MATH4240 Tutorial 7 Notes

1 Mean Recurrence Time

- The *mean recurrence time* of a state x is the expected time $m_x = E_x(T_x)$ where the chain enters x again when started at x. A transient state x has mean recurrence time $m_x = \infty$
- A recurrent state x is *positive recurrent* if $m_x < \infty$, *null recurrent* if $m_x = \infty$
- If x is positive recurrent and $x \to y$, then y is also positive recurrent.
- If the state space is finite, then there must be at least one positive recurrent state. In particular, a finite irreducible closed set must consists only of positive recurrent states.

Theorem 1.1. *On a state y of a Markov chain, with $N_n(y) = \sum_{i=1}^n \mathbf{1}_y(X_i)$ being (the random variable of) the number of visits to state y up to time n,*

- $\lim_n \frac{1}{n} N_n(y) = \mathbf{1}_{T_y<\infty}/m_y$ *with probability* 1
- $\lim_n \frac{1}{n} E_x(N_n(y)) = \lim_n \frac{1}{n}\sum_{k=1}^n P^k(x,y) = \rho_{xy}/m_y$ *for each state* x

Note that we do not need many conditions on the chain for these to hold.

Moreover, if y is transient, both limits are 0; and if the chain is irreducible and recurrent, $\mathbf{1}_{T_y<\infty} \equiv 1$ and $\rho_{xy} = 1$, and so you have the results that are noted in the lecture.

Theorem 1.2 (Kac's recurrence, Markov chain version)**.** *For an* ***irreducible*** *Markov chain, it has stationary distribution if and only if it is positive recurrent, in which case the stationary distribution exists and is unique, and is given as $\pi(x) = 1/m_x$ for each state x.*

Remark 1. If the chain is null recurrent instead, then $\pi(x) = 0$ for all x and so π is not a probability distribution.

Remark 2. The convergence $\frac{1}{n} N_n(y) \to 1/m_y$ can be seen as the *ergodic* behavior of the chain. In fact, you can show the following result by (formally) decomposing $\sum_t f(X_t) = \sum_y f(y) \sum_t \mathbf{1}_y(X_t)$[1]:

Theorem 1.3 (Ergodic theorem for Markov chain). *If a Markov chain is irreducible and* ***positive recurrent****, and $f : S \to \mathbb{R}$ is a bounded function on the state space, then with the (unique) stationary distribution π,*

$$\lim_T \frac{1}{T} \sum_{t=0}^{T-1} f(X_t) = E_\pi(f(X)) \coloneqq \sum \pi(x) f(x)$$

with probability 1.

That is, the *time* average converges to the *spatial* average (weighted by the unique stationary distribution). This means, in some *averaged* sense, that the chain will traverse the state space repeatedly.

Remark 3. Recall that we need a few strong conditions to ensure the existence of $\lim_n P^n$:

Theorem 1.4 (The theorem from previous lectures). *Assuming*

- 1 *is a simple eigenvalue of* P
- *the associated left eigenvector can be normalized as a probability density*

[1] For a detailed proof, see Norris, *"Markov Chain"*, Thm. 1.10.2. For the case where the chain is not necessarily positive recurrent, see these slides by M. Merle.

- *all other eigenvalues of P have modulus less than* 1

then $\lim_n P^n$ *exists (and takes some form that you can compute explicitly).*

However, the theorem above implies $\lim_n \frac{1}{n}\sum P^n$, the limit of P^n in the average sense, *always* exists (without any strong condition), even if the (strong) limit $\lim P^n$ may not.

Example 1. Consider the symmetric random walk on $\mathbb{Z}$, with $P(x, x+1) = P(x, x-1) = 1/2$.

We have already shown (in previous tutorial) that it is irreducible and recurrent. However, it does not have a stationary distribution: such distribution π would satisfy $\pi(x) = \frac{1}{2}(\pi(x+1) + \pi(x-1))$[2] and so $\pi(x) = \pi(0)$ for all $x \in \mathbb{Z}$, which does not form a stationary *probability* distribution.

This implies that the mean recurrence time for every state is infinite, and the chain is null recurrent.

Example 2. Consider a fair coin with equal probability of head and tail $P(H) = P(T) = 1/2$. What is the expected number of tosses to see a head followed by a tail?

There are many approaches to handle this (e.g. one-step equations, martingales), but we will solve this as a question on the mean recurrence time. We can track the latest two tosses as a Markov chain with transition matrix

$$P = \begin{array}{c} \\ HH \\ HT \\ TH \\ TT \end{array}\begin{array}{c} \begin{array}{cccc} HH & HT & TH & TT \end{array} \\ \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix} \end{array}$$

If the chain starts at the state HT, then the expected number of tosses needed is exactly the mean recurrence time of the state HT, so we just need to find the stationary distribution, which is unique as the chain is irreducible and finite, thus positive recurrent.

It is easy to see that P has constant column sum 1, so $(1, 1, 1, 1)$ is a left eigenvector associated to the eigenvalue 1. This implies that unique stationary distribution is $\pi = (1/4, 1/4, 1/4, 1/4)$. In particular, $m_{HT} = \frac{1}{\pi(HT)} = 4$.

We can also compute the number of tosses to see two consecutive heads.

From the calculation, we have $m_{HH} = 4$. However, this is not the what we want: if the chain starts at HH, it has probability $P(H) = 1/2$ to reach HH again, with the first H from the (assumed) initial state. If we want to start tossing with a blank record, we need to ignore the effect of the initial state, which is the same as (artificially) assuming that the first toss is tail. This means that, if the expected tosses needed is t, by total probability

$$4 = 1 \cdot P(X_1 = HH \mid X_0 = HH) + (1 + t) \cdot P(X_1 \neq HH \mid X_0 = HH)$$

so $t = 6$.

You can generalize this approach to a longer pattern and a more general coin, which the analysis becomes more complicated, but the principle remains the same.

2 Detailed Balanced Equation

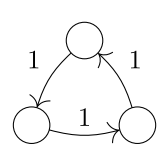
For a Markov chain with state space S and transition function P, the *detailed balance equation* is

$$\pi(i)P(i, j) = \pi(j)P(j, i)$$

for $i, j \in S$. That is, the probability flux from i to j is the same as the flux from j to i. We also say that P and π are in *detailed balance*.

Easy to see that, if a probability distribution π satisfies the detailed balance equation, then it must be a stationary distribution. However, *the converse does not hold*: the (unique) stationary distribution $(1/3, 1/3, 1/3)$ of the Markov chain with transition matrix

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$



does not satisfy the detailed balance equation. In fact, it is easy to show the following result:

[2]If we write $(\nabla f)(x) = f(x+1) - f(x)$ for function f, then this is (discrete) Laplace equation $\nabla^2\pi := \nabla(\nabla\pi) = 0$.

Lemma 2.1. *Let P be the transition function of an irreducible Markov chain. Suppose there exist states i, j such that $P(i,j) \neq 0$ and $P(j,i) = 0$. Then the detailed balance equation has no nontrivial solution.*

This means that, if there is an "irreversible transition" in the chain, the detailed balance equation gives you no information (and you need to work on the whole stationary equation).

Detailed balance is a *strong* condition and gives a simpler way to find the stationary distribution of a Markov chain, as it is (usually) easier to solve than the whole stationary equation.

Note however that detailed balance does not give any information on the uniqueness of stationary distribution, nor the (non)existence of stationary distribution if the equation has no probability density solution (for which you still need positive recurrence).

Example 3. Consider the birth-and-death chain on $S = \{ 0, \ldots, d \}$ (or $S = \mathbb{N}$), with $p_x = P(x, x+1)$, $r_x = P(x,x)$, $q_x = P(x, x-1)$, with appropriate definitions on the boundary points. We will assume that the chain is irreducible by assuming $p_x, q_x > 0$ for appropriate $x \in S$.

As noted in the textbook, the stationary distribution exists if and only if

$$\sum \pi_x < \infty \quad \text{with} \quad \pi_x = \prod_{i=0}^{x-1} \frac{p_i}{q_{i+1}} = \frac{p_0}{p_x} \frac{1}{\gamma_x}$$

adapting the notation $\gamma_x = \prod_{i=1}^{x} \frac{q_i}{p_i}$ defined for the absorption probability. In such case, the stationary distribution is unique and takes the form

$$\pi(x) = \pi_x / \sum_i \pi_i$$

I believe the stationary distribution is computed in the lecture for the special case where p_x, q_x are constants, but not for the general case. Here, we will find the stationary distribution with detailed balance equation.

The detailed balance equation reads

$$\pi(i) p_i = \pi(i+1) q_{i+1} \quad \text{so} \quad \pi(i+1) = \pi(i) \frac{p_i}{q_{i+1}}$$

Induction implies

$$\pi(x) = \pi(0) \prod_{j=0}^{x-1} \frac{p_i}{q_{i+1}} = \pi(0) \pi_x$$

and $\sum \pi(x) = \pi(0) \sum \pi_x$. If $\sum \pi_x < \infty$, we can obtain a normalized (so a probability density) solution with $\pi(0) = 1/\sum \pi_x$, which yields exactly the stationary distribution above.

Unfortunately, the detailed balance equation is not sufficient to show that no other stationary distribution exists, nor able to determine what happens if $\sum \pi_x = \infty$.[3]

Example 4 (MCMC, specifically Metropolis–Hastings). Let P be a transition function of an irreducible Markov chain on state space S, and π be a probability distribution on S that is positive everywhere. Assume that $P(i,j) \neq 0$ implies $P(j,i) \neq 0$.

Let us consider a new Markov chain on state space S with transition function

$$P_{\mathrm{MH}}(i,j) = \begin{cases} P(i,j)A(i,j) & \text{if } i \neq j \\ 1 - \sum_{k \neq i} P(i,k)A(i,k) & \text{if } i = j \end{cases} \quad \text{with} \quad A(i,j) = \min(1, \frac{\pi(j)P(j,i)}{\pi(i)P(i,j)})$$

Then the new chain is irreducible, and for two distinct states i, j,

- if $\pi(j)P(j,i) > \pi(i)P(i,j)$, we have $A(i,j) = 1$, $A(j,i) = \frac{\pi(i)P(i,j)}{\pi(j)P(j,i)}$, so

  $$\pi(i) P_{\mathrm{MH}}(i,j) = \pi(i)P(i,j) = \pi(j) \frac{\pi(i)P(i,j)}{\pi(j)P(j,i)} P(j,i) = \pi(j) P_{\mathrm{MH}}(j,i)$$

  The same holds if $\pi(j)P(j,i) < \pi(i)P(i,j)$.

- if $\pi(j)P(j,i) = \pi(i)P(i,j)$, then $P(i,j) = P(j,i) = 0$ or $A(i,j) = A(j,i) = 1$, in both cases

  $$\pi(j) P_{\mathrm{MH}}(j,i) = \pi(j)P(j,i)A(j,i) = \pi(i)P(i,j)A(i,j) = \pi(i) P_{\mathrm{MH}}(i,j)$$

so π satisfies the detailed balance equation for the new chain, and so is a stationary distribution.

Note that π appears only via the ratio $\pi(j)/\pi(i)$ and so is not needed to be normalized.

[3] Although, if you go through the proof in the textbook, you can see that for this chain the stationary equation *is equivalent to* the detailed balance equation (which is not a common phenomenon), so it suffices to consider *only* the latter.

3 (Optional) Proof of Bounded Convergence Theorem

I believe the following theorem is used in lecture, but only a scratch of proof is mentioned in the lecture slides:

Theorem 3.1 (BCT / Tannery's theorem). *Let $\{f_n : \mathbb{N} \to \mathbb{R}\}_n$ be a sequence of functions that satisfies*

- *(Uniformly bounded) there exists constant $M \geq 0$ such that $|\, f_n(x) \,| \leq M$ for all n, x*
- *(Pointwise convergence) for each x, $f(x) \coloneqq \lim_n f_n(x)$ exists (and is finite)*

Then for every $p : \mathbb{N} \to \mathbb{R}$ satisfying $p(x) \geq 0$ for all x and $\sum_x p(x) < \infty$, we have

$$\lim_n \sum_x p(x) f_n(x) = \sum_x p(x) f(x) \left(= \sum_x \lim_n p(x) f_n(x) \right)$$

That is, summation and (pointwise) limit can be interchanged.

This is in fact a special case of the more general *Lebesgue dominated convergence theorem* (which you can ask the lecturers of MATH4050, MATH5010, etc. for details). For completeness, we here give a (more detailed) proof of this result:

Proof. We first show that $\sum p(x)f(x)$ is well-defined (and finite). For each x, since $|\, f_n(x) \,| \leq M$ and $f_n(x) \to f(x)$, we have $|\, f(x) \,| \leq M$. So

$$\sum |\, p(x)f(x) \,| \leq \sum p(x) \,|\, f(x) \,| \leq M \sum p(x) < \infty$$

and thus $\sum p(x)f(x) \in \mathbb{R}$ is well-defined.

Let $\epsilon > 0$.

Since $\sum p(x) < \infty$, there exists $X \in \mathbb{N}$ such that $\sum_{x>X} p(x) < \frac{\epsilon}{3M}$. Then for all n,

$$\left| \sum_{x>X} p(x) f_n(x) \right| \leq \sum_{x>X} p(x) M < \epsilon/3$$

Similarly, $\left| \sum_{x>X} p(x)f(x) \right| < \epsilon/3$.

Also, as $f_n(x) \to f(x)$ for all x, there exists $N \in \mathbb{N}$ such that for all $n \geq N$ and $x \leq X$, $|\, f_n(x) - f(x) \,| < \epsilon/3$.

So, for all $n \geq N$,

$$\begin{aligned}
&\left| \sum_x p(x) f_n(x) - \sum_x p(x) f(x) \right| \\
&\leq \left| \sum_{x \leq X} p(x) f_n(x) - \sum_{x \leq X} p(x) f(x) \right| + \left| \sum_{x>X} p(x) f_n(x) \right| + \left| \sum_{x>X} p(x) f(x) \right| \\
&< \sum_{x \leq X} p(x) \,|\, f_n(x) - f(x) \,| + \epsilon/3 + \epsilon/3 \\
&\leq \frac{\epsilon}{3} \sum_x p(x) + \epsilon/3 + \epsilon/3 \\
&= \epsilon
\end{aligned}$$

This implies the convergence. □