lecture 20 NO. 05/05/2020.

$P(\theta, \sigma^2|X) \propto P(\theta|\sigma^2, X)\, P(\sigma^2|X)$ 구하기 힘들다, 대신 $k(\sigma^2|X)$로

$\propto P(\theta|\sigma^2, X)\, k(\sigma^2|X)$ ← last class

$P(\sigma^2|X) = c \cdot k(\sigma^2|X)$ Drawing from $k(\sigma^2|X)$ approximately in the goal.

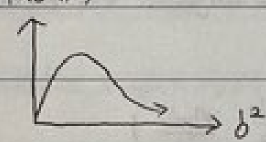
produce called "grid sampling."

Ⓘ Create a "grid" by picking σ^2_{min}, σ^2_{max}, Δ → grid resolution.

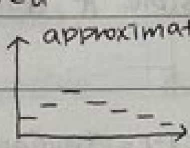
$G = \{\sigma^2_{min}, \sigma^2_{min}+\Delta, \sigma^2_{min}+2\Delta, \dots, \sigma^2_{max}\}$

Ⓘⓘ Compute $C_{approx} = \dfrac{1}{\sum_{\sigma^2 \in G} k(\sigma^2|X)}$ "$P(\sigma^2|X) \propto C_{approx}\, k(\sigma^2|X)$"

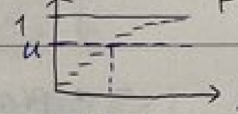
of $P(\sigma^2|X)$



[실제]



[grid sampling]



σ^2_{samp}.

Ⓘⓘⓘ Compute the approximate CDF $F(\sigma_0^2|X) := P(\sigma^2 \le \sigma_0^2|X)$

$= \sum_{\{\sigma^2 \le \sigma_0^2 : \sigma^2 \in G\}} C_{approx}\, k(\sigma^2|X)$

Ⓘⓥ Draw u from $U(0,1)$ via runif(0,1) then locate $\sigma^2_{samp} := \min_{\sigma^2 \in G} \{F(\sigma^2|X) \ge u\}$

Sample from the prior

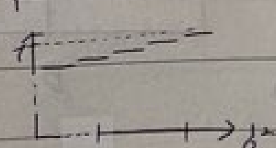
① Draw σ^2_{samp} from $k(\sigma^2|X)$ via grid sampling.

② Draw θ_{samp} from $P(\theta|X, \sigma^2 = \sigma^2_{samp})$

③ Return $\langle \theta_{samp}, \sigma^2_{samp} \rangle$

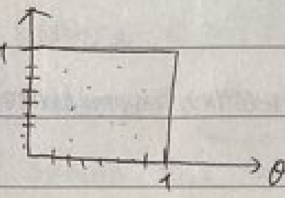
Disadvantages of Grid sampling.

① How to pick $\theta_{min}, \theta_{max}, \Delta$? In one dimension, with known $Supp[\theta]$, no problem. But in many dimensions with possibly unknown $Supp[\theta]$, problems.

②  Numerically unstable / inaccurate in the tails.

NO.

③ Curse of Dimensionality



$\Delta = 0.2$ # points in each dim = 5.

points in whole grid $5^2 = 25$.

Imagine 10^5 points in each dimension. 10 dimensions. $10^{5^{10}} = 10^{50}$ not possible

If you want 10^5 total points $\sqrt[10]{10^5} \approx 2$ points in each dimension.

" 10^9 (billion) .. $\sqrt[10]{10^9} \approx 8$ points "

⇒ Grid Sampling impossible if $\dim[\vec{\theta}]$ is large. We need another procedure.

- $P(\theta) = N(\mu_0, \tau^2)$, $P(\sigma^2) = \text{InvGamma}\left(\frac{n_0}{2}, \frac{n_0\sigma_0^2}{2}\right)$

$P(\theta, \sigma^2 | x)$ was a mess due to non-conjugacy. Technically it's "semi-conjugate."

$\Rightarrow P(\theta | x, \sigma^2) = N\left(\frac{\frac{n\bar{x}}{\sigma^2} + \frac{1}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\right)$ & $P(\sigma^2 | x, \theta) = \text{InvGamma}\left(\frac{n_0+n}{2}, \frac{n_0\sigma_0^2 + n\hat{\sigma}^2}{2}\right.$

Imagine the following sampling:

⓪ Let $\theta_0 = 0$ and $\sigma_0^2 = 1$.

calc. using θ_0

1a Draw σ_1^2 from rinvgamma$\left(\frac{n_0+n}{2}, \frac{n_0\sigma_0^2 + n\hat{\sigma}_{\theta_0}^2}{2}\right)$

1b Draw θ_1 from rnorm(,)

2a Draw σ_2^2 from rinvgamma(,)

2b Draw θ_2 from rnorm(,)

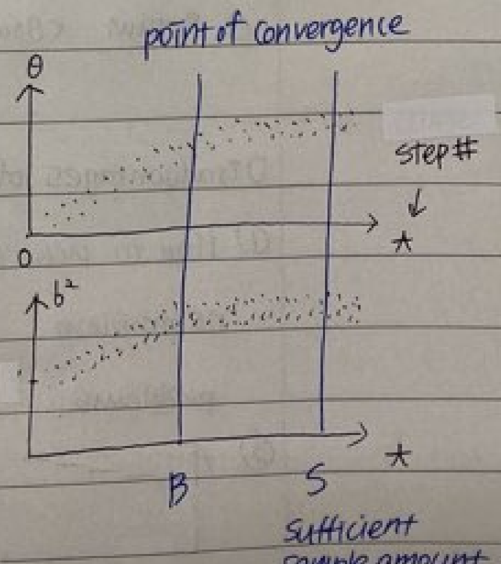
⋮

repeat until "convergence".

$\left\langle \begin{matrix}\theta_0 \\ \sigma_0^2\end{matrix} \right\rangle$, step 1 $\left\langle \begin{matrix}\theta_1 \\ \sigma_1^2\end{matrix} \right\rangle$, step 2 $\left\langle \begin{matrix}\theta_2 \\ \sigma_2^2\end{matrix} \right\rangle$, ...

"B" is called the "burn-in"

need some time to get the point (operating right)



go through all order

go through all p dimension.

Systematic Sweep Gibbs Sampler for $\vec{\theta} = [\theta_1, \theta_2, \ldots, \theta_p]$

⓪ Initialize $\vec{\theta}_0 = [\theta_{0,1}, \theta_{0,2}, \ldots, \theta_{0,p}]$

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

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

⋮

①ⓟ Sample $\theta_{1,p}$ from $P(\theta_p \mid \theta_1 = \theta_{1,1}, \theta_2 = \theta_{1,2}, \ldots, \theta_{p-1} = \theta_{1,p-1}, X)$

② Record $[\theta_{1,1}, \ldots, \theta_{1,p}]$

③ Repeat Step 1, 2 S times.

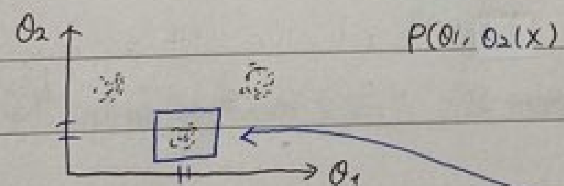
all other dimension.

Reminder: you need all conditional distribution $P(\theta_j \mid \vec{\theta}_{-j})$ $\forall j$.

If you are missing some $P(\theta_j \mid \vec{\theta}_{-j})$, you'll always have $k(\theta_j \mid \vec{\theta}_{-j})$

and step 1_j could be a grid sample step.

Problems with the Gibbs sampler



no mess in that □ region

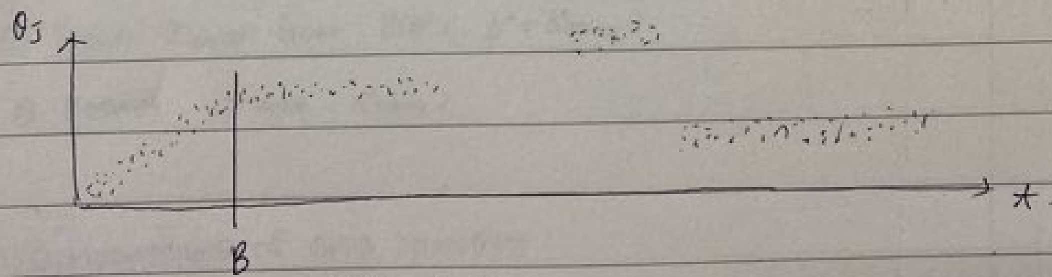
Ⓐ It's possible to not reveal while posterior

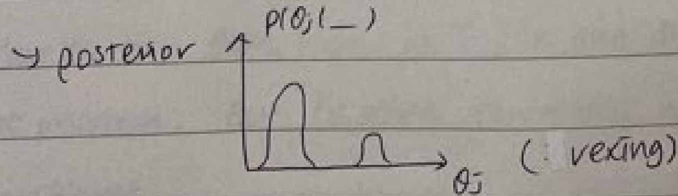
If this is your true posterior,

the Gibbs Sampler will only find

one pocket. e.g.

problem gets worse if p large.





(vexing)