An introduction to Markov chains

This lecture will be a general overview of basic concepts relating to Markov chains, and some properties useful for Markov chain Monte Carlo sampling techniques. In particular, we'll be aiming to prove a "Fundamental Theorem" for Markov chains.

1 What are Markov chains?

Definition. Let $\{X_0, X_1, ..\}$ be a sequence of random variables and $Z = 0, \pm 1, \pm 2, ...$ be the state space, i.e. all possible values the random variable can take on. Then $\{X_0, X_1, ..\}$ is called a *discrete-time Markov chain* if

$$P(X_{n+1} = i_{n+1} | X_n = i_n), i \in Z$$

That is, the state at time step $n+1$ is dependent only on the state at time n.

The definition can describe a random walk on a graph where the vertices are the state space Z, and the edges are weighted by the transition probabilities:

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

Definition. A *homogeneous* Markov chain is one that does not evolve in time; that is, its transition probabilities are independent of the time step n. Then we have the "n-step" transition probabilities

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

and we have

$$p_{ij}^{(0)} = \begin{cases} 1 & : i = j \\ 0 & : i \neq j \end{cases}$$

Now we can define a theorem.

Theorem. Chapman-Kolmogorov equation.

$$p_{ij}^{(m)} \sum_{k \in Z} p_{ik}^{(r)} p_{kj}^{(m-r)} \forall r \in \mathbb{N} \cup \{0\}$$

Proof.

$$\begin{aligned}
p_{ij} = P(X_m = j|X_0 = i) &= \sum_{k \in Z} P(X_m = j, X_r k|X_o = i) \\
&= \sum_{k \in Z} P(X_m = j|X_r = k, X_0 = i)P(X_r = k|X_0 = i) \\
&= \sum_{k \in Z} P(X_m = j|X_r = k)P(X_r = k|X_0 = i) \\
&= \sum_{k \in Z} p_{ik}^{(r)} p_{kj}^{(m-r)}
\end{aligned}$$

We can write this as a matrix for convenience:

$$\mathbf{P}^{(m)} = ((p_{ij}^{(m)}))$$

Corollary.

$$\mathbf{P}^{(m)} = \mathbf{P}^m$$

Proof. Chapman-Kolmogorov in matrix form gives us

$$\begin{aligned}
\mathbf{P}^{(m)} &= \mathbf{P}^{(r)}\mathbf{P}^{(m-r)} \forall r \in \mathbb{N} \cup \{0\} \\
\mathbf{P}^{(2)} &= \mathbf{P} \times \mathbf{P} = \mathbf{P}^2 \\
\mathbf{P}^{(3)} &= \mathbf{P} \times \mathbf{P}^2 = \mathbf{P}^3 \\
\mathbf{P}^{(m)} &= \mathbf{P}^m, m \geq 2, \text{then} \\
\mathbf{P}^{(m+1)} &= \mathbf{P} \times \mathbf{P}^m = \mathbf{P}^{m+1}
\end{aligned}$$

2 Several definitions

A Markov Chain is completely determined by its transition probabilities and its initial distribution.

An *initial distribution* is a probability distribution

$$\{\pi_i = P(X_0 = i)|i \in \mathbb{Z}\}$$

such that $\sum_i \pi_i = 1$.

A distribution is *stationary* if it satisfies $\pi = \pi\mathbf{P}$.

The *period* of state i is defined as

$$d_i = \gcd\{m \in \mathbb{Z}|p_{ii}^{(m)} > 0\}$$

that is, the gcd of the numbers of steps that it can take to return to the state.

If $d_i = 1$, the state is *aperiodic*– it can occur at non-regular intervals.

A state j is *accessible* from a state i if the system, when started in i, has a nonzero probability of eventually transitioning to j, or more formally if there exists some $n \geq 0$ such that

$$Pr(X_n = j|X_0 = i) > 0.$$

We write this as $(i \to j)$.

We define the *first-passage time* (or "hitting time") probabilities as

$$f_{ij}^{(m)} = P(X_m = j, X_k \neq j, 0 < k < m-1 | X_0 = i), i, j \in \mathbb{Z}.$$

that is, the time step at which we first reach state j.

We denote the expected "return time" as

$$\mu_{ij} = \sum m = 1^{\infty} m f_{ij}^{(m)}$$

A state is *recurrent* if

$$\sum_{m=1}^{\infty} f_{ij}^{(m)} = 1$$

(and *transient* if the sum is greater than 1).

It is *positive-recurrent* if $\mu_{ii} < \infty$. That is, we expect to return to the state in a finite number of time steps.

3 Fundamental Theorem of Markov Chains

Theorem. For any irreducible, aperiodic, positive-recurrent Markov chain $\mathbf{P}$ there exists a unique stationary distribution $\{\pi_j, j \in \mathbb{Z}\}$.

Proof. We know that for any m,

$$\sum_{i=0}^{m} p_{ij}^{(m)} \leq \sum_{i=0}^{\infty} p_{ij}^{(m)} \leq 1.$$

If we take the limit as $m \to \infty$:

$$\lim_{m\to\infty} \sum_{i=0}^{m} p_{ij}^{(m)} = \sum_{i=0}^{\infty} \pi_j \leq 1.$$

This implies that for any M, $\sum_{i=0}^{M} \pi_j \leq 1$.

Now we can use Chapman-Kolmogorov:

$$p_{ij}^{(m+1)} = \sum_{i=0}^{\infty} p_{ik}^{(m)} p_{kj} \geq \sum_{i=0}^{M} p_{ik}^{(m)} p_{kj}$$

and take the limit again as $m, M \to \infty$

$$\pi_j \geq \sum_{k=0}^{\infty} \pi_k p_{kj}$$

Now for the sake of contradiction, let's assume that strict inequality holds for some state j. Then if we sum over all of these states, we have

$$\begin{aligned}\sum_{j=0}^{\infty} \pi_j &> \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \pi_k p_{kj} \\ &= \sum_{k=0}^{\infty} \pi_k \sum_{j=0}^{\infty} p_{kj} \\ &= \sum_{k=0}^{\infty} \pi_k\end{aligned}$$

but this is a contradiction. So, for some state j, equality must hold.

$$\pi_j = \sum_{k=0}^{\infty} \pi_k p_{kj}$$

Thus, a unique stationary distribution exists.

We can separately prove that we're guaranteed to converge to the stationary distribution, but this proof is somewhat more involved.

References

[1] Aaron Plavnick. The fundamental theorem of Markov chains. *VIGRE REU at UChicago*, 2008.