

Definition 1 (Markov Property). Let S be a finite or countable set (state space). Let $\{X_0, X_1, \dots\}$ be a sequence of random variables whose ranges are in S . The sequence is said to satisfy the Markov property if

$$\mathbf{P}\{X_{n+1} = j \mid X_0 = i_0, \dots, X_n = i_n\} = \mathbf{P}\{X_{n+1} = j \mid X_n = i_n\} = P_{ij}$$

for all $i_0, \dots, i_n, j \in S$ and $n \in \mathbb{N}$. The probability P_{ij} is called the transition probability from i to j . The matrix P whose (i, j) -entry equals P_{ij} is called the transition matrix.

Definition 2 (Markov Chain). A sequence of random variables $\{X_t\}_{t=0}^\infty$ is a Markov chain if it satisfies the Markov property.

Definition 3 (Initial Distribution). Let $\{X_t\}$ be a Markov chain. The distribution of X_0 is called the initial distribution.

1 High-Order Transitions

Definition 4 (High-Order Transition Probabilities). The n -th order transition probability from i to j is defined as

$$P_{ij}^{(n)} := \mathbf{P}\{X_{n+m} = j \mid X_m = i\}.$$

Theorem 1 (Chapman-Kolmogorov's Identity). For any $n, m \in \mathbb{N}$, the $(n+m)$ -th order transition probability can be decomposed thus:

$$P_{ij}^{(n+m)} = \sum_{k \in S} P_{ik}^{(n)} P_{kj}^{(m)}.$$

2 Classification of States

Definition 5 (Persistent/Transient). Define

$$f_{ij} := \mathbf{P}\{\exists n \geq 1 \text{ s.t. } X_n = j \mid X_0 = i\}.$$

A state $i \in S$ is said to be persistent if $f_{ii} = 1$. A state $i \in S$ is said to be transient if $f_{ii} < 1$.

Remark 1. The interpretation of f_{ij} is “the probability that j will be visited at some point in time starting from i .”

Theorem 2. *The following statements are equivalent:*

- (i) *State $i \in S$ is persistent.*
- (ii) $\mathbf{P}\{X_n = i \text{ i.o.} \mid X_0 = i\} = 1.$
- (iii) $\sum_{n \geq 1} P_{ii}^{(n)} = +\infty.$

The following statements are equivalent:

- (i) *State $i \in S$ is transient.*
- (ii) $\mathbf{P}\{X_n = i \text{ i.o.} \mid X_0 = i\} = 0.$
- (iii) $\sum_{n \geq 1} P_{ii}^{(n)} < \infty.$

3 Classification of Markov Chains

Definition 6 (Irreducible). *A Markov chain is called irreducible if $\forall i, j \in S$ $\exists n \in \mathbb{N}$ such that $P_{ij}^{(n)} > 0$.*

Theorem 3. *Assume that a Markov chain is irreducible, then either all states are transient, or all states are persistent.*

Corollary 3.1. *If S is finite and irreducible, then the Markov chain is always persistent.*

4 Stationary Distribution

Definition 7 (Stationary Distribution). *Let P_{ij} be transition probabilities. A distribution $\pi = \{\pi_i\}_{i \in S}$ is said to be a stationary distribution of a Markov chain if $\forall i, j \in S$,*

$$\sum_{i \in S} \pi_i P_{ij} = \pi_j.$$

Remark 2. *If S is finite, then we can let $\pi = (\pi_1, \dots, \pi_{|S|})$ and rewrite the condition as $\pi P = \pi$ where P is the transition matrix.*

Definition 8 (Period). *A state i has period $t = \gcd\{n \geq 1 \mid P_{ii}^{(n)} > 0\}$.*

Definition 9 (Aperiodic). *A Markov chain is said to be aperiodic if all states have period one.*

Definition 10. *Suppose that a Markov chain is irreducible and aperiodic. Also suppose that a stationary distribution $\{\pi_i\}$ exists. Then,*

- (i) The chain is persistent.
- (ii) The limit of high-order probabilities converge:

$$\lim_{n \rightarrow \infty} P_{ij}^{(n)} = \pi_j \quad \forall i, j \in S.$$

- (iii) All π_i are positive.
- (iv) The stationary distribution is unique.

Remark 3. In this theorem, the effect of the initial distribution wears off as $n \rightarrow \infty$ if a stationary distribution exists.

Theorem 4. If an irreducible and aperiodic Markov chain has no stationary distribution, then

$$\lim_{n \rightarrow \infty} P_{ij}^{(n)} = 0 \text{ as } n \rightarrow \infty$$

5 Markov Chain Monte Carlo

5.1 Metropolis-Hastings Algorithm

Let π be a distribution. We want to construct a Markov chain such that its stationary distribution is π . That is, we want to find transition probabilities P_{ij} 's such that

$$\sum_i \pi(i) P_{ij} = \pi(j) \quad \forall j. \quad (\text{Global Balance Equation})$$

Since the global balance equation requires a summation over all states i , it is not easy to construct. However, we can find P_{ij} 's such that

$$\pi(i) P_{ij} = \pi(j) P_{ji} \quad \forall i, j. \quad (\text{Detailed Balance Equation})$$

Notice that detailed balance equation implies global balance equation since

$$\sum_i \pi(i) P_{ij} = \sum_i \pi(j) P_{ji} = \pi(j) \sum_i P_{ji} = \pi(j).$$

Since the detailed balance equation is only a requirement for each pair of states i and j , it is much more easy to construct the required P_{ij} and P_{ji} . This motivates the Metropolis-Hastings algorithm.

Theorem 5 (Metropolis-Hastings Algorithm). Given a target distribution π , construct a Markov chain as follows. Arbitrarily fix an initial state I_0 .

1. Generate $J \sim \{P_{ij}\}$ where $\{P_{ij}\}$ is an arbitrary distribution (transition probabilities) with the same support as π . Let j be the generated state. That is,

$$\mathbf{P}\{J = j \mid I_n = i\} = P_{ij}.$$

Distribution $\{P_{ij}\}$ is called the proposal distribution.

2. Set the value of I_{n+1} according to the following rule:

$$I_{n+1} = \begin{cases} j & \text{with probability } \alpha(i, j) \\ i & \text{otherwise} \end{cases}$$

where

$$\alpha(i, j) = \min \left\{ \frac{\pi(j)P_{ji}}{\pi(i)P_{ij}}, 1 \right\}.$$

Then, the Markov chain $\{I_n\}$ has stationary distribution π .

Remark 4. We can also summarize the result thus: The transition probabilities $\{\tilde{P}_{ij}\}$ where $\tilde{P}_{ij} := \alpha(i, j)P_{ij}$ has stationary distribution π .

Remark 5. In many applications, the proposal distribution is not drawn from a chain, but from an iid distribution. Thus, the ratio $(\pi(i)P_{ij})/(\pi(j)P_{ji})$ simplifies to $(\pi(i)f(i))/(\pi(j)f(j))$ with f being the proposal density.

Proof. (of Theorem 5) We only have to show that the constructed transition probabilities satisfies the detailed balance equation. Let i, j be an arbitrary pair of states. We have $\tilde{P}_{ij} = \alpha(i, j)P_{ij}$. Consider $\pi(i)$:

$$\pi(i)\tilde{P}_{ij} = \pi(i) \min \left\{ \frac{\pi(j)P_{ji}}{\pi(i)P_{ij}}, 1 \right\} P_{ij} = \begin{cases} \pi(j)P_{ji} & \text{if } \pi(j)P_{ji} \leq \pi(i)P_{ij} \\ \pi(i)P_{ij} & \text{if } \pi(j)P_{ji} \geq \pi(i)P_{ij} \end{cases}.$$

On the other hand, consider $\pi(j)$:

$$\pi(j)\tilde{P}_{ji} = \pi(j) \min \left\{ \frac{\pi(i)P_{ij}}{\pi(j)P_{ji}}, 1 \right\} P_{ji} = \begin{cases} \pi(i)P_{ij} & \text{if } \pi(i)P_{ij} \leq \pi(j)P_{ji} \\ \pi(j)P_{ji} & \text{if } \pi(i)P_{ij} \geq \pi(j)P_{ji} \end{cases}.$$

Therefore, the detailed balance equation is satisfied. \square

5.2 Gibbs Sampling

Gibbs sampling is a variation of the Metropolis-Hastings algorithm that makes sampling high-dimensional distributions more efficient.

For illustration, let $\pi(a, b)$ be a distribution on $A \times B$ for $a \in A$ and $b \in B$. Following the Metropolis-Hasting algorithm, we can view the pair (a, b) as a single state and perform the sampling. However, if $\pi(a | b)$ and $\pi(b | a)$ are easy to sample from, then we can sample a and b in alternation.

Theorem 6 (Gibbs Sampling). *Given a target distribution $\pi(a, b)$, construct a Markov chain as follows. Arbitrarily fix an initial state (a_0, b_0) . In the iteration at time t , perform the following:*

1. Generate a_{t+1} from $\pi(a | b_t)$.
2. Generate b_{t+1} from $\pi(b | a_{t+1})$

Then, Markov Chain $\{(a_t, b_t)\}_t$ has stationary distribution $\pi(a, b)$.

Proof. Following the proof of Theorem 5, note that $\pi(a | b)$ and $\pi(b | a)$ are proposal distributions. Thus, we only have to show that under these proposal distributions, we always accept the proposed state, i.e., the $\alpha(i, j)$ in the proof of Theorem 5 is always 1.

Let (a, b) be the current state and let $a' \sim \pi(a | b)$. Consider the following:

$$\begin{aligned} \alpha((a, b), (a', b)) &= \min \left\{ \frac{\pi(a', b) \pi(a | b)}{\pi(a, b) \pi(a' | b)}, 1 \right\} \\ &= \min \left\{ \frac{\pi(a' | b) \pi(b) \pi(a | b)}{\pi(a | b) \pi(b) \pi(a' | b)}, 1 \right\} = 1. \end{aligned}$$

Similarly, we can also show that the proposed $b' \sim \pi(b | a)$ is also always accepted. \square

Remark 6. *Something remarkable about Gibbs sampling is that there is no “rejection,” that is, we always “accept” the proposed state through clever choice of proposal distribution $\pi(a | b)$ and $\pi(b | a)$. In practice, we often know how to sample from $\pi(a | b)$ and $\pi(b | a)$ directly, but do not know the joint distribution $\pi(a, b)$. Thus, Gibbs sampling provides a very efficient way of sampling $\pi(a, b)$.*