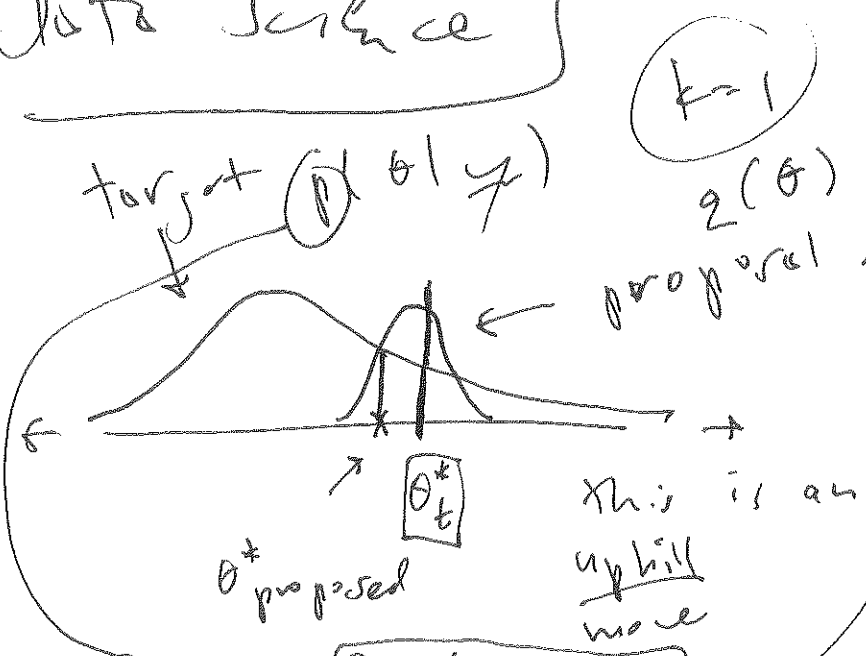


this Metropolis & Gibbs  
 fine: sampling; Big-Data  
 Data Science

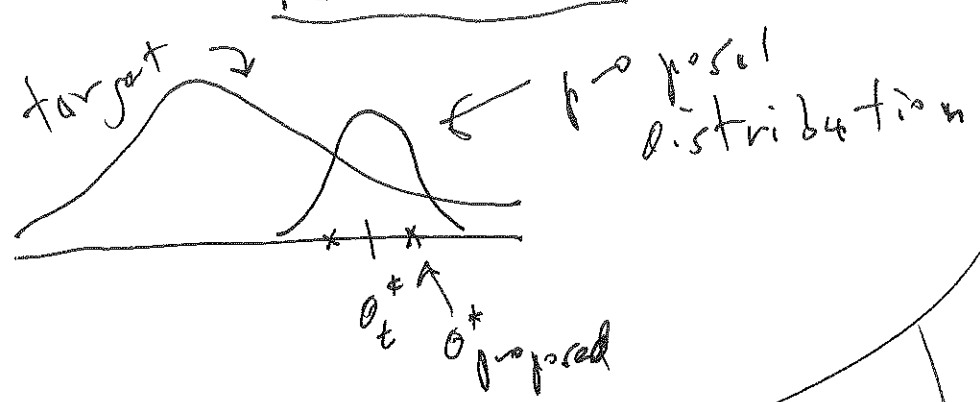
Metropolis  
 Sampling  
 AMS 206  
 20 Mar 18  
 ①



must be symmetric  
 PSD  $\sigma$   
 $N(\theta_t^*, \sigma^2)$

$$\frac{p(\theta_t^* \text{ proposed} | y)}{p(\theta_t^* | y)} = r(\theta_t^* \text{ proposed}, \theta_t^*)$$

rule 1: accept all uphill moves  
 $\theta_{t+1}^* = \theta_t^* \text{ proposed}$



$$p(\theta|y) = c p(\theta) \cdot \ell(\theta|y)$$

~~$$\frac{p(\theta_t^* \text{ proposed} | y)}{p(\theta_t^* | y)}$$~~

rule 2:  
 accept a downhill move with probability  
 $r(\theta_t^* \text{ proposed}, \theta_t^*)$   
 $\theta_{t+1}^* = \theta_t^* \text{ proposed}$

~~$$\frac{p(\theta_t^* \text{ proposed}) \cdot \ell(\theta_t^* \text{ proposed} | y)}{p(\theta_t^* | y) \cdot \ell(\theta_t^* | y)}$$~~

~~$$\frac{p(\theta_t^*) \cdot \ell(\theta_t^* | y)}{p(\theta_t^* \text{ proposed}) \cdot \ell(\theta_t^* \text{ proposed} | y)}$$~~

rule 2: if reject under rule 1,

(2)

$$\theta_{th}^* = \theta_t^*$$

if PSD  $\sigma$  is small,

too

converge  
target

sticky

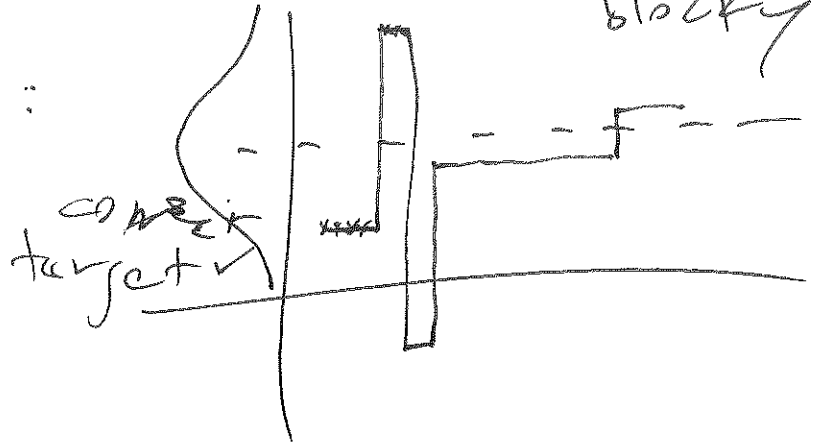


if PSD is big:

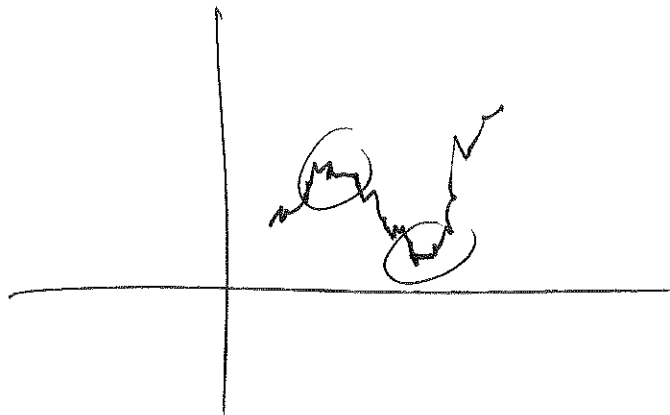
$\sigma$  too

converge  
target

blocky



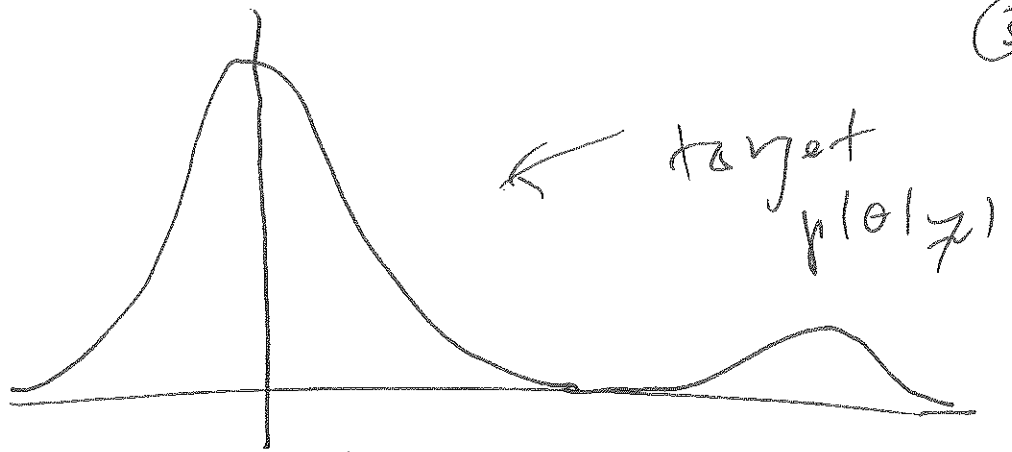
with optimal  
PSD



keep track of ~~rejection~~ acceptance rate of (k=1)  
algorithm  $\hat{\pi}$ ; optimal PSD +  $\hat{\pi} = 0.45$

Achilles Heel  
of MCMC

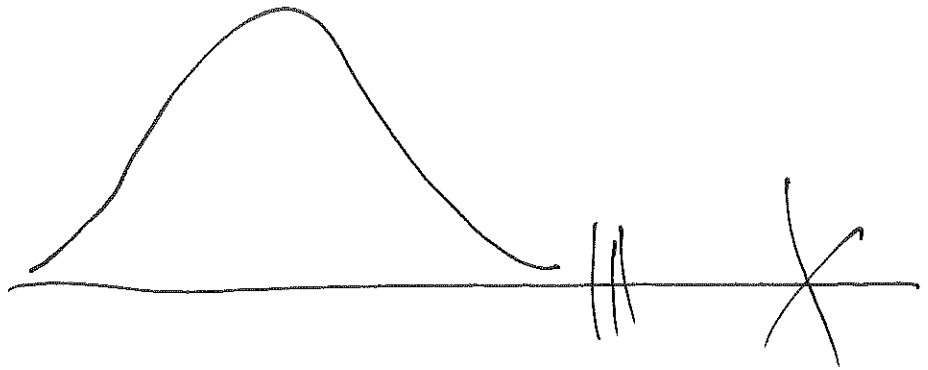
hard to sample



(3)

$n = 100,000$

start here  
many samples



Gibbs sampling

NB to note

$$\theta = (\mu, \sigma, \tau)$$

we want  $p(\mu, \sigma, \tau | z, B \dots)$ ; hard

Suppose  
con sample  
from

$$\left. \begin{aligned} & p(\mu | \tau, \sigma, r, B \dots) \\ & p(\sigma | \mu, \tau, r, B \dots) \\ & p(r | \mu, \sigma, B \dots) \end{aligned} \right\} \text{full-conditional dist. of } p(\sigma | \mu, \tau, r, B \dots)$$

intuitively

(1.1) sample  $\mu_1^*$

from  $p(\mu | \tau, \sigma_0, r_0)$

(1.2) sample  $\sigma_1^*$

from  $p(\sigma | \tau, \mu_1^*, r_0)$

(1.3) sample  $r_1^*$

from  $p(r | \mu_1^*, \sigma_1^*)$

	$\mu$	$\sigma$	$r$
0	$\mu_0$	$\sigma_0$	$r_0$
1	$\mu_1^*$	$\sigma_1^*$	$r_1^*$
2		$\sigma_2^*$	

(2.1)

(2.2) repeat

(2.3)

⋮ repeat

acceptance rate 1