Target vs. Proposal Distributions

Pretend this proposal distribution allows **good mixing**. What happens if we change it?
"White noise" appearance is a sign of good mixing.

I used the program Tracer to create this plot: http://tree.bio.ed.ac.uk/software/tracer/
Target vs. Proposal Distributions

Proposal distributions with smaller variance...

**Disadvantage:** robot takes smaller steps, more time required to explore the same area

**Advantage:** robot seldom refuses to take proposed steps
If step size is too small, large-scale trends will be apparent.
Target vs. Proposal Distributions

Proposal distributions with larger variance...

Advantage: robot can potentially cover a lot of ground quickly

Disadvantage: robot often proposes a step that would take it off a cliff, and refuses to move
Chain is spending long periods of time “stuck” in one place.

“Stuck” robot is indicative of step sizes that are too large (most proposed steps would take the robot “off the cliff”).
Metropolis Algorithm

\[ R = \min \left\{ \frac{p(D|\theta^*)p(\theta^*)}{p(D|\theta)p(\theta)}, 1 \right\} \]

probability of accepting proposal "move"

being a probability, maximum value of \( R \) must be 1

Posterior ratio


symmetrical -> Metropolis
Hastings ratio

\[
\left(\frac{1}{2}\right)^{(1)} = \frac{1}{2}
\]

\[
\left(\frac{1}{2}\right)^{(1)} = \frac{1}{2}
\]
Hastings ratio

\[ \left( \frac{2}{3} \right) (1) = \frac{2}{3} \]

\[ \left( \frac{1}{3} \right) (1) = \frac{1}{3} \]
Hastings ratio

\[
\left(\frac{2}{3}\right) \left[ \left(1 \right) \left(\frac{1/3}{2/3}\right) \right] = \frac{1}{3}
\]

\[
\left(\frac{1}{3}\right) \left[ \left(1 \right) \left(\frac{2/3}{1/3}\right) \right] = \frac{1}{3}
\]
Metropolis-Hastings Algorithm

\[ R = \min \left\{ \frac{p(D|\theta^*) p(\theta^*)}{p(D|\theta) p(\theta)} \left\| \frac{q(\theta|\theta^*)}{q(\theta^*|\theta)} \right\}, 1 \right\} \]

Posterior ratio

Hastings ratio

Heated chains act as scouts for the cold chain

**Cold chain robot** can easily make this jump because it is uphill

**Hot chain robot** can also make this jump with high probability because it is only slightly downhill
Hot chain and cold chain robots swapping places

Swapping places means both robots can cross the valley, but this is more important for the cold chain because its valley is much deeper.
So, what’s all this got to do with phylogenetics?
Moving through treespace

The Larget-Simon move

**Step 1:**
Pick 3 contiguous edges randomly, defining two subtrees, X and Y

Moving through treespace

The Larget-Simon move

**Step 1:**
Pick 3 contiguous edges randomly, defining two subtrees, X and Y

**Step 2:**
Shrink or grow selected 3-edge segment by a random amount
Moving through treespace

The Larget-Simon move

**Step 1:**
Pick 3 contiguous edges randomly, defining two subtrees, X and Y

**Step 2:**
Shrink or grow selected 3-edge segment by a random amount

![Diagram showing the Larget-Simon move with two subtrees X and Y]
Moving through treespace

The Larget-Simon move

**Step 1:**
Pick 3 contiguous edges randomly, defining two subtrees, X and Y

**Step 2:**
Shrink or grow selected 3-edge segment by a random amount

**Step 3:**
Choose X or Y randomly, then reposition randomly
Moving through treespace

The Larget-Simon move

**Step 1:**
Pick 3 contiguous edges randomly, defining two subtrees, X and Y

**Step 2:**
Shrink or grow selected 3-edge segment by a random amount

**Step 3:**
Choose X or Y randomly, then reposition randomly

Proposed new tree:
3 edge lengths have changed and the topology differs by one NNI rearrangement
Moving through treespace

Current tree
log-posterior = -34256

Proposed tree
log-posterior = -32519
(better, so accept)
Marginal Posterior Distribution of $\kappa$

Histogram created from a sample of 1000 kappa values.

95% credible interval

lower = 2.907
upper = 3.604

mean = 3.234
Prior Distributions

- **Discrete uniform** for topologies
  - exceptions becoming more common
- **Beta** for proportions
- **Gamma** for branch lengths and other parameters with support $[0, \infty)$
  - Exponential is common special case
- **Dirichlet** for state frequencies and GTR relative rates
For topologies, a discrete Uniform distribution is common.
Beta(a,b) for proportions

e.g. proportion of invariable sites

\[ \text{Beta}(a,b) \text{ for proportions} \]

\[ \text{e.g. proportion of invariable sites} \]

\[ \text{mean} = \frac{a}{a + b} \]

\[ \text{variance} = \frac{ab}{(a + b)^2(a + b + 1)} \]
For base frequencies, Dirichlet(a,b,c,d) distribution

Flat prior:
\[ a = b = c = d = 1 \]

Informative prior:
\[ a = b = c = d = 300 \]

(stereo pairs)
How much information is present in data about nucleotide frequencies?

Data:
- 26S rDNA gene
- 34 species of green algae
- 1688 aligned nucleotide sites

Analysis:
- 2,000,000 generations
- last 1000 samples (of 2000 total)
- HKY+gamma model
Gamma distributions

Gamma distributions are ideal for parameters that range from 0 to infinity (e.g. branch lengths)

\[ a = \text{shape} \]
\[ b = \text{scale} \]
\[ \text{mean} = ab \]
\[ \text{variance} = ab^2 \]