Lec 21 Math 341 5/1/19 1

Disadvantages of Grid Sampling

(1) Numerically unstable... computers hate min. values and max. values

(2) How to pick $\theta_{min}, \theta_{max}, \Delta$? Easy in one dimension with known supports... Bad decision for $\theta_{min}, \theta_{max}, \Delta$? Miss points!!

(3) "Curse of dimensionality"
Let's say you want 10^5 points in the grid per dimension
if $\vec{\theta}$ has 10 dimensions ... $\Rightarrow 10^{50}$ points total. IMPOSSIBLE!

$$\underbrace{100,000,\ldots,000}_{50 \text{ zeroes}}$$

In 10 dimen if 10^5 points $\Rightarrow \sqrt[10]{10^5} \approx 2$ points in each dimension

If 10^9 points $\Rightarrow \sqrt[10]{10^9} \approx 8$ points in each dimension.

Need another solution!

Want $P(\theta, \sigma^2 | X)$ but have $P(\theta | X, \sigma^2) = N\left(\dfrac{\frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}, \dfrac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\right)$

$P(\sigma^2 | X, \theta) = \text{InvGamma}\left(\dfrac{n_0 + n}{2}, \dfrac{n_0 \sigma_0^2 + n \hat{\sigma}^2_{MLE}}{2}\right)$

Can we use these two to get posterior?

Imagine the following algorithm:

(I) Begin at θ_0 = some reasonable value like 0

(II) Draw σ_1^2 from $P(\sigma^2 | X, \theta = \theta_0) = \text{InvGamma}\left(\frac{n+\nu_0}{2}, \frac{n\hat{\sigma}^2_{MLE}+\nu_0\sigma_0^2}{2}\right)$

(III) Draw θ_1 from $P(\theta | X, \sigma^2 = \sigma_1^2) = N\left(\frac{\frac{n\bar{X}}{\sigma^2}+\frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}, \frac{1}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}\right)$

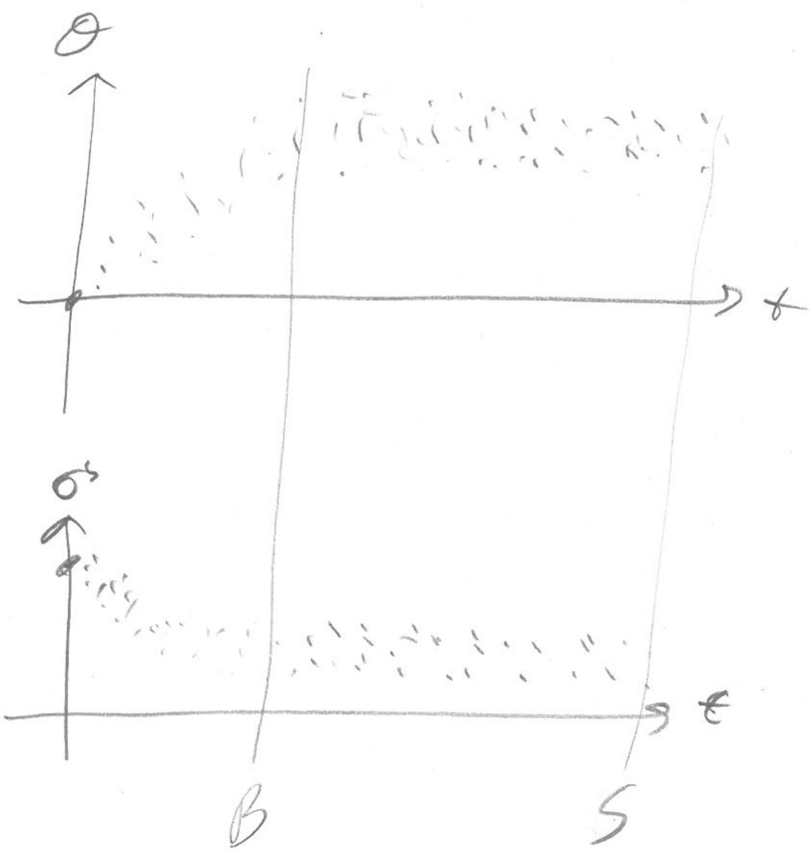
(IV) Draw σ_2^2 from $P(\sigma^2 | X, \theta = \theta_1)$

(V) Draw θ_2 from $P(\theta | X, \sigma^2 = \sigma_2^2)$

Keep repeating until "convergence." Then

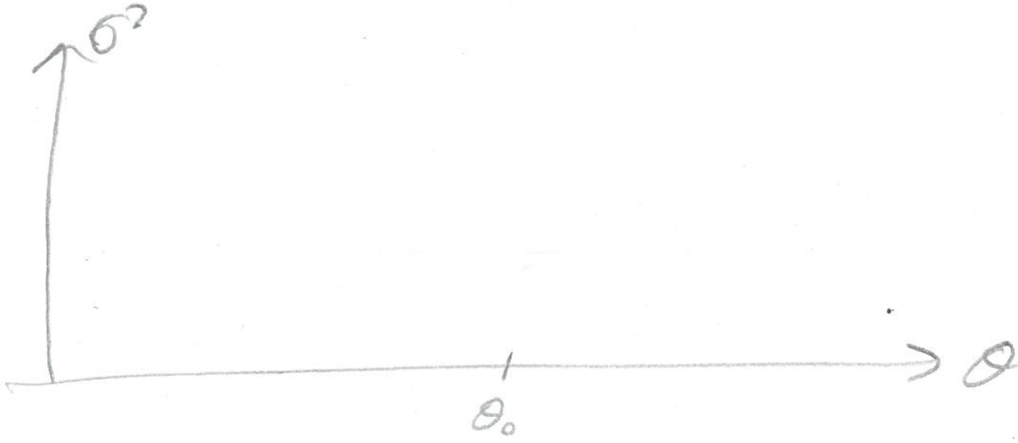
Results...

$$\begin{pmatrix}\theta_1 \\ \sigma_1^2\end{pmatrix}, \begin{pmatrix}\theta_2 \\ \sigma_2^2\end{pmatrix}, \ldots, \begin{pmatrix}\theta_t \\ \sigma_t^2\end{pmatrix}$$ where t is the current ~~iteration~~ #.



let B be the point at which both θ, σ^2 converged. This is called the "Burn-in" point.

let final iteration be S after you harvest "enough" samples.



$P(\sigma^2 | \theta = \theta_0)$

$P(\theta | \sigma^2 = \sigma_1^2)$

Systematic Sweep Gibbs Sampler for $P(\theta_1, \dots, \theta_p)$

(I) Introduce $\vec{\theta}_0 := \langle \theta_{0,1}, \theta_{0,2}, \dots, \theta_{0,p} \rangle$, a starting point

(II) Sample $\theta_{1,1}$ from $P(\theta_1 \mid \theta_2 = \theta_{0,2}, \dots, \theta_p = \theta_{0,p})$

Sample $\theta_{1,2}$ from $P(\theta_2 \mid \theta_1 = \theta_{1,1}, \theta_3 = \theta_{0,3}, \dots, \theta_p = \theta_{0,p})$

Sample $\theta_{1,3}$ from $P(\theta_3 \mid \theta_1 = \theta_{1,1}, \theta_2 = \theta_{1,2}, \theta_4 = \theta_{0,4}, \dots, \theta_p = \theta_{0,p})$

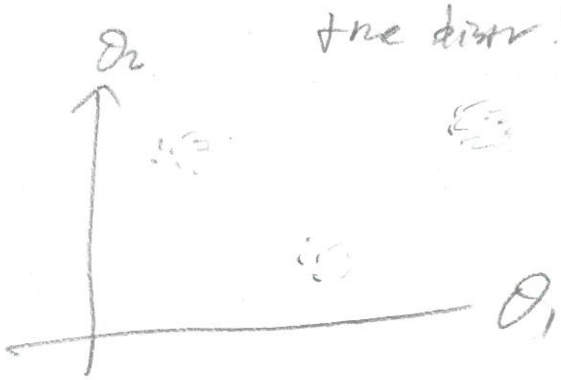
$\vdots$

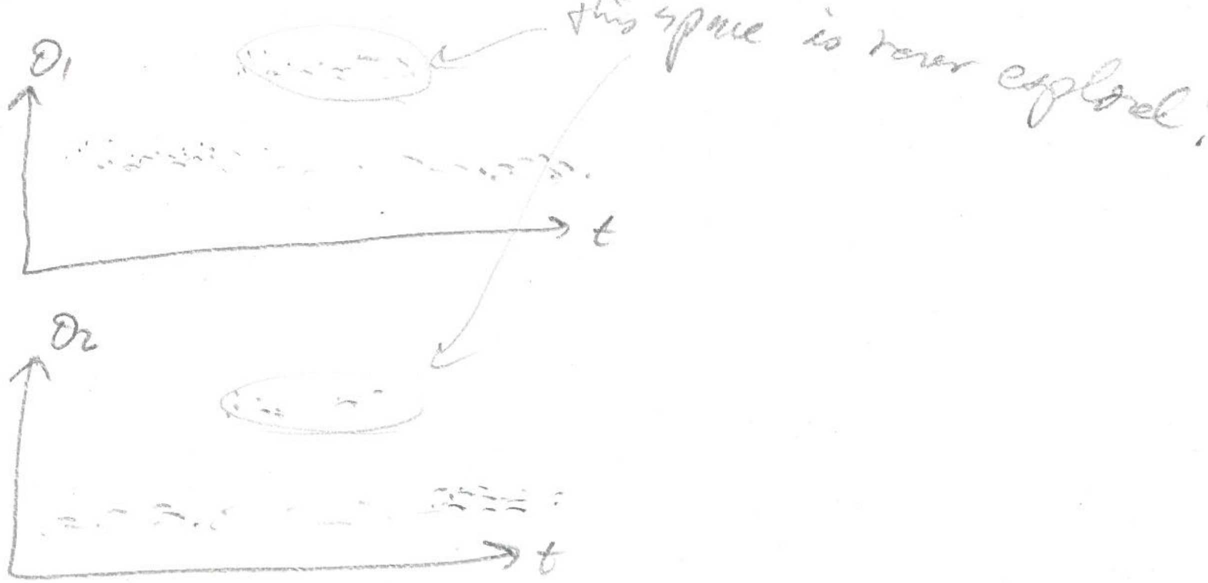
Sample $\theta_{1,p}$ from $P(\theta_p \mid \theta_1 = \theta_{1,1}, \dots, \theta_{p-1} = \theta_{1,p-1})$

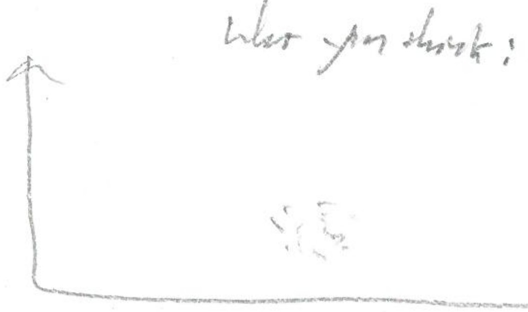
(III) Record $\vec{\theta}_1 := \langle \theta_{1,1}, \dots, \theta_{1,p} \rangle$ as a sample

(IV) Repeat II, III S times

Any problems with Gibbs Samplers? Yes...







this gets worse if $\dim[\vec{\theta}]$ is large....

Solution... start gibbs sampler in many different starting locations.

Another limitation: All cond. distr's

$$\forall_j \quad P(\theta_j \mid \theta_{-j}) \text{ are assumed known.}$$

If some are kernels... we can grid sample... but slow... or other methods