MCMC (Metropolis - Hastings)

you give me π

I want to construct Markov transition operator T
so $\pi T = \pi$

(we really want $\frac{1}{N} \sum_{k=1}^{N} f(X^{(k)}) \longrightarrow \int f(x)\pi(dx)$)

Metropolis Hastings
Pick some transition density $q(y|x)$
If q preserves π then you're done
$\left(\int q(y|x)\pi(dx) = \pi(y) \right)$

Usually you won't be able to pick q preserving π

How can we use q to generate a Markov chain
that does preserve π ?

Given $X^{(k)}$ generate $X^{(k+1)}$ as follows:

1. Generate $Y^{(k+1)} \sim q(y|X^{(k)})$

2. with probability $P_{acc} = \min\left\{1, \frac{\pi(Y^{(k+1)})\, q(X^{(k)}|Y^{(k+1)})}{\pi(X^{(k)})\, q(Y^{(k+1)}|X^{(k)})}\right\}$

set $X^{(k+1)} = Y^{(k+1)}$

otherwise set

$$X^{(k+1)} = X^{(k)}$$

Note that if q is in detailed balance w.r.t. π then $P_{acc} = 1$

(recall detailed balance is $q(y|x)\pi(x) = q(x|y)\pi(y)$)

Simple choice: $q(y|x) = \frac{e^{-(y-x)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} = q(x|y)$

$$y = x + N(0,\sigma^2)$$

in this case

$$P_{acc} = \min\left\{1, \frac{\pi(Y^{(k+1)})}{\pi(X^{(k)})}\right\}$$

Notice the symmetry

$$q(y|x) P_{acc}(x,y)\pi(x) = q(y|x)\min\left\{1, \frac{\pi(y)\, q(x|y)}{\pi(x)\, q(y|x)}\right\}\pi(x)$$

$$= \min\left\{q(y|x)\pi(x);\ \pi(y) q(x|y)\right\}$$

$$= q(x|y) P_{acc}(y,x)\pi(y)$$

$$Tf(x) = E[f(X^{(1)}) | X^{(0)} = x]$$

$$= f(x) P_{rej}(x) + \int f(y) P_{acc}(x,y) q(y|x)\, dy$$

$$p_{rej}(x) = \int (1 - p_{acc}(x,z))\, q(z|x)\, dz$$

(could write

$$p(y|x) = \delta(y-x)\, p_{rej}(x) + p_{acc}(x,y)\, q(y|x)$$

$$Tf(x) = \int f(y)\, p(y|x)\, dx$$)

Recall in terms of transition operator detailed balance requires

$$\int g(x)\, Tf(x)\, \pi(dx) = \int f(x)\, Tg(x)\, \pi(dx)$$

$\forall\, g, f$

$$\int g(x)\, Tf(x)\, \pi(dx) = \int g(x) f(x)\, p_{rej}(x)\, \pi(dx)$$

$$+ \iint f(y)\, q(y|x)\, p_{acc}(x,y)\, g(x)\, \pi(x)\, dx\, dy$$

$$= \int g(x) f(x)\, p_{rej}(x)\, \pi(dx)$$

$$+ \int f(y) \int g(x)\, q(x|y)\, p_{acc}(y,x)\, dx\, \pi(y)\, dy$$

$$= \int f(x)\, Tg(x)\, \pi(dx)$$

So M-H chain is in detailed balance w.r.t. π

Generators

For a discrete time process the generator is

$$\mathcal{L} = T - I$$

note

$$\pi T = \pi \iff \pi \mathcal{L} = 0$$

A very useful decomposition

$$f(X^{(t)}) = f(X^{(t-1)}) + \Big(\underbrace{E[f(X^{(t)}) \mid X^{(t-1)}] - f(X^{(t-1)})}_{\mathcal{L} f(X^{(t-1)})}\Big) + \Big(f(X^{(t)}) - E[f(X^{(t)}) \mid X^{(t-1)}]\Big)$$

$$\dots$$

$$= f(X^{(0)}) + \sum_{s=0}^{t-1} \mathcal{L} f(X^{(s)}) + \underbrace{\sum_{s=0}^{t-1} f(X^{(s+1)}) - E[f(X^{(s+1)}) \mid X^{(s)}]}_{M^{(t)}}$$

Let $\mathcal{F}_t$ be filtration generated by $X^{(t)}$

For $s<t$

$$E\left[f(X^{(t)}) - E[f(X^{(t)}) \mid X^{(t-1)}] \;\middle|\; \mathcal{F}_s \right]$$

$$= E[f(X^{(t)}) \mid \mathcal{F}_s] - E\left[E[f(X^{(t)}) \mid X^{(t-1)}] \;\middle|\; \mathcal{F}_s\right]$$

$$= E\left[E[f(X^{(t)}) \mid \mathcal{F}_{t-1}] \;\middle|\; \mathcal{F}_s\right]$$

$$- E\left[E[f(X^{(t)}) \mid X^{(t-1)}] \;\middle|\; \mathcal{F}_s\right]$$

$$= E\left[E[f(X^{(t)}) \mid X^{(t-1)}] \;\middle|\; \mathcal{F}_s\right] - \text{same} = 0$$

$(r<t)$

So $$E[M^{(t)} \mid \mathcal{F}_r] = \sum_{s=0}^{t-1} E\left[f(X^{(s+1)}) - E[f(X^{(s+1)}) \mid X^{(s)}] \;\middle|\; \mathcal{F}_r \right]$$

$$= \sum_{s=0}^{r-1} f(X^{(s+1)}) - E[f(X^{(s+1)}) \mid X^{(s)}]$$

$$= M^{(r)}$$

$M^{(t)}$ is a martingale

For a continuous time process

$$\mathcal{L}f(x) = \lim_{t \to 0} \frac{E[f(X^{(t)}) \mid X^{(0)} = x] - f(x)}{t}$$

$$= \lim_{t \to 0} \frac{T^t f(x) - f(x)}{t}$$

$$= \frac{d}{dt}\Big|_{t=0} T^t f(x)$$

ex) What is the generator of soln of

$$\frac{d}{dt} y = b(y), \quad y(0) = x \quad ?$$

$$T^t f(x) = f(y(t))$$

$$\mathcal{L} f(x) = f'(x)\, b(x)$$

on $\mathbb{R}^d$ $\quad \mathcal{L} f = \nabla f \cdot b$

Now let $u(t,x) = T^t f(x) = E[f(X^{(t)}) \mid X^{(0)} = x]$

Then $\partial_t u(t,x) = \lim_{h \to 0} \frac{T^{t+h} f(x) - T^t f(x)}{h}$

$$= \lim_{h \to 0} \frac{T^h T^t f(x) - T^t f(x)}{h}$$

$$= \mathcal{L} u(t,x)$$

$$u(0,x) = f(x)$$

$u(t,x) = e^{t\mathcal{L}} f(x)$ i.e. $T^t = e^{t\mathcal{L}}$

Diffusions

Recall the Euler discretization of the ODE

$$\frac{d}{dt} y = b(y) \qquad y(0) = y_0$$

$$y_h^{(k+1)} = y_h^{(k)} + h\, b(y_h^{(k)}) \qquad y_h^{(0)} = y_0$$

$h << 1$

or $$\frac{y_h^{(k+1)} - y_h^{(k)}}{h} = b(y_h^{(k)})$$

this is a consistent discretization of the ODE in the sense that:

$\| y_h^{(1)} - y(h) \| \longrightarrow 0$ faster than h as $h \to 0$

We could also check that the generator of $y_h^{(k)}$ converges to the generator of the ODE

$$\mathcal{L}_h f(y_0) = f(y_h^{(1)}) - f(y_0)$$

$$= f(y_0 + h\, b(y_0)) - f(y_0)$$

$$= h\, f'(y_0)\, b(y_0) + \frac{h^2}{2} f''(y_0)\, b^2(y_0)$$

$$+ O(h^3)$$

so $\frac{1}{h} \mathcal{L}_h f \to$ generator of ODE soln applied to f

Now let's add noise

$$X_h^{(k+1)} = X_h^{(k)} + h\, b(X_h^{(k)}) + h^\alpha \sigma(X_h^{(k)})\, \xi^{(k+1)}$$

$$X^{(0)} = x \qquad \xi^{(k)} \text{ i.i.d.} \quad P(\xi^{(k)} = \pm 1) = 1/2$$

$$(E[\xi^{(k)}] = 0)$$

$$\mathcal{L}_h f(x) = E[f(X^{(1)}) \mid X^{(0)} = x] - f(x)$$

$$= E\Big[f'(x)\big(h\, b(x) + \cancel{h^\alpha \sigma(x)\, \xi^{(1)}}\big)$$

$$+ \frac{f''(x)}{2}\big(h\, b(x) + h^\alpha \sigma(x)\, \xi^{(1)}\big)^2$$

$$+ \frac{f'''(x)}{6}\big(h\, b(x) + h^\alpha \sigma(x)\, \xi^{(1)}\big)^3$$

$$+ \text{higher order in } h \Big]$$

$$= h f'(x)\, b(x) + h^{2\alpha} \frac{f''(x)}{2} \sigma^2(x) \underbrace{E[(\xi^{(1)})^2]}_{=1}$$

$$+ \text{higher order than } h \text{ or } h^{2\alpha}$$

If I choose $\alpha > 1/2$ then

$$\frac{1}{h}\mathcal{L}_h f \longrightarrow \text{ODE generator}$$

If I choose $\alpha = 1/2$

then
$$\frac{1}{h}\mathcal{L}_h f(x) \xrightarrow[h\to 0]{} \underbrace{f'(x)\,b(x)} + \underbrace{\frac{f''(x)}{2}\sigma^2(x)}$$

$\longrightarrow$ generator for $y' = b(y)$ "diffusion term"

"drift" term

The limiting generator belongs to a continuous time process that we'll identify