Markov Chains

02 April, 2023

1 Basics

A **discrete-time stochastic process** is defined as a sequence of discrete random variables $\{X_n\}_{n \in \mathbb{N}_0}$, each taking values in a countable state space E

A discrete time stochastic process satisfying the **Markov condition** is called a **Markov Chain**

$$\Pr(X_n = j | X_{n-1} = i, X_{n-2} = x_{n-2}, \dots, X_0 = x_0) = \Pr(X_n = j | X_{n-1} = i)$$

for all $n \in \mathbb{N}$ and for all $x_0, \dots, x_{n-2}, i, j \in E$.

The Markov chain is **time-homogenous** if

$$\Pr(X_{n+1} = j | X_n = i) = \Pr(X_1 = j | X_0 = i)$$

for every $n \in \mathbb{N}_0$ and $i, j \in E$

1.1 Chapman-Kolmogorov (CK) equations

The **n-step** transition probability is

$$p_{ij}(n) = \Pr(X_{m+n} = j | X_m = i)$$

For a time homogenous Markov chain we have the **CK-equations**

$$p_{ij}(m+n) = \sum_{l \in E} p_{il}(m) p_{lj}(n)$$

where $m \in \mathbb{N} \cup \{0\}$ and $n \in \mathbb{N}$

The formula for **n-step transition matrix** follows:

$$P_{m+n} = P_m P_n$$

and in particular

$$P_n = P^n$$

1.2 First passage and hitting times

The **first passage time** is

$$T_j = \min\{n \in \mathbb{N} : X_n = j\}$$

In other words, $\{T_j = n\} = \{X_n = j, X_{n-1} \neq j, \cdots, X_1 \neq j\}$, if $X_n \neq j, \forall n \in \mathbb{N}$, then $T_j = \infty$.

The **first passage probability** is

$$f_{ij}(n) = \Pr(T_j = n | X_0 = i), n \in \mathbb{N}_0$$

from which the hitting probability follows

$$f_{ij} = \Pr(T_j < \infty | X_0 = i) = \sum_{n=0}^{\infty} f_{ij}(n)$$

With the **special case** being $f_{ij}(0) = 0$.

Decomposing the **n-step transition probability**

$$p_{ij}(n) = \sum_{l=1}^{n} f_{ij}(l) p_{jj}(n-l)$$

this is the same as starting from $l = 0$.

1.3 Generating Functions of Markov Chain

Recall the **probability generating function**

$$G_X(s) = \sum_{x=0}^{\infty} s^x \Pr(X = x)$$

where this holds on the support

$$\mathcal{S}_{\mathcal{X}} = \left\{ s \in \mathbb{R} : \sum_{x=0}^{\infty} |s|^x \Pr(X = x) < \infty \right\}$$

The generating functions here are

$$G_{p_{ij}(n)} = \sum_{n=0}^{\infty} p_{ij}(n) s^n$$
$$G_{f_{ij}(n)} = \sum_{n=0}^{\infty} f_{ij}(n) s^n$$

By arguing using equating coefficients and an identity, we have a **theorem**

$$G_{p_{ij}(n)} = \delta_{ij} + G_{f_{ij}(n)}(s) G_{p_{ij}(n)}$$

The identity used is decomposition of n-step transition probability.

2 Recurrence and Transience

A state j is **recurrent** if and only if

$$\sum_{n=1}^{\infty} p_{jj}(n) = \infty$$

A state j is **transient** if and only if

$$\sum_{n=1}^{\infty} p_{jj}(n) < \infty$$

Examples: Examples of transient, irreducible chains

The **number of periods** that the chain is in state j (or **number of visits** to j) is

$$N_j = \sum_{n=0}^{\infty} I_n(j)$$

where $I_n(j)$ is the indicator function taking value 1 if $X_n = j$ and 0 otherwise.

The **expected number of visits** to state j given $X_0 = j$ is

$$\mathbb{E}[N_j|X_0 = j] = \sum_{n=0}^{\infty} p_{jj}(n)$$

proof using generating functions:

Taking $s \to 1$ and using Abel's theorem, we can deduce...

2.1 Properties of recurrent/transient states

Theorem (Number of visits is geometric for transient states)

If j is transient, then

$$\Pr(N_j = n|X_0 = j) = f_{jj}^{n-1}(1 - f_{jj}), n \in \mathbb{N}$$

Let $i \neq j$, then

$$\Pr(N_j = n|X_0 = i) = \begin{cases} 1 - f_{ij} & n = 0 \\ f_{ij} f_{jj}^{n-1}(1 - f_{jj}) & n \geq 1 \end{cases}$$

Intuition is that the chain visits j for the first time and returns to it for $n - 1$ times, then leaves it. (note the N_j starts from 0)

Therefore, it follows that for $i \neq j$ by the mean of geometric distribution,

$$\mathbb{E}[N_j|X_0 = i] = \frac{f_{ij}}{1 - f_{jj}}$$

and

$$\mathbb{E}[N_j|X_0 = j] = \frac{1}{1 - f_{jj}}$$

Theorem(Unlikely to visit a transient state)

If j is transient, then

$$\lim_{n\to\infty} p_{ij}(n) = 0, \forall j \in E$$

Similar results hold for null recurrent states.

2.2 Mean recurrence time, null and positive recurrence

The **mean recurrence time** μ_j is

$$\mu_j = \mathbb{E}[T_j|X_0 = j] = \sum_{n=1}^{\infty} n f_{jj}(n)$$

where we recall that $\{T_j = n\} = \{X_n = j, X_{n-1} \neq j, \cdots, X_1 \neq j\}$.

Similarly, we can define the **mean first passage time** μ_{ij}:

$$\mu_{ij} = \mathbb{E}[T_j|X_0 = i] = \sum_{n=1}^{\infty} n f_{ij}(n)$$

those expectations can be finite or infinite; for transient states, they must be infinite.

For a recurrent state j, it is called **null recurrent** if $\mu_j = \infty$ and **positive recurrent** if $\mu_j < \infty$.

Theorem (unlikely to visit null recurrent state)

A state j is null recurrent, if and only if

$$\lim_{n\to\infty} p_{jj}(n) = 0$$

In addition if the above equation holds, then

$$\lim_{n\to\infty} p_{ij}(n) = 0, \forall i \neq j \in E$$

3 Aperiodicity and Ergodicity

The **period** of a state j is

$$d(j) = gcd\{n \in \mathbb{N} : p_{jj}(n) > 0\}$$

It is not necessarily true that $p_{jj}(d(j)) > 0$ (cf. Notes Pg. 36).

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix}$$

A state is **ergodic** if it is positive recurrent and aperiodic.

4 Communicating classes

- We say that a state j is **accessible** from state i if the chain can reach j at some time, written as $i \to j$.
- Two states i and j are **communicating** if there exists a state k such that $i \to k$ and $k \to j$, we write $i \leftrightarrow j$; this is an **equivalence relation**.
- If $i \neq j$, then $i \to j$ if and only if $f_{ij} > 0$.

4.1 Properties preserved by Communicating Classes

- Same period
- Same transience/recurrence
- Null recurrence

For a **set of states** C:

- C is **closed** if $\forall i \in C$, $j \notin C$, $p_{ij} = 0$
- C is **irreducible** if all states in the set communicate with each other

Therefore, an irreducible set of states share the same properties described above.

Theorem (Recurrence and closed)
If C is a communicating class of recurrent states, then C is closed.

Theorem (Stochastic matrix on closed states)
The stochastic matrix P restricted to a closed set of closed states C is still a stochastic matrix.

4.2 Decomposition of Chains

The state space can be partitioned into communicating classes.

$$E = T \cup \left(\bigcup_i C_i \right)$$

where T is the set of transient states and C_i's are irreducible closed sets of recurrent states.

4.3 Class Properties

The **classes** refer to communicating classes.

Theorem (Finite Chains have recurrent)
When state space is **finite**, at least one state is *recurrent* and all *recurrent* states are **positive**

Remark This combined with later results on stationarity makes a chain with finite state space particularly nice.

Remark It follows that there are no null recurrent states in a finite state space.

Theorem (Finite and closed)
If C is a finite, closed communicating class, then all states are positive recurrent.

Communicating class properties

Type of Class	Finite	Infinite
Closed	positive recurrent	positive/null recurrent, transient
Not closed	transient	transient

5 Gambler's Ruin

- Starting at state $i \in \{0, 1, \ldots, N\}$
- p of winning one unit and $1-p$ of losing one unit

- Assume successive games are independent

Question: What is the probability of reaching N before reaching 0?

Strategy: Use **first step analysis**:

- Define $V_i = \min\{n \in \mathbb{N}_0 : X_n = i\}$, the first time being present at state i, desired event is $V_N < V_0$
- Consider the conditional probability of starting at i and ending up at N

$$h_i = \Pr(V_N < V_0 \mid X_0 = i)$$

- Consider the first step

$$\begin{aligned} h_i &= \Pr(V_N < V_0 \mid X_1 = i+1, X_0 = i) \Pr(X_1 = i+1 \mid X_0 = i) \\ &\quad + \Pr(V_N < V_0 \mid X_1 = i-1, X_0 = i) \Pr(X_1 = i-1 \mid X_0 = i) \\ &= h_{i+1} p + h_{i-1}(1-p) \end{aligned}$$

Finish by solving this recurrence relation

$$h_i = \begin{cases} \dfrac{1-(q/p)^i}{1-(q/p)^N} & p \neq \dfrac{1}{2} \\ \dfrac{i}{N} & p = \dfrac{1}{2} \end{cases}$$

Stationarity

We are interested in the equilibrium states of a chain

- A **distribution** is a row vector λ with $\Sigma_j \lambda_j = 1$

- If $\lambda P = \lambda$ then it is called ***invariant***

5.1 Stationary distributions of irreducible chains

Theorem Every irreducible chain has a **stationary distribution** π if and only if all states are **positive recurrent**

- π is unique
- $\pi = \mu_i^{-1}$ the inverse of mean recurrence time

We first have some lemmas:

$$l_{ji}(n) = \Pr(X_n = i, T_j \geq n | X_0 = j)$$

being the probability that the chain reaches i in n steps without returning to j
Lemma (Decomposing the first hitting)

$$f_{jj}(m+n) = \sum_{i \in E, i \neq j} l_{ji}(m) f_{ij}(n)$$

from which $f_{jj}(m+n) \geq l_{ji}(m) f_{ij}(n)$ follows

Lemma (Formula for hitting) We also have the following recurrence relation for $l_{ji}(n+1)$

$$l_{ji}(n+1) = \sum_{r \in E, r \neq j} p_{ri} l_{jr}(n)$$

with $l_{ji}(1) = p_{ji}$.

Lemma (Existence): A positive recurrent chain has a stationary distribution.

Proof: (constructive)

- **(Step1 Construction)** Let $N_i(j)$ be the number of visits to state i before visiting state j (again); the sum of such numbers over i is equal to the hitting time T_j

- Define $\rho_i(j)$ to be the expected number of visits to the state i between two successive visits to state j (in this step the **recurrence** of the chain is used, as the T_j is finite with probability 1)

$$\begin{aligned} \rho_i(j) &= \mathbb{E}[N_i(j) | X_0 = j] \\ &= \sum_n \Pr(X_n = i, T_j \geq n | X_0 = j) \\ &= \sum_n l_{ji}(n) \end{aligned}$$

- Now the mean hitting time can be computed as

$$\begin{aligned} \mu_j &= \mathbb{E}\big[\sum_i N_i(j) | X_0 = j\big] \\ &= \sum_i \rho_i(j) \end{aligned}$$

- which can be written as sum of $\rho_i(j)$ by Tonelli and linearity of conditional expectation
- **(Step2 Finiteness)** Use a lemma to bound $\rho_i(j)$ so it's finite
- Namely write $\rho_i(j) = \sum_n l_{ji}(n)$ and bound using the fact that the chain is irreducible, so there exists $f_{ij}(n^*) > 0$, so $f_{jj}(m + n^*) \geq l_{ji}(m) f_{ij}(n^*)$
- **(Step3 Stationarity)** Use recurrence to show

$$\begin{aligned} \rho_i(j) &= \sum_n l_{ji}(n) \\ &= p_{ji} + \sum_{n=2} \sum_{r, r \neq j} p_{ri} l_{jr}(n-1) \\ &= p_{ji} \rho_j(j) + \sum_{n=1} \sum_{r \neq j} p_{ri} l_{jr}(n) \\ &= p_{ji} \rho_j(j) + \sum_{r, r \neq j} p_{ri} \sum_{n=1} l_{jr}(n) \\ &= \sum_r \rho_r(j) p_{ri} \end{aligned}$$

- which uses the fact that $\rho_j(j) = 1$
- This $\rho_i(j)$ does not necessarily give a probability vector when the chain is not positive recurrent.

- Now if the chain is positive recurrent, we have μ_j finite for every j, we have

$$\pi_i = \frac{\rho_i(j)}{\mu_j}$$

Therefore, we conclude that

Every **irreducible, recurrent** chain has a positive solution to the equation $\mathbf{x} = \mathbf{x}P$, which is unique up to a multiplicative constant (see PS2 for proof).
Moreover, the chain is

- positive recurrent if $\sum_i x_i < \infty$
- null recurrent if $\sum_i x_i = \infty$

Also, from the proof, we conclude 3 **identities**:

(Sum of expected visits)

$$\mu_j = \sum_i \rho_i(j)$$

(Sum of hitting prob)

$$\rho_i(j) = \sum_n l_{ji}(n)$$

(Sum of number of visits)

$$T_j = \sum_i N_i(j)$$

Lemma (Tail probability is expectation)
T is a nonnegative integer-valued random variable and A is an event with $\Pr(A) > 0$. Then

$$\mathbb{E}[T|A] = \sum_{n=0}^{\infty} \Pr(T \geq n|A)$$

Lemma If a stationary distribution exists, then the chain is positive recurrent and the distribution must be given by $\pi_i = \mu_i^{-1}$

proof: ...

In a **reducible chain**, the following results are useful:

- $\pi_i = 0$ for stationary distribution π if i is transient or null recurrent, so we can only compute the positive recurrent states and set the rest of π_i to 0
- A discrete time Markov chain with **finite** state space always has at least one stationary distribution.
- This distribution is **unique** unless it has two or more closed communicating classes.
- Every stationary distribution is a **linear combination** of the stationary distributions of the closed communicating classes, with coefficients adding up to 1.

5.2 Limiting Distribution

A distribution π is a **limiting distribution** of a chain if π satisfies

$$\lim_{n\to\infty} p_{ij}(n) = \pi_j$$

Theorem For an irreducible, aperiodic chain

$$\lim_{n\to\infty} p_{ij}(n) = \frac{1}{\mu_j}$$

It follows that for an irreducible, aperiodic, and positive recurrent state, the limiting distribution is its unique stationary distribution

$$\lim_{n\to\infty} p_{ij}(n) = \pi_j = \frac{1}{\mu_j}$$

Example of Chain with no limiting distribution

Consider the transistion matrix of two alternating states

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

the even and odd powers differ, but it has stationary distribution $\pi = (1/2, 1/2)$.

5.3 Ergodic Theorem

The **number of visits to i before time n** is defined as

$$V_i(n) = \sum_{k=0}^{n-1} \mathbf{1}_{\{X_k = i\}}$$

Theorem If a chain is irreducible, $V_i(n)/n$ denotes the proportion of time the chain spent in state i before time n

$$\Pr\left(\frac{V_i(n)}{n} \to \frac{1}{\mu_j} \text{ as } n \to \infty\right) = 1$$

5.4 Summary of properties of irreducible chains

1 Positive Recurrent

- Stationary distribution **exists** and is **unique**
- All mean recurrence times $\mu_j = \mathbb{E}[T_j | x_0 = j]$ are finite and $\pi_j = \dfrac{1}{\mu_j}$
- $V_i(n)/n \to \pi_i$
- If the chain is aperiodic, the limiting distribution is the stationary distribution

2 Null Recurrent

- All mean recurrence times are infinite
- No stationary distribution

- $V_i(n)/n \to 0$
- The limiting distribution is 0

3 Transient

- All mean recurrence times are infinite (any state is eventually never visited)
- No stationary distribution
- $V_i(n)/n \to 0$
- The limiting distribution is 0

6 Time reversibility

In this section, we assume the Markov chains are **irreducible and positive recurrent**, therefore there is a unique stationary distribution π.

The **reversed chain** for some $N \in \mathbb{N}$ is defined as

$$Y_n = X_{N-n}$$

Theorem (Reversed still Markov) The reversed chain is a Markov chain

$$\Pr(Y_{n+1} = j | Y_n = i) = \Pr(X_{N-n-1} = j | X_{N-n} = i) = \frac{\pi_j}{\pi_i} p_{ji}$$

A Markov chain (X_n) is called **time-reversible** if its transition matrix is the same as the transition matrix of its reversed chain.

Theorem A Markov chain is time-reversible if and only if

$$\pi_i p_{ij} = \pi_j p_{ji}$$

this condition is called **detailed balance**.

Theorem (Detailed balance implies positive recurrence)
For an irreducible chain, if **there is a vector** π such that the detailed balance equation holds for all i, j, then the chain is **time-reversible and positive recurrent** with stationary distribution π.

Proof: Note that the detailed balance conditions imply the chain has a stationary distribution, hence positive recurrent by previous theorems.