Hamiltonian MCMC-3

Friday, March 15, 2019 11:31 AM

Refs: 1206.1901
BDA3
1701.02434

- Connections to statistics:

As we discussed, the Hamiltonian can be interpreted as the total energy of a system.

Using the concept of canonical ensemble (mechanical system in thermal equilibrium with a heat bath at fixed temperature T) from statistical mechanics, we can relate the probability (density) distribution for a state x to its energy,

$$P(x) = \frac{1}{Z} \exp\left(-E(x)/T\right) \quad \text{---} \quad (1)$$

[Derivation follows from the definition of entropy (Boltzmann),

$$S(E) = k_B \log \Omega(E)$$

and literally counting individual states with energy E_i (see J. Mcgreevy's notes)]

In terms of the Hamiltonian function $H(p, q)$, we can use (1) to write the joint probability distribution,

$$P(p, q) \propto \exp\left(-H(p, q)\right) \quad \text{---} \quad (2)$$

(p: parameters of interest; q: auxiliary variable)

where we've set $T = 1$.

For a Bayesian analysis, we're interested in sampling the posterior, $P(q|y)$, which

we can write in terms of conditional distributions as follows,

$$P(q|y) \propto \int P(q|p,y)\, P(p)\, dp \qquad \text{---(3)}$$

$$\propto \int P(y|q,p)\, P(q|p)\, P(p)\, dp$$

where we've marginalized over the ancillary 'momentum' parameters in the first line and

[$P(A,B|C) = P(A|B,C)\, P(B|C)$]

used Bayes' theorem for the second.

Naive HMC sampler

In lieu of an algorithm, I'll summarize the steps in some detail.

The goal, like any MCMC sampler, is to draw new samples based on a sample, and then accept/reject steps with the Metropolis update.

Unlike the M-H sampler, for example, the proposal isn't a random walk but is guided by the gradient of the posterior density w.r.t its parameters.

Algorithm:

1.) Update p with a random initialization,

$$\vec{p}_0 \sim N(0, M_{ii})$$

2.) Next, we use the leapfog method (symplectic integrator; we'll discuss more about this choice of name later) to evaluate the Hamilton's equations. It relies on two 'tunable' parameters: step-size ϵ and number of steps L,

a) $\vec{p}(t+\epsilon/2) \leftarrow \vec{p}(t) + \frac{1}{2}\epsilon \frac{\partial \log p(q|y)}{\partial q}$

b) $\vec{q}(t+\epsilon) \leftarrow \vec{q}(t) + \epsilon M^{-1} \vec{p}(t)$ $\left(\frac{\partial H}{\partial p}\right)$

c) $\vec{p}(t+\epsilon) \leftarrow \vec{p}(t+\epsilon/2) + \frac{1}{2}\epsilon \frac{\partial \log p(q|y)}{\partial q}$ ——— (4)

We can combine a) and c) for all steps except the last, so we can write:

*

- $\vec{p}(t+\epsilon/2) \leftarrow \vec{p}_0 + \frac{1}{2}\epsilon \frac{\partial \log p(q|y)}{\partial q}$

- for (i in 1:L)

  { step b)

  $\vec{p}(t+\epsilon) = \vec{p}(t) + \epsilon \frac{\partial \log p(q|y)}{\partial q}$

  }

- Step c)

*

The accept-reject step follows the usual Metropolis update,

$q^{t+1} \begin{cases} q^* & , \quad \alpha(\cdot|q^t) \end{cases}$

$$q = \begin{cases} q^t & , \text{otherwise} \end{cases}$$ (after L steps)

where q^t, q^{t+1} are the initial and final states respectively, and the acceptance probability is given by,

$$\alpha(\cdot | q^t) = \min\left\{1, \exp\left(-H(q^*, p^*) + H(q^t, p^t)\right)\right\}$$

- We don't care about updating p since it is updated at the start of every iteration.

Comments:

1. The leapfrog integration steps preserves the the joint density $P(p, q | y)$. A formal proof is beyond the scope of this talk, but I'll try to motivate it intuitively: flat, low density (-ve gradient), high density (+ve gradient).

→ Density may change between iterations, however.

2. Metropolis → detailed balance → canonical density

Note:

$$P(p, q | y) = P(q | p, y) P(p)$$

Taking log on both sides,

$$\log P(p, q | y) = \log P(q | p, y) + \log P(p)$$

However, using the definition of canonical distribution from Eq. (2),

$$-K(p) - U(q) = \log P(q|p,y) + \log P(p)$$

$$\Rightarrow \boxed{U(q) = -\log P(q|p,y)}$$

$$\therefore \frac{\partial U}{\partial q} = -\frac{\partial \log P(q|p,y)}{\partial q}$$

The RHS is a negative log-likelihood implying that the 'potential' energy $U(q)$ increases as you 'roll' down the hill.