" Actually our family used to be pretty well-to-do... but then Pops went to a probability theory and stochastic processes conference in Vegas and decided to close the gap between theory and practice."

MATHEMATICS – PROBABILITIES & STATISTICS
DEPARTMENT OF MATHEMATICS

Stochastic Processes

NOTEBOOK

Author
Youthy WANG

December 14, 2022

Contents

Chapter 1

Probability Tools – Preliminaries

Here is a review of previous probability and statistic courses. Before learning stochastic processes, it is necessary to recap some principal probability tools.

1.1 The Inclusion-Exclusion Principle

Theorem 1. The Inclusion-Exclusion Principle (De Moivres Formula)

Suppose $A_1, A_2, \cdots, A_n$ are n events ($n \geq 2$), then we have

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \cdots + (-1)^{n+1} P\left(\bigcap_{i=1}^{n} A_i\right).$$

The following example shows the significance of above principle.

Example 1. Hat Problem (Bernoulli's Dislocation Arrangement)

Suppose that $n \geq 2$ people participated in a party. When entering the party, they took off their n hats. The hats are then mixed up and after the party, every man randomly select one. We say that a match occurs if a man selects his own hat.

`Qn`: What is the probability of no match?

`Ans.` Suppose X_i represents that the i^{th} people find his hat. Then, what we want to find is $p = 1 - P\left(\bigcup_{i=1}^{n} X_i\right)$. Apply the inclusion-exclusion principle, because $P\left(\bigcap_{j=1}^{k} X_{i_j}\right) = \dfrac{(n-k)!}{n!}$, $\forall i_1, \cdots, i_k$ (some people find their hats, others randomly). We then get $p = \binom{n}{2}\dfrac{(n-2)!}{n!} + \cdots + (-1)^n \binom{n}{n}\dfrac{1!}{n!} = \displaystyle\sum_{i=2}^{n} \frac{(-1)^i}{i!}$.

Remark. Recall in high school math, we can use the recursion of a sequence to solve the same question.

Another example is the renowned `Coupon Collector's Problem` in combinatorial probability. To solve such a question, we need another lemma.

Lemma

Assume that $X \geq 0$ is a discrete random variable with values in $\mathbb{N}$, then

$$\mathbb{E}(X) = \sum_{i=0}^{\infty} P(X > i).$$

The theorem also holds for continuous case. If $X \geq 0$ is a continuous RV,

$$\mathbb{E}(X) = \int_0^{\infty} P(X > x)dx.$$

Proof. By definition, $\mathbb{E}(X) = \sum_{i=0}^{\infty} iP(X = i) = \sum_{i=1}^{\infty}\sum_{j=i}^{\infty} P(X = j) = \sum_{i=1}^{\infty} P(X \geq i) = \sum_{i=0}^{\infty} P(X > i)$. (By using diagrams/forms, we can see the double sum more directly.)

For continuous case, $\int_0^{\infty} x f_X(x)dx = \int_0^{\infty}\int_0^{x} f_X(x)dtdx = \int_0^{\infty}\int_t^{\infty} f_X(x)dxdt$
$= \int_0^{\infty} P(X > t)dt.$ ■

Example 2. Coupon Collector's Problem

There are m different types of infinitely many coupons. Every time a person collects a coupon, it is independent of ones previously obtained, and a type j coupon with fixed probability p_j, $\sum_{j=1}^{m} p_j = 1$.

`Qn:` What is the expectation of the number of coupons one needs to collect in order to have all types?

`Ans.` Suppose Y_i represents the times one need to wait (i.e., the number of coupons one needs collecting previously) until getting type i coupon. Easily we know that $\{Y_i\}_{i=1}^{m}$ uniquely determines the sequence of coupon collection. What we require is that $\mathbb{E}(N) = \sum_{n=0}^{\infty} P(N > n) = \sum_{n=0}^{\infty} P(\exists\, i, Y_i > n) = \sum_{n=0}^{\infty} P\left[\bigcup_{i=1}^{m}(Y_i > n)\right]$.

By De Moivre's formula above, $P\left[\bigcup_{i=1}^{m}(Y_i > n)\right] = \sum_{i=1}^{m} P(Y_i > n) - \sum_{1\leq i<j\leq m} P(Y_i > n, Y_j > n) + \cdots + (-1)^{m+1}P\left[\bigcap_{i=1}^{m}(Y_i > n)\right] = \sum_{i=1}^{m}(1-p_i)^n - \sum_{1\leq i<j\leq m}(1-p_i-p_j)^n + \cdots$.

Thus, we get $\mathbb{E}(N) = \sum_{n=0}^{\infty} P(N > n) = \sum_{n=0}^{\infty}\left\{\sum_{i=1}^{m}(1-p_i)^n - \sum_{1\leq i<j\leq m}(1-p_i-p_j)^n + \cdots\right\}$
$\xlongequal[\text{geometry series}]{\text{change sum}} \sum_{i=1}^{m}\frac{1}{p_i} - \sum_{1\leq i<j\leq m}\frac{1}{p_i+p_j} + \cdots + (-1)^{m+1}\frac{1}{p_1+\cdots+p_m}.$

Proposition: A corollary from Coupon Collector's Problem

There is a combinatorial identity (given $N \in \mathbb{N}$)

$$\sum_{i=1}^{N} \frac{1}{i} = \sum_{i=1}^{N} \frac{(-1)^{i+1}}{i} \binom{N}{i}.$$

Remark. Three ways to prove this identity. One brutal way is to use `Induction`. Furthermore, applying $\int_0^1 x^i dx = \frac{1}{i+1}$ is another more elegant way. The most strange way is to use $p_i = \frac{1}{m}$ in the above case, and re-consider that question in `Geometric Distribution` way.

1.2 Indicator Functions

Definition

Given $A \subset \mathbb{R}$, we define then an `indicator function` on A, which is

$$\mathbb{1}_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

Indicator functions have may usage for make functions and proofs concise and simple. In the next we can see an example of order statistics which applies it.

Example 3. Order Statistics

Suppose $X_1, \cdots, X_n$ are i.i.d sampled with pdf $f(x)$ and CDF $F(x)$. Consider the ascending sorted rearrangement of such random variables (order statistics), $X_{(1)}, \cdots, X_{(n)}$..

`Qn`: What are the pdf of $X_{(m)}$ and joint pdf of $X_{(1)}, \cdots, X_{(m)}$, $1 \leq m \leq n$?

`Ans`. For pdf of $X_{(m)}$, see in previous STA2002 Notes. For joint pdf, consider joint pdf of $X_1, \cdots, X_m$, which is absolutely $f_{X_1,\cdots,X_m}(x_1, \cdots, x_m) = \prod_{i=1}^{m} f(x_i)$.

Nonetheless, when it comes to pdf of ordered sequences, the support of pdf changes.

Note that *supp* $\left[f_{X_{(1)},\cdots,X_{(m)}}(x_1, \cdots, x_m)\right] = \{(x_1, \cdots, x_m) | x_1 < \cdots < x_m\}$, we get then

$f_{X_{(1)},\cdots,X_{(m)}}(x_1, \cdots, x_m) = \left[n! \prod_{i=1}^{m} f(x_i)\right] \mathbb{1}_{x_1 < \cdots < x_m}$

1.3 Joint Probabilities and Independence

1.3.1 Joint Probabilities

We now discuss in general the transformation technique for the continuous case. Let (X_1, X_2) have a jointly continuous distribution with pdf $f_{X_1,X_2}(x_1, x_2)$ and support set S. Suppose the random variables Y_l and Y_2 are given by $Y_1 = u_1(X_1, X_2)$ and $Y_2 = u_2(X_1, X_2)$, where the functions define a one-to-one transformation that maps the set S in $\mathbb{R}^2$ onto a (two-dimensional) set T in $\mathbb{R}^2$ where T is the support of (Y_1, Y_2).

Then, Y_1 and Y_2 are jointly continuous with joint density function given by

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(u_1(X_1, X_2), u_2(X_1, X_2))|J|, \text{ where } J = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{bmatrix}.$$

More generally, for n variables, we have the same result.

1.3.2 Independent Random Variables

Theorem 2. Independence from pdf

Let the random variables X_1 and X_2 have supports S_1 and S_1, respectively, and have the joint pdf $f(x_1, x_2)$. Then X_1 and X_2 are independent ***if and only if*** $f(x_1, x_2)$ can be written as a product of a nonnegative function of X_1 and a nonnegative function of X_2 That is,

$$f(x_1, x_2) \equiv g(x_1)h(x_2),$$

where $g(x_1) > 0$, for $x_1 \in S_1$, and zero elsewhere, $h(x_2) > 0$, for $x_2 \in S_2$, and zero elsewhere.

(Proof is omitted here. Sketch: use the definition of independence.)

Apart from that, a vital investigation is that uncorrelated Rvs cannot imply independent RVs, even with assumption of normality.

Here is a counter-example.

Example 4. Uncorrelated Dependent Normal Random Variables

Let $X \sim N(0, 1)$, $\varepsilon \in \{\pm 1\}$ with probabilities $\frac{1}{2}$, respectively. $\varepsilon \perp\!\!\!\perp X$. Let $Y = \varepsilon X$.

Then, we can easily verify the conclusions below.

1. $Y \sim N(0, 1)$ by investigate CDF of Y.
2. $\mathbb{E}(XY) = \mathbb{E}(\varepsilon X^2) = \mathbb{E}(\varepsilon)\mathbb{E}(X^2) = 0$. (Imply Uncorrelated Rvs).
3. $\mathbb{E}(X^2Y^2) = \mathbb{E}(\varepsilon^2 X^4) = 3 \neq 1 = \mathbb{E}(Y^2)\mathbb{E}(X^2)$ (Dependence).

1.4 Introduction to Stochastic Processes

Here comes the definition of stochastic processes.

Definition: Stochastic Processes

A stochastic process $\{X_{(t)} : t \in T\}$ is a **collection of random variables**, i.e. for each $t \in T$, $X_{(t)}$ is a random variable.

The index t is often interpreted as time, so $X_{(t)}$ is called the *state* of the process at time t. The set T is called the *index* set of the process.

When T is a countable set, the stochastic process is said to be a *discrete-time* process. If T is an interval of the real line, the stochastic process is said to be a *continuous-time* process.

The *state space* of a stochastic process is the set of all possible values such that the random variables $X_{(t)}$ can assume.

There are four kinds of stochastic processes.

State Space\Time Parameter	Discrete	Continuous
Discrete	random walk; Markov chains	Poisson processes; Markov chains
Continuous	time series	Brownian motion; Stock prices;

Table 1.1: 4 types of stochastic processes

Theorem 3. Stirling's Formula

$$n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n},\, n \to \infty$$

One way to yield the above formula is to use the Poisson distribution and CLT, which is $e^{-n}\frac{n^n}{n!} = P(S_n = n) = P(n-1 < S_n \leq n) = P(-\frac{1}{\sqrt{n}} < \sqrt{n}(\overline{X}-1) \leq 0) \approx \int_{-1/\sqrt{n}}^{0} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \approx \frac{1}{\sqrt{2\pi n}}$, with $S_n = \sum_{i=1}^{n} X_i$ and $X_i \sim Pois(1)$.

1.4.1 Conditional Distributions

Recall the **theorem of iterated expectation and iterated variance** in STA2002 Notes:

$$\begin{cases} \mathbb{E}(X) = \mathbb{E}(X|Y) \\ \text{Var}(X) = \text{Var}(\mathbb{E}(X|Y)) + \mathbb{E}(\text{Var}(X|Y)) \end{cases}.$$

By finding a suitable instrumental random variable Y for which the conditional expectation $\mathbb{E}(X|Y)$ is available, then $\mathbb{E}(X)$ is found through the conditioning formula. Below are some applications by applying the theorems.

Example 5. Expectation of a Random Sum

Let $\{X_{(i)} : i = 1, 2, 3, \cdots\}$ be i.i.d.(independent identically distributed) r.v.s with common mean μ and common variance σ^2. Let $S_N = \sum_{i=1}^{N} X_{(i)}$, where N is a random integer and independent of X's.

`Qn`: Calculate $\mathbb{E}(S_N)$.

`Ans.` $\mathbb{E}(S_N) = \mathbb{E}(\mathbb{E}(S_N|N)) = \mathbb{E}(\mu N) = \mu\mathbb{E}(N)$, since $\mathbb{E}(S_N|N = n) = n\mathbb{E}(X_1) = n\mu$.

Another application of conditional distribution is to calculate probabilities by conditioning.

Lemma

Let A denote an arbitrary event and define the indicator random variable $\mathbb{1}_A$ as above shows(1.2). Then,

$$P(A) = \mathbb{E}(\mathbb{1}_A) = \mathbb{E}[\mathbb{E}(\mathbb{1}_A|Y)] = \begin{cases} \sum_y P(A|Y = y)P(Y = y) & \text{discrete} \\ \int_{\mathbb{R}} P(A|Y = y)f_Y(y)dy & \text{continuous} \end{cases}$$

1.4.2 Conditioning a Random Variable on an Event

Let X be a continuous random variable and A be an event with $P(A) > 0$, where A is an event related to X (e.g., $X \geq 1$, otherwise, it degenerates to trivial cases.).

Then the **conditional pdf** of X given A is defined as the nonnegative function $f_{X|A}$ that obeys

$$P(B|A) = \int_B f_{X|A}(x)dx, \text{ for all events } B \text{ defined by } X.$$

A special (and very important) case is when the event A explicitly involves the random variable X itself:

$$P(B|A) = \frac{P(X \in B, X \in A)}{P(X \in A)} = \frac{\int_{A\cap B} f_X(x)dx}{\int_A f_X(x)dx},$$

in which case

$$f_{X|A}(x) = \frac{f_X(x)}{P(A)} \cdot \mathbb{1}_{\{A\}},$$

which is simply a **rescaling** of the pdf of X over the set A.

Naturally, we have: if A is an event with $P(A) > 0$, then

$$\mathbb{E}[X|A] = \int_\Omega x f_{X|A}(x)dx.$$

Partition of A: it $\{A_j\}_{j=1}^N$ is a *partition* of A (explicitly involves the random variable X itself), we have

$$\mathbb{E}[X|A]P(A) = \sum_{j=1}^{N} \mathbb{E}[X|A_j]P(A_j),$$

which is also called **total expectation theorem**.

Example 6

Let $X \sim exp(\lambda)$. Find $\mathbb{E}(X|1 < X < 2)$.

`Ans`. By definition, $\mathbb{E}(X|1 < X < 2) = \dfrac{1}{P(1 < X < 2)} \displaystyle\int_0^\infty x\lambda e^{-\lambda x} \mathbb{1}_{\{1<x<2\}} dx$
$= \dfrac{\int_1^2 x\lambda e^{-\lambda x} dx}{\int_1^2 \lambda e^{-\lambda x} dx} = \lambda^{-1} + 1 - (e^\lambda - 1)^{-1}$.

The next example is much more complex, with the application in Poisson processes.

Example 7. Poisson Processes

People arrive at a sport event according to a Poisson process with rate $\lambda > 0$ per hour. Percent p of attendants are male and $q = 1 - p$ of them are female. During the first hour of the event, n people arrived ($n \geq 1$). Let Z be the arrival time of the first male person among these n people.

`Qn`: Calculate $\mathbb{E}(Z|A)$ where A is the event that there is at least one male among the n arrived people.

`Ans`. (Omitted here. Sketch: use properties of arrival time S_n and total expectation theorem.)

Remark. Note that in this sense, A does NOT explicitly involve Z. Nonetheless, we can still use the total expectation theorem. Since by definition of $f_{X|A}$, we can directly get $\displaystyle\int_B \sum_{j=1}^N P(A_j) f_{X|A_j}(x) dx = \int_B P(A) f_{X|A}(x) dx$, for all events B defined by X. Thus, total expectation theorem always holds.

Chapter 2

Discrete Time Markov Chains

2.1 Motivation and Introduction

We start with some instances. Consider a wanderer or drunkard and an endless street divided into blocks. In each of unit of time, he walks one block from street corner to corner, and each corner he may choose to go ahead with probability p or turn back with probability $1-p$.

Here is a mathematical representation of the random walk: let X_0 be the starting point and Y_n be the n^{th} step taken: then after n units of time, $X_n = X_0 + \sum_{i=1}^{n} Y_i$.

Here, we have a *discrete-time* stochastic process $\{X_n\}_{n\geq 0}$ with discrete state space $E = \{0, \pm 1, \pm 2, \cdots, \pm i, \cdots\}$.

If we then consider $P\left(X_{n+1}|X_1, \cdots, X_n\right)$, it is not hard to find that $P\left(X_{n+1}|X_1, \cdots, X_n\right) = P\left(X_{n+1}|X_n\right)$, which means $X_1, \cdots, X_{n-1}$ does not count in conditioning X_{n+1}, compared with X_n. This kind of property is called the `Markov Property`.

Definition: Markov Chain and Markov Property

A discrete state space stochastic process $\{X_n\}_{n\geq 0}$ is a `Markov Chain (MC)` if

$P\left(X_{n+1} = j|X_1 = i_1, \cdots, X_n = i_n\right) = P\left(X_{n+1} = j|X_n = i_n\right)$,

for all states $i_0, i_1, \cdots, i_n, j$ and all $n \geq 0$.

This property is referred as the `Markov Property`.

Remark. `Homogeneous` Markov Chain: $P\left(X_{n+1} = j|X_n = i\right) = p_{ij}$, independent of the time n.

2.1.1 Why We Want to Study Markov Chain?

Consider random work again. Set $X_0 = 0$. We want to know these questions below:

- Can this drunkard come back in future? (i.e., what is $P(\exists\ n \geq 1, \text{such that } S_n = 0)$?)

`Ans.` Only when $p = q = \dfrac{1}{2}$, he can for sure come back. (Details in 2.3)

- If the above probability is 1, what is the expected waiting time? (i.e., what is $\mathbb{E}(n)$?)

  `Ans.` $\mathbb{E}(n) = \infty$.

Another example is the *Ehrenfest model of diffusion*. It is a model proposed by the physicists P. and T. Ehrenfest, to describe the diffusion of molecules in a microscopic volume. Suppose that M molecules are distributed in a volume made with two equal rooms; and at each time point one of the molecules randomly decides to change its room.

We want to consider the number X_n at time point n of molecules in rooms.

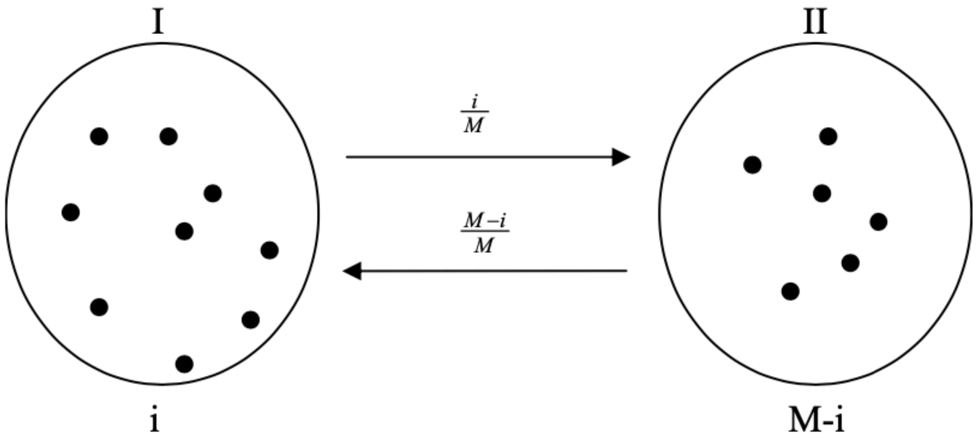


Figure 2.1: Illustration of Ehrenfest Model of Diffusion

A third example is the *highway problem*. Suppose on an infinitely long highway, there are two types of cars, say T and C. Now the information we get is: after one T which arrives, the next has probability 40% to be T and 60% to be C, and after one C which arrives, the next has probability 70% to be C and 30% to be T.

The question here is to calculate the percentage of two cars on the highway.

To get the solution for these questions, it is of significance to introduce and study a tool called **Markov Chains (MC)**.

2.1.2 Transition Probabilities

One-step transition probabilities: $p_{ij} = P\left(X_{n+1} = j | X_n = i\right)$. From this we can get the transition matrix.

For any i, the family $p_{i\cdot} = \{j \to p_{ij}\}$ is the conditional distribution of X_{n+1} given $X_n = i$; in particular $p_{ij} \geq 0$, and for all j, $\sum_{j=1}^{\infty} p_{ij} = 1$.

We can then get one-step transition matrix of the Markov Chain to be $\mathbf{P} = (p_{ij})$. Below are two examples for illustration.

Example 8. Transform a Process into a Markov Chain

Suppose that whether or not it rains today depends on previous weather conditions through the last two days. Specifically, suppose that if it

i. has rained for the past two days.
ii. rained today but not yesterday.
iii. rained yesterday but not today.
iv. has not rained in the past two days.

then it will rain tomorrow with probability $0.7; 0.5; 0.4; 0.2$, respectively

It is NOT a MC on the first sight, but by packaging two days into one new variable, namely $Y_n = \begin{pmatrix} X_{n-1} \\ X_n \end{pmatrix}$. We can get a four-state Markov chain with a transition probability matrix (consdier the combined new RV with $2 \times 2 = 4$ cases), $\mathbf{P} = \begin{bmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{bmatrix}$.

Example 9. A Gambling Model / Finite Random Walk Model

At each play of game, the gambler either wins \$1 with probability p, or loses \$1 with probability $1 - p$. The gambler quits playing either when he goes broke, or when he attains a fortune of \$N.

It is easy to see that the gambler's fortune is a Markov Chain having transition probabilities and matrix $\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 1-p & 0 & p & 0 & 0 & \cdots & 0 \\ 0 & 1-p & 0 & p & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

States 0 and N are called `absorbing states`, since once entered they are never left.

This is a random walk on a finite state space $\{0, 1, ..., N\}$ with absorbing barriers 0 and N.

2.1.3 Important Properties

Here are two vital properties, in respect to joint probabilities and conditional probabilities for MC, respectively.

Lemma: Joint Probabilities for MC

For a MC with transition probabilities $\{p_{ij}\}$,

$P(X_0 = i_0, X_1 = i_1, \cdots, X_n = i_n) = P(X_0 = i_0) \cdot p_{i_0 i_1} \cdots p_{i_{n-1} i_n}$, for $n \geq 1$ and $(i_0, i_1, \cdots, i_n) \in E_{n+1}$.

Proof. By rearranging the order, we have

$P(X_0 = i_0, X_1 = i_1, \cdots, X_n = i_n) = P(X_n = i_n, \cdots, X_1 = i_1, X_0 = i_0)$
$= P(X_n = i_n | X_{n-1} = i_{n-1} \cdots, X_1 = i_1, X_0 = i_0) \cdot P(X_{n-1} = i_{n-1} \cdots, X_1 = i_1, X_0 = i_0)$
$\xlongequal{\text{MC Property}} p_{i_{n-1}i_n} \cdot P(X_{n-1} = i_{n-1} \cdots, X_1 = i_1, X_0 = i_0)$. By iterations, we can get the conclusion. ■

Lemma: Conditional Probabilities for MC

For a MC $\{X_n\}_{n\geq 0}$, and $n > k_1 > \cdots > k_m \geq 0$,

$P(X_n = i_n | X_{k_1} = i_{k_1}, \cdots, X_{k_m} = i_{k_m}) = P(X_n = i_n | X_{k_1} = i_{k_1})$.

Proof. From our intuition, MC tells us that the latest probability matters, which is for sure this property.

A more rigorous way is to add back all middle terms, suppose $n-1 \geq \cdots \geq k_1 + 1$ are totally $l \geq 1$ numbers, with new sign $j_1 \leq \cdots \leq j_l$ we have

$P(X_n = i_n | X_{k_1} = i_{k_1}, \cdots, X_{k_m} = i_{k_m}) = \dfrac{P(X_n = i_n, X_{k_1} = i_{k_1}, \cdots, X_{k_m} = i_{k_m})}{P(X_{k_1} = i_{k_1}, \cdots, X_{k_m} = i_{k_m})}$
$= \dfrac{\sum_{(s_1,\cdots,s_l)\in E_{n+1}} P(X_n = i_n, X_{j_l} = s_l, \cdots, X_{j_1} = s_1, \cdots X_{k_1} = i_{k_1}, \cdots, X_{k_m} = i_{k_m})}{P(X_{k_1} = i_{k_1}, \cdots, X_{k_m} = i_{k_m})}$
$= \sum_{(s_1,\cdots,s_l)\in E_{n+1}} P(X_n = i_n | X_{j_l} = s_l) \cdots P(X_{j_1} = s_1 | X_{k_1} = |i_{k_1})$
$= \dfrac{\sum_{(s_1,\cdots,s_l)\in E_{n+1}} P(X_n = i_n | X_{j_l} = s_l) \cdots P(X_{j_1} = s_1 | X_{k_1} = |i_{k_1}) P(X_{k_1} = |i_{k_1})}{P(X_{k_1} = |i_{k_1})}$
$= \dfrac{\sum_{(s_1,\cdots,s_l)\in E_{n+1}} P(X_n = i_n, X_{j_l} = s_l, \cdots, X_{j_1} = s_1, \cdots X_{k_1} = i_{k_1})}{P(X_{k_1} = |i_{k_1})}$
$= \dfrac{\sum_{(s_1,\cdots,s_l)\in E_{n+1}} P(X_n = i_n, X_{k_1} = i_{k_1})}{P(X_{k_1} = |i_{k_1})} = P(X_n = i_n | X_{k_1} = i_{k_1})$. ■

2.2 Chapman-Kolmogorov Equation

Definition: n-step Transition Probability

Define $P_{ij}^{(n)} = P(X_{n+m} = j | X_m = i)$, $n \geq 1$ to be the n-step transition probability of MC $\{X_n\}_{n\geq 0}$

`Qn`: Why is $p_{ij}^{(n)}$ independent of n?

`Ans`. $P(X_{m+n} = j | X_m = i) = \dfrac{P(X_{m+n} = j, X_m = i)}{P(X_m = i)}$
$= \dfrac{\sum\limits_{\mathbf{k}\in E^{n-1}} P(X_{m+n} = j, X_{m+n-1} = k_{n-1}, \cdots X_{m+1} = k_1, X_m = i)}{P(X_m = i)}$
$\xlongequal{\text{Property I}} \sum\limits_{\mathbf{k}\in E^{n-1}} p_{ik_1} p_{k_1k_2} \cdots p_{k_{n-1}j} = (P^n)_{ij}$, where P is the one step transition probability matrix.

Proposition

The n-step transition probabilities are $p_{ij}^{(n)} = (P^n)_{ij}$.

From above definition and proofs, we can get the famous `Chapman-Kolmogorov Equation`.

Theorem 4. Chapman-Kolmogorov Equation

Suppose $p_{ij}^{(n)}$ is defined as above, then we have that

$$p_{ij}^{(n+m)} = \sum_{k \in E} p_{ik}^{(m)} p_{kj}^{(n)}, \ \forall n, m \geq 1 \text{ and } i, j \in E.$$

If matrix form is applied, the equation becomes

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

Proof. By the Ans. for above explanation of independence, we get directly $P^{(m)} = P^m$, $\forall m > 0$. ■

2.2.1 Probability Distribution of X_n at time n

Note that distribution of MC is determined by

(i) *initial distribution* of X_0,

(ii) *transition probability matrix* P.

Define $E = \{e_1, e_2, \cdots\}$, and we then give a notation about *pmf* of X_n to be $\mu_n = (P(X_n = e_1), P(X_n = e_2), \cdots)^T$ with the for as a row vector.

`How?`: Distribution function of X_n is $P(X_n = j) = \sum_{\mathbf{k}} P(X_n = j, X_{(n-1)} = k_{n-1}, \cdots, X_1 = k_1, X_0 = k_0) = \sum_{\mathbf{k}} P(X_0 = k_0) p_{k_0 k_1} \cdots p_{k_{n-1} j}$. Particularly, take $n = 1$, $P(X_1 = j) = \sum_i P(X_1 = j, X_0 = i) = \sum_i P(X_0 = i) p_{ij}$. That is $\mu_1 = \mu_0 P$. Similarly we can get $\mu_{n+1} = \mu_n P$.

Proposition: Distribution of X_n

Let μ_n be defined above, we then have $\mu_n = \mu_{n-k} P^k$, with any $0 \leq k \leq n$.

2.3 Classification of States

State i **leads** to j, denoted as $i \to j$, if $P_{ij}^{(n)} > 0$ for some $n \geq 1$. State i and j **communicate each other**, denoted as $i \leftrightarrow j$, if $i \to j$ and $j \to i$.

`Transitivity`: $i \to j$ and $j \to k$ give $i \to k$. $i \leftrightarrow j$ and $j \leftrightarrow k$ give $i \leftrightarrow k$.

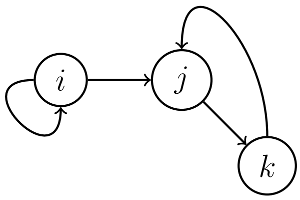
The state space can then be divided into distinct **communication classes**: each class contain states which communicate each other.

A Markov chain is said to be **irreducible** if all states communicate with each other (only one such class!)

2.3.1 Markov Graph

A directed graph (*multi-graph*) $G(V, E)$ is called `Markov Graph` if V is a collection of all states ($E = \{i, j, k, \cdots\}$) and all edges in E satisfy $(i, j) = i \to j$ iff $p_{ij} > 0$.

Below is an illustration of Markov graphs.



2.3.2 Recurrence and Transience

In this section, we try to classify states into two categories: recurrence and transience.

Definition: The First Hitting Time

Let state $i \in E$. The first **hitting time** to i is defined as

$$\tau_i = \inf\ \{n \geq 1 : X_n = i\}.$$

Definition: Recurrence and Transience

Let $f_i = \mathbb{P}(\tau_i < \infty | X_0 = i)$, meaning that the probability of re-enter state i starting from i and $f_{ii}^{(k)} = \mathbb{P}(\tau_i = k | X_0 = i)$, starting at state i, the first visit to state i happens at time k.

- State i is named **recurrent** if $f_i = 1$, namely, the process will definitely re-enter state i.
- State i is named **transient** if $f_i < 1$, i.e. if starting in state i, there is a positive probability that the process will not re-enter state i.

Theorem 5

State i is **recurrent** *exactly when* $\sum_{n=1}^{\infty} P_{ii}^{(n)} = \infty$.

Proposition

Below are two propositions necessary for this proof.

1. Following definitions above, we have $P_{ii}^{(n)} = \sum\limits_{k=1}^{n} f_{ii}^{(k)} P_{ii}^{(n-k)}$, $\forall n \geq 1$.
2. Define generating function: $U(\zeta) = \sum\limits_{n=0}^{\infty} P_{ii}^{(n)} \zeta^n$ and $F(\zeta) = \sum\limits_{n=1}^{\infty} f_{ii}^{(n)} \zeta^n$. (Note that $F(1) = f_i$).

   Under this definition, we have $U(\zeta) - 1 = F(\zeta)U(\zeta)$, $\forall\ 0 \leq \zeta < 1$.

Proof of Propositions. For proposition 1, $P_{ii}^{(n)} = P(X_n = i|X_0 = i) = \sum\limits_{k=1}^{n} P(X_n = i, \tau_i = k|X_0 = i) = \sum\limits_{k=1}^{n} P(X_n = i, X_k = i, X_{k-1} \neq i, \cdots, X_1 \neq i|X_0 = i) = \sum\limits_{k=1}^{n} P(X_n = i|X_k = i, X_{k-1} \neq i, \cdots, X_1 \neq i, X_0 = i)P(X_k = i, X_{k-1} \neq i, \cdots, X_1 \neq i|X_0 = i) = \sum\limits_{k=1}^{n} P(X_n = i|X_k = i) f_{ii}^{(k)}$.

For proposition 2, $U(\zeta) = \sum\limits_{n=0}^{\infty} P_{ii}^{(n)} \zeta^n \overset{\text{prop 1}}{=\!=\!=} 1 + \sum\limits_{n=1}^{\infty} \left(\sum\limits_{k=1}^{n} P_{ii}^{(n-k)} f_{ii}^{(k)} \right) \zeta^n = 1 + \sum\limits_{n=1}^{\infty} \sum\limits_{k=1}^{n} P_{ii}^{(n-k)} \zeta^{n-k} f_{ii}^{(k)} \zeta^k = 1 + \sum\limits_{k=1}^{\infty} \sum\limits_{n=k}^{\infty} P_{ii}^{(n-k)} \zeta^{n-k} f_{ii}^{(k)} \zeta^k = 1 + \sum\limits_{k=1}^{\infty} f_{ii}^{(k)} \zeta^k \sum\limits_{m=0}^{\infty} P_{ii}^{(m)} \zeta^m = 1 + F(\zeta)U(\zeta)$. ■

Proof of Theorem 5. $\lim\limits_{\zeta \to 1^-} U(\zeta) = \sum_{n=0}^{\infty} P_{ii}^{(n)} = U(1)$. Let $F(1) = f_i < 1$, then, from proposition 2, $\lim\limits_{\zeta \to 1^-} U(\zeta) = \dfrac{1}{1 - f_i} < \infty$ (transient i). Meantime, if $F(1) = f_i = 1$, then $\lim\limits_{\zeta \to 1^-} U(\zeta) = \dfrac{1}{1 - f_i} = \infty$ (recurrent i). ■

Theorem 6. Resolvent Equation

For $j, k \in E$, let $f_{jk}^{(n)} = P\{\tau_k = n|X_0 = j\}$, that is the probability that the chain visits state k for the first time at time n when started at state j. Apart from that, $\sum\limits_{n=1}^{\infty} f_{ij}^{(n)} = f_{ij}$. Then, we have

1. Recursion Formula $P_{jk}^{(n)} = \sum\limits_{k=1}^{n} f_{jk}^{(k)} P_{kk}^{(n-k)}$, $n \geq 1$.
2. Define the generating functions for $\zeta \geq 0$, $U_{jk}(\zeta) = \sum\limits_{n=0}^{\infty} P_{jk}^{(n)} \zeta^n$ and $F_{jk}(\zeta) = \sum\limits_{n=1}^{\infty} f_{jk}^{(n)} \zeta^n$.

   We have for $j, k \in E$, $U_{jk}(\zeta) - \mathbb{1}_{\{j=k\}} = F_{jk}(\zeta)U_{kk}(\zeta)$, $\forall\ 0 \leq \zeta < 1$.

Proofs for this theorem is similar to those above. Note that if $j \neq k$, $P_{jk}^{(0)} = 0$, the item starts from $P_{jk}^{(1)}s$.

Lemma

If state k is transient, $\forall\ j \in E$, $\lim_{n\to\infty} P_{jk}^{(n)} = 0$.

Proof. Since k is transient $\Leftrightarrow U_{kk}(1) < \infty$, and $F_{jk}(1) \leq 1$, we have $U_{jk}(1) = F_{jk}(1)U_{kk}(1) < \infty$. Namely, $\sum_{n=0}^{\infty} P_{jk}^{(n)}$ converges. ∎

A vital property for transience and recurrence is that they are **class properties**: states in the same class are all recurrent or all transient.

Theorem 7. Class Properties I

I State i is recurrent and $i \leftrightarrow j \Rightarrow$ State j is recurrent.

II State i is transient and $i \leftrightarrow j \Rightarrow$ State j is transient.

Proof. Suppose $i \leftrightarrow j$, $\exists\ k, m \in \mathbb{N}$ such that $P_{ij}^{(k)} > 0$ and $P_{ji}^{(m)} > 0$ by definition.

Since $\sum_{s=1}^{\infty} P_{ii}^{(s)} \geq \sum_{n=1}^{\infty} P_{ii}^{(n+k+m)} \geq \sum_{n=1}^{\infty} P_{jj}^{(n)} P_{ij}^{(k)} P_{ji}^{(m)} = P_{ij}^{(k)} P_{ji}^{(m)} \sum_{n=1}^{\infty} P_{jj}^{(n)}$. (Use the fact that the whole is greater than some part.) ∎

It is a very useful property since for many states whose transience and recurrence are difficult to analyze, we can alternatively analyze some simple states staying in the same class with them.

Example 10. Random Walk

Recall at the beginning of this chapter, a random walk example is released for motivating to study MC. One question is to ask whether the drunkard will for sure be back for some positions, which is the question about transience and recurrence.

`Ans`. Since all states clearly communicate (easy to verify), they are either all transient or all recurrent.

Now concentrate on 0 state and $\sum_{n=1}^{\infty} P_{00}^{(n)} = \sum_{n=1}^{\infty} \binom{2n}{n} p^n q^n \sim \sum_{n=1}^{\infty} \frac{[4p(1-p)]^n}{\sqrt{n\pi}}$.By Stirling's formula (3).

If $p = \frac{1}{2}$, $4p(1-p)$ attains its maximum 1 and $\sum_{n=1}^{\infty} \frac{[4p(1-p)]^n}{\sqrt{n\pi}} = \sum_{n=1}^{\infty} \frac{1}{\sqrt{n\pi}}$ diverges to ∞. By theorem above (5), all points are recurrent.

Otherwise $p \neq \frac{1}{2}$, $4p(1-p) < 1$ and $\sum_{n=1}^{\infty} \frac{[4p(1-p)]^n}{\sqrt{n\pi}} < \sum_{n=1}^{\infty} [4p(1-p)]^n < \infty$. All points are transient.

Remark. (i) We can also apply *limit ratio test* instead of invoking Stirling's formula.

(ii) Moreover, the symmetric random walk has all its states *null recurrent*. (Explanations: $f_{00}^{(2n)} = 2C_{n-1}p^n q^n$, where C_k is the k^{th} **Catalan number**. Combinatorial thought: arrange "1" and "-1", where the first k sum $s_k > 0$ or $s_k < 0$ for all k except for the last s_{2n}. Due to symmetry, choose all > 0 and let "-1" be in the last position. Then, we only arrange $(n-1)$ # "1" and "-1" to make every $s_k \geq 0$, totally C_{n-1} cases, and put the additional "1" before the position j which *firstly* makes $s_j = 0$. Test that it is a 1-to-1 mapping.)

Proposition

A finite state-space Markov chain has *at least one* recurrent state.

Proof. Argue by contradicts (reductio ad absurdum). Suppose no recurrent state for that chain, by lemma above (2.3.2), $1 = \sum_{j\in E} P_{ij}^{(n)} = \lim_{n\to\infty} \sum_{j\in E} P_{ij}^{(n)} = \sum_{j\in E} \lim_{n\to\infty} P_{ij}^{(n)} = 0$. Contradictory! ■

Remark. (i) All states in a **finite irreducible** Markov Chain are recurrent.

(ii) For infinite case, change of countable sum and limit cannot be always allowed. The proof in that case fail. As a counter-example, random walk with $p \neq \frac{1}{2}$ has all states transient.

2.4 Limiting Probabilities

The third example at the beginning of this chapter serves as an illustration for motivation of limiting probabilities.

Firstly, we need to know what is an *invariant (stationary) probability distribution.*

Definition: Invariant Probability Distribution

Any probability $\mu = (\mu_0, \mu_1, \cdots)$ on the state space is an invariant probability distribution, or a stationary probability distribution if $\mu = \mu P$.

Invariant (stationary) probability distribution is a special kind of distribution. Meanwhile, limiting probability is even more special since it is a case of stationary probability distributions (See below (8)).

Definition: Limiting Probability Distribution

A Markov chain is said to possess **limiting probabilities** $\{\pi_j\}$ if for all i, $P_{ij}^{(n)} \to \pi_j$ as $n \to \infty$.

We take interests in limiting probabilities since it has some wonderful properties deserving to be studied.

Theorem 8. Properties of Limiting Probabilities

Assume $\forall\, j \in E$, $P_{ij}^{(n)} \to \pi_j$ as $n \to \infty$, that is, limiting probabilities exist. The conclusions are

(I) Let $\pi = (\pi_j)$. Then, $\pi = \pi P$, namely, π is an **invariant distribution**.

(II) the MC **converges in distribution** to π, that is, for all j, $P(X_n = j) \to \pi_j$ as $n \to \infty$.

(III) The expected value of the **long-run proportions** of the MC converge also to π, namely, $\mathbb{E}\left(\frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\{X_k=j\}}\right) \to \pi_j$ as $n \to \infty$.

Proof. (I) For any state $j \in E$, $0 \le P_{ij}^{(n+1)} = \sum_{i\in E} P_{ij}^{(n)} P_{ij} \le 1$, thus, $\pi_j = \lim_{n\to\infty} P_{ij}^{(n+1)} = \lim_{n\to\infty} \sum_{k\in E} P_{ik}^{(n)} P_{kj} \xlongequal{\text{Lebesgue DCT}} \sum_{k\in E} \lim_{n\to\infty} P_{ik}^{(n)} P_{kj} = \sum_{i\in E} P_{kj}\pi_k$.

(II) $P(X_n = j) = \mu_n(j) = \sum_{k\in E} \mu_0(k) P_{kj}^{(n)} \le 1$, hence, when $n \to \infty$, RHS $\to \pi_j \sum_{k\in E} \mu_0(k) = \pi_j$.

(III) (*Cesaro Means*, proof from some elementary analysis techniques.) If $a_n \to a$, partial sum $\frac{a_1 + \cdots + a_n}{n} \to a$ as $n \to \infty$. Therefore, $\frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left(\mathbb{1}_{\{X_k=j\}}\right) = \frac{1}{n}\sum_{k=1}^{n} P(X_k = j) \to \pi_j$

■

Back to the previous example, the percentage of car C is $\pi_C = \frac{2}{3}$ (solutions from $\pi = \pi P$ and $\pi\vec{1} = 1$).

2.4.1 Conditions for a Markov Chain with Limiting Probabilities

1. Significance of `aperiodicity`: Consider a simple Markov chain with two states and transition matrix $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, we have $P^{2k+1} = P$ and $P^{2k} = I_2$, no way to have limiting probabilities;
2. Significance of `irreducibility`: Consider the Markov chain having states $0, 1, 2, 3$

and transition matrix $P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$, we have $P^n = P$, no way to have limiting probabilities since it depends on i.

2.4.2 Period of a State

Now we formally define the period of a state to check whether it is aperiodic.

Definition: Periodicity

(I) For all state i of a MC, $d(i) =$ period of state $i = g.c.d.\ \{n \in \{1, 2, \cdots\} : P_{ii}^{(n)} > 0\}$, here g.c.d. means the *greatest common divisor*.

(II) By convention, $d(i) = \infty$ when $P_{ii}^{(n)} = 0$ for all $n \geq 1$.

(III) A state i is said **aperiodic** if $d(i) = 1$.

Proposition: Class Properties II

Periodicity is also a class property: if state i has period d and $i \leftrightarrow j$, then state j also has period d.

2.4.3 Positive and Null Recurrence

Recall time of the first return to state i.

For a recurrent state i, $P(\tau_i < \infty | X_0 = i) = f_i = \sum_{n=1}^{\infty} f_{ii}^{(k)} = 1$, is a well-defined proper p.m.f. of $\mathcal{L}(\tau_i | X_0 = i)$.

Let $m_i = \mathbb{E}[\tau_i | X_0 = i] = \sum_{i=1}^{\infty} k f_{ii}^{(k)}$ be the mean recurrent time of state i. Then, there are two cases for m_i:

(i) If $m_i < \infty$, state i is said to be **positive recurrent**.

(ii) If $m_i = \infty$, state i is said to be **null recurrent**.

Remark (Class Properties III). Admittedly, positive/null recurrence are also class properties.

Note that *positive recurrent and aperiodic* states are called **ergodic**.

2.4.4 Basic Limit Theorem

The Basic Limit Theorem determine (almost) all the possible limits for the n-step probabilities $\{P_{ij}^{(n)}\}$ of a Markov chain.

Theorem 9. Basic Limit Theorem

(I) For a *transient* or *null recurrent* state j,

$$\lim_{n\to\infty} P_{ij}^{(n)} = 0, \text{ for all } i.$$

(II) For a *positive recurrent* state j with mean first recurrent m_j and peoriod $d \geq 1$,

$$\lim_{n\to\infty} P_{jj}^{(dn)} = \frac{d}{m_j}.$$

(III) For an *ergodic* state j with mean first recurrent m_j,

$$\lim_{n\to\infty} P_{ij}^{(n)} = \frac{f_{ij}}{m_j}, \text{ for all } i.$$

Proof. To prove this result, we firstly introduce several lemmas.

Lemma: (*I*)

Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of non-negative numbers, with $\sum\limits_{i=1}^{\infty} f_i = 1$ and $g.c.d \ \{n|f_n > 0\} = 1$. Define $\{u_m\}_{m=0}^{\infty}$ as: $u_m = \sum\limits_{i=1}^{m} f_i u_{m-i}$, $m \geq 1$ and $u_0 = 0$.

The conclusion is, $u_m \to \mu^{-1}$ as $m \to \infty$, where $\mu = \sum\limits_{k=1}^{\infty} k f_k$.

(*Proof of Lemma I.*) Firstly, let $A = \{n|f_n > 0\}$, and A^+ denote all positive linear combinations of elements in A (i.e., linear combinations with positive weights). Then, $\exists\ N \in \mathbb{N}$, such that $\{n \in \mathbb{N}|n > N\} \subset A^+$.

Since by *Bezout's lemma*, $\exists\ a_1, \cdots, a_r \in A, c_1, \cdots, c_r \in \mathbb{Z}$ such that $\sum\limits_{i=1}^{r} c_i a_i = 1$. Let $s = \sum\limits_{i=1}^{r} a_i$, we can choose large enough n, such that $n = xs + y$ (division with remainder, thus $y < s$), $x > \max\limits_{1\leq k\leq r}(|c_k|)y$, then $n = \sum\limits_{i=1}^{r}(x + c_k y)a_k \in A^+$.

Let $\eta = \limsup\limits_{k\to\infty} u_k$, then, since for every $k \in \mathbb{N}$, $0 \leq u_k \leq 1$, we have $\eta \leq 1$. Apart from that, $\exists\ b_1, b_2, \cdots$, such that $u_{b_k} \to \eta$. Construct doubly infinite sequence $u_n^{(k)} = \begin{cases} u_{n+b_k} & n \geq -b_k \\ 0, & n < -b_k \end{cases}$ for every $k \in \mathbb{N}$. Thus, by *selection theorem*, $u_n^{(k)}$ converges to w_n for any n, as $k \to \infty$.

Next, we prove that $w_n = \eta$, $\forall\ n \in \mathbb{Z}$. Since $0 \leq w_n \leq \eta$, $w_0 = \eta$, and also we have by definition, $u_n^{(k)} = \sum\limits_{m=1}^{\infty} f_m u_{n-m}^{(k)}$ for any n, k. Then we can prove that any $w_n = \sum\limits_{m=1}^{\infty} f_m w_{n-m}$,

$\forall\ n$ by Lebesgue D.C.T. Since $\eta = w_0 = \sum_{m=1}^{\infty} f_m w_{n-m} \leq \sum_{m=1}^{\infty} f_m \eta \leq \eta$, we know $\forall\ n \in A$, $w_{-n} = \eta$. Further, $\forall\ n \in A^+$, $w_{-n} = \eta$. It means that $w_{-n} = \eta$ for all $n > N$ with some N. Thus, $w_{-N} = \eta$ using the original equation, thus $w_{-N+1} = \eta$. Sequentially, we get $w_n = \eta$ for all $n \in \mathbb{Z}$.

Let $\rho_k = \sum_{n=k+1}^{\infty} f_n$, then $\rho_0 = 1$. Further, $1 = \sum_{m=0}^{N} \rho_m u_{N-m}$. Let N run through $b_1, b_2 \cdots$ to ∞. If $\sum \rho_k = \infty$, $\eta = 0$, it follows that $u_N \to 0$ are asserted. Otherwise $\eta = \mu^{-1}$.

Finally, it remains to show that this implies $u_N \to \eta$ for any subsequence when $N \to \infty$. Since by definintion, for every fixed k and sufficient large N, $u_{N-k} < \eta + \epsilon$, suppose under some approach $u_N \to \eta_0$, $1 = \sum_{m=0}^{N} \rho_m u_{N-m}$ gives $1 \leq \rho_0(\eta_0 - \eta) + \mu(\epsilon + \eta)$. Thus, $\eta_0 = \eta$, for otherwise $\eta_0 < \eta$, with sufficiently small ϵ, we get the contradiction.

Lemma: (*II*)

Let $\{b_n\}_{n=1}^{\infty}$ be a sequence of non-negative numbers, with $\sum_{i=1}^{\infty} b_i = b < \infty$ and $\{u_m\}_{m=0}^{\infty}$ be another sequence of non-negative numbers with limit u. Define $v_n = b_n + \sum_{k=1}^{n} u_k b_{n-k}$.

Then, $v_n \to bu$ as $n \to \infty$.

(*Proof of Lemma II.*) Denote $B_l = \sum_{m=1}^{l} b_m$, $\forall\ l \geq 1$. Since b_m is nonnegative, $B_l \leq b$, $\forall\ l \in \mathbb{N}$. Thus, $v_n = b_n + u_1 B_{n-1} + \sum_{k=2}^{n} (u_k - u_{k-1}) B_{n-k}$.

The latter can be written as $\sum_{k=2}^{\lfloor \frac{n}{2} \rfloor} (u_k - u_{k-1}) B_{n-k} + \sum_{k=\lfloor \frac{n}{2} \rfloor + 1}^{n} (u_k - u_{k-1}) B_{n-k}$. The 1$^{\text{st}}$ term converges to $(u - u_1)b$ while the 2$^{\text{nd}}$ term converges to 0 $(u_l - u_{l-1} \to 0,\ l \to \infty)$ as $n \to \infty$.

As a result, $\lim_{n \to \infty} v_n = ub$ (requiring detailed checking).

(*proof cont'd.*) Back to the main proof of BLT. Let j be an aperiodic recurrent state of the Markov chain (Xn). We can apply directly lemma (I) with $f_n = f_{jj}^{(n)}$, $u_n = P_j j^{(n)}$ and $\mu = m_j = \mathbb{E}(\tau_j | X_0 = j)$ to obtain claim (I), and (II) with $d = 1$.

For $d > 1$ case, introduce a new Markov chain (Y_k) with $Y_0 = X_0$ and $Y_k = X_{kd}$, then (Y_k) has transition matrix $Q = P^d$. Moreover, j is apeiodic for Q because if j had period $l \geq 2$, j would have period at least dl period for the original Markov chain (X_k), which is impossible. So apply the result above to the new Markov chain, $\lim_{n\infty} P_{jj}^{(dn)} = \lim_{n\infty} Q_{jj}^{(n)} = \frac{1}{m_j^Y} = \frac{d}{m_j}$, since $m_j^Y = \frac{1}{d} \sum_{k=1}^{\infty} k d f_{jj}^{(kd)} = \frac{1}{d} \sum_{n=1}^{\infty} n f_{jj}^{(n)} = \frac{m_j}{d}$. Thus, we finish checking claim (II).

As for claim (III), apply lemma (II) with $v_n = P_{ij}^{(n)}$, $b_n = f_{ij}^{(n)}$ and $u_n = P_{jj}^{(n)}$. The state j is here aperiodic recurrent, by the proved claims (II), $P_{jj}^{(n)} \to \dfrac{1}{m_j}$. We have then, $P_{ij}^{(n)} \to \dfrac{f_{ij}}{m_j}$. Claim (III) is established. ■

Remark. For an *ergodic* state j with mean first recurrent m_j in an *finite, irreducible* MC (*finite* + *irreducible* ⇒ all *recurrence*),

$$\lim_{n\to\infty} P_{ij}^{(n)} = \frac{1}{m_j}, \text{ for all } i.$$

Example 11. Bonus Malus System

- *Bonus Malus (Latin for Good-Bad) system*, is used to determine annual automobile insurance premiums in most of Europe and Asia.
- Each policyholder is given a positive integer valued state and the annual premium is a function of this state (along, of course, with the type of car being insured and the level of insurance).
- A policyholder's state changes from year to year in response to the number of claims made by that policyholder. Because lower numbered states correspond to lower annual premiums, a policyholder's state will usually decrease if he/she had no claims in the preceding year, and will generally increase if he or she had at least one claim (i.e., state(next year) = f(state(current year), claims)).

  - For a given Bonus Malus, let $s_i(k)$ denote the next state of a policyholder who was in sate i in the previous year and who made a total of k claims in that year.
  - If we suppose that the number of yearly claims made by a particular policyholder is a *Poisson random variable* with parameter λ, then the successive states of this policyholder will constitute a Markov chain with transition probabilities $p_{ij} = \sum_{k, s_i(k)=j} e^{-\lambda}\dfrac{\lambda^k}{k!}$.

Qn: Find the average annual premium paid bt policyholder (which can be obtained by considering limiting probabilities).

Ans. (Omitted).

2.5 Mean Time Spent in Transient States

Recall (6), define $s_{ij} = \lim_{\zeta\to 1^-} U_{ij}(\zeta) = \sum_{n=0}^{\infty} P_{ij}^{(n)} = \mathbb{E}\left[\sum_{n=0}^{\infty} \mathbb{1}_{(X_n=j)} \middle| X_0 = i\right]$, that is, the **expected number of visits** at state j starting from state i.

Thus, by definition, we directly have $s_{ij} = \mathbb{1}_{(i=j)} + f_{ij}s_{jj}$, which means $f_{ij} = \dfrac{s_{ij} - \mathbb{1}_{(i=j)}}{s_{jj}}$.

Lemma: Recurrence $\not\leftrightarrow$ Transience

For a *recurrent* state i and a *transient* state j, $p_{ij}^{(n)} = 0$ for all $n \geq 1$.

Proof. To prove this, we need another lemma, which says, for a *recurrent* i, if $i \to j$, then j is *recurrent*.

When $j \to i$, trivial case, see (7). Otherwise, if $j \not\to i$, $P(\text{not return to } i | X_0 = i) = 1 - f_i \geq f_{ij} > 0$, which gives $f_i < 1$, contradictory. Thus, $j \to i$ and j is recurrent.

Back to the previous lemma, $j \not\to i$ since $j \in T$ (transient), thus, $p_{ij}^{(n)} = 0$, for all $n \geq 1$ and obviously $s_{ij} = 0$. ■

Remark.

I Suppose i is a recurrent state and that $i \to j$. Then $f_{ij} = 1 = f_{ji}$.

(Reason: $f_{ij}(1 - f_{ji}) \leq 1 - f_i = 0$, but $f_{ij} \neq 0$. Since $i \leftrightarrow j$, same for f_{ij}.)

II A consequence of this proposition is that every *recurrent* state j will generate <u>a unique class of states</u>, reachable from j.

If k and l are two states of this class, $f_{kj} = f_{lj} = 1$, $k \leftrightarrow l$. In other words, it forms <u>a class of recurrent states</u> and $f_{ab} = 1$ for any tow states of this class.

Therefore, as shown below, all states can be decomposed into an unique partition of non-overlapping sets T, R_1, R_2, $\cdots$ such that T consists of all transient states and If j is in R_ν,then $f_{jk} = 1$ for all states k in R_ν while $f_{jk} = 0$ for all states k out of R_ν.

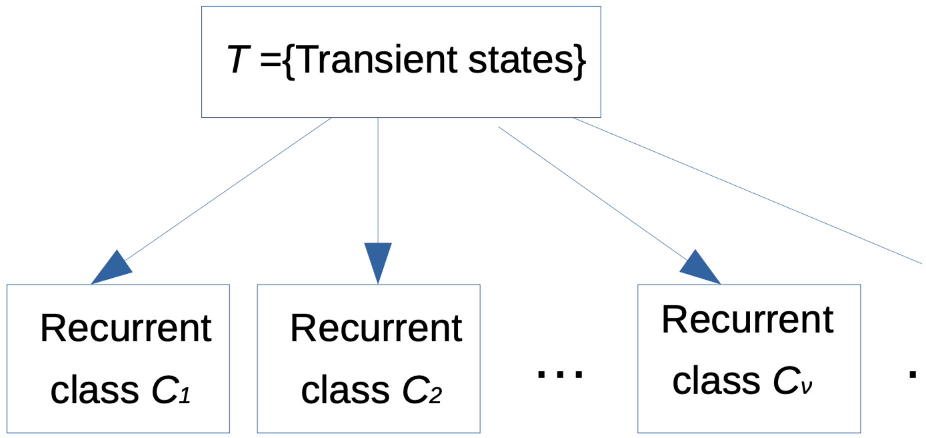


Figure 2.2: Partition of Non-overlapping Sets of States

2.5.1 Reducible Markov Chain with a Finite Number of Transient States

Consider a finite-state Markov chain with transient states $T = \{1, 2, \cdots, M\}$ and recurrent states: $R = \{M + 1, \cdots, N\}$.

Let P_T be the block of one state transition matrix P restricted to T, that is, $P = \begin{bmatrix} P_T & P_{TR} \\ P_{RT} & P_R \end{bmatrix}$. We want to determine s_{ij} for $i, j \in T$, which is called **mean time spent in transient states**.

Theorem 10. Mean Time Spent in Transient States

If $i, j \in T$, the expected number of visits at state j starting from state i,

$$s_{ij} = \mathbb{1}_{(i=j)} + \sum_{k=1}^{M} p_{ik} s_{kj}.$$

In matrix from, equivalently,

$$S_T = I + P_T S_T,$$

where $S_T := (s_{ij})_{M \times M}$

Proof. Note that by (4), $s_{ij} = \mathbb{1}_{(i=j)} + \sum_{n=1}^{\infty} p_{ij}^{(n)} = \mathbb{1}_{(i=j)} + \sum_{n=1}^{\infty} \sum_{k} p_{ik} p_{kj}^{(n-1)} = \mathbb{1}_{(i=j)} + \sum_{k} \sum_{n=0}^{\infty} p_{ik} p_{kj}^{(n)} = \mathbb{1}_{(i=j)} + \sum_{k} p_{ik} s_{kj} = \mathbb{1}_{(i=j)} + \sum_{k \in T} p_{ik} s_{kj}$, since $s_{kj} = 0$ if $k \in R$. ■

Remark. Since for $i, j \in T$, $(P_T^n)_{ij} \le p_{ij}^{(n)} \to 0$ for $n \to \infty$, $S_T = (I - P_T)^{-1}$.

2.5.2 Gambling Model

A direct application is to calculate mean time spent in *Gambling Model* (9), since it only has 2 recurrent states 0 and N.

Thus, it is direct to calculate $S_T = (I_{N-1} - P_T)^{-1}$ to get the expected amount of time the gambler has j units, where $j = 1, \cdots, N-1$.

As for **ruin probability**, which is $1 - a_i = P(\lim_{n\to\infty} X_n = 0 | X_0 = i)$, and $a_i = P(\lim_{n\to\infty} X_n = N | X_0 = i)$.

To calculate this, we need to use 1^{st} *step analysis*.

$a_0 = 0$ and $a_N = 1$, $a_i = pa_{i+1} + (1-p)a_{i-1}$, for $i \in \{1, \cdots, N-1\}$. From the recursive equation, easily we get $a_i = \begin{cases} \dfrac{1 - (\frac{1}{p} - 1)^i}{1 - (\frac{1}{p} - 1)^N}, & p \neq \dfrac{1}{2} \\ \dfrac{i}{N}, & p = \dfrac{1}{2} \end{cases}$.

Another question is to find $\lim_{n\to\infty} p_{ij}^{(n)}$. By BLT (9) and above results, we can have $\lim_{n\to\infty} p_{ij}^{(n)} = \begin{cases} 1 - a_i, & j = 0 \\ 0, & j = 1, \cdots, N-1 \\ a_i, & j = N \end{cases}$.

Under *simulation*: plotting trajectories of X_n (random), but when plotting the pmfs of X_n, it is relatively fixed, i.e., $\mu_0 = (0, \cdots, 1, \cdots, 0) \Rightarrow \lim_{n\to\infty} \mu_n = (1 - a_i, \cdots, a_i)$.

```
# Illustration of the gambler's Markov chain
# Copyright  Jeff Yao [SDS, CUHK-SZ]
##############################################
# Usage
# Step 1:  Run the codes of Parts A-B-C to define the 3 functions

# Step 2:  Define the gambler's chain using part ### D:
##         where winning  probability =p, n=100 = maximum fortune

# Step 3:  then call the programs with
#
#  -->  run.mc.plot(m,100)
#            [MC itself:  100 = number of steps, you may change it as
   fit]
#  -->  gambler.chain.dist(n,m)
##           [sequence of distributions
##  The MC always starts at the middle value X_0=(n+1)/2
################################################################

#### D : set up the transition matrix
#----------------------------------
p <- 19/38
# p <- 19/38
# p <- 20/38

q <- 1-p
n <- 100
m <- matrix(0, n+1, n+1)
for (i in 2:n) {
m[i,i-1] <- q
m[i,i+1] <- p
}
# Absorbing barriers on both sides:
m[1,1] <- 1
m[n+1,n+1] <- 1


#### A:  function displaying successive distributions
#----------------------------------
gambler.chain.dist <- function(n,m) {
c <- floor(n/2)    # center  point to start with

x <- matrix(0, 1, n+1)
x[1,c+1] <- 1
for (i in 1:2000) {
x <- x %*% m
plot((1:(n+1))[x>0], x[x>0], xlim=c(1,n+1),  xlab="State 'i'",
ylab="Distribution P(X_n='i') after n steps", col=2)

date_time<-Sys.time()     # control delay in plots
while((as.numeric(Sys.time()) - as.numeric(date_time))<0.06){} #dummy
   while loop
}
}
```

```
#### B:  dynamic view of the rouletter game
#-----------------------------------------------

#  a general Markov chain genenration code
#
   https://stephens999.github.io/fiveMinuteStats/simulating_discrete_chains_1.html
# simulate discrete Markov chains according to transition matrix P

run.mc.sim <- function(P, num.iters = 50 ) {

# number of possible states
num.states <- nrow(P)

# stores the states X_t through time
states      <- numeric(num.iters)

# initialize variable for first state
states[1]    <- floor(num.states/2)   # starts at middle

for(t in 2:num.iters) {

    # probability vector to simulate next state X_{t+1}
    p  <- P[states[t-1], ]

    ## draw from multinomial and determine state
    states[t] <-  which(rmultinom(1, 1, p) == 1)
}
return(states)
}

#### C.  plot
run.mc.plot <-function(P, num.iters =50) {

chain.states <- run.mc.sim(P,num.iters)

for (i in 1:num.iters) {
x <- chain.states[1:i]
plot((1:i)[x>0], x[x>0], xlim=c(1,num.iters),  xlab="Time n",
   ylab="Gambler's wealth",col=2)

date_time<-Sys.time()     # control delay in plots
while((as.numeric(Sys.time()) - as.numeric(date_time))<0.16){} #dummy
   while loop
}
```

Listing 2.1: Simulation of the Gambler's Markov Chain with R

2.6 Time Reversible Markov Chains

Consider a positive recurrent Markov chain $\{X_n : n \geq 0\}$ with transition probability matrix $P = (p_{ij})$ and invariant probability $\pi = (\pi_i)$;

From the Markov property $P(\text{future}|\text{present,past}) = P(\text{future}|\text{present})$, we have similarly

$$P(\text{past}|\text{present,future}) = P(\text{past}|\text{present}).$$

If we look the process backward in time, i.e. $\cdots, X_{n+1}, X_n, X_{n-1}, \cdots$, then the reversed process is again a Markov chain. Further, its transition probabilities q_{ij} is

$$q_{ij} = \frac{\pi_j p_{ji}}{\pi_i}.$$

If $q_{ij} = p_{ij}$, for all i and j, then the Markov chain is said to be **time reversible**.

`Meaning of Time Reversal`:

Looking forward, the asymptotic frequencies of consecutive pairs (i, j) and (j, i) are the same:

$$\lim_{T\to\infty} \frac{1}{T} \sum_{n=0}^{T} \mathbb{1}_{\{X_n=i, X_{n+1}=j\}} = \pi_i p_{ij} = \pi_j p_{ij} = \lim_{T\to\infty} \frac{1}{T} \sum_{n=0}^{T} \mathbb{1}_{\{X_n=j, X_{n+1}=i\}}.$$

In this stationary case, the distribution of (X_n, X_{n+1}) is the same as its time-reversed counter-part (X_{n+1}, X_n).

Lemma: A Techical Way to Find Invariant Probabilities

Any probability measure $\mu = (\mu_i)$ satisfying the *balanced equation* $\mu_i p_{ij} = \mu_j p_{ij}$, forall i, j, is an *invariant probability* for the Markov chain, i.e. $\mu P = \mu$.

Remark. Indeed for general positive recurrent Markov chains, it is difficult to find the invariant probability (limiting probabilities).

This lemma gives a way to determine the invariant probability since it is much easier to check the balanced equation for a candidate probability μ.

This principle is widely used in `MCMC` (Markov Chain Monte-Carlo) methods, e.g. for the Metropolis sampler and the Gibbs sampler.

Back to the example of *Ehrenfest model of diffusion*, since it is time-reversible (use definition), we can use the above technique to solve for limiting probabilities, which is

$$\pi_0 = \left[1 + \sum_{j=1}^{M} \frac{(M-j-1)\cdots(M-1)M}{j!}\right]^{-1} = \left(\frac{1}{2}\right)^M, \pi_i = \binom{M}{i}\left(\frac{1}{2}\right)^M, 0 \le i \le M.$$

(Other examples: RW with probabilities $p_{i,i+1} = \alpha_i > 0$; Page-Rank, etc.)

Chapter 3

Poisson Processes

3.1 Motivation and Introduction

Consider earthquake data: by invoking longitudes and latitudes $(\cdot, \cdot)$, people can determine the location of the earthquaking place. A common question is then to predict earthquake: **find differences** between two consecutive earthquaking places.

Which kinds of differences? For the simplest $1-$dim cases, the model is usually a queue. Precise addition of such "queues" needs **theory of point process** (e.g., how much more).

The simplest model of 1-dim case is **Posisson point process**. We need to figure out different meanings of *value* and *location*.

3.2 Recap of Exponential Distributions

For its pdf, CDF, mgf, memoryless property, and relationship with Gamma distributions, see in my first note of statistics.

Lemma: Independence Exponential Variables

Suppose $X_k \sim exp(\lambda_k)$ for $k = 1, \cdots, n$ and they are independent.

I. Their **minimum** is still a exponential r.v.: $\min\{X_1, \cdots, X_n\} \sim exp(\sum_{k=1}^{n} \lambda_k)$.

II. $P(X_1 < X_2) = \dfrac{\lambda_1}{\lambda_1 + \lambda_2}$; $\mathbb{E}(X_1 \mathbb{1}_{\{X_1 < X_2\}}) = \dfrac{\lambda_1}{(\lambda_1 + \lambda_2)^2}$.

III. Let R_n be the index of the observation realizing the minimum of the sequence, i.e. $R_n = i$ if $X_i = \min\{X_1, \cdots, X_n\}$.

The distribution of R_n is: $P(R_n = i) = \dfrac{\lambda_i}{\sum_{k=1}^{n} \lambda_k}$.

Proof. For II., with conditioning, $P(X_1 < X_2) = \mathbb{E}(\mathbb{1}_{\{X_1 < X_2\}}) = \mathbb{E}[\mathbb{E}(\mathbb{1}_{\{X_1 < X_2\}} | X_1)] =$

$\mathbb{E}[e^{-\lambda_2 X_1}] \xlongequal{\text{mgf}} \dfrac{\lambda_1}{\lambda_1 + \lambda_2}$. Another one can be proved with the same approach.

For III., $P(R_n = i) = P(X_i = \min\{X_1, \cdots, X_n\}) = P(X_i \le \min_{j \neq i}\{X_j\}) = P(X_i < \min_{j \neq i}\{X_j\}) \xlongequal{\text{II.}} \dfrac{\lambda_i}{\lambda_i + \sum_{k \neq i} \lambda_k}$. ■

3.3 The Poisson Process

3.3.1 Introductory Example: Phone Calls

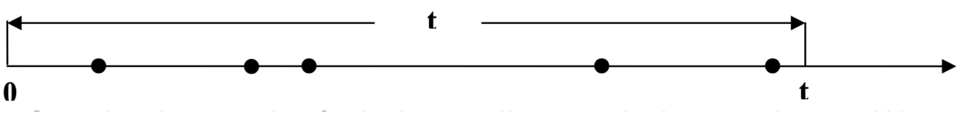


Figure 3.1: time axis

Consider the arrivals of telephone calls at a telephone exchange. We assume that we count the arrivals of calls N_t at time t, from the beginning of some time denoted as $t = 0$.

Also for each $t \ge 0$, we have a random variable N_t and thus we have a stochastic process $\{N_t \in \mathbb{N} : t \ge 0\}$. Note that the time parameter $t \in [0, \infty)$ is *continuous*, meanwhile, state space $E = \{0, 1, 2, \cdots\}$ is *discrete*.

Thus, a process is a mapping $N.(\cdot) : \Omega \times [0, \infty) \to \{0, 1, 2, \cdots\}$, such that $N_t(\omega)$ denotes the arrivals at time t with parameter ω for measurement of randomness.

One kind of process is called **counting process**, which is defined as:

- For $t \ge 0$, N_t has nonnegative integer values;
- For $s < t$, $N_t - N_s$ equals the number of events that have occurred in the interval $(s, t]$.

Note that the second property implies that $t \to N_t$ is increasing: if $s \ge t$, then $N_s \le N_t$.

Consider the next 2 counting processes with two essential concepts:

(1) $N_t = \#$ of phone calls in the time interval $(0, t]$.

(2) $N_t = \#$ of people who were born at time t.

Proposition: Two Important Concepts

- ⋆ **Independent Increments**: The numbers of events which occur in *disjoint time intervals are independent* (reasonable for (1) but NOT for (2)).
- ⋆ **Stationary Increments**: $N_{t+u} - N_{s+u}$ has *the same distribution* as $N_t - N_s$, for all $t > s > 0$, $u \ge 0$ (reasonable for (1) but NOT for (2) even the birth rate is kept to be constant).

3.3.2 Definition of Poisson Process

Here we give the formal definition of Poisson processes.

Definition: Poisson Processes

The *counting process* $\{N_t : t \geq 0\}$ is said to be a **Poisson process** having rate λ, if:

(1) $N_0 = 0$;

(2) The process has **independent increments**;

(3) The number of events in any interval of length t is **Poisson distributed** with mean λt. That is, for all $s, t \geq 0$,

$$P(N_{t+s} - N_s = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, n = 0, 1, \cdots,$$

which is independent of s.

From this we can directly know that $\mathbb{E}(N_t) = \mathbb{E}(N_t - N_0) = \lambda t$, which means $\lambda = \dfrac{\mathbb{E}(N_t)}{t}$ is the average rate of arrivals.

Lemma

Let N_h be the number of points in $(0, h]$ of a Poisson process $\{N_t : t \geq 0\}$ with rate λ. Then for small $h \sim 0$,

- $P(N_h = 0) = 1 - \lambda h + o(h)$;
- $P(N_h = 1) = \lambda h + o(h)$;
- $P(N_h \geq 2) = o(h)$.

That is, on a interval of very small length h, there will be almost no points. In other words, up to an error of $o(h)$, N_h could be considered as a $Ber(\lambda h)$ r.v. (Bernoulli).

So why do we need Poisson distribution? Here from the link of *Poisson approximation*, we may learn about the reason.

Lemma: Posisson Approximation

Let $X_1, \cdots, X_n$ be an i.i.d. sequence of $Ber(p)$ r.v's and assume that p is related to n such that $np \to \lambda > 0$. Then when $n \to \infty$, the distribution of $S_n = X_1 + \cdots + X_n$ converges to $Pois(\lambda)$.

3.4 Interarrival and Waiting Time Distributions

Firstly, we give the formal definition of interarrival time. Since we can know that with the sequence of interarrival time or waiting time, we can also describe a Poisson process (i.e. $\{N_t\}_{t\geq 0} \Leftrightarrow \{T_n\}_{n\geq 1} \Leftrightarrow \{S_n\}_{n\geq 1}$).

Definition: Interarrival Time and Arrival Time

Consider a Poisson process.

T_1 – the time of the first event;
T_n – the *elapsed time* between the $(n-1)^{\text{th}}$ and the n^{th} event, $n > 1$.

The sequence $\{T_n : n = 1, 2, \cdots\}$ is called *the sequence of* ***interarrival times***.

The **arrival time** S_n of the nth event, also called the **waiting time** until the nth event, is defined by $S_n = \sum_{i=1}^{n} T_i$.

The sequence $\{S_n : n = 1, 2, \cdots\}$ is called *the sequence of* ***arrival times***.

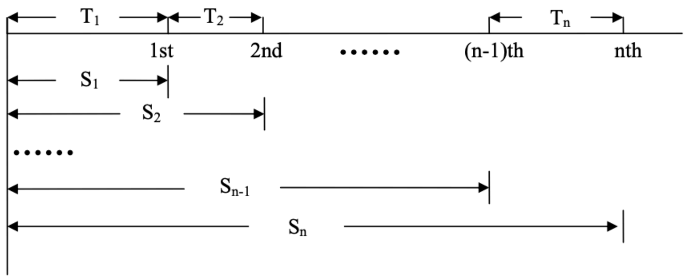


Figure 3.2: Illustration of interarrival and arrival time

Lemma

The interarrival times $\{T_n : n = 1, 2, \cdots\}$ is a sequence of **i.i.d.** $exp(\lambda)$ **(exponential) distributed random variables**.

Proof. By small $h-$method (The same approach is used below, too.)

Step 1: Find joint distribution of $(S_1, \cdots, S_n)$. When $x_1 < \cdots < x_n$, since by the meaning of *density* $P(x_1 < S_1 \leq x_1 + h_1, \cdots, x_n < S_n \leq x_1 + h_m) \sim f(x_1, \cdots, x_n) \prod_{i=1}^{n} h_i$, when $h_k \to 0$, $\forall\ k \in \mathbb{N}$, and furthermore,

LHS $= P(0$ arrival in $[0, x_1), 1$ arrival in $[x_1, x_1 + h_1); \cdots ; 0$ arrival in $[x_{n-1} + h_{n-1}, x_n)$, 1 arrival in $[x_n, x_n + h_n))$

$= [e^{-\lambda x_1} e^{-\lambda(x_2 - x_1 - h_1)} \cdots e^{-\lambda(x_n - x_{n-1} - h_{n-1})}](\lambda^n \prod_{i=1}^{n} h_i) \cdot [e^{-\lambda h_1} e^{-\lambda h_2} e^{-\lambda h_n}]$

$= \lambda^n e^{-\lambda x_n} \prod_{i=1}^{n} h_i \sim \lambda^n e^{-\lambda x_n} \prod_{i=1}^{n} h_i$

Thus, $f(x_1, \cdots, x_n) = \lambda^n e^{-\lambda x_n} \mathbb{1}_{(0 < x_1 < \cdots < x_n)}$.

Step 2: Change variables to find

$$g_{T_1,\cdots,T_n}(x_1,\cdots,x_n) = f_{S_1,\cdots,S_n}(x_1,\cdots,\sum_{k=1}^{n} x_k) \begin{vmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \ddots & \cdots \\ 1 & 1 & 1 & \cdots & 1 \end{vmatrix} = \lambda^n e^{-\lambda \sum\limits_{k=1}^{n} x_k}.$$

Then, the conclusion holds. ■

Remark. An easy consequence of the theorem is that that the n^{th} arrival time $S_n = T_1 + \cdots + T_n$ follows a gamma distribution $\Gamma(n,\lambda)$.

Its pdf is $f_{S_n}(t) = \dfrac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda x}$, while its mgf is $\mathbb{E}(e^{tS_n}) = \left(\dfrac{\lambda}{\lambda - t}\right)^n$, $t < \lambda$.

3.5 Conditional Distribution of Waiting Time

Now, we consider the (joint) conditional distribution of $\{S_n : n = 1, 2, \cdots\}$.

Suppose that exactly one event of a Poisson process has taken place by time t. Since a Poisson process possesses stationary and independent increments it seems reasonable that each interval in $[0,t]$ of equal length should have the same probability of containing the event. In other words, the time of the event should be uniformly distributed over $[0,t]$.

Lemma

(I) Conditional to $N_t = 1$, T_1 is **distributed uniformly** on $(0,t]$.

(II) Generally, provided that $N_t = n$, the joint distribution of the n arrival (waiting) times $S_1, S_2, \cdots, S_n$ is the same as that of **the order statistics from a sample of n i.i.d. random variables uniformly distributed** over the interval $(0,t]$, i.e. the density function is $f(s_1,\cdots,s_n|N_t = n) = \dfrac{n!}{t^n}\mathbb{1}_{(0<s_1<\cdots<s_n)}$.

Proof. To find $g(s|N_t = 1)$, still by using the meaning of *density*, we get $P(s < T_1 \le s + h|N_t = 1) \sim hg(s)$, and LHS $= \dfrac{P(s < T_1 \le s + h, N_t = 1)}{P(N_t = 1)}$

$= \dfrac{P(0 \text{ arrival in } (0,s], 1 \text{ arrival in } (s,s+h], 0 \text{ arrival in } (s+h,t])}{P(N_t = 1)} = \dfrac{h}{t}$,

which means that $g(s) = \dfrac{1}{t}$.

For the second part, similar approach gives that $f(s_1,\cdots,s_n|N_t = n) = \dfrac{n!}{t^n}\mathbb{1}_{(0<s_1<\cdots<s_n)}$. ■

Example 12. Insurance Claims

- Insurance claims arrive at P.P. times with rate λ;
- The successive claim amounts are independent r.v. having distribution G with mean μ, and are independent of the claim arrival times;
- Let S_i and C_i denote, respectively, the time and the amount of the i^{th} claim;
- The total discounted cost of all claims made up to time t, call it D(t), is defined by $D(t) = \sum_{i=1}^{N_t} e^{-\alpha S_i} C_i$.

`Qn`: Find $\mathbb{E}(D_t)$.

`Ans`. By conditioning $\mathbb{E}(D_t) = \mathbb{E}[\mathbb{E}(D_t|N_t)]$. Since $\mathbb{E}(D_t|N_t = n) = \sum_{i=1}^{n} \mu E(e^{-\alpha S_i}|N_t = n) = \sum_{i=1}^{n} \mu E(e^{-\alpha U_{(i)}}) = \sum_{i=1}^{n} \mu E(e^{-\alpha U_i}) = \dfrac{n\mu}{\alpha t}(1 - e^{-\alpha t})$.

Thus, $\mathbb{E}[\mathbb{E}(D_t|N_t)] = \dfrac{\lambda\mu}{\alpha}(1 - e^{-\alpha t})$.

3.6 Further Properties of Poisson Process

Recall for Poisson distribution, there is a theorem saying that two ibinomial distributions conditioned on a Poisson distribution are independent Posisson distributions (as follows).

Lemma: Coloring Theorem of Poisson Distribution

Assume that a counting variable N has Poisson distribution $Pois(\lambda)$ that the counted bowls are red with probability p and blue with probability $q = 1 - p$.

Let N_R and N_B be the counts of red and blue bowls, respectively. Then, $N_R \sim Pois(\lambda p)$, $N_B \sim Pois(\lambda q)$ and they are independent.

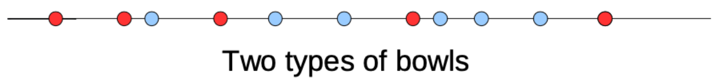


Figure 3.3: An illustration

Now, consider a Posisson process $\{N_t\}$ with rate λ. Suppose that the events can be classified into two types: type I with probability p (red bowls!) and type II with probability $q = 1 - p$ (blue bowls!), independently of all other events.

Let $N_t^{(1)}$ and $N_t^{(2)}$ be the numbers of type I and II events occurring in $[0, t]$, respectively. Note that $N_t = N_t^{(1)} + N_t^{(2)}$.

Then, we have the similar results as that of Posisson distributions.

Lemma

The counting processes $\{N_t^{(1)} : t \geq 0\}$ and $\{N_t^{(2)} : t \geq 0\}$ of type I and II events are **both Poisson processes** having respectively rates $p\lambda$ and $q\lambda$. Furthermore, *they are independent.*

Proof. Take $0 < t_1 < \cdots < t_n$, consider ΔN_{t_i}, and use coloring theorem above. ∎

Now, we consider a simple application with the P.P. model in the finance field (with above coloring lemma).

Example 13

- Nonnegative offers to buy an item that you want to sell arrive according to a P.P. with rate λ;
- Each offer is the value of a continuous r.v. X having density function $f(x)$;
- Once the offer is presented to you, you must either accept it or reject it and wait for the next offer;
- You incur costs at a rate c per unit time until the item is sold;
- Your objective is to *maximise your expected total return*, where the total return is equal to the amount received minus the total cost incurred;
- You employ the policy of accepting the first offer that is greater than some specified value y. (Such a type of policy, which we call a y-policy, can be shown to be optimal.)

Qn: What is the best value of y?

Ans. $\mathbb{E}[R(y)] = \mathbb{E}[X|X > y] - \mathbb{E}[cT_1] = \dfrac{1}{\overline{F}(y)} \displaystyle\int_y^\infty xf(x)dx - \dfrac{c}{\lambda\overline{F}(y)}$, where $\overline{F}(y) = P(X > y)$, and the time until an offer is accepted is an exponential random variable T_1 with rate $\lambda\overline{F}(y)$.

Let $\dfrac{d}{dy}\mathbb{E}[R(y)] = 0$, we get $\mathbb{E}[(X - y)_+] = \dfrac{c}{\lambda}$, since $\mathbb{E}[(X - y)_+]$ is decreasing with y, we compare $\mathbb{E}[X]$ with $\dfrac{c}{\lambda}$ to get the solution.

Proposition: Backcasting

Let $\{N_t : t \geq 0\}$ be a Poisson process with rate λ. Assume that $N_t = n$ for some $t > 0$ and $n \geq N$. The for $0 < s \leq t$, the *conditional distribution of N_s given $N_t = n$ is binomial* with parameters n and $p = \frac{s}{t}$.

Proof. Consider $P(N_s = m|N_t = n) = \dfrac{P(N_s = m, N_t = n)}{P(N_t = n)}$

$$= \frac{P(m \text{ arrivals in } (0,s], n-m \text{ arrivals in } (s,t])}{P(N_t = n)} = \binom{n}{m} \left(\frac{s}{t}\right)^m (1-\frac{s}{t})^{n-m}. \qquad \blacksquare$$

Proposition: Superposition

Let $\{N_{j,t} : t \geq 0\}$, $j = 1, 2, \cdots, m$, be m *independent Poisson processes* every with corresponding rate λ_j. Let $N_t = \sum_{k=1}^{m} N_{k,t}$. Then, $\{N_t : t \geq 0\}$ is a *Poisson process* with rate $\lambda = \sum_{k=1}^{m} \lambda_k$.

Proof. Check three definitions of Poisson process. $\blacksquare$

Here is an example with this superposition proposition.

Example 14

What is the probability that n events occur in one P.P. before m events have occurred in a second and independent P.P.?

Ans. Let $\{N_t^{(1)} : t \geq 0\}$ and $\{N_t^{(2)} : t \geq 0\}$ be two independent P.P having respective rates λ_1 and λ_2. $S_n^{(1)}$ denote the time of the n^{th} event in 1 and $S_m^{(2)}$ for the time of the m^{th} event in 2.

Noting that we can put the two events together (superposition), and then, this event will occur if and only if the first $n+m-1$ tosses result in n or more events 1 (with probability $\frac{\lambda_1}{\lambda_1+\lambda_2}$), we see that our desired probability is given by

$$P(S_n^{(1)} < S_m^{(2)}) = \sum_{k=n}^{n+m-1} \binom{n+m-1}{k} \frac{\lambda_1^k \lambda_2^{(n+m-1-k)}}{(\lambda_1+\lambda_2)^{(n+m-1)}}$$

Proposition: Decomposition

Let $\{N_t : t \geq 0\}$ be a Poisson process with rate λ. Suppose that every event can be independently classified as type j with probability p_j , for $j = 1, 2, \cdots, m$. Let $N_{j,t} = \#$ of type j events in $[0, t]$.

Then, $\{N_{j,t} : t \geq 0\}$, $j = 1, 2, \cdots, m$, are independent Poisson processes each with rate $\lambda_j = p_j \lambda$.

3.7 Compound Poisson Process

Definition: Compound Poisson Process

A stochastic process $\{X_t : t \geq 0\}$ is said to be a **compound Poisson process** if it can be represented as

$$X_t = \sum_{i=1}^{N_t} Y_i,$$

where $\{N_t : t \geq 0\}$ is a *Poisson process*, and $\{Y_i : i \geq 1\}$ is a family of independent and identically distributed random variables which is also independent of $\{N_t : t \geq 0\}$

Remark. Note that $t \to N_t$ is NOT necessarily increasing since behaviors of Y_i are unknown.

Thus, we can directly know that a compound Poisson process is NOT necessarily a Poisson process.

Here is an example of a compound Poisson process. Suppose that buses arrive at a sporting event in accordance with a Poisson process, and suppose that the numbers of fans in each bus are assumed to be i.i.d. Then $\{X_t : t \geq 0\}$ is a compound Poisson process, where N_t represents the number of buses arriving at the sporting event by time t, Y_i represents the number of fans in the i^{th} bus, and $X_t = \sum_{i=1}^{N_t} Y_i$ denotes the number of fans who have arrived by time t. The compound Poisson variable has some good properties like:

Proposition

For a compound Poisson process $\{X_t : t \geq 0\}$ with rate λ for $\{N_t : t \geq 0\}$,

$$\mathbb{E}(X_t) = \lambda t \mathbb{E}(Y), \mathrm{Var}(X_t) = \lambda t \mathbb{E}(Y^2),$$

where $\mathbb{E}(Y) = \mathbb{E}(Y_i)$ and $\mathbb{E}(Y^2) = \mathbb{E}(Y_i^2)$ for any i.

Proof. Omitted. ■

For the compound Poisson process, we have the following **Normal approximation**:

- Let $\{X_t : t \geq 0\}$ be a compound Poisson process built on *discrete random variable* Y_i, with pmf
$$P(Y = c_k) = p_k, k = 1, \cdots, m$$
- Let $N_{k,t} = \#$ of event by time t such that the corresponding value of Y_i is c_k. Then $\{N_{k,t} : t \geq 0\}, k = 1, \cdots, m$ are independent Poisson processes with respective rates $\lambda_k = \lambda p_k$.
- For large t, $N_{k,t}$ is approximately distributed as normal (CLT).
- Since the compound Poisson process is a linear combination of the decomposed Poisson processes:

$$X(t) = c_1 N_{1,t} + c_2 N_{2,t} + \cdots + c_m N_{m,t},$$

we can approximate the distribution of X_t by normal if t is large.

3.8 Inhomogeneous Poisson Process

- A standard P.P. has the **memoryless property**. In particular, the interarrival times are i.i.d. $exp(\lambda)$;
- This is of cause a very strong assumption and we need more general/flexible models without this memoryless property.

Definition: Inhomogeneous Poisson process

A counting process $\{N_t : t \geq 0\}$ is a **non-homogeneous Poisson process** with a continuous *intensity function* $\lambda(t) : t \geq 0$, if

(1) $N_0 = 0$;

(2) $\{N_t : t \geq 0\}$ has **independent** increments but NOT necessarily stationary;

(3) For infinitesmall positive $h > 0$,

$$P(N_{t+h} - N_t \geq 2) = o(h);$$

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

[In consequence, $N_{t+h} - N_t$ approximately follows $Ber(\lambda(t))$.]

Thus, we can estimate the inhomogeneous Poisson process by sum of Bernoulli distributions.

Lemma

For a Poisson P.P. with *intensity function* $\lambda(t)$, we have for $0 \leq s < t$,

$$N_t - N_s \sim Pois\left(\int_s^t \lambda\right).$$

To understand/prove above lemma, we need to invoke some knowledge from calculus and differential equations.

Firstly, let $p_m(t) = P(N_{(t)} = m)$, $\forall\, m \in \mathbb{N}$. Take $h \to 0^+$ (for all the explanations). Thus, $\frac{d}{dt}p_0(t) = \lim_{h\to 0^+} \frac{p_0(t+h) - p_0(t)}{h} = \lim_{h\to 0^+} \frac{1}{h}[P(N_{t+h} - N_t = 0)P(N_t = 0) - P(N_t = 0)] = \lim_{h\to 0^+} \left(-\lambda(t) + \frac{o(h)}{h}\right) p_0(t) = -\lambda(t)p_0(t)$.

From above, we first get that $p_0(t) = e^{-\int_0^t \lambda(\tau)d\tau}$, since $p_0(0) = 1$ from definintion.

Now, consider the case $n > 0$, for $P(N_t = n)$. This is slightly messier but repeats the same ideas. $p_n(t+h) = \sum_{k=0}^{n} p_k(t)P(N_{t+h} - N_t = n-k) = p_n(t)(1 - \lambda(t)h) + p_{n-1}(t)\lambda(t)h +$

$\sum_{k=0}^{n} p_k(t)o(h)$. Taking limits as above, $p_n'(t) = -\lambda(t)\left(p_n(t) - p_{n-1}(t)\right)$, $\forall\ n \geq 1$.

Solving this system of ODEs, we have $p_n(t) = e^{-\int_0^t \lambda(\tau)d\tau} \int_0^t \lambda(s) e^{\int_0^s \lambda(\tau)d\tau} p_{n-1}(s)ds$, since $p_n(0) = 0$, $\forall\ n > 0$. With the value of $p_0(t)$, we can consequentially get

$$p_n(t) = e^{-\int_0^t \lambda(\tau)d\tau} \frac{1}{n!} \left(\int_0^t \lambda(\tau)d\tau \right)^n .$$

Note that $\left[\frac{1}{n!} \left(\int_0^s \lambda(\tau)d\tau \right)^n \right]' = \lambda(s) \left[\frac{1}{(n-1)!} \left(\int_0^s \lambda(\tau)d\tau \right)^{n-1} \right]$.

Remark. It is also possible to approach the solution by moment-generating function (mgf).

3.8.1 Arrival Time

Consider an inhomogeneous Poisson process with intensity $\lambda(t)$, and the associated waiting times. Let T_1 denote the first arrival, then the waiting time is just the **exponential process**.

$$f_{T_1}(t) = \frac{d}{dt} P(N_t \geq 1) = \lambda(t) e^{-\int_0^t \lambda(\tau)d\tau}.$$

Chapter 4

Continuous Time Markov Chains (Markov Jump Processes)

A continuous-time Markov chain (CTMC) is a **continuous-time stochastic process** $X = \{X_{(t)}, t \geq 0\}$, taking values in a **discrete state space** E, and having *piecewise constant sample paths*. The time durations between jumps are exponentially distributed: at every jump time, the CTMC jumps from the current state to another state independently of the history.

To characterize a CTMC, instead of transition matrix, we have **jump matrix**, J, here. The characteristics are $J_{ij} \geq 0$, $\sum_j J_{ij} = 1$. Also we need a **holding time rate parameter** for each state that determines the holding time before each jump.

4.1 Introduction to MJP

Initially, here is an introductory example.

Example 15. Two Machines Example

Two machine with the same characteristics go up and down over time.

Consider one of them. When it is down, it takes d_i amount of time to be repaired, where d_i is exponentially distributed with mean 1 hour. $\{d_i : i = 1, 2, 3, \cdots\}$ is an i.i.d. sequence. When it is up, it will stay up u_i amount of time that is exponentially distributed with mean 10 hours. $\{u_i : i = 1, 2, 3, \cdots\}$ is an i.i.d. sequence.

Assume that up times and repair times are independent.

Qn: Can the problem be modeled by a CTMC?

Ans. As displayed above, we need t construct jump matrix and holding parameters. Let $X = \#$ of machines at up state. State space of X is $E = \{0, 1, 2\}$. Without doubt that the jump matrix is $\begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{11} & 0 & \frac{10}{11} \\ 0 & 1 & 0 \end{bmatrix}$. In the state 2 (suppose they do not become down simutaneously), two machines are competing to fail first. The holding time is

$\min(X_1, X_2) \sim exp(\frac{1}{10} + \frac{1}{10}) = exp(\frac{1}{5})$. At state 0, one of them is repaired, the holding time is $exp(1)$. The holding time for state 1 follows $exp(1 + \frac{1}{10}) = exp(\frac{11}{10})$.

Definition: Definition (I) of a MJP

Let

(1) $(\xi_n)_{n\geq 0}$ be a MC on a discrete state space E (with transition matrix Q);

(2) $(U_n)_{n\geq 0}$ be i.i.d. $\sim exp(1)$;

(3) family of holding rates be $(\lambda(i))_{i\in E}$.

A MJP pocessing with the above parameters is defined as

$$X_t = \xi_n,\ \ \sigma_n \leq t \leq \sigma_{n+1},$$

where $\sigma_n = \begin{cases} 0, & n = 0 \\ \sum_{i=1}^{n} \frac{U_i}{\lambda(\xi_{i-1})}, & n \geq 1 \end{cases}$

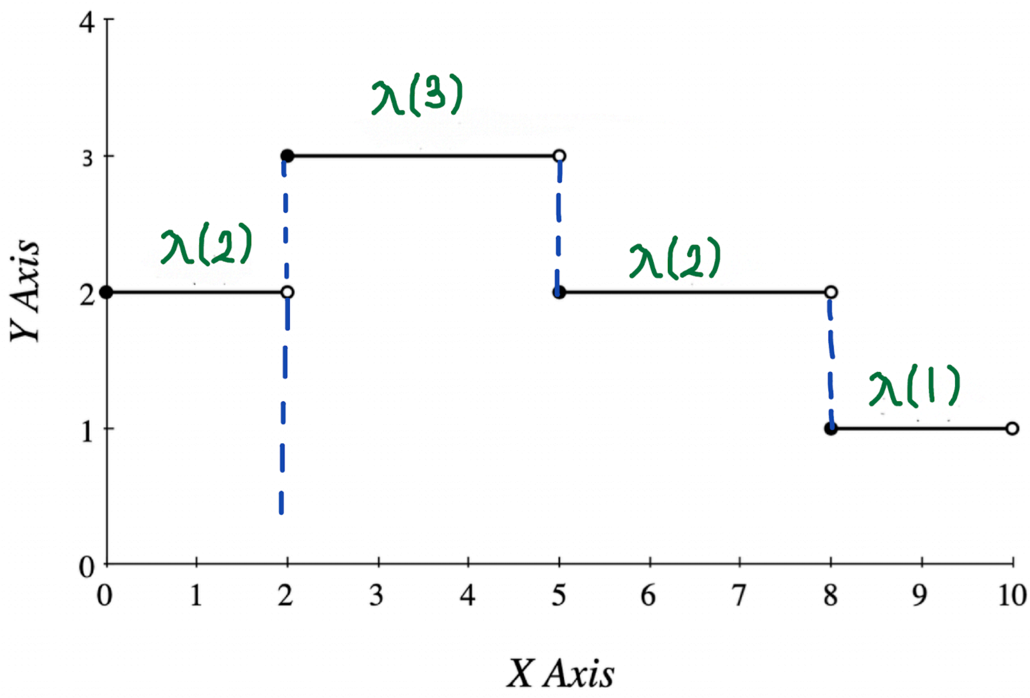


Figure 4.1: Illustration of MJP

Note that the holding time varies similarly like the inter-arrival time as in Poisson processes. Here we can have another different definitions from the previous standard one.

Definition: Definition (II) of a MJP

Let $(P_t)_{t\geq 0}$ be a family of transition matrix on a state space E, i.e., $\forall\, t \geq 0$, P_t is a transition matrix.

A MJP on E with transition kernel $(P_t)_{t\geq 0}$ is a process $(X_t)_{t\geq 0}$ on E satisfying: $\forall\, 0 \leq t_1 < t_2 < \cdots < t_n$, $P(X_{t_n} = i_n | X_{t_{n-1}} = i_{n-1}, \cdots, X_{t_1} = i_1, X_{t_0} = i_0) = P(X_{t_n} = i_n | X_{t_{n-1}} = i_{n-1}) = P_{t_n - t_{n-1}}(i_n, i_{n-1})$.

Remark. For $t > 0$, $P(X_t = j | X_0 = i) = P_t(i,j)$.

Example 16. MC along with PP

Suppose a MC $(\xi_n)_{n\geq 0}$ with 1-state transition matrix $Q \perp\!\!\!\perp$ a Posisson process $(N_t)_{t\geq 0}$ with rate λ. Let $X_t = \xi_{N_t}, t \geq 0$ in definintion (I) (a MJP with $\lambda(i) \equiv \lambda$).

Go back to definintion (II), $P_t(i,j) = \sum\limits_{n=0}^{\infty} e^{-\lambda t} \dfrac{(\lambda t)^n}{n!} Q^n(i,j)$.

Explanation: $P_t(i,j) = P(\xi_{N_t} = j | \xi_{N_0} = i) = \sum\limits_{n=0}^{\infty} P(\xi_n = j, N_t = n | \xi_0 = i) = \sum\limits_{n=0}^{\infty} P(\xi_n = j | \xi_0 = i) P(N_t = n)$.

Now here comes the common question between the two definintion: where are the parameters $\lambda_i, \forall\, i$ and Q in the information given from the second definintion.

Suppose Z_n is sampled to get values $\{na\}_{n=1}^{\infty}$, then, $Z_n = (X_{na})_{n\geq 0}$ is a MC. Nonetheless, with arbitrary discrete time $\{X_{t_n}\}_{n\geq 0}$, it is a time-inhomogeneous MC.

4.2 Chapman-Kolmogorov Equation

Theorem 11. Chapman-Kolmogorov Equation (general case)

The family of transition kernel $(P_s)_{s\geq 0}$ satisfies

$$P_{s+t} = P_s P_t = P_t P_s, \forall\, s,t \geq 0.$$

By convention, $P_0 = I$.

Remark. Applied to $(Z_n)_{n\geq 0}$ above we get Chapman-Kolmogorov equation for homogeneous discrete time MC in Chapter 2.

Proof. $P_{s+t}(i,j) = P(X_{s+t} = j | X_0 = i) = \sum\limits_{k\in E} P(X_{s+t} = j, X_t = k | X_0 = i) \overset{Bayes' thm}{=\!=\!=} \sum\limits_{k\in E} P(X_{s+t} = j | X_t = k, X_0 = i) P(X_t = k | X_0 = i) = P_s \circ P_t(i,j)$ ∎

Solve the functional equation given by Chapman-Kolmogorov equation (By taking *log* (if possible, e.g., true if P_t is diagonalizable), it will become Cauchy functional equation):

Note that we must assume P_t is continuous at $t = 0$, since there exist some dis-continuous functions satisfying Cauchy functional equation. With this continuity,

(1) P_t is a continuous function on $[0, \infty)$;

(2) $P_s = P_1^s$ for integers, rational numbers and then all real numbers on $[0, \infty)$, consequentially.

(If it is further assumed to be differentiable, the deduction is much easier: $P'_s = \lim_{t \to 0} \dfrac{P_{s+t} - P_s}{t} = P_s \lim_{t \to 0} \dfrac{P_t - I}{t} = P_s P'_0$, that is $P_t = Ce^{tP'_0}$. $P_s P_t = P_{s+t} \Rightarrow C = 1$. Let $G = P'_0$, we get the solution.)

To explicitly compute G, here we need one more lemma.

Lemma

Suppose that jump time $\{T_i\}$ satisfies:

(i) $T_0 = 0$,$T_1 = \inf\{t > 0 \big| X_t \neq X_0\}$;

(ii) $T_{n+1} = \inf\{t > T_n \big| X_t \neq X_{T_n}\}$, $\forall\, n \in \mathbb{N}$.

Assume that $\forall\, i \in E$, $P\left[T_1 < \infty | X_0 = i\right] = 1$, we have

(1) $\{X_{T_n}\}_{n \geq 0}$ is a Markov chain with transition matrix Q;

(2) Given $\{X_{T_n}\}_{n \geq 0}$, $T_1, T_2 - T_1, \cdots$, are mutually independent exponential distributions with rate parameters $\{\lambda(X_{T_0}), \lambda(X_{T_1}), \cdots\}$.

Remark. $\forall\, t > 0, i \neq j$, $P(T_1 > t, X_{T_1} = j \big| X_0 = i) = e^{-\lambda(i)t} Q(i, j)$.

Thus, we can deduce that $P_t = e^{tG}$, where G is a matrix named "generator" of the MJP.

$$G(i,j) = \begin{cases} \lim\limits_{t \to 0^+} \dfrac{P(X_t = j | X_0 = i)}{t} = \lambda(i) Q(i,j), & i \neq j \\ -\sum\limits_{l \neq i} G(i,l) = -\sum\limits_{l \neq i} \lambda(i) Q(i,l) = -\lambda(i), & i = j \end{cases}$$

Since it can be written as $\lim\limits_{t \to 0^+} \dfrac{P(X_t = j | X_0 = i)}{t} = \lim\limits_{t \to 0^+} \dfrac{P(T_1 < t, X_{T_1} = j | X_0 = i)}{t} = \lim\limits_{t \to 0^+} \dfrac{1 - e^{-\lambda(i)t}}{t} Q(i,j) \sim \lambda(i) Q(i,j)$, and $G(i,i) = \lim\limits_{t \to 0} \dfrac{P_t(i,i) - 1}{t} = -\lim\limits_{t \to 0} \dfrac{1}{t} \sum\limits_{j \neq i} P_t(i,j) = -\sum\limits_{l \neq i} G(i,l)$. (Note that $Q(i,i) = 0$ from lemma (I)).

Put G in "diagonalized"(Jordan) form: $J = U^{-1} G U$, with eigenvectors u_i, we have $e^{tG} u_j = e^{\lambda_j t} u_j$

4.2.1 Compute Generate Matrix from Transition Rate Diagram

We use an example to illustrate the method to use transition rate diagram.

Example 17

We have two operators and three phone lines. Calls arrival follows a Poisson process with rate $\lambda = 2$ calls/minute. John and Mary are the two operators. John's processing times are exponentially distributed with mean 6 minutes. Mary's processing times are exponentially distributed with mean 4 minutes. If each customer has a patience that is exponentially distributed with mean 10 minutes. An incoming call to an empty system always goes to John. Model this system by a CTMC.

The number on the arrow $i \to j$ is $G(i, j)$, and we can use the formula above for calculation. For instance, $J_{2,3} = P(X_{n+1} = 3|X_n = 2) = P(E_{2\to 3} < \min(E_{2\to 1M}, E_{2\to 1J})) = \frac{2}{2 + \frac{1}{4} + \frac{1}{6}}$, thus $G(2, 3) = \lambda(2)J(2, 3) = 2$, since $exp(\lambda(2)) = \min(E_{2\to 3}, E_{2\to 1M}, E_{2\to 1J})$.

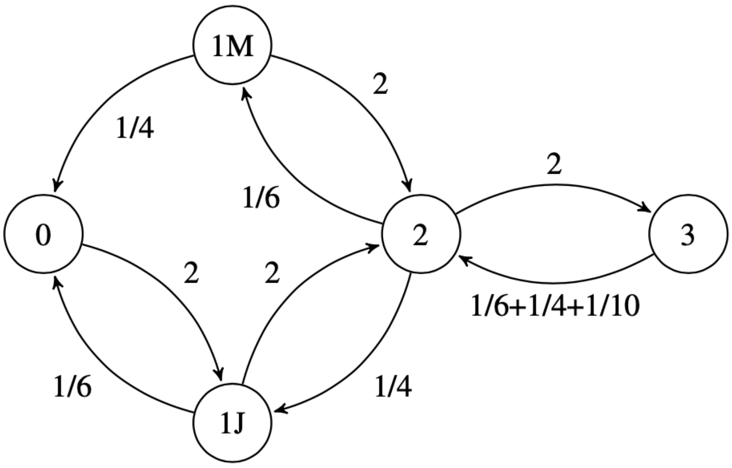


Figure 4.2: Illustration of Transition Rate Diagram

4.3 Limiting Behaviors & Stationary Distributions

The MJP $(X_t)_{t\geq 0}$ is called *irreducible/recurrent/transient/positive recurrent* if the embedded MC with transition matrix $Q = J$, (J is the jump matrix of MJP) being *irreducible/recurrent/transient/positive recurrent.*

Definition: Stationary Distribution

$\boldsymbol{\pi}$ is called a **stationary/invariant distribution** of MJP if $\forall\ t \geq 0$, $\boldsymbol{\pi} P_t = \boldsymbol{\pi}$ and $\sum\limits_{j\in E} \pi_j = 1$.

Lemma

$\boldsymbol{\pi}$ is a **stationary/invariant distribution** of MJP *exactly when* $\boldsymbol{\pi} G = 0$, where G is the generating matrix.

Proof. ($\Rightarrow$) Take derivatives;

($\Leftarrow$) Use the definintion of `expm`. ■

Next, we restrict to the case where only positive recurrent and irreducible MJPs are considered. Hence, it must have stationary probabilities.

Moreover, there is some limiting probability: $\frac{1}{T}\int_0^T \mathbb{1}_{\{X_t=i\}} dt = \frac{N_T(i)}{T} \to \pi(i)$ as $T \to \infty$ (special case with characteristic functions in SLLN).

To compute $\boldsymbol{\pi}$ with $\boldsymbol{\pi} G = \mathbf{0}$:

$$\sum_{k \neq i} \pi(k) G(k,i) = \pi(i) \sum_{k \neq i} G(i,k).$$

(Namely, for every state *total-in rate* = *total-out rate*, since at long run, stationary % $\pi(i)$ time X_t state at i).

Theorem 12. Ergodic Theorem/Strong Law of Large Numbers(SLLN)

Assume that a MJP $(X_t)_{t \geq 0}$ is **irreducible and positive** recurrent and $f : E \to \mathbb{R}$ is a real measurable **bounded** function.

$$P\left\{ \lim_{T \to \infty} \frac{1}{T} \int_0^T f \circ X(t) dt = \int_E f d\pi \right\} = 1,$$

where π is the unique stationary distribution.

Remark. (1) It can also be understood as *Long-run time average = Space average.*

(2) Note that the state space E is discrete here with measure π (which directly makes f simple functions or countable sum of characteristic/indicator functions), thus LHS above can be written as $\sum_{i \in E} f(i) \pi_i$.

(3) $f(i)$ = "cost" or "reward" for being in state i.

4.4 Queuing Theory

One vital application of MJP is to use them in some simple queues.

4.4.1 Introduction to Queues

Queue system: "**interarrival**" time (F as its distribution, i.e., $A_1, A_2 \cdots \sim F$ i.i.d) $\perp\!\!\!\perp$ "**service**" time (E as its distribution, i.e., $B_1, B_2 \cdots \sim G$ i.i.d). X_t denotes # people in this queue at time t.

Classification of queues (due to D. Kendall): (notation) $X/Y/k$ queue with $X, Y \in \{M, G\}$, where $X = \text{dist}F$ and $Y = \text{dist}E$. M = "*exponential* (refer to 'Markovian')$''$ and G = "*general*$''$ and k denotes # of servers. (e.g., M/G/k queue means that F is exponentially and E is arbitrarily distributed, with total k servers.)

4.4.2 Simple Queues as MJP

Example 18. M/M/1 Queues

Suppose we have an M/M/1 queue, meaning that we have Poisson arrival process with rate λ arrivals per minute and service times are i.i.d. exponentially distributed with rate μ. Assume that the buffer size is infinite. Let $X(t)$ be the number of customers in the system at time t. Then, $X = \{X(t) : t \geq 0\}$ is a CTMC with state space $S = \{0, 1, 2, \cdots\}$.

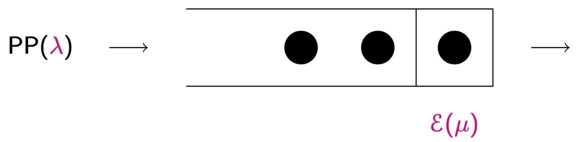


Figure 4.3: Illustration of M/M/1 Queuing Systems

To find the transition rate diagram is simple for this case.

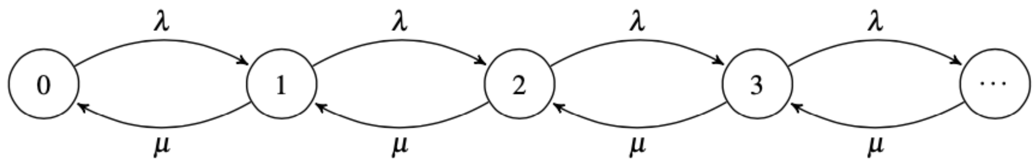


Figure 4.4: Transition Rate of M/M/1 Queuing Systems

Rate parameters: $\lambda(0) = \lambda$ and $\lambda(i) = \lambda + \mu, i \geq 1$.

Jump Matrix: $Q(0,1) = 1$, $Q(i, i-1) = \dfrac{\mu}{\mu + \lambda}$ and $Q(i, i+1) = \dfrac{\lambda}{\lambda + \mu}$.

Compare with arrival and service rates in M/M/1 queue to check recurrence and transience of $\{X_{(t)}\}$:

$$\begin{cases} \lambda < \mu, & \text{positive recurrent} \\ \lambda = \mu, & \text{null recurrent} \\ \lambda > \mu, & \text{transient} \end{cases}$$

Use rate-balanced equation, $\pi_n = \left(\dfrac{\lambda}{\mu}\right)^n \left(1 - \dfrac{\lambda}{\mu}\right)$.

Applications: long-run statistics (valid for bounded and nonnegative f): e.g., $f(x) = x : \mathbb{N} \to \infty$. Long run average of (X_t), that is # of people in the system, is

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f\circ X(t)dt \sim \int_E x d\pi = \sum_{n\geq 0} n\pi_n = \frac{\lambda}{\mu-\lambda}$$

.

Example 19. M/M/k Queues

Suppose we have an M/M/k queue, i.e., all conditions are the same as the M/M/1 queue except for there are total k servers.

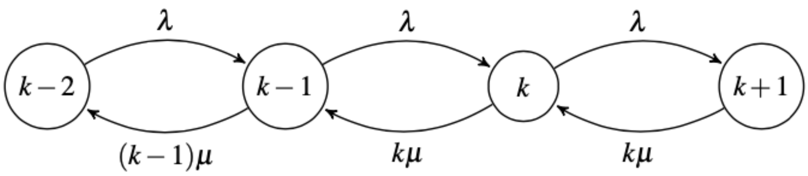


Figure 4.5: Transition Rate of M/M/k Queuing Systems

Rate parameters: $\lambda(i) = \lambda + \min\{i,k\}\mu$, $\forall\ i$.

Jump Matrix: $Q(i,i-1) = \dfrac{\min\{i,k\}\mu}{\min\{i,k\}\mu+\lambda}$ and $Q(i,i+1) = \dfrac{\lambda}{\lambda+\min\{i,k\}\mu}$, $\forall\ i$.

Compare with arrival and service rates (only large ones) in M/M/k queue to check recurrence and transience of $\{X_t\}$:

$$\begin{cases} \lambda < k\mu, & \text{positive recurrent} \\ \lambda = k\mu, & \text{null recurrent} \\ \lambda > k\mu, & \text{transient} \end{cases} .$$

Example 20. M/M/∞ Queues

Suppose we have an M/M/∞ queue, i.e., all conditions are the same as the M/M/1 queue except for there are total ∞ servers.

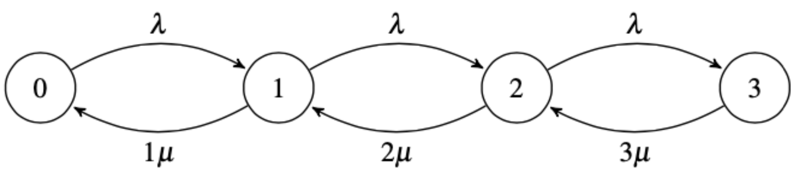


Figure 4.6: Transition Rate of M/M/∞ Queuing Systems

Rate parameters: $\lambda(i) = \lambda + i\mu, \forall\ i$.

Jump Matrix: $Q(i,i-1) = \dfrac{i\mu}{i\mu+\lambda}$ and $Q(i,i+1) = \dfrac{\lambda}{\lambda+i\mu}$, $\forall\ i$.

Always positive recurrent. Stationary distribution: $\pi_k = \frac{\lambda^k}{k!\mu^k}e^{-\frac{\lambda}{\mu}}, \forall\ k$. (Namely, $\boldsymbol{\pi} \sim Pois(\lambda/\mu)$)

4.4.3 Queuing Networks

Example 21. Tandem Queues

Suppose two queues are *connected in tandem* meaning that the queues are in series. Jobs are still arriving at rate α jobs per minute. The first queue's mean processing time is μ_1 and the second queue's mean processing time is μ_2 all $\perp\!\!\!\perp$. Define $X(t) = (X_1(t), X_2(t))$ where $X_i(t)$ is the number of jobs at station i at time t.

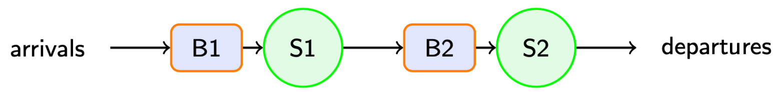


Figure 4.7: Illustration of Tandem Queuing Systems

For tandem queues, $(X_1(t), X_2(t))$ is a MJP with states in $\mathbb{N}^2$.

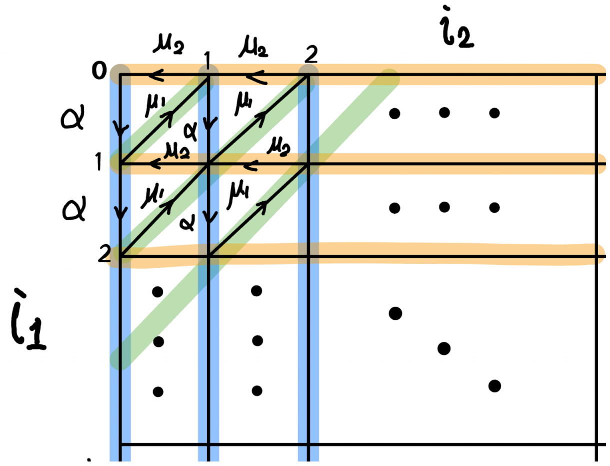


Figure 4.8: Transition Rates of Tandem Queuing Systems

Rate parameters: $\lambda(i_1, i_2) = \alpha + \mathbb{1}_{\{i_1 \geq 1\}}\mu_1 + \mathbb{1}_{\{i_2 \geq 1\}}\mu_2,\ \forall\ i_1, i_2$.

$$\text{Jump Matrix: } \begin{cases} Q[(i_1, i_2), (i_1+1, i_2)] = \dfrac{\alpha}{\lambda(i_1, i_2)}, \\ Q[(i_1, i_2), (i_1-1, i_2+1)] = \dfrac{\mathbb{1}_{\{i_1 \geq 1\}}\mu_1}{\lambda(i_1, i_2)}, \quad \forall\ i_1, i_2\,. \\ Q[(i_1, i_2), (i_1, i_2-1)] = \dfrac{\mathbb{1}_{\{i_2 \geq 1\}}\mu_2}{\lambda(i_1, i_2)}, \end{cases}$$

Remark. The tandem queue $\{X_t\}$ is positive recurrent if $\rho_1 = \frac{\alpha}{\mu_1} < 1$ and $\rho_2 = \frac{\alpha}{\mu_2} < 1$. It is also possible to find stationary distribution $(\pi(i_1, i_2))_{i_1 \geq 0, i_2 \geq 0}$, which is (from rate-balance equation)

$$\pi(i_1, i_2) = (1-\rho_1)\rho_1^{i_1}(1-\rho_2)\rho_2^{i_2} = \pi_{\rho_1} \otimes \pi_{\rho_2}.$$

[Decomposition]. Let $(I, J) \sim \boldsymbol{\pi}$ (stationary distribution of the tandem queue), then

(i) $I \perp\!\!\!\perp J$

(ii) $\begin{cases} I \sim \boldsymbol{\pi_1} \text{ of M/M/1 with } (\alpha, \mu_1), \\ J \sim \boldsymbol{\pi_2} \text{ of M/M/1 with } (\alpha, \mu_2). \end{cases}$

The tandem queue is irreducible (Irreducibility implies **at most** 1 invariant probability distribution.), but period 3, (recall that M/M/1 queue has period 2).

Example 22. Open Jackson Networks

A generalization to tandem queues: Totally there are J stations. For each station $j \in \{1, \cdots, J\}$:

- external arrivals $\sim$ P.P.(α_j)
- n_j servers; i.i.d. exponential service times with rate μ_j at station j
- unlimited waiting rooms
- When a job completes service, it is either routed to station k with probability P_{jk} , or exits the system with probability $1 - \sum_k P_{jk}$.

The routing matrix P is transient or $(I - P)$ is invertible.

Here we consider a simple three-station case with routing. There are two inspections at the end of station 2 and 3. Failure rates are 30% (route to 1) and 20% (route to 2), respectively. Then,

$$\begin{cases} \lambda_1 = \alpha + 0.3\lambda_2 \\ \lambda_2 = \lambda_1 + 0.2\lambda_3 \\ \lambda_3 = 0.7\lambda_2 \end{cases}.$$

Note that generally, we have $\boldsymbol{\lambda}^T = \boldsymbol{\alpha}^T + \boldsymbol{\lambda}^T P$. Eventually, we can compute $\pi(i_1, i_2, i_3) = \pi_{\rho_1} \otimes \pi_{\rho_2} \otimes \pi_{\rho_3}$, with $\rho_i = \dfrac{\lambda_i}{\mu_i}$, $i = 1, 2, 3$.

4.4.4 Little's Law

Suppose

- $\overline{L}$ = long-run average number of customers in the queue/system;
- λ = long-run average arrival rate (or throughput of the system);
- $\overline{W}$ = long-run average amount of time a customer waits in the queue/system.

Little's law promises the existence of the above quantities and shows their relations.

Theorem 13. Little's Law

If two quantities exist (well defined), the third quantity also exists. Furthermore, they satisfy

$$\overline{L} = \lambda \cdot \overline{W}.$$

(i.e.,*average number of customers in the queue =arrival rate* $\times$ *average waiting time*)

Remark. Above queues discussed satisfy the Little's law.

Chapter 5

Brownian Motion

5.1 Recap of Multivariate Normal Distributions

Proposition: Covariance Matrix

Let $\boldsymbol{X}$ be an $n \times 1$ **random vector** (no need to be Guassian), $\boldsymbol{\mu} = \mathbb{E}(\boldsymbol{X})$ and $\Sigma = Cov(\boldsymbol{X}) = \mathbb{E}[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T]$.

(i) Σ is symmetric and $\Sigma_{ij} = Cov(X_i, X_j)$.

(ii) Σ is positive semi-definite (i.e., $\Sigma \succeq 0$).

(iii) Let $A \in M_{m\times n}(\mathbb{R})$ and $\boldsymbol{b} \in \mathbb{R}^m$, $Cov(A\boldsymbol{X} + \boldsymbol{b}) = A\Sigma A^T$

Proof. Only consider (ii). $0 \le Cov(\boldsymbol{y}^T\boldsymbol{X}) \overset{\text{(iii)}}{=\!=} \boldsymbol{y}^T\Sigma\boldsymbol{y}$, $\forall\ \boldsymbol{y} \in \mathbb{R}^n$. ■

Definition: Multivariate Normal Distributions

Let $\boldsymbol{X}$ be an $n \times 1$ **random vector**, following multivariate normal distribution (denoted by $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \Sigma)$), if its p.d.f. is given by

$$f(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n|\Sigma|}} exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right\}, \forall\ \boldsymbol{x} \in \mathbb{R}^n,$$

where $\Sigma \in M_{n\times n}(\mathbb{R})$ is a *positive definite* matrix, and $|\Sigma| = det(\Sigma)$.

Corollary. In *a multi-variate normal vector* $\boldsymbol{X}$, its entries X_1, X_2 are independent if and only if $Corr(X_1, X_2) = 0$ (or $\Sigma_{ij} = 0$). (Uncorrelation in the multi-variate normal distributions implies independence.)

Properties of Multivariate Normal:

- **Linear Transformation**: Let $A \in M_{m\times n}(\mathbb{R})$ and $b \in \mathbb{R}^m$,

$$A\boldsymbol{X} + \boldsymbol{b} \sim N(A\boldsymbol{\mu} + \boldsymbol{b}, A\Sigma A^T).$$

  (proof with change of variables or just use m.g.f)

- $\mathbb{E}(\boldsymbol{X}) = \boldsymbol{\mu}$, $Cov(\boldsymbol{X}) = \mathbb{E}[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})^T] = \Sigma$.

  Its m.g.f. is $M_t(\boldsymbol{X}) = E(e^{\boldsymbol{X}^T \boldsymbol{t}}) = exp\left[\boldsymbol{\mu}^T \boldsymbol{t} + \frac{1}{2}\boldsymbol{t}^T \Sigma \boldsymbol{t}\right]$.

- Positive definiteness promises the existence of $\Sigma^{\frac{1}{2}} = Qdiag\{\sigma_1, \cdots, \sigma_n\}Q^T$, assume eigen-decomposition of Σ is $Qdiag\{\sigma_1^2, \cdots, \sigma_n^2\}Q^T$.

  Thus, by transformation $\boldsymbol{Z} = \Sigma^{-\frac{1}{2}}(\boldsymbol{X} - \boldsymbol{\mu})$ (standardization), we get the **standard multi-variate normal** $\boldsymbol{X} \sim N(\boldsymbol{0}, I_n)$.

5.1.1 Marginal and Conditional Distributions

Here are the general results for marginal and conditional distributions.

Proposition: Marginal Distributions of Multivariate Normal

Partition X into blocks $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X_1} \\ \boldsymbol{X_2} \end{bmatrix}$, $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix}$ of dimensions n_1, n_2, $n_1 + n_2 = n$.

- $\boldsymbol{X_1} \sim N(\boldsymbol{\mu_1}, \Sigma_{1,1})$ and $\boldsymbol{X_2} \sim N(\boldsymbol{\mu_2}, \Sigma_{2,2})$
- In particular, $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ are independent iff $\Sigma_{1,2} = \mathbf{O}_{n_1 \times n_2}$.

Proof. Let $A_1 = \begin{bmatrix} I_{n_1} & O_{n_1 \times n_2} \end{bmatrix}$, then $\boldsymbol{X_1} = A_1 \boldsymbol{X} \sim N(\boldsymbol{\mu_1}, \Sigma_{1,1})$, similarly for $\boldsymbol{X_2}$.

For independence, $\mathbb{E}[(\boldsymbol{X_1} - \boldsymbol{\mu_1})(\boldsymbol{X_2} - \boldsymbol{\mu_2})^T] = \Sigma_{1,2}$. ∎

Proposition: Conditional Distributions of Multivariate Normal

Partition X into blocks $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X_1} \\ \boldsymbol{X_2} \end{bmatrix}$, $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{bmatrix}$ and $\Sigma = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix}$ of dimensions n_1, n_2, $n_1 + n_2 = n$.

$\boldsymbol{X_1}|\boldsymbol{X_2} = \boldsymbol{x_2} \sim N(\boldsymbol{\mu_1} + \Sigma_{1,2}\Sigma_{2,2}^{-1}(\boldsymbol{x_2} - \boldsymbol{\mu_2}), \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1})$

Proof. The proof uses a *useful Gaussian trick.* In order to find the distribution of $\boldsymbol{X_1}|\boldsymbol{X_2} = \boldsymbol{x_2}$ directly, we first find matrix C such that $\boldsymbol{X_1} + C\boldsymbol{X_2}$ and $\boldsymbol{X_2}$ are independent. It is easy to find this C as $-\Sigma_{1,2}\Sigma_{2,2}^{-1}$.

With this C, $\boldsymbol{X_1}|\boldsymbol{X_2} = \boldsymbol{x_2}$ has the same distribution as $\boldsymbol{X_1} + C\boldsymbol{X_2} - C\boldsymbol{x_2}|\boldsymbol{X_2} = \boldsymbol{x_2}$, which is just

$$\boldsymbol{X_1} + C\boldsymbol{X_2} - C\boldsymbol{x_2} = \begin{bmatrix} I_{n_1} & C \end{bmatrix} \boldsymbol{X} - C\boldsymbol{x_2} \sim N(\boldsymbol{\mu_1} + C(\boldsymbol{\mu_2} - \boldsymbol{x_2}), \Sigma_{1,1} + C(\Sigma_{2,1} + \Sigma_{1,2}) + C^2\Sigma_{2,2}).$$

∎

5.2 Introduction to Brownian Motions

Historic Remarks:

- 1827: A botanist, *Robert Brown*, observed the **erratic motion** of grains of **pollen suspended in a liquid**. Later it was found such irregular motion comes from the extremely large number of collisions of the suspended pollen grains with molecules of the liquid.
- 1905: Physicist *Albert Einstein* investigated on the physics and mathematics theory (by proposing **random force**) of the random movement discovered by Robert Brown. Such movement was later named as the **Brownian Motion**.
- 1923: Mathematician *Norbert Wiener* completed the mathematical theory of Brownian motion. The standardized version of Brownian motion is also named the **Wiener process**.

Definition: Brownian Motion

A continuous time stochastic process $\{X_{(t)} : t \geq 0\}$ taking values in $\mathbb{R}$ is called a **Brownian motion** if the following conditions are satisfied.

1. $X_{(0)} = 0$;
2. The process has **stationary and independent** increments;
3. For every $t > 0$, $X_{(t)} \sim N(0, \sigma^2 t)$;
4. The sample path $X_{(t)}$ is continuous everywhere with probability 1.

Remark. (i) A Brownian motion with variance parameter $\sigma^2 = 1$ is called a **standard Brownian Motion** (or **Wiener Process**), usually denoted as $\{B_{(t)} : t \geq 0\}$ (or $\{W_{(t)} : t \geq 0\}$).

(ii) A standard **planar** or **2-d Brownian motion** is a two component process $\{(X_{(t)}, Y_{(t)}) : t \geq 0\}$ where $\{X_{(t)} : t \geq 0\}$ and $\{Y_{(t)} : t \geq 0\}$ are two **independent standard Brownian motions**.

5.2.1 Brownian Motion as Continuous Limit of Simple Random Walk

Consider a symmetric random walk on $\mathbb{Z}$:

$$X_n = Y_1 + Y_2 + \cdots + Y_n,$$

where Y_i are i.i.d. and $P(Y_i = \pm 1) = \frac{1}{2}$. Then, $P(X_n = k) = \dfrac{1}{2^n} \dfrac{n!}{(\frac{n+k}{2})!(\frac{n-k}{2})!}$, $\forall$ k such that $k \equiv n \pmod 2$]. $\mathbb{E}(X_k) = 0$ and $\mathrm{Var}(X_n) = n$.

Moreover, **independence of the increments** (i.e., $\forall$ $n > m \geq 0$, $X_n - X_m$ is independent of $X_1, \cdots, X_m$, hence from $X_u - X_v$ for $v < u < m$) and **stationary increments** (i.e., $\forall$ $n > m \geq 0$, $X_n - X_m$ has the same distribution as X_{n-m}) both hold.

Define then

$$B_{(t)}^{(n)} = \frac{1}{\sqrt{n}} X_{\lfloor nt \rfloor}, t \in \mathbb{R}, n \in \mathbb{N},$$

that is, the time is accelerated with a factor n while the space is compressed by a factor $\sqrt{n}$. Eventually, define

$$\{B_{(t)}\}_{t \geq 0} = \{\lim_{n \to \infty} B_{(t)}^{(n)}\}_{t \geq 0},$$

we conclude that $\{B_{(t)}\}_{t \geq 0}$ is the standard Brownian motion.

(***Explanations***): Firstly, taking limit perseves independent and stationary increments (need check!). What's more, invoking CLT, $B_{(t)}^{(n)} \sim N(0, t)$ (aslo check continuity of path).

5.2.2 Basic Properties of Brownian Motions

[***Property I***] The sample paths (motion trace) are **continuous** functions of t but **nowhere differentiable**.

To understand it, $\dfrac{B_{(t+h)} - B_{(t)}}{h} = N\left(0, \dfrac{1}{h}\right)$, for any $t, h > 0$, cannot take $h \to 0^+$.

[***Property II***] (**Markov Property**) $\mathcal{D}(B_{(t)} | B_{(u)}, 0 \leq u \leq s) = \mathcal{D}(B_{(t)} | B_{(s)})$, $s \leq t$.

[***Property III***] (*Joint Distribution*) Let $0 < t_1 < \cdots < t_n$, the joint distribution

$$(B_{(t_1)}, \cdots, B_{(t_n)}) \sim N(\mathbf{0}, \Sigma), \text{ where } (\Sigma)_{ij} = \min(t_i, t_j), \forall\ 1 \leq i, j \leq n.$$

Proof. $\Delta_i = B_{(t_i)} - B_{(t_{i-1})}$, $2 \leq i \leq n$ and $\Delta_1 = B_{(t_1)}$. They are independent and thus $(\Delta_1, \cdots, \Delta_n) \sim N(\mathbf{0}, \operatorname{diag}(\delta_1, \cdots, \delta_n))$, where $\delta_i = t_i - t_{i-1}$ and $\delta_1 = t_1$.

Since $(B_{(t_1)}, \cdots, B_{(t_n)}) = A(\Delta_1, \cdots, \Delta_n)$ for some $A \in M_{n \times n}(\mathbb{R})$, the target is normal and $\mathbb{E}((B_{(t_1)}, \cdots, B_{(t_n)})) = 0$ and $\operatorname{Cov}(B_{(t_i)}, B_{(t_j)}) = \operatorname{Var}(B_{\min(t_i, t_j)}) = \min(t_i, t_j)$. ∎

[***Property IV***] (*Conditional Distribution*). Let $0 < s \leq t$

(i) (**Forecasting**) $B_{(t)} | B_{(s)} = x \sim N(x, t - s)$;

(ii) (**Backcasting**) $B_{(s)} | B_{(t)} = y \sim N\left(\frac{s}{t} y, \frac{s}{t}(t - s)\right)$

Proof. Use conditional distributions of multi-variate normal. ∎

[***Property V***] (*Standard Brownian Motion*) Let $\left\{B_{(t)}\right\}_{t \geq 0}$ be a standard Brownian motion (SBM).

(i) (**Symmetric**) $\left\{-B_{(t)}\right\}_{t \geq 0}$ is also a SBM.

(ii) (**Scaling**) $\left\{cB_{(t/c^2)}\right\}_{t \geq 0}$ are SBMs, $\forall\ c \neq 0$.

(iii) (**Time Reversal**) $\left\{X_{(t)}\right\}_{t \geq 0}$ defined by $X_{(t)} = \begin{cases} 0, & t = 0 \\ tB_{(1/t)}, & t > 0 \end{cases}$ is a SBM.

Proof. Only for (iii), check 3 axioms of BM. Independence: let $s_2 > s_1 > t_2 > t_1$, w.l.o.g. $X_{(s_2)} - X_{(s_1)} = s_2 B_{(1/s_2)} - s_1 B_{(1/s_1)} \perp\!\!\!\perp X_{(t_2)} - X_{(t_1)} = t_2 B_{(1/t_2)} - t_1 B_{(1/t_1)}$. Since Bs are joint normal, we can get $X_{(s_2)} - X_{(s_1)}$ and $X_{(t_2)} - X_{(t_1)}$ can be obtained from joint normal (thus normal) as $s_2 t_2 \frac{1}{s_2} - s_1 t_2 \frac{1}{s_1} - s_2 t_1 \frac{1}{s_2} + s_1 t_1 \frac{1}{s_1} = 0$

For distribution of $X_{(t)} - X_{(s)} = tB_{(1/t)} - sB_{(1/s)} \sim N(0, t-s)$, since $\mathrm{Var}(tB_{(1/t)} - sB_{(1/s)}) = t + s - 2ts\frac{1}{t}$. ■

5.3 Hitting Time and Maximum Variable

Definition: Hitting Time

Let $\{B_{(t)} : t \geq 0\}$ be a standard Brownian motion.

For any $a > 0$, Let T_a, be the **first time the Brownian motion hits** a, i.e.

$$T_a = \inf\{t \geq 0 : B_{(t)} \geq a\}.$$

By continuity, it implies that $B_{(T_a)} = a$ and $B_{(t)} < a$ for any $t < T_a$.

Theorem 14. Reflection Principle

The stochastic process $\{B^{\star}_{(t),a} : t \geq 0\}$ defined by

$$B^{\star}_{(t),a} = \begin{cases} B_{(t)}, & t \leq T_a \\ 2a - B_{(t)}, & t > T_a \end{cases}$$

is also a *standard Brownian motion*.

Figure 5.1: Illustration of Reflection Principle

Definition: Running Maximum

The running maximum $\{M_{(t)}\}$ of a Brownian motion $\{B_{(t)}\}$ is defined by

$$M_{(t)} = \max_{0\leq s\leq t} B_{(s)}, t \geq 0.$$

Lemma: Joint Distribution of $(B_{(t)}, M_{(t)})$

For $0 < b \leq a$, $P(M_{(t)} \geq a, B_{(t)} \leq b) = P(B_{(t)} \geq 2a - b)$.

Proof. Suppose $M_{(t)} \geq a$, due to continuity $T_a \leq t$. Thus, $B^{\star}_{(t),a}$ is a SBM. $P(M_{(t)} \geq a, B_{(t)} \leq b) = P(T_a \leq t, B_{(t)\leq b}) = P(B^{\star}_{(t),a} \geq 2a - b) = P(B_{(t)} \geq 2a - b)$, due to the reflection principle. ■

Remark. Here are some corollaries from above lemma.

- The running maximum $M_{(t)}$ has the same distribution as $|B_{(t)}|$. In particular, its CDF is
$$P(M_{(t)} \leq x) = 2\phi\left(\frac{x}{\sqrt{t}}\right) - 1, x > 0.$$
- The hitting time T_a has the same distribution as $\dfrac{a^2}{B^2_{(1)}}$. In particular, its CDF is
$$P(T_a \leq t) = 2\left[1 - \phi\left(\frac{|a|}{\sqrt{t}}\right)\right], t > 0,$$
which is called a *Lévy distribution*.

Proof. Consider $P(M_{(t)} > x) = \displaystyle\int_x^\infty P(M_{(t)} = s)ds = \int_x^\infty P(M_{(t)} = s, B_{(t)} \leq s)ds$. Since $P(M_{(t)} = x, B_{(t)} \leq y) = -\dfrac{d}{dx}P(M_{(t)} \geq x, B_{(t)} \leq y) = -\dfrac{d}{dx}P(B_{(t)} \geq 2x - y) = \dfrac{2}{\sqrt{t}}\phi'\left(\dfrac{2x-y}{\sqrt{t}}\right)$. Thus, $P(M_{(t)} > x) = 2 - 2\phi\left(\dfrac{x}{\sqrt{t}}\right)$, which gives the solution. (Simple way: $P(M_{(t)} \geq x) = P(M_{(t)} \geq x, B_{(t)} \leq x) + P(M_{(t)} \geq x, B_{(t)} \geq x) = 2P((B_{(t)} \geq x))$.)

As for $P(T_a \leq t)$, it's equivalent to $P(M_{(t)} \geq a)$, for any $a > 0$ (thus it becomes $P(|tB_{(1)}| \geq a) = P\left(\dfrac{a^2}{B^2_{(1)}} \leq t\right)$). ■

Example 23. Brownian Motion and Gambler's MC

Find the probability of Brownian motion hitting A before $-B$ where $A > 0$, $B > 0$, (i.e., $P(T_A \leq T_{-B})$).

Understand BM as the limit of symmetric RW. Thus, with $-B$ and A be 2 absorbing states (9), it becomes a gambling model. $P(T_A \leq T_{-B}) = \dfrac{B}{A+B}$.

5.4 Variations on Brownian Motions

For Brownian motion, the rate of change in mean is zero and the rate of change in variance is a constant σ^2. In reality, the mean and variance of a stochastic process can evolve over time. In this section, we will introduce some important variations of Brownian motion.

5.4.1 Brownian Motion with Drift

Brownian motion with drift coefficient is a simple variant of standard BM, with the rate of change in mean becoming a non-zero constant.

Definition: Brownian Motion with Drift

$\{X_{(t)} : t \geq 0\}$ is a **Brownian motion process with drift** coefficient $\mu \in \mathbb{R}$ and variance parameter $\sigma^2 > 0$ is defined asif

$$X_{(t)} = \mu t + \sigma B_{(t)}.$$

The discretized version of Brwonian motion with drift is

$$\Delta X_{(t)} = X_{(t+\Delta t)} - X_{(t)} = \mu \Delta t + \sigma \Delta B_{(t)};$$

and the continuous version of Brwonian motion with drift is

$$dX_{(t)} = \mu dt + \sigma dB_{(t)}.$$

We can see that for Brownian motion with drift, the rate of change in mean is μ and the rate of change in variance is σ^2.

Conventionally, μ and σ are referred to as the **drift** and **volatility** parameters, respectively.

(*Joint Distribution*) Let $0 < t_1 < \cdots < t_n$, the joint distribution

$$(X_{(t_1)}, \cdots, X_{(t_n)}) \sim N(\mu(t_1, \cdots, t_n), \sigma^2 \Sigma), \text{ where } (\Sigma)_{ij} = \min(t_i, t_j), \forall\, 1 \leq i, j \leq n.$$

(*Conditional Distribution*) Let $0 < s \leq t$.

(i) (**Forecasting**) $X_{(t)} | X_{(s)} = x \sim N(x + \mu(t-s), \sigma^2(t-s))$;

(ii) (**Backcasting**) $X_{(s)} | X_{(t)} = y \sim N\left(\frac{s}{t} y, \sigma^2 \frac{s}{t}(t-s)\right)$

The Brownian motion with drift process can be generalized to the **Ito (diffusion) process**:

$$dX_t = \mu(X_{(t)}, t)dt + \sigma(X_{(t)}, t)dB_{(t)}.$$

That is, both drift term and volatility term are allowed to depend on $X_{(t)}$ and t.

5.4.2 Geometric Brownian Motion

Geometric Brownian motion is another simple variant of standard BM, using the exponential of Brownian motion with drift. Geometric Brownian motion is widely used to model the stock price at time t.

Definition: Geometric Brownian Motion

$\{S_{(t)} : t \geq 0\}$ is a **Geometric Brownian motion process** if

$$S_{(t)} = S_{(0)} \exp\left[\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_{(t)}\right].$$

Remark. Let $Z_{(t)} = \ln S_{(t)}$. Then, given $S_{(0)} = S_0$, $(Z_{(t)})_{t\geq 0}$ is a Brownian motion with drift process.

Lemma: Ito's Lemma

Suppose that $G(X_{(t)}, t)$ is differentiable, where $(X_{(t)})$ is a Ito's process. Then,

$$dG = \left[\mu(X_{(t)}, t)\frac{\partial G}{\partial X_{(t)}} + \frac{\partial G}{\partial t} + \frac{\sigma^2(X_{(t)}, t)}{2}\frac{\partial^2 G}{\partial X_{(t)}^2}\right]dt + \left[\sigma(X_{(t)}, t)\frac{\partial G}{\partial X_{(t)}}dB_{(t)}\right].$$

Geometric Brownian motion has the following continuous version:

$$dS_{(t)} = \mu S_{(t)}dt + \sigma S_{(t)}dB_{(t)}.$$

This implies that μ is the expected (relative) return of the stock per unit time, and σ is the standard deviation (or volatility) of the (relative) return of the stock per unit time. We shall highlight that σ is not the volatility of the stock price itself.

(*Moment Properties.*) Suppose $0 < l \leq t$,

$$\mathbb{E}[S_{(t)}|S_{(u)}, 0 \leq u \leq l] \xlongequal{Markovian} S_{(l)}\mathbb{E}[e^{X_{(t)}-X_{(l)}}],$$

where $X_{(s)}$ is the Brownian motion with drift $\mu - \dfrac{\sigma^2}{2}$ and volatility σ.

$\mathbb{E}(S_{(t)}) = S_{(0)}e^{\mu t}$, and $\mathrm{Var}(S_{(t)}) = S_{(0)}^2 e^{2\mu t}\left(e^{\sigma^2 t} - 1\right)$.

5.4.3 Application: Black-Scholes Formula

The aim of developing Black-Scholes formula is to determine a **fair price of an option** at any time t (for American option) with no **arbitrage** allowed to emerge, given **expiration date** T and **strike price** K. Idea behind is **pricing by replication**: *if there is a portfolio that perfectly replicates the option pay-off, then its value should be the option price.*

Terminologies:

- **Call option**: gives its holder the right nut not obligation to *buy a certain amount* of an underlying asset by a certain date for a certain price;
- **Put option**: gives its holder the right but not obligation to *sell a certain amount* of an underlying asset by a certain date for a certain price;
- **Expiration/Maturity date**: the date that an option expires;
- **Exercise/Strike price**: the buying or selling price;
- **American option**: may be exercised at any time before maturity;
- **European option**: can be exercised only at the maturity date.

The price in a **risk-neutral world**: $\mathbb{E}^Q[\text{NPV}|\text{past information}] = \text{current price}$. In the risk-neutral world, arbitrage will never occur.

BS-Formula Derived from PDE (Recap: Modeling (MAT3300))

Under the multi-period binomial model, there is a risk-free asset with one-period interest rate $r\Delta t$. There are n time steps ($n\Delta t = T$).

Assume further $d = 1 + r\Delta t - \sigma\sqrt{\Delta t} < 1 + r\Delta t < u = 1 + r\Delta t + \sigma\sqrt{\Delta t}$, which means the price of the stock (or other derivatives) will become dS_0 or uS_0 after one period Δt.

The probability to go up in the *risk-neutral world* should be $p = \dfrac{1 + r - d}{u - d} = \dfrac{1}{2}$ (which is also the price of the portofolio perfectly replicated the option-payoff).

$\begin{cases} C^+ = C(t_0 + \Delta t, uS_0) \\ C^- = C(t_0 + \Delta t, dS_0) \end{cases}$. The pricing must satisfy $C = \dfrac{1}{1 + r\Delta t} \cdot \dfrac{1}{2}(C^+ + C_-)$.

Up to $O((\Delta t)^{\frac{3}{2}})$, this can be transformed with Taylor expansion to be Black-Scholes PDE:

$$\begin{cases} C_t + rSC_S + \dfrac{1}{2}\sigma^2 S^2 C_{SS} - rC = 0 \\ C(T, S) = (K - S)_+ \end{cases}.$$

BS-Formula as Geometric BM

Note that the price of financial derivatives can be modified as geometric Brownian motion $(S_{(t)})_{t\geq 0}$. Let $r > 0$ be the fixed interest rate, we have then for any future price $S_{(t)}$, its present value is $e^{-rt}S_{(t)}$.

Conventionally, a fair price of the option is

$$C(t) = \mathbb{E}^Q[e^{r(t-T)} \cdot \text{Pay-off of the option}].$$

Thus, $C(t) = \mathbb{E}^Q[e^{r(t-T)} \cdot (S_T - K)_+] = e^{-r(T-t)}\mathbb{E}^Q[\cdot(S_{(t)}e^{X_{(T)}-X_{(t)}} - K)_+]$
$= e^{-r(T-t)} \int_{\ln(K/S_{(t)})}^{\infty} (S_{(t)}e^y - K) \frac{1}{\sqrt{2\pi(T-t)}\sigma} \exp\left[-\frac{(y-(\mu-\frac{\sigma^2}{2})(T-t))^2}{2\sigma^2(T-t)}\right] dy.$

By noting that $\mathbb{E}(S_{(t)}) = S_{(0)}e^{\mu t}$, it is not hard to see that $\mu = r$ in a risk-neutral world. Simplify the equation, we get

$$C(t) = S_{(t)}\phi(\sigma\sqrt{T-t} - a) - e^{-r(T-t)}K\phi(-a),$$

where $a = \dfrac{\ln(K/S_{(t)}) - (r - \frac{\sigma^2}{2})(T-t)}{\sigma\sqrt{T-t}}$, which is also the solution ($S_{(t)} = S$) from solving BS-PDE above with the fundamental solution of the diffusion equation.