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).

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.

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.

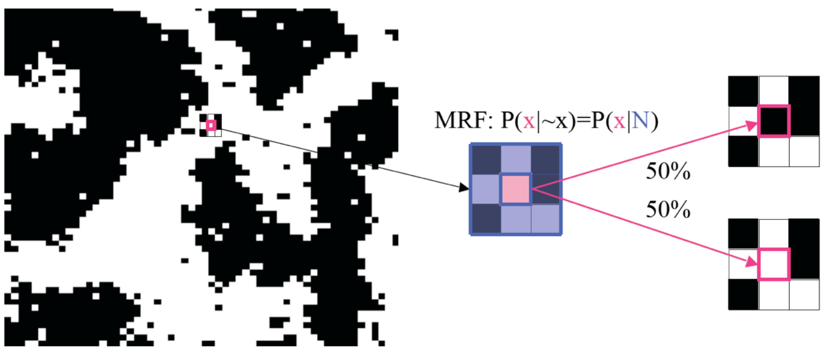
Gibbs sampling

- Gibbs sampling is special case of MH.
- The proposal distribution will be cycle over $p\left(z_n \middle| \left\{ z_{i \neq n} \right\}\right)$
- 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

Examples of Gibbs

- If one can specify the conditional distributions in a way that they can be sampled, Gibbs can be 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.

Examples of Gibbs



(From Dellaert and Zhu tutorial)

Examples of Gibbs

Weak Affinity to Neighbors

Strong Affinity to Neighbors

(From Dellaert and Zhu tutorial)

Consider a set of N variables, $z_1,\ z_1,\ ....,z_N$, Gibbs says

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

While not_bored

{

For i=1 to N

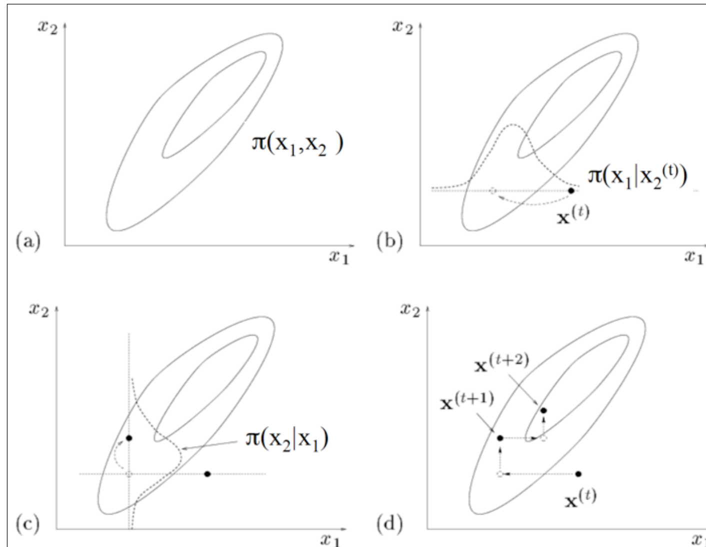
{

Sample $z_i^{(\tau+1)} \sim p\left(z_i \middle| z_1^{(\tau+1)},\dots,z_{i-1}^{(\tau+1)},z_{i+1}^{(\tau)},\dots,z_M^{(\tau)}\right)$

Always accept (emit $z = z_1^{(\tau+1)},\dots,z_{i-1}^{(\tau+1)},z_i^{(\tau+1)},z_{i+1}^{(\tau)},\dots,z_M^{(\tau)}$)

}

}



(From Dellaert and Zhu tutorial)

Gibbs as MH

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

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

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

Gibbs as MH

$$\mathrm{A}\left(\mathbf{z}^*, \mathbf{z}\right) = \min\left(1, \frac{p\left(\mathbf{z}^*\right) q_i\left(\mathbf{z}\middle|\mathbf{z}^*\right)}{p\left(\mathbf{z}\right) q_i\left(\mathbf{z}^*\middle|\mathbf{z}\right)}\right) \qquad \text{(def'n of A())}$$

$$= \min\left(1, \frac{p\left(\mathbf{z}^*_{\setminus i}\right) p\left(z^*_i\middle|\mathbf{z}^*_{\setminus i}\right) q_i\left(\mathbf{z}\middle|\mathbf{z}^*\right)}{p\left(\mathbf{z}_{\setminus i}\right) p\left(z_i\middle|\mathbf{z}_{\setminus i}\right) q_i\left(\mathbf{z}^*\middle|\mathbf{z}\right)}\right) \qquad \text{(def'n of ``bar'')}$$

$$= \min\left(1, \frac{p\left(\mathbf{z}^*_{\setminus i}\right) p\left(z^*_i\middle|\mathbf{z}^*_{\setminus i}\right) p\left(z_i\middle|\mathbf{z}^*_{\setminus i}\right)}{p\left(\mathbf{z}_{\setminus i}\right) p\left(z_i\middle|\mathbf{z}_{\setminus i}\right) p\left(z^*_i\middle|\mathbf{z}_{\setminus i}\right)}\right) \qquad \text{(Gibbs)}$$

$$= \min\left(1, 1\right) \qquad \text{(cancel colors using } \mathbf{z}^*_{\setminus i} = \mathbf{z}_{\setminus i} \text{, as only } z_i \text{ changes)}$$

$$= 1$$

Exploring the space

- Algorithms like Metropolis-Hastings exhibit "random walk behavior" if the step size (proposal variance) is small
- If the step size is too big, then you get rejected too often
- Adaptive methods exist (see slice sampling in Bishop)
- Another approach is to combine samplers with different properties

Combined samplers

1. Initialise $x^{(0)}$.
2. For $i = 0$ to $N - 1$
   - Sample $u \sim \mathcal{U}_{[0,1]}$.
   - If $u < \nu$

     Apply the MH algorithm with a global proposal.
   - else

     Apply the MH algorithm with a random walk proposal.