MCMC : basics

So far we can:

- Sample some standard r.v.
  (Gaussians, log normals, uniform etc)
- we can compute Gaussian approximations
- we can do importance sampling.

Another technique: $\underbrace{\text{Markov Chain}}_{\text{generate the samples using a Markov Chain.}}$ $\underbrace{\text{Monte Carlo.}}_{\text{use samples to compute expected values}}$

$$E_p[f(x)] \approx \frac{1}{N_e} \sum_{j=1}^{N_e} f(x_j)$$

$j = 1, \dots N_e$

What is a Markov Chain?

Define a "transition probability": $q(x_k | x_{k-1})$

A Markov Chain is a random sequence $\{x_1, x_2, \dots, x_{N_e}\}$ where $x_j \sim q(x_j | x_{j-1})$.

Examples: (i) $x_j = x_{j-1} + v_j$

v_j iid $N(0,1)$

Defines a Markov Chain

(ii) $x_j = f(x_{j-1}) + \sqrt{\Delta t}\, v_j$, v_j iid $v_j \sim N(0, I)$

L'63 or L'95 dynamics

Defines a Markov Chain

(iii) $x_j \sim N(0,1)$

also defines a MC but transition distribution is independent of x_j

This is an importance sampler! ⇒

Properties of Markov Chains

- Markov Chain specified by initial value (initial probability) & transition probability.

- $x_n = f(x_{n-1}) + \text{noise}$: next state depends only on immediately preceding state

- Irreducible Markov Chain:

  for any x and y, the chain can get from x to y in $n \geq 0$ steps and with non-zero probability

- Notation: $P_{xy}^{(n)}$ - n step transition probability
  is probability that $p(X_{t+n} = y \mid X_t = x)$

  Above statement becomes:

  For any x, y there exists an $n \geq 0$ such that $P_{xy}^{(n)} > 0$

- Aperiodic Markov chain

  For each x, define a "period d_x" by the greatest common divisor of the numbers $n \geq 0$ for which $P_{xx}^{(n)} > 0$. If $d_x = 1$, then the chain is aperiodic.

- Stationary distribution:

  What if $p(X_t = y \mid X_0 = x_0)$ becomes independent of x_0?

  ↳ The chain "forgets" where it started

Suppose in addition that

$$p(X_t = y \mid X_0 = x_0)$$

also becomes independent of t. Then

$$p(X_t = y \mid X_0 = x_0) = \phi$$

a distribution independent of t or x_0.

We call ϕ the <u>stationary distribution.</u>

- If a Markov Chain is irreducible & aperiodic, then it has a stationary distribution. This describes convergence of the chain. ⊛
- <u>Idea:</u> to sample a given pdf ϕ, run a Markov Chain whose stationary distribution is ϕ.

⊛ <u>Ergodic Theorem:</u>

Let $X_0, X_1, \ldots, X_n$ be samples of an irreducible, aperiodic Markov chain with stationary distribution ϕ.

Then:
$$\frac{1}{n}\sum_{k=1}^{n} u(x_k) \to E_\phi[u(x)] \quad \text{as } n \to \infty.$$

How can we construct a Markov chain whose stationary distribution is ϕ?

(Perron - Frobenius) Theorem

A Markov chain is said to preserve $\overbrace{f}^{\text{the distribution}}$ if

$$x_k \sim f \,, \quad \text{then} \quad x_{k+1} \sim f.$$

If $\overbrace{g}^{\text{the chain}}$ preserves f and is aperiodic and irreducible, then f is the stationary distribution.

Result: this will make our life easier!
We don't have to design chains which converge to ϕ.
We only have to design chains that preserve ϕ.

Detailed balance:

Suppose that:

$$\underbrace{q(x_{k-1} \mid x_k)\, \phi(x_k) = q(x_k \mid x_{k-1})\, \phi(x_{k-1})}_{\text{must hold for all } x_k,\, x_{k-1}}$$

"probability of going from x_k to x_{k-1}" = "probability to go from x_{k-1} to x_k"

must hold for all x_k, x_{k-1}

Integrate x_{k-1}:

$$\int q(x_{k-1} \mid x_k)\, \phi(x_k)\, dx_{k-1} = \int q(x_k \mid x_{k-1})\, \phi(x_{k-1})\, dx_{k-1}$$

$$\phi(x_k) = \int q(x_k \mid x_{k-1})\, \phi(x_{k-1})\, dx_{k-1}$$

This says that if $x_{k-1} \sim \phi$, then $x_k \sim \phi$ and, hence, all subsequent samples are also samples from ϕ.

↳ We can design Markov Chains that satisfy detailed balance (this is easy, we will see many examples later) and all of them have ϕ as their stationary distribution.

↳ **Note:** there are Markov Chains that do <u>not</u> satisfy detailed balance, but whose stationary distribution is also ϕ.
Detailed balance implies that the chain preserves ϕ but it's not necessary!
(balance does not need detail).
We will mostly deal with Markov Chains that satisfy detailed balance.

<u>Metropolis - Hastings:</u>

A (large) family of MCMC.

Pick a proposal distribution $q(x_k \mid x_{k-1})$ to propose a sample x'.

Then: $\quad x' \sim q(\cdot \mid x_{k-1})$ ↙ propose.

set: $\quad x_k = x'$ with prob. $\cancel{\alpha(x_{k-1}, x')}$ $\alpha(x' \mid x_{k-1})$ ↙ move/accept

$\quad\quad x_k = x_{k-1}$ with prob $1 - \alpha(\cancel{x_{k-1}}, \cancel{x'})$ ↙ stay/reject.

where $\cancel{\alpha(x, y) = \min\left(1, \frac{\phi(y)}{\phi(x)} \frac{q(x \mid y)}{q(y \mid x)}\right)}$.

$$\alpha(x' \mid x_{k-1}) = \min\left(1; \frac{\phi(x')\, q(x_{k-1} \mid x')}{\phi(x_{k-1})\, q(x' \mid x_{k-1})}\right)$$

Drop index on x_{k-1}.

We have:

$$\alpha(x'|x) = \min\left(1, \frac{\phi(x')\,q(x|x')}{\phi(x)\,q(x'|x)}\right)$$

Suppose: $\phi(x)\,q(x'|x) > \phi(x')\,q(x|x')$

$$\alpha(x'|x) = \frac{\phi(x')\,q(x|x')}{\phi(x)\,q(x'|x)}$$

$$\alpha(x'|x)\,q(x'|x)\,\phi(x) = \phi(x')\,q(x|x') \qquad (*)$$

We also have

$$\alpha(x|x') = \min\left(1, \frac{\phi(x)\,q(x'|x)}{\phi(x')\,q(x|x')}\right) = 1$$

Put $\overset{=1}{\alpha(x|x')}$ on lhs of $(*)$

$$\cancel{\alpha(x|x')\,\phi(x')\,q(x|x') = \phi(x')\,q(x|x')} \qquad (**)$$

$$\cancel{(*) = (**)}$$

$$(*) \quad \underbrace{\underset{\text{accept } x'}{\alpha(x'|x)}\,\underset{\text{propose } x'}{q(x'|x)}\,\phi(x)}_{\text{Move from } x \text{ to } x'} = \underbrace{\underset{\text{accept } x}{\overset{=1}{\alpha(x|x')}}\,\underset{\text{propose } x}{q(x|x')}}_{\text{move from } x' \text{ to } x}\,\phi(x')$$

Move from x to x'

$$p(x'|x)\,\phi(x) = p(x|x')\,\phi(x')$$

$\Rightarrow$ detailed balance!

MH-algorithm

propose $x' \sim q(\cdot|x_{k-1})$

accept with prob. $\alpha(x'|x_{k-1})$ $\quad x_k = x'$

(if $u \sim U[0,1]$, accept if $\alpha > u$)

else: $x_k = x_{k-1}$.

This works for many choices of q.
Depending on how you chose proposal, your MChain can converge slowly or quickly.
We will study this and try out a few strategies.

Summary:

We pick a proposal distribution q.

We can use MH algorithm to generate a Markov Chain with stationary distribution ϕ.

We can compute averages:

$$E_\phi[u(x)] \approx \frac{1}{Ne} \sum_{j=1}^{Ne} u(x_j).$$

Note: $\{x_j\}$ are not independent.

If chain is short, the distribution may be different from ϕ.

↳ MCMC does not produce independent samples of ϕ, but it produces samples close enough to ϕ and which are close enough to being independent. (JG)