PSTAT 213ABC Lecture Notes

Professor: Dr. Tomoyuki Ichiba and Dr. Raya Feldman

Fall 2020 - Spring 2021

Bryan Xu

Introduction

These are the lecture notes for PSTAT 213ABC - Probability Theory & Stochastic Processes, from 2020-2021 school year taught by Tomoyuki Ichiba and Raya Feldman. This course covers generating functions, discrete and continuous time Markov chains; random walks; branching processes; birth-death processes; Poisson processes, point processes.

Contents

1 Preliminaries

1.1 Random Variables

We start by reviewing what a random variable is.

Definition 1 (Random Variable)**.** On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we define a *random variable* as a function $X(\omega) : \Omega \to \mathbb{R} \cup \{\pm\infty\}$ from the sample space Ω to the extended real line $\mathbb{R} \cup \{\pm\infty\}$.

We associate it with probability $\mathbb{P}$ by

$$\mathbb{P}[X \in A] = \mathbb{P}[X(\omega) \in A] = \mathbb{P}[\{\omega : X(\omega) \in A\}),$$

for every Borel measurable set A. It induces a measure called *probability distribution* defined as

$$\mu(A) = \mathbb{P}[X(\omega) \in A]; \quad A \in \mathcal{B} = \sigma(\mathbb{R}),$$

where $\mathcal{B} = \sigma(\mathbb{R})$ is the Borel sigma-algebra.

Definition 2 (Finite Random Variable)**.** We say a random variable X is *finite* if $\mathbb{P}[X = \pm\infty] = 0$.

Now recall what the cumulative distribution function of a random variable is.

Definition 3 (Cumulative Distribution Function)**.** The *cumulative distribution function* F (or *c.d.f*) of X is defined by

$$\begin{aligned} F_X(x) &= \mathbb{P}[X \le x] = \mathbb{P}[X \in (-\infty, x]], \quad \forall x \in \mathbb{R} \\ &= \int_{-\infty}^{x} 1 \, d\mathbb{P}[\omega] = \int_{-\infty}^{x} \mathbb{P}[d\omega]. \end{aligned}$$

Mathematical expectation is in turn defined as

$$\mathbb{E}[X] := \int_{-\infty}^{\infty} x \, dF_X(x),$$

and given a measurable function $g:\mathbb{R}\to\mathbb{R}$, we have

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)\,dF(x) = \int_{-\infty}^{\infty} g(x)F(dx).$$

Variance is then defined to be

$$\mathrm{Var}[X] := \mathbb{E}[(X-\mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Definition 4. Let X and Y be two random variables. The random vector (X,Y) induces a probability distribution η by

$$\eta(A) := \mathbb{P}[(X,Y)\in A] = \mathbb{P}[\{\omega : (X(\omega),Y(\omega))\in A];\quad A\in\mathcal{B}\times\mathcal{B}.$$

Sometimes we write $(X,Y)^{-1}(A) = \{\omega : (X,Y)\in A\}$ for such A. Now we consider the joint probability cumulative distribution function, which is a two-dimensional analogue of the c.d.f:

$$F_2(x_1,x_2) := \mathbb{P}[X_1\le x_1, X_2\le x_2];\quad (x_1,x_2)\in\mathbb{R}^2.$$

The joint c.d.f. has the following properties, coming from the probability measure:

- (Monotonicity): If $x_1\le y_1$ and $x_2\le y_2$, then

$$F_2(x_1,x_2)\le F_2(y_1,x_2),\quad F_2(x_1,x_2)\le F_2(x_1,y_2).$$

- (Right continuity): For every $(x_1,x_2)\in\mathbb{R}^2$ and $(a,b)\in\mathbb{R}^2$ with $x_1\ge a$, $x_2\ge b$, we have

$$\lim_{x_1\to a^+} F_2(x_1,x_2) = F_2(a,x_2),\quad \lim_{x_2\to b^+} F_2(x_1,x_2) = F_2(x_1,b).$$

- (Left limits exist): Both $\lim_{x_1\to a^-} F_2(a,x_2)$ and $\lim_{x_2\to b^-} F_2(x_1,b)$ exist.

- (Consistency): We have

$$\lim_{x_2\to\infty} F_2(x_1,x_2) = F_X(x_1) = \mathbb{P}[X\le x_1],\quad \lim_{x_1\to\infty} F_2(x_1,x_2) = F_Y(x_2) = \mathbb{P}[Y\le x_2].$$

In general, for n–dimensional random vectors $(X_1,\dots,X_n)$, we consider the joint c.d.f. $F_n:\mathbb{R}^n\to[0,1]$. We are allowed to do this because of Kolmogorov:

Proposition 1 (Kolmogorov's Consistency Theorem)**.** Suppose that for every $n \geq 2$, a c.d.f. $F_n : \mathbb{R}^n \to [0,1]$ of n variables satisfy the consistency condition. That is, there exists a c.d.f. $F_m : \mathbb{R}^m \to [0,1]$ such that

$$\lim_{x_{m+1},\ldots,x_n \to \infty} F_n(x_1,\ldots,x_m,x_{m+1},\ldots,x_n) = F_m(x_1,\ldots,x_m),$$

for every $1 \leq m < n$. Then there exists a probability space and a sequence of random variables $X_1,\ldots,X_n$ on it such that $(X_1,\ldots,X_n)$ has the joint c.d.f. F_n.

1.2 Classic Examples

Here are typical examples of random sequences and random functions.

I.I.D. Sequences

We say the sequence $(X_1,\ldots,X_n)$ of random variables is i.i.d. (independently and identically distributed), if the c.d.f. F is common for all random variables. That is,

$$\mathbb{P}[X_i \leq x] = F(x),$$

for every $i = 1,2,\ldots,n$, and $x \in \mathbb{R}$, We must also have independence, or

$$\mathbb{P}[X_1 \leq x_1,\ldots,X_n \leq x_n] = \prod_{i=1}^{n} \mathbb{P}[X_i \leq x_i] = \prod_{i=1}^{n} F(x_i).$$

Random Walks

Consider the i.i.d. sequence $\{\xi_i, i \geq 1\}$ of random variables. We define the random walk as

$$X_n := \xi_1 + \cdots + \xi_n; \quad n \geq 1,$$

with $X_0 = 0$. This gives us the recursive relationship $X_{n+1} = X_n + \xi_{n+1}$, for $n \geq 0$. We call this an integrated process $I(1)$ in the study of time series analysis.

Discrete-Time Markov Chains

Let $\mathcal{S}$ be a finite set (our state space). We call the sequence $\{X_n, n \geq 0\}$ of $\mathcal{S}$–valued random variables a time-homogeneous Markov chain if there exists a transition probability matrix

$$M := (M_{i,k})_{i,k\in\mathcal{S}},$$

with $\sum_{k\in\mathcal{S}} M_{i,k} = 1$ for every $i \in \mathcal{S}$, such that

$$\mathbb{P}[X_{n+1} = k \mid X_n = i] = M_{i,k}; \quad i,k \in \mathcal{S}, \quad n = 0,1,2,\ldots$$

Martingales

The sequence $\{X_n, n \geq 0\}$ associated with filtration $\{\mathcal{F}_n, n \geq 0\}$ is called a martingale if

(1) X_n is $\mathcal{F}_n$–measurable,

(2) $\mathbb{E}[|X_n|] < \infty$,

(3) $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = X_n; \quad n \geq 0.$

Here, the filtration $\{\mathcal{F}_n, n \geq 0\}$ is an increasing sequence of sigma fields, i.e., $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, for $n \geq 0$.

Random Functions

Given a random walk $\{X_n, n \geq 1\}$, consider a random function

$$X_n(t, \omega) := \sum_{i=1}^{\lfloor nt \rfloor} X_i(\omega); \quad t \geq 0.$$

As a function of $t \geq 0$, $X_n(t) = X_n(t, \omega)$ is right continuous with left limits. We may also consider a continuous function by interpolating linearly the sample path of random walk. Then the resulting function is a random continuous function.

1.3 Perron-Frobenius (Preview)

Now, as $n \to \infty$, what will be the limiting object of our random element X_n? And how do we quantify the convergence of random elements? To do this, we examine discrete-time Markov chains.

Consider a Markov chain with transition probability $M = (M_{i,j})_{1 \leq i,j \leq n}$ on $\mathcal{S} = \{1, \ldots, n\}$ with $M_{i,j} \geq 0$ for every $i, j \in \mathcal{S}$ and

$$\sum_{j \in \mathcal{S}} M_{i,j} = 1,$$

for every $i \in \mathcal{S}$.

Proposition 2 (Perron-Frobenius)**.** Assume that there exists $k \in \mathbb{Z}^+$ such that M^k has all its positive entries. Then there exists a row vector $\pi := \begin{pmatrix} \pi_1 & \pi_2 & \dots & \pi_n \end{pmatrix}$ with positive entries summing to one such that

$$M^{\ell}_{i,j} = \mathbb{P}[X_{\ell+1} = j \mid X_1 = i], \quad i, j \in \mathcal{S}, \quad \ell \geq 1$$

converges to π_j as ℓ goes to infinity. That is,

$$\lim_{\ell \to \infty} M^{\ell}_{i,j} = \pi_j,$$

for every $i, j \in \mathcal{S}$. Moreover, π is the unique row vector such that $\sum_{j \in \mathcal{S}} \pi_j = 1$ and $\pi M = \pi$.

This theorem gives us an answer to the following question: How much time do we need to wait until the Markov chain is close to the stationary distribution? If we follow the proof of the theorem, we find that there exists $A > 0$ and $\epsilon \in (0,1)$ such that

$$\sup_{i,j \in \mathcal{S}} |M^{\ell}_{i,j} - \pi_j| \leq A(1-\epsilon)^{\ell}; \quad \ell > 1.$$

This indicates how our Markov chain becomes close to the stationary distribution in the long run. The problem is that A may be very big, and ϵ may be close to zero, which means that the speed of convergence will be slow. As it turns out, the decay of the distance to π is exponentially fast.

1.4 Expectation Rigorously

Remember that we can write expectation in terms of the corresponding probability measure $\mathbb{P}$:

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \int X \, d\mathbb{P}(\omega) = \mathbb{E}[X^+] - \mathbb{E}[X^-],$$

for a random variable $X : \Omega \to \mathbb{R} \cup \{\pm\infty\}$. Here $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$, for $x \in \mathbb{R}$. This way we interpret the probability of a measurable set A as a special case of expectation:

$$\mathbb{P}[A] = \mathbb{E}[\mathbb{1}_A(\omega)] = \int_{\Omega} \mathbb{1}_A(\omega) \, d\mathbb{P}(\omega) = 1 \cdot \mathbb{P}[A] + 0 \cdot \mathbb{P}[A^c].$$

Then there are different forms of convergence for limits of random variables.

Definition 5 (Forms of Convergence)**.** Assume that $X_n(\omega) \overset{n\to\infty}{\longrightarrow} X(\omega)$ for all $\omega \in \Omega$ (we can replace "all" by "almost sure").

(a) (*Monotone Convergence*): If $X_n(\omega) \geq 0$ and $X_n \leq X_{n+1}(\omega)$ for all n and ω, then

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

(b) (*Dominated Convergence*): If there exists a random variable Y such that $|X_n(\omega)| \leq Y(\omega)$ for all n and ω with $\mathbb{E}[|Y|] < \infty$, then

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

(c) (*Bounded Convergence*): In particular, if there exists $c > 0$ such that $|X_n(\omega)| \leq c$ for every n and ω, then

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

Proposition 3 (Fatou's Lemma)**.** If X_n, $n \in \mathbb{N}$ is a sequence of random variables such that $X_n(\omega) \geq Y(\omega)$ for all n and ω with $\mathbb{E}[|Y|] < \infty$, then

$$\mathbb{E}\left[\liminf_{n\to\infty} X_n\right] \leq \liminf_{n\to\infty} \mathbb{E}[X_n].$$

Since the converging sum can be written as an expectation of a random variable, taking derivatives with respect to a parameter means taking limits. Thus implicitly we have already used these limits of expectations.

Proposition 4 (Jensen's Inequality)**.** If $u(\cdot)$ is a convex function on $\mathbb{R}$ and X is a random variable, then

$$\mathbb{E}[u(X)] \geq u(\mathbb{E}[X]).$$

2 Generating Functions

2.1 Introduction

Definition 6 (Generating Function)**.** Let $a = \{a_n \mid n \in \mathbb{N}_0\}$ be an infinite sequence of numbers (real or complex) for which

$$G_a(s) := \sum_{n=0}^{\infty} a_n s^n \text{ converges for some } s \in \mathbb{R}.$$

We call $G_a(\cdot)$ the *generating function* of a.

We denote by $\mathscr{S}$ the family of such sequences. More precisely,

$$\mathscr{S} = \{a \mid G_a(s) \text{ is well defined for some } s \in \mathbb{R}\}.$$

We know that $G_a(s)$ is a power series. Hence it has the following properties:

- There exists a radius of convergence $R \geq 0$ such that $G_a(s)$ converges if $|s| < R$ and diverges if $|s| > R$.
- $G_a(s)$ is differentiable or integrable term-by-term with respect to s for $|s| < R$.
- (*Uniqueness*): If $G_a(s) = G_b(s)$ for some a, b for every $|s| < \tilde{R} \leq R$, then $a \cong b$. Moreover,

$$a_n = \frac{1}{n!} \frac{\partial^n}{\partial s^n} G_a(s)\bigg|_{s=0} = \frac{G^{(n)}(0)}{n!}; \quad n \in \mathbb{N}.$$

- (*Abel's Theorem*): If $a \in \mathscr{S}$ is nonnegative, and $G_a(s) < \infty$ for $|s| < 1$, then

$$\lim_{s \to 1} G_a(s) = \sum_{n=0}^{\infty} a_n.$$

A canonical example of generating functions is the probability generating function.

Definition 7 (Probability Generating Function)**.** Suppose that the support of a random variable X is a subset of $\mathbb{N}_0$ (discrete). Let $a_n = \mathbb{P}[X = n]$, and $n \in \mathbb{N}$. We define the *probability generating function* (or *p.g.f*) by

$$G_X(s) := \mathbb{E}[s^X] = \sum_{n=0}^{\infty} s^n \mathbb{P}[X = n]; \quad s \in [0, 1].$$

Because of the uniqueness of the generating function, the information of a p.g.f. $a \in \mathscr{S}$ with $a_n = \mathbb{P}[X = n]$ of $\mathbb{N}_0$−valued random variable is encoded into the generating function $G_X(\cdot)$.

Definition 8 (Convolution)**.** The *convolution* $c := a * b$ of two sequences $a, b \in \mathscr{S}$ is defined by

$$c_n := \sum_{k=0}^{\infty} a_k b_{n-k}, \quad n \in \mathbb{N}_0.$$

We can associate two $\mathbb{N}_0$−valued independent random variables X and Y by

$$a_n = \mathbb{P}[X = n], \quad b_n = \mathbb{P}[Y = n].$$

Example 1. Let $G_X(\cdot)$ be a p.g.f. of a $\mathbb{N}_0$−valued random variable X. Verify

(1) $G^{(k)}(1) = \mathbb{E}[X(X-1)(X-2)\cdots(X-k+1)]$. In particular, $G'(1) = \mathbb{E}[X]$.

(2) $\mathrm{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = G''(1) + G'(1) - ((G'(1))^2$.

Solution. (1) Let $s < 1$. We can calculate the kth derivative of G to get

$$G^{(k)}(s) = \sum_i s^{i-k} i(i-1)\cdots(i-k+1)f(i) = \mathbb{E}[s^{X-k} X(X-1)\cdots(X-k+1)].$$

Now we take $s \uparrow 1$, and by Abel's Theorem, we get

$$G^{(k)}(s) \to \sum_i i(i-1)\cdots(i-k+1)f(i) = \mathbb{E}[X(X-1)\cdots(X-k+1)].$$

(2) We have

$$\begin{aligned} \mathrm{Var}[X] &= \mathbb{E}[(X - E[X])^2] \\ &= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + (\mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + (\mathbb{E}[X])^2 \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\ &= G''(1) + G'(1) - (G'(1))^2, \end{aligned}$$

because $\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X]$.

□

Example 2. Say we have a box that contains N balls (b black balls and $N-b$ white balls). We select n balls without replacement. The probability of getting k black balls and $n-k$ white balls is

$$f(k) = \binom{N}{n}^{-1}\binom{b}{k}\binom{N-b}{n-k}, \quad 0 \le k \le \min(n,b),\ 0 \le n-k \le \min(n, N-b),$$

and $f(k) = 0$, otherwise. This is the *hypergeometric distribution*. What is the expectation and variance of the hypergeometric distribution from the p.g.f.?

Solution. First our □

Definition 9 (Joint Probability Generating Function)**.** The *joint probability generating function* of two $\mathbb{N}_0$–valued random variables X_1 and X_2 by

$$G_{X_1,X_2}(s_1,s_2) := \mathbb{E}[s_1^{X_1} \cdot s_2^{X_2}]; \quad s_1, s_2 \in [0,1].$$

Definition 10 (Moment Generating Function)**.** Given a random variable X, we call $M_X(t) := \mathbb{E}[e^{tX}] = G_X(e^t)$ the *moment generating function* (or *m.g.f.*) of X, if the expectation is finite for some $t \in \mathbb{R}$.

We can compute the m.g.f. from the p.g.f., and vice versa.

$$\begin{aligned}
M_X(t) = \sum_{k=0}^{\infty} e^{tk}\mathbb{P}[X=k] &= \sum_{k=0}^{\infty}\sum_{n=0}^{\infty}\frac{(tk)^n}{n!}a_k \\
&= \sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\frac{(tk)^n}{n!}a_k \\
&= \sum_{n=0}^{\infty}\frac{t^n}{n!}\left(\sum_{k=0}^{\infty}k^n a_k\right) \\
&= \sum_{n=0}^{\infty}\frac{t^n}{n!}\mathbb{E}[X^n] \\
&= \mathbb{E}\left[\sum_{n=0}^{\infty}\frac{t^n}{n!}X^n\right] \\
&= \mathbb{E}[e^{tX}] \\
&= G_X(e^t),
\end{aligned}$$

if it is finite, and hence, if $\mathbb{E}[X^k] < \infty$ for some $K \geq 1$, then because of the property of the power series, we view $M_X(t)$ as a generating function of $\mathbb{E}[X^n]/n!$, and we have

$$\frac{\partial^k}{\partial t^k} M_X(t)\bigg|_{t=0} = \mathbb{E}[X^k]; \quad k = 1, 2, \ldots, K$$

The m.g.f. is not limited to the class of $\mathbb{N}_0$–valued random variables. However, there does arise a possible issue of m.g.f.s about the finiteness of $\mathbb{E}[|X|^n]$ for some n. We can deal with these *heavy-tailed* cases by considering the Laplace transform for positive random variables.

As an example, consider a random variable X with $\mathbb{P}[X = \infty] > 0$. Then we must be careful about the summation:

$$\lim_{s \uparrow 1} G_X(s) = \sum_{k=0}^{\infty} \mathbb{P}[X = k] = 1 - \mathbb{P}[X = \infty].$$

2.2 Applications of Generating Functions

Coin Flipping Game

Say that two players A and B play a game of flipping a coin many times with probability of heads being p and probability of tails being $q = 1 - p$. Player A wins if there are m heads appear before n tails appear, and Player B wins otherwise. We want to compute

$$p_{m,n} = \mathbb{P}[A \text{ wins}.$$

2.3 Simple Random Walks

We can try to understand random walks now under the lens of generating functions to understand long-term behavior. For a simple random walk $S_n = X_1 + \cdots + X_n$, let us define two events:

$$A_n := \{\text{particle is at the origin after } n \text{ steps,}\}$$
$$B_n := \{\text{the first return to the origin occurs exactly after } n \text{ steps,}\}$$

for $n \geq 1$. We want to find

$$p_0(n) = \mathbb{P}[S_n = 0] = \mathbb{P}[A_n], \quad n = 1, 2, \ldots, \text{ and } p_0(0) = 1,$$

$$f_0(n) = \mathbb{P}[S_1 \neq 0, \ldots, S_n = 0] = \mathbb{P}[B_n], \quad n = 1, 2, \ldots$$

Recall that the *first return time* is defined as

$$T_0 := \min\{n \geq 1 \mid S_n = 0\}.$$

If we view these as sequences of numbers, then the generating function of $p_0(\cdot)$ and $f_0(\cdot)$ are defined by

$$P_0(s) = \sum_{n=0}^{\infty} p_0(n)s^n, \quad F_0(s) = \sum_{n=1}^{\infty} f_0(n)s^n = \mathbb{E}[s^{T_0}], \quad s \in (0,1).$$

We allow T_0 to be *defective*, meaning that $\mathbb{P}[T_0 = \infty] = 1 - F_0(1) > 0$. In other words, we do not have to have $\mathbb{P}[T_0 = \infty] = 0$ necessarily. We want to determine if this probability is strictly positive or not.

Proposition 5 (Return to the Origin)**.** For the simple random walk we have

$$P_0(s) = \frac{1}{4pqs^2}, \quad F_0(s) = 1 - \frac{1}{P_0(s)}; \quad s \in (0,1),$$

and

$$\mathbb{P}[T_0 < \infty] = \sum_{n=1}^{\infty} f_0(n) = F_0(1) = 1 - \frac{1}{P_0(1)} = 1 - |p-q|.$$

In particular, we have $\mathbb{P}[T_0 < \infty] = F_0(1) = 1$, if and only if the random walk is *persistent* ($p = q = 1/2$), and moreover in this case the simple random walk is *null-recurrent*, i.e.,

$$\mathbb{E}[T_0] = F_0'(1) = \sum_{n=1}^{\infty} n f_0(n) = \infty.$$

Proof. First observe that $A = A_n$ satisfies $\mathbb{P}[A \mid B_k] = p_0(n-k)$ and hence

$$p_0(n) = \mathbb{P}[A] = \sum_{k=1}^{n} \mathbb{P}[A \cap B_k] = \sum_{k=1}^{n} \mathbb{P}[A \mid B_k]\mathbb{P}[B_k] = \sum_{k=1}^{n} p_0(n-k) f_0(k),$$

for $n = 1, 2, \ldots$ Multiplying by s^n, and summing over $n = 0, 1, 2, \ldots$, we have

$$\begin{aligned} P_0(s) &= \sum_{n=0}^{\infty} p_0(n)s^n = 1 + \sum_{n=1}^{\infty} p_0(n)s^n \\ &= 1 + \sum_{n=1}^{\infty} \sum_{k=1}^{n} p_0(n-k) f_0(k) s^n \\ &= 1 + \sum_{n=1}^{\infty} \sum_{j=0}^{\infty} p_0(j) s^j f_0(k) s^k \\ &= 1 + P_0(s)F_0(s); \quad s \in (0,1). \end{aligned}$$

Thus we obtain $F_0(s) = 1 - (1/P_0(s))$ for $s \in (0,1)$. Now we shall compute $P_0(\cdot)$. Note that $p_0(n) = 0$ if n is odd and if n is even then the event A_n means that there are an equal numbers of steps to the left and right among n steps, or

$$p_0(n) = \binom{n}{n/2} p^{n/2} q^{n/2}; \quad n \geq 2 \text{ (and } n \text{ even).}$$

Now we can use Newton's generalized binomial expansion to find that

$$\begin{aligned} P_0(s) &= \sum_{n=0}^{\infty} p_0(n) s^n = \sum_{k=0}^{\infty} \binom{2k}{k} (pq)^k s^{2k} \\ &= \sum_{k=0}^{\infty} \binom{(1/2)+k-1}{k} (4pqs^2)^k \\ &= \frac{1}{\sqrt{1-4pqs^2}}; \quad s \in (0,1). \end{aligned}$$

Here we use the fact that

$$\binom{x}{n} := x(x-1)\cdots(x-n+1)/n!, \quad x \in \mathbb{R}.$$

Note that $pq = p(1-p) \leq 1/4$ for every $p \in (0,1)$. Also by using $q = 1-p$ and

$$|p-q|^2 = |2p-1|^2 = 1 + 4p^2 - 4p = 1 - 4pq,$$

we obtain

$$F_0(1) = 1 - \frac{1}{P_0(1)} = 1 - \sqrt{1-4pq} = 1 - |p-q|.$$

Thus $F_0(1) = 1$ is equivalent to $p = q = 1/2$. The expectation of T_0 is computed by direct differentiation of the generating function. □

Note that if $p > 1/2$, then the particle tends to stray a long way to the right. We can now generalize this to any level r as opposed to just 0.

Proposition 6. We have that

$$F_r(s) := \sum_{n=1}^{\infty} f_r(n) s^n = \left(\frac{1-\sqrt{1-4pqs^2}}{2qs} \right)^r; \quad s \in (0,1),$$

where $f_r(s) = \mathbb{P}[S_1 \neq r, S_2 \neq r, \ldots, S_n = r]$ is the probability that the first visit of level r occurs exactly at the nth step.

Proof. Observe that for $1 < r \in \mathbb{N}$, to visit the level r, the particle must go to the level 1 first in k steps, then go up by $r-1$ levels in $n-k$ steps for some $k = 1, 2, \ldots, n-1$. This implies

$$f_r(n) = \mathbb{P}[S_1 \neq r, S_2 \neq r, \ldots, S_{n-1} \neq r, S_n = r]$$

$$= \sum_{k=1}^{n-1} \mathbb{P}[S_1 \neq r, \ldots, S_{n-1} \neq r, S_n = r \mid T_1 = k]\mathbb{P}[T_1 = k] = \sum_{k=1}^{n-1} f_{r-1}(n-k) f_1(k),$$

where T_k is the first passage time of level $k = 1, 2, \ldots$, i.e.

$$T_k = \min\{n \mid S_n = k\}.$$

Note that $f_r(n) = 0$ for $n < r$. Multiplying by s^n and summing over n, we obtain

$$\begin{aligned} F_r(s) = \sum_{n=1}^{\infty} f_r(n)s^n &= \sum_{n=1}^{\infty} \sum_{k=1}^{n-1} f_{r-1}(n-k) f_1(k) s^n \\ &= \sum_{k=1}^{\infty} \left(\sum_{\ell=1}^{\infty} f_{r-1}(\ell) s^\ell \right) f_1(k) s^k \\ &= F_{r-1}(s) F_1(s), \end{aligned}$$

for $r > 1$ and $s \in (0,1)$. Iterating this relationship, we obtain

$$F_r(s) = [F_1(s)]^r; \quad r \in \mathbb{N},\ s \in (0,1).$$

Now observe that $f_1(1) = \mathbb{P}[S_1 = 1] = p$ and that for $n > 1$,

$$\begin{aligned} f_1(n) = \mathbb{P}[T_1 = n] &= \mathbb{P}[T_1 = n \mid X_1 = 1]\mathbb{P}[X_1 = 1] + \mathbb{P}[T_1 = n \mid X_1 = -1]\mathbb{P}[X_1 = -1] \\ &= \mathbb{P}[T_2 = n-1]q \\ &= q f_2(n-1), \end{aligned}$$

as $\mathbb{P}[T_1 = n \mid X_1 = 1] = \mathbb{P}[T_1 = n \mid T_1 = 1] = 0$ for $n \geq 1$. Using this relationship, we derive a quadratic equation for $F_1(s)$:

$$\begin{aligned} F_1(s) = \sum_{n=1}^{\infty} s^n f_1(n) &= ps + \sum_{n=2}^{\infty} s^n f_1(n) \\ &= ps + qs \sum_{n=2}^{\infty} s^{n-1} f_2(n-1) \\ &= ps + qs \sum_{\ell=1}^{\infty} s^\ell f_2(\ell) \\ &= ps + qs F_2(s) \\ &+ ps + qs(F_1(s))^r. \end{aligned}$$

We solve for $F_1(s)$ to get

$$F_1(s) = \frac{1 \pm \sqrt{1 - 4pqs^2}}{2qs}.$$

Among the two solutions, the + solution cannot be a PGF, because as $s \downarrow 0$, it diverges to ∞ but $\lim_{s\to 0} F_1(s) = 0$, if $F_1(\cdot)$ is a PGF. Therefore combining with the above relationship, we have

$$F_1(s) = \frac{1-\sqrt{1-4pqs^2}}{2qs}, \quad F_r(s) = \left(\frac{1-\sqrt{1-4pqs^2}}{2qs}\right)^r; \quad s \in (0,1).$$

□

In particular, by taking $r = 1$ and letting $s \to 1$ in the formula for $F_1(s)$, we obtain a probability of attaining the level 1 sometime as

$$\sum_{n=1}^{\infty} f_1(n) = F_1(1) = \frac{1-\sqrt{1-4pq}}{2q} = \frac{1-|p-q|}{2q} = \min(1, p/q).$$

2.4 Nonsimple Random Walks

We now want to consider a random walk $S_n = X_1 + \cdots + X_n$, $n \in \mathbb{N}$ where we assume the values of X_i is in the integers.

Definition 11 (Right/Left Continuity)**.** We say a random walk $S_n = X_1 + \cdots X_n$ is *right-continuous* if $\mathbb{P}[X_i \leq 1] = 1$ and is *left-continuous* if $\mathbb{P}[X_i \geq -1] = 1$ for every i.

The simple random walk is both left and right continuous. After we generalize it, does this still hold? The right-continuous random walk cannot jump more than one level, and so for this random walk, the particle cannot skip the milestones in the positive integer. In other words, if $S_n = b$ for some n and $1 \leq b \in \mathbb{N}$, there exists increasing random numbers $k_1, \ldots, k_b (= n)$, such that $S_{k_r} = r$ for every $r = 1, 2, \ldots, b$. For a right-continuous random walk we still have

$$F_b(z) = (F_1(z))^b; \quad b \geq 1,\ z \in \mathbb{C},$$

where $F_b(z) = \sum_{n=1}^{\infty} z^n f_b(n)$ as defined before.

Proposition 7 (Hitting Time Theorem)**.** For a right continuous random walk, the hitting time $T_b = \min\{n \mid S_n = b\}$ satisfies

$$\mathbb{P}[T_b = n] = \frac{b}{n} \cdot \mathbb{P}[S_n = b]; \quad n \geq 1,\ b \geq 1.$$

To prove this, we need to make use of the following result from complex analysis:

Proposition 8 (Lagrange Inversion Formula)**.** Assume $z = w/f(w)$ is an analytic function of w in a neighborhood of the origin. If g belongs to C^∞, then

$$g(w(z)) = g(0) + \sum_{n=1}^{\infty} \frac{z^n}{n!}\left[\frac{d^{n-1}}{du^{n-1}}[g'(u)(f(u)]^n\right]\Big|_{u=0}.$$

Proof. □

Proposition 9 (Spitzer's Identity)**.** For a right continuous random walk $\{S_n \mid n \in \mathbb{N}_0\}$ the maximum $M_n := \max\{S_i \mid 0 \le i \le n\}$ and $S_n \vee 0 = \max(S_n, 0) = (S_n)^+$ satisfy the identity:

$$\log\left(\sum_{n=0}^{\infty} t^n \mathbb{E}[u^{M_n}]\right) = \sum_{n=1}^{\infty} \frac{t^n}{n} \mathbb{E}\left[u^{S_n \vee 0}\right]; \quad |t|, |u| < 1.$$

2.5 Branching Processes

When we study growth of a population of cells or increase of neurons in a reactor or the spread of an epidemic, we may model the size of the population as a *branching process*. We assume each individual in each generation gives a random number of births. For a precise notation, let us write $\{X_{n,i} \mid i = 1, 2, \ldots\}$ be the number of births of an individual i at the nth generation for each $n = 0, 1, 2, \ldots$ We shall consider the total number Z_n of individuals at the nth generation for each $n = 0, 1, 2, \ldots$, with $Z_0 = 1$, $Z_1 = X_{0,1}$, and

$$Z_{n+1} := X_{n,1} + \cdots + X_{n,Z_n} = \sum_{i=1}^{Z_n} X_{n,i}; \quad n \ge 0.$$

Here we set $Z_0 = 1$ for simplicity. We assume the family sizes form a collection of independent random variables, and all family sizes have the common p.m.f. $f(\cdot)$ with mean μ and corresponding generating function

$$G(x) = \sum_{x=0}^{\infty} x^j f(j) = \mathbb{E}[x^{X_{n,i}}] = \mathbb{E}[x^{Z_1}] = G_1(x); \quad 0 \le x \le 1.$$

We define the p.g.f. $G_n(x) := \mathbb{E}[x^{Z_n}]$ of the nth generation size Z_n for $n \ge 0$.

Proposition 10. We have that

$$G_{m+n}(x) = G_m(G_n(x)), \quad G_n(x) = G_1(G_1(\cdots(G_1(x)\cdots)); \quad 0 \le x \le 1,\ m, n \in \mathbb{N}.$$

In particular,

$$G_{n+1}(x) = G_n(G(x)).$$

Proof. If we prove the $G_{n+1}(x) = G_n(G(x))$ case, then we can repeat the iteration process to get the other expressions. Now for $0 \le x \le 1$ and $n \ge 0$,

$$\begin{aligned}
G_{n+1}(x) &= \mathbb{E}[x^{Z_{n+1}}] \\
&= \mathbb{E}\left[\mathbb{E}\left[x^{\sum_{i=1}^{Z_n} X_{n,j}} \mid Z_n\right]\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{Z_n} \mathbb{E}[x^{X_{n,i}} \mid Z_n]\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{Z_n} x^{X_{n,i}}\right] \\
&= \mathbb{E}\left[(G(x))^{Z_n}\right] \\
&= G_n(G(x)).
\end{aligned}$$

□

Proposition 11 (Probability of Extinction)**.** The limiting probability $\eta := \lim_{n\to\infty} \mathbb{P}[Z_n = 0]$ of extinction exists and is the smallest non-negative root of the equation $s = G(s)$ for $0 \le s \le 1$. $\eta = 1$. If $\mu = 1$ and $\mathrm{Var}[Z_1] > 0$, then $\eta = 1$. If $\mu = 1$ and $\mathrm{Var}[Z_n] = 0$, then $\eta = 0$.

Proof. Since $\{Z_n = 0\} \subseteq \{Z_{n+1} = 0\}$ for every $n \ge 1$, we have

$$\mathbb{P}[\text{extinction}] = \mathbb{P}\left[\bigcup_{n=0}^{\infty} \{Z_n = 0\}\right] = \lim_{n\to\infty} \mathbb{P}[Z_n = 0] = \eta \in [0,1]$$

exists. Since the p.g.f. $G(\cdot)$ is continuous, taking the limits of both sides of

$$\mathbb{P}[Z_n = 0] = G_n(0) = G(G_{n-1}(0)) = G(\mathbb{P}[Z_{n-1} = 0]),$$

as $n \to \infty$, we obtain $\eta = G(\eta)$. Thus the limiting probability η is a non-negative root of the equation $s = G(s)$, for $0 \le s \le 1$.

Suppose now that there exists a real number $\psi \in [0,1]$ such that $\psi = G(\psi)$. Then since $G(\cdot)$ is a monotone increasing function,

$$\mathbb{P}[Z_1 = 0] = G(0) \le G(\psi) = \psi.$$

By induction, we claim that

$$\mathbb{P}[Z_n = 0] = G(\mathbb{P}[Z_{n-1} = 0]) \le G(\psi) = \psi.$$

Thus taking the limits again, we claim

$$\eta = \lim_{n\to\infty} \mathbb{P}[Z_n = 0] \le \psi.$$

This means that η is the smallest non-negative root of the equation. Note that

$$G''(x) = \mathbb{E}\Big[Z_1(Z_1-1)x^{Z_1-2}\cdot \mathbb{1}_{\{Z_1\geq 2\}}\Big] \geq 0; \quad 0 \leq x \leq 1,$$

implies that $G(\cdot)$ is a convex function with $G(1) = 1$. By this convexity and looking at the number of intersections between the curves $y = G(x)$ and $y = G(x)$ for $0 \leq x \leq 1$, we may conclude the proof. □

When $\mu = \mathbb{E}[Z_1] > 1$, the total population can explode quickly. How fast does it grow? To answer this question, we consider some normalization. If $\mathbb{E}[Z_1] > 1$ and $\eta < 1$, define

$$W_n := \frac{Z_n}{\mathbb{E}[Z_n]} = \frac{Z_n}{\mu^n}; \quad n \geq 0.$$

It follows from the expectation and variance of Z_n, we have

$$\mathbb{E}[W_n] = 1, \quad \mathrm{Var}[W_n] = \frac{\sigma^2(1-\mu^{-n})}{\mu^2-\mu} \overset{n\to\infty}{\longrightarrow} \frac{\sigma^2}{\mu^2-\mu}.$$

2.6 Characteristic Functions

Definition 12 (Characteristic Function)**.** We define the *characteristic function* of some random variable X as

$$\varphi_X(t) := \mathbb{E}[e^{itX}]; \quad t \in \mathbb{R}.$$

We can extend this definition of these functions for random vectors; e.g.,

$$\varphi_{X,Y}(s,t) = \mathbb{E}[e^{isX+itY}]; \quad (s,t) \in \mathbb{R}^2.$$

We understand the expectation of complex-valued random variables as

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \int_\Omega \cos(tX(\omega))\,d\mathbb{P}(\omega) + i\int_\Omega \sin(tX(\omega))\,d\mathbb{P}(\omega); \quad t \in \mathbb{R}.$$

Both the MGF and the characteristic function generate moments by differentiation under integration: for example,

$$\varphi_X^{(k)}(0) = \frac{\partial^k}{\partial t^k}\varphi_X(t)\Big|_{t=0} = \mathbb{E}\left[\frac{\partial^k}{\partial t^k}e^{itX}\right]\Big|_{t=0} = i^k\mathbb{E}[X^k]; \quad k \in \mathbb{N}.$$

However, we need to justify this interchange between differentiation and integration, and the characteristic function behaves better than the MGF because of the following theorem.

Proposition 12 (Bochner's Theorem)**.** The characteristic function satisfies

(a) $\varphi(0) = 1$, $|\varphi(t)| \le 1$ for every $t \in \mathbb{R}$,

(b) $t \mapsto \varphi(t)$ is uniformly continuous, and

(c) $\varphi(\cdot)$ is nonnegative definite, i.e., for every n and $(t_1, \ldots, t_n) \in \mathbb{R}^n$ and $(z_1, \ldots, z_n) \in \mathbb{C}^n$,

$$\sum_{j,k=1}^{n} \varphi(t_j - t_k) z_j \bar{z}_k \ge 0.$$

Since the MGF is equivalent to the Laplace transform, it is quite useful for nonnegative random variables, as we have versatile tools from the inverse Laplace transform at our disposal.

Proposition 13. For any fixed $a > 0$, the following three statements are equivalent:

(a) $|M(t)| < \infty$ for $|t| < a$,

(b) $\varphi(\cdot)$ is analytic on the strip $|\mathrm{Im}(z)| < a$,

(c) The moments $m_k = \mathbb{E}[X^k]$ exist for $k = 1, 2, \ldots$ and satisfy

$$\limsup_{k\to\infty} \left(\frac{|m_k|}{k} \right)^{1/k} \le \frac{1}{a}.$$

If one of them holds for $a > 0$, then the power series expansion for $M(\cdot)$ may be extended analytically to the strip $|\mathrm{Im}(t)| < a$ and hence, $\varphi(t) = M(it)$.

Just like for MGFs, for independent random variables X and Y,

$$\varphi_{X+Y}(t) = \mathbb{E}[e^{it(X+Y)}] = \mathbb{E}[e^{itX}] \cdot \mathbb{E}[e^{itY}] = \varphi_X(t)\varphi_Y(t); \quad t \in \mathbb{R},$$
$$\varphi_{X,Y}(s,t) = \mathbb{E}[e^{isX+itY}] = \varphi_X(s)\varphi_Y(t); \quad (s,t) \in \mathbb{R}^2.$$

Example 3 (Characteristic Function of Normal Distribution)**.** When we compute the characteristic function of the standard normal $N(0,1)$, we may not substitute $s = it$ into the MGF

$$M_X(s) = \mathbb{E}[e^{sX}] = \int_{\mathbb{R}} e^{sx} \cdot \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx = \int_{\mathbb{R}} e^{s^2/2} \cdot \frac{e^{-(x-s)^2/2}}{\sqrt{2\pi}}\, dx = e^{-s^2/2}; \quad s \in \mathbb{R}$$

without some justification. The justification here is that $M_X(\cdot)$ is bounded, and so $M_X(\cdot)$ may be extended analytically to the characteristic function by the theorem. If $Y = aX + b$ with $a, b \in \mathbb{R}$, then

$$\varphi_Y(t) = \mathbb{E}[e^{it(aX+b)}] = e^{itb}\mathbb{E}[e^{i(ta)X}] = e^{itb}\varphi_X(at); \quad t \in \mathbb{R}.$$

Applying this relation to the standard normal random variable X with mean 0 and variance 1, we obtain the characteristic function of $Y = \sigma X + \mu$ with $\sigma > 0$, $u \in \mathbb{R}$, i.e.,

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itx} \frac{e^{-x^2/2}}{\sqrt{2\pi}} \, dx = e^{-t^2/2}, \quad \varphi_Y(t) = e^{it\mu-(\sigma^2 t^2/2)}; \quad t \in \mathbb{R}.$$

Example 4 (Exp, Chi-Square, and Gamma Distributions)**.** The exponential, chi-square, and gamma distributions are in the same family. The characteristic function of the gamma distribution $\Gamma(\lambda, s)$ is defined by the PDF

$$f_X(x) = \frac{\lambda^2 x^{s-1} e^{-\lambda x}}{\Gamma(s)}; \quad x > 0$$

is given by

$$\varphi_X(t) = \int_0^\infty e^{itx} \frac{\lambda^2 x^{s-1} e^{-\lambda x}}{\Gamma(s)} \, dx = \left(\frac{\lambda}{\lambda - it} \right)^s; \quad t \in \mathbb{R}.$$

When $s = 1$, it is the exponential distribution with parameter λ. When $s = d/2$ and $\lambda = 1/2$, it is the centered chi-square distribution with d degrees of freedom.

Example 5 (Cauchy Distribution)**.** We can directly calculate the characteristic function of the Cauchy distribution through contour integration and the residue theorem:

$$\varphi_X(t) = \int_{\mathbb{R}} e^{itx} \frac{1}{\pi(1+x^2)} \, dx = e^{-|t|}; \quad t \in \mathbb{R}.$$

We can check it from the Fourier transform of $e^{-|t|}$.

The name characteristic function comes from the unique characterization of distribution, i.e.,

$$\mathbb{P}[X \in A] = \mathbb{P}[Y \in A] \text{ for every set } A \in \mathcal{B}.$$

$$\Longleftrightarrow \mathbb{E}[e^{itX}] = \varphi_X(\cdot) \equiv \varphi_Y(\cdot) = \mathbb{E}[e^{itY}]; \quad t \in \mathbb{R},$$

where $\mathcal{B} = \sigma(\mathbb{R})$ is the smallest Borel sigma field generated by open sets in the real line. Thus in order to find out the characteristic function of a new random variable Y, find out the characteristic function of Y and compare it to that of X to connect to a known random variable X. Because of the relationship between the MGF and characteristic function, this method of finding the distribution of new random variable Y for the MGF and the PGF is used for non-negative random variables and for discrete random variables.

In general, we have to take account for the discontinuity of distribution functions. Here is the main inversion theorem.

Proposition 14 (Inversion Theorem)**.** The characteristic function $\varphi(\cdot)$ and the cumulative distribution function $F(\cdot)$ of random variable X satisfy

$$\mathbb{P}[a < X < b] + \frac{1}{2}(\mathbb{P}[X = a] + \mathbb{P}[X = b]) = \overline{F}(b) - \overline{F}(a)$$
$$= \lim_{N\to\infty} \int_{-N}^{N} \frac{e^{-iat} - e^{-ibt}}{2\pi i t} \varphi(t)\, dt; \quad a, b \in \mathbb{R},$$

where for $x \in \mathbb{R}$,

$$\overline{F}(x) := \frac{1}{2}(F(x) + \lim_{y\uparrow x} F(y)) = \frac{1}{2}(\mathbb{P}[X \le x] + \mathbb{P}[X < x]) = \mathbb{P}[X \le x] - \frac{1}{2}\mathbb{P}[X = x].$$

Proof. Let us recall the following integrals from calculus:

$$\int_0^\infty \frac{\sin x}{x}\, dx = \int_0^\infty \sin x \left[\int_0^\infty e^{-xu}\, du\right] dx = \int_0^\infty \left[\int_0^\infty e^{-xu} \sin x\, dx\right] du = \int_0^\infty \frac{du}{1+u^2} = \frac{\pi}{2},$$

and hence, for every $\alpha \in \mathbb{R}$ we have

$$\int_0^\infty \frac{\sin \alpha x}{x}\, dx = \frac{\pi}{2} \cdot \operatorname{sgn}(\alpha) = \frac{\pi}{2} \cdot (-\mathbb{1}_{\{\alpha<0\}} + \mathbb{1}_{\{\alpha>0\}}).$$

Then using this integration result, for every $a, b \in \mathbb{R}$ and $N \in \mathbb{N}$, we may evaluate the integral

$$I(N, a, b) = \int_{-N}^{N} \frac{e^{-iat} - e^{-ibt}}{2\pi i t} \varphi(t)\, dt = \int_{-\infty}^{\infty} \left[\int_{-N}^{N} \frac{e^{it(x-a)} - e^{it(x-b)}}{2\pi i t}\right] dt\, dF(x)$$
$$= \int_{\mathbb{R}} \left(\frac{1}{\pi}\int_0^N \frac{\sin(t(x-a))}{t}\, dt - \frac{1}{\pi}\int_0^N \frac{\sin(t(x-b))}{t}\, dt\right) dF(x) = \int_{\mathbb{R}} (I_1 - I_2)\, dF(x),$$

where we use Fubini's Theorem to interchange the order of integration, since the integrand is bounded uniformly by $|a-b|$ for $t \in [-N, N]$ and $x \in \mathbb{R}$. As $N \to \infty$, the terms I_1, I_2 and I converge as in the table:

x	$\lim_{N\to\infty} I_1$	$\lim_{N\to\infty} I_2$	$\lim_{N\to\infty}(I_1 - I_2)$
$x < a$	$-1/2$	$-1/2$	0
$x = a$	0	$-1/2$	$1/2$
$a < x < b$	$1/2$	$-1/2$	1
$x = b$	$1/2$	0	$1/2$
$x > b$	$1/2$	$1/2$	0

Therefore we conclude that the inversion formula

$$\lim_{N\to\infty} I(N, a, b) = \mathbb{P}[a < X < b] + \frac{1}{2}(\mathbb{P}[X = a] + \mathbb{P}[X = b]); \quad a, b \in \mathbb{R}.$$

□

It follows from the right continuity of cumulative distribution function that two random variables X and Y have the same characteristic function if and only if they have the same distribution. Indeed, it is simple that if X and Y have the same cumulative distribution function, then they have the same characteristic function; on the other hand, if $\varphi_X(\cdot) \equiv \varphi_Y(\cdot)$, then the inversion theorem implies

$$\overline{F}_X(b) = \lim_{a\to-\infty}(\overline{F}_X(b) - \overline{F}_X(a)) = \lim_{a\to-\infty}(\overline{F}_Y(b) - \overline{F}_Y(a)) = \overline{F}_Y(b); \quad b \in \mathbb{R}.$$

Then $F_X(x) = F_Y(x)$ for every continuity point $x \in (\mathcal{J}_X \cup \mathcal{J}_Y)^c$, where

$$\mathcal{J}_X := \{x \in \mathbb{R} \mid F_X(x) - F_X(x-) > 0\}, \quad \mathcal{J}_Y := \{x \in \mathbb{R} \mid F_Y(x) - F_Y(x-) > 0\},$$

is an at most countable set of jump points of $F_X(\cdot)$ and $F_Y(\cdot)$. This implies X and Y have the same distribution.

Another application of the inversion theorem is about the PDF. Take $b = x$ and $a = x - h$ for $h > 0$ and observe that at the continuity points x and $x - h$ of $F(\cdot)$,

$$\frac{1}{h}(F(x)-F(x-h)) = \frac{F(x)-F(x-)}{2} - \frac{F(x-h)-F(x-h-)}{2} = \frac{1}{2\pi}\int_{\mathbb{R}} \frac{e^{ith}-1}{it} \cdot e^{-itx}\varphi(t)\,dt$$

for $x \in \mathbb{R}$. If $\int_{\mathbb{R}} |\varphi(t)|\,dt < \infty$, then by the dominated convergence theorem, the left-derivative of $F(\cdot)$ at x exists, i.e.,

$$\lim_{h\downarrow 0}\frac{1}{h}(F(x)-F(x-h)) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi(t)\,dt.$$

Similarly, we can work with right-derivatives.

Proposition 15. The probability density function $f(\cdot)$ and characteristic function $\varphi(\cdot)$ of a random variable satisfy

$$f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\varphi(t)\,dt$$

at every point x at which $f(\cdot)$ is differentiable.

Definition 13 (Convergence in Distribution)**.** Suppose that a sequence $F_n(\cdot)$, $n \geq 1$ of cumulative distribution functions converges to a CDF $F(\cdot)$ at every continuity point x of $F(\cdot)$. Then we say that F_n *converges in distribution* to F and we write $F_n \xrightarrow{d} F$ as $n \to \infty$. If $X_1, X_2, \ldots$ is a sequence of random variables with CDF $F_1, F_2, \ldots$, we say $X_n \xrightarrow{d} X$ if $F_n \xrightarrow{d} F$ as

$n \to \infty$.

Proposition 16 (Continuity Theorem)**.** Suppose $\{F_n(\cdot), n \geq 1\}$ is a sequence of distribution functions with the corresponding characteristic functions $\{\varphi_n(\cdot), n \geq 1\}$.

(a) If $F_n \xrightarrow{d} F$ as $n \to \infty$ for some distribution function F with characteristic function φ, then $\varphi(t) = \lim_{n\to\infty} \varphi_n(t)$ for every $t \in \mathbb{R}$.

(b) Conversely, if $\lim_{n\to\infty} \varphi_n(t) = \varphi(t)$ exists and is continuous at $t = 0$, then $\varphi(\cdot)$ is a characteristic function of some distribution function $F(\cdot)$ and $F_n \xrightarrow{d} F$.

2.7 Limit Theorems

Proposition 17. If X has a finite absolute moment of order k for some integer $k \geq 1$, then the characteristic function $\varphi(\cdot)$ of X has the following expansion in the neighborhood of $t = 0$:

$$\varphi(t) = \sum_{j=0}^{k} \frac{(it)^j}{j!} \mathbb{E}[X^j] + o(|t|^k) = \sum_{j=0}^{k-1} \frac{(it)^j}{j!} \mathbb{E}[X^j] + \frac{\theta_k}{k!} \mathbb{E}[|X|^k]|t|^k,$$

where θ_k is a constant with $|\theta_k| \leq 1$.

Proof. We only prove the $k = 1$ case. We apply the dominated convergence theorem to

$$(\varphi(t+h) - \varphi(t))/h = \int_{-\infty}^{\infty} (e^{i(t+h)x} - e^{itx})/h \, dF(x),$$

where the integrand is dominated by $|x|$, i.e.

$$\sup_{t\in\mathbb{R}} \left| \frac{e^{i(t+h)x} - e^{itx}}{h} \right| \leq |x|; \quad x \in \mathbb{R}.$$

Thus if $\mathbb{E}[|X|] < \infty$, then

$$\begin{aligned} \lim_{h\to 0} \frac{\varphi(t+h) - \varphi(t)}{h} &= \lim_{h\to 0} \int_{-\infty}^{\infty} \frac{e^{i(t+h)x} - e^{itx}}{h} \, dF(x) \\ &= \int_{-\infty}^{\infty} \lim_{h\to 0} \frac{e^{i(t+h)x} - e^{itx}}{h} \, dF(x) \\ &= \int_{-\infty}^{\infty} (ix) e^{itx} \, dF(x). \end{aligned}$$

□

Thus we have $\varphi(0) = 1$ and $\varphi'(t) = i\mathbb{E}[Xe^{itX}]$. Since $\varphi(\cdot)$ has a finite first derivative at the point $t = 0$, then we have the Taylor expansion

$$\varphi(t) = \varphi(0) + \varphi'(0)t + o(|t|) = 1 + \mathbb{E}[X]t + o(|t|).$$

Since $\varphi(\cdot)$ has a finite first derivative in the neighborhood of $t = 0$, then

$$\varphi(t) = \varphi(0) + \varphi'(\theta t)t = 1 + \mathbb{E}[iXe^{i\theta X}]t = 1 + \theta_1 \mathbb{E}[|X|]|t|; \quad |\theta| \le 1,$$

where

$$\theta_1 := \frac{\mathbb{E}[iXe^{i\theta X}]t}{(\mathbb{E}[|X|]|t|)}, \quad |\theta_1| \le 1.$$

Proposition 18 (Weak Law of Large Numbers)**.** Let $X_1, X_2, \ldots$ be independently, identically distributed random variables with finite means μ. Then

$$\frac{S_n}{n} \xrightarrow{d} \mu, \quad n \to \infty.$$

Proof. Computing the characteristic function $\varphi(t) = \mathbb{E}[e^{itX_1}]$ and expanding around $t = 0$,

$$\varphi(t/n) = \mathbb{E}[e^{itX_1/n}] = \varphi(0) + \frac{t}{n}\varphi'(0) + o\left(\frac{t}{n}\right) = 1 + \frac{it\mu}{n} + o\left(\frac{t}{n}\right).$$

Thus we conclude that

$$\varphi_n(t) = \mathbb{E}[e^{itS_n/n}] = \mathbb{E}\left[\prod_{i=1}^{n} e^{itX_1}\right] = (\mathbb{E}[e^{itX_1}])^n = \left(1 + \frac{i\mu t}{n} + o\left(\frac{t}{n}\right)\right)^n \underset{n\to\infty}{\longrightarrow} e^{i\mu t}; \quad t \in \mathbb{R}.$$

The right hand side is the characteristic function of a constant random variable μ. □

Proposition 19 (Central Limit Theorem)**.** Let $X_1, X_2, \ldots$ be a sequence of independently, identically distributed random variables with finite second moments $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2 > 0$ for every i. Let us define $Y := (X_i - \mu)/\sigma$ for every i. Then the scaled sample averages $\sqrt{n}\overline{Y}_n := (Y_1 + \cdots + Y_n)/\sqrt{n}$ converge to the standard normal distribution, i.e.,

$$\sqrt{n}\overline{Y}_n = \frac{(X_1 + \cdots + X_n) - n\mu}{\sqrt{n\sigma^2}} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} X, \quad n \to \infty,$$

where $X \sim \mathcal{N}(0,1)$. In terms of probability,

$$\lim_{n\to\infty} \mathbb{P}\left[\sqrt{n}\overline{Y}_n \le x\right] = \int_{-\infty}^{x} \frac{e^{-u^2/2}}{\sqrt{2\pi}} = \Phi(x); \quad x \in \mathbb{R}.$$

Proof. We employ the same strategy as above. We compute the characteristic functions of Y_1 and of $\sqrt{n}Y_n$:

$$\varphi_{Y_1}(t/\sqrt{n}) = \mathbb{E}[e^{i(t/\sqrt{n})Y_1}] = \varphi_{Y_1}(0) + \frac{t}{\sqrt{n}} \cdot \varphi'_{Y_1}(0) + \frac{1}{2} \cdot \frac{t^2}{n} \cdot \varphi''_{Y_1}(0) + o\left(\frac{t^2}{n}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right),$$

and hence we claim that

$$\mathbb{E}[e^{it\sqrt{n}Y_n}] = (\mathbb{E}[e^{i(t/\sqrt{n})Y_1})^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \overset{n\to\infty}{\longrightarrow} e^{-t^2/2}; \quad t \in \mathbb{R}.$$

□

For the IID random variables $\{X_n, n \geq 0\}$ and the cumulative sum $S_n = X_0 + \cdots + X_n$ with finite mean $\mathbb{E}[X_1] = 0$, let us define the expected number V_i of visits by the random walk S_n at site i and first passage time T_i at site i:

$$V_i := \mathbb{E}\left[\sum_{n=0}^{\infty} \mathbb{1}_{S_n = i}\right] = \sum_{n=0}^{\infty} \mathbb{P}[S_n = i], \quad T_i = \min\{n \mid S_n = i\}, \quad i \in \mathbb{Z}.$$

We presume that $T_i = \infty$ if there is no such n such that $S_n = i$. Note that $\mathbb{P}[S_n = i \mid T_i = t] = 0$ if $n < t$. Also because of the IID property and the definition of T_i, we have

$$\mathbb{P}[S_{m+t} = i \mid T_i = t] = \mathbb{P}[S_m = 0].$$

The law of large numbers says that for every $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}[|S_n| \leq n\epsilon] = 1,$$

and hence it follows from the definition of limits that there exist m such that $\mathbb{P}[|S_n| \leq n\epsilon] \geq \frac{1}{2}$ for every $n \geq m$. Thus for every $K > 0$ with $m \leq n$ and $n\epsilon \leq K$ we have

$$\mathbb{P}[|S_n| \leq K] \geq \mathbb{P}[|S_n| \leq n\epsilon] \geq \frac{1}{2}.$$

This gives us an important application to random walks using our limit theorems.

Proposition 20. The random walk $\{S_n, n \geq 0\}$ is persistent (or recurrent), i.e.,

$$\mathbb{P}[S_n = 0 \text{ for some } n \geq 1] = 1,$$

if the mean step size is zero.

3 Discrete-Time Markov Chains

3.1 Some Review

Let's recall some basic definitions and properties of Markov chains.

Definition 14 (Markov Chain)**.** Let us take a countable set $\mathcal{S}$ and consider a sequence $X := \{X_0, X_1, \dots\}$ of $\mathcal{S}$−valued random variables which satisfies the Markov property:

$$\mathbb{P}[X_n = s \mid X_0 = x_0, \dots, X_{n-1} = x_{n-1}] = \mathbb{P}[X_n = s \mid X_{n-1} = x_{n-1}]$$

for every $s, x_0, \dots, x_{n-1} \in \mathcal{S}$ and for every $n \geq 1$. Then we call X a *Markov chain*.

Let's check that this Markov property is equivalent to

$$\mathbb{P}[X_{n+1} = s \mid X_{n_1} = x_{n_1}, \dots, X_{n_k} = x_{n_k}] = \mathbb{P}[X_{n+1} = s \mid X_{n_k} = x_{n_k}]$$

for every $s, x_{n_1}, \dots, x_{n_k} \in \mathcal{S}$ and for every $0 \leq n_1 < n_2 < \cdots < n_k \leq n$. Indeed, taking $n_1 = 0 < n_2 = 1 < \cdots < n_k = n$ implies the Markov property with n being replaced by $n+1$. Conversely, we see that

$$\mathbb{P}[X_4 = x_4 \mid X_1 = x_1, X_3 = x_3] = \frac{\mathbb{P}[X_4 = x_4, X_3 = x_3, X_1 = x_1]}{\mathbb{P}[X_1 = x_1, X_3 = x_3]} = \mathbb{P}[X_4 = x_4 \mid X_3 = x_3].$$

Indeed we can verify that

$$\begin{aligned}
\mathbb{P}[X_4 = x_4, X_3 = x_3, X_1 = x_1] &= \sum_{(x_0, x_2) \in \mathcal{S}^2} \mathbb{P}[X_4 = x_4, X_3 = x_3, X_2 = x_2, X_1 = x_1, X_0 = x_0] \\
&= \sum_{(x_0, x_2) \in \mathcal{S}^2} \mathbb{P}[X_4 = x_4 \mid X_3, X_2, X_1, X_0] \cdot \mathbb{P}[X_3, X_2, X_1, X_0] \\
&= \sum_{(x_0, x_2) \in \mathcal{S}^2} \mathbb{P}[X_4 \mid X_3] \cdot \mathbb{P}[X_3, X_2, X_1, X_0] \\
&= \mathbb{P}[X_4 \mid X_3] \sum_{(x_0, x_2) \in \mathcal{S}^2} \mathbb{P}[X_3, X_2, X_1, X_0] \\
&= \mathbb{P}[X_4 \mid X_3] \cdot \mathbb{P}[X_1, X_3].
\end{aligned}$$

This is the specific case of $n = 3$, $n_1 = 1$, $n_2 = 3$, but we can similarly verify with additional indices that this holds in the general case.

Definition 15 (Time-Homogeneous Chain)**.** We say a Markov chain X is *time-homogeneous* if

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

for every $n, i, j \in \mathcal{S}$. We call the matrix $\mathbf{P} = (p_{i,j})_{(i,j)\in\mathcal{S}^2}$ the *transition probability matrix*. Note that $p_{i,j} \geq 0$ and $\sum_{j\in\mathcal{S}} p_{i,j} = 1$ for every $i \in \mathcal{S}$.

Definition 16 (n–step Transition Probability)**.** The *n-step transition probability* of a Markov chain is defined by

$$p_{i,j}(m, m+n) = \mathbb{P}[X_{m+n} = j \mid X_m = i]$$

for every $m, n \geq 0$, $i, j \in \mathcal{S}$. We write the matrix $\mathbf{P}(m, m+n) = p_{i,j}(m, m+n)$, $i, j \in \mathcal{S}$. If it is homogeneous, we have the probability of going from state i to state j in n steps

$$p_{i,j}(m, m+n) = \mathbb{P}[X_n = j \mid X_0 = i]$$

for $i, j \in \mathcal{S}$ and $m, n \geq 0$.

By the Markov property and conditional probability, we can obtain an important fact.

Proposition 21 (Chapman-Kolmogorov Equations)**.**

$$p_{i,j}(m, m+n+r) = \sum_k p_{i,k}(m, m+n) p_{k,j}(m+n, m+n+r).$$

Therefore, $\mathbf{P}(m, m+n+r) = \mathbf{P}(m, m+n)\mathbf{P}(m+n, m+n+r)$ and $\mathbf{P}(m, m+n) = \mathbf{P}^n$.

Proof. We have

$$\begin{aligned}
p_{i,j}(m, m+n-r) &= \sum_{k\in\mathcal{S}} \mathbb{P}[X_{n+m+r} = j, X_{n+m} = k \mid X_m = i] \\
&= \sum_{k\in\mathcal{S}} \mathbb{P}[X_{n+m+r} = j \mid X_{n+m} = k, X_m = i] \cdot \mathbb{P}[X_{n+m} = k \mid X_m = i] \\
&= \sum_{k\in\mathcal{S}} \mathbb{P}[X_{n+m+r} = j \mid X_{n+m} = k] \cdot \mathbb{P}[X_{n+m} = k \mid X_n = i] \\
&= \sum_{k\in\mathcal{S}} p_{k,j}(n+m, n+m+r) \cdot p_{i,k}(m, m+n)
\end{aligned}$$

for every $n, m, r \geq 0$ and $i, j \in \mathcal{S}$. In matrix form, we write this as

$$\mathbf{P}(m, m+n+r) = \mathbf{P}(m, m+n)\mathbf{P}(m+n, m+n+r),$$

and in the time-homoegeneous case we have $\mathbf{P}(m, m+n) = \mathbf{P}^n$. □

As a corollary, the marginal probability $\mu^{(n)} := (\mu_i^{(n)} := \mathbb{P}[X_n = i], i \in \mathcal{S})$ satisfies

$$\mu^{(m+n)} = \mu^{(m)}\mathbf{P}^n; \quad \mu^{(n)} = \mu^{(0)}\mathbf{P}^n; \quad m, n \geq 0.$$

This is an important conclusion that the random evolution of the chain is determined by the transition matrix $\mathbf{P}$ and the initial mass function $\mu^{(0)}$.

Example 6 (Simple Random Walk)**.** The simple random walk in $\mathbb{Z}$ with probabilities (p, q) of going to the left or right is a Markov chain on $\mathcal{S} = \mathbb{Z}$ with transition probability

$$p_{i,j}(n) := p_{i,j}(m, m+n) = \binom{n}{\frac{1}{2}(n+j-1)} p^{(n+j-i)/2} q^{(n-j+i)/2}$$

if $n + j - i$ is even and zero otherwise, for $m, n \geq 0$ and $i, j \in \mathbb{Z}$.

Example 7 (Branching Process)**.** The branching process Z_n is a Markov chain on $\mathcal{S} = \mathbb{N}_0$ with $p_{i,j} = \mathbb{P}[Z_{n+1} = j \mid Z_n = i]$ being the coefficient of s^j in

$$(G(s))^i = \mathbb{E}[s^{X_{1,1}+\cdots+X_{1,i}}]$$

for every $i, j \in \mathbb{N}_0$ where $X_{n,k}$ are IID random variables distributed in the offspring distribution, and $G(\cdot)$ is the generating function of the offspring distribution.

When we describe a stochastic process in an open system where some external effects or forces come into the system, the process itself may not be a Markov process. When this happens, we can enlarge the state space by including the external effects into our consideration.

Here is an artificial example due to Markov himself. Let $Y_1, Y_3, Y_5, \ldots$ be a sequence of IID random variable such that

$$\mathbb{P}[Y_{2k+1} = -1] = \mathbb{P}[Y_{2k+1} = 1] = \frac{1}{2}, \quad k = 0, 1, 2, \ldots,$$

and we define $Y_{2k} = Y_{2k-1}Y_{2k+1}$ for $k = 1, 2, \ldots$ For example, if $Y_1 = 1$, $Y_3 = 1$, $Y_5 = -1$, $Y_7 = 1$, then $Y_2 = 1$, $Y_4 = -1$, and $Y_6 = -1$. Now $\mathbb{E}[Y_{2k}Y_{2k+1}] = \mathbb{E}[Y_{2k-1}Y_{2k+1}^2] = \mathbb{E}[Y_{2k-1}] = 0$, and so the sequence $Y_1, Y_2, \ldots$ is pairwise independent. Hence $p_{i,j}(n) = \frac{1}{2}$ for all n and i, $j = \pm 1$, and it follows that the Chapman-Kolmogorov are satisfied. However, $\{Y_k, k \in \mathbb{N}\}$ is not Markov:

$$\mathbb{P}[Y_{2k+1} = 1 \mid Y_{2k} = -1] = \frac{1}{2} \neq \mathbb{P}[Y_{2k+1} = 1 \mid Y_{2k} = -1, Y_{2k-1} = 1] = 0.$$

Thus the Chapman-Kolmogorov equations are necessary for the Markov property, but not sufficient, for the same reason that pairwise independence is weaker than independence. Although Y_n is not a Markov chain, we can define a new process $Z_n := (Y_n, Y_{n+1})$ in $\mathcal{S} = \{\pm 1\}^2$. Z_n is a Markov chain with time-inhomogeneous transition probabilities, e.g.

$$\mathbb{P}[Z_{n+1} = (1,1) \mid Z_n = (1,1)] = \begin{cases} 1/2, & n \text{ is even} \\ 1, & n \text{ is odd} \end{cases}$$

$$\mathbb{P}[Z_{n+1} = (-1,-1) \mid Z_n = (1,-1)] = \begin{cases} 1/2, & n \text{ is even} \\ 1, & n \text{ is odd} \end{cases}$$

3.2 Classification of States

Recall how we analyzed the simple random walk, where we used generating functions to find out the probability distribution of the first passage time. With a similar spirit, for a homogeneous Markov chain in the countable state space $\mathcal{S}$, focusing on state $j \in \mathcal{S}$, let us define the sets

$$A_m := \{X_m = j\}, \quad B_m := \{X_r \neq j, 1 \leq r \leq m, X_m = j\}; \quad m \geq 1,$$

and the probability that the first visit to state j, starting from state i takes place at the nth step

$$f_{i,j}(n) := \mathbb{P}[B_n \mid X_0 = i] = \mathbb{P}[X_r \neq j, 1 \leq r \leq n-1, X_n = j \mid X_0 = i], \quad i,j \in \mathcal{S},\ n \geq 1,$$

and the probability of visiting state j, starting from state i,

$$f_{i,j} := \mathbb{P}[X_n = j \text{ for some } n \geq 1 \mid X_0 = i]; \quad i,j \in \mathcal{S}.$$

Definition 17 (Recurrent/Transient State)**.** We say that a state $i \in \mathcal{S}$ is called *recurrent* (or *persistent*), if $f_{i,i} = 1$. Otherwise, state i is called *transient*.

Observe by the Markov property and conditional probability that

$$\begin{aligned}p_{i,j}(m) = \mathbb{P}[A_m \mid X_0 = i] &= \sum_{r=1}^{m} \mathbb{P}[A_m \cap B_r \mid X_0 = i] \\ &= \sum_{r=1}^{m} \mathbb{P}[A_m \mid B_r, X_0 = i] \cdot \mathbb{P}[B_r \mid X_0 = i] \\ &= \sum_{r=1}^{m} \mathbb{P}[A_m \mid X_r = j] \cdot \mathbb{P}[B_r \mid X_0 = i] \\ &= \sum_{r=1}^{m} f_{i,j}(r) P_{j,j}(m-r),\end{aligned}$$

and hence we obtained a relationship between the generating functions $\mathscr{P}_{i,j}(s) := \sum_{n=1}^{\infty} s^n p_{i,j}(n)$ and $\mathscr{F}_{i,j}(s) := \sum_{n=0}^{\infty} s^n f_{i,j}(n)$ with $p_{i,j}(0) = \delta_{i,j} = \mathbb{1}_{i=j}$, and with $f_{i,j}(0) = 0$:

$$\mathscr{P}_{i,j}(s) = \delta_{i,j} + \sum_{m=1}^{\infty} s^m \sum_{r=1}^{m} f_{i,j}(r) P_{j,j}(m-r) = \delta_{i,j} + \mathscr{F}_{i,j}(s)\mathscr{P}_{j,j}(s); \quad i,j \in \mathcal{S}$$

for $|s| < 1$. This implies that for every persistent $j \in \mathcal{S}$, by Abel's Theorem,

$$\mathscr{F}_{j,j}(1) = f_{j,j} = 1 \Longleftrightarrow \mathscr{P}_{j,j}(s) = \frac{1}{1 - \mathscr{F}_{j,j}(s)} \underset{s \uparrow 1}{\longrightarrow} \infty = \sum_{n=0}^{\infty} p_{j,j}(n).$$

This implies that

$$\lim_{s \uparrow 1} \mathscr{P}_{i,j}(s) = \sum_{n=0}^{\infty} p_{i,j}(n) = \mathscr{F}_{i,j}(1)\mathscr{P}_{j,j}(1) < \infty; \quad i \in \mathcal{S}.$$

For the transient state $j \in \mathcal{S}$, i.e., $f_{j,j} < 1$ if and only if $\sum_{n=0}^{\infty} p_{j,j}(n) < \infty$ holds and

$$\lim_{s \uparrow 1} \mathscr{P}_{i,j}(s) = \sum_{n=0}^{\infty} p_{i,j}(n) = \mathscr{F}_{i,j}(1)\mathscr{P}_{j,j}(1) < \infty.$$

In particular, if j is transient, then $\lim_{n\to\infty} p_{i,j}(n) = 0$ for every $i \in \mathcal{S}$.

Definition 18 (Positive/Null-Recurrence)**.** Let $T_j := \min\{n \geq 1 \mid X_n = j\}$. For the mean recurrence time we have

$$\mu_i := \mathbb{E}[T_i \mid X_0 = i] = \begin{cases} \sum_{n=0}^{\infty} n f_{i,i}(n), & \text{if } i \text{ is recurrent} \\ \infty, & \text{if } i \text{ is transient} \end{cases}$$

State $i \in \mathcal{S}$ is *null recurrent* (or null persistent) if $\mu_i = \infty$ (by interpreting $1/\mu_i = 0$); state i is *positive recurrent* (or positive persistent) if $\mu_i < \infty$ (by interpreting $1/\mu_i > 0$).

Recall the renewal theorem: the limiting probability is the reciprocal of the mean recurrence time. In this sense we see the connection between the renewal theorem and the definition of null/positive recurrence.

Proposition 22. State $i \in \mathcal{S}$ is null recurrent if and only if $\lim_{n\to\infty} p_{i,i}(n) = 0$. If it is the case, $\lim_{n\to\infty} p_{j,i}(n) = 0$ for every $j \in \mathcal{S}$.

Definition 19 (Periodicity)**.** The *period* $d(i)$ of a state i is defined by $\gcd\{n \mid p_{i,i} > 0\}$ of the times at which return to the state i is possible. $p_{i,i}(n) = 0$ unless n is a multiple of $d(i)$. We call the state i *periodic* if $d(i) > 1$ and *aperiodic* if $d(i) = 1$.

A particular type of state gets a special name.

Definition 20 (Ergodic State)**.** We call a state $i \in \mathcal{S}$ *ergodic* if it is persistent, non-null recurrent, and aperiodic.

3.3 Classification of Chains

Definition 21 (Communication)**.** We say a state $i \in \mathcal{S}$ *communicates* with $j \in \mathcal{S}$, if there exists $m \geq 0$ such that $p_{i,j}(m) > 0$, and we write $i \to j$. If $i \to j$ and $j \to i$, we say i and j intercommunicate, and write $i \leftrightarrow j$.

Let's investigate some properties of communication.

- Recalling the definition of $f_{i,j} := \sum_{n=1}^{\infty} f_{i,j}(n)$. Thus $i \to j$ if and only if $f_{i,j} > 0$.
- If $i \leftrightarrow$ and $j \leftrightarrow k$, then there exist m and n such that $p_{i,j}(m) > 0$ and $p_{j,k}(n) > 0$. Thus by the Chapman-Kolmogorov equation,

$$\begin{aligned} p_{i,k}(m+n) &= \mathbb{P}[X_{m+n} = k \mid X_0 = i] \\ &\geq \mathbb{P}[X_{m+n} = k, X_m = j \mid X_0 = i] \\ &= \mathbb{P}[X_{m+n} = k \mid X_m = j, X_0 = i] \cdot \mathbb{P}[X_m = j \mid X_0 = i] \\ &= p_{i,j}(m) p_{j,k}(n) > 0, \end{aligned}$$

  that is, $i \to k$. Similarly, we can show that $k \to i$. Thus $i \leftrightarrow j$ and $j \leftrightarrow k$ imply $i \leftrightarrow k$ for every $i, j, k \in \mathcal{S}$.

- Suppose $i \leftrightarrow j$ for some $i, j \in \mathcal{S}$. Then there exist some integers k_1 and k_2 such that $p_{i,j}(k_1) > 0$ and $p_{j,i}(k_2) > 0$, and in particular, by the Chapman-Kolmogorov equation, $p_{i,i}(k_1 + k_2) > 0$ and $p_{j,j}(k_1 + k_2) > 0$. Then by definition of the periods $d(i) = \gcd(m \mid p_{i,i}(m) > 0)$ and $d(j) = \gcd(m \mid p_{j,j}(m) > 0)$ of state i and state j respectively, we claim that $k_1 + k_2 = 0$ modulo $d(i)$ and $k_1 + k_2 \equiv 0 \mod d(j)$. For every $m \in \mathbb{N}$ with $p_{i,i}(m) > 0$,

  $$p_{j,j}(m + k_1 + k_2) \geq p_{i,j}(k_2) p_{i,i}(m) p_{i,j}(k_1) > 0,$$

  and hence $\{m \mid p_{i,i}(m)\} \subseteq \{m \mid p_{j,j}(m + k_1 + k_2) > 0\}$. Combining this observation with the definition of $d(j)$ and $k_1 + k_2 \equiv 0 \mod d(j)$, we see

  $$d(i) = \gcd(m \mid p_{i,i}(m) > 0) \geq \gcd(m \mid p_{j,j}(m + k_1 + k_2)) > 0\} = d(j).$$

  By interchanging the roles of i and j, we can get $d(i) \leq d(j)$, and so if $i \leftrightarrow j$, then $d(i) = d(j)$.

- Similarly, if $i \leftrightarrow j$, then there exist k_1 and k_2 such that $p_{i,j}(k_1) > 0$ and $p_{j,i}(k_2) > 0$. This implies that for every $r \geq 0$

  $$p_{i,i}(k_1 + k_2 + r) \geq p_{i,j}(k_1) \cdot p_{j,j}(r) \cdot p_{j,i}(k_2) = \alpha p_{j,j}(r).$$

  Summing over r, we have

  $$\sum_{r=1}^{\infty} p_{i,i}(r) < \infty \implies \sum_{r=1}^{\infty} p_{j,j}(r) \leq \sum_{r=1}^{m+n+r-1} p_{j,j}(r) + \alpha \sum_{r=m+n+r}^{\infty} p_{i,i}(r) < \infty.$$

  In other words, if i is transient, then j is transient. Interchanging the roles of i and j, we can get that if j is transient, then i is transient. Thus under the condition $i \leftrightarrow j$ of communication, i is transient if and only if j is transient.

Proposition 23. If $i \leftrightarrow j$ for some $i, j \in \mathcal{S}$, then their periods are the same; that is, $d(i) = d(j)$. If, in addition, i is transient, then so is j.

Later on, we shall see that if $i \leftrightarrow j$, then i is null persistent if and only if j is null persistent. Since $p_{i,i}(0) = 1$, the relation $i \leftrightarrow j$ is an equivalence relation, and the state space $\mathcal{S}$ can be partitioned into the equivalence classes of $\leftrightarrow$. Within each equivalence class, all states are of the same type.

Definition 22 (Irreducible and Closed Sets)**.** We say a subset $C \subseteq \mathcal{S}$ is *irreducible* if every pair $i, j \in C$ of states in C intercommunicate, i.e. $i \leftrightarrow j$. We say a subset $C \subseteq \mathcal{S}$ is closed if $p_{i,j} = 0$ for every $i \in C$ and $j \notin C$. We call a set C aperiodic, persistent, null respectively, if every state $i \in C$ is aperiodic, persistent, and null respectively.

Let C_j, $j = 1, 2, \ldots$ be persistent equivalence classes of relation $\leftrightarrow$ in the state space $\mathcal{S}$. If there exists a non-closed set C_r for some r, then there exist $i \in C_r$ and $j \notin C_r$ such that $p_{i,j} > 0$ but then since $j \notin C_r$ means $j \not\to i$, we have a contradiction

$$\mathbb{P}[X_n \neq i \text{ for all } n \geq 1 \mid X_0 = 1] \geq \mathbb{P}[X_1 = j \mid X_0 = i] = p_{i,j} > 0$$

to that i is in the persistent set. Thus we have the following:

Proposition 24 (Decomposition of State Space)**.** The countable state space $\mathcal{S}$ is decomposed into the disjoint union of transient set T and irreducible, closed sets C_i, $i \in \mathbb{N}$ of persistent states, that is,

$$\mathcal{S} = T \cup \left(\bigcup_i C_i \right).$$

Particularly, if the state space $\mathcal{S}$ is finite and if all the states $j \in \mathcal{S}$ were transient, then for every $i \in \mathcal{S}$, $\lim_{n\to\infty} p_{i,j}(n) = 0$, and hence, it would be a contradiction:

$$0 = \lim_{n\to\infty} \sum_{j\in\mathcal{S}} p_{i,j}(n) = 1.$$

This means at least one state is persistent. Moreover, if a persistent state $i \in \mathcal{S}$ of such finite state space $\mathcal{S}$ were null persistent, then $\lim_{n\to\infty} p_{i,i}(n)$ and $\lim_{n\to\infty} p_{j,i}(n) = 0$ for every $j \in \mathcal{S}$, and thus it would be another contradiction: for the closed, persistent equivalence class to which state i belongs, and for every $k \in C_i \subseteq \mathcal{S}$, $i \leftrightarrow k$ and k is null-persistent, and so

$$0 = \lim_{n\to\infty} \sum_{k\in C_i} p_{i,k}(n) = 1.$$

Thus as a corollary, we have the following proposition:

Proposition 25. If the state space $\mathcal{S}$ is finite, then at least one state is persistent and all the persistent state is non-null.

Proposition 26. For an irreducible and aperiodic Markov chain $\{X_n, n \geq 0\}$,

$$\lim_{n\to\infty} p_{i,j}(n) = \lim_{n\to\infty} \mathbb{P}[X_n = j \mid X_0 = i] = \frac{1}{\mu_j}, \quad i, j \in \mathcal{S},$$

where μ_j is the mean recurrence time.

Proof. Let us assume that the Markov chain $\{X_n, n \geq 0\}$ is irreducible and aperiodic with transition probability matrix $\mathbf{P}$. We employ a coupling technique: take an independent Markov chain $\{Y_n, n \geq 0\}$ with the same transition probability matrix $\mathbf{P}$. Then the coupled Markov chain $Z := (Z_n := (X_n, Y_n), n \geq 0)$ with transition probability

$$\begin{aligned} p_{(i,j),(k,\ell)} &= \mathbb{P}[Z_{n+1} = (k,\ell) \mid Z_n = (i,j)] = \mathbb{P}[X_{n+1} = k, Y_{n+1} = \ell \mid X_n = i, Y_n = j] \\ &= \mathbb{P}[X_{n+1} = k \mid X_n = i] \cdot \mathbb{P}[Y_{n+1} = \ell \mid Y_n = j] = p_{i,k} p_{j,\ell}, \quad i,j,k,\ell \in \mathcal{S}. \end{aligned}$$

Since X is irreducible and aperiodic, there exists N_0 such that $p_{i,k}(n) p_{j,\ell}(n) > 0$ for every $n \geq N_0$. This implies Z is irreducible.

- Case: Transient. We have already proved the transient case.
- Case: Positive recurrent. Let us assume that the Markov chain X_n is irreducible, aperiodic, and positive recurrent with transition matrix $\mathbf{P}$. Since X is positive recurrent and irreducible, the stationary distribution $\pi := (\pi_i, i \in \mathcal{S})$ exists and so $\nu_{i,j} := \pi_i \pi_j$, $i,j \in \mathcal{S}$ is a stationary distribution of Z and Z is also positive recurrent. Recall $\pi_k = 1/\mu_k$, $k \in \mathcal{S}$.

  Fix $i,j,s \in \mathcal{S}$. Assume $X_0 = i$, $Y_0 = j$, and define $T = \min\{n \geq 1 \mid Z_n = (s,s)\}$. Since Z is positive recurrent, $\mathbb{P}[T < \infty \mid Z_0 = (i,j)] = 1$. Condtionally on $\{n \geq T\}$, X_m and Y_m for $m \geq n$ have the same distribution, independent of the initial points (i,j):

$$\begin{aligned} p_{i,k}(n) &= \mathbb{P}[X_n = k \mid X_0 = i] \\ &= \mathbb{P}[X_n = k, T \leq n \mid X_0 = i] + \mathbb{P}[X_n = k, T > n \mid X_0 = i] \\ &= \mathbb{P}[Y_n = k, T \leq n \mid Y_0 = j] + \mathbb{P}[X_n = k, T > n \mid X_0 = i] \\ &\leq \mathbb{P}[Y_n = k \mid Y_0 = j] + \mathbb{P}[T > n \mid X_0 = i] \\ &= p_{j,k}(n) + \mathbb{P}[T > n \mid Z_0 = (i,j)]. \end{aligned}$$

  Interchanging the roles of (i,j), we obtain that for every $i,j,k \in \mathcal{S}$,

$$|p_{i,k}(n) - p_{j,k}(n)| \leq \mathbb{P}[T > n \mid Z_0 = (i,j)] \overset{n\to\infty}{\longrightarrow} 0.$$

  This implies that for every finite subset $F \subset \mathcal{S}$,

$$\begin{aligned} \sum_{i\in\mathcal{S}} \pi_i |p_{i,k}(n) - p_{j,k}(n)| &= \left(\sum_{i\in F} + \sum_{i\in F^c} \right) |p_{i,k}(n) - p_{j,k}(n)| \\ &\leq \sum_{i\in F} |p_{i,k}(n) - p_{j,k}(n)| + 2 \sum_{i\in F^c} \pi_k \\ &\Longrightarrow \overset{n\to\infty}{\longrightarrow} 2 \sum_{i\in F^c} \pi_i, \end{aligned}$$

and then letting $F \uparrow \mathcal{S}$, the limit in the right must be 0. Thus for every $j,k \in \mathcal{S}$, we obtain

$$|\pi_k - p_{j,k}(n)| = \left|\sum_{i\in\mathcal{S}} \pi_i (p_{i,k}(n) - p_{j,k}(n))\right| \le \sum_{i\in\mathcal{S}} \pi_i |p_{i,k}(n) - p_{j,k}(n)| \overset{n\to\infty}{\longrightarrow} 0,$$

and we conclude the case for a positive recurrent Markov chain.

- Case: Null recurrent. Assume X is null recurrent.
  - If the coupled chain Z is transient, then it follows from the transition of Z,

    $$\lim_{n\to\infty} (p_{i,j}(n))^2 = \lim_{n\to\infty} \mathbb{P}[Z_n = (j,j) \mid Z_0 = (i,i)] = 0,$$

    and hence $\lim_{n\to\infty} p_{i,j}(n) = 0$ for every $i,j \in \mathcal{S}$.
  - If the coupled chain Z is positive recurrent, then it would yield a contradiction:

    $$\infty = \mathbb{E}[\min\{n \mid X_n = i\} \mid X_0 = i] \le \mathbb{E}[\min\{n \mid Z_n = (i,i)\} \mid Z_0 = (i,i)] < \infty,$$

    and hence Z cannot be positive recurrent.
  - If Z is null recurrent, if there exists (i,j) such that

    $$\lim_{n\to\infty} p_{i,j}(n) \ne 0,$$

    then by the diagonal argument, there exists a subsequence n_r such that $\lim_{r\to\infty} p_{i,j}(n_r)$ exists in $(0,1]$ (say α_i) and does not depend on i for every i,j. This implies that for every finite set $F \subset \mathcal{S}$,

    $$\sum_{k\in F} \alpha_k p_{k,j} = \lim_{r\to\infty} \sum_{k\in F} p_{i,k}(n_r) p_{k,j} \le \lim_{r\to\infty} p_{i,j}(n_r+1) = \lim_{r\to\infty} \sum_{k\in\mathcal{S}} p_{i,k} p_{k,j}(n_r) = \sum_{k\in\mathcal{S}} p_{i,k}\alpha_j = \alpha_j.$$

    Letting $F \uparrow \mathcal{S}$, then we claim that $\sum_{k\in\mathcal{S}} \alpha_k p_{k,j} \le \alpha_j$, $j \in \mathcal{S}$. If the strict inequality holds for some state in $\mathcal{S}$, then it would be a contradiction

    $$\sum_{k\in\mathcal{S}} \alpha_k = \sum_{k\in\mathcal{S}} \sum_{j\in\mathcal{S}} \alpha_k p_{k,j} < \sum_{j\in\mathcal{S}} \alpha_j.$$

    Thus $\sum_{k\in\mathcal{S}} \alpha_k p_{k,j} = \alpha_j$, for all $j \in \mathcal{S}$. But then this means the stationary probability exists, and it is a contradiction. Therefore

    $$\lim_{n\to\infty} p_{i,j}(n) = 0; \quad i,j \in \mathcal{S}.$$

□

Some corollaries:

Proposition 27.
- If the irreducible aperiodic chain is transient or null recurrent, then $\lim_{n\to\infty} p_{i,j}(n) = 0$ for every $i, j \in \mathcal{S}$.
- A persistent state is null if and only if $\lim_{n\to\infty} p_{i,i}(n) = 0$ for all $i \in \mathcal{S}$.

Proof.
- Immediate beacuse $\mu_j = \infty$.
- Let $C(i)$ be the irreducible closed set of states which contains the persistent state i. If $C(i)$ is aperiodic, then we apply the above theorem. If $C(i)$ is periodic with period $d(i)$, then $Y = \{Y_n = X_{nd}, n \geq 0\}$ is an aperiodic chain, and apply the theorem:

$$\lim_{n\to\infty} p_{j,j}(nd) = \lim_{n\to\infty} \mathbb{P}[Y_n = j \mid Y_0 = j] = \frac{d}{\mu_j} = 0; \quad j \in \mathcal{S}.$$

□

3.4 Reversibility and Long-Term Behavior

Now we want to examine the reversal of time-scale of irreducible, positive recurrent Markov chains X_n with transition probability matrix $\mathbf{P}$ and stationary distribution π under *equilibrium*. Assume $\mathbb{P}^\pi[X_n = j] = \pi_j$ for every $0 \leq n \leq N$ and $j \in \mathcal{S}$. We define the time-reversal $Y_n := X_{N-n}$ of $\{X_n, 0 \leq n \leq N\}$. Then

$$\begin{aligned}
\mathbb{P}^\pi[Y_{n+1} = i_{n+1} \mid Y_n = i_n, \ldots, Y_0 = i_0] &= \frac{\mathbb{P}^\pi[X_{N-n-1} = i_{n-1}, \ldots, X_N = i_0]}{\mathbb{P}^\pi[X_{N-n} = i_n, \ldots, X_N = i_0]} \\
&= \frac{\pi_{i_{n+1}} \cdot p_{i_{n+1},i_n} \cdots p_{i_1,i_0}}{\pi_{i_n} \cdot p_{i_n,i_{n-1}} \cdots p_{i_1,i_0}} \\
&= \frac{\pi_{i_{n+1}} \cdot p_{i_{n+1},i_n}}{\pi_{i_n}} \\
&= \mathbb{P}^\pi[Y_{n+1} = i_{n+1} \mid Y_n = i_n].
\end{aligned}$$

Thus the time-reversal $\{Y_n, 0 \leq n \leq N\}$ is a Markov chain with the transition probability

$$\mathbb{P}^\pi[Y_{n+1} = j \mid Y_n = i] = \frac{\pi_j}{\pi_i} \cdot \mathbf{P}_{j,i}; \quad i, j \in \mathcal{S}.$$

Definition 23 (Detailed Balance)**.** We say a nonnegative vector $\boldsymbol{\lambda} := \{\lambda_i, i \in \mathcal{S}\}$ and the transition probability matrix $\mathbf{P}$ are in *detailed balance* if

$$\lambda_i p_{i,j} = \lambda_j p_{j,i}; \quad \forall i, j \in \mathcal{S}.$$

Definition 24 (Reversibility)**.** An irreducible Markov chain X is called *reversible* if the transition of the chain X and that of its time-reversal Y are the same, i.e. the detailed balance equations hold:

$$\pi_i p_{i,j} = \pi_j p_{j,i}; \quad i, j \in \mathcal{S}.$$

In such a case we say that the irreducible chain X is reversible in equilibrium.

If this condition holds, then

$$\sum_{i \in \mathcal{S}} \pi_i p_{i,j} = \sum_i \pi_i p_{j,i} = \pi_j, \quad \forall j \in \mathcal{S}.$$

hence $\boldsymbol{\pi P} = \boldsymbol{\pi}$.

Proposition 28 (Detailed Balance Condition)**.** If the system of detailed balance conditions holds, then π is a stationary distribution. Moreover, X is reversible in equilibrium.

Example 8 (Ehrenfest Model)**.** Suppose we have m gas molecules in chamber A and chamber B connected by a very thin corridor. Let us denote X_n as the number of molecules in chamber A at time n. At each instance, pick a random molecule and move it from one chamber to another. Thus the transition of $\{X_n, n \geq 0\}$ is given by $p_{m,m-1} = 1 = p_{0,1}$ and

$$\mathbb{P}[X_{n+1} = i+1 \mid X_n = i] = p_{i,i+1} = 1 - \frac{i}{m}, \quad p_{i,i-1} = \frac{i}{m}; \quad i = 1, \ldots, m-1.$$

The detailed balance condition is that for each i,

$$\pi_i \left(1 - \frac{i}{m}\right) = \pi_{i+1} \cdot \frac{i+1}{m},$$

and the solution to this equation is $\pi_i = \binom{m}{i} \cdot \frac{1}{2^m}$, for $i = 0, 1, \ldots, m$. Indeed for each i,

$$\frac{m!}{i!(m-i)!} \cdot \frac{m-i}{m} = \frac{m!}{(i+1)!(m-i-1)!} \cdot \frac{i+1}{m}.$$

Thus $\pi_i = \binom{m}{i} \cdot \frac{1}{2^m}$, for $i \in \mathcal{S}$ is the stationary distribution.

Example 9 (Simple Random Walk)**.** The simple random walk with transition probabilities $p_{i,i+1} = p \in (0,1)$ and $p_{i,i-1} = q - 1 - p$ for $i = 1, \ldots, M-1$, and with $p_{0,1} = p$ and $p_{M,M-1} = q$. The detailed balance equations are $\pi_i p = \pi_{i+1} q$ for $i = 0, \ldots, M-1$. T he stationary distribution is $\boldsymbol{\pi} = (c(p/q))^i$, where c is a normalizing factor.

Proposition 29 (Perron-Frobenius Theorem)**.** For the transition probability matrix $\mathbf{P}$ of a finite-state, irreducible Markov chain with period d, the eigenvalues $(\lambda_1,\dots,\lambda_N)$ of $\mathbf{P}$ are $\lambda_1 = 1$, $\lambda_j = e^{2\pi\sqrt{-1}j/d}$, $j = 0,\dots,d-1$ and $\lambda_{d+1},\dots,\lambda_N$ lie inside the unit circle.

If $\lambda_1,\dots,\lambda_N$ are all distinct, $\mathbf{P}$ are decomposed as $B^{-1}\Lambda B$, where $\Lambda = \mathrm{diag}(\lambda_1,\dots,\lambda_N)$. If $d = 1$, in addition, then

$$\lim_{n\to\infty}\mathbf{P}^n = \lim_{n\to\infty}\mathbf{B}^{-1}\begin{pmatrix}1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0\end{pmatrix}\mathbf{B}.$$

In general, $\mathbf{P}$ can be represented by the Jordan canonical form $\mathbf{P} = \mathbf{B}^{-1}\mathbf{M}\mathbf{B}$ with $\mathbf{M}$ is the $(N \times N)$ block diagonal matrix, where the diagonal component $\mathbf{J}_i,\dots,$ corresponds to the eigenvalue λ_i and

$$\mathbf{M} := \begin{pmatrix}\mathbf{J}_1 & & & \\ & \mathbf{J}_2 & & \\ & & \ddots & \\ & & & \end{pmatrix}, \quad \mathbf{J}_i = \begin{pmatrix}\lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i\end{pmatrix}$$

Example 10. Consider the state space $\mathcal{S} = \{1,2,3\}$ and probability transition matrix

$$\mathbf{P} = \begin{pmatrix}1 & 0 & 0\\ 1/4 & 1/2 & 1/4\\ 0 & 0 & 1\end{pmatrix}.$$

The characteristic polynomial is $\lambda^3-(5/2)\lambda+2\lambda-1/2 = 0$, and hence the eigenvalues are $1/2, 1$, and 1. $\mathbf{P}$ is decomposed as

$$\mathbf{P} = \begin{pmatrix}0 & -1 & 2\\ 1 & 0 & 1\\ 0 & 1 & 0\end{pmatrix}\begin{pmatrix}1/2 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}-1/2 & 1 & -1/2\\ 0 & 0 & 1\\ 1/2 & 0 & 1/2\end{pmatrix},$$

and hence

$$\begin{aligned}\mathbf{P}^n &= \begin{pmatrix}0 & -1 & 2\\ 1 & 0 & 1\\ 0 & 1 & 0\end{pmatrix}\begin{pmatrix}(1/2)^n & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}-1/2 & 1 & -1/2\\ 0 & 0 & 1\\ 1/2 & 0 & 1/2\end{pmatrix}\\ &= \begin{pmatrix}1 & 0 & 0\\ \frac{1}{2}-\frac{1}{2^{n+1}} & \frac{1}{2^n} & \frac{1}{2}-\frac{1}{2^{n+1}}\\ 0 & 0 & 1\end{pmatrix}\end{aligned}$$

3.5 Branching Processes Again

Recall the branching process $\{Z_n, n \geq 0\}$ with $Z_0 = 1$ and $\mathbb{P}[Z_1 = k] = f(k)$, $k \geq 0$, $G(s) = \mathbb{E}[s^{Z_1}]$ and $f(0) + f(1) \in (0,1)$, $f(0) > 0$. Recall also that the extinction probability $\eta := \mathbb{P}[T = \min\{n \mid Z_n = 0\} < \infty]$ is the smallest non-negative solution of the equation $s = G(s)$, $s \in [0,1]$. We shall consider the condition $E_n = \{n < T < \infty\}$, i.e. the population is still alive at time n, and the conditional probability that $Z_n = j$ given the future extinction

$$p_j^{(n)} = \mathbb{P}[Z_n = j \mid E_n]; \quad n \geq 1, \quad j \geq 1.$$

Note that under the condition $0 < f(0) + f(1) < 1$, $0 < \mathbb{P}[E_n] < 1$, and the probability η of extinction satisfies $0 < \eta < 1$. Recall $\mathbb{P}[Z_n = 0] = G_n(0)$, and observe $\mathbb{P}[E_n] = \mathbb{P}[T < \infty] - \mathbb{P}[T \leq n] = \mathbb{P}[T < \infty] - \mathbb{P}[Z_n = 0] = \eta - G_n(0)$. Then

$$\mathbb{P}[Z_n = j, E_n] = \mathbb{P}[Z_n = j] \cdot \mathbb{P}[\textit{each of } j \textit{ descendants extinct in the future}] = \mathbb{P}[Z_n = j] n^j; \quad j \geq 1.$$

Computing the conditional probability generating function,

$$\begin{aligned}
G_n^\pi(s) = \mathbb{E}[s^{Z_n} \mid E_n] &= \sum_{j=0}^{\infty} p_j^{(n)} \cdot s^j \\
&= \sum_{j=0}^{\infty} s^j \frac{\mathbb{P}[Z_n = j, E_n]}{\mathbb{P}[E_n]} \\
&= \sum_{j=0}^{\infty} s^j \frac{\mathbb{P}[Z_n = j] \cdot n^j}{\mathbb{P}[E_n]} \\
&= \frac{G_n(s\eta) - G_n(0)}{\eta - G_n(0)} \\
&= 1 - H_n(s\eta) l \quad s \in [0,1),
\end{aligned}$$

where

$$H_n(s) := \frac{\eta - G_n(s)}{\eta - G_s(0)}, \quad \frac{H_n(s)}{H_{n-1}(s)} = \frac{h(G_{n-1}(s))}{h(G_{n-1}(0))}, \quad h(s) := \frac{\eta - G(s)}{\eta - s}, \quad 0 \leq s < \eta.$$

Since $G(\cdot)$ is convex and non-decreasing on $[0, \eta)$, h is non-decreasing and $G_{n-1}(\cdot)$ is non-decreasing. Then

$$H_n(s) = \frac{H_{n-1}(s) h(G_{n-1}(s))}{h(G_{n-1}(0))} \geq H_{n-1}(s); \quad s < \eta.$$

Taking the limits, we obtain

$$H(s\eta) = \lim_{n \to \infty} H_n(s\eta), \quad G^\pi(s) = \lim_{n \to \infty} G_n^\pi(s) = 1 - H(s\eta)$$

exist for $s \in (0,1]$. Therefore the limiting probability $\lim_{n\to\infty} p_j^{(n)} = \pi_j$ exists as the coefficient of s^j in the generating function $G^\pi(s)$ for every $j \geq 0$. Moreover, by taking the limits as $n \to \infty$ of

$$H_n(G(s)) = \frac{\eta - G_n(G(s))}{\eta - G_n(0)} = \frac{\eta - G(G_n(0))}{\eta - G_n(0)} \cdot \frac{\eta - G_{n+1}(s)}{\eta - G_{n+1}(0)} = h(G_n(0)) \cdot H_{n+1}(s),$$

we obtain

$$H(G(s)) = \lim_{n\to\infty} h(G_n(0)) \cdot H(s) = \lim_{s\uparrow\eta} \frac{\eta - G(s)}{\eta - s} \cdot H(s) = G'(\eta) \cdot H(s); \quad 0 \leq s < \eta.$$

Combining, we conclude the functional relationship

$$G^\pi(\eta^{-1}G(s\eta)) = 1 - H(G(s\eta)) = 1 - G'(\eta) \cdot H(s\eta) = 1 - G'(\eta) \cdot (1 - G'(s)) = G'(\eta)G^\pi(s) + 1 - G'(\eta).$$

- If $\mu = \mathbb{E}[Z_1] \leq 1$, then $\eta = 1$ and $G'(\eta) = \mu$. This means

$$G^\pi(G(s)) = \mu G^\pi(s) + 1 - \mu; \quad 0 \leq s < \eta.$$

- If $\mu = G'(\eta) \neq 1$, then

$$\lim_{s\uparrow\eta} H(s) = \lim_{s\uparrow\eta} H(G(s)) = \lim_{s\uparrow\eta} H(s) \cdot G'(s).$$

  This implies that $\lim_{s\uparrow\eta} H(s) = 0$, and hence

$$\lim_{s\to 1} G^\pi(s) = 1 - \lim_{s\uparrow\eta} H(s) = \sum_j \pi_j = 1.$$

  The distribution of Z_n, conditional on future extinction, converges as $n \to \infty$ to $\{\pi_j, j \geq 0\}$. We call $\mu > 1$ the supercritical case and $\mu < 1$ the subcritical case.

- If $\mu = 1$, the critical case, then $G'(\eta) = 1$ with $\eta = 1$, and hence $G^\pi(G(s)) = G^\pi(s)$. Since $G(s) > s$ for every $s < 1$, $G^\pi(s) = G^\pi(0) = 0$ for every $s < 1$. Thus $\pi_j = 0$ for every j.

We summarize as

$$\lim_{n\to\infty} \mathbb{P}[Z_n = j] = 0, \quad \lim_{n\to\infty} \mathbb{P}[Z_n = j \mid E_n] = 0.$$

If in addition $G''(1) < \infty$, then $\sigma^2 = \mathrm{Var}[Z_1] < \infty$ and

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{Z_n}{n\sigma^2} \leq y\right] = 1 - e^{-2y}.$$

4 Continuous-Time Markov Chains

4.1 Poisson Process

The *Poisson process* is a special case of birth processes, and a birth process is an example of a continuous-time Markov chain. Without loss of generality, we set $N(0) = 0$. The state space of $N(t)$, $t \geq 0$ is the whole non-negative integers ($N(t) \in \{0, 1, 2, \dots\}$), and $s \mapsto N(s)$ is non-decerasing, i.e. $N(s) \leq N(t)$ whenever $s \leq t$. We set the sample path $s \mapsto N(s)$, right continuous with left limits (aka cadlag). We call such a stochastic process a *counting process*.

Definition 25 (Poisson Process with intensity λ)**.** A counting process $\{N(t), t \geq 0\}$ taking values in $\mathcal{S} = \mathbb{N}_0$ is called a *Poisson process* with (constant) intensity $\lambda > 0$ if the conditional probability satisfies

$$\mathbb{P}[N(t+h) = n + m \mid N(t) = n] = \begin{cases} 1 - \lambda h + o(h), & m = 0 \\ \lambda h + o(h), & m = 1 \\ o(h), & m \geq 2 \end{cases} \quad t \geq 0$$

infinitesimally as $h \downarrow 0$, and $N(t) - N(s)$ is independent of the times of increments $N(\cdot)$ during $[0, s]$ for every $0 \leq s \leq t < \infty$.

Suppose $\{N(t), t \geq 0\}$ is a Poisson process with intensity λ. From the definition let us derive a differential equation for the MGF $G_t(\theta) = \mathbb{E}[e^{\theta N(t)}]$ at time $t \geq 0$ for $\theta < 0$: first using the conditioning and then using the independent increment property:

$$\begin{aligned} G_{t+h}(\theta) = \mathbb{E}[e^{\theta N(t+h)}] &= \mathbb{E}[e^{\theta N(t)}] \cdot \mathbb{E}[e^{\theta(N(t+h)-N(t))} \mid N(t)] \\ &= \mathbb{E}[e^{\theta N(t)}] \cdot \mathbb{E}[e^{\theta(N(t+h)-N(t))}] \\ &= \mathbb{E}[e^{\theta N(t)}(e^{\theta}\lambda h + (1 - \lambda h + o(h)))] \\ &= G_t(\theta) + \lambda h(e^{\theta} - 1)G_t(\theta) + o(h). \end{aligned}$$

Hence $G_t(\theta) = \exp(\lambda(e^{\theta} - 1)t)$, solving the differential equation

$$\frac{d}{dt} G_t(\theta) = \lim_{h \downarrow 0} \frac{G_{t+h}(\theta) - G_t(\theta)}{h} = \lambda(e^{\theta} - 1)G_t(\theta); \quad t \geq 0.$$

Hence we claim the marginal distribution of $N(t)$ is a Poisson distribution with mean λt, i.e.,

$$\mathbb{P}[N(t) = k] = \frac{(\lambda t)^k e^{-\lambda t}}{k!}; \quad k = 0, 1, 2, \dots, \quad t > 0.$$

Definition 26 (Interarrival of Events)**.** We define the *event times* $\{T_n, n \geq 0\}$ recursively with $T_0 = 0$ and $T_n = \inf\{t > 0 \mid N(t) = n\}$, and the *interarrival times* $X_n = T_n - T_{n-1}$, for $n \geq 1$.

If $N(\cdot)$ is a Poisson process with intensity λ, then $\mathbb{P}[X_1 > t] = \mathbb{P}[N(t) = 0] = e^{-\lambda t}$. That is, the first even time $T_1 = X_1$ is distributed exponentially with parameter λ. Moreover,

$$\mathbb{P}[X_{n+1} > t \mid X_1 = t_1, \dots, X_n = t_n] = \mathbb{P}[\text{no event in } [T, T+t]] = \mathbb{P}[N(T+t) - N(T) = 0] = e^{-\lambda t}$$

for every $0 < t_1 < \cdots < t_n$ and $T := t_1 + \cdots + t_n$ and every $t \geq 0$ and $n \geq 1$. Thus X_{n+1} is independent of $X_1, \dots, X_n$ and distributed exponentially with parameter λ for every $n \geq 1$. Thus T_j is the sum of IID exponential random variables, i.e., is Gamma distributed with parameter (λ, j) for $j \in \mathbb{N}$. Note that for every $j \in \mathbb{N}_0$ and $t > 0$,

$$N(t) \geq j \Longleftrightarrow T_j \leq t,$$

and hence

$$\mathbb{P}[N(t) = j] = \mathbb{P}[T_j \leq t < T_{j+1}] = \mathbb{P}[T_j \leq t] - \mathbb{P}[T_{j+1} \leq t] = \frac{(\lambda t)^j e^{-\lambda t}}{j!}.$$

4.2 Birth Process

By setting the intensity depend on state, we can generalize the Poisson process to a birth process.

Definition 27 (Birth Process with intensities $\{\lambda_n\}$)**.** A counting process $\{N(t), t \geq 0\}$ taking values in $\mathcal{S} = \mathbb{N}_0$ is called a *birth process* with intensities $\{\lambda_n > 0, n \geq 0\}$, if the condition probability satisfies

$$\mathbb{P}[N(t+h) = n + m \mid N(t) = n] = \begin{cases} 1 - \lambda_n h + o(h), & m = 0 \\ \lambda_n h + o(h), & m = 1 \\ o(h), & m \geq 2 \end{cases} \quad t \geq 0$$

infinitsimally as $h \downarrow 0$ and $N(t) - N(s)$ is independent of the times of increments of $N(\cdot)$ during $[0, s]$ for every $0 \le s \le t < \infty$.

Example 11 (Simple Birth Process/with Immigration)**.** Fix $\mu > 0$. Each individual gives a birth independently of one another with probability $\lambda h + o(h)$ in the infintiesimal small interval $(t, t+h)$ for every time $t \ge 0$, and no individual dies. Let the total population size be $N(t)$ at time $t \ge 0$. The number M of births in the interval $(t, t+h)$ satisfies the conditional probability

$$\mathbb{P}[M = m \mid N(t) = n] = \binom{n}{m}(\lambda h)^m (1-\lambda h)^{n-m} + o(h).$$

This corresponds to the birth process with $\lambda_n = n\lambda$.

To generalize this, we can take $\lambda_n = n\lambda + \lambda_0$ for $n \ge 0$. λ_0 is called the *immigration*. If $\lambda = 0$, then it just becomes a Poisson process with intensity λ_0.

Let us consider the transition probability $P_{i,j}(t) = \mathbb{P}[N(t) = j \mid N(0) = i]$ for $t \ge 0$ and $i, j \in \mathcal{S}$. It follows by definition that

$$P_{i,j}(t) = \mathbb{P}[N(t) = j \mid N(0) = i] = \mathbb{P}[N(s+t) = j \mid N(s) = i]; \quad s, t \ge 0.$$

Note that $P_{i,j}(t) = 0$ for $i > j$, because the simple path is non-decreasing. Moreover, for this birth process, it always moves up, if it jumps. With this observation, as $h \downarrow 0$, the infinitesimal change of conditional probability between time t and $t+h$ is

$$P_{i,j}(t+h) = \mathbb{E}[\mathbb{P}[N(t+h) = j \mid N(t)] \mid N(0) = i] = P_{i,j-1}(t) \cdot \lambda_{j-1} h + P_{i,j}(t) \cdot (1 - \lambda_j j) + o(h),$$

by conditioning on $N(t)$, and hence we obtain Kolmogorov's forward equation.

Proposition 30 (Kolmogorov's Forward Equation)**.**

$$\frac{d}{dt} P_{i,j}(t) = \lambda_{j-1} P_{i,j-1}(t) - \lambda_j P_{i,j}(t); \quad j \ge i$$

with $\lambda_{-1} = 0$ and $P_{i,j}(0) = \delta_{i,j}$.

This can be generalized to the *Fokker-Planck equation*. Similarly, we can condition on $N(h)$, and we get

$$P_{i,j}(t+h) = \mathbb{E}[\mathbb{P}[N(t+h) = j \mid N(h)] \mid N(0) = i] = P_{i+1,j}(t) \cdot \lambda_i h + P_{i,j}(t) \cdot (1 - \lambda_i h) + o(h),$$

and hence we get Kolmogorov's backward equation.

Proposition 31 (Kolmogorov's Backward Equation)**.**

$$\frac{d}{dt} P_{i,j}(t) = \lambda_i P_{i+1,j}(t) - \lambda_i P_{i,j}(t); \quad j \ge i$$

with $P_{i,j}(0) = \delta_{i,j}$.

Proposition 32. The forward equation has a unique, nonnegative solution. If $P_{i,j}(\cdot)$ is the unique solution to the forward equation, then it satisfies the backward equation and any nonnegative solution $\prod_{i,j}(\cdot)$ to the backward equation satisfies $P_{i,j}(\cdot) \le \prod_{i,j}(\cdot)$.

Proof. Note that for $i > j$, $P_{i,j}(t) = 0$, because the simple path is non-decreasing. Then the forward equation with $j = i$ is rewritten as

$$\frac{d}{dt} P_{i,i}(t) = -\lambda_i P_{i,i}(t),$$

and hence, by setting $P_{i,i}(0) = 1$ and solving the ODE, we obtain $P_{i,i}(t) = e^{-\lambda_i t}$ for $t \ge 0$. Let us derive a recursive relation for the Laplace transform

$$\widehat{P}_{i,j}(\theta) = \int_0^\infty e^{-\theta t} P_{i,j}(t)\, dt; \quad j \ge i,\ \theta \ge 0$$

of $P_{i,j}(\cdot)$ is a solution to the forward equation. By using integration by parts,

$$\begin{aligned}(\lambda_j + \theta)\widehat{P}_{i,j}(\theta) &= \int_0^\infty e^{-\theta t}\left(\lambda_j P_{i,j}(t) + \frac{d}{dt} P_{i,j}(t)\right) dt \\ &= \int_0^\infty e^{-\theta t} \lambda_{j-1} P_{i,j-1}(t)\, dt \\ &= \lambda_{j-1} \widehat{P}_{i,j-1}(\theta); \quad \theta \ge 0,\ j \ge i+1,\end{aligned}$$

and

$$\widehat{P}_{i,i}(\theta) = \int_0^\infty e^{-\theta t} e^{-\lambda_i t}\, dt = \frac{1}{\theta + \lambda_i}; \quad \theta \ge 0.$$

Thus we obtain the Laplace transform of the solution $P_{i,j}(\cdot)$ to the forward equation

$$\widehat{P}_{i,j}(\theta) = \frac{1}{\lambda_j} \prod_{k=1}^{j} \frac{\lambda_k}{\theta + \lambda_k}; \quad j \ge i.$$

By the uniqueness of the inverse Laplace transform, the solution $P_{i,j}(\cdot)$ is uniquely determined.

Similarly, any solution $\prod_{i,j}(\cdot)$ to the backward equation will have its Laplace transform $\widehat{\prod}_{i,j}(\theta) = \int_0^\infty e^{-\theta t} \prod_{i,j}(\cdot)\, dt$ and $\theta \ge 0$ satisfies

$$\widehat{\Pi}_{i,i}(t) = \frac{1}{(\theta + \lambda_i)}$$

and

$$(\lambda_i + \theta)\widehat{\Pi}_{i,j}(\theta) = \int_0^\infty e^{-\theta t} \lambda_i \widehat{\Pi}_{i+1,j}(t)\, dt = \lambda_i \widehat{\Pi}_{i+1,j}(\theta); \quad j \ge i+1.$$

Since the above Laplace transform $\widehat{P}_{i,j}(\theta)$ satisfies the same system of equations:

$$(\lambda_i+\theta)\widehat{P}_{i,j}(\theta) = (\lambda_i+\theta)\cdot\frac{1}{\lambda_j}\cdot\prod_{k=i}^{j}\frac{\lambda_k}{\theta+\lambda_k} = \frac{\lambda_j}{\lambda_i}\prod_{k=i+1}^{j}\frac{\lambda_k}{\theta+\lambda_k} = \lambda_i\widehat{P}_{i+1,j}(\theta); \quad j \ge i+1,$$

again by the uniqueness of inverse Laplace transforms, the solution $P_{i,j}(\cdot)$ of the forward equation satisfies the backward equation.

Finally, in matrix form, the Laplace transformed system of backward equation is rewritten as

$$(\theta\mathbf{I}+\mathbf{\Lambda})\widehat{\mathbf{\Pi}}(\theta) = \mathbf{I}+\mathbf{\Lambda}R\widehat{\mathbf{\Pi}}(\theta),$$

where $\widehat{\prod}(\theta) = (\widehat{\prod}_{i,j}(\theta))_{i,j\ge 0}$, $\mathbf{I}$ is the identity matrix, $\mathbf{\Lambda} = \mathrm{diag}(\lambda_0,\lambda_1,\dots)$, and $\mathbf{R} = (r_{i,j})_{i,j\ge 0}$ with $r_{i,i+1} = 1$ and $r_{i,j} = 0$ for $j \ne i+1$. The minimal solution to this system can be approximated by

$$\widehat{\mathbf{P}}^{(0)} = (\theta\mathbf{I}+\mathbf{\Lambda})^{-1}, \quad \widehat{\mathbf{P}}^{(n+1)} = (\theta\mathbf{I}+\mathbf{\Lambda})^{-1}(\mathbf{I}+\mathbf{\Lambda}R\widehat{\mathbf{P}}^{(n)}(\theta)); \quad n \ge 0,$$

in the sense that $\widehat{\mathbf{P}}^{(n)}(\theta) \le \widehat{\prod}(\theta)$, $n \ge 0$ in elementwise, $\lim_{n\to\infty}\widehat{\mathbf{P}}^{(n)}(\theta) = \widehat{\mathbf{P}}(\theta) = (\widehat{P}_{i,j}(\theta))_{i,j\ge 0}$ (thanks to monotonicity) from the Laplace transform of the solution of the forward equation, and hence $\widehat{P}_{i,j}(\theta) \le \widehat{\prod}_{i,j}(\theta)$ for every $i,j \ge 0$. This completes our proof. □

Here since the sample path is non-decreasing, the mathematical treatment becomes simple. The minimal solution for the forward and backward equations coincide, in general. The problem of determining uniqueness or non-uniqueness amounts to the explosion phenomena. The limit $T_\infty := \lim_{n\to\infty} T_n$ (time of explosion) of event times might be finite.

Definition 28 (Honest/Dishonest)**.** If $\mathbb{P}[T_\infty = \infty] = 1$, we call the counting process *honest*; if the time of explosion comes in a finite time with positive probability, i.e., $\mathbb{P}[T_\infty = \infty] < 1$, then we call the counting process *dishonest*.

Intuition is that if λ_n are large for large n, then the interarrival time length becomes short exepctedly and the birth process explodes in a finite time. For the birth process, we have a simple dichotomy. We have

$$\mathbb{E}[T_\infty] = \mathbb{E}\left[\sum_{n=1}^{\infty} X_n\right] = \sum_{n=1}^{\infty}\frac{1}{\lambda_{n-1}}.$$

Thus if $\sum_{n=0}^{\infty}\lambda_n^{-1} < \infty$, then $\mathbb{E}[T_\infty] < \infty$ and $\mathbb{P}[T_\infty = \infty] = 0$. For the case of $\sum_{n=0}^{\infty}\lambda_n^{-1} = \infty$,

$$\mathbb{E}[e^{-T_\infty}] = \lim_{n\to\infty}\mathbb{E}\left[\prod_{i=1}^{n} e^{-X_i}\right] = \lim_{n\to\infty}\prod_{i=1}^{n}\mathbb{E}[e^{-X_i}] = \lim_{n\to\infty}\prod_{i=1}^{n}\frac{1}{1+\lambda_{i-1}^{-1}} = 0.$$

Thus $\mathbb{P}[T_\infty = \infty] = 1$ if $\sum_{n=0}^{\infty} \lambda_n^{-1} = \infty$.

Let's fix constants T and $i \geq 0$. We can show that the birth process has the Markov property, that is, conditional on the event $\{N(T) = i\}$, the post-T evolution $\{N(t) - N(T), t > T\}$ is independent of the pre-T evolution $\{N(s) \mid 0 \leq s \leq T\}$. This means intuitively that the future and the past are independent conditionally on the present. The birth process, in fact, has a stronger property, known as the *strong Markov property*.

Definition 29 (Stopping Time)**.** We call the random time $\tau : \Omega \to \mathbb{R}^+ \cup \{\infty\}$ a *stopping time* for a birth process $\{N(t), t \geq 0\}$ if $\mathbb{1}_{\{\omega \mid \tau(\omega) \leq t\}}$ is a function of the values $\{N(s), 0 \leq s \leq t\}$ for every $t \geq 0$.

We may decide whether τ occurs or not by time t, knowing only the values $\{N(s) \mid 0 \leq s \leq t\}$ until time t. Note that the event times T_i, $i \geq 0$ are stopping times and a positive constant (as a random time) is a stopping time, however, both $T_4 - 2$ and $(T_1 + T_2)/2$ are not a stopping time. With some measure theory, we can show the following:

Proposition 33 (Strong Markov Property)**.** For a birth process $N(\cdot)$ and a stopping time τ for $N(\cdot)$,

$$\mathbb{P}[A \mid \{N(\tau) = i\} \cap B] = \mathbb{P}[A \mid \{N(\tau) = i\}]$$

where A is a post τ event which depends on $\{N(s), s > \tau\}$ and B is a pre τ event which depends on $\{N(s), s \leq \tau\}$.

When τ is a constant, i.e. $\tau = T$ for some $T > 0$, then the strong Markov property reduces to the (weak) Markov property.

4.3 Continuous-Time Markov Chains

Recall that the birth process has the non-decreasing, cadlag sample paths. In general, the stochastic process can both increase and decrease. We generalize the birth process in the previous section to discuss a continuous-time Markov chain in a countable state space $\mathcal{S}$. Similarly to the discrete case, we call a stochastic process a (continuous) Markov chain if for every $n \geq 1$, $t_0, t_1, \ldots, t_n > 0$ for every $i_1, \ldots, i_{n-1}, j \in \mathcal{S}$,

$$\mathbb{P}[X(t_n) = j \mid X(t_1) = i_1, \ldots, X(t_{n-1}) = i_{n-1}] = \mathbb{P}[X(t_n) = i_{n-1}].$$

Definition 30 (Transition Probability)**.** We denote the transition probability $P_{i,j}(s,t) = \mathbb{P}[X(t) = j \mid X(s) = i]$ for $0 \le s \le t < \infty$, and $i, j \in \mathcal{S}$. We call it (time) homogeneous if $P_{i,j}(s,t) = P_{i,j}(0, t-s)$ for every $0 \le s \le t < \infty$, $i, j \in \mathcal{S}$, and in such a case, we write $P_{i,j}(t-s) = \mathbb{P}[X(t-s) = j \mid X(0) = i]$. The matrix-valued function $\mathbf{P}_t = (P_{i,j}(t))_{i,j\in\mathcal{S}}$ and $t \ge 0$ determines the probability distribution of the continuous-time Markov chain.

We only consider time homogeneous chains. Since $\mathbb{P}[X(0) = i \mid X(0) = i] = 1$ and $\mathbb{P}[X(0) = j \mid X(0) = i] = 0$ if $j \ne i$, we have $\mathbf{P}_0 = \mathbf{I}$. Because of total probability, $\sum_{j\in\mathcal{S}} P_{i,j}(t) = 1$ for every $t \ge 0$, and by definition of conditional probability, we may derive the Chapman-Kolmogorov equations for the continuous case:

$$P_{i,j}(s+t) = \sum_{k\in\mathcal{S}} P_{i,k}(s)P_{k,j}(t) = \sum_{k\in\mathcal{S}} P_{i,k}(t)P_{k,j}(s).$$

Equivalently,

$$\mathbf{P}_{s+t} = \mathbf{P}_s\mathbf{P}_t = \mathbf{P}_t\mathbf{P}_s; \quad s,t \ge 0.$$

This property is called the *semigroup property*, and we say $\mathbf{P}_t$ for $t \ge 0$ is a *stochastic semigroup*.

Definition 31 (Standard Semigroup)**.** The semigroup $\mathbf{P}_t$, $t \ge 0$ is *standard* if $\lim_{t\downarrow 0} \mathbf{P}_t = \mathbf{I}$ element wisely, i.e., $\lim_{t\downarrow 0} P_{i,i}(t) = 1$ and $\lim_{t\downarrow 0} P_{i,j}(t) = 0$ if $j \ne i$.

We assume the map $t \mapsto P_{i,j}(t)$ is continuous for every $i, j \in \mathcal{S}$, and moreover, we assume its differentiability:

$$\lim_{h\downarrow 0} \frac{1}{h}(\mathbf{P}_h - I) = \mathbf{G} = (g_{i,j})_{i,j\in\mathcal{S}}.$$

We call such a limit the *generator* of a Markov chain.

Definition 32 (Uniform Semigroup)**.** The semigroup $\mathbf{P}_t$, $t \ge 0$ is *uniform* if $\lim_{t\downarrow 0} \mathbf{P}_t = \mathbf{I}$ uniformly among elements in $\mathcal{S}$.

- If $\mathcal{S}$ is finite, then the standard semigroup property and the uniform semigroup property are equivalent. In general, uniform semigroups are standard semigroups, but not vice versa.
- If $\mathbf{P}_t$ is standard, then there exists $h > 0$ such that $P_{i,i}(s) > 0$ for every $s \in [0,h]$, and then for every $t \ge 0$, choose n large enough so that $nh \ge t$ or $h \ge t/n$. For such n, we have by

$$P_{i,i}(t) \ge \{P_{i,i}(t/n)\}^n > 0.$$

Thus $P_{i,i}(t) > 0$ for every $t > 0$. For $j \neq i$, the following dichotomy is known as *Lévy dichotomy*: either $P_{i,j}(t) = 0$ for every $t > 0$ or $P_{i,j}(t) > 0$ for every $t > 0$ holds.

Assuming the differentiability of $P_{i,j}(t)$ with respect to t, we approximate $P_{j,j}(h) \approx 1 + g_{j,j}h$ and $P_{k,j}(h) \approx g_{k,j}h$ for small h, and hence,

$$P_{i,j}(t+h) = \sum_{k\in\mathcal{S}} P_{i,k}(t)P_{k,j}(h) \approx \sum_{k\neq j} P_{i,k}(h)g_{k,j}h + P_{i,j}(t)(1+g_{j,j}h); \quad t \geq 0.$$

This leads us to the Kolmogorov forward equation for the continuous case:

$$\frac{d}{dt}P_{i,j}(t) = \lim_{h\downarrow 0}\frac{1}{h}(P_{i,j}(t+h) - P_{i,j}(t)) = \sum_{k\in\mathcal{S}} P_{i,k}(t)g_{k,j},$$

or equivalently,

$$\frac{d}{dt}\mathbf{P}_t = \mathbf{P}_t\mathbf{G}; \quad t \geq 0.$$

Similarly, we form an approximation by conditioning on $X(h)$ first,

$$P_{i,j}(t+h) = \sum_{k\in\mathcal{S}} P_{i,k}(h)P_{k,j}(t) \approx \sum_{k\neq i} g_{i,k}hP_{k,j}(t) + (1+g_{i,i}h)P_{i,j}(t),$$

and this leads to the Kolmogorov backward equation:

$$\frac{d}{dt}P_{i,j}(t) = \lim_{h\downarrow 0}\frac{1}{h}(P_{i,j}(t+h) - P_{i,j}(t)) = \sum_{k\in\mathcal{S}} g_{i,k}P_{k,j}(t),$$

or equivalently,

$$\frac{d}{dt}\mathbf{P}_t = \mathbf{G}\mathbf{P}_t; \quad t \geq 0.$$

Example 12 (Birth Process)**.** The birth process has generator $\mathbf{G} = (g_{i,j})_{i,j\in\mathcal{S}}$ with $\mathcal{S} = \mathbb{N}_0$, $g_{i,i} = -\lambda_i$ and $g_{i,i+1} = \lambda_i$ and $g_{i,j} = 0$ for $j < i$ and $j > i+1$.

Proposition 34 (Kolmogorov Equations)**.** Suppose the Markov chain on a countable state space $\mathcal{S}$ has uniform semigroup $\mathbf{P}_t$. Then it is a unique solution to the forward equation and the backward equation with boundary condition $\mathbf{P}_0 = \mathbf{I}$. Moreover, it is a matrix exponential of $t\mathbf{G}$, i.e.,

$$\mathbf{P}_t = e^{t\mathbf{G}} := \sum_{k=0}^{\infty}\frac{(t\mathbf{G})^k}{k!}; \quad t \geq 0,$$

and $\mathbf{G}(\mathbf{1}') = \mathbf{0}'$.

Suppose $X(s) = i$ and let the holding time $U_i := \inf\{t \geq 0 \mid X(s+t) \neq i\}$ at state $i \in \mathcal{S}$. Then we have for every $x, y > 0$, because of the Markov property,

$$\mathbb{P}[U_i > x + y \mid U_i > x] = \mathbb{P}[U_i > x + y \mid X(s+x) = i] = \mathbb{P}[U_i > y].$$

This implies that the tail probability function $\overline{F}(x) = \mathbb{P}[U_i > x]$, $x > 0$ satisfies $\overline{F}(x+y) = \overline{F}(x) \cdot \overline{F}(y)$ for every $x, y > 0$. Solving this functional equation with $\lim_{u\to\infty} \overline{F}(u) = 0$, we obtain $\overline{F}(x) = e^{-g_{i,i}x}$, where $g_{i,i} = \frac{d}{dt}P_{i,i}(0)$. This means that the holding time U_i is distributed exponentially with parameter $-g_{i,i}(>0)$ for every $i \in \mathcal{S}$.

- If $g_{i,i} = -\infty$, then the state i does not hold the Markov chain and it moves out of state i instantaneously; if $g_{i,i} = 0$, then state i holds the Markov chain forever (state i is an absorbing state), and if $g_{i,i} \in (-\infty, 0)$, we say i is a stable state. Thus it is natural to think the condition $\sup_{i\in\mathcal{S}}(-g_{i,i}) < \infty$. Indeed this is equivalent to the uniform semigroup property.
- In a short time ($h > 0$) a jump occurs with probability $1 - P_{i,i}(h)$, and in that case, it jumps to state $j \in \mathcal{S}$ with probability $P_{i,j}(h)$. Thus we have

$$\mathbb{P}[\text{jumps to } j \mid \text{there is a jump}] \approx \frac{P_{i,j}(h)}{1 - P_{i,i}(h)} \xrightarrow{h\downarrow 0} -\frac{g_{i,j}}{g_{i,i}} = -h_{i,j}; \quad j \neq i.$$

4.4 Embedded Chains

Given a continuous Markov chain $X(t)$, $t \geq 0$, the event jump times are denoted by $T_1, T_2, \ldots$ with $T_0 = 0$. Then $Z_n = X(T_n)$, $t \geq 0$ forms a discrete-time Markov chain called a *jump chain*. The transition probability of Z is given by $h_{i,j}$, $i \neq j$, if $g_{i,i} \in (-\infty, 0)$. The holding time for each state i is distributed exponentially with parameter $-g_{i,i}$.

Conversely, if there is a discrete-time Markov chain Z_n, $n \geq 0$ in $\mathcal{S}$ with transition probability matrix $h_{i,j}$, we define the parameters $g_{i,i} \geq 0$, $i \in \mathcal{S}$ and $g_{i,i} = -h_{i,j}g_{i,i}$. Conditionally on the sample path of $Z_n = i_n$, define a sequence $U_0, U_1, \ldots$ of exponential random variables with parameters $-g_{i_0,i_0}, g_{i_1,i_1}, \ldots$, i.e.,

$$\mathbb{P}[U_n \geq t] = \exp(-g_{i_n,i_n}t); \quad t \geq 0,\ n \geq 0,$$

and define $T_{n+1} = U_0 + \cdots + U_n$, $n \geq 0$ and $T_0 = 0$. We may construct a minimal continuous-time Markov chain $X(\cdot)$ by

$$X(t) = \begin{cases} Z_n, & T_n \leq t < T_{n+1} \text{ for some } n \\ \infty, & \text{otherwise} \end{cases}$$

If $T_\infty = \lim_{n\to\infty} T_n < \infty$, then we call T_∞ an explosion time and move the Markov chain $X(\cdot)$ to a new state $\{\infty\}$, after time T_∞, by extending the state space $\mathcal{S}$ to the new state space $\mathcal{S}\cup\{\infty\}$. The continuous-time Markov chain $X(\cdot)$ constructed by this recipe does not explode if $\mathcal{S}$ is finite, or if $\sup_{i\in\mathcal{S}}(-g_{i,i}) < \infty$ or if Z starts from its persistent state $i \in \mathcal{S}$.

Proposition 35. If $g_{i,i} = 0$, then the state i is persistent. If $g_{i,i} < 0$, state i is persistent if and only if so is for the jump chain Z. Moreover,

$$\int_0^\infty P_{i,i}(t)\,dt = \infty$$

if and only if state i is persistent.

Definition 33 (Irreducible)**.** We say a continuous-time Markov chain is *irreducible* if for every $i, j \in \mathcal{S}$, then there exists t such that $P_{i,j}(t) > 0$.

Definition 34 (Stationary Distribution)**.** We say a probability vector $\boldsymbol{\pi} := (\pi_i,\ i \in \mathcal{S})$ with $\pi_i \geq 0$, $\sum_{j\in\mathcal{S}} \pi_j = 1$ is a *stationary distribution* if $\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}_t$ for every $t \geq 0$.

Since $\mathbf{P}_t = e^{t\mathbf{G}}$, $t \geq 0$, we observe

$$\boldsymbol{\pi}\boldsymbol{P}_t = \boldsymbol{\pi}\sum_{k=0}^{\infty}\frac{(t\mathbf{G})^k}{k!} = \boldsymbol{\pi} + \boldsymbol{\pi}\sum_{k=1}^{\infty}\frac{(t\mathbf{G})^k}{k!}; \quad t \geq 0.$$

Thus $\boldsymbol{\pi}\boldsymbol{P}_t = \boldsymbol{\pi}$ is equivalent $\boldsymbol{\pi}\boldsymbol{G} = 0$. This gives us an alternative definition for a stationary distribution.

Proposition 36. For an irreducible Markov chain with standard semigroup $\mathbf{P}_t$,

(a) If there exists a stationary distribution $\boldsymbol{\pi}$, then it is unique and $\lim_{t\to\infty} P_{i,j}(t) = \pi_j$ for every $i, j \in \mathcal{S}$.

(b) If there is no stationary distribution, then $\lim_{t\to\infty} P_{i,j}(t) = 0$ for every $i, j \in \mathcal{S}$.

Example 13 (Two-State Markov Chain)**.** Suppose that X is a Markov chain with generator with $\alpha, \beta > 0$,

$$G := \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}$$

for state space $\mathcal{S} = \{1, 2\}$. The system $\mathbf{P}_t = \mathbf{P}_t\mathbf{G}$ of the forward equations reads

$$P'_{1,1}(t) = -\alpha P_{1,1}(t) + \beta P_{1,2}(t), \quad P'_{1,2} = \alpha P_{1,1}(t) - \beta P_{1,2}(t)$$

$$P'_{2,1} = -\alpha P_{1,2}(t) + \beta P_{2,1}(t), \quad P'_{2,2}(t) = \alpha P_{1,2}(t) - \beta P_{2,2}(t)$$

Since this has finite state space, the chain is not exploding. Let us diagonalize the generator matrix $\mathbf{G}$:

$$\mathbf{G} = \boldsymbol{B}\boldsymbol{\Lambda}\boldsymbol{B}^{-1}, \quad \mathbf{B} = \begin{pmatrix} \alpha & 1 \\ -\beta & 1 \end{pmatrix}, \quad \boldsymbol{\Lambda} := \begin{pmatrix} -(\alpha+\beta) & 0 \\ 0 & 0 \end{pmatrix}$$

Thus the transition probability is given by

$$\mathbf{P}_t = e^{t\mathbf{G}} = \sum_{n=0}^{\infty} \frac{(t\mathbf{G})^n}{n!} = \mathbf{B}\begin{pmatrix} e^{-(\alpha+\beta)t} & 0 \\ 0 & 1 \end{pmatrix}\mathbf{B}^{-1} = \frac{1}{\alpha+\beta}\begin{pmatrix} \alpha h(t) + \beta & \alpha(1-h(t)) \\ \beta(1-h(t)) & \alpha + \beta h(t) \end{pmatrix},$$

where $h(t) = e^{-(\alpha+\beta)t}$, and its stationary distribution is $\boldsymbol{\pi} = (1-\rho, \rho)$, where $\rho = \frac{\alpha}{\alpha+\beta}$.

Example 14 (Uniformly Constructed Markov Chain)**.** Suppose that Z is a discrete-time Markov chain with transition matrix $(h_{i,j})_{i,j\in\mathcal{S}} = \mathbf{H}$, and let $N(\cdot)$ be a Poisson process with intensity λ and with the nth time of arrival of event T_n, $n \ge 0$, and $T_0 = 0$. Define $X(t) = Z_n$, if $T_n \le t < T_{n+1}$ for $t \ge 0$. Then the corresponding transition semigroup $\mathbf{P}_t$ is given by $\mathbf{P}_t = e^{\lambda t(\mathbf{H}-\mathbf{I}}$, because

$$\begin{aligned} P_{i,j}(t) = \mathbb{P}[X(t) = j \mid X(0) = i] &= \sum_{n=0}^{\infty} \mathbb{P}[X(t) = j, N(t) = n \mid X(0) = i] \\ &= \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} \mathbb{P}[Z_n = j \mid Z_0 = i] \\ &= e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n (\mathbf{H})^n_{i,j}}{n!}; \quad i,j \in \mathcal{S} \end{aligned}$$

4.5 Birth-and-Death Process

Let us consider the continuous *birth-and-death process* with birth rate λ_n and death rate μ_n. The generator $\mathbf{G} = (g_{i,j})_{i,j\ge 0}$ is given by

$$g_{i,i} = \begin{cases} -\lambda_0, & i = 0 \\ -(\lambda_i + \mu_i), & i \ge 1 \end{cases}, \quad g_{i,i+1} = \lambda_i,\ i \ge 0; \quad g_{i,i-1} = \mu_i,\ i \ge 1.$$

If $\lambda_0 = 0$, then there is no jump from the state 0, and hence, the state 0 is absorbing. If $\sup_n(\lambda_n + \mu_n) < \infty$, then it is uniform.

We want to find the stationary distribution $\boldsymbol{\pi}$. We may consider the detailed balance equation for the embedded process or simply $\boldsymbol{\pi}\boldsymbol{G} = \mathbf{0}$, that is,

$$-\lambda_0 \pi_0 + \mu_1 \pi_1 = 0, \quad \lambda_{n-1}\pi_{n-1} - (\lambda_n + \mu_n)\pi_n + \mu_{n+1}\pi_{n+1} = 0; \quad n \ge 1.$$

Solving this we see that if

$$C = \sum_{n=0}^{\infty} \frac{\lambda_0 \cdots \lambda_{n-1}}{\mu \cdots \mu_n} < \infty,$$

then the stationary distribution is given by $\pi_0 = C^{-1}$, and $\pi_n = \lambda_0 \cdots \lambda_{n-1} \pi_0 / (\mu_1 \cdots \mu_n)$ for $n \geq 1$.

Example 15 (Linear Birth-and-Death Process)**.** With $\lambda_n = n\lambda$ and $\mu_n = n\mu$ for some costants $\lambda, \mu > 0$ and with $\mathbb{P}[X(0) = I] = 1$ for some $I \geq 1$, let us compute $p_j(t) = \mathbb{P}[X(t) = j \mid X(0) = I]$. It does not yield uniform semigroup. By the forward equation,

$$p_j'(t) = p_{j-1}(t) \cdot \lambda(j-1) + p_{j+1}(t) \cdot \mu(j+1) - (\lambda + \mu) j p_j(t); \quad t \geq 0,$$

for $j \geq 1$. Multiplying by s^j and summing over j, we obtain a relationship for our familiar generating function $G(s,t) = \sum_{j=0}^{\infty} s^j p_j(t) = \mathbb{E}[s^{X(t)}]$, $s \in [0,1]$, $t \geq 0$:

$$\sum_{j=0}^{\infty} s^j p_j'(t) = \lambda s^2 \sum_{j=1}^{\infty} s^{j-2} p_{j-1}(t) \cdot (j-1) + \mu \sum_{j=0}^{\infty} s^j p_{j+1}(t) \cdot (j+1) - (\lambda + \mu) \sum_{j=0} j s^{j-1} p_j(t),$$

or equivalently, because we have

$$\partial_s G(s,t) = \sum_{j=1}^{\infty} j s^{j-1} p_j(t) = \sum_{j=1}^{\infty} s^{j-2} (j-1) p_{j-1}(t) = \sum_{j=0}^{\infty} s^j p_{j+1}(t),$$

we get the simplified expression

$$\partial_t G(s,t) = (\lambda s - \mu)(s-1) \partial_s G(s,t); \quad s \in [0,1],\ t \geq 0$$

with $G(s,0) = s^I$. The solution will be

$$G(s,t) = \begin{cases} \left(\frac{\lambda t(1-s)+s}{\lambda t(1-s)+1} \right)^I, & \mu = \lambda \\ \left(\frac{\mu(1-s)-(\mu-\lambda s)e^{-t(\lambda-\mu)}}{\lambda(1-s)-(\mu-\lambda s)e^{-t(\lambda-\mu)}} \right)^I, & \mu \neq \lambda \end{cases}$$

We can actually examine the simple birth-and-death process in two different ways.

Embedded Random Walk

Let $T = \inf\{t > 0 \mid X(s+t) \neq n,\ X(s) = n\}$ be the time to the next jump. It is an exponential random variable with parameter $n(\lambda + \mu)$, and at the jump time the transition is simple random walk, that is,

$$\mathbb{P}[X(s+T) = X(s) = 1 \mid X(s) = n] = \frac{\lambda}{\lambda + \mu} = 1 - \mathbb{P}[X(s+T) = X(s) + 1 \mid X(s) = n]; \quad n \geq 1.$$

Note that the continuous-time Markov chain cannot jump more than 1 at the same time.

Embedded Age-Dependent Branching Process

Each individual survives for the exponential random time with parameter $\lambda + \mu$, and dies without offspring with probability $\mu/(\lambda + \mu)$ and dies with two children with probability $\lambda/(\lambda + \mu)$. This is an age dependent branching process. The age distribution is exponential with parameter $\lambda + \mu$ and the generating function of then number N of offspring is given by

$$\mathscr{G}_N(x) := \sum_{k=0}^{\infty} x^k \mathbb{P}[N = k] = x^0 \cdot \frac{\mu}{\mu + \lambda} + x^2 \cdot \frac{\lambda}{\mu + \lambda} = \frac{\mu + x^2\lambda}{\lambda + \mu}.$$

We sum all the individuals and define the sum as $X(t)$ for $t \geq 0$. Recall that assuming $X(0) = 1$, and letting T be the life of the individual with PDF $f_T(\cdot)$, for $u \leq t$, we compute the conditional expectation

$$\mathbb{E}[s^{X(t)} \mid T = u] = \mathbb{E}\left[\prod_{i=1}^{N} \mathbb{E}[s^{X(t-u)} \mid N] \mid T = u\right] = \mathbb{E}[(G(s, t-u))^N] = \mathscr{G}_N(G(s, t-u)).$$

Thus we get

$$\begin{aligned} G(s,t) = \mathbb{E}[s^{X(t)}] &= \mathbb{E}[\mathbb{E}[s^{X(t)} \mid T]] \\ &= \int_0^\infty \mathbb{E}[s^{X(t)} \mid T = u] f_T(u)\,du \\ &= \int_0^\infty (s \cdot \mathbb{1}_{\{u>t\}} + \mathscr{G}_N(G(s, t-u)) \cdot \mathbb{1}_{\{u \leq t\}}) f_T(u)\,du \\ &= \int_0^t \mathscr{G}_N(G(s, t-u)) f_T(u)\,du + \int_t^\infty s f_T(u)\,du; \quad t \geq 0,\ 0 \leq s \leq 1. \end{aligned}$$

In particular, if $f_T(t) = \tilde{\lambda} e^{-\tilde{\lambda} t}$, we have by change of variables,

$$\int_0^t \mathscr{G}_N(G(s, t-u)) f_T(u)\,du = e^{-\tilde{\lambda} t} \int_0^t \mathscr{G}_N(G(s, v)) \tilde{\lambda} e^{\tilde{\lambda} v}\,dv,$$

and hence it follows from

$$\begin{aligned} \partial_t G(s,t) &= \partial_t \left(\int_0^t \mathscr{G}_N(G(s, t-u)) f_T(u)\,du + \int_t^\infty s f_T(u)\,du \right) \\ &= -\tilde{\lambda} \int_0^t \mathscr{G}_N(G(s, v)) \tilde{\lambda} e^{-\tilde{\lambda}(t-v)}\,dv + \tilde{\lambda} \mathscr{G}_N(G(s,t)) - \tilde{\lambda} s e^{-\lambda t} \\ &= \tilde{\lambda} \mathscr{G}(G(s,t)) - \tilde{\lambda}(G(s,t) - s e^{-\tilde{\lambda} t}) - \tilde{\lambda} s e^{-\tilde{\lambda} t}, \end{aligned}$$

that

$$\partial_t G(s,t) = \tilde{\lambda}(\mathcal{G}(G(s,t)) - G(s,t)); \quad t \geq 0,\ 0 \leq s \leq 1.$$

The boundary condition is $G(s,0) = s$ for $0 \leq s \leq 1$. Substituting $\mathcal{G}_N(x)$ with $\tilde{\lambda} = \lambda + \mu$, we obtain the forward equation for

$$\partial_t G(s,t) = \lambda[G(s,t)]^2 - (\lambda + \mu)G(s,t) + \mu.$$

This is a backward equation of the process with respect to t. Then going back to the functional relation, we can rewrite it as an integral equation:

$$\int_s^{G(s,t)} \frac{1}{\mathcal{G}_N(u) - u}\, du = \tilde{\lambda} t,$$

if $\mathcal{G}_N(u) \neq u$ for $s \leq u \leq G(s,t)$.

Proposition 37 (Dynkin's Theorem)**.**

$$G(1,t) = \sum_{j=0}^{\infty} \mathbb{P}[X(t) = j] = 1 \quad \text{iff} \quad \int_{1-\epsilon}^{1} \frac{1}{\mathcal{G}_N(u) - u}\, du = \infty \text{ for every } \epsilon \in (0,1).$$

Proof. Let us choose $s_0 \in (0,1)$ such that $\mathcal{G}_N(s) \neq s$ for every $s_0 < s < 1$, and then choose s_1, t_1 with $s_0 < s_1 < 1$ and $t_1 > 0$ such that $s_1 - \tilde{\lambda} t_1 > s_0$. It follows that $|G(s,t)| < \tilde{\lambda}$, and hence $G(s,t) \geq s - \tilde{\lambda} t > s_0$ for $s_1 < s < 1$ and $0 < t < t_1$. Also we have $|G(s,t)| < 1$ for every $|s| < 1$. Thus we have

$$\int_s^{G(s,t_1)} \frac{du}{\mathcal{G}_N(u) - u} = \tilde{\lambda} t_1, \quad s_1 < s < 1.$$

As $s \uparrow 1$, if

$$\int_{1-\epsilon}^{1} \frac{du}{\mathcal{G}_N(u) - u} < \infty,$$

we must have $G(1,t_1) < 1$. If instead

$$\int_{1-\epsilon}^{1} \frac{du}{\mathcal{G}_N(u) - u} = \infty,$$

then we must have $G(1,t_1) = 1$. Since we have either $G(1,t) = 1$ for every $t \geq 0$ or $G(1,t) < 1$ for every $t \geq 0$, we conclude our proof. □

If $\mathcal{G}_N'(1) < \infty$, by Taylor expansion $\mathcal{G}_N(u) - u = (\mathcal{G}_N'(u) - 1)(u-1) + o(u-1)$ in the neighborhood of $u = 1$, we have $G(1,t) = 1$.

4.6 Special Continuous-Time Markov Chains

Non-Homogeneous Simple Birth-and-Death Process

Consider the case of the birth-and-death process where the birth and death rates are dependent on the time t, that is, we set $\lambda_n = n\lambda(t)$ and $\mu_n = n\mu(t)$ for some functions λ and μ. If the functions are constant, obviously this becomes the time-homogeneous birth-and-death processes. Suppose the initial value is $X(0) = 1$. Then the transition probability is $p_j(t) = \mathbb{P}[X(t) = j]$ satisfies the Kolmogorov equation

$$\dot{p}_j(t) = (j-1)\lambda(t)p_{j-1}(t) - j(\lambda(t)+\mu(t))p_j(t) + (j+1)\mu(t)p_{j+1}(t); \quad j \geq 1,$$

where $\dot{p}_0(t) = \mu(t)p_1(t)$.

Following the derivation of the PGF of the homogeneous simple birth-and-death process, we replace the constants with their respective functions, the PGF $G(s,t)$ satisifes

$$\frac{\partial G}{\partial t}(s,t) = (s\lambda(t) - \mu(t))(s-1)\frac{\partial G}{\partial s}(s,t)$$

with the boundary condition $G(s,0) = s$.

Nonlinear Epidemic

Suppose we have $N+1$ individuals among which there is one sick individual and there are N healthy individuals at time 0. Let $X(t)$ be the number of healthy individuals, with $X(0) = N$. Assume there is no cure, and if $X(t) = n$, then the probability there of new infection in a short time is proportional to the possible encounters among the individuals, i.e.,

$$\mathbb{P}[X(t+h) = n+1 \mid X(t) = n] = \lambda n(N+1-n)h + o(h); \quad t \geq 0$$

as $h \downarrow 0$. Then the PGF

$$G(s,t) = \mathbb{E}[s^{X(t)}] = \sum_{k=0}^{N+1} s^k \mathbb{P}[X(t) = k]$$

satisfies

$$\begin{aligned} G(s,t+h) &= \mathbb{E}[\mathbb{E}[s^{X(t+h)} \mid X(t)]] \\ &= \sum_{k=0}^{N} s^k (1 - \lambda k(N+1-k)h)\mathbb{P}[X(t) = k] + \sum_{k=0}^{N} s^{k+1}\lambda k(N+1-k)h\mathbb{P}[X(t) = k] + o(h), \end{aligned}$$

and thus we have

$$\frac{1}{h}(G(s,t+h) - G(s,t)) = \sum_{k=0}^{N} s^{k+1}\lambda k(N+1-k)\mathbb{P}[X(t) = k] - \sum_{k=0}^{N} s^k \lambda k(N+1-k)\mathbb{P}[X(t) = k] + o(h)$$

for $s \in (-1,1)$ as $h \downarrow 0$. The LHS becomes the time-derivative of $G(s,t)$ and the RHS is represented by the first and second derivatives of $G(s,t)$ with respect to s:

$$\frac{\partial G}{\partial t}(s,t) = \lambda(1-s)\left(N \cdot \frac{\partial G}{\partial s}(s,t) - s \cdot \frac{\partial^2 G}{\partial s^2}(s,t)\right); \quad s \in (-1,1),\ t \geq 0$$

with the boundary condition $G(s,0) = s^N$ for some $s \in (-1,1)$. There is no explicit solution to this.

Birth-and-Death Process with Immigration

Consider the population dynamics that have different subpopulations that enter the system in different random times described by a Poisson process. Let $\{N_0(\cdot) = N(\cdot), N_1(\cdot), \ldots\}$ be the independent simple birth-and-death process and $\{I(t), t \geq 0\}$ is an independent Poisson process with parameter ν and the jump times $T_i = \inf\{t > 0 \mid I(t) = i\}$, $i \geq 1$, and $T_0 = 0$. Define the birth-and-death process with immigration by

$$Y(t) := \sum_{i=0}^{I(t)} N_i(t - T_i); \quad t \geq 0.$$

Since each $N_i(t - T_i)$ for $t \geq T_i$ is a simple birth-and-death process starting from time T_i, it represents a new subpopulation immigrates into the system at time T_i with initial value $N_i(0)$. For simplicity, suppose they have identical distribution with $N_i(0) = 1$. We want to derive the PGF of $Y(t)$.

Proposition 38. The conditional distribution of $T_1, \ldots, T_n$, conditional on the set $\{I(t) = n\}$ is the same as the joint distribution of the order statistics of n independent uniform random variables $U_1, \ldots, U_n$ on the interval $[0,t]$ for $t \geq 0$.

Proof. The joint distribution of $T_1, \ldots, T_n$ is obtained from the independent, exponential interarrival times property:

$$\mathbb{P}[T_1 \in dt_1, \ldots, T_n \in dt_n] = \nu^n e^{-\nu t_n} \cdot \mathbb{1}_{\{0<t_1<t_2<\cdots<t_n\}} dt_1 \cdots dt_n,$$

and by the Markov property, the conditional probability that $I(t) = n$, given the values of $T_1, \ldots, T_n$ are

$$\mathbb{P}[I(t) = n \mid T_1 = t_1, \ldots, T_n = t_n] = \mathbb{P}[I(t) = n \mid T_n = t_n] = \mathbb{P}[T_{n+1} - T_n > t - t_n] = e^{-\nu(t-t_n)},$$

for $n \geq 1$. By the definition of conditional probability, the conditional density of $T_1, \dots, T_n$ given $I(t) = n$ is

$$\begin{aligned}\mathbb{P}[T_1 \in dt_1, \dots, T_n \in dt_n \mid I(t) = n] &= \frac{\mathbb{P}[I(t) = n \mid T_1 = t_1, \dots, T_n = t_n] \cdot \mathbb{P}[T_1 \in dt_1, \dots, T_n \in dt_n]}{\mathbb{P}[I(t) = n]} \\ &= \frac{e^{-\nu(t-t_n)} \nu^n e^{-\nu t_n}}{(\nu t)^n e^{-\nu t}/n!} \cdot \mathbb{1}_{\{t_1 < \dots < t_n\}} \\ &= \frac{n!}{t^n} \cdot \mathbb{1}_{\{0 < t_1 < \dots < t_n\}}.\end{aligned}$$

This is the PDF of the order statistics of n independent uniform random variables $U_1, \dots, U_n$ on $[0, t]$. □

We use this to calculate $G_Y(\cdot)$:

$$\begin{aligned}G(s,t) = \mathbb{E}[s^{Y(t)}] &= \mathbb{E}[\mathbb{E}[s^{\sum_{i=1}^{I(t)} N_i(t-T_i)} \mid I(t)]] \cdot \mathbb{E}[s^{N_0(t)}] \\ &= \mathbb{E}[\mathbb{E}[s^{\sum_{i=1}^{n} N_i(t-U_i)}] \mid n = I(t)] \cdot G_N(s,t),\end{aligned}$$

where $G_N(s,t) = \mathbb{E}[s^{N(t)}]$ is the PGF of the simple birth-and-death process $N(\cdot)$ and the first term in the product can be further expanded by the i.i.d. property of $U_1, \dots, Y_k$ that

$$\begin{aligned}\mathbb{E}[\mathbb{E}[s^{\sum_{i=1}^{n} N_i(t-U_i)}] \mid n = I(t)] &= \mathbb{E}[(\mathbb{E}[s^{N_i(t-U_i)}])^{I(t)}] \\ &= \mathbb{E}[(\mathbb{E}[s^{N(t-U)}])^{I(t)}] \\ &= \mathbb{E}\left[\left(\int_0^t \frac{1}{t} G_N(t-u)\, du\right)^{I(t)}\right] \\ &= \exp\left(\nu t\left(\int_0^t \frac{1}{t} G_N(s, t-u)\, du - 1\right)\right) \\ &= \exp\left(\nu \int_0^t (G_N(s,u) - 1)\, du\right); \quad t \geq 0.\end{aligned}$$

Thus we conclude that

$$G(s,t) = G_N(s,t) \exp\left(\nu \int_0^t (G_N(s,t) - 1)\, du\right); \quad s \in (-1,1),\ t \geq 0.$$

4.7 Poisson Point Process

As a generalization of Poisson processes, we can study the randomly scattered (countable) points in $\mathbb{R}^d$. For example, stars in the sky, distribution of blackholes, distribution of plants, locations of crimes in a metropolitan area, etc.

Definition 35 (Homogeneous Poisson Process)**.** A countable subset Π of $\mathbb{R}^d$ is called a *homogeneous Poisson process* with constant intensity $\lambda > 0$, if for every measurable set $A \subseteq \mathbb{R}^d$, $N(A) = |\Pi \cap A|$ is a Poisson random variable with parameter $\lambda|A|$, and if for every n and every disjoint sets $A_1, \ldots, A_n$, $N(A_1), \ldots, N(A_n)$ are independent.

Definition 36 (Inhomogeneous Poisson Process)**.** A countable subset Π of $\mathbb{R}^d$ is called a *inhomogeneous Poisson process* with intensity $\lambda : \mathbb{R}^d \to \mathbb{R}_+$, if for every measurable set $A \subseteq \mathbb{R}^d$, $N(A) = |\Pi \cap A|$ is a Poisson random variable with parameter $\lambda|A|$, and if for every n and every disjoint sets $A_1, \ldots, A_n$, $N(A_1), \ldots, N(A_n)$ are independent.

Here $N(A) = |\Pi \cap A|$ is the number of points in the set A and $|A|$ is the volume of the set A with respect to the Lebesgue measure. We allow $|A| = \infty$ and in that case, we have $\mathbb{P}[|\Pi \cap A| = \infty] = 1$. As a special case $d = 1$, it is reduced to the Poisson process. Since

$$\mathbb{E}[N(A)] = \int_A \lambda(x)\, dx = \Lambda(A),$$

the function $\Lambda(\cdot)$ is called a mean measure.

Proposition 39 (Superposition)**.** Let Π_i for $i = 1, 2$ be independent Poisson processes with parameter λ_i. Then $\Pi_1 \cup \Pi_2$ is a Poisson point process with intensity $\lambda_1 + \lambda_2$.

Proposition 40 (Mapping Theorem)**.** Let $f : \mathbb{R}^d \to \mathbb{R}$ be a measurable function such that $\Lambda(f^{-1}(\{y\})) = 0$ for every $y \in \mathbb{R}^s$, i.e. no multiple points in $f(\Pi)$. Assume that $\Lambda(f^{-1}(B)) = \int_{f^{-1}(B)} \lambda(x)\, dx < \infty$ for every bounded measurable set B. Then $f(\Pi)$ is a Poisson point process with mean measure $\Lambda(f^{-1}(\cdot))$.

Example 16 (Polar Coordinates)**.** Consider the map $f : \mathbb{R}^2 \to \mathbb{R}$ of the polar coordinates, $f(x, y) = (r, \theta)$ where $r = \sqrt{x^2 + y^2}$ and $\theta = \arctan(y/x)$, and a Poisson point process with constant intensity λ. Then $f(\Pi)$ is a Poisson point process with the mean measure

$$\int_{f^{-1}(B)} \lambda\, dx\, dy = \int_{B \cap f(\mathbb{R}^2)} \lambda r\, dr\, d\theta$$

by change of variables. Here $f(\mathbb{R}^2) = \{(r, \theta) \mid r \geq 0,\ 0 \leq \theta < 2\pi\}$ is an infinite strip of width 2π, and so $f(\Pi)$ represents a Poisson point process is an infinte strip.

Proposition 41 (Conditional Property)**.** Let Π be a Poisson point process with intensity λ and A be a subset of $\mathbb{R}^d$ with $0 < \Lambda(A) < \int_A \lambda(x)\,dx < \infty$. Given $N(A) = |\Pi \cap A| = n$, the n random points in A have the same distribution as n points chosen independently at random in A, according the probability $\mathbb{Q}(B) = \Lambda(B)/\Lambda(A)$, for $B \subseteq A$.

When $\lambda(\cdot)$ is constant, then $\mathbb{Q}(B)$ is a uniform measure. This generalizes Proposition 38.

Proof. Let $A_1,\ldots,A_k$ be a partition of A. By the definition of Poisson point process, for every $(n_1,\ldots,n_k) \in \mathbb{N}_0^k$ with $n_1 + \cdots + n_k = n$, we have

$$\begin{aligned}\mathbb{P}[N(A_1) = n_1,\ldots,N(A_k) = n_k \mid N(A) = n] &= \frac{\mathbb{P}[N(A_1) = n]\cdots\mathbb{P}[N(A_k) = n_k]}{\mathbb{P}[N(A) = n]} \\ &= \frac{n!}{n_1!\cdots n_k!}\cdot\mathbb{Q}(A_1)^{n_1}\cdots\mathbb{Q}(A_k)^{n_k}.\end{aligned}$$

This multinomial distribution is the joint distribution of n points selected independently from A, according to $\mathbb{Q}$. □

Proposition 42 (Thinning/Coloring)**.** Each point at x in the Poisson process Π with intensity λ is colored with probability $\gamma_1(x)$ and not colored with probability $1 - \gamma_1(x) = \gamma_2(x)$. Then the random set Π_1 of colored points and the random set Π_2 of non-colored points are independent Poisson point process with intensity $\lambda\cdot\gamma_1$ and $\lambda_2\cdot\gamma_2$ respectively.

Proof. Using the above Theorem, we can compute the distribution directly. For every (n_1, n_2) with $n_1 + n_2 = n$, we have

$$\begin{aligned}\mathbb{P}[|\Pi_1| = n_1, |\Pi_2| = n_2] &= \mathbb{P}[|\Pi_1| = n_1, \Pi_2| = n_2 \mid |\Pi| = n]\cdot\mathbb{P}[|\Pi| = n] \\ &= \frac{n!}{n_1!n_2!}\left(\int_A \gamma_1(x)\,d\mathbb{Q}(x)\right)^{n_1}\cdot\left(\int_A \gamma_2(x)\,d\mathbb{Q}(x)\right)^{n_2}\cdot\frac{[\Lambda(A)]^n e^{-\Lambda(A)}}{n!} \\ &= \frac{\tilde{\Lambda}(A)^{n_1}e^{-\tilde{\Lambda}_1(A)}}{n_1!}\cdot\frac{\tilde{\Lambda}_2(A)^{n_2}e^{-\tilde{\Lambda}_2(A)}}{n_2!},\end{aligned}$$

where

$$\tilde{\Lambda}_i(A) = \int_A \gamma_i(x)\,d\mathbb{Q}(x)\cdot\Lambda(A) = \int_A \gamma_1(x)\cdot\lambda(x)\,dx,$$

because of the definition of the conditional probability measure $\mathbb{Q}$ for $i = 1,2$. Therefore we have concluded that Π_1 and Π_2 are independent Poisson point processes with intensity $\lambda\cdot\gamma_1$ and $\lambda\cdot\gamma_2$ respectively. □

Now we would like to characterize the empty set of Poisson point processes. This corresponds to the exponential tail of Poisson processes with intensity λ, i.e., $\mathbb{P}[N(t) = 0] = e^{-\lambda t}$ for $t \geq 0$.

Proposition 43 (Rényi's Theorem)**.** Let Π be a countable set of random points and $\lambda : \mathbb{R}^d \to \mathbb{R}_+$ is a measurable function with $\Lambda(A) = \int_A \lambda(x)\,dx$ for every measurable set $A \subseteq \mathbb{R}^d$. If $\mathbb{P}[\Pi \cap A = \emptyset] = e^{-\Lambda(A)}$ for every A of the form of finite union $\bigcup_k B_k(n)$ of some boxes $B_k(n)$ in the proof of the Superposition Proposition, then Π is a Poisson point process with intensity λ.

Proof. Let A be a bounded open set. We approximate $N(A) = |\Pi \cap A|$ by the sequence $T_n(A) = \sum_{k|B_k(n)\subseteq A} \mathbb{1}_{\{\Pi \cap B_k(n) \neq \emptyset\}}$. We want to compute the generating function by the independence property

$$\mathbb{E}[s^{T_n(A)}] = \prod_{k|B_k(n)\subseteq A} (s(1 - e^{-\Lambda(B_k(n)}) + e^{-\Lambda(B_k(n))}) = \prod_{k|B_k(n)\subseteq A} (s + (1-s)e^{-\Lambda(B_k(n))}).$$

Using convexity and Taylor series expansions, for every $\delta > 0$, we claim that there exists ϵ such that $e^{-(1-s)x} \leq s + (1-s)e^{-x} \leq e^{-(1-s)x\epsilon}$ for $0 < x < \delta$ with $\lim_{\delta\to 0} \epsilon = 1$. Using this inequality with $\delta = \max_{k|B_k(n)\subseteq A} \Lambda(B_k(n)) = M_n(A;\Lambda)$. □

5 Convergence of Random Variables

5.1 Introduction

Recall that in real analysis, we say a sequence of real numbers a_n converges to some number a as $n \to \infty$ ($a_n \to a$ as $n \to \infty$) if for all $\epsilon > 0$, there exists some $N = N_\epsilon$ such that $|a_n - a| < \epsilon$ for all $n \geq N_\epsilon$.

- If $a = \infty$, then $a_n \to \infty$ means that $\forall \Delta > 0$, there exists $N = N_\Delta$ such that $a_n > \Delta$ for all $n \geq N_\Delta$.
- Another way to prove $a_n \to a$ is the Cauchy criterion.

Why do we need to discuss the convergence of random variables separately? This is because $X_n \to X$ as $n \to \infty$ does not make sense. Random variables X_n and X are functions on a sample space $\Omega : X_n = X_n(\omega)$. Convergence of random variables follows much more closely the convergence of a sequence of functions. Recall the different forms:

- Pointwise: $\forall x \in [0,1]$, $f_n(x) \to f(x)$.
- Uniform: $\forall x \in [0,1]$, $\forall \epsilon > 0$, $\exists N = N_\epsilon$ such that $|f_n(x) - f(x)| < \epsilon$ for all $n \geq N_\epsilon$.
- Almost everywhere: There exists a set K in the Borel σ−algebra $\mathcal{B}([0,1])$, such that $\mu(K) = 1$ and $\forall x \in K$, $f_n(x) \to f(x)$.
- In L^p:
$$\left(\int_0^1 |f_n(x) - f(x)|^p \, d\mu(x)\right)^{1/p} \to 0, \quad p \geq 1.$$
- In norm: $\|f_n - f\| \to 0$.

In a probability theory, these different modes of convergence get physical interpretations.

Example 17. Consider the branching process. If each individual family has on average μ children, then the average population size is $\mathbb{E}[Z_n] = \mu^n$.

- If $\mu < 1$, then $\eta = \mathbb{P}[Z_n \to 0] = 1$, that is, almost surely the population size goes to 0: this population dies out. Also on average, $\mathbb{E}[Z_n] = \mu^n \to 0$.
- If $\mu = 1$, while the individual family size remains random ($\sigma^2 > 0$), then $\eta = \mathbb{P}[Z_n \to 0] = 1$, that is, almost surely the population size goes to 0, while the average $\mathbb{E}[Z_n] = \mu^n \to 1$. This is counter-intuitive.
- If $\mu > 1$, then $\mathbb{E}[Z_n] = \mu^n \to \infty$. The probability of extinction $\eta = \mathbb{P}[Z_n \to 0] < 1$, so that the sample space Ω splits into three sets: $\Omega = \Omega_1 \cup \Omega_2 \cup \Omega_3$. Here $\mathbb{P}[\Omega_1] = \eta$, $\mathbb{P}[\Omega_2] = 1 - \eta$, $\mathbb{P}[\Omega_3] = 0$ and $Z_n(\omega) \to 0$ for all $\omega \in \Omega_1$ and $Z_n(\omega) \to \infty$ for all $\omega \in \Omega_2$.

Definition 37 (Forms of Convergence)**.** Let $X_1, X_2, \ldots$ be random variables defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

- *Almost sure convergence*: We say $X_n \xrightarrow{a.s.} X$ iff

$$\mathbb{P}\left[\{\omega \in \Omega \mid X_n(\omega) \to X(\omega) \text{ as } n \to \infty\}\right] = 1.$$

- *Convergence in probability*: We say $X_n \xrightarrow{\mathbb{P}} X$ iff

$$\forall \epsilon > 0, \quad \mathbb{P}[\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \epsilon\}] \to 0 \text{ as } n \to \infty.$$

- *Convergence in rth mean*: We say $X_n \xrightarrow{r} X$ iff

$$\mathbb{E}[|X_n|^r] < \infty, \quad \forall n \text{ and } \mathbb{E}[|X_n - X|^r] \to 0 \text{ as } n \to \infty.$$

  If $r = 1$, this is simply called convergence in mean, and if $r = 2$, this is called convergence in quadratic mean.

- *Convergence in distribution*: We say $X_n \xrightarrow{\mathcal{D}} X$ iff

$$F_{X_n}(x) \to F_X(x) \text{ as } n \to \infty,$$

  for all real x at which F_X is continuous. except for any discontinuous points of $F_X(\cdot)$.

Let L_r denote the space of all RVs on $(\Omega, \mathcal{F}, \mathbb{P})$ with finite rth moment; that is,

$$L_r := \{Y : \Omega \to R \text{ such that } \mathbb{E}[|Y|^r] < \infty\}.$$

Then $\mathbb{E}[|X_n|^r] < \infty$ is equivalent to $X_n \in L_r$. Also

- $(\mathbb{E}[|X_n|^r])^{1/r} = \left(\int |x|^r \, dF_{X_n}(x)\right)^{1/r} = \|X_n\|_r$ is a norm for $r \geq 1$.

- $d(X,Y) = (\mathbb{E}[|X-Y|^r])^{1/r} = \|X-Y\|_r$ is a metric on L_r.

Thus for $r \geq 1$,

$$\{\text{convergence in } r\text{th mean}\} \iff \{L_r \text{ convergence}\} \iff \{\text{convergence in } L_r \text{ norm}\}.$$

Now we introduce the idea of *mutual convergence*. We use this when we have to prove convergence but do not have a good candidate for the limit X. Mutual convergence requires only information on pairs (X_n, X_m).

Proposition 44 (Mutual Convergence).

(a) $X_n \to X$ a.s. iff $\mathbb{P}[\{\omega \in \Omega : X_n(\omega) - X_m(\omega) \to 0 \text{ as } n, m \to \infty\}] = 1$.

(b) For $r \geq 1$, $X_n \xrightarrow{r} X$ iff $\mathbb{E}[|X_n|^r] < \infty$ for all n and $\mathbb{E}[|X_n - X_m|^r] \to 0$ when $n, m \to \infty$.

(c) $X_n \xrightarrow{\mathbb{P}} X$ iff $\forall \epsilon > 0$, $\mathbb{P}[\{\omega \in \Omega : |X_n(\omega) - X_m(\omega)| > \epsilon] \to 0$ as $n, m \to \infty$.

Proof. (a) This is trivial: we apply Cauchy's criterion for each real sequence $\{X_n(\omega)\}$.

(b) This asserts that the space L_r is complete in the L_r–norm. This is a well known fact; such spaces are called *Banach spaces*.

(c) Not trivial.

□

5.2 Borel-Cantelli Lemma

The Borel-Cantelli Lemma is the major tool for proving almost sure convergence. Recall the idea of the limit superior: Let $A_1, A_2, \ldots$ be an infinite sequence of events on our probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If we conduct an experiment (i.e. picking some ω in Ω), how many of the events A_k's occurred? More precisely, does $\omega \in A_k$ for finite or infinite number of events A_k?

Let $B_n = \bigcup_{k=n}^{\infty} A_k$. The sequence B_n is decreasing:

$$B_n \supset B_{n+1} \supset B_{n+2} \supset \cdots$$

Because decreasing sequences always have limits, we define

$$A^\infty := \lim_{n\to\infty} B_n = \bigcap_{n=1}^{\infty} B_n = \bigcap_{n=1} \bigcup_{k=n}^{\infty} A_k = \limsup_n A_n.$$

Proposition 45. $\omega \in A^\infty$ iff $\omega \in A_n$ for infintitely many values of n.

We interpret this as $A^\infty = \limsup_n A_n = \{A_n \text{ i.o.}\}$, that is, $\omega \in A^\infty$ iff $\omega \in A_k$ for infinitely many A_k's. This means that there exists an infinite subsequence $k_1, k_2, \ldots$ such that $\omega \in A_{k_m}$ for all $m = 1, 2, \ldots$.

Proof.

- $\omega \in A^\infty$ implies $\omega \in B_n = \bigcup_{k=n}^\infty A_k$ for all $n \geq 1$. Thus, for each n, there exists $k \geq n$ such that $\omega \in A_k$. We conclude that $\omega \in A_k$ for infinitely many k's.
- $\omega \notin A^\infty$ implies that there exists n such that $\omega \notin B_n$ for this particular n. Thus, $\omega \notin A_k$, for all $k \geq n$, i.e., ω can only be in $\{A_1, \ldots, A_{n-1}\}$ or in a subset of these (finite number of A_k's).

□

Proposition 46 (Borel-Cantelli Lemma, Part 1)**.** If $\sum_{n=1}^\infty \mathbb{P}[A_n] < \infty$, then $\mathbb{P}[A_n \text{ i.o.}] = \mathbb{P}[A^\infty] = 0$.

Proof. The decreasing sequence of events $\{B_n = \bigcup_{k=n}^\infty A_k\} \downarrow A^\infty$. Thus

$$0 \leq \mathbb{P}[A^\infty] \leq \mathbb{P}\left[\bigcup_{k=n}^\infty A_k\right] \leq \sum_{k=n}^\infty \mathbb{P}[A_k] \overset{n\to\infty}{\to} 0,$$

because it is the remainder of a convergent series. □

This is usually used to prove almost sure convergence of a sequence.

Example 18. We want to prove that $X_n \to X$ almost surely. We choose $\{a_k\} \downarrow 0$ and define events $A_k = \{\omega : |X_k(\omega) - X(\omega)| \geq a_k\}$. Assume that $\{a_k\}$ is chosen so that $\sum_{k=1}^\infty \mathbb{P}[A_k] < \infty$ (*). By Borel-Cantelli, $\mathbb{P}[A_k \text{ i.o.}] = 0$; that is, the probability of its complement is 1:

$$\mathbb{P}[\{A_k \text{ i.o.}^c\}] = \mathbb{P}[\{A_k \text{ happens for finitely many of } k\text{'s}\}] = 1.$$

Because a finite number of indices k has a finite maximum, say n_0, we conclude that with probability 1 A_k does not happen for $k > n_0$. That is, w.p.1 for $k > n_0$ we have

$$A_k^c = \{\omega : |X_k(\omega) - X(\omega)| < a_k\}.$$

Because $\{a_k\} \downarrow 0$ this means that w.p.1, $|X_k(\omega) - X(\omega)| \downarrow 0$.

Another way to see this is because $\{A_k \text{ i.o.}\} = A^\infty$, then $\{A_k \text{ i.o. }\}^c = (A^\infty)^c = \bigcup_{n=1} \bigcap_{k=n}^\infty A_k^c$. We now have

$$\mathbb{P}[\{A_k \text{ i.o.}\}^c] = \mathbb{P}[\{\omega : \exists n_0 \text{ s.t. } \forall k \geq n_0, \omega \in A_k^c\}] = \mathbb{P}[\{\omega : |X_k(\omega) - X(\omega) < a_k\}] = 1.$$

The key to proving this is to show (*) for a particular sequence $\{a_k\} \downarrow 0$. The Markov inequality often helps:

$$\mathbb{P}[|X_k - X| \geq a_k] = \mathbb{P}[|X_k - X|^p \geq a_k^p] \leq \mathbb{E}[|X_k - X|^p]/a_k^p, \text{ for } p \geq 1.$$

Proposition 47. If $X_n \xrightarrow{\mathbb{P}} X$, then there exists a non-random sequence $\{n_1, n_2, \ldots, n_k, \ldots\} \uparrow \infty$ such that $X_{n_k} \to X$ a.s. as $k \to \infty$.

Proof. Suppose $X_n \xrightarrow{\mathbb{P}} X$. Fix an integer $k > 0$ and choose $\epsilon = 1/k$. By definition of convergence in probability, for this ϵ, this holds:

$$\mathbb{P}[\{\omega : |X_n(\omega) - X(\omega)| > 1/k\}] \to 0 \text{ as } n \to \infty.$$

Thus there exists n_k such that $\mathbb{P}[\{\omega : |X_{n_k}(\omega) - X(\omega)| > 1/k\}] \leq 1/k^2$. We denote

$$A_k = \{\omega : |X_{n_k}(\omega) - X(\omega)| > 1/k\},$$

and so $\mathbb{P}[A_k] \leq 1/k^2$. By Borel-Cantelli, we can show that $X_{n_k} \to X$ a.s. as $k \to \infty$,

$$\sum_{k=1}^{\infty} \mathbb{P}[A_k] \leq \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty$$

implies that $\mathbb{P}[A_k \text{ i.o.}] = \mathbb{P}[\{\omega : |X_n(\omega) - X(\omega)| > 1/k i.o.\}] = 0$, that is, with probability 1 events A_k happen only for a finite number of k's and there exists n_0 such that $A_k^c = \{\omega : |X_{n_k}(\omega) - X(\omega)| \leq 1/k\}$ happens for all $k \geq n_0$ w.p.1. Thus we conclude that w.p.1, $X_{n_k} \to X$ as $k \to \infty$. □

Later we apply this to the SLLN. The converse of the Borel-Cantelli Lemma fails without additional assumptions.

Example 19. Take $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}$, and $\mathbb{P}$ is the Lebesgue measure. Consider a sequence $a_n \downarrow 0$ such that $a_n \geq 1/n$ (for example, $a_n = 1/n$). Now let $A_n = (0, a_n)$. Then $\mathbb{P}[A_n] = a_n$ and $\sum_{n=1}^{\infty} \mathbb{P}[A_n] \geq \sum_{n=1}^{\infty} \frac{1}{n} = \infty$. We then calculate

$$\mathbb{P}[A^\infty] : B_n = \bigcup_{k=n}^{\infty} A_k = \bigcup_{k=n}^{\infty} (0, a_k) = (0, a_n),$$

because $(0, a_k)$ is decreasing: $(0, a_k) \supset (0, a_{k+1}) \supset \cdots$. Thus

$$A^\infty = \bigcap_{n=1}^{\infty} (0, a_n) = \emptyset \implies \mathbb{P}[A^\infty] = 0.$$

This gives us an example of $\mathbb{P}[A^\infty] = 0$ but $\sum_{n=1}^{\infty} \mathbb{P}[A_n] = \infty$. Why is this true? This is because $A_n \supset A_{n+1}$, or in other words, $\mathbb{P}[A_n \mid A_{n+1}] = 1$. The events are dependent.

Proposition 48 (Borel-Cantelli Lemma, Part 2)**.** If the events A_n are independent, then $\sum_{n=1}^{\infty} \mathbb{P}[A_n] = \infty$ implies $\mathbb{P}[A_n \text{ i.o.}] = \mathbb{P}[A^\infty] = 1$.
(Note: As it turns out, it is sufficient for A_n to be pairwise independent.)

Proof.

- To show $\mathbb{P}[A_n \text{ i.o.}] = \mathbb{P}[\bigcap_{n=1} B_n] = 1$ is equivalent to showing that $\mathbb{P}[\bigcup_{n=1} B_n^c] = 1 - \mathbb{P}[\bigcap_{n=1} B_n] = 0$. But $0 \le \mathbb{P}[\bigcup_{n=1} B_n^c] \le \sum_{n=1}^{\infty} \mathbb{P}[B_n^c]$. If we show that $\mathbb{P}[B_n^c] = 0$ for all n, then $\sum_{n=1}^{\infty} \mathbb{P}[B_n^c] = 0$ and $\mathbb{P}[\bigcup_{n=1} B_n^c] = 0$ iff $\mathbb{P}[\{A_n \text{ i.o.}\}] = 1$.

- To conclude, we must show that $\mathbb{P}[B_N^c] = 0$ and $\mathbb{P}[\bigcup_{n=1} B_n^c] = 0$ iff $\mathbb{P}[A_n \text{ i.o.}] = 1$. Take $N > n$. We consider

$$\mathbb{P}\left[\bigcap_{k=n}^{N} A_k^c\right] = \prod_{k=n}^{N} \mathbb{P}[A_k^c] = \prod_{k=n}^{N} (1 - \mathbb{P}[A_k]).$$

□