STAT 206
2 May 21
DD extra OA

①

Use MC (IID) if you can; it will usually be more Monte Carlo efficient (especially with small k) than MCMC

(more info about post. dist with MC per clock-second than with MCMC)

---

but often MC is not available; then use MCMC

MC algorithms with large k either don't exist or are too MC-inefficient (in that case MCMC also wins)

Q: how large is "large k"?

A: with whatever k value you have in your problem, try using both MC & MCMC

& compare their Monte-Carlo efficiency

if MC not available you must use MCMC

method	
MC	
MCMC	

m = # monitoring draws

$$\widehat{MCSE}(\bar{\theta}^*) = \frac{\hat{\sigma}_\theta}{\sqrt{m}} \text{ for IID}$$

$\hat{\sigma}_\theta$ ← post. SD for θ

MC mean estimate of posterior mean

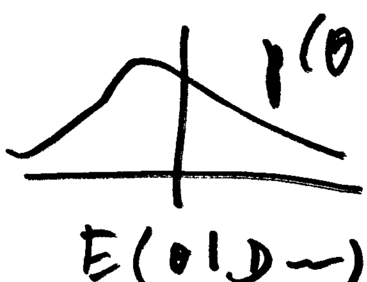
iteration #	θ
1	
2	
⋮	
m	

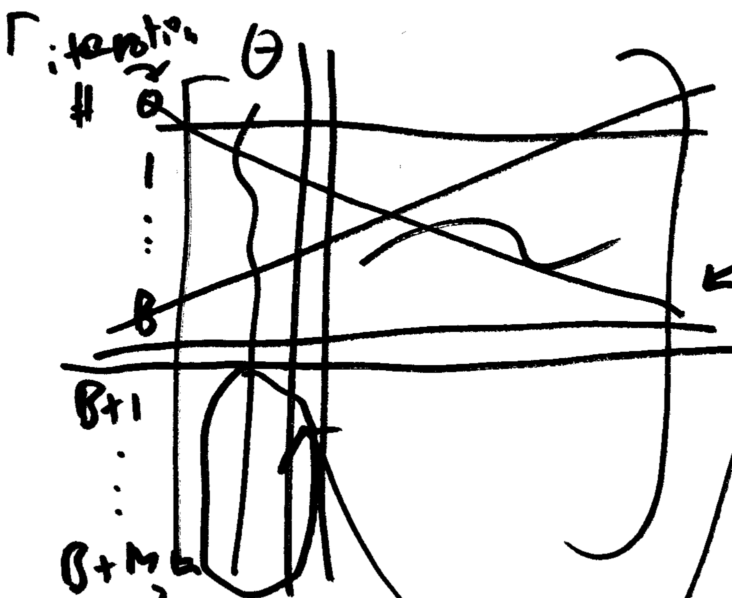
MC data set

mean $\bar{\theta}^* = \hat{E}(\theta \mid D \sim)$

SD $\hat{\sigma}_\theta$

posterior mean





MCMC data set

mean $\bar{\theta}^*$

SD $\hat{\sigma}_\theta$

last Fri: time series interlude

$AR_\rho(1)$ model

auto-regressive

order 1

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

$$\theta^*_{i+1} = \mu + \rho\, \theta^*_i + e_i$$

$\rho(\theta^*_{i+1}, \theta^*_i) \doteq \rho$

IID $N(0, \sigma^2)$ independent of $\{\theta^*\}$ process

but this is just a special case of the Markov chain idea

$\boxed{\text{fact:}}$ for many MCMC algorithms, the θ^*_i outcomes behave a lot like an $AR_\rho(1)$ (IID) ($i = B+1, \ldots, B+M_2$) if so,

$$\widehat{MCSE}(\bar{\theta}^*) = \frac{\hat{\sigma}_\theta}{\sqrt{M_2}} \cdot \sqrt{\frac{1+\hat{\rho}}{1-\hat{\rho}}}$$

if $\hat{\rho} = 0$: $\sqrt{} \doteq 1$

$AR_{\rho=0}(1)$ = MC IID sampling

$$\left(\widehat{MCSE}(\bar{\theta}^*)\right)^2 = \frac{\hat{\sigma}_\theta^2}{m_2}\boxed{\left(\frac{1+\hat{\rho}}{1-\hat{\rho}}\right)}$$

VIF = variance inflation factor

$=$ MC variance of $\bar{\theta}^*$

almost always with MCMC, $\rho > 0$ ($\hat{\rho} > 0$)

as $\rho \to +1$, the VIF $\uparrow \to \infty$ (!)

ρ	VIF
0	1 (same as IID)
0.5	3
0.99	199 (!)

① make m_1 IID MC draws:

$$\widehat{MCSE}(\bar{\theta}^*) = \frac{\hat{\sigma}_\theta}{\sqrt{m_1}}$$

② make m_2 MCMC draws: if $AR_\rho(1)$ works,

$$\widehat{MCSE}(\bar{\theta}^*_*) = \frac{\hat{\sigma}_\theta}{\sqrt{m_2}}\sqrt{\frac{1+\hat{\rho}}{1-\hat{\rho}}}$$

(5)

$$\frac{\cancel{\hat{\sigma}}}{\sqrt{m_1}} = \frac{\cancel{\hat{\sigma}}}{\sqrt{m_2}} \sqrt{\frac{1+\hat{\rho}}{1-\hat{\rho}}}$$ solve for m_2 in terms of m_1

$$\frac{m_2}{m_1} = \frac{1+\hat{\rho}}{1-\hat{\rho}}, \text{ i.e. } m_2 = m_1 \underbrace{\left(\frac{1+\hat{\rho}}{1-\hat{\rho}} \right)}_{VIF}$$

~~to get a completely relevant~~

MC efficiency comparison, suppose that the MC algorithm takes $s_{MC}(\underline{m_1})$ seconds of clock time to deliver m_1 draws, & the MCMC algorithm takes $s_{MCMC}(m_2)$ seconds to deliver m_2 draws

Bernstein–von Mises / Limit theorem

as $n \to \infty$ — an example of an asymptotic result

Q:) How can asymptotics help us when our n is finite?

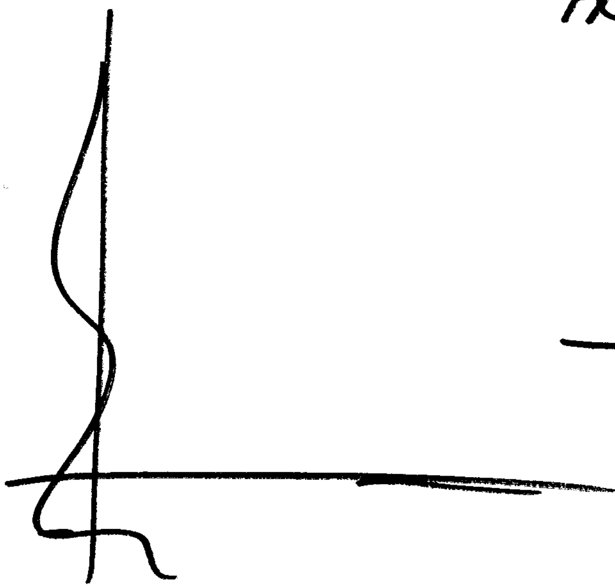
Ex. THT2 2(A) $n = 1{,}265$

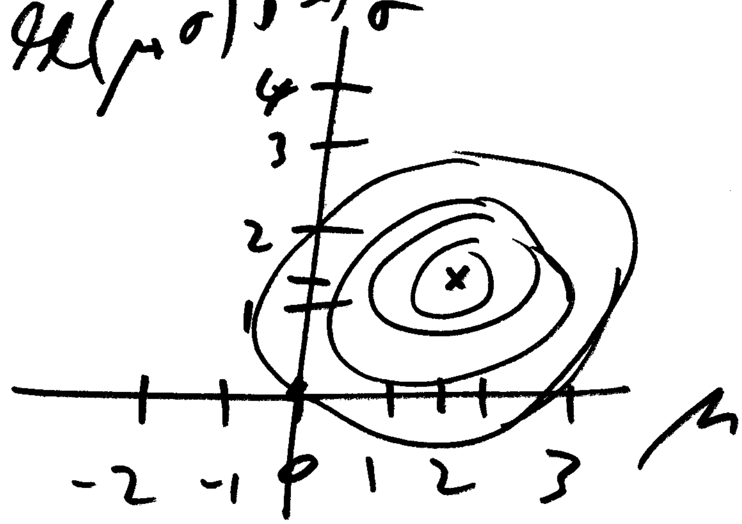
A:) Theorem also says (IF ...) when n is large, Bayes $\doteq$ Frequency

$\left| \text{Bayes answer} - \text{Frequency answer} \right| \leq \epsilon$

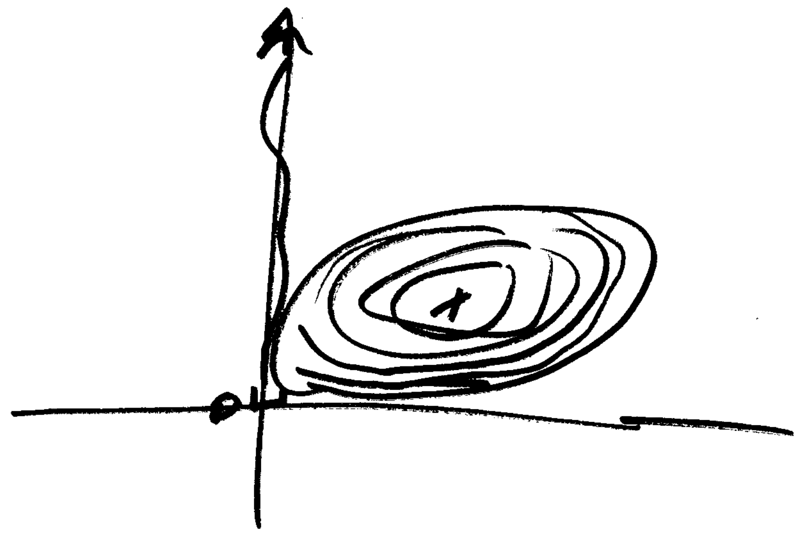
small "tolerance"

to find out how n & ϵ are related, you can either (a) do math or (b) simulate (better)





$$\rho(X, Y) = \frac{C(\overset{\$}{X}, \overset{\text{yrs.}}{Y})}{SD(X) \cdot SD(Y)}$$



$$= \frac{C(X, Y)}{\sqrt{V(X) \cdot V(Y)}}$$