

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*, 简称为  $\mathcal{S}$ 。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

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

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 (Markov 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

$$\mathcal{P}_{ij} > 0, \sum_{j=0}^{\infty} \mathcal{P}_{ij} = 1$$

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

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

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  $\mu \cdot \mathcal{P}$  by definition of matrix product.

*Proof.* Now we are asking for  $\mathcal{P}(X_{n+1} = s_j)$  (Notice this is different from 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

$$\begin{aligned} &= \sum_{k \in \mathcal{S}} \mathcal{P}(X_{n+1} = s_j \mid X_n = s_k) \mathcal{P}(X_n = s_k) \\ &= \sum_{k \in \mathcal{S}} \mu_k \cdot (\mathcal{P})_{kj} \end{aligned}$$

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  $\mathcal{P}$ . Then we have the prob. distribution of  $X_n$  equals  $\mu \cdot \mathcal{P}^n$ .

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  $\mu \cdot \mathcal{P}^n$  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时刻是指的上一个状态是  $\mu \cdot \mathcal{P}^{n-1}$ ，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  $\mathcal{P}$ , and  $\mu$  be a distribution on  $\mathcal{S}$ . We say that  $\mu$  is a *stationary distribution* if  $\mu \cdot \mathcal{P} = \mu$ .

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

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

$$\mu_{n+1} = \mu_n \cdot \mathcal{P} \xrightarrow{n \rightarrow \infty} \mu = \mu \cdot \mathcal{P}$$

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

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

and this consequently defines the *n-step transition (prob.) matrix*  $(\mathcal{P}^{(n)})_{ij} = \mathcal{P}_{ij}^n$  (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

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

A attempt should be made later and make up here

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

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

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.*

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

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

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

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 states  $i$  and  $j$  are *communicate* ( $i \leftrightarrow j$ ) if  $i$  is accessible from  $j$  and  $j$  is accessible from  $i$ .

Above properties 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

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

then by Chap.-Kolm. Equation we have

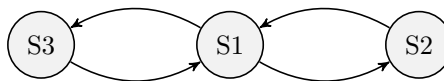
$$\mathcal{P}_{ik}^{m+n} = \sum_{\rho} \mathcal{P}_{i\rho}^m \cdot \mathcal{P}_{\rho k}^n \geq \mathcal{P}_{ij}^m \cdot \mathcal{P}_{jk}^n > 0$$

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

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

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  $(m+n)/d \in \mathbb{Z}$ ). 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  $d' = d$ .  $\square$

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

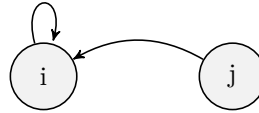
**Definition 9.** (Recurrence and Transience) Let  $f_i = \mathcal{P}(X_n = i \text{ for some } n \geq 1 | X_0 = i)$ . We say  $i$  is *recurrent* if  $f_i = 1$ , and  $i$  is *transient* if  $f_i < 1$ .

注意  $f_i$  的定义: 必须从零时刻时, 从该 state 开始.

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。

#### (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  $\mathcal{P}(\mathcal{N}_i = \infty | X_0 = i) = 1$

- If  $i$  is transient and  $X_0 = i$ , then  $\mathcal{N}_i \sim \text{Geom}(1 - f_i)$ . (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 state 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  $\mathbb{E}(\mathcal{N}_i | X_0 = i) = \infty$  (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.

$$\begin{aligned} \mathbb{E}[\mathcal{N}_i | X_0 = i] &= \mathbb{E}\left[\sum_{n=0}^{\infty} \mathbb{1}_{(X_n=i)} | X_0 = i\right] \\ &= \sum_{n=0}^{\infty} \mathbb{E}[\mathbb{1}_{(X_n=i)} | X_0 = i] = \sum_{n=0}^{\infty} \mathcal{P}_{ii}^n \end{aligned}$$

□

**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

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

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  $i \rightarrow j$ "
- If  $i \rightarrow j$  but  $j \not\rightarrow i$ , 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

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

Obviously there are three communicating classes:  $\{0\}$ ,  $\{N\}$  and  $\{1, \dots, N-1\}$ , where the first 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  $\mathcal{P}(R | X_0 = n)$ .

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

$$\begin{aligned} \mathcal{P}(\mu) &= \mathcal{P}(R | X_0 = \mu) = \mathcal{P}(R | X_1 = \mu+1)\mathcal{P}(X_1 = \mu+1 | X_0 = \mu) \\ &\quad + \mathcal{P}(R | X_1 = \mu-1)\mathcal{P}(X_1 = \mu-1 | X_0 = \mu) \end{aligned}$$

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

$$\alpha = \frac{a^N}{1 - a^N}, \quad \beta = \frac{1}{1 - a^N}$$

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

$$\mathcal{P}(n) = \begin{cases} 1 - \frac{a^n - 1}{a^N - 1}, & \text{if } \bar{p} \neq \frac{1}{2} \\ 1 - \frac{n}{N}, & \text{if } \bar{p} = \frac{1}{2} \end{cases} \quad (1)$$

*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 from 0. Since the recurrence and transience are class properties, so it is valid to just show 0 is feasible.

Starting from 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

$$\sum_{n=0}^{\infty} \frac{1}{\sqrt{\pi n}} \rightarrow \infty$$

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  $\vec{e}'_j$ s are the canonical basis of  $\mathcal{Z}^d$ . So we define the R.W at steps  $\vec{X}_i$  with probability

$$\mathcal{P}(\vec{X}_i = \vec{e}'_j) = \mathcal{P}(\vec{X}_i = -\vec{e}'_j) = \frac{1}{2d}$$

通俗讲, 在第  $i$  步时, 向  $\vec{e}_j$  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  $\vec{S}_0 = \vec{0}$  and  $\vec{S}_n = \text{position after } n \text{ steps} = \sum_{i=1}^n \vec{X}_i$ , then

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

then we can define the transition probability as

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

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

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

*Remark 8.* • The product of  $\vec{k}$  and  $\vec{S}_n$  is scalar product in  $\mathcal{R}^d$  of  $\vec{k}$  and  $\vec{S}_n$ . So

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

- 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 state  $i$  as  $n$  goes to infinity which is

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

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

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

*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,  $\pi_j = \lim_{n \rightarrow \infty} \mathcal{P}_{ij}^n$  always exists for all  $j$ , independent of  $i$ .



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

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

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

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

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

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

*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

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

*Remark 9.* • We always have  $P_{ii} = Q_{ii}$

- For a reversible M.C

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

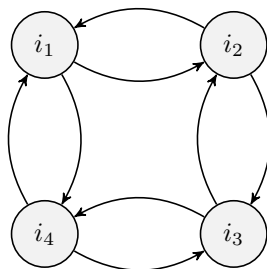
which is interpreted as the *fraction of jumps at stationarity*. This forms a so-called *detailed balance*.

**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  $(i_1, i_2, \dots, i_n)$ , we have

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

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  $S = \{0, 1, \dots, M\}$  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  $\pi$

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

## 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

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

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 from 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  $s = 0$ .

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

$$\begin{aligned} G''(1) &= \mathbb{E}(X^2) - \mathbb{E}(X) = \text{Var}(X) + (\mathbb{E}(X))^2 - \mathbb{E}(X) \\ \implies \text{Var}(X) &= G''(1) + G'(1) - (G'(1))^2 \end{aligned}$$

- 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 greater than 0.

**Proposition 7.** Let  $X_1, X_2, \dots, X_N$  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.*

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

□

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

$$\begin{aligned} G_{n+1}(s) &= \underbrace{G_{Z_{n+1}}(s) = G_{Z_n}(G_y(s))}_{\text{by above result}} \\ &= G_n(G_y(s)) = G_{n-1}(G_y(G_y(s))) = \dots = \underbrace{G_y \circ G_y \circ \dots \circ G_y(s)}_{n+1 \text{ times composition}} \end{aligned}$$

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

$$\{Z_n = 0\} \subset \{Z_{n+1} = 0\} \subset \{Z_{n+2} = 0\} \dots$$

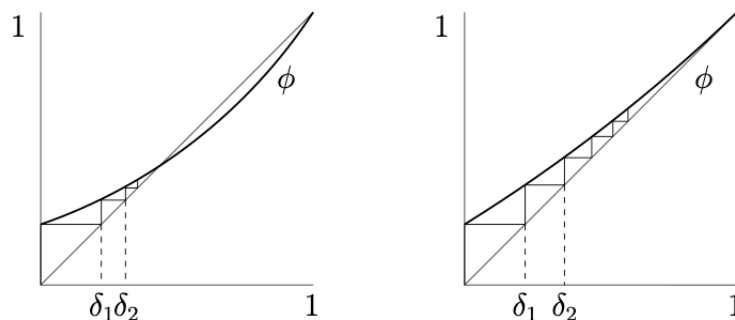
so the extinction event finally become

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

and consequently the probability of extinction becomes a limiting probability

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

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  $G'(1) = \mathbb{E}(Y) > 1$ ,  $\mathcal{P}_e = \lim_{n \rightarrow \infty} G_n(0) < 1$
- 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  $G''(1) = 0$ ,  $\text{Var}(y) = 0$ , then the process is deterministic and  $Z_n = 1 \forall n$ . 啥意思??

**Theorem 3.** Let  $\mu = \mathbb{E}(Y)$  and  $\sigma^2 = \text{Var}(Y)$  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)  $\mathcal{X} \sim \text{Exp}(\lambda)$  The pdf of a expo. distributed r.v is

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

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

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

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

for  $t < \lambda$  and infinity for  $t \geq \lambda$ , 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

$$\begin{aligned}\mathcal{P}(X \in [t, t+dt] | X > t) &= \frac{\mathcal{P}(X \in [t, t+dt])}{\mathcal{P}(X > t)} \\ &= \frac{\int_t^{t+dt} f(u)du}{1 - F(t)} \xrightarrow{dt \rightarrow 0} \frac{f(t)dt}{1 - F(t)} = r(t)dt\end{aligned}$$

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  $X_1 \sim \text{Expo}(\lambda_1)$  and  $X_2 \sim \text{Expo}(\lambda_2)$ . 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

$$\begin{aligned}\mathcal{P}(X > x) &= 1 - F_X(x) = \mathcal{P}(X_1 > x, X_2 > x) \\ &\stackrel{\text{independent}}{=} \mathcal{P}(X_1 > x)\mathcal{P}(X_2 > x) = e^{-\lambda_1 x}e^{-\lambda_2 x} \\ &= e^{-(\lambda_1 + \lambda_2)x}\end{aligned}$$

This means, the new r.v  $X$  is also a exponentially 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  $X_1, X_2, \dots, X_n$  are  $n$  independently r.v that follows  $X_i \sim \text{Exp}(\lambda_i)$ ,  $i \in [1, n]$ . Then

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

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  $\text{Exp}(\lambda)$ 。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  $X = \min(X_1, X_2) \sim \text{Exp}(2\lambda)$  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  $X' \sim \text{Exp}(\lambda)$ . 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  $X_i \sim \text{Exp}(\lambda_i)$ . Then

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

*Proof.*

$$\mathcal{P}(X_1 < X_2) = \int \int_{x_1 < x_2} f_{X_1}(x_1)f_{X_2}(x_2)dx_1dx_2$$

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

□

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

$$f_Z = \sum_{i=1}^n \left[ \prod_{j \neq i} \frac{\lambda_j}{\lambda_j - \lambda_i} \right] \cdot \lambda_i \cdot e^{-\lambda_i t}$$

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

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

with density function

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

[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  $t$ ).

### (a) The first description

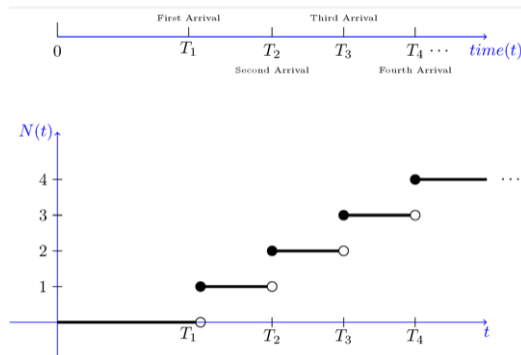
A poisson process with rate  $\lambda$  is a counting process where the time between 2 event is  $Exp(\lambda)$ . The time until  $n$  event happens is

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

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

Let the counting process to be  $N$  and

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



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

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

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

$$N(t) \sim \text{Poisson}(\lambda t)$$

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

### Example

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

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.v.s. (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 discription:

**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

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

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

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

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

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

$\Longleftrightarrow$

$N(t)$  is a counting process with iid inter arrival times  $\sim \text{Exp}(\lambda)$  (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  $\lambda_2$ . Then  $N(t)$  is a poisson process of rate  $\lambda_1 + \lambda_2$ .

Proof needed from note March 4.

**Proposition 13.** (Poisson Timing) 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 prob.  $1 - p$ . 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.



**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  $f_i$ ?