MATH495 Stochastic Process

Albert Peng

May 6, 2023

Contents

1 Intro

Stochastic Process: A random measurement evolving in time which could be discrete or continuous (Ex. Price of stock, temperature). We will study some basic building block of type of stochastic processes that are applied to model many more realistic processes.

1. **Markov Process**. Discrete time, taking finitely many values (Markv Chains). Or Continuous time.
2. **Poisson Process** and its generalizations.
3. **Brownian Motion**
4. **Stochastic Integral** with respect to Brownian motion.

2 Markov Processes

Definition. Stochastic Process. If X_t is value of the process at time $t \in T$, $\{X_t\}$ is the set of random variables.

We denote X_n for discrete values. In this case, **Markov Chain** $\{X_n\}_{n\in\mathbb{N}}$ where each X_n taies values on a set $S = \{\mathbb{N}\}$ (called state space of the process) and the following so-called Markov property holds:

$$\begin{aligned}\mathbb{P}\left[X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0\right] &= \mathbb{P}[X_{n+1} = j \mid X_n = i] \\ &= \mathbb{P}\left[X_1 = j | X_0 = i\right]\end{aligned}$$

This means that

1. Future and past are independent given present (memoryless)
2. The process is homogenenous, where given X_n, the process resets itself and we can assume it starts over in time.

Definition. The **Transition Probabilities** of the Markov chain is defined as

$$p(i,j) := \mathbb{P}[X_{n+1} = j \mid X_n = i] = \mathbb{P}[X_1 = j \mid X_0 = i]$$

and we can represent the Markov chain as a **transition matrix** of

$$P = \begin{bmatrix} p(1,1) & p(1,2) & ... & p(1,N) \\ p(2,1) & p(2,2) & \dots & p(2,N) \\ \vdots & \vdots & \ddots & \vdots \\ p(N,1) & p(N,2) & \dots & p(N,N) \end{bmatrix}$$

We can represent the transition probabilities as a directed grapp where vertices are states joined by arrows, named according to the corresponding transition probabilities.

Example. (Ehrenfest Chain) There are 2 urns with N balls. We pick one ball at random and move it to the other urn. X_n is number of balls in the first urn after the nth urn. At the beginning, there are 4 balls in the first and 3 balls in the second. What is the state space S, the transition probability and matrix, and the transition graph?

$S = \{0, ..., N\}$. With

$$p(i,j) = \begin{cases} \frac{i}{N}, & j = i-1 \\ \frac{N-i}{N}, & j = i+1 \\ 0, & \text{otherwise} \end{cases}, \; P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \dots \\ \frac{1}{N} & 0 & \frac{N-1}{N} & 0 & 0 & \dots \\ 0 & \frac{2}{N} & 0 & \frac{N-2}{N} & 0 & \dots \end{bmatrix}$$

Example. Let X_n be the invevntory at the end of day n, and D_{n+1} is the demand in day $n+1$.

$$\mathbb{P}[D_{n+1}=i] = \begin{cases} 0.3 & i=0 \\ 0.4 & i=1 \\ 0.2 & i=2 \\ 0.1 & i=3 \end{cases}$$

The policy is that if the inventory stock at the end of day n is ≤ 1, then we order enough to bring it to $S=5$. Thus, $S=\{0,1,2,3,4,5\}$.

$$X_{n+1} = \begin{cases} (X_n - D_{n+1})_+, & X_n > 1 \\ (5-D_{n+1})_+, & X_n \leq 1 \end{cases} \implies P = \begin{bmatrix} 0 & 0 & 0.1 & 0.2 & 0.4 & 0.3 \\ 0 & 0 & 0.1 & 0.2 & 0.4 & 0.3 \\ 0.3 & 0.4 & 0.3 & 0 & 0 & 0 \\ 0.1 & 0.2 & 0.4 & 0.3 & 0 & 0 \\ 0 & 0.1 & 0.2 & 0.4 & 0.3 & 0 \\ 0 & 0 & 0.1 & 0.2 & 0.4 & 0.3 \end{bmatrix}$$

Preview: $\mathbb{P}[X_3=j|X_0=i] = [P^3]_{i,j}$, the (i,j)th entry of the 3rd power of matrix P. Remark: In R, use package "expm" and function `P %^% n`

2.1 Multiple Transition Probabilities

By the **Law of Total Probability**, if $\{B_k\}$ is a partition of Ω,

$$\mathbb{P}[A] = \sum_k \mathbb{P}[A|B_k]\mathbb{P}[B_k] \to \mathbb{P}[A|C] = \sum_k \mathbb{P}[A|\underbrace{B_k, C}_{B_k \cap C}] \cdot \mathbb{P}[B_k|C]$$

Property 1:

$$\begin{aligned} \mathbb{P}[X_{n+m}=j|X_n=i, X_{n-1}=i_{n-1}, ...] &= \mathbb{P}[X_{n+m}=j|X_n=i] \\ &= \mathbb{P}[X_m=j|X_0=i] \end{aligned}$$

Proof. Induction in m. For $m=1$, this is the markov chain property. Now, suppose they are true for m.

$$\begin{aligned} \mathbb{P}[\underbrace{X_{n+m+1}=j}_{A}|\underbrace{X_n=i,...}_{C}] &= \sum_k \mathbb{P}[X_{n+m+i}=j|X_{n+1}=k, X_n=i] \cdot \mathbb{P}[X_{n+1}=k|X_n=i, X_{n-1}=...] \\ &= \sum_k \mathbb{P}[X_{n+m+1}=j|X_{n+1}=k, X_n=i] \cdot \mathbb{P}[X_{n+1}=k|X_n=i] \\ &= \mathbb{P}[X_{n+m+i}=j|X_n=i] \\ &= \sum \mathbb{P}[X_m=j|X_0=k]\mathbb{P}[X_1=k|X_0=i] \\ &= \sum \mathbb{P}[X_{m+1}=j|X_1=k] \cdot \mathbb{P}[X_1=k|X_0=i] \\ &= \mathbb{P}[X_{m+1}=j|X_0=i] \end{aligned}$$

■

We denote

$$p^{(m)}(i,j) := \mathbb{P}[X_m=j|X_0=i]$$

Theorem. [Chapman-Kolmoguvov Equations]

$$\mathbb{P}[X_{n+m} = j|X_0 = i] = \sum_k \mathbb{P}[X_m = k|X_0 = ki] \cdot \underbrace{\mathbb{P}[X_{n+m} = j|X_m = k]}_{\mathbb{P}[X_n=j|X_0=k]}$$

$$p^{(m+n)}(i,j) = \sum_k p^{(m)}(i,k) \cdot p^{(n)}(k,j) = P_{i,j}^{(n+m)}$$

Example. A gambler need N dollars but has only i dollars. He plays games that gives him 1 dollar with probability p and loose 1 dollar with proability q, and $p+q=1$. When his fortune is either N or 0, he stops. Now, suppose $N = 4$, $p = 0.4$, and he started with $X_0 = 1$. What is the probability the gambler is still playing after 20 games?

$\mathbb{P}[X_{20} \notin \{0,4\}|X_0 = 1] = 1 - \mathbb{P}[X_{20} = 4|X_0 = 1] - \mathbb{P}[X_{20} = 0|X_0 = 1] = 1 - [P^{20}]_{1,4} - [P^{20}]_{1,0}$

Now, we can also get $P[X_n = j]$ given P and $\alpha = [\mathbb{P}[X_0 = 1], ..., \mathbb{P}[X_0 = N]]$ simply using the total law of probability, where

$$\begin{aligned}\mathbb{P}[X_n = j] &= \sum_i \mathbb{P}[X_n = j|X_0 = i]\mathbb{P}[X_0 = i] \\ &= \sum_i P_{i,j}^n \alpha_i \implies \boxed{\mathbb{P}[X_n = j] = [\alpha P^n]_j}\end{aligned}$$

Thus, for $n_1 < ... < n_k$,

$$\mathbb{P}[X_{n_1} = i_1, ..., X_{n_k} = i_k] = [\alpha P^{n_1}]_{i_1} [P^{n_2 - n - 1}]_{i_1, i_2} \cdots [P^{n_k - n_{k-1}}]_{i_{k-1}, i_k}$$

Example. Suppose we are trying to calculate $\mathbb{P}[X_9 = 1|X_1 = 3, X_4 = 1, X_7 = 2]$, the long way is

$$\frac{\mathbb{P}[X_1 = 3, X_4 = 1, X_7 = 2, X_9 = 1]}{\mathbb{P}[X_1 = 3, x_4 = 1, X_7 = 2]} = \frac{[\alpha P]_3 [P^3]_{3,1} [P^3]_{1,2}, [P^2]_{2,1}}{[]\alpha P]_3, [P^3]_{3,1} [P^3]_{1,2}} = [P^2]_{2,1}$$

But notice the memoryless property so we simply have $[P^2]_{2,1}$

2.2 Long Run Behavior of Markov Chain

When does $\pi_j := \lim_{n\to\infty} \mathbb{P}[X_n = j]$ exist? Does it depend on the initial probabilities α? Now, we know that this problem is equivalent to finding if $\pi = \lim_{n\to\infty} \alpha P^n$ exists.

Proposition. $\lim_{n\to\infty} \mathbb{P}(X_n = j) = \pi_j$ if and only if $P_{i,j}^n = \mathbb{P}[X_n = j|X_0 = i] \to \pi_j$.

The vector $\pi = [\pi_1, ..., \pi_N]$ is called the **limiting distribution** or **ergodic distribution**.

Theorem. When there exists $\lim_\infty \mathbb{P}(X_n = j = \pi_j)$ for all j, we have $\pi P = \pi$

Proof.

$$\pi = \lim_{n\to\infty} \alpha P^n = \lim_{n\to\infty} \alpha P^{n+1} = \lim_{n\to\infty} \alpha P^n P = \pi P$$

■

Thus, π is a **left eigenvector** of the matrix P corresponding to eigenvalue 1 for the first expression.

Definition. A probability vector π satisfying $\pi P = \pi$ is called a **stationary, equilibrium, or invariant** probability distribution of the Markov Chain.

Theorem. The limiting distribution $pi = \lim_{n\to\infty} \alpha P^n$ is a stationary distribution $\pi P = \pi$.

If the initial distribution α is a stationary distribution q, then $[\mathbb{P}[X_n = 1], ..., \mathbb{P}[X_n = N]] = q \, \forall \, n$.

Proof. $[\mathbb{P}[X_n = 1], ..., \mathbb{P}[X_n = N]] = \alpha P^n = qP^n = (qP)P^{n-1} = qP^{n-1} = ... = qP^0 = qI_n = q$ ■

Claim: 1 is always an eigenvalue of P or P^T since $P[1 \cdots 1]^T = [1 \cdots 1]$ by definition of stochastic matrix where each row sums to 1.

Suppse $Pv = \lambda v$. Then, $\lambda \leq 1$, and with $|v - i| = \max_j |v_j|$

$$i\text{th entry} \rightarrow |\lambda||v_i| = \left|\sum_{j=1}^{N} p(i,j)v_j\right| \leq \sum_{j=1}^{N} |p(i,j)v_j| = \sum_{j=1}^{N} p(i,j)|v_j| \leq \sum_{j=1}^{N} p(i,j)|v_i|$$

$|\lambda||v_i| \leq |v_i| \rightarrow |\lambda| \leq 1$.

2.3 Existence of Limit

So, when does $\pi = \lim_{n\to\infty \alpha P^n}$ exist?

Method 1

We can exploit the eigen-decomposition of P. If $P = Q\Lambda Q^{-1}$, then $P^2 = Q\Lambda^2 Q^{-1}$. So, The limit is

$$Q \begin{bmatrix} \lim_{n\to\infty} \lambda_1^n & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \lim_{n\to\infty} \lambda_N^n \end{bmatrix} Q^{-1} \implies \exists \text{ if } |\lambda_i| \leq 1, \text{which holds}$$

For $\lim_{n\to\infty} \alpha P^n$ to be some π regardless of α, we need the eigenvalue $\lambda = 1$ to be simple. Then, $\lambda = 1, |\lambda_j| < 1 \, \forall j \neq 1$.

Method 2

A simpler criterion is to check if all entries of P^m are positive for some m. This type of matrix is called *regular*.

In summary, $q_i = \lim_{n\to\infty} \mathbb{P}[X_n = i]$ exists regardless of the initial distribution α. The stationary distribution $\pi P = \pi$ where P is transition matirx. A limiting distribution is stationary. Sufficent condition for existence of q and uniqueness if P^m has only positive entries, $\exists m = 1, 2, ...$ Another ssufficient condition is that P has eigendecomposition, $\lambda = 1$ is a simple eigenvalue, with a eigenvector v having all entries non-negative.

2.4 Recurrent and Transient States

Definition. A state is **accesible** from i if $P_{i,j}^m > 0$ for some $m \geq 0$. We then say i, j **communicate** if $i \to j$ and $j \to i$ are both possibe and write it as $i \leftrightarrow j$.

Claim: $i \leftrightarrow j$ is a equivalence relation:

- Reflexive: $i \leftrightarrow i$
- Symmetric: $i \leftrightarrow j \implies j \leftrightarrow i$
- Associative: $i \leftrightarrow j, j \leftrightarrow k \implies i \leftrightarrow k$

Definition. An **irreducible chain** is where all states communicate with each other. In other words, there exists only one communication class.

Definition. A state i is **recurrent** if $f_i := \mathbb{P}[\text{back to } i | X_0 = i] = 1$. Otherwise, we say i is **transient**.

Theorem. The state of a communication class are either all recurrent or all transient. Criteria for recurrence and transience

1. State i is recurrent iff $\sum_{n=0}^{\infty} P_{i,i}^n = \infty$.
2. State i is transient iff $\sum_{n=i}^{\infty} P_{i,i}^n < \infty$

Simply, $\mathbb{P}[N(i) = \infty | X_0 = i] = 1 \implies \mathbb{E}[N(i)|X_0 = i] = \infty$

Proof. Suppose $i \leftrightarrow j$ is current, and show that j is recurrent. In other words, prove $\sum_{n=0}^{\infty} p_{ii}^n = \infty \implies \sum_{n=0}^{\infty} p_{jj}^n = \infty$. Note that $i \leftrightarrow j \implies P_{i,j}^r > 0, P_{j,i}^m > 0$. Now,

$$P_{j,j}^{r+k+m} = \sum_{l,l'} P_{j,l}^r P_{l,l'}^k P_{l',j}^m \geq P_{j,i}^m P_{i,i}^k P_{i,j}^r$$

Then,

$$\sum_{k=0}^{\infty} \geq \sum_{k=0}^{\infty} P_{j,i}^m P_{i,i}^k P_{i,j}^r = P_{j,i}^m P_{i,j}^r \sum_{k=0}^{n} P_{i,i}^k = \infty \implies \sum_{n=0}^{\infty} P_{j,j}^n = \infty$$

■

Definition. A class is **recurrent** if at least one of its states is recurrent.

Implications:

- An irreducible (all states communicate) finite chain is recurrent.
- A finite chain has at least one recurrent class.
- A recurrent communication class $\mathcal{C}$ is **closed** if $\forall i \in \mathcal{C}, \forall j \notin \mathcal{C}, p_{i,j} = 0$
- A finite closed communication class must be recurrent.

2.5 Canonical Decomposition of Transition Matrix

Check picture on board

2.6 Transience Continued...

Proposition. The stationary distribution of a finite irreducible chain always exists and is unique.

Remark: The limiting distribution does not necessarily exist.

2.7 First Passage Time

Let $T_i = \min\{n > 0; X_n = i\}$, the earliest occurance of state i. Also let $N_n(i)$ = number of visited of state i up to time n

Theorem. Consider an irreducible, recurrent chain not necessarily finite. Then,

1. $$\frac{N_n(i)}{n} \xrightarrow[n\to\infty]{} \frac{1}{\mathbb{E}[T_i|X_0 = i]}$$
2. $$\frac{1}{N} \sum_{k=1}^{n} P_{l,i}^k \xrightarrow[n\to\infty]{} \frac{1}{\mathbb{E}[T_i|X_0 = i]}$$
3. $\pi_i := \frac{1}{\mathbb{E}[T_i|X_0=i]}$ is the unique stationary distribution

Proof. (1). Now, let $\tau_i(k)$ be the time of the k-th visit$= \sum_{j=1}^{k} T_i^{(j)}$. Then by the strong law of large numbers,

$$\lim_{k\to\infty} \frac{\tau_i(k)}{k} = \mathbb{E}[T_i^{(2)}] = \mathbb{E}[T_i|X_0 = i]$$

Also notice that

$$N_n(i) = \sum_{k=0}^{n} \mathbb{I}_{[X_k=i]} = \max\{k;\ \tau_i(k) \leq n\} = \text{number of } \{k; \tau_i(k) \leq n\} = N$$

Then,

$$\tau_i(N_n(i)) \leq n \leq \tau_i(N_n(i)+1) \implies \frac{\tau_i(N_n(i))}{N_n(i)} \leq \frac{n}{N_n(i)} \leq \frac{\tau_i(N_n(i)+1)}{N_n(i)+1} \cdot \frac{N_n(i)+1}{N_n(i)}$$

Which means that as $N_n(i) \to \infty$,

$$\frac{\tau_i(N_n(i))}{N_n(i)} = \mathbb{E}[T_i|X_0 = i], \frac{\tau_i(N_n(i)+1)}{N_n(i)+1} \cdot \frac{N_n(i)+1}{N_n(i)} = \mathbb{E}[T_i|X_0 = i] \cdot 1$$

Thus, by squeezing, $\lim_{n\to\infty} \frac{n}{N_n(i)} = \mathbb{E}[T_i|X_0 = i]$ ■

Proof. (2). Let $m := \frac{1}{\mathbb{E}[T_i|X_0=i]}$

$$\lim_{n\to\infty} \mathbb{E}\left[\frac{N_n(i)}{n}|X_0 = l\right] = \mathbb{E}\left[\lim_{n\to\infty} \frac{N_n(i)}{n}|X_0 = l\right] = \mathbb{E}[m|X_0 = i] = m$$

Then,

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}\left[\sum_{k=0}^{n} \mathbb{I}_{[X_k=i]}|X_0 = l\right] = \lim_{n\to\infty} \frac{1}{N} \sum_{i=1}^{n} \mathbb{E}[\mathbb{I}_{[X_k=i]}|X_0 = l] = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} P_{l,i}^{(k)}$$

■

Proof. (3).

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} P_{l,i}^{(k)} = m_i \implies \frac{1}{n} \sum_{k=1}^{m} \begin{bmatrix} P_{1,1}^k & \dots & P_{1,N}^k \\ \vdots & \ddots & \vdots \\ P_{N,1}^k & \dots & P_{N,N}^k \end{bmatrix} \xrightarrow{n\to\infty} \begin{bmatrix} m_1 & \dots & m_N \\ \vdots & \ddots & \vdots \\ m_1 & \dots & m_N \end{bmatrix}$$

... As a result,

$$\begin{bmatrix} m_1 & \dots & m_N \\ \vdots & \ddots & \vdots \\ m_1 & \dots & m_N \end{bmatrix} = \begin{bmatrix} m_1 & \dots & m_N \\ \vdots & \ddots & \vdots \\ m_1 & \dots & m_N \end{bmatrix} P \implies \pi = \pi P, \pi := [m_1, ..., m_n]$$

■

Corollary. Any irreducible and finite chine has a unique stationary distribution.

Note that if $P_{i,j}^m > 0 \forall i, j \exists m \implies$ stationary distribution exists

Theorem. Addition to theorem above:

$$\lim_{n\to\infty} \frac{\mathbb{E}[N_n(i)]}{n} = \frac{1}{\mathbb{E}[T_i|X_0 = i]}$$

Theorem. For a finite irreducible chain (thus recurrent), The stationary distribution π exists, is unique, and is given by

$$\pi_i = \frac{1}{\mathbb{E}[T_i|X_0 = i]}$$

Proof. First, show that $\pi P = \pi$.

Then, show $\sum_{i=1}^{N} \pi_i = 1, \pi \geq 0$. For $\pi P = \pi$, start with the theorem so that... ■

Theorem. For an irreducible, recurrent (not necessarily finite) chain that has a stationary distribution π,

$$\mathbb{E}[T_i | X_0 = i] = \pi_i$$

Proof. Use theorem 1 with initial distribution $\alpha = \pi$ to show that $\frac{\mathbb{E}[N_n(i)]}{n} = \pi_i \forall n$. So, $\pi_i = \frac{1}{\mathbb{E}[T_i|X_0=i]}$ ■

Example. A finite irreducible chain has a unique stationary distribution, that is not necessarily a limiting distribution.

(Consider an odd even sequence.)

2.8 Periodicity

Definition. The **period** of state i is the largest integer that divides all possible return times n,

$$d(i) = \gcd(n > 0; P_{ii}^n > 0)$$

If $d(i) = 1$, we say that i is aperiodic. $d(i) = \infty$ if $P_{ii} = 0 \forall n$.

Theorem. All states in the same communication class have the same period:

$$i \leftrightarrow j \implies d(i) = d(j)$$

Proof. $i \leftrightarrow j \implies d(i) \leq d(j)$: Let n be a possible returning time to j, $P_{j,j}^n > 0$. If $P_{i,j}^r > 0, P_{j,i}^s > 0$, $d(i) | r + s$ because $P_{i,i}^{r+s} > 0$.

Also, $P_{j,j}^n > 0 \implies d(i) | r + s + n$ since $P_{i,i}^{n+r+s} > 0$.

Together, it is obvious that $d(i)$ divides $(n + r + s) - (r + s) = n \implies d(i) \leq d(j)$.

Finally, we use the same argument to get $d(j) \leq d(i)$ ■

Theorem. **[Eryodic Theorem]** An aperiodic ($d(i) = 1$) finite irreducible chain $\implies$ limiting distribution exists $\implies \exists!$ stationary distribution $\pi P = \pi$, with

$$\lim_{n \to \infty} P_{i,j}^n = \pi_j \forall i, j$$

2.9 [TBD]

Suppose the chain has transient and recurrent states. We know the chain visits a transient state finite many times $N(i)$ before it is absorbed by one of the recurrent classes.

In general, we want to find $\mathbb{E}[N(i)|X_0 = i]$ and $\mathbb{P}[X_n \in \mathbb{R}_j | X_0 = i]$ for some recurrent class R_j

Given the canonical form of the matrix P, then for transient states $i, j \in \mathcal{T}$,

$$N(j) = \text{number of visits of state } j = \sum_{n=1}^{\infty} \mathbb{I}_{[X_n = j]} \text{ and } \mathbb{E}[N(j)|X_0 = i] = (I - Q)_{i,j}^{-1}$$

Proof. Denote $E_i[\cdot] = \mathbb{E}[\cdot|X_0 = i]$. Then,

$$\mathbb{E}_i[N(j)] = \mathbb{E}_i\left[\sum_{n=0}^{\infty} \mathbb{I}_{[X_n=j]}\right] = \sum_{n=0}^{\infty} \mathbb{P}[X_n = j|X_0 = i] = \sum_{n=0}^{\infty}[P^n]_{i,j} = \left[\sum_{n=0}^{\infty} P^n\right]_{i,j}$$

■

For Q being the top-left non-zero portion of the canonical representation, then for $i, j \in \mathcal{T}$,

$$\left[\sum_{n=0}^{\infty} P^n\right] = \left[\sum_{n=0}^{\infty} Q^n\right] = (I - Q)^{-1}_{i,j}$$

The number of steps before visits to recurrent classes can is essentially the sum $\sum_{i=1}^{n} N(i)$.

Corollary. Let N be the number of steps before the chain enters one of the recurrent classes. Then for $i \in \mathcal{J}$,

$$\mathbb{E}[N \mid X_0 = i] = \sum_j M_{i,j}, \text{ where } M = (I - Q)^{-1}$$

Corollary. Suppose $|R_i| = 1$. i.e., all recurrent states consist of only one element. Then, for $i \in \mathcal{T}, j$ recurrent, S being the part of canonical matrix to the right of Q,

$$\alpha(i, j) = \mathbb{P}(X_n = j \text{ eventually } |X_0 = j) = \left[(I - Q)^{-1}S\right]_{i,j}$$

$$= \mathbb{P}\left[\cup_{n=0}^{\infty}\{X_n = j\}|X_0 = i\right] = \lim_{n\to\infty} \mathbb{P}[X_n = j|X_0 = i], \text{ with } \{X_n = j\} \subseteq \{X_{n+1} = j\}$$

<u>Connection to Irreducible Chains</u> If i, j are recurrent, then

$$\mathbb{E}[\text{steps from } i \to j|X_0 = i]$$

Another approach we can take is to condition on the outcome of the first move.

$$\mathbb{P}(A|X_0 = i) = \sum_{j\in S} \mathbb{P}[A|X_1 = j, X_0 = i]\mathbb{P}[X_1 = j|X_0 = i] = \sum_{j\in S} \mathbb{P}[A|X_1 = j]P_{i,j}$$

$$\mathbb{E}[Y|X_0 = i] = \sum_{j\in S} \mathbb{E}[Y|X_1 = j]P_{i,j}$$

Looking back at the previous theorem: A finite irreducible, and aperiodic chain has a limiting distribution $\lim_{n\to\infty} \mathbb{P}[X_n = i] = \pi_i$

Proof. Consider another chain Y independent of X and starting in the stationary distribution π.

There exists a time T such that $X_T = Y_T$. Then

$$\begin{aligned}
\mathbb{P}[X_n = i] &= \mathbb{P}[X_n = i, T \le n] + \mathbb{P}[X_n = i, T > n] \\
&= \mathbb{P}[Y_n = i, T \le n] + \mathbb{P}[X_n = i, T > n] \\
&= \mathbb{P}[Y_n = i] - \mathbb{P}[Y_n = i, T > n] + \mathbb{P}[X_n = i, T > n] \\
&= \pi_i - \underbrace{\mathbb{P}[Y_n = i, T > n]}_{\le \mathbb{P}[T>0]\to 0} + \underbrace{\mathbb{P}[X_n = i, T > n]}_{\le \mathbb{P}[T>n]\to 0} \\
&\xrightarrow{x\to\infty} \pi_i
\end{aligned}$$

■

Theorem. Let S be irreducible, recurrent, and countable. If

$$\mathbb{E}[T_i|X_0 = i] < \infty,\ T_i = \min\{n > 0 : X_n = 1\}$$

then a stationary distribution exists

3 Poisson Process

Definition. Exponential distribution $T \sim \exp(\lambda)$ is a positive continuous random variable with density function

$$f(t) = \lambda e^{-\lambda t}$$

Note that

$$\mathbb{P}[T > t] = e^{-\lambda t},\ \mathbb{E}[T] = \frac{1}{\lambda}, Var[T] = \frac{1}{\lambda^2}$$

This also posesses the **memoryless** property, where

$$\mathbb{P}[T > t + s | T > t] = \mathbb{P}[T > s]$$

[**Exponential Races**] Let $T_1 \sim exp(\lambda_1), ..., T_n \sim exp(\lambda_n)$ be independent, then the minimum time to finish follows the distribution

$$\mathbb{P}[\min\{T_1, ..., T_n\} > t] \sim exp(\lambda + \cdots + \lambda_n)$$

Proof.

$$\begin{aligned}
\mathbb{P}[\min\{T_1, ..., T_n\} > t] &= \mathbb{P}[T_1 > t, ..., T_n > t] \\
&= \mathbb{P}[T_1 > t] \cdots \mathbb{P}[T_n > t] \\
&= e^{-\lambda_1 t} \cdots e^{-\lambda_n t} = e^{-(\lambda_1 + ... + \lambda_n)t}
\end{aligned}$$

■

$\mathbb{P}[T_i \text{ smallest}] = \frac{\lambda_i}{\lambda_1 + ... + \lambda_n}$

$T_1 \wedge ... \wedge T_N$ and I are independent

Proof. It is sufficient for us to prove $\mathbb{P}[I = 1, T_1 \wedge \cdots \wedge T_n \geq t] = \frac{\lambda_1}{\lambda_1 + \cdots + \lambda_n} e^{-(\lambda_1 + ... + \lambda_n)t}$. Here,

$$\begin{aligned}
LHS &= \mathbb{P}[T_1 \geq t, T_2 \geq T_1, ..., T_n \geq T_1] \\
&= \int_t^\infty \mathbb{P}[T_1 \geq t, T_2 \geq T_1, ..., T_n \geq T_1 | T_1 = x_1] \lambda_1 e^{-\lambda_1 x_1}\, dx_1 \\
&= \int_t^\infty \mathbb{P}[T_2 \geq x_1, ..., T_n \geq x_1] \lambda_1 e^{-\lambda_1 x_1}\, dx_1 \\
&= \int_t^\infty \mathbb{P}[T_2 \geq x_1] \cdots \mathbb{P}[T_n \geq x_1] \lambda_1 e^{-\lambda_1 x_1}\, dx_1 \\
&= \int_t^\infty e^{-\lambda_2 + ... + \lambda_n} \lambda_1 e^{-\lambda_1 x_1}\, dx_1 \\
&= \frac{\lambda_1}{\lambda_1 + ... + \lambda_n} e^{-(\lambda_1 + ... + \lambda_n)t}
\end{aligned}$$

■

Sum of exponentials. Let $\tau_1, .., \tau_n$ be independent dsitrbution following $exp(\lambda)$. Then,

$$T_n = \tau_1 + ... + \tau_n \sim Gamma(n, \lambda) \sim f_{T_n}(t) = \frac{\lambda^n e^{-\lambda t} t^{n-1}}{(n-1)!},\ t > 0$$

Poisson Distribution. $X \sim Pois(\lambda)$ if X is a discrete random variable taking values on $\{0, 1, ...\}$ according to the PMF

$$p_x(k) = \mathbb{P}[X = k] = e^{-\lambda} \frac{\lambda^k}{k!}$$

Note that this follows the property that

$$\mathbb{E}[X] = Var(X) = \lambda$$

The sum of poisson distribution where $X_i \sim Pois(\lambda_i)$ is given by

$$X_1 + ... + X_n \sim Pois(\lambda_1 + ... + \lambda_n)$$

Poisson process counts N_t, the number of jumps up to time t. In application, N_t counts the number of events happening up to time t.

Properties:

1. $N_0 = 0$
2. $N_t - N_s \sim Pois(\lambda(t - s))$, for some fixed λ called the rate of N.
3. If $t_1 < t_2 < ... < t_n$, then $N(t_2) - N(t_1), ..., N(t_n) - N(t_{n-1})$ are independent.

Example. If $\lambda = 2$, then

$$\begin{aligned}
\mathbb{P}[N_3 = 0, N_5 = 1, N_8 = 5] &= \mathbb{P}[N_3 = 0, N_5 - N_3 = 1, N_8 - N_5 = 4] \\
&= \mathbb{P}[N_3 = 0]\mathbb{P}[N_5 - N_3 = 1]\mathbb{P}[N_8 - N_5 = 4] \\
&= \mathbb{P}[Pois(6) = 0] \cdot \mathbb{P}[Pois(4) = 1] \cdot \mathbb{P}[Pois(6) = 4] \\
&= \left(e^{-6} \frac{6^0}{0!}\right) \left(e^{-4} \frac{4^1}{1!}\right) \left(e^{-6} \frac{6^4}{4!}\right)
\end{aligned}$$

Example. Let $\lambda = 2$. Find $\mathbb{E}[N_1 N_3]$. Note that N_1 and N_3 are not independent, but N_1 and N_3 are. Thus,

$$\begin{aligned}
\mathbb{E}[N_1 N_3] &= \mathbb{E}[N_1(N_1 + N_3 - N_1)] \\
&= \mathbb{E}[N_1^2] + \mathbb{E}[N_1(N_3 - N_1)] \\
&= Var(N_1) + \mathbb{E}[N_1]^2 + \mathbb{E}[N_1]\mathbb{E}[N_3 - N_1] \\
&= 2 + 2^2 + 2 \times 4 = 14
\end{aligned}$$

Similarly, we can rewrite

$$\mathbb{E}[N_1 N_2 N_4] = \mathbb{E}[N_1((N_2 - N_1) + N_1)((N_4 - N_2) + (N_2 - N_1) + N_1)]$$

For $T_j = \min\{t : N_t \geq j\}$, let $T_1, T_2, ...$ be arrival times of jump or events.

Then,$\tau_1, \tau_2, ...$ be the interarrival times, where $\tau_1 = T_1, \tau_j = T_j - T_{j-1}, j = 2, 3, ...$

Theorem. $\tau_1, ...,$ are exponential with rate λ and independent.

To verify, notice that $\mathbb{P}[\tau_1 > t] = \mathbb{P}[N_t = 0] = e^{-\lambda t} \implies \tau_1 \sim exp(\lambda)$. Then, $\mathbb{P}[\tau_2 > s | \tau_1 = t] = \mathbb{P}[N_{t+s} - N_t = 0] = e^{-\lambda s}$.

Application: we can construct or simulate a Poisson process by generating $\tau_1, .., \tau_n \sim exp(\lambda)$.

Generalizations. For **nonhomogeneous poison processes**, the likelihood of an event happening near time t may depend on time t. We want

$$\mathbb{P}[N_{t+\Delta} - N_t = 1] = \lambda(t)\Delta + o(\Delta), \text{ where } \frac{o(\Delta)}{\Delta} \xrightarrow[D\to 0]{} 0$$

For a poisson process, $\mathbb{P}[N_{t+\Delta} - N_t = 1] = e^{\lambda\Delta}\lambda\Delta = (1 - \lambda\Delta + \frac{(\lambda\Delta)^2}{2!} - ...)\lambda\Delta = \lambda\Delta + o(\Delta)$.

Definition. $\{N_t\}_{t\geq 0}$ is a **nonhomogeneous poisson process** with rate function $\lambda(t) : [0,\infty) \to [0,\infty)$ if

- $N_0 = 0$
- For $s < t$, $N_t - N_s \sim Pois\left(\int_s^t \lambda(u)\, du\right)$.
- If $t_1 < t_2 < ... < t_n$, then $N(t_2) - N(t_1), ..., N(t_n) - N(t_{n-1})$ are independent.

Exercise: $\mathbb{P}[N_{t+\Delta} - N_t = 1] = \lambda(t)\Delta + \theta(\Delta)$

To simulate this, start with a regular Poisson Process with rate $\lambda = 1$, then for homogenenous $\{N_t^h\}_{t\geq 0}$,

$$T(t) = \int_0^t \lambda(u)\, du,\ T'(t) = \lambda(t)$$

Then $N_t = N^h_{T(t)}$ is a nonhomogeneous Poisson Process with rate $\lambda(t)$. Then for $s < t$,

$$\mathbb{E}[N_t - N_s] = \mathbb{E}[N^h_{T(t)} - N^h_{T(s)}] = T(t) - T(s) = \int_s^t \lambda(u)\, du \implies N_t - N_s \sim Pois(T(t) - T(s))$$

Example. Student arrive at a cafeteria according to a nonhomogeneous poisson process with

$$\lambda(t) = \begin{cases} 100 + 100t, & 0 \leq t \leq 1 \\ 200, & 1 \leq t \leq 3 \\ 500 - 100t, & 3 \leq t \leq 4 \end{cases}$$

If $t = 0$ is 11:00AM, find the probability that at least 400 students arrive between 11:30 and 1:30.

Compound Poisson Process. $Y_1, Y_2, ...$ are i.i.d. with same distribution, independent of the times of jumps. Let N_T be the number of jumps up to time $t \sim Pois(\lambda)$.

$$X_t = \begin{cases} 0, & N_t = 0 \\ Y_1, & N_t = 1 \\ Y_1 + ... + Y_k, & N_t = k \end{cases} = \sum_{i=1}^{N_t} Y_i$$

Example. Cars are arriving at McDonalds between 12:00 and 1:00 according to a poisson process with a rate of 2 cars per minute. Suppose we want to determine the number of customers that McDonalds serves at any given time t.

$$Y_i = \begin{cases} 1, & p = 0.1 \\ 2, & p = 0.3 \\ 3, & p = 0.3 \\ 4 & p = 0.2 \\ 5 & p = 0.1 \end{cases} \implies \mathbb{E}(X_1) = \mathbb{E}\left(\sum_{i=1}^{N_1} Y_i\right) = \mathbb{E}(N_1)\mathbb{E}(Y_1)$$

$$
\begin{aligned}
Var(X_1) = Var\left(\sum_{i=1}^{N_t} Y_i\right) &= \mathbb{E}\left(Var\left(\sum_{i=1}^{N_T} Y_t | N_t\right)\right) + Var\left(\mathbb{E}\left(\sum_{i=1}^{N_t} Y_i | N_t\right)\right) \\
&= \mathbb{E}\left(\sum_{i=1}^{N_t} Var(Y_i)\right) + Var\left(\sum_{i=1}^{N} \mathbb{E}[Y_i]\right) \\
&= \mathbb{E}(N_t Var(Y_i)) + Var(N_t \mathbb{E}(Y_i)) \\
&= Var(Y_i)\mathbb{E}(N_t) + (\mathbb{E}[Y_i])^2 Var(N_t) \\
&= \lambda t Var(Y_1) + \lambda t(\mathbb{E}[Y_1])^2 \\
&= \boxed{\lambda t \mathbb{E}[Y_1^2]} = 1164
\end{aligned}
$$

Note that an important formual here is

$$\mathbb{E}[X_t] = \lambda t \mathbb{E}[Y_1]$$

Thinning of Poisson Process. We classifcy each event into one of k different types. The classification are independently done, and independent with respect to the time it happens at probability p_i for type i. Then, $N_i(t)$ is the number of events of type i up to time t.

Theorem.

$$N_i \sim Pois(\lambda p_i), \text{ all i.i.d.}$$

Example. Ellen catches fish according to poisson process with a rate of 2 fish per hour. $p = 0.4$ of fish are salmon and the rest of trouts. What is the probability she catches exactly 1 salmon and 2 trouts in 2.5 hours.

$$\mathbb{P}(N_S(2.5) = 1, N_T(2.5) = 2) = \mathbb{P}(N_S(2.5 = 1)) \cdot \mathbb{P}(N_T(2.5 = 2)) = e^{-2\times 0.4\times 2.5}\frac{2\times 0.4\times 2.5}{1} \times ... =$$

Rewording for thinning or possion process: Suppose we have poisson process with arrival rate λ and we thinned the arrival times so that T_i is kept with probability p and T_i is erased with probability $1-p$. The counting process of the remaining arrival times is also a poisson process with rate px

Then, $N_1(t), N_2(t), ..., N_j(t)$ are each a poisson process with respective rate $\lambda p_1, \lambda p_2, ..., \lambda p_n$ and they are independent of each other.

Example. Let probability of boy being born be 0.519. Assume births occur according to a poisson process with rate 2 births/hhour. On an 8 hour shift, what is the expectation and standard deviation on the number of female births?

$N_{\text{female}}(t) \sim PP(2 \times (1 - 0.519))$, $N_{\text{female}}(8) \sim Pois(8 \times 2 \times (1 - 0.519))$. $Var(N_f(8)) = 2.774$.

Example. From the example above, find the probability that only girls were born between 2-5 PM.

$\mathbb{P}(N_{\text{female}}(t) > 0, N_{\text{male}}(3) = 0) = \mathbb{P}(N_{\text{female}}(t) > 0)\mathbb{P}(N_{\text{male}}(3) = 0) = ... = 0.042$

Example. Assume 5 babies were born in 1 hour. Find the probability that two are boys.

$$5C2 \times (0.519)^2(1 - 0.519)^3$$

[Generalized Thinning Method] Consider a $N \sim PP(\lambda)$. Suppose $T_1, ...T_n$ are arrival times. Suppose we keep T_i with probability $p(T_i)$ or erase T_i with probability $1 - p_i$. If we get $\tilde{N}_t$ be the number of events by time t, then $\tilde{N}_t$ is a nonhomogeneous poisson process with rate function $\lambda p(t), t \geq 0$.

The intuition here is that

$$
\begin{aligned}
\mathbb{P}[\tilde{N}_{t+h} - \tilde{N} = 1] &= \mathbb{P}[N_{t+h} = 1 \text{ and keep it}] + o(h^2) \\
&= \mathbb{P}[N_{t+h} - N_t = 1]\mathbb{P}[\text{keep}] + o(h^2) \\
&= \lambda p(t)h + o(h^2)
\end{aligned}
$$

Example. Calls arrive in the call center according to Poisson process with rate of S per hour. Each call can last a random time with a uniform distribution over $(0, \frac{1}{6})$. What is the probability of having a call still in progress at time $t = 2$ hours, assuming no calls active at time 0?

4 Continuous Time Markov Chains

In continuous time, we can define markov chains as

$$\begin{aligned} P(X_{s+t} = j | X_s = i, ..., X_{s_m} = i_m) &= P(X_{s+t} = j | X_s = i) \\ &= P(X_t = j | X_0 = i) \end{aligned}$$

written as $\{X_t\}_{t \geq 0}$.

Definition. The **transition probability** is given by $p_{i,j}(t) := P(X_t = j | X_0 = i)$.

$$P(t) = \begin{bmatrix} p_{1,1}(t) & \dots & p_{1,N}(t) \\ \vdots & \ddots & \vdots \\ p_{N,1}(t) & \dots & p_{N,N}(t) \end{bmatrix}$$

Example. $\{N_t\}_{t \geq 0}$ is a homogenenous poisson process with rate λ is a CTMC with $p_{i,j}(t) = e^{\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!}$

Proposition. Suppose $\{N_t\}$ is a poisson process with rate λ and $\{Y_n\}_{n=0,1...}$ is an independent discrete time MC with transition probabilities $p_{i,j} = P(Y_1 = 1 \mid Y_0 = 1)$. Let

$$X_t = Y_{N_t} \begin{cases} Y_0, & 0 \leq t < T_1 \\ Y_1 & T_1 \leq t < T_2 \\ & \vdots \end{cases}$$

Then, $\{X_t\}_{t \geq 0}$ is a CTMC.

Proof.

$$\mathbb{P}[X_{t+s} = j \mid X_s = i] = \frac{\mathbb{P}[X_{t+s} = j, X_s = i]}{\mathbb{P}(X_s = i)} = \frac{\mathbb{P}[Y_{N_{t+s}} = j, Y_{N_s} = i]}{\mathbb{P}(X_s = i)}$$

Here,

$$\begin{aligned} \mathbb{P}(Y_{N_{s+t}} = j, Y_{N_s} = i) &= \sum_{l,m} \mathbb{P}[\underbrace{Y_{N_{t+s}}}_{=Y_{l+m}} = j, Y_l = i \mid N_s = l, N_{t+s} - N_s = m] \mathbb{P}(N_s = l) e^{-\lambda t} \frac{(\lambda t)^m}{m!} \\ &= \sum_{l,m} \mathbb{P}[Y_{l+m}, Y_l = i] \mathbb{P}[N_s = l] e^{-\lambda t} \frac{(\lambda t)^m}{m!} \\ &= \sum_{l,m} P_{i,j}^m \mathbb{P}(Y_l = i) \mathbb{P}(N_s = l) e^{-\lambda t} \frac{(\lambda t)^m}{m!} \\ &= \sum_{l,m} P_{i,j}^m \mathbb{P}(X_s = i, N_s = l) e^{-\lambda t} \frac{(\lambda t)^m}{m!} \\ &= \sum_m P_{i,j}^m e^{-\lambda t} \frac{(\lambda t)^m}{m!} \underbrace{\sum_l \mathbb{P}(X_s = i, N_s = l)}_{\mathbb{P}(X_s = i)} \end{aligned}$$

Then,

$$P(t) = \sum_{m=0}^{\infty} e^{\lambda t} \frac{(\lambda t)^m}{m!} []$$

■

Properties:

- $e^0 = I$
- $e^{cI} = e^c I$
- $e^A = \sum_{m=0}^{\infty} \frac{A^m}{m!}$
- $e^A \times e^B = e^{A+B}$ if $AB = BA$
- $e^{UAU^{-1}} = Ue^AU^{-1}$

Following the proof, we have

$$\sum_{m=0}^{\infty} e^{\lambda t} \frac{(\lambda t)^m}{m!} P^m = e^{-\lambda t} \sum_{m=0}^{\infty} \frac{(\lambda t)^m P^m}{m!} = e^{-\lambda t} e^{\lambda t P} = e^{\lambda t} I e^{\lambda t P} = e^{\lambda t I} e^{\lambda t P} = e^{(P-I)\lambda t}$$

Here, $Q := P - I$ is called the **infinitesimmal generator** of X.

Preview: Any CTMC is of the form Y_{N_t}, where $\{Y_n\}$ is DTMC and $\{N_t\}_{t\geq 0}$ is a poisson process.

$$X_t = \begin{cases} Y_0, & 0 \leq t \leq T_1 \sim exp(\lambda_{y_0}) \\ Y_1, & T_1 \leq t < T_1 + exp(\lambda_{y_1}) := T_2 \\ & \vdots \end{cases}$$

Claim: $\{X_t\}_{t\geq 0}$ is a CTMC with transition matrix $P(t) = e^{tQ}$, where Q is

$$Q_{i,j} = \begin{cases} -\lambda_i, & i = j \\ \lambda_i r(i,j), & i \neq j \end{cases}$$

With eigendecomposition $Q = U\Lambda U^{-1}$,

$$e^{tQ} = Ue^{t\Lambda}U^{-1}$$

4.1 Stationary Distribution

Definition. α is called a stationary distribution if $\alpha_t = \alpha \forall t$ or $\alpha P(t) = \alpha \forall t$.

Proposition. π is a stationary distribution $\iff \pi Q = 0$

Proof. $\Longleftarrow$: Suppose $\pi Q = 0$. $\pi P(t) = \pi e^{tQ} \implies (\pi P(t))' = \pi Q e^{tQ} = 0$. Thus $\pi P(t)$ is a constant $= \pi' \implies \pi P(0) = \pi$. ■

Proposition. Limiting Distribution is a stationary distribution.