Stochastic Modelling and Random Processes

Yiming MA

Contents

Chapter 1

Introduction

1.1 Motivation

Suppose we are modelling COVID. Let

- S be he number of the susceptible;
- I be the number of the infected;
- R be the number of the removed (those who have either recovered or died).

1.1.1 A Deterministic Model

A deterministic model might be

$$\begin{aligned}\dot{S} =& -\beta IS,\\ \dot{I} =& \beta IS - \gamma I,\\ \dot{R} =& \gamma I.\end{aligned}$$

But there are some problems in this model:

- S, I and R are integers, so it does not make sense to talk about $\dot{S}$, $\dot{I}$ and $\dot{R}$.
- There is variability in when contacts are made and lead to infection.

1.1.2 A Stochastic Model

A better model might be stochastic

$$\begin{aligned}\mathbb{P}S \to S-1 \;\&\; I \to I-1 \text{ in } \Delta t =& \beta IS\Delta t + o(\Delta t)\\ \mathbb{P}I \to I-1 \;\&\; R \to R+1 \text{ in } \Delta t =& \gamma I\Delta t + o(\Delta t).\end{aligned}$$

The problem of this model is that contacts are usually not made uniformally in the whole population.

1.1.3 A Network Model

We can use a network model, in which nodes represent individuals and edge weights represent contact rates, to avoid uniform contacts. But tis is unrealistic: the network is too big to represent 60 million people in the UK.

1.1.4 A Random Network Model

Based on the network model, we can make probability distributions on networks and derive probabilistic conclusions over the combination of stochastic dynamics and randomness of networks.

Chapter 2

Probability and Random Variables

2.1 Probability Theory

Suppose we are doing an experiment which have different random outcomes.

Definition 2.1.1 (Sample Spaces)**.** The **sample space** of the experiment is the set of all possible outcomes, denoted as Ω.

Definition 2.1.2 (Sigma Algebra)**.** The **σ-algebra** of subsets of Ω, denoted as $\mathcal{F}$, is a set of subsets of Ω which satisfies:

- $\Omega \in \mathcal{F}$;
- $A \in \mathcal{F} \implies A^c \in \mathcal{F}$;
- $\{A_i | i \in \mathcal{I}\} \subset \mathcal{F}$ with $\mathcal{I}$ being countable $\implies \bigcup_{i \in \mathcal{I}} A_i \in \mathcal{F}$.

Remark. We say $\mathcal{I}$ is countable if there exists a one-to-one map from $\mathcal{I}$ into $\mathbb{Z}$, so "countable" includes "finite".

Example 2.1.1. If Ω is countable, we usually take $\mathcal{F} = 2^{\Omega}$, which is the power set of Ω.

Example 2.1.2. When Ω is not countable, e.g. $[0, 1]$, if you allow Axiom of Choice [1], then there exist unmeasurable subsets, and we exclude them from $\mathcal{F}$, i.e. $\mathcal{F}$ is the set of all Lebesgue-measurable subsets on $[0, 1]$.

Definition 2.1.3 (Events)**.** The members of $\mathcal{F}$ are called **events**.

Definition 2.1.4 (Probability)**.** $\mathbb{P}[\cdot] : \mathcal{F} \mapsto \mathbb{R}$ is called a probability if

- $\mathbb{P}[A] \in [0, 1]$, $\forall A \in \mathcal{F}$;
- $\mathbb{P}[\Omega] = 1$;
- $\mathbb{P}[\cdot]$ satisfies the **countable additivity**: $\forall \{A_i | i \in \mathcal{I}\} \subset \mathcal{F}$, where $\mathcal{I}$ is a countable set, if A_i's are disjoint, then

$$\mathbb{P}\left[\bigcup_{i \in \mathcal{I}} A_i\right] = \sum_{i \in \mathcal{I}} \mathbb{P}\left[A_i\right].$$

[1] A Cartesian product of a collection of nonempty sets is nonempty.

Definition 2.1.5 (Independence). Say $A, B \in \mathcal{F}$ are **independent** if

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B].$$

Definition 2.1.6 (Conditional Probabilities). If $\mathbb{P}[B] > 0$, then the **conditional probability** $\mathbb{P}\left[A|B\right]$ is defined by

$$\mathbb{P}\left[A|B\right] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}, \quad \forall A \in \mathcal{F}.$$

Definition 2.1.7 (Partitions). $\{B_i | i \in \mathcal{I}\}$ is called a **partition** of the sample space Ω if:

- B_i's are **pairwise disjoint**: $B_i \cap B_j = \emptyset$, $\forall i, j \in \mathcal{I}$, $i \neq j$;
- $B_i \neq \emptyset$, $\forall i \in \mathcal{I}$;
- $\{B_i | i \in \mathcal{I}\}$ **covers** Ω: $\bigcup_{i \in \mathcal{I}} B_i = \Omega$.

Theorem 2.1.1 (The Law of Total Probability). *Let $\{B_i | i \in \mathcal{I}\}$ be a countable partition of Ω, with $B_i \in \mathcal{F}$ and $\mathbb{P}\left[B_i\right] > 0$, $\forall i \in \mathcal{I}$. Then $\forall A \subset \mathcal{F}$, we have*

$$\mathbb{P}\left[A\right] = \sum_{i \in \mathcal{I}} \mathbb{P}\left[A|B_i\right] \mathbb{P}\left[B_i\right].$$

Theorem 2.1.2 (Bayes' Rule). *For any events A and B, if $\mathbb{P}[A] > 0$ and $\mathbb{P}[B] > 0$, then*

$$\mathbb{P}\left[B|A\right] = \frac{\mathbb{P}\left[A|B\right] \mathbb{P}[B]}{\mathbb{P}[A]}.$$

Furthermore, if $\{B_i | i \in \mathcal{I}\}$ is a countable partition of Ω, with $B_i \in \mathcal{F}$ and $\mathbb{P}\left[B_i\right] > 0$, $\forall i \in \mathcal{I}$, then

$$\mathbb{P}\left[B_i|A\right] = \frac{\mathbb{P}\left[A|B_i\right] \mathbb{P}[B_i]}{\sum_{i \in \mathcal{I}} \mathbb{P}\left[A|B_i\right] \mathbb{P}\left[B_i\right]}.$$

Example 2.1.3. Suppose the true positive rate $\mathbb{P}\left(\text{tests positive}|\text{has COVID}\right)$ is 0.99 and the false positive rate $\mathbb{P}\left(\text{tests positive}|\text{does not have COVID}\right)$ is 0.01. Suppose in the population, the probability of getting contracted with COVID is 0.001, i.e. $\mathbb{P}\left(\text{has COVID}\right) = 0.001$, what is the probability that a person has COVID given his/her test is positive?

Sol.

$$\begin{aligned}
&\mathbb{P}\left(\text{has COVID}|\text{tests positive}\right) \\
=\ & \frac{\mathbb{P}\left(\text{tests positive}|\text{has COVID}\right)\mathbb{P}\left(\text{has COVID}\right)}{\mathbb{P}(\text{tests positive})} \\
=\ & \mathbb{P}\left(\text{tests positive}|\text{has COVID}\right)\mathbb{P}\left(\text{has COVID}\right) \\
& \div \Big[\mathbb{P}\left(\text{tests positive}|\text{has COVID}\right)\mathbb{P}\left(\text{has COVID}\right) \\
& +\mathbb{P}\left(\text{tests positive}|\text{does not have COVID}\right)\mathbb{P}\left(\text{does not have COVID}\right)\Big] \\
=\ & \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.01 \times (1 - 0.001)} \\
\approx\ & 0.090.
\end{aligned}$$

2.2 Random Variables

Definition 2.2.1 (Measurable Functions)**.** Let $(\Omega, \mathcal{F}), \mathbb{P})$ and $(\mathbb{R}, \Sigma, \mathcal{L})$ be two measurable spaces, where $\mathcal{L}$ is the Lebesgue measure. For any function $f : \Omega \mapsto \mathbb{R}$, if it satisfies $\forall A \in \Sigma$, $f^{-1}(A) \in \mathcal{F}$, then f is said to be **measurable**.

Remark. $(\Omega, \mathcal{F}, \mathbb{P})$ and $(\mathbb{R}, \Sigma, \mathcal{L})$ can be generalized:

Let (X, Σ) and (Y, T) be measurable spaces, meaning that X and Y are sets equipped with respective σ-algebras Σ and T. A function $f : X \mapsto Y$ is said to be **measurable** if for every $E \in T$ the pre-image of E under f is in Σ; i.e. $\forall E \in T$,

$$f^{-1}(E) := \{x \in X | f(x) \in E\} \in \Sigma.$$

Definition 2.2.2 (Random Variables)**.** A **random variable** is a measurable function $X : \Omega \mapsto \mathbb{R}$.

Definition 2.2.3 (Cumulative Distribution Functions)**.** The **cumulative distribution function** of a random variable X is defined as

$$F(x) = \mathbb{P}\left[X \leq x\right]$$

Definition 2.2.4 (Discrete Random Variables)**.** If $X(\Omega)$ is countable, then X is called **discrete**.

Definition 2.2.5 (Probability Mass Functions)**.** The **probability mass function** of a discrete random variable X is defined as

$$\pi(x) = \mathbb{P}\left[X = x\right], \quad \forall x \in X(\Omega).$$

Definition 2.2.6 (Continuous Random Variables & Probability Density Functions)**.** For a random variable X, if its cumulative distribution function satisfies

$$F(x) = \int_{-\infty}^{x} f(y)\,\mathrm{d}y$$

for some $f \in \mathcal{L}^1(\mathbb{R})$, then X is said to be **continuous**, and f is its **probability density function**.

Remark. It is possible to have mixtures. For example, X can have a positive probability on a particular point and continuous parts on other points.

Definition 2.2.7 (Expectation)**.** The **expectation** of a random variable X is

$$\mathbb{E}[X] := \int_{\Omega} X\,\mathrm{d}\mathbb{P} = \begin{cases} \sum_{x \in X(\Omega)} x\pi(x) & X \text{ is discrete} \\ \int_{X(\Omega)} xf(x)\,\mathrm{d}x & X \text{ is continuous} \end{cases}$$

Remark. The expectation may be infinite or even undefined.

Definition 2.2.8 (Variance)**.** The **variance** of a random variable X is

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

Definition 2.2.9 (Covariance)**.** The **covariance** of two random variables X and Y is

$$\mathrm{Cov}[X, Y] = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Definition 2.2.10 (Uncorrelated Random Variables)**.** If $\mathrm{Cov}[X, Y] = 0$, then X and Y are called uncorrelated.

Proposition 2.2.1. *If X and Y are two independent random variables, then they are also uncorrelated. But the opposite is generally not true, except for Gaussians.*

We can extend to random viarbles taking values in $\mathbb{R}^n$.

- For cumulative distribution functions, use the component-wise $\leq$ instead.
- For $\mathrm{Var}[X]$, use $\mathbb{E}[(X - \mathbb{E}[X])(X - \mathbb{E}[X])^T]$ which is a $n \times n$ matrix.
- For $\mathrm{Cov}[X, Y]$, use $\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])^T]$.
- X and Y are independent if events $\{X \leq x\}$, $\{Y \leq y\}$ are independent, $\forall x, y$.
  - For X, Y being discrete, this is equivalent to $\pi(x, y) = \pi^X(x)\pi^Y(y)$.
  - For X, Y being continuous, this is equivalent to $f(x, y) = f^X(x)f^Y(y)$.

Theorem 2.2.1 (The Weak Law of Large Numbers)**.** *Let X_k, $k = 1, 2, \cdots, X_n, \cdots$ be independent and identically distributed random variables with $\mu = \mathbb{E}[X_k] < \infty$, then*

$$\bar{X}_n := \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{d} \mu, \text{ as } n \to \infty,$$

*where $\xrightarrow{d}$ means **convergence in distribution** [2]. This means the CDF of $\bar{X}_n$ converges to the CDF of μ.*

Equivalently,

$$\mathbb{E}\left[g\left(\bar{X}_n\right)\right] \to g(\mu), \text{ as } n \to \infty,$$

*for any bounded and continuous function g. This type of convergence is called the **weak onvergence**.*

*Or $\bar{X}_n$ converges to μ **in probability** ($\bar{X}_n \xrightarrow{\mathbb{P}} \mu$):*

$$\mathbb{P}\left[\left|\bar{X}_n - \mu\right| > \epsilon\right] \to 0, \text{ as } n \to \infty, \ \forall \epsilon > 0.$$

Theorem 2.2.2 (The Strong Law of Large Numbers)**.** *Let X_k, $k = 1, 2, \cdots, X_n, \cdots$ be independent and identically distributed random variables with $\mu = \mathbb{E}[X_k] < \infty$, then*

$$\bar{X}_n \xrightarrow{a.s.} \mu,$$

*where the **almost surely convergence** means*

$$\mathbb{P}\left[\lim_{n\to\infty} \bar{X}_n = \mu\right] = 1.$$

Theorem 2.2.3 (Central Limit Theorem)**.** *Let X_k, $k = 1, 2, \cdots, X_n, \cdots$ be independent and identically distributed random variables with $\mu = \mathbb{E}[X_k] < \infty$ and $0 < \sigma^2 := \mathrm{Var}[X_k] < \infty$, then*

$$\frac{\sqrt{n}}{\sigma}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}(0, 1).$$

[2] Also called **convergence in law**.

Theorem 2.2.4 (Large Deviation Principle). *Let X_k, $k = 1, 2, \cdots, X_n, \cdots$ be independent and identically distributed random variables. For any interval $J \subset \mathbb{R}$,*

$$\mathbb{P}\left[\bar{X}_n \in J\right] \approx \exp\left(-n \min_{x \in J} I(x)\right),$$

meaning

$$\frac{1}{n} \log \mathbb{P}\left[\bar{X}_n \in J\right] \to -\min_{x \in J} I(x).$$

If we know the probability distribution of X_k, an explicit expression for the rate function can be obtained. This is given by a Legendre–Fenchel transformation,

$$I(x) = \sup_{\theta > 0} \theta x - \lambda(\theta),$$

*where $\lambda(\theta) = \log \mathbb{E}\left[e^{\theta X_k}\right]$ is called the **cumulant generating function (CGF)**.*

Definition 2.2.11 (Stochastic Processes). A **stochastic process** $\{X(t) | t \in T\}$ is a collection of random variables. That is, for each $t \in T$, $X(t)$ is a random variable.

- The index t is often interpreted as time and, as a result, we refer to $X(t)$ as the **state** of the process at time t.
- The set T is called the **index set** of the process.
  - When T is a countable set, the process is said to be a **discrete-time** process.
  - If T is an inverval of the real line, the stochastic process is said to be a **continuous-time** process.
- The **state space** of a stochastic process is defined as the set of all possible values that the random variables $X(t)$ can assume.

Chapter 3

Discrete-Time Markov Chain

3.1 Discrete-Time Markov Chains

Definition 3.1.1 (Discrete-Time Stocastic Processes)**.** A **discrete-time stochastic process** with state space S is a sequence $\{Y_n|n \in \mathbb{N}\}$ of random variables taking values in S.

Definition 3.1.2 (Discrete-Time Markov Chains)**.** Let $\{X_n|n \in \mathbb{N}\}$ be a discrete-time stochastic process with a discrete state space S. The process is called a **Markov chain**, if for all $A \subset S$, $n \in \mathbb{N}$ and $s_0, \cdots, s_n \in S$,

$$\mathbb{P}\left[X_{n+1} \in A | X_n = s_n, \cdots X_0 = s_0\right] = \mathbb{P}\left[X_{n+1} \in A | X_n = s_n\right].$$

Proposition 3.1.1. *For any Markov chain $\{X_n|n \in \mathbb{N}\}$, conditional on the present, the past and the future are independent, i.e. $\forall n \in \mathbb{N}_+$, $\forall s_n \in S$, $X_{n+1}|X_n = s$ and $X_{n-1}|X_n = s$ are independent.*

Proof.

$$\begin{aligned}
&\mathbb{P}\left[X_{n+1} = s_{n+1}, X_{n-1} = s_{n-1} | X_n = s_n\right] \\
&= \frac{\mathbb{P}\left[X_{n-1} = s_{n-1}, X_n = s_n, X_{n+1} = s_{n+1}\right]}{\mathbb{P}\left[X_n = s_n\right]} \\
&= \mathbb{P}\left[X_{n-1} = s_{n-1}\right] \cdot \mathbb{P}\left[X_n = s_n | X_{n-1} = s_{n-1}\right] \cdot \mathbb{P}\left[X_{n+1} = s_{n+1} | X_n = s_n, X_{n-1} = s_{n-1}\right] \cdot \frac{1}{\mathbb{P}\left[X_n = s_n\right]} \\
&= \mathbb{P}\left[X_{n-1} = s_{n-1}\right] \cdot \mathbb{P}\left[X_n = s_n | X_{n-1} = s_{n-1}\right] \cdot \mathbb{P}\left[X_{n+1} = s_{n+1} | X_n = s_n\right] \cdot \frac{1}{\mathbb{P}\left[X_n = s_n\right]} \\
&= \mathbb{P}\left[X_{n-1} = s_{n-1}\right] \cdot \frac{\mathbb{P}\left[X_{n-1} = s_{n-1} | X_n = s_n\right] \cdot \mathbb{P}\left[X_n = s_n\right]}{\mathbb{P}\left[X_{n-1} = s_{n-1}\right]} \cdot \mathbb{P}\left[X_{n+1} = s_{n+1} | X_n = s_n\right] \cdot \frac{1}{\mathbb{P}\left[X_n = s_n\right]} \\
&= \mathbb{P}\left[X_{n-1} = s_{n-1} | X_n = s_n\right] \cdot \mathbb{P}\left[X_{n+1} = s_{n+1} | X_n = s_n\right]
\end{aligned}$$

□

3.1.1 Homogeneity

Definition 3.1.3 (Homogeneity)**.** A Markov chain $\{X_n|n \in \mathbb{N}\}$ is **homogeneous** if for all $A \subset S$, $n \in \mathbb{N}$ and $s \in S$,

$$\mathbb{P}\left[X_{n+1} \in A | X_n = s\right] = \mathbb{P}\left[X_1 \in A | X_0 = s\right].$$

Example 3.1.1 (Random Walk with Boundaries)**.** Let $\{X_n|x \in \mathbb{N}\}$ be a **simple random walk** on $S = \{1, \cdots L\}$ with $p(x, y) = p\delta_{y,x+1} + q\delta_{y,x-1}$. The boundary conditions are

- **periodic** if $p(L,1)=p$, $p(1,L)=q$,
- **absorbing** if $p(L,L)=1$, $p(1,1)=1$,
- **closed** if $p(L,L)=p$, $p(1,1)=q$,
- **reflecting** if $p(L,L-1)=1$, $p(1,2)=1$.

3.1.2 Transition Matrices and Transition Functions

Definition 3.1.4 (One-Step Transition Matrices)**.** For a homogeneous discrete-time Markov chain $\{X_n|n\in\mathbb{N}\}$ taking values in $\{s_1,s_2,s_3,\cdots,s_n,\cdots\}$, its **one-step transition matrix** P is defined as

$$P_{i,j}=\mathbb{P}\left[X_{n+1}=s_j|X_n=s_i\right].$$

Remark. The sum of each row of a one-step transition matrix is 1, i.e.

$$P\left|1\right\rangle=\left|1\right\rangle.$$

Proposition 3.1.2. *Let $\pi_0(\cdot)$ be the probability mass function of X_0, then*

$$\mathbb{P}\left[X_0=s_0,X_1=s_1,\cdots,X_n=s_n\right]=\pi_0(s_0)P_{s_0,s_1}\cdots P_{s_{n-1},s_n}.$$

If we use a row vector $\langle\boldsymbol{\pi}_0|$ to represent the probability distribution of X_0, such that $\langle\boldsymbol{\pi}_0|_i=\mathbb{P}\left[X_0=s_i\right]$, then the probability distribution of X_n can be represented as

$$\langle\boldsymbol{\pi}_n|=\langle\boldsymbol{\pi}_0|P^n.$$

Definition 3.1.5 (Transition Functions)**.** The transition matrix of $\{X_n|n\in\mathbb{N}\}$ can be written into the **transition function** $p_n(x,y)$ instead:

$$p_n(x,y):=\mathbb{P}\left[X_n=y|X_0=x\right].$$

3.1.3 Chapman-Kolmogorov Equations

Theorem 3.1.1 (Chapman-Kolmogorov Equations)**.** *For a homegeneous discrete-time Markov chain $\{X_n|n\in\mathbb{N}\}$, its transition function fulfills the **Chapman-Kolmogorov equations***

$$p_{k+n}(x,y)=\sum_{z\in S}p_k(x,z)p_n(z,y)\quad \text{for all } k,n\geq 0,\ x,y\in S.$$

Remark. In matrix form, the Chapman-Kolmogorov equations read

$$P_{n+k}=P_nP_k\quad\text{and in particular}\quad P_{n+1}=P_nP_1.$$

Corollary 3.1.1. *Let P_n be the n-step transition matrix of a homogeneous discrete-time Markov chain $\{X_n|n\in\mathbb{N}\}$, then*

$$P_n=P^n\quad\&\quad P_0=I.$$

3.1.4 Stationary Distributions

Definition 3.1.6 (Stationarity)**.** Let $\{X_n | n \in \mathbb{N}\}$ be a homogeneous discrete-time Markov chain with state space S. The distribution $\pi(x)$, $x \in S$, is called **stationary** if for all $y \in S$

$$\sum_{x \in S} \pi(x) p(x, y) = \pi(y),$$

or

$$\langle \boldsymbol{\pi} | \, P = \langle \boldsymbol{\pi} | \, .$$

Remark. If $\boldsymbol{\pi}$ is a stationary distribution, then it is a left eigenvector with eigenvalue 1.

Remark. To solve the stationary distributions, we can solve

$$\begin{cases} \langle \boldsymbol{\pi} | \, P & = \langle \boldsymbol{\pi} | \\ \langle \boldsymbol{\pi} | \mathbf{1} \rangle & = 1 \end{cases}$$

Theorem 3.1.2. *Every homogeneous finite discrete-time Markov chain has a stationary distribution.*

Proof. Let

$$\Delta = \left\{ \langle \boldsymbol{\pi} | \, \middle| \, \boldsymbol{\pi}_i \geq 0, \langle \boldsymbol{\pi} | \mathbf{1} \rangle = 1 \right\}$$

Then $P_{i,j} \geq 0$ and $P \, |\mathbf{1}\rangle = |\mathbf{1}\rangle$, so $\boldsymbol{\pi} \in \Delta \implies \langle \boldsymbol{\pi} | \, P \in \Delta$. Notice that Δ is compact and convex, and P is continuous (linear), so by the Brouwer's Fixed-Point Theorem, P has a fixed point $\langle \boldsymbol{\pi}^* | \in \Delta$, such that $\langle \boldsymbol{\pi}^* | \, P = \langle \boldsymbol{\pi}^* |$. □

Remark. There can be more than one stationary distributions. For example, if a Markov chian has two parts with no transitions between them, then let $\langle \boldsymbol{\pi}_1 |$ and $\langle \boldsymbol{\pi}_2 |$ be stationary probabilities of them, and any convex combination of $\langle \boldsymbol{\pi}_1 |$ and $\langle \boldsymbol{\pi}_2 |$ is a stationary distribution.

Definition 3.1.7 (Cycles)**.** A **cycle** is a closed path in S along the graph of allowed transitions by P, and its length is greater than 0.

Definition 3.1.8 (Transience & Recurrence)**.** Say $i \in S$ is **transient** if there does not exists any cycle through i. Otherwise, i is **recurrent**.

Definition 3.1.9 (Communication)**.** Say $i, j \in S$ communicate with each other if there exist a cycle through i and j, denoted as $i \leftrightsquigarrow j$.

Proposition 3.1.3. *Communication is an equivalent relation on the set of all recurrent states:*

- $i \leftrightsquigarrow i$, $\forall i \in S$;
- $i \leftrightsquigarrow j \iff j \leftrightsquigarrow i$, $\forall i, j \in S$;
- $i \leftrightsquigarrow j \leftrightsquigarrow k \implies i \leftrightsquigarrow k$, $\forall i, j, k \in S$.

Definition 3.1.10 (Classes, Comminucating Components)**.** A **class** (also called a **comminucating component**) is a set of all comminucating states in the state space.

Definition 3.1.11 (Absorbing Communicating Components)**.** A communicating component is **absorbing** or **closed** if it is impossible to leave it.

Remark. The transition graph for a Markov chain can be quotiented using the communicating relation "$\leftrightsquigarrow$", and the resulting graph is acyclic (a DAG), in which the basal communicating components are the absorbing ones.

Proposition 3.1.4. *Any finite discrete-time Markov chain has at least an absorbing communicating class.*

We can restrict a Markov chain to an absorbing component, and the result of the restriction is called irreducible.

Definition 3.1.12 (Irreducibility)**.** A Markov chain is called **irreducible**, if its states belong to one comminucating component.

Now, here is another proof of Theorem 3.1.2: For $x \in A$, where A is an absorbing component of the Markov chain, let μ_x be the mean time to return to x given that the chain starts in x. For an arbitrary state $y \in S$, let $\gamma_x(y)$ be the mean time in y before returning to x given the chain starts in x. Then $\gamma_x(y) \geq 0$ and $\langle \gamma_x | P = \langle \gamma_x |$. Normalize $\langle \gamma_x |$ by letting $\pi_x(y) = \gamma_x(y)/T_x$, then

$$\langle \boldsymbol{\pi}_x | P = \langle \boldsymbol{\pi}_x | \quad \text{and} \quad \langle \boldsymbol{\pi}_x | \mathbf{1} \rangle = 1.$$

So $\boldsymbol{\pi}_x$ is a stationary probability.

Remark. $\boldsymbol{\pi}_x = \boldsymbol{\pi}_y$ if and only if x, y are in the same absorbing component A, so denote it by $\boldsymbol{\pi}_A$.

Remark.

$$\boldsymbol{\pi}_A(x) = \begin{cases} \frac{1}{T_x} & x \in A \\ 0 & x \notin A \end{cases}$$

Remark. $\boldsymbol{\pi}_A$ is dynamic-system ergodic.

Definition 3.1.13 (Dynamic-System Ergodic)**.** A stationary distribution is **dynamic-system ergodic** if it is not a convex combination of other stationary probabilities.

Theorem 3.1.3 (Dynamic-System Ergodic Theorem)**.** *If $\boldsymbol{\pi}$ is DS-ergodic, then $\forall x \in S$ with $\pi(x) > 0$, then the fraction of times $0, \cdots, T-1$ spent in any $y \in X$ given that the process starts in x, converges almost surely to $\pi(y)$, as $T \to \infty$.*

Remark. This provides a great interpretation of $\boldsymbol{\pi}_A$.

Proposition 3.1.5. *If a finite discrete-time Markov chain has a unique absorbing component A, then the fraction of time spent in y given that the process starts anywhere converges almost surely to $\boldsymbol{\pi}_A(y)$, and $\boldsymbol{\pi}_A$ is the only stationary probability.*

Remark. DS-theorists would say such a Markov chain is **uniquely ergodic**.

Remark. Suppose there are more than one absorbing components, say A_j's, then there are "commutors probabilities" C_{z,A_j}, for which absorbing component A_j, you eventually land in if the process starts in $z \in X$. Then the fraction of time spent in y converges to $\boldsymbol{\pi}_{A_j}(y)$ with probability C_{z,A_j}, as $T \to \infty$.

Definition 3.1.14 (Stationary Markov Chains)**.** Say a homogeneous Markov chain is **stationary** if $\pi(x) := \mathbb{P}\left[X(0) = x\right]$ is a stationary probability.

3.1.5 Reverse Markov Chains

The Markov property also implies

$$\mathbb{P}\left[X_m = s_m, \cdots, X_n = s_n\right] = \mathbb{P}\left[X_m = s_m\right] \tilde{P}_{s_n, s_n - 1} \cdots \tilde{P}_{s_{m+1}, s_m}$$

with

$$\tilde{P}_{i,j} := \mathbb{P}\left[X_{k-1} = j | X_k = i\right] = \frac{\mathbb{P}\left[X_k = i, X_{k-1} = j\right]}{\mathbb{P}\left[X_k = i\right]} = \frac{\mathbb{P}\left[X_{k-1} = j\right]}{\mathbb{P}\left[X_k = i\right]} P_{j,i}.$$

But in general the reverse chain is not homogeneous even if P is, but if the Markov chain is stationary with stationary probability $\boldsymbol{\pi}$ then

$$\tilde{P}_{i,j} = \frac{\boldsymbol{\pi}_j}{\boldsymbol{\pi}_i} P_{k,i},$$

so the reverse chain is also homogeneous; furthermore, it is stationary with the same $\boldsymbol{\pi}$.

Definition 3.1.15 (Reversible Markov Chains)**.** Say a stationary Markov chain is **reversible**, if

$$\boldsymbol{\pi}_i P_{i,j} = \boldsymbol{\pi}_j P_{j,i}, \ \forall i, j \in S$$

The above equation is referred as the condition of **detailed balance**.

Remark. If a Markov chain is reversible, then $P = \tilde{P}$.

Proposition 3.1.6. *Note that if $\boldsymbol{\pi} \in \Delta$ satisfying detailed balance for P then $\boldsymbol{\pi}$ is stationary for P and the Markov chain with the initial probability $\boldsymbol{\pi}$ is reversible.*

Definition 3.1.16 (Stochastic-Process Ergodicity)**.** Say a transition matrix P is **stochastic -process ergodic** if $\exists \boldsymbol{\pi} \in \Delta$ such that

$$\forall \boldsymbol{\pi} \in \Delta, \ \lim_{n \to \infty} \boldsymbol{\pi} P^n = \boldsymbol{\pi}.$$

(All senses of convergence are equivalent in a finite state space.)

Remark. This notion tells us not only that fraction of time in state y converges almost surely to $\pi(y)$ for any initial condition, but also that the probability of $X_n = y$ converges to $\pi(y)$, as $n \to \infty$.

Remark. There are finite discrete-time Markov chain with a unique absorbing component (so DS ergodic) which are not SP-ergodic.

The obstruction is periodicity: let $T_A = \text{hcf}\{\text{lengths of all cycles in } A\}$, which is called the **period** of A. If $T_A > 1$ then A decomposed into "cyclic classes" $A_1, \cdots, A_T$.

Theorem 3.1.4 (Stochastic-Process Ergodicity)**.** *A homogeneous finite discrete-time Markov chian is SP ergodic if it has a unique absorbing component and it is aperiodic ($T_A = 1$).*

Proof. The proof is by the Perron-Frobenius theorem: for an irreducible, aperiodic, nonnegative matrix, there is a unique and simple eigenvalue of maximum modulus and it has a possitive eigenvector.

In our case, the largest modulus of eigenvalue P can have is 1 (conservation of probability or the Gershgorin Circile Theorem) and we konw it has an eigenvalue 1 ($P|\mathbf{1}\rangle = |\mathbf{1}\rangle$). We can restrict to the unique absorbing component so 1 is simple and has a postive left eigenvector $\langle\boldsymbol{\pi}|$ which we normalize to $\langle\boldsymbol{\pi}|\,\mathbf{1}\rangle = 1$. Then let $\boldsymbol{\pi}_0$ be the initial probability and decompose P into its eigenvectors and generalized eigenvectors,

$$\boldsymbol{\pi}_0 P^n = a\boldsymbol{\pi}_0 + \text{terms corresponding to eigenvalues within the unit disc.}$$

Since terms corresponding to eigenvalues within the unit disc go to 0 exponentially as $n \to \infty$ and $\langle\boldsymbol{\pi}_0|\mathbf{1}\rangle = 1$, $P|\mathbf{1}\rangle = |\mathbf{1}\rangle$, we have $a = 1$. Hence the Markov chain is SP ergodic. □

3.1.6 Monte Carlo Markov Chains

Given a probability distribution $\boldsymbol{\pi}$ on a state space S and a random variable $X : S \mapsto \mathbb{R}$, we may want to compute the mean of X. Or, we might be given an unnormalized probability $\tilde{\boldsymbol{\pi}}$ and want $\mathbb{E}[X]$ or $Z = \langle\tilde{\boldsymbol{\pi}}|\mathbf{1}\rangle$ which is the normalization constan of $\tilde{\boldsymbol{\pi}}$.

Such problems arise in statistical mechanics where the probability of being in state x is proportional to $\exp(-\beta H(x))$, where H refers to the "energy" and $\beta = 1/(k_B T)$ refers to the "coolness". Then

$$Z = \sum_{x \in S} e^{-\beta H(x)}.$$

and

$$\mathbb{E}[X] = \frac{1}{Z}\sum_{x\in S} H(x)e^{-\beta H(x)}.$$

They also arise in statistical inference for Bayesian model comparison. Let M denote the model and μ some constant, then Bayesian inference gives

$$\mathbb{P}\left[\mu|\text{data}, M\right] = \frac{\mathbb{P}\left[\text{data}|M,\mu\right]\mathbb{P}\left[\mu\right]}{Z(M)}.$$

To solve such problems, we can design a Markov chain with unique absorbing component on which $\tilde{\boldsymbol{\pi}}$ is stationary. Then the fraction of time spent in state x by a typical realization converges to $\pi(x)$, as $T\to\infty$, and the time-average of $\mathbb{E}\left[X_n\right]\to Z$.

The easiest way to achieve $\tilde{\boldsymbol{\pi}}$ stationary is to choose P so that

$$\tilde{\boldsymbol{\pi}}_i P_{i,j} = \tilde{\boldsymbol{\pi}}_j P_{j,i}, \forall i,j\in S,$$

and $P\left|\mathbf{1}\right\rangle = \left|\mathbf{1}\right\rangle$ and $P_{i,j}\geq 0$. We can do this by taking any proposal transition probabilities $Q_{i,j}$ and use accpetance probabilities $A_{i,j}$ like

$$\text{Metropolis-Hastings}\quad A_{i,j} = \begin{cases} 1 & \tilde{\boldsymbol{\pi}}_j Q_{j,i} \geq \tilde{\boldsymbol{\pi}}_i Q_{i,j} \\ \frac{\tilde{\boldsymbol{\pi}}_j P_{j,i}}{\tilde{\boldsymbol{\pi}}_i P_{i,j}} & \text{Otherwise}\end{cases}$$

and

$$\text{Heatbath}\quad A_{i,j} = \frac{\tilde{\boldsymbol{\pi}}_j Q_{j,i}}{\tilde{\boldsymbol{\pi}}_i Q_{i,j} + \tilde{\boldsymbol{\pi}}_j Q_{j,i}},$$

and set

$$P_{i,j} = \begin{cases} Q_{i,j}A_{i,j}, & i\neq j \\ 1-\sum_{k\neq i} Q_{i,k}A_{i,k} & i=j\end{cases}$$

We might as well take $Q_{i,i}=0, \forall i\in S$. We require $Q_{i,i}$ have a unique absorbing component. Then with the above choices, P has detailed balance for $\tilde{\boldsymbol{\pi}}$:

$$\frac{\tilde{\boldsymbol{\pi}}_i Q_{i,j}A_{i,j}}{\tilde{\boldsymbol{\pi}}_j Q_{j,i}A_{j,i}} = 1.$$

We can do this without rejection: given i, let $w_{i,j} = Q_{i,j}A_{i,j}$, $W_i = \sum_{j\in S} w_{i,j}$ and $P_{i,j} = \frac{w_{i,j}}{W_i}$ and weight time spent in i by W_i.

Definition 3.1.17 (Mixing Time)**. Mixing time** is defined as

$$T(\epsilon) := \min\{T\in\mathbb{Z}_+ | d(\sigma P^n, n)\leq\epsilon, \forall n\geq T, \forall\sigma\in\Delta\},$$

using the total variation distance $d(x,y) := \frac{1}{2}\|x-y\|_1$.

Definition 3.1.18 (Zero Charge)**.** Define the **zero charge norm** of a matrix P as

$$\|P\|_Z = \sup_{\substack{v\neq\mathbf{0}\\ v\mathbf{1}=0}} \frac{\|vP\|_1}{\|v\|_1},$$

where Z is called **zero charge**.

Remark. If $\|P\|_Z<1$, then P is a contraction on Δ, so we can get the SP ergodicity. In particular, $\|P^n\|_Z\leq\|P\|_Z^n\to 0$ exponentially.

More generally, if $\|P^n\|_Z\leq Cr^n$ with $r\in(0,1)$, then

$$T(\epsilon)\leq\frac{|\log\epsilon/C|}{|\log r|}$$

Remark. One can get such bounds on $\|P\|^n$ from Dobrushion's ergodicity coefficient:

$$\|P\|_Z = 1 - \min_{i,j} \sum_k \min\{P_{i,k}, P_{j,k}\} = \frac{1}{2} \max_{i,j} \sum_k |P_{i,k} - P_{j,k}|.$$

But what we really want is to know $d(\mu_T, \pi)$, where

$$\mu_T = \frac{1}{T} \sum_{n=0}^{T-1} \delta_{X_n}.$$

Typically, $d(\mu_T, \pi)$ is $O(T^{-1/2})$ by the central limit theorem generalization.

We can reformulate the problem into finding $\min T$, such that $\mathbb{P}\left[d(\mu_T, \pi) \geq \epsilon\right] < \eta$. Large deviation theorem implies:

$$\mathbb{P} \leq C \exp\left(-\frac{T\epsilon^2}{K + \frac{1}{2}}\right),$$

where $K = \|(I - P)^{-1}\|_Z$. So $T \sim (K + \frac{1}{2})\epsilon^{-2} \log \frac{C}{\eta}$.

3.2 Countable Markov Chains

One can extend much of what we have done for finite discrete-time Markov chains to the countably infinite case, e.g. simple random walk on $\mathbb{Z}$, but some results become more subtle. For example, SRW is not SP ergodic, despite being irreducible. (Actually, it even fails to have a stationary probability.) Also, it is not periodic; it has a period 2. So, one has to refine various concepts.

Definition 3.2.1 (Transience & Recurrence)**.** Let $T_x := \inf\{t \geq 1 : X_n = x\}$. Then the state $x \in S$ is

- **transient**, if
$$\mathbb{P}\left[T_x = \infty | X_0 = x\right] > 0;$$
- **null recurrence** if
$$\mathbb{P}\left[T_x < \infty | X_0 = x\right] = 1 \quad \& \quad \mathbb{E}\left[T_x | X_0 = x\right] = \infty;$$
- **positive recurrence** if
$$\mathbb{P}\left[T_x < \infty | X_0 = x\right] = 1 \quad \& \quad \mathbb{E}\left[T_x | X_0 = x\right] < \infty.$$

Remark. If $x \in S$ is transient, then with probability 1 that X_n comes back to x only finite many times.

Proposition 3.2.1. *A communicating class is either null recurrent or positive recurrent.*

Theorem 3.2.1. *An absorbing component has a stationary probability if and only if it is positive recurrent. Furthermore, it is unique, and*

$$\pi(x) = \frac{1}{\mathbb{E}[T_x | X_0 = x]}$$

Chapter 4

Continuous-Time Markov Chain

4.1 Continuous-Time Markov Chains

Definition 4.1.1 (Continuous-Time Markov Chains)**.** Suppose $\{X(t) : t \in T\}$ ($T = \mathbb{R}$ or $\mathbb{R}_+$) is a stochastic process with a countable state space S, i.e. $X : T \mapsto S$, satisfying the Markov property:

$$\mathbb{P}\left[X(t_{n+1}) \in A | X(t_n) = s_n, \cdots, X(t_1) = s_1\right] = \mathbb{P}\left[X(t_{n+1}) \in A | X(t_n) = s_n\right],$$

for all $A \subset S$, $n \in \mathbb{N}$, $t_1 < \cdots < t_{n+1} \in [0, \infty)$ and $s_1, \cdots, s_n \in S$.

NOTE: we restrict to $X : \mathbb{R} \mapsto S$ that are piecewise constant and right continuous.

Definition 4.1.2 (Homegeneity)**.** Say a continuous-time Markov chain is **homogeneous** if

$$\mathbb{P}\left[X(t+u) \in A | X(u) = s\right] = \mathbb{P}\left[X(t) \in A | X(0) = s\right],$$

for all $A \subset S$, $u, t \in T$, and $s \in S$.

Definition 4.1.3 (Transition Matrices)**.** The **transition matrix** for a homogeneous continuous-time Markov chain depends on the time t:

$$(P_t)_{i,j} = \mathbb{P}\left[X(t) = j | X(0) = i\right].$$

Theorem 4.1.1 (The Chapman-Kolmogorov Equation)**.** *The transition matrix of a homogeneous continuous-time Markov chain satisfies the* ***Chapman-Koklmogorov equation****:*

$$P_{t+u} = P_t P_u = P_u P_t \quad \& \quad P_0 = I.$$

4.1.1 Rate Matrices

Definition 4.1.4 (Rate Matrices)**.** Suppose P_t is differentiable with respect to t at $t = 0$. Then

$$G := \left.\frac{\mathrm{d}P_t}{\mathrm{d}t}\right|_{t=0}$$

is called the **rate matrix** or the **generator**.

Remark. The elements of G satisfies

- $\sum_j G_{i,j} = 0$, $\forall i \in S$;
- $G_{i,j} \geq 0$, for $j \neq i$;

- $G_{i,i} = -\sum_{j \neq i} G_{i,j}$.

Notice that

$$\frac{P_{t+u} - P_t}{u} = \frac{P_t P_u - P_t}{u} = \frac{P_t(P_u - I)}{u} = P_t \frac{P_u - I}{u}$$

and

$$\frac{P_{t+u} - P_t}{u} = \frac{P_u P_t - P_t}{u} = \frac{(P_u - I)P_t}{u} = \frac{P_u - I}{u} P_t,$$

so

$$\lim_{u \to 0} \frac{P_{t+u} - P_t}{u} = P_t \left(\lim_{u \to 0} \frac{P_u - I}{u} \right) = \left(\lim_{u \to 0} \frac{P_u - I}{u} \right) P_t.$$

Proposition 4.1.1. *P_t evolves according to*

$$\frac{\mathrm{d}P_t}{\mathrm{d}t} = P_t G = G P_t,$$

which can also be written as

$$P_t = \exp(tG) \quad \textit{(defined by power series)}.$$

In particular, a probability

$$\langle \boldsymbol{\pi}_t | = \langle \boldsymbol{\pi}_0 | \, P_t$$

evolves by

$$\frac{\mathrm{d}\boldsymbol{\pi}_t}{\mathrm{d}t} = \langle \boldsymbol{\pi}_0 | \frac{\mathrm{d}P_t}{\mathrm{d}t} = \langle \boldsymbol{\pi}_0 | \, P_t G = \langle \boldsymbol{\pi}_t | \, G.$$

Corollary 4.1.1. *Conservation of probability $\langle \boldsymbol{\pi_t} | \mathbf{1} \rangle = 1$ implies $\langle G | \mathbf{1} = 0$.*

Theorem 4.1.2 (The Master Equation)**.** *The i-th element of $\boldsymbol{\pi}_t$, denoted as $\boldsymbol{\pi}_i$, satisfies the **master equation**:*

$$\begin{aligned} \frac{\mathrm{d}\boldsymbol{\pi}_i}{\mathrm{d}t} &= \left(\langle \boldsymbol{\pi} | \, G \right)_i \\ &= \sum_j \boldsymbol{\pi}_j G_{j,i} \\ &= \sum_{j \neq i} \left(\boldsymbol{\pi}_j G_{j,i} \right) + \boldsymbol{\pi}_i G_{i,i} \\ &= \underbrace{\sum_{j \neq i} \left(\boldsymbol{\pi}_j G_{j,i} \right)}_{\text{“gain”}} - \underbrace{\sum_{j \neq i} \left(\boldsymbol{\pi}_i G_{i,j} \right)}_{\text{“loss”}}. \end{aligned}$$

Remark. Its importance is exaggerated: it doesn't tell you correlations between states at different t for example.

4.1.2 Stationarity & Reversibility

Definition 4.1.5 (Stationarity)**.** Say $\boldsymbol{\pi} \in \Delta$ is **stationary** if

$$\langle \boldsymbol{\pi} | \, G = \langle \mathbf{0} | \, .$$

Definition 4.1.6 (Reversibility)**.** Say $\boldsymbol{\pi} \in \Delta$ is **reversible** if

$$\boldsymbol{\pi}_i G_{i,j} = \boldsymbol{\pi}_j G_{j,i}, \quad \forall i, j \in S.$$

Proposition 4.1.2. *Reversibility $\implies$ Stationarity.*

Proposition 4.1.3. *S is finite $\implies$ the existence of stationary $\boldsymbol{\pi}$.*

There is an analogous decomposition of S into transient and recurrent states, and of the set of recurrent states into communicating components.

We have the same definition of an absorbing component, and we have the following nice result.

Proposition 4.1.4. *For a continous-time Markov chain with a finite state space S, each absorbing component has a unique stationary probability $\boldsymbol{\pi}$.*

And the space of stationary $\boldsymbol{\pi}$ for the whole chain (up to normalization) is the span of those for its absorbing components.

Futhermore, 0 is a semisimple eigenvalue of G.

Theorem 4.1.3. *Suppose the continuous-time Markov chain has a finite state space S, and G has a unique absorbing component A, then the Markov chain is SP ergodic, which means*

$$\boldsymbol{\pi}_t \to \boldsymbol{\pi}_A, \text{ as } t \to \infty.$$

Remark. We didn't have to talk about aperiodicity, but in the dicrete-time case, we have to add the condition of it. This is because aperiodicity is automatically satisfied in continous time.

4.1.3 The Jump Chain

Definition 4.1.7 (Waiting Times)**.** Let $\{X(t) : t \in T\}$ be a continous-time Markov chain and $x \in S$ is a state, then the **waiting time** or the **holding time** is defined as

$$W_x := \inf\{t > 0 : X(t) \neq x | X(0) = x\}.$$

Proposition 4.1.5. *For a homogeneous Markov chain, the waiting time W_x is exponentially distributed with expectation $1/|G_{x,x}|$.*

Proof.

$$\begin{aligned}
\mathbb{P}\left[W_x > t + u | W_x > t\right] =& \mathbb{P}\left[W_x > t + u | X(s) = x,\ \forall s \leq t\right] \\
=& \mathbb{P}\left[W_x > t + u | X(t) = x\right] \quad \text{(by Markov property)} \\
=& \mathbb{P}\left[W_x > u | X(0) = x\right] \quad \text{(by homogeneity)} \\
=& \mathbb{P}\left[W_x > u\right]
\end{aligned}$$

So,

$$\mathbb{P}\left[W_x > t + u\right] = \mathbb{P}\left[W_x > u\right] \mathbb{P}\left[W_x > t\right].$$

Thus, $\exists \gamma \in \mathbb{R}$, s.t.

$$\mathbb{P}\left[W_x > t\right] = e^{-\gamma t}$$

($\mathbb{P}\left[W_x > 0\right] = 1$). Using

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}\left[W_x > t\right]\Big|_{t=0} = G_{x,x}$$

shows $\gamma = -G_{x,x}$. □

Definition 4.1.8 (Jump Times)**.** Define **jump times** as

$$J_{n+1} := \inf\{t > J_n : X(t) \neq X(J_n)\},\ J_0 := 0.$$

Remark. The jump times are an example of "stopping times".

Definition 4.1.9 (Stopping Times)**. Stopping times** $\{J_n : n \in \mathbb{N}\}$ are random variables such that $\{J_n \leq t\}$ is independent of $\{X(s) : s > t | X(s) : s \leq t\}$

Theorem 4.1.4. *Markov chains satisfy the strong Markov property: let T be a stopping time conditional on $X(T) = i$, then $\{X(T+t) : t \geq 0\}$ is Markov and independent of $\{X(s) : s \leq T\}$.*

Definition 4.1.10 (Jump Chains)**.** Let $\{X(t) : t \in \mathbb{R}\}$ be a Markov chain and $\{J_n : n \in \mathbb{N}\}$ be its jump times, then the corresponding **jump chain** is defined as

$$Y_n = X(J_n),\ n \in \mathbb{N}.$$

Proposition 4.1.6. *The jump chain is a discrete-time Markov chain, with the transition probability*

$$P_{i,j} = \begin{cases} 0, & G_{i,i} \neq 0 \ \& \ j = i \\ \frac{G_{i,j}}{|G_{i,i}|} & G_{i,i} \neq 0 \ \& \ j \neq i \\ \delta_{i,j} & G_{i,i} = 0 \end{cases}$$

Remark. We can make sample paths for the continous time Markov chain by making sample paths for the associated jump chain and choosing independent waiting times W_{Y_n} with mean $\frac{1}{|G_{Y_n Y_n}}$ and let $J_n = \sum_{0 \leq k < n} W_{Y_k}$.

4.1.4 Examples of Continuous-Time Markov Chains

Example 4.1.1 (Poisson Processes)**.** The Poisson process $\{X(t) : t \in \mathbb{N}\}$ with rate $\lambda > 0$ is a continuous-time Markov chain with $S = \mathbb{N}$, $X(0) = 0$, and the rate matrix

$$G_{i,j} = \begin{cases} \lambda & j = i+1 \\ -\lambda & j = i \end{cases}$$

It has

$$\mathbb{P}\left[X(t+u) = n+k | X(u) = n\right] = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad \forall n, k \in \mathbb{N}, \forall t, u \in \mathbb{R}_+.$$

Example 4.1.2 (Birth and Death Processes)**.** Suppose $\{X(t) : t \in \mathbb{N}\}$ has birth rates α_i and death rates β_i ($\beta_0 = 0$), $\forall i \in S := \mathbb{N}$. Its rate matrix is

$$G_{i,j} = \begin{cases} \alpha_i & j = i+1 \\ \beta_i & j = i-1 \\ -(\alpha_i + \beta_i) & j = i \end{cases}$$

Example 4.1.3 (The $M/M/1$ Server Queue)**.** The $M/M/1$ server queue is a birth and death process with $\alpha_i = \alpha$, $\beta_i = \beta$ for $i \neq 0$. i denotes the number of people in the queue.

Example 4.1.4 (The $M/M/\infty$ Server Queue)**.** The $M/M/\infty$ server queue is a birth and death process with $\alpha_i = \alpha$, $\beta_i = i\beta$. This is a model of a supermarket with lots of cash registers.

Example 4.1.5 (Population Growth)**.** Population growth can be modelled by a birth and death process with $\alpha_i = i\alpha$ and $\beta_i = i\beta$, where i denotes the size of the population.

4.2 Continuous-Time Markov Chains with Countably Infinite State Spaces

Define null and positive recurrence as in the discrete-time Markov chain case, but with $T_x := \inf\{t > J_1 : X(t) = x$ be the **first return time** to x.

Proposition 4.2.1. *Each positively recurrent absorbinig component has a unique stationary probability* $\boldsymbol{\pi}$*, and*

$$\pi(x) = \frac{\mathbb{E}[W_x]}{\mathbb{E}[T_x]} = \frac{\textit{mean waiting time}}{\textit{mean return time}}.$$

In continuous time, you can get "explosion":

Definition 4.2.1 (Explosion)**.** Let $J_\infty = \lim_{n\to\infty} J_n$. If $\mathbb{P}[J_\infty = \infty] < 1$, then the continuous-time Markov chain is called **explosive**.

Remark. Explosion means there is a positive probability for infinitely many events to happen in a bounded time.

Proposition 4.2.2. *If* $\sup_{i\in S} |G_{i,i}| < \infty$*, then the Markov chain is not explosive.*

Example 4.2.1 (An Example with Explosion)**.** Consider a birth and death process with $X(0) = 1$, $\alpha_i = i^2$, $\beta_i = 0$. Then

$$\begin{aligned}
\mathbb{E}[J_\infty] &= \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}[W_i] \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{\alpha_i} \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \frac{1}{i^2} \\
&= \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty
\end{aligned}$$

So $\mathbb{P}[J_\infty = \infty] = 0$.

4.3 Semi-Markov Chains

Take a discrete-time Markov chain and make a continuous-time process by waiting a time W_x in each state x independently of previous and future states (but W_x is not necessarily exponentially distributed). This continous-time process is called a **semi-Markov chain**.

Semi-Markov chains allows for latent period and more general variations of infectivity with time from infection, etc.

Chapter 5

Continuous State Space Markov Processes

5.1 Gaussian Processes

Definition 5.1.1 (Gaussian Processes)**.** Suppose $X : T \mapsto \mathbb{R}$ where T is an arbitrary domain, then it is called a **Gaussian process** if $\forall n \in \mathbb{N}_+$, $\forall t_1, \cdots, t_n \in T$, $(X(t_1), \cdots, X(t_n))$ follows a multivariate Gaussian distribution, i.e., its PDF has the form

$$f(x_1, \cdots, x_n) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right)$$

for some $\boldsymbol{\mu} = \begin{bmatrix}\mu_1, \cdots, \mu_n\end{bmatrix}^T$ and Σ being a $n \times n$ (symmetric) positive definite matrix.

Proposition 5.1.1. *A nice result is that there exist $m : T \mapsto \mathbb{R}$, $c : T \times T \mapsto \mathbb{R}$, such that*

$$\mu_i = m(t_i),\ \Sigma_{i,j} = c(t_i, t_j),$$

with c being "positive definite", i.e. such that Σ is positive definite, $\forall n \in \mathbb{N}_+$ and $\forall t_1, \cdots, t_n \in T$.

*This means we can specify a Gaussian process by giving its **mean function** $m(\cdot)$ and its **covariance function** $c(\cdot, \cdot)$.*

Example 5.1.1. Let $T = \mathbb{R}$, $m(t) = 0$, $c(t, t') = e^{-|t'-t|}$, then this process is called a **stationary Orstein-Uhlenbeck process**.

Remark. One can allow degenerate Gaussians, e.g. an Orstein-Uhlenbeck process with initial condition $X(0) = 0$, then

$$f(x_0) = \delta_0(x_0).$$

The best way to define a general multivariate Gaussian is through its characteristic function

$$\mathbb{E}\left[e^{i\boldsymbol{\theta}^T \boldsymbol{X}}\right] = e^{i\boldsymbol{\theta}^T \boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\theta}^T \Sigma \boldsymbol{\theta}}.$$

We can include vector valued Gaussian processes.

Definition 5.1.2 (Vector Valued Gaussian Processes)**.** Suppose $X : T \mapsto R^n$, then X is a **vector valued Gaussian process** if $X' : T \times \{1, \cdots, n\} \mapsto \mathbb{R}$ is a scaler valued Gaussian process.

Definition 5.1.3 (Stationarity)**.** If $T = \mathbb{R} \times K$, then say the Gaussian process is stationary if $m(t, k)$ is independent of t and $c(t, k; t', k')$ is a function only of $t - t'$, k and k'.

5.2 Markov Processes with $S = \mathbb{R}$

Definition 5.2.1 (Markov Processes)**.** Let $X : T \mapsto \mathbb{R}$, where T can be either $\mathbb{N}$ or $\mathbb{R}$, X is called a **Markov process** if it satisfies the **Markoc property**:

$$\mathbb{P}\left[X(t_{n+1}) \in A | X(t_n) = x_n, \cdots, X(t_1) = x_1\right] = \mathbb{P}\left[X(t_{n+1}) \in A | X(t_n) = x_n\right],$$

for any $A \subset S$, $n \in \mathbb{N}$, and $t_{n+1} > t_n > \cdots > t_1$.

Definition 5.2.2 (Homogeneity)**.** Say a Markov process is **homogeneous** if $\mathbb{P}[X(t) \in A | X(t') = x] = \mathbb{P}[X(t - t') \in A | X(0) = x]$, $\forall t, t' \in T$.

Remark. It's unlikely that $\mathbb{P}[X(t) = y | X(0) = x] > 0$, so instead specify $\mathbb{P}[X(t) \in A | X(0) = x]$ for a measurable subset $A \subset \mathbb{R}$ as

$$\int_A p_t(x, y)\, \mathrm{d}y$$

for a transition density p_t.

Definition 5.2.3 (Transition Densities)**.** $p_t : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}$ is called a **transition density** for a homogeneous Markov process if

$$\mathbb{P}\left[X(t) \in A | X(0) = x\right] = \int_A p_t(x, y)\, \mathrm{d}y,$$

where $A \subset \mathbb{R}$ is any measurable set.

Remark. There is a technical problem that perhaps $\mathbb{P}[X(0) = x] = 0$, so we need to extend the concept of conditional probability. Any choice of interpretation of conditional probability such that

$$\mathbb{P}[X(t) \in A] = \int \mathbb{P}\left[X(t) \in A | X(0) = x)\right] \mathrm{d}\mathbb{P}[X(0) \leq x] \qquad \text{(Riemann–Stieltjes integral)}$$

will do.

Theorem 5.2.1 (The Chapman-Kolmogrov Equation)**.** *The Markov property and homogeneity imply the* ***Chapman-Kolmogrov equation****:*

$$p_{t+u}(x, y) = \int_{\mathbb{R}} p_t(x, z) p_t(z, y)\, \mathrm{d}z.$$

5.2.1 Jump Processes

Definition 5.2.4 (Jump Processes)**.** $\{X(t) : t \in T\}$ is a **jump process** if there is a **jump rate density** $r(x, y)$ with the **exit rate**

$$R(x) = \int r(x, y)\, \mathrm{d}y \leq M < \infty, \forall x \in \mathbb{R},$$

where M is a constant.

The transition density satisfies

$$p_{\Delta t}(x, y) = r(x, y)\Delta t + (1 - R(x)\Delta t)\delta(y - x) + o(\Delta t), \text{ as } \Delta t \to 0.$$

Theorem 5.2.2 (The Kolmogrov-Feller Equation)**.** *The* ***Kolmogrov-Feller equation*** *for a jump process X is*

$$\frac{\partial}{\partial t} p_t(x, y) = \int_{\mathbb{R}} p_t(x, y) r(z, y) - p_t(x, y) r(y, z)\, \mathrm{d}z.$$

5.2.2 Diffusion Processes

The Brownian Motion

One example of the diffusion process is the Brownian motion (also called the Wiener process).

Definition 5.2.5 (The Brownian Motion)**.** The **Brownian motion** $B : \mathbb{R}_+ \mapsto \mathbb{R}$ is the Gaussian process with mean $m(t) = 0$ and covariance $c(t, t') = \min(t, t')$ and almost surely continuous paths.

Proposition 5.2.1. *The Brownian motion is a Markov process with independent increments: $\forall t_1 < \cdots, t_n \in \mathbb{R}_+$,*

$$\left\{X(t_{k+1}) - X(t_k) | k = 1, \cdots, n-1\right\}$$

are independent random variables.

Furthermore, the increments are stationary: $X(t) - X(s)$ and $X(t-s) - X(0) = X(t-s)$ have the same distribution, for $t > s$. So $B(t)$ is homogeneous.

Proposition 5.2.2. *The transition density $p_t(x, y)$ of a Brownian motion is $\mathcal{N}(y - x, t)$, which satisfies the heat equation*

$$\frac{\partial}{\partial t} p_t(x, y) = \frac{1}{2} \frac{\partial^2}{\partial y^2} p_t(x, y).$$

with initial condition

$$p_0(x, y) = \delta(y - x).$$

Remark. $B(t)$ is not stationary, and $B(t) \sim \mathcal{N}(0, t)$.

Proposition 5.2.3. *$B(t)$ is scale invariant: $B(\lambda t)$ and $\sqrt{\lambda} B(t)$ have the same distribution.*

Proposition 5.2.4. *$B(t)$ is continuous almost surely, but it is also almost surely nowhere differentiable.*

Let $\xi_{t,h} := \frac{B(t+h) - B(t)}{h} \sim \mathcal{N}(0, \frac{1}{h})$, then $\xi_{t,h}$ can take arbitrarily large values as $h \to 0$. But we can informally talk about and use the limit process $\xi_t := \lim_{h \to 0} \xi_{t,h}$, which is called the **Gaussian white noise**. It can be considered as a limiting case of a Gaussian process with mean $m(t) = 0$ and covariance $c(t, t') = \delta(t - t')$. Then $B(t) = \int_0^t \xi_{t'} \, \mathrm{d}t'$. Or, we can write it as a stochastic differential equation

$$\frac{\mathrm{d}B}{\mathrm{d}t} = \xi, \; B(0) = 0.$$

Generators as Operators

For a continuous-time Markov chain on a conutable state space S, for any function $f : S \mapsto \mathbb{R}$,

$$\mathbb{E}[f(X(t))] = \sum_{x \in S} \boldsymbol{\pi}_t(x) f(x) = \left\langle \boldsymbol{\pi}_t | \boldsymbol{f} \right\rangle.$$

So

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}[f(X(t))] = \frac{\mathrm{d}}{\mathrm{d}t} \left\langle \boldsymbol{\pi}_t | \boldsymbol{f} \right\rangle = \left\langle \boldsymbol{\pi}_t \, | G | \, \boldsymbol{f} \right\rangle = \mathbb{E}[Gf(X(t))].$$

This allows use to think of the generator G as acting on the function $f : S \mapsto \mathbb{R}$ by

$$(Gf)(x) = \sum_{y \neq x} G_{x,y} \left(f(y) - f(x) \right).$$

Extend this to $S = \mathbb{R}$ by replacing matrices and vectors by operators and functions. For the Brownina motion,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[f(X(t))] =& \int_{\mathbb{R}} \frac{\partial p_t(x,y)}{\partial t} f(y)\,\mathrm{d}y \\ =& \frac{1}{2}\int_{\mathbb{R}} \frac{\partial^2 p_t(x,y)}{\partial y^2} f(y)\,\mathrm{d}y \\ =& \mathbb{E}[(\mathcal{L}f)(X(t))]\end{aligned}$$

If f is chosen to be twice differentiable and f & $f' \to 0$ as $x \to \pm\infty$, then integration by parts gives

$$(\mathcal{L}f)(x) = \frac{1}{2}f''(x).$$

Notice that $\mathcal{L}$ is linear in this case.

Definition 5.2.6 (Generators)**.** We call $\mathcal{L}$ a **generator** on functions.

For a jump process on $\mathbb{R}$,

$$(\mathcal{L}f)(x) = \int_{\mathbb{R}} r(x,y)[f(y) - f(x)]\,\mathrm{d}y.$$

We can obtain the Brownian motion as a scaling limit of a jump process. Take a jump process $X(t)$ with $X(0) = 0$, and $r(x,y) = q(y-x)$ such that

$$\int_{\mathbb{R}} zq(z)\,\mathrm{d}z = 0, \quad \int_{\mathbb{R}} z^2 q(z)\,\mathrm{d}z = \sigma^2 \in (0,\infty).$$

Then $\forall T > 0$, with,

$$\frac{\epsilon}{\sigma} X(\frac{t}{\epsilon^2})\Big|_{t\in[0,T]} \xrightarrow{\mathrm{d}} B(t), \text{ as } \epsilon \to 0.$$

General Diffusions on $\mathbb{R}$

Chapter 6

Stochastic Particle Systems

Chapter 7

Networks

Chapter 8

Random Networks