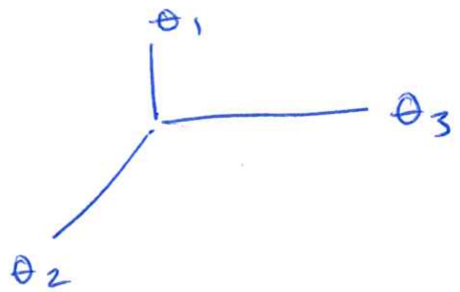
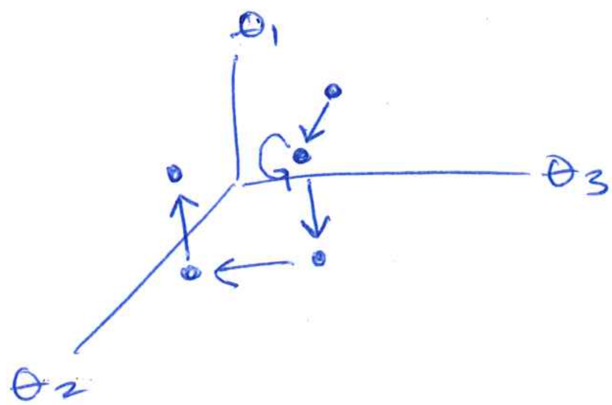
Recap:

* We want samples from $p(\theta_1, \ldots, \theta_n | y)$

* We view "parameter space" $\nearrow$ "state space" as a physical space: e.g. $n = 3$



* We view MCMC as a particle moving through parameter space:



* The "history" (histogram) of where the particle has been in $\underline{\theta}$ space approximates $p(\underline{\theta} | y)$.

similarly the history in θ_i approximates $p(\theta_i | y)$.

Metropolis-Hastings algorithm

Let $\pi(\theta)$ be the target distr.

1. sample $\theta^* \sim J(\theta \mid \theta^{(s)})$
2. ~~accep~~ compute the acceptance ratio:

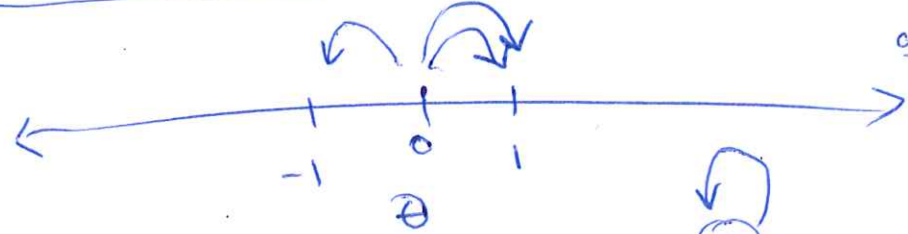
$$r = \frac{\pi(\theta^*)}{\pi(\theta^{(s)})} \times \frac{J(\theta^{(s)} \mid \theta^*)}{J(\theta^* \mid \theta^{(s)})}$$

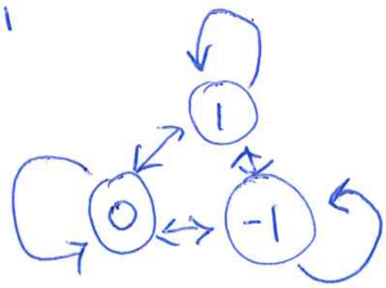
3. set $\theta^{(s+1)} = \theta^*$ w/ prob $\min(1, r)$
   otherwise $\theta^{(s+1)} = \theta^{(s)}$

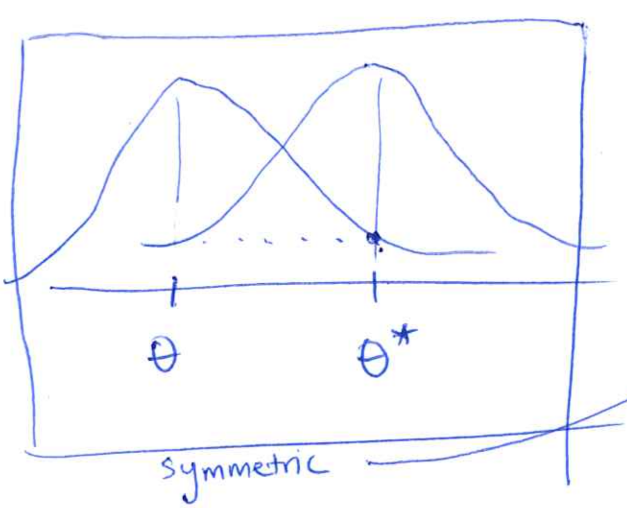
* We correct for asymmetry in the proposal.

Examples

asymmetric





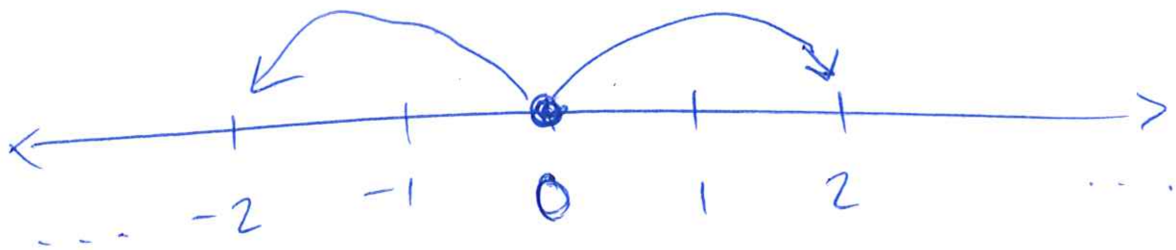


$$\exp\left\{\frac{-1}{2\sigma^2}(\theta - \theta^*)^2\right\}$$

symmetric

Examples

$\theta \in \mathbb{Z}$



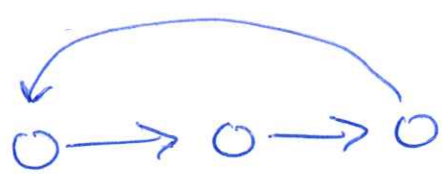
Let $J(\theta \mid \theta^{(s)}) = \theta^{(s)} \pm 2$

Our Markov chain is reducible.

ex) Imagine we cannot stay in the same state (in the ex above)

periodic

ex2



Ex Imagine discrete $\theta \in \left\{ \begin{matrix} A \\ B \\ C \end{matrix} \right.$

If we don't continue sampling state A

$Pr(\theta^{(s)} \in A) \longrightarrow 0$ as $s \to \infty$

NOT Recurrent

$\pi(\theta) :=$ stationary distr.

$p_o(\theta) :=$ posterior distr. "target" of MH MCMC.

Let θ be discrete r.v.

T.P. $\pi(\theta) = p_o(\theta)$

Put another way:

I want t.p. that if $pr(\theta^{(s)} = \theta) = p_o(\theta)$

then $pr(\theta^{(s+1)} = \theta) = p_o(\theta)$

because if p_o is <u>a</u> stationary distr., it is <u>the</u> stationary distr. by uniqueness.

Let θ_A & θ_b be two values of θ, s.t.

$$p_o(\theta_a) \cdot J_s(\theta_b | \theta_a) \geq p_o(\theta_b) \cdot J_s(\theta_a | \theta_b)$$

Then under MH, the probability $\theta^{(s)} = \theta_a$ and $\theta^{(s+1)} = \theta_b$ is given by

$$\underbrace{p_o(\cancel{\theta_a})}_{\text{prob ending up in } \theta_a} \cdot \underbrace{J_s(\cancel{\theta_b | \theta_a})}_{\text{prob proposing } \theta_b \text{ from } \theta_a} \cdot \underbrace{\frac{p_o(\theta_b) \cdot J_s(\theta_a | \theta_b)}{\cancel{p_o(\theta_a)} \cdot \cancel{J_s(\theta_b | \theta_a)}}}_{\text{prob accepting}}$$

$$= \boxed{p_o(\theta_b) \cdot J_s(\theta_a | \theta_b)}$$

another way: $pr(\theta^{(s)} = \theta_b, \theta^{(s+1)} = \theta_a)$

PROOF CONTINUED

$$\boxed{\begin{array}{l}\text{So } \forall\ \theta_a, \theta_b \\ pr(\theta^{(s)} = \theta_a, \theta^{(s+1)} = \theta_b) = pr(\theta^{(s)} = \theta_b, \theta^{(s+1)} = \theta_a)\end{array}} \quad \star$$

T. P. $Pr(\theta^{(s+1)} = \theta) = p_0(\theta)$

$$= pr(\theta^{(s)} = \theta)$$

<u>PROOF</u>

$$\begin{aligned} Pr(\theta^{(s+1)} = \theta) &= \sum_{\theta_a} Pr(\theta^{(s+1)} = \theta, \theta^{(s)} = \theta_a) \\ &= \sum_{\theta_a} Pr(\theta^{(s+1)} = \theta_a, \theta^{(s)} = \theta) \quad \text{by } \star \\ &= Pr(\theta^{(s)} = \theta) \end{aligned}$$

$\square$

In summary, if we choose $J(\theta^* | \theta^{(s)})$ such that the Markov chain is ergodic, it will have a unique stationary distr. that <u>is</u> the posterior we want to approximate <u>under</u> the Metropolis-Hastings algorithm.