Introduction to Stochastic Processes, Fall 2019

随机过程引论笔记

何憾[1]

版本 1.0.0 Beta

[1]访问我的个人主页, 查看其他笔记

Contents

Chapter 1

Discrete-time Markov chains

In this note, we denote by $\mathbb{N}$ all the non-negative integers, i.e., $\mathbb{N} = \{0, 1, 2, \cdots\}$. We work throughout with a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

1.1 Introductions

1.1.1 Definition and basic properties

Let I be a countable set. Each $i \in I$ is called a state and I is called the state-space. We say that $\lambda = (\lambda_i)_{i \in I}$ is a *measure* on I if $0 \leqslant \lambda_i < \infty$ for all $i \in I$. If in addition the total mass $\sum_{i \in I} \lambda_i$ equals 1, then we call λ a *distribution*.

Recall that a random variable X with values in I is a function $X : \Omega \to I$. If we set

$$\lambda_i = \mathbb{P}(X = i) = \text{ for each } i \in I$$

Then λ defines a *distribution*, the distribution of X. We think of X as modelling a random state which takes the value i with probability λ_i.

We say that a matrix $P = (p_{ij})_{i,j \in I}$ is **stochastic** if every row $(p_{ij}) :_{j \in I}$ is a distribution. We shall now formalize the rules for a Markov chain by a definition in terms of the corresponding matrix P.

We say $(X_n)_{n\geqslant 0}$ is Markov chain with **initial distribution** λ and **transition matrix** P, if X_0 has distribution λ and for any $n\in\mathbb{N}$, conditional on $X_n=i$, X_{n+1} has distribution $(p_{ij})_{j\in I}$ and is independent of $X_0,\cdots,X_{n-1}$. More explicitly, these conditions state that

(i) for any $i\in I$, $\mathbb{P}(X_0=i)=\lambda_i$.

(ii) for any $n\geqslant 1$ and $j,i,i_0,\cdots,i_{n-1}$ in I.

$$\begin{aligned}&\mathbb{P}(X_{n+1}=j \mid X_n=i, X_0=i_0,\cdots,X_{n-1}=i_{n-1})\\ &\quad=\mathbb{P}(X_{n+1}=j\mid X_n=i)=\ p_{ij}\,.\end{aligned}$$

We say that $(X_n)_{n\geqslant 0}$ is Markov(λ,P) for short. In formulating (ii) we have restricted our attention to the **temporally homogeneous** case in which the transition probability $\mathbb{P}(X_{n+1}=j\mid X_n=i)$ does not depend on the time n.

It is in terms of properties (i) and (ii) that most real-world examples are seen to be Markov chains. But mathematically the following result appears to give a more comprehensive description, and it is the key to some later calculations.

Theorem 1.1. *A discrete-time random process $(X_n)_{n\geqslant 0}$ is Markov(λ,P) if and only if for any $n\geqslant 1$ and $i_1,\cdots,i_n\in I$,*

$$\mathbb{P}(X_0=i_1,X_1=i_2,\cdots,X_n=i_n)=\lambda_{i_1}p_{i_1i_2}p_{i_2i_2}\cdots p_{i_{n-1}i_n}\,.$$

Proof. Note that for any $n\geqslant 1$ and $i_0,\cdots,i_{n+1}$ in I,

$$\begin{aligned}&\mathbb{P}(X_{n+1}=i_{n+1}\mid X_0=i_1,\cdots,X_n=i_n)\\ &=\frac{\mathbb{P}(X_0=i_1,\cdots,X_n=i_n,X_{n+1}=i_{n+1})}{\mathbb{P}(X_0=i_1,\cdots,X_n=i_n)}=p_{i_ni_{n+1}}\,.\quad\square\end{aligned}$$

The next result reinforces the idea that Markov chains have no memory. We write $\delta_i=(\delta_{ij})_{j\in I}$ for the unit mass at i, where

$$\delta_{ij}=\begin{cases}1 & \text{if } i=j\\ 0 & \text{otherwise.}\end{cases}$$

Theorem 1.2 (Markov property). *Let $(X_n)_{n\geqslant 0}$ be Markov(λ, P). Then, conditional on $X_m = i$, $(X_{m+n})_{n\geqslant 0}$ is Markov(δ_i, P), and is independent of the random variables $X_0, \cdots, X_m$.*

Proof. It usffices to show that for any event $A \in \sigma(X_0, \cdots, X_m)$, and $i_m, \cdots, i_{m+n}$ in I,

$$\begin{aligned}\mathbb{P}\left(\{X_m\right. &= i_m, \cdots, X_{m+n} = i_{m+n}\} \cap A \mid X_m = i) \\ &= \delta_{ii_m} p_{i_m i_{m+1}} \cdots p_{i_{m+n-1} i_{m+n}} \mathbb{P}\left(A \mid X_m = i\right)\end{aligned} \tag{1.1}$$

Since $\cup_{n=1}^{\infty} \sigma(X_m, \cdots, X_{m+n})$ is a π-systerm generating $\sigma\left(X_{m+n}\right)_{n\geqslant 0}$, (1.1) implies the desired result. By the same reason, we only need to consider the case of elementary events

$$A = \{X_0 = i_1, \cdots, X_m = i_m\} \ .$$

Note that

$$\begin{aligned}\mathbb{P}\left(X_0\right. &= i_1, \cdots, X_{m+n} = i_{m+n}) \\ &= \delta_{ii_m} p_{i_m i_{m+1}} \cdots p_{i_{m+n-1} i_{m+n}} \times \mathbb{P}\left(X_0 = i_1, \cdots, X_m = i_m\right) \ ,\end{aligned}$$

Which follows from Theorem 1.1. □

1.1.2 Multistep transition probabilities

Assume $(X_n)_{n\geqslant 0}$ is Markov chain with transition matrix P. The transition probability $p_{i,j} = P\left(X_{n+1} = j \mid X_n = i\right)$ gives the probability of going from i to j in one step. Our goal in this section is to compute the probability of going from i to j in n steps:

$$p_{ij}^{(n)} = \mathbb{P}\left(X_{n+m} = j \mid X_m = i\right)$$

By Markov property, the definition above is independent of m. We denote $P^{(n)} := \left(p_{ij}^{(n)}\right)_{i,j\in I}$, and call it the n *step transition matrix.*

Proposition 1.3 (Chapman-Kolmogorov equation). *For any $n, m \in \mathbb{N}_+$, $P^{(m+n)} = P^{(m)} P^{(n)}$. In other words, for any $i, j \in I$,*

$$p_{ij}^{(m+n)} = \sum_{k \in I} p_{ik}^{(m)} p_{kj}^{(n)} .$$

Proof. We do this by breaking things down according to the state at time m.

$$\mathbb{P}_i (X_{m+n} = j) = \sum_{k \in I} \mathbb{P}_i(X_m = k) \mathbb{P}_i (X_{m+n} = j \mid X_m = k)$$

By Markov property, $\mathbb{P}_i (X_{m+n} = j \mid X_m = k) = p_{kj}^{(n)}$, thus we get the C-K equation. □

Note that $P^{(1)} = P$, by induction we get:

Corollary 1.4. *For all $n \in \mathbb{N}_+$, $P^{(n)} = P^n$. In other words, for any $i, j \in I$,*

$$p_{ij}^{(n)} = \sum_{i_1, \cdots, i_{n-1} \in I} p_{ii_1} p_{i_1 i_2} \cdots p_{i_{n-1} j} .$$

Corollary 1.5. *Let $(X_n)_{n \geqslant 0}$ be Markov(λ, P), then for any n the distribution of X_n is λP^n, where we regard distributions (and measures) as row vectors whose components are indexed by I.*

1.1.3 Class structure

It is sometimes possible to break a Markov chain into smaller pieces, each of which is relatively easy to understand, and which together give an understanding of the whole. This is done by identifying the communicating classes of the chain.

Let $(X_n)_{n \geqslant 0}$ be a Markov chain with transition matrix P. Define the **hitting time** of state j is a random variable $\Omega \to \{0, 1, 2, \cdots\} \cup \{\infty\}$ and given by

$$\tau_j \coloneqq \inf\{n \geqslant 0 : X_n = j\} .$$

where we agree that the infimum of the empty set $\varnothing$ is ∞.

We say that i **leads to** j and write $i \to j$ if

$$\mathbb{P}_i\left(\tau_j < \infty\right) > 0\,.$$

We say i **communicates with** j and write $i \leftrightarrow j$ if both $i \to j$ and $j \to i$.

Proposition 1.6. *For distinct states i and j, i leads to j is and only if $p_{ij}^{(n)} > 0$ for some $n \geqslant 0$.*

Proof. Observe that

$$p_{ij}^{(n)} \leqslant \mathbb{P}_i\left(\tau_j < \infty\right) \leqslant \sum_{n=0}^{\infty} p_{ij}^{(n)}\,. \qquad \square$$

Clearly, from Proposition 1.6 we see that $i \to j$ and $j \to k$ imply $i \to k$. Also $i \to i$ forany state i. So "$\leftrightarrow$" satisfies the conditions for an *equivalence relation* on I, and thus partitions I into **communicating classes**.

Definition 1.1. We say that a nonempty set $A \subset I$ is **closed** if it is impossible to get out, i.e., for any $i \in A$, $i \to j$ implies that $j \in A$. A state $i \in I$ is **absorbing** if $\{i\}$ is a closed class.

Remark. It's easy to see that $A \subset I$ be closed, then for any communicating class C, either $C \subset A$ or $C \cap A = \varnothing$.

Definition 1.2. A chain or transition matrix P is called **irreducible**, if I is a single communicating class, i.e., any two state communicate with each other.

Proposition 1.7. *I is irreducible if and only if all the nonempty proper subset of I is not closed.*

Proof. If $A \subsetneqq I$ is closed, pick $i \in I\backslash A$ and $j \in A$, then $[i] \cap [j] = \varnothing$. I has at least two communicating class, so is reducible.

If I is irreducible, for any nonempty proper subset A. Pick any $i \notin A$, $j \notin A$, since I is irreducible, i leads to j. Thus A is not closed. $\square$

The advantage of closed set is that we can reduce the state sapce. Suppose A is closed, then for all $i \in A$,

$$\mathbb{P}_i\left(X_n \in A, \forall\, n \in \mathbb{N}\right) = 1\,.$$

Thus $P|_A = (p_{ij})_{i,j\in A}$ is a transition matrix. If $\mathbb{P}(X_0 \in A) = 1$, then, in fact, $(X_n)_{n\geqslant 0}$ is a Markov chain with transition matrix $P|_A$. In addition, if A is a class, then $P|_A$ is irrducible.

EXERCISE

¶ EXERCISE 1.1. Show that every transition matrix on a finite state-space has at least one closed communicating class. Find an example of a transition matrix with no closed communicating class.

1.2 First step analysis

Let $(X_n)_{n\geqslant 0}$ be a Markov chain with transition matrix P. The **hitting time** of a subset A of I is the random variable $\tau_A : \Omega \to \{0, 1, 2, \cdots\} \cup \{\infty\}$ given by

$$\tau_A = \inf\{n \geqslant 0 : X_n \in A\} ,$$

where we agree that the infimum of the empty set $\varnothing$ is ∞. The probability starting from i that $(X_n)_{n\geqslant 0}$ ever hits A is then

$$h_i = \mathbb{P}_i\left(\tau_A < \infty\right) .$$

When A is a closed class, h_i is called the absorption probability. The mean time taken for $(X_n)_{n\geqslant 0}$ to reach A is given by

$$t_i = \mathbb{E}_i\left(\tau_A\right) = \sum_{n=0}^{\infty} n\mathbb{P}_i\left(\tau_A = n\right) + \infty\mathbb{P}_i\left(\tau_A = \infty\right) .$$

We shall often write less formally

$$h_i = \mathbb{P}_i(\text{ hit } A),\ t_i = \mathbb{E}_i(\text{ time to hit } A) .$$

1.2.1 Hitting probabilities

Theorem 1.8. *The vector of hitting probabilities $(h_i)_{i\in I}$ is the minimal non-negative solution to the system of linear equations*

$$\begin{cases} h_i = 1, & \text{for } i \in A. \\ h_i = \sum_{j\in I} p_{ij} h_j, & \text{for } i \notin A. \end{cases} \tag{1.2}$$

(Minimality means that if $x = (x_i)_{i\in I}$ is another solution with $x_i \geqslant 0$ for all i, then $x_i \geqslant h_i$ for all i.)

Proof. First we show that $(h_i)_{i\in I}$ satisfies (1.2). Clearly, if $X_0 = i \in A$, then $\tau_A = 0$, so $h_i = 1$.

If $X_0 = i \notin A$, then $\tau_A \geqslant 1$,

$$h_i = \mathbb{P}_i\left(\tau_A < \infty\right) = \sum_{j \in I} p_{ij}\, \mathbb{P}_i\left(\tau_A < \infty \mid X_1 = j\right) \,,$$

by the Markov property

$$\mathbb{P}_i\left(\tau_A < \infty \mid X_1 = j\right) = \mathbb{P}_i\left(\tau_A - 1 < \infty \mid X_1 = j\right) = \mathbb{P}_j\left(\tau_A < \infty\right) = h_j \,.$$

Thus when $i \notin A$,

$$h_i = \sum_{j \in I} p_{ij}\, h_j \,.$$

This method is called sometimes *first step analysis*.

Suppose now that $x = (x_i)_{i \in I}$ is any solution to (1.2). Then $h_i = x_i = 1$ for $i \in A$. For $i \notin A$,

$$x_i = \sum_{j \in I} p_{ij} x_j = \sum_{j \in A} p_{ij} + \sum_{j \notin A} p_{ij} x_j$$

Substitute for x_j to obtain

$$\begin{aligned} x_i &= \sum_{j \in A} p_{ij} + \sum_{j \notin A} p_{ij} \left(\sum_{k \in A} p_{jk} + \sum_{k \notin A} p_{jk} x_k \right) \\ &= \mathbb{P}_i\left(X_1 \in A\right) + \mathbb{P}_i\left(X_1 \notin A, X_2 \in A\right) + \sum_{j \notin A} \sum_{k \notin A} p_{ij} p_{jk} x_k \,. \end{aligned}$$

By repeated substitution for x in the final term we obtain after n steps

$$\begin{aligned} x_i = \mathbb{P}_i\left(X_1 \in A\right) + \cdots + \mathbb{P}_i\left(X_1 \notin A, \cdots, X_{n-1} \notin A, X_n \in A\right) \\ + \sum_{j_1 \notin A} \cdots \sum_{j_n \notin A} p_{ij_1} p_{j_1 j_2} \cdots, p_{j_{n-1} j_n} x_{j_n} \,. \end{aligned}$$

Now if x is non-negative, so is the last term on the right, and the remaining terms sum to $\mathbb{P}_i\left(\tau_A \leqslant n\right)$. So $x_i \geqslant \mathbb{P}_i\left(\tau_A \leqslant n\right)$ for all n and then

$$x_i \geqslant \lim_{n \to \infty} \mathbb{P}_i\left(\tau_A \leqslant n\right) = \mathbb{P}_i\left(\tau_A < \infty\right) = h_i. \qquad \square$$

Remark. In fact, using first step analysis, we have

$$\sum_{k\in I} p_{ik}\mathbb{P}_k(\tau_A < \infty) = \mathbb{P}_i(\sigma_A < \infty), \quad \text{for all } i \in I.$$

where $\sigma_A := \inf\{n \geqslant 1 : X_n \in A\}$, is called **the first passage time of** A.

¶ Example 1.2 (Birth-death chain). Consider the Markov chain with diagram in Figure 1.1 where, for $i = 1, 2, \cdots$, we have $0 < b_i = 1 - d_i < 1$.

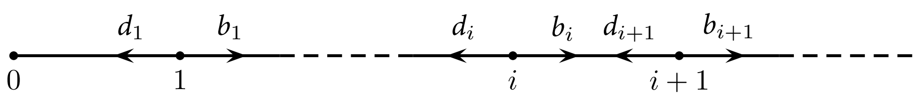


Figure 1.1: brith-death chain

Let $p_{01} = 1$. we wish to calculate the absorption probability starting from i. Such a chain may serve as a model for the size of a population, recorded each time it changes, b_i being the probability that we get a birth before a death in a population of size i. Then $h_i = \mathbb{P}_i(\text{ hit } 0)$ is the *extinction probability* starting from i.

We write down the usual system of equations

$$\begin{aligned} h_0 &= 1 \\ h_i &= b_i h_{i+1} + d_i h_{i-1}, \qquad \text{for } i = 1, 2, \cdots \end{aligned}$$

This recurrence relation has variable coefficients so the usual technique fails. But consider $u_i = h_{i-1} - h_i$, then $b_i u_{i+1} = d_i u_i$, so

$$u_{i+1} = \left(\frac{d_i}{b_i}\right) u_i = \left(\frac{d_i d_{i-1} \cdots d_1}{b_i b_{i-1} \cdots b_1}\right) u_1 = \gamma_i \, u_1$$

where the final equality defines γ_i for $i \geqslant 1$. Then $u_1 + \cdots + u_i = h_0 - h_i$, so

$$h_i = 1 - u_1 \left(\gamma_0 + \cdots + \gamma_{i-1}\right) = 1 - R_{i-1}$$

where $\gamma_0 = 1$, and where the final equality defines R_{i-1}. At this point u_1 remains to be determined.

(i) In the case $R := \sum_{i=0}^{\infty} \gamma_i = \infty$, the restriction $0 \leqslant h_i \leqslant 1$ forces $u_1 = 0$ and $h_i = 1$ for all $i \geqslant 0$.

(ii) But if $R = \sum_{i=0}^{\infty} \gamma_i < \infty$, then we can take $u_1 > 0$ so long as $h_i \geqslant 0$, that is $u_1 \leqslant \frac{1}{R}$. Thus the minimal non-negative solution occurs when $u_1 = \frac{1}{R}$ and then
$$h_i = 1 - \frac{R_i}{R}$$
In this case, for $i = 1, 2, \cdots$, we have $h_i < 1$, so the population survives with positive probability.

¶ Example 1.3 (Gamblers' ruin). Consider the Markov chain with diagram in Figure 1.2, where $0 < p = 1 - q < 1$. The transition probabilities are
$$p_{i,i-1} = q, p_{i,i+1} = p \quad \text{for } i = 1, 2, \cdots$$
Imagine that you enter a casino with a fortune of £1 and gamble, £1 at a time, with probability p of doubling your stake and probability q of losing it. The resources of the casino are regarded as infinite, so there is no upper limit to your fortune. But what is the probability that you leave broke?

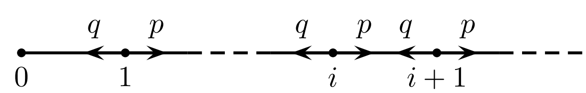


Figure 1.2: gamblers' ruin

Set $h_i = \mathbb{P}_i(\text{ hit } 0)$, then h is the minimal non-negative solution to
$$h_0 = 1$$
$$h_i = ph_{i+1} + qh_{i-1}, \text{ for } i = 1, 2, \cdots$$
Now we can see this model is particular case of brith-death chain(We dont's care the transition probability in state 0). So by Example 1.2,
$$R = \sum_{i=0}^{\infty} \left(\frac{q}{p}\right)^i.$$

(i) If $p > q$, the solution is $h_i = (\frac{q}{p})^i$. for all $i \geqslant 0$.

(ii) If $p < q$, which is the case in most successful casinos, we have $h_i = 1$ for all i. Even if $p = q$, i.e., you find a *fair* casino, you are certain to end up broke (But the mean breaking time is infinity, which we will see later). This apparent paradox is called *gamblers' ruin*.

¶ Example 1.4 (带时滞的生灭链). 考虑 $\mathbb{N}$ 上的马氏链，给定 $c \geqslant 0$, $\alpha \geqslant 0$. 令 $p_{01} = 1$, 对任意 $i \geqslant 1$,

$$p_{i,i+1} = \frac{1}{2}, \quad p_{i,i-1} = \frac{\exp\left(-ci^{-\alpha}\right)}{2}, \quad p_{ii} = \frac{1 - \exp\left(-ci^{-\alpha}\right)}{2}$$

我们来计算这个带有“时滞”的生灭链的灭绝概率. 令 $h_i := \mathbb{P}_i(\tau_0 < \infty)$, 则

$$h_i = p_{i,i+1}h_{i+1} + p_{i,i-1}h_{i-1} + p_{ii}h_i$$

等式两边同时减去 h_i, 得到

$$h_i - h_{i+1} = \frac{p_{i,i-1}}{p_{i,i+1}}\left(h_{i-1} - h_i\right)$$

将 $p_{i,i+1}$ 看为 b_i, 将 $p_{i,i-1}$ 视为 d_i, 则从生灭链的灭绝概率计算中我们知道, h_i 全为 1 当且仅当 $R = \infty$, 其中

$$R = 1 + \sum_{k=1}^{\infty} \exp\left(-c\sum_{j=1}^{k} j^{-\alpha}\right).$$

1.2.2 Mean hitting times

Theorem 1.9. *The vector of mean hitting times $(t_i)_{i\in I}$ is the minimal non-negative solution to the system of linear equations*

$$\begin{cases} t_i = 0, & \text{for } i \in A. \\ t_i = 1 + \sum_{j\notin A} p_{ij}t_j, & \text{for } i \notin A. \end{cases} \tag{1.3}$$

Proof. First we show that $(t_i)_{i\in I}$ satisfies (1.3). If $X_0 = i \in A$, then $\tau_A = 0$, so $t_i = 0$.

If $X_0 = i \notin A$, then $\tau_A \geqslant 1$, so, by the Markov property,

$$\mathbb{E}_i\left(\tau_A \mid X_1 = j\right) = 1 + \mathbb{E}_i\left(\tau_A - 1 \mid X_1 = j\right) = 1 + \mathbb{E}_j\left(\tau_A\right),$$

and

$$\begin{aligned} t_i = \mathbb{E}_i\left(\tau_A\right) &= \sum_{j \in I} \mathbb{E}_i\left(\tau_A \mid X_1 = j\right) \\ &= \mathbb{P}_i\left(X_1 = j\right) = 1 + \sum_{j \notin A} p_{ij} t_j\,. \end{aligned}$$

Suppose now that $y = (y_i)_{i \in I}$ is any solution to (1.3). Then $t_i = y_i = 0$ for $i \in A$. For $i \notin A$,

$$\begin{aligned} y_i &= 1 + \sum_{j \notin A} p_{ij} y_j \\ &= 1 + \sum_{j \notin A} p_{ij}\left(1 + \sum_{k \notin A} p_{jk} y_k\right) \\ &= \mathbb{P}_i\left(\tau_A \geqslant 1\right) + \mathbb{P}_i\left(\tau_A \geqslant 2\right) + \sum_{j \notin A} \sum_{k \notin A} p_{ij} p_{jk} y_k \end{aligned}$$

By repeated substitution for y in the final term we obtain after n steps

$$y_i = \mathbb{P}_i\left(\tau_A \geqslant 1\right) + \cdots + \mathbb{P}_i\left(\tau_A \geqslant n\right) + \sum_{j_1 \notin A} \cdots \sum_{j_n \notin A} p_{ij_1} p_{j_1 j_2} \cdots p_{j_{n-1} j_n} y_{j_n}$$

So, if y is non-negative,

$$y_i \geqslant \mathbb{P}_i\left(\tau_A \geqslant 1\right) + \cdots + \mathbb{P}_i\left(\tau_A \geqslant n\right)$$

and, letting $n \to \infty$

$$y_i \geqslant \sum_{n=1}^{\infty} \mathbb{P}_i\left(\tau_A \geqslant n\right) = \mathbb{E}_i\left(\tau_A\right) = t_i \qquad \square$$

Remark. In fact, using first step analysis, we find that

$$1 + \sum_{k \in I} p_{ik} \mathbb{E}_k(\tau_A) = \mathbb{E}_i(\sigma_A), \quad \text{for all } i \in I\,,$$

where $\sigma_A \coloneqq \inf\{n \geqslant 1 : X_n \in A\}$.

EXERCISE

¶ EXERCISE 1.5. Let $A, B \subset I$ and $A \cap B = \varnothing$. Let x_i be the probability starting from i that $(X_n)_{n \geqslant 0}$ hits A before hitting B. In other words,

$$x_i := \mathbb{P}_i(\tau_A < \tau_B)\,, \text{ for any } i \in I\,.$$

(i) Sohw that

$$\sum_{j \in I} p_{ij} \mathbb{P}_i(\tau_A < \tau_B) = \mathbb{P}_i(\sigma_A < \sigma_B)\,.$$

(ii) Show that $(x_i)_{i \in I}$ is the minimal non-negative solution to the system of linear equations

$$\begin{cases} x_i = 1, & \text{for } i \in A. \\ x_i = 0, & \text{for } i \in B. \\ x_i = \sum_{j \in I} p_{ij} x_j, & \text{for } i \notin A, B. \end{cases}$$

¶ EXERCISE 1.6. Let $G_{ij} = \mathbb{E}_i(V_j)$. (The definition of V_j is given by (1.4.2))

(i) Using first step analysis, show that for any j is fixed, $(G_{ij})_{i \in I}$ is the minimal non-negative solution to the system of linear equations

$$G_{ij} = \sum_{k \in I} p_{ik} G_{kj} + \delta_{ij}\,.$$

(ii)* Using *cycle trick*, show that for i is fixed, $(G_{ij})_{j \in I}$ is the solution to the system of linear equations

$$G_{ij} = \sum_{k \in I} G_{ik} p_{kj} + \delta_{ij}\,.$$

In fact, that is $G = PG + I = GP + I$, or $G = (I - P)^{-1}$. We can see this from $G = \sum_{n=0}^{\infty} P^n$.

¶ Exercise 1.7. $A \subset I$. Let

$$G_{ij}^A \coloneqq \mathbb{E}_i \left(\sum_{0 \leqslant n < \tau_A} 1_{\{X_n = j\}} \right)$$

(称为区域格林函数, 见 Oct 8 课堂笔记). Show that

$$\begin{cases} G_{ij}^A = 1, & \text{for } i \in A \text{ or } j \in A. \\ G_{ij}^A = \sum_{k \notin A} p_{ik}\, G_{kj}^A + \delta_{ij} & \\ \qquad = \sum_{k \notin A} G_{ik}^A\, p_{kj} + \delta_{ij}, & \text{for } i, j \notin A. \end{cases} \tag{1.4}$$

In fact that is $G^A|_{A^c} = G^A|_{A^c} P|_{A^c} + \mathrm{I}|_{A^c} = P|_{A^c} G^A|_{A^c} + \mathrm{I}|_{A^c}$. Then $G^A|_{A^c} = (\mathrm{I}|_{A^c} - P|_{A^c})^{-1}$. We can see this from $G^A|_{A^c} = \sum_{n=0}^{\infty} (P|_{A^c})^n$.

1.3 Strong Markov property

In Section 1.1 we proved the Markov property: for any time m, conditional on $X_m = i$, the process after time m begins *afresh* from i. Suppose, instead of conditioning on $X_m = i$, we simply waited for the process to hit state i, at the random time τ_i. What can one say about the process after time τ_i? What if we replaced τ_i by a more general random time, for example $\tau_i - 1$? In this section we shall identify a class of random times at which a version of the Markov property does hold. This class will include τ_i but not $\tau_i - 1$, after all, the process after time $\tau_i - 1$ jumps straight to i, so it does not simply begin afresh.

Definition 1.3. A random variable $\tau : \Omega \to \mathbb{N} \cup \{\infty\}$ is called a **stopping time** if the event $\{T \leqslant n\} \in \sigma(X_0, \cdots, X_n)$ for any $n \in \mathbb{N}_+$.

Obviously, $\{T \leqslant n\} \in \sigma(X_0, \cdots, X_n)$ can be replaced by $\{T = n\} \in \sigma(X_0, \cdots, X_n)$ for any $n \in \mathbb{N}_+$. Intuitively, τ is a stopping time if by watching the process, you know at the time when τ occurs. If asked to stop at τ, you know when to stop.

¶ Example 1.8.

(i) The first hitting time τ_A is a stopping time because

$$\{\tau_A = n\} = \{X_0 \notin A, \ldots, X_{n-1} \notin A, X_n \in A\}$$

(ii) The first passage time σ_A is a stopping time because

$$\{\sigma_A = n\} = \{X_1 \notin A, \ldots, X_{n-1} \notin A, X_n = j\}$$

(iii) The *last exit time* of $A \subset I$

$$L_A = \sup\{n \geq 0 : X_n \in A\}$$

is not in general a stopping time, because the event $\left\{L^A = n\right\}$ depends on whether $(X_{n+m})_{m\geq 1}$ visits A or not.

We shall show that the Markov property holds at stopping times τ. That is, if $\tau < \infty$ and we konw the state of the chain when it stops, i.e., $X_\tau = i$, then the history $(X_0, \cdots, X_\tau)$ and the future $(X_{\tau+n})_{n\geqslant 0}$ are independent, and the future $(X_{\tau+n})_{n\geqslant 0}$ is Markov chain starting at i with the previous tansition matrix.

On the event $\{\tau < \infty\}$, we define $(X_0, \cdots, X_\tau) = (X_0, \cdots, X_n)$ when $\{\tau = n\}$, for all $n \geqslant 0$. Since $\{\tau = n\} \in \sigma(X_0, \cdots, X_n)$, it is well-defined. We regard $(X_0, \cdots, X_\tau)$ as an random orbit in $\cup_{n\geqslant 0} I^{n+1}$ equipped with a σ-algebra consisting of all the subset of $\cup_{n\geqslant 0} I^{n+1}$. The first problem is, What are the events determined by $(X_0, \cdots, X_\tau)$, i.e., what is $\sigma(X_0, \cdots, X_\tau)$?

For any $A \subset \cup_{n\geqslant 0} I^{n+1}$, let $A_n = A \cap I^{n+1}$, then $A = \cup_{n\geqslant 0} A_n$. Then

$$\{(X_0, \cdots, X_\tau) \in A\} = \biguplus_{n\geqslant 0} \{(X_0, \cdots, X_n) \in A_n\} \cap \{\tau = n\}\,.$$

Let $\mathcal{F}_n = \sigma(X_0, \cdots, X_n)$ for all $n \geqslant 0$. Thus we have

$$\sigma(X_0, \cdots, X_\tau) = \{B \in \mathcal{F} : B \cap \{\tau = n\} \in \mathcal{F}_n \text{ for all } n \geqslant 0\}\,. \tag{1.5}$$

Theorem 1.10 (Strong Markov property). *Let $(X_n)_{n\geq 0}$ be Markov(λ, P) and let τ be a stopping time of $(X_n)_{n\geq 0}$. Then, conditional on $\tau < \infty$ and $X_\tau = i$, $(X_{\tau+n})_{n\geq 0}$ is Markov(δ_i, P) and independent of $(X_0, X_1, \ldots, X_\tau)$*

Proof. We only need to prove that for any $B \in \sigma(X_0, X_1, \ldots, X_\tau)$,

$$\begin{aligned} &\mathbb{P}\left(\{X_\tau = j_0, X_{\tau+1} = j_1, \ldots, X_{\tau+n} = j_n\} \cap B \mid \tau < \infty, X_\tau = i\right) \\ &= \mathbb{P}_i\left(X_0 = j_0, X_1 = j_1, \ldots, X_n = j_n\right) \mathbb{P}\left(B \mid \tau < \infty, X_\tau = i\right) \end{aligned}$$

Which is equivalent to

$$\begin{aligned} &\mathbb{P}\left(\{X_\tau = j_0, X_{\tau+1} = j_1, \ldots, X_{\tau+n} = j_n\} \cap B \cap \{\tau < \infty\} \cap \{X_\tau = i\}\right) \\ &= \mathbb{P}_i\left(X_0 = j_0, X_1 = j_1, \ldots, X_n = j_n\right) \mathbb{P}\left(B \cap \{\tau < \infty\} \cap \{X_\tau = i\}\right) \end{aligned}$$

It's sufficient to prove that for any $m \geqslant 0$,

$$\begin{aligned} &\mathbb{P}\left(\{X_\tau = j_0, X_{\tau+1} = j_1, \ldots, X_{\tau+n} = j_n\} \cap B \cap \{\tau = m\} \cap \{X_\tau = i\}\right) \\ &= \mathbb{P}_i\left(X_0 = j_0, X_1 = j_1, \ldots, X_n = j_n\right) \mathbb{P}\left(B \cap \{\tau = m\} \cap \{X_\tau = i\}\right)\,, \end{aligned}$$

i.e.,

$$\begin{aligned}&\mathbb{P}\left(\{X_m = j_0, X_{m+1} = j_1, \ldots, X_{m+n} = j_n\} \cap B \cap \{\tau = m\} \cap \{X_m = i\}\right) \\ &\quad = \mathbb{P}_i\left(X_0 = j_0, X_1 = j_1, \ldots, X_n = j_n\right) \mathbb{P}\left(B \cap \{\tau = m\} \cap \{X_m = i\}\right),\end{aligned}$$

Note that $B \cap \{\tau = m\} \in \sigma(X_0, \cdots, X_m)$ so, by the Markov property at time m we konw the indentity holds. So we comlpete the proof. □

¶ Example 1.9 (去掉时滞). We now consider an application of the strong Markov property to a Markov chain $(X_n)_{n\geq 0}$ observed only at certain times. In the first instance suppose that *we observe the original chain only when it moves*. Let $S_0 = 0$ and for $m = 0, 1, 2, \cdots$

$$S_{m+1} = \inf\{n \geqslant S_m + 1 : X_n \neq X_{S_m}\}$$

Assume there are no absorbing states. It's easy to find that the random times S_m for $m \geqslant 0$ are stopping times. The resulting process $(Z_n)_{n\geqslant 0}$ is given by $Z_n = X_{S_n}$ and, by the strong Markov property

$$\begin{aligned}&\mathbb{P}\left(Z_{m+1} = i_{m+1} \mid Z_0 = i_1, \ldots, Z_m = i_m\right) \\ &= \mathbb{P}\left(X_{S_{m+1}} = i_{m+1} \mid X_{S_0} = i_1, \ldots, X_{S_m} = i_m\right) \\ &= \mathbb{P}_{i_m}\left(X_{S_1} = i_{m+1}\right) = \widetilde{p}_{i_m i_{m+1}}\end{aligned}$$

where $\tilde{p}_{ii} = 0$ and, for $i \neq j$, $\tilde{p}_{ij} = p_{ij}/(1 - p_{ii})$. Thus $(Z_m)_{m\geq 0}$ is a Markov chain with transition matrix $\tilde{P}$.

¶ Example 1.10 (限制观测窗口). A second example of a similar type arises if $J \subset I$ is some subset of the state-space and we observe the chain only when it takes values in J. The resulting process $(Y_m)_{m\geq 0}$ may be obtained formally by setting $Y_m = X_{T_m}$, where

$$T_0 = \inf\{n \geq 0 : X_n \in J\}$$

and, for $m = 0, 1, 2, \cdots$

$$T_{m+1} = \inf\{n > T_m : X_n \in J\}$$

Let us assume that $\mathbb{P}\left(T_m < \infty\right) = 1$ for all m. For each m we can check easily that T_m, the time of the m th visit to J, is a stopping time. So the strong Markov property applies to show, for $i_1, \ldots, i_{m+1} \in J$, that

$$
\begin{aligned}
&\mathbb{P}\left(Y_{m+1} = i_{m+1} \mid Y_0 = i_1, \ldots, Y_m = i_m\right) \\
&= \mathbb{P}\left(X_{T_{m+1}} = i_{m+1} \mid X_{T_0} = i_1, \ldots, X_{T_m} = i_m\right) \\
&= \mathbb{P}_{i_m}\left(X_{T_1} = i_{m+1}\right) = \bar{p}_{i_m i_{m+1}}
\end{aligned}
$$

where $p_{ij} = \mathbb{P}_i(\sigma_J = j)$ for all $i, j \in J$.

¶ EXAMPLE 1.11. Let $(X_n)_{n\geqslant 0}$ be gamblers' ruin We know from Example

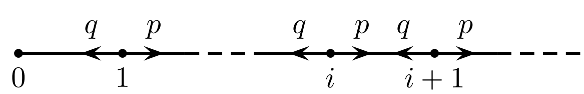


Figure 1.3: gamblers' ruin

1.3 the probability of hitting 0 starting from 1. Here we obtain the complete distribution of the time to hit 0 starting from 1 in terms of its *probability generating function*. Set

$$\tau_j = \inf\{n \geq 0 : X_n = j\}$$

and, for $0 \leqslant s < 1$,

$$\phi(s) = \mathbb{E}_1\left(s^{\tau_0}\right) = \sum_{n=1}^{\infty} \mathbb{P}_1\left(\tau_0 = n\right) s^n .$$

Using first step analysis,

$$
\begin{aligned}
\phi(s) = \mathbb{E}_1\left(s^{\tau_0}\right) &= p\,\mathbb{E}_1\left(s^{\tau_0} \mid X_1 = 2\right) + q\,\mathbb{E}_1\left(s^{\tau_0} \mid X_1 = 0\right) \\
&= ps\,\mathbb{E}_1\left(s^{\tau_0 - 1} \mid X_1 = 2\right) + q\,\mathbb{E}_1\left(s \mid X_1 = 0\right) \\
&= ps\,\mathbb{E}_2\left(s^{\tau_0}\right) + qs
\end{aligned}
$$

Now we try to compute $\mathbb{E}_2\left(s^{\tau_0}\right)$. Under $\mathbb{P}_2$, we have $\tau_0 < \tau_1$. Apply the strong Markov property at τ_1 to see that conditional on $\tau_1 < \infty$ (and of course $X_{\tau_1} = 1$), we have $\tau_0 - \tau_1$ is independent of τ_1.

$$\begin{aligned}
\mathbb{E}_2\left(s^{\tau_0}\right) &= \mathbb{E}_2\left(s^{\tau_1} s^{\tau_0-\tau_1} \mid \tau_1 < \infty\right)\mathbb{P}_2\left(\tau_1 < \infty\right) \\
&= \mathbb{E}_2\left(s^{\tau_1} 1_{\{\tau_1<\infty\}}\right)\mathbb{E}_2\left(s^{\tau_0-\tau_1} \mid \tau_1 < \infty\right) \\
&= \mathbb{E}_2\left(s^{\tau_1}\right)\mathbb{E}_2\left(s^{\tau_0-\tau_1} \mid \tau_1 < \infty\right)
\end{aligned}$$

The space is translation invarint, so the distribution of τ_1 under P_2 coincides with the distribution of τ_0 under P_1, so $\mathbb{E}_2\left(s^{\tau_1}\right) = \phi(s)$. On the other hand,by strong Markov property, conditional on $\tau_1 < \infty$, $\tau_0 - \tau_1$ has the same distribution of τ_0 under P_1, thus $\mathbb{E}_2\left(s^{\tau_0-\tau_1} \mid \tau_1 < \infty\right) = \phi(s)$. Then $\mathbb{E}_2\left(s^{\tau_0}\right) = \phi(s)^2$.

Thus $\phi = \phi(s)$ satisfies

$$ps\phi^2 - \phi + qs = 0 \tag{1.6}$$

and this equation has two solutions: $(1 \pm \sqrt{1-4pqs^2})/2ps$. Since $\phi(0) \leqslant 1$, and ϕ is continuous we are forced to take the negative root at $s = 0$ and stick with it for all $0 \leq s < 1$. Thus

$$\phi(s) = \frac{1-\sqrt{1-4pqs^2}}{2ps}. \tag{1.7}$$

(i) To recover the distribution of τ_0 we expand the square-root as a power series:

$$\begin{aligned}
\phi(s) &= \frac{1}{2ps}\left\{1-\left(1+\frac{1}{2}\left(-4pqs^2\right)+\frac{1}{2}\left(-\frac{1}{2}\right)\left(-4pqs^2\right)^2/2!+\ldots\right)\right\} \\
&= qs + pq^2s^3 + \ldots \\
&= s\mathbb{P}_1\left(\tau_0 = 1\right) + s^2\mathbb{P}_1\left(\tau_0 = 2\right) + s^3\mathbb{P}_1\left(\tau_0 = 3\right) + \ldots
\end{aligned}$$

The first few probabilities $\mathbb{P}_1\left(\tau_0 = 1\right), \mathbb{P}_1\left(\tau_0 = 2\right), \cdots$ are readily checked from first principles.

(ii) On letting $s \uparrow 1$, we have $\phi(s) \to \mathbb{P}_1(\tau_0 < \infty)$, so

$$\mathbb{P}_1(\tau_0 < \infty) = \frac{1-\sqrt{1-4pq}}{2p} = \begin{cases} 1\,, & \text{if } p \leq q\,. \\ q/p\,, & \text{if } p > q\,. \end{cases}$$

(iii) We can also find the mean hitting time using

$$\mathbb{E}_1(\tau_0) = \lim_{s\uparrow 1} \phi'(s)\,.$$

It is only worth considering the case $p \leqslant q$, where the mean hitting time has a chance of being finite. Differentiate (1.6) to obtain

$$2ps\phi\phi' + p\phi^2 - \phi' + q = 0\,,$$

so

$$\phi'(s) = \left(p\phi(s)^2 + q\right)/(1-2ps\phi(s)) \to 1/(1-2p) = 1/(q-p) \quad \text{as } s \uparrow 1\,.$$

EXERCISE

¶ Exercise 1.12. τ, σ both are stopping times. Then

(i) $\tau \wedge \sigma$, $\tau \vee \sigma$, $\tau + \sigma$ are stopping times.

(ii) Assume $\sigma \geqslant \tau$, give an example such that $\sigma - \tau$ is not stopping time.

¶ Exercise 1.13. Let $\xi_1, \xi_2, \ldots$ be independent identically distributed random variables with $\mathbb{P}(\xi_1 = 2) = \mathbb{P}(\xi_1 = -1) = 1/2$, and set $X_0 = 1, X_n = X_0 + \xi_1 + \cdots + \xi_n$ for $n \geq 1$. Show that the probability generating function $\phi(s) = \mathbb{E}(s^{\tau_0})$ now satisfies

$$s\phi^3 - 2\phi + s = 0$$

1.4 Recurrence and transience

Let $(X_n)_{n\geqslant 0}$ be a Markov chain with transition matrix P. We say that a state i is **recurrent** if

$$\mathbb{P}_i\left(X_n = i \text{ i.o. }\right) = 1\,,$$

is **transient** if

$$\mathbb{P}_i\left(X_n = i \text{ i.o. }\right) = 0\,.$$

Thus a recurrent state is one to which you keep coming back and a transient state is one which you eventually leave for ever. We shall show that every state is either recurrent or transient.

1.4.1 Decomposing orbit by excursions

Recall that **the first passage time** to state i is the random variable T_i defined by

$$T_i = \inf\left\{n \geqslant 1 : X_n = i\right\}\,,.$$

where $\inf\varnothing = \infty$. We now define inductively the r th passage time $T_i^{(r)}$ to state i by

$$T_i^{(0)} = 0, \quad T_i^{(1)} = T_i$$

and, for $r = 0, 1, 2, \cdots$

$$T_i^{(r+1)} = \inf\left\{n \geqslant T_i^{(r)} + 1 : X_n(\omega)\right\}\,.$$

The **length of the rth excursion** to i is then

$$\sigma_i^{(r)} = \begin{cases} T_i^{(r)} - T_i^{(r-1)}\,, & \text{if } T_i^{(r-1)} < \infty\,. \\ \infty\,, & \text{otherwise}\,. \end{cases}$$

Our analysis of recurrence and transience will rest on finding the joint distribution of these excursion lengths.

Lemma 1.11. *For $r = 2, 3, \cdots$, conditional on $T_i^{(r-1)} < \infty$, $\sigma_i^{(r)}$ is independent of $\left(X_0, \cdots, X_{T_i^{(r-1)}}\right)$ and*

$$\mathbb{P}\left(\sigma_i^{(r)} = n \mid T_i^{(r-1)} < \infty\right) = \mathbb{P}_i\left(T_i = n\right)$$

Proof. Apply the strong Markov property at the stopping time $T = T_i^{(r-1)}$. It is automatic that $X_T = i$ on $T < \infty$. So, conditional on $T < \infty$, $(X_{T+n})_{n\geqslant 0}$ is Markov (δ_i, P) and independent of $X_0, X_1, \ldots, X_T$. But

$$\sigma_i^{(r)} = \inf\{n \geqslant 1 : X_{T+n} = i\}$$

so $\sigma_i^{(r)}$ is the first passage time of $(X_{T+n})_{n\geqslant 0}$ to state i. □

Corollary 1.12. *For any positive integer r and $n_1, \cdots, n_r \in \mathbb{N}_+$,*

$$\mathbb{P}_i\left(\sigma_i^{(1)} = n_1, \sigma_i^{(2)} = n_2, \cdots, \sigma_i^{(r)} = n_r\right) = \prod_{s=1}^{r} \mathbb{P}_i\left(\sigma_i = n_s\right) .$$

Moreover, $\{\sigma_i^{(r)}\}_{r=1}^{\infty}$ are i.i.d. r.v.'s under $\mathbb{P}_i$ if $\mathbb{P}_i(\sigma_i < \infty) = 1$.

Also, we can compute the distribution of $T_i^{(r)}$ in terms of the **return probability**

$$\rho_{ij} = \mathbb{P}_i\left(\sigma_j < \infty\right) .$$

Then

Corollary 1.13. *For any $r \in \mathbb{N}^+$, We have*

$$\mathbb{P}\left(T_i^{(r)} < \infty\right) = \rho_{ii}^{r-1}\mathbb{P}\left(\sigma_i < \infty\right).$$

Proof. When $r = 1$ the result is true. Suppose inductively that it is true for r, then

$$\begin{aligned}
\mathbb{P}\left(T_i^{(r+1)} < \infty\right) &= \mathbb{P}\left(T_i^{(r)} < \infty \text{ and } \sigma_i^{(r+1)} < \infty\right) \\
&= \mathbb{P}\left(\sigma_i^{(r+1)} < \infty \mid T_i^{(r)} < \infty\right)\mathbb{P}\left(T_i^{(r)} < \infty\right) \\
&= \mathbb{P}_i\left(\sigma_i < \infty\right)\mathbb{P}\left(T_i^{(r)} < \infty\right) \\
&= \rho_{ii}(\rho_{ii})^{r-1}\mathbb{P}\left(\sigma_i < \infty\right) = (\rho_{ii})^{r}\mathbb{P}\left(\sigma_i < \infty\right) .
\end{aligned}$$

So by induction the result is true for all $r \in \mathbb{N}_+$. □

1.4.2 Visits number and recurrence

Let us introduce **the number of visits** V_i **to** i, which may be written in terms of indicator functions as

$$V_i = \sum_{n=0}^{\infty} 1_{\{X_n = i\}} .$$

It's easy to observe that if the Markov chain starts at state i, i.e.,$X_0 = i$, then

$$\{V_i > r\} = \{T_i^{(r)} < \infty\}.$$

Then $\mathbb{P}_i(V_i > n) = (\rho_{ii})^n$ for any n. Thus we have shown

Theorem 1.14. *Under $\mathbb{P}_i$, the number of visits V_i is* geometric$(1 - \rho_{ii})$.

As a consequence,

$$\mathbb{P}_i(V_i = \infty) = \begin{cases} 1, & \rho_{ii} = 1. \\ 0, & \rho_{ii} < 1. \end{cases}$$

In particular, every state is either transient or recurrent. Besides,

$$\mathbb{E}_i(V_i) = \sum_{r=0}^{\infty} \mathbb{P}_i(V_i > r) = \sum_{r=0}^{\infty} (\rho_{ii})^r = \begin{cases} \infty, & \rho_{ii} = 1. \\ \frac{1}{1-\rho_{ii}}, & \rho_{ii} < 1. \end{cases}$$

On the other hand, V_i is the sum of indicator functions, so

$$\mathbb{E}_i\left(V_i\right) = \sum_{n=0}^{\infty} \mathbb{P}_i\left(X_n = i\right) = \sum_{n=0}^{\infty} p_{ii}^{(n)} .$$

Finally, we have got the necessary and sufficient condition of recurrence.

Theorem 1.15. *The following dichotomy holds:*

(i) i *is recurrent* $\Leftrightarrow \rho_{ii} = 1 \Leftrightarrow \sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$.

(ii) i *is transient* $\Leftrightarrow \rho_{iii} < 1 \Leftrightarrow \sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty$.

Another proof using first passage decomposition is given in Exercise 1.21.

First we show that recurrence and transience are *class properties*.

Theorem 1.16. *Suppose state i is recurrent, and $i \to j$, then*

(i) $j \to i$, *and* $\mathbb{P}_j(\tau_i < \infty) = 1$.

(ii) j *is recurrent.*

Proof. First we show that $\mathbb{P}_j(\tau_i < \infty) = 1$. Since $\mathbb{P}_i(V_i = \infty) = 1$, there must be $\mathbb{P}_i(V_i = \infty, \tau_j < \infty) = \mathbb{P}_i(\tau_j < \infty)$. On the other hand,

$$\{V_i = \infty, \tau_j < \infty\} = \Big\{ \sum_{n \geqslant \tau_j} 1_{\{X_n = i\}} = \infty, \tau_j < \infty \Big\}$$

Thus, by strong Markov property,

$$\mathbb{P}_i\left(V_i = \infty\,, \tau_j < \infty\right) = \mathbb{P}_i(\tau_j < \infty)\,\mathbb{P}_j(V_i = \infty)$$

Thus

$$\mathbb{P}_j(V_i = \infty) = 1\,.$$

But $\mathbb{P}_j(V_i = \infty) \leqslant \mathbb{P}_j(\tau_i = \infty) \leqslant 1$, so we get $\mathbb{P}_j(\tau_i = \infty) = 1$. Particluarly, j leads to i.

So $j \leftrightarrow i$, there exist $l, m \geqslant 0$ with $p_{ij}^{(l)} > 0$ and $p_{ji}^{(m)} > 0$, and, for all $r \geqslant 0$

$$p_{jj}^{(l+r+m)} \geqslant p_{ij}^{(l)} p_{ii}^{(r)} p_{ji}^{(m)}$$

So

$$\infty = \sum_{r=0}^{\infty} p_{ii}^{(r)} \leqslant \frac{1}{p_{ij}^{(n)} p_{ji}^{(m)}} \sum_{r=0}^{\infty} p_{kk}^{(l+r+m)}$$

Hence j is also recurrent by Theorem 1.15. Another proof of (ii), using the partition of orbits by excuision , can be found in Exercise 1.22. □

Corollary 1.17. *Recurrence and transience are class properties, and every recurrent class is closed.*

1.4.3 Absorption probability test

The following theorem relates the recurrence with absorbtion probability. Intuitively, a state i is recurrent, then the chain must return to i with probability 1 from any state i leading to. And the converse is ture.

Theorem 1.18. *State i is recurrent if and only if for any state j such that i leads to j, there must be* $\mathbb{P}_j(\tau_i < \infty) = 1$.

Proof. Sufficiency : If $i \to j$ implies $\mathbb{P}_j(\tau_i < \infty) = 1$, since

$$\mathbb{P}_i(\sigma_i < \infty) = \sum_{j \in I} p_{ij} \mathbb{P}_j(\tau_i < \infty) \ ,$$

and for any j such that $p_{ij} > 0$, $\mathbb{P}_j(\tau_i < \infty) = 1$. Thus $\mathbb{P}_i(\sigma_i < \infty) = 1$.

Necessity : See Theorem 1.16. □

We will need the following corollary in Section 1.8, which asserts that irreducibile and recurrent chain will visit any state with probability one, no matter what the initial distribution is.

Corollary 1.19. *Suppose P is irreducible and recurrent. Then for all state j we have* $\mathbb{P}(\sigma_j < \infty) = 1$,

Proof. By total probability formula we have

$$\mathbb{P}(\sigma_j < \infty) = \sum_{i \in I} \mathbb{P}(X_0 = i) \mathbb{P}_i(\sigma_j < \infty)$$

Note that i is recurrent, $\mathbb{P}_i(\sigma_i < \infty) = 1$. and by Theorem 1.18 we konw that $\mathbb{P}_i(\sigma_j < \infty) = \mathbb{P}_i(\tau_j < \infty) = 1$, for all $j \neq i$. So $\mathbb{P}(\sigma_j < \infty) = 1$. □

<u>Absorption probability test</u> : If P is irreducible, then P is recurrent if and only if for any state i, (1.8) has a unique solution: $x_j = 1, \forall\, j \in I$.

$$\begin{cases} x_i = 1\,. \\ x_j = \sum_{k \in I} p_{ik} x_k, \ \forall\, j \neq i\,. \\ x_j \geqslant 0, \ \forall\, j \in I\,. \end{cases} \tag{1.8}$$

¶ EXAMPLE 1.14 (Recurrence of the birth-death chain). Let $(X_n)_{n\geqslant 0}$ be a irreducible birth-death chain with birth probability b_i and death probability d_i. Let $R_i = 1+\sum_{k=1}^{i}\frac{d_1\cdots d_k}{b_1\cdots b_k}$, $R = 1+\sum_{k=1}^{\infty}\frac{d_1\cdots d_k}{b_1\cdots b_k}$. By Example 1.2,

(i) if $R<\infty$, $x_i \coloneqq 1-\frac{R_{i-1}}{R}$ is a solution not equals 1, so the Markov chain is transient ;

(ii) if $R=\infty$, then $x_i = 1, \forall i\in\mathbb{N}_+$, the chain is recurrent.

Therefore, the Markov chain is recurrent if and only if $R=\infty$.

¶ EXAMPLE 1.15 (λ-biased random walk on homogeneous tree $\mathbb{T}^d$). Let $(X_n)_{n\geqslant 0}$ be the λ-biased random walk on homogeneous tree $\mathbb{T}^d$. Let

$$Y_n = |X_n| \text{ for each } n\in\mathbb{N}\,,$$

then $(Y_n)_{n\geqslant 0}$ is a birth-death chain with birth probability $b_i = \frac{d}{\lambda+d}$. $(X_n)_{n\geqslant 0}$ hits it's root if and only if $(Y_n)_{n\geqslant 0}$ hits 0, thus the recurrence of $\{X_n\}$ coincides with $\{Y_n\}$. By Example 1.14, $R = 1+\sum_{k=1}^{\infty}(\frac{\lambda}{d})^k$, hence $\{X_n\}$ is recurrent if and only if $\lambda\geqslant d$.

¶ EXAMPLE 1.16. 我们来考虑 Example 1.4 中马氏链的常返性. 这个马氏链是不可约的, 其常返性质等价于吸收概率方程组的解是否恒为 1. 由 Example 1.4 中的讨论, 我们知道该马氏链常返当且仅当 $R=\infty$, 其中

$$R = 1+\sum_{k=1}^{\infty}\exp\left(-c\sum_{j=1}^{k}j^{-\alpha}\right)$$

(i) 若 $c=0$, 则 $R=\infty$, 马氏链常返, 它就是带反射壁的简单随机游动.

(ii) 若 $c>0$. 当 $a>1$ 时,

$$\sum_{j=1}^{k}j^{-\alpha} < \sum_{j=1}^{\infty}j^{-\alpha} < \infty, \exp\left(-c\sum_{j=1}^{k}j^{-\alpha}\right)\geqslant\varepsilon>0$$

于是 $R=\infty$, 马氏链常返.

(iii) 若 $c>0$. 当 $a<1$ 时

$$\sum_{j=1}^{k} j^{-\alpha} \approx \frac{k^{1-\alpha}}{1-\alpha}, \ \exp\left(-c\sum_{j=1}^{k} j^{-\alpha}\right) \approx \exp\left(-\frac{c}{1-\alpha}k^{1-\alpha}\right).$$

于是 $R<\infty$, 马氏链常返.

(iv) 若 $c>0$. 当 $a=1$ 时

$$\sum_{j=1}^{k} j^{-1} \approx \log k, \exp\left(-c\sum_{j=1}^{k} j^{-\alpha}\right) \approx k^{-c}$$

于是当且仅当 $c \leqslant 1$ 时 $R=\infty$.

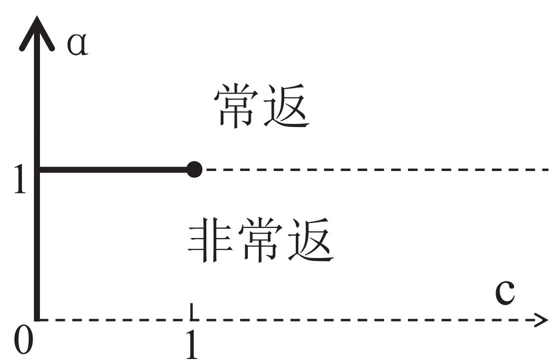


Figure 1.4: 参数取值与常返

于是, 常返性依赖于参数的关系如 Figure 1.4 所示, 其中实线表示常返, 虚线表示非常返.

¶ Example 1.17 (Simple random walk on $\mathbb{Z}^d$). The fact is when $d=1,2$, the walk is recurrent, but when $d \geqslant 3$, the walk is transient.

Suppose we start at 0. It is clear that we cannot return to 0 after an odd number of steps, so $p_{00}^{(2n+1)}=0$ for all n. Assume we return to 0 after $2n$ steps. Of these 2n steps there must be l_r up, l_r down in the rth direction, with $l_1+\cdots+l_d=n$. By counting the ways in which this can be done, we obtain

$$p_{00}^{(2n)} = P_0\left(S_{2n}=0\right) = \sum_{l_1+\cdots+l_d=n} \frac{(2n)!}{\left(l_1!\right)^2\cdots\left(l_d!\right)^2}\cdot\frac{1}{(2d)^{2n}},$$

where l_r take values in $\mathbb{N}$ for all $r=1,\cdots,d$. We will prove

$$P_0\left(S_{2n}=0\right)\sim Cn^{-\frac{d}{2}}.$$

(i) when $d=1$, using Stirling formular $n!\sim(n/e)^n\sqrt{2\pi n}$ we know

$$p_{00}^{(2n)} = \frac{(2n)!}{(n!)^2}\cdot\frac{1}{2^{2n}}\sim\frac{1}{\sqrt{\pi n}}$$

So $\mathbb{E}_0 V_0 = \sum_{n=0}^{\infty} p_{00}^{(n)} = \infty$.

(ii) When $d=2$,

$$\begin{aligned} P_0\left(S_{2n}=0\right) &= \sum_{l_1+l_2=n}\frac{(2n)!}{\left(l_1!\right)^2\left(l_2!\right)^2}\frac{1}{4^{2n}} = \frac{(2n)!}{n!n!}\frac{1}{4^{2n}}\sum_{l_1+l_2=n}\frac{n!}{l_1!l_2!}\frac{n!}{l_2!l_1!} \\ &= \binom{2n}{n}\frac{1}{4^{2n}}\sum_{l=0}^{n}\binom{n}{l}\binom{n}{n-l} = \frac{1}{4^{2n}}\binom{2n}{n}^2\sim\frac{1}{\pi n} \end{aligned}$$

Also, we have $\mathbb{E}_0 V_0 = \sum_{n=0}^{\infty} p_{00}^{(n)} = \infty$.

(iii) When $d=3$,

$$\begin{aligned} P_0\left(S_{2n}=0\right) &= \sum_{l_1+l_2+l_3=n}\frac{(2n)!}{\left(l_1!\right)^2\left(l_2!\right)^2\left(l_3!\right)^2}\frac{1}{6^{2n}} \\ &= \frac{1}{2^{2n}}\frac{(2n)!}{n!n!}\sum_{l_1+l_2+l_3=n}\left(\frac{n!}{l_1!l_2!l_3!}\frac{1}{3^n}\right)^2 \end{aligned}$$

Note that

$$\sum_{l_1+l_2+l_3=n}\binom{n}{l_1\,l_2\,l_3}\left(\frac{1}{3}\right)^n = 1$$

the left-hand side being the total probability of all the ways of placing n balls randomly into three boxes. For the case where $n=3m$, we have

$$\frac{n!}{l_1!l_2!l_3!}\leqslant\frac{n!}{(m!)^3}\sim 3^n\frac{3\sqrt{3}}{2\pi n}$$

for all l_1, l_2, l_3. So

$$p_{00}^{(2n)} \leq \binom{2n}{n}\left(\frac{1}{2}\right)^{2n}\binom{n}{m\,m\,m}\left(\frac{1}{3}\right)^{n} \sim \left(\frac{3}{\pi n}\right)^{3/2} \quad \text{as } n \to \infty$$

by Stirling's formula. Hence, $\sum_{m=0}^{\infty} p_{00}^{(6m)} < \infty$ by comparison with $\sum_{n=0}^{\infty} n^{-3/2}$. But $p_{00}^{(6m)} \geq (1/6)^2 p_{00}^{(6m-2)}$ and $p_{00}^{(6m)} \geq (1/6)^4 p_{00}^{(6m-4)}$ for all m so we must have

$$\mathbb{E}_0 V_0 = \sum_{n=0}^{\infty} p_{00}^{(n)} < \infty\,,$$

and the walk is transient.

(iv) In the last, For the case when $d \geqslant 4$ we can use the same method with $d = 3$.

EXERCISE

¶ Exercise 1.18. Prove that every finite closed class is recurrent.

¶ Exercise 1.19. 称 $G_{ij} \coloneqq \mathbb{E}_i(V_j) = \sum_{n=0}^{\infty} p_{ij}^{(n)}$ 为格林函数 (见章复熹老师讲义). 证明

(i) $\mathbb{P}_i(T_j^{(r)} < \infty) = \mathbb{P}_i(V_j \geqslant r) = \rho_{ij}(\rho_{jj})^{r-1}$, 其中 $j \neq i$ 且 r 为正整数.

(ii) 先用 (i) 中的结论, 证明对任何 i, j, $G_{ij} = \mathbb{P}_i(\tau_j < \infty) G_{jj}$. 再直接证明它, 不要用 (ii) 中的结论.

(iii) 状态 j 非常返当且仅当 $\mathbb{E}(V_j) < \infty$ 对任何初始分布 λ 成立, 状态 j 常返当且仅当 $\mathbb{E}(V_j) < \infty$ 对任何初始分布 λ, $\mathbb{P}_\lambda(\tau_j < \infty)$ 成立.

¶ Exercise 1.20. π is an invariant distribution of P. If state j is transient, then $\pi_j = 0$.

¶ Exercise 1.21 (First passage decomposition). Denote by σ_j the first passage time to state j and set

$$f_{ij}^{(n)} = \mathbb{P}_i\left(\sigma_j = n\right)$$

Justify the identity

$$p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)} \quad \text{for } n \geqslant 1$$

and deduce that

$$P_{ij}(s) = \delta_{ij} + F_{ij}(s) P_{jj}(s)$$

where

$$P_{ij}(s) = \sum_{n=0}^{\infty} p_{ij}^{(n)} s^n, \quad F_{ij}(s) = \sum_{n=0}^{\infty} f_{ij}^{(n)} s^n$$

Hence show that $\mathbb{P}_i(T_i < \infty) = 1$ if and only if

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty .$$

¶ Exercise 1.22. We should point that decomposing the orbits into many excursions is a very useful idea. Assume state i is recurrent and $(X_n)_{n\geqslant 0}$. start at i. Thus $T_i^{(k)} < \infty$ for any $k \in \mathbb{N}_+$. Then we can define $Z_k = (X_{T_i^{(k-1)}}, \cdots, X_{T_i^{(k)}})$, is the k-th excursion. Z_k take values in $\{(i_0, \cdots, i_n) : i_0 = i_n = i; i_k \neq i, 0 < k < n; n \in \mathbb{N}_+\}$.

(i) Show that conditional on $X_0 = i$, $\{Z_k\}$ are i.i.d. random vectors taking values in $\cup_{n\geqslant 0} I^{n+1}$.

(ii) Using (i) to show that if i leads to j, then

  (a) $\mathbb{P}_i(\sigma_j < \sigma_i) > 0$.

  (b) $\mathbb{P}_i(V_j = \infty) = 1$.

  (c) $\mathbb{P}_j(V_j = \infty) = 1$ and $\mathbb{P}_i(\tau_j < \infty) = 1$.

(iii) Read the proof of idea and Example 1.35.

1.5 Invariant distributions

1.5.1 Definitions and examples

Many of the long-time properties of Markov chains are connected with the notion of an invariant distribution or measure. Remember that a distribution λ is any row vector $(\lambda_i)_{i\in I}$ with non-negative entries. We say λ is an **invariant distribution** for P if

$$\lambda P = \lambda . \tag{1.9}$$

When λ is a measure satisfying (1.9), we call it **invariant measure**.

Proposition 1.20. *Let $(X_n)_{n\geqslant 0}$ be Markov (π, P) and suppose that π is invariant for P. Then $(X_{m+n})_{n\geqslant 0}$ is also Markov (π, P).*

Proof. Clearly, $\mathbb{P}(X_m = i) = (\pi P^m)_i = \pi_i$ for all i.

On the other hand, conditional on $X_{m+n} = i$, X_{m+n+1} is independent of $(X_m, X_{m+1}, \cdots, X_{m+n})$ and has distribution $(p_{ij})_{j\in I}$. □

Based on this property, invariant distribution is also called **stationary distribution**. Sometimes it also called **equilibrium**, the reason is given in Theorem 1.39.

为了从直观上理解不变分布，我们引入“概率流”的观点．把每个状态理解为一个位置．假设我们有大量的粒子，先将每个粒子独立地按照分布 λ 放在某个随机位置，根据强大数定律，位置 i 上的粒子量 (相对于所有粒子的比例) 应该为单个粒子位于状态的概率，即 λ_i．现在，让所有粒子独立地按照转移矩阵 P 跳跃一步，那么，仍然根据强大数定律，从位置 i 跳跃至位置 j 的粒子量就应该为当个粒子最初位于位置 i 一步以后位于位置 j 的概率，即 $\lambda_i p_{ij}$．因此，我们观测到的现象是，在这一步转移中，从 i 到 j 的粒子流量为 $\lambda_i p_{ij}$，我们也说从 i 到 j 的概率流为 $\lambda_i p_{ij}$．

若 π 为不变分布，则它任意位置 i 上的粒子量保持不变，因此，在 π 诱导的概率流中，对任意位置，从其他位置流入位置 i 的总流量 $\sum_{j\neq i} \pi_j p_{ji}$ 与

从 i 位置流出的总流量 $\pi_i(1-p_{ii})$ 是相等的，即 $\sum_j \pi_j p_{ji} = \pi_i$. 进一步，自然应该有进出任意子集 A 的概率流相等：

Proposition 1.21. *π is invariant for P is and only if for any $A \subset I$,*

$$\sum_{i \notin A, j \in A} \pi_i p_{ij} = \sum_{i \in A, j \notin A} \pi_i p_{ij}\,. \tag{1.10}$$

Proof. Sufficiency is obvious. To show necessity, we observe that

$$\begin{aligned}\sum_{i \notin A, j \in A} \pi_i p_{ij} + \sum_{i \in A, j \in A} \pi_i p_{ij} &= \sum_{j \in A, i \in I} \pi_i p_{ij} = \sum_{j \in A} \pi_j\,.\\ \sum_{i \in A, j \notin A} \pi_j p_{ji} + \sum_{i \in A, j \in A} \pi_i p_{ij} &= \sum_{i \in A, j \in I} \pi_i p_{ij} = \sum_{i \in A} \pi_i\,.\end{aligned}$$

Since $\sum_{i \in A, j \in A} \pi_i p_{ij} < \infty$, (1.10) holds. □

Remark. This proposition may be not ture for invariant measure λ, since $\sum_{i \in A, j \in A} \lambda_i p_{ij}$ can be ∞.

¶ Example 1.23 (Birth-death chain). Consider the existence od invariant distributions for birth-death chain mentioned in Example 1.2. $b_0 = 1$. If π is an invariant distribution, by Proposition 1.21, $\pi_0 = \pi_1 d_1$ and

$$\pi_i d_i = \pi_{i-1} b_{i-1}, \ \forall\, i \geqslant 1\,.$$

Thus

$$\pi_i = \frac{b_{i-1}}{d_i} \pi_{i-1} = \cdots = \frac{b_0 b_1 \cdots b_{i-1}}{d_1 d_2 \cdots d_i} \pi_0 = \hat{\gamma}_i \pi_0, \ \forall\, i \geqslant 1\,.$$

where the final equality defines $\hat{\gamma}_i$ for $i \geqslant 1$. Let $\hat{\gamma}_0 = 1$ and $\hat{R} = \sum_{i=0}^{\infty} \hat{\gamma}_i$. Then we have $\hat{R}\pi_0 = 1$.

(i) In the case $\hat{R} = \infty$, this controducts that π is a distribution. So the chian has no invariant distribution.

(ii) But if $\hat{R} < \infty$, then let $\pi_i = \frac{\hat{\gamma}_i}{\hat{R}}$ for all $i \geqslant 0$. It's easy to check that π is an invariant distribution, and is unique!

¶ EXAMPLE 1.24. A transition matrix P is said to be **doubly stochastic** if its COLUMNS sum to 1, or in symbols $\sum_{i\in I} p_{ij} = 1$. The ajective "doubly" refers to the fact that by its definition a transition probability matrix has ROWS that sum to 1, i.e., $\sum_{j\in I} p_{ij} = 1$.

(i) If P is a doubly stochastic transition probability, then the uniform measure, $\pi_i = 1$ for all i, is invariant.

(ii) If P is a doubly stochastic transition probability for a Markov chain with N states, then the uniform distribution, $\pi_i = \frac{1}{N}$ for all i, is invariant.

¶ EXAMPLE 1.25. 设 $I = \mathbb{N}_+$, 转移概率为: 对于 $i \geqslant 1$, $p_{i,i-1} = \lambda/(\lambda+1)$, $p_{i,i+k} = p_k/(\lambda+1)$, $\forall k \geqslant 1$; $p_{01} = 1$. 其中 $1 = \sum_{k=1}^{\infty} p_k < m := \sum_{k=1}^{\infty} kp_k < \lambda$. 求该马氏链的不变分布.

解. 首先, 出入 0 的概率流相等, 即 $\pi_0 = \pi_1\lambda/(\lambda+1)$. 其次, 对任意 $i \geqslant 2$, 令 $A = \{n : n \geqslant i\}$, 进入 A 的概率流为 $\sum_{j=1}^{i-1}\sum_{k=i}^{\infty} \pi_j p_{j,k} = \sum_{j=1}^{i-1} \pi_j f_{i-j}/(\lambda+1)$, 其中 $f_r := \sum_{k=r}^{\infty} p_k$; 出 A 的概率流为 $\pi_i\lambda/(\lambda+1)$, 于是

$$\pi_i = \frac{1}{\lambda}\sum_{j=1}^{i-1} \pi_j f_{i-j}, \quad \forall i \geqslant 2$$

对 i 求和

$$1 - \pi_0 - \pi_1 = \frac{1}{\lambda}\sum_{i=2}^{\infty}\sum_{j=1}^{i-1} \pi_j f_{i-j} = \frac{1}{\lambda}\sum_{j=1}^{\infty}\sum_{i=j+1}^{\infty} \pi_j f_{i-j} = \frac{1}{\lambda}\sum_{j=1}^{\infty} \pi_j \sum_{r=1}^{\infty} f_r = \frac{m}{\lambda}(1-\pi_0)$$

其中

$$\sum_{r=1}^{\infty} f_r = \sum_{r=1}^{\infty}\sum_{k=r}^{\infty} p_k = \sum_{k=1}^{\infty}\sum_{r=1}^{k} p_k = \sum_{k=1}^{\infty} kp_k = m$$

因此解出

$$\pi_0 = \frac{\lambda - m}{\lambda - m + \lambda + 1}, \ \pi_1 = \frac{\lambda - m}{\lambda - m + \lambda + 1}\frac{\lambda}{\lambda+1}$$

根据上面的递推式得到 π_i, $i \geqslant 2$. □

1.5.2 Existence and uniqueness of invariant measures

In this subsection, we will discuss the existence and uniqueness of invariant measures and distributions for general Markov chains. Recall that a measure λ is invariant iff for any i, j in I,

$$\lambda_i = \sum_{j \in I} \lambda_i p_{ij}$$

It's obvious that if the $c\lambda$ is invariant for any constant c. So without loss of generality we pick some $k \in I$ and set $\lambda_k = 1$.

A sufficient condition for existence of invariant measures

In Theorem 1.8, we had proved that the absorption probabilities is the minimal non-negative solutions for (1.2), by iterating over the equations system. Thus we guess, is the invarint measure the minimal non-negative solution of the equations system above ?

For each $i \neq k$, we have

$$\begin{aligned}
\lambda_i &= \sum_{i_1 \in I} \lambda_{i_1} p_{i_1 i} = \sum_{i_1 \neq k} \lambda_{i_1} p_{i_1 i} + p_{ki} \\
&= \sum_{i_1, i_2 \neq k} \lambda_{i_2} p_{i_2 i_1} p_{i_1 i} + \left(p_{ki} + \sum_{i_1 \neq k} p_{k i_1} p_{i_1 i} \right) \\
&= \sum_{i_1, \cdots, i_n \neq k} \lambda_{i_n} p_{i_n i_{n-1}} \cdots p_{i_1 i} \\
&\quad + \left(p_{ki} + \sum_{i_1 \neq k} p_{k i_1} p_{i_1 i} + \cdots + \sum_{i_1, \cdots, i_{n-1} \neq k} p_{k i_{n-1}} \cdots p_{i_2 i_1} p_{i_1 i} \right) .
\end{aligned}$$

Hence we obtain

$$\begin{aligned}
\lambda_i \geqslant\ & \mathbb{P}_k \left(X_1 = i \text{ and } \sigma_k > 1 \right) + \mathbb{P}_k \left(X_2 = i \text{ and } \sigma_k > 2 \right) \\
& + \cdots + \mathbb{P}_k \left(X_n = i \text{ and } \sigma_k > n \right) .
\end{aligned}$$

Letting $n \to \infty$,

$$\lambda_i \geqslant \mathbb{E}_k \sum_{n=1}^{\infty} 1_{\{X_n = i, n < \sigma_k\}} = \mathbb{E}_k \sum_{1 \leqslant n < \sigma_k} 1_{\{X_n = i\}} .$$

On the other hand, we have $\lambda_k = 1$. So we rewrite the right-hand-side as following

$$\lambda_i \geqslant \mathbb{E}_k \sum_{0 \leqslant n < \sigma_k} 1_{\{X_n = i\}}\,, \text{ for all } i \in I\,.$$

In right-hand-side, the sum of indicator functions serves to count the number of times n at which $X_n = i$ before the first passage time σ_k, it is exactly **the expected time spent in i between visits to k**. We give it a notation:

$$\gamma_i^k \coloneqq \mathbb{E}_k \sum_{n=0}^{\sigma_k - 1} 1_{\{X_n = i\}}\,.$$

Obviously, $\gamma_k^k = 1$. Then We will check if $\gamma^k = \left(\gamma_i^k\right)_{i \in I}$ is a invariant measure for P.

Theorem 1.22. *Suppose k is recurrent. Then $\gamma^k = \left(\gamma_i^k\right)_{i \in I}$ is an invariant measure for P.*

Remark. Why is this true? This is called the "*cycle trick*". γ_i^k is the expected number of visits to i in $\{0, \cdots, \sigma_k - 1\}$. Multiplying by P moves us forward one unit in time so $\gamma^k P_{\cdot i}$ is the expected number of visits to i in $\{1, \cdots, \sigma_k\}$. We need the condition k is recurrent, then $X_{\sigma_k} = X_0 = k$, it follows that $\gamma^k P = \gamma^k$.

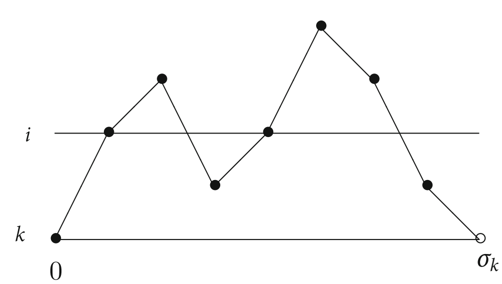


Figure 1.5: cycle trick

Proof. For any $j \in I$,

$$\begin{aligned}
\sum_{i\in I} \gamma_i^k p_{ij} &= \sum_{i\in I} p_{ij} \mathbb{E}_k \sum_{n=0}^{\sigma_k - 1} 1_{\{X_n = i\}} = \sum_{i\in I} \sum_{n=0}^{\infty} p_{ij} \mathbb{P}_k(X_n = i,\ n < \sigma_k) \\
&= \sum_{i\in I} \sum_{n=0}^{\infty} \mathbb{P}_k(X_n = i, X_{n+1} = j,\ n < \sigma_k) \\
&= \sum_{n=0}^{\infty} \mathbb{P}_k(X_{n+1} = j,\ n < \sigma_k) = \sum_{n=1}^{\infty} \mathbb{P}_k(X_n = j,\ n \leqslant \sigma_k) \\
&= \mathbb{E}_k \sum_{n=1}^{\sigma_k} 1_{\{X_n = j\}}
\end{aligned}$$

Since k is recurrent, under $\mathbb{P}_k$ we have $\sigma_k < \infty$, then $X_0 = X_{\sigma_k} = k$. So we have

$$\sum_{n=0}^{\sigma_k - 1} 1_{\{X_n = j\}} = \sum_{n=1}^{\sigma_k} 1_{\{X_n = j\}},\ \mathbb{P}_k\text{-a.s.}$$

So we get $\sum_{i\in I} \gamma_i^k p_{ij} = \gamma_j^k$, in other words, $\gamma^k P = \gamma^k$. □

Remark. Without the condition that k is recurrent, we only have

$$\sum_{n=0}^{\sigma_k - 1} 1_{\{X_n = j\}} = \sum_{n=1}^{\sigma_k} 1_{\{X_n = j\}} + 1_{\{X_0 = j,\ \sigma_k = \infty\}}.$$

So in this case, γ^k is not always invariant .

Now we have a sufficient condition for the existence of invariant measure.

Corollary 1.23. *If Markov chain has a recurrent state, then invaiant measures exist.*

Remark. *The existence of an invariant measure does not guarantee recurrence, even if the chain is irreducible.* For a counterexample,

¶ EXAMPLE 1.26. Consider the simple symmetric random walk on $\mathbb{Z}^3$, which is transient by Example 1.17, but has an invariant measure π given by $\pi_i = 1$ for all i.

A sufficient condition for uniqueness of invariant measures

Let $k \in I$ be recurrent and C is the communicating class containing k. Then C is closed, so there must be $\gamma_i^k = 0$ for all $i \notin C$. Hence $(\gamma_i^k)_{i\in C}$ is invariant for $P|_C$, and $P|_C$ is irreducible. So if I has two distinct recurrent class, the invariant measures for P is not unique. Therefore, now we only need to discuss the case that P is irreducible and recurrent.

First, we point that the invariant measures for an irreducible Markov chain has an important property:

Lemma 1.24. *Let P be irreducible. If λ is an invariant measure for P, then $\lambda = 0$, or $0 < \lambda < \infty$. Paricularly, if π is an invariant distribution for P, then $\pi_i > 0$ for all i.*

Proof. Let λ is a invariant measure for P. If there exists some $i \in I$ such that $\lambda_i > 0$, for any $j \in I$, pick n satisfying $p_{ij}^{(n)} > 0$,then $\lambda_j \geqslant \lambda_i p_{ij}^{(n)} > 0$. □

Now, we give a sufficient condition for the uniqueness of invariant measures:

Theorem 1.25. *Let P be irreducible and recurrent. Let λ be an invariant measure for P with $\lambda_k = 1$ for some $k \in I$, then $\lambda = \gamma^k$. In other words, P has an unique invariant measures up to scalar multiples.*

Proof. We have proved $\lambda \geqslant \gamma^k$. Since P is recurrent, γ^k is invariant by Theorem 1.22. So $\mu = \lambda - \gamma^k$ is also an invariant measure for P. Since P is irreducible, by Lemma 1.24, since $\mu_k = 0$ we konw $\mu \equiv 0$, so $\lambda = \gamma^k$. □

Remark. When I has only one recurrent class, but is not irreducible, the invariant measure may be not unique up to scalar mutiples. For an example,

¶ EXAMPLE 1.27. Consider the chain on $\mathbb{Z} \cup \{\infty\}$. Let $p_{i,i+1} = 1$ for $i \in \mathbb{Z}$, and , and $p_{\infty\infty} = 1$. Then P is transient, and 1_∞, $1_\mathbb{Z}$ both are invariant.

Corollary 1.26. *Let P be irreducible and recurrent. Then if P has an invariant distribution, it must be unique.*

¶ Example 1.28. Consider the asymmetric random walk on $\mathbb{Z}$ with transition probabilities $p_{i,i-1} = q < p = p_{i,i+1}$ In components the invariant measure equation $\lambda P = \lambda$ reads

$$\lambda_i = \lambda_{i-1}p + \lambda_{i+1}q$$

This is a recurrence relation for λ with general solution

$$\lambda_i = A + B(\frac{p}{q})^i$$

So, in this case, there is a two-parameter family of invariant measures uniqueness up to scalar multiples does not hold.

1.5.3 Existence and uniqueness of invariant distributions

Obviously, if P has a invariant measures λ and $\Lambda := \sum_{i\in I} \lambda_i < \infty$. Then let

$$\pi_i = \frac{\lambda_i}{\Lambda} \text{ for all } i \in I\,.$$

Then π is an invariant distribution for P. Note that

$$\sum_{j\in I} \gamma_j^i = \sum_{j\in I} \mathbb{E}_i \sum_{0\leqslant n<\sigma_i} 1_{\{X_n=j\}} = \mathbb{E}_i \sum_{0\leqslant n<\sigma_i} \sum_{j\in I} 1_{\{X_n=j\}} = \mathbb{E}_i\left(\sigma_i\right)$$

is exactly **the expected return time to** i when startingfrom i, and we give it a notation

$$m_i := \mathbb{E}_i\left(\sigma_i\right) = \sum_{j\in I} \gamma_j^i\,.$$

Now, it's natural to introduce the following definition :

Definition 1.4. We say a state i is **positive recurrent** if $m_i < \infty$, and a recurrent state which fails to have this stronger property is called **null recurrent**.

Thus, if I has a positive recurrent i, by normalizing γ^i, we get an invariant distribution $\pi = \frac{\gamma^i}{m_i}$. Let C be the communicating class containing i, which is closed, then $\sum_{i\in C}\pi_i = 1$, and $(\pi_i)_{i\in C}$ is invariant for irreducible transition matrix $P|_C$. Thus we focus on the case that P is irreducible first.

The next theorem says that positive recurrence is a class property. And an irreducible chain has invariant distribution $\Leftrightarrow$ it has unique invariant distribution $\Leftrightarrow$ it is positive recurrent.

Theorem 1.27. *Let P be irreducible. Then the following are equivalent:*

(i) *every state is positive recurrent.*

(ii) *some state i is positive recurrent.*

(iii) P *has an invariant distribution.*

(iv) P *has an unique invariant distribution π and $\pi_i = \frac{1}{m_i}$ for all $i \in I$.*

Proof. $1 \Rightarrow 2$ is obvious.

$2 \Rightarrow 3$. If i is positive recurrent, so P is recurrent. By Thoerem 1.22, γ^i is then invariant. But

$$\sum_{j\in I}\gamma^i_j = m_i < \infty\,.$$

So $\pi = \frac{1}{m_i}\gamma^i$ defines an invariant distribution for P.

$3 \Rightarrow 1$. Take any state k, Since P is irreducible and $\sum_{i\in I}\pi_i = 1$ we have $\pi > 0$ by Lemma 1.24. Hence $\pi_k > 0$. Set $\lambda = \frac{1}{\pi_k}\pi$, then λ is invariant with $\lambda_k = 1$. By Theorem 1.25, $\gamma^k \leqslant \lambda$, so

$$m_k = \sum_{i\in I}\gamma^k_i \leqslant \sum_{i\in I}\frac{\pi_i}{\pi_k} = \frac{1}{\pi_k} < \infty$$

and k is positive recurrent.

$3 \Leftrightarrow 4$. First, $4 \Rightarrow 3$ is obvious. Second, we prove $3 \Rightarrow 4$. Since 3 implies 1, we konw P is recurrent, so the invariant distribution must be unique.

At the same time, assume π is a invariant distribution, from Theorem 1.25, $\frac{1}{\pi_k}\pi = \gamma^k$ for all $k \in I$. Therefore $m_k = \sum_{i \in I} \gamma_i^k = \sum_{i \in I} \frac{\pi_i}{\pi_k} = \frac{1}{\pi_k}$. we have $\pi_k = \frac{1}{m_k}$ for all $k \in I$. □

Now we can give a necessary and sufficient condition for the existence and uniqueness of invariant distributions.

Theorem 1.28. *$(X_n)_{n \geqslant 0}$ be Markov chian on I with transition matrix P.*

(i) P has invariant distributions if and only if I has a positive recurrent class. In this case, let π one invariant distributions, then $\pi_i = \frac{\pi([i])}{m_i}$ for all $i \in I$, where $[i]$ is the communicating class containg i.

(ii) P has unique invariant distribution, if and only if I has unique positive recurrent class. In this case, let π the invariant distributions, and C is the unique positive recurrrent class. Then $\pi_i = \frac{1}{m_i}$ for all $i \in C$, and $\pi_i = 0$ for all $i \in I \backslash C$.

¶ Example 1.29 (Simple symmetric random walk on $\mathbb{Z}$). The simple symmetric random walk on $\mathbb{Z}$ is clearly irreducible and, by Example 1.17 it is also recurrent. Consider the measure $\lambda_i = 1$ for all $i \in \mathbb{Z}$. Then

$$\lambda_i = \frac{1}{2}\lambda_{i-1} + \frac{1}{2}\lambda_{i+1}$$

so λ is invariant. Now Theorem 1.25 forces any invariant measure to be a scalar multiple of λ. Since $\sum_{i \in \mathbb{Z}} \lambda_i = \infty$, there can be no invariant distribution and the walk is therefore null recurrent, by Theorem 1.27.

¶ Example 1.30 (Success-run chain). Consider a *success-run chain* on $\mathbb{N}$, whose transition probabilities are given by

$$p_{i,i+1} = p_i > 0, \quad p_{i0} = q_i = 1 - p_i > 0 \ \ \forall\, i \geqslant 0\,.$$

Note that For any $n \geqslant 1$,

$$\mathbb{P}_0(\sigma_0 > n) = p_0 \cdots p_{n-1}\,.$$

Thus, the chian is recurrent if and only if $p = \prod_{i=0}^{\infty} p_i = 0$. On the other hand, we compute that

$$\mathbb{E}_0(\sigma_0) = 1 + \sum_{n=1}^{\infty} p_0 \cdots p_{n-1} \, .$$

- If $\mathbb{E}_0(\sigma_0) = \infty$, the chain is mull recurrent.
- If $\mathbb{E}_0(\sigma_0) < \infty$, the chain is positive recurrent.

EXERCISE

¶ EXERCISE 1.31. Show that Any Markov chian on finite space I has a invariant distribution.

¶ EXERCISE 1.32. Let P be a stochastic matrix on a finite set I. Show that a distribution π is invariant for P if and only if $\pi(I-P+A) = a$, where $A = (a_{ij} : i, j \in I)$ with $a_{ij} = 1$ for all i and j, and $a = (a_i : i \in I)$ with $a_i = 1$ for all i. Deduce that if P is irreducible then $I - P + A$ is invertible. *Note that this enables one to compute the invariant distribution by any standard method of inverting a matrix.*

¶ EXERCISE 1.33. Prove that for any $i, j \in I$,

$$\gamma_j^i = \frac{\mathbb{P}_i(\sigma_j < \sigma_i)}{\mathbb{P}_j(\sigma_i < \sigma_j)}$$

1.6 Time reversal and detailed balance condition

For Markov chains, the past and future are independent given the present. This property is symmetrical in time and suggests looking at Markov chains with time running backwards. On the other hand, convergence to equilibrium (see Theorem 1.39) shows behaviour which is asymmetrical in time: *a highly organised state such as a point mass decays to a disorganised one, the invariant distribution*. This is an example of entropy increasing. It suggests that if we want complete time-symmetry we must begin in equilibrium.

We want to show that a Markov chain in equilibrium, run backwards, is again a Markov chain (the transition matrix may however be different). When discussing the time revarsal, without loss of genereality, we assume P is irreducible, since all the mass is concentrated at the closed positive recurrent classes when the chain in equilibrium.

Theorem 1.29. *Let P be irreducible and have an invariant distribution π. Suppose that $(X_n)_{0\leqslant n\leqslant N}$ is Markov(π, P) and set $Y_n = X_{N-n}$. Then $(Y_n)_{0\leqslant n\leqslant N}$ is Markov$(\pi, \hat{P})$, where $\hat{P} = (\hat{p}_{ij})_{i,j\in I}$ is given by*

$$\hat{p}_{ji} = \frac{\pi_i p_{ij}}{\pi_j} \quad \text{for all } i, j \in I\,.$$

and $\widehat{P}$ is also irreducible with invariant distribution π.

Proof. We need to calculate the conditional probability. For any $0 < n < N$,

$$\begin{aligned}
&\mathbb{P}\left(Y_{n+1} = j \mid Y_n = i, Y_{n-1} = i_{n-1}, \cdots, Y_0 = i_0\right)\\
&= \mathbb{P}\big(X_{N-(n+1)} = j \mid X_{N-n} = i, X_{N-n+1} = i_{n-1}, \cdots, X_N = i_0\big)\\
&= \mathbb{P}\big(X_{N-(n+1)} = j \mid X_{N-n} = i\big) = \frac{\pi_i p_{ij}}{\pi_j}\,.
\end{aligned}$$

This shows $(Y_n)_{0\leqslant n\leqslant N}$ is a Markov chain with the indicated transition probability. It's easy to check π is invariant for $\hat{P}$. □

The chain $(Y_n)_{0\leqslant n\leqslant N}$ is called the **time-reversal** of $(X_n)_{0\leqslant n\leqslant N}$. We say that $(X_n)_{n\geqslant 0}$ is reversible if, for all $N \geqslant 1$, $(X_{N-n})_{0\leqslant n\leqslant N}$ is also Markov

(λ, P), which is equilivant to

$$\pi_i p_{ij} = \pi_j p_{ji}, \, , \quad \text{for all } i, j \, .$$

Definition 1.5. A stochastic matrix P and a measure λ are said to be in **detailed balance** if

$$\lambda_i p_{ij} = \lambda_j p_{ji} \, , \quad \text{for all } i, j \, .$$

λ is called **symmetric measure** (配称测度) and P is **symmetrizable** (可配称). In addition, if π is distribution, P and π in detailed balance, we say π is an **reverable distribution** for P.

Remark.

(i) Obviously, if P and λ are in detailed balance,then λ is invariant for P. So when a solution λ to the detailed balance equations exists, it is often easier to find by the detailed balance equations than by the equation $\lambda = \lambda P$.

(ii) Let P be an irreducible stochastic matrix and π is the invariant distribution. Let $(X_n)_{n\geqslant 0}$ be Markov(π, P). $(X_n)_{n\geqslant 0}$ is reversible if and only if P and π are in detailed balance.

Theorem 1.30 (Kolmogorov cycle condition). *P is irreducible and has stationary distribution π. Then π is reverable if and only if the cycle condition is satisfied: given any cycle of states $i_0, i_1, \cdots, i_n = i_0$, we have*

$$p_{i_0 i_1} \cdots p_{i_{n-1} i_n} = p_{i_n i_{n-1}} \cdots p_{i_1 i_0} \, .$$

Proof. *Necessity* : Detailed balance implies

$$\pi_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} = \pi_{i_n} p_{i_n i_{n-1}} \cdots p_{i_1 i_0} \, .$$

Since $i_n = i_0$, and $\pi_{i_0} > 0$, the cycle condition holds.

Sufficiency : Suppose that the cycle condition holds. Let $i \in I$ and set $\lambda_i = 1$. For $j \neq i$ in I let $i = i_0, \cdots, i_k = j$ be a path from i to j with $p_{i_{n-1},i_n} > 0$ for $1 \leqslant n \leqslant k$ (and hence $p_{i_n,i_{n-1}} > 0$ for $1 \leqslant n \leqslant k$). Let

$$\lambda_j = \frac{p_{i_0,i_1} \cdots p_{i_{k-1},i_k}}{p_{i_k,i_{k-1}} \cdots p_{i_1,i_0}}.$$

The first step is to show that λ_j is well defined, i.e., is independent of the path chosen. Let $l_0 = i, \cdots, l_m = j$ be another path from i to j with $p_{l_{n-1},l_n} > 0$ for $1 \leqslant n \leqslant m$ (and hence $p_{l_n,l_{n-1}} > 0$ for $1 \leqslant n \leqslant m$). Combine these to get a loop that begins and ends at i. Thus, by cycle condition

$$p_{i_0,i_1} \cdots p_{i_{k-1},i_k} p_{l_m l_{m-1}} \cdots p_{l_1,l_0} = p_{l_0,l_1} \cdots p_{l_{m-1} l_m} p_{i_k,i_{k-1}} \cdots p_{i_1,i_0}$$

Which means that

$$\frac{p_{i_0,i_1} \cdots p_{i_{k-1},i_k}}{p_{i_k,i_{k-1}} \cdots p_{i_1,i_0}} = \frac{p_{l_0,l_1} \cdots p_{l_{m-1} l_m}}{p_{l_m l_{m-1}} \cdots p_{l_1,l_0}}.$$

This shows that the definition is independent of the path chosen. Obciously, λ and P are in detailed balance. Since P is irreducible, has a invariant distribution π, λ is a scalar mutipile of π. Then π and P are in detailed balance, so π is reverable. $\square$

1.7 Long-Run behavior of irreducible chains (I): egodic theorem

Ergodic theorems concern the limiting behaviour of averages over time. We shall prove a theorem which identifies for Markov chains the long-run proportion of time spent in each state. An essential tool is the following ergodic theorem for independent random variables which is a version of the strong law of large numbers.

Theorem (Strong law of large numbers). *Let $\{\xi_n\}$ be a sequence of independent, identically distributed, random variables with $\mathbb{E}\xi_1$ exists. Then as $n \to \infty$,*

$$\frac{\xi_1 + \cdots + \xi_n}{n} \to \mathbb{E}\xi_1 \ \ a.s.$$

We denote by $V_i(n)$ the **number of visits to i before** n,

$$V_i(n) := \sum_{k=0}^{n-1} 1_{\{X_k=i\}}$$

Then $\frac{V_i(n)}{n}$ can be interpreted as the *proportion of time before n spent in state i.*, or the *average number of times the chian appears at state i* . The following result gives the long-run proportion of time spent by a Markov chain in each state.

Theorem 1.31 (Ergodic theorem). *$(X_n)_{n\geqslant 0}$ is Markov(λ, P) and P is irreducible. Then as $n \to \infty$,*

$$\frac{V_i(n)}{n} \to \frac{1}{m_i}\,,\ \mathbb{P}\text{-a.s.} \tag{1.11}$$

where $m_i = \mathbb{E}_i\left(\sigma_i\right)$ is the expected return time to state i.

Proof. If P is transient, then, with probability 1, the total number V_i of visits to i is finite, so

$$\frac{V_i(n)}{n} \leqslant \frac{V_i}{n} \to 0 = \frac{1}{m_i}\,.$$

Suppose then that P is recurrent. Fix a state i, we have $\mathbb{P}(\sigma_i < \infty) = 1$. So $(X_{\sigma_i+n})_{n\geqslant 0}$ is Markov (δ_i, P) and independent of $(X_0, \cdots, X_{\sigma_i})$ by the strong Markov property. The long-run proportion of time spent in i is the same for $(X_{\sigma_i+n})_{n\geqslant 0}$ and $(X_n)_{n\geqslant 0}$, so it suffices to consider the case $\lambda = \delta_i$.

Write $\sigma_i^{(r)}$ for the length of the r th excursion to i, as in Section 1.4. By Corollary 1.12 , the non-negative random variables $\sigma_i^{(1)}, \sigma_i^{(2)}, \cdots$ are independent and identically distributed with $\mathbb{E}_i(\sigma_i) = m_i$. Notw that

$$\sigma_i^{(1)} + \cdots + \sigma_i^{(V_i(n)-1)} \leqslant n - 1$$

the left-hand side being the time of the last visit to i before n. Also

$$\sigma_i^{(1)} + \cdots + \sigma_i^{(V_i(n))} \geqslant n$$

the left-hand side being the time of the first visit to i after $n-1$. Hence

$$\frac{\sigma_i^{(1)} + \cdots + \sigma_i^{(V_i(n)-1)}}{V_i(n)} \leqslant \frac{n}{V_i(n)} \leqslant \frac{\sigma_i^{(1)} + \cdots + \sigma_i^{(V_i(n))}}{V_i(n)} \tag{1.12}$$

By the strong law of large numbers

$$\mathbb{P}_i\left(\frac{\sigma_i^{(1)} + \cdots + \sigma_i^{(n)}}{n} \to m_i \text{ as } n \to \infty\right) = 1$$

and, since P is recurrent $\mathbb{P}_i(V_i(n) \to \infty \text{ as } n \to \infty) = 1$. So, letting $n \to \infty$ in (1.12), we get $\mathbb{P}_i\left(\frac{n}{V_i(n)} \to m_i \text{ as } n \to \infty\right) = 1$, which implies

$$\mathbb{P}_i\left(\frac{V_i(n)}{n} \to \frac{1}{m_i} \text{ as } n \to \infty\right) = 1. \qquad \square$$

Ergodic theorem implies the *uniqueness* of invariant distribution for irreducible Markov chain.

Corollary 1.32. *P is irreducible, π is an invariant distribution for P. Then $\pi_i = \frac{1}{m_i}$ for all state i.*

Proof. Note that for any state i,

$$\mathbb{E}_\pi \frac{V_i(n)}{n} = \mathbb{E}_\pi \frac{1}{n} \sum_{k=0}^{n-1} 1_{\{X_k = i\}} = \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{P}_\pi (X_k = i) = \pi_i$$

From ergodic theorem we konw $V_i(n)/n \to \pi_i$ $\mathbb{P}_\pi$-a.s., by domainted limits theorem we konw

$$\lim_{n\to\infty} \mathbb{E}_\pi \frac{V_i(n)}{n} = \mathbb{E}_\pi \left(\lim_{n\to\infty} \frac{V_i(n)}{n} \right) = \frac{1}{m_i}$$

Thus $\pi_i = \frac{1}{m_i}$. ☐

Corollary 1.33. *$(X_n)_{n\geqslant 0}$ is Markov (λ, P), and P is irreducible, π is the (unique) stationary distribution for P. Then we have*

$$\sum_{i\in I} \left| \frac{V_i(n)}{n} - \pi_i \right| \to 0\,,\ \mathbb{P}\text{-a.s.}$$

Proof. Given $\varepsilon > 0$, choose $J \subset I$ finite so that $\sum_{i\notin J} \pi_i < \epsilon$. By Theorem 1.31 and Corollary 1.32, we have $\sum_{i\in J} \left| \frac{V_i(n)}{n} - \pi_i \right| \to 0$, $\mathbb{P}$-a.s. And

$$\sum_{i\in I} \left| \frac{V_i(n)}{n} - \pi_i \right| = \sum_{i\in J} \left| \frac{V_i(n)}{n} - \pi_i \right| + \sum_{i\notin J} \left| \frac{V_i(n)}{n} - \pi_i \right|$$

Note that

$$\begin{aligned}
\sum_{i\notin J} \left| \frac{V_i(n)}{n} - \pi_i \right| &\leqslant \sum_{i\notin J} \frac{V_i(n)}{n} + \sum_{i\notin J} \pi_i \\
&\leqslant \sum_{i\notin J} \frac{V_i(n)}{n} - \sum_{i\notin J} \pi_i + 2\epsilon \\
&\leqslant \sum_{i\in J} \left| \frac{V_i(n)}{n} - \pi_i \right| + 2\epsilon
\end{aligned}$$

Let $n \to \infty$, since ϵ is arbitrary we have $\sum_{i\in I} \left| \frac{V_i(n)}{n} - \pi_i \right| \to 0\,,\ \mathbb{P}$-a.s. ☐

Remark. Denote $\mathcal{M}$ as all the distribution on I. For any $\mu, \nu \in \mathcal{M}$, define

$$d_{TV}(\mu, \nu) := \frac{1}{2} \sum_{i \in I} |\mu_i - \nu_i| \tag{1.13}$$

and call it the **total variation distance** between μ and ν. It's easy to compute that

$$d_{TV}(\mu, \nu) = \sup_{A \subset I} |\mu(A) - \nu(A)| = \sup_{0 \leqslant f_i \leqslant 1, \forall i} \left| \sum_i (\mu_i - \nu_j) f_i \right| \tag{1.14}$$

So Corollary 1.33 means that the "statistics" distribution is convergent to the stationary distribution in d_{TV}, with probability one.

Back to the proof of Corollary 1.33, in fact we had got the following conclusion:

Proposition 1.34. *$\mathcal{M}$ as all the distribution on I. For any $\nu_n, \mu \in \mathcal{M}$, $\nu_n \to \mu$ in d_{TV} if and only if $\nu_n\{i\} \to \mu\{i\}$ for all $i \in I$.*

Now using Corollary 1.33, we have the following

Theorem 1.35 (Ergodic theorem). *$(X_n)_{n \geqslant 0}$ is Markov (λ, P) is irreducible and positive recurrent. π is the (unique) invariant distribution for P. For any bounded function $f : I \to \mathbb{R}$ we have*

$$\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \to \int_I f \, \mathrm{d}\pi \,, \ \mathbb{P}\text{-a.s.} \tag{1.15}$$

Remark. 将 f 看做对每个状态的观测, 那么遍历定理是在说, 观测的时间平均 (的极限) 等于空间平均.

Proof. Note that

$$\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \frac{1}{n} \sum_{k=0}^{n-1} \sum_{i \in I} 1_{\{X_k = i\}} f(i) = \sum_{i \in I} \frac{V_i(n)}{n} f(i)$$

and $\int_I f \, \mathrm{d}\pi = \sum_{i \in I} \pi_i f(i)$. Then we have

$$\left| \frac{1}{n} \sum_{k=0}^{n-1} f\left(X_k\right) - \int_I f \, \mathrm{d}\pi \right| \leqslant \sum_{i \in I} \left| \frac{V_i(n)}{n} - \pi_i \right| |f(i)| \,.$$

Therefore, as $n \to \infty$,

$$\left| \frac{1}{n} \sum_{k=0}^{n-1} f\left(X_k\right) - \int_I f \, \mathrm{d}\pi \right| \to 0 \,, \ \mathbb{P}\text{-a.s.} \qquad \square$$

Remark.

(i) In fact, for any $f \in L_1(\pi)$, this theorem holds.

The key idea here is that by *breaking the path at the return times to some state i, we get a sequence of i.i.d. random variables to which we can apply the law of large numbers.*

Proof. With out loss of generality, we assume f is nonnegative. Take any state i , let $T_i^{(0)} = 0$ and $T_i^{(k)} = \inf\{n \geqslant T_i^{(k-1)} + 1 : X_n = i\}$ be the k-th passage time.

$$W_k = \sum_{m=T_i^{(k)}}^{T_i^{(k+1)}-1} f\left(X_m\right) \,, \ k \geqslant 0$$

By the strong Markov property, the random variables $W_1, W_2, \cdots$ are i.i.d.. When computing $\mathbb{E}W_1$, we want to change the order of summation and integration, so we use Fubini theorem :

$$\begin{aligned}
\mathbb{E}W_1 = \mathbb{E} \sum_{m=T_i^{(1)}}^{T_i^{(2)}-1} f\left(X_m\right) &= \mathbb{E}_i \sum_{m=0}^{\sigma_i - 1} f\left(X_m\right) \\
&= \mathbb{E}_i \sum_{j \in I} V_j(\sigma_i) f(j) = \sum_{j \in I} \mathbb{E}_i V_j(\sigma_i) \, f(j) \\
&= \sum_{j \in I} \gamma_j^i f(j) = \sum_{j \in I} \frac{\pi_j}{\pi_i} \, f(j) = \frac{1}{\pi_i} \, \mathbb{E}_\pi f < \infty \,.
\end{aligned}$$

Using the law of large numbers for $W_1, W_2, \cdots$, since W_0 is finete a.s., then as $r \to \infty$

$$\frac{1}{r}\sum_{k=0}^{r} W_k = \frac{1}{r}\sum_{m=0}^{T_i^{(r)}-1} f\left(X_m\right) \to \frac{1}{\pi_i}\int_I f \,\mathrm{d}\pi \ , \quad \mathbb{P}\text{-a.s.}$$

For any $n \in \mathbb{N}_+$, let $V_i(n) = r$, then we have $T_i^{(r-1)} \leqslant n-1 < T_i^{(r)}$, so

$$\frac{1}{n}\sum_{m=0}^{n-1} f(X_m) = \frac{r-1}{n}\cdot\frac{1}{r-1}\sum_{m=0}^{T_i^{(r-1)}-1} f(X_m) + \frac{1}{n}\sum_{m=T_i^{(r-1)}}^{n-1} f(X_m) \tag{1.16}$$

Using Theorem 1.31, as $n \to \infty$, we have

$$\frac{r-1}{n} = \frac{V_i(n)-1}{n} \to \pi_i \ , \quad \mathbb{P}\text{-a.s.}$$

So as $n \to \infty$, $r \to \infty$, hence as $n \to \infty$

$$\frac{1}{r-1}\sum_{m=0}^{T_i^{(r-1)}-1} f(X_m) \to \frac{1}{\pi_i}\int_I f \,\mathrm{d}\pi \ , \quad \mathbb{P}\text{-a.s.}$$

Thus the first term on RHS in (1.16) converges to $\mathbb{E}_\pi f$ a.s.. We only need to show that the second term converges to zero. Note that

$$\frac{1}{n}\sum_{m=T_i^{(r-1)}}^{n-1} f(X_m) \leqslant \frac{1}{n}\sum_{m=T_i^{(r-1)}}^{T_i^{(r)}-1} f(X_m) = \frac{W_{r-1}}{n}\ ,$$

and using strong law of large number, as $r \to \infty$ we have

$$\frac{W_{r-1}}{n} = \frac{r-1}{n}\,\frac{Y_{r-1}}{r-1} - \frac{r-2}{n}\,\frac{Y_{r-2}}{r-2} \to \int_I f \,\mathrm{d}\pi - \int_I f \,\mathrm{d}\pi = 0\,, \ \mathbb{P}\text{-a.s.}$$

where $Y_r \coloneqq \sum_{k=1}^{r} W_k$. Now we get the desired result. $\square$

(ii) For what reason this theorem is called "ergodic"? In fact , this theorem has a close links with the following assertion: $A \subset I^\infty$ and $\mathbb{P}_\pi(X \in A) > 0$, where $X = (X_n)_{n\geqslant 0}$. Let $Y_k \coloneqq (X_{n+k})_{n\geqslant 0}$, then

$$\mathbb{P}_\pi(\exists\ k \text{ s.t. } Y_k \in A) = 1\,.$$

This is “ergodic”. In fact, for any integrable $f : I^\infty \to \mathbb{R}$, we have

$$\frac{1}{n}\sum_{k=0}^{n-1} f(Y_k) \to \mathbb{E}_{\mathbb{P}_\pi} f(Y_0)\,.$$

¶ Example 1.34. Assume (X_n) is irreducible with stationary distribution π. We give a “genelization” of Theorem 1.31, denote

$$V_{ij}(n) = \sum_{m=0}^{n-1} 1_{\{X_m = i, X_{m+1}=j\}}\,.$$

Then with probability one

$$\frac{V_{ij}(n)}{n} \to \pi_i p_{ij}\,, \quad \text{as } n \to \infty$$

To see this, let $Y_n = (X_n, X_{n+1})$, then Y_n is a Markov chain on $\widetilde{I} := \{(i,j) : p_{ij} > 0\}$, and $\{\widetilde{\pi}_{(i,j)} = \pi_i p_{ij} : (i,j) \in \widetilde{I}\}$ is a stationary distribution. Then $V_{ij}(n) = \sum_{m=0}^{n-1} 1_{\{Y_m = (i,j)\}}$. Using Theorem 1.31 we get the required result.

¶ Example 1.35 (Eestimating the transition probability). We consider now the statistical problem of estimating an unknown recurrent transition matrix P on the basis of observations of the corresponding Markov chain. Consider, to begin, the case where we have n observations $(X_0, \cdots, X_{n-1})$. The log-likelihood function is given by

$$l(P) = \log\left(\lambda_{X_0} p_{X_0 X_1} \cdots p_{X_{n-1} X_n}\right) = \sum_{i,j \in I} V_{ij}(n) \log p_{ij}$$

up to a constant independent of P, where $V_{ij}(n)$ is the number of transitions from i to j. before time n, i.e.,

$$V_{ij}(n) = \sum_{m=0}^{n-1} 1_{\{X_m = i, X_{m+1}=j\}}\,.$$

A standard statistical procedure is to find the maximum likelihood estimate $\widehat{P}$, which is the choice of P maximizing $l(P)$. Since P must satisfy the linear constraint $\sum_j p_{ij} = 1$ for each i, and $p_{ij} \geqslant 0$. Using Lagrange multipliers, we first try to maximize

$$l(P) + \sum_{i,j\in I} \lambda_i p_{ij}$$

and then choose $(\lambda_i : i \in I)$ to fit the constraints. Thus we find

$$\widehat{p}_{ij} = \frac{V_{ij}(n)}{V_i(n)}\,,$$

which is the proportion of jumps from i which go to j.

We now turn to consider the *consistency* of this sort of estimate, that is to say whether $\widehat{p}_{ij} \to p_{ij}$ with probability one as $n \to \infty$. To see this, we only need to combine Theorem 1.31 and Example 1.34.

Another way is using the idea in remark of ergodic theorem. For any $k = 1, 2, \cdots$, let

$$\xi_k := 1_{\{X_{T_i^{(k)}+1}=j\}}\,.$$

Then $V_{ij}(n) = \sum_{k=1}^r \xi_k$, where $r = \sum_{m=1}^{n-1} 1_{\{X_m=i\}} = V_i(n) - 1_{\{x_0=i\}}$. Using strong Markov property we konw $\{\xi_k\}$ are i.i.d. r.v.'s with mean p_{ij}. So, by the strong law of large numbers, with probability one

$$\frac{\sum_{k=1}^r \xi_k}{r} \to p_{ij}\,, \quad \text{as } r \to \infty$$

Note that when $n \to \infty$, $V_i(n) \to \infty$. Thus

$$\frac{V_{ij}(n)}{V_i(n)} = \frac{\sum_{k=1}^r \xi_k}{r+1} \to p_{ij}\,, \quad \text{as } n \to \infty$$

EXERCISE

¶ Exercise 1.36. $(X_n)_{n\geqslant 0}$ is Markov(λ, P). Show that as $n \to \infty$,

$$\frac{V_i(n)}{n} \to \frac{\mathbb{P}(\tau_i < \infty)}{m_i}\,,\ \mathbb{P}\text{-a.s.}$$

where $m_i = \mathbb{E}_i(\sigma_i)$ is the expected return time to state i.

1.8 Long-run behavior of irreducible chains (II): limits of n-step transition probability

In this section, we shall investigate the limiting behaviour of the n-step transition probabilities $p_{ij}^{(n)}$ as $n \to \infty$ for an Markov chain.

(i) If i doesn't lead to j, $p_{ij}^{(n)} = 0$ for all n.

(ii) If j is transient, by $\sum_n p_{ij}^{(n)} \leqslant \sum_n p_{jj}^{(n)} < \infty$ we have $p_{ij}^{(n)} \to 0$ as $n \to \infty$.

(iii) If $i \to j$ and i, j is recurrent, then i, j are communicating. In this case, we can assume P is irreducible.

(iv) If $i \to j$, i is transient, j is recurrent. By first decomposition (see Exercise 1.21) we have

$$p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)} \text{ for } n \geqslant 1.$$

Then the limitting behaviour of $p_{ij}^{(n)}$ can be determined by the limitting behaviour of $p_{jj}^{(n)}$. (See Exercise 1.40) When discussing $p_{jj}^{(n)}$ we can assume P is irreducible.

Therefore, in this section we always assume P is irreducible.

1.8.1 Periodicity

From ergodic theorem we konw that, if P is irreducible, then for any state i, we have

$$\frac{V_j(n)}{n} \to \frac{1}{m_j}, \ \mathbb{P}_i - \text{a.s.}$$

By Lebesgue domainted convergence theorem

$$\frac{1}{n}\sum_{k=0}^{n-1} p_{ij}^{(k)} \to \frac{1}{m_j}.$$

It's natual to guess that if $p_{ij}^{(n)} \to \frac{1}{m_j}$ as $n \to \infty$. If P is positive recurrent, that is $p_{ij}^{(n)} \to \pi_j$, where π is the unique invariant distribution for P. If P is null recurrent or transient, that is $p_{ij}^{(n)} \to 0$. Let's see some examples first.

¶ EXAMPLE 1.37. Let I be finite. Suppose for some $i \in I$ that

$$p_{ij}^{(n)} \to \pi_j \text{ as } n \to \infty \text{ for all } j \in I.$$

Then $\pi = (\pi_j : j \in I)$ is an invariant distribution.

As we saw in Example 1.37, if the state-space is finite and if for some i the limit exists for all j, then it must be an invariant distribution. But, as the following example shows, the limit does not always exist.

¶ EXAMPLE 1.38. Consider the two-state chain with transition matrix

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

Then $P^2 = I$, so $P^{2n} = I$ and $P^{2n+1} = P$ for all n. Thus $p_{ij}^{(n)}$ fails to converge for all i, j.

The behaviour of the chain in Example 1.38 is connected with its *periodicity*.

Definition 1.6. An integer d is called the **period** of state i if d is the largest common divisor of $\left\{n \geqslant 0 : p_{ii}^{(n)} > 0\right\}$. A state i is called **aperiodic** if i has period one.

Lemma 1.36. *If state i has period d, then $p_{ii}^{(nd)} > 0$ for sufficiently large n,i.e., there is $N_i \in \mathbb{N}_+$, for any $n \geqslant N_i$, $p_{ii}^{(nd)} > 0$.*

Proof. Note that d is the largest common divisor of $\left\{n \geqslant 0 : p_{ii}^{(n)} > 0\right\}$, so there exists some $n_1, \cdots, n_k$ such that $p_{ii}^{(n_k)} > 0$, and $d = (n_1, \cdots, n_k)$. By Bezout identity, there exist some $a_, \cdots, a_k \in \mathbb{Z}$ such that

$$a_1 \frac{n_1}{d} + \cdots a_k \frac{n_k}{d} = 1\,.$$

For any n, Let $m(\frac{n_1}{d}+\cdots+\frac{n_k}{d}) \leqslant n < (m+1)(\frac{n_1}{d}+\cdots+\frac{n_k}{d})$, then $\delta \coloneqq n - m(\frac{n_1}{d}+\cdots+\frac{n_k}{d}) < (\frac{n_1}{d}+\cdots+\frac{n_k}{d})$. We have

$$\begin{aligned} n &= m(\frac{n_1}{d}+\cdots+\frac{n_k}{d}) + \delta(a_1\frac{n_1}{d}+\cdots a_k\frac{n_k}{d}) \\ &= (m+a_1\delta)\frac{n_1}{d}+\cdots(m+a_k\delta)\frac{n_k}{d} \\ &=: b_1\frac{n_1}{d}+\cdots b_k\frac{n_k}{d}\,. \end{aligned}$$

where $b_1 = m + a_1\delta, \cdots, b_k = m + a_k\delta$. For sufficiently large n, we can let $b_1, \cdots, b_k \geqslant 1$, so

$$p_{ii}^{(nd)} \geqslant (p_{ii}^{(n_1)})^{b_1}\cdots(p_{ii}^{(n_1)})^{b_k} > 0\,. \qquad \square$$

Lemma 1.37. *periodicity is a class property, i.e., i communicates with j , then they have the same period.*

Proof. Assume i has period d_i and j has period d_j. Since i communicates with j, there exists some positive integers m and l such that $p_{ij}^{(m)} > 0$ and $p_{ji}^{(l)} > 0$. So when n is sufficiently large,

$$p_{ii}^{(m+nd_j+l)} \geqslant p_{ij}^{(m)}p_{jj}^{(nd)}p_{ji}^{(l)} > 0\,.$$

Then we have $d_i \mid d_j$. Samely, $d_j \mid d_i$, So $d_i = d_j$. $\square$

Therefore, we can say a ireducible Markov chain has period d, that means every state has peeiod d.

Theorem 1.38. *Let P be irreducible has period d. There is a partition*

$$I = D_0 \cup D_1 \cup \ldots \cup D_{d-1} \tag{1.17}$$

such that (setting $D_{nd+r} = D_r$)

(i) $p_{ij}^{(n)} > 0$ *only if $i \in D_r$ and $j \in D_{r+n}$ for some r.*

(ii) *For any $r = 0, \cdots d-1$ and $i, j \in D_r$, $p_{ij}^{(nd)} > 0$ for all sufficiently large n, i.e., there is $N_{ij} \in \mathbb{N}_+$ such that for any $n \geqslant N_{ij}$ we have $p_{ij}^{(nd)} > 0$.*

Remark. From (i) we can see chat P can write as block diagonal matrix as the following form:

$$\begin{bmatrix} 0 & P|_{D_0,D_1} & & & \\ 0 & 0 & P|_{D_1,D_2} & & \\ \vdots & \vdots & \ddots & \ddots & \\ 0 & 0 & & \ddots & P|_{D_{d-2}D_{d-1}} \\ P|_{D_{d-1}D_0} & 0 & \cdots & & 0 \end{bmatrix}$$

From (ii), We can see that $D_0, \cdots D_{d-1}$ is all the communicating class of P^d, also P^d can write as block diagonal matrix as the following form:

$$\begin{bmatrix} P^d|_{D_0} & & \\ & \ddots & \\ & & P^d|_{D_{d-1}} \end{bmatrix}$$

Proof. Take any state k fixed, and for any $r = 0, \cdots d-1$,

$$D_r \coloneqq \{j : \exists\, n \text{ s.t. } p_{kj}^{(nd+r)} > 0\}.$$

Obviously, $\cup_{r=0}^{d-1} D_r = I$. Let $0 \leqslant r_1 \neq r_2 \leqslant d-1$, We need to show that $D_{r_1} \cap D_{r_2} = \varnothing$. If not, assume $p_{kj}^{(n_1 d+r_1)} > 0$ and $p_{kj}^{(n_2 d+r_2)} > 0$. Since P is irreducible, there is some integer l such that $p_{jk}^{(l)} > 0$. Then $p_{kk}^{(n_1 d+r_1+l)} > 0$ and $p_{kk}^{(n_2 d+r_2+l)} > 0$. So $d \mid r_1 + l$ and $d \mid r_2 + l$, then $d \mid (r_1 - r_2)$, $r_1 = r_2$. We arrive at a contradiction. Hence we have a partition.

Let $i \in D_r$ and $p_{ij}^{(n)} > 0$, we show that $j \in D_{r+n}$. Since $i \in D_r$, there exists n_1 such that $p_{kj}^{(n_1 d+r)} > 0$, so $p_{kj}^{(n_1 d+r+n)} \geqslant p_{ki}^{(n_1 d+r)} p_{ij}^{(n)} > 0$. Hence we have $j \in D_{r+n}$.

Let $i, j \in D_r$, we prove that $p_{ij}^{(nd)} > 0$ for sufficiently large n. Assume $p_{ik}^{(l)} > 0$, $p_{ki}^{(n_1 d+r)} > 0$ and $p_{kj}^{(n_2 d+r)} > 0$. Then $p_{ii}^{(n_1 d+r+l)} > 0$. We have $d \mid r+l$ because i has period d. Then for sufficiently large n, $p_{ij}^{(l+nd+n_2 d+r)} \geqslant p_{ik}^{(l)} p_{kk}^{(nd)} p_{kj}^{(n_2 d+r)} > 0$. Let $r+l+n_2 d = n_3 d$, that is $p_{ij}^{(nd+n_3 d)} > 0$. So $p_{ij}^{(nd)} > 0$ for sufficiently large n. $\square$

Theorem 1.38 表明从某一块区域中的任意状态出发, 走一步必然到下一块区域中. 应该强调, 仅有 *(i)* 成立, 而 *(ii)* 不成立不能保证 d 是马氏链的周期.

¶ EXAMPLE 1.39. Figure 1.6 中马氏链周期为 6, 但是其中的分块也满足 Theorem 1.38 的 (i).

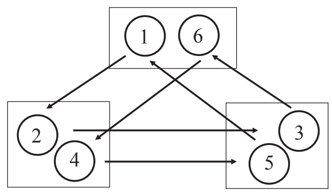


Figure 1.6: Counterexample

1.8.2 Limitting behavior of $p_{ij}^{(n)}$

Here is the main result of this section. The method of proof, by *coupling* two Markov chains, is ingenious.

Theorem 1.39 (Convergence to equilibrium). *Let P be irreducible, positive recurrent and aperiodic, π is the unique invariant distribution. Then for all $i, j \in I$,*

$$p_{ij}^{(n)} \to \pi_j\,, \text{ as } n \to \infty$$

In fact, suppose that $(X_n)_{n \geqslant 0}$ is Markov (λ, P), where λ is any initial distribution. Then $\mathbb{P}_{X_n} \to \pi$ in total variance distance d_{TV}, i.e.,

$$\sum_{j \in I} |\mathbb{P}(X_n = j) - \pi_j| \to 0, \quad \text{as } n \to \infty\,.$$

Proof. We use a coupling argument. Let $(Y_n)_{n \geqslant 0}$ be Markov (π, P) and independent of $(X_n)_{n \geqslant 0}$. Fix a reference state b and set

$$\tau = \inf\{n \geqslant 1 : X_n = Y_n = b\}\,.$$

Step 1. We show $\mathbb{P}(T < \infty) = 1$. The process $W_n = (X_n, Y_n)$ is a Markov chain on $I \times I$ with transition probabilities

$$\widetilde{p}_{(i,k)(j,l)} = p_{ij}p_{kl}$$

and initial distribution

$$\mu_{(i,k)} = \lambda_i \pi_k$$

since P is aperiodic, for all states i, j, k, l we have

$$\tilde{p}^{(n)}_{(i,k)(j,l)} = p^{(n)}_{ij} p^{(n)}_{kl} > 0$$

for all sufficiently large n; so P is irreducible. Also, P has an invariant distribution given by

$$\tilde{\pi}_{(i,k)} = \pi_i \pi_k \, .$$

So $\tilde{P}$ is positive recurrent. But τ is the first passage time of W_n to (b, b) so $\mathbb{P}(T < \infty) = 1$.

Step 2. Set

$$Z_n = \begin{cases} X_n, & \text{if } n \leqslant \tau. \\ Y_n, & \text{if } n > \tau. \end{cases}$$

The diagram below illustrates the idea. It's easy to show that $(Z_n)_{n \geqslant 0}$ is Markov (λ, P).

Step 3. We have

$$\mathbb{P}\left(Z_n = j\right) = \mathbb{P}\left(X_n = j \text{ and } n < T\right) + \mathbb{P}\left(Y_n = j \text{ and } n \geqslant T\right) \, .$$

Hence

$$\begin{aligned} \sum_{j \in I} \left|\mathbb{P}\left(X_n = j\right) - \pi_j\right| &= \sum_{j \in I} \left|\mathbb{P}\left(Z_n = j\right) - \mathbb{P}\left(Y_n = j\right)\right| \\ &= \sum_{j \in I} \left|\mathbb{P}\left(X_n = j, n \leqslant \tau\right) - \mathbb{P}\left(Y_n = j, n \leqslant \tau\right)\right| \\ &\leqslant \sum_{j \in I} \mathbb{P}\left(X_n = j, n \leqslant \tau\right) + \mathbb{P}\left(Y_n = j, n \leqslant \tau\right) \\ &= 2\mathbb{P}\left(n \leqslant \tau\right). \end{aligned}$$

and $\mathbb{P}(n \leqslant \tau) \to 0$ as $n \to \infty$. $\square$

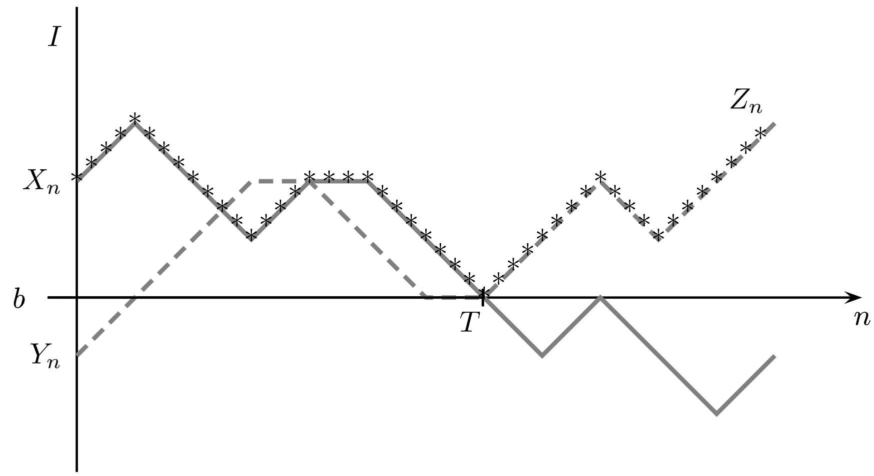


Figure 1.7: X_n, Y_n and Z_n

To understand this proof one should see what goes wrong when P is not aperiodic. Consider the two-state chain of Example 1.38 which has $(1/2, 1/2)$ as its unique invariant distribution. $(X_n)_{n\geqslant 0}$ We start $(X_n)_{n\geqslant 0}$ from 0 and $(Y_n)_{n\geqslant 0}$ with equal probability from 0 or 1. However, if $Y_0 = 1$, then, because of periodicity, $(X_n)_{n\geqslant 0}$ and $(Y_n)_{n\geqslant 0}$ will never meet, and the proof fails.

Theorem 1.40. *Let P be irreducible, aperiodic and null recurrent. Then for all $i, j \in I$,*

$$p_{ij}^{(n)} \to 0\,, \text{ as } n \to \infty\,.$$

In fact, for any initial distribution λ, suppose that $(X_n)_{n\geqslant 0}$ is Markov (λ, P), then

$$\mathbb{P}\left(X_n = j\right) \to 0, \quad \text{as } n \to \infty\,.$$

Proof. Return to the coupling argument used in Theorem 1.39 , only now let $(Y_n)_{n\geqslant 0}$ be Markov (μ, P), where μ is to be chosen later. Set $W_n = (X_n, Y_n)$. As before, aperiodicity of $(X_n)_{n\geqslant 0}$ ensures irreducibility of $(W_n)_{n\geqslant 0}$.

If $(W_n)_{n\geqslant 0}$ is transient then, on taking $\mu=\lambda$, we obtain

$$\mathbb{P}\left(X_n=j\right)^2=\mathbb{P}\left(W_n=(j,j)\right)\to 0$$

as required.

Now assume $(W_n)_{n\geqslant 0}$ is recurrent. Using the same argument in Theorem 1.39, we have $\mathbb{P}(\tau<\infty)=1$, we have that

$$\left|\mathbb{P}\left(X_n=j\right)-\mathbb{P}\left(Y_n=j\right)\right|\to 0\,.$$

We exploit this convergence by taking $\mu=\lambda P^k$ for $k\in\mathbb{N}_+$ so that

$$\left|\mathbb{P}\left(X_n=j\right)-\mathbb{P}\left(X_{n+k}=j\right)\right|\to 0\,. \tag{1.18}$$

($\star$) If we can prove that for any $\epsilon>0$, there is some $K=K_\epsilon$ such that

$$\min_{0\leqslant k\leqslant K}\mathbb{P}\left(X_{n+k}=j\right)<\epsilon,\quad\text{for any } n\in\mathbb{N}_+\,.$$

Then from (1.18) we konw for the ϵ given previously, there is $N=N_\epsilon>0$ such that for any $n\geqslant N$, there holds

$$\max_{0\leqslant k\leqslant K}\left|\mathbb{P}\left(X_n=j\right)-\mathbb{P}\left(X_{n+k}=j\right)\right|<\epsilon\,. \tag{1.19}$$

We can see that $\mathbb{P}\left(X_n=j\right)<2\epsilon$ for any $n\geqslant N$. So, $\mathbb{P}\left(X_n=j\right)\to 0$.

If ($\star$) doesn't hold, there is some $\epsilon_0>0$, for any positive integer K, there exist a $m=m(K)$ such that

$$\mathbb{P}\left(X_{m+k}=j\right)\geqslant\epsilon_0,\ k=0,1,\cdots,K\,.$$

Note that for any $k=0,1,2,\cdots K$,

$$\begin{aligned}&\mathbb{P}\left(X_{m+k}=j\right)\mathbb{P}_j\left(\sigma_j>K-k\right)\\&=\mathbb{P}\left(\sup\{\,n:X_n=j,\,n\leqslant m+K\,\}=m+k\right),\end{aligned}$$

and

$$\sum_{k=0}^{K}\mathbb{P}\Big(\sup\{\,n:X_n=j,\,n\leqslant m+K\,\}=m+k\Big)=1$$

So we have

$$\begin{aligned}1 &= \sum_{k=0}^{K} \mathbb{P}\left(X_{m+k}=j\right) \mathbb{P}_j\left(\sigma_j > K-k\right) \\ &\geqslant \epsilon_0 \sum_{k=0}^{K} \mathbb{P}_j\left(\sigma_j > K-k\right) = \epsilon_0 \sum_{k=0}^{K} \mathbb{P}_j\left(\sigma_j > k\right) .\end{aligned}$$

Since K is arbitrary, we have $\sum_{k=0}^{\infty} \mathbb{P}_j\left(\sigma_j > k\right) < \infty$. But P is null recurrent, we have

$$\mathbb{E}_j\left(\sigma_j\right) = \sum_{k=0}^{\infty} \mathbb{P}_j\left(\sigma_j > k\right) = \infty .$$

This is a contradiction. □

Combine Theorem 1.39 and Theorem 1.40, we have

Theorem 1.41. *Let P be irreducible, aperiodic. Then for all $i, j \in I$,*

$$p_{ij}^{(n)} \to \frac{1}{m_j}, \text{ as } n \to \infty . \tag{1.20}$$

In fact, for any initial distribution λ, suppose that $\left(X_n\right)_{n\geqslant 0}$ is Markov (λ, P), then

$$\mathbb{P}\left(X_n = j\right) \to \frac{1}{m_j}, \quad \text{as } n \to \infty .$$

We move on now to the cases that were excluded in the last theorem, where $\left(X_n\right)_{n\geqslant 0}$is periodic or transient or null recurrent. Here is a complete description of limiting behaviour for irreducible chains. This generalizes Theorem 1.39 in two respects since we require neither aperiodicity nor the existence of an invariant distribution.

Theorem 1.42. *Let P be irreducible of period d and let $D_0, D_1, \ldots, D_{d-1}$ be the partition obtained in Theorem 1.38. For any $i \in D_0$ and $r = 0, 1, \ldots, d-1$, and $j \in D_r$ we have*

$$p_{ij}^{(nd+r)} \to \frac{d}{m_j} \quad \text{as } n \to \infty.$$

In fact, Let λ be a distribution with $\sum_{i \in D_0} \lambda_i = 1$. Suppose that $(X_n)_{n \geqslant 0}$ is Markov(λ, P). Then for $j \in D_r$

$$\mathbb{P}\left(X_{nd+r} = j\right) \to \frac{d}{m_j}, \quad \text{as } n \to \infty$$

Proof. We reduce to the aperiodic case. Set $\nu = \lambda P^r$, by Theorem 1.38 we have

$$\sum_{i \in C_r} \nu_i = 1\,.$$

Set $Y_n = X_{nd+r}$, then $(Y_n)_{n \geqslant 0}$ is Markov $\left(\nu, P^d\right)$ and, by Theorem 1.38, P^d is irreducible and aperiodic on C_r. For $j \in C_r$ the expected return time of $(Y_n)_{n \geqslant 0}$ to j is m_j/d. By Theorem 1.41, in the aperiodic case, we ahve

$$\mathbb{P}\left(X_{nd+r} = j\right) = \mathbb{P}\left(Y_n = j\right) \to \frac{d}{m_j} \text{ as } n \to \infty$$

so the theorem holds in general. □

Corollary 1.43. *P is irreducible, then*

$$\limsup_{n \to \infty} p_{ij}^{(n)} = \frac{d}{m_j}\,, \tag{1.21}$$

where d is the period of P.

EXERCISE

¶ Exercise 1.40. Assume state $i, j \in I$ such that i is transient and j is recurrent.

(i) j is aperiodic, show that

$$p_{ij}^{(n)} \to \frac{\rho_{ij}}{m_j}\,, \text{ as } n \to \infty\,.$$

(ii) j has period d, show that

$$p_{ij}^{(nd+r)} \to \frac{d}{m_j}\left[\sum_{k=0}^{\infty} f_{ij}^{(kd+r)}\right] \quad \text{as} \quad n \to \infty$$

for any $r = 0, 1, \cdots, d-1$, where $f_{ij}^{(n)} = \mathbb{P}_i(\sigma_j = n)$.

Hint: Note that in Exercise 1.21 we proved that

$$p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)} \quad \text{for } n \geqslant 1$$

When j is apeiodic, for fixed k we have $p_{jj}^{(n-k)} \to 1/m_j$. Then use Lebesgue domainted convergence theorem: Let μ be a finite measure on $\mathbb{N}$ such that $\mu(k) = f_{ij}^{(k)}$. For any n, let $g_n(k) = p_{jj}^{(n-k)}$ for $k \leqslant n$ and $= 0$ for $k > n$, $g(k) = 1/m_j$ is constant function, thus we have $g_n \to g$ a.e. Note that $g_n \leqslant 1$, by Lebesgue domainted convergence theorem:

$$p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)} = \int_{\mathbb{N}} g_n \, d\mu \to \int_{\mathbb{N}} g \, d\mu = \sum_{k=1}^{\infty} f_{ij}^{(k)} \frac{1}{m_j} = \frac{\rho_{ij}}{m_j} \, .$$

Chapter 2

Continuous-time Markov chains

2.1 Right Continuous random processes

Let I be a countable set. A random vectors

$$X : (\Omega, \mathcal{F}) \to (I^{[0,\infty)}, \mathcal{P}(I)^{[0,\infty)})\,;\ \omega \mapsto (X_t(\omega))_{t\geqslant 0}$$

is called a **continuous-time random process with values in** I. Equivalently, a continuous-time random process is a family of random variables $X_t : \Omega \to I$ for all $t \geqslant 0$. Sometimes, we need

$$X : (\Omega \times [0,\infty)\,, \mathcal{F} \times \mathcal{B}[0,\infty)) \to \mathbb{R}\,;\ (\omega, t) \mapsto X_t(\omega)$$

is measurable.

We are going to consider ways in which we might specify the probabilistic behaviour (or law) of $(X_t)_{t\geqslant 0}$. These should enable us to find, at least in principle, any probability connected with the process, such as $\mathbb{P}\left(X_t = i \text{ for some } t\right)$. There are subtleties in this problem not present in the discrete-time case. They arise because, for a countable disjoint union

$$\mathbb{P}\left(\bigcup_n A_n\right) = \sum_n \mathbb{P}\left(A_n\right) .$$

whereas for an uncountable union $\cup_{t\geqslant 0} A_t$ there is no such rule. But

$$\{X_t = i \text{ for some } t \in [0,\infty)\} = \bigcup_{t\geqslant 0}\{X_t = i\}$$

To avoid these subtleties as far as possible we shall restrict our attention to processes $(X_t)_{t\geqslant 0}$ which are *right-continuous*.

Definition 2.1. A continuous-time random process $(X_t)_{t\geqslant 0}$ with values in I is called **right-continuous** if for any fixed $\omega \in \Omega$, $[0,\infty) \to I; t \mapsto X_t(\omega)$ is right-continuous, where I equipped with discrete topology.

Clearly, $(X_t)_{t\geqslant 0}$ is right-continuous means that, for each $\omega \in \Omega$ and $t \geqslant 0$, there exists $\delta > 0$, depending on ω and t, such that

$$X_s(\omega) = X_t(\omega)\,, \quad \text{for all } s \in [t, t+\delta]\,. \tag{2.1}$$

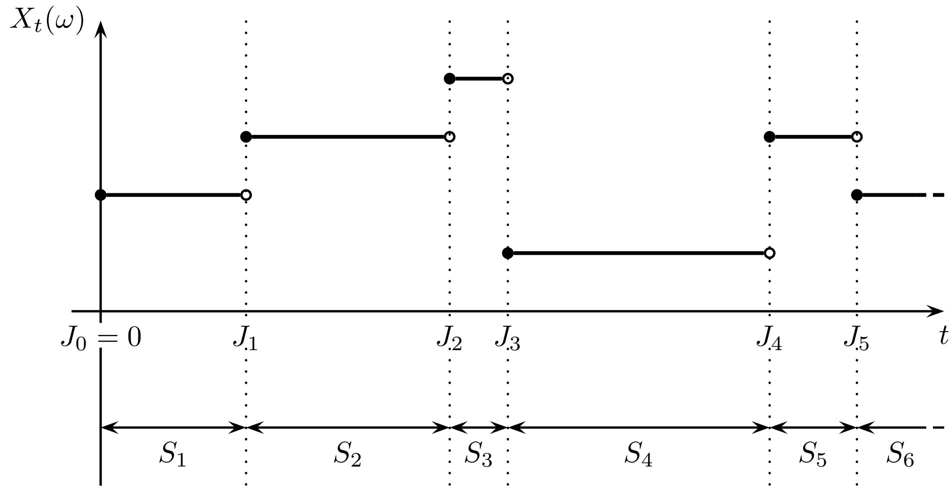
For example, we can now deduce that

$$\{X_t = i \text{ for some } t \in [0,\infty)\} = \bigcup_{t\geqslant 0}\{X_t = i\} = \bigcup_{r\in\mathbb{Q}_+}\{X_r = i\}\,.$$

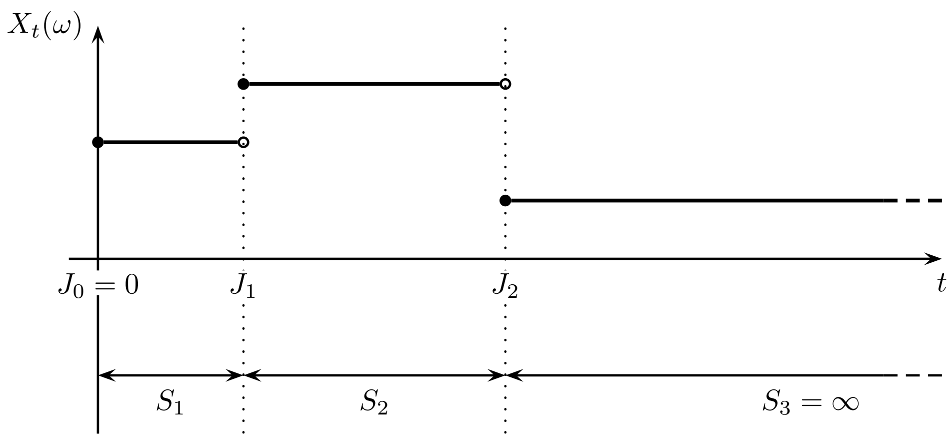
and following this we can compute the probability of $\{X_t = i \text{ for some } t \in [0,\infty)\}$.

Three possibilities for the sorts of path Every path $t \mapsto X_t(\omega)$ of a right-continuous process must remain constant for a while in each new state, so there are three possibilities for the sorts of path we get.

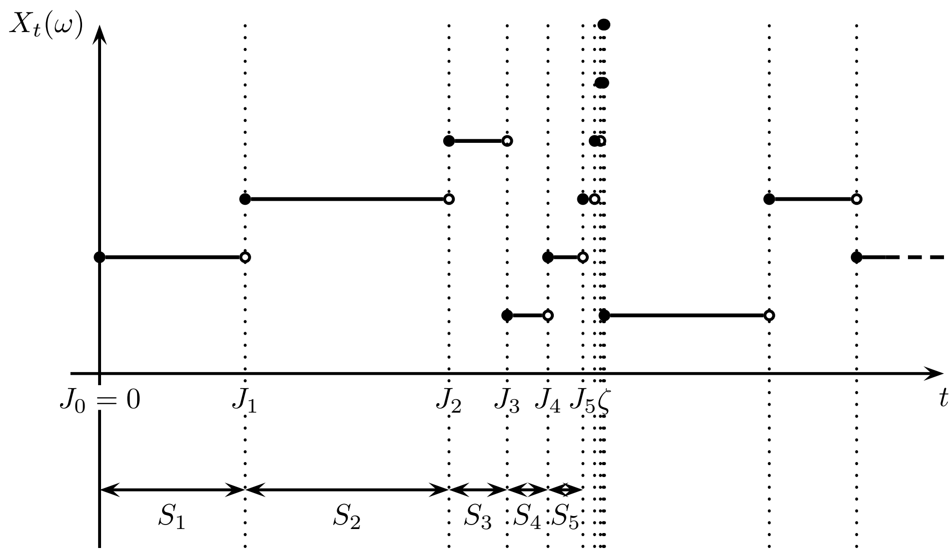
In the first case the path makes infinitely many jumps, but only finitely many in any interval $[0,t]$.



The second case is where the path makes finitely many jumps and then becomes stuck in some state forever:



In the third case the process makes infinitely many jumps in a finite interval; this is illustrated below. In this case, after the explosion time ζ the process starts up again; it may explode again, maybe infinitely often, or it may not.



Jump times, holding times We call $J_0, J_1, \cdots$ the *jump times* of $(X_t)_{t\geqslant 0}$ and $S_1, S_2, \cdots$ the *holding times*. They are obtained from $(X_t)_{t\geqslant 0}$ by

$$J_0 = 0, \quad J_{n+1} = \inf\{t \geqslant J_n : X_t \neq X_{J_n}\}$$

for $n = 0, 1, \cdots$, where $\inf \varnothing = \infty$, and, for $n = 1, 2, \cdots$

$$S_n = \begin{cases} J_n - J_{n-1} & \text{if } J_{n-1} < \infty \\ \infty & \text{otherwise.} \end{cases}$$

Note that right-continuity forces $S_n > 0$ for all n. If $J_{n+1} = \infty$ for some n, we define $X_\infty = X_{J_n}$, the final value, otherwise X_∞ is undefined.

Explosion time, embedded chain The *(first) explosion time* ζ is defined by

$$\zeta = \sup_n J_n = \sum_{n=1}^{\infty} S_n$$

The discrete-time process $(Y_n)_{n\geqslant 0}$ given by $Y_n = X_{J_n}$ is called the *jump process* of $(X_t)_{t\geqslant 0}$, or the *embedded chain (jump chain)* if it is a discrete-time Markov chain. This is simply the sequence of values taken by $(X_t)_{t\geqslant 0}$ up to explosion.

Minimal process We shall not consider what happens to a process after explosion. So it is convenient to adjoin to I a new state, ∞ (or ∂) say, and require that $X_t = \infty$ (or $X_t = \partial$) if $t \geqslant \zeta$. Any process satisfying this requirement is called *minimal*. The terminology "minimal" does not refer to the state of the process but to the interval of time over which the process is active.

Note that a minimal process may be reconstructed from its holding times and jump process. Thus by specifying the joint distribution of $S_1, S_2, \cdots$ and $(Y_n)_{n\geqslant 0}$ we have another "countable' specification of the probabilistic behaviour of $(X_t)_{t\geqslant 0}$. For example,

$$\{X_t = i \text{ for some } t \in [0,\infty)\} = \{Y_n = i \text{ for some } n \in \mathbb{N}\}\ .$$

Consider now the method of describing a minimal right-cont-continuous process $(X_t)_{t\geqslant 0}$ via its jump process $(Y_n)_{n\geqslant 0}$ and holding times $(S_n)_{n\geqslant 1}$. Let us take

$$\mathcal{F} = \sigma\left(X_t : t \geqslant 0\right) \text{ and } \mathcal{G} = \sigma\left((Y_n)_{n\geqslant 0}, (S_n)_{n\geqslant 1}\right)\ .$$

Firstly, for all $i \in I$

$$\{X_t = i\} = \bigcup_{n\geqslant 0} \{Y_n = i\} \cap \{J_n \leqslant t < J_{n+1}\} \in \mathcal{G}\,,$$

which deduces that $\mathcal{F} \subset \mathcal{G}$. Intuitively, $\mathcal{G} \subset \mathcal{F}$, but there's a little bit of troubles to prove it rigorously.

Denote by $\mathcal{F}_t$ the σ-algebra generated by $\{X_s : s \leqslant t\}$. We say that a random variable T with values in $[0,\infty]$ is a **stopping time of** $(X_t)_{t\geqslant 0}$ if $\{T \leqslant t\} \in \mathcal{F}_t$ for all $t \geqslant 0$. Note that this certainly implies

$$\{T < t\} = \bigcup_n \left\{T \leqslant t - \frac{1}{n}\right\} \in \mathcal{F}_t \quad \text{for all} \quad t \geqslant 0.$$

We define for stopping times T,

$$\mathcal{F}_T = \{A \in \mathcal{F} : A \cap \{T \leqslant t\} \in \mathcal{F}_t \text{ for all } t \geqslant 0\} \,.$$

This turns out to be the correct way to make precise the notion of sets which "depend only on $\{X_t : t \leqslant T\}$", somwtimes we use '$\sigma(\{X_t : t \leqslant T\})$' to indicate this.

Lemma 2.1. *Let T be a stopping time of $(X_t)_{t\geqslant 0}$. Then X_T is $\mathcal{F}_T$-measurable.*

Proof. In order to show X_T is $\mathcal{F}_T$-measurable, it suffices to show that for each $i \in I$,

$$\{X_T = i\} \cap \{T \leqslant t\} \in \mathcal{F}_t \,.$$

Clearly, $\{X_T = i\} \cap \{T < t\} \in \mathcal{F}_t$, so we need to prove

$$\{X_T = i\} \cap \{T < t\} \in \mathcal{F}_t \,.$$

Since $(X_t)_{t\geqslant 0}$ is right-continuous, on $\{T < t\}$ there exists an (random) $N \geqslant 0$ such that, for all $n \geqslant N$, there exists $k \geqslant 1$,

$$\frac{k-1}{2^n} \leqslant T < \frac{k}{2^n} < t \quad \text{and} \quad X_{\frac{k}{2^n}} = X_T.$$

Hence

$$\begin{aligned} &\{X_T = i\} \cap \{T < t\} \\ &= \bigcup_{N=0}^{\infty} \bigcap_{n=N}^{\infty} \bigcup_{k=1}^{[2^m t]} \left\{X_{\frac{k}{2^n}} = i\right\} \cap \left\{\frac{k-1}{2^n} \leqslant T < \frac{k}{2^n}\right\} \in \mathcal{F}_t \end{aligned}$$

so X_T is $\mathcal{F}_T$-measurable. □

Lemma 2.2. *Let S and T be stopping times of $(X_t)_{t\geqslant 0}$. Then*

$$\{S > T\} \in \mathcal{F}_T \,,\ \{S \leqslant T\} \in \mathcal{F}_T \,.$$

Proof. We have

$$\{S > T\} \cap \{T \leqslant t\} = \bigcup_{s\in\mathbb{Q}_+, s\leqslant t} \{T \leqslant s\} \cap \{S > s\} \in \mathcal{F}_t \,,$$

so $\{S > T\} \in \mathcal{F}_T$, and so $\{S \leqslant T\} \in \mathcal{F}_T$. □

Lemma 2.3. *For each $n \geqslant 0$, the jump time J_n is a stopping time of $(X_t)_{t\geqslant 0}$.*

Proof. Obviously, $J_0 = 0$ is a stopping time. Assume inductively that J_n is a stopping time. Then for all $t > 0$,

$$\{J_{n+1} < t\} = \bigcup_{s\in\mathbb{Q}_+, s\leqslant t} \{J_n < s\} \cap \{X_s \neq X_{J_n}\} \in \mathcal{F}_t$$

$$\{J_{n+1} = t\} = \bigcup_{s\in\mathbb{Q}_+, s\leqslant t} \{J_n < s\} \bigcap_{\substack{s'\in\mathbb{Q}_+ \\ s\leqslant s'<t}} \{X_s = X_{J_n}\} \in \mathcal{F}_t\,.$$

so J_{n+1} is a stopping time and the induction proceeds. $\square$

Now, by the three lemmas above, we $(Y_n)_{n\geq 0}$ and $(S_n)_{n\geq 1}$ are $\mathcal{F}$-measurable. Thus $\mathcal{G} \subset \mathcal{F}$. Therefore we get $\mathcal{F} = \mathcal{G}$, i.e.,

$$\sigma\left(X_t : t \geqslant 0\right) = \sigma\left((Y_n)_{n\geqslant 0}, (S_n)_{n\geqslant 1}\right)\,. \tag{2.2}$$

2.2 Poisson processes

Poisson processes are some of the simplest examples of continuous-time Markov chains. We shall also see that they may serve as *building blocks* for the most general continuous-time Markov chain. Moreover, a Poisson process is the natural probabilistic model for any uncoordinated stream ofdiscrete events in continuous time. So we shall study Poisson processes first, both as a gentle warm-up for the general theory and because they are useful in themselves. We shall begin with a definition in terms of jump chain and holding times.

Definition 2.2. A right-continuous process $(N_t)_{t \geqslant 0}$ with values in $\{0, 1, 2, \cdots\}$ is a **Poisson process of rate** λ, where $\lambda \in (0, \infty)$, if its holding times $S_1, S_2, \cdots$ are independent exponential random Variables of parameter λ and its jump chain is given by $Y_n = n$.

Remark. Given the distribution of the jump times and jump chian, the law of $(N_t)_{t \geqslant 0}$ is uniquely determined.

Here is the diagram of the Poisson process,

$$0 \xrightarrow{\lambda} 1 \xrightarrow{\lambda} 2 \xrightarrow{\lambda} 3 \xrightarrow{\lambda} 4 \longrightarrow \cdots$$

Figure 2.1: Poisson process of rate λ

and the associated Q-matrix is given by

$$Q = \begin{pmatrix} -\lambda & \lambda & & & \\ & -\lambda & \lambda & & \\ & & \ddots & \ddots & \\ & & & & \end{pmatrix} \tag{2.3}$$

Using the strong law of large numbers, we have $\mathbb{P}\left(J_n \to \infty\right) = 1$, so there is no explosion in Poisson process.

Why is the process defined above called "Poisson"?

Theorem 2.4. *For each* $t > 0$, N_t *has a* Poisson(λt) *distribution.*

Proof. For any $n \in \mathbb{N}$,

$$\mathbb{P}(N_t = n) = \mathbb{P}(J_n \leqslant t < J_{n+1}) = \int_0^t f_{J_n}(s)\mathbb{P}\left(S_{n+1} > t - s\right) ds$$

Note that $J_n = S_1 + \cdots S_n$ has a gamma(n, λ) distribution, that is

$$f_{J_n}(s) = \lambda e^{-\lambda s} \cdot \frac{(\lambda s)^{n-1}}{(n-1)!} 1_{\{s>0\}} .$$

So we have

$$\begin{aligned}\mathbb{P}(N(t) = n) &= \int_0^t \lambda e^{-\lambda s} \frac{(\lambda s)^{n-1}}{(n-1)!} \cdot e^{-\lambda(t-s)} ds \\ &= \frac{\lambda^n}{(n-1)!} e^{-\lambda t} \int_0^t s^{n-1} ds = e^{-\lambda t} \frac{(\lambda t)^n}{n!} ,\end{aligned}$$

which proves the desired result. □

Construct a Poisson process A simple way to construct a Poisson process of rate λ is to take a sequence $S_1, S_2, \cdots$ of independent exponential random Variables of parameter λ, to set $J_0 = 0$, $J_n = S_1 + \cdots + S_n$, then set

$$N_t = n \quad \text{if} \quad J_n \leqslant t < J_{n+1} \tag{2.4}$$

or equivalently

$$N_t = \sum_{n=1}^{\infty} 1_{\{J_n \leqslant t\}} \quad \text{for all } t \geqslant 0 . \tag{2.5}$$

The following diagram illustrates a typical path, and it's easy to check that $\{J_n\}$ is exactly the jump times of the right-continuous process $(N_t)_{t \geqslant 0}$.

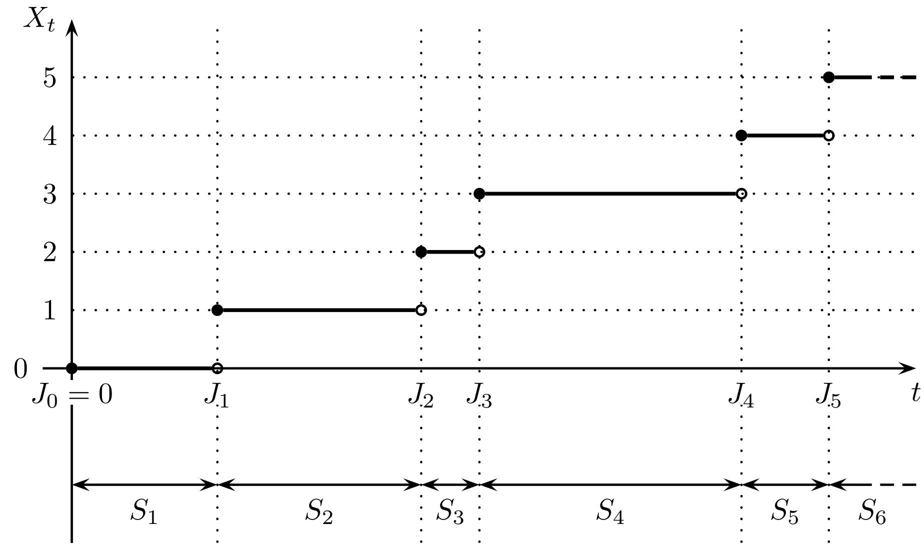


Figure 2.2: Construct a Poisson process

Markov property We now show how the memoryless property of the exponential holding times, leads to a memoryless property of the Poisson process.

Theorem 2.5 (Markov property). *Let $(N_t)_{t\geqslant 0}$ be a Poisson process of rate λ. Then, for any $s > 0, (N_{s+t} - N_s)_{t\geqslant 0}$ is also a Poisson process of rate λ, independent of $\{N_t : t \leqslant s\}$.*

Proof. Let $\widetilde{N}_t = N_{t+s} - N_s$ for all $t \geqslant 0$. Clearly, $(\widetilde{N}_t)_{t\geqslant 0}$ is a integer valued, increasing, right-continuous process starting at 0. Its jump times are given by $\widetilde{J}_0 = 0$ and

$$\widetilde{J}_n = J_{N(s)+1} - s \quad \text{ for } n \geqslant 1\,.$$

Hence the holding times are

$$\widetilde{S}_1 = J_{N(s)+1} - s \quad \text{ and } \quad \widetilde{S}_n = S_{N(s)+n} \text{ for } n \geqslant 2\,.$$

It suffices to show that $\{\widetilde{S}_n\}$ is sequence of i.i.d r.v.'s with exponential distribution of parameter λ, and independent of $(N_t)_{0\leqslant t\leqslant s}$.

For any real numbers $t_1,\cdots,t_n\in(0,\infty)$, $0\leqslant s_1<\cdots<s_m\leqslant s$, and non-negative integers $0\leqslant k_1\leqslant\cdots\leqslant k_{m-1}<k_m=k$, [1]

$$\begin{aligned}&\mathbb{P}\left(\widetilde{S}_1>t_1,\cdots,\widetilde{S}_n>t_n;\, N(s)=k, N(s_1)=k_1,\cdots,N(s_m)=k_m\right)\\ &=\mathbb{P}\left(J_{k+1}-s>t_1, S_{k+i}>t_i, 2\leqslant i\leqslant n;\, J_k\leqslant s_m, J_{k_j}\leqslant s_j<J_{k_j+1}, 1\leqslant j<m\right)\,,\end{aligned}$$

we denote by A the event $\{J_{k_j}\leqslant s_j<J_{k_j+1}, 1\leqslant j<m\}$, then

$$\begin{aligned}&\mathbb{P}\left(\widetilde{S}_1>t_1,\cdots,\widetilde{S}_n>t_n;\, N(s)=k, N(s_1)=k_1,\cdots,N(s_m)=k_m\right)\\ &=\mathrm{e}^{-\lambda(t_2+\cdots+t_n)}\times\mathbb{P}\big(\{S_{k+1}>t_1+s-J_k\}\cap\{J_k\leqslant s_m\}\cap A\big)\,,\end{aligned}$$

and let $A'=\{x_1+\cdots x_{k_j}\leqslant s_j<x_1+\cdots x_{k_j+1}\,, 1\leqslant j<m\}$,

$$\begin{aligned}&\mathbb{P}\left(\{S_{k+1}>t_1+s-J_k\}\cap\{J_k\leqslant s_m\}\cap A\right)\\ &=\int_{x_1,\cdots,x_m>0}\mathbb{P}\big(S_{k+1}>t_1+s-\textstyle\sum x_j\big)1_{\{\sum x_j\leqslant s_m\}}1_{A'}\,\mathrm{d}x_1\cdots\mathrm{d}x_m\end{aligned}$$

by memoryless property of the exponential distributions,

$$\begin{aligned}&=\int_{x_1,\cdots,x_m>0}\mathbb{P}\big(S_{k+1}>t_1\big)\mathbb{P}\big(S_{k+1}>s-\textstyle\sum x_j\big)1_{\{\sum x_j\leqslant s_m\}}1_{A'}\,\mathrm{d}x_1\cdots\mathrm{d}x_m\\ &=e^{-\lambda t_1}\int_{x_1,\cdots,x_m>0}\mathbb{P}\big(S_{k+1}>s-\textstyle\sum x_j\big)1_{\{\sum x_j\leqslant s_m\}}1_{A'}\,\mathrm{d}x_1\cdots\mathrm{d}x_m\\ &=e^{-\lambda t_1}\,\mathbb{P}\left(\{S_{k+1}>s-J_k\}\cap\{J_k\leqslant s_m\}\cap A\right)\\ &=e^{-\lambda t_1}\,\mathbb{P}\left(\{J_k\leqslant s_m<s<J_{k+1}\}\cap A\right)\,.\end{aligned}$$

Thus we get

$$\begin{aligned}&\mathbb{P}\left(\widetilde{S}_1>t_1,\cdots,\widetilde{S}_n>t_n;\, N(s)=k, N(s_1)=k_1,\cdots,N(s_m)=k_m\right)\\ &=\mathrm{e}^{-\lambda(t_1+t_2+\cdots+t_n)}\times\mathbb{P}\big(\{J_k\leqslant s_m\leqslant s<J_{k+1}\}\cap A\big)\,,\\ &=\mathrm{e}^{-\lambda(t_1+t_2+\cdots+t_n)}\times\mathbb{P}\big(N(s)=k, N(s_1)=k_1,\cdots,N(s_m)=k_m\big)\,.\end{aligned}$$

[1] We let $k_{m-1}<k_m=k$, in ordet that we can write $\{N(s)=k, N(s_1)=k_1,\cdots,N(s_m)=k_m\}$ as $\{J_{k+1}>s\}\cap\{J_{k_j}\leqslant s_j<J_{k_j+1}, 1\leqslant j<m\}$ and the second event is in $\sigma(S_1,\cdots,S_k)$.

Thus $\{\widetilde{S}_n\}$ is sequence of i.i.d r.v.'s with exponential distribution of parameter λ, and independent of $\{N_t : t \leqslant s\}$. $\square$

In fact, we shall see in Section ?? by an argument in essentially the same spirit that the result also holds with s replaced by any stopping time T of $(N_t)_{t\geqslant 0}$.

Theorem (Strong Markov property). *Let $(N_t)_{t\geqslant 0}$ be a Poisson process of rate λ and let T be a stopping time of $(N_t)_{t\geqslant 0}$. Then, conditional on $T < \infty$,*

$$(N_{T+t} - N_T)_{t\geqslant 0}$$

is also a Poisson process of rate λ, independent of $(N_s)_{s\leqslant T}$.

Alternative definitions We come to the key result for the Poisson process, which gives two conditions equivalent to the jump chain/holding time characterization which we took as our original definition. Thus we have three alternative definitions of the same process.

Theorem 2.6. *Let $(N_t)_{t\geqslant 0}$ be an increasing, right-continuous, integer-valued process starting from 0. Let $\lambda \in (0, \infty)$, then the following three conditions are equivalent:*

(i) (jump chain/holding time definition) the holding times $S_1, S_2, \ldots$ of $(N_t)_{t\geqslant 0}$ are independent exponential random Variables of parameter λ and the jump chain is given by $Y_n = n$ for all n;

(ii) (transition probability definition) $(N_t)_{t\geqslant 0}$ has stationary independent increments and, for each $t > 0$, N_t has Poisson(λt) distribution.

(iii) (infinitesimal definition) $(N_t)_{t\geqslant 0}$ has independent increments and, as $h \downarrow 0$, uniformly in t, [2]

$$\mathbb{P}(N_{t+h} - N_t = 0) = 1 - \lambda h + o(h),\ \mathbb{P}(N_{t+h} - N_t = 1) = \lambda h + o(h)\,.$$

[2] Uniformly in t implies " stationary increments ".

If $(N_t)_{t\geqslant 0}$ satisfies any of these conditions then it is called a Poisson process of rate λ.

Proof. (i) $\Leftrightarrow$ (ii). If (i) holds, then, by the Markov property, Theorem 2.5, $(N_t)_{t\geqslant 0}$ has stationary independent increments. By Theorem 2.4, we have proved (ii). If (ii) holds, the finite-dimensional distributions of $(N_t)_{t\geqslant 0}$ is uniquely determined, and hence the distribution of jump chain and holding times.

(ii) $\Leftrightarrow$ (iii). If (ii) holds, then for any t, $h \geqslant 0$,

$$\begin{aligned}\mathbb{P}\left(X_{t+h}-X_t=0\right)&=\mathbb{P}\left(X_h=0\right)=e^{-\lambda h}=1-\lambda h+o(h)\\ \mathbb{P}\left(X_{t+h}-X_t=1\right)&=\mathbb{P}\left(X_h=1\right)=\lambda he^{-\lambda h}=\lambda h+o(h)\,.\end{aligned}$$

which implies (iii). To show the other hand, if (iii) holds, then, for $i \geqslant 2$, we have $\mathbb{P}\left(X_{t+h}-X_t=i\right)=o(h)$ as $h\downarrow 0$, uniformly in t. Set $p_{0j}(t)=\mathbb{P}\left(X_t=j\right)$. Then, for $j=1,2,\ldots$,

$$\begin{aligned}p_{0j}(t+h)=\mathbb{P}\left(X_{t+h}=j\right)&=\sum_{i=0}^{j}\mathbb{P}\left(X_{t+h}-X_t=i\right)\mathbb{P}\left(X_t=j-i\right)\\ &=\big(1-\lambda h+o(h)\big)p_{0j}(t)+\big(\lambda h+o(h)\big)p_{0j-1}(t)+o(h)\,,\end{aligned}$$

so

$$\frac{p_{0j}(t+h)-p_{0j}(t)}{h}=-\lambda p_{0j}(t)+\lambda p_{0j-1}(t)+o(1)\,.$$

since this estimate is uniform in t we can put $t=s-h$ to obtain for all $s\geqslant h$,

$$\frac{p_{0j}(s)-p_{0j}(s-h)}{h}=-\lambda p_{0j}(s-h)+\lambda p_{0j-1}(s-h)+o(1)$$

Now let $h\downarrow 0$ to see that $p_{0j}(t)$ is first continuous and then differentiable and satisfies the differential equation

$$p_{0j}'(t)=-\lambda p_{0j}(t)+\lambda p_{0j-1}(t)\,.$$

By a simpler argument we also find

$$p_{00}'(t)=-\lambda p_{00}(t)\,.$$

Since $X_0 = 0$ we have initial conditions

$$p_{00}(0) = 1,\ p_{0j}(0) = 0 \quad \text{for } j \geqslant 1\,.$$

This system of equations has a unique solution given by

$$p_{0j}(t) = e^{-\lambda t}\frac{(\lambda t)^j}{j!} \quad \text{for } j \geqslant 0\,.$$

Hence N_t has a Poisson(λt) distribution. If $(X_t)_{t\geqslant 0}$ satisfies (iii), then $(X_{s+t} - X_s)_{t\geqslant 0}$ satisfies (iii), so the above argument shows $N_{t+s} - N_s$ has has a Poisson(λt) distribution for any s, which implies (ii). □

The differential equations which appeared in the proof are really the *forward equations* for the Poisson process. To make this clear, consider the possibility of starting the process from i at time 0, writing $\mathbb{P}_i$ as a reminder, and set

$$p_{ij}(t) = \mathbb{P}_i\left(X_t = j\right)$$

Then, by spatial homogeneity $p_{ij}(t) = p_{0j-i}(t)$, and we could rewrite the differential equations as

$$\begin{aligned} p'_{i0}(t) &= -\lambda p_{i0}(t)\ , & p_{i0}(0) &= \delta_{i0} \\ p'_{ij}(t) &= \lambda p_{i,j-1}(t) - \lambda p_{ij}(t)\,, & p_{ij}(0) &= \delta_{ij} \end{aligned}$$

or, in matrix form, for Q as above,

$$P'(t) = P(t)Q, \quad P(0) = I\,.$$

Theorem 2.6 contains a great deal of information about the Poisson process of rate λ. It can be useful when trying to decide whether a given process is a Poisson process as it gives you three alternative conditions to check, and it is likely that one will be easier to check than another. On the other hand it can also be useful when answering a question about a given Poisson process as this question may be more closely connected to one definition than another. For example, you might like to consider the difficulties in approaching the next result using the jump chain/holding time definition.

2.2.1 Compound Poisson Processes

Suppose that between 12: 00 and 1: 00 cars arrive at a fast food resturant according to a Poisson process $(N_t)_{t\geqslant 0}$ with rate λ. Let Y_n be the number of people in the n th vehicle which we assume to be i.i.d. and independent o $(N_t)_{t\geqslant 0}$. Having introduced the Y_n's, it is natural to consider the sum of the Y_n's we have seen up to time t :

$$S(t) = Y_1 + \cdots + Y_{N(t)} \tag{2.6}$$

where we set $S(t) = 0$ if $N(t) = 0$. In the motivating example, $S(t)$ gives the number of customers that have arrived up to time t.

Lemma 2.7. *Let $Y_1, Y_2, \cdots$ be i.i.d, N be an nonnegative integer valued r.v. independent with $\{Y_n\}$ and, let $S = Y_1 + \cdots + Y_N$.*

(i) If $\mathbb{E}\left|Y_1\right|, \mathbb{E}N < \infty$, then $\mathbb{E}S = \mathbb{E}N \cdot \mathbb{E}Y_1$.

(ii) If $\mathbb{E}Y_i^2, \mathbb{E}N^2 < \infty$, then $\operatorname{Var}(S) = \mathbb{E}N \operatorname{Var}(Y_i) + \operatorname{Var}(N)\left(\mathbb{E}Y_i\right)^2$.

Remark. Why is this reasonable? The first of these is natural since if $N = n$ is nonrandom $\mathbb{E}S = \mathbb{E}N \cdot \mathbb{E}Y_1$, (i) then results by setting $n = \mathbb{E}N$. This fact is known as Wald's equation. The formula in (ii) is more complicated but it clearly has two of the necessary properties:

If $N = n$ is nonrandom, $\operatorname{Var}(S) = n \operatorname{Var}(Y_i)$

If $Y_i = c$ is nonrandom $\operatorname{Var}(S) = c^2 \operatorname{Var}(N)$

Combining these two observations, we see that $\mathbb{E}N \operatorname{Var}(Y_i)$ is the contribution to the variance from the variability of the Y_i, while $\operatorname{Var}(N)\left(\mathbb{E}Y_i\right)^2$ is the contribution from the variability of N.

Proof. Breaking things down according to the value of N,

$$\begin{aligned}
\mathbb{E}S &= \sum_{n=0}^{\infty} \mathbb{E}[S \mid N = n]\, \mathbb{P}(N = n) \\
&= \sum_{n=0}^{\infty} n\, \mathbb{E}Y_i\, \mathbb{P}(N = n) = \mathbb{E}N \cdot \mathbb{E}Y_i
\end{aligned}$$

For the second formula we note that when $N = n, S = X_1 + \cdots + X_n$ has $\operatorname{Var}(S) = n \operatorname{Var}(Y_i)$ and hence,

$$\mathbb{E}\left(S^2 | N = n\right) = n \operatorname{Var}(Y_i) + (n\mathbb{E}Y_i)^2$$

Computing as before we get

$$\begin{aligned}
\mathbb{E}S^2 &= \sum_{n=0}^{\infty} \mathbb{E}\left(S^2 | N = n\right) \cdot P(N = n) \\
&= \sum_{n=0}^{\infty} \left\{ n \cdot \operatorname{Var}(Y_i) + n^2 (\mathbb{E}Y_i)^2 \right\} \cdot P(N = n) \\
&= (\mathbb{E}N) \cdot \operatorname{Var}(Y_i) + \mathbb{E}N^2 \cdot (\mathbb{E}Y_i)^2
\end{aligned}$$

To compute the Variance now, we observe that

$$\begin{aligned}
\operatorname{Var}(S) &= \mathbb{E}S^2 - (\mathbb{E}S)^2 \\
&= (\mathbb{E}N) \cdot \operatorname{Var}(Y_i) + \mathbb{E}N^2 \cdot (\mathbb{E}Y_i)^2 - (\mathbb{E}N \cdot \mathbb{E}Y_i)^2 \\
&= (\mathbb{E}N) \cdot \operatorname{Var}(Y_i) + \operatorname{Var}(N) \cdot (\mathbb{E}Y_i)^2
\end{aligned}$$

where in the last step we have used $\operatorname{Var}(N) = \mathbb{E}N^2 - (\mathbb{E}N)^2$ to combine the second and third terms. □

Corollary 2.8. *For any t, $\mathbb{E}S(t) = \lambda t\, \mathbb{E}Y_1$, and $\operatorname{Var} S(t) = \lambda t\, \mathbb{E}Y_1^2$.*

Proof. Note that in the special case of the Poisson, we have $EN(t) = \lambda t$ and $\operatorname{Var} N(t) = \lambda t$, so the result follows. □

2.2.2 Thinning

We will use the discrete i.i.d r.v.'s $\{Y_n\}$, independent with $(N_t)_{t\geqslant 0}$ to split the Poisson process into several, where Y_n take values in a conutable set I. For given $j \in I$, let $N_j(t)$ be the number of $n \leqslant N(t)$ with $Y_i = j$, that is

$$N_j(t) = \sum_{n=1}^{\infty} 1_{\{Y_n = j, n \leqslant N(t)\}} = \sum_{n=1}^{\infty} 1_{\{J_n \leqslant t, Y_n = j\}} \tag{2.7}$$

The somewhat remarkable fact is

Theorem 2.9. *$(N_j(t))_{t\geqslant 0}$ is Poisson process with rate λp_j, where $p_j = \mathbb{P}(Y_1 = j)$. Moreover, $\{N_j(t)\}_{j\in I}$ is independent.*

Proof. To begin, we suppose that $P\,(Y_i = 1) = p$, $P\,(Y_i = 2) = q$, and $p+q = 1$, so there are only two Poisson processes to consider, namely, $N_1(t)$ and $N_2(t)$. For $0 = t_0 < t_1 < \cdots < t_n < \infty$ and $k_j,\, m_j \in \mathbb{N}$ for $1 \leqslant j \leqslant n$,

$$\begin{aligned}&\mathbb{P}\big(N_1(t_j) - N_1(t_{j-1}) = k_j\,, N_2(t_j) - N_2(t_{j-1}) = m_j\,; 1 \leqslant j \leqslant n\big)\\ =\;& \mathbb{P}\big(N(t_j) - N(t_{j-1}) = k_j + m_j; \xi_j = k_j\,; 1 \leqslant j \leqslant n\big)\end{aligned}$$

where

$$\xi_j \coloneqq \sum_{s_{j-1}<n\leqslant s_j} 1_{\{Y_n=1\}}\,, \text{ and } s_j \coloneqq \sum_{i=1}^{j} k_i + m_i\,.$$

Clearly, $\{\xi_j\}$ is independent r.v.'s with Binomial$(k_j + m_j, p)$ distribution, respectively, and independent of $\{N(t)\}$. Hence

$$\begin{aligned}&= \mathbb{P}\big(N(t_j) - N(t_{j-1}) = k_j + m_j; 1 \leqslant j \leqslant n\big) \times \mathbb{P}\big(\xi_j = k_j\,; 1 \leqslant j \leqslant n\big)\\ &= \prod_{j=1}^{n} \mathbb{P}\big(N(t_j) - N(t_{j-1}) = k_j + m_j\big) \times \mathbb{P}\big(\xi_j = k_j\big)\\ &= \prod_{j=1}^{n} \mathrm{e}^{-\lambda(t_j-t_{j-1})}\frac{[\lambda(t_j - t_{j-1})]^{k_j+m_j}}{(k_j + m_j)!} \times \frac{(k_j + m_j)!}{k_j!m_j!}p^{k_j}q^{m_j}\,.\\ &= \prod_{j=1}^{n} \mathrm{e}^{-\lambda p(t_j-t_{j-1})}\frac{[\lambda p(t_j - t_{j-1})]^{k_j}}{k_j!} \times \prod_{j=1}^{n} \mathrm{e}^{-\lambda q(t_j-t_{j-1})}\frac{[\lambda q(t_j - t_{j-1})]^{m_j}}{m_j!}\end{aligned}$$

Thus $N_1(t)$ is Poisson Process with rate λp, $N_2(t)$ is Poisson Process with rate λq, and $N_1(t)$, $N_2(t)$ are independent.

Using the same method, one can show that for $Y_1 = 1, \cdots, m$ with probability $p_1, \cdots, p_m$ such that $p_i + \cdots p_m = 1$, the theorem holds. Then the general case follows. $\square$

Remark. We say that $\{N(t)\}_{t\geqslant 0}$ is a *(nonhomogeneous) Poisson process* with rate $\lambda(\cdot)$, if $N(0) = 0$, and

- $N(t)$ has independent increments and,
- $N(t)-N(s)$ is Poisson with mean $\int_s^t \lambda(r)dr$.

The thinning results generalizes easily to the nonhomogeneous case : Suppose that in a Poisson process with rate λ, we keep a point that lands at time t with probability $p(t)$. Then the result is a nonhomogeneous Poisson process with rate $\lambda p(\cdot)$.

¶ Example 2.1. Ellen catches fish at times of a Poisson process with rate 2 per hour. 40% of the fish are salmon(鲑鱼), while 60% of the fish are trout(鳟鱼). What is the probability she will catch exactly 1 salmon and 2 trout if she fishes for 2.5 hours?

The total number of fish she catches in 2.5 hours is Poisson with mean 5, so the number of salmon and the number of trout are independent Poissons with means 2 and 3. Thus the probability of interest is

$$e^{-2}\frac{2^1}{1!}\cdot e^{-3}\frac{3^2}{2!}\,.$$

Further theory A *Poisson point process on a measure space* $(S,\mathcal{S},\mu)$ a random mapping $m:\mathcal{S}\to\{0,1,\ldots\}$ that for each ω is a measure on $\mathcal{S}$ and has the following property : if $A_1,\ldots,A_n$ are disjoint sets with $\mu(A_i)<\infty$, then

$$m(A_1),\cdots,m(A_n)$$

are independent and have Poisson distributions with means $\mu(A_i)$. μ is called the *intensity measure* of the process.

If $\mu(S)<\infty$ then it follows from Theorem 2.9 that we can construct m by the following recipe: let $X_1,X_2,\ldots$ be i.i.d. elements of S with distribution $\nu(\cdot)=\mu(\cdot)/\mu(S)$, let N be an independent Poisson random variable with mean $\mu(S)$, and let

$$m(A)=\#\{j\leqslant N: X_j\in A\}\ ,\ \text{for all } A\in\mathcal{S}\,.$$

To extend the construction to infinite measure spaces, e.g. , $S = \mathbf{R}^d, S =$ Borel sets, $\mu =$ Lebesgue measure, divide the space up into disjoint sets of finite measure and put independent Poisson processes on each set.

2.2.3 Superposition

Taking one Poisson process and splitting it into two or more by using an i.i.d. sequence $\{Y_n\}$ is called thinning. Going in the other direction and adding up a lot of independent processes is called *superposition*. Since a Poisson process can be split into independent Poisson processes, it should not be too surprising that when the independent Poisson processes are put together, the sum is Poisson with a rate equal to the sum of the rates.

Theorem 2.10. *Suppose for $j \in I$, $(N_j(t))_{t\geqslant 0}$ are independent Poisson processes with rates $\lambda_j > 0$, respectively, and $\lambda := \sum_{j\in I} \lambda_j < \infty$. Then*

$$N(t) := \sum_{j\in I} N_j(t)$$

is a Poisson process with rate λ.

Proof. Firstly, for $0 = t_0 < t_1 < \cdots t_n < \infty$, we have

$$N(t_k) - N(t_{k-1}) = \sum_{j\in I} N_j(t_k) - N_j(t_{k-1})\,, \text{ for } j \in I\,.$$

Clearly, $\{N(t_k) - N(t_{k-1}) : k = 1, \cdots, n\}$ are independent and by Proposition**??**, they are Poisson r.v.'s, with parameter $\lambda(t_k - t_{k-1})$, respectively. It's easy to see that for any ω and t, there are finite many $j \in I$, $N_j(t) > 0$, thus $N(t)$ is right-continuous. $\square$

2.2.4 Conditioning

Next we establish some relations between Poisson processes and the uniform distribution. Notice that the conclusions are independent of the rate of the process considered. The results say in effect that the jumps of a Poisson process are as randomly distributed as possible.

Theorem 2.11. *Let $(N_t)_{t\geqslant 0}$ be a Poisson process. Then, conditional on the event $\{X_t = n\}$, the jump times $J_1, \ldots, J_n$ have joint density function*

$$f(t_1, \ldots, t_n) = \frac{n!}{t^n} 1_{\{0<t_1<\cdots<t_n<t\}} \tag{2.8}$$

Thus, conditional on $\{N_t = n\}$, the jump times $J_1, \ldots, J_n$ have the same distribution as an ordered sample of size n from the uniform$(0, t)$.

Proof. To compute the joint density, we need to find some f such that for all $A \in \mathcal{B}^n$,

$$\mathbb{P}((J_1, \ldots, J_n) \in A \mid N_t = n) = \int_A f(t_1, \ldots, t_n)\, \mathrm{d}t_1 \ldots \mathrm{d}t_n .$$

First, we compute the joint density for $(J_1, \cdots, J_n)$. Since the holding times $S_1, \ldots, S_{n+1}$ have joint density function

$$\lambda^{n+1} e^{-\lambda(s_1+\ldots+s_{n+1})} 1_{\{s_1,\ldots,s_{n+1}>0\}} ,$$

so the jump times $J_1, \ldots, J_{n+1}$ have joint density function

$$\lambda^{n+1} e^{-\lambda t_{n+1}} 1_{\{0<t_1<\ldots<t_{n+1}\}} ,$$

hence

$$\begin{aligned}
&\mathbb{P}((J_1, \ldots, J_n) \in A \mid X_t = n) \\
&= \frac{\mathbb{P}((J_1, \ldots, J_n) \in A \text{ and } J_n \leqslant t < J_{n+1})}{\mathbb{P}(X_t = n)} \\
&= \frac{\lambda^n \exp\{-\lambda t\}}{(\lambda t)^n / n! \exp\{-\lambda t\}} \int_A 1_{\{0<t_1<\ldots<t_n<t\}} \mathrm{d}t_1 \ldots \mathrm{d}t_n \\
&= \int_A \frac{n!}{t^n} 1_{\{0<t_1<\cdots<t_n<t\}}\, \mathrm{d}t_1 \ldots \mathrm{d}t_n
\end{aligned}$$

as required. □

Corollary 2.12. *If $s < t$ and $0 \leqslant m \leqslant n$, then*

$$\mathbb{P}(N(s) = m \mid N(t) = n) = \binom{n}{m} \left(\frac{s}{t}\right)^m \left(1 - \frac{s}{t}\right)^{n-m} \tag{2.9}$$

Or, conditional on $N(t) = n$, the distribution of $N(s)$ is binomial$(n, s/t)$.

2.3 Birth processes

A birth process is a generalization of a Poisson process in which the parameter λ is allowed to depend on the current state of the process. The data for a birth process consist of *birth rates* $q_j \geqslant 0$ where $j \in \mathbb{N}$.

We begin with a definition in terms of jump chain and holding times. A minimal right-continuous process $(X_t)_{t\geqslant 0}$ with values in $\{0, 1, 2, \cdots\} \cup \{\infty\}$ is a **birth process of rates** $(q_j)_{j\geqslant 0}$ if, conditional on $X_0 = i$, its holding times $S_1, S_2, \cdots$ are independent exponential random variables of parameters $q_i, q_{i+1}, \cdots$, respectively, and its jump chain is given by $Y_n = i + n$.

The flow diagram and the Q-matrix is given by

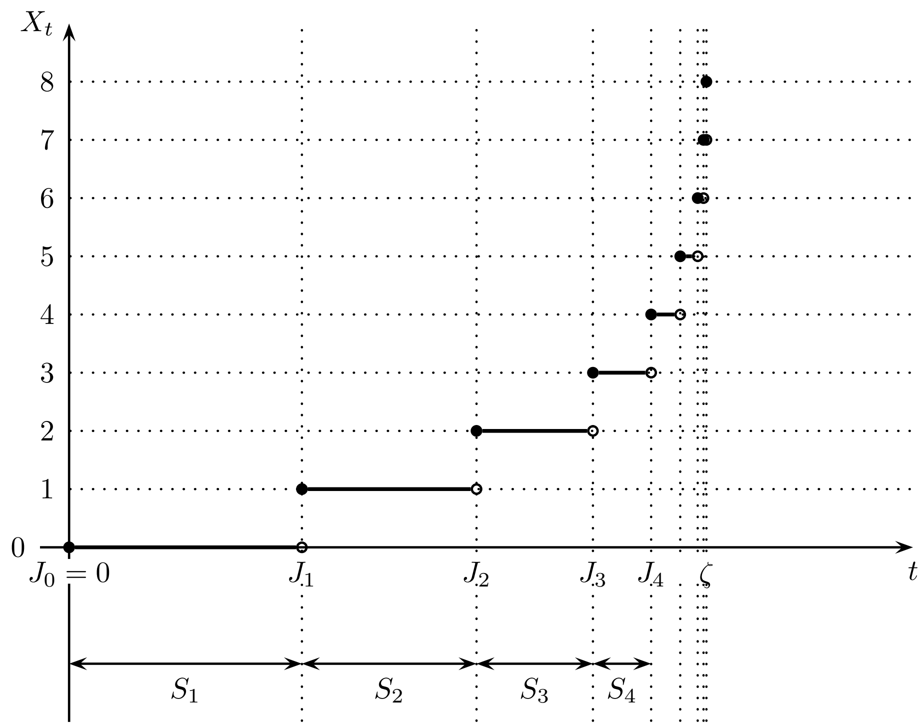
$$0 \xrightarrow{q_0} 1 \xrightarrow{q_1} 2 \xrightarrow{q_2} 3 \xrightarrow{q_3} 4 \text{ --- - - -}$$

and

$$Q = \begin{pmatrix} -q_0 & q_0 & & & \\ & -q_1 & q_1 & & \\ & & -q_2 & q_2 & \\ & & & \ddots & \ddots \\ & & & & \end{pmatrix}$$

Much of the theory associated with the Poisson process goes through for birth processes with little change, except that some calculations can no longer be made so explicitly.

Explosion of brith process The most interesting new phenomenon present in birth processes is the possibility of explosion. For certain choices of birth rates, a typical path will make infinitely many jumps in a finite time, as shown in the following diagram. The convention of setting the process to equal ∞ after explosion is particularly appropriate for birth processes!



In fact, Theorem 2.13 tells us exactly when explosion will occur.

Theorem 2.13. *Let* $(X_t)_{t\geqslant 0}$ *be a birth process of rates* $(q_j : j \geqslant 0)$ *starting from* 0.

(i) *If* $\sum_{j=0}^{\infty} \frac{1}{q_j} < \infty$, *then* $\mathbb{P}(\zeta < \infty) = 1$.

(ii) *If* $\sum_{j=0}^{\infty} \frac{1}{q_j} = \infty$, *then* $\mathbb{P}(\zeta = \infty) = 1$.

Proof. Using Proposition**??**. □

The proof of the Markov property for the Poisson process is easily adapted to give the following generalization.

Theorem 2.14 (Markov property). *Let $(X_t)_{t\geqslant 0}$ be a birth process of rates $(q_j : j \geqslant 0)$. Then, conditional on $X_s = i, (X_{s+t})_{t\geqslant 0}$ is a birth process of rates $(q_j : j \geqslant 0)$ starting from i and independent of $(X_s : s \leqslant s)$*

Proof. Let $\widetilde{X}_t = X_{t+s}$ for each $t \geqslant 0$. Then $(\widetilde{X}_t)_{t\geqslant 0}$ is a right-continuous process valued in $\{0, 1, 2, \cdots\} \cup \{\infty\}$. Denote $(\widetilde{J}_n)_{n\geqslant 0}$ the jump times of $(\widetilde{X}_t)_{t\geqslant 0}$ and

$$N(t) := \sum_{n=1}^{\infty} 1_{\{J_n \leqslant t\}}, \quad \text{for all } t \geqslant 0.$$

Then $\widetilde{J}_0 = 0$ and

$$\widetilde{J}_n = J_{N(s)+n} - s, \text{ for } n \geqslant 2.$$

The holding times of $(\widetilde{X}_t)_{t\geqslant 0}$ is given by

$$\widetilde{S}_1 = J_{N(s)+1} - s\ ;\ \widetilde{S}_n = S_{N(s)+n}, \text{ for } n \geqslant 2.$$

Hence, on $\{N(s) = k\} = \{J_k \leqslant s < J_{k+1}\}$,

$$\widetilde{S}_1 = J_{k+1} - s\ ;\ \widetilde{S}_n = S_{k+n}, \text{ for } n \geqslant 2,$$

and

$$X_t = i - k + \sum_{n=1}^{k} 1_{\{J_n \leqslant t\}}, \text{ for } t \leqslant s.$$

We show now that $(\widetilde{X}_t)_{t\geqslant 0}$ is a birth process of rates $(q_j : j \geqslant 0)$ and

$$\sigma(X_t : t \leqslant s) \text{ is independent of } \sigma(\widetilde{X}_t : t \geqslant 0) = \sigma(\widetilde{S}_n : n \geqslant 1),$$

First of all, it's easy to show that for each $A \in \sigma(X_t : t \leqslant s)$,

$$\begin{aligned}&\mathbb{P}(\{\widetilde{S}_1 > t_1, \cdots, \widetilde{S}_n > t_n\} \cap A \cap \{N(s) = k, X_s = i\})\\ =&\mathrm{e}^{-q_i t_1} \cdots \mathrm{e}^{-q_{i+n-1} t_n} \mathbb{P}(\cap A \cap \{N(s) = k, X_s = i\})\end{aligned}$$

Then sum over $k = 0, 1, 2, \cdots$, and divide $\mathbb{P}(X_s = i)$, we get

$$\begin{aligned}&\mathbb{P}(\{\widetilde{S}_1 > t_1, \cdots, \widetilde{S}_n > t_n\} \cap A \mid \{X_s = i\})\\ &= \mathrm{e}^{-q_i t} \cdots \mathrm{e}^{-q_{i+n-1} t}\ \mathbb{P}(A \mid \{X_s = i\})\\ &= \mathbb{P}_i(S_1 > t_1, \cdots, S_n > t_n)\ \mathbb{P}(A \mid X_s = i).\end{aligned}$$

We have completed the proof. $\square$

We shall shortly prove a theorem on birth processes which generalizes the key theorem on Poisson processes. First we must see what will replace the Poisson probabilities. In Theorem 2.6 these arose as the unique solution of a system of differential equations, which we showed were essentially the forward equations. Now we can still write down the forward equation

$$P'(t) = P(t)Q, \quad P(0) = I$$

or, in components

$$p'_{i0}(t) = -p_{i0}(t)q_0, \quad p_{i0}(0) = \delta_{i0}$$

and, for $j = 1, 2, \cdots$

$$p'_{ij}(t) = p_{i,j-1}(t)q_{j-1} - p_{ij}(t)q_j, \quad p_{ij}(0) = \delta_{ij}$$

Moreover, these equations still have a unique solution; it is just not as explicit as before. For we must have

$$p_{i0}(t) = \delta_{i0}e^{-q_0 t}$$

which can be substituted in the equation

$$p'_{i1}(t) = p_{i0}(t)q_0 - p_{i1}(t)q_1, \quad p_{i1}(0) = \delta_{i1}$$

and this equation solved to give

$$p_{i1}(t) = \delta_{i1}e^{-q_1 t} + \delta_{i0}\int_0^t q_0 e^{-q_0 s}e^{-q_1(t-s)}ds$$

Now we can substitute for $p_{i1}(t)$ in the next equation up the hierarchy and find an explicit expression for $p_{i2}(t)$, and so on.

Theorem 2.15. *Let $(X_t)_{t\geqslant 0}$ be an increasing, right-continuous process with values in $\{0, 1, 2, \cdots\} \cup \{\infty\}$. Let $0 \leqslant q_j < \infty$ for all $j \geqslant 0$. Then the following three conditions are equivalent:*

(i) (jump chain/holding time definition) conditional on $X_0 = i$, the holding times $S_1, S_2, \cdots$ are independent exponential random variables of parameters $q_i, q_{i+1}, \cdots$ respectively and the jump chain is given by $Y_n = i + n$ for all n.

(ii) (infinitesimal definition) for all $t, h \geqslant 0$, conditional on $X_t = i$, X_{t+h} is independent of $\{X_s : s \leqslant t\}$ and, as $h \downarrow 0$, uniformly in t

$$\begin{aligned}\mathbb{P}\left(X_{t+h} = i \mid X_t = i\right) &= 1 - q_i h + o(h)\,, \\ \mathbb{P}\left(X_{t+h} = i+1 \mid X_t = i\right) &= q_i h + o(h)\,.\end{aligned}$$

(iii) (transition probability definition) for all $n = 1, 2, \cdots$, all times $0 \leq t_0 \leqslant \cdots \leqslant t_{n+1}$ and all states $i_0, \cdots, i_{n-1}, i, j$

$$\mathbb{P}\left(X_{t_{n+1}} = j \mid X_{t_n} = i, X_{t_0} = i_0, \cdots, X_{t_{n-1}} = i_{n-1}\right) = p_{ij}\left(t_{n+1} - t_n\right)$$

where $\left(p_{ij}(t)\right)_{i,j\in\mathbb{N}}$ is the unique solution of the forward equations.

If $(X_t)_{t\geqslant 0}$ satisfies any of these conditions, then it is called a birth process of rates $(q_j : j \geqslant 0)$.

Proof. (i) $\Rightarrow$ (ii). If (i) holds, then, by the Markov property for any $t, h \geqslant 0$ conditional on $X_t = i, X_{t+h}$ is independent of $(X_s : s \leqslant t)$ and, as $h \downarrow 0$ uniformly in t

$$\begin{aligned}\mathbb{P}\left(X_{t+h} \geqslant i+1 \mid X_t = i\right) &= \mathbb{P}\left(X_h \geqslant i+1 \mid X_0 = i\right) \\ &= \mathbb{P}\left(J_1 \leqslant h \mid X_0 = i\right) = 1 - e^{-q_i h} = q_i h + o(h)\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}\left(X_{t+h} \geqslant i+2 \mid X_t = i\right) &= \mathbb{P}\left(X_h \geqslant i+2 \mid X_0 = i\right) \\ &= \mathbb{P}\left(J_2 \leqslant h \mid X_0 = i\right) \leqslant \mathbb{P}\left(S_1 \leqslant h \text{ and } S_2 \leqslant h \mid X_0 = i\right) \\ &= \left(1 - e^{-q_i h}\right)\left(1 - e^{-q_{i+1} h}\right) = o(h)\end{aligned}$$

which implies (ii).

(ii) $\Rightarrow$ (iii). If (ii) holds, then certainly for $k = i+2, i+3, \cdots$

$$\mathbb{P}\left(X_{t+h} = k \mid X_t = i\right) = o(h) \quad \text{as } h \downarrow 0, \text{ uniformly in } t$$

Set $p_{ij}(t) = \mathbb{P}\left(X_t = j \mid X_0 = i\right)$. Then, for $j = 1, 2, \cdots$

$$\begin{aligned}
p_{ij}(t+h) &= \mathbb{P}\left(X_{t+h} = j \mid X_0 = i\right) \\
&= \sum_{k=i}^{j} \mathbb{P}\left(X_t = k \mid X_0 = i\right) \mathbb{P}\left(X_{t+h} = j \mid X_t = k\right) \\
&= p_{ij}(t)\left(1 - q_j h + o(h)\right) + p_{i,j-1}(t)\left(q_{j-1}h + o(h)\right) + o(h) \\
\frac{p_{ij}(t+h) - p_{ij}(t)}{h} &= p_{i,j-1}(t) q_{j-1} - p_{ij}(t) q_j + O(h)
\end{aligned}$$

As in the proof of Theorem 2.6 we can deduce that $p_{ij}(t)$ is differentiable and satisfies the differential equation

$$p_{ij}'(t) = p_{i,j-1}(t) q_{j-1} - p_{ij}(t) q_j$$

By a simpler argument we also find

$$p_{i0}'(t) = -p_{i0}(t) q_0$$

Thus $(p_{ij}(t) : i, j = 0, 1, 2, \cdots)$ must be the unique solution to the forward equations. If $(X_t)_{t \geqslant 0}$ satisfies (ii) then certainly

$$\mathbb{P}\left(X_{t_{n+1}} = i_{n+1} \mid X_0 = i_0, \cdots, X_{t_n} = i_n\right) = \mathbb{P}\left(X_{t_{n+1}} = i_{n+1} \mid X_{t_n} = i_n\right)$$

(iii) $\Rightarrow$ (i). See the proof of Theorem 2.6. $\square$

¶ Example 2.2 (Yule process). Consider a population in which each individual gives birth after an exponential time of parameter λ, all independently. If i individuals are present then the first birth will occur after an exponential time of parameter $i\lambda$ Then we have $i+1$ individuals and, by the memoryless property, the process begins afresh. Thus the size of the population performs a birth process with rates $q_i = i\lambda$. Let X_t denote the number of individuals at time t and suppose $X_0 = 1$. Write $J = J_1$ for the time of the first birth.

(i) Note that

$$\begin{aligned}\mathbb{E}\, X_t &= \mathbb{E}\, X_t 1_{\{J \leqslant t\}} + \mathbb{E}\, X_t 1_{\{J > t\}} \\ &= \int_0^t \lambda e^{-\lambda s} \mathbb{E}\left[X_t \mid J = s\right] \mathrm{d}s + e^{-\lambda t}\end{aligned}$$

Put $\mu(t) = \mathbb{E} X_t$, then $\mathbb{E}\left[X_t \mid J = s\right] = 2\mu(t-s)$, so

$$\mu(t) = \int_0^t 2\lambda e^{-\lambda s} \mu(t-s) ds + e^{-\lambda t}$$

and setting $r = t - s$

$$e^{\lambda t}\mu(t) = 2\lambda \int_0^t e^{\lambda r}\mu(r) dr + 1$$

By differentiating we obtain $\mu'(t) = \lambda\mu(t)$, so the mean population size grows exponentially:

$$\mathbb{E} X_t = e^{\lambda t}\,.$$

(ii) Let $\phi(t,z) = \mathbb{E}\, z^{X_t}$ for all $t \geqslant 0$ and $|z| < 1$, then

$$\begin{aligned}\mathbb{E}\, z^{X_t} &= \mathbb{E}\, z^{X_t} 1_{\{J \leqslant t\}} + \mathbb{E}\, z^{X_t} 1_{\{J > t\}} \\ &= \int_0^t \lambda e^{-\lambda s}\, \mathbb{E}\left[\, z^{X_t} \mid J = s\right] \mathrm{d}s + e^{-\lambda t}\end{aligned}$$

Note that $\mathbb{E}\left[\, z^{X_t} \mid J = s\right] = \phi(t-s,z)^2$, hence

$$\phi(t,z) = z e^{-\lambda t} + \int_0^t \lambda e^{-\lambda s} \phi(t-s,z)^2 ds$$

Make a change of variables $u = t - s$ in the integral and deduce that

$$\frac{\partial \phi}{\partial t} = \lambda \phi(\phi - 1)\,,$$

so

$$\phi(t,z) = \frac{z e^{-\lambda t}}{1 - z[1 - e^{-\lambda t}]}\,.$$

We can deduce that, for $n = 1, 2, \cdots$

$$\mathbb{P}\left(X_t = n\right) = e^{-\lambda t}\left[1 - e^{-\lambda t}\right]^{n-1}.$$

2.4 Jump chain and holding times

This section begins the theory of continuous-time Markov chains proper, which will occupy the remainder of this chapter and the whole of the next. The approach we have chosen is to introduce continuous-time chains in terms of the joint distribution of their jump chain and holding times. This provides the most direct mathematical description. It also makes possible a number of constructive realizations of a given Markov chain, which we shall describe, and which underlie many applications.

Let I be a countable set. The basic data for a continuous-time Markov chain on I are given in the form of a **Q-matrix**. Recall that a Q-matrix on I is any matrix $Q=(q_{ij})_{i,j\in I}$ which satisfies the following conditions:

(i) $q_{ij}\geqslant 0$ for all $i\neq j$,

(ii) $-q_{ii}=\sum_{j\neq i} q_{ij}<\infty$ for all i.

We will sometimes find it convenient to write q_i as an alternative notation for $-q_{ii}$.

We are going to describe a simple procedure for obtaining from a Q matrix Q a stochastic matrix Π. The **jump matrix** $\Pi=(\pi_{ij})_{i,j\in I}$ of Q is defined by

$$\pi_{ij}=\begin{cases}\dfrac{q_{ij}}{q_i} & \text{if } j\neq i \text{ and } q_i\neq 0\\ 0 & \text{if } j\neq i \text{ and } q_i=0\end{cases} \qquad \pi_{ii}=\begin{cases}0 & \text{if } q_i\neq 0\\ 1 & \text{if } q_i=0\end{cases} \tag{2.10}$$

This procedure is best thought of *row by row.* For each $i\in I$ we take, where possible, the off-diagonal entries in the i th row of Q and scale them so they add up to 1 , putting a 0 on the diagonal. This is only impossible when the off-diagonal entries are all 0, then we leave them alone and put a 1 on the diagonal.

Here is the definition of a continuous-time Markov chain in terms of its jump chain and holding times.

Definition 2.3. A minimal right-continuous process $(X_t)_{t\geqslant 0}$ on I is called a **continuous-time Markov chain** with *initial distribution* λ and *generator matrix* Q, if its jump chain $(Y_n)_{n\geqslant 0}$ is discrete-time Markov(λ, Π) and for each $n \geqslant 1$, conditional on $Y_0, \cdots, Y_{n-1}$, its holding times $S_1, \cdots, S_n$ are independent exponential random variables of parameters $q(Y_0), \cdots, q(Y_{n-1})$ respectively. We say $(X_t)_{t\geqslant 0}$ is Markov(λ, Q) for short.

Remark. That is, given a set of the form

$$B = \{Y_0 = i_0, \cdots, Y_n = i_n, S_1 > s_1, \cdots, S_n > s_n\}$$

Our jump chain/holding time definition of the continuous-time chain $(X_t)_{t\geqslant 0}$ is saying that for such events

$$\mathbb{P}(B) = \lambda_{i_0}\pi_{i_0 i_1}\cdots\pi_{i_{n-1}i_n}e^{-q_{i_0}s_1}\cdots e^{-q_{i_{n-1}}s_n} \tag{2.11}$$

Then, this definition uniquely determines a probability measure $\mathbb{P}$ on

$$\sigma\left((Y_n)_{n\geqslant 0}, (S_n)_{n\geqslant 1}\right) = \sigma\left((X_t)_{t\geqslant 0}\right)$$

Moreover, one can show that conditional on $\{Y_{n-1} = i\}$, S_n is independnt of $(Y_n)_{n\geqslant 0}$ from (2.11).

Strong Markov property As for Poisson processes and birth processes, we shall deduce the Markov property from the jump chain/holding time definition. In fact, we shall give the strong Markov property as this is a fundamental result and the proof is not much harder. However, the proof of both results really requires the precision of measure theory, so we have omitted it, one can find a proof on *Markov Chians* written by J.R.Norris.

Theorem (Strong Markov property). *Let $(X_t)_{t\geqslant 0}$ be Markov(λ, Q) and let T be a stopping time of $(X_t)_{t\geqslant 0}$. Then, conditional on $T < \zeta$ and $X_T = i$,*

$$(X_{T+t})_{t\geqslant 0}$$

is Markov(δ_i, Q) and independent of $(X_t)_{t\leqslant T}$.

2.4.1 Construct a Markov chain

First construction We can construct such a process as follows: let $(Y_n)_{n\geqslant 0}$ be discrete-time Markov(λ, Π) and let $T_1, T_2, \cdots$ be independent exponential random variables of parameter 1, independent of $(Y_n)_{n\geqslant 0}$. Set

$$S_n = \frac{T_n}{q\left(Y_{n-1}\right)}\,, \;\; J_n = S_1 + \cdots + S_n, \quad \text{for } n \geqslant 1\,. \tag{2.12}$$

and define

$$X_t = \begin{cases} Y_n & \text{if } J_n \leqslant t < J_{n+1} \text{ for some } n \\ \partial & \text{otherwise} \end{cases}, \;\text{for } t \geqslant 0\,. \tag{2.13}$$

Then $(X_t)_{t\geqslant 0}$ has the required properties, that is one can check that, $(X_t)_{t\geqslant 0}$ is minimal right-continuous process, its jump times is exactly $(J_n)_{n\geqslant 0}$ and its jump chian is exactly $(Y_n)_{n\geqslant 1}$, and (2.11) holds.

We shall now describe two further constructions. You will need to understand these constructions in order to identify processes in applications which can be modelled as Markov chains. Both constructions make direct use of the entries in the Q-matrix, rather than proceeding first via the jump matrix.

Second construction We begin with an initial state $X_0 = Y_0$ with distribution λ, and with an array $\left\{T_n^j : n \geqslant 1, j \in I\right\}$ of independent exponential random variables of parameter 1. Then, inductively for $n \in \mathbb{N}$, on $\{Y_n = i\}$, we set

$$S_{n+1}^j = \frac{T_{n+1}^j}{q_{ij}}\,, \quad \text{for } j \neq i\,, \quad S_{n+1} = \inf_{j\neq i} S_{n+1}^j \tag{2.14}$$

and

$$Y_{n+1} = \begin{cases} j & \text{if} \quad S_{n+1}^j = S_{n+1} < \infty \\ i & \text{if} \quad S_{n+1} = \infty \end{cases} \tag{2.15}$$

Then, conditional on $Y_n = i$, the random variables S_{n+1}^j are independent exponentials of parameter q_{ij} for all $j \neq i$. So, conditional on $Y_n = i$, S_{n+1}

is exponential of parameter $q_i = \sum_{j \neq i} q_{ij}$, Y_{n+1} has distribution $(\pi_{ij})_{j \in I}$, and S_{n+1} and Y_{n+1} are independent, and independent of $Y_0, \cdots, Y_n$ and $S_1, \cdots, S_n$, as required. This construction shows why we call q_i *the rate of leaving* i and q_{ij} *the rate of going from* i *to* j.

Our third and final construction of a Markov chain with generator matrix Q and initial distribution λ is based on the Poisson process.

Third construction Imagine the state-space I as a labyrinth(迷宫) of chambers and passages, each passage shut off by a single door which opens briefly from time to time to allow you through in one direction only. Suppose the door giving access to chamber j from chamber i opens at the jump times of a Poisson process of rate q_{ij} and you take every chance to move that you can, then you will perform a Markov chain with Q-matrix Q.

In more mathematical terms, we begin with an initial state $X_0 = Y_0$ with distribution λ, and with a family of independent Poisson processes

$$\left\{ \left(N_t^{ij} \right)_{t \geqslant 0} : i, j \in I, i \neq j \right\}$$

$\left(N_t^{ij} \right)$ having rate q_{ij}. Then set $J_0 = 0$ and define inductively for $n \in \mathbb{N}$,

$$J_{n+1} = \inf \left\{ t > J_n : N_t^{Y_n j} \neq N_{J_n}^{Y_n j} \text{ for some } j \neq Y_n \right\},$$

$$Y_{n+1} = \begin{cases} j & \text{if } J_{n+1} < \infty \text{ and } N_{J_{n+1}}^{Y_n j} \neq N_{J_n}^{Y_n j} \\ i & \text{if } J_{n+1} = \infty \end{cases}.$$

The first jump time of $\left(N_t^{ij} \right)$ is exponential of parameter q_{ij}. So by Proposition**??**, conditional on $Y_0 = i$, J_1 is exponential of parameter $q_i = \sum_{j \neq i} q_{ij}$, and Y_1 has distribution $(\pi_{ij} : j \in I)$, and J_1 and Y_1 are independent.

Now suppose T is a stopping time of $(X_t)_{t \geqslant 0}$. If we condition on X_0 and on the processes $\left(N_t^{kl} \right)_{t \geqslant 0}$ for $(k, l) \neq (i, j)$, which are independent of N_t^{ij} then $\{T \leqslant t\}$ depends only on $\left\{ N_s^{ij} : s \leqslant t \right\}$. So, by the strong Markov property of the Poisson process $\tilde{N}_t^{ij} := N_{T+t}^{ij} - N_T^{ij}$ is a Poisson process of rate q_{ij} independent of $\left\{N_s^{ij} : s \leqslant T\right\}$, and independent of X_0 and $\left(N_t^{kl}\right)_{t\geqslant 0}$ for $(k,l) \neq (i,j)$. Hence, conditional on $T < \infty$ and $X_T = i$, $(X_{T+t})_{t\geqslant 0}$ has the same distribution as $(X_t)_{t\geqslant 0}$ and is independent of $\{X_s : s \leqslant T\}$.

In particular, we can take $T = J_n$ to see that, conditional on $J_n < \infty$ and $Y_n = i$, S_{n+1} is exponential of parameter q_i, Y_{n+1} has distribution $(\pi_{ij})_{j\in I}$, and S_{n+1} and Y_{n+1} are independent, and independent of $Y_0, \cdots, Y_n$ and $S_1, \cdots, S_n$. Hence $(X_t)_{t\geqslant 0}$ is Markov (λ, Q) and, moreover, $(X_t)_{t\geqslant 0}$ has the strong Markov property.

The conditioning on which this argument relies requires some further justification, especially when the state-space is infinite, so we shall not rely on this third construction in the development of the theory.

EXERCISE

¶ EXERCISE 2.3. $(X_t)_{t\geqslant 0}$ is Markov(λ, Q), let $(Y_n)_{n\geqslant 0}$ be the jump chain and $(S_n)_{n\geqslant 0}$ the holding times. Then there exists $T_1, T_2, \cdots$ that are independent exponential random variables of parameter 1, independent of $(Y_n)_{n\geqslant 0}$, satisfying

$$S_n = \frac{T_n}{q\left(Y_{n-1}\right)} \quad \text{for } n \geqslant 1\,.$$

2.4.2 Explosion

We saw in the special case of birth processes that, although each holding time is strictly positive, one can run through a sequence of states with shorter and shorter holding times and end end up taking infinitely many jumps in a finite time. This phenomenon is called *explosion*. Recall that for a process with jump times $J_0, J_1, J_2, \cdots$ and holding times $S_1, S_2, \cdots$, the explosion time ζ is given by

$$\zeta = \sup_n J_n = \sum_{n=1}^{\infty} S_n$$

We say that a Q -matrix Q is **explosive** if, for the associated Markov chain

$$\mathbb{P}_i(\zeta < \infty) > 0 \quad \text{for some } i \in I \tag{2.16}$$

Otherwise Q is **non-explosive**. It is a simple consequence of the Markov property for $(Y_n)_{n\geqslant 0}$ that under $\mathbb{P}_i$ the process $(X_t)_{t\geqslant 0}$ is Markov (δ_i, Q). The next result gives simple conditions for nonexplosion and covers many cases of interest.

Theorem 2.16. *Let $(X_t)_{t\geqslant 0}$ be Markov(λ, Q). Then $(X_t)_{t\geqslant 0}$ does not explode if any one of the following conditions holds:*

(i) $\sup_{i\in I} q_i < \infty$*, particulatly, I is finite.*

(ii) $X_0 = i$*, and i is recurrent for the jump chain.*

Proof. By Exercise 2.3, there exists $T_1, T_2, \cdots$ are independent exponential(1) and independent of $(Y_n)_{n\geqslant 0}$. In cases (i), let $q = \sup_i q_i < \infty$ then

$$q\zeta \geqslant \sum_{n=1}^{\infty} T_n = \infty$$

In case (ii), since i is recurrent for $(Y_n)_{n\geqslant 0}$, so $(Y_n)_{n\geqslant 0}$ visits i infinitely often, at times $N_1, N_2, \cdots$, say. Then

$$\zeta \geqslant \sum_{k=1}^{\infty} S_{N_k+1}\,,$$

it suffices to show that $(S_{N_k+1})_{k\geqslant 1}$ is i.i.d. r.v.'s sequence with distribution expoential(q_i). For any $t_1, \cdots, t_k > 0$,

$$\begin{aligned}
&\mathbb{P}\big(S_{N_1+1} > t_1, \cdots, S_{N_k+1} > t_k\big) \\
&= \sum_{n_1,\cdots,n_k} \mathbb{P}\big(S_{n_1+1} > t_1, \cdots, S_{n_k+1} > t_k\,; N_1 = n_1, \cdots, N_k = n_k\big) \\
&= \sum_{n_1,\cdots,n_k} \mathbb{P}\big(N_1 = n_1, \cdots, N_k = n_k\big)\mathrm{e}^{-q_i(t_1+\cdots t_k)} = \mathrm{e}^{-q_i(t_1+\cdots t_k)}\,,
\end{aligned}$$

as required. $\square$

2.5 Forward and backward equations

Although the definition of a continuous-time Markov chain in terms of its jump chain and holding times provides a clear picture of the process, it does not answer some basic questions. For example, we might wish to calculate

$$p_{ij}(t) \coloneqq \mathbb{P}_i\left(X_t = j\right).$$

In this section we shall obtain two more ways of characterizing a continuous-time Markov chain, which will in particular give us a means to find $p_{ij}(t)$. Let

$$P(t) \coloneqq (p_{ij}(t))_{i,j\in I}$$

Then $P(t)$ is a substochastic matrix (if Q non-explosive, $P(t)$ is stochastic), and $\{P(t)\}$ is called transition matrix.

Proposition 2.17 (C-K equation). *For transition matrix $\{P(t)\}$, we have $P(0) = I$, and for each $t, s \geqslant 0$,*

$$P(t+s) = P(t)P(s). \tag{2.17}$$

Proof. For any $i, j \in I$,

$$\begin{aligned} p_{ij}(t+s) = \mathbb{P}_i\left(X_{t+s} = j\right) &= \sum_{k\in I} \mathbb{P}_i\left(X_t = k, X_{t+s} = j\right) \\ &= \sum_{k\in I} \mathbb{P}_i\left(X_t = k\right) \mathbb{P}_i\left(X_{t+s} = j \mid X_t = k\right) \\ &= \sum_{k\in I} \mathbb{P}_i\left(X_t = k\right) \mathbb{P}_k\left(X_s = j\right) = \sum_{k\in I} p_{ik}(t)p_{kj}(s), \end{aligned}$$

and we have used Markov property. ☐

Lemma 2.18. *$(X_t)_{t\geqslant 0}$ is Markov(λ, Q), let $(J_n)_{n\geqslant 1}$ be the jump times. Then, as $t \to 0$,*

(i) $\mathbb{P}_i\big(t < J_1\big) = 1 - q_i t + o(t)$.

(ii) $\mathbb{P}_i\big(J_1 \leqslant t < J_2\big) = q_i t + o(t)$.

(iii) $\mathbb{P}_i\big(t \geqslant J_2\big) = o(t)$.

Proof. (i) is trival, since $\mathbb{P}_i\big(t < J_1\big) = \mathrm{e}^{-q_i t} = 1 - q_i t + o(t)$.

To show (ii), note that for $j \neq i$,

$$\begin{aligned}\mathbb{P}_i\big(J_1 \leqslant t < J_2, Y_1 = j\big) &= \int_0^t q_i \mathrm{e}^{-q_i s} \pi_{ij} \mathrm{e}^{-q_j(t-s)}\,\mathrm{d}s \\ &= q_{ij}\mathrm{e}^{-q_i t}\int_0^t \mathrm{e}^{(q_j - q_i)s}\,\mathrm{d}s = q_{ij} t + o(t)\,,\end{aligned}$$

since $\sum_{j\neq i} q_{ij} < \infty$, we get

$$\mathbb{P}_i\big(J_1 \leqslant t < J_2\big) = \sum_{j\neq i} \mathbb{P}_i\big(J_1 \leqslant t < J_2, Y_1 = j\big) = q_i t + o(t)\,.$$

Obviously, (i) and (ii) imply (iii). □

Proposition 2.19. $\{P(t)\}$ *satisfies*

(i) $P(t)$ *is uniformly continuous. Indeed , for given* $i, j \in I$

$$|p_{ij}(t+h) - p_{ij}(t)| \leqslant 1 - e^{-q_i h} \quad \textit{for all } t, h \geqslant 0,\ .$$

(ii) $P'(0) = Q$. *In other words*

$$\lim_{t\downarrow 0} \frac{p_{ij}(t) - \delta_{ij}}{t} = q_{ij} \textit{ for all } i, j\,.$$

Proof. To show (i), note that

$$\begin{aligned}|p_{ij}(t+h) - p_{ij}(t)| &= \left|\sum_{k\in I} p_{ik}(h)p_{kj}(t) - p_{ij}(t)\right| \\ &= \left|\sum_{k\neq i} p_{ik}(h)p_{kj}(t) - (1 - p_{ii}(h))\,p_{ij}(t)\right| \\ &\leqslant 1 - p_{ii}(h) \leqslant \mathbb{P}_i\,(J_1 \leqslant h) = 1 - e^{-q_i h}\,.\end{aligned}$$

To show (ii), note that for $i \neq j$

$$p_{ij}(t) = \mathbb{P}_i(X_t = j, J_1 \leqslant t < J_2) + \mathbb{P}_i(X_t = j, J_2 \leqslant t)\,,$$

we have shown the first term is $q_{ij}t + o(t)$ and the second is $o(t)$ in the proof of Lemma 2.18, so

$$\lim_{t\downarrow 0} \frac{p_{ij}(t)}{t} = q_{ij}$$

For $j = i$,

$$1 - p_{ii}(t) = P_i\left(X_t \neq i\right) = \mathbb{P}_i(J_1 \leqslant t < J_2) + o(t) = q_i t + o(t)\,,$$

as required. □

We come to the key result for continuous-time Markov chains. We shall present first a version for the case of finite state-space, where there is a simpler proof. In this case there are three alternative definitions, just as for the Poisson process.

Proposition 2.20. *Let $(X_t)_{t\geqslant 0}$ be a right-continuous process with values in a finite set I. Let Q be a Q-matrix on I with jump matrix Π. Then the following three conditions are equivalent:*

(i) (jump chain/holding time definition) conditional on $X_0 = i$, the jump chain $(Y_n)_{n\geqslant 0}$ of $(X_t)_{t\geqslant 0}$ is discrete-time Markov(δ_i, Π) and for each $n \geqslant 1$, conditional on $Y_0, \cdots, Y_{n-1}$, the holding times $S_1, \cdots, S_n$ are independent exponential r.v.'s of parameters $q\left(Y_0\right), \cdots, q\left(Y_{n-1}\right)$ respectively .

(ii) (infinitesimal definition) for all $t, h \geqslant 0$, conditional on $X_t = i$, X_{t+h} is independent of $\{X_s : s \leqslant t\}$ and, as $h \to 0$, uniformly in t,

$$\mathbb{P}\left(X_{t+h} = j \mid X_t = i\right) = \delta_{ij} + q_{ij}h + o(h)\,, \quad \text{for all } j\,.$$

(iii) (transition probability definition) for all $n \geqslant 1$*, all times* $0 \leqslant t_0 < t_1 < \cdots < t_{n+1}$ *and all states* $i_0, \cdots, i_{n-1}, i, j \in I$

$$\mathbb{P}\left(X_{t_{n+1}} = j \mid X_{t_n} = i,\, X_{t_0} = i_0, \cdots, X_{t_{n-1}} = i_{n-1}\right) = p_{ij}\left(t_{n+1} - t_n\right)$$

where $(p_{ij}(t) : i, j \in I, t \geqslant 0)$ *is the (unique) solution of the forward equation*

$$P'(t) = P(t)Q, \quad P(0) = I\,.$$

or the backward equation

$$P'(t) = QP(t), \quad P(0) = I\,.$$

If $(X_t)_{t \geqslant 0}$ *satisfies any of these conditions then it is called a Markov chain with generator matrix* Q*.*

Proof. (i) $\Rightarrow$ (ii). Suppose (i) holds, then, as $h \to 0$,

$$\begin{aligned}\mathbb{P}_i\left(X_h = i\right) &= \mathbb{P}_i\left(X_h = i, J_1 > h\right) + \mathbb{P}_i(X_h = i, J_2 \leqslant h)\\ &= e^{-q_i h} + o(h) = 1 + q_{ii}h + o(h)\,,\end{aligned}$$

where we used $\mathbb{P}_i(X_h = i, J_2 \leqslant h) \leqslant \mathbb{P}_i(J_2 \leqslant h) = o(h)$. And for $j \neq i$ we have

$$\begin{aligned}\mathbb{P}_i\left(X_h = j\right) &= \mathbb{P}_i\left(X_h = j, J_1 \leqslant h < J_2\right) + \mathbb{P}_i\left(X_h = j, J_2 \leqslant h\right)\\ &= q_{ij}\int_0^h \mathrm{e}^{-q_i s}\mathrm{e}^{-q_j(h-s)}\,\mathrm{d}s + o(h) = q_{ij}h + o(h)\,.\end{aligned}$$

Thus for every state j,

$$\mathbb{P}_i\left(X_h = j\right) = \delta_{ij} + q_{ij}h + o(h)\,.$$

Then by the Markov property, for any $t, h \geqslant 0$, conditional on $X_t = i$, X_{t+h} is independent of $\{X_s : s \leqslant t\}$ and, as $h \to 0$, uniformly in t

$$\mathbb{P}\left(X_{t+h} = j \mid X_t = i\right) = \mathbb{P}_i\left(X_h = j\right) = \delta_{ij} + q_{ij}h + o(h)\,.$$

(ii) $\Rightarrow$ (iii). Set $p_{ij}(t) = \mathbb{P}_i(X_t = j)$. If (ii) holds, then for all $t, h \geqslant 0$, as $h \to 0$, uniformly in t,

$$\begin{aligned} p_{ij}(t+h) &= \sum_{k \in I} \mathbb{P}_i(X_t = k)\, \mathbb{P}(X_{t+h} = j \mid X_t = k) \\ &= \sum_{k \in I} p_{ik}(t)\left(\delta_{kj} + q_{kj} h + o(h)\right), \end{aligned} \tag{2.18}$$

since I is finite we have

$$\frac{p_{ij}(t+h) - p_{ij}(t)}{h} = \sum_{k \in I} p_{ik}(t) q_{kj} + o(1)\,.$$

So, letting $h \downarrow 0$, we see that $p_{ij}(t)$ is differentiable on the right. Then by uniformity we can replace t by $t - h$ in the above and let $h \downarrow 0$ to see first that $p_{ij}(t)$ is continuous on the left, then differentiable on the left, hence differentiable, and satisfies the forward equations

$$p'_{ij}(t) = \sum_{k \in I} p_{ik}(t) q_{kj}, \quad p_{ij}(0) = \delta_{ij}$$

since I is finite, $p_{ij}(t)$ is then the unique solution by Proposition**??**. Also, if (ii) holds, then

$$\mathbb{P}\left(X_{t_{n+1}} = j \mid X_{t_n} = i, X_{t_0} = i_0, \cdots, X_{t_{n-1}} = i_{n-1}\right) = \mathbb{P}\left(X_{t_{n+1}} = j \mid X_{t_n} = i\right)$$

and, moreover, (ii) holds for $(X_{t_n + t})_{t \geqslant 0}$ so, by the above argument,

$$\mathbb{P}\left(X_{t_{n+1}} = j \mid X_{t_n} = i\right) = p_{ij}\left(t_{n+1} - t_n\right)$$

proving (iii).

(iii) $\Rightarrow$ (i). See the proof of Theorem 2.6. $\square$

Remark. It should emphsized that we have supposed that I is a finite set in Proposition2.20, how about I is infinite ?

In the proof above, we can see that (i) $\Rightarrow$ (ii) still holds, but (ii) $\Rightarrow$ (iii) becomes problematic : in (2.18), when I is infinite, we can not guarantee that $\sum_{k \in I} p_{ik}(t) r_{kj}(h)$ is $o(h)$, where $r_{kj}(h) \coloneqq p_{kj}(h) - \delta_{kj} - q_{kj} h = o(h)$.

We turn now to the case of infinite state-space. The backward equation may still be written in the form

$$P'(t) = QP(t), \quad P(0) = I$$

only now we have an infinite system of differential equations

$$p'_{ij}(t) = \sum_{k \in I} q_{ik} p_{kj}(t), \quad p_{ij}(0) = \delta_{ij}$$

A solution to the backward equation is any matrix $(p_{ij}(t) : i, j \in I)$ of differentiable functions satisfying this system of differential equations.

Theorem 2.21. *Let Q be a Q-matrix. Then the backward equation*

$$P'(t) = QP(t), \quad P(0) = I$$

has a minimal non-negative solution $\{P(t)\}$. This solution forms a matrix semigroup

$$P(s)P(t) = P(s+t) \quad \text{for all } s, t \geqslant 0\,.$$

We shall prove this result by a probabilistic method in combination with Theorem 2.22. Note that if I is finite we must have $P(t) = \mathrm{e}^t Q$ by Proposition**??**. We call $\{P(t) : t \geqslant 0\}$ the *minimal non-negative semigroup* associated to Q, or simply the *semigroup* of Q, the qualifications minimal and non- negative being understood.

Here is the key result for Markov chains with infinite state-space. There are just two alternative definitions now as the infinitesimal characterization becomes problematic for infinite state-space.

Theorem 2.22. *Let $(X_t)_{t \geqslant 0}$ be a minimal right-continuous process with values in I. Let Q be a Q-matrix on I with jump matrix Π and semigroup $\{P(t)\}$. Then the following conditions are equivalent:*

(i) (jump chain/holding time definition) conditional on $X_0 = i$, the jump chain $(Y_n)_{n \geqslant 0}$ of $(X_t)_{t \geqslant 0}$ is discretetime Markov(δ_i, Π) and for each $n \geqslant$

1, *conditional on* $Y_0, \cdots, Y_{n-1}$, *the holding times* $S_1, \cdots, S_n$ *are independent exponential random variables of parameters* $q(Y_0), \cdots, q(Y_{n-1})$ *respectively* .

(ii) *(transition probability definition) for all* $n \geqslant 1$, *all times* $0 \leqslant t_0 < t_1 < \cdots < t_{n+1}$ *and, all states* $i_0, \cdots, i_{n-1}, i, j$

$$\mathbb{P}\left(X_{t_{n+1}} = j \mid X_{t_n} = i, X_{t_0} = i_0, \cdots, X_{t_{n-1}} = i_{n-1}\right) = p_{ij}\left(t_{n+1} - t_n\right) .$$

If $(X_t)_t \geqslant 0$ *satisfies any of these conditions then it is called a Markov chain with generator matrix* Q *.*

Proof of Theorem 2.21 and 2.22 . We know that there exists a process $(X_t)_{t \geqslant 0}$ satisfying (i). So let us define $P(t) = (p_{ij}(t))$ by

$$p_{ij}(t) = \mathbb{P}_i\left(X_t = j\right) .$$

Step 1. We show that $\{P(t)\}$ satisfies the backward equation.

Conditional on $X_0 = i$ we have $J_1 \sim \operatorname{Exp}(q_i)$ and $X_{J_1} \sim (\pi_{ik})_{k \in I}$. Then conditional on $J_1 = s$ and $X_{J_1} = k$ we have $(X_{s+t})_{t \geqslant 0} \sim \operatorname{Markov}(\delta_k, Q)$. So

$$\mathbb{P}_i\left(X_t = j, t < J_1\right) = e^{-q_i t}\delta_{ij}$$

and

$$\mathbb{P}_i\left(J_1 \leqslant t, X_{J_1} = k, X_t = j\right) = \int_0^t q_i e^{-q_i s}\pi_{ik} p_{kj}(t-s)\,\mathrm{d}s$$

Therefore

$$\begin{aligned} p_{ij}(t) &= \mathbb{P}_i\left(X_t = j, t < J_1\right) + \sum_{k \neq i} \mathbb{P}_i\left(J_1 \leqslant t, X_{J_1} = k, X_t = j\right) \\ &= e^{-q_i t}\delta_{ij} + \sum_{k \neq i} \int_0^t q_i e^{-q_i s}\pi_{ik} p_{kj}(t-s)\,\mathrm{d}s \end{aligned} \tag{2.19}$$

Make a change of variable $u = t - s$ in each of the integrals, interchange sum and integral by monotone convergence and multiply by $e^{q_i t}$ to obtain

$$e^{q_i t} p_{ij}(t) = \delta_{ij} + \int_0^t \sum_{k \neq i} q_i e^{q_i u}\pi_{ik} p_{kj}(u) du \tag{2.20}$$

This equation shows, firstly, that $p_{ij}(t)$ is continuous in t for all i, j. Secondly, the integrand is then a uniformly converging sum of continuous functions, hence continuous, and hence $p_{ij}(t)$ is differentiable in t and satisfies

$$e^{q_i t}\left(q_i p_{ij}(t) + p'_{ij}(t)\right) = \sum_{k \neq i} q_i e^{q_i t} \pi_{ik} p_{kj}(t)$$

Recall that $q_i = -q_{ii}$ and $q_{ik} = q_i \pi_{ik}$ for $k \neq i$. Then, on rearranging, we obtain

$$p'_{ij}(t) = \sum_{k \in I} q_{ik} p_{kj}(t)\,. \tag{2.21}$$

So $P(t)$ satisfies the backward equation. The integral equation (2.19) is called the integral form of the backward equation.

Step 2. We show that if $\widetilde{P}(t)$ is another non-negative solution of the backward equation, then $P(t) \leqslant \widetilde{P}(t)$, hence $P(t)$ is the minimal non-negative solution. The argument used to prove (2.19) also shows that

$$\begin{aligned}
&\mathbb{P}_i\left(X_t = j, t < J_{n+1}\right) \\
&\quad = e^{-q_i t}\delta_{ij} + \sum_{k \neq i} \int_0^t q_i e^{-q_i s} \pi_{ik} \mathbb{P}_k\left(X_{t-s} = j, t - s < J_n\right) \mathrm{d}s\,.
\end{aligned} \tag{2.22}$$

On the other hand, if $\widetilde{P}(t)$ satisfies the backward equation, then, by reversing the steps from (2.19) to 2.21, it also satisfies the integral form:

$$\widetilde{p}_{ij}(t) = e^{-q_i t}\delta_{ij} + \sum_{k \neq i} \int_0^t q_i e^{-q_i s} \pi_{ik} \widetilde{p}_{kj}(t - s)\,\mathrm{d}s\,. \tag{2.23}$$

If $\widetilde{P}(t) \geqslant 0$, then

$$\mathbb{P}_i\left(X_t = j, t < J_0\right) = 0 \leqslant \tilde{p}_{ij}(t) \quad \text{ for all } i, j \text{ and } t\,.$$

Let us suppose inductively that

$$\mathbb{P}_i\left(X_t = j, t < J_n\right) \leqslant \widetilde{p}_{ij}(t) \quad \text{ for all } i, j \text{ and } t\,,$$

then by comparing (2.22) and (2.23) we have

$$\mathbb{P}_i\left(X_t = j, t < J_{n+1}\right) \leqslant \widetilde{p}_{ij}(t) \quad \text{ for all } i, j \text{ and } t\,,$$

and the induction proceeds. Hence

$$p_{ij}(t) = \lim_{n\to\infty} \mathbb{P}_i\left(X_t = j, t < J_n\right) \leqslant \widetilde{p}_{ij}(t) \quad \text{for all } i, j \text{ and } t\,.$$

Step 3. Since $(X_t)_{t\geqslant 0}$ does not return from ∞ we have

$$\begin{aligned} p_{ij}(s+t) = \mathbb{P}_i\left(X_{s+t} = j\right) &= \sum_{k\in I} \mathbb{P}_i\left(X_{s+t} = j | X_s = k\right) \mathbb{P}_i\left(X_s = k\right) \\ &= \sum_{k\in I} \mathbb{P}_i\left(X_s = k\right) \mathbb{P}_k\left(X_t = j\right) = \sum_{k\in I} p_{ik}(s) p_{kj}(t) \end{aligned}$$

by the Markov property. Hence $\{P(t) : t \geqslant 0\}$ is a matrix semigroup. This completes the proof of Theorem 2.21.

Step 4. Suppose, as we have throughout, that $(X_t)_{t\geqslant 0}$ satisfies (i). Then, by the Markov property

$$\begin{aligned} &\mathbb{P}\left(X_{t_{n+1}} = j \mid X_{t_n} = i, X_{t_0} = i_0, \cdots, X_{t_{n-1}} = i_{n-1}\right) \\ &\qquad = \mathbb{P}_i\left(X_{t_{n+1}-t_n} = j\right) = p_{ij}\left(t_{n+1} - t_n\right) \end{aligned}$$

so $(X_t)_{t\geqslant 0}$ satisfies (ii). We complete the proof of Theorem 2.22 by the usual argument that (ii) must now imply (i) : if (ii) holds, the finite-dimensional distributions of $(X_t)_{t\geqslant 0}$ is uniquely determined, and hence the distribution of jump chain and holding times. □

So far we have said nothing about the forward equation in the case of infinite state-space. Remember that the finite state-space results (Proposition**??**) are no longer valid. The forward equation may still be written

$$P'(t) = P(t)Q, \quad P(0) = I$$

now understood as an infinite system of differential equations

$$p'_{ij}(t) = \sum_{k\in I} p_{ik}(t) q_{kj}, \quad p_{ij}(0) = \delta_{ij}$$

A solution is then any matrix $\left(p_{ij}(t)\right)_{i,j\in I}$ of differentiable functions satisfying this system of equations. We shall show that the semigroup $\{P(t)\}$ of Q

does satisfy the forward equations, by a probabilistic argument resembling Step 1 of the Proof of Theorem 2.21 and 2.22.

This time, instead of conditioning on the first event, we condition on the last event before time t. The argument is a little longer because there is no reverse-time Markov property to give the conditional distribution. We need the following time-reversal identity.

Lemma 2.23. *We have*

$$\begin{aligned} & q_{i_n}\mathbb{P}\left(J_n \leqslant t < J_{n+1} \mid Y_0 = i_0, Y_1 = i_1, \cdots, Y_n = i_n\right) \\ & = q_{i_0}\mathbb{P}\left(J_n \leqslant t < J_{n+1} \mid Y_0 = i_n, \cdots, Y_{n-1} = i_1, Y_n = i_0\right) \end{aligned}$$

Proof. Conditional on $Y_0 = i_0, \cdots, Y_n = i_n$, the holding times $S_1, \cdots, S_{n+1}$ are independent with $S_k \sim \operatorname{Exp}\left(q_{i_{k-1}}\right)$. So the left-hand side is given by

$$\int_{\Delta(t)} q_{i_n} \exp\left\{-q_{i_n}\left(t - s_1 - \cdots - s_n\right)\right\} \prod_{k=1}^{n} q_{i_{k-1}} \exp\left\{-q_{i_{k-1}} s_k\right\} \mathrm{d}s_k$$

where

$$\Delta(t) = \left\{(s_1, \cdots, s_n) : s_1 + \cdots + s_n \leqslant t \text{ and } s_1, \cdots, s_n \geqslant 0\right\}.$$

On making the substitutions $u_1 = t - s_1 - \cdots - s_n$ and $u_k = s_{n-k+2}$, for $k = 2, \cdots, n$, we obtain

$$\begin{aligned} & q_{i_n}\mathbb{P}\left(J_n \leqslant t < J_{n+1} \mid Y_0 = i_0, \cdots, Y_n = i_n\right) \\ & = \int_{\Delta(t)} q_{i_0} \exp\left\{-q_{i_0}\left(t - u_1 - \cdots - u_n\right)\right\} \prod_{k=1}^{n} q_{i_{n-k+1}} \exp\left\{-q_{i_{n-k+1}} u_k\right\} du_k \\ & = q_{i_0}\mathbb{P}\left(J_n \leqslant t < J_{n+1} \mid Y_0 = i_n, \cdots, Y_{n-1} = i_1, Y_n = i_0\right). \quad \square \end{aligned}$$

Theorem 2.24. *We have The minimal non-negative solution* $(P(t) : t \geqslant 0)$ *of the backward equation is also the minimal non-negative solution of the forward equation*

$$P'(t) = P(t)Q, \quad P(0) = I$$

Proof. Let $(X_t)_{t\geqslant 0}$ denote the minimal Markov chain with generator matrix Q. By Theorem 2.22

$$\begin{aligned}p_{ij}(t) &= \mathbb{P}_i\left(X_t = j\right)\\ &= \sum_{n=0}^{\infty}\sum_{k\neq j}\mathbb{P}_i\left(J_n \leqslant t < J_{n+1}, Y_{n-1} = k, Y_n = j\right) .\end{aligned}$$

Now by Lemma 2.23 for $n \geqslant 1$, we have

$$\begin{aligned}&\mathbb{P}_i\left(J_n \leqslant t < J_{n+1} \mid Y_{n-1} = k, Y_n = j\right)\\ &= \frac{q_i}{q_j}\mathbb{P}_j\left(J_n \leqslant t < J_{n+1} \mid Y_1 = k, Y_n = i\right)\\ &= \frac{q_i}{q_j}\int_0^t q_j e^{-q_j s}\mathbb{P}_k\left(J_{n-1} \leqslant t-s < J_n \mid Y_{n-1} = i\right)\mathrm{d}s\\ &= q_i\int_0^t e^{-q_j s}\,\frac{q_k}{q_i}\,\mathbb{P}_i\left(J_{n-1} \leqslant t-s < J_n \mid Y_{n-1} = k\right)\mathrm{d}s\end{aligned}$$

where we have used the Markov property of $(Y_n)_{n\geqslant 0}$ for the second equality. Hence

$$\begin{aligned}p_{ij}(t) &= \delta_{ij}e^{-q_i t} + \sum_{n=1}^{\infty}\sum_{k\neq j}\int_0^t \mathbb{P}_i\left(J_{n-1} \leqslant t-s < J_n \mid Y_{n-1} = k\right)\\ &\qquad\qquad \times \mathbb{P}_i\left(Y_{n-1} = k, Y_n = j\right) q_k e^{-q_j s}\,\mathrm{d}s\\ &= \delta_{ij}e^{-q_i t} + \sum_{n=1}^{\infty}\sum_{k\neq j}\int_0^t \mathbb{P}_i\left(J_{n-1} \leqslant t-s < J_n, Y_{n-1} = k\right) q_k \pi_{kj} e^{-q_j s}\,\mathrm{d}s\\ &= \delta_{ij}e^{-q_i t} + \int_0^t \sum_{k\neq j} p_{ik}(t-s) q_{kj} e^{-q_j s}\,\mathrm{d}s \qquad (2.24)\end{aligned}$$

where we have used monotone convergence to interchange the sum and integral at the last step. This is the integral form of the forward equation.

Now make a change of variable $u = t - s$ in the integral and multiply by $e^{q_j t}$ to obtain

$$p_{ij}(t)e^{q_j t} = \delta_{ij} + \int_0^t \sum_{k\neq j} p_{ik}(u) q_{kj} e^{q_j u} du \qquad (2.25)$$

We know by equation (2.20) that $e^{q_i t} p_{ik}(t)$ is increasing for all i, k. Hence either $\sum_{k \neq j} p_{ik}(u) q_{kj}$ converges uniformly for $u \in [0, t]$ or

$$\sum_{k \neq j} p_{ik}(u) q_{kj} = \infty \quad \text{for all } u \geqslant t$$

The latter would contradict (2.25) since the left-hands ide is finite for all t so it is the former which holds. We know from the backward equation that $p_{ij}(t)$ is continuous for all i, j; hence by uniform convergence the integrand in (2.25) is continuous and we may differentiate to obtain

$$p'_{ij}(t) + p_{ij}(t) q_j = \sum_{k \neq j} p_{ik}(t) q_{kj} \tag{2.26}$$

Hence $P(t)$ solves the forward equation.

To establish minimality let us suppose that $\widetilde{p}_{ij}(t)$ is another solution of the forward equation; then we also have

$$\widetilde{p}_{ij}(t) = \delta_{ij} e^{-q_i t} + \sum_{k \neq j} \int_0^t \widetilde{p}_{ik}(t-s) q_{kj} e^{-q_j s} \, \mathrm{d}s$$

A small variation of the argument leading to (2.24) shows that, for $n \geqslant 0$

$$\begin{aligned} &\mathbb{P}_i \left(X_t = j, t < J_{n+1} \right) \\ &\quad = \delta_{ij} e^{-q_i t} + \sum_{k \neq j} \int_0^t \mathbb{P}_i \left(X_t = j, t < J_n \right) q_{kj} e^{-q_j s} \, \mathrm{d}s \,. \end{aligned} \tag{2.27}$$

If $\widetilde{P}(t) \geqslant 0$, then

$$\mathbb{P} \left(X_t = j, t < J_0 \right) = 0 \leqslant \tilde{p}_{ij}(t) \quad \text{for all } i, j \text{ and } t$$

Let us suppose inductively that

$$\mathbb{P}_i \left(X_t = j, t < J_n \right) \leqslant \widetilde{p}_{ij}(t) \quad \text{for all } i, j \text{ and } t \,.$$

then by comparing (2.26) and (2.27) we obtain

$$\mathbb{P}_i \left(X_t = j, t < J_{n+1} \right) \leqslant \widetilde{p}_{ij}(t) \quad \text{for all } i, j \text{ and } t \,,$$

and the induction proceeds. Hence

$$p_{ij}(t) = \lim_{n\to\infty} \mathbb{P}_i\left(X_t = j, t < J_n\right) \leqslant \widetilde{p}_{ij}(t) \quad \text{for all } i, j \text{ and } t\,.$$

□

Chapter 3

Continuous-time Markov chains(II)

This chapter brings together the discrete-time and continuous-time theories, allowing us to deduce analogues, for continuous-time chains, of all the results given in Chapter 1. A reasonable understanding of Chapter 1 is required here, but, given such an understanding, this chapter should look reassuringly familiar.

3.1 Basic properties

3.1.1 Class structure

A first step in the analysis of a continuous-time Markov chain $(X_t)_{t\geqslant 0}$ is to identify its class structure. We emphasise that we deal only with rninimal chains, those that die after explosion. Then the class structure is simply the discrete-time class structure of the jump chain $(Y_n)_{n\geqslant 0}$, as discussed in Subection 1.1.3 before.

We say that i **leads to** j and write $i \to j$, if

$$\mathbb{P}_i\left(X_t = j \text{ for some } t \geqslant 0\right) > 0$$

We say i **communicates with** j and write $i \leftrightarrow j$ if both $i \to j$ and $j \to i$. The notions of *communicating class*, *closed class*, *absorbing state* and *irreducibility* are inherited from the jump chain.

Theorem 3.1. *For distinct states i and j, the following are equivalent:*

(i) $i \to j$ *;*

(ii) $i \to j$ *for the jump chain ;*

(iii) $q_{i_0 i_1} q_{i_1 i_2} \cdots q_{i_{n-1} i_n} > 0$ *for some states* $i_0, i_1, \cdots, i_n$ *with* $i_0 = i$ *and* $i_n = j$.

(iv) $p_{ij}(t) > 0$ *for all* $t > 0$.

(v) $p_{ij}(t) > 0$ *for some* $t > 0$.

Proof. Implications (iv) ⇒ (v) ⇒ (i) ⇒ (ii) are clear.

(ii) ⇒ (iii). $i \to j$ for the jump chain , then there are states $i_0, i_1, \cdots, i_n$ with $i_0 = i, i_n = j$ and $\pi_{i_0 i_1} \pi_{i_1 i_2} \cdots \pi_{i_{n-1} i_n} > 0$, which implies (iii).

(iii) ⇒ (iv). If $q_{ij} > 0$, then

$$p_{ij}(t) \geqslant \mathbb{P}_i\left(Y_1 = j, J_1 \leqslant t, S_2 > t\right) = \left(1 - e^{-q_i t}\right) \pi_{ij} e^{-q_j t} > 0\,.$$

for all $t > 0$, so if (iii) holds, then

$$p_{ij}(t) \geqslant p_{i_0 i_1}(t/n) \cdots p_{i_{n-1} i_n}(t/n) > 0$$

for all $t > 0$, and (iv) holds. □

3.1.2 Hitting probabilities

Let $(X_t)_{t \geqslant 0}$ be a Markov chain with generator matrix Q. The hitting time of a subset A of I is the random variable τ_A defined by

$$\tau_A(\omega) = \inf\{t \geqslant 0 : X_t(\omega) \in A\}$$

with the usual convention that $\inf \varnothing = \infty$. We emphasise that $(X_t)_{t\geqslant 0}$ is minimal. So if $\tau_A^{(Y)}$ is the hitting time of A for the jump chain, then

$$\left\{\tau_A^{(Y)} < \infty\right\} = \{\tau_A < \infty\}$$

and on this set we have

$$\tau_A = J_{\tau_A^{(Y)}}$$

The probability, starting from i, that $(X_t)_{t\geqslant 0}$ ever hits A is then

$$h_i = \mathbb{P}_i\left(\tau_A < \infty\right) = \mathbb{P}_i\left(\tau_A^{(Y)} < \infty\right)$$

When A is a closed class, h_i is called the *absorption probability.* Since the hitting probabilities are those of the jump chain we can calculate them as in Section 1.2.

Theorem 3.2. *The vector of hitting probabilities $(h_i)_{i\in I}$ is the minimal non-negative solution to the system of linear equations*

$$\begin{cases} h_i = 1, & \text{for } i \in A. \\ h_i = \sum_{j\in S} \pi_{ij} h_j, & \text{for } i \notin A. \end{cases} \tag{3.1}$$

Proof. Apply Theorem 1.8 to the jump chain. □

3.1.3 Hitting times

The average time taken, starting from i, for $(X_t)_{t\geqslant 0}$ to reach A is given by

$$t_i = \mathbb{E}_i\left(\tau_A\right), \quad \text{for } i \in I\,.$$

In calculating t_i we have to take account of the holding times so the relationship to the discrete-time case is not quite as simple.

Theorem 3.3. *Q is non-explosive. The vector of mean hitting times $(t_i)_{i\in I}$ is the minimal non-negative solution to the system of linear equations*

$$\begin{cases} t_i = 0, & \text{for } i \in A. \\ t_i = \frac{1}{q_i} + \sum_{j\neq i} \pi_{ij} t_j, & \text{for } i \notin A. \end{cases} \tag{3.2}$$

Proof. First we show that $t = (t_i)_{i \in I}$ satisfies (3.2). If $X_0 = i \in A$, then $\tau_A = 0$, so $t_i = 0$. If $X_0 = i \notin A$, then $\tau_A \geqslant J_1$, and

$$t_i = \mathbb{E}_i(\tau_A) = \mathbb{E}_i(J_1) + \sum_{j \neq i} \mathbb{E}(\tau_A - J_1 \mid Y_1 = j)\, \mathbb{P}_i(Y_1 = j)\ ,$$

by the strong Markov property of $(X_t)_{t \geqslant 0}$,

$$\mathbb{E}_i(\tau_A - J_1 \mid Y_1 = j) = \mathbb{E}_j(\tau_A) = t_j\ ,$$

thus

$$t_i = \frac{1}{q_i} + \sum_{j \neq i} \pi_{ij} t_j\ .$$

Suppose now that $y = (y_i)_{i \in I}$ is another solution to (3.2). Then $t_i = y_i = 0$ for $i \in A$. Suppose $i \notin A$, then

$$\begin{aligned} y_i &= \frac{1}{q_i} + \sum_{j \notin A} \pi_{ij} y_j = \frac{1}{q_i} + \sum_{j \notin A} \pi_{ij} \left(\frac{1}{q_j} + \sum_{k \notin A} \pi_{jk} y_k \right) \\ &= \mathbb{E}_i(S_1) + \mathbb{E}_i\left(S_2 1_{\left\{\tau_A^{(Y)} \geqslant 2\right\}} \right) + \sum_{j \notin A} \sum_{k \notin A} \pi_{ij} \pi_{jk} y_k\ . \end{aligned}$$

By repeated substitution for y in the final term we obtain after n steps

$$y_i = \mathrm{E}_i(S_1) + \cdots + \mathbb{E}_i\left(S_n 1_{\left\{\tau_A^{(Y)} \geqslant n\right\}} \right) + \sum_{j_1, \cdots, j_n \notin A} \pi_{ij_1} \cdots \pi_{j_{n-1} j_n} y_{j_n}\ .$$

Since y is non-negative

$$y_i \geqslant \sum_{m=1}^{n} \mathbb{E}_i\left(S_m 1_{\tau_A^{(Y)} \geqslant m} \right) = \mathbb{E}_i\left(\sum_{m=1}^{\tau_A^{(Y)} \wedge n} S_m \right)\ ,$$

where we use the notation $\tau_A^{(Y)} \wedge n$ for the minimum of $\tau_A^{(Y)}$ and n. Now

$$\sum_{m=1}^{\tau_A^{(Y)}} S_m = \tau_A \wedge \zeta\ .$$

since Q is non-explosive, by monotone convergence,

$$y_i \geqslant \mathbb{E}_i(\tau_A \wedge \zeta) = \mathbb{E}_i(\tau_A) = t_i,$$

as required. $\square$

3.2 Recurrence and transience

Let $(X_t)_{t\geqslant 0}$ be Markov chain with generator matrix Q. Recall that we insist $(X_t)_{t\geqslant 0}$ be minimal. We say a state i is **recurrent** if

$$\mathbb{P}_i\left(\{t \geqslant 0 : X_t = i\} \text{ is unbounded }\right) = 1\,.$$

We say that i is **transient** if

$$\mathbb{P}_i\left(\{t \geqslant 0 : X_t = i\} \text{ is unbounded }\right) = 0\,.$$

Note that if $(X_t)_{t\geqslant 0}$ can explode starting from i, then i is certainly NOT recurrent. The next result shows that, like class structure, recurrence and transience are determined by the jump chain.

Theorem 3.4. *State $i \in I$ is recurrent for the jump chain $(Y_n)_{n\geqslant 0}$, if and only if i is recurrent for $(X_t)_{t\geqslant 0}$.*

Proof. Suppose i is recurrent for $(Y_n)_{n\geqslant 0}$. Starting at i, $(X_t)_{t\geqslant 0}$ does not explode, so $J_n \to \infty$ by Theorem 2.16. Also

$$X_{J_n} = Y_n = i$$

infinitely often, so $\{t \geqslant 0 : X_t = i\}$ is unbounded, with probability 1.

Suppose i is transient for $(Y_n)_{n\geqslant 0}$. If $X_0 = i$ then

$$N = \sup\{n \geqslant 0 : Y_n = i\} < \infty$$

so $\{t \geqslant 0 : X_t = i\}$ is bounded by J_{N+1}, which is finite with probability 1, because $(Y_n)_{n\leqslant N}$ cannot include an absorbing state. $\square$

Corollary 3.5. *We have,*

(i) *every state is either recurrent or transient ;*

(ii) *recurrence and transience are class properties.*

The next result gives continuous-time analogues of the conditions for recurrence and transience found in Theorem 1.15. We denote by σ_i *the first passage time of* $(X_t)_{t\geqslant 0}$ *to state* i, defined by

$$\sigma_i = \inf\{t \geqslant J_1 : X_t = i\} \ .$$

Theorem 3.6. *The following dichotomy holds:*

(i) $q_i = 0$ *or* $\mathbb{P}_i(\sigma_i < \infty) = 1 \Leftrightarrow i$ *is recurrent* $\Leftrightarrow \int_0^\infty p_{ii}(t)dt = \infty$.

(ii) $q_i > 0$ *and* $\mathbb{P}_i(\sigma_i < \infty) < 1 \Leftrightarrow i$ *is transient* $\Leftrightarrow \int_0^\infty p_{ii}(t)dt < \infty$.

Proof. It suffices to show (i). Note that if $q_i = 0$, then $(X_t)_{t\geqslant 0}$ cannot leave i, so i is recurrent, $p_{ii}(t) = 1$ for all t, and $\int_0^\infty p_{ii}(t)dt = \infty$.

Suppose then that $q_i > 0$. Let $\sigma_i^{(Y)}$ denote the first passage time of the jump chain $(Y_n)_{n\geqslant 0}$ to state i. Then, we have

$$\sigma_i = J_{\sigma_i^{(Y)}} \ ,$$

hence

$$\mathbb{P}_i\left(\sigma_i^{(Y)} < \infty\right) = \mathbb{P}_i(\sigma_i < \infty) \ .$$

Thus i is recurrent if and only if $\mathbb{P}_i(\sigma_i < \infty) = 1$, by Theorem 1.15 and the corresponding result for the jump chain.

Claim: Let $G_{ij}^{(Y)}$ be the Green function for the jump chain, then

$$\int_0^\infty p_{ii}(t)\,\mathrm{d}t = \frac{1}{q_i} G_{ii}^{(Y)} \ . \tag{3.3}$$

So i is recurrent if and only if $\int_0^\infty p_{ii}(t)\,\mathrm{d}t = \infty$, by Theorem 1.15 an dt he corresponding result for the jump chain. To establish (3.3), we use Fubini's theorem,

$$\begin{aligned}
\int_0^\infty p_{ii}(t)\,\mathrm{d}t &= \int_0^\infty \mathbb{E}_i\left(1_{\{X_i=i\}}\right)\mathrm{d}t = \mathbb{E}_i \int_0^\infty 1_{\{X_i=i\}}\,\mathrm{d}t \\
&= \mathbb{E}_i \sum_{n=0}^\infty S_{n+1} 1_{\{Y_n=i\}} = \sum_{n=0}^\infty \mathbb{E}_i(S_{n+1} \mid Y_n = i)\,\mathbb{P}_i(Y_n = i) \\
&= \frac{1}{q_i}\sum_{n=0}^\infty \mathbb{P}_i(Y_n = i) = \frac{1}{q_i} G_{ii}^{(Y)} \ .
\end{aligned}$$

□

Let $\mathfrak{T}_i$ denote the the time $(X_t)_{t\geqslant 0}$ stayed at i, that is

$$\mathfrak{T}_i = \int_0^\infty 1_{\{X_t=i\}}\,\mathrm{d}t = \sum_{n=0}^{\infty} S_{n+1}1_{\{Y_n=i\}}\,.$$

Theorem 3.7. *Suppose $X_0 = i$ and i is transient. Then $\mathfrak{T}_i$ is a exponential random variable with parameter $q_i(1-\rho_{ii})$, where $\rho_{ii} = \mathbb{P}_i(T_i < \infty)$.*

Proof. Let V_i be the visiting times for the embedded chain to state i, that is

$$V_i = \sum_{n=0}^{\infty} 1_{\{Y_n=i\}}\,.$$

We have shown that in Subsection 1.4.2, $V_i \sim \text{geometric}(1-\rho_{ii})$.

Let $\xi_0 = 0$ and $\xi_k = \inf\{n \geqslant \xi_{k-1}+1 : Y_n = i\}$ for $k \geqslant 1$, is the k-th passage time for $(Y_n)_{n\geqslant 0}$ to state i. It's easy to see that

$$\mathfrak{T}_i = \sum_{k=1}^{V_i} S_{\xi_k+1}\,.$$

Claim : Let $\{\eta_n\}$ be i.i.d. exponential r.v.'s with parameter q_i, and $\{\eta_n\}$ is independent with V_i, then

$$\mathfrak{T}_i \overset{d}{=} \sum_{k=1}^{V_i} \eta_k\,.$$

To show this, note that

$$\mathbb{P}_i\left(S_{\xi_k+1} \leqslant t_k; 1 \leqslant k \leqslant n \mid V_i = n\right) = \mathbb{P}_i\left(\eta_k \leqslant t_k; 1 \leqslant k \leqslant n\right)$$

Thus, conditional on $\{V_i = n\}$, $\mathfrak{T}_i$ and $\sum_{k=1}^{V_i} S_{\xi_k+1}$ have the same distribution. Since n is arbitary,Which deduce that the claim above is ture.

By Proposition**??**, $\sum_{k=1}^{V_i} \eta_k \sim \text{Exp}(q_i(1-\rho_{ii}))$, so the same for $\mathfrak{T}_i$. $\square$

Finally, we show that recurrence an dt ransience are determined by any discrete-time sampling of $(X_t)_{t\geqslant 0}$, sometime is calle dt he *h-skeleton*.

Theorem 3.8. *Let $h > 0$ be given and, set $Z_n = X_{nh}$ for $n \geqslant 0$. Then $i \in I$ is recurrent for $(X_t)_{t\geqslant 0}$ if and only if i is recurrent for $(Z_n)_{n\geqslant 0}$.*

Proof. If $i \in I$ is recurrent for $(Z_n)_{n\geqslant 0}$, obviously for $(X_t)_{t\geqslant 0}$. ,

To show the necessity, note that, for $t \in [nh, (n+1)h]$, we have the estimate

$$p_{ii}((n+1)h) \geqslant e^{-q_i h} p_{ii}(t)$$

which follows from the Markov property. Then, by monotone convergence

$$\int_0^\infty p_{ii}(t)\,\mathrm{d}t \leqslant h e^{q_i h} \sum_{n=1}^\infty p_{ii}(nh)$$

an dt he result follows by Theorem 3.6 and Theorem 1.15. □

3.3 Invariant distributions

In discrete time a measure is invariant if $\lambda P = \lambda$. However there is no first $t > 0$, in continuous time we need the stronger notion: a measure λ on I is said to be **invariant** for $\{P(t) : t \geqslant 0\}$ if

$$\lambda P(t) = \lambda\,, \quad \text{for all } t \geqslant 0\,. \tag{3.4}$$

In addition, if λ is a distribution, it is called a **invariant distribution**.

The last condition is difficult to check since it involves all of the $\{P(t)\}$ and, as we have seen in the previous section, the $P(t)$ are not easy to compute.

Proposition 3.9. *Let $(X_t)_{t\geqslant 0}$ be Markov(λ, Q) and λ is an invariant measure for the semigroup $\{P(t)\}$. Then for any $s \geqslant 0$, $(X_{t+s})_{t\geqslant 0}$ is also Markov(λ, Q).*

Proof. Firstly, for each $i \in I$,

$$\mathbb{P}\left(X_s = i\right) = (\lambda P(s))_i = \lambda_i\,.$$

On the other hand, by the Markov property, we have that $(X_{s+t})_{t\geqslant 0}$ is Markov(δ_i, Q) conditional on $\{X_s = i\}$. $\square$

Based on this property, invariant distribution is also called **stationary distribution**.

Lemma 3.10. *Let Q be irreducible. If λ is an invariant measure for $\{P(t)\}$, then $\lambda = 0$, or $0 < \lambda < \infty$. Paricularly, if λ is an invariant distribution for $P(t)$, then $\lambda_i > 0$ for all i.*

Proof. Since Q is ireducible, for any $i, j \in I$, we have

$$p_{ij}(t) > 0\,, \quad \text{for all } t > 0\,.$$

Thus if $\lambda_i > 0$

$$\lambda_j \geqslant \lambda_i p_{ij}(t) > 0\,.$$ $\square$

3.3.1 Existence and uniqueness of invariant measures

In continuous-time, by iterating the equation (3.4) we get nothing. However, we still define

$$\Gamma_j^i := \mathbb{E}_i \int_0^{\sigma_i} 1_{\{X_s=j\}} \,\mathrm{d}s = \int_0^{\infty} \mathbb{P}_i \left(X_s = j, s \leqslant \sigma_i\right) \mathrm{d}s$$ [1]

for any $i, j \in I$. Then we will check if $\Gamma^i = (\Gamma_j^i)_{j\in I}$ is a invariant measure for $P(t)$. Moreover, it's easy to see that $\Gamma_i^i = \frac{1}{q_i}$.

Theorem 3.11. *Suppose i is recurrent. Then $\Gamma^i = (\Gamma_i^j)_{j\in I}$ is an invariant measure for $\{P(t)\}$.*

Remark. By the "*cycle trick*", it's easy to see that the theorem is ture. Γ_j^i is the expected time staying in state j in $[0, \sigma_i]$. Multiplying by $P(t)$ moves us forward t in time so $\left(\Gamma^i P(t)\right)_j$ is the expected time staying in state j in $[t, \sigma_i + t]$. We need the condition i is recurrent, then $\sigma_i < \infty$ and $(X_{\sigma_i+t})_{t\geqslant 0}$ is Markov(δ_i, Q), then the expected time staying in state j in $[0, t]$ and $[\sigma, \sigma_i + t]$ consides, so we have $\Gamma_j^i = \left(\Gamma^i P(t)\right)_j$.

Proof. For any $k \in S$,

$$\begin{aligned}
\sum_{j\in S} \Gamma_i^j p_{jk}(t) &= \sum_{j\in S} \int_0^{\infty} p_{jk}(t) \mathbb{P}_i \left(X_s = j, s \leqslant \sigma_i\right) \mathrm{d}s \\
&= \sum_{j\in S} \int_0^{\infty} \mathbb{P}_i \left(X_s = j, X_{s+t} = k, s \leqslant \sigma_i\right) \mathrm{d}s \\
&= \int_0^{\infty} \mathbb{P}_i \left(X_{s+t} = k, s \leqslant \sigma_i\right) \mathrm{d}s \\
&= \int_t^{\infty} \mathbb{P}_i \left(X_u = k, u \leqslant \sigma_i + t\right) du \\
&= \mathbb{E}_i \int_t^{\sigma_i+t} 1_{\{X_u=k\}} \,\mathrm{d}u
\end{aligned}$$

[1] In fact, we omitted that $X : [0, \infty) \times \Omega \to I\,;\ (t, \omega) \mapsto X_t(\omega)$ is measurable.

Since i is recurrent, under $\mathbb{P}_i$ we have $\sigma_i < \infty$, and by strong Markov property, $(X_{\sigma_i+t})_{t\geqslant 0}$ is Markov(δ_i, Q), thus

$$\mathbb{E}_i \int_{\sigma_i}^{\sigma_i+t} 1_{\{X_u=k\}} du = \mathbb{E}_i \int_0^t 1_{\{X_{u+\sigma_i}=k\}} du = \mathbb{E}_i \int_0^t 1_{\{X_u=k\}} du = \Gamma_k^i \,.$$

So we get $\sum_{j\in S} \Gamma_j^i p_{jk}(t) = \Gamma_k^i$, in other words, $\Gamma^i P(t) = \Gamma^i$ for all $t \geqslant 0$. □

Remark. Without the condition that i is recurrent, we can see Γ^i is not always invariant from the proof above.

Now we have a sufficient condition for the existence of invariant measure, how about uniqueness?

Theorem 3.12. *Let Q be irreducible and recurrent, then $\{P(t)\}$ has an unique invariant measure up to scalar multiples.*

Proof. Since Q irreducible and recurrent, it's h-skeleton is irreducible and recurrent. The invariant measure of $P(t)$ is obviously invariant for $P(h)$, but $P(h)$ has an unique invariant measure. □

3.3.2 Existence and uniqueness of invariant distributions

Firstly, if $\{P(t)\}$ has an invariant measure λ and $\Lambda := \sum_{i\in S} \lambda_i < \infty$, we defnie

$$\tilde{\lambda}_i = \frac{\lambda_i}{\Lambda}\,, \quad \text{for all } i \in I\,.$$

Then $\tilde{\lambda}$ is an invariant distribution for $\{P(t)\}$. Suppose that i isrecurrent now,

$$\begin{aligned}\sum_{j\in S} \Gamma_j^i &= \sum_{j\in S} \mathbb{E}_i \int_0^{\sigma_i} 1_{\{X_t=j\}}\, \mathrm{d}t = \mathbb{E}_i \int_0^{\sigma_i} \sum_{j\in S} 1_{\{X_t=j\}}\, \mathrm{d}t \\ &= \mathbb{E}_i \int_0^{\sigma_i} 1_{\{t<\zeta\}}\, \mathrm{d}t = \mathbb{E}_i(\sigma_i \wedge \zeta)\,.\end{aligned} \tag{3.5}$$

Note that the chain starting at i will not explode, thus

$$\sum_{j\in S}\Gamma_j^i = \mathbb{E}_i(\sigma_i)\,, \tag{3.6}$$

is exactly **the expected return time to** i when starting from i, and write $\mathbb{E}_i\sigma_i$ as m_i for short. We shall introduce the following definition .

Definition 3.1. We say a state i is **positive recurrent** if $q_i = 0$, or $m_i < \infty$, and a recurrent state which fails to have this stronger property is called **null recurrent**.

Thus, if I has a positive recurrent i, by normalizing Γ^i, we get an invariant distribution $\lambda = \frac{\Gamma^i}{m_i}$. The next theorem says that positive recurrence is a class property, and an irreducible chain has invariant distribution iff it is positive recurrent.

Theorem 3.13. *Let Q be irreducible. Then the following are equivalent:*

(i) *every state is positive recurrent.*

(ii) *some state i is positive recurrent.*

(iii) *$P(t)$ has an invariant distribution.*

(iv) *$P(t)$ has an unique invariant distribution λ and $\lambda_i = \frac{1}{q_i m_i}$ for all $i \in I$.*

Proof. (i) $\Rightarrow$ (ii) is obvious.

(ii) $\Rightarrow$ (iii). If i is positive recurrent, so Q is recurrent. So Γ^i is then invariant, and

$$\sum_{j\in I}\Gamma_j^i = m_i < \infty\,.$$

So $\lambda = \frac{1}{m_i}\Gamma^i$ defines an invariant distribution for P.

(iii) $\Rightarrow$ (iv) $\Rightarrow$ (i). 遍历定理. □

3.3.3 Jump chain

Note that $\{P(t)\}$ satisfies the forward equation $P'(t) = P(t)Q$, thus if λ is an invariant measure, $\lambda P(t) = \lambda$, taking the derivative and supposing the order of summation and differentiation can be exchanged, we get

$$\lambda Q = \lambda P(t)Q = \lambda P'(t) = 0\,.$$

As we can see, the equation " $\lambda Q = 0$ " describes the balance of 速率流. Thus, it's natural to ask that, what is the relationship between these two equations

$$\lambda Q = 0 \ \text{ and, } \ \lambda P(t) = \lambda \ \text{ for all } t\,.$$

A first result is that the equation $\lambda Q = 0$ tie-up with measures invariant for the jump matrix.

Theorem 3.14. *Let Q be a Q-matrix with jump matrix Π. λ, μ be two measures on I, such that $\mu_i = \lambda_i q_i$. Then λ is invariant for $\{P(t)\}$ if and only if μ is invariant for Π.*

Proof. Note that $q_i\left(\pi_{ij} - \delta_{ij}\right) = q_{ij}$ for all i, j, so

$$(\mu(\Pi - I))_j = \sum_{i\in I} \mu_i\left(\pi_{ij} - \delta_{ij}\right) = \sum_{i\in I} \lambda_i q_{ij} = (\lambda Q)_j\,. \qquad \square$$

Corollary 3.15. *Q is irreducible, and λ is a measure so that $\lambda Q = 0$, then either $\lambda = 0$ or $\lambda > 0$.*

proof. If $\lambda_i > 0$, for any j, since Q is irreducible there exists $i = i_0, \cdots, i_n = j$ so that

$$q_{i_0 i_1} \cdots q_{i_{n-1} i_n} > 0\,.$$

Thus $\lambda_j q_j \geqslant \lambda_i q_{i_0 i_i} \cdots q_{i_{n-1} i_n} > 0$. $\square$

Second proof. Let us exclude the trivial case $I = \{i\}$. Then we have $q_i > 0$ for all $i \in I$. Since μ is invariant for Π, and Π is irreducible, then either $\mu = 0$ or $\mu > 0$, so the corollary follows. □

Corollary 3.16. *Suppose that Q is irreducible and recurrent. Then Q has an invariant measure λ which is unique up to scalar multiples.*

Proof. Let us exclude the trivial case $I = \{i\}$; then irreducibility forces $q_i > 0$ for all i. We have Π is irreducible and recurrent, so Π has an invariant measure μ, which is unique up to scalar multiples. So, by Theorem 3.14 we can take $\lambda_i = \frac{\mu_i}{q_i}$ to obtain an invariant measure unique up to scalar multiples. □

We will discussing the following three questions.

- What's the relationship between the invariant measures for Markov chain and that for jump chain ?
- What's the relationship between the invariant distributions for Markov chain and that for jump chain ?
- If λ is a distribution on I, can we use the equation "$\lambda Q = 0$" to replace "$\lambda P(t) = \lambda$, $\forall\, t$" ?

Theorem 3.17. *Q is irreducible and recurrent, λ, μ are measures on I and $\mu_i = q_i\lambda_i$ for all $i \in I$. Then we have*

$$\lambda P(t) = \lambda \text{ for all } t \geqslant 0 \Leftrightarrow \mu\Pi = \mu \Leftrightarrow \lambda Q = 0\,. \tag{3.7}$$

Proof. By Corollary 3.16, Theorem 3.11 and Theorem 1.22, it suffices to show that

$$\Gamma_j^i = \frac{\gamma_j^i}{q_j} \tag{3.8}$$

where, γ_j^i is the expected time in j between visits to i for the jump chain.

Claim : If Q is irreducible, then (3.8) holds for all $i, j \in I$.

To see this, denote by $\sigma_i^{(Y)}$ the first passage time of the jump chain to state i. Using Fubini's theorem we have

$$\begin{aligned}
\Gamma_j^i &= \mathbb{E}_i \int_0^{\sigma_i} 1_{\{X_s=j\}}\, \mathrm{d}s = \mathbb{E}_i \sum_{n=0}^{\infty} S_{n+1} 1_{\left\{Y_n=j, n<\sigma_i^{(Y)}\right\}} \\
&= \sum_{n=0}^{\infty} \mathbb{E}_i \left(S_{n+1} \mid Y_n = j,\, n < \sigma_i^{(Y)}\right) \mathbb{P}_i \left(Y_n = j,\, n < \sigma_i^{(Y)}\right) \\
&= \frac{1}{q_j} \mathbb{E}_i \sum_{n=0}^{\infty} 1_{\left\{Y_n=j, n<\sigma_i^{(Y)}\right\}} = \frac{\gamma_j^i}{q_j}\,.
\end{aligned}$$

□

Theorem 3.18. *Q is irreducible, then λ, μ are measures on I and $\mu_i = q_i \lambda_i$ for all $i \in I$.*

- *Suppose λ is the invariant distribution for $\{P(t)\}$, or equivalently, (X_t) is positive recurrent, and μ can be normalized, then (Y_n) is positive recurrent.*
- *Suppose μ is the invariant distribution for Π , or equivalently, (Y_n) is positive recurrent, and λ can be normalized, then (X_t) is positive recurrent.*

Proof. (X_t) is positive recurrent or (Y_n) is positive recurrent both implies Q is recurrent, so by Theorem 3.17 the desired result is trival. □

Theorem 3.19. *Q is irreducible, λ is a distribution on I, then*

$$\lambda P(t) = \lambda \text{ for all } t \Leftrightarrow \lambda Q = 0 \text{ and } Q \text{ is non-explosive}\,. \tag{3.9}$$

Proof. If λ is the invariant distribution for $\{P(t)\}$, then Q is recurrent, so Q is non-explosive. By Theorem 3.17, $\lambda Q = 0$ follows.

If $\lambda Q = 0$, then by Theorem 3.14 μ is invariant for Π, where μ defined by $\mu_i = q_i \lambda_i$ for all i. So, in Section 1.5.2 we have shown that

$$\frac{\mu_j}{\mu_i} \geqslant \gamma_j^i \quad \text{for all } i, j\,.$$

Thus by (3.8),

$$\sum_j \Gamma_j^i = \sum_j \frac{\gamma_j^i}{q_j} \leqslant \sum_j \frac{\mu_j}{\mu_i} = \sum_j \frac{\lambda_j}{q_i \lambda_i} = \frac{1}{q_i \lambda_i} < \infty \,,$$

and since Q is non-explosive, by 3.5 we have

$$m_i = \sum_j \Gamma_j^i < \infty$$

So Q is positive recurrent, and hence recurrent, by Theorem 3.17 λ is the invariant distribution for $\{P(t)\}$. □

Counterexamples First we give a counterexample such that we can not give up the condition that Q is non-explosive in Theorem 3.19.

¶ Example 3.1. *Q is irreducible, π is a distribution on I, $\pi Q = 0$ but Q is explosive.*

Consider the birth-death process with the following diagram, where $q_{i,i+1} = 2 \times 3^i$ and $q_{i,i-1} = 3^i$ for $i \geqslant 1$. By detailed balance condition, we can find

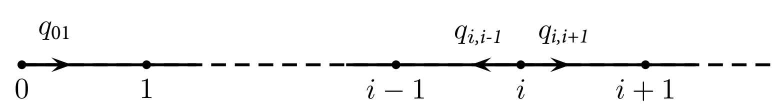


Figure 3.1: Birth-death process

the (unique) invariant distribution

$$\pi_i = \frac{1}{3} \times \left(\frac{2}{3}\right)^i \,, \text{ for all } i \geqslant 0\,.$$

However, we show that the Markov chain $(X_t)_{t \geqslant 0}$ will explode when starting at 0. Firstly, denote $G_{ii}^{(Y)}$ the Green function of the jump chain (Y_n) at state

i, then

$$\zeta = \int_0^\infty 1_{\{t<\zeta\}} dt = \int_0^\infty \sum_i 1_{\{X_t=i\}} dt = \sum_i \int_0^\infty 1_{\{X_t=i\}} dt\,.$$

Thus

$$\mathbb{E}_0\zeta = \sum_i \mathbb{E}_0 \int_0^\infty 1_{\{X_t=i\}} dt = \sum_i \mathbb{E}_i V_i = \sum_i \frac{G_{ii}^{(Y)}}{q_i}\,.$$

It's easy to find that $\{G_{ii}^{(Y)}\}_{i\geqslant 0}$ are uniformly bounded, thus $\mathbb{E}_0\zeta < \infty$ and Q is explosive.

¶ EXAMPLE 3.2. *Q is irreducible and positive recurrent, but Π is null recurrent.*

Consider the birth-death chain $(Y_n)_{n\geqslant 0}$ with the following diagram, where $b_0 = 1, b_1 = d_1 = \frac{1}{2}$ and $b_i = \frac{i-1}{2i-1}$, $d_i = \frac{i}{2i-1}$ for $i \geqslant 2$.

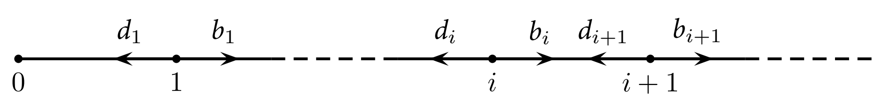


Figure 3.2: brith-death chain

Firstly, note that $d_i > b_i$ for $i \geqslant 2$, so $(Y_n)_{n\geqslant 0}$ is recurrent by Example 1.14. To show it is null recurrent, we use detailed balanced conditions,

$$\pi_i \frac{i}{2i-1} = \pi_{i-1} \frac{i-2}{2i-3}\,, \qquad \text{for } i \geqslant 3\,.$$

so

$$\frac{\pi_i}{\pi_2} = \prod_{k=3}^{i} \frac{k-2}{k} \frac{2k-1}{2k-3} = \frac{2}{i(i-1)} \times \frac{2i-1}{3} = O(\frac{1}{i})\,,$$

Thus $\sum_i \pi_i = \infty$, $(Y_n)_{n\geqslant 0}$ is null recurrent.

However, Let (Y_n) be the jump chain of $(X_t)_{t\geqslant 0}$ and $q_i = i$ for all $i \geqslant 1$, the invariant distribution of $(X_t)_{t\geqslant 0}$ is $\lambda_i = \frac{\pi_i}{q_i} = \frac{\pi_i}{i}$, so we have

$$\frac{\lambda_i}{\lambda_2} = \frac{2}{i} \times \frac{2}{i(i-1)} \times \frac{2i-1}{3} = O(\frac{1}{i^2})\,,$$

Thus $\sum_i \lambda_i < \infty$, $(X_t)_{t\geqslant 0}$ is positive recurrent.

¶ Example 3.3. *Q is irreducible and null recurrent, but Π is positive recurrent.*

Let $(Y_n)_{n\geqslant 0}$ be a positive recurrent birth-death chain with invariant distribution π. Let (Y_n) be the jump chain of $(X_t)_{t\geqslant 0}$ and $q_i = \pi_i$ for all $i \geqslant 0$. Then $\lambda_i = \frac{\pi_i}{q_i} = 1$ is an invariant measure for $\{X_t\}$, and therefore $\{X_t\}$ is null recurrent.

3.4 Time reversal

Time reversal of continuous-time chains has the same features found in the discrete-time case. Reversibility provides a powerful tool in the analysis of Markov chains, as we shall see in Section 3.6. Note in the following result how time reversal interchanges the roles of backward and forward equations. This echoes our proof of the forward equation, which rested on the time reversal identity of Lemma 2.23.

A small technical point arises in time reversal: right-continuous processes become left-continuous processes. For the processes we consider, this is unimportant. We could if we wished redefine the time-reversed process to equal its right limit at the jump times, thus obtaining again a right-continuous process. We shall suppose implicitly that this is done, and forget about the problem.

Theorem 3.20. *Let Q be irreducible and non-explosive and suppose that Q has an invariant distribution λ. Let $T \in (0, \infty)$ be given and let $(X_t)_{0 \leqslant t \leqslant T}$ be Markov(λ, Q). Set $\widehat{X}_t = X_{T-t}$. Then the process $(\widehat{X}_t)_{0 \leqslant t \leqslant T}$ is Markov$(\lambda, \widehat{Q})$, where $\widehat{Q} = (\widehat{q}_{ij})_{i,j \in I}$ is given by $\lambda_i \widehat{q}_{ij} = \lambda_j q_{ji}$. Moreover, $\widehat{Q}$ is also irreducible and non-explosive with invariant distribution λ.*

Proof. Clearly, $(\widehat{X}_t)_{0 \leqslant t \leqslant T}$ is minimal right-continuous process with values in I. For $0 \leqslant t_0 < \cdots < t_{n+1} \leqslant T$, we have

$$
\begin{aligned}
&\mathbb{P}\left(\widehat{X}_{t_{n+1}} = j \mid \widehat{X}_{t_n} = i, \widehat{X}_{t_0} = i_0, \cdots, \widehat{X}_{t_{n-1}} = i_{n-1}\right) \\
&= \mathbb{P}\left(X_{T-t_{n+1}} = j \mid X_{T-t_n} = i, X_{T-t_0} = i_0, \cdots, X_{T-t_{n-1}} = i_{n-1}\right) \\
&= \mathbb{P}\left(X_{T-t_{n+1}} = j \mid X_{T-t_n} = i\right) = \frac{\lambda_j p_{ji}(t_{n+1} - t_n)}{\lambda_i} .
\end{aligned}
$$

Define $\widehat{P}(t) = (\widehat{p}_{ij}(t))_{i,j \in I}$ by

$$\lambda_i \widehat{p}_{ij}(t) = \lambda_j p_{ji}(t) \quad \text{for all } i, j \in I ,$$

then $\widehat{P}(t)$ is an irreducible stochastic matrix with invariant distribution λ and we can rewrite the forward equation of $\{P(t)\}$ transposed as

$$\widehat{P}'(t) = \widehat{Q}\widehat{P}(t)\,.$$

But this is the backward equation for $\widehat{Q}$, which is itself a Q-matrix, and $\widehat{P}(t)$ is then its minimal non-negative solution. Hence $(\widehat{X}_t)_{0\leqslant t\leqslant T}$ is Markov$(\lambda, \widehat{Q})$.

The chain $(\widehat{X}_t)_{0\leqslant t\leqslant T}$ is called the *time-reversal* of $(X_t)_{0\leqslant t\leqslant T}$. ☐

Reversibility and detailed balance Let $(X_t)_{t\geqslant 0}$ be Markov (λ, Q), with Q irreducible and non-explosive. We say that $(X_t)_{t\geqslant 0}$ is **reversible** if, for all $T > 0, (X_{T-t})_{0\leqslant t\leqslant T}$ is also Markov(λ, Q).

A Q-matrix Q and a measure λ are said to be in detailed balance if

$$\lambda_i q_{ij} = \lambda_j q_{ji} \quad \text{for all } i, j\,.$$

Clearly, If Q and λ are in **detailed balance** then $\lambda Q = 0$.

Theorem 3.21. *Let Q be an irreducible and non-explosive Q-matrix and let λ be a distribution. Suppose that $(X_t)_{t\geqslant 0}$ is Markov(λ, Q). Then the following are equivalent:*

(i) $(X_t)_{t\geqslant 0}$ *is reversible;*

(ii) Q *and* λ *are in detailed balance.*

Proof. Both (i) and (ii) imply that λ is invariant for Q. Then both (i) and (ii) are equivalent to the statement that $\widehat{Q} = Q$ in Theorem 3.20. ☐

3.5 Long-run behavior

We now investigate the limiting behaviour of $p_{ij}(t)$ as $t \to \infty$ and its relation to invariant distributions. You will see that the situation is analogous to the case of discrete-time, only there is no longer any possibility of periodicity. We shall need the following estimate of uniform continuity for the transition probabilities.

3.5.1 Ergodic theorem

Long-run averages for continuous-time chains display the same sort of behaviour as in the discrete-time case, and for similar reasons. Here is the result.

Theorem 3.22 (Ergodic theorem). *Let Q be irreducible and let λ be any distribution. $(X_t)_{t\geq 0}$ is Markov(λ, Q), then as $t \to \infty$,*

$$\frac{1}{t}\int_0^t 1_{\{X_s=i\}}\,\mathrm{d}s \to \frac{1}{q_i m_i} \quad \text{a.s.}$$

where $m_i = \mathbb{E}_i(\sigma_i)$ is the expected return time to state i.

Proof. Suppose $q_i > 0$ for all $i \in I$. If Q is transient then the total time spent in any state i is finite, clearly

$$\frac{1}{t}\int_0^t 1_{\{X_s=i\}}\,\mathrm{d}s \leq \frac{1}{t}\int_0^\infty 1_{\{X_s=i\}}\,\mathrm{d}s \to 0 = \frac{1}{q_i m_i}\,.$$

Suppose then that Q is recurrent and fix a state i. Then $(X_t)_{t\geq 0}$ hits i with probability 1 and the long-run proportion of time in i equals the longrun proportion of time in i after first hitting i. So, by the strong Markov property (of the jump chain), it suffices to consider the case $\lambda = \delta_i$.

Denote by $T_i^{(n)}$ the time of the n th return to i, and by $M_i^{(n)}$ the length of the n th visit to i. That is, for $n \in \mathbb{N}$, setting $T_i^{(0)} = 0$, we have

$$\begin{aligned} M_i^{(n+1)} &= \inf\left\{t > T_i^{(n)} : X_t \neq i\right\} - T_i^{(n)}\,, \\ T_i^{(n+1)} &= \inf\left\{t > T_i^{(n)} + M_i^{(n+1)} : X_t = i\right\}\,. \end{aligned}$$

Now, for $T_i^{(n)} \leq t < T_i^{(n+1)}$ we have

$$\frac{M_i^{(1)} + \cdots + M_i^{(n)}}{T_i^{(n+1)}} \leq \frac{1}{t}\int_0^t 1_{\{X,=i\}}\,\mathrm{d}s \leq \frac{M_i^{(1)} + \cdots + M_i^{(n+1)}}{T_i^{(n)}}\,.$$

Denote by $L_i^{(n)}$ the length of the n th excursion to i, that is,

$$L_i^{(n)} = T_i^{(n+1)} - T_i^{(n)}\,.$$

By the strong Markov property (of the jump chain) at the stopping times $T_i^{(n)}$ for $n \in \mathbb{N}$, we find that $L_i^{(1)}, L_i^{(2)}, \cdots$ are independent and identically distributed with mean m_i, and that $M_i^{(1)}, M_i^{(2)}, \cdots$ are independent and identically distributed with mean $\frac{1}{q_i}$. Hence, by the strong law of large numbers, as $n \to \infty$

$$\frac{T_i^{(n)}}{n} = \frac{L_i^{(1)} + \cdots + L_i^{(n)}}{n} \to m_i \quad \text{a.s.}$$
$$\frac{M_i^{(1)} + \cdots + M_i^{(n)}}{n} \to \frac{1}{q_i} \quad \text{a.s.}$$

and hence

$$\frac{M_i^{(1)} + \cdots + M_i^{(n)}}{T_i^{(n)}} \to \frac{1}{m_i q_i} \quad \text{a.s.}$$

So on letting $t \to \infty$ we have, with probability 1

$$\frac{1}{t}\int_0^t 1_{\{X,=i\}}\,\mathrm{d}s \to \frac{1}{m_i q_i}\,. \qquad \square$$

Corollary 3.23. *Let Q be irreducible and positive recurrent with the invariant distribution λ. $(X_t)_{t\geq 0}$ is Markov(λ, Q), $f : I \to \mathbb{R}$ is a bounded function. Then as $t \to \infty$,*

$$\frac{1}{t}\int_0^t f(X_s)\,\mathrm{d}s \to \int_I f\,\mathrm{d}\lambda \quad \text{a.s.}$$

Proof. In the positive recurrent case we can write

$$\frac{1}{t}\int_0^t f(X_s)\,ds - \int_I f\,\mathrm{d}\lambda = \sum_{i\in I} f_i\left(\frac{1}{t}\int_0^t 1_{\{X,s=i\}}ds - \lambda_i\right)\,.$$

where $\lambda_i = \frac{1}{m_i q_i}$. We conclude that

$$\frac{1}{t}\int_0^t f(X_s)\,ds \to \int_I f\,\mathrm{d}\lambda \quad \text{as } t \to \infty\,.$$

with probability 1, by the same argument as was used in the proof of Theorem 1.35. □

3.5.2 Convergence to equilibrium

Theorem 3.24 (Convergence to equilibrium). *Let Q be an irreducible non-explosive Q-matrix with semigroup $\{P(t)\}$, and having an invariant distribution λ. Then for all states i, j we have*

$$p_{ij}(t) \to \lambda_j \quad \text{as} \quad t \to \infty\,.$$

Proof. Let $(X_t)_{t\geq 0}$ be Markov (δ_i, Q). Fix $h > 0$ and consider the h-skeleton $Z_n = X_{nh}$. Then

$$\mathbb{P}\left(Z_{n+1} = j \mid Z_n = i, Z_0 = i_0, \cdots, Z_{n-1} = i_{n-1}\right) = p_{ij}(h)$$

so $(Z_n)_{n\geq 0}$ is discrete-time Markov$(\delta_i, P(h))$. By Theorem 3.1 irreducibility implies $p_{ij}(h) > 0$ for all i, j so $P(h)$ is irreducible and aperiodic. Clearly, λ is invariant for $P(h)$. So, by discrete-time convergence to equilibrium, for all $i, j \in I$

$$p_{ij}(nh) \to \lambda_j \quad \text{as} \quad n \to \infty\,.$$

Thus we have a lattice of points along which the desired limit holds; we fill in the gaps using uniform continuity, Proposition2.19. Fix a state i, given $\epsilon > 0$ we can find $h > 0$ so that

$$1 - e^{-q_i t} \leq \frac{\epsilon}{2} \quad \text{for } 0 \leq t \leq h$$

and then find N, so that

$$|p_{ij}(nh) - \lambda_j| \leq \epsilon/2 \quad \text{for } n \geq N$$

For $t \geq Nh$ we have $nh \leq t < (n+1)h$ for some $n \geq N$ and

$$|p_{ij}(t) - \lambda_j| \leq |p_{ij}(t) - p_{ij}(nh)| + |p_{ij}(nh) - \lambda_j| \leq \varepsilon$$

by Proposition2.19. Hence

$$p_{ij}(t) \to \lambda_j \quad \text{as} \quad n \to \infty\,. \qquad \square$$

The complete description of limiting behaviour for irreducible chains in continuous time is provided by the following result. We do not give the details.

Theorem 3.25. *Let Q be an irreducible non-explosive Q-matrix and let λ be any distribution. Suppose that $(X_t)_{t\geqslant 0}$ is Markov(λ, Q). Then for all states i*

$$\mathbb{P}(X_t = i) \to \frac{1}{q_i m_i} \quad \text{as} \quad t \to \infty\,.$$

Proof. Let $(Y_n)_{n\geqslant 0}$ be the embedded chain, and let

$$\sigma_i^{(Y)} = \inf\{n \geqslant 1 : Y_n = i\}\,.$$

It follows from Theorem 1.41, by the same argument we used in the proof Theorem 3.24,

$$\mathbb{P}(X_t = i) \to \frac{1}{\mathbb{E}_i \sigma_i^{(Y)}}$$

Note that $m_i = \mathbb{E}_i \sigma_i = \frac{1}{q_i}\mathbb{E}_i \sigma_i^{(Y)}$, we have completed the proof now. $\square$

3.6 Queues and queueing networks

Queues form in many circumstances and it is important to be able to predict their behaviour. The basic mathematical model for queues runs as follows: there is a succession of customers wanting service; on arrival each customer must wait until a server is free, giving priority to earlier arrivals; it is assumed that the times between arrivals are independent random variables of the same distribution, and the times taken to serve customers are also independent random variables, of some other distribution.

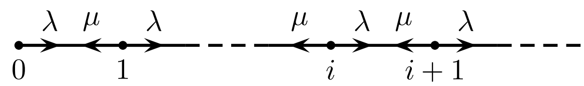
The main quantity of interest is the random process $(X_t)_{t \geqslant 0}$ recording the number of customers in the queue at time t. This is always taken to include both those being served and those waiting to be served.

In cases where inter-arrival times and service times have exponential distributions, $(X_t)_{t \geqslant 0}$ turns out to be a continuous-time Markov chain, so we can answer many questions about the queue.

In each example we shall aim to describe some salient features of the queue in terms of the given data of arrival-time and service-time distributions.

- We shall find conditions for the stability of the queue .
- In the stable case find means to compute the equilibrium distribution of queue length.
- We shall also look at the random times that customers spend waiting and the length of time that servers are continuously busy.

¶ Example 3.4 (M/M/1 queue). This is the simplest queue of all. The code means: memoryless inter-arrival times/memoryless service times/one server. Let us suppose that the inter-arrival times are exponential of parameter λ, and the service times are exponential of parameter μ. Then the number of customers in the queue $(X_t)_{t \geqslant 0}$ evolves as a Markov chain with the following diagram:



- The M/M/1 queue thus evolves like a random walk, except that it does not take jumps below 0. We deduce that

  (i) if $\lambda > \mu$ then $(X_t)_{t\geqslant 0}$ is transient, that is $X_t \to \infty$ as $t \to \infty$. Thus if $\lambda > \mu$ the queue grows without limit in the long term ;

  (ii) when $\lambda < \mu$, $(X_t)_{t\geqslant 0}$ is positive recurrent with invariant distribution
  $$\pi_i = \left(1 - \frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^i , \text{ for } i \geqslant 0 .$$

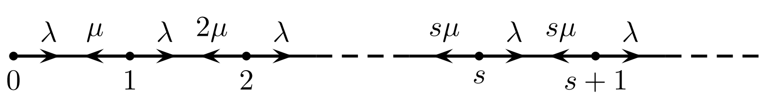
- When $\lambda < \mu$ the average number of customers in the queue in equilibrium is given by
$$\mathbb{E}_\pi\left(X_t\right) = \sum_{i=1}^{\infty} \mathbb{P}_\pi\left(X_t \geqslant i\right) = \sum_{i=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^i = \frac{\lambda}{\mu - \lambda} .$$

- Also, the mean time to return to 0 is given by
$$m_0 = \frac{1}{q_0 \pi_0} = \frac{\mu}{\lambda(\mu - \lambda)} ,$$
so the mean length of time that the server is continuously busy is given by
$$m_0 - \frac{1}{q_0} = \frac{1}{\mu - \lambda} .$$

- Another quantity of interest is the mean waiting time for a typical customer, when $\lambda < \mu$ and the queue is in equilibrium. Conditional on finding a queue of length i on arrival, this is $\frac{i+1}{\mu}$, so the overall mean waiting time is
$$\frac{\mathbb{E}_\pi\left(X_t + 1\right)}{\mu} = \frac{1}{\mu - \lambda} .$$

Thus, once the queue size is identified as a Markov chain, its behaviour is largely understood. Even in more complicated examples where exact calculation is limited, once the Markovian character of the queue is noted we know what sort of features to look for transience and recurrence, convergence to equilibrium, long-run averages, and so on.

¶ EXAMPLE 3.5 (M/M/s queue). This is a variation on the last example where there is one queue but there are s servers. Let us assume that the arrival rate is λ and the service rate by each server is μ. Then if i servers are occupied, the first service is completed at the minimum of i independent exponential times of parameter μ. The first service time is therefore exponential of parameter $i\mu$. The total service rate increases to a maximum sμ when all servers are working. We emphasise that the queue size includes those customers who are currently being served. The queue size $(X_t)_{t\geqslant 0}$ performs a Markov chain with the following diagram:



So this time we obtain a birth-and-death chain. It is transient in the case $\lambda > s\mu$ and otherwise recurrent. To find an invariant measure we look at the detailed balance equations

$$\pi_i q_{i,i+1} = \pi_{i+1} q_{i+1,i}$$

Hence

$$\frac{\pi_i}{\pi_0} = \begin{cases} \dfrac{1}{i!}(\dfrac{\lambda}{\mu})^i & \text{for } i = 0, 1, \cdots, s \\ \dfrac{1}{s^{i-s}s!}(\dfrac{\lambda}{\mu})^i & \text{for } i \geqslant s+1\,. \end{cases}$$

The queue is therefore positive recurrent when $\lambda < s\mu$. There are two cases when the invariant distribution has a particularly nice form: when $s = 1$ we are back to Example 3.4 and the invariant distribution is geometric of parameter $\frac{\lambda}{\mu}$:

$$\pi_i = (1 - \frac{\lambda}{\mu})(\frac{\lambda}{\mu})^i \, .$$

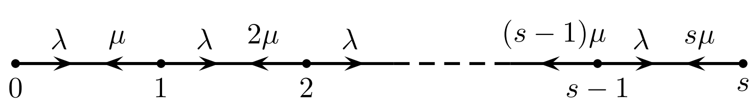
When $s = \infty$ we normalize π by taking $\pi_0 = \mathrm{e}^{-\frac{\lambda}{\mu}}$ so that

$$\pi_i = \mathrm{e}^{-\frac{\lambda}{\mu}} \frac{1}{i!} (\frac{\lambda}{\mu})^i \, .$$

and the invariant distribution is Poisson of parameter $\frac{\lambda}{\mu}$.

The number of arrivals by time t is a Poisson process of rate λ. Each arrival corresponds to an increase in X_t, and each departure to a decrease. Let us suppose that $\lambda < s\mu$, so there is an invariant distribution, and consider the queue in equilibrium. The detailed balance equations hold and $(X_t)_{t \geqslant 0}$ is non-explosive, so by Theorem 3.21 for any $T > 0, (X_t)_{0 \leqslant t \leqslant T}$ and $(X_{T-t})_{0 \leqslant t \leqslant T}$ have the same law. It follows that, in equilibrium, the number of departures by time t is also a Poisson process of rate λ. This is slightly counter-intuitive, as one might imagine that the departure process runs in fits and starts depending on the number of servers working. Instead, it turns out that the process of departures, in equilibrium, is just as regular as the process of arrivals.

¶ Example 3.6 (Telephone exchange). A variation on the M/M/s queue is to turn away customers who cannot be served immediately. This might serve as a simple model for a telephone exchange, where the maximum number of calls that can be connected at once is s: when the exchange is full, additional calls are lost. The maximum queue size or *buffer size* is s and we get the following modified Markov chain diagram :



We can find the invariant distribution of this finite Markov chain by solving the detailed balance equations, as in the last example. This time we get a truncated Poisson distribution

$$\pi_i = \frac{(\lambda/\mu)^i}{i!} \Big/ \sum_{j=0}^{s} \frac{(\lambda/\mu)^j}{j!} \text{ for } i = 0, \cdots, s\,.$$

By the ergodic theorem, the long-run proportion of time that the exchange is full, and hence the long-run proportion of calls that are lost, is given by

$$\pi_s = \frac{(\lambda/\mu)^s}{s!} \Big/ \sum_{j=0}^{s} \frac{(\lambda/\mu)^j}{j!}$$

This is known as *Erlang's formula.* [2]

¶ EXAMPLE 3.7 (Queues in series). Suppose that customers have two service requirements: they arrive as a Poisson process of rate λ to be seen first by server A, and then by server B. For simplicity we shall assume that the service times are independent exponentials of parameters α and β respectively. What is the average queue length at B?

Let us denote the queue length at A by $(X_t)_{t\geqslant 0}$ and that by B by $(Y_t)_{t\geqslant 0}$ Then $(X_t)_{t\geqslant 0}$ is simply an M/M/1 queue. If $\lambda > \alpha$, then $(X_t)_{t\geqslant 0}$ is transient so there is eventually always a queue at A and departures form a Poisson process of rate α. If $\lambda < \alpha$, then, by the reversibility argument of Example 3.5, the process of departures from A is Poisson of rate λ, provided queue A is in equilibrium. The question about queue length at B is not precisely formulated: it does not specify that the queues should be in equilibrium; indeed if $\lambda \geqslant \alpha$ there is no equilibrium.

Nevertheless, we hope you will agree to treat arrivals at B as a Poisson process of rate $\alpha \wedge \lambda$. Then, by Example 3.4, the average queue length at B when $\alpha \wedge \lambda < \beta$, in equilibrium, is given by $(\alpha \wedge \lambda)/(\beta - (\alpha \wedge \lambda))$. If, on the other hand, $\alpha \wedge \lambda > \beta$, then $(Y_t)_{t\geqslant 0}$ is transient so the queue at B grows without limit.

[2] 严格证明见应随考前题

There is an equilibrium for both queues if $\lambda < \alpha$ and $\lambda < \beta$. The fact that in equilibrium the output from A is Poisson greatly simplifies the analysis of the two queues in series. For example, the average time taken by one customer to obtain both services is given by

$$\frac{1}{\alpha - \lambda} + \frac{1}{\beta - \lambda}\,.$$

Chapter 4

Brownian Motion

Motivation

Much of probability theory is devoted to describing the *macroscopic picture* emerging in random systems defined by a host of *microscopic random effects*. Brownian motion is the macroscopic picture emerging from a particle moving randomly in d-dimensional space without making very big jumps. On the microscopic level, at any time step, the particle receives a random displacement, caused for example by other particles hitting it or by an external force, so that, if its position at time zero is S_0, its position at time n is given as $S_n = S_0 + \sum_{i=1}^{n} X_i$, where the displacements $X_1, X_2, X_3, \ldots$ are assumed to be independent, identically distributed random variables with values in $\mathbb{R}^d$. The process $\{S_n : n \geqslant 0\}$ is a *random walk*, the displacements represent the microscopic inputs. When we think about the macroscopic picture, what we mean is questions such as:

- Does S_n drift to infinity?
- Does S_n return to the neighbourhood of the origin infinitely often?
- What is the speed of growth of $\max\{|S_1|, \ldots, |S_n|\}$ as $n \to \infty$?

- What is the asymptotic number of windings of $\{S_n : n \geqslant 0\}$ around the origin?

It turns out that not all the features of the microscopic inputs contribute to the macroscopic picture. Indeed, if they exist, only the *mean* and *covariance* of the displacements are shaping the picture. In other words, all random walks whose displacements have the same mean and covariance matrix give rise to the same macroscopic process, and even the assumption that the displacements have to be independent and identically distributed can be substantially relaxed. This effect is called *universality*, and the macroscopic process is often called a *universal object*. It is a common approach in probability to study various phenomena through the associated universal objects.

If the jumps of a random walk are sufficiently tame to become negligible in the macroscopic picture, in particular if it has finite mean and variance, any continuous time stochastic process $\{B(t) : t \geqslant 0\}$ describing the macroscopic features of this random walk should have the following properties:

(i) for all times $0 \leqslant t_1 \leqslant t_2 \leqslant \ldots \leqslant t_n$ the random variables

$$B\left(t_n\right) - B\left(t_{n-1}\right), B\left(t_{n-1}\right) - B\left(t_{n-2}\right), \ldots, B\left(t_2\right) - B\left(t_1\right)$$

are independent; we say that the process has *independent increments*,

(ii) the distribution of the increment $B(t+h) - B(t)$ does not depend on t; we say that the process has *stationary increments*,

(iii) the process $\{B(t) : t \geqslant 0\}$ has almost surely *continuous paths*.

It follows (with some work) from the central limit theorem that these features imply that there exists a vector $\mu \in \mathbb{R}^d$ and a matrix $\Sigma \in \mathbb{R}^{d \times d}$ such that

(iv) for every $t \geqslant 0$ and $h \geqslant 0$ the increment $B(t+h)-B(t)$ is multivariate normally distributed with mean $h\mu$ and covariance matrix $h\Sigma\Sigma^{\mathrm{T}}$.

Hence any process with the features (i)-(ii) above is characterised by just three parameters,

- the *initial distribution*, i.e. the law of $B(0)$,
- the *drift vector* μ
- the *diffusion matrix* Σ.

The process $\{B(t) : t \geqslant 0\}$ is called a **d-dimensional Brownian motion with drift μ and diffusion matrix Σ**. If the drift vector is zero, and the diffusion matrix is the identity we simply say the process is a **Brownian motion**. If $B(0) = 0$, i.e. the motion is started at the origin, we use the term **standard Brownian motion**.

Suppose we have a standard Brownian motion $\{B(t) : t \geqslant 0\}$. If X is a random variable with values in $\mathbb{R}^d$, μ a vector in $\mathbb{R}^d$ and Σ a $d \times d$ matrix, then it is easy to check that $\{\tilde{B}(t) : t \geqslant 0\}$ given by

$$\tilde{B}(t) = X + \mu t + \Sigma B(t), \text{ for } t \geqslant 0$$

is a process with the properties (i)-(iv) with initial distribution X, drift vector μ and diffusion matrix Σ. Hence the macroscopic picture emerging from a random walk with finite variance can be fully described by a standard Brownian motion.

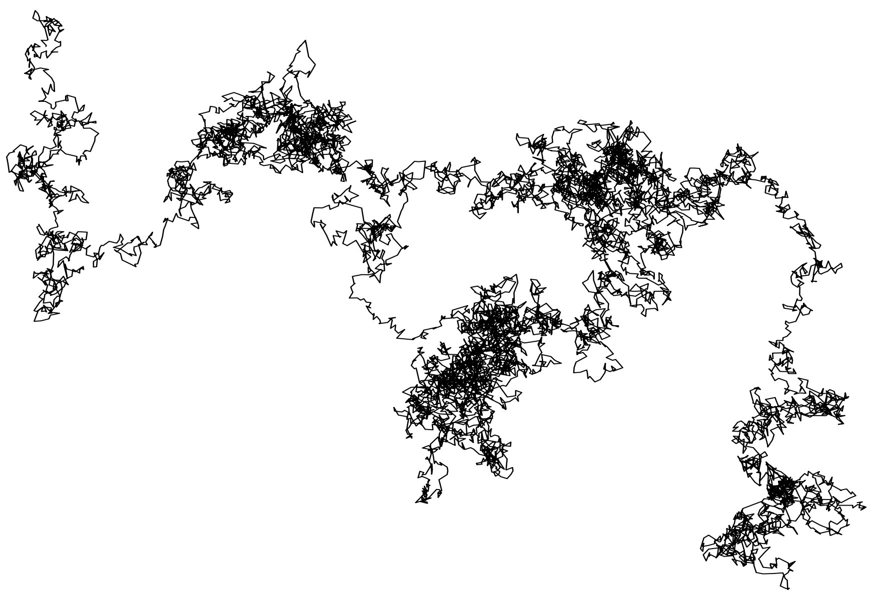


Figure 4.1: The range of a planar Brownian motion $\{B(t) : 0 \leqslant t \leqslant 1\}$.

4.1 Definition and existence

In this section, we focus on one-dimensional, or linear, Brownian motion as a random function. We start with Paul Lévy's construction of Brownian motion and discuss two fundamental sample path properties: continuity and differentiability.

4.1.1 Introduction

Definition 4.1. A real-valued stochastic process $\{B(t) : t \geqslant 0\}$ is called a (**linear**) **Brownian motion** if the following holds.

(i) The process has independent increments, that is, for any given times $0 \leqslant t_1 < t_2 < \cdots < t_n < \infty$ the increments

$$B\left(t_n\right) - B\left(t_{n-1}\right), B\left(t_{n-1}\right) - B\left(t_{n-2}\right), \cdots, B\left(t_2\right) - B\left(t_1\right)$$

are independent random variables,

(ii) For all $t \geqslant 0$ and $h > 0$, the increments $B(t+h) - B(t)$ are normally distributed with expectation 0 and variance h,

(iii) Almost surely, the function $t \mapsto B(t)$ is continuous.

We say $\{B(t) : t \geqslant 0\}$ starting at $x \in \mathbb{R}$, if $B(0) = x$, and $\{B(t) : t \geqslant 0\}$ is called a **standard Brownian motion** if $x = 0$.

Throughout this note we write $\mathbb{P}_x$ or $\mathbb{P}^x$ for the probability measure which makes $\{B(t) : t \geqslant 0\}$ a Brownian motion started in x and $\mathbb{E}_x$ or $\mathbb{E}^x$ for the corresponding expectation.

We will address the nontrivial question of the existence of a Brownian motion later, for the moment let us step back and look at some technical points. We have defined Brownian motion as a stochastic process $\{B(t)\}$ which is just a family of (uncountably many) random variables

$$\omega \mapsto B(t, \omega)$$

defined on a single probability space $(\Omega, \mathcal{F}, \mathbb{P})$. At the same time, a stochastic process can also be interpreted as a random function with the sample functions defined by

$$t \mapsto B(t, \omega).$$

The *sample path properties* of a stochastic process are the properties of these random functions, and it is these properties we most interested.

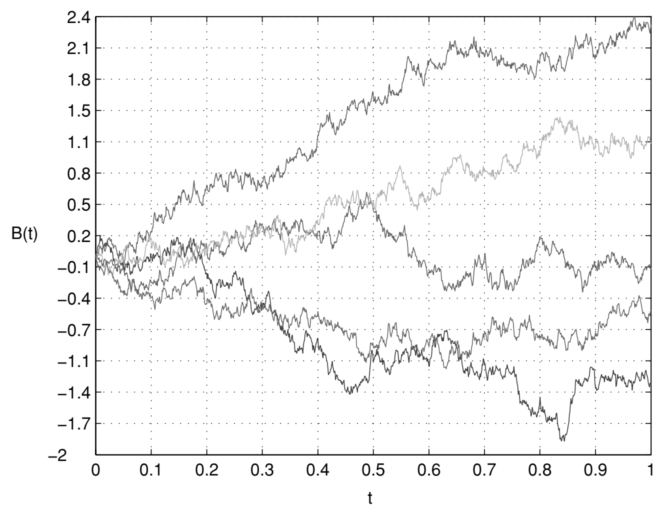


Figure 4.2: Graphs of five sampled Brownian motions

Remark. (i) By the finite-dimensional distributions of a stochastic process $\{B(t)\}$, we mean the laws of all the finite dimensional random vectors

$$\left(B\left(t_1\right), B\left(t_2\right), \cdots, B\left(t_n\right)\right), \text{ for all } 0 \leqslant t_1 < t_2 < \cdots < t_n$$

To describe these joint laws it suffices to describe the joint law of $B(0)$ and the increments

$$\big(B\left(t_1\right) - B(0), B\left(t_2\right) - B\left(t_1\right), \cdots, B\left(t_n\right) - B\left(t_{n-1}\right)\big).$$

This is what we have done in the first three items of the definition, which specify the finite-dimensional distributions of Brownian motion.

(ii) However, the last item, almost sure continuity, is also crucial, and this is information which goes beyond the finite-dimensional distributions of the process in the sense above, technically because the set

$$\{\omega \in \Omega : t \mapsto B(t, \omega) \text{ continuous } \}$$

is, in general, NOT in the σ-algebra generated by $\{B(t) : t \geqslant 0\}$.

¶ EXAMPLE 4.1. Suppose that $\{B(t) : t \geqslant 0\}$ is a Brownian motion and U is uniformly distributed on $[0, 1]$ and independent with $\{B(t)\}$. Then the process $\{\tilde{B}(t) : t \geqslant 0\}$ defined by

$$\tilde{B}(t) = \begin{cases} B(t) & \text{if } t \neq U \\ 0 & \text{if } t = U \end{cases}$$

has the same finite-dimensional distributions as a Brownian motion, but is discontinuous if $B(U) \neq 0$, i.e. with probability one, and hence this process is not a Brownian motion.

We see that, if we are interested in the sample path properties of a stochastic process, we may need to specify more than just its finite-dimensional distributions.

Suppose $\mathscr{P}$ is a property a function might or might not have, like continuity, differentiability, etc. We say that a process $\{X(t) : t \geqslant 0\}$ has property $\mathscr{P}$ almost surely, if there exists $N \in \mathcal{F}$ such that $\mathbb{P}(N) = 0$ and

$$\{\omega \in \Omega : t \mapsto X(t, \omega) \text{ doesn't have property } \mathscr{P}\} \subset N\,.$$

Note that the set on the left need not lie in $\mathcal{F}$. In fact, we can construct a probability space $(\Omega, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$, which is the completion of $(\Omega, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$, then all subset of a null set lies in $\tilde{\mathcal{F}}$.

4.1.2 Lévy's construction

It is a substantial issue whether the conditions imposed on the finite-dimensional distributions in the definition of Brownian motion allow the process to have continuous sample paths, or whether there is a contradiction. Now we show that there is no contradiction and, fortunately, Brownian motion exists.

Firstly, recall that a random vector $X = (X_1, \cdots, X_n)$ is called a **Gaussian random vector** if there exists an $n \times m$ matrix A, and an n-dimensional vector b such that

$$X^T = AY + b\,,$$

where Y is an m-dimensional vector with independent standard normal entries. As we konw, a Gaussian random vector X has independent entries if and only if its covariance matrix is diagonal. In other words, the entries in a Gaussian vector are uncorrelated if and only if they are independent.

Theorem 4.1 (Wiener's theorem). *Standard Brownian motion exists.*

Sketch : we construct Brownian motion as a uniform limit of continuous functions, to ensure that it automatically has continuous paths. We first construct Brownian motion on the interval $[0,1]$ as a random element on the space C$[0,1]$ of continuous functions on $[0,1]$. The idea is to construct the right joint distribution of Brownian motion step by step on the finite sets

$$\mathcal{D}_n = \left\{ \frac{k}{2^n} : 0 \leqslant k \leqslant 2^n \right\}$$

of dyadic points. We then interpolate the values on $\mathcal{D}_n$ linearly and check that the uniform limit of these continuous functions exists and is a Brownian motion.

Proof. **Step 1**. Let $\mathcal{D} = \bigcup_{n=0}^{\infty} \mathcal{D}_n$ and let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space on which a collection $\{Z_t : t \in \mathcal{D}\}$ of independent, standard normally distributed r.v.'s can be defined. Let $B(0) := 0$ and $B(1) := Z_1$. For each $n \in \mathbb{N}$ we define the random variables $B(d), d \in \mathcal{D}_n$ such that

(i) for all $r < s < t$ in $\mathcal{D}_n$ the random variable $B(t) - B(s)$ is normally distributed with mean zero and variance $t - s$, and is independent of $B(s) - B(r)$

(ii) the vectors $(B(d) : d \in \mathcal{D}_n)$ and $(Z_t : t \in \mathcal{D} \backslash \mathcal{D}_n)$ are independent.

Note that we have already done this for $\mathcal{D}_0 = \{0, 1\}$. Proceeding inductively we may assume that we have succeeded in doing it for some $n - 1$. We then define $B(d)$ for $d \in \mathcal{D}_n \backslash \mathcal{D}_{n-1}$ by

$$B(d) = \frac{1}{2}\left[B\big(d - \frac{1}{2^n}\big) + B\big(d + \frac{1}{2^n}\big)\right] + \frac{Z_d}{2^{\frac{n+1}{2}}}$$

Note that the first summand is the linear interpolation of the values of B at the neighbouring points of d in $\mathcal{D}_{n-1}$. Therefore $B(d)$ is independent of $(Z_t : t \in \mathcal{D} \backslash \mathcal{D}_n)$ and the second property is fulfilled.

Moreover, as

$$\frac{1}{2}\left[B\big(d - \frac{1}{2^n}\big) + B\big(d + \frac{1}{2^n}\big)\right]$$

depends only on $(Z_t : t \in \mathcal{D}_{n-1})$, it is independent of Z_d. By our induction assumptions both terms are normally distributed. Then their sum, and their difference

$$B(d) - B\big(d - \frac{1}{2^n}\big)\,,\ B\big(d + \frac{1}{2^n}\big) - B(d)$$

are independent and normally distributed with mean zero and variance $\frac{1}{2^n}$.

Indeed, all increments

$$B(d) - B\big(d - \frac{1}{2^n}\big)\,,\ d \in \mathcal{D}_n \backslash \{0\},$$

are independent. To see this, it suffices to show that they are pairwise independent, as the vector of these increments is Gaussian. We have seen in the previous paragraph that pairs

$$B(d) - B\big(d - \frac{1}{2^n}\big)\,,\ B\big(d + \frac{1}{2^n}\big) - B(d)$$

with $d \in \mathcal{D}_n \backslash \mathcal{D}_{n-1}$ are independent. The other possibility is that the increments are over intervals separated by some $d \in \mathcal{D}_{n-1}$. Choose $d \in \mathcal{D}_j$ with this property and minimal j, so that the two intervals are contained in $\left[d-2^{-j}, d\right]$, respectively $\left[d, d+2^{-j}\right]$. By induction the increments over these two intervals of length 2^{-j} are independent, and the increments over the intervals of length 2^{-n} are constructed from the independent increments $B(d)-B\left(d-2^{-j}\right)$, respectively $B\left(d+2^{-j}\right)-B(d)$ using a disjoint set of variables $\left(Z_t : t \in \mathcal{D}_n\right)$. Hence they are independent and this implies the first property, and completes the induction step.

Step 2. Having thus chosen the values of the process on all dyadic points, we interpolate between them. Formally, for each n denote by $\left(B_n(t)\right)_{t \geqslant 0}$ the continuous process obtained by linear interpolation from $\{B(d) : d \in D_n\}$.

$$B_n(t)=\begin{cases} B(d) \text{ for } t=d \in \mathcal{D}_n \\ \text{linear in between} \end{cases}$$

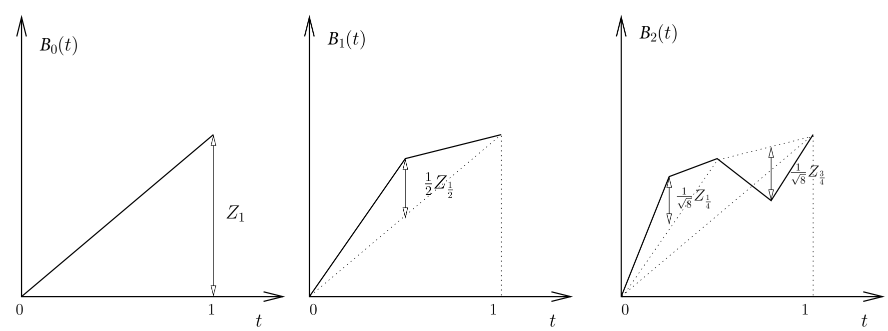


Figure 4.3: The first three steps in the construction of Brownian motion

Then, set $F_n(t)=B_n(t)-B_{n-1}(t)$, so

$$F_n(t)=\begin{cases} \dfrac{Z_d}{2^{\frac{n+1}{2}}}, \text{ for } t=d \in \mathcal{D}_n \backslash \mathcal{D}_{n-1} \\ 0, \text{ for } t=d \in \mathcal{D}_{n-1} \\ \text{linear between consecutive points in } \mathcal{D}_n \end{cases}$$

Clearly,

$$\|F_n\|_\infty = \sup\left\{\frac{|Z_d|}{2^{(n+1)/2}} : d \in \mathcal{D}_n \backslash \mathcal{D}_{n-1}\right\}.$$

So for large n,

$$\mathbb{P}\left(\|F_n| \geqslant \frac{c\sqrt{n}}{2^{\frac{n+1}{2}}}\right) \leqslant 2^n \mathbb{P}\left(|Z_1| \geqslant c\sqrt{n}\right) \leqslant 2^n \exp\left(\frac{-c^2 n}{2}\right).$$

Take $c > \sqrt{2\log 2}$, by the Borel-Cantelli lemma there exists a random (but almost surely finite) N such that

$$\|F_n\|_\infty < c\sqrt{n}2^{-\frac{n}{2}}, \quad \text{for all } n \geqslant N.$$

This upper bound implies that, almost surely, the sequence $\{B_n\}$ is uniformly convergent on $[0,1]$. We denote the continuous limit by

$$\{B(t) : t \in [0,1]\}.$$

Step 3. We check that the increments of this process have the right finite-dimensional distributions, namely *Brownian distribution.* This follows directly from the properties of B on the dense set $\mathcal{D} \subset [0,1]$ and the continuity of the paths. Indeed, suppose that $t_1 < t_2 < \cdots < t_n$ are in $[0,1]$. We find $\{t_{i,k}\}$ in $\mathcal{D}$ so that $t_{i,k} \uparrow t_k$ as $i \to \infty$, and infer from the continuity of B that, for $1 \leqslant i \leqslant n-1$,

$$B\left(t_{i+1}\right) - B\left(t_i\right) = \lim_{k\uparrow\infty} B\left(t_{i+1,k}\right) - B\left(t_{i,k}\right).$$

Using the characteristic function of Gsussian r.v., we have that the increments $B\left(t_{i+1}\right) - B\left(t_i\right)$ are independent Gaussian r.v's with mean 0 and variance $t_{i+1} - t_i$.

We have thus constructed a continuous process $B : [0,1] \to \mathbb{R}$ with the same finite-dimensional distributions as Brownian motion. Take a sequence $B_0, B_1, \cdots$ of indepen- dent $\mathbf{C}[0,1]$-valued random variables with the distribution of this process, and define $\{B(t) : t \geqslant 0\}$ by gluing together the parts, more precisely by

$$B(t) = B_{[t]}(t - [t]) + \sum_{i=0}^{[t]-1} B_i(1), \text{ for all } t \geqslant 0.$$

This defines a continuous random function $B : [0, \infty) \to \mathbb{R}$ and one can see easily from what we have shown so far that it is a standard Brownian motion. $\square$

Remark. If Brownian motion is constructed as a family $\{B(t) : t \geqslant 0\}$ of random variables on some probability space Ω, it is sometimes useful to know that the mapping $(t, \omega) \mapsto B(t, \omega)$ is measurable on the product space $[0, \infty) \times \Omega$. We point that this can be achieved by Lévy's construction.

4.2 Gaussian processes, Lévy processes

In this section, we introduce two classical processes, Gaussian processes, Lévy processes and we will see that standard Brownian motion is a special case of these two.

4.2.1 Gaussian processes

Definition 4.2. A continuous-time stochastic process $\{X(t) : t \geqslant 0\}$ is called **Gaussian process**, if for any given times $0 \leqslant t_1 < \cdots < t_n < \infty$, the vector

$$\big(X(t_1), \cdots, X(t_n)\big)$$

is a Gaussian random vector.

Recall that to describe a Gaussian random vector, one needs only give the means and the covariances. Hence the finite-dimensional distributions of a Gaussian process are determined by the numbers

$$m_t = \mathbb{E}X(t) \quad \text{and} \quad \Gamma_{s,t} = \mathrm{Cov}(X(s), X(t))\,.$$

Standard Brownian motion as a Gaussian process Suppose $\{B(t)\}$ is a standard Brownian motion and $0 \leqslant t_1 < t_2 < \cdots < t_n$, then we can write $B(t_1), \cdots, B(t_n)$ as linear combinations of the independent standard normal random variables

$$\frac{B(t_j) - B(t_{j-1})}{\sqrt{t_j - t_{j-1}}}, \quad j = 1, \ldots, n\,.$$

where $t_0 = 0$. Hence $\{B(t)\}$ is a Gaussian process with mean zero. For $s < t$,

$$\begin{aligned}
\mathbb{E}\left(B(s)B(t)\right) &= \mathbb{E}\left[B(s)\big(B(s) + B(t) - B(s)\big)\right] \\
&= \mathbb{E}B(s)^2 + \mathbb{E}\left[B(s)\left(B(t) - B(s)\right)\right] \\
&= s + \mathbb{E}\left[B(s)\right]\mathbb{E}\left[B(t) - B(s)\right] = s\,,
\end{aligned}$$

which gives the general rule

$$\mathrm{Cov}\,(B(s), B(t)) = s \wedge t\,.$$

The description of Brownian motion as a Gaussian process describes only the finite-dimensional distributions, however, as we have pointed, one cannot tell from the finite-dimensional distributions alone whether or not the paths are continuous. Therefore, we have the following equivalent definition of Brownian motion.

Theorem 4.2. $\{B(t)\}$ *is a standard Brownian motion if and only if*

(i) $\{B(t)\}$ *is a Gaussian process.*

(ii) $\mathbb{E}B(t) = 0$ *and* $\mathrm{Cov}(B(s), B(t)) = s \wedge t$ *for all* t*,* s*.*

(iii) Almost surely, the function $t \mapsto B(t)$ *is continuous.*

4.2.2 Lévy processes

$\{X(t) : t \geqslant 0\}$ is a continuous-time processes. We say that it has *independent increments* if for each $n \in \mathbb{N}$ and each $0 \leqslant t_1 < t_2 < \cdots < t_n < \infty$ the random variables

$$X(t_2) - X(t_1) \cdots , X(t_n) - X(t_{n-1})$$

are independent, and that it has *stationary increments* if for any t, $s \geqslant 0$,

$$X\,(t+s) - X\,(s) \overset{\mathrm{d}}{=} X\,(t) - X(0)$$

Definition 4.3. $\{X(t)\}$ is called a **Lévy process** if

(i) $X(0) = 0$ a.s.

(ii) $\{X(t)\}$ has independent, stationary increments.

(iii) $\{X(t)\}$ is *stochastically continuous* (or *continuous in probability*), i.e. for all $\epsilon > 0$ and for all $t \geqslant 0$

$$\lim_{h\to 0} \mathbb{P}\left(|X(t+h) - X(t)| > \epsilon\right) = 0$$

Note that in the presence of (i) and (ii), (iii) is equivalent to the condition

$$\lim_{h\downarrow 0} \mathbb{P}(|X(h)| > \epsilon) = 0\,.$$

We have already discussed an example of a Lévy process with right-continuous paths, the Poisson process.

Brownian motion as a Lévy processs As we can see, Brownian motion with drift μ and variance σ^2 starting at origin, is the Lévy process with continuous paths : (i),(ii) are clear, and

$$\mathbb{P}(|B(h)| > \epsilon) = \mathbb{P}\left(|N(\mu,\sigma^2)| > \frac{\epsilon}{\sqrt{h}}\right) \to 0\,,\ \text{as } h \downarrow 0\,.$$

Suprisingly, the converse is ture.

Theorem 4.3. *Suppose $\{X(t)\}$ is a Lévy process with continuous paths. Then there exists μ and σ^2 such that $\{X(t)\}$ is a Brownian motion with drifr μ and variance σ^2, starting in 0.*

Proof. All we need to show is that $X(1)$ has a normal distribution. Let

$$X_{n,j} = X\big(\frac{j}{n}\big) - X\big(\frac{j-1}{n}\big)\,,\ \text{for } j = 1,2,\cdots$$

then

$$X(1) = X_{n,1} + \cdots + X_{n,n}\,.$$

We shall use Lindeberg-Feller's theorem to show the desired result.

Step 1. We claim that all the moments of $X(t)$ are finite. To see this, let

$$M = \max_{0\leqslant t\leqslant 1} |X(t)|\ ,$$

since $t \mapsto X(t)$ is continuous, we have $M < \infty$ a.s. , so there exists some positive integer k such that

$$\mathbb{P}(M \geqslant k) \leqslant \frac{1}{2}\,.$$

Then using continuity of the paths, by stopping at the first time t that $|X_t| = nk$, we can see that

$$\mathbb{P}(M \geqslant (n+1)k \mid M \geqslant nk) \leqslant \frac{1}{2}\,.$$

and hence

$$\mathbb{P}(M \geqslant nk) \leqslant \frac{1}{2^n}\,.$$

Thus all the moments of M are finite. Of course all the moments of $X(t)$ are finite for a given t.

Step 2. Let $\mathbb{E}X_1 = \mu$, and $\mathrm{Var}\,(X_1) = \sigma^2$. Since the increments are independent and stationary,

$$\mathbb{E}X_{n,j} = \frac{\mu}{n}\,,\ \mathrm{Var}\,(X_{n,j}) = \frac{\sigma^2}{n}\,,\ \text{for all } j = 1, \cdots, n$$

Let $\xi_{n,j} = X_{n,j} - \mathbb{E}\,X_{n,j}$, then $\mathbb{E}\xi_{n,j} = 0$ and

$$\sum_{j=1}^{n} \mathbb{E}\,\xi_{n,j}^2 = \sigma^2$$

It suffices to check the Lindeberg's condition : for any $\epsilon > 0$,

$$\sum_{j=1}^{n} \mathbb{E}\,\xi_{n,j}^2 1_{\{|\xi_{n,j}|>\epsilon\}} \to 0\,.$$

Define $N_n = \sum_{j=1}^{n} 1_{\{|\xi_{n,j}|>\epsilon\}}$, note that $|\xi_{n,j}| \leqslant |X_{n,j}| + |\mathbb{E}X_{n,j}|$ we have

$$\sum_{j=1}^{n} \xi_{n,j}^2 1_{\{|\xi_{n,j}|>\epsilon\}} \leqslant \big(2M + \frac{|\mu|}{n}\big)^2 N_n$$

By C-B-S inequality, Lindeberg's condition follows from

$$\mathbb{E}N_n^2 \to 0\,. \tag{4.1}$$

Step 3. To show (4.1), put

$$M_n = \max_{1 \leqslant j \leqslant n} |X_{n,j}| \ .$$

Continuity of the paths implies that $M_n \to 0$ a.s., and hence for every $\epsilon > 0$,

$$\mathbb{P}\left(M_n \leqslant \epsilon\right) \to 1\, .$$

Since the increments are independent and stationary, we have

$$\mathbb{P}\left(M_n \leqslant \epsilon\right) = \left[1 - \mathbb{P}\left(|X_{n,1}| > \epsilon\right)\right]^n \leqslant \mathrm{e}^{-n\mathbb{P}(|X_{n,1}|>\epsilon)}$$

Therefore, for every $\epsilon > 0$, $n\,\mathbb{P}\left(|X_{n,1}| > \epsilon\right) \to 0$, which deduce that

$$n\,\mathbb{P}\left(|\xi_{n,1}| > \epsilon\right) \to 0\, .$$

Then $\mathbb{E}N_n = n\mathbb{P}\left(|\xi_{n,1}| \geqslant \epsilon\right) \to 0$ and

$$\mathrm{Var}(N_n) = \sum_{j=1}^{n} \mathrm{Var}(1_{\{|\xi_{n,j}|>\epsilon\}}) \leqslant \sum_{j=1}^{n} \mathbb{E}1_{\{|\xi_{n,j}|>\epsilon\}} = \mathbb{E}N_n \to 0,$$

and (4.1) follows. □

4.3 Simple invariance properties

The simplest invariance property of Brownian motion is that, standard Brownian motion is symmetric.

Proposition 4.4 (Symmetry). $\{B(t) : t \geqslant 0\}$ *is a standard Brownian motion, then* $\{-B(t) : t \geqslant 0\}$ *is also a standard Brownian motion.*

Sometimes, we need the following invariance property to reverse the increments.

Proposition 4.5 (Increments reversal). $\{B(t) : t \geqslant 0\}$ *is a standard Brownian motion, then for any given* $t > 0$,

$$\{B(t-s) - B(t) : 0 \leqslant s \leqslant t\} \tag{4.2}$$

is also a standard Brownian motion.

Many natural sets that can be derived from the sample paths of Brownian motion are in some sense *random fractals*. An intuitive approach to fractals is that they are sets which have an interesting geometric structure at all scales. A key role in this behaviour is played by the very simple *scaling invariance* property of Brownian motion, which we now formulate. It identifies a transformation on the space of functions, which changes the individual Brownian random functions but leaves their distribution unchanged.

Theorem 4.6 (Scaling invariance). $\{B(t) : t \geqslant 0\}$ *is a standard Brownian motion and let* $a > 0$. *Then the process* $\{W(t) : t \geqslant 0\}$ *defined by*

$$W(t) = \frac{1}{a} B\left(a^2 t\right), \text{ for } t \geqslant 0 \tag{4.3}$$

is also a standard Brownian motion.

Proof. Continuity of the paths, independence and stationarity of the increments remain unchanged under the scaling. It easy to observe that

$$X(t) - X(s) = \frac{1}{a}\left(B\left(a^2 t\right) - B\left(a^2 s\right)\right)$$

is normally distributed with expectation 0 and variance $t - s$. □

Scaling invariance has many useful consequences. As an example, let $a < 0 < b$, and look at

$$\tau_{(a,b)} = \inf \{t \geqslant 0 : B(t) \in \{-a, b\}\},$$

the first exit time of a standard Brownian motion from the interval $[a, b]$. Then, with

$$W(t) = \frac{1}{b} B\left(b^2 t\right)$$

we have

$$\begin{aligned} \mathbb{E}\,\tau_{(a,b)} &= b^2 \mathbb{E} \inf\{t \geqslant 0 : W(t) = \frac{a}{b} \text{ or } W(t) = 1\} \\ &= b^2 \mathbb{E}\,\tau_{(\frac{a}{b},1)} \end{aligned}$$

which implies that $\mathbb{E}\,\tau_{(-b,b)}$ is a constant multiple of b^2. Also

$$\mathbb{P}\left(B(t) \text{ exits } [a, b] \text{ at } a\right) = \mathbb{P}\Big(X(t) \text{ exits } [\frac{a}{b}, 1] \text{ at } 1\Big)$$

is only a function of the ratio $\frac{a}{b}$. The scaling invariance property will be used extensively in all the following sections, and we shall often use the phrase that a fact holds 'by Brownian scaling' to indicate this.

A further useful invariance property of Brownian motion, *invariance under time inversion*, can be identified easily. As above, the transformation on the space of functions changes the individual Brownian random functions without changing the distribution.

Theorem 4.7 (Time inversion). *Suppose $\{B(t) : t \geqslant 0\}$ is a standard Brownian motion. Then the process $\{X(t) : t \geqslant 0\}$ defined by*

$$X(t) = \begin{cases} 0 & \text{for } t = 0, \\ tB\left(\frac{1}{t}\right) & \text{for } t > 0, \end{cases}$$

is also a standard Brownian motion.

Proof. Obviously, $\{X(t) : t \geqslant 0\}$ is also a Gaussian process and the Gaussian random vectors $(X(t_1), \ldots, X(t_n))$ have expectation zero. The covariances, for $t > 0, h \geqslant 0$, are given by

$$\begin{aligned} \operatorname{Cov}(X(t+h), X(t)) &= (t+h)\, t \ \operatorname{Cov}[B(\frac{1}{t+h}), B(\frac{1}{t})] \\ &= t(t+h)\frac{1}{t+h} = t\,. \end{aligned}$$

Hence the law of all the finite-dimensional distributions

$$(X(t_1), X(t_2), \cdots, X(t_n)), \quad \text{for } 0 \leqslant t_1 \leqslant \cdots \leqslant t_n\,,$$

are the same as for Brownian motion. The paths of $t \mapsto X(t)$ are clearly continuous for all $t > 0$ and in $t = 0$ we use the following two facts:

First, as the set $\mathbb{Q}$ of rationals is countable, the distribution of $\{X(t) : t > 0, t \in \mathbb{Q}\}$ is the same as for a Brownian motion, and hence

$$\lim_{\substack{t \downarrow 0 \\ t \in \mathbb{Q}}} X(t) = 0 \ \text{ almost surely }.$$

Second, $\mathbb{Q} \cap (0, \infty)$ is dense in $(0, \infty)$ and $\{X(t) : t \geqslant 0\}$ is almost surely continuous on $(0, \infty)$, so that

$$\lim_{t \downarrow 0} X(t) = 0 \ \text{ almost surely }.$$

Hence $\{X(t) : t \geqslant 0\}$ has almost surely continuous paths, and is a Brownian motion. $\square$

Remark. The symmetry inherent in the time inversion property becomes more apparent if one considers the Ornstein-Uhlenbeck diffusion $\{X(t) : t \in \mathbb{R}\}$, which is given by

$$X(t) = e^{-t} B\left(e^{2t}\right) \ \text{ for all } t \in \mathbb{R}$$

This is a Markov process (this will be explained properly in Section 4.6), such that $X(t)$ is standard normally distributed for all t. It is a diffusion with a drift towards the origin proportional to the distance from the origin. Unlike Brownian motion, the Ornstein-Uhlenbeck diffusion is time reversible: The time inversion formula gives that $\{X(t) : t \geqslant 0\}$ and $\{X(-t) : t \geqslant 0\}$ have the same law. For t near $-\infty$, $X(t)$ relates to the Brownian motion near time 0, and for t near ∞, $X(t)$ relates to the Brownian motion near ∞.

Time inversion is a useful tool to relate the properties of Brownian motion in a neighbourhood of time $t = 0$ to properties at infinity. To illustrate the use of time inversion we exploit Theorem 1.9 to get an interesting statement about the long-term behaviour from an easy statement at the origin.

Corollary 4.8 (Law of large numbers). *Almost surely,*

$$\lim_{t\to\infty} \frac{B(t)}{t} = 0\,.$$

Proof. Let $\{X(t) : t \geqslant 0\}$ be the time inversion of $\{B(t)\}$. Thus, it's easy to see that

$$\lim_{t\to\infty} \frac{B(t)}{t} = \lim_{t\to\infty} X\big(\frac{1}{t}\big) = 0 \quad \text{a.s.}$$

□

4.4 Nondifferentiability of Brownian motion

One manifestation is that the paths of Brownian motion have no intervals of monotonicity.

Theorem 4.9. *Almost surely, Brownian motion $\{B(t)\}$ doesn't have a monotone interval.*

Proof. Given an interval $[a, b] \subset [0, \infty]$, if it is an interval of monotonicity, we pick numbers

$$a = a_0 < a_1 < \cdots < a_n = b$$

and divide $[a, b]$ into n sub-intervals $[a_i, a_{i+1}]$. Each increment

$$B(a_i) - B(a_{i-1}) \ , i = 1 \cdots , n$$

has to have the same sign. As the increments are independent, this has probability

$$2 \cdot \frac{1}{2^n}$$

and taking $n \to \infty$ shows that the probability that $[a, b]$ is an interval of monotonicity must be zero.

Taking a countable union gives that, almost surely, there is no nondegenerate interval of monotonicity with rational endpoints, but each nondegenerate interval would have a nondegenerate rational sub-interval. $\square$

Proposition 4.10. *Almost surely,*

$$\limsup_{n\to\infty} \frac{B(n)}{\sqrt{n}} = +\infty \, , \quad \textit{and} \quad \liminf_{n\to\infty} \frac{B(n)}{\sqrt{n}} = -\infty$$

Proof. We clearly have, by Fatou's lemma,

$$\mathbb{P}\left(B(n) > c\sqrt{n} \text{ i.o.}\right) \geqslant \limsup_{n\to\infty} \mathbb{P}\left(B(n) > c\sqrt{n}\right)$$

By the scaling property, the expression in the limsup equals $\mathbb{P}(B(1) > c)$, which is positive. Denote $X_n = B(n) - B(n-1)$ for all n then

$$\left\{ \frac{1}{\sqrt{n}} B(n) > c \quad \text{i.o.} \right\} = \left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j > c \quad \text{i.o.} \right\}$$

is an tail event with respect to $\{X_n\}$. Hence the Kolmogorov 0-1 law gives that,

$$\mathbb{P}\left(B(n) > c\sqrt{n} \quad \text{i.o.}\right) = 1\,.$$

Taking the intersection over all positive integers c gives the first part of the statement and the second part is proved analogously. □

For a function f, we define the upper and lower right derivatives

$$D^* f(t) = \limsup_{h \downarrow 0} \frac{f(t+h) - f(t)}{h}$$

and

$$D_* f(t) = \liminf_{h \downarrow 0} \frac{f(t+h) - f(t)}{h}$$

We now show that for any fixed time t, almost surely, Brownian motion is not differentiable at t. For this we use Proposition4.10 and the invariance under time inversion.

Proposition 4.11. *Fix $t \geqslant 0$, then, almost surely, Brownian motion is not differentiable at t. Moreover,*

$$D^* B(t) = +\infty \quad \text{and} \quad D_* B(t) = -\infty$$

Proof. Without loss of generality, we assume $t = 0$. Or we can consider $\{\widetilde{B}(s)\}$, defined by $\widetilde{B}(s) = B(t+s) - B(t)$.

For a standard Brownian motion $\{B(t)\}$, we construct a further Brownian motion $\{X(t)\}$ by time inversion. Then

$$D^* X(0) \geqslant \limsup_{n \to \infty} \frac{X\left(\frac{1}{n}\right) - X(0)}{\frac{1}{n}} = \limsup_{n \to \infty} nX\left(\frac{1}{n}\right) = \limsup_{n \to \infty} B(n)\,.$$

By Proposition4.10, the we find that $D^*X(0)$ is infinite. Similarly, one can see that $D_*X(0) = -\infty$, showing that X is not differentiable at 0. $\square$

While the previous proof shows that every t is almost surely a point of nondifferentiability for the Brownian motion, this does not imply that almost surely every t is a point of non-differentiability for the Brownian motion!

Remark. The behaviour of Brownian motion at a fixed time $t > 0$ reflects the behaviour at typical times in the following sense.

Suppose $\mathscr{P} = \{f \in \mathbf{C}[0,\infty) : f \text{ is differentable at } 0\}$. Then we have shown that for any $t \geqslant 0$,

$$\{B(t+s) - B(t) : s \geqslant 0\} \in \mathscr{P} \quad \text{a.s.} \tag{4.4}$$

In fact, denote $D = \{t : \{B(t+s) - B(t) : s \geqslant 0\} \notin \mathscr{P}\}$, then almost surely, D has Lebesgue measure zero. To see this, using the joint measurability mentioned in the remark of Theorem 4.1 and Fubini's theorem,

$$\mathbb{E}\int_0^\infty 1_D \,\mathrm{d}t = \int_0^\infty \mathbb{P}\left(\{B(t+s) - B(t) : s \geqslant 0\} \notin \mathscr{P}\right) \mathrm{d}t = 0\,.$$

Thus

$$\lambda(D) = 0 \qquad \text{a.s.}$$

where λ is the Lebesgue measure.

Theorem 4.12 (Paley, Wiener and Zygmund). *Almost surely, Brownian motion is nowhere differentiable. Moreover, almost surely,*

$$\textit{either}\ \ D^*B(t) = +\infty \ \textit{or}\ \ D_*B(t) = -\infty \ \textit{or both, for each}\ t \geqslant 0\,.$$

Proof. Suppose that there is a $t_0 \in [0,1]$ such that $-\infty < D_*B(t_0) \leqslant D^*B(t_0) < \infty$, then

$$\limsup_{h\downarrow 0} \frac{|B(t_0+h) - B(t_0)|}{h} < \infty\,.$$

Using the boundedness of Brownian motion on $[0,2]$, this implies that for some finite constant M there exists t_0 with

$$\sup_{h\in[0,1]} \frac{|B(t_0+h)-B(t_0)|}{h} \leqslant M\,.$$

It suffices to show that this event has probability zero for any M. From now on fix M. If t_0 is contained in the binary interval $\left[\frac{k-1}{2^n},\frac{k}{2^n}\right]$ for some $n>2$, and $k\geqslant 1$, then for all $1\leqslant j\leqslant 2^n-k$ the triangle inequality gives

$$\begin{aligned}
&\left|B\left(\frac{k+j}{2^n}\right)-B\left(\frac{k+j-1}{2^n}\right)\right| \\
&\leqslant \left|B\left(\frac{k+j}{2^n}\right)-B(t_0)\right| + \left|B(t_0)-B\left(\frac{k+j-1}{2^n}\right)\right| \leqslant \frac{(2j+1)M}{2^n}\,.
\end{aligned}$$

Define events

$$\Omega_{n,k} := \left\{ \left|B\left(\frac{k+j}{2^n}\right)-B\left(\frac{k+j-1}{2^n}\right)\right| \leqslant \frac{2j+1}{2^n}M \quad \text{for } j=1,2,3 \right\},$$

then by independence of the increments and the scaling property, for $1\leqslant k\leqslant 2^n-3$

$$\begin{aligned}
\mathbb{P}(\Omega_{n,k}) &\leqslant \prod_{j=1}^{3} \mathbb{P}\left(\left|B\left(\frac{k+j}{2^n}\right)-B\left(\frac{k+j-1}{2^n}\right)\right| \leqslant \frac{2j+1}{2^n}M\right) \\
&\leqslant \mathbb{P}\left(|B(1)| \leqslant \frac{7M}{\sqrt{2^n}}\right)^3 \leqslant \left(\frac{7M}{\sqrt{2^n}}\right)^3\,.
\end{aligned}$$

Hence

$$\mathbb{P}\left(\bigcup_{k=1}^{2^n-3}\Omega_{n,k}\right) \leqslant 2^n\left(\frac{7M}{\sqrt{2^n}}\right)^3 = \frac{(7M)^3}{\sqrt{2^n}}\,,$$

which is summable over all n. Hence, by the B-C lemma,

$$\begin{aligned}
&\mathbb{P}\left(\text{there is } t_0\in[0,1] \text{ with } \sup_{h\in[0,1]} \frac{|B(t_0+h)-B(t_0)|}{h} \leqslant M\right) \\
&\leqslant \mathbb{P}\left(\bigcup_{k=1}^{2^n-3}\Omega_{n,k} \text{ for infinitely many } n\right) = 0\,.
\end{aligned}$$

□

待补充: 二次变差

4.5 Continuity properties of Brownian motion

The definition of Brownian motion already requires that the sample functions are continuous almost surely. This implies that on the interval $[0,1]$ (or any other compact interval) the sample functions are uniformly continuous, i.e. there exists some (random) function φ with $\lim_{h\downarrow 0}\varphi(h)=0$ called a *modulus of continuity* of the function $B:[0,1]\to\mathbb{R}$ such that

$$\limsup_{h\downarrow 0}\sup_{t\in[0,1-h]}\frac{|B(t+h)-B(t)|}{\varphi(h)}\leqslant 1\,.$$

Can we achieve such a bound with a deterministic function φ, i.e. is there a nonrandom modulus of continuity for the Brownian motion ? The answer is yes, as the following theorem shows.

Theorem 4.13. *There exists a constant $C>0$ such that, almost surely, for every sufficiently small $h>0$,*

$$|B(t+h)-B(t)|\leqslant C\sqrt{h\log\frac{1}{h}}\quad\textit{for all}\;\; t\in[0,1-h]\,. \tag{4.5}$$

First proof. This follows quite elegantly from Lévy's construction of Brownian motion. Recall the notation introduced there and that we have represented Brownian motion as a series

$$B(t)=\sum_{n=0}^{\infty}F_n(t)$$

where each F_n is a piecewise linear function. We have shown that for any $c>\sqrt{2\log 2}$ there exists a (random) $N\in\mathbb{N}$, such that,

$$\|F_n\|_\infty<c\sqrt{n}2^{-\frac{n}{2}}\,,\quad\text{for all } n\geqslant N\,.$$

Now for each $t,t+h\in[0,1]$,

$$|B(t+h)-B(t)|\leqslant\sum_{n=0}^{\infty}|F_n(t+h)-F_n(t)|\;,$$

Suppose that h is (again random and) small enough, then the positive integer ℓ, defined by $h \in (\frac{1}{2^\ell}, \frac{1}{2^{\ell-1}}]$, exceeds N. Hence

$$\sum_{n=\ell+1}^{\infty} |F_n(t+h) - F_n(t)| \leqslant 2 \sum_{n=\ell+1}^{\infty} \|F_n\|_\infty \leqslant 2c \sum_{n=\ell+1}^{\infty} \sqrt{n} 2^{-\frac{n}{2}}$$

and there exists constants C_1, C_2 such that

$$\sum_{n=\ell+1}^{\infty} \sqrt{n} 2^{-\frac{n}{2}} \leqslant C_1 \sqrt{\ell}\, 2^{-\frac{\ell}{2}} \leqslant C_2 \sqrt{h \log \frac{1}{h}}\,.$$

Hence, using the mean-value theorem, we get for all $\ell > N$ that $|B(t+h) - B(t)|$ is bounded by

$$2 \sum_{n=0}^{\ell} \left\|F_n'\right\|_\infty h + C_2 \sqrt{h \log \frac{1}{h}}.$$

We now suppose that h is (again random and) small enough that the first summand is smaller than $\sqrt{h \log \frac{1}{h}}$, and hence we get (4.5). □

Remark. 注意这里我们对 Levy 构造的布朗运动证明了

For any $h > 0$, let

$$\operatorname{Osc}(h) \coloneqq \{|B(t) - B(s)| : t, s \in [0,1]\,, |t-s| \leqslant h\}\,.$$

It's easy to see that Theorem 4.5 hols if and only if there exists a constant $C > 0$ and almost surely,

$$\lim_{h \to 0} \frac{\operatorname{Osc}(h)}{\sqrt{h \log \frac{1}{h}}} \leqslant C\,. \tag{4.6}$$

Following this, we will give another proof by using Theorem 4.23.

Second proof. Firstly, take a positive integer n and let

$$I_{n,m} = \left[\frac{m-1}{2^n}, \frac{m}{2^n}\right]$$

for $m = 1, 2, \cdots, 2^n$ and denote

$$\Delta_{n,m} = \sup_{t \in I_{n,m}} \left| B(t) - B(\frac{m-1}{2^n}) \right| .$$

By Brownian scaling, for a standard Brownian motion $\{W(t)\}$, $\Delta_{n,m}$ is indetically distributed to

$$\frac{M \vee \widetilde{M}}{\sqrt{2^n}}$$

where

$$M := \max_{0 \leqslant t \leqslant 1} W(t) \text{ and } \widetilde{M} := \max_{0 \leqslant t \leqslant 1} -W(t) .$$

Thus for given $x > 0$,

$$\mathbb{P}\left(\Delta_{n,m} > \frac{2x}{\sqrt{2^n}}\right) = \mathbb{P}\left(M \vee \widetilde{M} > 2x\right)$$

and note that

$$\{M \vee \widetilde{M} > 2x\} \subset \{M > x\} \cup \{\widetilde{M} > x\} ,$$

we have

$$\begin{aligned} \mathbb{P}\left(\Delta_{n,m} > \frac{2x}{\sqrt{2^n}}\right) &\leqslant 2\mathbb{P}\left(M > x\right) = 2\ \mathbb{P}\left(|B(1)| > x\right) \\ &= 4 \int_x^\infty (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t \leqslant \exp\left(-\frac{x^2}{2}\right) . \end{aligned}$$

So

$$\mathbb{P}\left(\text{ exists } m \text{ s.t. } \Delta_{n,m} > \frac{2x}{\sqrt{2^n}}\right) \leqslant 2^n \exp\left(-\frac{x^2}{2}\right) = \exp\left(n \log 2 - \frac{x^2}{2}\right) .$$

In order that

$$\sum_{n=1}^\infty \mathbb{P}\left(\text{ exists } m \text{ s.t. } \Delta_{n,m} > \frac{2x_n}{\sqrt{2^n}}\right) < \infty ,$$

we let $x_n = c\sqrt{n}$ where $c > \sqrt{2 \log 2}$, by B-C lemma we have

$$\mathbb{P}\left(\text{ exists } m \text{ s.t. } \Delta_{n,m} > 2c\sqrt{\frac{n}{2^n}} \ \text{ i.o. } \right) = 0 .$$

That is,

$$\mathbb{P}\left(\liminf_{n\to\infty}\left\{\sup_{1\leqslant m\leqslant 2^n}\Delta_{n,m}\leqslant 2c\sqrt{\frac{n}{2^n}}\right\}\right)=1\,. \tag{4.7}$$

One can show that (4.6) follows from .(4.7). □

Remark. One can determine the constant c in the best possible modulus of continuity $\varphi(h)=c\sqrt{h\log(1/h)}$ precisely. Indeed, our proof of the lower bound yields a value of $c=\sqrt{2}$ which turns out to be optimal. This striking result is due to Paul Lévy.

Theorem (Lévy's modulus of continuity). *Almost surely,*

$$\limsup_{h\perp 0}\sup_{0\leqslant t\leqslant 1-h}\frac{|B(t+h)-B(t)|}{\sqrt{2h\log(1/h)}}=1$$

Recall that a function $f:[0,\infty)\to\mathbb{R}$ is said to be **locally α-Hölder continuous**, if there exists $\delta>0$ and $C>0$ such that

$$|f(x)-f(y)|\leqslant c|x-y|^{\alpha},\qquad \text{for all } x,y\geqslant 0 \text{ with } |y-x|<\delta\,.$$

We refer to $\alpha>0$ as the *Hölder exponent* and to $C>0$ as the *Hölder constant.*

Clearly, α-Holder continuity gets stronger, as the exponent α gets larger. The results of this chapter so far indicate that, for Brownian motion, the transition between paths which are α-Hölder continuous and paths which are not happens at $\alpha=\frac{1}{2}$.

Corollary 4.14. *For any $\alpha<\frac{1}{2}$, then, almost surely, Brownian motion is locally α-Hölder continuous.*

4.6 Markov property, strong Markov property

In this section we discuss the Markov property and strong Markov property of Brownian motion. For the discussion of the Markov property we include higher dimensional Brownian motion, which can be defined easily by requiring the characteristics of a linear Brownian motion in every component, and independence of the components.

Definition 4.4. If $B_1, \cdots, B_d$ are independent linear Brownian motions started in $x_1, \cdots, x_d$, then the stochastic process $\{B(t) : t \geqslant 0\}$ given by

$$B(t) = (B_1(t), \cdots, B_d(t))^{\mathrm{T}}$$

is called a d**-dimensional Brownian motion** started in $(x_1, \cdots, x_d)^{\mathrm{T}}$. The d-dimensional Brownian motion started in the origin is also called d**-dimensional standard Brownian motion**.

4.6.1 Markov peocesses

Suppose now that $\{X(t) : t \geqslant 0\}$ is a stochastic process. Intuitively, the *Markov property* says that if we know the process $\{X(t) : t \geqslant 0\}$ on the interval $[0, s]$, for the prediction of the future $\{X(t) : t \geqslant s\}$ this is as useful as just knowing the endpoint $X(s)$. Moreover, a process is called a *time-homogeneous* Markov process if it starts afresh at any fixed time s.

Slightly more precisely this means that, supposing the process can be started in any point $X(0) = x \in \mathbb{R}^d$, the time-shifted process $\{X(s+t) : t \geqslant 0\}$ has the same distribution as the process started in $X(s) \in \mathbb{R}^d$.

A function $\mathfrak{p} : [0, \infty) \times \mathbb{R}^d \times \mathcal{B}^d \to \mathbb{R}$, is **a Markov transition kernel** provided

(i) for each $A \in \mathcal{B}^d$, $\mathfrak{p}(\cdot, \cdot, A)$ is a measurable function on $[0, \infty) \times \mathbb{R}^d$;

(ii) for all $t > 0$ and $x \in \mathbb{R}^d$, $\mathfrak{p}(t, x, \cdot)$ is a probability measure on $\mathcal{B}^d$; [1]

[1] When integrating a function f with respect to this measure, we write $\int f(y)\mathfrak{p}(t, x, dy)$.

(iii) for all $A \in \mathcal{B}^d$, $x \in \mathbb{R}^d$ and $t, s > 0$,

$$\mathfrak{p}(t+s, x, A) = \int_{\mathbb{R}^d} \mathfrak{p}(t, y, A)\mathfrak{p}(s, x, dy)\,, \tag{4.8}$$

This is the so-called *Chanpmen-Kolmogorov Equation.*

Definition. An adapted process $\{X(t) : t \geqslant 0\}$ is called a (**time-homogeneous**) **Markov process** with transition kernel P with respect to $\{\mathcal{F}(t)\}$ if, almost surely,

$$\mathbf{P}[\, X(t) \in A \mid \mathcal{F}(s)\,] = \mathfrak{p}(t-s, X(s), A)$$

for all $A \in \mathcal{B}^d$ and $t \geqslant s \geqslant 0$.

Observe that $\mathfrak{p}(t, x, A)$ is the probability that the process takes a value in A at time t, if it is started at the point x. Recalling the Markov chains, we recognise the pattern behind this definition: the Markov transition kernel plays the role of the transition matrix in this setup. The next two examples are easy consequences of the Markov property for Brownian motion.

¶ Example 4.2. d-dimensional Brownian motion is a Markov process and, the transition probability $\mathfrak{p}(t, x, \cdot)$ is a Gaussian distribution with mean x and covariance matrix $t\, I_d$.

Note that the C-K equation is just the fact that the sum of two independent Gaussian random vectors is a Gaussian random vector with the sum of the covariance matrices.

The transition kernel of d-dimensional Brownian motion is described by probability measures $\mathfrak{p}(t, x, \cdot)$ with densities denoted by $p(t, x, y)$ or $p_t(x, y)$, given by

$$p(t, x, y) = \phi_t(y - x) = (2\pi t)^{-\frac{d}{2}} \exp\left(-\frac{|x-y|^2}{2t}\right)\,, \text{ for } x, y \in \mathbb{R}^d\,.$$

¶ Example 4.3. The one-dimensional *refected Brownian motion* $\{X(t) : t \geqslant 0\}$ defined by

$$X(t) = |B(t)|\,, \quad \text{for all } t \geqslant 0\,,$$

is a Markov process. Moreover, its transition kernel $\mathfrak{p}(t, x, \cdot)$ is the law of $|Y|$ for Y normally distributed with mean x and variance t, which we call *the modulus normal distribution with parameters x and t.*

We giving a straight formulation of the facts for a Brownian motion.

Theorem 4.15 (Markov property). *Suppose that $\{B(t) : t \geqslant 0\}$ is a d-dimensional Brownian motion started in $x \in \mathbb{R}^d$. Let $s > 0$, then the process $\{B(t+s) - B(s) : t \geqslant 0\}$ is again a Brownian motion started in the origin and it is independent of the process $\{B(t) : 0 \leqslant t \leqslant s\}$.*

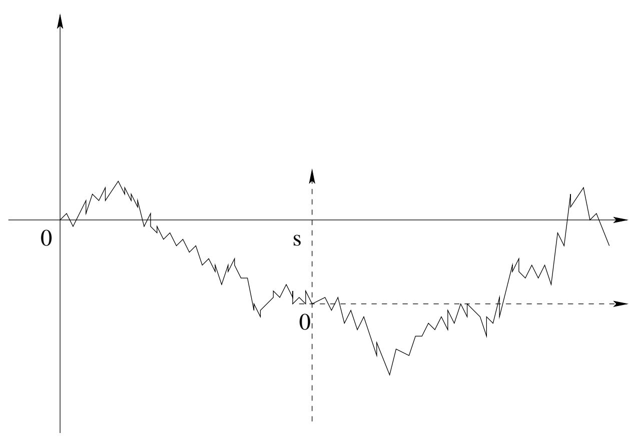


Figure 4.4: Brownian motion starts afresh at time s.

Proof. It is easy to check that $\{B(t+s) - B(s) : t \geqslant 0\}$ satisfies the definition of a d-dimensional Brownian motion. The independence statement follows directly from the independence of the increments of a Brownian motion. $\square$

4.6.2 More about Markov properties*

We intend to improve Markov property slightly, but first we introduce some useful terminology.

A **filtration** on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a family $(\mathcal{F}(t) : t \geqslant 0)$ of σ-algebras such that $\mathcal{F}(s) \subset \mathcal{F}(t) \subset \mathcal{F}$ for all $s < t$.

A probability space together with a filtration is called a **filtered probability space**.

A stochastic process $\{X(t) : t \geqslant 0\}$ defined on a filtered probability space with filltration $(\mathcal{F}(t) : t \geqslant 0)$ is called **adapted** if $X(t)$ is $\mathcal{F}(t)$-measurable for any $t \geqslant 0$.

Suppose we have a d-dimensional Brownian motion $\{B(t) : t \geqslant 0\}$ defined on some probability space, then we can define a filtration $(\mathcal{F}(t) : t \geqslant 0)$ by letting

$$\mathcal{F}(t) = \sigma(B(s) : 0 \leqslant s \leqslant t)$$

be the σ-algebra generated by the random variables $B(s)$, for $0 \leqslant s \leqslant t$. With this definition, $\{B(t)\}$ is obviously adapted to the filtration. Intuitively, this σ-algebra contains all the information available from observing the process up to time t.

By Markov property, the process $\{B(t+s) - B(s) : t \geqslant 0\}$ is independent of $\mathcal{F}(s)$. In a first step, we improve this and allow a slightly larger (augmented) σ -algebra $\mathcal{F}^+(s)$ defined by

$$\mathcal{F}^+(s) = \bigcap_{t>s} \mathcal{F}(t)$$

Clearly, the family $(\mathcal{F}^+(t) : t \geqslant 0)$ is again a filtration and $\mathcal{F}^+(s) \supset \mathcal{F}(s)$, but intuitively $\mathcal{F}^+(s)$ is a bit larger than $\mathcal{F}(s)$, allowing an additional infinitesimal glance into the future.

Theorem 4.16. *For every $s \geqslant 0$ the process $\{B(t+s) - B(s) : t \geqslant 0\}$ is independent of the σ-algebra $\mathcal{F}^+(s)$. In other words, conditional on $\mathcal{F}^+(s)$ the process $\{B(t+s) - B(s) : t \geqslant 0\}$ is a standard Brownian motion.*

Proof. Pick a strictly decreasing sequence $\{s_n : n \in \mathbb{N}\}$ converging to s. By continuity for any t

$$B(t+s) - B(s) = \lim_{n\to\infty} B\left(t + s_n\right) - B\left(s_n\right) .$$

By Markov property, for any $t_1, \cdots, t_m \geqslant 0$, the vector

$$\begin{aligned}&\left(B\left(t_1+s\right)-B(s), \cdots, B\left(t_m+s\right)-B(s)\right)\\&=\lim_{j \uparrow \infty}\left(B\left(t_1+s_j\right)-B\left(s_j\right), \cdots, B\left(t_m+s_j\right)-B\left(s_j\right)\right) .\end{aligned}$$

is independent of $\mathcal{F}^{+}(s)$, and so is the process $\{B(t+s)-B(s) : t \geqslant 0\}$. $\square$

We now look at the **germ σ-algebra** $\mathcal{F}^{+}(0)$, which heuristically comprises all events defined in terms of Brownian motion on an infinitesimal small interval to the right of the origin.

Theorem 4.17 (Blumenthal's 0-1 law). *Let $x \in \mathbb{R}^d$ and $A \in \mathcal{F}^{+}(0)$. Then*

$$\mathbb{P}_x(A) \in \{0,1\}. \tag{4.9}$$

Proof. Using Theorem 4.16 for $s=0$, we see that

$$\mathcal{F}(0)=\sigma(B(t) : t \geqslant 0)$$

is independent of $\mathcal{F}^{+}(0)$. This applies in particular to $A \in \mathcal{F}^{+}(0)$, which therefore is independent of itself, hence has probability zero or one. $\square$

As a first application we show that a standard linear Brownian motion has positive and negative values and zeros in every small interval to the right of 0. We have studied this remarkable property of Brownian motion already by different means.

Theorem 4.18. *Suppose $\{B(t) : t \geqslant 0\}$ is a linear Brownian motion. Define*

$$\tau=\inf\{t>0 : B(t)>0\} \quad \text{and} \quad \sigma_0=\inf\{t>0 : B(t)=0\},$$

then

$$\mathbb{P}_0\{\tau=0\}=\mathbb{P}_0\{\sigma_0=0\}=1 .$$

Proof. The event

$$\{\tau = 0\} = \bigcap_{n=1}^{\infty} \{ \text{ there is } r \in \mathbb{Q} \cap (0, 1/n) \text{ such that } B(r) > 0\}$$

is clearly in $\mathcal{F}^+(0)$. Hence we just have to show that this event has positive probability. This follows, as

$$\mathbb{P}_0\{\tau \leqslant t\} \geqslant \mathbb{P}_0\{B(t) > 0\} = 1/2$$

for $t > 0$. Hence

$$\mathbb{P}_0\{\tau = 0\} \geqslant 1/2$$

and we have shown the first part.

The same argument works replacing $B(t) > 0$ by $B(t) < 0$ and from these two facts $\mathbb{P}_0\{\sigma = 0\} = 1$ follows, using the intermediate value property of continuous functions. □

A further application is a $0-1$ law for the tail σ-algebra of a d-dimensional Brownian motion. Define $\mathcal{F}'(t) = \sigma(B(s) : s \geqslant t)$. Let

$$\mathcal{T} = \bigcap_{t \geqslant 0} \mathcal{F}'(t) \tag{4.10}$$

be the σ-algebra of all **tail events**.

Theorem 4.19 (Zero-one law for tail events). *Let $x \in \mathbb{R}^d$ and suppose $A \in \mathcal{T}$ is a tail event. Then*

$$\mathbb{P}_x(A) \in \{0, 1\}.$$

Proof. It suffices to look at the case $x = 0$. Under the time inversion of Brownian motion, the tail σ-algebra is mapped on the germ σ-algebra, which contains only sets of probability zero or one, by Blumenthal's $0 - 1$ law. □

4.6.3 The strong Markov property

Heuristically, the Markov property states that Brownian motion is started anew at each deterministic time instance. It is a crucial property of Brownian motion that this holds also for an important class of random times. These random times are called *stopping times*.

The basic idea is that a random time T is a stopping time if we can decide whether $\{T \leqslant t\}$ by just knowing the path of the stochastic process up to time t. Think of the situation that T is the first moment where some random event related to the process happens.

Definition 4.5. A random variable T with values in $[0, \infty]$, defined on a probability space with filtration $\{\mathcal{F}(t)\}$ is called a **stopping time with respect to** $\{\mathcal{F}(t)\}$, if

$$\{T \leqslant t\} \in \mathcal{F}(t), \qquad \text{for every } t \geqslant 0\,.$$

Theorem (Strong Markov property). $\{B(t)\}$ *is a d-dimensional Brownian motion, T is a stopping time with respect to $\{\mathcal{F}(t)\}$. Conditional on $\{T < \infty\}$, the process*

$$\{B(T+t) - B(T) : t \geqslant 0\} \tag{4.11}$$

is a d-dimensional standard Brownian motion independent of $\{B(t) : t \leqslant T\}$.

Remark. An alternative form of the strong Markov property is that, for any bounded measurable $f : \mathbf{C}\left([0,\infty), \mathbb{R}^d\right) \to \mathbb{R}$ and $x \in \mathbb{R}^d$, we have almost surely

$$\mathbb{E}_x[f\{B(T+t)\} \mid \mathcal{F}(T)] = \mathbb{E}_{B(T)}[f\{\tilde{B}(t)\}] \tag{4.12}$$

where the expectation on the right is with respect to a d-dimensional Brownian motion $\{\tilde{B}(t) : t \geqslant 0\}$ started in the fixed point $B(T)$.

4.7 Reflection principle and sample paths

In this section, we always let $\{B(t)\}$ be a one-dimensional standard Brownian motion, we will use the strong Markov property to show a very useful tools, the reflection principle. We then exploit these facts to get finer properties of Brownian sample paths.

Theorem 4.20 (Reflection principle). *Given $x \in \mathbb{R}$, Let τ_x be the first time to hit x, that is*

$$\tau_x = \inf\{t \geqslant 0 : B(t) = x\}\,. \tag{4.13}$$

Then the process $\left\{\widetilde{B}(t) : t \geqslant 0\right\}$, defined by

$$\widetilde{B}(t) = B(t)1_{\{t \leqslant \tau_x\}} + \big(2x - B(t)\big)1_{\{t > \tau_x\}}\,, \tag{4.14}$$

is also a standard Brownian motion.

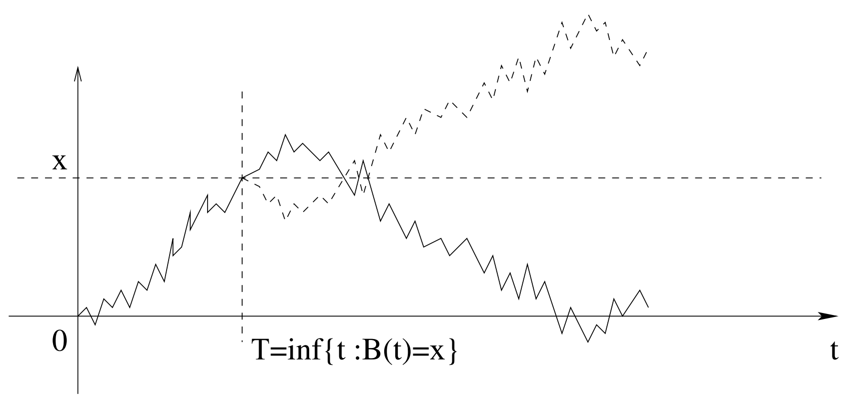


Figure 4.5: Reflection principle

Proof. Obviously, $\{\widetilde{B}(t)\}$ has continuous sample paths. It suffices to show the finite-dimensional distribution of $\{\widetilde{B}(t)\}$ consides with $\{B(t)\}$, Suppose $x \neq 0$, for any $0 = t_0 < t_1 < \cdots < t_n < \infty$ and $A_1, \cdots, A_n \in \mathcal{B}$,

$$\mathbb{P}(\widetilde{B}(t_j) \in A_j, 1 \leqslant j \leqslant n) = \mathbb{P}(B(t_j) \in A_j, 1 \leqslant j \leqslant n, \tau_x \geqslant t_n)$$
$$+ \sum_{m=1}^{n} \mathbb{P}(B(t_j) \in A_j, 1 \leqslant j < m;\ t_{m-1} \leqslant \tau_x < t_m;\ B(t_j) \in 2x - A_j, m \leqslant j \leqslant n)\,.$$

Note that, by strong Markov property we have

$$\begin{aligned}
&\mathbf{P}\big[B(t_j) \in A_j, j < m;\ t_{m-1} \leqslant \tau_x < t_m;\ B(t_j) \in 2x - A_j, j \geqslant m \mid \mathcal{F}(\tau_x)\big] \\
&= 1_{\{B(t_j) \in A_j, j<m;\, t_{m-1} \leqslant \tau_x < t_m\}} \int_{2x-A_m} \cdots \int_{2x-A_n} p_{t_m - \tau_x}(x, y_m) \cdots p_{t_n - t_{n-1}}(y_{n-1}, y_n) dy_m \cdots dy_n \\
&= 1_{\{B(t_j) \in A_j, j<m;\, t_{m-1} \leqslant \tau_x < t_m\}} \int_{A_m} \cdots \int_{A_n} p_{t_m - \tau_x}(x, y_m) \cdots p_{t_n - t_{n-1}}(y_{n-1}, y_n) dy_m \cdots dy_n \\
&= \mathbf{P}\big[B(t_j) \in A_j, j < m;\ t_{m-1} \leqslant \tau_x < t_m;\ B(t_j) \in A_j, j \geqslant m \mid \mathcal{F}(\tau_x)\big]\,,
\end{aligned}$$

Thus

$$\begin{aligned}
&\mathbb{P}(B(t_j) \in A_j, 1 \leqslant j < m;\ t_{m-1} \leqslant \tau_x < t_m;\ B(t_j) \in 2x - A_j, m \leqslant j \leqslant n) \\
&= \mathbb{P}(B(t_j) \in A_j, 1 \leqslant j < m;\ t_{m-1} \leqslant \tau_x < t_m;\ B(t_j) \in A_j, m \leqslant j \leqslant n)
\end{aligned}$$

and

$$\mathbb{P}(\widetilde{B}(t_j) \in A_j, 1 \leqslant j \leqslant n) = \mathbb{P}(B(t_j) \in A_j, 1 \leqslant j \leqslant n)\,. \qquad \square$$

4.7.1 The first hitting time

Theorem 4.21. *$\{B(t)\}$ is a standard Brownian motion. For any $x > 0$,*

$$\mathbb{P}(\tau_x \leqslant t) = 2\,\mathbb{P}(B(t) > x) = \mathbb{P}(|B(t)| > x)\,, \quad t > 0\,.$$

Proof. We have

$$\{\tau_x \leqslant t\} = \{\tau_x \leqslant t, B(t) \geqslant x\} \cup \{\tau_x \leqslant t, B(t) < x\}\ .$$

Note that

$$\{\tau_x \leqslant t, B(t) \geqslant x\} = \{B(t) \geqslant x\}\,,$$

and

$$\{\tau_x \leqslant t, B(t) < x\} = \{\tilde{\tau}_x \leqslant t, \widetilde{B}(t) > x\} = \{\widetilde{B}(t) > x\}\,,$$

where $\tilde{B}(t)$ is defined in (4.14), and $\tilde{\tau}_x$ is the first hitting time of $\{\tilde{B}(t)\}$, clearly $\tilde{\tau}_x = \tau_x$. By Reflection principle, we have

$$\mathbb{P}\left(\tau_x \leqslant t\right) = \mathbb{P}(B(t) \geqslant x) + \mathbb{P}(\widetilde{B}(t) > x) = \mathbb{P}(B(t) > x)\,. \qquad \square$$

Corollary 4.22. *For any $x > 0$, τ_x is a continuous random variable with density*

$$\rho(u) = \frac{x}{\sqrt{2\pi u^3}} e^{-\frac{x^2}{2u}}\, 1_{\{u>0\}} \tag{4.15}$$

Proof. Firstly,

$$\mathbb{P}(\tau_x < \infty) = \lim_{t\to\infty} \mathbb{P}(|B(1)| > x/\sqrt{t}\,) = 1\,.$$

On the other hand,

$$\begin{aligned}\mathbb{P}\left(\tau_x \leqslant t\right) = 2\mathbb{P}\left(B_t > x\right) &= 2\int_x^\infty \frac{1}{\sqrt{2\pi t}} e^{-\frac{v^2}{2t}}\, dv \\ &= \int_0^t \frac{x}{\sqrt{2\pi u^3}} e^{-\frac{x^2}{2u}}\, du, \quad \big(\text{let } u = x^2 t/v^2\big)\end{aligned}$$

Thus we complete the proof. $\square$

4.7.2 Maximum

$\{B(t)\}$ is a standard Brownian motion. Let

$$M(t) \coloneqq \sup_{0\leqslant s\leqslant t} B(s) = \max_{0\leqslant s\leqslant t} B(s)\,.$$

A priori it is not at all clear what the distribution of this random variable is, but we can determine it as a consequence of the reflection principle.

Theorem 4.23. *For eaxh $t > 0$, the distribution of $M(t)$ is given by*

$$\mathbb{P}(M(t) > x) = \mathbb{P}(|B(t)| > x)\,, \quad \textit{for all } x > 0\,.$$

Proof. Note that

$$\mathbb{P}(M(t) > x) = \mathbb{P}(\tau_x < t)\,,$$

so the desired result follows from Theorem 4.21. $\square$

Remark. For each $t > 0$, $\{M(t)\}$ is a continuous random variable, in fact, $M(t)$ has a modulus normal distribution with parameters 0 and t.

We now exploit the Markov property to study the local and global maxima of a standard Brownian motion.

Theorem 4.24. *$\{B(s) : 0 \leqslant s \leqslant t\}$ is a standard Brownian motion, then almost surely, the global maximum is attained at a unique time in $(0, t)$.*

Proof. Without loss of generality, let $t = 1$. For any given $r \in \mathbb{Q} \cap [0, 1]$, let

$$m_1 \coloneqq \max_{0 \leqslant t \leqslant r} B(t)\,,\ m_2 \coloneqq \max_{r \leqslant t \leqslant 1} B(t)\,.$$

Note that

$$m_2 = B(r) + \max_{0 \leqslant t \leqslant 1-r} (B(r+t) - B(r)) =: B(r) + \tilde{m}_2\,,$$

applying the Markov property at time r we see that $\tilde{m}_2$ is independent of $\mathcal{F}(r)$. Since

$$\tilde{m}_1 - B(r) = \max_{0 \leqslant t \leqslant r} B(t) - B(r) = \max_{0 \leqslant t \leqslant r} B(r-t) - B(r) =: \tilde{m}_1\,,$$

Thus

$$\{m_2 = m_1\} = \{\tilde{m}_2 = \tilde{m}_1\}\ ,$$

$\tilde{m}_2$, $\tilde{m}_1$ are two indenpendent continuous random variable, so hence this event has probability 0.

Hence, almost surely, for any rational number $q \in [0, 1]$ the maximum in $[0, q]$ and in $[q, 1]$ are different. Note that, if the global maximum is attained for two points $t_1 < t_2$ there exists a rational number $t_1 < q < t_2$ for which the maximum in $[0, q]$ and in $[q, 1]$ agree, which is a contradiction.

Now, it's suffices to show that $B(1) \neq M(1)$. Let

$$W(t) = B(1-t) - B(1)\,, \quad \text{for } t \in [0, 1]$$

then $W(t)$ is a standard Brownian motion, and

$$W(t) \leqslant W(1) \text{ for all } t \in (0,1) \Leftrightarrow B(t) \leqslant 0 \text{ for all } t \in (0,1),$$

But $M(1)$ is a continuous r.v., this event has probability 0. $\square$

Corollary 4.25. *$\{B(s) : 0 \leqslant s \leqslant t\}$ is a standard Brownian motion, then almost surely,*

(i) every local maximum is a strict local maximum.

(ii) the set of times where the local maxima are attained is countable and dense.

Proof. (i). By the statement just proved, almost surely, all nonoverlapping pairs of nondegenerate compact intervals with rational endpoints have different maxima. If Brownian motion however has a non-strict local maximum, there are two such intervals where Brownian motion has the same maximum.

(ii)) In particular, almost surely, the maximum over any nondegenerate compact interval with rational endpoints is not attained at an endpoint. Hence every such interval contains a local maximum, and the set of times where local maxima are attained is dense. As every local maximum is strict, this set has at most the cardinality of the collection of these intervals. $\square$

Corollary 4.26 (Large law of number). *$\{B(t)\}$ is a standard Brownian motion, then*

$$\frac{B(t)}{t} \to 0 \quad a.s.$$

Proof. By SLLN we have

$$\frac{B(n)}{n} \to 0 \quad \text{a.s.}$$

Then let

$$X_n = \sup_{n-1 \leqslant t \leqslant n} (B(t) - B(n)), \quad Y_n = \inf_{n-1 \leqslant t \leqslant n} (B(t) - B(n)),$$

Then by Markov property, $\{X_n\}$ is a i.i.d. sequence and $X_1 = M_1$, thus

$$\frac{X_n}{n} \to 0 \quad \text{and} \quad \frac{Y_n}{n} \to 0 \quad \text{a.s.}$$

Therefore

$$\frac{B(t)}{t} \to 0 \quad \text{a.s.} \qquad \square$$

We now recited a famous theorem of Paul Lévy, which shows that the difference of the maximum process of a Brownian motion and the Brownian motion itself is a reflected Brownian motion. See Figure 4.6.

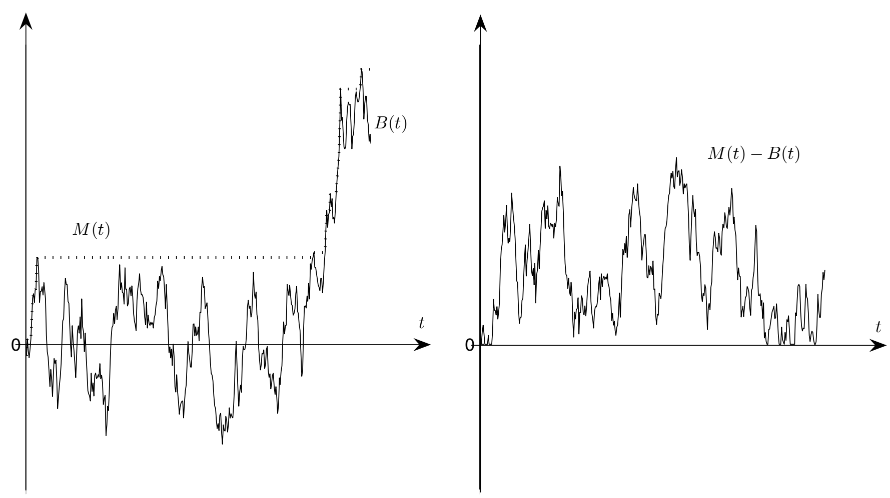


Figure 4.6: $\{B(t) : t \geqslant 0\}$ and $\{M(t) - B(t) : t \geqslant 0\}$.

Theorem 4.27 (Paul Lévy). *Let $\{M(t) : t \geqslant 0\}$ be the maximum process of a standard Brownian motion $\{B(t) : t \geqslant 0\}$. Then, the process*

$$\{M(t) - B(t) : t \geqslant 0\} \tag{4.16}$$

is a reflected Brownian motion.

Proof. The main step is to show that the process $\{M(t) - B(t) : t\}$ is a Markov process and its Markov transition kernel $P(t, x, \cdot)$ has modulus normal distribution with parameters x and t. Once this is established, it is immediate that the finite-dimensional distributions of this process agree with those of a reflected Brownian motion. Obviously, $\{M(t) - B(t)\}$ has almost surely continuous paths.

For the main step, fix $s, t \geqslant 0$ and observe that $M(s + t) = M(s) \vee (B(s) + \hat{M}(t))$, and hence

$$\begin{aligned} M(s+t) - B(s+t) &= \big(M(s) \vee (B(s) + \tilde{M}(t))\big) - B(s+t) \\ &= \big(M(s) - B(s)\big) \vee \tilde{M}(t) - \big(B(s+t) - B(s)\big) \end{aligned}$$

where $\tilde{M}(t) = \max_{0 \leqslant u \leqslant t} \tilde{B}(u)$ and $\tilde{B}(u) = B(u + t) - B(s)$. That is

$$M(s+t) - B(s+t) = \big(M(s) - B(s)\big) \vee \tilde{M}(t) - \tilde{B}(t) \,.$$

To finish, it suffices to check, for every $y \geqslant 0$, that $y \vee \tilde{M}(t) - \tilde{B}(t)$ has the same distribution as $|y + N(0, t)|$. Let $W(u) = \tilde{B}(t-u) - \tilde{B}(t)$ for $0 \leqslant u \leqslant t$, then it suffices to show

$$\begin{aligned} y \vee \tilde{M}(t) - \tilde{B}(t) &= \big(y - \tilde{B}(t)\big) \vee \big(\tilde{M}(t) - \tilde{B}(t)\big) \\ &= \big(y + W(t)\big) \vee M^{(W)}(t) \stackrel{d}{=} |y + N(0,t)| \,. \end{aligned}$$

For any $a > 0$, by Reflection priciple,

$$\begin{aligned} &\mathbb{P}\Big(\big(y + W(t)\big) \vee M^{(W)}(t) > a \Big) \\ &= \mathbb{P}\left(y + W(t) > a\right) + \mathbb{P}\left(W(t) < a - y, \tau_a^{(W)} < t\right) \\ &= \mathbb{P}\left(y + W(t) > a\right) + \mathbb{P}\left(W(t) < a - y, \tau_a^{(W)} < t\right) \\ &= \mathbb{P}\left(y + W(t) > a\right) + \mathbb{P}\left(W(t) > a + y\right) \\ &= \mathbb{P}\left(|y + N(0,t)| > a\right) \,. \end{aligned}$$

□

EXERCISE

¶ Exercise 4.4. For $x > y$, compute

$$\mathbb{P}\left(B(t) \leqslant y < x \leqslant M(t)\right) .$$

Then show that $(M(t), B(t))$ is a continuous random vector.

¶ Exercise 4.5. Let $\lambda(t) = \arg\max_{0\leqslant s\leqslant t} B(s)$, notice $\lambda(t)$ is well-defined, since all the local maximum of Brownian motion are distinct. Show that

$$\mathbb{P}(\lambda(t) \leqslant s) = \frac{2}{\pi} \arcsin \sqrt{\frac{s}{t}} . \tag{4.17}$$

¶ Exercise 4.6. For any $x > 0$ show that

$$\mathbb{P}\left(\sup_{0\leqslant u\leqslant t} |B_u| > x\right) \leqslant \frac{2t}{x^2}$$

4.7.3 Zero points of Brownian Motion

In this section we will investigate the (random) set

$$\mathcal{Z} = \{t \geqslant 0 : B(t) = 0\} . \tag{4.18}$$

Because the path $B(t)$ is continuous in t, the set $\mathcal{Z}$ is closed. Furthermore, With probability one the Lebesgue measure of $\mathcal{Z}$ is 0, because Fubini's theorem implies that the expected Lebesgue measure of $\mathcal{Z}$ is 0:

$$\begin{aligned}\mathbb{E}|\mathcal{Z}| = \mathbb{E}\int_0^\infty 1_{\{B(t)=0\}}dt = \int_0^\infty \mathbb{E}1_{\{B(t)=0\}}dt \\ = \int_0^\infty \mathbb{P}\left\{B_t = 0\right\} dt \\ = 0 ,\end{aligned}$$

where $|\mathcal{Z}|$ denotes the Lebesgue measure of $\mathcal{Z}$. Observe that for any fixed (nonrandom) $t > 0$, the probability that $t \in \mathcal{Z}$ is 0, because $\mathbb{P}(B(t) = 0) = 0$. Hence, because $\mathbb{Q}_+$ is countable,

$$\mathbb{P}\left\{\mathbb{Q}_+ \cap \mathcal{Z} \neq \varnothing\right\} = 0$$

First, we introduce the famous *arcsine law.* Define the last zero point of $\{B(t)\}$ in $[0,t]$ as L_t, that is

$$L(t) \coloneqq \sup\{s \in [0,t] : B(s) = 0\} \,. \tag{4.19}$$

Theorem 4.28 (Arcsin law). *For any* $s \in [0,t]$,

$$\mathbb{P}(L(t) \leqslant s) = \mathbb{P}(\mathcal{Z} \cap [s,t] = \varnothing) = \frac{2}{\pi} \arcsin \sqrt{\frac{s}{t}} \,. \tag{4.20}$$

Proof. Note that

$$\begin{aligned}
&\mathbb{P}\left(L(t) \leqslant s\right) \\
&= \int p_s(0,x)\mathbb{P}\big(B(s+u) \neq 0 \text{ for each } 0 \leqslant u \leqslant t-s \mid B(s) = x\big)dx \\
&= \int p_s(0,x)\mathbb{P}_x\big(\tau_0 > t-s\big)dx = \int p_s(0,x)\mathbb{P}_0\big(\tau_x > t-s\big)dx \\
&= \int p_s(0,x)\mathbb{P}\big(|B(t-s)| \leqslant |x|\big)dx = \int p_s(0,x)\mathbb{P}\big(\sqrt{t-s}|Z_1| \leqslant |x|\big)dx \,,
\end{aligned}$$

where Z_1 has distribution $N(0,1)$, and let Z_2 has identical distribution and independent with Z_1, then

$$\int p_s(0,x)\mathbb{P}\big(\sqrt{t-s}|Z_1| \leqslant |x|\big)dx = \mathbb{P}\left(\sqrt{t-s}|Z_1| \leqslant \sqrt{s}|Z_2|\right) \,,$$

letting $(Z_1, Z_2) = (R\cos\Theta, R\sin\Theta)$, we konw that Θ uniformly distributed in $[0, 2\pi]$, (and R has a *Rayleigh distribution*), thus

$$\begin{aligned}
&\mathbb{P}\left(\sqrt{t-s}|Z_1| \leqslant \sqrt{s}|Z_2|\right) = \mathbb{P}\left(\sqrt{t-s}|\sin\Theta| \leqslant \sqrt{s}|\cos\Theta|\right) \\
&= \mathbb{P}\left(|\tan\Theta| \leqslant \sqrt{\frac{s}{t-s}}\right) = \frac{2}{\pi}\arcsin\sqrt{\frac{s}{t}} \,. \quad \square
\end{aligned}$$

Corollary 4.29. *Let* $\sigma_0 \coloneqq \inf\{t > 0 : B_t = 0\}$, *then* $\mathbb{P}(\sigma_0 = 0) = 1$

Proof. By the arcsine law we have, for any $t > 0$,

$$\mathbb{P}(0 < L(t) < t) = 1 \,.$$

Thus $\sigma_0 \leqslant t$ a.s. Clearly $\sigma_0 = 0$ a.s.. $\square$

Theorem 4.30. *Almost surely, $\mathcal{Z}$ is a perfect set.*[2]

Proof. That $\mathcal{Z}$ is closed from path-continuity, as noted earlier. Note that

$$\mathcal{Z}(\omega) = \{0\} \cup C_-(\omega) \cup C_+(\omega)$$

where

$$C_-(\omega) = \{t > 0 : \text{ exists } \ t_n \uparrow t \ \text{ s.t. } B(t_n, \omega) = 0\} \,,$$
$$C_+(\omega) = \{t > 0 : \text{ exists } \ \delta > 0 \ \text{ s.t. } B(s, \omega) \neq 0 \ \text{ for all } \ t - \delta < s < t\} \,.$$

Obviously, $C_-(\omega) \subset d(\mathcal{Z}(\omega))$, it suffices to show that $C_+(\omega) \subset d(\mathcal{Z}(\omega))$ almost surely.

Fix a rational number $q > 0$, and define ν_q by

$$\nu_q = \inf\{t \geqslant q : B(t) = 0\} \,,$$

then one can see that, almost surely,

$$C_+(\omega) \subset \{\nu_q(\omega) : q \in \mathbb{Q}_+\}$$

it suffices to show that, almost surely, $\nu_q \in d(\mathcal{Z})$. By the Strong Markov Property, the process

$$B(\nu_q + t) - B(\nu_q) = B(\nu_q + t)$$

is a standard Brownian motion. Consequently, by Corollary 4.29,

$$\inf\{t > 0 : B(\nu_q + t) = 0\} = 0 \qquad \text{a.s.}$$

Thus $\nu_q \in d(\mathcal{Z})$ a.s. as required. $\square$

Remark. It can be shown (this is not especially difficult) that every compact perfect set of Lebesgue measure zero is homeomorphic to the *Cantor set*. Thus, with probability one, the set of zeros of the Brownian path $B(t)$ in the unit interval is a homeomorphic image of the Cantor set.

[2] In general topology, a subset of a topological space is perfect if it is closed and has no isolated points.

Fractal dimension*

How "big" is the set $\mathcal{Z}$? To discuss this we need to discuss the notion of a dimension of a set. There are two similar notions of dimension, *Hausdorff dimension* and *box dimension*, which can give fractional dimensions to sets. (There is a phrase "fractal dimension" which is used a lot in scientific literature. As a rule, the people who use this phrase are not distinguishing between Hausdorff and box dimension and could mean either one.) The notion of dimension we will discuss here will be that of box dimension, but all the sets we will discuss have Hausdorff dimension equal to their box dimension.

Suppose we have a bounded set A in d -dimensional space $\mathbb{R}^d$. Suppose we cover A with d -dimensional balls of diameter ϵ. How many such balls are needed? If A is a line segment of length 1 (one-dimensional set), then ϵ^{-1} such balls are needed. If A is a two-dimensional square, however, on the order of ϵ^{-2} such balls are needed. One can see that for a standard k -dimensional set, we need ϵ^{-k} such balls. This leads us to define the (box) dimension of the set A to be the number D such that for small ϵ the number of balls of diameter ϵ needed to cover A is on the order of ϵ^{-D}.

¶ EXAMPLE 4.7. Consider the fractal subset of $[0, 1]$, the Cantor set. The Cantor set A can be defined as a limit of approximate Cantor sets A_n. We start with $A_0 = [0, 1]$. The next set A_1 is obtained by removing the open middle interval $(1/3, 2/3)$, so that

$$A_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]$$

The second set A_2 is obtained by removing the middle thirds of the two intervals in A_1, hence

$$A_2 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right]$$

In general A_{n+1} is obtained from A_n by removing the "middle third" of each interval. The Cantor set A is then the limit of these sets A_n

$$A = \bigcap_{n=1}^{\infty} A_n$$

Note that A_n consists of 2^n intervals each of length 3^{-n}. Suppose we try to cover A by intervals of length 3^{-n}

$$\left[\frac{k-1}{3^n}, \frac{k}{3^n}\right]$$

We need 2^n such intervals. Hence the dimension D of the Cantor set is the number such that $2^n = \left(3^{-n}\right)^{-D}$, i.e.,

$$D = \frac{\ln 2}{\ln 3} \approx 0.631$$

Now consider the set $\mathcal{Z}$ and consider $\mathcal{Z}_1 = \mathcal{Z} \cap [0,1]$. We will try to cover $\mathcal{Z}_1$ by one-dimensional balls (i.e., intervals) of diameter (length) $\epsilon = 1/n$. For ease we will consider the n intervals

$$\left[\frac{k-1}{n}, \frac{k}{n}\right], \quad k = 1, 2, \ldots n$$

How many of these intervals are needed to cover $\mathcal{Z}_1$? Such an interval is needed if $\mathcal{Z}_1 \cap [(k-1)/n, k/n] \neq \varnothing$. Note that, by arcsince law,

$$P(k,n) = \mathbb{P}\left(\mathcal{Z}_1 \cap \left[\frac{k-1}{n}, \frac{k}{n}\right] \neq \varnothing\right) = 1 - \frac{2}{\pi} \arctan \sqrt{k-1}$$

Therefore, the expected number of the intervals needed to cover Z_1 looks like

$$\sum_{k=1}^{n} P(k,n) = \sum_{k=1}^{n} \left[1 - \frac{2}{\pi} \arctan \sqrt{k-1}\right]$$

To estimate the sum, we need to consider the Taylor series for arctan $(1/t)$ at $t = 0$ (which requires remembering the derivative of arctan),

$$\arctan \frac{1}{t} = \frac{\pi}{2} - t + O\left(t^2\right)$$

In other words, for x large,

$$\arctan x \approx \frac{\pi}{2} - \frac{1}{x}$$

Hence

$$\sum_{k=1}^{n} P(k,n) \approx 1 + \sum_{k=2}^{n} \frac{2}{\pi\sqrt{k-1}} \approx \frac{2}{\pi}\int_1^n (x-1)^{-1/2}dx \approx \frac{4}{\pi}\sqrt{n}$$

Hence it takes on the order of $\sqrt{n}$ intervals of length $1/n$ to cover Z_1, or, in other words,

Theorem. *The fractal dimension of the zero set $\mathcal{Z}$ is $1/2$.*

4.8 Brownian motion as a continuous martingale

In the previous section we have taken a particular feature of Brownian motion, the Markov property, and introduced an abstract class of processes, the Markov processes, which share this feature. In this section we follow a similar plan, taking a different feature of Brownian motion, the martingale property, as a starting point.

Definition. $(\Omega, \mathcal{F}, \{\mathcal{F}(t)\}, \mathbb{P})$ is a filtered probability space. Then a process $\{M(t)\}$ is called a **martingale** with respect to filtration $\{F(t)\}$ and probability measure $\mathbb{P}$, [3] if it is an adapted integrable process and, for each $0 \leqslant s \leqslant t$,

$$\mathbf{E}\left[M(t) \mid \mathcal{F}(s)\right] = M(s) \quad \text{a.s.} \tag{4.21}$$

The process is called a **submartingale** if $\geqslant$ holds, and a **supermartingale** if $\leqslant$ holds in the display above. A **continuous martingales** means that, almost surely, their sample paths are continuous.

Remark. Intuitively, a martingale is a process where the current state $M(t)$ is always the best prediction for its further states. In this sense, martingales describe *fair games*. If $\{M(t) : t \geqslant 0\}$ is a martingale, the process $\{|M(t)| : t \geqslant 0\}$ need not be a martingale, but it still is a submartingale, as a simple application of the triangle inequality shows.

¶ Example 4.8. For a linear Brownian motion $\{B(t)\}$ we have

$$\begin{aligned}\mathbf{E}\left[B(t) \mid \mathcal{F}(s)\right] &= \mathbf{E}\left[B(t) - B(s) \mid \mathcal{F}(s)\right] + B(s)\\ &= \mathbb{E}(B(t) - B(s)) + B(s) = B(s)\,,\end{aligned}$$

for $s \leqslant t$, thus $\{B(t)\}$ is a continuous martingale.

[3] Sometimes we say $(M(t), \mathcal{F}(t), \mathbb{P})$ is a martingale for short.

¶ EXAMPLE 4.9. For a linear Brownian motion $\{B(t)\}$, one can show that

$$\left\{\exp\left(\sigma B(t) - \frac{\sigma^2 t}{2}\right) : t \geqslant 0\right\} \tag{4.22}$$

is a continuous martingale. Moreover, by taking derivatives $\frac{\partial^n}{\partial \sigma^n}$ at 0, that the following processes are martingales.

$$\left\{B(t)^2 - t : t \geqslant 0\right\} \tag{4.23}$$

$$\left\{B(t)^3 - 3tB(t) : t \geqslant 0\right\}, \tag{4.24}$$

$$\left\{B(t)^4 - 6tB(t)^2 + 3t^2 : t \geqslant 0\right\} \tag{4.25}$$

We now state a useful facts about martingales, which we will exploit extensively: the *optional stopping theorem*.

Theorem (Doob's Optional stopping theorem). *If $(M(t), \mathcal{F}(t), \mathbb{P})$ is a martingale, then, for every stopping time T,*

$$\left\{M(t \wedge T), \mathcal{F}(t), \mathbb{P}\right\}$$

is a martingale. If, in addition, $\{M(t \wedge T)\}$ is uniformly integrable, then

$$\mathbb{E}M(T) = \mathbb{E}M(0) \tag{4.26}$$

4.8.1 Exitting probability and Exitting time

We now use the martingale property and the optional stopping theorem to obtain exit probabilities and expected exit times for a linear Brownian motion. In next section we shall explore the corresponding problem for higher-dimensional Brownian motion using harmonic functions.

Theorem 4.31. *$\{B(t)\}$ is a linear Brownian motion starting at $x \in (a, b)$. Let $\tau := \inf\{t \geqslant 0 : B(t) \in \{a, b\}\} = \tau_a \wedge \tau_b$, then*

$$\mathbb{P}_x\big(B(\tau) = a\big) = \frac{x-a}{b-a} \quad \text{and} \quad \mathbb{P}_x\big(B(\tau) = b\big) = \frac{b-x}{b-a}\,. \tag{4.27}$$

Proof. Note that

$$|B(t \wedge \tau)| \leqslant |a| \vee |b|\,, \text{ for all } t$$

by Doob's Optional stopping theorem we have that

$$\mathbb{E}_x(B(\tau)) = \mathbb{E}_x(B(0)) = x$$

Thus

$$\begin{aligned} a\,\mathbb{P}_x\big(B(\tau) = a\big) + b\,\mathbb{P}_x\big(B(\tau) = b\big) &= x \\ \mathbb{P}_x\big(B(\tau) = a\big) + \mathbb{P}_x\big(B(\tau) = b\big) &= 1 \end{aligned}$$

Solving this we obtain (4.27). □

Lemma 4.32. *The random variable τ has a moment generating function, that is, $E^x\left(e^{\theta\tau}\right)$ is finite for sufficiently small θ. In particular it has finite moments of all orders.*

Proof. We have the crude inequality

$$\sup_{x\in(a,b)} \mathbb{P}_x(\tau > 1) \leqslant \sup_{x\in(a,b)} \mathbb{P}_x(B(1) \in (a,b))$$

The right side can be expressed in terms of a normal distribution but it is sufficient to see that it is strictly less than 1, and we denote it by δ.

The next step is a basic argument using the Markovian character of the process. For any $x \in (a,b)$ and $n \geqslant 1$,

$$\mathbb{P}_x(\tau > n) \leqslant \mathbb{P}_x\left(B(1), \cdots, B(n) \in [a,b]\right) \leqslant \delta^n\,, \text{ for all } n\,.$$

In fact the argument above yields more. For any ϵ such that $e^\epsilon\delta < 1$, we have

$$\mathbb{E}_x e^{\epsilon\tau} \leqslant \sum_{n=1}^{\infty} e^{\epsilon n}\mathbb{P}_x(n-1 < \tau \leqslant n) \leqslant \sum_{n=1}^{\infty} e^{\epsilon n}\delta^{n-1} < \infty\,.$$

□

Theorem 4.33. *$\{B(t)\}$ is a linear Brownian motion starting at $x \in (a,b)$. Let $\tau := \inf\{t \geqslant 0 : B(t) \in \{a,b\}\} = \tau_a \wedge \tau_b$, then*

$$\mathbb{E}_x \tau = (x-a)(b-x) \tag{4.28}$$

Proof. Note that

$$|B(t\wedge\tau)^2 - t\wedge\tau| \leqslant |a| \vee |b| + \tau\,, \text{ for all } t$$

by Doob's Optional stopping theorem we have that

$$\mathbb{E}_x(B(\tau)^2 - \tau) = \mathbb{E}_x(B(0)^2) = x^2$$

Thus

$$\mathbb{E}_x \tau = \mathbb{E}_x \left(B(\tau)^2\right) - x^2 = (x-a)(b-x)\,,$$

as required, and in order to compute $\mathbb{E}_x \left(B(\tau)^2\right)$ we used (4.27). □

For more reading meterial, see K.L.Chung, *Green, Brown, and Probability & Brownian Motion on the Line.*

4.9 Potential theory

In this chapter we explore the relation of pde and Brownian motion. This approach will be particularly useful for d-dimensional Brownian motion for $d > 1$. It allows us to study the fundamental questions of transience and recurrence of Brownian motion, investigate the classical Dirichlet problem.

Theorem 4.34. *For an interval (a, b) and $f : \{a, b\} \to \mathbb{R}$, then the Dirichlet problem*

$$\begin{cases} \phi'' = 0 & \text{in } (a,b) \\ \phi = f & \text{in } \{a,b\} \end{cases} \tag{4.29}$$

has a (unique) sulution given by

$$\phi(x) = \mathbb{E}_x[f(B(\tau))] \,. \tag{4.30}$$

where $\tau = \tau_a \wedge \tau_b = \inf\{t \geqslant 0 : B(t) \in \{a, b\}\}$.

Proof. To show (4.30) solving Dirichlet problem, note that

$$\phi(x) = f(a)\mathbb{P}_x\left(B(\tau) = a\right) + f(b)\mathbb{P}_x\left(B(\tau) = b\right) \,.$$

We will not use Theorem 4.31, but use strong Markov property to show that

$$\mathbb{P}_x\left(B(\tau) = b\right) = \frac{x-a}{b-a} \quad \text{and} \quad \mathbb{P}_x\left(B(\tau) = a\right) = \frac{b-x}{b-a} \,,$$

then

$$\phi(x) = f(a)\frac{b-x}{b-a} + f(b)\frac{x-a}{b-a}$$

satisfying (4.29).

Note that for δ such that $a < x - \delta < x < x + \delta < b$, let $\xi = \tau_{x-\delta} \wedge \tau_{s+\delta}$, by strong Markov property,

$$\begin{aligned}
&\varphi(x) \coloneqq \mathbb{P}_x\left(B(\tau) = b\right) \\
&= \mathbb{P}_x(B(\xi) = x - \delta)\mathbb{P}_{x-\delta}\left(B(\tau) = b\right) + \mathbb{P}_x(B(\xi) = x + \delta)\mathbb{P}_{x+\delta}\left(B(\tau) = b\right) \\
&= \frac{1}{2}\varphi(x + \delta) + \frac{1}{2}\varphi(x - \delta) \,,
\end{aligned}$$

By induction we have, for $x = \frac{k}{2^n}a + (1 - \frac{k}{2^n})b$, $0 \leqslant k \leqslant 2^n$, $\varphi(x) = (x-a)/(b-a)$. On the other hand, for $a \leqslant x_1 < x_2 \leqslant b$,

$$\varphi(x_1) = \mathbb{P}_{x_1}(\tau_{x_2} < \tau_a)\,\mathbb{P}_{x_2}(B(\tau) = b) \leqslant \mathbb{P}_{x_2}(B(\tau) = b) = \varphi(x_2)$$

thus φ is monotone, so for all $x \in [a, b]$ we have

$$\varphi(x) = \frac{x-a}{b-a}\,. \qquad \square$$

Theorem 4.35. *For an interval (a, b) and g is a bounded continuous function in (a, b), the Poisson problem*

$$\begin{cases} \psi'' = -2g & \text{in } (a,b) \\ \psi = 0 & \text{in } \{a,b\} \end{cases} \tag{4.31}$$

has a (unique) sulution given by

$$\psi(x) = \mathbb{E}_x\left[\int_0^\tau g(B(t))\mathrm{d}t\right]. \tag{4.32}$$

Proof. To show (4.32) solving the *Poisson problem*, we need a lemma

Lemma (Schwarz). *f and g are two continuous functions on (a, b). If*

$$\lim_{h\to 0}\frac{f(x+h) - 2f(x) + f(x-h)}{h^2} = g(x)\,, \text{ for all } x \in (a,b)$$

Then $f \in C^{(2)}(a, b)$ and $f'' = g$.

Step 1. We show that ψ is continuous in (a, b) first. For given $\delta > 0$ and $y \in (x - \delta, x + \delta) \subset (a, b)$,

$$\psi(y) = \mathbb{E}_y\int_0^\xi g(B_t)\,dt + \mathbb{P}_y(B(\xi) = x-\delta)\,\psi(x-\delta) + \mathbb{P}_y(B(\xi) = x+\delta)\,\psi(x+\delta)\,,$$

Note that $\mathbb{P}_y(B(\xi) = x - \delta)$ is continuous on $(x-\delta, x+\delta)$, thus when $y \to x$,

$$|\psi(y) - \psi(x)| \leqslant \left|\mathbb{E}_y\int_0^\xi g(B_t)\,dt\right| + \left|\mathbb{E}_y\int_0^\xi g(B_t)\,dt\right| + o(1)$$

by Theorem 4.33, $\mathbb{E}_y \xi \leqslant \delta^2$ for all y, thus

$$\limsup_{y\to x} |\psi(y)-\psi(x)| \leqslant 2\|g\|_\infty \delta^2$$

letting $\delta \to 0$ we have $\psi(y)\to\psi(x)$, so ψ is continuous on (a,b).

Step 2. we show $\psi'' = -2g$. Observe that

$$\mathbb{E}_x \int_0^\tau g\left(B_t\right) dt = \mathbb{E}_x \int_0^\xi g\left(B_t\right) dt + \mathbb{E}_x \int_\xi^\tau g\left(B_t\right) dt\,,$$

and by strong Markov property we have

$$\mathbb{E}_x \int_\sigma^\tau g\left(B_t\right) dt = \frac{1}{2}(\psi(x+\delta)+\psi(x-\delta))\,,$$

Thus

$$\psi(x+\delta)-2\psi(x)+\psi(x-\delta) = -2\mathbb{E}_x \int_0^\xi g\left(B_t\right) dt\,.$$

by Theorem 4.33, $\mathbb{E}_x \xi = \delta^2$, thus

$$\left|\frac{\psi(x+\delta)-2\psi(x)+\psi(x-\delta)}{\delta^2}+2g(x)\right| = \frac{2}{\delta^2}\left|g(x)\mathbb{E}_x\xi - \mathbb{E}_x\int_0^\xi g\left(B_t\right) dt\right|$$

$$\leqslant \frac{2}{\delta^2}\,\mathbb{E}_x \int_0^\xi \left|g(x)-g\left(B_t\right)\right| dt \;\leqslant\; \frac{2}{\delta^2} \sup_{x-\delta\leqslant y\leqslant x+\delta} |g(x)-g(y)|\,\mathbb{E}_x\xi$$

$$= 2 \sup_{x-\delta\leq y\leq x+\delta} |g(x)-g(y)| \stackrel{\delta\to 0}{\longrightarrow} 0\,.$$

Thus $\psi \in C^{(2)}(a,b)$ and $\psi'' = -2g$.

Step 3. Finally, we are supposed to show that ψ is continuous in $[a,b]$, which can be proved by the same method used in step 1. □

In fact, the teo theorems can be extended to higner dimension. Let D be a bounded region (region means non-empty, open, and connected set) in $\mathbb{R}^d$, with boundary ∂D. (在高维时，我们需要对区域边界 ∂D 加一些所谓的“正则性”假设，以保证当 $x \to \partial D$ 时，$B(\tau_{\partial D}) \to a$. 当然，光滑的边界都是正则的. 从上面的证明可以看出，为了保证解是连续到边界的，正则性假设是至关重要的.)

The proof of the following theorem can be found in K.L.Chung, *Green, Brown, and Probability & Brownian Motion on the Line.*

Theorem 4.36. *Suppose D is a regoin in $\mathbb{R}^d$ with a smooth boundary ∂D, set $\tau = \tau_{\partial D} = \inf\{t \geqslant 0 : B(t) \in \partial D\}$.*

- *f is a continouos function on ∂D. The Dirichlet problem*

$$\begin{cases} \Delta\phi = 0 & \text{in } D \\ \phi = f & \text{in } \partial D \end{cases} \tag{4.33}$$

  has a (unique) sulution given by

$$\phi(x) = \mathbb{E}_x[f(B(\tau))]\,. \tag{4.34}$$

- *g is a bounded continuous function on D, the Poisson problem*

$$\begin{cases} \Delta\psi = -2g & \text{in } D \\ \psi = 0 & \text{in } \partial D \end{cases} \tag{4.35}$$

  has a (unique) sulution given by

$$\psi(x) = \mathbb{E}_x\left[\int_0^\tau g(B(t))\mathrm{d}t\right]\,. \tag{4.36}$$

4.9.1 Recurrence and Transience

In this subection we discuss whether the Brownian motion keeps returning to the origin. We have already answered this question for one-dimensional Brownian motion; if $B(t)$ is a standard (one-dimensional) Brownian motion, then $B(t)$ is recurrent, i.e., there are arbitrarily large times t with $B(t) = 0$.

Now suppose $\{B(t)\}$ is a d-dimensional standard Brownian motion. Let $0 < r < R < \infty$ and let D be the annulus

$$D = \left\{x \in \mathbb{R}^d : r < \|x\| < R\right\}$$

with boundary

$$\partial D = \left\{x \in \mathbb{R}^d : \|x\| = r \text{ or } \|x\| = R\right\}\,.$$

Suppose $x \in D$, let $\tau = \tau_{\partial D} = \inf\{t : B(t) \in \partial D\}$, and

$$\phi(x) := \mathbb{P}_x\left(\tau_r < \tau_R\right) = \mathbb{E}_x(f(B(\tau)))\,, \text{ where } f(x) = \begin{cases} 1, & \|x\| = r\,. \\ 0, & \|x\| = R\,. \end{cases}$$

Thus ϕ is the solution of the Dirichelet problem

$$\begin{cases} \Delta\phi(x) = 0\,, \quad x \text{ in } D\,. \\ \phi(x) = 1,\ \|x\| = r \ \text{ and } \ \phi(x) = 0,\ \|x\| = R\,. \end{cases}$$

To find ϕ we first note that the symmetry of Brownian motion implies

$$\phi(x_1) = \phi(x_2) \text{ for any } \|x_1\| = \|x_2\|\,.$$

Thus we define ν by,

$$\nu(\rho) = \phi(x)\,, \text{ for } \|x\| = \rho\,.$$

Clearly, ν is well-defined. Then one can check that

$$\Delta\phi(x) = 0 \Rightarrow \nu''(\rho) + \frac{d-1}{\rho}\nu'(\rho) = 0$$

is given by

$$\nu(\rho) = \begin{cases} c_1\rho + c_2, & d = 1 \\ c_1 \ln\rho + c_2, & d = 2 \\ c_1\rho^{2-d} + c_2, & d \geqslant 3 \end{cases}$$

From boundary conditions, we have

$$\varphi(x) = \begin{cases} \dfrac{R - \|x\|}{R - r}\,, & d = 1 \\ \dfrac{\log R - \log\|x\|}{\log R - \log r}\,, & d = 2 \\ \dfrac{R^{2-d} - \|x\|^{2-d}}{R^{2-d} - r^{2-d}}\,, & d \geqslant 3 \end{cases} \tag{4.37}$$

Consider now the two-dimensional case. Let $x \in \mathbb{R}^2$ and suppose that a Brownian motion starts at x. Take any $r > 0$, and ask the question: What is the probability that the Brownian motion never returns to the disc of radius r about 0 ? The argument above gives us the probability of reaching the circle of radius R before reaching the disc. The probability we are interested in is therefore

$$\lim_{R\to\infty} \mathbb{P}_x\left(\tau_r < \tau_R\right) = \lim_{R\to\infty} \frac{\log R - \log\|x\|}{\log R - \log r} = 1\,.$$

Hence, with probability one the Brownian motion always returns to the disc of radius r and hence it returns infinitely often and at arbitrarily large times.

Does it ever return to the point 0, i.e., are there times t with $B(t) = 0$? Again, start the walk at $x \neq 0$. If there is a positive probability of reaching 0, then there must be an R such that the probability of reaching 0 before reaching the circle of radius R is positive. But this latter probability can be written as

$$\lim_{r\to 0} \mathbb{P}_x\left(\tau_r < \tau_R\right) = \lim_{r\to 0} \frac{\log R - \log\|x\|}{\log R - \log r} = 0\,.$$

Hence the Brownian motion never actually returns to 0. To summarize, the Brownian motion in two dimensions returns arbitrarily close to 0 infinitely often, but never actually returns to 0. We say that the Brownian motion in two dimensions is *neighborhood recurrent* but *not point recurrent*.

Now consider $d \geqslant 3$. Again we take $r > 0$ and ask what is the probability that the Brownian motion starting at x never returns to the ball of radius r If $|x| > r$, this is given by

$$\lim_{R\to\infty} \frac{r^{2-d} - \|x\|^{2-d}}{r^{2-d} - R^{2-d}} = 1 - \left(\frac{r}{\|x\|}\right)^{d-2} < 1\,.$$

since the probability is less than 1 , we can see that eventually the Brownian motion escapes from any ball around the origin and hence goes off to infinity. We say that in this case the Brownian motion is *transient*.

4.10 Brownian bridge, Ornstein-Uhlenbeck Process

4.10.1 Brownian bridge

In the most common formulation, the Brownian bridge process is obtained by taking a standard Brownian motion process $\{B(t)\}$, restricted to the interval $[0,1]$, and conditioning on the event that $B(1) = 0$. Since $B(0) = 0$ also, the process is tied down at both ends, and so the process in between forms a bridge (albeit a very jagged one).

The Brownian bridge turns out to be an interesting stochastic process with surprising applications, including a very important application to statistics. In terms of a definition, however, we will give a list of characterizing properties as we did for standard Brownian motion and for Brownian motion with drift and scaling.

Definition 4.6. A **Brownian bridge** is a Guassian stochastic process $\{X(t) : t \in [0,1]\}$ that satisfies the following properties:

(i) For all $t, s \in [0,1]$, $\mathbb{E}X(t) = 0$ and $\mathrm{Cov}\,(X(t), X(s)) = t \wedge s - ts$

(ii) With probability 1, $t \mapsto X(t)$ is continuous on $[0,1]$.

Naturally, the first question is whether there exists such a process. The answer is yes, of course, otherwise why would we be here? But in fact, we will see several ways of constructing a Brownian bridge from a standard Brownian motion.

¶ Example 4.10. Suppose that $\{B(t)\}$ is a standard Brownian motion, and let

$$X(t) = B(t) - tB(1)\,, \quad \text{for } t \in [0,1]\,.$$

Then $\{X(t) : t \in [0,1]\}$ is a Brownian bridge.

Conversely to the construction above, we can build a standard Brownian motion on the time interval $[0,1]$ from a Brownian bridge.

Suppose that $\{X(t) : t \in [0,1]\}$ is a Brownian bridge, and Z is a random variable with a standard normal distribution, independent of $\{X(t) : t \in [0,1]\}$. Let

$$B(t) = X(t) + tZ\,, \quad \text{for } t \in [0,1]\,.$$

Then $\{B(t) : t \in [0,1]\}$ is a standard Brownian motion.

Here's another way to construct a Brownian bridge from a standard Brownian motion.

¶ EXAMPLE 4.11. Suppose that $\{B(t)\}$ is a standard Brownian motion. Define $X(1) = 0$ and

$$X(t) = (1-t)B\left(\frac{t}{1-t}\right)\,, \text{ for } t \in [0,1)\,.$$

Then $\{X(t) : t \in [0,1]\}$ is a Brownian bridge.

Conversely, we can construct a standard Brownian motion from a Brownian bridge.

Suppose that $\{X(t) : t \in [0,1]\}$ is a Brownian bridge. Define

$$B(t) = (1+t)B\left(\frac{t}{1+t}\right)\,, \text{ for } t \in [0,\infty)\,.$$

Then $\{B(t)\}$ is a standard Brownian motion.

¶ EXAMPLE 4.12. Suppose that $\{B(t) : t \in [0,1]\}$ is a standard Brownian motion. Conditional on $\{B(1) - 0\}$, the joint density of $(B(s), B(t))$ is

$$\begin{aligned}
\tilde{p}(s,t;x,y) &= \frac{p(s,t,1;x,y,0)}{p(1;0)} \\
&= \frac{1}{\sqrt{2\pi s}} e^{-\frac{x^2}{2s}} \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{(y-x)^2}{2(t-x)^2}} \frac{1}{\sqrt{2\pi(1-t)}} e^{-\frac{y^2}{2(1-t)}} \Big/ \left(\frac{1}{\sqrt{2\pi}}\right) \\
&= C \exp\left\{-\frac{x^2}{2s} - \frac{(y-x)^2}{2(t-s)} - \frac{y^2}{2(1-t)}\right\} \\
&= C \exp\left\{-\frac{1}{2}\left(x^2 \frac{t}{s(t-s)} - 2xy\frac{1}{(t-s)} + y^2 \frac{1-s}{(t-s)(1-t)}\right)\right\}.
\end{aligned}$$

This shows that the distribution of $(X(s), X(t))$ is the same to $(B(s), B(t))$ conditional on $B(1) = 0$. Therefore, Brownian bridge is obtained by taking a standard Brownian motion process $\{B(t)\}$, restricted to the interval $[0, 1]$, and conditioning on the event that $B(1) = 0$.

The General Brownian Bridge The processes constructed above (in several ways) is the standard Brownian bridge. It's a simple matter to generalize the process so that it starts at a and ends at b, for arbitrary $a, b \in \mathbb{R}$.

Suppose that $\{Z(t) : t \in [0, 1]\}$ is a Brownian bridge let a, $b \in \mathbb{R}$ and deffine

$$X(t) = (1 - t)a + tb + Z(t), \text{ for } t \in [0, 1] .$$

Then $\{X_t, : t \in [0, 1]\}$ is *a Brownian bridge from a to b.*

Of course, any of the constructions above for standard Brownian bridge can be modified to produce a general Brownian bridge. Here are the characterizing properties.

Proposition 4.37. *The Brownian bridge process $\{X(t) : t \in [0, 1]\}$ from a to b is characterized by the following properties:*

(i) $X_0 = a$ *and* $X_1 = b$ *(each with probability 1).*

(ii) $\{X(t) : t \in [0, 1]\}$ *is a Gaussian process.*

(iii) $\mathbb{E}X(t) = (1 - t)a + tb$ *for* $t \in [0, 1]$ *and* $\mathrm{Cov}\,(X(s), X(t)) = s \wedge t - st$ *for* $s, t \in [0, 1]$.

(iv) *With probability* 1, $t \mapsto X(t)$ *is continuous on* $[0, 1]$.

Applications : The Empirical Distribution Function We start with a problem that is one of the most basic in statistics. Suppose that T is a real-valued random variable with an unknown distribution. Let F denote the distribution function of T, i.e., $F(x) = \mathbb{P}(T \leqslant x)$ for $x \in \mathbb{R}$. Our goal is to construct an estimator of F.

Naturally our first step is to sample from the distribution of T. This generates a sequence $T_1, T_2, \cdots$ of independent variables, each with the distribution of F. For $n \in \mathbb{N}$ and $x \in \mathbb{R}$, the natural estimator of $F(x)$ based on the first n sample values is

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{T_i \leqslant x\}} \,.$$

which is simply the proportion of the first n sample values that fall in the interval $(-\infty, t]$. Appropriately enough, F_n is known as the *empirical distribution finction* corresponding to the sample of size n. Note that

$$1_{\{T_1 \leqslant x\}}, 1_{\{T_2 \leqslant x\}}, \cdots$$

is a sequence of independent, identically distributed indicator variables (and hence is a sequence of Bernoulli trials), and corresponds to sampling from the distribution of $1_{\{T \leqslant x\}}$.

The estimator $F_n(x)$ is simply the sample mean of the first n of these variables. The numerator, the number of the original sample variables with values in $(-\infty, t]$, has the binomial distribution with parameters n and $F(x)$. Like all sample means from independily distributed samples, $F_n(x)$ satisfites some basic and important properties. A summary is given below, but to make of some of these facts, you need to reall the mean and variance of the indicator variable that we are sampling from :

$$\mathbb{E}\, 1_{\{T \leqslant x\}} = F(x) \quad \text{and} \quad \mathrm{Var}(1_{\{T \leqslant x\}}) = F(x)[1 - F(x)] \,.$$

For fixed $x \in \mathbb{R}$,

(i) $F_n(x)$ is an unbised estimater of $F(x)$.

(ii) $\mathrm{Var}\,(F_n(x)) = \frac{F(x)[1-F(x)]}{n}$, so $F_n(x)$ is a consistent estimater of $F(x)$.

(iii) By SLLN $F_n(x) \to F(x)$ as $n \to \infty$ with probability 1.

(iv) By CLT, $\sqrt{n}\,[F_n(x) - F(x)]$ has mean 0 and variance $F(x)[1 - F(x)]$ and converges weakly to the normal distribution with these parameters as $n \to \infty$.

The propositions above gives us a great deal of information about $F_n(x)$ for fixed x, but now we want to let t vary and consider the expression in (iv), namely $x \mapsto \sqrt{n}\,[F_n(x) - F(x)]$, as a random process for each $n \in \mathbb{N}$. The key is to consider a very special distribution first.

Suppose that T uniformly distributed on the interval $[0, 1]$, in this case the distribution function is simply $F(t) = t$ for $t \in [0, 1]$, so we have the sequence of stochastic processes $\{X_n(t) : t \in [0, 1]\}$ for $n \in \mathbb{N}$ where

$$X_n(t) = \sqrt{n}\,[F_n(t) - t] \ .$$

Of course, the previous results apply, for fixed $t \in [0, 1]$, the process $X_n(t)$ has mean function 0, variance function $t(1 - t)$, the distribution $X_n(t)$ converges to the corresponding normal distribution as $n \to \infty$. Here is the new bit of information, the covariance function of $\{X_n(t) : t \in [0, 1]\}$ is the same as that of the Brownian bridge :

$$\mathrm{Cov}(X_n(t), X_n(s)) = s \wedge t - st\,, \ \text{for } s, t \in [0, 1]\,.$$

In fact, we have

Theorem. $\{X_n(t) : t \in [0, 1]\}$ *converges weakly to a Brownian bridge as* $n \to \infty$.

4.10.2 Ornstein-Uhlenbeck Process

Suppose that $\{B(t)\}$ is a standard Brownian motion. For $t \in \mathbb{R}$, let

$$X(t) = \mathrm{e}^{-\alpha t} B(\mathrm{e}^{2\alpha t})\,.$$

Then

(i) $\{X(t)\}$ is a Gaussian process.

(ii) For any $t, s \in \mathbb{R}$,

$$\mathbb{E}X(t) = 0 \quad \text{and} \quad \mathrm{Cov}(X(s), X(t)) = \mathrm{e}^{-\alpha|s-t|}\,.$$

(iii) Almost surely, $t \mapsto X(t)$ is continuous on $\mathbb{R}$.

A process $\{X(t)\}$ satisfying (i),(ii),(iii) is called **Ornstein-Uhlenbeck process**, or **OU process** for short.

Note that for any t, $X(t)$ is normally distributed with paramaters 0 and 1, thus, standard normal distribution is the invariant distribution of OU process.

Moreover, we consider the specific OU process, $X(t) = \mathrm{e}^{-\alpha t}B(\mathrm{e}^{2\alpha t})$. Then for given $s \in \mathbb{R}$ and $t > 0$,

$$\begin{aligned} X(t+s) &= e^{-\alpha(t+s)}B\left(e^{2\alpha(t+s)}\right) \\ &= e^{-\alpha(t+s)}B\left(e^{2\alpha s}\right) + e^{-\alpha(t+s)}\left(B\left(e^{2\alpha(t+s)}\right) - B\left(e^{2\alpha s}\right)\right) \\ &= e^{-\alpha t}X(s) + Y \end{aligned}$$

where Y is defined by the last equation. Obviously, Y is independent of $X(s)$ and $Y \sim N(0, 1-\mathrm{e}^{-2\alpha t})$. Thus we deduce that $\{X(t)\}$ is a Markov process with the transition density

$$\mathfrak{p}(t,x,y) = \frac{1}{\sqrt{2\pi\left(1-e^{-2\alpha t}\right)}}\exp\left\{-\frac{1}{2\left(1-e^{-2\alpha t}\right)}\left(y - e^{-\alpha t}x\right)^2\right\}$$

and

$$\begin{aligned} &\lim_{t\to\infty}\mathfrak{p}(t,x,y) = \frac{1}{\sqrt{2\pi}}e^{-y^2/2} = \phi(y)\,, \\ &\int \mathfrak{p}(t,x,y)\phi(x)\,\mathrm{d}x = \phi(y)\,. \end{aligned}$$