在随机过程中，将随机变量看做关于时间的变量，定义其所有可以取值的集合为 *State Space*，简写为 $\mathcal{S}$。且目前只讨论 $\mathcal{S}$ 是有限或者是有限可数的情况。通常也可以将其结论推广至无限应用。

1. **Discrete Time Markov Chain**

   (a) **Basic Notations and Definition**

   **Definition 1.** (Markov Chain 马尔科夫链) Let $\{X_n\}_{n\in\mathcal{N}}$ be a series of r.v in $\mathcal{S}$. Then we say $\{X_n\}_{n\in\mathcal{N}}$ is a Markov Chain or has Markov properties if: $\forall\, n\in\mathcal{N}, \forall\, x_i\in\mathcal{S}$

   $$\mathcal{P}(X_{n+1}=x_{n+1}\mid X_n=x_n,\, X_{n-1}=x_{n-1},\,\dots,X_0=x_0)=\mathcal{P}(X_{n+1}\mid X_n=x_n)$$

   This means the next behavior only depends on the current status and nothing to do with how it get here. For example, the random walk, your position next second depends only on where you are now.

   注意以上的符号用法中，我们用 $X_n$ 的下标 $n$ 指代时刻 $n$，即 $t=n$ 时（假定时间是离散的）并假设无论在哪一个时刻在给定的 *state* 下，都有固定的概率分布，该性质称之为 *time homogeneity*。数学定义为

   **Definition 2.** (Time Homogeneity) We say a Markov chain is time homogeneity if

   $$\forall\,(x,y)\in\mathcal{S}^2,\ \mathcal{P}(X_{n+1}=x\,|\,X_n=y)$$

   is all the same for all time $n$.

   Summary: Markov Chain 的两点非常强的假设

   - 下一时刻的状态只依赖于当前状态，与之前到达路径无关 (Markvo Property)
   - 无论哪个时刻，两两状态间的概率不变 (Time Homogeneity)

   对于概率通常用 $\mathcal{P}_{ij}$ 表示，意为目前在 *state* $=X_n=i$ 状态下，下一步到 $X_{n+1}=j$ 的概率。且由于是概率分布，有结论

   $$p_{ij}>0,\ \ \sum_{j=0}^{\infty}p_{ij}=1$$

   The second property can be proved made up later. Then consequently we can define the *transition matrix* of M.C as the collection of all element $\mathcal{P}_{ij}$ as

   $$(\underset{\sim}{\mathcal{P}})_{ij}=\mathcal{P}_{ij}$$

   即矩阵的第 $i$ 行 $j$ 列个元素是从 *state* $i$ 到 *state* $j$ 的概率。因此行向量各个元素和为 1。注意 *transition matrix* 一定是方阵。

   *Remark* 1. *Let the prob distribution of $X_n$ (at time $n$) is $\mu=(\mu_1,\dots,\mu_N)$. Then the prob. distribution of $X_{n+1}$ is $\mu\cdot\underset{\sim}{\mathcal{P}}$ by definition of matrix product.*

   *Proof.* Now we are asking for $\mathcal{P}(X_{n+1}=s_j)$(Notice this is different form asking the M.C properties). First we have

   $$\mathcal{P}(X_{n+1}=s_j)=\sum_{k\in\mathcal{S}}\mathcal{P}(X_{n+1}=s_j,\,X_n=s_k)$$

   this is because $\{X_n=s_k\}_{k\in\mathcal{S}}$ form a complete set of mutually exclusive event. Then by conditional probability we have

   $$=\sum_{k\in\mathcal{S}}\mathcal{P}(X_{n+1}=s_j\,|\,X_n=s_k)\mathcal{P}(X_n=s_k)$$

   $$=\sum_{k\in\mathcal{S}}\mu_k\cdot(\underset{\sim}{\mathcal{P}})_{kj}$$

   which is exactly the matrix product. □

值得注意的是，这里的矩阵乘法是向量右乘矩阵，而非一般的左乘。最后一步的 $\mu$ 的角标指的是第 $k$ 个 state。

*Remark* 2. (When State space is infinite) When $\mathcal{S}$ is not finite or countable, then we can generalize all concepts to infinite case. For example the transition matrix with infinite rows and column.

*Remark* 3. If we start with $X_0$ under an initial distribution $\mu = (\mu_1, \mu_2)$(consider two states case, $\mu$ is a prob distribution on $\mathcal{S}$) with corresponding transition matrix $\underset{\sim}{\mathcal{P}}$. Then we have the prob. distribution of $X_n$ equals $\mu \cdot \underset{\sim}{\mathcal{P}}^n$.

Does the initial prob distribution the same as a row in the transition matrix?

个人理解①所谓 *initial distribution* 是说目前整个随机过程还并未开始，该概率分布指的是在第一个时刻进入其中某个 *state* 的概率，与 *transition matrix* 中元素代表的概率不同。②无需纠结于 Remark3。形式上看似 $\mu \cdot \underset{\sim}{\mathcal{P}}^n$ 并非只依赖于上一时刻的状态（因为至少有 $\mu$ 在，看似依赖于初始分布）。其实所谓的只依赖于上一刻是指的上一个状态是 $\mu \cdot \underset{\sim}{\mathcal{P}}^{n-1}$，即需要上一时刻的概率分布必须是充分的。一直右乘同一个矩阵说明性质 *time homo.*。

**Definition 3.** (Stationary Distribution) Let $\{X_n\}_{n\in\mathcal{N}}$ be a M.C with transition matrix $\underset{\sim}{\mathcal{P}}$, and $\mu$ be a distribution on $\mathcal{S}$. We say that $\mu$ is a *stationary distribution* if $\mu \cdot \underset{\sim}{\mathcal{P}} = \mu$.

**Proposition 1.** *If $\{X_n\}_{n\in\mathcal{N}}$ converges to a distribution $\mu$ as $n \to \infty$, then $\mu$ is stationary.*

*Proof.* Let $\mu_k$ be the prob. distribution of $X_k$. Then

$$u_{n+1} = \mu_n \cdot \underset{\sim}{\mathcal{P}} \overset{n\to\infty}{\longrightarrow} \mu = \mu \cdot \underset{\sim}{\mathcal{P}}$$

Thus stationary. □

**Definition 4.** (Graph Representation of a transition matrix) Using a *directed graph* where each node represents a state and we draw arrow from $x$ to $y$ is $\mathcal{P}_{xy} > 0$, with weight $\mathcal{P}_{xy}$. This graph is called a *transition diagram.*

(b) **Chapman-Kolmogorov Equation**

**Definition 5.** (Generalized Remark3) For $n \in \mathcal{N}$, we define the *n-step transition probability* as

$$\mathcal{P}^n_{ij} = \mathcal{P}(X_n = j \mid X_0 = i)$$

and this consequently defines the *n-step transition (prob.) matrix* $(\underset{\sim}{\mathcal{P}}^{(n)})_{ij} = \mathcal{P}^n_{ij}$（等式左边的 $(n)$ 指的是 *step n*，右边的 $n$ 是指数）.

*Remark* 4. One can also show (by induction on $n$ and appeal to time homo.) that

$$\forall\ k \in \mathcal{N},\ \ \mathcal{P}(X_{k+n} = j \mid X_k = i) = \mathcal{P}^n_{ij}$$

A attempt should be made later and make up here

**Definition 6.** (Chapman-Kolmogorov Equation) The Chapman-Kolmogorov Equation provides a method fpr solving the n-step probability. The equation is

$$\mathcal{P}^{n+m}_{ij} = \sum_{k=0}^{\infty} \mathcal{P}^n_{ik} \cdot \mathcal{P}^m_{kj}$$

Interpretation of this equation is, starting from state $i$ and go through state $k$ at $t = n$ on the way to state $j$ where finally land at $t = m + n$.

*Proof.*

$$\mathcal{P}_{ij}^{n+m} = \mathcal{P}(X_{n+m} = j \mid X_0 = o)$$

$$= \sum_{k=0}^{\infty} \mathcal{P}(X_{n+m} = j,\, X_n = k \mid X_0 = i)$$

this is simply the total probability in case of conditional probability. Then

$$= \sum_{k=0}^{\infty} \mathcal{P}(X_{n+m} = j \,|X_n = k, X_0 = i) \cdot \mathcal{P}(X_n = k \mid X_0 = i)$$

this is achieved by

$$\mathcal{P}(A \cap B \mid C) = \frac{\mathcal{P}(A \cap B \cap C)}{\mathcal{P}(C)} = \frac{\mathcal{P}(A \mid B \cap C)\mathcal{P}(B \cap C)}{\mathcal{P}(C)} = \mathcal{P}(A \mid B \cap C) \cdot \mathcal{P}(B \mid C)$$

then by the definition of step transition we have

$$= \sum_{k=0}^{\infty} \mathcal{P}_{ik}^{n} \cdot \mathcal{P}_{kj}^{m}$$

□

(c) **Classification of States(4.3 Ross)**

- (Accessibility) A state is *accessible* from state $i$ if $\mathcal{P}_{ij}^n > 0$ for some $n \geq 0$.
- (Community) Two sates $i$ and $j$ are *communicate* (i ↔ j) if $i$ is accessible form $j$ and $j$ is accessible from $ii$.

Above properties are can easily be checked by transition diagram. Communication is an *equivalence* relation which satisfies *reflexivity, symmetric and transitivity.* Reflexivity and symmetric are trivial by definition. Here prove transitivity.

*Proof.* (Transitivity) Want to prove: if $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$. The given condition can be interpreted as

$$\exists\ m \quad s.t\ \mathcal{P}_{ij}^m > 0, \ \ \exists\ n \quad s.t.\ \mathcal{P}_{jk}^n > 0$$

then by Chap.-Kolm. Equation we have

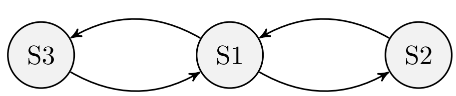
$$\mathcal{P}_{ik}^{m+n} = \sum_{\rho}^{\infty} \mathcal{P}_{i\rho}^{m} \cdot \mathcal{P}_{\rho k}^{n} \geq \mathcal{P}_{ij}^{m} \cdot \mathcal{P}_{jk}^{n} > 0$$

the $\geq$ is because the sum must include the case $\rho = j$, that is $j$ is just one case in the summation. Vice versa need to show $k$ can achieve $i$ in the same way. □

Therefore we can partition the space state into *communicating class.*

**Definition 7.** (Irreducibility) A M.C is said to be *irreducible* if there is only one communicating class (e.g all states communicates with each other).

For example the one dimensional random walk.



**Definition 8.** (Period) The state $i$ has *period* $d$ if $\mathcal{P}_{ii}^n = 0$ whenever $n$ is not divisible by $d$, and $d$ is the biggest integer with this property. If $d = 1$ then, $i$ is called *aperiodic*. Formally the definition is

$$p = gcd\{n : \mathcal{P}(X_n = i) \,|\, X_0 = i\}$$

where *gcd* is the greatest common divisor. Or another more understandable interpretation is: *The period of a state i is the greatest common denominator (gcd) of all integers n > 0, for which* $\mathcal{P}_{ii}^n > 0$.

Notice the *"period"* usually referred to a particular state. Each state may have different period.

**Proposition 2.** *If $i$ has period $d$ and $i \leftrightarrow j$, then $j$ also has period $d$.*

*Proof.* By community we have

$$\exists\, n,\, m \geq 0 \; s.t. \; \mathcal{P}_{ij}^n > 0 \,\&\, \mathcal{P}_{ji}^m > 0$$

Since $i$ has period $d$, so we have, $\forall\, k$ s.t. $\mathcal{P}_{jj}^k > 0$, we have

$$\mathcal{P}_{ii}^{m+n} \geq \mathcal{P}_{ij}^n \mathcal{P}_{ji}^m > 0 \implies d \,|\, m + n$$

$$\mathcal{P}_{ii}^{m+n+k} \geq \mathcal{P}_{ij}^n \mathcal{P}_{jj}^k \mathcal{P}_{ji}^m > 0 \implies d \,|\, m + n + k$$

both from C-K-E. The | means devisable by (e.g $d \,|\, m + n$ means $(m + n)/d \in \mathcal{Z}$). Then we have

$$d \,|\, k \implies d \leq d'$$

where $d'$ is the period of $j$. By symmetric we can show $d' \leq d$ so pushing $d' = d$. □

*The consequence of the proposition is that the periodicity is a class property (i.e all states in the same class has this property).*

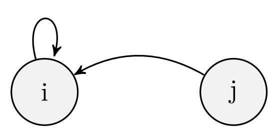
**Definition 9.** (Recurrence and Transience) Let $f_i = \mathcal{P}(X_n = i \text{ for some } n \geq 1 \,|\, X_0 = i)$. We say $i$ is *recurrent* if $f_i = 1$, and $i$ is *transient* if $f_i < 1$.

注意 $f_i$ 的定义：必须从零时刻时，从该 state 开始 . $f_i$ 实则只是一个描述性的定义，实际操作中无太大意义。

This definition measure the ability of a M.C to revisit a state. Another good property is, the state is either transient or recurrence, so we can split the state space into recurrent and transient spaces.

*Remark* 5. If $\mathcal{P}_{ii} = 1$ and $i$ is recurrent, then $i$ is called a absorbing state.

The absorbing state $i$ usually looks like




注意定义中 $\mathcal{P}_{ii} = 1$ 意思是下一步只能到自己，recurrent 意为一定可以回来。所以回来则留下不离开 $i$ 为 obsorbing。

(d) **Properties of Recurrence and Transience**

We first define the random variable $\mathcal{N}_i$ as

$$\mathcal{N}_i = \#\{X_n = i \,|\, n \geq 0\}$$

and it has a range $\mathcal{N} \cup \infty$. It can also be represented as

$$\mathcal{N}_i = \sum_{n=0}^{\infty} \mathbb{1}_{(X_n = i)}$$

简而言之，是一个计数过程。随机过程不断进行，每当其中一个时刻 $X_k = i$，$\mathcal{N}_i$ 就加一。且 $\mathcal{N}_i$ 的和中要求 $n = 0$ 开始，暗指了必须从时刻 0 开始计数。where $\mathbb{1}_{(X_n=i)}$ is 1 if $X_n = i$ or 0 if $X_n \neq i$, just like a counting process or accumulator. Also if we think of $\mathbb{1}_{(X_n=i)}$ as a random variable, it follows a *Bernoulli distribution* (either is or not). So

$$\mathbb{E}(\mathbb{1}_{(X_n=i)}) = 1 \cdot \mathcal{P}(X_n = i) + 0 \cdot \mathcal{P}(X_n \neq i) = \mathcal{P}(X_n = i)$$

注意，上式的期望值当 $n$ 取之不同时也不相同。

**Proposition 3.** • *If $i$ is recurrent, then $\mathcal{P}(\mathcal{N}_i = \infty \mid X_0 = i) = 1$*

- *If $i$ is transient and $X_0 = i$, then $\mathcal{N}_i \sim$ Geom(1 - $f_i$). (i.e, $\mathcal{P}(\mathcal{N}_i = m) = f_i^{m-1}(1 - f_i)), \mathbb{E}(\mathcal{N}_i) = 1/(1 - f_i)$.*

Interpretation of above proposition is, a recurrent sate is revisited $\infty$ many times, while for a transient state, then chain doesn't revisit the state after a certain times (and the probability to revisit is $f_i$). proof made up later Jan 13th.

*Remark* 6. • If state $i$ is recurrent, then $\mathbb{E}(\mathcal{N}_i \mid X_0 = i) = \infty$ (if transient, $\mathbb{E} < \infty$).

- If state space is finite, then some state must be recurrent.

The first one is trivial may have some formal proof here. The proof for second one is —- we assume all states are transients, so there is no state visited after a certain time, so contradicted.

Now we use *n-step transition probability* to characterize recurrence and transience.

**Proposition 4.** *We say state $i$ is recurrent if*

$$\sum_{n=0}^{\infty} \mathcal{P}_{ii}^n = \infty$$

*is transient if*

$$\sum_{n=0}^{\infty} \mathcal{P}_{ii}^n < \infty$$

Less than $\infty$ means the probability is a concrete number (finite).

*Proof.* Only need to prove the expectation is the sum of n-step transition matrix.

$$\mathbb{E}[\mathcal{N}_i \mid X_0 = i] = \mathbb{E}[\sum_{n=0}^{\infty} \mathbb{1}_{(X_n=i)} \mid X_0 = i]$$

$$= \sum_{n=0}^{\infty} \mathbb{E}[\mathbb{1}_{(X_n=i)} \mid X_0 = i] = \sum_{n=0}^{\infty} \mathcal{P}_{ii}^n$$

□

**Collary 1.** *If $i$ is recurrent and $i \leftrightarrow j$, then $j$ is also recurrent*

*Proof.* The recurrent here is used to be integrated into the C.K.E. By comm. and rec. we have $\exists\, k$, $m$ s.t. $\mathcal{P}_{ij}^k > 0$, $\mathcal{P}_{ji}^m > 0$. Then $\forall\, n > 0$ we have

$$\mathcal{P}_{jj}^{m+n+k} = \sum_{s,t} \mathcal{P}_{js}^m \mathcal{P}_{st}^n \mathcal{P}_{tj}^k \geq \mathcal{P}_{ji}^m \mathcal{P}_{ii}^n \mathcal{P}_{ij}^k$$

then we check the characterized recurrent definition

$$\sum_{N=0}^{\infty} \mathcal{P}_{jj}^N \geq \sum_{n=0}^{\infty} \mathcal{P}_{jj}^{m+n+k} \geq \sum_{n=0}^{\infty} \mathcal{P}_{ji}^m \mathcal{P}_{ii}^n \mathcal{P}_{ij}^k \geq \mathcal{P}_{ji}^m (\sum_{n=0}^{\infty} \mathcal{P}_{ii}^n) \mathcal{P}ij^k = \infty$$

the first inequity is because the LHS start from 0 while the RHS start actually from $m + k$. Then middle term of the the last term is infinity by recurrence of $i$. So we finally generate that the sum is infinity and thus $j$ is recurrent. □

So our consequence are

- Recurrence and transience are class properties (all states in the communicating class are all either transient or recurrent).
- The collory can be relaxed as "$i$ is recurrent and $i \to j$"
- If $i \to j$ but $j \nrightarrow i$, then $i$ is transient

2. **Model: Gambler's Ruin**

The scenario is: a gambler has \$$n$ initially (capital) and play a game with probability $\bar{p}$ of winning \$1 and $1-\bar{p}$ of losing \$1 otherwise. The gambler plays until he reach a goal of having \$$N \geq n$, and clearly his wealth follows a Markov Chain.

The transition diagram is one dimensional with $N+1$ states (n $\in [0, N]$). The corresponding transition matrix is

$$\underset{\sim}{\mathcal{P}} = \begin{pmatrix} 1 & 0 & 0 & ... & 0 & 0 & 0 \\ 1-p & 0 & p & ... & 0 & 0 & 0 \\ 0 & 1-p & 0 & p & ... & & 0 \\ 0 & 0 & 1-p & 0 & p & ... & 0 \\ ... & ... & ... & ... & ... & ... & ... \\ 0 & 0 & ... & ... & 0 & 0 & 1 \end{pmatrix}$$

Obviously there are three communicating classes: $\{0\}$, $\{N\}$ and $\{1, ..., N-1\}$, where the fist two are aperiodic and the last class is of period 2. Then a reasonable question is, let $R$ be the event that the "gambler goes broke", and asking for $\mathcal{P}(R \mid X_0 = n)$.

<u>**Model:**</u> Now let $\mathcal{P}(n) = \mathcal{P}(R \mid X_0 = n)$. Then the boundary conditions are $\mathcal{P}(0) = 1$ and $\mathcal{P}(N) = 0$. Then for $1 \leq n \leq N-1$ we have

$$\mathcal{P}(\mu) = \mathcal{P}(R \mid X_0 = \mu) = \mathcal{P}(R \mid X_1 = \mu+1)\mathcal{P}(X_1 = \mu+1 \mid X_0 = \mu)$$
$$+\mathcal{P}(R \mid X_1 = \mu-1)\mathcal{P}(X_1 = n-1 \mid X_0 = \mu)$$

This can be generated by conditioning on $R \cap X_0 = \mu+1$ and $\mu - 1$Then we have

$$\mathcal{P}(\mu) = \mathcal{P}(\mu+1) \cdot \bar{p} + \mathcal{P}(\mu-1) \cdot (1-\bar{p})$$

which is a 2nd order linear recurrence. Consider the Characteristic equation in $x$, we have

$$x = x^2 \cdot \bar{p} + 1 \cdot (1-\bar{p}) \implies \bar{p}x^2 - x + (1-\bar{p}) = 0$$

check $\Delta = (2\bar{p}-1)^2 \geq 0$. So two cases are

- $\Delta > 0 \equiv \bar{p} \neq 1/2$ : Two roots are $x_1 = 1$ and $x_2 = (1-\bar{p})/\bar{p} = a$ (a for short). Then the 'general solution' is
$$x = \mathcal{P}(n) = \alpha \cdot x_1^n + \beta \cdot x_2^n$$
where $\alpha$ & $\beta$ need to be determined by the boundary conditions. Plugging in the boundary condition
$$\alpha = \frac{a^N}{1-a^N},\ \beta = \frac{1}{1-a^N}$$
so the general solution is
$$\mathcal{P}(n) = \frac{a^N - a^n}{a^N - 1} = 1 - \frac{a^n - 1}{a^N - 1}$$

- $\Delta = 0 \equiv \bar{p} = 1/2$: Similar to case in differential equation, when encounter with repeated root, we assume the general solution to be
$$\mathcal{P}(n) = \alpha \cdot x^n + \beta \cdot nx^n$$
and using the boundary condition solving for coefficients. We get
$$\mathcal{P}(n) = 1 - \frac{n}{N}$$

So in summary, the solution is
$$\mathcal{P}(n) = \begin{cases} 1 - \frac{a^n - 1}{a^N - 1}, \text{if } \bar{p} \neq \frac{1}{2} \\ 1 - \frac{n}{N}, \text{if } \bar{p} = \frac{1}{2} \end{cases} \tag{1}$$

*Remark* 7. This type of problem is typically interpreted as a one dimensional random walk with absorbing boundaries.

3. **Model: Random walk in $\mathcal{Z}^1$**

We consider the (discrete) uniform random walk and questioning is the state recurrent or transient?

**Solution:** First we will need a tool called *Stirling's formula*, which is a approximation of $n!$. That is
$$n! \;\approx\; \sqrt{2\pi n}\,\left(\frac{n}{e}\right)^n$$

As we showed previously, the way to prove transience or recurrence is by showing the sum of all n-step transition probability is $\infty$ or not. Let's consider the 1 dimensional case starting form 0. Since the recurrence and transience are class properties, so it is valid to just show 0 is feasible.

Starting form 0. If we want to come back to 0, we must take $2n$ steps (since one dimensional) and easy to see $\mathcal{P}_{00}^{2n+1} = 0$. Thus
$$\mathcal{P}_{00}^{2n} = \left(\frac{1}{2}\right)^n \cdot \binom{2n}{n} = \frac{\binom{2n}{n}}{2^{2n}}$$

which means, we have to move $2n$ steps in total and want $n$ to be forward and $n$ to be backwards to get back to 0. The summation becomes
$$\sum_{n=0}^{\infty} \mathcal{P}_{00}^{n} = \sum_{n=0}^{\infty} \frac{\binom{2n}{n}}{2^{2n}} = \sum_{n=0}^{\infty} \frac{(2n)!}{(n!)^2 \cdot 2^{2n}}$$

apply the stirling's formula to replace the factorial, we get
$$\sum_{n=0}^{\infty} \frac{1}{\sqrt{\pi n}} \to \infty$$

we can compare to the series $1/n$ (we know $1/n$ is divergent and $1/\sqrt{n}$ is greater than it), so the series diverges and therefore the symmetric 1-dimension random walk is recurrent. Also we can show that the asymmetric 1-dimension random walk is transient.

4. **Model: Random walk in $\mathcal{Z}^d$**

By means of *Characteristic function*. Notations: we define $\overrightarrow{e}'_j s$ are the canonical basis of $\mathcal{Z}^d$. So we define the R.W at steps $\overrightarrow{X}_i$ with probability
$$\mathcal{P}(\overrightarrow{X}_i = \overrightarrow{e}_j) = \mathcal{P}(\overrightarrow{X}_i = -\overrightarrow{e}_j) = \frac{1}{2d}$$

通俗讲，在第 $i$ 步时，向 $\overrightarrow{e}_j$ 方向走的概率为 1/2d（2d 是因为有 d 个维度，则有 2d 个方向）。Then let $\overrightarrow{S}_0 = \overrightarrow{0}$ and $\overrightarrow{S}_n$ = position after n steps $= \sum_{i=1}^{n} \overrightarrow{X}_i$, then

$$\overrightarrow{S}_{n+1} = \overrightarrow{S}_n + \overrightarrow{X}_{n+1}$$

then we can define the transition probability as

$$\mathcal{P}(\overrightarrow{S}_{n+1} = \overrightarrow{S} \pm \overrightarrow{e}_j \,|\, \overrightarrow{S}_n = \overrightarrow{S}) = \frac{1}{2d}$$

**Definition 10.** For $\overrightarrow{k} \in \mathcal{R}^d = (k_1, ..., k_d)$, we can define $\phi(\overrightarrow{k})$, that is the characteristic function of $\overrightarrow{S}_n$ evaluated at $\overrightarrow{k}$,s.t

$$\phi_n(\overrightarrow{k}) = \mathbb{E}(e^{i\overrightarrow{k}\cdot\overrightarrow{S}_n})$$

*Remark* 8.

- The product of $\overrightarrow{k}$ and $\overrightarrow{S}_n$ is scalar product in $\mathcal{R}^d$ of $\overrightarrow{k}$ and $\overrightarrow{S}_n$. So

$$\overrightarrow{k} \cdot \overrightarrow{S}_n = \sum_{i=1}^{n} \overrightarrow{k} \cdot \overrightarrow{X}_i$$

- The Euler's formula can be applied here

$$e^{i\theta} = \cos\theta + i\sin\theta$$

**Theorem 1.** *The symmetric R.W in $\mathcal{Z}^d$ is recurrent if $d = 1, 2$ and is transient if $d \geq 3$.*

*Proof.* make up later Jan 22nd □

5. **Limiting probability**

The limiting probability is simply the proba. distribution of being at sate $i$ as $n$ goes to infinity which is

$$\pi_j = \lim_{n\to\infty} \mathcal{P}(X_n = j \,|\, X_0 = i)$$

this also accounts for the *proportion of time the M.C stays in each state in the long run.* Let $m_i$ be the *mean # of transaction returning to i, starting from i.* Then we have the proposition that the proportion of time that the M.C stay at i must equal to $1/m_i$ in the long run, so $\pi_i = 1/m_i$. Also notice, the limiting distribution depends does not depends on the initial state, so we can also write is as

$$\pi_j = \lim_{n\to\infty} \mathcal{P}(X_n = j), \ \forall$$

*Proof.* text book chapter 4 P216. incredible □

**Definition 11.** A state $i$ is said to be positive recurrent if $m_i < \infty$; it is said to be null recurrent if $m_i = \infty$

It is easy to check that the one dimension symmetric random walk is null recurrent by

$$m_i = \sum_n 2n \cdot P(T = 2n) = \sum_n \frac{1}{\sqrt{\pi n}} = \infty$$

Notice that: ① Positive and null recurrence are both class properties ② If the state space is finite, then all recurrent states are positive recurrent.

**Definition 12.** An aperiodic positive recurrent state is called *ergodic*. A ergodic M.C is a M.C whose states are all ergodic.

**Theorem 2.** *For an irreducible ergodic M.C, $\pi_j = \lim_{n\to\infty} \mathcal{P}_{ij}^n$ always exists for all $j$, independent of $i$.*

In addition: ① $\underline{\pi}$ is the unique solution to

$$\underline{\pi} \cdot \underset{\sim}{\mathcal{P}} = \underline{\pi} \quad \& \quad \sum_{j \in \mathcal{S}} \pi_j = 1$$

② $\pi_j = 1/m_j$, where $m_j$ is the mean return time to $j$ ($\pi_i > 0$). ③ $\pi_j$ by its definition can be written as

$$\pi_j = \lim_{n \to \infty} \frac{\#\text{visits to } j \text{ by total } n \text{ transition}}{n} = \text{long run proportion of time spent at } j$$

注意,对于①来说,如果想用不等式解出 distribution,那么需要 check 前提条件: irreducible and positive re--current. 其实对于大部分情况来说很好检查: irreducible 只需要看是否只有一个 communicating class (直接检查是否 transition matrix 没有 0 在其中); 对于 positive recurrent 来说, 往往遇到的情况都是 finite 的 state space, 根据之前的结论可知, recurrent state 必然是 positive recurrent。

**Definition 13.** (Probability for inverse M.C) Assume the M.C has been operating for a long time with stationary distribution $\pi$ the limiting probability?. Let $Y_n = X_{N-n}$, take $N$ as the time that M.C has already operated. Then we define the time inverse transition probability for $Y_n$ as

$$\mathcal{Q}_{ij} = \mathcal{P}(Y_n = j \,|\, Y_{n-1} = i) = \frac{\pi_j \mathcal{P}_{ji}}{\pi_i}$$

*Proof.* $Y_n$ is also a M.C Notes on Jan.31 □

由于 $Y_n$ 是从 $X$ 的序列 tracing back, 所以证明时注意站在 $X$ 的视角注意角标。Notice the existence of stationary distribution is necessary for the inverse M.C to be homogeneous.

**Definition 14.** We say the M.C is time reversible if

$$\mathcal{Q}_{ij} = \mathcal{P}_{ij}$$

*Remark* 9.

- We always have $P_{ii} = Q_{ii}$
- For a reversible M.C

$$\pi_i \mathcal{P}_{ij} = \pi_j \mathcal{P}_{ji}$$

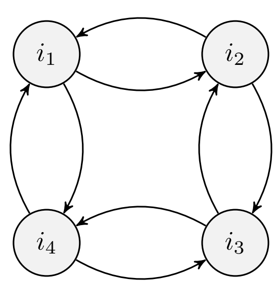
  which is interpreted as the *fraction of jumps at stationarity.* This forms a so-called *detailed balance.*

**Proposition 5.** *Let $X_n$ be a ergodic irreducible M.C. If we can find a series of $x_i \geq 0$, s.t $x_i \mathcal{P}_{ij} = x_j \mathcal{P}_{ji}$ and $\sum x_i = 1$, then $x_i = \pi_i$ and M.C is time reversible.*

**Proposition 6.** *(Kolmogorov's criteria): An ergodic MC is time reversible **iff** for $\forall$ finite sequence of states $(i_1, i_2, ..., i_n)$, we have*

$$\mathcal{P}_{i_1 i_2} \mathcal{P}_{i_2 i_3} ... \mathcal{P}_{i_n i_1} = \mathcal{P}_{i_1 i_n} \mathcal{P}_{i_n i_{n-1}} ... \mathcal{P}_{i_2 i_1}$$

*visualization is*



prove Feb 3rd starting from the stationary distribuion.

**Example: Ehrenfest Chain**

有两个罐子,一共 M 个球。每一步随机从其中一个罐子中拿一个球放到另一个罐子中。Let $\mathcal{S} = \{0, 1, ..., M\}$ and $X_n$ be the # of balls in urn 1. The probability is defined as

$$\mathcal{P}_{01} = 1 \quad \mathcal{P}_{M,M-1} = 1 \quad \mathcal{P}_{i,i+1} = \frac{M-i}{M} \quad \mathcal{P}_{i,i-1} = \frac{i}{M}$$

The chain is ergodic. The stationary distribution can be found by either solving the detailed balance equation or guess and check. The result turn out to be a binomial distribution of $\pi$

$$\pi \sim Bino(M, \frac{1}{2})$$

6. **Barnching Process**

**Definition 15.** (Branching Process) Let $Z_i$ be the # of individuals at generation $i$, $y_{n,i}$ be the # of offsprings of $i$ th individual at generation $n$. Then the *branching process* is defined as the sequence of r.v. $\{Z_n\}_{n\geq 0}$ such that

$$\begin{cases} Z_0 = 1 \\ Z_{n+1} = \sum_{i=1}^{Z_n} y_{n,i} \end{cases} \tag{2}$$

This is called the reproduction law.

- $Z_n$ defines a discrete time M.C on $\mathcal{N}$.
- $\{0\}$ defines a absorbing $\rightarrow$ recurrent class. We assume the probability $\mathcal{P}(y_{n,i} = 0) > 0$ and 0 is accessible form all other states, so all other states are transient.
- Since the finite set of transient states can only be visited finite times. So we can expect, in the long run, $Z_n$ is either $\infty$ or 0 as $n \rightarrow \infty$.

**Definition 16.** (Probability Generation Function) Let $X$ be a r.v on $\mathcal{N}$. Then the probability generating function of $X$ is

$$G_X(s) = \mathbb{E}_X(s^X) = \sum_{k=0}^{\infty} \mathcal{P}(X = k)s^k$$

- We have, in branching process, $X$ is a discrete r.v. so we sum up the PMF for expectation.
- The radius of convergence is either greater than or equal to 1, since when $s = 1$, $\sum \mathcal{P} = 1$, which means the radius of convergence is at least 1 (may use comparison test of series), so $G(s)$ is well defined $\forall\ s \in [-1, 1]$.
- By induction, similar to the moment generating function, we can have

$$G^{(k)}(0) = \mathcal{P}(x = k) \cdot k!$$

注意该结果不需要泰勒展开，直接求导且令 $s = 0$。(Notice k is the $k$th derivative).
- Similar to the moment generating function, the probability distribution is uniquely determined by the PGF, or equivalently, two r.v that have same PGF follows same prob. distribution.
- Properties: If $X$ and $Y$ are two independent r.v on $\mathcal{N}$, then $\forall s \in [-1, 1]$

$$G_{X+Y}(s) = G_X(s) \cdot G_Y(s)$$

- If we set $s = 1$, we can work out several properties

$$G(1) = \mathbb{E}(1^X) = 1 \quad G'(1) = \mathbb{E}(X)$$

recall $Var(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$, then

$$G''(1) = \mathbb{E}(X^2) - \mathbb{E}(X) = Var(X) + (\mathbb{E}(X))^2 - \mathbb{E}(X)$$
$$\implies Var(X) = G''(1) + G'(1) - (G'(1))^2$$

- The PGF is non-decreasing and convex on $s \in [0, 1]$. Other? It is easy to check that the first and second derivative are all grater than 0.

**Proposition 7.** *Let $X_1, X_2, ..., X_N$ be iid r,v with PGF $G_X(s)$. Let $N$ be a r.v independent of $X$. Let $T = \sum_{i=1}^{N} X_i$. Then we have*

$$G_T = G_N(G_X(s))$$

*Proof.*

$$\begin{aligned} G_T(s) &= \mathbb{E}(s^T) = \mathbb{E}_N(\mathbb{E}(s^T \mid N)) \\ &= \sum_{n=0}^{\infty} \mathbb{E}(s^T \mid N = n) \cdot \mathcal{P}(N = n) \\ &= \sum_{n=0}^{\infty} \mathbb{E}(s^{\sum X_i} \mid N = n) \cdot \mathcal{P}(N = n) \\ &= \sum_{n=0}^{\infty} (G_X(s))^n \cdot \mathcal{P}(N = n) \\ &= \mathbb{E}_N(G_X(s)^N) = G_N(G_X(s)) \end{aligned}$$

□

Then we apply this result to the Branching process. Since $Z_{n+1} = \sum_{i=1}^{Z_n} y_{n,i}$, then

$$G_{n+1}(s) = \underbrace{G_{Z_{n+1}}(s) = G_{Z_n}(G_y(s))}_{\text{by above result}}$$

$$= G_n(G_y(s)) = G_{n-1}(G_y(G_y(s))) = ... = \underbrace{G_y \circ G_y \circ ... \circ G_y(s)}_{\text{n+1 times composition}}$$

then we concludes that

$$G_n(s) = G_1^{(n)}(s)$$

where RHS means the n-th iterate of the generating function. Notice that since the 0th generation is defined to be 1 in the branching process, so the distribution of $Z_1$ is the same as $Y$ Is this correct?

## Survival Probability (prob. of extinction)

Let $\mathcal{P}_e = \mathcal{P}(Extinction)$ (survival probability = 1 - $\mathcal{P}_e$). The extinction event becomes $\bigcup_{n=0}^{\infty}\{Z_n = 0\}$ and

$$\{Z_n = 0\} \subset \{Z_{n+1} = 0\} \subset \{Z_{n+1} = 0\}......$$

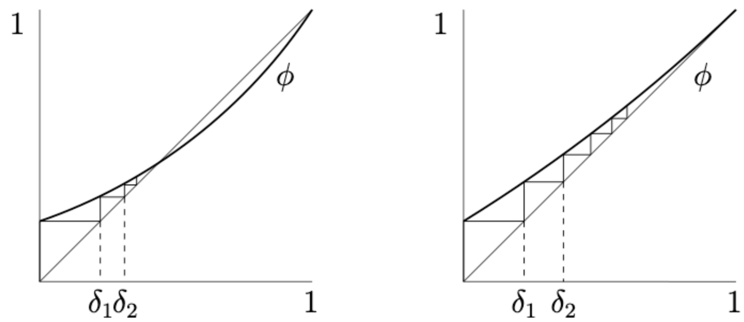
so the extinction event finally become

$$\lim_{n\to\infty} Z_n = 0$$

and consequently the probability of extinction becomes a limiting probability

$$\mathcal{P}_e = \lim_{n\to\infty} \mathcal{P}(Z_n = 0) = \lim_{n\to\infty} G_{Z_n}(0) = \lim_{n\to\infty} G_n(0) = \lim_{n\to\infty} G_1^{(n)}(0)$$

Notice the last $(n)$ means $n$ times composition. The above result shows that we can get to the extinction probability via studying $G_1$, where $G_1$ is the generating function of the reproduction law. Then we can visualize the $G_n(0)$ from the graph of $G_y$ Two graph needed here and question about the graph




Assume $p_0 = G_1(0) > 0$ and by convexity

- If $G'(1) = \mathbb{E}(Y) > 1$, $\mathcal{P}_e = \lim_{n\to\infty} G_n(0) < 1$
- If $G'(1) = \mathbb{E}(Y) < 1$, $\mathcal{P}_e = 1$
- If $G'(1) = \mathbb{E}(Y) = 1$ : If $G''(1) > 0$, same graph as for $G'(1) < 1$; If $G''(1) = 0$, $Var(y) = 0$, then the process is deterministic and $Z_n = 1\ \forall n$. 啥意思??

**Theorem 3.** *Let $\mu = \mathbb{E}(Y)$ and $\sigma^2 = Var(Y)$ where $Y$ is the reproduction law. Then $\eta$, the probability of extinction is the smallest non-negative root of $s = G_Y(s)$ and*

$$\eta = 1\,,\ \text{if } \mu < 1 \text{ or } [\mu = 1 \text{ and } \sigma^2 > 0]$$

$$\eta < 1\,,\ \text{if } \mu > 1$$

*Remark* 10. 该定理符合猜测：每一个 individual 都要生产出足够多的后代（$\mu > 1$）以保证 population 延续（renew），即使得不一定灭绝。

对于 discrete M.C 的介绍到此为止。之后自学 MCMC 和 HMM。

## 7. Exponential Distribution and Poisson Process

**Definition 17.** (Exponential distribution) $\mathcal{X} \sim Exp(\lambda)$ The pdf of a expo. distributed r.v is

$$f(x) = \lambda e^{-\lambda x} \implies \mathbb{E}(X) = \frac{1}{\lambda},\ Var(X) = \frac{1}{\lambda^2}$$

$$F(X) = \mathcal{P}(X \le x) = \int_0^x \lambda e^{-\lambda x} dx = 1 - e^{-\lambda x}$$

for $x$ defined only on $[0, \infty)$. Its MGF is

$$\phi(t) = \frac{\lambda}{\lambda - t}$$

for $t < \lambda$ and infinity for $t \ge \lambda$, since the integral goes to infinity.

**Proposition 8.** *(Memoryless property) In short*

$$\mathcal{P}(X > s + t \,|\, X > s) = \mathcal{P}(X > t)$$

具体证明解释查看 *302 note*。该性质说明已知目前正在进行，这对下一个时刻是否还会进行没有任何影响。

一个运算的例子

$$\mathcal{P}(X > 5 \,|\, X > 2) = \mathcal{P}(X > 3) = e^{-3\lambda}$$

*Remark* 11. We can associate the exponential distribution as a measure of lifetime, from which we can qualify a *failure rate* as follows.

**Set up the failure rate**

Let $X$ represent the life time of some object (physical life, battery life etc.). Then the probability that the object can still be alive for another $dt$ time given that it has already lived for t is

$$\mathcal{P}(X \in [t, t+dt] \,|\, X > t) = \frac{\mathcal{P}(X \in [t, t+dt])}{\mathcal{P}(X > t)}$$

$$= \frac{\int_t^{t+dt} f(u)du}{1 - F(t)} \xrightarrow{dt \to 0} \frac{f(t)\,dt}{1 - F(t)} = r(t)\,dt$$

where we define the *failure rate function (or hazard rate function )* as

$$r(t) = \frac{f(t)}{1 - F(t)}$$

in this context where $X$ is exponentially distributed $r(t) = \lambda$. So the failure rate function for an exponential r,v is constant which does not depend on time. This is another version of the non-memory property. We can prove that the exponential r.v is the only real continuous memoryless r.v. Feb.14th note. For discrete cases, the geometric distribution also has memoryless property.

**Min. of two exponential r.v**

Let $X_1 \sim Expo(\lambda_1)$ and $X_2 \sim Expo(\lambda_2)$. We define a new r,v $X$ which is $X = Min.(X_1, X_2)$. We are interested in the prob. distribution of $X$ so it is neat here to find the prob. of $X > x$, this is

$$\mathcal{P}(X > x) = 1 - F_X(x) = \mathcal{P}(X_1 > x, X_2 > x)$$

$$\underbrace{=}_{independent} \mathcal{P}(X_1 > x)\mathcal{P}(X_2 > x) = e^{-\lambda_1 x} e^{-\lambda_2 x}$$

$$= e^{-(\lambda_1 + \lambda_2)x}$$

This means, the new r.v $X$ is also a exponetially distributed r.v with $\lambda = \lambda_1 + \lambda_2$. Then we can generalize this to the case of $n$ expo. r.v which is

**Proposition 9.** *Let $X_1, X_2, ..., X_n$ are $n$ independently r.v that follows $X_i \sim Exp(\lambda_i)$, $i \in [1, n]$. Then*

$$X = \min(X_1, X_2, ..., X_n) \sim Exp(\sum_{i=1}^{n} \lambda_i)$$

例: 现在有三个顾客，其中两个正在前台被服务。假设每个顾客的服务时间为随即变量且均服从指数分布 $Exp(\lambda)$。问：预计三个顾客都完成服务离开的时间是多长？

解: Time until the 1st customer leave is a r.v $X = min(X_1, X_2) \sim Exp(2\lambda)$ and $\mathbb{E}(X) = 1/2\lambda$. The clock then restart with the second customer fill the position and again the r.v is $X$. Finally the sole customer left and the time as a r.v following $X' \sim Exp(\lambda)$. So the expected time is $1/2\lambda + 1/2\lambda + 1/\lambda$.

**Proposition 10.** *If $X_1$ and $X_2$ are independent and follows $X_i \sim Exp(\lambda_i)$. Then*

$$\mathcal{P}(X_1 < X_2) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

*Proof.*

$$\mathcal{P}(X_1 < X_2) = \int\int_{x_1 < x_2} f_{X_1}(x_1) f_{X_2}(x_2) dx_1 dx_2$$

$$
\begin{aligned}
&= \int_0^\infty \lambda_2 e^{-\lambda x}\, dx_2 \int_0^{x_2} \underbrace{\lambda_1 e^{-\lambda_1 x} dx_1}_{=\mathcal{P}(X_1 \le x_2) = 1 - e^{-\lambda_1 x_2}} \\
&= \int_0^\infty \lambda_2 e^{-\lambda_2 x_2}\, dx_2 - \int_0^\infty \lambda_2 e^{-(\lambda_1+\lambda_2)x_2} dx_2 \\
&= 1 - \frac{\lambda_2}{\lambda_1+\lambda_2} = \frac{\lambda_1}{\lambda_1+\lambda_2}
\end{aligned}
$$

□

**Proposition 11.** *Let $\mathcal{Z} = \sum_{i=1}^n X_i$ where $X$ and $Y$ are expo. distributed r.v with parameter $\lambda_i$. Then the probability density function is*

$$
f_{\mathcal{Z}} = \sum_{i=1}^n [\prod_{j \ne i} \frac{\lambda_j}{\lambda_j - \lambda_i}] \cdot \lambda_i \cdot e^{-\lambda_i t}
$$

*then if we know all $X_i$s are i.i.d with same $\lambda$, then*

$$
\mathcal{Z} = \sum_{i=1}^n X_i \sim \Gamma(n, \lambda)
$$

*with density function*

$$
f(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \quad \mathbb{E}(\mathcal{Z}) = \frac{n}{\lambda} \quad Var(Z) = \frac{n}{\lambda^2}
$$

Graph here needed to show the changes of parameter.

8. **The Poisson Process**

Two ways to describe the process:

- To generate the process via exponentially distributed r.v. as "arrival time"
- Emphasizes on the properties of the process as counting process (i.e a stochastic process $\mathcal{N}(t)$ that represents the total number of events by time t).

无论是哪一种解释的方式，$\lambda$ 所谓的' rate '是指单位的时间段内，有多少次事件发生。如果所给的信息是关于 $t$ 时长的 rate，而相关的问题是 $t$ 时长的倍数的时间发生概率，则也要相应的扩大 rate 的倍数。

(a) **The first description**

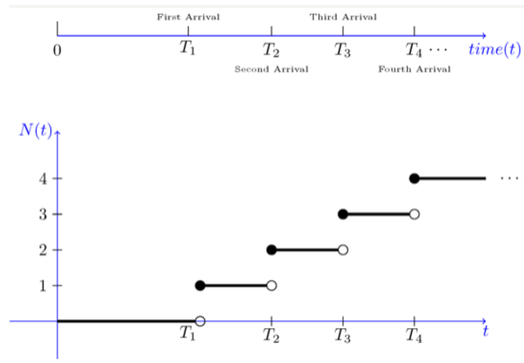
A poisson process with rate $\lambda$ is a counting process where the time between 2 event is $Exp(\lambda)$. The time until $n$ event happens is

$$
T = \sum_{i=1}^n X_i, \ \ X_i \sim Exp(\lambda)
$$

where all $X_i$ are *i.i.d.* The $\lambda$ here represent the rate of job completion (i.e the average job taks $1/\lambda$ time unit /job, so rate fo job completion is $1/1/\lambda$ Why?)

Let the counting process to be $N$ and

$$
N(t) = \#\text{ of jobs completed by time t}, t \ge 0
$$




We can conclude by intuition that

$$\mathbb{E}(N(t)) = \lambda t$$

**PMF of N(t)**

For $n \geq 1$, consider $\{N(t) \geq n\} \equiv \{\sum_{i=1}^{n} X_i \leq t\}$. Both means $n$ jobs are completed by $t$. We have already derived the sum of exp. r.v follows Gamma distribution. Thus

$$\mathcal{P}(N(t) \geq n) = \mathcal{P}(\sum_{i=1}^{n} X_i \leq t)$$

Proof made up here.

Finally we get

$$\mathcal{P}(N(t) = m) = \frac{(\lambda t)^m e^{-\lambda t}}{m!}, m \geq 0$$

which is the poisson distribution. So we conclude

**Theorem 4.** *If N(t) is a poisson process then*

$$N(t) \sim Poisson(\lambda t)$$

*with mean $\lambda t$ and variance $\lambda t$. This matches our anticipation.*

**Example**

- $\mathbb{E}(S_4 \mid N(1) = 2) = 1 + \mathbb{E}(S_2) = 1 + 2/\lambda$
- $\mathbb{E}(N(4) - N(2) \mid N(1) = 3) = \mathbb{E}(N(4) - N(2)) = \mathbb{E}(N(4-2)) = \mathbb{E}(N(2)) = 2\lambda$

对于以上问题可以直接画一个 t 的实轴分析。严谨证明在 latex 这个文件夹下的 pdf 里面

**Proposition 12.**
- *Independent Increments: If $(s_1, s_2)$ and $(t_1, t_2)$ are disjoint time interval, then $N(t_2) - N(t_1)$ and $N(s_2) - N(s_1)$ are independent r.vs. (This can be generalized to more than 2 intervals)*
- *Stationary Increments: The difference between $N(s)$ and $N(s+t)$ is independent of $s$, in other words, same as $N(t)$.*

(b) **Second discrption:**

**Definition 18.** (Poisson process) A counting process $\{N(t), t \geq 0\}$ is a poisson process of rate $\lambda$, if it satisfies the following axioms:

- N(0) = 0
- It has independent increment
- $\mathcal{P}(N(t+h) - N(t) = 1) = \lambda h + o(h)$
- $\mathcal{P}(N(t+h) - N(t) \geq 2) = o(h)$, or equivalently, $\mathcal{P}(N(t+h) - N(t) = 0) = 1 - \lambda h + o(h)$

where we define the $o(h)$ as the function $f$ which satisfy

$$\lim_{h\to 0}\frac{f(h)}{h}=0$$

what is the function $o(h)$ means?

The last axioms when we prove it, we can not treat $N(t)$ as a r.v following poisson distribution since that is what we are trying to show. We can only use theses axioms to prove it.

*Remark* 12. A consequence of this 2nd definition is that

$$N(s+t)-N(s)\sim Poisson(\lambda t)$$

This result is useful to show the equivalent of two definition of poisson process. proof is in note March 2.

**Theorem 5.** *We conclude from the above prove that two definition are equivalent that*

$$\{N(t),t\geq 0\}\ \textit{is a Poisson process of rate}\ \lambda\ \textit{(by def. 2)}$$

$$\Longleftrightarrow$$

$$N(t)\textit{is a counting process with iid inter arrival times}\sim Exp(\lambda)\textit{(by def. 1)}$$

Proof needed from March 2

**Theorem 6.** *(Superposition) Let $\{N_1(t),t\geq 0\}$ and $\{N_2(t),t\geq 0\}$ be two independent Poisson process, with rates $\lambda_1$ and $\lambda_2$. Let $N(t)=N_1(t)+N_2(t)$. Then $N(t)$ is a poisson process of rate $\lambda_1+\lambda_2$.*

Proof needed from note March 4.

**Proposition 13.** *(Poisson Thining) Let $\{N(t),t\geq 0\}$ be a poisson process of rate $\lambda$ and that its events are independently labelled.: events are of either type 1 with prob. $p$ or type 2 with porb. $1-p$. Let $N_1(t)=\#$ type 1 event by time $t$ and $N_2(t)$ similarly. Then $N_1$ and $N_2$ are independent poisson process of rate $\lambda p$ and $\lambda(1-p)$ respectively.*

Proof as exercises, by showing the 5 axioms.

9. **Conditional distribution of arrival time**

We are trying to find the joint distribution of first $n$ arrival time for the first $n$ event, conditioning on $\mathcal{N}(t)=n$. Let $S_1,S_2,...,S_n$ *i.i.d*, which are in **ascending "order"** be the r.v. for arrival time. (Do not get confused with order stat. here $s$'s are not ordered statistics) Then

$$f_{S_1,S_2,...,S_n}(s_1,s_2,...,s_n\,|\,\mathcal{N}(t)=n)$$
$$=\frac{\int_t^\infty \lambda e^{-\lambda s_1}\lambda e^{-\lambda(s_2-s_1)}...\lambda e^{-\lambda(s_{n+1}-s_n)}\,ds_{n+1}}{(\lambda t)^n e^{-\lambda t}/n!}$$

to achieve this equality we need to convert the conditional event to be

$$\frac{\mathcal{P}(S_1<s_1,S_2<s_2,...,S_n<s_n\ ,\ s_n<t<S_{n+1})}{\mathcal{P}(N(t))=n}$$

where the denominator is exactly the poisson distribution r.v.. For the numerator, since we have fixed the value of the $n$ events as a "known" value and we want to constrain the $(n+1)$th event to happen after $t$. So this can be treated as an integral w.r.t $s_{n+1}$. After simplification we get

$$=\frac{n!}{t^n}$$

This is the CDF or the PDF? should be PDF right? Then with some proof and knowledge of ordered statistics we can conclude:

**Theorem 7.** *The $n$ i.i.d exponential r.v. $S_1, S_2, ..., S_n$ conditioning on its corresponding poisson process (appealing to the first interpretation of poisson process), have the same distribution of $n$ i.i.d Uniform r.v distributed on $[0, t]$. In other words, the $n$ poisson events are uniformly distributed on time interval $[0, t]$.*

A proof of the ordered stat is needed. However refer to the files in Math303 to set up a new doc. for order stat is necessary and important.

值得注意的是，只有在给定的条件下的 $n$ 个泊松事件才是均匀分布的（在给出证明后会更加明显）。

**Proposition 14.** *Let $\mathcal{N}_i(t) = \#$ of type $i$ events happed by time $t$. Then $\{\mathcal{N}_i\}_{i=1}^k$ are independent poisson r.v with $\mathbb{E}(\mathcal{N}_i(t)) = \lambda \int_0^t \mathcal{P}_i(s)ds$ where $\mathcal{P}_i(s)$ is defined as the probability distribution (cdf) of being type $i$ at time $s$.* *Proof Mar 9th*

容易理解为什么是 CDF 而非 PDF。诚然时间是连续的，但是概率并非是关于时间的概率，在每个时刻，我们都可以讨论该事件是何种类的概率。

对泊松过程的讨论到此为止。继续可以学习 non-homon. poisson process. 对于 inter-arrival time 不是指数分布的情况，继续学习 renewal process 更新过程。

10. **Continuous Time Marcov Chain (CTMC)**

**Definition 19.** Let $\{X(t)\}_{t\geq 0}$ be a series of r.v, which all taking values in discrete state space. We say $\{X(t)\}_{t\geq 0}$ is a continuous-time M.C if

$$\mathcal{P}\big(X(s+t) = j \,|\, X(s=i), X(u) = x(u)\big) = \mathcal{P}(X(s+t) = j \,|\, X(s) = i)$$

where $u \in (0, s)$, $\forall s, t \geq 0$ and $\forall$ state $i, j$.

*Remark* 13. In CTMC we always assume *stationarity* and *time homogeneous*:

- Stationarity: $\mathcal{P}(X(s+t) = j \,|\, X(s) = i)$ for all time $s$. This generate the time-homo property.
- Time-Homo.: $\mathcal{P}(X(s+t) = j \,|\, X(s) = i) = \mathcal{P}(X(t) = j \,|\, X(0) = i)$, special case emphasizes on 0.

注意：此处的性质表明，转移概率与从哪一个 state 起始无关，与从哪一个时刻起始无关，但是与间隔时长有关。因此，在 CTMC 中，某种程度上我们可以将转移概率表示为关于间隔时长的函数，即

$$\mathcal{P}_{ij} = \mathcal{P}_{ij}(t) = \mathcal{P}r(X(s+t) = j \,|\, X(s) = i), \forall\, s \geq 0,\, i, j \in \mathcal{S}$$

**Proposition 15.** *(Time memoryless of CTMC) Let CTMC is in state $i$ at time $a$ and let $T_i$ be the additional time (comparing to $a$) until CTMC leaves $i$, then*

$$T_i \sim Exp(v_i)$$

*where $v_i$ is different from the starting state of MC at time $a$.* *Proof Mar 13th*

注意，逻辑是由于随机过程服从马尔科夫性质，所以必然等待时间是指数分布。反之亦然？

*Remark* 14. $T_i$ must also be independent of the next state the MC jumps to, otherwise tye MC property would be violated, because the past waiting in a state would entirely affects the next outcome. ?

**Definition 20.** A CTMC can be fully determined by two component:

- $T_i \sim Exp(v_i)$, which characterize the waiting time at state $i$
- $\mathcal{P}_{ij}$, the transition probability form $i$ to $j$, which must satisfy $\mathcal{P}_{ii} = 0$ and $\sum_j \mathcal{P}_{ij} = 1$.

Embeded chain?

11. **Example: Birth and Death process**

(a) **Model set up**

Let $\{X(t)\}_{t\geq 0} \in \mathcal{S} =: N^{+}$ be series of r.v that represent the population size. The population can only be affected by death or birth where both can be modeled by exponential distribution 个人理解，此处是强行让两个 r.v 服从指数分布，当然这也是合理的假设。注意由于 MC 性质我们需要的是 total waiting time 服从指数分布. Let $\lambda_n$ be the birth rate of occurrence when population size is $n$ and $\mu_n$ be the death rate of occurrence. Given $(\lambda_0, \lambda_1, ...), (\mu_0 = 0, \mu_1, \mu_2, ...)$. Then define this CTMC with

- $v_0 = \lambda_0$, $v_n = \lambda_n + \mu_n$

This is by time change in population $= \min(Exp(\lambda_n), Exp(\mu_n)) \sim Exp(\lambda_n + \mu_n)$;

- $\mathcal{P}_{01} = 1$, $\mathcal{P}_{n,n+1}(\lambda_n/(\lambda_n + \mu_n))$ and $\mathcal{P}_{n,n-1}(\mu_n/(\lambda_n + \mu_n))$

this is by

$$\mathcal{P}(\text{Bith time} < \text{Death time}) = \frac{\lambda_n}{\lambda_n + \mu_n}$$

same for death one. 注意，transition probability 不都是由时间决定的，在此处模型里恰好可以而已。在下面的 general case 会有讨论。

(b) **Local Behavior of Birth & Death process**

Recall the second interpretation of poisson process, which, in summary, is saying that in a give period $t$ and occurrence rate, the counts can only increase by 1, decrease by 1 or keep status quo.

$X(t+h)$:
- $\mathcal{P}(X(t) - 1) = \mu_n h + o(h)$
- $\mathcal{P}(X(t) + 1) = \lambda_n h + o(h)$
- $\mathcal{P}(X(t)) = 1 - (\lambda_n + \mu_n)h + o(h)$

and the probability of increasing and decreasing by 2 or more is $o(h)$. Since the waiting time is exponentially distributed, so the occurrence of either birth or death can be modeled by poisson process with the probability above. proof needed march 16.

Once we set up this, we can model the process with different type of birth and death rate. For example, the population size is in linear growth with immigration, we can have

$$\lambda_n = n\lambda + \theta,\ n \geq 1 \quad \mu_n = n\mu,\ n \geq 0$$

so we have

$$X(t+h) = \begin{cases} X(t) - 1 & w.p & X(t)\mu h + o(h) \\ X(t) + 1 & w.p & (X(t)\lambda + \theta)h + o(h) \\ X(t) & w.p & 1 - \big(X(t)(\lambda + \mu) + \theta\big)h + o(h) \end{cases}$$

(c) **Find M(t)** $= \mathbb{E}[X(t)]$

基本思路是，由上述讨论的搭配的结论找的一个合适的 ODE，通过解 ODE 的到 $M(t)$ 的解。这可以理解因为上述结论中牵涉到非常小的 $h$，我们可以想到通过变形将其于导数形式相联系。推导过程如下:

$$M(t+h) = \mathbb{E}[X(t+h)] = \mathbb{E}_{X(t)}[\mathbb{E}[X(t+h) \mid P(X(t))]]$$
$$= \mathbb{E}\Big[\sum \mathcal{P}(X(t+h) = i \mid X(t)) \cdot i\Big]$$

注意，(b) 中的结论由证明可知，其已经是条件概率。通过计算可得

$$= \mathbb{E}[X(t) + (X(t)(\lambda - \mu) + \theta)h + o(h)]$$

$$\Longrightarrow \frac{M(t+h)-M(t)}{h} = M(t)(\lambda-\mu)+\theta+\underbrace{o(h)}_{\text{goes to 0 as h}\to 0}$$

then we get a ODE

$$M'(t) = M(t)(\lambda-\mu)+\theta$$

using integrating factor we find

$$\frac{1}{\lambda-\mu}\ln[(\lambda-\mu)M(t)+\theta] = t+D$$

where D depends on the initial conditions. Here we assume $M(0)=i$ and finally we get

$$M(t) = ie^{(\lambda-\mu)t}+\frac{\theta}{\lambda-\mu}(e^{(\lambda-\mu)t}-1)$$

*Remark* 15. When $\lambda > \mu$ (i.e single birth rate > single death rate) then $M(t)$ grows exponentially. If $\lambda < \mu$, then $\lim_{t\to\infty} M(t) = \frac{\theta}{\mu-\lambda}$ and if $\lambda = \mu$, then $M(t) = \theta t + i$.

(d) **More general case for determining the transition probability**

上述例子是已经给定了出生和死亡发生的 rate 从而可以直接定义 transition probability。现在讨论更 general 的情况。

Let use the same notations as discussion above.

**Proposition 16.** *Let $q_{ij} = v_i \cdot p_{ij}$. If we know $q_{ij}$, then we know both $v_i$ and $p_{ij}$.*

Where $q_{ij}$ is called the *instantaneous transition rates*. It means: Since $v_i$ is the rate that the transition happens when the CTMC is in state $i$, and $\mathcal{P}_{ij}$ is the probability of transiting to $j$ form $i$, so $q_{ij}$ is thought to be the rate that the CTMC transit to $j$ when it is in state $i$.

*Proof.*

$$v_i = v_i \cdot \sum_{i\neq j} p_{ij} = \sum_{i\neq j} v_i p_{ij} = \sum_{i\neq j} q_{ij}$$

then

$$p_{ij} = \frac{q_{ij}}{v_i} = \frac{q_{ij}}{\sum_{i\neq j} q_{ij}}$$

□

*Remark* 16. We will always assume, when state space is finite, the $v_i$'s are small enough, so that the probability a time interval $[a, b]$ contains $\infty$ many transitions is 0. Otherwise, the process is called to be *explosive*, and may not be defined for all time.

**Lemma 1.** *As $h \to 0$, $\mathcal{P}_{ii}(h) = 1 - v_i h + o(h)$, which is equivalent to*

$$\lim_{h\to 0}\frac{\mathcal{P}_{ii}(h)-\mathcal{P}_{ii}(0)}{h} = \lim_{h\to 0}\frac{\mathcal{P}_{ii}(h)-1}{h} = -v_i$$

*and therefore*

$$\frac{d}{dh}\mathcal{P}_{ii}(h)\bigg|_{h=0} = -v_i$$

**Lemma 2.** *For $i \neq j$, $\mathcal{P}_{ij}(h) = q_{ij}h + o(h)$*

$$\lim_{h\to 0}\frac{\mathcal{P}_{ij}(h)-\mathcal{P}_{ij}(0)}{h} = \lim_{h\to 0}\frac{\mathcal{P}_{ij}(h)}{h} = q_{ij}$$

*and therefore*

$$\frac{d}{dh}\mathcal{P}_{ij}(h)\bigg|_{h=0} = q_{ij}$$

proof Wed 19 Fri 20

These lemma leads to the *Kolmogorov Backward equation.*

**Theorem 8.** *(Kolmogorov Backward equation)*

$$\mathcal{P}'_{ij}(t) = \sum_{k \neq i} q_{ik} \mathcal{P}_{kj}(t) - v_i \mathcal{P}_{ij}(t)$$

**Theorem 9.** *(Kolmogorov forward equation)*

$$\mathcal{P}'_{ij}(t) = \sum_{k \neq j} \mathcal{P}_{ik}(t) q_{kj} - v_j \mathcal{P}_{ij}(t)$$

The transition probability in the above equation $\mathcal{P}_{kj}$ and $\mathcal{P}_{ik}$ indicate what forward and backward means.

*Remark* 17. The Kolmo. equations define a set of linear ODE's that in theory are solved by $e^{Qt}$, the matrix exponential. This can only be done explicitly in a few cases but the Kom. equations are also useful in studying limiting behaviours and stationary distribution.

(e) **Limiting probability in CTMC**

Same as the discrete case, we are looking for the long-term behavior of the CTMC. The limiting probability is defined, if exits, as

$$\mathcal{P}_j = \lim_{t \to \infty} \mathcal{P}_{ij}(t), \ \forall i \in \mathcal{S}$$

which means independent of the initial state.

*Remark* 18. If $\lim_{t\to\infty} \mathcal{P}_{ij}(t)$ exists, then $\lim_{t\to\infty} \mathcal{P}'_{ij}(t) = 0$.
If we apply the kolmo. backward equation, it indicates that this is always true, so it does not provide any additional information. If we apply the kolmo. forward equation, it suggests

$$\text{①} v_j \mathcal{P}_j = \sum_{k \neq j} \mathcal{P}_k q_{kj} \quad \text{②} \sum_j \mathcal{P}_j = 1$$

Equation ①: LHS is $v_j \mathcal{P}_j$, where $v_i$ is the rate leaving state $j$ and $\mathcal{P}_j$ is the probability of being in state $j$, so the product means the overall rate at which the CTMC *leaves* state j in the long run. RHS, $q_{kj} = v_k p_{kj}$ is the rate at which the CTMC enter $j$ form $k$ and $\mathcal{P}_k$ is the probability of being in state $k$ in the long run, so sum up all the cases of $k$, it means the overall rate at which CTMC *enters* state $j$ in the long run. So it is

$$\underbrace{v_j \mathcal{P}_j}_{\text{rate out j}} = \underbrace{\sum_{k \neq j} \mathcal{P}_k q_{kj}}_{\text{rate in j}}$$

So in summary, in the long run the rate of joining state $j$ is the same as the rate leaving state $j$, so ① + ② is a *balanced equation* (but not detailed balanced, coming later).

The two equation above can also be expressed in the following way. Let $\pi = (\mathcal{P}_0, \mathcal{P}_1, ...)$ the limiting prob. distribution and $\mathcal{Q}$ to be the *intensity matrix*, which is

$$\mathcal{Q} = \begin{pmatrix} -v_1 & q_{21} & q_{31} & \dots \\ q_{12} & -v_2 & q_{32} & \dots \\ q_{13} & q_{23} & -v_3 & \dots \\ \dots & \dots & \dots & -v_4 \end{pmatrix}$$

then ① $\equiv \pi\mathcal{Q} = 0$ (i.e more to one side the summation). Using these new stuff defined, the two komlo. equations can be rewritten to be

$$\mathcal{P}'_{ij}(t) = (\mathcal{Q}\mathcal{P}(t))_{ij} \quad \textit{(backward)}$$

$$\mathcal{P}'_{ij}(t) = (\mathcal{P}(t)\mathcal{Q})_{ij} \quad \textit{(forward)}$$

where $\mathcal{P}(t)$ is defined as the matrix

$$\mathcal{P}(t) = \begin{pmatrix} \mathcal{P}_{00}(t) & \mathcal{P}_{01}(t) & \mathcal{P}_{02}(t) & \dots \\ \mathcal{P}_{10}(t) & \mathcal{P}_{11}(t) & \mathcal{P}_{12}(t) & \dots \\ \mathcal{P}_{20}(t) & \mathcal{P}_{21}(t) & \mathcal{P}_{22}(t) & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix}$$

(f) **Stationary Distribution**

**Definition 21.** The distribution $\pi$ is said to be stationary in CTMC if

$$\mathcal{P}(X(t) = y \,|\, X(0) \sim \pi) = \pi_y$$

where

$$\mathcal{P}(X(t) = y \,|\, X(0) \sim \pi) = \sum_x \mathcal{P}(X(0) = x \,|\, X(0) \sim \pi) \cdot \mathcal{P}(X(t) = y \,|\, X(0) = x)$$

$$= \sum_x \pi_x \mathcal{P}_{xy}(t) = \pi_y$$

The first line means, at $t = 0$ the distribution is $\pi$ and after $t$ the distribution is still $\pi$. The second line can be thought of as a generalization of the total probability law in conditional probability. Detail is as follows. The total probability law is

$$\mathcal{P}(A) = \sum_i \mathcal{P}(A \,|\, B_i)\mathcal{P}(B_i), \ \bigcup_i B_i = \Omega$$

then for the conditional probability we can have

$$\mathcal{P}(A \,|\, B) = \sum_i \mathcal{P}(A \,|\, B_i)\mathcal{P}(B_i \,|\, B), \ \bigcup_i B_i = B$$

其实所谓条件只是限定了一个新的概率空间（或是事件域），进而将新的概率空间进行拆分。在拆分之后欲使得每一个子条件获得取值则必须使其发生，即要乘以 $\mathcal{P}(B_i \,|\, B)$。

**Theorem 10.** *The distribution $\pi$ is stationary if and only if $\pi\mathcal{Q} = 0$.*

Proof Lecture 31

**Definition 22.** For a CTMC $\{X(t)\}_{t\geq 0}$, two states $x, y \in \mathcal{S}$ is communicate ($x \leftrightarrow y$) if $\mathcal{P}_{xy}(s) > 0$ and $\mathcal{P}_{yx}(s) > 0$ for some $s, t \geq 0$. The CTMC is said to be irreducible if there is only one communicating class.

*Remark* 19. If $x \leftrightarrow y$, then $\mathcal{P}_{xy}(t) > 0 \, \forall t > 0$.

**Theorem 11.** *Let $\{X(t)\}_{t\geq 0}$ be an irreducible CTMC, then*

- *There is no stationary distribution (i.e $\pi Q = 0$ does not have a valid solution) and $\lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{1}_{X(s)=y} ds = 1$ almost surely goes to one for $\forall\, y \in \mathcal{S}$*
- *There is a unique stationary distribution and $\lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{1}_{\{X(s)=y\}} ds = \pi_y$ almost surely $\forall\, y \in \mathcal{S}$.*

The expression of the limit means the proportion the CTMC staying in state $y$, and $\mathbb{1}$ has the same meaning as before. Proof omitted

**Proposition 17.** *The two cases in theorem 11 can happen in the following cases:*

- *Alternative (1) happens when all states are transient or all states are null-recurrent.*
- *Alternative (2) happens when all states are positive recurrent, with mean recurrence time*

$$-\frac{1}{\mathcal{Q}_{yy}\pi_y} = \frac{1}{v_y\pi_y}$$

(g) **Time Reversibility**

**Definition 23.** We say that $X$ is time reversible if it has a stationary distribution $\pi$ and

$$\pi_i q_{ij} = \pi_j q_{ji}$$

for $\forall$ states $i$ and $j$.

**Questions**

1. ~~Does the initial prob. distribution have to be a row of the transition matrix?~~
2. ~~For remark 3, we can keep multiplying the transition matrix because of time homo.?~~
3. For the definition of $f_i$, we if we want to get back to $i$, there can be many way to achieve. So for proposition 3 the second property, how do we define the $f_i$?