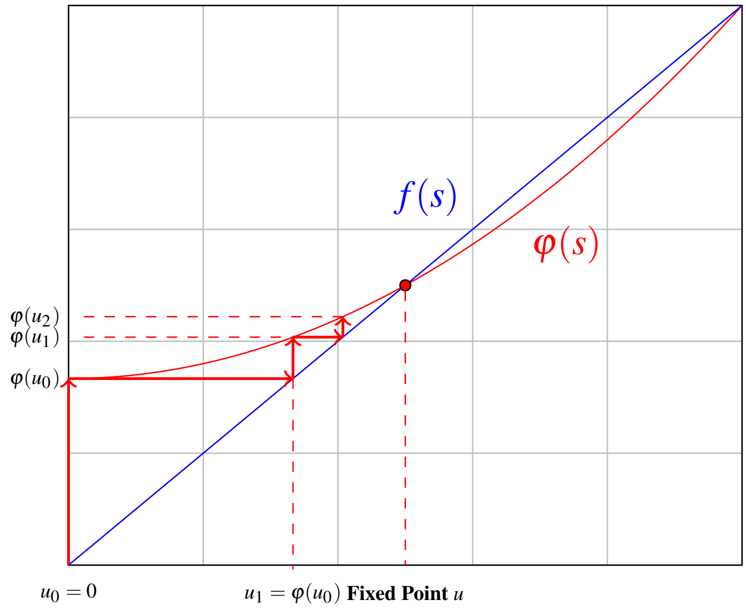


Figure 1.24.4: *Fixed Point Iteration*

Thus, we conclude that

$$\mathbb{E}(Y) > 1 \Longrightarrow \textbf{Extinction with certain porbability less than 1 and } u \textbf{ is a unique solution}$$

also, we can think of $\mathbb{E}(Y) > 1$ as **on average, there are more than 1 offspring**, so the population will probably explode, which diminish the chance to wipe out the whole population. (*Thanos has left the chat.*)

$$\mathbb{E}(Y) \leq 1 \Longrightarrow \textbf{Extinction happes for sure (with prob. 1)}$$

we can think of $\mathbb{E}(Y) \leq 1$ as **on average, there is less than or equal to 1 offspring**, so there is always a risk to have the population to die out.

1.25 Review on DTMC

Preliminaries

Probability space, stochastic process, index set, state space...

1.25.1 Basic Properties

Definition 1.25.1 — Discrete-Time Markov Chain (DTMC) and its transition matrix. A discrete-time stochastic process $\{X_n\}_{n=0,1,2,...}$ is called a discrete-time Markov Chain (DTMC) with transition matrix

$$P = [P_{ij}]_{i,j \in S}$$

if for any $j, i, i_{n-1}, \dots, i_0 \in S$,

$$\mathbf{P}(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P_{ij} = \mathbf{P}(X_{n+1} = j | X_n = i)$$

Proposition 1.25.1 — Properties of Transition Matrix.

$$P_{ij} \geq 0 \text{ and } \sum_{j \in S} P_{ij} = 1 \iff \text{there exists a legitimate DTMC}$$

Proposition 1.25.2 — Multi-Step Transition Probabilities.

$$P^{(n)} = P^n$$

and **C-K Equation:** $P^{(n+m)} = P^{(n)}P^{(m)}$, more frequently,

$$P_{ij}^{n+m} = \sum_{k \in S} P_{ik}^{(m)} P_{kj}^{(n)}$$

R We also have graphical representation (this does not require justification)

Conditional Probability

1. $\mathbb{E}(g(X)|Y = y)$ is a number
2. $\mathbb{E}(g(X)|Y)$ is a random variable of Y

R Make sure you check the properties of conditional expectations. Especially, the **iterated expectation**

$$\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X)$$

Distribution of X_n and $\mathbb{E}(f(X_n))$

Given an initial distribution $\mu = (\mu(0), \mu(1), \dots)$ (a distribution of X_0), then

$$\mu_n = \mu P^n$$

as the distribution at time n (X_n). For $f = (f(0), f(1), \dots)$, then

$$\begin{aligned}\mathbb{E}(f(X_n)) &= \mu P^n f^{'} \\ &= \mu_n f^{'} \\ &= \mu f^{(n)^{'}}\end{aligned}$$

where

$$f^{(n)^{'}} = \begin{pmatrix} \mathbb{E}(f(X_n)|X_0 = 0) \\ \mathbb{E}(f(X_n)|X_0 = 1) \\ \vdots \end{pmatrix}$$

and μ and P completely characterize a DTMC.

R We have **row vectors** as distribution and function vector (on state) is a column vector when used.

1.25.2 Classification and Class Properties

$$T_y = \min\{n \geq 1 : X_n = y\} \text{ first visit (revisit) time}$$

$$\rho_{xy} = \mathbf{P}(T_y < \infty) = \mathbf{P}(T_y < \infty | X_0 = x)$$

we say $x \to y$ if $\rho_{xy} > 0$. We know that $x \leftrightarrow y$ is an equivalence relation. This implies that there is a disjoint partition of states into communicating classes. When there is only one a communicating class, we call the Chain **irreducible**. To identity the classes, we can find the loops.

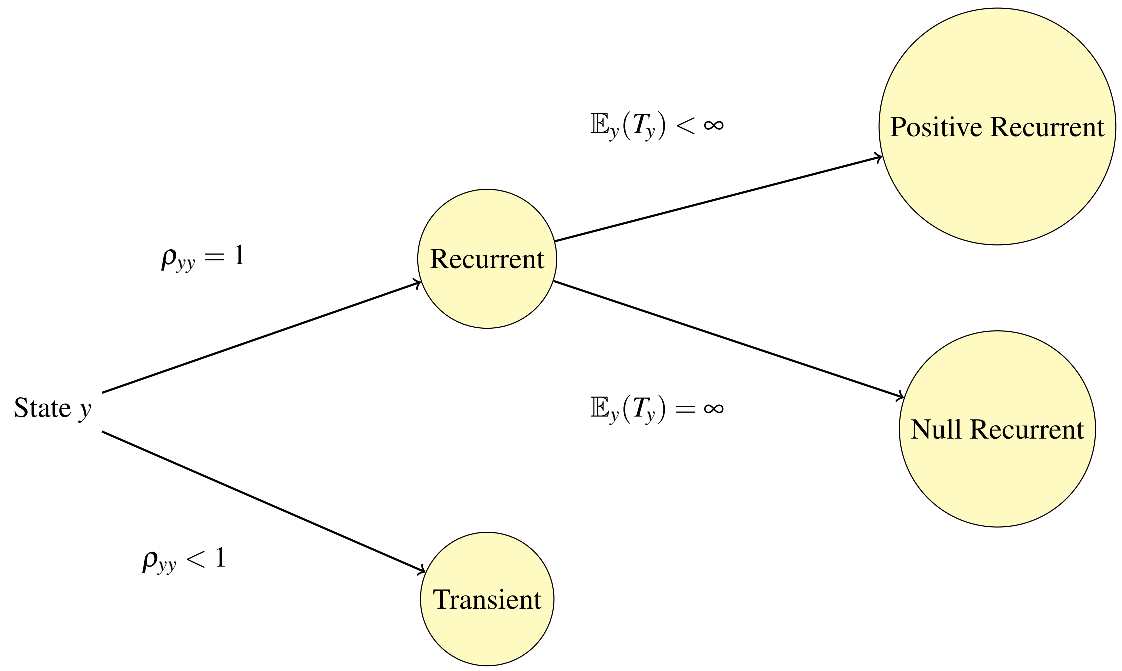


Figure 1.25.1: *Summary*

Positive Recurrence/Null Recurrence/Transience

Positive recurrent, null recurrent, transience are class properties!

Criteria for Recurrence vs. Transience

Recurrent	Transient
$\rho_{yy} = \mathbf{P}_y(T_y = \infty) = 1$	$\rho_{yy} < 1$
$\mathbf{P}(N(y) = \infty) = 1$	$\mathbf{P}(N(y) < \infty) = 1$
$\mathbb{E}_y(N(y)) = \infty$	$\mathbb{E}_y(N(y)) < \infty$
$\sum_{n=1}^{\infty} P_{yy}^n = \infty$	$\sum_{n=1}^{\infty} P_{yy}^n < \infty$

Proposition 1.25.3
1. $\rho_{xy} > 0, \rho_{yx} < 1 \implies x$ is transient
2. x is recurrent and $\rho_{xy} > 0 \implies \rho_{yx} = 1$

Definition 1.25.2 — Closed Set. If $i \in A, j \notin A \implies P_{ij} = 0$, equivalently, $\rho_{ij} = 0$.

Proposition 1.25.4 A finite and closed class is positive recurrent.

Corollary 1.25.5 A DTMC with finite DTMC does not have null recurrent state/class.

Proposition 1.25.6 If a class is recurrent, then it is closed.

Criteria for Positive Recurrent State y

1. $\mathbb{E}_y(T_y) < \infty$ (hard to do most of the time)
2. There exists a stationary distribution π such that $\pi(y) > 0$.

Decomposition of the State Space S

$$S = T \cup R_1 \cup R_2 \cup \dots$$

where T is the transient states (con contain classes and stand-alone states) and R_i are the recurrent (closed) classes.

Proposition 1.25.7 — Starting from x, how many visits to y on average?.

$$\mathbb{E}_x(N(y)) = \frac{\rho_{xy}}{1-\rho_{yy}}$$

Strong Markov Property

$$\{X_{T_y+k}\}_{k=0,1,2...}$$

behaves like the MC starting from $X_0 = y$.

Period

$$d(x) = \gcd\{n \geq 1 : P_{xx}^n > 0\}$$

If $d(x) = 1 \Longrightarrow x$ is aperiodic.
If all states are aperiodic, then the MC is aperiodic. Easiest way to check is

$$P_{xx} > \Longrightarrow x \text{ aperiodic}$$

the other direction is false, check the example in previous part.

Proposition 1.25.8 Period is a class property.

1.25.3 Stationary Distribution and Limiting Behaviour

Definition 1.25.3 — Stationary Distribution. A probability distribution $\pi = (\pi_0, \pi_1, \dots)$ is called a stationary distribution (invariant distribution) of the DTMC $\{X_n\}_{n=0,1,...}$ with transition matrix P if

1. $\pi = \pi P$ as a system of equations
2. $\sum_{i \in S} \pi_i = 1$ by the definition of probability distribution

Proposition 1.25.9 If the DTMC is irreducible, recurrent, then

$$\begin{aligned}\mu_x(y) &= \sum_{n=0}^{\infty} \mathbf{P}_x(X_n = y, T_x > n) \\ &= \mathbb{E}_x(\text{number of visits to } y \text{ before returing to } x)\end{aligned}$$

give a stationary measure where $\mu_x(x) = 1$. It gives stationary distribution if and only if

$$\sum_{y \in S} \mu_x(y) = \mathbb{E}_x(T_x) < \infty$$

which means x is a positive recurrent state.

Main Theorems

Theorem 1.25.10 — Convergence Theorem. Suppose I,A,S. Then,

$$P_{xy}^n \longrightarrow_{n\to\infty} \pi(y), \forall x, y \in S$$

no-matter where you starts, only depends on the target states.

> "The limiting transition probability, hence also the limiting distribution, does not depend on where we start. (Under the conditions of I,A,S)"

Or we can write

$$\lim_{n\to\infty} P_{xy}^n = \pi(y), \forall x, y \in S \Longrightarrow \lim_{n\to\infty} \mathbf{P}(X_n = y) = \pi(y)$$

Theorem 1.25.11 — Asymptotic Frequency. Suppose I, R, if $N_n(y)$ is the **number of visits to** y **up to time** n, then

$$\frac{N_n(y)}{n} \underset{n\to\infty}{\longrightarrow} \frac{1}{\mathbb{E}_y(T_y)}$$

where

$$T_y = \min\{n \geq 1 : X_n = y\}$$

we consider $\frac{N_n(y)}{n}$ as the fraction of time spent in y (up to time n).

"Long run fraction of time spent in y is $\frac{1}{\mathbb{E}_y(T_y)}$"

where $\mathbb{E}_y(T_y)$ is the expected revisit time to y given that we start with y, which is also the "expected cycle length".

Theorem 1.25.12 — How to find a stationary distribution?. Suppose I and S, then,

$$\pi(y) = \frac{1}{\mathbb{E}_y(T_y)}$$

Corollary 1.25.13 — Nicest Case. Suppose I, A, S, (R), then

$$\pi(y) = \lim_{n\to\infty} P^n_{xy} = \lim_{n\to\infty} \frac{N_n(y)}{n} = \frac{1}{\mathbb{E}_y(T_y)}$$

$$\begin{aligned}\textbf{Stationary Distribution} &= \textbf{Limiting transition probability}\\ &= \textbf{Long-run fraction of time}\\ &= \frac{1}{\textbf{Expected revisit time}}\end{aligned}$$

Theorem 1.25.14 — Long-run Average. Suppose I, S, and $\sum_x |f(x)|\pi(x) < \infty$. Then,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} f(X_m) = \sum_x f(x)\pi(x) = \pi f'$$

Proposition 1.25.15 If MC is I, then

S $\iff$ positive recurrent

In this case, $\pi(x) > 0$ for all $x \in S$

Detailed Balance Condition

$$\pi(x)P_{xy} = \pi(y)P_{yx}, \forall x, y \in S$$

we know that

Detailed balance condition $\Longrightarrow$ stationary distribution

Detailed balance condition $\overset{\text{tridiagonal transition matrix}}{\Longleftarrow}$ stationary distribution

Time-reversed Chain

For fixed n,

$$Y_m = X_{n-m}, m = 0, 1, \ldots, n$$

Proposition 1.25.16 $\{Y_m\}$ is a DTMC if X starts from a stationary distribution.

Definition 1.25.4 — Time-reversable. X is called time-reversable if $\{X_n\} \overset{d}{=} \{Y_m\}$.

Theorem 1.25.17 Time-reversible $\iff$ detailed balance condition holds

Metropolis-Hastings Algorithm (Omitted)

1.25.4 Exit Distribution and Exit Time

Exit Distribution

We have

$$S = C \cup A \cup B$$

$$h(x) = \mathbf{P}_x(V_A < V_B)$$

is the unique solution of the system of equations

$$\begin{cases} h(x) = \sum_y P_{xy} h(y) & x \in C \\ h(a) = 1 & x \in A \\ h(b) = 0 & x \in B \end{cases}$$

Matrix Formula: In the proof we have seen $h' = (I-Q)^{-1} R'_A$, where

$$\begin{array}{cc} & \begin{array}{ccc} C & & A \quad B \end{array} \\ \begin{array}{c} C \\ \\ A \\ B \end{array} & \begin{pmatrix} Q & | & R \\ -- & -- & -- \\ & | & \\ & | & \end{pmatrix} \end{array} \qquad R'_A = \begin{pmatrix} \sum_{y \in A} P_{x_1, y} \\ \sum_{y \in A} P_{x_2, y} \\ \vdots \end{pmatrix} = \begin{pmatrix} \mathbf{P}(X \in A | X_0 = x_1) \\ \mathbf{P}(X \in A | X_0 = x_2) \\ \vdots \end{pmatrix}$$

Exit Time

$$S = A \cup S$$

we have $g(x) = \mathbb{E}_x(V_A)$ is the unique solution of

$$\begin{cases} g(x) = 1 + \sum_y P_{xy} g(y) & x \in C \\ g(a) = 0 & a \in A \end{cases}$$

and its **matrix form:**

$$g' = (I-Q)^{-1} \mathbf{1}'$$

1.25.5 Two Examples

Simple Random Walk Without Barrier

1. $p \neq \frac{1}{2} \implies$ transient
2. $p = \frac{1}{2} \implies$ null recurrent

If there is a reflecting barrier, $p < \frac{1}{2} \implies$ positive recurrent.

Branching Process

Definition 1.25.5 — Generating Function.

$$\varphi(s) = \mathbb{E}(s^{\eta}) = \sum_{k=0}^{\infty} P_k s^k, s \in [0,1]$$

Need to know the properties of g.f.s

Galton-Watson Process

Given $Y \sim \{P_n\}$ as the distribution of number of offsprings for 1 individual.

$$\mathbb{E}(Y) = \mu \Longrightarrow \mathbb{E}(X_n) = \mu^n$$

Given assumption $X_0 = 1$, we have

$$u_n = \varphi(u_{n-1})$$

extinction before