Metropolis Hastings

September 14, 2021

The Metropolis-Hastings (MH) algorithm is a sampling algorithm that generates samples from a *target distribution* $\pi(\cdot)$. MH is most effective in the case where $\pi(\cdot)$ is intractable. Let's express $\pi(\cdot)$ as the following fraction,

$$\pi(\cdot) = \frac{f(\cdot)}{K}$$

where $f(\cdot)$ is some tractable density function and K is an intractable normalizing constant. MH enables us to sample from $\pi(\cdot)$ without knowing anything about the intractable normalizing constant K!

The idea behind MH is to construct a Markov Chain whose stationary distribution is $\pi(\cdot)$. How do we do this? From Markov Chain theory we know that if a Markov Chain is reversible, the detailed balance equation holds,

$$\pi(x)p(x, y) = \pi(y)p(y, x)$$

where $\pi(\cdot)$ is the stationary distribution and $p(x, y)$ is the transition kernel. Hence, if we can find a transition kernel $p(x, y)$ which satisfies this reversibility condition, we can sample from the target distribution $\pi(\cdot)$.

We start by defining some **tractable** *proposal distribution* $q(x, y)$, which generates a proposal y given some current state x (this is something you pick). Since $q(x, y)$ is a transition kernel, does it satisfy the reversibility condition? Let's derive a general solution. Assume $q(x, y)$ **does not** satisfy the reversibility condition (if it does, then we just have the Metropolis algorithm). We have the following,

$$\pi(x)q(x, y) > \pi(y)q(y, x) \qquad (w.l.g)$$

An intuitive interpretation of what this equation is saying is basically we are moving from state x to state y more than we are vice versa; in short, detailed balance is not satisfied. To correct this, we introduce a probability $\alpha(x, y)$ that the move is actually made,

$$\pi(x)q(x, y)\alpha(x, y) = \pi(y)q(y, x)$$

Rearranging we obtain,

$$\alpha(x, y) = \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}$$

Reiterating, we interpret $\alpha(x, y)$ as the probability of moving from state x to state y. Since it is a probability, it cannot exceed 1 so the final definition is,

$$\alpha(x, y) = \min[\frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}, 1]$$

Note that $\alpha(x, y)$ **is** the transition kernel $p(x, y)$ that we were after! It satisfies the detailed balance equation by definition; however, does it deal with the pesky intractable normalizing constant K? Yes! We have that,

$$\begin{aligned}\frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} &= \frac{\frac{f(y)q(y,x)}{K}}{\frac{f(x)q(x,y)}{K}} \\ &= \frac{f(y)q(y, x)}{f(x)q(x, y)}\end{aligned}$$

So we no longer have to worry about K (yay!). Ergo, we can simply use $\alpha(x, y)$ as our transition kernel for our Markov Chain and rest assured that as long as we obtain enough samples, they will be distributed according to the stationary distribution; voila, we have samples from the intractable distribution $\pi(\cdot)$!

Algorithm 1 Metropolis Hastings

```
Initialize x_0
for j = 1, 2, ..., N do
    Generate y from q(x^(j), ·) and u from U(0, 1)
    if u ≤ α(x^(j), y) then
        x^(j+1) = y
    else
        x^(j+1) = x^(j)
    end if
end for
Return x^(1), x^(2), ..., x^(N)
```

1 Notes

- $\pi(x)$ is the *target distribution*
  - this is the distribution we want to generate samples from
  - $\pi(x) = \frac{f(x)}{K}$; the normalizing constant K is usually intractable!
- $q(x, y)$ is the *proposal distribution*
  - generates a proposal y given the current state x
  - **you** pick the distribution! (which means you are able to sample from it easily, provided you don't use some intractable distribution)
- $\alpha(x, y) = \min[\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1]$
  - this **is** the transition kernel that will satisfy detailed balance
  - probability of moving from state x to state y
  - notice that

$$\begin{aligned}\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)} &= \frac{\frac{f(y)q(y,x)}{K}}{\frac{f(x)q(x,y)}{K}} \\ &= \frac{f(y)q(y,x)}{f(x)q(x,y)}\end{aligned}$$

    so we don't have to worry about the intractable normalizing constant (yay!)
- $\pi(x)p(x, y) = \pi(y)p(y, x)$ is the *reversibility condition*
  - necessarily, $\pi(\cdot)$ is the stationary distribution
  - $p(x, y)$ is a transition kernel