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 $\implies \bigcup\limits_{i \in \mathcal{I}} A_i \in \mathcal{F}$.

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

[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 $\mathbb{P}[B_i] > 0$, $\forall i \in \mathcal{I}$. Then $\forall A \subset \mathcal{F}$, we have*

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

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 $\mathbb{P}[B_i] > 0$, $\forall i \in \mathcal{I}$, then

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

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. $\mathbb{P}(\text{has COVID}) = 0.001$, what is the probability that a person has COVID given his/her test is positive?

Sol.

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

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

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

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

$$P_{n+k} = P_n P_k \quad \text{and in particular} \quad P_{n+1} = P_n P_1.$$

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 Markoc 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. $P | \mathbf{1} \rangle = | \mathbf{1} \rangle \implies P$ has an eigenvalue 1, so P also has a left eigenvector with eigenvalue 1, i.e. $\langle \boldsymbol{\pi} | \neq \langle \mathbf{0} |$. Then normalize it to make $\langle \boldsymbol{\pi} | \mathbf{1} \rangle = 1$. □

Remark. There can be more than one stationary distributions, and a convex combination of two stationary distributions is also a stationary distribution.

Chapter 4

Continuous-Time Markov Chain

Chapter 5

Continuous State Space Markov Processes

Chapter 6

Stochastic Particle Systems

Chapter 7

Networks

Chapter 8

Random Networks