Markov chain Monte Carlo methods

- The approximations of expectation so far have assumed that the samples are independent draws.
- This sounds good, but in high dimensions, we do not know how to get **good** independent samples from the distribution.
- MCMC methods drop this requirement.
- Basic intuition
  - If you have **finally** found a region of high probability, stick around for a bit, enjoy yourself, grab some more samples.

Markov chain Monte Carlo methods

- Samples are conditioned on the previous one (this is the Markov chain).
- MCMC is generally a good hammer for complex, high dimensional, problems.
- Main downside is that it is not "plug-and-play"
  - Doing well requires taking advantage to the structure of your problem
  - MCMC tends to be expensive (but take heart---there may not be any other solution, and at least your problem is being solved).

Metropolis Example

We want samples $z^{(1)}, z^{(2)}, \ldots.$

Again, write $p(z) = \tilde{p}(z)/Z$

Assume that $q\left(z \middle| z^{(prev)}\right)$ can be sampled easily

Also assume that $q(\;)$ is symmetric, i.e., $q\left(z_A \middle| z_B\right) = q\left(z_B \middle| z_A\right)$

For example, $q\left(z \middle| z^{(prev)}\right) \sim \mathbb{N}\left(z;\, z^{(prev)}, \sigma^2\right)$

Metropolis Example

While not_bored

{

Sample $q\left(z \middle| z^{(prev)}\right)$

Accept with probability $A\left(z, z^{(prev)}\right) = \min\left(1, \frac{\tilde{p}(z)}{\tilde{p}\left(z^{(prev)}\right)}\right)$

If accept, emit z, otherwise, emit $z^{(prev)}$.

}

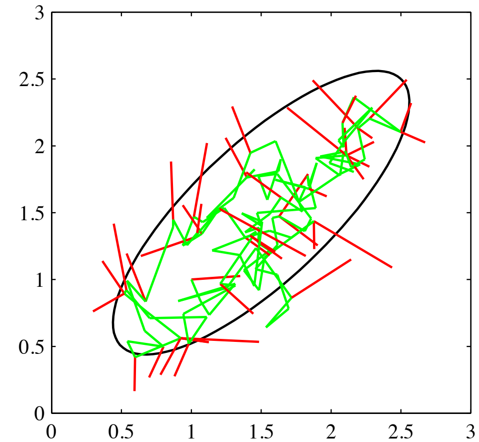
Metropolis Example

Note that

$$A\left(z, z^{(prev)}\right) = \min\left(1, \frac{\tilde{p}(z)}{\tilde{p}\left(z^{(prev)}\right)}\right) = \min\left(1, \frac{p(z)}{p\left(z^{(prev)}\right)}\right)$$

We do not need to normalize $p(z)$

Metropolis Example



Green follows accepted proposals
Red are rejected moves.

Markov chain view

Denote an initial probability distribution by $p\left(z^{(1)}\right)$

Define transition probabilities by:

$$T\left(z^{(prev)}, z\right) = p\left(z \middle| z^{(prev)}\right) \qquad \text{(a probability distribution)}$$

$T = T_m(\ )$ can change over time, but for now, assume that it it is always the same (homogeneous chain)

A given chain evolves from a sample of $p\left(z^{(1)}\right)$, and is an instance from an essemble of chains.

Stationary Markov chains

- Recall that our goal is to have our Markov chain emit samples from our target distribution.
- This implies that the distribution being sampled at time *t+1* is the same as that of time *t* (stationary).
- If our stationary (target) distribution is *p()*, then if imagine an ensemble of chains, they are in each state with (long-run) probability *p()*.
  - On average, a switch from s1 to s2 happens as often as going from s2 to s1, otherwise, the percentage of states would not be stable
- If our stationary (target) distribution is *p()*, what do the transition probabilities look like?

Detailed balance

- Detailed balance is defined by:

$$p(z)T(z,z') = p(z')T(z',z)$$

(We assume that $T(\bullet) > 0$)

- Detailed balance is a sufficient condition for a stationary distribution.
- Detailed balance is also referred to as reversibility.

Detailed balance implies stationary

$$p(z) = \sum_{z'} p(z')T(z',z) \qquad \text{(marginalization)}$$

If we have detailed balance, then

$$p(z')T(z',z) = p^{(prev)}(z)T(z,z')$$

So,

$$p(z) = \sum_{z'} p(z')\,T(z',z) = \sum_{z} p^{(prev)}(z)T(z,z') = p^{(prev)}(z')$$

Hence, detailed balance implies the distribution is stationary.

Detailed balance (cont)

- Detailed balance (for p()) means that *if* our chain was generating samples from *p()*, it would continue to due so.
  - We will address how it gets there shortly
- Does the Metropolis algorithm have detailed balance?

Metropolis Example

While not_bored

{

Sample $q\left(z \middle| z^{(prev)}\right)$

Accept with probability $A\left(z, z^{(prev)}\right) = \min\left(1, \dfrac{\tilde{p}(z)}{\tilde{p}\left(z^{(prev)}\right)}\right)$

If accept, emit z, otherwise, emit $z^{(prev)}$.

}

Same as $\dfrac{p(z)}{p\left(z^{(prev)}\right)}$

Metropolis Example

Recall that in Metropolis, $A(z,z') = \min\left(1, \frac{p(z)}{p(z')}\right)$

$$\begin{aligned}
p(z')q(z|z')A(z,z') &= q(z|z')\min(p(z'),p(z)) \\
&= q(z'|z)\min(p(z'),p(z)) \\
&= p(z)q(z'|z)\min\left(\frac{p(z')}{p(z)},1\right) \\
&= p(z)q(z'|z)\min\left(1,\frac{p(z')}{p(z)}\right) \\
&= p(z)q(z'|z)A(z',z)
\end{aligned}$$

Ergodic chains

- Different starting probabilities will give different chains
- We want our chains to converge (in the limit) to the same stationary state, regardless of starting distribution.
- Such chains are called ergodic, and the common stationary state is called the equilibrium state.
- Ergodic chains have a unique equilibrium.

When do our chains converge?

- Important theorem tells us that (for finite state spaces*) our chains converge to equilibrium under two relatively weak conditions.
- (1) Irreducible
  - We can get from any state to any other state
- (2) Aperiodic
  - The chain does not get trapped in cycles
- These are true for detailed balance which is sufficient, but not necessary for convergence.

*Infinite or uncountable state spaces introduces additional complexities.

Intuition behind ergodic chains

Let $p^{(t)}(z)$ be the distribution at some time (e.g., initial distribution)

Let $p^{*}(z)$ be the stationary distribution

Let $p^{(t)}(z) = p^{*}(z) - \mathrm{q}^{(t)}(z)$

Note that the elements of $p^{(t+1)}(z)$ and $p^{*}(z)$ sum to one, and thus the elements of $q(z)$ sum to zero.

Note also that $q(z)$ is not a probablity.

Intuition behind ergodic chains

Let $p^{(t)}(z)$ be the distribution at some time (e.g., initial distribution)

Let $p^{*}(z)$ be the stationary distribution

Let $p^{(t)}(z) = p^{*}(z) - q^{(t)}(z)$

$$
\begin{aligned}
p^{(t+1)}(z) &= \sum_{z'} p^{(t)}(z')\, T(z,z') \\
&= \sum_{z'} p^{*}(z')\, T(z,z') - \sum_{z'} q^{(t)}(z')\, T(z,z') \\
&= p^{*}(z) - q^{(t+1)}(z)
\end{aligned}
$$

Intuition behind ergodic chains

$$
\begin{aligned}
p^{(t+1)}(z) &= \sum_{z'} p^{(t)}(z')\, T(z,z') \\
&= \sum_{z'} p^{*}(z')\, T(z,z') - \sum_{z'} q^{(t)}(z')\, T(z,z') \\
&= p^{*}(z) - q^{(t+1)}(z)
\end{aligned}
$$

Claim that $\left|q^{(t+1)}(z)\right| < \left|q^{(t)}(z)\right|$