Review Sheet

EE 226A

Ahmed Shakil

Contents

1 Discrete-Time Markov Chains

1.1 Definition of a Markov Chain

Definition 1.1: Markov chain

A Markov chain is a process $(X_n)_{n \geqslant 0}$ satisfying

$$\Pr\{X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} = \Pr\{X_{n+1} = j \mid X_n = i\}$$

for all $n \geqslant 1$ and $j, i, i_{n-1}, i_0 \in \mathcal{S}$. A Markov chain is said to be **temporally homogeneous** if there are numbers $(p_{ij})_{i,j \in \mathcal{S}}$ such that

$$\Pr\{X_{n+1} = j \mid X_n = i\} = p_{ij}$$

for all $n \geqslant 0$ and all states $i, j \in \mathcal{S}$. The numbers $(p_{ij})_{i,j \in \mathcal{S}}$ are generically referred to as the **transition probabilities** of the Markov chain.

Definition 1.2: Transition matrix

$$P = \begin{bmatrix} p_{00} & p_{01} & \cdots \\ p_{10} & p_{11} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

The matrix P is called the transition matrix. It is a stochastic matrix, which means it is square with non-negative entries, whose rows sum to one. A Markov chain and transition matrix are equivalent representations of each other.

1.1.1 The Chapman-Kolmogorov equations

Proposition 1.3: The Chapman-Kolmogorov Equations

We define **multi-step transition probabilities**

$$P^n_{ij} := \Pr\{X_{n+m} = j \mid X_m = i\}, \quad n, m \geqslant 0.$$

The Chapman-Kolmogorov equations give a recursive formula for computing the n-step transition probabilities.

For all $m, n \geqslant 0$ and states i, j,

$$P^{m+n}_{ij} = \sum_k P^m_{ik} P^n_{kj}.$$

In particular, we have

$$P^n = \begin{bmatrix} P^n_{00} & P^n_{01} & \cdots \\ P^n_{10} & P^n_{11} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

where P^n is the transition matrix P raised to the n^{th} power. Also note $P^{m+n} = P^m P^n$.

Definition 1.4: Irreducible

A **class** of states is a nonempty set of states such that every state in the set can communicate with one another. These classes form an equivalence relation and partition the state space. We say a Markov chain is irreducible if there is only one class.

Definition 1.5: Periodicity

For a state i, define its **period**

$$d(i) \coloneqq \gcd\{n \geqslant 1 : P_{ii}^n > 0\}.$$

States with period 1 are called **aperiodic**. Periodicity is a class property.

Proposition 1.6

Define the **first return time** for state $j \in \mathcal{S}$ as

$$T_j \coloneqq \inf\{n \geqslant 1 : X_n = j\}.$$

Let $(X_n)_{n \geqslant 0}$ be a Markov chain with transition matrix P. Conditioned on the event $\{T_j < \infty\}$, the process $(X_{n+T_j})_{n \geqslant 0}$ is a Markov chain with transition matrix P and starting state j and is independent of $X_0, \ldots, X_{T_j}$.

Proposition 1.7

A state j is **recurrent** if $\Pr\{T_j < \infty \mid X_0 = j\} = 1$, and it is called **transient** if $\Pr\{T_j < \infty \mid X_0 = j\} < 1$. Recurrence and transience are class properties.

Lemma 1.8

State i is recurrent if and only if $\sum_{n=1}^{\infty} P_{ii}^n = \infty$.

1.2 Markov Limit Theorems

Theorem 1.9: Strong Law of Large Numbers for Markov Chains

Define $N_j(n)$, $n \geqslant 1$, to be the number of transitions into state j, up to and including time n. More precisely,

$$N_j(n) \coloneqq \#\{1 \leqslant k \leqslant n : X_k = j\}.$$

Also define the expected first return time to be

$$\mu_{jj} \coloneqq \mathbb{E}[T_j | X_0 = j].$$

Let $(X_n)_{n\geqslant 0}$ be a Markov chain starting in state $X_0 = i$. If $i \leftrightarrow j$, then

$$\frac{N_j(n)}{n} \to \frac{1}{\mu_{jj}} \quad \text{a.s.}$$

Corollary 1.10

For an irreducible Markov chain $(X_n)_{n\geqslant 0}$, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} P_{ij}^{k} = \frac{1}{\mu_{jj}}.$$

1.2.1 Stationary distributions and the issue of convergence

Definition 1.11: Stationary Distribution

A probability distribution $(\pi_j)_{j\in\mathcal{S}}$ is a stationary distribution for a Markov chain with transition probability matrix P if $\pi_j = \sum_i \pi_i p_{ij}$ for each $j \in \mathcal{S}$. Equivalently in matrix notation, $\pi = \pi P$, when π is considered as a row vector.

Definition 1.12: Positive and Null recurrence

A recurrent state j is positive recurrent if $\mu_{jj} < \infty$, or null recurrent if $\mu_{jj} = \infty$. Positive and null recurrence are class properties.

It is important to note stationary distributions aren't necessarily unique. It is important to note that stationary distribution does not always exist.

Theorem 1.13

An irreducible Markov chain satisfies exactly one of the following:

1. All states are transient, or all states are null recurrent. In this case, $\frac{1}{n}\sum_{k=1}^{n} P_{ij}^{k} \to 0$ as $n \to \infty$ for all states i, j, and no stationary distribution exists.

2. All states are positive recurrent. In this case, a unique stationary distribution exists and is given by $\pi_j = \frac{1}{\mu_{jj}} = \lim_{n\to\infty} \frac{1}{n}\sum_{k}^{n} P_{ij}^{k}$.

Theorem 1.14

Let $(X_n)_{n\geqslant 0}$ be an irreducible, aperiodic, and positive recurrent Markov chain with stationary distribution π. Then

$$\lim_{n\to\infty} \sum_{j} \left|P_{ij}^{n} - \pi_j\right| = 0 \text{ for all } i \in \mathcal{S}.$$

In particular, $P_{ij}^{n} = \pi_j$ for all i, j.

1.3 Reversibility and Spectral Gap

Definition 1.15: Reversible Markov chain

A Markov chain with transition matrix P and stationary distribution π is reversible if the transition probabilities satisfy

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

In this case, we say (P, π) is reversible for convenience.

The equations above are called the detailed balance equations. These equations are also sufficient for reversibility. Hence, if there is a probability distribution π such that the equations are satisfied, then the Markov chain is reversible with stationary distribution π.

1.3.1 Spectral gap and trend to equilibrium

Definition 1.16: Total variation distance

For probability measures μ, ν on a measurable space $(\Omega.\mathcal{F})$, we define their total variation distance

$$\|\mu - \nu\|_{TV} := \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|.$$

Also note that when Ω is countable, we have

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{\omega} |\mu(\omega) - \nu(\omega)|.$$

Definition 1.17: Spectral gap

Define $\mathrm{Var}_\pi(f) = Var(f(X))$ for $X \sim \pi$ and $f : \mathcal{S} \to \mathbb{R}$. Also define the function $Pf : \mathcal{S} \to \mathbb{R}$ via the matrix-vector multiplication

$$(Pf)(i) = \sum_j p_{ij} f(j), \quad i \in \mathcal{S}.$$

For a reversible Markov Chain (P, π), define the spectral gap $\gamma := 1 - \lambda_2$, where λ_2 is the smallest number satisfying

$$\mathrm{Var}_\pi(Pf) \leqslant \lambda_2 \mathrm{Var}_\pi(f) \text{ for all } f : \mathcal{S} \to \mathbb{R} \text{ with } \mathrm{Var}_\pi(f) < \infty.$$

Theorem 1.18

If (P, π) is a reversible Markov chain with spectral gap γ, then

$$\|P^n_{i\bullet} - \pi\|^2_{TV} \leqslant \frac{(1-\gamma)^n}{\pi_i} \text{ for each } n \geqslant 0 \text{ and each state } i.$$

Moreover, if the Markov chain is irreducible, aperiodic and has a finite state space, then $\gamma > 0$.

Definition 1.19

Let μ, π be probability distributions on state space $\mathcal{S}$. We say that μ has density h with respect to π (written $d\mu = h d\pi$) if

$$\mu_j = h(j)\pi_j \quad \forall j.$$

Lemma 1.20

Let (P, π) be a reversible Markov chain. If $d\mu = h d\pi$, then $P^n h$ is the density of μP^n with respect to π.

2 Martingales

2.1 Definitions and Examples

Definition 2.1

Let $(X_n)_{n\geqslant 0}$ be a stochastic process. A process $(M_n)_{n\geqslant 0}$ is said to be a **martingale with respect to** $(X_n)_{n\geqslant 0}$ if $(M_n)_{n\geqslant 0}$ is adapted to $(X_n)_{n\geqslant 0}$ and, for each $n \geqslant 0$,

(i) $\mathbb{E}|M_n| < \infty$;

(ii) $\mathbb{E}[M_{n+1} \mid X_0, \ldots, X_n] = M_n$.

If the equality in *(ii)* is replaced by $\geqslant$ or $\leqslant$, then the process is said to be a **submartingale** or **supermartingale**, respectively. The phrase “adapted to” means that M_n is a measurable function $(X_0, \ldots, X_n)$ for each $n \geqslant 0$.

Proposition 2.2

If $(M_n)_{n\geqslant 0}$ is a martingale with respect to $(X_n)_{n\geqslant 0}$, then for all $m > n$,

$$\mathbb{E}[M_m \mid X_0, \ldots, X_n] = M_n.$$

If $(M_n)_{n\geqslant 0}$ is a submartingale or supermartingale, then the equality above is $\geqslant$ or $\leqslant$, respectively.

2.2 Stopping Times

Definition 2.3

A nonnegative integer-valued random variable T is a **stopping time** with respect to $(X_n)_{n\geqslant 0}$ if, for each $n \geqslant 0$, the occurrence of the event $\{T \leqslant n\}$ is determined entirely by $(X_0, \ldots, X_n)$. In other words, the indicator $1_{\{T\leqslant n\}}$ is a measurable function of $(X_0, \ldots, X_n)$.

Definition 2.4: Stopped process

Let $(X_n)_{n\geqslant 0}$ be a process, and T be a stopping time. If $(Y_n)_{n\geqslant 0}$ is adapted to $(X_n)_{n\geqslant 0}$, then the process $(Y_{T\wedge n})_{n\geqslant 0}$ is called the **stopped process**. Note that the stopped process satisfies $Y_{T\wedge n} = Y_n$ for $n \leqslant T$, and $Y_{T\wedge n} = Y_T$ for $n > T$.

2.2.1 Stopping times and martingales

Proposition 2.5

If $(M_n)_{n\geqslant 0}$ is a martingale and T is a stopping time, both with respect to $(X_n)_{n\geqslant 0}$, then the stopped process $(M_{T\wedge n})_{n\geqslant 0}$ is also a martingale with respect to $(X_n)_{n\geqslant 0}$.

Proposition 2.6

If $(M_n)_{n\geqslant 0}$ is a submartingale and T is a stopping time, both with respect to $(X_n)_{n\geqslant 0}$, then

$$\mathbb{E}[M_0] \leqslant \mathbb{E}[M_{T\wedge n}] \leqslant \mathbb{E}[M_n] \quad n \geqslant 0.$$

Proposition 2.7: Optimal Stopping Theorem

Let $(M_n)_{n\geqslant 0}$ be a submartingale and T be a stopping time, both with respect to $(X_n)_{n\geqslant 0}$. If there is a constant $k < \infty$ such that any one of the following hold

i.) $T \leqslant k$ a.s.; or

ii.) $|M_n| \leqslant k$ a.s. for each n, and $\Pr\{T < \infty\} = 1$; or

iii.) $\mathbb{E}[T] < \infty$ and $|M_n - M_{n-1}| \leqslant k$ a.s. for each $n \geqslant 1$,

then

$$\mathbb{E}[M_0] \leqslant \mathbb{E}[M_T].$$

The inequality above is an equality when $(M_n)_{n\geqslant 0}$ is a martingale.

Theorem 2.8: Wald's Identity

Let $(Y_n)_{n\geqslant 1}$ be adapted to $(X_n)_{n\geqslant 1}$. Assume Y_{n+1} is independent of $(X_1, \ldots, X_n)$ for each $n \geqslant 1$, $\sup_{n\geqslant 1} \mathbb{E}|Y_n| < \infty$, and $\mathbb{E}[Y_n] = \mu$ for all $n \geqslant 1$. If $T \geqslant 1$ is a stopping time with respect to $(X_n)_{n\geqslant 1}$ satisfying $\mathbb{E}[T] < \infty$, then

$$\mathbb{E}\left[\sum_{n=1}^{T} Y_n\right] = \mu\mathbb{E}[T].$$

3 Poisson Processes

3.1 The Exponential Distribution

Definition 3.1: Exponential distribution

The **exponential distribution with rate** $\lambda > 0$, denote $\text{Exp}(\lambda)$, has density

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & t \geqslant 0 \\ 0 & t < 0. \end{cases}$$

For $T \sim \text{Exp}(\lambda)$, the distribution function is given by

$$\Pr\{T \leqslant t\} = \begin{cases} 1 - e^{-\lambda t} & t \geqslant 0 \\ 0 & t < 0. \end{cases}$$

We note that $\mathbb{E}[T] = 1/\lambda$ and $\text{Var}(T) = 1/\lambda^2$. An important property of exponential random variables is the **memoryless property**. In particular, if $T \sim \text{Exp}(\lambda)$, then

$$\Pr\{T > t + s | T > t\} = \frac{\Pr\{T > t + s\}}{\Pr\{T > t\}} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = \Pr\{T > s\}.$$

Definition 3.2: Erlang distribution

If, $T_1, \ldots, T_k$ are i.i.d. exponential random variables with rate λ, then their sum $T = T_1 + \cdots + T_k$ has an **Erlang** distribution, with density

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t} \frac{(\lambda t)^{k-1}}{(k-1)!} & t \geqslant 0 \\ 0 & t < 0. \end{cases}$$

3.2 Poisson Processes

Definition 3.3: Poisson Process

Let $\tau_1, \tau_2, \ldots$ be i.i.d. exponential random variables with rate $\lambda > 0$ and, for $n \geqslant 1$, define $T_n = \tau_1 + \tau_2 + \cdots + \tau_n$, with the convention that $T_0 = 0$. For each $t \geqslant 0$, define the random variable $N_t = \sup\{n \geqslant 0 : T_n \leqslant t\}$. The process $(N_t)_{t \geqslant 0}$ is called a **Poisson process** with rate λ.

This is best thought of as an example of a counting process. A **counting process** is a random process $(N_t)_{t \geqslant 0}$, such that (i) N_t is a non-negative integer for each time $t \geqslant 0$; (ii) the sample paths $t \mapsto N_t(\omega)$ are non-decreasing in t; and (iii) the sample paths $t \mapsto N_t(\omega)$ are right-continuous.

Definition 3.4: Poisson distribution

A random variable X is said to be Poisson distributed with mean $\lambda \geqslant 0$ ($X \sim \text{Poisson}(\lambda)$) if X has probability

mass function
$$\Pr\{X = k\} = e^{-\lambda}\frac{\lambda^k}{k!},\ k = 0, 1, 2, \dots$$

Proposition 3.5

If $(N_t)_{t\geqslant 0}$ is a Poisson process with rate $\lambda \geqslant 0$, then for each $t \geqslant 0$, we have $N_t \sim \text{Poisson}(\lambda t)$.

Theorem 3.6

Let $(N_t)_{t\geqslant 0}$ be a Poisson process with rate λ. For any finite collection of distinct time instants $0 = t_0 < t_1 < \cdots < t_k$, the increments $(N_{t_1} - N_{t_0}), \dots, (N_{t_k} - N_{t_{k-1}})$ are independent with $(N_{t_i} - N_{t_{i-1}}) \sim \text{Poisson}(\lambda(t_i - t_{i-1}))$ for each $1 \leqslant i \leqslant k$.

Theorem 3.7: Characterization of Poisson Processes

If $(N_t)_{t\geqslant 0}$ is a Poisson process, then the following hold:

1. $N_0 = 0$;
2. $N_t \sim \text{Poisson}(\lambda t)\ \forall t \geqslant 0$;
3. $(N_t)_{t\geqslant 0}$ has independent increments.

Conversely, if these properties hold for a counting process $(N_t)_{t\geqslant 0}$, then it is a Poisson process.

3.3 Conditioning on Arrivals

Definition 3.8

Let $X_1, X_2, \dots, X_k$ be a collection of random variables. The **order statistics** $X_{(1)}, \dots, X_{(k)}$ are the random variables defined by sorting the realizations of $X_1, X_2, \dots, X_k$ into increasing order.

Theorem 3.9

Let $(N_t)_{t\geqslant 0}$ be a Poisson process with arrivals $(T_i)_{i\geqslant 1}$. Conditioned on the event $\{N_t = n\}$, the vector of arrival times $(T_1, \dots, T_n)$ has the same distribution as that of order statistics $(U_{(1)}, \dots, U_{(n)})$, where $U_i \sim \text{Unif}(0, t),\ 1 \leqslant i \leqslant n$ are independent.

Theorem 3.10

Let $(N_t)_{t\geqslant 0}$ be a Poisson process with rate λ and corresponding arrivals $(T_n)_{n\geqslant 1}$. For a Borel set $B \subset [0, \infty)$, let $|B|$ denotes its Lebesgue volume, and let $N(B)$ denote the number of arrivals in B; i.e.,
$$N(B) = \#\{n \geqslant 1 : T_n \in B\}.$$

If $B_1, B_2, \dots \subset [0, \infty)$ are disjoint, bounded Borel sets, then $N(B_1), N(B_2), \dots$ are independent, with $N(B_i) \sim$

Poisson($\lambda|B_i|$).

Theorem 3.11: Slivnyak's Theorem

Let $(N_t)_{t\geqslant 0}$ be a Poisson process with rate λ and let $x \in (0, \infty)$. Conditioned on one arrival at time x, the other arrivals form an (unconditional) rate-λ Poisson process.

4 Continuous-Time Markov Chains

4.1 Definitions and Constructions

Definition 4.1

A process $(X_t)_{t\geqslant 0}$ taking values in $\mathcal{S}$ is a temporally homogeneous **continuous-time Markov chain** if:

(i) given any initial state $X_0 = i \in \mathcal{S}$, the sample paths $t \mapsto X_t$ are a.s. right-continuous (with respect to the discrete topology on $\mathcal{S}$); and

(ii) for any choice of discrete time instants $0 \leqslant t_1 < \cdots < t_k < t \leqslant s$ and states $i_1, i_2, \ldots, i_k, i, j \in \mathcal{S}$, we have the Markov property

$$\Pr\{X_s = j \mid X_t = i, X_{t_k} = i_k, \ldots, X_{t_1} = i_1\} = \Pr\{X_{s-t} = j \mid X_0 = i\}.$$

Theorem 4.2

Let $(X_t)_{t\geqslant 0}$ be a continuous-time Markov chain. The transition probabilities satisfy

$$P^{s+t} = P^s P^t \text{ for all } s, t \geqslant 0,$$

and $\lim_{t\downarrow 0} P^t = I$. In other words, the transition probabilities $(P^t)_{t\geqslant 0}$ form a **Markov semigroup**.

Theorem 4.3

Let $(X_t)_{t\geqslant 0}$ be a continuous-time Markov chain with initial non-absorbing state $X_0 = i$. The holding time $T = \inf\{t \geqslant 0 : X_t \neq i\}$ has distribution $T \sim \mathrm{Exp}(\lambda_i)$ for $\lambda_i \geqslant 0$ satisfying

$$P^h_{ii} = 1 - h\lambda_i + o(h).$$

Moreover, the next state X_T is independent of T and has distribution

$$p_{ij} := \Pr\{X_T = j \mid X_0 = i\} = \lim_{h\downarrow 0} \frac{P^h_{ij}}{1 - P^h_{ii}}, \quad j \neq i.$$

Theorem 4.4

Let $(X_t)_{t\geqslant 0}$ be a continuous-time Markov chain with transition probabilities $(P^t)_{t\geqslant 0}$, starting in non-absorbing state $X_0 = i$, and let $T = \inf\{t \geqslant 0 : X_t \neq i\}$ denote the time of the first transition. Conditioned on T and $X_T = j$, the process $(X_{T+t})_{t\geqslant 0}$ is a continuous-time Markov chain with transition probabilities $(P^t)_{t\geqslant 0}$ and starting state j.

Definition 4.5

The transition probabilities $(p_{ij})_{i,j\in\mathcal{S}}$ (with $p_{ii} = 0$) define a discrete-time Markov chain, known as the **embedded chain**. The parameters $(\lambda_i)_{i\in\mathcal{S}}$ are called the **transition rates** for the Markov chain, λ_i is precisely the rate at which the process transitions out of state i.

Lemma 4.6

Let $(p_{ij})_{i,j\in\mathcal{S}}$ be transition probabilities for a discrete-time Markov chain $(X_n)_{n\geqslant 0}$ starting in non-absorbing state $X_0 = i$.

(i) The random variable $N \coloneqq \inf\{n \geqslant 0 : X_n \neq i\}$ is geometric with distribution

$$\Pr\{N = k \mid X_0 = i\} = p_{ii}^{k-1}(1 - p_{ii}), \quad k \geqslant 1$$

(ii) The random variable X_N is independent of N, and has distribution

$$\Pr\{X_N = j \mid X_0 = i\} = \frac{p_{ii}}{(1 - p_{ii})}, \quad j \neq i.$$

4.2 The Infinitesimal Generator

Definition 4.7

The **infinitesimal generator** for a continuous-time Markov chain $(X_t)_{t\geqslant 0}$ with transition rates $(\lambda_i)_{i\in\mathcal{S}}$ is a matrix Q with entries

$$q_{ij} \coloneqq [Q]_{ij} = \begin{cases} \lambda_i p_{ij} & \text{for } j \neq i \\ -\lambda_i & \text{for } j = i, \end{cases}$$

where $(p_{ij})_{i,j\in\mathcal{S}}$ are the transition probabilities for the embedded chain. In particular, $\lambda_i = \sum_{j\neq i} q_{ij}$.

The numbers $(q_{ij})_{i,j\in\mathcal{S}}$ are called the **jump rates** for the Markov chain. Essentially, q_{ij} describes the rate at which the Markov chain with infinitesimal generator Q transitions from state i to state j ($j \neq i$).