Lecture 10:

①

Topics:

I) Hamiltonian Markov chain

II) Taylor, Ashdown & Hobson application.

Suppose d is some data vector drawn from some density $p(d|\theta)$ where $\theta = (\theta_1, \dots, \theta_n)$ is a parameter vector. Let $\pi(\theta)$ denote a prior density for θ.

Goal: Produce posterior samples

$$\theta^{(1)}, \dots, \theta^{(N)} \sim p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}.$$

Problem: if the dimension n of $\theta \in \mathbb{R}^n$ is large then Metropolis-Hastings chains typically have low acceptance rates.

Hamiltonian Monte Carlo (HMC)

... an MCMC chain which attempts to solve the low acceptance rate problem when θ is high dimensional.

The main requirement for this method is the ability to compute

$$\nabla_\theta \log p(\theta|d) \quad \& \quad p(\theta|d)$$

quickly (where $p(\theta|d)$ only needs to be computed up to a unknown normalizing factor).

②

To define the HMC sample, start by defining a momentum vector

$$p := (p_1, \dots, p_n) \in \mathbb{R}^n$$

with corresponding mass

$$m := (m_1, \dots, m_n) \in \mathbb{R}^n \quad \text{s.t. } m_i > 0.$$

Now define the Hamiltonian as follows

$$H(\theta, p) := -\log p(\theta|d) + \sum_{i=1}^{n} \frac{|p_i|^2}{2m_i}$$

$$= -\log p(\theta|d) - \log \pi(p) + c$$

$$= -\log\left[p(\theta|d)\pi(p)\right] + c$$

where $\pi(p)$ is the density of $p \sim N(0, \mathrm{diag}(m))$.

You can think of $-\log \pi(p)$ as the kinetic energy of the system (θ, p) and $-\log p(\theta|d)$ as potential energy.

If we are thinking about (θ, p) as describing a physical system at time $t=0$, then we can evolve the system, producing (θ^t, p^t) for $t \in [0, \infty)$, with Hamiltonian dynamics:

$$\text{(H1)} \quad \frac{d\theta^t}{dt} = \nabla_p H(\theta^t, p^t)$$

$$\text{(H2)} \quad \frac{dp^t}{dt} = -\nabla_\theta H(\theta^t, p^t)$$

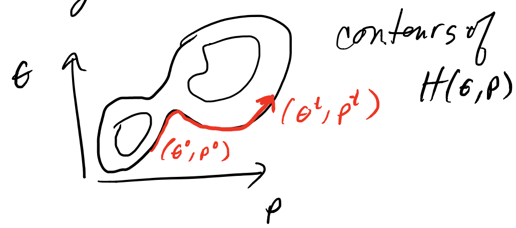
Under this evolution $H(\theta^t, p^t)$ is preserved

$H(\theta^t, p^t) \equiv H(\theta^0, p^0) \ \forall t \in [0, \infty)$ since

$$\frac{d}{dt} H(\theta^t, p^t) = \nabla_\theta H(\theta^t, p^t) \frac{d\theta^t}{dt} + \nabla_p H(\theta^t, p^t) \frac{dp^t}{dt}$$

$$= 0.$$

∴ (θ^t, p^t) moves along a contour of the log likelihood given by $H(\theta,p)+c$



One can also show that (H1) & (H2) preserve volumn in "phase space $(\theta,p)\in\mathbb{R}^n\times\mathbb{R}^n$" meaning that for any set $\Omega\subset\mathbb{R}^n\times\mathbb{R}^n$ one has

$$vol(\Omega)=vol(\mathcal{H}^t(\Omega))$$

where

$$\mathcal{H}^t(\Omega) := \left\{\mathcal{H}^t(\theta,p)\in\mathbb{R}^n\times\mathbb{R}^n \;\middle|\; (\theta,p)\in\Omega\right\}$$

$\mathcal{H}^t(\theta,p) = (\theta^t,p^t)$ evolved by (H1)(H2) with initial condition $(\theta^0,p^0):=(\theta,p)$.

To see this simply note that $\mathcal{H}^t(\theta,p)$ is a div free evolution. i.e. that

$$\operatorname{div}\left((\nabla_p H, -\nabla_\theta H)^T\right) = \sum_{i=1}^n \frac{\partial^2}{\partial\theta_i\partial p_i}H(\theta,p) - \sum_{i=1}^n \frac{\partial^2}{\partial p_i\partial\theta_i}H(\theta,p)$$

$$=0$$

Why is this useful?

Let $X \equiv (\theta^0,p^0)$ be a random vector with some density $p_X(x)$.

Transform $X\mapsto Y$ as follows

$$Y=\mathcal{H}^t(X)\equiv(\theta^t,p^t) \text{ under (H1) (H2)}$$

the change of variables formula says

$$p_X(x)\,\partial x = p_Y(y)\,\partial y$$

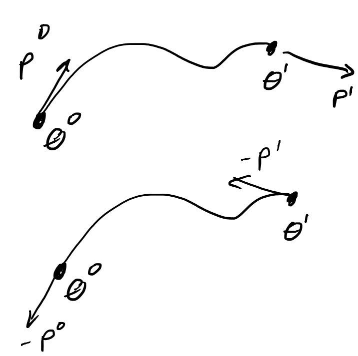
$$\therefore\quad p_X(x)=p_Y(y)\left|\frac{\partial y}{\partial x}\right|$$

jacobian

$$= p_Y(\mathcal{H}^t(x))\underbrace{\left|\frac{\partial\mathcal{H}^t(x)}{\partial x}\right|}_{=1 \text{ by vol preserving property}}$$

i.e. $p_X(\theta^0,p^0)=p_Y(\theta^t,p^t)$

Finally we note that (H1) & (H2) is reversable in that flipping the sign on momentum p runs the chain in reverse.



Let $s \equiv (\theta, p)$ denote general state vectors. Let L be the operator on state vectors defined by evolving (H1) & (H2) to some fixed time t. (5)

$$\therefore \quad Ls \equiv L(\theta,p)^T := \mathcal{H}^t(\theta,p)$$

Also let F denote the "momentum sign flip operator" so that

$$Fs \equiv F(\theta,p)^T := (\theta,-p).$$

Note that our previous results on (H1) & (H2) imply

$$L^{-1}s = FLFs \quad \text{(Reversability)}$$

$$\left|\frac{\partial Ls}{\partial s}\right| = 1 \quad \text{(vol preserving)}$$

$$H(Ls) = H(s) \quad \text{(conservation of energy)}$$

we also have

$$F^{-1}s = Fs$$

$$\left|\frac{\partial Fs}{\partial s}\right| = 1 \quad \text{(vol preserving)}$$

$$H(Fs) = H(s).$$

Finally define the "momentum Randomizing operator" R_β (parameter $\beta \in [0,1]$) as follows

$$R_\beta s \equiv R_\beta(\theta,p) := (\theta, p\sqrt{1-\beta} + \sqrt{\beta}\,p^*)$$

$$p^* \sim N(0, \mathrm{diag}(m))$$

(6)

Now we can construct a markov chain $s^{(1)}, s^{(2)}, \ldots$ for state vectors

$$s^{(i)} \equiv (\theta^{(i)}, p^{(i)})$$

as follows.

Hamiltonian Markov Chain

for $i = 1, 2, \ldots$

This is basically a proposal.

$$\alpha := \min\left(\frac{\exp(-H(FLs^{(i)}))}{\exp(-H(s^{(i)}))}, 1\right)$$

$$\tilde{s} := \begin{cases} FLs^{(i)} & \text{if } u \le \alpha \\ s^{(i)} & \text{if } u > \alpha \end{cases}$$

$$s^{(i+1)} := R_\beta F \tilde{s}$$

then an additional Random walk step.

end

The key for showing this works (i.e. has invariant density $e^{-H(s)} = e^{-H(\theta,p)}$) is the Reversability of L & the vol preserving of L & F.

In this form it is hard to see what is going on. We can simplify the above algorithm when $\beta = 1$. Notice the momentum components of $s^{(i)}$ are eventually discarded so we can write the algorithm in terms of $\theta^{(i)}$ alone.

Hamiltonian Markov Chain $\beta = 1$

for $i = 1, 2, \ldots$

  simulate $p^* \sim \mathcal{N}(0, \mathrm{diag}(m))$

  $\tilde{s} := (\theta^{(i)}, p^*)$

  $$\alpha := \min\left(\frac{\exp(-H(L\tilde{s}))}{\exp(-H(\tilde{s}))}, 1 \right)$$

  $$\theta^{(i+1)} := \begin{cases} (L\tilde{s})_\theta & \text{if } u \le \alpha; \\ \tilde{s}_\theta & \text{if } u > \alpha. \end{cases}$$

end

This is basically RBF $\tilde{s}$ from the previous step

where S_θ extracts the parameter coordinates from S.

Using the exact Hamiltonian evolution operator L, since $H(L\tilde{s}) = H(\tilde{s})$, α is always 1 so this MCMC always accepts the proposal.

To get an idea of what is going on expand out the Hamiltonian

$$H(\tilde{s}) = H(\theta^{(i)}, p^*)$$

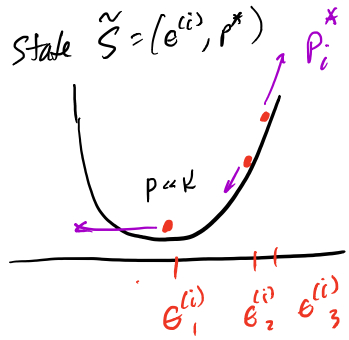
$$= -\log P(\theta^{(i)} \mid d) + \underbrace{\sum_{i=1}^{n} \frac{(P_i^*)^2}{2m_i}}$$

Potential energy (T)

Kinetic energy (K)

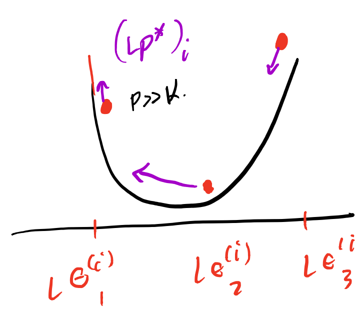
This will be a $\chi^2_{\frac{n}{2}}$ which is how $-\log P(\theta^{(i)} \mid d)$ show approximately vary when $\theta^{(i)} \sim P(\theta \mid d)$

State $\tilde{S} = (\theta^{(i)}, p^*)$



The Hamiltonian is evolved with L which where Kinetic & potential energy are exchanged.

$$H(L\tilde{s}) = -\log P(L\theta^{(i)} \mid d) + \sum_{i=1}^{n} \frac{(LP_i^*)^2}{2m_i}$$



$L\tilde{S} = (L\theta^{(i)}, Lp^*)$

This changes the Potential energy $-\log P(\theta \mid d)$ by an amount expected from samples $\theta \sim P(\theta \mid d)$

Leapfrog discretizing L

(9)

Tuning Parameters

(10)

* $\beta > 0$ can be useful (see Culpepper 2011)
* The typical suggestion is to set diag(m) (or a general cov matrix M) to the posterior cov matrix of θ.
* $L^{\text{steps}}, \varepsilon$ ...

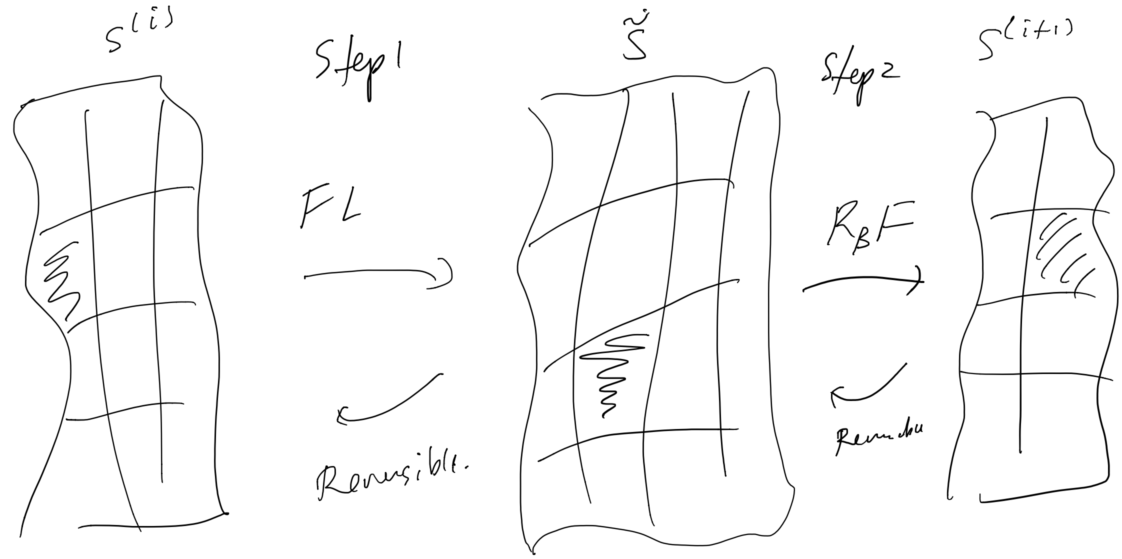
CMB applications

11

12

Look Ahead HMC

Showing detailed balance for HMC.



$$P(\tilde{s} \mid s^{(i)})\, \underbrace{e^{-H(s^{(i)})}}_{p(s^{(i)})}$$

$$= \begin{cases} \alpha\, P(s^{(i)}) & \tilde{S} = FL\, s^{(i)} \\ (1-\alpha) P(s^{(i)}) & \tilde{S} = s^{(i)} \end{cases}$$

Vol preserving

$P(s^{(i)} \in A_p)$

$\approx Vol(A_p)\, e^{-H(s^{(i)})}$

$= Vol(B_p)\, e^{-H(s^{(i)})}$

$$= P(s^{(i)} \mid \tilde{s}) P(\tilde{s})$$

$$P(s^{(i+1)} \mid \tilde{s}) P(\tilde{s})$$

$$= P_\beta\left(s^{(i+1)} + \sqrt{1-\beta} F \tilde{S}\right) P(\tilde{s})$$

$$\stackrel{?}{=} P_\beta\left(\tilde{S} + \sqrt{1-\beta}\, F\, s^{(i+1)}\right) P\left(s^{(i+1)}\right)$$

$$p(\tilde{s})\, e^{-\frac{\left|s_p^{(i+1)} - \sqrt{1-\beta}\, \tilde{S}_p\right|^2}{2\beta m}} = e^{-\frac{|\tilde{S}_p|^2}{2m}}\, e^{-\frac{|s_p^{(i+1)}|^2}{2m\beta}}\, e^{-\frac{\sqrt{1-\beta}}{\beta m} s_p^{(i+1)} \tilde{S}_p}$$

$$e^{-\frac{1-\beta}{\beta} \frac{|\tilde{S}_p|^2}{m}}$$