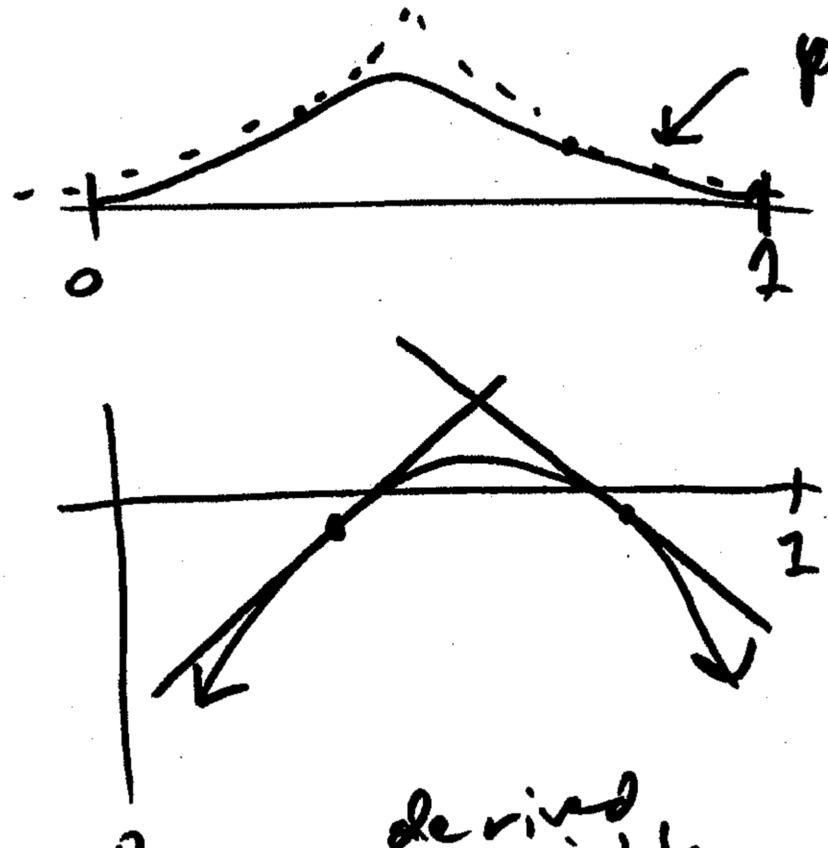
Handwritten notes 8 Feb



log concave density

derived variable

$$I(\theta^* \le .15) = \begin{cases} 1 & \text{if } \theta^* \le .15 \\ 0 & \text{else} \end{cases}$$

θ^*	$I(\theta^* \le .15)$
below .15	1
	⋮
	1
above .15	0
	⋮
	0

sort

31,200

mean

Monte Carlo data set

$$
\text{time} \quad \theta^k = \begin{bmatrix} \theta_1^k \\ \theta_2^k \\ \vdots \end{bmatrix}
$$

(iteration #)

we will allow θ_2^k to depend on θ_1^k, k

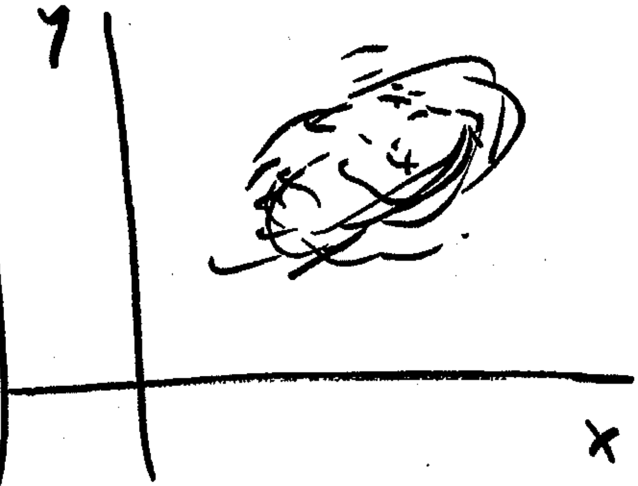
θ_3^k to depend on θ_2^k, etc.

$$
p(\theta \mid y) = c \, p(\theta) \, \ell(\theta \mid y)
$$

dim k

k-dimensional integral

$$\begin{pmatrix}\theta_0^*\\ \theta_1^*\\ \theta_2^*\\ \vdots\end{pmatrix} \qquad \begin{bmatrix} y_1 & x_1\\ \vdots & \vdots \\ y_n & x_n\end{bmatrix}$$



$$r = \frac{\hat{C}(x,y)}{\sqrt{\hat{V}(x)\hat{V}(y)}}$$

$$C(x,y) = E\left[(x-\mu_x)(y-\mu_y)\right]$$

$$r = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\left[\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2\right]\left[\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2\right]}}$$

sample correlation

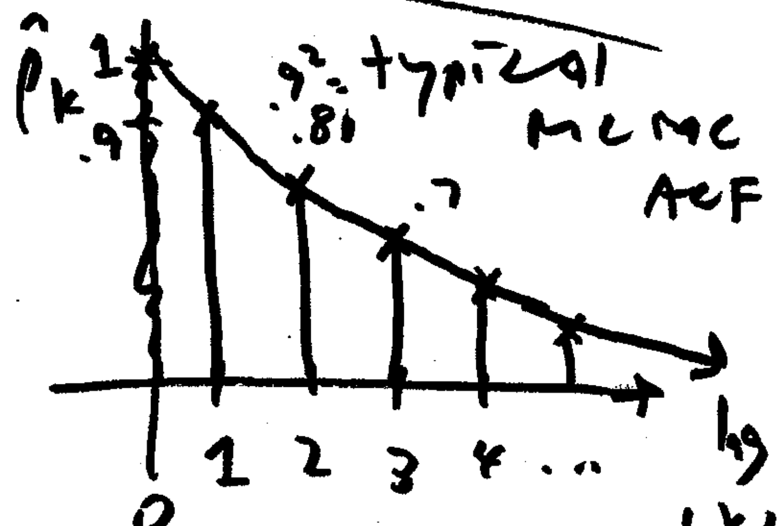
R: corr(x, y)

$$\begin{bmatrix} \theta^*_0 & \theta^*_1 \\ \theta^*_1 & \theta^*_2 \\ \theta^*_2 & \theta^*_3 \\ \vdots & \vdots \\ \theta^*_{m-1} & \theta^*_m \\ \cancel{\theta^*_m} & \cancel{\theta^*_{m+1}} \end{bmatrix}$$

→ calculate sample correlation (r) of these 2 columns: this is $\hat{\rho}_1$

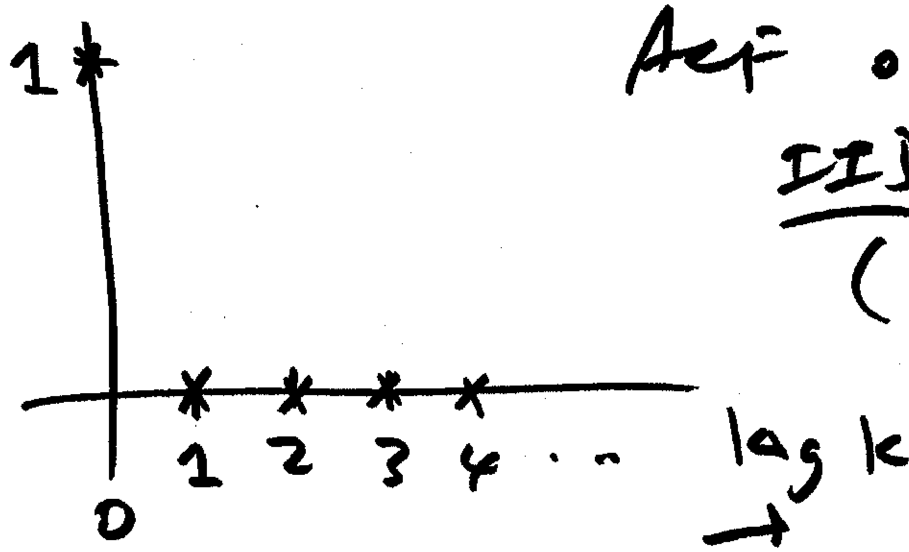
1st order autocorrelation of the time series $\{\theta^*_t\}$

collect all the $\hat{\rho}_k$ together for $k = 0, 1, 2, \ldots$ & plot: <u>autocorrelation function</u> (ACF)

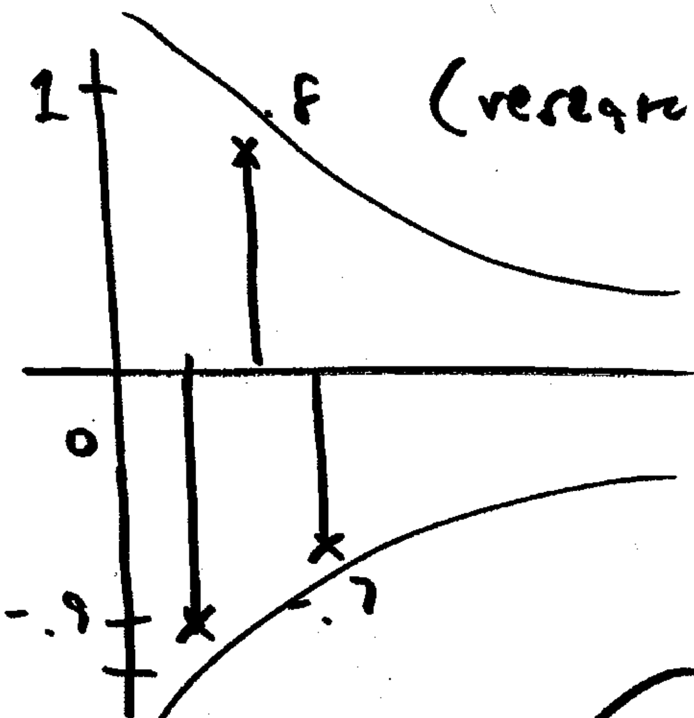


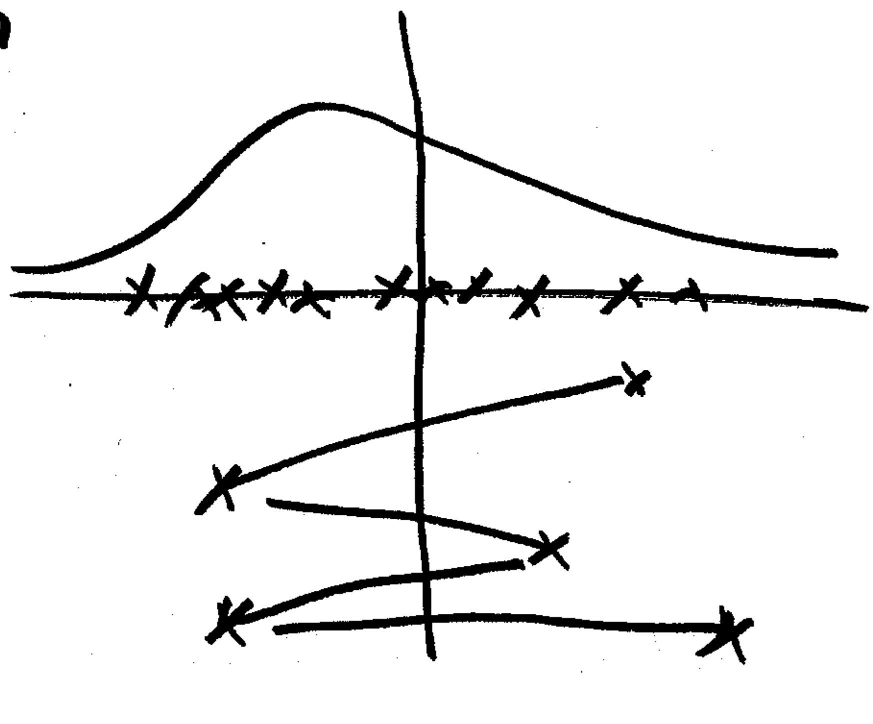
this chain is not <u>mixing well</u>

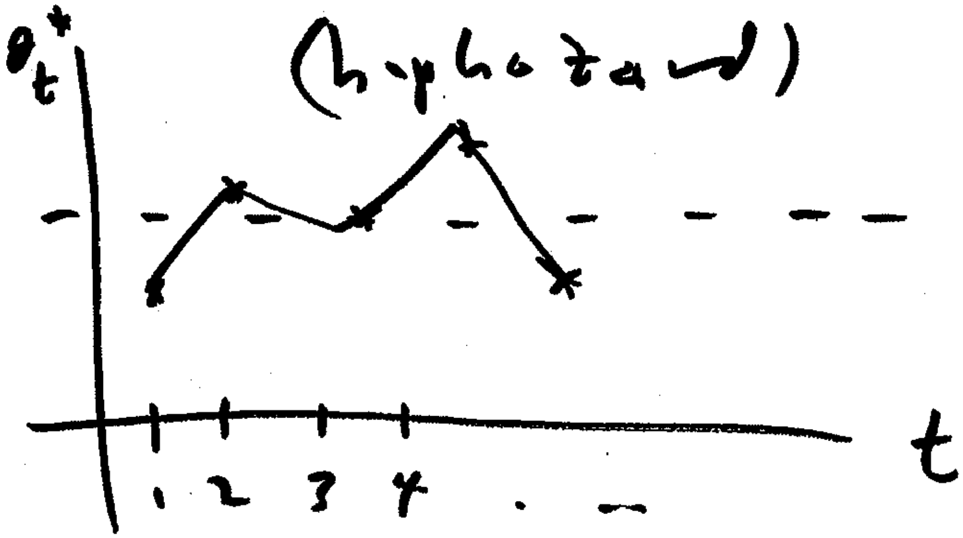
Acf of
IID samples
(white noise)



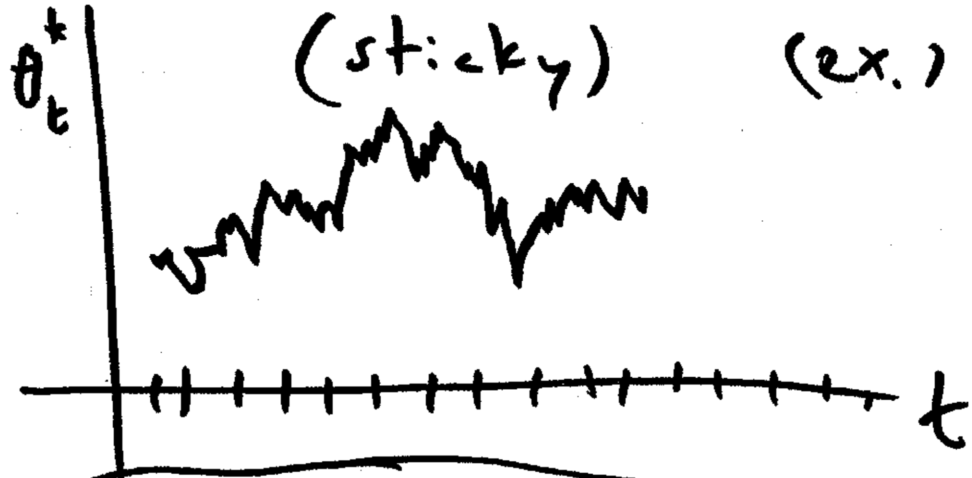
(research) Acf of
highly
efficient
MCMC
sampler





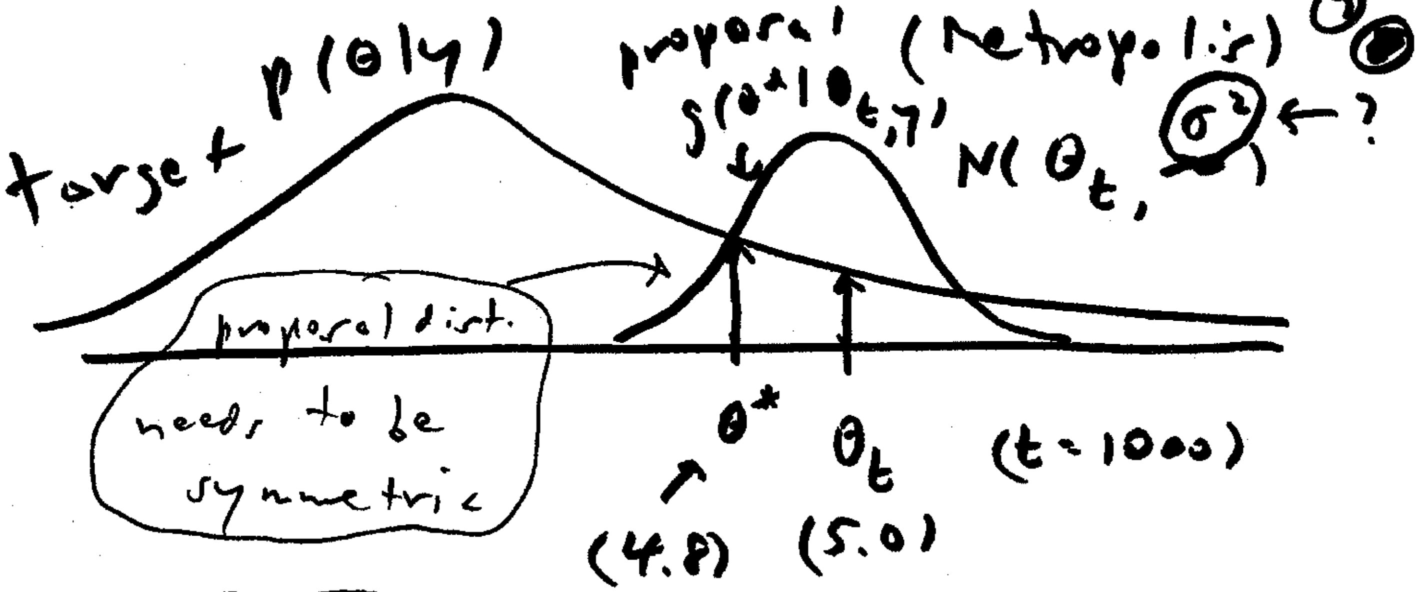


time series plot for white noise



time series plot with high positive ACF

acceptance rate of Metropolis algorithm is high but moves are small



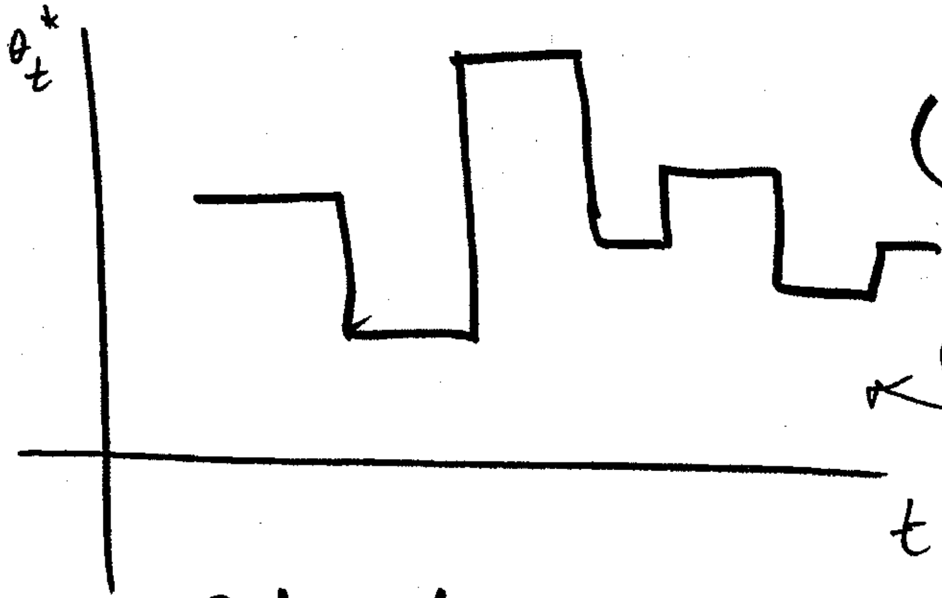
if $\boxed{\dfrac{p(\theta^*|y)}{p(\theta_t|y)}} > 1$ accept with prob. 1

(accept all uphill moves)

if $\boxed{\dfrac{p(\theta^*|y)}{p(\theta_t|y)}} \le 1$ accept with prob. $\boxed{\dfrac{p(\theta^*|y)}{p(\theta_t|y)}}$

if accept $\theta_{t+1} = \theta^*$

if reject $\theta_{t+1} = \theta_t$ (!)



(blocky)

another time series plot with high + autocorrelation (for a different reason)

(try to make really big moves & fail most of the time)

$$\sigma^2 \sim \chi^{-2}(\nu, s^2)$$

$$(y_i \mid \sigma^2) \overset{iid}{\sim} N(\mu, \sigma^2) \quad i = 1, \dots, n$$

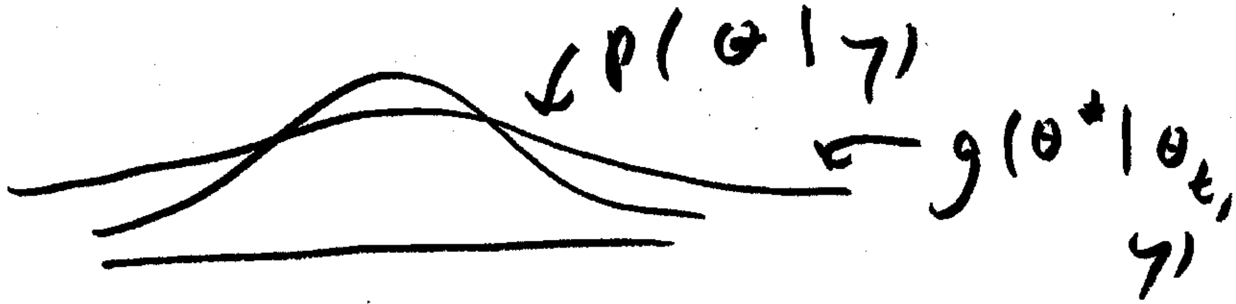
μ known

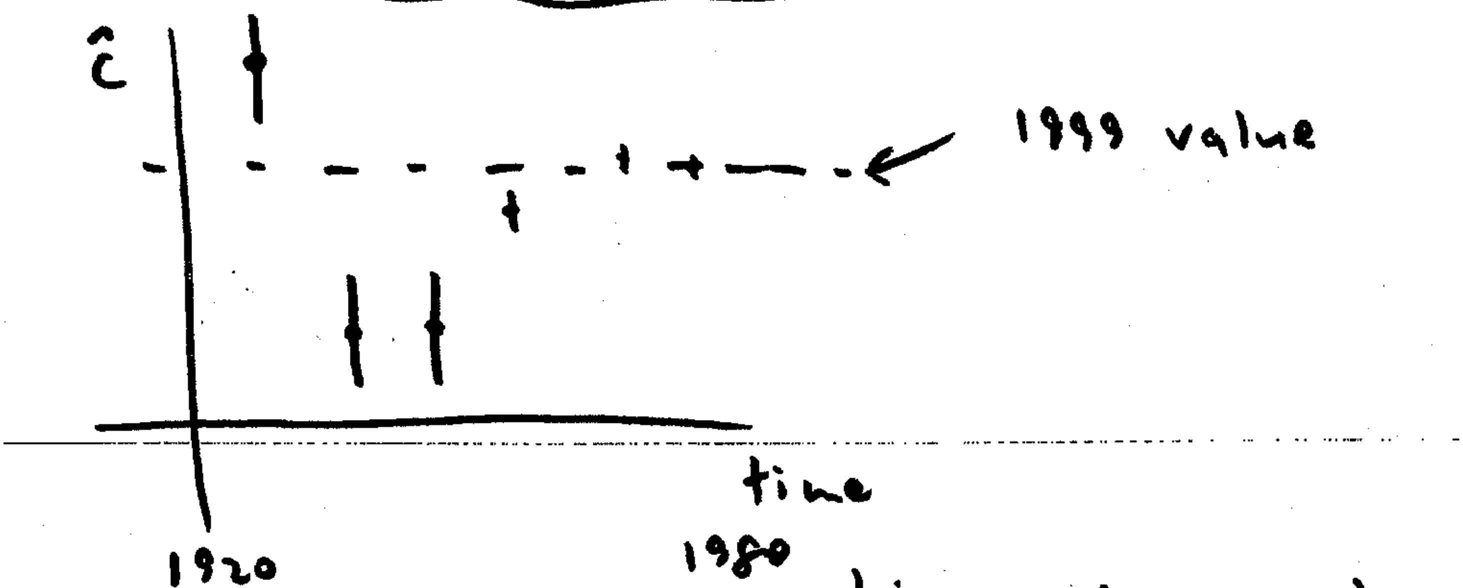
$$(\sigma^2 \mid y) \sim \chi^{-2}\left(\nu + n, \frac{\nu s^2 + n\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - \mu)^2\right)}{\nu + n}\right)$$

$$\ell(\sigma^2 \mid y) = c \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(y_i - \mu)^2}$$

$$\left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}} = (\sigma^2)^{-\frac{1}{2}}$$

$$= c\,(\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{\sum (y_i - \mu)^2}{2\sigma^2}\right)$$





$$Y_{ij} = c + b_i + e_{ij}$$

bias

b_i iid $N(0, \sigma_b^2)$

e_{ij} iid $N(0, \sigma_e^2)$

i: time point

j: replication

random-effects

<u>meta</u>-<u>analysis</u> model

$$\theta = (c, \sigma_b^2, \sigma_e^2)$$

$$i = 1, \ldots, I$$

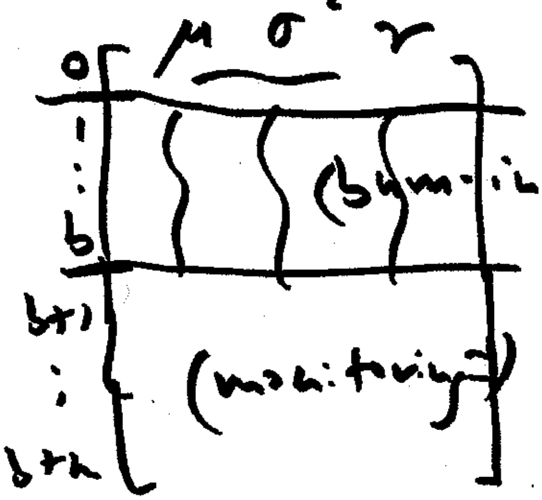
$$j = 1, \ldots, n_i$$

$$\left[\; c \quad \sigma_b^2 \quad \sigma_e^2 \;\Big|\; \underbrace{b_1 \quad b_2 \quad \cdots \quad b_I}_{A \,=\, \text{auxiliary variables}} \;\right]$$

$(\mu, \sigma^2, \nu) \sim p(\cdot)$

$(Y_i \mid \mu, \sigma^2, \nu) \overset{iid}{\sim} t_\nu(\mu, \sigma^2)$

$i = 1, \dots, n$



MCMC data set

(μ, σ^2, ν) update all at once

$$\left\{ \begin{array}{ll} \mu & \text{update by itself} \\ \sigma^2 & \text{''} \\ \nu & \text{''} \end{array} \right\}$$

(μ, σ^2), then ν

$\theta = (\mu, \sigma^2, \nu) \rightarrow$ 3 full conditionals

$$\left.\begin{array}{l} \boxed{p(\mu \mid y, \sigma^2, \nu)} \\ p(\sigma^2 \mid y, \mu, \nu) \\ p(\nu \mid y, \mu, \sigma^2) \end{array}\right\}$$

easier to work with than $p(\mu, \sigma^2, \nu \mid y)$ & also than

$\boxed{p(\mu \mid y)}$, $p(\sigma^2 \mid y)$, $p(\nu \mid y)$

$$\begin{array}{ccc} \mu & \sigma^2 & \nu \\ \mu_0 & \sigma_0^2 & \nu_0 \\ \mu_1^* & (\sigma_1^2)^* & \nu_1^* \\ \vdots & \vdots & \vdots \end{array}$$

$\mu_1^* \sim p(\mu \mid y, \sigma_0^2, \nu_0)$

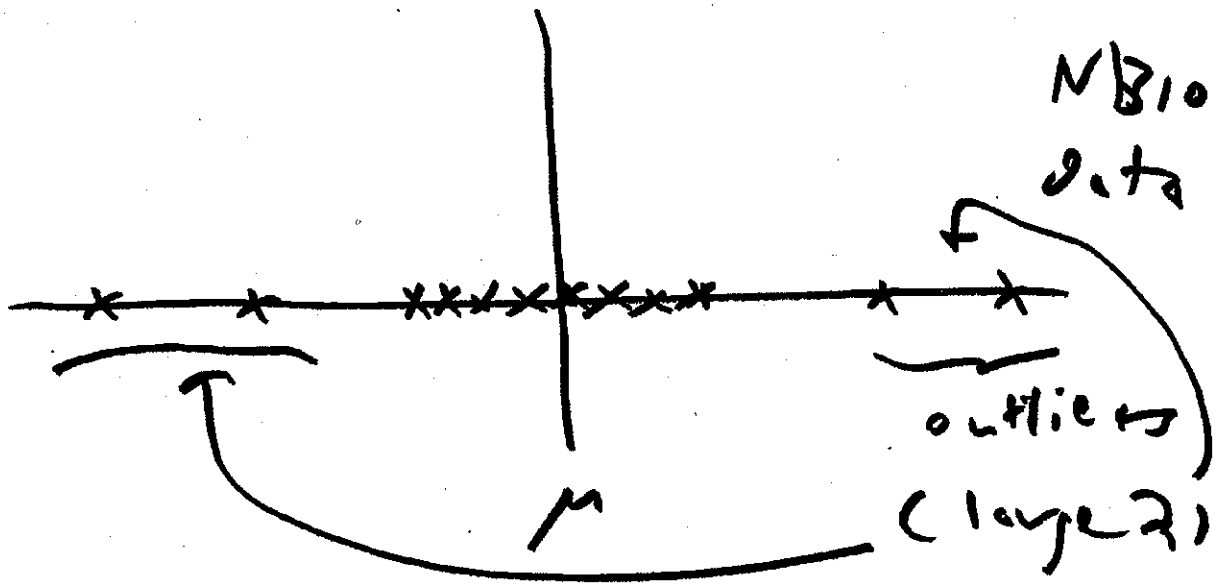
$(\sigma_1^2)^* \sim p(\sigma^2 \mid y, \mu_1^*, \nu_0)$

$\nu_1^* \sim p(\nu \mid y, \mu_1^*, (\sigma_1^2)^*)$

$\vdots$

stochastic relaxation

$$\{Y\} \longleftrightarrow \begin{Bmatrix} X \\ Y \mid X \end{Bmatrix} \quad \begin{Bmatrix} \lambda \\ Y \mid \lambda \end{Bmatrix}$$



$$\overleftarrow{\quad \theta \quad}\!\!\rightarrow$$

$$\left[\begin{array}{ccc|ccc} \mu & \sigma^2 & \nu & \lambda_1 & \cdots & \lambda_n \\ & & & & \text{(latent)} & \end{array}\right]$$

$\lambda_n \leftarrow 100$

$$p(\mu,\sigma^2,r \mid y) = c\, p(\mu,\sigma^2,r)\,\ell(\mu,\sigma^2,r \mid y)$$

?

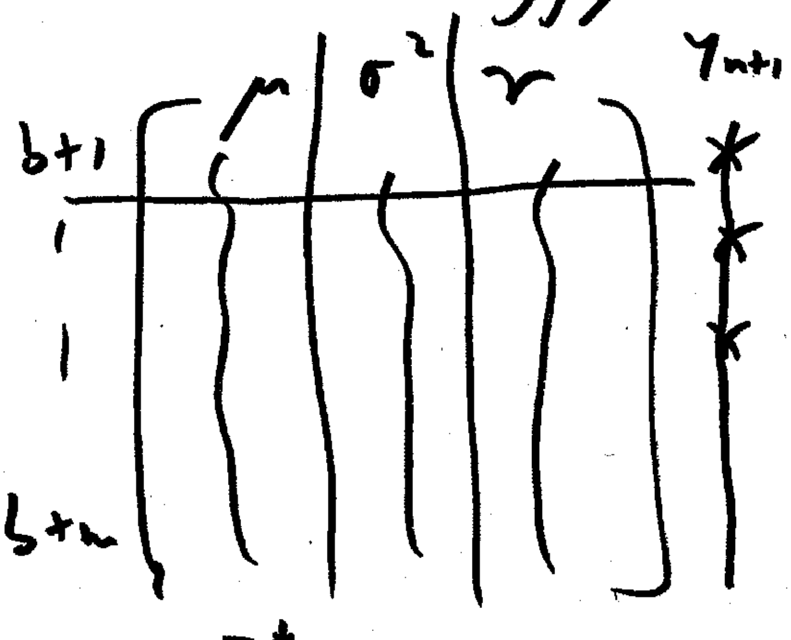
$$\boxed{\iiint p(\mu,\sigma^2,r)\,\ell(\mu,\sigma^2,r \mid y)\,d\mu\,d\sigma^2\,dr}$$

$$p(\mu \mid y) = \iint p(\mu,\sigma^2,r \mid y)\,d\sigma^2\,dr$$

$\vdots$

$$p(y_{n+1} \mid y) = \int p(y_{n+1} \mid \theta)\,p(\theta \mid y)\,d\theta$$

$$\iiint$$



mean $\bar{\mu}^*$

$$\hat{E}(\mu \mid y) = \bar{\mu}^*$$

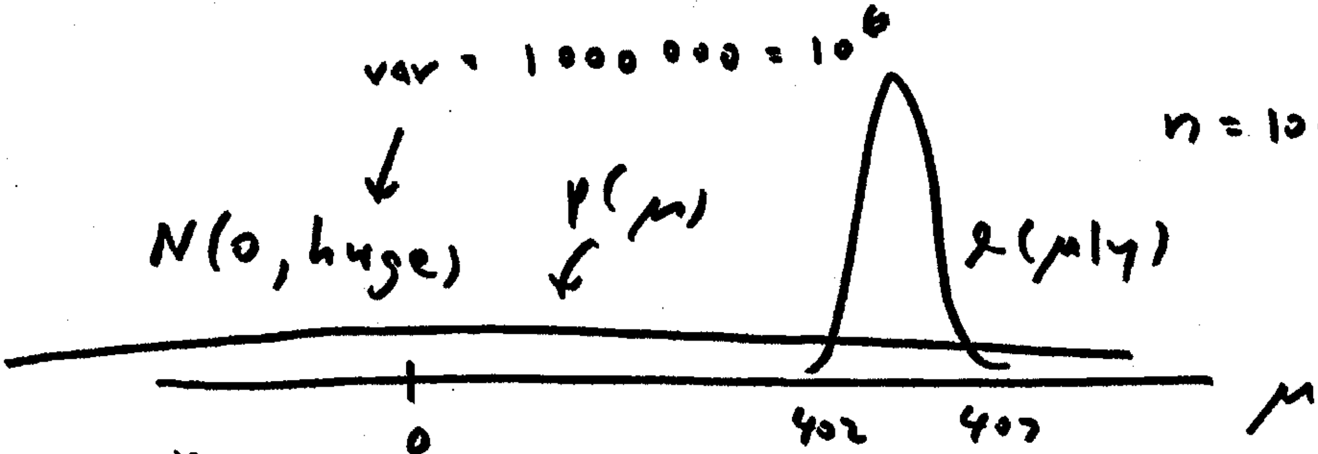
etc

$\cancel{p(y_{n+1} \mid y) =}$

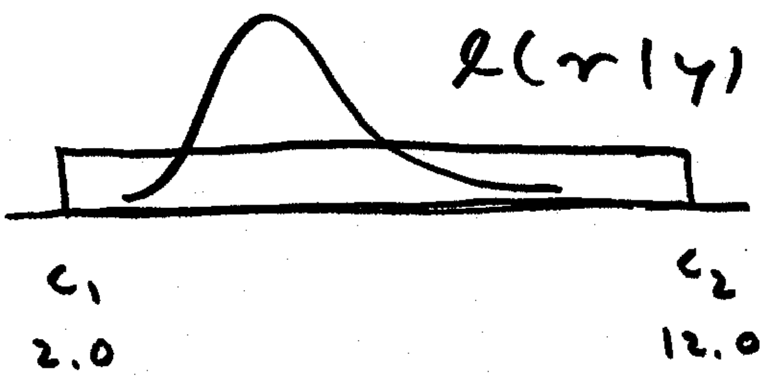
$$y_{n+1} \leftrightarrow \left\{ \begin{matrix} \theta \\ y_{n+1} \mid \theta \end{matrix} \right\}$$

16

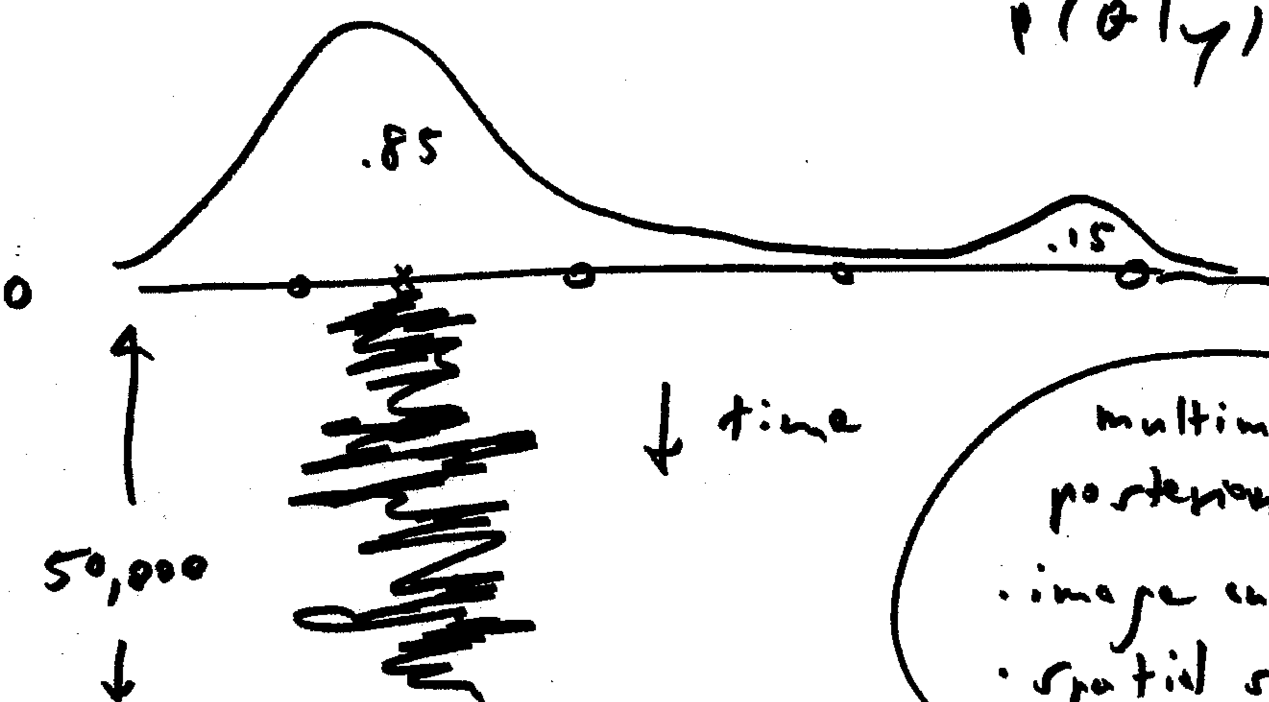
$$\left[\begin{array}{ccc|c|c} \mu & \sigma^2 & r & Y_{n+1} & \eta = \frac{\sqrt{\sigma^2}}{\mu} \\ 1 & 2 & & & \sqrt{2}/1 \end{array}\right]$$



prior SD = 1000   prior precision = 10^{-6}

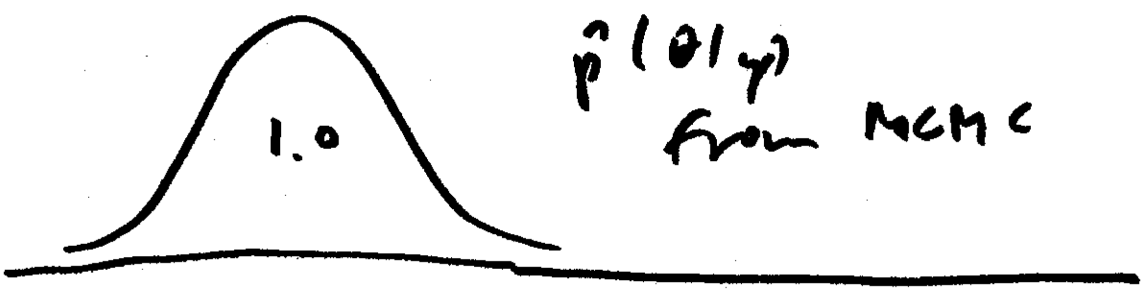


MCMC's bête noire:

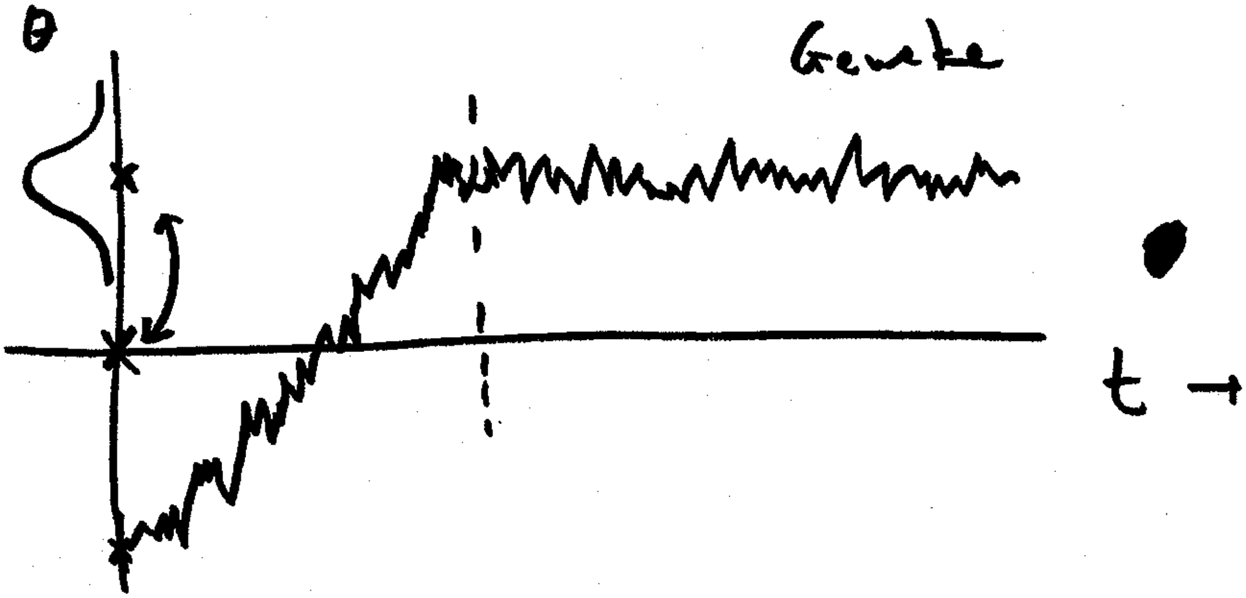


multimodal posteriors:

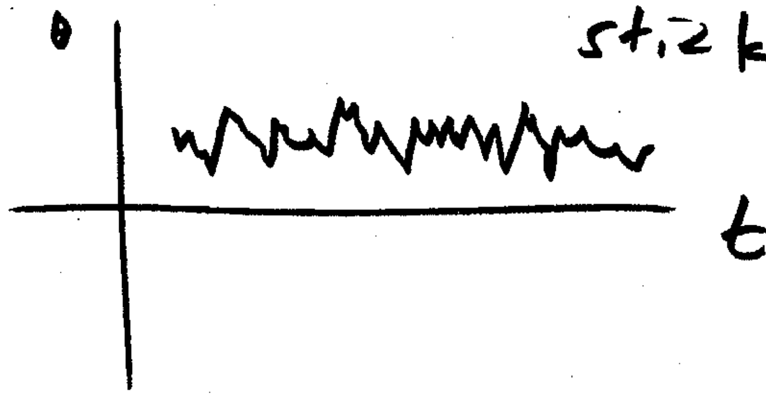
- image analysis
- spatial statistics



Geweke



$$\begin{array}{c|c|}
 & \theta \\
1001 & \\
1002 & \\
1003 & \\
\vdots & \\
6000 &
\end{array}$$



sticky

subsample every 10^{th} obs. (say): thinning chain by a factor of 10

thinned chain will have less AC

only reason to do this: save disk space