Markov Chain Monte Carlo

Samuel Brody

April 2009

Outline

Outline

Markov Chain

- Represents a process (series of states).
- Probability of transition to a new state (X_{t+1}) depends only on the current state (X_t).

Markov Chain

What's it good for?

- Many processes can be described or approximated by a MC.
- Ignoring long range dependencies simplifies the problem.
- In many cases the advantages outweigh the loss of information.

Markov Chain

Notation:

- $\pi_t(x)$ - the probability of being in state x at time t.
- π_0 - starting probabilities.
- $P_{x \to y}$ - the transition probability from state x to state y.
- $\pi_{t+1} = \pi_t \cdot P$

Stationary Distribution

- MCs which fulfill certain conditions converge to a *Stationary Distribution*.
- The probability of being in state X does not depend on where you started.

  $\pi_t = \pi_{t'}$ for all t, t'.
- or

  $\pi^* = \pi^* \cdot P$

Conditions for Stationary Dist.

The MC must be:

- Irreducible:
  every state reachable from every other - no dead ends.
- Aperiodic:
  no cycle of fixed length between two states.

A sufficient condition is:

- Reversibility:
  $P_{x \to y} \cdot \pi^*(x) = P_{y \to x} \cdot \pi^*(y)$

Reversibility

- Reversibility $\Rightarrow$ unique stationary distribution
- Proof:

$$P_{x\to y}\cdot\pi^*(x)=P_{y\to x}\cdot\pi^*(y)$$

$$(\pi^*\cdot P)_x=\sum_y\pi^*(y)\cdot P_{y\to x}=\sum_y\pi^*(x)\cdot P_{x\to y}=\pi^*(x)\cdot\sum_y P_{x\to y}=\pi^*(x)$$

Outline

1. Markov Chains
   - What are they?
   - Stationary Distribution
     - What is it?
     - Why is it useful?
     - Conditions
2. MCMC
   - What's it for?
   - How it's done
     - Metropolis
     - Hasting
     - Common proposal distributions
   - Summary

Markov Chain Monte Carlo

- A way to sample from a complex distribution
- Useful for calculating:
  - typical states
  - expected values
  - simulation

Example

Protein Folding:

- given a specific folding, we know how to calculate the energy level/stability

We want to know

- the most likely folding(s)
- expected value of some property (e.g., toxicity, binding)

Monte Carlo Integration:

If we have a representative sample of a distribution, we can approximate these values

$$E_{p(x)}[f(x)] \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i)$$

Markov Chain Monte Carlo

The Idea:

- create a Markov Chain with the desired stationary distribution
- run it until it converges
- then start sampling the states

Metropolis Algorithm

Remember - we know how to calculate the stability function $s(x)$ for a given folding x.

The Algorithm:

- pick a (simple) transition distribution $Q_{x \to y}$
- start from a random state x_0
- in each step, sample a potential new state x^*_{i+1} from $Q_{x_i \to ?}$
- if $\frac{s(x^*_{i+1})}{s(x_i)} > 1$, move to the new (more stable) state:

  $x_{i+1} = x^*_{i+1}$ w.p. 1
- otherwise, move to the new state with probability $\frac{s(x^*_{i+1})}{s(x_i)}$:

$$
\begin{aligned}
x_{i+1} &= x^*_{i+1} && \text{w.p. } \frac{s(x^*_{i+1})}{s(x_i)} \\
x_{i+1} &= x_i && \text{w.p. } 1 - \frac{s(x^*_{i+1})}{s(x_i)}
\end{aligned}
$$

Why it works

Claim: if we choose a Q that is symmetric ($Q_{x \to y} = Q_{y \to x}$) the chain will converge to the desired distribution π.

Proof:

- The transition probability of our chain

$$T_{x \to y} = Q_{x \to y} \cdot min\left[\frac{s(y)}{s(x)}, 1\right]$$

$$\frac{s(y)}{s(x)} = \frac{\pi(y)}{\pi(x)}$$

- so

$$T_{x \to y} = Q_{x \to y} \cdot min\left[\frac{\pi(y)}{\pi(x)}, 1\right]$$

Proof - cont.

- in our construction

$$T_{x\to y} = Q_{x\to y} \cdot min\left[\frac{\pi(y)}{\pi(x)}, 1\right]$$

- we want to show

$$T_{x\to y} \cdot \pi(x) = T_{y\to x} \cdot \pi(y)$$

- i.e.,

$$Q_{x\to y} \cdot min\left[\frac{\pi(y)}{\pi(x)}, 1\right] \cdot \pi(x) = Q_{y\to x} \cdot min\left[\frac{\pi(x)}{\pi(y)}, 1\right] \cdot \pi(y)$$

- if $\pi(x) = \pi(y)$ then we have

$$Q_{x\to y} \cdot 1 \cdot \cancel{\pi(x)} = Q_{y\to x} \cdot 1 \cdot \cancel{\pi(y)}$$

  which is true because Q is symmetric.

- if $\pi(x) < \pi(y)$ then

$$Q_{x\to y} \cdot \pi(x) = Q_{y\to x} \cdot \frac{\pi(x)}{\cancel{\pi(y)}} \cdot \cancel{\pi(y)}$$

Hastings' Contribution

- The symmetry requirement for Q is restrictive.
- Often, transitions in one direction are more probable.
- Hastings modified the algorithm to use the transition probability

$$min\left[\frac{s(y)\cdot Q_{y\rightarrow x}}{s(x)\cdot Q_{x\rightarrow y}},1\right]$$

- With this modification, Q can be non-symmetric (similar proof).

The Metropolis-Hastings Algorithm

- pick any transition distribution $Q_{x \to y}$
- start from a random state x_0
- in each step, sample a potential new state x^*_{i+1} from $Q_{x_i \to ?}$
- if $\frac{s(y) \cdot Q_{y \to x}}{s(x) \cdot Q_{x \to y}} > 1$, move to the new (more stable) state:

  $x_{i+1} = x^*_{i+1}$ w.p. 1

- otherwise, move to the new state with probability $\frac{s(y) \cdot Q_{y \to x}}{s(x) \cdot Q_{x \to y}}$:

$$
\begin{aligned}
x_{i+1} &= x^*_{i+1} \quad &\text{w.p. } \frac{s(y) \cdot Q_{y \to x}}{s(x) \cdot Q_{x \to y}} \\
x_{i+1} &= x_i \quad &\text{w.p. } 1 - \frac{s(y) \cdot Q_{y \to x}}{s(x) \cdot Q_{x \to y}}
\end{aligned}
$$

Which Q to choose?

We want the distribution to be:

- Well-Mixing: move around the parameter space
- High Acceptance: a good probability that transitions happen

Common Qs

Independence: $Q_{x \to y} = f(y)$

- independent of current state
- non-symmetric
- usually use a Normal/Gaussian distribution, with tunable variance

Random Walk: $Q_{x \to y} = f(y - x)$

- commonly f is symmetric around zero:

$$f(y - x) = f(x - y)$$

- can use Metropolis version

MCMC

What's it for?

- obtaining a representative sample from a complex distribution

Why?

- to estimate typical states and expected values

How?

- build a Markov Chain that (eventually) behaves like the target distribution

Keep in mind:

- burn-in period
- proposal distribution