NB10 case study

$(\mu\, \sigma\, \nu \mid [pM]\, \mathcal{B}) \sim p(\mu\, \sigma\, \nu \mid \mathcal{B})$

$(Y_i \mid \mu\, \sigma\, \nu\, [SM\, \mathbb{T}]\, \mathcal{B}) \overset{IID}{\sim} t_\nu(\mu, \sigma^2)$

$(i = 1, \dots, n)$

STAT 206
2 Mar 21

Lecture ①

BIC

DIC

Gaussian model doesn't fit

$\nexists$ a conj. prior for this SM

we want a LI prior

hard to implement in this SM

rejection sampling (von Neumann)

there does not exist

(1953)
(1942) A. Rosenbluth
M. Rosenbluth
(Metropolis et. al.)
(not Ulam)
(MC)
(1949)

Metropolis Hastings (MH) algorithm

Markov Chain Monte Carlo (MCMC)

Chemical Physics

analogy engine "similar to, in all relevant ways"

(1970) Hastings Biometrika (general AI)

late 1980s

D. Geman & Geman (1984) → Gibbs Sampling (J. W. Gibbs)

special case of MH alg.

IEEE Transactions

late 1980s Geman & A. Gelfand

Gelfand & A.F.M. Smith (1990) Gibbs sampling (JASA)

(1) tower of Babel

(2) personal computers became fast & inexpensive enough to make MCMC practical

~~(then)~~

5,000 MCMC iterations took all night on your PC (12 hours) 8P → 8A von Neumann's rejection sampling

IID draws from

$p(\underset{\sim}{\theta} \mid \text{——})$ $\leftarrow$ $\underset{\sim}{\theta} = (\theta_1, \dots, \theta_k)$

works well when k is small (e.g., in special cases, $k \le 10$)

takes too long in clock time for large k

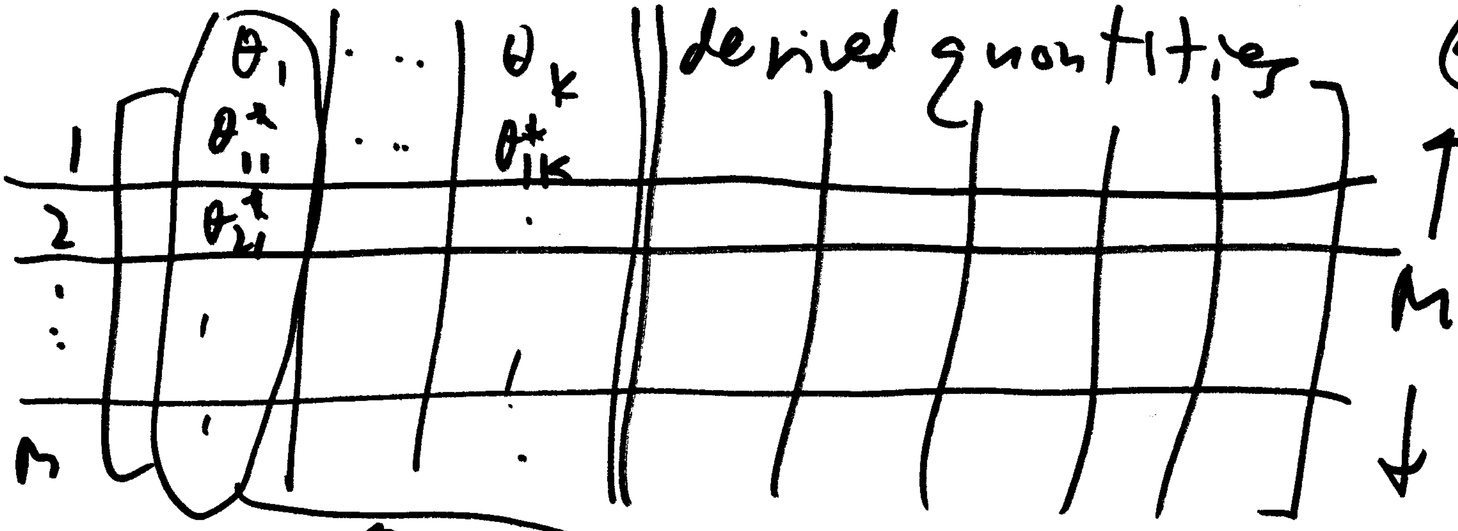
doesn't scale as $k \uparrow$

MC method

M-H insight

anything you want to know about $p(\underset{\sim}{\theta}_k \mid \text{——})$ (*) can be estimated to arbitrary accuracy by (a) finding a way to make lots of random ~~IID~~ draws from (*) and (b) making graphical and numerical summaries of the draws

metropolis et al. insight (independence not needed)



MC data set

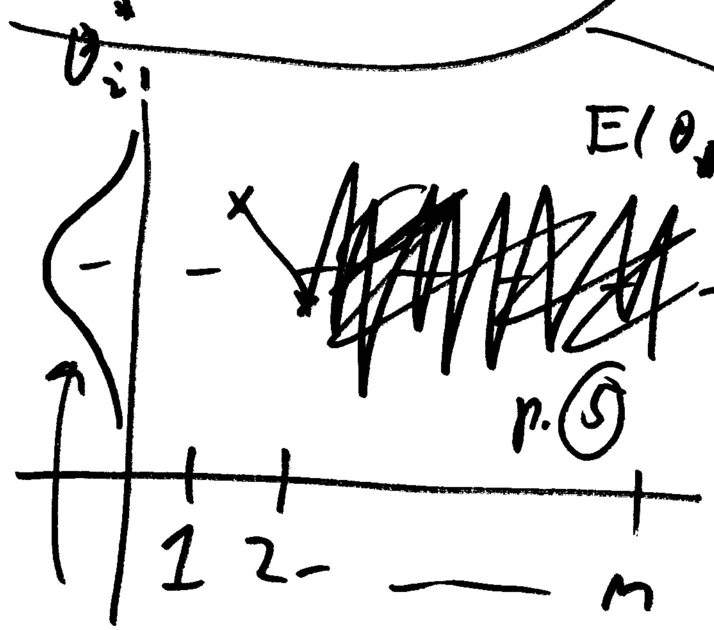
$\theta^*_1 = (\theta^*_{11}, \ldots, \theta^*_{m1})$

iteration # (like time index)

$E(\theta_1 | \text{—})$

IID

MC



p. 5

$\theta^*_{i+1,1}$ IID

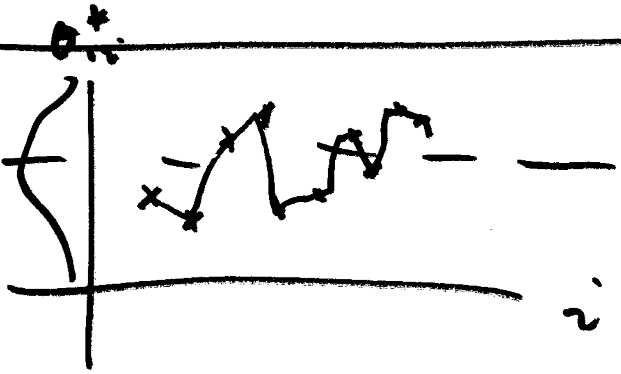
$p(\theta_1 | \text{—})$ does not depend on

$$(\theta^*_{i,1}, \theta^*_{i-1,1}, \ldots, \theta^*_{1,1})$$

1st order Markov chain: $\theta^*_{i+1,1}$ is allowed to depend on $\theta^*_{i,1}$ but not on any previous $\theta^*_{\cdot,1}$

nice Markov chains: settle down, after (a few) initial iterations (starting with an initial value), into a steady-state behavior called the <u>equilibrium</u> or <u>stationary</u> distribution of the Markov chain

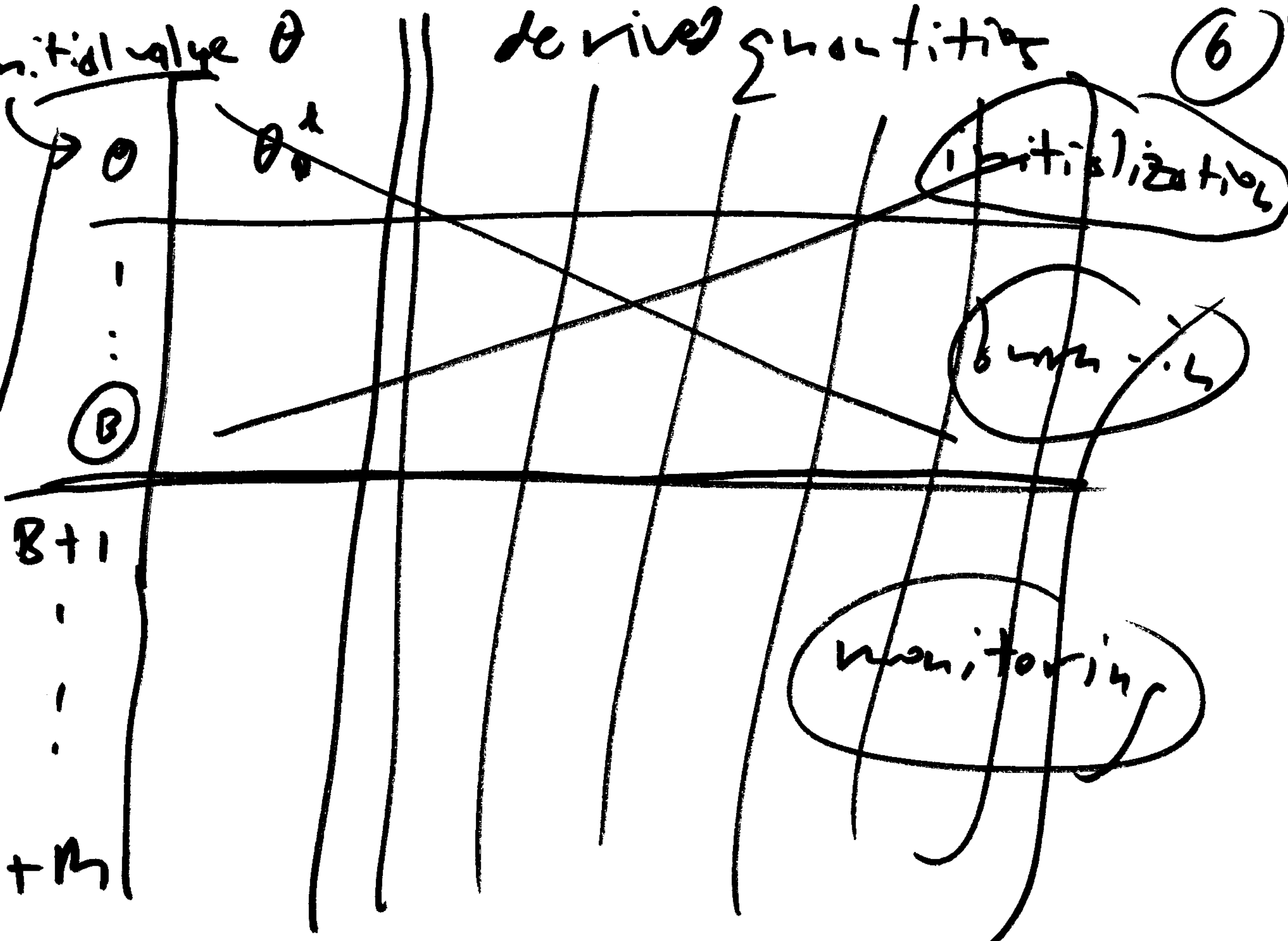
take $t = 1$, $\theta_{t+1} = \theta_{1,1}$



haphazard, no pattern

(k = 1) MCMC data set

initial value θ | derived quantities

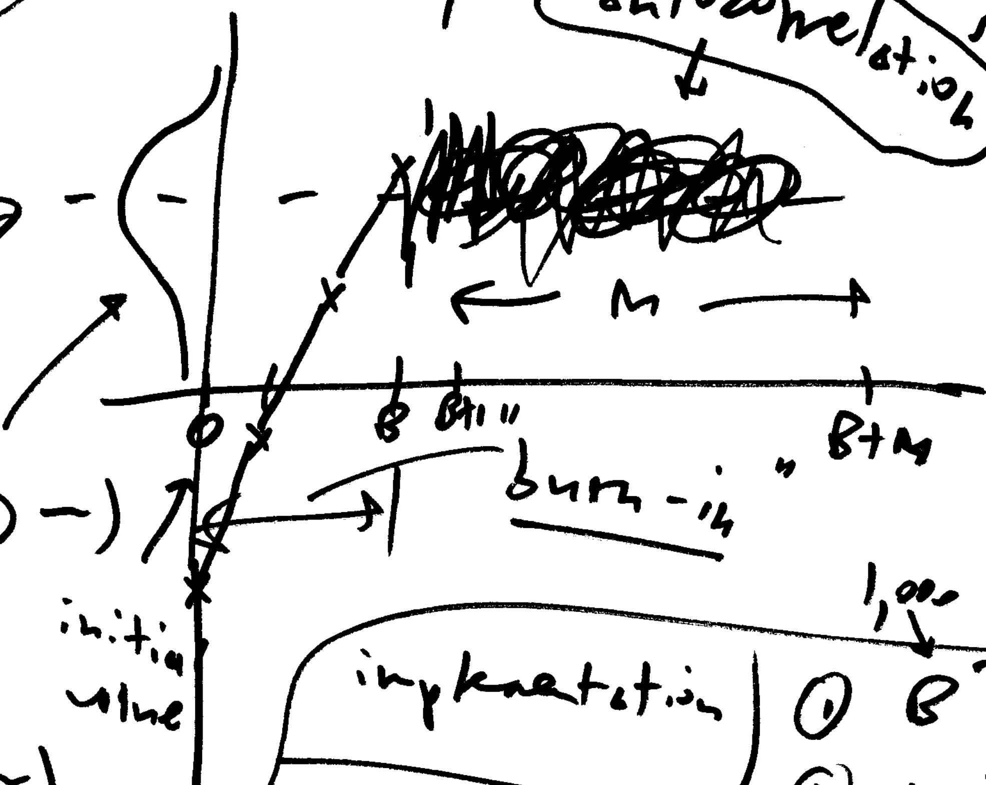


each row depends on previous row

θ^*

MCMC sample of θ values

autocorrelation



$p(\theta \mid D\sim)$

$E(\theta \mid D\sim)$

implementation

1,000

① B? B unnecessary okay

② M?

100,000

③ initial value?