Markov Chain Monte Carlo Ⅰ

Woojoo Lee

Definition:
A stochastic process $\{X_n : n \in T\}$ is a collection of random variables.

X_n take values in the set called "state space".

T is called the index set. (T often denotes time.)

Q) Consider a sequence of weather.

- What is the state space ?
- What is the index set ?

Markov chain

Markov chain is a stochastic process where the distribution of X_n depends only on X_{n-1}. For simplicity, in our class, we assume that the state space is discrete and the index set is $T = \{0, 1, 2, \ldots\}$.

Definition

$\{X_n : n \in T\}$ is a Markov chain if

$$P(X_n = x | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}) = P(X_n = x | X_{n-1} = i_{n-1})$$

for all x, $i_0, \cdots, i_{n-1}$ and n.

Intuitively, we can say that given the current state, the future and past states are independent.

More generally, for each n and m and for any states, Markov Chain implies

$$P(X_{n+m} = x | X_n = i_n, X_{n-1} = i_{n-1}, \ldots) = P(X_{n+m} = x | X_n = i_n).$$

Some examples of Markov chain

Q) Are i.i.d. sequences examples of Markov chain ?

Q) Is the random walk a Markov chain ? Note that the random walk is a diverging process.

For more details, refer to Wikipedia "Examples of Markov chains".

Key theorem

An irreducible, ergodic Markov chain has a unique stationary distribution π. The limiting distribution exists and is equal to π.

For any bounded function $h(\cdot)$, as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} h(X_i) \to Eh(X)$$

where X follows π.

Our main question is:

- How can we contruct Makov chains that converge to π ?

The key quantity to analyze Markov chain is the transition probability. (We assumes their homogeneity with respect to time.)

The transition probabilities are:

$$p_{ij} = P(X_{n+1} = j | X_n = i)$$

The matrix $\mathbf{P}$ whose (i, j)th element is p_{ij} is called the transition matrix.

- $p_{ij} \geq 0$
- $\sum_j p_{ij} = 1$

Let

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

and $\mathbf{P}^n$ be the matrix whose (i, j) element is $p_{ij}(n)$. $p_{ij}(n)$ are called the n-step transition probabilities.

Chapman-Kolmogorov equation:

$$p_{ij}(m+n) = \sum_k p_{ik}(m) p_{kj}(n).$$

Proof)

Chapman-Kolmogorov equation is nothing but matrix multiplication in terms of computation.

→ $$\mathbf{P}^{n+m} = \mathbf{P}^m \mathbf{P}^n$$

In particular, $\mathbf{P}^n = \mathbf{P}^{n-1}\mathbf{P}$

This implies $\mathbf{P}^n = \mathbf{P} \cdots \mathbf{P}$ (n times)

Q) Let $\mathbf{P} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$. Interpret $\mathbf{P}^2$ and $\mathbf{P}^3$.

Suppose that μ_0 is the initial distribution.
Then, the marginal probabilities

$$\mu_n = (P(X_n = 1), P(X_n = 2), \ldots, P(X_n = N))$$

are given by

$$\mu_n = \mu_0 \mathbf{P}^n.$$

Q) What is the meaning of μ_n ?

Q) Prove this.

Application of Markov chain to a population modelling

- We want to model a population in the city vs suburbs.
- Currently, 60% lives in city.
- It turns out that each year (1) 5% of citizes move to suburbs,
- and (2) 3% of suburbanities move to city.
- Does this population settle down to a steady state ?

Definition:
We say that π is a stationary distribution if $\pi = \pi\mathbf{P}$.

Remark) If at any time the chain has distribution π, then it will continue to have distribution π.

Definition:

We say that a chain has limiting distribution if $\mathbf{P}^n \to \begin{pmatrix} \pi \\ \pi \\ \vdots \\ \pi \end{pmatrix}$ for some π.

This implies that $\lim_{n\to\infty} \mathbf{P}^n_{ij}$ exists and is independent of i.

Another interpretation for the stationary distribution

Define

$$I_n = \begin{cases} 1 & \text{if } X_n = j \\ 0 & \text{if } X_n \neq j \end{cases}$$

Number of visits to j in first N transitions $= \sum_{n=1}^{N} I_n$.

Derive $\mathrm{E}\left(\sum_{n=1}^{N} I_n \,|\, X_0 = i\right) = \sum_{n=1}^{N} p_{ij}^{(n)}$ and interpret this result.

Simulating a Markov chain.

Consider a row vector $\mu_n = (\mu_n(1), \ldots, \mu_n(N))$ where $\mu_n(i) = P(X_n = i)$. (The cardinality of the state space is N, here.)

1. Draw $X_0 \sim \mu_0$.
2. Suppose that we obtained i from the previous step.
   Draw $X_1 \sim$ ith row of $\mathbf{P}$.
3. Suppose that we obtained j from the previous step.
   Draw $X_2 \sim$ jth row of $\mathbf{P}$.
4. $\vdots$

Definition:
We say that i reaches j if

$$p_{ij}(n) > 0$$

for some n. (Notation: $i \to j$).

We say that i and j communicate ($i \leftrightarrow j$) if $i \to j$ and $j \to i$.

Q) Does $i \leftrightarrow j$ imply $j \leftrightarrow i$?

Q) Do $i \leftrightarrow j$ and $j \leftrightarrow k$ imply $i \leftrightarrow k$?

If all states communicate with each other, we say that the Markov chain is irreducible.

Definition:
We say that a Markov chain is irreducible if

$$p_{ij}(m) > 0$$

for some m for each pair of (i, j). (Note that m can be different for different pairs.)

Q)

$$\mathbf{P} = \begin{pmatrix} 1/3 & 2/3 & 0 & 0 \\ 2/3 & 1/3 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Is this Markov chain irreducible ?

Note that the fact that a Markov chain has a stationary distribution does not imply its convergence.

Example: Let $\pi = (1/3, 1/3, 1/3)$ and $\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$

Q) Explain why this is an example of periodic chain.

Q) Is π a stationary distribution ?

Q) Does this chain have a limit ?

Remark)Every finite irreducible Markov Chain has a unique stationary distribution.

Example: $\mathbf{P}=\begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix}$

Q) Find the stationary distribution.

Q) Is it useful to know the eigenvalues of $\mathbf{P}$?

Q) Compute $\mathbf{P}^n$ for large n.

Definition:
The period of a state i is defined by

$$d(i) = gcd\{n : p_{ii}(n) > 0\}.$$

A state i is periodic if $d(i) > 1$ and aperiodic if $d(i) = 1$.

A Markov chain is aperiodic if all states are aperiodic.

Remark) If the finite Markov Chain is irreducible and aperiodic, then

$$\lim_{n\to\infty} \mathbf{P}_{ij}^{n} \to \pi_j \text{ for all } i \text{ and } j.$$

Definition:
We say that state i is recurrent if

$$P(X_n = i \text{ for some } n \geq 1 | X_0 = i) = 1.$$

Otherwise, we say that state i is transient.

Remark) State i is recurrent if $\sum_n p_{ii}(n) = \infty$.

Definition:
Suppose that state $X_0 = i$. Define the recurrence time

$$T_{ij} = \min\{n > 0 : X_n = j\}.$$

The mean recurrence time of a recurrence state i is defined by

$$m_i = E(T_{ii}).$$

A recurrent state is null if $m_i = \infty$. Otherwise, it is called non-null (positive).

Theorem (without proof):
A recurrent state is null if and only if $p_{ii}(n) \to 0$ as $n \to \infty$.

Definition:
We say that a state is ergodic if it is recurrent, non-null and aperiodic. A chain is ergodic if all its states are ergodic.

Key theorem revisited

An irreducible, ergodic Markov chain has a unique stationary distribution π. The limiting distribution exists and is equal to π.

For any bounded function $h(\cdot)$, as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} h(X_i) \to Eh(X)$$

where X follows π.