MATH4240 Tutorial 2 Notes

1 Definitions

A *stochastic process* X is a collection of random values on a common probability space that take values in a common space: $X_t = X(t) : \Omega \to S$ for each t in an index set $\mathcal{T}$, typically interpreted as time.

For now, we will assume that

- index set $\mathcal{T}$ is natural number $\mathbb{N}$
- *state space* S is a countable set, usually finite

In particular, we may write the process as a sequence of random variables $X_0, X_1, \ldots$.

Definition 1. A *(discrete) Markov chain* is a (discrete time) stochastic process $\{X_n\}$ (on discrete state space) with the *Markov property*: for all $n \geq 0$, $t_0 < \ldots < t_{n+1}$, $x_0, \ldots, x_{n+1} \in S$,

$$P\left(X_{t_{n+1}} = x_{n+1} \;\middle|\; X_{t_n} = x_n \text{ and } X_{t_i} = x_i, \forall i < n \right) = P\left(X_{t_{n+1}} = x_{n+1} \;\middle|\; X_{t_n} = x_n \right)$$

That is, given precisely the current state, the behavior of the chain in the future is independent of the past.

We will also assume that the chain is *(time) homogeneous*: for all $n, m \in \mathbb{N}$ and $x, y \in S$,

$$P\left(X_{n+1} = y \mid X_n = x \right) = P\left(X_{m+1} = y \mid X_m = x \right)$$

For such chain, the *(one-step) transition matrix* is

$$P_{xy} = P(x, y) = P\left(X_{n+1} = y \mid X_n = x \right)$$

The distribution π_n of the chain at time n is described via the pmf

$$\pi_n(x) = P\left(X_n = x\right), \; \forall x \in S$$

In particular, its *initial distribution* is π_0. We *will* treat the pmf as a *row vector*, so that $\pi_n = \begin{pmatrix}\pi_n(1) & \ldots & \pi_n(k)\end{pmatrix}$ (assuming the state space is $\{ 1, \ldots, k \}$).

Occasionally, the initial distribution is deterministic (i.e. $\pi_0(x_0) = 1$ for some $x_0 \in S$), in which case we will write the corresponding probability as $P_{x_0}(E) = P\left(E \mid X_0 = x_0\right)$ and expectation as $E_{x_0}(Y) = E\left(Y \mid X_0 = x_0\right)$.

Remark 1. We can write the Markov property in a (slightly) more general way: on $T \geq 0$, $E_{<T}$ is an event that depends only on X_t with $t < T$, $E_{\geq T}$ is an event that depends only on X_t with $t \geq T$, $x \in S$, then

$$P\left(E_{\geq T} \mid X_T = x \text{ and } E_{<T} \right) = P\left(E_{\geq T} \mid X_T = x \right)$$

Note that we *need* to have the *precise* information of the present state, i.e. $X_T = x$. Quoting Chung from *Green, Brown, and Probability & Brownian Motion on the Line* (p.30),

> This is often described by words like: “the past has no aftereffect on the future when we know the present.” But beware of such non-technical presentation (sometimes required in funding applications because bureaucrats can’t read mathematics). Big mistakes have been made through misunderstanding the exact meaning of the words “when the present is known”.

Example 1 (Two-state Markov chain). Suppose in Hong Kong the weather of a specific day depends only on the weather of the day before, and we observed the following pattern

$$\begin{aligned} P(\text{ sunny today } \mid \text{ sunny yesterday }) &= 1-p \\ P(\text{ not sunny today } \mid \text{ sunny yesterday }) &= p \\ P(\text{ sunny today } \mid \text{ not sunny yesterday }) &= q \\ P(\text{ not sunny today } \mid \text{ not sunny yesterday }) &= 1-q \end{aligned}$$

for some $p, q \in [0,1]$. Then we can model the weather as a Markov chain, with the transition matrix represented as

$$P = \begin{array}{c} \\ S \\ N \end{array} \!\!\begin{array}{c} \begin{array}{cc} S & N \end{array} \\ \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} \end{array}$$

with $S =$ sunny, $N =$ not sunny. Note that the headings (S, N) are there just to keep track on the entries.

Example 2 (Typical chains). Here are some chains that are mentioned in the textbook:

- (birth-death chain) on X_n denoting the number of people in a region, assuming that at each time unit people are born / immigrate and die / emigrate with probabilities that depend only on the population, then

$$P(i,j) = \begin{cases} p_i & \text{if } j = i+1 \\ q_i & \text{if } j = i-1 \\ r_i & \text{if } j = i \\ 0 & \text{otherwise} \end{cases}$$

  with $p_i + q_i + r_i = 1$ for each i

- (random walk) $X_n = X_{n-1} + \xi_n$ where $\xi_1, \xi_2, \ldots$ are iid random variables

- (queuing chain) on a queue of X_n customers at time n, with iid ξ_n being the number of new customers arrived in time $(n, n+1]$, assuming one customer can be served for each time unit,

$$X_{n+1} = X_n + \xi_n - \begin{cases} 0 & \text{if } X_n = 0 \\ 1 & \text{if } X_n \neq 0 \end{cases}$$

- (branching chain) on the number X_n of particles at time n, assuming at each unit time a particle may independently split into identical particles, with the number of descendants distributed according to some distribution f, then

$$P(x,y) = P(\ \xi_1 + \ldots + \xi_x = y \mid \text{iid } \xi_1, \ldots, \xi_x \sim f\ )$$

2 Computations

As shown in lecture, for a Markov chain,

- the n-step transition probability is the n-th power of the (one-step) transition matrix:

$$P(\ X_{k+n} = y \mid X_k = x\ ) = (P^n)_{xy}$$

- given the initial distribution π_0, the distribution at time n is its product with the transition matrix:

$$\pi_n = \pi_{n-1} P = \ldots = \pi_0 P^n$$

Definition 2. On a given chain X, a random variable T with value in $\mathbb{N} \cup \{\infty\}$ is a *stopping time* if for each $n \in \mathbb{N}$, the event $\{T \leq n\}$ can be determined by $X_0, X_1, \ldots, X_n$ (possibly with randomness independent of the chain).

As a special case, for $A \subseteq S$, the *hitting time* T_A of A is the first time X_t enters A *after* the initial time:

$$T_A = \min\{\ t \geq 1 \ \mid\ X_t \in A\ \}$$

If $A = \{y\}$ is singleton, we also write $T_y = T_{\{y\}}$. Note that we start from 1 instead of 0, and (by convention) $T_A = \infty$ if $X_t \notin A$, $\forall t \geq 1$.

We also define $\rho_{xy} = P\left(\ T_y < \infty \ \mid\ X_0 = x\ \right)$ as the probability of visiting y when starting at x, so

$$P\left(X_n \neq y, \forall n \geq 1 \,|\, X_0 = x\right) = P_x\left(T_y = \infty\right) = 1 - \rho_{xy}$$

Example 3. Consider the two-state chain from Example 1, and assume that $p, q \neq 0$. Then

- given that $X_0 = S$, on $n \geq 1$ we have

  $$P_S\left(T_N = n\right) = P\left(X_1 = \ldots = X_{n-1} = S,\, X_n = N \,|\, X_0 = S\right) = P(S,S)^{n-1}P(S,N) = (1-p)^{n-1}p$$

  and so the hitting time T_N is geometrically distributed: $(T_N \mid X_0 = S) \sim \text{Geom}(p)$

- given that $X_0 = N$, we have

  $$P_N\left(T_N = 1\right) = P\left(X_1 = N \,|\, X_0 = N\right) = P(N,N) = 1 - q$$

  and on $n \geq 2$ we have

  $$P_N\left(T_N = n\right) = P\left(X_1 = \ldots = X_{n-1} = S,\, X_n = N \,|\, X_0 = N\right) = P(N,S)P(S,S)^{n-2}P(S,N) = q(1-p)^{n-2}p$$

  and so $(T_N \mid X_0 = N)$ is a mixture of distributions, and we can represent it as $(T_N \mid X_0 = N) = 1 + BG$ with independent $B \sim \text{Bernoulli}(q)$ and $G \sim \text{Geom}(p)$.

Example 4. Consider the two-state chain from Example 1, and assume that $0 < p + q < 2$. Given the initial distribution π_0, what is the probability distribution of the weather n days later?

We write the initial distribution as $\pi_0 = \begin{pmatrix} P\left(X_0 = S\right) & P\left(X_0 = N\right) \end{pmatrix} = \begin{pmatrix} \pi_0(S) & 1 - \pi_0(S) \end{pmatrix}$. Then $\pi_n = \pi_0 P^n$ with P being the transition matrix. To compute P^n, we can diagonalize P as

$$P = VDV^{-1} \quad \text{where} \quad D = \begin{pmatrix} 1 & 0 \\ 0 & 1-p-q \end{pmatrix},\ V = \begin{pmatrix} 1 & -p \\ 1 & q \end{pmatrix}$$

which implies

$$\pi_n = \pi_0 V D^n V^{-1} = \begin{pmatrix} \frac{q}{p+q} + (\pi_0(S) - \frac{q}{p+q})(1-p-q)^n & \frac{p}{p+q} + (\frac{q}{p+q} - \pi_0(S))(1-p-q)^n \end{pmatrix}$$

If we take limit $n \to \infty$, we have

$$\lim_n P^n = \lim_n \begin{pmatrix} \frac{q}{p+q} + \frac{p}{p+q}(1-p-q)^n & \frac{p}{p+q}(1-(1-p-q)^n) \\ \frac{p}{p+q}(1-(1-p-q)^n) & \frac{p}{p+q} + \frac{q}{p+q}(1-p-q)^n \end{pmatrix} = \begin{array}{c c} & \begin{array}{cc} S & N \end{array} \\ \begin{array}{c} S \\ N \end{array} & \begin{bmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{q}{p+q} & \frac{p}{p+q} \end{bmatrix} \end{array}$$

and so the limiting transition matrix exists. Also, $\lim_n \pi_n = \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \end{pmatrix}$ exists.

Example 5. Ehrenfest chain is a simplified model on gas dynamics.

We distribute d balls to two boxes. At each time unit, we pick one ball at random uniformly (among all d, independent of the past) and move it to the other box.

Let X_n denote the number of balls in the given box at time n. Then $\{X_n\}$ forms a Markov chain on state space $S = \{\ 0, \ldots, d\ \}$ with transition probability

$$P(x,y) = \begin{cases} \frac{d-x}{d} & \text{if } y = x+1 \\ \frac{x}{d} & \text{if } y = x-1 \\ 0 & \text{otherwise} \end{cases} \quad \text{for } x, y \in S$$

Given that $X_0 = x_0 \in \{\ 1, \ldots, d-1\ \}$, what is the probability that $X_2 = x_0$?

There are only two possibilities: $X_1 = x_0 - 1$, or $X_1 = x_0 + 1$. So, by Markov property,

$$\begin{aligned} P\left(\ X_2 = x_0 \ \mid\ X_0 = x_0\ \right) &= P(x_0 - 1, x_0)P(x_0, x_0 - 1) + P(x_0 + 1, x_0)P(x_0, x_0 + 1) \\ &= \frac{d - x_0 + 1}{d}\frac{x_0}{d} + \frac{x_0 + 1}{d}\frac{d - x_0}{d} \\ &= \frac{d + 2x_0(d - x_0)}{d^2} \end{aligned}$$

With a similar computation, the same result holds if $x_0 = 0$ or $x_0 = d$.