

Stochastics Processes

Labix

May 19, 2023

Abstract

Contents

1	Stochastic Processes	2
1.1	Markov Chains	2
1.2	Communicating Classes	3
1.3	Hitting Times	3
1.4	Strong Markov Property	4
1.5	Recurrence and Transience	4
2	Branching Process	7
2.1	Branching Process	7
3	Invariant Distributions	8
3.1	Basics of Invariant Distributions	8
3.2	Uniqueness of Invariant Distribution	9
3.3	Convergence to Equilibrium	11
3.4	Time Reversal and Invariant Distributions	11
3.5	The Ergodic Theorem	12

1 Stochastic Processes

1.1 Markov Chains

Definition 1.1.1: Stochastic Matrix

A matrix is a stochastic matrix if

- $0 \leq p_{ij} \leq 1$ for all $i, j \in I$
- For all $i \in I$, we have $\sum_{j \in I} p_{ij} = 1$

In particular, we use a stochastic matrix to represent the full set of probability of transitions between states.

Definition 1.1.2: Markov Chain

Let X_0, X_1, \dots, X_n be a sequence of random variables with common image $I = \{E_0, \dots, E_k\}$ with $X_0 \sim \lambda$ where λ is some fixed distribution. We say $(X_n)_{n \geq 0}$ is a markov chain if

$$\mathbb{P}(X_{n+1} = E_{n+1} | X_0 = E_0, \dots, X_n = E_n) = \mathbb{P}(X_{n+1} = E_{n+1} | X_n = E_n)$$

for all $n \in \mathbb{N}$.

In other words, Markov chains are a sequence of random variables where the next step does not depends on what states you visited previously, but only on the state now. There is an important property that makes studying Markov chains worth while.

Theorem 1.1.3: Markov Property

Let $(X_n)_{n \geq 0}$ be a markov chain. Suppose that $X_m = E_m$ is given. Then $(X_{m+n})_{n \geq 0}$ is a Markov chain.

Notice that the transition probabbility $\mathbb{P}(X_{n+1} = j | X_n = i)$ still depedns on i, j, n . We further restrict out study of Markov chains to the following type.

Definition 1.1.4: Homogenous Markov Chains

We say that a Markov chain $(X_n)_{n \geq 0}$ is homogenous if

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

This means that homogenous Markov chains are time independent.

Definition 1.1.5: Transition Matrix

Define the transition matrix of a homogenous Markov chain with state space $\{1, \dots, n\}$ to be

$$P = \begin{pmatrix} \mathbb{P}(X_1 = 1 | X_0 = 1) & \cdots & \mathbb{P}(X_1 = n | X_0 = 1) \\ \vdots & \ddots & \vdots \\ \mathbb{P}(X_1 = 1 | X_0 = n) & \cdots & \mathbb{P}(X_1 = n | X_0 = n) \end{pmatrix}$$

Lemma 1.1.6

The transition matrix P of a homogenous Markov chain is a Stochastic matrix.

Theorem 1.1.7

Let $(X_n)_{n \geq 0}$ be Markov. Then for all $m, n \geq 0$, we have

- $\mathbb{P}(X_n = j | X_0 = i) = \mathbb{P}(X_1 = j | X_0 = i)^n$
- $\mathbb{P}(X_n = j) = \sum_{i \in I} \mathbb{P}(X_0 = i) \mathbb{P}(X_n = j | X_0 = i)$

This theorem is saying that $\mathbb{P}(X_n = j | X_0 = i)$ is equal to the i, j th entry of the matrix P^n if P is the transition matrix. To find out the total probability of reaching the j th state at the n th step regardless of the starting point, you sum up the $\mathbb{P}(X_n = j | X_0 = i)$ multiplying the probability that you start at state i .

We often like to create new Markov chains from old (at least from exams). If we want to show that a sequence $(X_n)_{n \geq 0}$ of random variables is Markov, try and write $X_{n+1} = X_n + \text{some term only related to } n + 1$.

1.2 Communicating Classes**Definition 1.2.1: Talks and Communicates**

Let i, j be states. We say that i talks to j which is $i \rightarrow j$ if there exists $n \in \mathbb{N}$ such that

$$\mathbb{P}(X_n = j | X_0 = i) > 0$$

We say that two states i and j communicate if $i \leftrightarrow j$.

Definition 1.2.2: Communicating Class

Let $C \subseteq I$. C is a communicating class if $\forall i, j \in C, i \leftrightarrow j$ and $\forall i \in C$ and $\forall k \in I \setminus C, i \not\rightarrow k$

Definition 1.2.3: Closed and Absorbing

We say that a class is closed if no states in the class talks to any states outside of that class.
We say that a class is absorbing if it forms a closed class by itself.

Definition 1.2.4: Irreducible Markov Chains

A Markov chain is said to be irreducible if for all $i, j \in I, i \leftrightarrow j$.

1.3 Hitting Times**Definition 1.3.1: Stopping Times**

Let $(X_n)_{n \geq 0}$ be Markov. Then a random variable

$$T : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$$

is a stopping time if for all $n \geq 0$, the event $\{T = n\}$ depends only on X_0, \dots, X_n .

Definition 1.3.2: Hitting Times

Let $(X_n)_{n \geq 0}$ be Markov. The hitting time of a subset $A \subseteq I$ is a random variable

$$H_A : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$$

defined by

$$H_A(\omega) = \inf\{n \geq 0 | X_n(\omega) \in A\}$$

We use

$$h_i^A = \mathbb{P}(H^A < \infty | X_0 = i)$$

to denote the corresponding probability. We use

$$k_i^A = E_i[H_A] = \sum_{k=0}^{\infty} k \mathbb{P}(H_A = k | X_0 = i) + \infty \mathbb{P}(H_A = \infty | X_0 = i)$$

to denote the expectation.

Proposition 1.3.3

We can find the probabilities of all hitting times by solving the system of linear equations

$$\begin{cases} \mathbb{P}(H_A < \infty | X_0 = i) = 1 & i \in A \\ \mathbb{P}(H_A < \infty | X_0 = i) = \sum_{j \in I} \mathbb{P}(X_1 = j | X_0 = i) \mathbb{P}(H_A < \infty | X_0 = j) & i \notin A \end{cases}$$

Proposition 1.3.4

We can find the expected number of hitting times by solving the system of linear equations

$$\begin{cases} E_i[H_A] = 0 & i \in A \\ E_i[H_A] = 1 + \sum_{j \notin A} \mathbb{P}(X_1 = j | X_0 = i) E_j[H_A] & i \notin A \end{cases}$$

1.4 Strong Markov Property

Theorem 1.4.1: Strong Markov Property

Let $(X_n)_{n \geq 0}$ be Markov(λ, P) and T a stopping time of $(X_n)_{n \geq 0}$. Then conditional on both $\{X_T = i\}$ and $\{T < \infty\}$, we have $(X_{T+n})_{n \geq 0}$ is Markov(δ_i, P) and independent of X_0, \dots, X_T .

1.5 Recurrence and Transience

Definition 1.5.1: Recurrence and Transience

A state $i \in I$ is recurrent if $\mathbb{P}_i(X_n = i \text{ for infinitely many } n) = 1$. A state $i \in I$ is transient if $\mathbb{P}_i(X_n = i \text{ for infinitely many } n) = 0$.

Definition 1.5.2: k -th Passage Time

The first passage time of state i is a stopping time such that

$$T_i(\omega) = \inf\{n \geq 1 | X_n(\omega) = i\}$$

The k -th passage time is a stopping time such that

$$T_i^{(k)}(\omega) = \inf\{n \geq T_i^{(k-1)} | X_n(\omega) = i\}$$

The k -th excursion time is defined by

$$S_i^{(k)} = \begin{cases} T_i^{(k)} - T_i^{(k-1)} & \text{if } T_i^{(k-1)} < \infty \\ \infty & \text{otherwise} \end{cases}$$

Intuitively, $T_i^{(k)}$ outputs the time it takes for the Markov chain to reach state i for the k th time and S_i^k outputs the number of steps taken between consecutive visits.

Lemma 1.5.3

For $k = 2, 3, \dots$ and if $T_i^{(k-1)} < \infty$, then $S_i^{(k)}$ is independent of $\{X_m : m \leq T_i^{(k-1)}\}$ and

$$\mathbb{P}(S_i^{(k)} = n | T_i^{(k-1)} < \infty) = \mathbb{P}_i(T_i = n)$$

Definition 1.5.4: Visit Counting Function

Let $i \in I$ be a state. Define

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

the number of visits ever to state i . We use V_i to denote $\sum_{k=0}^{\infty} 1_{\{X_k=i\}}$.

Lemma 1.5.5

If V_i counts the number of visits to state i , then

$$E_i(V_i) = \sum_{k=0}^{\infty} P_{ii}^{(k)}$$

Lemma 1.5.6

For $k = 0, 1, 2, \dots$, we have

$$\mathbb{P}_i(V_i > k) = (\mathbb{P}_i(T_i < \infty))^k$$

This lemma means that V_i asserts a geometric distribution.

Theorem 1.5.7: Characteristic of Recurrent and Transient States

Let $(X_n)_{n \geq 0}$ be a Markov chain. Let T_i be the first passage time of state i . If $\mathbb{P}_i(T_i < \infty) = 1$, then i is recurrent and $\sum_{k=1}^{\infty} \mathbb{P}(X_1 = k | X_0 = k)^n = \infty$. If $\mathbb{P}_i(T_i < \infty) < 1$, then i is transient and $\sum_{k=1}^{\infty} \mathbb{P}(X_1 = k | X_0 = k)^n < \infty$.

Theorem 1.5.8

Let C be a communicating class. Then either all states in C are transient or all are recurrent.

Theorem 1.5.9

Let $C \subset I$ be a class of a Markov chain. Then

- Every recurrent class is closed
- Every finite closed class is recurrent

Theorem 1.5.10

If P is irreducible and recurrent then for all $j \in I$,

$$\mathbb{P}(T_j < \infty) = 1$$

Definition 1.5.11: Component Independent Simple Random Walk on \mathbb{Z}^d

A component independent simple random walk on \mathbb{Z}^d for $d \in \mathbb{N}$ is defined as

$$X_{n+1} = X_n + Z_{n+1}$$

where $Z_{n+1} = (Z_{n+1}^1, \dots, Z_{n+1}^d)$ with

$$Z_m^j = \begin{cases} 1 & \text{with probability } p_j \\ -1 & \text{with probability } q_j \end{cases}$$

where Z_m^j are independent for all j, m .

Proposition 1.5.12

Let $(X_n)_{n \geq 0}$ be a CISRW on \mathbb{Z}^d . If there exists $j \in \{1, \dots, d\}$ such that $p_j \neq \frac{1}{2}$ then (X_n) is transient.

Proposition 1.5.13

Let $(X_n)_{n \geq 0}$ be a CISRW on \mathbb{Z}^d and $p_j = q_j = \frac{1}{2}$ for all j . Then

- If $d \leq 2$ then all states are recurrent
- If $d \geq 3$ then all states are transient

2 Branching Process

2.1 Branching Process

Definition 2.1.1: Branching Process

Let $(X_n)_{n \geq 0}$ be the number of individuals in a population at time n where $X_n \in \{0\} \cup \mathbb{N}$. At every time step each individual in the population gives birth to a random number of offspring. Thus X_{n+1} is defined to be

$$X_{n+1} = \sum_{k=1}^{X_n} Z_k^n$$

where $Z_k^n \sim Z$ and $\mathbb{P}(Z \geq 0) = 1$. Then $(X_n)_{n \geq 0}$ is said to be a branching process.

Proposition 2.1.2

Define $G(s) = E[s^Z]$ and $F_{n+1}(s) = E[S^{X_{n+1}} | X_0 = 1]$ for $s \in (0, 1)$. Then

$$F_n(s) = G(F_{n-1}(s))$$

Proposition 2.1.3

For a braching process $(X_n)_{n \geq 0}$ with off-spring distribution Z where $E[Z] = \mu$, we have

$$E[X_n] = \mu^n$$

Theorem 2.1.4: Extinction Probability

The extinction probability is the smallest non-negative root of $G(\alpha) = \alpha$.

Theorem 2.1.5

Suppose that $G(0) > 0$, $\mu = E[Z] = G'(1)$. Then $\mu \leq 1$ implies certain extinction. $\mu > 1$ implies uncertain extinction.

3 Invariant Distributions

3.1 Basics of Invariant Distributions

Definition 3.1.1: Invariant Distribution

A measure $\lambda = (\lambda_i : i \in I)$ with non-negative entries is said to be an invariant measure of P if

- $\lambda P = P$
- $\pi_i \in [0, 1]$ for all i

If in addition it satisfies $\sum_{i \in I} \pi_i = 1$ then λ is said to be invariant distribution.

The following theorem shows that after applying a Markov process to an invariant distribution, the new distribution will be the same as the old one. Notice here that right multiplication of a distribution gives the next step on the Markov chain instead of the usual left multiplication.

Theorem 3.1.2

Let $(X_n)_{n \geq 0}$ be Markov(π, P), where π is an invariant distribution for P . Then $\forall m \geq 0$, $(X_{n+m})_{n \geq 0}$ is Markov(π, P).

Proof. By the Markov property, the new sequence will be Markov. But we do not yet know its distribution. We have that

$$\begin{aligned} \mathbb{P}(X_m = i) &= (\pi P^m)_i && \text{(this denotes the } i\text{th entry of } P) \\ &= (\pi P P^{m-1})_i \\ &= (\pi P^{m-1})_i \end{aligned}$$

Thus by induction, we see that $\mathbb{P}(X_m = i) = \pi_i$ thus we are done. \square

Theorem 3.1.3

Let I be finite. Suppose that for some $i \in I$,

$$\lim_{n \rightarrow \infty} P_{ij}^n = \pi_j$$

for all $j \in I$. Then $\pi = (\pi_j : j \in I)$ is an invariant distribution.

Let us look at an example.

Example 3.1.4: Gene Mutation

Recall the gene mutation example that has transition matrix $P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$. Recall that

$$P^n = \begin{pmatrix} \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n & \frac{\alpha}{\alpha+\beta} - \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n \\ ? & ? \end{pmatrix}$$

As $n \rightarrow \infty$, observe that the first entry tends to $\frac{\beta}{\alpha+\beta}$ the second entry tends to $\frac{\alpha}{\alpha+\beta}$. This means that $\pi = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$ is an invariant distribution for the gene mutation model.

We can also find the invariant distribution directly: Let $\pi = (\pi_1, \pi_2)$ be an invariant distribu-

tion for P . We have that

$$\begin{aligned}\pi P &= \pi \\ \begin{pmatrix} \pi_1(1-\alpha) + \pi_2\beta & \pi_1\alpha + \pi_2(1-\beta) \end{pmatrix} &= \begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix} \\ \begin{pmatrix} \pi_2\beta - \pi_1\alpha & \pi_1\alpha - \pi_2\beta \end{pmatrix} &= 0\end{aligned}$$

Notice that these two equations mean the same thing. (In general if you have an $n \times n$ transition matrix, only the first $n - 1$ equations will be useful and the last one will be redundant.) This is why we need to use the fact that $\pi_1 + \pi_2 = 1$ as our final equation.

Now solving it, we have that $\pi = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)$ which is the exact same answer as the one given in the first method.

A third way to think about it is that notice that $P^T \pi^T = \pi^T$ is the equation that we are attempting to solve. Recall that 1 is necessarily an eigenvector of a transition matrix which means that this equation really is just a question of finding eigenvectors from the eigenvalue 1.

The above examples gives a lot of practical information when calculating the invariant distribution of a transition matrix.

Observe that in general, solving the equation $\pi P = \pi$ does not give a unique invariant distribution. We require additional constraints.

3.2 Uniqueness of Invariant Distribution

Definition 3.2.1: Expected Time Spent

Define the expected time spent in state i in between visits to state k by

$$\gamma_i^k = E_k \left(\sum_{n=0}^{T_k-1} \mathcal{X}_{X_n=i} \right)$$

This uses k as a reference, and as we keep going back to k , this records how much time we have spent in i in the process of leaving and returning to k .

Theorem 3.2.2

Let P be irreducible and recurrent. Then

- $\gamma_k^k = 1$
- $\gamma^k = (\gamma_i^k : i \in I)$ is an invariant measure satisfying $\gamma^k P = \gamma^k$
- $0 < \gamma_i^k < \infty \forall i \in I$

Proof.

- Trivial
- Notice that $\{T_k \geq n\} = \{T_k < n\}^c$ depends only on X_0, \dots, X_{n-1} , thus by the Markov property at time $n - 1$ we have

$$\mathbb{P}_k(X_{n-1} = i, X_n = j, T_k \geq n) = \mathbb{P}_k(X_{n-1} = i) \mathbb{P}_k(X_n = j, T_k \geq n | X_{n-1} = i)$$

- Since P is irreducible, for all $i \in I$, there exists $m, n \in \mathbb{N}$ such that $P_{ik}^n > 0$ and $P_{kj}^m > 0$. Then

$$\begin{aligned}\gamma_i^k &= \sum_{j \in I} \gamma_j^k P_{ji}^m \\ &\geq \gamma_k^k P_{ki}^m > 0 \\ &> 0\end{aligned}$$

We also have that

$$\begin{aligned}\gamma_i^k &= \sum_{n=1}^{T_k} \mathcal{X}_{X_n=i} \\ &< \infty\end{aligned}$$

Since $T_k < \infty$.

□

Theorem 3.2.3

Let P be irreducible and λ an invariant measure and $\lambda_k = 1$ for some k . Then $\lambda \geq \gamma^k$. If in addition we also have that P is recurrent, then $\lambda = \gamma^k$.

Proof.

□

This theorem asserts that as long as P is irreducible and recurrent, then any invariant measure is exactly equal to the γ^k we defined, which means that there is really only one invariant measure up to rescaling. In other words, irreducibility and recurrence guarantees uniqueness of the invariant measure. However, notice that this still does not assure that we can normalize the invariant measure to become an invariant distribution.

Definition 3.2.4: Positive Recurrent

A state $i \in I$ is positive recurrent if it is recurrent and

$$E_i[T_i] < \infty$$

A state which is not positive recurrent but is recurrent is called null recurrent.

Theorem 3.2.5

Let P be irreducible. Then the following are equivalent.

- Every state is positive recurrent
- Some state i is positive recurrent
- P has an invariant distribution π . Moreover, $\pi_i = \frac{1}{E_i[T_i]}$

Essentially, combining these conditions, we required that the transition matrix P is irreducible and positive recurrent in order for an invariant distribution to exist and is unique.

3.3 Convergence to Equilibrium

Definition 3.3.1: Periodicity

Let $i \in I$ be a state. Define the periodicity of i to be

$$\gcd\{n \geq 0 \mid P_{ii}^n > 0\}$$

A state i is said to be aperiodic if $d = 1$.

Proposition 3.3.2: L

Let $i \in I$ be a state. Then i is aperiodic if and only if there exists $N \in \mathbb{N}$ such that $P_{ii}^n > 0$ for all $n \geq N$.

Lemma 3.3.3

Suppose that a Markov chain is irreducible and $i \in I$ is aperiodic. Then for all $j, k \in I$, there exists $M \in \mathbb{N}$ such that

$$P_{jk}^m > 0$$

for all $m > M$. In other words, aperiodicity is a class property.

Theorem 3.3.4: Convergence to Equilibrium

Let P be irreducible and aperiodic. Suppose that P has an invariant distribution π . Let λ be any distribution. If $(X_n)_{n \geq 0}$ is Markov(λ, P), then we have

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = j) = \pi_j$$

In other words, $\lim_{n \rightarrow \infty} \mathbb{P}_{ij}^n = \pi_j$ for all $i, j \in I$.

3.4 Time Reversal and Invariant Distributions

Theorem 3.4.1

Let P be irreducible and have invariant distribution π . Suppose that $(X_n)_{0 \leq n \leq N}$ is Markov(π, P). Define $Y_n = X_{N-n}$ for each N . Then $(Y_n)_{0 \leq n \leq N}$ is Markov(π, P') where P' is given by

$$\pi_j P'_{ji} = \pi_i P_{ij}$$

and P' is also irreducible with invariant distribution π .

Definition 3.4.2: Time Reversal

The above sequence of random variables $(Y_n)_{0 \leq n \leq N}$ is called the time reversal of $(X_n)_{0 \leq n \leq N}$.

Definition 3.4.3: Reversible

If the time reversal sequence of random variables $(Y_n)_{0 \leq n \leq N}$ of $(X_n)_{0 \leq n \leq N}$ are both Markov(λ, P), then we say that $(X_n)_{0 \leq n \leq N}$ is reversible.

Definition 3.4.4: Detailed Balance

A stochastic matrix P and a measure λ is said to be in detailed balance if

$$\lambda_i p_{ij} = \lambda_j p_{ji}$$

for all $i, j \in I$.

Lemma 3.4.5

If P and λ are in detailed balance then λ is an invariant measure for P .

Theorem 3.4.6

Let P be an irreducible stochastic matrix. Let λ be a distribution. Let $(X_n)_{n \geq 0}$ is Markov(λ, P). Then $(X_n)_{n \geq 0}$ is reversible if and only if P and λ are in detailed balance.

This provides a good way to check whether a transition matrix and a measure is an invariant distribution.

3.5 The Ergodic Theorem

Recall that $V_i(n)$ is the visit counting function.

Theorem 3.5.1: The Ergodic Theorem

Let P be irreducible and let λ be any distribution. If $(X_n)_{n \geq 0}$ is Markov(λ, P) then

$$\mathbb{P} \left(\lim_{n \rightarrow \infty} \frac{V_i(n)}{n} = \frac{1}{E(T_i)} \right) = 1$$

Corollary 3.5.2

Let $(X_n)_{n \geq 0}$ be a positive recurrent Markov chain. Then for any $f : I \rightarrow \mathbb{R}$ such that $E_\pi(|f(x)|) < \infty$, we have

$$\mathbb{P} \left(\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \rightarrow E_\pi(f(x)) \right) = 1$$

Notice that $E_\pi(f(x)) = \sum_{i \in I} \pi_i f(i)$ in the corollary.

We wish that given an invariant distribution, we can construct a Markov chain so that we can use the Ergodic theorem to get estimates of $E_\pi(f(x))$.

Theorem 3.5.3: Markov Chain Monte Carlo Process

Describe a procedure as follows.

Step 1: Define $X_0 = i \in I$.

Step 2: Set $m = 0$, $Y = i$, N to be something big.

Step 3: While $m \leq N$,

1. Sample $Y \sim q_{X_m}$
2. Calculate $\alpha_{X_m, Y} = \min \left\{ 1, \frac{\pi_Y q_{Y, X_m}}{\pi_{X_m} q_{X_m, Y}} \right\}$, the acceptance probability
3. Sample $U \sim \text{Uniform}(0, 1)$
4. If $(U < \alpha_{X_m, Y})$, set $X_{m+1} = Y$. Else set $X_{m+1} = X_m$.

Step 4: Output $\{X_0, \dots, X_N\}$.