Lecture 20 05/05/2020

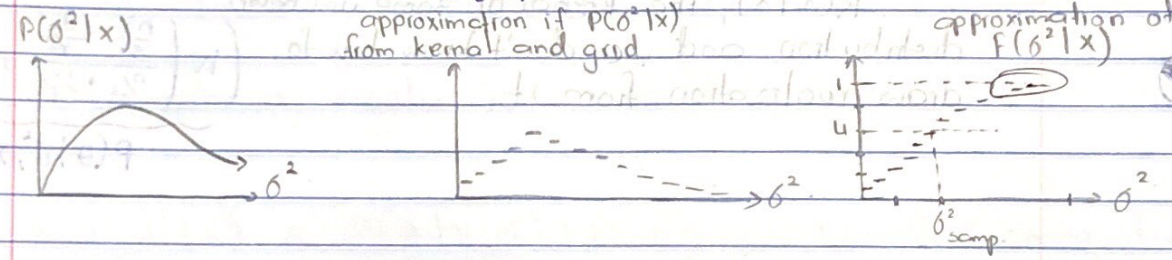
$P(\theta, \sigma^2 | X) \propto P(\theta | \sigma^2, X)\, P(\sigma^2 | X)$

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

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

(I) Create a "grid" by picking σ^2_{min}, σ^2_{max}, $\Delta \to$ grid resolution
$G = [\sigma^2_{min}, \sigma^2_{min} + \Delta, \sigma^2_{min} + 2\Delta, \ldots, \sigma^2_{max}]$

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



(III) Compute the approximate CDF $F(\sigma_0^2 | X) = P(\sigma^2 \leq \sigma_0^2 | X)$
$= \sum_{\{\sigma^2 \leq \sigma_0^2 : \sigma^2 \in G\}} C_{approx}\, k(\sigma^2 | X)$

(IV) Draw u from $U(0,1)$ via runif(0,1) then locate $\sigma^2_{samp} = \min_{\sigma^2 \in G} [F(\sigma^2|X) \geq u]$

Sample from this posterior

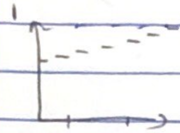
(1) Draw σ^2_{samp} from $k(\sigma^2|X)$ via grid sampling.

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

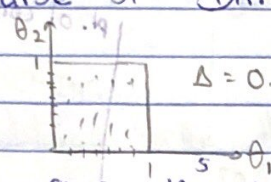
(3) Return $\langle \theta_{samp}, \sigma^2_{samp} \rangle$

Disadvantages of Grid sampling

① How to pick θ_{min}, θ_{max}, Δ? In one dimension, with known Supp[θ], no problem, but in many dimentions with possibly unknown supp[θ], problems.

②  Numerically unstable / inaccurate in the tails.

③ Curse of Dimensionality.



$\Delta = 0.2$ # pts in each dim = 5.
pts in whole grid $5^2 = 25$.

Imagine 10 points in each dimension. 10 dimensions $10^{10} = 10^{10}$ not possible. If you want 10^5 total points $\sqrt[10]{10^5} \approx 2$ points in each dimension.
If you want 10^9 (billion) total points $\sqrt[10]{10^9} \approx 8$ pts in each.

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

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

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

$$\Rightarrow P(\theta \mid x, \sigma^2) = N\left(\frac{\frac{n\bar{x}}{\sigma^2} + \frac{\mu_0}{\tau^2}}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}}\right) \;\&\; P(\sigma^2 \mid x, \theta) =$$

$$\text{InvGamma}\left(\frac{n_0 + n}{2}, \frac{n_0\sigma_0^2 + n\hat{\sigma}^2_{MLE}}{2}\right)$$

Imagine the following sampling:

calc. using θ_0

(0) Let $\theta_0 = 0$ and $\sigma_0^2 = 1$

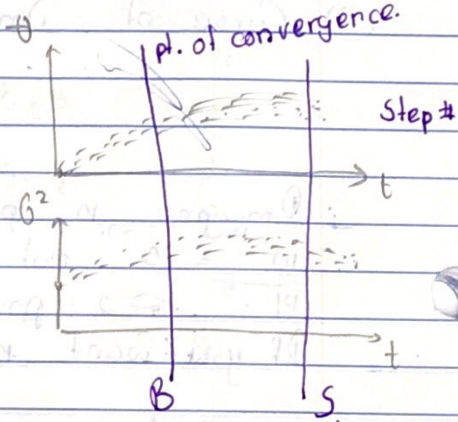
(1a) Draw σ_1^2 from rinvgamma$\left(\frac{n_0+n}{2}, \frac{n_0\sigma_0^2 + n\hat{\sigma}_0^2}{2}\right)$

(1b) Draw θ_1 from rnorm(,)

(2a) Draw σ_2^2 from rinvgamma(,)

(2b) Draw θ_2 from rnorm(,)

⋮



repeat until "convergence"

$$\begin{pmatrix}\theta_0\\ \sigma_0^2\end{pmatrix}, \underset{\text{step 1}}{\begin{pmatrix}\theta_1\\ \sigma_1^2\end{pmatrix}}, \underset{\text{step 2}}{\begin{pmatrix}\theta_2\\ \sigma_2^2\end{pmatrix}}, \dots$$

"B" is called the "burn-in"

need some time to get the point. (operating right)

Sufficient sample amound

---

go through all order — go through all p dimention

Systematic sweep Gibbs sampler for $\vec{\theta} = [\theta_1, \theta_2, \dots, \theta_p]$.

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

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

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

⋮

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

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

③ Repeat step 1,2 s times.

Reminder: You need all conditional distribution

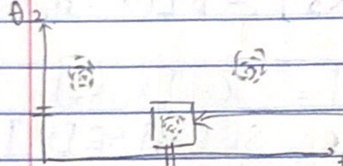
$P(\theta_j | \vec{\theta}_{-j})\ \forall j$

all other dimention

If you are missing some $P(\theta_j | \vec{\theta}_{-j})$, you'll always have $K(\theta_j | \vec{\theta}_{-j})$ and step 1j could be a grid sample step.

Problems with the Gibbs sampler.

Ⓐ It's possible to not reveal while posterior. If this is your true posterior, the Gibbs sampler will only find one pocket. e.g.



no mess in that □ region.

Problem gets worse if p large,