Markov Chain Monte Carlo

Robert Finn

Feb. 22, 2018

A Markov chain $\{X^{(t)}\}$ is a sequence of dependent random variables

$$X^{(0)}, X^{(1)}, X^{(2)}, \ldots, X^{(t)}, \ldots$$

such that the probability distribution of $X^{(t)}$ given the past variables depends only on $X^{(t-1)}$. This conditional probability distribution is called a *transition kernel* or a *Markov kernel* K; that is,

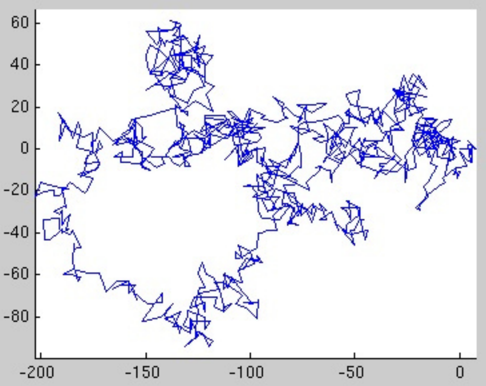
$$X^{(t+1)} \mid X^{(0)}, X^{(1)}, X^{(2)}, \ldots, X^{(t)} \sim K(X^{(t)}, X^{(t+1)})$$

For example, a simple *random walk* Markov chain satisfies

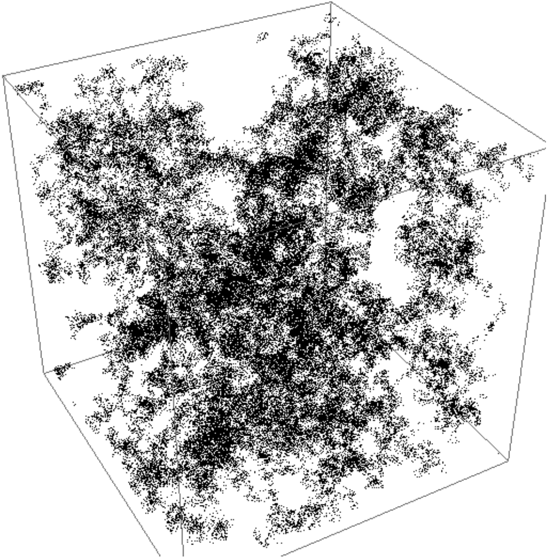
$$X^{(t+1)} = X^{(t)} + \epsilon_t \,,$$

where $\epsilon_t \sim \mathcal{N}(0,1)$, independently of $X^{(t)}$; therefore, the Markov kernel $K(X^{(t)}, X^{(t+1)})$ corresponds to a $\mathcal{N}(X^{(t)}, 1)$ density.

Markov Chains



Markov Chains



- Markov chains encountered in Markov chain Monte Carlo (MCMC) settings enjoy a very strong stability property.
- A stationary probability distribution exists by construction for those chains; that is, there exists a probability distribution f such that if $X(t) \sim f$, then $X(t+1) \sim f$.
- The kernel and stationary distribution satisfy the equation

$$\int_{\mathcal{X}} K(x,y) f(x) \mathrm{d}x = f(y).$$

- The existence of a stationary distribution (or stationarity) imposes a preliminary constraint on K called *irreducibility*
- The kernel K allows for free moves all over the state-space;
- No matter the starting value $X(0)$, the sequence $\{X(t)\}$ has a positive probability of eventually reaching any region of the state-space.
- A sufficient condition is that $K(x, \cdot) > 0$ everywhere.
- The existence of a stationary distribution has major consequences on the behavior of the chain $\{X(t)\}$.
- This implies that most of the chains involved in MCMC algorithms are recurrent, i.e., they will return to any arbitrary nonnegligible set an infinite number of times.

Markov Chains

- The existence of a stationary distribution (or stationarity) imposes a preliminary constraint on K called *irreducibility*
- The kernel K allows for free moves all over the state-space;
- No matter the starting value $X(0)$, the sequence $\{X(t)\}$ has a positive probability of eventually reaching any region of the state-space.
- A sufficient condition is that $K(x, \cdot) > 0$ everywhere.
- The existence of a stationary distribution has major consequences on the behavior of the chain $\{X(t)\}$.
- This implies that most of the chains involved in MCMC algorithms are recurrent, i.e., they will return to any arbitrary nonnegligible set an infinite number of times.

Markov Chains

- The existence of a stationary distribution (or stationarity) imposes a preliminary constraint on K called *irreducibility*
- The kernel K allows for free moves all over the state-space;
- No matter the starting value $X(0)$, the sequence $\{X(t)\}$ has a positive probability of eventually reaching any region of the state-space.
- A sufficient condition is that $K(x, \cdot) > 0$ everywhere.
- The existence of a stationary distribution has major consequences on the behavior of the chain $\{X(t)\}$.
- This implies that most of the chains involved in MCMC algorithms are recurrent, i.e., they will return to any arbitrary nonnegligible set an infinite number of times.

Markov Chains

- The existence of a stationary distribution (or stationarity) imposes a preliminary constraint on K called *irreducibility*
- The kernel K allows for free moves all over the state-space;
- No matter the starting value $X(0)$, the sequence $\{X(t)\}$ has a positive probability of eventually reaching any region of the state-space.
- A sufficient condition is that $K(x, \cdot) > 0$ everywhere.
- The existence of a stationary distribution has major consequences on the behavior of the chain $\{X(t)\}$.
- This implies that most of the chains involved in MCMC algorithms are recurrent, i.e., they will return to any arbitrary nonnegligible set an infinite number of times.

Markov Chains

- The existence of a stationary distribution (or stationarity) imposes a preliminary constraint on K called *irreducibility*
- The kernel K allows for free moves all over the state-space;
- No matter the starting value $X(0)$, the sequence $\{X(t)\}$ has a positive probability of eventually reaching any region of the state-space.
- A sufficient condition is that $K(x, \cdot) > 0$ everywhere.
- The existence of a stationary distribution has major consequences on the behavior of the chain $\{X(t)\}$.
- This implies that most of the chains involved in MCMC algorithms are recurrent, i.e., they will return to any arbitrary nonnegligible set an infinite number of times.

Markov Chains

- The existence of a stationary distribution (or stationarity) imposes a preliminary constraint on K called *irreducibility*
- The kernel K allows for free moves all over the state-space;
- No matter the starting value $X(0)$, the sequence $\{X(t)\}$ has a positive probability of eventually reaching any region of the state-space.
- A sufficient condition is that $K(x, \cdot) > 0$ everywhere.
- The existence of a stationary distribution has major consequences on the behavior of the chain $\{X(t)\}$.
- This implies that most of the chains involved in MCMC algorithms are recurrent, i.e., they will return to any arbitrary nonnegligible set an infinite number of times.

- In the case of recurrent chains, the stationary distribution is also a limiting distribution in the sense that the limiting distribution of $X(t)$ is f for almost any initial value $X(0)$.
- This property is also called *ergodicity*.
- This has major consequences from a simulation point of view in that, if a given kernel K produces an ergodic Markov chain with stationary distribution f, generating a chain from this kernel K will eventually produce simulations from f.

- In the case of recurrent chains, the stationary distribution is also a limiting distribution in the sense that the limiting distribution of $X(t)$ is f for almost any initial value $X(0)$.
- This property is also called *ergodicity*.
- This has major consequences from a simulation point of view in that, if a given kernel K produces an ergodic Markov chain with stationary distribution f , generating a chain from this kernel K will eventually produce simulations from f.

- In the case of recurrent chains, the stationary distribution is also a limiting distribution in the sense that the limiting distribution of $X(t)$ is f for almost any initial value $X(0)$.
- This property is also called *ergodicity*.
- This has major consequences from a simulation point of view in that, if a given kernel K produces an ergodic Markov chain with stationary distribution f , generating a chain from this kernel K will eventually produce simulations from f.

- In particular, for integrable functions h, the standard average

$$\frac{1}{T}\sum_{t=1}^{T} h(X^{(t)}) \longrightarrow \mathbb{E}_f[h(X)]$$

- which means that the Law of Large Numbers that lies at the basis of Monte Carlo methods can also be applied in MCMC settings.

Algorithm 4 Metropolis–Hastings

Given $x^{(t)}$,

1. Generate $Y_t \sim q(y|x^{(t)})$.
2. Take

$$X^{(t+1)} = \begin{cases} Y_t & \texttt{with probability} \quad \rho(x^{(t)}, Y_t), \\ x^{(t)} & \texttt{with probability} \quad 1 - \rho(x^{(t)}, Y_t), \end{cases}$$

where

$$\rho(x, y) = \min\left\{ \frac{f(y)}{f(x)} \frac{q(x|y)}{q(y|x)}, 1 \right\}.$$

A generic R implementation is straightforward, assuming a generator for $q(y|x)$ is available as geneq(x). If x[t] denotes the value of $X^{(t)}$,

```
> y=geneq(x[t])
> if (runif(1)<f(y)*q(y,x[t])/(f(x[t])*q(x[t],y))){
+     x[t+1]=y
+     }else{
+       x[t+1]=x[t]
+       }
```

MCMC

The distribution q is called the instrumental (or proposal or candidate) distribution and the probability $\rho(x, y)$ the Metropolis-Hastings acceptance probability. It is to be distinguished from the acceptance rate, which is the average of the acceptance probability over iterations,

$$\overline{\rho} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \rho(X^{(t)}, Y_t) = \int \rho(x, y) f(x) q(y|x)\, \mathrm{d}y \mathrm{d}x$$

Note that Algorithm 4 satisfies what is known as the *detailed balance condition*

$$f(x)K(y|x) = f(y)K(x|y)$$

from which we can deduce that f is the stationary distribution of the chain $\{X(t)\}$ by integrating each side of the equality in x.

Example 6.1. Recall Example 2.7, where we used an Accept–Reject algorithm to simulate a beta distribution. We can just as well use a Metropolis–Hastings algorithm, where the target density f is the $\mathcal{B}e(2.7, 6.3)$ density and the candidate q is uniform over $[0, 1]$, which means that it does not depend on the previous value of the chain. A Metropolis–Hastings sample is then generated with the following R code:

```
> a=2.7; b=6.3; c=2.669 # initial values
> Nsim=5000
> X=rep(runif(1),Nsim)  # initialize the chain
> for (i in 2:Nsim){
+   Y=runif(1)
+   rho=dbeta(Y,a,b)/dbeta(X[i-1],a,b)
+   X[i]=X[i-1] + (Y-X[i-1])*(runif(1)<rho)
+   }
```