Proof of Metropolis

i) Does the algorithm give detailed balance, i.e.

$$\pi_i P_{ij} = \pi_j P_{ji}$$

Proof Let $X(i) = \vec{\theta}^{\,i}$ be a Markov Chain.

- Recall for Metropolis

$$\rho(\vec{\theta}^{\,i}, \vec{\theta}^{\,j}) = \min\left(1, \frac{\pi(\vec{\theta}^{\,j} | \vec{y})}{\pi(\vec{\theta}^{\,i} | \vec{y})}\right)$$

MH gave $\dfrac{\pi(\vec{\theta}^{\,j}|\vec{y})}{J(\vec{\theta}^{\,j}|\vec{\theta}^{\,i})} \Big/ \dfrac{\pi(\vec{\theta}^{\,i}|\vec{y})}{J(\vec{\theta}^{\,i}|\vec{\theta}^{\,j})}$

For non-sym. J

$$\rho(\vec{\theta}^{\,i}, \vec{\theta}^{\,j}) = \min\left(1, \frac{P(\vec{y}|\vec{\theta}^{\,j})\,\pi_0(\vec{\theta}^{\,j})}{P(\vec{y}|\vec{\theta}^{\,i})\,\pi_0(\vec{\theta}^{\,i})} \frac{J(\vec{\theta}^{\,i}|\vec{\theta}^{\,j})}{J(\vec{\theta}^{\,j}|\vec{\theta}^{\,i})}\right)$$

- So for Metropolis

$$P_{ij} = \rho(\vec{\theta}^{\,i}, \vec{\theta}^{\,j})\, J(\vec{\theta}^{\,j} | \vec{\theta}^{\,i}), \quad i \neq j$$

$$P_{ii} = 1 - \sum_{i \neq j}^{N_s} P_{ij}$$

- Note: If J is symmetric (J is Gaussian)

$$P_{ji} = \rho(\vec{\theta}^{\,j}, \vec{\theta}^{\,i})\, J(\vec{\theta}^{\,i} | \vec{\theta}^{\,j}) = \rho(\vec{\theta}^{\,j}, \vec{\theta}^{\,i})\, J(\vec{\theta}^{\,j} | \vec{\theta}^{\,i})$$

- Then

$$\pi_i P_{ij} = \pi_j P_{ji}$$

$$\rho(\vec{\theta}^{\,i}, \vec{\theta}^{\,j})\, J(\vec{\theta}^{\,j} | \vec{\theta}^{\,i})$$

$i \; ij \quad j \; ji$

$$\Rightarrow \quad \pi_i \, p(\vec{\theta}^i, \vec{\theta}^j) \, J(\vec{\theta}^j | \vec{\theta}^i) = \pi_j \, p(\vec{\theta}^j, \vec{\theta}^i) \, J(\vec{\theta}^i | \vec{\theta}^j)$$

$$\Rightarrow \quad \pi_i \, p(\vec{\theta}^i, \vec{\theta}^j) = \pi_j \, p(\vec{\theta}^j, \vec{\theta}^i)$$

$$\Rightarrow \quad \pi_i \min\left(1, \frac{\pi(\vec{\theta}^j | \vec{y})}{\pi(\vec{\theta}^i | \vec{y})}\right) = \pi_j \min\left(1, \frac{\pi(\vec{\theta}^i | \vec{y})}{\pi(\vec{\theta}^j | \vec{y})}\right)$$

- If $\pi(\vec{\theta}_i | y) < \pi(\vec{\theta}_j | y)$

$$\Rightarrow \quad \pi_i \cdot 1 = \pi_j \frac{\pi(\vec{\theta}^i | \vec{y})}{\pi(\vec{\theta}^j | \vec{y})}$$

- If $\pi_i = \pi(\vec{\theta}^i | \vec{y})$ + $\pi_j = \pi(\vec{\theta}^j | \vec{y})$

$$\Rightarrow \quad \pi(\vec{\theta}^i | \vec{y}) \cdot 1 = \pi(\vec{\theta}^j | \vec{y}) \cdot \frac{\pi(\vec{\theta}^i | \vec{y})}{\pi(\vec{\theta}^j | y)}$$

$$\Rightarrow \quad \pi(\vec{\theta}^i | \vec{y}) = \pi(\vec{\theta}^i | \vec{y}) \quad \checkmark$$

ii) Is the stationary distribution the posterior?

- Let $\pi_q = \pi(\vec{\theta}^q | \vec{y})$ + let

$$P_{q-1,q} = P(X_q = \vec{\theta}^q \mid X_{q-1} = \vec{\theta}^{q-1})$$

- Note: $\min(x, z) = x \min\left(1, \frac{z}{x}\right) = z \min\left(1, \frac{x}{z}\right)$

- If detailed balance holds

$$\pi(\vec{\theta}^{\,q-1}|\vec{y})\, P_{q-1,q} = \pi(\vec{\theta}^{\,q-1}|\vec{y})\, J(\vec{\theta}^{\,q}|\vec{\theta}^{\,q-1})\, \rho(\vec{\theta}^{\,q}, \vec{\theta}^{\,q-1})$$

note: $J(\vec{\theta}^{\,q-1}|\vec{\theta}^{\,q}) = J(\vec{\theta}^{\,q}|\vec{\theta}^{\,q-1})$

also $$\pi(\vec{\theta}^{\,q-1}|\vec{y})\, J(\vec{\theta}^{\,q}|\vec{\theta}^{\,q-1})\, \min\left(1, \frac{\pi(\vec{\theta}^{\,q}|\vec{y})}{J(\vec{\theta}^{\,q}|\vec{\theta}^{\,q-1})} \Big/ \frac{\pi(\vec{\theta}^{\,q-1}|\vec{y})}{J(\vec{\theta}^{\,q-1}|\vec{\theta}^{\,q})}\right)$$

$$\rightarrow = \underbrace{\pi(\vec{\theta}^{\,q-1}|\vec{y})\, J(\vec{\theta}^{\,q}|\vec{\theta}^{\,q-1})}_{=x}\, \min\left(1, \frac{\overbrace{\pi(\vec{\theta}^{\,q}|\vec{y}) \cdot J(\vec{\theta}^{\,q-1}|\vec{\theta}^{\,q})}^{z}}{\underbrace{\pi(\vec{\theta}^{\,q-1}|\vec{y}) \cdot J(\vec{\theta}^{\,q}|\vec{\theta}^{\,q-1})}_{x}}\right)$$

$$= \pi(\vec{\theta}^{\,q}|\vec{y})\, J(\vec{\theta}^{\,q-1}|\vec{\theta}^{\,q})\, \min\left(1, \frac{\pi(\vec{\theta}^{\,q-1}|\vec{y})\, J(\vec{\theta}^{\,q}|\vec{\theta}^{\,q-1})}{\pi(\vec{\theta}^{\,q}|\vec{y})\, J(\vec{\theta}^{\,q-1}|\vec{\theta}^{\,q})}\right)$$

$$= \pi(\vec{\theta}^{\,q}|\vec{y})\, J(\vec{\theta}^{\,q-1}|\vec{\theta}^{\,q})\, \rho(\vec{\theta}^{\,q-1}, \vec{\theta}^{\,q})$$

$$= \pi(\vec{\theta}^{\,q}|\vec{y})\, P_{q,q-1}$$

$\square$

$$\Rightarrow \boxed{\pi(\vec{\theta}^{\,q-1}|\vec{y})\, P_{q-1,q} = \pi(\vec{\theta}^{\,q}|\vec{y})\, P_{q,q-1}}$$

<u>Adaptive Metropolis</u> (AM)

- The proposal, J, has a large effect on sampling

- AM adapts the covariance for J by:
  - for every q steps

$$\underset{\sim}{V}_q = S_p \, \mathrm{cov}\left(\vec{\theta}^0, \vec{\theta}^1, \dots, \vec{\theta}^{q-1}\right) + \eta \underset{\sim}{I}_p , \quad 0 < \eta \ll 1$$

  - η is a "jitter" term
  - S_p is a scaling factor (we will use $S_p = 2.38^2 / p$)

- In practice, we cannot generate $\underset{\sim}{V}_q$ at every step explicitly.
- Instead, we sequentially update:

$$\underset{\sim}{V}_{q+1} = \frac{q-1}{q} \underset{\sim}{V}_q + \frac{S_p}{q} \Big[q \, \bar{\theta}^{q-1} \bar{\theta}^{q-1\,T} - (q+1) \, \bar{\theta}^{q} \bar{\theta}^{q\,T} + \vec{\theta}^{q} \vec{\theta}^{q\,T} + \eta \underset{\sim}{I}_p \Big]$$

where

$$\bar{\theta}^q = \frac{q}{q+1} \bar{\theta}^{q-1} + \frac{1}{q+1} \vec{\theta}^q$$

- In practice, we <u>only update</u> every q_0 steps