Hamiltonian Monte Carlo

Bradley Gram-Hansen

July 12, 2017

Abstract

These are note on aspects of Hamiltonian monte carlo methods.]

0.1 Radford Neals intro to HMC MCMC

0.1.1 Metropolis-hastings algorithm

The Metropolis hastings algorithm is the work horse of Monte Carlo Markov Chain (MCMC) methods, It relies on a simple reject and accept criteria for determining whether or not we should proceed to some next state. If our state is accepted, but does not satisfy the full criteria then a probability is associated with that state, and if that probability exceeds the one sampled from a given cut off, usually determined by sampling from a uniform distribution then it is accepted, else rejected.

Algorithm 1 Metropolis-Hasting algorithm

1: $x^0 \sim P_0(x)$ ▷ Where $p(x)$ is the proposed initial distribution
2: **for** $s = 0, 1, 2, \ldots$ **do**
3: $x \leftarrow x^s$
4: $x^{'} \sim Q(x^{'}|x)$
5: $a = \frac{\tilde{P}(x')Q(x|x')}{\tilde{P}(x)Q(x'|x)}$ ▷ Acceptance ratio
6: $r = min(1, a)$ ▷ Acceptance condition
7: $u \sim U(0, 1)$
8: $x^{s+1} = \begin{cases} x' \text{ if } u < r \\ x^s \text{ if } u \geq r \end{cases}$
9: **end for**

0.1.2 HMC for MCMC

In a top level view Hamilton Monte Carlo for Monte Carlo Markov Chain (HMC MCMC) is a two step process. In step one we define a Hamiltonian function in terms of the probability distribution from which we wish to sample from. We introduce a position variable, q and momentum variable p, where p is an auxiliary variable that typically has a Gaussian distribution. All p's are assumed independent. In step two, the HMC alternates simple updates for the momentum variables with Metropolis updates. This enables us to propose a new state by computing a trajectory according to Hamiltonian dynamics, implemented with the leapfrog method.

Prerequisites

The Hamiltonian of a physical system is defined completely with respect to the position q and p momentum variables, which span the phase space of the system. The Hamiltonian is the Legendre transform of the Lagrangian and gives us the total energy in the system. It is defined as follows:

$$H(\mathbf{q}, \mathbf{p}) = \sum_{i=1}^{d} \dot{q}_l p_l - L(\mathbf{q}, \dot{\mathbf{q}}(\mathbf{q}, \mathbf{p})) \tag{1}$$

where d is the system dimensions, and so the full state space with has $2d$ dimensions. Thus, for simplicity, if we set $d = 1$ we can derive the Hamiltonian equations as follows:

$$\frac{\partial H}{\partial p} = \dot{q} + p\frac{\partial \dot{q}}{\partial p} - \frac{\partial L}{\partial \dot{q}}\frac{\partial \dot{q}}{\partial p} = \dot{q} \tag{2}$$

and

$$\frac{\partial H}{\partial q} = p\frac{\partial \dot{q}}{\partial q} - \frac{\partial L}{\partial q} - \frac{\partial L}{\partial \dot{q}}\frac{\partial \dot{q}}{\partial q} = -\frac{\partial L}{\partial q} = -\dot{p} \tag{3}$$

and the process is the same for more than one dimension. We can write 1 more succinctly as:

$$H(\mathbf{q}, \mathbf{p}) = K(p) + U(q) \tag{4}$$

where $K(p)$ represents our kinetic energy and $U(q)$ is the potential energy.

Within the HMC MCMC framework the "positions", q, are the variables of interest and for each position variable we have to create a fictitious "momentum", p. For compactness let $z = (q, p)$. The potential energy $U(q)$ will be the minus of the log of the probabilty density for the distribution of the position variables we wish to sample, plus **any** constant that is convenient. The kinetic energy will represents the dynamics of our variables, for which a popular form of $K(p) = \frac{p^T M^{-1} p}{2}$, where M is symmetric, positive definite and typically diagonal. This form of $K(p)$ corresponds to a minus the log probability of the zero mean Gaussian distribution with covariance matrix M. For this choice we can write the Hamiltonian equations, for any dimension d, as follows:

$$\dot{q}_i = \frac{dq_i}{dt} = [M^{-1}p]_i \tag{5}$$

$$\dot{p}_i = \frac{dp_i}{dt} = -\frac{\partial U}{\partial q_i} \tag{6}$$

To view the Hamiltonian in terms of probabilities, we use the concept of the canonical distribution from Statistical mechanics to construct our pdf. Thus, the distribution that we wish to sample from can be related to the potential energy via the canonical distribution as:

$$P(z) = \frac{1}{Z}\exp\left(\frac{-E(z)}{T}\right) \tag{7}$$

As the Hamiltonian is just an energy function we may can insert 4 into our canonical distribution 7 which gives us the joint density:

$$P(q, p) = \frac{1}{Z}\exp(-U(q))\exp(-K(p) \tag{8}$$

where $T = 1$ is fixed. Where we characterise the posterior distribution for the model parameters using the potential energy function"

$$U(q) = -\log[\pi(q)L(q|D)] \tag{9}$$

where $\pi(q)$ is the prior distribution, and $L(q|D)$ is the likelihood, not the Lagrangian, of the given data D.

0.1.3 The Algorithm

The leapfrog method

The leapfrog method enables reduced error and allows us to dicretize Hamiltons equations, so that we can implement them numerically. We start with a state at $t = 0$ and then evaluate at a subsequent time $t + \epsilon, \ldots, t + n\epsilon$, where ϵ is the step in which we increase and n is the number of time steps.

$$p_i(t + \frac{\epsilon}{2}) = p_i(t) - \left(\frac{\epsilon}{2}\right) \frac{\partial U(q(t))}{\partial q_i} \tag{10}$$

$$q_i(t + \epsilon) = q_i(t) + \epsilon \frac{\partial K(p + \frac{\epsilon}{2})}{dp_i} \tag{11}$$

$$p_i(t + \epsilon) = p_i(t + \frac{\epsilon}{2}) - \left(\frac{\epsilon}{2}\right) \frac{\partial U}{\partial q_i} \tag{12}$$

$$\tag{13}$$

[1]

For the leapfrog method the local error, error after one step, is of $\mathcal{O}(\epsilon^2)$ and a global error, error after simulating for some fixed time interval s, which requires $\frac{s}{\epsilon}$ is $\mathcal{O}(\epsilon^3)$

Some initial points of notice

Neals implementation of the HMC can only be used to sample from continuous distributions on $\mathbb{R}^d$ for which a density function can be evaluated.
We must be able to compute the partial derivative of the log of the density function. The derivaties must exists at the points at which they are evaluated [**Automatic differention**]

HMC samples from the canonical distribution for q and p. q has the distribution of interest as specified by the potential $U(q)$. The distribution of the p's can be chosen by us and are independent of the q's. The p components are specified to be independent, with component p_i having variance m_i. The kinetic energy $K(p) = \sum_{i=1}^{d} \frac{p_i^2}{2m_i}(q(t + \epsilon))$

The steps

1. Step 1: Changes only the momentum
2. Step 2: May change both position and momentum

Both steps leave the canonical distribution of (q,p) invariant, hence the distribution remains invariant.

In **Step 1** we first draw the p_i randomly from their Gaussian distribution independently of the current values of the position values.

[1] For the usual choice of kinetic energy, we have $\frac{\partial K(p+\frac{\epsilon}{2})}{dp_i} = \frac{p_i(t+\frac{\epsilon}{2})}{m_i}$

next, in **Step 2** a Metropolis update is performed, using the Hamiltonian dynamics to propose a new state. Starting with the current state (q, p), Hamiltonian dynamics is simulated for L steps using the leapfrog method, with a stepsize of ϵ. L and ϵ are parameters of the mnodel that need to be tuned.

The momentum vairables at the end of this L-step trajectory are then negated, giving s proposed state $(q^*.p^*)$. This proposed state is accepted as the next state in the Markov Chain with probability:

$$\min[1, \exp(-H(q^*, p^*) + H(p, q)] = \min[1, \exp(-U(q^*) + U(q) - K(p^*) + K(p))] \tag{14}$$

If the proposed state is rejected, then the next state is the same as the current state and is counted again when calculating the expected value of some function.