Monte Carlo simulations

Instead of simulating the dynamics of a physical system we will generate samples $\{X_k\}$ directly from the Boltzmann distribution $\pi(x) = \frac{1}{Z} e^{-\beta E(x)}$, $\beta = \frac{1}{k_B T}$

$$Z = \sum_x e^{-\beta E(x)} = e^{-\beta F}$$

Let's begin simple...

Ex Calculation of an integral

$$\int_a^b f(x)\,dx$$

Deterministic methods, e.g.

$$\int_a^b f(x)\,dx = \sum_{k=1}^{N} f(x_k)\, w_k + O\left(\frac{1}{N^p}\right)$$

e.g. $\frac{h}{2} f(x_1) + hf(x_2) + \dots + hf(x_{N-1}) + \frac{h}{2} f(x_N) + O\left(\frac{1}{N^2}\right)$

$x_k = a + (k-1)h$

In d-dimensions: Error $\sim O\left(\frac{1}{N^{p/d}}\right)$

Monte Carlo calculation: Draw $X_k \sim U[a,b]$

$$\int_a^b f(x)\,dx = \frac{b-a}{N} \sum_{k=1}^{N} f(x_k) + O\left(\frac{1}{\sqrt{N}}\right)$$

crude MC wins in high dimensions

Importance sampling:

Draw $x_k \sim g(\cdot)$ instead of uniformly

$$\int_a^b f(x)\,dx = \int_a^b \frac{f(x)}{g(x)}\, g(x)\,dx \approx \frac{1}{N}\sum_k \frac{f(x_k)}{g(x_k)}$$

works best if $f(x)$ ang $g(x)$ are similar.

If we choose $g(x) \propto f(x)$ all terms in the sum are equal $\Rightarrow$ zero-variance estimate! (Only possible if $f(x) \geq 0$.)

This optimal choice requires knowledge of

$$g(x) = \frac{f(x)}{\int_a^b f(x)\,dx}$$ i.e. the result we are after!

So, not useful in practice.

Still, we should choose $g(x)$ to be similar to $f(x)$.

Note: need $g(x) > 0$ whenever $f(x) \neq 0$

In many cases we are interested in estimating averages

$$\langle A \rangle = \int A(x)\pi(x)\,dx \approx \frac{1}{N}\sum_{k=1}^{N} A(x_k) = \bar{A}_N\,, \quad x_k \sim \pi(\cdot)$$

Often slowly varying   Often sharply peaked

<u>Law of large numbers:</u> $\bar{A}_N \xrightarrow{N\to\infty} \langle A \rangle = \mu_A$

$$\lim_{N\to\infty} P(|\bar{A}_N - \mu_A| > \varepsilon) = 0 \qquad \forall \varepsilon > 0 \quad \text{(Weak law)}$$

$$P(\lim_{N\to\infty} \bar{A}_N = \mu_A) = 1 \qquad \text{(Strong law)}$$

Can be proved under various conditions.
For independent identically distributed (i.i.d) samples the requirements are $\langle A(x_i)\rangle = \mu_A$ is finite for the weak law and $\langle |A(x_i)| \rangle < \infty$ for the strong law.

<u>Central limit theorem</u>

$$\sqrt{N}(\bar{A}_N - \mu_A) \sim N(0, \sigma^2) \quad \text{when } N \to \infty$$

for iid samples $\underbrace{\sigma^2}_{\text{with finite variance}} = \mathrm{Var}\, A = \langle (A-\mu_A)^2 \rangle = \langle A^2 \rangle - \langle A \rangle^2$

How do we sample from $\pi(\cdot)$?
There are many techniques for sampling from elementary distributions.
But for a Boltzmann distribution $\pi(x) = \frac{1}{Z} e^{-\beta E(x)}$?
Where x maybe contains $100 - 10^6$ degrees of freedom?

Markov Chain Monte Carlo MCMC

Idea: Construct a Markov chain whose limiting distribution is π.

$$X_0 \to X_1 \to \dots \to X_N$$

$$\lim_{t\to\infty} P(x,t) = \pi(x)$$

Markov property: $P(X_N, X_{N-1}, \dots, X_1, X_0) =$

$$= P(X_N | X_{N-1}) P(X_{N-1}, \dots, X_0) =$$

$$= P(X_N | X_{N-1}) P(X_{N-1} | X_{N-2}) \cdots P(X_1 | X_0) P(X_0)$$

The Markov chain is fully described by the initial distribution $P(X_0)$ and the transition probability $P(X_{t+1}|X_t) \equiv W(X_{t+1}|X_t)$.
We assume that W does not depend on time: we have a time-homogeneous Markov chain.

The probability $P(x,t) \equiv P(x = X_t)$ of finding x at t then obeys a <u>Master equation</u>:

$$P(x,t+1) = \sum_{x'} W(x|x')P(x',t) \quad \text{or equivalently}$$

$$P(x,t+1) = P(x,t) + \sum_{x' \neq x} W(x|x')P(x',t) - W(x'|x)P(x,t)$$

Note: For continuous variables the sums $\sum_x$ have to be replaced by $\int dx$!

Convergence to a unique equilibrium distribution $\pi(x)$

We want that $\lim_{t \to \infty} P(x,t) = \pi(x)$.

$\Rightarrow$ The following requirements need to hold:

- π is invariant: $\pi(x) = \sum_{x'} W(x|x')\pi(x')$

  or thinking of π as a column vector $\pi = W\pi$

  Since $W(x|x) = 1 - \sum_{y \neq x} W(y|x)$ (the probability to stay at x)

  this is equivalent to the condition of

  global balance $\sum_{x'} W(x|x')\pi(x') - W(x'|x)\pi(x) = 0$

  We also require that π is a probability, $\pi(x) \geq 0$, $\sum_x \pi(x) = 1$.

- The Markov chain is irreducible:

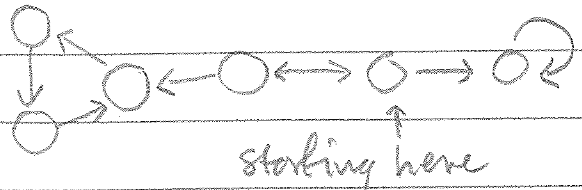
  Any state x (for which $\pi(x) > 0$) can be reached from any other state y, with $\pi(y) > 0$, in a finite number of steps.

  Mathematically: $P(x,t|y,0) > 0$ for some $t > 0$ and all x, y such that $\pi(x) > 0$, $\pi(y) > 0$.

- The chain must also be aperiodic

If these properties hold the chain is said to be ergodic.

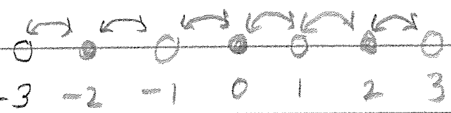
Ex



Not irreducible
$\Rightarrow$ reducible

No unique stationary distribution

Ex Random walk on all integers, choose left or right with equal probability, starting at 0 at $t=0$



Even sites may only be visited at even times
Odd   odd

Period 2, so not aperiodic.

In this example there is also another problem: the solution to the global balance eq is $\pi(x) = \text{const}$ but cannot be normalized $\sum_x \pi(x) = \infty \Rightarrow$ No equilibrium

Let T_x be the (random) time to return to x.

The Markov chain is recurrent if $P(T_x < \infty) = 1$ for all x and positive recurrent if $\langle T_x \rangle < \infty$. (a stronger condition)

For an irreducible, aperiodic Markov chain all states are positive recurrent iff a stationary distribution $\pi > 0$ exists, and then $\pi(x) = \frac{1}{\langle T_x \rangle}$.

Given a distribution $\pi(x)$ that we want to sample from, how can we choose the transition probabilities $W(x/x')$ such that π is the limiting equilibrium distribution?

Global balance $\sum_{x'} W(x|x')\pi(x') = \sum_{x'} W(x'|x)\pi(x)$

gives a lot of freedom to choose W.
Usually, it is replaced by a sufficient condition:

Detailed balance $W(x|x')\pi(x') = W(x'|x)\pi(x)$

$$\Leftrightarrow \quad \frac{W(x/x')}{W(x'/x)} = \frac{\pi(x)}{\pi(x')}$$ (still leaves a lot of freedom)

The resulting Markov chain is then called reversible.

In fact, if we define an inner product $\langle f, g\rangle = \sum_x \frac{f(x)g(x)}{\pi(x)}$
then W will be symmetric: $\langle f, Wg\rangle = \langle Wf, g\rangle$
so its eigenvalues will be real.
One can show that all eigenvalues are ≤ 1
with equality only for the unique equilibrium distribution as eigenvector.

Metropolis-Hastings algorithm

$W(x|x') = A(x|x')T(x|x')$ for $x' \neq x$

where $T(x|x')$ is a trial transition probability for $x' \to x$ and $A(x|x')$ is an acceptance probability for the move

The detailed balance condition becomes

$$\frac{W(x|x')}{W(x'|x)} = \frac{T(x|x')}{T(x'|x)} \frac{A(x|x')}{A(x'|x)} = \frac{\pi(x)}{\pi(x')}$$

Given $T(\cdot|\cdot)$ and $\pi(\cdot)$ we can solve for $A(\cdot|\cdot)$:

$$\frac{A(x|x')}{A(x'|x)} = \frac{T(x'|x)}{T(x|x')} \frac{\pi(x)}{\pi(x')}$$

In many cases $T(x|x') = T(x'|x)$ so that the trial transition probabilities cancel each other.
The generalization to include $T(\cdot|\cdot)$ is due to Hastings.
Then we only must require $T(x|y) > 0$ whenever $T(y|x) > 0$.
The most common choice for the acceptance probability is the Metropolis choice

$$A(x|x') = \min(1, A(x|x')/A(x'|x)) = \min\left(1, \frac{T(x'|x)\pi(x)}{T(x|x')\pi(x')}\right)$$

Another common choice is (Barker)

$$A(x|x') = \frac{A(x|x')}{A(x|x') + A(x'|x)} = \frac{T(x'|x)\pi(x)}{T(x'|x)\pi(x) + T(x|x')\pi(x')}$$

In terms of statistical efficiency (maximal number of accepted moves) the Metropolis choice is optimal among all other ones.

Note that the normalization constant drops out in the ratio $\pi(x)/\pi(x')$, which is a big advantage of MCMC, since it is often unknown.

For example, the Boltzmann probability $\pi(x) = \frac{1}{Z} e^{-E(x)/T}$

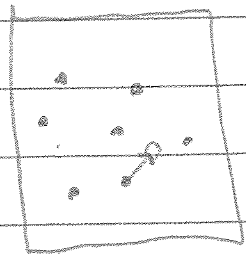
$\Rightarrow \pi(x)/\pi(x') = e^{-\Delta E/T}$, so it only involves the energy difference $\Delta E = E(x) - E(x')$ which typically is much faster to calculate than the total energy of a state $E(x)$.

The Metropolis-Hastings algorithm

Initialize x_0, then repeat for $t = 0, 1, \ldots$

1. Propose a trial move $x_t \to x^*$ with probability $T(x^* | x_t)$
2. compute $r = \dfrac{T(x_t | x^*)\,\pi(x^*)}{T(x^* | x_t)\,\pi(x_t)}$
3. If $r > 1$ or if $r > u$, where $u \sim U[0,1)$ then accept x^*:
   Set $x_{t+1} \leftarrow x^*$
   otherwise reject:
   Set $x_{t+1} \leftarrow x_t$ (no change)
4. Continue with next iteration, i.e. goto 1 unless some stopping criterion is met.

<u>Ex</u> N interacting particles in a box



$X = \{\bar{r}_i\}_1^N$

Trial moves: Pick one particle at random
and try to move it a random distance $\Delta \bar{r} \in [-\delta, \delta]^d$

The maximal step length δ can be tuned to optimize performance

Too large step length $\rightarrow$ most trial moves will be rejected

Too small step length $\rightarrow$ slow motion, long correlation times

Rule of thumb: Aim for an average acceptance probability
$\approx 0{,}25 - 0{,}5$

A MC sweep correspond to one trial attempt per particle

<u>Ex</u> Ising model: N interacting spins on a lattice

$X = \{S_i = \pm 1\}_1^N$

$$\begin{matrix} \uparrow & \downarrow & \downarrow & \uparrow \\ \uparrow & \uparrow & \uparrow & \downarrow \\ \uparrow & \uparrow & \uparrow & \uparrow \end{matrix}$$

$$H = -J \sum_{\langle ij \rangle} S_i S_j - h \sum_i S_i$$

Trial moves: Pick a random spin i
and try to flip it $S_i \leftarrow -S_i$

Energy change: $\Delta E = +J\, 2S_i \sum_{j\, nn\, i} S_j + h\, 2 S_i$

sum over nearest neighbours

Instead of drawing the spins randomly
one can go through the lattice sequentially.

Note: there are much smarter algorithms
for the Ising model.