STAT 333 WINTER 2021

Applied Probability

(Stochastic Process 1)

Instructor: Yi Shen

Lecture Notes

by Justin Li

Contents

1. Preparation

1.1 Probability Space

Definition 1.1.1 — Probability Space. consists of a triplet $(\Omega, \mathscr{E}, \mathbf{P})$ where:

1. Ω is the **sample space**, the collection of all possible outcomes of a random experiment. e.x.$\{1,2,....,6\}$, $\{H,T\}$, $\{$sunny, cloudy, rainy$\}$

2. $\mathscr{E}$ is the **σ-algebra** or **σ-field**, the collection of all the **"events"**. An **event** E is a subset of Ω, for which we can talk about probability. e.x. $\{1,3,5\} \subseteq \{1,2,....,6\}$

3. $\mathbf{P}$ is the **probability (measure)**, a set function (a mapping from events to real numbers). i.e. $\mathbf{P} : \mathscr{E} \to \mathbb{R}$ or $\mathbf{E} \longmapsto \mathbf{P}(\mathbf{E})$. A probability needs to satisfy the probability axioms:

   (a) $\forall \mathbf{E} \in \mathscr{E},\ 0 \leq \mathbf{P}(\mathbf{E}) \leq 1$

   (b) $\mathbf{P}(\Omega) = 1$

   (c) For countable, disjoint events E_1,E_2,....., we have

$$\mathbf{P}\left(\bigcup_{i=1}^{\infty} \mathbf{E_i}\right) = \sum_{i=1}^{\infty} \mathbf{P}(\mathbf{E_i})$$

Definition 1.1.2 — Random Variable.
A **random variable** $\mathbf{X}$ is mapping from Ω to $\mathbb{R}$. i.e. $\mathbf{X} : \Omega \to \mathbb{R}$ or $\omega \longmapsto \mathbf{X}(\omega)$

1.2 Stochastic Processes

Basic understading:

Process: change/evolve over time.

Stochastic: random

Then we can say it as

1. A sequence/family of random variables (simple, take it as the definition)
2. A random function ((hard to formulate)

Definition 1.2.1 — Stochastic Process.
A stochastic process $\{X_t\}_{t\in T}$ is a collection of random variables defined on a common probability space where T is an **index set**. In most cases, T corresponds to **time**. There are two common types of T corresponds to time as below:

$$\textbf{Discrete: } \{0,1,2,.....\}$$
$$\textbf{Continuous: } [0,\infty)$$

In discrete-time case, we typically write $\{X_n\}_{n=0,1,2,...}$

Definition 1.2.2 — States.
The possible values of X_t with $t \in T$ are called the **sates** of the process. Their collection is call the **state space**, denoted by S. The state space can be either discrete or continuous. In this course, we will focus on discrete state space. We can relabel the states in S to get the **standardized state space**:

$S^* = \{0,1,2,.....\}$ **(Countable State Space)** *or* $S^* = \{0,1,2,.....,n\}$ **(Finite State Space)**

1.3 Simple Random Walk

■ **Example 1.1** Let $X_0, X_1, ...$ be independent and identically distributed random variables following certain distribution. Then $\{X_n\}_{n=0,1,...}$ is a stochastic process, and sometimes called **"White noise"**. ■

■ **Example 1.2 — Simple Random Walk.**
Let $X_1, X_2, ...$ be independent and identically distributed with each of them

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

for $i = 1,2,.....$ Define $S_0 = 0$ and $S_1 = X_1$, $S_2 = X_1 + X_2$,....., $S_n = \sum_{i=1}^{n} X_i$ for $n \geq 1$. Then $\{S_n\}_{n=0,1,...}$ is a stochastic process , with state space $S^* = Z$. $\{S_n\}_{n=0,1...}$ is called a **simple random walk**. Note that $S_n = S_{n-1} + X_n$ so

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

Question: Why do we need the notion of stochastic process? Why don't we just look at the joint distribution of $(X_1, ..., X_n)$
Answer: The joint distribution of a large number of random variables is very complicated, because it does not take advantage of the special structure of T, which is time.

For the simple random walk, the full distribution of $(S_0, S_1, ..., S_n)$ is complicated for large n. However, as we have seen, the structure is actually simple if we focus on adjacent terms:

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

where we note that S_n and X_{n+1} are independent. By introducing time into the framework, things can often be greatly simplified. For the simple random walk, we find that if we know S_n, the distribution of S_{n+1} will not depend on the history S_i for $i = 0, ..., n-1$. This is a very useful properties, and it motivate the notion of **Markov Chain**. ∎

2. Discrete-Time Markov Chain

2.1 Review on Conditional Probability

Definition 2.1.1 — Conditional Probability.
The **conditional probability** of an event B given an event A with $P(A) > 0$ is given by

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} = \frac{P(B,A)}{P(A)} \quad \text{where P(A,B) denotes probability of A and B}$$

Theorem 2.1.1

Let $A_1, A_2, \ldots$ be **disjoint** events s.t. $\bigcup_{i=1}^{\infty} A_i = \Omega$ (we say A_i the **partition** of Ω)

1. Law of total probability: $P(B) = \sum_{i=1}^{\infty} P(B \mid A_i) \cdot P(A_i)$

Proof: Note that $B \cap A_i$ are disjoint and $\bigcup_{i=1}^{\infty} (B \cap A_i) = B$, hence

$$\begin{aligned} P(B) &= \sum_{i=1}^{\infty} P(B \cap A_i) \\ &= \sum_{i=1}^{\infty} P(B \mid A_i) \cdot P(A_i) \end{aligned}$$

2. Bayes' Rule : $P(A_i \mid B) = \dfrac{P(B \mid A_i) \cdot P(A_i)}{\sum_{j=1}^{\infty} P(B \mid A_j) \cdot P(A_j)}$

Proof:

$$P(A_i \mid B) = \frac{P(A_i \cap B)}{P(B)} \quad \textbf{by definition of conditional probability}$$
$$= \frac{P(B \mid A_i) \cdot P(A_i)}{\sum_{j=1}^{\infty} P(B \mid A_j) \cdot P(A_j)} \quad \textbf{by law of total probability}$$

■ **Remark 2.1** Recall that two events A and B are called **independent**, if $P(A \cap B) = P(A) \cdot P(B)$
We say A and B are **independent** by $A \perp\!\!\!\perp B$. If $A \perp\!\!\!\perp B$ and $P(A) > 0$, then we have $P(B \mid A) = P(B)$

2.2 Discrete-Time Markov Chains

Definition 2.2.1 — Discrete-Time Markov Chains.
A **discrete-time** stochastic process $\{X_n\}_{n=0,1\ldots}$ is called a **discrete-time markov chains (DTMC)** with transition matrix $P = \{P_{ij}\}_{i,j \in S^*}$ (S^* is state space), if for any n and any $i, j, i_{n-1}, \ldots, i_0 \in S^*$

$$P(\underbrace{X_{n+1} = j}_{\textbf{future}} \mid \underbrace{X_n = i}_{\textbf{current state}}, \underbrace{X_{n-1} = i_{n-1}, \ldots, X_0 = i_0}_{\textbf{history/past}}) = P_{i,j}$$

Intuition: given the present/current state, the history and the future are **independent**. Equivalently, the past influences the future only through the current state.

■ **Remark 2.2** More generally, the Markov property can be defined as

$$P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \underbrace{P(X_{n+1} = j \mid X_n = i)}_{\textbf{can depend on n}}$$

In addition to this property, the definition we use requires that $P(X_{n+1} = j \mid X_n = i)$ does not depend on n. This is called **time-homogeneity**. In this course, we always focus on the **time-homogeneity DTMC** (default setting).

2.3 Transition Matrix

$$P = \{P_{ij}\}_{i,j \in S^*} = \begin{pmatrix} P_{00} & P_{01} & \ldots & P_{0j} & \ldots \\ P_{10} & P_{11} & \ldots & P_{1j} & \ldots \\ \vdots & \vdots & & \vdots & \\ P_{i0} & P_{i1} & \ldots & P_{ij} & \ldots \\ \vdots & \vdots & & \vdots & \end{pmatrix}$$

where **i** is row : **starting/initial/current state** and **j** is column : **ending /target/net state**

Properties of (one-step) transition matrix
1. $P_{ij} \geq 0$ for all $i, j \in S$
2. Row sums of P are always 1, $\sum_{j \in S} P_{ij} = 1$ for all $i \in S$

$$\textbf{Reason}: \sum_{j \in S} P_{ij} = \sum_{j \in S^*} P(X_{n+1} = j \mid X_n = i) = P(X_{n+1} \in S \mid X_n = i) = 1$$

■ Example 2.1 — Simple Random Walk revisited.
For $P_{ij} = P(S_{n+1} = j \mid S_n = i)$, recall that $S_{n+1} = \underbrace{S_n + X_{n+1}}_{\perp\!\!\!\perp}$ with

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

Then we have

$$\begin{aligned} P_{ij} &= P(S_n + X_{n+1} = j \mid S_n = i) \\ &= P(X_{n+1} = j - i \mid S_n = i) \\ &= P(X_{n+1} = j - i) \quad \textbf{since they are } \perp\!\!\!\perp \\ &= \begin{cases} p & \textbf{if } j = i+1 \\ 1-p & \textbf{if } j = i-1 \\ 0 & \textbf{otherwise} \end{cases} \end{aligned}$$

That is

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

Note that we have not shown that the simple random walk is **DTMC**, this is left as an exercise. ■

Exercise 2.3.1 A simple random walk is Discrete-Time Markov Chain. ■

■ Example 2.2 — Ehrenfest's Urn.
We have two urn A, B, there are total M balls in urn, Each time, we pick one ball uniformly randomly and put it into the opposite urn. Let's define:

1. X_n number of balls in urn A after step n
2. State Space: $S = \{0, ..., M\}$
3. $X_n = i$ means i balls in A and $M - i$ balls in B

4.

$$
\begin{aligned}
P_{ij} &= P(X_{n+1} = j \mid X_n = i) \\
&= \begin{cases} \dfrac{i}{M} & \textbf{if } j = i+1 \textbf{ (The ball is from } A\textbf{)} \\ \dfrac{M-i}{M} & \textbf{if } j = i-1 \textbf{ (The ball is from } B\textbf{)} \\ M & \textbf{if } j \neq i \pm 1 \end{cases}
\end{aligned}
$$

Then, the transition matrix in $\mathbb{R}^{(M+1)\times(M+1)}$ is

$$
P = \begin{pmatrix}
0 & 1 & & & & \\
\frac{1}{M} & 0 & \frac{M-1}{M} & & & \\
 & \frac{2}{M} & 0 & \frac{M-2}{M} & & \\
 & & \frac{3}{M} & 0 & \ddots & \\
 & & & \ddots & 0 & \frac{1}{M} \\
 & & & & 1 & 0
\end{pmatrix}
$$

■

2.4 Multi-step transition probability

Theorem 2.4.1 — Chapman-Kolmogorov Equation (C-K Equation).
Question: What if we are interested in the behavior of the **DTMC** in n steps rather that 1 step?

That is

$$
P_{ij}^{(n)} = P(X_n = j \mid X_0 = i) = P(X_{m+n} = j \mid X_m = j) = ? \qquad \text{for } m = 1, 2, \ldots
$$

Note that $P_{ij}^{(1)} = P_{ij}$ by definition, start with $n = 2$:

$$
P^{(2)} = P(X_2 = j \mid X_0 = i) = \sum_{k \in S} P(X_2 = j \mid X_0 = i,\, X_i = k) \cdot P(X_1 = k \mid X_0 = i)
$$

This is nothing else but the conditional version of the law of total provability.
Details for $P^{(2)}$:

$$
\begin{aligned}
P(X_2 = j \mid X_0 = i) &= \sum_{k \in S} P(X_2 = j,\, X_i = k \mid X_0 = i) \\
&= \sum_{k \in S} \frac{P(X_2 = j,\, X_1 = k,\, X_0 = i)}{P(X_0 = i)} \\
&= \sum_{k \in S} \frac{P(X_2 = j,\, X_1 = k,\, X_0 = i)}{P(X_1 = k,\, X_0 = i)} \cdot \frac{P(X_1 = k,\, X_0 = i)}{P(X_0 = i)} \\
&= \sum_{k \in S} P(X_2 = j \mid X_0 = i,\, X_i = k) \cdot P(X_1 = k \mid X_0 = i) \\
&= \sum_{k \in S} P_{ik} \cdot P_{kj} \\
&= (P \cdot P)_{ij} = (P^2)_{ij}
\end{aligned}
$$

Thus, if we have $P^{(2)} = \left\{P_{ij}^{(2)}\right\}_{ij\in S}$ where $P^{(2)}$ is 2-step transition matrix. Then $P^{(2)} = P^2$

Now, in general, for $n, m = 0, 1, 2...$

$$\begin{aligned}
P_{ij}^{(m+n)} &= P(X_{m+n} = j \mid X_0 = i) \\
&= \sum_{k\in S} P(X_{m+n} = j \mid X_m = k,\ X_0 = i) \cdot P(X_{m+n} = j \mid X_0 = i) \\
&= \sum_{k\in S} P(X_{m+n} = j \mid X_m = k) \cdot P(X_m = k \mid X_0 = i) \\
&= \sum_{k\in S} P_{kj}^{(n)} \cdot P_{ik}^{(m)} \\
&= \sum_{k\in S} P_{ik}^{(m)} \cdot P_{kj}^{(n)} \\
&= (P^{(m)} \cdot P^{(n)})_{ij}
\end{aligned}$$

This gives us that $P^{(m+n)} = P^{(m)} \cdot P^{(n)}$, we called **C-K Equation**

As a result , we have

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

This gives us that

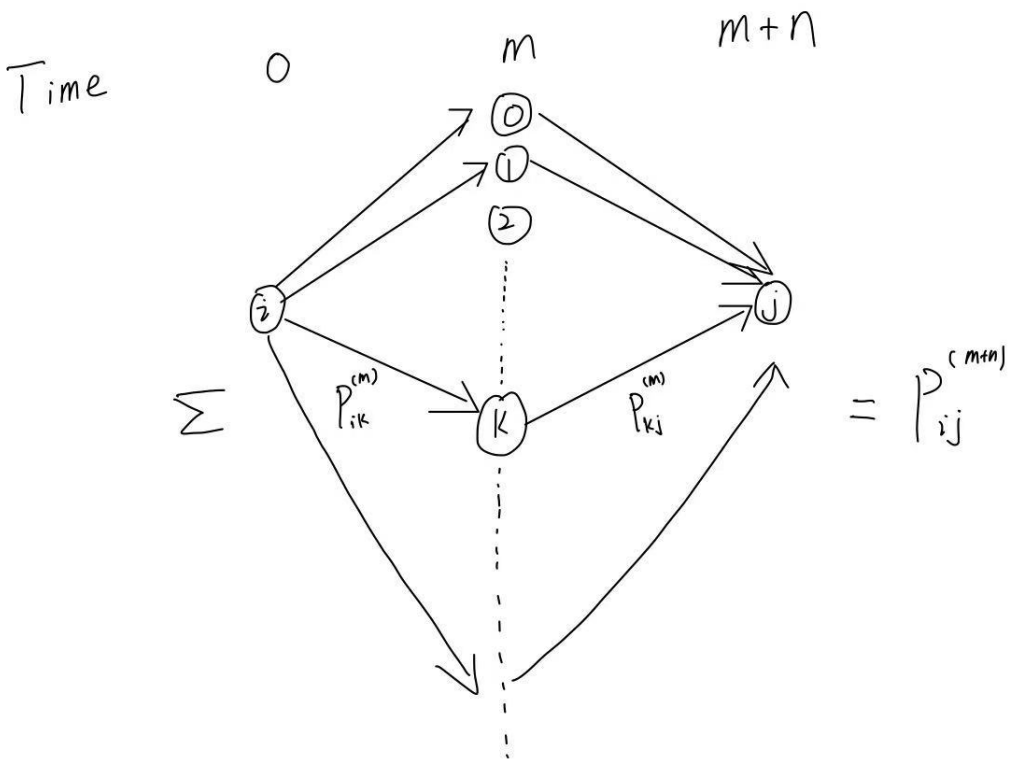
$$P^{(n)} = P^n$$

for all $n \in \mathbb{N}$, but we should know the $P^{(n)}$ is n-step transition matrix with $P^{(n)} = \left\{P_{ij}^{(n)}\right\}_{i,j\in S}$ and $P_{ij}^{(n)} = P(X_n = j \mid X_0 = i)$. P^n is the n-th power of the one step transition matrix $P^n = \underbrace{P \cdot P \cdots P}_{\textbf{n terms}}$

Intuition behind C-K Equation:
Condition at time m (on X_m) and sum up all the probabilities

2.5 Visualization of Discrete Time Markov Chain

DTMC can be presented by weighted directed graph
1. **states** $\longrightarrow$ **nodes**
2. possible (one-step) transition $\longrightarrow$ **edges**: draw an edge from i to j iff $P_{ij} > 0$
3. transition probability $\longrightarrow$ weight.



■ **Example 2.3** Let $S = \{0,1,2\}$ with

$$P = \begin{array}{c} \\ 0 \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{ccc} 0 & 1 & 2 \end{array} \\ \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \\ \frac{2}{5} & \frac{3}{5} & 0 \end{pmatrix} \end{array}$$

Here is the directed graph:

■

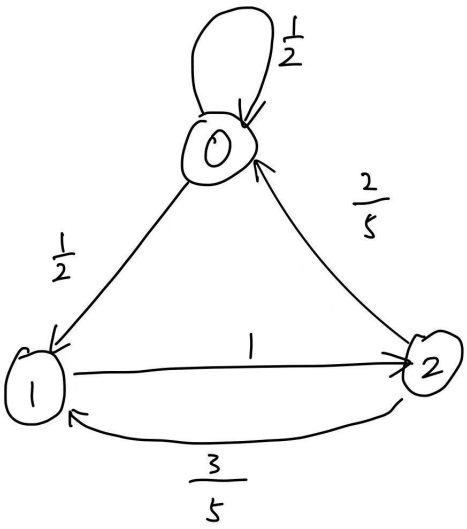
■ **Example 2.4 — Simple Random Walk Again.**

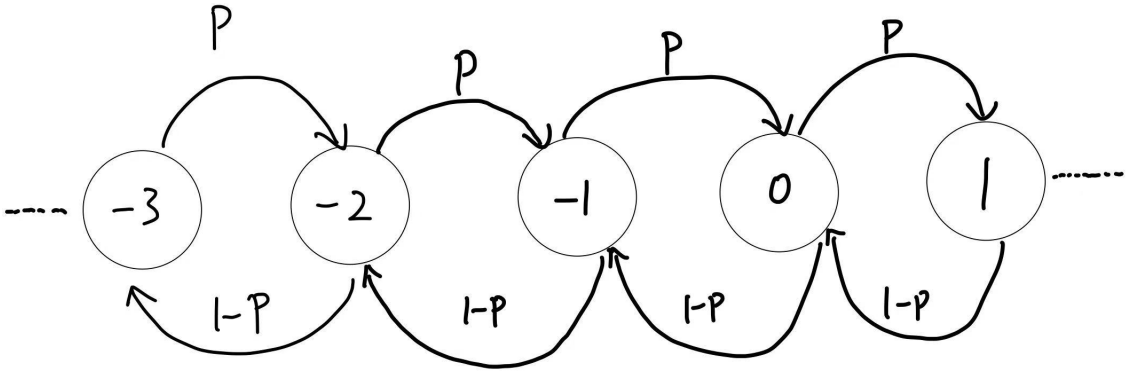
■

2.6 Distribution of X_n

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

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





If the **DTMC** starts from i for sure (i.e. $P(X_0 = i) = 1$), then the $P_{ij}^{(n)}$ is also the probability that $X_n = j$. Hence $\left\{P_{ij}^{(n)}\right\}_{j\in S}$ is the distribution of X_n. That is the i-th row of the $P^{(n)}$ is the distribution of X_n if the chain starts from i

$$P^n = \begin{pmatrix} & & & \\ - & - & - & i & - & - & - \\ & & & \end{pmatrix}$$

what if the chain has a random starting state?

Definition 2.6.1 — Initial Distribution. Let $\mu(i) = P(X_0 = i)$, then the vector $\mu = (\mu(0), \mu(1), \cdots, \mu(i))$ gives the distribution of X_0 and this called the **initial distribution** of this **DTMC**

■ **Remark 2.3** This is the initial distribution of the initial state X_0.

Definition 2.6.2 Similarly, we can define $\mu_n = (\mu_n(0), \mu_n(1), \cdots)$ to be the distribution of X_n, where $\mu_n(i) = P(X_n = i)$. Here we think μ_n as a **row vector** representing a distribution. Sometimes, we also write $\mu_n(X_n = i)$, in this case we think μ_n as probability.

■ **Remark 2.4** Note $\mu_0 = \mu$

■ **Remark 2.5 — Property of μ_n.**
The row vector μ_n represent a distribution, hence we have :
1. $\mu_n(i) \geq 0$ for any $i \in S$
2. $\sum_{i \in S} \mu_n(i) = 1$ i.e. $\mu_n \cdot \mathbb{I} = 1$ **where** $\mathbb{I} = [1, 1, 1, \cdots]^T$

Fact 2.6.1 Given μ and P, we have $\mu_n = \mu \cdot P^n$
Proof: For any $j \in S$:

$$\begin{aligned}
\mu_n(j) = P(X_n = j) &= \sum_{\textbf{all i}} P(X_n = j \mid X_0 = i) \cdot P(X_0 = i) \\
&= \sum_{\textbf{all i}} \mu(i) \cdot P_{ij}^{(n)} \\
&= (\mu \cdot P^{(n)})_j \\
&= (\mu \cdot P^n)_j
\end{aligned}$$

Thus, we have $\mu_n = \mu \cdot P^n$

■ **Remark 2.6** We see that the distribution of a **DTMC** is completely determined by two things:
1. Transition matrix P
2. Initial distribution μ

3. Expectation of DTMC

3.1 Conditional Expectation and $E(f(X_n))$

Definition 3.1.1 — Conditional Expectation.
Let X, Y be the discrete random variables. If $P(Y = y) > 0$, then the conditional distribution of X given $Y = y$ is defined by

$$P(X = x \mid Y = y) = \frac{P(X = x,\ Y = y)}{P(Y = y)}$$

where $P(X = x \mid Y = y)$ is called the **conditional probability mass function** (con'd pmf), denoted by

$$f_{X|Y=y}(x) \quad \textbf{or} \quad f_{X|Y}(x \mid y)$$

Proposition 3.1.1 — Conditional pmf is a a legitimate pmf.
For any y s.t. $P(Y = y) > 0$

$$\forall x,\ \ f_{X|Y=y}(x) \geq 0$$

and

$$\sum_{\textbf{all x}} f_{X|Y=y}(x) = 1$$

The continuous case is similar (with density functions replacing mass functions)

■ **Remark 3.1** Since hte conditional pmf is a legitimate pmf, the conditional distribution is also a legitimate probability distribution (just potentially different from the unconditionally distribution). As a result, we can define expectation under the conditional distribution.

Definition 3.1.2 — Conditional Expectation. Let X, Y be the discrete random variables, and g be a

function. Then the conditional expectation of $g(X)$ given $P(Y=y)>0$ is defined as

$$E(g(X) \mid Y=y) = \sum_{\textbf{all x}} g(x)P(X=x \mid Y=y)$$

■ **Remark 3.2** Fixed y, the conditional expectation is nothing else but the expectation under the conditional distribution.

■ **Remark 3.3** Different ways to understand conditional expectations (**very important**)

1. Fix y, $E(g(x)) \mid Y=y)$ is a number.
2. As y changes, $h(y) = E(g(x) \mid Y=y)$ is a function of y.
3. Since Y is a random variable, we can define $E(g(x) \mid Y) = h(Y)$. Thus $E(g(x) \mid Y)$ is a function of Y, hence also a random variable

$$E(g(X) \mid Y)_\omega = E(g(X) \mid Y = Y(\omega)) \quad \omega \in \Omega$$

That is , the random variable $E(g(X) \mid Y)$ takes value $E(g(X) \mid Y=y)$ when $Y=y$.

Proposition 3.1.2 — Linearity of Conditional Expectation.

$$E(aX+b \mid Y=y) = a \cdot E(X \mid Y=y) + b$$

$$E(X+Z \mid Y=y) = E(X \mid Y=y) + E(Z \mid Y=y)$$

Proposition 3.1.3 — Plug in Property.

$$E(g(X,Y) \mid Y=y) = E(g(X,y) \mid Y=y)$$

Proof (discrete case) :

$$\begin{aligned}
E(g(X,Y) \mid Y=y) &= \sum_{x_i}\sum_{y_j} g(x_i,y_i) \cdot E(X=x_i,\ Y=y_j \mid Y=y) \\
&\Longrightarrow \begin{cases} 0 & \textbf{if } y_j \neq y \\ \frac{P(X=x_i,\ Y=y_j)}{P(Y=y)} = P(X=x_i,\ Y=y_j) & \textbf{if } y_j = y \end{cases} \\
&\Longrightarrow E(g(X,Y) \mid Y=y) \\
&= \sum_{x_i} g(x_i,y) \cdot P(X=x_i \mid Y=y) \\
&= E(\underbrace{g(X,y)}_{\textbf{is a function of X}} \mid Y=y)
\end{aligned}$$

In particular,

$$E(g(X)h(Y) \mid Y=y) = E(g(X)h(y) \mid Y=y) = h(y) \cdot E(g(x) \mid Y=y)$$

In the random variable form

$$E(g(X)h(Y) \mid Y=y) = h(Y) \cdot E(g(x) \mid Y=y)$$

Proposition 3.1.4 If $X \perp\!\!\!\perp Y$, then $E(g(X), \mid Y) = E(g(X))$
Reason: independent $\Longrightarrow$ conditional distribution is the same as the unconditional distribution.

Proposition 3.1.5 — Law of Iterated Expectation.

$$E(\underbrace{E(X \mid Y)}_{\text{r.v, function of Y}}) = E(X)$$

Proof (discrete case): When $Y = y_j$, then the random variable

$$E(X \mid Y) = E(X \mid Y = y_j) = \sum_{x_i} P(X = x_i \mid Y = y_j)$$

This happens with $P(Y = y_j)$, thus

$$\begin{aligned} E(E(X \mid Y)) &= \sum_{y_j} E(X \mid Y = y_j) \cdot P(Y = y_j) \\ &= \sum_{y_j} \left(\sum_{x_j} x_i \cdot P(X = x_i \mid Y = y_j) \right) P(Y = y_j) \\ &= \sum_{x_i} x_i \sum_{y_j} P(X = x_i \mid Y = y_j) \cdot P(Y = y_j) \\ &= \sum_{x_i} x_i \cdot P(X = x_i) \quad \textbf{by law of total probability} \\ &= E(X) \end{aligned}$$

■ **Example 3.1** Let Y be the number of claims received by an insurance company, X be some random parameter, such that

$$Y \mid X \sim Poi(X),\ X \sim Exp(\lambda)$$

Find $E(Y)$
Solution: Note that

$$E(Y) = E(E(Y \mid X))$$

Since $Y \mid X \sim Poi(X)$ for any x, then $E(Y \mid X = x) = x$, so we have $E(Y \mid X) = X$. Hence, we have

$$E(Y) = E(E(Y \mid X)) = E(X) = \frac{1}{\lambda}$$

as desired. ■

3.2 Expectation of $f(X_n)$

Let f be a function from S to $\mathbb{R}$. Think $f(X_n)$ as a reward/penalty we receive at step n according to the state. Average reward/penalty at step n is $E(f(X_n))$

How to find $E(f(X_n))$?

Approach 1:

$$E(f(X_n)) = \sum_{i\in S} f(i)\cdot \underbrace{P(X_n = i)}_{\mu_n(i)} = \sum_{i\in S} f(i)\mu_n(i) = \mu_n \cdot f^T$$

where $f = (f(0), f(1), ...)$ is the row vector giving all the values of f in different states. Recall that $\mu_n = u\cdot P^n$ where μ is row vector represent the initial distribution and P is transition matrix.
Then we get

$$E(f(X_n)) = \mu_n \cdot f^T = \mu \cdot P^n \cdot f^T$$

Question: What happens if we calculate $P^n \cdot f^T$ first?
Answer: This correspond to finding $E(f(X_n) \mid X_0 = i)$ where $i \in S$ first. Approach 2:

$$\begin{aligned} E(f(X_n)) &= E(E(f(X_n) \mid X_0)) \\ &= \sum_{i\in S} E(f(X_n) \mid X_0 = i)\cdot P(X_0 = i) \\ &= \sum_{i\in S} E(F(X_n) \mid X_0 = i)\cdot \mu(i) \\ &\coloneqq \sum_{i\in S} f^n(i)\cdots \mu(i) \\ &= \mu \cdot \left(f^{(n)}\right)^T \end{aligned}$$

where

$$\left(f^{(n)}\right)^T = \begin{pmatrix} E(f(X_n) \mid X_0 = 0) \\ E(f(X_n) \mid X_0 = 1) \\ \vdots \end{pmatrix}$$

How can we find $\left(f^{(n)}\right)^T$?

$$\begin{aligned} (f^{(n)})^T(i) &= E(f(X_n) \mid X_0 = i) \\ &= \sum_j f(j)\cdot P(X_n = j \mid X_0 = i) \\ &= \sum_j P_{ij}^{(n)} \cdot f(j) \\ &= \left(P^n \cdot f^T\right)_i \end{aligned}$$

Thus, $\left(f^{(n)}\right)^T = P^n \cdot f^T$. Going back to $(*)$, we have

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

which agrees with what we get from approach 1.
In summary, the $E(f(X_n)) = \mu \cdot P^n \cdot f^T$, calculate μP^n first: $E(f(X_n)) = \mu_n \cdot f^T$ starting at n, find the distribution of X_n. If calculate $P^n \cdot f^T$ first:

$$E(f(X_n)) = \mu \cdot \begin{pmatrix} E(f(X_n) \mid X_0 = 0) \\ E(f(X_n) \mid X_0 = 1) \\ \vdots \end{pmatrix}$$

starting at 0, find $E(f(X_n) \mid X_0 = i)$ where $i \in S$. In both cases, row vectors are distributions: $\mu, \mu_1, \mu_2 ...$ while column vectors are functions f^T, $\left(f^{(n)}\right)^T$

3.3 Stationary Distribution

Definition 3.3.1
A probability distribution $\pi = (\pi_0, \pi_1, ...)$ is called a **stationary distribution** (invariant distribution) of a **DTMC** $\{X_n\}_{n=0,1,2,..}$ with transition matrix P if:
(1) $\pi = \pi \cdot P$ $\longleftarrow$ systems of equations (Stationary Condition)
(2) $\sum_{i \in S} \pi_i = 1$ $\longleftarrow$ Normalization Condition, also written as $\pi \cdot \mathbb{I} = 1$ where $\mathbb{I} = (1, 1, 1..., 1)^T$

Why such π is called stationary?
Assume the DTMC starts from the initial distribution $\mu = \pi$, then the distribution of X_1 is $\mu_1 = \mu \cdot P = \pi P$
The distribution of X_2: $\mu_2 = \mu \cdot P^2 = \pi \cdot P \cdot P = \pi = \mu$ and so on. Thus $\mu_n = \mu = \pi$. If the **DTMC** starts from a stationary distribution, its distribution will never change. $\Longrightarrow$ "Stationary" ("Invariant")

■ **Example 3.2** An election has two states: ground state 0 and excited state 1. Let $X_n = \{0, 1\}$ be its state at time n. At each step, the election changes state with probability α if it's in state 0, with probability β if it's in state 1. Then $\{X_n\}$ is a **DTMC**, it's transition matrix is

$$P = \begin{matrix} \\ 0 \\ 1 \end{matrix} \begin{matrix} \begin{matrix} 0 & 1 \end{matrix} \\ \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \end{matrix}$$

Goal: Solve for the stationary distribution.
Solution: solve $\pi = \pi \cdot P$:

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

Then we have

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

We have two equations and two unknows; however, note that they are linearly dependent. From (1), we have $\alpha\pi_0 = \beta\pi_1 \Longrightarrow \frac{\pi_0}{\pi_1} = \frac{\beta}{\alpha}$. However, we don't get their values by only using (1). This is where we need $\pi_0 + \pi_1 = 1$, then

$$\pi_0 = \frac{\beta}{\alpha+\beta} \quad \textbf{and} \quad \pi_1 = \frac{\alpha}{\alpha+\beta}$$

Then, there exists one unique stationary distribution

$$\pi = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta} \right)$$

■

■ **Remark 3.4** The may to solve for the stationary distribution in the example above is typical.
1. Use $\pi = \pi \cdot P$ to get proportion among different component of π.

$$(--\pi--) \cdot \begin{pmatrix} \vdots \\ ----P---- \\ \vdots \end{pmatrix} = (--\pi--)$$

Reason: (1) The system of equation is always linearly dependent. (2) This system of equation is homogeneous. (if π is a solution, then $a\pi = (a\pi_0, a\pi_1, ...)$ is also a solution)
2. Use $\pi \cdot \mathbb{I} = 1$ to normalize to get exact values.

■ **Remark 3.5** We note that $\pi = \pi P$, this implies that π is the transpose of an eigenvector of P with eigenvalue 1.

Question: Existence and uniqueness of π? Converge to π?

4. Properties of DTMC

4.1 Recurrence and Transience

Definition 4.1.1 — First (Re)Visit to y.
Let $y \in S$ be a state, define T_y to be

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

to be the time of the first (re)visit to y, and we define

$$\rho_{yy} = P_y(\underbrace{T_y < \infty}_{\text{the DTMC ever (re)visit y}}) = P(T_y < \infty \mid X_0 = y)$$

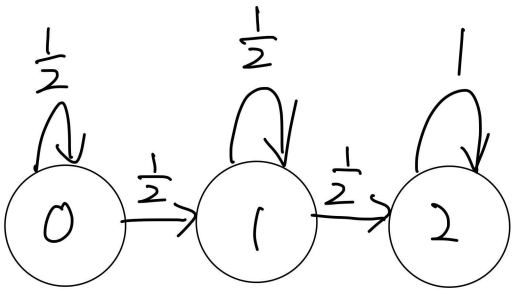
where P_y is the conditional probability under the condition $X_0 = y$

Definition 4.1.2 — Recurrent and Transient State.
A state $y \in S$ is called **recurrent**, if the $\rho_{yy} = 1$ (always return to y), a state $y \in S$ called **transient**, if $\underbrace{\rho_{yy} < 1}_{1-\rho_{yy}=P_y(T_y=\infty)>0}$ (never revisit y again with pass probability)

■ **Example 4.1** Consider a **DTMC** with

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

Graphical Representation:



Assume $X_0 = 0$, two possibilities for X_1:

$$X_1 = 0 \Longrightarrow T_0 = 1$$
$$X_1 = 1 \Longrightarrow T_0 = \infty \qquad \textbf{because states 1 and 2 will never go to } 0$$

Then $\rho_{00} = P_0(T_0 < \infty) = P_0(X_1 = 0) = \frac{1}{2} < 1$, so by definition the state 0 is transient. Similarly, the state 1 is also transient. Assume $X_0 = 2$, then $P(X_1 = 2 \mid X_0 = 2) = 1$, so $T_2 = 1$. Then $\rho_{22} = P_2(T_2 < \infty) = 1$, by definition the state 2 is recurrent. ∎

∎ **Remark 4.1** This is an example where recurrence and transience can be directly checked by definition. However, this is very rare, as the distribution of T_i is very hard to deride in general. Thus, we need better criteria for recurrence/transience.

4.2 Communication

Definition 4.2.1 — x communicates to y.
Let $x, y \in S$ (possible the same state). x is said to **communicate** to y, or "y is **accessible** from x". Denoted by $x \to y$, if starting from x, the probability that the chain eventually (re)visits state y is positive. i.e.

$$\rho_{xy} \coloneqq P_x(T_y < \infty) > 0$$

Note that this is equivalent to say

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

or say “x can go to y".

Lemma 4.2.1 — Transitivity of Communication.
If $x \to y$ and $y \to z$, then $x \to z$.

Proof: $x \to y \Longrightarrow \exists m \geq 1$ s.t. $P^m_{xy} > 0$ and $y \to z \Longrightarrow \exists n \geq 1$ s.t. $P^n_{yz} > 0$. Then by **C-K equation**

$$P^{m+n}_{xz} = \sum_{k \in S} P^m_{xk} P^n_{kz} \geq \underbrace{P^m_{xy}}_{>0} \underbrace{P^n_{yz}}_{>0} > 0$$

Then we have $x \to z$

Intuition: $P^m_{xy} P^n_{yz}$ specifies one way to go from x to z in $m+n$ steps (via y). While the quantity P^{m+n}_{xz}

is the total probability to go from x to z in $m+n$ steps.

Theorem 4.2.2
If $\rho_{xy} > 0$ but $\rho_{yx} < 1$, then x is transient

Proof: Define $k = \min\{k : P_{xy}^k > 0\}$ to be the smallest length of a path from x to y. Since $P_{xy}^k > 0$, there exists states $y_1, ..., y_{k-1}$ s.t.

$$P_{xy_1} P_{y_1 y_2} \dots P_{y_{k-1} y} > 0$$

Moreover, none of $y_1, ..., y_{k-1}$ is x, since otherwise this is not the shortest path. Once we are in state y, with probability $1 - P_{yx} > 0$, we will never go back to x.

$$P(T_x = \infty) \geq P_{xy_1} P_{y_1 y_2} \dots P_{y_{k-1} y} \cdot (1 - P_{yx})$$

where $P_{xy_1} P_{y_1 y_2} \dots P_{y_{k-1} y}$ is the path going from x to y without returning to x and $(1 - P_{yx})$ corresponds to the idea of once in y, it never goes back. This is one way not to visit x again (via y). Then $P_{xx} = P_x(T_x < \infty) < 1$, so x is transient.

Corollary 4.2.3
If x is recurrent and $\rho_{xy} > 0$, then $\rho_{yx} = 1$. (the contrapositive of the theorem 2.11.2)

Definition 4.2.2 — Communicating Class.
A set of states $C \subseteq S$ is called a **communicating class**, if it satisfies the followings:

1. $\forall i, j \in C$, $i \to j$ and $j \to i$
2. $\forall i \in C$ and $j \notin C$, $i \nrightarrow j$ or $j \nrightarrow i$

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

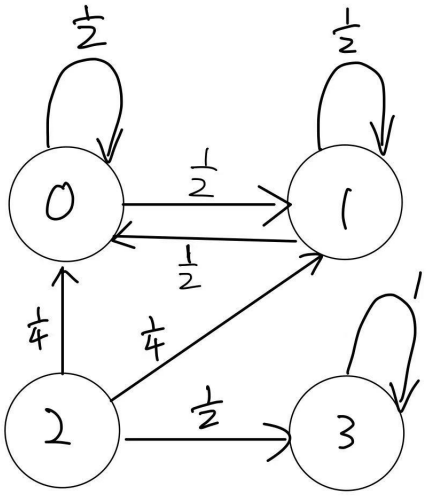
Communication and Communicating Class in graphs:

$i \to j$: we can go from i to j by following the arrows (directed edges)

How to find the classes: "find the loops"

■ **Example 4.2**

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



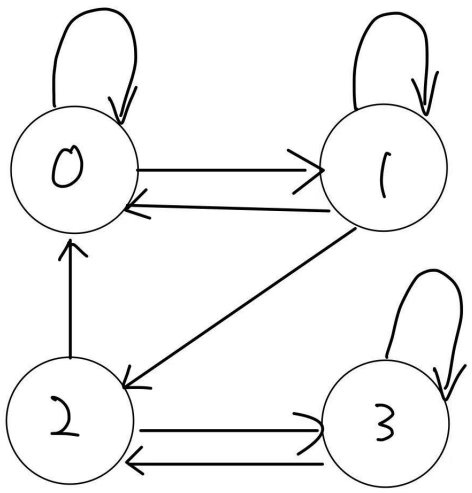
$P_{01} > 0$, $P_{10} > 0 \quad \Longrightarrow 0, 1$ **in a same class**

State 2 not in any class, because $\forall i \in S$, $P_{i2} = 0$

State 3 forms its own class, because $3 \to 3$, **but** $\forall i \in S$ **and** $i \neq 3$, $P_{3i=0}$

Therefore, there are two classes: $\{0,1\}$ and $\{3\}$, but state 2 which does not belong to any class. ∎

■ **Example 4.3**



$$P_{01},\ P_{12},\ P_{20} > 0 \quad \Longrightarrow \quad 0,\ 1,\ 2 \ \textbf{in the same class}$$
$$P_{23} > 0,\ P_{32} > 0 \quad \Longrightarrow \quad 2,\ 3 \ \textbf{in the same class}$$

Then, by transitivity we have $\{0,\ 1,\ 2,\ 3\}$ all in the same class. ■

Definition 4.2.3 — Irreducible Markov Chain.
A **DTMC** is called **irreducible**, if all states are in the same class. In other words:

$$\forall\, i,\ j\ \in S\,,\ \ i \longleftrightarrow j$$

A set B is called irreducible, if $i \longleftrightarrow j$, $\forall i, j \in B$

Theorem 4.2.4
Let $i, j \in C$ be in a same communicating class C, then j is recurrent/ transient if and only if i is recurrent/ transient ("Recurrent/ Transient are class properties")

Proof: will be given later in the class.

■ **Remark 4.2** As a result of this theorem, we can call a class recurrent/ transient, if all its states are recurrent/ transient. $\iff$ one state in the class is recurrent/ transient.
Hence, in order to check if a class is recurrent/ transient, we just need to check one state in that class.

Definition 4.2.4 — Closed Set.
A set A of states is called **closed**, if $i \in A, j \notin A \Longrightarrow P_{ij} = 0 \iff \ \ i \in A,\ j \notin A \Longrightarrow i \nrightarrow j$

4.3 Decomposition of the State Space

Theorem 4.3.1
The state space S can be written as a disjoint union

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

where T is the set of all transient states. (not necessarily one class), and R_i for $i = 1, 2...$ are closed recurrent classes.

Proof: First, collect all the transient states and put them into T. For each recurrent state, it must belong to a recurrent class, since at least if communicates to itself. Take one class containing it. Put these classes together and remove the identical ones. Denote the left by $R_1, R_2, ...$, then we have

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

What is left to prove: (1) $R_1, R_2, ...$ are disjoint and (2) $R_1, R_2, ...$ are closed.
For (1) Suppose there are R_m, R_n such that $R_m \cap R_n \neq \emptyset$. Let $i \in R_m \cap R_n$, then for any $j \in R_m$ and $k \in R_n$ we have $i \longleftrightarrow j$ and $i \longleftrightarrow k \underbrace{\Longrightarrow}_{\textbf{transitivity}} j, k$ in the same class. Since this holds for any $j \in R_m$ and $k \in R_n$, R_n and R_m are the same class, contradicting the construction. Then we have $R_1, R_2, ...$ are disjoint.

For (2) Suppose there exists R_k which is not closed, then exists $i \in R_k$ and $j \notin R_k$ such that $P_{ij} > 0 \implies i \to j \iff P_{ij} > 0$ Since $j \notin R_k$ and ij. Hence, $ji \iff P_{ji} = 0 < 1$. Then, the state i is transient. **Contradiction!**

4.4 More Recurrence and Transience

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

Theorem 4.4.1 — Strong Markov Property of (time-homogeneous) DTMC.
The process $\{X_{T_y+k}\}_{k=0,1,..}$ behaves like **DTMC** with initial state y. (Forget the history and restart from y)

Note: X_{T_y+k} is a random variable defined as $X_{T_y+k}(\omega) = X_{T_y(\omega)+k}(\omega)$

Proof: (Use T for T_y) It's suffices to show

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

for all $n,\ x_{n-1}, \ldots,\ x_1 \neq y,\ x_0 \in S$. Indeed, we have

$$\begin{aligned}
&P(X_{T+1} = z \mid X_T = y,\ T = n,\ X_{n-1} = x_{n-1}, \ldots,\ X_0 = x_0) \\
&= P(X_{n+1} = z \mid X_T = y,\ T = n,\ X_{n-1} = x_{n-1}, \ldots,\ X_0 = x_0) \\
&= P(X_{n+1} = z \mid X_n = y,\ X_{n-1} = x_{n-1}, \ldots,\ X_0 = x_0) \\
&= P(X_{n+1} = z \mid X_n = y) \quad \textbf{by property of DTMC} \\
&= P_{yz}
\end{aligned}$$

Definition 4.4.1
Define $T_y^1 = T_y$ and for $k \geq 2$, define

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

to be the time of the $k-th$ (re)visit of state y.

By **Strong Markov Property**

$$P_y(T_y^k < \infty) = (\rho_{yy})^k$$

where T_y^k is (re)visits y for at least k times and $(\rho_{yy})^k$ represents revisits y for the first time
There are two possibilities:

1. y is **transient** $\iff \rho yy < 1$, then $\rho_{yy}^k \to a$ as $k \to \infty$, hence

$$\begin{aligned}
P_y(\text{visits y for infinite number of times}) = 0 &\iff P_y(\text{only visits y for finite number of times}) = 1 \\
&\iff P_y(\text{there exists a last visit to y}) = 1
\end{aligned}$$

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

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

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

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

This is the **c.d.f** of $\textbf{Geo}(1-\rho_{yy})$, hence

$$N(y) \mid X_0 = y \sim \textbf{Geo}(1-\rho_{yy})$$

"keep trying until leaving y forever"

$$E_y N(y) = \frac{\rho_{yy}}{1-\rho_{yy}}$$

then

$$\rho_{yy} < 1 \implies E_y N(y) < \infty \qquad \rho_{yy} = 1 \implies E_y N(y) = \infty$$

Therefore, we have

$$y \text{ is transient} \iff E_y N(y) < \infty \quad \textbf{and} \quad y \text{ is recurrent} \iff E_y N(y) = \infty$$

More generally, we have

Lemma 4.4.2

$$E_x N(y) = \frac{\rho_{xy}}{1-\rho_{yy}}$$

Proof:

$$\begin{aligned}
E_x N(y) &= E_x(N(y) \mid T_y < \infty) \cdot P_x(T_y < \infty) + E_x(N(y) \mid T_y = \infty) \cdot P_x(T_y = \infty) \\
&= E_x(N(y) \mid T_y < \infty) \cdot P_x(T_y < \infty) \\
&= \rho_{xy} \cdot (E_y(N(y)) + 1) \\
&= \rho_{xy} \cdot \left(\frac{\rho_{yy}}{1-\rho_{yy}} + 1 \right) \\
&= \rho_{xy} \cdot \frac{1}{1-\rho_{yy}} \\
&= \frac{\rho_{xy}}{1-\rho_{yy}}
\end{aligned}$$

4.5 A short Review of Indicator

Let A be an event, then $\mathbb{I}_A$ is a random variable given by

$$\mathbb{I}_A(\omega) = \begin{cases} 1 & \textbf{If } \omega \in A \\ 0 & \textbf{If } \omega \notin A \end{cases}$$

Then $E(\mathbb{I}_A) = 1 \cdot P(A) + 0 \cdot P(A^c) = P(A)$

Lemma 4.5.1

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

Proof:

$$\begin{aligned}
N(y) = \sum_{n=1}^{\infty} \mathbb{I}_{X_n=y} \implies E_x N(y) = E_x\left(\sum_{n=1}^{\infty} \mathbb{I}_{X_n=y} \right) &= \sum_{n=1}^{\infty} E_x(\mathbb{I}_{X_n=y}) \\
&= \sum_{n=1}^{\infty} P_x(X_n = y) \\
&= \sum_{n=1}^{\infty} P_{xy}^n
\end{aligned}$$

Combining this with the previous results, we have

$$y \text{ is transient} \iff \sum_{n=1}^{\infty} P_{yy}^n < \infty$$

$$y \text{ is recurrent} \iff \sum_{n=1}^{\infty} P_{yy}^n = \infty$$

Theorem 4.5.2 — Transience/Recurrence are Class Properties.

Proof: Assume x, y are in the same class ($x \longleftrightarrow y$) and x is recurrent. We show y is recurrent. Since $x \to y$ and $y \to x$, there are $m, n \in \mathbb{Z}^+$ such that

$$P_{xy}^{(n)} > 0 \quad P_{yx}^n > 0$$

Note that

$$\begin{aligned} P_{yy}^{n+m+k} &= P(X_{n+m+k} = y \mid X_0 = y) \\ &\geq P(X_{n+m+k} = y,\ X_{n+k} = x,\ X_n = x \mid X_0 = y) \\ &= P_{yx}^n \cdot P_{xx}^k \cdot P_{xy}^m \end{aligned}$$

Since x is recurrent, consider $k = l - m - n$ then we have

$$\sum_{l=1}^{\infty} P_{yy}^l \geq \sum_{l=m+n+1}^{\infty} P_{yy}^l = \sum_{k=1}^{\infty} P_{yy}^{n+m+k} \geq \sum_{k=1}^{\infty} P_{yx}^n P_{xx}^k P = \underbrace{P_{yx}^n}_{>0} \underbrace{P_{xy}^m}_{>0} \cdot \underbrace{\left(\sum_{k=1}^{\infty} P_{xx}^k \right)}_{=\infty} = \infty$$

Therefore, y is recurrent, so recurrence is a class property.
As a result, transience is also a class property. (transience $\iff$ not recurrence)

Lemma 4.5.3 — A finite closed set has at least one recurrence state.

Proof: Let C be a finite closed set. Suppose all the states in C are transient. Then for any states $x, y \in C$ we have

$$E_x(N(y)) = \frac{\rho_{xy}}{1 - \rho_{yy}} < \infty \implies \sum_{y \in C} E_x(N(y)) < \infty$$

However,

$$\sum_{y \in C} E_x(N(y)) = E_x\left(\sum_{y \in C} N(y) \right) = E_x\left(\sum_{y \in C} \sum_{n=1}^{\infty} \mathbb{I}_{X_n = y} \right) = E_x\left(\sum_{n=1} \sum_{y \in C} \mathbb{I}_{X_n = y} \right)$$

Since C is closed, starting from x, at any times n, X_n must be in one of the states in C. Hence, one indicator takes value 1, the rest are 0, so

$$\sum_{y \in C} \mathbb{I}_{X_n = y} = 1$$

Thus

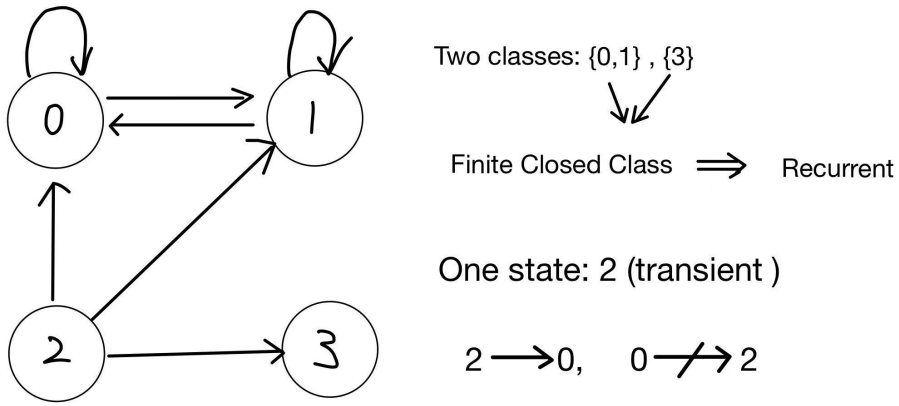
$$E_x\left(\sum_{y\in C}\sum_{n=1}^{\infty}\mathbb{I}_{X_n=y}\right)=E_x\left(\sum_{n=1}^{\infty}1\right)=\infty$$

This is a **contradiction!** Hence, we conclude that there must be at least one recurrent state.

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

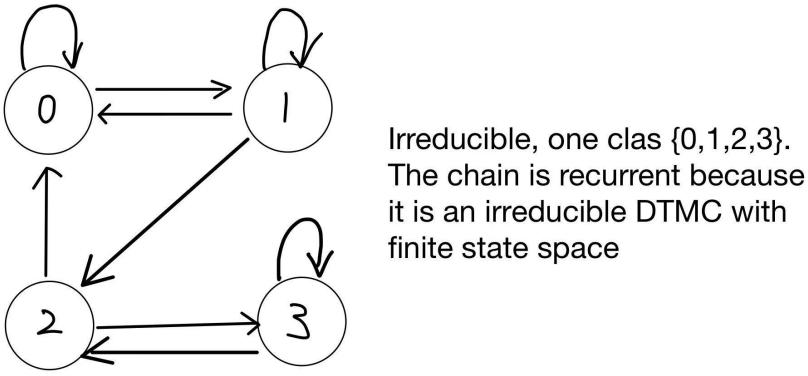
Theorem 4.5.4 — A Finite Closed Class Must be Recurrent.
In particular, on irreducible **DTMC** with finite state space is recurrent.

■ **Example 4.4**



■

■ **Example 4.5**



■

4.6 Existence of Stationary Measure

In this part, we show that on irreducible and recurrent (property of the chain since it is irreducible) DTMC "almost" has a stationary distribution. If the state space is finite, then it has a stationary distribution

Definition 4.6.1
A row vector $\mu^* = (\mu^*(0), \mu^*(1), ...)$ is called a **stationary measure**/invariant measure, if $\mu^*(i) \geq 0$ for all $i \in S$ and $\mu^* P = \mu^*$

■ **Remark 4.3** As we can see, a stationary measure is a stationary distribution without normalization. If $\sum_{i=0}^{\infty} \mu^*(i) < \infty$, then it can be normalized to get a stationary distribution.

$$\text{Normalization}: \ \mu(i) = \frac{\mu^*(i)}{\sum_j \mu^*(j)}$$

Theorem 4.6.1
Let $\{X_n\}_{n=0,1,...}$ be an irreducible and recurrent **DTMC** with transition matrix P. Let $x \in S$ and $T_x = \min\{n \geq 1 : \ X_n = x\}$, then

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

defines a stationary measure with $0 < \mu_x(y) < \infty$ for all $y \in S$

Proof: Define $\bar{P}^n_{xy} = P_x(X_n = y, \ T_x > n)$, then $\mu_x(y) = \sum_{n=0}^{\infty} \bar{P}^n_{xy}$. Now we have two cases.

Case 1: For $z \neq x$

$$(\mu_x P)(z) = \sum_y \mu_x(y) P_{yz} = \sum_{n=0}^{\infty} \sum_y \bar{P}^n_{xy} P_{yz}$$

where

$$\sum_y \bar{P}^n_{xy} P_{yz} = \sum_y P_x(X_n = y, \ T_x > n, \ X_{n+1} = z) = P_x(T_x > n+1, \ X_{n+1} = z) = \bar{P}^{n+1}_{zx}$$

Then

$$\sum_{n=0}^{\infty} \sum_y \bar{P}^n_{xy} P_{yz} = \sum_{n=0}^{\infty} \bar{P}^{n+1}_{zx} = \sum_{n=0}^{\infty} \bar{P}^n_{zx} = \mu_x(z)$$

Case 2: For $z = x$, similarly, we have

$$\sum_{n=0}^{\infty} \sum_y \bar{P}^n_{xy} P_{yx} = \sum_y P_x(X_n = y, \ T_x > n, \ X_{n+1} = x) = P_x(T_x = n+1)$$

and

$$(\mu_x P)(x) = \sum_{n=0}^{\infty} \sum_y \bar{P}^n_{xy} P_{yx} = \sum_{n=0}^{\infty} P_x(T_x = n+1) = 1$$

Since x is recurrent, so $(\mu_x P)(x) = 1 = \mu_x(x) = \sum_{n=0}^{\infty} P_x(X_n = x,\ T_x = n)$. Combine these two parts, we have

$$(\mu_x P)(z) = \mu_x(z) \quad \text{for any } z \in S \implies \mu_x P = \mu_x$$

Next we will show $0 < \mu_x(y) < \infty$ for any $y \in S$, first

$$1 = \mu_x(x) = (\mu_x P^n)(x) = \sum_z \mu_x(z) P^n_{zn} \geq \mu_x(y) \underbrace{P^n_{yx}}_{>0} \textbf{ for any } n$$

Since the chain is irreducible, $P^n_{yx} > 0$ for same n, so $\mu_x(y) < \infty$.
Second, recall that we have proved earlier that if $x \to y$, then there is a way visit y before returning to x, so

$$P_x(\text{number of visits to y before returning to } x \geq 1) > 0$$

this implies that

$$\underbrace{E_x}_{\mu_x(y)}(\text{number of visits to y before returning to } x \geq 1) > 0$$

which completes the proof.

■ **Remark 4.4** Note that

$$\begin{aligned}
\mu_x(y) &= \sum_{n=0}^{\infty} P_x(X_n = y,\ T_x > n) \\
&= \sum_{n=0}^{\infty} E_x(\mathbb{I}_{X_n=y},\ \mathbb{I}_{T_x>n}) \\
&= E_x \sum_{n=0}^{T_x=1} \mathbb{I}_{X_n=y} \\
&= E_x(\textbf{number of visits to y before returning to x})
\end{aligned}$$

Cut the Markov Chain into different "cycles" according to visits to state x.

$$\mu_x(y) = E_x(\textbf{number of visits to y before returning to x})$$

In particular, $\mu_x(x) = 1$

4.7 Periodicity

Definition 4.7.1 — Periodicity.
The **period** of state x is defined as

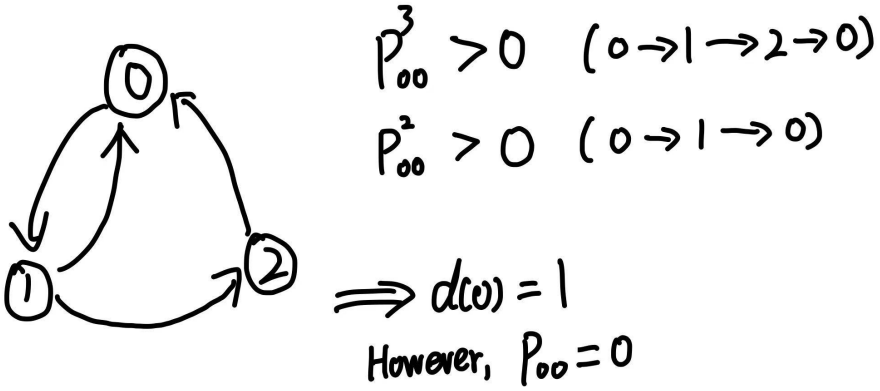
$$d(x) \coloneqq \gcd\{n \geq 1 : P^n_{xx} > 0\}$$

■ **Remark 4.5** We are taking the gcd of the number of steps by which the process returns to x with **positive probability** (x can go back to x in n steps), not the number of steps this probability is 1 (x **must** go back to x in n steps). There is no guarantee that the chain will be in state x at time $d(x)$. Indeed, we don't even always have $P^{d(x)}_{xx} > 0$. The $d(x) = d$ means: If n is not a multiple of d, then $P^n_{xx} = 0$.

Definition 4.7.2 — Aperiodic.
If $d(x) = 1$, we say state x is **aperiodic**, we call this **MC aperiodic**

Note: If $P_{xx} > 0$, then x is obviously aperiodic. However, note that the converse is not true. x is aperiodic does not imply $P_{xx} > 0$.

■ **Example 4.6**



■

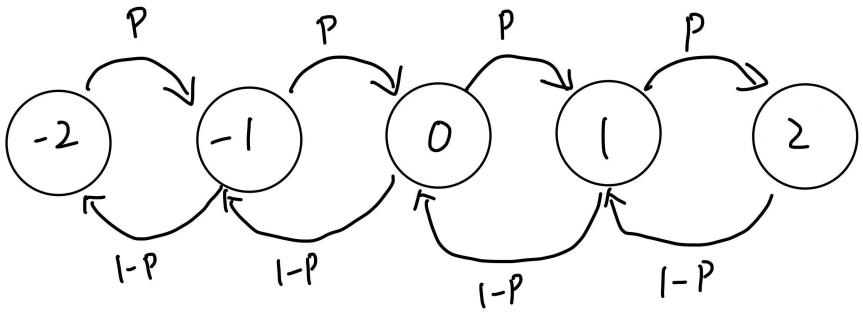
■ **Example 4.7** Simple random walk revisited. Consider a **symmetric** simple random walk ($p = \frac{1}{2}$).

$$\begin{cases} P_{00}^n = 0 & \textbf{if n is odd} \\ P_{00}^n = \binom{n}{\frac{n}{2}} \cdot \left(\frac{1}{2}\right)^n & \textbf{if n is even} \end{cases}$$

Note: $\binom{n}{\frac{n}{2}}$ is the number of ways to get $\frac{n}{2}$ steps to the left and $\frac{n}{2}$ steps to the right. We have that

$$d(0) = \gcd\{2,4,6,...\} = 2$$

Then we have $d(i) = 2$ for all $i \in \mathbb{Z}$.



From this graph, it's easy to see that $d(0) = d(i) = 2$ also for $p \neq \frac{1}{2}$. i.e. $p \in (0,1)$ ■

Lemma 4.7.1 — Period is a Class Property.

$$x \longleftrightarrow y \quad \Longrightarrow \quad d(x) = d(y)$$

Proof: Since $x \to y$ and $y \to x$, there exists n, m such that $P_{xy}^m > 0$ and $P_{yx}^n > 0$. Then we have

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

Moreover, for any l such that $P_{yy}^l > 0$, we have $P_{xx}^{m+n+l} \geq P_{xx}^m P_{yy}^l P_{yx}^n$. As a result, $d(x) \mid (m+n)$ and $d(x) \mid (m+n+l)$ so by **DIC** we have $d(x) \mid l$. Since this holds for all l such that $P_{yy}^l > 0$, $d(x)$ is a common divisor of $\{l : P_{yy}^l > 0\}$. Recall that $d(y)$ is the largest common divisor of the same set, $d(x) \leq d(y)$, similarly we have $d(y) \leq d(x)$, then $d(x) = d(y)$ as desired.

4.8 Convergence Theorem

Conditions:

I: The DTMC is irreducible
A: The DTMC is aperiodic
R: All the states are recurrent
S: There exists a stationary distribution π

Lemma 4.8.1
If there exists a stationary distribution π such that $\pi(y) > 0$, then state y is recurrent.

Proof: Assume $\{x_n\}_{n=0,1,..}$ is a **DTMC** starts from the stationary distribution π, then $P(X_n = y) = \underbrace{\pi(y)}_{>0}$ where $n = 0, 1, ..., $ so

$$\begin{aligned}
\infty = \sum_{n=1}^{\infty} P(X_n = y) = \sum_{n=1}^{\infty} E(\mathbb{I}_{X_n=y}) &= E\left(\sum_{n=1}^{\infty} \mathbb{I}_{X_n=y}\right) \\
&= E(N(y)) \\
&= \sum_{x \in S} E_x(N(y)) \cdot \pi(x) \\
&= \sum_{x \in S} \pi(x) \cdot \frac{\rho_{xy}}{1-\rho_{xy}} \\
&\leq \sum_{x \in S} \pi(x) \cdot \frac{1}{1-\rho_{yy}} \\
&= \frac{1}{1-\rho_{yy}}
\end{aligned}$$

Then $\rho_{yy} = 1$, so y is recurrent.

Corollary 4.8.2
If y is transient, then $\pi(y) = 0$ for any stationary distribution π

Proof: Take the contrapositive of the **Lemma 2.17.2**

Corollary 4.8.3
If a **DTMC** is irreducible and there exists a stationary distribution π, then all the state are recurrent.

Proof: Since there exists a stationary distribution π, so exists a state y such that $\pi(y) > 0$, so y is recurrent. The irreducible implies all the states are recurrent.

Lemma 4.8.4
If y is aperiodic, then there exists n_0 such that $P^n_{yy} > 0$ for all $n \geq n_0$

Proof: We use a fact from number theory, a corollary of Bezout's Lemma: If we have a set I of coprime numbers, then exist integers $i_1, ..., i_m \in I$ and n_0 such that for any $n \geq n_0$ n can be written as

$$n = a_1 i_i + ... + a_n i_m$$

where $a_1, ..., a_m$ are positive integers. Hence, $I = \{n \geq 1 : P^n_{yy} > 0\}$, by aperiodicity, use the above fact, we have there exists $n_0 \in \mathbb{Z}^+$ such that for any $n \geq n_0$ there exists $a_1, ..., a_m > a$ and $i_1, ..., i_m \in I$ such that

$$n = a_1 i_1 + ... + a_m i_m$$

Then we have

$$P^n_{yy} \geq \underbrace{P^{i_1}_{yy} P^{i_1}_{yy} ..... P^{i_1}_{yy}}_{a_1 \textbf{ terms}} \underbrace{P^{i_1}_{yy} P^{i_2}_{yy} ..... P^{i_2}_{yy}}_{a_2 \textbf{ terms}} ..... \underbrace{P^{i_m}_{yy} P^{i_m}_{yy} ..... P^{i_m}_{yy}}_{a_m \textbf{ terms}} > 0$$

which completes the proof.

Theorem 4.8.5 — Convergence.
If a **DTMC** is irreducible, aperiodic and exists a stationary distribution π, then

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

Proof: Consider two independent **DTMC** $\{X_n\}$ and $\{Y_n\}$ which have the same transition matrix P, and arbitrary initial distributions. It's easy to show that $Z_n = (X_n, Y_n)$ is also a **DTMC**, with transition matrix $\overline{P}_{(x_1,y_1),(x_2,y_2)} = P_{x_1x_2} P_{y_1y_2}$. Next we show that under the conditions that a **DTMC** is irreducible and aperiodic, $\{Z_n\}$ is also irreducible. Since $\{X_n\}$ is irreducible, for any x_1, x_2 there exists k such that $P^k_{x_1x_2} > 0$. Similarly, for any y_1, y_2 there exists l such that $P^l_{y_1y_2} > 0$. Since the **DTMC** is aperiodic, by the **Lemma 2.17.5** we have $P^m_{x_1x_2} > 0$ and $P^m_{y_1y_2} > 0$ for all m large enough. That is there exists M such that $P^m_{x_1x_2} > 0$ and $P^m_{y_1y_2} > 0$ for all $m \geq M$. Then for $n \geq M + \max\{k, l\}$ we have

$$P^n_{x_1x_2} \geq P^k_{x_1x_2} P^{n-k}_{x_1x_2} > 0 \quad \textbf{and} \quad P^n_{y_1y_2} \geq P^l_{y_1y_2} P^{n-l}_{y_1y_2} > 0$$

This gives us that

$$\overline{P}_{(x_1,y_1),(x_2,y_2)} = P_{x_1x_2}P_{y_1y_2} > 0$$

Since this holds for all (x_1,y_1) and (x_2,y_2), so **DTMC** $\{Z_n\}$ is irreducible. Moreover, $\{Z_n\}$ is recurrent, to see this we note that $\overline{\pi}_{(x,y)} = \pi(x)\pi(y)$ is a stationary distribution of $\{Z_n\}$. Take x such that $\pi(x) > 0$, then $\overline{\pi(x,x)} > 0$. By the **Lemma 2.17.2** the state (x,x) is recurrent. Since $\{Z_n\}$ is irreducible, so all the states in $\{Z_n\}$ are recurrent. Now define $T = \min\{n \geq 0 : X_n = Y_n\}$, the first time that the chain meets. Also define

$$\underbrace{V(x,x)}_{\geq T} = \{n \geq 0 : X_n = Y_n = x\} = \min\{n \geq 0 : Z_n = (x,x)\} \leq \min\{n \geq 1 : Z_n = (x,x)\} = T(x,x)$$

and we see that

$$\begin{aligned} P(T(x,x) < \infty) &= E(P(T(x,x) < \infty \mid (x_0,y_0))) \quad \textbf{for any } x_0, y_0 \\ &= P(T(x,x) < \infty \mid X_0 = x_0, Y_0 = y_0) \\ &= \rho_{(x_0,y_0),(x,x)} \end{aligned}$$

$\{Z_n\}$ irreducible $\Longrightarrow (x,x) \to (x_0,y_0)$ and $\rho_{(x,x),(x_0,y_0)} > 0$. And we see (x,x) is recurrent $\Longrightarrow \rho_{(x_0,y_0),(x,x)} = 1$. Thus $P(T(x,x) < \infty) = 1$, so

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

for some state. Therefore, we have proved that the two independent **DTMC** $\{X_n\}$ and $\{Y_n\}$ will eventually meet. (with prob 1)
For any state $y \in S$, by discussing the values of T and X_T, we have for any n

$$\begin{aligned} P(X_n = y, T \leq n) &= \sum_{m=0}^{n} \sum_{x \in S} P(T = m, X_m = x, X_n = y) \\ &= \sum_{m=0}^{n} \sum_{x \in S} P(T = m, X_m = x) \cdot P(X_n = y \mid X - m = x) \\ &= \sum_{m=0}^{n} \sum_{x \in S} P(T = m, X_m = x) \cdot P_{xy}^{n-m} \\ &= \sum_{m=0}^{n} \sum_{x \in S} P(T = m, Y_m = x) \cdot P(Y_n = y \mid Y_m = x) \\ &= P(Y_n = y, T \leq n) \end{aligned}$$

"After meeting, they have the same distribution", then

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

Recall that this holds for any initial distribution of $\{X_n\}$ and $\{Y_n\}$. Take $x_0 = x$, $Y_0 \sim \pi$, then by the above we have

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

that is

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

which completes the proof.

■ **Remark 4.6** Note: $\pi(y)$ does not depend on the starting state x, the stationary distribution is unique. The limiting transition probability, hence also the limiting distribution, does not depend on where we start.

$$\lim_{n\to\infty} P_{xy}^n = \pi(y) \quad \Longrightarrow \quad \lim_{n\to\infty} P(X_n = y) = \pi(y)$$

Theorem 4.8.6 — Asymptotic Frequency.
Suppose a **DTMC** is aperiodic and all states are recurrent. If $N_n(y)$ is the number of visits to y up to time n, then

$$\lim_{n\to\infty} \frac{N_n(y)}{n} = \frac{1}{E_y(T_y)}$$

where recall that $T_y = \min\{n \geq 1 : X_n = y\}$.
"Long-run function of time spent in y is $\frac{1}{E_y(T_y)}$" and the $\frac{1}{E_y(T_y)}$ is the expected cycle length.

Proof: We chop the time line into different cycles. Let $T_y^{(1)}, T_y^{(2)}, \ldots..$ be the times that the chain (re)visits y after time 0. By the strong Markov property, $T_y^{(2)} - T_y^{(1)}, T_y^{(3)} - T_y^{(2)}, \ldots..$ are **i.i.d r.v.s** We recall the strong law of large number: $X_1, X_2, \ldots$ **i.i.d** then $\frac{\sum_{i=1}^n X_i}{n} \to E[X_1]$. By **SLLN**,

$$\frac{T_y^{(k)} - T_y^{(1)}}{k-1} = \frac{\sum_{i=1}^{k-1} T_y^{(i+1)} - T_y^{(i)}}{k-1} \to E\left(T_y^{(i+1)} - T_y^{(i)}\right) = E_y(T_y)$$

with negligible changes, this implies

$$\frac{T_y^{(k)}}{k} \to E_y(T_y) \quad \textbf{as } k \to \infty \quad \textbf{(almost surely)}$$

Note that

$$\underbrace{\frac{T_y^{(N_n(y))}}{N - n(y)}}_{\to E_y(T_y)} \leq \frac{n}{N_n(y)} < \underbrace{\frac{T_y^{(N_n(y)+1}}{N_n(y)+1} \cdot \frac{N_n(y)+1}{N_n(y)}}_{\to E_y(T_y)}$$

Then, we have that

$$\frac{n}{N_n(y)} \to E_y(T_y) \quad \textbf{with prob 1}$$

This gives us that

$$\frac{N_n(y)}{n} \to \frac{1}{E_y(T_y)} \quad \textbf{with prob 1}$$

as desired.

4.9 Cycle Length and the Uniqueness of Stationary Distribution

Theorem 4.9.1

If a **DTMC** is irreducible and there exists a stationary distribution, then

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

In particular, the stationary distribution is unique.

Proof: Since irreducible and stationary distribution implies all states are recurrent, we can apply the above theorem and get

$$\frac{N_n(y)}{n} \to \frac{1}{E_y(T_y)}$$

take the expectation both sides

$$E\left(\frac{N_n(y)}{n}\right) \to \frac{1}{E_y(T_y)} \qquad \textbf{Dominated Convergence Theorem}$$

This gives us that

$$E(N_n(y)) = E\left(\sum_{m=1}^{n} \mathbb{I}_{X_n=y}\right) = \sum_{m=1}^{n} E(\mathbb{I}_{X_n=y}) = \sum_{m=1}^{n} P(X_m = y)$$

This result holds for any initial distribution. Now assume the chain starts from the stationary distribution π. Then we have

$$P(X_m = y) = P(X_0 = y) = T(y) \quad \Longrightarrow \quad E(N_n(y)) = nT(y)$$

Thus,

$$E\left(\frac{N_n(y)}{n}\right) = \pi(y) = \frac{1}{E_y(T_y)}$$

Corollary 4.9.2

If a **DTMC** is irreducible, aperiodic and has a stationary distribution. Then

$$\pi(y) = \lim_{n\to\infty} P_{xy}^n = \lim_{n\to\infty} \frac{N_n(y)}{n} = \frac{1}{E_y(T_y)}$$

i.e.

Stationary distribution= limiting transition prob=Long-run fraction of time= 1/expected cycle length

"Everything exists and things are all equal"

4.10 Long Run Average

Lemma 4.10.1
If a **DTMC** is irreducible and has a stationary distribution , then $\pi(x) > 0$ for all $x \in S$

Proof: Since π is a stationary distribution,

$$\sum_z \pi(z) = 1 \quad \Longrightarrow \quad \pi(y) > 0 \ \textbf{for some} \ y \in S$$

Since markov chain is irreducible for any state $x \in S$, $y \to x$

$$\Longrightarrow \ \exists\, n \in \mathbb{N} \ \textbf{s.t.} \ P^n_{yx} > 0$$

Since π is a stationary distribution, then $\pi = \pi \cdot P^n$. Hence, we have

$$\pi(x) = \sum_{z \in S} \pi(z) \cdot P^n_{zx} \geq \underbrace{\pi(y)}_{>0} \cdot \underbrace{P^n_{yx}}_{>0} > 0$$

Theorem 4.10.2 — Long Run Average.
If a **DTMC** is irreducible and has a stationary distribution, and a function f satisfies $\sum_x |f(x)|\pi(x) < \infty$, then

$$\underbrace{\lim_{n\to\infty} \underbrace{\frac{1}{n} \underbrace{\sum_{m=1}^{n} f(x_m)}_{\text{reward from 1 to n}}}_{\text{average reward per step}}}_{\text{long-run average reward per step}} = \sum_x f(x)\pi(x) = \pi \cdot f^T = \pi \cdot [f(0), f(1), ...]^T$$

Proof: Since irreducible and stationary distribution implies all states are recurrent, from the result for the existence of stationary measure, we have for $y \in S$

$$\mu_x(y) = \sum_{n=0}^{\infty} P_x(X_n = y, T_x > n) = E_x(\text{number of visits to y before returning to x}) = E_x N_{T_x}(y)$$

Moreover, note that $\sum_y E_x N_{T_x}(y) = E_x(T_x) = \frac{1}{\pi(x)}$ and $\pi(x) > 0 \Longrightarrow E_x(T_x) < \infty$. Then we can see $\{\pi_x(y)\} = \{E_x(N_{T_x}(y))\}_{y\in S}$ is negligible, and $\left\{\frac{E_x(N_{T_x}(y))}{E_x(T_x)}\right\}_{y\in S}$ gives a stationary distribution. Moreover, irreducible and has a stationary distribution implies the stationary distribution is unique. Hence,

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

Recall that $E_x(T_x) = \frac{1}{\pi(x)}$, we have

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

Again, we use the **"cycle trick"**. The reward collected in $k-th$ cycle (defined by returns to x) is

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

and we have

$$\begin{aligned} E(Y_k) &= \sum_{y\in S} E_x(N_{T_x}(y)) \cdot f(y) \quad \textbf{Strong Markov Property} \\ &= \frac{\sum_{y\in S} \pi(y) f(y)}{\pi(x)} \\ &= \frac{\pi \cdot f^T}{\pi(x)} \end{aligned}$$

the average reward over time is

$$\frac{\sum_{k=2}^{N_x(x)} Y_k + \textbf{negligible term}}{\sum_{k=2}^{N_n(x)} (T_k - T_{k-1}) + \textbf{negligible term}}$$

where $T_k = T_x^{(k)}$ is the time of the k-th (re)visit to x. The negligible terms come from the first and the last cycles. Strong law of the large number:

$$\frac{\sum_{k=2}^{N_x(x)} Y_k + \textbf{negligible term}}{\sum_{k=2}^{N_n(x)} (T_k - T_{k-1}) + \textbf{negligible term}} = \frac{\frac{1}{N_n(x)-1}\sum_{k=2}^{N_n(x)} Y_k}{\frac{1}{N_n(x)-1}\sum_{k=2}^{N_x(x)} (T_k - T_{k-1})} \to \frac{E(Y_k)}{E_x(T_x)} = \frac{\pi \cdot f^T/\pi(x)}{1/\pi(x)} = \pi \cdot f^T$$

4.11 Application of the Main Theorems

■ **Example 4.8**

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

Irreducible: $P_{03}, P_{32}, P_{21}, P_{10} > 0$
Aperiodic: $P_{00} > 0$
Recurrent: Irredicible **DTMC** with finite state space.
Solve for stationary distributions:

$$\begin{cases} \pi P = P \\ \pi \mathbb{I} = 1 \end{cases} \quad \Longrightarrow \quad \pi = \left(\frac{19}{110}, \frac{30}{110}, \frac{40}{110}, \frac{21}{110} \right)$$

By previous result, we have limiting transition prob

$$\lim_{n\to\infty} P_{xy}^n = \pi(y) \quad \textbf{for example:} \quad \lim_{n\to\infty} P_{12}^n = \pi(2) = \frac{4}{11}$$

Note again that this limit does not depend on the initial state x

Long-run fraction of time visiting 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 $\pi(0) = \frac{19}{110}$.

Expected time that the chain visit state y again, given if starts from y: $E_y(T_y) = \frac{1}{\pi(y)}$. For example, given the chain starts from state 3, the expected time that it returns to state 3 is $\frac{1}{\pi(3)} = \frac{110}{21}$

Long-run average: For example, we are looking at an inventory model, and the holding cost for state x is $2x$, then the average holding cost per unit of time in the long-run is

$$\pi \cdot f^T = \frac{173}{55}$$

■ **Remark 4.7** Typically, the stationary distribution is easy to find. So the above results are usually used to find the other related quatities.

■

4.12 Roles of different conditions

Irreduciblity is related to the uniqueness of the stationary distribution.

$$\textbf{irreducible} \implies \textbf{stationary distribution is unique (if exists)}$$

Example

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

Both $(1,0)$ and $(0,1)$ are stationary distributions. Then, any convex combination of them:

$$a(1,0) + (1-a)(0,1) = (a, 1-a) \quad a \in [0,1]$$

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

$$\lim_{n\to\infty} P^n_{01} = 0 \qquad \lim_{n\to\infty} P^n_{11} = 1$$

Aperiodicity is related to the existence of $\lim_{n\to\infty} P^n_{xy}$

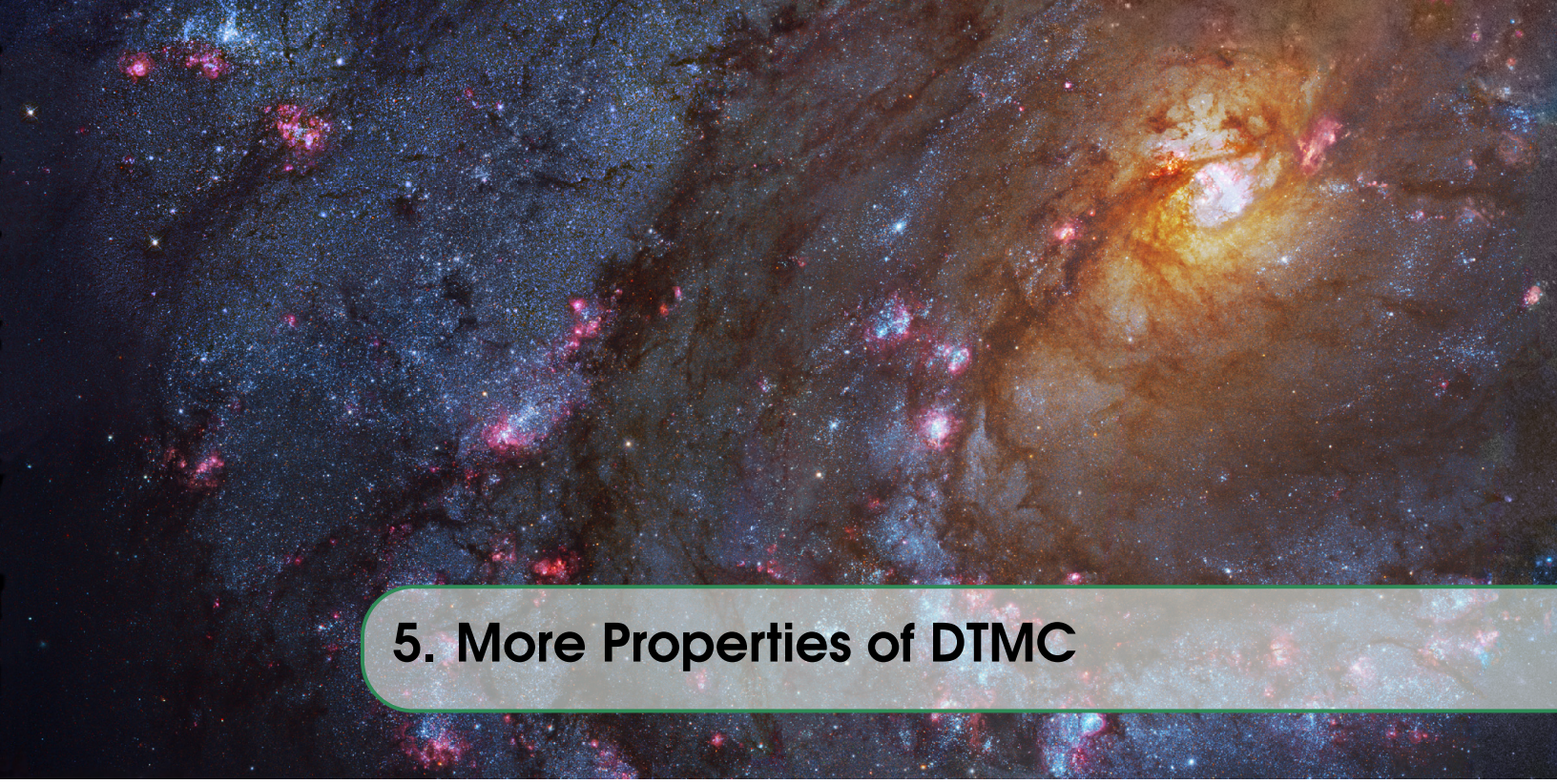
$$y \textbf{ aperiodic} \implies \lim_{n\to\infty} P^n_{xy} \textbf{ exists}$$

■ **Example 4.9**

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

so we can see $d(0) = d(1) = 2$. Note that $P^2 = I \implies P^{2n} = I$ and $P^{2n+1} \neq I$.
Thus, $\lim_{n\to\infty} P^n_{xy}$ does not exists for $x, y \in \{0,1\}$ ■

Recurrence:is related to the existence of a stationary distribution. The MC is (irreducible and) recurrent, implies a stationary measure exists. "positive recurrence" implies stationary distribution exists



5. More Properties of DTMC

5.1 Detailed Balance Condition

Definition 5.1.1 — Detailed Balance Condition.
A distribution $\pi = \{\pi(x)\}_{x\in S}$ is said to satisfy the **detailed balanced condition** if

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

Proposition 5.1.1 Detailed Balance Condition $\Longrightarrow$ Stationary Distribution

Proof: Note that the $(\pi P)_j$ is defined as $(\pi P)_j = \sum_{i\in S} \pi(i)P_{ij}$. Since π satisfies the detailed balance condition, so we have $\pi(i)P_{ij} = \pi(j)P_{ji}$ for all $i, j \in S$. Then we have that

$$(\pi P)_j = \sum_{i\in S} \pi(i)P_{ij} = \sum_{i\in S} \pi(j)P_{ji} = \pi(j)\sum_{i\in S} P_{ji} = \pi(j)\sum_{i\in S} P(X_1 = j \mid X_0 = i) = \pi(j)\cdot 1 = \pi(j)$$

That is $\pi(j) = (\pi P)_j$ for any $j \in S$, so we have $\pi = \pi P$ as desired, which shows that π is a stationary distribution.

■ **Remark 5.1** Detailed Balance Condition- Balance between each pair of states: probability flow $x \to y$ equals the probability flow $y \to x$

5.2 Time Reversibility

Definition 5.2.1 — Reversed Process.
Let $\{X_n\}_{n=0,1,2\ldots}$ be **DTMC**. Fix n, then the process $\{Y_m\}_{m=0,1,2\ldots,n}$ given by $Y_n = X_{n-m}$, is called the **reversed process** of $\{X_n\}$

■ **Remark 5.2** In general, the reversed process of a **DTMC** is not necessarily a **DTMC** But it's still a **DTMC**

in the following case.

Theorem 5.2.1
If $\{X_n\}_{n=0,1,...}$ 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

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

Proof: To show that $\{Y_m\}$ is a **DTMC**, we check the Markov property

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

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

$$P(Y_{m+1} = i_{m+1} \mid Y_m = i_m, ..., Y_0 = i_0) = \frac{\pi(i_{m+1})P_{i_{m+1},i_m}}{\pi(i_m)}$$

This shows:
1. The transition prob does not depend on the history $i_{m-1}, ..., i_0$ when i_m is given. Hence, $\{Y_m\}$ is a **DTMC**
2. The transiton prob is given by $\widehat{P_{ij}} = P(Y_{m+1} = j \mid Y_m = j) = \frac{\pi(j)P_{ji}}{\pi(i)}$

■ **Remark 5.3** We can check that $\hat{P} = \left\{\hat{P}_{ij}\right\}_{i,j \in S}$ is indeed a valid transition matrix: $\hat{P}_{ij} \geq 0$ and $\sum_{j \in S} \hat{P}_{ij} = 1$

Definition 5.2.2 — Time-Reversible DTMC.
A DTMC $\{X_n\}_{n=0,1,..}$ is called **time-reversible** if its reversed chain $\{Y_m ::= X_{n-m}\}_{m=0,...,n}$ has the **same distribution** as $\{X_n\}_{n=0,1,..}$ for all n.

Note: **Same distribution** in here means: not only marginal distributions, but also joint distributions.
$\Longrightarrow$ Distributes as two DTMCs (same initial distribution, same transition matrix)

■ **Remark 5.4** If a DTMC is time-reversible, then its reversed process is clearly a DTMC, but the converse is **FALSE** (reversedd process is DTMC **does not implies** it has the same distribution as the original DTMC - time-reversibility)
Intuitively, this difference is related to the difference between the detailed balance condition and the stationary condtion.

Proposition 5.2.2
A DTMC $\{X_n\}_{n=0,1,2..}$ is **time-reversible** if and only if it satisfies the **detailed balance condition**

Proof: $\Longleftarrow$ Assume the detailed balance condition, then $\{X_n\}$ starts from the stationary distribution and $\pi(i)P_{ij} = \pi(j)P_{ji}$. Then we have $\{Y_m\}_{m=0,1,..,n}$ is a DTMC and $Y_0 = X_n \sim \pi$ and transition probability is given by

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

Therefore, $\{X_n\}$ and $\{Y_m\}$ are two DTMCs with the same initial distribution and same transition matrix, hence have the same distribution.

$\Longrightarrow$ Assume $\{X_n\}$ is **time-reversible**. Then by definition, X_0 and $X_n = Y_0$ have the same distribution. This holds for all n. Then X_0 follows a stationary distribution π. Moreover, by **time-reversibility**

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

so we have $\pi(i)P_{ij} = \pi(j)P_{ji}$ for all $i, j \in S$, which is the detailed balance condition.

5.3 Metropolis-Hastings Algorithm

Goal: Suppose from the distribution $\pi = \{\pi(x)\}_{x\in S}$ when a direct sampling is hard to implement.

An "**MCMC**" (Markov Chain Monte Carlo) algorithm-Idea: Construct a DTMC which is easy to simulate, then modify to get another DTMC, for which π is the stationary distribution. Then wait for long enough for the distribution of the DTMC to approach the stationary distribution π

Algorithm:

-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 Q_{xy}

2. Accept this move with probability

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

If the move is rejected, stay in x and wait for a longtime, then sample from this MC

Why this algorithm gives a DTMC having π as its stationary distribution?

Reason: The transition matrix of the modified MC is given by

$$P_{xy} = Q_{xy} \cdot r_{xy} \quad \textbf{with } x \neq y \qquad \textbf{and} \qquad P_{xx} = 1 - \sum_{y \neq x} P_{xy}$$

We will check the detailed balance condition for any two states $x, y \in S$, by symmetric, assume $\pi(y)Q_{yx} \geq \pi(x)Q_{xy}$, then we have $r_{xy} = 1$ and

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

Hence, we have

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

Thus, the detailed balance condition holds, π is a stationary distribution.

In order to use the **convergence theorem**, we still need the conditions: irreducible and aperiodic.
It needs to be guaranteed by construction.

For example, $Q_{xy} > 0$ whenever $Q_{yx} > 0$. Aperiodic is almost satisfied, because the rejection rate is typically positive, then $P_{xx} > 0$. Then, by **convergence theorem**

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

5.4 Exit Distribution

Temporary behavior of the DTMC

1. If a DTMC starts from a transient state and will eventualy enter a recurrent class, when will this happen? (exit time/absorption time)

2. If there are more than one recurrent class, which one will the chain enter? (exit probability/absorption probability)

Basic Setting (exit probability):
Let $A, B \subseteq S$ where S is the state space, $C = S \setminus (A \cup B)$ is finite.

Question: Starting from a state in C, what is the probability that the chain exits C by entering A or B?

Mathematical Formulation:

$$V_A = \min\{n \geq 0 : \ X_n \in A\} \qquad \textbf{and} \qquad V_B = \min\{n \geq 0 : \ X_n \in B\}$$

Then what is $P_x(V_A < V_B)$?

■ **Example 5.1**

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

$A = \{3\}$, $B = \{4\}$ and $C = \{1,2\}$. Note that $P_{33} = P_{44} = 1$ is not important, as we are only interested in the chain before it hits 3 or 4. Now let

$$h(1) = P_1(V_3 < V_4) \quad \textbf{and} \quad h(2) = P_2(V_3 < V_4)$$

Discuss what happens in the first step:

$$h(1) = P_1(V_3 < V_4) = \sum_{x=1}^{4} P(V_3 < V_4 \mid X_1 = x, X_0 = 1) \cdot P(X_1 = x \mid X_0 = 1) \quad \textbf{by law of total probability}$$

where

$$P(V_3 < V_4 \mid X_1 = x, X_0 = 1) = \begin{cases} P_1(V_3 < V_4) = h(1) & \textbf{if } x = 1 \\ P_2(V_3 < V_4) = h(2) & \textbf{if } x = 2 \\ 1 & \textbf{if } x = 3 \\ 0 & \textbf{if } x = 4 \end{cases}$$

Then

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

Similarly, we have

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

Solving the system of equation we get

$$h(1) = 0.7 \quad \textbf{and} \quad h(2) = \frac{7}{8}$$

The idea used to solve the example is called "First-step analysis" ■

Theorem 5.4.1

Let $S = A \cup B \cup C$, where A, B, C are disjoint sets, and C is finite. If $P_x(V_A \wedge V_B < \infty) > 0$ ($\wedge$ means $\min\{V_A, V_B\}$) for all $x \in C$, then

$$h(x) := P_x(V_A < V_B), \quad x \in C$$

is the unique solution of the system of equations

$$h(x) = \sum_{y \in S} P_{xy} h(y) \quad x \in C$$

with boundary conditions

$$h(a) = 1 \quad a \in A \quad \textbf{and} \quad h(b) = 0 \quad b \in B$$

Proof: By first-step analysis, we have

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

We see that the boundary conditions hold trivially. Hence, we need to check the uniqueness. Note that the system of equations can be written as

$$h^T = Q\cdot h^T + R_A^T$$

where

$$h = (h(x_1), h(x_2), ...) \quad \textbf{for } x_1,x_2,...\in C \qquad Q = \{P_{xy}\}_{x,y\in C}$$

and

$$R'_A = \begin{pmatrix} \sum_{y\in A} P_{x_1 y} \\ \sum_{y\in A} P_{x_2 y} \\ \vdots \end{pmatrix}$$

The reason is that

$$h(x) = \sum_{y\in S} P_{xy}\cdot h(y) = \underbrace{\sum_{y\in C} P_{xy}\cdot h(y)}_{(Q\cdot h')_{(x)}} + \underbrace{\sum_{y\in A} P_{xy}}_{R'_A(x)}$$

This gives us that

$$I\cdot h^T = Q\cdot h^T + R_A^T \quad\Longrightarrow\quad (I-Q)h^T = R_A^T \quad\Longrightarrow\quad h^T = (I-Q)^{-1}R_A^T$$

is unique as long as $I-Q$ is intertible. Now note that we can modify the transition matrix. **(see the picture below)** Since we are only interested in observing the chain before it hits A or B, changing the transition probabilities going out of states in A or B will not change the result of this problem. After this change, A and B are absorbing and all the states in C be cause transient. (because $P_x(V_A \wedge V_B < \infty) > 0$). As a result,

$$0 = \lim_{n\to\infty} P_x(X_n^T \in C) = \lim_{n\to\infty} \sum_{y\in C} ((P^T)^n)_{xy} = \lim_{n\to\infty} \underbrace{\sum_{y\in C} (Q^n)_{xy}}_{X'_0,X'_1,...X'_n\in C}$$

Since this is true for any $x\in C$, we have $\lim_{n\to\infty} Q^n = 0$. Then all the eigenvalues of Q have norm smaller than 1. Hence, there does not exists a non-zero column vector f^T s.t.

$$f^T = Qf^T \quad\Longleftrightarrow\quad (I-Q)f^T = 0$$

so $I-Q$ is invertible.

■ **Remark 5.5** We see that the function h in the above theorem satisfies

$$h(x) = \sum_y P_{xy}\cdot h(y) = E_x(h(x_1)) = h^{(1)}(x) \quad x\in C$$

Definition 5.4.1 — Harmonic Function.
A function h is called **harmonic** at state x if

$$h(x) = E_x(h(x_1)) = h^{(1)}(x)$$

and h is called **harmonic** in $A \subseteq S$ if

$$h(x) = h^{(1)}(x) \qquad \textbf{for any } x \in A$$

Definition 5.4.2 — Matrix Formula.
In the proof we have seen that $h^T = (I-Q)^{-1}R_A^T$. This is the matrix formula to calculate

$$P_x(V_A < V_B) = h(x)$$

5.5 Exit Time

Basic Setting:
Let $S = A \cup C$ where A, C are disjoint and C is finite. Define

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

V_A is called **exit time** (from C) absorption time/hitting time (for A). We want to know $E_x(V_A) = E(V_A \mid X_0 = x)$ for $x \in C$.

■ **Example 5.2**

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

Let $C = \{1,2\}$ and $A = \{3,4\}$. We want to know

$$g(1) := E(V_A \mid X_0 = 1) \quad \textbf{and} \quad g(2) := E(V_A \mid X_0 = 2)$$

Note that $g(3) = g(4) = 0$. Some as for the exit probability. the idea is the first step analysis.

$$g(1) = E(V_A \mid X_0 = 1) = \sum_{x=1}^{4} E(V_A \mid X_1 = x, X_0 = 1) \cdot P(X_1 = x \mid X_0 = 1)$$

where

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

Note that the "1" corresponds to the time already passed from time 0 to time 1. Then

$$g(1) = 0.25(g(1)+1) + 0.6(g(2)+1) + 0.15 \cdot 1 = 1 + 0.25 \cdot g(1) + 0.6 \cdot g(2)$$

Similarly, we have

$$g(2) = 1 + 0.2 \cdot g(2)$$

Solving for $g(1)$ and $g(2)$, we get $(g(1), g(2)) = (\frac{7}{3}, \frac{5}{4})$. Then, starting from state 1, the expected time until the chain visits 3 or 4 is $\frac{7}{3}$. If start at 2, the expected time until the chain visits 3 or 4 is $\frac{5}{4}$ ∎

Theorem 5.5.1

Let $S = A \cup C$ where A, C are disjoint, and let C is finite. If $P_x(V_A < \infty) > 0$ for any $x \in C$, then $g(x) = E_x(V_A)$ for $x \in C$ is the unique solution of the system of equations

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

with boundary conditions $g(a) = 0$ for all $a \in A$

Proof: By first step analysis, we have

$$g(x) = \sum_{y \in C} P_{xy}(g(y) + 1) + \sum_{y \in A} P_{xy} \cdot 1 = 1 + \sum_{y \in C} P_{xy} g(y) = 1 + \sum_{y \in S} P_{xy} g(y) \quad \textbf{by boundary condition}$$

Uniqueness: Rewrite the system of equations in vector-matrix form:

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

Then we have $g^T = \mathbb{I}^T + Q g^T$, so

$$I g^T = \mathbb{I}^T + Q g \implies (I - Q) g^t = \mathbb{I}^T \implies g^T = (I - Q)^{-1} \mathbb{I}^T$$

we are looking at the exactly the same matrix $I - Q$ as in the exit probability part. By the previous theorem, we know $I - Q$ is invertible. As a result, g^T is the unique solution.

5.6 Positive Recurrence and Null Recurrence

Infinite State Space

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

Definition 5.6.1 — Positive Recurrence.

A state x is called **positive recurrent** if $E_x(T_x) < \infty$.

Definition 5.6.2 — Null Recurrence.

A recurrent state x is called **null recurrent** if $E_x(T_x) = \infty$

How is it possible that an (almost surely finite) random variable has an infinite mean?

■ **Example 5.3** Let X be a random variable s.t. $X = 2^n$ with probability 2^{-n} for $n = 1, 2, ...$ Since $\sum_{n=1}^{\infty} 2^{-n} = 1$, so $P(X < \infty) = 1$. However

$$E(X) = 2 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} + ... = 1 + 1 + ... + \infty$$

■

5.7 Positive Recurrence = Existence of Stationary Distribution

Theorem 5.7.1
For an irreducible DTMC, 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:
(3) $\Longrightarrow$ (1): Trivial
(1) $\Longrightarrow$ (2): Let x be positive recurrent. Recall that x is recurrent and the chain is irreducible, it gives us a stationary measure.

$$\mu_x(y) = \sum_{n=0}^{\infty} P_x(X_n = y, T_x > n) = E(\textbf{the number of visits to } y)$$

before returning to x, given $X_0 = x$, for $y \in S$. Moreover, this stationary measure can be normalized to a stationary distribution if and only if $\sum_y \mu_x(y) < \infty$. Recall that

$$\begin{aligned}
\sum_{y \in S} \mu_x(y) = \sum_{y \in S} \sum_{n=0}^{\infty} P_x(X_n = y, T_x > n) &= \sum_{y \in S} \sum_{n=0}^{\infty} E_x(\mathbb{I}_{X_n = y} \mathbb{I}_{T_x > n}) \\
&= E_x\left(\sum_{n=0}^{\infty} \mathbb{I}_{T_x > n} \cdot \underbrace{\sum_{y \in S} \mathbb{I}_{X_n = y}}_{=1} \right) \\
&= E_x\left(\sum_{n=0}^{\infty} \mathbb{I}_{T_x > n} \right) \\
&= E_x(T_x) < \infty
\end{aligned}$$

Since x is positive recurrent, so $\pi(y) = \dfrac{\mu_x(y)}{E_x(T_x)}$ gives a stationary distribution.

(2) $\Longrightarrow$ (3): Recall that we have irreducible and stationary distribution implies

$$\pi(x) = \frac{1}{E_x(T_x)} > 0$$

for any $x \in S$. As a result, $E_x(T_x) < \infty$ for all $x \in S$.

Corollary 5.7.2
Positive recurrence and null recurrence are class property.

Proof: Let C be a class and $x \in C$ is positive recurrent. Since C is recurrent, so it's closed. Since for any $y \in C$, the chain starting from y will only move in C, we can focus on C and consider the chain restricted on C with transition matrix $P\,|_C = \{P_{xy}\}_{x,y\in C}$, that is

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

The restricted chain is irreducible, and has a positive recurrent state x. (Note that $E_x(T_x)\,|_P = E_x(T_x)\,|_{P|_C}$). By the previous theorem, all its states are positive recurrent, then the states in C is positive recurrent. Since both positive recurrence and recurrence are class properties, so null recurrence is also class properties.

Corollary 5.7.3
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 show that the result for the case where the chain is irreducible (Otherwise, we can consider the chain restricted on the closed class containing x).
$\Longrightarrow$ By previous theorem, since x is positive recurrent, there exists a stationary distribution π. Recall that we also has irreduciblestationary distribution implies $\pi(x) > 0$ for all x, hence we have $\pi(x) > 0$.

$\Longleftarrow$ Already given by the previous theorem

Corollary 5.7.4
A DTMC with finite state space must have at least one positive recurrent state.

Proof: Again, we can assune the DTMC is irreducible. We already know it must be recurrent. Take a state x, then

$$\mu_x(y) = \sum_{n=0}^{\infty} P_x(X_n = y, T_x > n)$$

gives a stationary measure. Moreover, since there are only finitely many terms, the summation $\sum_{y\in S} \mu_x(y)$ is

trivially finite. This gives us that $\{\mu_x(y)\}_{y\in S}$ is normaligable, and

$$\left\{\pi(y) := \frac{\mu_x(y)}{\sum_{y\in S}\mu_x(y)}\right\}_{y\in S}$$

is a stationary distribution. Then x must be positive recurrent.

Corollary 5.7.5
A DTMC with finite state space does not have null recurrent state.

Proof: Suppose there is a null recurrent state, hence there exists null recurrent class. Since it is recurrent, it is closed. Consider the chain restricted to the class, the restricted chain is irreducible and null recurrent. However, since it only has a finite a number of states, it must has a positive recurrent state. This is a contradiction, so there is no null recurrent state.

"A null recurrent class must have an infinite number of states."

Intuition:

$$\frac{1}{E_x(T_x)} = \lim_{n\to\infty}\frac{N_n(x)}{n} \qquad \textbf{long-run faction spent on } x$$

and

$$E_x(T_x) = \infty \quad \textbf{long-run faction is } 0$$

This can happen only if there are infinitely many such states.

5.8 Simple Random Walk Examples

■ **Example 5.4** Let $S=\mathbb{Z}$, $P_{x,x+1}=p$ and $P_{x,x-1}=1-p=q$ where $p\in(0,1)$. Then this DTMC is irreducible with period 2.

Claim: The simple random walk is transient for $p\neq\frac{1}{2}$, it's null recurrent for $p=\frac{1}{2}$.
Proof:
When $p\neq\frac{1}{2}$, by symmetry, assume $p>\frac{1}{2}$, note that

$$X_n = Y_1 + ... + Y_n$$

where $Y_1, Y_2, ...$ are **i.i.d** and

$$Y_n\begin{cases}1 & \textbf{prob=}p\\ -1 & \textbf{prob=}1-p\end{cases} \qquad \textbf{and} \qquad E(Y_1) = 1\cdot p + (-1)(1-p) = 2p-1>0$$

By strong law of large number, we have

$$\frac{X_n}{n} = \frac{1}{n}\sum_{m=1}^{n} Y_m \underbrace{\longrightarrow}_{\textbf{almost surely}} E(Y_1) = 2p-1>0$$

as $n\to\infty$. Then we have $X_n\to\infty$ as $n\to\infty$ almost surely. Therefore, for any state $x\geq 0$ (in particular, for state), there is a lat visit to x, so 0 is transient. Hence, $\{X_n\}$ is transient.

When $p = \frac{1}{2}$, recall that a state x is recurrent if and only if $\sum_{n=0}^{\infty} P_{xx}^{(n)} = \infty$. For $x = 0$, we have

$$P_{00}^{(2n)} = \binom{2n}{n}\left(\frac{1}{2}\right)^n\left(\frac{1}{2}\right)^n = \binom{2n}{n}\left(\frac{1}{4}\right)^n$$

Then $P_{00}^{2n+1} = 0$ since it has period 2. This is hard to compute, but we have a good way to approximate. By Stirling's formula we have

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

Hence, we have

$$\binom{2n}{n} = \frac{(2n)!}{n!n!} \sim \frac{\sqrt{2\pi}e^{-2n}(2n)^{2n+\frac{1}{2}}}{(\sqrt{2\pi}e^{-n}n^{n+\frac{1}{2}})^2} = \frac{1}{\sqrt{2\pi}}2^{2n+\frac{1}{2}}\frac{1}{\sqrt{n}}$$

This gives us that

$$\binom{2n}{n}\left(\frac{1}{4}\right)^n \sim \frac{1}{\sqrt{\pi n}} > \frac{1}{n} \quad \textbf{for } n \geq 4$$

Then we have

$$\sum_{n=1}^{\infty}\binom{2n}{n}\left(\frac{1}{4}\right)^n = \infty$$

State 0 is recurrent. Next, we show that 0 is not positive recurrent by showing that there does not exists a stationary distribution. Consider the system of equations $\pi P = \pi$. That is

$$\pi(x) = \frac{1}{2}\pi(x-1) + \frac{1}{2}\pi(x+1) \quad \Longrightarrow \quad \pi(x+1) - \pi(x) = \pi(x) - \pi(x-1)$$

Since this holds for all $x \in \mathbb{Z}$, $\pi(x)$ is an arithmetic series. The general form is $\pi(x) = \pi(0) + ax$ where $a = \pi(1) - \pi(0)$. Also, we know $\pi(x) \in [0,1]$ for all $x \in \mathbb{Z}$, this forces $a = 0$. This implies $\pi(x) = \pi(0)$ for $x \in \mathbb{Z}$. If $\pi(0) = 0$, $\pi = (0,0,....)$. If $\pi(0) > 0$, $\sum \pi(x) = \infty \neq 1$. Then, the normalization condition $\sum \pi(x) = 1$ can never hold. Then a stationary distribution does not exists, so chain is not positive recurrent. Since it's recurrent, so it must be null recurrent. ∎

■ **Example 5.5** Sample random walk with a reflecting barrier

Reflecting barrier at 0: $S = \overline{\mathbb{Z}^+} = \{0,1,2....\}$, $P_{x,x+1} = p$, $P_{x,x-1} = 1-p$ for $x \geq 1$ and $P_{0,1} = 1$.

Claim: If $p < \frac{1}{2}$, then this chain is positive recurrent. (An example of positive recurrent class having infinitely many states.)
Proof: We solve for the stationary distribution. Since only $P_{x,x+1}$ and $P_{x,x-1}$ are non-zero, we can see the detailed balance condition, then we have

$$\pi(0)\cdot 1 = \pi(1)\cdot(1-p) \quad \Longrightarrow \quad \pi(1) = \frac{1}{1-p}\pi(0)$$

and we also note that

$$\pi(x)\cdot p = \pi(x+1)\cdot(1-p) \quad x = 1,2,3... \quad \Longrightarrow \quad \pi(x+1) = \frac{p}{1-p}\pi(x)$$

Hence, we have

$$\pi(x) = \left(\frac{p}{1-p}\right)^{x-1} \cdot \frac{1}{1-p}\pi(0)$$

This is a geometric series, so

$$\sum \pi(x) < \infty$$

and a stationary distribution exists, if and only if $\frac{p}{1-p} < 1 \iff p < \frac{1}{2}$ ∎

■ **Remark 5.6** The reflected sample random walk is positive recurrent if $p < \frac{1}{2}$, it's null recurrent if $p = \frac{1}{2}$, it's transient if $p > \frac{1}{2}$

6. Branching Process

6.1 Branching Process (Galton Watson Process)

Basic Setup: Consider a population. Start from one common ancestor $X_0 = 1$. Each individual, at the end of its life, produces a random number Y of offspring. The distribution of Y is given by $P(Y = k) = p_k$ for $k = 0, 1, ...$ with $p_k \geq 0$ and $\sum_{k=1}^{\infty} p_k = 1$. The number of offspring of different individuals are independent.

Let the X_n be the number of individuals in the n-th generation. Then

$$X_{n+1} = Y_1^{(n)} + Y_2^{(n)} + ... + Y_n^{(n)}$$

where $Y_1^{(n)}, ..., Y_n^{(n)}$ are independent copies of Y, $Y_i^{(n)}$ is the number of offspring of the i-th individual in the n-th generation.

Expectation: Expected population size in the n-th generation $E(X_n)$. By **A1** we have $E(X_n) = E(X_0)$

6.2 Extinction Probability and Generating Function

Extinction (population eventually dies out) Probability: $\{X_n\}$ is a DTMC. State 0 is absorbing, all the other states are transient (as long as $p_0 > 0$). However, it does not mean that the population size goes to infinity with positive probability, then the probability of extinction is smaller than 1. To find the probability of extinction, we introduce a mathematical tool: generating function.

Definition 6.2.1 — Generating Function.
Let $\{p_0, p_1, ...\}$ be a distribution on $\{0, 1, ...\}$. Let η be a random variable following $\{p_0, p_1, ...\}$. That

is $P(y=i)=p_i$. The generating function of η or of $\{p_0,p_1,...\}$, is defined by

$$\varphi(s)=E(s^{\eta})=\sum_{k=0}^{\infty}p_k s^k \quad 0\le s\le 1$$

Proposition 6.2.1 — Properties of Generating Function.
1. $\varphi(0)=p_0$,$\varphi(1)=1$
2. Generating function determines the distribution

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

Reason: Taylor's expansion:

$$\varphi(s)=p_0+p_1s+p_2s^2+.....+p_{k-1}s^{k-1}+p_ks^k+p_{k+1}s^{k+1}+.....$$

and

$$\frac{d^k\varphi(s)}{ds^k}=k!p_k+\underbrace{(...)}_{\ge 0}\cdot s+\underbrace{(...)}_{\ge 0}\cdots^2+.... \quad (*)$$

Then

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

Also, from $(*)$ we have

$$\frac{d^k\varphi(s)}{ds^k}\ge 0 \quad \textbf{for any k and } s\in[0,1]$$

In particular, $\varphi(s)$ is non-decreasing and convex

3. Let $\eta_1,....,\eta_n$ be independent random variables with generating function $\varphi_1,....,\varphi_n$, then for $x=\eta_1+....+\eta_n$ has generating function $\varphi_x(s)=\varphi_1(s).....\varphi_n(s)$

Proof:

$$\varphi_x(s)=E(s^x)=E(s^{\eta_1}.....s^{\eta_n})=E(s^{\eta_1}).....E(s^{\eta_n})=\varphi_1(s).....\varphi_n(s)$$

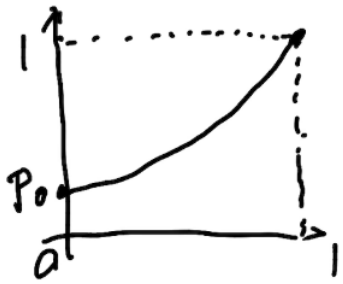
4. Moments

$$\frac{d^k\varphi(s)}{ds^k}\bigg|_{s=1}=\frac{d^kE(s^{\eta})}{ds^k}\bigg|_{s=1}=E\left(\frac{d^ks^{\eta}}{ds^k}\bigg|_{s=0}\right)=E(\eta(\eta-1)....(\eta-k+1)s^{\eta-k}\bigg|_{s=1})=E(\eta...(\eta-k+1))$$

In particular, $E(\eta)=\varphi'(1)$ and

$$Var(\eta)=E(\eta(\eta-1))+E(\eta)-(E(\eta))'=\varphi''(1)+\varphi'(1)-(\varphi(1))^2$$

The graph of a generating function:



6.3 Extinction Probability - Dynamics

Back to extinction probability. Define

$$N = \min\{n : X_n = 0\} \quad \textbf{extinction time}$$

so $\mu_n = P(N \leq n) = P(X_n = 0)$ (extinction before time n). Note that μ_n is non-decreasing and bounded from above, hence

$$\mu := \lim_{n\to\infty} \mu_n = P(N < \infty) = P(\textbf{population eventually dies out}) = \textbf{extinction probability}$$

is well defined.
Our goal is fine μ. The key step is to note that we have the following relation between μ_n and μ_{n-1}:

$$\mu_n = \sum_{k=0}^{\infty} p_k \mu_{n-1}^k = \varphi(\mu_{n-1})$$

where φ is generating function of $\{p_0, p_1, ...\}$ or equivalently, the generating function of Y.

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

$$\begin{aligned}
\mu_n = P(N \leq n) &= \sum_k P(N \leq n \mid X_1 = k) \cdot P(X_1 = k) \\
&= \sum_k \frac{P(N_1 \leq n-1, ..., N_k \leq n-1 \mid X_1 = k)}{p_k} \\
&= \sum_k p_k \mu_{n-1}^k \\
&= \varphi(\mu_k - 1)
\end{aligned}$$

where N_m is the number of steps for the sub-population m to dies out.

The problem becomes: with initial value $\mu_0 = 0$ (since $X_0 - 1$) and relation $\mu_n = \varphi(\mu_{n-1})$, what is

$$\lim_{n\to\infty} \mu_n = \mu$$

6.4 Extinction Probability - Result

Theorem 6.4.1
The extinction probability μ is the smallest intersection of $\varphi(s)$ and $f(s) = s$ or equivalently, it's the smallest solution of $\varphi(s) = s$ between 0 and 1

■ **Remark 6.1** For the above theorem we have 2 cases, $\mu < 1$ and $\mu = 1$.
Question: How to tell whether we are in case 1 or in case 2?
Answer: Check the derivative of φ at $s = 1$ (if $\varphi'(1) > f'(1) = 1$: $\mu < 1$, otherwise $\mu = 1$)

Moreover, recall that we know $\varphi'(1) = E(Y)$, then we conclude that $E(Y) > 1$ implies extinction happens with certain probability smaller than 1.

Intuitively, on average, more than 1 offspring for each individual implies the population can explode with positive probability. That means $E(Y \leq 1)$ implies extinction happens for sure on average, less than or equal to 1 offspring.

7. Basic Distributions

7.1 Exponential Distribution

If $T \sim Exp(\lambda)$, then the **C.D.F**

$$F(t) = \begin{cases} 0 & t < 0 \\ 1 - e^{-\lambda t} & t \geq 0 \end{cases}$$

also the **P.D.F** $f(t) = \lambda e^{-\lambda t} \cdot \mathbb{I}_{t \geq 0}$, also $E(T) = \frac{1}{\lambda}$ and $Var(T) = \frac{1}{\lambda^2}$
Then if $S \sim Exp(1)$, then $\frac{s}{\lambda} \sim Exp(\lambda)$

Proposition 7.1.1 — Memory-less Property.
If $T \sim Exp(\lambda)$, then

$$P(T > t+s \mid T > t) = P(T > s)$$

"How much time we still need to wait does not depend on how long we have been waiting"

Proof:

$$P(T > t+s \mid T > t) = \frac{P(T > t+s)}{P(T > t)} = \frac{1 - (1 - e^{-\lambda(t+s)})}{1 - (1 - e^{-\lambda t})} = e^{-\lambda s} = P(T > s)$$

Proposition 7.1.2 — Minimum of independent exponential random variables.
If $S \sim Exp(\lambda)$, $T \sim Exp(\mu)$ and $S \perp\!\!\!\perp T$, then $Z := \underbrace{\min\{S,T\}}_{S \wedge T} \sim Exp(\lambda + \mu)$.

Moreover, $P(S = Z) = P(S \leq T) = \frac{\lambda}{\lambda + \mu}$

Proof:

$$P(Z > t) = P(S > t, T > t) = P(S > t) \cdot P(T > t) = e^{-\lambda t} \cdot e^{-\mu t} = e^{-(\lambda+\mu)t}$$

for $t \geq 0$, so we have $Z \sim Exp(\lambda + \mu)$

Note that

$$P(s \leq T) = E(P(s \leq T \mid s)) = \int e^{-\mu s} \cdot \lambda e^{-\lambda s} ds = \frac{\lambda}{\lambda + \mu}$$

Corollary 7.1.3
If $T_1, ..., T_n$ are independent random variables with $T_i \sim Exp(\lambda_i)$, then $Z := \min\{T_1, ..., T_n\} \sim Exp(\lambda_1 + \lambda_2 + ...\lambda_n)$ and

$$P(Z = T_i) = P(T_i \leq T_1, ..., T_i \leq T_n) = \frac{\lambda_i}{\lambda_1 + ... + \lambda_n}$$

Proof: Similar as above.

"Competition" among independent exponential random variables will result in an exponential random variable with the parameter being the sum of the parameters. The probability that the $i-th$ exponent wins is $\frac{\lambda_i}{\lambda_1 + ... + \lambda_n}$

7.2 Poisson Distribution

For $X \sim Poi(\lambda)$, the **P.D.F**

$$P(X = n) = \frac{e^{-\lambda} \lambda^n}{n!}$$

for $n = 0, 1, ...$ It has $E(X) = \lambda$ and $Var(X) = \lambda$.

Theorem 7.2.1
If $X_1, ..., X_n$ are independent, $X_i \sim Poi(\lambda_i)$, then $X_1 + ... + X_n \sim Poi(\lambda_1 + ... + \lambda_n)$

Proof: Consider the generating function φ_{X_i} of X_i, so

$$\varphi_{x_i}(s) = \sum_{k=0}^{\infty} P(X_i = k) s^k = \sum_{k=0}^{\infty} \frac{e^{-\lambda_i} \lambda_i^k}{k!} s^k = \sum_{k=0}^{\infty} \frac{e^{-\lambda_i} (\lambda_i s)^k}{k!} = e^{-\lambda_i} e^{\lambda_i s} \underbrace{\sum_{k=0}^{\infty} \frac{e^{-\lambda_i s} (\lambda_i s)^k}{k!}}_{=1} = e^{-\lambda_i (1-s)}$$

By independence, the generating function of $X = X_1 + ... + X_n$ $\varphi(s)$ is

$$\varphi(s) = \varphi_{x_1}(s) ..... \varphi_{x_n}(s) = e^{-\lambda_1 (1-s)} ..... e^{-\lambda_n (1-s)} = e^{-(\lambda_1 + \lambda_2 + ... + \lambda_n)(1-s)}$$

This is the generating function of $Poi(\lambda_1 + \lambda_2 + ... + \lambda_n)$. Since the generating function determines the distribution, we conclude that $X_1 + ... + X_n \sim Poi(\lambda_1 + ... + \lambda_n)$

7.3 Counting Process

DTMC is a discrete-time process: $\{0,1,2.....\}$
WE also want to consider the case where time is continuous: $T=[0,\infty)$, we use $\{X_t\}_{t\geq 0}$ or $\{X(t)\}_{t\geq 0}$. The simplest continuous-time process is counting process, which count the number of occurrence of certain events up to time t

Definition 7.3.1
Let $0\leq S_1\leq S_2\leq .....$ be the time of occurrence of some events. Then the process $\{N(t)\}_{t\geq 0}$ given by

$$N(t)=\{n:S_n\leq t\}=\sum_{n=1}^{\infty}\mathbb{I}_{S_n\leq t}$$

is called the counting process (of events S_n with $n=1,2....$)
Equivalently, $N(t)=n$ if and only if $S_n\leq t<S_{n+1}$

■ **Example 7.1** Calls arrive at a calling center. Let S_n be the arrive time of the n-th call and $N(t)$ be the number of calls received before time t

Other examples: Cars passing a speed reader, atoms having radioaction decay. ■

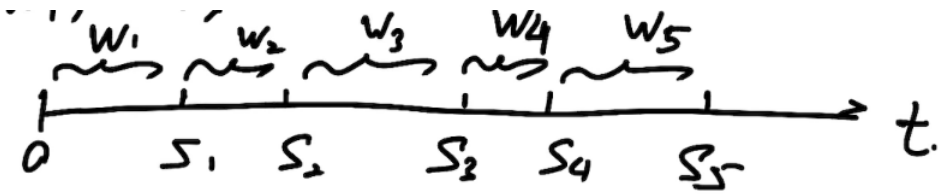
Properties of a counting process:

1. $N(t)\geq 0$ for any $t\geq 0$
2. $N(t)$ takes integer values
3. $N(t)$ is non-decreasing: $N(t_1)\leq N(t_2)$ if $t_1\leq t_2$
4. $N(t)$ is right-continuous: $N(t)=\lim_{s\to t^+}N(s)$
5. We also assume $N(0)=0$ (no event happens at time 0) and $N(t)$ only has jumps with size 1 (no two events happen at exactly the same time)

7.4 Poisson Process

Definition 7.4.1 — Interarrival Time.
Let $W_1=S_1$, $W_n=S_n-S_{n-1}$: the time between the $n-1$th event and the n-th event. The $W_1,W_2,.....$ are called interarrival times. The picture is shown below

The graph of a generating function:



Definition 7.4.2 — Renewal Process.
A **renewal process** is a counting process for which the interarrival times. $W_1, W_2, ...$ **i.i.d** All the three examples above of counting process can be reasonably modeled as renewal process.

Definition 7.4.3 — Homogeneous Poisson Process.
The **Poisson Process** $\{N(t)\}_{t\geq 0}$ is the renewal process for which the interarrival times are exponentially distributed. Thta is $W_n \sim Exp(\lambda)$ **i.i.d**. The parameter λ is called the **intensity/rate** of $\{N(t)\}_{t\geq 0}$.

Notation:

$$\underbrace{\{N(t)\}_{t\geq 0}}_{\textbf{process}} \sim Poi(\lambda t) \qquad \textbf{and} \qquad \underbrace{N(t)}_{\textbf{r.v}} \sim Poi(\lambda t)$$

7.5 Basic Properties of Poisson Processes

1. **Continuous-Time Markov Property**:

$$P(N(t_m) = j \mid N(t_{m-1}) = i, N(t_{m-2}) = i_{m-2}, ..., N(t_1) = i_1) = P(N(t_m) = j \mid N(t_{m-1}) = i)$$

for any $m, t_1 < t_2 < ..... < t_m, i_1, i_2, ..., i_{m-2}, i, j \in S$

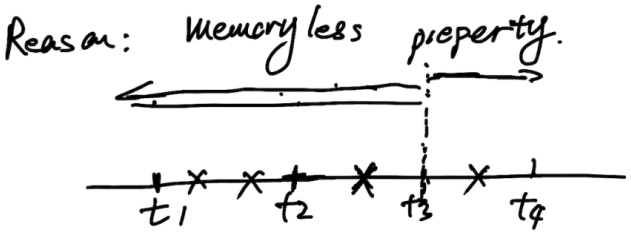
Fact: The Poisson Process is the only renewal process having the markov property.

Reason: $N(t_{m-1}) = i$ only tells i events happened before t_{m-1}, it does not tell when the last event occurred. The "history", $N(t_{m-2})$,..... tells more about when the last event occurred.

Hence, we have **Markov Property $\implies$ Given how many events occurred before, when the last event occurred has no influence on when the next event will occur. i.e. How long we have waited for the next event has no influence on how long we still need to wait.**

$\impliedby$ Memory-less Property. Since the exponential distribution is the clearly (continuous-time) distribution which is memory-less, the Poisson process is the only renewal process which has the Markov property.

2. **Independent Increments**:
If $t_1 < t_2 \leq t_3 < t_4$, then $N(t_2) - N(t_1) \perp\!\!\!\perp N(t_4) - N(t_3)$



7.6 Poisson Increments

Proposition 7.6.1

$$N(t_2) - N(t_1) \sim Poi(\lambda(t_2 - t_1))$$

In particular,

$$N(t) = N(t) - N(0) \sim Poi(\lambda t)$$

Reason: By the memory-less property of exponential **r.v.s** it suffices to show

$$N := N(t_2 - t_1) \sim Poi(\lambda(t_2 - t_1))$$

Proof: Note that

$$\begin{aligned} N = n &\iff S_n \le t_2 - t_1 < S_{n+1} \\ &\iff W_1 + ... + W_n \le t_2 - t_1 \quad \textbf{and} \quad W_1 + ... + W_n + W_{n+1} > t_2 - t_1 \end{aligned}$$

Fact: $W_1, ..., W_n$ **i.i.d r.v.s** following $Exp(\lambda)$, then

$$W_1 + ... + W_n \sim Erlang(n, \lambda) \quad (\textbf{Erlang Distribution - special type of Gamma distribution})$$

The **c.d.f** of $Erlang(n, \lambda)$ is

$$F(x) = 1 - \sum_{k=1}^{n-1} \frac{1}{k!} e^{-\lambda x} (\lambda x)^k$$

Hence here, we have

$$P(W_1 + ... + W_n \le t_2 - t_1) = 1 - \sum_{k=1}^{n-1} \frac{1}{k!} e^{-\lambda(t_2 - t_1)} (\lambda(t_2 - t_1))^k$$

Similarly, we have

$$P(\underbrace{W_1 + ... + W_n + W_{n+1}}_{\textbf{Erlang(n+1,}\lambda\textbf{)}} \le t_2 - t_1) = 1 - \sum_{k=1}^{n} \frac{1}{k!} e^{-\lambda(t_2 - t_1)} (\lambda(t_2 - t_1))^k$$

and

$$P(N = n) = P(W_1 + ... + W_n \le t_2 - t_1) - P(W_1 + ... + W_n + W_{n+1} \le t_2 - t_1) = \underbrace{\frac{1}{n!} e^{-\lambda(t_2 - t_1)} (\lambda(t_2 - t_1))^n}_{\textbf{p.m.f of } Poi(\lambda(t_2 - t_1)) \textbf{ at } n}$$

Therefore, we have $N \sim Poi(\lambda(t_2 - t - 1))$

■ **Remark 7.1**

1. As a result of the Poisson increment property, $N(1) = Poi(\lambda)$, $E(N(1)) = \lambda$. This is why λ is called the "intensity/rate" of the process: it is the expected number of arrivals/occurrence in one unit of time.

2. Note that hte distribution of the increments, together with the independence of the increments, uniquely, determines the distribution of the process:

$$(N(t_1),N(t_2),.....,N(t_k)) \iff (N(t_1)-N(t_0),N(t_2)-N(t_1),.....,N(t_k)-N(t_{k-1}))$$

and each $N(t_i)-N(t_{i-1})$ is independent. That means $Poi(\lambda t_1) \perp\!\!\!\perp Poi(\lambda(t_2-t_1)) \perp\!\!\!\perp ..... \perp\!\!\!\perp Poi(\lambda(t_k-t_{k-1}))$ Then, we can equivalently define the Poisson process as follows

Definition 7.6.1 — Alternative Definition of Poisson Process.
$\{N(t)\}_{t\geq 0}$ is a Poisson process if

(1) $N(0)=0$
(2) $N(t)-N(s) \sim Poi(\lambda(t-s))$ for $0 \leq s \leq t$
(3) $t_0 < t_1 < .... < t_n$, then

$$N(t_1)-N(t_0),N(t_2)-N(t_1),.....,N(t_k)-N(t_{k-1})$$

are independent.

In this case, our original definition becomes a property of Poisson Process: Poisson process are counting processes at events with **i.i.d** exponential interarrival times.

7.7 Combining Poisson Processes

Theorem 7.7.1
Let $\{N_1(t)\}_{t\geq 0}$ and $\{N_2(t)\}_{t\geq 0}$ be two independent Poisson processes, with intensities λ_1,λ_2 respectively. Then

$$N(t) = N_1(t)+N_2(t)$$

is also Poisson process, with intensity $\lambda = \lambda_1+\lambda_2$

Proof: Check the alternative definition.

(1) $N(0) = N_1(0)+N_2(0) = 0$
(2) For $0 \leq s < t$

$$N(t)-N(s) = N_1(t)-N_1(s)+N_2(t)-N_2(s)$$

By the property of Poisson **r.v.s** $N(t)-N(s) \sim Poi((\lambda_1+\lambda_2)(t-s))$
(3) For $t_0 < t_1 < .... < t_n$

$$\begin{aligned}&(N(t_1)-N(t_0),.....,N(t_n)-N(t_{n-1})\\ &= (N_1(t_1)-N_1(t_0)+N_2(t_1)-N_2(t_0),.....,N_1(t_n)-N_1(t_{n-1})+N_2(t_n)-N_2(t_{n-1}))\end{aligned}$$

The operators are jointly independent since $N_i(t_1)-N_i(t_0),.....,N_i(t_n)-N_i(t_{n-1})$ are jointly independent for $i=1,2$, and terms of N_1 and terms of N_2 are also independent. Then, we have $\{N(t)\}_{t\geq 0}$ is a Poisson process with intensity $\lambda_1+\lambda_2$, completes the proof.

In general, let $\{N_1(t)\}_{t\geq 0}, ...\{N_k(t)\}_{t\geq 0}$ be independent Poisson processes with intensities $\lambda_1, ..., \lambda_n$, then

$$N(t) = \sum_{i=1}^{k} N_i(t)$$

is a Poisson process with intensity $\sum_{i=1}^{k} \lambda_i$

7.8 Splitting Poisson Processes

Consider a Poisson process with intensity λ as the counting process of the events with **i.i.d** exponential interarrival times. For each events, makr it with "1" with prob p, with "2" with prob $1-p$. The marking of different events are independent.

Let N_1 and N_2 be the counting process of the event with marks "1" and "2", respectively.

Theorem 7.8.1
The $\{N_1(t)\}_{t\geq 0}$ and $\{N_2(t)\}_{t\geq 0}$ are **independent** Poisson processes with intensities $p\lambda$ and $(1-p)\lambda$ respectively. (Intuition: This is the inverse procedure of combining two independent Poisson processes into one Poisson process)

Proof: Again we check the alternative definition.

(1) $N_1(0) = 0, N_2(0) = 0$
(2) Since $N(t) - N(s) = N(t-s) \sim Poi(\lambda(t-s))$ are the splitting rule does not change ever time, it suffices to consider the case where $s = 0$. Consider the joint distribution:

$$\begin{aligned} P(N_1(t) = m, N_2(t) = n) &= P(N_1(t) = m, N_2(t) = n \mid N(t) = m+n) \cdot P(N(t) = m+n) \\ &= \binom{m+n}{m} p^m (1-p)^n \cdot e^{-\lambda t} \frac{(\lambda t)^{m+n}}{(m+n)!} \\ &= \underbrace{e^{-p\lambda t} \frac{(p\lambda t)^m}{m!}}_{\textbf{p.m.f of } Poi(p\lambda t) \textbf{ at } m} \cdot \underbrace{e^{-(1-p)\lambda t} \frac{((1-p)\lambda t)^n}{n!}}_{\textbf{p.m.f of } Poi((1-p)\lambda t) \textbf{ at } n} \end{aligned}$$

This means
1. $N_1 \perp\!\!\!\perp N_2(t)$
2. $N_1(t) \sim Poi(p\lambda t)$ and $N_2(t) \sim Poi((1-p)\lambda t)$

Therefore, we have the Poisson increments property that we want.

(3) Since $N(t)$ has independent increments, and the marking is also independent, $N_1(t)$ and $N_2(t)$ also have independent increments. Then $\{N_1(t)\}$ and $\{N_2(t)\}$ are Poisson process with intensities $p\lambda$ and $(1-p)\lambda$ respectively. Note that $N_1(t) \perp\!\!\!\perp N_2(t)$, $t \geq 0$ is not yet enough for N_1 and N_2 to be independent as two

processes. We need

$$(N_1(t_0),.....,N_1(t_n)) \perp\!\!\!\perp (N_2(t_0),.....,N_2(t_n))$$

for all n and $t_0 < t_1 < ..... < t_n$. This follows from the independent increment property. For example, $N_1(t_0) \perp\!\!\!\perp N_2(t_0)$, $N_1(t_0) \perp\!\!\!\perp (N_2(t_1) - N_2(t_0))$

■ **Remark 7.2** When we are only interested in N_1, "splitting" is also called "thinning"

7.9 Order Statistics

Definition 7.9.1 — Order Statistics.
Let $X_1, X_2, ..., X_n$ be (typically i.i.d) r.v.s, the **order statistics** of $X_1, X_2, ..., X_n$ are r.v.s defined as follows:

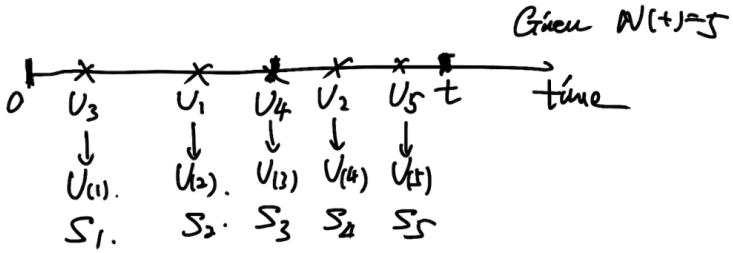
$$X_{(1)} = \min\{X_1, ..., X_n\}$$

$X_{(2)}$ be the 2nd smallest of $\{X_1, ..., X_n\}$,....., $X_{(n)} = \max\{X_1, ..., X_n\}$. Note that

$$(X_{(1)}, X_{(2)}, ..., X_{(n)})$$

is $(X_1, X_2, ..., X_n)$ rearranged in non-decreasing order.

Conditional on $N(t) = n$, the occurrence/arrival times before t are distributed as the order statistics of n i.i.d uniform $[0,t]$ random variables.



Theorem 7.9.1
Let $\{N(t)\}_{t\geq 0}$ be a Poisson process with intensity λ. Condition on $N(t) = n$, the occurrence times of the events in $[0,t]$ are distributed as the order statistics of n i.i.d uniformly distributed r.v.s on $[0,t]$. That is

$$(S_1, ..., S_n \mid N(t) = u) = (U_{(1)}, ..., U_{(n)})$$

where S_i is the time of the i-th event, the U_i is i.i.d $Unif[0,t]$, and $U_{(1)}, ..., U_{(n)}$ are the order statistics of $U_1, ..., U_n$

Proof: Let $a_1 < b_1 < a_2 < b_2 < ..... < a_n < b_n < t$, then

$$\begin{aligned}
&P(S_i \in (a_i, b_i],\ i = 1, ..., n \mid N(t) = n) \\
&= \frac{P(S_i \in (a_i, b_i],\ i = 1, ..., n,\ N(t) = n)}{P(N(t) = 0)} \\
&= \frac{P(N(a_1) = 0, N(b_1) - N(a_1) = 1, N(a_2) - N(b_1) = 0, N(b_2) - N(a_1) = 1, ..., N(t) - N(b_n) = 0)}{P(N(t) = 0)}
\end{aligned}$$

Note that the numerator is equal to

$$= e^{-\lambda a_1} \cdot \lambda(b_1 - a_1) \cdot e^{-\lambda(b_1 - a_1)} \cdot e^{-\lambda(t - b_n)} = \underbrace{e^{-\lambda(a_1 + (b_1 - a_1) + (a_2 - b_1) + ... + (t - b_n))}}_{= e^{-\lambda t}} \cdot \lambda^n \prod_{i=1}^{n} (b_i - a_i)$$

This gives us that

$$P(S_i \in (a_i, b_i],\ i = 1, ..., n \mid N(t) = n) = \frac{e^{-\lambda t} \lambda^n \prod_{i=1}^{n} (b_i - a_i)}{e^{-\lambda t} \frac{(\lambda t)^n}{n!}} = \frac{n!}{t^n} \prod_{i=1}^{n} (b_i - a_i)$$

divide both sides by $\prod_{i=1}^{n} (b_i - a_i)$ and take limits $b_i \to a_i$, we have the conditional **p.d.f**

$$f_{S_1, ..., S_n | N(t) = n}(a_1, ..., a_n) = \frac{n!}{t^n} \mathbb{I}_{a_1 < a_2 < .... < a_n}$$

this is the **p.d.f** of $(U_{(1)}, ..., U_{(n)})$

How to simulate the () events of a Poisson process until the time t:

Method 1: Simulate i.i.d $Exp(\lambda)$, until their sum exceeds t

Method 2: Simulate $a\ Poi(\lambda t)$, denote it as N. Then simulate N i.i.d $Unif[0,t]$ rearrange them by increasing order.

Corollary 7.9.2

$$N(s) \mid N(t) = n \sim Bin\left(n, \frac{s}{t}\right)$$

for $s \leq t$.
Proof:

$$\begin{aligned}
P(N(s) = k \mid N(t) = n) &= P(s_1, s_2, ..., s_k \leq s, S_{k+1}, ..., S_n > s \mid N(t) = n) \\
&= P(U_{(1)}, U_{(1)}, ..., U_{(k)} \leq s, U_{(k+1)}, ..., U_{(n)} > S) \\
&= P(\textbf{k out of n i.i.d unif[0,t], r.v.s are} \leq \textbf{S}) \\
&= \binom{n}{k} \left(\frac{s}{t}\right)^k \left(1 - \frac{s}{t}\right)^{n-k}
\end{aligned}$$

Hence, we have $N(s) \mid N(t) = n \sim Bin\left(n, \frac{s}{t}\right)$ for $s \leq t$.

7.10 Nonhomogeneous Poisson Process and Compound Poisson Process

Definition 7.10.1 The $\{N(t)\}_{t\geq 0}$ is a (nonhomogeneous) Poisson process with rate $\lambda(r)$, if

(1) $N(0)=0$
(2) $N(t)-N(s)\sim Poi\big(\int_s^t \lambda(r)dr\big)$
(3) $N(t)$ has independent increment i.e. $N(t_1)-N(t_0),.....,N(t_n)-N(t_{n-1})$ are independent for $t_0<t_1<....<t_n$

■ **Remark 7.3** We see that this definition is basically the same as the alternative definition of (homogeneous) Poisson process with (2) replacing the original condition $N(t)-N(s)\sim Poi(\lambda(t-s))$. Indeed, since $\lambda(t-s)=\int_s^t \lambda dr$, the regarded as a special case of the nonhomogeneous Poisson process where the rate function $\lambda(r)=\lambda$ with $r\geq 0$. In order to betten understand the nonhomogeneous Poisson process, consider

$$\begin{aligned} P(\textbf{there exists at one point in a small interval } (t,t+At]) &= P(N(t+\Delta t)-N(t)\geq 1)\\ &= 1-P(N(t+\Delta t)-N(t)=0)\\ &= 1-e^{\int_t^{t+\Delta\lambda(r)dr}} \end{aligned}$$

when At is small ($At\to 0$), we have $\int_t^{t+At}\lambda(r)dr\to 0$. Then

$$e^{\int_t^{t+\Delta t}\lambda(r)dr}=1-\int_t^{t+\Delta t}\lambda(r)dr+0\left(\int_t^{t+\Delta t}\lambda(r)dr\right)$$

so we have

$$P(N(t+\Delta t)-N(t)\geq 1)=\int_t^{t+\Delta t}\lambda(r)dr+0\left(\int_t^{t+\Delta t}\lambda(r)dr\right)\approx\lambda(t)\Delta t$$

when Δt is small and $\lambda(t)$ is continuous. "attractioness" of location t

Homogeneous Poisson Process: All the locations are equally attractive.
Non-Homogeneous Poisson Process: Some locations are more attractive than some other locations.

■ **Example 7.2** If we want to modeled the calls received by a call center in a relatively long time period, then a non-homogeneous Poisson process is more suitable. People are more likely to call in some hours (9am) than in some other hours (3am). What are still true and what are no longer true for non-homogeneous Poisson processes:

Still True: Counting process, Markov property, Independent increments, Poisson increments, Combining and splitting, Order Statistics Property

Not True: Exponential interarrivial time, Renewal process

However, the order statistics property is no longer uniform r.v.s, but i.i.d r.v.s with density $f(s)=\dfrac{\lambda(s)}{\int_0^t\lambda(s)ds}$ for $s\in[0,t]$ ■

7.11 Compound Poisson Process

Each arrival/occurrence is associated with a quantity. Quantities associated with different arrivals /occurrence are i.i.d. We are interested in the total quantity up to time t.

$$S(t) = Y_1 + ... + Y_{N(t)}$$

and each Y_i is i.i.d

■ **Example 7.3** Claims arrive at an insurance company. The number of claims can be modelled by a Poisson process. The total amount of claims can be models by a compound Poisson process. The mean and variance of $S(t)$ can be calculated by the following result: ■

Proposition 7.11.1
Let X, Y be two r.v.s, then $Var(X) = E(Var(X \mid Y)) + Var(E(X \mid Y))$ where $Var(X \mid Y) = E((X - E(X \mid Y))^2 \mid Y)$

Proof:

$$\begin{aligned} &Var(X) = E((X - E(X))^2) \\ &= E\left[((X - E(X \mid Y) + (E(X \mid Y) - E(X))^2)\right] \\ &= E[(X - E(X \mid Y))^2] + E[(E(X \mid Y) - E(X))^2] + 2E[(X - E(X \mid Y)) \cdot (E(X \mid Y) - E(X))] \end{aligned}$$

Note that

$$\begin{aligned} E[(X - E(X \mid Y)) \cdot (E(X \mid Y) - E(X))] &= E[E[(X - E(X \mid Y)) \cdot (E(X \mid Y) - E(X)) \mid Y]] \\ &= E[(E(X \mid Y) - E(X)) \cdot \underbrace{E(X - E(X \mid Y) \mid Y)}_{=E(X|Y) - E(E(X|Y)|Y)}] \\ &= E[(E(X \mid Y) - E(X)) \cdot 0] \\ &= 0 \end{aligned}$$

also we have

$$E((X - E(X \mid Y))^2) = E[E[(X - E(X \mid Y)^2) \mid Y]] = E(Var(X \mid Y))$$

Combining all the results above, finishes the proof.

Theorem 7.11.2
Let $Y_1, Y_2, ...$ be i.i.d, r.v.s and N is a non-negative integer valued r.v., independent of $\{Y_i\}$, let $S = Y_1 + ... + Y_N$, then

(1) If $E(Y_1) = \mu$, $E(N) < \infty$, then $E(S) = \mu \cdot E(N)$
(2) If $Var(Y_1) = \sigma^2$, $Var(N) < \infty$, then $Var(S) = \sigma E(N) + \mu^2 Var(N)$

In particular, if $N \sim Poi(\lambda)$, then $Var(S) = \lambda E(Y_1^2)$ where $E(Y_1^2) = (E(Y_1))^2 + Var(Y_1)$

Proof: For (1), this is simply the basic identity. For (2), we note that $Var(S) = Var(\sum_{i=1}^{N} Y_i)$ and

$$E\left(\sum_{i=1}^{N} Y_i \mid N\right) = N \cdot E(Y_1) = \mu N$$

$$Var\left(\sum_{i=1}^{N} Y_i \mid N\right) = N \cdot Var(Y_1) = \sigma^2 N$$

Then

$$Var\left(\sum_{i=1}^{N} Y_i \mid N = n\right) = Var\left(\sum_{i=1}^{n} Y_i \mid N = n\right) = Var\left(\sum_{i=1}^{n} Y_i\right) = N \cdot Var(Y_1)$$

so we have

$$var(S) = Var\left(E(\sum_{i=1}^{N} Y_i \mid N)\right) + E\left(Var\left(\sum_{i=1}^{N} Y_i \mid N\right)\right) = Var(\mu N) + E(\sigma^2 N) = \mu^2 Var(N) + \sigma^2 E(N)$$

when $N \sim Poi(\lambda)$, then

$$Var(S) = \mu^2\lambda + \sigma^2\lambda = \lambda\left(E(Y_1)^2 + Var(Y_1)\right) = \lambda E(Y_1^2)$$

Corollary 7.11.3
For a Poisson process with rate λ, $S(t)$ has mean $\lambda t\mu$, variance is $\lambda t E(Y_1^2)$, which is $\sigma^2\lambda t$

7.12 Epilogue

What's next?