# STAT 530: Metropolis-Hastings Algorithm

Mar. 22, 2010

## Recap: given $\pi$ seek $T$ such that $\pi T = \pi$

E.g., $\pi$ is posterior distribution
(whole, or full conditional)

**Limitation: Implementing $T$**

- can't involve evaluating $\pi$
- but can involve evaluating $\pi_i/\pi_j$

Strategy: Start with any transition $J$, and 'fix it' to have $\pi$ as the stationary distribution

## MH transition $T$ defined by

**Given $Z^{(t)} = a$:**

1. Simulate $b$ according to $J_a$.
2. Evaluate

$$k = \min\left\{\frac{\pi_b}{\pi_a}\frac{J_{ba}}{J_{ab}}, 1\right\}$$

3. Set

$$Z^{(t+1)} \leftarrow \begin{cases} b & \text{with probability } k \\ a & \text{with probability } 1-k \end{cases}$$

## Prove it works

For any $a, b$ need $\pi_a T_{ab} = \pi_b T_{ba}$

## More 'continuous-looking' version

**Given $z^{(t)}$:**

1. Simulate $z^*$ according to $J(z^{(t)}; \cdot)$
2. Evaluate

$$k = \min\left\{\frac{\pi(z^*)}{\pi(z^{(t)})}\frac{J(z^*; z^{(t)})}{J(z^{(t)}; z^*)}, 1\right\}$$

3. Set

$$z^{(t+1)} \leftarrow \begin{cases} z^* & \text{with probability } k \\ z^{(t)} & \text{with probability } 1-k \end{cases}$$

## When **won't** this work well?

High autocorrelations, i.e., $z^{(t)}$, $z^{(t+1)}$ typically close, because...

## Random walk Metropolis-Hastings (RWMH)

$J(z^{(t)}; \cdot) \equiv N(z^{(t)}, V)$

Convenient symmetry: $J(z^{(t)}; z^*) = J(z^*; z^{(t)})$

Simple interpretation:
add noise to current state, then accept or reject.

See demo...

## Mix and match...

For instance say going after $p(\theta|D) = p(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5|D)$

Perhaps the particular form of the likelihood and prior plus experimentation leads you to:

Update $\theta_1$ using RWMH with 'jump variance' $\tau_1^2$

Update $\theta_2$ using Gibbs sampling

Update $(\theta_3, \theta_4)$ using RWMH with 'jump variance' $\text{diag}(\tau_3^2, \tau_4^2)$

Update $\theta_5$ using MH with some other choice of $J$

## Independence sampler

In some cases it may be possible to come up with an approximation to the posterior (full conditional or whole) which can be directly sampled.

e.g. $p(\beta|D)$ arising in logistic regression, Poisson regression, etc. Take $J(\beta^{(t)}; \cdot) \equiv N(\hat{\beta}, \hat{V})$.

## RWMH and parameterization

Say $z^* \sim N(z^{(t)}, \tau^2)$ doesn't work particularly well.

Might consider 'adding noise' in a different parameterization. E.g. $\log z^* \sim N(\log z^{(t)}, \tau^2)$

May well improve (or worsen) performance. But have to be careful...