Summary so far

- Under reasonable (easily checked and/or arranged) conditions, our chains converge to an equilibrium state.
- Easiest way to prove (or check) that this is the case is to show detailed balance.
- To use MCMC for sampling a distribution, we simply ensure that our target distribution is the equilibrium state.
- Variations on MCMC are mostly about improving the speed of convergence for particular situations.

Summary so far

- The time it takes to get reasonably close to equilibrium (where samples come from the target distribution) is called "burn in" time.
  - I.E., how long does it take to forget the starting state.
  - There is no general way to know when this has occurred.
- The average time it takes to visit a state is called "hit time".
- What if we really want independent samples?
  - We can take every N^{th} sample (some theories about how long to wait exist, but it depends on the algorithm and distribution)

Metropolis-Hastings MCMC method

- Like Metropolis, but now q() is not symmetric.

Metropolis-Hastings MCMC method

While not_bored

{

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

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

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

}

Does Metropolis-Hastings converge to the target distribution?

If Metropolis-Hastings has detailed balance, then it converges to the target distribution under weak conditions.

(The converse is not true, but generally samplers of interest will have detailed balance).

How can we show this?

Does Metropolis-Hastings have detailed balance?

To show detailed balance we need to show

$p(z')q(z|z')A(z,z') = p(z)q(z'|z)A(z',z)$

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

Metropolis-Hastings comments

- Again it does not matter if we use unnormalized probabilities.
- It should be clear that the previous version, where q() is symmetric, is a special case.

Say a bit more about q(). It can be anything, but you need to specify the reverse move.

Reversible Jump MH

- Suppose the dimension of your problem is not known (e.g., you want to estimate the number of clusters).
- Sampling now includes "jumping" changes probability space
- Requires a modification to Metropolis Hastings
  - Reversible jump MCMC, Green 95, 03
- RJMCMC is only about sampling. It does not tell you the best number of dimensions (e.g., how many clusters).
  - This must come from either the prior or the likelihood.

Perhaps include formula?

More pedantic about model selection?

Gibbs sampling

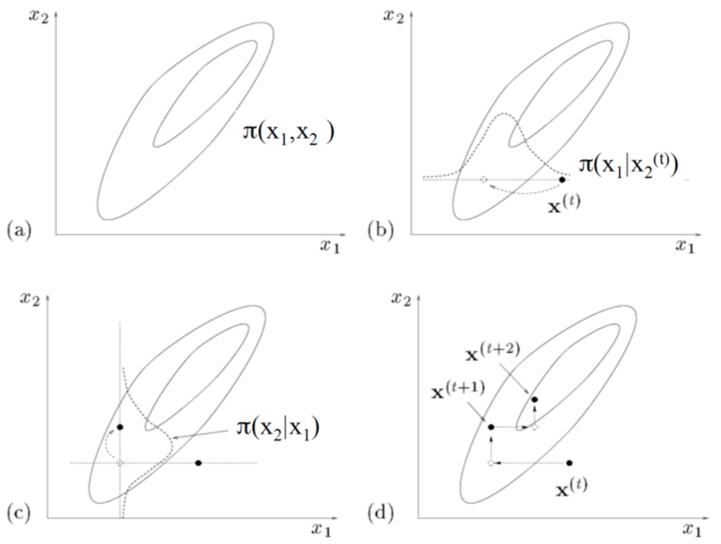
- Gibbs sampling is another special case of MH.
- You might notice that the transition function, T(), varies (cycles) over time.
  - This is a relaxation of our assumption used to provide intuition about convergence
  - However, it still OK because the concatenation of the T() for a cycle converge

Consider a set of N variables, $x_1,\ x_1,\ \ldots,x_N$, Gibbs says

Initialize $\left\{z_i^{(0)} : i = 1,\ldots,N\right\}$

Consider doing the image denoising example first, to help motivate this arguably odd looking formula.

While not_bored
{
    For i=1 to N
    {
        Sample $z_i^{(\tau+1)} \sim p\left(z_i \middle| z_1^{(\tau+1)},\ldots,z_{i-1}^{(\tau+1)}, z_{i+1}^{(\tau)},\ldots,z_M^{(\tau)}\right)$
        Always accept (emit $z = z_1^{(\tau+1)},\ldots,z_{i-1}^{(\tau+1)}, z_i^{(\tau+1)}, z_{i+1}^{(\tau)},\ldots,z_M^{(\tau)}$)
    }
}



(From Dellaert and Zhu tutorial)

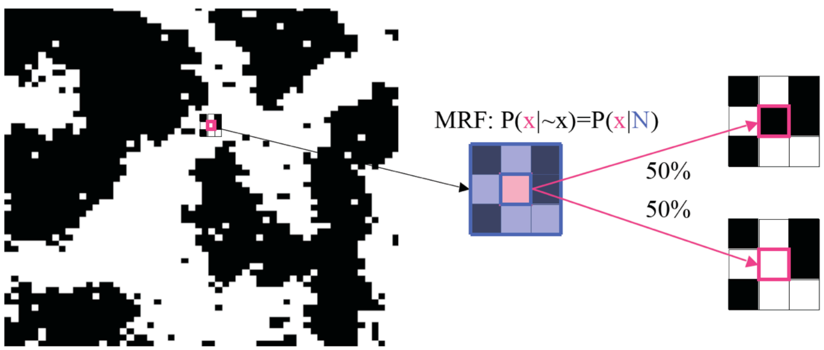
Examples of Gibbs

- If one can specify the conditional distributions so that they can be sampled, Gibbs is often a very good method.
- Typical examples include symmetric systems like the Markov random fields we had for images.
  - With a Markov property, the conditional probability can be quite simple.

Remember our method for fixing noise in images?

We iterated greedy assignment. Likelihood always goes up, and we head straight to a local minimum.

Examples of Gibbs



(From Dellaert and Zhu tutorial)

Examples of Gibbs

Weak Affinity to Neighbors

Strong Affinity to Neighbors

(From Dellaert and Zhu tutorial)

Gibbs as MH

To see Gibbs as MH, consider that if was MH, then our proposal distribution, q(), for a given variable, i, would be

$$q_i\left(\mathbf{z}\middle|\mathbf{z}^*\right) = p\left(z_i\middle|\mathbf{z}^*_{\backslash i}\right) \quad \text{and} \quad q_i\left(\mathbf{z}^*\middle|z\right) = p\left(z^*_i\middle|\mathbf{z}_{\backslash i}\right)$$

And we have $\mathbf{z}_{\backslash i} = \mathbf{z}^*_{\backslash i}$ because only i changes.

Gibbs as MH

$$\begin{aligned}
\mathrm{A}\left(\mathbf{z}^*,\mathbf{z}\right) &= \frac{p\left(\mathbf{z}^*\right) q_i\left(\mathbf{z}\middle|\mathbf{z}^*\right)}{p\left(\mathbf{z}\right) q_i\left(\mathbf{z}^*\middle|\mathbf{z}\right)} \\
&= \frac{p\left(\mathbf{z}^*_{\backslash i}\right) p\left(z^*_i\middle|\mathbf{z}^*_{\backslash i}\right) q_i\left(\mathbf{z}\middle|\mathbf{z}^*\right)}{p\left(\mathbf{z}_{\backslash i}\right) p\left(z_i\middle|\mathbf{z}_{\backslash i}\right) q_i\left(\mathbf{z}^*\middle|\mathbf{z}\right)} \\
&= \frac{p\left(\mathbf{z}^*_{\backslash i}\right) p\left(z^*_i\middle|\mathbf{z}^*_{\backslash i}\right) p\left(z_i\middle|\mathbf{z}^*_{\backslash i}\right)}{p\left(\mathbf{z}_{\backslash i}\right) p\left(z_i\middle|\mathbf{z}_{\backslash i}\right) p\left(z^*_i\middle|\mathbf{z}_{\backslash i}\right)} \\
&= 1
\end{aligned}$$

$$q_i\left(\mathbf{z}\middle|\mathbf{z}^*\right) = p\left(\mathbf{z}_i\middle|\mathbf{z}^*_{\backslash i}\right)$$

$$\text{and} \quad q_i\left(\mathbf{z}^*\middle|\mathbf{z}\right) = p\left(\mathbf{z}^*_i\middle|\mathbf{z}_{\backslash i}\right)$$

$$\text{and} \quad \mathbf{z}_{\backslash i} = \mathbf{z}^*_{\backslash i}$$